ASPECTS OF BAYESIAN CHANGE-POINT ANALYSIS

BY

ANITA CARINA SCHOEMAN

THESIS

Submitted in fulfilment of the requirements for the degree of

PHILOSOPHIAE DOCTOR

IN

THE FACULTY OF NATURAL AND AGRICULTURAL SCIENCES

DEPARTMENT OF MATHEMATICAL STATISTICS

UNIVERSITY OF THE ORANGE FREE STATE

BLOEMFONTEIN

NOVEMBER 2000

PROMOTER: PROF. DR. P.C.N. GROENEWALD

ACKNOWLEDGEMENTS

I would like to acknowledge the assistance given to me during the preparation of this thesis:

A special note of appreciation towards Prof. P.C.N. Groenewald, to whom I am indepted for his encouragement, guidance, numerous suggestions and patience.

I would also like to thank Mrs. G. van Jaarsveld for typing this thesis so expertly.

Words cannot adequately express my gratitude towards my family for their interest and ongoing support — especially my parents and my aunt. You gave me so many opportunities in my life.

To my husband, Johan, my sincere thanks for all your patience, sacrifices, understanding and support during the completion of this study.

Everything I have and what I am is by the grace of our Lord, who is the Kingdom, the Power and the Glory.

Anita Schoeman

November 2000

CONTENTS

I. INTRODUCTION

II. CHANGES IN THE NORMAL MODEL

CHAPTER 1

INTRODUCTION

1.1 WHAT IS STRUCTURAL CHANGE?

It is generally recognised that a physical entity experiences structural change as it evolves over time, e.g. the society and its economic behaviour changes over time and an economic policy that was once ineffective may become effective.

In many fields of empirical science, theories have been proposed arguing that a behavioral relationship changes over time. Such a change occurs, e.g. in the demand and supply characteristics of a product over its life cycle. In the new product stage, demand may have high income elasticity and low price elasticity, but in the standardized stage the price elasticity increases while income elasticity decreases. Within each stage, however, the demand relationship may be stable enough to be described by regression with constant parameters. In such a case, a switching regression model may be appropriately used.

On the other hand, data with considerable noise may suggest the use of a regression model in which parameters follow random walk patterns or some steady and systematic changes over all the sample periods. There are many ways in which parameters of regression models are assumed to change.

An example of structural change is pointed out by Nordhaus and Samuelson (1985). In the late 1940's many Keynesian economists emphasized the role of fiscal policy, i.e. changes in taxes and government expenditures, as the key to controlling the business cycle and these advocates of fiscal policy tended to slight the role of money. In the 1948 edition of a leading college textbook on economics this view was reflected with the words: "Today few economists regard Federal Reserve monetary policy as a panacea for controlling the business cycle". The 1985 edition of the textbook, thirty-seven years later, states: "Money is the most powerful and useful tool that macroeconomic policy makers have at their disposal" and "in the U.S. today, the central bank (Federal Reserve system) is the most important factor in the making of macroeconomic policy".

Structural change may be defined as a change in one or more parameters of the model in question. In the example of the effectiveness of monetary policy above, we may observe that the once significant interest-rate variable in the investment function or in the demand-for-money question has changed to insignificant.

Given data then, we need to make inferences on join points (or the points at which parameters change) if they are unknown and on parameters within each regime. Inferences on both join points (change-points) and parameter changes are important, because they will provide evidence on which a particular theory is correct or not. Furthermore, poor forecasts of econometrics models may result if the models do not account for structural shift.

Bayesian procedures are applied to solve inferential problems of structural change. Among the various methodological approaches within Bayesian inference, emphasis is put on the analysis of the posterior distribution itself, since the posterior distribution can be used for conducting hypothesis testing as well as obtaining a point estimate.

Diagrammatically, the statistical inference process can be described as follows:

MATHEMATICAL MODEL

$f(\boldsymbol{X}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Omega$

$\downarrow$

Prior Information

$\Pi(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Omega$

$\downarrow$

Sample Data

$\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$

$\downarrow$

Posterior Distribution

$\Pi(\boldsymbol{\theta}|\boldsymbol{X}), \boldsymbol{\theta} \in \Omega$

$\downarrow$

Posterior Inferences

$\swarrow \qquad \downarrow \qquad \searrow$

Estimation Test of hypothesis Prediction

1.2 HISTORY OF STRUCTURAL CHANGE IN STATISTICS

For some 40 years beginning with Page, statisticians have been studying and developing probability models that account for a changing distribution of random variables. The goal has been to develop inference procedures that will estimate the parameters, test hypothesis about the parameters and forecast future observations.

Page (1954, 1955, 1957) found methods for detecting change in the distribution of a sequence of independent random variables and these tests are based on cumulative sums called cusums. Page's objective was to find efficient methods for quality control and his approach was non-parametric. Quandt (1958) devoted attention to the problem of fitting lines or curves to data which suggest abrupt changes in parameter values from one range of the independent variable to another. He suggested tests of the hypothesis that a sudden change in behaviour has occurred at an estimated join point and described maximum likelihood estimation procedures for the model parameters and the join point.

In the next decade there was a burst of activity which was based mostly on parametric statistical procedures. Chernoff and Zacks (1964) and Kander and Zacks (1966) studied sequences of normal random variables and found a Bayesian test to detect a change in the mean. Other contributions have been made by Sprent (1961), Robison (1964), Hudson (1966) and Gardner (1969). Bhattacharya and Johnson (1968) determined the sampling properties of the tests found by Chernoff and Zacks (1964) and Kander and Zacks (1966).

During the late 1960's and early 1970's, Hinkley (1969, 1971) studied structural change in sequences of random variables and in linear regression models, and he employed a non-Bayesian parametric method. For example, to detect change, the likelihood ratio test was used and maximum likelihood was used for estimating the parameters of binomial and normal sequences. The asymptotic properties of these procedures were also studied by him. His work appears to have stimulated the study of structural change.

In one of the earliest Bayesian contributions, Bacon and Watts (1971) introduced the transition function to model "smooth" changes in the regression function. Prior to the Bacon and Watts study, the change was represented by a shift point $m = 1, 2, \ldots, n-1$, where n is the number of observations. That is, suppose that the first m random variables $X_1, X_2, \ldots, X_m$ have a common distribution and the remaining $X_{m+1}, \ldots, X_n$ have another (distinct) common distribution, where $m = 1, 2, \ldots, n-1$. Thus the change-point indexes where or when the change occurs. The transition function allows one to model structural change, allowing the number of the change (either abrupt or smooth) to be incorporated into the model. Bacon and Watts (1971) found exact small-sample inferences for the parameters of the transition function and their method was to be adopted in later research.

Motivated by Bacon and Watts (1971), the decade of the 1970's was a time of many Bayesian contributions. Structural change of univariate and multivariate linear models received most of the attention. Ferreira (1975), Holbert and Broemeling (1977) and Chin Choy and Broemeling (1980) all studied two-phase regression problems. The observations are the Y_i's and the X_i's are the corresponding values of the independent variable. Assuming a normal distribution for the errors, two problems are solved. First, assuming that a change ($m \leq n-1$) has (or will) occurred, the parameters are estimated by finding their marginal posterior distribution. Secondly, detecting a change in the parameters is examined by testing the hypothesis $1 \leq m \leq n-1$. The test is based on the marginal posterior distribution of the change-point. Based on the Bayesian approach, Holbert (1973) and Holbert and Broemeling (1977) assigned a uniform proper prior distribution to the change-point m and an improper prior to the unknown regression parameters. They derived the posterior distribution of m for a number of cases. Ferreira (1975) also assigned a vague-type prior distribution to the unknown regression parameters and assigned three different prior distributions for the change-point m and the regression parameters. The Chin Choy and Broemeling (1980) paper gives a Bayesian way to detect a future shift in the parameters of a general linear model and the test is contrasted to a sequential test of Smith (1975).

Another important contribution to structural change was the intervention analysis of Box and Tiao (1975), who found a way to study changes in the mean of a time series represented by an ARMA process. They represented the change by a transfer function which allows a very general class of intervention effects, and the statistical analysis is based on the time series techniques pioneered by Box and Jenkins (1970).

In the 1980's, Menzefricke (1981) examined a changing linear model, with a change in the precision parameter at an unknown change-point, from the Bayesian viewpoint. Hsu (1982) examined a linear model that exhibited changes in regression parameters and precision at an unknown change-point. He assumed that the observations follow an exponential power distribution and used numerical integration to evaluate the posterior distributions of the regression and precision parameters. Salazar (1980, 1982) considered changes in the multivariate linear model using a change-point parameter. Moen (1983) developed a detailed analysis of the multivariate linear model and Tsurumi, *et al.* (1984) developed a gradual switching multivariate regression model with stochastic constraints.

Moen, Salazar and Broemeling (1985) generalized the work of Chin Choy and Broemeling (1980), who investigated the change in the regression parameters of univariate linear models. They studied the case in which the assumption was made that there has been a single shift in the regression matrix of a multivariate linear model at some unknown point m. Broemeling, *et al.* (1987) discussed Bayesian inference of two-phase linear multiple regression models, presenting Bayesian inference of two-phase multivariate linear regression models. They also reviewed the Bayesian analysis of multivariate regression models, first with natural conjugate priors and then with diffuse priors. They also made a posterior analysis of a join point and parameter shift.

Till the 1990's, very few studies on changing linear models consider the problem of a change in the precision parameter at an unknown change-point. Ng (1990) used the Bayesian approach to examine the linear model in which both the mean and the precision change exactly once at an unknown point in time. Furthermore he also generalized the results of Menzefricke (1981) and obtained the Bayesian predictive distribution of k future observations in a closed form, as a mixture of multivariate t. Wang and Lee (1993) considered a Bayesian approach to detect a change-point in the intercept of simple linear regression. The Jeffrey's non-informative prior is employed and compared with the uniform prior in Bayesian analysis.

Most of the mentioned analyses are under the assumption of exactly one change-point. Complications arise due to the changing dimensions of the parameter space if the number of change-points are unknown. Barry and Hartigan (1992) propose a product partition model for multiple change-points. Groenewald (1993) considered a general Bayes procedure for the examination of possible change-points in the linear model. Provision is made for the possibility of no, one or more than one change-point under the assumption of homogeneity of error variance. In linear regression, certain components which may be the cause of a change-point, can be examined. His results are in terms of posterior probabilities over a class of conjugate priors.

1.3 BAYES FACTORS

The Bayesian approach to hypothesis testing was developed by Jeffreys (1935, 1961) as a major part of his program for scientific inference. The centerpiece was a number, now called the Bayes factor, which is the posterior odds of the null hypothesis when the prior probability on the null is one-half. Jeffreys was concerned with the comparison of predictions made by two competing scientific theories. In his approach, statistical models are introduced to represent the probability of the data according to each of the two theories and Bayes' theorem is used to compute the posterior probability that one of the theories is correct.

According to Kass and Raftery (1993), often lost from the controversy however, are the practical aspects of the Bayesian methods: how conclusions may be drawn from them, how they can provide answers when non-Bayesian methods are hard to construct, what their strengths and limitations are.

Kass and Raftery (1993) begin with data D assumed to have arisen under one of the two hypotheses H_1 and H_2 according to a probability density $pr(D|H_1)$ and $pr(D|H_2)$. Given a priori probabilities $pr(H_1)$ and $pr(H_2) = 1 - pr(H_1)$, the data produce aposteriori probabilities $pr(H_1|D)$ and $pr(H_2|D) = 1 - pr(H_1|D)$. Since any prior opinion gets transformed to a posterior opinion through consideration of the data, the transformation itself represents the evidence provided by the data. In fact, the same transformation is used to obtain the posterior probability, regardless of the prior probability. Once they convert to the odds scale (odds = probability/(1 - probability)), the transformation takes a simple form. From Bayes' Theorem they obtain

$$pr(H_k|D) = \frac{pr(D|H_k)pr(H_k)}{pr(D|H_1)pr(H_1) + pr(D|H_2)pr(H_2)}, \quad (k = 1, 2)$$

so that

$$\frac{pr(H_1|D)}{pr(H_2|D)} = \frac{pr(D|H_1)pr(H_1)}{pr(D|H_2)pr(H_2)},$$

and the transformation is simply multiplication by

$$B_{12} = \frac{pr(D|H_1)}{pr(D|H_2)}, \tag{1.1}$$

which is the Bayes factor.

Thus, in words, posterior odds = Bayes factor $\times$ prior odds, and the Bayes factor is the ratio of the posterior odds of H_1 to its prior odds, regardless of the value of the prior odds. (The terminology is apparently due to Good (1983) who attributes the method to Turing in addition to, and independently of, Jeffreys at about the same time.) When the hypotheses H_1 and H_2 are equally probable a priori so that $pr(H_1) = pr(H_2) = \frac{1}{2}$, the Bayes factor is equal to the posterior odds in favor of H_1. The two hypotheses may well not be equally likely a priori, however.

In the simplest case, when the two hypotheses are single distributions with no free parameters, B_{12} is the likelihood ratio. In other cases, when there are unknown parameters under either or both of the hypotheses, the Bayes factor is still given by (1.1) and in a sense it continues to have the form of a likelihood ratio. Then the densities $pr(D|H_k)$ $(k = 1, 2)$ are obtained by integrating over the parameter space, so that in equation (1.1),

$$pr(D|H_k) = \int pr(D|\theta_k, H_k)\Pi(\theta_k|H_k)d\theta_k, \tag{1.2}$$

where θ_k is the parameter under H_k, $\Pi(\theta_k|H_k)$ is its prior density and $pr(D|\theta_k, H_k)$ is the probability density of D given the value of θ_k, or the likelihood function of θ (θ_k may be a vector with dimension d_k).

Note that the prior distributions $\Pi(\theta_k|H_k)$, $k = 1, 2$ are necessary, although considered both good and bad. Good, because it is a way of including other information about the values of the parameters. Bad, because these prior densities may be hard to set when there is no such information.

The quantity $pr(D|H_k)$ in (1.2) is the marginal probability of the data, since it is obtained by integrating the joint density of (D, θ_k) given D over θ_k. It is also the predictive probability of the data, i.e. the probability of seeing the data that actually were observed, calculated before any data became available. It is also sometimes called a marginal likelihood or an integrated likelihood. Note that, as in computing the likelihood ratio statistic, but unlike in some other applications of likelihood, all constants appearing in the definition of the likelihood $pr(D|\theta_k, H_k)$ must be retained when computing B_{12} (B_{12} is closely related to the likelihood ratio statistic, in which the parameters θ_k are eliminated by maximization rather than by integration).

In the Bayesian approach to model selection or hypothesis testing with models or hypotheses of differing dimensions, it is typically not possible to utilize standard non-informative prior distributions which has led Bayesians to use conventional proper prior distributions or crude approximations to Bayes factors (e.g. the Bayesian information criterion (BIC) developed by Schwarz (1978)). So Berger and Pericchi (1995) introduced a new criterion called the intrinsic Bayes factor, which is fully automatic in the sense of requiring only standard non-informative priors for its computation, and yet seems to correspond to very reasonable actual Bayes factors. The criterion can be used for nested or non-nested models, and for multiple model comparison and prediction. From another perspective, the development suggests a general definition of a "reference prior" for model comparison.

Berger and Pericchi (1995) proposed a completely general method of testing and model selection that will be argued to be essentially equivalent to the conventional proper prior approach (the conventional prior approach of Jeffreys (1961)), but without the need to determine a reasonable proper prior. Unlike the BIC criterion, which starts with an asymptotic approximation to the Bayes factor and then simply ignores the term involving the prior, Berger and Pericchi's (1995) approach can be thought of as automatically "correcting" BIC by inserting a reasonable value for the term that BIC ignores.

So Berger and Pericchi (1995) considered models $M_1, M_2, \ldots, M_p$, with the data $\boldsymbol{X}$ having density $f_i(\boldsymbol{x}|\boldsymbol{\theta}_i)$ under model M_i. The parameter vectors $\boldsymbol{\theta}_i$ are unknown and are of dimension k_i. Bayesian model selection proceeds by selecting prior distributions $\Pi_i(\boldsymbol{\theta}_i)$ for the parameters of each model, together with prior probabilities p_i of each model being true. The posterior probability that M_i is true is then

$$P(M_i|\boldsymbol{x}) = \left(\sum_{j=1}^{p} \frac{p_j}{p_i} B_{ji}\right)^{-1}, \tag{1.3}$$

where B_{ji}, the Bayes factor of M_j to M_i, is defined by

$$B_{ji} = \frac{m_j(\boldsymbol{x})}{m_i(\boldsymbol{x})} = \frac{\int f_j(\boldsymbol{x}|\boldsymbol{\theta}_j)\Pi_j(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j}{\int f_i(\boldsymbol{x}|\boldsymbol{\theta}_i)\Pi_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i} \tag{1.4}$$

where $m_j(\boldsymbol{x})$ is the marginal or predictive density of $\boldsymbol{X}$ under M_j.

Computing B_{ji} requires specification of $\Pi_i(\boldsymbol{\theta}_i)$ and $\Pi_j(\boldsymbol{\theta}_j)$. Often in Bayesian analysis, one can effectively use non-informative (or default) priors $\Pi_i^N(\boldsymbol{\theta}_i)$. Three common choices are the "uniform" prior $\Pi_i^U(\boldsymbol{\theta}_i) \propto 1$, the Jeffreys prior $\Pi_i^J(\boldsymbol{\theta}_i) \propto (\det(I_i(\boldsymbol{\theta}_i)))^{\frac{1}{2}}$, where $I_i(\boldsymbol{\theta}_i)$ is the expected Fisher information matrix corresponding to M_i, and the reference prior $\Pi_i^R(\boldsymbol{\theta}_i)$, definitions of which can be found in Bernardo (1979) and Berger and Bernardo (1992). Using any of the Π_i^N in (1.4) would yield

$$B_{ji}^N = \frac{m_j^N(\boldsymbol{x})}{m_i^N(\boldsymbol{x})} = \frac{\int f_j(\boldsymbol{x}|\boldsymbol{\theta}_j)\Pi_j^N(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j}{\int f_i(\boldsymbol{x}|\boldsymbol{\theta}_i)\Pi_i^N(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i}. \tag{1.5}$$

The difficulty with this solution is that the Π_i^N are typically improper and hence defined only up to arbitrary constants c_i. Hence B_{ji}^N is defined only up to $\frac{c_j}{c_i}$, which is itself arbitrary.

A common solution to this problem is to use part of the data as a training sample. Let $\boldsymbol{x}(\ell)$

denote the part of the data to be so used and $\boldsymbol{x}(-\ell)$ represent the remainder of the data. The idea is that $\boldsymbol{x}(\ell)$ will be used to convert the $\Pi_i^N(\boldsymbol{\theta}_i)$ to proper posterior distributions

$$\Pi_i^N(\boldsymbol{\theta}_i|\boldsymbol{x}(\ell)) = f_i(\boldsymbol{x}(\ell)|\boldsymbol{\theta}_i)\Pi_i^N(\boldsymbol{\theta}_i)/m_i^N(\boldsymbol{x}(\ell)), \tag{1.6}$$

where $f_i(\boldsymbol{x}(\ell)|\boldsymbol{\theta}_i)$ is the marginal density of $\boldsymbol{X}(\ell)$ under M_i and

$$m_i^N(\boldsymbol{x}(\ell)) = \int f_i(\boldsymbol{x}(\ell)|\boldsymbol{\theta}_i)\Pi_i^N(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i. \tag{1.7}$$

The idea is to then compute the Bayes factors with the remainder of the data, $\boldsymbol{x}(-\ell)$, using the $\Pi_i^N(\boldsymbol{\theta}_i|\boldsymbol{x}(\ell))$ as priors. The result is easily shown to be

$$\begin{aligned} B_{ji}(\ell) &= \frac{\int f_j(\boldsymbol{x}(-\ell)|\boldsymbol{\theta}_j, \boldsymbol{x}(\ell))\Pi_j^N(\boldsymbol{\theta}_j|\boldsymbol{x}(\ell))d\boldsymbol{\theta}_j}{\int f_i(\boldsymbol{x}(-\ell)|\boldsymbol{\theta}_i, \boldsymbol{x}(\ell))\prod_i^N(\boldsymbol{\theta}_i|\boldsymbol{x}(\ell))d\boldsymbol{\theta}_i} \\ &= B_{ji}^N \cdot B_{ij}^N(\boldsymbol{x}(\ell)), \end{aligned} \tag{1.8}$$

where

$$B_{ij}^N(\boldsymbol{x}(\ell)) = \frac{m_i^N(\boldsymbol{x}(\ell))}{m_j^N(\boldsymbol{x}(\ell))}. \tag{1.9}$$

Clearly (1.8) removes the arbitrariness in the choice of constant multiples of the Π_i^N: the arbitrary ratio $\frac{c_j}{c_i}$ that multiplies B_{ji}^N would be cancelled by the ratio $\frac{c_i}{c_j}$ that would then multiply $B_{ij}^N(\boldsymbol{x}(\ell))$. Note that, while the first motivating expression in (1.8) seems to require the conditional distribution of $\boldsymbol{x}(-\ell)$ given $\boldsymbol{x}(\ell)$, the second expression only utilizes the typically much simpler marginal densities of $\boldsymbol{x}(\ell)$.

The above use of a training sample makes sense only if the $m_i^N(\boldsymbol{x}(\ell))$ in (1.7) are finite. This is formalized in a definition.

Definition: A training sample, $\boldsymbol{x}(\ell)$, will be called proper if $m_i^N(\boldsymbol{x}(\ell)) < \infty$ for all M_i, and minimal if it is proper and no subset is proper.

The training sample idea has been informally used many times. More formal developments of the idea can be found in Lempers (1971), Atkinson (1978), Geisser and Eddy (1979), Spiegelhalter and Smith (1982) and Gelfand, Dey and Chang (1992), although not all these works utilize the idea with ordinary Bayes factors. Other references and the general asymptotic behavior of training sample methods can be found in Gelfand and Dey (1993). Aitkin (1991) can also be considered to be a training sample method; it takes the entire sample $\boldsymbol{x}$ as a training sample to obtain $\Pi_i^N(\boldsymbol{\theta}_i|\boldsymbol{x})$ and then uses this as the prior in (1.8) to compute the Bayes factor. This double use of the data is, of course, not consistent with usual Bayesian logic and the method violates the asymptotic criterion rather severely.

Independently of the work of Berger and Pericchi (1995), De Vos (1993) has proposed a training sample method for linear models that is similar to their proposal.

For a given data set $\boldsymbol{x}$, there will typically be many minimal training samples. Let

$$\chi_T = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(L)\}$$

denote the set of all minimal training samples, $\boldsymbol{x}(\ell)$. Clearly the Bayes factor $B_{21}(\ell)$, as defined in (1.8), will depend on choice of the minimal training sample. To eliminate this dependence and increase stability, a natural idea is to average the $B_{21}(\ell)$ over all $\boldsymbol{x}(\ell) \in \chi_T$. This average can be done either arithmetically or geometrically, leading to the arithmetic intrinsic Bayes factor (AIBF) and geometric intrinsic Bayes factor (GIBF) defined respectively by

$$B_{21}^{AI} = \frac{1}{L}\sum_{\ell=1}^{L} B_{21}(\ell) = B_{21}^N \cdot \frac{1}{L}\sum_{\ell=1}^{L} B_{12}^N(\boldsymbol{x}(\ell)), \tag{1.10}$$

$$B_{21}^{GI} = \left(\prod_{\ell=1}^{L} B_{21}(\ell)\right)^{\frac{1}{L}} = B_{21}^N \cdot \left(\prod_{\ell=1}^{L} B_{12}^N(\boldsymbol{x}(\ell))\right)^{\frac{1}{L}} \tag{1.11}$$

where the $B_{12}^N(\boldsymbol{x}(\ell))$ are defined in (1.9). Note that $B_{21}^{GI} \leq B_{21}^{AI}$, since the geometric mean is less than or equal to the arithmetic mean (for positive variables). Thus B_{21}^{GI} will favor the simpler (nested) model to a greater extent than will B_{21}^{AI}. Also note that $B_{12}^{AI} = \frac{1}{B_{21}^{AI}}$ and not as in (1.10) with the indices reversed. The asymmetry arises because of M_1 being nested within M_2. For B_{21}^{GI} there is no problem, as reversing the indices in (1.11) clearly results in $\frac{1}{B_{21}^{GI}}$. Berger and Pericchi (1998) also define a median intrinsic Bayes factor (MIBF).

O'Hagan (1995) advocated the fractional Bayes factor (FBF), a new variant of the partial Bayes factor, on grounds of consistency, simplicity, robustness and coherence. In general, the partial Bayes factor divides the data into two parts, $\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{z})$. The first part $\boldsymbol{y}$ is used as a training sample to provide information about $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ and the second part $\boldsymbol{z}$ is used for model comparison. To avoid the arbitrariness of choosing a particular $\boldsymbol{y}$ or having to consider all possible subsets of a given size, O'Hagan defined a simplified form of the partial Bayes factor as follows. Let $b = \frac{m}{n}$. If both m and n are large, the likelihood $f_i(\boldsymbol{y}|\boldsymbol{\theta}_i)$ based only on the training sample $\boldsymbol{y}$ will be approximate to the full likelihood $f_i(\boldsymbol{x}|\boldsymbol{\theta}_i)$ raised to the power b. The FBF is then

$$B^F = \frac{m_1(b)}{m_2(b)} \tag{1.12}$$

where

$$m_i(b) = \frac{m_i}{m_i^b} = \frac{\int \Pi_i(\boldsymbol{\theta}_i) f_i(\boldsymbol{x}|\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i}{\int \Pi_i(\boldsymbol{\theta}_i) f_i(\boldsymbol{x}|\boldsymbol{\theta}_i)^b d\boldsymbol{\theta}_i}. \tag{1.13}$$

If $\Pi_i(\boldsymbol{\theta}_i) = c_i h_i(\boldsymbol{\theta}_i)$, h_i a function whose integral over the $\boldsymbol{\theta}_i$-space converges, the indeterminate constant c_i cancels out, leaving

$$m_i(b) = \frac{\int h_i(\boldsymbol{\theta}_i) f_i(\boldsymbol{x}|\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i}{\int h_i(\boldsymbol{\theta}_i) f_i(\boldsymbol{x}|\boldsymbol{\theta}_i)^b d\boldsymbol{\theta}_i}. \tag{1.14}$$

So O'Hagan (1995) proposes using a fractional part of the entire likelihood, $[f(\boldsymbol{x}|\boldsymbol{\theta})]^b$, instead of a training sample. This tends to produce a more stable answer than use of a particular training sample, but will fail the asymptotic criterion, unless $b \propto \frac{1}{n}$ as the sample size n grows. The behavior of fractional Bayes factors for such b is well worth study, although it appears to be quite difficult to decide on a specific choice of b. O'Hagan suggested $b = \frac{m}{n}$, where m is the minimal sample size (when it is unique). Other suggestions are $\frac{1}{\sqrt{n}}$ and $\frac{\log(n)}{n}$.

Another approach is the imaginary training sample device of Spiegelhalter and Smith (1982). The basic idea, a variation on a theme of Good (1947), is to imagine that a data set is available which: (1) involves the smallest possible sample size permitting a comparison of M_0 and M_1 and (2) provides maximum possible support for M_0. On the strength of (2), such a data set would lead to $B_{01} > 1$ (data have provided evidence in favour of M_0). But on the basis of (1), one could only have $B_{01} = 1 + \epsilon$, where $\epsilon > 0$ is rather small (that is, such evidence as exists must be very weak, since the data set is necessarily very small). Suppose that E_0, E_1 denote the design matrices for regression models M_0, M_1 occurring in the "thought experiment" generating the imaginary training sample, which leads to an F-statistic value $F = 0$ (giving maximum support to M_0). Spiegelhalter and Smith (1982) deduce that $\frac{c_0}{c_1} = \left[\frac{|E_1'E_1|}{|E_0'E_0|}\right]^{-\frac{1}{2}}$. The device of an imaginary training sample therefore provides a general solution to the problem of assigning a value to the hitherto undefined ratio of constants. Booth and Smith (1982) applied this to change-point analysis.

In this study we will eliminate hyperparameters by integrating out or by introducing limits where possible or by determining partial Bayes factors. Some of the disadvantages of the Intrinsic Bayes factor (IBF) is the definition of a minimal sample in change-point analysis, incoherency, especially with the Arithmetic Intrinsic Bayes factor (AIBF), and that more computer intensive calculations is required. Berger's solution to the coherency problem is to use weighted averages.

The fractional Bayes factor looks better on grounds of coherency, robustness, simplicity and consistency and the problem of a minimal sample is not present, except that the posterior probabilities can sometimes be very sensitive to the choice of the training fraction b, as will be shown in some of the examples.

Another version of the partial Bayes factor is the "Posterior Bayes Factor" (PBF), suggested by Aitkin (1991). Here the marginal likelihood of the data is replaced by the expected value, or posterior mean of the likelihood, with expectation taken with respect to the posterior distribution.

Thus the Bayes factor in favour of model M_1 when compared to model M_2, is given by

$$B_{12}^{p} = \frac{\bar{\boldsymbol{L}}_1}{\bar{\boldsymbol{L}}_2}, \tag{1.15}$$

where

$$\bar{\boldsymbol{L}}_j = \int f_j(\boldsymbol{x}|\boldsymbol{\theta}_j)\Pi(\boldsymbol{\theta}_j|\boldsymbol{x})d\boldsymbol{\theta}_j \tag{1.16}$$

$$= \frac{\int [f(\boldsymbol{x}|\boldsymbol{\theta}_j)]^2\Pi(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j}{\int f(\boldsymbol{x}|\boldsymbol{\theta}_j)\Pi(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j} \tag{1.17}$$

One of the main criticisms against the PBF is the reuse of the data (see discussion of Aitkin's paper), once for estimation, and then for model comparisons. We will not make use of this method further on, except in the illustrative example.

Illustrative Example 1

To illustrate and compare the partial Bayes factors mentioned in the previous section, we will consider the simplest example of a change-point and calculate the following five partial Bayes factors:

(i) Improper priors, (ii) Imaginary training sample, (iii) Fractional B.F., (iv) Intrinsic BF and (v) Posterior BF.

Let $X_1, X_2, \ldots, X_n$ be a sequence of normal random variables with known variance $\sigma^2 = 1$. Under model $M_0 : X_i \sim N(\mu_0, 1)$, $i = 1, \ldots, n$, and under model

$$M_k : \begin{cases} X_i \sim N(\mu_1, 1); & i = 1, \ldots, k \\ X_i \sim N(\mu_2, 1); & i = k+1, \ldots, n, \end{cases} \tag{1.18}$$

where k is fixed.

We assume the vague improper prior, $\Pi(\mu_j) \propto 1, \quad j = 0, 1, 2$ throughout the example.

The marginal likelihood under M_0 is then

$$m_0 = n^{-\frac{1}{2}} (2\pi)^{-\frac{n-1}{2}} e^{-\frac{1}{2}S^2}, \tag{1.19}$$

and under M_k,

$$m_k = [k(n-k)]^{-\frac{1}{2}} (2\pi)^{-\frac{n-2}{2}} e^{-\frac{1}{2}(S_1^2 + S_2^2)}, \tag{1.20}$$

where

$$S^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2, \quad S_1^2 = \sum_{i=1}^{k} (x_i - \overline{x}_1)^2, \quad S_2^2 = \sum_{i=k+1}^{n} (x_i - \overline{x}_2)^2.$$

(i) Just using the improper prior, the Bayes factor is

$$B_{0k} = \frac{m_0}{m_k} = \left[\frac{k(n-k)}{n}\right]^{\frac{1}{2}} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}\frac{k(n-k)}{n}(\overline{x}_1 - \overline{x}_2)^2}. \tag{1.21}$$

(ii) For the imaginary training sample method, the minimal sample is $n = 2$ with $k = 1$, while perfect support for M_0 means that $x_1 = x_2$. In this case the Bayes factor should be approximately one, but according to (1.21) we have $B_{0k} = (2\pi)^{-\frac{1}{2}}$, so let

$$\begin{aligned} B_{0k}^I &= B_{0k}(2\pi)^{\frac{1}{2}} \\ &= \left[\frac{k(n-k)}{n}\right]^{\frac{1}{2}} e^{-\frac{1}{2}\frac{k(n-k)}{n}(\overline{x}_1 - \overline{x}_2)^2}. \end{aligned} \tag{1.22}$$

(iii) The fractional Bayes factor is defined as

$$B_{0k}^{F}=\frac{m_0}{m_0^b}\cdot\frac{m_k^b}{m_k}, \tag{1.23}$$

with

$$m_0^b=\int[f(\boldsymbol{x}|\mu_0)]^b\Pi(\mu_0)d\mu_0$$

and similarly for m_k^b. For $b=\frac{2}{n}$ (the minimal sample size over n), we get

$$B_{0k}^{F}=\left(\frac{n}{2}\right)^{\frac{1}{2}}e^{-\frac{(n-2)k(n-k)}{2n^2}(\overline{x}_1-\overline{x}_2)^2}. \tag{1.24}$$

(iv) For the arithmetic intrinsic Bayes factor we have

$$B_{0k}^{AI}=B_{0k}\overline{B}_{ko}(\ell),$$

where

$$\overline{B}_{k0}(\ell)=\frac{1}{L}\sum_{j=1}^{L}B_{k0j}(\ell),\quad \ell=2 \tag{1.25}$$

and $B_{k0j}(\ell)$ is the Bayes factor calculated from two observations, x_{1j} before or at k, and x_{2j} after k, and $j=1,\ldots,L$ indexes all possible pairs (x_{1j},x_{2j}). Then

$$B_{k0j}(\ell)=(2\pi)^{\frac{1}{2}}e^{\frac{1}{2}S_j^2(\ell)}, \tag{1.26}$$

where

$$S_j^2(\ell)=\frac{1}{2}(x_{1j}-x_{2j})^2,\quad j=1,\ldots,L. \tag{1.27}$$

Then the AIBF is

$$B_{0k}^{AI}=\left[\frac{2k(n-k)}{n}\right]^{\frac{1}{2}}e^{-\frac{k(n-k)}{2n}(\overline{x}_1-\overline{x}_2)^2}\cdot\frac{1}{L}\sum_{j=1}^{L}e^{\frac{1}{2}S_j^2(\ell)}. \tag{1.28}$$

Similarly for the Geometric and Median IBF's.

(v) The Posterior Bayes factor follows from equations (1.15) to (1.17) and is given by

$$B^P_{0k} = e^{-\frac{k(n-k)}{2n}(\overline{x}_1-\overline{x}_2)^2}. \tag{1.29}$$

All these Bayes factors are based on the quantity

$$d^2 = \frac{k(n-k)}{n}(\overline{x}_1 - \overline{x}_2)^2, \tag{1.30}$$

and when the data supports model M_0 perfectly, $d^2 = 0$ and we would expect the Bayes factor to be large, increasing with n.

When $d^2 = 0$, we have the following:

$$\max B_{0k} = \left[\frac{k(n-k)}{2n\pi}\right]^{\frac{1}{2}},$$

$$\max B^I_{0k} = \left[\frac{k(n-k)}{n}\right]^{\frac{1}{2}},$$

$$\max B^F_{0k} = \left(\frac{n}{2}\right)^{\frac{1}{2}},$$

$$\max B^{AI}_{0k} = \left[\frac{2k(n-k)}{n}\right]^{\frac{1}{2}} \frac{1}{L}\sum_{j=1}^{L} e^{\frac{1}{2}S^2_j(\ell)}$$

and

$$\max B^P_{0k} = 1. \tag{1.31}$$

B^P_{0k} is always smaller or equal to one and obviously unrealistic. Further, $B^{AI}_{0k} > B^I_{0k}$ unless all observations in the sequence are equal, and $B^I_{0k} > B_{0k}$. Also notice that the maximum value of B^F_{0k} does not depend on the position of the change-point, while B^{AI}_{0k} still depends on the variation between observations.

From a frequentist viewpoint, it is interesting to examine the long-term properties of these Bayes factors. For d given in (1.30), the expectation under model M_k is

$$E\left[e^{-\frac{1}{2}d^2}\right] = \frac{1}{\sqrt{2}}e^{-\frac{1}{4}\delta^2}, \tag{1.32}$$

where $\delta = \mu_1 - \mu_2$.

The expected values of the Bayes factors are then as follows:

$$E[B_{0k}] = \left[\frac{k(n-k)}{4\pi n}\right]^{\frac{1}{2}} \pi^{-1}e^{-\frac{1}{4}\delta^2},$$

$$E[B_{0k}^I] = \sqrt{2\pi}E[B_{0k}],$$

$$E[B_{0k}^F] = \left(\frac{n}{2}\right)^{\frac{1}{2}} e^{-\frac{n-2}{4n}\delta^2}.$$

and

$$E[B_{0k}^P] = \frac{1}{\sqrt{\pi}}e^{-\frac{1}{4}\delta^2}. \tag{1.33}$$

The expectation of B_{0k}^{AI} is too complex, but is larger than that of B_{0k}^I.

Figure 1.1 shows the four Bayes factor expectations from (1.33) as a function of $\delta = \mu_1 - \mu_2$ for $n = 100$ and $k = 50$. The values are surprisingly low for $\delta = 0$ with the maximum of $B_{0k}^F = 5.025$. On the other hand, for $\delta > 0$, the Bayes factors decrease very slowly, showing a difficulty to discriminate between the two models. The ordinary Bayes factor and Posterior Bayes factor are obviously too low for δ close to zero.

Illustrative example 2

When σ^2 is unknown, the partial Bayes factors give intuitively better results. Consider the same two models as before, with vague prior, $\Pi(\mu_j, \sigma^2) \propto \frac{1}{\sigma^2}, \quad j = 0, 1, 2.$

(i) The ordinary Bayes factor is

$$B_{0k} = \pi^{-\frac{1}{2}} \left[\frac{k(n-k)}{n} \right]^{\frac{1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \frac{(S_1^2+S_2^2)^{\frac{n-2}{2}}}{(S^2)^{\frac{n-1}{2}}}. \tag{1.34}$$

(ii) The imaginary training sample method. With $n = 3, k = 1$ and $S^2 = S_1^2 + S_2^2$ $(\bar{x}_1 = \bar{x}_2)$ we have

$$B_{0k} = \frac{1}{\pi} \left[\frac{2}{3} \right]^{\frac{1}{2}} [S^2]^{-\frac{1}{2}}.$$

For this to be equal to one, our Bayes factor should be

$$B_{0k}^{I} = \pi^{\frac{1}{2}} \left[\frac{3k(n-k)}{2n} \right]^{\frac{1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \left[\frac{S_1^2+S_2^2}{S^2} \right]^{\frac{n-2}{2}}. \tag{1.35}$$

(iii) For the fractional Bayes factor with $b = \frac{3}{n}$ we have

$$B_{0k}^{F} = \pi^{\frac{1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \left[\frac{S_1^2 + S_2^2}{S^2} \right]^{\frac{n-3}{2}}. \tag{1.36}$$

(iv) The Posterior Bayes factor is

$$B_{0k}^{P} = \pi^{-\frac{1}{2}} \left[\frac{k(n-k)}{n} \right]^{\frac{1}{2}} \frac{\Gamma\left(n-\frac{1}{2}\right)}{\Gamma(n-1)} \frac{(S_1^2+S_2^2)^{n-1}}{(S^2)^{n-\frac{1}{2}}}. \tag{1.37}$$

The Intrinsic Bayes factor will not be considered here.

These Bayes factors depend on the ratio between S^2 and $S_1^2 + S_2^2$. The maximums, when $S^2 = S_1^2 + S_2^2$, is

$$\max B_{0k}^{F} = \pi^{\frac{1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)};$$

$$\max B_{0k}^{I} = \max B_{0k}^{F}\left[\frac{3k(n-k)}{2n}\right]^{\frac{1}{2}},$$

$$\max B_{0k} = \max B_{0k}^{I}\pi^{-1}\left(\frac{2}{3}\right)^{\frac{1}{2}}(S^2)^{-\frac{1}{2}}$$

and

$$\max B_{0k}^{P} = \max B_{0k}\frac{\Gamma\left(n-\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma(n-1)\Gamma\left(\frac{n-1}{2}\right)}. \tag{1.38}$$

Notice that $\max B_{0k}$ and $\max B_{0k}^{P}$ depend on S^2, and that $\max B_{0k}^{I} > \max B_{0k}^{F}$ for $n > 3$, which is the opposite of what happened in example 1 (see figure 1.1).

In Figure 1.2 the maximum Bayes factors are plotted as a function of n when $k = \frac{n}{2}$ and $S^2 = 1$. Now the B_{0k}^{I} is the highest while B_{0k}^{F} is the most conservative. The other two of course are decreasing functions of S^2.

The above two simple examples illustrate the widely different answers that can be obtained, depending on the partial Bayes factor used. All claimed to be reasonably "objective", but the interpretation of the weight of evidence in favour of a particular model clearly differs, and this can change from one application to the next.

The purpose of this study is the Bayesian detection and estimation of change-points in a variety of statistical models. We try to be as objective as possible so that the result can be considered as default, with little subjective prior input. One way of doing this is by use of partial Bayes factors. We will consider only the intrinsic and fractional Bayes factors for this purpose, as they are applicable in most cases and seem to give the most sensible answers in a variety of conditions. In the conclusion some of their advantages and disadvantages experienced during this study will be discussed.

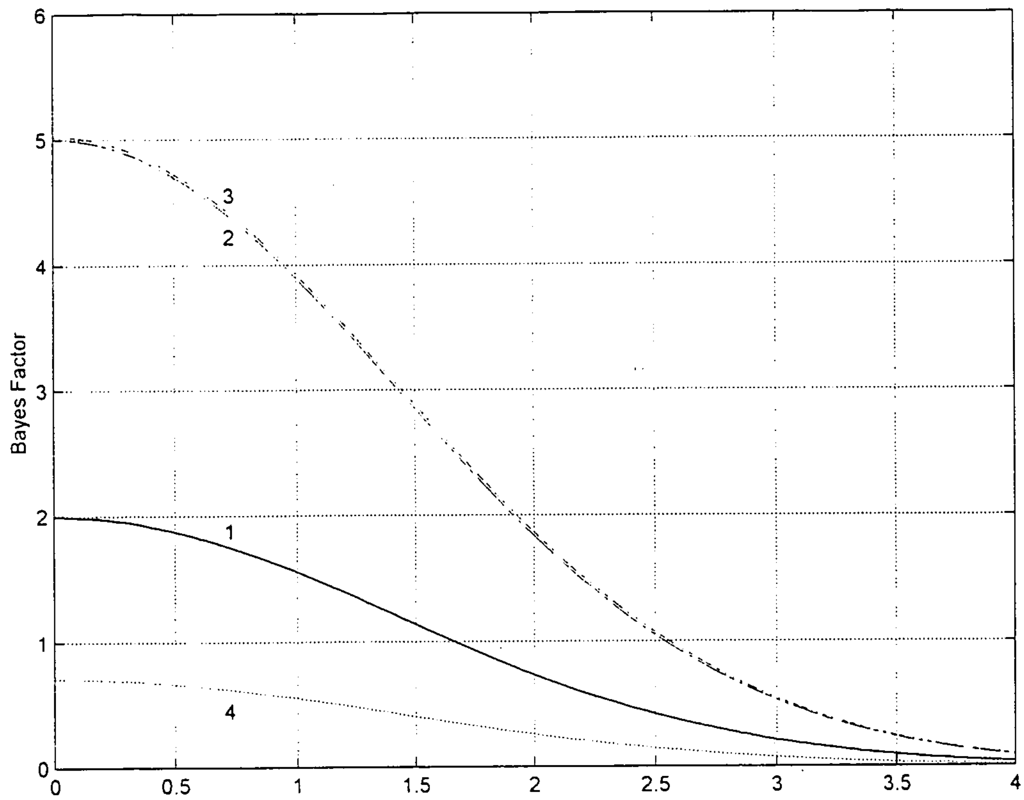


Figure 1.1: Expected Bayes factors as a function of $\delta = \mu_1 - \mu_2$ for $n = 100, k = 50$. (1) Usual Bayes factor, (2) Imaginary training sample, (3) Fractional Bayes factor, (4) Posterior Bayes factor.

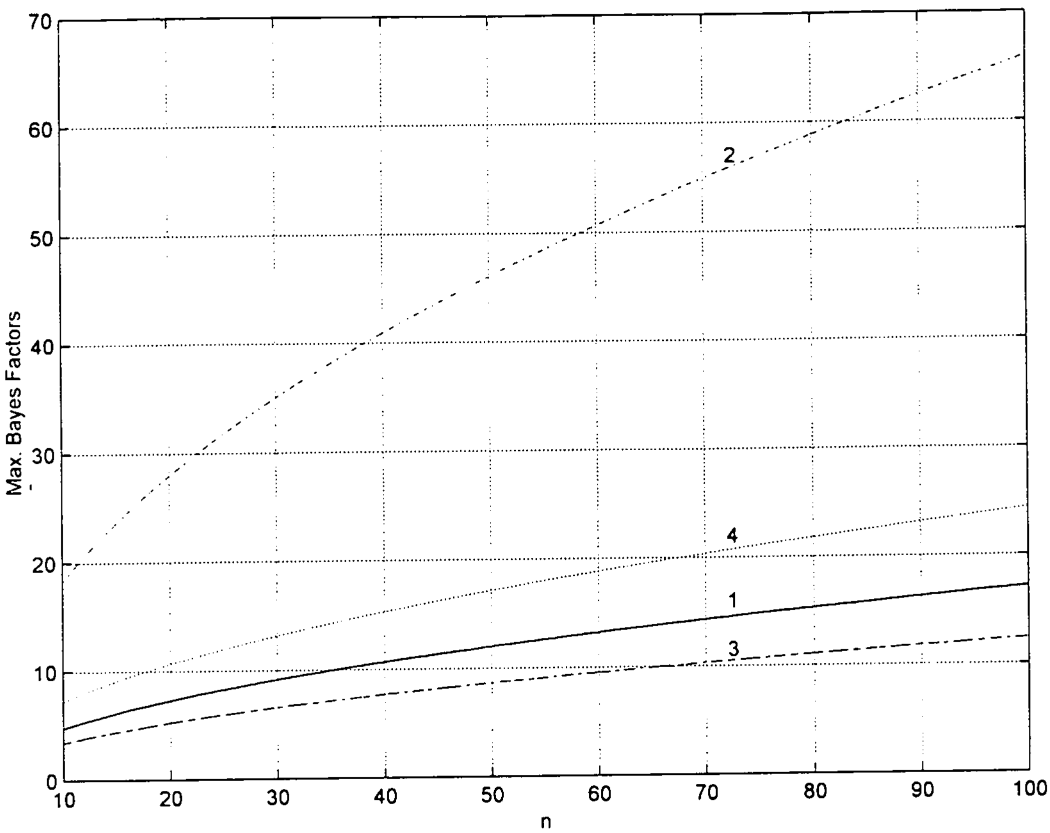


Figure 1.2: Maximum possible Bayes factor as a function of n for $k = \frac{n}{2}, S^2 = 1$ (1) Usual Bayes factor, (2) Imaginary training sample, (3) Fractional Bayes factor, (4) Posterior Bayes factor.

In chapter 2 the multivariate normal model will be examined with a change in the mean vector, a change in the mean vector and covariance matrix and a change in only the covariance matrix. We will also consider component analysis, multiple change-points, Bayes factors and the univariate case. Four illustrations will be used in chapter 2. In chapter 3 changes in the linear model will be examined, with seven illustrations, while changes in some other models (with eleven illustrations) will be considered in chapter 4. The hazard rate will be studied in chapter 5 (with two illustrations), while a conclusion, a summary of other methods and applications in the literature and possible extensions will be given in chapter 6.

CHAPTER 2

CHANGES IN THE NORMAL MODEL

2.1 INTRODUCTION

The problem of a change in the mean of random variables at an unknown time point has been addressed extensively in the literature. Broemeling (1974) considered a Bayesian analysis of a univariate normal model with known or unknown variance where (1) both the means μ_1 and μ_2 are known, (2) μ_1 is known and μ_2 is unknown and (3) both the means are unknown. Holbert and Broemeling (1977) also considered a change in the normal means and estimated the change-point in a sequence of independent random variables from a Bayesian viewpoint. Theoretical results and numerical examples were given.

Lee and Heghinian (1977) also made a Bayesian study about a shift in the mean of a set of independent univariate normal random variables with unknown common variance. The marginal and joint posterior distributions of the unknown time point and the amount of shift are derived. Booth and Smith (1982) looked at changes of the mean in the univariate as well as the multivariate normal sequence and Broemeling and Tsurumi (1987) also considered a change in the mean (both μ_1, μ_2 unknown) of normal sequence, both from a Bayesian viewpoint. Smith (1975) considered a Bayesian analysis of a univariate normal case with a changing mean where (1) all the parameters are known, (2) μ_1 known, μ_2 unknown and the reciprocal of the variance known, (3) μ_1 known, μ_2 unknown and the reciprocal of the variance unknown and (4) all the parameters unknown.

Within a non-Bayesian framework, this same problem of a changing mean has been discussed by Chernoff and Zacks (1964), Kander and Zacks (1966) and Sen and Srivastava (1973), although the emphasis and the objectives differed from those in the paper of Smith (1975). More non-Bayesian work on this problem includes that of Page (1954, 1955, 1957) using cumulative sums, and Hinkley (1970), using asymptotic arguments based on maximum likelihood estimates and likelihood ratio tests. Gardner (1969) considered the problem of detecting changes in the mean of independent unit variance normal random variables when the times of change are assigned an apriori distribution. Two situations were considered: The unknown amounts of change are (1) arbitrary, or (2) successively plus and minus the same unknown quantity. Sen and Srivastava (1975) and Bhattacharya and Brockwell (1976) also studied the problem of a change in the mean of a univariate normal distribution, all from non-Bayesian viewpoints.

The problem of a change in both the mean and variance as well as just a change in the variance at an unknown time point has, however, been covered less widely. Smith (1975) considered the normal case and gives posterior probabilities for the point in time when both a variance change and mean change occurs. Menzefricke (1981) also used a Bayesian approach to analyze a sequence of independent normal random variables in which the precision may have been subjected to one change at an unknown point in time. Posterior distributions were found both for an unknown point in time at which the change occurred and for the magnitude of the change.

From a non-Bayesian viewpoint, Hsu (1977,1979) examined the problem of testing whether there has been a change in the variance at an unknown time point by using sampling theory and applied the theory to stock-return data. It was of interest whether the uncertainty in the stock market was increased at some point during the Watergate events, an increase in uncertainty being measured by an increase in variance. Hsu (1977, 1979) extended the work of Miller, Wichern and Hsu (1971), who gave a Bayesian treatment of a similar problem. These papers, however, assumed that the mean of the process is known. A paper by Davis (1979) dealt with robust methods for the detection of a change in the variance.

In this chapter we want to find the marginal mass function of k, given the data, i.e. $\pi(k/X)$, and the posterior distributions of the parameters in the multivariate normal distribution for no, one or more than one change-point for the cases where just the mean changes, the mean and variance change and just the variance changes. We will also consider component analysis of the mean vector and covariance matrix and determine Bayes factors for the three cases. The univariate case will also be looked at, including autocorrelation for this case.

Finally, we will draw comparisons between our results and some of the results from the references cited above and look at a few applications.

2.2 EXACTLY ONE CHANGE-POINT IN THE MULTIVARIATE MODEL

2.2.1 A CHANGE IN THE MEAN

2.2.1.1 VARIANCE Σ KNOWN

Broemeling (1974) considered the univariate normal sequence with common known variance $\sigma^2 = 1$ and suppose that the first $k (k \in I_{n-1})$ observations have mean μ_1 and the remaining $n - k$ have mean μ_2, where $-\infty < \mu_1 < \mu_2 < \infty$. With μ_1 known, the improper vague prior $\Pi_0(\mu_2) \propto \text{const}(\mu_2 \in (\mu_1, \infty))$ and $\Pi_0(\mu_2) = 0$ otherwise, is used. He estimated the change-point k by getting the posterior distribution $\Pi(k)$ by considering the cases where (1) both means are known, (2) μ_1 is known and (3) both means are unknown. Here we consider a multivariate generalization of the case where both means are unknown.

Suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independent normal random vectors such that

$$\boldsymbol{X}_i \sim \begin{cases} iidN(\boldsymbol{\mu}_1, \Sigma); & i = 1, \ldots, k \\ iidN(\boldsymbol{\mu}_2, \Sigma); & i = k+1, \ldots, n-1 \end{cases} \tag{2.2.1}$$

where $n \geq 2p - 1, \boldsymbol{x}_i \in I\!R^p$ and $1 \leq k \leq n - 1$.

Furthermore the mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are unknown, with $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. The known covariance matrix Σ remains unchanged through the change of the mean at an unknown k.

The sequence is changing in the mean, where the first k values have a mean vector $\boldsymbol{\mu}_1$ and the remaining $n - k$ values have a different mean vector $\boldsymbol{\mu}_2$.

Assuming that a change has taken place, one will want to be able to detect the change and to estimate it as well as the other parameters of the model. To do a Bayesian analysis, one should choose prior densities. Let the marginal prior density of k be $\Pi(k) = \dfrac{1}{n-1}$ and the marginal prior density of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ be such that

$$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \sim iidN(\boldsymbol{\theta}, \Phi)$$

where $\boldsymbol{\theta}$ and Φ are assumed known.

The joint distribution of $\boldsymbol{X}(n \times p) = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n]$ conditional on $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and the change having taken place at $k(1 \leq k \leq n-1)$ is given by

$$f(X|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, k, \Sigma) = \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Sigma|^{-n/2} e^{-\frac{1}{2}\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)} e^{-\frac{1}{2}\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)}$$

where

$$\sum^{k_1} = \sum_{i=1}^{k} \qquad \text{and} \qquad \sum^{k_2} = \sum_{i=k+1}^{n}.$$

We want to find the marginal posterior distribution of k, given the data, i.e. $\Pi(k|X)$. Furthermore we want to find the posterior distributions of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, i.e. $\Pi(\boldsymbol{\mu}_1|k, X)$ and $\Pi(\boldsymbol{\mu}_2|k, X)$.

The marginal distribution of X is given by

$$f(X|k, \boldsymbol{\theta}, \Phi, \Sigma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(X|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, k, \Sigma)\Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)d\boldsymbol{\mu}_1\, d\boldsymbol{\mu}_2 = \left(\frac{1}{2\pi}\right)^{\frac{np}{2}+p} |\Sigma|^{-n/2}|\Phi|^{-1}$$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_{1\Sigma})'(k_1\Sigma^{-1}+\Phi^{-1})(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_{1\Sigma}) + \boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta} + \sum^{k_1}\boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{1\Sigma}'(k_1\Sigma^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{1\Sigma}\right]} d\boldsymbol{\mu}_1$$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}[(\boldsymbol{\mu}_2 - \hat{\boldsymbol{\mu}}_{2\Sigma})'(k_2\Sigma^{-1}+\Phi^{-1})(\boldsymbol{\mu}_2 - \hat{\boldsymbol{\mu}}_{2\Sigma}) + \boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta} + \sum^{k_2}\boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{2\Sigma}'(k_2\Sigma^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{2\Sigma}]} d\boldsymbol{\mu}_2$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Sigma|^{-n/2}|\Phi|^{-1}|k_1\Sigma^{-1}+\Phi^{-1}|^{-\frac{1}{2}}|k_2\Sigma^{-1}+\Phi^{-1}|^{-\frac{1}{2}}$$

$$e^{-\frac{1}{2}\left[2\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta} + \sum^{k_1}\boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i + \sum^{k_2}\boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{1\Sigma}'(k_1\Sigma^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{1\Sigma} - \hat{\boldsymbol{\mu}}_{2\Sigma}'(k_2\Sigma^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{2\Sigma}\right]}$$

where $k_1 = k, \quad k_2 = n-k$ and

$$\hat{\boldsymbol{\mu}}_{i\Sigma} = (k_i\Sigma^{-1} + \Phi^{-1})^{-1}(k_i\Sigma^{-1}\overline{\boldsymbol{x}}_{ik} + \Phi^{-1}\boldsymbol{\theta}), \quad i = 1, 2. \tag{2.2.2}$$

The posterior marginal mass function of k is then

$$\Pi(k|X, \boldsymbol{\theta}, \Phi, \Sigma) = \frac{f(X|k, \boldsymbol{\theta}, \Phi, \Sigma)\Pi(k)}{\sum_{k=1}^{n-1} f(X|k, \boldsymbol{\theta}, \Phi, \Sigma)\Pi(k)} \tag{2.2.3}$$

where $\pi(k) = \frac{1}{n-1}$.

We want to find the posterior distributions of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ where

$$\Pi(\boldsymbol{\mu}_j|X, \boldsymbol{\theta}, \Phi, \Sigma) = \sum_{k=1}^{n-1} \Pi(\boldsymbol{\mu}_j|k, X, \boldsymbol{\theta}, \Phi, \Sigma)\Pi(k|X, \boldsymbol{\theta}, \Phi, \Sigma), \quad j = 1, 2. \tag{2.2.4}$$

and

$$\boldsymbol{\mu}_i|k, X, \boldsymbol{\theta}, \Phi, \Sigma \sim N(\hat{\boldsymbol{\mu}}_{i\Sigma}, (k_i\Sigma^{-1} + \Phi^{-1})^{-1}), \qquad i = 1, 2.$$

Also notice from (2.2.4) that the unconditional posteriors of $\boldsymbol{\mu}_j (j = 1, 2)$, (given $X, \boldsymbol{\theta}, \Phi, \Sigma$) are mixtures of normal distributions.

Up to this stage the hyperparameters Φ and $\boldsymbol{\theta}$ were assumed known. In practice this is not usually true. In fact little may be known about the distributions of the means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. If the number of change-points is fixed as in this section, vague improper priors can be used. So if we let $\Phi^{-1} \to 0$ in equation (2.2.3), the posterior of k simplifies to

$$\Pi(k|X, \Sigma) = \frac{(k_1 k_2)^{-p/2} e^{-\frac{1}{2}tr(S_{1k}+S_{2k})\Sigma^{-1}}}{\sum_{k=1}^{n-1} (k_1 k_2)^{-p/2} e^{-\frac{1}{2}tr(S_{1k}+S_{2k})\Sigma^{-1}}}$$

where

$$S_{ik} = \sum^{k_i} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{ik})(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{ik})' \text{ and } \overline{\boldsymbol{x}}_{ik} = \frac{\sum^{k_i} \boldsymbol{x}_i}{k_i}. \tag{2.2.5}$$

Furthermore

$$\boldsymbol{\mu}_i | k, X, \Sigma \sim N\left(\overline{\boldsymbol{x}}_{ik}, \frac{1}{k_i}\Sigma\right), \qquad i = 1, 2.$$

Another way of dealing with the problem of unknown hyperparameters is to use a hierarchical model where the second stage priors are vague. For example, let

$$\Pi(\boldsymbol{\theta}) \propto 1$$

and

$$\Pi(\Phi) \propto |\Phi|^{-\frac{p+1}{2}}. \tag{2.2.6}$$

We know that for $i = 1, 2$

$$\hat{\boldsymbol{\mu}}_{i\Sigma}'(k_i\Sigma^{-1} + \Phi^{-1})\hat{\boldsymbol{\mu}}_{i\Sigma} = k_i^2\overline{\boldsymbol{x}}_{ik}'\Sigma^{-1}(k_i\Sigma^{-1} + \Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{ik}$$
$$-2k_i\overline{\boldsymbol{x}}_{ik}'\Sigma^{-1}(k_i\Sigma^{-1} + \Phi^{-1})^{-1}\Phi^{-1}\boldsymbol{\theta} + \boldsymbol{\theta}'\Phi^{-1}(k_i\Sigma^{-1} + \Phi^{-1})^{-1}\Phi^{-1}\boldsymbol{\theta}$$

so that

$$f(X|\Phi, k, \Sigma) = \left(\frac{1}{2\pi}\right)^{\frac{pn}{2}} |\Sigma|^{-\frac{n}{2}} |k_1\Sigma^{-1} + \Phi^{-1}|^{-\frac{1}{2}} |k_2\Sigma^{-1} + \Phi^{-1}|^{-\frac{1}{2}} \cdot |\Phi|^{-1}$$

$$e^{-\frac{1}{2}\left[\sum^{k_1} \boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i + \sum^{k_2} \boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i + k_1^2\overline{\boldsymbol{x}}_{1k}'\Sigma^{-1}(k_1\Sigma^{-1} + \Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{1k} + \right.}$$

$$\left. k_2^2\overline{\boldsymbol{x}}_{2k}'\Sigma^{-1}(k_2\Sigma^{-1} + \Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{2k}\right]$$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left\{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'[2I-(k_1\Phi\Sigma^{-1}+I)^{-1}-(k_2\Phi\Sigma^{-1}+I)^{-1}]^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})-\left[k_1(k_1I+\Sigma\Phi^{-1})^{-1}\overline{\boldsymbol{x}}_{1k}\right.\right.} +$$

$$\left.\left. k_2(k_2I+\Sigma\Phi^{-1})^{-1}\overline{\boldsymbol{x}}_{2k}\right]'\left[2I-(k_1\Phi\Sigma^{-1}+I)^{-1}-(k_2\Phi\Sigma^{-1}+I)^{-1}\right]\left[k_1(k_1I+\Sigma\Phi^{-1})^{-1}\overline{\boldsymbol{x}}_{1k}+k_2(k_2I+\Sigma\Phi^{-1})^{-1}\overline{\boldsymbol{x}}_{2k}\right]\right\} d\boldsymbol{\theta}$$

where

$$\hat{\theta} = [2I - (k_1\Phi\Sigma^{-1} + I)^{-1} - (k_2\Phi\Sigma^{-1} + I)^{-1}]^{-1}[k_1(k_1 I + \Sigma\Phi^{-1})^{-1}\overline{\boldsymbol{x}}_{1k} + k_2(k_2 I + \Sigma\Phi^{-1})^{-1}\overline{\boldsymbol{x}}_{2k}], \tag{2.2.7}$$

so that

$$f(X|k,\Sigma) = \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{\frac{p(n-1)}{2}} |\Sigma|^{-\frac{n}{2}} \frac{1}{n-1} |k_1\Sigma^{-1} + \Phi^{-1}|^{-\frac{1}{2}} |k_2\Sigma^{-1} + \Phi^{-1}|^{-\frac{1}{2}}$$

$$|2\Phi^{-1} - \Phi^{-1}(k_1\Sigma^{-1} + \Phi^{-1})^{-1}\Phi^{-1} - \Phi^{-1}(k_2\Sigma^{-1} + \Phi^{-1})^{-1}\Phi^{-1}|^{-\frac{1}{2}} |\Phi|^{-\left(\frac{p+3}{2}\right)}$$

$$e^{-\frac{1}{2}\{[k_1\Phi^{-1}(k_1\Sigma^{-1}+\Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{1k}+k_2\Phi^{-1}(k_2\Sigma^{-1}+\Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{2k}]'} .$$

$$[2\Phi^{-1}-\Phi^{-1}(k_1\Sigma^{-1}+\Phi^{-1})^{-1}\Phi^{-1}-\Phi^{-1}(k_2\Sigma^{-1}+\Phi^{-1})^{-1}\Phi^{-1}]^{-1}[k_1\Phi^{-1}(k_1\Sigma^{-1}+\Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{1k}+$$

$$k_2\Phi^{-1}(k_2\Sigma^{-1}+\Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{2k}] + \sum^{k_1} \boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i + \sum^{k_2} \boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i +$$

$$k_1^2\overline{\boldsymbol{x}}_{1k}'\Sigma^{-1}(k_1\Sigma^{-1}+\Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{1k}+k_2^2\overline{\boldsymbol{x}}_{2k}'\Sigma^{-1}(k_2\Sigma^{-1}+\Phi^{-1})^{-1}\Sigma^{-1}\overline{\boldsymbol{x}}_{2k}\} d\Phi. \tag{2.2.8}$$

This integral is analytically intractable and numerically very complex. In fact this integral probably does not exist, since the full conditional distribution of Φ^{-1} is singular for $p > 2$. So Φ is not estimable. One possible simplification is to assume that Φ is proportional to Σ, that is $\Phi^{-1} = \delta\Sigma^{-1}$, where $\delta > 0$. This is a strong assumption reducing the unknown matrix to a single unknown parameter δ. If we use the prior $\Pi(\delta) \propto \frac{1}{\delta}$, replacing (2.2.6), then (2.2.8) reduces to

$$f(X|k,\Sigma) = \int (2\pi)^{\frac{p(n-1)}{2}} \frac{1}{n-1} |\Sigma^{-1}|^{\frac{n+p+1}{2}} \delta^{p-\frac{3}{2}} (2k_1k_2 + \delta k_1 + \delta k_2)^{-\frac{1}{2}}$$

$$e^{-\frac{1}{2}tr\left\{\left(2-\frac{\delta}{k_1+\delta}-\frac{\delta}{k_2+\delta}\right)^{-1}\left[\frac{k_1\delta}{k_1+\delta}\overline{\boldsymbol{x}}_{1k}+\frac{k_2\delta}{k_2+\delta}\overline{\boldsymbol{x}}_{2k}\right]\left[\frac{k_1\delta}{k_1+\delta}\overline{\boldsymbol{x}}_{1k}+\frac{k_2\delta}{k_2+\delta}\overline{\boldsymbol{x}}_{2k}\right]'+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'+\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'+k_1^2(k_1+\delta)^{-1}\overline{\boldsymbol{x}}_{1k}\overline{\boldsymbol{x}}_{1k}'+k_2^2(k_2+\delta)^{-1}\overline{\boldsymbol{x}}_{2k}\overline{\boldsymbol{x}}_{2k}'\right\}\Sigma^{-1}} d\delta,$$

which can be integrated numerically.

Another possible solution is to use Gibbs sampling. To ensure proper posteriors a proper prior can be put on δ, e.g. $\delta \sim \Gamma(a,b)$. The full conditional distributions are as follows:
$\boldsymbol{\mu}_i|X,k,\boldsymbol{\theta},\delta,\boldsymbol{\mu}_j \sim N\left(\frac{k_i\overline{\boldsymbol{x}}_{ik}+\delta\boldsymbol{\theta}}{k_i+\delta},\frac{1}{k_i+\delta}\Sigma\right)$, where if $i=1$, then $j=2$ and vice versa.

$$\delta|X,k,\boldsymbol{\theta},\boldsymbol{\mu}_1,\boldsymbol{\mu}_2 \sim \text{Gamma }\left(a+p,\tfrac{1}{2}tr\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\theta})(\boldsymbol{\mu}_1-\boldsymbol{\theta})'+(\boldsymbol{\mu}_2-\boldsymbol{\theta})(\boldsymbol{\mu}_2-\boldsymbol{\theta})'+b\right),$$

$$\Pi(k|X,\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\delta,\boldsymbol{\theta}) = \frac{e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)'\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)+\sum^{k_2}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)\right]}}{\sum_{k=1}^{n-1}e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)'\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)+\sum^{k_2}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)\right]}}$$

and

$$\boldsymbol{\theta}|X,\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\delta,k) \sim N\left(\frac{\boldsymbol{\mu}_1+\boldsymbol{\mu}_2}{2},\frac{1}{2}\delta^{-1}\Sigma\right). \tag{2.2.9}$$

2.2.1.2 VARIANCE Σ UNKNOWN

Broemeling (1974) also considered the same model as mentioned in paragraph 2.2.1.1, but with common unknown variance and the means μ_1 and μ_2 known. He used the Jeffrey vague density $\Pi_0(\sigma^2) \propto \frac{1}{\sigma^2}$, $\sigma^2 > 0$ or $\Pi_0(\sigma^2) = 0$ otherwise and determined the posterior distribution $\pi(k|\boldsymbol{X})$, where k is the change-point.

Smith (1975) considered the univariate normal model $X_i \sim N(x|\theta_1), i = 1,\ldots,k$ (k the change-point) and $X_i \sim N(x|\theta_2)$, $i = k+1, \ldots, n$ where $N(x|\theta_1) \neq N(x|\theta_2)$, from a Bayesian viewpoint. Taking $\theta_j = (\mu_j, \sigma_j^{-2})$, the mean and the reciprocal of the variance, he consider the special case $\mu_2 \neq \mu_1$ (μ_1 known, μ_2 unknown) and unknown $\sigma_1^{-2} = \sigma_2^{-2} = \sigma^{-2}$ with standard vague prior assignments for μ_2 and σ^{-2}.

Once again suppose that

$$\boldsymbol{X}_i \sim \begin{cases} iidN(\boldsymbol{\mu}_1, \Sigma); & i = 1, \ldots, k \\ iidN(\boldsymbol{\mu}_2, \Sigma); & i = k+1, \ldots, n \end{cases} \tag{2.2.10}$$

where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ are still unknown, but the covariance matrix Σ is now unknown but still remains unchanged through the structural change at an unknown k. Notice that n must be larger than or equal to $2p+1$ for the parameters to be estimable for all $1 \leq k \leq n-1$.

The marginal prior densities of $k, \boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ conditional on Σ, are still as in the previous section. Furthermore let the marginal prior density of Σ^{-1} be $W(v, \Gamma)$, i.e.

$$\Pi(\Sigma^{-1}) \propto |\Sigma^{-1}|^{\frac{v-p-1}{2}} e^{-\frac{1}{2}tr\Gamma^{-1}\Sigma^{-1}}. \tag{2.2.11}$$

We once again want to find $\Pi(k|X)$ as well as the posterior distributions of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, i.e. $\Pi(\boldsymbol{\mu}_1|k, X)$ and $\Pi(\boldsymbol{\mu}_2|k, X)$.

The joint distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ conditional on $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma$ and the change having taken place at $k (1 \leq k \leq n-1)$ is the same as in section 2.2.1.1, while the marginal distribution of X is given by

$$f(X|k, \boldsymbol{\theta}, \Phi) = \iiint f(X|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, k, \Sigma)\Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)\Pi(\Sigma^{-1})d\boldsymbol{\mu}_1, d\boldsymbol{\mu}_2 d\Sigma^{-1} = \iiint \left(\frac{1}{2\pi}\right)^{\frac{np}{2}+p}$$

$$|\Sigma^{-1}|^{\frac{n+v-p-1}{2}} |\Phi|^{-1} e^{-\frac{1}{2}tr\Gamma^{-1}\Sigma^{-1}} e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) + (\boldsymbol{\mu}_1 - \boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\theta})\right]}$$

$$e^{-\frac{1}{2}\left[\sum^{k_2}(\boldsymbol{x}_i-\mu_2)'\Sigma^{-1}(\boldsymbol{x}_i-\mu_2)+(\mu_2-\theta)'\Phi^{-1}(\mu_2-\theta)\right]}d\boldsymbol{\mu}_1 d\boldsymbol{\mu}_2 d\Sigma^{-1}.$$

By using the previous result in section 2.2.1.1 where Σ was known, it now follows that

$$f(X|k,\boldsymbol{\theta},\Phi)\propto\int_{\Sigma^{-1}>0}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}|\Sigma^{-1}|^{\frac{n+v-p-1}{2}}|\Phi|^{-1}|k_1\Sigma^{-1}+\Phi^{-1}|^{-\frac{1}{2}}|k_2\Sigma^{-1}+\Phi^{-1}|^{-\frac{1}{2}}$$

$$e^{-\frac{1}{2}\left[2\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta}+\sum^{k_1}\boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_{1\Sigma}'(k_1\Sigma^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{1\Sigma}+tr\Gamma^{-1}\Sigma^{-1}\right.}$$

$$\left.+\sum^{k_2}\boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_{2\Sigma}'(k_2\Sigma^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{2\Sigma}\right]d\Sigma^{-1}$$

where $\hat{\boldsymbol{\mu}}_{1\Sigma}$ and $\hat{\boldsymbol{\mu}}_{2\Sigma}$ are as in (2.2.2).

This integral seems intractable, so once again let $\Phi^{-1}=\delta\Sigma^{-1}$ so that

$$f(X|k,\boldsymbol{\theta},\delta)=\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}\delta^p(k_1+\delta)^{-\frac{p}{2}}(k_2+\delta)^{-\frac{p}{2}}\int\left|\Sigma^{-1}\right|^{\frac{n+v-p-1}{2}}e^{-\frac{1}{2}tr[\Gamma^{-1}+2\delta\boldsymbol{\theta}\boldsymbol{\theta}'}$$

$$^{+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'+\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'-(k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}'-(k_2+\delta)\hat{\boldsymbol{\mu}}_{2\delta}\hat{\boldsymbol{u}}_{2\delta}']\Sigma^{-1}}d\Sigma^{-1}$$

$$=\quad c_{1p}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}(k_1+\delta)^{-\frac{p}{2}}(k_2+\delta)^{-\frac{p}{2}}\delta^p|\Gamma^{-1}+2\delta\boldsymbol{\theta}\boldsymbol{\theta}'+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'+\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'$$

$$-(k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}'-(k_2+\delta)\hat{\boldsymbol{\mu}}_{2\delta}\hat{\boldsymbol{\mu}}_{2\delta}'|^{-\left(\frac{v+n}{2}\right)}$$

$$(2.2.12)$$

where
$$\hat{\boldsymbol{\mu}}_{i\delta}=(k_i+\delta)^{-1}(k_i\overline{\boldsymbol{x}}_{ik}+\delta\boldsymbol{\theta}),\qquad i=1,2,\tag{2.2.13}$$

$$c_{1p}=2^{\frac{p(v+n)}{2}}\Gamma_p(\tfrac{1}{2}(v+n))\tag{2.2.14}$$

and

$$\Gamma_p(\frac{1}{2}n) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(n+1-i)\right).$$

The constants $c_{\ell p}$, if independent of k, are of no interest in this section, but are defined throughout this section as they will be needed in later sections. The marginal mass function of k is then

$$\Pi(k|X, \boldsymbol{\theta}, \delta) = \frac{f(X|k, \boldsymbol{\theta}, \delta)\Pi(k)}{\sum_{k=1}^{n-1} f(X|k, \boldsymbol{\theta}, \delta)\Pi(k)}. \quad (2.2.15)$$

Furthermore,

$$\Pi(\boldsymbol{\mu}_1|X, \boldsymbol{\theta}, \delta) = \sum_{k=1}^{n-1} \Pi(\boldsymbol{\mu}_1|k, X, \boldsymbol{\theta}, \delta)\Pi(k|X, \boldsymbol{\theta}, \delta)$$

where

$$\Pi(\boldsymbol{\mu}_1|k, X, \boldsymbol{\theta}, \delta) \propto \int_{\Sigma^{-1}>0} \Pi(\boldsymbol{\mu}_1|k, \Sigma^{-1}, X, \boldsymbol{\theta}, \delta)\Pi(\Sigma^{-1}|k, X, \boldsymbol{\theta}, \delta)d\Sigma^{-1}$$

and

$$\Pi\left(\boldsymbol{\mu}_1|k, \Sigma^{-1}, X, \boldsymbol{\theta}, \delta\right) \propto f\left(X|\boldsymbol{\mu}_1, \Sigma^{-1}, k, \boldsymbol{\theta}, \delta\right)\Pi\left(\boldsymbol{\mu}_1|k, \Sigma^{-1}, \boldsymbol{\theta}, \delta\right)$$

so that

$$\boldsymbol{\mu}_1|k, \Sigma^{-1}, X, \boldsymbol{\theta}, \delta \sim N\left(\hat{\boldsymbol{\mu}}_{1\delta}, \frac{\Sigma}{k+\delta}\right),$$

while

$$\Pi(\Sigma^{-1}|k, X, \boldsymbol{\theta}, \delta) \propto f(X|k, \Sigma^{-1}, \boldsymbol{\theta}, \delta)\Pi(\Sigma^{-1})$$

so that

$$\Sigma^{-1}|k, X, \boldsymbol{\theta}, \delta \sim W\left(v+n, \left[\Gamma^{-1} + 2\delta\boldsymbol{\theta}\boldsymbol{\theta}' + \sum^{k_1} \boldsymbol{x}_i\boldsymbol{x}_i' + \right.\right.$$

$$\left.\left.\sum^{k_2} \boldsymbol{x}_i\boldsymbol{x}_i' - (k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}' - (k_2+\delta)\hat{\boldsymbol{\mu}}_{2\delta}\hat{\boldsymbol{\mu}}_{2\delta}'\right]^{-1}\right).$$

Therefore

$$\Pi(\boldsymbol{\mu}_1|k,X,\boldsymbol{\theta},\delta) \propto c_{2p}(k_1+\delta)^{p/2}|\Gamma^{-1}+2\delta\boldsymbol{\theta}\boldsymbol{\theta}'+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'+\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'-(k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}'$$

$$-(k_2+\delta)\hat{\boldsymbol{\mu}}_{2\delta}\hat{\boldsymbol{\mu}}_{2\delta}'+(k_1+\delta)(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_{1\delta})(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_{1\delta})'|^{-\left(\frac{v+n+1}{2}\right)}$$

where

$$c_{2p}=2^{\frac{p(v+n+1)}{2}}\Gamma_p\left(\tfrac{1}{2}(v+n+1)\right). \tag{2.2.16}$$

Therefore $\boldsymbol{\mu}_i$ has a multivariate t-distribution,

$$\boldsymbol{\mu}_i|k,X,\boldsymbol{\theta},\delta\sim t_p(v+n+1-p,\hat{\boldsymbol{\mu}}_{i\delta},T_{i\delta}), \qquad i=1,2$$

where

$$T_{i\delta}=(v+n+1-p)(k_i+\delta)\left[\Gamma^{-1}+2\delta\boldsymbol{\theta}\boldsymbol{\theta}'+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'+\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'-(k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}'-(k_2+\delta)\hat{\boldsymbol{\mu}}_{2\delta}\hat{\boldsymbol{\mu}}_{2\delta}'\right]^{-1} \tag{2.2.17}$$

and

$$\Pi(\boldsymbol{\mu}_i|X,\boldsymbol{\theta},\delta)=\sum_{k=1}^{n-1}\Pi(\boldsymbol{\mu}_i|k,X,\boldsymbol{\theta},\delta)\pi(k|X,\boldsymbol{\theta},\delta), \qquad i=1,2$$

can be obtained.

Up to stage, in this section, the hyperparameters δ and $\boldsymbol{\theta}$ were assumed known. While the number of change-points is fixed at one (as we assume), vague improper priors can be used by letting $\delta\longrightarrow 0,\Gamma^{-1}\longrightarrow 0$ and $v\longrightarrow 0$ in equation (2.2.12), the posterior of k simplifies to

$$\Pi(k|X)=\frac{(k_1k_2)^{-p/2}\left[|S_{1k}+S_{2k}|\right]^{-n/2}}{\sum_{k=1}^{n-1}(k_1k_2)^{-p/2}[|S_{1k}+S_{2k}|]^{-n/2}}, \tag{2.2.18}$$

assuming $\pi(k)=\frac{1}{n-1}\forall k$. Furthermore, for $i=1,2$,

$$\boldsymbol{\mu}_i|k, X \sim t_p(n-p+1, \overline{\boldsymbol{x}}_{ik}, T_i) \tag{2.2.19}$$

where

$$T_i = k_i(n-p+1)(S_{1k}+S_{2k})^{-1}$$

and

$$\Sigma^{-1}|k, X \sim W\left(n, \left[\sum^{k_1} \boldsymbol{x}_i\boldsymbol{x}_i' + \sum^{k_2} \boldsymbol{x}_i\boldsymbol{x}_i' - k_1\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}' - k_2\hat{\boldsymbol{\mu}}_{2\delta}\hat{\boldsymbol{\mu}}_{2\delta}'\right]^{-1}\right).$$

Another way of dealing with the problem of unknown hyperparameters is to use a hierarchical model where the second stage priors are vague. For example let $\Pi(\boldsymbol{\theta}) \propto 1$ and $\Pi(\delta) \propto \frac{1}{\delta}$. From (2.2.12) and after completing the square it follows that

$$\begin{aligned} f(X|k,\delta) &= c_{1p}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}(k_1+\delta)^{-\frac{p}{2}}(k_2+\delta)^{-\frac{p}{2}}\delta^{p-1}\int_0^\infty |\Gamma^{-1} + \sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i' + \sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i' - (k_1+\delta)^{-1} \\ &\quad k_1^2\overline{\boldsymbol{x}}_{1k}\overline{\boldsymbol{x}}_{1k}' - (k_2+\delta)^{-1}k_2^2\overline{\boldsymbol{x}}_{2k}\overline{\boldsymbol{x}}_{2k}' - \hat{\boldsymbol{\theta}}_\delta(2\delta - (k_1+\delta)^{-1}\delta^2 - (k_2+\delta)^{-1}\delta^2)\hat{\boldsymbol{\theta}}_\delta' + \\ &\quad (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_\delta)(2\delta - (k_1+\delta)^{-1}\delta^2 - (k_2+\delta)^{-1}\delta^2)(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_\delta)'|^{-\left(\frac{v+n}{2}\right)}d\boldsymbol{\theta} \end{aligned}$$

so that

$$\begin{aligned} &\Pi(k|X,\delta) \\ &\propto c_{1p}\delta^{p-\frac{3}{2}}2^{-\frac{np}{2}}(k_1+\delta)^{-\left(\frac{p-1}{2}\right)}(k_2+\delta)^{-\left(\frac{p-1}{2}\right)}\Gamma\left(\frac{v+n-p}{2}\right)\left[\Gamma\left(\frac{v+n}{2}\right)\right]^{-1}\pi^{-\frac{p(n-1)}{2}} \\ &[2k_1k_2+\delta k_1+\delta k_2]^{-\frac{1}{2}}|T_{\delta\Gamma}|^{-\left(\frac{v+n-p}{2}\right)}, \qquad k=1,\ldots,n-1 \end{aligned} \tag{2.2.20}$$

where

$$\hat{\boldsymbol{\theta}}_\delta = [2-(k_1+\delta)^{-1}\delta - (k_2+\delta)^{-1}\delta]^{-1}\left[k_1(k_1+\delta)^{-1}\overline{\boldsymbol{x}}_{1k} + k_2(k_2+\delta)^{-1}\overline{\boldsymbol{x}}_{2k}\right], \tag{2.2.21}$$

c_{1p} is given in (2.2.14) and

$$T_{\delta\Gamma} = \Gamma^{-1} + \sum^{k_1} \boldsymbol{x}_i\boldsymbol{x}_i' + \sum^{k_2} \boldsymbol{x}_i\boldsymbol{x}_i' - (k_1+\delta)^{-1}k_1^2\overline{\boldsymbol{x}}_{1k}\overline{\boldsymbol{x}}_{1k}' - (k_2+\delta)^{-1}k_2^2\overline{\boldsymbol{x}}_{2k}\overline{\boldsymbol{x}}_{2k}' -$$

$$\hat{\boldsymbol{\theta}}_\delta[2\delta - (k_1+\delta)^{-1}\delta^2 - (k_2+\delta)^{-1}\delta^2]\hat{\boldsymbol{\theta}}_\delta'. \qquad (2.2.22)$$

This can be integrated numerically. Note that δ can't approach zero as in (2.2.18). In that case $\boldsymbol{\theta}$ disappeared when $\Gamma^{-1} \to 0$. In this case, $\boldsymbol{\theta}$ is integrated out before Γ^{-1} approaches zero. This means that the prior structural relationship between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, namely that they come from exchangeable priors, is retained, and expressed through δ. However, we can still let $\Gamma^{-1} \to 0$ and $v \to 0$ in equation (2.2.20).

When using Gibbs sampling, the full conditional distributions of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \delta, k$ and $\boldsymbol{\theta}$ are the same as in (2.2.9), with additionally (let $\Gamma^{-1} \to 0$ and $v \to 0$),

$$\Sigma^{-1}|k, X, \boldsymbol{\theta}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \delta \sim W\left(n+2, \left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)(\boldsymbol{x}_i - \boldsymbol{\mu}_1)' + \delta(\boldsymbol{\mu}_1 - \boldsymbol{\theta})(\boldsymbol{\mu}_1 - \boldsymbol{\theta})' + \sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)(\boldsymbol{x}_i - \boldsymbol{\mu}_2)' + \delta(\boldsymbol{\mu}_2 - \boldsymbol{\theta})(\boldsymbol{\mu}_2 - \boldsymbol{\theta})'\right]^{-1}\right). \qquad (2.2.23)$$

2.2.2 A CHANGE IN THE MEAN AND VARIANCE

In the same univariate normal model as mentioned in paragraph 2.2.1.2, Smith (1975) also looked at the special case where the means $\mu_1 \neq \mu_2$ and the reciprocals of the variances $\sigma_1^{-2} \neq \sigma_2^{-2}$ are all unknown. He derived the posterior distribution of k.

Menzefricke (1981) examined the model $X_i \sim N(\mu_1, \sigma_1^{-2}),\ \ i = 1, \ldots, k,\ X_i \sim N(\mu_2, \sigma_2^{-2}),\ \ i = k+1, \ldots, n$, with unknown means μ_j and precisions $\sigma_j^2, \quad j = 1, 2$. He derived posterior distributions of the change-point k and $\tau = \dfrac{\sigma_2^2}{\sigma_1^2}$ the magnitude of the change in variance, under the assumptions (1) μ_1 and μ_2 are unknown and (2) μ_1 and μ_2 are known.

2.2.2.1 CASE 1.

Suppose again that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independent normal random vectors such that

$$\boldsymbol{X}_i \sim \begin{cases} iidN(\boldsymbol{\mu}_1, \Sigma_1); & i = 1, \ldots, k \\ iidN(\boldsymbol{\mu}_2, \Sigma_2); & i = k+1, \ldots, n \end{cases} \tag{2.2.24}$$

where $n > 2p,\ \boldsymbol{x}_i \in R^p$ and $p + 1 \leq k \leq n - p - 1$.

Furthermore the mean vectors $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ and the covariance matrices $\Sigma_1 \neq \Sigma_2$ are unknown. The marginal prior density of k is still uniform, while the marginal prior densities of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are such that

$$\boldsymbol{\mu}_i \sim N(\boldsymbol{\theta}_i, \Phi_i), \quad i = 1, 2.$$

Let the marginal prior density of Σ_1^{-1} and Σ_2^{-1} be

$$\Sigma_i^{-1} \sim W(v, \Gamma), \qquad i = 1, 2. \tag{2.2.25}$$

We once again want to find $\Pi(k|X)$ as well as the posterior distributions of the unknown parameters.

The joint distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ conditional on $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2$ and the change having taken place at k, is given by

$$f(X|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2, k) = \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Sigma_1|^{-\frac{k_1}{2}} |\Sigma_2|^{-\frac{k_2}{2}}$$

$$e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) + \sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\right]}.$$

Then

$$f(X|k, \boldsymbol{\theta}_i, \Phi_i, \Gamma, v) = \iiiint f(X|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, k, \Sigma_1, \Sigma_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \Phi_1, \Phi_2)$$

$$\Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)\Pi(\Sigma_1^{-1})\Pi(\Sigma_2^{-1})d\boldsymbol{\mu}_1 d\boldsymbol{\mu}_2 d\Sigma_1^{-1} d\Sigma_2^{-1}$$

$$= \left[\frac{1}{\Gamma_p(\frac{1}{2}v)}\right]^2 |2\Gamma|^{-v} \int\int \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Phi_1|^{-\frac{1}{2}}|\Phi_2|^{-\frac{1}{2}}|\Sigma_1|^{-\frac{k_1}{2}}|\Sigma_2|^{-\frac{k_2}{2}}|k_1\Sigma_1^{-1}+\Phi_1^{-1}|^{-\frac{1}{2}}|k_2\Sigma_2^{-1}+\Phi_2^{-1}|^{-\frac{1}{2}}\cdot$$

$$\left|\Sigma_1^{-1}\right|^{\frac{v-p-1}{2}}\left|\Sigma_2^{-1}\right|^{\frac{v-p-1}{2}} e^{-\frac{1}{2}\left[\boldsymbol{\theta}_1'\Phi_1^{-1}\boldsymbol{\theta}_1+\boldsymbol{\theta}_2'\Phi_2^{-1}\boldsymbol{\theta}_2+\sum^{k_1}\boldsymbol{x}_i'\Sigma_1^{-1}\boldsymbol{x}_i+tr(\Gamma^{-1}\Sigma_1^{-1})-\hat{\boldsymbol{\mu}}_{1\Sigma_1\boldsymbol{\theta}_1}'(k_1\Sigma_1^{-1}+\right.}$$

$$\left.\Phi_1^{-1})\hat{\boldsymbol{\mu}}_{1\Sigma_1\boldsymbol{\theta}_1}+\sum^{k_2}\boldsymbol{x}_i'\Sigma_2^{-1}\boldsymbol{x}_i+tr(\Gamma^{-1}\Sigma_2^{-1})-\hat{\boldsymbol{\mu}}_{2\Sigma_2\boldsymbol{\theta}_2}'(k_2\Sigma_2^{-1}+\Phi_2^{-1})\hat{\boldsymbol{\mu}}_{2\Sigma_2\boldsymbol{\theta}_2}\right] d\Sigma_1^{-1}d\Sigma_2^{-1} \quad (2.2.26)$$

and

$$\hat{\boldsymbol{\mu}}_{i\Sigma_i\boldsymbol{\theta}_i} = (k_i\Sigma_i^{-1}+\Phi_i^{-1})^{-1}(k_i\Sigma_i^{-1}\overline{\boldsymbol{x}}_{ik}+\Phi_i^{-1}\boldsymbol{\theta}_i), \quad i=1,2. \quad (2.2.27)$$

This integral is numerically very complex and analytically intractable. As before, a possible solution is to assume that Φ_1 is proportional to Σ_1 and that Φ_2 is proportional to Σ_2, i.e. $\Phi_i^{-1} = \delta_i\Sigma_i^{-1}$, where $\delta_i > 0$ and $i = 1,2$.

It then follows that

$$f(X|k,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\delta_1,\delta_2,\Gamma,v)$$

$$= \left[\frac{1}{\Gamma_p(\frac{1}{2}v)}\right]^2 |2\Gamma|^{-v}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}} c_{3p}c_{4p}\delta_1^{\frac{p}{2}}\delta_2^{\frac{p}{2}}(k_1+\delta_1)^{-\frac{p}{2}}(k_2+\delta_2)^{-\frac{p}{2}}|\Gamma^{-1}+\delta_1\boldsymbol{\theta}_1\boldsymbol{\theta}_1'+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'-$$

$$\hat{\boldsymbol{\mu}}_{1\Sigma_1\delta_1}'(k_1+\delta_1)\hat{\boldsymbol{\mu}}_{1\Sigma_1\delta_1}'|^{-\left(\frac{v+k_1}{2}\right)}|\Gamma^{-1}+\delta_2\boldsymbol{\theta}_2\boldsymbol{\theta}_2'+\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'-\hat{\boldsymbol{\mu}}_{2\Sigma_2\delta_2}(k_2+\delta_2)\hat{\boldsymbol{\mu}}_{2\Sigma_2\delta_2}'|^{-\left(\frac{v+k_2}{2}\right)} \quad (2.2.28)$$

where

$$\hat{\boldsymbol{\mu}}_{i\Sigma_i\delta_i} = (k_i+\delta_i)^{-1}(k_i\overline{\boldsymbol{x}}_{ik}+\delta_i\boldsymbol{\theta}_i), \quad i=1,2, \quad (2.2.29)$$

$$c_{3p} = 2^{\frac{p(v+k_1)}{2}}\Gamma_p\left(\frac{1}{2}(v+k_1)\right) \quad (2.2.30)$$

and

$$c_{4p} = 2^{\frac{p(v+k_2)}{2}} \Gamma_p \left(\frac{1}{2}(v + k_2) \right). \tag{2.2.31}$$

The marginal posterior mass function of k is then

$$\Pi(k|X, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \delta_1, \delta_2) = \frac{f(X|k, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \delta_1, \delta_2)\Pi(k)}{\sum_{k=p+1}^{n-p-1} f(X|k, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \delta_1, \delta_2)\Pi(k)} \tag{2.2.32}$$

and for $i = 1, 2$

$$\Sigma_i^{-1}|k, X, \boldsymbol{\theta}_i, \delta_i \sim W\left(v + k_i, \left[\Gamma^{-1} + \delta_i \boldsymbol{\theta}_i \boldsymbol{\theta}_i' + \sum^{k_i} \boldsymbol{x}_i \boldsymbol{x}_i' - \hat{\boldsymbol{\mu}}_{i\Sigma_i\delta_i}(k_i + \delta_i)\hat{\boldsymbol{\mu}}_{i\Sigma_i\delta_i}\right]^{-1}\right).$$

We want to find the conditional posterior distributions of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ where, for $j = 1, 2$,

$$\Pi(\boldsymbol{\mu}_j|X, \boldsymbol{\theta}_j, \delta_j, \Sigma_j) = \sum_{k=1}^{n-1} \Pi(\boldsymbol{\mu}_j|k, X, \boldsymbol{\theta}_j, \delta_j, \Sigma_j)\Pi(k|X, \boldsymbol{\theta}_j, \delta_j, \Sigma_j) \tag{2.2.33}$$

where

$$\boldsymbol{\mu}_j|k, X, \boldsymbol{\theta}_j, \delta_j, \Sigma_j \sim N\left(\hat{\boldsymbol{\mu}}_{j\delta_j}, \frac{\Sigma_j}{k_j + \delta_j}\right)$$

and

$$\hat{\boldsymbol{\mu}}_{j\delta_j} = (k_j + \delta_j)^{-1}(k_j \overline{\boldsymbol{x}}_{jk} + \delta_j \boldsymbol{\theta}_j). \tag{2.2.34}$$

Notice from (2.2.33) that the posterior of $\boldsymbol{\mu}_j$, given $X, \boldsymbol{\theta}_j, \delta_j, \Sigma_j (j = 1, 2)$ are mixtures of normal distributions.

The above distributions are still functions of the unknown hyperparameters $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \delta_1$ and δ_2. Also, (2.2.33) and (2.2.34) are functions of the Σ_j. Under the assumption of having a change-point, vague improper priors can be used. By letting $\delta_1 \to 0, \quad \delta_2 \to 0, \quad \Gamma^{-1} \to 0$

and $v \to 0$ in equation (2.2.28), the posterior of k simplifies to

$$\Pi(k|X) = \frac{c_{5p}(k_1k_2)^{-p/2}|S_{1k}|^{-\frac{k_1}{2}}|S_{2k}|^{-\frac{k_2}{2}}}{\sum_{k=p+1}^{n-p-1} c_{5p}(k_1k_2)^{-p/2}|S_{1k}|^{-\frac{k_1}{2}}|S_{2k}|^{-\frac{k_2}{2}}} \tag{2.2.35}$$

where

$$c_{5p} = 2^{\frac{pn}{2}}\Gamma_p\left(\frac{1}{2}k_1\right)\Gamma_p\left(\frac{1}{2}k_2\right).$$

It is interesting to note that if vague priors were assumed from the beginning of the analysis, the answer is slightly different from (2.2.35), namely the loss of one degree of freedom. Then we would have

$$\Pi(k|X) \propto (k_1k_2)^{-p/2}\Gamma_p\left(\frac{k_1-1}{2}\right)\Gamma_p\left(\frac{k_2-1}{2}\right)|S_{1k}|^{-\frac{k_1-1}{2}}|S_{2k}|^{-\frac{k_2-1}{2}}. \tag{2.2.36}$$

Furthermore, for $j = 1, 2$,

$$\Sigma_j^{-1}|k, X \sim W\left(k_j, \left[\sum^{k_j}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{jk})(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{jk})'\right]^{-1}\right).$$

Also notice that ($j = 1, 2$)

$$\boldsymbol{\mu}_j|k, X \sim t_p(k_j + 1 - p, \overline{\boldsymbol{x}}_{jk}, k_j\left(k_j + 1 - p)S_{jk}^{-1}\right).$$

Another way of dealing with the problem of unknown hyperparameters is to use a hierarchical model where, once again, the second stage priors are vague. For example let $\Pi(\boldsymbol{\theta}_i) \propto 1$ and $\Pi(\delta_i) \propto \frac{1}{\delta_i}$, $\quad i = 1, 2$.

We know from (2.2.28), after completing the square and integrating out $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, that

$$
\begin{aligned}
f(X|k,\Gamma,v,\delta_1,\delta_2) \quad = \quad & c_{6p}\left[\frac{1}{\Gamma_p(\frac{1}{2}v)}\right]^2 |\Gamma|^{-v} 2^{-\frac{p(v+n)}{2}} \pi^{-\frac{p(n-2)}{2}} (\delta_1\delta_2)^{\frac{p-1}{2}} (k_1k_2)^{-\frac{1}{2}} \\
& [(k_1+\delta_1)(k_2+\delta_2)]^{-(\frac{p-1}{2})} \Gamma\left(\tfrac{v+k_2-p}{2}\right) \left[\Gamma\left(\tfrac{v+k_1}{2}\right)\Gamma\left(\frac{v+k_2}{2}\right)\right]^{-1} \Gamma\left(\tfrac{v+k_1-p}{2}\right) \\
& |\Gamma^{-1}+S_{1k}|^{-(\frac{v+k_1-p}{2})} |\Gamma^{-1}+S_{2k}|^{-(\frac{v+k_2-p}{2})}
\end{aligned}
\tag{2.2.37}
$$

where

$$
c_{6p} = 2^{\frac{p(2v+n)}{2}} \Gamma_p\left(\frac{1}{2}(v+k_1)\right) \Gamma_p\left(\frac{1}{2}(v+k_2)\right) \tag{2.2.38}
$$

so that, with $\Gamma^{-1} \to 0$ and $v \to 0$ it follows that

$$
\begin{aligned}
\Pi(k|X,\delta_1,\delta_2) \quad \propto \quad & c_{5p}(\delta_1\delta_2)^{\frac{p-1}{2}} (k_1k_2)^{-\frac{1}{2}} [(k_1+\delta_1)(k_2+\delta_2)]^{-\left(\frac{p-1}{2}\right)} \Gamma\left(\tfrac{k_1-p}{2}\right) \Gamma\left(\tfrac{k_2-p}{2}\right) \\
& \left[\Gamma\left(\tfrac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)\right]^{-1} |S_{1k}|^{-(\frac{k_1-p}{2})} |S_{2k}|^{-(\frac{k_2-p}{2})}.
\end{aligned}
\tag{2.2.39}
$$

It is possible to put a vague prior on Γ^{-1} in equation (2.2.37) to preserve the prior exchangeability assumption between Σ_1 and Σ_2. This will result in the integral of a function of the form of a multivariate Beta distribution of the second kind. We will not pursue that here further, but will give the equivalent result in the univariate case in paragraph 2.7.2.

Another possible solution is to use Gibbs sampling, where the full conditional distributions (for fixed Γ and v) are as follows (j=1,2):

$$
\boldsymbol{\mu}_j| \text{ rest of parameters} \sim N\left(\frac{k_j\overline{\boldsymbol{x}}_{jk}+\delta_j\boldsymbol{\theta}_j}{k_j+\delta_j}, \frac{1}{k_j+\delta_j}\Sigma_j\right),
$$

$$\Pi(k|\text{ rest of parameters}) \quad \propto \quad |\Sigma_1|^{-\frac{k_1}{2}}|\Sigma_2|^{-\frac{k_2}{2}}.$$

$$e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)+\sum^{k_2}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)\right]},$$

$\boldsymbol{\theta}_j|$ rest of parameters $\sim N(\boldsymbol{\mu}_j, \delta_j^{-1}\Sigma_j)$,

$\delta_j|$ rest of parameters $\sim$ Gamma $\left(a+\frac{p}{2};\quad \frac{1}{2}tr\Sigma_j^{-1}(\boldsymbol{\mu}_j-\boldsymbol{\theta}_j)(\boldsymbol{\mu}_j-\boldsymbol{\theta}_j)'+b\right)$,

$$\Sigma_j^{-1}\mid \text{rest of parameters} \sim W\left(v_j+k_j, \left[\Gamma^{-1}+\sum^{k_j}(\boldsymbol{x}_i-\boldsymbol{\mu}_j)(\boldsymbol{x}_i-\boldsymbol{\mu}_j)'\right]^{-1}\right). \tag{2.2.40}$$

Once again, to ensure proper posteriors, a proper prior must be put on $\delta_j (j = 1, 2)$, e.g. $\delta_j \sim \Gamma(a, b)$.

With reference to a more general model, a proper prior can be put on $\Phi_i^{-1}(i = 1, 2)$ so as to obtain proper posteriors, e.g. $\Phi_i^{-1} \sim W(\rho, \Psi)$. So once again Gibbs sampling can be applied, where the full conditional distributions are now as follows:

$\Pi(k|$ rest of parameters$)$

$$= \frac{|\Sigma_1|^{-\frac{k_1}{2}}|\Sigma_2|^{-\frac{k_2}{2}} e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)+\sum^{k_2}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)\right]}}{\sum_{k=p+1}^{n-p-1}|\Sigma_1|^{-\frac{k_1}{2}}|\Sigma_2|^{-\frac{k_2}{2}} e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_1)+\sum^{k_2}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_2)\right]}},$$

$\boldsymbol{\theta}_j|$ rest of parameters $\sim N(\boldsymbol{\mu}_j, \Phi_j)$,

$\Phi_j^{-1}|$ rest of parameters $\sim W(\rho+1, [\Psi^{-1}+(\boldsymbol{\mu}_j-\boldsymbol{\theta}_j)(\boldsymbol{\mu}_j-\boldsymbol{\theta}_j)']^{-1})$,

$$\Sigma_j^{-1} \mid \text{rest of parameters} \sim W\left(v_j + k_j, \left[\Gamma^{-1} + \sum^{k_j}\left(\boldsymbol{x}_i - \boldsymbol{\mu}_j\right)\left(\boldsymbol{x}_i - \boldsymbol{\mu}_j\right)'\right]^{-1}\right)$$

and

$$\boldsymbol{\mu}_j | \text{ rest of parameters} \sim N\left((k_j\Sigma_j^{-1} + \Phi_j^{-1})^{-1}(k_j\Sigma_j^{-1}\overline{\boldsymbol{x}}_{jk} + \Phi_j^{-1}\boldsymbol{\theta}_j), (k_j\Sigma_j^{-1} + \Phi_j^{-1})^{-1}\right). \tag{2.2.41}$$

2.2.2.2 CASE 2

Here we consider the same model as in case 1, but with different prior assumptions. Let the marginal prior density of k be the same as previously, while the marginal prior density of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ be such that $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \sim N(\boldsymbol{\theta}, \Phi)$. Let the marginal prior density of Σ_1^{-1} be

$$\Pi(\Sigma_1^{-1}) \propto |\Sigma_1^{-1}|^{\frac{v-p-1}{2}} e^{-\frac{1}{2} tr \Gamma^{-1}\Sigma_1^{-1}} \tag{2.2.42}$$

and assume furthermore that Σ_2 is proportional to Σ_1, i.e.

$$\Sigma_2^{-1} = \gamma \Sigma_1^{-1}. \tag{2.2.43}$$

Notice that in case 2, a structural relationship is assumed a priori between the two variances, which is not present in case 1. We once again want to find $\pi(k|X)$, $\Pi(\boldsymbol{\mu}_1|k, X)$ and $\Pi(\boldsymbol{\mu}_2|k, X)$.

The joint distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ conditional on $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2$ and the change having taken place at k, is given by

$$f(X|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \gamma, k) = \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Sigma_1^{-1}|^{\frac{k_1}{2}} |\gamma\Sigma_1^{-1}|^{\frac{k_2}{2}} e^{-\frac{1}{2}\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)}$$

$$e^{-\frac{1}{2}\gamma\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)}.$$

The marginal distribution of X is given by

$$f(X|k,\boldsymbol{\theta},\Phi,\gamma) = \iiint f(X|\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,k,\Sigma_1,\Sigma_2)\Pi(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\Pi(\Sigma_1^{-1})\Pi(\Sigma_2^{-1})d\boldsymbol{\mu}_1 d\boldsymbol{\mu}_2 d\Sigma_1^{-1}$$

$$= \int \frac{1}{\Gamma_p\left(\frac{1}{2}v\right)}|2\Gamma|^{-\frac{1}{2}v}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}|\Phi|^{-1}\gamma^{\frac{pk_2}{2}}|\Sigma_1^{-1}|^{\frac{v+n-p-1}{2}}|k_1\Sigma_1^{-1}+\Phi^{-1}|^{-\frac{1}{2}}|k_2\gamma\Sigma_1^{-1}+\Phi^{-1}|^{-\frac{1}{2}}$$

$$e^{-\frac{1}{2}\left[tr\Gamma^{-1}\Sigma_1^{-1}+2\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta}+\sum^{k_1}\boldsymbol{x}_i'\Sigma_1^{-1}\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_{1\Sigma_1}(k_1\Sigma_1^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{1\Sigma_1}+\right.}$$

$$\left.\sum^{k_2}\boldsymbol{x}_i'\gamma\Sigma_1^{-1}\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_{2\gamma\Sigma_1}(k_2\gamma\Sigma_1^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{2\gamma\Sigma_1}\right]d\Sigma_1^{-1}$$

where

$$\hat{\boldsymbol{\mu}}_{1\Sigma_1} = (k_1\Sigma_1^{-1}+\Phi^{-1})^{-1}(k_1\Sigma_1^{-1}\overline{\boldsymbol{x}}_{1k}+\Phi^{-1}\boldsymbol{\theta}) \tag{2.2.44}$$

and

$$\hat{\boldsymbol{\mu}}_{2\gamma\Sigma_1} = (k_2\gamma\Sigma_1^{-1}+\Phi^{-1})^{-1}(k_2\gamma\Sigma_1^{-1}\overline{\boldsymbol{x}}_{2k}+\Phi^{-1}\boldsymbol{\theta}). \tag{2.2.45}$$

This integral is numerically very complex and analytically intractable. A possible solution is to assume that Φ is proportional to Σ_1, i.e. $\Phi^{-1}=\delta\Sigma_1^{-1}$ where $\delta>0$.

It then follows that

$$f(X|k,\boldsymbol{\theta},\delta,\gamma)$$

$$= \frac{1}{\Gamma_p(\frac{1}{2}v)}|2\Gamma|^{-\frac{1}{2}v}c_{1p}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}\gamma^{\frac{pk_2}{2}}\delta^p\left[(k_1+\delta)(k_2\gamma+\delta)\right]^{-\frac{p}{2}}|\Gamma^{-1}+2\delta\boldsymbol{\theta}\boldsymbol{\theta}'+$$

$$\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'-(k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}'+\gamma\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'-(k_2\gamma+\delta)\hat{\boldsymbol{\mu}}_{2\gamma\delta}\hat{\boldsymbol{\mu}}_{2\gamma\delta}'|^{-\left(\frac{v+n}{2}\right)} \tag{2.2.46}$$

where $\hat{\boldsymbol{\mu}}_{1\delta}$ is as in (2.2.11), c_{1p} is given by (2..2.14) and

$$\hat{\boldsymbol{\mu}}_{2\gamma\delta} = (k_2\gamma + \delta)^{-1}(k_2\gamma\overline{\boldsymbol{x}}_{2k} + \delta\boldsymbol{\theta}).$$

The marginal mass function of k is then

$$\Pi(k|X,\boldsymbol{\theta},\delta,\gamma) = \frac{f(X|\boldsymbol{\theta},\delta,\gamma,k)\Pi(k)}{\sum_{k=p+1}^{n-p-1} f(X|\boldsymbol{\theta},\delta,\gamma,k)\Pi(k)} \tag{2.2.47}$$

and

$$\Sigma_1^{-1}|k,X,\boldsymbol{\theta},\delta,\gamma \sim W\left(v+n, \left[\Gamma^{-1} + 2\delta\boldsymbol{\theta}\boldsymbol{\theta}' + \sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i' + \gamma\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i' - (k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}' - \right.\right.$$

$$\left.\left.(k_2\gamma+\delta)\hat{\boldsymbol{\mu}}_{2\gamma\delta}\hat{\boldsymbol{\mu}}_{2\gamma\delta}'\right]^{-1}\right).$$

With the prior $\Pi(\gamma) \propto \frac{1}{\gamma}$, $0 < \gamma < A$ where the upper limit on γ can be chosen large enough and is just to ensure a proper posterior for γ, it follows that

$$\Pi(\gamma|k,X,\boldsymbol{\theta},\delta) \propto \gamma^{\frac{pk_2}{2}-1}(k_2\gamma+\delta)^{-p/2}|\Gamma^{-1} + 2\delta\boldsymbol{\theta}\boldsymbol{\theta}' + \sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i' + \gamma\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i' - (k_1+\delta)\hat{\boldsymbol{\mu}}_{1\delta}\hat{\boldsymbol{\mu}}_{1\delta}' -$$

$$(k_2\gamma+\delta)\hat{\boldsymbol{\mu}}_{2\gamma\delta}\hat{\boldsymbol{\mu}}_{2\gamma\delta}'\Big|^{-(\frac{v+n}{2})}.$$

We want to find the posterior distribution of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ where (for $j = 1,2$)

$$\Pi(\boldsymbol{\mu}_j|X,\boldsymbol{\theta},\delta,\gamma,\Sigma_1) = \sum_{k=1}^{n-1}\Pi(\boldsymbol{\mu}_j|k,X,\boldsymbol{\theta},\delta,\gamma,\Sigma_1)\Pi(k|X,\boldsymbol{\theta},\delta,\gamma,\Sigma_1) \tag{2.2.48}$$

and where

$$\boldsymbol{\mu}_1|k,X,\boldsymbol{\theta},\delta,\gamma,\Sigma_1 \sim N\left(\hat{\boldsymbol{\mu}}_{1\delta}, \frac{\Sigma_1}{k_1+\delta}\right)$$

and

$$\boldsymbol{\mu}_2|k, X, \boldsymbol{\theta}, \delta, \gamma, \Sigma_1 \sim N\left(\hat{\boldsymbol{\mu}}_{2\delta\gamma}, \frac{\Sigma_1}{k_2\gamma + \delta}\right).$$

Also notice from (2.2.48) that the posteriors of $\boldsymbol{\mu}_j$, given $X, \boldsymbol{\theta}, \delta, \gamma$ and $\Sigma_1 (j = 1, 2)$ are mixtures of normal distributions.

Under the assumption of having a change-point, vague improper priors can once again be used. By letting $\delta \to 0$, $\Gamma^{-1} \to 0$ and $v \to o$ in equation (2.2.47), the posterior of k simplifies to

$$\Pi(k|X, \gamma) = \frac{\gamma^{\frac{pk_2}{2}}(k_1k_2\gamma)^{-p/2}|S_{1k} + \gamma S_{2k}|^{-n/2}}{\sum_{k=p+1}^{n-p-1} \gamma^{\frac{pk_2}{2}}(k_1k_2\gamma)^{-p/2}|S_{1k} + \gamma S_{2k}|^{-n/2}},$$

assuming $\pi(k) = \frac{1}{n-2p+1}, k = p+1, \cdots, n-p+1$.

Furthermore

$$\Sigma_1^{-1}|k, X, \gamma \sim W(n, [S_{1k} + \gamma S_{2k}]^{-1})$$

and

$$\Pi(\gamma|k, X) \propto \gamma^{\frac{pk_2}{2}-1}(k_2\gamma)^{-p/2}|S_{1k} + \gamma S_{2k}|^{-n/2}, \quad 0 < \gamma < A.$$

Also notice that

$$\boldsymbol{\mu}_1|k, X, \gamma \sim t_p(n+1-p, \overline{\boldsymbol{x}}_{1k}, k_1(n+1-p)(S_{1k} + \gamma S_{2k})^{-1})$$

and

$$\boldsymbol{\mu}_2|k, X, \gamma \sim t_p(n+1-p, \overline{\boldsymbol{x}}_{2k}, \gamma k_2(n+1-p)(S_{1k} + \gamma S_{2k})^{-1}).$$

Another way of dealing with the problem of unknown hyperparameters is to use once again a hierarchical model where the second stage priors are vague. For example let $\Pi(\boldsymbol{\theta}) \propto 1$ and $\Pi(\delta) \propto \frac{1}{\delta}$.

We know from (2.2.46), after completing the square, that

$$f(X|k,\delta,\gamma)$$
$$= \frac{1}{\Gamma_p\left(\frac{1}{2}v\right)}|2\Gamma|^{-\frac{1}{2}v}c_{1p}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}\gamma^{\frac{pk_2}{2}}\delta^p(k_1+\delta)^{-\frac{p}{2}}(k_2\gamma+\delta)^{-\frac{p}{2}}\int|\Gamma^{-1}+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'+\gamma\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'$$

$$-(k_1+\delta)^{-1}k_1^2\overline{\boldsymbol{x}}_{1k}\overline{\boldsymbol{x}}_{1k}'-(\gamma k_2+\delta)^{-1}k_2^2\gamma^2\overline{\boldsymbol{x}}_{2k}\overline{\boldsymbol{x}}_{2k}'-\hat{\boldsymbol{\theta}}_{\delta\gamma}[2\delta-(\gamma k_2+\delta)^{-1}\delta^2-(k_1+\delta)^{-1}\delta^2]\hat{\boldsymbol{\theta}}_{\delta\gamma}'$$

$$+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{\delta\gamma})[2\delta-(k_2\gamma+\delta)^{-1}\delta^2-(k_1+\delta)^{-1}\delta^2](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{\delta\gamma})'|^{-\left(\frac{v+n}{2}\right)}\Pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

so that

$$\Pi(k|X,\gamma,\delta)$$
$$=\frac{1}{\Gamma_p(\frac{1}{2}v)}|2\Gamma|^{-\frac{1}{2}v}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}c_{1p}\gamma^{\frac{pk_2}{2}}\delta^{p-\frac{1}{2}}(k_1+\delta)^{-\left(\frac{p-1}{2}\right)}(\gamma k_2+\delta)^{-\left(\frac{p-1}{2}\right)}$$

$$\Gamma\left(\frac{v+n-p}{2}\right)\left[\Gamma\left(\frac{v+n}{2}\right)\right]^{-1}\pi^{\frac{p}{2}}\left[2\gamma k_1k_2+\delta k_1+\delta\gamma k_2\right]^{-\frac{1}{2}}|T_{\delta\gamma\Gamma}|^{-\left(\frac{v+n-p}{2}\right)} \qquad (2.2.49)$$

where c_{1p} is given by (2.2.14),

$$\hat{\boldsymbol{\theta}}_{\delta\gamma}=[2-(\gamma k_2+\delta)^{-1}\delta-(k_1+\delta)^{-1}\delta]^{-1}[k_1(k_1+\delta)^{-1}\overline{\boldsymbol{x}}_{1k}+\gamma k_2(\gamma k_2+\delta)^{-1}\overline{\boldsymbol{x}}_{2k}]$$

and

$$T_{\delta\gamma\Gamma} = \Gamma^{-1}+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i'+\gamma\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i'-(k_1+\delta)^{-1}k_1^2\overline{\boldsymbol{x}}_{1k}\overline{\boldsymbol{x}}_{1k}'-(\gamma k_2+\delta)^{-1}k_2^2\gamma^2\overline{\boldsymbol{x}}_{2k}\overline{\boldsymbol{x}}_{2k}'-$$

$$\hat{\boldsymbol{\theta}}_{\delta\gamma}[2\delta-(\gamma k_2+\delta)^{-1}\delta^2-(k_1+\delta)^{-1}\delta^2]\hat{\boldsymbol{\theta}}_{\delta\gamma}',\quad 0<\gamma<A.$$

Furthermore, by letting $\Gamma^{-1}\to 0$ and $v\to 0$, it follows that

$$\Pi(\gamma|k,X)\propto\int\delta^{p-\frac{3}{2}}\gamma^{\frac{pk_2-2}{2}}(k_1+\delta)^{-\frac{(p-1)}{2}}(\gamma k_2+\delta)^{-\left(\frac{p-1}{2}\right)}[2\gamma k_1k_2+\delta k_1+\delta\gamma k_2]^{-\frac{1}{2}}|T_{\delta\gamma}|^{-\left(\frac{n-p}{2}\right)}d\delta$$

where $T_{\delta\gamma}$ is the same as $T_{\delta\gamma\Gamma}$, but with Γ^{-1} omitted. This can be integrated numerically. Another possible solution is to use Gibbs sampling, where the full conditional distributions are as follows:

$$\boldsymbol{\mu}_1|k, X, \boldsymbol{\theta}, \delta, \Sigma_1, \boldsymbol{\mu}_2 \sim N\left(\frac{k_1\overline{\boldsymbol{x}}_{1k} + \delta\boldsymbol{\theta}}{k_1+\delta}, \quad \left(\frac{1}{k_1+\delta}\right)\Sigma_1\right),$$

$$\boldsymbol{\mu}_2|k, X, \boldsymbol{\theta}, \delta, \Sigma_1, \gamma, \boldsymbol{\mu}_1 \sim N\left(\frac{k_2\overline{\boldsymbol{x}}_{2k} + \delta\boldsymbol{\theta}}{k_2+\delta}, \quad \left(\frac{1}{\gamma k_2+\delta}\right)\Sigma_1\right),$$

$$\Pi(k|X, \Sigma_1, \gamma, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \frac{e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) + \gamma\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\right]}}{\sum_{k=p+1}^{n-p-1} e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) + \gamma\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\right]}},$$

$$\delta|k, X, \boldsymbol{\theta}, \Sigma_1, \gamma, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \sim \text{Gamma}\left(a+p, \quad b + \tfrac{1}{2}tr\Sigma_1^{-1}\left[(\boldsymbol{\mu}_1 - \boldsymbol{\theta})(\boldsymbol{\mu}_1 - \boldsymbol{\theta})' + (\boldsymbol{\mu}_2 - \boldsymbol{\theta})(\boldsymbol{\mu}_2 - \boldsymbol{\theta})'\right]\right)$$

where a proper prior $\delta \sim \Gamma(a, b)$ is put on δ to ensure that the simulation process won't go out of control. Furthermore

$$\boldsymbol{\theta}|k, \Sigma_1, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \delta \sim N\left(\frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2}, \frac{1}{2}\delta^{-1}\Sigma_1\right),$$

$$\gamma|k, X, \Sigma_1, \boldsymbol{\mu}_1\ \boldsymbol{\mu}_2, \delta \sim \quad \text{Gamma} \quad \left(\frac{pk_2}{2}, \frac{1}{2}\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\right)$$

and by letting $\Gamma^{-1} \to 0$ and $v \to 0$ it follows that

$$\Sigma_1^{-1}|k, X, \gamma, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \delta \quad \sim \quad W\left(n+2, \left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)(\boldsymbol{x}_i - \boldsymbol{\mu}_1)' + \delta(\boldsymbol{\mu}_1 - \boldsymbol{\theta})(\boldsymbol{\mu}_1 - \boldsymbol{\theta})' + \delta \sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\right]^{-1}\right). \tag{2.2.50}$$

With reference to a more general model, a proper prior can be put on Φ^{-1} so as to obtain proper posteriors, e.g. $\Phi^{-1} \sim W(\rho, \Psi)$. So once again Gibbs sampling can be applied, where the full conditional distributions are now as follows:

$$\Pi(k|X, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \gamma) =$$

$$\frac{\gamma^{\frac{pk_2}{2}} e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) + \gamma\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\right]}}{\sum_{k=p+1}^{n-p-1} \gamma^{\frac{pk_2}{2}} e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) + \gamma\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\right]}},$$

$$\boldsymbol{\mu}_1|k, X, \boldsymbol{\theta}, \Sigma_1, \gamma, \Phi, \boldsymbol{\mu}_2 \sim N((k_1\Sigma_1^{-1} + \Phi^{-1})^{-1}(k_1\Sigma_1^{-1}\overline{\boldsymbol{x}}_{1k} + \Phi^{-1}\boldsymbol{\theta}), (k_1\Sigma_1^{-1} + \Phi^{-1})^{-1}),$$

$$\boldsymbol{\mu}_2|k, X, \boldsymbol{\theta}, \Sigma_1, \gamma, \Phi, \boldsymbol{\mu}_1 \sim N((k_2\gamma\Sigma_1^{-1} + \Phi^{-1})^{-1}(k_2\gamma\Sigma_1^{-1}\overline{\boldsymbol{x}}_{2k} + \Phi^{-1}\boldsymbol{\theta}), (k_2\gamma\Sigma_1^{-1} + \Phi^{-1})^{-1}),$$

$$\boldsymbol{\theta}|k, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \gamma, \Phi \sim N\left(\frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2}, \frac{1}{2}\Phi\right),$$

$$\gamma|k, X, \Sigma_1, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \sim \quad \text{Gamma} \quad \left(\frac{pk_2}{2}, \frac{1}{2}\sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)\right),$$

$$\Phi^{-1}|k, \boldsymbol{\theta}, \Sigma_1, \gamma, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \sim W(\rho + 2, [\Psi^{-1} + (\boldsymbol{\mu}_1 - \boldsymbol{\theta})(\boldsymbol{\mu}_1 - \boldsymbol{\theta})' + (\boldsymbol{\mu}_2 - \boldsymbol{\theta})(\boldsymbol{\mu}_2 - \boldsymbol{\theta})']^{-1})$$

and after letting $\Gamma^{-1} \to 0$ and $v \to 0$,

$$\Sigma_1^{-1}|k, X, \gamma, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \sim W\left(n, \left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)(\boldsymbol{x}_i - \boldsymbol{\mu}_1)' + \right.\right.$$

$$\left.\left.\gamma \sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu}_2)(\boldsymbol{x}_i - \boldsymbol{\mu}_2)'\right]^{-1}\right). \tag{2.2.51}$$

2.2.3 A CHANGE IN THE VARIANCE

In his study of the same model mentioned in paragraph 2.2.2, Menzefricke (1981) also considered the case $\mu = \mu_1 = \mu_2$ where μ is unknown, i.e. just a change in the variance occurred. Suppose again that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independent normal random vectors such that

$$\boldsymbol{X}_i \sim \begin{cases} iN(\boldsymbol{\mu}, \Sigma_1); & i = 1, \ldots, k \\ iN(\boldsymbol{\mu}, \Sigma_2); & i = k+1, \ldots, n \end{cases} \tag{2.2.52}$$

where $n \geq 2p$, $\boldsymbol{x}_i \in I\!R^p$ and $p \leq k \leq n - p$.

Furthermore $\boldsymbol{\mu}$ (i.e. $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$) and the covariance matrices Σ_1 and $\Sigma_2 (\Sigma_1 \neq \Sigma_2)$ are unknown. The marginal prior densities of k and $\boldsymbol{\mu}$ are the same as previously. The joint distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ conditional on $\boldsymbol{\mu}, \Sigma_1, \Sigma_2$ and the change having taken place at k is given by

$$f(X|k, \boldsymbol{\mu}, \Sigma_1, \Sigma_2) = \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Sigma_1|^{-\frac{k_1}{2}} |\Sigma_2|^{-\frac{k_2}{2}}$$

$$e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i - \boldsymbol{\mu})'\Sigma_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}) + \sum^{k_2}(\boldsymbol{x}_i - \boldsymbol{\mu})'\Sigma_2^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right]}$$

with the priors on Σ_1 and Σ_2,

$$\Pi(\Sigma_i^{-1}) \propto |\Sigma_i^{-1}|^{\frac{v-p-1}{2}} e^{-\frac{1}{2} tr \Gamma^{-1} \Sigma_i^{-1}}, \quad i = 1, 2. \tag{2.2.53}$$

After completing the square, it follows that

$f(X,\Sigma_1,\Sigma_2|\Phi,k,\boldsymbol{\theta},\Gamma,v)$

$$= \int_{-\infty}^{\infty} \frac{1}{[\Gamma_p(\frac{1}{2}v)]^2} |2\Gamma|^{-v} \left(\frac{1}{2\pi}\right)^{\frac{p(n+1)}{2}} |\Sigma_1^{-1}|^{\frac{k_1+v-p-1}{2}} |\Sigma_2^{-1}|^{\frac{k_2+v-p-1}{2}} |\Phi|^{-\frac{1}{2}} e^{-\frac{1}{2}tr\Gamma^{-1}\Sigma_1^{-1}}$$

$$e^{-\frac{1}{2}tr\Gamma^{-1}\Sigma_2^{-1}} \int e^{\frac{1}{2}[(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2\theta})'(k_1\Sigma_1^{-1}+k_2\Sigma_2^{-1}+\Phi^{-1})(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2\theta})} + \sum^{k_1} \boldsymbol{x}_i'\Sigma_1^{-1}\boldsymbol{x}_i + \sum^{k_2} \boldsymbol{x}_i'\Sigma_2^{-1}\boldsymbol{x}_i$$

$$+\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta} - \hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2\theta}'(k_1\Sigma_1^{-1}+k_2\Sigma_2^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2\theta}]d\boldsymbol{\mu}$$

where

$$\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2\theta} = \left[k_1\Sigma_1^{-1}+k_2\Sigma_2^{-1}+\Phi^{-1}\right]^{-1}\left[k_1\Sigma_1^{-1}\overline{\boldsymbol{x}}_{1k}+k_2\Sigma_2^{-1}\overline{\boldsymbol{x}}_{2k}+\Phi^{-1}\boldsymbol{\theta}\right].$$

It now follows that

$$f(X|k,\boldsymbol{\theta},\Phi,\Gamma,v) = \frac{1}{[\Gamma_p(\frac{1}{2}v)]^2}|2\Gamma|^{-v}\iint\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}|k_1\Sigma_1^{-1}+k_2\Sigma_2^{-1}+\Phi^{-1}|^{-\frac{1}{2}}|\Sigma_1^{-1}|^{\frac{k_1+v-p-1}{2}}$$

$$|\Sigma_2^{-1}|^{\frac{k_2+v-p-1}{2}}|\Phi|^{-\frac{1}{2}}\, e^{-\frac{1}{2}\left[tr\Gamma^{-1}\Sigma_1^{-1}+tr\Gamma^{-1}\Sigma_2^{-1}\right]}.$$

$$e^{-\frac{1}{2}\left[\sum^{k_1}\boldsymbol{x}_i'\Sigma_1^{-1}\boldsymbol{x}_i+\sum^{k_2}\boldsymbol{x}_i'\Sigma_2^{-1}\boldsymbol{x}_i+\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta}-\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2\theta}'(k_1\Sigma_1^{-1}+k_2\Sigma_2^{-1}+\Phi^{-1})\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2\theta}\right]}d\Sigma_1^{-1}d\Sigma_2^{-1}. \tag{2.2.54}$$

In this equation the matrix integral is intractable and Φ^{-1} is inestimable. A first simplification can be to put a uniform prior on $\boldsymbol{\mu}$, i.e. to let $\Phi^{-1}\to 0$. A second step could be to let $\Gamma^{-1}\to 0$ and $v\to 0$ under the assumption that $p\le k\le n-p$, which is equivalent to putting vague priors on Σ_1^{-1} and Σ_2^{-1}. (The situation where a vague prior is put on Γ^{-1} is discussed in the univariate case in paragraph 2.7.3). Then (2.2.54) reduces to

$$f(X|k) \propto \iint \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |k_1\Sigma_1^{-1} + k_2\Sigma_2^{-1}|^{-\frac{1}{2}} |\Sigma_1^{-1}|^{\frac{k_1-p-1}{2}} |\Sigma_2^{-1}|^{\frac{k_2-p-1}{2}}$$

$$e^{-\frac{1}{2}\left[\sum^{k_1} \boldsymbol{x}_i'\Sigma_1^{-1}\boldsymbol{x}_i + \sum^{k_2} \boldsymbol{x}_i'\Sigma_2^{-1}\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2}'(k_1\Sigma_1^{-1} + k_2\Sigma_2^{-1})\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2}\right]} d\Sigma_1^{-1} d\Sigma_2^{-1}$$

where

$$\hat{\boldsymbol{\mu}}_{\Sigma_1\Sigma_2} = \left[k_1\Sigma_1^{-1} + k_2\Sigma_2^{-1}\right]^{-1} \left[k_1\Sigma_1^{-1}\overline{\boldsymbol{x}}_{1k} + k_2\Sigma_2^{-1}\overline{\boldsymbol{x}}_{2k}\right].$$

The only way to integrate this analytically is to make a further simplification as in the previous sections, assuming that $\Sigma_2^{-1} = \gamma\Sigma_1^{-1}$. This means that the prior on Σ_2^{-1} is replaced by a prior on γ where $\Pi(\gamma) \propto \frac{1}{\gamma}$. Then

$$f(X|k,\gamma) = c_{7p}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}} (k_1 + k_2\gamma)^{-\frac{p}{2}} \gamma^{\frac{p(k_2-p-1)}{2}} |T_\gamma|^{-\left(\frac{n-p-2}{2}\right)} \tag{2.2.55}$$

where

$$c_{7p} = 2^{\frac{p(n-p-2)}{2}} \Gamma_p\left(\frac{1}{2}(n-p-2)\right),$$

$$T_\gamma = \sum^{k_1} \boldsymbol{x}_i\boldsymbol{x}_i' + \gamma\sum^{k_2} \boldsymbol{x}_i\boldsymbol{x}_i' - \hat{\boldsymbol{\mu}}_\gamma(k_1 + k_2\gamma)\hat{\boldsymbol{\mu}}_\gamma'$$

and

$$\hat{\boldsymbol{\mu}}_\gamma = (k_1 + k_2\gamma)^{-1}(k_1\overline{\boldsymbol{x}}_{1k} + k_2\gamma\overline{\boldsymbol{x}}_{2k}).$$

The marginal mass function of k is then

$$\Pi(k|X,\gamma) = \frac{f(X|k,\gamma)\Pi(k)}{\sum_{k=p}^{n-p} f(X|k,\gamma)\Pi(k)}.$$

Furthermore

$$\Pi(\boldsymbol{\mu}|X,\gamma) = \sum_{k=p}^{n-p} \Pi(\boldsymbol{\mu}|k,X,\gamma)\Pi(k|X,\gamma)$$

where

$$\Pi(\boldsymbol{\mu}|k,X,\gamma) \propto \int_0^\infty \Pi(\boldsymbol{\mu}|k,\Sigma_1^{-1},X,\gamma)\Pi(\Sigma_1^{-1}|k,X,\gamma)d\Sigma_1^{-1}$$

and

$$\boldsymbol{\mu}|k,\Sigma_1^{-1},X,\gamma \sim N\left(\hat{\boldsymbol{\mu}}_\gamma, \frac{\Sigma_1}{k_1+\gamma k_2}\right)$$

and

$$\Sigma_1^{-1}|k,X,\gamma \sim W(n-p-2, T_\gamma^{-1}).$$

Therefore

$$\boldsymbol{\mu}|k,X,\gamma \sim t_p(n-2p-1, \hat{\boldsymbol{\mu}}_\gamma, (n-2p-1)(k_1+\gamma k_2)T_\gamma^{-1})$$

and also notice that

$$\Pi(\gamma|k,X) \propto (k_1+k_2\gamma)^{-\frac{p}{2}}\gamma^{p\frac{(k_2-p-1)}{2}-1}|T_\gamma|^{-(\frac{n-p-2}{2})}.$$

Another possible solution is to use Gibbs sampling. Once again let $\Phi^{-1} \sim W(\rho,\Psi)$, but now it is not necessary to include the restriction $\Sigma_2^{-1} = \gamma\Sigma_1^{-1}$. The full conditional distributions are then as follows:

$$\Pi(k|X,\boldsymbol{\mu},\gamma,\Sigma_1) \propto \gamma^{\frac{k_2+v-p-1}{2}} e^{-\frac{1}{2}\left[\sum^{k_1}(\boldsymbol{x}_i-\boldsymbol{\mu})'\Sigma_1^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) + \gamma\sum^{k_2}(\boldsymbol{x}_i-\boldsymbol{\mu})'\Sigma_1^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right]},$$

$$\boldsymbol{\mu}|k,X,\boldsymbol{\theta},\gamma,\Sigma_1,\Phi \sim N(\hat{\boldsymbol{\mu}}_{\Sigma_1\gamma\theta}, (k_1\Sigma_1^{-1}+k_2\gamma\Sigma_1^{-1}+\Phi^{-1})^{-1})$$

where

$$\hat{\mu}_{\Sigma_1\gamma\theta} = \left[k_1\Sigma_1^{-1} + k_2\gamma\Sigma_1^{-1} + \Phi^{-1}\right]^{-1}\left[k_1\Sigma_1^{-1}\overline{x}_{1k} + k_2\gamma\Sigma_1^{-1}\overline{x}_{2k} + \Phi^{-1}\theta\right],$$

$$\theta|\Phi, \mu \sim N(\mu, \Phi),$$

$$\Phi^{-1}|\mu, \theta \sim W\left(p + 2 + \rho, [(\mu - \theta)(\mu - \theta)' + \Psi^{-1}]^{-1}\right),$$

$$\gamma|k, X, \mu, \Sigma_1 \sim \text{Gamma}\left(\frac{p(n-k+v-p-1)}{2}, \frac{1}{2}tr\left\{\left[\Gamma^{-1} + \sum^{k_2}(x_i - \mu)(x_i - \mu)'\right]\Sigma_1^{-1}\right\}\right)$$

and

$$\begin{aligned}\Sigma_1^{-1}|k, X, \mu, \gamma \quad &\sim \quad W\Bigg(2v + n - p - 1, \Bigg[\Gamma^{-1} + \Gamma^{-1}\gamma + \sum^{k_1}(x_i - \mu)(x_i - \mu)' \\ &\qquad + \gamma\sum^{k_2}(x_i - \mu)(x_i - \mu)'\Bigg]^{-1}\Bigg). \qquad (2.2.56)\end{aligned}$$

2.3. NO OR ONE CHANGE-POINT IN THE MULTIVARIATE MODEL

In section 2.2 three normal models are discussed under the assumption of exactly one change-point. The fixed number of change-points simplifies matters greatly, the parameter spaces are of the same dimensions and the interpretation of parameters remains the same under all models and for all values of k. The possibility of no change is however very important in most applications and should not be ignored.

The problem of no change versus one or more changes in the normal model has been covered less widely. Broemeling and Tsurumi (1987), however, consider no change versus a change in the linear model, while Diaz (1982) also studied the case of no change. A possible reason for the few references on no change versus a change is perhaps because of the problem caused by the differing dimensions.

In this section we will add the option of a model with no change (say $k = n$) to the previous models, and assign it the prior probability q. In the applications q will usually be taken as $\frac{1}{2}$, so as to give equal prior weight to "no change" versus "a change".

The marginal prior density of k is then

$$\Pi(k) = \begin{cases} q & for \quad k = n \\ \frac{1-q}{n-1} & for \quad k = 1, 2, \ldots, n-1 \end{cases} . \tag{2.3.1}$$

The case $k = n$ has fewer parameters than for $1 \leq k \leq n-1$, and the marginal of X is the same for all three normal models. To conform with the prior specifications in paragraph 2.2, let (for $i = 1, \ldots, n$)

$$X_i \sim iN(\boldsymbol{\mu}_0, \Sigma) \quad \text{if} \quad k = n$$

where

$$\boldsymbol{\mu}_0 \sim N(\boldsymbol{\theta}, \Phi) \quad \text{and} \quad \Sigma^{-1} \sim W(v, \Gamma).$$

We let $\Pi(\boldsymbol{\theta}) \propto 1$ (as in paragraph 2.2) and justify it by arguing that $\boldsymbol{\theta}$ has the same interpretation here as in the case of one change-point, although $\boldsymbol{\theta}$ is not estimable in this case. In fact, putting a vague prior on $\boldsymbol{\theta}$ here is equivalent to a vague prior on $\boldsymbol{\mu}_0$.

The marginal of X is then independent of Φ and given by

$$f(X|k = n, \Gamma, v) = \pi^{-\frac{p(n-1)}{2}} n^{-\frac{1}{2}} \frac{\Gamma_p\left[\frac{1}{2}(n + v - 1)\right]}{\Gamma_p(\frac{1}{2}v)} |\Gamma^{-1}|^{\frac{1}{2}v} |S_n + \Gamma^{-1}|^{-\left(\frac{n+v-1}{2}\right)}. \tag{2.3.2}$$

If $\Gamma^{-1} \to 0$ and $v \to 0$, (2.3.2) reduces to

$$f(X|k = n) = c_n n^{-\frac{1}{2}} |S_n|^{-\left(\frac{n-1}{2}\right)} \tag{2.3.3}$$

where

$$c_n = \pi^{-\frac{p(n-1)}{2}} \Gamma_p\left[\frac{1}{2}(n-1)\right]$$

and

$$S_n = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)'.$$

The marginal posterior distributions of $\boldsymbol{\mu}_0$ and Σ^{-1} under no change-point follows directly as

$$\boldsymbol{\mu}_0 | k = n, X \sim t(n-p, \overline{\boldsymbol{x}}_n, n(n-p)S_n^{-1})$$

and

$$\Sigma^{-1} | k = n, X \sim W(n-1, S_n^{-1}).$$

2.3.1 A POSSIBLE CHANGE IN THE MEAN

The posterior distribution of k is given by

$$\Pi(k = n | X, \delta) = \frac{q f(X | k = n)}{\sum_{k=1}^{n} \left[\frac{1-q}{n-1} f(X | k, \delta) \right] + q f(X | k = n)}$$

$$\Pi(k | X, \delta) = \frac{\frac{1-q}{n-1} f(X | k, \delta)}{\sum_{k=1}^{n} \left[\frac{1-q}{n-1} f(X | k, \delta) \right] + q f(X | k = n)}, \quad k = 1, \ldots, n-1 \tag{2.3.4}$$

where $f(X | k = n)$ is given in (2.3.3) and $f(X | k, \delta)$ in (2.2.20).

In the case of exactly one change-point, we let $\delta \to 0$, i.e. putting a vague prior on $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. But δ can't approach zero when we are comparing the case "no change" with "one change", because of the different dimensions. This can be seen from (2.3.4) where $\Pi(k = n | X, \delta) \to 1$ as $\delta \to 0$.

If $\delta \to \infty$, $\Pi(k = n | X, \delta) \to q$, so the posterior probability for no change is crucially dependent on δ. This is an example of Lindley's paradox (Lindley (1957)) where, for fixed data, the Bayes factor in favour of a certain hypothesis can be manipulated to any degree by appropriate choices of hyperparameters. The limiting results above however make intuitive sense. If $\delta \to 0$, it means an infinitely large prior variance for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, which makes it impossible to detect a change-point from a fixed finite data set, so the probability for no change goes to one. On the other hand, if $\delta \to \infty$, the two means are virtually the same, and the data can't give any information. So the posterior equals the prior probability.

In general, we will consider three possible solutions:

(a) Examine the posterior probabilities as a function of δ. The sensitivity of these probabilities to δ can then be determined, as well as certain upper and lower bounds.

(b) Use a vague prior on δ, say $\Pi(\delta) \propto \frac{1}{\delta}$ over a reasonable finite range, $0 < \delta < A$, and integrate δ out. However, the posterior probabilities can also be sensitive to the choice of A.

(c) Use partial Bayes factors. This approach will be treated in paragraph 2.5.

2.3.2. A POSSIBLE CHANGE IN THE MEAN AND VARIANCE

For case 1 in paragraph 2.2.2 where $\boldsymbol{\mu}_i \sim N(\boldsymbol{\theta}_i, \Phi_i)$ and $\Phi_i^{-1} = \delta_i \Sigma_i^{-1}$ for $i = 0, 1, 2$, $f(X|k, \delta_1, \delta_2, \Gamma, v)$ is given in (2.2.37). Together with equations (2.3.2) and (2.3.4), the posterior probability of k follows as

$$
\begin{aligned}
\Pi(k = n|X, \Gamma, v) \quad &\propto \quad q n^{-\frac{1}{2}} \Gamma_p\left(\tfrac{1}{2}(n + v - 1)\right) |S_n + \Gamma^{-1}|^{-\left(\frac{n+v-1}{2}\right)}, \\
\Pi(k|X, \delta_1, \delta_2, \Gamma, v) \quad &\propto \quad \tfrac{1-q}{n-1} \tfrac{1}{\Gamma_p(\frac{1}{2}v)} |\Gamma|^{-\frac{1}{2}v} 2^{\frac{-p(v+n)}{2}} c_{6p} \pi^{\frac{p}{2}} (k_1 k_2)^{-\frac{1}{2}} (\delta_1 \delta_2)^{\frac{p-1}{2}} \\
&\quad [(k_1 + \delta_1)(k_2 + \delta_2)]^{-\left(\frac{p-1}{2}\right)} \Gamma\left(\tfrac{v+k_1-p}{2}\right) \Gamma\left(\tfrac{v+k_2-p}{2}\right) \left[\Gamma\left(\tfrac{v+k_1}{2}\right) \Gamma\left(\tfrac{v+k_2}{2}\right)\right]^{-1} \\
&\quad |\Gamma^{-1} + S_{1k}|^{-\left(\frac{v+k_1-p}{2}\right)} |\Gamma^{-1} + S_{2k}|^{-\left(\frac{v+k_2-p}{2}\right)}, \quad k = 1, \ldots, n-1
\end{aligned}
\tag{2.3.5}
$$

where c_{6p} is given by (2.2.38).

A possible simplification would be to let $\delta_1 = \delta_2$. Notice that if $\delta_1 = \delta_2 \to \infty$, then (2.3.5), ignoring the normalizing constant, approaches a finite limit which corresponds to a vague improper prior on $\Phi_i, i = 1, 2$. For case 2 where $\Pi(\Sigma_1^{-1})$ is the same as in case 1, but $\Sigma_2^{-1} = \gamma\Sigma_1^{-1}$, $f(X|k, \delta, \gamma)$ follows from (2.2.49).

$\Pi(k|X, \delta, \gamma, \Gamma, v)$ then follows as in (2.3.4). The posterior distribution of k is still dependent on Γ, v. Their effects will be examined in the univariate case.

2.3.3 A POSSIBLE CHANGE IN THE VARIANCE

With vague priors on $\boldsymbol{\mu}_i$ and $\Sigma_2^{-1} = \gamma\Sigma_1^{-1}$ in equation (2.2.54), it follows that

$$\begin{aligned} f(X|k, \gamma, \Gamma, v) &= \int_{\Sigma_1^{-1}>0} \left[\frac{1}{\Gamma_p(\frac{1}{2}v)}\right] |2\Gamma|^{-\frac{1}{2}v}(2\pi)^{-\frac{np}{2}}|k_1 + k_2\gamma|^{-\frac{1}{2}}\gamma^{\frac{p(k_2+v-p-1)}{2}}|\Sigma_1^{-1}|^{\frac{n+2v-2p-3}{2}} \\ &\quad e^{-\frac{1}{2}tr\left[\Gamma^{-1}+\Gamma^{-1}\gamma+\sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i' + \gamma\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i' - (k_1+k_2\gamma)\hat{\boldsymbol{\mu}}_\gamma\hat{\boldsymbol{\mu}}_\gamma'\right]\Sigma_1^{-1}} d\Sigma_1^{-1} \end{aligned}$$

where

$$\hat{\boldsymbol{\mu}}_\gamma = [k_1 + k_2\gamma]^{-1}[k_1\overline{\boldsymbol{x}}_{1k} + k_2\gamma\overline{\boldsymbol{x}}_{2k}]$$

so that

$$\begin{aligned} f(X|k, \gamma, \Gamma, v) &= \left[\frac{1}{\Gamma_p(\frac{1}{2}v)}\right] |\Gamma|^{-\frac{1}{2}v}\pi^{-\frac{np}{2}}2^{\frac{-p(n+v)}{2}}|k_1 + k_2\gamma|^{-\frac{1}{2}}\gamma^{\frac{p(k_2+v-p-1)}{2}}c_{8p} \cdot \\ &\quad \left|\Gamma^{-1} + \Gamma^{-1}\gamma + \sum^{k_1}\boldsymbol{x}_i\boldsymbol{x}_i' + \gamma\sum^{k_2}\boldsymbol{x}_i\boldsymbol{x}_i' - (k_1+k_2\gamma)\hat{\boldsymbol{\mu}}_\gamma\hat{\boldsymbol{\mu}}_\gamma'\right|^{-\left(\frac{n+2v-p-2}{2}\right)} \end{aligned}$$

where

$$c_{8p} = 2^{\frac{p(n+2v-p-2)}{2}}\Gamma_p\left(\frac{1}{2}(n + 2v - p - 2)\right).$$

Now it's possible to let $\Gamma^{-1}, v \to 0$ when comparing with the mass function for no change-point. Then

$$\Pi(k = n|X) \propto \pi^{\frac{p}{2}} n^{-\frac{1}{2}} \Gamma_p \left[\frac{1}{2}(n-1)\right] |S_n|^{-\left(\frac{n-1}{2}\right)}$$

$$\Pi(k|X, \gamma) \propto c_{7p} 2^{-\frac{np}{2}} \left[k_1 + k_2\gamma\right]^{-\frac{1}{2}} \gamma^{\frac{p(k_2-p-1)}{2})} \left|\sum^{k_1} \boldsymbol{x}_i \boldsymbol{x}_i' + \gamma \sum^{k_2} \boldsymbol{x}_i \boldsymbol{x}_i' - (k_1 + k_2\gamma) \hat{\boldsymbol{\mu}}_\gamma \hat{\boldsymbol{\mu}}_\gamma'\right|^{-\frac{(n-p-2)}{2}} \tag{2.3.6}$$

where $c_{7p} = 2^{\frac{p(n-p-2)}{2}} \Gamma_p \left(\frac{1}{2}(n-p-2)\right)$.

2.4 COMPONENT ANALYSIS

It is of importance to see which component has the most influence in causing the change to occur at a fixed k. For the case where we have a change in the mean, let $M(p \times 2) = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2] \sim N(\boldsymbol{\theta} \mathbf{1}_2', \text{H})$ where $\text{H} = I_2 \otimes \Phi$ in the model

$$\underset{(p\times 1)}{\boldsymbol{X}_i} \; |\boldsymbol{\mu}_1, \Sigma \sim N(\boldsymbol{\mu}_1, \Sigma), \quad i = 1, \ldots, k$$

$$\underset{(p\times 1)}{\boldsymbol{X}_i} \; |\boldsymbol{\mu}_2, \Sigma \sim N(\boldsymbol{\mu}_2, \Sigma), \quad i = k+1, \ldots, n$$

where $\Sigma^{-1} \sim W(v, \Gamma)$.

The joint distribution of X conditional in M and Σ can be written in matrix notation as

$$\underset{(p\times n)}{f(X|\boldsymbol{\mu}, \Sigma)} \propto \frac{1}{|\Sigma|^{n/2}} \text{ etr } \left[-\frac{1}{2}(X - ME_k)'\Sigma^{-1}(X - ME_k)\right]$$

where

$$X = \underset{(p\times k)}{[X_1} \quad \underset{(p\times(n-k))}{X_2]}$$

and

$$E_k = \begin{pmatrix} \mathbf{1}'_k & 0 \\ 0 & \mathbf{1}'_{n-k} \end{pmatrix}.$$

The posterior distribution of M is

$$\Pi(M|X,k,\theta,\Phi) \propto \frac{1}{|\Phi|}\int \frac{1}{|\Sigma|^{\frac{n}{2}}} etr\left[-\frac{1}{2}\left\{(X-ME_k)'\Sigma^{-1}(X-ME_k)+(M-\boldsymbol{\theta}\mathbf{1}'_2)'\Phi^{-1}\right.\right.$$

$$\left.\left.(M-\boldsymbol{\theta}\mathbf{1}'_2)\right\}\right] etr\left[-\frac{1}{2}\Sigma^{-1}\Gamma^{-1}\right]|\Sigma^{-1}|^{\frac{1}{2}(\rho-p+1)}d\Sigma^{-1}. \tag{2.4.1}$$

When using a vague prior on M and $\Sigma^{-1}(\Phi^{-1}\to 0, \quad \Gamma^{-1}\to 0, \quad v\to 0)$

$$\Pi(M|X,k) \propto |X(I_n - J_n^{-1})X' + (M-\overline{X}_k)J_2(M-\overline{X}_k)'|^{-\frac{n}{2}}$$

where

$$J_2 = \underset{(2\times 2)}{E_kE'_k} = \begin{pmatrix} k & 0 \\ 0 & n-k \end{pmatrix},$$

$$J_n^{-1} = E'_k(E_kE'_k)^{-1}E_k = \begin{pmatrix} \frac{1}{k}\mathbf{1}_k\mathbf{1}'_k & 0 \\ 0 & \frac{1}{n-k}\mathbf{1}_{n-k}\mathbf{1}'_{n-k} \end{pmatrix},$$

$$\overline{X}_k = XE_k(E_kE'_k)^{-1} = [\overline{\boldsymbol{x}}_1 \quad \overline{\boldsymbol{x}}_2]$$

and

$$P = [X(I_n - J_n^{-1})X']^{-1}$$

so that

$$M \sim \text{Matrix } T(PJ_2^{-1}, n, \overline{X}_k). \tag{2.4.2}$$

If we let $\boldsymbol{m}_i = [\mu_{1i} \quad \mu_{2i}]$ be the i-th row of M, then

$$\begin{aligned}\Pi(\boldsymbol{m}_i|k,X) &\propto \left|P_{ii.2}^{-1} + (\boldsymbol{m}_i - \overline{\boldsymbol{x}}_{ki})J_2(\boldsymbol{m}_i - \overline{\boldsymbol{x}}_{ki})'\right|^{-\left(\frac{n-p+1}{2}\right)} \\ &\propto [1 + P_{ii.2}(\boldsymbol{m}_i - \overline{\boldsymbol{x}}_{ki})J_2(\boldsymbol{m}_i - \overline{\boldsymbol{x}}_{ki})']^{-\left(\frac{n-p+1}{2}\right)}\end{aligned}$$

where

$$\overline{\boldsymbol{x}}_{ki} = [x_{1i} \; x_{2i}] \quad and \quad P_{ii.2} = P_{ii} - P_{i2}P_{22}^{-1}P_{2i}$$

is the Schur complement of the (i,i)th element of P. Finally, let $\boldsymbol{\Delta} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = M\boldsymbol{c}$, i.e.

$\Delta_i = \boldsymbol{m}_i\boldsymbol{c} = \mu_{1i} - \mu_{2i}$ where $\underset{(2\times 1)}{\boldsymbol{c}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, so that it follows that

$$\Pi(\Delta_i|k,X) \propto \left[(n-p-1) + (\Delta_i - \overline{\boldsymbol{x}}_{ki}\boldsymbol{c})\left(\frac{\boldsymbol{c}'J_2^{-1}\boldsymbol{c}}{P_{ii.2}(n-p-1)}\right)^{-1}(\Delta_i - \overline{\boldsymbol{x}}_{ki}\boldsymbol{c})\right]^{-\left(\frac{n-p}{2}\right)}.$$

This is a generalized t-distribution with

$$E(\Delta_i|k,X) = \overline{\boldsymbol{x}}_{ki}\boldsymbol{c} = \overline{x}_{1i} - \overline{x}_{2i} \tag{2.4.3}$$

and

$$\text{var}(\Delta_i|k,X) = \frac{n}{k(n-k)(n-p-3)} \cdot P_{ii.2}^{-1} \tag{2.4.4}$$

so that

$$\Pi(\Delta_i|k,X) = \frac{1}{B\left(\frac{1}{2}, \frac{n-p-1}{2}\right)}\left(\frac{k(n-k)}{n}P_{ii.2}\right)^{\frac{1}{2}}\left[1 + \frac{k(n-k)P_{ii.2}}{n}(\Delta_i - (\overline{x}_{i1} - \overline{x}_{i2}))^2\right]^{-\left(\frac{n-p}{2}\right)} \tag{2.4.5}$$

represents the marginal posterior distribution of the difference between the i-th components of the means for a fixed change-point.

If $p = 1$ and we write

$$\underset{(2\times 1)}{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \sim N(\theta \mathbf{1}_2, \Phi) \quad ,$$

and

$$\underset{(n\times 1)}{\widetilde{X}} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \begin{matrix} (k\times 1) \\ ((n-k)\times 1) \end{matrix} ,$$

then

$$\mathrm{Var}(\boldsymbol{\mu}|\widetilde{X}, k) = \frac{\widetilde{X}'(I_n - J_k^{-1})\widetilde{X}}{(n-4)} J_2^{-1}.$$

If $\Delta = \mu_1 - \mu_2$ it follows that

$$\Pi(\Delta|\widetilde{X}, k) \propto \left[1 + \frac{k_1 k_2}{n\widetilde{X}'(I_n - J_k^{-1})\widetilde{X}}(\Delta - (\overline{x}_1 - \overline{x}_2))^2\right]^{-(\frac{n-1}{2})}$$

so that

$$\mathrm{Var}(\Delta) = \frac{n\widetilde{X}'(I_n - J_k^{-1})\widetilde{X}}{(n-4)k_1 k_2}.$$

In the above analysis, with a vague prior on M, no prior relationship between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ is assumed. This results in the $\Delta_i's$ being uncorrelated posteriori, each distributed as in equation (2.4.5).

To retain the prior relationship between the means, let $\Phi^{-1} = \delta\Sigma^{-1}$ and integrate $\boldsymbol{\theta}$ out in equation (2.4.1). In this case the $\Delta_i's$ are correlated and their marginal posteriors are functions of δ.

Then, with $\Gamma^{-1}, v \longrightarrow 0$, it follows that

$$\begin{aligned}\Pi(M|k, X, \delta) &\propto \int\!\!\int |\Sigma^{-1}|^{\frac{n-p+1}{2}} etr\left[-\frac{1}{2}\Sigma^{-1}\left\{(X - ME_k)(X - ME_k)' + \right.\right. \\ &\qquad \left.\left.\delta(M - \boldsymbol{\theta}\mathbf{1}_2')(M - \boldsymbol{\theta}\mathbf{1}_2')\right\}\right] d\Sigma^{-1} d\boldsymbol{\theta} \\ &\propto |XX' - XE_k'K^{-1}E_kX' + (M - XE_k'K^{-1})K(M - XE_k'K^{-1})|^{-\frac{n+1}{2}}\end{aligned}$$

where

$$K = J_2 + \delta I_2 - \frac{1}{2}\delta\mathbf{1}\mathbf{1}' = \begin{pmatrix} k + \frac{1}{2}\delta & -\frac{1}{2}\delta \\ -\frac{1}{2}\delta & n - k + \frac{1}{2}\delta \end{pmatrix}$$

so that

$$M \sim \text{ Matrix } T(P_\delta K^{-1}, n+1, XE_k'K^{-1}) \qquad (2.4.6)$$

where

$$P_\delta = [X(I_n - E_k'K^{-1}E_k)X']^{-1}.$$

It now follows that

$$\Pi(\boldsymbol{m}_i|k, X, \delta) \propto \left[(n-p) + \left(\boldsymbol{m}_i - (XE_k'K^{-1})_{(i)}\right)\left(\tfrac{K^{-1}}{P_{\delta ii.2}(n-p)}\right)^{-1}\left(\boldsymbol{m}_i - (XE_k'K^{-1})_{(i)}\right)'\right]^{-\left(\frac{n-p+2}{2}\right)}$$

where $A_{(i)}$ denotes the i-th row of the matrix A.

Then we have a generalized t-distribution,

$$\Pi(\Delta_i|k, X, \delta) \propto \left[(n-p) + \left(\frac{\boldsymbol{c}'K^{-1}\boldsymbol{c}}{(n-p)P_{\delta ii.2}}\right)^{-1}(\Delta_i - (XE_k'K^{-1})_{(i)}\boldsymbol{c})^2\right]^{-\left(\frac{n-p+1}{2}\right)}$$

where

$$c'K^{-1}c = \frac{n}{k(n-k)+\frac{1}{2}\delta n}$$

and

$$(XE_k'K^{-1})_{(i)}c = \left[\frac{1}{k(n-k)+\frac{1}{2}\delta n}\right][k(n-k)\overline{x}_{1i} - k(n-k)\overline{x}_{2i}].$$

Therefore

$$E(\Delta_i) = \left[\frac{k(n-k)}{k(n-k)+\frac{1}{2}\delta n}\right](\overline{x}_{1i} - \overline{x}_{2i}) \tag{2.4.7}$$

and

$$\text{Var}(\Delta_i) = \frac{nP_{\delta ii.2}^{-1}}{(n-p-2)[k(n-k)+\frac{1}{2}\delta n]}. \tag{2.4.8}$$

Notice that (2.4.7) reduces to (2.4.3) when $\delta \to 0$, but there is a difference of one degree of freedom between (2.4.8) and (2.4.4).

To determine the influence of each component on the specific position of a change-point, the expected differences in the means, Δ_i, can be standardised and compared. So if

$$D_i = E(\Delta_i)(\text{Var}(\Delta_i))^{-\frac{1}{2}}, \quad i = 1, \ldots, p, \tag{2.4.9}$$

we can use

$$I_i = \frac{D_i}{\Sigma D_i} \tag{2.4.10}$$

as a measure of the proportion of the overall influence.

2.5 BAYES FACTORS

2.5.1 A CHANGE IN THE MEAN

The regular Bayes factor for the model as in (2.2.10) follows from (2.2.20) and (2.3.3).

$$B_{k0} = \frac{(k+\delta)^{-\frac{p}{2}}(n-k+\delta)^{-\frac{p}{2}}\delta^{\frac{p}{2}}[2\delta-(k+\delta)^{-1}\delta^2-(n-k+\delta)^{-1}\delta^2]^{-\frac{1}{2}}|T_{\delta\Gamma}|^{-(\frac{n-p}{2})}}{(n+\delta)^{-\frac{p}{2}}[\delta-(n+\delta)^{-1}\delta^2]^{-\frac{1}{2}}|T_{0\delta}|^{-(\frac{n-p}{2})}}; \tag{2.5.1}$$

where $T_{\delta\Gamma}$ is given in (2.2.22) and $T_{0\delta}$ is the same, but with $\Gamma^{-1}=0$ and where B_{k0} denotes the Bayes factor in favour of a change-point at k versus no change. To calculate the partial Bayes factor, we use the improper priors $\Pi(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \propto 1$ and $\Pi(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(\frac{p+1}{2})}$, so that the whole sample Bayes factor is

$$B_{k0}^N = \frac{f(X|k)}{f(X|k=n)} = \frac{(2\pi)^{\frac{p}{2}}k^{-\frac{p}{2}}(n-k)^{-\frac{p}{2}}|S_{1k}+S_{2k}|^{-(\frac{n-2}{2})}}{n^{-\frac{p}{2}}|S_n|^{-(\frac{n-2}{2})}}. \tag{2.5.2}$$

Note that the difficulty with this solution is that the non-informative priors are typically improper and hence defined only up to arbitrary constants c_0. Hence B_{k0}^N is defined only up to $\frac{c_k}{c_0}$, which is itself arbitrary.

For this model, the minimal sample size if there are no change-points is $p+1$ and if there is exactly one change-point it would be $p+2$.

Booth and Smith (1982) also considered changes of the mean in normal sequences. Considering the multivariate case, they use the principle of the imaginary training sample method and define the minimal sample as $2p+1$ when $k=p$. Their result for B_{k0} is the same as ours in (2.5.2).

In order to calculate the Intrinsic Bayes factor, the minimal sample must be defined. The size of the minimal sample will be $m_\ell = m_k + 1$, where $m_k = p+1$ is the number of observations in the minimal sample that must be on the same side of the change-point k. The set $\boldsymbol{x}(\ell)$ consists of the two disjoint subsets $\boldsymbol{x}_{m_k}(\ell)$ and $\boldsymbol{x}_1(\ell)$. In the Bayes factor for the minimal sample $B_{0k}^N(\boldsymbol{x}(\ell))$, the term $k(n-k)$ always reduces to $p+1$ so that

$$B_{0k}^N(\boldsymbol{x}(\ell)) = \left[\frac{2\pi(p+2)|S_{m_\ell}|}{(p+1)|S_{m_k}|}\right]^{-\frac{p}{2}} \tag{2.5.3}$$

where

$$S_{m_k} = \sum_j (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{m_k})(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{m_k})'$$

where the summation is over the subset $\boldsymbol{x}_{m_k}(\ell)$ and $\overline{\boldsymbol{x}}_{m_k} = \frac{1}{m_k}\sum_j \boldsymbol{x}_j$ and $S_{m_\ell} = \sum_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{m_\ell})\ (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{m_\ell})'$ where the summation is over the set $\boldsymbol{x}(\ell)$.

The Arithmetic Intrinsic Bayes factor would then be

$$B_{k0}^{AI} = B_{k0}^{N} B_{0k}^{NA}(\boldsymbol{x}(\ell)) \tag{2.5.4}$$

where

$$B_{0k}^{NA}(\boldsymbol{x}(\ell)) = \frac{1}{L}\sum_{\ell=1}^{L} B_{0k}^{N}(\boldsymbol{x}(\ell)) \quad \text{where} \quad k = 1, \ldots, n-1$$

and where B_{k0}^{N} is given by (2.5.2) and $B_{0k}^{N}(\boldsymbol{x}(\ell))$ by (2.5.3).

Furthermore the geometric and the median IBF follows by inserting

$$B_{0k}^{NG}(\boldsymbol{x}(\ell)) = \left[\prod_\ell B_{0k}^{N}(\boldsymbol{x}(\ell))\right]^{\frac{1}{L}}$$

or

$$B_{0k}^{NM}(\boldsymbol{x}(\ell)) = \text{ Med}\left[B_{0k}^{N}(\boldsymbol{x}(\ell))\right].$$

To get the Fractional Bayes factor, as described in chapter 1, for this model, it follows that

$$m_k(b) = \frac{m_k}{m_k^b} = \prod_{j=1}^{p}\Gamma\left(\frac{n-1-j}{2}\right)\left[\prod_{j=1}^{p}\Gamma\left(\frac{bn-1-j}{2}\right)\right]^{-1} \pi^{-\frac{np(1-b)}{2}} b^{\frac{pbn}{2}} |S_{1k}+S_{2k}|^{-\frac{n(1-b)}{2}}$$

and

$$m_0(b) = \frac{m_0}{m_0^b} = \prod_{j=1}^{p}\Gamma\left(\frac{n-j}{2}\right)\left[\prod_{j=1}^{p}\Gamma\left(\frac{nb-j}{2}\right)\right]^{-1} \pi^{-\frac{np(1-b)}{2}} b^{\frac{pbn}{2}} |S_n|^{-\frac{n(1-b)}{2}} \tag{2.5.5}$$

so that

$$B_{k0}^F = \frac{m_k(b)}{m_0(b)} = \frac{\Gamma\left(\frac{n-1-p}{2}\right)\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{nb-1-p}{2}\right)}\left[\frac{|S_{1k}+S_{2k}|}{|S_n|}\right]^{-\frac{n(1-b)}{2}}, \quad k = 1, \ldots, n-1. \tag{2.5.6}$$

Notice that k plays no role in equation (2.5.6) except in the S_{ik}'s.

The minimal sample for a change in the mean is $p+2$. From (2.5.6) it follows that nb must be greater than p so that we must have $b > \frac{p}{n}$. If $b = \frac{p+2}{n}$ (which always satisfies $b > \frac{p}{n}$), (2.5.6) reduces to

$$B_{k0}^F = \frac{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{p}{2}\right)}\left[\frac{|S_{1k}+S_{2k}|}{|S_n|}\right]^{\frac{n-p-2}{2}}.$$

Also notice that $\Pi(k=n|X) = \left[1 + \frac{1-q}{q(n-1)}\sum_j B_{j0}\right]^{-1}$ and

$$\Pi(k|X) = \frac{1-q}{q(n-1)}B_{k0}\left[1 + \frac{1-q}{q(n-1)}\sum_j B_{j0}\right]^{-1}, \quad k = 1, \ldots, n-1. \tag{2.5.7}$$

2.5.2 A CHANGE IN THE MEAN AND VARIANCE

The regular Bayes factor for the model as in (2.2.24) follows from (2.3.5) in paragraph 2.3.2 so that with $\Gamma^{-1}, v \longrightarrow 0$ it follows that

$$B_{k0} = \left\{\pi^{\frac{p}{2}}(\delta_1\delta_2)^{\frac{p-1}{2}}[(k+\delta_1)(n-k+\delta_2)]^{-\frac{p}{2}}\Gamma\left(\frac{k-p}{2}\right)\Gamma\left(\frac{n-k-p}{2}\right)\left[\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n-k}{2}\right)\right]^{-1}\right.$$

$$\left.\left[1-(k+\delta_1)^{-1}\delta_1\right]^{-\frac{1}{2}}\left[1-(n-k+\delta_2)^{-1}\delta_2\right]^{-\frac{1}{2}}|T_{\delta_1}|^{-\frac{k-p}{2}}|T_{\delta_2}|^{-\frac{n-k-p}{2}}\right\}$$

$$\div\left\{(n+\delta)^{-\frac{p}{2}}\delta^{\frac{p-1}{2}}\Gamma\left(\frac{n-p}{2}\right)\left[\Gamma\left(\frac{n}{2}\right)\right]^{-1}[1-(n+\delta)^{-1}\delta]^{-\frac{1}{2}}|T_{0\delta}|^{-\left(\frac{n-p}{2}\right)}\right\}.$$

To calculate the partial Bayes factors, the whole sample Bayes factor, with the improper priors $\Pi(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \propto 1$ and $\Pi(\Sigma_i^{-1}) \propto |\Sigma_i^{-1}|^{-\frac{p+1}{2}} (i = 1, 2)$, is

$$B_{k0}^N = \frac{(2\pi)^{\frac{p}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\frac{k-j}{2}\right) \prod_{j=1}^{p} \Gamma\left(\frac{n-k-j}{2}\right) [k(n-k)]^{-\frac{p}{2}} [S_{1k}|^{-\left(\frac{k-1}{2}\right)} |S_{2k}|^{-\left(\frac{n-k-1}{2}\right)}}{\prod_{j=1}^{p} \Gamma\left(\frac{n-1-j}{2}\right) n^{-\frac{p}{2}} |S_n|^{-\left(\frac{n-2}{2}\right)}}. \tag{2.5.8}$$

In this case the minimal sample will be $m_\ell = 2m_k$, where $m_k = p + 1$ is the number of observations in the minimal sample which must lie on each side of the change-point k. The set $\boldsymbol{x}(\ell)$ consists of the two disjointed subsets before the change-point, $\boldsymbol{x}^1_{m_k}(\ell)$ and after the change-point, $\boldsymbol{x}^2_{m_k}(\ell)$. So

$$B_{0k}^N(\boldsymbol{x}(\ell)) = \frac{\prod_{j=1}^{p} \Gamma\left(\frac{2p+1-j}{2}\right) 2^{-\frac{p}{2}} |S_{m_\ell}|^{-p}}{(2\pi)^{\frac{p}{2}} \pi^{\frac{p(p-1)}{4}} \left[\prod_{j=1}^{p} \Gamma\left(\frac{p+1-j}{2}\right)\right]^2 (p+1)^{-\frac{p}{2}} \left[|S^1_{m_k}||S^2_{m_k}|\right]^{-\frac{p}{2}}} \tag{2.5.9}$$

where

$$S^i_{m_k} = \sum_j \left(\boldsymbol{x}_j - \overline{\boldsymbol{x}}^i_{m_k}\right)\left(\boldsymbol{x}_j - \overline{\boldsymbol{x}}^i_{m_k}\right)' \quad \text{and} \quad \overline{\boldsymbol{x}}^i_{m_k} = \frac{1}{m_k} \sum_j \boldsymbol{x}_j$$

and where the summation is over the subset $\boldsymbol{x}^i_{m_k}(\ell)$, $i = 1, 2$.

The Arithmetic Intrinsic Bayes factor for this model would then be the same as in equation (2.5.4), but where B_{k0}^N is as in (2.5.8) and $B_{0k}^N(\boldsymbol{x}(\ell))$ is as in (2.5.9).

To get the Fractional Bayes factor for this model, it follows that

$$\begin{aligned} m_k(b) &= \frac{m_k}{m_k^b} = \pi^{\frac{-np(1-b)}{2}} \prod_{j=1}^{p} \Gamma\left(\frac{k-j}{2}\right) \prod_{j=1}^{p} \Gamma\left(\frac{n-k-j}{2}\right) \\ &\left[\prod_{j=1}^{p} \Gamma\left(\frac{bk-j}{2}\right) \prod_{j=1}^{p} \Gamma\left(\frac{b(n-k)-j}{2}\right)\right]^{-1} b^{\frac{pbn}{2}} |S_{1k}|^{-\frac{k(1-b)}{2}} |S_{2k}|^{-\frac{(n-k)(1-b)}{2}} \end{aligned}$$

and $m_0(b)$ is the same as previously

so that

$$B_{k0}^F = \frac{m_k(b)}{m_0(b)} = \frac{\Gamma_p\left(\frac{k-1}{2}\right)\Gamma_p\left(\frac{n-k-1}{2}\right)\Gamma_p\left(\frac{nb-1}{2}\right)}{\Gamma_p\left(\frac{bk-1}{2}\right)\Gamma_p\left(\frac{b(n-k)-1}{2}\right)\Gamma_p\left(\frac{n-1}{2}\right)} \cdot \left[\frac{|S_{1k}|^{k_1}|S_{2k}|^{k_2}}{|S_n|^n}\right]^{-\frac{(1-b)}{2}}. \tag{2.5.10}$$

Notice from (2.5.10) that bk must be greater than p and that $b(n-k)$ must be greater than p so that we must have $b > \frac{p}{k}$ and $b > \frac{p}{n-k}$ respectively. The minimal sample is $2p+2$, but $b = \frac{2p+2}{n}$ does not always satisfy the above restrictions. If we take $b = \frac{2p+2}{n}$ and we want to consider all the possible values of k where $p+1 \leq k \leq n-p-1$, it follows that n must be greater or equal to 8 and $p < \frac{n}{4} - 1 + \sqrt{\frac{n^2-8n}{16}}$. For example, if $n = 40$, p must be smaller or equal to 17 for the fractional Bayes factor to be determined for any $p+1 \leq k \leq n-p-1$. If the above condition on p is not satisfied, we can either increase the fraction b or reduce the number of possible change-point positions being considered.

2.5.3 A CHANGE IN THE VARIANCE

The regular Bayes factor for the model as in (2.2.52) follows from paragraph 2.2.3 so that

$$B_{k0} = \frac{\prod_{j=1}^{p}\Gamma\left(\frac{n-p-1-j}{2}\right)\gamma^{\frac{p(n-k-p-1)}{2}}[k+(n-k)\gamma]^{-\frac{p}{2}}|T_\gamma|^{-\left(\frac{n-p-2}{2}\right)}}{2^{\frac{p(p+2)}{2}}\prod_{j=1}^{p}\Gamma\left(\frac{n+1-j}{2}\right)(n+\delta)^{-\frac{p}{2}}\delta^{\frac{p-1}{2}}\Gamma\left(\frac{n-p}{2}\right)\left[\Gamma\left(\frac{n}{2}\right)\right]^{-1}\pi^{\frac{p}{2}}[1-(n+\delta)^{-1}\delta]^{-\frac{1}{2}}|T_{0\delta}|^{-\left(\frac{n-p}{2}\right)}} \tag{2.5.11}$$

where $T_{0\delta}$ and T_γ follow from (2.2.22) and (2.2.55) respectively.

To calculate the partial Bayes factors, the whole sample Bayes factor, with the improper priors $\Pi(\mu) \propto 1$, $\Pi\left(\Sigma_i^{-1}\right) \propto \left|\Sigma_i^{-1}\right|^{-\frac{p+1}{2}}$ $(i = 1,2)$ and $\Pi(\gamma) \propto \frac{1}{\gamma}$ is

$$B_{k0}^N = \frac{\int 2^{\frac{p(p+2)}{2}}\prod_{j=1}^{p}\left(\frac{n-p-j}{2}\right)[k+(n-k)\gamma]^{-\frac{p}{2}}\gamma^{\frac{p(n-k-p-1)}{2}-1}|T_\gamma|^{-\left(\frac{n-p-2}{2}\right)}d\gamma}{\prod_{j=1}^{p}\Gamma\left(\frac{n-1-j}{2}\right)n^{-\frac{p}{2}}|S_n|^{-\left(\frac{n-2}{2}\right)}}. \tag{2.5.12}$$

For this case the minimal sample is exactly the same as in the previous model (2.2.24), but with

$$B_{0k}^N(\boldsymbol{x}(\ell)) = \frac{\prod_{j=1}^{p}\left(\frac{2p+1-j}{2}\right)|S_{m_\ell}|^{-p}}{\int 2^{\frac{p(2-p)}{2}}\prod_{j=1}^{p}\left(\frac{p+2-j}{2}\right)(1+\gamma)^{-\frac{p}{2}}\gamma^{-1}|T_\gamma(\ell)|^{-\left(\frac{p+1}{2}\right)}d\gamma} \tag{2.5.13}$$

where

$$T_\gamma(\ell) = \sum_j \boldsymbol{x}_j\boldsymbol{x}_j' + \gamma\sum_r \boldsymbol{x}_r\boldsymbol{x}_r' - \hat{\boldsymbol{\mu}}_{\gamma\ell}(p+1)(1+\gamma)\hat{\boldsymbol{\mu}}_{\gamma\ell}'$$

and where

$$\hat{\boldsymbol{\mu}}_{\gamma\ell} = (1+\gamma)^{-1}\left(\overline{\boldsymbol{x}}_{m_k}^1 + \gamma\overline{\boldsymbol{x}}_{m_k}^2\right)$$

and where $\sum_j$ is the summation over the subset $\boldsymbol{x}_{m_k}^1(\ell)$ and $\sum_r$ is the summation over the subset $\boldsymbol{x}_{m_k}^2(\ell)$ with $\overline{\boldsymbol{x}}_{m_k}^1, \overline{\boldsymbol{x}}_{m_k}^2$ and S_{m_ℓ} the same as before.

The Arithmetic Intrinsic Bayes factor for this model would then be the same as in equation (2.5.4), but where B_{k0}^N is as in (2.5.12) and $B_{0k}^N(\boldsymbol{x}(\ell))$ is as in (2.5.13).

To get the Fractional Bayes factor for this model it follows, after integrating γ first, that

$$m_k(b) = b^{\frac{bnp}{2}}\pi^{-\frac{pn(1-b)}{2}}.$$

$$\frac{\Gamma_p\left(\frac{k_1}{2}\right)\Gamma_p\left(\frac{k_2}{2}\right)\int_{-\infty}^{\infty}\left[S_{1k}^2 + k_1(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{1k})'(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{1k})\right]^{-\frac{k_1}{2}}\left[S_{2k}^2 + k_2(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{2k})'(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{2k})\right]^{-\frac{k_2}{2}}d\boldsymbol{\mu}}{\Gamma_p\left(\frac{bk_1}{2}\right)\Gamma_p\left(\frac{bk_2}{2}\right)\int_{-\infty}^{\infty}\left[S_{1k}^2 + k_1(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{1k})'(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{1k})\right]^{-\frac{bk_1}{2}}\left[S_{2k}^2 + k_2(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{2k})'(\boldsymbol{\mu}-\overline{\boldsymbol{x}}_{2k})\right]^{-\frac{bk_2}{2}}d\boldsymbol{\mu}} \tag{2.5.14}$$

and $m_0(b)$ is the same as previously, so that

$$B_{k0}^F = \frac{m_k(b)}{m_0(b)}.$$

follows directly.

The minimal sample for this case is also $2p+2$ and b has the same conditions as for a change in the mean and variance in section 2.5.2.

The posterior probabilities follow from equation (2.2.3). If we let $P(M_0) = \frac{1}{2}$ and $P(M_k) = \frac{1}{2m}$, $k = 1, \dots, m$ be the prior probabilities for no change-point and a change-point at k respectively, with m possible positions for the change-point, then the posterior probabilities are given by

$$P(M_0|\boldsymbol{x}) = \left[1 + \frac{1}{m}\sum_{j=1}^{m} B_{0j}^{-1}\right]^{-1} \text{ and } P[M_k|\boldsymbol{x}] = B_{0k}^{-1}\left[m + \sum_{j=1}^{m} B_{0j}^{-1}\right]^{-1}. \tag{2.5.15}$$

2.5.4 MULTIPLE CHANGE-POINTS

In the case of a maximum of R possible change-points, let $\Pi(\boldsymbol{k}|r) = \left[\binom{h}{r}\right]^{-1}$ and $\Pi(r) = \frac{1}{R+1}$, $r = 0, \dots, R$, where h is the number of possible change-points for the particular model. Then

$$\Pi(r=0|y) = \left[1 + \sum_{r=1}^{R}\left[\binom{h}{r}\right]^{-1} B_{\boldsymbol{k}0}^{r}\right]^{-1} \tag{2.5.16}$$

where $B_{\boldsymbol{k}0}^{r}$ is the Bayes factor in favour of the partition $\boldsymbol{k}$ out of r change-points when compared to no change-point, and also

$$\Pi(\boldsymbol{k}, r|y) = \left[hB_{0k}^{1} + \sum_{j=2}^{n-1} B_{j\boldsymbol{k}}^{1,r} + \frac{\Pi(k|r=2)}{\Pi(k|r=1)}\sum_{\boldsymbol{j}} B_{\boldsymbol{j}\boldsymbol{k}}^{2,r} + \cdots\right]^{-1} \tag{2.5.17}$$

where in general $B_{\boldsymbol{j}\boldsymbol{k}}^{s,r}$ is the Bayes factor for partition $\boldsymbol{j}$ out of s change-points against partition $\boldsymbol{k}$ out of r change-points.

For example if $R = 2$, we get that

$$\Pi(k, r=1|y) = \left[hB_{0k}^1 + \sum_{j=2}^{n-1} B_{jk}^1 + \frac{2}{h-1}\sum_{\boldsymbol{j}} B_{\boldsymbol{j}k}^{2,1}\right]^{-1} \tag{2.5.18}$$

and

$$\Pi(\boldsymbol{k}, r=2|y) = \left[\frac{1}{2}h(h-1)B_{0\boldsymbol{k}}^2 + \frac{h-1}{2}\sum_{j=2}^{n-1} B_{j\boldsymbol{k}}^{1,2} + \sum_{\boldsymbol{j}} B_{\boldsymbol{j}\boldsymbol{k}}^2\right]^{-1} \tag{2.5.19}$$

where the following relationships hold for the Bayes factors:

$$B_{jk}^1 = \frac{B_{j0}^1}{B_{k0}^1}, \quad B_{\boldsymbol{j}k}^{2,1} = \frac{B_{\boldsymbol{j}0}^2}{B_{k0}^1}, \quad B_{j\boldsymbol{k}}^{1,2} = \frac{B_{j0}^1}{B_{\boldsymbol{k}0}^2}, \quad B_{\boldsymbol{j}\boldsymbol{k}}^2 = \frac{B_{\boldsymbol{j}0}^2}{B_{\boldsymbol{k}0}^2}. \tag{2.5.20}$$

So the posterior probabilities can be written as

$$\begin{aligned} \Pi(k, r=1|y) &= \left[hB_{0k}^1 + \frac{1}{B_{k0}^1}\sum_{j=2}^{n-1} B_{j0}^1 + \frac{2}{h-1}\frac{1}{B_{k0}^1}\sum_{\boldsymbol{j}} B_{\boldsymbol{j}0}^2\right]^{-1} \\ &= B_{k0}^1\left[h + \Sigma B_{j0}^1 + \frac{2}{h-1}\Sigma B_{\boldsymbol{j}0}^2\right]^{-1} \end{aligned} \tag{2.5.21}$$

and

$$\Pi(k, r=2|y) = B_{\boldsymbol{k}0}^2\left[\tfrac{h(h-1)}{2} + \tfrac{h-1}{2}\sum_{j=2}^{n-1} B_{j0}^1 + \sum_{\boldsymbol{j}} B_{j0}^2\right]^{-1}. \tag{2.5.22}$$

If $R = 2$ and we consider only a subset of say L of the $\frac{h(h-1)}{2}$ combinations of two change-points, the prior probabilities become

$$\Pi(k = n) = \frac{h(h-1)}{2[h(h-1)+L]},$$

$$\Pi(k|r=1) = \frac{h-1}{2[h(h-1)+L]}, \quad k = 2, \ldots, n-1$$

and

$$\Pi(\boldsymbol{k}|r=2) = \frac{1}{h(h-1)+L}, \quad \boldsymbol{k} \in \boldsymbol{k}(L) \tag{2.5.23}$$

and the formulae for the posterior probabilities stay the same as above.

For R change-points with $b = \frac{p+R+1}{n}$, the fractional BF follows as a generalization of (2.5.6) as

$$B^r_{0\boldsymbol{k}} = \frac{\Gamma_p\left(\frac{n-1}{2}\right)\Gamma_p\left(\frac{p+R-r}{2}\right)}{\Gamma_p\left(\frac{p+R}{2}\right)\Gamma_p\left(\frac{n-r-1}{2}\right)}\left(\frac{|S|}{\left|\sum_{i=1}^{r+1} S_i\right|}\right)^{-\frac{n-p-R-1}{2}} \tag{2.5.24}$$

for a change in the mean.

2.6 MULTIPLE CHANGE-POINTS (CP'S)

Apart from section 2.5.4, the previous section results are derived for models with at most one change-point. This theory can regularly be extended to any number of change-points. In this section we give the results when the actual number of change-points, $r \geq 0$, is unknown. As the number of change-points increases, there is an increase of complexity with the computational aspects. Different approaches have been proposed.

Non-Bayesian approach by Lombard (1987) uses rank tests to test for multiple change-points. Hartigan (1990) introduced product partition models, which assume that observations in different components of a random partition of the data are independent. If the probability distribution of random partitions is in a certain product form prior to making the observations, it is also in product form given the observations. The product model thus provides a convenient machinery for allowing the data to weight the partitions likely to hold. Inference about particular future observations may then be made by first conditioning on the partition and then averaging over all partitions. Barry and Hartigan (1992) applied these models with special computational simplicity to change-point problems, where the partitions divide the sequence of observations into components within which different regimes hold. They show, with appropriate selection of prior product models, that the observations can eventually determine approximately the true partition.

Barry and Hartigan (1993) model the process in which a sequence of observations undergoes sudden changes (multiple changes) at unknown times, by supposing that there is an underlying sequence of parameters partitioned into contiguous blocks of equal parameter values. The beginning of each block is said to be a change-point. Observations are then assumed to be independent in different blocks given the sequence of parameters.

Gupta and Chen (1996) applied the Schwarz information criterion together with the binary segmentation procedure to detect change-points in a set of geological data and the changes in the frequencies of pronouns in the plays of Shakespeare. The emphasis here was on the exploratory data analysis rather than the theoretical statistical investigation.

Carlin, Gelfand and Smith (1992) presented a general approach to hierarchical Bayes change-point models. In particular, desired marginal posterior densities are obtained utilizing the Gibbs sampler, an iterative Monte Carlo method, avoiding sophisticated analytic and numerical high dimensional integration procedures.

Markov chain Monte Carlo (MCMC) methods for Bayesian computation have however been restricted to problems where the joint distribution of all variables has a density with respect to some fixed standard underlying measure. They have therefore not been available for application to Bayesian model determination where the dimensionality of the parameter vector is typically not fixed. Green (1995) proposes a new framework for the construction of reversible Markov chain samplers that jump between parameter subspaces of differing dimensionality, which is flexible and entirely constructive. He illustrated the methodology with applications to multiple change-point analysis in one and two dimensions.

The use of MCMC simulation techniques has made feasible the routine Bayesian analysis of many complex high-dimensional problems. However, one area which has received relatively little attention is that of comparing models of possibly different dimensions, where the essential difficulty is that of computing the high-dimensional integrals needed for calculating the normalization constants for the posterior distribution under each model specification

(Raftery (1995)).

Phillips and Smith (1996) show how methodology developed recently by Grenander and Miller (1991, 1994) can be exploited to enable routine simulation-based analysis for this type of problem. Model uncertainty is accounted for by introducing a joint prior probability distribution over both the set of possible models and the parameters of those models. Inference can then be made by simulating realizations from the resulting posterior distribution using an iterative jump-diffusion sampling algorithm. The essential features of this simulation approach are that discrete transitions, or jumps, can be made between models of different dimensionality, and within a model of fixed dimensionality the conditional posterior appropriate for that model is sampled. Model comparison or choice can then be based on the simulated approximation to the marginal posterior distribution over the set of models.

Lee (1998) considered the problem of estimating the number of change-points in a sequence of independent random variables in a Bayesian framework. They found that, under mild assumptions and with respect to a suitable prior distribution, the posterior mode of the number of change-points converges to the true number of change-points in the frequentist sense. Furthermore, the posterior mode of the locations of the change-points is shown to be within $O_p(\log n)$ of the true locations of the change-points (n is the sample size).

2.6.1 A CHANGE IN THE MEAN

Suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independent normal random vectors such that

$$\boldsymbol{X}_i \sim \begin{cases} iN(\boldsymbol{\mu}_1, \Sigma); & i = 1, \ldots, k_1 \\ iN(\boldsymbol{\mu}_2, \Sigma); & i = k_1+1, \ldots, k_2 \\ \quad\vdots & \\ iN(\boldsymbol{\mu}_{r+1}, \Sigma); & i = k_r+1, \ldots, n \end{cases} \tag{2.6.1}$$

where

$$n \geq r+p+1, \quad \boldsymbol{X}_i \in I\!R^p \quad \text{and} \quad 1 \leq k_1 < k_2 < \cdots < k_r \leq n-1.$$

The mean vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{r+1}$ are unknown, with $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \neq \cdots \neq \boldsymbol{\mu}_{r+1}$. The unknown covariance matrix Σ remains unchanged through the r structural changes at unknown k's, where r is fixed and known for the moment. Let the marginal prior densities of $\boldsymbol{k} = [k_1, \ldots, k_r]$ be uniform, i.e. $\Pi(\boldsymbol{k}|r) = \left({}^{n-1}C_r\right)^{-1}$, (where n is large enough to allow all permutations, i.e. $n \geq rp + p + 1$) and the marginal prior density of Σ^{-1} be $W(v, \Gamma)$ and $\boldsymbol{\mu}_1 \cdot \ldots, \boldsymbol{\mu}_{r+1} \sim iN(\boldsymbol{\theta}, \Phi)$.

The joint distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ conditional on $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{r+1}$ and the changes having taken place at $\boldsymbol{k}$ is given by

$$f(X|\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{r+1}, \boldsymbol{k}, r, \Sigma) = \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{j=1}^{r+1}\sum_{i=k_{j-1}+1}^{k_j} (\boldsymbol{x}_i - \boldsymbol{\mu}_j)'\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)},$$

with

$$k_0 = 0,$$

$$k_{r+1} = n$$

and

$$n_j = k_j - k_{j-1}, \qquad j = 1, \ldots, r+1.$$

Notice that the partitioning of $\boldsymbol{k}$ must be so that $n_j \geq p + 1$ for at least one j.

Furthermore

$$\Pi(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{r+1}) = \left(\frac{1}{2\pi}\right)^{\frac{p(r+1)}{2}} |\Phi|^{-\left(\frac{r+1}{2}\right)} e^{-\frac{1}{2}\sum_{j=1}^{r+1}(\boldsymbol{\mu}_j - \boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\mu}_j - \boldsymbol{\theta})}$$

so that

$$f(X|\boldsymbol{k}, \boldsymbol{\theta}, \Phi, r, \Gamma, v) = \int \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} |\Sigma^{-1}|^{\left(\frac{n+v-p-1}{2}\right)} |\Phi|^{-\left(\frac{r+1}{2}\right)} \prod_{j=1}^{r+1} \left|n_j\Sigma^{-1} + \Phi^{-1}\right|^{-\frac{1}{2}}$$

$$e^{-\frac{1}{2}\left[(r+1)\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta} + \sum_{j=1}^{r+1}\sum_{i=k_{j-1}+1}^{k_j} \boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i + \sum_{j=1}^{r+1} \hat{\boldsymbol{\mu}}_{j\Sigma}' \left(n_j\Sigma^{-1} + \Phi^{-1}\right) \hat{\boldsymbol{\mu}}_{j\Sigma}\right]}$$

$$e^{-\frac{1}{2}tr\Gamma^{-1}\Sigma^{-1}} d\Sigma^{-1}$$

where for $j = 1, \ldots, r+1$,

$$\hat{\boldsymbol{\mu}}_{j\Sigma} = \left(n_j\Sigma^{-1} + \Phi^{-1}\right)^{-1}\left(n_j\Sigma^{-1}\overline{\boldsymbol{x}}_j + \Phi^{-1}\boldsymbol{\theta}\right)$$

and

$$\overline{\boldsymbol{x}}_j = \frac{1}{n_j}\sum_{i=k_{j-1}+1}^{k_j}\boldsymbol{x}_i.$$

As in paragraph 2.2.1, let $\Phi^{-1} = \delta\Sigma^{-1}$ so that

$$f(X|\boldsymbol{k},\boldsymbol{\theta},\delta,r,\Gamma,v) = c_{1p}\left(\frac{1}{2\pi}\right)^{\frac{np}{2}}\delta^{\frac{p(r+1)}{2}}\prod_{j=1}^{r+1}(n_j+\delta)^{-\frac{p}{2}}\left|\Gamma^{-1} + (r+1)\delta\boldsymbol{\theta}\boldsymbol{\theta}'\right.$$

$$\left. + \sum_{j=1}^{r+1}\sum_{i=k_{j-1}+1}^{k_j}\boldsymbol{x}_i\boldsymbol{x}_i' - \sum_{j=1}^{r+1}(n_j+\delta)\hat{\boldsymbol{\mu}}_{j\delta}\hat{\boldsymbol{\mu}}_{j\delta}'\right|^{-\left(\frac{n+v}{2}\right)} \tag{2.6.2}$$

where $\hat{\boldsymbol{\mu}}_{j\delta} = (n_j+\delta)^{-1}(n_j\overline{\boldsymbol{x}}_j + \delta\boldsymbol{\theta})$, for $j = 1, \ldots, r+1$

and c_{1p} is as in (2.2.14).

The marginal mass function of $\boldsymbol{k}$ is then

$$\Pi(\boldsymbol{k}|X,\boldsymbol{\theta},\delta,r,\Gamma,v) = \frac{f(X|\boldsymbol{k},\boldsymbol{\theta},\delta,r,\Gamma,v)\Pi(\boldsymbol{k}|r)}{\sum\limits_{\boldsymbol{k}} f(X|\boldsymbol{k},\boldsymbol{\theta},\delta,r,\Gamma,v)\Pi(\boldsymbol{k}|r)} \tag{2.6.3}$$

where $\sum\limits_{\boldsymbol{k}}$ is the summation over all combinations of $\boldsymbol{k}$.

The probability that the j-th change-point out of r lies at observation i follows from equation (2.6.3) as

$$P[k_j = i|r,X,\boldsymbol{\theta},\delta,\Gamma,v] = \sum_{k_j=i}\Pi(\boldsymbol{k}|X,\boldsymbol{\theta},\delta,r,\Gamma,v), \qquad j = 1,\ldots,r,$$

where $\sum_{k_j=i}$ defines all those partitions of $\boldsymbol{k}$ for which the j-th change-point lies at i.

While the number of change-points is fixed, vague improper priors can be used. By letting $\delta \to 0$, $\Gamma^{-1} \to 0$ and $v \to 0$ in equation (2.6.3), the posterior of k simplifies to

$$\Pi(\boldsymbol{k}|X, r) = \frac{\prod_{j=1}^{r+1} (n_j)^{-\frac{p}{2}} \left[\sum_{j=1}^{r+1} trS_j\right]^{-\frac{n}{2}}}{\sum_{\boldsymbol{k}} \prod_{j=1}^{r+1} (n_j)^{-\frac{p}{2}} \left[\sum_{j=1}^{r+1} trS_j\right]^{-\frac{n}{2}}} \tag{2.6.4}$$

where

$$S_j = \sum_{i=k_{j-1}+1}^{k_j} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_j)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_j)'.$$

Furthermore

$$\boldsymbol{\mu}_j|\boldsymbol{k}, X \sim t_p(n + 1 - p, \overline{x}_j, T_j) \tag{2.6.5}$$

where

$$T_j = n_j(n + 1 - p)S_j^{-1}.$$

A generalization of (2.2.15), when using a hierarchical model with vague second stage priors $\Pi(\boldsymbol{\theta}) \propto 1$ and $\Pi(\delta) = \frac{1}{\delta}$, follows from (2.6.2) after integrating out $\boldsymbol{\theta}$, as

$$\begin{aligned} f(X|\boldsymbol{k}, r, \Gamma, v) \;\; = \;\; & \int \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} \delta^{\frac{p(r+1)}{2} - \frac{3}{2}} \prod_{j=1}^{r+1} (n_j + \delta)^{-\frac{p}{2}} \Pi^{\frac{p}{2}} \Gamma\left(\frac{v + n - p}{2}\right) \left[\Gamma\left(\frac{v + n}{2}\right)\right]^{-1} \\ & \left[r + 1 - \sum_{j=1}^{r+1} (n_j + \delta)^{-1}\delta\right]^{-\frac{1}{2}} |T_{\delta\Gamma}|^{-\left(\frac{v+n-p}{2}\right)} \, d\delta \end{aligned}$$

where

$$T_{\delta\Gamma} = \Gamma^{-1} + \sum_{j=1}^{r+1} \sum_{i=k_{j-1}+1}^{k_j} \boldsymbol{x}_i\boldsymbol{x}_i' - \sum_{j=1}^{r+1} (n_j + \delta)^{-1} n_j^2 \overline{x}_j \overline{x}_j' - \hat{\boldsymbol{\theta}}_\delta[(r + 1)\delta - (n_j + \delta)^{-1}\delta^2]\hat{\boldsymbol{\theta}}_\delta'$$

(2.6.6)

and

$$\hat{\boldsymbol{\theta}}_\delta = \left[r + 1 - \sum_{j=1}^{r+1}(n_j+\delta)^{-1}\delta\right]^{-1}\left[\sum_{j=1}^{r+1} n_j(n_j+\delta)^{-1}\overline{x}_j\right].$$

This can be integrated numerically. Note that we can't let $\delta \to 0$, while we can still let $\Gamma^{-1} \to 0$ and $v \to 0$ in the above equation, so that

$$\Pi(\boldsymbol{k}|X,\delta,r) \propto \delta^{\frac{p(r+1)-1}{2}} \prod_{j=1}^{r+1}(n_j+\delta)^{-\frac{p}{2}} \left[r+1-\sum_{j=1}^{r+1}(n_j+\delta)^{-1}\delta\right]^{-\frac{1}{2}} |T_\delta|^{-\frac{n-p}{2}} \tag{2.6.7}$$

where T_δ is given in (2.6.6) with Γ^{-1} omitted.

If we assume that the number of change-points (r) is fixed, Gibbs sampling can be used instead. The full conditional distributions are as follows:

$$\boldsymbol{\mu}_j|X,\boldsymbol{k},\boldsymbol{\theta},\delta,\Sigma \sim N\left(\frac{n_j\overline{x}_j+\delta\boldsymbol{\theta}}{n_j+\delta}, \frac{1}{n_j+\delta}\Sigma\right), \qquad j = 1,\ldots,r+1,$$

$$\delta|X,\boldsymbol{k},\boldsymbol{\theta},\boldsymbol{\mu}_j,\Sigma \sim \text{Gamma}\left(p, \frac{1}{2}tr\Sigma^{-1}\sum_{j=1}^{r+1}(\boldsymbol{\mu}_j-\boldsymbol{\theta})(\boldsymbol{\mu}_j-\boldsymbol{\theta})'\right),$$

$$\Pi(\boldsymbol{k}|X,\boldsymbol{\mu}_j,\delta,\Sigma,\boldsymbol{\theta}) = \frac{e^{-\frac{1}{2}\sum_{j=1}^{r+1}\sum_{i=k_{j-1}+1}^{k_j}(\boldsymbol{x}_i-\overline{x}_j)'\Sigma^{-1}(\boldsymbol{x}_i-\overline{x}_j)}}{\sum_{k} e^{-\frac{1}{2}\sum_{j=1}^{r+1}\sum_{i=k_{j-1}+1}^{k_j}(\boldsymbol{x}_i-\overline{x}_j)'\Sigma^{-1}(\boldsymbol{x}_i-\overline{x}_j)}},$$

$$\boldsymbol{\theta}|X,\boldsymbol{\mu}_j,\delta,k,\Sigma \sim N\left(\frac{\sum_{j=1}^{r+1}\boldsymbol{\mu}_j}{r+1}, \frac{1}{r+1}\delta^{-1}\Sigma\right)$$

and, after letting $v \to 0$ and $\Gamma^{-1} \to 0$, it follows that

$$\Sigma^{-1}|X,\boldsymbol{\mu}_j,\boldsymbol{\theta},\delta,k \sim W\left(n+r+1,\left[\delta\sum_{j=1}^{r+1}(\boldsymbol{\mu}_j-\boldsymbol{\theta})(\boldsymbol{\mu}_j-\boldsymbol{\theta})'\right.\right.$$

$$\left.\left.+\sum_{j=1}^{r+1}\sum_{i=k_{j-1}+1}^{k_j}(\boldsymbol{x}_i-\boldsymbol{\mu}_j)(\boldsymbol{x}_i-\boldsymbol{\mu}_j)'\right]^{-1}\right). \qquad (2.6.8)$$

Up to this stage r was assumed known, which is not always the case in practice. So let $R = \max\{r\}$ where R is fixed.

Let the marginal prior density of r be uniform, i.e.

$$\Pi(r) = \frac{1}{R+1} \text{ and } \Pi(\boldsymbol{k}^{(r)}/r) = ({}^{n-1}C_r)^{-1} \text{ where } \boldsymbol{k}^{(r)} = [k_1,\ldots,k_r].$$

Considering the same model as in (2.6.1), with $n \geq Rp+p+1$.

Because of dimensions differing between models, we can't let $\delta \to 0$. If we let $\Gamma^{-1} \to 0$ and $v \to 0$, it follows from equation (2.6.7) that the marginal mass function of $\boldsymbol{k}$ in (2.6.3) will now be

$$\Pi(\boldsymbol{k}^{(r)}|X,\delta) = \frac{f(X|\boldsymbol{k},\delta,r)\Pi(\boldsymbol{k}^{(r)}|r)}{\sum_{r=0}^{R}\sum_{\boldsymbol{k}^{(r)}} f(X|\boldsymbol{k},\delta,r)\Pi(\boldsymbol{k}^{(r)}|r)}, \qquad r = 0,1,\ldots,R \qquad (2.6.9)$$

where $\sum_{\boldsymbol{k}^{(r)}}$ is the summation over all partitions of $\boldsymbol{k}^{(r)}$.

The probability for exactly i change-points, where $i = 0,\ldots,R$, is then given by

$$P[r=i|X,\delta] = \sum_{\boldsymbol{k}^{(i)}} \Pi(\boldsymbol{k}^{(r)}|X,\delta). \qquad (2.6.10)$$

If δ must be eliminated, we must integrate over equation (2.6.7).

Notice that we can't apply standard Gibbs, as the number of change-points is unknown and the dimensions will differ. However, we can use the MCMC of Green (1995), a reversible-jump Metropolis-Hastings algorithm, or the MCMC of Phillips and Smith (1996), an iterative jump-diffusion sampling algorithm.

2.6.2 A CHANGE IN THE MEAN AND VARIANCE

Suppose that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are again independent normal random vectors but such that

$$\boldsymbol{X}_i \sim \begin{cases} iN(\boldsymbol{\mu}_1, \Sigma_1); & i = 1, \ldots, k_1 \\ iN(\boldsymbol{\mu}_2, \Sigma_2); & i = k_1 + 1, \ldots, k_2 \\ \vdots & \\ iN(\boldsymbol{\mu}_{r+1}, \Sigma_{r+1}); & i = k_r + 1, \ldots, n-1 \end{cases}$$

where $n \geq (r+1)(p+1), \quad \mathbf{X}_i \in I\!R^p$ and $n_j \geq p+1$ for all j.

The mean vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{r+1}$ with $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \neq \cdots \neq \boldsymbol{\mu}_{r+1}$ and the covariance matrices $\Sigma_1 \neq \Sigma_2 \neq \cdots \neq \Sigma_{r+1}$ are unknown. The marginal prior density of $\boldsymbol{k}$ is uniform for fixed r, while the marginal prior densities are such that for $j = 1, \ldots, r+1$

$$\boldsymbol{\mu}_j \sim iN(\boldsymbol{\theta}_j, \Phi_j) \text{ and } \Sigma_j^{-1} \sim W(v, \Gamma).$$

Then

$$f(X|\boldsymbol{k}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{r+1}, \Phi_1, \ldots, \Phi_{r+1})$$

$$= \int \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} \prod_{j=1}^{r+1} |\sum_j|^{-\frac{(n_j+v_j-p-1)}{2}} |n_j\Sigma_j^{-1} + \Phi_j^{-1}|^{-\left(\frac{r+1}{2}\right)} |\Phi_j|^{-\left(\frac{r+1}{2}\right)} e^{-\frac{1}{2}(r+1)\boldsymbol{\theta}_j'\Phi_j^{-1}\boldsymbol{\theta}_j}$$

$$\prod_{j=1}^{r+1} e^{-\frac{1}{2}\left[\sum_{i=k_{j-1}+1}^{k_j} \boldsymbol{x}_i'\Sigma_j^{-1}\boldsymbol{x}_i + \Gamma^{-1}\Sigma_j^{-1} - \hat{\boldsymbol{\mu}}_{j\Sigma_j\theta_j}' \left(n_j\Sigma_j^{-1} + \Phi_j^{-1}\right) \hat{\boldsymbol{\mu}}_{j\Sigma_j\theta_j}\right]} d\Sigma_j^{-1}$$

where

$$\hat{\boldsymbol{\mu}}_{j\Sigma_j\theta_j} = \left(n_j\Sigma_j^{-1} + \Phi_j^{-1}\right)^{-1} \left(n_j\Sigma_j^{-1}\overline{\boldsymbol{x}}_j + \Phi_j^{-1}\boldsymbol{\theta}_j\right).$$

Assume that $\Phi_j^{-1} = \delta_j \sum_j^{-1}$ for $j = 1, \ldots, r+1$ and where $\delta_j > 0$. It then follows that

$$f(X|\boldsymbol{k},\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_{r+1},\delta_1,\ldots,\delta_{r+1}) = \left(\frac{1}{2\pi}\right)^{\frac{np}{2}} c_{9p}\prod_{j=1}^{r+1}\left\{\delta_j^{\frac{p}{2}}(n_j+\delta_j)^{-\frac{p}{2}}\left|\delta_j\boldsymbol{\theta}_j\boldsymbol{\theta}_j'+\right.\right.$$

$$\left.\left.\sum_{i=k_{j-1}+1}^{k_j}\boldsymbol{x}_i\boldsymbol{x}_i'+\Gamma^{-1}-\hat{\boldsymbol{\mu}}_{j\Sigma_j\delta_j}(n_j+\delta_j)\hat{\boldsymbol{\mu}}_{j\Sigma_j\delta_j}'\right|^{-\frac{(v_j+n_j)}{2}}\right\} \tag{2.6.11}$$

where

$$\hat{\boldsymbol{\mu}}_{j\Sigma_j\delta_j} = (n_j+\delta_j)^{-1}(n_j\bar{\boldsymbol{x}}_j+\delta_j\boldsymbol{\theta}_j)$$

and

$$c_{9p} = \prod_{j=1}^{r+1}\left[2^{\frac{p(v_j+n_j)}{2}}\pi^{\frac{p(p-1)}{4}}\prod_{i=1}^{p}\left(\frac{v_j+n_j+1-i}{2}\right)\right].$$

The mass function of k is then proportional to (2.6.11).

Under the assumption of having a fixed number of change-points, vague improper priors can be used. By letting $\delta_j \to 0$, $\Gamma^{-1} \to 0$ and $v_j \to 0$ in equation (2.6.11), the posterior of $\boldsymbol{k}$ simplifies to

$$\Pi(\boldsymbol{k}|X) \propto c_{10p}\prod_{j=1}^{r+1}\left\{n_j^{-\frac{p}{2}} tr S_j^{-\frac{n_j}{2}}\right\}$$

where

$$c_{10p} = \prod_{j=1}^{r+1}\Gamma_p\left(\frac{n_j}{2}\right).$$

Furthermore

$\Sigma_j^{-1}|\boldsymbol{k},X \sim$ Wishart $\left(n_j, S_j^{-1}\right)$ and $\boldsymbol{\mu}_j|\boldsymbol{k},X \sim t_p\left(n_j+1-p,\bar{\boldsymbol{x}}_j,n_j(n_j+1-p)S_j^{-1}\right)$.

With the hierarchical model $\Pi(\boldsymbol{\theta}_j) \propto 1$ and $\Pi(\delta_j) \propto \dfrac{1}{\delta_j}$ for $j = 1,\ldots,r+1$ it follows, after integrating out $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_{r+1}$, that

$$f(X|\boldsymbol{k},\delta_1,\ldots,\delta_{r+1}) = \pi^p\left(\frac{1}{2\pi}\right)^{\frac{np}{2}} c_{11p}\prod_{j=1}^{r+1}\left\{\delta_j^{\frac{p-1}{2}}(n_j+\delta_j)^{-\frac{p}{2}}\Gamma\left(\frac{v_j+n_j-p}{2}\right)\left[\Gamma\left(\frac{v_j+n_j}{2}\right)\right]^{-1}\right.$$

$$\left.\left[1-(n_j+\delta_j)^{-1}\delta_j\right]^{-\frac{1}{2}}|\Gamma^{-1}+S_j|^{-\left(\frac{v_j+n_j-p}{2}\right)}\right\} \tag{2.6.12}$$

where

$$c_{11p} = 2^{\frac{p\left(\sum_{j=1}^{r+1} v_j+n\right)}{2}}\prod_{j=1}^{r+1}\Gamma_p\left(\frac{v_j+n_j}{2}\right).$$

By letting $\Gamma^{-1}\to 0$ and $v_j \to 0$ $(j=1,\ldots,r+1)$, it follows that

$$\Pi(\boldsymbol{k}|X,\delta_j) \propto c_{10p}\prod_{j=1}^{r+1}\delta_j^{\frac{p-1}{2}}k_j^{-\frac{1}{2}}(n_j+\delta_j)^{-\left(\frac{p-1}{2}\right)}\Gamma\left(\frac{n_j-p}{2}\right)\left[\Gamma\left(\frac{n_j}{2}\right)\right]^{-1}n_j^{-\frac{1}{2}}|S_j|^{-\left(\frac{n_j-p}{2}\right)}.$$

Note that when r is not fixed, the rest follows as in equation (2.6.9). Also notice that the full conditional distributions, in order to apply Gibbs sampling, are similar to those given in paragraph 2.2.2.

Extension of the case of a change in the variance alone to several change-points is also possible (see paragraph 2.3.3), but because of the multiple numerical integration we will omit the details here.

2.7 THE UNIVARIATE CASE

The results for the univariate case can directly be derived from the previous sections. The reason why it's treated in a separate section is to look more critically at certain model assumptions in a simpler setting. We will also examine the special cases of autocorrelation and model comparisons.

The general area of non-sequential statistical inference about change-points has been explored mainly for location-change in a univariate distribution. From a Bayesian approach, Broemeling (1974), Smith (1975) and Booth and Smith (1982) and from a non-Bayesian approach Chernoff and Zacks (1964), Gardner (1969), Hinkley (1970), Sen and Srivastava (1975), Bhattacharya and Brockwell (1976) and Bhattacharya (1987) have studied the problem of the change in mean of a univariate normal distribution. Non-parametric methods for location change in a univariate distribution have been proposed by Battacharya and Johnson (1968), Sen and Srivastava (1975) and Darkhovskv (1976). Hinkley, *et al.* (1980) and Zacks (1983) discussed these and other related contributions.

Smith (1975) and Bhattacharya (1987) also considered a change in the mean and variance of a univariate normal distribution.

2.7.1 A CHANGE IN THE MEAN

In the univariate case, when $p = 1$, the marginal prior density of Σ^{-1}, i.e.

$$\Pi(\Sigma^{-1}) \propto |\Sigma^{-1}|^{\frac{v-p-1}{2}} e^{-\frac{1}{2} tr \Gamma^{-1}\Sigma^{-1}}$$

is replaced by

$$\sigma^2 \sim IG(\alpha, \beta).$$

Furthermore

$$\mu_0, \mu_1, \mu_2 \sim iN(\theta, \phi)$$

and

$$\phi = \frac{\sigma^2}{\delta}$$

and (2.3.2) is replaced by

$$f(X|k = n, \alpha, \beta) = \pi^{-\frac{(n-1)}{2}} n^{-\frac{1}{2}} \frac{\Gamma\left[\frac{1}{2}(n + 2\alpha - 1)\right]}{\Gamma(\alpha)} [2\beta]^{\alpha} [s_n^2 + 2\beta]^{-\frac{n+2\alpha-1}{2}}$$

where

$$s_n^2 = \sum_{i=1}^{n}(x_i - \overline{x}_n)^2.$$

It also follows from (2.2.20) that

$$f(X|k,\delta,\alpha,\beta) \quad = \quad c_1 2^{-\frac{n}{2}}\delta^{-\frac{1}{2}}\Gamma\left(\frac{2\alpha+n-1}{2}\right)\left[\Gamma\left(\frac{2\alpha+n}{2}\right)\right]^{-1}\pi^{-\left(\frac{n-1}{2}\right)}$$

$$[2k_1k_2 + \delta k_1 + \delta k_2]^{-\frac{1}{2}}[T_{\delta\beta}]^{-\left(\frac{2\alpha+n-1}{2}\right)}$$

where

$$T_{\delta\beta} = 2\beta + \sum_{i=1}^{k} x_i^2 + \sum_{i=k+1}^{n} x_i^2 - (k_1+\delta)^{-1}k^2\overline{x}_{1k}^2 - (k_2+\delta)^{-1}k_2^2\overline{x}_{2k}^2$$

$$-\hat{\theta}_\delta^2(2\delta - (k_2+\delta)^{-1}\delta^2 - (k_1+\delta)^{-1}\delta^2), \tag{2.7.1}$$

$\hat{\theta}_\delta$ is the same as in (2.2.21) and c_1 is the same as c_{1p} but with $p = 1$.

To summarize, for $\beta \to 0$ and $\alpha \to 0$

$$f(X|k=n) = n^{-\frac{1}{2}}(s_n^2)^{-\frac{n-1}{2}},$$

$$f(X|k,\delta) = \delta^{-\frac{1}{2}}[2k_1k_2 + \delta k_1 + \delta k_2]^{-\frac{1}{2}}[T_\delta]^{-\frac{n-1}{2}},$$

where T_δ is the same as in equation (2.7.1) but with $\beta = 0$.

The posterior distributions $\Pi(k = n|X,\delta)$ and $\Pi(k|X,\delta)$ follow from (2.3.4). Under the assumption of exactly one change-point and $\delta \to 0$, the posteriors follow as

$$\mu_1|k,X \sim t(n,\overline{x}_{1k},T_1), \quad \mu_2|k,X \sim t(n,\overline{x}_{2k},T_2) \text{ and } \sigma^2|k,X \sim IG\left(\frac{n}{2},\ \frac{s_{1k}^2+s_{2k}^2}{2}\right)$$

where

$$T_1 = \frac{kn}{s_{1k}^2 + s_{2k}^2}, \quad T_2 = \frac{n(n-k)}{s_{1k}^2 + s_{2k}^2} \text{ and } s_{1k}^2 = \sum^{k_1}(x_i - \overline{x}_{1k})^2, \quad s_{2k}^2 = \sum^{k_2}(x_i - \overline{x}_{2k})^2.$$

Furthermore $\Pi(k|X)$ reduces to

$$\Pi(k|X) \propto \frac{(s_{1k}^2 + s_{2k}^2)^{-\frac{n}{2}}}{(k_1 k_2)^{\frac{1}{2}}}. \tag{2.7.2}$$

Under the assumption of no change-point, the posteriors are

$$\mu_0 | k = n, \delta, X \sim t(n-1, \overline{x}_n, T)$$

and

$$\sigma^2 | k = n, X \sim IG\left(\frac{n-1}{2}, \frac{s_n^2}{2}\right)$$

where $T = \dfrac{n(n-1)}{s_n^2}$.

Note from equation (2.7.2) that the distribution of $\dfrac{s_{1k}^2 + s_{2k}^2}{\sigma^2}$ is chi-squared with $n-2$ degrees of freedom under model M_0, independent of k. Although the expectation of $(s_{1k}^2 + s_{2k}^2)^{-\frac{n}{2}}$ does not exist, it does not depend on k and $\Pi(k|X)$ is proportional to $(k_1 k_2)^{-\frac{1}{2}}$, which is a minimum when $k = \frac{n}{2}$. So the posterior distribution of k would be u-shaped. This makes sense since failing to find an obvious change-point in the data, the analysis concludes that the change-point should be near the end points of the sequence where it is less detectable. If that happens, the model is probably wrong and the model incorporating the possibility of no change should be used.

2.7.2 A CHANGE IN THE MEAN AND VARIANCE

2.7.2.1 CASE 1

In this case the prior assumptions are as in (2.2.25).

If $p = 1$, (2.2.25) is replaced by

$$\sigma_i^2 \sim IG(\alpha, \beta).$$

Furthermore

$\mu_i \sim iN(\theta_i, \phi_i), \quad i = 0, 1, 2$ and $\phi_i = \frac{\sigma_i^2}{\delta_i}$.

After integrating θ_1 and θ_2 out, (2.2.28) reduces to

$$\begin{aligned} f(X|k, \alpha, \beta) &= c_6(\Gamma(\alpha))^{-2}\beta^{2\alpha}2^{\frac{2\alpha-n}{2}}\pi^{-\left(\frac{n-2}{2}\right)}(k_1k_2)^{-\frac{1}{2}}\Gamma\left(\frac{2\alpha+k_2-1}{2}\right)\Gamma\left(\frac{2\alpha+k_1-1}{2}\right) \\ & \left[\Gamma\left(\frac{2\alpha+k_1}{2}\right)\Gamma\left(\frac{2\alpha+k_2}{2}\right)\right]^{-1}\left[2\beta+s_{1k}^2\right]^{-\left(\frac{2\alpha+k_1-1}{2}\right)}\left[2\beta+s_{2k}^2\right]^{-\left(\frac{2\alpha+k_2-1}{2}\right)} \end{aligned}$$

where c_6 is the same as c_{6p} but with $p = 1$ and $v = 2\alpha$. Note that in the univariate case, the posterior probability is independent of δ_1 and δ_2.

Furthermore, if we place a vague prior on β, $\Pi(\beta)\ \alpha\ \frac{1}{\beta}$, it follows that (see Gradshteyn and Ryzhik (1980, p285-286)

$$f(X|k = n) = \pi^{-\left(\frac{n-1}{2}\right)}n^{-\frac{1}{2}}\Gamma\left(\frac{n-1}{2}\right)(s_n^2)^{\alpha} \qquad (2.7.3)$$

and

$$\begin{aligned} f(X|k, \alpha) &= c_6\pi^{-\left(\frac{n-2}{2}\right)}(k_1k_2)^{-\frac{1}{2}}B\left(2\alpha, \frac{n-2}{2}\right)\left(\frac{s_{1k}^2}{2}\right)^{-\left(\frac{k_1+2\alpha-1}{2}\right)}\left(\frac{s_{2k}^2}{2}\right)^{-\left(\frac{k_2+2\alpha-1}{2}\right)} \\ & [\Gamma(\alpha)]^{-2}2^{\frac{2\alpha-n}{2}}\Gamma\left(\frac{2\alpha+k_2-1}{2}\right)\Gamma\left(\frac{2\alpha+k_1-1}{2}\right)\left[\Gamma\left(\frac{2\alpha+k_1}{2}\right)\Gamma\left(\frac{2\alpha+k_2}{2}\right)\right]^{-1} \\ & {}_2F_1\left(\frac{k_1+2\alpha-1}{2};\ 2\alpha;\ \frac{n+4\alpha-2}{2};\ 1-\frac{k_2+2\alpha-1}{s_{1k}^2}\right) \end{aligned} \qquad (2.7.4)$$

where ${}_2F_1(\cdot\ ,\ \cdot\ ;\ \cdot\ ;\ \cdot)$ is the standard hypergeometric function.

The posterior distributions $\Pi(k = n|X)$ and $\Pi(k|X, \alpha)$ follow from (2.3.4). Furthermore, under the assumption of exactly one change-point and $\alpha, \beta \to 0$, the posteriors follow as

$$\mu_1|k, X \sim t(k_1, \overline{x}_{1k}, T_3)$$
$$\mu_2|k, X \sim t(k_2, \overline{x}_{2k}, T_4)$$

where $T_3 = \dfrac{k_1^2}{s_{1k}^2}$ and $T_4 = \dfrac{k_2^2}{s_{2k}^2}$ and $\sigma_1^2|k, X \sim IG\left(\frac{k_1}{2}; \frac{s_{1k}^2}{2}\right)$ and $\sigma_2^2|k, X \sim IG\left(\frac{k_2}{2}; \frac{s_{2k}^2}{2}\right)$.

Furthermore (from 2.2.24) $\Pi(k|X)$ reduces to

$$\Pi(k|X) \propto \Gamma\left(\tfrac{k_1-1}{2}\right)\Gamma\left(\tfrac{k_2-1}{2}\right)\left[\Gamma\left(\tfrac{k_1}{2}\right)\Gamma\left(\tfrac{k_2}{2}\right)\right]^{-1}(k_1k_2)^{-\frac{1}{2}}\left(s_{1k}^2\right)^{-\left(\frac{k_1-1}{2}\right)}\left(s_{2k}^2\right)^{-\left(\frac{k_2-1}{2}\right)},$$
$$k = 2, \ldots, n-2. \qquad (2.7.5)$$

As in the case of change-points in the mean alone, the expectation of $(s_{1k}^2)^{-(\frac{k_1-1}{2})}(s_{2k}^2)^{-(\frac{k_2-1}{2})}$ does not exist under model M_0, but substituting in the individual expectations, we have approximately that $\Pi(k|X) \propto (k_1-1)^{\frac{k_1-1}{2}}(k_2-1)^{\frac{k_2-1}{2}}$ under M_0. In contrast with the previous case, this posterior is unimodal with maximum at $k = \frac{n}{2}$.

2.7.2.2 CASE 2

In case 2, with $p = 1$, the prior assumptions (2.2.42) and (2.2.43) reduces to

$$\sigma_1^2 \sim IG(\alpha, \beta)$$

and

$$\sigma_2^2 = \gamma\sigma_1^2.$$

Furthermore

$$\mu_i \sim iN(\theta, \phi)$$

and

$$\phi = \frac{\sigma_1^2}{\delta}.$$

and

$$\phi = \frac{\sigma_1^2}{\delta}.$$

(2.2.32) is replaced by

$$f(X|k,\gamma,\alpha,\beta) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha}\pi^{-\left(\frac{n-1}{2}\right)}\gamma^{\frac{k_2}{2}}\Gamma\left(\frac{2\alpha+n-1}{2}\right)2^{\alpha}$$

$$\int_0^{\infty}\delta^{-\frac{1}{2}}[2\gamma k_1k_2+\delta k_1+\delta\gamma k_2]^{-\frac{1}{2}}[t_{\delta\gamma\beta}]^{-\left(\frac{2\alpha+n-1}{2}\right)}$$

where

$$t_{\delta\gamma\beta} = 2\beta+\sum^{k_1}x_i^2+\gamma\sum^{k_2}x_i^2-(k_1+\delta)^{-1}k_1^2\overline{x}_{1k}^2-[\gamma k_2+\delta]^{-1}k_2^2\gamma^2\overline{x}_{2k}^2-$$

$$[2\delta-(\gamma k_2+\delta)^{-1}\delta^2-(k_1+\delta)^{-1}\delta^2]\hat{\theta}_{\delta\gamma}^2$$

and $\hat{\theta}_{\delta\gamma}$ is given below (2.2.49).

If we once again place a vague prior on $\beta, \Pi(\beta) \propto \frac{1}{\beta}$, it follows that (see Gradshteyn and Ryzhik (1980, p286) $f(X|k=n)$ is given by (2.7.3) and that

$$f(X|k,\gamma,\alpha,\delta) = \frac{1}{\Gamma(\alpha)}\pi^{-\left(\frac{n-1}{2}\right)}\gamma^{\frac{k_2}{2}}\Gamma\left(\frac{2\alpha+n-1}{2}\right)B\left(\alpha,\frac{n-1}{2}\right)$$

$$\delta^{\frac{1}{2}}[2\gamma k_1k_2+\delta k_1+\delta\gamma k_2]^{-\frac{1}{2}}(t_{\delta\gamma})^{\alpha} \qquad (2.7.6)$$

where $t_{\delta\gamma}$ is the same as $t_{\delta\gamma\beta}$ but with 2β omitted.

The posterior distributions $\Pi(k=n|X)$ and $\Pi(k|X,\gamma,\alpha,\delta)$ follows from (2.3.4).

Under the assumption of exactly one change-point and $\delta \rightarrow 0$ the posteriors become

$$\mu_1|k,X,\gamma \sim t\,(n,\ \ \overline{x}_{1k},\ \ T_5)$$
$$\mu_2|k,X,\gamma \sim t\,(n,\ \ \overline{x}_{2k},\ \ T_6)$$

where

$$T_5 = \frac{k_1n}{s_{1k}^2+\gamma s_{2k}^2} \quad \text{and} \quad T_6 = \frac{\gamma k_2 n}{s_{1k}^2+\gamma s_{2k}^2}$$

and

$$\sigma_1^2|k, X, \gamma \sim IG\left(\frac{n}{2}, \quad \frac{s_{1k}^2 + \gamma s_{2k}^2}{2}\right)$$

and

$$\Pi(\gamma|k, X) \propto \gamma^{\frac{k_2}{2}-1}[k_2\gamma]^{-\frac{1}{2}}\left[s_{1k}^2 + \gamma s_{2k}^2\right]^{-\frac{n}{2}}.$$

Furthermore $\Pi(k|X, \gamma)$ reduces to

$$\Pi(k|X, \gamma) \propto \frac{\gamma^{\frac{k_2-1}{2}}\left[s_{1k}^2 + \gamma s_{2k}^2\right]^{-\frac{n}{2}}}{(k_1 k_2)^{\frac{1}{2}}}.$$

This is again compatible with the case of a change in the mean alone, giving a u-shaped posterior when there is no evidence of a change-point. The shape is asymmetrical, depending on the value of γ.

2.7.3 A CHANGE IN THE VARIANCE

If $p = 1$, (2.2.53) is replaced by

$\sigma_i^2 \sim IG(\alpha, \beta), \quad i = 0, 1; \quad \sigma_2^{-2} = \gamma\sigma_1^{-2}$ and $\mu \sim N(\theta, \phi)$.

If a vague prior $\Pi(\beta) \propto \frac{1}{\beta}$ is placed on β, it follows that $f(X|k = n)$ is given by (2.7.3) and that

$$\begin{aligned} f(X|k, \alpha, \beta, \gamma) &= (2\pi)^{-\left(\frac{n-1}{2}\right)}(k_1 + \gamma k_2)^{-\frac{1}{2}}\gamma^{\frac{k_2}{2}}\frac{\beta^\alpha}{\Gamma(\alpha)}\Gamma\left(\frac{n + 2\alpha - 1}{2}\right)\cdot \\ & \left\{\beta + \frac{1}{2}\left[s_{1k}^2 + \gamma s_{2k}^2 + \frac{\gamma k_1 k_2(\overline{x}_{1k} - \overline{x}_{2k})^2}{k_1 + \gamma k_2}\right]\right\}^{-\left(\frac{n+2\alpha-1}{2}\right)} \end{aligned}$$

so that it follows, after integrating out β (see Gradshteyn and Ryzhik (1980, p285), that

$$\begin{aligned} f(X|k, \gamma) &= \pi^{-\left(\frac{n-1}{2}\right)}(k_1 + \gamma k_2)^{\frac{n-2}{2}}\gamma^{\frac{k_2}{2}}\Gamma\left(\frac{n-1}{2}\right)[(k_1 + \gamma k_2)(s_{1k}^2 + \gamma s_{2k}^2) + \\ & \gamma k_1 k_2(\overline{x}_{1k} - \overline{x}_{2k})^2]^{-\left(\frac{n-1}{2}\right)}. \end{aligned} \qquad (2.7.7)$$

The posterior distributions $\Pi(k = n|X)$ and $\Pi(k|X, \gamma, \alpha)$ follows from (2.3.4). Under the assumption of exactly one change-point it follows that

$$\Pi(k|X,\gamma) \propto \gamma^{\frac{k_2}{2}} (k_1 + \gamma k_2)^{\frac{n-2}{2}} \left[(k_1 + \gamma k_2)(s_{1k}^2 + \gamma s_{2k}^2) + \gamma k_1 k_2 (\overline{x}_{1k} - \overline{x}_{2k})\right]^{-\left(\frac{n-1}{2}\right)}. \quad (2.7.8)$$

When no evidence of a change is present, i.e. $k_2 s_{1k}^2 \simeq k_1 s_{2k}^2$ and $\overline{x}_{1k} \simeq \overline{x}_{2k}$, then approximately $\Pi(k) \propto \gamma^{\frac{k_2}{2}} (k_1 + \gamma k_2)^{-\frac{n}{2}} k_1^{\frac{n-1}{2}}$. This is a unimodal function while $\gamma > 1$ with mode at $k = f(\gamma)$ as long as $f(\gamma) < n$ where

$$f(\gamma) = \frac{-n(1 - \gamma \ell n\gamma - \gamma) - \sqrt{n^2(1 - \gamma \ell n \gamma - \gamma)^2 - 4n\gamma \ell n\gamma(\gamma - 1)(n-1)}}{2\ell n\gamma(\gamma - 1)}.$$

Otherwise the maximum will be at $k = n - 1$.

Under the assumption of exactly one change-point with $\alpha, \beta \to 0$, the posteriors become

$$\mu|k, X, \gamma \sim t(n-3, \quad \hat{\mu}_\gamma, \quad T_7)$$

and

$$\sigma_1^2|k, X, \gamma \sim IG\left(\frac{n-3}{2}, \quad \frac{T_\gamma}{2}\right)$$

where

$$T_7 = (n-3)(k_1 + \gamma k_2) T_\gamma^{-1}.$$

2.7.4 BAYES FACTORS AND MODEL COMPARISONS

In this section we'll compare the different models that have been discussed so far for a given k (or unknown k) to see which model best fits the data. In paragraph 2.5, Bayes factors are discussed given a certain model. The corresponding Bayes factors in the univariate case follows directly from there.

We will consider the following four models with a maximum of one change-point, $2 \leq k \leq n-2$, where M_3 is the so-called encompassing model:

$$M_0 : X_i \sim N(\mu_1, \sigma_1^2); \qquad i = 1, \dots, n$$

$$\begin{aligned} M_1 : \ & X_i \sim N(\mu_1, \sigma_1^2); \quad & i = 1, \dots, k \\ & X_i \sim N(\mu_1, \sigma_2^2); \quad & i = k+1, \dots, n \end{aligned}$$

$$\begin{aligned} M_2 : \ & X_i \sim N(\mu_1, \sigma_1^2); \quad & i = 1, \dots, k \\ & X_i \sim N(\mu_2, \sigma_1^2); \quad & i = k+1, \dots, n \end{aligned}$$

$$\begin{aligned} M_3 : \ & X_i \sim N(\mu_1, \sigma_1^2); \quad & i = 1, \dots, k \\ & X_i \sim N(\mu_2, \sigma_2^2); \quad & i = k+1, \dots, n. \end{aligned}$$

Notice that in paragraph 2.5, B_{ij} defines the Bayes factor for a change at i compared to a change at j for a certain model. In this paragraph, B_{ijk} will denote the Bayes factor of model i versus model j for a fixed k, where $i, j = 0, 1, 2, 3$. When k is omitted, B_{ij} would refer to the Bayes factor when summed over k.

2.7.4.1 USUAL BAYES FACTORS

If comparing between models, it is important to make the prior distributions as exchangeable as possible. So the prior assumptions on the parameters of some of the four models will be slightly different from the previous sections. We assume that, for all four models where applicable,

$$\mu_1, \mu_2 \sim N(\theta, \phi) \quad \text{and} \quad \sigma_1^2, \sigma_2^2 \sim IG(\alpha, \beta)$$

where $\phi = \frac{\sigma_1^2}{\delta}$ and $\Pi(\theta) \propto 1$.

Then, with $k_2 = n - k_1$,

$$f_3(X|k, \delta, \alpha, \beta) = \pi^{-\frac{n}{2}} 2^{\frac{4\alpha-1}{2}} \delta^{\frac{1}{2}} (k_1 + 2\alpha)^{\frac{1}{2}} \left[\frac{\beta^\alpha}{\Gamma(\alpha)}\right]^2 \Gamma\left(\frac{k_1+2\alpha}{2}\right) \Gamma\left(\frac{k_2+2\alpha}{2}\right)$$

$$\int_{-\infty}^{\infty} \left[2\beta + s_{2k}^2 + k_2(\mu_2 - \bar{x}_2)^2\right]^{-\left(\frac{k_2+2\alpha}{2}\right)}$$

$$\left[2\beta + s_{1k}^2 + \frac{k_1\frac{\delta}{2}}{k_1+\frac{\delta}{2}}(\mu_2-\bar{x}_1)^2\right]^{-\frac{k_1+2\alpha}{2}} d\mu_2, \tag{2.7.9}$$

which is an integral of a multiple t distribution. Furthermore, if we place a vague prior on β, $\Pi(\beta) \propto \frac{1}{\beta}$, as with θ; both being parameters with common meaning for all models, then (for the integral, see Gradshteyn and Ryzhik (1980, p286)

$$\begin{aligned}
f_3(X|k,\delta,\alpha) &= \pi^{-\frac{n}{2}} 2^{\frac{n+4\alpha-1}{2}} \delta^{\frac{1}{2}} (k_1+2\alpha)^{\frac{1}{2}} \frac{1}{B(\alpha,\alpha)} B\left(\frac{k_1+2\alpha}{2}, \frac{k_2+2\alpha}{2}\right) \\
&\int_0^\infty \left[s_{2k}^2 + k_2(\mu_2-\bar{x}_2)^2\right]^{-\left(\frac{k_2+2\alpha}{2}\right)} \left[s_{1k}^2 + \frac{k_1\delta}{2k_1+\delta}(\mu_2-\bar{x}_1)^2\right]^{-\left(\frac{k_1+2\alpha}{2}\right)} \\
&{}_2F_1\left(\frac{k_2+2\alpha}{2},\; 2\alpha;\; \frac{n+4\alpha}{2};\; 1 - \frac{s_{1k}^2 + \frac{k_1\delta}{2k_1+\delta}(\mu_2-\bar{x}_2)^2}{s_{2k}^2 + k_2(\mu_2-\bar{x}_1)^2}\right) d\mu_2.
\end{aligned}$$

This however, is a very complex integral as well as a function of the two hyperparameters δ and α.

For the other models it follows that

$$\begin{aligned}
f_2(X|k,\delta,\alpha,\beta) &= 2^\alpha \pi^{-\left(\frac{n-1}{2}\right)} \left[\frac{\delta}{k_1k_2+\frac{n\delta}{2}}\right]^{\frac{1}{2}} \frac{\beta^\alpha}{\Gamma(\alpha)} \Gamma\left(\frac{n-1}{2}+\alpha\right) \\
&\left[2\beta + s_{1k}^2 + s_{2k}^2 + \frac{k_1k_2\frac{\delta}{2}(k_1+\frac{\delta}{2})(\bar{x}_1-\bar{x}_2)^2}{(k_1+\frac{\delta}{2})(k_1k_2+\frac{n\delta}{2})}\right]^{-\left(\frac{n-1}{2}+\alpha\right)}
\end{aligned} \tag{2.7.10}$$

and

$$\begin{aligned}
f_2(X|k,\delta) &= \pi^{\frac{n-1}{2}} \left[\frac{\delta}{k_1k_2+\frac{n\delta}{2}}\right]^{\frac{1}{2}} \Gamma\left(\frac{n-1}{2}\right) \cdot \\
&\left[s_{1k}^2 + s_{2k}^2 + \frac{k_1k_2\frac{\delta}{2}\left(k_1+\frac{\delta}{2}\right)(\bar{x}_1-\bar{x}_2)^2}{(k_1+\frac{\delta}{2})(k_1k_2+\frac{n\delta}{2})}\right]^{-\left(\frac{n-1}{2}\right)}.
\end{aligned}$$

$$f_1(X|k,\alpha,\beta) = \pi^{-\frac{n}{2}}2^{2\alpha}[\tfrac{\beta^\alpha}{\Gamma(\alpha)}]^2\Gamma(\tfrac{k_1+2\alpha}{2})\Gamma(\tfrac{k_2+2\alpha}{2})\int_{-\infty}^{\infty}[2\beta+s_{1k}^2+k_1(\mu-\bar{x}_1^2)^2]^{-\frac{k_1+2\alpha}{2}}\cdot$$

$$[2\beta+s_{2k}^2+k_2(\mu-\bar{x}_2)^2]^{-(\frac{k_2+2\alpha}{2})}d\mu \qquad (2.7.11)$$

and

$$f_1(X|k,\alpha) = \pi^{-(\frac{n-1}{2})}\frac{1}{B(\alpha,\alpha)}\Gamma(\frac{n-1}{2})\int_0^\infty \gamma^{\frac{k_1}{2}+\alpha-1}(1+\gamma)^{-2\alpha}(\gamma k_1+k_2)^{-\frac{1}{2}}\cdot$$

$$\left[\gamma s_{1k}^2+s_{2k}^2+\frac{\gamma k_1k_2(\bar{x}_1-\bar{x}_2)^2}{\gamma k_1+k_2}\right]^{-(\frac{n-1}{2})}d\gamma.$$

$$f_0(X|k,\alpha,\beta) = 2^\alpha\pi^{-(\frac{n-1}{2})}n^{-\frac{1}{2}}\frac{\beta^\alpha}{\Gamma(\alpha)}\Gamma\left(\frac{n+2\alpha-1}{2}\right)[2\beta+s_n^2]^{-\left(\frac{n+2\alpha-1}{2}\right)} \qquad (2.7.12)$$

and

$$f_0(X|k) = \pi^{-(\frac{n-1}{2})}n^{-\frac{1}{2}}\Gamma(\frac{n-1}{2})[s_n^2]^{-(\frac{n-1}{2})}.$$

Note that these results are slightly different from those in paragraph 2.7.2 and 2.7.3 for model M_0 and M_1. Also, with a vague prior on β, the parameter α disappears if the model assumes no change in the variance. Similarly, the parameter δ disappears in models with no change in the mean. This result is equal to the one obtained by simply putting a vague improper prior on any first-stage parameter (μ or σ^2) which doesn't change in the model.

The Bayes factors for a given k are then

$$B_{ijk} = \frac{f_i(X|k,\delta,\alpha)}{f_j(X|k,\delta,\alpha)}.$$

The unconditional Bayes factors would be, with uniform prior on k,

$$B_{ij} = \frac{\sum_k f_i(X|k,\delta,\alpha)}{\sum_k f_j(X|k,\delta,\alpha)}.$$

2.7.4.2 THE INTRINSIC BAYES FACTOR (IBF)

Berger and Pericchi (1995, 1996, 1997) proposed using all possible minimal training samples and averaging the resulting Bayes factors.

For the Intrinsic Bayes factor we use the vague priors as in paragraph 2.5 where $\Pi(\sigma_1, \sigma_2) \propto \frac{1}{\sigma_1^2 \sigma_2^2}$ and $\Pi(\mu_1, \mu_2) \propto 1$. Let $m_i^N(x)$ denote the marginal density of the whole data set for model i when using the above non-informative priors. Then

$$m_3^N(x) = \frac{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{n-k-1}{2}\right)}{(k_1 k_2)^{\frac{1}{2}} \pi^{\frac{n-2}{2}} (s_{1k}^2)^{\frac{k_1-1}{2}} (s_{2k}^2)^{\frac{k_2-1}{2}}} \tag{2.7.13}$$

where s_{1k}^2 and s_{2k}^2 are defined in paragraph 2.7.1.

Also

$$m_2^N(x) = \frac{\Gamma\left(\frac{n-2}{2}\right)}{(k_1 k_2)^{\frac{1}{2}} \pi^{\frac{n-2}{2}} (s_{1k}^2 + s_{2k}^2)^{\frac{n-2}{2}}}, \tag{2.7.14}$$

$$\begin{aligned} m_1^N(x) \quad &= \quad \Gamma\left(\frac{n-1}{2}\right) \pi^{-\left(\frac{n-1}{2}\right)} \int_0^\infty \gamma^{\frac{k}{2}-1} (n-k+k\gamma)^{\frac{n}{2}-1} \{k_1 \gamma^2 s_{1k}^2 + [k_2 s_{1k}^2 + k_1 s_{2k}^2 + \\ &\qquad k_1 k_2 (\overline{x}_1 - \overline{x}_2)^2]\gamma + k_2 s_{2k}^2\}^{-\left(\frac{n-1}{2}\right)} d\gamma, \end{aligned} \tag{2.7.15}$$

which can be numerically integrated, and

$$m_0^N(x) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{n^{\frac{1}{2}} \pi^{\frac{n-1}{2}} (s_n^2)^{\left(\frac{n-1}{2}\right)}}. \tag{2.7.16}$$

It then follows that

$$B_{31}^N(x) = \frac{m_3^N(x)}{m_1^N(x)}, \tag{2.7.17}$$

$$B_{32}^N(x) = \frac{B\left(\frac{k_1-1}{2}, \frac{k_2-1}{2}\right)\left[s_{1k}^2 + s_{2k}^2\right]^{\frac{n-2}{2}}}{\left(s_{1k}^2\right)^{\frac{k_1-1}{2}}\left(s_{2k}^2\right)^{\frac{k_2-1}{2}}} \tag{2.7.18}$$

and

$$B_{30}^N(x) = \frac{\Gamma\left(\frac{k_1-1}{2}\right)\Gamma\left(\frac{k_2-1}{2}\right) n^{\frac{1}{2}}\pi^{\frac{1}{2}}\left(s_n^2\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)(k_1k_2)^{\frac{1}{2}}\left(s_{1k}^2\right)^{\frac{k_1-1}{2}}\left(s_{2k}^2\right)^{\frac{k_2-1}{2}}}. \tag{2.7.19}$$

The minimal sample size is $\ell = 4$ with $\ell(x) = (x_{t_1}, x_{t_2}, x_{t_3}, x_{t_4})$ where $t_1 < t_2 \leq k < t_3 < t_4$ and $x_{t_1} \neq x_{t_2}, \quad x_{t_3} \neq x_{t_4}$. Then if $m_i(\ell(x))$ is the marginal density of the minimal sample,

$$m_3(\ell(x)) = \frac{1}{2s_1^2(k)s_2^2(k)}, \tag{2.7.20}$$

where $$s_1^2(k) = \tfrac{1}{2}(x_{t_1} - x_{t_2})^2 \text{ and } s_2^2(k) = \tfrac{1}{2}(x_{t_3} - x_{t_4})^2,$$

$$m_2(\ell(x)) = \frac{1}{2\pi(s_1^2(k) + s_2^2(k))}, \tag{2.7.21}$$

$$m_1(\ell(x)) = \frac{s_1^2(k) + s_2^2(k)}{\sqrt{2}\pi s_1^2(k)s_2^2(k)\left[(s_1^2(k) + s_2^2(k))^2 + 2(\overline{x}_1 - \overline{x}_2)^2\right]} \tag{2.7.22}$$

where $\overline{x}_1 = \frac{1}{2}(x_{t_1} + x_{t_2})$, $\overline{x}_2 = \frac{1}{2}(x_{t_3} + x_{t_4})$ and

$$m_0(\ell(x)) = \frac{1}{4\pi(s^2(n))^{\frac{3}{2}}} \tag{2.7.23}$$

where $s^2(n) = \sum_{i=1}^{4}(x_{t_i} - \overline{x}_t)^2$.

It then follows that

$$B_{13}^N(x(\ell)) = \frac{\sqrt{2}(s_1^2(k)+s_2^2(k))}{\pi[s_1^2(k)+s_2^2(k))^2+2(\overline{x}_1-\overline{x}_2)^2]}, \tag{2.7.24}$$

$$B_{23}^N(x(\ell)) = \frac{s_1^2(k)s_2^2(k)}{\pi(s_1^2(k)+s_2^2(k))} \tag{2.7.25}$$

and

$$B_{03}^N(x(\ell)) = \frac{s_1^2(k)s_2^2(k)}{2\pi(s^2(n))^{\frac{3}{2}}}. \tag{2.7.26}$$

Then

$$B_{ji}^A = B_{ji}^N(x)\cdot\frac{\bar{B}_{i0}^N(x(\ell))}{\bar{B}_{j0}^N(x(\ell))}$$

is the Arithmetic Intrinsic Bayes factor where

$$\bar{B}_{ij}^N(x(\ell)) = \frac{1}{L}\sum_{\ell=1}^{L} B_{ij}^N(x(\ell)).$$

Similarly the geometric and median intrinsic Bayes factors can be obtained.

2.7.4.3 THE FRACTIONAL BAYES FACTOR (FBF)

For the fractional Bayes factor, with the same priors as the intrinsic, it follows that

$$m_0^F = \pi^{-\frac{n(1-b)}{2}} b^{\frac{nb}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{nb-1}{2}\right)} (s_n^2)^{-\frac{n(1-b)}{2}}, \tag{2.7.27}$$

$$m_1^F = \pi^{-\frac{n(1-b)}{2}} b^{\frac{nb}{2}} \frac{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)\int\left[s_{1k}^2+k_1(\mu-\overline{x}_{1k})^2\right]^{-\frac{k_1}{2}}\left[s_{2k}^2+k_2(\mu-\overline{x}_{2k})^2\right]^{-\frac{k_2}{2}}d\mu}{\Gamma\left(\frac{bk_1}{2}\right)\Gamma\left(\frac{bk_2}{2}\right)\int\left[s_{1k}^2+k_1(\mu-\overline{x}_{1k})^2\right]^{-\frac{bk_1}{2}}\left[s_{2k}^2+k_2(\mu-\overline{x}_{2k})^2\right]^{-\frac{bk_2}{2}}d\mu}, \tag{2.7.28}$$

$$m_2^F = \pi^{-\frac{n(1-b)}{2}} b^{\frac{nb}{2}} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{nb-2}{2}\right)} (s_{1k}^2 + s_{2k}^2)^{-\frac{n(1-b)}{2}} \tag{2.7.29}$$

and

$$m_3^F = \frac{\Gamma\left(\frac{k_1-1}{2}\right)\Gamma\left(\frac{k_2-1}{2}\right)\Pi^{-\frac{n(1-b)}{2}} b^{\frac{nb}{2}} \left(s_{1k}^2\right)^{-\frac{k_1(1-b)}{2}} \left(s_{2k}^2\right)^{-\frac{k_2(1-b)}{2}}}{\Gamma\left(\frac{bk_1-1}{2}\right)\Gamma\left(\frac{bk_2-1}{2}\right)}. \tag{2.7.30}$$

The FBF's, B_{ij}^F, are then given by $B_{ij}^F = \dfrac{m_i^F}{m_j^F}$, $\quad i, j = 0, 1, 2, 3.$

2.7.5 AUTOCORRELATION

In all previous models the sequence of random variables were assumed to be conditionally independent. Here we will assume that the observations are correlated with common correlation $\text{Cor}(X_i, X_{i+t}) = \rho^t, \quad -1 < \rho < 1$ and variance σ^2 under both model M_0 (no change) and model M_1 (one change-point). The theory can easily be extended to multiple change-points. MacNeill, Tang and Jandhyala (1991) examined the annual discharges of the Nile river at Aswan for possible change-points under the assumption of serial correlation. For a more general analysis in multiple regression, see Garisch and Groenewald (1999).

Under $M_0(k = n)$ we have the model

$$\boldsymbol{X}(n \times 1) \sim N(\mathbf{1}_n \mu_0, \sigma^2 R), \tag{2.7.31}$$

where $\mathbf{1}_n$ is a $(n \times 1)$ vector of ones and

$$\{R_{ij}\} = \rho^{|i-j|}, \quad i, j = 1, \ldots, n. \tag{2.7.32}$$

The prior specifications are

$$\mu_0 \sim N(\theta, \delta\sigma^2), \quad \sigma^2 \sim IG(\alpha, \beta), \quad \rho \sim U(-1,1) \text{ and } \Pi(\theta) \propto 1. \tag{2.7.33}$$

Under M_1 with change-point at k we have

$$\boldsymbol{X} \sim N(M_k\boldsymbol{\mu}, \sigma^2 R),$$

where

$$M_k = \begin{pmatrix} \mathbf{1}_k & 0 \\ 0 & \mathbf{1}_{n-k} \end{pmatrix} \text{ and } \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \tag{2.7.34}$$

with

$$\boldsymbol{\mu} \sim N(\mathbf{1}_2\theta, \delta\sigma^2 I_2) \text{ and } \Pi(k|k \neq n) = \frac{1-q}{n-1}, \quad k = 1, \cdots, n-1 (n \geq 5). \tag{2.7.35}$$

Under M_0, the joint marginal of $\boldsymbol{X}$ and ρ reduces to

$$f(X, \rho|k = n, \alpha, \beta) = 2^{\alpha - \frac{1}{2}} \pi^{-(\frac{n-1}{2})} n^{-\frac{1}{2}} \frac{\beta^\alpha \Gamma\left(\frac{n+2\alpha}{2}\right)}{\Gamma(\alpha)} |R|^{-\frac{1}{2}} \left[2\beta + \boldsymbol{x}' R^{-1} H_0 R^{-1} \boldsymbol{x}\right]^{-(\frac{n+2\alpha}{2})} \tag{2.7.36}$$

where $H_0 = R - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$.

Then, if $\alpha, \beta \longrightarrow 0$, the marginal posterior distribution of ρ under M_0 is

$$\Pi(\rho|X, k = n) \propto |R|^{-\frac{1}{2}} [\boldsymbol{x}' R^{-1} H_0 R^{-1} \boldsymbol{x}]^{-\frac{n}{2}}. \tag{2.7.37}$$

Under M_1 it follows that

$$f(X,\rho|k,\delta,\alpha,\beta) \;=\; 2^{-\alpha-\frac{3}{2}}\pi^{-(\frac{n-1}{2})}\delta^{-\frac{1}{2}}\frac{\beta^{\alpha}\Gamma(\frac{n+2\alpha-1}{2})}{\Gamma(\alpha)}|R|^{-\frac{1}{2}}|N_\delta|^{-\frac{1}{2}}.$$

$$[2\beta+\boldsymbol{x}'R^{-1}H_\delta R^{-1}\boldsymbol{x}]^{-(\frac{n+2\alpha-1}{2})} \tag{2.7.38}$$

where $H_\delta = R - M_k N_\delta^{-1} M_k'$, $N_\delta = M_k' R^{-1} M_k + \frac{1}{\delta} J$ and

$$J = I_2 - \mathbf{1}_2\mathbf{1}_2'. \tag{2.7.39}$$

In this case the posterior of ρ (for $\alpha, \beta \to 0$ as well as $\delta \to \infty$) is given by

$$\Pi(\rho|X,k,\delta) \propto |R|^{-\frac{1}{2}}|M_k'R^{-1}M_k|^{-\frac{1}{2}}\left[\boldsymbol{x}'R^{-1}(R - M_k(M_k'R^{-1}M_k)^{-1}M_k')R^{-1}\boldsymbol{x}\right]^{-(\frac{n-1}{2})}. \tag{2.7.40}$$

Under both models we must integrate numerically over ρ in equations (2.7.36) and (2.7.38) to find the marginal of X.

The posterior distribution of k follows as

$$\Pi(k=n|X,\delta) \propto q f(X|k=n)$$

and

$$\Pi(k|X,\delta) \propto \frac{1-q}{n-1} f(X|k,\delta). \tag{2.7.41}$$

Also notice that

$$\sigma^2|k,X,\delta,\rho \sim IG\left(\frac{n-1}{2}+\alpha;\quad \beta+\frac{1}{2}\boldsymbol{x}'R^{-1}H_\delta R^{-1}\boldsymbol{x}\right)$$

and

$$\sigma^2|k=n,X,\rho \sim IG\left(\frac{n}{2}+\alpha;\;\; \beta+\frac{1}{2}\boldsymbol{x}'R^{-1}H_0 R^{-1}\boldsymbol{x}\right)$$

and that

$$\boldsymbol{\mu}|k,\delta,\rho,X \sim t_2(n+2\alpha-1, N_\delta^{-1}M_k'R^{-1}\boldsymbol{x},\ \ T_1)$$

where

$$T_1 = (n+2\alpha-1)[2\beta+\boldsymbol{x}'R^{-1}H_\delta R^{-1}\boldsymbol{x}]^{-1}N_\delta$$

and

$$\mu_0|k=n,\rho,X \sim t(n+2\alpha-1, \tfrac{1}{n}\mathbf{1}_n'R^{-1}\boldsymbol{x},\ \ T_2),$$

where

$$T_2 = n(n+2\alpha-1)(2\beta+\boldsymbol{x}'R^{-1}H_0R^{-1}\boldsymbol{x})^{-1}. \tag{2.7.42}$$

The unconditional (of k) posterior of ρ is then

$$\Pi(\rho|X,\delta)\ \alpha\ \sum_{i=1}^{n}\Pi(k=i|X,\delta)\Pi(\rho|X,k=i,\delta). \tag{2.7.43}$$

2.8 SUMMARY OF APPROACHES IN THE LITERATURE

As mentioned, Broemeling (1974) considered the univariate normal sequence with a change in the mean (only one change-point at k) and with common variance σ^2. He supposed that the first $k(k=1,\ldots,n-1)$ have mean μ_1 and the remaining $n-k$ have mean μ_2, where (different from our study) $-\infty<\mu_1<\mu_2<\infty$. A uniform prior was put on k, i.e. $\Pi(k)=\frac{1}{n-1}$. He considered four cases: Case 1 with known μ_1 and μ_2 and $\sigma^2=1$ and derived

$$\Pi(k|X)\propto e^{-\frac{1}{2}\left[\sum_{i=1}^{k}(X_i-\mu_1)^2+\sum_{i=k+1}^{n}(X_i-\mu_2)^2\right]},\qquad k=1,\ldots,n-1.$$

Case 2 with μ_1 known, μ_2 unknown and $\sigma^2=1$ and with improper vague prior $\Pi(\mu_2)\propto$ const, $\mu_2\in(\mu_1,\infty)$, resulted in the posterior density for k as (with $k_1=k, k_2=n-k$)

$$\Pi(k|X,\mu_1)\propto\frac{\{1-\Phi[(\mu_1-\bar{x}_2)\sqrt{k_2}]\}}{\sqrt{k_2}}e^{-\frac{k}{2}(\mu_1-\bar{x}_1)^2}e^{\frac{1}{2}(k_1\bar{x}_1^2+k_2\bar{x}_2^2]},$$

where $\Phi(X)$ is the standard normal distribution function of X. Case 3 has both μ_1 and μ_2 unknown and $\sigma^2 = 1$ and with the prior $\Pi(\mu_1, \mu_2) \propto$ const $(-\infty < \mu_1 < \mu_2 < \infty)$. For this case the posterior density for k is

$$\Pi(k|X) \propto (k_1 k_2)^{-\frac{1}{2}} \cdot E_\times \Phi[k(X - \bar{x}_1)] e^{\frac{1}{2}[k_1 \bar{x}_1^2 + k_2 \bar{x}_2^2]}$$

where $E_\times$ denotes expectation with respect to a normal distribution with mean $\bar{x}_2$ and variance k_2^{-1}.

Case 4 with both μ_1 and μ_2 known and σ^2 unknown and with the prior $\Pi(\sigma^2) \propto \frac{1}{\sigma^2}(\sigma^2 > 0)$, the posterior density for k is

$$\Pi(k|X, \mu_1, \mu_2) \propto \left[\sum_{i=1}^{k}(X_i - \mu_1)^2 + \sum_{i=k+1}^{n}(X_i - \mu_2)^2\right]^{-\frac{n}{2}}.$$

Smith (1975) also considered the univariate normal sequence with only one change-point at k and considered the cases with (1) all parameters known, (2) μ_1 known, μ_2 unknown and $\sigma_1^{-2} = \sigma_2^{-2} = \sigma^{-2}$ known, (3) μ_1 known, μ_2 unknown and $\sigma_1^{-2} = \sigma_2^{-2} = \sigma^2$ unknown, (4) $\mu_1, \mu_2, \sigma_1^{-2} = \sigma_2^{-2} = \sigma^{-2}$ unknown and (5) $\mu_1, \mu_2, \sigma_1^{-2}, \sigma_2^{-2}$ unknown. The results for the first three cases corresponds with those of Broemeling, while the results for case 4 is the same as ours in (2.7.2) and the results of case 5 with

$$\Pi(k|X) \propto \Gamma\left(\frac{1}{2}k_1 + \frac{1}{2}\right)\Gamma\left(\frac{1}{2}k_2 + \frac{1}{2}\right)[k_1 k_2]^{-\frac{1}{2}}(s_{1k}^2)^{-\frac{1}{2}(k_1+1)}(s_{2k}^2)^{-\frac{1}{2}(k_2+1)}$$

corresponds closely with those in (2.7.5). There is a difference in the constants and the exponent of s_{ik}^2 as Smith starts with improper priors and we start with proper priors, in which case the integration over the hyperparameters causes the loss of one degree of freedom.

Menzefricke (1981) studied the three cases with a change in the precision where (1) μ_1 and μ_2 unknown, (2) μ_1 and μ_2 known and (3) $\mu = \mu_1 = \mu_2$ unknown. In case 1 the priors are $\sigma_i^2 \sim IG\left(\frac{\alpha_i}{2}, \frac{\beta_i}{2}\right)$, $\mu_i \sim N(\theta_i, \phi_i)$ where $\phi_i = \frac{\sigma_i^2}{\delta_i}$ $(i = 1, 2)$ and $p(k)$ is any discrete distribution $(k = 1, \ldots, n)$. The posterior distribution of k, conditional on θ_i, is as follows:

$$p(k|\mathbf{X}) \propto p(k) B\left(\frac{\alpha_1 + k_1}{2}, \frac{\alpha_2 + k_2}{2}\right)(\delta_1 + k_1)^{-\frac{1}{2}}(\delta_2 + k_2)^{-\frac{1}{2}} B_{1k}^{-(\frac{\alpha_1 + k_1}{2})} B_{2k}^{-(\frac{\alpha_2 + k_2}{2})}$$

where

$$
\begin{aligned}
B(a,b) &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad a,b>0, \\
B_{1k} &= \frac{k_1\delta_1}{\delta_1+k_1}(\theta_1-\bar{x}_{1k})^2+\frac{s_{1k}^2}{k_1}+\beta_1, \\
B_{2k} &= \frac{k_2\delta_2}{\delta_2+k_2}(\theta_2-\bar{x}_{2k})^2+\frac{s_{2k}^2}{k_2}+\beta_2
\end{aligned}
$$

and

$$s_{ik}^2=\sum^{k_i}(x_i-\bar{x}_{ik})^2.$$

The posterior distribution of $\gamma=\dfrac{\sigma_2^2}{\sigma_1^2}$ conditional on k,θ_1 and θ_2 is given by

$$p(\gamma^{-1}|k,X)=\frac{\gamma\, B_{1k}^{\frac{\alpha_1+k_1}{2}}\left(B_{2k}\gamma^{-1}\right)^{\frac{\alpha_2+k_2}{2}}}{B\left[\frac{\alpha_1+k_1}{2},\frac{\alpha_2+k_2}{2}\right]\left(B_{1k}+B_{2k}\gamma^{-1}\right)^{\frac{\alpha_1+\alpha_2+n}{2}}},$$

which is a Beta-distribution of the second kind, a well-known result (e.g. Box and Tiao (1973)).

For vague priors Menzefricke (1981) let $\Pi(\sigma_1^2,\sigma_2^2)=\frac{1}{\sigma_1^2\sigma_2^2}$, $\alpha_1=\alpha_2=-1$ and $\beta_1=\beta_2=0$. His degrees of freedom differs from ours in equation (2.7.3) as we let $\beta_1,\beta_2,\alpha_1,\alpha_2\to 0$.

In case 3 (change in variance only), the priors on σ_1^2 and σ_2^2 are the same as previously. Further $\mu\sim N\left(\theta,\frac{\sigma_1^2}{\delta}\right)$. The posterior distribution of γ is

$$g(\gamma|k,X)=\gamma^{\frac{k_2-2}{2}}(k_1+\gamma k_2)^{-\frac{1}{2}}\left\{k_2s_{2k}^2\gamma^2+k_1k_2\gamma\left[\frac{s_{2k}^2}{k_2}+\frac{s_{1k}^2}{k_1}+(\bar{x}_{1k}-\bar{x}_{2k})^2\right]+k_1s_{1k}^2\right\}^{-\frac{n}{2}}$$

and the posterior distribution of k is

$$p(k|X)\propto p(k)\int_0^\infty g(\gamma|k,X)d\gamma.$$

Booth and Smith (1982) also considered changes of the mean in normal sequences for the univariate and multivariate case. For the latter case he supposed that under M_0 $y_i \sim N_p(\boldsymbol{\mu}_1, \Sigma)$ for $i = 1, \ldots, n$ and under M_k that $y_i \sim N_p(\boldsymbol{\mu}_1, \Sigma)$ for $i = 1, \ldots, k$ and $y_i \sim N_p(\boldsymbol{\mu}_2, \Sigma)$ for $i = k+1, \ldots, n$. For the unknown parameters $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma$ the improper priors $p(\boldsymbol{\mu}_j|\Sigma) = c_j(2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}$ and $p(\Sigma) = c|\Sigma|^{-(\frac{p+1}{2})}$ were used. With the change-point as close to the end-points as possible he developed the Bayes factors $B_{k_0} = c_2\left(\frac{n}{k_1 k_2}\right)^{\frac{p}{2}}\left(\frac{|S_n|}{|S_{1k}+S_{2k}|}\right)^{\frac{n}{2}}$, which is the same as our (2.4.10). He defined the minimal calibrating sample as $n = 2p+1$ when $k = p$. If this gives $\bar{y}_{k_1} = \bar{y}_{k_2}$ so that $S_{1k} + S_{2k} = S_n$, one should want approximately $B_{k_0} = 1$ with $S_{jk} = \sum_{i=1}^{k_i}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{k_j})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{k_j})'$. This leads to the choice $c_2 = \left(\frac{p(p+1)}{2p+1}\right)^{\frac{p}{2}}$.

2.9 APPLICATIONS

In the applications that follow in this and later chapters, the prior probability for no change is taken as $q = \frac{1}{2}$ when compared with a single possible change-point. The rest of the probability is uniformly distributed over the number of possible values of the change-point. When multiple change-points are considered, up to a maximum of R, the prior mass is again uniformly distributed among the possible number of change-points i.e. $\Pi(r) = \frac{1}{R+1}$, $r = 0, 1, \ldots, R$.

The data for all the examples are given in Appendix A.

EXAMPLE 2.9.1

The measurements on male Egyptian skulls of Thompson, A. and Randall-Maciver, R. (1905), as given in Hand, *et al.*, (1994) from 5 epochs are to be analysed with a view to deciding whether there are any differences between the measurements from the epochs and if they show any changes with time. A steady change of head shape with time would indicate interbreeding with immigrant populations. Measurements are: x_1 = maximum breadth, x_2 = basibregmatic height, x_3 = basialveolar length and x_4 = nasal height. The time periods are 4000 BC(1), 3300 BC(2), 1850 BC(3), 200 BC(4) and AD150(5), so that there is four possible change-points. There are 30 observations for each epoch.

If the presence of exactly one change in the mean is assumed, then the posterior probability of k (by using (2.2.18)) is given in Table 2.1,

Table 2.1

k	1	2	3	4
prob	0	0.8399	0.1601	0

showing that a change is most likely to have occurred at $k = 2$.

For the possibility of no change against one change, the posterior probability of k (by using (2.3.3), (2.3.4) and (2.2.20) with $\delta = 10$ (the approximate mean of the unconditional posterior of δ - see Figure 2.3) is given by Table 2.2,

Table 2.2

k	No change	1	2	3	4
prob	6.5206×10^{-7}	0	0.7961	0.2038	0.0001

giving a very low probability for no change and once again indicating a change at $k = 2$, which means a change between 3300 BC and 1850 BC.

The unconditional marginal posteriors of the components of the means μ_1 and μ_2, which follow from (2.2.19), are given by Figure 2.1.

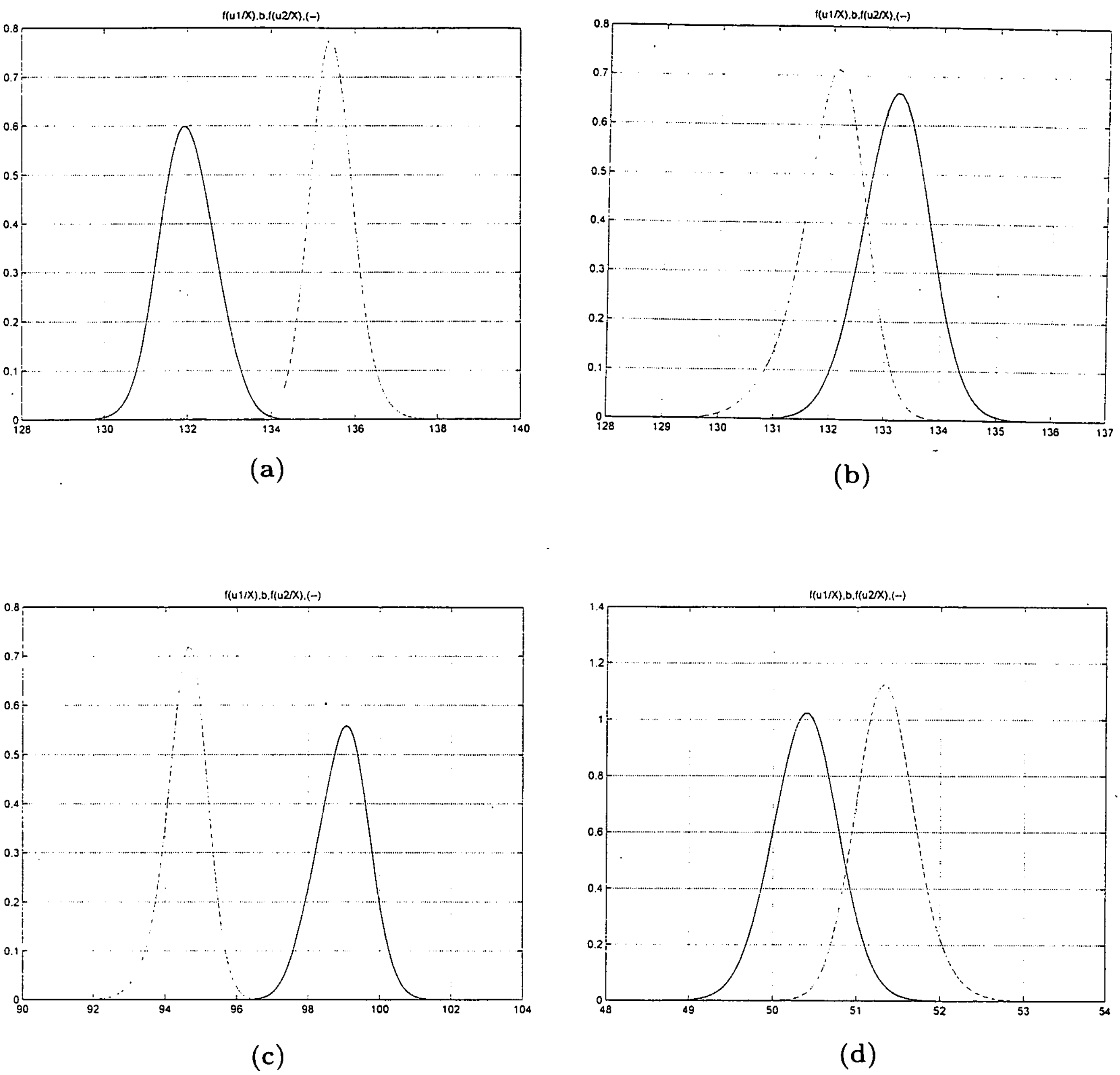


Figure 2.1: The unconditional marginal posteriors of the four components of the mean before (-) and after (- - -) the change-point for example 2.9.1

The distributions of the differences Δ_i between the components of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, which follows from (2.4.5), are given in Figure 2.2.

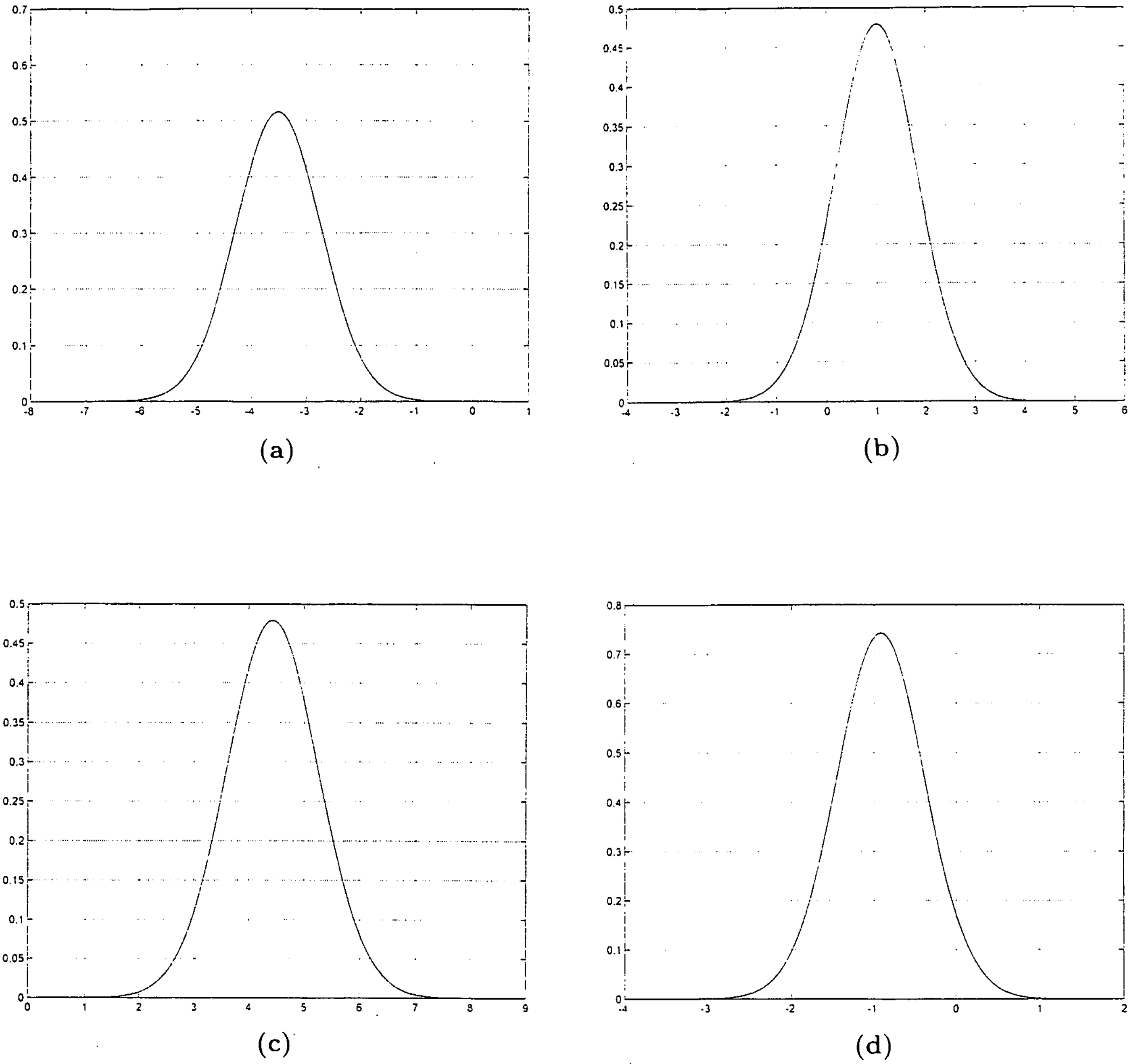


Figure 2.2: The distributions of the differences Δ_i between the components of μ_1 and μ_2 for example 2.9.1.

The unconditional and conditional (given $k = 2$) posteriors of δ which follows from (2.4.5), are given by Figure 2.3, while $\Pi(k|X)$ as a function of δ is given in Figure 2.4 for $k = 2$ and $k = 3$, which shows that the posterior probabilities are not very sensitive to δ.

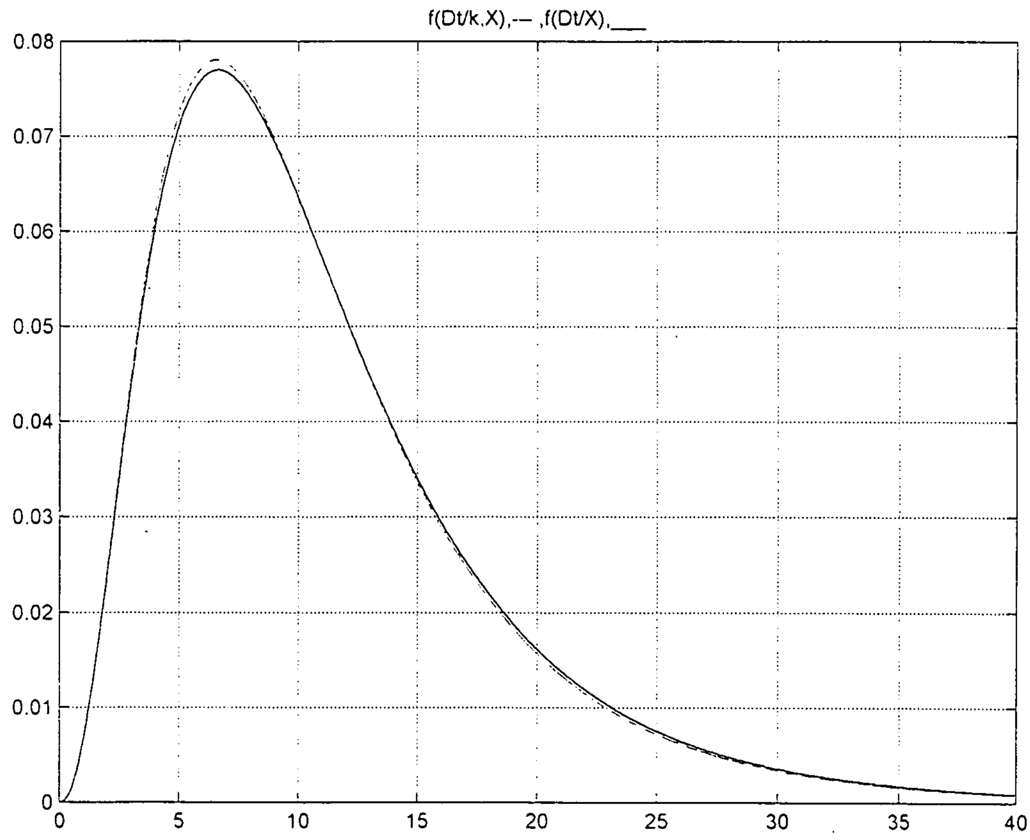
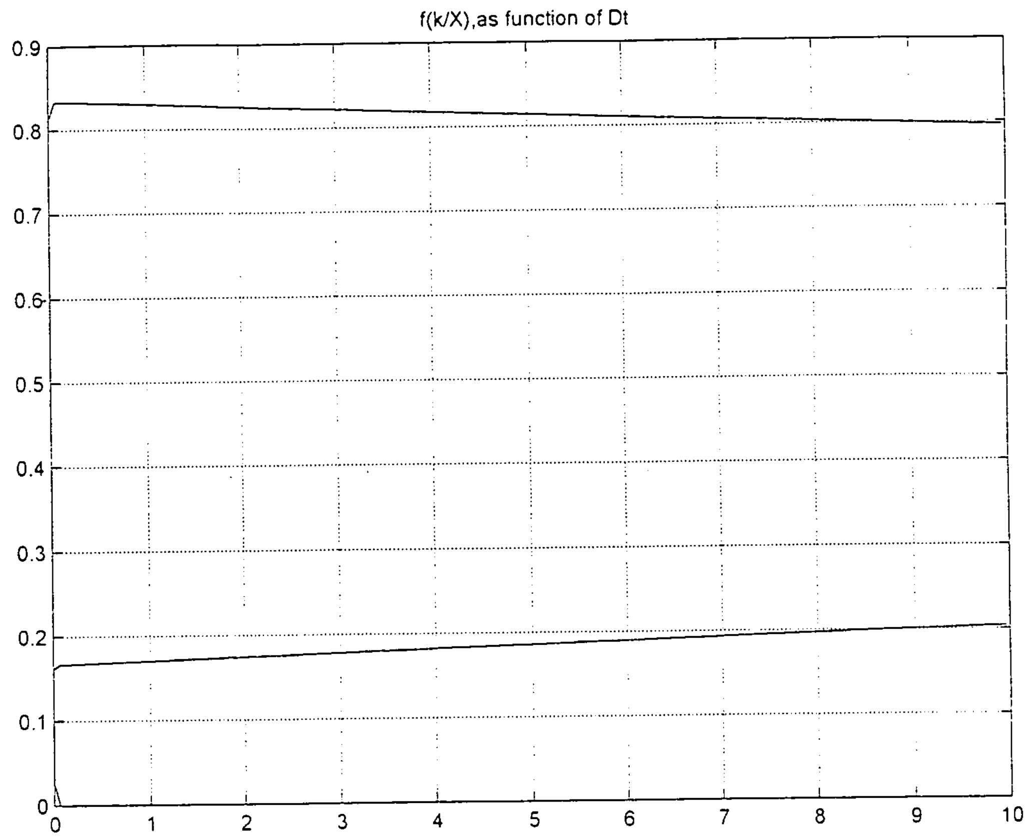


Figure 2.3: The unconditional (-) and conditional (given $k = 2$) (- -) posteriors of δ for example 2.9.1



δ

Figure 2.4: $\Pi(k|X)$ as a function of δ for example 2.9.1 for $k = 2$ (top) and $k = 3$ (bottom).

Still under the assumption of one change-point, it is of importance to see which component has the most influence in causing the change to occur at a fixed k. By using (2.4.3) and (2.4.4) in (2.4.9) and (2.4.10) we obtained the results given in Table 2.3., row I_1.

Table 2.3

Variables	X_1	X_2	X_3	X_4
I_1	35.5080	9.4688	41.6669	13.3563
I_2	35.4933	9.5407	41.5164	13.4495

For $\delta = 10$ and by using (2.4.7) and (2.4.8) in (2.4.9) and (2.4.10) we obtained the results given by Table 2.3, row I_2. The results indicate that X_1, maximum breadth, and X_3, basialveolar length, are the variables mostly responsible for causing a change-point at $k = 2$.

The posterior probabilities that follow from the FBF (equations (2.5.6) and (2.5.7)) are given in Table 2.4, where $k = 0$ denotes "no change".

Table 2.4

k	0	1	2	3	4
Prob.	5.3068×10^{-6}	0	0.8293	0.1707	0

which corresponds reasonably well with the results in Table 2.2.

The Intrinsic BF's, following from (2.5.3) and (2.5.4), are given by Table 2.5.

Table 2.5

B_{k0}	1	2	3	4
Arithmetic IBF	7.0990×10^5	6.7493×10^8	7.4676×10^4	4.8849×10^9
Median IBF	0.0390	6.4252×10^3	1.4257×10^3	0.2221
Geometric IBF	2.1235×10^{-5}	85.1779	18.4967	1.6977×10^{-4}

From Table 2.6, giving the posterior probabilities (from 1.3), we can see that the median and geometric IBF's correspond with the previous result, but the arithmetic IBF gives a somewhat different result. Berger and Pericchi (1998) point out that the arithmetic IBF can be unstable and that the median IBF is much more stable.

Table 2.6

k	0	1	2	3	4
P_A	0	0.0001	0.8785	0.1214	0
P_M	0.0005	0	0.8180	0.1815	0
P_G	0.0371	0	0.7911	0.1718	0

Under the assumption of two change-points, some probabilities for the pair (k_1, k_2) in (2.6.4) are given by Table 2.7.

Table 2.7

k_1, k_2	probability
1,3	0.0073
2,4	0.1549
2,3	0.8277

The posterior probabilities that follow from the FBF for no and 2 change-points is given by Table 2.8.

Table 2.8

0	(3,4)	(1,2)	(1,3)	(2,4)	(2,3)
0.0002	0.0024	0.0046	0.0088	0.1641	0.8199

By using (2.5.16) and (2.5.17) with $R = 2$ in (2.6.9) and (2.6.10), the posterior probabilities follow as $P(r = 0|\boldsymbol{x}) = 1.629 \times 10^{-4}$, $P(r = 1|\boldsymbol{x}) = 0.9685$ and $P(r = 2|\boldsymbol{x}) = 0.0315$. This indicates only one change-point after the second time period (between 3300 BC and 1850 BC).

It is interesting to note that the classical F-test for the equality of multivariate means indicates significant differences ($p < 0.01$) between all successive time periods, with the difference between the second and the third period the most significant ($p < 0.001$).

EXAMPLE 2.9.2

For the case of a multiple change in the mean, we considered the Colorado data used by Chernoff (1973) to illustrate his well-known "Chernoff faces". This data is $n = 55$ observations on 12 variables representing mineral contents from a 4500-foot core drilled from a

Colorado mountainside. We considered only 5 variables, which are the same 5 variables that Srivastava and Worsley (1986) and Gupta and Chen (1996) have used.

For $R = 7$ and by using (2.5.24) in (2.5.16) and (2.5.17) where B_{0k} is the FBF and where $P[r = j|X] = \sum_{k} \Pi(k, r = j|y]$, the posterior probabilities for a change (given by (2.5.24)) are given in Table 2.9, showing a high probability for 5 or 6 change-points.

Table 2.9

No	Post. prob.	Maximum prob. points						
0	0							
1	0	24						
2	0	20	34					
3	0.0105	20	26	34				
4	0.1202	20	26	34	48			
5	0.4166	20	24	26	32	48		
6	0.4346	18	23	27	34	43	48	
7	0.0315	20	23	27	34	41	43	48

The sets of 5 change-points with highest posterior probability are given in Table 2.10 and Figure 2.5 shows the marginal distributions of the 5 change-points, given there are 5 change-points.

Table 2.10

					Probability
20	24	26	32	34	0.0772
20	23	27	34	48	0.0821
20	24	27	34	48	0.0840
20	24	26	34	48	0.0894
20	24	26	32	48	0.0956

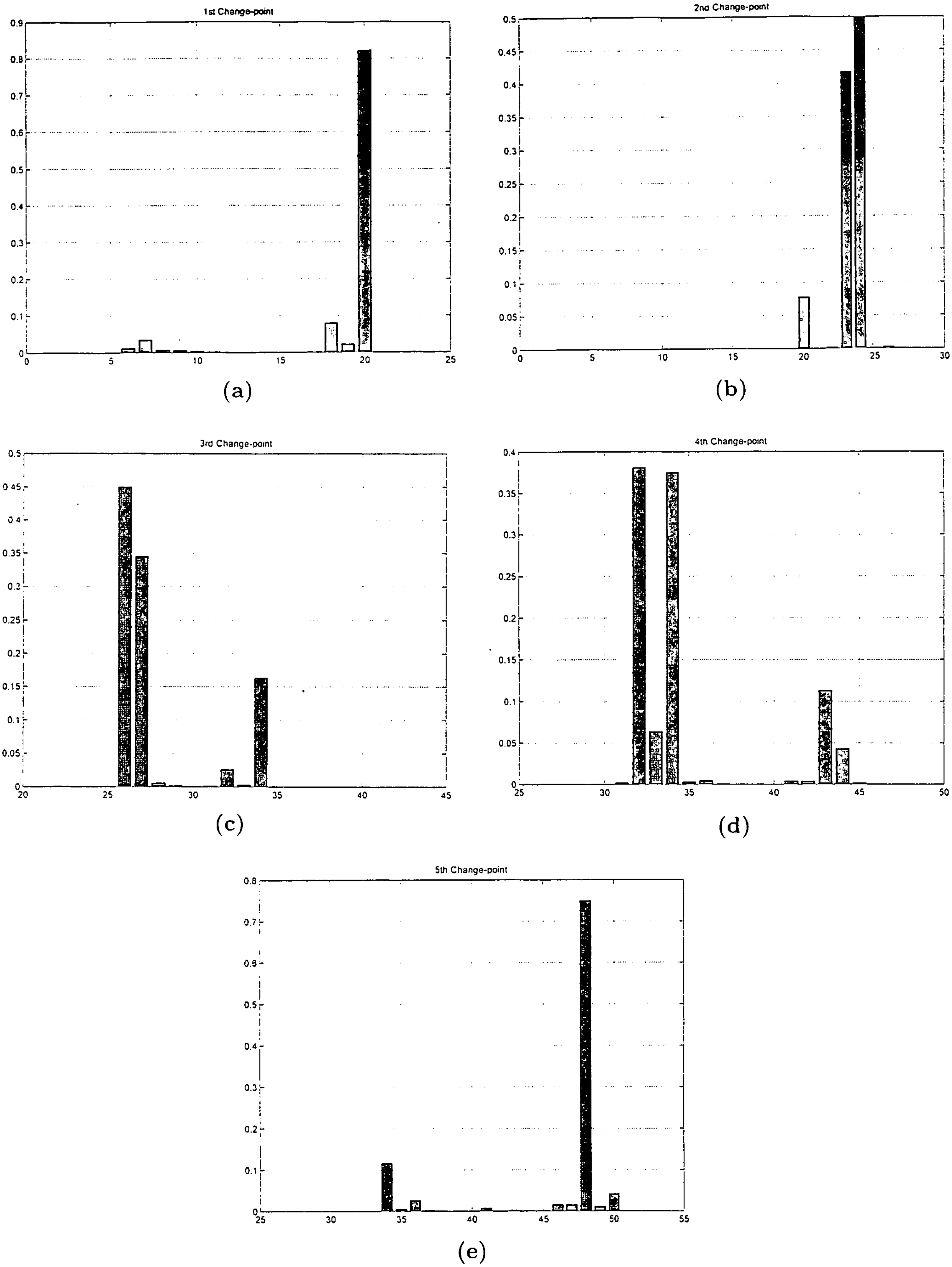


Figure 2.5: The marginal posterior distributions of the 5 change-points for example 2.9.2

Gupta and Chen (1996) gave the following summary given in table 2.11 of their own results and that of Chernoff (1973) and Srivastava and Worsley (1986).

Chernoff found 4 change-points, Srivastava and Worsley found five, while Gupta and Chen found twelve. The only change-points common to all analyses are 20 and 32.

Table 2.11: Change-points for the mineral contents of a core sample

Chernoff	Srivastava and Worsley	Gupta and Chen	Schoeman
		6	
	12	12	
	18	18	
20			20
24	24	24	24
			26
		28	
32	32	32	32
	34	34	
35			
		39	
		41	
		43	
		46	
		48	48

EXAMPLE 2.9.3

To illustrate our results of a change in the mean and variance, we will use the Friday closing prices collected by Chen and Gupta (1991) from January, 1990 through December, 1991 for two stocks (Exxon and General Dynamics). Chen and Gupta (1991) tested the hypothesis of a change in the variance. The weekly rates of return for these two stocks will be analyzed for a single change-point, where the weekly rates of returns $X_t = [X_{t_1} \quad X_{t_2}]$ and X_{t_i} = Current

Friday closing price — Previous Friday closing price. There are 103 observations and the original data is given in Appendix A.

For case 2 in paragraph 2.2.2.2 where we got the prior assumptions $\Sigma_2^{-1} = \gamma\Sigma_1^{-1}$ and $\Phi^{-1} = \delta\Sigma_1^{-1}$, we took $\gamma = 0.4$ and $\delta = 10$ in equations (2.3.3), (2.3.4) and (2.2.49). The probability for no change was 3.5807×10^{-5} and we got a maximum probability of 0.4144 at the 28th observation. For case 1, by using (2.3.5), we also get a maximum at the 28th observation if we assume $\delta = \delta_1 = \delta_2$. Notice that the value of the maximum depends on δ.

In Figure 2.6, where $\Pi(k|X)$ is a function of δ for $\gamma = 0.4$, we can see that the posterior probability is quite robust with respect to δ and the change-point seems to be in the region of $k = 26$ to 28.

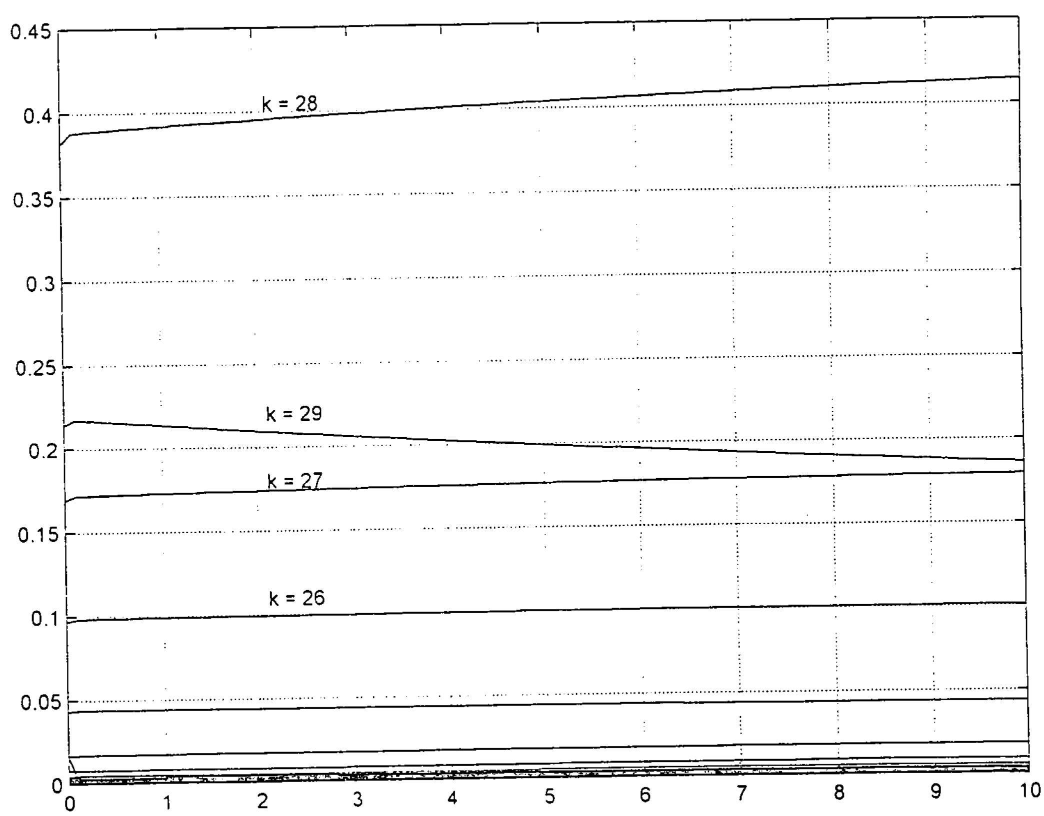


Figure 2.6: $\Pi(k|X)$ as a function of δ for example 2.9.3

In Figure 2.7, where $\pi(k|X)$ is a function of γ for $\delta = 10$, it is clear that for a small γ the 28th observation gives a maximum probability and if we choose γ to be larger than one, that the 66th observation gives the maximum probability. Note that Chen and Gupta (1991) mainly got the 66th observation as a change-point for the return series, but also got more change-points, under which the 27th and 28th observations. Also note that Chen and Gupta (1991) used the difference divided by the previous closing price, while we just used the difference. It is obvious from Figure 2.7 that the posterior probability is sensitive to the value of γ. It is important to remember that these probabilities are determined under the rather strict assumption of $\Sigma_2^{-1} = \gamma\Sigma_1^{-1}$.

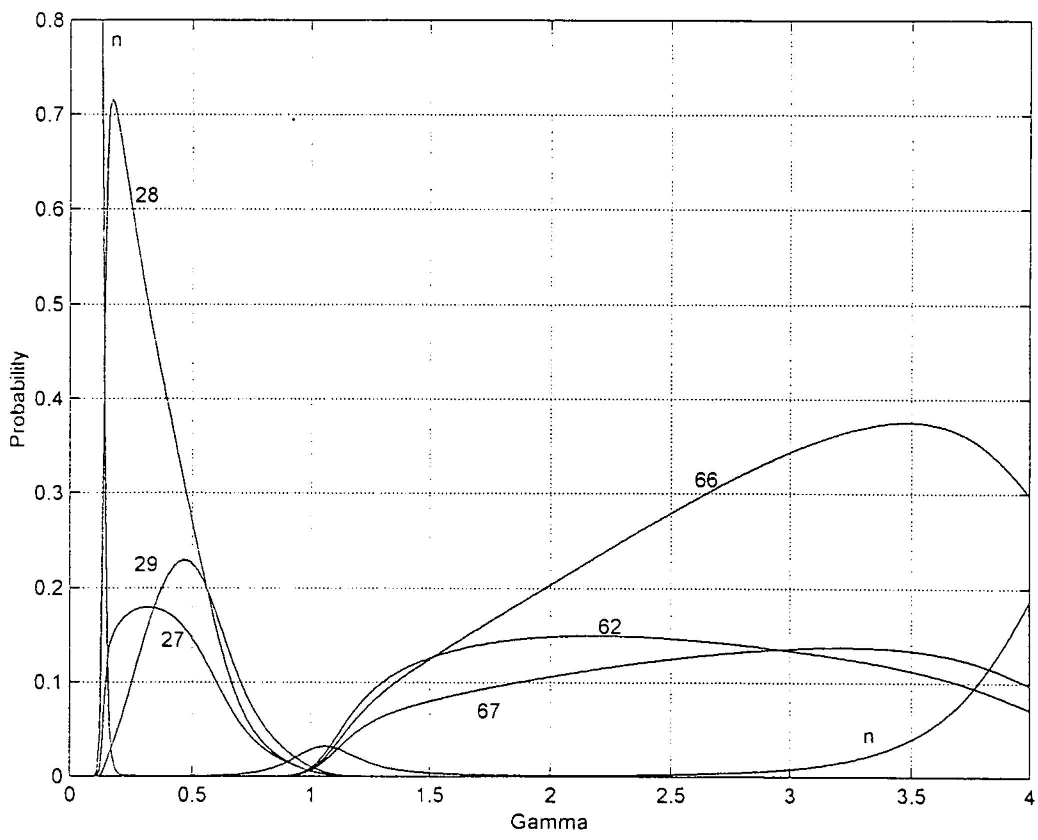


Figure 2.7: $\Pi(k)$ **as a function of** γ **for example 2.9.3**

Next, if we assume exactly one change-point then, according to (2.2.36), the posterior probability distribution of k is given as in Figure 2.8 with a maximum at $k = 52$. This is completely different from the previous results. The marginal posterior of the elements of Σ_1 and Σ_2 is shown in Figure 2.9 and Figure 2.10, where $\sigma^2_{11(1)}$ and $\sigma^2_{11(2)}$ denote the variance of the first element of X before and after the change-point. Similarly for the variance of the second element of X. The posterior means are given as $E(\sigma^2_{11(1)}|X) = 1.6310$, $E(\sigma^2_{11(2)}|X) = 1.4489$, $E(\sigma^2_{22(1)}|X) = 1.9532$ and $E(\sigma^2_{22(2)}|X) = 8.9495$. These results indicate that the assumption of proportional co-variance matrices used in the previous paragraph may not be valid. That could be the reason for the discrepancy between the results. In this paragraph no assumptions about the structures of the co-variance matrices were made.

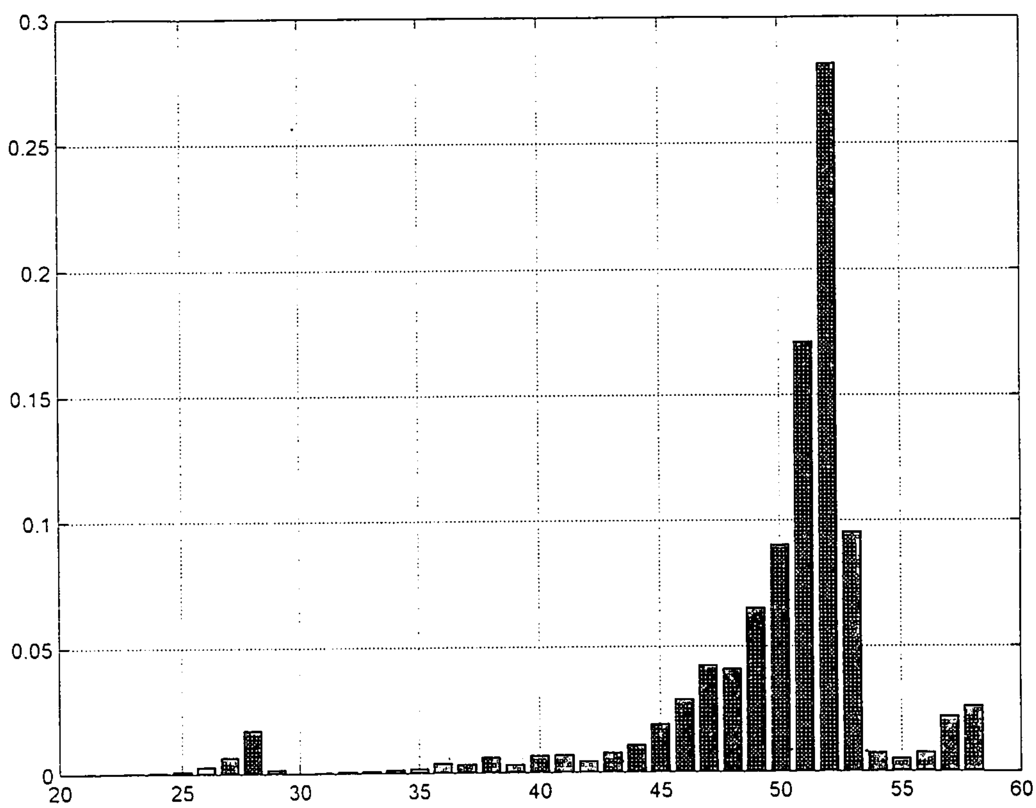


Figure 2.8: The posterior probability distribution of k for example 2.9.3

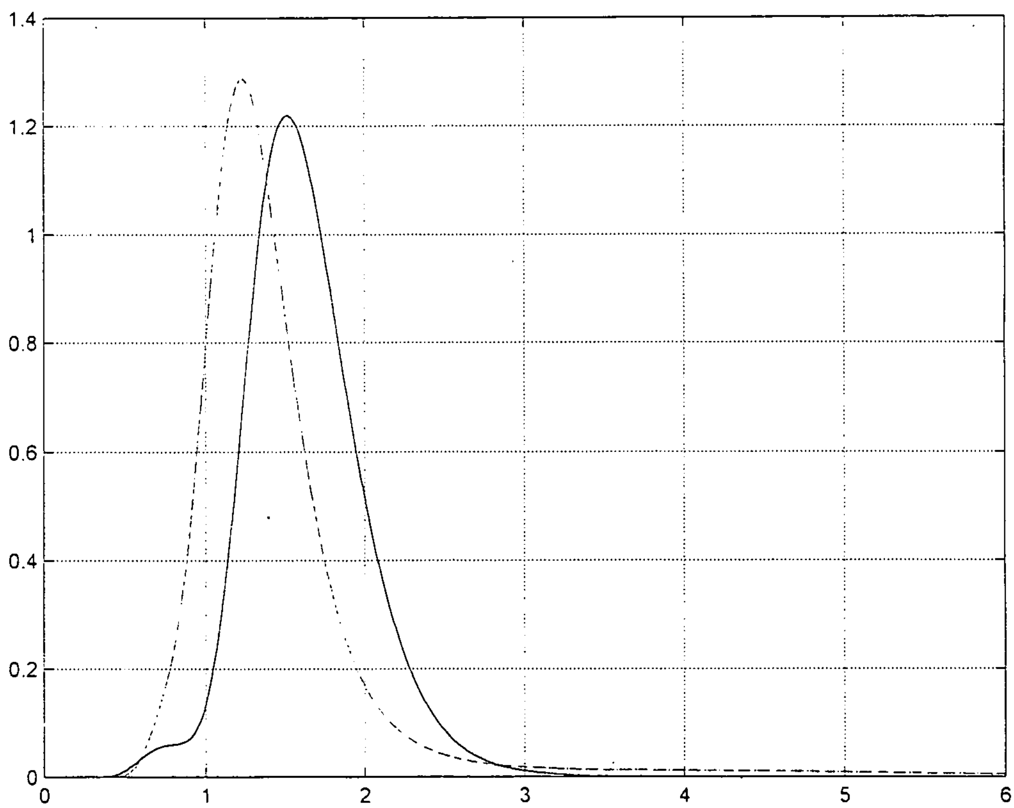


Figure 2.9: The marginal posterior of σ_{11}^2 before ($-$) and after ($--$) the change-point for example 2.9.3

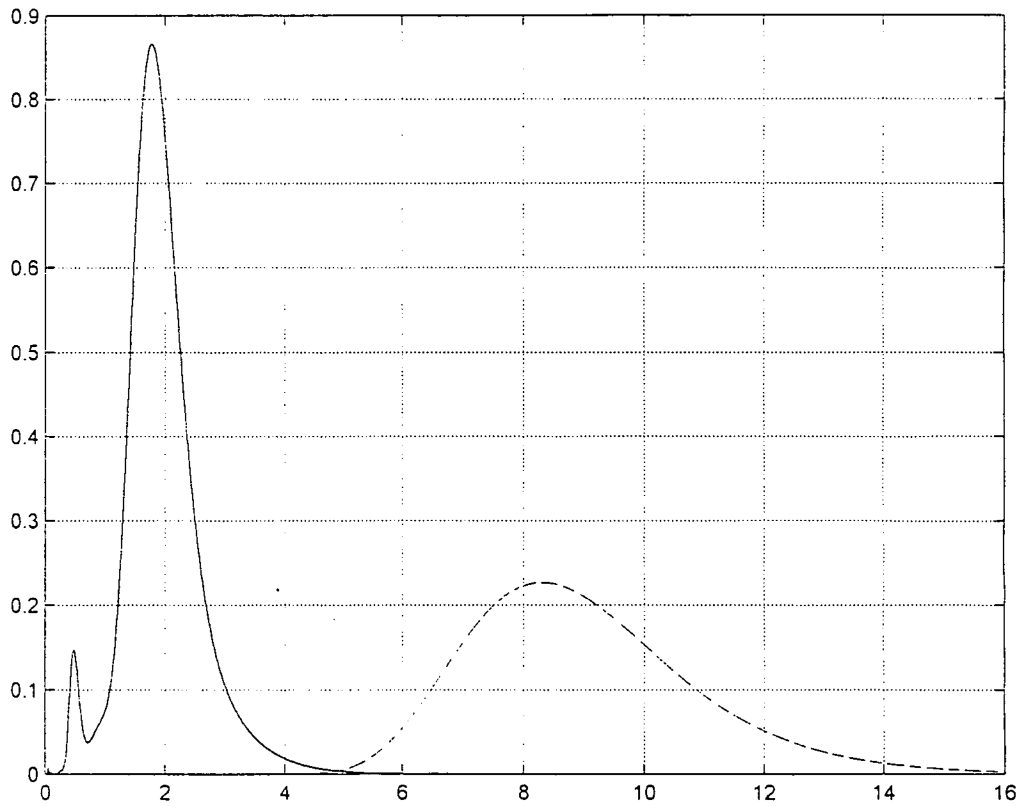


Figure 2.10: The marginal posterior of σ_{22}^2 before $(-)$ and after $(--)$ the change-point for example 2.9.3

For the FBF in (2.5.10) with $b = 0.11$ (see discussion below (2.5.10)) it follows that the probability for no change is 0.0033 and that the maximum probability is 0.2568 for the 52nd observation, as can be seen in Figure 2.11. This corresponds with the previous results as in Figure 2.8. Note from (2.5.10) that k and $n - k$ must be larger than $\frac{1}{b}$.

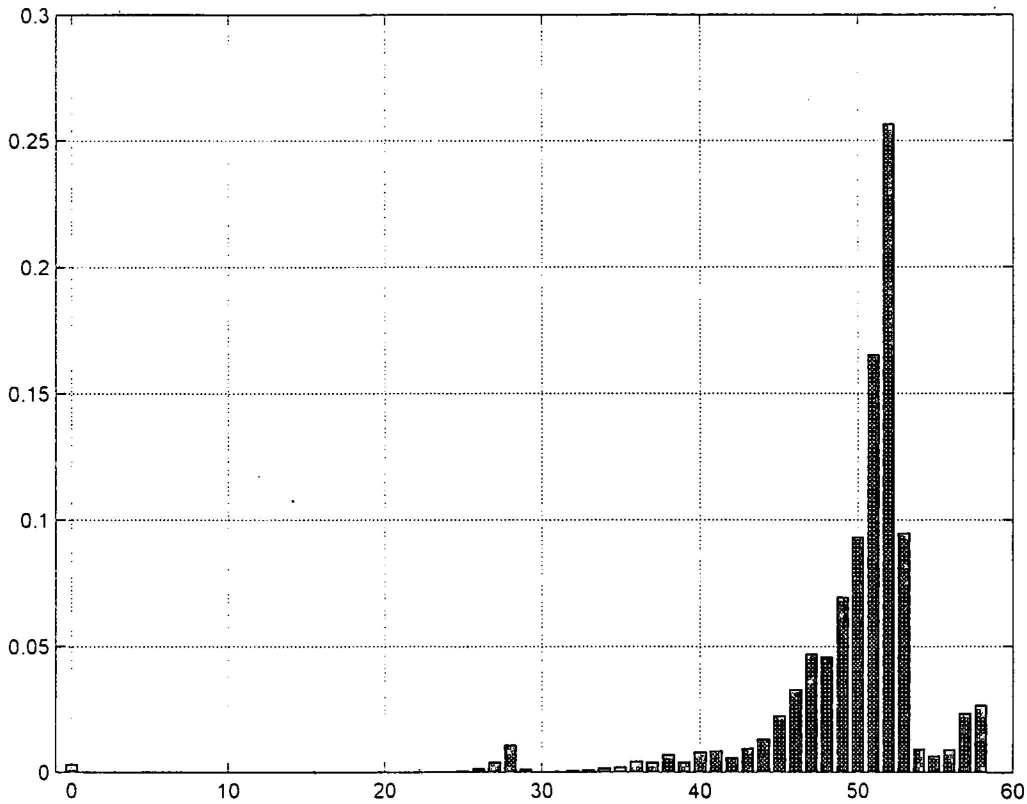


Figure 2.11: Posterior probability distribution of change-point — FBF for example 2.9.3

The posterior probability of $\Pi(k = n|X)$ as a function of b is given in Figure 2.12. Note that the probability of no change is smaller than 0.015 for any $b < 0.5$.

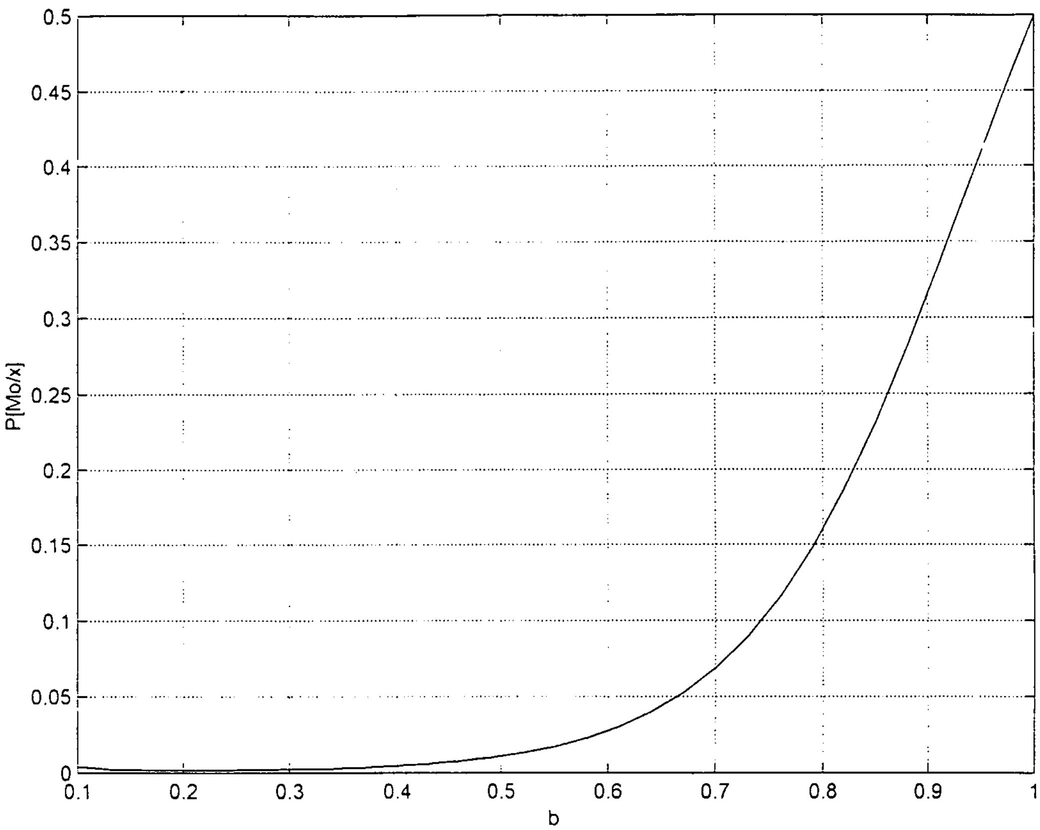


Figure 2.12: The posterior probability of $\Pi(k = n|X)$ as a function of b for example 2.9.3

Our conclusion is that a change-point occurs in the mean and variance of the weekly returns in the region of $k = 52$, which is January 1991.

EXAMPLE 2.9.4

For a change in the variance we will use the weekly closing values of the Dow-Jones Industrial Average from July 1, 1971 through August 2, 1974, studied by Hsu (1979). The data were extracted from Daily Stock Price Record: New York Stock Exchange, published quarterly by Standard and Poor's Co., New York, N.Y. The weekly closing values will be analyzed for a single change-point, where the weekly closing values $X_t = [X_{t_1} \quad X_{t_2}]$ and X_{t_i} = (Current weekly closing value — Previous weekly closing value)/Previous weekly closing value. There are 161 observations and the original data is given in Appendix A. A plot of the data is given in Figure 2.13.

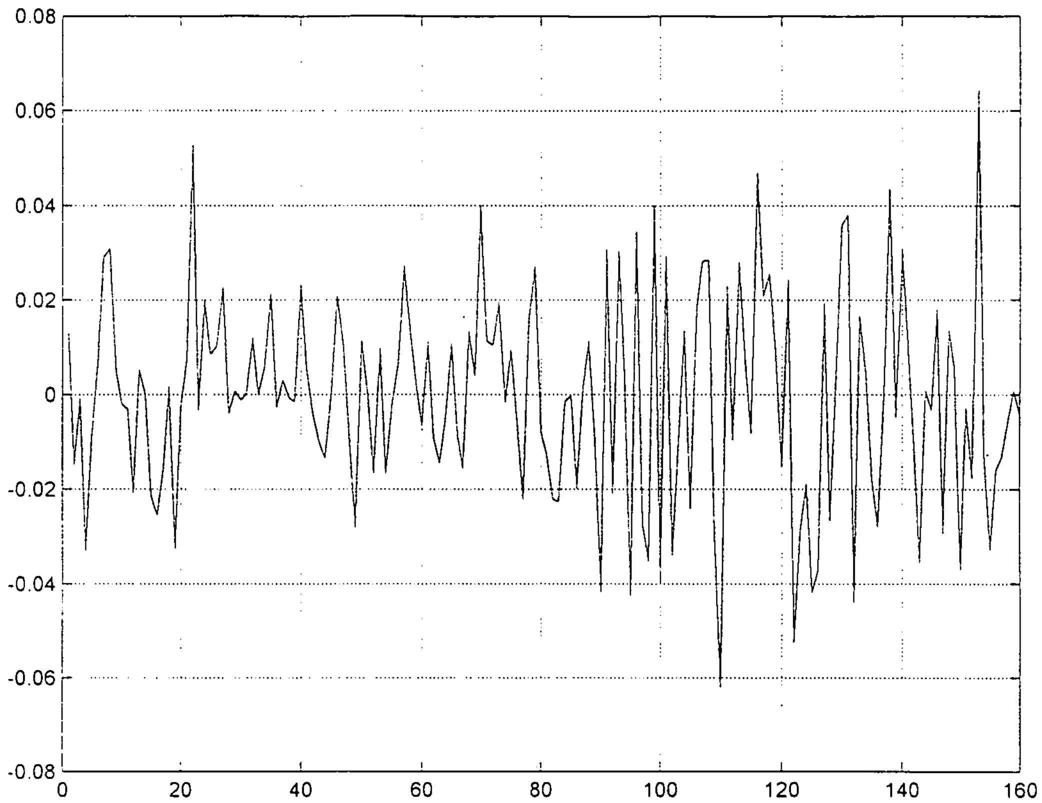


Figure 2.13: A plot of the weekly closing values of the Dow-Jones Industrial Average from July 1, 1971 through August 2, 1974

Chen and Gupta (1997) also used this data, performing a change-point analysis using the SIC procedure. According to them the stock price started to change at the 91st time point, which corresponds to the calender week of March 19-23, 1973. Their conclusion matched Hsu's (1977, 1979).

Worsley (1986) also considered this data and found a single change-point in late February.

For the FBF which follows from (2.5.14) with $b = 0.11$, it follows that $\Pi(k = n|X) = 0.0007$ and that the maximum probability is 0.2088 for the 89th observation, as can be seen in Figure 2.14.

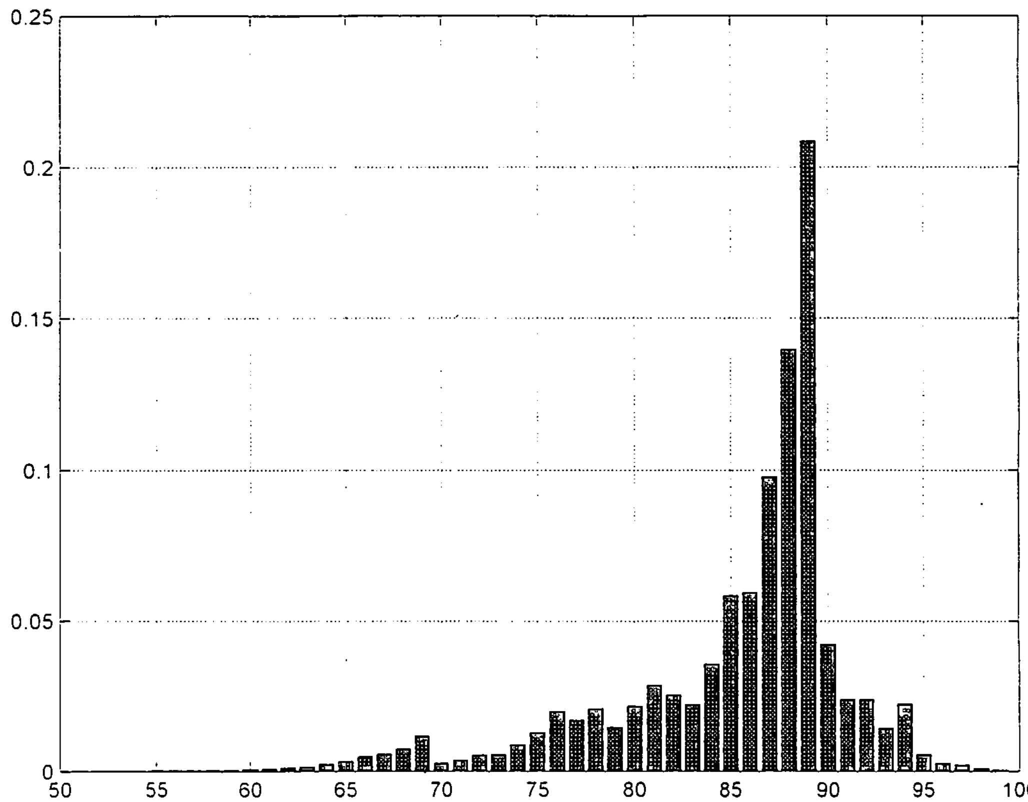


Figure 2.14: Posterior probability distribution of change-point — FBF for example 2.9.4

In Figure 2.15, where $\Pi(k|X)$ is plotted as a function of γ, it is clear that for $\gamma \simeq 0.36$ the result is similar to the above obtained result. Notice that the probability is very sensitive when γ is very small or close to 1, and $\pi(k = n|X) \longrightarrow 1$ as $\gamma \longrightarrow 0$ or 1.

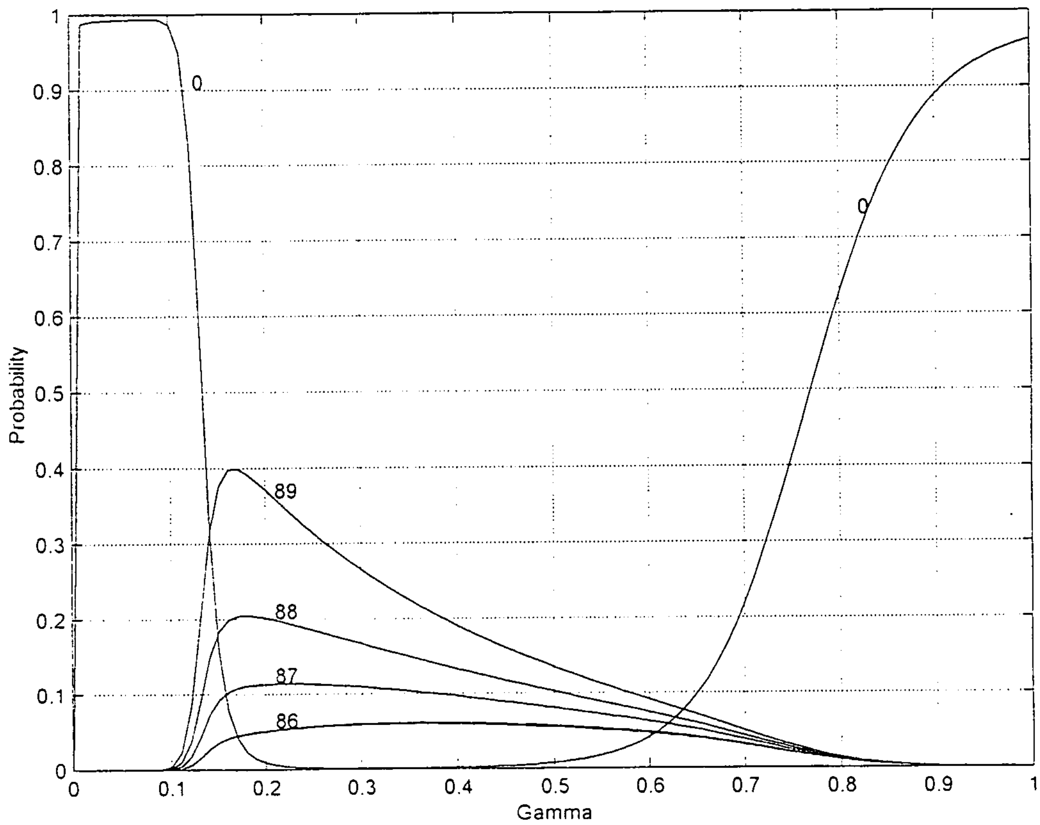


Fig 2.15: $\Pi(k|X)$ **as a function of** γ **for example 2.9.4**

For Gibbs sampling, where the full conditional distributions are given by (2.2.56), the maximum probability is 0.2324 at the 89th observation, as can be seen in Figure 2.16. This is very similar to our results for the FBF (Figure 2.14).

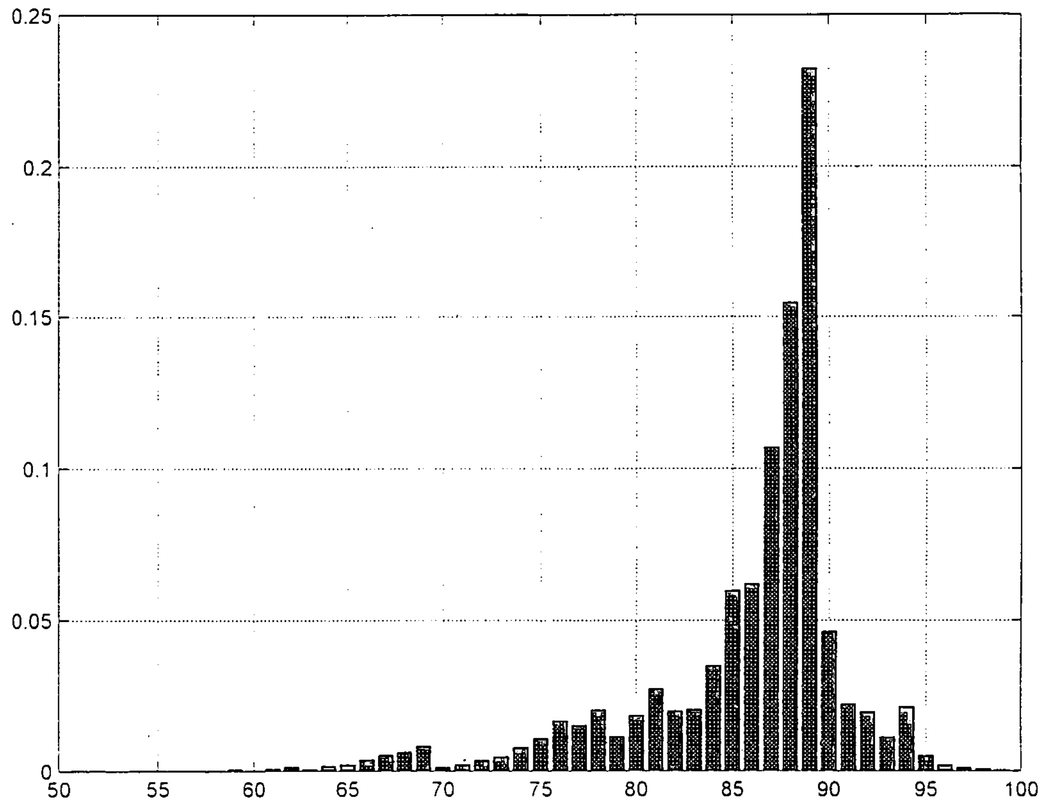


Figure 2.16: Posterior probability distribution of change-point — Gibbs for example 2.9.4

The unconditional marginals of the variances before and after the change-point are given by Figure 2.17. The expected values of the variances are 2.574×10^{-4} and 7.767×10^{-4} respectively, with 95% credibility intervals of $(1.86 - 3.52) \times 10^{-4}$ and $(5.55 - 11.00) \times 10^{-4}$ respectively. According to these posteriors an estimated value for γ is 0.33.

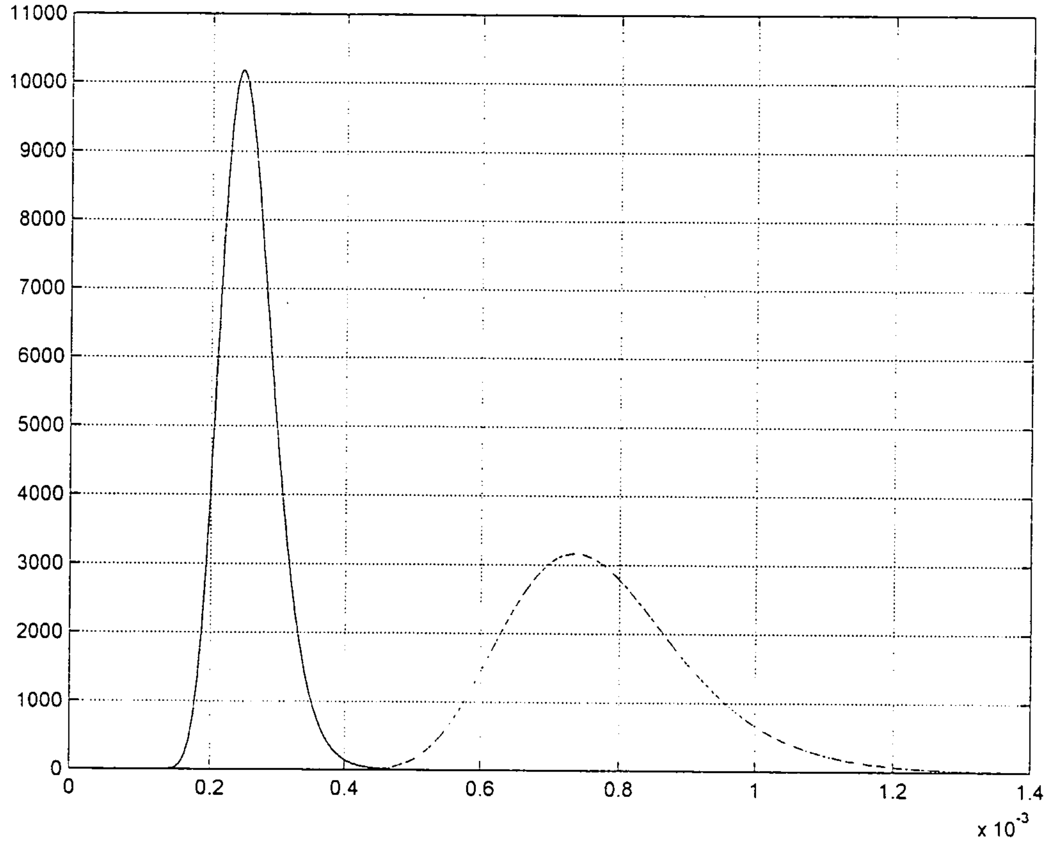


Figure 2.17: **The unconditional marginals of the variances before (–) and after (– –) the change-point for example 2.9.4**

CHAPTER 3

CHANGES IN THE LINEAR MODEL

3.1 INTRODUCTION

Change-point problems has been addressed extensively in the literature and as some authors referred to change-point while others refer to shift point or switchpoint, it's necessary to distinguish between the concepts. When it is assumed that the change occurs at a discrete point between two consecutive observations, with no functional relationship between the parameters before and after the change, we'll refer to it as a change-point. In the case of a switchpoint (which can be continuous), say t, we have the condition that $f_1(y|t) = f_2(y|t)$, where f_1 and f_2 are the models before and after the switchpoint.

It is also important to distinguish between a change in the univariate regression versus a multivariate regression. The univariate change-point can be a function of time or an independent variable, while in multiple regression the change-point only makes sense if the observations are just taken over time. It doesn't make sense in multivariate regression to look for change-points as a function of independent variables.

Considering the linear model $y_i = \alpha + \boldsymbol{x}_i\beta + \epsilon_i$ with respect to possible change-points, it is obvious that there are a number of different ways in which changes can occur. The first question is if any changes have occurred at all and secondly, if there are changes, how many? Then changes can occur in the parameter vector $\boldsymbol{\beta}$, or in α or in both. Another possibility is a change in the error variance, with or without a change in the parameters.

Two kinds of change-point problems have been dealt with in the literature. The first one is that of testing for the null hypothesis of no change versus the existence of a change occurring at some unknown time in a sequence of i.i.d. normal random variables (Page (1955), Chernoff and Zacks (1964), Gardner (1969), Hawkins (1977), Worsley (1979)), or in a simple linear model (Quandt (1958), Farley and Hinich (1970), Maronna and

Yohai (1978)), or in a general linear model (Worsley (1983), Jandhayala and MacNeill (1991)), all from a non-Bayesian viewpoint.

Considerable attention has been devoted, e.g. by Hinkley (1969), Hudson (1966), Quandt (1958, 1960) and Robison (1964), to the problem of fitting lines or curves to data which suggest abrupt changes in parameter values from one range of the independent variables to another. Tests of the hypothesis that a sudden change in behaviour has occurred at an estimated join point have been suggested by Quandt (1958) and Hinkley (1969) and maximum likelihood estimation procedures for the model parameters and the join point have been described by Hinkley (1969), Quandt (1958) and Robison (1964). Assuming a change does occur, Quandt (1958) estimated the switch-point m and the regression parameters by a maximum likelihood technique, and Hinkley (1969, 1971), under the assumption that the two-phase regression model is continuous, estimated and made inferences about the abscissa of the intersection. In the case of two constant means the emphasis of published work, in particular that of Page (1954, 1955, 1957) on cumulative sum schemes, has been on testing the null hypothesis $H_0 : \theta_0 = \theta_1$ against the two-mean alternative.

The (second) problem of estimating the point at which the change occurs, has among others been addressed by Schulze (1982) and Zacks (1982).

Still from a non-Bayesian viewpoint two-phase linear models, which is a generalization of the shift problem for a normal sequence, has been studied by Quandt (1958) and Hinkley (1969, 1971), while some other studies of two-phase regression problems have been considered by Quant (1960), Sprent (1961), Hudson (1966), Feder (1975), Farley, *et al.* (1975), Brown, *et al.* (1975), Holbert (1982), Hsu (1982) and McAleer and Fisher (1982). Harrison and Stevens (1976), Swamy and Mehta (1975) and Farley and Hinich (1970) have all studied these models. Poirier (1976) gives a review of the literature concerning the prediction of a future observation when the model has changed.

Many authors have studied the change-point problem associated with regression models. Brown, Durbin and Evans (1975) introduced a method of recursive residuals to test for change-points in multiple regression models. Hawkins (1989) used a union-intersection approach to test changes in a linear regression model. Kim (1994) considered a test for a change-point in linear regression by using the likelihood ratio statistic and studied the asymptotic behavior of the LRT statistic. Chen (1998) studied the change-point problem for simple linear regression model, as well as for the multiple linear regression model mainly by using Schwarz Information criterion, SIC (Schwarz, 1978).

Relatively little has appeared in literature about change in the multivariate linear model. However, Sen and Srivastava (1973, 1975, a,b,c) proposed tests for detecting change in means and examined the exact and asymptotic properties of the test statistics. Salazar (1980, 1982) considered changes in the multivariate linear model using a change-point parameter. Moen (1982) developed a detailed analysis of the multivariate linear model and Tsurumi, *et al.* (1984) developed a gradual switching multivariate regression model with stochastic constraints. Booth and Smith (1982) considered a Bayesian approach to retrospective identification of change-points and studied changes of the mean in the univariate and multivariate normal sequences as well as changes of coefficients in regression models.

From a Bayesian viewpoint, Chin Choy and Broemeling (1980), Holbert and Broemeling (1977), Ferreira (1975) and Bacon and Watts (1971) studied the two-phase regression model, a simplification of the linear model when $p = 2$. Holbert and Broemeling (1977) and Smith and Cook (1980) estimated the point at which the change occurs. Chernoff and Zacks (1964) and Bhattacharya and Johnson (1968) have discussed the same problem as Page (154, 1955, 1957) within a Bayesian framework.

Bacon and Watts (1971), Ferreira (1975), Holbert and Broemeling (1977), Chin Choy and Broemeling (1980), Moen, *et al.* (1985), Smith and Cook (1990) and Kim (1991) also looked at change-points in linear models. An overview and numerous references can be found in Broemeling and Tsurumi (1987).

When there is no change in the precision parameter at the switchpoint, the linear model reduces to the one studied by, among others, Ferreira (1975), Holbert and Broemeling (1977), Chin Choy and Broemeling (1980) and Land and Broemeling (1983). Chin Choy and Broemeling (1980) derived the Bayesian posterior distributions of the switch-point, the regression parameters and the precision parameter, generalizing the studies of Ferreira (1975) and Holbert and Broemeling (1977). Land and Broemeling (1983) considered the prediction problem and derived the Bayesian predictive distribution of k future observations.

Stephens (1994) discussed the use of a sampling-based technique, the Gibbs sampler, in multiple change-point problems and demonstrate how it can be used to reduce the computational load involved considerably. Carlin, Gelfand and Smith (1992) presented this new Bayesian analysis. They used Gibbs sampling to resample (repeatedly sample) from the joint posterior distribution of all the parameters in a change-point model. Their study was an advancement in the sense that earlier use of numerical and analytical approximations could be avoided with their Monte Carlo Markov Chain resampling technique. Broemeling and Gregurich (1996) approached the same problem, but using a direct sampling approach in conjunction with analytical reductions, whereby standard random number generators can be used to directly generate samples from the posterior distribution. In this way, convergence issues with Gibbs sampling can be avoided and the posterior analysis simplified.

Broemeling and Gregurich (1996) confine their study to the fixed sample size version. Using a direct resampling process, a Bayesian approach is developed for the analysis of the shift point problem. In many problems it is straight forward to isolate the marginal posterior distribution of the shift point parameter and the conditional distribution of some of the parameters given the shift point and the other remaining parameters. When this is possible, a direct sampling approach is easily implemented whereby standard random number generators can be used to generate samples from the joint posterior distribution of all the parameters in the model. This technique is illustrated with examples involving one shift from Poisson processes and regression models.

Wang and Lee (1993) considered a Bayesian approach to detect a change-point in the intercept of simple linear regression. He employed the Jeffrey's non-informative prior and compared it with the uniform prior in Bayesian analysis.

Up to this stage most of the mentioned analyses are under the assumption of exactly one change-point as complications arise due to the changing dimensions of the parameter space if the number of change-points is unknown. Barry and Hartigan (1992) propose a product partition model for multiple change-points. Groenewald (1993) considered a general Bayes procedure for the examination of possible change-points in the linear model and made provision for no, one or more than one change-point under the assumption of homogeneity of error variance. In linear regression, certain components which may be the cause of a change-point, can be examined. The results are in terms of posterior probabilities over a class of conjugate priors.

3.2 NO OR ONE CHANGE-POINT IN THE LINEAR MODEL

3.2.1 A CHANGE IN THE REGRESSION COEFFICIENT WITH CONSTANT VARIANCE

Consider a linear model where a change may have occurred,

$$\begin{aligned} y_i &= \boldsymbol{x}_i\boldsymbol{\beta}_0 + e_i; \quad i = 1, \ldots, n \\ \text{or} \qquad y_i &= \begin{cases} \boldsymbol{x}_i\boldsymbol{\beta}_1 + e_i; & i = 1, \ldots, k \\ \boldsymbol{x}_i\boldsymbol{\beta}_2 + e_i; & i = k+1, \ldots, n \end{cases} \end{aligned} \tag{3.2.1}$$

where $p \leq k \leq n - p$.

The $\boldsymbol{x}_i = [1 \quad x_{1i} \cdots x_{p-1,i}]$ in the given model are $1 \times p$ known vectors of regressor variables, the $\boldsymbol{\beta}_i = [\beta_0 \quad \beta_1 \cdots \beta_{p-1}]'$ are $p \times 1$ unknown parameter vectors and the e_i are i.i.d. normal random variables with mean zero and variance $\sigma^2 > 0$.

If $k = n$, no change has occurred, while exactly one change has occurred somewhere if $1 \leq k \leq n - 1$. So if $1 \leq k \leq n - 1$, $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ where k is unknown.

Let the marginal prior mass function of k be

$$\Pi(k) = \begin{cases} q; & k = n \\ \frac{1-q}{n-2p+1}; & p \leq k \leq n-p \end{cases} \tag{3.2.2}$$

and let $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ be assigned normal-gamma densities with $\boldsymbol{\beta}_0 \in I\!R^p$, $\boldsymbol{\beta}_0|\sigma^2 \sim N(\boldsymbol{\theta}_0, \sigma^2\Phi_{11})$ and $\boldsymbol{\beta}' = [\boldsymbol{\beta}_1', \ \boldsymbol{\beta}_2'] \in R^{2p}$, $\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\theta}, \sigma^2\Phi)$ so that

$$\Pi(\boldsymbol{\beta}_0|\sigma^2) = |\Phi_{11}|^{-\frac{1}{2}} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{p}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}_0-\boldsymbol{\theta}_0)'\Phi_{11}^{-1}(\boldsymbol{\beta}_0-\boldsymbol{\theta}_0)} \tag{3.2.3}$$

and

$$\Pi(\boldsymbol{\beta}|\sigma^2) = |\Phi|^{-\frac{1}{2}} \left(\frac{1}{2\pi\sigma^2}\right)^{p} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})}. \tag{3.2.4}$$

Furthermore the marginal prior density of σ^2 if $k = n$ is

$$\Pi(\sigma^2) = \frac{\gamma_1^{\alpha_1}}{\Gamma(\alpha_1)} \left(\frac{1}{\sigma^2}\right)^{\alpha_1+1} e^{-\frac{\gamma_1}{\sigma^2}}, \quad \text{(i.e. } \sigma^2 \sim IG(\alpha_1, \gamma_1), \sigma^2 > 0)$$

while if $k \neq n$,

$$\Pi(\sigma^2) = \frac{\gamma^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} e^{-\frac{\gamma}{\sigma^2}}, \quad \text{(i.e. } \sigma^2 \sim IG(\alpha, \gamma)). \tag{3.2.5}$$

In addition, $\boldsymbol{\theta}' = [\boldsymbol{\theta}_0', \boldsymbol{\theta}_0']$ where $\boldsymbol{\theta}_0 \in I\!R^p$ and the covariance matrix Φ is a positive-definite matrix of order $2p$ where

$$\Phi = \begin{bmatrix} \Phi_{11} & 0 \\ 0 & \Phi_{11} \end{bmatrix}$$

with Φ a $2p \times 2p$ matrix and Φ_{11} a $p \times p$ matrix.

The likelihood function for $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ and σ^2 follows from (3.2.1) as

$$L(k, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2|\boldsymbol{y}) = \begin{cases} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta}_0)'(\boldsymbol{y} - X\boldsymbol{\beta}_0)}; & k = n \\ \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}_k - X_k\boldsymbol{\beta})'(\boldsymbol{y}_k - X_k\boldsymbol{\beta})}; & p \leq k \leq n - p \end{cases} \tag{3.2.6}$$

where

$$\boldsymbol{y} = \boldsymbol{y}_k = \begin{bmatrix} \boldsymbol{y}_{1k} \\ \boldsymbol{y}_{2k} \end{bmatrix} : n \times 1,$$

$$\boldsymbol{y}_{1k} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} : k \times 1$$

and

$$\boldsymbol{y}_{2k} = \begin{bmatrix} y_{k+1} \\ \vdots \\ y_n \end{bmatrix} : (n - k) \times 1.$$

Furthermore $X = \begin{bmatrix} X_{1k} \\ X_{2k} \end{bmatrix}$ and $X_k = \begin{bmatrix} X_{1k} & 0 \\ 0 & X_{2k} \end{bmatrix} : n \times 2p$

where $X_{1k} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_k \end{bmatrix} : k \times p$ and $X_{2k} = \begin{bmatrix} \boldsymbol{x}_{k+1} \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} : (n - k) \times p.$

The posterior distribution if $k = n$, according to Bayes' theorem, is

$$\Pi(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma^2, k = n|\boldsymbol{y}, \alpha_1, \gamma_1, \Phi_{11}) \;\alpha\; q|\Phi_{11}|^{-\frac{1}{2}}(2\pi)^{-(\frac{p+n}{2})} \frac{\gamma_1^{\alpha_1}}{\Gamma(\alpha_1)} \left(\frac{1}{\sigma^2}\right)^{\alpha_1 + \frac{p+n}{2} + 1} e^{-\frac{1}{\sigma^2}[\gamma + \frac{1}{2}A]}$$

where

$$A = Q_1(n) + J_1(n),$$

$$Q_1(n) = (\boldsymbol{\beta}_0 - \widehat{\widehat{\boldsymbol{\beta}}})'(X'X + \Phi_{11}^{-1})(\boldsymbol{\beta}_0 - \widehat{\widehat{\boldsymbol{\beta}}}), \tag{3.2.7}$$

$$J_1(n) = \boldsymbol{y}'\boldsymbol{y} + \boldsymbol{\theta}_0'\Phi_{11}^{-1}\boldsymbol{\theta}_0 - (X'\boldsymbol{y} + \Phi_{11}^{-1}\boldsymbol{\theta}_0)'\widehat{\widehat{\boldsymbol{\beta}}} \tag{3.2.8}$$

and
$$\widehat{\widehat{\boldsymbol{\beta}}} = (X'X + \Phi_{11}^{-1})^{-1}(X'\boldsymbol{y} + \Phi_{11}^{-1}\boldsymbol{\theta}_0), \tag{3.2.9}$$

which results in

$$\Pi(\boldsymbol{\beta}_0, \theta_0, \sigma^2, k = n|\boldsymbol{y}, \alpha_1, \gamma_1, \Phi_{11}) \propto q|\Phi_{11}|^{-\frac{1}{2}}(2\pi)^{-\frac{p+n}{2}} \frac{\gamma_1^{\alpha_1}}{\Gamma(\alpha_1)} \left(\frac{1}{\sigma^2}\right)^{\alpha_1 + \frac{p+n}{2} + 1} \cdot$$

$$e^{-\frac{1}{2\sigma^2}\left[2\gamma_1 + \left(\boldsymbol{\beta}_0 - \widehat{\widehat{\boldsymbol{\beta}}}\right)' \left(X'X + \Phi_{11}^{-1}\right) \left(\boldsymbol{\beta}_0 - \widehat{\widehat{\boldsymbol{\beta}}}\right) + y'y + \boldsymbol{\theta}_0'\Phi_{11}^{-1}\boldsymbol{\theta}_0 - \left(X'y + \Phi_{11}^{-1}\boldsymbol{\theta}\right)' \widehat{\widehat{\boldsymbol{\beta}}}\right]}. \tag{3.2.10}$$

The posterior distribution if $k \neq n$ is then likewise

$$\Pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2, k, \boldsymbol{\theta}|\boldsymbol{y}, \alpha, \gamma, \Phi) \quad \propto \quad \frac{1-q}{n-2p+1}|\Phi|^{-\frac{1}{2}}(2\pi)^{-\frac{2p+n}{2}}$$

$$\frac{\gamma^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha + \frac{2p+n}{2} + 1} e^{-\frac{1}{2\sigma^2}[2\gamma + B]} \tag{3.2.11}$$

where

$$B = Q(k) + J(k),$$

$$Q(k) = (\boldsymbol{\beta} - \widehat{\widehat{\boldsymbol{\beta}}}_k)'(X_k'X_k + \Phi^{-1})(\boldsymbol{\beta} - \widehat{\widehat{\boldsymbol{\beta}}}_k), \tag{3.2.12}$$

$$J(k) = \boldsymbol{y}_k'\boldsymbol{y}_k + \boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta} - (X_k'\boldsymbol{y}_k + \Phi^{-1}\boldsymbol{\theta})'\hat{\hat{\boldsymbol{\beta}}}_k \tag{3.2.13}$$

and
$$\hat{\hat{\boldsymbol{\beta}}}_k = (X_k'X_k + \Phi^{-1})^{-1}(X_k'\boldsymbol{y}_k + \Phi^{-1}\boldsymbol{\theta}), \tag{3.2.14}$$

so that

$$\Pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2, \boldsymbol{\theta}, k|\boldsymbol{y}, \alpha, \gamma, \Phi) \propto \frac{1-q}{n-2p+1}|\Phi|^{-\frac{1}{2}}(2\pi)^{-\frac{2p+n}{2}}\frac{\gamma^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{2p+n}{2}+1}$$
$$e^{-\frac{1}{2\sigma^2}\left[2\gamma+(\boldsymbol{\beta}-\hat{\hat{\boldsymbol{\beta}}}_k)'(X_k'X_k+\Phi^{-1})(\boldsymbol{\beta}-\hat{\hat{\boldsymbol{\beta}}}_k)+\boldsymbol{y}_k'\boldsymbol{y}_k+\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta}-(X_k'\boldsymbol{y}_k+\Phi^{-1}\boldsymbol{\theta})'\hat{\hat{\boldsymbol{\beta}}}_k\right]}. \tag{3.2.15}$$

Note that $\boldsymbol{\theta} = J_p'\boldsymbol{\theta}_0$ where $J_p = [I_p \ I_p]$ so that $\Phi^{-1}\boldsymbol{\theta} = \Phi^{-1}J_p'\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta} = 2\boldsymbol{\theta}_0'\Phi_{11}^{-1}\boldsymbol{\theta}_0$.

Our interest is the marginal posterior mass function of k, which can be obtained from the joint posterior density of the parameters by eliminating all parameters except k. Therefore, by integrating out $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ and σ^2 it follows that

$$\Pi(k = n|\Phi_{11}, \alpha_1, \gamma_1, \boldsymbol{\theta}_0, \boldsymbol{y}) \propto \frac{\gamma_1^{\alpha_1} q|\Phi_{11}|^{-\frac{1}{2}}\Gamma\left(\alpha_1 + \frac{n}{2}\right)}{\Gamma(\alpha_1)(2\pi)^{\frac{n}{2}}|X'X + \Phi_{11}^{-1}|^{\frac{1}{2}}[2\gamma_1 + J_1(n)]^{\left(\frac{2\alpha_1+n}{2}\right)}} \tag{3.2.16}$$

and

$$\Pi(k|\Phi, \alpha, \gamma, \boldsymbol{\theta}, \boldsymbol{y}) \propto \frac{(1-q)\gamma^\alpha|\Phi|^{-\frac{1}{2}}\Gamma\left(\alpha + \frac{n}{2}\right)}{(n-2p+1)\Gamma(\alpha)(2\pi)^{\frac{n}{2}}|X_k'X_k + \Phi^{-1}|^{\frac{1}{2}}[2\gamma + J(k)]^{\left(\frac{2\alpha+n}{2}\right)}}, \tag{3.2.17}$$

for $k = p, \cdots, n-p$.

These expressions ((3.2.16) and (3.2.17)) still leave unknown parameters $\alpha, \gamma, \alpha_1, \gamma_1, \boldsymbol{\theta}_0$

and Φ_{11}. We proceed to put a vague prior to $\boldsymbol{\theta}_0$ and integrating to leave expressions with unknown $\Phi_{11}, \alpha, \gamma, \alpha_1$ and γ_1.

From (3.2.16), after completing the square, it follows that

$$2\gamma_1 + J_1(n) = (\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\beta}})'[\Phi_{11} + (X'X)^{-1}]^{-1}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\beta}}) + 2\gamma_1 + \boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{\beta}}'X'X\widehat{\boldsymbol{\beta}}$$

where

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} \tag{3.2.18}$$

so that

$$\Pi(k = n|\Phi_{11}, \alpha_1, \gamma_1, \boldsymbol{y}) \propto$$

$$\frac{\gamma_1^{\alpha_1} q|\Phi_{11}|^{-\frac{1}{2}}\Gamma(\alpha_1 + \frac{n}{2})[(2\alpha_1 + n - p)\pi]^{\frac{p}{2}}\Gamma(\frac{2\alpha_1+n-p}{2})\left|\frac{(2\alpha_1+n-p)N_1(n)}{M_1(n)}\right|^{-\frac{1}{2}}}{\Gamma(\alpha_1)(2\pi)^{\frac{n}{2}}|X'X + \Phi_{11}^{-1}|^{\frac{1}{2}}\Gamma(\frac{2\alpha_1+n}{2})}$$

$$\cdot[M_1(n)]^{-(\frac{2\alpha_1+n}{2})} \tag{3.2.19}$$

where

$$M_1(n) = 2\gamma_1 + \boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{\beta}}'X'X\widehat{\boldsymbol{\beta}}, \quad N_1(n) = [\Phi_{11} + (X'X)^{-1}]^{-1}. \tag{3.2.20}$$

From (3.2.17), after completing the square, it follows that

$$2\gamma + J(k) = (\boldsymbol{\theta}_0 - \tilde{\boldsymbol{\beta}}_k)'J_p[\Phi + (X_k'X_k)^{-1}]^{-1}J_p'(\boldsymbol{\theta}_0 - \tilde{\boldsymbol{\beta}}_k) + M(k)$$

where

$$\tilde{\boldsymbol{\beta}}_k = [J_p(\Phi + (X_k'X_k)^{-1})^{-1}J_p']^{-1}J_p\Phi^{-1}(X_k'X_k + \Phi^{-1})^{-1}X_k'\boldsymbol{y}_k, \tag{3.2.21}$$

$$M(k) = 2\gamma + \boldsymbol{y}_k'H_A\boldsymbol{y}_k - \tilde{\boldsymbol{\beta}}_k[\Phi_{11} + (X_{1k}'X_{1k})^{-1} + (X_{2k}'X_{2k})^{-1}] \tag{3.2.22}$$

and

$$H_A = I_n - X_k[X_k'X_k + \Phi^{-1}]^{-1}X_k' \tag{3.2.23}$$

so that

$$\Pi(k|\Phi,\alpha,\gamma,\boldsymbol{y}) \propto \frac{(1-q)\gamma^{\alpha}|\Phi|^{-\frac{1}{2}}\Gamma(\frac{2\alpha+n-p}{2})[(2\alpha+n-p)\Pi]^{\frac{p}{2}}}{(n-2p+1)\Gamma(\alpha)(2\pi)^{\frac{n}{2}}|X_k'X_k+\Phi^{-1}|^{\frac{1}{2}}}$$

$$\left|\frac{(2\alpha+n-p)J_p[\Phi+(X_k'X_k)^{-1}]^{-1}J_p'}{2\gamma+\boldsymbol{y}_k'\boldsymbol{y}_k-\tilde{\boldsymbol{\beta}}_k'X_k'X_k\tilde{\boldsymbol{\beta}}_k}\right|^{-\frac{1}{2}}(2\gamma+\boldsymbol{y}_k'\boldsymbol{y}_k-\hat{\boldsymbol{\beta}}_k'X_k'X_k\hat{\boldsymbol{\beta}}_k)^{-(\frac{2\alpha+n}{2})}. \tag{3.2.24}$$

Note that if $\Phi = \begin{bmatrix} \Phi_1 & 0 \\ 0 & \Phi_2 \end{bmatrix}$ then $J_p\Phi^{-1}J_p' = \Phi_1^{-1} + \Phi_2^{-1}$ and

$$J_p[\Phi + (X_k'X_k)^{-1}]^{-1}J_p' = [\Phi_1 + (X_{1k}'X_{1k})^{-1}]^{-1} + [\Phi_2 + (X_{2k}'X_{2k})^{-1}]^{-1} = C$$

and

$$\begin{aligned} J_p\Phi^{-1}(X_k'X_k+\Phi^{-1})^{-1}X_k'\boldsymbol{y} &= \Phi_1^{-1}(X_{1k}'X_{1k}+\Phi_1^{-1})^{-1}X_{1k}'\boldsymbol{y}_{1k} \\ &\quad +\Phi_2^{-1}(X_{2k}'X_{2k}+\Phi_2^{-1})^{-1}X_{2k}'\boldsymbol{y}_{2k} = D \end{aligned}$$

so that

$$\boldsymbol{\theta}_0|\boldsymbol{y},k,\Phi,\sigma^2 \sim N(C^{-1}D, \sigma^2C^{-1}) \tag{3.2.25}$$

and

$$\boldsymbol{\theta}_0|\boldsymbol{y},k,\Phi \sim t\left(C^{-1}D, n+2\alpha-p, \frac{(2\alpha+n-p)C}{2\gamma+\boldsymbol{y}_k'\boldsymbol{y}_k-\widehat{\boldsymbol{\beta}}_kX_k'X_k\widehat{\boldsymbol{\beta}}_k}\right). \tag{3.2.26}$$

Furthermore, the posterior of $\boldsymbol{\theta}_0$ unconditional of k is given by

$$\begin{aligned} \Pi(\boldsymbol{\theta}_0|\boldsymbol{y},\alpha,\gamma,\Phi) \propto\ & \Pi(\boldsymbol{\theta}_0|\boldsymbol{y},\Phi,k=n)\Pi(k=n|\Phi_{11},\alpha_1,\gamma_1,\boldsymbol{y}) + \\ & \sum_{k=p}^{n-p}\Pi(\boldsymbol{\theta}_0|\boldsymbol{y},\alpha,\gamma,k)\Pi(k|\Phi,\alpha,\gamma,\boldsymbol{y}). \end{aligned} \tag{3.2.27}$$

Note that all the equations are still dependent on the unknown parameters Φ, α and γ.

3.2.1.1 EXACTLY ONE CHANGE-POINT IN THE MODEL

Assuming now that we have established the existence of a structural change, we can obtain probabilities of a change at $k = 1, 2, \ldots, n-1$. The behaviour of the other parameters of the model can then be analyzed, given the change at k, where $1 \leq k \leq n-1$.

Using equation (3.2.17), ignoring all constants, we have

$$\Pi(k|\boldsymbol{y}, \Phi, \alpha, \gamma, \boldsymbol{\theta}) \propto |X_k' X_k + \Phi^{-1}|^{-\frac{1}{2}}[2\gamma + J(k)]^{-\frac{2\alpha+n}{2}} \tag{3.2.28}$$

where, from (3.2.13) and (3.2.14), it follows that

$$\begin{aligned} J(k) &= \boldsymbol{y}_k' \boldsymbol{y}_k + \boldsymbol{\theta}' \Phi^{-1} \boldsymbol{\theta} - (X_k' \boldsymbol{y}_k + \Phi^{-1}\boldsymbol{\theta})' \hat{\hat{\boldsymbol{\beta}}}_k \\ &= [\boldsymbol{y}_k - X_k \hat{\hat{\boldsymbol{\beta}}}_k]' \boldsymbol{y}_k + [\boldsymbol{\theta} - \hat{\hat{\boldsymbol{\beta}}}_k]' \Phi^{-1} \boldsymbol{\theta} \end{aligned} \tag{3.2.29}$$

where $\hat{\hat{\boldsymbol{\beta}}}_k$ is given in (3.2.14).

The marginal posterior mass function in (3.2.28) can therefore be expressed in a different way, i.e.

$$\Pi(k|\Phi, \boldsymbol{y}, \alpha, \gamma, \boldsymbol{\theta}) \propto |X_k' X_k + \Phi^{-1}|^{-\frac{1}{2}}[2\gamma + (\boldsymbol{y}_k - X_k \hat{\hat{\boldsymbol{\beta}}}_k)' \boldsymbol{y}_k + (\boldsymbol{\theta} - \hat{\hat{\boldsymbol{\beta}}}_k)' \Phi^{-1} \boldsymbol{\theta}]^{-\frac{2\alpha+n}{2}}. \tag{3.2.30}$$

Once the change-point has been determined at, say $k = k^*$, we are now interested in the distributions of $\boldsymbol{\beta}$ and σ^2.

Using (3.2.11) it follows that

$$\Pi(\boldsymbol{\beta}|k=k^*,\boldsymbol{y},\alpha,\gamma,\Phi) \propto \int_{-\infty}^{\infty}\int_{0}^{\infty}\Pi(\boldsymbol{\beta},\boldsymbol{\theta}_0,\sigma^2,k^*|\boldsymbol{y},\alpha,\gamma,\Phi)d\sigma^{-2}d\boldsymbol{\theta}_0$$

$$\propto \int_{-\infty}^{\infty}\int_{0}^{\infty}\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{2p+n}{2}+1} e^{-\frac{1}{2\sigma^2}\left[2\gamma+\left(\boldsymbol{\beta}-\hat{\hat{\boldsymbol{\beta}}}_k\right)'\left(X_k'X_k+\Phi^{-1}\right)\left(\boldsymbol{\beta}-\hat{\hat{\boldsymbol{\beta}}}_k\right)\right.}$$

$$\left.{}^{+\boldsymbol{y}_k'\boldsymbol{y}_k+\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta}-\left(X_k'\boldsymbol{y}_k+\Phi^{-1}\boldsymbol{\theta}\right)'\hat{\hat{\boldsymbol{\beta}}}_k\right]} d\sigma^{-2}d\boldsymbol{\theta}_0$$

$$\propto \int_{-\infty}^{\infty}[2\gamma+(\boldsymbol{\beta}-\hat{\hat{\boldsymbol{\beta}}}_k)'(X_k'X_k+\Phi^{-1})(\boldsymbol{\beta}-\hat{\hat{\boldsymbol{\beta}}}_k)+\boldsymbol{y}_k'\boldsymbol{y}_k+\boldsymbol{\theta}'\Phi^{-1}\boldsymbol{\theta}$$

$$-(X_k'\boldsymbol{y}_k+\Phi^{-1}\boldsymbol{\theta})'\hat{\hat{\boldsymbol{\beta}}}_k]^{-(\frac{2\alpha+2p+n}{2})}d\boldsymbol{\theta}_0$$

$$\propto \int_{-\infty}^{\infty}[U+(\boldsymbol{\theta}_0-\boldsymbol{b})'2\Phi_k^{-1}(\boldsymbol{\theta}_0-\boldsymbol{b})]^{-\frac{2\alpha+2p+n}{2}}d\boldsymbol{\theta}_0$$

$$\propto \int_{-\infty}^{\infty}\left[1+\frac{1}{(2\alpha+n+p)}(\boldsymbol{\theta}_0-\boldsymbol{b})'\frac{J_p\Phi^{-1}J_p'(2\alpha+n+p)}{U}(\boldsymbol{\theta}_0-\boldsymbol{b})\right]^{-[\frac{2\alpha+n+2p}{2}]}d\boldsymbol{\theta}_0$$

$$U^{-[\frac{(2\alpha+n+p)+p}{2}]}$$

where

$$U=[\boldsymbol{y}_k-X_k\boldsymbol{\beta}]'[\boldsymbol{y}_k-X_k\boldsymbol{\beta}]+\boldsymbol{\beta}'\Phi^{-1}\boldsymbol{\beta}+2\gamma-\frac{1}{2}\boldsymbol{\beta}'(I_2\otimes\Phi_{11}^{-1})\boldsymbol{\beta}$$

and

$$\boldsymbol{b}=\frac{1}{2}J_p\boldsymbol{\beta}=\frac{1}{2}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2) \tag{3.2.31}$$

and which is the integral of a multivariate t-distribution with $2\alpha+n+p$ degrees of freedom, location vector $\boldsymbol{b}$ and precision matrix $\dfrac{J_pA^{-1}J_p'(2\alpha+n+p)}{U}$ so that

$$\Pi(\boldsymbol{\beta}|k=k^*,\boldsymbol{y},\alpha,\gamma,\Phi) \propto U^{-(\frac{2\alpha+n+p}{2})} \tag{3.2.32}$$

where

$$\begin{aligned} U \;=\; & \left\{\boldsymbol{\beta} - \left[X_k'X_k + \Phi^{-1} - \frac{1}{2}J_p'\Phi_k^{-1}J_p\right]^{-1} X_k'\boldsymbol{y}_k\right\}' . \\ & \left[X_k'X_k + \Phi^{-1} - \Phi^{-1}J_p'(J_p\Phi^{-1}J_p')^{-1}J_p\Phi^{-1}\right] . \\ & \left\{\boldsymbol{\beta} - \left[X_k'X_k + \Phi^{-1} - \frac{1}{2}J_p'\Phi_k^{-1}J_p\right]^{-1} X_k'\boldsymbol{y}_k\right\} + 2\gamma + \boldsymbol{y}_k'\boldsymbol{y}_k \end{aligned}$$

so that $\Pi(\boldsymbol{\beta}|k=k^*,\boldsymbol{y},\alpha,\gamma,\Phi)$ is a multivariate t-distribution with $2\alpha+n-p$ degrees of freedom, location vector $[X_k'X_k + \Phi^{-1} - \frac{1}{2}J_p\Phi_k^{-1}J_p]^{-1}X_k'\boldsymbol{y}_k$ and precision matrix

$$\frac{\left[X_k'X_k + \Phi^{-1} - \frac{1}{2}J_p'\Phi_k^{-1}J_p\right](2\alpha+n-p)}{2\gamma + \boldsymbol{y}_k'\boldsymbol{y}_k - \boldsymbol{y}_k'X_k\left[X_k'X_k + \Phi^{-1} - \frac{1}{2}J_p'\Phi_{11}^{-1}J_p\right]X_k'\boldsymbol{y}_k'} . \tag{3.2.33}$$

Furthermore, by using (3.2.24) and (3.2.32), it follows that

$$\Pi(\boldsymbol{\beta}|\boldsymbol{y},\alpha,\gamma,\Phi) = \sum_{k=1}^{n-1} \Pi(k|\boldsymbol{y},\alpha,\gamma,\Phi)\Pi(\boldsymbol{\beta}|k,\boldsymbol{y},\alpha,\gamma,\Phi) \tag{3.2.34}$$

where

$$\begin{aligned} \Pi(k|\boldsymbol{y},\alpha,\gamma,\Phi) \;\propto\; & |X_k'X_k + \Phi^{-1}|^{-\frac{1}{2}} \left|\frac{(2\alpha+n-p)J_p[\Phi + (X_k'X_k)^{-1}]^{-1}J_p'}{2\gamma + \boldsymbol{y}_k'\boldsymbol{y}_k - \tilde{\boldsymbol{\beta}}_k'X_k'X_k\tilde{\boldsymbol{\beta}}_k}\right|^{-\frac{1}{2}} \\ & \left[2\gamma + \boldsymbol{y}_k'\boldsymbol{y}_k - \widehat{\boldsymbol{\beta}}_k'X_k'X_k\widehat{\boldsymbol{\beta}}_k\right]^{-(\frac{2\alpha+n}{2})} \end{aligned} \tag{3.2.35}$$

and $\Pi(\boldsymbol{\beta}|k=k^*,\boldsymbol{y},\alpha,\gamma,\Phi)$ is given by (3.2.32).

The marginal posterior density of σ^2, conditional on k, is

$$\begin{aligned}\Pi(\sigma^2|\alpha,\gamma,k=k^*,\boldsymbol{y},\Phi) &\propto \int\int_{-\infty}^{\infty}\Pi(\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}_0|k=k^*,\boldsymbol{y})d\boldsymbol{\beta}d\boldsymbol{\theta}_0\\ &\propto \int\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{2p+n}{2}+1}e^{-\frac{1}{2\sigma^2}[2\gamma+J(k)]}\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}Q(k)}d\boldsymbol{\beta}d\boldsymbol{\theta}_0\end{aligned}$$

where $Q(k)$ is as in (3.2.12) and $J(k)$ is given by (3.2.13), so that

$$\Pi(\sigma^2|k=k^*,\boldsymbol{y},\alpha,\Phi,\gamma)\propto\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{n}{2}+1}\int_{\infty}^{\infty}e^{-\frac{1}{2\sigma^2}[2\gamma+J(k)]}d\boldsymbol{\theta}_0$$

where

$$\begin{aligned}&\int_{-\infty}^{\infty}e^{-\frac{1}{2\sigma^2}[2\gamma+J(k)]}d\boldsymbol{\theta}_0\\ &= e^{-\frac{1}{2\sigma^2}\left[2\gamma+\boldsymbol{y}_k'\boldsymbol{y}_k-\widehat{\boldsymbol{\beta}}_k'X_k'X_k\widehat{\boldsymbol{\beta}}_k\right]}(2\pi)^p\left|J_p[\Phi+(X_k'X_k)^{-1}]^{-1}J_p'\right|^{-\frac{1}{2}}(\sigma^2)^{\frac{p}{2}}.\end{aligned}$$

This results in

$$\Pi(\sigma^2|\boldsymbol{y},k^*,\alpha,\gamma)\propto\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{n-p}{2}+1}e^{-\frac{1}{2\sigma^2}\left[2\gamma+\boldsymbol{y}_k'\boldsymbol{y}_k-\widehat{\boldsymbol{\beta}}_k'X_k'X_k\widehat{\boldsymbol{\beta}}_k\right]}\tag{3.2.36}$$

which is a IG$\left(\alpha+\dfrac{n-p}{2};\dfrac{1}{2}\left(2\gamma+\boldsymbol{y}_k'\boldsymbol{y}_k-\widehat{\boldsymbol{\beta}}_k'X_k'X_k\widehat{\boldsymbol{\beta}}_k\right)\right)$.

The marginal posterior density of σ^2, unconditional of k, is

$$\Pi(\sigma^2|\boldsymbol{y},\alpha,\gamma,\Phi)=\sum_{k=1}^{n-1}\Pi(k|\boldsymbol{y},\alpha,\gamma,\Phi_{11})\Pi(\sigma^2|k,\boldsymbol{y},\alpha,\gamma)\tag{3.2.37}$$

where

$\Pi(k|\boldsymbol{y},\alpha,\gamma,\Phi_{11})$ is given by (3.2.24) and (3.2.35) and $\Pi(\sigma^2|k,\boldsymbol{y},\alpha,\gamma)$ is given by (3.2.36).

Note that the distributions are still dependent of α, γ and Φ_{11}.

Since the number of change-points is known, we can put vague priors on the parameters $\boldsymbol{\beta}$ and σ^2, that is if $\alpha, \gamma, \Phi_{11}^{-1} \to 0$ in (3.2.4) to (3.2.5), then $\Pi(\boldsymbol{\beta}, k, \sigma^2) \propto \frac{1}{\sigma^2}$. The expressions for $Q(k)$ and $J(k)$ becomes

$$J(k) = \boldsymbol{y}_k' H_k \boldsymbol{y}_k$$

and

$$Q(k) = [\boldsymbol{\beta} - (X_k'X_k)^{-1}(X_k'\boldsymbol{y}_k)]'[X_k'X_k][\boldsymbol{\beta} - (X_k'X_k)^{-1}(X_k'\boldsymbol{y}_k)]$$

where

$$H_k = I_n - X_k(X_k'X_k)^{-1}X_k'. \tag{3.2.38}$$

The posterior of $\boldsymbol{\beta}|\boldsymbol{y}, k$ becomes a multivariate t-distribution with $n - 2p$ degrees of freedom, location vector $[X_k'X_k]^{-1}X_k'\boldsymbol{y}_k = \widehat{\boldsymbol{\beta}}_k$ and precision matrix $\dfrac{|X_k'X_k|(n-2p)}{\boldsymbol{y}_k'H_k\boldsymbol{y}_k}$, and the posterior of σ^2 becomes a $\mathrm{IG}\left(\dfrac{n-2p}{2}; \dfrac{1}{2}\boldsymbol{y}_k'H_k\boldsymbol{y}_k\right)$.

The marginal posterior mass function of k will be

$$\Pi(k|\boldsymbol{y}) \alpha |X_k'X_k|^{-\frac{1}{2}}(\boldsymbol{y}'H_k\boldsymbol{y})^{-\frac{n-2p}{2}}. \tag{3.2.39}$$

3.2.2 A CHANGE IN THE VARIANCE WITH CONSTANT REGRESSION COEFFICIENT

Consider the linear model where a change may have occurred,

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + e_i; \quad i = 1, \ldots, n$$

or

$$\begin{aligned} y_i &= \boldsymbol{x}_i\boldsymbol{\beta} + e_i; \quad i = 1, \ldots, k \\ y_i &= \boldsymbol{x}_i\boldsymbol{\beta} + u_i; \quad i = k+1, \ldots, n \end{aligned} \tag{3.2.40}$$

where $1 \leq k \leq n-1$ where $n \geq p+1$.

The $\boldsymbol{x}_i$'s and $\boldsymbol{\beta}$ are the same as previously and e_i are i.i.d. normal random variables with mean zero and variance $\sigma_1^2 > 0$ if $i = 1, \ldots, k$, and u_i is distributed with mean zero and variance $\sigma_2^2 > 0$ if $i = k+1, \ldots, n$.

Exactly one change has occurred somewhere if $1 \leq k \leq n-1$. So if $1 \leq k \leq n-1$, $\sigma_1^2 \neq \sigma_2^2$, where k is unknown. Furthermore, if $k = n$, i.e. no change has occurred, $\sigma^{-2} \sim \Gamma(\alpha, \gamma)$ and if $k \neq n$ (i.e. $1 \leq k \leq n-1$ and exactly one change has occurred), $\sigma_1^{-2}, \sigma_2^{-2} \sim \Gamma(\alpha, \gamma)$, independently.

Let the marginal prior mass function of k be as in (3.2.2) and let $\boldsymbol{\beta}$ be assigned a normal density with $\boldsymbol{\beta} \in I\!\!R^p \sim N(\boldsymbol{\theta}, \Phi)$ and also let

$$\Sigma_0 = \sigma^2 I_n, \ \underset{(2\times 2)}{\sum}1 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \quad \text{and} \quad \underset{(n\times n)}{\sum}k = \begin{pmatrix} \sigma_1^2 I_k & 0 \\ 0 & \sigma_2^2 I_{n-k} \end{pmatrix} \quad \text{where } \sigma^2 > 0$$

so that

$$\Pi(\boldsymbol{\beta}) \propto |\Phi|^{-\frac{1}{2}} \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})}. \tag{3.2.41}$$

Furthermore the marginal prior density of σ_1^2, σ_2^2 is

$$\begin{aligned} \Pi(\sigma_1^2, \sigma_2^2) &= \frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2} \left(\frac{1}{\sigma_1^2\sigma_2^2}\right)^{\alpha+1} e^{-\gamma\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)} \\ &\propto |\Sigma_1|^{-(\alpha+1)} e^{-\gamma tr \Sigma_1^{-1}}, \qquad (1 \leq k \leq n-1) \end{aligned}$$

and

$$\Pi(\sigma^2) = \frac{\gamma^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} e^{-\frac{\gamma}{\sigma^2}}, \qquad (k=n), \tag{3.2.42}$$

The likelihood function for $\boldsymbol{\beta}_1, \sigma_1^2$ and σ_2^2 is from (3.2.6) for $k = n$

$$L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}- X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})}$$

while for $1 \leq k \leq n-1$ it follows that

$$L(k, \boldsymbol{\beta}_1, \sigma_1^2, \sigma_2^2|\boldsymbol{y}) = (2\pi\sigma_1^2)^{-\frac{k}{2}}(2\pi\sigma_2^2)^{-\frac{n-k}{2}} e^{-\frac{1}{2}(\boldsymbol{y}- X\boldsymbol{\beta})'\sum_k^{-1}(\boldsymbol{y}- X\boldsymbol{\beta})}. \tag{3.2.43}$$

The conjugate priors for both cases are given by

$$\begin{aligned}
\Pi(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \boldsymbol{\theta}, k|\boldsymbol{y}, \alpha, \gamma, \Phi) \quad &\propto \quad q(2\pi)^{-\frac{p}{2}}\left(\frac{1}{\sigma^2}\right)^{\alpha+1}|\Phi|^{-\frac{1}{2}}\frac{\gamma^\alpha}{\Gamma(\alpha)}e^{-\frac{\gamma}{\sigma^2}} \\
&\qquad e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})}, \quad (k=n) \\
&\propto \quad \frac{1-q}{n-1}(2\pi)^{-\frac{p}{2}}|\Phi|^{-\frac{1}{2}}|\Sigma_1|^{-(\alpha+1)}\frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2} \\
&\qquad e^{-\gamma tr\Sigma_1^{-1}} e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})}, \quad (1 \leq k \leq n-1).
\end{aligned} \tag{3.2.44}$$

The posterior distribution if $k = n$, according to Bayes' theorem, is

$$\begin{aligned}
\Pi(\boldsymbol{\beta}_1, \sigma^2, k=n|\boldsymbol{y}, \alpha, \gamma, \Phi) \quad &\propto \quad (2\pi)^{-\left(\frac{n+p}{2}\right)}|\Phi|^{-\frac{1}{2}}q\frac{\gamma^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{n}{2}+1} e^{-\frac{\gamma}{\sigma^2}} \\
&\qquad e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})} e^{-\frac{1}{2}(\boldsymbol{y}- X\boldsymbol{\beta})'\Sigma^{-1}(\boldsymbol{y}-X\boldsymbol{\beta})}.
\end{aligned} \tag{3.2.45}$$

Furthermore the marginal posterior mass function of $k = n$ is

$$\begin{aligned}
\Pi(k=n|\boldsymbol{y},\alpha,\gamma) &= \iiint \Pi(\boldsymbol{\beta},\sigma^2,k=n|\boldsymbol{y},\alpha,\gamma,\Phi)d\boldsymbol{\theta}d\boldsymbol{\beta}d\sigma^2 \\
&= \int (2\pi)^{-\frac{n}{2}} q\frac{\gamma^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{n}{2}+1} e^{-\frac{\gamma}{\sigma^2}} \int e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta})} d\boldsymbol{\beta}d\sigma^2 \\
&= (2\pi)^{-\frac{n}{2}} q\frac{\gamma^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{n}{2}+1} e^{-\frac{\gamma}{\sigma^2}} \int e^{-\frac{1}{2\sigma^2}\left[\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)'X'X(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})+\boldsymbol{y}'H\boldsymbol{y}\right]} d\boldsymbol{\beta}d\sigma^2
\end{aligned}$$

where

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}, \tag{3.2.46}$$

$$H = I - X(X'X)^{-1}X' \tag{3.2.47}$$

so that

$$\boldsymbol{y}'H\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{\beta}}'X'X\widehat{\boldsymbol{\beta}}.$$

Therefore

$$\begin{aligned}
\Pi(k=n|\boldsymbol{y},\alpha,\gamma) &= q(2\pi)^{-\left(\frac{n-p}{2}\right)}\frac{\gamma^\alpha}{\Gamma(\alpha)}|X'X|^{-\frac{1}{2}}\int\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{n-p}{2}+1} e^{-\frac{1}{\sigma^2}[\gamma+\frac{1}{2}\boldsymbol{y}'H\boldsymbol{y}]}d\sigma^2 \\
&= q(2\pi)^{-\left(\frac{n-p}{2}\right)}\frac{\gamma^\alpha}{\Gamma(\alpha)}|X'X|^{-\frac{1}{2}}\Gamma\left(\frac{n+2\alpha-p}{2}\right)\left[\gamma+\frac{1}{2}\boldsymbol{y}'H\boldsymbol{y}\right]^{-\left(\frac{n+2\alpha-p}{2}\right)}.
\end{aligned} \tag{3.2.48}$$

The posterior distribution if $1 \le k \le n-1$ according to Bayes' theorem is

$$\Pi(\boldsymbol{\beta},\sigma_1^2,\sigma_2^2,k|\boldsymbol{y},\alpha,\gamma,\boldsymbol{\theta},\Phi) = \frac{1-q}{n-1}(2\pi)^{-\left(\frac{n-p}{2}\right)}(\sigma_1^2)^{-\frac{k}{2}}(\sigma_2^2)^{-\frac{n-k}{2}}\frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2}$$

$$e^{-\gamma tr\Sigma_1^{-1}}|\Sigma_1|^{-(\alpha+1)}|\Phi|^{-\frac{1}{2}}e^{-\frac{1}{2}(\boldsymbol{y}-X\boldsymbol{\beta})'\Sigma_k^{-1}(\boldsymbol{y}-X\boldsymbol{\beta})}e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\theta})'\Phi^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})}. \tag{3.2.49}$$

Furthermore the marginal posterior mass function, if $1 \leq k \leq n-1$, is

$$\Pi(k \neq n|\boldsymbol{y}, \alpha, \gamma, \Phi) = \iiint \frac{1-q}{n-2p+1}(2\pi)^{-\left(\frac{n-p}{2}\right)} \frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2}|\Phi|^{-\frac{1}{2}} \left(\frac{1}{\sigma_1^2\sigma_2^2}\right)^{\alpha+1}$$

$$e^{-\gamma\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}(\sigma_1^2)^{-\frac{k}{2}}(\sigma_2^2)^{-\left(\frac{n-k}{2}\right)} e^{-\frac{1}{2}[\boldsymbol{y}-X\boldsymbol{\beta}]'\,\Sigma_k^{-1}[\boldsymbol{y}-X\boldsymbol{\beta}]}$$

$$\int e^{-\frac{1}{2}[\boldsymbol{\theta}-\boldsymbol{\beta}]'\,\Phi^{-1}[\boldsymbol{\theta}-\boldsymbol{\beta}]} d\boldsymbol{\theta} d\boldsymbol{\beta} d\sigma_1^2 d\sigma_2^2$$

$$= \iint \frac{1-q}{n-2p+1}(2\pi)^{-\frac{n}{2}} \frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2}\left(\frac{1}{\sigma_1^2}\right)^{\frac{k}{2}+\alpha+1}\left(\frac{1}{\sigma_2^2}\right)^{\frac{n-k}{2}+\alpha+1} e^{-\frac{\gamma}{\sigma_1^2}}\, e^{-\frac{\gamma}{\sigma_2^2}}$$

$$\int e^{-\frac{1}{2}[\boldsymbol{y}-X\boldsymbol{\beta}]'\,\Sigma_k^{-1}[\boldsymbol{y}-X\boldsymbol{\beta}]} d\boldsymbol{\beta} d\sigma_1^2 d\sigma_2^2$$

$$= \iint \frac{1-q}{n-2p+1}(2\pi)^{-\frac{n}{2}} \frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2}\left(\frac{1}{\sigma_1^2}\right)^{\frac{k}{2}+\alpha+1}\left(\frac{1}{\sigma_2^2}\right)^{\frac{n-k}{2}+\alpha+1} e^{-\frac{\gamma}{\sigma_1^2}}\, e^{-\frac{\gamma}{\sigma_2^2}}$$

$$\int e^{-\frac{1}{2}\left[\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_k\right)' X'\Sigma_k^{-1}X\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_k\right)+\boldsymbol{y}'H_{\Sigma k}\boldsymbol{y}\right]} d\boldsymbol{\beta} d\sigma_1^2 d\sigma_2^2$$

where

$$\widehat{\boldsymbol{\beta}}_k = \left(X'\,\Sigma_k^{-1}\,X\right)^{-1} X'\,\Sigma_k^{-1}\,\boldsymbol{y} \tag{3.2.50}$$

and

$$H_{\Sigma k} = \Sigma_k^{-1} - \Sigma_k^{-1}X\left(X'\Sigma_k^{-1}X\right)^{-1}X'\Sigma_k^{-1} \tag{3.2.51}$$

so that

$$\boldsymbol{y}'H_{\Sigma k}\boldsymbol{y} = \boldsymbol{y}'\Sigma_k^{-1}\boldsymbol{y} - \boldsymbol{y}'\Sigma_k^{-1}X\left(X'\Sigma_k^{-1}X\right)^{-1}X'\Sigma_k^{-1}\boldsymbol{y}.$$

Therefore

$$\Pi(k|\boldsymbol{y},\alpha,\gamma) \propto \iint \frac{1-q}{n-2p+1}(2\pi)^{-\frac{n-p}{2}}\frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2}\left(\frac{1}{\sigma_1^2}\right)^{\frac{k}{2}+\alpha+1}\left(\frac{1}{\sigma_2^2}\right)^{\frac{n-k}{2}+\alpha+1}$$

$$e^{-\frac{\gamma}{\sigma_1^2}}\, e^{-\frac{\gamma}{\sigma_2^2}}\, e^{-\frac{1}{2}\boldsymbol{y}'H_{\Sigma k}\boldsymbol{y}}\left|X'\Sigma_k^{-1}X\right|^{-\frac{1}{2}} d\sigma_1^2 d\sigma_2^2 \tag{3.2.52}$$

where

$$\boldsymbol{y}'H_{\Sigma k}\boldsymbol{y} = \frac{1}{\sigma_1^2}y_{1k}'y_{1k}+\frac{1}{\sigma_2^2}y_{2k}'y_{2k}-\left[\frac{1}{\sigma_1^2}y_{1k}'X_{1k}+\frac{1}{\sigma_2^2}y_{2k}'X_{2k}\right]'$$

$$\left[X'\Sigma_k^{-1}X\right]^{-1}\left[\frac{1}{\sigma_1^2}X_k'y_{1k}+\frac{1}{\sigma_2^2}X_{2k}'y_{2k}\right]$$

with

$$X'\Sigma_k^{-1}X = \frac{1}{\sigma_1^2}X_{1k}'X_{1k}+\frac{1}{\sigma_2^2}X_{2k}'X_{2k}.$$

If we let $\sigma_2^2 = \frac{\sigma_1^2}{\delta}$, it follows that $\Sigma_k = \sigma_1^2\Delta_k$ and $\boldsymbol{y}'H_{\Sigma k}\boldsymbol{y} = \frac{1}{\sigma_1^2}\boldsymbol{y}'H_{k\Delta}\boldsymbol{y}$

where

$$H_{k\Delta} = \Delta_k - \Delta_k X(X'\Delta_k X)^{-1}X'\Delta_k \tag{3.2.53}$$

and

$$\Delta_k = \begin{pmatrix} I_k & 0 \\ 0 & \frac{1}{\delta}I_{n-k} \end{pmatrix}$$

so that, with prior $\pi(\delta) \propto \frac{1}{\delta}$,

$$\Pi(k|\boldsymbol{y},\alpha,\gamma) = \frac{1-q}{n-1}(2\pi)^{-\frac{n-p}{2}}\frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2}\left|X'\Delta_k X\right|^{-\frac{1}{2}}\delta^{\frac{n-k}{2}+\alpha-1}$$

$$\int_0^\infty\int_0^\infty\left(\frac{1}{\sigma_1^2}\right)^{\left(2\alpha-\frac{p}{2}+\frac{n}{2}\right)+1} e^{-\frac{1}{\sigma_1^2}\left\{\frac{1}{2}y'H_{k\Delta}\boldsymbol{y}+\gamma+\delta\gamma\right\}} d\sigma_1^2 d\delta$$

$$= \frac{1-q}{n-1}(2\pi)^{-\left(\frac{n-p}{2}\right)\frac{\gamma^{2\alpha}}{\Gamma(\alpha)^2}}\int_0^\infty \left|X'\Delta_k X\right|^{-\frac{1}{2}}\delta^{\frac{n-k}{2}+\alpha-1}\Gamma\left(\frac{n+4\alpha-p}{2}\right)$$

$$\left\{\frac{1}{2}\boldsymbol{y}'H_{k\Delta}\boldsymbol{y}+\gamma+\delta\gamma\right\}^{-\frac{n+4\alpha-p}{2}} d\delta.$$

Note that δ must be eliminated by numerical integration. To summarize, it follows that

$$\Pi(k|\boldsymbol{y},\alpha,\gamma) \propto q|X'X|^{-\frac{1}{2}}\Gamma\left(\frac{n-p+2\alpha}{2}\right)\left\{\gamma+\frac{1}{2}\boldsymbol{y}'H\boldsymbol{y}\right\}^{-\frac{n-p+2\alpha}{2}}, \text{ if } k=n$$

and

$$\begin{aligned}\Pi(k|\boldsymbol{y},\alpha,\gamma,\delta) \quad &\propto \quad \frac{1-q}{n-2p+1}\frac{\gamma^{\alpha}}{\Gamma(\alpha)}|X'\Delta_k X|^{-\frac{1}{2}}\Gamma\left(\frac{n+4\alpha-p}{2}\right)\delta^{\frac{n-k}{2}+\alpha-1}. \\ &\left\{\frac{1}{2}\boldsymbol{y}'H_{k\Delta}\boldsymbol{y}+\gamma+\delta\gamma\right\}^{-\frac{n+4\alpha-p}{2}}, \text{ if } 1\le k\le n-1. \qquad (3.2.54)\end{aligned}$$

3.2.2.1 EXACTLY ONE CHANGE-POINT IN THE MODEL

Assuming once again that we have established the existence of a structural change, we can now obtain probabilities of a change at $k=1,2,\ldots,n-1$. We can then analyze the behaviour of the other parameters of the model, given the change occurred at k where $1\le k\le n-1$.

From (3.2.49) it follows that, with $\Phi^{-1}\longrightarrow 0$,

$$\boldsymbol{\beta}|\boldsymbol{y},\sigma_1^2,\sigma_2^2,k \sim N(\widehat{\boldsymbol{\beta}},(X'\Sigma_k^{-1}X)^{-1})$$

and that

$$\begin{aligned}\boldsymbol{\beta}|y,k \quad &\propto \quad \left[\gamma+\frac{1}{2}(\boldsymbol{y}_{1k}-X_{1k}\boldsymbol{\beta})'(\boldsymbol{y}_{1k}-X_{1k}\boldsymbol{\beta}\right]^{-\frac{k+2\alpha}{2}}. \\ &\left[\gamma+\frac{1}{2}\left(\boldsymbol{y}_{2k}-X_{2k}\boldsymbol{\beta}\right)'\left(\boldsymbol{y}_{2k}-X_{2k}\boldsymbol{\beta}\right)\right]^{-\frac{n-k+2\alpha}{2}} \qquad (3.2.55)\end{aligned}$$

which is the product of two t-distributions.

Using (3.2.54), ignoring all constants and letting $\alpha, \gamma \to 0$, it follows that

$$\Pi(k|y, \delta) \propto |X'\Delta_k X|^{-\frac{1}{2}} \delta^{\frac{n-k}{2}-1} \{\boldsymbol{y}' H_{k\Delta} \boldsymbol{y}\}^{-\frac{n-p}{2}}. \tag{3.2.57}$$

3.2.3 A CHANGE IN THE REGRESSION COEFFICIENT AND THE VARIANCE

Consider the linear model where a change may have occurred,

$$y_i = \boldsymbol{x}_i \boldsymbol{\beta}_0 + e_i; \quad i = 1, \ldots, n$$

or

$$\begin{aligned} y_i &= \boldsymbol{x}_i \boldsymbol{\beta}_1 + e_i; \quad i = 1, \ldots, k, \\ y_i &= \boldsymbol{x}_i \boldsymbol{\beta}_2 + e_i; \quad i = k+1, \ldots, n. \end{aligned} \tag{3.2.58}$$

The x_i's and $\boldsymbol{\beta}_i$ are the same as previously, and e_i are i.i.d normal random variables with $e_i \sim N(0, \sigma_1^2)$ if $i = 1, \ldots, k$ and $e_i \sim N(0, \sigma_2^2)$ if $i = k+1, \ldots, n$ with $\sigma_1^2, \sigma_2^2 > 0$ and $p + 1 \leq k \leq n - p - 1$ with $n \geq 2p + 2$.

Furthermore if $k = n$, $\sigma^{-2} \sim \Gamma(\alpha_0, \gamma_0)$ and if $k \neq n$, $\sigma_1^{-2} \sim \Gamma(\alpha_1, \gamma_1)$ and $\sigma_2^{-2} \sim \Gamma(\alpha_2, \gamma_2)$.

Let the marginal prior mass function of k be as in (3.2.2) and let $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ be assigned a normal with

$$\boldsymbol{\beta}_0 | \sigma^2 \in I\!R^p \sim N(\boldsymbol{\theta}_0, \sigma^2 \Phi_0) \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 \Phi_{11} & 0 \\ 0 & \sigma_2^2 \Phi_{22} \end{pmatrix} \right)$$

so that

$$\Pi(\boldsymbol{\beta}_0|k=n,\Phi_0,\sigma^2) \propto |\Phi_0|^{-\frac{1}{2}}(2\pi\sigma^2)^{-\frac{p}{2}}\ e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}_0-\boldsymbol{\theta}_0)'\Phi_0^{-1}(\boldsymbol{\beta}_0-\boldsymbol{\theta}_0)}$$

and

$$\begin{aligned}\Pi(\boldsymbol{\beta}|k\neq n,\Phi_{11},\Phi_{22},\sigma_1^2,\sigma_2^2) \quad \propto \quad & |\Phi_{11}|^{-\frac{1}{2}}(2\pi\sigma_1^2)^{-\frac{p}{2}}\ e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}_1-\boldsymbol{\theta}_1)'\Phi_{11}^{-1}(\boldsymbol{\beta}_1-\boldsymbol{\theta}_1)} \\ & (2\pi\sigma_2^2)^{-\frac{p}{2}}\ e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}_2-\boldsymbol{\theta}_2)'\Phi_{22}^{-1}(\boldsymbol{\beta}_2-\boldsymbol{\theta}_2)}|\Phi_{22}|^{-\frac{1}{2}}\ .\end{aligned} \tag{3.2.59}$$

The marginal prior densities of σ^2 and of σ_1^2,σ_2^2 are then respectively

$$\Pi(\sigma^2|k=n,\gamma_0,\alpha_0) = \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)}\left(\frac{1}{\sigma^2}\right)^{\alpha_0+1} e^{-\frac{\gamma_0}{\sigma^2}}$$

and

$$\Pi(\sigma_1^2,\sigma_2^2|k\neq n,\gamma_1,\gamma_2,\alpha_1,\alpha_2) = \frac{\gamma_1^{\alpha_1}}{\Gamma(\alpha_1)}\left(\frac{1}{\sigma_1^2}\right)^{\alpha_1+1} e^{-\frac{\gamma_1}{\sigma_1^2}}\frac{\gamma_2^{\alpha_2}}{\Gamma(\alpha_2)}\left(\frac{1}{\sigma_2^2}\right)^{\alpha_2+1} e^{-\frac{\gamma_2}{\sigma_2^2}}. \tag{3.2.60}$$

With uniform priors on the $\boldsymbol{\theta}$'s, the marginal posterior mass function if $k=n$ is

$$\Pi(k=n|\boldsymbol{y},\gamma_0,\alpha_0) \propto q\frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)}|X'X|^{-\frac{1}{2}}\Gamma\left(\frac{2\alpha_0+p+n}{2}\right)\left[\gamma_0+\frac{1}{2}\boldsymbol{y}'H\boldsymbol{y}\right]^{-\left(\frac{2\alpha_0+p+n}{2}\right)} \tag{3.2.61}$$

where H is given by (3.2.47). The joint posterior distribution if $k\neq n$ according to Bayes' theorem is

$$
\begin{aligned}
\Pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1^2, \sigma_2^2, k|\boldsymbol{y}, k \neq n) \;\propto\; & \frac{1-q}{n-2p+1} |\Phi_{11}|^{-\frac{1}{2}} |\Phi_{22}|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}-p} \\
& \left(\frac{1}{\sigma_1^2}\right)^{\frac{p+k}{2}+\alpha_1+1} \left(\frac{1}{\sigma_2^2}\right)^{\frac{p+n-k}{2}+\alpha_2+1} \frac{\gamma_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\gamma_2^{\alpha_2}}{\Gamma(\alpha_2)}\, e^{-\frac{\gamma_1}{\sigma_1^2}} e^{-\frac{\gamma_2}{\sigma_2^2}} \\
& e^{-\frac{1}{2\sigma_1^2}(\boldsymbol{\beta}_1-\boldsymbol{\theta}_1)'\Phi_{11}^{-1}(\boldsymbol{\beta}_1-\boldsymbol{\theta}_1)}\; e^{-\left(\frac{1}{2\sigma_2^2}\right)(\boldsymbol{\beta}_2-\boldsymbol{\theta}_2)'\Phi_{22}^{-1}(\boldsymbol{\beta}_2-\boldsymbol{\theta}_2)} \\
& e^{-\frac{1}{2\sigma_1^2}[\boldsymbol{y}_{1k}-X_{1k}\boldsymbol{\beta}_1]'[\boldsymbol{y}_{1k}-X_{1k}\boldsymbol{\beta}_1]}\; e^{-\frac{1}{2\sigma_2^2}[\boldsymbol{y}_{2k}-X_{2k}\boldsymbol{\beta}_2]'[\boldsymbol{y}_{2k}-X_{2k}\boldsymbol{\beta}_2]}.
\end{aligned}
\tag{3.2.62}
$$

The marginal posterior mass function if $k \neq n$ is

$$
\begin{aligned}
\Pi(k \neq n|\boldsymbol{y}, \alpha_1, \alpha_2, \gamma_1, \gamma_2) \;\propto\; & \frac{1-q}{n-2p+1} \frac{\gamma^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\gamma_2^{\alpha_2}}{\Gamma(\alpha_2)} \left(\frac{2\alpha_1+p+k}{2}\right) \Gamma\left(\frac{2\alpha_2+p+n-k}{2}\right) \\
& |X_{1k}'X_{1k}|^{-\frac{1}{2}} |X_{2k}'X_{2k}|^{-\frac{1}{2}} \left[\gamma_1 + \frac{1}{2}\boldsymbol{y}_{1k}' H_1 \boldsymbol{y}_{1k}\right]^{-\left(\frac{2\alpha_1+p+k}{2}\right)} \\
& \left[\gamma_2 + \frac{1}{2}\boldsymbol{y}_{2k}' H_2 \boldsymbol{y}_{2k}\right]^{-\left[\frac{2\alpha_2+p+n-k}{2}\right]}
\end{aligned}
\tag{3.2.63}
$$

where

$$
H_i = I - X_{ik}'(X_{ik}' \quad X_{ik})^{-1} X_{ik} \quad \text{for} \quad i = 1, 2.
$$

3.2.3.1 EXACTLY ONE CHANGE-POINT IN THE MODEL

From (3.2.62) it follows that

$$\begin{aligned}\boldsymbol{\beta}_1|\gamma_1, \alpha_1 &\propto \int \left(\frac{1}{\sigma_1^2}\right)^{\frac{k}{2}+\alpha_1+1} e^{-\frac{\gamma_1}{\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}[X_{1k}\boldsymbol{\beta}_1 - \boldsymbol{y}_{1k}]'[X_{1k}\boldsymbol{\beta}_1 - \boldsymbol{y}_{1k}]} d\sigma_1^2 \\ &\propto \Big\{\gamma_1 + \frac{1}{2}\left(\boldsymbol{\beta}_1 - \left(X_{1k}'X_{1k}\right)^{-1}X_{1k}'X_{1k}(\boldsymbol{\beta}_1 - (X_{1k}'X_{1k})^{-1}X_{1k}'\boldsymbol{y}_{1k}\right)\Big) \\ &\quad + \boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}\Big\}^{-\frac{k}{2}+\alpha_1}\end{aligned}$$

which is a t-distribution with $k-p+2\alpha$ degrees of freedom, precision $\dfrac{k-p+2\alpha}{\gamma_1+\boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}} \cdot X_{1k}'X_{1k}$ and mean $(X_{1k}'X_{1k})^{-1}X_{1k}'\boldsymbol{y}_{1k}$.

Furthermore

$$\sigma_1^2 \sim IG\left(\frac{k-p}{2}+\alpha_1, \frac{1}{2}\boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}+\gamma_1\right)$$

and

$$\sigma_2^2 \sim IG\left(\frac{n-k-p}{2}+\alpha_2, \frac{1}{2}\boldsymbol{y}_{2k}'H_2\boldsymbol{y}_{2k}+\gamma_2\right). \tag{3.2.64}$$

If $\alpha_i, \gamma_i \to 0 (i=1,2)$ in (3.2.63), ignoring all constants, it follows that

$$\begin{aligned}\Pi(k|y) &\propto |X_k'X_k|^{-\frac{1}{2}}\Gamma\left(\frac{p+k}{2}\right)\Gamma\left(\frac{p+n-k}{2}\right)[\boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}]^{-\left(\frac{p+k}{2}\right)} \\ &\quad [\boldsymbol{y}_{2k}'H_2\boldsymbol{y}_{2k}[^{-\left(\frac{p+n-k}{2}\right)}.\end{aligned} \tag{3.2.65}$$

3.3 BAYES FACTORS

3.3.1 A CHANGE IN THE REGRESSION COEFFICIENT

Consider the models

$$M_k \quad : \quad \begin{cases} y_i = X_i\boldsymbol{\beta}_1 + e_i; & i = 1, \ldots, k \\ y_i = X_i\boldsymbol{\beta}_2 + e_i; & i = k+1, \ldots, n \end{cases} .$$

$$M_0 \quad : \quad y_i = X_i\boldsymbol{\beta}_0 + e_i; \quad i = 1, \ldots, n$$

with

$\boldsymbol{\beta}$ a $p \times 1$ vector of unknown parameters and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1' \; \boldsymbol{\beta}_2']$ and $e_i \sim N(0, \sigma^2 I_n)$. Also bring in the (Jeffrey's) priors $\Pi_0(\boldsymbol{\beta}_0, \sigma^2) \propto \frac{1}{\sigma^2}$ and $\Pi_1(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2) \propto \frac{1}{\sigma^2}$. Furthermore consider a minimal sample of size $\ell = r + s$ with r observations before k and s observations after k.

All parameters are identifiable and we have proper posteriors if $\ell = 2p+1$ and $r, s \geq p$. Let $r = p$, so that $s = p+1$. The marginal densities for the whole sample under model M_k will be

$$\begin{aligned} m_k^N(x) &= \iint \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}_k - X_k\boldsymbol{\beta})'(\boldsymbol{y}_k - X_k\boldsymbol{\beta})} \frac{1}{\sigma^2} d\boldsymbol{\beta} d\sigma^2 \\ &= \iint \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2\sigma^2}[\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_k]' X_k' X_k [\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_k]} e^{-\frac{1}{2\sigma^2}\boldsymbol{y}_k' H_k \boldsymbol{y}_k} d\boldsymbol{\beta} d\sigma^2 \end{aligned}$$

where $H_k = I_{2p} - X_k'(X_k'X_k)^{-1}X_k$ so that

$$\begin{aligned} m_k^N(x) &= \int (2\pi)^{-\left(\frac{n-2p}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{n-2p}{2}+1} e^{-\frac{1}{2\sigma^2}\boldsymbol{y}_k' H_k \boldsymbol{y}_k} |X_k'X_k|^{-\frac{1}{2}} |d\sigma^2 \\ &= \pi^{-\left(\frac{n-2p}{2}\right)} \Gamma\left(\frac{n-2p}{2}\right) [\boldsymbol{y}_k' H_k \boldsymbol{y}_k]^{-\left(\frac{n-2p}{2}\right)} |X_k'X_k|^{-\frac{1}{2}}. \qquad (3.3.1) \end{aligned}$$

For the minimal sample it follows directly that

$$m_k(\boldsymbol{x}(\ell)) = [\boldsymbol{y}_k'(\ell) H_k(\ell) \boldsymbol{y}_k'(\ell)]^{-\frac{1}{2}} \, |X_k'(\ell) X_k(\ell)|^{-\frac{1}{2}}$$

where $\boldsymbol{y}_k(\ell)$, $X_k(\ell)$ and $H_k(\ell)$ denote the values from the minimal sample of size $2p+1$, with p observations before or at k and $p+1$ after k.

Under model M_0 it follows that

$$\begin{aligned} m_0^N(\boldsymbol{x}) &= \int_0^\infty \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1-\frac{p}{2}} (2\pi)^{\frac{p}{2}} |\boldsymbol{X}_n' \boldsymbol{X}_n|^{-\frac{1}{2}} \; e^{-\frac{1}{2\sigma^2}[\boldsymbol{y}' H \boldsymbol{y}]} d\sigma^2 \\ &= \pi^{-\left(\frac{n-p}{2}\right)} |\boldsymbol{X}'\boldsymbol{X}|^{-\frac{1}{2}} \Gamma\left(\frac{n-p}{2}\right) [\boldsymbol{y}' H \boldsymbol{y}]^{-\frac{n-p}{2}} \end{aligned}$$

and

$$m_0(\boldsymbol{x}(\ell)) = \pi^{-\left(\frac{p+1}{2}\right)} \Gamma\left(\frac{p+1}{2}\right) [\boldsymbol{y}'(\ell) H(\ell) \boldsymbol{y}(\ell)]^{-\frac{p+1}{2}} |X'(\ell) X(\ell)|^{-\frac{1}{2}}. \tag{3.3.2}$$

Furthermore

$$B_{k0}^N(\boldsymbol{x}) = \frac{\pi^{\frac{p}{2}} \Gamma\left(\frac{n-2p}{2}\right) [\boldsymbol{y}_k' H_k \boldsymbol{y}_k]^{-\left(\frac{n-2p}{2}\right)} |X_k' X_k|^{-\frac{1}{2}}}{\Gamma\left(\frac{n-p}{2}\right) |X'X|^{-\frac{1}{2}} [\boldsymbol{y}' H \boldsymbol{y}]^{-\left(\frac{n-p}{2}\right)}} \tag{3.3.3}$$

and

$$B_{0k}^N(\boldsymbol{x}(\ell)) = \frac{\pi^{-\frac{p+1}{2}} \Gamma\left(\frac{p+1}{2}\right) [y'(\ell) H(\ell) y(\ell)]^{-\left(\frac{p+1}{2}\right)} |X'(\ell) X(\ell)|^{-\frac{1}{2}}}{[\boldsymbol{y}_k'(\ell) H_k(\ell) \boldsymbol{y}_k(\ell)]^{-\frac{1}{2}} |X_k'(\ell) X_k(\ell)|^{-\frac{1}{2}}} \tag{3.3.4}$$

so that

$$B_{k0}(\ell) = B_{k0}^N(\boldsymbol{x}) B_{0k}^N(\boldsymbol{x}(\ell))$$

and

$$B_{k0}^{AI} = B_{k0}^{N}(\boldsymbol{x}) \cdot \frac{1}{L} \sum_{\ell=1}^{L} B_{0k}^{N}(\boldsymbol{x}(\ell)). \tag{3.3.5}$$

Typically the Arithmetic Intrinsic Bayes factor does not satisfy the coherency condition that $B_{ij} = \frac{B_{i0}}{B_{j0}}$ as mentioned by Berger and Pericchi (1996). The encompassing model approach suggested by Berger and Pericchi is also not applicable in this change-point situation, as the only minimal sample satisfying all models is the whole sample. So we will define all Bayes factors between competing change-point models as relative to the no change models, i.e. let $B_{ij} = \frac{B_{i0}}{B_{j0}}$.

For the Fractional Bayes factor it follows that

$$m_k(b) = \frac{m_k}{m_k^b} = b^{\frac{nb}{2}} \pi^{-\frac{n(1-b)}{2}} \frac{\Gamma\left(\frac{n-2p}{2}\right)}{\Gamma\left(\frac{nb-2p}{2}\right)} \left(\boldsymbol{y}_k' H_k \boldsymbol{y}_k\right)^{-\frac{n(1-b)}{2}}, \tag{3.3.6}$$

$$m_0(b) = \frac{m_0}{m_0^b} = b^{\frac{nb}{2}} \pi^{-\frac{n(1-b)}{2}} \frac{\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{nb-p}{2}\right)} \left(\boldsymbol{y}' H \boldsymbol{y}\right)^{-\frac{n(1-b)}{2}} \tag{3.3.7}$$

and therefore, with $b = \frac{2p+1}{n}$, it follows that

$$B_{0k}^{F} = \frac{m_0(b)}{m_k(b)} = \frac{\pi^{\frac{1}{2}} \Gamma\left(\frac{n-p}{2}\right) \left[\boldsymbol{y}' H \boldsymbol{y}\right]^{-\left(\frac{n-2p-1}{2}\right)}}{\Gamma\left(\frac{p+1}{2}\right) \Gamma\left(\frac{n-2p}{2}\right) \left[\boldsymbol{y}_k' H_k \boldsymbol{y}_k\right]^{-\left(\frac{n-2p-1}{2}\right)}} \tag{3.3.8}$$

3.3.2 A CHANGE IN THE VARIANCE

For $1 \leq k \leq n-1$ and with the priors $\Pi(\beta) \propto 1$, $\Pi(\sigma_1^2) \propto \frac{1}{\sigma_1^2}$ and $\Pi(\sigma_2^2) \propto \frac{1}{\sigma_2^2}$ it follows that

$$\begin{aligned} m_k^N(x) &= \iiint \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} (\sigma_1^2)^{-\frac{k}{2}} (\sigma_2^2)^{-\frac{n-k}{2}} \frac{1}{\sigma_1^2} \frac{1}{\sigma_2^2}\, e^{-\frac{1}{2}(y - X\boldsymbol{\beta})' \Sigma_k^{-1} (y - X\boldsymbol{\beta})} d\boldsymbol{\beta} d\sigma_1^2 d\sigma_2^2 \\ &= \iint (2\pi)^{-\frac{n-p}{2}} (\sigma_1^2)^{-\frac{k}{2}-1} (\sigma_2^2)^{-\frac{n-k}{2}-1}\, e^{-\frac{1}{2}\left[\boldsymbol{y}_k' H_{\Sigma k} \boldsymbol{y}_k\right]} |X' \Sigma_k^{-1} X|^{-\frac{1}{2}} d\sigma_1^2 d\sigma_2^2 \end{aligned}$$

where Σ_k is defined above (3.2.41) and

$$H_{\Sigma k} = \Sigma_k^{-1} - \Sigma_k^{-1} X \left(X' \Sigma_k^{-1} X\right)^{-1} X' \Sigma_k^{-1}. \tag{3.3.9}$$

By letting $\sigma_2^2 = \dfrac{\sigma_1^2}{\delta}$ and $\pi(\delta) \propto \frac{1}{\delta}$, it follows that

$$\begin{aligned} m_k^N(x) &= \iint |X'\Delta_k X|^{-\frac{1}{2}} (2\pi)^{-\frac{n-p}{2}} \delta^{\frac{n-k}{2}-1} (\sigma_1^2)^{-\left(\frac{n-p}{2}\right)-1} e^{-\frac{1}{2\sigma_1^2}[\boldsymbol{y}_k' H_{k\Delta} \boldsymbol{y}_k]} d\sigma_1^2 d\delta \\ &= \int (\pi)^{-\left(\frac{n-p}{2}\right)} \Gamma\left(\frac{n-p}{2}\right) \delta^{\frac{n-k}{2}-1} |X'\Delta_k X|^{-\frac{1}{2}} [\boldsymbol{y}' H_{k\Delta} \boldsymbol{y}]^{-\frac{n-p}{2}} d\delta. \end{aligned} \tag{3.3.10}$$

Furthermore

$$B_{k0}^N(\boldsymbol{x}) = \int \frac{\delta^{\left(\frac{n-k}{2}\right)-1} |X'\Delta_k X|^{-\frac{1}{2}} [\boldsymbol{y}' H_{k\Delta} \boldsymbol{y}]^{-\left(\frac{n-p}{2}\right)} d\delta}{|X'X|^{-\frac{1}{2}} [\boldsymbol{y}' H \boldsymbol{y}]^{-\left(\frac{n-p}{2}\right)}}. \tag{3.3.11}$$

For the minimal sample $n = p + 1$ it follows that

$$m_k(\boldsymbol{x}(\ell)) = \int |X' D_\delta^{-1}(\ell) X|^{-\frac{1}{2}} \delta^{\frac{p+1-k}{2}-1} [\boldsymbol{y}'(\ell) H_D(\ell) \boldsymbol{y}(\ell)]^{-\frac{1}{2}} d\delta \tag{3.3.12}$$

where $D_\delta(p)$ and $H_D(p)$ are defined for the minimal sample with at least one observation on either side of k.

Then

$$B_{0k}^N(x(\ell)) = \frac{\int |X' D_\delta(\ell) X|^{-\frac{1}{2}} \delta^{\frac{p+1-k}{2}-1} [\boldsymbol{y}'(\ell) H_D(\ell) \boldsymbol{y}(\ell)]^{-\frac{1}{2}} d\delta}{|X'(\ell) X(\ell)|^{-\frac{1}{2}} |\boldsymbol{y}'(\ell) H(\ell) \boldsymbol{y}(\ell)|^{-\frac{1}{2}}}.$$

The Arithmetic Intrinsic Bayes factor follows as in (3.3.6).

For the Fractional Bayes factor

$$m_k(b) = \frac{\int |X'\Delta_k^{-1}X|^{-\frac{1}{2}}[\boldsymbol{y}'H_{k\Delta}\boldsymbol{y}]^{-\frac{n-p}{2}}\delta^{\frac{n-k}{2}-1}d\delta}{\int [\boldsymbol{y}'H_{k\Delta}\boldsymbol{y}]^{-\frac{nb-p}{2}}|X'\Delta_k^{-1}X|^{-\frac{1}{2}}\delta^{\frac{b(n-k)}{2}-1}d\delta} \tag{3.3.13}$$

and

$$m_0(b) = [\boldsymbol{y}'H\boldsymbol{y}]^{-\frac{n(1-b)}{2}} \tag{3.3.14}$$

so that

$$B_{0k}^F = \frac{m_0(b)}{m_k(b)} = \frac{[\boldsymbol{y}'H\boldsymbol{y}]^{-\frac{n(1-b)}{2}}\int [\boldsymbol{y}'H_{k\Delta}\boldsymbol{y}]^{-\frac{nb-p}{2}}|X'\Delta_k^{-1}X|^{-\frac{1}{2}}\delta^{\frac{b(n-k)}{2}-1}d\delta}{\int |X'\Delta_k^{-1}X|^{-\frac{1}{2}}[\boldsymbol{y}'H_{k\Delta}\boldsymbol{y}]^{-\frac{n-p}{2}}\delta^{\frac{n-k}{2}-1}d\delta} \tag{3.3.15}$$

where we can take $b = \frac{p+1}{n}$.

3.3.3 A CHANGE IN THE REGRESSION COEFFICIENT AND THE VARIANCE

Consider the same model as in (3.2.58) with the vague priors $\Pi(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2) \propto \frac{1}{\sigma_1^2\sigma_2^2}$ so that

$$\begin{aligned} m_k^N(x) &= \iiiint (2\pi)^{-\frac{n}{2}}(\sigma_1^2)^{-\frac{k}{2}-1}(\sigma_2^2)^{-\frac{n-k}{2}-1}\, e^{-\frac{1}{2\sigma_1^2}(\boldsymbol{y}_{1k}-X_{1k}\boldsymbol{\beta}_1)'(\boldsymbol{y}_{1k}-X_{1k}\boldsymbol{\beta}_1)} \\ &\quad e^{-\frac{1}{2\sigma_2^2}(\boldsymbol{y}_{2k}-X_{2k}\boldsymbol{\beta}_2)'(\boldsymbol{y}_{2k}-X_{2k}\boldsymbol{\beta}_2)} d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_2 d\sigma_1^2 d\sigma_2^2 \\ &= \pi^{-\frac{n-2p}{2}}|X_k'X_k|^{-\frac{1}{2}}\Gamma\left(\frac{k-p}{2}\right)\Gamma\left(\frac{n-k-p}{2}\right)[\boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}]^{-\frac{k-p}{2}}[\boldsymbol{y}_{2k}'H_2\boldsymbol{y}_{2k}]^{-\frac{n-k-p}{2}} \end{aligned} \tag{3.3.16}$$

where $\quad H_i = I - X_{ik}'(X_{ik}'X_{ik})X_{ik}, \quad i = 1, 2.$ (3.3.17) and where $X_k, \boldsymbol{y}_{ik}, X_{ik}$ are as defined in section (3.2.1).

With the minimal sample $n = 2p+2$ it follows that $k = n-k = p+1$ and

$$m_k(x(\ell)) = |X_k'(\ell)X_k(\ell)|^{-\frac{1}{2}}[\boldsymbol{y}_{1k}'(\ell)H_1(\ell)\boldsymbol{y}_{1k}(\ell)]^{-\frac{1}{2}}[\boldsymbol{y}_{2k}'(\ell)H_2(\ell)\boldsymbol{y}_{2k}(\ell)]^{-\frac{1}{2}} \tag{3.3.18}$$

where $X_k(\ell)$, etc. are defined as before with $p+1$ observations on either side of k.

Furthermore

$$B_{0k}^N(x) = \frac{\pi^{-\frac{p}{2}}|X'X|^{-\frac{1}{2}}\Gamma\left(\dfrac{n-p}{2}\right)[\boldsymbol{y}'H\boldsymbol{y}]^{-\frac{n-p}{2}}}{|X_k'X_k|^{-\frac{1}{2}}\Gamma\left(\frac{k-p}{2}\right)\Gamma\left(\frac{n-k-p}{2}\right)[\boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}]^{-\frac{k-p}{2}}[\boldsymbol{y}_{2k}'H_2\boldsymbol{y}_{2k}]^{-\left(\frac{n-k-p}{2}\right)}} \tag{3.3.19}$$

and

$$B_{0k}^N(\boldsymbol{x}(\ell)) = \frac{\pi^{-\frac{p+2}{2}}\Gamma\left(\frac{p+2}{2}\right)[\boldsymbol{y}'(\ell)H(\ell)\boldsymbol{y}(\ell)]^{-\frac{p+2}{2}}|X'(\ell)X(\ell)|^{-\frac{1}{2}}}{[\boldsymbol{y}_{1k}'(\ell)H_1(\ell)\boldsymbol{y}_{1k}(\ell)]^{-\frac{1}{2}}[\boldsymbol{y}_{2k}'(\ell)H_2(\ell)\boldsymbol{y}_{2k}(\ell)]^{-\frac{1}{2}}}. \tag{3.3.20}$$

For the Fractional Bayes factor, with $nb = 2p+2$ and $kb = p+1 = b(n-k)$ $m_0(b)$ is given as

$$m_0(b) = \frac{\pi^{-\left(\frac{n-2p}{2}\right)+1}\Gamma\left(\frac{n-p}{2}\right)[\boldsymbol{y}'H\boldsymbol{y}]^{-\left(\frac{n-2p}{2}\right)+1}}{b^{-\frac{nb}{2}}\Gamma\left(\frac{p+2}{2}\right)}$$

and

$$m_k(b) = b^{\frac{nb}{2}}\pi^{-\left(\frac{n-2p}{2}\right)}\Gamma\left(\frac{k-p}{2}\right)\Gamma\left(\frac{n-k-p}{2}\right)[\boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}]^{-\frac{k-p-1}{2}}[\boldsymbol{y}_{2k}'H_2\boldsymbol{y}_{2k}]^{-\frac{n-k-p-1}{2}} \tag{3.3.21}$$

so that

$$B_{0k}^F = \frac{m_0(b)}{m_k(b)} = \frac{\pi\Gamma\left(\frac{n-p}{2}\right)(\boldsymbol{y}'H\boldsymbol{y})^{-\frac{n-2p}{2}+1}}{\Gamma\left(\frac{p+2}{2}\right)\Gamma\left(\frac{k-p}{2}\right)\Gamma\left(\frac{n-k-p}{2}\right)[\boldsymbol{y}_{1k}'H_1\boldsymbol{y}_{1k}]^{-\frac{k-p-1}{2}}[\boldsymbol{y}_{2k}'H_2\boldsymbol{y}_{2k}]^{-\left(\frac{n-k-p-1}{2}\right)}}. \tag{3.3.22}$$

As in equation (2.5.15) of chapter 2 the posterior probabilities follow for all models.

3.4 MULTIPLE CHANGE-POINTS

Results in this chapter can readily be extended to more than one change-point. Assuming r change-points in the regression coefficient vector

$$y_i \sim \begin{cases} \boldsymbol{x}_i\boldsymbol{\beta}_1 + e_i; & i = 1, \ldots, k_1 \\ \boldsymbol{x}_i\boldsymbol{\beta}_2 + e_i; & i = k_1 + 1, \ldots, k_2 \\ \vdots & \\ x_i\boldsymbol{\beta}_{r+1} + e_i; & i = k_r + 1, \ldots, n \end{cases} \tag{3.4.1}$$

where $p \leq k_1 < k_2 < \cdots < k_r \leq n - p$ and $k_{i+1} - k_i \geq p$ for $i = 1, \ldots, r - 1$, with the strict inequality holding at least once. So $n \geq p(r+1) + 1$. As in section (2.6.1), the marginal prior density of $\boldsymbol{k} = [k_1, \ldots, k_r]$ is uniform over all possible permutations.

As in section (3.2.1), let $\boldsymbol{\beta}' = [\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_{r+1}'] \in I\!R^{(r+1)p}$ and $\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\theta}, \sigma^2\Phi)$ where $\boldsymbol{\theta} = I_{r+1} \otimes \boldsymbol{\theta}_0$, $\boldsymbol{\theta}_0 \in I\!R^p$ and $\Phi = I_{r+1} \otimes \Phi_{11}$. Further as in (3.2.5) let $\sigma^2 \sim IG(\alpha, \gamma)$.

Furthermore let

$$\boldsymbol{y}_k = \begin{bmatrix} \boldsymbol{y}_{1k} \\ \boldsymbol{y}_{2k} \\ \vdots \\ \boldsymbol{y}_{(r+1)k} \end{bmatrix} \quad \text{where} \quad \boldsymbol{y}_{ik} = \begin{bmatrix} y_{k_{i-1}+1} \\ \vdots \\ y_{k_i} \end{bmatrix}$$

and $X_k = \text{diag}\{X_{ik}\} : n \times (r+1)p$ where $X_{ik} = \begin{bmatrix} x_{k_{i-1}+1} \\ \vdots \\ x_{k_i} \end{bmatrix} : (k_i - k_{i-1}) \times p.$

$$(3.4.2)$$

With the above notation, the joint posterior of all the parameters is exactly as in (3.2.15) and the marginal posterior of k conditional on Φ, α, γ is given in (3.2.28), where $J_p = [I_p \ldots I_p] : p \times (r+1)p$.

The marginals of $\boldsymbol{\beta}$ and σ^2 are given in (3.2.34) and (3.2.36) respectively, given $\boldsymbol{k}$.

If r is unknown, let the marginal prior density of r be uniform, i.e. $\Pi(r) = \frac{1}{R+1}$, $r = 0, \ldots, R$ and $\boldsymbol{k}^{(r)} = [k_1, \ldots, k_r]$ where $\boldsymbol{k}^{(r)}|r$ is uniformly distributed over all possible permutations for any given r.

The posterior distribution of r is then given by

$$\Pi(r|\Phi, \alpha, \gamma, \boldsymbol{\theta}, \boldsymbol{y}) \propto \begin{cases} \Pi(k = n|\Phi_{11}, \alpha, \gamma, \boldsymbol{\theta}_o, \boldsymbol{y}) & \text{for} \quad r = 0 \\ \sum_k \Pi(k|r = 1, \Phi, \alpha, \gamma, \boldsymbol{\theta}, \boldsymbol{y}) & for \quad r = 1 \\ \vdots & \\ \sum_{\boldsymbol{k}} \Pi(k|r = R, \Phi, \alpha, \gamma, \boldsymbol{\theta}, \boldsymbol{y}) & \text{for} \quad r = R \end{cases} \tag{3.4.3}$$

where $\Pi(k = n|\Phi_{11}, \alpha, \gamma, \boldsymbol{\theta}_0, \boldsymbol{y})$ is given by (3.2.16) with q replaced by $\frac{1}{R+1}$ and $\Pi(k|r, \Phi, \alpha, \gamma, \boldsymbol{\theta}, \boldsymbol{y})$ is given by (3.2.17) with $\frac{1-q}{n-2p+1}$ replaced by $\frac{1}{R+1}$ times one over the number of permutations of $\boldsymbol{k}$ given r.

The marginal prior density of σ^2 and σ_1^2, σ_2^2 is given by (3.2.60), while the marginal posterior mass function if $k \neq n$ is given by (3.2.63). With $\alpha_i, \gamma_i \to 0$ the posterior distribution of $k \neq n|y$ is given by (3.2.65).

The Fractional Bayes Factors (FBF) of O'Hagan (1995, 1997) for model m_0 against M_r^{kr} from data $\boldsymbol{y}$ is denoted by B_{0r}^F where

$$B_{0r}^F = \frac{m_0(b)}{m_r(b)}, \tag{3.4.4}$$

where b is the training fraction of the likelihood and

$$m_r(b) = \frac{\int f(\boldsymbol{y}|\boldsymbol{\beta}_r, \sigma^2, r, \boldsymbol{k}_r)\Pi(\boldsymbol{\beta}_r)\Pi(\sigma^2)d\boldsymbol{\beta}_r d\sigma^2}{\int [f(\boldsymbol{y}|\boldsymbol{\beta}_r, \sigma^2, r, \boldsymbol{k}_r]^b\Pi(\boldsymbol{\beta}_r)\Pi(\sigma^2)d\boldsymbol{\beta}_r d\sigma^2}.$$

We now have the following prior distributions:

$\Pi(\boldsymbol{\beta}_r) \propto 1$ and $\Pi(\sigma^2) \propto \frac{1}{\sigma^2}$.

For the FBF (a change in the regression coefficient), $m_0(b)$ is given by equation (3.3.7), while for $b = \frac{(R+1)p+1}{n}$

$$m_r(b) = b^{\frac{nb}{2}} \pi^{-\frac{n(1-b)}{2}} \frac{\Gamma\left(\frac{n-(r+1)p}{2}\right)}{\Gamma\left(\frac{nb-(r+1)p}{2}\right)} (y' H_k y)^{-\frac{n(1-b)}{2}}$$

so that

$$B_{r0}^F = \frac{m_r(b)}{m_0(b)}. \tag{3.4.5}$$

Berger and Pericchi (1995, 1996, 1997) proposed using all possible minimal training samples and averaging the resulting Bayes factors. The Arithmetic Intrinsic Bayes Factor for model M_s against model M_r is defined by $\mathrm{AIB}_{sr}^{k_s,k_r} = B_{sr}^{k_s,k_r}(\boldsymbol{y}) \cdot \frac{1}{L} \sum_{\ell=1}^{L} B_{rs}^{k_r,k_s}(\boldsymbol{y}(\ell))$, where $B_{sr}^{k_s,k_r}(\boldsymbol{y})$ is the usual Bayes factor with the whole sample and improper priors and $y(\ell)$ represents a minimal training sample. The geometric and median IBF are defined similarly.

Suppose we use the same vague prior distributions as before, then

$$B_{sr}^{k_s,k_r}(\boldsymbol{y}) = m_s^{k_s}(\boldsymbol{y})/m_r^{k_r}(\boldsymbol{y}). \tag{3.4.6}$$

For a fixed number of change-points the minimal sample size for a particular partition $\boldsymbol{k}_r$ is $p(r+1)+1$. Let χ_{k_r} be the set of minimal samples for model $M_r^{k_r}$, then the set of minimal samples for comparing $M_r^{k_r}$ and $M_r^{k_r'}$ is $\chi_{k_r} \cap \chi_{k_r'}$. The computation of this set is complex and the only minimal sample valid for all possible models is the whole data set.

So we will only apply the IBF for comparing model m_0 with $M_r^{k_r}$ with r fixed and define

$$B_{0r}^{k_r} = \frac{1}{B_{r0}^{k_r}} \text{ and } B_{sr}^{k_s,k_r} = \frac{B_{s0}^{k_s}}{B_{r0}^{k_r}}. \tag{3.4.7}$$

3.5 COMPONENT ANALYSIS

As in section 2.4, we can analize the effect of individual components on the position of the estimated change-point. Considering the case of a single change in the regression coefficients with constant variance as in paragraph 3.2.1.1., the posterior distribution of $\boldsymbol{\beta}' = [\boldsymbol{\beta}_2' \;\; \boldsymbol{\beta}_2']$ is given below (3.2.38) for fixed k, i.e.

$$\boldsymbol{\beta}|\boldsymbol{y}, k \sim t\left(n-2p, \widehat{\boldsymbol{\beta}}_k, \left(\frac{(n-2p)X_k'X_k}{\boldsymbol{y}_k' H_k \boldsymbol{y}_k}\right)\right).$$

Let $\boldsymbol{\Delta} = C\boldsymbol{\beta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$, where $C(p \times 2p) = [I_p \;\; -I_p]$, then $\boldsymbol{\Delta}$ represents the differences between corresponding components of the regression parameter vector before and after the change-point. The distribution of $\boldsymbol{\Delta}$ follows then directly from above as

$$\begin{aligned}
\boldsymbol{\Delta}|\boldsymbol{y},k &\sim t\left(n-2p, C\widehat{\boldsymbol{\beta}}_k, \frac{(n-2p)CX_k'X_kC'}{\boldsymbol{y}_k'H_k\boldsymbol{y}_k}\right) \\
&\sim t\left(n-2p, \widehat{\boldsymbol{\beta}}_1-\widehat{\boldsymbol{\beta}}_2, \left(\frac{(n-2p)X'X}{\boldsymbol{y}_k'H_k\boldsymbol{y}_k}\right)\right) . \qquad (3.5.1)
\end{aligned}$$

The marginal posterior of the i-th component of $\boldsymbol{\Delta}$ is then

$$\Delta_i|y,k \sim t\left(n-2p, \widehat{\beta}_{1i}-\widehat{\beta}_{2i}, \left(\frac{(n-2p)(X'X)_{ii}}{\boldsymbol{y}_k'H_k\boldsymbol{y}_k}\right)\right), \quad i=0,1,\ldots,p-1 \qquad (3.5.2)$$

where

$$\widehat{\boldsymbol{\beta}}_j = (X_{jk}'X_{jk})^{-1}X_{jk}'\boldsymbol{y}_{jk}, \quad j=1,2.$$

As in (2.4.9) and (2.4.10), the influence of each component on a specific change-point can be compared by standardizing as

$$\begin{aligned}
D_i &= E(\Delta_i)(\mathrm{Var}(\Delta_i))^{-\frac{1}{2}} \\
&= (\widehat{\beta}_{1i}-\widehat{\beta}_{2i})\left(\frac{\boldsymbol{y}_k'H_k\boldsymbol{y}_k(X'X)_{ii}^{-1}}{n-2p-2}\right)^{-\frac{1}{2}}, \quad i=0,\ldots,p-1 \qquad (3.5.3)
\end{aligned}$$

with

$$I_i = \frac{D_i}{\sum_{i=0}^{p-1} D_i}. \qquad (3.5.4)$$

For a change in mean and variance, comparisons between regression components can be done similarly, but to find the influence of the variance alone on the change-point as compared to the mean, the multiple models method of section 2.7.4 can be used for the linear model with fixed k.

In his paper, Groenewald (1993) derives posterior probabilities of change-points for three special cases:

(a) Multiple change-points in the parameter vector $\boldsymbol{\beta}$ or α under homogeneity of error variance,

(b) changes in $\boldsymbol{\beta}$, excluding α, under homogeneity of error variance and

(c) distinguishing between changes in $\boldsymbol{\beta}$ alone, α alone or changes in both.

Writing the model again as $y = \alpha \boldsymbol{e} + X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\beta}$ is $(p \times 1)$, Groenewald (1993) is interested in distinguishing between changes in $\boldsymbol{\beta}$ alone, α alone or both. Here, due to the notational complexity, only one possible change-point is considered, but the same arguments can be extended to any number of change-points. The error variance is assumed constant.

3.6 SWITCHPOINT (CONTINUOUS CHANGE-POINT)

3.6.1 EXACTLY ONE SWITCHPOINT

Let $\boldsymbol{y}(n \times 1) \sim N(X_k\boldsymbol{\beta}_k, \sigma^2 I_n)$ where

$$\underset{(4\times 1)}{\boldsymbol{\beta}_k} = \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{pmatrix}, \quad \underset{(n\times 4)}{X_k} = \begin{pmatrix} X_{1k} & 0 \\ 0 & X_{2k} \end{pmatrix},$$

$$\underset{(k\times 2)}{X_{1k}} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_k \end{pmatrix}, \quad \underset{((n-k)\times 2)}{X_{2k}} = \begin{pmatrix} 1 & x_{k+1} \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}. \qquad (3.6.1)$$

We assume that $x_1 < x_2 < \cdots < x_n$ with switchpoint at x_0, where $x_k \leq x_0 \leq x_{k+1}$, $k = 2, \ldots, n-2$, so k is the largest integer for which $x_k \leq x_0$.

Further, we must have, at the switchpoint, that

$$\alpha_1 + \beta_1 x_0 = \alpha_2 + \beta_2 x_0.$$

So

$$\begin{aligned} \alpha_2 &= \alpha_1 + (\beta_1 - \beta_2)x_0 \\ &= \alpha_1 - \tau x_0, \quad \text{where } \tau = \beta_2 - \beta_1. \end{aligned}$$

Thus

$$\boldsymbol{\beta}_k = \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \alpha_1 - \tau x_0 \\ \beta_1 + \tau \end{pmatrix} = A\boldsymbol{\beta} + J_0 \tau$$

where

$$A = \begin{pmatrix} I_2 \\ I_2 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix}, \quad J_0 = \begin{pmatrix} 0 \\ 0 \\ -x_0 \\ 1 \end{pmatrix}. \tag{3.6.2}$$

For vague priors, let

$$\Pi(\beta) \propto 1, \quad \Pi(\tau) \propto 1, \quad \Pi(\sigma^2) \propto \frac{1}{\sigma^2}, \quad x_0 \sim U(x_2, x_{n-1}), \qquad \text{i.e.}$$

$$x_0 | k \sim U(x_k, x_{k+1}), \quad \Pi(k) = \frac{x_{k+1} - x_k}{x_{n-1} - x_2}, \quad k = 2, \ldots, n-2,$$

i.e. proportional to the length of the interval. (3.6.3)

The likelihood function is

$$f(\boldsymbol{y}|\boldsymbol{\beta},\tau,\sigma^2,x_0(\text{or } k)) = (2\pi\sigma^2)^{-\frac{n}{2}}\ e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-X_k\boldsymbol{\beta}_k)'(\boldsymbol{y}-X_k\boldsymbol{\beta}_k)}$$

where

$$X_k\boldsymbol{\beta}_k = X\boldsymbol{\beta} + \tilde{\boldsymbol{x}}_k\tau, \qquad \underset{(n\times 2)}{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},$$

$$\underset{(n\times 1)}{\tilde{\boldsymbol{x}}_k} = \begin{pmatrix} \mathbf{0} \\ -x_0 1_{n-k} + \boldsymbol{x}_{2k} \end{pmatrix}^{k\times 1}_{(n-k)\times 1}, \quad x_{2k} = \begin{pmatrix} x_{k+1} \\ \vdots \\ x_n \end{pmatrix}. \tag{3.6.4}$$

Then

$$\begin{aligned}(\boldsymbol{y} - X_k\boldsymbol{\beta}_k)'(\boldsymbol{y} - X_k\boldsymbol{\beta}_k) &= (\boldsymbol{\beta} - (X'X)^{-1}X'(\boldsymbol{y} - \tau\tilde{\boldsymbol{x}}_k))'X'X(\boldsymbol{\beta} - (X'X)^{-1}X' \\ &\quad (\boldsymbol{y} - \tau\tilde{\boldsymbol{x}}_k)) + (\boldsymbol{y} - \tau\tilde{\boldsymbol{x}}_k)'H(\boldsymbol{y} - \tau\tilde{\boldsymbol{x}}_k)\end{aligned}$$

where

$$H = I - X(X'X)^{-1}X'.$$

So

$$f(\boldsymbol{y}|\tau,\sigma^2,x_0) = (2\pi\sigma^2)^{-\frac{n-2}{2}}|X'X|^{-\frac{1}{2}}\ e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\tau\tilde{\boldsymbol{x}}_k)'\,H\,(\boldsymbol{y}-\tau\tilde{\boldsymbol{x}}_k)}$$

and then

$$\begin{aligned}(\boldsymbol{y} - \tau\tilde{\boldsymbol{x}}_k)'H(\boldsymbol{y} - \tau\tilde{\boldsymbol{x}}_k) &= \tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k(\tau - (\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-1}\boldsymbol{y}'H\tilde{\boldsymbol{x}}_k)^2 \\ &\quad + \boldsymbol{y}'H\boldsymbol{y} - (\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-1}(\boldsymbol{y}'H\tilde{\boldsymbol{x}}_k)^2\end{aligned}$$

so that

$$
\begin{aligned}
f(\boldsymbol{y}|\sigma^2, x_0) &= (2\pi\sigma^2)^{-\left(\frac{n-3}{2}\right)}(\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}}|X'X|^{-\frac{1}{2}} \\
& \quad e^{-\frac{1}{2\sigma^2}\left[\boldsymbol{y}'H\boldsymbol{y}-\boldsymbol{y}'H\tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k H\boldsymbol{y}\right]}
\end{aligned} \tag{3.6.5}
$$

and

$$
\begin{aligned}
f(\boldsymbol{y}|x_0, k) &= \pi^{-\frac{n-3}{2}}(\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}}|X'X|^{-\frac{1}{2}}\Gamma\left(\frac{n-3}{2}\right) \\
& \quad [\boldsymbol{y}'(H - H\tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k H)\boldsymbol{y}]^{-\frac{n-3}{2}}.
\end{aligned} \tag{3.6.6}
$$

As

$$
\Pi(x_0|\boldsymbol{y}, k) \propto f(\boldsymbol{y}|x_0, k)\Pi(x_0|k)
$$

it follows that

$$
\Pi(x_0|\boldsymbol{y}, k) \propto (\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}}\left[\boldsymbol{y}'H(H^{-1} - \tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k')H\boldsymbol{y}\right]^{-\frac{n-3}{2}} \tag{3.6.7}
$$

with

$x_k \leq x_0 \leq x_{k+1}$.

3.6.2 NO OR ONE SWITCHPOINT

Under the assumption of a switchpoint, model M_{x_0}, the marginal likelihood of the data is given by

$$
\begin{aligned}
m_{x_0} &= \int (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}- X_k\boldsymbol{\beta}_k)'(\boldsymbol{y} - X_k\boldsymbol{\beta}_k)} d\boldsymbol{\beta} d\tau d\sigma^2 dx_0 \\
&= \int (2\pi)^{-\left(\frac{n-2}{2}\right)}(\sigma^2)^{-\frac{n}{2}}|X'X|^{-\frac{1}{2}}\, e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}- \tau\tilde{\boldsymbol{x}}_k)'H(\boldsymbol{y}- \tau\tilde{\boldsymbol{x}}_k)} d\tau d\sigma^2 dx_0 \\
&= \int (2\pi)^{-\left(\frac{n-3}{2}\right)}(\sigma^2)^{-\frac{n-1}{2}}|X'X|^{-\frac{1}{2}}(\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}} \\
&\quad e^{-\frac{1}{2\sigma^2}\left[\boldsymbol{y}'H\left(H^{-1}-\tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k'\right)H\boldsymbol{y}\right]} d\sigma^2 dx_0 \\
&= \Gamma\left(\frac{n-3}{2}\right)|X'X|^{-\frac{1}{2}}\pi^{-\left(\frac{n-3}{2}\right)}\sum_k \int_{xk}^{x_k+1} (\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}}\,\cdot \\
&\quad [\boldsymbol{y}'H(H^{-1} - \tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k')H\boldsymbol{y}]^{-\frac{n-3}{2}} dx_0 \qquad (3.6.8)
\end{aligned}
$$

and the fractional marginal likelihood is

$$
\begin{aligned}
m_{x_0}^b &= \int (2\pi)^{-\frac{nb}{2}}(\sigma^2)^{-\frac{nb+2}{2}}\, e^{-\frac{b}{2\sigma^2}(\boldsymbol{y}- X_k\boldsymbol{\beta}_k)'(\boldsymbol{y}- X_k\boldsymbol{\beta}_k)} d\boldsymbol{\beta} d\tau d\sigma^2 dx_0 \\
&= \int (2\pi)^{-\frac{nb-3}{2}}(\sigma^2)^{-\left(\frac{nb-1}{2}\right)} b^{-\frac{3}{2}}|X'X|^{-\frac{1}{2}}\,(\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}}\,\cdot \\
&\quad e^{-\frac{b}{2\sigma^2}\left[\boldsymbol{y}'H\left(H^{-1}-\tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k'\right)H\boldsymbol{y}\right]} d\sigma^2 dx_0 \\
&= \Gamma\left(\frac{nb-3}{2}\right)\pi^{-\left(\frac{nb-3}{2}\right)} b^{-\frac{3}{2}}|X'X|^{-\frac{1}{2}}\sum_k \int_{x_k}^{x_{k+1}} (\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}} \\
&\quad [\boldsymbol{y}'H(H^{-1} - \tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k')H\boldsymbol{y}]^{-\left(\frac{nb-3}{2}\right)} b^{-\left(\frac{nb-3}{2}\right)} dx_0 \\
&= \Gamma\left(\frac{nb-3}{2}\right)|X'X|^{-\frac{1}{2}}\pi^{-\left(\frac{nb-3}{2}\right)} b^{-\frac{nb}{2}}\sum_k \int_{x_k}^{x_{k_1}} (\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}} \\
&\quad [\boldsymbol{y}'H(H^{-1} - \tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k'H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k')H\boldsymbol{y}]^{-\frac{nb-3}{2}} dx_0.
\end{aligned}
$$

(3.6.9)

It follows that

$$m_{x_0}(b) = \Gamma\left(\frac{n-3}{2}\right)\pi^{-\frac{n-3}{2}}b^2 \cdot \frac{\sum_k \int_{x_k}^{x_{k+1}} (\tilde{x}_k' H \tilde{x}_k)^{-\frac{1}{2}} [y'H(H^{-1} - \tilde{x}_k(\tilde{x}_k' H \tilde{x}_k)^{-1}\tilde{x}_k') Hy]^{-\frac{n-3}{2}} dx_0}{\sum_k \int_{x_k}^{x_{k+1}} (\tilde{x}_k' H \tilde{x}_k)^{-\frac{1}{2}} [y'H(H^{-1} - \tilde{x}_k(\tilde{x}_k' H \tilde{x}_k)^{-1}\tilde{x}_k') Hy]^{-\frac{1}{2}} dx_0}. \tag{3.6.10}$$

Under no change, model M_0,

$$\begin{aligned} m_0 &= \int (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n+2}{2}} e^{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)} d\beta d\sigma^2 \\ &= \int (2\pi)^{-\frac{n-2}{2}} (\sigma^2)^{-\frac{n}{2}} |X'X|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2} y'Hy} d\sigma^2 \\ &= \Gamma\left(\frac{n-2}{2}\right) \pi^{-\frac{n-2}{2}} |X'X|^{-\frac{1}{2}} [y'Hy]^{-\frac{n-2}{2}} \end{aligned}$$

and

$$\begin{aligned} m_0^b &= \int (2\pi)^{-\frac{nb-2}{2}} (\sigma^2)^{-\frac{nb}{2}} b^{-1} |X'X|^{-\frac{1}{2}} e^{-\frac{b}{2\sigma^2} y'Hy} d\sigma^2 \\ &= \Gamma\left(\frac{nb-2}{2}\right) \pi^{-\frac{nb-2}{2}} |X'X|^{-\frac{1}{2}} b^{-\frac{nb}{2}} [y'Hy]^{-\frac{nb-2}{2}} \end{aligned}$$

so that

$$m_0(b) = \Gamma\left(\frac{n-2}{2}\right) \pi^{-\frac{n-4}{2}} b^2 [y'Hy]^{-\left(\frac{n-4}{2}\right)} \tag{3.6.11}$$

and the Fractional Bayes factor in favour of no change, with $b = \frac{4}{n}$, is

$$B_{oi} = \frac{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)[\boldsymbol{y}'H\boldsymbol{y}]^{-\frac{n-4}{2}}}{\Gamma\left(\frac{n-3}{2}\right)\frac{\sum_k Q_{n-3}(k)}{\sum_k Q_1(k)}} \tag{3.6.12}$$

where

$$Q_a(k) = \int_{x_k}^{x_{k+1}} (\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-\frac{1}{2}}[\boldsymbol{y}'H(H^{-1} - \tilde{\boldsymbol{x}}_k(\tilde{\boldsymbol{x}}_k' H\tilde{\boldsymbol{x}}_k)^{-1}\tilde{\boldsymbol{x}}_k')H\boldsymbol{y}]^{-\frac{a}{2}} dx_0. \tag{3.6.13}$$

Also

$$P[\text{No switchpoint}] = [1 + B_{i0}]^{-1}. \tag{3.6.14}$$

The probability of a switchpoint in a particular interval follows from equation (3.6.7).

3.7 AUTOCORRELATION

Garisch and Groenewald (1999) considered the linear model with correlated errors

$$\boldsymbol{y} = X\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim N(\boldsymbol{0}, \sigma^2 R)$$

where $\boldsymbol{y} : n \times 1 = [y_1, y_2, \ldots, y_n]'$ is an ordered sequence,

$$X : n \times p, \quad \boldsymbol{\beta} : p \times 1 \text{ and } \{R\}_{ij} = \rho^{|i-j|}, \quad i, j = 1, 2, \ldots, n.$$

Suppose there are r structural changes to the model at $k_1, k_2, \ldots, k_r$, $r \in \{0, 1, \ldots, T\}$ and $T \leq \frac{n}{p} - 1$.

Denote this model by $M_r^{\boldsymbol{k}_r}$, where $\boldsymbol{k}_r = [k_1, \ldots, k_r]$. Then $\boldsymbol{y} = X_r\boldsymbol{\beta}_r + \epsilon$, where $X_r : n \times p(r+1) = \text{diag}\{X^{(i)}\}$, $X^{(i)} : k_i - k_{i-1} \times p$ where

$$k_0 = 0, \quad k_{r+1} = n, \quad i = 1, \ldots, r+1 \text{ and } \boldsymbol{\beta}_r' : 1 \times p(r+1) = [\boldsymbol{\beta}^{(1)'}, \ldots, \boldsymbol{\beta}^{(r+1)'}]. \tag{3.7.1}$$

Suppose we have the following prior distributions:

$$\boldsymbol{\beta}_r \sim N(\mathbf{1}'_{r+1} \otimes \boldsymbol{\theta}, \sigma^2 I_{r+1} \otimes \Phi), \quad \Pi(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \Pi(\boldsymbol{\theta}) \propto 1,$$

$\rho \sim U(-1;1)$ where $\mathbf{1}_{r+1} : (r+1) \times 1 = [1, \ldots, 1], \boldsymbol{\theta}' : 1 \times p$ and $\Phi : p \times p$. (3.7.2)

Under model $M_r^{\boldsymbol{k}_r}$, the data are related to the parameters as follows:

$$\boldsymbol{y}|\boldsymbol{\beta}_r, \sigma^2, \rho, r, \boldsymbol{k}_r \sim N(X_r\boldsymbol{\beta}_r, \sigma^2 R). \tag{3.7.3}$$

We assume that the prior probability for $M_r^{\boldsymbol{k}_r}$ is $P(M_r^{\boldsymbol{k}}) = P(\boldsymbol{k}|r)P(r)$ where $P(\boldsymbol{k}|r) = [\text{no. of partitions of } \boldsymbol{k}]^{-1}$ and $P(r) = \frac{1}{T+1}, \quad r = 0, 1, \ldots, T$.

Let $B_{0r}^{\boldsymbol{k}_r}(\boldsymbol{y})$ denote the Bayes factor for model M_0 (no change-point) against $M_r^{\boldsymbol{k}_r}$ from data $\boldsymbol{y}$. Then $B_{0r}^{\boldsymbol{k}_r}(\boldsymbol{y}) = D_0^{\boldsymbol{k}_0}/D_r^{\boldsymbol{k}_r}$, where

$$\begin{aligned} D_r^{\boldsymbol{k}_r} &= (r+1)^{-\frac{p}{2}}|\Phi|^{-\frac{r}{2}} \int_{-1}^{1} |R|^{-\frac{1}{2}}|X_r'R^{-1}X_r + \frac{r}{r+1}(I_{r+1} \otimes \Phi^{-1})|^{-\frac{1}{2}} \times \\ &\quad \left\{\boldsymbol{y}'R^{-1}\boldsymbol{y} - \boldsymbol{y}'R^{-1}X_r(X_r'R^{-1}X_r + \frac{r}{r+1}(I_{r+1} \otimes \Phi^{-1}))^{-1}X_r'R^{-1}\boldsymbol{y}\right\}^{-\frac{1}{2}(n-p)} d\rho. \end{aligned}$$

The posterior probability for model $M_r^{\boldsymbol{k}_r}$ can now be calculated for Φ known. For fixed T,

$$P(M_r^{\boldsymbol{k}_r}|\boldsymbol{y}) = P(\boldsymbol{k}_r|r)(B_{0r}^{\boldsymbol{k}_r})^{-1} \left[\sum_{j=0}^{T} P(\boldsymbol{k}_j|j) \sum_{k_j} (B_{0j}^{k_j})^{-1}\right]^{-1}. \tag{3.7.4}$$

The Fractional Bayes Factors of O'Hagan (1995, 1997), $B_{0r}^{b,\boldsymbol{k}_r}(\boldsymbol{y})$ is given by

$$B_{0r}^{b,\boldsymbol{k}_r}(\boldsymbol{y}) = \frac{m_0^{\boldsymbol{k}_0}}{m_r^{\boldsymbol{k}_r}}$$

and

$$m_r^{\boldsymbol{k}_r} = \pi^{\frac{n}{2}(b-1)} b^{\frac{nb}{2}} \Gamma\left(\tfrac{1}{2}(n - p(r+1))\right) / \Gamma\left(\tfrac{1}{2}(nb - p(r+1))\right) \times$$

$$\frac{\int_{-1}^{1} |R|^{-\frac{1}{2}} |X_r' R^{-1} X_r|^{-\frac{1}{2}} \left[\boldsymbol{y}' R^{-1} \boldsymbol{y} - \boldsymbol{y}' R^{-1} X_r (X_r' R^{-1} X_r)^{-1} X_r' R^{-1} \boldsymbol{y}\right]^{-\frac{1}{2}[n-p(r+1)]} d\rho}{\int_{-1}^{1} |R|^{-\frac{b}{2}} |X_r' R^{-1} X_r|^{-\frac{1}{2}} \left[\boldsymbol{y}' R^{-1} \boldsymbol{y} - \boldsymbol{y}' R^{-1} X_r (X_r' R^{-1} X_r)^{-1} X_r' R^{-1} \boldsymbol{y}\right]^{-\frac{1}{2}[nb-p(r+1)]} d\rho}. \tag{3.7.5}$$

The posterior probability for model $M_r^{\boldsymbol{k}}$ can now be calculated using the FBF's.

The arithmetic IBF for model M_s against model M_r is defined by

$$AIB_{sr}^{k_s,k_r} = B_{sr}^{k_s,k_r}(\boldsymbol{y}) \, \frac{1}{L} \sum_{\ell=1}^{L} B_{rs}^{k_r,k_s}(\boldsymbol{y}(\ell)),$$

where $B_{sr}^{k_s,k_r}(\boldsymbol{y})$ is the usual Bayes factor with the whole sample and improper priors and $\boldsymbol{y}(\ell)$ represents a minimal training sample.

Suppose we use the same vague prior distributions as before, then

$$B_{sr}^{k_s,k_r}(\boldsymbol{y}) = \frac{m_s^{k_s}(\boldsymbol{y})}{m_r^{k_r}(\boldsymbol{y})}$$

where

$$m_r^{k_r}(\boldsymbol{y}) = (2\pi)^{-\frac{1}{2}[n-(r+1)p]} \left(\frac{1}{2}\right)^{-\frac{1}{2}[n-(r+1)p]+1} \Gamma\left(\tfrac{1}{2}(n-(r+1)p)\right)$$

$$\times \int |R|^{-\frac{1}{2}} |X_r' R^{-1} X_r|^{-\frac{1}{2}} [\boldsymbol{y}' R^{-1} \boldsymbol{y} - \boldsymbol{y}' R^{-1} X_r (X_r' R^{-1} X_r)^{-1} X_r' R^{-1} \boldsymbol{y}]^{-\frac{1}{2}[n-(r+1)p)} d\rho. \tag{3.7.6}$$

The posterior density of ρ, conditional on k, also follows as

$$\Pi(\rho|\boldsymbol{y}, k) \propto |R|^{-\frac{1}{2}} |X_r' R^{-1} X_r|^{-\frac{1}{2}} [\boldsymbol{y}' R^{-1} \boldsymbol{y} - \boldsymbol{y}' R^{-1} X_r (X_r' R^{-1} X_r)^{-1} X_r'$$

$$R^{-1} \boldsymbol{y}]^{-\frac{n-(r+1)p}{2}}, \quad -1 < \rho < 1. \tag{3.7.7}$$

3.8 APPLICATIONS

EXAMPLE 3.8.1

Quandt's data (1958) will be analyzed to see whether we can detect a change in the regression coefficient. This data consists of a sequence of 20 (X, Y) pairs simulated by Quandt, where the first twelve and last eight of these are obeying respectively,

$$\begin{aligned} Y_i &= 2.5 + 0.7X_i + u_i, \quad i = 1, \dots, 12 \\ Y_j &= 5 + 0.5X_j + u_j, \quad j = 13, \dots, 20. \end{aligned}$$

Here the u's are independent standard normal variates. Quandt's data is plotted in Figure 3.1, where '0' denotes the first twelve and '$*$' the last eight observations.

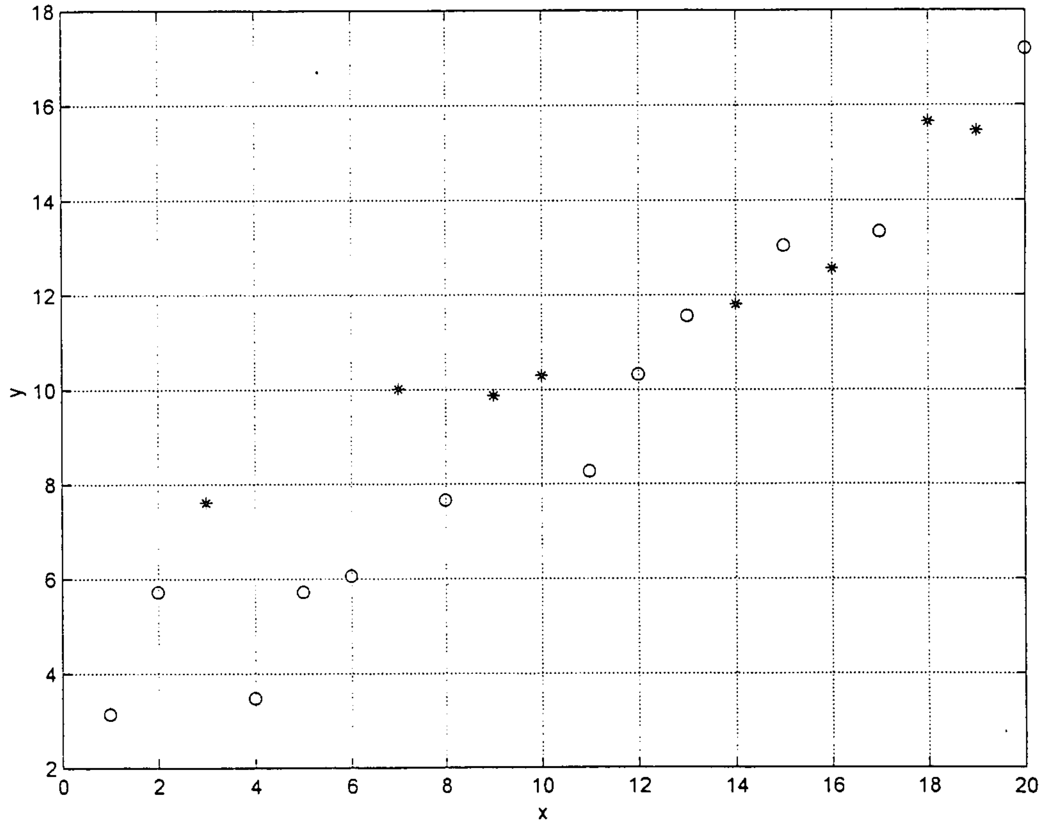


Figure 3.1: Quandt's data

Quandt's data was analyzed by a number of people to illustrate methods of change-point detection. Amongst others, Ferreira (1975), Holbert and Broemeling (1977), Chin Choy and Broemeling (1980), Land and Broemeling (1983) and Wang and Lee (1993) analyzed the data. All of them assumed that there is exactly one change-point.

We will calculate the posterior probability of the position of the change-point, assuming one exists, as well as the probability of no change by using the FBF and the intrinsic Bayes factor.

Using equation (3.2.39), the posterior probability distribution of k is given in Figure 3.2. The maximum probability is 0.5353 at the 12th observation. Holbert and Broemeling (1977) got a maximum probability of 0.5051, while Chin Choy and Broemeling (1980), by using proper priors, got a maximum probability of 0.6844 (both at the 12th observation).

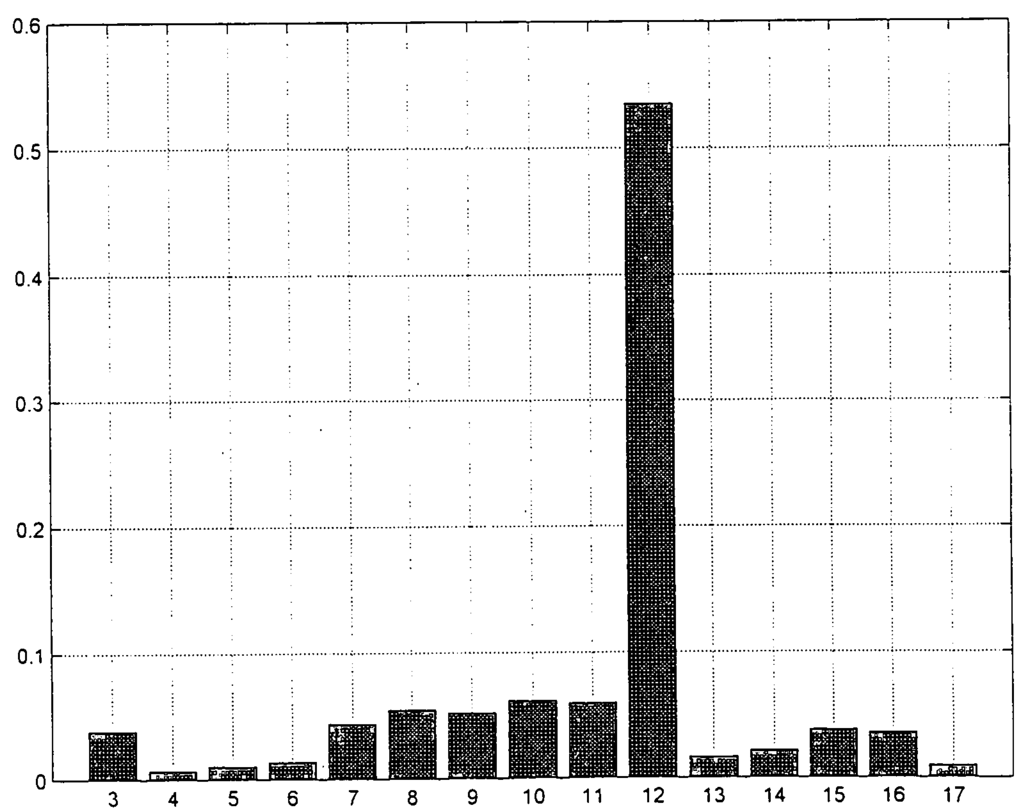


Figure 3.2: Posterior probability distribution of change-point for example 3.8.1

The results for the FBF (equation (3.3.8) with (2.5.7)) and the three intrinsic Bayes factors (arithmetic, median and geometric) from (3.3.3) and (3.3.4) are given in Table 3.1 for no change and for $k = 12$. In all cases $k = 12$ was by far the highest probability. Although the FBF gives the highest probability for $k = 12$, it also has the highest probability of 0.1894 for no change among the four methods used. The three intrinsic Bayes factors give very similar answers.

Table 3.1

	No change	$k = 12$
FBF	0.1894	0.6114
Arithmetic IBF	0.0128	0.5284
Median IBF	0.0067	0.5317
Geometric IBF	0.0055	0.5323

Figure 3.3 gives the conditional ($k = 12$) and unconditional marginals of the five parameters. The conditional forms of the equations are given below (3.2.38). A summary of the results is given in Table 3.2.

Table 3.2: **Summary of the results of the unconditional marginals of $\alpha_1, \alpha_2, \beta_1, \beta_2$ and σ^2, as well as those of Chin Choy and Broemeling (CCB)(1980).**

	α_1	α_2	β_1	β_2	σ^2
CCB - mean	2.36	5.45	0.67	0.52	0.83
- 95% HPD	1.24 – 3.42	3.8 – 6.77	0.55 – 0.80	0.41 – 0.63	0.52 – 1.85
Schoeman - mean	2.3345	5.5306	0.6812	0.5031	1.3369
- 95% HPD	0.839 – 3.611	3.349 – 7.722	0.543 – 0.839	0.344 – 0.656	0.489 – 2.569

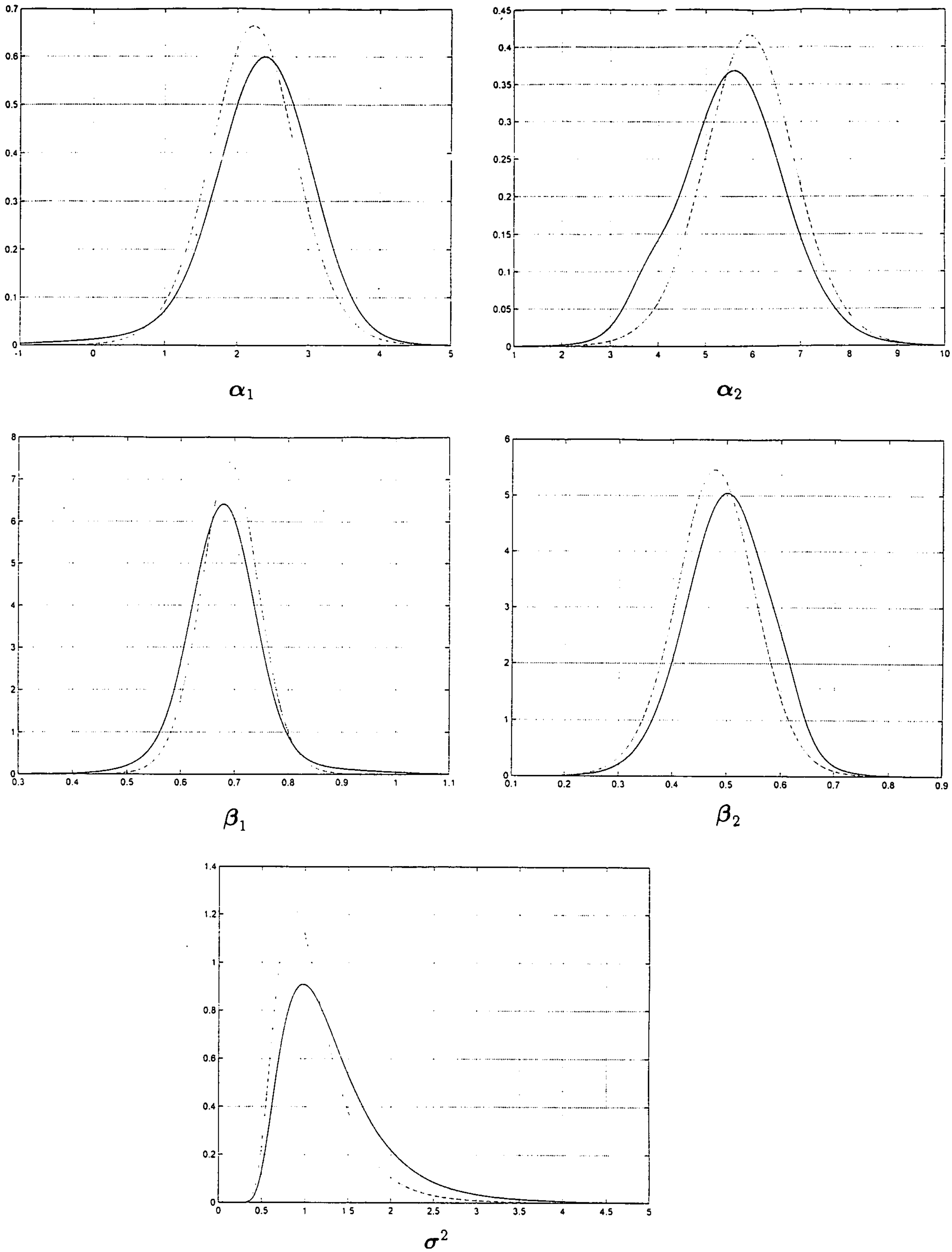


Figure 3.3: The conditional (−−) and unconditional (−) marginals posteriors of $\alpha_1, \alpha_2, \beta_1, \beta_2, \sigma^2$, respectively, for example 3.8.1

EXAMPLE 3.8.2

For component analysis as in paragraph 3.5 we will use Brownlee's stack loss data, as given in Hand, *et al.* (1994). These classic data are observations from 21 days' operation of a plant for the oxidation of ammonia as a stage in the production of nitric acid. The carrier variables are air flow (X_1), cooling water inlet temperature (° C) (X_2) and acid concentration (%) (X_3). The response variable is stack loss (Y), which is the percentage of the ingoing ammonia that escapes unabsorbed.

From using the FBF in equation (3.3.8), Table 3.3 gives the probability for no change-point and of the three highest probability change-points.

Table 3.3

k	Probability
No change	2.29×10^{-6}
4	0.4740
5	0.5069
6	0.0187

If we assume one change-point the probability, by using equation (3.2.39), for $k = 4$ is 0.8023, followed by a probability of 0.1742 for $k = 5$ and a probability of 0.0208 for $k = 6$. Although these probabilities seem quite different from that from the FBF in Table 3.3, the differences are not that extreme. Out of 16 possible change-points, both methods give a posterior probability of about 0.98 for a change at either $k = 4$ or $k = 5$.

The distribution of σ^2 and of $\sigma^2|k = 4$ is given in Figure 3.4 according to the information below equation (3.2.38). These distributions are almost identical since most posterior probability mass is concentrated at $k = 4$.

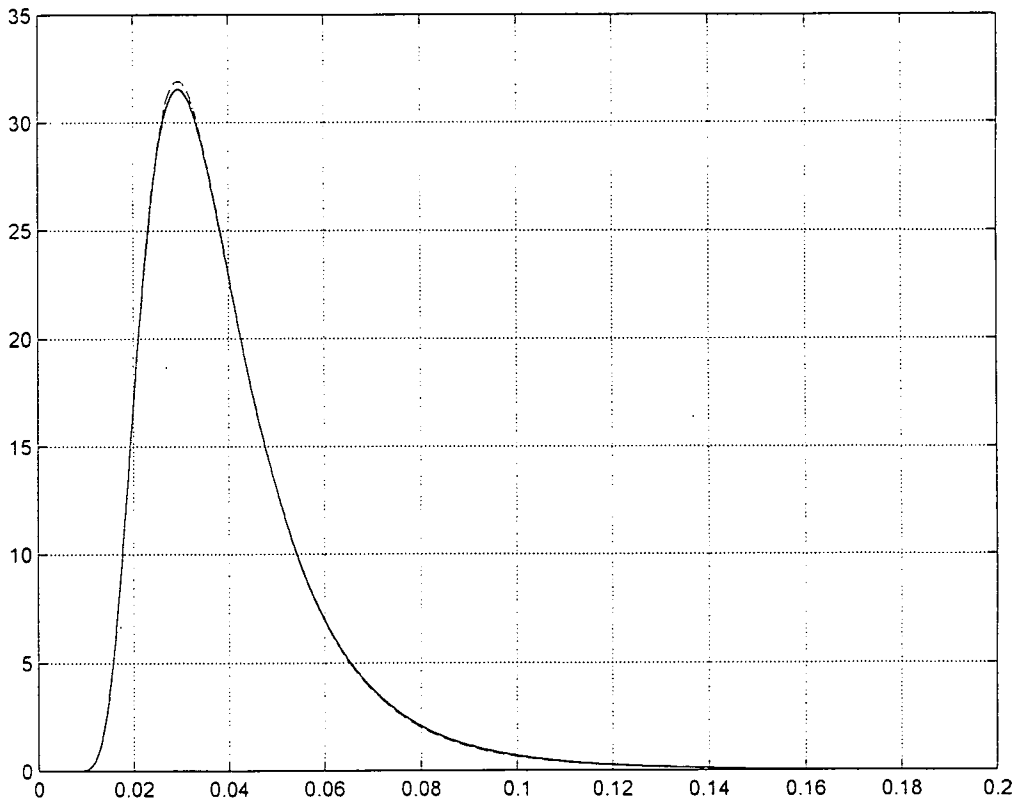


Figure 3.4: The distribution of $\sigma^2(-)$ and $\sigma^2|k = 4(--)$ for example 3.8.2

The conditional ($k = 4$) and unconditional marginal posteriors of the components of the differences $\boldsymbol{\Delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$, by using equation (3.5.2), is given by Figure 3.5. Notice that the unconditional distribution of the differences has two modes. The major mode is due to the conditional density of Δ_i given $k = 4$ (with probability 0.8023) and the minor mode is caused by the conditional density of Δ_i when $k = 5$ (with probability 0.1742).

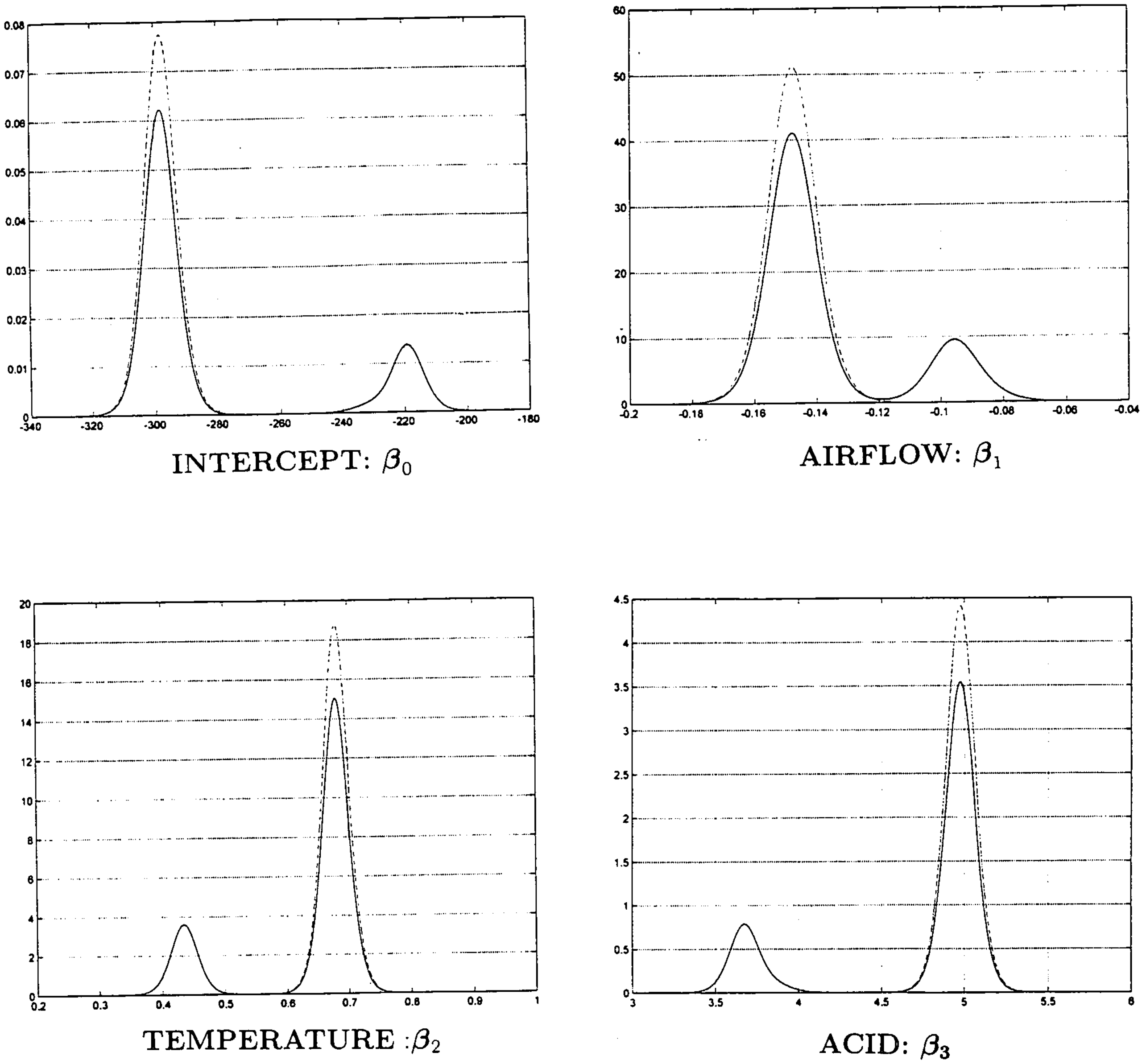


Figure 3.5: The conditional (−−) and unconditional (−) marginal posteriors of the components of the differences Δ for example 3.8.2

The standardized mean differences between the four regression coefficients before and after a change at $k = 4$ (equation (3.5.3)) and the influence of each component in causing the change-point (equation (3.5.4)) are given in Table 3.4.

Table 3.4

	D	I
β_0	54.3293	35.36
β_1	17.7707	11.57
β_2	29.9382	19.48
β_3	51.6203	33.59

So it seems that there is a major change in the relationship between stack loss and the carrier variables after the fourth or fifth day, with major contributor the % acid (X_3), and the intercept.

EXAMPLE 3.8.3

Distances and heights (in metres) achieved by winners of Olympic jumping events from 1896 to 1988 will be analyzed under the assumption of a linear improvement over time in performance to see whether we can detect a change in the regression coefficient. The four events are high jump, pole vault, long jump and triple jump. There were no Olympic games in 1916, 1940 and 1944. Caussinus and Lyazrhi (1997) also analyzed the data.

By using equation (3.4.5) and (2.5.17), the probability for no change, one change and two changes are given by Table 3.5, with $b = \frac{3p+1}{n} = \frac{7}{n}$ in the FBF.

Table 3.5: Posterior probabilities for $k = 0, 1, 2$ change-points

	No change	$k = 1$	$k = 2$
High jump	0.0003	0.9169	0.0828
Pole vault	0	0.1574	0.8425
Long jump	0.0820	0.8288	0.0891
Triple jump	0.2786	0.6340	0.0874

By using equation (3.2.39), the probabilities of the change-point positions for the four events, given there is one change-point, are given by Figure 3.6.

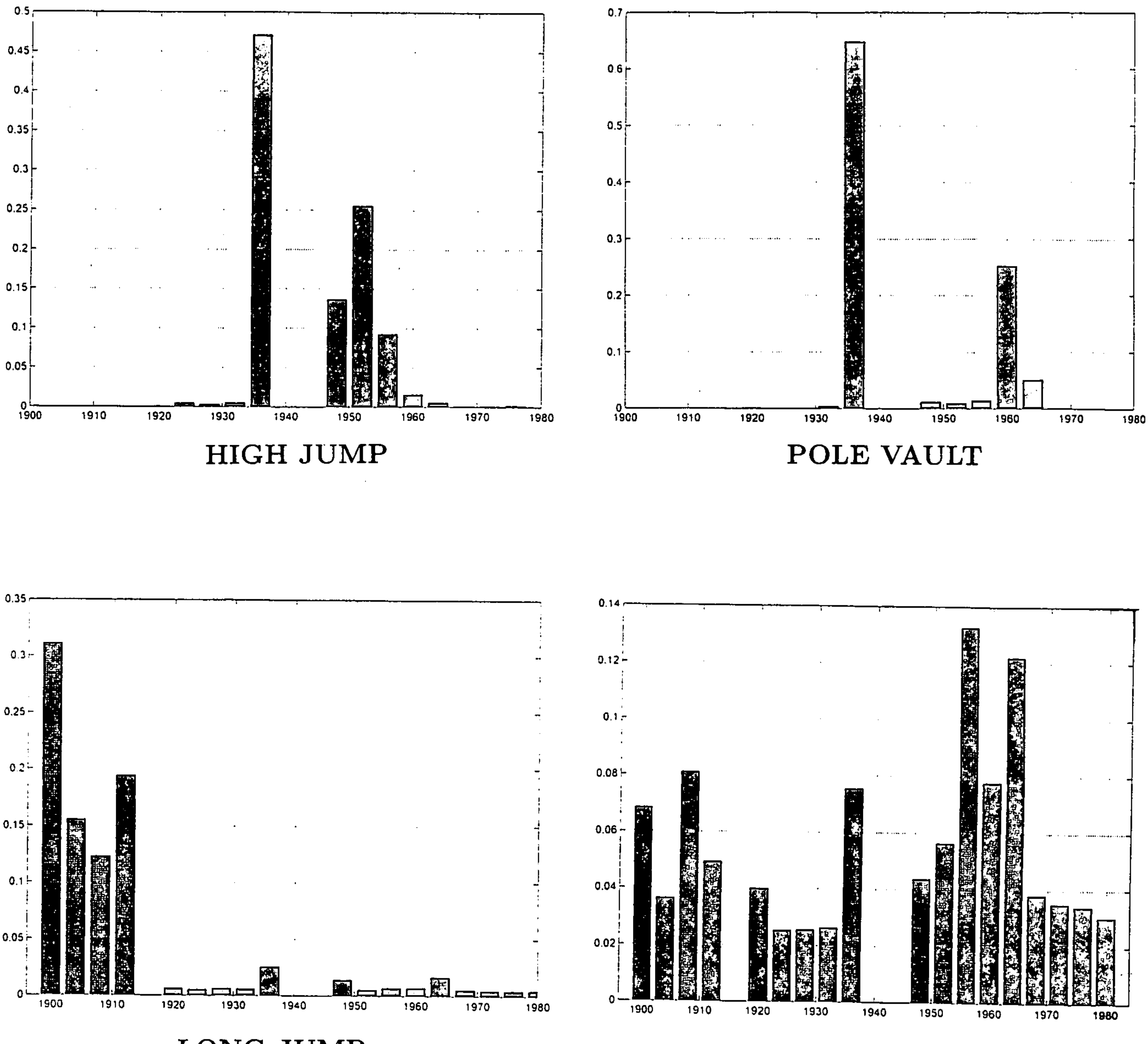


Figure 3.6: The probabilities of the change-point positions for the four events, given there is one change-point, for example 3.8.3

The two highest probabilities of the change-points, given there are two change-points, are given by Table 3.6. This follows from equation (2.5.17).

Table 3.6

	Years		Probability
High jump	1900	1936	0.0778
	1936	1968	0.0647
Pole vault	1908	1960	0.3570
	1912	1960	0.1993
Long jump	1900	1964	0.1708
	1900	1956	0.0651
Triple jump	1936	1968	0.1457
	1936	1964	0.1203

Figure 3.7 gives the men's Olympic performances in the pole vault. Caussinus and Lyazrhi (1997) decided that two change-points have occured, one after 1908 and another after 1960, which are the same as our results. According to them the second change-point corresponds to a sudden, or at least very rapid, improvement in the equipment.

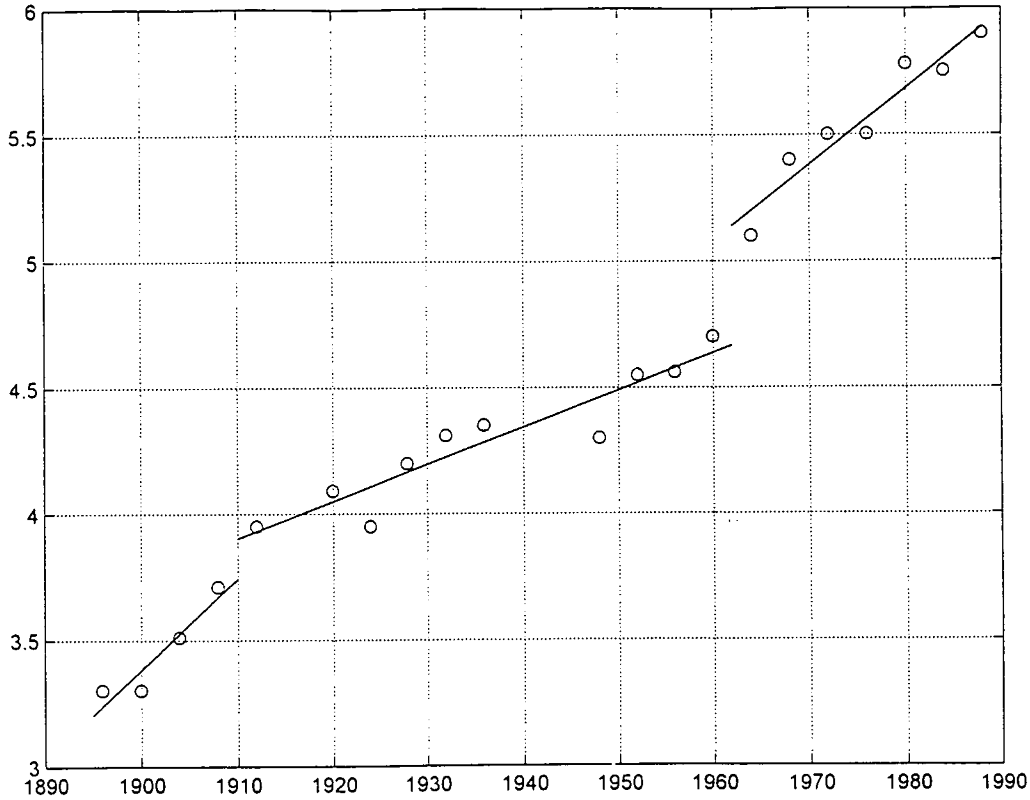


Figure 3.7: Men's Olympic performances in the pole vault, for example 3.8.3

In the long jump, Caussinus and Lyazrhi (1997) found that there are two change-points (1912 and 1936) and one outlier (1968). According to them the two change-points correspond to the two war periods, while the outlying performance in 1968 (Mexico) is well known. According to our results, only one change-point appears in this sequence with a probability of 0.8288. The probability of two change-points is only 0.0891. The triple jump is the event most likely to have no change-point.

EXAMPLE 3.8.4

We will use the windmill data of Jaglekar, Schuenemeyer and LaRiccia (1989), as given in Hand, *et al.* (1994), as an example of a switchpoint. For the windmill data direct current output (Y) was measured against wind velocity (x, miles per hour). There were 25 observations recorded. Assuming a linear relationship between current output of the windmill and wind velocity, we want to determine whether and where there is a switchpoint.

By using the FBF in equation (3.6.12) and (3.6.14), the probability of no change was 8.945×10^{-5}. Assuming a change has occurred, the posterior distribution of the switch point, by using equation (3.6.7), is given in Figure 3.8. The mean is 4.4264, with a probability of 0.5586 for the interval 4.1 – 4.6 mph. The slope of the regression line is 0.626 before the switch point and 0.148 after the switch point, which means that the increase in generated output becomes more than four times slower when windspeeds exceed about 4.4 mph.

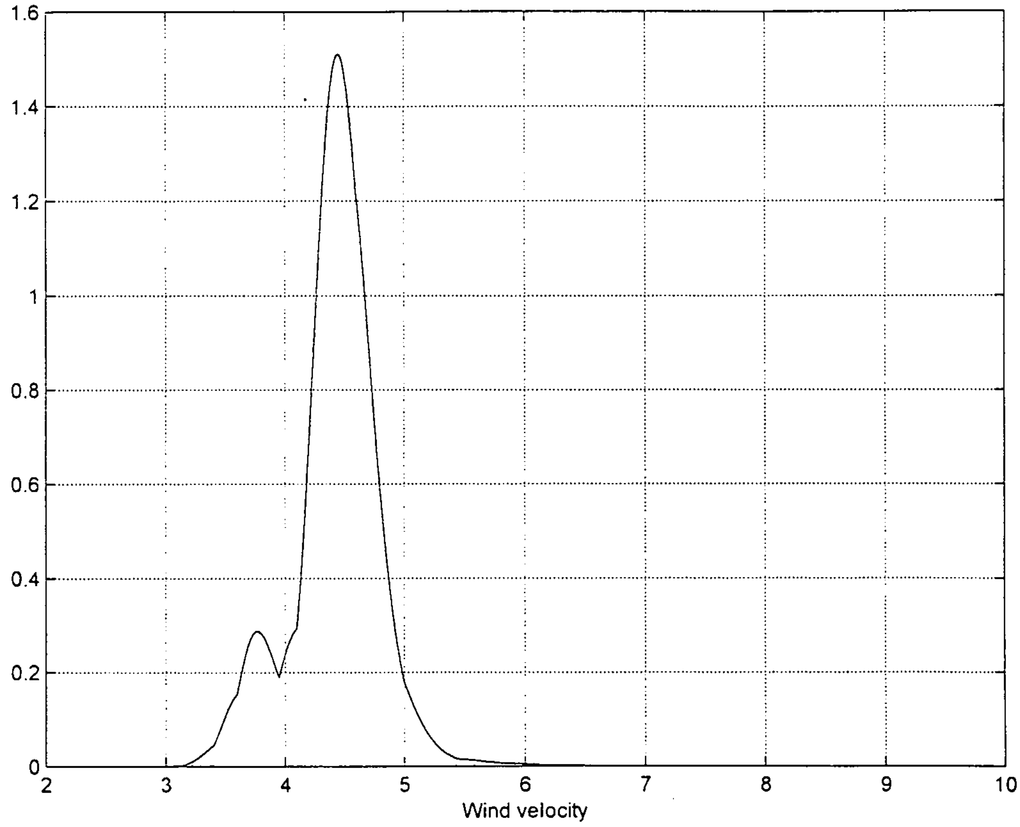


Figure 3.8: The posterior distribution of the switchpoint, assuming a change has occured, for example 3.8.4.

EXAMPLE 3.8.5

As an example of a change in the regression coefficient over time, we will use the data of monthly dollar volume of sales (in millions) on the Boston Stock Exchange (BSE) and the combined New York American Stock Exchange (NYAMSE) from January 1967 to November 1969.

Holbert (1982) analyzed this same data set to illustrate the estimation of the change-point in two-phase regression by calculating the posterior density of the change-point. He found out that the maximum posterior density occurred at position 24, which was corresponding to the calender month of December of 1968 and concluded that it is a change-point caused by the abolition of give-ups (commission splitting) in December of 1968.

Chen (1998) took the same data to illustrate the SIC method for locating the switching change-point in linear regression, which corresponded to time point 23, hence the regression model change-point started at the time point 24. This conclusion coincides with the one drawn by Holbert (1982) using his method.

Figure 3.9 is the scatter plot of the BSE versus NYAMSE, with circles indicating the pairs of (BSE, NYAMSE) before December 1968 and stars indicating the pairs of (BSE, NYAMSE) on and after December 1968. In this scatter plot, the two regression lines are also plotted, with the line having dashes indicating the regression line before December 1968 and the solid line indicating the regression line after December 1968.

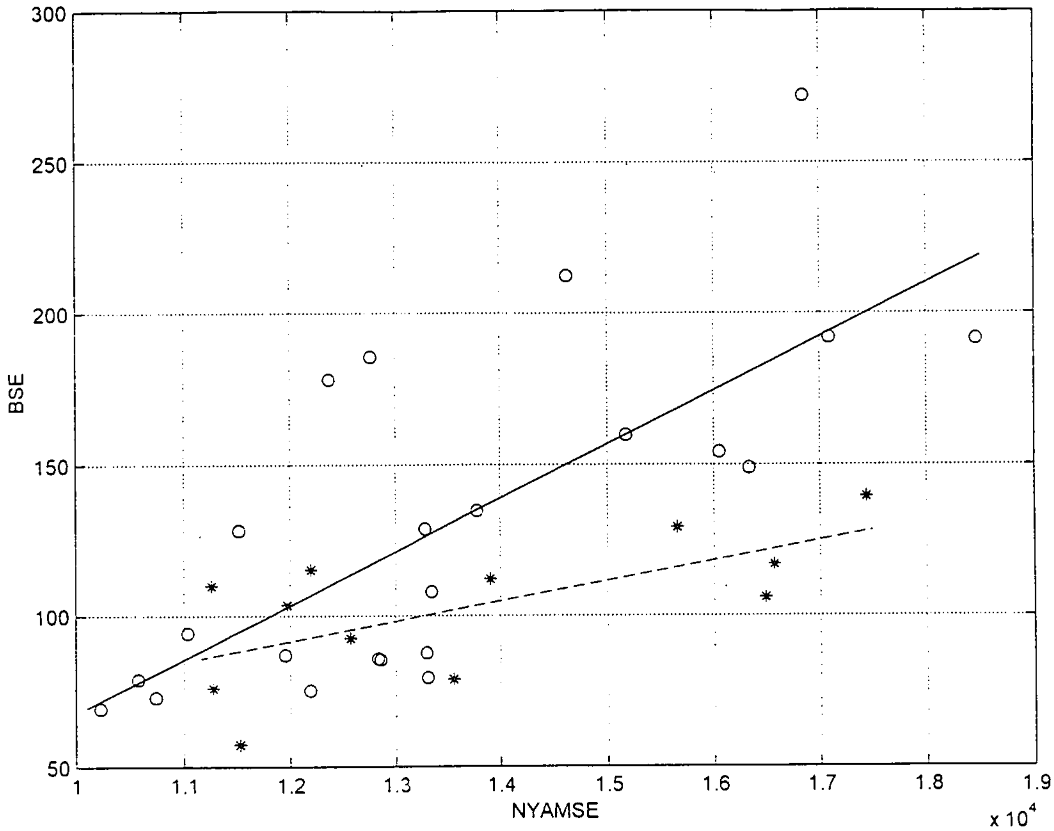


Figure 3.9: Scatter plot and regression lines of BSE versus NYAMSE, given $k = 23$, for example 3.8.5

The posterior probabilities from the Arithmetic Intrinsic Bayes factor (IBF), Median IBF and Geometric IBF, together with the posterior probabilities by assuming a change-point (using equation (3.2.39)) and by using the FBF, is given in Table 3.7.

Table 3.7

Time point	AIBF	MIBF	GIBF	(3.2.39)	FBF
0	0.0091	0.0033	0.0032	–	0.0781
23	0.3087	0.3105	0.3105	0.3115	0.4768

The posterior measures of the parameters, given a change-point occured, is given in Table 3.8.

Table 3.8

	CONDITIONAL, $k = 23$				UNCONDITIONAL			
	Expected value		95% cred. int.		Expected value		95% cred. int.	
	Before cp	After cp	Before cp	After cp.	Before cp	After cp	Before cp	After cp
β_0	-110.3096	11.0752	-199- - 22	-114- +136	-53.7681	-9.6231	-233- + 199	-141- +144
β_1	0.0178	0.0067	0.0113 - 0.0242	-0.0023 - 0.0157	0.0123	0.0091	-0.0101- -0.0262	-0.0025 -0.0194
σ^2	1183.0		697-1910		1347.3		759 - 2231	

As an example of a change in the variance, we will once again use the data of the BSE versus NYAMSE, but with a rearrangement of the X matrix in increasing order.

The posterior probability $\Pi(k|y, \alpha, \gamma, \delta)$, by using equation (3.2.53), is sensitive to all three parameters. Figure 3.10 gives the posterior probabilities of $k = 3, 4, 5, 6$ and of no change-point as a function of δ then $\alpha = 2, \gamma = 0.1$. Approximately the same figure is obtained as a function of α and as a function of γ.

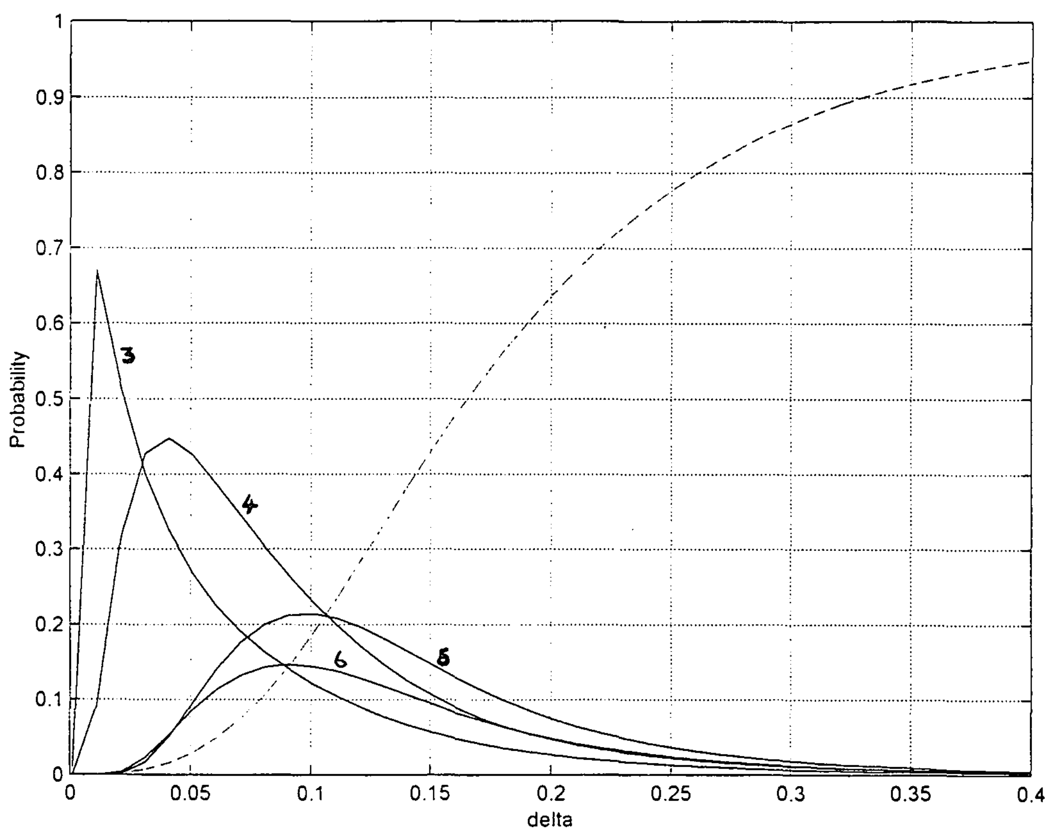


Figure 3.10: Posterior probabilities of $k = 3, 4, 5, 6(-)$ and of no change-point $(--)$ as a function of δ when $\alpha = 2, \gamma = 0.1$, for example 3.8.5

Figure 3.11 gives the posterior probabilities of the position of the change-point by (a) assuming a change-point did occur (using equation (3.2.57)), (b) using proper priors, using equation (3.2.54) with $\alpha = 2, \gamma = 0.1, \delta = 0.11$ and (c) by using the FBF in equation (3.3.15). For both the FBF (c) and the proper priors when $\delta = 0.11$ (b), the probability of no change is 0.2228.

There is no strong evidence of a change-point in the variance when the NYAMSE changes. If there is one change-point, it looks to be somewhere between a NYAMSE volume of 11040 and 11280.

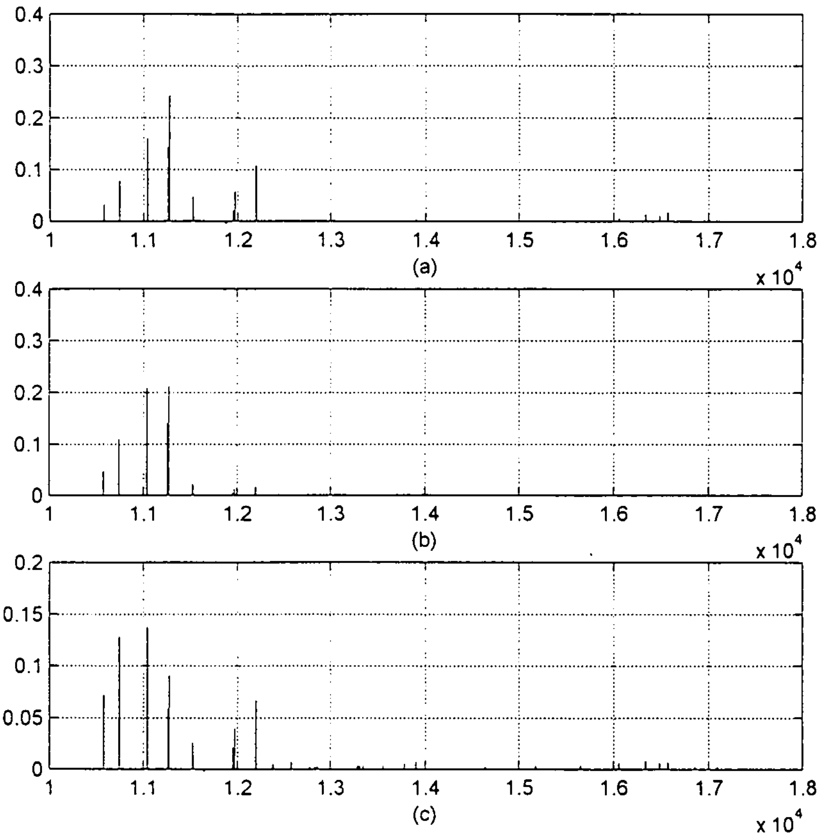


Figure 3.11: Posterior probabilities of position of the change-point using (a) equation (3.2.57), (b) equation (3.2.54) with $\alpha = 2$, $\gamma = 0.1$, $\delta = 0.11$ and (c) the FBF in equation (3.3.15), for example 3.8.5

EXAMPLE 3.8.6

As an example of a change in the mean and variance, we will use the data of raw cotton imports into the UK by weight, for each year 1770-1800, as given in Hand *et al.* (1994). Figure 3.12 represents a plot of this cotton imports in the 18th century.

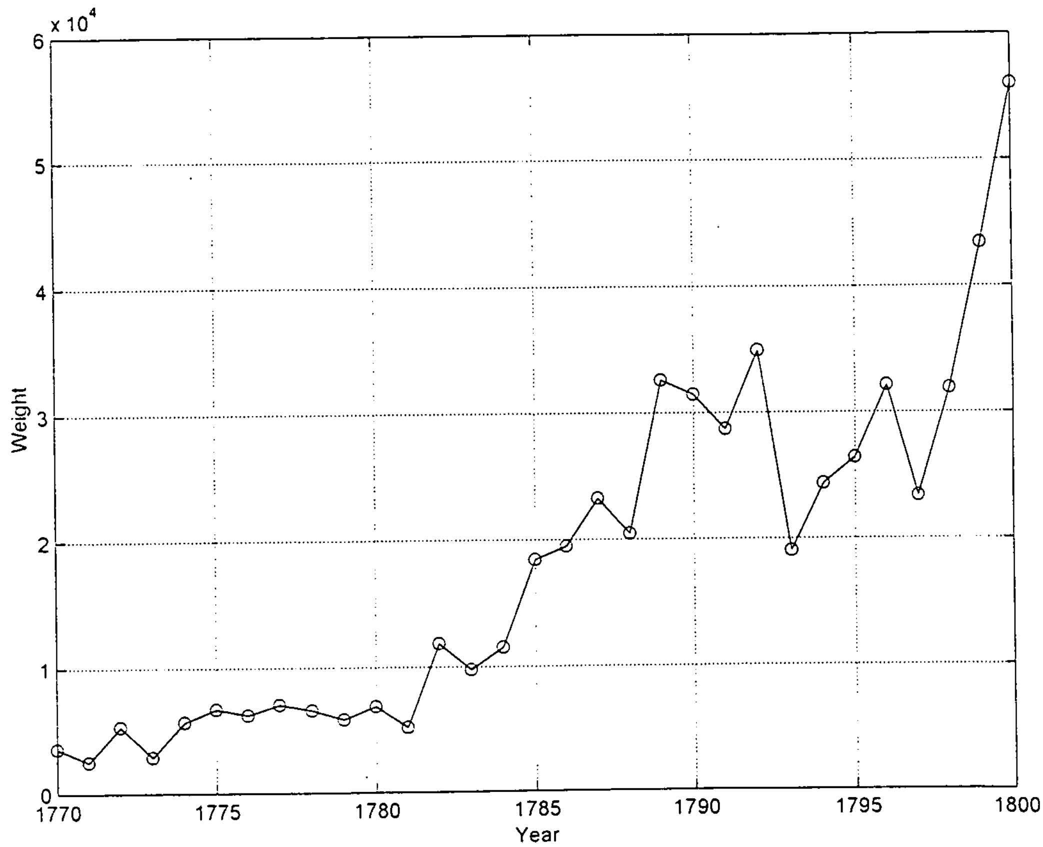


Figure 3.12: Cotton imports into the UK in the 18th century

First, to select between the four models no change (M_0), the slope changes (M_1), the variance changes (M_2) and both the slope and variance change (M_3), the FBF is used. Using equations (3.3.6), (3.3.13) and (3.3.21) together with (3.3.7), the FBF and posterior probability for each model is calculated, using a prior probability of $\frac{1}{4}$ for each model. The results are given in Table 3.9.

Table 3.9

Model	M_0	M_1	M_2	M_3
Post. prob.	0.0006	0.0069	0.0103	0.9822

This shows overwhelming evidence for a change in both the slope and variance.

Figure 3.13 shows the posterior probabilities of a change-point in the slope and variance for the cotton import data by (a) assuming a change-point and (b) using the FBF.

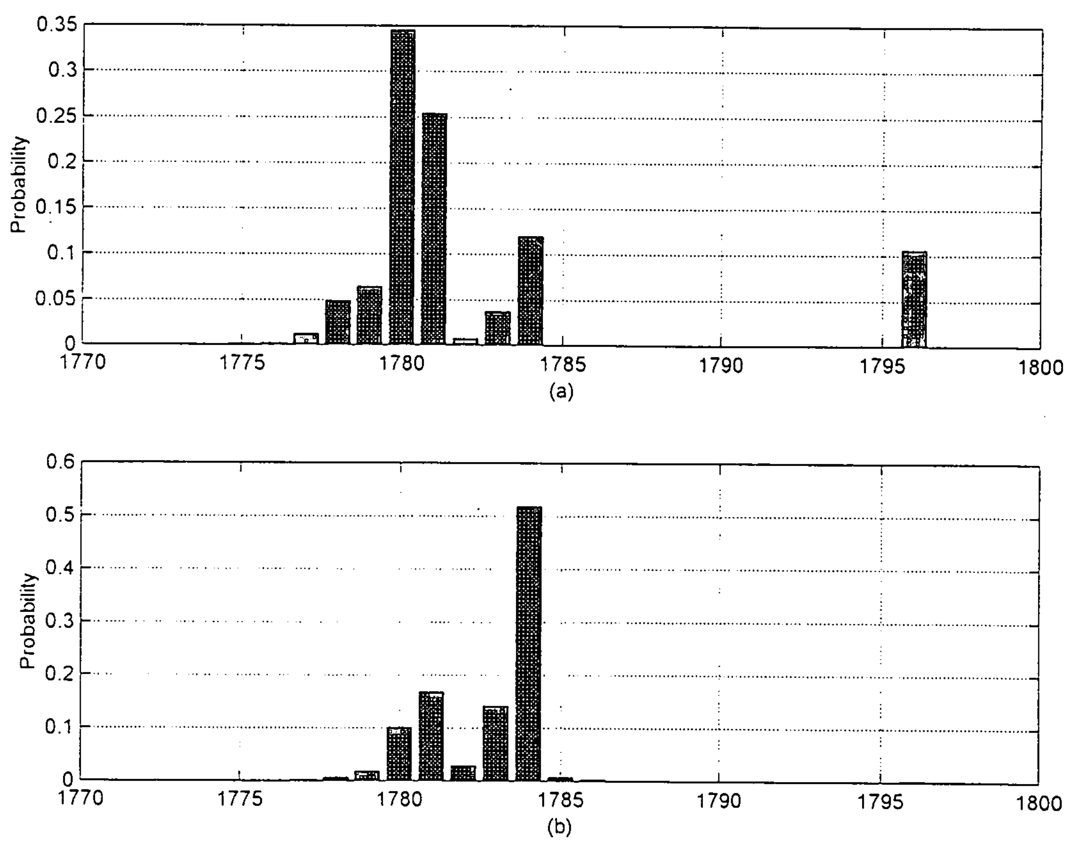


Figure 3.13: **Posterior probability of a change-point in the slope and variance for cotton import data by (a) assuming a change-point and (b) using the FBF, for example 3.8.6**

Assuming only model M_3 or M_0, the posterior probability for no change is 0.0007, using the FBF in equation (3.3.22). Assuming a change-point, equation (3.2.65) indicates a change-point after 1780 or 1781 (probability of 0.345 and 0.254 respectively in Figure 3.13 (a). However, the FBF favours a change-point after 1784 (probability of 0.518), as shown in Figure 3.13(b).

EXAMPLE 3.8.7

We will use the values (in millions of pounds) of British exports for each year 1820 - 1850 as an example of correlated data, given in Hand, *et al.*. Figure 3.14 represents a plot of this British exports in the 19th century.

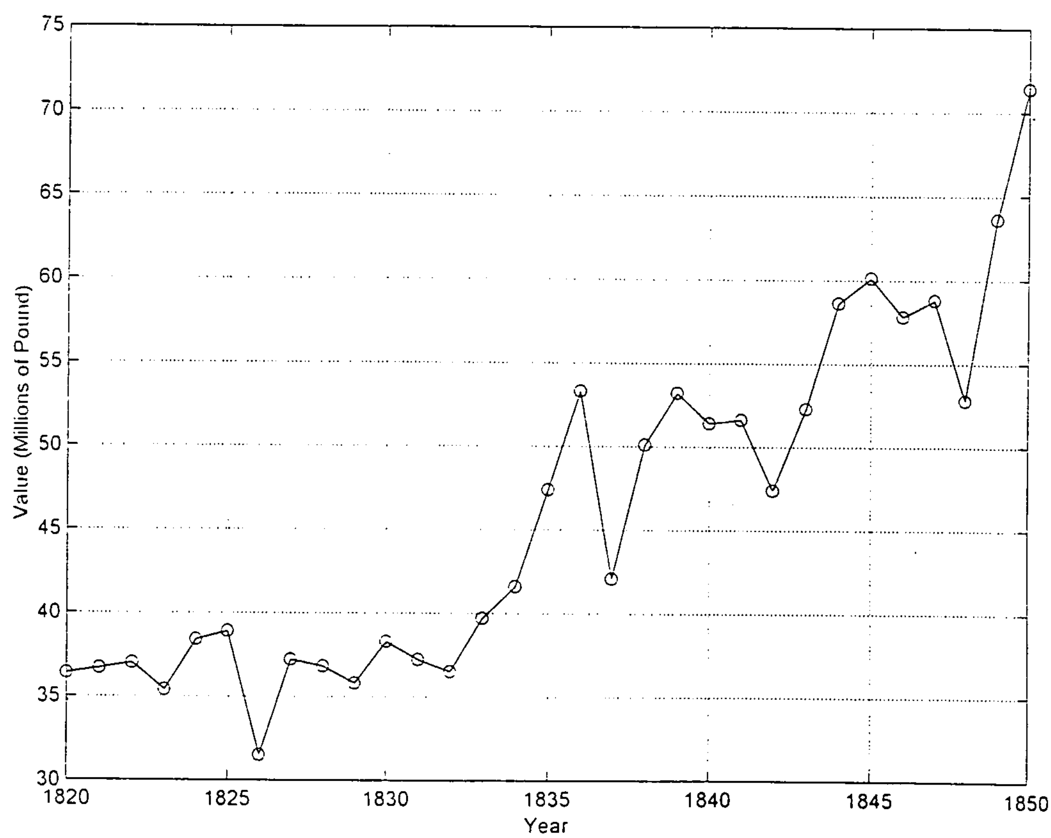


Figure 3.14: British exports in the 19th century

To investigate the effect of correlation, we calculated the FBF in favour of no change, when $\rho = 0$, by using equation (3.3.8) and resulted in a posterior probability of 0.0020 with maximum change-point probability $P(k = 14) = 0,1445$. If $\rho \neq 0$, the posterior probability, by using the FBF in equation (3.7.5), for no change is 0.8042 with a maximum change-point probability $P(k = 6) = 0.0329$. So the probability that there is a change-point depends heavily on whether there is autocorrelation.

Figure 3.15 gives the posterior density of the autocorrelation coefficient ρ (by using equation (3.7.7)), given a change-point (unconditional of k), assuming no change and the unconditional posterior of ρ. The expected value given a change-point is 0.344, given no change-point is 0.529 and for the unconditional case is 0.493, showing that the estimated correlation is much smaller when a change-point is assumed.

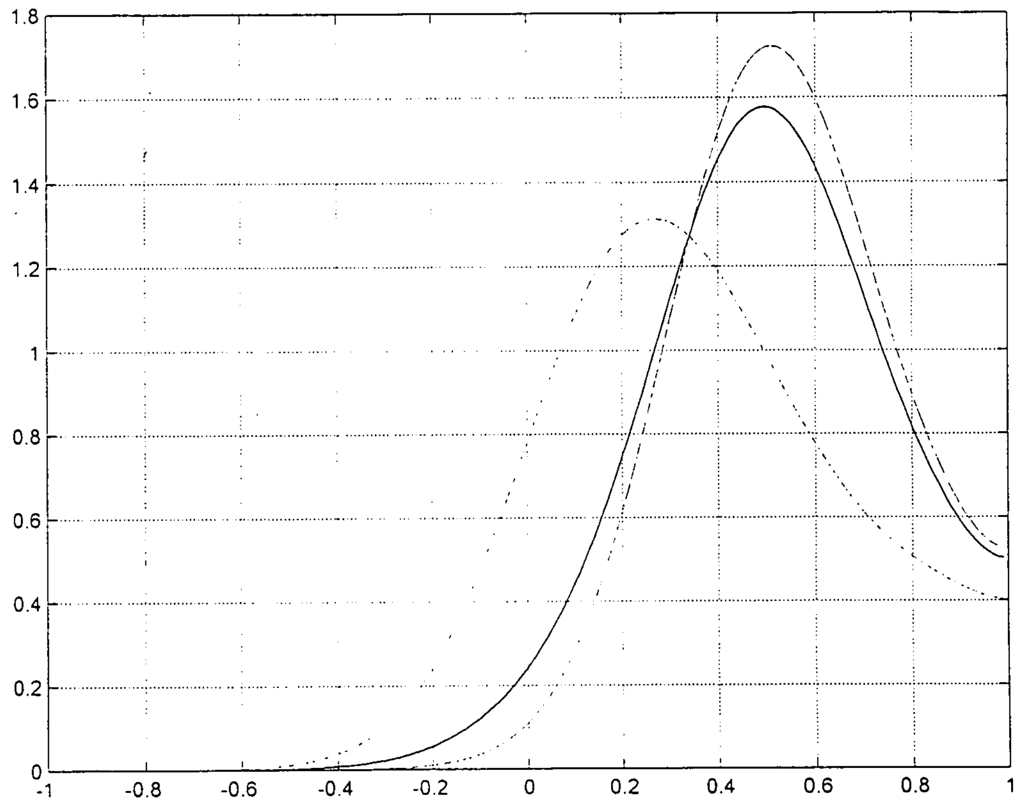


Figure 3.15: Posterior density of the autocorrelation coefficient ρ, given a change-point $(-.-.$, unconditional of $k)$, assuming no change $(--)$ and the unconditional posterior of $\rho(-)$ for example 3.8.7

To see which model would best fit the data, let's find the FBF when comparing a model with a change-point but no autocorrelation with a model with autocorrelation but no change-point (Equations (3.3.6) and (3.7.5)). Figure 3.16 shows the FBF in favour of no change-point as a function of the autocorrelation coefficient ρ. The maximum, $B = 0.835$, occurs at $\rho = 0.375$. This shows mild evidence in favour of the change-point model.

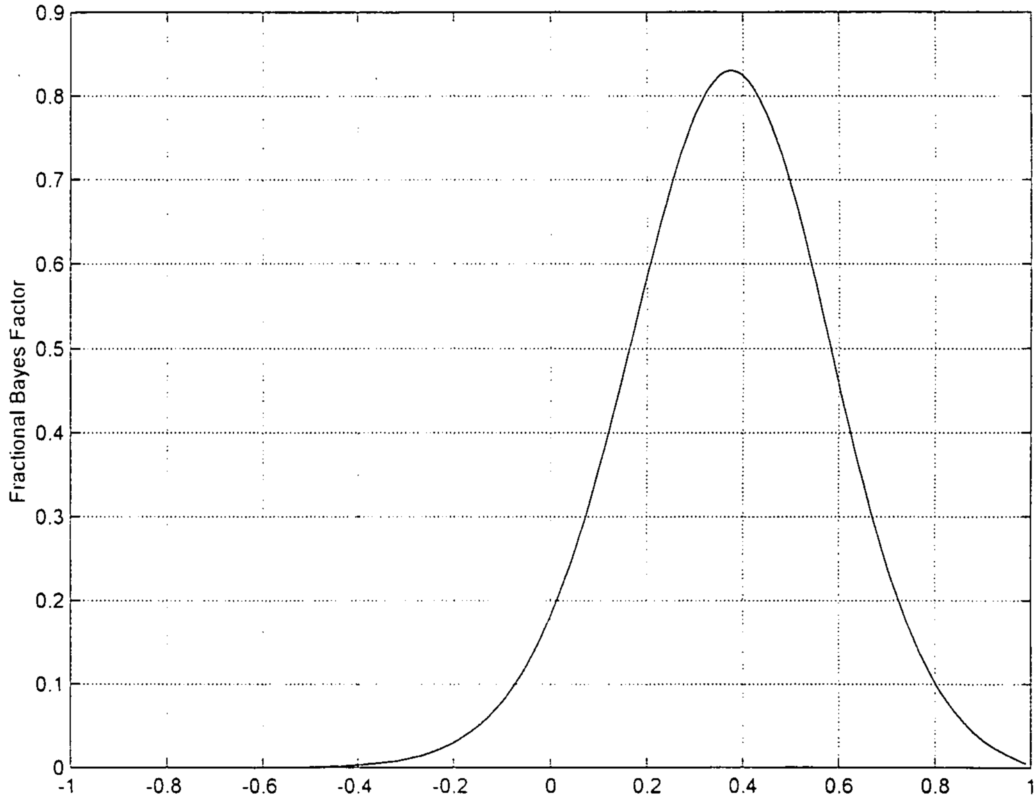


Figure 3.16: FBF in favour of no change-point as a function of the autocorrelation, for example 3.8.7

The model that actually fit the British export data best is one with a change-point after 1834 ($k = 14$) and autocorrelation of $\rho = 0.110$.

CHAPTER 4

CHANGES IN SOME OTHER MODELS

4.1 CHANGE-POINT IN BERNOULLI TYPE EXPERIMENTS

4.1.1 THE BINOMIAL DISTRIBUTION

Hinkley and Hinkley (1970) considered the binomial parameter in the problem of making inferences about the change-point in a sequence of zero-one variables. They derived the asymptotic distribution of the maximum likelihood estimate of the change-point in computable form using random walk results. They also obtained the asymptotic distributions of likelihood ratio statistics for testing hypotheses about the change-point.

Smith (1975) considered a Bayesian approach to the problem of making inferences about the change-point. He considered the binomial distribution where a sequence of random variables $X_1, \ldots, X_n$ is said to have a change-point at $k(1 \leq k \leq n)$ if $X_i \sim$ Bin $(x|p_1)$ $(i = 1, \ldots, k)$ and $X_i \sim$ Bin $(x|p_2)$ $(i = k+1, \ldots, n)$. He considered the cases where (a) p_1 and p_2 are known, (b) p_1 known, p_2 unknown and (c) p_1 and p_2 unknown.

We will consider the case where both parameters are unknown.

Let the sequence $X_1, \ldots, X_n$ be such that

$$X_i \sim \text{ Bin } (p_1, m_i), \quad i = 1, \ldots, k$$

and

$$X_i \sim \text{ Bin } (p_2, m_i), \quad i = k+1, \ldots, n$$

so that

$$f(\boldsymbol{x}|p_1, p_2, k) = \left[\prod_{i=1}^{n} \begin{pmatrix} m_i \\ x_i \end{pmatrix}\right] p_1^{y_1}(1-p_1)^{\sum\limits_{i=1}^{k} m_i - y_1} p_2^{y_2}(1-p_2)^{\sum\limits_{i=k+1}^{n} m_i - y_2}$$

where

$$y_1 = \sum_{i=1}^{k} x_i \text{ and } y_2 = \sum_{i=k+1}^{n} x_i \tag{4.1.1}$$

Let k have a uniform prior and let the prior of p_1, p_2 be

$$\Pi(p_1, p_2) = B(\alpha, \beta)^{-2} p_1^{\alpha-1}(1-p_1)^{\beta-1} p_2^{\alpha-1}(1-p_2)^{\beta-1}.$$

Then

$$\begin{aligned} f(\boldsymbol{x}|k,\alpha,\beta) &= \iint f(\boldsymbol{x}|p_1,p_2,k)\Pi(p_1,p_2)\Pi(k)dp_1dp_2 \\ &= \prod_{i=1}^{n} \begin{pmatrix} m_i \\ x_i \end{pmatrix} B(y_1+\alpha, \sum_{i=1}^{k} m_i - y_1 + \beta) B(y_2 + \alpha, \sum_{i=k+1}^{n} m_i - y_2 + \beta) \end{aligned} \tag{4.1.2}$$

so that

$$\pi(k|\boldsymbol{x},\alpha,\beta) = \frac{f(\boldsymbol{x}|k,\alpha,\beta)}{\sum_{i=1}^{n-1} f(\boldsymbol{x}|i,\alpha,\beta)}. \tag{4.1.3}$$

For a vague prior a number of options are available, such as $\alpha, \beta = 0, \frac{1}{2}$ or 1. For a symmetrical prior we can put $\alpha = \beta$.

Furthermore

$$p_1|\boldsymbol{x},k,\alpha,\beta \sim B\left(y_1 + \alpha, \sum_{i=1}^{k} m_i - y_1 + \beta\right)$$

and

$$p_2|\boldsymbol{x},k,\alpha,\beta \sim B(y_2 + \alpha, \sum_{i=k+1}^{n} m_i - y_2 + \beta).$$

Also

$$\Pi(p_i|\boldsymbol{x},\alpha,\beta) = \sum_{k} \Pi(p_i|\boldsymbol{x},k,\alpha,\beta)\Pi(k|\boldsymbol{x},\alpha,\beta), \quad i = 1, 2. \tag{4.1.4}$$

For $\tau = \frac{p_1}{p_2}$ it follows that

$$\Pi(\tau|\boldsymbol{x},k,\alpha,\beta) \propto \tau^{y_1+\alpha} \int_0^q p_2^{y+2\alpha-2}(1-\tau p_2)^{\sum_1^k m - y_1 + \beta - 1} (1-p_2)^{\sum_{k+1}^n m - y_2 + \beta - 1} dp_2 \tag{4.1.5}$$

where $q = \min\left(1, \frac{1}{\tau}\right)$.

Considering the possibility of no change, let $p \sim B(\alpha, \beta)$ so that

$$\begin{aligned} f(\boldsymbol{x}|k=n,\alpha,\beta) &= \int \left[\prod_{i=1}^n \begin{pmatrix} m_i \\ x_i \end{pmatrix}\right] p^y (1-p)^{\sum_{i=1}^n m_i - y} [B(\alpha,\beta)]^{-1} p^{\alpha-1}(1-p)^{\beta-1} dp \\ &= \prod_{i=1}^n \begin{pmatrix} m_i \\ x_i \end{pmatrix} B(\alpha+y, \sum_{i=1}^n m_i - y + \beta)[(B(\alpha,\beta)]^{-1}. \end{aligned}$$

Furthermore, if we let $\Pi(k) = \begin{cases} q, & k=n \\ \frac{1-q}{n-1}, & k=1,\ldots,n-1 \end{cases}$ (4.1.6)

it follows that

$$\begin{aligned} \Pi(k=n|\boldsymbol{x},\alpha,\beta) &= \frac{qf(\boldsymbol{x}|k=n,\alpha,\beta)}{\sum_{k=1}^n \frac{1-q}{n-1} f(\boldsymbol{x}|k,\alpha,\beta) + qf(\boldsymbol{x}|k=n,\alpha,\beta)} \\ &= \left[1 + \frac{1-q}{q(n-1)} \sum_{j=1}^{n-1} B_{j0}\right]^{-1} \end{aligned}$$

and

$$\Pi(k|\boldsymbol{x},\alpha,\beta) = B_{k0}\left[\frac{q(n-1)}{1-q} + \sum_{j=1}^{n-1} B_{j0}\right]^{-1}, \quad k = 1 \cdots n-1 \tag{4.1.7}$$

where

$$B_{k0} = \frac{f(\boldsymbol{x}|k,\alpha,\beta)}{f(\boldsymbol{x}|k=n,\alpha,\beta)}$$

$$= \frac{B(y_1+\alpha, \sum_{i=1}^{k} m_i - y_1 + \beta) B(y_2+\alpha, \sum_{i=k+1}^{n} m_i - y_2 + \beta) B(\alpha,\beta)^{-1}}{B(\alpha + y, \sum_{i=1}^{n} m_i - y + \beta)}. \tag{4.1.8}$$

We can use the proper priors $\alpha, \beta \to \frac{1}{2}$ or $\alpha, \beta \to 1$ in (4.1.8), otherwise we can use partial Bayes factors with $\alpha, \beta \to 0$.

If we let $\alpha, \beta \to 0$, for the FBF it follows that

$$m_0(b) = \frac{m_0}{m_0^b} = \frac{B(y, \sum_{i=1}^{n} m_i - y)}{B(by, b(\sum_{i=1}^{n} m_i - y))} \tag{4.1.9}$$

where $y = y_1 + y_2$ and

$$m_k(b) = \frac{m_k}{m_k^b} = \frac{B\left(y_1, \sum_{i=1}^{k} m_i - y_1\right) B\left(y_2, \sum_{i=k+1}^{n} m_i - y_2\right)}{B\left(by_1, b\left(\sum_{i=1}^{k} m_i - y_1\right)\right) B\left(by_2, b(\sum_{i=k+1}^{n} m_i - y_2)\right)} \tag{4.1.10}$$

so that

$$B_{0k}^F = \frac{m_0(b)}{m_k(b)},$$

$y, y_1, y_2 > 0$ and $b \geq \frac{2}{n}$, since the minimal sample size is not unique. When $m_i = 1$, $i = 1, \ldots, n$, it becomes a Bernoulli sequence and

$$B_{0k}^F = \frac{B(y, n-y) B(by, b(k-y_1)) B(by_2, b(n-k-y_2))}{B(nb, b(n-y)) B(y_1, k-y_1) B(y_2, n-k-y_2)} \tag{4.1.11}$$

where $0 < y_1 < k$, $0 < y_2 < n-k$.

For multiple change-points $\boldsymbol{k} = [k_1, \ldots, k_r]$ it follows directly that

$$m_{\boldsymbol{k}}(b) = \frac{\prod_{j=1}^{r+1} B(y_j, m_j - y_j)}{\prod_{j=1}^{r+1} B(by_j, b(m_j - y_j))} \tag{4.1.12}$$

where

$$y_j = \sum_{i=k_{j-1}+1}^{k_j} x_i, \quad m_j = \sum_{i=k_{j-1}+1}^{k_j} m_i, \quad b \geq \frac{r+1}{n}$$

and $k_0 = 0, \quad k_{r+1} = n$, as long as $y_j, m_j - y_j > 0, \quad j = 1, \ldots, r+1$.

4.1.2 THE NEGATIVE BINOMIAL MODEL

Suppose that

$X_i \sim$ Neg Bin (r, p) with fixed r so that, for $\boldsymbol{x} = x_1, \ldots, x_n$ and $y = \sum_{i=1}^{n} x_i$,

$$f(\boldsymbol{x}|p) = \left[\prod_{i=1}^{n} \binom{x_i - 1}{r - 1}\right] p^{nr}(1-p)^{y-rn}, \quad x_i = r, r+1, \ldots$$

Furthermore, let

$$\Pi(p|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

i.e. $p \sim \text{Beta}(\alpha, \beta)$ and

$$f(\boldsymbol{x}|\alpha, \beta) = \left[\prod_{i=1}^{n} \binom{x_i - 1}{r - 1}\right] \tfrac{B(nr+\alpha, y-nr+\beta)}{B(\alpha,\beta)}. \tag{4.1.13}$$

If $X_i \sim$ Neg Bin $(r, p_1), \quad i = 1, \ldots, k$

and $X_i \sim$ Neg Bin $(r, p_2), \quad i = k+1, \ldots, n$

it follows that

$$f(\boldsymbol{x}|p_1,p_2,k) = \left[\prod_{i=1}^{n} \begin{pmatrix} x_i - 1 \\ r-1 \end{pmatrix}\right] p_1^{kr}(1-p_1)^{y_1-kr}p_2^{(n-k)r}(1-p_2)^{y_2-(n-k)r}$$

so that

$$f(\boldsymbol{x}|k,\alpha,\beta) \propto \frac{B(kr+\alpha, y_1-kr+\beta)}{B(\alpha,\beta)^2} B((n-k)r+\alpha, y_2-(n-k)r+\beta). \tag{4.1.14}$$

Furthermore

$$\Pi(k|\alpha,\beta,\boldsymbol{x}) \propto B(kr+\alpha, y_1-kr+\beta)B((n-k)r+\alpha, y_2-(n-k)r+\beta). \tag{4.1.15}$$

For the usual Bayes factor we have

$$B_{0k} = \frac{\left[\prod_{i=1}^{n} \begin{pmatrix} x_i - 1 \\ r-1 \end{pmatrix}\right] B(nr+\alpha, y-nr+\beta)B(\alpha,\beta)}{B(kr+\alpha, y_1-kr+\beta)B((n-k)r+\alpha, y_2-(n-k)r+\beta)}. \tag{4.1.16}$$

For the fractional Bayes factor $(\alpha,\beta \to 0)$ it follows that

$$m_0 = B(nr, y-nr)$$

and

$$m_0^b = B(bnr, b(y-nr))$$

so that, with a minimal sample of 2 and $b = \frac{2}{n}$, it follows that

$$m_0(b) = \frac{B(nr, y-nr)}{B(2r, 2\bar{x}-r))}. \tag{4.1.17}$$

Furthermore

$$m_k(b) = \frac{B(kr, y_1 - kr)B((n-k)r, y_2 - (n-k)r)}{B\left(\frac{2kr}{n}, 2\left(\bar{x}_1 - \frac{kr}{n}\right)\right) B\left(\frac{2(n-k)r}{n}, 2\left(\bar{x}_2 - \frac{(n-k)r}{n}\right)\right)}, \tag{4.1.18}$$

$$B^F_{0k} = \frac{m_0(b)}{m_k(b)}, \tag{4.1.19}$$

$p_i|k, \boldsymbol{x} \sim B(k_i r, y_i - k_i r)$; $i = 1, 2$ and the unconditional posterior of p_i $(\alpha, \beta \to 0)$ will be as in (4.1.4).

The results for the Geometric distribution $f(x|p) = p(1-p)^{x-1}$, $x = 1, 2, \ldots$ follows directly from the Negative Binomial when putting $r = 1$.

4.1.3 THE MULTINOMIAL MODEL

Consider the distribution

$\boldsymbol{X}_i(q \times 1) \sim$ Multinomial $(\boldsymbol{p}, m)$, $i = 1, \ldots, n$

where

$$\begin{aligned} f(X|\boldsymbol{p}) &= \prod_{i=1}^{n} \left[\frac{m!}{x_{1i}! \ldots x_{qi}!} p_1^{x_{1i}} \cdots p_q^{x_{qi}}\right] \\ &= \prod_{i=1}^{n} \left[\frac{m!}{x_{1i}! \cdots x_{qi}!}\right] \prod_{j=1}^{q} p_j^{y_j} \end{aligned} \tag{4.1.20}$$

where $y_j = \sum_i^n x_{ji}$, $j = 1, \ldots, q$.

Furthermore, let the prior on $\boldsymbol{p}$ be

$$\Pi(\boldsymbol{p}) = \frac{\Gamma\left(\sum_{j=1}^{q} \alpha_j\right)}{\prod_j^q \Gamma(\alpha_j)} \prod_j^q p_j^{\alpha_j - 1}, \quad \boldsymbol{p} \sim \text{ Dirichlet } (\alpha_1, \ldots, \alpha_q).$$

It then follows that the marginal distribution of $\boldsymbol{y}$ is

$$f(\boldsymbol{y}|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{j=1}^{q} \alpha_j\right)}{\prod_{i}^{q} \Gamma(\alpha_j)} \prod_{i=1}^{n} \left[\frac{m!}{x_{1i}! \cdots x_{ki}!}\right] \frac{\prod^{q} \Gamma(\alpha_j + y_j)}{\Gamma(\Sigma\alpha_j + mn)}. \tag{4.1.21}$$

Under the assumption of a change-point, let $\boldsymbol{X}_i(q \times 1) \sim$ Multinomial $(\boldsymbol{p}_1, m)$, $\quad i = 1, \ldots, k$ and $\boldsymbol{X}_i(q \times 1) \sim$ Multinomial $(\boldsymbol{p}_2, m)$, $\quad i = k+1, \ldots, n$.

It follows that

$$f(\boldsymbol{y}|p_1, p_2, k) = \prod_{i=1}^{n} \left[\frac{m!}{x_{1i}! \cdots x_{ki}!}\right] \prod_{j=1}^{q} p_{1j}^{y_{1j}} \prod_{j=1}^{q} p_{2j}^{y_{2j}}$$

where $y_{1j} = \sum_{i=1}^{k} x_{ji}$ and $y_{2j} = \sum_{i=k+1}^{n} x_{ji}, \quad j = 1, \ldots, q$ (4.1.22)

and with identical independent Dirichlet priors,

$$\Pi(p_1, p_2) = \frac{\Gamma^2\left(\sum_{j=1}^{q} \alpha_j\right)}{\prod^{q} \Gamma(\alpha_j)^2} \prod_{j=1}^{q} p_{1j}^{\alpha_j - 1} \prod_{j=1}^{q} p_{2j}^{\alpha_j - 1} \tag{4.1.23}$$

it follows that

$$f(\boldsymbol{y}|\boldsymbol{\alpha}) = \prod_{i=1}^{n} \left[\frac{m!}{x_{1i}! \cdots x_{ki}!}\right] \left\{\frac{\Gamma\left(\sum_{j}^{q} \alpha_j\right)}{\prod_{j}^{q} \Gamma(\alpha_j)}\right\}^2 \frac{\prod^{q} \Gamma(\alpha_j + y_{1j}) \prod^{q} \Gamma(\alpha_j + y_{2j})}{\Gamma(\Sigma\alpha_j + mk)\Gamma(\Sigma\alpha_j + m(n-k))}.$$

Then

$$\Pi(k|\boldsymbol{y}, \boldsymbol{\alpha}) \propto \frac{\prod^{q} \Gamma(\alpha_j + y_{1j}) \prod^{q} \Gamma(\alpha_j + y_{2j})}{\Gamma(\Sigma\alpha_j + mk)\Gamma(\Sigma\alpha_j + m(n-k))}. \tag{4.1.24}$$

The usual Bayes factor in favour of no change will be

$$B_{0k} = \frac{\prod^{q} \Gamma(\alpha_j + y_j) \prod_{j}^{q} \Gamma(\alpha_j)\Gamma(\Sigma\alpha_j + mk)\Gamma(\Sigma\alpha_j + m(n-k))}{\Gamma(\Sigma\alpha_j + mn)\Gamma(\sum_{j=1}^{q} \alpha_j) \prod^{q} \Gamma(\alpha_j + y_{1j}) \prod^{q} \Gamma(\alpha_j + y_{2j})}, \quad k = 1, \ldots, n-1 \tag{4.1.25}$$

and for the fractional Bayes factor it follows that

$$m_0^b = \frac{\prod^{q} \Gamma(by_j)}{\Gamma(bmn)}$$

so that

$$m_0(b) = \frac{\prod^{q} \Gamma(y_j)\Gamma(bmn)}{\Gamma(mn) \prod^{q} \Gamma(by_j)} \tag{4.1.26}$$

and

$$m_k^b = \frac{\prod^{q} \Gamma(by_{1j}) \prod^{q} \Gamma(by_{2j})}{\Gamma(bmk)\Gamma(bm(n-k))}$$

so that

$$m_k(b) = \frac{\prod^{q} \Gamma(y_{1j}) \prod^{q} \Gamma(y_{2j})\Gamma(bmk)\Gamma(bm(n-k))}{\Gamma(mk)\Gamma(m(n-k)) \prod^{q} \Gamma(by_{1j}) \prod^{q} \Gamma(by_{2j})} \tag{4.1.27}$$

and

$$B_{0k}^F = \frac{m_0(b)}{m_k(b)} = \frac{\prod^{q} \Gamma(y_j)\Gamma(bmn)\Gamma(mk)\Gamma(m(n-k)) \prod^{q} \Gamma(by_{1j}) \prod^{q} \Gamma(by_{2j})}{\Gamma(mn) \prod^{q} \Gamma(by_j) \prod^{q} \Gamma(y_{1j}) \prod^{q} \Gamma(y_{2j})\Gamma(bmk)\Gamma(bm(n-k))} \tag{4.1.28}$$

for given b. For $\alpha_j = 0, \quad j = 1, \ldots, q$, the minimal sample size is not unique, so it is not clear what b should be. The minimal sample size depends on the observed values. However

B_{0k}^F is finite for any $b > 0$ in (4.1.28) and any k for which $y_{ij} > 0$ for all i, j. The influence of b is illustrated in example 4.3.3. Notice that $\Sigma\Sigma x_{ji} = nm$, $\Sigma p_i = 1$ and that $\sum_j^q x_{ji} = m$. The posterior probabilities for k follow from (4.1.7), with $q = \frac{1}{2}$.

Furthermore, conditional on k,

$$\boldsymbol{p}_i|k, \boldsymbol{\alpha}, \boldsymbol{y} \sim \text{Dirichlet}\ (y_{i1} + \alpha_1, \ldots, y_{iq} + \alpha_q), \quad i = 1, 2. \tag{4.1.29}$$

The unconditional posterior of p (with $\alpha_j \to 0$) follows as

$$\Pi(\boldsymbol{p}_i|\boldsymbol{y}) \propto \sum_{k=1}^{n-1} \Pi(\boldsymbol{p}_i|k, \boldsymbol{y})\Pi(k|\boldsymbol{y}). \tag{4.1.30}$$

4.1.4 THE MARKOV CHAIN MODEL

Carlin, Gelfand and Smith (1992) were the first to examine the Markov chain change-point problem from a Bayesian viewpoint, using Gibbs sampling. They suppose a sequence of n observations $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ from a process which is a p-state stationary Markov chain having either transition matrix A or precisely one change to a transition matrix B. The entries of A are $a_{ij} = P(Y_{t+1} = j|Y_t = i)$ whence $a_{ij} \geq 0$, $\Sigma_j a_{ij} = 1$; similarly for B with entries b_{ij}. They take independent Dirichlet priors on the rows of A and similar for B. The multinomial change-point problem occurs as a special case when the Y_i's are independent and $a_{ij} = a_j$ and $b_{ij} = b_j$.

Let's start with the simplest Markov chain, i.e. when $p = 2$.

$$A = \begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \left[\begin{array}{cc} a_{11} & 1 - a_{11} \\ 1 - a_{22} & a_{22} \end{array}\right] \end{array}.$$

Let the Beta prior before the change be

$$\Pi(A) = D_{\lambda_1}(a_1)D_{\lambda_2}(a_2)$$

where

$$D_{\lambda_1}(a_1) = \frac{\Gamma(\lambda_{11}+\lambda_{12})}{\Gamma(\lambda_{11})\Gamma(\lambda_{12})} a_{11}^{\lambda_{11}-1}(1-a_{11})^{\lambda_{12}-1}, \tag{4.1.31}$$

and after the change be

$$\Pi(B) = D_{\gamma_1}(b_1)D_{\gamma_2}(b_2).$$

Under model M_0 of no change,

$$f(\boldsymbol{y}|A) = p(y_1)\prod_{t=1}^{n-1} a_{y_t,y_{t+1}}$$

and where $p(y_1)$ is the initial state probabilities, the joint distribution is

$$\begin{aligned} f(\boldsymbol{y},A) &= f(y|A)\Pi(A) \\ &= p(y_1)\prod_{t=1}^{n-1} a_{y_t,y_{t+1}} \frac{1}{B(\lambda_{11},\lambda_{12})} a_{11}^{\lambda_{11}-1}(1-a_{11})^{\lambda_{12}-1}\,. \\ &\quad \frac{1}{B(\lambda_{21},\lambda_{22})} a_{22}^{\lambda_{22}-1}(1-a_{22})^{\lambda_{21}-1}. \end{aligned} \tag{4.1.32}$$

Example:

If $Y = 1 \;\; 1 \;\; 2 \;\; 1 \;\; 2 \;\; 2 \;\; 1 \;\; 1 \;\; 2$, then

$$\begin{aligned} \prod_{t=1}^{n-1} a_{y_t,y_{t+1}} &= a_{11}(1-a_{11})(1-a_{22})(1-a_{11})a_{22}(1-a_{22})a_{11}(1-a_{11}) \\ &= a_{11}^2(1-a_{11})^3 a_{22}(1-a_{22})^2 \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{z}_1 &= (z_{11} \;\; z_{12}) = (2 \;\; 3) \\ \boldsymbol{z}_2 &= (z_{21} \;\; z_{22}) = (2 \;\; 1) \end{aligned}$$

where z_{ij} is the number of transitions from i to j in the $n-1$ steps.

So

$$f(\boldsymbol{y}, A) \propto a_{11}^{\lambda_{11}+z_{11}-1}(1-a_{11})^{\lambda_{12}+z_{12}-1}a_{22}^{\lambda_{22}+z_{22}-1}(1-a_{22})^{\lambda_{21}+z_{21}-1}p(y_1)$$

and

$$\Pi(A|\boldsymbol{y}) = D_{\lambda_1+\boldsymbol{z}_1}(a_1)D_{\lambda_2+\boldsymbol{z}_2}(a_2).$$

With vague priors, $\lambda_{11} = \lambda_{12} = \lambda_{21} = \lambda_{22} = 1$:

$$\begin{aligned} \Pi(A|\boldsymbol{y}) &= \Pi(a_{11}|y)\Pi(a_{22}|y) \\ &= \text{Beta}(z_{11}+1, z_{12}+1) \text{ Beta}(z_{21}+1, z_{22}+1). \end{aligned}$$

The marginal likelihood of $\boldsymbol{y}$ (with $\lambda_{ij} = 0$) is

$$\begin{aligned} m_0 &= \iint a_{11}^{z_{11}-1}(1-a_{11})^{z_{12}-1}a_{22}^{z_{22}-1}(1-a_{22})^{z_{21}-1}da_{11}da_{22} \\ &= \prod_{i=1}^{2} \frac{\Gamma(z_{i1})\Gamma(z_{i2})}{\Gamma(z_{i1}+z_{i2})} \end{aligned}$$

and

$$m_0^b = \prod_{i=1}^{2} \frac{\Gamma(bz_{i1})\Gamma(bz_{i2})}{\Gamma(bz_{i1}+bz_{i2})}$$

where b can be $\frac{1}{\sqrt{n}}$ and where all $z_{ij} > 0$ (unless $\lambda_{ij} > 0 \quad \forall \; i, j$). This is not really usable unless the chain is long enough to have transitions from every state to every other state.

<u>Under model M_k</u> it follows that

$$f(\boldsymbol{y}|A, B, k) = p(y_1) \prod_{t=1}^{k-1} a_{y_t, y_{t+1}} \prod_{t=k}^{n-1} b_{y_t, y_{t+1}}$$

and

$$\begin{aligned} f(y, A, B, k) \;\; \propto \;\; & f(\boldsymbol{y}|A, B, k)a_{11}^{\lambda_{11}-1}(1-a_{11})^{\lambda_{12}-1}a_{22}^{\lambda_{22}-1}(1-a_{22})^{\lambda_{21}-1} \\ & b_{11}^{\gamma_{11}-1}(1-b_{11})^{\gamma_{12}-1}b_{22}^{\lambda_{22}-1}(1-b_{22})^{\lambda_{21}-1}\Pi(k) \end{aligned}$$

where $2 \leq k \leq n-1$ for D_λ and D_γ proper densities.

Furthermore

$$\Pi(A|\boldsymbol{y}, k) = D_{\lambda_1+\boldsymbol{z}'_1}(a_1) D_{\lambda_2+\boldsymbol{z}'_2}(a_2)$$

and

$$\Pi(B|\boldsymbol{y}, k) = D_{\gamma_1+\boldsymbol{z}''_1}(b_1) D_{\gamma_2+\boldsymbol{z}''_2}(b_2)$$

and

$$\Pi(k|\boldsymbol{y}) \propto \Pi(k) \prod_{i=1}^{2} B(\lambda_{i1} + z'_{i1}, \lambda_{i2} + z'_{i2}) B(\gamma_{i1} + z''_{i1}, \gamma_{i2} + z''_{i2}), \tag{4.1.33}$$

and all $z'_{ij}, z''_{ij} > 0$, where z'_{ij} is the number of transitions in $Y_1, \ldots, Y_k$ and z''_{ij} is the number in $Y_k, \ldots, Y_n$.

The marginal likelihood of $\boldsymbol{y}$ (with $\gamma_{ij} = \lambda_{ij} = 0$) is

$$m_k = \prod_{i=1}^{2} B(z'_{i1}, z'_{i2}) B(z''_{i1}, z''_{i2}) \tag{4.1.34}$$

and

$$m_k^b = \prod_{i=1}^{2} B(bz'_{i1}, bz'_{i2}) B(bz''_{i1}, bz''_{i2}). \tag{4.1.35}$$

The fractional BF is then

$$B_{0k}^F = \frac{m_0 m_k^b}{m_0^b m_k}.$$

The ordinary BF with $\lambda_{ij} = \gamma_{ij} = 1$ (uniform prior) is

$$B_{0k} = \prod_{i=1}^{2} \frac{B(z_{i1}+1, z_{i2}+1)}{B(z'_{i1}+1, z'_{i2}+1) B(z''_{i1}+1, z''_{i2}+1)}. \tag{4.1.36}$$

To generalize to $p > 2$ $\quad (\lambda_{ij} = \gamma_{ij} = 1)$, it follows that

$$m_0 = \Gamma^p(p) \prod_{i=1}^{p} \frac{\prod_{j=1}^{p} \Gamma(z_{ij}+1)}{\Gamma\left(\sum_{i=j}^{p} z_{ij} + p\right)} \tag{4.1.37}$$

and

$$m_k = \Gamma^{2p}(p) \prod_{i=1}^{p} \frac{\prod_{j=1}^{p} \Gamma(z'_{ij}+1)}{\Gamma\left(\sum_{j} z'_{ij} + p\right)} \prod_{i}^{p} \frac{\prod_{j=1}^{p} \Gamma(z''_{ij}+1)}{\Gamma\left(\sum_{j} z''_{ij} + p\right)} \tag{4.1.38}$$

so that

$$B_{0k} = \frac{m_0}{m_k}.$$

Also notice that

$$\Pi(k|y) \propto m_k \Pi(k).$$

The marginal posterior of say a_{11} is, for given k,

$$a_{11}|k \sim \text{Beta}\left(z'_{11} + 1, \sum_{j=2}^{p} z'_{1j} + p - 1\right) \tag{4.1.39}$$

and similarly for the other elements.

Multiple change-points

Let's consider the transition matrices $A_1, A_2, \ldots, A_{R+1}$ where R is the maximum number of possible change-points. Then

$$\Pi(A_\ell) = \prod_{i=1}^{p} D_{\lambda_{\ell i}}(a_{\ell i}), \quad \ell = 1, \ldots, R+1$$

and where

$$D_{\lambda_{\ell_i}}(a_{\ell_i}) = \frac{\Gamma\left(\sum_j^p \lambda_{\ell_{ij}}\right)}{\prod^p \Gamma(\lambda_{\ell_{ij}})} \prod_{j=1}^p a_{\ell_{ij}}^{\lambda_{\ell_{ij}}-1}. \tag{4.1.40}$$

Under $M_{\boldsymbol{k}}^r$, where $\boldsymbol{k} = (k_1 \ldots, k_r)$ and $r = 0, \ldots, R$, it follows that

$$f(\boldsymbol{y}|r, \boldsymbol{k}, A_\ell, \ell = 1, \ldots, r+1) = \prod_{\ell=0}^r \left(\prod_{t=k_\ell}^{k_{\ell+1}-1} a_{\ell+1|y_t, y_{t+1}} \right) p(y_1)$$

where $k_0 = 1, \quad k_{r+1} = n$.

With $\lambda_{\ell ij} = 1$, $\Pi(r) = \frac{1}{R+1}$ and $\boldsymbol{k}|r$ uniform, it follows that

$$f(\boldsymbol{y}, r, \boldsymbol{k}, A_\ell) = \prod_i^p \prod_j^p a_{1ij}^{z_{1ij}} \prod_i^p \prod_j^p a_{2ij}^{z_{2ij}} \prod_i^p \prod_j^p a_{r+1,i,j}^{z_{r+1,ij}} p(y_1)[\Gamma(p)]^{p(R+1)} \frac{1}{R+1} \Pi(k|r) \tag{4.1.41}$$

and that

$$f(\boldsymbol{y}|k, r) = p(y_1)[\Gamma(p)]^{p(R+1)} \prod_{\ell=1}^r \prod_i^p \frac{\prod_j^p \Gamma(z_{\ell ij}+1)}{\Gamma\left(\sum_j^p z_{\ell ij} + p\right)}. \tag{4.1.42}$$

Under m_0, it follows that

$$f(y) = p(y_1)[\Gamma(p)]^p \prod_i^p \frac{\prod_j^p \Gamma(z_{0ij}+1)}{\Gamma(\Sigma z_{0ij} + p)}$$

where $z_{\ell ij}$ is the number of transitions from state i to state j of the observations between k_ℓ and $k_{\ell+1} - 1$. Then

$$B_{0\boldsymbol{k}}^r = \frac{1}{[\Gamma(p)]^{pR}} \frac{\prod \frac{\Pi\Gamma(z_{0ij}+1)}{\Gamma(\Sigma z_{0ij}+p)}}{\prod\prod \frac{\Pi\Gamma(z_{\ell ij}+1)}{\Gamma(\Sigma z_{0ij}+p)}}. \tag{4.1.43}$$

4.2 EXPONENTIAL TYPE MODELS

4.2.1 THE POISSON MODEL

Raftery and Akman (1986) developed a Bayesian approach to estimate and test for a Poisson process with a change-point, considering the change-point to be continuous. Carlin, Gelfand and Smith (1992) presented a general approach to hierarchical Bayes change-point models. In particular, desired marginal posterior densities are obtained utilizing the Gibbs sampler. They include an application to changing Poisson processes, applied to the coal mining disaster data of Jarrett (1979). Raftery and Akman (1986) also use the coal mining disaster data.

There has been speculation that the number of cases of diarrhoea-associated haemolytic uraemic syndrome increased abruptly during the early part of the 1980's. Henderson and Matthews (1993) investigate this hypothesis and applied change-point models for Poisson variables to two series of data from regional referral units in Newcastle upon Tyne and Birmingham.

Using a direct resampling process, Broemeling and Gregurich (1996) developed a Bayesian approach for the analysis of the change-point problem. They implemented a direct sampling approach whereby standard random number generators can be used to generate samples from the joint posterior distribution of all the parameters in the model. They illustrated this technique with examples involving one shift for Poisson processes and regression models.

Let's first consider the model with exactly one discrete change-point

$$X_i \sim \text{ Poisson } (\lambda_1); \quad i = 1, \ldots, k$$

$$X_i \sim \text{ Poisson } (\lambda_2); \quad i = k+1, \ldots, n.$$

The likelihood function is

$$L(\lambda_1, \lambda_2, k|y) = \frac{e^{-\lambda_1 k}}{\prod_{i=1}^{n} x_i!} \lambda_1^{y_1} \; e^{-(n-k)\lambda_2} \lambda_2^{y_2}, \quad 1 \leq k \leq n-1$$

where

$$y_1 = \sum_{i=1}^{k} x_i \text{ and } y_2 = \sum_{i=k+1}^{n} x_i. \tag{4.2.1}$$

Assuming λ_1, λ_2, k are independent a priori and that the prior densities have the conjugate form

$$\Pi(\lambda_1, \lambda_2 | \alpha, \beta) = \frac{\beta^{2\alpha}}{\Gamma^2(\alpha)} \lambda_1^{\alpha-1} \lambda_2^{\alpha-1} e^{-\beta(\lambda_1+\lambda_2)} \tag{4.2.2}$$

and we have a uniform prior on k so that

$$f(y|k, \alpha, \beta) \propto \frac{\Gamma(\alpha + y_1)\Gamma(\alpha + y_2)}{(k + \beta)^{\alpha+y_1}(n - k + \beta)^{\alpha+y_2}} \tag{4.2.3}$$

and

$$\Pi(k|y, \alpha, \beta) = \frac{f(y|k, \alpha, \beta)}{\sum_{k=1}^{n-1} f(y|k, \alpha, \beta)}.$$

If we let $\alpha \rightarrow 0$ and $\beta \rightarrow 0$ so that $\pi(\lambda_1, \lambda_2) \propto \frac{1}{\lambda_1 \lambda_2}$, it follows that

$$\Pi(k|y) \propto \Gamma(y_1)\Gamma(y_2) k^{-y_1} (n - k)^{-y_2}. \tag{4.2.4}$$

Furthermore

$$\lambda_1 | y, k, \alpha, \beta \sim \Gamma(\alpha + y_1, k + \beta) \tag{4.2.5}$$

and

$$\lambda_2 | y, k, \alpha, \beta \sim \Gamma(\alpha + y_2, n - k + \beta). \tag{4.2.6}$$

Also

$$\Pi(\lambda_i|y,\alpha,\beta) = \sum_k \Pi(\lambda_i|y,k,\alpha,\beta)\Pi(k|y,\alpha,\beta), \quad j=1,2. \tag{4.2.7}$$

For $\tau = \frac{\lambda_1}{\lambda_2}$ it follows that

$$\Pi(\tau|y,k,\alpha,\beta) \propto \left[1 + \frac{k}{n-k+\beta}\tau\right]^{-(2\alpha+y)} \tau^{\alpha+y_1-1} \tag{4.2.8}$$

so that

$$\frac{2(\alpha+y_2)k}{2(\alpha+y_1)(n-k+\beta)}\tau|y,k \sim F_{v_1,v_2} \tag{4.2.9}$$

where

$$v_1 = 2(\alpha+y_1) \text{ and } v_2 = 2(\alpha+y_2) \tag{4.2.10}$$

and that

$$\Pi(\tau|y) = \sum_k \Pi(\tau|y,k,\alpha,\beta)\Pi(k|y,\alpha,\beta). \tag{4.2.11}$$

Considering the possibility of no change, let $\Pi(k) = \begin{cases} q; & k=n \\ \frac{1-q}{n-1}; & k=1,\ldots,n-1 \end{cases}$

so that

$$f(y|k=n,\alpha,\beta) = \frac{\beta^\alpha\Gamma(\alpha+y)}{\Gamma(\alpha)\prod^n x_i!(n+\beta)^{\alpha+y}}$$

where

$$y = \sum_{i=1}^n x_i. \tag{4.2.12}$$

Then the posterior probability of no change follows as

$$\Pi(k=n|y) \quad = \quad \frac{qf(y|k=n)}{\sum^n \frac{1-q}{n-1}f(y|k) + qf(y|k=n)}$$

$$= \left[1 + \frac{1-q}{q(n-1)} \sum_{k=1}^{n-1} B_{k0}\right]^{-1}$$

(4.2.13)

and

$\Pi(k|y)$ is the same as in (4.1.7) where

$$B_{k0} = \frac{f(y|k,\alpha,\beta)}{f(y|k=n,\alpha,\beta)} = \frac{\beta^{\alpha}\Gamma(\alpha+y_1)\Gamma(\alpha+y_2)(n+\beta)^{\alpha+y}}{\Gamma(\alpha)(k+\beta)^{\alpha+y_1}(n-k+\beta)^{\alpha+y_2}\Gamma(\alpha+y)}. \tag{4.2.14}$$

We can't let $\alpha, \beta \longrightarrow 0$ (using vague priors) because of the normalizing constants, but can do it through partial Bayes factors.

With vague priors, for the fractional BF it follows that

$$m_0 = \frac{\Gamma(y)}{\Pi x_i! n^y}, \tag{4.2.15}$$

$$m_0^b = \frac{\Gamma(by)}{(nb)^{by}(\Pi x_i!)^b},$$

$$m_0(b) = \frac{\Gamma(y) b^{by} n^{-y(1-b)}}{\Gamma(by)(\Pi x_i!)^{1-b}} \tag{4.2.16}$$

and

$$m_k(b) = \frac{m_k}{m_k^b} = \frac{\Gamma(y_1)\Gamma(y_2) b^{by} k_1^{y_1(b-1)} k_2^{y_2(b-1)}}{\Gamma(by_1)\Gamma(by_2)(\Pi x_i!)^{1-b}} \tag{4.2.17}$$

so that

$$B_{0k}^F = \frac{m_0(b)}{m_k(b)} = \frac{B(by_1, by_2)}{B(y_1, y_2)} \left(\frac{k_1}{n}\right)^{y_1(1-b)} \left(\frac{k_2}{n}\right)^{y_2(1-b)}. \tag{4.2.18}$$

If we let $b = \frac{2}{n}$ it follows that

$$B_{0k}^F = \frac{n^{\frac{-y(n-2)}{n}} B\left(\frac{2y_1}{n}, \frac{2y_2}{n}\right)}{k_1^{\frac{-y_1(n-2)}{n}} k_2^{\frac{-y_2(n-2)}{n}} B(y_1, y_2)}. \tag{4.2.19}$$

For r possible change-points, we have

$$B_{0\boldsymbol{k}}^F = \frac{\Gamma(y) n^{-y(1-b)} \prod_{i=1}^{r+1} \Gamma(by_i)}{\Gamma(by) \prod_{i=1}^{r+1} \Gamma(y_i) \prod_{i=1}^{r+1} k_i^{y_i(b-1)}} \tag{4.2.20}$$

where $b = \frac{r+1}{n}$ and $\boldsymbol{k} = (k_1 \ldots k_{r+1})$ with $y_i > 0, \quad i = 1, \ldots, r+1$.

The posterior is as before in (2.5.16) and (2.5.17).

For the intrinsic BF for one change-point it follows that

$$B_{k0}^N = B(y_1, y_2) \left(\frac{k}{n}\right)^{-y_1} \left(1 - \frac{k}{n}\right)^{-y_2}, \tag{4.2.21}$$

and with minimal sample size of 2 we have

$$B_{0k}(\ell) = \frac{1}{B(x_{(1)}, x_{(2)})} \left(\frac{1}{2}\right)^{x_{(1)}+x_{(2)}} \tag{4.2.22}$$

where $x_{(1)}$ is an observation before the change-point, and $x_{(2)}$ an observation after the change-point.

The intrinsic BF then follows from (1.10) and (1.11).

4.2.2 THE GAMMA MODEL

Diaz (1982) considered a Bayesian detection of a change of scale parameter in sequence of independent gamma random variables. He stated two problems: the first is the detection of the change, while the second is the estimation of the change-point and the two scale parameters under the assumption that a change has occurred. He assumed α known but uses ordinary Bayes factors with improper priors, which does not allow for the indeterminancy created by the improper priors.

Hsu (1979) presented a classical (non-Bayesian) asymptotic solution to the above mentioned first problem and made a review of other solutions proposed that can be used only for large samples.

Let's consider the model

$$
\begin{aligned}
X_i &\sim \Gamma(\alpha, \beta_1); \quad i = 1, \ldots, k \\
X_i &\sim \Gamma(\alpha, \beta_2); \quad i = k+1, \ldots, n \qquad (k = 1, \ldots, n-1).
\end{aligned}
$$

Assume α is known

The likelihood function will be

$$f(x|\beta_1, \beta_2, k) = \frac{\beta_1^{k_1\alpha}\beta_2^{k_2\alpha}}{\Gamma^n(\alpha)} \prod_{i=1}^{n} x_i^{\alpha-1} e^{-\beta_1 y_1 - \beta_2 y_2} \tag{4.2.23}$$

where $y_1 = \sum_{i=1}^{k_1} x_i$ and $y_2 = \sum_{i=k_1+1}^{n} x_i$, $k_1 = k$, $k_2 = n-k$.

The priors on β_1 and β_2 are assumed to be independent Gamma densities

$$\Pi(\beta_1, \beta_2) \propto \beta_1^{a-1} e^{-d\beta_1} \beta_2^{a-1} e^{-d\beta_2}.$$

Then

$$f(\boldsymbol{x}|k, \alpha, a, d) = \frac{\Gamma(k_1\alpha + a)\Gamma(k_2\alpha + a)}{(y_1 + d)^{k_1\alpha+a}(y_2 + d)^{k_2\alpha+a}} \frac{d^{2a}}{\Gamma^n(\alpha)\Gamma^2(a)} (\prod_i x_i)^{\alpha-1} \tag{4.2.24}$$

where $k_1 = k$ and $k_2 = n - k$. Then, with uniform prior on k,

$$\Pi(k|y, \alpha, a, d) \propto f(\boldsymbol{x}|k, \alpha, a, d).$$

If $a, d \to 0$, i.e. $\pi(\beta_1, \beta_2) \propto \frac{1}{\beta_1\beta_2}$, it follows that

$$\Pi(k|y,\alpha) \propto \frac{\Gamma(k_1\alpha)\Gamma(k_2\alpha)}{y_1^{k_1\alpha} y_2^{k_2\alpha}} \tag{4.2.25}$$

and that

$$\beta_i|y,\alpha,k \sim \Gamma(k_i\alpha, y_i), \qquad i = 1, 2. \tag{4.2.26}$$

Under the assumption of no change,

$X_i \sim \Gamma(\alpha, \beta), \quad i = 1, \ldots, n$, so

$$f(x|\alpha) = \frac{\Gamma(n\alpha + a)d^a}{(y+d)^{n\alpha+a}\Gamma^n(\alpha)\Gamma(a)} \prod_i x_i^{\alpha-1}. \tag{4.2.27}$$

The usual Bayes factor in favour of no change

$$B_{0k} = \frac{\Gamma(n\alpha + a)\Gamma(a)(y_1+d)^{k_1\alpha+a}(y_2+d)^{k_2\alpha+a}}{(y+b)^{n\alpha+a}d^a\Gamma(k_1\alpha+a)\Gamma(k_2\alpha+a)}. \tag{4.2.28}$$

If we let $a, d \to 0$, the Fractional Bayes factor follows as

$$m_0(b) = \frac{m_0}{m_0^b} = \frac{\Gamma(n\alpha)}{\Gamma(bn\alpha)\Gamma^{(1-b)n}(\alpha)}(\Pi x_i)^{(1-b)(\alpha-1)} y^{-(1-b)n\alpha} \tag{4.2.29}$$

and

$$m_k(b) = \frac{m_k}{m_k^b} = \frac{\Gamma(k_1\alpha)\Gamma(k_2\alpha)}{\Gamma(bk_1\alpha)\Gamma(bk_2\alpha)\Gamma^{(1-b)n}(\alpha)}(\Pi x_i)^{(1-b)(\alpha-1)} y_1^{-(1-b)k_1\alpha} y_2^{-(1-b)k_2\alpha} \tag{4.2.30}$$

so that

$$B_{0k}^F = \frac{m_0(b)}{m_k(b)} = \frac{\Gamma(n\alpha)\Gamma(bk_1\alpha)\Gamma(bk_2\alpha)y_1^{(1-b)k_1\alpha}y_2^{(1-b)k_2\alpha}}{\Gamma(k_1\alpha)\Gamma(k_2\alpha)\Gamma(bn\alpha)y^{(1-b)n\alpha}} \tag{4.2.31}$$

where we'll again take $b = \frac{2}{n}$.

The generalization to multiple change-points follows directly as

$$B^r_{0\boldsymbol{k}} = \frac{\Gamma(n\alpha)}{y^{(1-b)n\alpha}} \prod_{j=1}^{r+1} \frac{\Gamma(bk_j\alpha)}{\Gamma(k_j\alpha)} y_j^{(1-b)k_j\alpha}$$

for r change-points.

Assume α unknown

When α is unknown, there is no standard way to eliminate α. Then the following possible solutions can be considered, i.e.:

(1) Determine $\Pi(k|y,\alpha)$ and plot it as a function of α for $k = 1, \ldots, n-1$. In this way upper and lower bounds can be established.

(2) Estimate α by some empirical Bayes procedure and then replace α by $\hat{\alpha}_k$.

(3) Putting a prior on α on a bounded interval, $0 < \alpha < K$ and integrate numerically.

Consider again the likelihood function in (4.2.23). Assuming now a change-point at k, the moments estimator of α from the two groups are $\frac{\bar{x}_1^2}{s_1^2}$ and $\frac{\bar{x}_2^2}{s_2^2}$ respectively, where $\bar{x}_i$ and s_i^2 are the sample means and variances. Our estimator is then the weighted mean

$$\hat{\alpha}_k = \frac{k_1\frac{\bar{x}_1^2}{s_1^2} + k_2\frac{\bar{x}_2^2}{s_2^2}}{n}. \qquad (4.2.32)$$

Under no change the estimator will be $\hat{\alpha}_0 = \frac{\bar{x}^2}{s^2}$. So for every value of k we will have a different estimator of α.

If we assume a change in α as well as in β, then

$$f(\boldsymbol{x}|\alpha_1,\alpha_2,\beta_1,\beta_2,k) = \frac{\beta_1^{k_1\alpha}\beta_2^{k_2\alpha}}{\Gamma^{k_1}(\alpha_1)\Gamma^{k_2}(\alpha_2)} \left(\prod_1^{k_1} x_i\right)^{\alpha_1-1} \left(\prod_{k+1}^{n} x_i\right)^{\alpha_2-1} e^{-\beta_1 y_1 - \beta_2 y_2}. \qquad (4.2.33)$$

Then

$$f(\boldsymbol{x}|\alpha_1, \alpha_2, k) = \frac{\Gamma(k_1\alpha_1)\Gamma(k_2\alpha_2)\left(\prod_{1}^{k_1} x_i\right)^{\alpha_1 - 1}\left(\prod_{k_2}^{n} x_i\right)^{\alpha_2 - 1}}{y_1^{k_1\alpha_1} y_2^{k_2\alpha_2}\Gamma^{k_1}(\alpha_1)\Gamma^{k_2}(\alpha_2)} \tag{4.2.34}$$

and for a given k the α's can be estimated as above so that

$$\hat{\alpha}_{1k} = \frac{\bar{x}_1^2}{s_1^2} \text{ and } \hat{\alpha}_{2k} = \frac{\bar{x}_2^2}{s_2^2} \tag{4.2.35}$$

or integrated numerically.

The Fractional Bayes factor follows then similarly to equation (4.2.29) to (4.2.31). The posterior probabilities follows as in (4.1.7).

4.2.2.1 THE EXPONENTIAL MODEL

When $\alpha = 1$ (fixed) in paragraph 4.2.2, the result for the standard exponential distribution follows directly from (4.2.23) to (4.2.31).

If censored observations $x_1^*, x_2^*, \ldots, x_m^*$ are present, y_1 and y_2 are replaced by $y_1 + y_1^*$ and $y_2 + y_2^*$ in equations (4.2.23) to (4.2.31), where

$$y_1^* = \sum_{i=1}^{k^*} x_i^*, \quad y_2^* = \sum_{i=k^*+1}^{m} x_i^*$$

and k^* are the number of censored observations below k.

Worsley (1986) gives classical tests for change-points in the more general setting of exponential families of random variables, but with particular emphasis on the exponential distribution.

4.2.2.1.1 THE LEFT TRUNCATED EXPONENTIAL MODEL

A variation on the exponential distribution is the truncated exponential distribution. Jani and Pandya (1999) observed a sequence of independent lifetimes $X_i, \ldots, X_k, X_{k+1}, \ldots, X_n$ from left truncated exponential populations. They let

$$f(\boldsymbol{x}|\beta) = \beta e^{-\beta(x-\eta)}, \quad x \geq \eta$$

with reliability function

$$\begin{aligned} R(t) &= P[X \geq t] \\ &= \beta e^{\beta\eta} \int_t^{\infty} e^{-\beta x} dx \\ &= e^{-\beta(t-\eta)}. \end{aligned} \tag{4.2.36}$$

Assuming a change-point in the reliability at k so that $R_1(t) = e^{-\beta_1}(t-\eta)$ for $X_1, \ldots, X_k$ and $R_2(t) = e^{-\beta_2}(t-\eta)$ for $X_{k+1}, \ldots, X_n$, Jani and Pandya (1999) assume marginal prior distributions of the reliability levels at a common prefixed time τ to be log inverse gamma distributions, i.e.

$$\Pi(r_{i\tau}) = \frac{b_i^{a_i}}{\Gamma(a_i)} r_{i\tau}^{b_i-1} \left[\ell n \left(\frac{1}{r_{i\tau}}\right)\right]^{a_i-1}, \quad a_i, b_i > 0; \; 0 \leq r_{i\tau} \leq 1; i = 1, 2. \tag{4.2.37}$$

They put a vague prior $\Pi(\eta) = \frac{1}{\tau}, \; 0 \leq \eta \leq \tau$ on η. The marginal posterior distributions on the change-point k were obtained as

$$\Pi(k|\boldsymbol{x}) = \frac{M_k J(k_1)}{\sum\limits_{k_1=1}^{n-1} M_k J(k_1)} \tag{4.2.38}$$

where

$$J(k_1) = \frac{\int_0^C (\tau-\eta)^{a_1+a_2}}{(A^{k_1+a_1} B^{k_2+a_2})} d\eta$$

with

$$A = y_1 - k_1\eta + b_1(\tau-\eta), \quad B = y_2 - k_2\eta + b_2(\tau-\eta),$$

$$C = \min\ (X_1, X_2, \ldots, X_n, \tau) \text{ and } M_k = \Gamma(k_1 + a_1)\Gamma(k_2 + a_2). \tag{4.2.39}$$

To find the fractional Bayes factor for the problem of a possible change in the parameter β, let's first consider the model with a change-point at k, that is

$$f(\boldsymbol{x}|\beta_1, \beta_2, \eta, k) = \beta_1^{k_1}\beta_2^{k_2}e^{-\beta_1(y_1 - k_1\eta)}e^{-\beta_2(y_2 - k_2\eta)} \tag{4.2.40}$$

with the priors

$$\Pi(\eta) \propto 1, \quad \eta \geq 0 \text{ and } \Pi(\beta_1, \beta_2) \propto \tfrac{1}{\beta_1\beta_2}, \quad \beta_1, \beta_2 > 0. \tag{4.2.41}$$

Then

$$\Pi(k|\boldsymbol{x}) \propto f(\boldsymbol{x}|k) \propto \int_0^{\min(x)} \frac{\Gamma(k_1)\Gamma(k_2)}{(y_1 - k_1\eta)^{k_1}(y_2 - k_2\eta)^{k_2}} d\eta = m_k, \qquad k = 1, \ldots, n-1 \tag{4.2.42}$$

and

$$m_k^b = \int_0^{\min(x)} \frac{\Gamma(bk_1)\Gamma(bk_2)}{[b(y_1 - k_1\eta)]^{k_1}[b(y_2 - k_2\eta)]^{k_2}} d\eta. \tag{4.2.43}$$

Under the model with no change-point we have

$$m_0^b = \int_0^{\min(x)} \frac{\Gamma(bk)}{[b(y - k\eta)]^n} d\eta \tag{4.2.44}$$

so that

$$m_0(b) = \frac{m_0}{m_0^b} \quad \text{and} \quad m_k(b) = \frac{m_k}{m_k^b}$$

and

$$B_{0k}^F = \frac{m_0(b)}{m_k(b)} \text{ as before.}$$

The conditional posteriors of the parameters under the change-point model follow as

$$\beta_i|\boldsymbol{x}, \eta, k \sim \Gamma(k_i, y_i - k_i\eta), \qquad i = 1, 2$$

and

$$\Pi(\eta|\boldsymbol{x}, k) \propto \frac{\Gamma(k_1)\Gamma(k_2)}{(y_1 - k_1\eta)^{k_1}(y_2 - k_2\eta)^{k_2}}, \quad 0 \le \eta < \min(\boldsymbol{x}).$$

The unconditional densities can be obtained by averaging over $\Pi(k|\boldsymbol{x})$ in (4.2.42).

4.2.2.1.2 AN EPIDEMIC CHANGE

Ramanayake and Gupta (1998) and Yao (1993) considered a sequence of independent exponential random variables that is susceptible to a change in the mean. They wanted to test whether the mean has been subjected to a change after an unknown point, for an unknown duration in the sequence, before returning to the original value. They called this the epidemic change model. The likelihood ratio statistic was derived.

Let $X_1, \ldots, X_n$ be a sequence of independent exponential random variables. Consider the models M_0 and M_1 with the following means:

$$M_0 : \theta_i = \theta_0; \quad i = 1, \ldots, n$$

and

$$M_1 : \theta_i = \begin{cases} \theta_1; & i \le k_1 \\ \theta_2; & k_1 < i \le k_2 \\ \theta_1; & k_2 < i \le n, \end{cases}$$

where k_1, k_2 are the unknown change-points such that $1 \le k_1 < k_2 \le n - 1$ while θ_1 and θ_2 are the unknown parameters such that $\theta_1, \theta_2 > 0$.

The density function under M_1 will be

$$f(x_i|\theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_1} e^{-\frac{x_i}{\theta_1}}, & i \le k_1 \text{ or } i > k_2 \\ \frac{1}{\theta_2} e^{-\frac{x_i}{\theta_2}}, & k_1 < i \le k_2. \end{cases} \tag{4.2.45}$$

Let the priors be uniform $1 \leq k_1 < k_2 \leq n-1, \quad \Pi(\theta_1, \theta_2) \propto \frac{1}{\theta_1\theta_2}$.

The likelihood function will be

$$f(\boldsymbol{x}|\theta_1, k_1, k_2, \theta_2) \propto \theta_1^{-(k_1+n-k_2)} e^{-\frac{t_1}{\theta_1}} \theta_2^{-(k_2-k_1)} e^{-\frac{t_2}{\theta_2}} \tag{4.2.46}$$

where

$$t_1 = \sum_{i=1}^{k_1} x_i + \sum_{i=k_2+1}^{n} x_i \quad \text{and} \quad t_2 = \sum_{i=k_1+1}^{k_2} x_i.$$

The joint posterior of k_1, k_2 is

$$\Pi(k_1, k_2|\boldsymbol{x}) \propto \Gamma(k_2 - k_1) t_2^{-(k_2-k_1)} \Gamma(n + k_1 - k_2) t_1^{-(n+k_1-k_2)}, \qquad 1 \leq k_1 < k_2 \leq n-1. \tag{4.2.47}$$

For the Fractional Bayes factor it follows that

$$m_{k_1k_2} = \Pi(k_1, k_2|\boldsymbol{x})$$

and that

$$m^b_{k_1k_2} = \Gamma(bk_2 - bk_1) t_2^{-(bk_2-bk_1)} \Gamma[b(n + k_1 - k_2)] t_1^{-b(n+k_1-k_2)}. \tag{4.2.48}$$

$m_0(b)$ follows as in (4.2.29), with $\alpha = 1$.

The only difference between this model and the exponential model with two change-points is the condition that the mean returns to the initial value after the second change-point.

The conditional and unconditional posteriors of $\delta = \theta_2 - \theta_1$ follow from numerical integration of (4.2.46), where

$$\Pi(\delta|\boldsymbol{x}, k_1, k_2) \propto \int_{-\delta}^{\infty} f(\boldsymbol{x}|\theta, k_1, k_2, \delta)\Pi(\theta_1)d\theta_1$$

and

$$\Pi(\delta|\boldsymbol{x}) \propto \sum_{\boldsymbol{k}} \Pi(\delta|\boldsymbol{x}, k_1, k_2). \tag{4.2.49}$$

With censored observations, t_1 and t_2 are replaced by $t_1 + t_1^*$ and $t_2 + t_2^*$ in equations (4.2.46) to (4.2.49), where t_1^* and t_2^* are again the partial sums of the censored observations as in paragraph 4.2.2.1.

4.3 APPLICATIONS

EXAMPLE 4.3.1

As an example of the Binomial model, we will use the simulated set of observations first discussed by Page (1955). This data set was also analyzed by Smith (1975) and Pettitt (1979). Forty observations were taken, the first twenty were simulated with $p_1 = 0.5$ and the last twenty were simulated with $p_2 = 0.84$. The full data set consists therefore of $n = 40$ observations with a change-point at $k = 20$.

The posterior probability of k, given a change-point, by using equation (4.1.3) is given in Figure 4.1. A maximum probability of 0.2130 at $k = 17$ is obtained with uniform priors $\alpha, \beta = 1$. Smith (1975) also got a change-point at $k = 17$, with probability of 0.367, but assumed that $p_1 < p_2$.

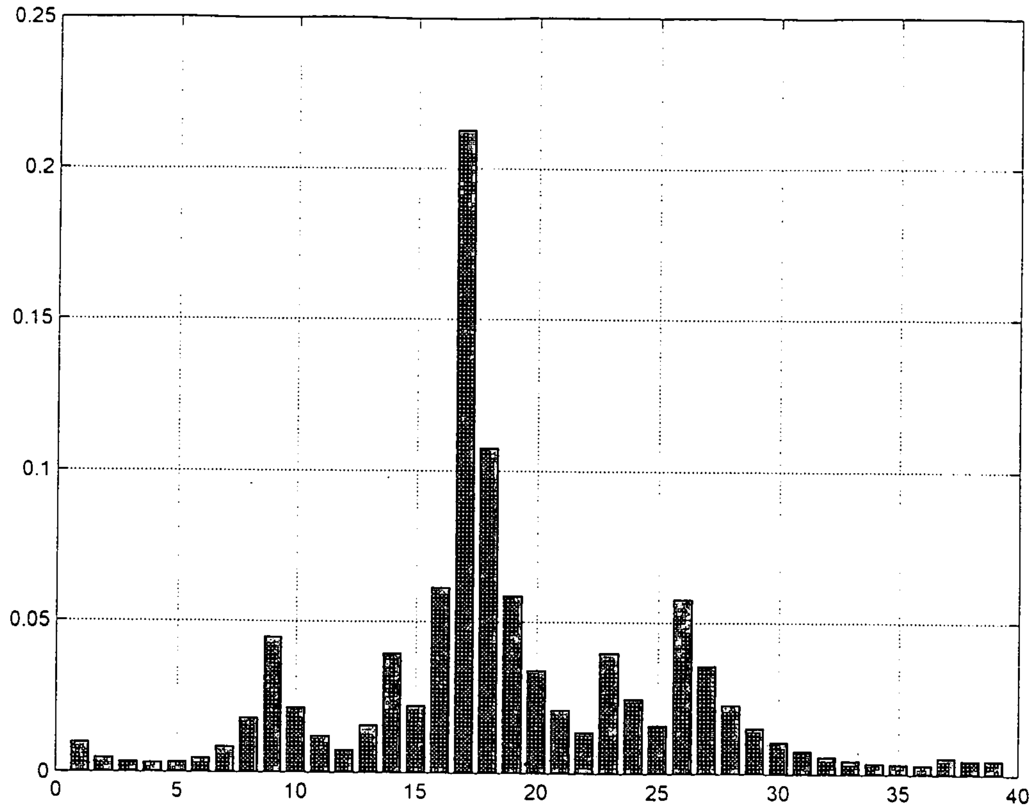


Figure 4.1: Posterior probability of k, given a change-point for example 4.3.1

By using equations (4.1.6) and (4.1.7) with $\alpha, \beta = 1$, the posterior probability for no change is 0.2003 and the posterior probability for $k = 17$ is 0.1703.

By using the FBF in equation (4.1.17) to (4.1.19), the posterior probability for no change is 0.5368 and the posterior probability for $k = 17$ is 0.1072.

The conditional (given $k = 17$) and unconditional posteriors of p_1 and p_2, by using equation (4.1.4), are given in Figure 4.2, while the posterior distribution of $\tau = \frac{p_1}{p_2}$, given $k = 17$ (by using equation (4.1.5)), is given in Figure 4.3.

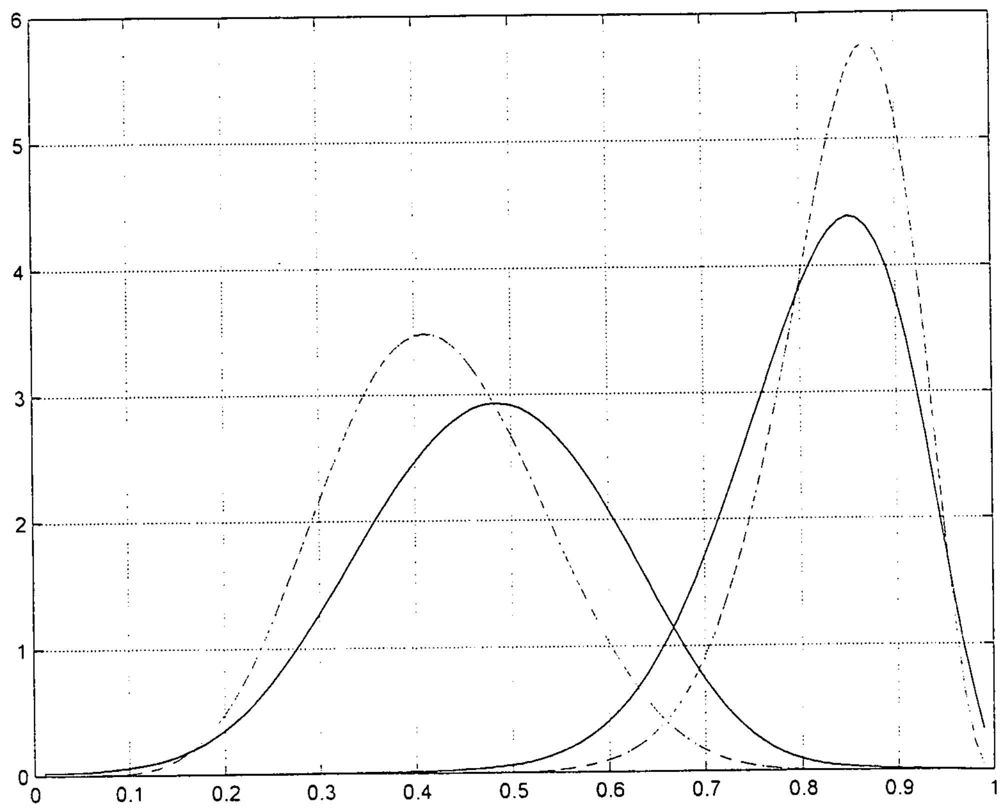


Figure 4.2: The conditional (given $k = 17$), $--$) and unconditional ($-$) posteriors of p_1 and p_2 for example 4.3.1

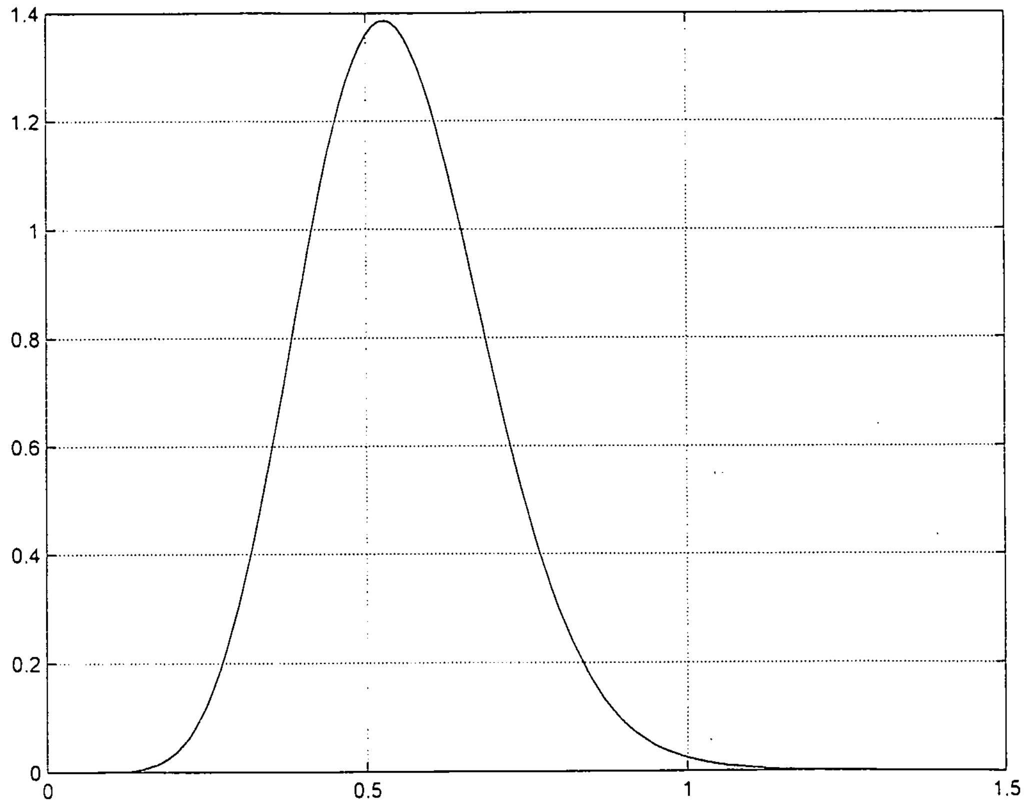


Figure 4.3: The posterior distribution of $\tau = \frac{p_1}{p_2}$, given $k = 17$ for example 4.3.1.

We get the same change-point as Smith (1975) ($k = 17$), but the FBF finds little evidence of a change-point. Pettitt (1979) and Page (1955) found significance for a one-sided test of no change against a change at $k = 17$.

EXAMPLE 4.3.2

As a second example of the Binomial model, we will use the Lindisfarne Scribe's data given originally by Ross (1950) and subsequently analyzed by Silvey (1956). The data refer to the number of occurrences of present indicative third person singular endings "$-s$" and "$-\partial$" for different sections of Lindisfarne. It is believed different scribes used the endings "$-s$"

and "$-\partial$" in different proportions.

According to Pettitt (1979) a change occurred after the 6th section, but Smith (1980) found evidence of two change-points, after the 6th and 7th sections. Pettitt used a non-parametric approach, while Smith used a Bayesian approach. Kashiwagi (1991) also considered the Scribe's data.

Stephens (1994) also analyzed the Lindisfarne Scribe's data assuming two change-points, but the data given in Stephens differs from that given by Pettitt. Pettitt gives 18 data points, while Stephens only gives 13. We will consider the data given by Stephens.

Firstly considering the possibility of a maximum of three change-points, using equation (4.1.12), we get the probabilities as in Table 4.1.

Table 4.1

Number of change-points	0	1	2	3
Probability	0.0155	0.4437	0.3363	0.2046

Assuming two change-points, the most likely pair seems to be (4,5) with probability of 0.2559.

Assuming one change-point, both equation (4.1.7) with $\alpha, \beta = 1$ and the FBF (equations (4.1.9) and (4.1.10)) indicate a change after the 5th or 6th section. The probabilities are given in Table 4.2. For his data set, Pettitt (1979) also suggested a change-point after the sixth section.

Table 4.2

Change-point	0	4	5	6	7	8
Usual BF	0.0536	0.0013	0.4710	0.3717	0.0635	0.0199
FBF	0.0337	0.0018	0.4614	0.3916	0.0770	0.0250

Our conclusion is that a change did occur in the Scribe's data, possibly after the 5th or 6th sections. Two changes are also a possibility, but then after the 4th and again after the 5th sections.

EXAMPLE 4.3.3

We will use the data set for the 269 cricket test match outcomes between England and Australia up to the end of the 1989 season as an example of the multinomial model. This data set, given by Colwell, Jones and Gillett (1990), takes the value E, A or D depending on whether England win, Australia win or the match is drawn.

Assuming one change-point, Figure 4.4 gives the posterior probability for the position of the change-point, using equation (4.1.24) with $\alpha_j = 0, \quad j = 1,2,3$. The position seems to be somewhere between the 46th and 51st test with maximum probability of 0.0120 at $k = 47$ (in 1897).

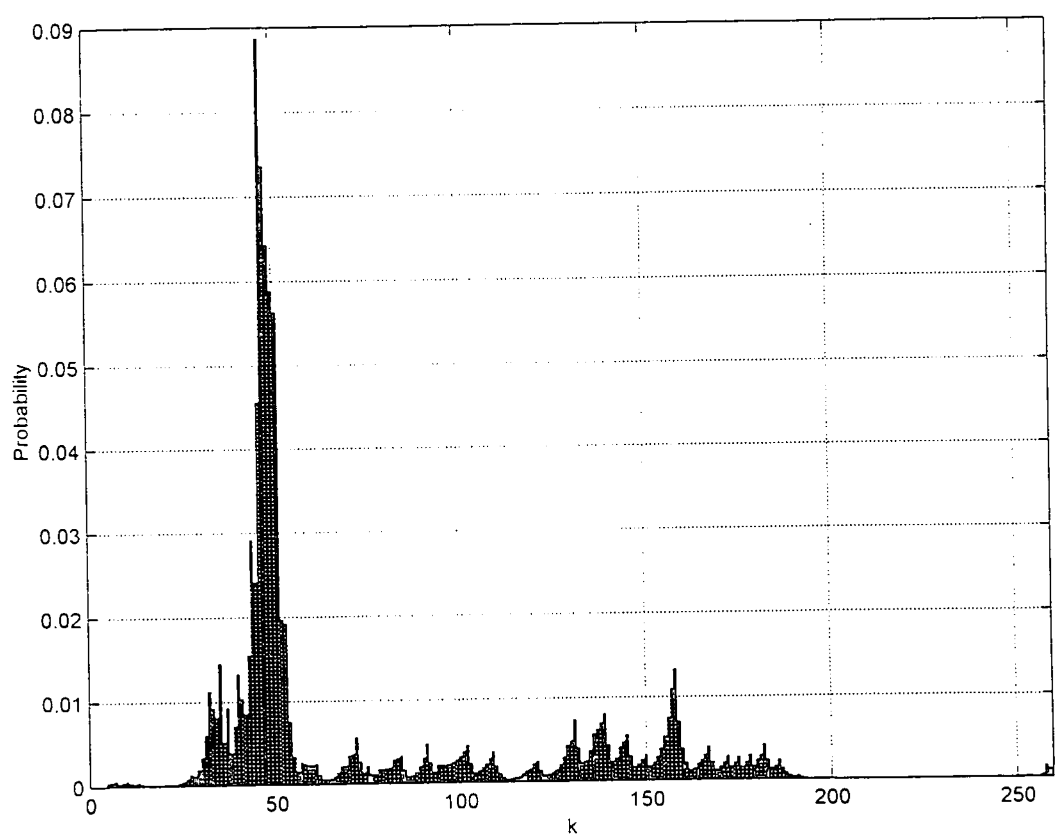


Figure 4.4: Posterior probability of k given a change-point for example 4.3.3

As far as the existence of a change-point is concerned, the FBF (using equation (4.1.28)) is very sensitive to the value of b and seen in Figure 4.5, where the posterior probability of no change is plotted as a function of b. It varies from a maximum (for $b < 0.5$) of 0.9343 at $b = \frac{1}{n} = 0.0037$ to a minimum of 0.1334 at $b = \frac{56}{n} = 0.2082$. O'Hagan (1995) suggested $b = \frac{1}{\sqrt{n}}$ when robustness is a serious concern. In that case we would have $b = 0.061$ with a probability of 0.2162 for no change. This is higher than the probability for any specific change-point, but much lower than the prior probability of 0.5, and we would consider the evidence for a change-point inconclusive. When using equation (4.1.25) with proper prior $\alpha_j = \frac{1}{2}, \quad j = 1, 2, 3$, the posterior probability of no change is 0.0115 and the maximum probability of 0.0877 is also at $k = 47$, indicating stronger evidence of a change-point.

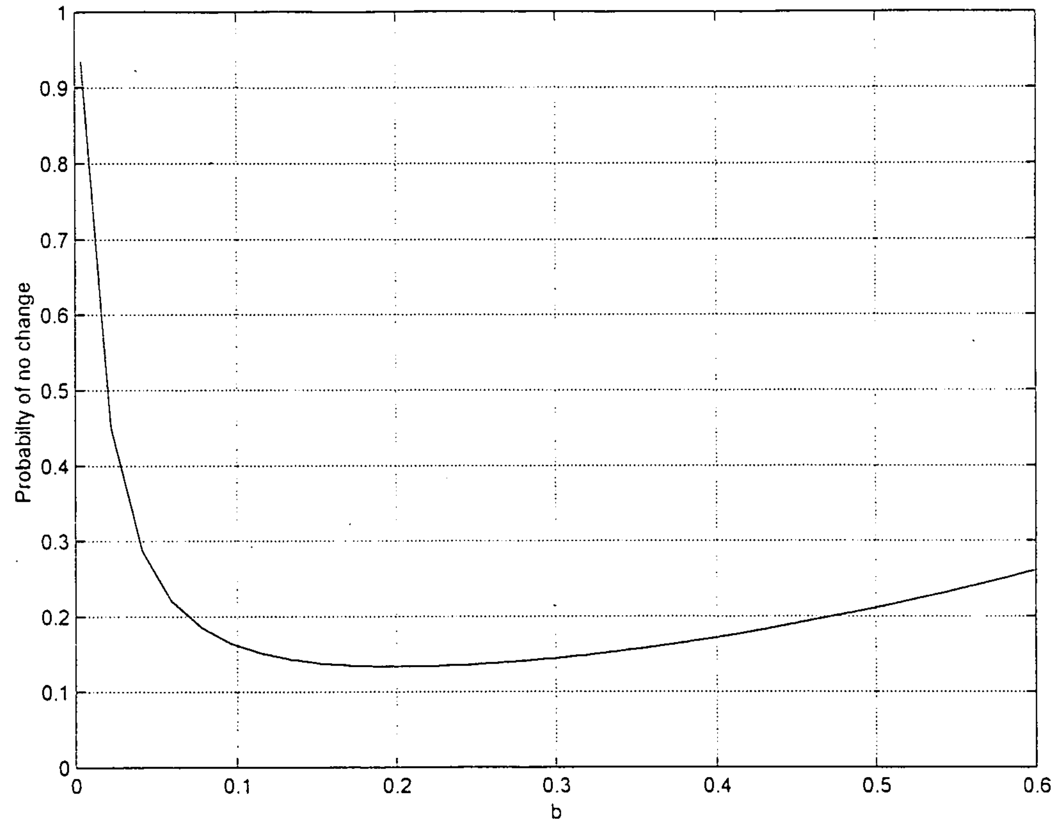


Figure 4.5: Posterior probability of no change as a function of training fraction b for example 4.3.3

Assuming a change in 1897 ($k = 47$), Figure 4.6(a) shows the marginal posterior of p_1, the probability of an Australian win, before and after the change-point. Figure 4.6(b) shows the same for p_2, the probability of an England win. We see that the mean probability of an Australian win has increased from 0.319 to 0.387, while the mean probability of an England win has decreased from 0.553 to 0.279. The mean probability of a draw has also increased from 0.128 to 0.333.

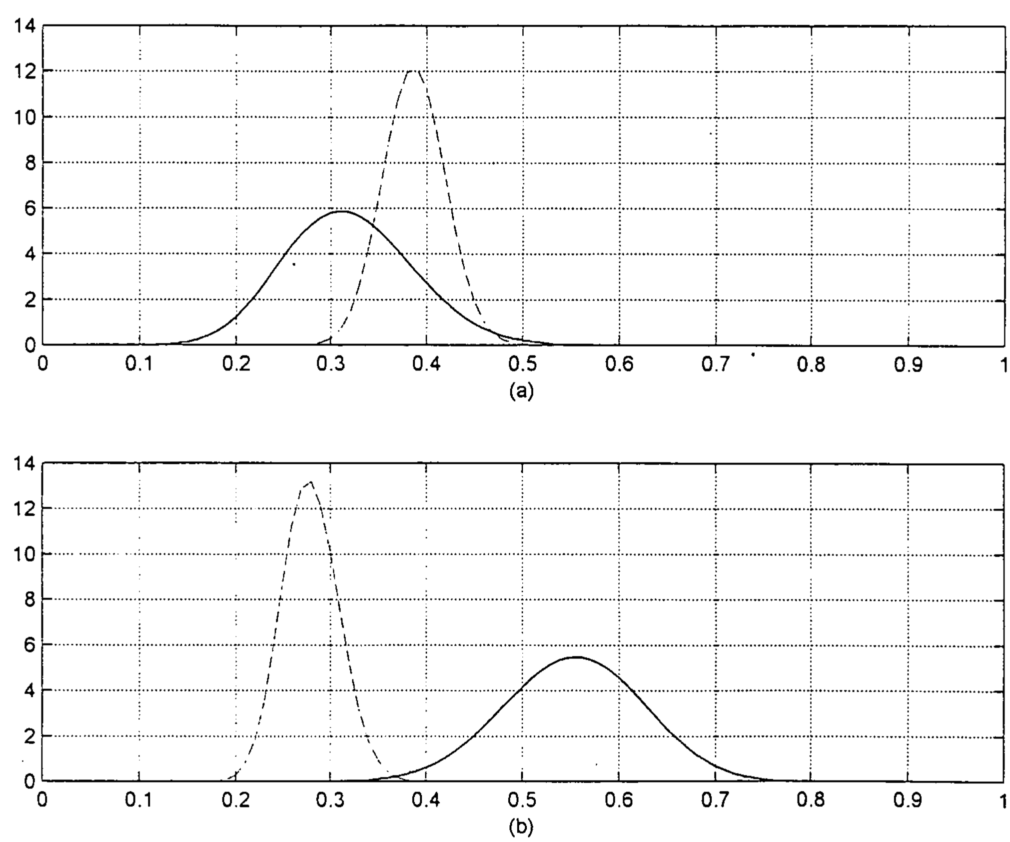


Figure 4.6: Posterior density functions of the probability of an Australian win (a) and an England win (b) before (–) and after (– –) a change-point at 1897 ($k = 47$) for example 4.3.3

EXAMPLE 4.3.4

If we model the Australian /England cricket test results of the previous example (4.3.3) as a Markov chain (as was done by Colwell, *et al.* (1990)), the resulting posterior probabilities, assuming a change-point according to equation (4.1.33), is shown in Figure 4.7. However when using the Bayes factor (equation (4.1.36)), the posterior probability for no change is 0.609, while the maximum probability for a change-point is 0.0053 at $k = 175$. Thus there is little evidence of a change-point. There is also no evidence of two or more change-points.

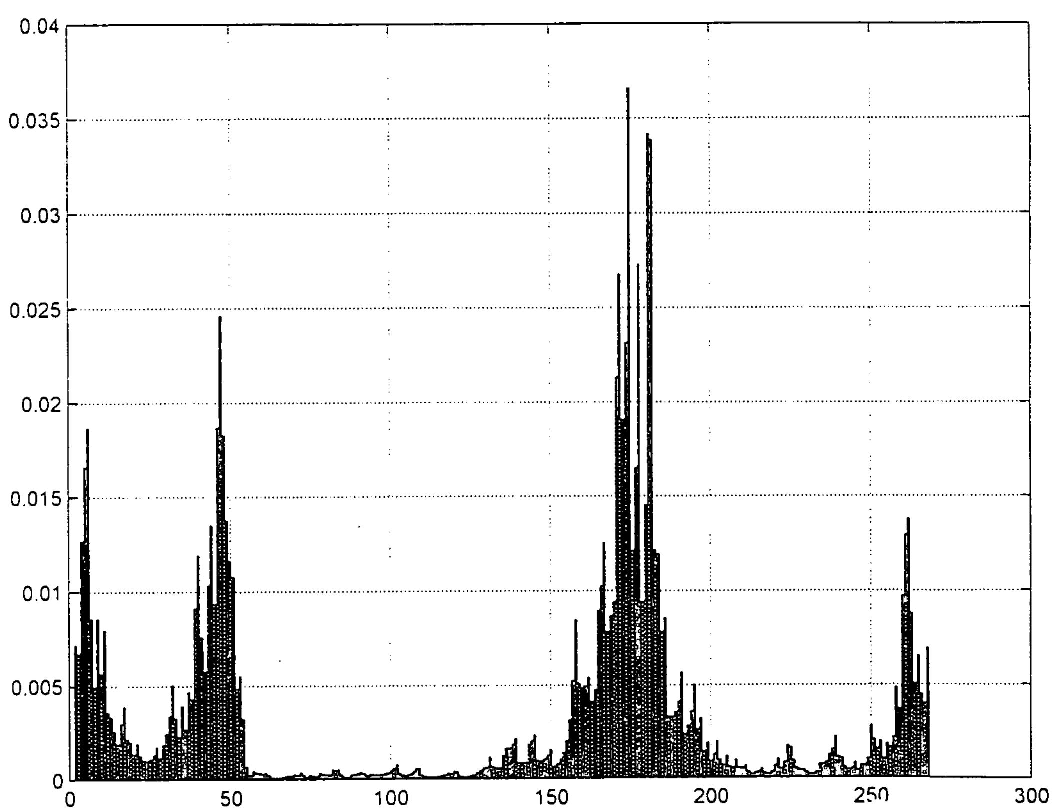


Figure 4.7: Posterior probability of k given a change-point for example 4.3.4

EXAMPLE 4.3.5

We will use the 50 observations simulated by Carlin, Gelfand and Smith (1992) as a second example of the Markov chain model. They considered a three-state stationary Markov chain where

$$A = \begin{pmatrix} 0.7000 & 0.1500 & 0.1500 \\ 0.3333 & 0.3333 & 0.3333 \\ 0.3333 & 0.3333 & 0.3333 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0.3333 & 0.3333 & 0.3333 \\ 0.1500 & 0.7000 & 0.1500 \\ 0.3333 & 0.3333 & 0.3333 \end{pmatrix},$$

and the change-point is after the 35th observation.

The posterior according to equation (4.1.36) is given in Figure 4.8 with a probability of no change of 0.152 and a maximum probability of 0.220 at $k = 33$. Carlin, Gelfand and Smith (1992) got a maximum probability of 0.34, also at $k = 33$. However, they did not consider a probability for no change.

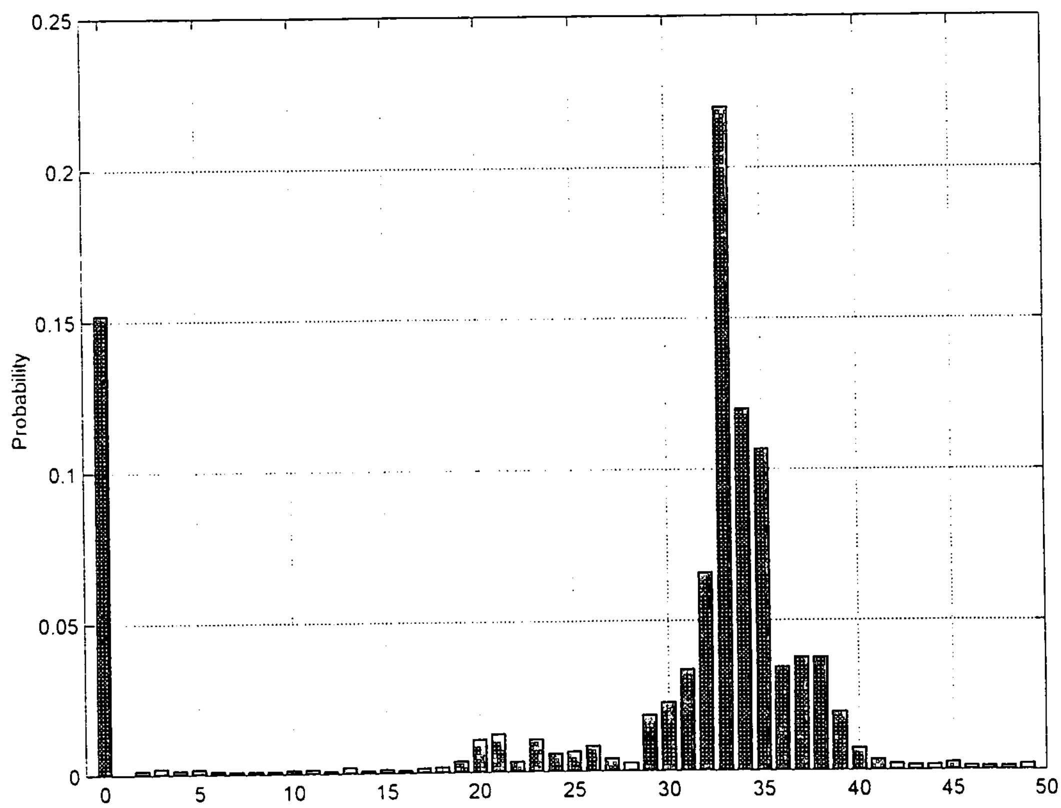


Figure 4.8: Posterior probability of a change-point for example 4.3.5

The marginal posterior density functions of the first elements of A and B, a_{11} and b_{11}, is shown in Figure 4.9, from equation (4.1.39). The expected values are $E(a_{11}|\boldsymbol{y}) = 0.652$ and $E(b_{11}|\boldsymbol{y}) = 0.306$, which corresponds well with the true values of 0.7 and 0.3333. Notice that the marginal posterior of b_{11} is one-tailed. This is because there are no transitions from state 1 to state 1 after $k = 33$.

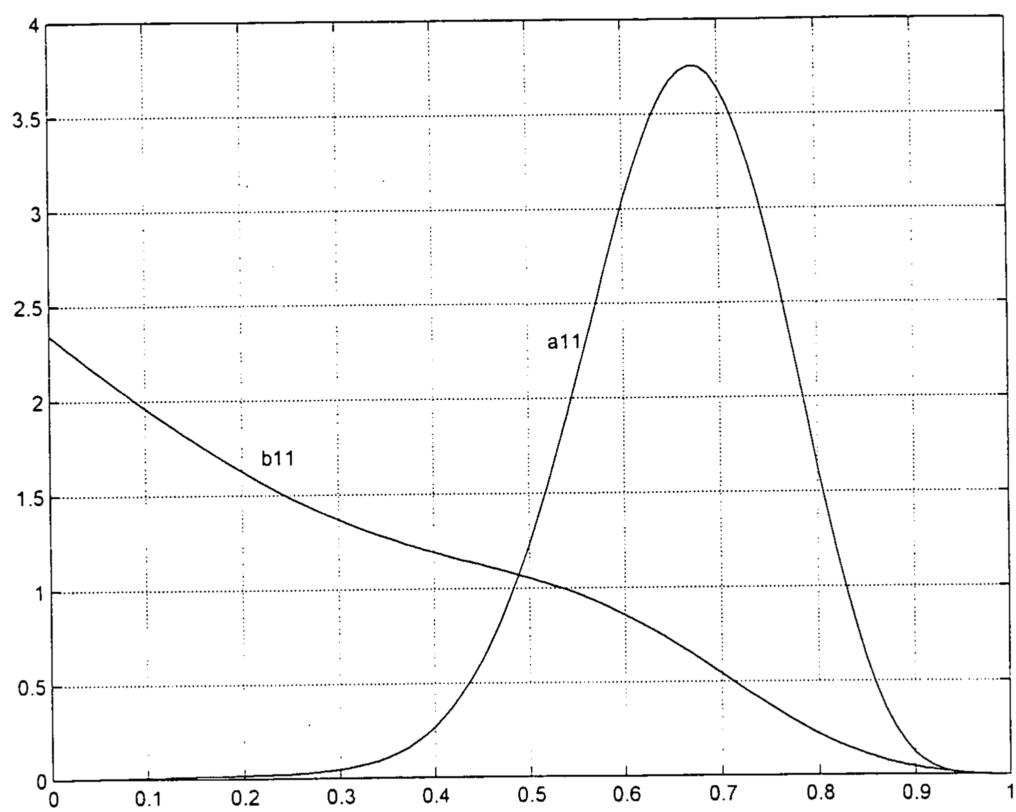


Figure 4.9: Unconditional posterior density functions of a_{11} and b_{11} for example 4.3.5

EXAMPLE 4.3.6

As an example of the Poisson model, we will use the diarrhoea-associated haemolytic uraemic syndrome (HUS) data used by Henderson and Matthews (1993). HUS is a severe, life threatening illness which predominantly affects infants and young children (Levin and Barret (1984)). There has been concern that the incidence of HUS has apparently increased sharply during the 1980's (Tarr, *et al.* (1989), Coad, *et al.* (1991)). The annual frequency of cases of HUS treated in two specialist centres in Newcastle upon Tyne and Birmingham from 1970 to 1989 is considered.

By using equation (4.2.4), we get an obvious change in both sets of data: For Newcastle at $k = 15$ (1984) and for Birmingham at $k = 11$ (1980).

The posterior probabilities assuming at most one change-point is given in Table 4.3. All approaches give essentially the same answers. Considering models with up to three change-points, by using equation (4.2.20), the data seems to favour a single change-point as seen in Table 4.4. Figure 4.10 shows the magnitude of the change $\frac{\lambda_1}{\lambda_2}$ (from equation (4.2.8)) for Newcastle and Birmingham.

So it seems that the change occurred later in Newcastle than in Birmingham and the magnitude of the change is also greater in Birmingham with a mean increase of over 6 times compared to an increase of about 5 times in Newcastle.

Henderson and Matthews (1993) compared the models from 0 to 3 possible change-points pairwise and concluded that there are 2 change-points for Birmingham at 11 and 16 (1980, 1985) and 3 change-points for Newcastle at 1,7 and 15 (1970, 1976 and 1984). Our results, however, from Table 4.4, show no evidence of this.

Table 4.3: Posterior probabilities assuming at most one change-point

		$P[\text{No change}\|X]$	$P[k = 15\|X]$
Newcastle	FBF	1.7×10^{-11}	0.9834
	AIBF	3.3×10^{-12}	0.9861
	MIBF	1.7×10^{-11}	0.9746
	GIBF	2.4×10^{-11}	0.9832
		$P[\text{No change}\|X]$	$P[k = 11\|X]$
Birmingham	FBF	1.9×10^{-13}	0.9508
	AIBF	2.7×10^{-14}	0.9591
	MIBF	8.4×10^{-14}	0.9694
	GIBF	1.9×10^{-13}	0.9648

Table 4.4: **Posterior probabilities for multiple change-points, using the FBF**

	No change	1 change-point	2 change-points	3 change-points
Newcastle	0	0.4763	0.2311	0.2926
Birmingham	0	0.4325	0.3991	0.1684

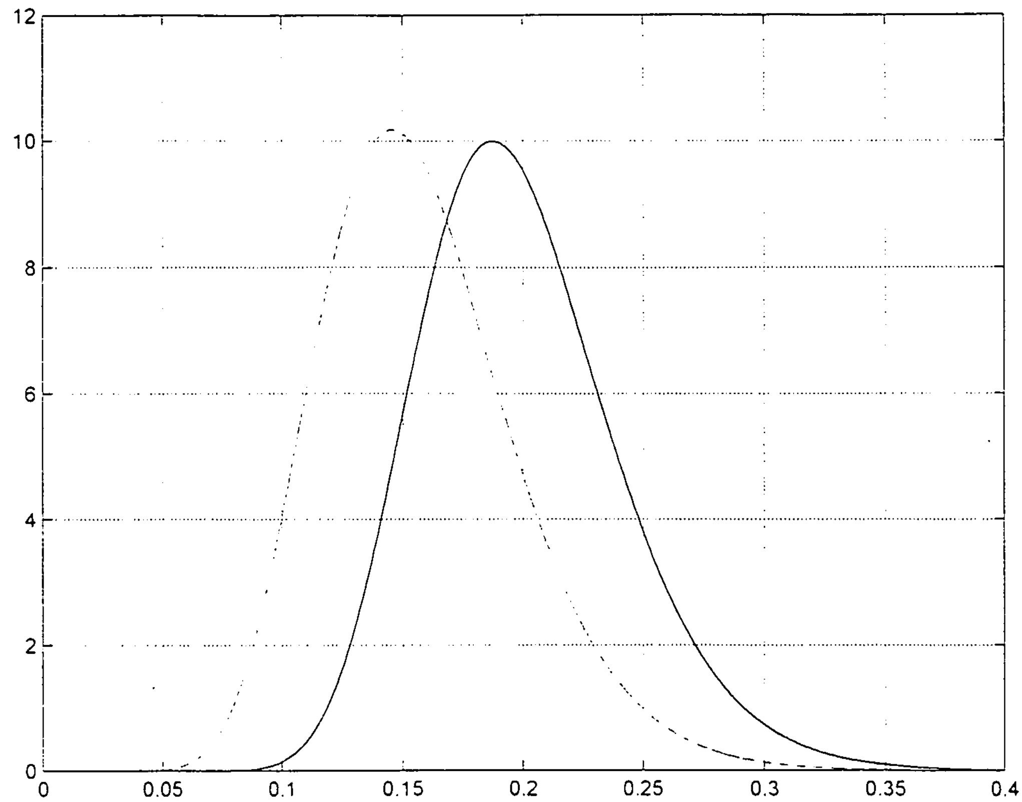


Figure 4.10: **Posterior density of the magnitude of the change, $\frac{\lambda_1}{\lambda_2}$, for Newcastle (−) and Birmingham (−−), for example 4.3.6.**

EXAMPLE 4.3.7

As a second example of the Poisson model we will use the much analyzed data set of annual number of British coal-mining disasters during the 112-year period 1851-1962 gathered by

Maguire, *et al.* (1952), extended and corrected by Jarrett (1979). Frequentist change-point investigations appear in Worsley (1986) and in Siegmund (1988), while Raftery and Akman (1986) apply their Bayesian model to investigate a continuous single change-point. Broemeling and Gregurich (1996) investigated a discrete single change-point, while Carlin, Gelfand and Smith (1992) used Gibbs sampling in examining for a single change-point. Green (1995) considered multiple change-points with reversible jump.

Assuming one change-point, Carlin, Gelfand and Smith (1992) found a maximum probability of 0.42 at 1891 ($k = 41$) with $\alpha = \frac{1}{2}$ and $\beta = 0$. The same result is obtained from equation (4.2.3). The FBF (equation (4.2.18)) gives a probability of 0.2366 at 1891, while the three fractional Bayes factors (equations (4.2.21) and (4.2.22)) yield 0.2409, 0.2575 and 0.2412.

Allowing for at most 4 change-points, the posterior probabilities from equations (4.2.20) and (2.5.17) are given in Table 4.5, together with the results of Green (1995), who used the reversible jump algorithm and a Poisson prior on k with mean 3.

Table 4.5

Posterior probability	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r \geq 5$
From FBF	5.3×10^{-14}	0.2089	0.3367	0.2620	0.1924	—
Green (1995)	0	0.157	0.348	0.266	0.149	0.080

The evidence points to 2 change-points with maximum probability at $\boldsymbol{k} = [41, 97]$, which is 1891 and 1947. Worsley (1986) and Raftery and Akman (1986) give some possible historic reasons for the possible change-points. According to Worsley changes in the coal-mining regulations during 1896 may have reduced the probability of accidents. According to Raftery and Akman a fairly abrupt decrease around 1887-1895 may be associated with changes in the coal industry around that time, namely a severe decline in labour productivity starting at the end of the 1980's, and the emergence of the Miner's Federation at the end of 1889. The change in 1947 may be due to changes in labour practices just after the war.

The joint posterior of k_1 and k_2 is shown in Figure 4.11. The posterior mass is clearly concentrated around $\boldsymbol{k}$ given above. Figure 4.12 shows the posterior distributions of λ_1, λ_2

and λ_3 with means of respectively 3.10, 1.07 and 0.27. So the number of disasters has been reduced significantly each time.

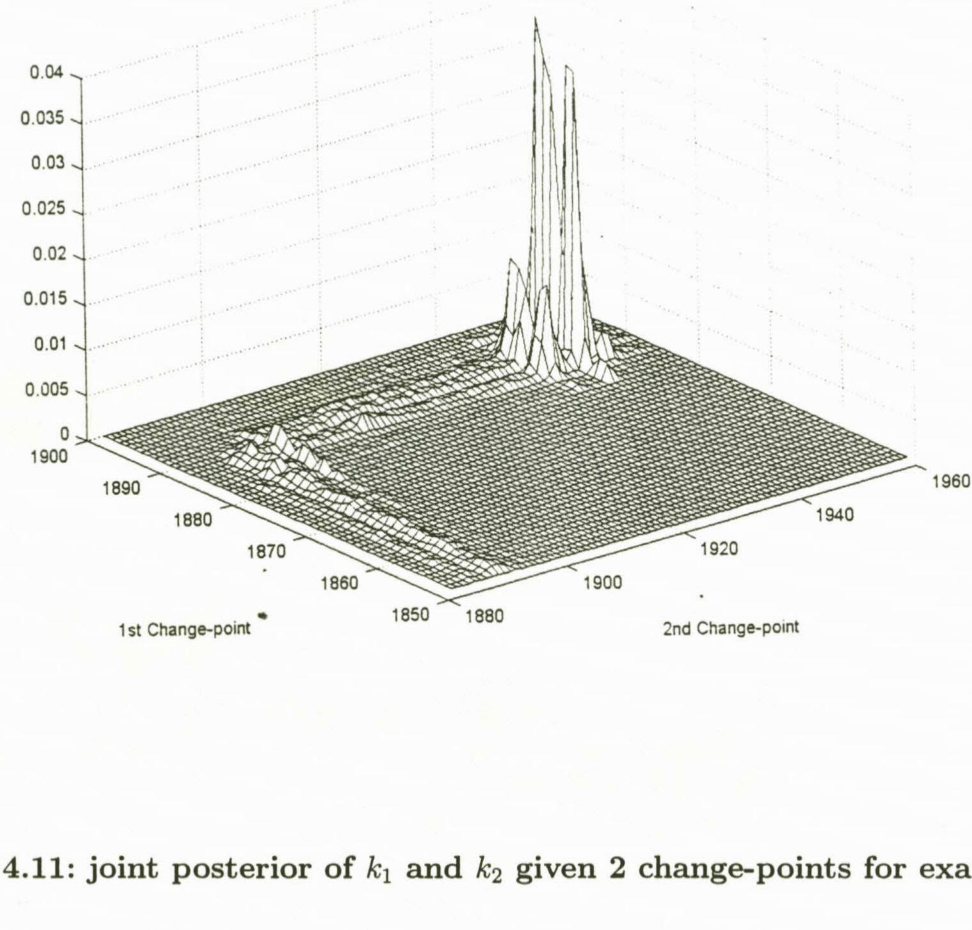


Figure 4.11: joint posterior of k_1 and k_2 given 2 change-points for example 4.3.7

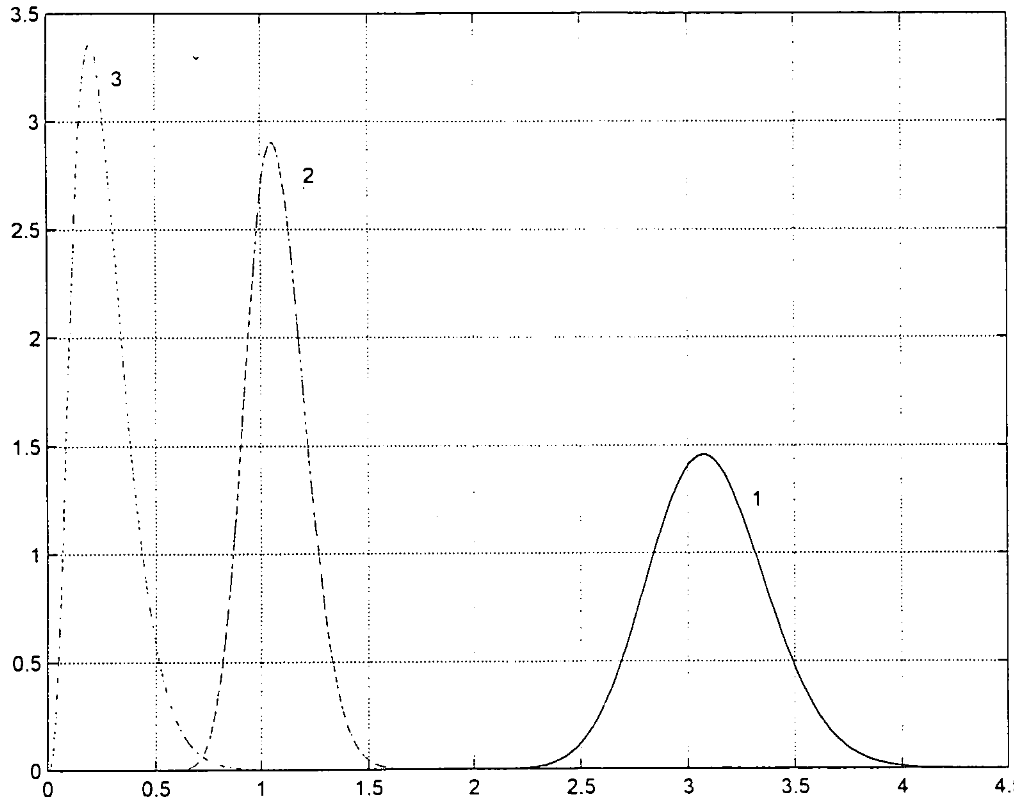


Figure 4.12: Marginal posteriors of $\lambda_1(1), \lambda_2(2)$ and $\lambda_3(3)$, the mean yearly disaster rates before, between and after 2 change-points for example 4.3.7

EXAMPLE 4.3.8

For an example of an exponential model we can also use the coal-mining disaster data as given in Jarrett (1979) in terms of time between disasters. Worsley (1986) and Raftery and Akman (1986) both found a single change-point during 1890. Assuming one change-point, we find a maximum probability of 0.2477 at $k = 124$, using equation (4.2.25), which is also during 1890. Using the FBF (equation (4.2.31)), the same probability is 0.2439 with probability of no change 10^{-14}. Raftery and Akman found the probability of no change to be 1.58×10^{-14}.

Assuming two change-points, the results are again corresponding with that in example 4.3.7, with maximum probability at $k = [124, 187]$, that is during 1890 and during 1947. The marginal posterior probability that $k = 124$ is the first change-point, is 0.1404 and the marginal posterior probability that $k = 187$ is the second change-point, is 0.2526, while the joint probability is 0.0493.

EXAMPLE 4.3.9

We will use the 20 observations artificially generated by Diaz (1982) with $\alpha = 2$ and unitary scale parameter $\beta = 1$ as an example of the Gamma model. Diaz considered three different situations. In the three cases, the first eight observations remained unchanged, while in the first case the last twelve observations of the previous list were taken, in the second and third they were multiplied by 1.5 and 3 respectively (we only multiplied by 3), in order to produce a change in the scale parameter after the eighth observation with different relative magnitude. Diaz used a prior probability of no change of 0.5.

Figure 4.13 gives the posterior probability of change-point position (with $q = 0.5$) for the Diaz data when the data has: (a) no change-point, $\alpha = 2$ (using equation (4.2.31)); (b) no change-point, α estimated (using equations (4.2.31) and (4.2.32)); (c) a change-point at $k = 8$, $\alpha = 2$ (using equation (4.2.31)); (d) a change-point at $k = 8$, α estimated (using equations (4.2.31) and (4.2.32)).

Diaz (1982) found a probability of 0.39808 for no change in the first case and a probability of 0.055 in the second case (when the last 12 observations were multiplied by 3), with a probability of 0.290 at $k = 6$.

It appears that the method of estimating the unknown α is working well when comparing Figures (a) and (b), and Figures (c) and (d).

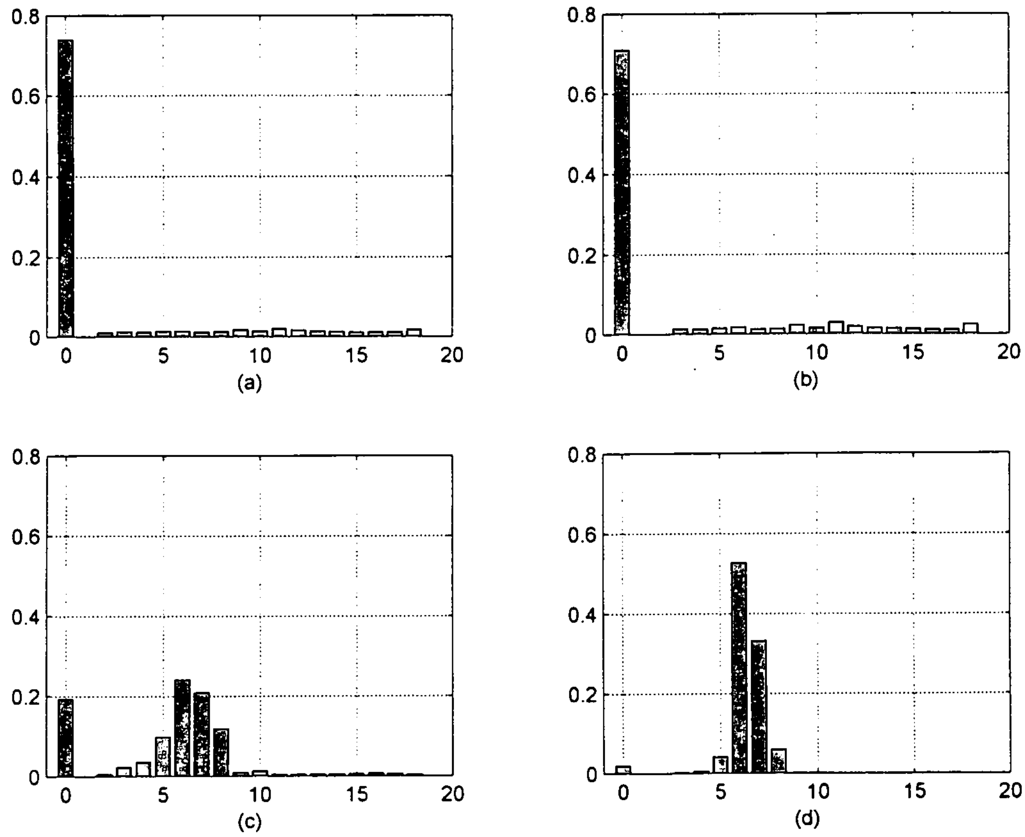


Figure 4.13: **Posterior probability of change-point position for Diaz data for (a) no change-point, $\alpha = 2$; (b) no change-point, α estimated; (c) change-point at $k = 8$, $\alpha = 2$; (d) change-point at $k = 8$, α estimated**

EXAMPLE 4.3.10

We will use the arrival times of aircraft reported to a control sector during a certain time interval as an example of the exponential model. The sample period was 16/00/00 - 24/00/00 GMT (noon - 8 P.M. New York time) on April 30, 1969. Tabulated values are in seconds from the start of the sample period. The data were originally collected by the Federal Aviation Administration, National Aviation Facilities Experimental Center, Atlantic City, New Jersey. The data seems to fit the exponential model and Hsu (1979) and Diaz (1982) also analyzed the data.

We found no evidence of change in the interarrival times, which is the same conclusion as Hsu (1979) and Diaz (1982). For the FBF with $\alpha = 1$, the probability for no change is 0.8849. With unknown α replaced by $\hat{\alpha}_k$ from equation (4.2.32), the probability for no change is 0.8543.

We also added successive observations together, creating a Gamma sequence with $\alpha = 2, 3, \ldots$. The posterior probabilities for no change stayed essentially the same as above for both known and unknown value of α.

EXAMPLE 4.3.11

As an example of an epidemic change, we will use the Stanford heart transplant data given by Kalbfleisch and Prentice (1980). This data is the survival times of potential heart transplant recipients from their date of acceptance into the Stanford heart transplant program. Ramanayake and Gupta (1998) analyzed the same data, but does not seem to make any distinction between censored and uncensored observations. Our data set contains 45 uncensored and 24 censored observations. Figure 4.14 shows the censored and uncensored observations.

The FBF (using equation (4.2.48)) yields a probability of no change of 0.1084, while the maximum probability for an epidemic change is 0.1319 at the pair $\boldsymbol{k} = [1, 23]$, which are the ages 19 and 48. This shows weak evidence of an epidemic change.

Since there is little reason to believe that an epidemic change model is applicable in this situation, we also considered the exponential model with two change-points, but without the epidemic model restriction. The probability for no change is 0.0268 with a maximum probability of 0.0206 at $\boldsymbol{k} = [4, 23]$, showing again little evidence of change-points, since the probability of no change, although small, is larger than for any particular change-point.

In fact it looks as if the Stanford heart transplant data has only one change-point, at $k = 23$ (48 years) with probability of 0.3004, against a probability of 0.0192 for no change-point.

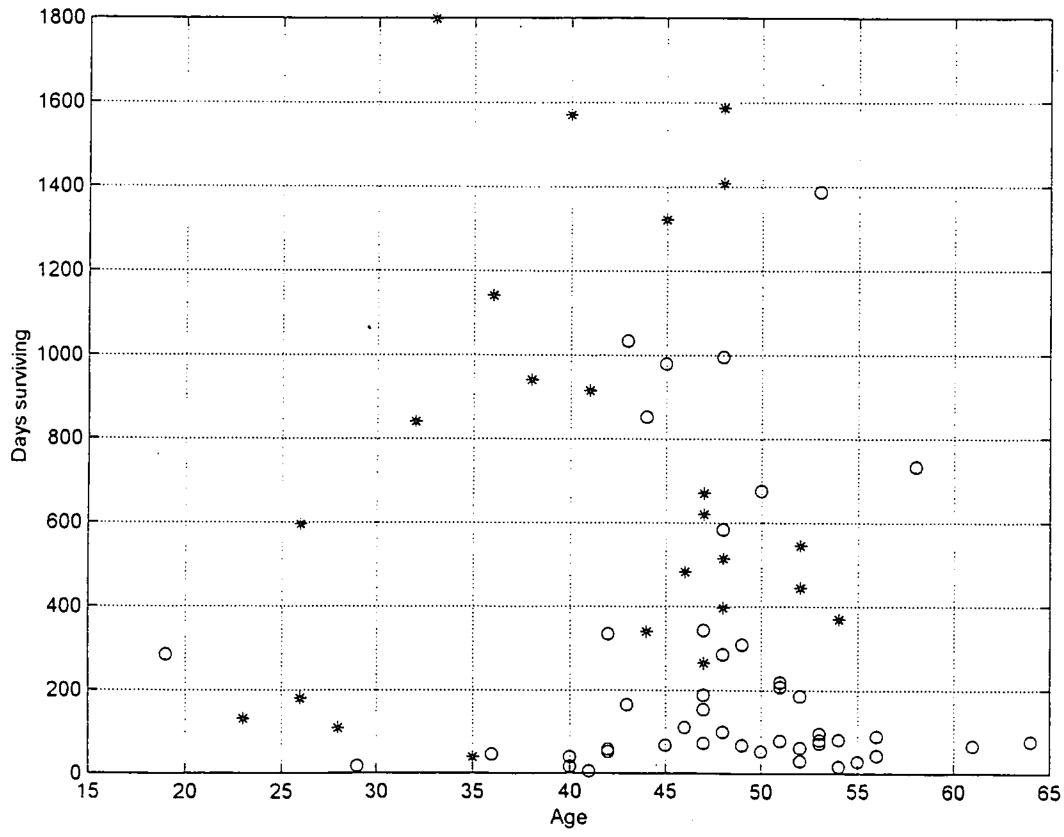


Figure 4.14: **Stanford heart transplant data, censored** $(*)$ **and uncensored** (0) **observations**

CHAPTER 5

THE HAZARD RATE

5.1 INTRODUCTION

If a random variable X represents the lifetime or time to failure of a unit, then the reliability of the unit at time t is defined to be

$$R(t) = P[X > t] = 1 - F_X(t). \tag{5.1}$$

The same function with the notation $S(t) = 1 - F_X(t)$, is called the survival function in biomedical applications.

Properties of a distribution that were previously studied, such as the mean and variance, are still important in the reliability area, but an additional property that is quite useful is the hazard function (HF) or failure-rate function. The hazard function, $h(t)$, for a *pdf* is defined to be

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}. \tag{5.2}$$

The HF may be interpreted as the instantaneous failure rate or the conditional density of failure at time t, given that the unit has survived until time t,

$$f(t|X \geq t) = h(t). \tag{5.3}$$

An increasing HF at time t indicates that the unit is more likely to fail in the next increment of time $(t, t + \Delta t)$ than it would be in an interval of the same length at an earlier age. That is, the unit is wearing out or deteriorating with age.

Similarly, a decreasing HF means that the unit is improving with age. A constant hazard function occurs for the exponential distribution and it reflects the no-memory property of that distribution.

If $X \sim \text{Exp}(\theta)$, then

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{\theta e^{-\theta t}}{e^{-\theta t}} = \theta. \tag{5.4}$$

In this case the failure rate is the reciprocal of the mean time to failure and it does not depend on the age of the unit. This assumption may be reasonable for certain types of electrical components, but it would tend not to be true for mechanical components. However, the no wear out assumption may be reasonable over some restricted time span. The exponential distribution has been an important model in the life-testing area, due in part to its simplicity. The Weibull distribution is a generalization of the exponential distribution and it is much more flexible.

If $X \sim \text{Wei}(\theta, \beta)$, then

$$\begin{aligned} h(t) &= \frac{\beta\theta^{-\beta}t^{\beta-1}e^{-\left(\frac{t}{\theta}\right)^{\beta}}}{e^{-\left(\frac{t}{\theta}\right)^{\beta}}} \\ &= \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}. \end{aligned} \tag{5.5}$$

This reduces to the exponential case for $\beta = 1$. For $\beta > 1$, the Weibull HF is an increasing function of t and for $\beta < 1$, the HF is a decreasing function of t.

One typical form of HF in the area of life-testing is a U-shaped or bathtub-shaped HF. For example, a unit may have a fairly high failure rate when it is first put into operation, due to the possible presence of manufacturing defects. If the unit survives the early period, then a nearly constant HF may apply for some period, where the causes of failure are occurring "at random". Later on the failure rate may begin to increase as wear out or old age becomes a factor. In life sciences the early failures are associated with the "infant mortality" effect.

Mitra and Basu (1995) considered the problem of estimating change-points in various non-monotonic aging models. A general methodology for consistent estimation of the change-point is developed and applied to non-monotonic aging models based on the hazard rate function as well as on the mean residual life function. All the other references considered a change in a constant hazard rate, while Mitra and Basu (1995) considered changing survival functions.

A more general model is dealt with by Basu, Ghosh and Joshi (1998) and Ghosh and Joshi (1992), assuming the failure rate $r(t)$ to have the shape of the "first" part of the "bathtub" model, i.e. $r(t)$ is decreasing for $t < \tau$ and is constant for $t \geq \tau$. The asymptotic distribution of one of the estimates proposed earlier has been investigated. This leads to a test for the hypothesis $H_0 : \tau \leq \tau_0$ vs $H_1 : \tau > \tau_0$ (where $\tau_0 > 0$). Asymptotic expression for the power of this test under Pitman alternatives is derived. Kulasekera and Saxena (1991) consider "bathtub" and "upside down bathtub" shaped hazard functions and based their estimator of the change-point on kernel estimators of the density function.

A frequently recurring question posed by leukemia researchers concerns a test of a constant failure rate against the alternative of a failure rate involving a single change-point. In answer to this question, Matthews and Farewell (1982) derived and simulated a likelihood ratio test appropriate for the stated alternative. Consideration is given also to tests based on alternatives in the log gamma family. Other classical approaches include Matthews, Farewell and Pyke (1985), Yao (1986) and Pham and Nguyen (1990, 1993).

Nguyen, Rogers and Walker (1984) discussed the estimation of parameters in hazard rate models with a change-point. Due to the irregularity of the models, the classical maximum likelihood method and the method of moments cannot be used. A consistent estimator of the change-point is obtained by examining the properties of the density represented as a mixture. The performance of the estimator is checked via simulation.

Achcar and Bolfarine (1989) developed a Bayesian approach to the problem of a constant hazard with a single change-point using non-informative reference priors. They also present a generalization for the comparison for two treatments.

Ghosh, *et al.* (1993) examine the asymptotics of a Bayesian approach for the same model, but under the assumption of a lower hazard rate after the change-point. Ghosh, Joshi and Mukhopadhyay (1996) assumed the hazard rate $h(t)$ of a lifetime random variables to be a constant equal to a up to time τ and another constant equal to b thereafter. The parameters τ and (a, b) are assumed to be independent apriori, with τ having a uniform prior on $[t_1, t_2]$, $0 < t_1 < t_2 < \infty$, while the prior of (a, b) is assumed to be smooth. They proved that the marginal posterior mode of τ is n-consistent; the marginal posterior mass of τ is concentrated around an n^{-1} neighbourhood of the unknown parameter value; the posterior distribution of (a, b) can be approximated by a Normal distribution; a, b, τ are asymptotically independent aposteriori and one can approximate the posterior mean and variance of (a, b) by easily computable quantities. The accuracies of these approximations were examined by a simulation study.

Ghosal, Ghosh and Samanta (1999) considered a family of models that arise in connection with a sharp change in the hazard rate corresponding to a high initial hazard rate dropping to a more stable or slowly changing rate at an unknown change-point k. Although the Bayes estimates are well behaved and are asymptotically efficient, it is difficult to compute them as the posterior distributions are generally very complicated. They obtained a simple first order asymptotic approximation to the posterior distribution of k. They judged the accuracy of the approximation through simulation. Zacks (1972) considers a constant failure rate (exponential) that changes at an unknown time point to an increasing (Weibull) rate.

5.2 THE EXPONENTIAL MODEL

A constant hazard function occurs for the exponential distribution and it reflects the no-memory property of that distribution mentioned earlier.

If $X \sim \text{Exp}(\theta)$, then

$$\begin{aligned} f(t|\theta) &= \theta e^{-\theta t}, \\ F(t|\theta) &= 1 - e^{-\theta t} \end{aligned} \tag{5.6}$$

so that the survival function is

$$S(t|\theta) = e^{-\theta t} \tag{5.7}$$

and the hazard function follows as

$$h(t|\theta) = \theta. \tag{5.8}$$

Under the models

M_0 : No change in hazard rate,

$$f(t|\theta_0) = \theta_0 e^{-\theta_0 t}, \quad t > 0,$$

M_τ: Change at τ (continuous)

$$f(t|\theta_1, \theta_2, \tau) = \begin{cases} \theta_1 e^{-\theta_1 t}; & 0 < t \leq \tau \\ \theta_2 e^{-\theta_1 \tau - \theta_2 (t-\tau)}; & \tau < t \end{cases} . \tag{5.9}$$

Notice that the data is not considered sequentially over time as before, but arranged in order of magnitude.

Achcar and Bolfarine (1989) consider a discrete prior on τ with the observed values as support, and a vague prior,

$$\Pi(\theta_1, \theta_2) \propto \frac{1}{\theta_1 \theta_2}, \quad \theta_1, \theta_2 > 0, \tag{5.10}$$

on the parameters. They also derive the non-informative Jeffreys prior in the case of uncensored and censored observations, assuming τ known, and show that the above prior is a reasonable approximation to the Jeffreys prior. Ghosh, *et al.* (1993, 1996) consider the assumptions one needs to make about the prior in order to obtain an analytical approximation to the posterior density of τ. The prior in (5.10) would satisfy their conditions if $0 < c < \theta_2 < \theta_1 < \infty$ and τ is uniform on $[t_1, t_{n-1}]$ where $0 < t_1 < \tau < t_{n-1} < \infty$.

We will start by assuming proper conjugate priors on the parameters

$$\theta_0, \theta_1, \theta_2 \sim iid\Gamma(\alpha, \beta) \tag{5.11}$$

and

$$\tau \sim U(t_1, t_{n-1}) \tag{5.12}$$

where $t_1 < t_2 < \cdots < t_n$.

For censored observations $t_1^* < t_2^* < \cdots < t_m^*$ it follows that:

If $t_i^* > \tau$

$$\begin{aligned} F(t_i^*|\theta_1, \theta_2, \tau) &= 1 - e^{-\theta_1 \tau} + \int_\tau^{t_i^*} \theta_2 e^{-\theta_1 \tau - \theta_2 (t - \tau)} dt \\ &= 1 - e^{-\theta_1 \tau} + e^{-\theta_1 \tau + \theta_2 \tau}(e^{-\theta_2 \tau} - e^{-\theta_2 t_i^*}) \\ &= 1 - e^{-\theta_1 \tau - \theta_2 (t_i^* - \tau)} \end{aligned}$$

and if $t_i^* \leq \tau$

$$F(t_i^*|\theta_1, \theta_2, \tau) = 1 - e^{-\theta_1 t_i^*}. \tag{5.13}$$

Under Model M_τ the likelihood function follows as

$$\begin{aligned} f(\boldsymbol{t}, \boldsymbol{t}^*|\theta_1, \theta_2, \tau) &= \prod_{i=1}^{k_1} f(t_i|\theta) \prod_{k_1+1}^{n} f(t_i|\theta_1, \theta_2, \tau) \prod_{i=1}^{k_1^*}(1 - F(t_i^*|\theta_1)) \cdot \prod_{k_1^*+1}^{m} (1 - F(t_i^*|\theta_1, \theta_2, \tau)) \\ &= \theta_1^{k_1} e^{-\theta_1 y_1} \theta_2^{k_2} e^{-k_2 \theta_1 \tau - \theta_2 (y_2 - k_2 \tau)} e^{-\theta_1 y_1^*} e^{-\theta_1 k_2^* \tau - \theta_2 (y_2^* - k_2^* \tau)} \\ &= \theta_1^{k_1} \theta_2^{k_2} e^{-\theta_1 [s_1 + r_2 \tau]} e^{-\theta_2 [s_2 - r_2 \tau]} \end{aligned} \tag{5.14}$$

where k_1 and k_2 are, respectively, the number of uncensored observations before and after the change point τ. Similarly we have k_1^* and k_2^* for the censored observations. Further,

$$t_1 < \cdots < t_{k_1} < \tau < t_{k_1+1} < \cdots < t_{k_1+k_2} = t_n,$$

$$s_1 = y_1 + y_1^* = \sum_{i=1}^{k_1} t_i + \sum_{i=1}^{k_1^*} t_i^*,$$

$$s_2 = y_2 + y_2^* = \sum_{i=k_1+1}^{n} t_i + \sum_{i=k_1^*+1}^{m} t_i^*$$

and

$$r_2 = k_2 + k_2^*. \tag{5.15}$$

The joint distribution is given by

$$f(\boldsymbol{t}, \boldsymbol{t}^*, \theta_1, \theta_2, \tau) = \frac{1}{t_{n-1} - t_1} \frac{\beta^{2\alpha}}{\Gamma^2(\alpha)} \theta_1^{\alpha+k_1-1} \theta_2^{\alpha+k_2-1} \; e^{-\beta\theta_1 - \beta\theta_2} \; e^{-\theta_1[s_1 + r_2\tau]} \; e^{-\theta_2[s_2 - r_2\tau]}$$

so that

$$f(\boldsymbol{t}, \boldsymbol{t}^*, \tau) \propto \frac{1}{t_{n-1} - t_1} \Gamma(\alpha + k_1)\Gamma(\alpha + k_2)[s_1 + r_2\tau + \beta]^{-(\alpha+k_1)}[s_2 - r_2\tau + \beta]^{-(\alpha+k_2)}. \tag{5.16}$$

With the priors (5.11), the posterior density of the change-point is given by

$$\Pi(\tau|\boldsymbol{t}, \boldsymbol{t}^*) \propto \Gamma(k_1 + \alpha)\Gamma(k_2 + \alpha)[s_1 + r_2\tau + \beta]^{-(k_1+\alpha)}[s_2 - r_2\tau + \beta]^{-(k_2+\alpha)}, \quad x_k \leq \tau < x_{k+1} \tag{5.17}$$

where $x_1 < x_2 < \cdots < x_{n+m}$ are all the ordered observations and k is running from the index of the first to the $(n-2)$th uncensored observation. So the posterior of τ has a discontinuity at each observed value of the variable. The posterior with vague prior (5.10) follows directly by letting $\alpha, \beta \to 0$ in (5.17). Also, the posteriors of θ_1 and θ_2, conditional on τ, follows as

$$\theta_i | \boldsymbol{t}, \boldsymbol{t}^*, \tau \sim \quad \text{Gamma}(k_i, z_i), \quad i = 1, 2 \tag{5.18}$$

where

$$z_1 = s_1 + r_2\tau \text{ and } z_2 = s_2 - r_2\tau. \tag{5.19}$$

For the distribution of $\delta = \frac{\theta_2}{\theta_1}$ it follows that

$$f(\boldsymbol{t}, \boldsymbol{t}^* | \theta_2, \delta, \tau) \propto \delta^{k_2} \theta_2^n e^{-\theta_1[s_1 + r_2\tau]} e^{-\delta\theta_1[s_2 - r_2\tau]} \tag{5.20}$$

and with the priors

$$\Pi(\theta, \delta) \propto \frac{1}{\delta\theta} \text{ and } \Pi(\tau) = \frac{1}{t_{n-1} - t_1}, \tag{5.21}$$

the posterior of δ follows as

$$\begin{aligned} \Pi(\delta | \boldsymbol{t}, \boldsymbol{t}^*, \tau) &\propto \delta^{k_2 - 1}[s_1 + \delta s_2 + \tau r_2(1 - \delta)]^{-n} \\ &= \frac{1}{B(k_1, k_2)} \left(\frac{z_2}{z_1}\right)^{k_2} \delta^{k_2 - 1} \left(1 + \frac{z_2}{z_1}\delta\right)^{-n}, \quad \delta > 0, \end{aligned} \tag{5.22}$$

a Beta Type II distribution, so that $\frac{k_1 z_2}{k_2 z_1}\delta$ has a F-distribution with $2k_1$ and $2k_2$ degrees of freedom.

The unconditional distributions of θ_1, θ_2 and δ follow by averaging over the marginal posterior of τ, i.e.

$$\Pi(\delta | \boldsymbol{t}, \boldsymbol{t}^*) = \sum_k \int_{x_k}^{x_{k+1}} \Pi(\delta | \boldsymbol{t}, \boldsymbol{t}^*, \tau) \Pi(\tau | \boldsymbol{t}, \boldsymbol{t}^*) d\tau. \tag{5.23}$$

In the above analysis the assumption of exactly one change-point in the hazard rate was made. To examine the evidence in favour of this assumption, we will consider the Bayes factor when comparing this model M_τ in (5.9) with the model M_0 which assumes a constant hazard with no change.

For the Fractional Bayes factor it follows that

$$m_\tau = \sum_{i=1}^{n-2} \int_{t_i}^{t_{i+1}} \Pi(\tau | \boldsymbol{t}, \boldsymbol{t}^*) d\tau, \tag{5.24}$$

$$m_\tau^b = \sum_{i=1}^{n-2} \int_{t_i}^{t_{i+1}} \frac{\Gamma(bk_1)\Gamma(bk_2)}{b^{nb} z_1^{bk_1} z_2^{bk_2}} d\tau, \tag{5.25}$$

$$m_0 = \frac{\Gamma(n)}{s^n} \tag{5.26}$$

and

$$m_0^b = \frac{\Gamma(bn)}{b^{nb} s^{nb}}. \tag{5.27}$$

If we let $b = \frac{2}{n}$,

$$m_0(b) = \frac{\Gamma(n)}{s^{n-2}}. \tag{5.28}$$

Up to this stage we assumed that there are no restrictions on the hazard rates θ_1 and θ_2. If we know more accurately that there is a decrease in the hazard rate after the change-point, the prior will change to

$$\Pi(\theta_1, \theta_2) \propto \frac{1}{\theta_1 \theta_2}, \quad 0 < \theta_2 < \theta_1 < \infty. \tag{5.29}$$

Then the joint distribution will be

$$f(\boldsymbol{t}, \boldsymbol{t}^*, \tau) \propto \int_0^\infty \int_0^{\theta_1} \theta_1^{k_1 - 1} \theta_2^{k_2 - 1} e^{-\theta_1 z_1 - \theta_2 z_2} d\theta_2 d\theta_1 \tag{5.30}$$

so that the posterior distribution follows as

$$\Pi(\tau|\boldsymbol{t},\boldsymbol{t}^*) \propto z_2^{-k_2} \int_0^\infty \Gamma_{\text{inc}}(\theta_1 z_2; k_2)\theta_1^{k_1-1} e^{-\theta_1 z_1} d\theta_1 \tag{5.31}$$

where $\Gamma_{\text{inc}}(\cdot, k_2)$ is the incomplete Gamma function and which can be integrated numerically.

5.3 HAZARD RATES FROM COMBINATIONS OF DENSITY FUNCTIONS

Unfortunately none of the common standard distributions will accommodate a U-shaped hazard rate. To obtain the wanted shape, different combinations of the Weibull hazard rate can be applied. If $X \sim \text{Wei}(\theta, \beta)$ then $h(x) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1}$. This reduces to the exponential case for $\beta = 1$. For $\beta > 1$, it is an increasing function of x and if $\beta < 1$ it is a decreasing function of x.

5.3.1 DECREASING HAZARD RATE

Firstly we will consider a lifetime distribution with a strictly decreasing hazard rate up to a change point τ after which it remains constant. Such a model is important in situations where equipment have high infant mortality rate and manufacturers are interested in the point at which the surviving equipment can be considered more reliable.

In general, if $f_1(\cdot)$ and $f_2(\cdot)$ are non-increasing continuous density functions with survival functions $S_1(\cdot)$ and $S_2(\cdot)$ respectively, and

$$f(x|\tau,\theta_1,\theta_2) = \begin{cases} f_1(x|\theta_1); & x \leq \tau \\ \frac{S_1(\tau|\theta_1)}{S_2(\tau|\theta_2)} f_2(x|\theta_2); & x > \tau, \end{cases} \tag{5.32}$$

then $h(x_1) \geq h(x_2)$ for $x_1 < x_2$ if $\theta_2 = g(\theta_1)$ such that

$$f_1(\tau|\theta_1)S_2(\tau|\theta_2) = f_2(\tau|\theta_2)S_1(\tau|\theta_1).$$

This last condition defines the assumption that there is no jump in the density or hazard functions at the point of discontinuity.

We will assume a Weibull density with decreasing hazard rate for $f_1(\cdot)$ and an exponential density for $f_2(\cdot)$.

Then

$$f(t|\tau,\lambda,\beta) = \begin{cases} \lambda\beta t^{\beta-1}e^{-\lambda t^{\beta}}; & t \leq \tau \\ \lambda\beta\tau^{\beta-1}e^{-\lambda\tau^{\beta-1}[(1-\beta)\tau+\beta t]}; & t > \tau \end{cases} \tag{5.33}$$

with hazard rate

$$h(x) = \begin{cases} \lambda\beta t^{\beta-1}; & t \leq \tau \\ \lambda\beta\tau^{\beta-1}; & t > \tau \end{cases} \tag{5.34}$$

where $\lambda > 0$ and $0 < \beta < 1$.

The likelihood function is given by

$$L(\lambda,\beta,\tau|\boldsymbol{t},\boldsymbol{t}^*) = \lambda^n\beta^n\tau^{-k_1(1-\beta)}\left(\prod_{i=1}^{k} t_i\right)^{\beta-1} e^{-\lambda[s_{1\beta}+r_2(1-\beta)\tau^{\beta}+\beta\tau^{\beta-1}s_2]} \tag{5.35}$$

with notation as in (5.15) and

$$s_{1\beta} = y_{1\beta} + y_{1\beta}^* = \sum_{i=1}^{k_1} t_i^{\beta} + \sum_{i=1}^{k_1^*} t_i^{*\beta}. \tag{5.36}$$

With the priors

$$\lambda_1 \sim \Gamma(\alpha,\gamma), \quad \beta \sim U(0,1) \text{ and } \tau \sim U(t_1,t_{n-1}) \tag{5.37}$$

the joint distribution is

$$f(\boldsymbol{t}, \boldsymbol{t}^*, \lambda_1, \beta, \tau) = \lambda_1^{n+\alpha-1} e^{-\lambda_1[s_{1\beta}+r_2(1-\beta)\tau^\beta+\beta\tau^{\beta-1}s_2+\gamma]} \beta^n \left(\prod^{k_1} t_i\right)^{-(1-\beta)} \cdot$$

$$\tau^{-k_2(1-\beta)} \cdot \frac{\gamma^\alpha}{\Gamma(\alpha)} \frac{1}{t_{n-1}-t_1} \tag{5.38}$$

so that

$$f(\boldsymbol{t}, \boldsymbol{t}^*, \beta, \tau) = \Gamma(n+\alpha)[s_{1\beta} + r_2(1-\beta)\tau^\beta + \beta\tau^{\beta-1}s_2 + \gamma]^{(n+\alpha)} \beta^n \left(\prod^{k_1} t_i\right)^{-(1-\beta)} \tau^{-k_2(1-\beta)}. \tag{5.39}$$

If $\alpha, \beta \to 0$, the posterior distribution of τ follows as

$$\Pi(\tau|\boldsymbol{t}, \boldsymbol{t}^*) \propto \int_0^1 \left(\tau^{k_2} \prod^{k_1} t_i\right)^{-(1-\beta)} \left[\frac{1}{\beta}s_{1\beta} + r_2\left(\frac{1-\beta}{\beta}\right)\tau^\beta + \tau^{\beta-1}s_2\right]^{-n} d\beta. \tag{5.40}$$

The marginal likelihood under M_τ will be

$$m_\tau = \frac{\Gamma(n+\alpha)\gamma^\alpha}{(t_{n-1}-t_1)\Gamma(\alpha)} \int_{t_1}^{t_{n-1}} \Pi(\tau|\boldsymbol{t}, \boldsymbol{t}^*) d\tau. \tag{5.41}$$

Under M_0 the joint distribution follows as

$$f(\boldsymbol{t}, \lambda_1, \beta) = \lambda_1^n \beta^n \left(\prod^n t_i\right)^{\beta-1} e^{-\lambda_1 y_\beta} \frac{\gamma^\alpha}{\Gamma(\alpha)} \lambda_1^{\alpha-1} e^{-\gamma\lambda_1} \tag{5.42}$$

where

$$y_\beta = \sum_i^n t_i^\beta$$

so that

$$f(t,\beta) = \frac{\Gamma(n+\alpha)\gamma^{\alpha}}{\Gamma(\alpha)}[y_{\beta}+\gamma]^{-(n+\alpha)}\beta^{n}(\Pi t_i)^{\beta-1}. \tag{5.43}$$

The marginals under the two hypotheses are then, with vague priors $\alpha, \beta \to 0$,

$$m_0 = \int_0 \beta^n \left(\prod^n t_i\right)^{\beta-1} y_{\beta}^{-n} d\beta \tag{5.44}$$

and

$$m_{\tau} = \frac{1}{(t_{n-1}-t_1)}\int_{t_1}^{t_{n-1}}\int_0^1 \left(\tau^{k_2}\prod^{k_i} t_i\right)^{-(1-\beta)} \left[\frac{y_{1\beta}}{\beta} + k\left(\frac{1-\beta}{\beta}\right)\tau^{\beta} + \tau^{\beta-1}y_2\right]^{-n} d\beta d\tau \tag{5.45}$$

so that

$$B_{0\tau} = \frac{m_0}{m_{\tau}}.$$

In this case we consider the ordinary $B_{0\tau}$ as reasonable, since we have proper priors on β and τ. The parameter λ which has an improper prior appears under both models. Thus it's not required to calculate a Fractional Bayes factor in this case.

5.3.2 INCREASING HAZARD RATE

In the situation where an initially constant hazard rate later changes to an increasing rate due to wear-out or ageing, we have a combination of an exponential distribution and a Weibull distribution with increasing hazard rate.

Let's consider the model M_{τ} with hazard rate

$$h(t) = \begin{cases} \lambda; & t \leq \tau \\ \lambda\left(\frac{t}{\tau}\right)^{\beta-1}; & t > \tau \end{cases}. \tag{5.46}$$

Then

$$f(t|\lambda,\beta,\tau) = \begin{cases} \lambda e^{-\lambda t}; & t \leq \tau \\ \lambda e^{-\lambda\tau\left(\frac{\beta-1}{\beta}\right)\left(\frac{t}{\tau}\right)^{\beta-1}} e^{-\frac{\lambda t}{\beta}\left(\frac{t}{\tau}\right)^{\beta}}; & t > \tau \end{cases}. \tag{5.47}$$

It then follows that

$$f(\boldsymbol{t},\boldsymbol{t}^*|\lambda,\beta,\tau) = \lambda^n \tau^{-k_2(\beta-1)} \left(\prod_{k_1+1}^{n} t_i\right)^{\beta-1} e^{-\lambda s_1} e^{-\frac{\lambda\tau}{\beta}\left[r_2(\beta-1)+\frac{s_{2\beta}}{\tau^\beta}\right]} \tag{5.48}$$

where

$$s_1 = y_1 + y_1^* = \sum_{i=1}^{k_1} t_i + \sum_{i=1}^{k_1^*} t_i^*; \quad s_{2\beta} = y_{2\beta} + y_{2\beta}^* = \sum_{k_1+1}^{n} t_i^\beta + \sum_{k_1^*+1}^{m} t_i^{*\beta}$$

and $r_2 = k_2 + k_2^*$ (as in 5.15).

Let the priors be

$$\tau \sim U(t_1, t_{n-1}), \quad \pi(\lambda) \propto \frac{1}{\lambda} \text{ and } \Pi(\beta) \propto \frac{1}{\beta}, \quad 1 < \beta < K. \tag{5.49}$$

We put an upper limit on β, since the likelihood function is unbounded if $\beta \to \infty$.

The posterior of τ is

$$\Pi(\tau|\boldsymbol{t},\boldsymbol{t}^*) \propto \int_1^K \frac{1}{\beta}\tau^{-k_2(\beta-1)} \left(\prod_{k_1+1}^{n} t_i\right)^{\beta-1} \left[s_1 + \tau r_2\left(\frac{\beta-1}{\beta}\right) + \frac{\tau^{1-\beta}}{\beta} s_{2\beta}\right]^{-n} d\beta \tag{5.50}$$

and for the Fractional Bayes factor the fractional marginal likelihood is

$$m_\tau^b = \sum_{i=1}^{n-2} \int_{t_i}^{t_i+1} \int_1^K \frac{1}{\beta} \tau^{-bk_2(\beta-1)} \left(\prod_{i=k_1+1}^{n} t_i \right)^{b(\beta-1)} \left[b \left\{ s_1 + \tau r_2 \left(\frac{\beta-1}{\beta} \right) + \frac{\tau^{1-\beta}}{\beta} s_{2\beta} \right\} \right]^{-nb} \Gamma(nb) d\beta d\tau. \tag{5.51}$$

This model can then be compared with any of the previous models like (5.24), (5.26) and (5.45).

5.3.3 THE BATHTUB HAZARD RATE

A lifetime distribution having support $[0, \infty)$ is said to be a Bathtub Hazard Rate (BHR) distribution if there exists a $t_0 > 0$ such that $h(t)$ is non-increasing on $[0, t_0)$ and non-decreasing on $[t_0, \infty)$.

The point t_0 need not be unique and we will consider a model with decreasing failure rate on $[0, \tau_1]$, a constant failure rate on $(\tau_1, \tau_2]$ and an increasing failure rate on (τ_2, ∞). Assuming no jumps at the points of discontinuity, the density function can be written as

$$f(t|\lambda, \beta_1, \beta_2, \tau_1, \tau_2) = \begin{cases} \lambda\beta_1 t^{\beta_1-1} e^{-\lambda t^{\beta_1}}; & 0 \le t \le \tau_1 \\ \lambda\beta_1 \tau_1^{\beta_1-1} e^{-\lambda\tau_1^{\beta_1}(1-\beta_1)} e^{-\lambda\beta_1\tau_1^{\beta_1-1}t}; & \tau_1 < t \le \tau_2 \\ At^{\beta_2-1} e^{-Bt^{\beta_2}}; & t_2 < t < \infty \end{cases} \tag{5.52}$$

with $\lambda > 0, \quad 0 < \beta_1 \le 1, \quad 1 \le \beta_2 \le K$ and

$$A = \beta_2 B e^{-\lambda\beta_1 \frac{\beta_2-1}{\beta_2} \tau_1^{\beta_1-1} \tau_2 - \lambda\tau_1^{\beta_1}(1-\beta_1)},$$

$$B = \lambda \frac{\beta_1}{\beta_2} \frac{\tau_1^{\beta_1-1}}{\tau_2^{\beta_2-1}}. \tag{5.53}$$

The corresponding hazard rate is

$$h(t) = \begin{cases} \lambda\beta_1 t^{\beta_1 - 1}; & 0 \le t \le \tau_1 \\ \lambda\beta_1 \tau_2^{\beta_1 - 1}; & \tau_1 < t \le \tau_2 \\ \beta_2 B t^{\beta_2 - 1}; & \tau_2 < t < \infty. \end{cases} \tag{5.54}$$

The priors are

$$\Pi(\lambda) = \frac{1}{\lambda}, \quad \Pi(\beta_1) = 1, \quad 0 < \beta_1 \le 1,$$

$$\Pi(\beta_2) \propto \frac{1}{\beta_2}, \quad 1 \le \beta_2 \le K,$$

$$\Pi(\tau_2|\tau_1) = U(\tau_1, b), \quad \Pi(\tau_1) = U(a, b). \tag{5.55}$$

In practice, under the assumption that both change-points fall within the range of observed data, we will take $a = t_1$ and b as either t_{n-1} or t_{n-2}, ensuring at least one observation in each section of the density function.

Then

$$f(\boldsymbol{t}, \boldsymbol{t}^* | \lambda, \beta_1, \beta_2, \tau_1, \tau_2) = \left(\prod_{i=1}^{k_1} t_i \right)^{\beta_1 - 1} \left(\prod_{i=k_1+k_2+1}^{n} t_i \right)^{\beta_2 - 1} e^{-\lambda C}$$

where

$$C = (r_2 + r_3)\tau_1^{\beta_1}(1 - \beta_1) + \beta_1 \tau_1^{\beta_1 - 1} \left(r_3 \frac{\beta_2 - 1}{\beta_2} \tau_2 + s_2 + \frac{1}{\beta_2} \tau_2^{1-\beta_2} s_{3\beta_2} \right) + s_{1\beta_1},$$

$$r_i = k_i + k_i^*, \quad i = 1, 2, 3 \tag{5.56}$$

and

$$s_{1\beta_1} = \sum_{i=1}^{r_1} x_i^{\beta_1}, \quad s_2 = \sum_{i=r_1+1}^{r_1+r_2} x_i, \quad s_{3\beta_2} = \sum_{i=r_1+r_2+1}^{n+m} x_i^{\beta_2}, \tag{5.57}$$

where $x_1 < x_2 < \cdots < x_{n+m}$ are again all the ordered observations.

Then

$$f(\boldsymbol{t}, \beta_1, \beta_2, \tau_1, \tau_2) \propto \beta_1^n \tau_1^{(r_2+r_3)(\beta_1-1)} \tau_2^{-r_3(\beta_2-1)} \left(\prod_{i=1}^{k_1} t_i \right)^{\beta_1 - 1} \left(\prod_{i=n_2+1}^{n} t_i \right)^{\beta_2 - 1} C^{-n} \frac{1}{\beta_2}. \quad (5.58)$$

The conditional distribution of λ is Gamma (n, C). For model comparisons, the fractional marginal likelihood as in (1.13) can be obtained by a MCMC sampling scheme. As mentioned by Gilks (1995), notice from the denominator of (1.13) that

$$m^b = E[L^{1-b}(\theta|\boldsymbol{x})] \quad (5.59)$$

with expectation with respect to the proper density $\Pi^*(\theta) \propto L^b(\theta|\boldsymbol{x})\Pi(\theta)$. So we perform MCMC iterations on the density and estimate m^b as the average of $L^{1-b}(\theta|\boldsymbol{x})$ over the generated samples.

To obtain the marginal posterior distributions, Gibbs sampling can be employed, simulating from the conditional distributions successively and averaging. After integrating λ out, the full conditional distributions of the other four parameters are given by

$$\Pi(\beta_1|\boldsymbol{t}, \boldsymbol{t}^*, \beta_2, \tau_1, \tau_2) \propto \beta_1^n \tau_1^{\beta_1(k_2+k_3)} T_1^{\beta_1 - 1} C^{-n}, \quad 0 < \beta_1 \leq 1, \quad (5.60)$$

$$\Pi(\beta_2|\boldsymbol{t}, \boldsymbol{t}^*, \beta_1, \tau_1, \tau_2) \propto \frac{1}{\beta_2} \tau_2^{-k_3(\beta_2-1)} T_3^{\beta_2 - 1} C^{-n}, \quad 1 \leq \beta_2 \leq K, \quad (5.61)$$

$$\Pi(\tau_1|\boldsymbol{t}, \boldsymbol{t}^*, \beta_1, \beta_2, \tau_2) \propto \tau_1^{-(k_2+k_3)(\beta_1-1)} T_1^{\beta_1 - 1} C^{-n}, \quad t_1 \leq \tau_1 < t_{[\tau_2]^-}, \quad (5.62)$$

$$\Pi(\tau_2|\boldsymbol{t}, \boldsymbol{t}^*, \beta_1, \beta_2, \tau_1) \propto \tau_2^{-k_3(\beta_2-1)} T_3^{\beta_2 - 1} C^{-n}, \quad t_{[\tau_2]^+} < \tau \leq t_{n-1}, \quad (5.63)$$

where

$$T_1 = \prod_{i=1}^{k_1} t_i, \quad T_3 = \prod_{i=k_1+k_2+1}^{n} t_i \tag{5.64}$$

and where k_1, k_2 and k_3 are the number of observations smaller or equal to τ_1, between τ_1 and τ_2 and larger than τ_2 respectively. Similarly we have k_1^*, k_2^* and k_3^* for censored observations.

5.4 APPLICATIONS

EXAMPLE 5.4.1

We will consider a data set given in Matthews and Farewell (1982) as well as in Achcar and Bolfarine (1989). The data specifies the times (in days) from remission induction to relapse for 84 patients with acute non-lymphoblastic leukemia. There are 51 uncensored and 33 censored observations.

Under the Exponential model the Fractional Bayes factor in favour of model M_0 (no change) when compared with model M_τ is $B_{0_\tau} = 0.0628$ when $b = \frac{2}{n}$ (equations (5.24) to (5.28)). This translates to a posterior probability for the change-point model of $P = 0.9409$ with equal prior weights. This indicates reasonably strong evidence in favour of the change-point model. In Figure 5.1 the Fractional Bayes factor is shown as a function of b. As b approaches one, so will B_{0_τ}, as there is then little or no information left for model comparison. However, for b reasonably small ($b < 0.5$) the Bayes factor is fairly robust ($B_{0_\tau} < 0.1$) with a minimum of 0.0367 at $b = \frac{8}{n}$.

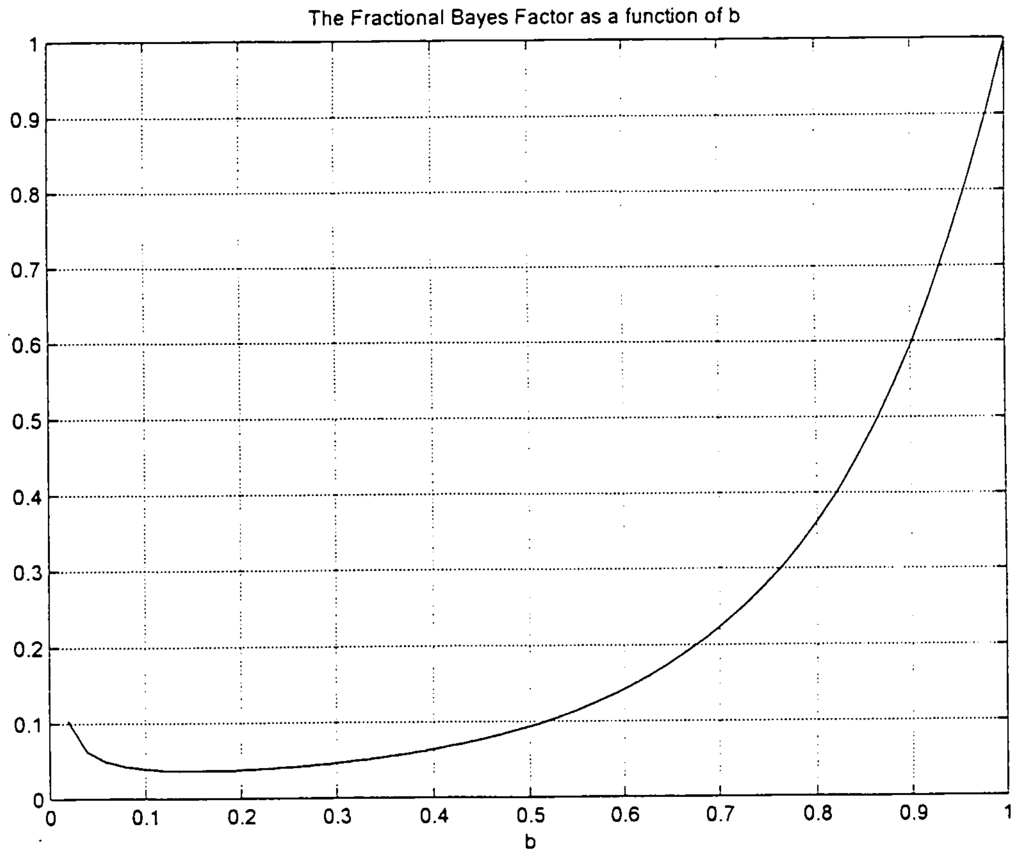


Figure 5.1: The Fractional Bayes Factor as a function of b for example 5.4.1

According to Matthews and Farewell (1982) the classical significance level when testing the null hypotheses of no change against the change-point alternative is approximately 0.001.

We also compared the change-point model M_τ with a Weibull model with decreasing hazard rate. The Fractional Bayes factor in favour of M_τ is 53.64, indicating a much better fit to the data than a Weibull model.

Assuming a change-point, Figure 5.2 shows the posterior distribution of τ. The discontinuities occur at each of the uncensored observations.

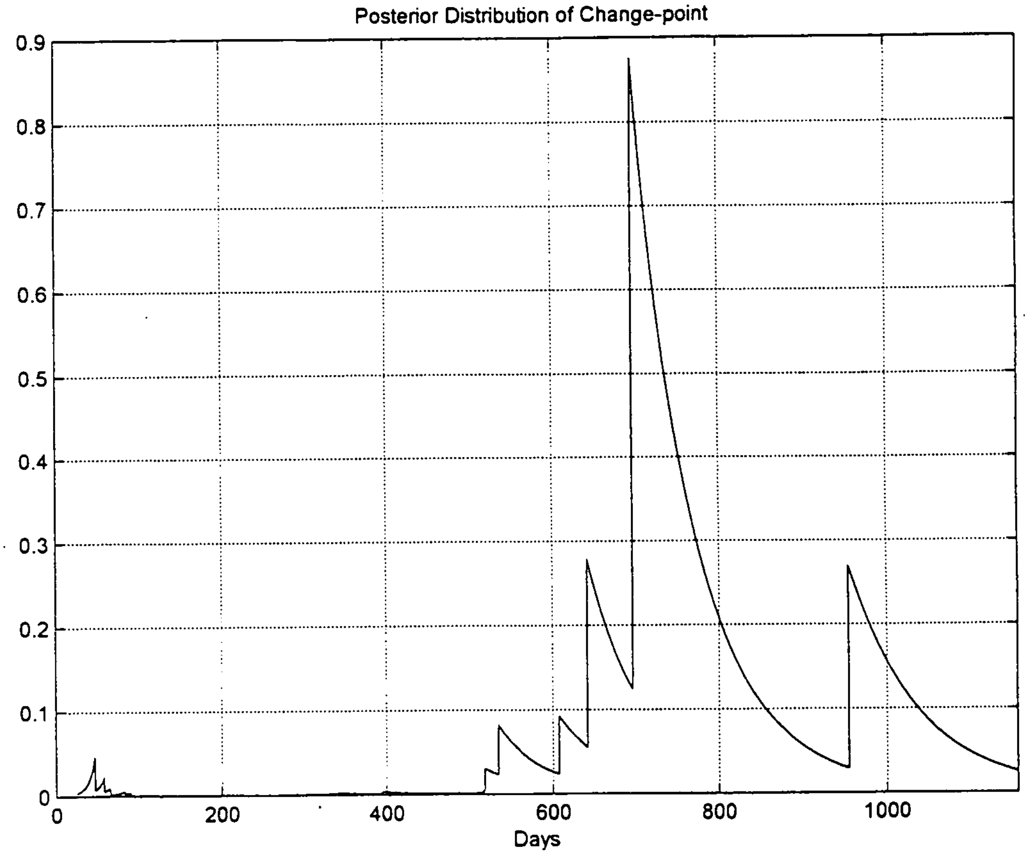


Figure 5.2: The posterior distribution of the change-point τ for example 5.4.1

The mode is at 697 days. This is also the maximum likelihood estimator of τ (Matthews and Farewell (1982)).

The unconditional posterior means of λ_1 and λ_2 are given in Table 5.1 together with the asymptotic approximations according to Ghosh, *et al.* (1996) and the maximum likelihood estimators from Matthews and Farewell (1982).

Table 5.1: Estimates of λ_1 and λ_2 for example 5.4.1

	$\lambda_1 \times 10^3$	$\lambda_2 \times 10^3$
Post. Means	2.005	0.3364
Approximation (GJM)	2.081	0.2877
MLE (MF)	2.040	0.4300

In Figure 5.3 the conditional, given $\tau = 697$, and unconditional posteriors of $\delta = \dfrac{\lambda_2}{\lambda_1}$ is shown according to (5.22) and (5.23).

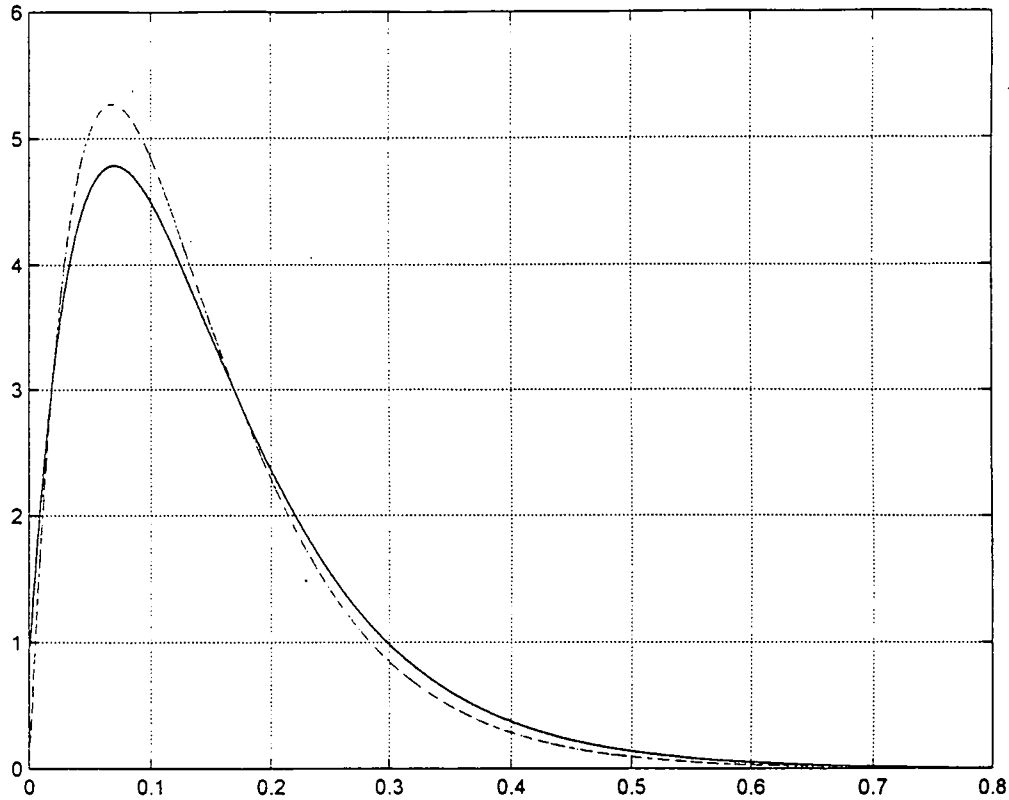


Figure 5.3: Conditional (given $\tau = 697$) posterior $(---)$ and unconditional posterior $(-)$ of $\delta = \lambda_2/\lambda_1$ for example 5.4.1

The unconditional posterior of δ has a mean of 0.1512 and 95% HPD interval (0.007; 0.393). The maximum likelihood estimator is $\hat{\delta} = 0.2108$.

Assuming both τ and λ_1 known, and equal to the MLE's, Achcar & Bolfarine (1989) report the posterior mode of δ as 0.1405, and with only τ known, as 0.1374. They also give in this case a 95% HPD interval of (0.1253; 0.3540) by using a normal approximation. We consider this interval as too narrow. Their unconditional posterior mode is 0.25.

EXAMPLE 5.4.2

In Appendix A the failure times (hours) of 107 units of a piece of electronic equipment are given. The data is from Juran and Gryna (1980) and is also analysed by Schneider, *et al.* (1990). It is suggested by Schneider, *et al.* that there is an early failure period of approximately 20 hours and a wear-out period of about 100 hours, the period between being fairly constant.

We compared the non-decreasing hazard rate model (5.46) with the bathtub model (5.54) by means of the MCMC calculation (5.59) of the Fractional Bayes factor. After 10 000 simulations from each model and with $b = \frac{3}{n}$, the estimated Bayes factor was calculated as $B = 3.79$ in favour of model (5.54), moderate evidence in favour of the bathtub hazard rate model. Compared to a constant hazard rate model and the single abrupt change model (5.9), the Bayes factors in favour of the BHR model are, respectively, 1.2×10^9 and 1.3×10^3.

Assuming the BHR model, 15 000 Gibbs samples were kept after burn-in, and the marginal posterior densities of the five parameters obtained. Some posterior measures are given in Table 5.2 with the two estimated change-points at 46.9 and 95.5 hours. The position of the second change-point, τ_2, can be estimated much more accurately than that of τ_1 as can be seen from the 95% HPD Interval. Also some contours of the joint density of τ_1 and τ_2 are shown in Figure 5.4.

Table 5.2: Posterior measures of five parameters for example 5.4.2

	Mean	Mode	Variance	95% Interval
λ	0.2071	0.1990	0.0015	0.137 – 0.289
β_1	0.7763	0.7780	0.0081	0.604 – 0.958
β_2	11.8259	10.6300	23.8981	4.450 – 20.400
τ_1	46.922	26.100	483.37	10.80 – 89.40
τ_2	95.546	97.100	82.99	78.70 – 112.80

Figure 5.5 shows the expectation of the unconditional distribution of the hazard rate (5.54) with the 95% credibility interval, and in Figure 5.6 the empirical survival function $S(t) = 1 - F(t)$ is plotted together with the posterior expectation and 95% credibility interval.

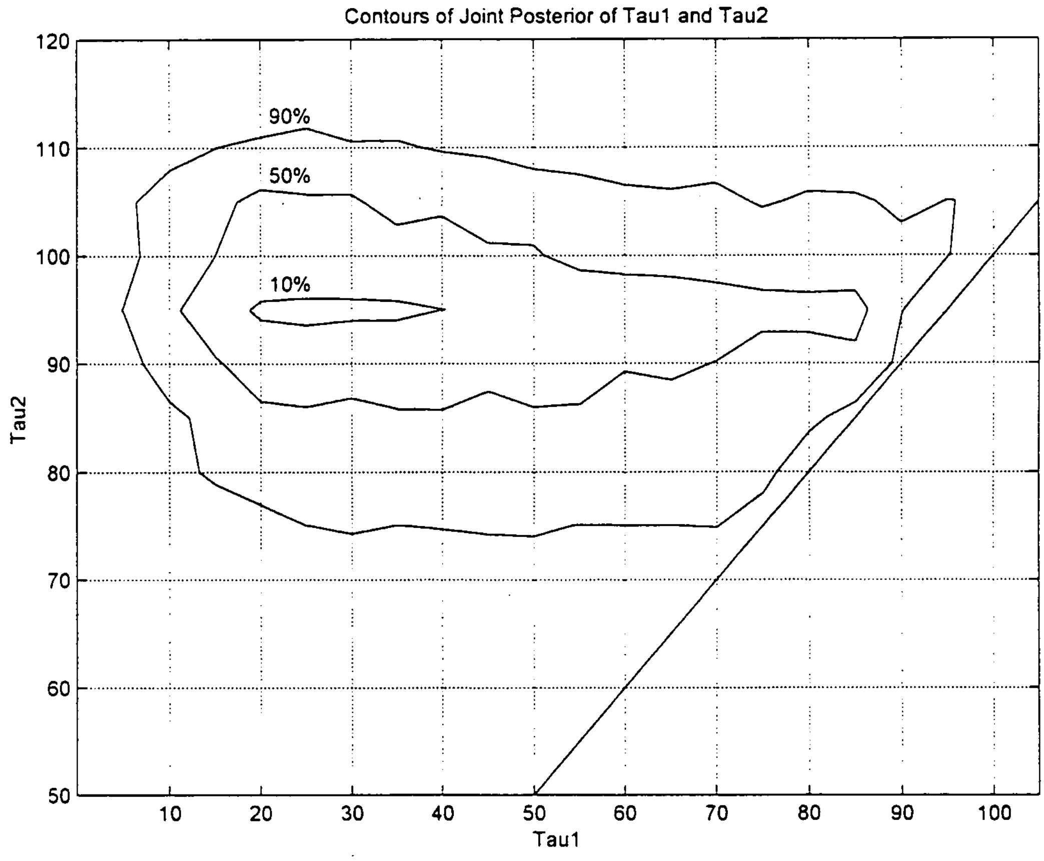


Figure 5.4: Contours of the joint posterior density of the change-points τ_1 and τ_2 for example 5.4.2

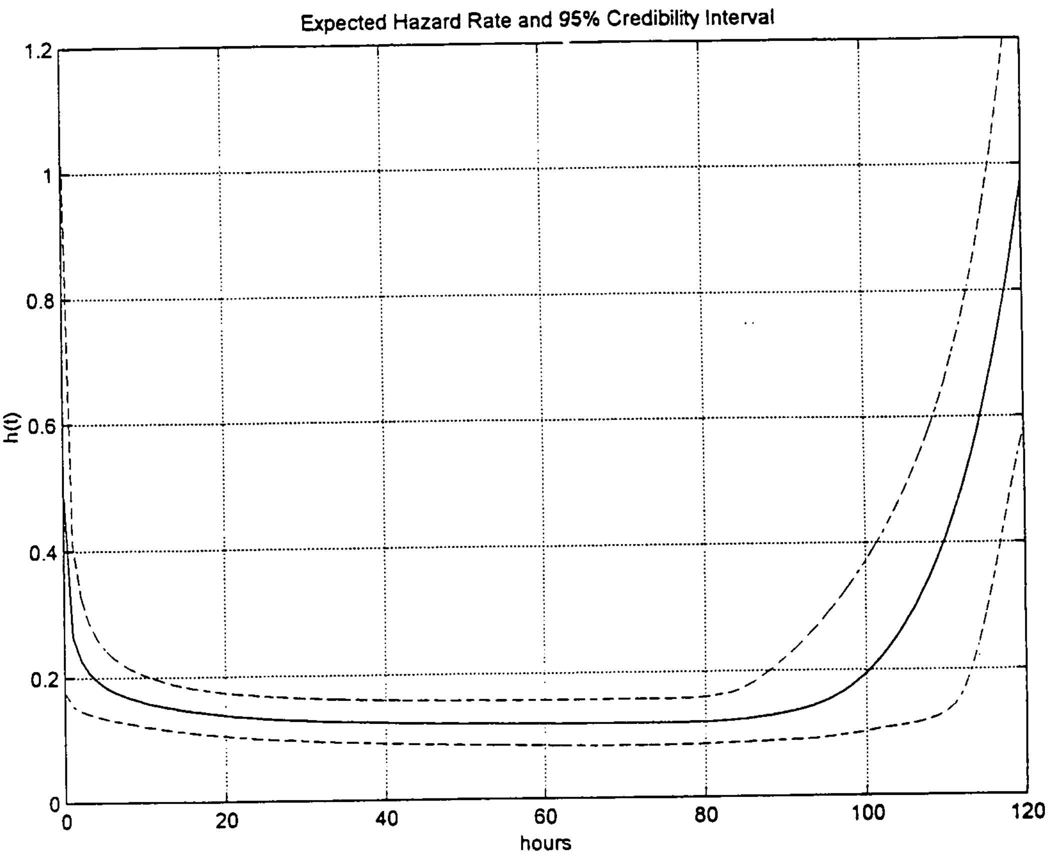


Figure 5.5: **Estimated Hazard Rate with 95% Credibility Interval for example 5.4.2**

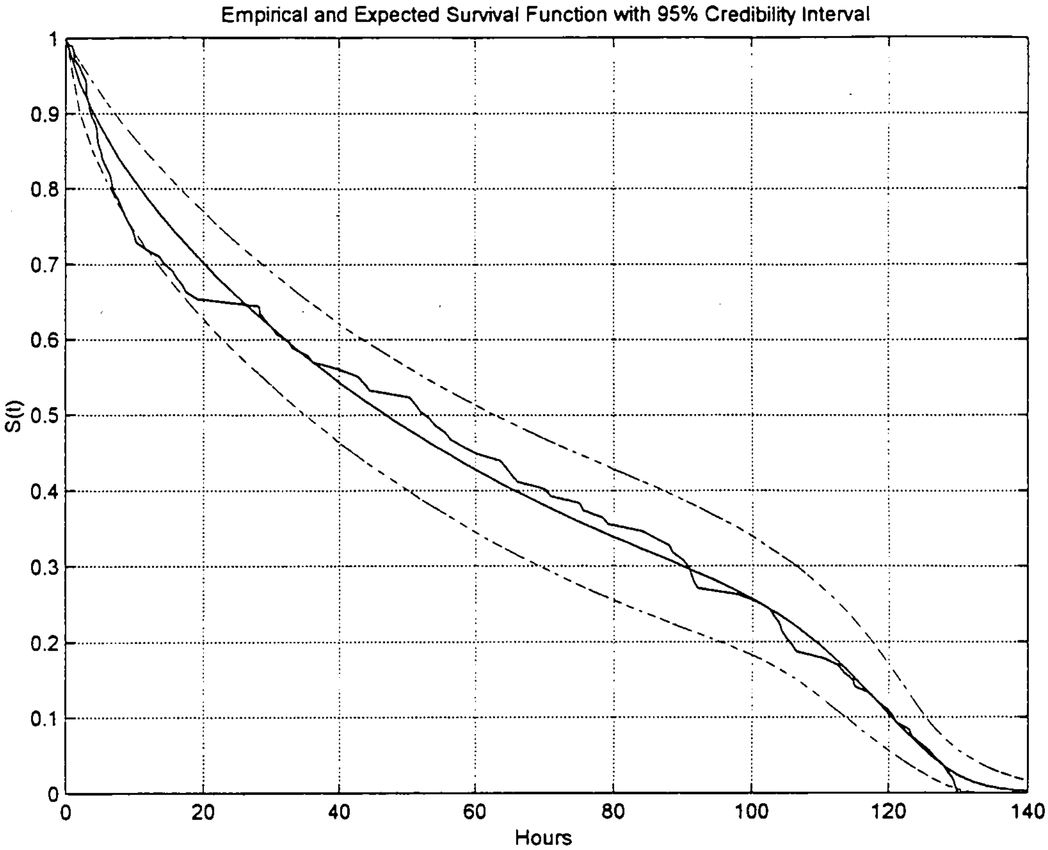


Figure 5.6: Empirical and Estimated Survival Function with 95% Credibility Interval for example 5.4.2

CHAPTER 6

CONCLUSIONS AND SUMMARY OF OTHER APPROACHES

6.1 CONCLUSIONS

In this study we have attempted to use a pure Bayesian approach for the detection and estimation of change-points in a sequence of random variables from a variety of models. The emphasis was on obtaining "default" results, or at least results with as little as possible subjective prior input. It is well known that the Bayes factor is very sensitive to the prior, especially when comparing models of different dimensions. When the number of change-points is assumed known, the use of improper priors is not a problem. However, if there is uncertainty about the number of change-points, we must be able to specify reasonable proper priors under each model or revert to partial Bayes factors. We have looked at two of them, the Intrinsic and Fractional Bayes factors. The advantages and disadvantages of these two Bayes factors are discussed in some detail by O'Hagan (1997) and Berger and Pericchi (1998). We will just mention a few which are particularly relevant in change-point analysis.

Firstly, the Intrinsic Bayes factors generally violate the multiple comparison coherence condition, so B_{12} is defined as $B_{12} = \frac{1}{B_{21}}$ where M_2 is the more complex model. In general an encompassing model is usually defined to avoid this problem, and a minimal sample is then defined with respect to all possible submodels. However, in change-point analysis a minimal sample for all possible positions of the change-point will be the whole sample, so an encompassing model will not be possible. In our analysis we have considered only the minimal sample for each fixed change-point when compared to the no change model and defined $B_{ij} = \frac{B_{i0}}{B_{j0}}$. The shortcoming is that B_{i0} and B_{j0} are not calculated from the same set of minimal samples.

Secondly, the Intrinsic Bayes factor are computationally very intensive, especially for multiple change-points. For example, if $n > 100$ and the number of change-points is larger than one, the number of combinations of change-points and minimal samples gets unmanageably large and use a lot of computer time. This problem can be reduced by taking a random sample of minimal samples instead of all possible minimal samples.

In contrast, the Fractional Bayes factor (FBF) is relatively simple to calculate, even for large samples and multiple change-points. Also the FBF is invariant to transformations of the data, while IBF's are not always. The FBF does not suffer from the incoherency problem. So in general we prefer to use the FBF in change-point analysis. In the examples given throughout this study, both methods gave similar and sensible answers.

The one major practical problem with the FBF is the choice of the training fraction b. We have seen from the examples that the posterior probabilities can be highly sensitive to the value of b, but it seems possible to obtain a lower limit on the Bayes factor in favour of no change (see for example Figure 4.5). O'Hagan (1995, 1997) suggested $b = \frac{m}{n}$, where m is the minimal training sample size. For increased robustness when n increases, O'Hagan also have two other suggestions, namely $b = \sqrt{\frac{m}{n}}$ and $b = \frac{m \log n}{n \log m}$. However, in some models the minimal training sample is not unique, and our suggestion is to examine the Bayes factor B_{ok} or the posterior probability for no change over a range of values for b to check the robustness of your answer.

6.2 OTHER APPROACHES IN THE LITERATURE

Here we want to summarize some approaches, Bayesian, quasi-Bayesian and non-Bayesian in the literature that were not dealt with in this study in any detail.

MCMC methods: In chapters two and three we have looked at the Gibbs sampling scheme for deriving the posterior distribution of a change-point for certain models. However, the analytical result was usually available under the assumption of exactly one change-point for certain models. When the number of change-points is unknown however, the ordinary Gibbs sampling algorithm is not applicable. Papers that deal with Markov chain Monte Carlo methods when the chain has to move between models of different dimensions include a Gibbs sampler-based approach by Rotondi and Pagliano (1994), the reversible jump algorithm of Green (1995) and the jump diffusion sample algorithm of Phillips and Smith (1996). These methods seem promising when the number of possible change-points are reasonably large.

Product partition models: This is also an approach to the detection of multiple change-points. Product partition models were studied by Hartigan (1990) and Barry and Hartigan (1992, 1993). According to Hartigan (1990), product partition models assume that observations in different components of a random partition of the data are independent given the partition. If the probability distribution of random partitions is in a certain product form prior to making the observations, it is also in product form given the observations. The product model thus provides convenient machinery for allowing the data to weight the partitions likely to hold and inference about particular future observations may then be made by first conditioning on the partition and then averaging over all partitions. Barry and Hartigan (1992) show, with appropriate selection of prior product models, that the observations can eventually determine approximately the true partition. Barry and Hartigan (1993) show that the parameter values may be estimated exactly in $O(n^3)$ calculations, or to an adequate approximation by Markov sampling techniques that are $O(n)$ in the number of observations. The Markov sampling computations are thus practicable for long sequences.

Information criteria: The Schwarz information criterion has been applied by Gupta and Chen (1996), Ramanayake and Gupta (1998) and Chen and Gupta (1997, 1999). Gupta and Chen (1996) applied the Schwarz information criterion together with the binary segmentation procedure to detect change-points in a set of geological data and the changes in the frequencies of pronouns in the plays of Shakespeare. Ramanayake and Gupta (1998) considered a sequence of independent exponential random variables that is susceptible to a change in the means. They tested whether the means have been subjected to an epidemic change after an unknown point, for an unknown duration in the sequence and derived the likelihood ratio statistic and a likelihood ratio type statistic. Chen and Gupta (1997) explored testing and locating multiple variance change-points in a sequence of independent Gaussian random variables (assuming known and common mean). A binary procedure combined with the Schwarz information criterion (SIC) is used to search all of the possible variance change-points existing in the sequence. Chen and Gupta (1999) studied the testing and estimation of a single change-point in means and variances of a sequence of independent Gaussian normal random variables. The SIC is defined as $\mathrm{SIC}(n) = -2\log L(\widehat{\theta}) + p\log n$ and Chen and Gupta's criteria is to reject M_0, the model of no change, if $\mathrm{SIC}(n) > \mathrm{SIC}(k)$ for some k. The position of the change-point is estimated by $\hat{k}$ such that $\mathrm{SIC}(\hat{k}) = \min_k \mathrm{SIC}(k)$. Akaike's and Schwarz's criteria are also used by Yao (1988), Caussinus and Lyazrhi (1997) and Xiong and Milliken (2000).

Decision theory: Decision theory was considered by Lyazrhi (1994). He studied the change-point problem for normal regression models as the problem of choosing the hypothesis H_o of no change or one of the hypotheses H_i that one or more parameters change after the ith observation. The observations are often associated with a known increasing sequence τ_i (for example τ_i is the date of the ith observation). It then seems natural to introduce a quadratic loss function involving $(\tau_i - \tau_j)^2$ for selecting H_i instead of the true hypothesis H_j. A Bayes optimal invariant procedure is derived within such a framework and compared to previous proposals. Rukhin (1996) also considered the change-point problem as a multiple decision problem and show that a positive limit of the minimum Bayes risk for the uniform prior exists for any loss. Its explicit form and some inequalities are derived for the zero-one loss function. A multiple decision (non-Bayesian) treatment was used by Haccou and Meelis (1988).

Sequential analysis: An approach closely related to a decision theoretic one is sequential analysis. This includes optimal stopping rules and quality control. All the models described in our study is for fixed sample size, but can easily be applied sequentially. However, models that are developed specifically for sequential detection were considered by Zacks and Barzily (1981), Kenett and Pollak (1983), Zacks (1983, 1991), Zhang (1995) and Zacks (1995). Sequential methods are especially important in problems of statistical control of processes with stochastic input, early warning of changes in the distributions and tracking of processes. Zacks and Barzily (1981) discussed the determination of a stopping rule for the detection of the time of an increase in the success probability of a sequence of independent Bernoulli trials. Both success probabilities are assumed known. A Bayesian approach is applied and the distribution of the location of the shift in the success probability is assumed geometric and the success probabilities are assumed to have a known joint prior distribution. The costs involved are penalties for late or early stoppings. The nature of the optimal dynamic programming solution is discussed and a procedure for obtaining a suboptimal stopping rule is determined.

Kenett and Pollak (1983) supposed that one is monitoring a sequence of observations for a possible increase in the probability of a rare event and that it is not possible to immediately stop the process under observation or influence it to return to its normal state. According to them one would then desire a scheme which takes advantage of observations occuring after a detection of a change is proclaimed. They developed a modification of Page's CUSUM Procedure, taking account of these additional observations. Zacks (1991) provided a development of a Bayesian tracking algorithm, which estimates parameters of the posterior distributions of the means under the AMOC model. They also cited other tracking procedures, which are based on adaptive Kalman filtering. Zhang (1995) discussed and applied Bayesian and likelihood approaches to on-line detecting change-points in time series to analyze biomedical data. Using a linear dynamic model, the Bayesian analysis outputs the conditional posterior probability of a change at time $t-1$, given the data up to time t and the status of changes occured before time $t-1$. The likelihood method is based on a change-point regression model and tests whether there is no change-point. Zacks (1995) developed sequential stopping rules for testing reliability systems having a random number of change-points in their hazard rate functions. The failure process follows the empirical Bayes model of Littlewood (1981). Sarkar and Meeker (1997) presented a Bayesian on-line change detection algorithm for the cases when there are multiple jumps and when there is a trend in the process output parameter. A decision theory based method has been formulated to determine the optimum inspection interval for process control applications. Karunamuni and Zhang (1996) considered an empirical Bayes stopping rule for detecting a change in distribution when the prior is not completely known.

Non-Bayesian papers on optimal stopping include Pollak (1985), Lai (1995) and Yakir (1997). Papers that deal specifically with quality control and are based on the "Minimum Description Length" (MDL) criterion are Seki and Hashimoto (1996), Suzuki and Enkawa (1995) and Suzuki (2000).

Non-parametric approaches to the change-point problem include Pettitt (1979), Wolfe and Schechtman (1984), Lombard (1987), Eastwood (1993) and Aly and BuHamra (1996).

6.3 OTHER APPLICATIONS

Time series models are not considered in any detail in this study, except for the regression model with autocorrelated errors in paragraph 2.7.5 and 3.7. There we considered only a change in the regression coefficients. Many variations and extensions of time series models are of course possible. For example, consider a AR(1) process,

$$y_i = \beta y_{i-1} + \epsilon_i; \quad i = 1, \ldots, n, \tag{6.3.1}$$

with a possible change in β. We can also have

$$y_i = \alpha + \beta y_{i-1} + \epsilon_i; \quad i = 1, \ldots, n, \tag{6.3.2}$$

with a possible change in the mean α, or in β or in both. We can also consider a moving average, MA(1) process,

$$y_i = \beta + \epsilon_i - \phi\epsilon_{i-1}; \quad i = 1, \ldots, n, \tag{6.3.3}$$

with possible changes in β and/or ϕ.

Broemeling and Tsurumi (1987) consider some aspects of the three models given above. In general,very little has appeared about changes in time series with regard to changes in the covariance or correlation structure of a time series; however, there has been some studies about a change in the mean of a process. For example, Box and Tiao (1975) have introduced changes in the mean of a very general time series

$$y_i = f_i(\theta, \xi) + \epsilon_i; \quad i = 1, \ldots, n \tag{6.3.4}$$

where f is a function of unknown parameters θ and ξ, where the latter are called intervention effects which change the mean of the process. The errors ϵ_i are assumed to follow and ARMA process.

Tsurumi (1976) and Salazar (1980) used a transition function to represent both abrupt and smooth changes in linear models. Tsurumi considered a simultaneous equations model, while Salazar studied smooth changes in a regression model with autocorrelated errors. Ilmakunnas and Tsurumi (1984, 1985) used a Bayesian approach in the case where the change-point is known, but there is a possible change in the autocorrelation. Ohtani (1982) considered two possibilities: a change in the regression parameters and a change in the autocorrelation coefficient of the error distribution; thus there were two shift points to consider simultaneously. Salazar (1982) also studied changes in α and β for model (6.3.2), while Cook (1983) examined the multivariate case.

Kim, Cho and Lee (2000) considered the problem of testing parameter constancy in GARCH (1,1) models from a frequentist viewpoint. The GARCH (1,1) process is such that

$$\epsilon_i = z_i \sigma_i$$

where $\sigma_i^2 = \omega + \alpha\epsilon_{i-1}^2 + \beta\sigma_{i-1}^2; \quad i = 1, \ldots, n$ (6.3.5)

and $E(z_i) = 0$ and $E(z_i^2) = 1$. A cusum of squares test is proposed in analogy of Inclàn and Tiao's (1994) statistic. Its limiting distribution is derived via using the invariance principle for martingale sequences.

Spatial statistics: Spatial statistics was considered by Stephens and Smith (1992) and Raftery (1995). Stephens and Smith (1992) and Mascarenhas and Prado (1980) formulated the problem of edge-detection as a statistical change-point problem using a Bayesian approach. Practical applications of image analysis abound in agronomy (remote sensing), astronomy (study of galaxies), industrial processing (automated manufacturing and quality control, medicine (internal body imaging) and the military (intelligence, reconnaissance, defence/ offence systems), relating variously to imaging technologies such as photography, tomography, radiography, etc. They have shown that the Gibbs sampler provides an effective procedure for the required Bayesian calculations and illustrated the use of the method for "quick and dirty" image segmentation. According to Raftery (1995) in estimating and testing for change-points in one-dimensional stochastic processes, model-based approaches have often succeeded. Once a model is specified, the problem can be solved with fairly simple Bayesian methods. When it is assumed that there is only one change-point, Bayesian analyses can take especially simple forms. A fairly general solution is given and the possibility of exact finite-sample inference is illustrated with the change-point Poisson process. Then a general approximate approach based on the Laplace method for integrals applied to Bayes factors is described. When the number of change-points is not known in advance, the Bayesian approach often proceeds most naturally using state-space models. The state-space model approach to change-points was introduced by Harrison and Stevens (1976) under the name multi-process Kalman Filter. This name describes a general computational strategy. However, the multi-process Kalman Filter does not always work well and a different Bayesian state-space modeling approach is reviewed. A different approach uses time series models that can generate change-points but are not specified in terms of them. One class of such models, the Markov Transition Distribution (MTD) models is described and its capacity for representing change-points without the need to specify in advance that they may be present is illustrated with a simple simulation. In two dimensions the problem is much harder, because instead of a single change-point we have a whole change curve. The fully Bayesian approach is then much more difficult. Raftery (1995) reviewed a semi-parametric approach in which change curves are modeled by principal curves, a family of non-parametric smooth curves (Hastie and Stuetzle (1989)). The set of potential edge elements (which may be pixels in an image or events in a spatial point process) is then modeled as a mixture of distributions each of which is centered around a different principal curve. These are estimated using generalizations of traditional cluster analysis methods.

Analysis-of-variance models have received little attention in the literature with respect to the detection of change-points and there seems to be scope for further research in this direction. Hirotsu (1997) assume that the two-way data y_{ij} are independently distributed according either to normal $N(\mu_{ij}, 1)$ or Poisson $P(\mu_{ij})$, so that the density function is given by

$$f(y_{ij}) = \exp[\{y_{ij}\theta_{ij} - \psi(\theta_{ij})\} + c(y_{ij})]$$

where

$$\theta_{ij} = \begin{cases} \theta + \alpha_i + \beta_j + \gamma; & (i \leq I, j \leq J) \\ \theta + \alpha_i + \beta_j; & \text{(otherwise)}, \end{cases} \tag{6.3.6}$$

with (I, J) an unknown two-way change-point. For the Poisson model we have $a(\mu_{ij}) = \log(\mu_{ij})$ and for the Normal model $a(\mu_{ij}) = \mu_{ij}$.

Hirotsu derived exact null and alternative distributions of the two-way maximally selected χ^2 for interaction between the ordered rows and columns for each of the normal and Poisson models respectively. The method is one of the multiple comparison procedures for ordered parameters and is useful for defining a block interaction or a two-way change-point model as a simple alternative to the two-way additive model. He described the construction of a confidence region for the two-way change-point. An important application is found in a dose-response clinical trial with ordered categorical responses, where detecting the dose level which gives significantly higher responses than the lower doses can be formulated as a problem of detecting a change in the interaction effects.

Multi-path change-point problems were considered by Joseph and Wolfson (1992), Kiuchi, Hartigan, Holford, Rubinstein and Stevens (1995), Joseph, Vandal and Wolfson (1996), Joseph, Wolfson, du Berger and Lyle (1997), BuHamra (1997) and Bélisle, Joseph, MacGibbon, Wolfson and du Berger (1998). Multi-path data refer to repeated measurements, say N independent sequences of random variables, each sequence possibly containing a change-point. According to Joseph and Wolfson (1992), this extension allowed the effective use of bootstrap and empirical Bayes methods, both of which is not feasible in the single-path context. Two classes of these multi-path change-point problems are considered by them. If the change-point is assumed to occur at the same position in each sequence, then the terminology "fixed-tau multi-path change-point" is used. In other cases, one may expect the change-point to occur at random positions in each sequence, according to some distribution, a "random-tau multi-path change-point" problem. Kiuchi, *et al.* (1995) proposed empirical Bayes and hierarchical Bayes change-point models to estimate the distribution of the time before AIDS diagnosis when a rapid decline in the $T4$ cell count begins. Results using the EM Algorithm and Markov chain Monte Carlo indicate that the mean change-point occurs approximately 1 year before diagnosis with a standard deviation of 9 months. Detection of a change-point may indicate that an AIDS diagnosis is increasingly likely for an individual HIV-positive but AIDS-free.

According to Joseph, *et al.* (1996) several measurements on the same variable in many experiments may be taken over time, a geographic region or some other index set. It is often of interest to know if there has been a change over the index set in the parameters of the distribution of the variable. Frequently, the data consists of a sequence of correlated random variables and there may also be several experimental units under observation, each providing a sequence of data. A problem in ascertaining the boundaries between the layers in geological sedimentary beds is used to introduce the model and then to illustrate the proposed methodology. It is assumed that, conditional on the change-point, the data from each sequence arise from an autoregressive process that undergoes a change in one or more of its parameters. Unconditionally, the model then becomes a mixture of non-stationary autoregressive processes. Maximum likelihood methods are used and results of simulations to evaluate the performance of these estimators under practical conditions are given.

Joseph, *et al.* (1997) present a Bayesian multi-path change-point model, which facilitates the comparison of baseline measurements to post-intervention values within each individual, eliminating the need to explicitly model the effects of baseline means. The main aim of their paper is to show how the ensemble of sample paths may be used to make Bayesian inference about the distribution of the times or locations of change. The position of the population change-point is modelled by using a Dirichlet prior for the probabilities of each possible discrete change-point. Two applications are shown, one Poisson model with 285 sequences, each of length 8 and a Normal model with 75 sequences, each of length 10. Bèlisle, *et al.* (1998) used the same Bayesian model as Joseph, *el al.* (1997) to analyze neuron spike train data. The data consists of counts of electrical discharges after a stimulus was applied to a neuron. A Poisson model was used and there were 35 data sequences.

BuHamra (1997) proposed four non-parametric test statistics for the change-point problem with repeated measures data. In a Monte Carlo simulation study, critical values for the proposed test statistics are simulated and the performances of the proposed tests are compared with the performance of some competitive tests in terms of asymptotic behavior and power.

Asymptotic theory was considered by Jandhyala and Minogue (1993), Ghosh, Joshi and Mukhopadhyay (1996), Lee (1998) and Ghosal, Ghosh and Samanta (1999). Jandhyala and Minogue (1993) derived simple elegant expressions for the covariance kernels of residual partial sum limit processes under the assumption of a polynomial regression model. A numerical method of solving Fredholm integral equations is derived, which is shown to provide solutions that are uniformly close to the analytical solutions. This numerical procedure is applied to compute quantiles for the asymptotic distributions of Bayes-type statistics derived to test for change in an arbitrary parameter of a general polynomial regression model. Ghosh, *et al.* (1996) examined the asymptotics of a Bayesian approach to the problem of a constant hazard with a single change-point, under the assumption of a lower hazard rate after the change-point. Lee (1998) found that the posterior mode of the number of change-points converges to the true number of change-points in the frequentist sense under mild assumptions and with respect to a suitable prior distribution. Ghosal, *et al.* (1999) considered a family of models that arise in connection with sharp change in hazard rate corresponding to high initial hazard rate dropping to a more stable or slowly changing rate at an unknown change-point. Although the Bayes estimates are well behaved and are asymptotically efficient, it is difficult to compute them as the posterior distributions are generally very complicated. They obtained a simple first order asymptotic approximation to the posterior distribution of the change-point.

APPENDIX A

Example 2.9.1: Male Egyptian skulls

c4000 BC				c3300 BC				c1850 BC			
MB	BH	BL	NH	MB	BH	BL	NH	MB	BH	BL	NH
131	138	89	49	124	138	101	48	137	141	96	52
125	131	92	48	133	134	97	48	129	133	93	47
131	132	99	50	138	134	98	45	132	138	87	48
119	132	96	44	148	129	104	51	130	134	106	50
136	143	100	54	126	124	95	45	134	134	96	45
138	137	89	56	135	136	98	52	140	133	98	50
139	130	108	48	132	145	100	54	138	138	95	47
125	136	93	48	133	130	102	48	136	145	99	55
131	134	102	51	131	134	96	50	136	131	92	46
134	134	99	51	133	125	94	46	126	136	95	56
129	138	95	50	133	136	103	53	137	129	100	53
134	121	95	53	131	139	98	51	137	139	97	50
126	129	109	51	131	136	99	56	136	126	101	50
132	136	100	50	138	134	98	49	137	133	90	49
141	140	100	51	130	136	104	53	129	142	104	47
131	134	97	54	131	128	98	45	135	138	102	55
135	137	103	50	138	129	107	53	129	135	92	50
132	133	93	53	123	131	101	51	134	125	90	60
139	136	96	50	130	129	105	47	138	134	96	51
132	131	101	49	134	130	93	54	136	135	94	53
126	133	102	51	137	136	106	49	132	130	91	52
135	135	103	47	126	131	100	48	133	131	100	50
134	124	93	53	135	136	97	52	138	137	94	51
128	134	103	50	129	126	91	50	130	127	99	45
130	130	104	49	134	139	101	49	136	133	91	49
138	135	100	55	131	134	90	53	134	123	95	52
128	132	93	53	132	130	104	50	136	137	101	54
127	129	106	48	130	132	93	52	133	131	96	49
131	136	114	54	135	132	98	54	138	133	100	55
124	138	101	46	130	128	101	51	138	133	91	46

c200 BC				cAD150			
MB	BH	BL	NH	MB	BH	BL	NH
137	134	107	54	137	123	91	50
141	128	95	53	136	131	95	49
141	130	87	49	128	126	91	57
135	131	99	51	130	134	92	52
133	120	91	46	138	127	86	47
131	135	90	50	126	138	101	52
140	137	94	60	136	138	97	58
139	130	90	48	126	126	92	45
140	134	90	51	132	132	99	55
138	140	100	52	139	135	92	54
132	133	90	53	143	120	95	51
134	134	97	54	141	136	101	54
135	135	99	50	135	135	95	56
133	136	95	52	137	134	93	53
136	130	99	55	142	135	96	52
134	137	93	52	139	134	95	47
131	141	99	55	138	125	99	51
129	135	95	47	137	135	96	54
136	128	93	54	133	125	92	50
131	125	88	48	145	129	89	47
139	130	94	53	138	136	92	46
144	124	86	50	131	129	97	44
141	131	97	53	143	126	88	54
130	131	98	53	134	124	91	55
133	128	92	51	132	127	97	52
138	126	97	54	137	125	85	57
131	142	95	53	129	128	81	52
136	138	94	55	140	135	103	48
132	136	92	52	147	129	87	48
135	130	100	51	136	133	97	51

Example 2.9.2: Colorado mountainside data

Z_1	Z_8	Z_9	Z_{10}	Z_{12}
320	060	020	250	370
280	060	040	210	420
260	060	010	250	440
305	050	050	260	250
290	050	020	210	510
275	050	020	230	570
280	080	020	270	400
300	120	010	280	300
250	070	030	250	330
285	070	010	240	280
280	060	020	370	300
300	120	060	250	200
280	150	010	280	280
305	130	010	300	260
230	270	030	250	240
325	160	010	280	170
270	160	010	290	330
250	120	001	260	330
260	270	080	480	330
270	180	040	450	220
325	600	080	660	250
315	410	200	600	260
335	360	080	590	170
310	640	240	630	190
410	760	440	800	001
360	770	260	770	010
310	660	380	640	010
420	620	520	680	001

Z_1	Z_8	Z_9	Z_{10}	Z_{12}
415	370	220	340	001
420	630	510	580	001
450	690	570	630	001
395	580	530	560	010
380	350	320	400	270
430	340	340	360	200
410	170	170	170	060
520	210	190	190	180
385	140	200	260	020
535	110	230	270	070
550	050	230	270	030
510	190	150	230	110
510	140	100	150	040
385	050	050	300	050
505	001	200	130	030
470	160	300	380	060
465	260	440	500	060
400	330	400	390	040
415	220	190	270	010
435	370	360	500	010
370	130	080	330	030
380	070	001	050	030
430	130	070	300	020
420	050	100	350	050
425	100	010	340	010
250	001	001	050	001
520	770	570	800	570

Example 2.9.3: Friday closing prices

Obs.	Exxon	General Dynamics	Obs.	Exxon	General Dynamics
1	48 - 06	45 - 05	27	47 - 07	32 - 01
2	47 - 06	41 - 05	28	48 - 06	31 - 05
3	48 - 05	41 - 04	29	48 - 05	31 - 06
4	46 - 06	39 - 07	30	49 - 02	27 - 03
5	47 - 07	39 - 06	31	53 - 01	29
6	48 - 02	37	32	51 - 05	27 - 04
7	48	36 - 06	33	52 - 01	26 - 01
8	47 - 01	36 - 02	34	48 - 07	26 - 02
9	46 - 06	38	35	50	24 - 07
10	46 - 03	37 - 06	36	50 - 07	25 - 05
11	47 - 04	37 - 07	37	51 - 03	25 - 03
12	46 - 01	37 - 05	38	51 - 02	26 - 04
13	46 - 01	37 - 04	39	49	23 - 05
14	46 - 01	37 - 02	40	49 - 07	25 - 04
15	45 - 06	37 - 03	41	48 - 04	22 - 07
16	46 - 03	37 - 02	42	49 - 07	20 - 06
17	45	35 - 04	43	47 - 02	22 - 03
18	46 - 04	34 - 04	44	49 - 05	23 - 04
19	47 - 07	33 - 04	45	50 - 03	23 - 07
20	47 - 06	34 - 05	46	50 - 03	22 - 05
21	46 - 04	34 - 03	47	51 - 01	23
22	47 - 06	35	48	50 - 05	23 - 03
23	47	36	49	49 - 02	24 - 05
24	47 - 07	35 - 05	50	50 - 06	25 - 03
25	47 - 07	33 - 04	51	50 - 06	26
26	47 - 07	32	52	51 - 05	25

Example 2.9.4: New York Stock Exchange Data

890.19	901.80	888.51	887.78	858.43	850.61
856.02	880.91	908.15	912.75	911.00	908.22
889.31	893.98	893.91	874.85	852.37	839.00
840.39	812.94	810.67	816.55	859.59	856.75
873.80	881.17	890.20	910.37	906.68	907.44
906.38	906.68	917.59	917.52	922.79	942.43
939.87	942.88	942.28	940.70	962.60	967.72
963.80	954.17	941.23	941.83	961.54	971.25
961.39	934.45	945.06	944.69	929.03	938.06
922.26	920.45	926.70	951.76	964.18	965.83
959.36	970.05	961.24	947.23	943.03	953.27
945.36	930.46	942.81	946.42	984.12	995.26
1005.57	1025.21	1023.43	1033.19	1027.24	1004.21
1020.02	1047.49	1039.36	1026.19	1003.54	980.81
979.46	979.23	959.89	961.32	972.23	963.05
922.71	951.01	931.07	959.36	963.20	922.19
953.87	927.89	895.17	930.84	893.96	920.00
888.55	879.82	891.71	870.11	885.99	910.90
936.71	908.87	852.38	871.84	863.49	887.57
898.63	886.36	927.90	947.10	971.25	978.63
963.73	987.06	935.28	908.42	891.33	854.00
822.25	838.05	815.65	818.73	848.02	880.23
841.48	855.47	859.39	843.94	820.40	820.32
855.99	851.92	878.05	887.83	878.13	846.68
847.54	844.81	859.90	834.64	845.90	850.44
818.84	816.65	802.17	853.72	843.09	815.39
802.41	791.77	787.23	787.94	784.57	752.58

Example 3.8.1: Quandt's data set

Obs. No. (i)	1	2	3	4	5	6	7	8	9	10
x_i	4	13	5	2	6	8	1	12	17	20
y_i	3.473	11.555	5.714	5.710	6.046	7.650	3.140	10.312	13.353	17.197

Obs. No. (i)	11	12	13	14	15	16	17	18	19	20
x_i	15	11	3	14	16	10	7	19	18	9
y_i	13.036	8.264	7.612	11.802	12.551	10.296	10.014	15.472	15.65	9.871

Example 3.8.2: Brownlee's stack lost data

Stack loss

	X_1	X_2	X_3	Y
1	80	27	58.9	4.2
2	80	27	58.8	3.7
3	75	25	59.0	3.7
4	62	24	58.7	2.8
5	62	22	58.7	1.8
6	62	23	58.7	1.8
7	62	24	59.3	1.9
8	62	24	59.3	2.0
9	58	23	58.7	1.5
10	58	18	58.0	1.4
11	58	18	58.9	1.4
12	58	17	58.8	1.3
13	58	18	58.2	1.1
14	58	19	59.3	1.2
15	50	18	58.9	0.8
16	50	18	58.6	0.7
17	50	19	57.2	0.8
18	50	19	57.9	0.8
19	50	20	58.0	0.9
20	56	20	58.2	1.5
21	70	20	59.1	1.5

Example 3.8.3: Olympic jumping events

	Year	High jump	Pole vault	Long jump	Triple jump
1	1896	1.81	3.30	6.34	13.72
2	1900	1.90	3.30	7.19	14.43
3	1904	1.80	3.51	7.34	14.33
4	1908	1.90	3.71	7.48	14.92
5	1912	1.93	3.95	7.60	14.76
6	1920	1.94	4.09	7.15	14.50
7	1924	1.98	3.95	7.45	15.53
8	1928	1.94	4.20	7.74	15.21
9	1932	1.97	4.31	7.64	15.72
10	1936	2.03	4.35	8.06	16.00
11	1948	1.98	4.30	7.82	15.40
12	1952	2.04	4.55	7.57	16.22
13	1956	2.11	4.56	7.83	16.34
14	1960	2.16	4.70	8.12	16.81
15	1964	2.18	5.10	8.07	16.85
16	1968	2.24	5.40	8.90	17.39
17	1972	2.23	5.50	8.24	17.35
18	1976	2.25	5.50	8.34	17.29
19	1980	2.36	5.78	8.54	17.35
20	1984	2.35	5.75	8.54	17.56
21	1988	2.38	5.90	8.72	17.61

Example 3.8.4: Windmill data

Wind velocity, x	DC output, y
2.45	0.123
2.70	0.500
2.90	0.653
3.05	0.558
3.40	1.057
3.60	1.137
3.95	1.144
4.10	1.194
4.60	1.562
5.00	1.582
5.45	1.501
5.80	1.737
6.00	1.822
6.20	1.866
6.35	1.930
7.00	1.800
7.40	2.088
7.85	2.179
8.15	2.166
8.80	2.112
9.10	2.303
9.55	2.294
9.70	2.386
10.00	2.236
10.20	2.310

Example 3.8.5: NYAMSE and BSE values

Time point	Calendar Month	NYAMSE	BSE
1	Jan. 1967	10581.6	78.8
2	Feb. 1967	10234.3	69.1
3	Mar. 1967	13299.5	87.6
4	Apr. 1967	10746.5	72.8
5	May 1967	113310.7	79.4
6	Jun. 1967	12835.5	85.6
7	Jul. 1967	12194.2	75.0
8	Aug. 1967	12860.4	85.3
9	Sep.1967	11955.6	86.9
10	Oct. 1967	13351.5	107.8
11	Nov. 1967	13285.9	128.7
12	Dec. 1967	13784.4	134.5
13	Jan. 1968	16336.7	148.7
14	Feb. 1968	11040.5	94.2
15	Mar. 1968	11525.3	128.1
16	Apr. 1968	16056.4	154.1
17	May 1968	18464.3	191.3
18	Jun. 1968	17092.2	191.9

Time point	Calendar Month	NYAMSE	BSE
19	Jul. 1968	15178.8	159.6
20	Aug. 1968	12774.8	185.5
21	Sep. 1968	12377.8	178.0
22	Oct. 1968	16856.3	271.8
23	Nov. 1968	14635.3	212.3
24	Dec. 1968	17436.9	139.4
25	Jan. 1969	16482.2	106.0
26	Feb. 1969	13905.4	112.1
27	Mar. 1969	11973.7	103.5
28	Apr. 1969	12573.6	92.5
29	May 1969	16566.8	116.9
30	Jun. 1969	13558.7	78.9
31	Jul. 1969	11530.9	57.4
32	Aug. 1969	11278.0	75.9
33	Sep. 1969	11263.7	109.8
34	Oct. 1969	15649.5	129.2
35	Nov. 1969	12197.1	115.1

Example 3.8.6: Cotton imports in the 18th century

Year	Imports	Year	Imports
1770	3612	1786	19475
1771	2547	1787	23250
1772	5307	1788	20467
1773	2906	1789	32576
1774	5707	1790	31448
1775	6694	1791	28707
1776	6216	1792	34907
1777	7037	1793	19041
1778	6569	1794	24359
1779	5861	1795	26401
1780	6877	1796	32126
1781	5199	1997	23354
1782	11828	1798	31881
1783	9736	1799	43379
1784	11482	1800	56011
1785	18400		

Example 3.8.7: British exports in the 19th century

Year	Exports	Year	Exports
1820	36.4	1836	53.3
1821	36.7	1837	42.1
1822	37.0	1838	50.1
1823	35.4	1839	53.2
1824	38.4	1840	51.4
1825	38.9	1841	51.6
1826	31.5	1842	47.4
1827	37.2	1843	52.3
1828	36.8	1844	58.6
1829	35.8	1845	60.1
1830	38.3	1846	57.8
1831	37.2	1847	58.8
1832	36.5	1848	52.8
1833	39.7	1849	63.6
1834	41.6	1850	71.4
1835	47.4		

Example 4.3.1: Page's data

Obs. no.	1	2	3	4	5	6	7	8	9	10
Binomial	0	1	1	1	0	0	0	0	0	1
Obs. no.	11	12	13	14	15	16	17	18	19	20
Binomial	1	1	0	0	1	0	0	1	1	1
Obs. no.	21	22	23	24	25	26	27	28	29	30
Binomial	1	1	0	1	1	0	1	1	1	1
Obs. no	31	32	33	34	35	36	37	38	39	40
Binomial	1	1	1	1	1	1	0	1	1	1

Example 4.3.2: Lindisfarne Scribe's data

	No. of occurrences for the following values of i:												
	1	2	3	4	5	6	7	8	9	10	11	12	13
m_i	12	26	31	24	28	34	39	46	41	19	17	17	16
x_i	21	36	44	30	52	45	48	57	48	22	20	21	20

Example 4.3.3 and 4.3.4: Cricket test match outcomes between England and Australia

A	E	A	E	D	A	A	D	A	A	E	E	A	D	E	D	E	E
A	A	E	E	E	E	E	E	E	A	E	E	E	E	A	A	E	D
E	D	E	E	A	A	E	E	A	E	E	A	A	A	A	D	A	D
D	D	E	A	A	A	A	D	D	A	A	E	E	E	A	E	A	E
D	D	E	D	A	E	A	A	A	E	A	A	D	D	A	E	E	E
E	D	D	E	A	A	A	A	A	A	A	A	D	D	A	A	A	E
A	D	D	D	D	E	E	E	E	E	A	E	A	D	D	A	E	A
E	E	E	A	E	D	D	A	E	E	A	A	A	D	D	A	E	A
A	D	D	A	A	A	D	A	A	A	A	A	A	E	D	D	D	D
E	A	E	E	E	D	D	A	E	E	D	A	A	D	A	A	D	A
E	A	D	D	E	A	D	D	D	D	A	D	D	D	D	E	A	D
A	D	D	D	E	D	D	E	D	D	E	E	A	D	E	A	A	A
D	A	A	E	A	D	D	D	A	D	E	E	E	D	E	E	A	E
E	E	A	A	A	D	A	D	E	E	E	D	D	A	A	E	D	A
D	D	E	E	E	D	D	E	A	D	A	A	D	A	A	D		

Example 4.3.5: Sequence of 50 observations for the three-state Markov chain

1	1	2	2	1	1	1	1	1	1	2	3	3	2
3	2	1	1	2	1	1	3	3	1	1	1	3	2
1	1	1	1	1	3	2	2	3	1	2	2	2	2
2	2	2	3	2	3	2	2						

Example 4.3.6: Annual numbers of cases of HUS at each referral centre

Obs.	Year	No. of cases at		Obs.	Year	No. of cases at	
		Newcastle	Birmingham			Newcastle	Birmingham
1	1970	6	1	11	1980	4	1
2	1971	1	5	12	1981	0	7
3	1972	0	3	13	1982	4	11
4	1973	0	2	14	1983	3	4
5	1974	2	2	15	1984	3	7
6	1975	0	1	16	1985	13	10
7	1976	1	0	17	1986	14	16
8	1977	8	0	18	1987	8	16
9	1978	4	2	19	1988	9	9
10	1979	1	1	20	1989	19	15

Example 4.3.7: British coal-mining disaster data by year, 1851-1962

YEAR	COUNT	YEAR	COUNT	YEAR	COUNT	YEAR	COUNT
1851	4	1881	2	1911	0	1941	4
1852	5	1882	5	1912	1	1942	2
1853	4	1883	2	1913	1	1943	0
1854	1	1884	2	1914	1	1944	0
1855	0	1885	3	1915	0	1945	0
1856	4	1886	4	1916	1	1946	1
1857	3	1887	2	1917	0	1947	4
1858	4	1888	1	1918	1	1948	0
1859	0	1889	3	1919	0	1949	0
1860	6	1890	2	1920	0	1950	0
1861	3	1891	2	1921	0	1951	1
1862	3	1892	1	1922	2	1952	0
1863	4	1893	1	1923	1	1953	0
1864	0	1894	1	1924	0	1954	0
1865	2	1895	1	1925	0	1955	0
1866	6	1896	3	1926	0	1956	0
1867	3	1897	0	1927	1	1957	1
1868	3	1898	0	1928	1	1958	0
1869	5	1899	1	1929	0	1959	0
1870	4	1900	0	1930	2	1960	1
1871	5	1901	1	1931	3	1961	0
1872	3	1902	1	1932	3	1962	1
1873	1	1903	0	1933	1		
1874	4	1904	0	1934	1		
1875	4	1905	3	1935	2		
1876	1	1906	1	1936	1		
1877	5	1907	0	1937	1		
1878	5	1908	3	1938	1		
1879	3	1909	2	1939	1		
1880	4	1910	2	1940	2		

Example 4.3.8: **Time intervals in days between explosions in mines, from 15 March 1851 to 22 March 1962 (to be read down columns)**

157	65	53	93	127	176	22	1205	1643	312
123	186	17	24	218	55	61	644	54	536
2	23	538	91	2	93	78	467	326	145
124	92	187	143	0	59	99	871	1312	75
12	197	34	16	378	315	326	48	348	364
4	431	101	27	36	59	275	123	745	37
10	16	41	144	15	61	54	456	217	19
216	154	139	45	31	1	217	498	120	156
80	95	42	6	215	13	113	49	275	47
12	25	1	208	11	189	32	131	20	129
33	19	250	29	137	345	388	182	66	1630
66	78	80	112	4	20	151	255	292	29
232	202	3	43	15	81	361	194	4	217
826	36	324	193	72	286	312	224	368	7
40	110	56	134	96	114	354	566	307	18
12	276	31	420	124	108	307	462	336	1358
29	16	96	95	50	188	275	228	19	2366
190	88	70	125	120	233	78	806	329	952
97	225	41	34	203	28	17	517	330	632

Example 4.3.9: Diaz data (to be read across rows)

2.0777	2.1089	0.4033	2.0729	1.3243
1.5223	3.0164	4.0225	3.3887	0.8362
3.3298	1.0387	1.2537	1.3364	1.2291
1.0502	1.7754	3.9709	1.9282	0.2673

Example 4.3.10: Aircraft arrival times

467	761	792	812	926	1,100	1,147
1,163	1,398	1,462	1,487	1,749	1,865	2,004
2,177	2,208	2,279	2,609	2,682	2,733	2,818
2,837	2,855	2,868	3,089	3,209	3,223	3,233
3,272	3,399	2,595	3,634	3,650	3,851	4,176
4,304	4,391	4,453	4,539	4,748	4,839	5,049
5,202	5,355	5,551	5,598	5,640	5,702	5,935
6,000	6,192	6,435	6,474	6,600	6,810	6,824
7,168	7,181	7,202	7,218	7,408	7,428	7,720
7,755	7,835	7,958	8,307	8,427	8,754	8,819
8,904	8,938	8,980	9,048	9,237	9,268	9,513
9,635	9,750	9,910	9,929	10,167	10,254	10,340
10,624	10,639	10,669	10,889	11,386	11,515	11,651
11,727	11,737	11,844	11,928	12,168	12,657	12,675
12,696	12,732	13,092	13,281	13,536	13,556	13,681
13,710	14,008	14,151	14,601	14,877	14,927	15,032
15,134	15,213	15,491	15,589	15,600	15,631	15,674
15,797	15,953	16,089	16,118	16,215	16,394	16,503
16,515	16,537	16,570	16,597	16,619	16,693	17,314
17,516	17,646	17,770	17,897	17,913	17,922	18,174
18,189	18,328	18,345	18,499	18,521	18,588	19,117
19,150	19,432	19,662	19,758	19,789	19,831	19,978
20,119	20,312	20,346	20,449	20,455	20,604	20,675
20,817	20,898	21,245	21,386	21,562	22,022	22,056
22,095	22,182	22,554	22,764	22,955	22,993	23,025
23,117	23,321	23,341	23,650	23,766	23,879	23,888
24,458	24,889	24,930	24,967	25,224	25,312	25,477
25,498	25,712	25,721	25,884	25,919	25,985	26,196
26,459	26,468	26,494	26,505	26,554	26,906	27,003
27,437	27,661	27,675	27,697	27,721	27,734	27,802
27,971	28,116	29,746				

Example 4.3.11: Stanford heart transplant data

Uncensored observations (45)				Censored observations (24)	
Age	Surv. Time	Age	Surv. Time	Age	Surv. Time
41	5	53	96	35	39
40	16	48	100	28	109
54	16	46	110	23	131
29	17	47	153	26	180
55	28	43	165	47	265
52	30	52	186	44	340
40	39	47	188	54	370
56	43	51	207	48	397
36	45	51	219	52	445
42	51	48	285	46	482
50	53	19	285	48	515
42	58	49	308	52	545
52	61	42	334	26	596
61	66	47	342	47	620
45	68	48	583	47	670
49	68	50	675	32	841
53	72	58	733	41	915
47	72	44	852	38	941
64	77	45	979	36	1141
51	78	48	995	45	1321
53	80	43	1032	48	1407
54	81	53	1386	40	1571
56	90			48	1586
				33	1799

Example 5.4.1: **Ordered remission durations for 84 patients with acute non-lymphoblastic leukemia**

Uncensored observations (51)							
24	46	57	57	64	65	82	89
90	90	111	117	128	143	148	152
166	171	186	191	197	209	223	230
247	249	254	258	264	269	270	273
284	294	304	304	332	341	393	395
487	510	516	518	518	534	608	642
697	955	1160					
Censored observations (33)							
68	119	182	182	182	182	182	182
182	182	182	182	182	182	182	182
182	182	182	182	182	182	182	182
182	182	583	1310	1538	1634	1908	1996
2057							

Example 5.4.2: Failure times for a piece of electronic equipment

1.0	6.4	19.2	54.2	88.4	114.98
1.2	6.8	28.1	55.6	89.9	115.1
1.3	6.9	28.2	56.4	90.8	117.4
2.0	7.2	29.0	58.3	91.1	118.3
2.4	7.9	29.9	60.2	91.5	119.7
2.9	8.3	30.6	63.7	92.1	120.6
3.0	8.7	32.4	64.6	97.9	121.0
3.1	9.2	33.9	65.3	100.8	122.9
3.3	9.8	35.3	66.2	102.6	123.3
3.5	10.2	36.1	70.1	103.2	124.5
3.8	10.4	40.1	71.0	104.0	125.8
4.3	11.9	42.8	75.1	104.3	126.6
4.6	13.8	43.7	75.6	105.0	127.7
4.7	14.4	44.5	78.4	105.8	128.4
4.8	15.6	50.4	79.2	106.5	129.2
5.2	16.2	51.2	84.1	110.7	129.5
5.4	17.0	52.0	86.0	112.6	129.9
5.9	17.5	53.3	87.9	113.5	

BIBLIOGRAPHY

[1.] Achcar, J.A. and Bolfarine, H. (1989): Constant hazard against a change-point alternative. A Bayesian approach with censored data. *Comm. Stat. - Theory Meth.*, 18(10), 3801-3819.

[2.] Aitkin, M. (1991): Posterior Bayes factors (with discussion). *J. Royal Stat. Soc.* B., 53, 111-142.

[3.] Akman, V.E. and Raftery, A.E. (1986): Bayesian analysis of a poisson process with a change-point. *Biometrika.*, 73(1), 85-89.

[4.] Aly, E.A.A. and BuHamra, S.S. (1996): Rank tests for two change-points. *Computational Stat. and Data Analysis.*, 22, 363-372.

[5.] Anscombe, F.J. (1960): Rejection of outliers. *Technometrics.*, 2, 123-146.

[6.] Atkinson, A.C. (1978): Posterior probabilities for choosing a regression model. *Biometrika.*, 65, 39-48.

[7.] Bacon, D.W. and Watts, D.G. (1971): Estimating the transition between two intersecting straight lines. *Biometrika.*, 58, 525-534.

[8.] Barry, D. and Hartigan, J.A. (1992): Product partition models for change-point problems. *Ann. Stat.*, 20, 260-279.

[9.] Barry, D. and Hartigan, J.A. (1993): A Bayesian analysis for change-point problems. *J. Amer. Stat. Ass.*, 88(421), 309-319.

[10.] Barry, D. and Hartigan, J.A. (1995): Change-points in 0-1 sequences, with an application to predicting divisional winners in major league baseball. Applied change-point problems in statistics. Nova Science Publishers, 1-14.

[11.] Basu, A.P., Ghosh, J.K. and Joshi, S.N. (1988): On estimating change-point in a Hazard rate, Statistical decision theory and related topics, (S.S. Gupta and J.O. Berger eds) Springer-Verlag, New York, 239-252.

[12.] Bélisle, P., Joseph, L., MacGibbon, B., Wolfson, D.B. and du Berger, R. (1998): Change-point analysis of neuron spike train data. *Biometrics,*, 54, 113-123.

[13.] Berger, J.O. and Berry, D.A. (1988): The relevance of stopping rules in statistical inference. In Statistical Decision Theory and Related Topics IV, eds. S.S. Gupta and J.O. Berger., 1, 29-47.

[14.] Berger, J.O. and Delampady, M. (1987): Testing precise hypothesis. *Stat. Sci.*, 3, 317-352.

[15.] Berger, J. and Pericchi, L. (1995): "The Intrinsic Bayes Factor for Linear Models", Bayesian Statistics V, eds. J.M. Bernardo, *et al.*, London: Oxford University Press, 23-42.

[16.] Berger, J. and Pericchi, L. (1996): The Intrinsic Bayes Factor for model selection and prediction. *J. Amer. Stat. Ass.*, 91, 109-122.

[17.] Berger, J. and Pericchi, L. (1997): "On the justification of default and Intrinsic Bayes Factors", Modeling and Prediction, eds. J.C. Lee *et al.*, New York: Springer-Verlag, 276-293.

[18.] Berger, J. and Pericchi, L. (1998): Accurate and stable Bayesian model selection: the median intrinsic Bayes factor. *Sankhya B.*, 60, 1-18.

[19.] Bhattacharya, P.K. (1987): Maximum likelihood estimation of a change-point in the distribution of independent random variables: General Multiparameter case. *J. Multiv. Analysis.*, 23, 183-208.

[20.] Bhattacharya, P.K. and Brockwell, P.J. (1976): The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrsch. Verw. Gebiete.*, 37, 51-75.

[21.] Bhattacharya, P.K. and Johnson, R.A. (1968): Non-parametric tests for shift at unknown time point. *Ann. Math. Stat.*, 39, 1731-1743.

[22.] Booth, N.B. and Smith, A.F.M. (1982): A Bayesian approach to retrospective identification of change-points. *J. Econo.*, 19, 7-22.

[23.] Box, G.E.P. and Tiao, G.C. (1962): A further look at robustness via Bayes' theorem. *Biometrika.*, 49, 419-432.

[24.] Box, G.E.P. and Tiao, G.C. (1964): A Bayesian approach to the importance of assumptions applied to the comparison of variances. *Biometrika.*, 51, 153-167.

[25.] Box, G.E.P. and Tiao, G.C. (1965): A change in level of a non-stationary time series. *Biometrika.*, 52, 181-192.

[26.] Box, G.E.P. and Tiao, G.C. (1968): A Bayesian approach to some outlier problems. *Biometrika.*, 55, 119-129.

[27.] Box, G.E.P. and Tiao, G.C. (1975): Intervention analysis with applications to economic and environmental problems. *J. Amer. Stat. Ass.*, 70, 70-79.

[28.] Broemeling, L.D. (1972): Bayesian procedures for detecting a change in a sequence of random variables. *Metron.*, 30, 1-14.

[29.] Broemeling, L.D. (1974): Bayesian inferences about a changing sequence of random variables. *Comm. Stat.*, 3, 243-255.

[30.] Broemeling, L.D. (1977): Forecasting future values of a changing sequence. *Comm. Stat.*, A6, 87-102.

[31.] Broemeling, L.D. and Gregurich, M.A. (1996): On a Bayesian approach for the shift point problem. *Comm. Stat. - Theory Meth.*, 25(10), 2267-2279.

[32.] Broemeling, L.D. and Tsurumi, H. (1987): Econometrics and structural change. Marcel Dekker Inc., New York.

[33.] Brown, R.L., Durbin, J. and Evans, J.M. (1975): Techniques for testing the constancy of regression relations over time (with discussion). *J. Royal Stat. Soc.* B., 149-192.

[34.] Butlamra, S.S. (1997): Testing for a change in repeated measures data. *Comm. Stat. - Simula.*, 26(3), 841-872.

[35.] Carlin, B.P., Gelfand, A.E. and Smith, A.F.M. (1992): Hierarchical Bayesian analysis of change-point problems. *Appl. Stat.*, 41, 389-405.

[36.] Caussinus, H. and Lyazrhi, F. (1997): Choosing a linear model with a random number of change-points and outliers. *Ann. Inst. Stat.* Ma., 49(4), 761-775.

[37.] Chen, J. (1998): Testing for a change-point in linear regression models. *Comm. Stat. Theory Meth.*, 27(10), 2481-2493.

[38.] Chen, J. and Gupta, A.K. (1997): Testing and locating variance change-points with application to stock prices. *J. Amer. Stat. Ass.*, 92(438), 739-747.

[39.] Chen, J. and Gupta, A.K. (1999): Change-point analysis of a Gaussian model. *Stat. Papers.*, 40, 323-333.

[40.] Chen, J. and Gupta, A.K. (2000): Parametric statistical change-point analysis. Birkhäuser.

[41.] Chernoff, H. (1973): The use of faces to represent points in k-dimensional space graphically. *J. Amer. Stat. Ass.*, 68, 361-368.

[42.] Chernoff, H. and Zacks, S. (1964): Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Math. Stat.*, 35, 999-1018.

[43.] Chin Choy, J.H. (1977): A Bayesian analysis of a changing linear model, doctoral dissertation, Oklahoma State University, Stillwater, Okla.

[44.] Chin Choy, J.H. and Broemeling, L. (1980): Some Bayesian inferences for a changing linear model. *Technometrics.*, 22, 71-78.

[45.] Coad, N.A.G., Marshall, T., Rowe, B. and Taylor, C.M. (1991): Changes in the postenteropathic form of the hemolytic uremic syndrome in children. *Clin. Nephrol.*, 35, 10-16.

[46.] Colwell, D., Jones, B. and Gillett, J. (1990): A Markov chain in cricket (MCC!). *The Mathematical Gazette.*, 183-185.

[47.] Consonni, G. and Veronese, P. (1991): Bayes Factors for linear models and improper priors. Technical Report no. 34. Universita L. Bocconi, Milano.

[48.] Cook, J.P. (1983): A Bayesian analysis of autoregressive processes: time and frequency domain, Unpublished Ph.D. dissertation, Oklahoma State University, Stillwater, Okla.

[49.] Daniel, C. (1960): Locating outliers in factorial experiments. *Technometrics.*, 2, 149-156.

[50.] Darkhovskv, B.S. (1976): A non-parametric method for the aposteriori detection of the "disorder" time of a sequence of independent random variables. *Theory Prob. Appl.*, 21, 178-183.

[51.] Davis, W.W. (1979): Robust methods for detection of shifts of the innovation variance of a time series. *Technometrics.*, 21, 313-320.

[52.] De Vos, A.F. (1993): A fair comparison between regression models of different dimension. Technical report, The Free University, Amsterdam.

[53.] Dey, D.K. and Purkayastha, S. (1997): Bayesian approach to change-point problems. *Comm. Stat. – Theory Meth.*, 26(8), 2035-2047.

[54.] Diaz, J. (1982): Bayesian detection of a change of scale parameter in sequences of independent gamma random variables. *J. Econo.*, 19, 23-30.

[55.] Eastwood, V.R. (1993): Some non-parametric methods for change-point problems. *Canadian J. Stat.*, 21(2), 209-222.

[56.] Farley, J.U. and Hinick, M.J. (1970): A test for shifting regression model. *J. Amer. Stat. Ass.*, 65, 1320-1329.

[57.] Farley, J.U., Hinick, M.J. and McGuire, T.W. (1975): Some comparisons of tests for a shift in the slopes of multivariate time series. *J. Econo.*, 3, 297-313.

[58.] Feder, P. (1975): The log likelihood ratio in segmented regression. *Ann. Stat.*, 3, 84-97.

[59.] Ferreira, P.E. (1975): A Bayesian analysis of a switching regression model: a known number of regimes. *J. Amer. Stat. Ass.*, 70, 370-374.

[60.] Gardner, L.A. (1969): On detecting changes in the mean of normal variables. *Ann. Math. Stat.*, 40, 116-126.

[61.] Garisch, I. and Groenewald, P.C.N. (1999): The Nile revisited: Change-point analysis with autocorrelation. *Bayesian Stat.*, 6, 753-760.

[62.] Gebhardt, F. (1964): On the risk of some strategies for outlying observations. *Ann. Math. Stat.*, 35, 1524-1536.

[63.] Geisser, S. and Eddy, W.F. (1979): A predictive approach to model selection. *J. Amer. Stat. Ass.*, 74, 153-160.

[64.] Gelfand, A.E. and Dey, D.K. (1994): Bayesian model choice: Asymptotics and exact calculations. *J. Royal Stat. Soc.* B., 56, 501-514.

[65.] Gelfand, A.E., Dey, D.K. and Chang, H. (1992): Model determination using predictive distributions with implementations via sampling-based methods. In Bayesian Statistics 4 (eds. J.M. Bernardo *et al.*), Oxford University Press, 147-167.

[66.] Ghosal, S., Ghosh, J.K. and Samanta, T. (1999): Approximation of the posterior distribution in a change-point problem. *Ann. Inst. Stat. Math.*, 51(3), 479-497.

[67.] Ghosh, J.K. and Joshi, S.N. (1992): On the asymptotic distribution of an estimate of the change-point in a failure rate. *Comm. Stat. - Theory Meth.*, 21(12), 3571-3588.

[68.] Ghosh, J.K., Joshi, S.N. and Mukhopadhyay, C. (1993): A Bayesian approach to the estimation of the change-point in a Hazard rate. Advances in reliability, (A.P. Basu ed.) Elsevier Science Publishers, North-Holland, 141-170.

[69.] Ghosh, J.K., Joshi, S.N. and Mukhopadhyay, C. (1996): Asymptotics of a Bayesian approach to estimating change-point in a hazard rate. *Comm. Stat. - Theory Meth.*, 25(12), 3147-3166.

[70.] Gilks, W.R. (1995): Discussion of the paper by O'Hagan. *J. Royal Stat. Soc.* B., 97, 118-120.

[71.] Gombay, E., Horvath, L. and Huskova, M. (1996): Estimators and tests for change in variances. *Stat. and Decisions.*, 14, 145-159.

[72.] Good, I.J. (1947-50): Probability and the weighing of evidence. London: Griffin. New York: Haffner.

[73.] Good, I.J. (1983): Explicativity, corroboration and the relative odds of hypotheses (# 846). In Good thinking - The foundations of probability and its applications. University of Minnesota Press, Minneapolis, MN.

[74.] Gradshteyn, I.S. and Ryzhik, I.M. (1980): Table of integrals, series, and products, Fourth Edition, Academic Press, London.

[75.] Green, P.J. (1995): Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika.*, 82(4), 711-732.

[76.] Grenander, U. and Miller, M.I. (1991): Jump-diffusion processes for abduction and recognition of biological shapes. Technical report, Electronic signals and systems research laboratory, Washington University.

[77.] Grenander, U. and Miller, M.I. (1994): Representations of knowledge in complex systems. *J. Royal Stat. Soc.* B., 56, 549-603.

[78.] Groenewald, P.C.N. (1993): Bayes analysis of change-points in the linear model. Proceedings of the section on Bayesian statistical science, *Amer. Stat. Ass.*, 45-50.

[79.] Grubbs, F.E. (1950): Sample criteria for testing outlying observations. *Ann. Math. Stat.*, 21, 27-58.

[80.] Gupta, A.K. and Chen, J. (1996): Detecting changes of mean in multidimensional normal sequences with applications to literature and geology. *Comput. Stat.*, 11, 211-221.

[81.] Haccou, P. and Meelis, E. (1988): Testing for the number of change-points in a sequence of exponential random variables. *J. Stat. Comput. Simul.*, 30, 285-298.

[82.] Hand, D.J., Daly, F., Runn, A.D., McConway, K.J. and Ostrowski, E. (1994): A handbook of small data sets. Chapman and Hall, London.

[83.] Harrison, P.J. and Stevens, C.F. (1976): Bayesian forecasting (with discussion). *J. Royal Stat. Soc.* B., 38, 205-247.

[84.] Hartigan, J.A. (1990): Partition models. *Comm. Stat. - Theory Meth.*, 19(8), 2745-2756.

[85.] Hastie, T. and Stuetzle, W.X. (1989) : Principal curves. *J. Amer. Stat. Ass.*, 84, 502-516.

[86.] Hawkins, D.M. (1977): Testing a sequence of observation for a shift in location. *J. Amer. Stat. Ass.*, 72, 180-186.

[87.] Hawkins, D.L. (1989): A $U-I$ approach to retrospective testing for shifting parameters in a linear model. *Comm. Stat. - Theory Meth.*, 18, 3117-3134.

[88.] Henderson, R. and Matthews, J.N.S. (1993): An investigation of change-points in the annual number of cases of haemolytic uraemic syndrome. *Appl. Stat.*, 42(3), 461-471.

[89.] Hinkley, D.V. (1969): Inference about the intersection in two-phase regression. *Biometrika.*, 56, 495-504.

[90.] Hinkley, D.V. (1970): Inference about the change-point in a sequence of random variables. *Biometrika.*, 57, 1-17.

[91.] Hinkley, D.V. (1971): Inference in two-phase regression. *J. Amer. Stat. Ass.*, 66, 736-793.

[92.] Hinkley, D.V., Chapman, P. and Runger, G. (1980): Change-point problems. Tech. Report 382, University of Minnesota.

[93.] Hinkley, D.V. and Hinkley, E. (1970): Inference about the change-point in a sequence of binomial variables. *Biometrika.*, 57(3), 477-488.

[94.] Hirotsu, C. (1997): Two-way change-point model and its application. *Austral. J. Stat.*, 39(2), 205-218.

[95.] Holbert, D. (1973): A Bayesian analysis of shifting sequences with applications to two-phase regression. Ph.D. dissertation, Oklahoma State University, Stillwater, Okla.

[96.] Holbert, D. (1982): A Bayesian analysis of a switching linear model. *J. Econo.*, 19, 77-88.

[97.] Holbert, D. and Broemeling, L. (1977): Bayesian inference relating to shifting sequences and two-phase regression. *Comm. Stat. - Theory Meth.* A., 6, 165-275.

[98.] Hsu, D.A. (1977): Tests for variance shift at an unknown time point. *Appl. Stat.*, 26, 279-284.

[99.] Hsu, D.A. (1979): Detecting shifts of parameters in gamma sequences with applications to stock price and air traffic flow analysis. *J. Amer. Stat. Ass.*, 74, 31-40.

[100.] Hsu, D.A. (1982): A Bayesian robust detection of shift in risk structure of stock market returns. *J. Amer. Stat. Ass.*, 77, 29-39.

[101.] Hudson, D.J. (1966): Fitting segmented curves whose joint points have to be estimated. *J. Amer. Stat. Ass.*, 61, 1097-1129.

[102.] Ilmakunnas, P. and Tsurumi, H. (1984): Testing for parameter shifts in a regression model with two regimes of autocorrelated errors. *Economic Studies Quarterly.*, 35, 46-56.

[103.] Ilmakunnas, P. and Tsurumi, H. (1985): Testing the Lucas hypothesis on output-inflation trade-off. *J. Business and Econo. Stat.*, 3, 43-53.

[104.] Inclán, C. and Tiao, G. (1994): Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Stat. Ass.*, 89, 913-923.

[105.] Jandhyala, V.K. and MacNeill, I.B. (1991): Tests for parameter changes at unknown times in linear regressin models. *J. Stat. Plan. Infer.*, 27, 291-316.

[106.] Jandhyala, V.K. and Minogue, C.D. (1993): Distributions of Bayes type change-point statistics under polynomial regression. *J. Stat. Plan. Infer.*, 37, 271-290.

[107.] Jani, P.N. and Pandya, M. (1999): Bayes estimation of shift point in left truncated exponential sequence. *Comm. Stat. - Theory Meth.*, 28(11), 2623-2639.

[108.] Jarrett, R.G. (1979): A note on the intervals between coal-mining disasters. *Biometrika.*, 66(1), 191-193.

[109.] Jeffreys, H. (1935): Some tests of significance, treated by the theory of probability. *Proc. Camb. Phil. Soc.*, 31, 203-222.

[110.] Jeffreys, H. (1961): Theory of Probability, Third Edition. Oxford University Press.

[111.] Joseph, L., Vandal, A.C. and Wolfson, D.B. (1996): Estimation in the multi-path change-point problem for correlated data. *Canadian J. Stat.*, 24, 37-53.

[112.] Joseph, L. and Wolfson, D.B. (1992): Estimation in multi-path change-point problems. *Comm. Stat. - Theory Meth.*, 21(4), 897-913.

[113.] Joseph, L., Wolfson, D.B., Du Berger, R. and Lyle, R.M. (1997): Analysis of panel data with change-points. *Statistica Sinica.*, 7, 687-703.

[114.] Juran, J. and Gryna, F. (1980): Quality planning and analysis. McGraw-Hill, New York.

[115.] Kalbfleisch, J.D. and Prentice, R.L. (1980): The statistical analysis of failure time data. Wiley, New York.

[116.] Kander, Z. and Zacks, S. (1966): Test procedures for possible changes in parameters of statistical distributions occuring at unknown time points. *Ann. Math. Stat.*, 37, 1696-1710.

[117.] Karunamuni, R.J. and Zhang, S. (1996): Empirical Bayes detection of a change in distribution. *Ann. Inst. Stat. Math.*, 48(2), 229-246.

[118.] Kashiwagi, N. (1991): Bayesian detection of structural changes. *Ann. Inst. Stat. Math.*, 43(1), 77-93.

[119.] Kass, R.E. and Raftery, A.E. (1995): Bayes Factors. *J. Amer. Stat. Ass.*, 90, 773-795.

[120.] Kenett, R. and Pollak, M. (1983): On sequential detection of a shift in the probability of a rare event. *J. Amer. Stat. Ass.*, 78(382),389-395.

[121.] Kim, D.C. (1991): A Bayesian significance test of the stationarity of regression parameters. *Biometrika.*, 78, 667-675.

[122.] Kim, D. (1994): Tests for a change-point in linear regression. *IMS Lecture Notes – Monograph Series.*, 23, 170-176.

[123.] Kim, S., Cho, S. and Lee, S. (2000): On the CUSUM test for parameter changes in GARCH (1,1) models. *Comm. Stat. - Theory Meth.*, 29(2), 445-462.

[124.] Kiuchi, A.S., Hartigan, J.A., Holford, T.R., Rubinstein, P. and Stevens, C.E. (1995): Change-points in the series of $T4$ counts prior to AIDS. *Biometrics.*, 51, 236-248.

[125.] Kulasekera, K.B. and Saxena, K.M.L. (1991): Estimation of change-point in failure rate models. *J. Stat. Plan. Infer.*, 29, 111-124.

[126.] Lai, T.L. (1995): Sequential change-point detection in quality control and dynamical systems. *J. Royal Stat. Soc.* B., 57(4), 613-658.

[127.] Land, M. and Broemeling, L.D. (1983): Bayesian forecasting with changing linear models. *Comm. Stat. - Theory Meth.*, 12, 1421-1430.

[128.] Lee, C. (1998): Bayesian estimation of the number of change-points. *Stat. Sinica.*, 8, 923-939.

[129.] Lee, A.F.S. and Heghinian, S.M. (1977): A shift of the mean level in a sequence of independent normal random variables. A Bayesian approach. *Technometrics.*, 19, 503-506.

[130.] Lempers, F.B. (1971): Posterior probabilities of alternative linear models. Rotterdam: University Press.

[131.] Levin, M. and Barrett, T.M. (1984): Haemolytic uraemic syndrome. *Arch. Dis. Chldhd.*, 59, 397-400.

[132.] Lindley, D.V. (1957): A statistical paradox. *Biometrika.*, 44, 187-192.

[133.] Littlewood, B. (1981): Stochastic reliability-growth: A model for fault removal in computer programs and hardware design, IEEE Transactions on reliability R-30, No. 4, 313-320.

[134.] Lombard, F. (1987): Rank tests for change-point problems. *Biometrika.*, 74, 615-624.

[135.] Lyazrhi, F. (1994): Bayesian criteria for discriminating among regression models with one possible change-point. Technical report. Université Paul Sabatier. Toulouse.

[136.] MacNeill, I.B., Tang, S.M. and Jandhyala, V.K. (1991): A search for the source of the Nile's change-points. *Environmetrics.*, 2(3), 341-375.

[137.] Maguire, B.A., Pearson, E.S. and Wynn, A.H.A. (1952): The time intervals between industrial accidents. *Biometrika.*, 38, 168-180.

[138.] Maronna, R. and Yohai (1978): A bivariate test for the detection of a systematic change in means. *J. Amer. Stat. Ass.*, 73, 640-645.

[139.] Mascarenhas, N.D.A. and Prado, L.O.C. (1980): A Bayesian approach to edge detection in images, IEEE Transactions on Automatic Control, AC-25, 36-43.

[140.] Matthews, D.E. and Farewell, V.T. (1982): On testing for a constant Hazard against a change-point alternative. *Biometrics.*, 38, 463-468.

[141.] Matthews, D.E., Farewell, V.T. and Pyke, R. (1985): Asymptotic score statistic processes and tests for constant Hazard against a change-point alternative. *Ann. Stat.*, 13, 583-591.

[142.] McAleer, M. and Fisher, G. (1982): Testing separate regression models subject to specification error. *J. Econo.*, 19, 125-146.

[143.] Menzefricke, U. (1981): A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. *Appl. Stat.*, 30, 141-146.

[144.] Miller, R.B., Wichern, D.W. and Hsu, D.A. (1971): Changes in variances in time series models, Technical Report No. 284, October 1971, Department of Statistics, The University of Wisconsin, Madison, Wisconsin.

[145.] Mitra, M. and Basu, S.K. (1995): Change-point estimation in non-monotonic aging models. *Ann. Inst. Stat. Math.*, 47(3), 483-491.

[146.] Moen, D.H. (1983): The Bayesian analysis of structural change in multivariate linear models. Doctoral dissertation. Oklahoma State University, Stillwater, Okla.

[147.] Moen, D.H., Salazar, D. and Broemeling, L.D. (1985): Structural changes in multivariate regression models. *Comm. Stat. - Theory Meth.* A., 14, 1757-1768.

[148.] Ng, V.M. (1990): Bayesian analysis of linear models exhibiting changes in mean and precision at an unknown time point. *Comm. Stat. - Theory Meth.*, 19, 111-120.

[149.] Nguyen, H.T., Rogers, G.S. and Walker, E.A. (1984): Estimation in change-point hazard rate models. *Biometrika.*, 71(2), 299-304.

[150.] Nordhaus, N.D. and Samuelson, D.A. (1985): Thoughts on rewriting "economics". New York Times, Sunday, February 17, pp. F1, F10.

[151.] O'Hagan, A. (1991): Contribution to the discussion of "Posterior Bayes factors". *J. Royal Stat. Soc.* B., 53, 137.

[152.] O'Hagan, A. (1995): Fractional Bayes factors for model comparison. *J. Royal Stat. Soc.* B., 57(1), 99-138.

[153.] O'Hagan, A. (1997): Properties of intrinsic and fractional Bayes factors. *Test.*, 6, 101-118.

[154.] Ohtani, K. (1982): Bayesian estimation of the switching regression model with autocorrelated errors. *J. Econo.*, 18, 251-261.

[155.] Page, E.S. (1954): Continuous inspection scheme. *Biometrika.*, 41, 100-114.

[156.] Page, E.S. (1955): A test for change in a parameter occuring at an unknown point. *Biometrika.*, 42, 525-527.

[157.] Page, E.S. (1957): On problems in which a change in parameter occurs at an unknown point. *Biometrika.*, 44, 248-252.

[158.] Pettitt, A.N. (1979): A non-parametric approach to the change-point problem. *Appl. Stat.*, 28(2), 126-135.

[159.] Pham, T.D. and Nguyen, H.T. (1990): Strong consistency of the maximum likelihood estimators in the change-point Hazard rate model. *Stat.*, 21(2), 203-216.

[160.] Pham, T.D. and Nguyen, H.T. (1993): Bootstrapping the change-point of a Hazard rate. *Ann. Inst. Stat. Math.*, 45, 331-340.

[161.] Phillips, D.B. and Smith, A.F.M. (1996): Bayesian model comparison via jump diffusions. In Markov Chain Monte Carlo in Practice (eds. W.R. Gilks, S. Richardson and D.J. Spiegelhalter), 215-239. London: Chapman and Hall.

[162.] Poirier, D.J. (1976): The econometrics of structural change. North Holland Publishing Company, Amsterdam.

[163.] Pollak, M. (1985): Optimal detection of a change in distribuiton. *Ann. Stat.*, 13, 206-227.

[164.] Quandt, R.E. (1958): The estimation of the parameters of a linear regression system obeying two separate regimes. *J. Amer. Stat. Ass.*, 53, 873-880.

[165.] Quandt, R.E. (1960): Tests of the hypothesis that a linear regression obeys two separate regimes. *J. Amer. Stat. Ass.*, 55, 324-330.

[166.] Raftery, A.E. (1995): Change-point and change modelling in stochastic processes and spatial statistics. Applied change-point problems in statistics. Nova Science Publishers.

[167.] Raftery, A.E. (1995): Hypothesis testing and model selection. In Markov Chain Monte Carlo in Practice (eds. W.R. Gilks, S. Richardson and D.J. Spiegelhalter), 163-187. London: Chapman and Hall.

[168.] Raftery, A.E. and Akman, V.E. (1986): Bayesian analysis of a Poisson process with a change-point. *Biometrika.*, 73(1), 85-89.

[169.] Ramanayake, A. and Gupta, A.K. (1998): Tests for an Epidemic change in a sequence of Exponentially distributed random variables. Technical Report, Bowling Green State University.

[170.] Robison, D.E. (1964): Estimates for the points of intersection of two polynomial regressions. *J. Amer. Stat. Ass.*, 59, 214-224.

[171.] Ross, A.S.C. (1950): Philological probability problems. *J.Royal Stat. Soc.* B., 12, 19-40.

[172.] Rotondi, R. and Pagliano, P. (1994): An identification of a multiple change-point sequence via a Gibbs sampler-based approach. Technical Report: IAMI 94.7.

[173.] Rukhin, A.L. (1996): Change-point analysis as a multiple decision problem. *Stat. and Decisions.*, 14, 103-114.

[174.] Salazar, D. (1980): The analysis of structural changes in time series and multivariate linear models. Ph.D. dissertation, Oklahoma State University, Stillwater, Oklahoma.

[175.] Salazar, D. (1982): Structural changes in time series models. *J. Econo.*, 19, 147-164.

[176.] Sanso, B. and Pericchi, L.R. (1995): Calculating intrinsic Bayes factors using Monte Carlo. Technical Report. Universidad Simón Bolivar, Caracas, Venezuela.

[177.] Sarkar, P. and Meeker, W.Q. (1997): A Bayesian on-line change detection algorithm with process monitoring applications. Technical report # 97-1. Department of Statistics. Iowa State University.

[178.] Schneider, H., Lin, B-S. and O'cinneide, C. (1990): Comparisons of non-parametric estimators for the renewal function. *Appl. Stat.*, 39, 55-61.

[179.] Schulze, U. (1982): Estimation in segmented regression. *Stat.*, 13(2), 295-317.

[180.] Schwarz, G. (1978): Estimating the dimension of a model. *Ann. Stat.*, 6, 461-464.

[181.] Seki, Y. and Hashimoto, T. (1996): A study on change detection problem of Normal population based on MDL criterion. *J. Japan Industr. Manag. Ass.*, 47, 192-198.

[182.] Sen, A.K. and Srivastava, M.S. (1973): On multivariate tests for detecting change in mean. *Sankhyā* A., 35, 173-186.

[183.] Sen, A.K. and Srivastava, M.S. (1975): On tests for detecting change in mean. *Ann. Stat.*, 3, 98-108.

[184.] Siegmund, D. (1986): Boundary crossing probabilities and statistical applications. *Ann. Stat.*, 14, 361-404.

[185.] Silvey, S.D. (1958): The Lindisfarne Scribe's problem. *J. Royal Stat. Soc.* B., 20, 93-101.

[186.] Smith, A.F.M. (1975): A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika.*, 62, 407-416.

[187.] Smith, A.F.M. and Cook, D.G. (1980): Straight lines with a change-point: a Bayesian analysis of some renal transplant data. *Appl. Stat.*, 29, 180-189.

[188.] Spiegelhalter, D.J. and Smith, A.F.M. (1982): Bayes factors for linear and log-linear models with vague prior information. *J. Royal Stat. Soc.* B., 44, 377-387.

[189.] Sprent, P. (1961): Some hypotheses concerning two-phase regression lines. *Biometrika.*, 17, 634-645.

[190.] Srivastava, M.S. and Worsley, K.J. (1986): Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Stat. Ass.*, 81, 199-204.

[191.] Stephens, D.A. (1994): Bayesian retrospective multiple change-point identification. *Appl. Stat.*, 43(1), 159-178.

[192.] Stephens, D.A. and Smith, A.F.M. (1992): Bayesian edge-detection in images via change-point methods. Technical Report, Imperial College. 92-02.

[193.] Sugiura, N. and Ogden, R.T. (1994): Testing change-points with linear trend. *Comm. Stat. - Theory Meth.*, 23(2), 287-322.

[194.] Suzuki, H. (2000): Recognition of unnatural patterns in manufacturing processes using the minimum description length criterion. *Comm. Stat. - Simula.*, 29(2), 583-601.

[195.] Suzuki, H. and Enkawa, T. (1995): Estimation of a change-point in a fraction defective based on minimum description length criterion. *J. Japan Soc. Qual. Control.*, 25, 181-190.

[196.] Swamy, P.A.V.B. and Mehta, J.S. (1975): Bayesian and non-Bayesian analyzes of switching regressions and of random coefficient regression models. *J. Amer. Stat. Ass.*, 76, 593-602.

[197.] Tarr, P.I., Neill, M.A., Allen, J., Siccardi, C.J., Watkins, S.L. and Hickman, R.O. (1989): The increasing incidence of the hemolytic-uremic syndrome in King county, Washington: lack of evidence for ascertainment bias. *Am. J. Epidem.*, 129, 582-586.

[198.] Thompson, W.R. (1935): On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Ann. Math. Stat.*, 6, 214-219.

[199.] Tiao, G.C. and Tan, W.Y. (1966): Bayesian analysis of random-effect models in the analysis of variance. II. Effect of autocorrelated errors. *Biometrika.*, 53, 477-495.

[200.] Tsurumi, H. (1976): A Bayesian test of the product life cycle hypothesis applied to Japanese crude steel production. *J. Econo.*, 5, 175-189.

[201.] Tsurumi, H. and Sheflin, N. (1984): Bayesian tests of a parameter shift under heteroscedasticity: Weighted – T vs. Double-T approaches. *Comm. Stat. – Theory Meth.*, 13(8), 1003-1013.

[202.] Wang, C. and Lee, C. (1993): Bayesian analysis for a change in the intercept of simple linear regression. *Comm. Stat. – Theory Meth.*, 22(11), 3031-3050.

[203.] Wolfe, D.A. and Chen, Y. (1990): The change-point problem in a multinomial sequence. *Comm. Stat – Theory Meth.*, 19(2), 603-618.

[204.] Wolfe, D.A. and Schechtman, E. (1984): Non-parametric statistical procedures for the change-point problem. *J. Stat. Plan. Infer.*, 9, 389-396.

[205.] Worsley, K.J. (1979): On the likelihood ratio test for shift in location of normal population. *J. Amer. Stat. Ass.*, 74, 36-57.

[206.] Worsley, K.J. (1983): Testing for a two-phase multiple regression. *Technometrics.*, 25, 35-42.

[207.] Worsley, K.J. (1986): Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika.*, 73(1), 91-104.

[208.] Xiong, C. and Milliken, G.A. (2000): Change-points in stochastic ordering. *Comm. Stat. – Theory Meth.*, 29, 381-399.

[209.] Yakir, B. (1997): A note on optimal detection of a change in distribution. *Ann. Stat.*, 25, 2117-2126.

[210.] Yao, Q. (1993): Test of change-points with epidemic alternatives. *Biometrika.*, 80, 170-191.

[211.] Yao, Y.C. (1986): Maximum likelihood estimation in Hazard rate models with change-point. *Comm. Stat. - Theory Meth.*, 15, 2455-2466.

[212.] Yao, Y.C. (1988): Estimating the number of change-points by Schwarz' criterion. *Stat. Prabab. Lett.*, 6, 181-187.

[213.] Zacks, S. (1972): On detecting the epoch at which the failure rate of a reliability system starts to increase. Tech. Memo., Serial TM. -80320, Program in Logistics, George Washington University.

[214.] Zacks, S. (1982): Classical and Bayesian approaches to the change-point problem: Fixed samples and sequential procedures. *Statistique et Analyse des Données.*, 7(1), 48-68.

[215.] Zacks, S. (1983): Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. In Recent Advances in Statistics: Herman Chernoff Festschrift, 245-269, Academic Press, New York/London.

[216.] Zacks, S. (1991): Detection and change-point problems. Handbook of sequential analysis (eds. B.K. Gosh and P.K. Sen), 531-562.

[217.] Zacks, S. (1995): Sequential testing of reliability system with change-points. Applied change-point problems in statistics. Nova Science Publishers.

[218.] Zacks, S. and Barzily, Z. (1981): Bayes procedures for detecting a shift in the probability of success in a series of Bernoulli trials. *J. Stat. Plan. Infer.*, 5, 107-119.

[219.] Zellner, A. and Tiao, G.C. (1964): Bayesian analysis of the regression model with autocorrelated errors. *J. Amer. Stat. Ass.*, 59, 763-778.

[220.] Zhang, H. (1995): Detecting change-points and monitoring biomedical data. *Comm. Stat. - Theory Meth.*, 24(5), 1307-1324.

SUMMARY

In chapter one we looked at the nature of structural change and defined structural change as a change in one or more parameters of the model in question. Bayesian procedures can be applied to solve inferential problems of structural change. Among the various methodological approaches within Bayesian inference, emphasis is put on the analysis of the posterior distribution itself, since the posterior distribution can be used for conducting hypothesis testing as well as obtaining a point estimate. The history of structural change in statistics, beginning in the early 1950's, is also discussed. Furthermore the Bayesian approach to hypothesis testing was developed by Jeffreys (1935, 1961), where the centerpiece was a number, now called the Bayes factor, which is the posterior odds of the null hypothesis when the prior probability on the null is one-half. According to Kass and Raftery (1993) this posterior odds = Bayes factor $\times$ prior odds and the Bayes factor is the ratio of the posterior odds of H_1 to its prior odds, regardless of the value of the prior odds. The intrinsic and fractional Bayes factors are defined and some advantages and disadvantages of the IBF's are discussed.

In chapter two changes in the multivariate normal model are considered. Assuming that a change has taken place, one will want to be able to detect the change and to estimate its position as well as the other parameters of the model. To do a Bayesian analysis, prior densities should be chosen. Firstly the hyperparameters are assumed known, but as this is not usually true, vague improper priors are used (while the number of change-points is fixed). Another way of dealing with the problem of unknown hyperparameters is to use a hierarchical model where the second stage priors are vague. We also considered Gibbs sampling and gave the full conditional distributions for all the cases. The three cases that are studied is

(1) a change in the mean with known or unknown variance,

(2) a change in the mean and variance by firstly using independent prior densities on the different variances and secondly assuming the variances to be proportional and

(3) a change in the variance.

The same models above are also considered when the number of change-points are unknown.

In this case vague priors are not appropriate when comparing models of different dimensions. In this case we revert to partial Bayes factors, specifically the intrinsic and fractional Bayes factors, to obtain the posterior probabilities of the number of change-points. Furthermore we look at component analysis, i.e. determining which components of a multivariate variable are mostly responsible for the changes in the parameters. The univariate case is then also considered in more detail, including multiple model comparisons and models with autocorrelated errors. A summary of approaches in the literature as well as four examples are included.

In chapter three changes in the linear model, with

(1) a change in the regression coefficient and a constant variance,

(2) a change in only the variance and

(3) a change in the regression coefficient and the variance, are considered. Bayes factors for the above mentioned cases, multiple change-points, component analysis, switch-point (continuous change-point) and autocorrelation are included, together with seven examples.

In chapter four changes in some other standard models are considered. Bernoulli type experiments include the Binomial model, the Negative binomial model, the Multinomial model and the Markov chain model. Exponential type models include the Poisson model, the Gamma model and the Exponential model. Special cases of the Exponential model include the left truncated exponential model and the Exponential model with epidemic change. In all cases the partial Bayes factor is used to obtain posterior probabilities when the number of change-points is unknown. Marginal posterior densities of all parameters under the change-point model are derived. Eleven examples are included.

In chapter five change-points in the hazard rate are studied. This includes an abrupt change in a constant hazard rate as well as a change from a decreasing hazard rate to a constant hazard rate or a change from a constant hazard rate to an increasing hazard rate. These hazard rates are obtained from combinations of Exponential and Weibull density functions. In the same way a bathtub hazard rate can also be constructed. Two illustrations are given.

Some concluding remarks are made in chapter six, with discussions of other approaches in the literature and other possible applications not dealt with in this study.

KEYWORDS: autocorrelation, Bayesian analysis, change-point, component analysis, Fractional Bayes Factor, Gibbs sampling, Intrinsic Bayes factor, linear model, multiple change-point, multivariate normal model, structural change, switchpoint.

OPSOMMING

In hoofstuk een het ons gekyk na die aard van strukturele verandering en definieer strukturele verandering as 'n verandering in een of meer parameters van die model. Bayes prosedures kan toegepas word om inferensiële probleme t.o.v. strukturele verandering op te los. Onder die verskillende benaderings binne Bayes inferensie is klem geplaas op die analise van die posterior verdeling, aangesien die posterior verdeling gebruik kan word vir die uitvoer van hipotese toetsing sowel as die verkryging van 'n puntbenadering. Die geskiedenis van strukturele verandering in statistiek, beginnende in die vroeë 1950's is ook bespreek. Die Bayes benadering tot hipotese toetsing is ontwikkel deur Jeffreys (1935, 1961), waarvan die hoofresultaat 'n getal was (nou na verwys as die Bayes faktor), wat die posterior kansverhouding van die nulhipotese is wanneer die prior waarskynlikheid op die nulhipotese een-half is. Volgens Kass en Raftery (1993) is hierdie posterior kansverhouding = Bayes faktor $\times$ prior kansverhouding en die Bayes faktor is die verhouding van die posterior kansverhouding van H_1 tot sy prior kansverhouding, ongeag van wat die waarde van die prior kansverhouding is. Die Intrinsieke en Fraksionele Bayes faktore is gedefinieer en sommige voordele en nadele van die Intrinsieke Bayes faktore is bespreek.

In hoofstuk twee is gekyk na die meerveranderlike normaalmodel. Met die veronderstelling dat 'n verandering plaasgevind het, wil mens in staat wees om die verandering op te spoor en om sy posisie te bepaal, sowel as die ander parameters van die model. Om 'n Bayes analise uit te voer, moet prior digthede gekies word. Eerstens is aanvaar dat die hiperparameters bekend is, maar aangesien dit normaalweg nie waar is nie, is vae onegte priors gebruik (terwyl die aantal breekpunte vas is). 'n Ander manier om die probleem van onbekende hiperparameters te hanteer, is om 'n hiërargiese model met vae tweede fase priors te gebruik. Gibbs steekproefneming is behandel en die volle voorwaardelike verdelings vir al die gevalle is gegee. Die drie gevalle is:

(1) 'n verandering in die gemiddeld met bekende en onbekende variansie,

(2) 'n verandering in die gemiddeld en variansie deur eers onafhanklike priordigthede op

die verksillende variansies te gebruik en verder ook te veronderstel dat die variansies proporsioneel is en

(3) 'n verandering in die variansie.

Dieselfde modelle is ook bestudeer as die aantal breekpunte onbekend is. In hierdie geval is vae priors nie toepaslik in die vergelyking van modelle met verskillende dimensies nie. Dus het ons ons gewend tot parsiële Bayes faktore, in besonder die Intrinsieke en Fraksionele Bayes faktore, om sodoende die posterior waarskynlikhede van die getal breekpunte te verkry. Verder is gekyk na komponentanalise om te bepaal watter komponent van die meerveranderlike variaat grootliks verantwoordelik is vir die verandering in die parameters. Die eenveranderlike geval is dan ook in meer besonderhede behandel, insluitende die vergelyking tussen 'n groep van moontlike modelle en modelle met outogekorreleerde foute. 'n Opsomming van benaderings in die literatuur, asook vier voorbeelde, is ingesluit.

In hoofstuk drie is veranderings in die lineêre model met

(1) 'n verandering in die regressie koeffisiënt en 'n konstante variansie,

(2) 'n verandering in slegs die variansie en

(3) 'n verandering in die regressie koeffisiënt en die variansie

behandel. Bayes faktore vir die genoemde gevalle, meervoudige breekpunte, komponentanalise, kontinue breekpunte en outokorrelasie, tesame met sewe voorbeelde, is ingesluit.

In hoofstuk vier is veranderings in ander standaard modelle behandel. Bernoulli tipe eksperimente sluit in die Binomiaalmodel, die Negatief Binomiaalmodel, die Multinomiale model en die Markov ketting model. Eksponensiële tipe modelle sluit in die Poisson model, die Gamma model en die Eksponensiële model. Spesiale gevalle van die Eksponensiële model sluit in die links afgeknotte eksponensiële model en die Eksponsensiële model met epidemiese verandering. In al die gevalle is parsiële Bayes faktore gebruik om posterior waarskynlikhede te verkry wanneer die aantal breekpunte onbekend is. Rand posterior digthede van al die parameters onder die breekpuntmodel is afgelei. Elf voorbeelde is ingesluit.

In hoofstuk vyf is breekpunte in die gevaarkoers bestudeer. Dit sluit in 'n skielike verandering in die konstante gevaarkoers, sowel as 'n verandering vanaf 'n afnemende gevaarkoers na 'n konstante gevaarkoers of 'n verandering vanaf 'n konstante gevaarkoers na 'n toenemende gevaarkoers. Hierdie gevaarkoerse is verkry vanuit kombinasies van Eksponensiële en Weibull digtheidsfunksies. 'n "Bathtub" gevaarkoers kan op dieselfde wyse gekonstrueer word. Twee illustrasies is gegee.

In hoofstuk ses is gevolgtrekkings gemaak, met 'n bespreking van ander benaderings in die literatuur asook ander moontlike toepassings wat nie in hierdie studie behandel is nie.