BAYESIAN INFERENCE FOR LINEAR AND NONLINEAR FUNCTIONS OF POISSON AND BINOMIAL RATES

by

LIZANNE RAUBENHEIMER

THESIS

Submitted in fulfillment of the requirements for the degree

PHILOSOPHIAE DOCTOR

IN

THE FACULTY OF NATURAL AND AGRICULTURAL SCIENCES

DEPARTMENT OF MATHEMATICAL STATISTICS AND ACTUARIAL SCIENCE

UNIVERSITY OF THE FREE STATE

PROMOTER: PROF. A. J. VAN DER MERWE

BLOEMFONTEIN
JANUARY 2012

"We [statisticians] will all be Bayesians in 2020, and then we can be a united profession."

Dennis Lindley in Smith (1995)

Acknowledgements

I would like to acknowledge and express my gratitude to the following people for their wonderful support and contributions to the making of this thesis:

- My promoter, Prof. Abrie van der Merwe, for his help and advice.
- A special thanks to Prof. Sarah Radloff for all the support, encouragement, numerous suggestions, advice and guidance.
- My family for their encouragement and understanding during the time it took to complete this thesis, especially my parents, Alwyn and Annelise – without their love and support this would never have been possible.
- Prof. Piet Groenewald for his contribution and help and thank you to Dr. Isabelle Garisch for being such an inspirational lecturer.

Last, but the most important, to my Heavenly Father all the credit, without God in my life, I would never be able to get this far.

Contents

List of Figures

List of Tables

List of Research Outputs

A list of research outputs related to this thesis is given below.

- Raubenheimer, L. & Van der Merwe, A. J. (2011a). Bayesian estimation of functions of binomial rates. *South African Statistical Journal*, 45(1), 41 - 64.
- Raubenheimer, L. & Van der Merwe, A. J. (2011b). Bayesian estimation of the ratio and product of two Poisson rates. *Proceedings of the 53rd annual conference of the South African Statistical Association*, 88 - 99.
- Raubenheimer, L. & Van der Merwe, A. J. (Accepted). Bayesian inference on nonlinear functions of Poisson rates. *South African Statistical Journal*, to appear in 2012.
- Raubenheimer, L. & Van der Merwe, A. J. (Submitted). Bayesian estimation of linear functions of Poisson parameters. *Communications in Statistics - Theory and Methods*, submitted in 2011.

List of Abbreviations

ARL	Average Run Length
DWI	Driving While Intoxicated
FAR	False Alarm Rate
HPD	Highest Posterior Density
LCL	Lower Control Limit
MCMC	Markov Chain Monte Carlo
MIR	Minimum Infection Rate
MLE	Maximum Likelihood Estimate
WMCM	Weighted Monte Carlo Method
WNV	West Nile Virus
SAW	Square-and-add Walter
SIR	Sampling Importance Re-sampling
UCL	Upper Control Limit

List of Distributions

- **Bernoulli distribution**

  Parameters: p where $0 < p < 1$

  Probability function: $p(x) = p^x (1-p)^{1-x}$, $x = 0, 1$

  Moments: $E(X) = p$, $Var(X) = p(1-p)$

- **Beta distribution**

  Parameters: α, β where $\alpha > 0$ and $\beta > 0$

  Probability density function: $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$

  Moments: $E(X) = \frac{\alpha}{\alpha+\beta}$, $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

  Notation: $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{1}{B(\alpha,\beta)}$

- **Beta-binomial distribution**

  Parameters: α, β where $\alpha > 0$ and $\beta > 0$

  Probability function: $p(x) = \left(\begin{array}{c} n \\ x \end{array} \right) \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(\alpha+\beta+n)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, $x = 0, 1, 2, \ldots, n$

  Moments: $E(X) = \frac{n\alpha}{\alpha+\beta}$, $Var(X) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$

- **Beta Prime distribution** (Beta distribution of the second kind)

  Parameters: α, β where $\alpha > 0$ and $\beta > 0$

  Probability density function: $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1+x)^{-\alpha-\beta}$, $0 < x < 1$

  Moments: $E(X) = \frac{\alpha}{\beta-1}$ if $\beta > 1$, $Var(X) = \frac{\alpha(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2}$ if $\beta > 2$

- **Binomial distribution**

  Parameters: n, p where n is a positive integer and $0 < p < 1$

  Probability function: $p(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}$, $x = 0, 1, 2, \ldots, n$

Moments: $E(X) = np, Var(X) = np(1-p)$

- **Gamma distribution**

  Parameters: α, β where $\alpha > 0$ and $\beta > 0$

  Probability density function: $f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0$

  Moments: $E(X) = \frac{\alpha}{\beta}, Var(X) = \frac{\alpha}{\beta^2}$

  Notation: $\Gamma(\alpha) = (\alpha-1)!$

- **Geometric distribution**

  Parameters: p where $0 < p < 1$ with $q = 1-p$

  Probability function: $p(x) = p(1-p)^{x-1}, x = 1, 2, \ldots$

  Moments: $E(X) = \frac{1}{p}, Var(X) = \frac{q}{p^2}$

- **Poisson distribution**

  Parameters: λ where $\lambda > 0$

  Probability function: $p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, x = 0, 1, 2, \ldots$

  Moments: $E(X) = \lambda, Var(X) = \lambda$

- **Poisson-gamma distribution** (Negative binomial distribution)

  Parameters: α, β where $\alpha > 0$ and $\beta > 0$

  Probability function: $p(x) = \frac{\Gamma(\alpha+x)}{\Gamma(\alpha)x!} \left(\frac{\beta}{\beta+1}\right)^{\alpha} \left(\frac{1}{\beta+1}\right)^{x}, x = 0, 1, 2, \ldots$

  Moments: $E(X) = \frac{\alpha}{\beta}, Var(X) = \frac{\alpha(\beta+1)}{\beta^2}$

- **Uniform distribution**

  Parameters: a, b where $a < b$

  Probability density function: $f(x) = \frac{1}{b-a}, a < x < b$

  Moments: $E(X) = 1/2(a+b), Var(X) = 1/12(b-a)^2$

Chapter 1

Introduction

1.1 Overview

Reverend Thomas Bayes (1702 - 1761) is known for having formulated the well-known Bayes' theorem. This work was published after his death. In 1763 the following paper "An Essay Towards Solving a Problem in the Doctrine of Chances" was published by the late Bayes (1763), communicated by Richard Price in a letter to John Canton. This paper indicated how to make statistical inferences that build upon earlier knowledge, and how to combine this earlier knowledge with current data in a way that updates the degree of belief. This "earlier knowledge" is called the "prior belief" and this "updated belief" is called the "posterior belief". This updating process is called Bayesian inference. The Bayes rule can be expressed as: posterior distribution $\propto$ likelihood function $\times$ prior distribution. When we specify Bayesian models, we have to decide on prior distributions for unknown parameters. As mentioned by Robert (2001), the most critical and most criticised point of Bayesian analysis deals with the choice of the prior distribution. Choosing the prior distribution is the key to Bayesian inference, but it is also a very difficult part. Gill (2008) mentions that while it is coy to say "everyone is a Bayesian, some of us know it," most researchers tell us about their prior knowledge even if it is not put directly in the form of a prior distribution. How do we choose a prior distribution? It depends on the information given to you and also the decision of being subjective or objective in the information that you would like to introduce to the problem.

One can use a conjugate prior, this is where the prior distribution and the posterior distribution both belong to the same family of distributions. Conjugate prior distributions are usually associated with a specific type of sampling distribution that always allows for their deviation. If no prior information is available, we can have an objective viewpoint when choosing the prior distribution. Such priors are known as noninformative priors, also known as objective, vague or flat priors. A noninformative prior which is often used, is the uniform prior, Thomas Bayes used a uniform prior on the binomial parameter. Some other noninformative priors are the Jeffreys prior, reference prior and the probability matching prior. Berger (1985) states the following: "We should indeed argue that noninformative prior

Bayesian analysis is the single most powerful method of statistical analysis."

1.2 Objectives

The main objectives of this thesis can be summarised as follows:

- to provide an overview of some noninformative priors;
- to derive the probability matching prior for the following cases: the product of different powers of binomial proportions, a linear combination of binomial proportions, the product of different powers of Poisson rates and linear functions of Poisson rates;
- to derive the reference prior in the case of the ratio of two Poisson rates;
- to compare the performance of the probability matching prior in the above mentioned cases to other noninformative priors and to classical (frequentist) methods;
- to show the properness of the probability matching posterior for the following cases: the product of different powers of binomial proportions and a linear combination of binomial proportions;
- to propose a Bayesian method for the estimation of binomial rates from pooled samples, and compare the results to classical (frequentist) methods;
- to propose Bayesian methods for the p - chart and the c - chart, and compare the Bayesian results to the results from the classical (frequentist) method;
- to investigate Bayesian tolerance intervals for the binomial and Poisson distributions.

1.3 Contributions

Given the objectives, we can summarise the contribution of the thesis in the field of objective Bayesian statistics as the following:

- the derivation of the probability matching prior for the product of different powers of k binomial proportions using the method by Datta & Ghosh (1995) and showing that the posterior distribution is proper;
- the derivation of the probability matching prior for a linear combination of binomial proportions using the method by Datta & Ghosh (1995) and showing that the posterior distribution is proper;
- the derivation of the probability matching prior for the product of different powers of k Poisson rates, this has been derived by Kim (2006), but Kim used the method by Tibshirani (1989) where we used the method by Datta & Ghosh (1995);

- the derivation of the reference prior for the ratio of two Poisson rates using the method by Berger & Bernardo (1992);
- the derivation of the probability matching prior for linear functions of Poisson rates using the method by Datta & Ghosh (1995);
- the application of an objective Bayesian method for the estimation of binomial rates from pooled samples;
- the application of objective Bayesian methods for the p - chart and the c - chart;
- the application of objective Bayesian methods for the construction of tolerance intervals for the binomial and Poisson distributions.

1.4 Thesis outline

The probability matching prior will be derived for several different cases. We will use the method imposed by Datta & Ghosh (1995) to derive the probability matching prior. Datta & Ghosh (1995) derived the differential equation which a prior must satisfy if the posterior probability of a one sided credibility interval for a parametric function and its frequentist probability agree up to $O\left(n^{-1}\right)$ where n is the sample size. They proved that the agreement between the posterior probability and the frequentist probability holds if and only if $\sum_{i=1}^{k} \frac{\partial}{\partial p_i}\left\{\eta_i\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = 0$, where $\pi\left(\underline{p}\right)$ is the probability matching prior for $\underline{p}$, the vector of unknown parameters. Let $\nabla_t\left(\underline{p}\right) = \left[\begin{array}{ccc} \frac{\partial}{\partial p_1} t\left(\underline{p}\right) & \cdots & \frac{\partial}{\partial p_k} t\left(\underline{p}\right) \end{array}\right]'$, then $\eta\left(\underline{p}\right) = \frac{F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right)}{\sqrt{\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right)}} = \left[\begin{array}{ccc} \eta_1\left(\underline{p}\right) & \cdots & \eta_k\left(\underline{p}\right) \end{array}\right]'$, where $F^{-1}\left(\underline{p}\right)$ is the inverse of $F\left(\underline{p}\right)$, the Fisher information matrix of $\underline{p}$ and $t\left(\underline{p}\right)$ is the parameter of interest. The above mentioned method is in the case where we deal with the binomial distribution. When we deal with the Poisson distribution, the method is exactly the same. The only difference in the notation will be to replace each $\underline{p}$ with $\underline{\lambda}$.

In **Chapter 2** we will look into Bayesian inference on the product of different powers of k binomial parameters. The parameter of interest is $\psi = \prod_{i=1}^{k} p_i^{a_i}$, and appears in applications to system reliability. We will derive the probability matching prior for this case by using the method by Datta & Ghosh (1995), and also evaluate the properness of the posterior distribution when the probability matching prior is used. In this chapter we will compare the performance of the probability matching prior, Jeffreys prior and uniform prior when $\psi_1 = p_1 p_2$, $\psi_3 = p_1^2 p_2$ and $\psi_4 = p_1 p_2^2$, for different values of p_1, p_2, n_1 and n_2. A comparison is also made between Bayesian and frequentist procedures using the observed values from Harris (1971) for $\psi_1 = p_1 p_2$ and $\psi_2 = p_1 p_2 p_3$. Another simulation study will be considered for $\psi_1 = p_1 p_2$, comparing the probability matching prior and the Jeffreys prior where

the binomial distribution is used and where the Poisson approximation to the binomial distribution is used.

In **Chapter 3** Bayesian interval estimation for linear functions of binomial parameters will be considered. The parameter of interest is $\theta = \sum_{i=1}^{k} a_i p_i$. We will derive the probability matching prior for this case by using the method by Datta & Ghosh (1995), and also evaluate the properness of the posterior distribution when the probability matching prior is used. In this chapter we will compare the performance of the probability matching prior, Jeffreys prior and uniform prior when $\theta = p_1 - p_2$, for different values of p_1, p_2, n_1 and n_2. The probability matching , Jeffreys and uniform priors will be compared to some well known classical methods by Roths & Tebbs (2006) for the cases where $n_1 = n_2 = 10$ and $n_1 = n_2 = 20$. The Jeffreys, uniform and probability matching priors will also be applied to a real problem to assess if male and female insects transmit the Mal de Rio Cuarto virus to susceptible maize plants at similar rates.

In **Chapter 4** our interest is to make Bayesian inferences on nonlinear functions of Poisson rates. Kim (2006) derived a noninformative (probability matching) prior for $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson rates, thereby obtaining approximate point estimates and Bayesian credibility intervals of the reliability of systems of k independent parallel components. We will derive the probability matching prior for this case by using the method by Datta & Ghosh (1995). Kim (2006) used the method by Tibshirani (1989) to derive the probability matching prior. Price & Bonett (2000) used noninformative priors for small and large values of λ_i $(i = 1, 2)$ to construct credibility intervals for $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates. Using the method of Berger & Bernardo (1992), the reference prior for the ratio of two Poisson rates is also obtained. Simulation studies will be done to compare the probability matching, Jeffreys and uniform priors in the cases where $\xi_1 = \prod_{i=1}^{k} \lambda_i$, $\xi_3 = \lambda_1^2 \lambda_2$ and $\xi_4 = \lambda_1^3 \lambda_2$. A simulation study will be done to compare the probability matching, Jeffreys and uniform priors and three other priors in the case where $\xi_2 = \lambda_1 \lambda_2$, this can be applied to the reliability of independent parallel components systems. In two sample situations it may be of interest to test or to construct confidence intervals for the ratio of two Poisson rates. A further simulation study will be done where we compare the uniform and Jeffreys (probability matching and reference) priors when $\nu = \lambda_1/\lambda_2$.

In **Chapter 5** our interest is to make Bayesian inferences on linear functions of Poisson rates, in general we can define such a linear contrast as $\delta = \sum_{i=1}^{k} a_i \lambda_i$, where a_i is the known coefficient value. Stamey & Hamilton (2006) considered four interval estimators for linear functions of Poisson rates, a Wald interval, a t interval with Satterthwaite's degrees of freedom and two Bayesian intervals using noninformative priors. We will consider another Bayesian interval using a probability matching prior. The probability matching prior will be derived by using the method proposed by Datta & Ghosh (1995). Krishnamoorthy & Thomson (2004) addressed the problem of hypothesis testing about two Poisson means. They compared the conditional test (C - test) to a test based on estimated p - values (E - test). We will use four different Bayesian methods, the Jeffreys prior, the probability matching prior,

a third prior which is proportional to $\lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and a fourth prior which is proportional to $\lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$ and compare these to their results.

In **Chapter 6** Bayesian estimation for binomial rates from pooled samples will be considered. The performance of Bayesian credibility intervals for the difference of two binomial proportions estimated from pooled samples will be investigated. These results will be compared to the results obtained by Biggerstaff (2008). Biggerstaff (2008) used asymptotic methods to derive Wald, profile score and profile likelihood ratio intervals.

Bayesian process control for the p - chart will be considered in **Chapter 7**. Control chart limits, average run lengths and false alarm rates will be determined, and the results for the proposed Bayesian method will be compared to the results obtained from the classical method. Chakraborti & Human (2006) examined the effects of parameter estimation for the p - chart using the classical method, our results will be compared to the results obtained by them.

In **Chapter 8** we discuss Bayesian process control for the c - chart. Control chart limits, average run lengths and false alarm rates will be determined, and the results for the proposed Bayesian method will be compared to the results obtained from the classical method. Chakraborti & Human (2008) studied the c - chart using the classical method, our results will be compared to the results obtained by them.

Bayesian tolerance intervals for the binomial and Poisson distributions will be studied in **Chapter 9**. The Jeffreys prior will be used for the Bayesian tolerance intervals.

Chapter 10 contains the conclusions and this chapter concludes by looking at possible shortcomings/ drawbacks of this thesis and at possible future research in this area.

The Appendices are found towards the end of this thesis. The Appendices contain additional theorems and proofs which are well known, additional simulation results and some data, and MATLAB® code used for simulation. All simulation studies have been done in MATLAB®, and all graphs were constructed in MATLAB®.

Appendix A contains the derivation of the inverse of the Fisher information matrix for k binomial rates and MATLAB® code for the simulation studies done in Chapter 2, Estimation for the Product of Binomial Rates.

Appendix B contains the derivation of the maximum likelihood estimate (MLE) for p_i, some additional simulation results and MATLAB® code for the simulation studies done in Chapter 3, Estimation for a Linear Function of Binomial Rates.

Appendix C contains the derivation of the inverse of the Fisher information matrix for k Poisson rates and MATLAB® code for the simulation studies done in Chapter 4, Estimation for the Ratio and Product of Poisson Rates.

Appendix D contains MATLAB® code for the simulation studies done in Chapter 5, Estimation for Linear Functions of Poisson Rates.

Appendix E contains the derivation of the inverse of the Fisher information matrix for M indepen-

dent binomial random variables from pooled samples, the data used in the example and MATLAB® code for the simulation studies done in Chapter 6, Estimation for Binomial Rates from Pooled Samples.

Appendix F contains MATLAB® code for the calculations and simulation studies done in Chapters 7 and 8, Bayesian Process Control for the p - chart and Bayesian Process Control for the c - chart, respectively. This appendix also contains the MATLAB® code for the simulation studies done in Chapter 9.

1.5 The Binomial and Poisson Distributions

The binomial distribution is an example of a discrete probability distribution, since the associated binomial random variable can take on only discrete values. We will start by defining Bernoulli trials and then show how we obtain a binomial random variable. Bernoulli trials are trials where: each trial has two possible outcomes, a success or a failure; the probability of success for each trial is the same, denoted by p and the probability of failure is denoted by $1-p$; the trials are independent. A binomial experiment is an experiment which consists of n independent Bernoulli trials. A binomial random variable, X, counts the number of successes in n trials of a binomial experiment, where the probability of success is p. The parameters of the binomial distribution are n and p, $X \sim Bin(n,p)$. The average (mean) number of successes in n trials is given by $\mu_X = E(X) = np$ and the variance for the number of successes is given by $\sigma_X^2 = Var(X) = np(1-p)$. The probability function for the binomial distribution is given by: $P(X=x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}$ for $x = 0,1,2\ldots,n$. Figure 1.1 shows binomial distribution bar graphs.

The Poisson distribution is another discrete probability distribution. It is appropriate when the probability of an event occurring is very small. It is a useful model for the number of events per unit time, or area, or volume. A Poisson random variable, X, has a Poisson distribution if it counts the number of successes per unit time, area, distance, etc. The probability of success must be the same for each unit of time, area, distance, etc. This probability is usually fairly small. The number of successes in each unit of time, area, distance, etc., is independent of the number that occur in any other unit. The Poisson distribution has a single parameter, λ, $X \sim P(\lambda)$. The mean and the variance of a Poisson distribution are given by $\mu_X = E(X) = \lambda$ and $\sigma_X^2 = Var(X) = \lambda$, respectively. The probability function of the Poisson distribution is given by: $P(X=x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0,1,2,\ldots$. Figure 1.2 shows Poisson distribution bar graphs.

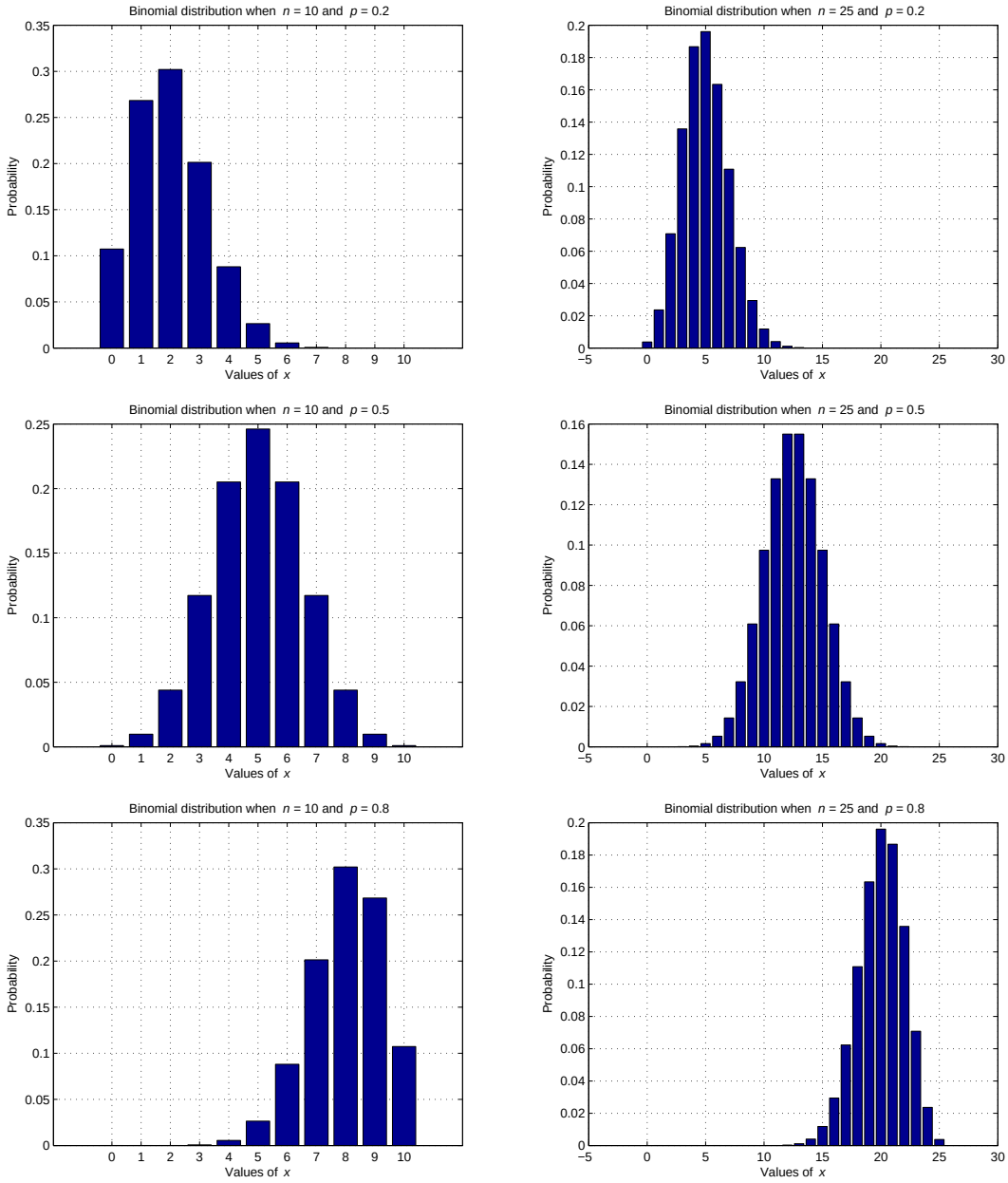


Figure 1.1: Binomial distribution bar graphs.

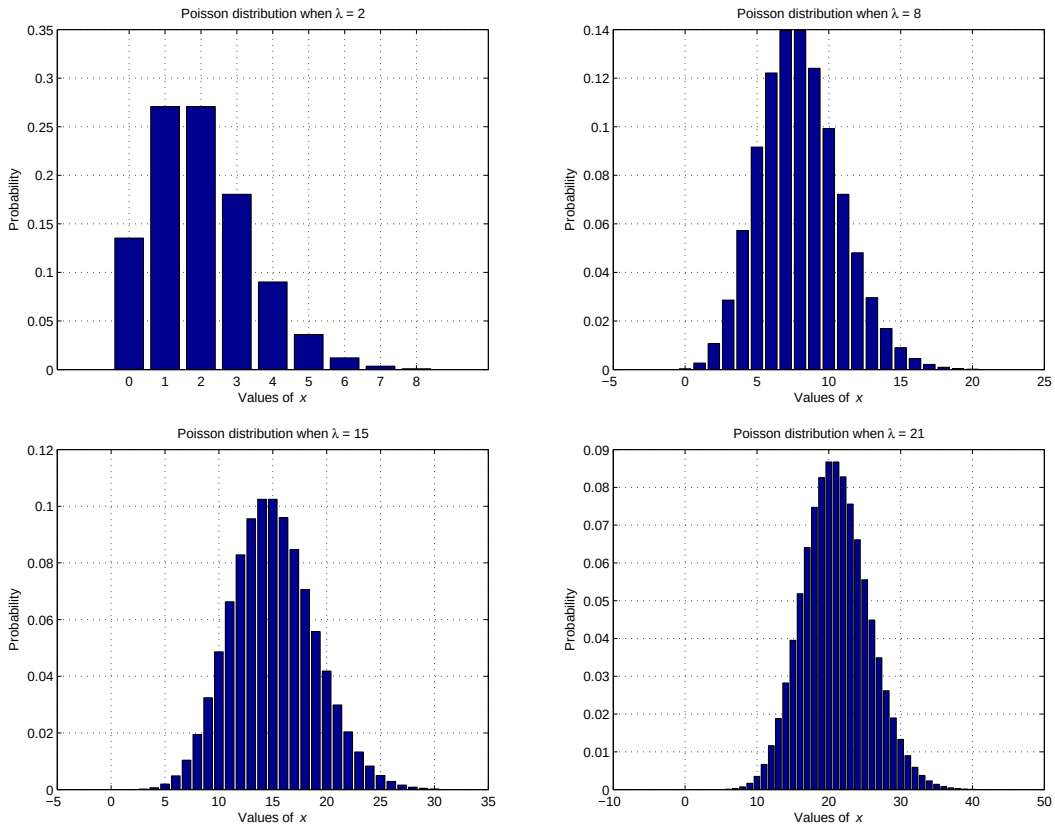


Figure 1.2: Poisson distribution bar graphs.

1.5.1 Functions of Binomial Proportions

Assume that $X_1, X_2, \ldots, X_k$ are independent binomial random variables with $X_i \sim Bin(n_i, p_i)$ for $i = 1, 2, \ldots, k$. Therefore $P(X_i = x_i) = \begin{pmatrix} n_i \\ x_i \end{pmatrix} p_i^{x_i} (1-p_i)^{n_i - x_i}$ for $x_i = 0, 1, \ldots, n_i$.

The likelihood function is given by

$$\begin{aligned} L(p_1, p_2 \ldots, p_k \,|x_1, x_2 \ldots, x_k) &= L\left(\underline{p} \,|x_1, x_2 \ldots, x_k\right) \\ &= \prod_{i=1}^{k} \begin{pmatrix} n_i \\ x_i \end{pmatrix} p_i^{x_i} (1-p_i)^{n_i - x_i}. \end{aligned}$$

- **Product of Binomial Proportions**

The parameter $\psi = \prod_{i=1}^{k} p_i^{a_i}$, the product of different powers of k binomial parameters, appears in applications to system reliability. If a system consists of k components in parallel, then the probability of system failure is $\psi = \prod_{i=1}^{k} p_i$ where p_i is the probability that the i^{th} component will fail. Also if a system requires that at least one of each of k types of components must be employed and that these components are needed in parallel, then the probability of failure of an $m-$component system is

$\psi = \prod_{i=1}^{k} p_i^{a_i}$, where $k < m$, a_i is the number of components of type i and $\sum_{i=1}^{k} a_i = m$. The product of binomial proportions can thus be used to estimate the reliability of a parallel system.

- **Linear Function of Binomial Proportions**

The parameter of interest in this case is, $\theta = \sum_{i=1}^{k} a_i p_i$ a linear combination of binomial proportions. Due to their important practical value, linear functions of binomial proportions have received some attention recently (Price & Bonett, 2004; Tebbs & Roths, 2008). The difference between two binomial proportions can appear in applications in epidemiology and medical research, just to mention two areas. For example, if one wants to compare the proportions of households without health insurance in two countries.

1.5.2 Functions of Poisson Parameters

Consider a sample from k Poisson populations. Let X_i be an observation from population i. Then $X_1, X_2, \ldots, X_k$ will be independent Poisson distributions such that $X_i \sim P(\lambda_i)$, for $i = 1, 2, \ldots, k$, where λ_i is the expected number of events per unit sample. Therefore $P(X_i = x_i) = \dfrac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}$ for $x_i = 0, 1, 2, \ldots$.

The likelihood function is given by

$$\begin{aligned} L(\lambda_1, \lambda_2 \ldots, \lambda_k \,|x_1, x_2 \ldots, x_k) &= L(\underline{\lambda} \,|x_1, x_2 \ldots, x_k) \\ &= \prod_{i=1}^{k} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}. \end{aligned}$$

- **Product and Ratio of Poisson Parameters**

The parameter of interest in this case will be $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson rates. This appears in the reliability of systems of k independent parallel components. The parameter $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson parameters appears in applications to system reliability. If a system consists of k components in parallel, then the probability of system failure is $\psi = \prod_{i=1}^{k} \left(\frac{\lambda_i}{n_i}\right)^{a_i}$ where $p_i = \frac{\lambda_i}{n_i}$ is the probability that the i^{th} component will fail. Also if a system requires that at least one of each of k types of components must be employed and that these components are needed in parallel, then the probability of failure of an m- component system is $\psi = \prod_{i=1}^{k} \left(\frac{\lambda_i}{n_i}\right)^{a_i}$, where $k < m$, a_i is the number of components of type i and $\sum_{i=1}^{k} a_i = m$. Another function that will be used is $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates. The ratio of two Poisson means can be used to compare incident rates of a disease in a control group and a treatment group, where the product of Poisson parameters can be used to estimate the reliability of a parallel system.

- **Linear Contrast of Poisson Parameter**

In this case the interest is in a linear combination of Poisson rates. In general we can define a linear contrast as $\delta = \sum_{i=1}^{k} a_i \lambda_i$, where a_i is the known coefficient value. In the case of a linear contrast of

Poisson parameters $\sum_{i=1}^{k} a_i = 0$. When $\sum_{i=1}^{k} a_i \neq 0$, we will consider the average of the Poisson rates. A linear combination of Poisson parameters can be used to estimate the number of fatal vehicle accidents driving while under the influence of alcohol on the different public holidays, and also to see if less (or more) such accidents occur during the summer public holidays than during the winter public holidays.

1.6 Bayesian Methods

As mentioned in Section 1.1, the Bayes rule can be expressed as: posterior distribution $\propto$ likelihood function $\times$ prior distribution. When we specify Bayesian models, we have to decide on prior distributions for unknown parameters. As mentioned by Robert (2001), the most critical and most criticised point of Bayesian analysis deals with the choice of the prior distribution. Choosing the prior distribution is the key to Bayesian inference, but it is also a very difficult part. We will investigate a number of noninformative priors. A noninformative prior is used when little or no prior information is available. Noninformative priors are often improper, which means that the prior is not integrable. Some improper priors are integrable, but not integrable to one, these priors are also regarded as improper since it will be equal to a constant. This is not problematic, as long as the posterior distribution results into a proper distribution.

1.6.1 The Probability Matching Prior

A probability matching prior is a prior distribution under which the posterior probabilities match their coverage probabilities. The fact that the resulting Bayesian posterior intervals of level $1-\alpha$ are also good frequentist confidence intervals at the same level is a very desirable situation. As mentioned, we will use the method by Datta & Ghosh (1995) to derive the probability matching prior in these cases. Datta & Ghosh (1995) derived the differential equation which a prior must satisfy if the posterior probability of a one sided credibility interval for a parametric function and its frequentist probability agree up to $O\left(n^{-1}\right)$ where n is the sample size. In this thesis the probability matching prior will be denoted by π_{PM}.

1.6.2 The Jeffreys Prior

Jeffreys (1939) argued that if there is no prior information about the unknown parameter, then there is also no information about any one-to-one transformation of the parameter, and therefore the rule for determining a prior should give a similar result if it is applied to the transformed parameter. The Jeffreys prior is proportional to the square root of the determinant of the Fisher information matrix and is given by

$$\pi_J \propto |F|^{\frac{1}{2}}.$$

1.6.3 The Uniform Prior

When using a uniform prior, one assigns a prior distribution to the unknown parameter on the interval (a, b) using the uniform distribution. Bayes himself used a uniform prior on the binomial parameter. In general the uniform prior is denoted as

$$\pi_U \propto \text{constant.}$$

1.6.4 The Reference Prior

The reference prior was introduced by Bernardo (1979) and Berger & Bernardo (1992). As mentioned by Pearn & Wu (2005) the reference prior maximises the difference in information about the parameter provided by the prior and posterior. The reference prior is derived in such a way that it provides as little information as possible about the parameter. As in the case of the Jeffreys prior, the reference prior method is derived from the Fisher information matrix. In this thesis the reference prior will be denoted by π_R.

Chapter 2

Estimation for the Product of Binomial Rates

2.1 Introduction

In this chapter the probability matching prior for the product of k binomial parameters will be derived. In the case of two independently distributed binomial random variables, the Jeffreys, uniform and probability matching priors for the product of the parameters are compared. This research is an extension of the work by Kim (2006), who derived the probability matching prior for the product of k independent Poisson rates.

Assume that $X_1, X_2, \ldots, X_k$ are independent binomial random variables with $X_i \sim Bin(n_i, p_i)$ for $i = 1, 2, \ldots, k$, where the parameter of interest is $\psi = \prod_{i=1}^{k} p_i^{a_i}$, $a_i \in (-\infty, \infty)$. The parameter $\psi = \prod_{i=1}^{k} p_i^{a_i}$, the product of different powers of k binomial parameters, appears in applications to system reliability. If a system consists of k components in parallel, then the probability of system failure is $\psi = \prod_{i=1}^{k} p_i$ where p_i is the probability that the i^{th} component will fail. Also if a system requires that at least one of each of k types of components must be employed and that these components are needed in parallel, then the probability of failure of an $m-$component system is $\psi = \prod_{i=1}^{k} p_i^{a_i}$, where $k < m$, a_i is the number of components of type i and $\sum_{i=1}^{k} a_i = m$. The probability of system failure is also studied in cases where at least one of two types of components are required to be employed and where three components in parallel are needed. The weighted Monte Carlo method is used for the simulation from the posterior distribution in the case of the probability matching prior.

From a Bayesian perspective a prior is needed for the parameter ψ. Common noninformative priors in multiparameter problems such as Jeffreys priors can have features that have an unexpectedly dramatic effect on the posterior distribution. It is for this reason that the probability matching prior for ψ will be derived in Theorem 2.1.

Datta & Ghosh (1995) derived the differential equation which a prior must satisfy if the posterior probability of a one sided credibility interval for a parametric function and its frequentist probability agree up to $O\left(n^{-1}\right)$ where n is the sample size. They proved that the agreement between the posterior probability and the frequentist probability holds if and only if $\sum_{i=1}^{k} \frac{\partial}{\partial p_i} \left\{ \eta_i \left(\underline{p}\right) \pi \left(\underline{p}\right) \right\} = 0$, where

$\pi\left(\underline{p}\right)$ is the probability matching prior for $\underline{p}$, the vector of unknown parameters.
Let $\nabla_t\left(\underline{p}\right) = \left[\begin{array}{ccc} \frac{\partial}{\partial p_1} t\left(\underline{p}\right) & \cdots & \frac{\partial}{\partial p_k} t\left(\underline{p}\right) \end{array}\right]'$, then $\eta\left(\underline{p}\right) = \frac{F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right)}{\sqrt{\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right)}} = \left[\begin{array}{ccc} \eta_1\left(\underline{p}\right) & \cdots & \eta_k\left(\underline{p}\right) \end{array}\right]'$, where $F\left(\underline{p}\right)$ is the Fisher information matrix of $\underline{p}$ and $F^{-1}\left(\underline{p}\right)$ is the inverse of the Fisher information matrix. Reasons for using the probability matching prior is that it provides a method of constructing accurate frequentist intervals and it could also be useful for comparative purposes in Bayesian analysis. From Wolpert (2004), Berger states that frequentist reasoning will play an important role in finally obtaining good general priors for estimation and prediction. Some statisticians argue that frequency calculations are an important part of applied Bayesian statistics (see Rubin, 1984). Rubin (1984) states that the applied Bayesian statistician's tool-kit should be more extensive and include tools that may be usefully labeled frequency calculations. The applied statistician should be Bayesian in principle and calibrated to the real world in practice - appropriate frequency calculations help to define such a tie (Rubin, 1984).

2.2 Probability Matching Prior for the Product of Different Powers of k Binomial Parameters

A probability matching prior is a prior distribution under which the posterior probabilities match their coverage probabilities. The fact that the resulting Bayesian posterior intervals of level $1-\alpha$ are also good frequentist confidence intervals at the same level is a very desirable situation. See also Severini et al. (2002) and Bayarri & Berger (2004) for general discussion. By using the method of Datta & Ghosh (1995) the following theorem is proved.

Theorem 2.1. The probability matching prior for $\psi = \prod_{i=1}^{k} p_i^{a_i}$, the product of different powers of k binomial parameters, is given by

$$\pi_{PM}\left(\underline{p}\right) \quad = \quad \pi_{PM}\left(p_1, p_2, \ldots, p_k\right) \propto \left\{\sum_{i=1}^{k} \frac{a_i^2\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}} \prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}. \tag{2.1}$$

Proof. Assume that $X_1, X_2, \ldots, X_k$ are independent binomial random variables with $X_i \sim Bin\left(n_i, p_i\right)$ for $i = 1, 2, \ldots, k$.

Therefore $P\left(X_i = x_i\right) = \left(\begin{array}{c} n_i \\ x_i \end{array}\right) p_i^{x_i}\left(1-p_i\right)^{n_i - x_i}$ for $x_i = 0, 1, \ldots n_i$.

The likelihood function is given by

$$\begin{aligned} L(p_1, p_2 \ldots, p_k | x_1, x_2 \ldots, x_k) &= L\left(\underline{p} | x_1, x_2 \ldots, x_k\right) \\ &= \prod_{i=1}^{k} \left(\begin{array}{c} n_i \\ x_i \end{array} \right) p_i^{x_i} (1-p_i)^{n_i - x_i}. \end{aligned}$$

The derivation of the inverse of the Fisher information matrix is given in Appendix A in Theorem A.1. The inverse of the Fisher information matrix is given by

$$F^{-1}\left(\underline{p}\right) = \left[\begin{array}{ccc} p_1(1-p_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & p_k(1-p_k) \end{array} \right].$$

We are interested in a probability matching prior for $t\left(\underline{p}\right) = \psi = \prod_{i=1}^{k} p_i^{a_i}$, the product of different powers of k binomial parameters.

Now

$$\begin{aligned} \nabla_t'\left(\underline{p}\right) &= \left[\begin{array}{cccc} \frac{\partial t(\underline{p})}{\partial p_1} & \frac{\partial t(\underline{p})}{\partial p_2} & \cdots & \frac{\partial t(\underline{p})}{\partial p_k} \end{array} \right] \\ &= \left[\begin{array}{cccc} a_1 p_1^{a_1 - 1} \prod_{i \neq 1}^{k} p_i^{a_i} & a_2 p_2^{a_2 - 1} \prod_{i \neq 2}^{k} p_i^{a_i} & \cdots & a_k p_k^{a_k - 1} \prod_{i \neq k}^{k} p_i^{a_i} \end{array} \right] \\ &= \left[\begin{array}{cccc} \frac{a_1}{p_1} \prod_{i=1}^{k} p_i^{a_i} & \frac{a_2}{p_2} \prod_{i=1}^{k} p_i^{a_i} & \cdots & \frac{a_k}{p_k} \prod_{i=1}^{k} p_i^{a_i} \end{array} \right] \\ &= \left[\begin{array}{cccc} \frac{a_1}{p_1} & \frac{a_2}{p_2} & \cdots & \frac{a_k}{p_k} \end{array} \right] \prod_{i=1}^{k} p_i^{a_i}. \end{aligned}$$

Also

$$\begin{aligned} \nabla_t'\left(\underline{p}\right) F^{-1}\left(\underline{p}\right) &= \left[\begin{array}{cccc} \frac{a_1}{p_1} \prod_{i=1}^{k} p_i^{a_i} & \frac{a_2}{p_2} \prod_{i=1}^{k} p_i^{a_i} & \cdots & \frac{a_k}{p_k} \prod_{i=1}^{k} p_i^{a_i} \end{array} \right] \\ &\quad \times \left[\begin{array}{ccc} p_1(1-p_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & p_k(1-p_k) \end{array} \right] \\ &= \left[\begin{array}{cccc} a_1(1-p_1) \prod_{i=1}^{k} p_i^{a_i} & a_2(1-p_2) \prod_{i=1}^{k} p_i^{a_i} & \cdots & a_k(1-p_k) \prod_{i=1}^{k} p_i^{a_i} \end{array} \right] \\ &= \left[\begin{array}{cccc} a_1(1-p_1) & a_2(1-p_2) & \cdots & a_k(1-p_k) \end{array} \right] \prod_{i=1}^{k} p_i^{a_i} \end{aligned}$$

and

$$\begin{aligned}\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right) &= \left[\begin{array}{cccc} a_1(1-p_1)\prod_{i=1}^{k}p_i^{a_i} & a_2(1-p_2)\prod_{i=1}^{k}p_i^{a_i} & \cdots & a_k(1-p_k)\prod_{i=1}^{k}p_i^{a_i}\end{array}\right] \\ &\quad \times \left[\begin{array}{c} \frac{a_1}{p_1}\prod_{i=1}^{k}p_i^{a_i} \\ \frac{a_2}{p_2}\prod_{i=1}^{k}p_i^{a_i} \\ \vdots \\ \frac{a_k}{p_k}\prod_{i=1}^{k}p_i^{a_i}\end{array}\right] \\ &= \left(\prod_{i=1}^{k}p_i^{a_i}\right)^2\sum_{i=1}^{k}\frac{a_i^2(1-p_i)}{p_i}.\end{aligned}$$

Define

$$\begin{aligned}\eta'\left(\underline{p}\right) &= \frac{\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)}{\sqrt{\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right)}} \\ &= \left[\begin{array}{cccc} \frac{a_1(1-p_1)}{\sqrt{\sum_{i=1}^{k}\frac{a_i^2(1-p_i)}{p_i}}} & \frac{a_2(1-p_2)}{\sqrt{\sum_{i=1}^{k}\frac{a_i^2(1-p_i)}{p_i}}} & \cdots & \frac{a_k(1-p_k)}{\sqrt{\sum_{i=1}^{k}\frac{a_i^2(1-p_i)}{p_i}}}\end{array}\right] \\ &= \left[\begin{array}{cccc}\eta_1\left(\underline{p}\right) & \eta_2\left(\underline{p}\right) & \cdots & \eta_k\left(\underline{p}\right)\end{array}\right].\end{aligned}$$

The prior $\pi\left(\underline{p}\right)$ is a probability matching prior if and only if the differential equation $\sum_{i=1}^{k}\frac{\partial}{\partial p_i}\left\{\eta_i\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\}=0$ is satisfied.

Let

$$\pi\left(\underline{p}\right) = \left\{\sum_{i=1}^{k}\frac{a_i^2(1-p_i)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k}\left(a_i(1-p_i)\right)^{-1}$$

then

$$\begin{aligned}\eta_1\left(\underline{p}\right)\pi\left(\underline{p}\right) &= a_1(1-p_1)\prod_{i=1}^{k}\left(a_i(1-p_i)\right)^{-1} \\ &= \prod_{i\neq 1}^{k}\left(a_i(1-p_i)\right)^{-1}\end{aligned}$$

therefore

$$\frac{\partial}{\partial p_1}\left\{\eta_1\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = \frac{\partial}{\partial p_1}\left\{\prod_{i\neq 1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\right\} = 0$$

and

$$\begin{aligned}\eta_2\left(\underline{p}\right)\pi\left(\underline{p}\right) &= a_2\left(1-p_2\right)\prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1} \\ &= \prod_{i\neq 2}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\end{aligned}$$

therefore

$$\frac{\partial}{\partial p_2}\left\{\eta_2\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = \frac{\partial}{\partial p_2}\left\{\prod_{i\neq 2}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\right\} = 0$$

and

$$\begin{aligned}\eta_k\left(\underline{p}\right)\pi\left(\underline{p}\right) &= a_k\left(1-p_k\right)\prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1} \\ &= \prod_{i\neq k}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\end{aligned}$$

therefore

$$\frac{\partial}{\partial p_k}\left\{\eta_k\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = \frac{\partial}{\partial p_k}\left\{\prod_{i\neq k}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\right\} = 0.$$

We can therefore conclude that

$$\sum_{i=1}^{k}\frac{\partial}{\partial p_i}\left\{\eta_i\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = 0.$$

The differential equation will be satisfied if $\pi\left(\underline{p}\right)$ is

$$\pi_{PM}\left(\underline{p}\right) \propto \left\{\sum_{i=1}^{k}\frac{a_i^2\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1} \qquad \text{for } 0\leq p_i\leq 1.$$

□

The joint posterior distribution when using the probability matching prior is given by

$$\begin{aligned}
\pi_{PM}\left(\underline{p}|data\right) &\propto \pi_{PM}\left(\underline{p}\right)\times L\left(\underline{p}|data\right)\\
&\propto \left\{\sum_{i=1}^{k}\frac{a_i^2\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\times\prod_{i=1}^{k}\left(\begin{array}{c}n_i\\x_i\end{array}\right)p_i^{x_i}\left(1-p_i\right)^{n_i-x_i}\\
\therefore \pi_{PM}\left(\underline{p}|data\right) &\propto \left\{\sum_{i=1}^{k}\frac{a_i^2\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k}a_i^{-1}p_i^{x_i}\left(1-p_i\right)^{n_i-x_i-1} \qquad \text{for } 0\leq p_i\leq 1. \qquad (2.2)
\end{aligned}$$

When $a_i = 1$, the probability matching prior for $\psi = \prod_{i=1}^{k} p_i$, will be

$$\pi_{PM}\left(\underline{p}\right) \propto \left\{\sum_{i=1}^{k}\frac{\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k}\left(1-p_i\right)^{-1}. \qquad (2.3)$$

When $a_i = 1$, for $i = 1,2,\ldots,k$, the posterior distribution in the case of the probability matching prior is given by

$$\pi_{PM}\left(\underline{p}|data\right) \propto \left\{\sum_{i=1}^{k}\frac{\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k}p_i^{x_i}\left(1-p_i\right)^{n_i-x_i-1} \qquad \text{for } 0\leq p_i\leq 1. \qquad (2.4)$$

When $a_i = 1$ and $k = 2$ the probability matching prior for $\psi = \prod_{i=1}^{2} p_i$, will be

$$\begin{aligned}
\pi_{PM}\left(\underline{p}\right) &\propto \left\{\sum_{i=1}^{2}\frac{\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{2}\left(1-p_i\right)^{-1}\\
&= \left[\frac{\left(1-p_1\right)}{p_1}+\frac{\left(1-p_2\right)}{p_2}\right]^{\frac{1}{2}}\left(1-p_1\right)^{-1}\left(1-p_2\right)^{-1} \qquad (2.5)
\end{aligned}$$

When $a_i = 1$, for $i = 1,2$, the posterior distribution in the case of the probability matching prior is given by

$$\pi_{PM}\left(\underline{p}|data\right) \propto \left\{\sum_{i=1}^{2}\frac{\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{2}p_i^{x_i}\left(1-p_i\right)^{n_i-x_i-1} \qquad \text{for } 0\leq p_i\leq 1. \qquad (2.6)$$

Theorem 2.2. $\pi_{PM}\left(\underline{p}|data\right)$ is a proper posterior distribution if $x_i < n_i$, for $i = 1,2,\ldots,k$.

Proof.

$$\begin{aligned}\sum_{i=1}^{k}\frac{1-p_i}{p_i} &= \frac{\left(\frac{1-p_1}{p_1}\right)\prod_{i=1}^{k}p_i+\left(\frac{1-p_2}{p_2}\right)\prod_{i=1}^{k}p_i+\cdots+\left(\frac{1-p_k}{p_k}\right)\prod_{i=1}^{k}p_i}{\prod_{i=1}^{k}p_i}\\ &= \frac{\left(\frac{\prod_{i=1}^{k}p_i}{p_1}-\prod_{i=1}^{k}p_i\right)+\left(\frac{\prod_{i=1}^{k}p_i}{p_2}-\prod_{i=1}^{k}p_i\right)+\cdots+\left(\frac{\prod_{i=1}^{k}p_i}{p_k}-\prod_{i=1}^{k}p_i\right)}{\prod_{i=1}^{k}p_i}\end{aligned}$$

$$\begin{aligned}\therefore\sum_{i=1}^{k}\frac{1-p_i}{p_i} &= \frac{\sum_{i=1}^{k}\left(\frac{\prod_{i=1}^{k}p_i}{p_i}-\prod_{i=1}^{k}p_i\right)}{\prod_{i=1}^{k}p_i}\\ &= \frac{\sum_{i=1}^{k}\left(\frac{\prod_{i=1}^{k}p_i}{p_i}\right)}{\prod_{i=1}^{k}p_i}-\frac{\sum_{i=1}^{k}\left(\prod_{i=1}^{k}p_i\right)}{\prod_{i=1}^{k}p_i}\\ &= \frac{\sum_{i=1}^{k}\left(\frac{\prod_{i=1}^{k}p_i}{p_i}\right)}{\prod_{i=1}^{k}p_i}-\frac{k\left(\prod_{i=1}^{k}p_i\right)}{\prod_{i=1}^{k}p_i}\\ &= \frac{\sum_{i=1}^{k}\left(\frac{\prod_{i=1}^{k}p_i}{p_i}\right)}{\prod_{i=1}^{k}p_i}-k.\end{aligned}$$

We can therefore conclude that

$$\frac{\sum_{i=1}^{k}\left(\frac{\prod_{i=1}^{k} p_i}{p_i}\right)}{\prod_{i=1}^{k} p_i} - k \;<\; \frac{\sum_{i=1}^{k}\frac{\prod_{i=1}^{k} p_i}{p_i}}{\prod_{i=1}^{k} p_i}$$

since k is positive. We can thus conclude that

$$\sum_{i=1}^{k}\frac{1-p_i}{p_i} \;<\; \frac{\sum_{i=1}^{k}\frac{\prod_{i=1}^{k} p_i}{p_i}}{\prod_{i=1}^{k} p_i} < \frac{k}{\prod_{i=1}^{k} p_i}.$$

Therefore

$$\left\{\sum_{i=1}^{k}\frac{(1-p_i)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k} p_i^{x_i}(1-p_i)^{n_i-x_i-1} \;<\; \left\{\frac{k}{\prod_{i=1}^{k} p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k} p_i^{x_i}(1-p_i)^{n_i-x_i-1}$$

$$\therefore \left\{\sum_{i=1}^{k}\frac{(1-p_i)}{p_i}\right\}^{\frac{1}{2}}\prod_{i=1}^{k} p_i^{x_i}(1-p_i)^{n_i-x_i-1} \;<\; k^{\frac{1}{2}}\prod_{i=1}^{k} p_i^{x_i-\frac{1}{2}}(1-p_i)^{n_i-x_i-1}$$

and each

$$\int_0^1 p_i^{x_i-\frac{1}{2}}(1-p_i)^{n_i-x_i-1}\,dp_i \;=\; Beta\left(x_i+\frac{1}{2}, n_i-x_i\right)$$

converges if $x_i < n_i$ for $i = 1,\ldots,k$. Therefore $\pi_{PM}\left(\underline{p}\,|data\right)$ is a proper posterior distribution if $x_i < n_i$, for $i = 1,2,\ldots,k$. □

2.3 The Jeffreys and Uniform Priors for the Product of k Binomial Parameters

The Jeffreys prior, on the other hand, is proportional to the square root of the determinant of the Fisher information matrix and is given by

$$\begin{aligned} \pi_J\left(\underline{p}\right) &\propto \left|F\left(\underline{p}\right)\right|^{\frac{1}{2}} \\ &= \left(\prod_{i=1}^{k} \frac{1}{p_i\left(1-p_i\right)}\right)^{\frac{1}{2}}. \end{aligned} \tag{2.7}$$

The joint posterior distribution when using the Jeffreys prior is given by

$$\begin{aligned} \pi_J\left(\underline{p}|data\right) &\propto \pi_J\left(\underline{p}\right) \times L\left(\underline{p}|data\right) \\ &\propto \left(\prod_{i=1}^{k} \frac{1}{p_i\left(1-p_i\right)}\right)^{\frac{1}{2}} \times \prod_{i=1}^{k} \left(\begin{array}{c} n_i \\ x_i \end{array}\right) p_i^{x_i}\left(1-p_i\right)^{n_i-x_i} \\ \therefore \pi_J\left(\underline{p}|data\right) &\propto \prod_{i=1}^{k} p_i^{x_i-\frac{1}{2}}\left(1-p_i\right)^{n_i-x_i-\frac{1}{2}} \qquad \text{for } 0 \leq p_i \leq 1. \end{aligned} \tag{2.8}$$

The posterior distribution of $\underline{p}$ is thus the product of k independently distributed $Beta\left(x_i+\frac{1}{2}, n_i-x_i+\frac{1}{2}\right)$ variates.

The reference prior for the bivariate binomial has been derived, from Yang & Berger (1997) the reference prior for $\psi = \prod_{i=1}^{2} p_i$ is given by

$$\pi_R\left(\underline{p}\right) \propto \left(1-p_1\right)^{-\frac{1}{2}} p_2^{\frac{1}{2}}\left(1-p_2\right)^{-\frac{1}{2}}\left(1-p_1p_2\right)^{-\frac{1}{2}}. \tag{2.9}$$

From Equation 2.5 the probability matching prior for this case was given by

$$\pi_{PM}\left(\underline{p}\right) \propto \left[\frac{\left(1-p_1\right)}{p_1} + \frac{\left(1-p_2\right)}{p_2}\right]^{\frac{1}{2}}\left(1-p_1\right)^{-1}\left(1-p_2\right)^{-1}. \tag{2.10}$$

From Equations 2.9 and 2.10 it can be seen that the probability matching prior is not the same as the reference prior when considering the product of two binomial proportions.

Theorem 2.3. If $k = 1$ and $a_i = 1$, the probability matching prior in the case of the product of different powers of k binomial parameters becomes the Jeffreys prior.

Proof. From Equation 2.1 the probability matching prior for $\psi = \prod_{i=1}^{k} p_i^{a_i}$, is given by

$$\pi_{PM}\left(\underline{p}\right) \propto \left\{\sum_{i=1}^{k} \frac{a_i^2\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}} \prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}.$$

Let $k = 1$, $p_i = p$ and $a_i = 1$, then the probability matching prior simplifies to

$$\begin{aligned}\pi_{PM}(p) &\propto \frac{(1-p)^{\frac{1}{2}}}{p^{\frac{1}{2}}}(1-p)^{-1} \\ &= \frac{1}{(1-p)^{\frac{1}{2}}\, p^{\frac{1}{2}}} \\ &\propto \pi_J(p).\end{aligned}$$

□

Corollary 2.1. From Yang & Berger (1997) it can be seen that the Jeffreys prior is the same as the reference prior for the one-dimensional case. We can therefore, from Theorem 2.3, conclude that the Jeffreys prior, probability matching prior and the reference prior are all the same for the one-dimensional case, i.e. when we have a single p.

The uniform prior is proportional to a constant and is given by

$$\pi_U\left(\underline{p}\right) \propto \text{constant.} \tag{2.11}$$

The joint posterior is

$$\pi_U\left(\underline{p}\,|data\right) \propto \prod_{i=1}^{k} p_i^{x_i}\left(1-p_i\right)^{n_i-x_i} \qquad \text{for } 0 \leq p_i \leq 1. \tag{2.12}$$

The joint posterior distribution of $\underline{p}$ is thus the product of k independently distributed $Beta\left(x_i+1, n_i-x_i+1\right)$ variates.

2.4 The Weighted Monte Carlo Method in the Case of the Probability Matching Prior for $\psi = \prod_{i=1}^{k} p_i^{a_i}$

In this section a weighted Monte Carlo method is described which will be used for simulation from the posterior distribution in the case of the probability matching prior. This method is especially suitable

for computing Bayesian credible intervals. It does not require knowing the closed form of the marginal posterior distribution of the parameter of interest, only the kernel of the posterior distribution of the parameters is needed.

As mentioned by Smith & Gelfand (1992), Chen & Shao (1999), Guttman & Menzefricke (2003), Skare et al. (2003), Kim (2006) and Li (2007) the weighted Monte Carlo (sampling - importance re-sampling (SIR)) algorithm aims at drawing a random sample from a target distribution π, by first drawing a sample from a proposal distribution q, and from this a smaller sample is drawn with sample probabilities proportional to the importance ratios π/q. For the algorithm to be efficient, it is important that q is a good approximation for π. This means that q should not have too light tails when compared to π. In the case of credibility intervals it is not even necessary to draw the smaller sample. The weights (sample probabilities) are, however, important.

If a uniform prior is put on $\underline{p}$, using Equation 2.12, the posterior (proposal) distribution is

$$q\left(\underline{p}|data\right) \propto \prod_{i=1}^{k} p_i^{x_i}\left(1-p_i\right)^{n_i-x_i} \qquad 0 \leq p_i \leq 1.$$

In the case of the probability matching prior, using Equation 2.2, the posterior (target) distribution is

$$\pi_{PM}\left(\underline{p}|data\right) \propto \left\{\sum_{i=1}^{k} \frac{a_i^2\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}} \left\{\prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\right\} \left\{\prod_{i=1}^{k} p_i^{x_i}\left(1-p_i\right)^{n_i-x_i}\right\} \qquad 0 \leq p_i \leq 1.$$

The sample probabilities are therefore proportional to

$$\frac{\pi_{PM}\left(\underline{p}|data\right)}{q\left(\underline{p}|data\right)} = \left\{\sum_{i=1}^{k} \frac{a_i^2\left(1-p_i\right)}{p_i}\right\}^{\frac{1}{2}} \left\{\prod_{i=1}^{k}\left(a_i\left(1-p_i\right)\right)^{-1}\right\} \qquad 0 \leq p_i \leq 1$$

and the normalised weights are

$$\omega_l = \frac{\left\{\sum_{i=1}^{k} \frac{a_i^2\left(1-p_i^{(l)}\right)}{p_i^{(l)}}\right\}^{\frac{1}{2}} \left\{\prod_{i=1}^{k} a_i^{-1}\left(1-p_i^{(l)}\right)^{-1}\right\}}{\sum_{l=1}^{n}\left[\left\{\sum_{i=1}^{k} \frac{a_i^2\left(1-p_i^{(l)}\right)}{p_i^{(l)}}\right\}^{\frac{1}{2}} \left\{\prod_{i=1}^{k} a_i^{-1}\left(1-p_i^{(l)}\right)^{-1}\right\}\right]} \qquad l = 1,2,\ldots,n$$

where n is the number of simulations. A straightforward way of doing the weighted Monte Carlo (WMC) method was proposed by Chen & Shao (1999).

The Monte Carlo method:

- **Step 1**

  Obtain a Monte Carlo sample $\left\{\left(p_1^{(l)}, p_2^{(l)} \ldots, p_k^{(l)}\right); l = 1,2,\ldots,n\right\}$ from the proposal distribution $q\left(\underline{p}|data\right)$ and calculate $\psi^{(l)} = \prod_{i=1}^{k}\left(p_i^{(l)}\right)^{a_i}$ for $l = 1,2,\ldots,n$.

- **Step 2**

  Sort $\left\{\psi^{(l)}, (l = 1,2,\ldots,n)\right\}$ to obtain the ordered values $\psi^{[1]} \leq \psi^{[2]} \leq \cdots \leq \psi^{[n]}$.

- **Step 3**

  Each simulated ψ value has an associated weight. Therefore compute the weighted function $\omega_{(l)}$ associated with the l^{th} ordered $\psi^{[l]}$ value.

- **Step 4**

  Add the weights up from left to right (from the first on) until one obtains $\sum_{l=1}^{n_1}\omega_{(l)} = \alpha/2$. Write down the corresponding $\psi^{[n_1]}$ value and denote it as $\psi_{(\alpha/2)}$. Add the weights up from right to left (from the last back) until one obtains $\sum_{l=n_2}^{n}\omega_{(l)} = \alpha/2$. Write down the corresponding $\psi^{[n_2]}$ value and denote it as $\psi_{(1-\alpha/2)}$.

- **Step 5**

  The $100\,(1-\alpha)\,\%$ Bayesian credible interval is: $\left(\psi_{(\alpha/2)}, \psi_{(1-\alpha/2)}\right)$.

2.5 Example and Simulation Studies

2.5.1 Example - Reliability of Independent Parallel Components System (Kim, 2006)

Consider the following example for the probability of failure of independent parallel components system, using the observed data values from Harris (1971). One has to assume that the two systems consist of two and three components in parallel, respectively. The respective probabilities of system failure will then be $\psi_1 = p_1 p_2$ and $\psi_2 = p_1 p_2 p_3$. ψ_1 is the product of two binomial parameters, and ψ_2 is the product of three binomial parameters. When manufacturing the components system, one may be interested in an upper bound of the confidence interval on the system failure. The upper bound of the confidence interval, $\psi_{(1-\alpha)}$, for each system failure, which is given by $P\left(0 \leq \psi_j \leq \psi_{(1-\alpha)}\right) = 1-\alpha$ will be estimated for $j = 1,2$. The estimate of $\psi_{(1-\alpha)}$ is therefore the upper end point of a one-sided $(1-\alpha)\,100\%$ confidence interval for ψ_j. The methods used to obtain the upper limit of the confidence

interval are: the likelihood ratio method by Madansky (1965); randomised limit based method by Harris (1971); Bayesian method by Kim (2006). The last two columns in Table 2.1 are obtained from the probability matching prior and the Jeffreys prior for the product of k binomial parameters. Kim approximated the binomial distribution by the Poisson distribution and obtained a probability matching prior for $\tilde{\theta} = \prod_{i=1}^{k} \lambda_i$, the product of k Poisson rates. The prior is $\pi(\underline{\lambda}) \propto \sqrt{\sum_{i=1}^{k} \lambda_i^{-1}}$. A simulated value for ψ is then obtained from the linear relationship between $\tilde{\theta}$ and ψ, namely $\psi = \tilde{\theta} / \prod_{i=1}^{k} n_i$. Comparisons between these five estimates are made in Table 2.1. The values for Madansky's and Harris' methods are from Harris (1971) and the values for the Bayesian method are from Kim (2006).

The effectiveness of the comparisons between the five methods in Table 2.1 is rather restricted, since the five methods are all approximate and we do not have the exact confidence coefficient.

Table 2.1: Upper confidence limits for $\prod_{i=1}^{k} p_i$ with confidence coefficient $1 - \alpha = 0.9$.

Sample sizes n_1, n_2	**Observed** x_1, x_2	**Madansky's Method**	**Harris' Method**	**Bayesian Method**	**Probability Matching Prior**	**Jeffreys Prior**
100, 100	3, 5	0.00433	0.00416	0.00406	0.00393	0.00355
100, 100	1, 4	0.00188	0.00184	0.00172	0.00167	0.00145
100, 100	2, 2	0.00168	0.00170	0.00157	0.00155	0.00131
150, 150	3, 3	0.00133	0.00128	0.00124	0.00120	0.00107
Sample sizes n_1, n_2, n_3	**Observed** x_1, x_2, x_3	**Madansky's Method**	**Harris' Method**	**Bayesian Method**	**Probability Matching Prior**	**Jeffreys Prior**
100, 100, 100	1, 2, 1	0.000019	0.000027	0.000021	0.000021	0.000013
100, 100, 100	2, 3, 5	0.000133	0.000145	0.000132	0.000129	0.000102

As mentioned the last two columns are added to Table 2 of Kim (2006) and give $\psi_{(1-\alpha)}$ for the probability matching prior and the Jeffreys prior of $\psi_1 = p_1 p_2$ and $\psi_2 = p_1 p_2 p_3$. The values of $\psi_{(1-\alpha)}$ in the case of the probability matching prior compare well with those obtained by the other researchers while it seems that the Jeffreys prior tends to somewhat under estimate the upper confidence limit.

2.5.2 Simulation Study I - Comparison of Four Priors for $\psi_1 = p_1 p_2$

In this section comparisons will be made between the following four priors:

1. $\pi_{PM}(\underline{p}) \propto \left\{ \sum_{i=1}^{2} \frac{(1-p_i)}{p_i} \right\}^{\frac{1}{2}} \prod_{i=1}^{2} (1-p_i)^{-1}$;

2. $\pi_J\left(\underline{p}\right) \propto \left(\prod_{i=1}^{2} \frac{1}{p_i(1-p_i)}\right)^{\frac{1}{2}}$;

3. $\pi_{PM}\left(\underline{\lambda}\right) \propto \sqrt{\sum_{i=1}^{2} \lambda_i^{-1}}$;

4. $\pi_J\left(\underline{\lambda}\right) \propto \left(\prod_{i=1}^{2} \lambda_i\right)^{-\frac{1}{2}}$,

for the following binomial distributions:

1. $n_1 = 10, p_1 = 0.4$ and $n_2 = 12, p_2 = 0.6$;

2. $n_1 = 20, p_1 = 0.4$ and $n_2 = 24, p_2 = 0.6$;

3. $n_1 = 40, p_1 = 0.4$ and $n_2 = 48, p_2 = 0.6$.

The priors denoted by π_{PM} (1 and 3) are probability matching priors while those denoted by π_J (2 and 4) are Jeffreys priors. The parameter of interest is $\psi_1 = p_1 p_2$. The Poisson parameter is $\lambda_i = n_i p_i$, $i = 1, 2$. The two priors (3 and 4) using the Poisson approximation will be derived and discussed in Chapter 4. The results are given in Tables 2.2 and 2.3, where the number of simulations is equal to 1 000 in both cases.

Table 2.2: Frequentist coverage probabilities for 0.95 posterior quantile of $\psi_1 = p_1 p_2$.

	Binomial			**Poisson**		
$1-\alpha = 0.95$	**1**	**2**	**3**	**1**	**2**	**3**
n_1	10	20	40	10	20	40
p_1	0.4	0.4	0.4	0.4	0.4	0.4
λ_1				4	8	16
n_2	12	24	48	12	24	48
p_2	0.6	0.6	0.6	0.6	0.6	0.6
λ_2				7.2	14.4	28.8
λ_0				[4 7.2]	[8 14.4]	[16 28.8]
# x vectors	1 000	1 000	1 000	1 000	1 000	1 000
# $\lambda's$				1 000	1 000	1 000
π_{PM}	0.953	0.954	0.95	0.933	0.933	0.949
π_J	0.926	0.944	0.946	0.913	0.917	0.949

Table 2.3: Frequentist coverage probabilities for 0.05 posterior quantile of $\psi_1 = p_1 p_2$.

	Binomial			**Poisson**		
$\alpha = 0.05$	**1**	**2**	**3**	**1**	**2**	**3**
n_1	10	20	40	10	20	40
p_1	0.4	0.4	0.4	0.4	0.4	0.4
λ_1				4	8	16
n_2	12	24	48	12	24	48
p_2	0.6	0.6	0.6	0.6	0.6	0.6
λ_2				7.2	14.4	28.8
λ_0				[4 7.2]	[8 14.4]	[16 28.8]
# x vectors	1 000	1 000	1 000	1 000	1 000	1 000
# $\lambda's$				1 000	1 000	1 000
π_{PM}	0.048	0.047	0.052	0.048	0.055	0.064
π_J	0.027	0.031	0.042	0.038	0.043	0.054

From Tables 2.2 and 2.3 it is clear that the priors π_{PM} are better than the Jeffreys priors, π_J, in most of the situations. It is surprising that $\pi_{PM}(\underline{\lambda})$ performs better than $\pi_J(\underline{p})$ since $\pi_{PM}(\underline{\lambda})$ is the probability matching prior for the Poisson distribution. The latter will be a good approximation to the binomial distribution if n is large and p is small. However the values used in Tables 2.2 and 2.3 are $p_1 = 0.4$ and $p_2 = 0.6$, which are quite large. As expected and although this is a limited experiment it seems that $\pi_{PM}(\underline{p})$ is the best prior of the four.

2.5.3 Simulation Study II - A comparison of the Jeffreys, Uniform and Probability Matching priors for $\psi_1 = p_1 p_2$

In this section a more extensive simulation study is done and coverage probabilities are obtained for $\psi_1 = p_1 p_2$, the product of two binomial parameters. For comparison purposes the following priors will be used:

1. the Jeffreys prior: $\pi_J(\underline{p}) \propto \prod_{i=1}^{2} p_i^{-\frac{1}{2}} (1-p_i)^{-\frac{1}{2}}$;
2. the uniform prior: $\pi_U(\underline{p}) \propto$ constant;
3. the probability matching prior: $\pi_{PM}(\underline{p}) \propto \left\{ \sum_{i=1}^{2} (1-p_i) p_i^{-1} \right\}^{\frac{1}{2}} \prod_{i=1}^{2} (1-p_i)^{-1}$.

The parameter values for the binomial distribution are $n_1 = n_2 = 10$, $n_1 = n_2 = 20$, $n_1 = n_2 = 30$ and $p_i = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ (for $i = 1,2$). The average length and standard deviation of the intervals are also given. The number of X variates and the number of simulations are equal to 1 000. The average length and standard deviation of the intervals are calculated by using the following

formulas, where I_i is the interval length and n^* the number of intervals:

$$\text{average length} = \frac{1}{n^*}\sum_{i=1}^{n^*} I_i$$

and

$$\text{standard deviation} = \sqrt{\frac{1}{n^*-1}\sum_{i=1}^{n^*}\left(I_i-\bar{I}\right)^2},$$

Tables 2.4 to 2.6 contain coverage probabilities, mean lengths and standard deviations for the Jeffreys, uniform and probability matching priors when $n_1 = n_2 = 10$, for several choices of p_1 and p_2. These results will be summarised in Figures 2.1 to 2.3 where box plots are constructed. MATLAB® was used to construct the box plots in this chapter. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points the algorithm considers not to be outliers, and the outliers are plotted individually.

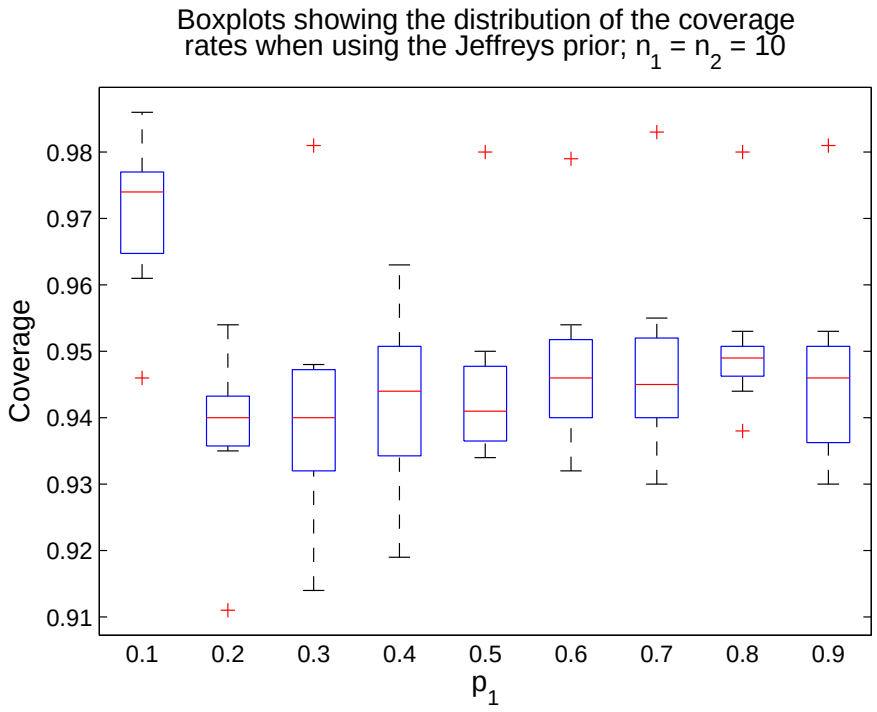


Figure 2.1: Box plot summarising the coverage rates of the 95% credibility intervals for $\psi_1 = p_1 p_2$ for $p_1 = 0.1 : 0.1 : 0.9$, using the Jeffreys prior; $n_1 = n_2 = 10$.

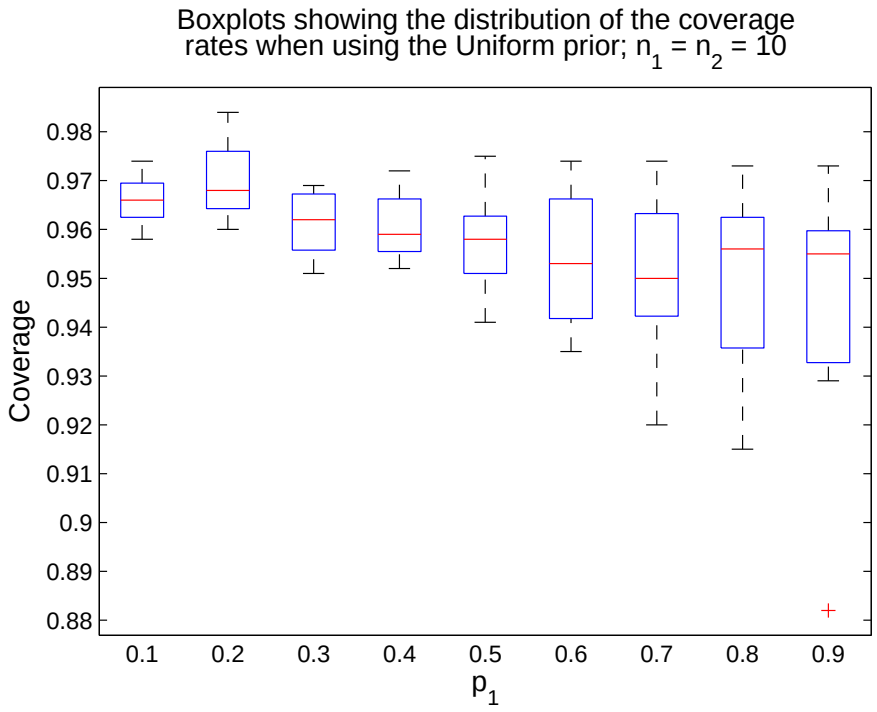


Figure 2.2: Box plot summarising the coverage rates of the 95% credibility intervals for $\psi_1 = p_1 p_2$ for $p_1 = 0.1 : 0.1 : 0.9$, using the uniform prior; $n_1 = n_2 = 10$.

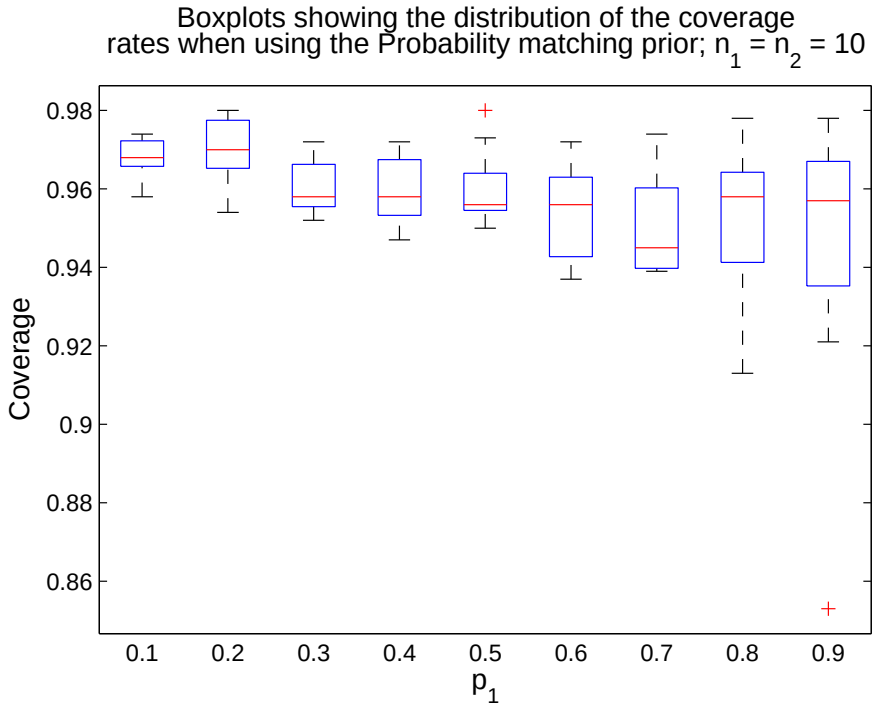


Figure 2.3: Box plot summarising the coverage rates of the 95% credibility intervals for $\psi_1 = p_1 p_2$ for $p_1 = 0.1 : 0.1 : 0.9$, using the probability matching prior; $n_1 = n_2 = 10$.

From Figure 2.1 we see that when the Jeffreys prior is used, the median coverage probability is below 0.95 for all values of p_1, apart from $p_1 = 0.1$. When $p_1 = 0.1$ the median coverage probability is above 0.97. From Figure 2.2 we see that when the uniform prior is used, the median coverage probability is above 0.95 for all values of p_1, apart from $p_1 = 0.7$. When $p_1 = 0.7$ the median coverage probability is equal to 0.95. From Figure 2.3 we see that when the probability matching prior is used, the median coverage probability is above 0.95 for all values of p_1, apart from $p_1 = 0.7$. When $p_1 = 0.7$ the median coverage probability is just below to 0.95. From these figures it seems that the performance of the uniform and probability matching priors are very similar, where there is over coverage most of the time. With the Jeffreys prior it seems that there is under coverage most of the time.

Table 2.4: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the Jeffreys prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 10$.

	Jeffreys prior $n_1 = n_2 = 10$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9860	0.9440	0.9810	0.9630	0.9800	0.9790	0.9830	0.9800	0.9810	0.9752
(b)	0.0721	0.1073	0.1367	0.1631	0.1910	0.2187	0.2477	0.2745	0.3016	0.1903
(c)	0.0460	0.0614	0.0710	0.0780	0.0806	0.0848	0.0904	0.0935	0.0968	0.0781
0.2 (a)	0.9460	0.9110	0.9140	0.9370	0.9410	0.9470	0.9450	0.9490	0.9300	0.9356
(b)	0.1048	0.1507	0.1848	0.2231	0.2558	0.2931	0.3269	0.3526	0.3863	0.2531
(c)	0.0606	0.0771	0.0830	0.0882	0.0932	0.0940	0.0956	0.0940	0.0962	0.0869
0.3 (a)	0.9700	0.9400	0.9290	0.9260	0.9380	0.9320	0.9440	0.9530	0.9440	0.9418
(b)	0.1330	0.1850	0.2257	0.2702	0.3049	0.3409	0.3737	0.4086	0.4428	0.2983
(c)	0.0684	0.0790	0.0887	0.0903	0.0899	0.0879	0.0868	0.0838	0.0789	0.0837
0.4 (a)	0.9610	0.9360	0.9330	0.9190	0.9440	0.9460	0.9340	0.9500	0.9340	0.9397
(b)	0.1662	0.2230	0.2734	0.3169	0.3487	0.3886	0.4148	0.4496	0.4781	0.3399
(c)	0.0796	0.0880	0.0925	0.0905	0.0858	0.0835	0.0782	0.0675	0.0619	0.0808
0.5 (a)	0.9660	0.9350	0.9350	0.9440	0.9350	0.9510	0.9490	0.9480	0.9530	0.9462
(b)	0.1955	0.2613	0.3057	0.3473	0.3890	0.4238	0.4457	0.4733	0.4990	0.3712
(c)	0.0844	0.0935	0.0892	0.0841	0.0797	0.0713	0.0643	0.0542	0.0436	0.0738
0.6 (a)	0.9740	0.9430	0.9440	0.9430	0.9370	0.9450	0.9550	0.9440	0.9370	0.9469
(b)	0.2273	0.2845	0.3457	0.3886	0.4218	0.4493	0.4740	0.4917	0.5053	0.3987
(c)	0.0887	0.0957	0.0898	0.0831	0.0735	0.0603	0.0482	0.0408	0.0347	0.0683
0.7 (a)	0.9750	0.9380	0.9400	0.9480	0.9470	0.9540	0.9510	0.9380	0.9500	0.9490
(b)	0.2506	0.3219	0.3770	0.4189	0.4520	0.4741	0.4887	0.4972	0.5019	0.4203
(c)	0.0928	0.0961	0.0857	0.0761	0.0648	0.0494	0.0372	0.0333	0.0358	0.0635
0.8 (a)	0.9760	0.9540	0.9470	0.9500	0.9500	0.9400	0.9300	0.9490	0.9460	0.9491
(b)	0.2754	0.3585	0.4133	0.4515	0.4753	0.4909	0.4955	0.4932	0.4799	0.4371
(c)	0.0943	0.0941	0.0852	0.0691	0.0535	0.0426	0.0358	0.0392	0.0547	0.0632
0.9 (a)	0.9800	0.9430	0.9480	0.9530	0.9340	0.9400	0.9420	0.9470	0.9460	0.9481
(b)	0.3045	0.3813	0.4382	0.4809	0.4975	0.5039	0.5012	0.4815	0.4464	0.4484
(c)	0.0996	0.0988	0.0807	0.0591	0.0422	0.0332	0.0381	0.0492	0.0657	0.0630
mean (a)	0.9704	0.9382	0.9412	0.9426	0.9451	0.9482	0.9481	0.9509	0.9468	**0.9480**
(b)	0.1922	0.2526	0.3001	0.3401	0.3707	0.3981	0.4187	0.4358	0.4490	**0.3508**
(c)	0.0794	0.0871	0.0851	0.0798	0.0737	0.0674	0.0639	0.0617	0.0631	**0.0735**

From Table 2.4 the overall average for the coverage probabilities is equal to 0.9480, which is slightly below the nominal level of 0.95. When $p_1 = 0.1$ and $p_2 = 0.1$ the coverage rate is equal to 0.9860, which is much higher than the nominal level, and when looking at $p_1 = 0.4$ and $p_2 = 0.4$ the coverage rate is equal to 0.9190, which is much lower than the nominal level. When using the Jeffreys prior, for $\psi_1 = p_1 p_2$, we have under coverage except when one of the values of p is 0.1.

Table 2.5: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the uniform prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 10$.

	Uniform prior $n_1 = n_2 = 10$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9640	0.9720	0.9680	0.9660	0.9650	0.9740	0.9620	0.9620	0.9590	0.9658
(b)	0.0932	0.1252	0.1582	0.1827	0.2103	0.2391	0.2668	0.2936	0.3226	0.2102
(c)	0.0435	0.0551	0.0591	0.0689	0.0708	0.0724	0.0725	0.0750	0.0732	0.0656
0.2 (a)	0.9710	0.9600	0.9670	0.9720	0.9750	0.9730	0.9740	0.9730	0.9730	0.9709
(b)	0.1246	0.1651	0.1992	0.2274	0.2640	0.2893	0.3193	0.3543	0.3802	0.2582
(c)	0.0535	0.0663	0.0713	0.0748	0.0765	0.0776	0.0793	0.0787	0.0743	0.0725
0.3 (a)	0.9660	0.9620	0.9630	0.9610	0.9520	0.9530	0.9490	0.9640	0.9550	0.9583
(b)	0.1541	0.1992	0.2341	0.2715	0.3096	0.3361	0.3655	0.4000	0.4258	0.2995
(c)	0.0609	0.0720	0.0746	0.0786	0.0812	0.0775	0.0769	0.0687	0.0650	0.0728
0.4 (a)	0.9630	0.9650	0.9690	0.9570	0.9590	0.9640	0.9640	0.9580	0.9550	0.9616
(b)	0.1818	0.2309	0.2718	0.3084	0.3467	0.3782	0.4043	0.4298	0.4550	0.3341
(c)	0.0675	0.0756	0.0772	0.0771	0.0760	0.0678	0.0645	0.0605	0.0542	0.0689
0.5 (a)	0.9610	0.9680	0.9620	0.9560	0.9480	0.9620	0.9500	0.9560	0.9620	0.9583
(b)	0.2139	0.2586	0.3048	0.3438	0.3734	0.4059	0.4295	0.4551	0.4778	0.3625
(c)	0.0701	0.0774	0.0788	0.0773	0.0709	0.0635	0.0575	0.0457	0.0376	0.0643
0.6 (a)	0.9690	0.9650	0.9520	0.9540	0.9410	0.9510	0.9630	0.9350	0.9430	0.9526
(b)	0.2382	0.2926	0.3340	0.3744	0.4063	0.4341	0.4528	0.4705	0.4849	0.3875
(c)	0.0725	0.0778	0.0786	0.0722	0.0648	0.0556	0.0447	0.0399	0.0297	0.0596
0.7 (a)	0.9580	0.9840	0.9580	0.9670	0.9560	0.9350	0.9430	0.9360	0.9340	0.9523
(b)	0.2674	0.3242	0.3664	0.4013	0.4303	0.4492	0.4684	0.4789	0.4848	0.4079
(c)	0.0741	0.0770	0.0736	0.0658	0.0556	0.0494	0.0372	0.0296	0.0277	0.0545
0.8 (a)	0.9740	0.9760	0.9570	0.9590	0.9580	0.9380	0.9400	0.9150	0.9290	0.9496
(b)	0.2910	0.3503	0.3925	0.4296	0.4523	0.4706	0.4788	0.4808	0.4767	0.4247
(c)	0.0721	0.0795	0.0705	0.0601	0.0499	0.0381	0.0288	0.0252	0.0329	0.0508
0.9 (a)	0.9680	0.9760	0.9510	0.9520	0.9620	0.9430	0.9200	0.9370	0.8820	0.9434
(b)	0.3180	0.3764	0.4251	0.4574	0.4770	0.4853	0.4854	0.4764	0.4576	0.4398
(c)	0.0742	0.0772	0.0651	0.0518	0.0397	0.0307	0.0275	0.0321	0.0407	0.0488
mean (a)	0.9660	0.9698	0.9608	0.9604	0.9573	0.9548	0.9517	0.9484	0.9436	**0.9570**
(b)	0.2091	0.2580	0.2984	0.3329	0.3633	0.3875	0.4078	0.4266	0.4406	**0.3472**
(c)	0.0654	0.0731	0.0721	0.0696	0.0650	0.0592	0.0543	0.0506	0.0484	**0.0620**

From Table 2.5 the overall average for the coverage probabilities is equal to 0.9570, which is slightly above the nominal level of 0.95. When $p_1 = 0.1$ and $p_2 = 0.1$ the coverage rate is equal to 0.9640, which is higher than the nominal level, but this value is closer to the nominal level than the coverage reached by the Jeffreys prior for this case. When looking at $p_1 = 0.4$ and $p_2 = 0.4$ the coverage rate is equal to 0.9570, which is almost equal to the nominal level, and this value is greater than the coverage reached by the Jeffreys prior for this case. When using the uniform prior, for $\psi_1 = p_1 p_2$, we have over coverage except when p_1 and p_2 are both large.

Table 2.6: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the probability matching prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 10$.

	Probability matching prior $n_1 = n_2 = 10$									
$\downarrow p_2 \quad p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9670	0.9540	0.9630	0.9720	0.9730	0.9630	0.9580	0.9650	0.9670	0.9647
(b)	0.0933	0.1304	0.1543	0.1789	0.2101	0.2428	0.2672	0.2941	0.3216	0.2103
(c)	0.0435	0.0558	0.0608	0.0636	0.0684	0.0716	0.0743	0.0722	0.0700	0.0645
0.2 (a)	0.9720	0.9720	0.9720	0.9690	0.9800	0.9720	0.9740	0.9780	0.9780	0.9741
(b)	0.1256	0.1645	0.1960	0.2342	0.2616	0.2895	0.3218	0.3490	0.3799	0.2580
(c)	0.0519	0.0656	0.0679	0.0759	0.0770	0.0779	0.0792	0.0788	0.0738	0.0720
0.3 (a)	0.9680	0.9630	0.9650	0.9580	0.9530	0.9630	0.9610	0.9610	0.9570	0.9610
(b)	0.1557	0.1994	0.2329	0.2696	0.3038	0.3351	0.3688	0.3958	0.4216	0.2981
(c)	0.0612	0.0719	0.0779	0.0773	0.0804	0.0744	0.0754	0.0697	0.0676	0.0729
0.4 (a)	0.9730	0.9790	0.9560	0.9510	0.9560	0.9590	0.9600	0.9640	0.9670	0.9628
(b)	0.1827	0.2317	0.2725	0.3112	0.3428	0.3720	0.4020	0.4329	0.4580	0.3340
(c)	0.0662	0.0762	0.0794	0.0805	0.0764	0.0705	0.0657	0.0568	0.0506	0.0691
0.5 (a)	0.9740	0.9700	0.9700	0.9470	0.9610	0.9490	0.9390	0.9580	0.9620	0.9589
(b)	0.2077	0.2617	0.3024	0.3448	0.3745	0.4000	0.4285	0.4548	0.4762	0.3612
(c)	0.0673	0.0787	0.0754	0.0787	0.0709	0.0680	0.0590	0.0473	0.0374	0.0647
0.6 (a)	0.9680	0.9660	0.9580	0.9610	0.9550	0.9560	0.9420	0.9550	0.9500	0.9568
(b)	0.2402	0.2930	0.3360	0.3786	0.4061	0.4305	0.4506	0.4717	0.4855	0.3880
(c)	0.0731	0.0793	0.0777	0.0699	0.0654	0.0560	0.0485	0.0344	0.0286	0.0592
0.7 (a)	0.9620	0.9800	0.9520	0.9570	0.9560	0.9430	0.9400	0.9460	0.9400	0.9529
(b)	0.2661	0.3231	0.3622	0.4062	0.4319	0.4520	0.4678	0.4795	0.4870	0.4084
(c)	0.0737	0.0790	0.0751	0.0662	0.0552	0.0452	0.0366	0.0294	0.0252	0.0540
0.8 (a)	0.9670	0.9770	0.9580	0.9540	0.9500	0.9370	0.9450	0.9130	0.9210	0.9469
(b)	0.2943	0.3512	0.4001	0.4290	0.4543	0.4699	0.4794	0.4831	0.4753	0.4263
(c)	0.0717	0.0781	0.0673	0.0594	0.0485	0.0378	0.0276	0.0246	0.0315	0.0496
0.9 (a)	0.9580	0.9680	0.9540	0.9670	0.9570	0.9420	0.9390	0.9270	0.8530	0.9406
(b)	0.3227	0.3841	0.4210	0.4588	0.4744	0.4852	0.4852	0.4768	0.4595	0.4409
(c)	0.0740	0.0744	0.0673	0.0504	0.0407	0.0287	0.0255	0.0309	0.0387	0.0478
mean (a)	0.9677	0.9699	0.9609	0.9596	0.9601	0.9538	0.9509	0.9519	0.9439	**0.9576**
(b)	0.2098	0.2599	0.2975	0.3346	0.3622	0.3863	0.4079	0.4264	0.4405	**0.3472**
(c)	0.0647	0.0732	0.0721	0.0691	0.0648	0.0589	0.0546	0.0493	0.0470	**0.0615**

From Table 2.6 the overall average for the coverage probabilities is equal to 0.9576, which is slightly above the nominal level of 0.95 and a bit higher than the overall average when using the uniform prior. When $p_1 = 0.1$ and $p_2 = 0.1$ the coverage rate is equal to 0.9670, which is a bit higher than the nominal level, but this value is closer to the nominal level than the coverage reached by the Jeffreys prior for this case and is almost exactly the same as the result from the uniform prior. When looking at $p_1 = 0.4$ and $p_2 = 0.4$ the coverage rate is equal to 0.9510, which is almost equal to the nominal level. When using the probability matching prior, for $\psi_1 = p_1 p_2$, we have over coverage

except for larger values of p_1 and p_2 where we have under coverage. For instance when $p_1 = 0.9$ and $p_2 = 0.9$ the coverage rate is equal to 0.8530, which is way below the nominal level of 0.95.

In general, we can conclude that, when $n_1 = n_2 = 10$ and for different values of p_1 and p_2 the Jeffreys prior produce coverage rates below the nominal level and that the uniform and probability matching priors produce coverage rates above the nominal level. Where the overall average coverage rate for the Jeffreys prior is equal to 0.9480, the overall average coverage rate for the uniform prior is equal to 0.9570 and the overall average coverage rate for the probability matching prior is equal to 0.9576. The average interval lengths of the Jeffreys prior are also generally larger than that of the uniform and probability matching priors. The uniform and probability matching priors give smaller standard deviation values for the interval lengths than the Jeffreys prior.

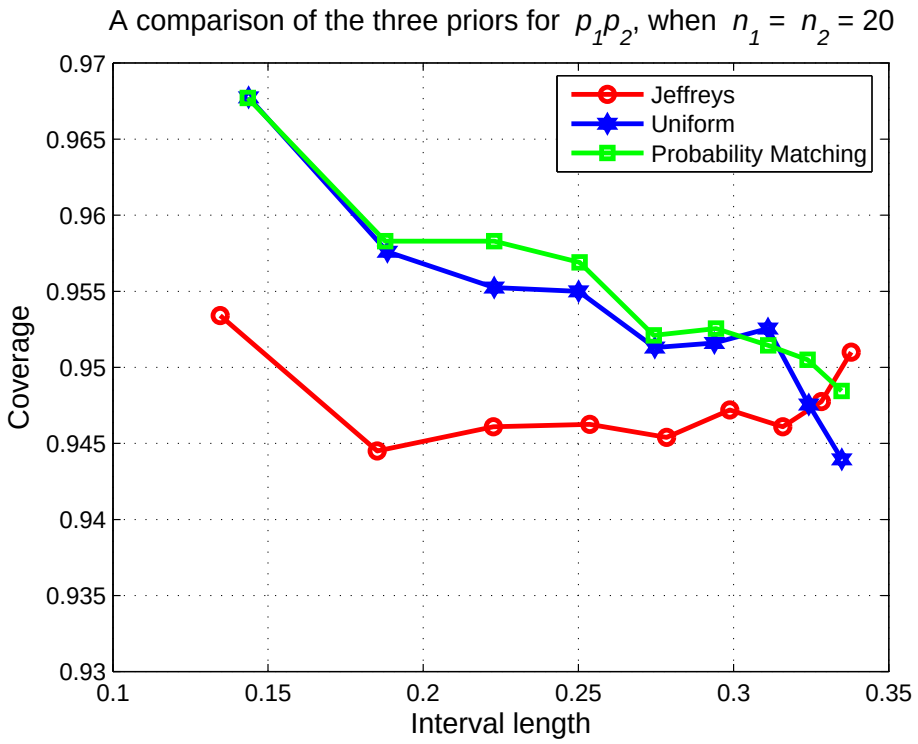


Figure 2.4: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1p_2$ against the interval length for the three priors when $n_1 = n_2 = 20$.

Figure 2.4 shows line plots of the coverage rates obtained when using the Jeffreys, uniform and probability matching priors for $\psi_1 = p_1p_2$ when $n_1 = n_2 = 20$. The values plotted are averages over the nine possible values for p_2. The Jeffreys prior underestimates the coverage probabilities while the uniform prior and the probability matching prior tend to overestimate the coverage probabilities.

Table 2.7: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the Jeffreys prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 20$.

	Jeffreys prior $n_1 = n_2 = 20$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9430	0.9570	0.9510	0.9450	0.9440	0.9500	0.9540	0.9620	0.9670	0.9526
(b)	0.0451	0.0714	0.0925	0.1141	0.1340	0.1578	0.1799	0.1962	0.2208	0.1346
(c)	0.0259	0.0323	0.0389	0.0436	0.0494	0.0536	0.0561	0.0602	0.0640	0.0471
0.2 (a)	0.9550	0.9370	0.9320	0.9390	0.9470	0.9360	0.9410	0.9500	0.9560	0.9437
(b)	0.0701	0.1018	0.1338	0.1623	0.1880	0.2134	0.2406	0.2653	0.2903	0.1851
(c)	0.0327	0.0392	0.0463	0.0481	0.0514	0.0548	0.0554	0.0547	0.0543	0.0485
0.3 (a)	0.9520	0.9470	0.9480	0.9590	0.9390	0.9570	0.9410	0.9500	0.9550	0.9498
(b)	0.0931	0.1315	0.1656	0.1994	0.2264	0.2548	0.2849	0.3104	0.3390	0.2228
(c)	0.0404	0.0450	0.0473	0.0481	0.0503	0.0485	0.0513	0.0469	0.0429	0.0467
0.4 (a)	0.9540	0.9440	0.9300	0.9470	0.9320	0.9640	0.9490	0.9430	0.9410	0.9449
(b)	0.1117	0.1603	0.1973	0.2339	0.2603	0.2913	0.3164	0.3410	0.3663	0.2532
(c)	0.0423	0.0495	0.0512	0.0506	0.0477	0.0437	0.0428	0.0391	0.0335	0.0445
0.5 (a)	0.9570	0.9410	0.9390	0.9630	0.9460	0.9360	0.9350	0.9460	0.9470	0.9456
(b)	0.1359	0.1883	0.2262	0.2644	0.2930	0.3190	0.3397	0.3618	0.3811	0.2788
(c)	0.0499	0.0526	0.0525	0.0460	0.0451	0.0404	0.0363	0.0289	0.0225	0.0416
0.6 (a)	0.9580	0.9470	0.9390	0.9350	0.9400	0.9520	0.9610	0.9430	0.9510	0.9473
(b)	0.1557	0.2134	0.2557	0.2879	0.3178	0.3394	0.3588	0.3725	0.3851	0.2985
(c)	0.0522	0.0524	0.0496	0.0463	0.0402	0.0339	0.0270	0.0213	0.0180	0.0379
0.7 (a)	0.9480	0.9400	0.9480	0.9360	0.9550	0.9370	0.9610	0.9440	0.9490	0.9464
(b)	0.1804	0.2401	0.2849	0.3159	0.3405	0.3586	0.3702	0.3758	0.3770	0.3159
(c)	0.0574	0.0549	0.0486	0.0420	0.0345	0.0283	0.0185	0.0180	0.0213	0.0359
0.8 (a)	0.9580	0.9460	0.9480	0.9470	0.9570	0.9450	0.9310	0.9400	0.9560	0.9476
(b)	0.1985	0.2687	0.3084	0.3426	0.3614	0.3734	0.3757	0.3706	0.3588	0.3287
(c)	0.0619	0.0560	0.0465	0.0379	0.0284	0.0206	0.0179	0.0232	0.0315	0.0360
0.9 (a)	0.9630	0.9490	0.9470	0.9570	0.9470	0.9470	0.9390	0.9530	0.9470	0.9499
(b)	0.2199	0.2918	0.3373	0.3664	0.3816	0.3846	0.3773	0.3577	0.3238	0.3378
(c)	0.0642	0.0568	0.0449	0.0326	0.0222	0.0173	0.0229	0.0345	0.0463	0.0380
mean (a)	0.9542	0.9453	0.9424	0.9476	0.9452	0.9471	0.9458	0.9479	0.9521	**0.9475**
(b)	0.1345	0.1853	0.2224	0.2541	0.2781	0.2991	0.3159	0.3279	0.3380	**0.2617**
(c)	0.0474	0.0487	0.0473	0.0439	0.0410	0.0379	0.0365	0.0363	0.0371	**0.0418**

From Table 2.7 the overall average for the coverage probabilities is equal to 0.9475, which is slightly below the nominal level of 0.95. When $p_1 = 0.6$ and $p_2 = 0.1$ the coverage rate is equal to 0.9500, which is equal to nominal level, and when looking at $p_1 = 0.8$ and $p_2 = 0.8$ the coverage rate is equal to 0.9400, which is below the nominal level. When using the Jeffreys prior, for $\psi_1 = p_1 p_2$, we have under coverage except when one of the values of p is 0.1. These results are similar to the results obtained when $n_1 = n_2 = 10$.

Table 2.8: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the uniform prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 20$.

	Uniform prior $n_1 = n_2 = 20$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9730	0.9650	0.9700	0.9650	0.9700	0.9740	0.9710	0.9770	0.9640	0.9699
(b)	0.0534	0.0790	0.0991	0.1220	0.1441	0.1677	0.1870	0.2073	0.2320	0.1435
(c)	0.0241	0.0311	0.0357	0.0403	0.0436	0.0452	0.0495	0.0503	0.0524	0.0413
0.2 (a)	0.9600	0.9500	0.9630	0.9490	0.9570	0.9660	0.9550	0.9650	0.9600	0.9583
(b)	0.0771	0.1103	0.1375	0.1635	0.1933	0.2154	0.2394	0.2637	0.2908	0.1879
(c)	0.0312	0.0389	0.0408	0.0465	0.0480	0.0483	0.0494	0.0503	0.0488	0.0447
0.3 (a)	0.9740	0.9510	0.9590	0.9560	0.9480	0.9460	0.9650	0.9590	0.9480	0.9562
(b)	0.1004	0.1405	0.1706	0.1980	0.2281	0.2546	0.2800	0.3045	0.3296	0.2229
(c)	0.0352	0.0428	0.0433	0.0469	0.0475	0.0481	0.0445	0.0414	0.0399	0.0433
0.4 (a)	0.9670	0.9470	0.9470	0.9570	0.9560	0.9460	0.9560	0.9580	0.9650	0.9554
(b)	0.1234	0.1647	0.2009	0.2277	0.2563	0.2834	0.3074	0.3327	0.3556	0.2502
(c)	0.0401	0.0454	0.0461	0.0452	0.0438	0.0426	0.0392	0.0351	0.0296	0.0408
0.5 (a)	0.9580	0.9640	0.9440	0.9360	0.9520	0.9460	0.9630	0.9500	0.9430	0.9507
(b)	0.1451	0.1909	0.2293	0.2583	0.2847	0.3103	0.3327	0.3529	0.3710	0.2750
(c)	0.0433	0.0465	0.0477	0.0472	0.0419	0.0388	0.0313	0.0264	0.0218	0.0383
0.6 (a)	0.9690	0.9610	0.9620	0.9560	0.9380	0.9380	0.9480	0.9470	0.9440	0.9514
(b)	0.1655	0.2162	0.2537	0.2828	0.3086	0.3283	0.3477	0.3636	0.3757	0.2936
(c)	0.0471	0.0472	0.0442	0.0424	0.0392	0.0338	0.0258	0.0210	0.0170	0.0353
0.7 (a)	0.9680	0.9580	0.9460	0.9510	0.9450	0.9560	0.9500	0.9420	0.9430	0.9510
(b)	0.1905	0.2435	0.2790	0.3091	0.3316	0.3489	0.3614	0.3686	0.3719	0.3116
(c)	0.0484	0.0477	0.0452	0.0395	0.0326	0.0255	0.0198	0.0154	0.0178	0.0324
0.8 (a)	0.9640	0.9540	0.9420	0.9600	0.9530	0.9510	0.9420	0.9270	0.9110	0.9449
(b)	0.2087	0.2674	0.3050	0.3326	0.3502	0.3632	0.3687	0.3663	0.3574	0.3244
(c)	0.0516	0.0482	0.0438	0.0348	0.0279	0.0205	0.0154	0.0166	0.0249	0.0315
0.9 (a)	0.9570	0.9620	0.9560	0.9610	0.9480	0.9430	0.9370	0.9270	0.9110	0.9447
(b)	0.2320	0.2880	0.3308	0.3551	0.3706	0.3756	0.3710	0.3572	0.3322	0.3347
(c)	0.0538	0.0487	0.0404	0.0306	0.0220	0.0163	0.0185	0.0246	0.0338	0.0321
mean (a)	0.9656	0.9569	0.9543	0.9546	0.9519	0.9518	0.9541	0.9502	0.9432	**0.9536**
(b)	0.1440	0.1890	0.2229	0.2499	0.2742	0.2941	0.3106	0.3241	0.3351	**0.2604**
(c)	0.0416	0.0441	0.0430	0.0415	0.0385	0.0355	0.0326	0.0312	0.0318	**0.0378**

From Table 2.8 the overall average for the coverage probabilities is equal to 0.9536, which is almost equal to the nominal level of 0.95. When $p_1 = 0.6$ and $p_2 = 0.1$ the coverage rate is equal to 0.9740, which is higher than the nominal level, and also higher than the coverage reached by the Jeffreys prior for this case. When looking at $p_1 = 0.8$ and $p_2 = 0.8$ the coverage rate is equal to 0.9270, which is below the nominal level, and this value is also smaller than the coverage reached by the Jeffreys prior for this case. When using the uniform prior, for $\psi_1 = p_1 p_2$, we have over coverage except for larger values of p_1 and p_2 where we have under coverage. Similar results were obtained for $n_1 = n_2 = 10$.

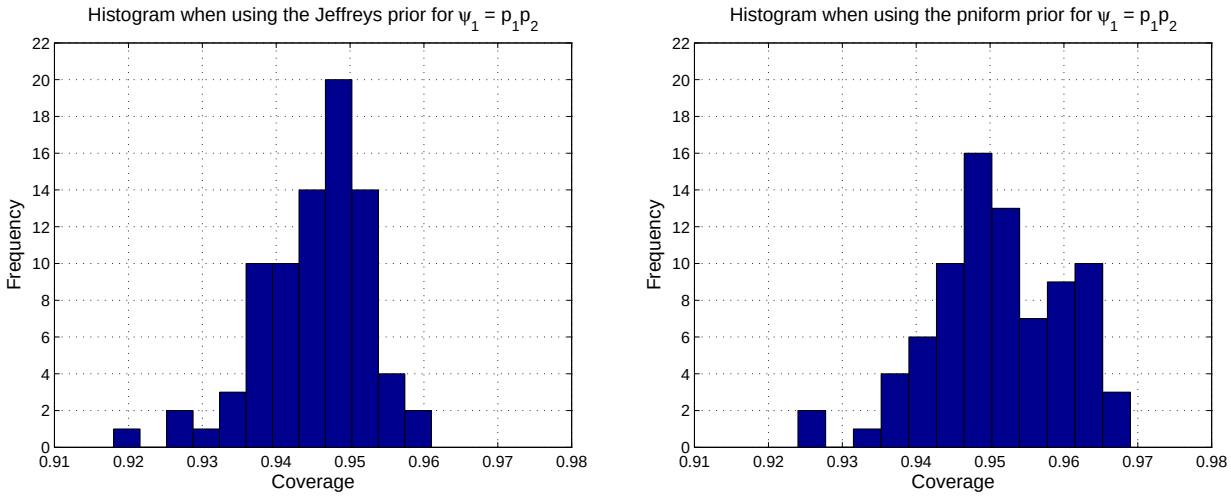
Table 2.9: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the probability matching prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 20$.

	Probability matching prior $n_1 = n_2 = 20$									
$\downarrow p_2 \quad p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9670	0.9550	0.9670	0.9740	0.9680	0.9780	0.9620	0.9730	0.9700	0.9682
(b)	0.0513	0.0776	0.0992	0.1224	0.1460	0.1663	0.1883	0.2082	0.2309	0.1433
(c)	0.0239	0.0315	0.0357	0.0395	0.0438	0.0471	0.0495	0.0494	0.0517	0.0413
0.2 (a)	0.9600	0.9560	0.9580	0.9620	0.9430	0.9570	0.9630	0.9560	0.9620	0.9574
(b)	0.0782	0.1083	0.1395	0.1612	0.1908	0.2179	0.2408	0.2623	0.2901	0.1877
(c)	0.0309	0.0385	0.0428	0.0445	0.0471	0.0480	0.0488	0.0505	0.0478	0.0443
0.3 (a)	0.9650	0.9670	0.9500	0.9630	0.9540	0.9640	0.9570	0.9720	0.9570	0.9610
(b)	0.0999	0.1378	0.1713	0.1985	0.2260	0.2560	0.2820	0.3038	0.3301	0.2228
(c)	0.0359	0.0407	0.0458	0.0459	0.0474	0.0452	0.0431	0.0412	0.0391	0.0427
0.4 (a)	0.9690	0.9580	0.9570	0.9590	0.9580	0.9600	0.9510	0.9470	0.9600	0.9577
(b)	0.1221	0.1652	0.2015	0.2298	0.2552	0.2849	0.3095	0.3317	0.3556	0.2506
(c)	0.0390	0.0445	0.0466	0.0451	0.0445	0.0415	0.0395	0.0348	0.0295	0.0405
0.5 (a)	0.9710	0.9550	0.9470	0.9590	0.9430	0.9410	0.9570	0.9610	0.9480	0.9536
(b)	0.1430	0.1915	0.2259	0.2593	0.2874	0.3090	0.3310	0.3514	0.3704	0.2743
(c)	0.0430	0.0473	0.0482	0.0436	0.0420	0.0387	0.0329	0.0276	0.0214	0.0383
0.6 (a)	0.9660	0.9530	0.9520	0.9440	0.9470	0.9500	0.9330	0.9500	0.9550	0.9500
(b)	0.1672	0.2152	0.2525	0.2834	0.3092	0.3302	0.3479	0.3638	0.3754	0.2939
(c)	0.0457	0.0477	0.0473	0.0433	0.0373	0.0332	0.0266	0.0196	0.0164	0.0352
0.7 (a)	0.9730	0.9560	0.9520	0.9390	0.9450	0.9520	0.9540	0.9460	0.9530	0.9522
(b)	0.1884	0.2420	0.2803	0.3102	0.3318	0.3493	0.3603	0.3685	0.3716	0.3114
(c)	0.0475	0.0492	0.0450	0.0403	0.0338	0.0260	0.0193	0.0152	0.0181	0.0327
0.8 (a)	0.9710	0.9560	0.9630	0.9430	0.9460	0.9530	0.9410	0.9390	0.9270	0.9488
(b)	0.2104	0.2635	0.3063	0.3311	0.3510	0.3639	0.3676	0.3657	0.3579	0.3241
(c)	0.0502	0.0494	0.0409	0.0362	0.0275	0.0208	0.0152	0.0171	0.0244	0.0313
0.9 (a)	0.9630	0.9770	0.9540	0.9620	0.9510	0.9410	0.9380	0.9260	0.9140	0.9473
(b)	0.2320	0.2889	0.3298	0.3561	0.3713	0.3756	0.3709	0.3570	0.3314	0.3348
(c)	0.0537	0.0479	0.0393	0.0297	0.0213	0.0172	0.0171	0.0243	0.0344	0.0317
mean (a)	0.9672	0.9592	0.9556	0.9561	0.9506	0.9551	0.9507	0.9522	0.9496	**0.9551**
(b)	0.1436	0.1878	0.2229	0.2502	0.2743	0.2948	0.3109	0.3236	0.3348	**0.2603**
(c)	0.0411	0.0441	0.0435	0.0409	0.0383	0.0353	0.0324	0.0311	0.0314	**0.0376**

From Table 2.9 the overall average for the coverage probabilities is equal to 0.9551, which is slightly above the nominal level of 0.95 and a bit higher than the overall average when using the uniform prior. When $p_1 = 0.6$ and $p_2 = 0.1$ the coverage rate is equal to 0.9780, which is higher than the nominal level, and is almost exactly the same as the result from the uniform prior, where the Jeffreys prior gave a coverage probability of 0.9500 in this case. When looking at $p_1 = 0.8$ and $p_2 = 0.8$ the coverage rate is equal to 0.9390, which is almost the same as the coverage reached by the Jeffreys prior for this case. When using the probability matching prior, for $\psi_1 = p_1 p_2$, we have

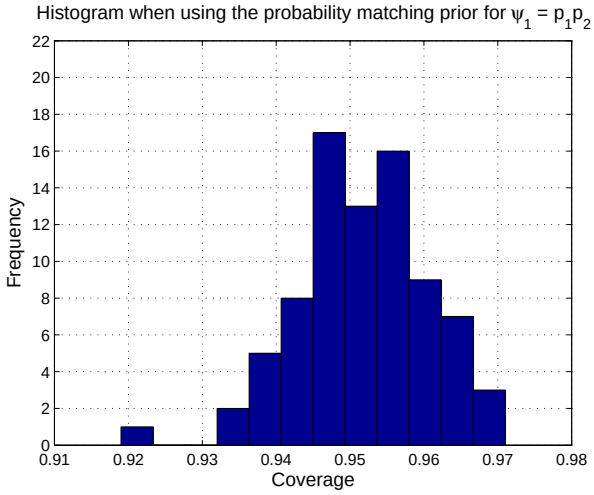
over coverage except for larger values of p_1 and p_2 where we have under coverage. For instance when $p_1 = 0.9$ and $p_2 = 0.9$ the coverage rate is equal to 0.9140, which is way below the nominal level of 0.95, but much better than the value of 0.8530 which was obtained when $n_1 = n_2 = 10$.

The average interval lengths of the Jeffreys prior are also generally larger than that of the uniform and probability matching priors. The uniform and probability matching priors give smaller standard deviation values for the interval lengths than the Jeffreys prior.



(a)

(b)



(c)

Figure 2.5: Histograms showing the distribution of the coverage rates of the 95% credibility intervals for $\psi_1 = p_1 p_2$ against the interval length for the three priors when $n_1 = n_2 = 30$, (a) the Jeffreys prior, (b) the uniform prior, (c) the probability matching prior.

Table 2.10: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the Jeffreys prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 30$.

	Jeffreys prior $n_1 = n_2 = 30$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9520	0.9450	0.9290	0.9350	0.9510	0.9400	0.9370	0.9440	0.9280	0.9401
(b)	0.0341	0.0537	0.0731	0.0912	0.1114	0.1275	0.1487	0.1668	0.1834	0.1100
(c)	0.0162	0.0216	0.0274	0.0311	0.0332	0.0374	0.0397	0.0419	0.0470	0.0328
0.2 (a)	0.9180	0.9500	0.9430	0.9390	0.9490	0.9470	0.9510	0.9430	0.9390	0.9421
(b)	0.0542	0.0827	0.1089	0.1300	0.1556	0.1764	0.1971	0.2248	0.2470	0.1530
(c)	0.0229	0.0265	0.0307	0.0342	0.0353	0.0373	0.0382	0.0381	0.0386	0.0335
0.3 (a)	0.9360	0.9460	0.9560	0.9430	0.9440	0.9520	0.9440	0.9480	0.9430	0.9458
(b)	0.0735	0.1075	0.1369	0.1617	0.1903	0.2124	0.2376	0.2605	0.2840	0.1849
(c)	0.0271	0.0308	0.0324	0.0352	0.0350	0.0344	0.0338	0.0318	0.0295	0.0322
0.4 (a)	0.9470	0.9520	0.9260	0.9460	0.9450	0.9380	0.9340	0.9480	0.9490	0.9428
(b)	0.0943	0.1293	0.1626	0.1918	0.2166	0.2437	0.2649	0.2865	0.3069	0.2107
(c)	0.0316	0.0331	0.0366	0.0332	0.0333	0.0322	0.0303	0.0261	0.0238	0.0311
0.5 (a)	0.9410	0.9350	0.9450	0.9400	0.9500	0.9390	0.9480	0.9490	0.9530	0.9444
(b)	0.1102	0.1549	0.1872	0.2166	0.2444	0.2637	0.2844	0.3029	0.3199	0.2316
(c)	0.0338	0.0364	0.0345	0.0332	0.0318	0.0286	0.0249	0.0204	0.0169	0.0289
0.6 (a)	0.9460	0.9520	0.9490	0.9440	0.9380	0.9510	0.9530	0.9470	0.9490	0.9477
(b)	0.1313	0.1782	0.2138	0.2426	0.2650	0.2838	0.3005	0.3118	0.3229	0.2500
(c)	0.0361	0.0364	0.0347	0.0307	0.0282	0.0233	0.0184	0.0144	0.0125	0.0261
0.7 (a)	0.9420	0.9600	0.9390	0.9570	0.9540	0.9540	0.9530	0.9470	0.9440	0.9500
(b)	0.1484	0.2014	0.2360	0.2648	0.2843	0.2988	0.3100	0.3147	0.3149	0.2637
(c)	0.0402	0.0360	0.0346	0.0292	0.0244	0.0193	0.0136	0.0123	0.0163	0.0251
0.8 (a)	0.9450	0.9430	0.9500	0.9490	0.9530	0.9470	0.9610	0.9390	0.9470	0.9482
(b)	0.1657	0.2230	0.2614	0.2866	0.3035	0.3129	0.3143	0.3094	0.2970	0.2749
(c)	0.0422	0.0390	0.0315	0.0266	0.0198	0.0147	0.0114	0.0158	0.0236	0.0250
0.9 (a)	0.9360	0.9410	0.9440	0.9490	0.9460	0.9530	0.9530	0.9510	0.9470	0.9467
(b)	0.1832	0.2448	0.2826	0.3078	0.3202	0.3223	0.3152	0.2973	0.2685	0.2824
(c)	0.0459	0.0401	0.0299	0.0222	0.0163	0.0129	0.0161	0.0233	0.0329	0.0266
mean (a)	0.9403	0.9471	0.9423	0.9447	0.9478	0.9468	0.9482	0.9462	0.9443	**0.9453**
(b)	0.1105	0.1528	0.1847	0.2103	0.2323	0.2491	0.2636	0.2750	0.2827	**0.2179**
(c)	0.0329	0.0333	0.0325	0.0306	0.0286	0.0267	0.0252	0.0249	0.0268	**0.0290**

From Table 2.10 the overall average for the coverage probabilities is equal to 0.9453, which is slightly below the nominal level of 0.95. When $p_1 = 0.6$ and $p_2 = 0.6$ the coverage rate is equal to 0.9510, which is almost equal to nominal level, and when looking at $p_1 = 0.3$ and $p_2 = 0.1$ the coverage rate is equal to 0.9290, which is below the nominal level. When using the Jeffreys prior, for $\psi_1 = p_1 p_2$, we have under coverage most of time.

Table 2.11: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the uniform prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 30$.

	Uniform prior $n_1 = n_2 = 30$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9590	0.9520	0.9560	0.9630	0.9560	0.9690	0.9640	0.9610	0.9630	0.9603
(b)	0.0398	0.0597	0.0786	0.0990	0.1168	0.1338	0.1533	0.1708	0.1906	0.1158
(c)	0.0172	0.0224	0.0258	0.0286	0.0316	0.0331	0.0351	0.0382	0.0389	0.0301
0.2 (a)	0.9410	0.9570	0.9510	0.9520	0.9470	0.9510	0.9630	0.9540	0.9570	0.9526
(b)	0.0588	0.0872	0.1116	0.1341	0.1573	0.1789	0.2013	0.2211	0.2442	0.1549
(c)	0.0228	0.0265	0.0297	0.0327	0.0334	0.0357	0.0360	0.0356	0.0358	0.0320
0.3 (a)	0.9630	0.9480	0.9400	0.9580	0.9470	0.9460	0.9480	0.9580	0.9670	0.9528
(b)	0.0784	0.1091	0.1392	0.1648	0.1901	0.2115	0.2347	0.2559	0.2805	0.1849
(c)	0.0258	0.0306	0.0321	0.0331	0.0333	0.0321	0.0330	0.0301	0.0268	0.0308
0.4 (a)	0.9500	0.9620	0.9490	0.9480	0.9450	0.9570	0.9480	0.9450	0.9450	0.9499
(b)	0.0994	0.1340	0.1647	0.1922	0.2143	0.2362	0.2607	0.2802	0.3015	0.2092
(c)	0.0292	0.0311	0.0332	0.0320	0.0335	0.0307	0.0287	0.0254	0.0218	0.0295
0.5 (a)	0.9640	0.9440	0.9610	0.9490	0.9330	0.9620	0.9530	0.9540	0.9690	0.9543
(b)	0.1155	0.1571	0.1902	0.2151	0.2381	0.2603	0.2791	0.2968	0.3142	0.2296
(c)	0.0319	0.0335	0.0322	0.0319	0.0315	0.0258	0.0226	0.0192	0.0146	0.0270
0.6 (a)	0.9610	0.9510	0.9400	0.9470	0.9550	0.9510	0.9430	0.9520	0.9590	0.9510
(b)	0.1358	0.1787	0.2111	0.2385	0.2605	0.2783	0.2931	0.3067	0.3173	0.2467
(c)	0.0340	0.0365	0.0343	0.0309	0.0271	0.0230	0.0192	0.0148	0.0122	0.0258
0.7 (a)	0.9610	0.9460	0.9620	0.9520	0.9420	0.9500	0.9260	0.9390	0.9380	0.9462
(b)	0.1545	0.2000	0.2341	0.2607	0.2786	0.2947	0.3028	0.3098	0.3117	0.2607
(c)	0.0357	0.0364	0.0314	0.0274	0.0239	0.0183	0.0144	0.0115	0.0144	0.0237
0.8 (a)	0.9580	0.9420	0.9460	0.9370	0.9480	0.9370	0.9460	0.9480	0.9520	0.9460
(b)	0.1707	0.2223	0.2576	0.2801	0.2961	0.3059	0.3095	0.3067	0.2977	0.2718
(c)	0.0384	0.0363	0.0304	0.0260	0.0197	0.0156	0.0116	0.0129	0.0191	0.0233
0.9 (a)	0.9650	0.9570	0.9540	0.9470	0.9500	0.9460	0.9410	0.9480	0.9240	0.9480
(b)	0.1898	0.2436	0.2791	0.3005	0.3137	0.3163	0.3126	0.2968	0.2713	0.2804
(c)	0.0390	0.0355	0.0286	0.0230	0.0157	0.0130	0.0136	0.0195	0.0269	0.0239
mean (a)	0.9580	0.9510	0.9510	0.9503	0.9470	0.9521	0.9480	0.9510	0.9527	**0.9512**
(b)	0.1159	0.1546	0.1851	0.2094	0.2295	0.2462	0.2608	0.2716	0.2810	**0.2171**
(c)	0.0304	0.0321	0.0308	0.0295	0.0277	0.0253	0.0238	0.0230	0.0234	**0.0273**

From Table 2.11 the overall average for the coverage probabilities is equal to 0.9512, which is almost equal to the nominal level of 0.95. When $p_1 = 0.6$ and $p_2 = 0.6$ the coverage rate is equal to 0.9510, which is higher than the nominal level, but equal to the coverage reached by the Jeffreys prior for this case. When looking at $p_1 = 0.3$ and $p_2 = 0.1$ the coverage rate is equal to 0.9560, which is above the nominal level, and this value is also higher than the coverage reached by the Jeffreys prior for this case. When using the uniform prior, for $\psi_1 = p_1 p_2$, we have over coverage except for larger values of p_1 and p_2.

Table 2.12: Coverage rate of the 95% credibility intervals for $\psi_1 = p_1 p_2$ using the probability matching prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 30$.

	Probability matching prior $n_1 = n_2 = 30$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9560	0.9670	0.9640	0.9520	0.9630	0.9580	0.9580	0.9590	0.9620	0.9599
(b)	0.0392	0.0595	0.0786	0.0986	0.1163	0.1352	0.1535	0.1712	0.1886	0.1156
(c)	0.0168	0.0204	0.0247	0.0292	0.0316	0.0352	0.0363	0.0387	0.0396	0.0303
0.2 (a)	0.9640	0.9470	0.9610	0.9470	0.9510	0.9640	0.9560	0.9480	0.9530	0.9546
(b)	0.0594	0.0864	0.1108	0.1350	0.1580	0.1778	0.2009	0.2226	0.2448	0.1551
(c)	0.0211	0.0263	0.0292	0.0316	0.0341	0.0345	0.0359	0.0366	0.0355	0.0317
0.3 (a)	0.9710	0.9530	0.9560	0.9570	0.9450	0.9470	0.9520	0.9560	0.9550	0.9547
(b)	0.0782	0.1099	0.1379	0.1643	0.1899	0.2118	0.2354	0.2558	0.2785	0.1846
(c)	0.0247	0.0293	0.0317	0.0324	0.0331	0.0344	0.0321	0.0309	0.0282	0.0308
0.4 (a)	0.9560	0.9450	0.9520	0.9490	0.9550	0.9330	0.9470	0.9550	0.9490	0.9490
(b)	0.0986	0.1347	0.1616	0.1902	0.2152	0.2359	0.2591	0.2812	0.3026	0.2088
(c)	0.0295	0.0321	0.0339	0.0334	0.0321	0.0318	0.0284	0.0248	0.0218	0.0298
0.5 (a)	0.9610	0.9430	0.9590	0.9520	0.9620	0.9480	0.9600	0.9550	0.9500	0.9544
(b)	0.1142	0.1589	0.1869	0.2160	0.2403	0.2596	0.2808	0.2979	0.3141	0.2298
(c)	0.0315	0.0331	0.0330	0.0326	0.0297	0.0271	0.0228	0.0196	0.0156	0.0272
0.6 (a)	0.9670	0.9490	0.9550	0.9570	0.9450	0.9460	0.9410	0.9530	0.9420	0.9506
(b)	0.1360	0.1779	0.2119	0.2388	0.2593	0.2791	0.2939	0.3068	0.3164	0.2467
(c)	0.0339	0.0350	0.0331	0.0300	0.0279	0.0227	0.0187	0.0141	0.0123	0.0253
0.7 (a)	0.9490	0.9460	0.9600	0.9490	0.9490	0.9390	0.9630	0.9470	0.9380	0.9489
(b)	0.1539	0.1994	0.2359	0.2606	0.2794	0.2939	0.3040	0.3098	0.3123	0.2610
(c)	0.0367	0.0353	0.0313	0.0282	0.0239	0.0189	0.0140	0.0111	0.0136	0.0237
0.8 (a)	0.9660	0.9630	0.9380	0.9500	0.9410	0.9470	0.9380	0.9450	0.9380	0.9473
(b)	0.1719	0.2228	0.2562	0.2798	0.2962	0.3059	0.3096	0.3064	0.2965	0.2717
(c)	0.0378	0.0336	0.0314	0.0257	0.0203	0.0140	0.0111	0.0130	0.0198	0.0230
0.9 (a)	0.9560	0.9590	0.9500	0.9510	0.9530	0.9540	0.9190	0.9490	0.9340	0.9472
(b)	0.1891	0.2444	0.2769	0.3000	0.3132	0.3176	0.3111	0.2977	0.2741	0.2804
(c)	0.0408	0.0357	0.0292	0.0212	0.0160	0.0117	0.0155	0.0190	0.0255	0.0239
mean (a)	0.9607	0.9524	0.9550	0.9516	0.9516	0.9484	0.9482	0.9519	0.9468	**0.9518**
(b)	0.1156	0.1549	0.1841	0.2093	0.2298	0.2463	0.2609	0.2722	0.2809	**0.2171**
(c)	0.0303	0.0312	0.0308	0.0294	0.0276	0.0256	0.0239	0.0231	0.0235	**0.0273**

From Table 2.12 the overall average for the coverage probabilities is equal to 0.9518, which is slightly above the nominal level of 0.95 and almost equal to the overall average when using the uniform prior. When $p_1 = 0.6$ and $p_2 = 0.6$ the coverage rate is equal to 0.9460, which is below the nominal level, and less than the coverage obtained from the Jeffreys and uniform priors in this case. When looking at $p_1 = 0.3$ and $p_2 = 0.1$ the coverage rate is equal to 0.9640, which is higher than the coverage reached by the Jeffreys and uniform priors for this case. When using the probability matching prior, for $\psi_1 = p_1 p_2$, we have over coverage most of the time when one of the values of p is 0.1.

The average interval lengths of the Jeffreys prior are also generally a bit larger than that of the uniform and probability matching priors. The uniform and probability matching priors give smaller standard deviation values for the interval lengths than the Jeffreys prior.

In Table 2.13 summary statistics (averages over the nine possible values of the parameter p_2) are given for the coverage probabilities, mean lengths and standard deviations for the 90% credibility intervals of $\psi_1 = p_1 p_2$.

Table 2.13: Coverage rate of the 90% credibility intervals for $\psi_1 = p_1 p_2$. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation. The values in this table are averages over the nine possible values of p_2.

		$n_1 = n_2 = 10$			$n_1 = n_2 = 20$		
p_1		**Jeffreys**	**Uniform**	**Probability matching**	**Jeffreys**	**Uniform**	**Probability matching**
0.1	(a)	0.936	0.932	0.935	0.874	0.926	0.923
	(b)	0.156	0.175	0.174	0.111	0.119	0.120
	(c)	0.069	0.057	0.057	0.040	0.036	0.036
0.2	(a)	0.882	0.931	0.933	0.890	0.913	0.919
	(b)	0.210	0.217	0.218	0.155	0.157	0.158
	(c)	0.075	0.063	0.063	0.042	0.038	0.037
0.3	(a)	0.887	0.920	0.916	0.892	0.906	0.907
	(b)	0.251	0.252	0.252	0.188	0.187	0.187
	(c)	0.074	0.063	0.063	0.041	0.037	0.037
0.4	(a)	0.897	0.914	0.910	0.888	0.914	0.906
	(b)	0.288	0.282	0.284	0.214	0.212	0.211
	(c)	0.070	0.060	0.060	0.038	0.035	0.035
0.5	(a)	0.894	0.917	0.909	0.891	0.907	0.902
	(b)	0.315	0.309	0.307	0.236	0.232	0.232
	(c)	0.065	0.056	0.057	0.036	0.033	0.033
0.6	(a)	0.903	0.913	0.910	0.887	0.906	0.901
	(b)	0.338	0.329	0.329	0.252	0.249	0.249
	(c)	0.060	0.051	0.052	0.033	0.030	0.030
0.7	(a)	0.893	0.909	0.905	0.890	0.906	0.902
	(b)	0.357	0.348	0.347	0.267	0.264	0.263
	(c)	0.056	0.047	0.048	0.031	0.028	0.028
0.8	(a)	0.897	0.902	0.898	0.887	0.900	0.902
	(b)	0.372	0.363	0.363	0.278	0.275	0.275
	(c)	0.056	0.044	0.045	0.030	0.027	0.027
0.9	(a)	0.899	0.896	0.900	0.888	0.901	0.898
	(b)	0.381	0.375	0.376	0.286	0.284	0.284
	(c)	0.057	0.043	0.042	0.033	0.027	0.027
Overall Mean	(a)	**0.899**	**0.915**	**0.913**	**0.887**	**0.909**	**0.907**
	(b)	**0.296**	**0.295**	**0.294**	**0.221**	**0.220**	**0.220**
	(c)	**0.065**	**0.054**	**0.054**	**0.036**	**0.032**	**0.032**

From Table 2.13 it seems that the coverage probabilities for the Jeffreys prior is, in general, somewhat smaller than 0.9 and that under coverage is larger for $n_1 = n_2 = 20$ than for $n_1 = n_2 = 10$. The uniform and probability matching priors, on the other hand, tend to give coverage probabilities larger than 0.9 and more so for the uniform prior. As can be expected the interval lengths and standard deviations are smaller for larger n.

It also seems that the probability matching prior gives the best results for $0.3 \leq p_i \leq 0.7$, $(i = 1,2)$. This also explains the good performance of the probability matching priors in Tables 2.2 and 2.3.

In Table 2.14 the overall averages are given for $n_1 = n_2 = 10$ and $n_1 = n_2 = 20$ for p_i = 0.3, 0.4, 0.5, 0.6 and 0.7, $(i = 1,2)$.

Table 2.14: Average coverage probabilities for $n_1 = n_2 = 10$ and $n_1 = n_2 = 20$ for p_i = 0.3, 0.4, 0.5, 0.6 and 0.7, $(i = 1,2)$.

$n_1 = n_2 = 10$			$n_1 = n_2 = 20$		
Jeffreys	**Uniform**	**Probability matching**	**Jeffreys**	**Uniform**	**Probability matching**
0.890	0.910	0.905	0.891	0.907	0.900

From Table 2.14 it can be seen that the probability matching prior is somewhat better than the uniform and Jeffreys priors. We conclude that all three priors are satisfactory in attaining the nominal coverage probabilities. In general the differences between the priors are quite small.

2.5.4 Simulation Study III - A comparison of the Jeffreys, Uniform and Probability Matching priors for $\psi_3 = p_1^2 p_2$ and $\psi_4 = p_1 p_2^2$

In this section a more extensive simulation study is done and coverage probabilities are obtained for $\psi_3 = p_1^2 p_2$ and $\psi_4 = p_1 p_2^2$. Assume that a system requires that at least one of two types of components be employed and that three components in parallel are needed. The probability of system failure will be $\psi_3 = p_1^2 p_2$ or $\psi_4 = p_1 p_2^2$.

For comparison purposes the following priors will be used:

1. the Jeffreys prior: $\pi_J\left(\underline{p}\right) \propto \prod_{i=1}^{2} p_i^{-\frac{1}{2}} (1-p_i)^{-\frac{1}{2}}$;

2. the uniform prior: $\pi_U\left(\underline{p}\right) \propto$ constant;

3. the probability matching prior: $\pi_{PM}\left(\underline{p}\right) \propto \left\{ \sum_{i=1}^{2} a_i^2 (1-p_i) p_i^{-1} \right\}^{\frac{1}{2}} \prod_{i=1}^{2} a_i^{-1} (1-p_i)^{-1}$.

Figures 2.6 and 2.7 give line plots of the coverage rates obtained when using the Jeffreys, uniform and probability matching priors for $\psi_3 = p_1^2 p_2$ and $\psi_4 = p_1 p_2^2$, respectively, when $n_1 = n_2 = 20$. The values plotted are averages over the nine possible values for p_1 and p_2.

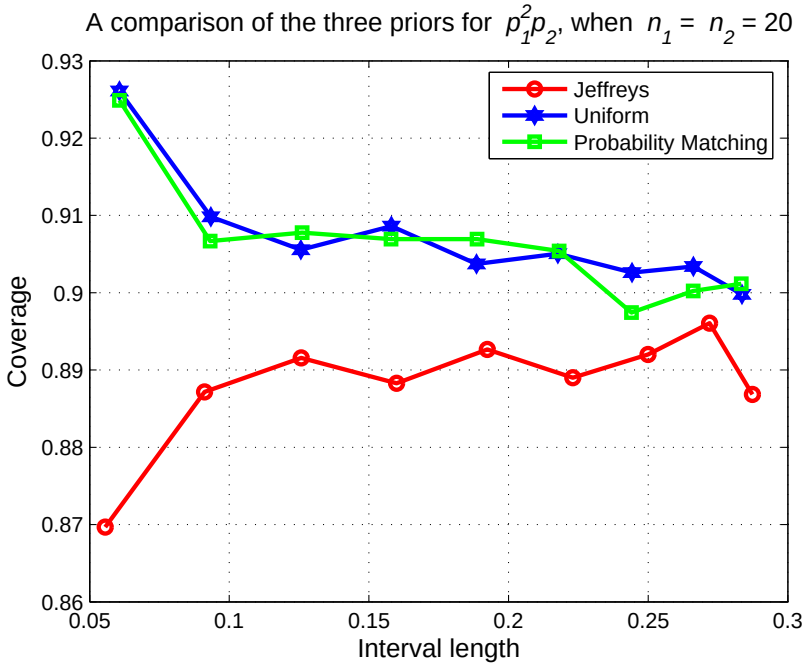


Figure 2.6: Coverage rate of the 90% credibility intervals for $\psi_3 = p_1^2 p_2$ against the interval length for the three priors when $n_1 = n_2 = 20$.

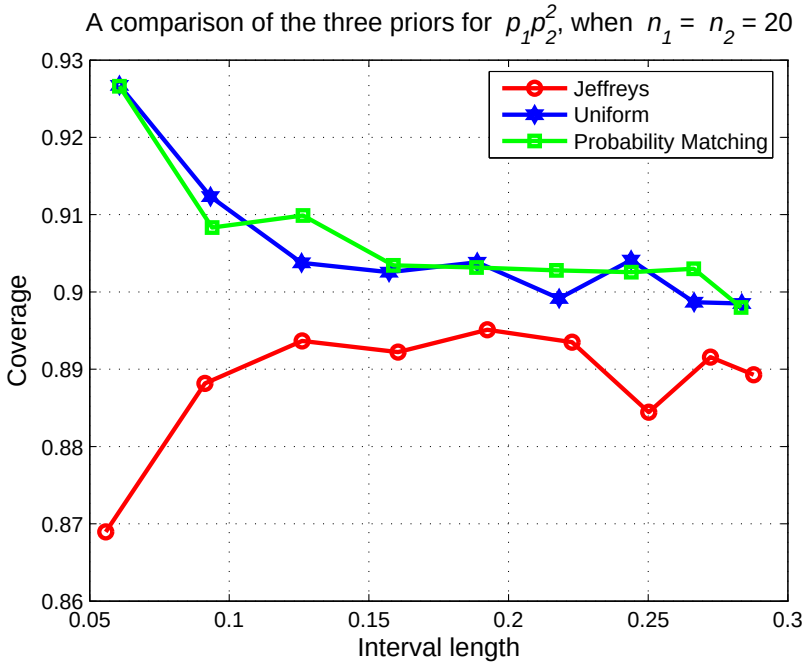


Figure 2.7: Coverage rate of the 90% credibility intervals for $\psi_4 = p_1 p_2^2$ against the interval length for the three priors when $n_1 = n_2 = 20$.

From Figures 2.6 and 2.7 it seems if the Jeffreys prior underestimates the coverage probabilities while the uniform prior and the probability matching prior tend to overestimate the coverage probabilities.

2.6 Conclusion

In this chapter the probability matching prior for the product of different powers of k binomial parameters, i.e. $\psi = \prod_{i=1}^{k} p_i^{a_i}$, was derived. An example and a number of simulation studies were considered. A weighted Monte Carlo was introduced, which was used for the simulation from the posterior distribution in the case of the probability matching prior. The example that was considered dealt with the probability of failure of independent parallel components system. Here the probability matching prior compared well with the other results, but the Jeffreys prior under estimated the upper confidence limit. In the first simulation study the probability matching prior, $\pi_{PM}\left(\underline{p}\right)$, performed the best of the four priors.

In the second simulation study a comparison was made between the Jeffreys, uniform and probability matching priors for the product of two binomial proportions. Different values of n_1, n_2, p_1 and p_2 were considered. The probability matching prior and the uniform prior gave similar results. In general, we can conclude that the coverage probabilities for the Jeffreys prior is below the nominal level. The uniform and probability matching priors, on the other hand, give coverage probabilities larger than the nominal level and more so for the uniform prior. The average interval lengths and standard deviations are smaller for larger values of n. In the last simulation study we compared the performance of the Jeffreys, uniform and probability matching priors for $\psi_3 = p_1^2 p_2$ and $\psi_4 = p_1 p_2^2$. Where one needs two types of components and three components in parallel are needed for a system to operate. Again the Jeffreys prior gave coverage rates below the nominal level, and the uniform and probability matching priors gave coverage rates above the nominal level.

Chapter 3

Estimation for a Linear Function of Binomial Rates

3.1 Introduction

Due to its important practical value, confidence interval construction for a linear function of binomial proportions has received some attention recently (Price & Bonett, 2004; Tebbs & Roths, 2008). In the first part of this chapter the probability matching prior for a linear function of binomial proportions will therefore be derived and in the latter part Bayesian credible intervals will be constructed for the difference between two binomial proportions.

Estimating the difference between two binomial proportions is a problem that occurs regularly in practice. There are some asymptotic procedures available for the construction of confidence intervals for the difference. A number of authors have studied the performance of these asymptotic procedures in circumstances where the samples are small. Some of them are Beal (1987), Newcombe (1998), Agresti & Caffo (2000) and Zhou et al. (2004). According to Roths & Tebbs (2006) asymptotic intervals are generally preferred to exact intervals. The reason for this, is that asymptotic intervals are often much easier to calculate than exact intervals and they can also produce acceptable results without wasteful conservatism. Roths & Tebbs (2006) showed how their intervals can be used adaptively in experiments conducted in stages over time. They concentrated on samples that are small.

3.2 The Probability Matching Prior for a Linear Combination of Binomial Proportions

The procedure of Datta & Ghosh (1995) will be used to derive the probability matching prior. The following theorem is proved.

Theorem 3.1. The probability matching prior for $\theta = \sum_{i=1}^{k} a_i p_i$, a linear combination of binomial proportions, is given by

$$\pi_{PM}\left(\underline{p}\right) = \pi_{PM}\left(p_1, p_2, \ldots, p_k\right) \quad \propto \quad \left\{ \sum_{i=1}^{k} a_i^2 p_i \left(1 - p_i\right) \right\}^{\frac{1}{2}} \prod_{i=1}^{k} p_i^{-1} \left(1 - p_i\right)^{-1}. \tag{3.1}$$

Proof. Assume that $X_1, X_2, \ldots, X_k$ are independent binomial random variables with $X_i \sim Bin\left(n_i, p_i\right)$ for $i = 1, 2, \ldots, k$.

It was shown in Appendix A in Theorem A.1 that the inverse of the Fisher information matrix for $n_1 = n_2 = \ldots = n_k$ is given by

$$F^{-1}\left(\underline{p}\right) \quad = \quad \begin{bmatrix} p_1\left(1-p_1\right) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & p_k\left(1-p_k\right) \end{bmatrix}.$$

We are interested in a probability matching prior for $t\left(\underline{p}\right) = \theta = \sum_{i=1}^{k} a_i p_i$, a linear combination of k binomial proportions.

Thus $t\left(\underline{p}\right) = a_1 p_1 + a_2 p_2 + \ldots + a_k p_k$.

Now

$$\frac{\partial t\left(\underline{p}\right)}{\partial p_i} \quad = \quad a_i \qquad\qquad \text{for } i = 1, 2, \ldots, k.$$

Further

$$\begin{aligned} \nabla_t'\left(\underline{p}\right) \quad &= \quad \begin{bmatrix} \frac{\partial t\left(\underline{p}\right)}{\partial p_1} & \frac{\partial t\left(\underline{p}\right)}{\partial p_2} & \cdots & \frac{\partial t\left(\underline{p}\right)}{\partial p_k} \end{bmatrix} \\ &= \quad \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix}. \end{aligned}$$

Also

$$\begin{aligned} \nabla_t'\left(\underline{p}\right) F^{-1}\left(\underline{p}\right) \quad &= \quad \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix} \begin{bmatrix} p_1\left(1-p_1\right) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & p_k\left(1-p_k\right) \end{bmatrix} \\ &= \quad \begin{bmatrix} a_1 p_1\left(1-p_1\right) & a_2 p_2\left(1-p_2\right) & \cdots & a_k p_k\left(1-p_k\right) \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned}\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right) &= \left[\begin{array}{cccc} a_1p_1\left(1-p_1\right) & a_2p_2\left(1-p_2\right) & \cdots & a_kp_k\left(1-p_k\right)\end{array}\right] \\ &\quad\times\left[\begin{array}{c} a_1\\ a_2\\ \vdots\\ a_k\end{array}\right] \\ &= \sum_{i=1}^{k}a_i^2p_i\left(1-p_i\right).\end{aligned}$$

Define

$$\begin{aligned}\eta'\left(\underline{p}\right) &= \frac{\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)}{\sqrt{\nabla_t'\left(\underline{p}\right)F^{-1}\left(\underline{p}\right)\nabla_t\left(\underline{p}\right)}} \\ &= \left[\begin{array}{cccc}\frac{a_1p_1(1-p_1)}{\sqrt{\sum_{i=1}^{k}a_i^2p_i(1-p_i)}} & \frac{a_2p_2(1-p_2)}{\sqrt{\sum_{i=1}^{k}a_i^2p_i(1-p_i)}} & \cdots & \frac{a_kp_k(1-p_k)}{\sqrt{\sum_{i=1}^{k}a_i^2p_i(1-p_i)}}\end{array}\right] \\ &= \left[\begin{array}{cccc}\eta_1\left(\underline{p}\right) & \eta_2\left(\underline{p}\right) & \cdots & \eta_k\left(\underline{p}\right)\end{array}\right].\end{aligned}$$

As before, the prior $\pi\left(\underline{p}\right)$ is a probability matching prior if and only if the differential equation $\sum_{i=1}^{k}\frac{\partial}{\partial p_i}\left\{\eta_i\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\}=0$ is satisfied.

Let

$$\pi\left(\underline{p}\right) = \left\{\sum_{i=1}^{k}a_i^2p_i\left(1-p_i\right)\right\}^{\frac{1}{2}}\prod_{i=1}^{k}p_i^{-1}\left(1-p_i\right)^{-1}$$

then

$$\begin{aligned}\eta_1\left(\underline{p}\right)\pi\left(\underline{p}\right) &= a_1p_1\left(1-p_1\right)\prod_{i=1}^{k}p_i^{-1}\left(1-p_i\right)^{-1} \\ &= a_1\prod_{i\neq 1}^{k}p_i^{-1}\left(1-p_i\right)^{-1}\end{aligned}$$

therefore

$$\frac{\partial}{\partial p_1}\left\{\eta_1\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = \frac{\partial}{\partial p_1}\left\{a_1\prod_{i\neq 1}^{k}p_i^{-1}\left(1-p_i\right)^{-1}\right\}=0$$

and

$$\begin{aligned}\eta_2\left(\underline{p}\right)\pi\left(\underline{p}\right) &= a_2 p_2\left(1-p_2\right)\prod_{i=1}^{k} p_i^{-1}\left(1-p_i\right)^{-1} \\ &= a_2\prod_{i\neq 2}^{k} p_i^{-1}\left(1-p_i\right)^{-1}\end{aligned}$$

therefore

$$\frac{\partial}{\partial p_2}\left\{\eta_2\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = \frac{\partial}{\partial p_2}\left\{a_2\prod_{i\neq 2}^{k} p_i^{-1}\left(1-p_i\right)^{-1}\right\} = 0$$

and

$$\begin{aligned}\eta_k\left(\underline{p}\right)\pi\left(\underline{p}\right) &= a_k p_k\left(1-p_k\right)\prod_{i=1}^{k} p_i^{-1}\left(1-p_i\right)^{-1} \\ &= a_k\prod_{i\neq k}^{k} p_i^{-1}\left(1-p_i\right)^{-1}\end{aligned}$$

therefore

$$\frac{\partial}{\partial p_k}\left\{\eta_k\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = \frac{\partial}{\partial p_k}\left\{a_k\prod_{i\neq k}^{k} p_i^{-1}\left(1-p_i\right)^{-1}\right\} = 0.$$

We can therefore conclude that

$$\sum_{i=1}^{k}\frac{\partial}{\partial p_i}\left\{\eta_i\left(\underline{p}\right)\pi\left(\underline{p}\right)\right\} = 0.$$

The differential equation is satisfied if $\pi\left(\underline{p}\right)$ is

$$\pi_{PM}\left(\underline{p}\right) \propto \left\{\sum_{i=1}^{k} a_i^2 p_i\left(1-p_i\right)\right\}^{\frac{1}{2}}\prod_{i=1}^{k} p_i^{-1}\left(1-p_i\right)^{-1} \qquad \text{for } 0\leq p_i\leq 1.$$

□

The joint posterior distribution when using the probability matching prior is given by

$$\begin{aligned}
\pi_{PM}\left(\underline{p}\,|data\right) &\propto \pi_{PM}\left(\underline{p}\right)\times L\left(\underline{p}\,|data\right)\\
&\propto \left\{\sum_{i=1}^{k} a_i^2 p_i(1-p_i)\right\}^{\frac{1}{2}} \prod_{i=1}^{k} p_i^{-1}(1-p_i)^{-1}\times\prod_{i=1}^{k}\left(\begin{array}{c} n_i \\ x_i \end{array}\right) p_i^{x_i}(1-p_i)^{n_i-x_i}\\
\therefore \pi_{PM}\left(\underline{p}\,|data\right) &\propto \left\{\sum_{i=1}^{k} a_i^2 p_i(1-p_i)\right\}^{\frac{1}{2}} \prod_{i=1}^{k} p_i^{x_i-1}(1-p_i)^{n_i-x_i-1} \qquad \text{for } 0\le p_i\le 1. \qquad (3.2)
\end{aligned}$$

If $k=2$, $a_1=1$, $a_2=-1$ and $a_3=a_4=\ldots=a_k=0$, then $\theta_1=p_1-p_2$ and the prior will be

$$\pi_{PM}(p_1,p_2) \propto \left\{\sum_{i=1}^{2} p_i(1-p_i)\right\}^{\frac{1}{2}} \prod_{i=1}^{2} p_i^{-1}(1-p_i)^{-1} \qquad (3.3)$$

and the joint posterior distribution will be

$$\pi_{PM}(p_1,p_2\,|data) \propto \left\{\sum_{i=1}^{2} p_i(1-p_i)\right\}^{\frac{1}{2}} \prod_{i=1}^{2} p_i^{x_i-1}(1-p_i)^{n_i-x_i-1} \qquad \text{for } 0\le p_i\le 1. \qquad (3.4)$$

Theorem 3.2. The posterior distribution, $\pi_{PM}(p_1,p_2\,|data)$, defined in Equation 3.4 is a proper distribution if $0<x_i<n_i$.

Proof. Since

$$\left\{\sum_{i=1}^{2} p_i(1-p_i)\right\}^{\frac{1}{2}} < \sum_{i=1}^{2} p_i^{\frac{1}{2}}(1-p_i)^{\frac{1}{2}}$$

it follows that

$$\left\{\sum_{i=1}^{2} p_i(1-p_i)\right\}^{\frac{1}{2}} \prod_{i=1}^{2} p_i^{x_i-1}(1-p_i)^{n_i-x_i-1} < \left\{\sum_{i=1}^{2} p_i^{\frac{1}{2}}(1-p_i)^{\frac{1}{2}}\right\}\prod_{i=1}^{2} p_i^{x_i-1}(1-p_i)^{n_i-x_i-1}.$$

Now

$$\int_0^1\int_0^1 \left\{\sum_{i=1}^{2} p_i^{\frac{1}{2}}(1-p_i)^{\frac{1}{2}}\right\}\left\{\prod_{i=1}^{2} p_i^{x_i-1}(1-p_i)^{n_i-x_i-1}\right\} dp_1dp_2$$

will converge if $x_i>0$ and $x_i<n_i$,$(i=1,2)$. □

3.3 The Jeffreys and Uniform Priors for a Linear Combination of Binomial Proportions

The Jeffreys prior, on the other hand, is proportional to the square root of the determinant of the Fisher information matrix and is given by

$$\begin{aligned} \pi_J\left(\underline{p}\right) &\propto \left|F\left(\underline{p}\right)\right|^{\frac{1}{2}} \\ &= \left(\prod_{i=1}^{k} \frac{1}{p_i\left(1-p_i\right)}\right)^{\frac{1}{2}}. \end{aligned} \tag{3.5}$$

The joint posterior distribution when using the Jeffreys prior is given by

$$\begin{aligned} \pi_J\left(\underline{p}|data\right) &\propto \pi_J\left(\underline{p}\right) \times L\left(\underline{p}|data\right) \\ &\propto \left(\prod_{i=1}^{k} \frac{1}{p_i\left(1-p_i\right)}\right)^{\frac{1}{2}} \times \prod_{i=1}^{k} \binom{n_i}{x_i} p_i^{x_i}\left(1-p_i\right)^{n_i-x_i} \\ \therefore \pi_J\left(\underline{p}|data\right) &\propto \prod_{i=1}^{k} p_i^{x_i-\frac{1}{2}}\left(1-p_i\right)^{n_i-x_i-\frac{1}{2}} \qquad \text{for } 0 \leq p_i \leq 1. \end{aligned} \tag{3.6}$$

The joint posterior distribution of $\underline{p}$ is thus the product of k independently distributed $Beta\left(x_i+\frac{1}{2}, n_i-x_i+\frac{1}{2}\right)$ variates.

If $k=1$ and $a_i=1$, $\pi_{PM}\left(p\right)$ becomes the Jeffreys prior. This will be shown in Theorem 3.3.

Theorem 3.3. If $k=1$ and $a_i=1$, the probability matching prior in the case of a linear combination of k binomial parameters becomes the Jeffreys prior.

Proof. From Equation 3.1 the probability matching prior for $\theta = \sum_{i=1}^{k} a_i p_i$, is given by

$$\pi_{PM}\left(\underline{p}\right) \propto \left\{\sum_{i=1}^{k} a_i^2 p_i\left(1-p_i\right)\right\}^{\frac{1}{2}} \prod_{i=1}^{k} p_i^{-1}\left(1-p_i\right)^{-1}.$$

Let $k=1$, $p_i=p$ and $a_i=1$, then the probability matching prior simplifies to

$$\begin{aligned} \pi_{PM}\left(p\right) &\propto \frac{\left(1-p\right)^{\frac{1}{2}}}{p^{\frac{1}{2}}}\left(1-p\right)^{-1} \\ &= \frac{1}{\left(1-p\right)^{\frac{1}{2}} p^{\frac{1}{2}}} \propto \pi_J\left(p\right). \end{aligned}$$

□

The uniform prior is proportional to a constant and is given by

$$\pi_U\left(\underline{p}\right) \propto \text{constant.} \tag{3.7}$$

The joint posterior is

$$\pi_U\left(\underline{p}|data\right) \propto \prod_{i=1}^{k} p_i^{x_i}\left(1-p_i\right)^{n_i-x_i} \qquad \text{for } 0 \leq p_i \leq 1. \tag{3.8}$$

The joint posterior distribution of $\underline{p}$ is thus the product of k independently distributed $Beta\left(x_i+1, n_i-x_i+1\right)$ variates.

3.4 Other Methods

Roths & Tebbs (2006) constructed confidence intervals for the difference between two binomial proportions, by first considering the Wald interval and the adjustment to the Wald interval proposed by Agresti & Caffo (2000). Then they focused on the intervals from Beal (1987) and the Haldane and Jeffreys-Perks intervals.

Assume that X_1 and X_2 are independent binomial random variables with $X_i \sim Bin\left(n_i, p_i\right)$ for $i = 1,2$. For $x_1 = 0,1,\ldots,n_1$ and $x_2 = 0,1,\ldots,n_2$, the maximum likelihood estimate (MLE) of $p_1 - p_2$ is equal to $\hat{p}_1 - \hat{p}_2$, where $\hat{p}_i = {}^{x_i}/_{n_i}$. The derivation of the MLE, $\hat{p}_i = {}^{x_i}/_{n_i}$, is given in Appendix B in Theorem B.1.

The statistic $\hat{p}_1 - \hat{p}_2$, when properly scaled and centered, converges to a standard Normal distribution,

$$\frac{\left(\hat{p}_1 - \hat{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{Var\left(\hat{p}_1 - \hat{p}_2\right)}} \xrightarrow{d} N\left(0,1\right)$$

where

$$Var\left(\hat{p}_1 - \hat{p}_2\right) = \frac{p_1\left(1-p_1\right)}{n_1} + \frac{p_2\left(1-p_2\right)}{n_2}$$

is the asymptotic variance of $\hat{p}_1 - \hat{p}_2$.

A large-sample $100\left(1-\alpha\right)\%$ confidence interval for $p_1 - p_2$ is the well known Wald interval

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Agresti & Caffo (2000) suggested an adjustment to the Wald interval to improve the coverage probability in small-sample situations. They suggested adding two pseudo-observations, one failure and one success, to each sample. Thereafter they computed the Wald limits based on the new data. To obtain the Agresti-Caffo interval, they replaced n_i with $n_i^* = n_i + 2$ and $\hat{p}_i$ with $\hat{p}_i^* = {}^{(x_i+1)}/_{n_i^*}$. The Agresti-Caffo interval can then be obtained as

$$(\hat{p}_1^* - \hat{p}_2^*) \quad \pm \quad z_{\alpha/2}\sqrt{\frac{\hat{p}_1^*\left(1-\hat{p}_1^*\right)}{n_1^*} + \frac{\hat{p}_2^*\left(1-\hat{p}_2^*\right)}{n_2^*}}.$$

The class of intervals from Beal (1987) is developed by assuming that the population proportions follow independent symmetric prior distributions. This leads to a family of intervals indexed by the hyperparameter of the prior. The well-known Haldane and Jeffreys-Perks intervals are special cases in this family (Roths & Tebbs, 2006).

Roths & Tebbs (2006) proposed data-driven Beal procedures, where prior hyperparameters are estimated using a parametric empirical Bayesian approach.

3.5 The Weighted Monte Carlo Method in the Case of the Probability Matching Prior for $\theta = \sum_{i=1}^{k} a_i p_i$

This method has been introduced in Section 2.4.

If a uniform prior is put on $\underline{p}$, using Equation 3.8, the posterior (proposal) distribution is

$$q\left(\underline{p}\,|data\right) \quad \propto \quad \prod_{i=1}^{k} p_i^{x_i}\left(1-p_i\right)^{n_i-x_i} \qquad\qquad 0 \leq p_i \leq 1.$$

In the case of the probability matching prior, using Equation 3.2, the posterior (target) distribution is

$$\pi_{PM}\left(\underline{p}\,|data\right) \quad \propto \quad \left\{\sum_{i=1}^{k} a_i^2 p_i\left(1-p_i\right)\right\}^{\frac{1}{2}} \prod_{i=1}^{k} p_i^{x_i-1}\left(1-p_i\right)^{n_i-x_i-1} \qquad\qquad 0 \leq p_i \leq 1.$$

The sample probabilities are therefore proportional to

$$\frac{\pi_{PM}\left(\underline{p}\,|data\right)}{q\left(\underline{p}\,|data\right)} \quad = \quad \left\{\sum_{i=1}^{k} a_i^2 p_i\left(1-p_i\right)\right\}^{\frac{1}{2}} \left\{\prod_{i=1}^{k} p_i^{-1}\left(1-p_i\right)^{-1}\right\} \qquad\qquad 0 \leq p_i \leq 1$$

and the normalised weights are

$$\omega_l = \frac{\left\{\sum_{i=1}^{k} a_i^2 p_i^{(l)}\left(1-p_i^{(l)}\right)\right\}^{\frac{1}{2}}\left\{\prod_{i=1}^{k}\left(p_i^{(l)}\right)^{-1}\left(1-p_i^{(l)}\right)^{-1}\right\}}{\sum_{l=1}^{n}\left[\left\{\sum_{i=1}^{k} a_i^2 p_i^{(l)}\left(1-p_i^{(l)}\right)\right\}^{\frac{1}{2}}\left\{\prod_{i=1}^{k}\left(p_i^{(l)}\right)^{-1}\left(1-p_i^{(l)}\right)^{-1}\right\}\right]} \qquad l = 1,2,\ldots,n$$

where n is the number of simulations.
The Monte Carlo method:

- **Step 1**

  Obtain a Monte Carlo sample $\left\{\left(p_1^{(l)}, p_2^{(l)} \ldots, p_k^{(l)}\right); l = 1,2,\ldots,n\right\}$ from the proposal distribution $q\left(\underline{p}|data\right)$ and calculate $\theta^{(l)} = \sum_{i=1}^{k} a_i p_i^{(l)}$ for $l = 1,2,\ldots,n$.

- **Step 2**

  Sort $\left\{\theta^{(l)}, (l = 1,2,\ldots,n)\right\}$ to obtain the ordered values $\theta^{[1]} \leq \theta^{[2]} \leq \cdots \leq \theta^{[n]}$.

- **Step 3**

  Each simulated θ value has an associated weight. Therefore compute the weighted function $\omega_{(l)}$ associated with the l^{th} ordered $\theta^{[l]}$ value.

- **Step 4**

  Add the weights up from left to right (from the first on) until one obtains $\sum_{l=1}^{n_1} \omega_{(l)} = \alpha/2$. Write down the corresponding $\theta^{[n_1]}$ value and denote it as $\theta_{(\alpha/2)}$. Add the weights up from right to left (from the last back) until one obtains $\sum_{l=n_2}^{n} \omega_{(l)} = \alpha/2$. Write down the corresponding $\theta^{[n_2]}$ value and denote it as $\theta_{(1-\alpha/2)}$.

- **Step 5**

  The $100(1-\alpha)\%$ Bayesian credible interval is: $\left(\theta_{(\alpha/2)}, \theta_{(1-\alpha/2)}\right)$.

3.6 Example and Simulation Studies

3.6.1 Simulation Study I - A comparison of Eight Methods for $\theta = p_1 - p_2$

In this section an extensive simulation study will be done and coverage probabilities will be obtained for $\theta = p_1 - p_2$. The two Bayesian methods, when using the Jeffreys and probability matching priors, will be compared with known classical procedures. The prior distribution and posterior distribution when using the probability matching prior are given in Equations 3.3 and 3.4, respectively. The prior distribution and posterior distribution when using the Jeffreys prior are given in Equations 3.5 and 3.6, respectively, where $k = 2$.

The Bayesian credible intervals for the probability matching prior are obtained by using the weighted Monte Carlo method as described in Section 3.5. The number of X variates and the number of simulations are equal to 1 000.

Tables 3.1 and 3.2 contain coverage probabilities, mean lengths and conditional mean length ratios for the six intervals discussed by Roths & Tebbs (2006) as well as the full (real) Bayesian procedures, by using Jeffreys prior (Bayes (Jef)) and the probability matching prior (Bayes (PMP)), when $n_1 = n_2 = 10$ and $n_1 = n_2 = 20$, respectively, for a number of choices for p_1 and p_2. The nominal confidence level is $1 - \alpha = 0.95$. The conditional mean length ratio is the ratio of the mean lengths for cases where the differences are covered and when they are not covered. A small value of the conditional mean length ratio is desirable.

Table 3.1: (a) Exact coverage probabilities, (b) mean lengths, and (c) conditional mean length ratios for $n_1 = n_2 = 10$. The nominal level is 0.95. WAL, Wald; AGC, Agresti-Caffo; HAL, Haldane; JFP, Jeffreys-Perks; MLE, Beal-MLE; MOM, Beal-MOM; Bayes (Jef), Bayesian procedure using the Jeffreys prior; Bayes (PMP), Bayesian procedure using the probability matching prior.

p_1	p_2		**WAL**	**AGC**	**HAL**	**JFP**	**MLE**	**MOM**	**Bayes (Jef)**	**Bayes (PMP)**
0.1	0.1	(a)	0.95000	0.99100	0.99100	0.99100	0.99100	0.99100	0.97000	0.99100
		(b)	0.45600	0.57800	0.41900	0.52300	0.47700	0.41900	0.53600	0.56840
		(c)	0.77500	0.90000	0.72700	0.86100	0.78200	0.72700	0.93200	0.94473
0.1	0.3	(a)	0.93900	0.96800	0.94300	0.94300	0.94600	0.94300	0.95000	0.97300
		(b)	0.63200	0.65700	0.58500	0.62600	0.63600	0.58500	0.63200	0.63854
		(c)	1.34800	1.10000	1.35800	1.18300	1.42000	1.35700	1.11900	1.11850
0.1	0.5	(a)	0.91100	0.96300	0.91500	0.93000	0.94900	0.91500	0.93900	0.95400
		(b)	0.67800	0.68200	0.64000	0.65700	0.66800	0.64000	0.65300	0.65098
		(c)	1.27300	1.02800	1.16900	1.09100	1.10400	1.16500	1.11000	1.01310
0.1	0.7	(a)	0.91500	0.94500	0.94500	0.94500	0.96000	0.94500	0.95600	0.91700
		(b)	0.61900	0.65600	0.61800	0.62400	0.62700	0.61900	0.61400	0.61652
		(c)	1.74100	0.99000	0.99000	0.99000	0.97600	0.98100	0.95100	0.87559
0.1	0.9	(a)	0.87000	0.95700	0.94900	0.95700	0.95700	0.94900	0.94600	0.87000
		(b)	0.41500	0.56800	0.51600	0.51800	0.51800	0.51700	0.48900	0.52713
		(c)	9.76600	0.77500	0.73700	0.71900	0.71300	0.73800	0.70900	0.80078
0.3	0.3	(a)	0.90500	0.96300	0.96300	0.96300	0.96300	0.96300	0.90900	0.95400
		(b)	0.75600	0.72700	0.69600	0.71400	0.73900	0.69600	0.71900	0.69898
		(c)	1.11500	1.08900	1.11800	1.11600	1.15800	1.10800	1.13400	1.11780
0.3	0.5	(a)	0.92200	0.96400	0.94900	0.94900	0.95800	0.94900	0.95400	0.95700
		(b)	0.79400	0.75000	0.73400	0.74100	0.75700	0.73500	0.73900	0.71538
		(c)	1.16200	1.09000	1.11700	1.11100	1.14700	1.11600	1.13600	1.07780
0.3	0.7	(a)	0.93200	0.95500	0.94100	0.94100	0.95700	0.94100	0.93500	0.95000
		(b)	0.75400	0.72700	0.71000	0.71300	0.72000	0.71000	0.70700	0.68957
		(c)	1.37700	0.99500	1.06300	1.06000	1.11900	1.06100	1.09700	0.96325
0.5	0.5	(a)	0.91200	0.95800	0.95800	0.95800	0.95800	0.95800	0.94400	0.96100
		(b)	0.83000	0.77100	0.76400	0.76800	0.78000	0.76500	0.76300	0.73764
		(c)	1.12900	1.10400	1.13300	1.13200	1.14600	1.13400	1.11700	1.11680
Overall Mean		(a)	**0.917**	**0.963**	**0.950**	**0.953**	**0.960**	**0.950**	**0.945**	**0.948**
		(b)	**0.659**	**0.680**	**0.631**	**0.654**	**0.658**	**0.632**	**0.650**	**0.649**
		(c)	**1.211**	**1.008**	**1.046**	**1.029**	**1.063**	**1.043**	**1.034**	**1.003**

From Table 3.1 we see that the Wald interval only reaches the nominal level once, that is when $p_1 = p_2 = 0.1$. For all other values of p_1 and p_2 the coverage is below the nominal level. The Jeffreys and probability matching priors compare relatively well with the other methods, with overall average coverage of 0.945 and 0.948, respectively. Note that the overall average value for the conditional mean length ratio for the probability matching prior is equal to 1.003, which is the smallest value if compared with the other methods.

Table 3.2: (a) Exact coverage probabilities, (b) mean lengths, and (c) conditional mean length ratios for $n_1 = n_2 = 20$. The nominal level is 0.95. WAL, Wald; AGC, Agresti-Caffo; HAL, Haldane; JFP, Jeffreys-Perks; MLE, Beal-MLE; MOM, Beal-MOM; Bayes (Jef), Bayesian procedure using the Jeffreys prior; Bayes (PMP), Bayesian procedure using the probability matching prior.

p_1	p_2		**WAL**	**AGC**	**HAL**	**JFP**	**MLE**	**MOM**	**Bayes (Jef)**	**Bayes (PMP)**
0.1	0.1	(a)	0.96000	0.98800	0.96100	0.98300	0.96100	0.96100	0.94800	0.97400
		(b)	0.35200	0.39600	0.33600	0.37000	0.36700	0.33600	0.38000	0.39128
		(c)	0.92200	0.93300	0.92500	0.89000	0.96900	0.92500	1.01300	0.96505
0.1	0.3	(a)	0.93500	0.96000	0.95200	0.95200	0.95200	0.95200	0.94100	0.96000
		(b)	0.46400	0.47300	0.44600	0.46000	0.46900	0.44600	0.46200	0.46414
		(c)	1.08400	1.05400	1.10400	1.08600	1.07200	1.10400	1.05700	1.07410
0.1	0.5	(a)	0.93700	0.95400	0.94300	0.95000	0.95000	0.94300	0.95200	0.95000
		(b)	0.49600	0.49600	0.48000	0.48600	0.49200	0.48000	0.48200	0.48345
		(c)	1.09900	1.03800	1.04200	1.05300	1.04200	1.03800	1.06500	0.98793
0.1	0.7	(a)	0.91300	0.95500	0.93300	0.93300	0.93700	0.93300	0.94500	0.94200
		(b)	0.46400	0.47300	0.45700	0.45900	0.46100	0.45900	0.45700	0.45785
		(c)	1.25200	0.93200	1.02000	1.01900	1.00300	1.02300	1.02900	0.90632
0.1	0.9	(a)	0.91300	0.95800	0.94300	0.94300	0.94300	0.94300	0.94100	0.90900
		(b)	0.34200	0.39400	0.36800	0.36900	0.36900	0.36900	0.36100	0.37578
		(c)	2.41900	0.78100	0.89900	0.89900	0.89800	0.89900	0.87900	0.80140
0.3	0.3	(a)	0.93100	0.95000	0.94700	0.94700	0.95500	0.94700	0.93800	0.94800
		(b)	0.55200	0.53800	0.52800	0.53400	0.55000	0.52800	0.53400	0.52510
		(c)	1.05900	1.04000	1.05500	1.05300	1.08300	1.05500	1.06600	1.04840
0.3	0.5	(a)	0.94200	0.95200	0.94600	0.94600	0.95100	0.94600	0.94900	0.95800
		(b)	0.57900	0.55900	0.55500	0.55600	0.56500	0.55500	0.55600	0.54606
		(c)	1.07900	1.04300	1.05500	1.05400	1.06600	1.05500	1.06300	1.03000
0.3	0.7	(a)	0.92800	0.94400	0.94400	0.94400	0.94400	0.94400	0.95000	0.94800
		(b)	0.55200	0.53800	0.53300	0.53300	0.53600	0.53300	0.52900	0.52428
		(c)	1.15400	1.03300	1.03900	1.03900	1.03700	1.03900	1.01900	1.03680
0.5	0.5	(a)	0.91900	0.95700	0.95700	0.95700	0.96100	0.95700	0.93800	0.95600
		(b)	0.60400	0.57800	0.57800	0.57800	0.58300	0.57800	0.57700	0.56406
		(c)	1.05600	1.06000	1.06800	1.06800	1.07200	1.06800	1.07300	1.06860
Overall Mean		(a)	**0.931**	**0.958**	**0.947**	**0.951**	**0.950**	**0.947**	**0.945**	**0.949**
		(b)	**0.489**	**0.494**	**0.476**	**0.483**	**0.488**	**0.476**	**0.482**	**0.481**
		(c)	**1.236**	**0.990**	**1.023**	**1.018**	**1.027**	**1.023**	**1.029**	**0.991**

From Table 3.2 we see that the Wald interval never reaches the nominal level, except when $p_1 = p_2 = 0.1$ the coverage probability obtained is equal to 0.96. For all other values of p_1 and p_2 the coverage is below the nominal level. The Jeffreys and probability matching priors compare relatively well with the other methods, with overall average coverage of 0.945 and 0.949, respectively. Note that the overall average value for the conditional mean length ratio for the probability matching prior is equal to 0.991, which is just larger than the smallest value of 0.990 obtained from the Agresti-Caffo

interval.

The differences among all the intervals are not too large, the only exception to this is the Wald interval. The results from the Bayesian procedures compare well with the other methods.

3.6.2 Simulation Study II - A comparison of the Jeffreys, Uniform and Probability Matching priors for $\theta_1 = p_1 - p_2$

In this section a more extensive simulation study is done and coverage probabilities are obtained for $\theta_1 = p_1 - p_2$, the difference between two binomial proportions. For comparison purposes the following priors will be used:

1. the Jeffreys prior: $\pi_J\left(\underline{p}\right) \propto \prod_{i=1}^{2} p_i^{-\frac{1}{2}} (1-p_i)^{-\frac{1}{2}}$;
2. the uniform prior: $\pi_U\left(\underline{p}\right) \propto$ constant;
3. the probability matching prior: $\pi_{PM}\left(\underline{p}\right) \propto \left\{\sum_{i=1}^{2} p_i(1-p_i)\right\}^{\frac{1}{2}} \prod_{i=1}^{2} p_i^{-1}(1-p_i)^{-1}$.

The parameter values for the binomial distribution are $n_1 = n_2 = 10$, $n_1 = n_2 = 20$, $n_1 = n_2 = 30$ and $p_i = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ (for $i = 1,2$). The average length and standard deviation of the intervals are also given. The number of X variates and the number of simulations are equal to 1 000. The average length and standard deviation of the intervals are calculated using the following formulas, where I_i is the interval length and n^* the number of intervals:

$$\text{average length} = \frac{1}{n^*}\sum_{i=1}^{n^*} I_i$$

and

$$\text{standard deviation} = \sqrt{\frac{1}{n^*-1}\sum_{i=1}^{n^*}\left(I_i - \bar{I}\right)^2}.$$

Tables 3.3 to 3.5 contain coverage probabilities, mean lengths and standard deviations for the Jeffreys, uniform and probability matching priors when $n_1 = n_2 = 10$, for several choices of p_1 and p_2. These results will be summarised in Figures 3.1 to 3.3. As in Chapter 2, MATLAB® was used to construct the box plots in this chapter. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points the algorithm considers not to be outliers, and the outliers are plotted individually.

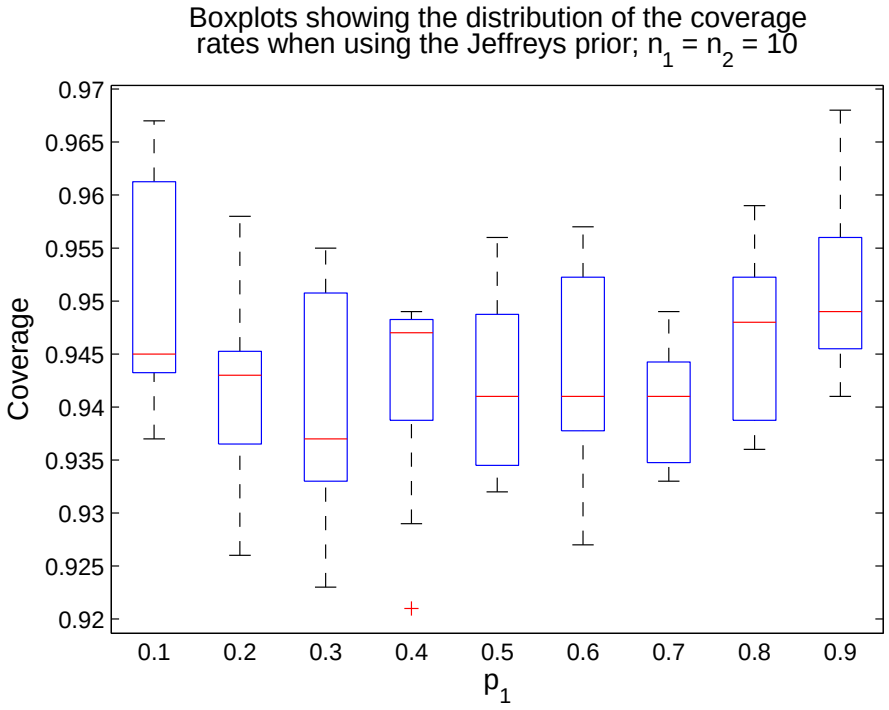


Figure 3.1: Box plot summarising the coverage rates of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ for $p_1 = 0.1 : 0.1 : 0.9$, using the Jeffreys prior; $n_1 = n_2 = 10$.

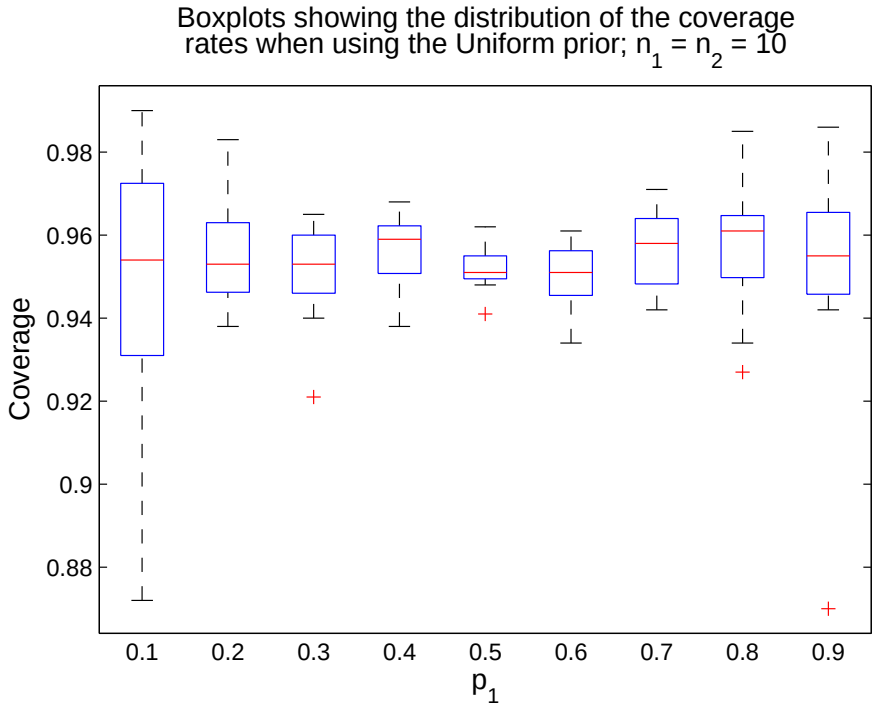


Figure 3.2: Box plot summarising the coverage rates of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ for $p_1 = 0.1 : 0.1 : 0.9$, using the uniform prior; $n_1 = n_2 = 10$.

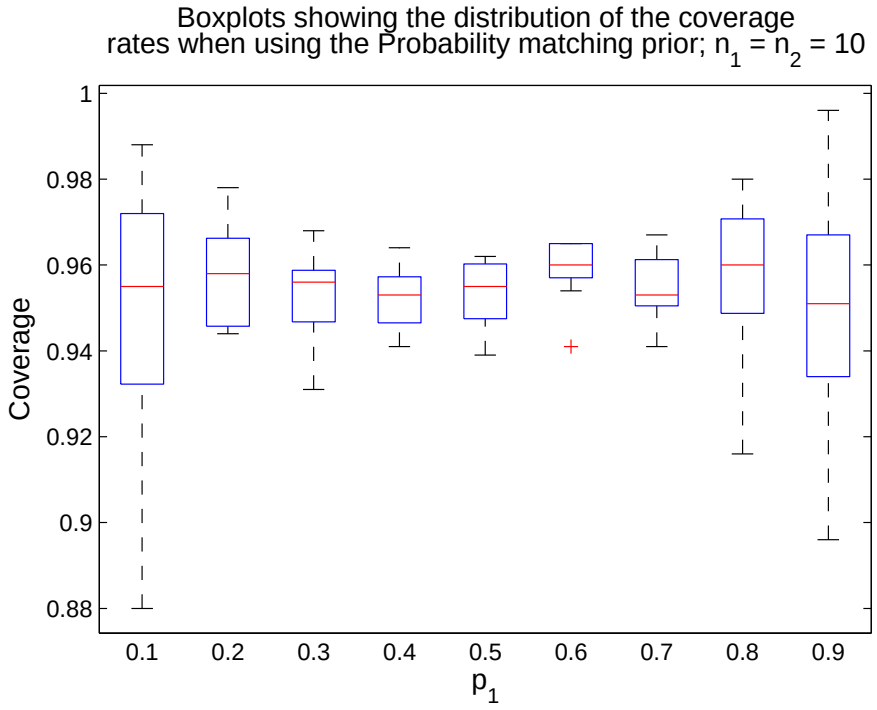


Figure 3.3: Box plot summarising the coverage rates of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ for $p_1 = 0.1 : 0.1 : 0.9$, using the probability matching prior; $n_1 = n_2 = 10$.

From Figure 3.1 we see that when the Jeffreys prior is used, the median coverage probability is below 0.95 for all values of p_1. When $p_1 = 0.5$ and $p_1 = 0.6$ the median coverage probability is just above 0.94. From Figure 3.2 we see that when the uniform prior is used, the median coverage probability is at or above 0.95 for all values of p_1. When $p_1 = 0.5$ and $p_1 = 0.6$ the median coverage probability is equal to 0.95. From Figure 3.3 we see that when the probability matching prior is used, the median coverage probability is at or above 0.95 for all values of p_1. When $p_1 = 0.5$ the median coverage probability is just below 0.96 and when $p_1 = 0.6$ the median coverage probability is equal to 0.96. From these figures it seems that the performance of the uniform and probability matching priors are very similar, where there is over coverage most of the time. With the Jeffreys prior it seems that there is under coverage most of the time.

Table 3.3: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the Jeffreys prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 10$.

	Jeffreys prior $n_1 = n_2 = 10$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9670	0.9450	0.9550	0.9470	0.9330	0.9460	0.9490	0.9480	0.9560	0.9496
(b)	0.5415	0.5938	0.6326	0.6502	0.6556	0.6393	0.6055	0.5669	0.4963	0.5980
(c)	0.9368	1.0399	1.1387	1.0696	1.0763	1.0859	0.9883	0.8115	0.7016	0.9832
0.2 (a)	0.9600	0.9370	0.9270	0.9290	0.9320	0.9270	0.9410	0.9590	0.9410	0.9392
(b)	0.5926	0.6451	0.6846	0.6969	0.7027	0.6925	0.6688	0.6230	0.5695	0.6528
(c)	1.0557	1.0933	1.1157	1.1097	1.0698	1.0907	0.9995	0.9142	0.8829	1.0368
0.3 (a)	0.9410	0.9460	0.9370	0.9440	0.9350	0.9370	0.9350	0.9360	0.9560	0.9408
(b)	0.6339	0.6803	0.7168	0.7350	0.7393	0.7280	0.7055	0.6706	0.6148	0.6916
(c)	1.1290	1.0902	1.1193	1.1231	1.0969	1.1200	1.0863	1.0515	0.9718	1.0876
0.4 (a)	0.9370	0.9440	0.9500	0.9420	0.9540	0.9570	0.9340	0.9390	0.9460	0.9448
(b)	0.6500	0.7008	0.7351	0.7544	0.7587	0.7494	0.7273	0.6925	0.6403	0.7121
(c)	1.1079	1.0963	1.1426	1.1500	1.1153	1.1457	1.0888	1.1126	1.0571	1.1129
0.5 (a)	0.9440	0.9430	0.9530	0.9480	0.9560	0.9380	0.9440	0.9380	0.9440	0.9453
(b)	0.6536	0.7068	0.7380	0.7580	0.7634	0.7597	0.7377	0.7055	0.6539	0.7196
(c)	1.1347	1.1006	1.1233	1.1078	1.1342	1.1410	1.0968	1.1343	1.0999	1.1192
0.6 (a)	0.9440	0.9350	0.9370	0.9480	0.9470	0.9410	0.9330	0.9510	0.9510	0.9430
(b)	0.6440	0.6962	0.7311	0.7470	0.7586	0.7547	0.7351	0.6964	0.6499	0.7126
(c)	1.0200	1.0627	1.1391	1.0958	1.1403	1.1444	1.1454	1.1007	1.0567	1.1006
0.7 (a)	0.9450	0.9260	0.9350	0.9490	0.9450	0.9510	0.9440	0.9530	0.9460	0.9438
(b)	0.6168	0.6726	0.7067	0.7281	0.7389	0.7388	0.7173	0.6803	0.6339	0.6926
(c)	0.9411	0.9973	1.0336	1.1030	1.1130	1.1365	1.1301	1.0928	1.1476	1.0772
0.8 (a)	0.9500	0.9580	0.9230	0.9210	0.9410	0.9410	0.9450	0.9400	0.9680	0.9430
(b)	0.5643	0.6231	0.6668	0.6963	0.7032	0.6972	0.6848	0.6416	0.5922	0.6522
(c)	0.8503	1.0873	1.0342	1.0193	1.0984	1.1307	1.1151	1.0598	1.0159	1.0457
0.9 (a)	0.9650	0.9430	0.9490	0.9490	0.9410	0.9560	0.9370	0.9520	0.9490	0.9490
(b)	0.4946	0.5617	0.6169	0.6444	0.6539	0.6511	0.6298	0.6004	0.5366	0.5988
(c)	0.7225	0.8060	0.9725	1.1036	1.0818	1.1244	1.1535	1.0474	0.9304	0.9936
mean (a)	0.9503	0.9419	0.9407	0.9419	0.9427	0.9438	0.9402	0.9462	0.9508	**0.9443**
(b)	0.5990	0.6534	0.6921	0.7123	0.7194	0.7123	0.6902	0.6530	0.5986	**0.6700**
(c)	0.9887	1.0415	1.0910	1.0980	1.1029	1.1244	1.0893	1.0361	0.9849	**1.0619**

From Table 3.3 the overall average for the coverage probabilities is equal to 0.9443, which is below the nominal level of 0.95. When $p_1 = 0.1$ and $p_2 = 0.9$ the coverage rate is equal to 0.9650, which is higher than the nominal level, and when looking at $p_1 = 0.4$ and $p_2 = 0.8$ the coverage rate is equal to 0.9210, which is much lower than the nominal level. When using the Jeffreys prior, for $\theta_1 = p_1 - p_2$, we have under coverage most of the time (60 out of the 81 values are below 0.95).

Table 3.4: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the uniform Prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 10$.

	Uniform prior $n_1 = n_2 = 10$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9900	0.9830	0.9590	0.9620	0.9580	0.9340	0.9420	0.9270	0.8700	0.9472
(b)	0.5693	0.6090	0.6374	0.6486	0.6514	0.6415	0.6179	0.5843	0.5247	0.6093
(c)	0.9331	1.0296	1.0799	1.0573	0.9973	0.9560	0.8905	0.8349	0.7980	0.9529
0.2 (a)	0.9830	0.9660	0.9650	0.9680	0.9540	0.9460	0.9580	0.9340	0.9470	0.9579
(b)	0.6113	0.6460	0.6728	0.6873	0.6890	0.6818	0.6568	0.6269	0.5815	0.6504
(c)	1.0224	1.0461	1.0888	1.0861	1.0240	0.9754	0.9406	0.8735	0.8246	0.9868
0.3 (a)	0.9690	0.9480	0.9530	0.9620	0.9410	0.9510	0.9460	0.9560	0.9490	0.9528
(b)	0.6372	0.6753	0.6994	0.7153	0.7161	0.7134	0.6939	0.6577	0.6196	0.6809
(c)	1.0678	1.0700	1.0941	1.0981	1.0738	1.0462	0.9661	0.9199	0.8842	1.0245
0.4 (a)	0.9540	0.9620	0.9630	0.9630	0.9480	0.9570	0.9490	0.9610	0.9420	0.9554
(b)	0.6515	0.6894	0.7151	0.7297	0.7334	0.7271	0.7075	0.6801	0.6438	0.6975
(c)	1.1109	1.0712	1.0911	1.1128	1.1065	1.0802	1.0240	0.9605	0.9755	1.0592
0.5 (a)	0.9590	0.9580	0.9500	0.9520	0.9500	0.9560	0.9710	0.9640	0.9550	0.9572
(b)	0.6516	0.6917	0.7184	0.7339	0.7400	0.7345	0.7189	0.6888	0.6508	0.7032
(c)	1.0000	1.0559	1.0845	1.1021	1.1131	1.1106	1.0918	1.0435	0.9892	1.0657
0.6 (a)	0.9400	0.9530	0.9480	0.9590	0.9500	0.9480	0.9530	0.9550	0.9590	0.9517
(b)	0.6417	0.6796	0.7105	0.7271	0.7339	0.7286	0.7159	0.6886	0.6515	0.6975
(c)	0.9571	0.9915	1.0415	1.1303	1.1197	1.1069	1.1046	1.0470	1.0633	1.0624
0.7 (a)	0.9350	0.9530	0.9400	0.9470	0.9620	0.9610	0.9620	0.9630	0.9610	0.9538
(b)	0.6180	0.6608	0.6901	0.7126	0.7182	0.7154	0.6982	0.6761	0.6372	0.6807
(c)	0.8833	0.9079	0.9615	1.0321	1.0708	1.0889	1.0741	1.0691	1.0717	1.0177
0.8 (a)	0.9190	0.9380	0.9580	0.9570	0.9510	0.9440	0.9670	0.9670	0.9790	0.9533
(b)	0.5819	0.6280	0.6626	0.6825	0.6897	0.6875	0.6723	0.6493	0.6068	0.6512
(c)	0.8284	0.8571	0.9447	0.9782	1.0005	1.0599	1.0719	1.0477	1.0397	0.9809
0.9 (a)	0.8720	0.9410	0.9210	0.9380	0.9540	0.9540	0.9630	0.9850	0.9860	0.9460
(b)	0.5298	0.5844	0.6181	0.6411	0.6535	0.6497	0.6386	0.6116	0.5719	0.6110
(c)	0.7985	0.8364	0.8866	0.9696	1.0444	1.0839	1.0636	1.0431	0.9463	0.9636
mean (a)	0.9468	0.9558	0.9508	0.9564	0.9520	0.9501	0.9568	0.9569	0.9498	**0.9528**
(b)	0.6103	0.6516	0.6805	0.6976	0.7028	0.6977	0.6800	0.6515	0.6098	**0.6646**
(c)	0.9557	0.9851	1.0303	1.0630	1.0611	1.0564	1.0252	0.9821	0.9547	**1.0126**

From Table 3.4 the overall average for the coverage probabilities is equal to 0.9528, which is slightly above the nominal level of 0.95. When $p_1 = 0.1$ and $p_2 = 0.9$ the coverage rate is equal to 0.8720, which is way below the nominal level, and when looking at $p_1 = 0.4$ and $p_2 = 0.8$ the coverage rate is equal to 0.9570, which is almost equal to the nominal level. When using the uniform prior, for $\theta_1 = p_1 - p_2$, we have over coverage when p_1 and p_2 are both small.

Table 3.5: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the probability matching prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 10$.

	Probability matching prior $n_1 = n_2 = 10$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9880	0.9780	0.9570	0.9640	0.9610	0.9610	0.9410	0.9160	0.8960	0.9513
(b)	0.5681	0.6059	0.6368	0.6511	0.6526	0.6424	0.6162	0.5761	0.5264	0.6084
(c)	0.9302	1.0146	1.0357	1.0538	1.0440	0.9885	0.8893	0.8159	0.8028	0.9527
0.2 (a)	0.9720	0.9730	0.9640	0.9530	0.9620	0.9650	0.9460	0.9390	0.9280	0.9558
(b)	0.6108	0.6488	0.6742	0.6859	0.6873	0.6796	0.6596	0.6244	0.5782	0.6499
(c)	1.0279	1.0389	1.0699	1.0483	1.0176	0.9749	0.9142	0.8598	0.8279	0.9755
0.3 (a)	0.9720	0.9640	0.9680	0.9610	0.9480	0.9410	0.9520	0.9600	0.9360	0.9558
(b)	0.6363	0.6733	0.6985	0.7157	0.7191	0.7081	0.6911	0.6570	0.6183	0.6797
(c)	1.0720	1.0656	1.0834	1.1163	1.0738	1.0270	0.9869	0.9399	0.8846	1.0277
0.4 (a)	0.9640	0.9550	0.9480	0.9520	0.9560	0.9580	0.9530	0.9520	0.9510	0.9543
(b)	0.6514	0.6860	0.7150	0.7287	0.7348	0.7280	0.7100	0.6823	0.6418	0.6975
(c)	1.0392	1.0708	1.0914	1.1063	1.1060	1.0919	1.0443	0.9901	0.9649	1.0561
0.5 (a)	0.9550	0.9610	0.9560	0.9480	0.9600	0.9580	0.9530	0.9570	0.9510	0.9554
(b)	0.6535	0.6886	0.7137	0.7358	0.7407	0.7339	0.7172	0.6895	0.6481	0.7023
(c)	1.0330	1.0088	1.0427	1.1077	1.1096	1.1102	1.0603	1.0764	1.0301	1.0643
0.6 (a)	0.9530	0.9580	0.9510	0.9560	0.9480	0.9650	0.9540	0.9630	0.9590	0.9563
(b)	0.6428	0.6834	0.7085	0.7289	0.7350	0.7309	0.7141	0.6882	0.6527	0.6983
(c)	1.0047	0.9945	1.0555	1.0739	1.1195	1.0891	1.1026	1.0662	1.0403	1.0607
0.7 (a)	0.9370	0.9460	0.9430	0.9420	0.9460	0.9540	0.9620	0.9700	0.9620	0.9513
(b)	0.6171	0.6624	0.6904	0.7100	0.7184	0.7154	0.7007	0.6751	0.6355	0.6806
(c)	0.8882	0.9252	1.0032	1.0333	1.0895	1.1240	1.1103	1.1032	1.1009	1.0420
0.8 (a)	0.9180	0.9450	0.9570	0.9410	0.9550	0.9650	0.9610	0.9730	0.9820	0.9552
(b)	0.5872	0.6276	0.6601	0.6784	0.6904	0.6893	0.6767	0.6510	0.6097	0.6523
(c)	0.8377	0.8562	0.9251	0.9796	1.0134	1.0793	1.0951	1.0636	1.0081	0.9842
0.9 (a)	0.8800	0.9440	0.9310	0.9560	0.9390	0.9600	0.9670	0.9800	0.9960	0.9503
(b)	0.5279	0.5806	0.6176	0.6421	0.6550	0.6499	0.6369	0.6053	0.5739	0.6099
(c)	0.7974	0.8363	0.8789	0.9664	0.9887	1.0489	1.0843	0.9935	0.9611	0.9506
mean (a)	0.9488	0.9582	0.9528	0.9526	0.9528	0.9586	0.9543	0.9567	0.9512	**0.9540**
(b)	0.6106	0.6507	0.6794	0.6974	0.7037	0.6975	0.6803	0.6499	0.6094	**0.6643**
(c)	0.9589	0.9790	1.0207	1.0540	1.0624	1.0593	1.0319	0.9898	0.9579	**1.0127**

From Table 3.5 the overall average for the coverage probabilities is equal to 0.9540, which is slightly above the nominal level of 0.95 and a bit higher than the overall average when using the uniform prior. When $p_1 = 0.1$ and $p_2 = 0.9$ the coverage rate is equal to 0.8800, which is way below the nominal level, this value is almost the same as the one obtained from the uniform prior. When looking at $p_1 = 0.4$ and $p_2 = 0.8$ the coverage rate is equal to 0.9410, which is below the nominal level of 0.95. When using the probability matching prior, for $\theta_1 = p_1 - p_2$, we have over coverage when p_1 and p_2 are both small.

In general, we can conclude that, when $n_1 = n_2 = 10$ and for different values of p_1 and p_2 the Jeffreys prior produce coverage rates below the nominal level and that the uniform and probability matching priors produce coverage rates above the nominal level, more so for the probability matching prior. Where the overall average coverage rate for the Jeffreys prior is equal to 0.9443, the overall average coverage rate for the uniform prior is equal to 0.9528 and the overall average coverage rate for the probability matching prior is equal to 0.9540. The average interval lengths of the Jeffreys prior are also generally larger than that of the uniform and probability matching priors. The uniform and probability matching priors give smaller standard deviation values for the interval lengths than the Jeffreys prior.

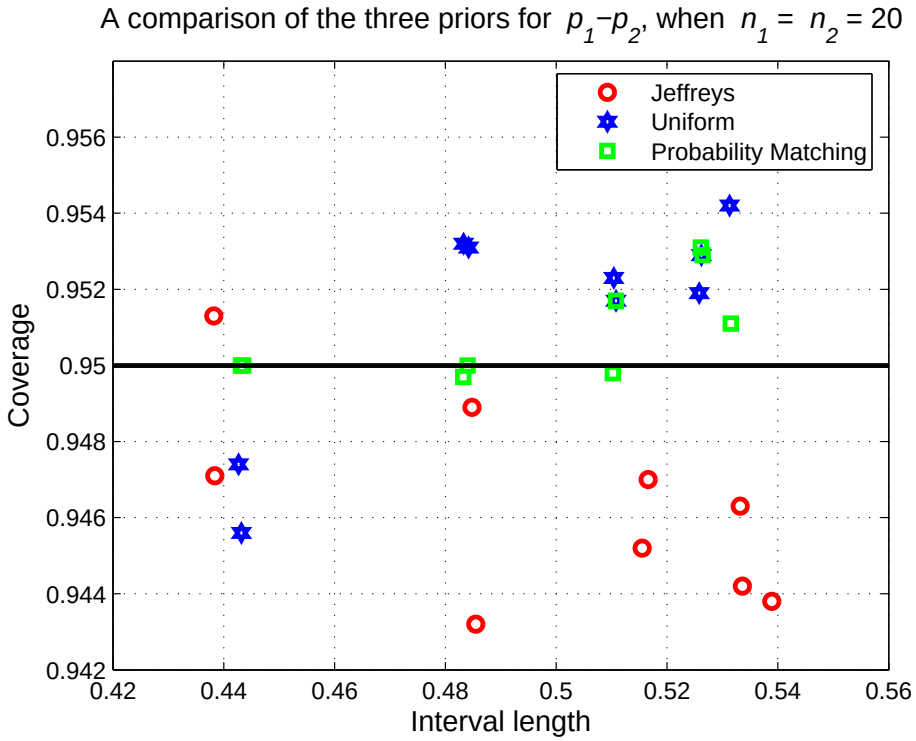


Figure 3.4: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ against the interval length for the three priors when $n_1 = n_2 = 20$.

Scatter plots are given in Figure 3.4 of the coverage rates obtained when using the Jeffreys, uniform and probability matching priors for $\theta_1 = p_1 - p_2$ when $n_1 = n_2 = 20$. The values plotted are averages over all possible values for p_1 and p_2. The Jeffreys prior underestimates the coverage probabilities while the uniform prior and the probability matching prior tends to overestimate the coverage probabilities. The actual simulation results are given in Appendix B in Tables B.1, B.2 and B.3.

Tables 3.6 to 3.8 contain coverage probabilities, mean lengths and standard deviations for the Jeffreys, uniform and probability matching priors when $n_1 = n_2 = 30$, for several choices of p_1 and p_2.

Table 3.6: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the Jeffreys prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 30$.

	Jeffreys prior $n_1 = n_2 = 30$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9480	0.9660	0.9540	0.9490	0.9560	0.9540	0.9520	0.9620	0.9530	0.9549
(b)	0.3058	0.3531	0.3811	0.3965	0.4039	0.3952	0.3779	0.3441	0.2977	0.3617
(c)	1.0454	1.0559	1.0389	1.0498	1.0228	1.0145	0.9989	0.9586	0.9429	1.0142
0.2 (a)	0.9530	0.9440	0.9270	0.9470	0.9560	0.9490	0.9470	0.9440	0.9400	0.9452
(b)	0.3552	0.3918	0.4203	0.4351	0.4383	0.4335	0.4160	0.3889	0.3467	0.4029
(c)	1.1206	1.0211	1.0339	1.0364	1.0151	1.0357	1.0207	0.9635	1.0566	1.0337
0.3 (a)	0.9500	0.9390	0.9520	0.9530	0.9500	0.9390	0.9490	0.9570	0.9450	0.9482
(b)	0.3820	0.4181	0.4443	0.4590	0.4636	0.4584	0.4428	0.4199	0.3761	0.4293
(c)	1.0687	1.0454	1.0437	1.0540	1.0293	1.0368	1.0631	1.0079	1.0096	1.0398
0.4 (a)	0.9530	0.9540	0.9520	0.9450	0.9460	0.9440	0.9330	0.9490	0.9460	0.9469
(b)	0.3995	0.4356	0.4593	0.4734	0.4773	0.4726	0.4574	0.4322	0.3966	0.4449
(c)	1.0802	1.0449	1.0398	1.0477	1.0299	1.0301	1.0194	1.0248	1.0101	1.0363
0.5 (a)	0.9360	0.9380	0.9470	0.9490	0.9420	0.9500	0.9570	0.9350	0.9510	0.9450
(b)	0.4020	0.4387	0.4639	0.4773	0.4808	0.4769	0.4618	0.4382	0.4028	0.4492
(c)	1.0378	1.0503	1.0445	1.0456	1.0449	1.0478	1.0521	1.0297	1.0300	1.0425
0.6 (a)	0.9560	0.9470	0.9550	0.9380	0.9450	0.9480	0.9460	0.9370	0.9520	0.9471
(b)	0.3967	0.4324	0.4568	0.4722	0.4768	0.4720	0.4591	0.4345	0.3988	0.4444
(c)	1.0140	1.0574	1.0280	1.0392	1.0352	1.0455	1.0568	1.0308	1.0421	1.0388
0.7 (a)	0.9410	0.9400	0.9410	0.9490	0.9510	0.9540	0.9650	0.9500	0.9360	0.9474
(b)	0.3783	0.4179	0.4422	0.4571	0.4643	0.4593	0.4427	0.4204	0.3823	0.4294
(c)	1.0063	1.0198	1.0360	1.0248	1.0383	1.0363	1.0334	1.0749	1.0435	1.0348
0.8 (a)	0.9570	0.9560	0.9410	0.9210	0.9450	0.9460	0.9520	0.9460	0.9410	0.9450
(b)	0.3474	0.3901	0.4183	0.4337	0.4381	0.4344	0.4190	0.3941	0.3524	0.4031
(c)	1.0137	1.0567	1.0497	1.0385	1.0343	1.0266	1.0228	1.0578	1.0633	1.0404
0.9 (a)	0.9530	0.9600	0.9510	0.9400	0.9520	0.9520	0.9400	0.9410	0.9350	0.9471
(b)	0.2969	0.3465	0.3781	0.3966	0.4037	0.3996	0.3830	0.3526	0.3074	0.3627
(c)	0.8401	0.9393	1.0102	1.0373	1.0221	1.0377	1.0349	1.0367	1.0519	1.0011
mean (a)	0.9497	0.9493	0.9467	0.9434	0.9492	0.9484	0.9490	0.9468	0.9443	**0.9474**
(b)	0.3626	0.4027	0.4294	0.4445	0.4496	0.4447	0.4289	0.4028	0.3623	**0.4142**
(c)	1.0252	1.0323	1.0361	1.0415	1.0302	1.0346	1.0336	1.0205	1.0278	**1.0313**

From Table 3.6 the overall average for the coverage probabilities is equal to 0.9474, which is below the nominal level of 0.95. When $p_1 = 0.1$ and $p_2 = 0.3$ the coverage rate is equal to 0.9500, which is equal to the nominal level, and when looking at $p_1 = 0.4$ and $p_2 = 0.8$ the coverage rate is equal to 0.9210, which is much lower than the nominal level and exactly the same as the value obtained for this case when $n_1 = n_2 = 10$.

Table 3.7: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the uniform prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 30$.

	Uniform prior $n_1 = n_2 = 30$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9540	0.9670	0.9640	0.9550	0.9640	0.9350	0.9490	0.9390	0.9240	0.9501
(b)	0.3174	0.3572	0.3840	0.3980	0.4004	0.3966	0.3788	0.3480	0.3052	0.3651
(c)	0.9941	1.0171	1.0599	1.0263	1.0269	0.9997	0.9512	0.8940	0.8219	0.9768
0.2 (a)	0.9560	0.9560	0.9630	0.9520	0.9560	0.9500	0.9580	0.9570	0.9530	0.9557
(b)	0.3583	0.3940	0.4181	0.4305	0.4349	0.4298	0.4138	0.3889	0.3487	0.4019
(c)	1.0201	1.0444	1.0400	1.0234	1.0345	1.0283	0.9836	0.9854	0.8938	1.0059
0.3 (a)	0.9490	0.9670	0.9490	0.9520	0.9460	0.9480	0.9670	0.9440	0.9420	0.9516
(b)	0.3846	0.4188	0.4415	0.4526	0.4570	0.4522	0.4395	0.4144	0.3788	0.4266
(c)	1.0337	1.0323	1.0429	1.0583	1.0408	1.0191	1.0590	0.9828	0.9701	1.0266
0.4 (a)	0.9570	0.9500	0.9480	0.9420	0.9380	0.9450	0.9530	0.9530	0.9460	0.9480
(b)	0.3984	0.4306	0.4528	0.4658	0.4701	0.4666	0.4542	0.4287	0.3961	0.4404
(c)	1.0154	1.0152	1.0347	1.0385	1.0386	1.0473	1.0106	0.9856	0.9916	1.0197
0.5 (a)	0.9620	0.9510	0.9500	0.9570	0.9550	0.9580	0.9420	0.9520	0.9420	0.9521
(b)	0.4035	0.4354	0.4569	0.4704	0.4744	0.4705	0.4569	0.4344	0.4026	0.4450
(c)	1.0162	1.0233	1.0228	1.0395	1.0376	1.0346	1.0344	1.0169	1.0113	1.0263
0.6 (a)	0.9350	0.9490	0.9550	0.9470	0.9440	0.9620	0.9480	0.9420	0.9540	0.9484
(b)	0.3952	0.4292	0.4524	0.4659	0.4702	0.4661	0.4529	0.4321	0.3962	0.4400
(c)	0.9964	0.9837	0.9992	1.0288	1.0362	1.0433	1.0324	1.0466	1.0326	1.0221
0.7 (a)	0.9470	0.9400	0.9480	0.9470	0.9410	0.9550	0.9500	0.9510	0.9540	0.9481
(b)	0.3792	0.4147	0.4393	0.4528	0.4573	0.4547	0.4410	0.4194	0.3824	0.4268
(c)	0.9218	0.9819	1.0014	1.0260	1.0383	1.0484	1.0478	1.0602	1.0450	1.0190
0.8 (a)	0.9340	0.9440	0.9650	0.9430	0.9590	0.9460	0.9500	0.9370	0.9450	0.9470
(b)	0.3484	0.3873	0.4133	0.4304	0.4354	0.4308	0.4182	0.3939	0.3595	0.4019
(c)	0.9080	0.9294	1.0159	0.9864	1.0418	1.0367	1.0505	1.0412	1.0545	1.0072
0.9 (a)	0.9480	0.9300	0.9370	0.9430	0.9500	0.9510	0.9540	0.9520	0.9650	0.9478
(b)	0.3040	0.3505	0.3773	0.3954	0.4017	0.3983	0.3838	0.3558	0.3205	0.3653
(c)	0.7935	0.8981	0.9474	1.0129	1.0005	1.0388	1.0705	1.0527	1.0002	0.9794
mean (a)	0.9491	0.9504	0.9532	0.9487	0.9503	0.9500	0.9523	0.9474	0.9472	**0.9499**
(b)	0.3654	0.4020	0.4262	0.4402	0.4446	0.4406	0.4266	0.4017	0.3656	**0.4125**
(c)	0.9666	0.9917	1.0183	1.0267	1.0328	1.0329	1.0267	1.0073	0.9801	**1.0092**

From Table 3.7 the overall average for the coverage probabilities is equal to 0.9499, which is just below the nominal level of 0.95. When $p_1 = 0.1$ and $p_2 = 0.3$ the coverage rate is equal to 0.9490, which is below the nominal level, where the Jeffreys prior obtained a coverage value equal to the nominal level for this case. When looking at $p_1 = 0.4$ and $p_2 = 0.8$ the coverage rate is equal to 0.9430, which is lower than the nominal level and lower than the value obtained for this case when $n_1 = n_2 = 10$.

Table 3.8: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the probability matching prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 30$.

	Probability matching prior $n_1 = n_2 = 30$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9630	0.9590	0.9470	0.9460	0.9590	0.9490	0.9430	0.9380	0.9440	0.9498
(b)	0.3197	0.3574	0.3825	0.3984	0.4010	0.3955	0.3792	0.3501	0.3049	0.3654
(c)	1.0060	1.0377	1.0308	1.0218	1.0058	0.9920	0.9288	0.8810	0.8212	0.9695
0.2 (a)	0.9550	0.9630	0.9560	0.9540	0.9530	0.9500	0.9590	0.9460	0.9370	0.9526
(b)	0.3588	0.3942	0.4176	0.4303	0.4360	0.4294	0.4140	0.3878	0.3499	0.4020
(c)	1.0431	1.0507	1.0553	1.0489	1.0287	1.0043	0.9865	0.9461	0.8762	1.0044
0.3 (a)	0.9480	0.9540	0.9490	0.9490	0.9530	0.9510	0.9470	0.9370	0.9580	0.9496
(b)	0.3842	0.4193	0.4414	0.4531	0.4576	0.4510	0.4378	0.4140	0.3787	0.4263
(c)	1.0551	1.0441	1.0567	1.0456	1.0393	1.0158	0.9925	0.9771	0.9282	1.0172
0.4 (a)	0.9540	0.9610	0.9580	0.9440	0.9520	0.9390	0.9510	0.9650	0.9570	0.9534
(b)	0.3962	0.4318	0.4543	0.4662	0.4694	0.4663	0.4528	0.4297	0.3953	0.4402
(c)	1.0502	1.0244	1.0330	1.0432	1.0335	1.0550	1.0445	1.0008	0.9791	1.0293
0.5 (a)	0.9410	0.9490	0.9610	0.9540	0.9500	0.9600	0.9380	0.9470	0.9550	0.9506
(b)	0.4035	0.4348	0.4573	0.4707	0.4741	0.4704	0.4579	0.4354	0.4018	0.4451
(c)	1.0265	1.0098	1.0264	1.0513	1.0398	1.0392	1.0379	1.0379	1.0203	1.0321
0.6 (a)	0.9500	0.9350	0.9480	0.9480	0.9540	0.9500	0.9380	0.9460	0.9460	0.9461
(b)	0.3973	0.4287	0.4526	0.4669	0.4714	0.4661	0.4535	0.4313	0.3963	0.4405
(c)	1.0090	0.9957	1.0466	1.0402	1.0429	1.0408	1.0408	1.0412	1.0431	1.0334
0.7 (a)	0.9440	0.9510	0.9580	0.9510	0.9580	0.9570	0.9430	0.9580	0.9710	0.9546
(b)	0.3786	0.4133	0.4361	0.4533	0.4582	0.4521	0.4407	0.4182	0.3845	0.4261
(c)	0.9329	0.9496	1.0117	1.0246	1.0316	1.0391	1.0348	1.0424	1.0354	1.0113
0.8 (a)	0.9330	0.9490	0.9490	0.9470	0.9550	0.9550	0.9450	0.9450	0.9590	0.9486
(b)	0.3516	0.3898	0.4161	0.4298	0.4340	0.4321	0.4182	0.3940	0.3572	0.4025
(c)	0.8936	0.9594	0.9613	1.0085	0.9878	1.0389	1.0411	1.0387	1.0576	0.9985
0.9 (a)	0.9260	0.9310	0.9460	0.9420	0.9410	0.9460	0.9500	0.9530	0.9550	0.9433
(b)	0.3030	0.3483	0.3802	0.3959	0.4011	0.3967	0.3830	0.3579	0.3159	0.3647
(c)	0.7942	0.9001	0.9737	0.9978	1.0304	1.0285	1.0368	1.0442	1.0239	0.9811
mean (a)	0.9460	0.9502	0.9524	0.9483	0.9528	0.9508	0.9460	0.9483	0.9536	**0.9498**
(b)	0.3659	0.4020	0.4265	0.4405	0.4448	0.4399	0.4263	0.4020	0.3649	**0.4125**
(c)	0.9790	0.9968	1.0217	1.0313	1.0266	1.0282	1.0160	1.0010	0.9761	**1.0085**

From Table 3.8 the overall average for the coverage probabilities is equal to 0.9498, which is just below the nominal level of 0.95 and almost equal to the overall average when the uniform prior was used. When $p_1 = 0.1$ and $p_2 = 0.3$ the coverage rate is equal to 0.9480, which is below the nominal level, where the Jeffreys prior obtained a coverage value equal to the nominal level for this case. When looking at $p_1 = 0.4$ and $p_2 = 0.8$ the coverage rate is equal to 0.9470, which is lower than the nominal level and a bit higher than the value obtained for this case when $n_1 = n_2 = 10$.

3.6.3 Example - Mal de Rio Cuarto Virus

In this section the adaptability of the intervals are shown in situations where data are collected in multiple stages. To illustrate this consider the data from Ornaghi et al. (1999) given in Table 3.9. The stages correspond to different dates on which insects were collected during maize planting season in Argentina (from October to November). The goal of this experiment was to assess if male and female insects transmit the Mal de Rio Cuarto virus to susceptible maize plants at similar rates.

Table 3.9: Number of test plants and numbers of virus-infected plants collected on five different dates during the maize plant season.

Stage	Gender	Number of test plants	Number infected
1	M	29	9
	F	31	5
2	M	57	4
	F	57	7
3	M	57	8
	F	57	16
4	M	24	2
	F	24	3
5	M	24	3
	F	24	2

Assume that, at a specific stage, the researchers want to estimate the difference $p_1 - p_2$, where p_1 is equal to the proportion infected plants for male insects and p_2 is the proportion infected plants for female insects. In Table 3.10 the 95% confidence intervals for $\theta_1 = p_1 - p_2$ are given for the six methods described by Roths & Tebbs (2006) and in Table 3.11 the 95% Bayesian credible intervals are given for the Jeffreys, uniform and probability matching priors.

Table 3.10: 95% confidence intervals for the difference in disease transmission probabilities among male and female insects.

Stage	Interval	Lower limit	Upper limit	Length
1	WAL	-0.063	0.361	0.425
	AGC	-0.070	0.351	0.421
	HAL	-0.065	0.347	0.412
	JFP	-0.067	0.350	0.417
	MLE	-0.072	0.354	0.427
	MOM	-0.065	0.347	0.412
2	WAL	-0.161	0.055	0.216
	AGC	-0.163	0.062	0.225
	HAL	-0.157	0.055	0.212
	JFP	-0.160	0.059	0.219
	MLE	-0.163	0.061	0.224
	MOM	-0.157	0.055	0.212
3	WAL	-0.288	0.007	0.295
	AGC	-0.283	0.012	0.295
	HAL	-0.281	0.009	0.290
	JFP	-0.282	0.011	0.293
	MLE	-0.283	0.012	0.295
	MOM	-0.281	0.009	0.290

According to Roths & Tebbs (2006) if suitable computing facilities are available, they would recommend either the Jeffreys-Perks or Beal-MLE interval, because their coverage probabilities are closer to the nominal level than those for the Haldane and Beal-MOM intervals and are not as conservative as the Agresti-Caffo interval. Roths & Tebbs (2006) only considered the first three stages, they compared the performance of the six intervals only for these stages, where the results are given in Table 3.10. We considered all five stages, and compared Bayesian methods for all five stages, these results are given in Table 3.11.

Table 3.11: 95% Bayesian credible intervals for the difference in disease transmission probabilities among male and female insects. Using the Jeffreys prior, uniform prior and probability matching prior.

Stage	Interval	Lower limit	Upper limit	Length
1	Jeffreys	-0.073	0.366	0.439
	Uniform	-0.057	0.348	0.405
	PMP	-0.078	0.346	0.423
2	Jeffreys	-0.175	0.056	0.231
	Uniform	-0.177	0.071	0.248
	PMP	-0.163	0.071	0.234
3	Jeffreys	-0.293	0.004	0.297
	Uniform	-0.267	0.016	0.284
	PMP	-0.280	0.011	0.290
4	Jeffreys	-0.227	0.138	0.365
	Uniform	-0.225	0.156	0.381
	PMP	-0.219	0.144	0.363
5	Jeffreys	-0.138	0.223	0.361
	Uniform	-0.136	0.232	0.369
	PMP	-0.141	0.228	0.368

From Table 3.11 it can be seen that the Bayesian credible intervals when using the Jeffreys prior compares well with the other methods. The lower limits of the intervals are however, in general, somewhat smaller and the interval lengths somewhat larger than those of the methods suggested by Roths & Tebbs (2006). From our simulation studies (Tables 3.1 and 3.2) it was clear that there is not much to choose between the Jeffreys and probability matching priors.

3.7 Conclusion

In this chapter the probability matching prior for a linear combination of binomial proportions, i.e. $\theta = \sum_{i=1}^{k} a_i p_i$, was derived. An example and two simulation studies were considered. A weighted Monte Carlo method was used for the simulation from the posterior distribution when the probability matching prior was used. The goal of this experiment, in the example, was to assess if male and female insects transmit the Mal de Rio Cuarto virus to susceptible maize plants at similar rates. In the first simulation study a comparison was made between the six intervals from Roths & Tebbs (2006) and two Bayesian intervals. For the two Bayesian intervals we used the Jeffreys and probability matching priors. In the second simulation study a comparison was made between the Jeffreys, uniform and probability matching priors for the difference between two binomial proportions. Different values of n_1, n_2, p_1 and p_2 were considered. The probability matching prior and the uniform prior gave similar results. In general, we can conclude that the coverage probabilities for the Jeffreys prior is below the nominal level. Limited simulation studies have shown that the probability matching prior achieves its sample frequentist coverage results somewhat better than in the case of the Jeffreys prior.

Chapter 4

Estimation for the Ratio and Product of Poisson Rates

4.1 Introduction

The Poisson distribution is often used as a probability model to describe the occurrence of rare events. For example the number of defects in items randomly selected from a production process may follow a Poisson distribution. Also the number of misprints counted on the first four pages of an early draft of a scientific paper. Events may also occur over time such as the number of radio-active decays in a fixed time interval, the number of injuries during a rugby match and the number of overseas telephone calls per hour. Research has been done improving statistical inferences on Poisson data. Methods for computing point and interval estimates of a single Poisson rate are, for example, discussed in Hald (1952), Guenther (1965) and Agresti & Coull (1998). Barker (2002) also made an attempt to find approximate confidence intervals for a single Poisson rate.

Our interest is to make Bayesian inferences on nonlinear functions of Poisson rates. Kim (2006) derived a noninformative (probability matching) prior for $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson rates, thereby obtaining approximate point and Bayesian credible intervals of the reliability of systems of k independent parallel components. The parameter $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson parameters appears in applications to system reliability. If a system consists of k components in parallel, then the probability of system failure is $\psi = \prod_{i=1}^{k} \left(\frac{\lambda_i}{n_i}\right)^{a_i}$ where $p_i = \frac{\lambda_i}{n_i}$ is the probability that the i^{th} component will fail. Also if a system requires that at least one of each of k types of components must be employed and that these components are needed in parallel, then the probability of failure of an m- component system is $\psi = \prod_{i=1}^{k} \left(\frac{\lambda_i}{n_i}\right)^{a_i}$, where $k < m$, a_i is the number of components of type i and $\sum_{i=1}^{k} a_i = m$. In two sample situations it may be of interest to test or to construct credibility intervals for the ratio of two Poisson rates. Price & Bonett (2000) used noninformative priors for small and large values of λ_i $(i = 1,2)$ to construct credibility intervals for

$\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates. According to them these improper priors worked well. Our main purpose of this chapter is to obtain probability matching priors for nonlinear functions of Poisson rates. The reference prior for the ratio of two Poisson rates will also be derived.

The parameter $\psi = \xi/(\prod_{i=1}^{k} n_i^{a_i})$ where $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$ is the product of different powers of k Poisson rates and appears in applications to system reliability. The probability matching prior for ξ will therefore be derived as well as the reference prior for the ratio λ_1/λ_2.

Datta & Ghosh (1995) derived the differential equation which a prior must satisfy if the posterior probability of a one sided credibility interval for a parametric function and its frequentist probability agree up to $O\left(n^{-1}\right)$, where n is the sample size. They proved that the agreement between the posterior probability and the frequentist probability holds if and only if $\sum_{i=1}^{k} \frac{\partial}{\partial \lambda_i} \{\eta_i(\underline{\lambda})\, \pi(\underline{\lambda})\} = 0$, where $\pi(\underline{\lambda})$ is the probability matching prior for $\underline{\lambda} = [\lambda_1\, \lambda_2 \ldots \lambda_k]'$, the vector of unknown parameters. Let $\nabla_t(\underline{\lambda}) = \left[\begin{array}{ccc} \frac{\partial}{\partial \lambda_1} t(\underline{\lambda}) & \cdots & \frac{\partial}{\partial \lambda_k} t(\underline{\lambda}) \end{array} \right]'$, where $t(\underline{\lambda})$ is a nonlinear function of Poisson parameters, then $\eta(\underline{\lambda}) = \frac{F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda})}{\sqrt{\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda})}} = \left[\begin{array}{ccc} \eta_1(\underline{\lambda}) & \cdots & \eta_k(\underline{\lambda}) \end{array} \right]'$, where $F^{-1}(\underline{\lambda})$ is the inverse of $F(\underline{\lambda})$, the Fisher information matrix of $\underline{\lambda}$.

4.2 Probability Matching Prior for the Product of Different Powers of k Poisson Rates

In this section we will derive the probability matching prior for the general case where the parameter of interest is $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$. Kim (2006) derived the probability matching prior for the case where $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$. Kim, however, used a different method to derive the probability matching prior. Kim's proof is based on the method of Tibshirani (1989), whilst our proof is based on the procedure of Datta & Ghosh (1995).

Consider a sample from k Poisson populations. Let X_i be an observation from population i. Then $X_1, X_2, \ldots, X_k$ will be independent Poisson distributions such that $X_i \sim P(\lambda_i)$, for $i = 1, 2, \ldots, k$, where λ_i is the expected number of events per unit sample.

Theorem 4.1. The probability matching prior for $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson rates, is given by

$$\pi_{PM}(\underline{\lambda}) = \pi_{PM}(\lambda_1, \lambda_2, \ldots, \lambda_k) \;\propto\; \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i^{-1} \right\}^{\frac{1}{2}}. \tag{4.1}$$

Proof. Assume that $X_1, X_2, \ldots, X_k$ are independent Poisson random variables with $X_i \sim P(\lambda_i)$, for $i = 1, 2, \ldots, k$.

Therefore $P(X_i = x_i) = \dfrac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}$ for $x_i = 0, 1, 2, \ldots$.

The likelihood function is given by

$$\begin{aligned} L(\lambda_1, \lambda_2 \ldots, \lambda_k | x_1, x_2 \ldots, x_k) &= L(\underline{\lambda} | x_1, x_2 \ldots, x_k) \\ &= \prod_{i=1}^{k} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}. \end{aligned}$$

The derivation of the inverse of the Fisher information matrix is given in Appendix C in Theorem C.1. The inverse of the Fisher information matrix is given by

$$F^{-1}(\underline{\lambda}) = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}.$$

We are interested in a probability matching prior for $t(\underline{\lambda}) = \xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson rates.

Now

$$\begin{aligned} \nabla_t'(\underline{\lambda}) &= \begin{bmatrix} \frac{\partial t(\underline{\lambda})}{\partial \lambda_1} & \frac{\partial t(\underline{\lambda})}{\partial \lambda_2} & \cdots & \frac{\partial t(\underline{\lambda})}{\partial \lambda_k} \end{bmatrix} \\ &= \begin{bmatrix} a_1 \lambda_1^{a_1 - 1} \prod_{i \neq 1}^{k} \lambda_i^{a_i} & a_2 \lambda_2^{a_2 - 1} \prod_{i \neq 2}^{k} \lambda_i^{a_i} & \cdots & a_k \lambda_k^{a_k - 1} \prod_{i \neq k}^{k} \lambda_i^{a_i} \end{bmatrix} \\ &= \begin{bmatrix} \frac{a_1}{\lambda_1} \prod_{i=1}^{k} \lambda_i^{a_i} & \frac{a_2}{\lambda_2} \prod_{i=1}^{k} \lambda_i^{a_i} & \cdots & \frac{a_k}{\lambda_k} \prod_{i=1}^{k} \lambda_i^{a_i} \end{bmatrix} \\ &= \begin{bmatrix} \frac{a_1}{\lambda_1} & \frac{a_2}{\lambda_2} & \cdots & \frac{a_k}{\lambda_k} \end{bmatrix} \prod_{i=1}^{k} \lambda_i^{a_i}. \end{aligned}$$

Also

$$\begin{aligned} \nabla_t'(\underline{\lambda}) F^{-1}(\underline{\lambda}) &= \begin{bmatrix} \frac{a_1}{\lambda_1} \prod_{i=1}^{k} \lambda_i^{a_i} & \frac{a_2}{\lambda_2} \prod_{i=1}^{k} \lambda_i^{a_i} & \cdots & \frac{a_k}{\lambda_k} \prod_{i=1}^{k} \lambda_i^{a_i} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix} \\ &= \begin{bmatrix} a_1 \prod_{i=1}^{k} \lambda_i^{a_i} & a_2 \prod_{i=1}^{k} \lambda_i^{a_i} & \cdots & a_k \prod_{i=1}^{k} \lambda_i^{a_i} \end{bmatrix} \\ &= \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix} \prod_{i=1}^{k} \lambda_i^{a_i} \end{aligned}$$

and

$$\begin{aligned}\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda}) &= \left[\begin{array}{cccc} a_1\prod_{i=1}^{k}\lambda_i^{a_i} & a_2\prod_{i=1}^{k}\lambda_i^{a_i} & \cdots & a_k\prod_{i=1}^{k}\lambda_i^{a_i}\end{array}\right] \\ &\quad\times\left[\begin{array}{c} \frac{a_1}{\lambda_1}\prod_{i=1}^{k}\lambda_i^{a_i} \\ \frac{a_2}{\lambda_2}\prod_{i=1}^{k}\lambda_i^{a_i} \\ \vdots \\ \frac{a_k}{\lambda_k}\prod_{i=1}^{k}\lambda_i^{a_i}\end{array}\right] \\ &= \left(\prod_{i=1}^{k}\lambda_i^{a_i}\right)^2\sum_{i=1}^{k}a_i^2\lambda_i^{-1}.\end{aligned}$$

Define

$$\begin{aligned}\eta'(\underline{\lambda}) &= \frac{\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})}{\sqrt{\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda})}} \\ &= \left[\begin{array}{cccc}\frac{a_1}{\sqrt{\sum_{i=1}^{k}a_i^2\lambda_i^{-1}}} & \frac{a_2}{\sqrt{\sum_{i=1}^{k}a_i^2\lambda_i^{-1}}} & \cdots & \frac{a_k}{\sqrt{\sum_{i=1}^{k}a_i^2\lambda_i^{-1}}}\end{array}\right] \\ &= \left[\begin{array}{cccc}\eta_1(\underline{\lambda}) & \eta_2(\underline{\lambda}) & \cdots & \eta_k(\underline{\lambda})\end{array}\right].\end{aligned}$$

The prior $\pi(\underline{\lambda})$ is a probability matching prior if and only if the differential equation $\sum_{i=1}^{k}\frac{\partial}{\partial\lambda_i}\{\eta_i(\underline{\lambda})\pi(\underline{\lambda})\}=0$ is satisfied.

Let

$$\pi(\underline{\lambda}) = \left\{\sum_{i=1}^{k}a_i^2\lambda_i^{-1}\right\}^{\frac{1}{2}}$$

then

$$\begin{aligned}\eta_1(\underline{\lambda})\pi(\underline{\lambda}) &= \frac{a_1}{\sqrt{\sum_{i=1}^{k}a_i^2\lambda_i^{-1}}}\left\{\sum_{i=1}^{k}a_i^2\lambda_i^{-1}\right\}^{\frac{1}{2}} \\ &= a_1\end{aligned}$$

therefore

$$\frac{\partial}{\partial\lambda_1}\{\eta_1(\underline{\lambda})\pi(\underline{\lambda})\} = \frac{\partial}{\partial\lambda_1}\{a_1\}=0$$

and

$$\begin{aligned}\eta_2(\underline{\lambda})\,\pi(\underline{\lambda}) &= \frac{a_2}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i^{-1}}}\left\{\sum_{i=1}^{k} a_i^2\lambda_i^{-1}\right\}^{\frac{1}{2}} \\ &= a_2\end{aligned}$$

therefore

$$\frac{\partial}{\partial\lambda_2}\{\eta_2(\underline{\lambda})\,\pi(\underline{\lambda})\} = \frac{\partial}{\partial\lambda_2}\{a_2\} = 0$$

and

$$\begin{aligned}\eta_k(\underline{\lambda})\,\pi(\underline{\lambda}) &= \frac{a_k}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i^{-1}}}\left\{\sum_{i=1}^{k} a_i^2\lambda_i^{-1}\right\}^{\frac{1}{2}} \\ &= a_k\end{aligned}$$

therefore

$$\frac{\partial}{\partial\lambda_k}\{\eta_k(\underline{\lambda})\,\pi(\underline{\lambda})\} = \frac{\partial}{\partial\lambda_k}\{a_k\} = 0.$$

We can therefore conclude that

$$\sum_{i=1}^{k}\frac{\partial}{\partial\lambda_i}\{\eta_i(\underline{\lambda})\,\pi(\underline{\lambda})\} = 0.$$

The differential equation will be satisfied if $\pi(\underline{\lambda})$ is

$$\pi_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2\lambda_i^{-1}\right\}^{\frac{1}{2}}.$$

□

The joint posterior distribution when using the probability matching prior is given by

$$\begin{aligned}
\pi_{PM}(\underline{\lambda}\,|data) &\propto \pi_{PM}(\underline{\lambda}) \times L(\underline{\lambda}\,|data) \\
&\propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i^{-1}\right\}^{\frac{1}{2}} \times \prod_{i=1}^{k} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} \\
\therefore \pi_{PM}(\underline{\lambda}\,|data) &\propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i^{-1}\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}. \qquad (4.2)
\end{aligned}$$

When $a_i = 1$, the probability matching prior for $\xi = \prod_{i=1}^{k} \lambda_i$, will be

$$\pi_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} \lambda_i^{-1}\right\}^{\frac{1}{2}}. \qquad (4.3)$$

The prior in Equation 4.3 was used in the simulation study in Section 2.5.2, where $k = 2$.

For $a_i = 1$, $(i = 1, 2, \ldots, k)$, the posterior distribution in the case of the probability matching prior is given by

$$\pi_{PM}(\underline{\lambda}\,|data) \propto \left\{\sum_{i=1}^{k} \lambda_i^{-1}\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}. \qquad (4.4)$$

Kim (2006) showed that the posterior distribution, $\pi_{PM}(\underline{\lambda}\,|data)$, from Equation 4.2 is a proper distribution.

Theorem 4.2. The posterior distribution for the ratio $\nu = \lambda_1/\lambda_2$ when using the probability matching prior is given by

$$\pi_{PM}(\nu\,|x_1, x_2) \propto \frac{1}{B\left(x_1 + \frac{1}{2}, x_2 + \frac{1}{2}\right)} \nu^{x_1 - \frac{1}{2}} \left(\frac{1}{\nu + 1}\right)^{x_1 + x_2 + 1} \qquad \text{for } \nu > 0 \qquad (4.5)$$

which is a beta distribution of the second kind.

Proof. If $a_1 = 1, a_2 = -1$ and $a_3 = a_4 = \ldots = a_k = 0$, it easily follows from Equation 4.1 that the probability matching prior in the case of $\nu = \lambda_1/\lambda_2$ is given by

$$\pi_{PM}(\lambda_1, \lambda_2) \propto \frac{(\lambda_1 + \lambda_2)^{\frac{1}{2}}}{\lambda_1^{\frac{1}{2}} \lambda_2^{\frac{1}{2}}}. \qquad (4.6)$$

Using Equation 4.2, the joint posterior distribution of λ_1 and λ_2 is given by

$$\pi_{PM}(\lambda_1,\lambda_2|x_1,x_2) \propto \frac{(\lambda_1+\lambda_2)^{\frac{1}{2}}}{\lambda_1^{\frac{1}{2}}\lambda_2^{\frac{1}{2}}} e^{-(\lambda_1+\lambda_2)}\lambda_1^{x_1}\lambda_2^{x_2}.$$

Let $\nu = {}^{\lambda_1}/_{\lambda_2}$, thus $\lambda_1 = \nu\lambda_2$ and $d\lambda_1 = \lambda_2 d\nu$, then

$$\begin{aligned}\pi_{PM}(\nu,\lambda_2|x_1,x_2) &\propto (\nu\lambda_2+\lambda_2)^{\frac{1}{2}} e^{-(\nu\lambda_2+\lambda_2)}(\nu\lambda_2)^{x_1-\frac{1}{2}}\lambda_2^{x_2-\frac{1}{2}}\lambda_2 \\ &= \lambda_2^{\frac{1}{2}}(\nu+1)^{\frac{1}{2}} e^{-\lambda_2(\nu+1)}\nu^{x_1-\frac{1}{2}}\lambda_2^{x_1-\frac{1}{2}}\lambda_2^{x_2-\frac{1}{2}}\lambda_2 \\ &= \nu^{x_1-\frac{1}{2}}(1+\nu)^{\frac{1}{2}}\lambda_2^{x_1+x_2+\frac{1}{2}} e^{-\lambda_2(1+\nu)}\end{aligned}$$

and

$$\begin{aligned}\pi_{PM}(\nu|x_1,x_2) &= \int_0^\infty \pi_{PM}(\nu,\lambda_2|x_1,x_2)\,d\lambda_2 \\ &= \nu^{x_1-\frac{1}{2}}(1+\nu)^{\frac{1}{2}} \underbrace{\int_0^\infty \lambda_2^{x_1+x_2+\frac{1}{2}} e^{-\lambda_2(1+\nu)} d\lambda_2}_{Gamma\left(\alpha=x_1+x_2+1\frac{1}{2},\beta=1+\nu\right)} \\ &\propto \nu^{x_1-\frac{1}{2}}(1+\nu)^{\frac{1}{2}} \left(\frac{(1+\nu)^{x_1+x_2+1\frac{1}{2}}}{\Gamma\left(x_1+x_2+1\frac{1}{2}\right)}\right)^{-1} \\ &\propto C\nu^{x_1-\frac{1}{2}}(1+\nu)^{-x_1-x_2-1} \qquad \text{for } \nu>0 \end{aligned} \tag{4.7}$$

which is a beta distribution of the second kind and $C = \frac{1}{B\left(x_1+\frac{1}{2},x_2+\frac{1}{2}\right)}$. □

4.3 The Jeffreys and Uniform Priors for the Product of Different Powers of k Poisson Rates

The Jeffreys prior is given by

$$\pi_J(\underline{\lambda}) \propto |F(\underline{\lambda})|^{\frac{1}{2}} = \left(\prod_{i=1}^{k}\lambda_i\right)^{-\frac{1}{2}} \tag{4.8}$$

where $F(\underline{\lambda})$ is the Fisher information matrix connected with the likelihood function. The prior in Equation 4.8 was used in the simulation study in Section 2.5.2, where $k = 2$.

When using the Jeffreys prior, the joint posterior distribution of $\underline{\lambda}$ is given by

$$\begin{aligned}
\pi_J(\underline{\lambda}\,|data) &\propto \pi_J(\underline{\lambda}) \times L(\underline{\lambda}\,|data) \\
&\propto \left(\prod_{i=1}^{k}\lambda_i\right)^{-\frac{1}{2}} \prod_{i=1}^{k}\lambda_i^{x_i}e^{-\lambda_i} \\
\therefore \pi_J(\underline{\lambda}\,|data) &\propto \prod_{i=1}^{k}\lambda_i^{x_i-\frac{1}{2}}e^{-\lambda_i}. \qquad (4.9)
\end{aligned}$$

The posterior distribution of $\underline{\lambda}$ is thus the product of k independently distributed *Gamma* $\left(x_i+\frac{1}{2},1\right)$ variates.

Theorem 4.3. The posterior distribution for the ratio $\nu = \lambda_1/\lambda_2$ when using the Jeffreys prior is given by

$$\pi_J(\nu\,|x_1,x_2) \propto \frac{1}{B\left(x_1+\frac{1}{2},x_2+\frac{1}{2}\right)}\nu^{x_1-\frac{1}{2}}\left(\frac{1}{\nu+1}\right)^{x_1+x_2+1} \qquad \text{for } \nu>0 \qquad (4.10)$$

which is a beta distribution of the second kind.

Proof. If $k=2$, it easily follows from Equation 4.8 that the Jeffreys prior is given by

$$\pi_J(\lambda_1,\lambda_2) \propto \left(\prod_{i=1}^{2}\lambda_i\right)^{-\frac{1}{2}}. \qquad (4.11)$$

Using Equation 4.9, the joint posterior distribution of λ_1 and λ_2 is given by

$$\pi_J(\lambda_1,\lambda_2\,|x_1,x_2) \propto e^{-(\lambda_1+\lambda_2)}\lambda_1^{x_1-\frac{1}{2}}\lambda_2^{x_2-\frac{1}{2}}.$$

Let $\nu = \lambda_1/\lambda_2$, thus $\lambda_1 = \nu\lambda_2$ and $d\lambda_1 = \lambda_2 d\nu$, then

$$\begin{aligned}
\pi_J(\nu,\lambda_2\,|x_1,x_2) &\propto e^{-(\nu\lambda_2+\lambda_2)}(\nu\lambda_2)^{x_1-\frac{1}{2}}\lambda_2^{x_2-\frac{1}{2}}\lambda_2 \\
&= e^{-\lambda_2(\nu+1)}\nu^{x_1-\frac{1}{2}}\lambda_2^{x_1-\frac{1}{2}}\lambda_2^{x_2-\frac{1}{2}}\lambda_2 \\
&= \nu^{x_1-\frac{1}{2}}\lambda_2^{x_1+x_2}e^{-\lambda_2(1+\nu)}
\end{aligned}$$

and

$$\begin{aligned}
\pi_J(\nu|x_1,x_2) &= \int_0^\infty \pi_J(\nu,\lambda_2|x_1,x_2)\,d\lambda_2 \\
&= \nu^{x_1-\frac{1}{2}} \underbrace{\int_0^\infty \lambda_2^{x_1+x_2} e^{-\lambda_2(1+\nu)} d\lambda_2}_{Gamma(\alpha=x_1+x_2+1,\beta=1+\nu)} \\
&\propto \nu^{x_1-\frac{1}{2}} \left(\frac{(\nu+1)^{x_1+x_2+1}}{\Gamma(x_1+x_2+1)}\right)^{-1} \\
&\propto C\nu^{x_1-\frac{1}{2}}(1+\nu)^{-x_1-x_2-1} \qquad \text{for } \nu>0
\end{aligned} \tag{4.12}$$

which is a beta distribution of the second kind and $C = \frac{1}{B\left(x_1+\frac{1}{2},x_2+\frac{1}{2}\right)}$. □

The uniform prior is given by

$$\pi_U(\underline{\lambda}) \propto \text{constant}. \tag{4.13}$$

When using the uniform prior, the joint posterior distribution of $\underline{\lambda}$ is given by

$$\pi_U(\underline{\lambda}|data) \propto \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}. \tag{4.14}$$

The posterior distribution of $\underline{\lambda}$ is thus the product of k independently distributed $Gamma(x_i+1,1)$ variates.

Theorem 4.4. The posterior distribution for the ratio $\nu = \lambda_1/\lambda_2$ when using the uniform prior is given by

$$\pi_U(\nu|x_1,x_2) \propto \frac{1}{B(x_1+1,x_2+1)} \nu^{x_1} \left(\frac{1}{\nu+1}\right)^{x_1+x_2+2} \qquad \text{for } \nu>0 \tag{4.15}$$

which is a beta distribution of the second kind.

Proof. If $k=2$ it easily follows from Equation 4.13 that the uniform prior in the case of $\nu = \lambda_1/\lambda_2$ is given by

$$\pi_U(\lambda_1,\lambda_2) \propto \text{constant}. \tag{4.16}$$

Using Equation 4.14, the joint posterior distribution of λ_1 and λ_2 is given by

$$\pi_U(\lambda_1,\lambda_2|x_1,x_2) \propto e^{-(\lambda_1+\lambda_2)} \lambda_1^{x_1} \lambda_2^{x_2}.$$

Let $\nu = \lambda_1/\lambda_2$, thus $\lambda_1 = \nu\lambda_2$ and $d\lambda_1 = \lambda_2 d\nu$, then

$$\begin{aligned}\pi_U(\nu,\lambda_2|x_1,x_2) &\propto e^{-(\nu\lambda_2+\lambda_2)}(\nu\lambda_2)^{x_1}\lambda_2^{x_2}\lambda_2 \\ &= e^{-\lambda_2(\nu+1)}\nu^{x_1}\lambda_2^{x_1}\lambda_2^{x_2}\lambda_2 \\ &= \nu^{x_1}\lambda_2^{x_1+x_2+1}e^{-\lambda_2(1+\nu)}\end{aligned}$$

and

$$\begin{aligned}\pi_U(\nu|x_1,x_2) &= \int_0^\infty \pi_U(\nu,\lambda_2|x_1,x_2)\,d\lambda_2 \\ &= \nu^{x_1}\underbrace{\int_0^\infty \lambda_2^{x_1+x_2+1}e^{-\lambda_2(1+\nu)}d\lambda_2}_{Gamma(\alpha=x_1+x_2+2,\beta=1+\nu)} \\ &\propto \nu^{x_1}\left(\frac{(\nu+1)^{x_1+x_2+2}}{\Gamma(x_1+x_2+2)}\right)^{-1} \\ &\propto C\nu^{x_1}(1+\nu)^{-x_1-x_2-2} \qquad \text{for } \nu>0 \end{aligned} \tag{4.17}$$

which is a beta distribution of the second kind and $C = \frac{1}{B(x_1+1,x_2+1)}$. □

Corollary 4.1. The Jeffreys and probability matching priors for $\nu = \lambda_1/\lambda_2$ have the same posterior distribution. As will be seen in the next section, the reference prior of $\nu = \lambda_1/\lambda_2$ also has the same posterior distribution as the Jeffreys and probability matching priors.

4.4 The Reference Prior

As mentioned, the Jeffreys prior is not always suitable for multiparameter problems. In recognition of this problem, Bernardo (1979) and Berger & Bernardo (1992) proposed the reference prior approach to the development of noninformative priors, the key feature of which was a possible dependence of the reference prior on specification of parameters of interest and nuisance parameters. As mentioned by Pearn & Wu (2005) the reference prior maximises the difference in information about the parameter provided by the prior and posterior. The reference prior is derived in such a way that it provides as little information as possible about the parameter. In this section the reference prior of Berger & Bernardo (1992) will be derived for the ratio of two Poisson rates. As in the case of the Jeffreys prior, the reference prior method is derived from the Fisher information matrix. Reference priors depend on the group ordering of the parameters. Berger & Bernardo (1992) recommended the reference prior which is based on having each parameter in its own group, i.e. having each conditional reference prior to be one dimensional.

Theorem 4.5. The reference prior of $\nu = {}^{\lambda_1}/_{\lambda_2}$ for the group ordering $\{\lambda_1, \lambda_2\}$ is given by

$$\pi_R(\lambda_1, \lambda_2) \propto \left\{ \frac{1}{\lambda_1 \lambda_2 (\lambda_1 + \lambda_2)} \right\}^{\frac{1}{2}}. \tag{4.18}$$

Proof. By making a transformation we will, first derive the reference prior, $\pi_R(\nu, \lambda_2)$. The Fisher information matrix $F(\nu, \lambda_2) = A' F(\lambda_1, \lambda_2) A$ where

$$A = \frac{\partial(\lambda_1, \lambda_2)}{\partial(\nu, \lambda_2)} = \begin{bmatrix} \lambda_2 & \nu \\ 0 & 1 \end{bmatrix}.$$

Therefore

$$\begin{aligned} F(\nu, \lambda_2) &= \begin{bmatrix} \lambda_2 & 0 \\ \nu & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\nu\lambda_2} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} \begin{bmatrix} \lambda_2 & \nu \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_2}{\nu} & 1 \\ 1 & \frac{(\nu+1)}{\lambda_2} \end{bmatrix}. \end{aligned}$$

Now

$$h_1 = \left| \frac{\lambda_2}{\nu} - \frac{\lambda_2}{(\nu+1)} \right| = \lambda_2 \left(\frac{1}{\nu(\nu+1)} \right)$$

and

$$\begin{aligned} \pi_R(\nu) &= |h_1|^{\frac{1}{2}} \propto \frac{1}{\nu^{\frac{1}{2}} (\nu+1)^{\frac{1}{2}}} \\ h_2 &= \left| \frac{1}{\lambda_2} (\nu+1) \right|. \end{aligned}$$

Therefore

$$\pi_R(\lambda_2 | \nu) = |h_2|^{\frac{1}{2}} \propto \left(\frac{1}{\lambda_2} \right)^{\frac{1}{2}}.$$

The joint prior for the group ordering $\{\nu, \lambda_2\}$ is given by

$$\pi_R(\nu, \lambda_2) = \pi_R(\nu)\, \pi_R(\lambda_2 | \nu) \propto \left(\frac{1}{\lambda_2} \right)^{\frac{1}{2}} \left(\frac{1}{\nu(\nu+1)} \right)^{\frac{1}{2}}$$

and the joint reference prior for the group ordering $\{\lambda_1, \lambda_2\}$ is given by

$$\pi_R(\lambda_1, \lambda_2) \propto \left\{ \frac{1}{\lambda_1 \lambda_2 (\lambda_1 + \lambda_2)} \right\}^{\frac{1}{2}}.$$

□

The reference prior is also a probability matching prior.

Theorem 4.6. The posterior distribution for the ratio $\nu = \lambda_1/\lambda_2$ when using the reference prior is given by

$$\pi_R(\nu \,|x_1, x_2) \propto \frac{1}{B\left(x_1 + \frac{1}{2}, x_2 + \frac{1}{2}\right)} \nu^{x_1 - \frac{1}{2}} \left(\frac{1}{\nu + 1} \right)^{x_1 + x_2 + 1} \qquad \text{for } \nu > 0 \tag{4.19}$$

which is a beta distribution of the second kind.

Proof. From Equation 4.18 the reference prior is given by

$$\pi_R(\lambda_1, \lambda_2) \propto \left\{ \frac{1}{\lambda_1 \lambda_2 (\lambda_1 + \lambda_2)} \right\}^{\frac{1}{2}}. \tag{4.20}$$

The joint posterior distribution of λ_1 and λ_2 is given by

$$\pi_R(\lambda_1, \lambda_2 \,|x_1, x_2) \propto (\lambda_1 + \lambda_2)^{-\frac{1}{2}} e^{-(\lambda_1 + \lambda_2)} \lambda_1^{x_1 - \frac{1}{2}} \lambda_2^{x_2 - \frac{1}{2}}.$$

Let $\nu = \lambda_1/\lambda_2$, thus $\lambda_1 = \nu\lambda_2$ and $d\lambda_1 = \lambda_2 d\nu$, then

$$\begin{aligned} \pi_R(\nu, \lambda_2 \,|x_1, x_2) &\propto (\nu\lambda_2 + \lambda_2)^{-\frac{1}{2}} e^{-(\nu\lambda_2 + \lambda_2)} (\nu\lambda_2)^{x_1 - \frac{1}{2}} \lambda_2^{x_2 - \frac{1}{2}} \lambda_2 \\ &= \lambda_2^{-\frac{1}{2}} (\nu + 1)^{-\frac{1}{2}} e^{-\lambda_2(\nu+1)} \nu^{x_1 - \frac{1}{2}} \lambda_2^{x_1 - \frac{1}{2}} \lambda_2^{x_2 - \frac{1}{2}} \lambda_2 \\ &= \nu^{x_1 - \frac{1}{2}} (\nu + 1)^{-\frac{1}{2}} \lambda_2^{x_1 + x_2 - \frac{1}{2}} e^{-\lambda_2(1+\nu)} \end{aligned}$$

and

$$\begin{aligned}
\pi_R(\nu\,|x_1,x_2) &= \int_0^\infty \pi_R(\nu,\lambda_2\,|x_1,x_2)\,d\lambda_2 \\
&= \nu^{x_1-\frac{1}{2}}(\nu+1)^{-\frac{1}{2}} \underbrace{\int_0^\infty \lambda_2^{x_1+x_2-\frac{1}{2}} e^{-\lambda_2(1+\nu)} d\lambda_2}_{Gamma\left(\alpha=x_1+x_2+\frac{1}{2},\beta=1+\nu\right)} \\
&\propto \nu^{x_1-\frac{1}{2}}(\nu+1)^{-\frac{1}{2}} \left(\frac{(\nu+1)^{x_1+x_2+\frac{1}{2}}}{\Gamma\left(x_1+x_2+\frac{1}{2}\right)}\right)^{-1} \\
&\propto C\nu^{x_1-\frac{1}{2}}(1+\nu)^{-x_1-x_2-1} \qquad \text{for } \nu>0 \qquad (4.21)
\end{aligned}$$

which is a beta distribution of the second kind and $C=\frac{1}{B\left(x_1+\frac{1}{2},x_2+\frac{1}{2}\right)}$. □

Corollary 4.2. The Jeffreys, probability matching and reference priors for $\nu=\lambda_1/\lambda_2$ have the same posterior distribution.

4.5 The Weighted Monte Carlo Method in the Case of the Probability Matching Prior for $\xi=\prod_{i=1}^k \lambda_i^{a_i}$

This method has been introduced in Section 2.4.

If a uniform prior is put on $\underline{\lambda}$, using Equation 4.14, the posterior (proposal) distribution is

$$q(\underline{\lambda}\,|data) \propto \prod_{i=1}^k \lambda_i^{x_i} e^{-\lambda_i}.$$

In the case of the probability matching prior, using Equation 4.2, the posterior (target) distribution is

$$\pi_{PM}(\underline{\lambda}\,|data) \propto \left\{\sum_{i=1}^k a_i^2\lambda_i^{-1}\right\}^{\frac{1}{2}} \prod_{i=1}^k \lambda_i^{x_i} e^{-\lambda_i}.$$

The sample probabilities are therefore proportional to

$$\frac{\pi_{PM}(\underline{\lambda}\,|data)}{q(\underline{\lambda}\,|data)} = \left\{\sum_{i=1}^k a_i^2\lambda_i^{-1}\right\}^{\frac{1}{2}}$$

and the normalised weights are

$$\omega_l = \frac{\left\{\sum_{i=1}^{k}\left(a_i^2\lambda_i^{-1}\right)^{(l)}\right\}^{\frac{1}{2}}}{\sum_{l=1}^{n}\left\{\sum_{i=1}^{k}\left(a_i^2\lambda_i^{-1}\right)^{(l)}\right\}^{\frac{1}{2}}} \qquad l = 1,2,\ldots,n$$

where n is the number of simulations.
The Monte Carlo method:

- **Step 1**

  Obtain a Monte Carlo sample $\left\{\left(\lambda_1^{(l)},\lambda_2^{(l)}\ldots,\lambda_k^{(l)}\right);l=1,2,\ldots,n\right\}$ from the proposal distribution $q\left(\underline{\lambda}\,|data\right)$ and calculate $\xi^{(l)}=\prod_{i=1}^{k}\left(\lambda_i^{(l)}\right)^{a_i}$ for $l=1,2,\ldots,n$.

- **Step 2**

  Sort $\left\{\xi^{(l)},(l=1,2,\ldots,n)\right\}$ to obtain the ordered values $\xi^{[1]}\leq\xi^{[2]}\leq\cdots\leq\xi^{[n]}$.

- **Step 3**

  Each simulated ξ value has an associated weight. Therefore compute the weighted function $\omega_{(l)}$ associated with the l^{th} ordered $\xi^{[l]}$ value.

- **Step 4**

  Add the weights up from left to right (from the first on) until one obtains $\sum_{l=1}^{n_1}\omega_{(l)}=\alpha/2$. Write down the corresponding $\xi^{[n_1]}$ value and denote it as $\xi_{(\alpha/2)}$. Add the weights up from right to left (from the last back) until one obtains $\sum_{l=n_2}^{n}\omega_{(l)}=\alpha/2$. Write down the corresponding $\xi^{[n_2]}$ value and denote it as $\xi_{(1-\alpha/2)}$.

- **Step 5**

  The $100\,(1-\alpha)\,\%$ Bayesian credible interval is: $\left(\xi_{(\alpha/2)},\xi_{(1-\alpha/2)}\right)$.

4.6 Simulation Studies

As mentioned, the parameter $\xi=\prod_{i=1}^{k}\lambda_i^{a_i}$, the product of different powers of k Poisson parameters appears in applications to system reliability. If a system consists of k components in parallel, then the probability of system failure is $\psi=\prod_{i=1}^{k}\left(\frac{\lambda_i}{n_i}\right)^{a_i}$ where $p_i=\frac{\lambda_i}{n_i}$ is the probability that the i^{th} component will fail. Also if a system requires that at least one of each of k types of components must be employed and that these components are needed in parallel, then the probability of failure of an m- component system is $\psi=\prod_{i=1}^{k}\left(\frac{\lambda_i}{n_i}\right)^{a_i}$, where $k<m$, a_i is the number of components of type i and $\sum_{i=1}^{k}a_i=m$.

4.6.1 Simulation Study I - Comparison of the Jeffreys, Uniform and Probability Matching Priors for $\xi_1 = \prod_{i=1}^{k} \lambda_i$

In Table 4.1 the frequentist coverage probabilities are given for $\xi_1 = \prod_{i=1}^{k} \lambda_i$ in the case of:

1. the Jeffreys prior, $\pi_J(\underline{\lambda}) \propto \left(\prod_{i=1}^{k} \lambda_i \right)^{-\frac{1}{2}}$;
2. the uniform prior, $\pi_U(\underline{\lambda}) \propto$ constant;
3. the probability matching prior, $\pi_{PM}(\underline{\lambda}) \propto \left\{ \sum_{i=1}^{k} \lambda_i^{-1} \right\}^{\frac{1}{2}}$;

50 000 samples were generated and from each sample 10 000 parameter values were simulated to obtain the Bayesian credible intervals in the case of the Jeffreys and uniform priors. For the probability matching prior only 20 000 samples were generated.

Table 4.1: Frequentist coverage probabilities for 5% and 95% posterior quantiles of $\xi_1 = \prod_{i=1}^{k} \lambda_i$.

		Jeffreys		**Uniform**		**Prob. Matching**	
$\underline{\lambda}$	ξ_1	5%	95%	5%	95%	5%	95%
[1 1 1]	1	0.0223	0.9512	0.1037	1.0000	0.0551	1.0000
[1 2 3]	6	0.0270	0.9133	0.0849	0.9975	0.0514	0.9819
[2 2 2]	8	0.0243	0.9142	0.0775	1.0000	0.0491	0.9674
[1 5 10]	50	0.0346	0.9576	0.0876	0.9967	0.0532	0.9862
[5 5 5]	125	0.0352	0.9069	0.0612	0.9650	0.0484	0.9417
[10 10 10]	1 000	0.0325	0.9253	0.0588	0.9625	0.0462	0.9490
[1 2 3 4 5]	120	0.0192	0.8861	0.0806	0.9922	0.0500	0.9742
[2 2 3 4 5]	240	0.0194	0.8657	0.0721	0.9847	0.0518	0.9586
[3 3 3 4 5]	540	0.0198	0.8647	0.0666	0.9753	0.0482	0.9499
[1 2 3 4 5 6 7 8]	40 320	0.0141	0.8560	0.0756	0.9877	0.0503	0.9704
[1 2 3 4 5 5 5 5]	15 000	0.0131	0.8442	0.0704	0.9870	0.0514	0.9693
[5 5 5 5 5 5 5 5]	390 625	0.0164	0.8466	0.0591	0.9614	0.0477	0.9458
[5 5 5 5 6 7 8 9]	1 890 000	0.0178	0.8617	0.0587	0.9614	0.0499	0.9455
[5 5 5 5 10 10 10 10]	6 250 000	0.0179	0.8729	0.0582	0.9627	0.0472	0.9472
[10 10 10 10 10 10 10 10]	10^8	0.0184	0.8880	0.0534	0.9546	0.0492	0.9507

From the simulation results in Table 4.1 it is clear that the probability matching prior performs better than the Jeffreys and uniform priors in most of the situations. As mentioned by Kim (2006) if each co-ordinate of the parameter vector $\underline{\lambda}$ is large, the frequentist coverage percentages obtained from using the probability matching prior is close to the desired level.

The simulation results are displayed in Figures 4.1 and 4.2. The inability of the Jeffreys and uniform priors to give good coverage probabilities is even more clear from these graphs.

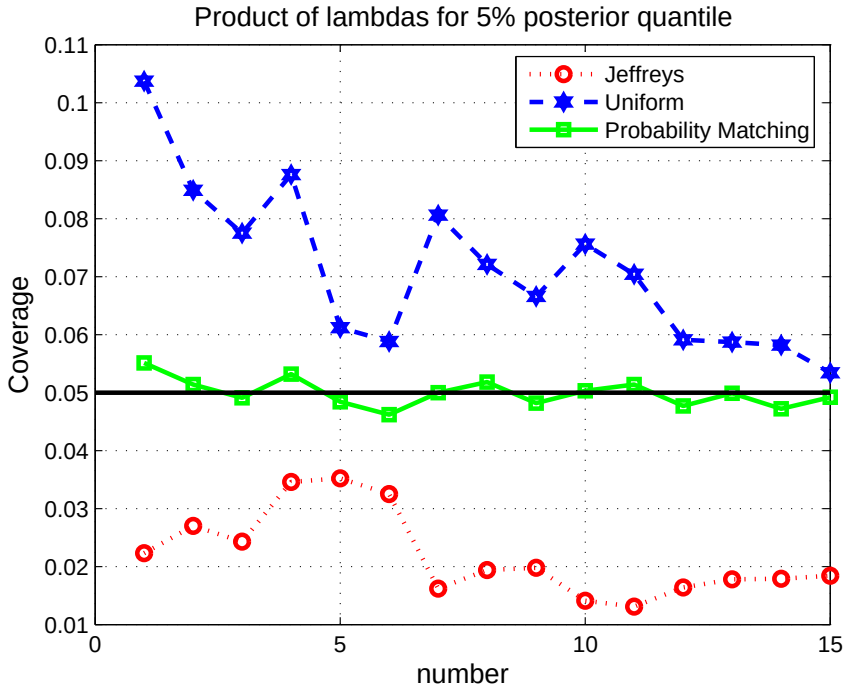


Figure 4.1: Illustration of the 5% quantiles of $\xi_1 = \prod_{i=1}^{k} \lambda_i$ in the same order as given in Table 4.1.

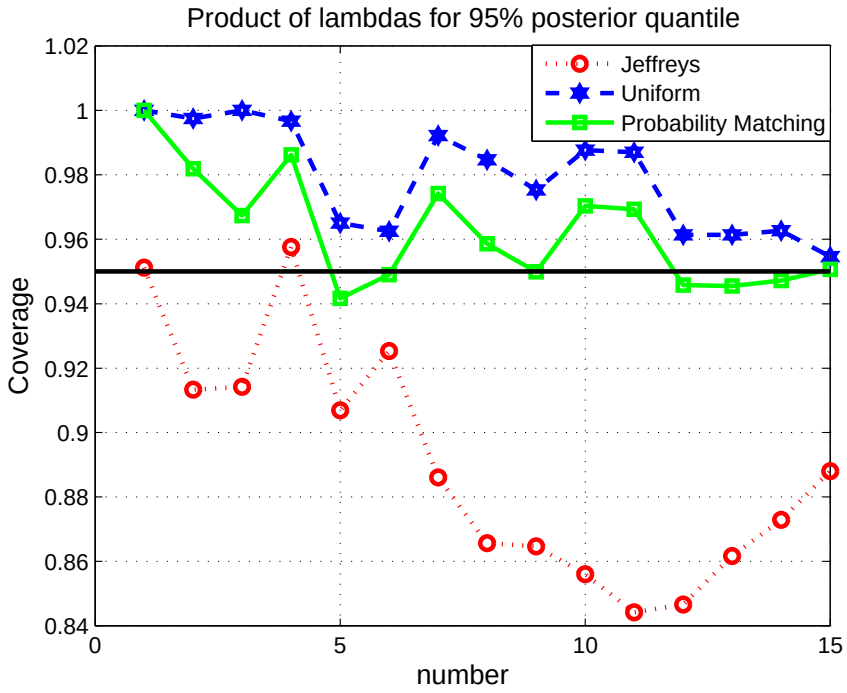


Figure 4.2: Illustration of the 95% quantiles of $\xi_1 = \prod_{i=1}^{k} \lambda_i$ in the same order as given in Table 4.1.

4.6.2 Simulation Study II - Comparing Six Priors for $\xi_2 = \lambda_1\lambda_2$ - Reliability of Independent Parallel Components System

In this example a simulation study is done for $\xi_2 = \lambda_1\lambda_2$, the product of two Poisson rates. The parameter values for the Poisson distributions are λ_i =2, 3, 4, 5, 6, 7, 8, 9, 10 (for $i = 1, 2$).

The priors that will be compared are:

1. the uniform prior: $\pi_U(\lambda_1, \lambda_2) \propto$ constant;
2. the Jeffreys prior: $\pi_J(\lambda_1, \lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-\frac{1}{2}}$;
3. the probability matching prior: $\pi_{PM}(\lambda_1, \lambda_2) \propto \left\{\sum_{i=1}^{2} \lambda_i^{-1}\right\}^{\frac{1}{2}}$;
4. $\pi_A(\lambda_1, \lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-\frac{1}{8}}$;
5. $\pi_B(\lambda_1, \lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-\frac{1}{4}}$;
6. $\pi_C(\lambda_1, \lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-\frac{3}{8}}$.

We know from experience (and this is also clear from Table 4.1) that the Jeffreys prior under covers while the uniform prior tends to over cover in the case of the product of Poisson rates. Priors (4), (5) and (6) are in between priors for (1) and (2) and it is for this reason that they are included in this simulation study. In general we will define these in between priors by $\pi(\lambda_1, \lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-a}$, such that $a = \frac{1}{8}$ for prior (4), $a = \frac{1}{4}$ for prior (5) and $a = \frac{3}{8}$ for prior (6).

The frequentist coverage percentages of the 95% HPD (highest posterior density) intervals and the interval lengths are displayed in Figure 4.3. The graphs are averages over λ_1 for $\lambda_2 = 2$ to 10. The coverage percentage of the probability matching prior is much better than those of the Jeffreys and uniform priors. It is also not impossible that other priors with $a = \frac{1}{5}$ or $a = \frac{1}{6}$ will give a very good coverage.

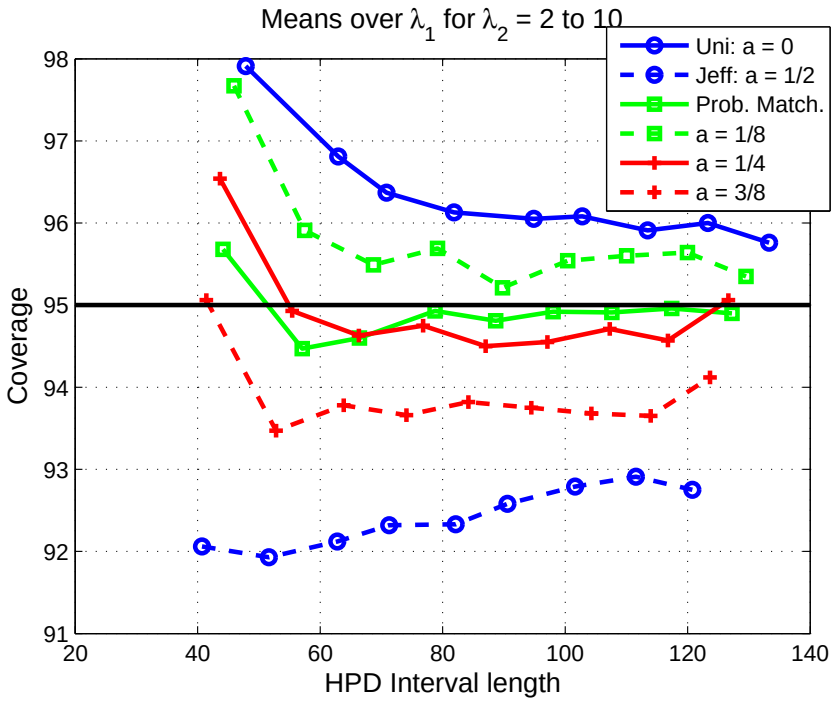


Figure 4.3: Illustration of the coverage percentages of the 95% HPD Intervals of $\xi_2 = \lambda_1\lambda_2$.

4.6.3 Simulation Study III - Comparing Priors for $\xi_3 = \lambda_1^2\lambda_2$ and $\xi_4 = \lambda_1^3\lambda_2$ - Reliability of Repeated Components System

Assume a system needs three components in parallel and at least one of each of two types of components must be used. If the first component is replicated, then the probability of failure is $\psi_3 = p_1^2 p_2$. Also if four components are needed and the first component is replicated three times, then the probability of failure is $\psi_4 = p_1^3 p_2$ where $p_i = \frac{\lambda_i}{n_i}$ $(i = 1,2)$. For further details see Kim (2006). In this simulation study the coverage probabilities of these different priors for the parameters $\xi_3 = \lambda_1^2\lambda_2$ and $\xi_4 = \lambda_1^3\lambda_2$ will therefore be looked at. The parameter values for the Poisson distribution are as in Section 4.6.2, i.e. λ_i =2, 3, 4, 5, 6, 7, 8, 9, 10 (for $i = 1,2$) and the priors that will be compared are:

1. the uniform prior: $\pi_U(\lambda_1,\lambda_2) \propto$ constant;
2. the Jeffreys prior: $\pi_J(\lambda_1,\lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-\frac{1}{2}}$;
3. the probability matching prior: $\pi_{PM}(\lambda_1,\lambda_2) \propto \left\{\sum_{i=1}^{2} a_i^2 \lambda_i^{-1}\right\}^{\frac{1}{2}}$.

The frequentist coverage probabilities as well as the interval lengths of the 95% Bayesian credible intervals for the above priors in the case of $\xi_3 = \lambda_1^2\lambda_2$ are given in Figure 4.4 and in Figure 4.5 the same graphs are given for $\xi_4 = \lambda_1^3\lambda_2$. The graphs are averages over λ_1 for $\lambda_2 = 2$ to 10.

The same patterns as in Figure 4.3 emerge from Figures 4.4 and 4.5, i.e. the Jeffreys prior underestimates the coverage probabilities while the uniform prior tends to overestimate the coverage probabilities. In general the probability matching prior seems to give the best coverage probabilities.

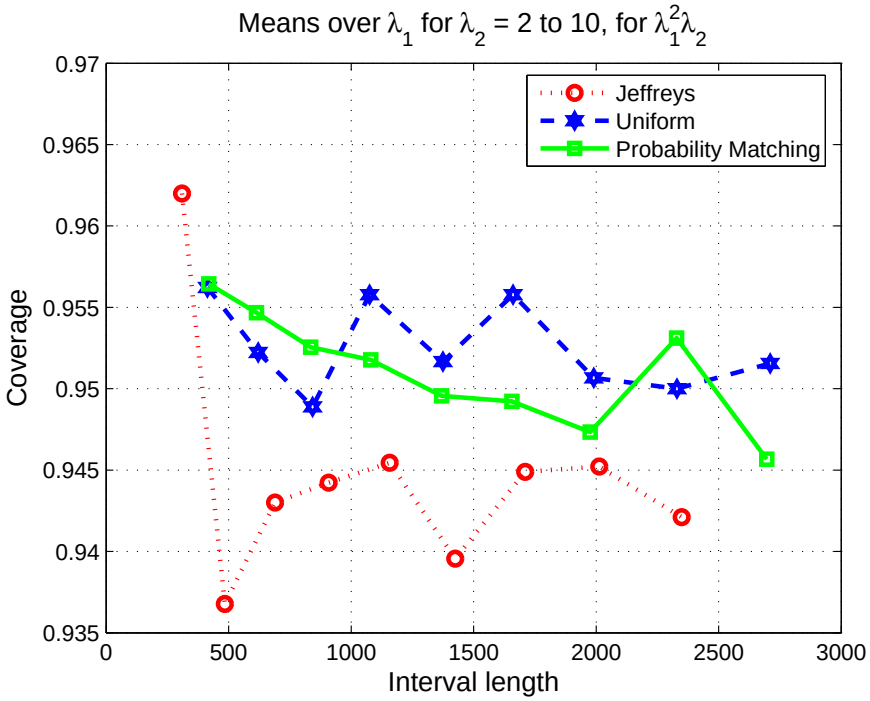


Figure 4.4: Illustration of the coverage probabilities of the 95% Bayesian credible intervals for $\xi_3 = \lambda_1^2\lambda_2$.

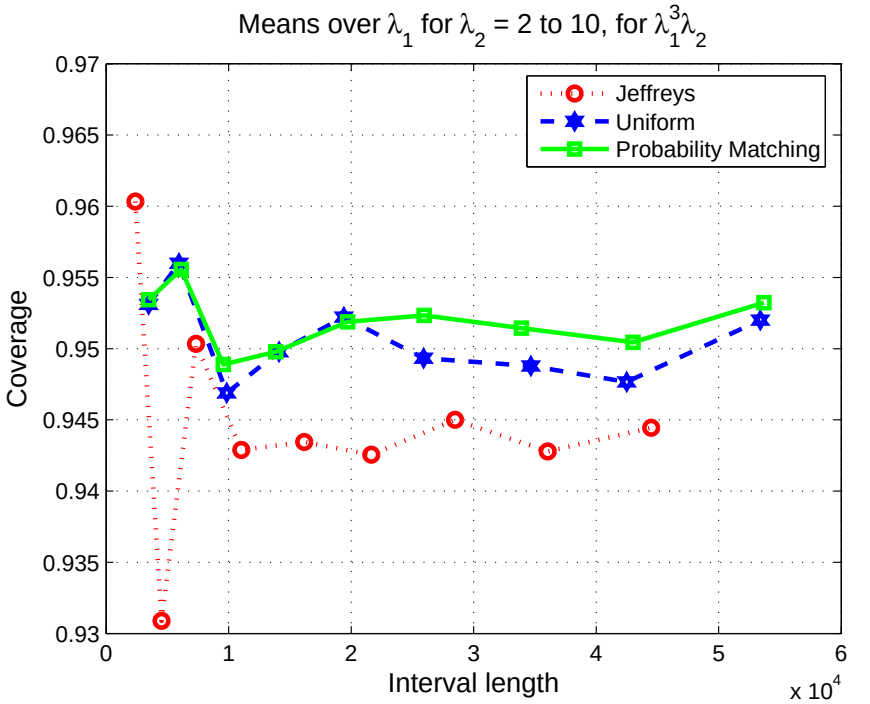


Figure 4.5: Illustration of the Coverage Probabilities of the 95% Bayesian credible intervals for $\xi_4 = \lambda_1^3\lambda_2$.

4.6.4 Simulation Study IV - Comparison of the Jeffreys, Uniform, Reference and Probability Matching Priors for $\nu = \lambda_1/\lambda_2$

The comparison of Poisson rates is of great interest in biological, agricultural and medical research. In two sample situations it may be of interest to test or to construct confidence intervals for the ratio of two Poisson rates. Gu et al. (2008) compared the properties of four approaches for testing the ratio of two Poisson rates. They considered asymptotically normal tests, tests based on approximate p - values, exact conditional tests and a likelihood ratio test.

Price & Bonett (2000), on the other hand, computed the exact coverage probabilities of the intervals of six classical methods and that of the Bayesian interval, using the Jeffreys prior, for small and large values of λ_i $(i = 1,2)$. They also looked at other plausible noninformative priors for λ_i such as $\pi(\lambda_i) \propto \lambda_i^{-1}$ and $\pi(\lambda_i) \propto$ constant and mentioned that these priors work almost as well as the Jeffreys prior ($\pi(\lambda_i) \propto \lambda_i^{-\frac{1}{2}}$). According to them the Jeffreys prior has the advantage of adding 0.5 to the sample data which would avoid the problem of sampling zeros. From their simulation studies they concluded that the noninformative Bayesian intervals (using the Jeffreys prior) is reasonable under classical evaluation. We, however, tend to differ from them, since we came to the conclusion that the Jeffreys prior cannot be used for testing the ratio $\nu = \lambda_1/\lambda_2$ or obtaining confidence intervals, especially if λ_2 is small. A prior that can be used for these purposes is the uniform prior. This will become clear from the following simulation study.

It was shown that the posterior distribution of $\nu = \lambda_1/\lambda_2$ in the case of the Jeffreys, probability matching and reference priors is a beta distribution of the second kind. This distribution can easily be transformed to an $F-$distribution with $2x_1 + 1$ and $2x_2 + 1$ degrees of freedom. In a similar way the posterior distribution of ν, using the uniform prior can be transformed to an $F-$distribution with $2x_1 + 2$ and $2x_2 + 2$ degrees of freedom. Bayesian credible intervals and coverage probabilities for ν can therefore be calculated exactly.

In this example a simulation study is done for $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates. The parameter values for the Poisson distributions are λ_1 =2, 5 and 10 and λ_2 =2, 3, 4, 5, 6, 7, 8, 9, 10. The priors that will be compared are:

1. the uniform prior: $\pi_U(\lambda_1, \lambda_2) \propto$ constant;

2. the Jeffreys (reference and probability matching) prior: $\pi_J(\lambda_1, \lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-\frac{1}{2}}$.

In Table 4.2 the coverage percentages are given for the 95% Bayesian equal-tail and HPD intervals in the case of the Jeffreys prior and in Table 4.3 the coverage percentages are given for the 95% Bayesian equal-tail and HPD intervals in the case of the uniform prior. Ten thousand samples were generated for each parameter combination.

Table 4.2: Coverage percentages of the 95% Bayesian credible intervals for $\nu = \lambda_1/\lambda_2$ in the case of the Jeffreys prior. (a) Coverage percentage, (b) mean length and (c) variance of interval length.

Equal-tail intervals										
$\downarrow\lambda_1$	$\lambda_2 \rightarrow$	2	3	4	5	6	7	8	9	10
2	(a)	95.92	94.16	94.46	94.72	95.24	95.20	95.60	95.72	95.30
	(b)	599.98	250.08	102.77	41.956	8.0429	3.7614	4.8311	1.5405	0.8661
	(c)	3.3E+6	1.5E+6	6.1E+5	2.5E+5	32752	8470	24634	1261.9	0.4405
5	(a)	95.40	94.08	94.38	94.94	94.76	94.30	94.32	95.06	94.40
	(b)	1381.8	599.82	208.86	87.282	23.004	13.445	6.6475	1.8303	1.5217
	(c)	1.5E+7	7.2E+6	2.5E+6	1.0E+6	2.4E+5	1.1E+5	46513	4.5629	1.5058
10	(a)	95.40	94.56	94.76	94.16	93.82	94.32	94.92	94.80	95.20
	(b)	2886.3	1054.9	3912.1	1727.3	58.639	13.765	10.434	11.167	2.4904
	(c)	5.7E+7	2.3E+7	8.9E+6	3.5E+6	1.2E+6	2.3E+5	2.1E+5	3.2E+5	3.5445
HPD intervals										
$\downarrow\lambda_1$	$\lambda_2 \rightarrow$	2	3	4	5	6	7	8	9	10
2	(a)	94.70	94.08	94.60	94.44	94.52	93.90	94.48	93.86	94.24
	(b)	154.31	65.525	27.853	12.064	3.1596	1.8068	1.9247	0.9932	0.7294
	(c)	2.0E+5	94 520	38 760	15 880	2 270.1	467.92	1 605.1	90.013	0.2881
5	(a)	95.18	94.80	95.28	94.08	94.84	94.84	94.86	95.26	94.36
	(b)	354.20	156.11	56.186	24.835	8.2267	4.9627	3.0205	1.5732	1.3301
	(c)	9.5E+5	4.5E+5	1.5E+5	62 546	17 057	6 322.3	2 888.1	2.2182	0.9420
10	(a)	94.92	94.88	95.20	95.00	94.74	94.88	95.18	95.62	95.32
	(b)	741.52	276.47	104.83	48.109	19.005	6.2733	4.7505	4.4775	2.2129
	(c)	3.6E+5	1.4E+5	5.5E+5	2.1E+5	78 638	11 609	1 071.7	15.581	2.2795

From Table 4.2 it is clear that the coverage percentages of the Jeffreys (reference and probability matching) priors are reasonably good (slight under coverage in some cases) but the mean lengths and variances of the Bayesian credible intervals are much too large. This is especially true if λ_2 is small. The uniform prior, on the other hand, also gives reasonably good coverage (slight over coverage) but the mean lengths and variances of the credibility intervals are much smaller.

Table 4.3: Coverage percentages of the 95% Bayesian credible intervals for $\nu = \lambda_1/\lambda_2$ in the case of the uniform prior. (a) Coverage percentage, (b) mean length and (c) variance of interval length.

Equal-tail intervals										
$\downarrow\lambda_1$	$\lambda_2 \rightarrow$	2	3	4	5	6	7	8	9	10
2	(a)	98.16	97.34	96.58	96.38	96.80	96.82	96.06	96.18	96.22
	(b)	21.572	10.130	5.4558	3.2781	2.1531	1.6649	1.3948	1.0562	0.9313
	(c)	1901.0	688.59	278.14	107.50	34.622	21.250	22.322	0.9839	3.4559
5	(a)	95.86	96.46	95.70	95.62	95.92	95.68	95.56	96.26	95.94
	(b)	42.687	20.474	10.341	6.1223	3.6942	2.7507	2.2019	1.7188	1.4644
	(c)	6813.1	2935.1	1075.7	415.72	83.159	38.849	38.722	2.0315	1.1217
10	(a)	96.22	95.76	96.00	95.76	94.84	95.28	95.54	95.90	95.44
	(b)	83.653	38.105	18.813	10.916	6.8133	4.9464	3.6860	3.0692	2.3601
	(c)	23576	9702.6	3713.5	1481.8	570.94	233.67	96.470	158.94	3.2491
HPD intervals										
$\downarrow\lambda_1$	$\lambda_2 \rightarrow$	2	3	4	5	6	7	8	9	10
2	(a)	97.84	97.32	98.42	98.28	98.58	98.46	98.62	98.70	98.68
	(b)	11.973	6.1556	3.6194	2.3565	1.6559	1.3211	1.1222	0.8942	0.7874
	(c)	451.92	165.09	70.507	26.729	9.3764	5.6677	5.6862	0.5540	1.0442
5	(a)	93.76	94.96	94.92	96.34	96.04	96.10	96.12	96.22	96.30
	(b)	23.666	12.28	6.8340	4.4143	2.9397	2.2680	1.8536	1.5016	1.2941
	(c)	1 603.5	706.45	267.98	105.99	24.740	11.811	11.087	1.2281	0.7298
10	(a)	93.48	94.44	94.96	94.84	94.62	94.64	95.42	96.36	95.92
	(b)	45.898	22.734	12.391	7.7628	5.2379	3.9842	3.1191	2.6106	2.1101
	(c)	5 557.1	2 337.2	923.02	377.39	147.07	61.495	27.084	37.362	2.0345

4.7 Conclusion

In this chapter the probability matching prior for the product of different powers of k Poisson rates, $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, was derived. The reference prior for the ratio of two Poisson rates was also derived. We considered a number of simulation studies. If one is interested in obtaining point estimates and Bayesian credible intervals for $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers Poisson rates and if $a_i \geq 0$ $(i = 1, 2, \ldots, k)$, then the probability matching prior is the best. If, on the other hand, one wants to obtain point estimates, credibility intervals or do hypothesis testing about $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates, then the uniform prior should be used.

Price & Bonett (2000) mentioned that the Jeffreys prior has the advantage of adding 0.5 to the sample data which will avoid sampling zeros. From our research it seems that adding 1 to the sample data (using the uniform prior) gives better results for $\nu = \lambda_1/\lambda_2$.

Chapter 5

Estimation for Linear Functions of Poisson Rates

5.1 Introduction

Research has been done on improving confidence intervals for discrete data. Barker (2002) made an attempt to find approximate confidence intervals for a single Poisson rate. Stamey & Hamilton (2006) considered four interval estimators for linear functions of Poisson rates, a Wald interval, a t interval with Satterthwaite's degrees of freedom and two Bayesian intervals using noninformative priors. We will consider another Bayesian interval using a probability matching prior. The probability matching prior will be derived by using the method proposed by Datta & Ghosh (1995).

Krishnamoorthy & Thomson (2004) addressed the problem of hypothesis testing about two Poisson means. They compared the conditional test (C - test) to a test based on estimated p - values (E - test). Krishnamoorthy & Thomson (2004) considered the size and the power of these tests. We will use Bayesian methods, using the Jeffreys prior, the probability matching prior and two other priors. The results obtained from the Bayesian methods will be compared to the results obtained by Krishnamoorthy & Thomson (2004).

The probability matching prior for a linear contrast of Poisson parameters is derived. This prior will be extended in such a way that it can be used as the probability matching prior for the average of Poisson parameters. A weighted Monte Carlo simulation method will be used to obtain Bayesian credible intervals in the case of the probability matching prior. An example and simulation studies will be considered.

5.2 Probability Matching Prior for a Linear Contrast of Poisson Parameters

Consider a sample from k Poisson populations. Let X_i be an observation from population i. Then $X_1, X_2, \ldots, X_k$ will be independent Poisson distributions such that $X_i \sim P(\lambda_i)$, for $i = 1, 2, \ldots, k$. Where λ_i is the expected number of events per unit sample. We assume that the interest is in a linear combination of Poisson rates. In general, we can define such a linear function of Poisson parameters as $\delta = \sum_{i=1}^{k} a_i \lambda_i$, where a_i is the known coefficient value.

Theorem 5.1. The probability matching prior for $\delta = \sum_{i=1}^{k} a_i \lambda_i$, a linear contrast of Poisson parameters (i.e. $\sum_{i=1}^{k} a_i = 0$), is given by

$$\pi_{PM}(\underline{\lambda}) = \pi_{PM}(\lambda_1, \lambda_2, \ldots, \lambda_k) \propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}}. \tag{5.1}$$

Proof. Assume that $X_1, X_2, \ldots, X_k$ are independent Poisson random variables with $X_i \sim P(\lambda_i)$. The likelihood function is given by

$$L(\underline{\lambda} | x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} e^{-\lambda_i} \frac{\lambda_i^{x_i}}{x_i!}.$$

It was shown in Appendix C in Theorem C.1 that the inverse of the Fisher information matrix is given by

$$F^{-1}(\underline{\lambda}) = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}.$$

We are interested in a probability matching prior for $t(\underline{\lambda}) = \delta = \sum_{i=1}^{k} a_i \lambda_i$, a linear contrast of Poisson parameters, where $\sum_{i=1}^{k} a_i = 0$.

Now

$$\begin{aligned} \nabla_t'(\underline{\lambda}) &= \begin{bmatrix} \frac{\partial t(\underline{\lambda})}{\partial \lambda_1} & \frac{\partial t(\underline{\lambda})}{\partial \lambda_2} & \cdots & \frac{\partial t(\underline{\lambda})}{\partial \lambda_k} \end{bmatrix} \\ &= \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix}. \end{aligned}$$

Also

$$\begin{aligned}\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda}) &= \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix} \\ &= \begin{bmatrix} a_1\lambda_1 & a_2\lambda_2 & \cdots & a_k\lambda_k \end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda}) &= \begin{bmatrix} a_1\lambda_1 & a_2\lambda_2 & \cdots & a_k\lambda_k \end{bmatrix} \\ &\quad \times \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} \\ &= \sum_{i=1}^{k} a_i^2\lambda_i.\end{aligned}$$

Define

$$\begin{aligned}\eta'(\underline{\lambda}) &= \frac{\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})}{\sqrt{\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda})}} \\ &= \begin{bmatrix} \frac{a_1\lambda_1}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i}} & \frac{a_2\lambda_2}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i}} & \cdots & \frac{a_k\lambda_k}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i}} \end{bmatrix} \\ &= \begin{bmatrix} \eta_1(\underline{\lambda}) & \eta_2(\underline{\lambda}) & \cdots & \eta_k(\underline{\lambda}) \end{bmatrix}.\end{aligned}$$

As before the prior $\pi(\underline{\lambda})$ is a probability matching prior if and only if the differential equation $\sum_{i=1}^{k} \frac{\partial}{\partial\lambda_i}\{\eta_i(\underline{\lambda})\pi(\underline{\lambda})\} = 0$ is satisfied.
Let

$$\pi(\underline{\lambda}) = \left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}}$$

then

$$\begin{aligned}\eta_1(\underline{\lambda})\,\pi(\underline{\lambda}) &= \frac{a_1\lambda_1}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i}}\left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \\ &= a_1\lambda_1\end{aligned}$$

therefore

$$\frac{\partial}{\partial\lambda_1}\{\eta_1(\underline{\lambda})\,\pi(\underline{\lambda})\} = \frac{\partial}{\partial\lambda_1}\{a_1\lambda_1\} = a_1$$

and

$$\begin{aligned}\eta_2(\underline{\lambda})\,\pi(\underline{\lambda}) &= \frac{a_2\lambda_2}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i}}\left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \\ &= a_2\lambda_2\end{aligned}$$

therefore

$$\frac{\partial}{\partial\lambda_2}\{\eta_2(\underline{\lambda})\,\pi(\underline{\lambda})\} = \frac{\partial}{\partial\lambda_2}\{a_2\lambda_2\} = a_2$$

and

$$\begin{aligned}\eta_k(\underline{\lambda})\,\pi(\underline{\lambda}) &= \frac{a_k\lambda_k}{\sqrt{\sum_{i=1}^{k} a_i^2\lambda_i}}\left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \\ &= a_k\lambda_k\end{aligned}$$

therefore

$$\frac{\partial}{\partial\lambda_k}\{\eta_k(\underline{\lambda})\,\pi(\underline{\lambda})\} = \frac{\partial}{\partial\lambda_k}\{a_k\lambda_k\} = a_k.$$

We can therefore conclude that

$$\sum_{i=1}^{k}\frac{\partial}{\partial\lambda_i}\{\eta_i(\underline{\lambda})\,\pi(\underline{\lambda})\} = a_1 + a_2 + \ldots + a_k = 0$$

since we are dealing with a linear contrast, i.e. $\sum_{i=1}^{k} a_i = 0$.

The differential equation will be satisfied if $\pi(\underline{\lambda})$ is

$$\pi_{PM}(\underline{\lambda}) \propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}}.$$

□

When using the probability matching prior, the joint posterior distribution of $\underline{\lambda}$ is given by

$$\begin{aligned} \pi_{PM}(\underline{\lambda}\,|data) &\propto \pi_{PM}(\underline{\lambda}) \times L(\underline{\lambda}\,|data) \\ &\propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}} \times \prod_{i=1}^{k} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} \\ \therefore \pi_{PM}(\underline{\lambda}\,|data) &\propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}. \end{aligned} \tag{5.2}$$

Corollary 5.1. If $\sum_{i=1}^{k} a_i \neq 0$, the following equation can be used for a probability matching prior

$$\pi_{PM*}(\underline{\lambda}) \propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{-1} \tag{5.3}$$

and the posterior distribution of $\underline{\lambda}$ will be

$$\pi_{PM*}(\underline{\lambda}\,|data) \propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i - 1} e^{-\lambda_i}. \tag{5.4}$$

The Jeffreys prior is given by

$$\pi_J(\underline{\lambda}) \propto |F(\underline{\lambda})|^{\frac{1}{2}} = \left(\prod_{i=1}^{k} \lambda_i \right)^{-\frac{1}{2}} \tag{5.5}$$

where $F(\underline{\lambda})$ is the Fisher information matrix connected with the likelihood function.

When using the Jeffreys prior, the posterior distribution of $\underline{\lambda}$ is given by

$$\begin{aligned}\pi_J(\underline{\lambda}\,|data) &\propto \pi_J(\underline{\lambda}) \times L(\underline{\lambda}\,|data) \\ &\propto \left(\prod_{i=1}^{k}\lambda_i\right)^{-\frac{1}{2}} \prod_{i=1}^{k}\lambda_i^{x_i}e^{-\lambda_i} \\ \therefore \pi_J(\underline{\lambda}\,|data) &\propto \prod_{i=1}^{k}\lambda_i^{x_i-\frac{1}{2}}e^{-\lambda_i}. \end{aligned} \tag{5.6}$$

The posterior distribution of $\underline{\lambda}$ is thus the product of k independently distributed *Gamma* $\left(x_i+\frac{1}{2},1\right)$ variates.

In some cases the exact posterior distribution of $\delta=\sum_{i=1}^{k}a_i\lambda_i$ can be derived. For example if $a_1=a_2=\ldots=a_k=\frac{1}{k}$. The following theorem can now be proved.

Theorem 5.2. If $\pi_J(\underline{\lambda}\,|data)=\prod_{i=1}^{k}\frac{e^{-\lambda_i}\lambda_i^{x_i-\frac{1}{2}}}{\Gamma\left(x_i+\frac{1}{2}\right)}$, then the posterior distribution of $\widetilde{\delta}=\frac{1}{k}\sum_{i=1}^{k}\lambda_i$ is

$$\pi_J\left(\widetilde{\delta}\,|data\right)=\frac{k^{\sum_{i=1}^{k}x_i+\frac{k}{2}}}{\Gamma\left(\sum_{i=1}^{k}x_i+\frac{k}{2}\right)}\widetilde{\delta}^{\sum_{i=1}^{k}x_i+\frac{k}{2}-1}e^{-k\widetilde{\delta}}. \tag{5.7}$$

Proof. $\delta=\sum_{i=1}^{k}a_i\lambda_i$. The moment generating function of δ is

$$\begin{aligned}M_\delta(t) &= E\left(e^{\delta t}\right)=E\left(e^{t\sum_{i=1}^{k}a_i\lambda_i}\right) \\ &= \left\{\prod_{i=1}^{k}\frac{1}{\Gamma\left(x_i+\frac{1}{2}\right)}\right\}\int_0^\infty\cdots\int_0^\infty e^{t\sum_{i=1}^{k}a_i\lambda_i}\prod_{i=1}^{k}\left\{e^{-\lambda_i}\lambda_i^{x_i+\frac{1}{2}-1}\right\}d\lambda_1\ldots d\lambda_k \\ &= C\left(\int_0^\infty e^{ta_1\lambda_1}e^{-\lambda_1}\lambda_1^{x_1+\frac{1}{2}-1}d\lambda_1\right)\cdots\left(\int_0^\infty e^{ta_k\lambda_k}e^{-\lambda_k}\lambda_k^{x_k+\frac{1}{2}-1}d\lambda_k\right)\end{aligned}$$

where

$$C = \left\{\prod_{i=1}^{k}\frac{1}{\Gamma\left(x_i+\frac{1}{2}\right)}\right\}.$$

Consider

$$\begin{aligned}I &= \int_0^\infty e^{ta_i\lambda_i}e^{-\lambda_i}\lambda_i^{x_i+\frac{1}{2}-1}d\lambda_i \\ &= \int_0^\infty e^{-\lambda_i(1-a_it)}\lambda_i^{x_i+\frac{1}{2}-1}d\lambda_i.\end{aligned}$$

Let $\lambda_i(1-a_it) = y, \therefore \lambda_i = \left(\frac{1}{1-a_it}\right) y$ and $d\lambda_i = \left(\frac{1}{1-a_it}\right) dy$.

$$\begin{aligned} \therefore I &= \int_0^\infty e^{-y} \left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}-1} y^{x_i+\frac{1}{2}-1} \left(\frac{1}{1-a_it}\right) dy \\ &= \left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}} \int_0^\infty e^{-y} y^{x_i+\frac{1}{2}-1} dy \\ &= \left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}} \Gamma\left(x_i+\frac{1}{2}\right). \end{aligned}$$

Therefore $M_\delta(t) = \prod_{i=1}^k \left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}}$.
If $a_1 = a_2 = \ldots = a_k = \frac{1}{k}$, then
$M_{\widetilde{\delta}}(t) = \left(\frac{k}{k-t}\right)^{\sum_{i=1}^k x_i+\frac{k}{2}}$, which is the moment generating function of a Gamma distribution.
Therefore

$$\pi_J\left(\widetilde{\delta}\,|\text{data}\right) = \frac{k^{\sum_{i=1}^k x_i+\frac{k}{2}}}{\Gamma\left(\sum_{i=1}^k x_i+\frac{k}{2}\right)} \widetilde{\delta}^{\sum_{i=1}^k x_i+\frac{k}{2}-1} e^{-k\widetilde{\delta}} \qquad 0 < \widetilde{\delta} < \infty.$$

□

5.3 The Weighted Monte Carlo Method in the Case of the Probability Matching Prior for $\delta = \sum_{i=1}^k a_i\lambda_i$

This method has been introduced in Section 2.4.

If a uniform prior is put on $\underline{\lambda}$, the posterior (proposal) distribution is

$$q(\underline{\lambda}\,|data) \propto \prod_{i=1}^k \lambda_i^{x_i} e^{-\lambda_i}.$$

In the case of the probability matching prior, using Equations 5.2 and 5.4, the posterior (target) distribution is

$\pi_{PM}(\underline{\lambda}\,|data) \propto \left(\sum_{i=1}^k a_i^2\lambda_i\right)^{\frac{1}{2}} \prod_{i=1}^k \lambda_i^{x_i} e^{-\lambda_i}$, if $\sum_{i=1}^k a_i = 0$,
or
$\pi_{PM*}(\underline{\lambda}\,|data) \propto \left(\sum_{i=1}^k a_i^2\lambda_i\right)^{\frac{1}{2}} \prod_{i=1}^k \lambda_i^{x_i-1} e^{-\lambda_i}$, if $\sum_{i=1}^k a_i \neq 0$.

The sample probabilities are therefore proportional to

$$\frac{\pi_{PM}(\underline{\lambda}\,|data)}{q(\underline{\lambda}\,|data)} = \left(\sum_{i=1}^{k} a_i^2\lambda_i\right)^{\frac{1}{2}} \qquad \text{if} \quad \textstyle\sum_{i=1}^{k} a_i = 0,$$

or

$$\frac{\pi_{PM*}(\underline{\lambda}\,|data)}{q(\underline{\lambda}\,|data)} = \left(\sum_{i=1}^{k} a_i^2\lambda_i\right)^{\frac{1}{2}} \prod_{i=1}^{k}\lambda_i^{-1} \qquad \text{if} \quad \textstyle\sum_{i=1}^{k} a_i \neq 0.$$

The normalised weights are

$$\omega_l = \frac{\left\{\sum\limits_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}}}{\sum\limits_{l=1}^{n}\left[\left\{\sum\limits_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}}\right]} \qquad \text{if} \quad \textstyle\sum_{i=1}^{k} a_i = 0 \quad \text{for } l = 1,2,\ldots,n,$$

or

$$\omega_l = \frac{\left\{\sum\limits_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}} \prod\limits_{i=1}^{k}\left(\lambda_i^{-1}\right)^{(l)}}{\sum\limits_{l=1}^{n}\left[\left\{\sum\limits_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}} \prod\limits_{i=1}^{k}\left(\lambda_i^{-1}\right)^{(l)}\right]} \qquad \text{if} \quad \textstyle\sum_{i=1}^{k} a_i \neq 0 \quad \text{for } l = 1,2,\ldots,n$$

where n is the number of simulations.
The Monte Carlo method:

- **Step 1**

  Obtain a Monte Carlo sample $\left\{\left(\lambda_1^{(l)}, \lambda_2^{(l)} \ldots, \lambda_k^{(l)}\right); l = 1,2,\ldots,n\right\}$ from the proposal distribution $q(\underline{\lambda}\,|data)$ and calculate $\delta^{(l)} = \sum\limits_{i=1}^{k} a_i\lambda_i^{(l)}$ for $l = 1,2,\ldots,n$.

- **Step 2**

  Sort $\left\{\delta^{(l)}, (l = 1,2,\ldots,n)\right\}$ to obtain the ordered values $\delta^{[1]} \leq \delta^{[2]} \leq \cdots \leq \delta^{[n]}$.

- **Step 3**

  Each simulated δ value has an associated weight. Therefore compute the weighted function $\omega_{(l)}$ associated with the l^{th} ordered $\delta^{[l]}$ value.

- **Step 4**

  Add the weights up from left to right (from the first on) until one obtains $\sum_{l=1}^{n_1} \omega_{(l)} = \alpha/2$. Write down the corresponding $\delta^{[n_1]}$ value and denote it as $\delta_{(\alpha/2)}$. Add the weights up from right to left (from the last back) until one obtains $\sum_{l=n_2}^{n} \omega_{(l)} = \alpha/2$. Write down the corresponding $\delta^{[n_2]}$ value and denote it as $\delta_{(1-\alpha/2)}$.

- **Step 5**

  The $100\,(1-\alpha)\,\%$ Bayesian credible interval is: $\left(\delta_{(\alpha/2)}, \delta_{(1-\alpha/2)}\right)$.

5.4 Example and Simulation Studies

5.4.1 Example

Stamey & Hamilton (2006) considered an example where they compared four intervals. We are going to compare two Bayesian intervals, using the Jeffreys prior and a probability matching prior, with the four intervals from Stamey & Hamilton (2006). They considered the number of fatal motor vehicle accidents involving driving while intoxicated (DWI) during six major holidays for the year 2000. They obtained the data from the Crash Records Bureau of the Texas Department of Public Safety. The data are given in Table 5.1.

Stamey & Hamilton (2006) used the following methods: a Wald interval, a *t* interval with Satterthwaite's degrees of freedom, and two Bayesian intervals using noninformative priors. In Table 5.2 the four methods used by Stamey & Hamilton (2006) to obtain 95% confidence intervals can be seen, as well as the two Bayesian methods that we considered. The purpose of this experiment was to estimate the number of DWI involved fatal accidents per holiday, and also to see if less accidents occur during the summer holidays than during the winter holidays. Please note that the data are taken from the Crash Records Bureau of the Texas Department of Public Safety, the summer holidays are therefore Memorial Day, July 4 and Labour Day, and the winter holidays are Thanksgiving, Christmas and New Year's Eve. Table 5.2 indicates the 95% confidence intervals for the two linear functions.

Table 5.1: Number of DWI involved fatal motor vehicle accidents during six major holidays (2000).

Holiday	**No. of Accidents**
Memorial Day	0
July 4	5
Labour Day	2
Thanksgiving	11
Christmas	8
New Year's Eve	9

Table 5.2: 95% confidence intervals for the contrasts for DWI - involved fatal motor vehicle accidents.

Contrast	**Wald**	**Student's *t***	**Bayes**	**Bayes with *t***
Average number of DWI accidents per holiday $\underline{a} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$	$(3.9, 7.77)$	$(3.87, 7.80)$	$(4.31, 8.35)$	$(4.29, 8.38)$
Winter vs. Summer $\underline{a} = (-1/3, -1/3, -1/3, 1/3, 1/3, 1/3)$	$(3.13, 10.87)$	$(3.07, 10.93)$	$(2.97, 11.03)$	$(2.91, 11.09)$

Contrast	**Bayes Jef**	**Bayes PMP**	**Bayes PMP***
Average number of DWI accidents per holiday $\underline{a} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$	$(4.48, 8.48)$		$(4.21, 8.09)$
Winter vs. Summer $\underline{a} = (-1/3, -1/3, -1/3, 1/3, 1/3, 1/3)$	$(3.14, 11.20)$	$(3.19, 11.34)$	$(3.17, 11.07)$

From Table 5.2 it can be seen that the Wald interval and the Student's t interval imply the average number of DWI - involved fatal accidents per holiday do not exceed four, while the Bayesian intervals imply that the average number exceeds four. The upper limits of the Bayesian methods exceed eight, while the Wald and Student's t upper limits are less than eight. For the contrast between the summer holidays and the winter holidays, it is noted that all the intervals are considerably greater than zero, which indicates that fatal accidents were more common in winter than in the summer. This was also noted by Stamey & Hamilton (2006). The Bayesian methods that we suggested compare well with the other intervals.

5.4.2 Simulation Study I

In this section we will look into the expected widths and the coverage probabilities of six methods for constructing confidence intervals. To examine the coverage percentages the following simulation procedure was proposed by Stamey & Hamilton (2006). They first created Poisson means λ_i, $i = 1, \ldots, k$, from a uniform distribution on the interval $(0, 5)$, for a given number of theoretical populations k. They then simulated $X_i \sim P(\lambda_i)$, $i = 1, \ldots, k$, and compared the confidence intervals for each of the four methods based on the drawn observations and on the specified contrast coefficients a_i, $i = 1, \ldots, k$. To obtain the coverage probabilities the percentages of times over 100 000 draws that each confidence interval contains the true value of the contrast $\delta = \sum_{i=1}^{k} a_i \lambda_i$ were calculated and to obtain the expected

widths, the average width of each interval was calculated.

Table 5.3: (a) Average coverage probabilities and (b) average widths for contrasts where $\lambda_i \in (0, 5)$.

Contrast		Wald	Student's *t*	Bayes	Bayes with *t*	Bayes Jef	Bayes PMP	Bayes PMP*
$k = 2$								
$(1, -1)$	(a)	91.1%	98.3%	96.9%	99.1%	96.1%	95.66%	95.32%
	(b)	8.40	10.09	9.28	11.27	9.44	10.16	9.59
$\left(\frac{1}{2}, \frac{1}{2}\right)$	(a)	91.2%	97.8%	96.7%	99.2%	95.6%		96.22%
	(b)	4.12	5.08	4.64	5.63	4.53		4.46
$k = 3$								
$\left(1, -\frac{1}{2}, -\frac{1}{2}\right)$	(a)	93.5%	96.4%	96.5%	98.2%	95.9%	96.06%	95.19%
	(b)	7.24	8.31	8.05	9.32	8.20	8.68	8.16
$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	(a)	91.6%	95.7%	96.7%	98.3%	94.8%		95.61%
	(b)	3.46	3.85	3.83	4.28	3.80		3.64
$k = 4$								
$\left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$	(a)	94.8%	96.7%	97.1%	98.2%	95.5%	96.16%	95.49%
	(b)	6.05	6.51	6.68	7.20	6.81	7.05	6.59
$\left(1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right)$	(a)	92.2%	94.8%	95.9%	97.5%	95.5%	96.44%	95.76%
	(b)	6.74	7.78	7.55	8.73	7.73	8.15	7.65
$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$	(a)	92.7%	95.1%	96.1%	97.4%	93.4%		95.73%
	(b)	3.02	3.25	3.38	3.60	3.33		3.14
$k = 5$								
$\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$	(a)	93.1%	94.8%	95.5%	96.6%	93.7%		95.52%
	(b)	2.72	2.87	3.00	3.17	2.99		2.82
$\left(1, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}\right)$	(a)	91.2%	93.8%	95.4%	97.1%	94.6%	96.88%	95.85%
	(b)	6.48	7.56	7.28	8.50	7.39	7.80	7.34
$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{2}, -\frac{1}{2}\right)$	(a)	94.6%	96.1%	96.9%	97.8%	96.7%	96.21%	95.56%
	(b)	5.53	5.89	6.10	6.51	6.21	6.38	5.99

From Table 5.3 it can be seen that the Wald interval is overall the poorest performer when all the Poisson rates are expected to be small, because the coverage never reached 95%. Stamey & Hamilton (2006) used the *t* distribution, to widen the intervals, but also did not get completely satisfactory results. When using the student's *t* distribution, they still got coverage results that are below nominal, but in most cases they were close to nominal, except for the case where $k = 5$. According to Stamey & Hamilton (2006) the Bayesian procedure based on the Jeffreys noninformative prior performed the best for the small-rate cases. The coverage is above or at nominal for every case. The interval is also in many cases narrower on average than the interval based on the student's *t*. The interval based on student's *t* using the Bayesian prior estimator also has coverage above nominal in every case, but is wider than the Bayesian interval using the standard normal coefficient. The interval procedures using the Jeffreys and probability matching priors, the last three columns from Table 5.3, compare well with

the other procedures. The average widths also compare well with the other procedures. In Table 5.3, the column Bayes PMP gives the results when the prior, $\pi_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i\right\}^{\frac{1}{2}}$, is used and the column Bayes PMP* gives the results when the prior, $\pi_{PM*}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{-1}$, is used.

Stamey & Hamilton (2006) also considered another simulation study, to see what impact larger expected counts would have on the intervals. They used exactly the same methods as in the previous simulation study, the only difference is that in this case the Poisson rates were generated from a uniform distribution on the interval $(5, 10)$. In the previous simulation study the Poisson rates were generated from a uniform distribution on the interval $(0, 5)$. The results (coverage percentages and interval widths) calculated by them are given in the first four columns of Table 5.4. In the last three columns the results are given when using the Jeffreys and probability matching priors. In Table 5.4, the column Bayes PMP gives the results when the prior, $\pi_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i\right\}^{\frac{1}{2}}$, is used and the column Bayes PMP* gives the results when the prior, $\pi_{PM*}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{-1}$, is used.

Table 5.4: (a) Average coverage probabilities and (b) average widths for contrasts where $\lambda_i \in (5, 10)$.

Contrast		**Wald**	**Student's *t***	**Bayes**	**Bayes with *t***	**Bayes Jef**	**Bayes PMP**	**Bayes PMP***
				***k* = 2**				
$(1,-1)$	(a)	95.1%	96.1%	95.9%	96.7%	94.8%	94.76%	93.48%
	(b)	15.01	15.68	15.53	16.22	15.69	16.24	15.66
$\left(\frac{1}{2},\frac{1}{2}\right)$	(a)	93.4%	94.6%	95.3%	96.5%	94.2%		95.09%
	(b)	7.51	7.84	7.76	8.11	7.70		7.58
				***k* = 3**				
$\left(1,-\frac{1}{2},-\frac{1}{2}\right)$	(a)	94.5%	95.5%	95.4%	96.3%	95.6%	95.02%	93.88%
	(b)	13.00	13.55	13.44	14.03	13.43	13.87	13.39
$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	(a)	93.9%	94.7%	95.5%	96.1%	95.9%		95.06%
	(b)	6.46	6.33	6.36	6.54	6.28		6.20
				***k* = 4**				
$\left(\frac{1}{2},\frac{1}{2},-\frac{1}{2},-\frac{1}{2}\right)$	(a)	95.1%	95.6%	95.8%	96.3%	95.5%	95.00%	94.52%
	(b)	10.68	10.91	11.04	11.26	11.12	11.27	10.91
$\left(1,-\frac{1}{3},-\frac{1}{3},-\frac{1}{3}\right)$	(a)	94.0%	95.1%	94.7%	95.8%	94.1%	95.00%	94.66%
	(b)	12.24	12.85	12.67	13.29	12.80	13.06	12.60
$\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$	(a)	94.2%	94.8%	95.4%	95.9%	94.8%		94.71%
	(b)	5.34	5.45	5.52	5.63	5.50		5.36
				***k* = 5**				
$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right)$	(a)	94.4%	94.8%	95.2%	95.6%	94.2%		94.74%
	(b)	4.78	4.86	4.94	5.02	4.91		4.76
$\left(1,-\frac{1}{4},-\frac{1}{4},-\frac{1}{4},-\frac{1}{4}\right)$	(a)	93.6%	95.0%	94.5%	95.7%	95.5%	95.46%	94.64%
	(b)	11.83	12.49	12.24	12.92	12.19	12.55	12.15
$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3},-\frac{1}{2},-\frac{1}{2}\right)$	(a)	95.0%	95.4%	95.7%	96.2%	95.3%	95.35%	94.42%
	(b)	9.75	9.94	10.08	10.28	10.1	10.27	9.93

As in the previous simulation study the Wald interval is again overall the poorest performer. Stamey & Hamilton (2006) could not clearly state whether the interval based on the Bayesian or the interval based on the student's t performed the best in this case. Both have coverages that are slightly below nominal for some contrasts, but in most cases the coverage is usually quite close to nominal. From the second last column of Table 5.4, the probability matching prior, compares well with the other procedures in the cases where there is a linear contrast of Poisson parameters. This results in coverage at or just above nominal level in each case. The interval widths also compare well with the other procedures used.

5.4.3 Simulation Study II - Comparing Two Poisson Means

Krishnamoorthy & Thomson (2004) considered the problem of hypothesis testing for two Poisson means. They compared the usual conditional test (C - test) to a test based on estimated p - values (E - test). The C - test is due to Przyborowski & Wilenski (1940) and it is based on the conditional distribution of X_1 given X_1+X_2, which follows a binomial distribution whose success probability is a function of the ratio λ_1/λ_2.

Here

$$X_1 = \sum_{i=1}^{n_1} X_{1i} \sim P(n_1\lambda_1),$$

independently distributed of

$$X_2 = \sum_{i=1}^{n_2} X_{2i} \sim P(n_2\lambda_2) \tag{5.8}$$

where $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ are independent samples, respectively from $P(\lambda_1)$ and $P(\lambda_2)$ distributions. Let k_1 and k_2 be the observed values of X_1 and X_2, respectively.

The p -value for testing

$$H_0 : \lambda_1 - \lambda_2 \leq d \quad \text{vs} \quad H_a : \lambda_1 - \lambda_2 > d \tag{5.9}$$

is $P\left(T_{X_1,X_2} \geq T_{k_1,k_2} | H_0\right)$ which involves the unknown parameter λ_2, and where $d \geq 0$ is a given number and

$$T_{X_1,X_2} = \frac{X_1/n_1 - X_2/n_2 - d}{\sqrt{\frac{X_1/n_1}{n_1} + \frac{X_2/n_2}{n_2}}}$$

is the pivot statistic for the testing problem given n_1, k_1, n_2 and k_2.

The observed value of the pivot statistic is given by

$$T_{k_1,k_2} = \frac{k_1/n_1 - k_2/n_2 - d}{\sqrt{\frac{k_1/n_1}{n_1} + \frac{k_2/n_2}{n_2}}}.$$

Krishnamoorthy & Thomson (2004) give the following estimate for λ_2

$$\hat{\lambda}_{2k} = \frac{k_1 + k_2}{n_1 + n_2} - \frac{dn_1}{n_1 + n_2}.$$

Using this $\hat{\lambda}_{2k}$, Krishnamoorthy & Thomson (2004) estimated the p - value by

$$P\left(T_{X_1,X_2} \geq T_{k_1,k_2} \,|H_0\right) = \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} \frac{e^{-n_1\left(\hat{\lambda}_{2k}+d\right)}\left[n_1\left(\hat{\lambda}_{2k}+d\right)\right]^{x_1}}{x_1!} \frac{e^{-n_2\hat{\lambda}_{2k}}\left(n_2\hat{\lambda}_{2k}\right)^{x_2}}{x_2!} \times I\left[T_{X_1,X_2} \geq T_{k_1,k_2}\right]$$

where $I[\cdot]$ denotes the indicator function. For given nominal level α, the test rule is to reject H_0 in Equation 5.9 whenever the estimated p - value is less than α. In view of Equation 5.8, without loss of generality, we can take $n_1 = n_2 = 1$.

Krishnamoorthy & Thomson (2004) compared the C - and E - tests by looking at the size and the power of the tests at different nominal levels and also at different values for λ_1 and λ_2. They found that the E - test is almost exact and that it is more powerful than the C - test. We will compare Bayesian procedures to the tests used in Krishnamoorthy & Thomson (2004). For the Bayesian procedures we will use the Jeffreys prior, the probability matching prior, a third prior: $\pi_A(\underline{\lambda}) \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and a fourth prior: $\pi_B(\underline{\lambda}) \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$.

From Theorem 5.1, the probability matching prior is given by

$$\pi_{PM}(\lambda_1,\lambda_2) \propto \left\{\sum_{i=1}^{2} a_i^2\lambda_i\right\}^{\frac{1}{2}} = \sqrt{\lambda_1+\lambda_2}.$$

When using the probability matching prior, the posterior distribution of λ_1, λ_2 is given by

$$\pi_{PM}(\lambda_1,\lambda_2\,|x_1,x_2) \propto \left\{\sum_{i=1}^{2} a_i^2\lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{2}\lambda_i^{x_i}e^{-\lambda_i}.$$

The Jeffreys prior, π_J, is given by

$$\pi_J(\lambda_1,\lambda_2) \propto |F(\lambda_1,\lambda_2)|^{\frac{1}{2}} = \left(\prod_{i=1}^{2}\lambda_i\right)^{-\frac{1}{2}} = \lambda_1^{-\frac{1}{2}}\lambda_2^{-\frac{1}{2}}$$

where $F(\lambda_1,\lambda_2)$ is the Fisher information matrix connected with the likelihood function.

When using the Jeffreys prior, the posterior distribution of λ_1, λ_2 is given by

$$\pi_J(\lambda_1, \lambda_2 | x_1, x_2) \propto \left(\prod_{i=1}^{2} \lambda_i\right)^{-\frac{1}{2}} \prod_{i=1}^{2} \lambda_i^{x_i} e^{-\lambda_i} = \prod_{i=1}^{2} \lambda_i^{x_i - \frac{1}{2}} e^{-\lambda_i}.$$

The posterior distribution of λ_1, λ_2 is thus the product of 2 independently distributed *Gamma* $\left(x_i + \frac{1}{2}, 1\right)$ variates.

The third prior, π_A, is given by

$$\pi_A(\lambda_1, \lambda_2) \propto \lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}.$$

When using this prior, the posterior distribution of λ_1, λ_2 is given by

$$\pi_A(\lambda_1, \lambda_2 | x_1, x_2) \propto \left(\prod_{i=1}^{2} \lambda_i\right)^{-\frac{1}{4}} \prod_{i=1}^{2} \lambda_i^{x_i} e^{-\lambda_i} = \prod_{i=1}^{2} \lambda_i^{x_i - \frac{1}{4}} e^{-\lambda_i}.$$

The posterior distribution of λ_1, λ_2 is thus the product of 2 independently distributed *Gamma* $\left(x_i + \frac{3}{4}, 1\right)$ variates.

The fourth prior, π_B, is given by

$$\pi_B(\lambda_1, \lambda_2) \propto \lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}.$$

When using this prior, the posterior distribution of λ_1, λ_2 is given by

$$\pi_B(\lambda_1, \lambda_2 | x_1, x_2) \propto \left(\prod_{i=1}^{2} \lambda_i\right)^{-\frac{3}{8}} \prod_{i=1}^{2} \lambda_i^{x_i} e^{-\lambda_i} = \prod_{i=1}^{2} \lambda_i^{x_i - \frac{3}{8}} e^{-\lambda_i}.$$

The posterior distribution of λ_1, λ_2 is thus the product of 2 independently distributed *Gamma* $\left(x_i + \frac{5}{8}, 1\right)$ variates.

Rice (1995) gives the following definition for the size of a test, which is also known as a type I error:

H_0 may be rejected when it is true. Such an error is called a type I error, and its probability is denoted by α.

In Figures 5.1 - 5.3, we compare the size of the tests using Bayesian procedures to the two tests from Krishnamoorthy & Thomson (2004). The Bayesian simulation procedure in the case of the probability matching prior is as discussed in Section 5.3. The size of the tests as a function of $\lambda = \lambda_1 = \lambda_2$ at the three different nominal level under the null hypothesis $H_0 : \lambda_1 - \lambda_2 = 0$ are given in Figures 5.1 to 5.3.

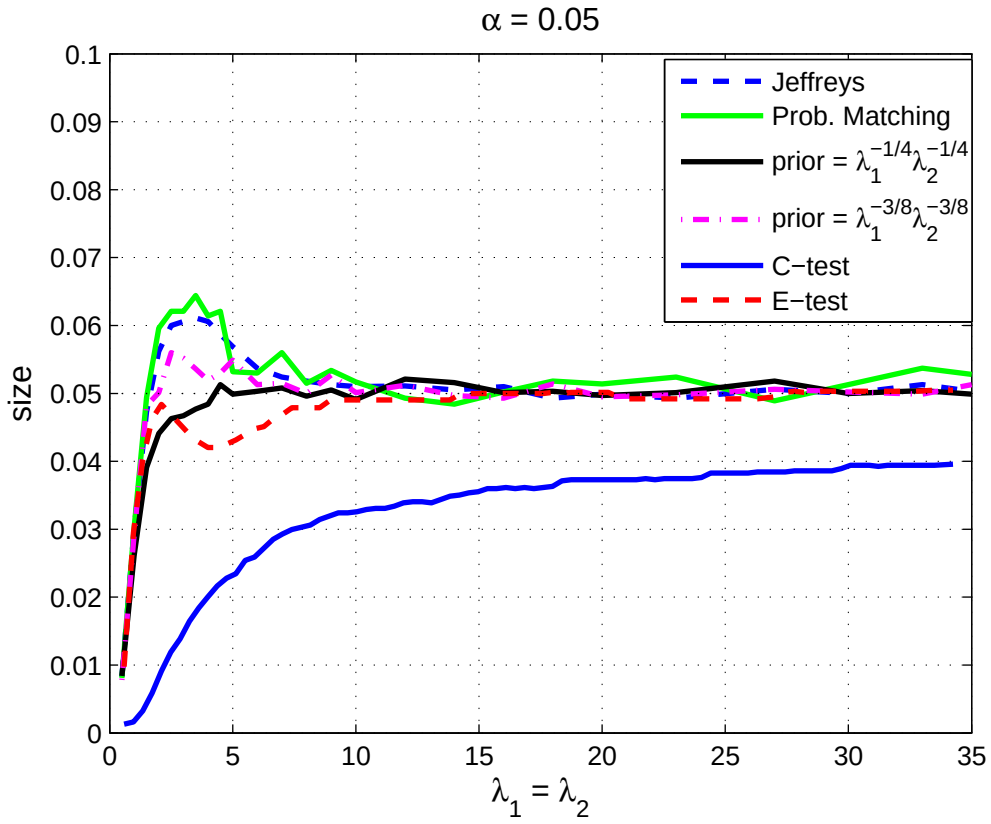


Figure 5.1: Size of the tests at the 5% nominal level.

From Figure 5.1 it can be seen that the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, reaches the nominal level when $\lambda_1 = \lambda_2 = 5$ and from there onwards it attains this level. Whereas the Jeffreys and probability matching priors reach the nominal level at $\lambda_1 = \lambda_2 = 2$, and then only at $\lambda_1 = \lambda_2 = 10$ again, from there onwards it attains this level. The prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$, is an improvement on the Jeffreys and probability matching priors. From Krishnamoorthy & Thomson (2004) the C - test never reaches the nominal level, and the E - test reaches the nominal level only at $\lambda_1 = \lambda_2 = 10$. We must, however, mention that the graphs for the E - and C - tests are scanned in using a MATLAB® program. This means that some small technical errors may occur in the graphs. In general one can say that the Jeffreys and probability matching priors tend to give type I error rates that are somewhat larger than the chosen alpha. The error rates of the priors $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$ seem to be more accurate.

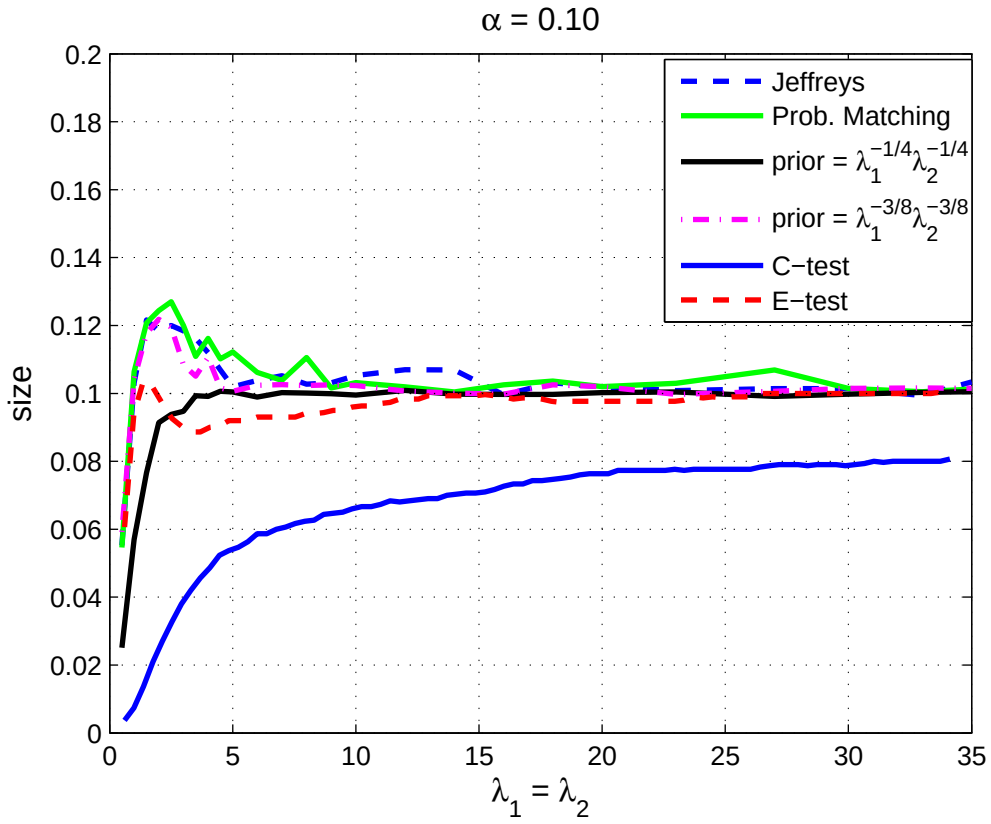


Figure 5.2: Size of the tests at the 10% nominal level.

From Figure 5.2 it can be seen that the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$, reaches the nominal level when $\lambda_1 = \lambda_2 = 4$ and from here onwards it attains this level. Whereas the Jeffreys prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$, reach the nominal level at $\lambda_1 = \lambda_2 = 1$, and then only at $\lambda_1 = \lambda_2 = 5$ again, from there onwards it attains this level. The probability matching prior follows a similar pattern, but it reaches the nominal level at $\lambda_1 = \lambda_2 = 1$, and then only at $\lambda_1 = \lambda_2 = 10$ again, from there onwards it attains this level. From Krishnamoorthy & Thomson (2004) the C - test never reaches the nominal level, and the E - test reaches the nominal level $\lambda_1 = \lambda_2 = 1$, and then only at $\lambda_1 = \lambda_2 = 13$ again, from there onwards it attains this level.

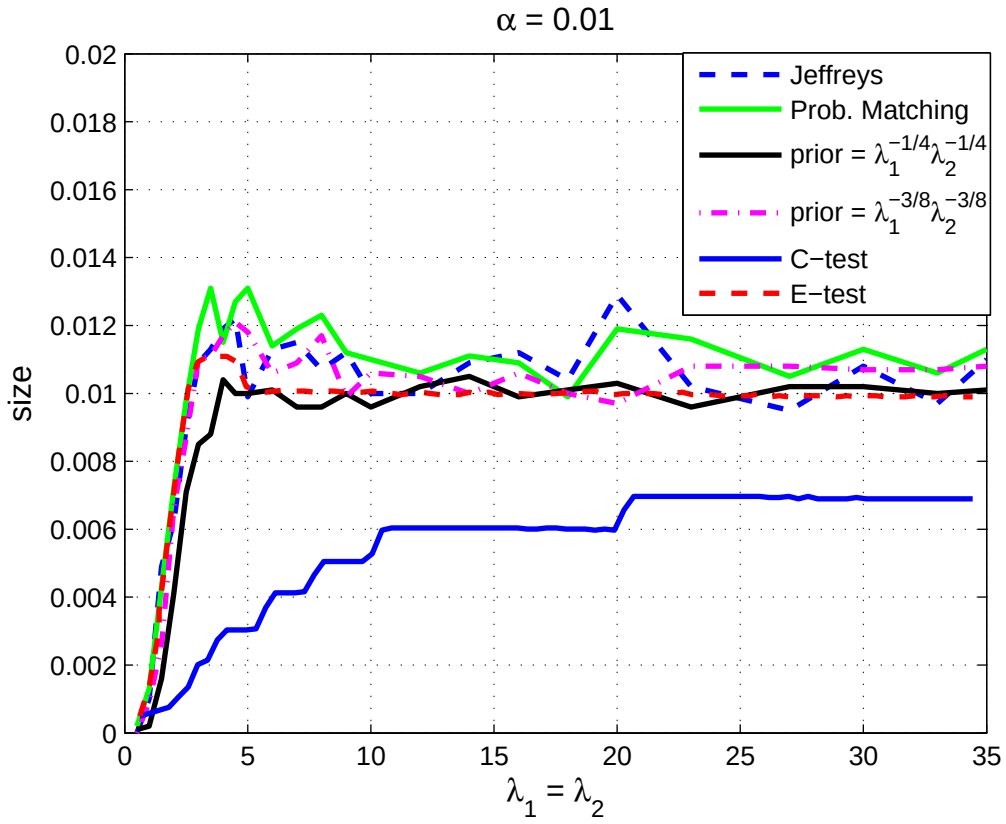


Figure 5.3: Size of the tests at the 1% nominal level.

From Figure 5.3 it can be seen that the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, reaches the nominal level when $\lambda_1 = \lambda_2 = 4.5$ and from here onwards it attains this level. Whereas the Jeffreys and probability matching priors never stay constant at the nominal level, it fluctuates most of the time above the nominal level. The prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$, follows a similar pattern but for a lesser extent than that of the Jeffreys and probability matching priors. Figure 5.3, however, enlarges the fluctuation of the observed error rate. A more direct comparison is to plot the error rate on the same scale. In terms of absolute deviations the Jeffreys and probability matching priors are not performing more poorly as α decreases. In fact the mean deviations for the Jeffreys prior from the nominal α values 0.01, 0.005 and 0.001 are 0.009, 0.0014 and 0.000412 respectively. From Krishnamoorthy & Thomson (2004) the C - test never reaches the nominal level, and the E - test also reaches the nominal level at $\lambda_1 = \lambda_2 = 3$, and then at $\lambda_1 = \lambda_2 = 5$ again, from there onwards it attains this level.

From Figures 5.1 to 5.3, it can be seen that the Bayesian procedures compare relatively well with the E - test from Krishnamoorthy & Thomson (2004). Out of the four Bayesian procedures, the procedures using the priors $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$ give the best results. The C - test is the poorest performer.

In Figures 5.4 to 5.6, we compare the power of the tests using Bayesian procedures to the two tests from Krishnamoorthy & Thomson (2004). The power of the tests as a function of λ_1 at the nominal level $\alpha = 0.05$ under the alternative hypothesis $H_a : \lambda_1 - \lambda_2 > 0$ are given in Figures 5.4 to 5.6.

Rice (1995) gives the following definition for a type II error:

H_0 may be accepted when it is false. Such an error is called a type II error, and its probability is denoted by β.

The probability that H_0 is rejected when it is false is called the power of the test, the power equals $1 - \beta$.

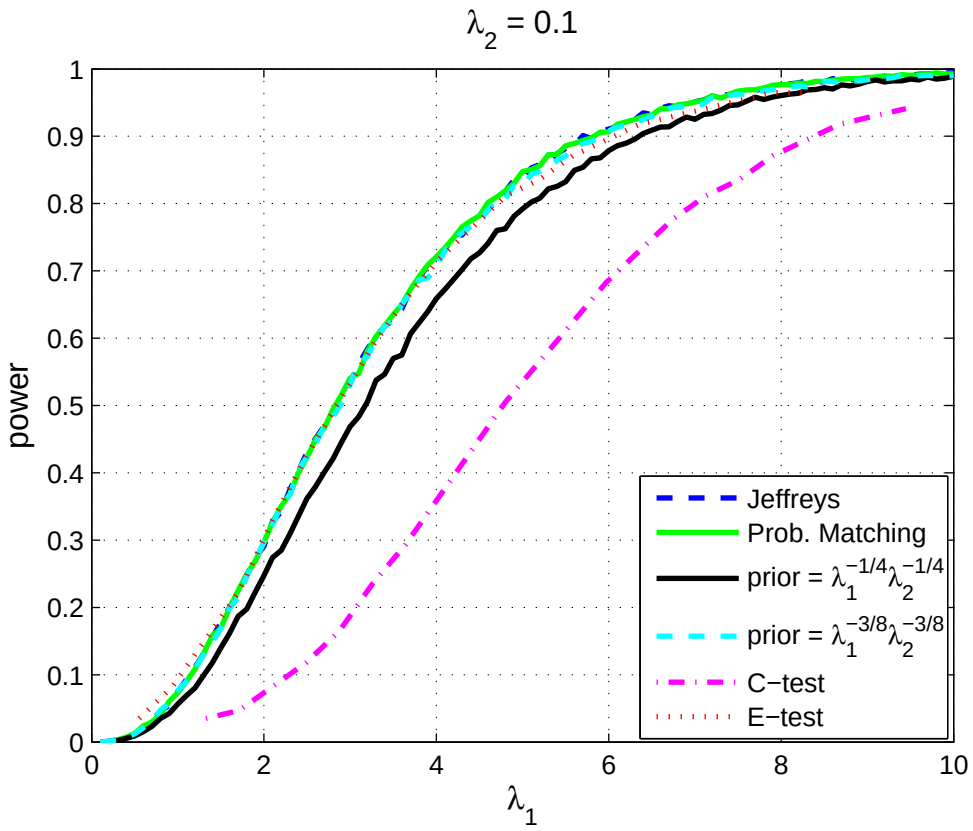


Figure 5.4: Power of the test as a function of λ_1 when $\lambda_2 = 0.1$.

From Figure 5.4 it can be seen that the power of the test when using the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$, is smaller than the power of the tests when using the Jeffreys prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$. The Jeffreys prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$, and the E- test give almost exactly the same results. From Krishnamoorthy & Thomson (2004) the power of the E- test is larger than the power of the C - test.

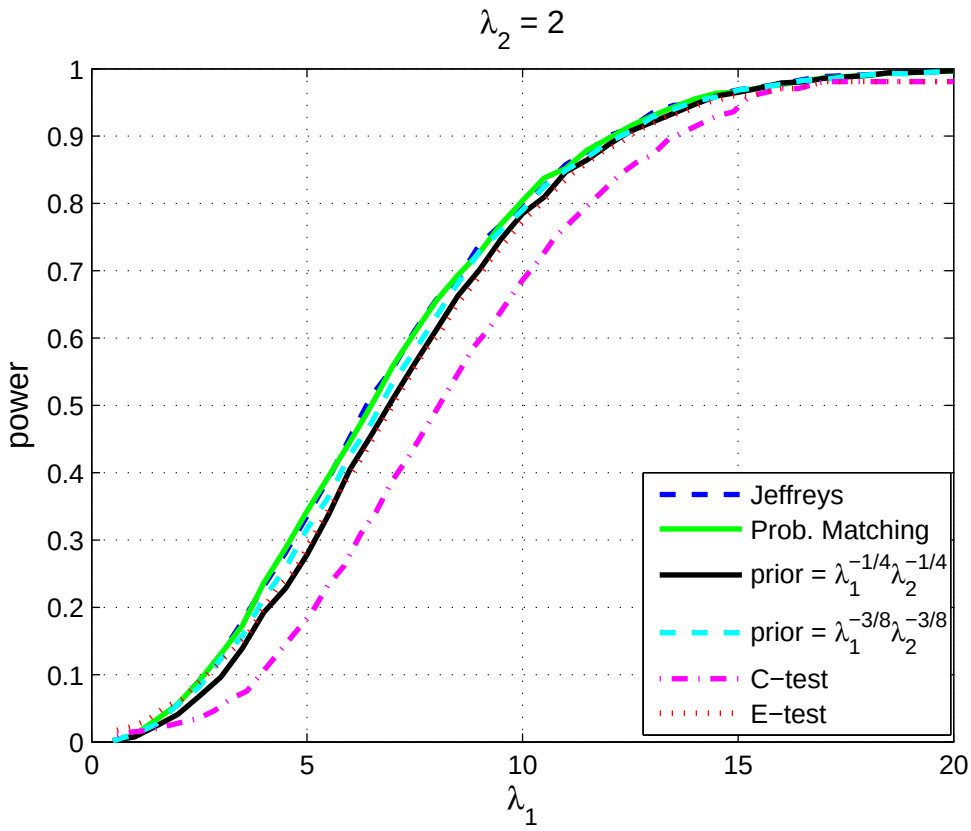


Figure 5.5: Power of the test as a function of λ_1 when $\lambda_2 = 2$.

From Figure 5.5 it can be seen that the power of the test when using the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$, is still a bit smaller than the power of the tests when using the Jeffreys prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$. The Jeffreys and probability matching priors give almost exactly the same results. The prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$, and the E- test give almost exactly the same results. From Krishnamoorthy & Thomson (2004) the power of the E - test is larger than the power of the C - test.

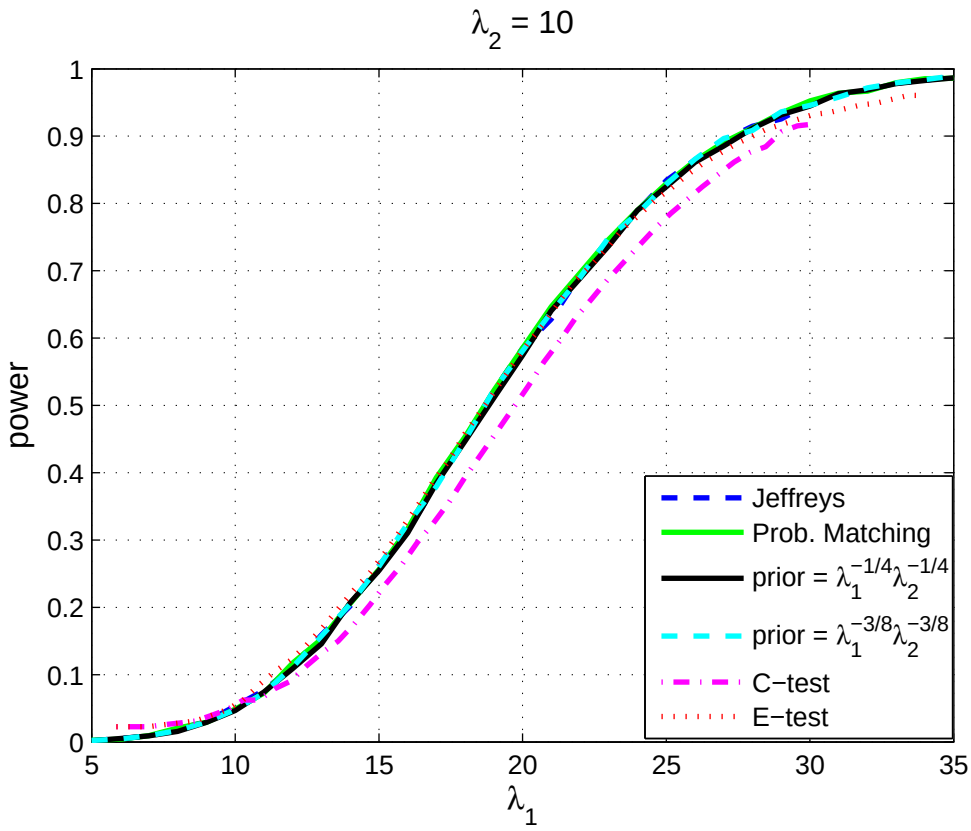


Figure 5.6: Power of the test as a function of λ_1 when $\lambda_2 = 10$.

From Figure 5.6 it can be seen that the power of the test when using the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$, is almost the same as the power of the tests when using the Jeffreys prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$. From Krishnamoorthy & Thomson (2004) the power of the E - test is still larger than the power of the C - test, but they are almost equal to each other. The four tests using Bayesian methods and the E - test all give almost exactly the same results. We can conclude from the Bayesian procedures and from the methods by Krishnamoorthy & Thomson (2004), that as the sample sizes increase, i.e. as the values of λ_1 and λ_2 increase, the powers of the tests are increasing. Also, as the values of λ_1 and λ_2 increase, the difference between the powers of the different tests are smaller. The Bayesian procedures compare well with the procedures by Krishnamoorthy & Thomson (2004). From Figures 5.4 - 5.6 it is also clear that the powers of Jeffreys and probability matching priors are larger than those of the E - test. This could be expected because the type I error rates of these two priors are usually somewhat larger that the chosen alpha value.

5.5 Conclusion

In this chapter the probability matching prior for a linear contrast of Poisson parameters, $\delta = \sum_{i=1}^{k} a_i \lambda_i$, (i.e. $\sum_{i=1}^{k} a_i = 0$) was derived. We also indicated what the probability matching prior should be when $\sum_{i=1}^{k} a_i \neq 0$. We compared the four approximate confidence intervals for linear contrasts of Poisson rates proposed by Stamey & Hamilton (2006) to confidence intervals using Bayesian procedures, when using the probability matching prior. Simulation studies have shown that the Wald interval performs the poorest. The probability matching prior also performs satisfactory. We then addressed the problem of hypothesis testing about two Poisson means, by looking at the size and power of different tests. We compared four Bayesian procedures to two procedures used by Krishnamoorthy & Thomson (2004). We used the Jeffreys prior, the probability matching prior, a third prior which is proportional to $\lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$ and a fourth prior which is proportional to $\lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$ and compared them to their results. The Bayesian procedures compared well with the procedures used by Krishnamoorthy & Thomson (2004). The C - test performed most poorly of the six tests.

Chapter 6

Estimation for Binomial Rates from Pooled Samples

6.1 Introduction

In this chapter we will look into confidence intervals for linear functions of binomial rates from pooled samples. We will investigate the performance of Bayesian credibility intervals for a single proportion as well as the difference of two binomial proportions estimated from pooled samples. In previous chapter we considered equal tail Bayesian credible intervals, in this chapter HPD intervals will be considered. Where the HPD interval will have a shorter interval length than the equal tail interval. Group testing has been used in many fields of study, as individual testing can be too time consuming and pooled testing is more cost-effective. Group testing is where units are pooled together and tested as a group rather than individually. Biggerstaff (2008) used asymptotic methods to derive Wald, profile score and profile likelihood ratio intervals. Biggerstaff (2008) also adapted the Wilson score-based interval of Newcombe. Tu et al. (1995) investigated the maximum likelihood estimator for equal pool sizes. Hepworth (1996) considered the sequential testing of groups of different sizes, by constructing exact confidence intervals for problems involving unequal sized groups. Hepworth (2005) also considered asymptotic interval estimation methods where groups are of different sizes. Hepworth (2005) investigated four methods, two based on the distribution of the maximum likelihood estimate (MLE), one on the score statistic and one on the likelihood ratio. Hepworth (2005) recommended the method based on the score statistic with a correction for skewness.

6.2 Prior Distribution for Binomial Proportions from Pooled Samples

Assume that the proportion of successes in a given population is p. We will refer to an infected individual as a success in a binomial trial. Using the notation from Biggerstaff (2008), let N individuals be sampled independently from the population, and then be grouped into pools. The size of a pool will be indicated by m_i, for $i = 1,2,\ldots,M$, where M is the number of distinct pool sizes, let n_i be the number of pools of size m_i, and let X_i be the number of the n_i pools that is positive. Assume that $X_1, X_2, \ldots, X_M$ are independent binomial random variables with $X_i \sim Bin\left(n_i, 1-(1-p)^{m_i}\right)$.

The likelihood function is given by

$$L(p\,|x_1,x_2,\ldots,x_M) \;\propto\; \prod_{i=1}^{M}\left\{[1-(1-p)^{m_i}]^{x_i}\,[(1-p)^{m_i}]^{n_i-x_i}\right\}.$$

The derivation of the Fisher information is given in Appendix E in Theorem E.1, derived by Walter et al. (1980). The Fisher information is given by

$$F(p) \;=\; \sum_{i=1}^{M}\left\{\frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}\right\}.$$

As mentioned in previous chapters, the Jeffreys prior is proportional to the square root of the determinant of the Fisher information and is given by

$$\begin{aligned}\pi_J(p) &\propto |F(p)|^{\frac{1}{2}} \\ &= \left(\sum_{i=1}^{M}\left\{\frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}\right\}\right)^{\frac{1}{2}}. \qquad (6.1)\end{aligned}$$

The joint posterior distribution when using the Jeffreys prior is given by

$$\begin{aligned}\pi_J(p\,|data) &\propto \pi_J(p)\times L(p\,|data) \\ &\propto \left(\sum_{i=1}^{M}\left\{\frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}\right\}\right)^{\frac{1}{2}} \\ &\quad\times\prod_{i=1}^{M}\left\{[1-(1-p)^{m_i}]^{x_i}\,[(1-p)^{m_i}]^{n_i-x_i}\right\} \qquad \text{for } 0\le p\le 1. \qquad (6.2)\end{aligned}$$

Theorem 6.1. The prior distribution

$$\pi_J(p\,|data) \;\propto\; \left(\sum_{i=1}^{M}\left\{\frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}\right\}\right)^{\frac{1}{2}}\prod_{i=1}^{M}\left\{[1-(1-p)^{m_i}]^{x_i}\,[(1-p)^{m_i}]^{n_i-x_i}\right\}$$

converges and exists if at least one $x_i > 0$. Where $x_i \in \{0, 1, \ldots, n_i\}$ and $0 < p \leq 1$.

Proof. Let at least one $x_i > 0$, say $x_1 > 0$, then

$$\begin{aligned}
\pi_J(p\,|data) &\propto \left(\sum_{i=1}^{M}\left\{\frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}\right\}\right)^{\frac{1}{2}} \prod_{i=1}^{M}\left\{[1-(1-p)^{m_i}]^{x_i}\,[(1-p)^{m_i}]^{n_i-x_i}\right\} \\
&= [(1-p)^{m_1}]^{n_1-x_1} \prod_{i=2}^{M}\left\{[1-(1-p)^{m_i}]^{x_i}\,[(1-p)^{m_i}]^{n_i-x_i}\right\} \\
&\quad \times \left[\sum_{i=1}^{M}\left\{\frac{[1-(1-p)^{m_1}]^{2x_1} m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}\right\}\right]^{\frac{1}{2}}
\end{aligned}$$

The first term ($i = 1$) of the sum in the square brackets has the form $[1-(1-p)^{m_1}]^{2x_1-1} m_1^2 n_1 (1-p)^{m_1-2}$.

$$\lim_{p\to 0^+} [1-(1-p)^{m_1}]^{2x_1-1} m_1^2 n_1 (1-p)^{m_1-2} = 0.$$

The other term ($i > 1$) has the form

$$\frac{[1-(1-p)^{m_1}]^{2x_1} m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}$$

and

$$\lim_{p\to 0^+} \frac{[1-(1-p)^{m_1}]^{2x_1} m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]} = 0.$$

Thus, $\pi_J(p\,|data) \to 0$ as $p \to 0^+$. We can therefore conclude that $\pi_J(p\,|data)$ is limited to $(0, 1]$ and therefore $\int_0^1 \pi_J(p\,|data)\,dp$ exists. □

Theorem 6.2. The prior distribution

$$\pi_J(p\,|data) \propto \left(\sum_{i=1}^{M}\left\{\frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}\right\}\right)^{\frac{1}{2}} \prod_{i=1}^{M}\left\{[1-(1-p)^{m_i}]^{x_i}\,[(1-p)^{m_i}]^{n_i-x_i}\right\}$$

converges and exists if all $x_i = 0$. Where $x_i \in \{0, 1, \ldots, n_i\}$ and $0 < p \leq 1$.

Proof. Consider two functions, $g(p)$ and $h(p)$.

Let

$$g(p) = \frac{1}{\sqrt{p}} \qquad \text{for} \quad 0 < p \leq 1$$

and

$$\begin{aligned} h(p) &= [(1-p)^{m_1}]^{n_1-x_1} \prod_{i=2}^{M} \left\{ [1-(1-p)^{m_i}]^{x_i} [(1-p)^{m_i}]^{n_i-x_i} \right\} \\ &\quad \times \left[\sum_{i=1}^{M} \left\{ \frac{[1-(1-p)^{m_1}]^{2x_1} m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]} \right\} \right]^{\frac{1}{2}} \end{aligned}$$

Then

$$\begin{aligned} \frac{h(p)}{g(p)} &= \sqrt{p} h(p) \\ &= \prod_{i=1}^{M} (1-p)^{m_i n_i} \times \left[\sum_{i=1}^{M} \left\{ \frac{m_i^2 n_i p (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]} \right\} \right]^{\frac{1}{2}} \end{aligned}$$

For each $i = 1, 2, \ldots, M$ we have.

$$\begin{aligned} \lim_{p \to 0^+} \frac{p(1-p)^{m_i-2}}{1-(1-p)^{m_i}} &= \lim_{p \to 0^+} \frac{(1-p)^{m_i-2} - p(m_i-2)(1-p)^{m_i-3}}{m_i (1-p)^{m_i-1}} \\ &= \frac{1}{m_i}. \end{aligned}$$

Thus

$$\lim_{p \to 0^+} \frac{h(p)}{g(p)} = \sqrt{\sum_{i=1}^{M} m_i n_i} > 0$$

Since $\int_0^1 g(p)\,dp$ converges, $\int_0^1 h(p)\,dp$ will also converge and we can conclude that $\int_0^1 \pi_J(p\,|data)\,dp$ exists and converges. □

If $M = 1$, $m_1 = m$, $n_1 = n$ and $x_1 = x$, it follows from Equation 6.1 that

$$\begin{aligned} \pi_J(p) &\propto \left\{ \frac{m^2 n (1-p)^{m-2}}{[1-(1-p)^m]} \right\}^{\frac{1}{2}} \\ &\propto [(1-p)^m]^{\frac{1}{2}-\frac{1}{m}} [1-(1-p)^m]^{-\frac{1}{2}}. \end{aligned} \tag{6.3}$$

The posterior distribution when using the Jeffreys prior is given by

$$\pi_J(p\,|data) \propto [(1-p)^m]^{n-x+\frac{1}{2}-\frac{1}{m}} [1-(1-p)^m]^{x-\frac{1}{2}} \qquad \text{for } 0 \le p \le 1. \tag{6.4}$$

Theorem 6.3. When $\theta = (1-p)^m$, the posterior distribution of θ will be $Beta\left(x+\frac{1}{2}, n-x+\frac{1}{2}\right)$, i.e.

$$\pi_J(\theta\,|data) \;\propto\; (1-\theta)^{n-x-\frac{1}{2}}\,\theta^{x-\frac{1}{2}}. \tag{6.5}$$

Proof. From Equation 6.4, the posterior distribution is given as

$$\pi_J(p\,|data) \;\propto\; [(1-p)^m]^{n-x+\frac{1}{2}-\frac{1}{m}}\,[1-(1-p)^m]^{x-\frac{1}{2}} \qquad \text{for } 0 \le p \le 1.$$

Let $\theta = (1-p)^m$, then $p = 1-\theta^{\frac{1}{m}}$, and

$$\left|\frac{dp}{d\theta}\right| \;=\; \frac{1}{m}\theta^{\frac{1}{m}-1}$$

$$\begin{aligned}
\pi_J(\theta\,|data) &\propto \left[\left(1-\left(1-\theta^{\frac{1}{m}}\right)\right)^m\right]^{n-x+\frac{1}{2}-\frac{1}{m}}\left[1-\left(1-\left(1-\theta^{\frac{1}{m}}\right)\right)^m\right]^{x-\frac{1}{2}}\frac{1}{m}\theta^{\frac{1}{m}-1}\\
&= \left[\left(\theta^{\frac{1}{m}}\right)^m\right]^{n-x+\frac{1}{2}-\frac{1}{m}}\left[1-\left(\theta^{\frac{1}{m}}\right)^m\right]^{x-\frac{1}{2}}\frac{1}{m}\theta^{\frac{1}{m}-1}\\
&= \theta^{n-x+\frac{1}{2}-\frac{1}{m}}(1-\theta)^{x-\frac{1}{2}}\frac{1}{m}\theta^{\frac{1}{m}-1}\\
&= \frac{1}{m}\theta^{n-x+\frac{1}{2}-\frac{1}{m}+\frac{1}{m}-1}(1-\theta)^{x-\frac{1}{2}}\\
\therefore \pi_J(\theta\,|data) &\propto (1-\theta)^{x-\frac{1}{2}}\,\theta^{n-x-\frac{1}{2}}.
\end{aligned} \tag{6.6}$$

□

Using Equation 6.6 and transformation, the posterior distribution for $p = 1-\theta^{\frac{1}{m}}$ can be determined, where $\left|\frac{d\theta}{dp}\right| = m(1-p)^{m-1}$.

$$\therefore \pi_J(p\,|data) \;=\; \frac{m}{B\left(x+\frac{1}{2}, n-x+\frac{1}{2}\right)}\,[(1-p)^m]^{n-x+\frac{1}{2}-\frac{1}{m}}\,[1-(1-p)^m]^{x-\frac{1}{2}}. \tag{6.7}$$

6.3 Example and Simulation Studies

6.3.1 Simulation Study I - Single Proportion

In this section we will consider a simulation study for proportions from pooled samples. A single proportion will be considered where $M = 1$, $M = 2$, $M = 3$ and $M = 4$. We will look at coverage, left noncoverage, right noncoverage, symmetry and interval length. Biggerstaff (2008) defines noncoverage symmetry as the difference in proportional noncoverage, i.e.

$$\text{Symmetry} = \frac{P[\text{Left noncoverage}] - P[\text{Right noncoverage}]}{P[\text{Left noncoverage}] + P[\text{Right noncoverage}]}$$

with a negative value indicating mostly right noncoverage and a positive value indicating mostly left noncoverage. A value of zero for symmetry indicates symmetric noncoverage.

6.3.1.1 Single proportion: $M = 1$

The posterior distribution is given in Equation 6.7 when the Jeffreys prior is used. So for 200 pools each of size 5, $n = 200$ and $m = 5$. The posterior distribution will be

$$\pi_J(p|data) = \frac{5}{B\left(x+\frac{1}{2}, 200.5-x\right)}\left[(1-p)^5\right]^{200.3-x}\left[1-(1-p)^5\right]^{x-\frac{1}{2}}.$$

The simulation results are given in Table 6.1. The average length and the mean of the expected value, as well as the mode can also be found by averaging the posterior measures over all values of X for a given p. This is given in Table 6.1 for $p = 0.001$, 0.0015, 0.002, 0.005 and 0.01, where the coverage is for the 95% HPD interval. Range X refers to the range of X values for which the posterior HPD interval contains the true value of p.

Table 6.1: Simulation results when $M = 1$, $n = 200$ and $m = 5$, for different values of p.

p	**Range** X	**Coverage**	**Left non-coverage**	**Right non-coverage**	**Symmetry**	**Length**	**Mean**	**Mode**
0.0010	0 - 4	0.9965	0.0035	0	1	0.0038	0.0015	0.0007
0.0015	0 - 5	0.9958	0.0042	0	1	0.0045	0.0020	0.0011
0.0020	0 - 5	0.9842	0.0158	0	1	0.0052	0.0025	0.0016
0.0050	2 - 10	0.9478	0.0405	0.0118	0.549	0.0084	0.0055	0.0045
0.0100	5 - 16	0.9497	0.0302	0.0201	0.200	0.0120	0.0105	0.0095

Now consider the case where we have 20 pools each of size 50, that is $n = 20$ and $m = 50$. The

posterior distribution will be

$$\pi_J(p|data) = \frac{50}{B\left(x+\frac{1}{2},20.5-x\right)}\left[(1-p)^{50}\right]^{20.48-x}\left[1-(1-p)^{50}\right]^{x-\frac{1}{2}}.$$

The simulation results are given in Table 6.2. The average length and the mean of the expected value, as well as the mode can also be found by averaging the posterior measures over all values of X for a given p. This is given in Table 6.2 for $p = 0.001$, 0.0015, 0.002, 0.005 and 0.01, where the coverage is for the 95% HPD interval. Range X refers to the range of X values for which the posterior HPD interval contains the true value of p.

Table 6.2: Simulation results when $M = 1$, $n = 20$ and $m = 50$, for different values of p.

p	**Range** X	**Coverage**	**Left non-coverage**	**Right non-coverage**	**Symmetry**	**Length** $\times 1\,000$	**Mean** $\times 1\,000$	**Mode** $\times 1\,000$
0.0010	0 - 3	0.9854	0.0146	0	1	3.888	1.513	0.689
0.0015	0 - 4	0.9878	0.0122	0	1	4.655	2.026	1.118
0.0020	0 - 5	0.9911	0.0089	0	1	5.389	2.540	1.575
0.0050	2 - 8	0.9359	0.0195	0.0446	-0.392	9.024	5.631	4.516
0.0100	5 - 12	0.9254	0.0188	0.0558	-0.496	13.919	10.815	9.521

6.3.1.2 Single proportion: $M = 2$

Let $M = 2$, where $m_1 = 5$, $m_2 = 10$, $n_1 = 100$ and $n_2 = 50$, that is 100 pools of size 5 and 50 pools of size 10. The posterior distribution is given in Equation 6.2 when the Jeffreys prior is used. The posterior distribution will therefore be

$$\begin{aligned}\pi_J(p|data) \propto & \left(\frac{2\,500(1-p)^3}{\left[1-(1-p)^5\right]}+\frac{5\,000(1-p)^8}{\left[1-(1-p)^{10}\right]}\right)^{\frac{1}{2}} \\ & \times\left[1-(1-p)^5\right]^{x_1}\left[1-(1-p)^{10}\right]^{x_2}(1-p)^{1\,000-5x_1-10x_2}. \qquad (6.8)\end{aligned}$$

We can calculate the 95% HPD interval, mean, mode and interval length of the posterior for all combinations of $\underline{X} = [x_1\ x_2]$ and find all those values of $\underline{X}$ which yield an interval that contains the true value of p.

The simulation results are given in Table 6.3, where the coverage is for the 95% HPD interval. The range is omitted as it involves a large number of combinations.

Table 6.3: Simulation results when $M = 2$, $m_1 = 5$, $m_2 = 10$, $n_1 = 100$ and $n_2 = 50$, for different values of p.

p	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Length**	**Mean**	**Mode**
0.0010	0.9817	0.0183	0	1	0.0034	0.0015	0.0007
0.0015	0.9824	0.0176	0	1	0.0051	0.0020	0.0011
0.0020	0.8496	0.0153	0.1351	-0.797	0.0056	0.0025	0.0016
0.0050	0.9486	0.0107	0.0407	-0.583	0.0085	0.0055	0.0045
0.0100	0.9520	0.0168	0.0321	-0.301	0.0123	0.0105	0.0095

Notice that the length, mean and mode are similar to the case when $M = 1$ in Table 6.1. Also, the coverage probability when $p = 0.002$ is very low compared to $M = 1$. That is because in the first case, $p = 0.002$ falls just inside the credibility interval when $x = 0$ which has a relatively high probability, while in the second case, it just misses the interval when $\underline{x} = [0\ 0]$.

6.3.1.3 Single proportion: $M = 3$

Let $M = 3$, where $m_1 = 10$, $m_2 = 25$, $m_3 = 50$, $n_1 = 20$, $n_2 = 8$ and $n_3 = 12$, that is 20 pools of size 10, 8 pools of size 25 and 12 pools of size 50. The posterior distribution is given in Equation 6.2 when the Jeffreys prior is used. The posterior distribution will therefore be

$$\begin{aligned}\pi_J(p\,|data) \;\propto\; & \left(\frac{2\,000(1-p)^8}{\left[1-(1-p)^{10}\right]}+\frac{5\,000(1-p)^{23}}{\left[1-(1-p)^{25}\right]}+\frac{30\,000(1-p)^{48}}{\left[1-(1-p)^{50}\right]}\right)^{\frac{1}{2}} \\ & \times\left[1-(1-p)^{10}\right]^{x_1}\left[1-(1-p)^{25}\right]^{x_2} \\ & \times\left[1-(1-p)^{50}\right]^{x_3}(1-p)^{1\,000-10x_1-25x_2-50x_3}.\end{aligned}$$

We calculate the 95% HPD interval, mean, mode and interval length of the posterior for all combinations of $\underline{X} = [x_1\ x_2\ x_3]$ and find all those values of $\underline{X}$ which yield an interval that contains the true value of p.

The simulation results are given in Table 6.4, where the coverage is for the 95% HPD interval. The range is omitted as it involves a large number of combinations.

Table 6.4: Simulation results when $M = 3$, $m_1 = 10$, $m_2 = 25$, $m_3 = 50$, $n_1 = 20$, $n_2 = 8$ and $n_3 = 12$, for different values of p.

p	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Length** $\times 1\,000$	**Mean** $\times 1\,000$	**Mode** $\times 1\,000$
0.0010	0.9842	0.0157	0	1	3.9119	1.5146	0.6877
0.0015	0.9862	0.0138	0	1	4.6881	2.0231	1.1163
0.0020	0.9892	0.0108	0	1	5.4009	2.5321	1.5731
0.0050	0.9458	0.0108	0.0434	-0.5999	8.8797	5.5911	4.5121
0.0100	0.9357	0.0176	0.0467	-0.4538	13.3565	10.7042	9.5117

6.3.1.4 Single proportion: $M = 4$

Let $M = 4$, where $m_1 = 5$, $m_2 = 10$, $m_3 = 25$, $m_4 = 50$, $n_1 = 20$, $n_2 = 40$, $n_3 = 12$ and $n_4 = 4$, that is 20 pools of size 5, 40 pools of size 10, 12 pools of size 25 and 4 pools of size 50. The posterior distribution is given in Equation 6.2 when the Jeffreys prior is used. The posterior distribution will therefore be

$$\begin{aligned}\pi_J(p\,|data) \;\propto\; & \left(\frac{500\,(1-p)^3}{\left[1-(1-p)^5\right]}+\frac{4\,000\,(1-p)^8}{\left[1-(1-p)^{10}\right]}+\frac{7\,500\,(1-p)^{23}}{\left[1-(1-p)^{25}\right]}+\frac{10\,000\,(1-p)^{48}}{\left[1-(1-p)^{50}\right]}\right)^{\frac{1}{2}} \\ & \times\left[1-(1-p)^5\right]^{x_1}\left[1-(1-p)^{10}\right]^{x_2}\left[1-(1-p)^{25}\right]^{x_3} \\ & \times\left[1-(1-p)^{50}\right]^{x_4}(1-p)^{1\,000-5x_1-10x_2-25x_3-50x_4}.\end{aligned}$$

We calculate the 95% HPD interval, mean, mode and interval length of the posterior for all combinations of $\underline{X} = [x_1\ x_2\ x_3\ x_4]$ and find all those values of $\underline{X}$ which yield an interval that contains the true value of p.

The simulation results are given in Table 6.5, where the coverage is for the 95% HPD interval. The range is omitted as it involves a large number of combinations.

Table 6.5: Simulation results when $M = 4$, $m_1 = 5$, $m_2 = 10$, $m_3 = 25$, $m_4 = 50$, $n_1 = 20$, $n_2 = 40$, $n_3 = 12$ and $n_4 = 4$, for different values of p.

p	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Length** $\times 1\,000$	**Mean** $\times 1\,000$	**Mode** $\times 1\,000$
0.0010	0.9829	0.0171	0	1	3.8933	1.5084	0.6858
0.0015	0.9843	0.0157	0	1	4.6541	2.0128	1.1143
0.0020	0.8519	0.0130	0.1351	-0.8245	5.3496	2.5175	1.5712
0.0050	0.9411	0.0170	0.0419	-0.4242	8.6859	5.5480	4.5101
0.0100	0.9458	0.0161	0.0381	-0.4043	12.7852	10.6019	9.5091

6.3.1.5 Single proportion: Averages

Results over all combinations are very similar, except that the coverage probabilities for $p = 0.002$ at $M = 2$ and $M = 4$ are very low, 0.8496 and 0.8519 respectively. The interval length increases with p, but stays reasonably constant when M increases. Table 6.6 gives the averages over p for the various pool combinations while Table 6.7 gives the averages over combinations for given value of p. Table 6.8 gives the overall averages.

Table 6.6: Averages over all values of p.

M	n	m	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Length** $\times 1\,000$
1	200	5	0.9748	0.0188	0.0064	0.7498	6.724
	20	50	0.9651	0.0148	0.0201	0.4225	7.377
2	[100 50]	[5 10]	0.9429	0.0157	0.0414	0.0639	6.981
3	[20 8 12]	[10 25 50]	0.9682	0.0138	0.0180	0.3893	7.247
4	[20 40 12 4]	[5 10 25 50]	0.9412	0.0158	0.0430	0.0694	7.074

Table 6.7: Averages over all pool combinations.

p	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Length** $\times 1\,000$
0.0010	0.9862	0.0138	0	1	3.747
0.0015	0.9873	0.0127	0	1	4.707
0.0020	0.9332	0.0128	0.0540	0.2757	5.280
0.0050	0.9438	0.0197	0.0365	-0.2899	8.710
0.0100	0.9317	0.0199	0.0384	-0.2909	12.8861

Table 6.8: Overall averages.

Coverage	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Length** $\times 1\,000$
0.9584	0.0158	0.0258	0.3390	7.0659

We considered the different pool size combinations which was used by Biggerstaff (2008). Table 6.9 gives the results from Biggerstaff (2008) and the results obtained by us using the Bayesian method. The first five intervals in Table 6.9 are from Biggerstaff (2008).

Table 6.9: Overall averages of coverage rates, noncoverages, symmetry and average lengths. Nominal coverage is 95%.

Interval	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Length** $\times 1\,000$
MIR	0.8070	0.0010	0.1920	-0.99	6.0000
Wald	0.8140	0.0027	0.1830	-0.97	6.5000
Likelihood ratio (LRT)	0.9660	0.0188	0.0150	0.11	7.6000
Profile score	0.9480	0.0476	0.0040	0.84	8.0000
Skewness corrected score	0.9660	0.0205	0.0136	0.20	7.8000
Bayesian	0.9584	0.0158	0.0258	0.34	7.0659

From Table 6.9 we see that the coverage rates obtained by the MIR and Wald intervals are far below the nominal level of 0.95, this was also stated by Biggerstaff (2008). The other four intervals give coverages close to the nominal level, with the profile score and the Bayesian intervals performing slightly better. The results obtained from the Bayesian method by us compare well with the results obtained from the other researcher. As mentioned, a value of zero symmetry indicates symmetric noncoverage, the symmetry observed from the Bayesian method is equal to 0.34 which indicates mostly left noncoverage.

6.3.2 Simulation Study II - Two Proportions

In this section we will consider a simulation study for proportions from pooled samples, for the difference between two proportions. We will consider two cases, $M_1 = M_2 = 1$ and $M_1 = M_2 = 2$. We will look at coverage, distal noncoverage, mesial noncoverage, symmetry and interval length. Left noncoverage is interpretable as distal noncoverage probability and right noncoverage is interpretable as mesial noncoverage. It is desirable that these should be equal.

6.3.2.1 Difference between two proportions: $M_1 = M_2 = 1$

We use simulation to determine the properties of the posterior distribution of the difference according to the following steps, for given values of p_1 and p_2, and for all possible values of x_1 and x_2:

- **Step 1**

  Calculate the probabilities of outcomes x_1 and x_2 using the binomial distribution, and thus $P(x_1, x_2) = P(x_1)P(x_2)$.

- **Step 2**

  Simulate a sample of 100 000 from each of the two marginal posteriors of p_1 and p_2, using the beta distribution, using Equation 6.7.

- **Step 3**

  Now construct a sample of 100 000 differences, $p_1 - p_2$, and sort them.

- **Step 4**

  Stepwise search the sorted sample for the shortest interval containing 95% of the observations, and record the interval, length and mean of the sample.

- **Step 5**

  This is now available for every combination of x_1 and x_2, as well as the probability. So for the given values of p_1 and p_2, find all the intervals that cover the true value of $p_1 - p_2$ and sum all the corresponding probabilities. This will give the coverage probability. In the same way we find the distal and mesial probabilities and the average length.

Table 6.10 gives results for $m_1 = m_2 = 50$ and $n_1 = n_2 = 20$. Other simulations with different number of pools and pool sizes gave very similar results.

Table 6.10: Coverage rates, distal, mesial, symmetry and average lengths for $p_1 - p_2$ when $M_1 = M_2 = 1$, $m_1 = m_2 = 50$ and $n_1 = n_2 = 20$. Nominal coverage is 95%.

$p_1 - p_2$	**Coverage**	**Distal**	**Mesial**	**Symmetry**	**Length**
$0.0015 - 0.0010$	0.9794	0.0160	0.0046	-0.5504	0.0075
$0.0020 - 0.0010$	0.9857	0.0103	0.0040	-0.4431	0.0081
$0.0050 - 0.0010$	0.9520	0.0383	0.0096	-0.5991	0.0108
$0.0100 - 0.0010$	0.9468	0.0430	0.0101	-0.6192	0.0151
$0.0020 - 0.0015$	0.9673	0.0240	0.0087	-0.4696	0.0086
$0.0050 - 0.0015$	0.9646	0.0287	0.0067	-0.6194	0.0113
$0.0100 - 0.0015$	0.9568	0.0354	0.0078	-0.6396	0.0155
$0.0050 - 0.0020$	0.9537	0.0327	0.0135	-0.4146	0.0118
$0.0100 - 0.0020$	0.9536	0.0322	0.0143	-0.3858	0.0158
$0.0100 - 0.0050$	0.9547	0.0290	0.0162	-0.2825	0.0177

The coverage rates given in Table 6.10 are all above the nominal level of 0.95, except for the case where we have $p_1 - p_2 = 0.0100 - 0.0010$. In this case the coverage rate is equal to 0.9468, which is close to the nominal level. For all the cases, we have negative values for symmetry, which indicates mostly right noncoverage.

6.3.2.2 Difference between two proportions: $M_1 = M_2 = 2$

The problem here is more complex and there are simply too many combinations of outcomes when M is larger than one. Also we cannot use the beta distribution to simulate from the marginal posteriors of the values of p.

We use the following steps:

- **Step 1**

  For a specific data set, say $\underline{x} = [x_{11}\, x_{12}\, x_{21}\, x_{22}]$, we know the form of the marginal posteriors of p_1 and p_2 as given in Equation 6.8, so we discretise them by calculating their values at small intervals (0.0001) and then normalise them.

- **Step 2**

  Now we have a probability for every discrete outcome of p_1 and p_2. Forming all possible combinations of p_1 and p_2 and their associated probabilities by using a grid, we have a distribution for $p_1 - p_2$ for the given $\underline{x}$.

- **Step 3**

  After sorting, we can now search for the shortest 95% interval for $p_1 - p_2$, using the associated probabilities and also calculate the mean.

- **Step 4**

  Steps 1 to 3 should be done for all possible values of $\underline{x}$. The probability of $\underline{x}$ is the product of the individual binomial probabilities for given p_1 and p_2 .

- **Step 5**

  For the given p_1 and p_2, we find all the values of $\underline{x}$ which yielded an interval that covers the true value of $p_1 - p_2$, and sum their probabilities. This will give the coverage probability. In the same way we find the distal and mesial probabilities and the average length.

This simulation was carried out for only one combination of pool sizes. They are: $n_{11} = 100$, $m_{11} = 5$, $n_{12} = 50$, $m_{12} = 10$; $n_{21} = 50$, $m_{21} = 10$, $n_{22} = 20$, $m_{22} = 25$. Since the values of p are small, it is not necessary to calculate steps 1 to 3 for all combinations of the x values. The probabilities for a large number of positive outcomes can be regarded as negligibly small. For example, for 100 pools ($n = 100$) of size 5 each ($m = 5$) and $p = 0.001$, the probability of more than 7 successes is less than 10^{-6}. And then it must still be multiplied with the probabilities of the other pools outcomes. Results are given in Table 6.11.

Table 6.11: Coverage rates, distal, mesial, symmetry and average lengths for $p_1 - p_2$ when $M_1 = M_2 = 2$, $n_{11} = 100$, $m_{11} = 5$, $n_{12} = 50$, $m_{12} = 10$; $n_{21} = 50$, $m_{21} = 10$, $n_{22} = 20$, $m_{22} = 25$. Nominal coverage is 95%.

$p_1 - p_2$	**Coverage**	**Distal**	**Mesial**	**Symmetry**	**Length**
0.0015 – 0.0010	0.9942	0.0043	0.0015	-0.4869	0.00848
0.0020 – 0.0010	0.9849	0.0108	0.0043	-0.4363	0.00892
0.0050 – 0.0010	0.9707	0.0207	0.0086	-0.4148	0.01215
0.0100 – 0.0010	0.9606	0.0299	0.0095	-0.5266	0.01442
0.0020 – 0.0015	0.9868	0.0098	0.0034	-0.4735	0.00935
0.0050 – 0.0015	0.9646	0.0266	0.0088	-0.5028	0.01160
0.0100 – 0.0015	0.9607	0.0268	0.0125	-0.3639	0.01474
0.0050 – 0.0020	0.9688	0.0216	0.0096	-0.3860	0.01196
0.0100 – 0.0020	0.9614	0.0271	0.0115	-0.4041	0.01504
0.0100 – 0.0050	0.9581	0.0270	0.0149	-0.2888	0.01666

The coverage rates given in Table 6.11 are all above the nominal level of 0.95. For $p_1 - p_2 = 0.0100 - 0.0050$, the coverage rate is equal to 0.9581 which is the closest to the nominal value for the cases in the table. For all the cases we have negative values for symmetry, which indicates mostly right noncoverage, the same was observed in Table 6.10.

6.3.3 Example - West Nile Virus

Biggerstaff (2008) considered an example where a comparison is made between West Nile virus (WNV) infection prevalences in field collected *Culex nigripalpus* mosquitoes trapped at different heights. Biggerstaff (2008) derived asymptotic confidence intervals for the difference between two proportions estimated from pooled samples, where the sizes of the pools are not equal. Biggerstaff (2008) considered seven confidence intervals: an interval based on the minimum infection rate (MIR), the Wald interval, the profile score interval, the skewness corrected score interval, the bias- and skewness-corrected score interval, square-and-add Walter (SAW) interval and the profile likelihood interval. Table 6.12 summarises the data given in Table E.1 from Biggerstaff (2008).

Table 6.12: Summary of *Culex nigripalpus* mosquitoes trapped at different heights of $6m$ and $1.5m$.

	Sample 1 **height** $= 6m$	**Sample 2** **height** $= 1.5m$
Total	2 021	1 324
Number of pools	53	31
Average pool size	38.1321	42.7097
Minimum pool size	1	5
Maximum pool size	50	100
Number of positive pools	7	1

We used the Jeffreys prior to construct a 95% Bayesian (HPD) interval for each sample. The results are shown in Table 6.13. Figure 6.1 shows a plot of the posterior distribution, using the posterior distribution defined in Equation 6.7, for the two samples.

Table 6.13: 95% intervals and interval lengths for the proportions (per 1 000) of the two samples.

	95% HPD Interval	**Length**	**95% Confidence Interval (Biggerstaff, 2008)**	**Length**
Sample 1 height $= 6m$	$(1.444, 6.959)$	5.515	$(1.653, 7.408)$	5.755
Sample 2 height $= 1.5m$	$(0.019, 3.002)$	2.983	$(0.044, 3.670)$	3.626

From Table 6.13 the Bayesian intervals are shorter than those obtained by Biggerstaff (2008).

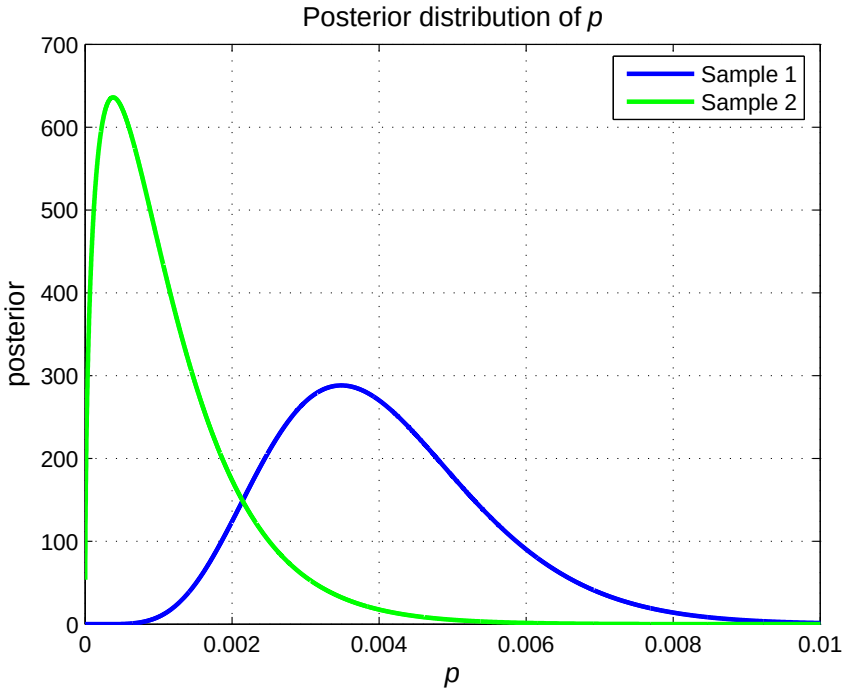


Figure 6.1: Posterior distribution of p.

For the mosquito data we draw random samples of 100 000 from each of the two posteriors mentioned above and calculate the difference between the two proportions. The histogram of the difference between the two proportions is then an approximation to the posterior distribution. We used the Jeffreys prior to construct a 95% Bayesian (HPD) interval for the difference between the two proportions. The results are shown in Table 6.14, the results for the first seven intervals are from Biggerstaff (2008). In Table 6.14 we see that zero is just included in the 95% Bayesian (HPD) Interval.

Table 6.14: 95% intervals and interval lengths for the difference between the two proportions (per 1 000).

	95% Interval	**Length**
MIR	$(-0.250, 5.667)$	5.920
Wald	$(-0.165, 6.182)$	6.347
Profile score	$(-0.746, 6.935)$	7.681
Skewness corrected score	$(-0.572, 6.824)$	7.396
Bias- and skewness-corrected score	$(-0.570, 6.825)$	7.395
Profile likelihood	$(-0.355, 6.729)$	7.084
Square-and-add Walter	$(-0.861, 6.852)$	7.713
Bayesian	$(-0.403, 6.528)$	6.931

The Bayesian interval compares relatively well with the others, all the intervals include 0. The MIR, Wald and Bayesian intervals give shorter interval lengths than the other intervals. The MIR and Wald intervals are known for giving poor coverage. So if we compare the Bayesian interval to the other five intervals, the Bayesian interval is the shortest one.

A normal approximation is fitted to the posterior in Figure 6.2. It seems that the posterior is slightly skew, the fit is thus not very good.

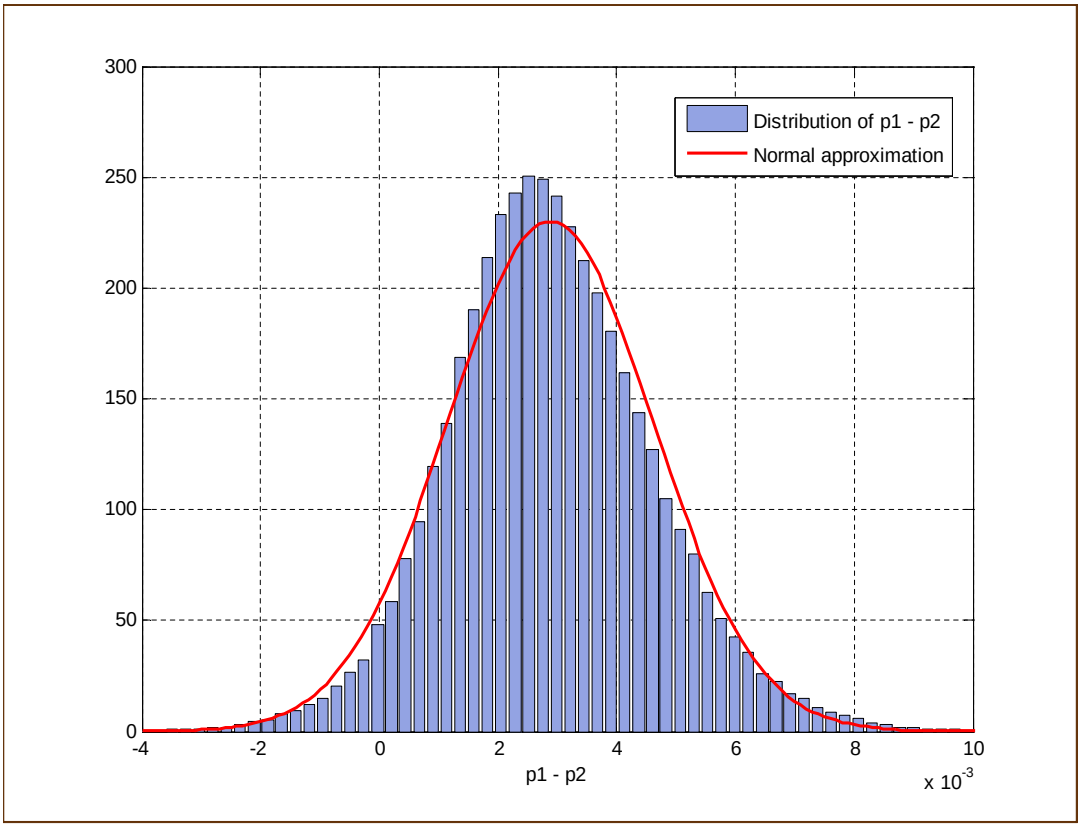


Figure 6.2: Posterior distribution of $p_1 - p_2$.

6.3.4 Simulation Study III

6.3.4.1 Coverage

Using the sample and pool sizes as given in Biggerstaff (2008) where $M_1 = 19$, with $p_1 = 0.004$, we simulated 10 000 outcomes of the 19×1 vector $\underline{x}_1$. This was done by simulating 19 binomial observations, each with a sample size and a different probability, since the pool sizes differ.

For each outcome of $\underline{x}_1$ the posterior distribution

$$\pi_J(p_1|data) \propto \left(\sum_{i=1}^{M_1}\left\{\frac{m_i^2 n_i (1-p_1)^{m_i-2}}{[1-(1-p_1)^{m_i}]}\right\}\right)^{\frac{1}{2}} \times \prod_{i=1}^{M_1}\left\{[1-(1-p_1)^{m_i}]^{x_i}[(1-p_1)^{m_i}]^{n_i-x_i}\right\}$$

is calculated over the range p_1 = 0.0001:0.0001:0.015 and normalised. Then the shortest 95% interval is calculated, as well as the mean and mode. This was done for the 10 000 samples and the percentages of coverage and noncoverage observed.

The same was done with the second sample where $M_2 = 16$, with $p_2 = 0.001$. The results are given in Table 6.15, where the mean, mode and length $(\times 1\,000)$ are the averages over all posteriors.

Table 6.15: Simulation results when $M_1 = 19$ with $p_1 = 0.004$ and $M_2 = 16$ with $p_2 = 0.001$, with samples of 10 000.

p	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Mean** $\times 1\,000$	**Mode** $\times 1\,000$	**Length** $\times 1\,000$
0.004	0.9285	0.0260	0.0455	-0.2727	4.2890	3.7546	5.5848
0.001	0.9870	0.0130	0	1	1.3706	0.7527	3.1931

Notice that when $p = 0.001$, (with $M_2 = 16$) there can be no right noncoverage. The lowest possible upper limit of the 95% credibility interval (when all values of x are zero) is 0.00131, which still covers the true value.

We cannot simulate from the posterior distribution of the difference between the two values of p, but we can look at the difference between the means of the 10 000 posteriors of the two individual values of p.

The mean difference between the posterior means is $\overline{p_1 - p_2} = 0.0029184$ (where the true difference between the values of p is 0.003), and the shortest 95% interval is $(-0.000457; 0.006283)$, which still includes zero.

A histogram of the differences between means is shown in Figure 6.3.

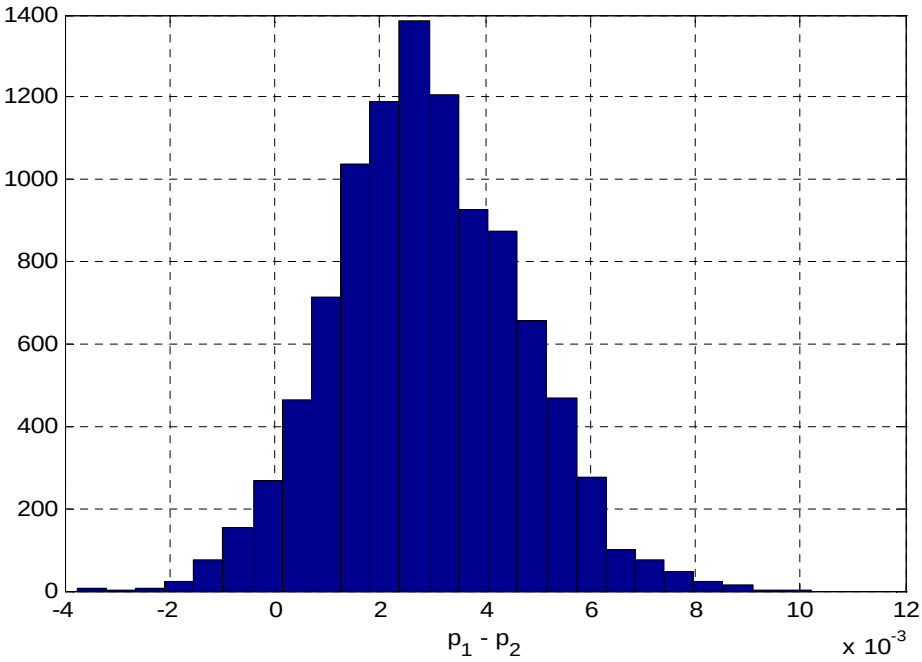


Figure 6.3: Histogram of $p_1 - p_2$.

This simulation was repeated, but now with samples of 20 000, the original 10 000 plus an additional 10 000. The results are given in Table 6.16, where the mean, mode and length $(\times 1\,000)$ are the averages over all posteriors.

Table 6.16: Simulation results when $M_1 = 19$ with $p_1 = 0.004$ and $M_2 = 16$ with $p_2 = 0.001$, with samples of 20 000.

p	**Coverage**	**Left noncoverage**	**Right noncoverage**	**Symmetry**	**Mean** $\times 1\,000$	**Mode** $\times 1\,000$	**Length** $\times 1\,000$
0.004	0.9294	0.0248	0.0458	-0.2975	4.3013	3.7666	5.5944
0.001	0.9871	0.0129	0	1	1.3690	0.7516	3.1903

The results changed slightly. The mean, mode and length $(\times 1\,000)$ are the averages over all posteriors.

6.3.4.2 Bayes factor

In this section we will briefly look at the Bayes factor and apply it to an example. Robert (2001) gives the following definition for the Bayes factor: The Bayes factor is the ratio of the posterior probabilities of the null and the alternative hypotheses over the ratio of the prior probabilities at the null and the alternative hypotheses.

Again using the sample and pool sizes of Biggerstaff (2008) with $p_1 = 0.004$ and $p_2 = 0.001$, we simulated 10 000 outcomes of the vectors $\underline{x}_1$ and $\underline{x}_2$. Here we used the $Beta\,(1/2, 1/2)$ prior for the values of p, so that under model M_0: $p_1 = p_2 = p$, we have

$$L(p\,|\underline{x}, M_0) \;=\; \prod_{i=1}^{2}\prod_{j=1}^{M_i}[1-(1-p)^{m_{ij}}]^{x_{ij}}\,[(1-p)^{m_{ij}}]^{n_{ij}-x_{ij}}\,.$$

The marginal likelihood is then

$$f(\underline{x}\,|M_0) \;=\; \frac{1}{\pi}\int_0^1 p^{-\frac{1}{2}}\,(1-p)^{-\frac{1}{2}}\,L(p\,|\underline{x}, M_0)\,dp.$$

Under the model $M_1 : p_1 \neq p_2$ we have

$$L(p_1, p_2\,|\underline{x}_1, \underline{x}_2, M_1) \;=\; \prod_{i=1}^{2}\prod_{j=1}^{M_i}[1-(1-p_i)^{m_{ij}}]^{x_{ij}}\,[(1-p_i)^{m_{ij}}]^{n_{ij}-x_{ij}}\,,$$

and

$$f(\underline{x}_1, \underline{x}_2\,|M_1) \;=\; \frac{1}{\pi^2}\int_0^1\int_0^1 p_i^{-\frac{1}{2}}\,(1-p_i)^{-\frac{1}{2}}\,L(p_1, p_2\,|\underline{x}_1, \underline{x}_2, M_1)\,dp_1 dp_2.$$

The Bayes factor in favour of M_0 is then

$$B_{01} \;=\; \frac{f(\underline{x}\,|M_0)}{f(\underline{x}_1, \underline{x}_2\,|M_1)}$$

with

$$P(M_0\,|x) \;=\; \left(1+\frac{1}{B_{01}}\right)^{-1}.$$

Using numerical integration, the Bayes factors and posterior probabilities were calculated and the histograms are shown in Figure 6.4.

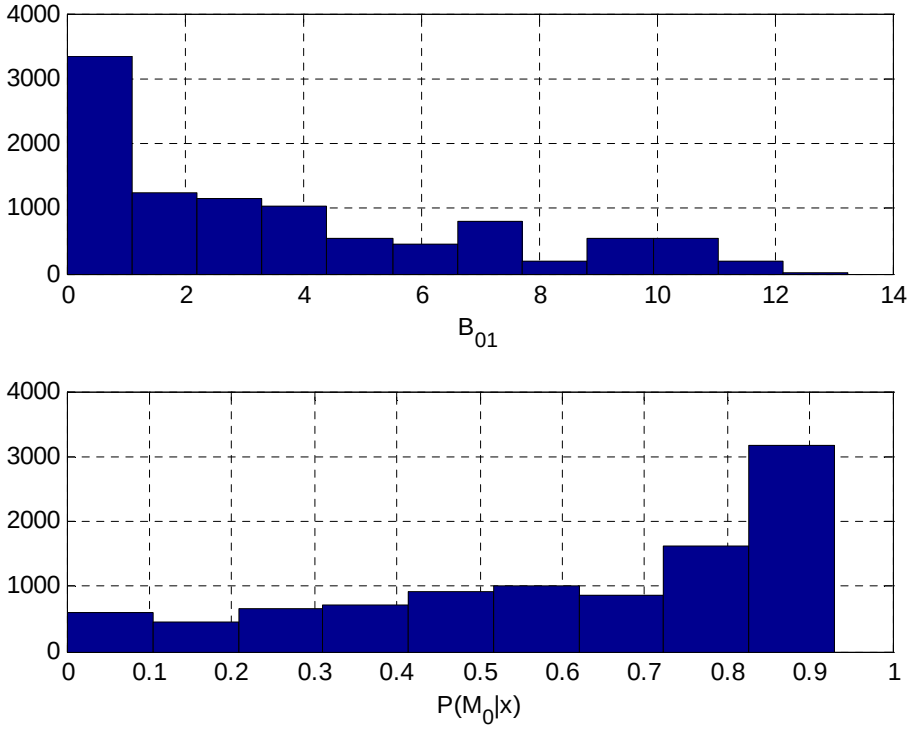


Figure 6.4: Histograms of the Bayes factor and posterior probabilities.

The mean of B_{01} is 3.6241 and the mean posterior probability is 0.6202, still favouring a single p slightly.

It is interesting to note that 626 of the 10 000 simulations gave the same result as the Biggerstaff (2008) data, 7 positives from the samples with p_1 and one positive from the samples with p_2, although not necessarily from samples with the same pool sizes. The range of posterior probabilities for the 626 simulations is $(0.6925, 0.7208)$, with mean of 0.7030. So the pools from which the positive observations come do not have a large affect on the posterior.

We find the Bayes factor a bit unsatisfactory. We would have liked it to have performed better in discriminating between the two groups. For example, when we have 11 against 2 positives for the two groups, the posterior probability of M_0 is still 0.5696.

The Bayes factor is a summary of the evidence provided by the data in favour of one scientific theory, represented by a statistical model, as opposed to another (Kass & Raftery, 1995).

Kass & Raftery (1995) gave the following categories for interpreting the Bayes factor, B_{10}:

$\log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 0.5	1 to 3.2	Not worth more than a bare mention
0.5 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Using these scales and categories to judge the evidence against M_0 for B_{01}, we obtain the following results:

- 85.12% of the time, the evidence was poor;
- 9.06% of the time, it was substantial;
- 5.49% of the time, it was strong;
- 0.33% of the time, it was decisive.

6.4 Conclusion

In this chapter we compared the proposed Bayesian method to results obtained by Biggerstaff (2008). The Jeffreys prior was used for the Bayesian method. Simulation studies were considered as well as an example. The Bayesian method compared well with the other results, and gave much better results that the Wald and minimum infection rate intervals. The Wald and the minimum infection rate intervals performed the poorest. We gave a brief overview of the Bayes factor, and applied it to an example. A $Beta(1/2, 1/2)$ prior was used for the Bayes factor. We found the Bayes factor a bit unsatisfactory. We would have liked it to have performed better in discriminating between the two groups.

Chapter 7

Bayesian Process Control for the p - chart

7.1 Introduction

Quality control is a process which is used to maintain the standards of products produced or services delivered. The binomial distribution is often used in quality control. The proportion, p, denotes the proportion of defective items in the population. The number of defective items in a random sample of size n, is denoted by x. The sample proportion of defectives items will then be $\hat{p} = x/n$, where $X \sim Bin(n, p)$. The number of nonconformities refers to the number of defectives and the number of conformities refers to the number of nondefectives.

There are three widely used attributes control charts, from Montgomery (1996):

- the control chart for the proportion of nonconforming or defective product produced by a process, and is called the p - chart;
- the control chart for nonconformities, or the c - chart deals with the number of defects or nonconformities observed;
- the control chart for nonconformities per unit, or the u - chart.

In this chapter we will look at the p - chart. Control chart limits, average run lengths and false alarm rates will be determined by using a Bayesian method. These results will be compared to the results obtained when using the classical (frequentist) method. Calabrese (1995) states that attributes control techniques, such as p - charts, plot statistics related to defective items and call for corrective action if the number of defectives becomes too large. The goal is to decide on the basis of the sample data whether the production process has shifted from an in-control state to an out-of-control state. If the production process shifted to an out-of-control state, the process should be inspected and repaired. Chakraborti & Human (2006) examined the effects of parameter estimation for the p - chart. Calabrese (1995) considered a Bayesian process control procedure with fixed samples sizes and sampling intervals where the proportion of defectives is the quality variable of interest.

Let X_i follow a binomial distribution with parameters n and p. Therefore $P(X_i = x_i) = \begin{pmatrix} n \\ x_i \end{pmatrix} p^{x_i}(1-p)^{n-x_i}$ for $x_i = 0,1,2,\ldots,n$. From Montgomery (1996) a control chart for the proportion of nonconformities with 3-sigma limits is defined as:

$$\begin{aligned} \text{UCL} &= p + 3\sqrt{\frac{p(1-p)}{n}} \\ \text{Centre line} &= p \\ \text{LCL} &= p - 3\sqrt{\frac{p(1-p)}{n}}. \end{aligned}$$

The above mentioned control chart is when the value for p is known. This case is also referred to as the "standard given" case. If there is no standard given value, then p should be estimated from the observed sample data. The proportion of nonconforming items from sample i is defined as

$$\hat{p}_i = \frac{X_i}{n} \qquad i = 1,2,\ldots,m$$

and then $\overline{p}$ is calculated, where $\overline{p}$ is the average of the sample proportions and is defined as

$$\overline{p} = \frac{\sum_{i=1}^{m} \hat{p}_i}{m}.$$

Where n is the size of each sample, and m is the number of samples. Again from Montgomery (1996) the control chart will be defined as:

$$\begin{aligned} \text{UCL} &= \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \\ \text{Centre line} &= \overline{p} \\ \text{LCL} &= \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}. \end{aligned}$$

The above mentioned control chart is when the value for p is unknown. This case is also referred to as the "no standard given" case. Chakraborti & Human (2006) state that when p is unknown, the common practice is to estimate the proportion in phase I of the study when the process is thought to be in-control. The size of each sample, n, is assumed to be equal, which is not always the case in practice. Chakraborti & Human (2006) give the following definition for phase I of the study: "Retrospective analysis used to obtain reference data; estimate any parameters". The control chart limits can then be used for future process monitoring in phase II. The following definition is given by Chakraborti & Human (2006) for phase II of the study: "Prospective monitoring of a process". The control chart limits will be used to determine average run lengths and false alarm rates. The mentioned control limits are 3-sigma limits and based on the assumption that the normal approximation to the binomial

distribution holds. We will introduce a Bayesian approach for the p - chart. For the Bayesian method, the predictive density will be used to determine the control chart. From a Bayesian point, we have to decide on a prior for this unknown value of p. This will be discussed in Section 7.2.

7.2 Prior Distribution, Posterior Distribution and Predictive density, $f(T|data)$

Menzefricke (2002) proposed a Bayesian approach to obtain control charts when there is parameter uncertainty, using a predictive distribution based on a Bayesian approach to derive the rejection region. Menzefricke (2002) assumed that the prior information on p, the proportion of defective items in the population, is a beta distribution and then the posterior distribution on p will also be a beta distribution. We will consider the Jeffreys prior.

In Chapters 2 and 3 we introduced noninformative priors which could be used in the case of the product of different powers of k binomial proportions and linear combinations of k binomial proportions, respectively. The performance of the Jeffreys, probability matching and uniform priors were evaluated in these two cases. In this section we will first evaluate the performance of the above mentioned priors in the case of a single proportion, and then we will choose a prior which will be used for process control for the p - chart.

From Theorems 2.1 and 3.1 the probability matching prior for p is given by

$$\pi_{PM}(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}. \tag{7.1}$$

The Jeffreys prior, on the other hand, is proportional to the square root of the determinant of the Fisher information matrix and is given by

$$\pi_{J}(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}. \tag{7.2}$$

From Equations 7.1 and 7.2 it is clear that the Jeffreys prior and the probability matching prior yield the same prior.

The uniform prior is proportional to a constant and is given by

$$\pi_{U}(p) \propto \text{constant}. \tag{7.3}$$

In Table 7.1, average coverage probabilities, mean lengths and standard deviations are given for $n = 10$, 20, 30 and 40. The averages are taken over values for $p = 0.1 : 0.1 : 0.9$.

Table 7.1: Coverage rate of the 95% credibility intervals for p using the Jeffreys and uniform priors. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n = 10$, 20, 30 and 40. Results are averages over values for $p = 0.1 : 0.1 : 0.9$.

		$n = 10$	$n = 20$	$n = 30$	$n = 40$
Jeffreys	(a)	0.9557	0.9550	0.9451	0.9515
	(b)	0.4532	0.3424	0.2864	0.2511
	(c)	0.0756	0.0427	0.0293	0.0224
Uniform	(a)	0.9532	0.9595	0.9605	0.9489
	(b)	0.4557	0.3429	0.2865	0.2512
	(c)	0.0569	0.0364	0.0262	0.0204

The performances of the priors are very similar, and therefore we decided to use the Jeffreys prior which is also the probability matching prior in this case. Ghosh (2011) states that the uniform prior has often been criticised due to its lack of invariance under one-to-one reparameterisation, and that the Jeffreys prior is invariant under one-to-one reparameterisation of parameters. We will therefore use the Jeffreys prior for the remainder of this chapter.

For the p - chart the likelihood follows as

$$L(p\,|data) \quad \propto \quad p^{\sum_{i=1}^{m} x_i}(1-p)^{mn-\sum_{i=1}^{m} x_i} \tag{7.4}$$

and as mentioned the Jeffreys prior will be used, which is given by

$$\pi_J(p) \quad \propto \quad p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}. \tag{7.5}$$

Combining Equations 7.4 and 7.5 it follows that the posterior distribution of p is a $Beta\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)$ distribution, i.e.

$$\pi_J(p\,|data) \quad = \quad \frac{1}{B\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)} p^{\sum_{i=1}^{m} x_i - \frac{1}{2}}(1-p)^{mn-\sum_{i=1}^{m} x_i - \frac{1}{2}}. \tag{7.6}$$

If the process remains stable, the control chart limits for a future sample of n Bernoulli trials which results in T successes can be derived. Given n and p, the distribution of T is binomial, and the

unconditional prediction distribution of T is

$$\begin{aligned} f(T\,|data) &= \int_0^1 f(T\,|p)\,\pi_J(p\,|data)\,dp \\ &= \int_0^1 \binom{n}{T} p^T (1-p)^{n-T} \frac{1}{B\left(\sum_{i=1}^m x_i+\frac{1}{2}, mn-\sum_{i=1}^m x_i+\frac{1}{2}\right)} p^{\sum_{i=1}^m x_i-\frac{1}{2}} (1-p)^{mn-\sum_{i=1}^m x_i-\frac{1}{2}} dp \\ &= \frac{\binom{n}{T}}{B\left(\sum_{i=1}^m x_i+\frac{1}{2}, mn-\sum_{i=1}^m x_i+\frac{1}{2}\right)} \int_0^1 p^{T+\sum_{i=1}^m x_i-\frac{1}{2}} (1-p)^{n-T+mn-\sum_{i=1}^m x_i-\frac{1}{2}} dp \\ &= \binom{n}{T} \frac{B\left(\sum_{i=1}^m x_i+\frac{1}{2}+T, mn-\sum_{i=1}^m x_i+\frac{1}{2}+n-T\right)}{B\left(\sum_{i=1}^m x_i+\frac{1}{2}, mn-\sum_{i=1}^m x_i+\frac{1}{2}\right)} \qquad 0 \le T \le n \end{aligned} \tag{7.7}$$

which is a beta-binomial distribution whose mean and standard deviation are easily derived. It is known as a beta-binomial distribution, because it is generated by the mixture of beta and binomial distributions. It is assumed that the sample size is the same for the posterior distribution and the future sample. The predictive distribution in Equation 7.7 can be used to obtain the control chart limits. The size of the rejection region, $R^\star(\alpha)$, is then defined as

$$\alpha = \sum_{R^\star(\alpha)} f(T\,|data)\,. \tag{7.8}$$

7.3 False Alarm Rates and Average Run lengths

If a point falls within the lower and upper control limits, the process is in-control, and if a point falls outside or on the lower or upper control limit, the process is out-of-control. When the process is out-of-control, an alarm is given. Let β denote the probability that a point plots in-control, i.e.

$$\begin{aligned} \beta &= P(\text{LCL} < \hat{p} < \text{UCL}\,|p) \\ &= P(n\text{LCL} < X_i < n\text{UCL}\,|p) \\ &= P(X_i < n\text{UCL}\,|p) - P(X_i \le n\text{LCL}\,|p)\,. \end{aligned} \tag{7.9}$$

The probability that a point plots out-of-control will therefore be $1-\beta$, i.e.

$$\begin{aligned} 1-\beta &= 1 - P(n\text{LCL} < X_i < n\text{UCL}\,|p) \\ &= P(X_i \ge n\text{UCL}\,|p) + P(X_i \le n\text{LCL}\,|p)\,. \end{aligned} \tag{7.10}$$

If the process is in-control, the probability that a point plots out-of-control is also known as the false alarm rate (FAR).

The probability that a point plots in-control is used to derive the run length distribution. The average run length (ARL) is calculated as:

$$\begin{aligned} \text{ARL} &= \frac{1}{P(\text{sample point plots out of control})} \\ &= \frac{1}{P(X_i \geq n\text{UCL}\,|\,p) + P(X_i \leq n\text{LCL}\,|\,p)}. \end{aligned} \tag{7.11}$$

Montgomery (1996) defines the average run length as the average number of points that must be plotted before a point indicates an out-of-control condition. If the process is in-control, the expected nominal value for the false alarm rate is 0.0027 and the expected nominal value for the average run length is $(0.0027)^{-1} = 370.3704$. That means even if the process remains in control, an out-of-control signal will be generated every 370 samples, on average.

7.4 Example and Simulation Studies

7.4.1 Simulation Study I

In this simulation study the unconditional average run length and unconditional false alarm rate will be compared using the classical (frequentist) method and the proposed Bayesian method. Two cases will be considered, one where $m = 4$ and $n = 5$ and the other case will be when $m = 2$ and $n = 10$. The values for m and n are the same as the values used by Chakraborti & Human (2006). Lower and upper control limits will be calculated for the two situations, and the unconditional average run length and unconditional false alarm rate will be determined for the two situations. The results will be given in Tables 7.2 to 7.5.

Using the control chart limits stated in in Section 7.1 we can determine the np control chart limits for the frequentist method:

$$\begin{aligned} n\text{UCL} &= n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})} \\ n\text{LCL} &= n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} \end{aligned}$$

with

$$\overline{p} = \frac{\sum_{i=1}^{m} \hat{p}_i}{m}$$

where

$$\hat{p}_i = \frac{X_i}{n} \qquad i = 1,2,\ldots,m.$$

The predictive density given in Equation 7.7 will be used to obtain the control chart limits when the Bayesian approach is used. The predictive density is a beta-binomial distribution with parameters $\sum_{i=1}^{m} x_i + \frac{1}{2}$ and $mn - \sum_{i=1}^{m} x_i + \frac{1}{2}$. The mean of the predictive density will therefore be

$$E(T|data) = \frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)}{mn+1}$$

and the variance will be

$$Var(T|data) = \frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)\left(mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)(mn+1+n)}{(mn+1)^2(mn+2)}.$$

Using the Bayesian method, the control chart limits are calculated as

$$n\text{UCL} = \frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)}{mn+1} + 3\sqrt{\frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)\left(mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)(mn+1+n)}{(mn+1)^2(mn+2)}}$$

$$n\text{LCL} = \frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)}{mn+1} - 3\sqrt{\frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)\left(mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)(mn+1+n)}{(mn+1)^2(mn+2)}}.$$

The values for nLCL and nUCL which will be used to calculate the average run length and the false alarm rate should be integer values, since we are working with the binomial distribution. Chakraborti & Human (2006) state that the following values should be used for nLCL and nUCL in calculations:

$$A = \max\{0, [n\text{LCL}]\}$$

and

$$B = \begin{cases} \min\{n\text{UCL} - 1, n\} & \text{if } n\text{UCL is an integer} \\ \min\{[n\text{UCL}], n\} & \text{if } n\text{UCL is not an integer} \end{cases}$$

where $[n\text{UCL}]$ denotes the largest integer not exceeding nUCL and $[n\text{LCL}]$ denotes the largest integer not exceeding nLCL.

Using Equations 7.9, 7.10 and 7.11 we can determine the average run length and the false alarm rate. After obtaining the control limits, the conditional false alarm rate (CFAR) has to be calculated

and then one can determine the conditional average run length (CARL) for each possible value of $\sum_{i=1}^{m} X_i$. After obtaining the values for CFAR and CARL, the unconditional false alarm rate and the unconditional average run length can be calculated. The false alarm rate will be calculated as $1-\beta$, where β is calculated as

$$\beta = \begin{cases} P(X_i \leq B) & \text{if } n\text{LCL is negative} \\ P(X_i \leq B) - P(X_i \leq A) & \text{if } n\text{LCL is positive} \end{cases}$$

and the in-control conditional average run length is calculated as

$$\text{ICARL} = \left(\frac{1}{1-\beta}\right) \times P\left(\sum_{i=1}^{m} X_i = \sum_{i=1}^{m} x_i\right)$$

where

$$P\left(\sum_{i=1}^{m} X_i = \sum_{i=1}^{m} x_i\right) = \left(\begin{array}{c} mn \\ \sum_{i=1}^{m} x_i \end{array}\right) 0.5^{\sum_{i=1}^{m} x_i} 0.5^{mn - \sum_{i=1}^{m} x_i}.$$

The unconditional false alarm rate will be determined by

$$\text{UFAR} = \sum_{j=0}^{mn} \left[(1-\beta) \times P\left(\sum_{i=1}^{m} X_i = \sum_{i=1}^{m} x_i\right) \right]_j \tag{7.12}$$

and the unconditional average run length will be determined by

$$\text{UARL} = \sum_{j=0}^{mn} [\text{ICARL}]_j . \tag{7.13}$$

For the first study we have $m = 4$ and $n = 5$, where m is the number of samples and n is the sample size. This was also considered by Chakraborti & Human (2006), where they assumed that $p_0 = p_1 = 0.5$, where p_0 is the known value of the proportion of nonconforming items and p_1 is the true proportion of nonconforming items in phase II. Table 7.2 gives the result for the classical method, which was also obtained by Chakraborti & Human (2006). Table 7.3 gives the results for the Bayesian method, proposed by us.

Table 7.2: Control limits, false alarm rate and average run length for p - chart when $m = 4$ and $n = 5$, using the classical (frequentist) method.

$\sum_{i=1}^{m} X_i$	$\overline{p}$	nLCL	nUCL	A	B	CFAR	ICARL
0	0.00	0.0000	0.0000	0	0	1.00000	9.5367E-07
1	0.05	-1.2120	1.7120	0	1	0.81250	2.3475E-05
2	0.10	-1.5125	2.5125	0	2	0.50000	0.0003624
3	0.15	-1.6453	3.1453	0	3	0.18750	0.0057983
4	0.20	-1.6833	3.6833	0	3	0.18750	0.0246430
5	0.25	-1.6547	4.1547	0	4	0.03125	0.4731400
6	0.30	-1.5741	4.5741	0	4	0.03125	1.1829000
7	0.35	-1.4496	4.9496	0	4	0.03125	2.3657000
8	0.40	-1.2863	5.2863	0	5	0.00000	∞
9	0.45	-1.0873	5.5873	0	5	0.00000	∞
10	0.50	-0.8541	5.8541	0	5	0.00000	∞
11	0.55	-0.5873	6.0873	0	5	0.00000	∞
12	0.60	-0.2863	6.2863	0	5	0.00000	∞
13	0.65	0.0504	6.4496	0	5	0.03125	2.3657000
14	0.70	0.4259	6.5741	0	5	0.03125	1.1829000
15	0.75	0.8453	6.6547	0	5	0.03125	0.4731400
16	0.80	1.3167	6.6833	1	5	0.18750	0.0246430
17	0.85	1.8547	6.6453	1	5	0.18750	0.0057983
18	0.90	2.4875	6.5125	2	5	0.50000	0.0003624
19	0.95	3.2880	6.2120	3	5	0.81250	2.3475E-05
20	1.00	5.0000	5.0000	5	5	1.00000	9.5367E-07
UFAR = 0.010209 UARL = ∞							

Table 7.3: Control limits, false alarm rate and average run length for p - chart when $m = 4$ and $n = 5$, using the Bayesian method.

$\sum_{i=1}^{m} X_i$	nLCL	nUCL	A	B	CFAR	ICARL
0	-0.9928	1.2308	0	1	0.81250	1.1738E-06
1	-1.5210	2.2353	0	2	0.50000	3.8147E-05
2	-1.7664	2.9569	0	2	0.50000	0.0003624
3	-1.8845	3.5511	0	3	0.18750	0.0057983
4	-1.9209	4.0638	0	4	0.03125	0.1478600
5	-1.8968	4.5159	0	4	0.03125	0.4731400
6	-1.8237	4.9190	0	4	0.03125	1.1829000
7	-1.7086	5.2800	0	5	0.00000	∞
8	-1.5557	5.6033	0	5	0.00000	∞
9	-1.3678	5.8916	0	5	0.00000	∞
10	-1.1463	6.1463	0	5	0.00000	∞
11	-0.8916	6.3678	0	5	0.00000	∞
12	-0.6034	6.5557	0	5	0.00000	∞
13	-0.2800	6.7086	0	5	0.00000	∞
14	0.08104	6.8237	0	5	0.03125	1.1829000
15	0.4841	6.8968	0	5	0.03125	0.4731400
16	0.9362	6.9209	0	5	0.03125	0.1478600
17	1.4489	6.8845	1	5	0.18750	0.0057983
18	2.0431	6.7664	2	5	0.50000	0.0003624
19	2.7647	6.5210	2	5	0.50000	3.8147E-05
20	3.7692	5.9927	3	5	0.81250	1.1738E-06
UFAR = 0.0041327 UARL = ∞						

From Table 7.2, we see that the unconditional false alarm rate, using the classical method, is equal to 0.010209 and is much higher than the nominal value of 0.0027. From Table 7.3 the unconditional false alarm rate, using the Bayesian method, is equal to 0.004132. This value is still higher than the nominal value of 0.0027, but it is much closer to the nominal value than the value obtained from the classical method. The unconditional average run length is equal to infinity when using the classical and the Bayesian method, which seems unrealistic. For values of $\sum_{i=1}^{m} X_i = 0$ and $\sum_{i=1}^{m} X_i = 20$ the frequentist method yields conditional false alarm rates of 1, with the Bayesian method one never has conditional false alarm rates of 1 in this case. The intervals are generally a bit wider when using the Bayesian method. Chakraborti & Human (2006) also state that even though one typically wants a high value for the average run length, infinity is not a practical choice. Thus 4 samples of size 5 each is not sufficient to control the false alarm rate.

For the second study we have $m = 2$ and $n = 10$. This was also considered by Chakraborti & Human (2006), where they assumed that $p_0 = p_1 = 0.5$, where p_0 is the known value of the proportion of nonconforming items and p_1 is the true proportion of nonconforming items in phase II. Table 7.4 gives the result for the classical method, which was also obtained by Chakraborti & Human (2006), Table 7.5 gives the results for the Bayesian method, proposed by us.

Table 7.4: Control limits, false alarm rate and average run length for p - chart when $m = 2$ and $n = 10$, using the classical (frequentist) method.

$\sum_{i=1}^{m} X_i$	$\overline{p}$	nLCL	nUCL	A	B	CFAR	ICARL
0	0.00	0.0000	0.000	0	0	1.00000	9.5367E-07
1	0.05	-1.5676	2.5676	0	2	0.94531	2.0177E-05
2	0.10	-1.8460	3.8460	0	3	0.82812	0.0002188
3	0.15	-1.8875	4.8875	0	4	0.62305	0.0017450
4	0.20	-1.7947	5.7947	0	5	0.37695	0.0122580
5	0.25	-1.6079	6.6079	0	6	0.17187	0.0860260
6	0.30	-1.3474	7.3474	0	7	0.05469	0.6759200
7	0.35	-1.0249	8.0249	0	8	0.01074	6.8821000
8	0.40	-0.6476	8.6476	0	8	0.01074	11.1830000
9	0.45	-0.2196	9.2196	0	9	0.00098	164.020000
10	0.50	0.2566	9.7434	0	9	0.00195	90.2130000
11	0.55	0.7804	10.2200	0	10	0.00098	164.020000
12	0.60	1.3524	10.6480	1	10	0.01074	11.1830000
13	0.65	1.9751	11.0250	1	10	0.01074	6.8821000
14	0.70	2.6526	11.3470	2	10	0.05469	0.6759200
15	0.75	3.3921	11.6080	3	10	0.17188	0.0860260
16	0.80	4.2053	11.7950	4	10	0.37695	0.0122580
17	0.85	5.1125	11.8870	5	10	0.62305	0.0017450
18	0.90	6.1540	11.8460	6	10	0.82813	0.0002188
19	0.95	7.4324	11.5680	7	10	0.94531	2.0177E-05
20	1.00	10.0000	10.0000	10	10	1.00000	9.5367E-07
UFAR = 0.019128 UARL = 455.9432							

Table 7.5: Control limits, false alarm rate and average run length for p - chart when $m = 2$ and $n = 10$, using the Bayesian method.

$\sum_{i=1}^{m} X_i$	nLCL	nUCL	A	B	CFAR	ICARL
0	-1.4788	1.9549	0	1	0.98926	9.6403E-07
1	-2.1860	3.6145	0	3	0.82812	2.3032E-05
2	-2.4565	4.8374	0	4	0.62305	0.0002908
3	-2.5302	5.8635	0	5	0.37695	0.0028841
4	-2.4780	6.7637	0	6	0.17187	0.0268830
5	-2.3322	7.5703	0	7	0.05469	0.2703700
6	-2.1109	8.3013	0	8	0.01074	3.4411000
7	-1.8245	8.9674	0	8	0.01074	6.8821000
8	-1.4800	9.5752	0	9	0.00098	123.02000
9	-1.0813	10.1290	0	10	0.00000	∞
10	-0.6307	10.6310	0	10	0.00000	∞
11	-0.1289	11.0810	0	10	0.00000	∞
12	0.4248	11.4800	0	10	0.00098	123.02000
13	1.0326	11.8250	1	10	0.01074	6.8821000
14	1.6987	12.1110	1	10	0.01074	3.4411000
15	2.4297	12.3320	2	10	0.05469	0.2703700
16	3.2363	12.4780	3	10	0.17188	0.0268830
17	4.1365	12.5300	4	10	0.37695	0.0028841
18	5.1626	12.4560	5	10	0.62305	0.0002908
19	6.3855	12.1860	6	10	0.82813	2.3032E-05
20	8.0451	11.4790	8	10	0.98926	9.6403E-07
UFAR = 0.0069015 UARL = ∞						

From Table 7.4, we see that the unconditional false alarm rate, using the classical method, is equal to 0.019128 and is much higher than the nominal value of 0.0027. From Table 7.5 the unconditional false alarm rate, using the Bayesian method, is equal to 0.0069015. This value is still higher than the nominal value of 0.0027, but it is much closer to the nominal value than the value obtained from the classical method. The unconditional average run length is equal to infinity when using the Bayesian method, which seems unrealistic. Whereas the unconditional average run length is equal to 455.9432 when using the classical method, which is more realistic and practical. For values of $\sum_{i=1}^{m} X_i = 0$ and $\sum_{i=1}^{m} X_i = 20$ the frequentist method yields conditional false alarm rates of 1, with the Bayesian method one never has conditional false alarm rates of 1 in this case. The intervals are generally slightly wider when using the Bayesian method.

7.4.2 Simulation Study II

We consider another study where different values for m and n are given. Since mn is rather large for each case, we only give the unconditional false alarm rate and the unconditional average run length for each case. Using the same methods that was used to obtain the results in Tables 7.2 to 7.5 and Equations 7.12 and 7.13 in Section 7.4.1, the unconditional false alarm rate and the unconditional average run length can be determined. The results are given in Table 7.6.

Table 7.6: Unconditional false alarm rates (UFAR) and unconditional average run lengths (UARL) for the p - chart for different values of m and n when $p_0 = p_1 = 0.5$.

			Frequentist		**Bayesian**	
m	n	mn	UFAR	UARL	UFAR	UARL
25	20	500	0.002577	470.7250	0.002137	548.9617
20	25	500	0.002905	379.2967	0.002875	567.4341
30	20	600	0.002482	478.4399	0.002120	545.2668
20	30	600	0.002976	413.9184	0.002299	519.5687
35	30	1 050	0.002658	445.4722	0.002283	516.1604
30	35	1 050	0.002753	410.7182	0.002352	465.0252
40	30	1 200	0.002608	453.1612	0.002277	517.4981
30	40	1 200	0.002841	378.6987	0.002371	473.6618

From Table 7.6, we conclude that for every combination of m and n the unconditional false alarm rate is smaller when using the Bayesian method, and that the unconditional average run length is larger when using the Bayesian method. Typically, one wants a smaller false alarm rate and a larger average run length. When using the frequentist method for $m = 30$ and $n = 35$ the false alarm rate is equal to 0.002753, this value is almost equal to the nominal value of 0.0027. When using the Bayesian method the false alarm rate is below the nominal value of 0.0027 for all combinations of m and n, except for $m = 20$ and $n = 25$.

7.4.3 Example

Consider the following example from Montgomery (1996), Example 6-1 on page 255. Chakraborti & Human (2006) also considered this example. Frozen orange juice concentrate is packed in 6-oz cardboard cans. These cans are formed on a machine by spinning them from cardboard stock and attaching a metal bottom panel. By inspection of a can, we may determine whether, when filled, it could possibly leak either on the side seam or around the bottom joint. Such a nonconforming can has an improper seal on either the side seam or the bottom panel. We wish to set up a control chart to improve the proportion of nonconforming cans produced by this machine. To establish the control chart, 30 samples of $n = 50$ cans each were selected at half-hour intervals over a three-shift period in which the machine was in continuous operation. Once the control chart was established, samples 15

and 23 were found to be out-of-control, and eliminated after further investigation. Revised limits were calculated using the remaining samples, with $m = 28$ and $n = 50$. Based on the revised control limits, sample 21 was found to be out-of-control. Since further investigations regarding sample 21 did not produce any reasonable or logical assignable cause, it was not discarded.

For this given data set $\sum_{i=1}^{28} x_i = 301$, the total number of nonconforming cans after discarding samples 15 and 23, is observed. Considering this example, we have, $n = 50$, $m = 28$ and $\sum_{i=1}^{m} x_i = 301$. As mentioned, we will use the Jeffreys prior for p, $\pi_J(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$. This will result in a beta posterior, $Beta\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)$. For this example the posterior distribution of p will be a $Beta(301.5, 1099.5)$. Figure 7.1 shows the posterior distribution of p.

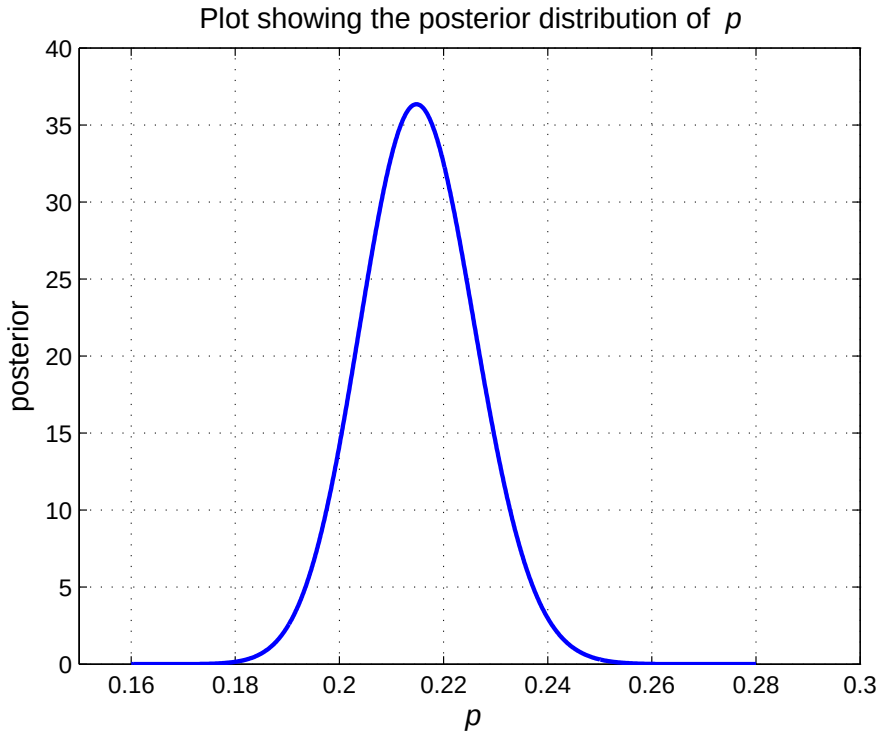


Figure 7.1: Posterior distribution of p, when $n = 50$, $m = 28$ and $\sum_{i=1}^{m} x_i = 301$.

Summary statistics for p:

mean = 0.2152 standard deviation = 0.0110 median = 0.2150

95% credibility interval = (0.1940; 0.2371).

Figure 7.2 shows a bar graph of the predictive density function, $f(T\,|data)$, where T is the number of nonconformities in a future sample. The predictive density will be used to obtain the control chart limits.

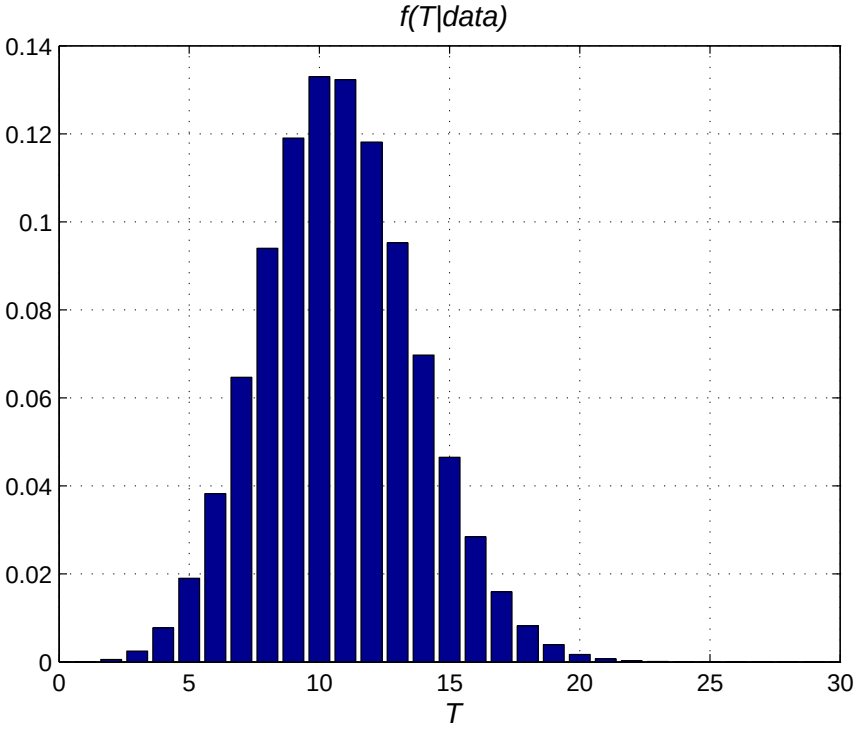


Figure 7.2: Bar graph of the predictive density of $f(T|data)$.

For the observed value, $\sum_{i=1}^{m} x_i = 301$, the chart's performance will be investigated by looking at the conditional average run length (CARL) and the conditional false alarm rate (CFAR). The control chart limits, conditional average run length and conditional average false alarm rate will be calculated using the classical (frequentist) method and using the proposed Bayesian method. Chakraborti & Human (2006) determined these values using the classical method, the results are given in Table 7.7. We used the proposed Bayesian method to determine these values, the results are also given in Table 7.7.

Calculations used to obtain the frequentist control limits:

$$\begin{aligned} n\text{UCL} &= n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})} \\ &= 50\left(\frac{301}{1\,400}\right) + 3\sqrt{50\left(\frac{301}{1\,400}\right)\left(1-\frac{301}{1\,400}\right)} \\ &= 19.4649 \end{aligned}$$

$$\begin{aligned} n\text{LCL} &= n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} \\ &= 50\left(\frac{301}{1\,400}\right) - 3\sqrt{50\left(\frac{301}{1\,400}\right)\left(1-\frac{301}{1\,400}\right)} \\ &= 2.0351 \end{aligned}$$

$$\begin{aligned} A &= \max\{0,[2.0351]\} \\ &= \max\{0,2\} = 2 \\ B &= \min\{[19.4649],50\} \\ &= \min\{19,50\} = 19. \end{aligned}$$

Calculations used to obtain the Bayesian control limits:

$$\begin{aligned} n\text{UCL} &= \frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)}{mn+1} + 3\sqrt{\frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)\left(mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)(mn+1+n)}{(mn+1)^2(mn+2)}} \\ &= \frac{50\left(301+\frac{1}{2}\right)}{1\,400+1} + 3\sqrt{\frac{50\left(301+\frac{1}{2}\right)\left(1\,400-301+\frac{1}{2}\right)(1\,400+1+50)}{(1\,400+1)^2(1\,400+2)}} \\ &= 19.6291 \end{aligned}$$

$$\begin{aligned} n\text{LCL} &= \frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)}{mn+1} - 3\sqrt{\frac{n\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)\left(mn - \sum_{i=1}^{m} x_i + \frac{1}{2}\right)(mn+1+n)}{(mn+1)^2(mn+2)}} \\ &= \frac{50\left(301+\frac{1}{2}\right)}{1\,400+1} - 3\sqrt{\frac{50\left(301+\frac{1}{2}\right)\left(1\,400-301+\frac{1}{2}\right)(1\,400+1+50)}{(1\,400+1)^2(1\,400+2)}} \\ &= 1.8913 \end{aligned}$$

$$\begin{aligned} A &= \max\{0,[1.8913]\} \\ &= \max\{0,1\} = 1 \\ B &= \min\{[19.6291],50\} \\ &= \min\{19,50\} = 19. \end{aligned}$$

Calculations used to obtain the average run lengths and false alarm rates:

$$\begin{aligned} \text{CFAR} &= 1 - [P(X_i \leq B \,|\, p = 0.2) - P(X_i \leq A \,|\, p = 0.2)] \\ \text{CARL} &= \frac{1}{\text{CFAR}}. \end{aligned}$$

Table 7.7: Lower control limits, upper control limits, conditional average run lengths (CARL) and conditional false alarm rates (CFAR) for $n = 50$, $m = 28$ and $\sum_{i=1}^{m} x_i = 301$.

Frequentist					
nLCL	nUCL	A	B	CARL	CFAR
2.0351	19.4649	2	19	450.8868	0.0022179
Bayesian					
nLCL	nUCL	A	B	CARL	CFAR
1.8913	19.6291	1	19	888.7981	0.0011251

From Table 7.7, we see that the Bayesian method gives a wider interval than the classical method. This results in a larger value for CARL and a smaller value for CFAR. The conditional false alarm rates (CFAR) obtained from both methods yield values smaller than the expected nominal value of 0.0027.

Using the same method as given in the simulation study in Section 7.4.1 the unconditional average run length and the unconditional false alarm rate can be determined. These results are given in Table 7.8.

Table 7.8: Unconditional average run lengths (UARL) and unconditional false alarm rates (UFAR) using the classical method and the Bayesian method.

Frequentist		**Bayesian**	
UARL	UFAR	UARL	UFAR
401.5103	0.0035	473.6610	0.0030

From Table 7.8, we see that the false alarm rate obtained from the Bayesian method is smaller than the false alarm rate obtained from the classical method. A smaller value for the false alarm rate is desired, since the false alarm rate is the probability of a signal being given by the control chart when the process is actually in-control. The average run length obtained from the Bayesian method is larger than the average run length obtained from the classical method. A larger value for the average run length is desired.

In Section 7.4 the mean and the variance of the predictive density was used to construct the 3-sigma limits. This method gave smaller values for the false alarm rate and larger values for the average run length, but for certain values average run lengths of infinity was observed, which is not satisfactory. This problem can be solved if the rejection region of size α is defined as the equal-tailed intervals of the predictive density where $\alpha = 0.0027$. This will be investigated in future research.

7.5 Conclusion

The usual operation of the p - chart was extended by introducing a Bayesian approach for the p - chart. The Jeffreys prior was used. We conclude that the proposed Bayesian method gives wider control limits than those obtained from the classical method. The Bayesian method gives larger values for the average run length and smaller values for the false alarm rate. A smaller value for the false alarm rate is desired.

Chapter 8

Bayesian Process Control for the c - chart

8.1 Introduction

In this chapter the c - chart will be studied. Control chart limits, average run lengths and false alarm rates will be determined by using a Bayesian method. These results will be compared to the results obtained when using the classical (frequentist) method.

The c - chart or the control chart for nonconformities is designed for the case where one deals with the number of defects or nonconformities observed. A control chart can be developed for the total or average number of nonconformities per unit, which is well modelled by the Poisson distribution. In most of the standard textbooks on Quality Control, see for example Montgomery (1996), the Poisson parameter is indicated by c. To be consistent with the text in this thesis, we will use λ to indicate the Poisson parameter. The inspection unit should be the same for each sample. Let X_i follow a Poisson distribution with parameter λ. Therefore $P(X_i = x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$ for $x_i = 0, 1, 2, \ldots$.

From Montgomery (1996) a control chart for nonconformities with 3-sigma limits is defined as:

$$\begin{aligned} \text{UCL} &= \lambda + 3\sqrt{\lambda} \\ \text{Centre line} &= \lambda \\ \text{LCL} &= \lambda - 3\sqrt{\lambda}. \end{aligned}$$

The above mentioned control chart is when the value of λ is known. This case is also referred to as the "standard given" case. If there is no standard given value, then λ should be estimated as the average number of nonconformities in an initial sample. This average number of nonconformities will

be indicated by $\overline{\lambda}$. Again from Montgomery (1996) the control chart will be defined as:

$$\begin{aligned} \text{UCL} &= \overline{\lambda} + 3\sqrt{\overline{\lambda}} \\ \text{Centre line} &= \overline{\lambda} \\ \text{LCL} &= \overline{\lambda} - 3\sqrt{\overline{\lambda}}. \end{aligned}$$

The above mentioned control chart is when the value of λ is unknown. This case is also referred to as the "no standard given" case. As mentioned in Chapter 7, when the parameter is unknown, the common practice is to estimate the parameter from phase I of the study. Where phase I is as defined in Chapter 7. The mentioned control limits are 3-sigma limits and based on the assumption that the normal approximation to the Poisson distribution holds. As stated by Montgomery (1996), if the value obtained for the LCL is negative, one should set the lower control limit equal to zero. Once the control chart is set, independent inspection units are selected, and the number of nonconformities in the each inspection unit is determined and plotted on the chart. If a point falls within the lower and upper control limit, the process is in-control. If a point falls outside or on the lower or upper control limit, the process is out-of-control. When this happens, an alarm or a signal is given.

Bayarri & García-Donato (2005) highlight the following concerns with using the classical approach:

- The Poisson model is often a very poor fit to this type of data;
- there is quite a long period in which the process is not controlled at all, namely the one used to estimate the parameter (i.e. phase I);
- previous information cannot be incorporated in any way.

Bayarri & García-Donato (2005) extended the usual operation of the u - chart, by introducing an empirical Bayesian model and a Bayesian sequential approach. We will introduce a Bayesian approach for the c - chart. From a Bayesian point, we have to decide on a prior distribution for this unknown value of λ. This will be discussed in Section 8.2.

8.2 Prior Distribution, Posterior Distribution and Predictive density, $f\left(x_f \,|data\right)$

In Chapters 4 and 5 we introduced noninformative priors which could be used in the case of the product of different powers of k Poisson rates and linear functions of k Poisson rates, respectively. The performance of the Jeffreys, probability matching and uniform priors were evaluated in these two chapters. In this section we will first evaluate the performance of the above mentioned priors in the

case of a single Poisson rate, and then we will choose a prior which will be used for process control for the c - chart.

From Theorems 4.1 and 5.1 and Corollary 5.1 the probability matching prior for λ is given by

$$\pi_{PM}(\lambda) \propto \lambda^{-\frac{1}{2}}. \tag{8.1}$$

The Jeffreys prior, on the other hand, is proportional to the square root of the determinant of the Fisher information matrix and is given by

$$\pi_{J}(\lambda) \propto \lambda^{-\frac{1}{2}}. \tag{8.2}$$

From Equations 8.1 and 8.2 it is clear that the Jeffreys prior and the probability matching prior yield the same prior.

The uniform prior is proportional to a constant and is given by

$$\pi_{U}(\lambda) \propto \text{constant}. \tag{8.3}$$

In Table 8.1, average coverage probabilities, mean lengths and standard deviations are given for $\lambda = 2:1:10$, $\lambda = 5:2:59$ and $\lambda = 0.1:0.1:20$. The averages are taken over the different values of λ.

Table 8.1: Coverage rate of the 95% credibility intervals for λ using the Jeffreys and uniform priors. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation, the results are averages over the different values of λ.

		$\lambda = 2:1:10$	$\lambda = 5:2:59$	$\lambda = 0.1:0.1:20$
Jeffreys	(a)	0.9494	0.9489	0.9502
	(b)	9.4265	21.3418	11.8281
	(c)	1.9434	1.9757	1.9176
Uniform	(a)	0.9503	0.95077	0.9491
	(b)	9.8728	21.5496	12.2147
	(c)	1.8482	1.9458	1.8368

The performances of the priors are very similar, and therefore we decided to use the Jeffreys prior which is also the probability matching prior in this case. As mentioned in Chapter 7, Ghosh (2011) states that the uniform prior has often been criticised due to its lack of invariance under one-to-one reparameterisation and that the Jeffreys prior is invariant under one-to-one reparameterisation of parameters. We will therefore use the Jeffreys prior for the remainder of this chapter.

As for the binomial case, in Chapter 7, the predictive density will be used to obtain the control chart. If independent inspection units are randomly selected at equally spaced time intervals, the number of nonconformities in the i^{th} inspection will follow a Poisson distribution with parameter λ. Therefore

$P(X_i = x_i) = \dfrac{\lambda^{x_i} e^{-\lambda}}{x_i!}$ for $x_i = 0, 1, 2, \ldots$ and $i = 1, \ldots, m$. The likelihood function will thus be

$$L(\lambda \,|data) \;\; \propto \;\; e^{-m\lambda} \lambda^{\sum_{i=1}^{m} x_i}. \tag{8.4}$$

As mentioned the Jeffreys prior will be used, which is given by

$$\pi_J(\lambda) \;\; \propto \;\; \lambda^{-\frac{1}{2}}. \tag{8.5}$$

Combining Equations 8.4 and 8.5 it follows that the posterior distribution of λ is a $Gamma\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, m\right)$ distribution, i.e.

$$\pi_J(\lambda \,|data) \;\; = \;\; \frac{m^{\sum_{i=1}^{m} x_i + \frac{1}{2}}}{\Gamma\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)} e^{-m\lambda} \lambda^{\sum_{i=1}^{m} x_i - \frac{1}{2}} \qquad \lambda > 0. \tag{8.6}$$

Denote the number of nonconformities in a future inspection unit by x_f, then the predictive density (if λ is known) is

$$f\left(x_f \,|\lambda\right) \;\; = \;\; e^{-\lambda} \frac{\lambda^{x_f}}{x_f!} \qquad x_f = 0, 1, \ldots, \infty \tag{8.7}$$

and the unconditional predictive density is

$$\begin{aligned}
f\left(x_f \,|data\right) \;\; &= \;\; \int_0^{\infty} f\left(x_f \,|\lambda\right) \pi_J(\lambda \,|data)\, d\lambda \\
&= \;\; \int_0^{\infty} e^{-\lambda} \frac{\lambda^{x_f}}{x_f!} \frac{m^{\sum_{i=1}^{m} x_i + \frac{1}{2}}}{\Gamma\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)} e^{-m\lambda} \lambda^{\sum_{i=1}^{m} x_i - \frac{1}{2}} d\lambda \\
&= \;\; \frac{m^{\sum_{i=1}^{m} x_i + \frac{1}{2}}}{\Gamma\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right) x_f!} \int_0^{\infty} e^{-(m+1)\lambda} \lambda^{x_f + \sum_{i=1}^{m} x_i - \frac{1}{2}} d\lambda \\
&= \;\; \frac{m^{\sum_{i=1}^{m} x_i + \frac{1}{2}}}{\Gamma\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right) x_f!} \frac{\Gamma\left(x_f + \sum_{i=1}^{m} x_i + \frac{1}{2}\right)}{(m+1)^{x_f + \sum_{i=1}^{m} x_i + \frac{1}{2}}} \qquad x_f = 0, 1, \ldots, \infty
\end{aligned} \tag{8.8}$$

which is a Poisson-gamma distribution whose mean and standard deviation are easily derived. It is known as a Poisson-gamma distribution, because it is generated by the mixture of Poisson and gamma

distributions. As before the predictive distribution in Equation 8.8 can be used to obtain the control chart limits. The size of the rejection region, $R^{\star}(\alpha)$, is then defined as

$$\alpha = \sum_{R^{\star}(\alpha)} f\left(x_f \,|data\right). \tag{8.9}$$

8.3 False Alarm Rates and Average Run Lengths

If a point falls within the lower and upper control limit, the process is in-control. If a point falls outside or on the lower or upper control limit, the process is out-of-control. When this happens, an alarm or a signal is given. Let β denote the probability of a "no-signal", i.e. the probability that a point plots in-control. The probability of a "no-signal" will thus be

$$\begin{aligned} \beta &= P\left(\text{LCL} < X_i < \text{UCL}|\lambda\right) \\ &= P\left(X_i < \text{UCL}|\lambda\right) - P\left(X_i \leq \text{LCL}|\lambda\right). \end{aligned} \tag{8.10}$$

The probability of a "signal" when the process is in-control, also known as the false alarm rate (FAR), will thus be

$$\begin{aligned} 1-\beta &= 1 - P\left(\text{LCL} < X_i < \text{UCL}|\lambda\right) \\ &= P\left(X_i \geq \text{UCL}|\lambda\right) + P\left(X_i \leq \text{LCL}|\lambda\right). \end{aligned} \tag{8.11}$$

Chakraborti & Human (2008) state that the performance of a control chart is usually judged on the basis of its run length distribution. Chakraborti & Human (2008) give the following definition for the run length distribution: "The run length distribution is the probability distribution of the random variable, say N, which denotes the number of inspection units that must be sampled before the first signal is observed on the chart".

Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the run lengths. Given λ and a stable process, the distribution of the run length $r^{\star}$ is geometric with parameter $1-\beta$, the probability of a "signal". The average run length (ARL) is calculated as:

$$\begin{aligned} \text{ARL} &= \frac{1}{P\left(\text{sample point plots out of control}\right)} \\ &= \frac{1}{P\left(X_i \geq \text{UCL}|\lambda\right) + P\left(X_i \leq \text{LCL}|\lambda\right)}. \end{aligned} \tag{8.12}$$

If the process is in-control, the expected nominal value for the false alarm rate is 0.0027 and the expected nominal value for the average run length is $(0.0027)^{-1} = 370.3704$.

8.4 Example and Simulation Study

8.4.1 Simulation Study

Our aim in this simulation study is to compare the unconditional average run lengths and unconditional false alarm rates using the frequentist method and a Bayesian method. Lower and upper control limits will be calculated for given m and λ values. The values used for m and λ are the same as the values used by Chakraborti & Human (2008).

As mentioned in Section 8.1 the control chart limits, using the frequentist method, are calculated as follow

$$\begin{aligned} \text{UCL} &= \overline{\lambda} + 3\sqrt{\overline{\lambda}} \\ \text{LCL} &= \overline{\lambda} - 3\sqrt{\overline{\lambda}} \end{aligned} \tag{8.13}$$

with the average number of nonconformities

$$\overline{\lambda} = \frac{\sum_{i=1}^{m} x_i}{m}.$$

These are usually referred to as 3-sigma control limits.

The predictive density given in Equation 8.8 will be used to obtain the control chart limits when using the Bayesian approach. The predictive density is a Poisson-gamma distribution with parameters $\sum_{i=1}^{m} x_i + \frac{1}{2}$ and m. The mean of the predictive density will therefore be

$$E\left(x_f \,|data\right) = \frac{\sum_{i=1}^{m} x_i + \frac{1}{2}}{m} \tag{8.14}$$

and the variance will be

$$Var\left(x_f \,|data\right) = \frac{\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)(m+1)}{m^2}. \tag{8.15}$$

Using the Bayesian method, the control chart limits are calculated as:

$$\begin{aligned} \text{UCL} &= \frac{\sum_{i=1}^{m} x_i + \frac{1}{2}}{m} + 3\sqrt{\frac{\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)(m+1)}{m^2}} \\ \text{LCL} &= \frac{\sum_{i=1}^{m} x_i + \frac{1}{2}}{m} - 3\sqrt{\frac{\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)(m+1)}{m^2}}. \end{aligned} \tag{8.16}$$

Since we are working with the Poisson distribution, which is a discrete distribution, the values for LCL and UCL which will be used to calculate average run lengths and false alarm rates should be integer values. Chakraborti & Human (2008) state that the following values should be used for LCL and UCL in calculations:

$$C = \max\{0, [\text{LCL}]\}$$

and

$$D = \begin{cases} [\text{UCL} - 1] & \text{if UCL is an integer} \\ [\text{UCL}] & \text{if UCL is not an integer} \end{cases}$$

where [UCL] denotes the largest integer not exceeding UCL and [LCL] denotes the largest integer not exceeding LCL.

Using Equations 8.10, 8.11 and 8.12 we can determine the average run length and the false alarm rate. After obtaining the control limits, the conditional false alarm rate (CFAR) and the conditional average run length (CARL) can be calculated. The false alarm rate will be calculated as $1-\beta$, where β is calculated as

$$\beta = P(X_i \leq D) - P(X_i \leq C)$$

and the in-control conditional average run length is calculated as

$$\text{CARL} = \frac{1}{1-\beta}.$$

The unconditional false alarm rate will be determined by

$$\text{UFAR} = \sum_{j=0}^{\infty} \left[(1-\beta) \times e^{-m\lambda} \frac{(m\lambda)^j}{j!} \right] \tag{8.17}$$

and the unconditional average run length will be determined by

$$\text{UARL} = \sum_{j=0}^{\infty}\left[\frac{1}{(1-\beta)}\times e^{-m\lambda}\frac{(m\lambda)^j}{j!}\right]. \tag{8.18}$$

Consider the following simulation study where λ =1, 2, 4, 6, 8, 10, 20, 50 and m =5, 10, 15, 20, 25, 30, 50, 100, 200, 300, 500, 1 000. Chakraborti & Human (2008) considered these values and used the classical (frequentist) method to obtain the unconditional false alarm rates and the unconditional average run lengths. We will consider these values and apply the proposed Bayesian method to obtain the unconditional false alarm rates and the unconditional average run lengths. The results for the given m and λ values, using the classical and the Bayesian methods, are given in Tables 8.2 to 8.4.

Table 8.2: Unconditional average run lengths and unconditional false alarm rates for m = 5, 10, 15, 20 and different values of λ.

	$m=5$				$m=10$			
	Frequentist		**Bayesian**		**Frequentist**		**Bayesian**	
λ	UARL	UFAR	UARL	UFAR	UARL	UFAR	UARL	UFAR
1	2.5045	0.40659	2.6232	0.38329	2.576	0.39043	2.6339	0.3808
2	6.5614	0.15938	6.9533	0.14566	6.8228	0.14847	6.9858	0.14393
4	38.595	0.032368	44.535	0.025926	40.389	0.027249	43.769	0.02463
6	167.59	0.013607	219.49	0.0085607	162.67	0.0096356	192.1	0.0077817
8	434.37	0.010584	792.49	0.0058451	370.43	0.0068362	512.95	0.0048803
10	399.14	0.0094211	1466.1	0.0047268	377.16	0.0063311	673.51	0.0043535
20	302.21	0.0079546	856.78	0.0038292	329	0.0052386	541.61	0.0034856
50	254.62	0.0068141	690.97	0.0031555	294.52	0.0046165	481.9	0.0030082
	$m=15$				$m=20$			
	Frequentist		**Bayesian**		**Frequentist**		**Bayesian**	
λ	UARL	UFAR	UARL	UFAR	UARL	UFAR	UARL	UFAR
1	2.608	0.38428	2.6418	0.3791	2.6185	0.38248	2.6443	0.37857
2	6.8754	0.14634	6.9853	0.14379	6.9361	0.14474	6.9953	0.14341
4	40.988	0.026006	43.523	0.024085	41.349	0.025274	43.218	0.024016
6	158.67	0.0087727	179.59	0.0076308	158.39	0.0081963	174.85	0.0072704
8	328.94	0.005955	425.7	0.0046938	310.69	0.005453	371.96	0.0045611
10	356.4	0.0053731	523.55	0.0040201	351.58	0.0047652	457.62	0.0038822
20	330.94	0.0045562	462.81	0.0033816	340	0.0041014	428.89	0.0033418
50	312.94	0.0039543	433.92	0.0029409	318.06	0.0037464	407.61	0.0029827

A smaller value for the false alarm rate is desired, since the false alarm rate is the probability of a signal being given by the control chart when the process is actually in-control. The Bayesian method

gives for all the values of λ smaller false alarm rates. When $\lambda = 50$ and $m = 5$ it can be seen from Table 8.2 that the false alarm rate (using the Bayesian method) is equal to 0.0031555 is much closer to the nominal value than the false alarm rate attained from using the frequentist method which is equal to 0.0068141. When $\lambda = 50$ and $m = 20$ that the false alarm rate (using the Bayesian method) is equal to 0.0029827 is much closer to the nominal value than the false alarm rate attained from using the frequentist method which is equal to 0.0037464.

Table 8.3: Unconditional average run lengths and unconditional false alarm rates for $m = 25$, 30, 50, 100 and different values of λ.

	$m = 25$				$m = 30$			
	Frequentist		**Bayesian**		**Frequentist**		**Bayesian**	
λ	UARL	UFAR	UARL	UFAR	UARL	UFAR	UARL	UFAR
1	2.6254	0.38127	2.6464	0.37816	2.6287	0.3807	2.6462	0.37813
2	6.9315	0.14473	7.0057	0.14311	6.9658	0.14388	7.0027	0.14308
4	41.759	0.024808	43.335	0.023791	41.752	0.024684	43.233	0.023735
6	155.09	0.0080105	170.87	0.0071288	155.84	0.0077513	168.62	0.0070453
8	302.03	0.0052127	350.97	0.0044467	290.85	0.0050244	333.77	0.0044389
10	339.39	0.0046143	424.55	0.003924	333.01	0.004441	399.47	0.0038335
20	336.52	0.0039022	406.22	0.0032668	334.42	0.0037541	392.9	0.0032327
50	325.11	0.0035391	395.25	0.0029413	332.14	0.00338	391.65	0.0028875
	$m = 50$				$m = 100$			
	Frequentist		**Bayesian**		**Frequentist**		**Bayesian**	
λ	UARL	UFAR	UARL	UFAR	UARL	UFAR	UARL	UFAR
1	2.6336	0.3799	2.6454	0.37819	2.6358	0.37954	2.6439	0.37839
2	6.988	0.14334	7.0282	0.14247	7.0331	0.14234	7.0556	0.14187
4	42.265	0.024105	43.176	0.023563	42.413	0.023885	43.015	0.023545
6	153.99	0.0073598	161.85	0.0069654	154.5	0.006958	158.1	0.006767
8	276.64	0.0046847	296.54	0.0043625	259.83	0.0044108	269.01	0.0042669
10	324	0.0040172	357.45	0.0036986	307.44	0.0038292	318.65	0.0036879
20	335.22	0.0034987	368.61	0.0032003	334.5	0.0033069	351.06	0.0031687
50	338.53	0.0031895	372.43	0.0029041	344.74	0.0030371	361.71	0.0028942

From Table 8.3, we see that the Bayesian method gives for all values of λ smaller false alarm rates than the classical method. The average run lengths are also larger for the Bayesian method. When $\lambda = 1$ and $m = 30$ it can be seen from Table 8.3 that the false alarm rate when using the Bayesian method equals 0.37813, this value is almost the same as the value of 0.3807 from the classical method. Looking at a larger value of λ, say $\lambda = 50$ and $m = 30$, the false alarm rate is 0.0028875 when using the Bayesian method, and when using the classical method the false alarm rate is equal to 0.00338.

The value from the Bayesian method is much closer to the nominal level of 0.0027 than that of the classical method. For this case the UFAR from the Bayesian method is almost 7% higher than the nominal value of 0.0027, whereas the UFAR from the classical method is almost 25% higher than the nominal value.

Table 8.4: Unconditional average run lengths and unconditional false alarm rates for $m = 200$, 300, 500, 1 000 and different values of λ.

	$m = 200$				$m = 300$			
	Frequentist		**Bayesian**		**Frequentist**		**Bayesian**	
λ	UARL	UFAR	UARL	UFAR	UARL	UFAR	UARL	UFAR
1	2.637	0.37937	2.6425	0.37858	2.6364	0.37946	2.6413	0.37875
2	7.0834	0.14127	7.089	0.14116	7.1067	0.14078	7.1102	0.1407
4	42.406	0.023876	42.824	0.023642	42.484	0.023833	42.837	0.023635
6	157.2	0.0066063	158.14	0.0065496	158.87	0.0064507	159.56	0.006406
8	252.13	0.0042008	255.61	0.0041436	248.14	0.0041381	249.75	0.0041097
10	294.93	0.0036522	299.73	0.0035914	289.59	0.0035884	292.14	0.003557
20	332.82	0.0031869	340.52	0.0031251	333.08	0.0031247	337.38	0.0030858
50	349.65	0.0029445	358.07	0.0028745	351.57	0.0029139	357.36	0.0028663
	$m = 500$				$m = 1\,000$			
	Frequentist		**Bayesian**		**Frequentist**		**Bayesian**	
λ	UARL	UFAR	UARL	UFAR	UARL	UFAR	UARL	UFAR
1	2.6369	0.37938	2.6408	0.37883	2.6377	0.37927	2.6402	0.37891
2	7.1298	0.14029	7.1314	0.14026	7.1473	0.13992	7.1477	0.13991
4	42.559	0.023791	42.789	0.023662	42.53	0.023807	42.709	0.023707
6	161.72	0.0062466	161.93	0.0062325	163.6	0.0061171	163.61	0.0061161
8	246.69	0.0040728	246.95	0.004069	246.69	0.0040539	246.69	0.0040539
10	286.42	0.0035221	286.81	0.0035167	285.82	0.0034996	285.82	0.0034996
20	334.18	0.0030545	335.71	0.0030381	337.94	0.0029702	338.13	0.0029677
50	356.42	0.002864	360.25	0.0028328	367.59	0.0027677	369.54	0.0027515

From Table 8.4, we see that the Bayesian method gives for all values of λ smaller false alarm rates than the classical method. The average run lengths are also larger for the Bayesian method. When $\lambda = 8$ and $m = 500$ it can be seen from Table 8.4 that the false alarm rate using the Bayesian method equals 0.004069, this value is almost the same as the value of 0.0040728 from the classical method. Looking at a larger value of $\lambda = 50$ and $m = 1\,000$, the false alarm rate is 0.0027515 when using the Bayesian method, and when using the classical method the false alarm rate is equal to 0.0027677. These two values are almost the same, and nearly equal to the nominal value.

From this simulation study we conclude that the Bayesian method give larger average run lengths and smaller false alarm rates than the classical method. This is true for smaller values of m. For $m = 500$ and $m = 1\,000$ the two methods yield almost the same results. According to Chakraborti & Human (2008) it appears that, for the classical method, one needs at least 300 to 500 inspection units to estimate the unknown standard to ensure that the c - chart performs as well when the standard is in fact known. For the Bayesian method the false alarm rates are smaller and the average run lengths are larger than for the classical method.

8.4.2 Example

Consider the following example from Montgomery (1996), Example 6-3 on page 277. Chakraborti & Human (2008) also considered this example. This example deals with the number of nonconformities observed in 26 successive samples of 100 printed circuit boards. The inspection unit is defined as 100 boards. The 26 samples contained 516 nonconformities, and λ is estimated by

$$\overline{\lambda} = \frac{516}{26} = 19.85.$$

It was found that units 6 and 20 were out-of-control, and they were eliminated after further investigation. Revised limits were calculated using the remaining samples, with $m = 24$ and $\sum_{i=1}^{m} x_i = 472$. The average number of nonconformities per inspection unit was recalculated as

$$\overline{\lambda} = \frac{472}{24} = 19.67.$$

Considering this example, we have $m = 24$ and $\sum_{i=1}^{m} x_i = 472$. As mentioned, we will use the Jeffreys prior for λ, $\pi_J(\lambda) \propto \lambda^{-\frac{1}{2}}$. This will result in a gamma posterior, $Gamma\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, m\right)$. For this example the posterior distribution of λ will be a $Gamma(472.5, 24)$. Figure 8.1 shows the posterior distribution of λ. Figure 8.2 shows a bar graph of the predictive density function, $f\left(x_f \,|data\right)$, where x_f is the number of nonconformities in a future inspection unit. The predictive density will be used to obtain the control chart limits.

Summary statistics for the posterior distribution of λ:

mean = 19.69 standard deviation = 0.9057 median = 19.67

95% credibility interval = (17.95; 21.50).

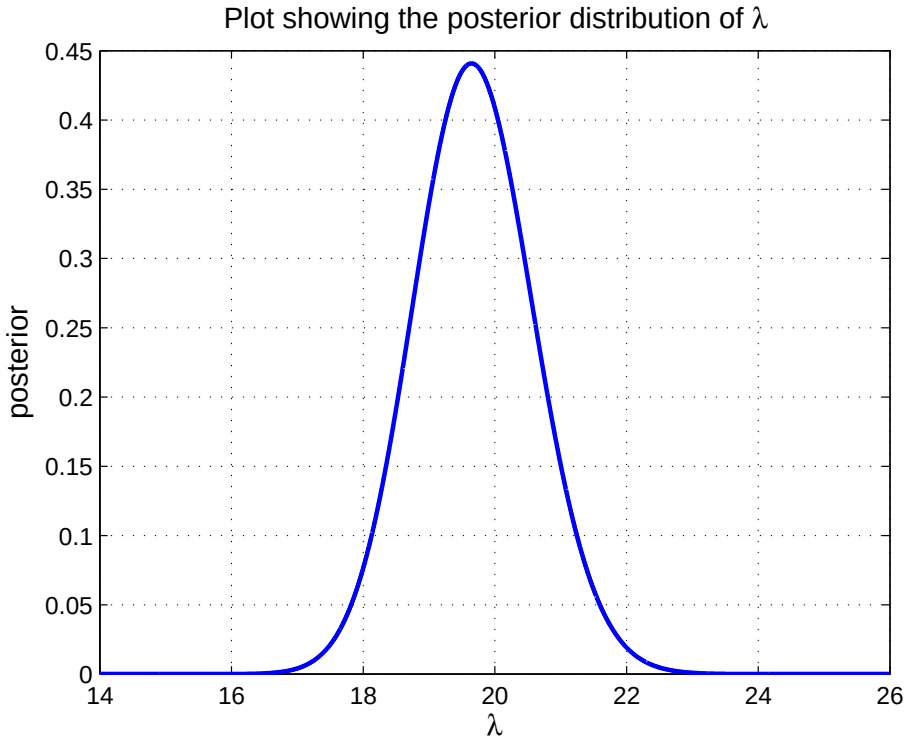


Figure 8.1: Posterior distribution of λ, when $m = 24$ and $\sum_{i=1}^{m} x_i = 472$.

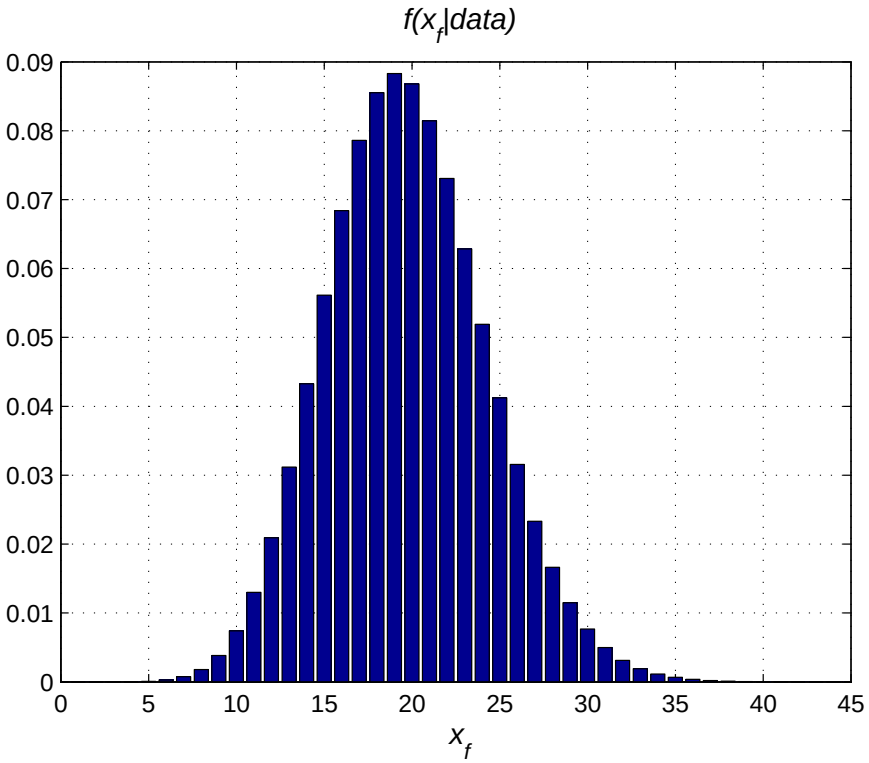


Figure 8.2: Bar graph of the predictive density of $f(x_f|data)$.

For the observed value, $\sum_{i=1}^{m} = 472$, the chart's performance will be investigated by looking at the conditional average run length (CARL) and the conditional false alarm rate (CFAR). The control chart limits, conditional average run length and conditional average false alarm rate will be calculated using the classical (frequentist) method and using a Bayesian method. Chakraborti & Human (2008) determined these values using the classical method, the results are given in Table 8.5. We used a Bayesian method to determine these values, the results are also given in Table 8.5. Chakraborti & Human (2008) determined the lower and upper control limits by using the limits given in Equation 8.13, we used the limits from Equation 8.16 to determine the lower and upper control limits. From Chakraborti & Human (2008), we use $\lambda = 20$ to calculate CARL and CFAR.

Calculations used to obtain the frequentist control limits:

$$\begin{aligned}
\text{UCL} &= \overline{\lambda} + 3\sqrt{\overline{\lambda}} \\
&= \frac{472}{24} + 3\sqrt{\frac{472}{24}} \\
&= 32.9708 \\
\text{LCL} &= \overline{\lambda} - 3\sqrt{\overline{\lambda}} \\
&= \frac{472}{24} - 3\sqrt{\frac{472}{24}} \\
&= 6.3625
\end{aligned}$$

$$\begin{aligned}
C &= \max\{0, [6.3625]\} \\
&= \max\{0, 6\} = 6 \\
D &= [32.9708] = 32.
\end{aligned}$$

Calculations used to obtain the Bayesian control limits:

$$\begin{aligned}
\text{UCL} &= \frac{\sum_{i=1}^{m} x_i + \frac{1}{2}}{m} + 3\sqrt{\frac{\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)(m+1)}{m^2}} \\
&= \frac{472 + \frac{1}{2}}{24} + 3\sqrt{\frac{\left(472 + \frac{1}{2}\right)(24+1)}{24^2}} \\
&= 33.2632
\end{aligned}$$

$$\begin{aligned}
\text{LCL} &= \frac{\sum_{i=1}^{m} x_i + \frac{1}{2}}{m} - 3\sqrt{\frac{\left(\sum_{i=1}^{m} x_i + \frac{1}{2}\right)(m+1)}{m^2}} \\
&= \frac{472 + \frac{1}{2}}{24} - 3\sqrt{\frac{\left(472 + \frac{1}{2}\right)(24+1)}{24^2}} \\
&= 6.1018
\end{aligned}$$

$$\begin{aligned}
C &= \max\{0, [6.1018]\} \\
&= \max\{0, 6\} = 6 \\
D &= [33.2632] = 33.
\end{aligned}$$

Calculations used to obtain the average run lengths and false alarm rates:

$$\begin{aligned}
\text{CFAR} &= 1 - [P(X_i \leq D|\lambda = 20) - P(X_i \leq C|\lambda = 20)] \\
\text{CARL} &= \frac{1}{\text{CFAR}}.
\end{aligned}$$

Table 8.5: Lower control limits, upper control limits, conditional average run lengths (CARL) and conditional false alarm rates (CFAR) for $m = 24$ and $\sum_{i=1}^{m} = 472$.

Frequentist					
LCL	UCL	C	D	CARL	CFAR
6.3625	32.9708	6	32	200.7	0.0049825
Bayesian					
LCL	UCL	C	D	CARL	CFAR
6.1018	33.2632	6	33	339.72	0.0029436

From Table 8.5, we see that the Bayesian method gives a wider interval than the classical method, this results in a larger value for CARL and a smaller value for CFAR. The conditional false alarm rate (CFAR) when the Bayesian method is used, is equal to 0.0029436. This value is much closer to the nominal value of 0.0027, than the conditional false alarm rate (CFAR) of 0.0049825 obtained from the classical method.

In Section 8.4 the mean and the variance of the predictive density was used to construct the 3-sigma limits. This method gave smaller values for the false alarm rate and larger values for the average run length. The problem for small values of λ can be solved if the rejection region of size α is defined as the equal-tailed intervals of the predictive density where $\alpha = 0.0027$. This will be investigated in future research.

8.5 Conclusion

The usual operation of the c - chart was extended by introducing a Bayesian approach for the c - chart. From an extended simulation study, using different values of λ and number of inspection units, we conclude that the suggested Bayesian approach gives larger values for the average run length and smaller values for the false alarm rate. A smaller value for the false alarm rate is desired, since the false alarm rate is the probability of a signal being given by the control chart when the process is actually in-control. The false alarm rates obtained from the Bayesian method were generally closer to the expected nominal value for the false alarm rate, 0.0027. The Bayesian method generally had wider control limits.

Bayarri & García-Donato (2005) give the following reasons for recommending a Bayesian analysis:

- Control charts are based on future observations, and Bayesian methods are very natural for prediction;
- uncertainty in the estimation of the unknown parameters is adequately handled;
- implementation with complicated models and in a sequential scenario poses no methodological difficulty, the numerical difficulties are easily handled via Monte Carlo methods;
- objective Bayesian analysis is possible without introduction of external information other than the model, but any kind of prior information can be incorporated into the analysis, if desired.

Chapter 9

Tolerance Intervals

9.1 Introduction

In this chapter we will introduce Bayesian tolerance intervals for the binomial and Poisson distributions. Tolerance intervals could be of interest in quality control. Wang & Tsung (2009) state the following: "The construction of tolerance intervals to measure discrete quality characteristics has been one of the major tasks in developing quality control systems used in manufacturing and pharmaceutical sectors". A tolerance interval gives information about a certain proportion or more of the population, with a given confidence level. This proportion is also referred to as the content of a tolerance interval. Whereas a confidence interval gives information about an unknown parameter. The goal of a tolerance interval is to contain at least a specified proportion of the population with a specified degree of confidence. In this chapter π will indicate the content, and $1-\alpha$ the confidence level. There are three main types of tolerance intervals. Van der Merwe & Hugo (2007) state the following three types:

- the $(\pi, 1-\alpha)$ tolerance interval, where π is the content and $1-\alpha$ the confidence level;
- the π-expectation tolerance interval, where π is the expected coverage of the interval. The π - expectation intervals focus on the prediction of one or a few future observations from the process;
- the fixed-in-advance tolerance interval, where the interval is constant and one wishes to estimate the proportion of process measurements it contains.

The above mentioned intervals can be one-sided or two-sided tolerance intervals. The one-sided interval can take the form $(-\infty, U)$ or (L, ∞), where U is called a one-sided upper tolerance limit and L a one-sided lower tolerance limit. The two-sided interval can take on two different types. From Krishnamoorthy & Mathew (2009) the one is constructed such that it would contain at least a proportion π of the population with confidence level $1-\alpha$, and the other type is constructed such that it would contain at least a proportion π of the centre of the population with confidence level $1-\alpha$. The latter is referred to as an equal-tailed tolerance interval. Our interest is to investigate Bayesian tolerance

limits and intervals for the binomial and Poisson distributions. The Jeffreys prior will be used for the Bayesian method.

9.2 Tolerance Intervals for the Binomial Distribution

Suppose we know that the distribution is a binomial distribution, but the value of the p is unknown. In the case of the binomial distribution, when using the Jeffreys prior from Equation 7.2, the posterior distribution will be

$$\pi_J(p|data) \propto p^{x-\frac{1}{2}}(1-p)^{n-x-\frac{1}{2}}. \tag{9.1}$$

The posterior distribution given in Equation 9.1 can now be used to obtain the required tolerance interval. In the simulation study we will consider confidence intervals for P_{95}, the 95^{th} percentile. A confidence interval for a percentile is called a tolerance interval.

9.2.1 Simulation Study - Tolerance Intervals for the Binomial Distribution

In this simulation study we consider tolerance intervals where the content is 0.95 and the level of confidence is 0.95. The binomial distribution is considered for $n = 10$, 20 and 30. The results for the two-sided and one-sided coverage are given in Tables 9.1 to 9.3.

Table 9.1: Interval estimation of the 95^{th} percentile of the binomial distribution for $n = 10$.

p	P_{95}	$P[X \leq P_{95}]$	**Two-sided coverage**	**One-sided coverage**
0.1	3	0.9872	0.9872	1.0000
0.2	4	0.9672	0.9672	1.0000
0.3	5	0.9527	0.9894	0.9718
0.4	7	0.9877	0.9413	0.9536
0.5	8	0.9893	0.9785	0.9453
0.6	8	0.9536	0.9520	0.9877
0.7	9	0.9718	0.9612	0.9894
0.8	10	1.0000	0.9936	0.9936
0.9	10	1.0000	0.9999	0.9999

Table 9.2: Interval estimation of the 95^{th} percentile of the binomial distribution for $n = 20$.

p	P_{95}	$P[X \leq P_{95}]$	**Two-sided coverage**	**One-sided coverage**
0.1	4	0.9568	0.9568	1.0000
0.2	7	0.9679	0.9785	0.9885
0.3	9	0.9520	0.9752	0.9645
0.4	12	0.9790	0.9776	0.9490
0.5	14	0.9793	0.9734	0.9423
0.6	16	0.9840	0.9754	0.9435
0.7	17	0.9645	0.9872	0.9829
0.8	19	0.9885	0.9785	0.9679
0.9	20	1.0000	0.9976	0.9887

Table 9.3: Interval estimation of the 95^{th} percentile of the binomial distribution for $n = 30$.

p	P_{95}	$P[X \leq P_{95}]$	**Two-sided coverage**	**One-sided coverage**
0.1	6	0.9742	0.9498	0.9576
0.2	10	0.9744	0.9800	0.9558
0.3	13	0.9599	0.9737	0.9698
0.4	16	0.9519	0.9616	0.9828
0.5	19	0.9506	0.9706	0.9786
0.6	22	0.9565	0.9745	0.9788
0.7	25	0.9698	0.9737	0.9599
0.8	27	0.9558	0.9864	0.9905
0.9	29	0.9576	0.9556	0.9922

When $n = 10$, we see from Table 9.1 that the two-sided coverage is above 0.95 for all values of p, except when $p = 0.4$. The one-sided coverage is also above 0.95 for all values of p, except when $p = 0.5$. When $n = 20$, we see from Table 9.2 that the two-sided coverage is above 0.95 for all values of p. The one-sided coverage is above 0.95 for most values of p, except when $p = 0.4$, 0.5 and 0.6. When $n = 30$, we see from Table 9.3 that the two-sided coverage is above 0.95 for all values of p, except when $p = 0.1$. The one-sided coverage is also above 0.95 for all values of p.

9.3 Tolerance Intervals for the Poisson Distribution

Suppose we know that the distribution is a Poisson distribution, but we do not know what the value of λ is. In the case of the Poisson distribution, when using the Jeffreys prior from Equation 8.2, the posterior distribution will be

$$\pi_J(\lambda \,|data) \quad \propto \quad e^{-\lambda}\lambda^{x-\frac{1}{2}} \qquad\qquad \lambda > 0. \tag{9.2}$$

The predictive density for a future observation, x_f, will then be

$$\begin{aligned} f\left(x_f\,|data\right) &= \int_0^\infty f\left(x_f\,|\lambda\right)\pi\left(\lambda\,|x\right)d\lambda \\ &= \int_0^\infty e^{-\lambda}\frac{\lambda^{x_f}}{x_f!}\frac{1}{\Gamma\left(x+\frac{1}{2}\right)}\lambda^{x-\frac{1}{2}}e^{-\lambda}d\lambda \\ &= \frac{\Gamma\left(x_f+x+\frac{1}{2}\right)}{\Gamma\left(x+\frac{1}{2}\right)x_f!}\frac{1}{2^{x_f+x+\frac{1}{2}}} \qquad x_f=0,1,\ldots,\infty \end{aligned} \tag{9.3}$$

which is a Poisson-gamma distribution, similar to the expression obtained in Equation 8.8. This predictive density can be used to construct the second type of tolerance interval mentioned in Section 9.1, the $\pi-$expectation tolerance interval. As mentioned, the $\pi-$expectation tolerance intervals focus on the prediction of one or a few future observations from the process. The posterior distribution given in Equation 9.2 can now be used to obtain the required tolerance interval.

9.3.1 Simulation Study - Tolerance Intervals for the Poisson Distribution

Consider the following study for the interval estimation of the 95^{th} percentile of the Poisson distribution. For given x, $x = 0,1,\ldots 50$, 10 000 values are simulated from the posterior distribution of λ. For each value of λ, the corresponding 95^{th} percentile (P_{95}) of the Poisson distribution is found. From the sorted 10 000 values of P_{95} the lower limit (2.5%) and upper limit (97.5%) is found in the case of the two-sided interval. For given λ, $\lambda = 1,\ldots,15$, the probabilities for all values of x which yielded an interval which contains the true λ are added to obtain the coverage probability. The results are given in Table 9.4 and plotted in Figures 9.1 and 9.2.

Table 9.4: Interval estimation of the 95^{th} percentile of the Poisson distribution.

λ	P_{95}	$P(X \leq P_{95})$	**Two-sided coverage**	**One-sided coverage**
1.0	3	0.9810	0.9963	1.0000
1.5	4	0.9814	0.9955	1.0000
2.0	5	0.9834	0.9955	1.0000
2.5	5	0.9580	0.9858	1.0000
3.0	6	0.9665	0.9383	0.9502
3.5	7	0.9733	0.9599	0.9698
4.0	8	0.9786	0.9736	0.9084
4.5	8	0.9597	0.9718	0.9389
5.0	9	0.9682	0.9796	0.9596
5.5	10	0.9747	0.9624	0.9734
6.0	10	0.9574	0.9626	0.9826
6.5	11	0.9661	0.9727	0.9570
7.0	12	0.9730	0.9576	0.9704
7.5	12	0.9573	0.9582	0.9797
8.0	13	0.9658	0.9690	0.9576
8.5	14	0.9726	0.9561	0.9256
9.0	14	0.9585	0.9567	0.9450
9.5	15	0.9665	0.9674	0.9597
10.0	15	0.9513	0.9626	0.9707
10.5	16	0.9604	0.9570	0.9496
11.0	17	0.9678	0.9672	0.9625
11.5	17	0.9542	0.9631	0.9723
12.0	18	0.9626	0.9584	0.9542
12.5	19	0.9694	0.9481	0.9654
13.0	19	0.9573	0.9491	0.9741
13.5	20	0.9649	0.9693	0.9585
14.0	20	0.9521	0.9691	0.9684
14.5	21	0.9604	0.9524	0.9516
15.0	21	0.9673	0.9625	0.9626

From Table 9.4 and Figures 9.1 and 9.2 it can be seen that the coverage rates are most of the time at or above the nominal value of 0.95. When $\lambda = 3$, 12.5 and 13 the coverage rates are below 0.95 for the two-sided case, and when $\lambda = 4$, 4.5, 8.5, 9 and 10.5 the coverage rates are below 0.95 for the one-sided case.

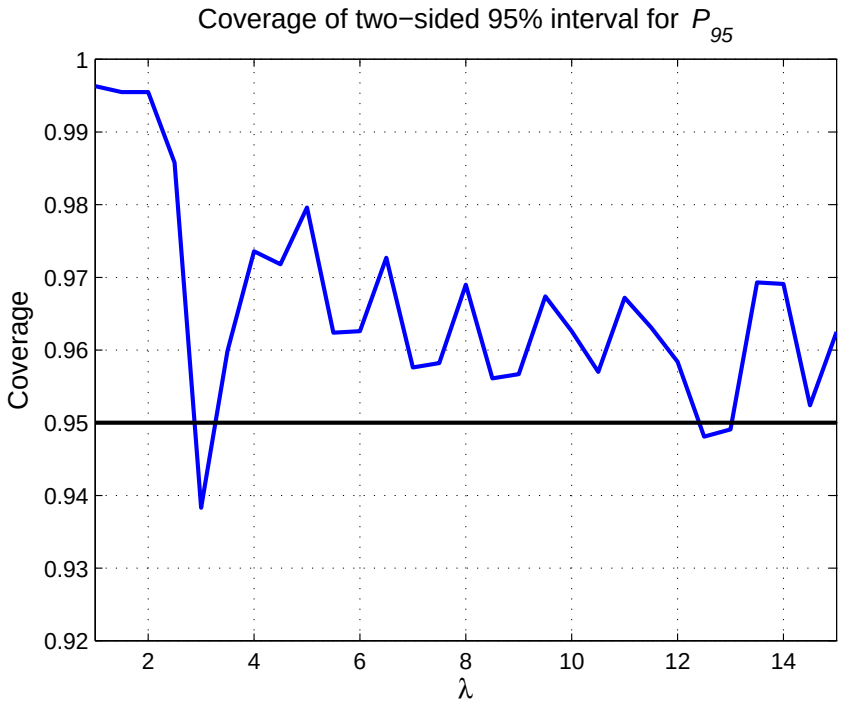


Figure 9.1: Coverage of two-sided 95% interval for P_{95}.

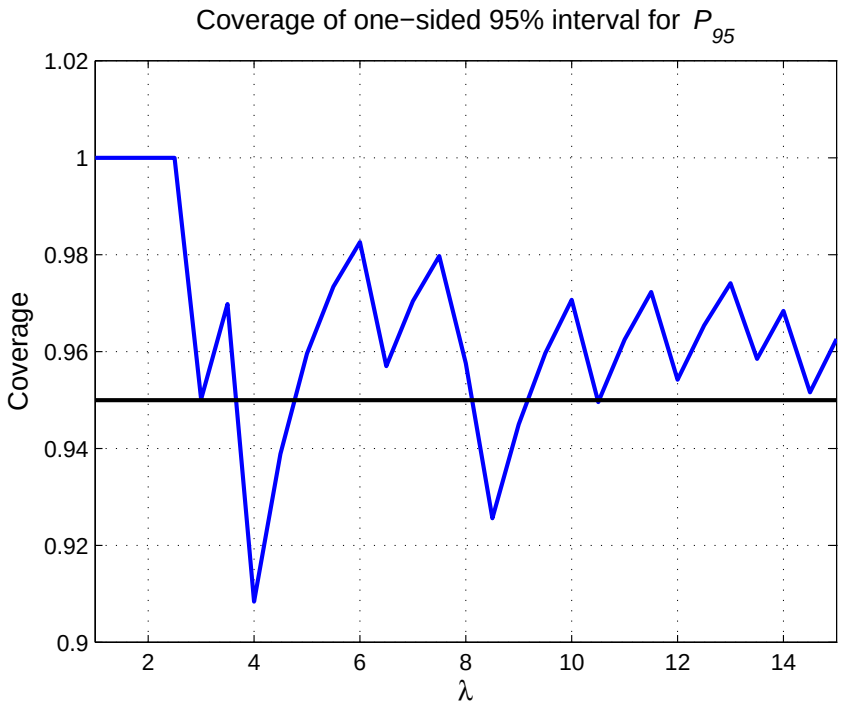


Figure 9.2: Coverage of one-sided 95% interval for P_{95}.

9.3.2 Example - Tolerance Intervals for the Poisson Distribution

Consider the example from Section 8.4.2. Recall that for this example, we had $m = 24$ and $\sum_{i=1}^{m} x_i = 472$. We construct the $(\pi, 1-\alpha)$ tolerance interval for this example, where $\pi = 0.95$ and $1-\alpha = 0.95$. The results gave the lower tolerance limit equal to 25, the upper tolerance limit equal to 29 and the coverage equal to 0.9772. Figure 9.3 shows the histogram of P_{95}.

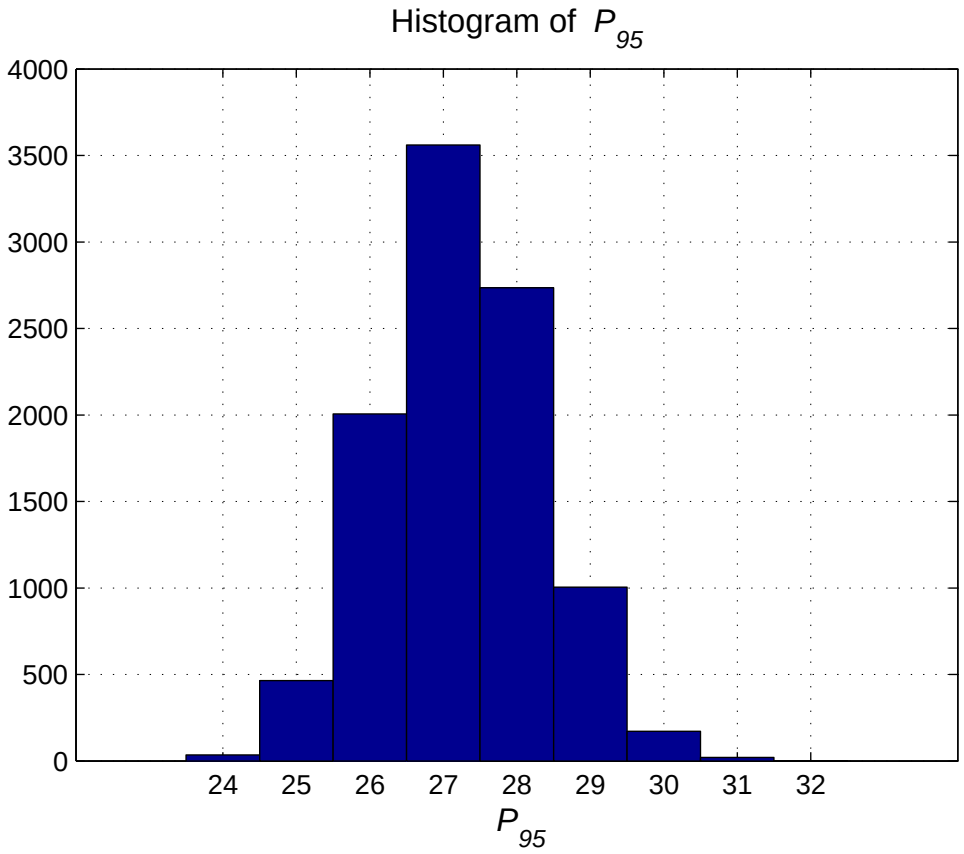


Figure 9.3: Histogram of P_{95}.

9.4 Conclusion

Bayesian tolerance intervals for the binomial and Poisson distributions were introduced, and from the simulation studies it was seen that the coverage rates obtained for one-sided and two-sided intervals were relatively good. For the two-sided and one-sided intervals the coverage rates were most of the time at or above 0.95, except in a few cases. We introduced tolerance intervals here, to show that tolerance intervals could be useful for applications in quality control.

Chapter 10

Conclusion

This chapter concludes with a summary of the conclusions of the chapters in the thesis, possible shortcomings and/or drawbacks of this thesis and possible future research in this area.

10.1 Summary of Conclusions

This thesis focused on the objective part of Bayesian statistics, by looking at a number of noninformative priors. A noninformative prior is used when no prior information is available. If no prior information is available it is impossible to justify the choice of a prior distribution on a subjective basis, in such a case a noninformative prior should be used.

In Chapter 2 the probability matching prior for the product of different powers of k binomial parameters, $\psi = \prod_{i=1}^{k} p_i^{a_i}$, was derived. It was also shown that when the probability matching prior is used, the resulting posterior distribution is proper. The performance of the probability matching prior was evaluated. The probability matching prior was compared to the Jeffreys and uniform priors. For simulation purposes it is much easier to simulate from the Jeffreys posterior or the uniform posterior, since the posterior distributions are known distributions for these two posteriors. This is not the case for the probability matching prior, we therefore introduced a weighted Monte Carlo method for simulation from the probability matching posterior. An example on the probability of failure of independent parallel components system was considered. The probability matching and Jeffreys priors were compared to other methods, where the probability matching prior compared well with the other results, but the Jeffreys prior under estimated the upper confidence limit. Three simulation studies were considered. In the first simulation study we compared four priors for the product of two binomial proportions. Two of the priors were probability matching priors, one of the two was a Poisson approximation to the binomial distribution, the other two priors were Jeffreys priors, where one of the two was also a Poisson approximation to the binomial distribution. The probability matching prior for the binomial distribution performed the best of the four priors. In the second simulation study a comparison was made between the Jeffreys, uniform and probability matching priors for the product of two binomial

proportions. Different values of n_1, n_2, p_1 and p_2 were considered. The probability matching prior and the uniform prior gave similar results. In general, we can conclude that the coverage probabilities for the Jeffreys prior is below the nominal level. The uniform and probability matching priors, on the other hand, give coverage probabilities larger than the nominal level and more so for the uniform prior. The average interval lengths and standard deviations are smaller for larger values of n. In the last simulation study we compared the performance of the Jeffreys, uniform and probability matching priors where one needs two types of components and three components in parallel are needed for a system to operate. Again the Jeffreys prior gave coverage rates below the nominal level, and the uniform and probability matching priors gave coverage rates above the nominal level.

The probability matching prior for a linear combination of binomial proportions, $\theta = \sum_{i=1}^{k} a_i p_i$, was derived in Chapter 3 and the properness of this posterior was shown. The weighted Monte Carlo method was again used for simulation from the probability matching posterior. In the example that was considered, the probability matching prior compared well with the other results. In the first simulation study a comparison was made between six other intervals and the Jeffreys and probability matching priors. In this study the results from the two Bayesian procedures compared well with the others, where the difference between the intervals were not too large except for the Wald interval. In the second simulation study a comparison was made between the Jeffreys, uniform and probability matching priors for the difference between two binomial proportions. Different values of n_1, n_2, p_1 and p_2 were considered. The probability matching prior and the uniform prior gave similar results. In general, we can conclude that the coverage probabilities for the Jeffreys prior is below the nominal level. Limited simulation studies have shown that the probability matching prior achieves its sample frequentist coverage results somewhat better than in the case of the Jeffreys prior.

In Chapter 4 the probability matching prior was derived for $\xi = \prod_{i=1}^{k} \lambda_i^{a_i}$, the product of different powers of k Poisson rates. This has been derived before by Kim (2006), but Kim used the method by Tibshirani (1989) where we used the method by Datta & Ghosh (1995). We derived the reference prior for the ratio of two Poisson rates. We considered a number of simulation studies in this chapter. If one is interested in obtaining point estimates and Bayesian credible intervals for the product of different powers of Poisson rates the probability matching prior is the best. It was shown in this chapter that the Jeffreys, probability matching and reference priors for the ratio of two Poisson rates have the same posterior distribution. From the simulation study we came to the conclusion that if one wants to obtain point estimates, credibility intervals or do hypothesis testing about the ratio of two Poisson rates, then the uniform prior should be used.

The probability matching prior for a linear contrast of Poisson parameters, $\delta = \sum_{i=1}^{k} a_i \lambda_i$, (i.e. $\sum_{i=1}^{k} a_i = 0$) was derived in Chapter 5 and it was also shown when $\sum_{i=1}^{k} a_i \neq 0$ what the probability matching prior should be. In a simulation study we compared the four approximate confidence intervals for linear contrasts of Poisson rates proposed by Stamey & Hamilton (2006) to confidence intervals using Bayesian procedures. The probability matching prior also performed satisfactory, where the

Wald interval was the poorest performer. We also addressed the problem of hypothesis testing about two Poisson means, by looking at the size and power of different tests. We compared four Bayesian procedures to two procedures used by Krishnamoorthy & Thomson (2004). We used the Jeffreys prior, the probability matching prior, a third prior which is proportional to $\lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and a fourth prior which is proportional to $\lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$and compared them to their results. The Bayesian procedures compared well with the E - test. The C - test performed the poorest of the six tests.

In Chapter 6 we proposed a Bayesian method for the estimation of binomial proportions from pooled samples, where the Jeffreys prior was used. Simulation studies were considered as well as an example. The Bayesian method compared well with the other results, and gave much better results than the Wald and minimum infection rate intervals. The Wald and the minimum infection rate intervals performed the poorest. We considered the Bayes factor, but we found the results from the Bayes factor a bit unsatisfactory.

We considered Bayesian process control for the p - chart and the c - chart in Chapters 7 and 8, respectively. The Jeffreys prior was used for the proposed Bayesian method and we derived the predictive density for both cases. The predictive density was used to obtained the control chart limits for the Bayesian method. In both these chapters we compared the results from the Bayesian procedure to the classical (frequentist) method. We compared the average run lengths and the false alarm rates. The Bayesian method gave smaller false alarm rates and larger average run lengths than the classical method. Smaller values for the false alarm rate are desired, and large values for the average run length.

In Chapter 9, Bayesian tolerance intervals for the binomial and Poisson distributions were introduced. From the simulation studies it was seen that the coverage rates obtained for one-sided and two-sided intervals were relatively good. Bayesian tolerance intervals were introduced to show that tolerance intervals could be useful for applications in quality control.

10.2 Shortcomings of the Thesis

One of the shortcomings is the limitation of the simulation studies. For the binomial case, we only considered certain values for n_i, mainly small samples. The largest value considered was $n_1 = n_2 = 30$.

Another shortcoming of this thesis is the applied side of the work done. Because of the theoretical nature of the thesis, simulation studies were used to compare the results, there was no room to apply the findings to an actual data set. It would have been interesting to apply some of the results to an actual real life problem.

In Chapters 6, 7, 8 and 9, the Jeffreys prior was the only prior used. Would other priors give similar results? The practical aspect of the work done was on quality control. Would a subjective approach not have been more interesting? This can only happen if one is actually involved with an actual case study in process control.

10.3 Possible Future Research

Possible further studies would be to investigate the reference prior in more detail. The reference prior was only derived and evaluated for the ratio for two Poisson rates. A possibility is to derive the reference prior for the product of different powers of k binomial rates and Poisson rates. The reference prior has been derived for the product of two binomial rates, and is given in Yang & Berger (1997). The performance of the reference prior will be compared to the other priors used in the thesis for the different cases in future research. Another area is to extend the Bayesian process control, and to investigate other priors. The Bayes factor could also be used in the studies considered.

References

Agresti, A. & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54(4), 280 – 288.

Agresti, A. & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119 – 126.

Barker, L. (2002). A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is ≤ 5. *The American Statistician*, 56(2), 85 – 89.

Bayarri, M. J. & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 19(1), 58 – 80.

Bayarri, M. J. & García-Donato, G. (2005). A Bayesian sequential look at u-control charts. *Technometrics*, 47(2), 142 – 151.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53, 370 – 418.

Beal, S. L. (1987). Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics*, 43(4), 941 – 950.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 2nd edition.

Berger, J. O. & Bernardo, J. M. (1992). On the development of reference priors. In *Bayesian Statistics 4*, Bernardo, J. M., Berger, J. O., Dawid, A. P. & Smith, A. F. M. (eds). Oxford University Press, Oxford, 35 - 60.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 41(2), 113 – 147.

Biggerstaff, B. J. (2008). Confidence intervals for the difference of two proportions estimated from pooled samples. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(4), 478 – 496.

Calabrese, J. M. (1995). Bayesian process control for attributes. *Management Science*, 41(4), 637 – 645.

Chakraborti, S. & Human, S. W. (2006). Parameter estimation and performance of the p - chart for attributes data. *IEEE Transactions on Reliability*, 55(3), 559 – 566.

Chakraborti, S. & Human, S. W. (2008). Properties and performance of the c - chart for attributes data. *Journal of Applied Statistics*, 35(1), 89 – 100.

Chen, M.-H. & Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1), 69 – 92.

Datta, G. S. & Ghosh, J. K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*, 82(1), 37 – 45.

Ghosh, M. (2011). Objective priors: an introduction for frequentists. *Statistical Science*, 26(2), 187 – 202.

Gill, J. (2008). *Bayesian Methods: A Social and Behavioral Sciences Approach*. Chapman & Hall/CRC, London, 2nd edition.

Gu, K., Ng, H. K. T., Tang, M. L., & Schucany, W. R. (2008). Testing the ratio of two Poisson rates. *Biometrical Journal*, 50(2), 283 – 298.

Guenther, W. C. (1965). *Concepts of Statistical Inference*. McGraw-Hill, New York, 1st edition.

Guttman, I. & Menzefricke, U. (2003). Posterior distributions for functions of variance components. *Test*, 12(1), 115 – 123.

Hald, A. (1952). *Statistical Theory with Engineering Applications*. John Wiley & Sons, New York.

Harris, B. (1971). Hypothesis testing and confidence intervals for products and quotients of Poisson parameters with applications to reliability. *Journal of the American Statistical Association*, 66(335), 609 – 613.

Hepworth, G. (1996). Exact confidence intervals for proportions estimated by group testing. *Biometrics*, 52(3), 1134 – 1146.

Hepworth, G. (2005). Confidence intervals for proportions estimated by group testing with groups of unequal size. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(4), 478 – 497.

Jeffreys, H. (1939). *Theory of Probability*. Oxford University Press, Oxford, 1st edition.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773 – 795.

Kim, H.-J. (2006). On Bayesian estimation of the product of Poisson rates with application to reliability. *Communications in Statistics - Simulation and Computation*, 35(1), 47 – 59.

Krishnamoorthy, K. & Mathew, T. (2009). *Statistical Tolerance Regions: theory, applications, and computation*. John Wiley & Sons, New Jersey.

Krishnamoorthy, K. & Thomson, J. (2004). A more powerful test for comparing two Poisson means. *Journal of Statistical Planning and Inference*, 119(1), 23 – 35.

Li, K.-H. (2007). Pool size selection for the sampling/importance resampling algorithm. *Statistica Sinica*, 17(3), 895 – 907.

Madansky, A. (1965). Approximate confidence limits for the reliability of series and parallel systems. *Technometrics*, 7(4), 495 – 503.

Menzefricke, U. (2002). On the evaluation of control chart limits based on predictive distributions. *Communications in Statistics - Theory and Methods*, 31(8), 1423 – 1440.

Montgomery, D. C. (1996). *Introduction to Statistical Quality Control*. John Wiley & Sons, New York, 3rd edition.

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine*, 17(8), 873 – 890.

Ornaghi, J. A., March, G. J., Boito, G. T., Marinelli, A., Beviacqua, J. E., Giuggia, J., & Lenardon, S. L. (1999). Infectivity in natural populations of *Delphacodes kuscheli* vector of "Mal Rio Cuarto" virus. *Maydica*, 44, 219 – 223.

Pearn, W. L. & Wu, C.-W. (2005). A Bayesian approach for assessing process precision based on multiple samples. *European Journal of Operational Research*, 165(3), 685 – 695.

Price, R. M. & Bonett, D. G. (2000). Estimating the ratio of two Poisson rates. *Computational Statistics & Data Analysis*, 34(3), 345 – 356.

Price, R. M. & Bonett, D. G. (2004). An improved confidence interval for a linear function of binomial proportions. *Computational Statistics & Data Analysis*, 45(3), 449 – 456.

Przyborowski, J. & Wilenski, H. (1940). Homogeneity of results in testing samples from Poisson series. *Biometrika*, 31(3/4), 313 – 323.

Raubenheimer, L. & Van der Merwe, A. J. (2011a). Bayesian estimation of functions of binomial rates. *South African Statistical Journal*, 45(1), 41 – 64.

Raubenheimer, L. & Van der Merwe, A. J. (2011b). Bayesian estimation of the ratio and product of two Poisson rates. *Proceedings of the 53rd annual conference of the South African Statistical Association*, 88 - 99.

Raubenheimer, L. & Van der Merwe, A. J. (Accepted). Bayesian inference on nonlinear functions of Poisson rates. *South African Statistical Journal*, to appear in 2012.

Raubenheimer, L. & Van der Merwe, A. J. (Submitted). Bayesian estimation of linear functions of Poisson parameters. *Communications in Statistics - Theory and Methods*, submitted in 2011.

Rice, J. (1995). *Mathematical Statistics and Data Analysis*. Duxbury Press, California, 2nd edition.

Robert, C. P. (2001). *The Bayesian Choice*. Springer, New York, 2nd edition.

Roths, S. A. & Tebbs, J. M. (2006). Revisiting Beal's confidence intervals for the difference of two binomial proportions. *Communications in Statistics - Theory and Methods*, 35(9), 1593 – 1609.

Rubin, D. B. (1984). Bayesian justifiable and relevant frequency calculations for the applied Statistician. *The Annals of Statistics*, 12(4), 1151 – 1172.

Severini, T. A., Mukerjee, R., & Ghosh, M. (2002). On an exact probability matching property of right-invariant priors. *Biometrika*, 89(4), 952 – 957.

Skare, Ø., Bølviken, E., & Holden, L. (2003). Improved sampling-importance resampling and reduced bias importance sampling. *Scandinavian Journal of Statistics*, 30(4), 719 – 737.

Smith, A. F. M. (1995). A conversation with Dennis Lindley. *Statistical Science*, 10(3), 305 – 319.

Smith, A. F. M. & Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2), 84 – 88.

Stamey, J. & Hamilton, C. (2006). A note on confidence intervals for a linear function of Poisson rates. *Communications in Statistics - Simulation and Computation*, 35(4), 849 – 856.

Tebbs, J. M. & Roths, S. A. (2008). New large-sample confidence intervals for a linear combination of binomial proportions. *Journal of Statistical Planning and Inference*, 138(6), 1884 – 1893.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*, 76(3), 604 – 608.

Tu, X. M., Litvak, E., & Pagano, M. (1995). On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: Application to HIV screening. *Biometrika*, 82(2), 287 – 297.

Van der Merwe, A. J. & Hugo, J. (2007). Bayesian tolerance intervals for the balanced two-factor nested random effects model. *Test*, 16(3), 598 – 612.

Walter, S. D., Hildreth, S. W., & Beaty, B. J. (1980). Estimation of infection rates in populations of organisms using pools of variable size. *American Journal of Epidemiology*, 112(1), 124 – 128.

Wang, H. & Tsung, F. (2009). Tolerance intervals with improved coverage probabilities for binomial and Poisson variables. *Technometrics*, 51(1), 25 – 33.

Wolpert, R. L. (2004). A conversation with James O. Berger. *Statistical Science*, 19(1), 205 – 218.

Yang, R. & Berger, J. O. (1997). *A catalog of noninformative priors*. Technical report, Duke University, ISDS Discusion Paper 97-42.

Zhou, X.-H., Tsao, M., & Qin, G. (2004). New intervals for the difference between two independent binomial proportions. *Journal of Statistical Planning and Inference*, 123(1), 97 – 115.

Appendix A - Estimation for the Product of Binomial Rates

A.1 Additional Theorem and Proof

Theorem A.1. The inverse of the Fisher information matrix for k binomial rates when $n_1 = n_2 = \ldots = n_k$, is given by

$$\begin{aligned} F^{-1}\left(\underline{p}\right) &= F^{-1}\left(p_1, p_2, \ldots, p_k\right) \\ &= \begin{bmatrix} p_1(1-p_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & p_k(1-p_k) \end{bmatrix} \end{aligned}$$

Proof. Assume that $X_1, X_2, \ldots, X_k$ are independent binomial random variables with $X_i \sim Bin\left(n_i, p_i\right)$ for $i = 1, 2, \ldots, k$.

Therefore $P\left(X_i = x_i\right) = \begin{pmatrix} n_i \\ x_i \end{pmatrix} p_i^{x_i} \left(1-p_i\right)^{n_i - x_i}$ $\left(x_i = 0, 1, \ldots n_i\right)$.

The likelihood function is given by

$$\begin{aligned} L\left(p_1, p_2 \ldots, p_k \,|x_1, x_2 \ldots, x_k\right) &= L\left(\underline{p}\,|x_1, x_2 \ldots, x_k\right) \\ &= \prod_{i=1}^{k} \begin{pmatrix} n_i \\ x_i \end{pmatrix} p_i^{x_i} \left(1-p_i\right)^{n_i - x_i}. \end{aligned}$$

To obtain the Fisher information matrix we take the logarithm of the likelihood function and dif-

ferentiate twice with respect to the unknown parameters.

$$\begin{aligned}
\log L &= \text{constant} + \sum_{i=1}^{k} x_i log\left(p_i\right) + \sum_{i=1}^{k} \left(n_i - x_i\right) log\left(1-p_i\right). \\
\log L &\propto \sum_{i=1}^{k} x_i log\left(p_i\right) + \sum_{i=1}^{k} \left(n_i - x_i\right) log\left(1-p_i\right). \\
\frac{\partial \log L}{\partial p_i} &= \frac{x_i}{p_i} - \frac{n_i - x_i}{1-p_i} \\
\frac{\partial^2 \log L}{\left(\partial p_i\right)^2} &= \frac{-x_i}{p_i^2} - \frac{\left(n_i - x_i\right)}{\left(1-p_i\right)^2} \\
\frac{\partial^2 \log L}{\partial p_i \partial p_j} &= 0 \qquad\qquad (i \neq j).
\end{aligned}$$

Therefore

$$\begin{aligned}
-E\left(\frac{\partial^2 \log L}{\left(\partial p_i\right)^2}\right) &= \frac{n_i p_i}{p_i^2} + \frac{n_i\left(1-p_i\right)}{\left(1-p_i\right)^2} \\
&= \frac{n_i}{p_i} + \frac{n_i}{\left(1-p_i\right)} \\
&= \frac{n_i}{p_i\left(1-p_i\right)} \qquad\qquad (i = 1, \ldots, k), \\
-E\left(\frac{\partial^2 \log L}{\partial p_i \partial p_j}\right) &= 0 \qquad\qquad (i \neq j).
\end{aligned}$$

The Fisher information matrix is then given by

$$\begin{aligned}
F\left(\underline{p}\right) &= F\left(p_1, p_2, \ldots, p_k\right) \\
&= \begin{bmatrix} -E\left(\frac{\partial^2 \log L}{\left(\partial p_1\right)^2}\right) & \cdots & -E\left(\frac{\partial^2 \log L}{\partial p_1 \partial p_k}\right) \\ \vdots & & \vdots \\ -E\left(\frac{\partial^2 \log L}{\partial p_k \partial p_1}\right) & \cdots & -E\left(\frac{\partial^2 \log L}{\left(\partial p_k\right)^2}\right) \end{bmatrix} \\
&= \begin{bmatrix} \frac{n_1}{p_1\left(1-p_1\right)} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{n_k}{p_k\left(1-p_k\right)} \end{bmatrix}.
\end{aligned}$$

If $n_1 = n_2 = \ldots = n_k$, n can be ignored for all practical purposes (i.e. $n = 1$) and the inverse of the Fisher Information matrix is then given by

$$
\begin{aligned}
F^{-1}(\underline{p}) &= F^{-1}(p_1, p_2, \ldots, p_k) \\
&= \begin{bmatrix} \frac{1}{p_1(1-p_1)} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{p_k(1-p_k)} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} p_1(1-p_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & p_k(1-p_k) \end{bmatrix}.
\end{aligned}
$$

□

A.2 MATLAB® Code

MATLAB® code for the simulation studies considered for the product of binomial rates

```
%Jeffreys and uniform prior for the product of binomial proportions
clc;
clear;
p1 = 0.1:0.1:0.9;
p2 = 0.1:0.1:0.9;
n1 = 20;
n2 = 20;
theta0 = (p1)'*(p2);
matrix = [];
for i = 1:9
    SI = [];
    AVG = [];
    COUNT = [];
    for j = 1:9
        count = 0;
        Ii = [];
        THETA = [];
            for k = 1:1000
              x1 = binornd(n1,p1(i));
              x2 = binornd(n2,p2(j));
```

```
                %p_1 = betarnd(x1+1,n1-x1+1,1,1000); %uniform prior
                %p_2 = betarnd(x2+1,n2-x2+1,1,1000); %uniform prior
                p_1 = betarnd(x1+0.5,n1-x1+0.5,1,1000); %Jeffreys prior
                p_2 = betarnd(x2+0.5,n2-x2+0.5,1,1000); %Jeffreys prior
                theta = (p_1).*(p_2);
                THETA = [theta];
                temp = sort(THETA);
                lower = temp(50);
                upper = temp(950);
                leng = upper - lower;
                Ii = [Ii leng];
                if lower < theta0(i,j)
                   if upper >theta0(i,j)
                      count = count + 1;
                   end
                end
             end
          avg = sum(Ii)/1000;
          si = sqrt(1/999*(sum(Ii.^2) - (sum(Ii))^2/1000));
          SI = [SI si];
          AVG = [AVG avg];
          COUNT = [COUNT count];
       end
       matrix = [matrix; COUNT./1000; AVG; SI;];

end
```

```
%Probability Matching prior for the product of binomial proportions
clc
clear
p1 = 0.1:0.1:0.9;
p2 = 0.1:0.1:0.9;
n1 = 20;
n2 = 20;
theta0 = (p1.^2)'*(p2);
%change theta0 to (p1)'*(p2) for diff sim (same for PMP)
```

```
matrix = [];
for i = 1:9
    SI = [];
    AVG = [];
    COUNT = [];
    for j = 1:9
        count = 0;
        Ii = [];
        THETA = [];
        vectors = [];
        CHOSEN = [];
        chosen = [];
        for k = 1:1000
            x1 = binornd(n1,p1(i));
            x2 = binornd(n2,p2(j));
            p_1 = betarnd(x1+1,n1-x1+1,1,1000);
            p_2 = betarnd(x2+1,n2-x2+1,1,1000);
            theta = (p_1.^2).*(p_2);
            pp1 = p_1.^(-0.5);
            pp2 = p_2.^(-0.5);
            ppp1 = (1-p_1).^(-1);
            ppp2 = (1-p_2).^(-1);
            pppp1 = 1-p_1;
            pppp2 = 1-p_2;
            PMP = pp1.* pp2.* ppp1.* ppp2.* sqrt((2*p_2.*(1 - p_1))
                  + (p_1.*(1 - p_2)));
            vectors = [PMP];
            total  =sum(vectors);
            weights = vectors/total;
            vectorweight = [theta; weights]';
            sorteds = sortrows(vectorweight);
            CHOSEN = [sorteds];
            temp = sort(CHOSEN);
            lower = temp(round(0.05*length(CHOSEN)));
            upper = temp(round(0.95*length(CHOSEN)));
            leng = upper - lower;
            Ii = [Ii leng];
```

```
            if lower < theta0(i,j)
                 if upper >theta0(i,j)
                    count = count + 1;
                end
            end
        end
        avg = sum(Ii)/1000;
        si = sqrt(1/999*(sum(Ii.^2) - (sum(Ii))^2/1000));
        SI = [SI si];
        AVG = [AVG avg];
        COUNT = [COUNT count];
    end
    matrix = [matrix; COUNT./1000; AVG; SI;];
end
```

Appendix B - Estimation for a Linear Function of Binomial Rates

B.1 Additional Theorem and Proof

Theorem B.1. The maximum likelihood estimate (MLE) of p_i is equal to $\hat{p}_i = {}^{x_i}/_{n_i}$.

Proof. Assume that X_1, X_2 are independent binomial random variables with $X_i \sim Bin(n_i, p_i)$ for $i = 1, 2$. Therefore $P(X_i = x_i) = \begin{pmatrix} n_i \\ x_i \end{pmatrix} p_i^{x_i} (1-p_i)^{n_i - x_i}$ for $x_i = 0, 1, \ldots, n_i$.

The likelihood function is given by

$$L(p_1, p_2 | x_1, x_2) = L(\underline{p} | x_1, x_2) = \prod_{i=1}^{2} \begin{pmatrix} n_i \\ x_i \end{pmatrix} p_i^{x_i} (1-p_i)^{n_i - x_i}.$$

We take the logarithm of the likelihood function and differentiate with respect to the unknown parameter.

$$\begin{aligned} \log L &= \sum_{i=1}^{2} \log \begin{pmatrix} n_i \\ x_i \end{pmatrix} + \sum_{i=1}^{2} x_i \log(p_i) + \sum_{i=1}^{2} (n_i - x_i) \log(1-p_i) \\ &\propto \sum_{i=1}^{2} x_i \log(p_i) + \sum_{i=1}^{2} (n_i - x_i) \log(1-p_i) \\ \frac{\partial \log L}{\partial p_i} &= \frac{x_i}{p_i} - \frac{n_i - x_i}{1-p_i} \qquad (i = 1, 2). \end{aligned}$$

Now set the differential equal to zero and solve the unknown parameter.

$$
\begin{aligned}
\frac{\partial \log L}{\partial p_i} &= 0 \\
\therefore \frac{x_i}{p_i} - \frac{n_i - x_i}{1 - p_i} &= 0 \\
\frac{x_i}{p_i} &= \frac{n_i - x_i}{1 - p_i} \\
\frac{1 - p_i}{p_i} &= \frac{n_i - x_i}{x_i} \\
\frac{1}{p_i} - 1 &= \frac{n_i - x_i}{x_i} \\
\frac{1}{p_i} &= \frac{n_i - x_i + x_i}{x_i} = \frac{n_i}{x_i} \\
\therefore \hat{p}_i &= \frac{x_i}{n_i}.
\end{aligned}
$$

□

B.2 Additional Simulation Results

Tables B.1 to B.3 contain coverage probabilities, mean lengths and standard deviations for the Jeffreys, uniform and probability matching priors when $n_1 = n_2 = 20$, for several choices of p_1 and p_2. These results are summarised in Figure 3.4.

Table B.1: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the Jeffreys prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 20$.

	Jeffreys prior $n_1 = n_2 = 20$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9420	0.9330	0.9470	0.9580	0.9560	0.9280	0.9560	0.9690	0.9530	0.9491
(b)	0.3784	0.4296	0.4642	0.4811	0.4851	0.4768	0.4522	0.4180	0.3622	0.4386
(c)	1.0463	1.0607	1.0773	1.0640	1.0689	1.0502	1.0724	0.8912	0.9360	1.0297
0.2 (a)	0.9500	0.9290	0.9480	0.9460	0.9430	0.9490	0.9550	0.9460	0.9560	0.9469
(b)	0.4285	0.4775	0.5048	0.5245	0.5285	0.5193	0.4988	0.4652	0.4167	0.4849
(c)	1.1025	1.0747	1.0413	1.0477	1.0903	1.0591	1.0426	1.0112	0.9047	1.0416
0.3 (a)	0.9570	0.9330	0.9400	0.9430	0.9430	0.9430	0.9470	0.9440	0.9520	0.9447
(b)	0.4633	0.5068	0.5337	0.5529	0.5543	0.5490	0.5293	0.4989	0.4541	0.5158
(c)	1.0923	1.0690	1.0713	1.0655	1.0616	1.0725	1.0336	1.0248	1.0463	1.0597
0.4 (a)	0.9360	0.9490	0.9470	0.9480	0.9420	0.9460	0.9340	0.9480	0.9520	0.9447
(b)	0.4781	0.5243	0.5511	0.5660	0.5706	0.5657	0.5494	0.5197	0.4776	0.5336
(c)	1.0482	1.0333	1.0556	1.0566	1.0548	1.0504	1.0708	1.0570	1.0530	1.0533
0.5 (a)	0.9410	0.9310	0.9480	0.9510	0.9500	0.9590	0.9390	0.9400	0.9440	0.9448
(b)	0.4839	0.5267	0.5560	0.5716	0.5763	0.5701	0.5555	0.5249	0.4831	0.5387
(c)	1.0449	1.0627	1.0716	1.0528	1.0563	1.0705	1.0599	1.0764	1.0385	1.0593
0.6 (a)	0.9530	0.9490	0.9500	0.9270	0.9490	0.9450	0.9310	0.9470	0.9500	0.9446
(b)	0.4752	0.5181	0.5495	0.5653	0.5700	0.5658	0.5517	0.5217	0.4815	0.5332
(c)	1.0430	1.0527	1.0551	1.0865	1.0789	1.0621	1.0557	1.0775	1.0538	1.0628
0.7 (a)	0.9480	0.9550	0.9420	0.9380	0.9400	0.9490	0.9350	0.9470	0.9460	0.9444
(b)	0.4558	0.4992	0.5324	0.5480	0.5555	0.5502	0.5340	0.5054	0.4617	0.5158
(c)	1.0308	1.0517	1.0405	1.0651	1.0491	1.0670	1.0639	1.0633	1.0710	1.0558
0.8 (a)	0.9570	0.9410	0.9530	0.9420	0.9380	0.9490	0.9490	0.9440	0.9460	0.9466
(b)	0.4189	0.4683	0.5003	0.5176	0.5270	0.5222	0.5044	0.4763	0.4280	0.4848
(c)	0.9592	0.9823	0.9977	1.0228	1.0479	1.0390	1.0817	1.0486	1.0968	1.0307
0.9 (a)	0.9400	0.9690	0.9480	0.9450	0.9330	0.9490	0.9610	0.9550	0.9630	0.9514
(b)	0.3640	0.4189	0.4572	0.4755	0.4831	0.4798	0.4638	0.4326	0.3792	0.4393
(c)	0.8765	0.9540	0.9961	0.9955	1.0547	1.0284	1.0775	1.1014	1.0311	1.0128
mean (a)	0.9471	0.9432	0.9470	0.9442	0.9438	0.9463	0.9452	0.9489	0.9513	**0.9463**
(b)	0.4384	0.4855	0.5166	0.5336	0.5389	0.5332	0.5155	0.4848	0.4382	**0.4983**
(c)	1.0271	1.0379	1.0452	1.0507	1.0625	1.0555	1.0620	1.0390	1.0257	**1.0451**

Table B.2: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the uniform prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 20$.

	Uniform prior $n_1 = n_2 = 20$									
$\downarrow p_2$ $p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9740	0.9690	0.9580	0.9490	0.9480	0.9460	0.9430	0.9440	0.9240	0.9506
(b)	0.3923	0.4367	0.4629	0.4803	0.4842	0.4741	0.4536	0.4209	0.3758	0.4423
(c)	0.9676	1.0371	1.0449	1.0394	0.9921	1.0366	0.9311	0.8651	0.8070	0.9690
0.2 (a)	0.9560	0.9490	0.9620	0.9490	0.9540	0.9490	0.9470	0.9620	0.9340	0.9513
(b)	0.4380	0.4780	0.5019	0.5155	0.5188	0.5129	0.4941	0.4644	0.4204	0.4827
(c)	1.0686	1.0450	1.0534	1.0389	1.0343	0.9982	0.9565	0.8844	0.8552	0.9927
0.3 (a)	0.9590	0.9670	0.9550	0.9570	0.9450	0.9540	0.9500	0.9380	0.9430	0.9520
(b)	0.4642	0.5017	0.5275	0.5413	0.5451	0.5406	0.5240	0.4985	0.4560	0.5110
(c)	1.0504	1.0524	1.0555	1.0522	1.0490	1.0321	1.0122	0.9810	0.9126	1.0220
0.4 (a)	0.9590	0.9510	0.9490	0.9530	0.9610	0.9630	0.9600	0.9590	0.9510	0.9562
(b)	0.4813	0.5158	0.5425	0.5548	0.5585	0.5549	0.5392	0.5164	0.4760	0.5266
(c)	1.0587	1.0463	1.0555	1.0557	1.0508	1.0202	1.0263	1.0081	0.9741	1.0328
0.5 (a)	0.9480	0.9420	0.9440	0.9640	0.9530	0.9540	0.9610	0.9560	0.9430	0.9517
(b)	0.4813	0.5214	0.5450	0.5588	0.5648	0.5590	0.5451	0.5199	0.4838	0.5310
(c)	1.0049	1.0330	1.0348	1.0496	1.0637	1.0567	1.0222	1.0392	1.0415	1.0384
0.6 (a)	0.9480	0.9570	0.9480	0.9500	0.9580	0.9550	0.9520	0.9600	0.9550	0.9537
(b)	0.4773	0.5129	0.5390	0.5540	0.5596	0.5553	0.5429	0.5155	0.4807	0.5264
(c)	0.9608	0.9647	1.0097	1.0322	1.0686	1.0582	1.0594	1.0365	1.0546	1.0272
0.7 (a)	0.9410	0.9510	0.9640	0.9500	0.9640	0.9470	0.9600	0.9570	0.9480	0.9536
(b)	0.4581	0.4976	0.5250	0.5389	0.5460	0.5431	0.5270	0.5032	0.4634	0.5114
(c)	0.9340	0.9465	0.9942	1.0320	1.0596	1.0622	1.0752	1.0599	1.0371	1.0223
0.8 (a)	0.9280	0.9520	0.9460	0.9450	0.9500	0.9520	0.9460	0.9370	0.9560	0.9458
(b)	0.4221	0.4689	0.4959	0.5133	0.5208	0.5172	0.5011	0.4754	0.4362	0.4834
(c)	0.8567	0.9191	0.9841	0.9774	1.0114	1.0454	1.0586	1.0481	1.0691	0.9967
0.9 (a)	0.8970	0.9400	0.9390	0.9500	0.9550	0.9560	0.9520	0.9660	0.9730	0.9476
(b)	0.3745	0.4247	0.4571	0.4755	0.4843	0.4789	0.4666	0.4353	0.3924	0.4432
(c)	0.7940	0.8525	0.9315	0.9697	1.0191	1.0431	1.0759	1.0611	0.9870	0.9704
mean (a)	0.9456	0.9531	0.9517	0.9519	0.9542	0.9529	0.9523	0.9532	0.9474	**0.9514**
(b)	0.4432	0.4842	0.5108	0.5258	0.5313	0.5262	0.5104	0.4833	0.4427	**0.4953**
(c)	0.9662	0.9885	1.0182	1.0275	1.0387	1.0392	1.0242	0.9982	0.9709	**1.0079**

Table B.3: Coverage rate of the 95% credibility intervals for $\theta_1 = p_1 - p_2$ using the probability matching prior. (a) Exact coverage probabilities, (b) mean lengths, (c) standard deviation for $n_1 = n_2 = 20$.

	Probability Matching prior $n_1 = n_2 = 20$									
$\downarrow p_2 \quad p_1 \rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**mean**
0.1 (a)	0.9630	0.9600	0.9540	0.9490	0.9490	0.9450	0.9470	0.9380	0.9210	0.9473
(b)	0.3940	0.4400	0.4623	0.4784	0.4842	0.4764	0.4576	0.4237	0.3733	0.4433
(c)	0.9729	1.0888	1.0505	1.0453	1.0162	0.9944	0.9348	0.8591	0.7910	0.9725
0.2 (a)	0.9670	0.9590	0.9440	0.9490	0.9470	0.9500	0.9310	0.9460	0.9430	0.9484
(b)	0.4358	0.4761	0.5020	0.5148	0.5183	0.5147	0.4951	0.4684	0.4191	0.4827
(c)	1.0401	1.0614	1.0439	1.0409	1.0099	1.0014	1.0306	0.8976	0.8504	0.9974
0.3 (a)	0.9570	0.9490	0.9570	0.9620	0.9400	0.9510	0.9510	0.9320	0.9320	0.9479
(b)	0.4638	0.5016	0.5279	0.5424	0.5449	0.5400	0.5225	0.4958	0.4572	0.5107
(c)	1.0612	1.0450	1.0621	1.0613	1.0372	1.0345	0.9886	0.9970	0.9534	1.0267
0.4 (a)	0.9400	0.9580	0.9430	0.9470	0.9630	0.9590	0.9560	0.9580	0.9490	0.9526
(b)	0.4813	0.5162	0.5409	0.5549	0.5595	0.5546	0.5395	0.5140	0.4751	0.5262
(c)	1.0318	1.0549	1.0565	1.0658	1.0606	1.0658	1.0058	1.0030	0.9951	1.0377
0.5 (a)	0.9640	0.9500	0.9550	0.9560	0.9580	0.9490	0.9500	0.9520	0.9510	0.9539
(b)	0.4826	0.5216	0.5466	0.5584	0.5638	0.5599	0.5455	0.5176	0.4825	0.5310
(c)	0.9931	1.0173	1.0424	1.0410	1.0617	1.0655	1.0425	1.0271	0.9915	1.0313
0.6 (a)	0.9550	0.9400	0.9380	0.9470	0.9590	0.9560	0.9600	0.9600	0.9530	0.9520
(b)	0.4772	0.5123	0.5391	0.5542	0.5595	0.5547	0.5411	0.5170	0.4808	0.5262
(c)	0.9836	0.9856	1.0179	1.0382	1.0616	1.0641	1.0534	1.0362	1.0462	1.0319
0.7 (a)	0.9410	0.9530	0.9560	0.9560	0.9470	0.9580	0.9560	0.9540	0.9530	0.9527
(b)	0.4586	0.4967	0.5218	0.5402	0.5470	0.5410	0.5270	0.5008	0.4664	0.5111
(c)	0.9659	0.9950	1.0194	1.0329	1.0607	1.0519	1.0560	1.0532	1.0385	1.0304
0.8 (a)	0.9370	0.9420	0.9510	0.9520	0.9470	0.9590	0.9580	0.9460	0.9680	0.9511
(b)	0.4236	0.4694	0.4955	0.5150	0.5211	0.5146	0.5012	0.4766	0.4385	0.4839
(c)	0.8743	0.9548	0.9642	1.0096	1.0216	1.0249	1.0509	1.0385	1.1034	1.0047
0.9 (a)	0.9260	0.9390	0.9500	0.9600	0.9500	0.9490	0.9560	0.9610	0.9800	0.9523
(b)	0.3742	0.4221	0.4559	0.4765	0.4851	0.4818	0.4675	0.4354	0.3952	0.4437
(c)	0.7965	0.8482	0.9648	0.9627	1.0463	1.0665	1.0775	1.0881	0.9740	0.9805
mean (a)	0.9500	0.9500	0.9498	0.9531	0.9511	0.9529	0.9517	0.9497	0.9500	**0.9509**
(b)	0.4435	0.4840	0.5102	0.5261	0.5315	0.5264	0.5108	0.4832	0.4431	**0.4954**
(c)	0.9688	1.0057	1.0246	1.0331	1.0418	1.0410	1.0267	1.0000	0.9715	**1.0126**

B.3 MATLAB® Code

MATLAB® code for the simulation studies considered for the difference between binomial rates

```
%Jeffreys and uniform prior for the difference between binomial proportions
clc;
clear;
p1 = 0.1:0.1:0.9;
p2 = 0.1:0.1:0.9;
n1 = 26;
n2 = 20;
matrix = [];
for i = 1:9
    AVG2 = [];
    AVG = [];
    COUNT = [];
    for j = 1:9
        count = 0;
        li = [];
        li2 = [];
        THETA = [];
        for k = 1:1000
            x1 = binornd(n1,p1(i));
            x2 = binornd(n2,p2(j));
            %p_1 = betarnd(x1+1,n1-x1+1,1,1000); %uniform prior
            %p_2 = betarnd(x2+1,n2-x2+1,1,1000); %uniform prior
            p_1 = betarnd(x1+0.5,n1-x1+0.5,1,1000); %Jeffreys prior
            p_2 = betarnd(x2+0.5,n2-x2+0.5,1,1000); %Jeffreys prior
            theta = p_1-p_2;
            THETA = [theta];
            temp = sort(THETA);
            lower = temp(25);
            upper = temp(975);
            leng = upper - lower;
            theta0 = p1(i) - p2(j);
            if lower < theta0
```

```
            if upper >theta0
               count = count + 1;
               li = [li leng];
            else
               li2 = [li2 leng];
            end
         else
           li2 = [li2 leng];
         end
      end
      avg = sum(li)/length(li);
      AVG = [AVG avg];
      avg2 = sum(li2)/length(li2);
      AVG2 = [AVG2 avg/avg2];
      COUNT = [COUNT count];
   end
   matrix = [matrix; COUNT./1000; AVG; AVG2;];
end
```

```
%Probability matching prior for the difference between binomial proportions
clc;
clear;
p1 = 0.1:0.1:0.9;
p2 = 0.1:0.1:0.9;
n1 = 10;
n2 = 30;
matrix = [];
for i = 1:9
   AVG2 = [];
   AVG = [];
   COUNT = [];
   for j = 1:9
      count = 0;
      li = [];
```

```
li2 = [];
THETA = [];
vectors = [];
CHOSEN = [];
chosen = [];
for k = 1:1000
     x1 = binornd(n1,p1(i));
     x2 = binornd(n2,p2(j));
     p_1 = betarnd(x1+1,n1-x1+1,1,1000);
     p_2 = betarnd(x2+1,n2-x2+1,1,1000);
     theta = p_1-p_2;
     pp1 = p_1.^(-1);
     pp2 = p_2.^(-1);
     ppp1 = (1-p_1).^(-1);
     ppp2 = (1-p_2).^(-1);
     pppp1 = 1-p_1;
     pppp2 = 1-p_2;
     PMP = pp1.* pp2.* ppp1.* ppp2.* sqrt((p_1.*(1 - p_1))
           + (p_2.*(1 - p_2)));
     vectors = [PMP];
     total =sum(vectors);
     weights = vectors/total;
     vectorweight = [theta; weights]';
     sorteds = sortrows(vectorweight);
     CHOSEN = [sorteds];
     temp = sort(CHOSEN);
     lower = temp(round(0.025*length(CHOSEN)));
     upper = temp(round(0.975*length(CHOSEN)));
     leng = upper - lower;
     theta0 = p1(i) - p2(j);
     if lower < theta0
       if upper >theta0
          count = count + 1;
          li = [li leng];
       else
          li2 = [li2 leng];
       end
```

```
            else
            li2 = [li2 leng];
            end
        end
        avg = sum(li)/length(li);
        AVG = [AVG avg];
        avg2 = sum(li2)/length(li2);
        AVG2 = [AVG2 avg/avg2];
        COUNT = [COUNT count];
    end
    matrix = [matrix; COUNT./1000; AVG; AVG2;];
end
```

The code can be adapted to be used for the example on the Mal de Rio Cuarto Virus.

Appendix C - Estimation for the Ratio and Product of Poisson Rates

C.1 Additional Theorem and Proof

Theorem C.1. The inverse of the Fisher information matrix for k Poisson is given by

$$\begin{aligned} F^{-1}(\underline{\lambda}) &= F^{-1}(\lambda_1, \lambda_2, \ldots, \lambda_k) \\ &= \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}. \end{aligned}$$

Proof. Assume that $X_1, X_2, \ldots, X_k$ are independent Poisson random variables with $X_i \sim P(\lambda_i)$, for $i = 1, 2, \ldots, k$.

Therefore $P(X_i = x_i) = \dfrac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}$ $(x_i = 0, 1, 2, \ldots)$.

The likelihood function is given by

$$\begin{aligned} L(\lambda_1, \lambda_2 \ldots, \lambda_k | x_1, x_2 \ldots, x_k) &= L(\underline{\lambda} | x_1, x_2 \ldots, x_k) \\ &= \prod_{i=1}^{k} \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}. \end{aligned}$$

To obtain the Fisher information matrix we take the logarithm of the likelihood function and differentiate twice with repsect to the unknown parameters.

$$
\begin{aligned}
\log L &= \text{constant} + \sum_{i=1}^{k} x_i log\left(\lambda_i\right) - \sum_{i=1}^{k} \lambda_i \\
\log L &\propto \sum_{i=1}^{k} x_i log\left(\lambda_i\right) - \sum_{i=1}^{k} \lambda_i \\
\frac{\partial \log L}{\partial \lambda_i} &= \frac{x_i}{\lambda_i} - 1 \\
\frac{\partial^2 \log L}{\left(\partial \lambda_i\right)^2} &= \frac{-x_i}{\lambda_i^2} \\
\frac{\partial^2 \log L}{\partial \lambda_i \partial \lambda_j} &= 0 \qquad\qquad (i \neq j).
\end{aligned}
$$

$(i = 1, \ldots, k)$

Therefore

$$
\begin{aligned}
-E\left(\frac{\partial^2 \log L}{\left(\partial \lambda_i\right)^2}\right) &= \frac{\lambda_i}{\lambda_i^2} = \frac{1}{\lambda_i} \qquad\qquad (i = 1, \ldots, k), \\
-E\left(\frac{\partial^2 \log L}{\partial \lambda_i \partial \lambda_j}\right) &= 0 \qquad\qquad (i \neq j.)
\end{aligned}
$$

The Fisher information matrix is then given by

$$
\begin{aligned}
F\left(\underline{\lambda}\right) &= F\left(\lambda_1, \lambda_2, \ldots, \lambda_k\right) \\
&= \begin{bmatrix} -E\left(\frac{\partial^2 \log L}{\left(\partial \lambda_1\right)^2}\right) & \cdots & -E\left(\frac{\partial^2 \log L}{\partial \lambda_1 \partial \lambda_k}\right) \\ \vdots & & \vdots \\ -E\left(\frac{\partial^2 \log L}{\partial \lambda_k \partial \lambda_1}\right) & \cdots & -E\left(\frac{\partial^2 \log L}{\left(\partial \lambda_k\right)^2}\right) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\lambda_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\lambda_k} \end{bmatrix}.
\end{aligned}
$$

The inverse of the Fisher information matrix is then given by

$$\begin{aligned} F^{-1}(\underline{\lambda}) &= F^{-1}(\lambda_1, \lambda_2, \ldots, \lambda_k) \\ &= \begin{bmatrix} \frac{1}{\lambda_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\lambda_k} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}. \end{aligned}$$

□

C.2 MATLAB® Code

MATLAB® code for the simulation studies considered for the product of Poisson rates

```
%PRODUCT OF TWO POISSON PARAMETERS, PROB MATCHING PRIOR
clc;
clear;
%Lambda2 runs from 2 to 10
L2=[2:10];
RESULTS1=[];
RESULTS2=[];
for u=1:9
    %Set the value of Lambda1
    l1=10;
    l2=L2(u);
    L=[l1;l2];
    %Set a1, a2 equal to 0.5, 1 or (1,-1) for product, square-root or ratio
    a1=1;
    a2=1;
    true=(l1^a1)*(l2^a2);
    m=500;
    Lo=[];
    Up=[];
    Leng=[];
    HPDlo=[];
```

```
HPDup=[];
HPDleng=[];
for i=1:m
    x=poissrnd(L);
    k=1000;
    r=ones(1,k);
    la=gamrnd((x+1)*r,1);
    w=sqrt((a1^2)./la(1,:)+(a2^2)./la(2,:));
    W=w/sum(w);
    theta=(la(1,:).^a1).*(la(2,:).^a2);
    [Y,I]=sort(theta);
    %Weighted Monte Carlo Method
    W2=W(I);
    C=cumsum(W2);
    J1=min(find(C>=0.024));
    J2=min(find(C>=0.975));
    lo=Y(J1);
    up=Y(J2);
    P=C(J2)-C(J1);
    leng=up-lo;
    Lo=[Lo;lo];
    Up=[Up;up];
    Leng=[Leng;leng];
    H=[];
    Y11=[];
    Y22=[];
    for j=0.01:0.001:0.04
        K1=min(find(C>=j));
        K2=min(find(C>=0.95+j));
        Y1=Y(K1);
        Y2=Y(K2);
        Y11=[Y11;Y1];
        Y22=[Y22;Y2];
        h=Y2-Y1;
        H=[H;h];
    end
    hpdleng=min(H);
```

```
        HPDleng=[HPDleng;hpdleng];
        Q=find(H==min(H));
        hpdlo=Y11(Q);
        hpdup=Y22(Q);
        HPDlo=[HPDlo;hpdlo];
        HPDup=[HPDup;hpdup];
    end
    cover=length(find(Lo<=true & Up>=true))*100/m;
    LENG=mean(Leng);
    VAR=cov(Leng);
    HPDcover=length(find(HPDlo<=true & HPDup>=true))*100/m;
    HPDLENG=mean(HPDleng);
    HPDVAR=cov(HPDleng);
    u+1
    results1=[cover;LENG;VAR];
    results2=[HPDcover;HPDLENG;HPDVAR];
    RESULTS1=[RESULTS1 results1]; %RESULTS1 is for equal-tail
    RESULTS2=[RESULTS2 results2]; %RESULTS2 is for shortest interval

end
```

```
%PRODUCT OF POISSON PARAMETERS FOR JEFFREYS TYPE PRIORS
clc;
clear;
%Lambda2 runs from 2 to 10
L2=[2:10];
RESULTS1=[];
RESULTS2=[];
for j=1:9

    %Set the value of Lambda1
    l1=2;
    l2=L2(j);
    L=[l1;l2];
    %Set a1, a2 equal to 0.5, 1 or (1,-1) for product, square-root or ratio
    a1=0.5;
    a2=0.5;
```

```
true=(l1^a1)*(l2^a2);
m=5000;
Lo=[];
Up=[];
HPDlo=[];
HPDup=[];
Leng=[];
HPDleng=[];
for i=1:m
    x=poissrnd(L);
    k=10000;
    w=ones(1,k);
    %Parameter in gamrnd depends on prior used
    la=gamrnd((x+1)*w,1);
    theta=(la(1,:).^a1).*(la(2,:).^a2);
    st=sort(theta);
    lo=st(0.025*k);
    up=st(0.975*k);
    leng=up-lo;
    Lo=[Lo;lo];
    Up=[Up;up];
    Sl=[];
    for w=1:499
        sl=st(10000-w)-st(500-w);
        Sl=[Sl;sl];
    end
    hpdleng=min(Sl);
    Leng=[Leng;leng];
    HPDleng=[HPDleng;hpdleng];
    J=find(Sl==min(Sl));
    hpdlo=st(500-J);
    hpdup=st(10000-J);
    HPDlo=[HPDlo;hpdlo];
    HPDup=[HPDup;hpdup];
end
cover=length(find(Lo<=true & Up>=true))*100/m;
LENG=mean(Leng);
```

```
    HPDcover=length(find(HPDlo<=true & HPDup>=true))*100/m;
    HPDLENG=mean(HPDleng);
    VAR=cov(Leng);
    HPDVAR=cov(HPDleng);
    j+1
    results1=[cover;LENG;VAR];
    results2=[HPDcover;HPDLENG;HPDVAR];
    RESULTS1=[RESULTS1 results1]; %RESULTS1 is for equal-tail
    RESULTS2=[RESULTS2 results2]; %RESULTS2 is for shortest interval

end
```

MATLAB® code for the simulation studies considered for the ratio of Poisson rates

```
%RATIO OF POISSON PARAMETERS USING THE F-DISTRIBUTION
%ONLY FOR RATIO
clc;
clear;
L2=[2:10];
RESULTS1=[];
RESULTS2=[];
for j=1:9

    l1=10;
    l2=L2(j);
    L=[l1;l2];
    true=l1/l2;
    m=5000;
    Lo=[];
    Up=[];
    HPDlo=[];
    HPDup=[];
    Leng=[];
    HPDleng=[];
    for i=1:m

        x=poissrnd(L);
        k=10000;
        %c=2 for uniform prior
        %c=1 for Jeffreys prior
```

```
        %c=0 for Prob Matching prior (Does not work if any x=0)
        c=2;
        a=(x(1)+c)/(x(2)+c);
        theta=a*frnd(2*x(1)+c,2*x(2)+c,k,1);
        st=sort(theta);
        lo=st(0.025*k);
        up=st(0.975*k);
        leng=up-lo;
        Lo=[Lo;lo];
        Up=[Up;up];
        Sl=[];
        for w=1:499
            sl=st(10000-w)-st(500-w);
            Sl=[Sl;sl];
        end
        hpdleng=min(Sl);
        Leng=[Leng;leng];
        HPDleng=[HPDleng;hpdleng];
        J=find(Sl==min(Sl));
        hpdlo=st(500-J);
        hpdup=st(10000-J);
        HPDlo=[HPDlo;hpdlo];
        HPDup=[HPDup;hpdup];
    end
    cover=length(find(Lo<=true & Up>=true))*100/m;
    LENG=mean(Leng);
    HPDcover=length(find(HPDlo<=true & HPDup>=true))*100/m;
    HPDLENG=mean(HPDleng);
    VAR=cov(Leng);
    HPDVAR=cov(HPDleng);
    j+3
    results1=[cover;LENG;VAR];
    results2=[HPDcover;HPDLENG;HPDVAR];
    RESULTS1=[RESULTS1 results1]; %RESULTS1 is for equal-tail
    RESULTS2=[RESULTS2 results2]; %RESULTS2 is for shortest interval

end
```

```
%Uniform and Jeffreys priors (ratio)
clc;
clear;
lambda1 = 2:1:10;
lambda2 = 2:1:10;
lambda11 = 1./lambda1;
lambda2a = 1./lambda2;
theta0 = lambda1'*lambda2a;
%theta0 = lambda11'*lambda2;
matrix = [];
for i = 1:9

    SI = [];
    AVG = [];
    COUNT = [];
    for j = 1:9

        count = 0;
        Ii = [];
        THETA = [];
        for k = 1:1000

           x1 = poissrnd(lambda1(i));
           x2 = poissrnd(lambda2(j));
           %lambda_1 = gamrnd(x1+1,1,1,1000); %Uniform prior
           %lambda_2 = gamrnd(x2+1,1,1,1000); %Uniform prior
           lambda_1 = gamrnd(x1+0.5,1,1,1000); %Jeffreys prior
           lambda_2 = gamrnd(x2+0.5,1,1,1000); %Jeffreys prior
           theta = lambda_1./lambda_2;
           THETA = [theta];
           temp = sort(THETA);
           lower = temp(25);
           upper = temp(975);
           leng = upper - lower;
           Ii = [Ii leng];
           if lower < theta0(i,j)

             if upper > theta0(i,j)
                count = count + 1;
             end
```

```
            end
        end
        avg = sum(Ii)/1000;
        si = sqrt(1/999*(sum(Ii.^2) - (sum(Ii))^2/1000));
        SI = [SI si];
        AVG = [AVG avg];
        COUNT = [COUNT count];
    end
    matrix = [matrix; COUNT./1000; AVG; SI;];
end
```

Appendix D - Estimation for Linear Functions of Poisson Rates

D.1 MATLAB® Code for Linear Combination - coverage

MATLAB® code for the example considered in this chapter

```
%Probability matching prior (DWI)
clc;
clear;
%DATA FOR PMP I
X=[0;5;2;11;8;9];
%DATA FOR PMPII
%X=[0.5;5;2;11;8;9];
Z=[];
W=[];
GG=[];
for i=1:10000

    %FOR PMPI
    G=gamrnd(X+0.5,1);
    %FOR PMPII
    %G=gamrnd(X,1);
    w=sqrt(sum(G));
    GG=[GG G];
    %FOR MEAN (can only use PMP II)
    %z=mean(G);
    %FOR DIFF BETWEEN WINTER AND SUMMER
    z=mean(G(4:6))-mean(G(1:3));
    W=[W;w];
    Z=[Z;z];
```

```
end
W=W/sum(W);
[Y,I]=sort(Z);
Wi=W(I);
cW=cumsum(Wi);
J1=max(find(cW<=0.025));
J2=min(find(cW>=0.975));
lo=Y(J1);
up=Y(J2);
leng=up-lo;
K1=[];
K2=[];
L=[];
Hpdlo=[];
Hpdup=[];
for j=0.01:0.001:0.04

    k1=max(find(cW<=j));
    k2=min(find(cW>=0.95+j));
    hpdlo=Y(k1);
    hpdup=Y(k2);
    l=hpdup-hpdlo;
    K1=[K1;k1];
    K2=[K2;k2];
    L=[L;l];
    Hpdlo=[Hpdlo;hpdlo];
    Hpdup=[Hpdup;hpdup];

end
hpdl=min(L);
q=find(L==hpdl);
HPDlo=Hpdlo(q);
HPDup=Hpdup(q);
J=[0.01:0.001:0.04];
TAILS=[J(q) 0.95+J(q)];
COVER=[lo up leng cW(J1) cW(J2)]
HPDCOVER=[HPDlo HPDup hpdl TAILS]
```

```
%Jeffreys prior (DWI)
clc;
clear;
X=[0;5;2;11;8;9];
Z=[];
for i=1:100000

    G=gamrnd(X+0.5,1);
    %FOR MEAN
    %z=mean(G);
    %FOR DIFF BETWEEN WINTER AND SUMMER
    z=mean(G(4:6))-mean(G(1:3));
    Z=[Z;z];

end
sZ=sort(Z);
lo=sZ(0.025*i);
up=sZ(0.975*i);
leng=up-lo;
HPDleng=[];
L=[];
U=[];
for j=0.01:0.001:0.04

    l=sZ(round(j*i));
    u=sZ(round(0.95*i+j*i));
    hpdleng=u-l;
    HPDleng=[HPDleng;hpdleng];
    L=[L;l];
    U=[U;u];

end
minleng=min(HPDleng);
K=find(HPDleng==minleng);
HPDLo=L(K);
HPDUp=U(K);
J=[0.01:0.001:0.04];
TAILS=[J(K) 0.95+J(K)];
COVER=[lo up leng]
HPDCOVER=[HPDLo HPDUp minleng TAILS]
```

MATLAB® code for the simulation studies considered for linear combinations of Poisson rates

```
%Jeffreys prior (linear comb Poiss)
clc;
clear;
RESULTS=[];
for i=1:10

    k=5;
    n=1000;
    lamb=5+rand(k,n)*5;
    true=lamb(1,:)-lamb(2,:)/4-lamb(3,:)/4-lamb(4,:)/4-lamb(5,:)/4;
    %true=lamb(1,:)/2+lamb(2,:)/2-lamb(3,:)/2-lamb(4,:)/2;
    %true=lamb(1,:)-0.5*lamb(2,:)-0.5*lamb(3,:);
    %true=mean(lamb);
    %true=lamb(1,:)-lamb(2,:);
    x=poissrnd(lamb);
    m=10000;
    G1=gamrnd(ones(m,1)*(x(1,:)+0.5),1);
    G2=gamrnd(ones(m,1)*(x(2,:)+0.5),1);
    G3=gamrnd(ones(m,1)*(x(3,:)+0.5),1);
    G4=gamrnd(ones(m,1)*(x(4,:)+0.5),1);
    G5=gamrnd(ones(m,1)*(x(5,:)+0.5),1);
    G=G1-G2/4-G3/4-G4/4-G5/4;
    %G=G1/2+G2/2-G3/2-G4/2;
    clear G1 G2 G3 G4 G5
    sG=sort(G);
    lo=sG(0.025*m,:);
    up=sG(0.975*m,:);
    leng=mean(up-lo);
    p=100*length(find(lo<=true & up>=true))/n;
    Hl=[];
    K2=[];
    K1=[];
    for j=0.01:0.001:0.04

        k1=sG(round(j*m),:);
        k2=sG(round((0.95+j)*m),:);
        hl=k2-k1;
```

```
        Hl=[Hl;hl];
        K1=[K1;k1];
        K2=[K2;k2];
    end
    [Y,I]=min(Hl);
    hpdlo=diag(K1(I,:))';
    hpdup=diag(K2(I,:))';
    hpdleng=mean(Y);
    hp=100*length(find(hpdlo<=true & hpdup>=true))/n;
    RESULTS=[RESULTS;[p leng hp hpdleng]];
end
mean(RESULTS)
```

```
%Probability matching prior (linear comb Poiss)
clc;
clear;
RESULTS=[];
for u=1:200
    k=2;
    n=10;
    lamb=5+rand(k,n)*5;
    %true=lamb(1,:)-lamb(2,:)/4-lamb(3,:)/4-lamb(4,:)/4-lamb(5,:)/4;
    %true=lamb(1,:)/2+lamb(2,:)/2-lamb(3,:)/2-lamb(4,:)/2;
    %true=lamb(1,:)-0.5*lamb(2,:)-0.5*lamb(3,:);
    %true=mean(lamb);
    true=lamb(1,:)-lamb(2,:);
    x=poissrnd(lamb);
    for v=1:k
        I=find(x(v,:)==0);
        x(v,I)=0.5;
    end
    Z=[];
    W=[];
    GG=[];
    m=4000;
```

```
for i=1:m
    %G=gamrnd(x+0.5,1);
    G=gamrnd(x,1);
    w=sqrt(G(1,:)+G(2,:));
    GG=[GG G];
    z=(G(1,:)-G(2,:));
    %z=mean(G);
    W=[W;w];
    Z=[Z;z];
end
clear GG
W=W./(ones(m,1)*sum(W));
[Y,I]=sort(Z);
for j=1:n
    Wi(:,j)=W(I(:,j),j);
end
cW=cumsum(Wi);
for j=1:n
    J1(j)=max(find(cW(:,j)<=0.025));
    J2(j)=max(find(cW(:,j)<=0.975));
    lo(j)=Y(J1(j),j);
    up(j)=Y(J2(j),j);
end
leng=mean(up-lo);
p=100*length(find(lo<=true & up>=true))/n;
K1=[];
K2=[];
L=[];
Hpdlo=[];
Hpdup=[];
for j=0.01:0.001:0.04
    for t=1:n
       k1(t)=max(find(cW(:,t)<=j));
       k2(t)=max(find(cW(:,t)<=0.95+j));
       hpdlo(t)=Y(k1(t),t);
       hpdup(t)=Y(k2(t),t);
```

```
            l(t)=hpdup(t)-hpdlo(t);
        end
        K1=[K1;k1];
        K2=[K2;k2];
        L=[L;l];
        Hpdlo=[Hpdlo;hpdlo];
        Hpdup=[Hpdup;hpdup];
    end
    for t=1:n
        hpdl(t)=min(L(:,t));
        q(t)=min(find(L(:,t)==hpdl(t)));
        HPDlo(t)=Hpdlo(q(t),t);
        HPDup(t)=Hpdup(q(t),t);
    end
    hl=mean(HPDup-HPDlo);
    hp=100*length(find(HPDlo<=true & HPDup>=true))/n;
    RESULTS=[RESULTS;[leng p hl hp]];
end
mean(RESULTS)
```

D.2 MATLAB® Code for the Power and Size of Tests

MATLAB® code for the simulation studies considered for the size of the tests

```
%Probability matching prior (size)
%FIRST OR SECOND PM PRIOR %ONE-SIDED TESTS
clc;
clear;
RESULTS=[];
for u=1:10
    k=2;
    n=1000;
    lamb=35;
    true = 0;
    alpha = 0.01;
    x=poissrnd(lamb,k,n);
```

```
        %FOR PMP II
        %for v=1:k
        %    I=find(x(v,:)==0);
        %    x(v,I)=0.5;
        %end
        Z=[];
        W=[];
        GG=[];
        m=5000;
        for i=1:m
            %FOR PMP I
            G=gamrnd(x+0.5,1);
            %FOR PMP II
            %G=gamrnd(x,1);
            w=sqrt(G(1,:)+G(2,:));
            GG=[GG G];
            z=(G(1,:)-G(2,:));
            %z=mean(G);
            W=[W;w];
            Z=[Z;z];
        end
        clear GG
        W=W./(ones(m,1)*sum(W));
        [Y,I]=sort(Z);
        for j=1:n
            Wi(:,j)=W(I(:,j),j);
        end
        cW=cumsum(Wi);
        for j=1:n
            J(j)=max(find(cW(:,j)<=(1-alpha)));
        up(j)=Y(J(j),j);
        end
        leng=mean(up);
        p=100*length(find(up>=true))/n;
        RESULTS=[RESULTS;[lamb 100-p leng]];
    end
```

```
mean(RESULTS)
```

```
%Jeffreys and other two priors (size)
%ONE-SIDED TESTS
clc;
clear;
RESULTS=[];
for i=1:10

    k=2;
    n=1000;
    lamb=35;
    true=0;
    alpha = 0.001;
    x=poissrnd(lamb,k,n);
    m=5000;
    %G1=gamrnd(ones(m,1)*(x(1,:)+0.625),1); %for prior 3/8
    %G2=gamrnd(ones(m,1)*(x(2,:)+0.625),1); %for prior 3/8
    %G1=gamrnd(ones(m,1)*(x(1,:)+0.75),1); %for prior 1/4
    %G2=gamrnd(ones(m,1)*(x(2,:)+0.74),1); %for prior 1/4
    G1=gamrnd(ones(m,1)*(x(1,:)+0.5),1);
    G2=gamrnd(ones(m,1)*(x(2,:)+0.5),1);
    G=G1-G2;
    clear G1 G2
    sG=sort(G);
    up=sG((1-alpha)*m,:);
    leng=mean(up);
    p=100*length(find(up>=true))/n;
    RESULTS=[RESULTS;[lamb 100-p leng]];

end
mean(RESULTS)
```

MATLAB® code for the simulation studies considered for the power of the tests

```
%Probability matching prior (power)
clc;
clear;
 R=[];
lamb2=2;
for lamb1=1:20

    RESULTS=[];
    for i=1:10

        n=1000;
        true=lamb1-lamb2;
        x1=poissrnd(lamb1,1,n);
        x2=poissrnd(lamb2,1,n);
        %FOR PMP II
        %I1=find(x1==0);
        %x1(I1)=0.5;
        %I2=find(x2==0);
        %x2(I2)=0.5;
        Z=[];
        W=[];
        GG=[];
        m=5000;
        for j=1:m
            %FOR PMP I
            G1=gamrnd((x1+0.5),1);
            G2=gamrnd((x2+0.5),1);
            %FOR PMP II
            %G1=gamrnd((x1),1);
            %G2=gamrnd((x2),1);
            z=G1-G2;
            w=sqrt(G1+G2);
            W=[W;w];
            Z=[Z;z];
            clear G1 G2

        end
        W=W./(ones(m,1)*sum(W));
```

```
        [Y,I]=sort(Z);
        for j=1:n
            Wi(:,j)=W(I(:,j),j);
        end
        cW=cumsum(Wi);
        for j=1:n
            J(j)=max(find(cW(:,j)<=0.05));
            lo(j)=Y(J(j),j);
        end
        power=100*length(find(lo>=0))/n;
        RESULTS=[RESULTS;[lamb1 power]];
    end
    M=mean(RESULTS);
    R=[R;M];
end
```

```
%Jeffreys and other two priors (power)
%ONE-SIDED TESTS
clc;
clear;
R=[];
lamb2=20;
for lamb1=10:0.5:55
    RESULTS=[];
    for i=1:10
        n=1000;
        true=lamb1-lamb2;
        x1=poissrnd(lamb1,1,n);
        x2=poissrnd(lamb2,1,n);
        m=5000;
        %G1=gamrnd(ones(m,1)*(x(1,:)+0.625),1); %for prior 3/8
        %G2=gamrnd(ones(m,1)*(x(2,:)+0.625),1); %for prior 3/8
        %G1=gamrnd(ones(m,1)*(x(1,:)+0.75),1); %for prior 1/4
        %G2=gamrnd(ones(m,1)*(x(2,:)+0.74),1); %for prior 1/4
```

```
            G1=gamrnd(ones(m,1)*(x1+0.5),1);
            G2=gamrnd(ones(m,1)*(x2+0.5),1);
            G=G1-G2;
            clear G1 G2
            sG=sort(G);
            lo=sG(0.05*m,:);
            power=100*length(find(lo>=0))/n;
            RESULTS=[RESULTS;[lamb1 power]];
        end
        M=mean(RESULTS);
        R=[R;M];
    end
```

Appendix E - Estimation for Binomial Rates from Pooled Samples

E.1 Additional Theorem and Proof

Theorem E.1. The Fisher information, derived by Walter et al. (1980), in the case of M independent binomial random variables, $X_i \sim Bin\left(n_i, 1-(1-p)^{m_i}\right)$, is given by

$$F(p) = \sum_{i=1}^{M} \frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]}$$

Proof. Assume that $X_1, X_2, \ldots, X_M$ are independent binomial random variables with $X_i \sim Bin\left(n_i, 1-(1-p)^{m_i}\right)$.

The likelihood function is given by

$$L(p|x_1, x_2, \ldots, x_M) \propto \prod_{i=1}^{M} \left\{ [1-(1-p)^{m_i}]^{x_i} [(1-p)^{m_i}]^{n_i-x_i} \right\}.$$

To obtain the Fisher information we take the logarithm of the likelihood function and differentiate twice with respect to the unknown parameter.

$$\begin{aligned}
\log L &= \text{constant} + \textstyle\sum_{i=1}^{M} \{x_i \log[1-(1-p)^{m_i}]\} + \sum_{i=1}^{M} m_i (n_i - x_i) \log(1-p) \\
&\propto \sum_{i=1}^{M} \{x_i \log[1-(1-p)^{m_i}]\} + \sum_{i=1}^{M} m_i (n_i - x_i) \log(1-p) \\
&= \sum_{i=1}^{M} \{x_i \log[1-(1-p)^{m_i}]\} + \log(1-p) \sum_{i=1}^{M} m_i (n_i - x_i)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \log L}{\partial p} &= \sum_{i=1}^{M} \frac{x_i m_i (1-p)^{m_i-1}}{[1-(1-p)^{m_i}]} - \frac{1}{(1-p)} \sum_{i=1}^{M} m_i (n_i - x_i) \\
&= \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{x_i m_i (1-p)^{m_i}}{[1-(1-p)^{m_i}]} - m_i (n_i - x_i) \right\} \\
&= \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{x_i m_i (1-p)^{m_i} - m_i (n_i - x_i)[1-(1-p)^{m_i}]}{[1-(1-p)^{m_i}]} \right\} \\
&= \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{x_i m_i (1-p)^{m_i} - m_i n_i + m_i x_i + m_i n_i (1-p)^{m_i} - m_i x_i (1-p)^{m_i}}{[1-(1-p)^{m_i}]} \right\} \\
&= \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{-m_i n_i + m_i x_i + m_i n_i (1-p)^{m_i}}{[1-(1-p)^{m_i}]} \right\} \\
&= \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{m_i x_i}{[1-(1-p)^{m_i}]} - m_i n_i \right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial p^2} &= \frac{1}{(1-p)^2} \sum_{i=1}^{M} \left\{ \frac{m_i x_i}{[1-(1-p)^{m_i}]} - m_i n_i \right\} \\
&\quad - \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{m_i x_i m_i (1-p)^{m_i-1}}{[1-(1-p)^{m_i}]^2} \right\} \\
\therefore -E\left(\frac{\partial^2 \log L}{\partial p^2} \right) &= -\frac{1}{(1-p)^2} \sum_{i=1}^{M} \left\{ \frac{m_i n_i [1-(1-p)^{m_i}]}{[1-(1-p)^{m_i}]} - m_i n_i \right\} \\
&\quad + \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{m_i^2 n_i [1-(1-p)^{m_i}](1-p)^{m_i-1}}{[1-(1-p)^{m_i}]^2} \right\} \\
&= \frac{1}{(1-p)} \sum_{i=1}^{M} \left\{ \frac{m_i^2 n_i (1-p)^{m_i-1}}{[1-(1-p)^{m_i}]} \right\} \\
\therefore F(p) = -E\left(\frac{\partial^2 \log L}{\partial p^2} \right) &= \sum_{i=1}^{M} \left\{ \frac{m_i^2 n_i (1-p)^{m_i-2}}{[1-(1-p)^{m_i}]} \right\}.
\end{aligned}$$

□

E.2 Data

Table E.1 contains the data used by Biggerstaff (2008) of *Culex nigripalpus* mosquitoes trapped at heights of $6m$ and $1.5m$. Where m_i is the pool size for $i = 1, 2, \ldots, M$, and M indicates the number of distinct pool sizes. Where n_i is the number of pools of size m_i and X_i indicates the number of the n_i pools that is positive.

Table E.1: Biggerstaff (2008) Example Data

Sample 1 height $= 6m$			Sample 2 height $= 1.5m$		
m_1	n_1	x_1	m_2	n_2	x_2
50	22	6	100	1	0
49	7	0	50	12	0
48	3	0	49	2	0
47	2	0	48	3	0
46	1	0	46	1	0
42	1	0	45	1	1
41	1	0	42	1	0
35	2	0	41	2	0
32	1	1	40	1	0
27	1	0	36	1	0
20	1	0	35	1	0
17	1	0	22	1	0
9	1	0	13	1	0
8	1	0	10	1	0
7	2	0	6	1	0
6	1	0	5	1	0
3	1	0			
2	1	0			
1	3	0			
$M_1 = 19$			$M_2 = 16$		

E.3 MATLAB® Code

E.3.1 MATLAB® Code for Biggerstaff (2008) Example

```
clear;
clc;
%BIGGERSTAFF DATA
m1=[50 49 48 47 46 42 41 35 32 27 20 17 9 8 7 6 3 2 1];
n1=[22 7 3 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 3];
x1=[6 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0];
m2=[100 50 49 48 46 45 42 41 40 36 35 22 13 10 6 5];
n2=[1 12 2 3 1 1 1 2 1 1 1 1 1 1 1 1];
x2=[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0];
M1=19;
M2=16;
%Plot of posteriors
for p1 = 0.000001:0.000001:0.015
    A1 = prod((((1-(1-p1).^m1)).^x1).*(((1-p1).^m1).^(n1-x1)));
    B1 = sqrt(sum((m1.^2).*n1.*((1-p1).^(m1-2))./(1-(1-p1).^m1)));
    f1 = A1*B1;
    F1 = [F1;[p1 f1]];
end
S1 = sum(0.000001*F1(:,2));
Fp1 = [F1(:,1),F1(:,2)/S1];
plot(Fp1(:,1),Fp1(:,2),'b','LineWidth',2)
title('Posterior distribution of \it{p}','Fontsize',12)
xlabel('\it{p}','Fontsize',12)
ylabel('posterior','Fontsize',12)
grid
hold on
F2 = [];
for p2 = 0.000001:0.000001:0.015
    A2 = prod((((1-(1-p2).^m2)).^x2).*(((1-p2).^m2).^(n2-x2)));
    B2 = sqrt(sum((m2.^2).*n2.*((1-p2).^(m2-2))./(1-(1-p2).^m2)));
    f2 = A2*B2;
    F2 = [F2;[p2 f2]];
end
S2 = sum(0.000001*F2(:,2));
```

```
Fp2 = [F2(:,1),F2(:,2)/S2];
plot(Fp2(:,1),Fp2(:,2),'g','LineWidth',2)
hold off
axis([0 0.01 0 700])
legend('Sample 1', 'Sample 2')
%SET M=M1 etc. FOR SAMPLE 1 AND M=M2 etc. FOR SAMPLE 2
M=M1;
m=m1;
n=n1;
x=x1;
F=[];
h=0.000001;
for p=0.000001:h:0.015
    A=prod((((1-(1-p).^m)).^x).*(((1-p).^m).^(n-x)));
    B=sqrt(sum((m.^2).*n.*((1-p).^(m-2))./(1-(1-p).^m)));
    f=A*B;
    F=[F;[p f]];
end
S=sum(h*F(:,2));
Fp=[F(:,1),F(:,2)/S];
sF=sum(Fp(:,2));
av=Fp(:,1)'*Fp(:,2)/sF;
T=find(Fp(:,2)==max(Fp(:,2)));
Mo=Fp(T,1);
CS=h*cumsum(Fp(:,2));
L=[];
LL=[];
for i=0.003:0.000001:0.047
    l1=max(find(CS<=i));
    l2=min(find(CS>=0.95+i));
    l=l2-l1;
    LL=[LL;[l1 l2]];
    L=[L;l];
end
Q=max(find(L==min(L)));
R=LL(Q,:);
Lo=Fp(R(1),1);
```

```
Up=Fp(R(2),1);
Le=Up-Lo;
%T=[LOWER LIMIT UPPER LIMIT MEAN MODE LENGHT]
T=[Lo Up av Mo Le]
```

E.3.2 MATLAB® Code for Simulation Studies

The programs given here is for $M = 2$, that is, two samples $n = [n_1\ n_2]$, with pool sizes $m = [m_1\ m_2]$.

There are three programs that must run in sequence. The results of the first two must be saved in .mat files. The first program calculates the probabilities of all combined outcomes, $x = [x_1\ x_2]$ for a given p.

The second program calculates the posterior density of p, its mean, mode, 95% HPD Interval and length of interval for each of the possible outcomes $x = [x_1\ x_2]$ used in program 1.

The third program combines the results of the first two for given p. For example, it adds up the probabilities of all outcomes $x = [x_1\ x_2]$ that yield a 95% HPD Interval that contains the given p. This gives the coverage probability. Also right and left non-coverage, expected mean, mode and length.

- Changing p: Change p in parts 1 and 3. The ranges of x_1 and x_2 must also be changed (in parts 1 and 2) to ensure that we include all reasonable values, so that the sum of the probabilities is close to 1, say 1.0000 to the nearest 4 decimals (S in part 1). It is impractical and unnecessary to try and run x from 0 to n.

- Changing n and m: Change the entries of n and m in parts 1 and 2. Again the ranges of x_1 and x_2 (in parts 1 and 2) must be checked and probably changed.

- Changing M: If we want M =3, we must add values to the entries of n and m and add another 'for' loop for x_3 in parts 1 and 2. Other changes in part 1 are that:

  ```
  x = [x1 x2 x3];
  S = sum(Px(:,4));
  ```

  The same for part 2. Remember that RESULT in part 2 now has an extra column. So in part 3 the column numbers of RESULT must be changed. Similarly for $M > 3$. For $M = 1$, just do the opposite. For $M > 3$ the running time would get very high, especially if the ranges of the x's have to be relatively wide.

```
%PART 1:
%BINOMIAL PROBABILITIES FOR MULTIPLE SAMPLES
clear;
clc;
```

```
%Change M, n, m and the number of x's as required
M=2;
n=[100 50];
m=[5 10];
%Choose p
p=0.002;
t=1-(1-p).^m;
Px=[];
for x1=0:7
    for x2=0:7
        x=[x1 x2];
        lnK=gammaln(n+1)-gammaln(x+1)-gammaln(n-x+1);
        px=exp(lnK).*(t.^x).*((1-t).^(n-x));
        ppx=prod(px);
        Px=[Px;[x ppx]];
    end
end
S=sum(Px(:,3))
%Choose the upper values for x1 and x2 as small as possible
%but so that S = 1. It depends on p. For example, if p = 0.001
%we can use 5 and 5, but if p = 0.01, we must have at least
%17 and 14.
%Now save the result.
save C:/- - - -/prob2002.mat Px

%PART 2:
%HPD INTERVAL OF p FOR MULTIPLE SAMPLES M FOR ALL X
clear;
clc;
%Change M, n, m and the number of x's as required
M=2;
m=[5 10];
n=[100 50];
RESULT=[];
Px=[];
%Use the same values for x1 and x2 as in onesamplecoverage1
```

```
for x1=0:7

    for x2=0:7

        x=[x1 x2];

        F=[];

    %If sum(x)=0 we need more accurate and slightly different calculations

        if sum(x)==0

           h=0.0000001;

           for p=0.00000001:h:0.002

               A=prod((((1-(1-p).^m)).^x).*(((1-p).^m).^(n-x)));

               B=sqrt(sum((m.^2).*n.*((1-p).^(m-2))./(1-(1-p).^m)));

               f=A*B;

               F=[F;[p f]];

           end

        else

           h=0.00001;

           for p=0.00001:h:0.035

               A=prod((((1-(1-p).^m)).^x).*(((1-p).^m).^(n-x)));

               B=sqrt(sum((m.^2).*n.*((1-p).^(m-2))./(1-(1-p).^m)));

               f=A*B;

               F=[F;[p f]];

           end

        end

        S=sum(h*F(:,2));

        Fp=[F(:,1),F(:,2)/S];

%Check the plots to see if the range used for p is enough.

%plot(Fp(:,1),Fp(:,2))

%grid

%pause

       sF=sum(Fp(:,2));

        av=Fp(:,1)'*Fp(:,2)/sF;

        T=find(Fp(:,2)==max(Fp(:,2)));

        Mo=Fp(T,1);

        CS=h*cumsum(Fp(:,2));

        L=[];
```

```
        LL=[];
        if sum(x)==0;
           l2=min(find(CS>=0.95));
           Lo=0; Up=Fp(l2);
           Le=Up;
           Mo=0;
        else
        for i=0.01:0.0001:0.04
             l1=max(find(CS<=i));
             l2=min(find(CS>=0.95+i));
             l=l2-l1;
             LL=[LL;[l1 l2]];
             L=[L;l];
        end
        Q=max(find(L==min(L)));
        R=LL(Q,:);
        Lo=Fp(R(1),1);
        Up=Fp(R(2),1);
        Le=Up-Lo;
    end
    T=[Lo Up av Mo Le];
    RESULT=[RESULT;[x T]];
  end
end
%RESULT stores [x1 x2 Lowerlimit Upperlimit Mean Mode Intervallength]
%Now save the results
save C:/- - - -/R2002.mat RESULT

%PART3:
%CALCULATE COVERAGE PROBABILITIES FOR GIVEN P
%Using the saved results from onesamplecoverage1 and 2
clear;
clc;
load R2002;
load prob2002;
%Below put in the right column number, depending on the number of x's.
```

```
lo=RESULT(:,3);
hi=RESULT(:,4);
ml=RESULT(:,7);
me=RESULT(:,5);
mo=RESULT(:,6);
whd=Px(:,3);
p=0.002;
q=find(lo<=p & hi>=p);
prob=sum(whd(q));
l=find(lo>p);
r=find(hi<p);
right=sum(whd(r));
left=1-prob-right;
sym=(left-right)/(right+left);
ML=whd'*ml;
ME=whd'*me;
MO=whd'*mo;
A=[p prob left right sym ML ME MO]
```

The following code was used for the third simulation study, to determine the coverage and Bayes factor.

```
%PART 1 - Coverage
%SIMULATE USING BIGGERSTAFF SAMPLE SIZES - INDIVIDUAL SAMPLES
clear;
clc
p1=0.004;
p2=0.001;
m1=[50 49 48 47 46 42 41 35 32 27 20 17 9 8 7 6 3 2 1];
n1=[22 7 3 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 3];
m2=[100 50 49 48 46 45 42 41 40 36 35 22 13 10 6 5];
n2=[1 12 2 3 1 1 1 2 1 1 1 1 1 1 1 1];
M1=19;
M2=16;
t1=(1-p1).^m1;
t2=(1-p2).^m2;
%SET M=M1 etc. FOR SAMPLE 1 AND M=M2 etc. FOR SAMPLE 2
M=M2;
m=m2;
```

```
n=n2;
pp=p2;
t=t2;
W=[];
c=10000;
for j=1:c
X=[];
for i=1:length(m)

    x=0:n(i);
    fx=(gamma(n(i)+1)./gamma(x+1)./gamma(n(i)-x+1)).*((1-t(i)).^x).*(t(i)
        .^(n(i)-x));
    Fx=cumsum(fx);
    r=rand(1,1);
    I=min(find(Fx>=r));
    x1=x(I);
    X=[X x1];

end
y=sum(X);
%break
x=X;
F=[];
h=0.0001;
for p=0.00001:h:0.015

    A=prod((((1-(1-p).^m)).^x).*(((1-p).^m).^(n-x)));
    B=sqrt(sum((m.^2).*n.*((1-p).^(m-2))./(1-(1-p).^m)));
    f=A*B;
    F=[F;[p f]];

end
S=sum(h*F(:,2));
Fp=[F(:,1),F(:,2)/S];
%plot(Fp(:,1),Fp(:,2))
%grid
%pause
sF=sum(Fp(:,2));
av=Fp(:,1)'*Fp(:,2)/sF;
T=find(Fp(:,2)==max(Fp(:,2)));
```

```
Mo=Fp(T,1);
CS=h*cumsum(Fp(:,2));
Me=F(min(find(CS>=0.5)),1);
L=[];
LL=[];
if y==0 Lo=0;

    l2=min(find(CS>=0.95));
    Up=Fp(l2,1);
    Le=Up;

else
for i=0.004:0.00001:0.03

    l1=min(find(CS>=i));
    l2=min(find(CS>=0.95+i));
    l=l2-l1;
    LL=[LL;[l1 l2]]; L=[L;l];

end
Q=max(find(L==min(L)));
R=LL(Q,:);
Lo=Fp(R(1),1);
Up=Fp(R(2),1);
Le=Up-Lo;
end
%T=[LOWER LIMIT UPPER LIMIT MEAN MODE MEDIAN LENGHT]
T=[Lo Up av Mo Me Le];
W=[W;[y T]];
end
J=find(W(:,2)<=pp & W(:,3)>=pp);
COVER=length(J)/c
J1=find(W(:,2)>pp);
J2=find(W(:,3)<pp);
LEFT=length(J1)/c;
RIGHT=length(J2)/c;
W22=W;

%PART 2 - Bayes factor
%BIGGERSTAFF SIMULATIONS OF BAYES FACTOR %NUMERICAL INTEGRATION
```

```
%NO CONSTANTS
%BETA(0.5,0.5) PRIOR ON P
clear;
clc;
p1=0.004;
p2=0.001;
m1=[50 49 48 47 46 42 41 35 32 27 20 17 9 8 7 6 3 2 1];
n1=[22 7 3 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 3];
m2=[100 50 49 48 46 45 42 41 40 36 35 22 13 10 6 5];
n2=[1 12 2 3 1 1 1 2 1 1 1 1 1 1 1 1];
M1=19;
M2=16;
t1=(1-p1).^m1; t2=(1-p2).^m2;
W=[];
c=10000;
for j=1:c
x1=[];
for i=1:length(m1)

    x=0:n1(i);
    fx1=(gamma(n1(i)+1)./gamma(x+1)./gamma(n1(i)-x+1)).*((1-t1(i)).^x).*(t1(i)
         .^(n1(i)-x));
    Fx1=cumsum(fx1);
    r1=rand(1,1);
    I1=min(find(Fx1>=r1));
    x11=x(I1);
    x1=[x1 x11];

end
y1=sum(x1);
x2=[];
for i=1:length(m2)

    x=0:n2(i);
    fx2=(gamma(n2(i)+1)./gamma(x+1)./gamma(n2(i)-x+1)).*((1-t2(i)).^x).*(t2(i)
         .^(n2(i)-x));
    Fx2=cumsum(fx2);
    r2=rand(1,1);
    I2=min(find(Fx2>=r2));
```

```
    x22=x(I2);
    x2=[x2 x22];
end
y2=sum(x2);
%UNDER M0
m=[m1 m2];
n=[n1 n2];
x=[x1 x2];
h=0.0001;
L0=[];
for p=0.0001:h:0.012

    lnL0=-0.5*(log(p)+log(1-p))+sum(x.*log((1-(1-p).^m)))+log(1-p).*sum(m.*(n-x));
    L0=[L0;exp(lnL0)];
end
p=0.00001:h:0.012;
%plot(p,L0)
%grid
%pause
f0=h*sum(L0);
%UNDER M1
f1=0;
for p1=0.0001:h:0.012
    for p2=0.0001:h:0.012
        A1=-0.5*(log(p1)+log(1-p1))+sum(x1.*log((1-(1-p1).^m1)))+log(1-p1)
             .*sum(m1.*(n1-x1));
        A2=-0.5*(log(p2)+log(1-p2))+sum(x2.*log((1-(1-p2).^m2)))+log(1-p2)
             .*sum(m2.*(n2-x2));
        lnL1=A1+A2;
        f1=f1+h*h*exp(lnL1);
   end
end
B01=f0/pi/f1;
P=(1+1./B01).^(-1);
W=[W;[y1 y2 B01 P]];
end
```

Appendix F - Bayesian Process Control and Tolerance Intervals

F.1 MATLAB® Code for the *p* - chart

MATLAB® code for the *p* - chart

```
%Binomial control chart
%Frequentist method
clc;
clear;
m = 30;
n = 40;
T = m*n;
p0 = 0.5;
p1 = 0.5;
u = 0:T;
p_hat = u/T;
%freq
nUCL = n*p_hat + 3*sqrt(n.*p_hat.*(1-p_hat));
nLCL = n*p_hat - 3*sqrt(n.*p_hat.*(1-p_hat));
c = fix(nLCL);
c(c==-1) = 0;
d = min(floor(nUCL),n);
neg = (nLCL(:) < 0);
beta = zeros(length(d),1);
beta(neg) = binocdf(d(neg),n,p0);
beta(~neg) = binocdf(d(~neg),n,p0)-binocdf(c(~neg),n,p0);
CFAR = 1-beta;
pU_u = binopdf(u,T,p0);
```

```
PU_U = pU_u;
FAR1 = CFAR'.*PU_U;
ICARL = PU_U./CFAR';
matrix = [u; p_hat; nLCL; nUCL; c; d; beta'; CFAR'; PU_U; FAR1; ICARL]';
FAR = sum(FAR1(:))
ARL0 = sum(ICARL(:))
```

```
%Binomial control chart
%Bayes method
clc;
clear;
m = 30;
n = 40;
T = m*n;
p0 = 0.5;
p1 = 0.5;
u = 0:T;
alpha = u + 0.5;
beta = m*n - u + 0.5;
aveBB = (n.*alpha)./(alpha+beta);
varianceBB = ((n.*alpha.*beta).*(alpha+beta+n))
              ./(((alpha+beta).^2).*(alpha+beta+1));
nUCL = aveBB + 3*sqrt(varianceBB);
nLCL = aveBB - 3*sqrt(varianceBB);
c = fix(nLCL);
c(c==-1) = 0;
c(c==-2) = 0;
d = min(floor(nUCL),n);
neg = (nLCL(:) < 0);
beta = zeros(length(d),1);
beta(neg) = binocdf(d(neg),n,p0);
beta(~neg) = binocdf(d(~neg),n,p0)-binocdf(c(~neg),n,p0);
CFAR = 1-beta;
pU_u = binopdf(u,T,p0);
PU_U = pU_u;
FAR1 = CFAR'.*PU_U;
```

```
ICARL = PU_U./CFAR';
matrix = [u; nLCL; nUCL; c; d; beta'; CFAR'; PU_U; FAR1; ICARL]';
FAR = sum(FAR1(:))
ARL0 = sum(ICARL(:))
```

F.2 MATLAB® Code for the *c* - chart

MATLAB® code for the *c* - chart

```
%Poisson control chart simulation
clc;
clear;
l = [1 2 4 6 8 10 20 50];
MMRUN = [];
MMPHI = [];
MMMED = [];
MMPHIM = [];
MMRUNb = [];
MMPHIb = [];
MMMEDb = [];
MMPHIMb = [];
for j = 1:length(l)
    lamb = l(j);
    m = 5; %change value of m
    Mrun = [];
    PHI = [];
    Mrunb = [];
    PHIb = [];
    for i = 1:10000
        z = poissrnd(lamb,m,1);
        x = sum(z);
        %frequentists limits
        c = x/m;
        lo = c- 3*sqrt(c);
        up = c + 3*sqrt(c);
        if lo <= 0;
            lo = 0;
```

```
    end
    if  up == floor(up)
       up = up - 1;
    else
       up = up;
    end
    TL = floor(lo);
    TU = floor(up);
    %Bayes limits
    ave = (x + 0.5)/(m);
    var = ((x + 0.5)*(m+1))/(m^2);
    lob = ave -3*sqrt(var);
    upb = ave +3*sqrt(var);
    if lob <= 0;
       lob = 0;
    end
    if  upb == floor(upb)
       upb = upb - 1;
    else
       upb = upb;
    end
    TLb = floor(lob);
    TUb = floor(upb);
    %Freq
    t = TL+1:TU;
    phi = 1-sum(poisspdf(t,lamb));
    PHI = [PHI;phi];
    meanrun = 1/phi;
    Mrun = [Mrun;meanrun];
    %Bayes
    tb = TLb+1:TUb;
    phib = 1-sum(poisspdf(tb,lamb));
    PHIb = [PHIb;phib];
    meanrunb = 1/phib;
    Mrunb = [Mrunb;meanrunb];
end
```

```
    %Freq
    MRUN = mean(Mrun);
    MMED = median(Mrun);
    MPHI = mean(PHI);
    MPHIM = median(PHI);
    MMRUN = [MMRUN;MRUN]; %average run length
    MMPHI = [MMPHI;MPHI]; %false alarm rate (FAR)
    MMMED = [MMMED;MMED]; %median run length
    MMPHIM = [MMPHIM;MPHIM];
    %Bayes
    MRUNb = mean(Mrunb);
    MMEDb = median(Mrunb);
    MPHIb = mean(PHIb);
    MPHIMb = median(PHIb);
    MMRUNb = [MMRUNb;MRUNb];
    MMPHIb = [MMPHIb;MPHIb];
    MMMEDb = [MMMEDb;MMEDb];
    MMPHIMb = [MMPHIMb;MPHIMb];

end
%results = [Freq; Bayes]
results = [MMRUN MMMED MMPHI MMPHIM;
           MMRUNb MMMEDb MMPHIb MMPHIMb];
```

F.3 MATLAB® Code for Tolerance Intervals

MATLAB® code for the Binomial Tolerance Intervals

```
%TOLERANCE INTERVALS FOR BINOMIAL
clc;
clear;
LIMITS=[];
N=10;
for x=0:N

    x
    X95=[];
    n=10000;
    for i=1:n
```

```
        p=betarnd(x+1,N-x+1);
        y=0:N;
        z=binocdf(y,N,p);
        x95=min(find(z>=0.95))-1;
        X95=[X95;x95];
    end
    H=[min(X95):max(X95)];
    M=hist(X95,H);
    hist(X95,H)
    grid
    cN=cumsum(M);
    I1=min(find(cN>=0.025*n));
    I2=min(find(cN>=0.975*n));
    I=min(find(cN>=0.95*n));
    lo=H(I1);
    up=H(I2);
    Up=H(I);
    cover=length(find(X95>=lo & X95<=up))/n;
    cover2=length(find(X95<=Up))/n;
    LIMITS=[LIMITS;[x lo up cover Up cover2]];
end
%Using the results obtained in the code above and let LIMITS = ZZ9,
 to obtain the coverage
N=10;
x=[0:N];
COVER=[];
for p=0.01:0.1:0.99
    cx=binocdf(x,N,p);
    x95=min(find(cx>=0.95))-1;
    prob=cx(x95+1);
    J=find(ZZ9(:,2)<=x95 & ZZ9(:,3)>=x95);
    J2=find(ZZ9(:,5)>=x95);
    px=binopdf(x,N,p);
    cover1=sum(px(J));
    cover2=sum(px(J2));
    COVER=[COVER;[p x95 prob cover1 cover2]];
```

```
end
```

MATLAB® code for the Poisson Tolerance Intervals

```
%TOLERANCE INTERVALS
clc;
clear;
LIMITS=[];
x=472;
m=24;
%for x=51:100
% x
X95=[];
n=10000;
for i=1:n

    lam=gamrnd(x+0.5,1/m);
    y=0:100;
    z=poisscdf(y,lam);
    x95=min(find(z>=0.95))-1;
    X95=[X95;x95];

end
H=[min(X95):max(X95)];
N=hist(X95,H);
hist(X95,H)
grid
cN=cumsum(N);
I1=min(find(cN>=0.025*n));
I2=min(find(cN>=0.975*n));
%I=min(find(cN>=0.95*n));
lo=H(I1);
up=H(I2);
%Lo=H(I);
cover=length(find(X95>=lo & X95<=up))/n;
LIMITS=[LIMITS;[x lo up cover]];
%end
%COVERAGE OF TOLERANCE LIMIT INTERVAL
clc;
```

```
clear;
%0.9626 at 10
%0.9541 at 5
lam=5;
true=10;
y=0:100;
LIMITS=[];
for x= 0:20

    x
    px=poisspdf(x,lam);
    X95=[];
    n=10000;
    for i=1:n

        lam2=gamrnd(x+0.5,1);
        z=poisscdf(y,lam2);
        x95=min(find(z>=0.95))-1;
        X95=[X95;x95];

    end
    H=[min(X95):max(X95)];
    N=hist(X95,H);
    cN=cumsum(N);
    I1=min(find(cN>=0.025*n));
    I2=min(find(cN>=0.975*n));
    lo=H(I1);
    up=H(I2);
    LIMITS=[LIMITS;[x px lo up]];

end
J=find(LIMITS(:,3)<=true & LIMITS(:,4)>=true);
COVER=sum(LIMITS(J,2))
%Using the results obtained in the code above and let LIMITS = Z or Z2,
 to obtain the coverage
x=[0:100];
COVER=[];
for lam=1:0.5:15
    cx=poisscdf(x,lam);
    x95=min(find(cx>=0.95))-1;
```

```
        p=cx(x95+1);
        %J=find(Z(:,2)<=x95 & Z(:,3)>=x95);
        J=find(Z2(:,2)>=x95);
        px=poisspdf(x,lam);
        cover=sum(px(J));
        COVER=[COVER;[lam x95 p cover]];
    end
```

Abstract

This thesis focuses on objective Bayesian statistics, by evaluating a number of noninformative priors. Choosing the prior distribution is the key to Bayesian inference. The probability matching prior for the product of different powers of k binomial parameters is derived in Chapter 2. In the case of two and three independently distributed binomial variables, the Jeffreys, uniform and probability matching priors for the product of the parameters are compared. This research is an extension of the work by Kim (2006), who derived the probability matching prior for the product of k independent Poisson rates. In Chapter 3 we derive the probability matching prior for a linear combination of binomial parameters. The construction of Bayesian credible intervals for the difference of two independent binomial parameters is discussed.

The probability matching prior for the product of different powers of k Poisson rates is derived in Chapter 4. This is achieved by using the differential equation procedure of Datta & Ghosh (1995). The reference prior for the ratio of two Poisson rates is also obtained. Simulation studies are done to compare different methods for constructing Bayesian credible intervals. It seems that if one is interested in making Bayesian inference on the product of different powers of k Poisson rates, the probability matching prior is the best. On the other hand, if we want to obtain point estimates, credibility intervals or do hypothesis testing for the ratio of two Poisson rates, the uniform prior should be used.

In Chapter 5 the probability matching prior for a linear contrast of Poisson parameters is derived, this prior is extended in such a way that it is also the probability matching prior for the average of Poisson parameters. This research is an extension of the work done by Stamey & Hamilton (2006). A comparison is made between the confidence intervals obtained by Stamey & Hamilton (2006) and the intervals derived by us when using the Jeffreys and probability matching priors. A weighted Monte Carlo method is used for the computation of the Bayesian credible intervals, in the case of the probability matching prior. In the last section of this chapter hypothesis testing for two means is considered. The power and size of the test, using Bayesian methods, are compared to tests used by Krishnamoorthy & Thomson (2004). For the Bayesian methods the Jeffreys prior, probability matching prior and two other priors are used.

Bayesian estimation for binomial rates from pooled samples are considered in Chapter 6, where the Jeffreys prior is used. Bayesian credibility intervals for a single proportion and the difference of two binomial proportions estimated from pooled samples are considered. The results are compared

to those from other methods. In Chapters 7 and 8, Bayesian process control for the p - chart and the c - chart are considered. The Jeffreys prior is used for the Bayesian methods. Control chart limits, average run lengths and false alarm rates are determined. The results from the Bayesian method are compared to the results obtained from the classical (frequentist) method. Bayesian tolerance intervals for the binomial and Poisson distributions are studied in Chapter 9, where the Jeffreys prior is used.

Keywords: Bayesian intervals; Binomial distribution; c - chart; Coverage probabilities (rates); Jeffreys prior; p - chart; Poisson distribution; Power of test; Size of test; Probability matching prior; Reference prior; Uniform prior; Weighted Monte Carlo method.

Opsomming

Hierdie tesis sal op objektiewe Bayes-statistiek fokus, deur 'n aantal nie-informatiewe priors te evalueer. Die waarskynlikheidsafparings prior sal in Hoofstuk 2 afgelei word vir die produk van verskillende magte van k binomiaal parameters. In die geval van twee en drie onafhanklik verdeelde binomiaalveranderlikes, sal die Jeffreys, uniforme en waarskynlikheidsafparings priors vir die produk van die parameters vergelyk word. Hierdie navorsing is 'n verlenging van die werk deur Kim (2006), wat die waarskynlikheidsafparings prior vir die produk van k onafhanklike Poisson parameters afgelei het. In Hoofstuk 3 sal ons die waarskynlikheidsafparings prior vir 'n lineêre kombinasie van binomiaal parameters aflei. Die samestelling van Bayes-vertrouensintervalle vir die verskil van twee onafhanklike binomiaal parameters word bespreek.

Die waarskynlikheidsafparings prior vir die produk van verskillende magte van k Poisson parameters word in Hoofstuk 4 afgelei. Dit word gedoen deur die differensiaalvergelyking prosedure van Datta & Ghosh (1995) te gebruik. Die verwysings prior word ook afgelei vir die verhouding van twee Poisson parameters. Simulasie studies word gedoen om die verskillende metodes vir die samestelling van Bayes-vertrouensintervalle te vergelyk. Dit blyk dat indien 'n mens geïnteresseerd is in Bayes-inferensie vir die produk van verskillende magte van k Poisson parameters, die waarskynlikheidsafparings prior die beste is. Aan die anderkant, indien ons puntberamers, geloofwaardigheidsintervalle of hipotesetoetsing wil doen vir die verhouding van twee Poisson parameters, moet die uniforme prior gebruik word.

In Hoofstuk 5 word die waarskynlikheidsafparings prior vir 'n lineêre kontras van Poisson parameters afgelei, hierdie prior word uitgebrei sodat dit ook die waarskynlikheidsafparings prior is vir die gemiddelde van Poisson parameters. Hierdie navorsing is 'n uitbreiding van die werk gedoen deur Stamey & Hamilton (2006). 'n Vergelyking word gemaak tussen die vertrouensintervalle verkry deur Stamey & Hamilton (2006) en die intervalle afgelei deur ons waneer die Jeffreys en waarskynlikheidsafparings priors gebruik word. 'n Geweegde Monte Carlo-metode word gebruik vir die berekening van die Bayes-vertrouensintervalle, in die geval van die waarskynlikheidsafparings prior. In die laaste afdeling van die hoofstuk word hipotesetoetsing vir twee gemiddeldes oorweeg. Die onderskeidingsvermoë en toetsgrootte, waar Bayes metodes gebruik word, sal vergelyk word met die toetse gebruik deur Krishnamoorthy & Thomson (2004). Die Jeffreys prior, waarskynlikheidsafparings prior en twee ander priors word vir die Bayes metodes gebruik.

Bayes-beraming vir binomiaal parameters van saamgevoegde steekproewe sal in Hoofstuk 6 oorweeg word, die Jeffreys prior sal hier gebruik word. Bayes-geloofwaardigheidsintervalle vir ’n enkele proporsie en die verskil tussen twee binomiaal proporsies beraam vanaf saamgevoegde steekproewe sal oorweeg word. Die resultate sal vergelyk word met die van ander metodes. In Hoofstukke 7 en 8, word Bayes proseskontrole vir die p - kaart en c - kaart bestudeer. Die Jeffreys prior sal gebruik word vir die Bayes metodes. Kontrolekaart grense, gemiddelde lopielengte en vals alarm koerse sal bereken word. Die resultate van die Bayes metode sal vergelyk word met die resultate van die klassieke (frekwentistiese) metode. Bayes toleransie-intervalle vir die binomiaal en Poisson verdelings word in Hoofstuk 9 bestudeer, waar die Jeffreys prior gebruik word.

Sleutelwoorde: Bayes intervalle; Binomiaalverdeling; c - kaart; Oordekkingswaarskynlikhede (koerse); Jeffreys prior; p - kaart; Poisson-verdeling; Onderskeidingsvermoë van toets; Toetsgrootte; Waarskynlikheidsafparings prior; Verwysings prior; Uniforme prior; Geweegde Monte Carlo-metodes.