A thesis submitted in fullfilment of the degree of Doctor of Philosophy in the Faculty of Natural and Agricultural Sciences, Department of Mathematical Statistics and Actuarial Science

Bayesian Control Charts Based on Predictive Distributions

by Ruaan van Zyl

January 2016

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
NATURAL AND AGRICULTURAL SCIENCES
NATUUR- EN LANDBOUWETENSKAPPE
MATHEMATICAL STATISTICS AND ACTUARIAL SCIENCE
WISKUNDIGE STATISTIEK EN AKTUARIËLE WETENSKAP

Promoter: Prof A.J. van der Merwe

Contents

List of Tables

List of Figures

Declaration

I hereby declare that this thesis submitted for the Philosophae Doctor (PhD) degree at the University of the Free State is my own original work, and has not been submitted at another university/faculty for any reward, recognition or degree. Every effort has been made to clearly reference the contributions of others within this thesis. I furthermore cede copyright of this thesis in favor of the University of the Free State.

The thesis was completed under the guidance of Professor A.J. van der Merwe from the University of the Free State, Bloemfontein, South Africa.

Acknowledgments

I would like to thank my promoter, Prof. van der Merwe, for all the support given in the three years worked on this study.

I would also like to thank Prof. Groenewald for the programmatic validation of results presented in this thesis.

I would then also like to thank my wife, Elzanne van Zyl, for the support given to me during this time.

Last but not least, I would like to thank God for providing me with the means to undertake this work.

Abstract/Opsomming

Abstract

Control charts are statistical process control (SPC) tools that are widely used in the monitoring of processes, specifically taking into account stability and dispersion. Control charts signal when a significant change in the process being studied is observed. This signal can then be investigated to identify issues and to find solutions. It is generally accepted that SPC are implemented in two phases, Phase I and Phase II. In Phase I the primary interest is assessing process stability, often trying to bring the process in control by locating and eliminating any assignable causes, estimating any unknown parameters and setting up the control charts. After that the process move on to Phase II where the control limits obtained in Phase I are used for online process monitoring based on new samples of data. This thesis concentrate mainly on implementing a Bayesian approach to monitoring processes using SPC. This is done by providing an overview of some non-informative priors and then to specifically derive the reference and probability-matching priors for the common coefficient of variation, standardized mean and tolerance limits for a normal population. Using the Bayesian approach described in this thesis SPC is performed, including derivations of control limits in Phase I and monitoring by the use of run-lengths and average run-lengths in Phase II for the common coefficient of variation, standardized mean, variance and generalized variance, tolerance limits for normal populations, two-parameter exponential distribution, piecewise exponential model and capability indices. Results obtained using the Bayesian approach are compared to frequentist results.

Keys: Statistical Process Control, Run-length, Control Charts, Non-informative Priors, Reference Priors, Probability-matching Priors, Bayes.

Opsomming

Beheer kaarte is statistiese beheer prosesse wat gebruik word om prosesse te monitor, deur veral na stabiliteit en verspreiding te kyk. Beheer kaarte gee 'n waarskuwing-sein as daar 'n bedeidende verandering in die proses wat bestudeer word opgemerk word. Hierdie sein kan dan ondersoek word om probleme te identifiseer en op te los. Dit word oor die algemeen aanvaar dat statististiese beheer prosesse in twee fases geimplementeer word. In Fase I word die stabiliteit van die proses geasseseer en die proses word in beheer gebring deur redes vir probleme te identifseer en op te los, onbekende parameters word bepaal en die beheer kaarte word opgestel. In Fase II word die beheer limiete wat in Fase I bereken is gebruik deur 'n voortdurende proses te monitor met nuwe data. Hierdie proefskrif handel grotendeels oor die implementeering van 'n Bayesiaanse metode om statistiese beheer toe te pas. Dit word gedoen deur nie-objektiewe priors te bereken, meer spesifiek die verwysingsprior en die waarskynlikheidsooreenstemmende prior te bereken vir die algemene koeffisient van variasie, die gestandardiseerde gemiddelde en toleransie limiete vir 'n normale populasie. Deur die gebruik van die Bayes metode uiteen gesit in hierdie proefskrif, insluitend die berekeninge van beheer limiete in Fase I en die monitering deur gebruik te maak van proses-lengte en gemidelde proses-lengte in Fase II vir die algemene koeffisient van variasie, gestandardiseerde gemiddelde, variansie en algemene variansie, toleransie limiete vir die normale populasie, twee-parameter eksponensiele verdeling, stuksgewysde eksponensiele model en vermoë indekse. Resultate deur die Bayes proses is dan vergelyk met resultate uit die klassieke statistiek.

1. General Introduction

1.1. Introduction

Control charts are statistical process control (SPC) tools that are widely used in the monitoring of processes, specifically taking into account stability and dispersion. Control charts signal when a significant change in the process being studied is observed. This signal can then be investigated to identify issues that can then be used to find a solution in order to reduce variation in the process and improve process stabilization. In general, a control chart is a two dimensional graph consisting of the values of a plotting (charting) statistic including the associated control limits namely the upper control limit (UCL) and the lower control limit (LCL) When a charting statistic plots above the UCL or below the LCL it is said that the control chart has signaled and the process is considered to be out of control.

It is a generally accepted notion that statistical process control are implemented in two Phases: Phase I (also called the retrospective phase) and Phase II (also called the prospective or monitoring phase). In Phase I the primary interest is assessing process stability, often trying to bring a process in control by locating and eliminating any assignable causes, estimating any unknown parameters and setting up the control charts. Once Phase I has been completed, the process moves on to Phase II where the control limits obtained in Phase I are used for online process monitoring based on new samples of data.

In this thesis we will mainly concentrate on implementing a Bayesian approach to monitor processes using statistical process control charts.

Bayarri and 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.

In order to implement a Bayesian approach, two priors will mainly be used:

1. The reference prior; and
2. The probability-matching prior.

1.1.1. The Reference Prior

In general, the derivation of the reference prior depends on the ordering of the parameters and how the parameter vector is divided into sub-vectors. As mentioned by Pearn and Wu (2005) the reference prior maximizes the difference in information (entropy) about the parameter provided by the prior and posterior. In other words, the reference prior is derived in such a way that it provides as little information possible about the parameter of interest. The reference prior algorithm is relatively complicated and, as mentioned, the solution depends on the ordering of the parameters and how the parameter vector is partitioned into sub-vectors. In spite of these difficulties, there is growing evidence, mainly through examples that reference priors provide "sensible" answers from a Bayesian point of view and that frequentist properties of inference from reference posteriors are asymptotically "good". As in the case of the Jeffreys' prior, the reference prior is obtained from the Fisher information matrix. In the case of a scalar parameter, the reference prior is the Jeffreys' prior.

1.1.2. The Probability-matching Prior

The reference prior algorithm is but one way to obtain a useful non-informative prior. Another type of non-informative prior is the probability-matching prior. This

prior has good frequentist properties. Two reasons for using probability-matching priors are that they provide a method for constructing accurate frequentist intervals, and that they could be potentially useful for comparative purposes in a Bayesian analysis.

There are two methods for generating probability-matching priors due to Tibshirani (1989) and Datta and Ghosh (1995).

Tibshirani (1989) generated probability-matching priors by transforming the model parameters so that the parameter of interest is orthogonal to the other parameters. The prior distribution is then taken to be proportional to the square root of the upper left element of the information matrix in the new parametrization.

Datta and Ghosh (1995) provided a different solution to the problem of finding probability-matching priors. They derived the differential equation that 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.

According to Datta and Ghosh (1995) $p\left(\underline{\theta}\right)$ is a probability-matching prior for $\underline{\theta}$,the vector of unknown parameters, if the following differential equation is satisfied:

$$\sum_{\alpha=1}^{m+1} \frac{\partial}{\partial \theta_{\alpha}} \left\{ \Upsilon_{\alpha}\left(\underline{\theta}\right) p\left(\underline{\theta}\right) \right\} = 0$$

where

$$\Upsilon\left(\underline{\theta}\right) = \frac{F^{-1}\left(\underline{\theta}\right) \nabla_{t}\left(\underline{\theta}\right)}{\sqrt{\nabla_{t}^{\prime}\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right) \nabla_{t}\left(\underline{\theta}\right)}} = \left[\begin{array}{cccc} \Upsilon_{1}\left(\underline{\theta}\right) & \Upsilon_{2}\left(\underline{\theta}\right) & \cdots & \Upsilon_{m+1}\left(\underline{\theta}\right) \end{array} \right]^{\prime}$$

and

$$\nabla_{t}\left(\underline{\theta}\right) = \left[\begin{array}{cccc} \frac{\partial}{\partial \theta_{1}} t\left(\underline{\theta}\right) & \frac{\partial}{\partial \theta_{2}} t\left(\underline{\theta}\right) & \cdots & \frac{\partial}{\partial \theta_{m+1}} t\left(\underline{\theta}\right) \end{array} \right]^{\prime}.$$

$t\left(\underline{\theta}\right)$ is a function of $\underline{\theta}$ and $F^{-1}\left(\underline{\theta}\right)$ is the inverse of the Fisher information matrix.

The method of Datta and Ghosh (1995) to obtain a probability matching prior will be used in this thesis.

1.2. Objectives and Contributions

The main objectives and contributions of this thesis can be summarized as follows:

- To provide an overview of some non-informative priors;
- To derive the probability-matching priors and reference priors for the following cases:
  - Common coefficient of variation;
  - Standardized mean;
  - One-sided upper tolerance limit for a Normal population;
  - Piece-wise exponential model.
- To propose a Bayesian approach to statistical process control, including the derivations of control limits in Phase I and monitoring by use of run-lengths and average run-lengths in Phase II for the following cases:
  - Common coefficient of variation;
  - Standardized mean;
  - Variance and generalized variance;
  - Tolerance limits in the case of Normal populations;
  - Two parameter exponential distribution;
  - Piecewise exponential model;
  - Capability indices.
- To compare results obtained from the Bayesian approach to statistical process control to the frequentist results;
- To show that the joint posterior distribution for the common coefficient of variation and the standardized mean is a proper distribution;
- To provide guidelines to the practitioners in implementing the Bayesian control chart for the variance and generalized variance;

- To compare two models of the piecewise exponential model using Bayes factors;
- To perform Bayesian hypothesis testing for capability indices.

1.3. Thesis Outline

In **Chapter 2** the medical data analyzed by Kang, Lee, Seong, and Hawkins (2007) is used to apply a Bayesian procedure to obtain control limits for the coefficient of variation. Reference and probability-matching priors are derived for a common coefficient of variation across the range of sample values. By simulating the posterior predictive density function of a future coefficient of variation it is shown that the control limits are effectively identical to those obtained by Kang, Lee, Seong, and Hawkins (2007) for the specific dataset they used. This chapter illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, predictive intervals and run-lengths in the case of the coefficient of variation. A simulation study shows that the 95% Bayesian confidence intervals for the coefficient of variation has the correct frequentist coverage.

In **Chapter 3** the same medical data analyzed by Kang, Lee, Seong, and Hawkins (2007) are used to apply a Bayesian procedure for obtaining control limits for the standardized mean. Reference and probability-matching priors are derived for a common standardized mean across the range of sample values. By simulating the posterior predictive density function of a future standardized mean it is shown that the inverse of the control limits for the standardized mean are effectively identical to those calculated by Kang et al. (2007) for the coefficient of variation. This chapter also illustrates the flexibility and unique features of the Bayesian simulation method for obtaining the posterior predictive distribution of $\delta = \frac{\mu}{\sigma}$ (the population standardized mean), predictive intervals and run-lengths for the future sample standardized means. A simulation study shows that the 95% Bayesian confidence intervals for δ has the correct frequentist coverage.

In **Chapter 4** the results of Human, Chakraborti, and Smit (2010) are extended and Phase I control charts are derived for the generalized variance when the mean vector and covariance matrix of multivariate normally distributed data are unknown and estimated from m independent samples, each of size n. In Phase II predictive distributions based on a Bayesian approach are used to construct Shewart-type control limits for the variance and generalized variance. The posterior distribution

is obtained by combining the likelihood (the observed data in Phase I) and the uncertainty of the unknown parameters via the prior distribution. By using the posterior distribution the unconditional predictive density functions are derived.

In **Chapter 5** air-lead data analyzed by Krishnamoorthy and Mathew (2009) are used to apply a Bayesian procedure to obtain control limits for the upper one-sided tolerance limit. Reference and probability matching priors are derived for the pth quantile of a normal distribution. By simulating the predictive density of a future upper one-sided tolerance limit, run-lengths and average run-lengths are derived. In the second part of this chapter control limits are derived for one-sided tolerance limits for the distribution of the difference between two normal random variables.

In **Chapter 6** and **Chapter 7** data that are the mileages for some military personnel carriers that failed in service given by Grubbs (1971) and Krishnamoorthy and Mathew (2009) are used to apply a Bayesian procedure to obtain control limits for the location and scale parameters, as well as for a one-sided upper tolerance limit in the case of the two-parameter exponential distribution. An advantage of the upper tolerance limit is that it monitors the location and scale parameter at the same time. By using Jeffreys' non-informative prior, the predictive distributions of future maximum likelihood estimators of the location and scale parameters are derived analytically. The predictive distributions are used to determine the distribution of the run-length and expected run-length. These chapters illustrates the flexibility and unique features of the Bayesian simulation method. In Chapter 6 it is assumed that the location parameter $0 < \mu < \infty$ while in Chapter 7 it is assumed that $-\infty < \mu < \infty$ for the two-parameter exponential distribution.

In **Chapter 8** data that are failure data on load-haul-dump (LHD) machines given by Kumar and Klefsjö (1992) and reported in Hamada, Wilson, Reese, and Martz (2008, page 201) are used to apply a Bayesian procedure to obtain control limits and control charts for a future observation X_f for the piecewise exponential model. Two models are considered, one where all μ's (scale parameters) across multiple systems are considered to be the same and the other one where all systems have different μ's. The two models are compared using Bayes factors to determine the best suited model.

In **Chapter 9** process capability indices C_{pl}, C_{pu} and C_{pk} are defined. Bayesian methods are applied to piston ring data for automobile engines studied by Chou (1994) for four suppliers. The method proposed by Ganesh (2009) for multiple test-

ing are then applied using a Bayesian procedure to $C_{pl}, C_{pu},$ and C_{pk} to determine whether significant differences between the four suppliers exist. A Bayesian control chart for C_{pk} is also implemented

2. A Bayesian Control Chart For a Common Coefficient of Variation for the Normal Distribution

2.1. Introduction

The monitoring of variability is a vital part of modern statistical process control (SPC). Shewart control charts are widely used statistical process control tools for detecting changes in the quality of a process. In most settings where the process is under control the process have readings that have a constant mean (μ) and constant variance (σ^2). In such settings the $\bar{X}$ chart is usually used to monitor the mean, and the R and S control charts the variance of the process.

In practice there are some situations though where the mean is not a constant and the usual statistical process control reduces to the monitoring of the variability alone. As a further complication it sometimes happens that the variance of the process is a function of the mean. In these situations the usual R and S charts can also not be used.

The proposed remedy depends on the nature of the relationship between the mean and the variance of the process. One common relationship that we will look at is where the mean and standard deviation of the process is directly proportional so that the coefficient of variation

$$\gamma = \frac{\sigma}{\mu} \tag{2.1}$$

is a constant. According to Kang, Lee, Seong, and Hawkins (2007) this is often the case in medical research. Scientists at the Clinical Research Organization, Quin-

tiles, also confirmed that the coefficient of variation of drug concentrations is often constant or approximately constant over a reasonable interval of drug concentration measurements. By using frequentist methods Kang, Lee, Seong, and Hawkins (2007) developed a Shewart control chart, equivalent to the S chart, for monitoring the coefficient of variation using rational groups of observations. The chart is a time-ordered plot of the coefficient of variation for successive samples. It contains three lines:

- A center line;
- The upper control limit (UCL);
- The lower control limit (LCL).

By using the posterior predictive distribution, a Bayesian procedure will be developed to obtain control limits for a future sample coefficient of variation. These limits will be compared to the classical limits obtained by Kang, Lee, Seong, and Hawkins (2007).

2.2. Frequentist Methods

Assume that X_j $(j = 1, 2, \ldots, n)$ are independently, identically normally distributed with mean μ and variance σ^2. $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ is the sample mean and $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2$ is the sample variance. The sample coefficient of variation is defined as

$$W = \frac{S}{\bar{X}}$$

Kang, Lee, Seong, and Hawkins (2007) suggested a control chart for the sample coefficient of variation, similar to that of the $\bar{X}$, R and S charts. They proposed two methods in developing these charts:

1. The use of the non-central t distribution:

It can be noted that

$$T = \frac{\sqrt{n}\bar{X}}{S} = \sqrt{n}W^{-1}$$

follows a non-central t distribution with $(n-1)$ degrees of freedom and non-centrality parameter, $\frac{\sqrt{n}}{\gamma}$. The cumulative distribution function of the coefficient of variation can therefore be computed from the non-central t distribution.

2. By deriving a canonical form of the distribution of the coefficient of variation.

Any one of these two methods can be used for obtaining control charts. Kang, Lee, Seong, and Hawkins (2007) used the second method to obtain control limits for the coefficient of variation chart for a selection of values of n and γ. According to them the probability of exceeding these limits is $\frac{1}{740}$ on each side when the process is in control.

In what follows, a more general distribution (than the canonical form of Kang, Lee, Seong, and Hawkins (2007)) will be given for $W = \frac{S}{\bar{X}}$.

Theorem 2.1. *The density function of the coefficient of variation $W = \frac{S}{\bar{X}}$ is given by*

$$f\left(w|\gamma\right) = \begin{cases} \frac{A(w)}{\left(n+\tilde{f}w^2\right)^{\frac{\tilde{f}+1}{2}}} I_{\tilde{f}}\left(\frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{0.5}}\right) & , w \geq 0 \\ \\ \frac{(-1)^{\tilde{f}-1}A(w)}{\left(n+\tilde{f}w^2\right)^{\frac{\tilde{f}+1}{2}}} I_{\tilde{f}}\left(\frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{0.5}}\right) & , w < 0 \end{cases} \tag{2.2}$$

where $\gamma = \frac{\sigma}{\mu}$, $\tilde{f} = n-1$,

$$A\left(w|\gamma\right) = \frac{\tilde{f}^{\frac{\tilde{f}}{2}}\sqrt{n}w^{\tilde{f}-1}}{2^{\frac{1}{2}\left(\tilde{f}-2\right)}\Gamma\left(\frac{\tilde{f}}{2}\right)\sqrt{2\pi}} \exp\left\{-\frac{\frac{1}{2}\left(n\tilde{f}w^2\right)}{\gamma^2\left(n+\tilde{f}w^2\right)}\right\}$$

and

$$I_{\tilde{f}}\left(\frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{0.5}}\right) = \int_0^\infty q^{\tilde{f}} \exp\left\{-\frac{1}{2}\left[q - \frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right]^2\right\} dq$$

is the Airy function (Iglewicz (1967)).

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Using a Bayesian procedure this distribution will be used for prediction purposes.

2.2.1. The Data

The example used by Kang, Lee, Seong, and Hawkins (2007) was that of patients undergoing organ transplantation, for which Cyclosporine is administered. For patients undergoing immunosuppressive treatment, it is vital to control the amount of drug circulating in the body. For this reason frequent blood assays were taken to find the best drug stabilizing level for each patient. The dataset consist of $\ddot{m} = 105$ patients and the number of assays obtained for each patient is $n = 5$. By doing a regression test they confirmed that there is no evidence that the coefficient of variation depends on the mean which implies that the assumption of a constant coefficient of variation is appropriate. They used the root mean square estimator $\hat{\gamma} = \sqrt{\frac{1}{\ddot{m}} \sum_{i=1}^{\ddot{m}} w_i^2} = \sqrt{\frac{0.593515}{105}} = 0.0752$ to pool the samples for estimating γ. $w_i = \frac{s_i}{\bar{x}_i}$ is the sample coefficient of variation for the ith patient. $\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij}$ and $s_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$ where x_{ij} is the jth blood assay for the ith patient. By substituting $\hat{\gamma}$ for γ in the distribution of W and by calculating the lower and upper $\frac{1}{740}$ percentage points, they obtained a LCL = 0.01218 and UCL = 0.15957. The chart was then applied to a fresh data set of 35 samples from a different laboratory. The data used by Kang, Lee, Seong, and Hawkins (2007) is given in Table 2.1.

2.3. Bayesian Procedure

By assigning a prior distribution to the unknown parameters the uncertainty in the estimation of the unknown parameters can adequately be handled. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution of γ. By using the posterior distribution the predictive distribution of a future coefficient of variation can be obtained. The predictive distribution on the other hand can be used to obtain control limits and to determine the distribution of the run-length. Determination of reasonable non-informative priors is however not an easy task. Therefore, in the next section, reference and probability matching priors will be derived for a common coefficient of variation across the range of sample values.

Table 2.1.: Data used by Kang, Lee, Seong, and Hawkins (2007)

m	$\overline{X}$	$w_i \times 100$	m	$\overline{X}$	$w_i \times 100$	m	$\overline{X}$	$w_i \times 100$
1	31.7	12.4	36	120.3	5.8	71	361.4	8.3
2	37.7	15.3	37	143.7	5.6	72	361.5	13.4
3	40.6	9.1	38	148.6	5.5	73	361.8	6.1
4	50.5	4.6	39	149.1	3.1	74	374.6	5.8
5	52	10.5	40	149.9	2	75	376.3	2.8
6	57.6	6.2	41	151	4.4	76	382.3	5.8
7	58.3	6.6	42	153.6	6.6	77	401.7	7.3
8	58.9	8.4	43	172.2	7.2	78	415.2	15.1
9	61.2	8.1	44	179.8	7.9	79	428.8	4.5
10	64.3	7	45	185.3	7.6	80	442.1	9.9
11	64.5	8.8	46	192.1	5.3	81	450.1	7.4
12	65.6	4.1	47	193.8	5.9	82	496.5	4.8
13	68	3.7	48	195.1	11	83	499.7	10
14	71.8	6.2	49	195.2	5.1	84	504.6	8.4
15	72.1	8.4	50	195.4	9.4	85	523.1	5
16	78.4	6.8	51	196.4	5.6	86	531.7	8.5
17	78.4	4.6	52	199.6	6.8	87	556.4	11.8
18	79.5	5.7	53	204.4	3.7	88	571.4	5.9
19	83.2	10.5	54	207.8	12.4	89	584.1	8.3
20	85.1	4.8	55	219	7.6	90	597.6	4.2
21	85.6	5.4	56	222.9	4.8	91	606.2	8.2
22	86	10.1	57	225.1	5.7	92	609	9.7
23	87.3	7.9	58	227.6	6.5	93	635.4	5.6
24	89.1	10.3	59	240.5	3.8	94	672.2	7.2
25	95.4	6.2	60	241.1	8.4	95	695.9	2.7
26	101.9	4.8	61	252.2	8.3	96	696.4	10.6
27	105.4	5.6	62	262.2	5.8	97	721.3	9.8
28	107.2	2.2	63	277.9	8.7	98	752	4.2
29	108.2	3.3	64	278.3	6.2	99	769.5	9.7
30	112	8.7	65	303.4	8.8	100	772.7	9.6
31	112.3	5.7	66	309.7	3.9	101	791.6	2
32	113.5	9.4	67	323.9	4.1	102	799.9	11.4
33	114.3	3.5	68	328.7	4.1	103	948.6	5.2
34	116.8	6	69	341.2	6.5	104	971.8	11.1
35	117.8	5.7	70	347.3	4.9	105	991.2	8.8

2.4. Reference and Probability-Matching Priors for a Common Coefficient of Variation

As mentioned the Bayesian paradigm emerges as attractive in many types of statistical problems, also in the case of the coefficient of variation.

Prior distributions are needed to complete the Bayesian specification of the model. Determination of reasonable non-informative priors in multi-parameter problems is not easy; common non-informative priors, such as the Jeffreys' prior can have features that have an unexpectedly dramatic effect on the posterior.

Reference and probability-matching priors often lead to procedures with good frequentist properties while returning to the Bayesian flavor. The fact that the resulting Bayesian posterior intervals of the level $1-\alpha$ are also good frequentist intervals at the same level is a very desirable situation.

See also Bayarri and Berger (2004) and Severine, Mukerjee, and Ghosh (2002) for a general discussion.

2.4.1. The Reference Prior

In this section the reference prior of Berger and Bernardo (1992) will be derived for a common coefficient of variation, γ, across the range of sample values.

Berger, Liseo, and Wolpert (1999) derived the reference prior for the coefficient of variation in the case of a single sample. From the medical example given in Kang, Lee, Seong, and Hawkins (2007) it is clear that the standard deviation of measurements is approximately proportional to the mean; that is, the coefficient of variation is constant across the range of means, which is an indication that the a reference prior for a constant coefficient of variation should be derived.

Theorem 2.2. *Let* $X_{ij} \sim N(\mu_i, \sigma_i^2)$ *where* $i = 1, 2, ..., \ddot{m}$, $j = 1, 2, ..., n$, *and the coefficient of variation is* $\gamma = \frac{\sigma_1}{\mu_1} = \frac{\sigma_2}{\mu_2} = ... = \frac{\sigma_{\ddot{m}}}{\mu_{\ddot{m}}}$.

The reference prior for the ordering $\{\gamma; (\sigma_1^2, \sigma_2^2, ..., \sigma_{\ddot{m}}^2)\}$ *is given by*

$$p_R\left(\gamma, \sigma_1^2, \sigma_2^2, ..., \sigma_{\ddot{m}}^2\right) \propto \frac{1}{|\gamma| \sqrt{\gamma^2 + \frac{1}{2}}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2}$$

Note: The ordering $\{\gamma; (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2)\}$ means that the coefficient of variation is the most important parameter while the $\ddot{m}$ variance components are of equal importance, but not as important as γ. Also, if $\ddot{m} = 1$, the above equation simplifies to the reference prior obtained by Berger, Liseo, and Wolpert (1999).

Proof. The proof is given in the Mathematical Appendices to this chapter. □

2.4.2. Probability-matching Priors

The reference prior algorithm is but one way to obtain a useful non-informative prior. Another type of non-informative prior is the probability-matching prior. This prior has good frequentist properties.

As mentioned in the introduction $p(\underline{\theta})$ is a probability-matching prior for $\underline{\theta} = [\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2]'$ the vector of unknown parameters, if the following differential equation is satisfied:

$$\sum_{\alpha=1}^{\ddot{m}+1} \frac{\partial}{\partial \theta_\alpha} \{\Upsilon_\alpha(\underline{\theta})\, p(\underline{\theta})\} = 0$$

where

$$\Upsilon(\underline{\theta}) = \frac{F^{-1}(\underline{\theta})\, \nabla_t(\underline{\theta})}{\sqrt{\nabla_t'(\underline{\theta})\, F^{-1}(\underline{\theta})\, \nabla_t(\underline{\theta})}} = \left[\begin{array}{cccc} \Upsilon_1(\underline{\theta}) & \Upsilon_2(\underline{\theta}) & \cdots & \Upsilon_{\ddot{m}+1}(\underline{\theta}) \end{array} \right]'$$

and

$$\nabla_t(\underline{\theta}) = \left[\begin{array}{cccc} \frac{\partial}{\partial \theta_1} t(\underline{\theta}) & \frac{\partial}{\partial \theta_2} t(\underline{\theta}) & \cdots & \frac{\partial}{\partial \theta_{\ddot{m}+1}} t(\underline{\theta}) \end{array} \right]'.$$

$t(\underline{\theta})$ is a function of $\underline{\theta}$ and $F^{-1}(\underline{\theta})$ is the inverse of the Fisher information matrix.

Theorem 2.3. *The probability-matching prior for the coefficient of variation γ and the variance components is given by*

$$p_M\left(\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) \propto \frac{1}{|\gamma|\left(1+2\gamma^2\right)^{\frac{1}{2}}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2} \propto \frac{1}{|\gamma|\sqrt{\gamma^2+\frac{1}{2}}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2}$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

From Theorems 2.2 and 2.3 it can be seen that the reference and probability-matching priors are equal and that Bayesian analysis using either of these priors will yield exactly the same results.

Note that the reference (probability matching) prior in terms of γ and the standard deviations, $\sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}$ is

$$p\left(\gamma, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}\right) \propto \frac{1}{|\gamma|\left(1+2\gamma^2\right)^{\frac{1}{2}}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-1} \propto \frac{1}{|\gamma|\sqrt{\gamma^2+\frac{1}{2}}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-1}.$$

2.4.3. The Joint Posterior Distribution

By combining the prior with the likelihood function the joint posterior distribution of $\gamma, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}$ can be obtained.

$$p\left(\gamma, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}} | data\right) \propto \prod_{i=1}^{\ddot{m}} \left(\sigma_i\right)^{-(n+1)} \exp\left\{-\frac{1}{2\sigma_i^2}\left[n\left(\bar{x}_i - \frac{\sigma_i}{\gamma}\right)^2 + (n-1)\, s_i^2\right]\right\} \times \frac{1}{|\gamma|\left(1+2\gamma^2\right)^{\frac{1}{2}}} \quad (2.3)$$

In Theorem 2.4 it will be proved that Equation (2.3) is proper and can be used for inferences.

Theorem 2.4. *The posterior distribution $p\left(\gamma, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}} | data\right)$ is a proper distribution.*

Proof. The proof is given in the Mathematical Appendices to this chapter. □

From Equation (2.3) the conditional posterior distributions follow easily as

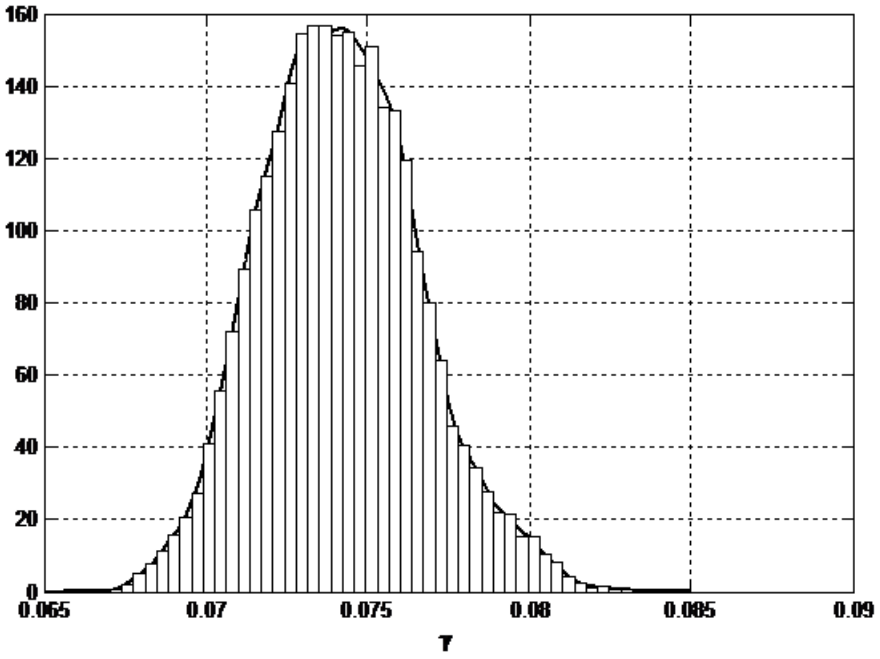
$$p\left(\sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}} | \gamma, data\right) \propto \prod_{i=1}^{\ddot{m}} \left(\sigma_i\right)^{-(n+1)} \exp\left\{-\frac{1}{2\sigma_i}\left[n\left(\bar{x}_i - \frac{\sigma_i}{\gamma}\right)^2 + (n-1)\, S_i^2\right]\right\} \text{ with } \sigma_i^2 > 0 \tag{2.4}$$

$$p\left(\gamma | \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}, data\right) \propto \frac{1}{|\gamma|\left(1+2\gamma^2\right)^{\frac{1}{2}}} \exp\left\{-\frac{n}{2}\sum_{i=1}^{\ddot{m}} \frac{1}{\sigma_i^2}\left(\bar{x}_i - \frac{\sigma_i}{\gamma}\right)^2\right\} \text{ with } -\infty < \gamma < \infty \tag{2.5}$$

For the medical example, $n = 5$ and $\ddot{m} = 105$.

By using the conditional posterior distributions (Equations (2.4) and (2.5)) and Gibbs sampling the unconditional posterior distribution of the coefficient of variation, $p\left(\gamma | data\right)$ can be obtained as illustrated in Figure 2.1.

Figure 2.1.: Histogram of the Posterior-Distribution of $\gamma = \frac{\sigma}{\mu}$



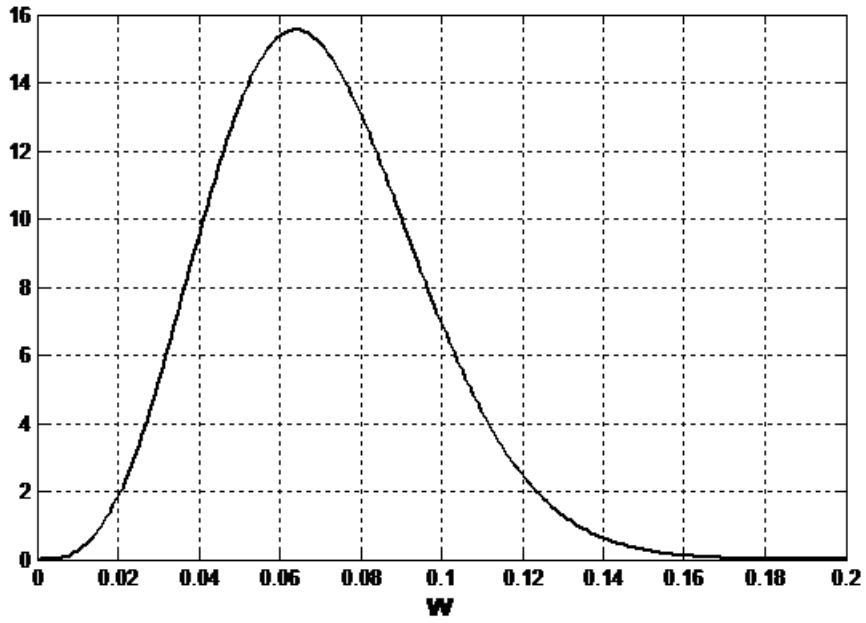
$mean\left(\gamma\right) = 0.0742$, $mode\left(\gamma\right) = 0.0739$, $median\left(\gamma\right) = 0.0741$, $var\left(\gamma\right) = 6.0467e^{-6}$

From a frequentist point of view Kang, Lee, Seong, and Hawkins (2007) mentioned that the best way to pool the sample coefficients of variation is to calculate the root mean square $\hat{\gamma} = \sqrt{\frac{1}{m}\sum_i w_i^2} = \sqrt{\frac{1}{105}(0.593515)} = 0.0752$.

It is interesting to note that the root mean square value is for all practical purposes equal to the mean of the posterior distribution of γ.

By substituting each of the simulated γ values of the posterior distribution into the conditional predictive density $f(w|\gamma)$ and using the Rao-Blackwell procedure the unconditional posterior predictive density $f(w|data)$ of a future sample coefficient of variation can be obtained. This is illustrated in Figure 2.2 for $n = 5$.

Figure 2.2.: Predictive Density $f(w|data)$ for $n = 5$



$mean(w) = 0.0698$, $mode(w) = 0.0640$, $median(w) = 0.0674$, $var(w) = 6.5408e^{-4}$
99.73% equal-tail interval = $(0.0115; 0.1582)$
99.73% HPD interval = $(0.0081; 0.1511)$

Kang, Lee, Seong, and Hawkins (2007) calculated lower (LCL=0.01218) and upper (UCL=0.15957) control limits which are very much the same as the 99.73% equal-tail prediction interval.

Kang, Lee, Seong, and Hawkins (2007) then applied their control chart to a new dataset of 35 patients from a different laboratory. Eight of the patients' coefficient of variation (based on five observations) lie outside the control limits. Since the

99.73% equal-tail prediction interval is effectively identical to the control limits of Kang, Lee, Seong, and Hawkins (2007) our conclusions are the same.

In what follows the Bayesian posterior predictive distribution will be used to derive the distribution of the run-length and the average run-length (ARL).

As mentioned the rejection region of size β ($\beta = 0.0027$) for the predictive distribution is

$$\beta = \int_{R(\beta)} f\left(w|data\right) dw.$$

In the case of the equal-tail interval, $R\left(\beta\right)$ represents those values of w that are smaller than 0.0115 or larger than 0.1582.

It is therefore clear that statistical process control is actually implemented in two phases. In Phase I the primary interest is to assess process stability. The practitioner must therefore be sure that the process is in statistical control before control limits can be determined for online monitoring in Phase II.

Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length. The run-length is defined as the number of future coefficients of variation, r until the control chart signals for the first time. (Note that r does not include the coefficient of variation when the control chart signals.) Given γ and a stable Phase I process, the distribution of the run-length r is geometric with parameter

$$\Psi\left(\gamma\right) = \int_{R(\beta)} f\left(w|\gamma\right) dw$$

where $f\left(w|\gamma\right)$ is the distribution of a future sample coefficient of variation given γ as defined in Equation (2.2).

The value of γ is of course unknown and the uncertainty is described by the posterior distribution.

The predictive distribution of the run-length or the average run-length can therefore easily be simulated. The mean and variance of r given γ are given by

$$E\left(r|\gamma\right) = \frac{1-\Psi\left(\gamma\right)}{\Psi\left(\gamma\right)}$$

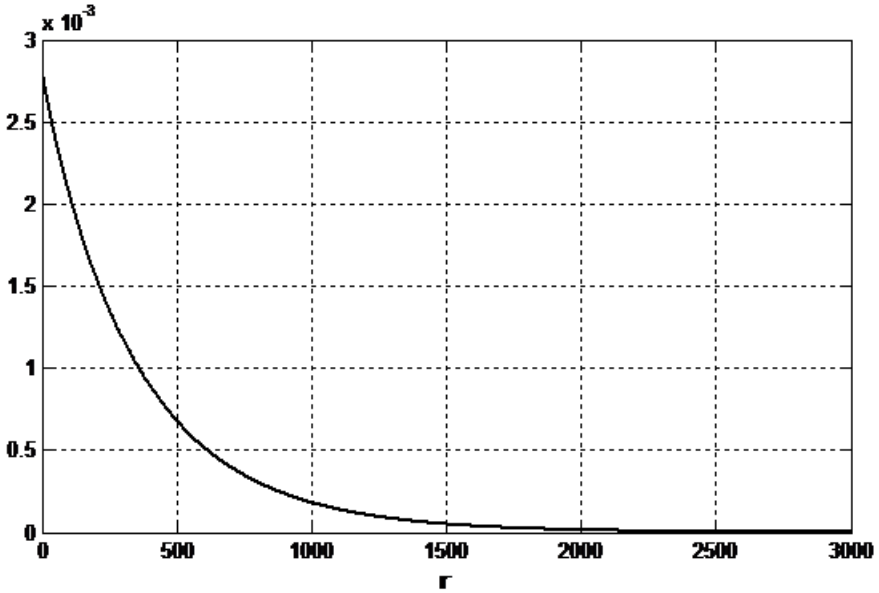
and

$$Var\left(r|\gamma\right) = \frac{1-\Psi\left(\gamma\right)}{\Psi^2\left(\gamma\right)}$$

The unconditional moments $E\left(r|data\right)$, $E\left(r^2|data\right)$ and $Var\left(r|data\right)$ can therefore easily be obtained by simulation or numerical integration. For further details see Menzefricke (2002, 2007, 2010b,a).

In Figure 2.3 the predictive distribution of the run-length is displayed and in Figure 2.4, the distribution of the average run-length is given for the 99.73% equal-tail interval.

Figure 2.3.: Predictive Distribution of the Run-Length $f\left(r|data\right)$ for $n = 5$

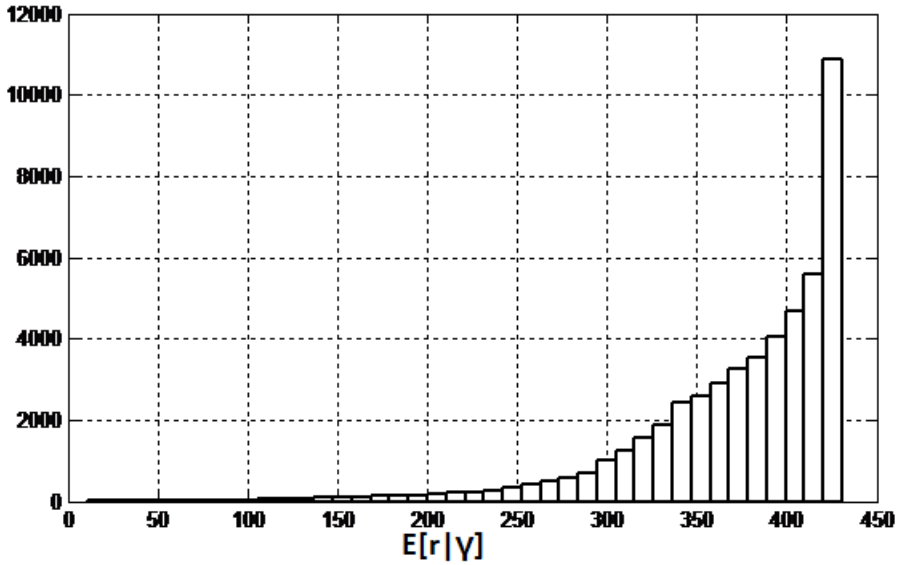


$E\left(r|data\right) = 373.1327,\ Median\left(r|data\right) = 252.420$

As mentioned for given γ, the run length r is geometric with parameter $\Psi\left(\gamma\right)$. The unconditional run length displayed in Figure 2.3 is therefore obtained using the Rao-Blackwell method, i.e., it is the average of the conditional "run-lengths". Figure 2.4

on the other hand is the distribution of $E(r|\gamma)$ for each simulated value of γ, i.e., the distribution of the expected run-length.

Figure 2.4.: Distribution of the Expected Run-Length



Mean = 373.1347, Median = 398.3835

From Figure 2.3 it can be seen that the expected run-length, $E(r|data) = 373.1327$, is the same as the ARL of 370 given by Kang, Lee, Seong, and Hawkins (2007). The median run-length, $Median(r|data) = 252.420$ is smaller than the mean run-length. This is clear from the skewness of the distribution. The means of Figure 2.3 and Figure 2.4 are the same as it should be.

In the case of the highest posterior density (HPD) limits, $\tilde{R}(\beta)$ represents those values of w that are smaller than 0.0081 and larger than 0.1551. The expected run-length is now $E(r|data) = 383.3208$ and $Median(r|data) = 239.850$. For the distribution of the expected run-length ($E(r|\gamma)$), the Mean = 383.5127 and the Median = 360.9072.

A comparison of the above given results show that the run-lengths do not differ much for equal-tail and HPD limits.

Further results are given in Table 2.2.

Table 2.2.: Descriptive Statistics

	Figure					
Descriptive Statistics	2.1 $p(\gamma\|data)$	2.2 $f(w\|data)$	2.3 $f(r\|data)$ Equal-tail	2.4 Exp Run Length Equal-tail	$f(r\|data)$ HPD Limits	Exp Run Length HPD Limits
Mean	*0.0742*	0.0698	373.1327	373.1347	383.3208	383.5127
Median	*0.0741*	0.0674	252.420	389.3835	239.850	360.9077
Mode	*0.0739*	0.0640	-	-	-	-
Variance	$6.0467e^{-6}$	$6.5408e^{-4}$	$1.4500e^{5}$	$3.448e^{3}$	$1.911e^{5}$	$2.3182e^{4}$
95% Equal-tail	(0.0697; 0.0794)	(0.0252; 0.1242)	(8; 1404)	(210.1463; 430.6457)	(7.08; 1584.60)	(139.5615; 735.6124)
95% HPD	(0.0695; 0.0792)	-	(0; 1284)	(258.4241; 430.8225)	(0; 1233.3)	(124.3392; 706.9055)
99.73% Equal-tail	-	(0.0115; 0.1582)	-	-	-	-
99.73% HPD	-	(0.0081; 0.1511)	-	-	-	-

2.5. Simulation Study

In this section a simulation study will be conducted to observe if the 95% Bayesian confidence interval for γ have the correct frequentist coverage.

For the simulation study the following combinations of parameter values will be used:

μ_i	10	20	30	40	50	60	$\cdots$	1000	1010	1020	$\cdots$	1050
σ_i	0.75	1.5	2.25	3.0	3.75		$\cdots$	75		$\cdots$		78.15

which means that $\gamma = \frac{\sigma_i}{\mu_i} = 0.075$, $i = 1, 2, \ldots, \ddot{m}$ and $\ddot{m} = 105$. These parameter combinations are representative of the parameter values of the medical dataset on patients undergoing organ transplantation analyzed by Kang, Lee, Seong, and Hawkins (2007). As mentioned, the dataset consist of $\ddot{m} = 105$ patients and the number of assays obtained for each patient is $n = 5$. As a common estimator of γ, the root mean square $\hat{\gamma} = 0.075$ was used.

For the above given parameter combinations a dataset can be simulated consisting of $\ddot{m}$ samples and $n = 5$ observations per sample. However, since we are only

interest in the sufficient statistics $\bar{X}_i$ and S_i these can be simulated directly, namely $\bar{X}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{n}\right)$ and $S_i^2 \sim \frac{\sigma_i^2 \chi_{n-1}^2}{n-1}$.

The simulated $\bar{X}_i$ and S_i^2 $(i = 1, 2, \ldots, \ddot{m})$ values are then substituted in the conditional posterior distributions $p(\sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}|\gamma, data)$ and $p(\gamma|\sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}})$ (Equations (2.4) and (2.5)). By using the conditional posterior distributions and Gibbs sampling the unconditional posterior distribution $p(\gamma|data)$ can be obtained. A confidence interval for γ will be calculated as follows:

Simulate $l = 10,000$ values of γ and sort the values in ascending order $\tilde{\gamma}_{(1)} \leq \tilde{\gamma}_{(2)} \leq \cdots \leq \tilde{\gamma}_{(l)}$.

Let $K_1 = \left[\frac{\alpha}{2} l\right]$ and $K_2 = \left[\left(1 - \frac{\alpha}{2}\right) l\right]$ where $[a]$ denotes the largest integer not greater than a. $\left\{\tilde{\gamma}_{(K_1)}, \tilde{\gamma}_{(K_2)}\right\}$ is then a $100(1-\alpha)\,\%$ Bayesian confidence interval for γ.

By repeating the procedure for $R = 3,000$ datasets it is found that the $3,000$ 95% Bayesian confidence intervals ($\alpha = 0.05$) cover the parameter $\gamma = 0.075$ in $2,841$ cases. An estimate of the frequentist probability of coverage is therefore $P\left\{\tilde{\gamma}_{(K_1)} < \gamma < \tilde{\gamma}_{(K_2)}\right\} = 0.9470$. Also $P\left\{\gamma \leq \tilde{\gamma}_{(K_1)}\right\} = 0.00217$ and $P\left\{\gamma \geq \tilde{\gamma}_{(K_2)}\right\} = 0.0313$.

2.6. Additional Details on Priors and Methods Used

Upon suggestion from external assessor further details on priors and methods used in this Chapter and throughout the thesis is provided in this section.

2.6.1. Reference Priors

Suppose the data $\underline{Y}$ depends on a $k \times 1$ vector of unknown parameters $\underline{\theta}$. The reference prior method is motivated by the notion of maximizing the expected amount of information about $\underline{\theta}$ provided by the data $\underline{Y}$. The expectation is $E\left[D\left(p\left(\underline{\theta}|\underline{Y}\right), p\left(\underline{\theta}\right)\right)\right]$ where

$$D\left(p\left(\underline{\theta}|\underline{Y}\right), p\left(\underline{\theta}\right)\right) = \int_{\theta} p\left(\underline{\theta}|\underline{Y}\right) \log\left[\frac{p\left(\underline{\theta}|\underline{Y}\right)}{p\left(\underline{\theta}\right)}\right] d\underline{\theta}$$

is the Kullback-Lieber divergence.

The actual reference prior method stems from a modification of the notion of maximizing the expected information provided by the data. Berger and Bernardo (1992) define $\underline{Z}_t = [\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_t]$ to be a vector containing data from t replications of an experiment. The first step in the reference prior method is to choose a prior distribution to maximize $E\left[D\left(p\left(\underline{\theta}|\underline{Y}\right), p\left(\underline{\theta}\right)\right)\right]$ for each t. The reference prior is then given as the limit of these priors. The algorithm for generating reference priors is described by Berger and Bernardo (1992) and Robert (2001). Only some of the features of the algorithm are described below:

1. Assume that the Fisher information matrix for $\underline{\theta}$, $F\left(\underline{\theta}\right)$, exists and is of full rank. Denote $S = F^{-1}\left(\underline{\theta}\right)$.
2. Separate the parameters into r groups of sizes $n_1, n_2, \ldots, n_r$ that correspond to their decreasing level of importance, i.e., $\underline{\theta} = \left[\underline{\theta}_{(1)} \vdots \underline{\theta}_{(2)} \vdots \ldots \vdots \underline{\theta}_{(r)}\right]$ where $\underline{\theta}_{(1)} = [\theta_1, \ldots, \theta_{N_1}]$, $\underline{\theta}_{(2)} = [\theta_{N_1+1}, \ldots, \theta_{N_2}]$ and $\underline{\theta}_{(r)} = \left[\theta_{N_{r-1}+1}, \ldots, \theta_k\right]$ with $N_i = \sum_{j=1}^{i} n_j$ for $j = 1, \ldots, r$. Note that $\underline{\theta}_{(1)}$ is the most important and $\underline{\theta}_{(r)}$ is the least.
3. Define, for $j = 1, \ldots, r$, $\underline{\theta}_{[j]} = \left[\underline{\theta}_{(1)}, \ldots, \underline{\theta}_{(j)}\right]$ and $\underline{\theta}^{[j]} = \left[\underline{\theta}_{(j+1)}, \ldots, \underline{\theta}_{(r)}\right]$ so that $\underline{\theta} = \left[\underline{\theta}_{[j]} \vdots \underline{\theta}^{[j]}\right]$.
4. Decompose the matrix S according to the r groups of sizes $n_1, n_2, \ldots, n_r$, i.e., $S = \begin{bmatrix} A_{11} & A_{21}' & \cdots & A_{r1}' \\ A_{21} & A_{22} & \cdots & A_{r2}' \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rr} \end{bmatrix}$ where A_{ij} is an $n_i \times n_j$ matrix.
5. Define S_j as the $N_j \times N_j$ matrix consisting of elements from the upper left corner of S with $S_r \equiv S$.
6. Let $H_j \equiv S_j^{-1}$. Then define h_j to be the $n_j \times n_j$ matrix contained in the upper lower right corner of H_j for $j = 1, \ldots, r$.
7. Define the $n_j \times N_{j-1}$ matrix $B_j = \left[A_{j1} \quad A_{j2} \quad \ldots \quad A_{j\,j-1}\right]$, for $j = 2, \ldots, r$, of sizes $(n_j \times N_{j-1})$.
8. It is straightforward to very that for $j = 2, \ldots, r$ $h_j = \left[A_{jj} - B_j H_{j-1} B_j'\right]^{-1}$ and $H_j = \begin{bmatrix} H_{j-1} + H_{j-1}B_j' h_j B_j H_{j-1} & -H_{j-1}B_j' h_j \\ -h_j B_j H_{j-1} & h_j \end{bmatrix}$.
9. Iteratively calculate $H_2, \ldots, H_r$ and hence $h_2, \ldots, h_r$ to obtain the ordered reference priors under asymptotic normality.

According to Bernardo (1998), the derivation of the ordered reference prior is greatly simplified if the $h_j(\underline{\theta})$ terms depend only on $\underline{\theta}_{[j]}$, and not on $\underline{\theta}^{[j]}$, then

$$p^l(\underline{\theta}) = \prod_{j=1}^{m} \frac{|h_j(\underline{\theta})|^{\frac{1}{2}}}{\int \left|h_j(\underline{\theta})^{\frac{1}{2}}\right| d\underline{\theta}_{[j]}}.$$

Often some of the integrals appearing in the integral are not defined. Berger and Bernardo (1992) then propose to derive the reference prior for compact sets of $\underline{\theta}^l$ of $\underline{\theta}$ and to consider the limit of the corresponding reference priors as l tends to infinity and $\underline{\theta}^l$ tends to $\underline{\theta}$.In general, the resulting limits do not depend on the choice of sequence of compact sets.

As in the case of the Jeffrey's prior, the reference prior method is derived from the Fisher information matrix. Berger and Bernardo (1992) recommended the reference prior be based on having each parameter in its own group, i.e., having each conditional reference prior only be one-dimensional. The notation $\{\theta_1, \theta_2, \theta_3\}$ means that the parameter θ_1 is the most important and θ_3 the least important.

2.6.2. Probability Matching Priors

The study of priors ensuring, up to the desired order of asymptotics, the approximate frequentist validity of posterior credible sets has received significant attention in recent years and a considerable interest is still continuing in this field. Bayesian credible sets based on these priors have approximately correct frequentist coverage as well. Such priors are generically known as probability matching priors, or matching priors in short. As noted by Tibshirani (1989) among others, study in this direction has several important practical implications with appeal to both Bayesians and Frequentists:

1. First, the ensuing matching priors are, in a sense, non-informative. The approximate agreement between the Bayesian and frequentist coverage probabilities of the associated credible sets provided an external validation for these priors. They can form the basis of an objective Bayesian analysis and are potentially useful for comparative purposes in subjective Bayesian analyses as well.
2. Second, Bayesian credible sets given by matching priors can also be interpreted as accurate frequentist confidence sets because of their approximately correct

frequentist coverage. Thus the exploration of matching priors provides a route for obtaining accurate frequentist confidence sets which are meaningful also to a Bayesian.

3. In addition, the research in this area has led to the development of a powerful and transparent Bayesian route, via a shrinkage argument, for higher order asymptotic frequentist computations.

Further, Berger states (in Wolpert (2004)) that frequentist reasoning will play an important role in finally obtaining good general objective priors for model selection. Indeed, some statisticians argue that frequency calculations are an important part of applied Bayesian statistics (see Rubin (1984) for example).

There are two methods for generating probability-matching priors due to Tibshirani (1989) and Datta and Ghosh (1995). Tibshirani (1989) generated probability-matching priors by transforming the model parameters so that the (single) parameter of interest is orthogonal to the other parameters. The prior distribution is then taken to be proportional to the square root of the upper left element of the information matrix in the new parametrization.

Datta and Ghosh (1995) provided a different solution to the problem of finding probability-matching priors. They derived the differential equation that 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.

The exact definition of Datta and Ghosh (1995) is as follows: Suppose $Y_1, Y_2, \ldots, Y_n$ are independently and identically distributed with density $f\left(y, \underline{\theta}\right)$ where $\underline{\theta} = \left[\theta_1 \theta_2, \ldots, \theta_k\right]'$ is a k-dimensional vector of parameters and the parameter of interest is $t\left(\underline{\theta}\right)$, which is a real-valued twice continuously differentiable parametric function. Consider a prior density for $\underline{\theta}$, $p\left(\underline{\theta}\right)$, which matches frequentist and posterior probability for $t\left(\underline{\theta}\right)$ as follows: For $-\infty < z < \infty$

$$P_\theta\left[n^{\frac{1}{2}}\left(t\left(\underline{\theta}\right) - t\left(\underline{\hat{\theta}}\right)\right)\frac{1}{\tau} \leq z\right] = P_{p(\underline{\theta})}\left[n^{\frac{1}{2}}\left(t\left(\underline{\theta}\right) - t\left(\underline{\bar{\theta}}\right)\right)\frac{1}{\tau} \leq z|\underline{Y}\right] + O_p\left(n^{-1}\right)$$

where $\underline{\hat{\theta}}$ is the posterior mode or maximum likelihood estimator of $\underline{\theta}$, τ^2 is the asymptotic posterior variance of $n^{\frac{1}{2}}\left[t\left(\underline{\theta}\right) - t\left(\underline{\hat{\theta}}\right)\right]$ up to $O_p\left(n^{-\frac{1}{2}}\right)$, $P\left(\cdot\right)$ is the joint

posterior measure of $\underline{Y} = [Y_1, Y_2, \ldots, Y_n]'$ under $\underline{\theta}$, and $P(\cdot|\underline{Y})$ is the posterior probability measure of $\underline{\theta}$ under the prior $P(\underline{\theta})$. According to Datta and Ghosh (1995), such a prior may be sought in an attempt to reconcile a frequentist and Bayesian approach or to find (in some cases validate) a non-informative prior, or to construct frequentist confidence sets.

Let

$$\nabla_{t(\underline{\theta})} = \left[\frac{\partial}{\partial\theta_1} t(\underline{\theta}), \ldots, \frac{\partial}{\partial\theta_k} t(\underline{\theta})\right]'$$

and

$$\eta(\underline{\theta}) = \frac{F^{-1}(\underline{\theta})\nabla_t(\underline{\theta})}{\sqrt{\nabla_t'(\underline{\theta})F^{-1}(\underline{\theta})\nabla_t(\underline{\theta})}} = [\eta_1(\underline{\theta}), \ldots, \eta_k(\underline{\theta})]'$$

where $F(\underline{\theta})$ is the Fisher information matrix and $F^{-1}(\underline{\theta})$ is its inverse. It is evident that $\eta'(\underline{\theta})F(\underline{\theta})\eta(\underline{\theta}) = 1$ for all $\underline{\theta}$. Datta and Ghosh (1995) proved that the agreement between the posterior probability and the frequentist probability holds if and only if

$$\sum_{\alpha=1}^{k} \frac{\partial}{\partial\theta_\alpha}\{\eta_\alpha(\underline{\theta})p(\underline{\theta})\} = 0.$$

Henceforth $p(\underline{\theta})$ is the probability-matching prior for $\underline{\theta}$, the vector of unknown parameters.

The method of Datta and Ghosh (1995) provides a necessary and sufficient condition that a prior distribution must satisfy in order to have the probability-matching prior property. They pointed out that their method is more general than Tibshirani's, but will yield equivalent results when the parameter of interest is defined to be the first parameter in an orthogonal parametrization.

2.6.3. Rao-Blackwell Method

Throughout this thesis, the Rao-Blackwell method is used to compute predictive distributions. In summary the Rao-Blackwell Theorem provides a process by which

a possible improvement in efficiency of an estimator can be obtained by taking its conditional expectation with respect to a sufficient statistic. In this thesis this is applied by taking the average over a large set of simulated conditional distributions. For a full explanation on the Rao-Blackwell theorem, refer to Rao (1945) and Blackwell (1947).

2.7. Conclusion

This chapter develops a Bayesian control chart for monitoring a common coefficient of variation across a range of sample values. In the Bayesian approach prior knowledge about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution. By using the posterior distribution the predictive distribution of a future coefficient of variation can be obtained.

Determination of reasonable non-informative priors in multi-parameter problems is not an easy task. The Jeffreys' prior for example can have a bad effect on the posterior distribution. Reference and probability-matching priors are therefore derived for a common coefficient of variation across a range of sample values. The theory and results are applied to a real problem of patients undergoing organ transplantation for which Cyclosporine is administered. This problem is discussed in detail by Kang, Lee, Seong, and Hawkins (2007). The 99.73% equal-tail prediction interval of a future coefficient of variation is effectively identical to the lower and upper control chart limits calculated by Kang, Lee, Seong, and Hawkins (2007). A simulation study shows that the 95% Bayesian confidence intervals for γ has the correct frequentist coverage.

The example illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, prediction intervals and run-lengths.

Mathematical Appendix to Chapter 2

Proof to Theorem 2.1

$W = \frac{S}{\bar{X}} = \frac{S/\sigma}{\bar{X}/\sigma} = \frac{V}{Y}$

Since $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} = \chi^2_{\tilde{f}}$, it follows that $V = \frac{S}{\sigma} \sim \frac{1}{\sqrt{\tilde{f}}}\chi_{\tilde{f}}$ and for given μ and σ, $Y = \frac{\bar{X}}{\sigma} \sim N\left(\gamma^{-1}, \frac{1}{n}\right)$ where $\gamma = \frac{\sigma}{\mu}$ is the population coefficient of variation.The joint distribution

$$f(v,y) = f(v)\, f(y) = \frac{\left(\tilde{f}\right)^{\frac{1}{2}\tilde{f}} v^{\tilde{f}-1} \exp\left\{-\frac{1}{2}v^2\tilde{f}\right\}}{2^{\frac{1}{2}\left(\tilde{f}-2\right)}\Gamma\left(\frac{\tilde{f}}{2}\right)} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}\left(y-\frac{1}{\gamma}\right)^2\right\}.$$

From this it follows that the joint distribution of W and Y is given by

$$\begin{aligned} f(w,y) &= \frac{\left(\tilde{f}\right)^{\frac{1}{2}\tilde{f}}(wy)^{\tilde{f}-1}\exp\left\{-\frac{1}{2}w^2y^2\tilde{f}\right\}|y|}{2^{\frac{1}{2}\left(\tilde{f}-2\right)}\Gamma\left(\frac{\tilde{f}}{2}\right)} \frac{\sqrt{n}}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}\left(y-\frac{1}{\gamma}\right)^2\right\} \\ &= \frac{\left(\tilde{f}\right)^{\frac{1}{2}\tilde{f}} w^{\tilde{f}-1}\sqrt{n}\exp\left\{-\frac{1}{2}\frac{n\tilde{f}w^2}{\gamma^2\left(n+\tilde{f}w\right)}\right\} y^{\tilde{f}} \exp\left\{-\frac{1}{2}\left[y\left(n+\tilde{f}w^2\right)^{\frac{1}{2}} - \frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right]^2\right\}}{2^{\frac{1}{2}\left(\tilde{f}-2\right)}\Gamma\left(\frac{\tilde{f}}{2}\right)\sqrt{2\pi}} \end{aligned}$$

Let

$$A(w) = \frac{\left(\tilde{f}\right)^{\frac{1}{2}\tilde{f}} w^{\tilde{f}-1}\exp\left\{-\frac{1}{2}\frac{n\tilde{f}w^2}{\gamma^2\left(n+\tilde{f}w\right)}\right\}}{2^{\frac{1}{2}\left(\tilde{f}-2\right)}\Gamma\left(\frac{\tilde{f}}{2}\right)\sqrt{2\pi}}$$

then

$$f(w) = A(w)\int_{-\infty}^{\infty} y^{\tilde{f}} \exp\left\{-\frac{1}{2}\left[y\left(n+\tilde{f}w^2\right)^{\frac{1}{2}} - \frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right]^2\right\} dy$$

Let $y\left(n+\tilde{f}w^2\right)^{\frac{1}{2}} = q$, then

$$f(w) = A(w)\left(n+\tilde{f}w^2\right)^{-\frac{1}{2}\left(\tilde{f}+1\right)}\int_{-\infty}^{\infty} q^{\tilde{f}} \exp\left\{-\frac{1}{2}\left[q - \frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right]^2\right\} dq$$

Therefore

$$f\left(w|\gamma\right)=\begin{cases}\frac{A(w)}{\left(n+\tilde{f}w^2\right)^{\frac{1}{2}\left(\tilde{f}+1\right)}}I_{\tilde{f}}\left(\frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right) & \text{if } w\geq 0\\ \\ \frac{(-1)^{\tilde{f}}A(W)}{\left(n+\tilde{f}w^2\right)^{\frac{1}{2}\left(\tilde{f}+1\right)}}I_{\tilde{f}}\left(\frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right) & \text{if } w<0\end{cases}$$

where

$$I_{\tilde{f}}\left(\frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right)=\int_{\infty}^{\infty}q^{\tilde{f}}\exp\left\{-\frac{1}{2}\left[q-\frac{n}{\gamma\left(n+\tilde{f}w^2\right)^{\frac{1}{2}}}\right]^2\right\}dq$$

and $\tilde{f}=n-1$.

Proof of Theorem 2.2

The likelihood function is given by

$$L\left(\gamma,\sigma_1^2,\sigma_2^2,...,\sigma_{\ddot{m}}^2|data\right)\propto\prod_{i=1}^{\ddot{m}}\left(\sigma_i^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_i^2}\left[n\left(\bar{x}_i-\frac{\sigma_i}{\gamma}\right)^2+\left(n-1\right)s_i^2\right]\right\}$$

where $\bar{x}_i=\frac{1}{n}\sum_{j=1}^{n}x_{ij}$ and $\left(n-1\right)s_i^2=\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2$.

By differentiating the log likelihood function, $LogL$, twice with respect to the unknown parameters and taking expected values the Fisher information matrix can be obtained.

$$LogL\propto\log L\left(\gamma,\sigma_1^2,\sigma_2^2,...,\sigma_{\ddot{m}}^2|data\right)=-\frac{n}{2}\sum_{i=1}^{\ddot{m}}\log\sigma_i^2-\frac{1}{2}\sum_{i=1}^{\ddot{m}}\frac{1}{\sigma_i^2}\left[n\left(\bar{x}_i-\frac{\sigma_i}{\gamma}\right)^2+\left(n-1\right)s_i^2\right]$$

and

$$\frac{\partial^2 LogL}{\left(\partial\sigma_i^2\right)^2}=\frac{n}{2}\left(\frac{1}{\sigma_i^2}\right)^2-\frac{2n\bar{x}_i^2}{2\left(\sigma_i^2\right)^3}+\frac{3n\bar{x}_i}{4\gamma\sigma_i^5}-\frac{\left(n-1\right)s_i^2}{\left(\sigma_i^2\right)^3}.$$

Therefore

$$-E\left[\frac{\partial^2 LogL}{(\partial\sigma_i^2)^2}\right] = \frac{n}{2}\left(\frac{1}{\sigma_i^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\} \text{ where } i = 1, 2, ..., \ddot{m}.$$

Also

$$-E\left[\frac{\partial^2 LogL}{\partial\sigma_i^2\partial\sigma_l^2}\right] = 0 \text{ where } i = 1, 2, ..., \ddot{m},\ l = 1, 2, ..., \ddot{m} \text{ and } i \neq l.$$

Further

$$\frac{\partial^2 LogL}{(\partial\gamma)^2} = \sum_{i=1}^{\ddot{m}}\left(\frac{2n\bar{x}_i}{\sigma_i\gamma^3} - \frac{3n}{\gamma^4}\right)$$

and

$$-E\left[\frac{\partial^2 LogL}{(\partial\gamma)^2}\right] = \frac{\ddot{m}n}{\gamma^4}$$

If $LogL$ is differentiated with respect to σ_i^2 and γ we get

$$\frac{\partial^2 LogL}{\partial\sigma_i^2\partial\gamma} = \frac{n\bar{x}_i}{2\gamma^2\sigma_i^3}$$

and

$$-E\left[\frac{\partial^2 LogL}{\partial\sigma_i^2\partial\gamma}\right] = \frac{-n}{2\gamma^3\sigma_i^2}.$$

The Fisher information matrix then follows as

$$F\left(\gamma, \sigma_1^2, \sigma_2^2, ..., \sigma_{\ddot{m}}^2\right) = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}$$

where

$$F_{11} = \frac{\ddot{m}n}{\gamma^4}\ ,\ F_{12} = F_{21}^{'} = \begin{bmatrix} \frac{-n}{2\gamma^3\sigma_1^2} & \frac{-n}{2\gamma^3\sigma_2^2} & \cdots & \frac{-n}{2\gamma^3\sigma_{\ddot{m}}^2} \end{bmatrix}$$

and

$$F_{22} = \begin{bmatrix} \frac{n}{2}\left(\frac{1}{\sigma_1^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\} & 0 & \cdots & 0 \\ 0 & \frac{n}{2}\left(\frac{1}{\sigma_2^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n}{2}\left(\frac{1}{\sigma_{\ddot{m}}^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\} \end{bmatrix}.$$

To calculate the reference prior for the ordering $\{\gamma; (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2)\}$, $F_{11\cdot 2}$ must first be calculated and then $|F_{22}|$. Now

$$F_{11\cdot 2} = F_{11} - F_{12}F_{22}^{-1}F_{21} = \frac{\ddot{m}n}{\gamma^4} - \frac{\ddot{m}n}{\gamma^4}\left(\frac{1}{1+2\gamma^2}\right) = \frac{2\ddot{m}n}{\gamma^2\left(1+2\gamma^2\right)} = h_1$$

and

$$p\left(\gamma\right) \propto h_1^{\frac{1}{2}} \propto \frac{1}{|\gamma|\sqrt{\gamma^2+\frac{1}{2}}}.$$

Also

$$|F_{22}| = \left(\frac{n}{2}\left\{1+\frac{1}{2\gamma^2}\right\}\right)^{\ddot{m}} \prod_{i=1}^{\ddot{m}}\left(\frac{1}{\sigma_i^2}\right)^2 = h_2$$

which means that

$$p\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2 | \gamma\right) \propto h_2^{\frac{1}{2}} \propto \prod_{i=1}^{\ddot{m}} \left(\frac{1}{\sigma_i^2}\right).$$

Therefore the reference prior for the ordering $\{\gamma; (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2)\}$ is

$$p_R\left(\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) = p(\gamma)\, p\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2 | \gamma\right) \propto \frac{1}{|\gamma| \sqrt{\gamma^2 + \frac{1}{2}}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2}. \quad (2.6)$$

Proof of Theorem 2.3

Using the previously derived Fisher information matrix we can calculate

$$F^{-1}(\underline{\theta}) = F^{-1}\left(\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) = \begin{bmatrix} F^{11} & F^{12} & F^{13} & \cdots & F^{1,\ddot{m}+1} \\ F^{21} & F^{22} & F^{23} & \cdots & F^{2,\ddot{m}+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F^{\ddot{m}+1,1} & F^{\ddot{m}+1,2} & F^{\ddot{m}+1,3} & \cdots & F^{\ddot{m}+1,\dot{m}+1} \end{bmatrix}.$$

Let

$$t(\underline{\theta}) = t\left(\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) = \gamma.$$

Since

$$\nabla'(\underline{\theta}) = \begin{bmatrix} \frac{\partial}{\partial \gamma} t(\underline{\theta}) & \frac{\partial}{\partial \sigma_1^2} t(\underline{\theta}) & \cdots & \frac{\partial}{\partial \sigma_{\ddot{m}}^2} t(\underline{\theta}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

we have that

$$\begin{aligned} \nabla'(\underline{\theta}) &= \begin{bmatrix} F^{11} & F^{12} & \cdots & F^{1,\ddot{m}+1} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\gamma^2\left(1+2\gamma^2\right)}{2n\ddot{m}} & \frac{\gamma\sigma_1^2}{n\ddot{m}} & \frac{\gamma\sigma_2^2}{n\ddot{m}} & \cdots & \frac{\gamma\sigma_m^2}{n\ddot{m}} \end{bmatrix} \end{aligned}$$

and

$$\sqrt{\nabla_t'\left(\underline{\theta}\right)F^{-1}\left(\underline{\theta}\right)\nabla_t\left(\underline{\theta}\right)}=\left\{\frac{\gamma^2\left(1+2\gamma^2\right)}{2n\ddot{m}}\right\}^{\frac{1}{2}}.$$

Further

$$\Upsilon'\left(\underline{\theta}\right)=\frac{\nabla_t'\left(\underline{\theta}\right)F^{-1}\left(\underline{\theta}\right)}{\sqrt{\nabla_t'\left(\theta\right)F^{-1}\left(\underline{\theta}\right)\nabla_t\left(\underline{\theta}\right)}}=\left[\begin{array}{cccc}\Upsilon_1\left(\underline{\theta}\right) & \Upsilon_2\left(\underline{\theta}\right) & \cdots & \Upsilon_{\ddot{m}+1}\left(\underline{\theta}\right)\end{array}\right]$$

where

$$\Upsilon_1\left(\underline{\theta}\right)=\frac{\gamma\left(1+2\gamma^2\right)^{\frac{1}{2}}}{\left(2n\ddot{m}\right)^{\frac{1}{2}}},$$

$$\Upsilon_2\left(\underline{\theta}\right)=\frac{\left(2\right)^{\frac{1}{2}}\sigma_1^2}{\left\{n\ddot{m}\left(1+2\gamma^2\right)\right\}^{\frac{1}{2}}},$$

$$\Upsilon_3\left(\underline{\theta}\right)=\frac{\left(2\right)^{\frac{1}{2}}\sigma_2^2}{\left\{n\ddot{m}\left(1+2\gamma^2\right)\right\}^{\frac{1}{2}}}$$

and

$$\Upsilon_{\ddot{m}+1}\left(\underline{\theta}\right)=\frac{\left(2\right)^{\frac{1}{2}}\sigma_{\ddot{m}}^2}{\left\{n\ddot{m}\left(1+2\gamma^2\right)\right\}^{\frac{1}{2}}}.$$

The prior

$$p_M\left(\underline{\theta}\right)=p_M\left(\gamma,\sigma_1^2,\sigma_2^2,\ldots,\sigma_{\ddot{m}}^2\right)\propto\frac{1}{\left|\gamma\right|\left(1+2\gamma^2\right)^{\frac{1}{2}}}\prod_{i=1}^{\ddot{m}}\sigma_i^{-2} \tag{2.7}$$

is therefore a probability-matching prior since

$$\frac{\partial}{\partial\gamma}\left\{\Upsilon_1\left(\underline{\theta}\right)p_M\left(\underline{\theta}\right)\right\}+\frac{\partial}{\partial\sigma_1^2}\left\{\Upsilon_2\left(\underline{\theta}\right)p_M\left(\underline{\theta}\right)\right\}+\cdots+\frac{\partial}{\partial\sigma_{\ddot{m}}^2}\left\{\Upsilon_{\ddot{m}+1}\left(\underline{\theta}\right)p_M\left(\underline{\theta}\right)\right\}=0.$$

Proof of Theorem 2.4

The joint posterior distribution given in Equation (2.3) can be written as

$$\begin{aligned} p\left(\gamma,\sigma_1,\sigma_2,\ldots,\sigma_{\ddot{m}}|data\right) \quad &\propto \quad \tfrac{1}{|\gamma|\sqrt{\gamma^2+\frac{1}{2}}}\prod_{i=1}^{\ddot{m}}\{\exp\left[-\tfrac{n}{2\gamma^2}\left(1-\tfrac{\bar{x}_i^2}{D_i^2}\right)\right]\left(\tfrac{1}{\sigma_i}\right)^{n+1}\times \\ &\qquad \exp\left[-\tfrac{n}{2}D_i^2\left(\tfrac{1}{\sigma_i}-\tfrac{\bar{x}_i}{D_i^2\gamma}\right)^2\right]\} \end{aligned}$$

where $\bar{x}_i=\frac{1}{n}\sum_{j=1}^{n}x_{ij}$, $D_i^2=\frac{1}{n}\sum_{j=1}^{n}x_{ij}^2$ and $\left(n-1\right)s_i^2=\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2$.

Therefore

$$\begin{aligned} p\left(\gamma|data\right) \quad &= \quad \int_0^\infty\cdots\int_0^\infty p\left(\gamma,\sigma_1,\sigma_2,\ldots,\sigma_{\ddot{m}}|data\right)d\sigma_1 d\sigma_2\ldots d\sigma_{\ddot{m}} \\ &\propto \quad \tfrac{1}{|\gamma|\sqrt{\gamma^2+\frac{1}{2}}}\exp\left\{-\tfrac{n}{2\gamma^2}\sum_{i=1}^{\ddot{m}}\left(1-\tfrac{\bar{x}_i^2}{D_i^2}\right)\right\}\times \\ &\qquad \prod_{i=1}^{\ddot{m}}\left\{\int_0^\infty\left(\tfrac{1}{\sigma_i}\right)^{n+1}\exp\left[-\tfrac{n}{2}D_i^2\left(\tfrac{1}{\sigma_i}-\tfrac{\bar{x}_i}{D_i^2\gamma}\right)^2\right]d\sigma_i\right\} \end{aligned}$$

which is proper. As mentioned by Berger et al. (1999) it is usually the case that the reference and Jeffreys' priors will yield proper posterior distributions.

3. Bayesian Control Charts for the Standardized Normal Mean

3.1. Introduction

As mentioned in Chapter 2 there are some situations in practice where the mean is not a constant and the usual SPC reduces to the monitoring of the variability alone. As a further complication it sometimes happens that the variance of the process is a function of the mean.

The proposed remedy depends on the nature of the relationship between the mean and the variance of the process. In Chapter 2 we looked at the coefficient of variation as one remedy. Another common relationship that we will look at is when the mean and standard deviation is directly proportional so that the standardized mean ($\delta = \frac{\mu}{\sigma}$) is a constant. According to Kang, Lee, Seong, and Hawkins (2007) this is often the case in medical research.

Scientists at the Clinical Research Organization, Quintiles, also confirmed that the standardized mean or coefficient of variation of drug concentrations is constant or approximately constant. By using frequentist methods, Kang, Lee, Seong, and Hawkins (2007), developed a Shewart control chart, equivalent to the S chart, for monitoring the coefficient of variation using rational groups of observations. The control chart developed by them is a time-ordered plot of the coefficient of variation for successive samples. It contains three lines:

- A center line;
- The upper control limit (UCL);
- The lower control limit (LCL).

By using the posterior predictive distribution in this chapter, a Bayesian procedure will be developed to obtain control limits for a future sample standardized mean.

These limits will be compared to the classical limits obtained by Kang, Lee, Seong, and Hawkins (2007).

3.2. Frequentist Methods

Assume that X_j $(j = 1, 2, \ldots, n)$ are independently, identically normally distributed with mean μ and variance σ^2. $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ is the sample mean and $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2$ is the sample variance. The sample coefficient of variation is defined as

$$W = \frac{S}{\bar{X}}$$

and the sample standardized mean as

$$W^{-1} = \frac{\bar{X}}{S}.$$

Kang, Lee, Seong, and Hawkins (2007) suggested a control chart for the sample coefficient of variation, similar to that of the $\bar{X}$, R and S charts. By deriving a canonical form for the distribution of the coefficient of variation they obtained control limits for a selection of values of n and $\gamma = \frac{\sigma}{\mu}$. The probability of exceeding these limits is $\frac{1}{740}$ on each side when the process is in control.

In this chapter the emphasis will rather be on the inverse of the coefficient of variation, i.e., the standardized mean. From a statistical point of view it is easier to handle the standardized mean than the coefficient of variation.

It is well known that

$$T = \frac{\sqrt{n}\bar{X}}{S} = \sqrt{n}W^{-1}$$

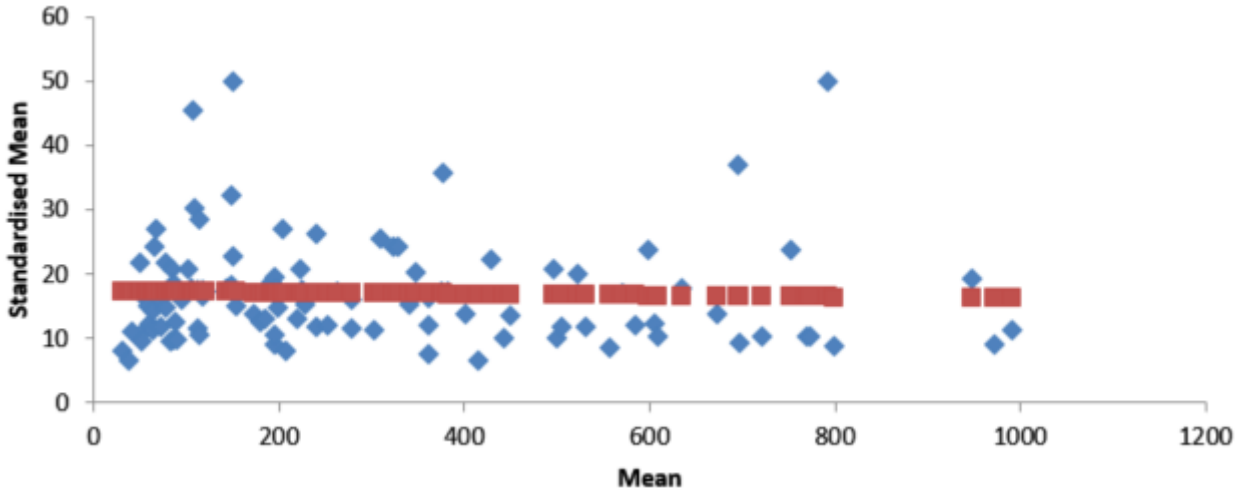
follows a non-central t distribution with $(n-1)$ degrees of freedom and non-centrality parameter $\sqrt{n}\delta$. Inferences about a future standardized mean can therefore be made if δ is known.

3.2.1. The Data

The example used by Kang, Lee, Seong, and Hawkins (2007) and also described in Chapter 2 of this thesis, will be used for this chapter.

As mentioned, in this chapter the emphasis will be on the standardized mean $\delta = \frac{\mu}{\sigma}$. By using the predictive distribution, a Bayesian procedure will be developed to obtain control limits for a future sample standardized mean. Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length. In the last section of this chapter a simulation study will be conducted to evaluate the accuracy of our Bayesian procedure. A plot of the sample standardized means against the sample means are given in Figure 3.1.

Figure 3.1.: Scatter Plot of Sample Standardized Mean Versus Sample Mean



From Figure 3.1 and the least squares regression line it is clear that a common standardized mean assumption is appropriate for the Phase I Cyclosporine data. The analysis of variance test in Table 3.1 confirms that there is no evidence that the standardized mean depends on the mean. A common standardized mean control chart is therefore justified for ongoing control.

Table 3.1.: Regression Test of Dependence of Standardized Mean on Mean

Source	SS	df	MSS	F	P-value
Regression	8.7093	1	8.7093	0.1272	0.722
Error	7052.7	103	68.4731		
Total	7061.4	104	67.8984		

Note that the p-value of 0.722 is larger than the p-value of 0.245 calculated by Kang, Lee, Seong, and Hawkins (2007) for the coefficient of variation which might be an indication that a common standardized mean is more appropriate to use than a common coefficient of variation.

3.3. Bayesian Procedure

The non-central t distribution can be used to make inferences about a future standardized mean if δ is known. In practice δ is usually unknown.

By assigning a prior distribution to the unknown parameters the uncertainty in the estimation of the unknown parameters can adequately be handled. The information contained in the prior is combined with the likelihood to obtain the posterior distribution of δ. By using the posterior distribution the predictive distribution of a future standardized mean can be obtained. The predictive distribution on the other hand can be used to obtain control limits and to determine the distribution of the run length. Determination of reasonable non-informative priors is however not an easy task. Therefore, in the next section, reference and probability matching priors will be derived for a common standardized mean across the range of sample values.

3.4. Reference and Probability-Matching Priors for a Common Standardized Mean

As mentioned the Bayesian paradigm emerges as attractive in many types of statistical problems, also in the case of the standardized mean.

Prior distributions are needed to complete the Bayesian specification of the model. Determination of reasonable non-informative priors in multi-parameter problems

is not easy; common non-informative priors, such as the Jeffreys' prior can have features that have an unexpectedly dramatic effect on the posterior.

Reference and probability-matching priors often lead to procedures with good frequency properties while returning to the Bayesian flavor. The fact that the resulting Bayesian posterior intervals of the level $1-\alpha$ are also good frequentist intervals at the same level is a very desirable situation.

See also Bayarri and Berger (2004) and Severine, Mukerjee, and Ghosh (2002) for a general discussion.

3.4.1. The Reference Prior

In this section the reference prior of Berger and Bernardo (1992) will be derived for a common standardized mean, δ, across the range of sample values.

Bernardo (1998) derived the reference prior for the standardized mean in the case of a single sample. From the medical example given in Kang, Lee, Seong, and Hawkins (2007) it is clear that the standard deviation of measurements is approximately proportional to the mean; that is, the standardized mean is constant across the range of means, which is an indication that the a reference prior for a common standardized mean should be derived.

Theorem 3.1. *Let* $X_{ij} \sim N(\mu_i, \sigma_i^2)$ *where* $i = 1, 2, \ldots, \ddot{m}$*;* $j = 1, 2, \ldots, n$ *and the standardized mean is* $\delta = \frac{\mu_1}{\sigma_1} = \frac{\mu_2}{\sigma_2} = \cdots \frac{\mu_{\ddot{m}}}{\sigma_{\ddot{m}}}$*. The reference prior for the ordering* $\{\delta, (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2)\}$ *is given by*

$$p_R\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) \propto \left(1 + \frac{1}{2}\delta^2\right)^{-\frac{1}{2}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2} \tag{3.1}$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Note: The ordering $\{\delta, (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2)\}$ means that the standardized mean is the most important parameter while the $\ddot{m}$ variance components are of equal importance, but not as important as δ. Also if $\ddot{m} = 1$, Equation (3.1) simplifies to the reference prior obtained by Bernardo (1998).

3.4.2. Probability-Matching Priors

As mentioned in the introduction $p(\underline{\theta})$ is a probability-matching prior for $\underline{\theta} = [\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2]'$ the vector of unknown parameters, if the following differential equation is satisfied:

$$\sum_{\alpha=1}^{\ddot{m}+1} \frac{\partial}{\partial \theta_\alpha} \{\Upsilon_\alpha(\underline{\theta})\, p(\underline{\theta})\} = 0$$

where

$$\Upsilon(\underline{\theta}) = \frac{F^{-1}(\underline{\theta})\, \nabla_t(\underline{\theta})}{\sqrt{\nabla_t'(\underline{\theta})\, F^{-1}(\underline{\theta})\, \nabla_t(\underline{\theta})}} = \left[\begin{array}{cccc} \Upsilon_1(\underline{\theta}) & \Upsilon_2(\underline{\theta}) & \cdots & \Upsilon_{\ddot{m}+1}(\underline{\theta}) \end{array} \right]'$$

and

$$\nabla_t(\underline{\theta}) = \left[\begin{array}{cccc} \frac{\partial}{\partial \theta_1} t(\underline{\theta}) & \frac{\partial}{\partial \theta_2} t(\underline{\theta}) & \cdots & \frac{\partial}{\partial \theta_{\ddot{m}+1}} t(\underline{\theta}) \end{array} \right]'.$$

$t(\underline{\theta})$ is a function of $\underline{\theta}$ and $F^{-1}(\underline{\theta})$ is the inverse of the Fisher information matrix.

Theorem 3.2. *The probability-matching prior for the standardized mean, δ, and the variance components is given by*

$$p_M\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) \propto \left(1 + \frac{1}{2}\delta^2\right)^{-\frac{1}{2}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2}$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

From Theorem 3.1 and Theorem 3.2 it can be seen that the reference and probability-matching priors are equal and the Bayesian analysis using either of these priors will yield exactly the same results.

Note that the reference (probability-matching) prior in terms of δ and the standard

deviations, $\sigma_1, \sigma_{2,} \ldots, \sigma_{\ddot{m}}$ is

$$p\left(\delta, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}\right) = \left(1 + \frac{1}{2}\delta^2\right)^{-\frac{1}{2}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-1}.$$

3.4.3. The Joint Posterior Distribution

By combining the prior distribution with the likelihood function the joint posterior distribution of $\delta, \sigma_1, \sigma_2, \ldots, \ddot{m}$ can be obtained:

$$p\left(\delta, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}} | data\right) \propto \left(1 + \frac{1}{2}\delta^2\right)^{-\frac{1}{2}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-(n+1)} \exp\left\{-\frac{1}{2\sigma_i^2}\left[n\left(\bar{x}_i - \sigma_i\delta\right)^2 + (n-1)\, s_i^2\right]\right\} \tag{3.2}$$

where $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ and $(n-1)\, s_i^2 = \sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2$.

In Theorem 3.3 it will be proved that the joint posterior distribution is proper and can be used for inferences.

Theorem 3.3. *The posterior distribution is $p\left(\delta, \sigma_1, \sigma_2, \ldots, \sigma_m | data\right)$ is a proper posterior distribution.*

Proof. The proof is given in the Mathematical Appendices to this chapter. □

The conditional posterior distributions follow easily from the joint posterior distribution:

$$p\left(\delta | \sigma_1, \sigma_2 \ldots, \sigma_{\ddot{m}}, data\right) \propto \left(1 + \frac{1}{2}\delta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{n}{2}\sum_{i=1}^{\ddot{m}} \frac{1}{\sigma_1^2}\left(\bar{x}_i - \sigma_i\delta\right)^2\right\} \tag{3.3}$$

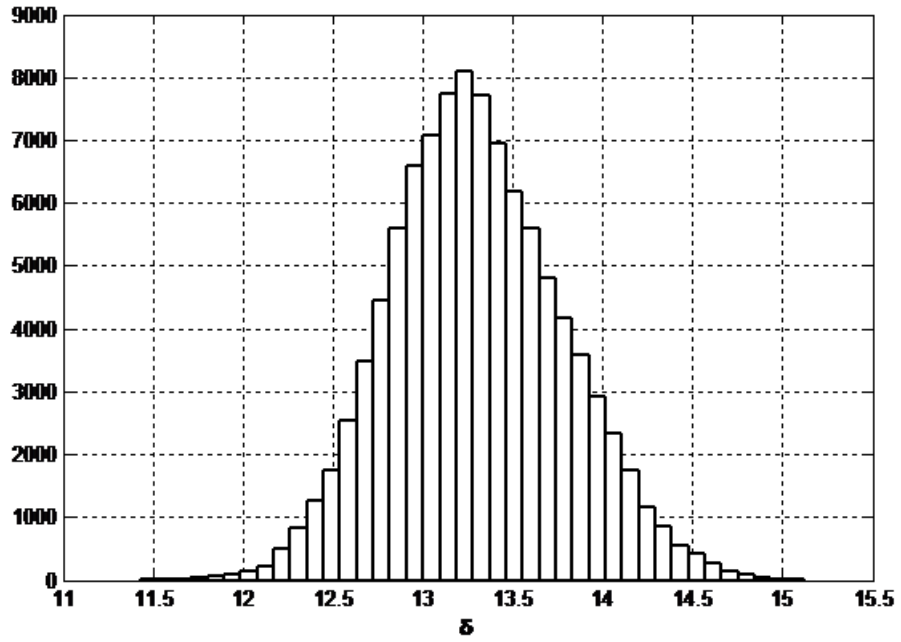
and

$$p\left(\sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}} | \delta, data\right) \propto \prod_{i=1}^{\ddot{m}} \left(\sigma_i^{-(n+1)} \exp\left\{-\frac{1}{2\sigma_i^2}\left[n\left(\bar{x}_i - \sigma_i\delta\right)^2 + (n-1)\, s_i^2\right]\right\}\right). \tag{3.4}$$

By using the conditional posterior distributions (Equations (3.3) and (3.4)) and Gibbs sampling the unconditional posterior distributions can be obtained.

In Figure 3.2 the unconditional posterior distribution of δ (the standardized mean), $p(\delta|data)$ for the medical data is illustrated ($\ddot{m} = 105$ and $n = 5$).

Figure 3.2.: Histogram of the Posterior Distribution of $\delta = \frac{\mu}{\sigma}$



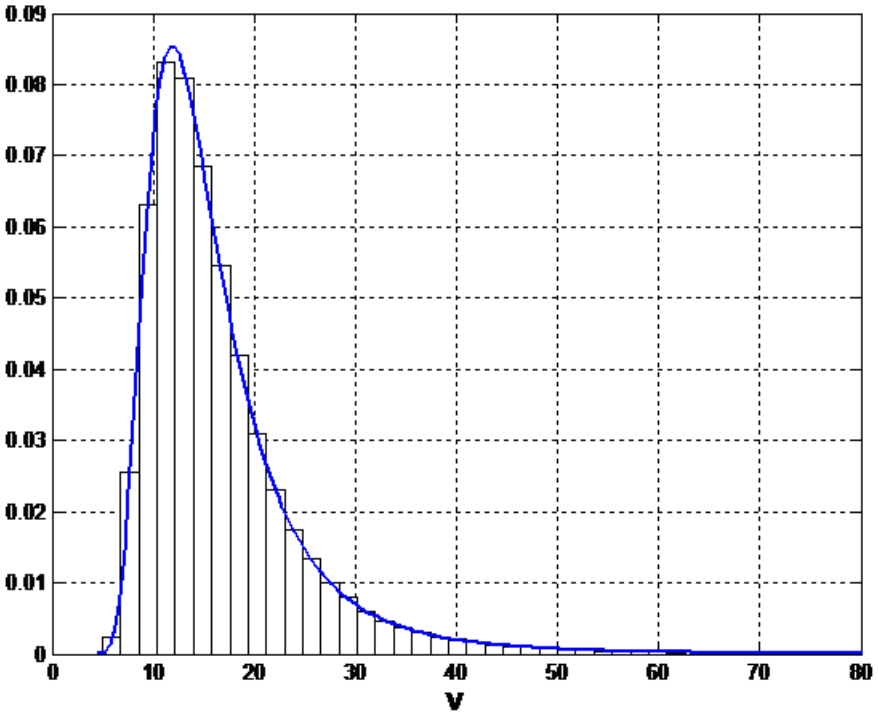
$mean(\delta) = 13.2984$, $mode(\delta) = 13.230$, $median(\delta) = 13.2733$, $var(\delta) = 0.2296$

$\{mean(\delta)\}^{-1} = (13.2984)^{-1} = 0.075197$ can therefore be used as an estimate for the common coefficient of variation, γ. As mentioned in Chapter 2, Kang, Lee, Seong, and Hawkins (2007) used the weighted root mean square estimator $\hat{\gamma} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} w_i^2} = \sqrt{\frac{0.593515}{105}} = 0.0752$ to pool the samples for estimating γ. $w_i = \frac{s_i}{\bar{x}_i}$ is the sample coefficient of variation. It is interesting to note that $\{mean(\delta)\}^{-1}$ is for all practical purposes the same as $\hat{\gamma}$.

Since $T = \sqrt{n}W^{-1} = \sqrt{n}V$ follows a non-central t distribution with $(n-1)$ degrees of freedom and a non-centrality parameter $\sqrt{n}\delta$, $f(V|\delta)$ can be obtained. By substituting each of the simulated δ values of the posterior distribution of δ in $f(V|\delta)$ and using the Rao-Blackwell procedure (averaging the conditional distributions), the unconditional posterior predictive density of $f(V|data)$ (a future standardized mean) can be obtained. This is illustrated by the smooth curve in Figure 3.3.

The histogram in Figure 3.3 is obtained in the following way. Define a future sample mean as $\bar{X}_f$ and a future sample standard deviation as S_f. Since $V = \frac{\bar{X}_f}{S_f} = \frac{\tilde{Z}}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}$ where $\tilde{Z} \sim N\left(\delta, \frac{1}{n}\right)$, the histogram of the distribution of V is obtained by simulating δ from the its posterior distribution and then $\tilde{Z} \sim N\left(\delta, \frac{1}{n}\right)$. Simulate now a χ^2_{n-1} random variable and calculate V and repeat the process a large number of times. It is clear that the two distributions are the same.

Figure 3.3.: Predictive Density $f(V|data)$ for $n = 5$



$mean(V) = 16.6671$, $mode(V) = 11.910$, $median(V) = 14.276$,
$Var(V) = 75.0494$
99.73% Equal-tail Interval $= (6.212; 83.365)$
99.73% HPD Interval $= (5.176; 67.777)$

According to this the 99.73% equal-tail interval for a future sample coefficient of variation is $\left[(83.365)^{-1}; (6.212)^{-1}\right] = [0.011995; 0.1609787]$. For a 99.73% equal-tail control chart for the coefficient of variation, Kang, Lee, Seong, and Hawkins (2007) calculated the lower control limit as 0.1218, the upper control limit as 0.15957 and as central line they used the root-mean square value $\hat{\gamma} = 0.075$. The frequentist limits calculated by them are for all practical purposes the same as our Bayesian

control limits.

As mentioned Kang, Lee, Seong, and Hawkins (2007) then applied their control chart to a new dataset of 35 patients from a different laboratory. Eight of the patients' coefficient of variation (based on five observations) lie outside the control limits. Since the 99.73% equal-tail prediction interval is effectively identical to the control limits of Kang, Lee, Seong, and Hawkins (2007) our conclusions are the same.

Also the rejection region of size β ($\beta = 0.0027$) for the predictive distribution is

$$\beta = \int_{R(\beta)} f\left(V|data\right) dV.$$

In the case of the equal-tail interval, $R\left(\beta\right)$ represents those values of V that are smaller than 6.212 or larger than 83.365.

It is therefore clear that statistical process control is actually implemented in two phases. In Phase I the primary interest is to assess process stability. The practitioner must therefore be sure that the process is in statistical control before control limits can be determined for online monitoring in Phase II.

Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length. The run-length is defined as the number of future standardized means, r until the control chart signals for the first time (Note that r does not include that standardized mean when the control chart signals). Given δ and a stable Phase I process, the distribution of the run-length r is geometric with parameter

$$\Psi\left(\delta\right) = \int_{R(\beta)} f\left(V|\delta\right) dV$$

where $f\left(V|\delta\right)$ is the distribution of a future standardized mean ($\sqrt{n}V$ is a non-central t distribution with $(n-1)$ degrees of freedom and a non-centrality parameter $\sqrt{n}\delta$). The value of δ is of course unknown and its uncertainty is described by its posterior distribution. The predictive distribution of the run-length and the average run-length can therefore easily be obtained.

The mean and variance of r given δ are given by

$$E\left(r|\delta\right) = \frac{1-\psi\left(\delta\right)}{\psi\left(\delta\right)}$$

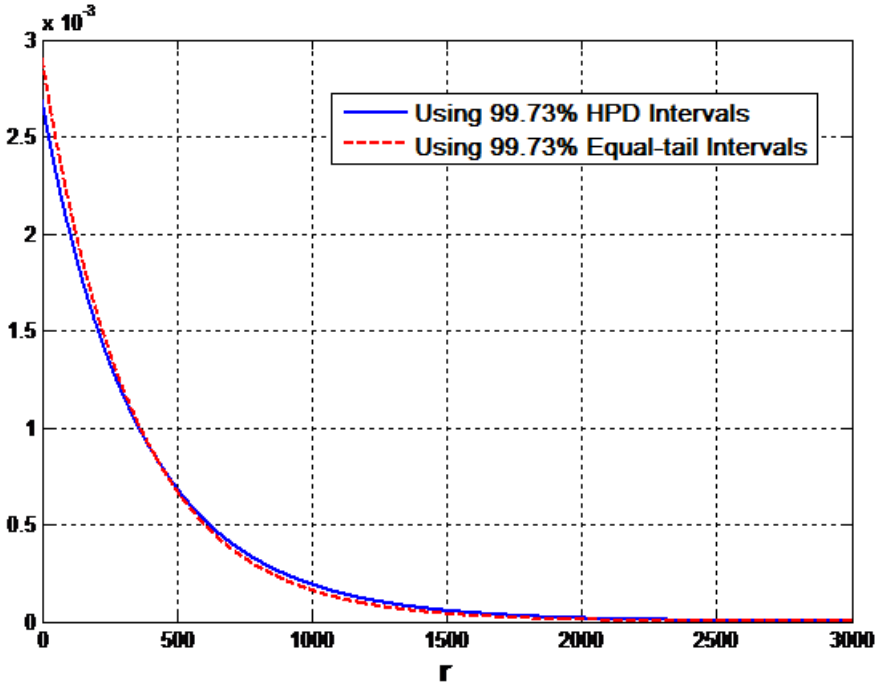
and

$$Var\left(r|\delta\right) = \frac{1-\psi\left(\delta\right)}{\psi^2\left(\delta\right)}.$$

The unconditional moments $E\left(r|data\right)$, $E\left(r^2|data\right)$ and $Var\left(r|data\right)$ can therefore easily be obtained by simulation or numerical integration. For further details see Menzefricke (2002, 2007, 2010b,a).

In Figure 3.4 the predictive distributions of the run-length is displayed for the 99.73% equal-tail interval as well as for the 99.73% HPD interval. As mentioned for given δ, the run-length, r, is geometric with parameter $\psi\left(\delta\right)$. The unconditional run-length given in Figure 3.4 are therefore obtained by the Rao-Blackwell method, i.e., the average of a large number of conditional run-lengths.

Figure 3.4.: Predictive Density of Run Length $f(r|data)$with $n = 5$

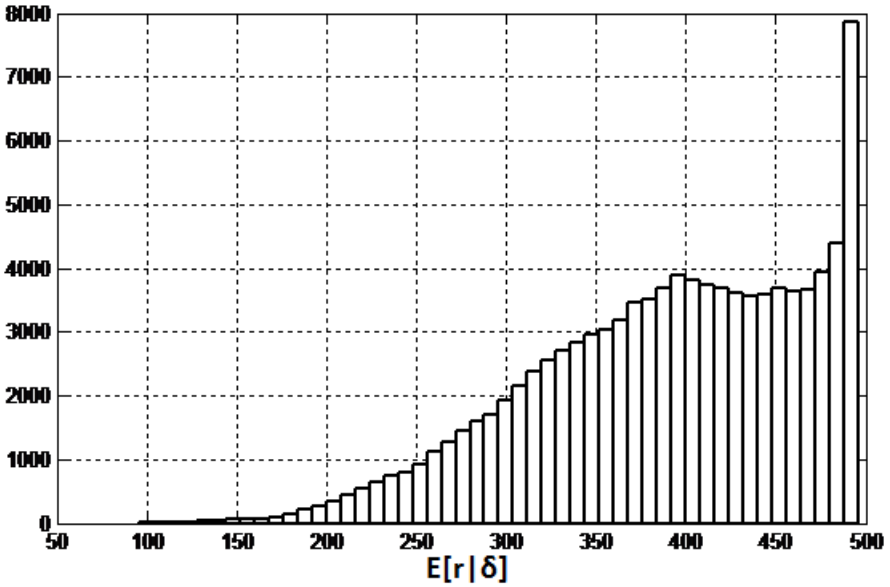


99.73% Equal-tail Interval:
$E(r|data) = 385.943$, $Median(r|data) = 261.27$
99.73% HPD Interval:
$E(r|data) = 347.625$, $Median(r|data) = 238.50$

From the figure it can be seen that the expected run-length $E(r|data) = 385.943$ for the 99.73% equal-tail interval is somewhat larger than the ARL of 370 given by Kang, Lee, Seong, and Hawkins (2007). The median run-length, $Median(r|data) = 261.27$ on the other hand is smaller than the mean run-length. This is clear from the skewness of the distribution. From Figure 3.4 it is clear that there is some difference between the mean and median obtained from an equal-tail interval and HPD interval.

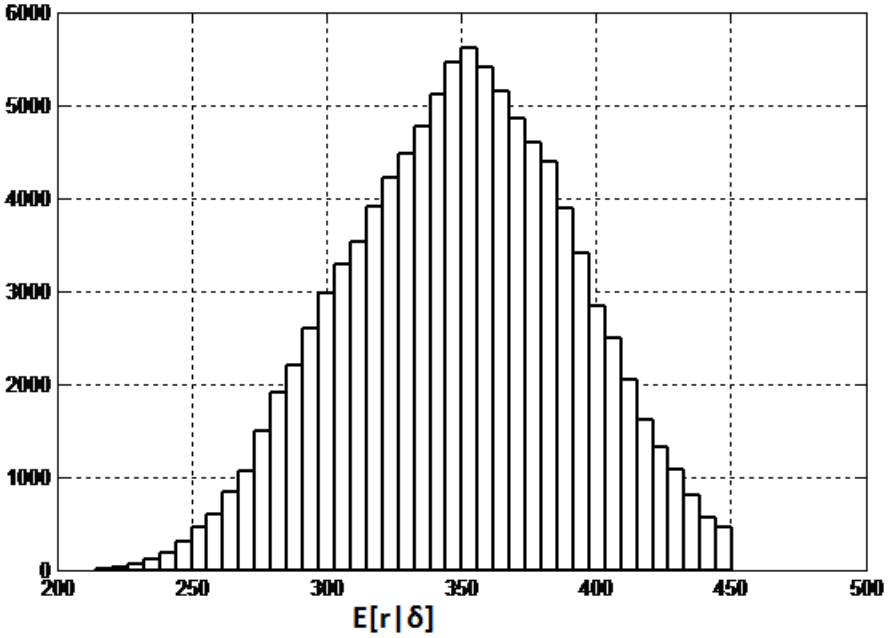
Figure 3.5 illustrates the distribution of $E(r|data)$ for each simulated value of δ, i.e., the distribution of the expected run-length in the case of the 99.73% equal-tail interval while Figure 3.6 displays the distribution of the expected run-length for the 99.73% HPD interval.

Figure 3.5.: Distribution of the Expected Run-Length, Equal-tail Interval, $n = 5$



$Mean = 389.63$ and $Median = 398.21$

Figure 3.6.: Distribution of the Expected Run-Length, HPD Interval, $n = 5$



$Mean = 349.505$ and $Median = 350.711$

As it should be the corresponding mean run-lengths in Figure 3.4 are for all practical

purposes the same as the mean run-lengths in Figures 3.5 and 3.6.

Further descriptive statistics for Figure 3.2 to Figure 3.6 are given in Table 3.2.

Table 3.2.: Descriptive Statistics for Medical Data

	Figure					
Descriptive Statistics	3.2 $p(\delta \mid data)$	3.3 $f(V \mid data)$	3.4 $f(r \mid data)$ Equal Tail	3.5 Exp Run Length Equal Tail	3.4 $f(r \mid data)$ HPD Limits	3.6 Exp Run Length HPD Limits
Mean	13.2984	16.6671	385.943	389.630	347.625	349.505
Median	13.2733	14.276	261.27	398.210	238.500	350.711
Mode	13.230	11.910	-	-	-	-
Variance	0.2296	75.0494	$1.5538e^5$	$5.6568e^3$	$1.2292e^5$	$1.7614e^3$
95% Equal Tail	(12.4056; 14.2805)	(7.661; 37.540)	(8.41; 1469.42)	(226.052; 494.735)	(7.70; 1298.75)	(267.230; 428.919)
95% HPD	(12.3796; 14.2500)	(6.830; 32.703)	(0; 1183.63)	(252.514; 495.817)	(0; 1050.56)	(268.621; 430.002)
99.73% Equal Tail	-	(6.212; 83.365)	-	-	-	-
99.73% HPD	-	(5.179; 67.777)	-	-	-	-

3.5. Simulation Study

In this section a simulation study will be conducted to observe if the 95% Bayesian confidence intervals for δ have the correct frequentist coverage.

For the simulation study the following combinations of parameters will be used:

μ_i	10	20	30	40	50	60	$\cdots$	1000	1010	1020	$\cdots$	1050
σ_i	0.75	1.5	2.25	3.0	3.75		$\cdots$	75		$\cdots$		78.15

which means that $\delta = \frac{\mu_i}{\sigma_i} = 13.3333$, $\gamma = 0.075$, $i = 1, 2, \ldots, \ddot{m}$ and $\ddot{m} = 105$. These parameter combinations are representative of the parameter values of the medical dataset on patients undergoing organ transplantation analyzed by Kang, Lee, Seong, and Hawkins (2007). As mentioned, the dataset consist of $\ddot{m} = 105$ patients and the number assays obtained for each patient is $n = 5$. As a common estimate for γ, the weighted mean square $\hat{\gamma} = 0.075$ was used.

For the above given parameter combination a dataset can be simulated consisting of $\ddot{m}$ samples and $n = 5$ observations per sample. However since we are only interested in the sufficient statistics $\bar{X}_i$ and S_i these can be simulated directly, namely $\bar{X}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{n}\right)$ and $S_i^2 \sim \frac{\sigma_i^2 \chi^2_{n-1}}{n-1}$.

The simulated $\bar{X}_i$ and S_i^2 $(i = 1, 2, \ldots, \ddot{m})$ values are then substituted in the conditional posterior distributions given in Equation (3.3) and Equation (3.4). By using the conditional posterior distributions and Gibbs sampling the unconditional posterior distribution $p(\delta|data)$ can be obtained. A confidence interval for δ will be calculated as follows: Simulate $l = 10,000$ values of δ and sort the values in ascending order $\tilde{\delta}_{(1)} \leq \tilde{\delta}_{(2)} \leq \cdots \leq \tilde{\delta}_{(l)}$.

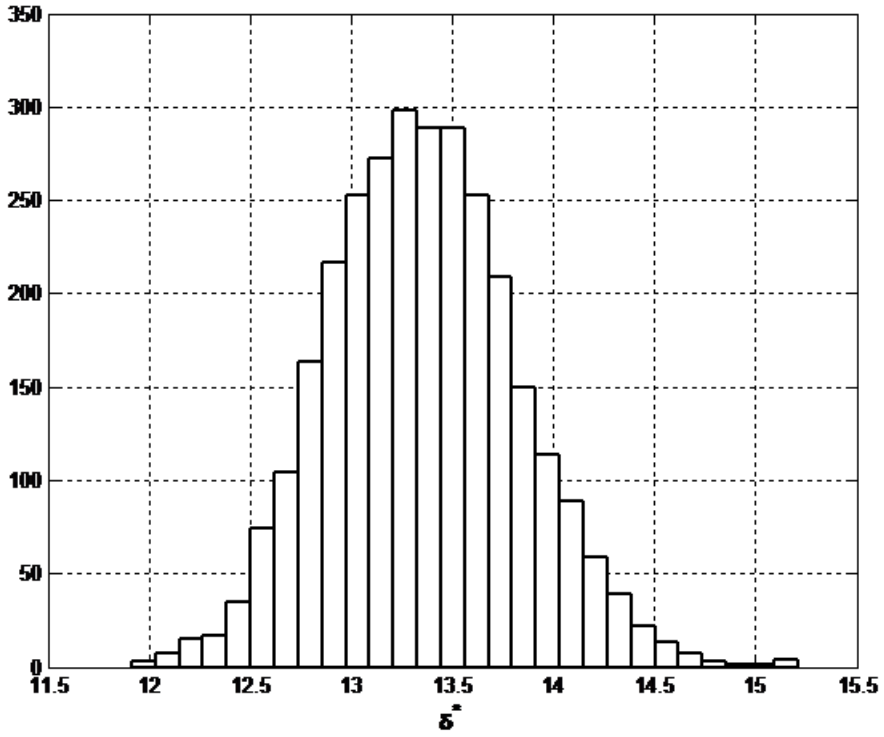
Let $K_1 = \left[\frac{\alpha}{2} l\right]$ and $K_2 = \left[\left(1 - \frac{\alpha}{2}\right) l\right]$ where $[a]$ denotes the largest integer not greater than a. $\left\{\tilde{\delta}_{(K_1)}, \tilde{\delta}_{(K_2)}\right\}$ is then a $100(1-\alpha)\%$ Bayesian confidence interval for δ. By repeating the procedure for $R = 3,000$ datasets it is found that the $3,000$, 95% Bayesian confidence intervals ($\alpha = 0.05$) cover the true parameter value $\delta = 13.333$ in $2,841$ cases.

An estimate of the frequentist probability of coverage is therefore $P\left\{\tilde{\delta}_{(K_1)} \leq \delta \leq \tilde{\delta}_{(K_2)}\right\} = 0.9470$. Also, $P\left\{\delta \leq \tilde{\delta}_{(K_1)}\right\} = 0.0313$ and $P\left\{\delta \geq \tilde{\delta}_{(K_2)}\right\} = 0.0217$.

For each dataset the posterior mean, δ^* of the $l = 10,000$ simulated δ values is calculated as well as $d = 13.3333 - \delta^*$, the difference between the posterior mean

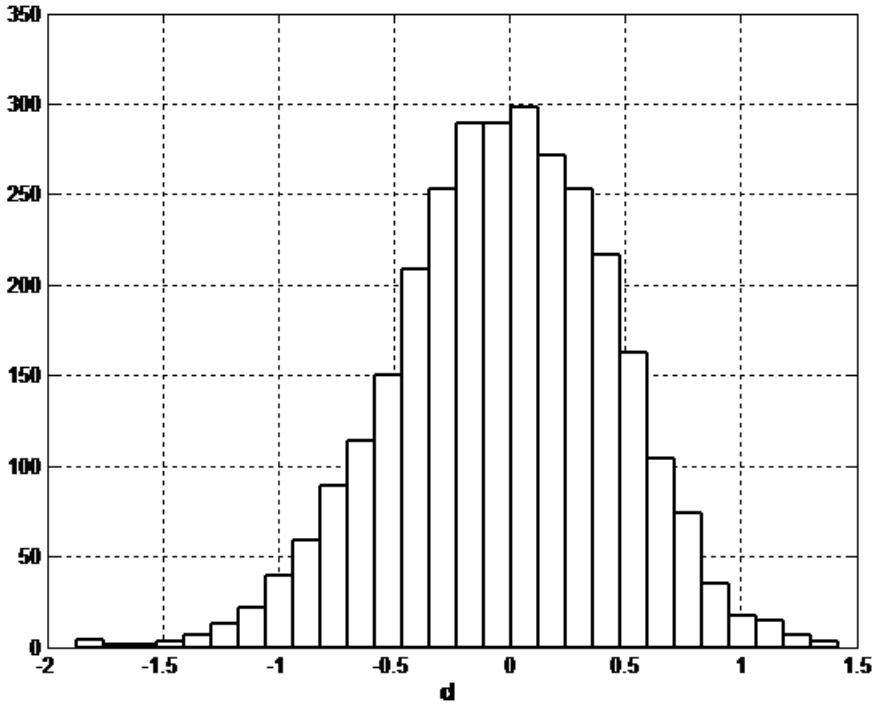
and the true parameter value. The histograms of the $R = 3,000$ δ^* and d values are displayed in Figure 3.7 and Figure 3.8. The histogram of the posterior means δ^* (Figure 3.7) is for all practical purposes symmetrical. For the d values (Figure 3.8), the histogram is slightly skew to the left.

Figure 3.7.: Histogram of Posterior Means δ^*



$Mean\,(\delta^*) = 13.356$; $Median\,(\delta^*) = 13.3439$; $Var\,(\delta^*) = 0.2177$; 95% Interval = $(12.492; 14.298)$

Figure 3.8.: Histogram of $d = \delta - \delta^*$



$Mean\,(d) = -0.0227$; $Median\,(d) = -0.0106$; $Var\,(d) = 0.2177$; 95% Interval = $(-0.968; 0.838)$

3.6. Conclusion

This chapter develops a Bayesian control chart for monitoring a common standardized mean across a range of sample values. In the Bayesian approach prior knowledge about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution. By using the posterior distribution the predictive distribution of a future standardized mean can be obtained.

Determination of reasonable non-informative priors in multi-parameter problems is not an easy task. The Jeffreys' prior for example can have a bad effect on the posterior distribution. Reference and probability-matching priors are therefore derived for a constant standardized mean across a range of sample values. The theory

and results are applied to a real problem of patients undergoing organ transplantation for which Cyclosporine is administered. This problem is discussed in detail by Kang, Lee, Seong, and Hawkins (2007). The 99.73% equal-tail prediction interval of a future coefficient of variation (inverse of the standardized mean) is effectively identical to the lower and upper control chart limits calculated by Kang, Lee, Seong, and Hawkins (2007). A simulation study shows that the 95% Bayesian confidence intervals for δ has the correct frequentist coverage.

The example illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, prediction intervals and run-lengths.

Mathematical Appendix to Chapter 3

Proof of Theorem 3.1

The likelihood function is given by

$$L\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2 | data\right) \propto \prod_{i=1}^{\ddot{m}} \left(\sigma_i^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_i^2}\left[n\left(\bar{x}_i - \delta\sigma_i\right)^2 + (n-1)\, S_i^2\right]\right\}$$

where $\bar{x}_i = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ and $(n-1)\, S_i^2 = \sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2$.

By differentiating the log likelihood function $\tilde{l}$ twice with respect to the unknown parameters and taking expected values, the Fisher Information matrix can be obtained.

$$\tilde{l} \propto \log L\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2 | data\right) = -\frac{n}{2}\sum_{i=1}^{\ddot{m}} \log \sigma_i^2 - \frac{1}{2}\sum_{i=1}^{\ddot{m}} \frac{1}{\sigma_i^2}\left[n\left(\bar{x}_i - \delta\sigma_i\right)^2 + (n-1)\, S_i^2\right]$$

and

$$\frac{\partial^2 \tilde{l}}{\left(\partial\sigma_i^2\right)^2} = \frac{n}{2}\left(\frac{1}{\sigma_i^2}\right)^2 - \frac{2n\bar{x}_i^2}{2\left(\sigma_i^2\right)^3} + \frac{3n\bar{x}_i\delta}{4\sigma_i^5} - \frac{(n-1)\, S_i^2}{\left(\sigma_i^2\right)^3}.$$

Therefore

$$-E\left[\frac{\partial^2\tilde{l}}{(\partial\sigma_i^2)^2}\right] = \frac{n}{2}\left(\frac{1}{\sigma_i^2}\right)^2\left\{1+\frac{1}{2}\delta^2\right\} \text{ where } i = 1, 2, \ldots \ddot{m}.$$

Also

$$-E\left(\frac{\partial^2\tilde{l}}{\partial\sigma_i^2\partial\sigma_l^2}\right) = 0.$$

Further

$$\frac{\partial^2\tilde{l}}{(\partial\delta)^2} = -\frac{1}{2}\sum_{i=1}^{\ddot{m}}\frac{1}{\sigma_i^2}\left(2n\sigma_i^2\right) = -n\ddot{m}$$

$$\therefore -E\left(\frac{\partial^2\tilde{l}}{(\partial\delta)^2}\right) = n\ddot{m}.$$

If we differentiate $\tilde{l}$ with respect to δ and σ_i^2 we get

$$\frac{\partial^2\tilde{l}}{\partial\delta\partial\sigma_i^2} = \frac{-n\bar{x}_i}{2\sigma_i^3}$$

and

$$-E\left(\frac{\partial^2\tilde{l}}{\partial\delta\partial\sigma_i^2}\right) = \frac{n\delta}{2\sigma_i^2} \; i = 1, 2, \ldots, \ddot{m}.$$

The Fisher Information matrix now follows as

$$F\left(\underline{\theta}\right) = F\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}$$

where

$$F_{11} = n\ddot{m},\; F_{12} = F_{21}' = \left[\frac{n\delta}{2\sigma_1^2} \quad \frac{n\delta}{2\sigma_2^2} \quad \cdots \quad \frac{n\delta}{2\sigma_{\ddot{m}}^2}\right]$$

and

$$F_{22} = \begin{bmatrix} \frac{n}{2}\left(\frac{1}{\sigma_1^2}\right)^2\left\{1+\frac{1}{2}\delta^2\right\} & 0 & \cdots & 0 \\ 0 & \frac{n}{2}\left(\frac{1}{\sigma_2^2}\right)^2\left\{1+\frac{1}{2}\delta^2\right\} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n}{2}\left(\frac{1}{\sigma_{\ddot{m}}^2}\right)^2\left\{1+\frac{1}{2}\delta^2\right\} \end{bmatrix}.$$

To calculate the reference prior for the ordering $\{\delta, (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2)\}$, $F_{11\cdot 2}$ must first be calculated and then $|F_{22}|$. Now

$$F_{11\cdot 2} = F_{11} - F_{12}F_{22}^{-1}F_{21} = n\ddot{m} - \frac{1}{2}n\ddot{m}\delta^2\left(1+\frac{1}{2}\delta^2\right)^{-1} = \frac{n\ddot{m}}{\left(1+\frac{1}{2}\delta^2\right)} = h_1$$

and

$$p(\delta) \propto h_1^{\frac{1}{2}} \propto \left(1+\frac{1}{2}\delta^2\right)^{-\frac{1}{2}}.$$

Also

$$|F_{22}| = \left\{\frac{n}{2}\left(1+\frac{1}{2}\delta^2\right)\right\}^{\ddot{m}} \prod_{i=1}^{\ddot{m}}\left(\frac{1}{\sigma_i^2}\right)^2 = h_2.$$

This means that

$$p\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2|\delta\right) \propto h_2^{\frac{1}{2}} = \prod_{i=1}^{\ddot{m}}\left(\frac{1}{\sigma_i^2}\right).$$

Therefore the reference prior for the ordering $\{\delta, (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2)\}$ is

$$p_R\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) = p(\delta)\, p\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2|data\right) \propto \left(1+\frac{1}{2}\delta^2\right)^{-\frac{1}{2}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2}.$$

Proof of Theorem 3.2

The inverse of the Fisher Information matrix is given by

$$F^{-1}\left(\underline{\theta}\right) = F^{-1}\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) = \begin{bmatrix} F^{11} & F^{12} & F^{13} & \cdots & F^{1,\ddot{m}+1} \\ F^{21} & F^{22} & F^{23} & \cdots & F^{2,\ddot{m}+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F^{\ddot{m}+1,1} & F^{\ddot{m}+1,2} & F^{\ddot{m}+1,3} & \cdots & F^{\ddot{m}+1,\ddot{m}+1} \end{bmatrix}.$$

Let

$$t\left(\underline{\theta}\right) = t\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) = \delta.$$

Since

$$\nabla_t^{'}\left(\underline{\theta}\right) = \begin{bmatrix} \frac{\partial}{\partial\delta} t\left(\underline{\theta}\right) & \frac{\partial}{\partial\sigma_1^2} t\left(\underline{\theta}\right) & \cdots & \frac{\partial}{\partial\sigma_{\ddot{m}}^2} t\left(\underline{\theta}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

we have that

$$\nabla_t^{'}\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right) = \begin{bmatrix} F^{11} & F^{12} & \cdots & F^{1,\ddot{m}+1} \end{bmatrix} = \begin{bmatrix} \frac{2+\delta^2}{2\ddot{m}n} & \frac{-\delta\sigma_1^2}{\ddot{m}n} & \frac{-\delta\sigma_2^2}{\ddot{m}n} & \cdots & \frac{-\delta\sigma_{\ddot{m}}^2}{\ddot{m}n} \end{bmatrix}$$

and

$$\sqrt{\nabla_t^{'}\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right) \nabla_t\left(\underline{\theta}\right)} = \left(\frac{2+\delta^2}{2\ddot{m}n}\right)^{\frac{1}{2}}.$$

Further

$$\Upsilon^{'}\left(\underline{\theta}\right) = \frac{\nabla_t^{'}\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right)}{\sqrt{\nabla_t^{'}\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right) \nabla_t\left(\underline{\theta}\right)}} = \begin{bmatrix} \Upsilon_1\left(\underline{\theta}\right) & \Upsilon_2\left(\underline{\theta}\right) & \cdots & \Upsilon_{\ddot{m}}\left(\underline{\theta}\right) \end{bmatrix}$$

where

$$\Upsilon_1\left(\underline{\theta}\right) = \left(\frac{2+\delta^2}{2\ddot{m}n}\right)^{\frac{1}{2}},$$

$$\Upsilon_2(\underline{\theta}) = \frac{-\sqrt{2}\delta\sigma_1^2}{(\ddot{m}n)^{\frac{1}{2}}(2+\delta^2)^{\frac{1}{2}}},$$

$$\Upsilon_3(\underline{\theta}) = \frac{-\sqrt{2}\delta\sigma_2^2}{(\ddot{m}n)^{\frac{1}{2}}(2+\delta^2)^{\frac{1}{2}}}$$

and

$$\Upsilon_{m+1}(\underline{\theta}) = \frac{-\sqrt{2}\delta\sigma_{\ddot{m}}^2}{(\ddot{m}n)^{\frac{1}{2}}(2+\delta^2)^{\frac{1}{2}}}.$$

The prior

$$p_M(\underline{\theta}) = p_M\left(\delta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_{\ddot{m}}^2\right) \propto \frac{1}{(2+\delta^2)^{\frac{1}{2}}} \prod_{i=1}^{\ddot{m}} \sigma_i^{-2}$$

is therefore a probability matching prior since

$$\frac{\partial}{\partial\delta}\left\{\Upsilon_1(\underline{\theta})\, p_M(\underline{\theta})\right\} + \frac{\partial}{\partial\sigma_1^2}\left\{\Upsilon_2(\underline{\theta})\, p_M(\underline{\theta})\right\} + \cdots + \frac{\partial}{\partial\sigma_{\ddot{m}}^2}\left\{\Upsilon_{\ddot{m}+1}(\underline{\theta})\, p_M(\underline{\theta})\right\} = 0.$$

Proof of Theorem 3.3

The joint posterior distribution given in Equation (3.2) can be written as

$$p(\delta, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}|data) \propto p(\delta, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}|data) \propto \left(1+\tfrac{1}{2}\delta^2\right)^{-\frac{1}{2}} \prod_{i=1}^{\ddot{m}} \exp\left[-\tfrac{n\delta^2}{2}\left(1-\tfrac{\bar{x}_i^2}{D_i^2}\right)\right]$$
$$\times \left(\tfrac{1}{\sigma_i}\right)^{n+1} \times \exp\left[-\tfrac{n}{2}D_i^2\left(\tfrac{1}{\sigma_i} - \tfrac{\bar{x}_i\delta}{D_i^2}\right)\right]$$

where $\bar{x}_i = \frac{1}{n}\sum_{j=1}^n x_{ij}$ and $D_i^2 = \frac{1}{n}\sum_{j=1}^n x_{ij}^2$.

Therefore

$$p(\delta|data) = \int_0^\infty \cdots \int_0^\infty p(\delta, \sigma_1, \sigma_2, \ldots, \sigma_{\ddot{m}}|data)\, d\sigma_1 d\sigma_2 \ldots d\sigma_{\ddot{m}}$$

$$\propto \left(1+\frac{1}{2}\delta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{n\delta^2}{2}\sum_{i=1}^{\ddot{m}}\left(1-\frac{\bar{x}_i^2}{D_i^2}\right)\right\}\prod_{i=1}^{\ddot{m}}\left\{\left(\frac{1}{\sigma_i}\right)^{n+1}\exp\left[-\frac{n}{2}D_i^2\left(\frac{1}{\sigma_i}-\frac{\bar{x}_i\delta}{D_i^2}\right)^2\right]\right\}$$

which is proper. As mentioned by Berger, Liseo, and Wolpert (1999) it is usually the case that the reference and Jeffreys' priors will yield proper posterior distributions.

4. Bayesian Control Charts for the Variance and Generalized Variance for the Normal Distribution

4.1. Introduction

Quality control is a process which is used to maintain the standards of products produced or services delivered. As mentioned by Human, Chakraborti, and Smit (2010) monitoring spread is important in practice since it is an indicator of process quality and also the spread must be first monitored before the mean control chart should be constructed and examined. If any of the sample variances plot on or outside the control limits, they should be examined and if they are discarded, revised values should be calculated for the estimators as well as for the control limits. Although the aim of this chapter is to monitor sample variances and generalized variances that are too large, i.e., upper one-sided control limits, it is also important to look at samples with very small variances, i.e., two-sided control-charts. Observations with very small variances or no variance at all seem suspicious and could have been tampered with (i.e., artificially changed) to suit the quality control procedure.

It is nowadays commonly accepted by most statisticians that statistical process control should be implemented in two phases:

1. Phase I where the primary interest is to assess process stability; and
2. Phase II where online monitoring of the process is done.

As in Human, Chakraborti, and Smit (2010) the Phase I control charts will be constructed using the false alarm probability (FAP) which is the overall probability of at least one false alarm. The nominal false alarm probability value that will be used is $FAP_0 = 0.05$. The main difference between our method and that of Human,

Chakraborti, and Smit (2010) is in Phase II where we are using a Bayesian procedure to study the in-control run-length.

In this chapter, control chart limits will be determined for the sample variance, S^2, and the generalized variance $|S|$. Average run-lengths and false alarm rates will also be calculated in the Phase II setting, using a Bayesian predictive distribution.

4.2. An Example

The data presented in Table 4.1 represents measurements of inside diameters and represent the number of 0.0001 inches above 0.7500 inches (Duncan (1965)). The measurements are taken in samples of $j = 1, 2, \ldots, n$ each ($n = 5$) over time. Also shown in Table 4.1 are the sample variances, S_i^2 for $i = 1, 2, \ldots m$ samples ($m = 10$). These data will be used to construct a Shewart type Phase I upper control chart for the variance, and also to calculate the run-length for future samples of size $n = 5$ taken repeatedly from the process.

From the data in Table 4.1 the sample variances are calculated by

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^{m} (y_{ij} - \bar{y}_i)^2 .$$

The pooled sample variance is then determined as

$$S_p^2 = \frac{1}{m} \sum_{i=1}^{m} S_i^2 = 10.72.$$

Table 4.1.: Data for Constructing a Shewart-Type Phase I Upper Control Chart for the Variance

Sample Number/ Time (i)	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	s_i^2
1	15	11	8	15	6	16.5
2	14	16	11	14	7	12.3
3	13	6	9	5	10	10.3
4	15	15	9	15	7	15.2
5	11	14	11	12	5	11.3
6	13	12	9	6	10	7.5
7	10	15	12	4	6	19.8
8	9	12	9	8	8	2.7
9	8	12	14	9	10	5.8
10	10	10	9	14	14	5.8

4.3. Statistical Calculation of the Control Limits in Phase I

The upper control limit, using the data by Duncan (1965) will be obtained as described by Human, Chakraborti, and Smit (2010).

It is well known that

$$\frac{(n-1)\,S_i^2}{\sigma^2} \sim \chi^2_{n-1}$$

Also, if the underlying distribution is Normal,

$$\frac{m\,(n-1)\,S_p^2}{\sigma^2} \sim \chi^2_{m(n-1)} = \sum_{i=1}^{m} \chi^2_{n-1}.$$

Define

$$Y_i = \frac{(n-1)\,S_i^2/\sigma^2}{m\,(n-1)\,S_p^2/\sigma^2} = \frac{X_i}{\sum_{i=1}^{m} X_i}$$

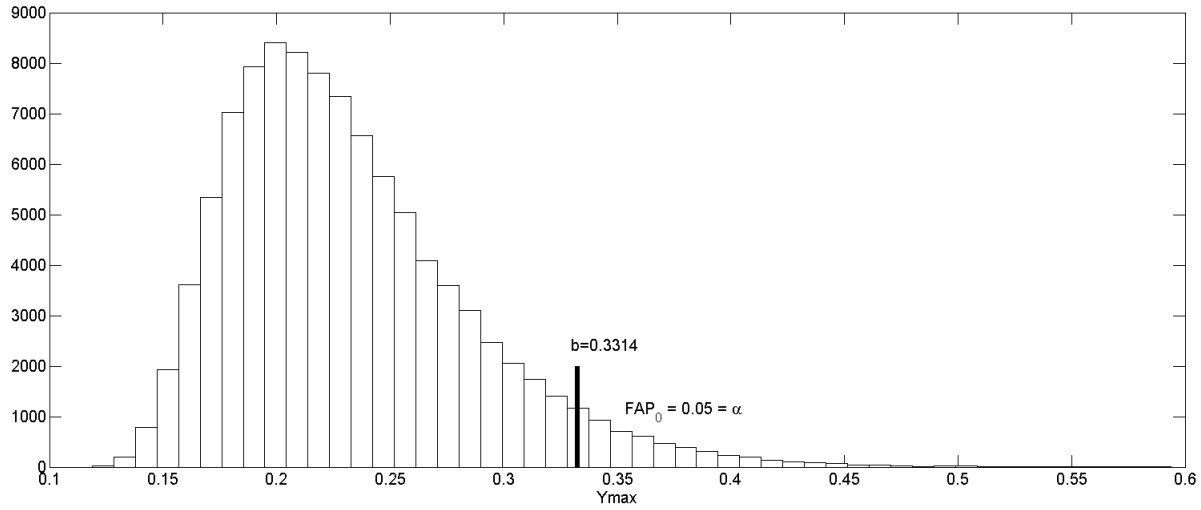
where $X_i \sim \chi^2_{n-1}$ $(i = 1, 2, \ldots, m)$.

The distribution of $Y_{max} = max\,(Y_1, Y_2, \ldots, Y_m)$ obtained from $100,000$ simulations is illustrated in Figure 4.1. The value b is then calculated such that the false alarm probability is at a level of 0.05 (also shown in the figure).

The upper control limit is then determined as:

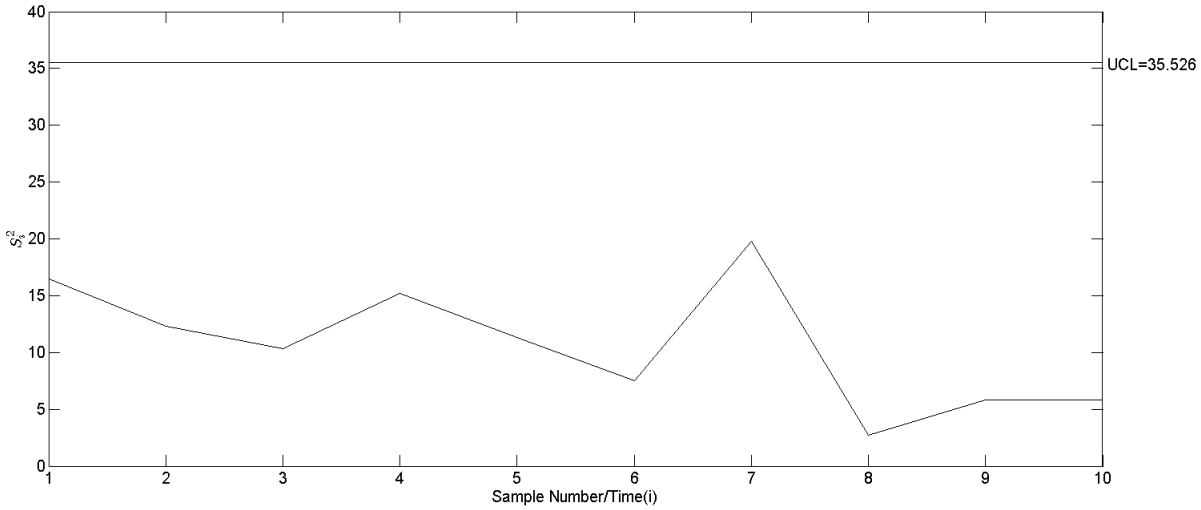
$$UCL = mbS_p^2 = 10\,(0.3314)\,(10.72) = 35.526.$$

Figure 4.1.: Distribution of $Y_{max} = max\,(Y_1, Y_2, \ldots, Y_m)$ (100,000 simulations)



The data from Duncan (1965) are presented visually in Figure 4.2. The figure includes the upper control limit as determined above.

Figure 4.2.: Shewart-type Phase I Upper Control Chart for the Variance - $FAP_0 = 0.05$



In the case of a two-sided control chart the joint distribution of $Y_{min} = min\,(Y_1, Y_2, \ldots, Y_m)$ and $Y_{max} = max\,(Y_1, Y_2, \ldots, Y_m)$ must first be simulated. The equal-tail values given in Table 2, Page 868 of Human, Chakraborti, and Smit (2010) are calculated in such a way that the false alarm probability does not exceed 0.05. For $m = 10$ and $n = 5$ it follows that the lower control limit, $LCL = ma_1 S_p^2 = (10)\,(0.0039)\,(10.72) = 0.4181$ and the upper control limit, $UCL = mb_1 S_p^2 = (10)\,(0.3599)\,(10.72) = 38.581$. Since the correlation coefficient between Y_{min} and Y_{max} for $m = 10$ and $n = 5$ is -0.2731, the shortest two-sided control limits such that the false alarm probability does not exceed 0.05 is given by $L\tilde{C}L = (10)\,(0.0002)\,(10.72) = 0.0214$ and $U\tilde{C}L = (10)\,(0.3314)\,(10.72) = 35.5261$. This interval is 7% shorter than the equal-tail interval.

4.4. Upper Control Limit for the Variance in Phase II

In the first part of this section, the upper control limit in a Phase II setting will be derived using the Bayesian predictive distribution.

Theorem 4.1. *Assume* $Y_{ij} \sim^{iid} N\,(\mu_i, \sigma^2)$ *where* Y_{ij} *denotes the* j^{th} *observation from the* i^{th} *sample where* $i = 1, 2, \ldots, m$ *and* $j = 1, 2, \ldots, n$. *The mean* μ_i *and variance* σ^2 *are unknown.*

Using the Jeffreys' prior $p(\mu_1,\mu_2,\ldots,\mu_m,\sigma^2) \propto \sigma^{-2}, \sigma^2 > 0, -\infty < \mu_i < \infty, i = 1,2,\ldots,m$ *it can be proved that the posterior distribution of* σ^2 *is given by*

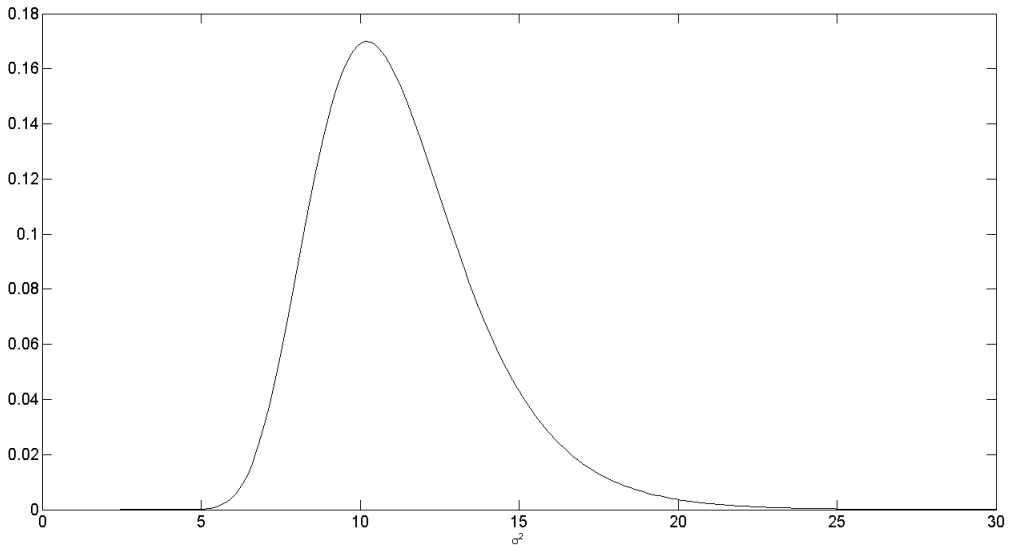
$$p\left(\sigma^2|data\right) = \left(\frac{\tilde{S}}{2}\right)^{\frac{1}{2}k} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}(k+2)} \exp\left(-\frac{\tilde{S}}{2\sigma^2}\right), \sigma^2 > 0 \tag{4.1}$$

an Inverse Gamma distribution with $k = m(n-1)$ *and* $\tilde{S} = m(n-1)S_p^2$.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

The posterior distribution given in Equation 4.1 is presented in Figure 4.3.

Figure 4.3.: Distribution of $p(\sigma^2|data)$-Simulated Values



A predictive distribution derived using a Bayesian approach will be used to obtain the control limits in a Phase II setting. Let S_f^2 be the sample variance of a future sample of n observations from the Normal distribution. Then for a given σ^2 it follows that

$$\frac{(n-1)\,S_f^2}{\sigma^2} = \frac{vS_f^2}{\sigma^2} \sim \chi_v^2$$

which means that

$$f\left(S_f^2|\sigma^2\right) = \left(\frac{v}{2\sigma^2}\right)^{\frac{1}{2}v} \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(S_f^2\right)^{\frac{1}{2}v-1} \exp\left(-\frac{vS_f^2}{2\sigma^2}\right) \tag{4.2}$$

Theorem 4.2. *If S_f^2 is the sample variance of a future sample of n observations from the Normal distribution then the unconditional predictive density of S_f^2 is given by*

$$f\left(S_f^2|data\right) = S_p^2 F_{n-1,m(n-1)} \tag{4.3}$$

where S_p^2 is the pooled sample variance and $F_{n-1,m(n-1)}$ the F-distribution with $n-1$ and $m\,(n-1)$ degrees of freedom.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

The upper control limit in the Phase II setting is then derived as

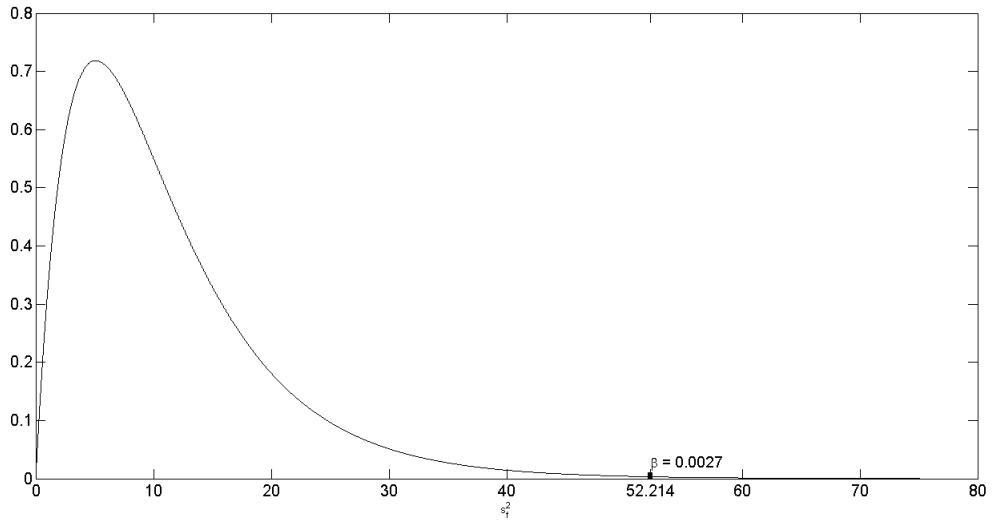
$$S_p^2 F_{n-1,m(n-1)}\,(1-\beta)\,.$$

At $\beta = 0.0027$ we therefore obtain the upper control limit as

$$S_p^2 F_{n-1,m(n-1)}\,(0.9973) = (10.72)\,(4.8707) = 52.214.$$

The distribution of the predictive density of S_f^2 including the derived upper control limit is presented in Figure 4.4.

Figure 4.4.: Distribution of $f\left(S_f^2|data\right)$



Assuming that the process remains stable, the predictive distribution for S_f^2 can also be used to derive the distribution of the run-length, that is the number of samples until the control chart signals for the first time.

The resulting rejection region of size β using the predictive distribution for the determination of the run-length is defined as

$$\beta = \int_{R(\beta)} f\left(S_f^2|data\right) dS_f^2$$

where

$$R\left(\beta\right) = (52.214, \infty),$$

is the upper one-sided control limit.

Given σ^2 and a stable process, the distribution of the run-length r is Geometric with parameter

$$\psi\left(\sigma^2\right) = \int_{R(\beta)} f\left(S_f^2|\sigma^2\right) dS_f^2 \tag{4.4}$$

where $f\left(S_f^2|\sigma^2\right)$ given in Equation (4.2) is the predictive distribution of a future sample variance given σ^2.

The value of the parameter σ^2 is however unknown and its uncertainty is described by the posterior distribution $p\left(\sigma^2|data\right)$ defined in Equation (4.1).

Theorem 4.3. *For a given σ^2 the parameter of the Geometric distribution is*

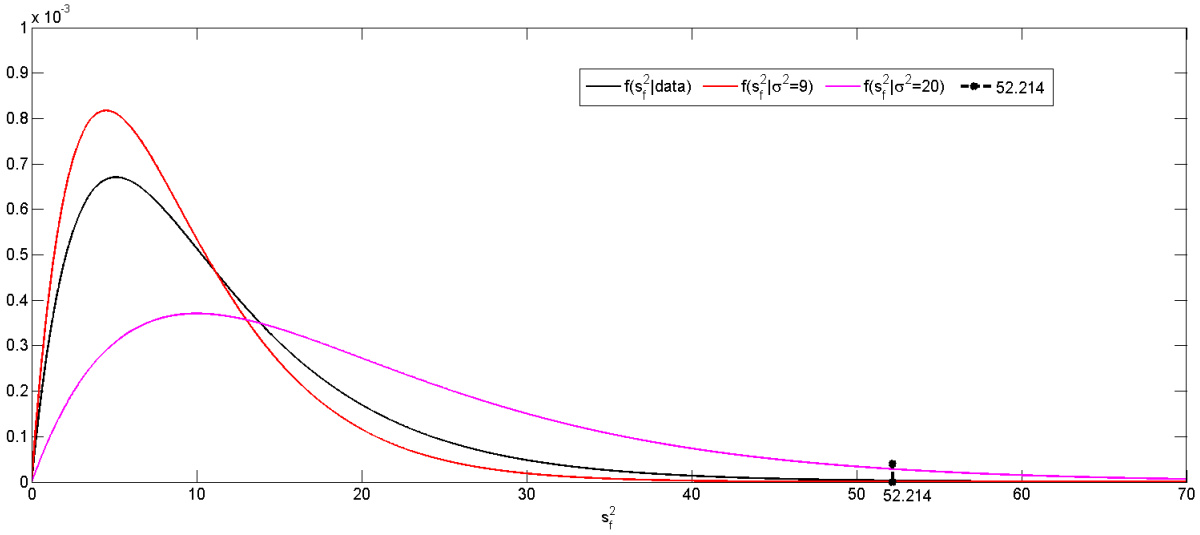
$$\psi\left(\sigma^2\right) = \psi\left(\chi^2_{m(n-1)}\right) \text{ for given } \chi^2_{m(n-1)}$$

which means that it is only dependent on $\chi^2_{m(n-1)}$ and not on σ^2.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

In Figure 4.5 the distributions of $f\left(S_f^2|data\right)$ and $f\left(S_f^2|\sigma^2\right)$ for $\sigma^2 = 9$ and $\sigma^2 = 20$ are presented to show the different shapes of the applicable distributions.

Figure 4.5.: Distributions of $f\left(S_f^2|\sigma^2\right)$ and $f\left(S_f^2|data\right)$ showing $\psi\left(\sigma_1^2\right)$



As mentioned, by simulating σ^2 from $p\left(\sigma^2|data\right)$ the probability density function $f\left(S_f^2|\sigma^2\right)$ as well as the parameter $\psi\left(\sigma^2\right)$ can be obtained. This must be done for each future sample. Therefore by simulating a large number of σ^2 values from the posterior distribution a large number of $\psi\left(\sigma^2\right)$ values can be obtained. A large number of geometric and run-length distributions with different parameter values ($\psi\left(\sigma_1^2\right)$, $\psi\left(\sigma_2^2\right)$,...,$\psi\left(\sigma_l^2\right)$) will therefore be available. The unconditional run-length distribution is obtained by using the Rao-Blackwell method, i.e., the average of the conditional run-length distributions.

In Table 4.2(a) results for the run-length at $\beta = 0.0027$ for $n = 5$ and different values for m are presented for the upper control limit for the variance. The table present the mean, median, 95% equal-tail interval and calculated β value to obtain a run-length of 370 (the expected run-length at $\beta = 0.0027$ is $\frac{1}{0.0027} \approx 370$ if σ^2 is known).

In the case of the diameter example the mean run-length is 29754 and the median run-length 1354. The reason for these large values is the uncertainty in the parameter estimate because of the small sample size and number of samples ($n = 5$ and $m = 10$). To get a mean run-length of 370, β must be 0.0173 instead of 0.0027.

For a two-sided control chart, the upper control limit is $UCL = S_p^2 F_{n-1,m(n-1)}\left(1 - \frac{\beta}{2}\right)$ and the lower control limit is $LCL = S_p^2 F_{n-1,m(n-1)}\left(\frac{\beta}{2}\right)$. $R\left(\beta\right)$ are all those values larger than UCL and smaller than LCL. For the diameter example $UCL =$

$(10.72)(5.4445) = 58.365$ and $LCL = (10.72)(0.02583) = 0.2769$. In this case the mean run-length is 500.

In Table 4.2(b) results for the run-length at $\beta = 0.0027$ for $n = 5$ and different values of m are presented for a two-sided control chart.

Table 4.2.: Mean and Median Run-length at $\beta = 0.0027$ for $n = 5$ and Different Values of m

(a) Upper Control Limit

m	n	Mean	Median	95% Equal-tail Interval	Calculated β for Mean Run-length of 370
10	5	29 754	1 354	(54;117 180)	0.0173
50	5	654	470	(121;2 314)	0.0044
100	5	482	411	(156;1 204)	0.0035
200	5	422	391	(197;829)	0.0031
500	5	389	379	(244;596)	0.0028
1 000	5	379	374	(274;517)	0.0028
5 000	5	371	370	(322;428)	0.0027
10 000	5	370	370	(335;410)	0.0027

(b) Upper and Lower Control Limit

m	n	Mean	Median	95% Equal-tail Interval
10	5	500	552	(92;661)
50	5	399	427	(184;492)
100	5	385	398	(227;474)
200	5	377	383	(263;459)
500	5	373	375	(300;432)
1 000	5	371	372	(320;416)
5 000	5	370	370	(347;391)
10 000	5	370	370	(354;385)

From Table 4.2 it can be noted that as the number of samples increase (larger m) the mean and median run-lengths converges to the expected run-length of 370.

Further, define $\bar{\psi}(\sigma^2) = \frac{1}{l}\sum_{i=1}^{l}\psi(\sigma_i^2)$. From Menzefricke (2002) it is known that if $l \to \infty$, then $\bar{\psi}(\sigma^2) \to \beta = 0.0027$ and the harmonic mean of the unconditional run-length will be $\left(\frac{1}{\beta}\right) = \frac{1}{0.0027} = 370$. Therefore it does not matter how small m and n is, the harmonic mean of the run-length will be $\frac{1}{\beta}$ if $l \to \infty$.

4.5. Phase I Control Charts for the Generalized Variance

Assume $\underline{\mathrm{Y}}_{ij} \sim^{idd} N(\underline{\mu}_i, \Sigma)$ where $\underline{\mathrm{Y}}_{ij}\,(p \times 1)$ denotes the jth observation vector from the ith sample, $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The mean vector $\underline{\mu}_i\,(p \times 1)$ and covariance matrix, $\Sigma\,(p \times p)$ are unknown.

Define $\underline{\bar{Y}}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ and $A_i = \sum_{j=1}^{n}\left(\underline{\mathrm{Y}}_{ij} - \underline{\bar{\mathrm{Y}}}_i\right)\left(\underline{\mathrm{Y}}_{ij} - \underline{\bar{Y}}_i\right)' \ (i = 1, 2, \ldots, m)$.

From this it follows that

$$\underline{\bar{Y}}_i \sim N\left(\underline{\mu}_i, \frac{1}{n}\Sigma\right), (i = 1, 2, \ldots, m),$$

$$A_i = (n-1)\,S_i \sim W_p\,(n-1, \Sigma),$$

$$A = \sum_{i=1}^{m} A_i \sim W_p\,(m\,(n-1), \Sigma)$$

and

$$S_p = \frac{1}{m\,(n-1)}A.$$

The generalized variance of the ith sample is defined as the determinant of the sample covariance matrix, i.e., $|S_i|$.

Define

$$T_i = \frac{|A_i|}{|\sum_{i=1}^{m} A_i|} = \frac{|A_i^*|}{|\sum_{i=1}^{m} A_i^*|}$$

where $A_i^* \sim W_p\left(n-1, I_p\right)$.

Also

$$T = max\left(T_1, T_2, \ldots, T_m\right) = max\left(T_i\right), i = 1, 2, \ldots, m$$

Now

$$T_i = \frac{|A_i|}{|\sum_{i=1}^{m} A_i|} = \frac{|S_i|}{m^p\,|S_p|}$$

Therefore a $(1-\beta)\,100\%$ upper control limit for $|S_i|$ $(i = 1, 2, \ldots, m)$ is $m^p\,|S_p|\,T_{1-\beta}$.

Figure 4.6 presents a histogram of 100,000 simulated values of $max\left(T_i\right)$ for the two-dimensional case (p=2,m=10 and n=6). The upper control limit as presented in Table 4.3 is given on the figure. Table 4.3 also presents the upper control limit for the one-dimensional (p=1) and the three-dimensional (p=3) situations as well as the lower control limit for all three dimensions.

Figure 4.6.: Histogram of $max\,(T_i)$-100,000 Simulations

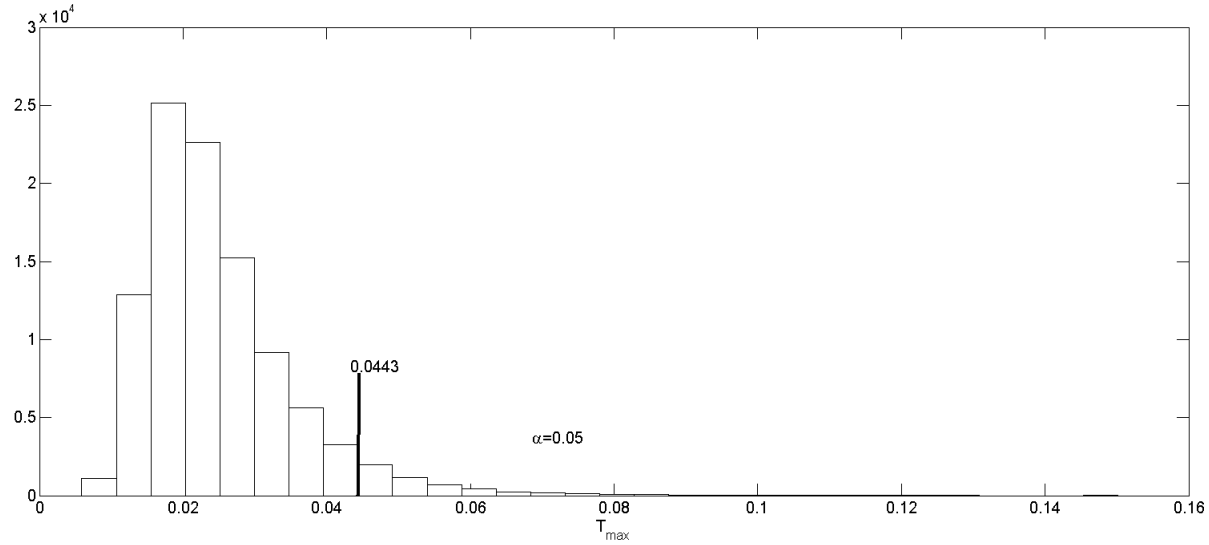


Table 4.3.: Upper 95% Control Limit, $T_{0.95}$ for $T = max\,(T_i)$ for the Generalized Variance in Phase I for $m = 10$, $n = 6$ and $p = 1, 2$ and 3

p	m	n	$T_{0.95}$	$T_{0.025}$	$T_{0.975}$
1	10	6	0.30259	0.00665	0.32655
2	10	6	0.04429	0.00014	0.05122
3	10	6	0.00445	1.822967e^{-6}	0.00544

By using a Bayesian procedure a predictive distribution will be derived to obtain control chart limits in Phase II.

Using the Jeffreys' prior

$$p\,(\underline{\mu}, \Sigma) \propto |\Sigma|^{-\frac{1}{2}(p+1)} \quad -\infty < \underline{\mu} < \infty, \Sigma > 0$$

the posterior distribution of Σ is derived as

$$|\Sigma|\,|data \sim |A| \prod_{i=1}^{p} \left(\frac{1}{\chi^2_{m(n-1)+1-i}} \right) \tag{4.5}$$

and the predictive distribution of a future sample generalized variance $|S_f|$ given Σ as

$$|S_f|\,|\Sigma \sim \left|\frac{1}{n-1}\Sigma\right| \prod_{i=1}^{p} \chi^2_{n-i} \tag{4.6}$$

By combining Equation (4.5) and Equation (4.6) the unconditional predictive distribution is given by

$$\left|S_f^2\right|\,|data \sim \left(\frac{1}{n-1}\right)^p |A| \left(\prod_{i=1}^{p} \frac{n-i}{m\,(n+1)+1-i}\right) F^* \tag{4.7}$$

where

$$F^* = \prod_{i=1}^{p} F_{n-i,m(n-1)+i-i}.$$

Equation (4.7) can be used to obtain the control chart limits.

Similarly for the variance, the rejection region of size β is defined as

$$\beta = \int_{R(\beta)} f\left(|S_f|\,|data\right) d\,|S_f|\,.$$

Given Σ and a stable process, the distribution of the run-length r is Geometric with parameter

$$\psi\left(|\Sigma|\right) = \int_{R(\beta)} f\left(|S_f|\,|\Sigma\right)$$

where $f\left(\left|S_f\right|\,|\Sigma\right)$ is given in Equation 4.6.

Theorem 4.4. *For a given value of $|\Sigma|$, and simulated from the posterior distribution described in Equation (4.5) the parameter of the Geometric distribution is*

$$\psi\left(|\Sigma|\right) = P\left\{\prod_{i=1}^{p}\chi^2_{n-i} \geq \left(\prod_{i=1}^{p}\chi^2_{m(n-1)+1-i}\right)\left(\prod_{i=1}^{p}\frac{n-i}{m\left(n-1\right)+1-i}\right)F^*_{1-\beta}\right\}$$

for a given $\prod_{i=1}^{p}\chi^2_{m(n-1)+1-i}$.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

For further details see Menzefricke (2002, 2007, 2010b,a).

Mean and median run-length results at $\beta = 0.0027$ for $n = 50$, $m = 50$ and 100 for the one-, two- and three-dimensional cases are presented in Table 4.4.

Table 4.4.: Mean and Median Run-length at $\beta = 0.0027$ for $n = 50$, $m = 50$ and 100 and $p = 1, 2$ and 3

p	m	n	Mean	Median	95% Equal Tail Interval
1	50	50	482	414	(185;1 198)
1	100	50	431	402	(197;841)
2	50	50	466	404	(162;1 128)
2	100	50	423	396	(205;819)
3	50	50	461	407	(165;1 063)
3	100	50	424	399	(209;786)

4.6. Guidelines for Practitioners for the Implementation of the Proposed Control Chart

It was mentioned in the introductory section and as by Chakraborti, Human, and Graham (2008) that it is now generally accepted that statistical process control should be implemented in two phases. Phase I is the so-called retrospective phase and Phase II the prospective or monitoring phase.

The construction of Phase I control charts should be considered as a multiple testing problem. The distribution of a set of dependent variables (ratios of chi-square

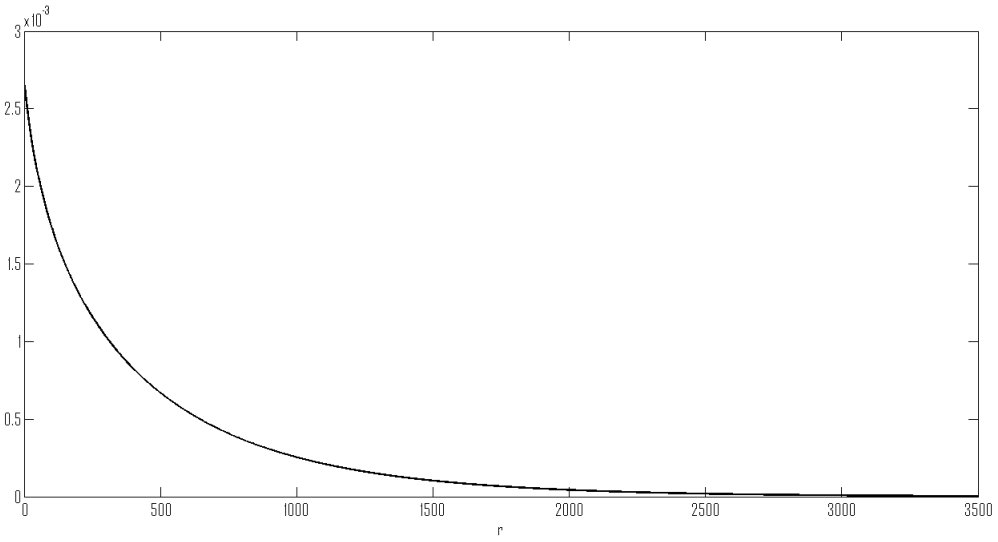
random variables) can therefore be used to calculate the control limits so that the false alarm probability is not larger than FAP_0. Tables are provided by Human, Chakraborti, and Smit (2010) for the charting constants for S^2 for each Phase I chart, for a FAP_0 of 0.01 and 0.05 respectively. These tables can easily be implemented by practitioners. Further the charting constants can also be adjusted so that the difference between the upper and lower control limit is a minimum. Tables of the "adjusted" charting constants for a FAP_0 and 0.05 can therefore also be made available to practitioners. Similar tables can be drawn up for the generalized variance.

Since the Phase I control charting problem is considered to be a multiple hypothesis testing problem, Bartlett's test can be used instead of two-sided control charts. Bartlett's test is a likelihood ratio test and it is uniformly the most powerful test for testing the homogeneity of variances. Under the null hypothesis it is the ratio of the geometric and arithmetic mean of a set of independent chi-square random variables. Tables of critical values can therefore be easily calculated. A similar test is available for the generalized variance.

Once the Phase I control chart has been calculated, statistical process control should be moved to Phase II. According to Chakraborti, Human, and Graham (2008) Phase II control chart performance should be measured in terms of some attribute of the run-length. This is exactly what the aim of this chapter is. Predictive distributions based on a Bayesian approach are used to construct Shewart-type control limits for the variance and generalized variance. By using Monte Carlo simulation methods the distributions of the run-length and average run-length can easily be obtained. It has also been shown that an increase in the number of samples m and sample size n leads to a convergence of the run-length towards the expected value of 370 at $\beta = 0.0027$. In the case of small sample sizes and number of samples the average run-length can be very large because of the uncertainty in the estimation of the parameter σ^2. This is especially the case for the upper control limit. If practitioners are interested in a certain average run-length, they are advised to adjust the nominal value of β to obtain that specific run-length. For $m = 10$ and $n = 5$, $\beta = 0.0173$ will give an average run-length of 370. For the "diameter" example, the upper control limit will then be $UCL = (3.406)(10.72) = 36.512$. In Figure 4.7 the predictive distribution of the run-length is displayed for the $(1 - 0.0173)\,100\% = 98.27\%$ two-sided control limit. As mentioned for given σ^2 the run-length r is geometric with parameter $\psi(\sigma^2)$. The unconditional run-length as given in Figure 4.7 is therefore

obtained using the Rao-Blackwell method, i.e., the average of a large number of unconditional run-lengths.

Figure 4.7.: Predictive Distribution of the "Run-length" $f\left(r|data\right)$for $m = 10$ and $n = 5$ - Two-sided Control Chart

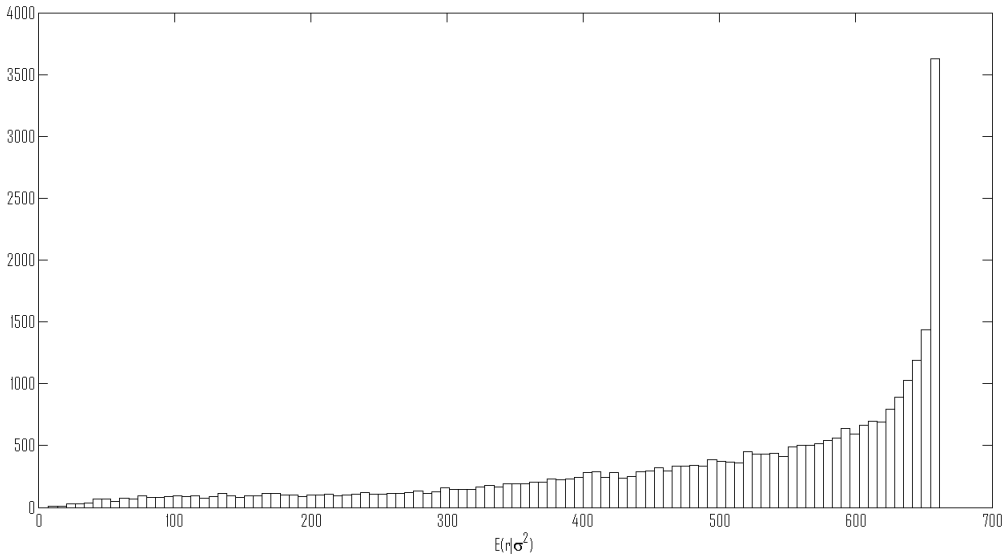


$$E\left(r|data\right) = 498.6473;\ Median\left(r|data\right) = 319;\ Var\left(r|data\right) = 274473.1449$$

$$95\%\ Equal - tail = [9; 1961]$$

In Figure 4.8 the distribution of the average run-length is given.

Figure 4.8.: Distribution of the Average "Run-length" - Two-sided Control Chart



$Mean = 500;\ Median = 552;\ Variance = 25572.95$

$95\%\ Equal - tail = [92; 661]$

Also the harmonic mean of the run-length is $\frac{1}{\beta}$. Therefore if $\beta = 0.0027$, the harmonic mean is $\frac{1}{0.0027} = 370.37$ and if $\beta = 0.0173$, the harmonic mean is $\frac{1}{0.0173} = 57.8$ and the arithmetic mean is 370.

4.7. Conclusion

Phase I and Phase II control chart limits have been constructed using Bayesian methodology. In this chapter we have seen that due to Monte Carlo simulation the construction of control chart limits using the Bayesian paradigm are handled with ease. Bayesian methods allow the use of any prior to construct control limits without any difficulty. It has been shown that the uncertainty in unknown parameters are handled with ease in using the predictive distribution in the determination of control chart limits. It has also been shown that an increase in number of samples m and the sample size n leads to a convergence in the run-length towards the expected value of 370 at $\beta = 0.0027$.

Mathematical Appendix to Chapter 4

Proof of Theorem 4.1

The likelihood function, i.e., the distribution of the data is

$$L\left(\mu_1, \mu_2, \ldots, \mu_m, \sigma^2 | data\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}mn} \prod_{i=1}^{m} \prod_{j=1}^{n} \exp\left\{-\frac{1}{2}\left(y_{ij} - \mu_i\right)^2 / \sigma^2\right\}.$$

Deriving the posterior distribution as Poster $\propto$ Likelihood $\times$ Prior, and using the Jeffreys' prior it follows that

$$\mu_i | \sigma^2, data \sim N\left(\bar{y}_i, \frac{\sigma^2}{n}\right), i = 1, 2, \ldots, m$$

and

$$p\left(\sigma^2 | data\right) = \left(\frac{\tilde{S}}{2}\right)^{\frac{1}{2}k} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}(k+2)} \exp\left(-\frac{\tilde{S}}{2\sigma^2}\right), \sigma^2 > 0$$

an Inverse Gamma distribution with $k = m\left(n-1\right)$ and $\tilde{S} = m\left(n-1\right) S_p^2$.

Proof of Theorem 4.2

For a given σ^2 it follows that

$$\frac{\left(n-1\right) S_f^2}{\sigma^2} = \frac{v S_f^2}{\sigma^2} \sim \chi_v^2,$$

which means that

$$f\left(S_f^2|\sigma^2\right) = \left(\frac{v}{2\sigma^2}\right)^{\frac{1}{2}v} \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(S_f^2\right)^{\frac{1}{2}v-1} \exp\left(-\frac{vS_f^2}{2\sigma^2}\right)$$

where $v = n-1$ and $S_f^2 > 0$.

The unconditional predictive density of S_f^2 is given by

$$\begin{aligned} f\left(S_f^2|data\right) &= \int_0^\infty f\left(S_f^2|\sigma^2\right) p\left(\sigma^2|data\right) d\sigma^2 \\ &= \frac{(v)^{\frac{1}{2}v}\left(\tilde{S}\right)^{\frac{1}{2}k}\left(S_f^2\right)^{\frac{1}{2}v-1}\Gamma\left(\frac{v+k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{v}{2}\right)\left(\tilde{S}+vS_f^2\right)^{\frac{1}{2}(v+k)}} \qquad S_f^2 > 0 \end{aligned}$$

where $v = n-1$, $k = m\left(n-1\right)$ and $\tilde{S} = kS_p^2 = m\left(n-1\right)S_p^2$.

$$\therefore f\left(S_f^2|data\right) = S_p^2 F_{n-1,m(n-1)}$$

Proof of Theorem 4.3

For a given σ^2

$$\begin{aligned}
\psi(\sigma^2) &= P\left(S_f^2 > S_p^2 F_{n-1,m(n-1)}(1-\beta)\right) \\
&= P\left(\frac{\sigma^2\chi^2_{n-1}}{n-1} > S_p^2 F_{n-1,m(n-1)}(1-\beta)\right) && \text{for given } \sigma^2 \\
&= P\left(\frac{m(n-1)S_p^2}{\chi^2_{m(n-1)}}\frac{\chi^2_{n-1}}{n-1} > S_p^2 F_{n-1,m(n-1)}(1-\beta)\right) && \text{for given } \chi^2_{m(n-1)} \\
&= P\left(\chi^2_{n-1} > \tfrac{1}{m}\chi^2_{m(n-1)} F_{n-1,m(n-1)}(1-\beta)\right) && \text{for given } \chi^2_{m(n-1)} \\
&= \psi\left(\chi^2_{m(n-1)}\right) && \text{for given } \chi^2_{m(n-1)}
\end{aligned}$$

Proof of Theorem 4.4

For a given $|\Sigma|$

$$\begin{aligned}
\psi(|\Sigma|) &= P\left\{|S_f| > \left(\tfrac{1}{n-1}\right)^p |A| \left(\textstyle\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i} F^*_{1-\beta}\right)\right\} \\
&= P\left\{\left|\tfrac{1}{n-1}\Sigma\right| \textstyle\prod_{i=1}^p \chi^2_{n-i} \geq \left(\tfrac{n}{n-1}\right)^p |A| \left(\textstyle\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i}\right) F^*_{1-\beta}\right\} \\
&= P\left\{|A| \textstyle\prod_{i=1}^p \left(\frac{1}{\chi^2_{m(n-1)+1-i}}\right) \prod_{i=1}^p \chi^2_{n-i} \geq |A| \left(\textstyle\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i}\right) F^*_{1-\beta}\right\} \\
&= P\left\{\textstyle\prod_{i=1}^p \chi^2_{n-i} \geq \left(\prod_{i=1}^p \chi^2_{m(n-1)+1-i}\right) \left(\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i}\right) F^*_{1-\beta}\right\}
\end{aligned}$$

for a given $\prod_{i=1}^p \chi^2_{m(n-1)+1-i}$.

5. Bayesian Control Charts for Tolerance Limits in the Case of Normal Populations

5.1. Introduction

Krishnamoorthy and Mathew (2009) and Hahn and Meeker (1991) defined a tolerance interval as an interval that is constructed in such a way that it will contain a specified proportion or more of the population with a certain degree of confidence. The proportion is also called the content of the tolerance interval. As opposed to confidence intervals that give information on unknown population parameters, a one-sided upper tolerance limit for example provides information about a quantile of the population. According to Hahn and Meeker (1991) tolerance intervals would be of importance in obtaining limits on the process capability of a product manufactured in large quantities. Further application examples of tolerance intervals include statistical process control, wood manufacturing, clinical and industrial applications, environmental monitoring and assessment and for exposure data analysis. For more applications see Krishnamoorthy and Mathew (2009) and Hugo (2012).

In the first part of this chapter we will look at a one-sided upper tolerance limit for a single normal population and in the second part of the chapter at one-sided tolerance limits for the difference between two normal populations.

5.2. One-sided Tolerance Limits for a Normal Population

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ population. The maximum likelihood estimators of the unknown mean, μ and unknown variance, σ^2 are the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$. Using the same notation as given in Krishnamoorthy and Mathew (2009), the pth quantile of a $N(\mu, \sigma^2)$ population is

$$q_p = \mu + z_p\sigma \tag{5.1}$$

where z_p denotes the pth quantile of a standard normal distribution.

A $1-\alpha$ upper confidence limit for q_p is a $(p, 1-\alpha)$ one-sided upper tolerance limit for the normal distribution. By using the posterior predictive distribution a Bayesian procedure will be developed to obtain control limits for a one-sided upper tolerance limit in the case of future samples.

There do not appear to be many papers on control charts for tolerance intervals from a Bayesian point of view. Hamada (2002) derived Bayesian tolerance interval control limits for np, p, c and u charts which control the probability content at a specified level with a given confidence while we are deriving posterior predictive intervals. It is therefore clear that our Bayesian method differs substantially from his.

5.3. Bayesian Procedure

By assigning a prior distribution to the unknown parameters the uncertainty in the estimation of the unknown parameters can adequately be handled. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution of q_p. By using the posterior distribution the predictive distribution of a future sample one-sided upper tolerance limit can be obtained. The predictive distribution on the other hand can be used to obtain control limits and to determine the distribution of the run-length and expected run-length. Determination of reasonable non-informative priors is however not an easy task. Therefore, in the next

section, reference and probability matching priors will be derived for $q_p = \mu + z_p\sigma$ the pth quantile of a $N(\mu, \sigma^2)$ distribution.

5.4. Reference and Probability-Matching Priors for $q_p = \mu + z_p\sigma$

As mentioned the Bayesian paradigm emerges as attractive in many types of statistical problems, also in the case of q_p, the pth quantile of a $N(\mu, \sigma^2)$ population.

Prior distributions are needed to complete the Bayesian specification of the model. Determination of reasonable non-informative priors in multi-parameter problems is not easy; common non-informative priors, such as the Jeffreys' prior can have features that have an unexpectedly dramatic effect on the posterior.

Reference and probability-matching priors often lead to procedures with good frequency properties while returning the Bayesian flavor. The fact that the resulting Bayesian posterior intervals of the level $1 - \alpha$ are also good frequentist intervals at the same level is a very desirable situation.

See also Bayarri and Berger (2004) and Severine, Mukerjee, and Ghosh (2002) for a general discussion.

5.4.1. The Reference Prior

In this section the reference prior of Berger and Bernardo (1992) will be derived for $q_p = \mu + z_p\sigma$.

The following theorem can be proved:

Theorem 5.1. *The reference prior for the ordering* $\{q_p, \sigma^2\}$ *is given by* $p_R(q_p, \sigma^2) \propto \sigma^{-2}$.

In the (μ, σ) *parametrization this corresponds to* $p_R(\mu, \sigma) \propto \sigma^{-2}$.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Note: The ordering $\{q_p, \sigma^2\}$ means that the parameter $q_p = \mu + z_p\sigma$ is a more important parameter than σ^2.

5.4.2. Probability-Matching Priors

Using the method of Datta and Ghosh (1995) the following theorem will be proved:

Theorem 5.2. *The probability-matching prior for q_p and σ^2 is $p_M(q_p, \sigma^2) \propto \sigma^{-2}$.*

Proof. The proof is given in the Mathematical Appendices to this chapter. □

5.4.3. The Posterior Distribution

As mentioned, by combining the information contained in the prior with the likelihood function the posterior distribution can be obtained. Since our non-informative prior for q_p in the (μ, σ^2) parametrization is $p(\mu, \sigma^2) \propto \sigma^{-2}$, it follows that the posterior distribution of σ^2 has an Inverse Gamma distribution which means that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ and $\mu|\sigma^2, data \sim N\left(\bar{X}, \frac{\sigma^2}{n}\right)$.

The posterior distribution of q_p is therefore equal to

$$\bar{X} + \frac{Z + z_p\sqrt{n}}{U}\frac{S}{\sqrt{n}} = \bar{X} + \frac{1}{\sqrt{n}}t_{n-1}\left(z_p\sqrt{n}\right)S$$

where $Z \sim N(0,1)$ and independently distributed of $U^2 \sim \frac{\chi^2_{n-1}}{n-1}$. Thus a $(p, 1-\alpha)$ upper tolerance limit is given by

$$\bar{X} + k_1 S = \bar{X} + t_{n-1;1-\alpha}\left(z_p\sqrt{n}\right)\frac{S}{\sqrt{n}}$$

where $t_{n-1;1-\alpha}(z_p\sqrt{n})$ denotes the $1-\alpha$ quantile of a non-central t distribution with $n-1$ degrees of freedom and non-centrality parameter $z_p\sqrt{n}$. $\bar{X} + k_1 S$ is an exact tolerance limit (i.e., has the correct coverage probability) and as mentioned by Krishnamoorthy and Mathew (2009) is the same solution that is obtained by the frequentist approach. The tolerance factor k_1, which is derived from the non-central t-distribution, can be obtained from Table B1 in Krishnamoorthy and Mathew (2009).

In this chapter we are firstly interested in the predictive distribution of a future sample one-sided upper tolerance limit. By using the predictive distribution a Bayesian procedure will be developed to obtain control limits for a future sample one-sided upper tolerance limit. Assuming that the process remains stable, the predictive

distribution can be used to derived the distribution of the run-length and average run-length.

5.5. A Future Sample One-sided Upper Tolerance Limit

Consider a future sample of m observations from the $N(\mu, \sigma^2)$ population: $X_{1f}, X_{2f}, \ldots, X_{mf}$. The future sample mean is defined as $\bar{X}_f = \frac{1}{m}\sum_{j=1}^{m} X_{jf}$ and a future sample variance by $S_f^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(X_{jf} - \bar{X}_f\right)^2$.

A $(p, 1-\alpha)$ upper tolerance limit for the future sample is defined as

$$\tilde{q} = \bar{X}_f + \tilde{k}_1 S_f \tag{5.2}$$

where

$$\tilde{k}_1 = \frac{1}{\sqrt{m}} t_{m-1;1-\alpha}\left(z_p\sqrt{m}\right).$$

Although the posterior predictive distribution of $\tilde{q}$ can easily be obtained by simulation, the exact mean and variance can be derived analytically. The following theorem can now be proved.

Theorem 5.3. *The exact mean and variance of* $\tilde{q} = \bar{X}_f + \tilde{k}_1 S_f$ *is given by*

$$E(\tilde{q}|data) = \bar{X} + \tilde{k}_1 \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\sqrt{n-1}}{\sqrt{m-1}} S$$

and

$$Var(\tilde{q}|data) = \left(\frac{m+n}{nm}\right)\left(\frac{n-1}{n-3}\right) + \tilde{k}_1^2 \left\{\frac{n-1}{n-3} - \frac{\Gamma^2\left(\frac{m}{2}\right)\Gamma^2\left(\frac{n-2}{2}\right)(n-1)}{\Gamma^2\left(\frac{m-1}{2}\right)\Gamma^2\left(\frac{n-1}{2}\right)(m-1)}\right\} S^2$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Corollary 5.4. *If $m = n$, then*

$$E\left(\tilde{q}|data\right) = \bar{X} + \tilde{k}_1 \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)} S$$

and

$$Var\left(\tilde{q}|data\right) = \frac{2}{n}\left(\frac{n-1}{n-3}\right) + \tilde{k}_1^2\left(\frac{n-1}{n-3} - \frac{\Gamma^2\left(\frac{n}{2}\right)\Gamma^2\left(\frac{n-2}{2}\right)}{\Gamma^4\left(\frac{n-1}{2}\right)}\right) S^2.$$

5.6. The predictive distribution of $\tilde{q} = \bar{X}_f + \tilde{k}_1 S_f$

As mentioned the posterior predictive distribution of $\tilde{q}$ can easily be simulated. This can be done in the following way:

$$\tilde{q}|\sigma^2, S_f^2, data \sim N\left(\bar{X} + \tilde{k}_1 S_f, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right).$$

Therefore

$$f\left(\tilde{q}|\sigma^2, S_f^2, data\right) = \left(\frac{mn}{\sigma^2\left(m+n\right)2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{mn}{2\sigma^2\left(m+n\right)}\left[\tilde{q} - \left(\bar{X} + \tilde{k}_1 S_f\right)\right]^2\right\}. \tag{5.3}$$

The unconditional predictive distribution can be obtained by first simulating σ^2 and then S_f. From the posterior distribution it follows that $\sigma^2 \sim \frac{(n-1)S^2}{\chi^2_{n-1}}$ and given σ^2, $S_f \sim \left\{\frac{\sigma^2\chi^2_{m-1}}{m-1}\right\}^{\frac{1}{2}}$. Substitute the simulated σ^2 and S_f values in equation (5.3) and draw the Normal density function. Repeat the procedure l times and average the l simulated normal density functions (Rao-Blackwell method) to obtain the unconditional predictive density function $f\left(\tilde{q}|data\right)$. In the example that follows $l = 100,000$.

5.7. Example

According to Krishnamoorthy and Mathew (2009) one-sided upper tolerance limits can commonly be used to assess the pollution level in a work place or in a region. The data in Table 5.1 represent air lead levels collected by the National Institute of Occupational Safety and Health at a laboratory, for health hazard evaluation. The air lead levels were collected from $n = 15$ different areas within the facility.

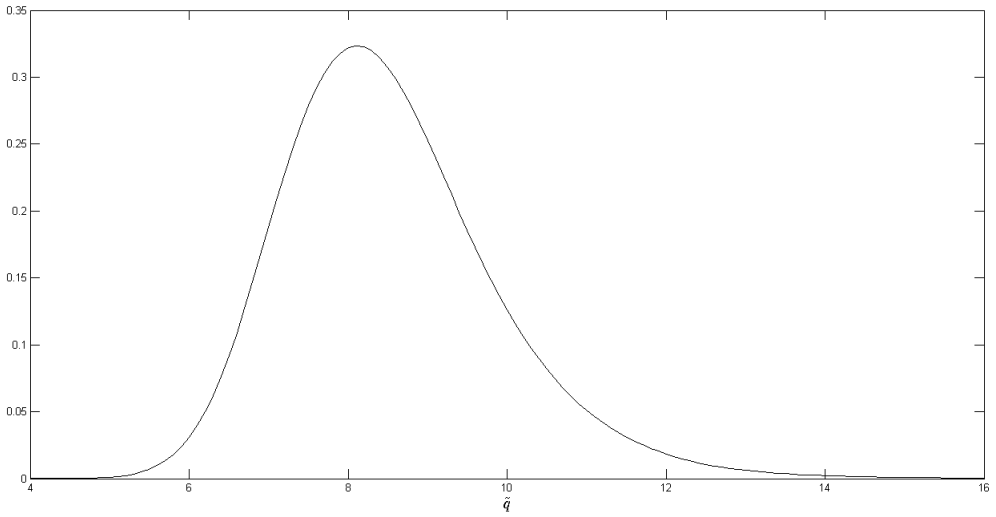
Table 5.1.: Air Lead Levels ($\mu g/m^3$)

200	120	15	7	8	6	48	61
380	80	29	1,000	350	1,400	110	

A Normal distribution fitted the log-transformed lead levels quite well. The sample mean and standard deviation of the log-transformed data are calculated as $\bar{X} = 4.3329$ and $S = 1.7394$. For $n = 15$, $1 - \alpha = 0.90$, $p = 0.95$ and using the non-central t distribution in MATLAB, $k_1 = \frac{1}{\sqrt{n}} t_{n-1,1-\alpha}\left(z_p\sqrt{n}\right) = 2.3290$. A $(0.95, 0.90)$ upper tolerance limit for the air lead level is $\bar{X} + k_1 S = 8.3840$.

In the first part of this chapter we are however interested in the predictive distribution of $\tilde{q} = \bar{X}_f + \tilde{k}_1 S_f$ the tolerance limit for a future sample of $m = n = 15$ observations. Using the simulated procedure described in Section 5.6, the predictive distribution is illustrated in Figure 5.1.

Figure 5.1.: Predictive Density Function of a Future Tolerance Limit $\tilde{q} = \bar{X}_f + \tilde{k}_1 S_f$



Mean = 8.5214, Mode = 8.2741

The mean of the predictive distribution of $\tilde{q}$ is somewhat larger and the mode somewhat smaller than 8.384 the sample upper tolerance limit of the air lead level.

In Table 5.2 it is shown that the calculated means and variances from the simulation and formulae are for all practical purposes the same.

Table 5.2.: Mean and Variance of $\tilde{q}$

	$E(\tilde{q}\|data)$	$Var(\tilde{q}\|data)$
From simulated $\tilde{q}$	8.5214	1.8981
Using Formulae	8.5427	1.8950

In Table 5.3 confidence limits for $\tilde{q}$ are given

Table 5.3.: Confidence Limits for $\tilde{q}$

	95% Left One-sided	95% Right One-sided	95% Two-sided
Left Limit	6.5683	-	6.2421
Right Limit	-	11.0320	11.6827

5.8. Control Chart for a Future One-sided Upper Tolerance Limit

By using the predictive distribution a Bayesian procedure will be developed to obtain a control chart for a future one-sided upper tolerance limit. Assuming the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length. From Figure 5.1 it follows a 99.73% upper control limit for $\tilde{q} = \bar{X}_f + \tilde{k}_1 S_f$ is 13.7. Therefore the the rejection region of size β ($\beta = 0.0027$) for the predictive distribution is

$$\beta = \int_{R(\beta)} f\left(\tilde{q}|data\right) d\tilde{q}$$

where $R\left(\beta\right)$ represents those values of $\tilde{q}$ that are larger than 13.7.

The run-length is defined as the number of future $\tilde{q}$ values (r) until the control chart signals for the first time (Note that r does not include that $\tilde{q}$ value when the control chart signals). Given μ and σ^2 and a stable Phase I process, the distribution of the run-length r is Geometric with parameter

$$\psi\left(\mu, \sigma^2\right) = \int_{R(\beta)} f\left(\tilde{q}|\mu, \sigma^2\right) d\tilde{q}$$

where $f\left(\tilde{q}|\mu, \sigma^2\right)$ is the distribution of a future $\tilde{q}$ given that μ and σ^2 are known. The values of μ and σ^2 are however unknown and the uncertainty of these parameter values are described by their joint posterior distribution $p\left(\mu, \sigma^2|data\right)$. By simulating μ and σ^2 from $p\left(\mu, \sigma^2|data\right)$, the probability density function $f\left(\tilde{q}|\mu, \sigma^2\right)$ (for the charting statistic $\tilde{q}$) can be obtained in the following way:

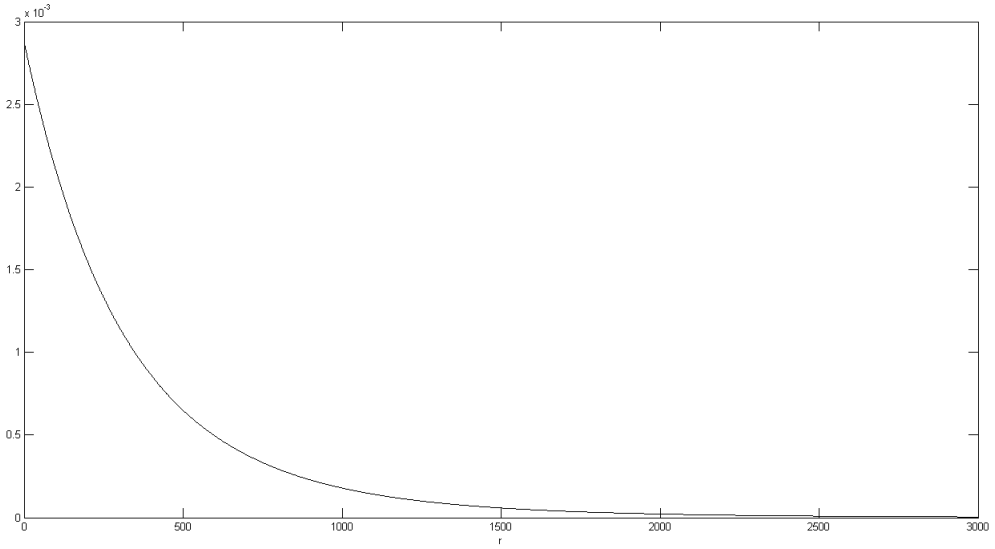
1. $\tilde{q}|\mu, \sigma^2, \chi^2_{m-1} \sim N\left(\mu + k_1\sigma\frac{\sqrt{\chi^2_{m-1}}}{\sqrt{m-1}}, \frac{\sigma^2}{m}\right)$.
2. The next step is to simulate $l = 100,000$ χ^2_{m-1} values to obtain l normal density functions for given μ and σ^2.
3. By averaging the l density functions (Rao-Blackwell method), $f\left(\tilde{q}|\mu, \sigma^2\right)$ can be obtained.

This must be done for each future sample. In other words for each future sample μ and σ^2 must first be simulated from $p\left(\mu, \sigma^2|data\right)$ and then the steps described in (item 1), (item 2) and (item 3).

The unconditional moments $E(r|data)$, $E(r^2|data)$ and $Var(r|data)$ can therefore easily be obtained by simulation or numerical integration. For further details refer to Menzefricke (2002, 2007, 2010b,a).

In Figure 5.2 the predictive distribution of the run-length is displayed for the 99.73% upper control limit. As mentioned for given μ and σ^2 the run-length r is Geometric with parameter $\psi(\mu, \sigma^2)$. The unconditional run-length as given in Figure 5.2 is therefore obtained using the Rao-Blackwell method, i.e., the average of a large number of conditional run-lengths

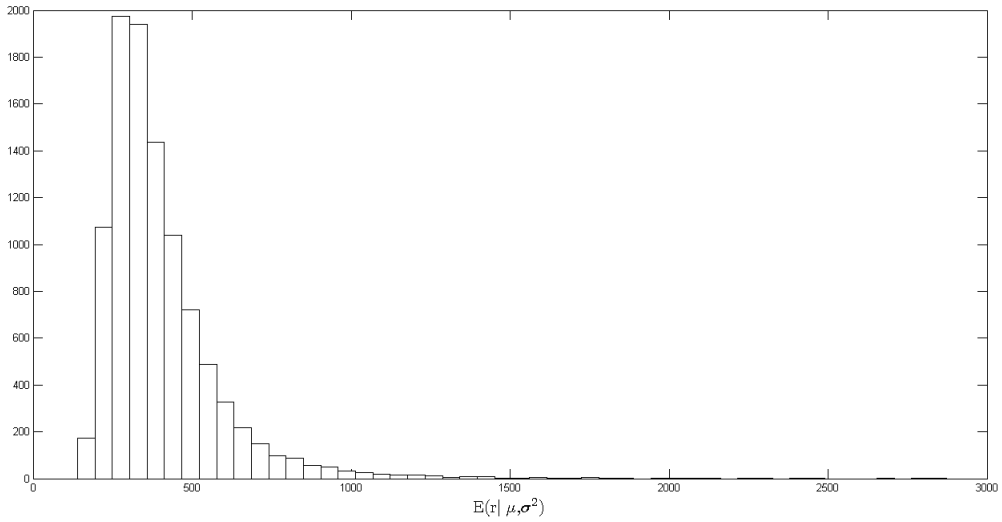
Figure 5.2.: Predictive Distribution of the "Run-length" $f(r|data)$ for $n = m = 15$



$E(r|data) = 396.27438, Median(r|data) = 251, Var(r|data) = 2.0945e^5$
95% Equal-tail Interval = $(8; 1644)$ Length = 1636
95% HPD Interval = $(3; 1266)$ Length = 1263

In Figure 5.3 the distribution of the average run-length is given.

Figure 5.3.: Distribution of the Average "Run-length"



$Mean = 400.1084, Median = 353.5753, Var = 3.5605e^4$
95% Equal-tail Interval = $(202.1834; 874.6607)$ Length = 672.4773
95% HPD Interval = $(163.4804; 566.0842)$ Length = 402.6038

For known μ and σ the expected run-length is $\frac{1}{0.0027} = 370$. If μ and σ^2 are unknown and estimated from the posterior distribution the expected run-length will usually be larger than 370 - especially if the sample size is small.

5.9. One-sided Tolerance Limits for the Difference Between Two Normal Populations

In the second part of this chapter, tolerance limits for the difference between normal distributions will be considered.

Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and independently distributed from $X_2 \sim N(\mu_2, \sigma_2^2)$ where μ_1 and μ_2 are the unknown population means and σ_1^2 and σ_2^2 the unknown population variances. If z_p is the pth quantile of a standard normal distribution then

$$L_p = \mu_1 - \mu_2 - z_p\sqrt{\sigma_1^2 + \sigma_2^2} \tag{5.4}$$

is the $1-p$ quantile of the distribution of $X_1 - X_2$. As mentioned by Krishnamoorthy

and Mathew (2009) a $(p, 1-\alpha)$ lower tolerance limit for the distribution of $X_1 - X_2$ is a $1-\alpha$ lower confidence limit of L_p.

Suppose $\bar{X}_i$ and S_i^2 are the sample mean and variance of a random sample of n_i observations from $N(\mu_i, \sigma_i^2)$, $i = 1, 2$. By using exact distributional results Krishnamoorthy and Mathew (2009) constructed a $(1-\alpha)$ lower confidence limit for L_p as

$$\bar{X}_1 - \bar{X}_2 - t_{n_1+n_2-2,1-\alpha}\left(z_p\sqrt{v_1}\right)\frac{S_d}{\sqrt{v_1}} \tag{5.5}$$

where

$$v_1 = \frac{\sigma_1^2 + \sigma_2^2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{5.6}$$

and

$$S_d^2 = \frac{(\sigma_1^2 + \sigma_2^2)}{n_1 + n_2 - 2}\left\{\frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{\sigma_2^2}\right\} \tag{5.7}$$

which is an exact $(p, 1-\alpha)$ lower tolerance limit for the distribution of $X_1 - X_2$ if the variance ratio is known. This is usually not the case, but the problem can be overcome by making use of the following "semi-Bayesian" approach.

Equation (5.5) is derived from the fact that the pivotal quantity

$$\frac{\bar{X}_1 - \bar{X}_2 - L_p}{S_d} \sim \frac{Z + z_p\sqrt{v_1}}{\sqrt{v_1}\sqrt{\chi^2_{n_1+n_2-2}/(n_1+n_2-2)}}N_2 \tag{5.8}$$

where $Z \sim N(0,1)$ and independent of $\chi^2_{n_1+n_2-2}$ (See equation (2.4.4) in Krishnamoorthy and Mathew (2009)).

From equation (5.8) it follows that

$$L_p|data, \chi^2_{n_1+n_2-2}, \sigma_1^2, \sigma_2^2 \sim N\left\{\left(\bar{X}_1 - \bar{X}_2\right) - \frac{z_p S_d}{\sqrt{\chi^2_{n_1+n_2-2}/(n_1+n_2-2)}}, \frac{S_d^2}{v_1 \chi^2_{n_1+n_2-2}/(n_1+n_2-2)}\right\} \tag{5.9}$$

if the non-informative prior $p\left(\sigma_1^2, \sigma_2^2\right) \propto \sigma_1^{-2}\sigma_2^{-2}$ is used then the posterior distribution of σ_i^2 $(i = 1, 2)$ is an Inverse-Gamma distribution which means that

$$\sigma_i^2 \sim \frac{(n_i - 1)\, S_i^2}{\chi^2_{n_i-1}} \quad (i = 1, 2)$$

$p\left(L_p|data\right)$ - the unconditional posterior distribution of L_p can be obtained by using the following simulation procedure:

1. Simulate σ_i^2 $(i = 1, 2)$ and substitute the simulated variance components in S_d^2 and v_1.
2. Simulate $\chi^2_{n_1+n_2-2}$ and substitute the value in equation (5.9).
3. Draw the normal density function.
4. Repeat the procedure l times and average the l simulated normal density functions (Rao-Blackwell method) to obtain $p\left(L_p|data\right)$.

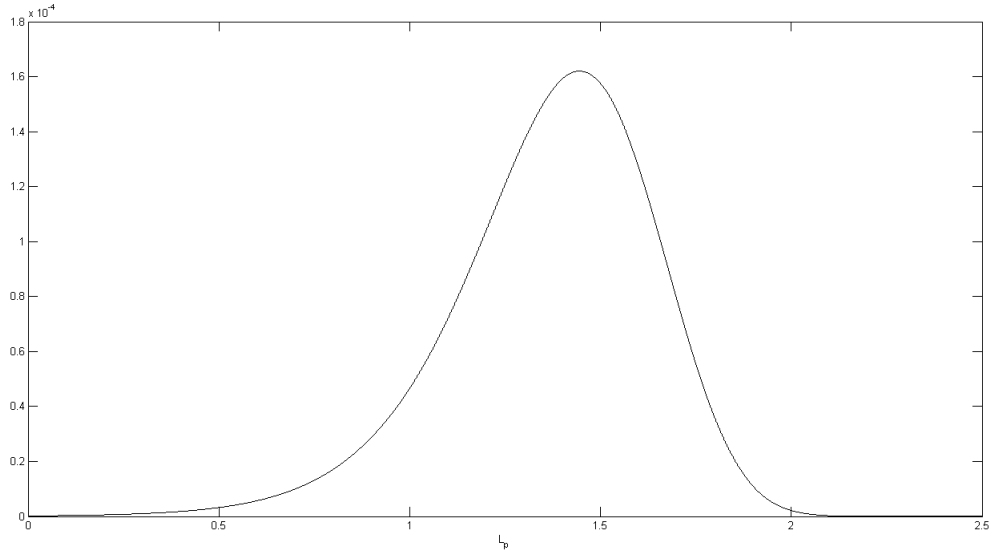
$l = 100,000$ is used in the example that follows.

5.10. Example - Breakdown Voltage - Power Supply

This example is taken from Krishnamoorthy and Mathew (2009) and it is concerned with the proportion of times the breakdown voltage of X_1 of a capacitor exceeds the voltage output X_2 of a transverter (power supply). A sample of $n_1 = 50$ capacitors yielded $\bar{X}_1 = 6.75kV$ and $S_1^2 = 0.123$. The voltage output from $n_2 = 20$ transverters produced $\bar{X}_2 = 4.00kV$ and $S_2^2 = 0.53$.

Using the simulation procedure as described the unconditional posterior distribution $p\left(L_p|data\right)$ is obtained and illustrated in Figure 5.4 for $z_p = z_{0.95} = 1.6449$.

Figure 5.4.: $p\left(L_p|data\right)$ - Unconditional Posterior Distribution of L_p



$$p\left(L_p < 0.8895\right) = 0.05$$
$$Mean = 1.3671,\ Median = 1.3954,\ Mode = 1.4448,\ Var = 0.0714$$

A $(0.95, 0.95)$ lower tolerance limit for the distribution of $X_1 - X_2$ is therefore $q_p = 0.8895$.

In what follows it will be shown that a more formal Bayesian procedure gives for all practical purposes the same estimate for the lower tolerance limit.

If the objective prior $p\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) \propto \sigma_1^{-2}\sigma_2^{-2}$ is used then it is well known that the posterior distribution L_p (equation (5.4)) given the variance components is normal with mean

$$E\left(L_p^B|\sigma_1^2, \sigma_2^2, data\right) = \bar{X}_1 - \bar{X}_2 - z_p\sqrt{\sigma_1^2 + \sigma_2^2}$$

and variance

$$Var\left(L_p^B|\sigma_1^2, \sigma_2^2, data\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Also

$$\sigma_i^2 \sim \frac{\left(n_i - 1\right) S_i^2}{\chi^2_{n_i - 1}}. \quad (i = 1, 2)$$

From this the posterior distribution follows:

$$L_p^B|data, U_1, U_2 \sim N\left\{\left(\bar{X}_1 - \bar{X}_2\right) - z_p\sqrt{\frac{S_1^2}{U_1^2} + \frac{S_2^2}{U_2^2}}, \frac{S_1^2}{n_1 U_1^2} + \frac{S_2^2}{n_2 U_2^2}\right\} \tag{5.10}$$

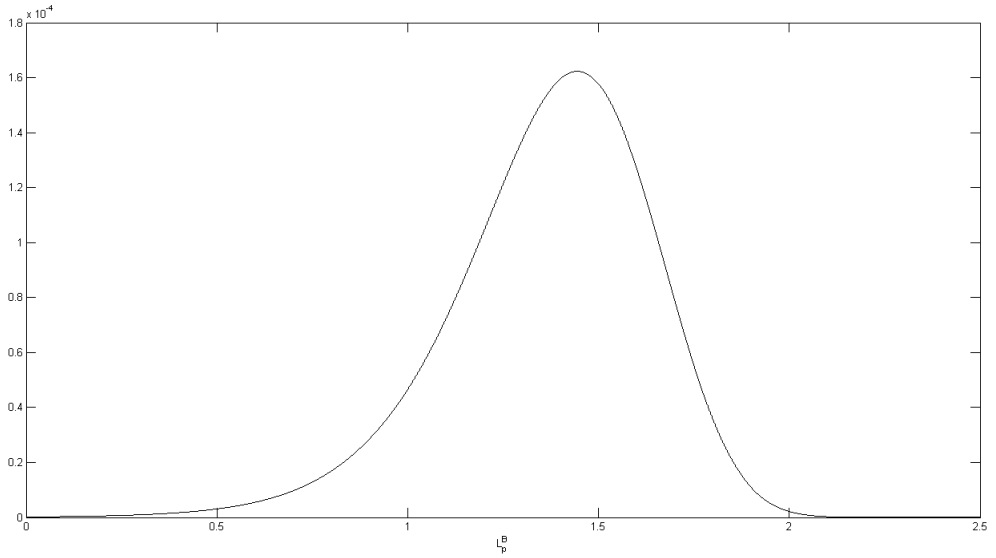
where

$$U_i^2 \sim \frac{\chi^2_{n_i-1}}{n_i - 1}.$$

Equation (5.10) is exactly the same as the generalized variable approached described by Krishnamoorthy and Mathew (2009).

By simulating 100,000 $U_i^2 \sim \frac{\chi^2_{n_i-1}}{n_1-1}$ $(i = 1, 2)$ values and using the Rao-Blackwell method the unconditional posterior distribution of L_p^B is obtained and given in Figure 5.5.

Figure 5.5.: $p\left(L_p^B|data\right)$



$$p\left(L_p^B < 0.88861\right) = 0.05$$
$$Mean = 1.3682,\ Median = 1.3964,\ Mode = 1.4462,\ Var = 0.0713$$

Krishnamoorthy and Mathew (2009) obtained the $(0.95, 0.95)$ lower tolerance limit as 0.8870.

Since the two methods give for all practical purposes the same estimate for the lower tolerance limit, only the first method will be used to obtain the predictive distribution of a future one-sided tolerance limit.

5.11. The Predictive Distribution of a Future One-sided Lower Tolerance Limit for $X_1 - X_2$

By using the predictive distribution a Bayesian procedure will be developed to obtain a control chart for a future one-sided lower tolerance limit. Assuming the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length.

Consider a future sample of m_1 observations from the $N(\mu_1, \sigma_1^2)$ population $X_{1f}^{(1)}, X_{2f}^{(1)}, \ldots, X_{m_1 f}^{(1)}$. The future sample mean is defined as $\bar{X}_f^{(1)} = \frac{1}{m_1}\sum_{j=1}^{m_1} X_{jf}^{(1)}$ and a future sample variance by $S_f^{2\,(1)} = \frac{1}{m_1-1}\sum_{j=1}^{m_1}\left(X_{jf}^{(1)} - \bar{X}_f^{(1)}\right)^2$. Similar for a future sample of m_2 observations from $N(\mu_2, \sigma_2^2)$, $\bar{X}_f^{(2)} = \frac{1}{m_2}\sum_{j=1}^{m_2} X_{jf}^{(2)}$ and $S_f^{2\,(2)} = \frac{1}{m_2-1}\sum_{j=1}^{m_2}\left(X_{jf}^{(2)} - \bar{X}_f^{(2)}\right)$.

Thus using the exact distributional results a $(p, 1-\alpha)$ lower tolerance limit for the difference between two future samples is defined as

$$q_f = \bar{X}_f^{(1)} - \bar{X}_f^{(2)} - t_{m_1+m_2-2;1-\alpha}\left(z_p\sqrt{v_1^f}\right)\frac{S_d^f}{\sqrt{v_1^f}} \tag{5.11}$$

If the sample sizes of the two future samples are the same, i.e., $m_1 = m_2 = m$, then equation (5.11) simplifies to

$$\begin{aligned} q_f &= \bar{X}_f^{(1)} - \bar{X}_f^{(2)} - t_{2(m-1);1-\alpha}\left(z_p\sqrt{m}\right)\frac{S_d^f}{\sqrt{m}} \\ &= \bar{X}_f^{(1)} - \bar{X}_f^{(2)} - \tilde{\tilde{k}}S_d^f \end{aligned}$$

where

$$\tilde{\tilde{k}} = \frac{1}{\sqrt{m}}t_{2(m-1);1-\alpha}\left(z_p\sqrt{m}\right),$$

$$v_1^f = m$$

and

$$S_d^{2\,f} = \frac{(\sigma_1^2 + \sigma_2^2)}{2}\left\{\frac{S_f^{2\,(1)}}{\sigma_1^2} + \frac{S_f^{2\,(2)}}{\sigma_2^2}\right\}.$$

We are interested in the predictive distribution of q_f. This cannot be derived analytically, but can easily be obtained by simulation.

Now

$$q_f|\sigma_1^2, \sigma_2^2, S_f^{2\,(1)}, S_f^{2\,(2)}, \mu_1, \mu_2 \sim N\left\{(\mu_1 - \mu_2) - \tilde{\tilde{k}} S_d^f, \frac{1}{m}\left(\sigma_1^2 + \sigma_2^2\right)\right\}.$$

since

$$\bar{X}_f^{(1)} - \bar{X}_f^{(2)} \sim N\left\{\mu_1 - \mu_2, \frac{1}{m}\left(\sigma_1^2 + \sigma_2^2\right)\right\}.$$

Also if the prior $p(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-2}\sigma_2^{-2}$ is used,

$$(\mu_1 - \mu_2)\,|data, \sigma_1^2, \sigma_2^2 \sim N\left\{\bar{X}_1 - \bar{X}_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right\}.$$

Therefore

$$q_f|\sigma_1^2, \sigma_2^2, S_f^{2\,(1)}, S_f^{2\,(2)}, data \sim N\left\{\left(\bar{X}_1 - \bar{X}_2\right) - \tilde{\tilde{k}} S_d^f, \frac{1}{m}\left(\sigma_1^2 + \sigma_2^2\right) + \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right\} \tag{5.12}$$

The unconditional predictive distribution of q_f can be obtained by first simulating

σ_i^2 and then $S_f^{2\,(i)}$ $(i = 1, 2)$. As before it follows from the posterior distribution that

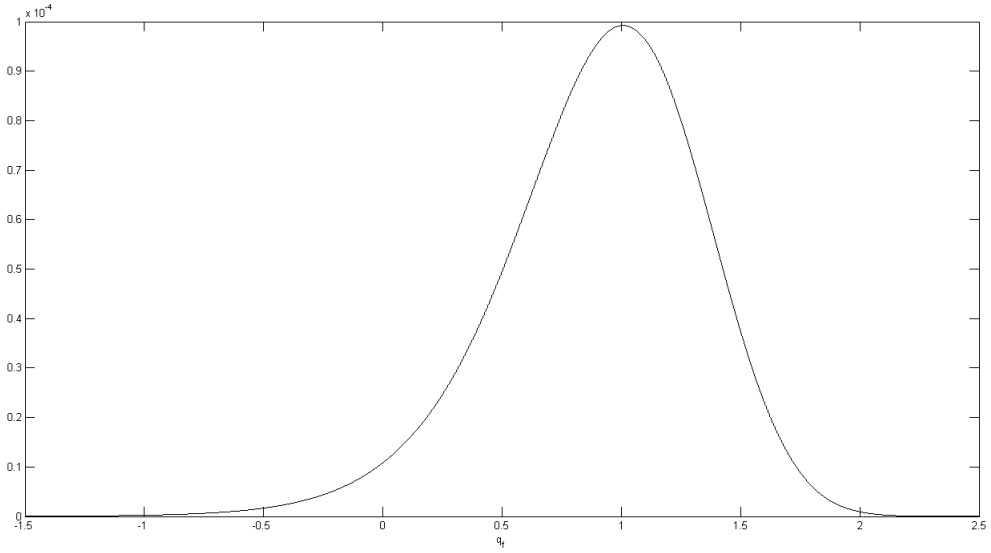
$$\sigma_i^2 \sim \frac{(n_i - 1)\, S_i^2}{\chi^2_{n_i - 1}} \quad (i = 1, 2)$$

and given σ_i^2

$$S_f^{2\,(i)} \sim \frac{\sigma_i^2 \chi^{2\,(i)}_{m-1}}{m - 1} \quad (i = 1, 2)$$

Substitute the simulated σ_i^2 and $S_f^{2\,(i)}$ values in equation (5.12) and draw the normal density function. Repeat the procedure l times and average the l simulated density functions (Rao-Blackwell method) to obtain the unconditional predictive density function $f\,(q_f|data)$. The predictive density function for the Breakdown Voltage example given in Section 5.10 for $m_1 = m_2 = 20$ is illustrated in Figure 5.6.

Figure 5.6.: $f\,(q_f|data)$- Predictive Density of the Lower Tolerance Limit, q_f



$\alpha = 0.05$, $\tilde{\tilde{k}} = 2.2096$
$Mean = 0.9060$, $Median = 0.9427$, $Mode = 1.0081$, $Var = 0.1834$
95% Equal-tail Interval $= (-0.0417; 1.6406)$
95% HPD Interval $= (0.0081; 1.6703)$
99.73% Equal-tail Interval $= (-0.8068; 1.9614)$
$p\,(q_f \leq -0.6258) = 0.0027$
$p\,(q_f \geq 1.8955) = 0.0027$

The mean of the predictive distribution of q_f is 0.9060 and is somewhat larger than 0.08895 the sample lower tolerance limit for the break down voltage data.

5.12. Control Chart for a Future One-sided Lower Tolerance Limit

As before, a Bayesian procedure will be developed to obtain a control chart for a future one-sided lower tolerance limit. Assuming the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length.

From Figure 5.6 it follows that for a 99.73% two-sided control chart the lower control limit is $LCL = -0.8068$ and the upper control limit is $UCL = 1.9614$. Therefore the rejection region of size β ($\beta = 0.0027$) for the predictive distribution is

$$\beta = \int_{R(\beta)} f\left(q_f|data\right) dq_f$$

where $R\left(\beta\right)$ represents those values of q_f that are smaller than LCL and larger than UCL.

As mentioned, the run-length is defined as the number of future q_f values (r) until the control chart signals for the first time. Given μ_i, σ_i^2 $(i = 1, 2)$, the distribution of the run-length r is geometric with parameter

$$\psi\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) = \int_{R(\beta)} f\left(q_f|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) d_{qf}$$

where $f\left(q_f|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right)$ is the distribution of q_f given that μ_i,σ_i^2 $(i = 1, 2)$ are known. These parameter values are however unknown but can be simulated from their joint posterior distribution $p\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2|data\right)$ as follows:

$$\sigma_i^2|data \sim \frac{\left(n_i - 1\right) S_i^2}{\chi_{n_i-1}^2} \quad (i = 1, 2)$$

and

$$\mu_i|\sigma_i^2, data \sim N\left(\bar{X}_i, \frac{\sigma_i^2}{n_1}\right) \quad (i = 1, 2)\,.$$

1. From this it follows that

$$q_f|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, S_f^{2\,(1)}, S_f^{2\,(2)} \sim N\left\{(\mu_1 - \mu_2) - \tilde{\tilde{k}} S_d^{(f)}, \frac{1}{m}\left(\sigma_1^2 + \sigma_2^2\right)\right\}$$

2. The next step is to simulate $l = 100,000$ $S_f^{2\,(i)}$ values for given σ_i^2 $(i = 1, 2)$. $S_f^{2\,(i)} \sim \frac{\sigma_i^2 \chi_{m-1}^{2\,(i)}}{m-1}$ $(i = 1, 2)$.
3. By averaging the l density functions (Rao-Blackwell method), $f\left(q_f|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right)$ can be obtained.

This must be done for each future sample. In other words for each future sample μ_i and σ_i^2 $(i = 1, 2)$ must first be simulated from their joint posterior distribution $p\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2|data\right)$ and then the steps described in (1), (2) and (3) must follow.

The mean of the predictive distribution of the run-length for the 99.73% two-sided control limits is $E\left(r|data\right) = 201500$ much larger than the 370 that one would have expected if $\beta = 0.0027$. The reason for this large average run-length is mainly the small future sample sizes $(m_1 = m_2 = 20)$. The median run-length = 3770. Define$\bar{\psi}\left(\mu_1, \mu_2, \sigma_1, \sigma_2\right) = \frac{1}{\tilde{m}} \sum_{i=1}^{\tilde{m}} \psi\left(\mu_1^{(i)}, \mu_2^{(i)}, \sigma_1^{2\,(i)}, \sigma_2^{2\,(i)}\right)$. According to Menzefricke (2002) $\bar{\psi}\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\right) \to \beta$ if $\tilde{m} \to \infty$. The harmonic mean of $r = \frac{1}{\beta}$ and if $\beta = 0.0027$, the harmonic mean $= \frac{1}{0.0027} = 370$.

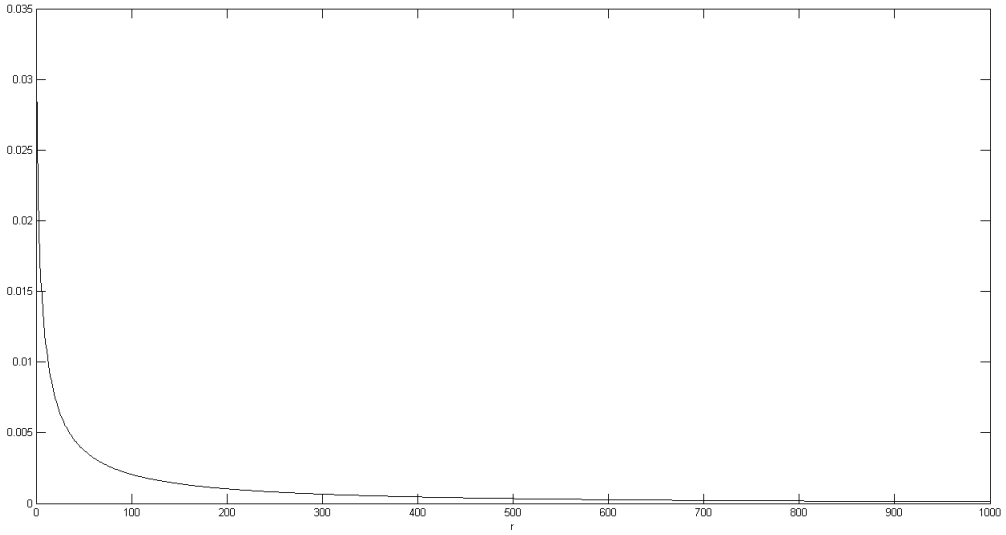
In Table 5.4 the average "run-length" for different values of β is given.

Table 5.4.: β Values and Corresponding Average Run-length

β	0.0027	0.012	0.016	0.02	0.024	0.028	0.03	0.032	0.036	0.04
$E\left(r|data\right)$	201500	4044	1785	959	633	438	400	340	240	218

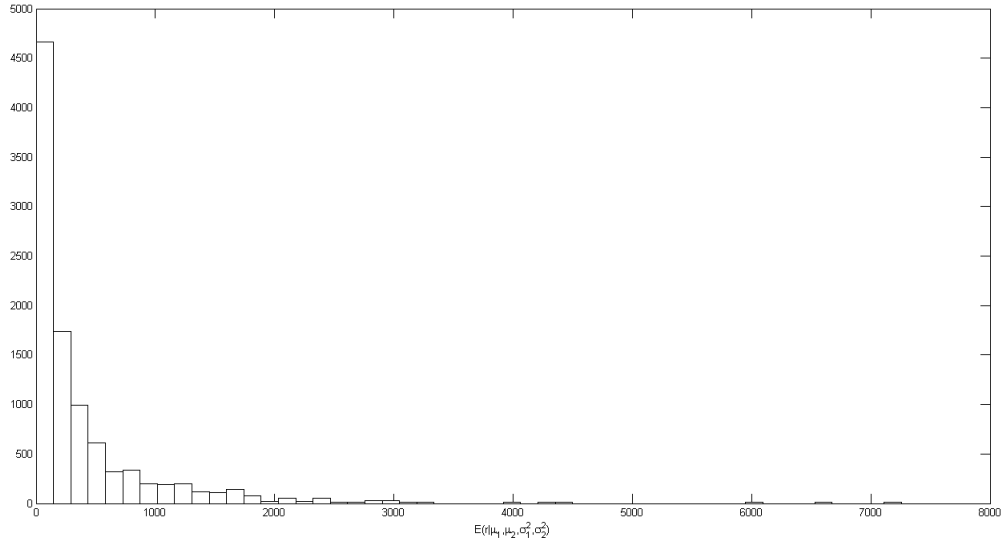
In Figure 5.7 the distribution of the run-length for $m_1 = m_2 = 20$ and $\beta = 0.03$ is illustrated and in Figure 5.8 the histogram of the expected run-length is given.

Figure 5.7.: Predictive Distribution of the Run-length $f\left(r|data\right)$ for $m_1 = m_2 = m = 20$



$E\left(r|data\right) = 400.41$,$Median\left(r|data\right) = 79.51, Var\left(r|data\right) = 1.8668e^6$
95% HPD Interval $= (0; 1688)$

Figure 5.8.: Distribution of the Average Run-length



$Mean = 400.41, Median = 138.79, Var = 7.0775e^5$
95% HPD Interval $= (1.11; 1552.1)$

5.13. Conclusion

The first part of this chapter develops a Bayesian control chart for monitoring a upper one-sided tolerance limit across a range of sample values. In the Bayesian approach prior knowledge about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution. By using the posterior distribution the predictive distribution of a upper one-sided tolerance limit can be obtained.

Determination of reasonable non-informative priors in multi-parameter problems is not an easy task. The Jeffreys' prior for example can have a bad effect on the posterior distribution. Reference and probability matching priors are therefore derived for the pth quantile of a normal distribution. The theory and results have been applied to air-lead level data analyzed by Krishnamoorthy and Mathew (2009) to illustrate the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, prediction intervals and run-lengths.

In the second part of this chapter the Bayesian procedure has been extended to control charts of one-sided tolerance limits for a distribution of the difference between two independent normal variables.

Mathematical Appendix to Chapter 5

Proof of Theorem 5.1

Assume $X_i\,(i = 1, 2, \ldots, n)$ are independently and identically Normally distributed with mean μ and variance σ^2. The Fisher information matrix for the parameter vector $\underline{\theta} = [\mu, \sigma^2]'$ is given by

$$F\left(\mu, \sigma^2\right) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix}.$$

Let $q_p = \mu + z_p\sigma = t\,(\mu, \sigma^2) = t\,(\underline{\theta})$.

To obtain the reference prior, the Fisher information matrix $F\left(t\left(\underline{\theta}\right),\sigma\right)$ must first be derived.

Let

$$A=\begin{bmatrix}\frac{\partial\mu}{\partial t(\underline{\theta})} & \frac{\partial\mu}{\partial\sigma^2}\\ \\ \frac{\partial\sigma^2}{\partial t(\underline{\theta})} & \frac{\partial\sigma^2}{\partial\sigma^2}\end{bmatrix}=\begin{bmatrix}1 & -\frac{1}{2}\frac{z_p}{\sigma}\\ \\ 0 & 1\end{bmatrix}.$$

Now

$$F\left(t\left(\underline{\theta}\right),\sigma^2\right)=A'F\left(\mu,\sigma^2\right)A=\begin{bmatrix}\frac{n}{\sigma^2} & -\frac{nz_p}{2\sigma^3}\\ \\ -\frac{nz_p}{2\sigma^3} & \frac{nz_p^2}{4\sigma^4}+\frac{n}{2\sigma^4}\end{bmatrix}=\begin{bmatrix}F_{11} & F_{12}\\ \\ F_{21} & F_{22}\end{bmatrix}$$

and the inverse

$$F^{-1}\left(t\left(\underline{\theta}\right),\sigma^2\right)=\frac{2\sigma^6}{n^2}\begin{bmatrix}\frac{n}{2\sigma^4}\left(\frac{z_p^2}{2}+1\right) & \frac{nz_p}{2\sigma^3}\\ \\ \frac{nz_p}{2\sigma^3} & \frac{n}{\sigma^2}\end{bmatrix}=\begin{bmatrix}F^{11} & F^{12}\\ \\ F^{21} & F^{22}\end{bmatrix}.$$

Therefore

$$F^{11}=\frac{\sigma^2}{n}\left(\frac{z_p^2}{2}+1\right),$$

$$\left(F^{11}\right)^{-1}=\frac{n}{\sigma^2}\left(\frac{z_p^2}{2}+1\right)^{-1}=h_1$$

and

$$p\left(t\left(\underline{\theta}\right)\right)\propto h_1^{\frac{1}{2}}\propto \text{ constant because it does not contain } t\left(\underline{\theta}\right).$$

Further

$$h_2 = F_{22} = \frac{n}{2\sigma^4}\left(\frac{z_p^2}{2} + 1\right)$$

and

$$p\left(\sigma^2 | t\left(\underline{\theta}\right)\right) \propto h_2^{\frac{1}{2}} \propto \sigma^{-2}.$$

Therefore the reference prior for the ordering $\{t\left(\underline{\theta}\right), \sigma^2\} = \{q_p, \sigma^2\}$ is $P_R\left(q_p, \sigma^2\right) \propto \sigma^{-2}$.

In the (μ, σ^2) parametrization this corresponds to $P_R\left(\mu, \sigma^2\right) = p\left(t\left(\underline{\theta}\right), \sigma^2\right)\left|\frac{\partial t(\underline{\theta})}{\partial \mu}\right|$.

Since $\left|\frac{\partial t(\underline{\theta})}{\partial \mu}\right| = 1$, it follows that $P_R\left(\mu, \sigma^2\right) \propto \sigma^{-2}$.

Proof of Theorem 5.2

Let

$$t\left(\underline{\theta}\right) = t\left(\mu, \sigma^2\right) = q_p$$

and

$$\nabla_t'\left(\underline{\theta}\right) = \left[\frac{\partial}{\partial q_p} t\left(\underline{\theta}\right) \quad \frac{\partial}{\partial \sigma^2} t\left(\underline{\theta}\right)\right] = \left[1 \quad 0\right].$$

Also

$$\nabla_t'\left(\underline{\theta}\right) F^{-1}\left(t\left(\underline{\theta}\right), \sigma^2\right) = \left[F^{11} \quad F^{12}\right]$$

and

$$\sqrt{\nabla_t'\left(\underline{\theta}\right) F^{-1}\left(t\left(\underline{\theta}\right), \sigma^2\right) \nabla_t\left(\underline{\theta}\right)} = \left(F^{11}\right)^{\frac{1}{2}}.$$

Further

$$\Upsilon'(\underline{\theta}) = \frac{\nabla_t'(\underline{\theta}) F^{-1}(t(\underline{\theta}), \sigma^2)}{\sqrt{\nabla_t'(\underline{\theta}) F^{-1}(t(\underline{\theta}), \sigma^2) \nabla_t(\underline{\theta})}} = \begin{bmatrix} \Upsilon_1(\underline{\theta}) & \Upsilon_2(\underline{\theta}) \end{bmatrix}$$

where

$$\Upsilon_1(\underline{\theta}) = \left(F^{11}\right)^{\frac{1}{2}} = \frac{\sigma}{\sqrt{n}} \left(\frac{z_p^2}{2} + 1\right)^{\frac{1}{2}}$$

and

$$\Upsilon_2(\underline{\theta}) = \frac{F^{12}}{\sqrt{F^{11}}} = \frac{\sigma^2 z_p}{\sqrt{n}} \left(\frac{z_p^2}{2} + 1\right)^{-\frac{1}{2}}.$$

According to Datta and Ghosh (1995) a prior $P_M(\underline{\theta}) = P_M(q_p, \sigma^2)$ will be a probability matching prior if the following differential equation is satisfied

$$\frac{\partial}{\partial q_p}\{\Upsilon_1(\underline{\theta}) P_M(\underline{\theta})\} + \frac{\partial}{\partial \sigma^2}\{\Upsilon_2(\underline{\theta}) P_M(\underline{\theta})\} = 0$$

.

It is therefore clear that if $P_M(\underline{\theta}) \propto \sigma^{-2}$ the differential equation is satisfied.

Proof of Theorem 5.3

It is well known that if $Y \sim \chi_u^2$, then

$$E(Y^r) = \frac{2^r \Gamma\left(\frac{u}{2} + r\right)}{\Gamma\left(\frac{u}{2}\right)}.$$

Also since $\bar{X}_f | \mu, \sigma^2 \sim N\left(\mu, \frac{\sigma^2}{m}\right)$ and $S_f \sim \left\{\frac{\sigma^2 \chi_{m-1}^2}{m-1}\right\}^{\frac{1}{2}}$ for given σ^2, it follows that

$$E\left(\tilde{q} | \mu, \sigma^2\right) = \mu + \frac{\tilde{k}_1 \sqrt{2}\sigma}{\sqrt{m-1}} \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)}$$

and

$$E\left(\tilde{q}^2|\mu,\sigma^2\right)=\mu^2+2\tilde{k}_1\mu\frac{\sigma\sqrt{2}\Gamma\left(\frac{m}{2}\right)}{\sqrt{m-1}\Gamma\left(\frac{m-1}{2}\right)}+\sigma^2\left(\frac{1}{m}+\tilde{k}_1^2\right).$$

From the posterior distribution it follows that $\mu|\sigma^2,data\sim N\left(\bar{X},\frac{\sigma^2}{n}\right)$ and $\sigma\sim\left\{\frac{(n-1)S^2}{\chi^2_{n-1}}\right\}$ given the data. Therefore

$$E\left(\tilde{q}|data\right)=\bar{X}+\tilde{k}_1\sqrt{\frac{n-1}{m-1}}\frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}S \tag{5.13}$$

and

$$E\left(\tilde{q}^2|data\right)=\bar{X}^2+2\tilde{k}_1\bar{X}\sqrt{\frac{n-1}{m-1}}\frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}S+\left(\frac{1}{n}+\frac{1}{m}+\tilde{k}_1^2\right)\left(\frac{n-1}{n-3}\right)S \tag{5.14}$$

By making use of Equations (5.13) and (5.14) and the fact that

$$Var\left(\tilde{q}|data\right)=E\left(\tilde{q}^2|data\right)-\left\{E\left(\tilde{q}|data\right)\right\}^2$$

it follows that

$$Var\left(\tilde{q}|data\right)=\left(\frac{m+n}{nm}\right)\left(\frac{n-1}{n-3}\right)+\tilde{k}_1^2\left\{\frac{n-1}{n-3}-\frac{\Gamma^2\left(\frac{m}{2}\right)\Gamma^2\left(\frac{n-2}{2}\right)(n-1)}{\Gamma^2\left(\frac{m-1}{2}\right)\Gamma^2\left(\frac{n-1}{2}\right)(m-1)}\right\}S^2.$$

6. Two-Parameter Exponential Distribution

6.1. Introduction

The two-parameter exponential distribution plays an important role in engineering, life testing and medical sciences. In these studies where the data are positively skewed, the exponential distribution is as important as the normal distribution is in sampling theory and agricultural statistics. Researchers have studied various aspects of estimation and inference for the two-parameter exponential distribution using either the frequentist approach or the Bayesian procedure.

However, while parameter estimation and hypothesis testing related to the two-parameter exponential distribution are well documented in the literature, the research on control charts has received little attention. Ramalhoto and Morais (1999) developed a control chart for monitoring the scale parameter while Sürücü and Sazak (2009) presented a control chart scheme in which moments are used. Mukherjee, McCracken, and Chakraborti (2014) on the other hand proposed several control charts and monitoring schemes for the location and the scale parameters of the two-parameter exponential distribution.

In this chapter control charts for the location and scale parameters as well as for a one-sided upper tolerance limit will be developed by deriving their predictive distributions and using a Bayesian procedure.

6.2. Preliminary and Statistical Results

In this section the same notation will be used as given in Krishnamoorthy and Mathew (2009).

The two-parameter exponential distribution has the probability density function

$$f(x;\mu,\theta) = \frac{1}{\theta}\exp\left\{-\frac{(x-\mu)}{\theta}\right\} \qquad x > \mu,\ \mu > 0,\ \theta > 0$$

where μ is the location parameter and θ the scale parameter.

Let $X_1, X_2, \ldots, X_n$ be a sample of n observations from the two-parameter exponential distribution. The maximum likelihood estimators for μ and θ are given by

$$\hat{\mu} = X_{(1)}$$

and

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - X_{(1)}\right) = \bar{X} - X_{(1)}$$

where $X_{(1)}$ is the minimum or the first order statistic of the sample. It is well known (see Johnson and Kotz (1970); Lawless (1982); Krishnamoorthy and Mathew (2009)) that $\hat{\mu}$ and $\hat{\theta}$ are independently distributed with

$$\frac{(\hat{\mu}-\mu)}{\theta} \sim \frac{\chi^2_2}{2n} \text{ and } \frac{\hat{\theta}}{\theta} \sim \frac{\chi^2_{2n-2}}{2n}. \tag{6.1}$$

6.3. Bayesian Procedure

If a sample of n observations are drawn from the two-parameter exponential distribution, then the likelihood function is given by

$$L(\mu,\theta|data) = \left(\frac{1}{\theta}\right)^n \exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)\right\}.$$

As prior the Jeffreys' prior

$$p(\mu,\theta) \propto \theta^{-1}$$

will be used.

The following theorems can now be proved.

Theorem 6.1. *The joint posterior distribution of μ and θ is*

$$\begin{aligned} p(\theta,\mu|data) &\propto p(\mu,\theta)\,L(\mu,\theta|data) \\ &= K_1\left(\tfrac{1}{\theta}\right)^{n+1}\exp\left\{-\tfrac{n}{\theta}(\bar{x}-\mu)\right\} \qquad 0<\mu<x_{(1)},\ 0<\theta<\infty \end{aligned} \tag{6.2}$$

where

$$K_1 = \frac{n^n}{\Gamma(n-1)}\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1}-\left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Theorem 6.2. *The posterior distribution of μ is*

$$p(\mu|data) = (n-1)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1}-\left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}(\bar{x}-\mu)^{-n} \quad 0<\mu<x_{(1)}$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Theorem 6.3. *The posterior distribution of θ is*

$$\begin{aligned} p(\theta|data) &= \textstyle\int_0^\infty p(\theta,\mu|data)\,d\mu \\ &= K_1\tfrac{1}{n}\left(\tfrac{1}{\theta}\right)^n\left\{\exp\left(-\tfrac{n\hat{\theta}}{\theta}\right)-\exp\left(-\tfrac{n\bar{x}}{\theta}\right)\right\} \qquad 0<\theta<\infty \end{aligned} \tag{6.3}$$

Proof. The proof follows easily from the proof of Theorem 6.1. □

Theorem 6.4. *The conditional posterior distribution of μ given θ is*

$$\begin{aligned} p(\mu|\theta,data) &= \tfrac{p(\theta,\mu|data)}{p(\theta|data)} \\ &= K_2\exp\left(\tfrac{n\mu}{\theta}\right) \end{aligned} \tag{6.4}$$

where

$$K_2 = \frac{n}{\theta}\left\{\exp\left(\frac{nx_{(1)}}{\theta}\right) - 1\right\}^{-1}.$$

Proof. The proof follows easily from the proof of Theorem 6.1. □

Theorem 6.5. *The conditional posterior distribution of θ given μ is*

$$\begin{aligned} p(\theta|\mu, data) &= \frac{p(\theta,\mu|data)}{p(\mu|data)} \\ &= K_3\left(\frac{1}{\theta}\right)^{n+1}\exp\left\{-\frac{n}{\theta}(\bar{x}-\mu)\right\} \qquad 0<\theta<\infty \end{aligned} \tag{6.5}$$

where

$$K_3 = \frac{\{n(\bar{x}-\mu)\}^n}{\Gamma(n)}.$$

Proof. Since

$$\begin{aligned} K_3^{-1} &= \int_0^\infty \left(\frac{1}{\theta}\right)^{n+1}\exp\left\{-\frac{n}{\theta}(\bar{x}-\mu)\right\}d\theta \\ &= \frac{\Gamma(n)}{\{n(\bar{x}-\mu)\}^n}, \end{aligned}$$

the result follows. □

6.4. The Predictive Distributions of Future Sample Location and Scale Maximum Likelihood Estimators, $\hat{\mu}_f$ and $\hat{\theta}_f$

Consider a future sample of m observations from the two-parameter exponential population: $X_{1f}, X_{2f}, \ldots, X_{mf}$. The future sample mean is defined as $\bar{X}_f = \frac{1}{m}\sum_{j=1}^{m} X_{jf}$. The smallest value in the sample is denoted by $\hat{\mu}_f$ and $\hat{\theta}_f = \bar{X}_f - \hat{\mu}_f$. To obtain control charts for $\hat{\mu}_f$ and $\hat{\theta}_f$ their predictive distributions must first be derived.

The following theorems can now be proved.

Theorem 6.6. *The predictive distribution of a future sample location maximum likelihood estimator, $\hat{\mu}_f$, is given by*

$$f\left(\hat{\mu}_f|data\right) = \begin{cases} K^* \left\{ \left[\frac{1}{n\left(\bar{x}-\hat{\mu}_f\right)} \right]^n - \left[\frac{1}{m\hat{\mu}_f + n\bar{x}} \right]^n \right\} & 0 < \mu_f < x_{(1)} \\ \\ K^* \left\{ \left[\frac{1}{m\left(\hat{\mu}_f - x_{(1)}\right) + n\hat{\theta}} \right]^n - \left[\frac{1}{m\hat{\mu}_f + n\bar{x}} \right]^n \right\} & x_{(1)} < \hat{\mu}_f < \infty \end{cases} \tag{6.6}$$

where

$$K^* = \frac{n^n \left(n-1\right) m}{\left(n+m\right)\left\{ \left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1} \right\}}.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Theorem 6.7. *The mean of $\hat{\mu}_f$ is given by*

$$E\left(\hat{\mu}_f|data\right) = \bar{x} - \tilde{K}L\left(1-a\right)$$

and the variance by

$$Var\left(\hat{\mu}_f|data\right) = \left\{ \frac{n^3}{m^2\left(n-1\right)^2\left(n-2\right)} + \left(1-a\right)^2 \right\} \tilde{K}M - \left(1-a\right)^2 \tilde{K}^2 L^2$$

where

$$a = \frac{n}{m\left(n-1\right)},$$

$$\tilde{K} = \left(n-1\right)\left\{ \left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1} \right\}^{-1},$$

$$L = \left(\frac{1}{n-2}\right)\left\{ \left(\frac{1}{\hat{\theta}}\right)^{n-2} - \left(\frac{1}{\bar{x}}\right)^{n-2} \right\}$$

and

$$M = \left(\frac{1}{n-3}\right)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-3} - \left(\frac{1}{\bar{x}}\right)^{n-3}\right\}.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Theorem 6.8. *The predictive distribution of a future sample scale maximum likelihood estimator, $\hat{\theta}_f$, is given by*

$$\begin{aligned} f\left(\hat{\theta}_f|data\right) \quad = \quad & m^{m-1}n^{n-1}\tfrac{\Gamma(m+n-2)}{\Gamma(m-1)\Gamma(n-1)}\left\{\left(\tfrac{1}{\hat{\theta}}\right)^{n-1} - \left(\tfrac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}\hat{\theta}_f^{m-2}\times \\ & \left\{\left(\tfrac{1}{m\hat{\theta}_f+n\hat{\theta}}\right)^{m+n-2} - \left(\tfrac{1}{m\hat{\theta}_f+n\bar{x}}\right)^{m+n-2}\right\} \qquad \hat{\theta}_f > 0. \end{aligned} \tag{6.7}$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Theorem 6.9. *The mean and variance of $\hat{\theta}_f$ is given by*

$$E\left(\hat{\theta}_f|data\right) = \frac{n\,(m-1)}{m\,(n-1)}\tilde{K}L$$

and variance

$$Var\left(\hat{\theta}_f|data\right) = \frac{n^2\,(m-1)}{m\,(n-1)}\left\{\frac{\tilde{K}M}{n-2} - \frac{(m-1)}{m\,(n-1)}\tilde{K}^2L^2\right\}.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

6.5. Example

The following data is given in Grubbs (1971) as well as in Krishnamoorthy and Mathew (2009). The failure mileages given in Table 6.1 fit a two-parameter exponential distribution.

Table 6.1.: Failure Mileages of 19 Military Carriers

162	200	271	302	393	508	539	629	706	777
884	1008	1101	1182	1463	1603	1984	2355	2880	

For this data, the maximum likelihood estimates are $\hat{\mu} = x_{(1)} = 162$, $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - x_{(1)}\right) = \bar{x} - x_{(1)} = 835.21$ and $n = 19$.

As mentioned in the introductory section, the aim of this chapter is to obtain control charts for location and scale maximum likelihood estimators as well as for a one-sided upper tolerance limit.

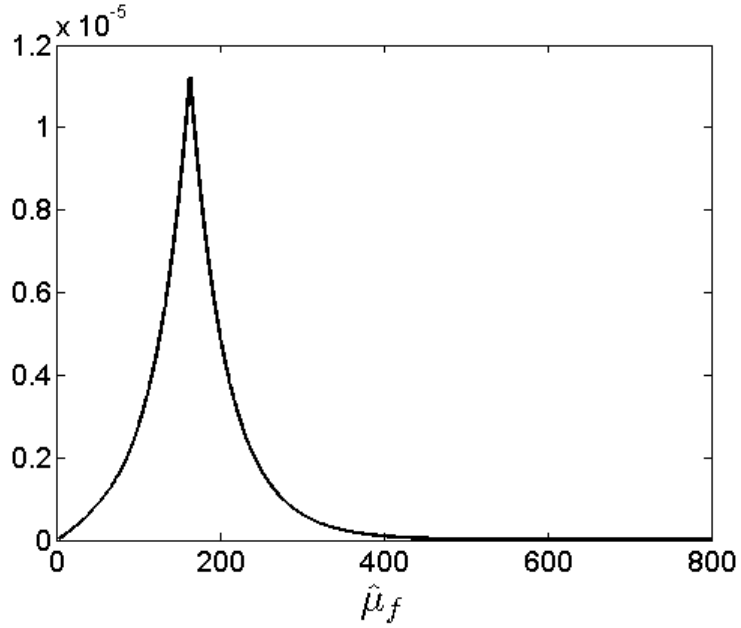
Therefore in the next section a control chart for a future location maximum likelihood estimator will be developed.

6.6. Control Chart for $\hat{\mu}_f$

By using the predictive distribution (defined in Equation (6.6)) a Bayesian procedure will be developed in Phase II to obtain a control chart for $\hat{\mu}_f$. Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length.

For the example given in Table 6.1 (failure mileage data) the predictive distribution, $f\left(\hat{\mu}_f|data\right)$ for $m = 19$ future data is illustrated in Figure 6.1.

Figure 6.1.: Distribution of $\hat{\mu}_f$, $n = 19$, $m = 19$



$Mean\,(\hat{\mu}_f) = 168.78$, $Median\,(\hat{\mu}_f) = 163.91$, $Mode\,(\hat{\mu}_f) = 162$, $Var\,(\hat{\mu}_f) = 3888.7$
95% $Interval\,(\hat{\mu}_f) = (55.18; 317.45)$

99.73% $Interval\,(\hat{\mu}_f) = (13.527; 489.52)$

From Figure 6.1 it follows that for a 99.73% two-sided control chart the lower control limit is $LCL = 13.527$ and the upper control limit is $UCL = 489.52$.

Let $R\,(\beta)$ represents those values of $\hat{\mu}_f$ that are smaller than LCL and larger than UCL.

The run-length is defined as the number of future $\hat{\mu}_f$ values (r) until the control chart signals for the first time (Note that r does not include the $\hat{\mu}_f$ value when the control chart signals). Given μ and θ and a stable Phase I process, the distribution of the run-length r is Geometric with parameter

$$\psi\,(\mu, \theta) = \int_{R(\beta)} f\,(\hat{\mu}_f|\mu, \theta)\, d\hat{\mu}_f$$

where

$$f\,(\hat{\mu}_f|\mu, \theta) = \frac{m}{\theta} \exp\left\{-\frac{m}{\theta}\,(\hat{\mu}_f - \mu)\right\} \qquad \hat{\mu}_f > \mu$$

i.e., the distribution of $\hat{\mu}_f$ given that of μ and θ are known. See also Equation (6.1). The values of μ and θ are however unknown and the uncertainty of these parameter values are described by their joint posterior distribution $p(\theta, \mu|data)$ given in Equation (6.2).

By simulating μ and θ from $p(\theta, \mu|data) = p(\theta|\mu, data)\, p(\mu|data)$ the probability density function of $f(\hat{\mu}_f|\mu, \theta)$ as well as the parameter $\psi(\mu, \theta)$ can be obtained. This must be done for each future sample. In other words, for each future sample μ and θ must first be simulated from $p(\theta, \mu|data)$ and then $\psi(\mu, \theta)$ calculated. Therefore, by simulating all possible combinations of μ and θ from their joint posterior distribution a large number of $\psi(\mu, \theta)$ values can be obtained. Also, a large number of geometric distributions, i.e., a large number of run-length distributions each with a different parameter value $(\psi(\mu_1, \theta_1), \psi(\mu_2, \theta_2), \ldots, \psi(\mu_m, \theta_m))$ can be obtained.

The unconditional moments, $E(r|data)$, $E(r^2|data)$ and $Var(r|data)$ can therefore be obtained by simulation or numerical integration. For further details refer to Menzefricke (2002, 2007, 2010b,a).

The mean of the predictive distribution of the run-length for the 99.73% two-sided control limits is $E(r|data) = 37526$, much larger than the 370 that one would have expected if $\beta = 0.0027$. The reason for this large average run-length is the small sample size and large variation in the data. The median run-length = 1450. Define $\tilde{\psi}(\mu, \theta) = \frac{1}{m}\sum_{i=1}^{m} \psi(\mu_i, \theta_i)$. From Menzefricke (2002) it follows that if $m \to \infty$, then $\tilde{\psi}(\mu, \theta) \to \beta$ and the harmonic mean of $r = \frac{1}{\beta}$. For $\beta = 0.0027$ the harmonic mean would therefore be 370.

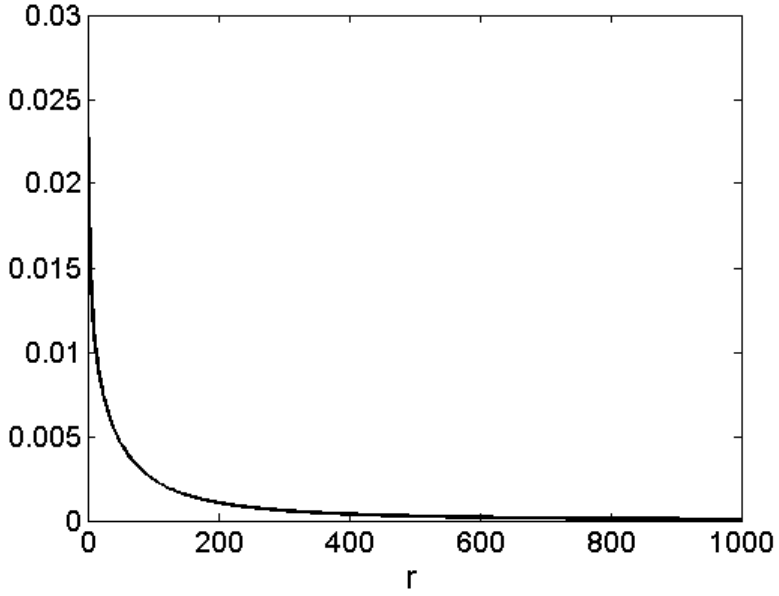
In Table 6.2 the average run-length for different values of β are given.

Table 6.2.: β Values and Corresponding Average Run-length

β	0.0027	0.003	0.005	0.007	0.009	0.01	0.015	0.02	0.025	0.0258	0.03
$E(r\|data)$	37526	32610	10231	4655	2807	2313	987.6	594.2	399.7	369.67	280.1

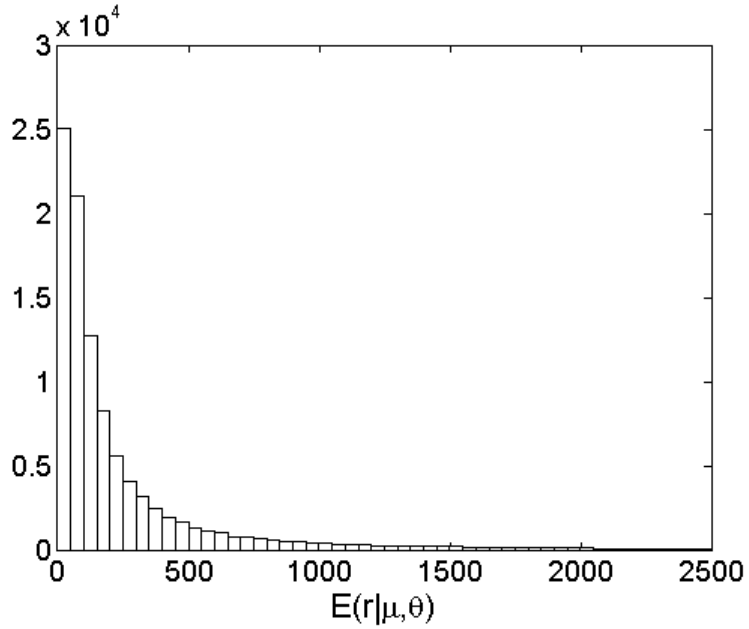
In Figure 6.2 the distribution of the run-length for $n = 19$, $m = 19$ and $\beta = 0.0258$ is illustrated and in Figure 6.3 the histogram of the expected run-length is given.

Figure 6.2.: Run-length, $n = 19$, $m = 19$, $\beta = 0.0258$



$E\,(r|data) = 369.67$, $Median(r|data) = 71.22$, $Var\,(r|data) = 6.5275 \times 10^6$
$95\%\ interval\,(r|data) = (0; 1242.2)$

Figure 6.3.: Expected Run-length, $n = 19$, $m = 19$, $\beta = 0.0258$



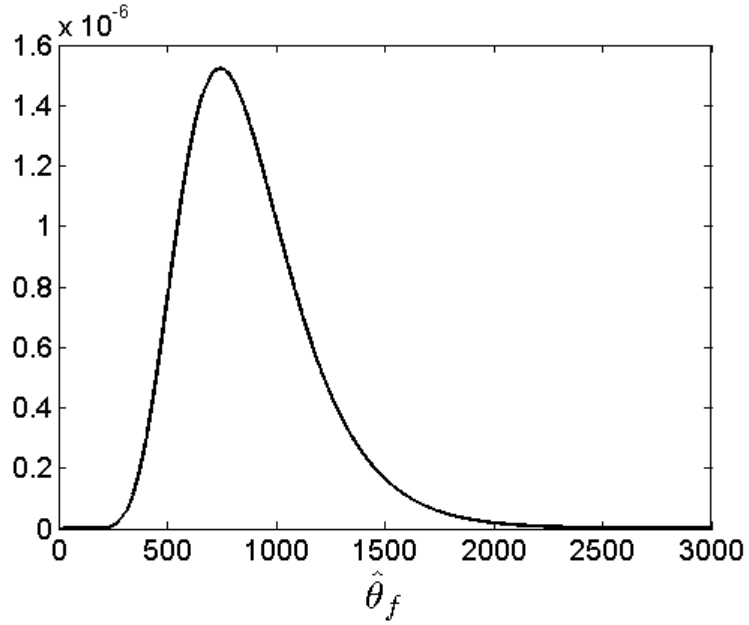
$Mean = 364.1$, $Median = 111.93$, $Var = 3.7157 \times 10^6$, $Mean\,(\psi) = 0.0261$
$95\%\ interval = (0; 1194.3)$

It is therefore clear that β must be adjusted to get a run-length of 370.

6.7. Control Chart for $\hat{\theta}_f$

In this section a control chart for $\hat{\theta}_f$, a future scale maximum likelihood estimator, will be developed. The predictive distribution $f\left(\hat{\theta}_f|data\right)$ given in Equation (6.7) is displayed in Figure 6.4 for the example previously given and $m = 19$.

Figure 6.4.: Predictive Distribution of $\hat{\theta}_f$, $n = 19$, $m = 19$



$Mean\left(\hat{\theta}_f\right) = 876.73$, $Median\left(\hat{\theta}_f\right) = 829.23$, $Mode\left(\hat{\theta}_f\right) = 743.34$

For a 99.73% two-sided control chart the lower control limit is LCL = 297.5 and the upper control limit is UCL = 2278. $\tilde{R}(\beta)$ represents those values of $\hat{\theta}_f$ that are smaller than LCL and larger than UCL. Given θ and a stable Phase I process, the distribution of the run-length is Geometric with parameter

$$\psi(\theta) = \int_{\tilde{R}(\beta)} f\left(\hat{\theta}_f|\theta\right) d\hat{\theta}_f$$

where $f\left(\hat{\theta}_f|\theta\right)$ is defined in Equation (6.1). As before the value of θ is unknown, but can be simulated from the posterior distribution $p(\theta|data)$ given in Equation (6.3)

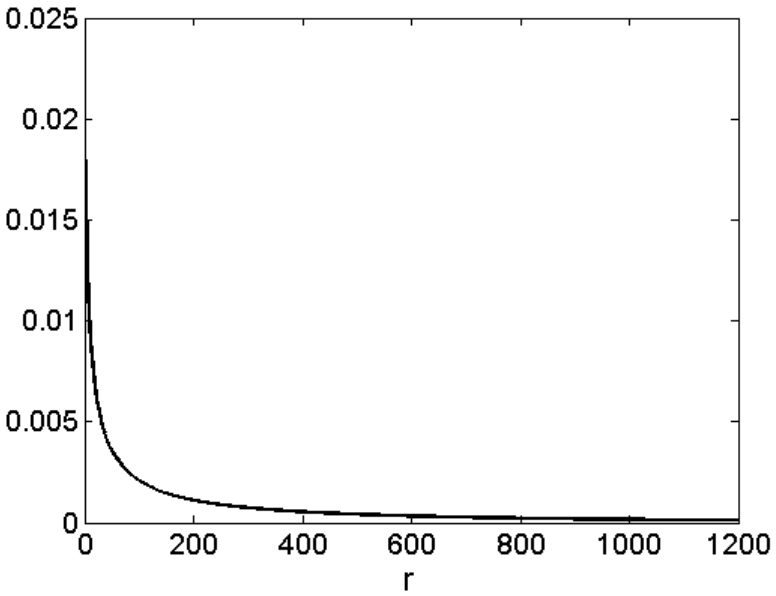
or equivalently by first simulating μ from $p(\mu|data)$ and then θ from $p(\theta|\mu, data)$. By simulating θ the probability density function $f\left(\hat{\theta}_f|\theta\right)$ Equation (6.3) as well as the parameter $\psi(\theta)$ can be obtained. As mentioned previously, this must be done for each future sample. The mean of the predictive distribution of the run-length for the 99.73% two-sided control limits is $E(r|data) = 8188.6$. As in the case of $\hat{\mu}_f$, this is much larger than the 370 that one would have expected if $\beta = 0.0027$. In Table 6.3 the average run-length versus probabilities β are given.

Table 6.3.: β Values and Corresponding Average Run-length

β	0.0027	0.003	0.005	0.007	0.009	0.01	0.015	0.018	0.02	0.025
$E(r\|data)$	10939.2	9029.8	3517.7	1913	1238.2	1010.7	512.1	372.4	311.2	211.8

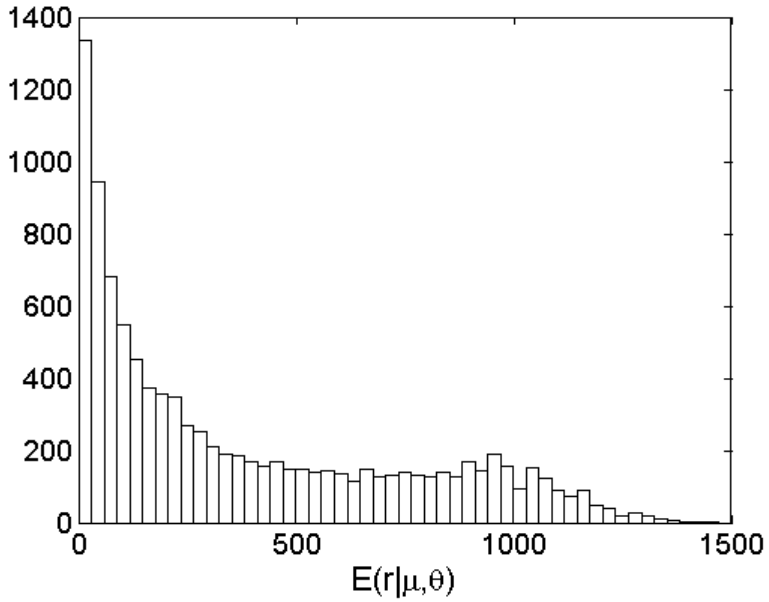
In Figure 6.5 the distribution of the run-length for $n = 19$, $m = 19$ and $\beta = 0.018$ is illustrated and in Figure 6.6 the histogram of the expected run-length is given.

Figure 6.5.: Run-length, $n = 19$, $m = 19$, $\beta = 0.018$



$E(r|data) = 372.4$, $Median(r|data) = 127.4$, $Var(r|data) = 3.9515 \times 10^5$
$95\%\ interval(r|data) = (0; 1602.5)$

Figure 6.6.: Expected Run-length, $n = 19$, $m = 19$, $\beta = 0.018$



$Mean = 372.4$, $Median = 238.38$, $Var = 1.2698 \times 10^5$, $Mean\,(\psi) = 0.0177$
95% $interval = (7.809; 1096.7)$

From Table 6.3 it is clear how β should be adjusted to get a specific run-length. For example if one is interested in a run-length of a 1,000 then β should be 0.01.

6.8. A Bayesian Control Chart for a One-sided Upper Tolerance Limit

As mentioned in the introduction, a confidence interval for a quantile is called a tolerance interval. A one-sided upper tolerance limit is therefore a quantile of a quantile. It can easily be shown that the p quantile of a two-parameter exponential distribution is given by $q_p = \mu - \theta \ln(1-p)$. By replacing the parameters by their generalized pivotal quantities (GPQs), Krishnamoorthy and Mathew (2009) showed that a GPQ for q_p can be obtained as

$$G_{q_p} = \hat{\mu} - \left[\frac{\chi_2^2 + 2n\ln(1-p)}{\chi_{2n-2}^2}\right]\hat{\theta}.$$

Let $E_{p;\alpha}$ denotes the α quantile of $E_p = \frac{\chi^2_2 + 2n\ln(1-p)}{\chi^2_{2n-2}}$, then

$$\hat{\mu} - E_{p;\alpha}\hat{\theta} = \hat{\mu} - \tilde{k}_2\hat{\theta} \tag{6.8}$$

is a $1-\alpha$ upper confidence limit for q_p, which means that $(p, 1-\alpha)$ is an upper tolerance limit for the *exponential* (μ, θ) distribution.

An advantage of the upper tolerance limit is that it monitors the location and scale parameters of the two-parameter exponential distribution at the same time.

It is shown in Roy and Mathew (2005) and Krishnamoorthy and Mathew (2009) that the upper and lower tolerance limits are actually exact, which means that they have the correct frequentist coverage probabilities.

6.9. The Predictive Distribution of a Future Sample Upper Tolerance Limit

From Equation (6.8) it follows that a future sample upper tolerance limit is defined as

$$U_f = \hat{\mu}_f - \tilde{k}_2\hat{\theta}_f \text{ where } \hat{\mu}_f > \mu \text{ and } \hat{\theta}_f > 0.$$

From Equation (6.1) it follows that

$$f\left(\hat{\mu}_f|\mu,\theta\right) = \left(\frac{m}{\theta}\right)\exp\left\{-\frac{m}{\theta}\left(\hat{\mu}_f - \mu\right)\right\} \;\; \hat{\mu}_f > \mu$$

which means that

$$f\left(U_f|\mu,\theta,\hat{\theta}_f\right) = \left(\frac{m}{\theta}\right)\exp\left\{-\frac{m}{\theta}\left[U_f - \left(\mu - \tilde{k}_2\hat{\theta}_f\right)\right]\right\} \qquad U_f > \mu - \tilde{k}_2\hat{\theta}_f \tag{6.9}$$

From Equation (6.9) it can be seen that the derivation of the unconditional predictive density function $f\left(U_f|data\right)$ will be quite complicated. An approximation of

the density function can however be obtained by using the following Monte Carlo simulation procedure:

1. Simulate μ and θ from $p(\theta, \mu|data)$. This can be achieved by first simulating μ from $p(\mu|data)$ defined in Equation (6.2) and then θ from $p(\theta|\mu, data)$ defined in Equation (6.5).
2. For given θ, simulate $\hat{\theta}_f$ from $\frac{\chi^2_{2m-2}}{2m}\theta$.
3. Substitute the simulated μ, θ, and $\hat{\theta}_f$ values in Equation (6.9) and draw the exponential distribution.

Repeat this procedure l times and obtain the average of the l simulated exponential density functions (Rao-Blackwell method) to obtain the unconditional predictive density $f(U_f|data)$.

Although (as mentioned) the derivation of the exact unconditional predictive density will be quite complicated the exact moments can be derived analytically. The following theorem can be proved:

Theorem 6.10. *The exact mean and variance of $U_f = \hat{\mu}_f - \tilde{k}_2\hat{\theta}_f$, a future sample tolerance limit, is given by*

$$E(U_f|data) = \bar{x} + \tilde{K}L(aH - 1) \tag{6.10}$$

and

$$Var(U_f|data) = \frac{n^2}{m^2(n-1)(n-2)}\left\{J + \frac{H^2}{n-1}\right\}\tilde{K}M + (1-aH)^2\left\{\tilde{K}M - \tilde{K}^2L^2\right\} \tag{6.11}$$

where

$$H = \left\{1 - \tilde{k}_2(m-1)\right\}$$

and

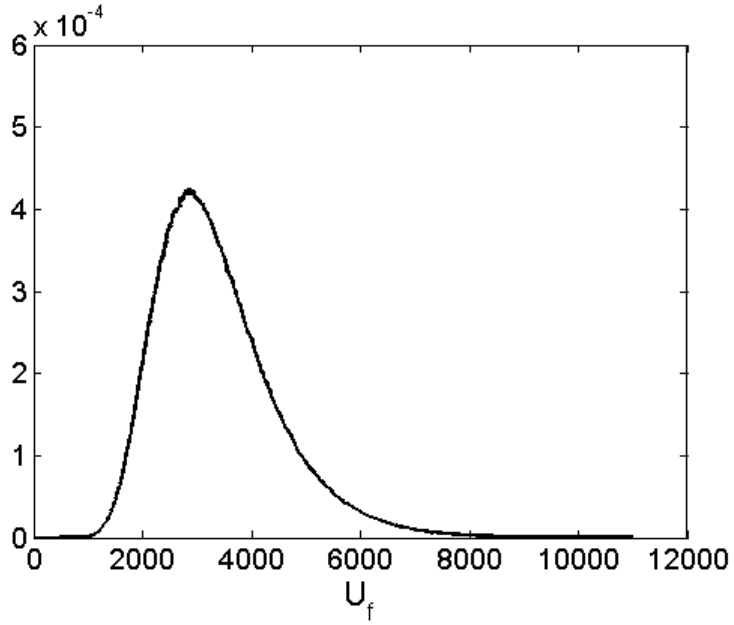
$$J = 1 + \tilde{k}_2^2 (m - 1) .$$

$\tilde{K}$, L, M and a are defined in Theorem 6.7.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

For the failure mileage data given in Table 6.1, $\tilde{k}_2 = -3.6784$ if $m = 19$ and by using $10,000,000$ Monte Carlo simulations as described in (item 1), (item 2) and (item 3), the unconditional predictive density function can be obtained and is illustrated in Figure 6.7. Figure 6.7 is therefore the distribution of an upper future tolerance limit for the mileages of the next 19 military personnel carriers that will fail in service.

Figure 6.7.: Predictive Density of $f(U_f|data)$



$$Mean(U_f) = 3394.7,\ Mode(U_f) = 2900,\ Median(U_f) = 3211.5,$$
$$Var(U_f) = 1.2317 \times 10^6$$
$$95\%\ interval(U_f) = (1736.5; 6027)$$

$$99.73\%\ interval(U_f) = (1249.05, 7973)$$

As in the previous sections the predictive distribution can be used to derive the run-length and average run-length. From Figure 6.7 it follows that for a 99.73%

two-sided control chart the lower control limit is LCL = 1249.05 and the upper control limit is UCL = 7973. $R^*(\beta)$ therefore represents those values of U_f that are smaller than LCL and larger than UCL. As before, the run-length is defined as the number of future U_f values (r) until the control chart signals for the first time. Given μ and θ the distribution of the run-length r is Geometric with parameter

$$\tilde{\psi}(\mu,\theta) = \int_{R^*(\beta)} f(U_f|\mu,\theta)\, dU_f$$

where $f(U_f|\mu,\theta)$ is the distribution of a future U_f given that μ and θ are known. As mentioned before the values of μ and θ are however unknown and the uncertainty of these parameter values are described by their joint posterior distribution $p(\mu,\theta|data)$.

By simulating μ and θ from $p(\mu,\theta|data)$, the probability density function $f(U_f|\mu,\theta)$ can be obtained from Equations (6.1) and (6.9) in the following way.

I. $f\left(U_f|\mu,\theta,\chi^2_{2m-2}\right) = \left(\frac{m}{\theta}\right)\exp\left\{-\frac{m}{\theta}\left[U_f - \left(\mu - \tilde{k}_2\frac{\chi^2_{2m-2}}{2m}\theta\right)\right]\right\}$.

II. The next step is to simulate $l^* = 100000$ χ^2_{2m-2} values to obtain l^* exponential density functions for given μ and θ.

III. By averaging the l^* density functions (Rao-Blackwell method) $f(U_f|\mu,\theta)$ can be obtained and also $\tilde{\psi}(\mu,\theta)$.

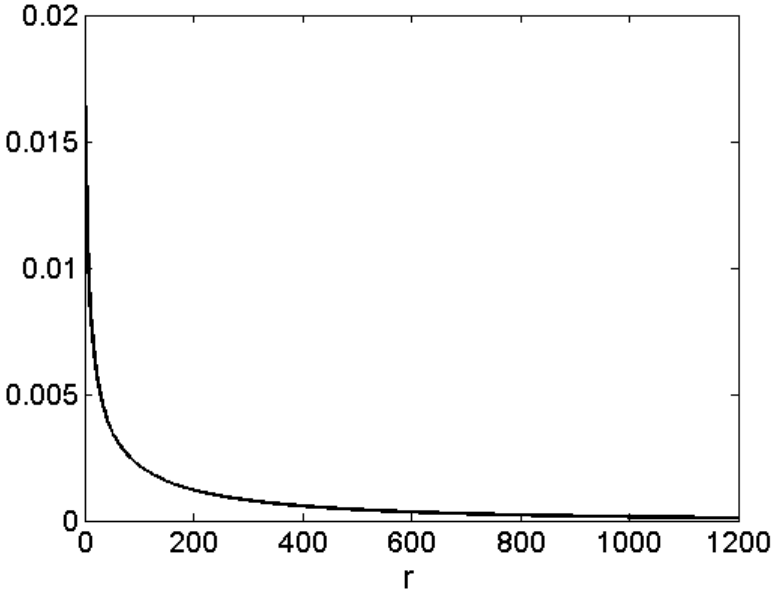
This must be done for each future sample. In other words, for each future sample μ and θ must be simulated from $p(\mu,\theta|data)$ and then the steps described in (item I.), (item II.) and (item III.). The mean of the predictive distribution of the run-length for the 99.73% two-sided control limits is $E(r|data) = 1.1709 \times 10^{11}$, much larger than the 370 that would have been expected for $\beta = 0.0027$. In Table 6.4 the average run-lengths versus probabilities β are given.

Table 6.4.: β Values and Corresponding Average Run-length

β	0.007	0.009	0.01	0.015	0.018	0.02	0.025
$E(r\|data)$	10000	1240.2	1020.7	530.1	374.2	300.2	208.8

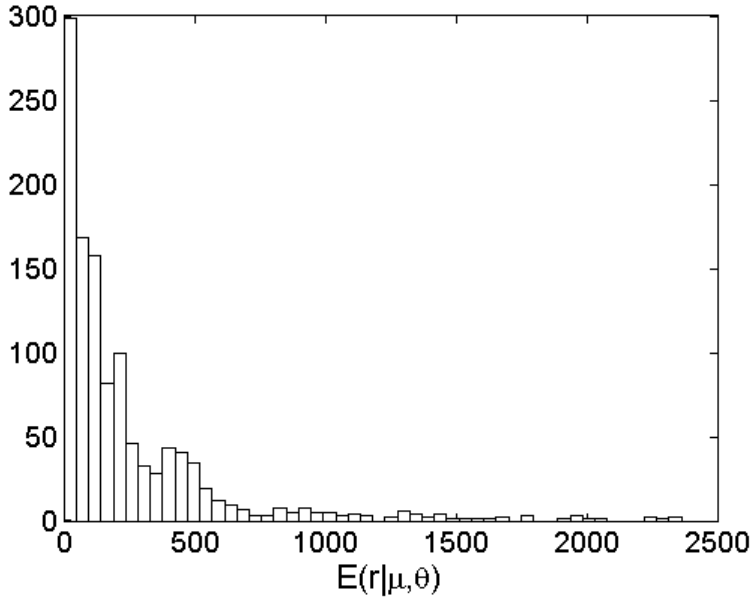
In Figure 6.8 the distribution of the run-length for $n = 19$, $m = 19$ and $\beta = 0.018$ is illustrated and Figure 6.9 the histogram of the expected run-length is given. The reason for $\beta = 0.018$ is because it gives and average run-length of approximately 370.

Figure 6.8.: Distribution of Run-length when $\beta = 0.018$



$E\left(r|data\right) = 374.2$, $Median\left(r|data\right) = 132.1$, $Var\left(r|data\right) = 5.8236 \times 10^5$
95% $interval\left(r|data\right) = (0; 1803.6)$

Figure 6.9.: Expected Run-length when $\beta = 0.018$



$Mean = 374.2$, $Median = 248.03$, $Var = 2.0351 \times 10^5$
95% $interval = (3.5541; 1463.2)$

6.10. Conclusion

This chapter develops a Bayesian control chart for monitoring the scale parameter, location parameter and upper tolerance limit of a two-parameter exponential distribution. In the Bayesian approach prior knowledge about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution. By using the posterior distribution the predictive distributions of $\hat{\mu}_f$, $\hat{\theta}_f$ and U_f can be obtained.

The theory and results described in this chapter have been applied to the failure mileages for military carriers analyzed by Grubbs (1971) and Krishnamoorthy and Mathew (2009). The example illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions and run-lengths for $\hat{\mu}_f$, $\hat{\theta}_f$ and U_f.

Mathematical Appendix to Chapter 6

Proof of Theorem 6.1

$$\begin{aligned} K_1^{-1} &= \int_0^{\infty}\int_0^{x_{(1)}} \left(\tfrac{1}{\theta}\right)^{n+1} \exp\left\{-\tfrac{n}{\theta}\left(\bar{x}-\mu\right)\right\} d\mu d\theta \\ &= \int_0^{\infty} \left(\tfrac{1}{\theta}\right)^{n+1} \exp\left\{-\tfrac{n}{\theta}\bar{x}\right\} \left[\int_0^{x_{(1)}} \exp\left(\tfrac{n\mu}{\theta}\right) d\mu\right] d\theta \end{aligned}$$

Since

$$\int_0^{x_{(1)}} \exp\left(\frac{n\mu}{\theta}\right) d\mu = \frac{\theta}{n}\left\{\exp\left(\frac{nx_{(1)}}{\theta}\right) - 1\right\}$$

it follows that

$$\begin{aligned} K_1^{-1} &= \tfrac{1}{n}\int_0^{\infty} \left(\tfrac{1}{\theta}\right)^{n} \exp\left\{-\tfrac{n\hat{\theta}}{\theta}\right\} d\theta - \tfrac{1}{n}\int_0^{\infty} \left(\tfrac{1}{\theta}\right)^{n} \exp\left\{-\tfrac{n\bar{x}}{\theta}\right\} d\theta \\ &= \left(\tfrac{1}{n}\right)^{n} \left(\tfrac{1}{\hat{\theta}}\right)^{n-1} \Gamma\left(n-1\right) - \left(\tfrac{1}{n}\right)^{n} \left(\tfrac{1}{\bar{x}}\right)^{n-1} \Gamma\left(n-1\right) \\ &= \left(\tfrac{1}{n}\right)^{n} \Gamma\left(n-1\right) \left\{\left(\tfrac{1}{\hat{\theta}}\right)^{n-1} - \left(\tfrac{1}{\bar{x}}\right)^{n-1}\right\}. \end{aligned}$$

Since $K_1 = \left(K_1^{-1}\right)^{-1}$ the theorem follows.

Proof of Theorem 6.2

$$\begin{aligned} p(\mu|data) &= \int_0^\infty p(\mu,\theta|data)\,d\theta \\ &= K_1 \int_0^\infty \left(\tfrac{1}{\theta}\right)^{n+1} \exp\left\{-\tfrac{n}{\theta}(\bar{x}-\mu)\right\} d\theta \\ &= K_1 \left\{\tfrac{1}{n(\bar{x}-\mu)}\right\}^n \Gamma(n). \end{aligned}$$

By substituting K_1 the result follows.

Proof of Theorem 6.6

From Equation (6.1) it follows that

$$\hat{\mu}_f|\mu,\theta \sim \frac{\chi_2^2}{2m}\theta + \mu \qquad \hat{\mu}_f > \mu.$$

Therefore

$$f(\hat{\mu}_f|\mu,\theta) = \left(\frac{m}{\theta}\right) \exp\left\{-\frac{m}{\theta}(\hat{\mu}_f - \mu)\right\} \qquad \hat{\mu}_f > \mu.$$

Further

$$f(\hat{\mu}|data) = \int\int f(\hat{\mu}|\mu,\theta)\, p(\theta|\mu,data)\, p(\mu|data)\, d\theta d\mu$$

where

$$p(\theta|\mu,data) = \frac{\{n(\bar{x}-\mu)\}^n}{\Gamma(n)} \left(\frac{1}{\theta}\right)^{n+1} \exp\left\{-\frac{n}{\theta}(\bar{x}-\mu)\right\} \qquad 0 < \theta < \infty$$

and

$$p(\mu|data) = (n-1)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1} (\bar{x}-\mu)^{-n} \qquad 0 < \mu < x_{(1)}.$$

Now

$$\begin{aligned} f\left(\hat{\mu}_f|\mu, data\right) &= \int_0^\infty f\left(\hat{\mu}_f|\mu,\theta\right) p\left(\theta|\mu, data\right) d\theta \\ &= \frac{m\{n(\bar{x}-\mu)\}^n}{\Gamma(n)} \int_0^\infty \left(\frac{1}{\theta}\right)^{n+2} \exp\left\{-\frac{1}{\theta}\left[m\left(\hat{\mu}_f-\mu\right)+n\left(\bar{x}-\mu\right)\right]\right\} d\theta \\ &= \frac{n^{n+1}(\bar{x}-\mu)^n m}{\left[m\left(\hat{\mu}_f-\mu\right)+n(\bar{x}-\mu)\right]^{n+1}} \qquad \hat{\mu}_f > 0 \end{aligned}$$

and

$$\begin{aligned} f\left(\hat{\mu}_f|data\right) &= \int_0^{\hat{\mu}_f} f\left(\hat{\mu}_f,\mu|data\right) d\mu \qquad 0 < \hat{\mu}_f < x_{(1)} \\ &= \int_0^{x_{(1)}} f\left(\hat{\mu}_f,\mu|data\right) d\mu \qquad x_{(1)} < \hat{\mu}_f < \infty \end{aligned}$$

where

$$f\left(\hat{\mu}_f,\mu|data\right) = \frac{n^{n+1}\left(n-1\right)m}{\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}} \left\{\left(m\hat{\mu}_f + n\bar{x}\right) - \mu\left(n+m\right)\right\}^{-(n-1)}.$$

Therefore

$$\begin{aligned} f\left(\hat{\mu}_f|data\right) &= K^*\left\{\left[\frac{1}{n\left(\bar{x}-\hat{\mu}_f\right)}\right]^n - \left[\frac{1}{m\hat{\mu}_f+n\bar{x}}\right]^n\right\} \qquad 0 < \hat{\mu}_f < x_{(1)} \\ &= K^*\left\{\left[\frac{1}{m\left(\hat{\mu}_f-x_{(1)}\right)+n\hat{\theta}}\right]^n - \left[\frac{1}{m\hat{\mu}_f+n\bar{x}}\right]^n\right\} \qquad x_{(1)} < \hat{\mu}_f < \infty \end{aligned}$$

and

$$K^* = \frac{n^n\left(n-1\right)m}{\left(n+m\right)}\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}.$$

Proof of Theorem 6.7

Expected Value of $\hat{\mu}_f$

It follows from Equation (6.1) that

$$\hat{\mu}_f \sim \theta \frac{\chi_2^2}{2m} + \mu.$$

Therefore

$$E\left(\hat{\mu}_f | \mu, \theta\right) = \frac{\theta}{m} + \mu.$$

From Equation (6.2) it follows that

$$p\left(\theta | \mu, data\right) = \frac{\left\{n\left(\bar{x} - \mu\right)\right\}^n}{\Gamma\left(n\right)} \left(\frac{1}{\theta}\right)^{n+1} \exp\left\{-\frac{n}{\theta}\left(\bar{x} - \mu\right)\right\}$$

which means that

$$E\left(\theta | \mu, data\right) = \frac{n\left(\bar{x} - \mu\right)}{\left(n - 1\right)}$$

and therefore

$$\begin{aligned} E\left(\hat{\mu}_f | \mu, data\right) &= \tfrac{n(\bar{x}-\mu)}{m(n-1)} + \mu \\ &= \left(1 - a\right)\mu + a\bar{x} \end{aligned} \tag{6.12}$$

where

$$a = \frac{n}{m\left(n - 1\right)}.$$

Also from Equation (6.2) it follows that

$$p\left(\mu | data\right) = \tilde{K}\left(\bar{x} - \mu\right)^{-n} \qquad 0 < \mu < x_{(1)}$$

where

$$\tilde{K} = \left(n - 1\right)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}.$$

Now

$$E\left(\mu|data\right) = -E\left\{\left(\bar{x}-\mu\right)|data\right\} + \bar{x}$$

and

$$E\left\{\left(\bar{x}-\mu\right)|data\right\} = \frac{\tilde{K}}{(n-2)}\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-2} - \left(\frac{1}{\bar{x}}\right)^{n-2}\right\} = \tilde{K}L$$

which means that

$$E\left(\mu|data\right) = -\tilde{K}L + \bar{x}. \tag{6.13}$$

Substitute Equation (6.13) in (6.12) and the result follows as

$$E\left(\hat{\mu}_f|data\right) = \bar{x} - \tilde{K}L\left(1-a\right).$$

Variance of $\hat{\mu}_f$

From Equation (6.1) it also follows that

$$Var\left(\hat{\mu}_f|\mu, data\right) = \theta^2\frac{4}{4m^2} = \frac{\theta^2}{m^2}.$$

Further

$$Var\left(\hat{\mu}_f|\mu, data\right) = E_{\theta|\mu}\left\{Var\left(\hat{\mu}_f|\mu,\theta\right)\right\} + Var_{\theta|\mu}\left\{E\left(\hat{\mu}_f|\mu,\theta\right)\right\}.$$

Since

$$E\left(\theta^2|\mu, data\right) = \frac{\left\{n\left(\bar{x}-\mu\right)\right\}^2}{(n-1)(n-2)}$$

and

$$Var\left(\theta|\mu, data\right) = \frac{\left\{n\left(\bar{x}-\mu\right)\right\}^2}{(n-1)^2(n-2)}$$

it follows that

$$\begin{aligned} Var\left(\hat{\mu}_f|\mu, data\right) &= \tfrac{1}{m^2}\tfrac{\{n(\bar{x}-\mu)\}^2}{(n-1)(n-2)} + \tfrac{1}{m^2}\tfrac{\{n(\bar{x}-\mu)\}^2}{(n-1)^2(n-2)} \\ &= \tfrac{n^3(\bar{x}-\mu)^2}{m^2(n-1)^2(n-2)}. \end{aligned}$$

Also

$$Var\left(\hat{\mu}_f|data\right) = E_\mu\left\{Var\left(\hat{\mu}_f|\mu, data\right)\right\} + Var_\mu\left\{E\left(\hat{\mu}_f|\mu, data\right)\right\}.$$

Since

$$E\left(\bar{x}-\mu\right)^2 = \tilde{K}M$$

where

$$M = \left(\frac{1}{n-3}\right)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-3} - \left(\frac{1}{\bar{x}}\right)^{n-3}\right\}$$

it follows that

$$E_\mu\left\{Var\left(\hat{\mu}_f|\mu, data\right)\right\} = \frac{n^3}{m^2\left(n-1\right)^2\left(n-2\right)}\tilde{K}M.$$

Further

$$Var_\mu\left\{E\left(\hat{\mu}_f|\mu, data\right)\right\} = \left(1-a\right)^2 Var\left(\mu|data\right)$$

and

$$\begin{aligned} Var\left(\mu|data\right) &= E_\mu\left\{\mu - \left(\bar{x} - \tilde{K}L\right)\right\}^2 \\ &= E_\mu\left\{\left(\bar{x}-\mu\right)^2 - 2\left(\bar{x}-\mu\right)\tilde{K}L + \tilde{K}^2L^2\right\} \\ &= \tilde{K}M - 2\tilde{K}L\tilde{K}L + \tilde{K}^2L^2 = \tilde{K}M - \tilde{K}^2L^2. \end{aligned}$$

Therefore

$$Var_{\mu}\left\{E\left(\hat{\mu}_f|\mu, data\right)\right\} = (1-a)^2\left\{\tilde{K}M - \tilde{K}^2L^2\right\}$$

and

$$Var\left(\hat{\mu}_f|data\right) = \left\{\frac{n^3}{m^2\left(n-1\right)^2\left(n-2\right)} + (1-a)^2\right\}\tilde{K}M - (1-a)^2\tilde{K}^2L^2.$$

Proof of Theorem 6.8

From Equation (6.1) it follows that

$$\hat{\theta}_f|\theta \sim \frac{\chi^2_{2m-2}}{2m}\theta \qquad \hat{\theta}_f > 0.$$

Therefore

$$f\left(\hat{\theta}_f|\theta\right) = \left(\frac{m}{\theta}\right)^{m-1}\frac{\left(\hat{\theta}_f\right)^{m-2}\exp\left(-\frac{m}{\theta}\hat{\theta}_f\right)}{\Gamma\left(m-1\right)} \qquad \hat{\theta}_f > 0.$$

The posterior distribution of θ is

$$p\left(\theta|data\right) = K_1\left(\frac{1}{\theta}\right)^n\left\{\exp\left(-\frac{n}{\theta}\hat{\theta}\right) - \exp\left(-\frac{n}{\theta}\bar{x}\right)\right\}$$

where

$$K_1 = \frac{n^{n-1}}{\Gamma\left(n-1\right)}\left\{\left(\frac{1}{\theta}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}$$

(see Theorem 6.3).

The unconditional predictive density function of $\hat{\theta}_f$ is therefore given by

$$
\begin{aligned}
f\left(\hat{\theta}_f|data\right) &= \int_0^\infty f\left(\hat{\theta}_f|\theta\right) p\left(\theta|data\right) d\theta \\
&= \frac{m^{m-1}}{\Gamma(m-1)}\hat{\theta}_f K_1 \int_0^\infty \left(\frac{1}{\theta}\right)^{m+n-1} \left\{\exp\left[-\frac{1}{\theta}\left(m\hat{\theta}_f + n\hat{\theta}\right)\right] - \exp\left[-\frac{1}{\theta}\left(m\hat{\theta}_f + n\bar{x}\right)\right]\right\} d\theta \\
&= m^{m-1}n^{n-1}\frac{\Gamma(m+n-2)}{\Gamma(m-1)\Gamma(n-1)} \left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1} \left(\hat{\theta}_f\right)^{m-2} \\
&\quad \times \left\{\left(\frac{1}{m\hat{\theta}_f + n\hat{\theta}}\right)^{m+n-2} - \left(\frac{1}{m\hat{\theta}_f + n\bar{x}}\right)^{m+n-2}\right\} \qquad \hat{\theta}_f > 0.
\end{aligned}
$$

Proof of Theorem 6.9

Expected Value of $\hat{\theta}_f$

From Equation (6.1) it follows that

$$\hat{\theta}_f|\mu,\theta \sim \frac{\chi^2_{2m-2}}{2m}\theta.$$

Therefore

$$E\left(\hat{\theta}_f|\mu,\theta\right) = \frac{(m-1)}{m}\theta$$

and

$$Var\left(\hat{\theta}_f|\mu,\theta\right) = \frac{(m-1)}{m^2}\theta^2.$$

By using $p\left(\theta|\mu,data\right)$ (given in Equation [6.5]) it follows that

$$E\left(\theta|\mu,data\right) = \frac{n\left(\bar{x}-\mu\right)}{(n-1)}$$

and therefore

$$E\left(\hat{\theta}_f|\mu,data\right) = \frac{(m-1)}{m}\frac{n\left(\bar{x}-\mu\right)}{(n-1)}.$$

Since

$$p\left(\mu|data\right) = \tilde{K}\left(\bar{x}-\mu\right)^{-n} \qquad 0 < \mu < x_{(1)}$$

it follows that

$$E\left(\bar{x}-\mu\right)=\tilde{K}L$$

and

$$E\left(\hat{\theta}_f|\mu,data\right)=\frac{n\left(m-1\right)}{m\left(n-1\right)}\tilde{K}L.$$

Variance of $\hat{\theta}_f$

$$\begin{aligned} Var\left(\hat{\theta}_f|\mu,data\right) &= E_{\theta|\mu}\left\{Var\left(\hat{\theta}_f|\mu,\theta\right)\right\}+Var_{\theta|\mu}\left[E\left(\hat{\theta}_f|\mu,\theta\right)\right] \\ &= \tfrac{(m-1)}{m^2}\tfrac{\{n(\bar{x}-\mu)\}^2}{(n-1)(n-2)}+\tfrac{(m-1)^2}{m^2}\tfrac{\{n(\bar{x}-\mu)\}^2}{(n-1)^2(n-2)} \\ &= \tfrac{(m-1)n^2}{m^2(n-1)^2(n-2)}\left(m+n-2\right)\left(\bar{x}-\mu\right)^2. \end{aligned}$$

Further

$$Var\left(\hat{\theta}_f|data\right)=E_\mu\left\{Var\left(\hat{\theta}_f|\mu,data\right)\right\}+Var_\mu\left\{E\left(\hat{\theta}_f|\mu,data\right)\right\}$$

and

$$E_\mu\left\{Var\left(\hat{\theta}_f|data\right)\right\}=\frac{\left(m-1\right)n^2}{m^2\left(n-1\right)^2\left(n-2\right)}\left(m+n-2\right)\tilde{K}M.$$

Also

$$Var_\mu\left\{E\left(\hat{\theta}_f|\mu,data\right)\right\}=\frac{\left(m-1\right)^2}{m^2}\frac{n^2}{\left(n-1\right)^2}Var\left(\mu|data\right)$$

and therefore

$$\begin{aligned} Var_{\mu}\left\{E\left(\hat{\theta}_f|\mu, data\right)\right\} &= \tfrac{(m-1)^2}{m^2}\tfrac{n^2}{(n-1)^2}E\left\{\mu-\left(\bar{x}-\tilde{K}L\right)\right\}^2 \\ &= \tfrac{(m-1)^2n^2}{m^2(n-1)^2}E\left\{(\bar{x}-\mu)^2-2\tilde{K}L\left(\bar{x}-\mu\right)+\tilde{K}^2L^2\right\} \\ &= \tfrac{(m-1)^2n^2}{m^2(n-1)^2}\left\{\tilde{K}M-2\tilde{K}L\tilde{K}L+\tilde{K}^2L^2\right\} \\ &= \tfrac{(m-1)^2n^2}{m^2(n-1)^2}\left\{\tilde{K}M-\tilde{K}^2L^2\right\}. \end{aligned}$$

From this it follows that

$$\begin{aligned} Var\left(\hat{\theta}_f|data\right) &= \tfrac{(m-1)n^2}{m^2(n-1)^2(n-2)}\left(m+n-2\right)\tilde{K}M+\tfrac{(m-1)^2n^2}{m^2(n-1)^2}\left\{\tilde{K}M-\tilde{K}^2L^2\right\} \\ &= \tfrac{n^2(m-1)}{m(n-1)}\left\{\tfrac{\tilde{K}M}{(n-2)}-\tfrac{(m-1)}{m(n-1)}\tilde{K}^2L^2\right\}. \end{aligned}$$

Proof of Theorem 6.10

Proof of $E\left(U_f|data\right)$

From Equation (6.9) it follows that

$$f\left(U_f|\mu,\theta,\hat{\theta}_f\right)=\left(\frac{m}{\theta}\right)\exp\left\{-\frac{m}{\theta}\left[U_f-\left(\mu-\tilde{k}_2\hat{\theta}_f\right)\right]\right\} \qquad U_f>\mu-\tilde{k}_2\hat{\theta}_f$$

which means that

$$E\left(U_f|\mu,\theta,\hat{\theta}_f\right)=\frac{\theta}{m}+\mu-\tilde{k}_2\hat{\theta}_f.$$

Since

$$\hat{\theta}_f|\theta,\mu\sim\frac{\chi^2_{2m-2}}{2m}\theta$$

it follows that

$$E\left(U_f|\mu,\theta\right)=\mu+\frac{\theta}{m}\left\{1-\tilde{k}_2\left(m-1\right)\right\}=\mu+\frac{\theta}{m}H.$$

Also since the posterior distribution

$$p\left(\theta|\mu, data\right) = \frac{\left\{n\left(\bar{x}-\mu\right)\right\}^{n}}{\Gamma\left(n\right)}\left(\frac{1}{\theta}\right)^{n+1}\exp\left\{-\frac{n}{\theta}\left(\bar{x}-\mu\right)\right\} \qquad 0<\theta<\infty$$

it follows that

$$E\left(\theta|\mu, data\right) = \frac{n\left(\bar{x}-\mu\right)}{n-1}$$

and therefore

$$E\left(U_f|\mu, data\right) = a\bar{x}H + \mu\left(1-aH\right)$$

where

$$a = \frac{n}{m\left(n-1\right)}.$$

Further

$$p\left(\mu|data\right) = \left(n-1\right)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}\left(\bar{x}-\mu\right)^{-n} \qquad 0<\mu<x_{(1)}$$

and therefore

$$E\left(\mu|data\right) = \bar{x} - \tilde{K}L$$

which means that

$$E\left(U_f|data\right) = \bar{x} + \tilde{K}L\left(aH-1\right).$$

Proof of $Var\left(U_f|data\right)$

From Equation (6.9) it follows that

$$Var\left(U_f|\mu, \theta, \hat{\theta}_f\right) = \left(\frac{\theta}{m}\right)^{2}.$$

Now

$$\begin{aligned} Var\left(U_f|\mu,\theta\right) &= Var_{\hat{\theta}_f}\left\{E\left(U_f|\mu,\theta,\hat{\theta}_f\right)\right\} + E_{\hat{\theta}_f}\left\{Var\left(U_f|\mu,\theta,\hat{\theta}_f\right)\right\} \\ &= Var_{\hat{\theta}_f}\left(\tfrac{\theta}{m}+\mu-\tilde{k}_2\hat{\theta}_f\right) + E_{\hat{\theta}_f}\left\{\left(\tfrac{\theta}{m}\right)^2\right\} \\ &= \tilde{k}_2^2 Var\left(\tfrac{\chi^2_{2m-2}}{2m}\theta\right) + \left(\tfrac{\theta}{m}\right)^2 \\ &= \left(\tfrac{\theta}{m}\right)^2\left\{\tilde{k}_2^2\left(m-1\right)+1\right\} = \left(\tfrac{\theta}{m}\right)^2 J. \end{aligned}$$

Further

$$\begin{aligned} Var\left(U_f|\mu,data\right) &= E_{\theta|\mu}\left\{Var\left(U_f|\mu,\theta\right)\right\} + Var_{\theta|\mu}\left\{E\left(U_f|\mu,\theta\right)\right\} \\ &= E_{\theta|\mu}\left\{\left(\tfrac{\theta}{m}\right)^2 J\right\} + Var_{\theta|\mu}\left\{\mu+\tfrac{\theta}{m}H\right\} \\ &= \tfrac{J}{m^2}E\left(\theta^2|\mu,data\right) + \tfrac{H^2}{m^2}Var\left(\theta|\mu,data\right). \end{aligned}$$

Since

$$E\left(\theta^2|\mu,data\right) = \frac{\left\{n\left(\bar{x}-\mu\right)\right\}^2}{\left(n-1\right)\left(n-2\right)}$$

and

$$Var\left(\theta|\mu,data\right) = \frac{\left\{n\left(\bar{x}-\mu\right)\right\}^2}{\left(n-1\right)^2\left(n-2\right)}$$

it follows that

$$Var\left(U_f|\mu,data\right) = \frac{n^2\left(\bar{x}-\mu\right)^2}{m^2\left(n-1\right)\left(n-2\right)}\left\{J+\frac{H^2}{n-1}\right\}.$$

Finally

$$E_\mu\left\{Var\left(U_f|\mu,data\right)\right\} = \frac{n^2}{m^2\left(n-1\right)\left(n-2\right)}\left\{J+\frac{H^2}{n-1}\right\}E\left(\bar{x}-\mu\right)^2.$$

From $p(\mu|data)$ it follows that

$$E(\bar{x}-\mu)^2 = \tilde{K}M$$

and

$$E_\mu\{Var(U_f|\mu, data)\} = \frac{n^2}{m^2(n-1)(n-2)}\left\{J + \frac{H^2}{n-1}\right\}\tilde{K}M.$$

Also

$$\begin{aligned} Var_\mu\{E(U_f|\mu, data)\} &= Var\{a\bar{x}H + \mu(1-aH)\} \\ &= (1-aH)^2 Var(\mu|data) \\ &= (1-aH)^2\left(\tilde{K}M - \tilde{K}^2L^2\right). \end{aligned}$$

Therefore

$$Var(U_f|data) = \frac{n^2}{m^2(n-1)(n-2)}\left\{J + \frac{H^2}{n-1}\right\}\tilde{K}M + \left(1-aH^2\right)\left\{\tilde{K}M - \tilde{K}^2L^2\right\}.$$

7. Two-Parameter Exponential Distribution if the Location Parameter Can Take on Any Value Between Minus Infinity and Plus Infinity

7.1. Introduction

In this chapter the same notation will be used as in the previous chapter with the exception that the location parameter can now take on values between $-\infty$ and ∞, similarly as in some literature, see for example Johnson and Kotz (1970).

Therefore the two-parameter exponential distribution has the probability density function

$$f(x;\mu,\theta) = \frac{1}{\theta}\exp\left\{-\frac{(x-\mu)}{\theta}\right\} \qquad x > \mu,\ -\infty < \mu < \infty,\ \theta > 0$$

where μ is the location parameter and θ the scale parameter.

As before, let $X_1, X_2, \ldots, X_n$ be a sample of n observations from the two-parameter exponential distribution. The maximum likelihood estimators for μ and θ are given by

$$\hat{\mu} = X_{(1)}$$

and

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - X_{(1)}\right) = \bar{X} - X_{(1)}$$

where $X_{(1)}$ is the minimum or the first order statistic of the sample. It is well known (see Johnson and Kotz (1970); Lawless (1982); Krishnamoorthy and Mathew (2009)) that $\hat{\mu}$ and $\hat{\theta}$ are independently distributed with

$$\frac{(\hat{\mu}-\mu)}{\theta} \sim \frac{\chi^2_2}{2n} \text{ and } \frac{\hat{\theta}}{\theta} \sim \frac{\chi^2_{2n-2}}{2n}. \tag{7.1}$$

Let $\hat{\mu}_0$ and $\hat{\theta}_0$ be observed values of $\hat{\mu}$ and $\hat{\theta}$ then it follows from Equation (7.1) that a generalized pivotal quantity (GPQ) for μ is given by

$$G_\mu = \hat{\mu}_0 - \frac{\chi^2_2}{\chi^2_{2n-2}}\hat{\theta}_0 \tag{7.2}$$

and a GPQ for θ is given by

$$G_\theta = \frac{2n\hat{\theta}_0}{\chi^2_{2n-2}} \tag{7.3}$$

From a Bayesian perspective it will be shown that G_μ and G_θ are actually the posterior distributions of μ and θ if the prior $p(\mu,\theta) \propto \theta^{-1}$ is used.

7.2. Bayesian Procedure

In this section it will be shown that the Bayesian procedure is the same as the generalized variable approach.

If a sample of n observations are drawn from the two-parameter exponential distribution, then the likelihood function is given by

$$L(\mu,\theta|data) = \left(\frac{1}{\theta}\right)^n \exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)\right\}.$$

As prior the Jeffreys' prior

$$p(\mu, \theta) \propto \theta^{-1}$$

will be used.

The joint posterior distribution of μ and θ is

$$\begin{aligned} p(\theta, \mu|data) &\propto p(\mu, \theta) L(\mu, \theta|data) \\ &= K_1 \left(\tfrac{1}{\theta}\right)^{n+1} \exp\left\{-\tfrac{n}{\theta}(\bar{x} - \mu)\right\} \quad -\infty < \mu < x_{(1)},\ 0 < \theta < \infty \end{aligned} \tag{7.4}$$

It can easily be shown that

$$K_1 = \frac{n^n \left(\hat{\theta}\right)^{n-1}}{\Gamma(n-1)} \text{ where } \hat{\theta} = \bar{x} - x_{(1)}.$$

The posterior distribution of μ is

$$\begin{aligned} p(\mu|data) &= \int_0^\infty p(\theta, \mu|data)\, d\theta \\ &= (n-1)\left(\hat{\theta}\right)^{n-1} \left(\tfrac{1}{\bar{x}-\mu}\right)^n \quad -\infty < \mu < x_{(1)} \end{aligned} \tag{7.5}$$

and the posterior distribution of θ is

$$\begin{aligned} p(\theta|data) &= \int_{-\infty}^{x_{(1)}} p(\theta, \mu|data)\, d\mu \\ &= K_2 \left(\tfrac{1}{\theta}\right)^n \exp\left\{-\tfrac{n\hat{\theta}}{\theta}\right\} \text{ where } 0 < \theta < \infty \end{aligned} \tag{7.6}$$

an Inverse Gamma distribution where

$$K_2 = \frac{n^{n-1}\left(\hat{\theta}\right)^{n-1}}{\Gamma(n-1)}.$$

The conditional posterior distribution of θ given μ is given by

$$\begin{aligned} p(\theta|\mu, data) &= \tfrac{p(\theta,\mu|data)}{p(\mu|data)} \\ &= K_3\left(\tfrac{1}{\theta}\right)^{n+1}\exp\left\{-\tfrac{n}{\theta}(\bar{x}-\mu)\right\} \text{ where } 0<\theta<\infty \end{aligned} \tag{7.7}$$

an Inverse Gamma distribution where

$$K_3 = \frac{\{n(\bar{x}-\mu)\}^n}{\Gamma(n)}.$$

Also the conditional posterior distribution of μ given θ is

$$\begin{aligned} p(\mu|\theta, data) &= \tfrac{p(\theta,\mu|data)}{p(\theta|data)} \\ &= \tfrac{n}{\theta}\exp\left\{-\tfrac{n}{\theta}\left(x_{(1)}-\mu\right)\right\} \text{ where } -\infty<\mu<x_{(1)} \end{aligned} \tag{7.8}$$

The following theorem can now easily be proved:

Theorem 7.1. *The distribution of the generalized pivotal quantities G_μ and G_θ defined in Equations (7.2) and (7.3) are exactly the same as the posterior distributions $p(\mu|data)$ and $p(\theta|data)$ given in Equations (7.5) and (7.6).*

Proof. The proof is given in the Mathematical Appendices to this chapter. □

7.3. The Predictive Distributions of Future Sample Location and Scale Maximum Likelihood Estimators, $\hat{\mu_f}$ and $\hat{\theta_f}$

As before consider a future sample of m observations from the two-parameter exponential population: $X_{1f}, X_{2f}, \ldots, X_{mf}$. The future sample mean is defined as $\bar{X_f} = \frac{1}{m}\sum_{j=1}^{m} X_{jf}$. The smallest value in the sample is denoted by $\hat{\mu}_f$ and $\hat{\theta}_f = \bar{X}_f - \hat{\mu}_f$. To obtain control charts for $\hat{\mu}_f$ and $\hat{\theta}_f$ their predictive distributions must first be derived. If $-\infty < \mu < \infty$, then the descriptive statistics, posterior distributions and predictive distributions will be denoted by a tilde (~).

Theorem 7.2. *The predictive distribution of a future sample location maximum likelihood estimator, $\hat{\mu}_f$ is given by*

$$\tilde{f}\left(\hat{\mu}_f|data\right) = \begin{cases} \tilde{K}^* \left[\frac{1}{n\left(\bar{x}-\hat{\mu}_f\right)}\right]^n & -\infty < \hat{\mu}_f < x_{(1)} \\ \\ \tilde{K}^* \left[\frac{1}{n\hat{\theta}+m\left(\hat{\mu}_f - x_{(1)}\right)}\right]^n & x_{(1)} < \hat{\mu}_f < \infty \end{cases} \tag{7.9}$$

where

$$\tilde{K}^* = \frac{n^n\left(n-1\right)m}{n+m}\left(\hat{\theta}\right)^{n-1}.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

The reason why $\tilde{f}\left(\hat{\mu}_f|data\right)$ (Equation (7.9)) differs from $f\left(\hat{\mu}_f|data\right)$ (Equation (6.6), previous chapter) is that for Equation (6.6) it is assumed that $0 < \mu < \infty$ which results in a posterior distribution of $p\left(\mu|data\right) = \left(n-1\right)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}\left(\bar{x}-\mu\right)^{-n}$, while for Equation (7.9) it is assumed that $-\infty < \mu < \infty$ and the posterior distribution for μ is therefore $\tilde{p}\left(\mu|data\right) \propto \left(n-1\right)\left(\hat{\theta}\right)^{n-1}\left(\bar{x}-\mu\right)^{-n}$.

Theorem 7.3. *The mean and variance of $\hat{\mu}_f$ is given by*

$$\tilde{E}\left(\hat{\mu}_f|data\right) = \bar{x} - \frac{mn-m-n}{m\left(n-2\right)}\hat{\theta} \tag{7.10}$$

and

$$\tilde{Var}\left(\hat{\mu}_f|data\right) = \frac{n^3\left(n-2\right)+\left(mn-m-n\right)^2}{m^2\left(n-1\right)\left(n-3\right)\left(n-2\right)^2}\left(\hat{\theta}\right)^2 \tag{7.11}$$

Proof. By deleting the term $\left(\frac{1}{\bar{x}}\right)$ in Equations (6.6) and (6.7) from the previous chapter Equation (7.10) and Equation (7.11) follows. □

Theorem 7.4. *The predictive distribution of a future sample scale maximum likelihood estimator, $\hat{\theta}_f$ is given by*

$$\tilde{f}\left(\hat{\theta}_f|data\right) = \frac{\Gamma\left(m+n-2\right)m^{m-1}\left(n\hat{\theta}\right)^{n-1}\left(\hat{\theta}_f\right)^{m-2}}{\Gamma\left(m-1\right)\Gamma\left(n-1\right)\left(m\hat{\theta}_f+n\hat{\theta}\right)^{m+n-2}} \quad 0<\hat{\theta}_f<\infty \tag{7.12}$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Corollary 7.5. $\hat{\theta}_f|data \sim \hat{\theta}\frac{n}{m}\frac{(m-1)}{(n-1)}F_{2m-2,2n-2}$.

Proof.

$$\hat{\theta}_f|\theta \sim \frac{\chi^2_{2m-2}}{2m}\theta$$

and

$$\theta|data \sim \hat{\theta}\frac{2n}{\chi^2_{2n-2}}.$$

Therefore

$$\hat{\theta}_f|data \sim \frac{\chi^2_{2m-2}}{2m}\frac{2n}{\chi^2_{2n-2}}\hat{\theta},$$

$$\frac{1}{\hat{\theta}}\frac{2m}{2n}\frac{2n-2}{2m-2}\hat{\theta}_f \sim F_{2m-2;2n-2}$$

and

$$\hat{\theta}_f \sim \hat{\theta}\frac{n}{m}\frac{(m-1)}{(n-1)}F_{2(m-1);2(n-1).}$$

□

Theorem 7.6. *The mean and variance of $\hat{\theta}_f$ is given by*

$$\tilde{E}\left(\hat{\theta}_f|data\right) = \frac{(m-1)}{m}\frac{n}{(n-2)}\hat{\theta} \tag{7.13}$$

and

$$\widetilde{var}\left(\hat{\theta}_f|data\right) = \frac{n^2(m-1)}{m^2(n-2)^2(n-3)}(n+m-3)\hat{\theta}^2 \tag{7.14}$$

Proof. By deleting the term $\left(\frac{1}{\bar{x}}\right)$ in Theorem 6.9 from the previous chapter, Equation 7.13 and Equation 7.14 follow. □

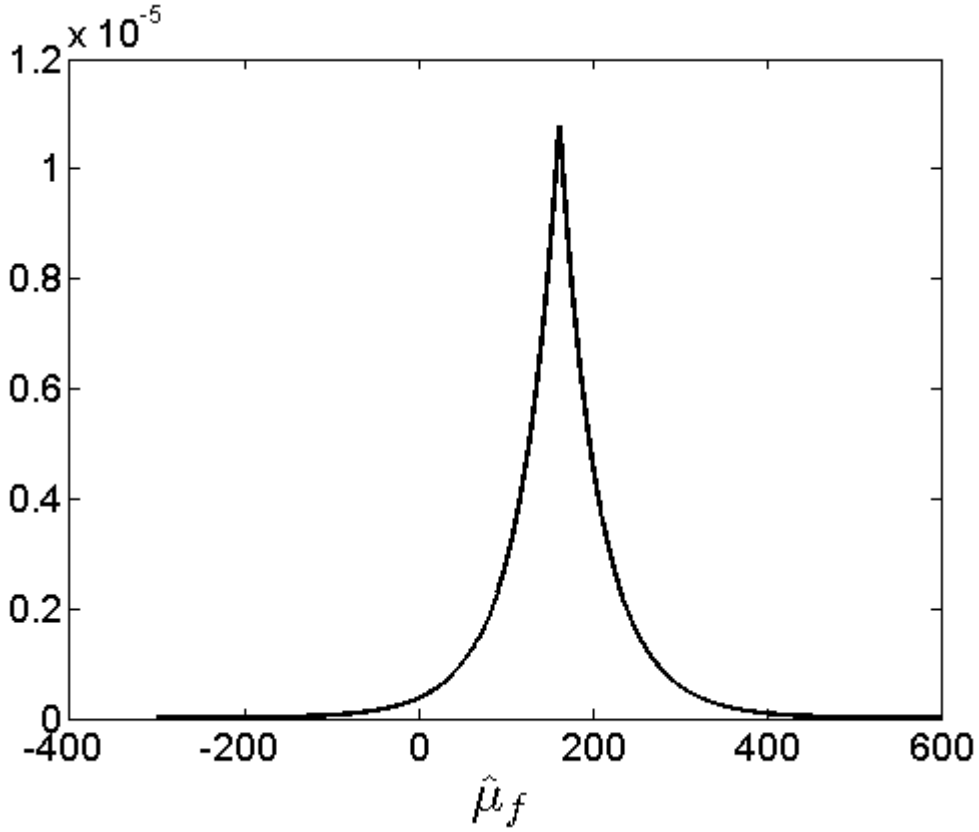
7.4. Example

The same example as in the previous chapter will be used for illustration purposes.

7.4.1. The Predictive Distribution of $\hat{\mu}_f$ $(-\infty < \mu < \infty)$

By using Equation (6.6) the predictive distribution $\tilde{f}(\hat{\mu}_f|data)$ for $m = 19$ future failure mileage data is illustrated in Figure 7.1.

Figure 7.1.: Distribution of $\hat{\mu}_f$, $n = 19$, $m = 19$



$\tilde{mean}(\hat{\mu}_f) = \tilde{median}(\hat{\mu}_f) = \tilde{mode}(\hat{\mu}_f) = 162$, $\tilde{var}(\hat{\mu}_f) = 5129.2$
$95\%\ \tilde{\text{interval}}(\hat{\mu}_f) = (11.2; 312.8)$
$99.73\%\ \tilde{\text{interval}}(\hat{\mu}_f) = (-154.85; 477.36)$
$96.08\%\ \tilde{\text{interval}}(\hat{\mu}_f) = (-2; 326)$

For $-\infty < \mu < \infty$, $n = 19$, $m = 19$, the predictive distribution $\tilde{f}(\hat{\mu}_f|data)$ is symmetrical. $\tilde{mean}(\hat{\mu}_f) = \tilde{median}(\hat{\mu}_f) = \tilde{mode}(\hat{\mu}_f) = 162$. $\tilde{Mode}(\hat{\mu}_f)$ for $-\infty < \mu < \infty$ is exactly the same as that for $0 < \mu < \infty$. A further comparison of Figure 7.1 and Figure 6.1 (from previous chapter) shows that $\tilde{var}(\hat{\mu}_f) = 5129.2$ is somewhat larger than $var(\hat{\mu}_f) = 3888.7$, when $0 < \mu < \infty$. Also the predictive intervals are somewhat wider. A dissatisfactory aspect for $-\infty < \mu < \infty$ is that the $99.73\%\ \tilde{\text{interval}}(\hat{\mu}_f) = (-154.85; 477.36)$, i.e., contains negative values.

In Table 7.1 descriptive statistics are given for the run-length and expected run-length.

Table 7.1.: Descriptive Statistics for the Run-length and Expected Run-length in the Case of $\hat{\mu}_f$; $-\infty < \mu < \infty$ and $\beta = 0.039$ for $n = 19$ and $m = 19$

Descriptive Statistics	$f(r\|data)$ Equal Tail	Expected Run-length Equal Tail
$\tilde{mean}$	367.84	373.34
$\tilde{median}$	44.50	69.93
$\tilde{var}$	6.663×10^7	1.414×10^8
95% $\tilde{\text{interval}}$	$(0; 740.50)$	$(0; 724.64)$

A comparison of Table 7.1 with Figure 6.2 and Figure 6.3 from the previous chapter shows that if $-\infty < \mu < \infty$ the expected (mean) run-length ≈ 370 if $\beta = 0.0392$ while this is the case if $\beta = 0.0258$ for $0 < \mu < \infty$. Also the 95% intervals are somewhat shorter for $-\infty < \mu < \infty$.

7.4.2. A Comparison of the Predictive Distributions for $\hat{\theta}_f$

In Figure 7.2 comparisons are made between

$$\tilde{f}\left(\hat{\theta}_f|data\right) = \frac{\Gamma(m+n-2)\, m^{m-1}\left(n\hat{\theta}\right)^{n-1}\left(\hat{\theta}_f\right)^{m-2}}{\Gamma(m-1)\,\Gamma(n-1)\left(m\hat{\theta}_f + n\hat{\theta}\right)^{m+n-2}} \quad 0 < \hat{\theta}_f < \infty$$

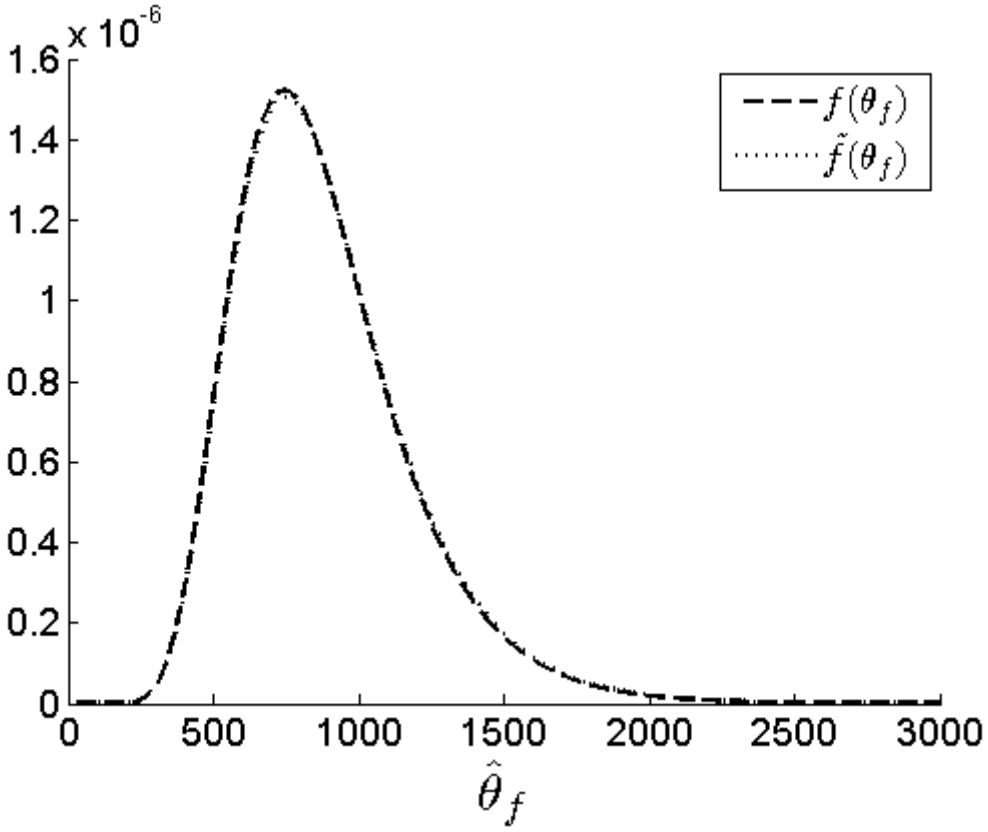
and

$$f\left(\hat{\theta}_f|data\right) = m^{m-1}n^{n-1}\frac{\Gamma(m+n-2)}{\Gamma(m-1)\,\Gamma(n-1)}\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}\left(\hat{\theta}_f\right)^{m-2}$$

$$\times\left\{\left(\frac{1}{m\hat{\theta}_f + n\hat{\theta}}\right)^{m+n-2} - \left(\frac{1}{m\hat{\theta}_f + n\bar{x}}\right)^{m+n-2}\right\} \quad 0 < \hat{\theta}_f < \infty.$$

As mentioned $\tilde{f}\left(\hat{\theta}_f|data\right)$ denotes the predictive density function if $-\infty < \mu < \infty$ and $f\left(\hat{\theta}_f|data\right)$ denotes the predictive density function if $0 < \mu < \infty$.

Figure 7.2.: $f\left(\hat{\theta}_f|data\right)$, $n = 19$, $m = 19$



In Table 7.2 descriptive statistics are given for $\tilde{f}\left(\hat{\theta}_f|data\right)$ and $f\left(\hat{\theta}_f\right)$.

Table 7.2.: Descriptive Statistics of $\tilde{f}\left(\hat{\theta}_f|data\right)$ and $f\left(\hat{\theta}_f|data\right)$

Descriptive Statistics	$\tilde{f}\left(\hat{\theta}_f\|data\right)$	$f\left(\hat{\theta}_f\|data\right)$
$Mean\left(\hat{\theta}_f\right)$	884.34	876.98
$Median\left(\hat{\theta}_f\right)$	835.2	829.1
$Mode\left(\hat{\theta}_f\right)$	747.3	743.3
$Var\left(\hat{\theta}_f\right)$	95037	91991
95% $Equal-tail\ Interval\left(\hat{\theta}_f\right)$	(430; 1622) Length=1192	(428; 1601) Length=1173
95% $HPD\ Interval\left(\hat{\theta}_f\right)$	(370.8; 1500.5) Length=1129.7	(369.2; 1482.6) Length=1113.4

Also, the exact means and variances are given by:

$$\tilde{E}\left(\hat{\theta}_f|data\right) = \frac{n\,(m-1)}{m\,(n-2)}\hat{\theta} = 884.34,$$

$$\tilde{Var}\left(\hat{\theta}_f|data\right) = \frac{n^2\,(m-1)}{m^2\,(n-2)^2\,(n-3)}\,(n+m-3)\,\hat{\theta}^2 = 95042,$$

$$E\left(\hat{\theta}_f|data\right) = \frac{n\,(m-1)}{m\,(n-2)}\tilde{K}L = 976.98,$$

$$Var\left(\hat{\theta}_f|data\right) = \frac{n^2\,(m-1)}{m\,(n-1)}\left\{\frac{\tilde{K}M}{n-2} - \frac{m-1}{m\,(n-1)}\tilde{K}^2L^2\right\} = 91991$$

where

$$\tilde{K} = (n-1)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1},$$

$$L = \frac{1}{n-2}\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-2} - \left(\frac{1}{\bar{x}}\right)^{n-2}\right\}$$

and

$$M = \frac{1}{n-3}\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-3} - \left(\frac{1}{\bar{x}}\right)^{n-3}\right\}.$$

From Figure 7.2 and Table 7.2 it can be seen that $\tilde{f}\left(\hat{\theta}_f|data\right)$ and $f\left(\hat{\theta}_f|data\right)$ are for all practical purposes the same. Also the exact means and variances coincide with the the numerical values. It therefore seems that whether the assumption is

that $-\infty < \mu < \infty$ or that $0 < \mu < \infty$ does not play a big role in the prediction of $\hat{\theta}_f$.

In Table 7.3 comparisons are made between the run-lengths and expected run-lengths in the case of $\hat{\theta}_f$ for $-\infty < \mu < \infty$ and $0 < \mu < \infty$.

Table 7.3.: Descriptive Statistics for the Run-lengths and Expected Run-lengths in the case of $\hat{\theta}_f$, $-\infty < \mu < \infty$, $0 < \mu < \infty$ and $\beta = 0.018$

Descriptive Statistics	$-\infty < \mu < \infty$		$0 < \mu < \infty$	
	$f(r\|data)$ Equal Tail	Expected Run-Length Equal Tail	$f(r\|data)$ Equal Tail	Expected Run-Length Equal Tail
Mean	369.29	370.42	375.25	375.25
Median	122.8	229.28	127.4	238.38
Variance	3.884×10^5	1.2572×10^5	3.9515×10^5	1.2698×10^5
95% Interval	$(0; 1586)$	$(7.423; 1089.7)$	$(0; 1602.5)$	$(7.809; 1096.7)$

From Table 7.3 it is clear that the corresponding statistics for $-\infty < \mu < \infty$ and $0 < \mu < \infty$ are for all practical purposes the same. So with respect to $\hat{\theta}_f$ it does not really matter whether it is assumed that μ is positive or not.

7.5. Phase I Control Chart for the Scale Parameter in the Case of the Two-parameter Exponential Distribution

As mentioned before, statistical quality control is implemented in two phases. In Phase I the primary interest is to assess process stability. Phase I is the so-called retrospective phase and Phase II the prospective or monitoring phase. The construction of Phase I control charts should be considered as a multiple testing problem. The distribution of a set of dependent variables (ratios of chi-square random variables) will therefore be used to calculate the control limits so that the false alarm probability (FAP) is not larger than $FAP_0 = 0.05$. To obtain control limits in Phase I, more than one sample is needed. Therefore in the example that follows there will be $m^* = 5$ samples or subgroups each of size $n = 10$.

Example

The data in Table 7.4 are simulated data obtained from the following two-parameter exponential distribution:

$$f(x_{ij};\theta,\mu) = \frac{1}{\theta}\exp\left\{-\frac{x_{ij}-\mu_i}{\theta}\right\},\ i = 1,2,\ldots,m^*,\ j = 1,2,\ldots,n;\ x_{ij} > \mu_i$$

$$\theta = 8;\ \mu_i = 2i;\ m^* = 5;\ n = 10.$$

Table 7.4.: Simulated Data for the Two-parameter Exponential Distribution

μ_i				
2	**4**	**6**	**8**	**10**
3.6393	18.7809	9.3759	10.7846	16.5907
2.7916	4.2388	32.6582	35.5781	17.7079
18.5094	4.3502	7.3084	18.2721	12.1376
2.7249	9.7827	6.5463	32.6032	11.8333
5.6664	5.7823	9.1002	26.6535	23.4186
20.6199	19.6218	8.2193	9.5539	15.7106
12.2267	10.9065	8.3750	10.9127	16.4669
6.8282	4.7042	13.4873	17.1883	13.4918
2.3474	5.8635	9.3791	8.4085	12.7471
2.2859	4.3308	20.1200	34.9469	12.2616

From the simulated data in Table 7.4 we have

$$\hat{\theta}_{i\cdot} = \bar{X}_{i\cdot} - X_{(1,i)} = \left[\begin{array}{ccccc} 5.4780 & 4.5974 & 5.9107 & 12.0817 & 3.4023 \end{array}\right],$$

$$\sum_{i=1}^{m^*} \hat{\theta}_{i\cdot} = 31.4701$$

and

$$\hat{\theta} = \frac{1}{m^*} \sum_{i=1}^{m^*} \hat{\theta}_{i\cdot} = 6.2940.$$

It is well known that $\hat{\theta}_i \sim \frac{\theta}{2n}\chi^2_{2n-2} = \frac{\theta}{2n} Y_i$ and therefore $\sum_{i=1}^{m^*} \hat{\theta}_i \sim \frac{\theta}{2n} \sum_{i=1}^{m^*} Y_i$.
Let

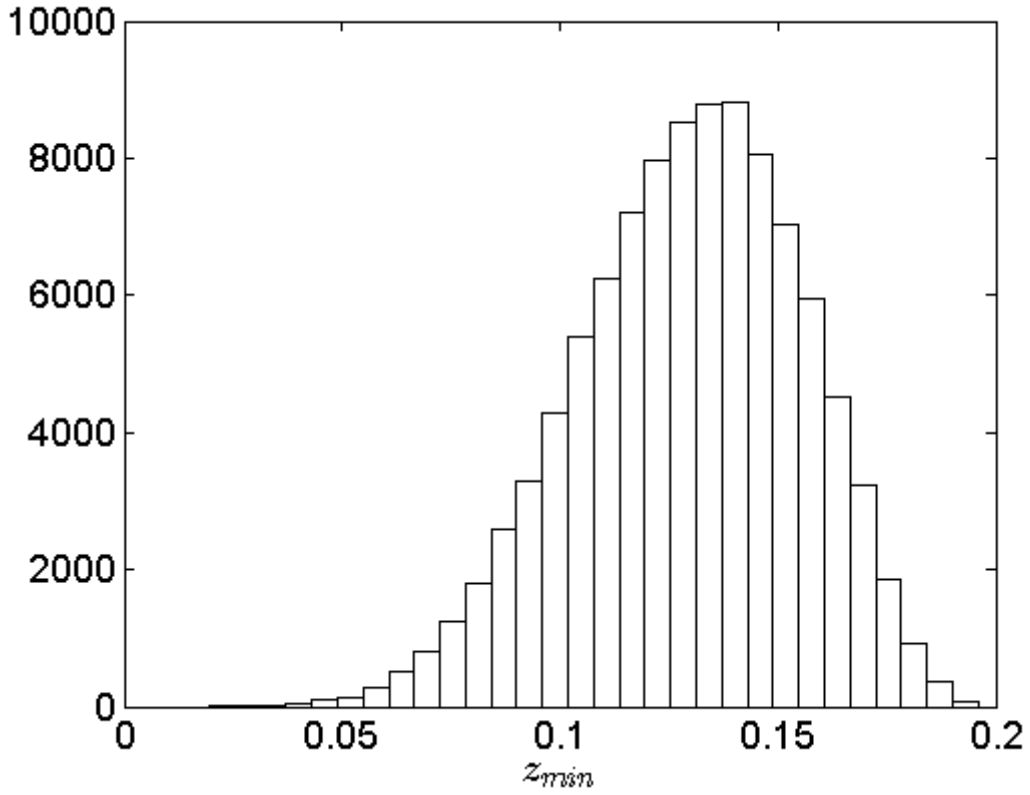
$$Z_1 = \frac{\hat{\theta}_i}{\sum_{i=1}^{m^*} \hat{\theta}_i} = \frac{\frac{\theta}{2n} Y_i}{\frac{\theta}{2n} \sum_{i=1}^{m^*} Y_i} = \frac{Y_i}{\sum_{i=1}^{m^*} Y_i} \quad i = 1, 2, \ldots, m^*$$

where

$$Y_i \sim \chi^2_{2n-2}.$$

To obtain a lower control limit for the data in Table 7.4, the distribution of $Z_{min} = min\,(Z_1, Z_2, \ldots, Z_{m^*})$ must be obtained.

Figure 7.3.: Distribution of $Z_{min} = min\,(Z_1, Z_2, \ldots, Z_{m^*})$, 100 000 simulations



The distribution of Z_{min} obtained from 100,000 simulations is illustrated in Figure 7.3. The value $Z_{0.05} = 0.0844$ is calculated such that the FAP is at a level of 0.05. The lower control limit is then determined as

$$LCL = Z_{0.05} \sum_{i=1}^{m} \hat{\theta}_i = (0.0844)(31.4701) = 2.656.$$

Since $\hat{\theta}_i > 2.656$ $(i = 1, 2, \ldots, m^*)$ it can be concluded that the scale parameter is under statistical control.

7.6. Lower Control Limit for the Scale Parameter in Phase II

In the first part of this section, the lower control limit in a Phase II setting will be derived using the Bayesian predictive distribution.

The following theorems can easily be proved:

Theorem 7.7. *For the two-parameter exponential distribution*

$$f\left(x_{ij}; \theta, \mu_i\right) = \left(\frac{1}{\theta}\right) \exp\left\{-\frac{1}{\theta}\left(x_{ij} - \mu_i\right)\right\} \quad i = 1, 2, \ldots, m^*,\ j = 1, 2, \ldots, n,\ \textit{and}\ x_{ij} > \mu_i$$

the posterior distribution of the parameter θ given the data is given by

$$p\left(\theta | data\right) = \frac{\left(n\hat{\theta}\right)^{m^*(n-1)}}{\Gamma\left[m^*\left(n-1\right)\right]} \left(\frac{1}{\theta}\right)^{m^*(n-1)+1} \exp\left(-\frac{n\hat{\theta}}{\theta}\right) \quad \theta > 0$$

an Inverse Gamma Distribution.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Theorem 7.8. *Let $\hat{\theta}_f$ be the maximum likelihood estimator of the scale parameter in a future sample of n observations, then the predictive distribution of $\hat{\theta}_f$ is*

$$f\left(\hat{\theta}_f | data\right) = \frac{\Gamma\left[m^*\left(n-1\right)+n-1\right]}{\Gamma\left(n-1\right)\Gamma\left[m^*\left(n-1\right)\right]} \frac{\left(\hat{\theta}_f\right)^{n-2}}{\left(\hat{\theta}_f + \hat{\theta}\right)^{m^*(n-1)+n-1}} \quad \hat{\theta}_f > 0$$

which means that

$$\hat{\theta}_f|data \sim \frac{\hat{\theta}}{m^*}F_{2(n-1);2m^*(n-1)}$$

where

$$\hat{\theta} = \sum_{i=1}^{m^*} \hat{\theta}_i.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

At $\beta = 0.0027$ the lower control limit is obtained as $\frac{\hat{\theta}}{m^*}F_{2(n-1);2m^*(n-1)}\left(0.0027\right) = \frac{31.4701}{5}\left(0.29945\right) = 1.8847$ for $m^* = 5$ and $n = 10$.

Assuming that the process remains stable, the predictive distribution for $\hat{\theta}_f$ can also be used to derive the distribution of the run-length, that is the number of samples until the control chart signals for the first time.

The resulting region of size β using the predictive distribution for the determination of the run-length is defined as

$$\beta = \int_{R(\beta)} f\left(\hat{\theta}_f|data\right) d\hat{\theta}_f$$

where

$$R\left(\beta\right) = \left(0; 1.8847\right)$$

is the lower one-sided control interval.

Given θ and a stable process, the distribution of the run-length r is Geometric with parameter

$$\psi\left(\theta\right) = \int_{R(\beta)} f\left(\hat{\theta}_f|\theta\right) d\hat{\theta}_f$$

where $f\left(\hat{\theta}|\theta\right)$ is the distribution of a future sample scale parameter estimator given θ.

The value of the parameter θ is however unknown and its uncertainty is described by the posterior distribution $p(\theta|data)$.

The following theorem can also be proved.

Theorem 7.9. *For given θ the parameter of the Geometric distribution is*

$$\psi(\theta) = \psi\left(\chi^2_{2m^*(n-1)}\right)$$

for given $\chi^2_{2m^*(n-1)}$ which means that the parameter is only dependent on $\chi^2_{2m^*(n-1)}$ and not on θ.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

As mentioned, by simulating θ from $p(\theta|data)$ the probability density function of $f\left(\hat{\theta}_f|\theta\right)$ as well as the parameter $\psi(\theta)$ can be obtained. This must be done for each future sample. Therefore, by simulating a large number of θ values from the posterior distribution a large number of $\psi(\theta)$ values can be obtained. A large number of Geometric and run-length distributions with different parameter values $(\psi(\theta_1), \psi(\theta_2), \ldots, \psi(\theta_l))$ will therefore be available. The unconditional run-length distribution is obtained by using the Rao-Blackwell method, i.e., the average of the conditional run-length distributions.

In Table 7.5 results for the run-length at $\beta = 0.0027$ for $n = 10$ and different values for m^* are presented for the lower control limit of the scale parameter estimator.

Table 7.5.: Two Parameter Exponential Run Lengths Results

n	m^*	mean (ARL)	median (ARL)	Mean (PDF)	Median (PDF)	One-sided		Two-sided	
						Low	High	Low	High
10	5	1040.878616	561.9041434	1016.107825	365	120.668063	3285.604745	90.53140623	4875.25148
10	6	859.2946037	528.043143	850.6496099	340	132.1636771	2633.779211	103.814474	3585.574629
10	7	752.7840665	502.7053642	747.964747	330	137.1365507	2163.362738	107.5044166	2913.68755
10	8	692.3117213	478.8131128	688.5627995	320	147.7899272	1904.506951	121.7607631	2547.589774
10	9	644.1233737	467.3792855	640.6357323	310	149.1918895	1707.552491	119.5874038	2198.554651
10	10	616.4008914	456.271994	613.0915372	305	154.9653041	1581.526121	128.5788067	2094.950512
10	11	582.5441167	445.4805119	579.3980237	295	159.4756421	1466.351811	133.3858464	1874.789144
10	12	563.8471571	445.4805119	560.8517002	295	165.742153	1340.987775	139.7090749	1681.424117
10	13	542.293566	440.1999559	539.3968128	290	170.6404787	1264.446813	145.0341248	1534.216048
10	14	523.2584573	429.8629367	520.4260911	285	170.6404787	1193.031076	145.0341248	1466.351811
10	15	514.4424077	429.8629367	511.6810866	285	177.4499555	1142.597024	149.1918895	1381.311827
10	16	498.7108368	424.804032	496.0401877	285	179.2039115	1064.049573	150.6101839	1283.085571
10	17	493.4525335	419.816607	490.7979663	280	179.2039115	1064.049573	154.9653041	1264.446813
10	18	491.9807528	424.804032	489.3787352	285	188.2980341	1019.999561	164.1476967	1228.123456
10	19	479.4053832	414.8994965	476.8198625	280	184.5942927	978.1005367	159.4756421	1175.931545
10	20	470.1638992	410.0515567	467.6270667	275	188.2980341	951.3034035	164.1476967	1110.378304
10	50	408.6804318	386.8108266	406.4862335	265	235.6714776	654.3843849	214.7137467	726.2252763
10	100	389.0488816	377.9609703	386.9577637	260	267.625879	548.0561023	248.3896595	583.4710484
10	500	374.3715597	373.6276458	372.3573508	255	318.8260483	440.1999559	308.3784626	450.8374316
10	1000	372.9795051	373.6276458	370.9756458	255	333.4301277	414.8994965	326.0292806	424.804032
10	5000	371.9360097	373.6276458	369.9394557	255	352.8387861	391.3293518	348.8501358	395.9116345
10	10000	371.7307984	373.6276458	369.7352037	255	356.8826129	386.8108266	356.8826129	386.8108266

mean(ARL) and median(ARL) refer to results obtained from the expected run-length while mean(PDF) and median(PDF) refer to results obtained from the probability density function of the run-length

From Table 7.5 it can be seen that as the number of samples increase (larger m^*) the mean and median run-lengths converge to the expected run-length of 370.

Further define $\bar{\psi}(\theta) = \frac{1}{l}\sum_{i=1}^{l}\psi(\theta_i)$. From Menzefricke (2002) it is known that as $l \to \infty$, $\bar{\psi}(\theta) \to \beta = 0.0027$ and the harmonic mean of the unconditional run-length will be $\left(\frac{1}{\beta}\right) = \frac{1}{0.0027} = 370$. Therefore it does not matter how small m^* and n is, the harmonic mean of the run-length will always be $\frac{1}{\beta}$ if $l \to \infty$. In the case of the simulated example the mean run-length is 1040.88 and the median run-length 561.90. The reason for these large values is the uncertainty in the parameter estimate because of the small sample size and number of samples ($n = 10$ and $m^* = 5$). β however can easily be adjusted to get a mean run-length of 370.

7.7. A Comparison of the Predictive Distributions and Control Charts for a One-sided Upper Tolerance Limit, U_f

As mentioned in the previous chapter a future sample tolerance limit is defined as

$$U_f = \hat{\mu}_f - \tilde{k}_2\hat{\theta}_f \text{ where } \hat{\mu}_f > \mu \text{ and } \hat{\theta}_f > 0.$$

Also

$$f(\hat{\mu}_f|\mu,\theta) = \left(\frac{m}{\theta}\right)\exp\left\{-\frac{m}{\theta}(\hat{\mu}_f - \mu)\right\} \quad \hat{\mu}_f > \mu$$

which means that

$$f\left(U_f|\mu,\theta,\hat{\theta}_f\right) = \left(\frac{m}{\theta}\right)\exp\left\{-\frac{m}{\theta}\left[U_f - \left(\mu - \tilde{k}_2\hat{\theta}_f\right)\right]\right\} \qquad U_f > \mu - \tilde{k}_2\hat{\theta}_f \quad (7.15)$$

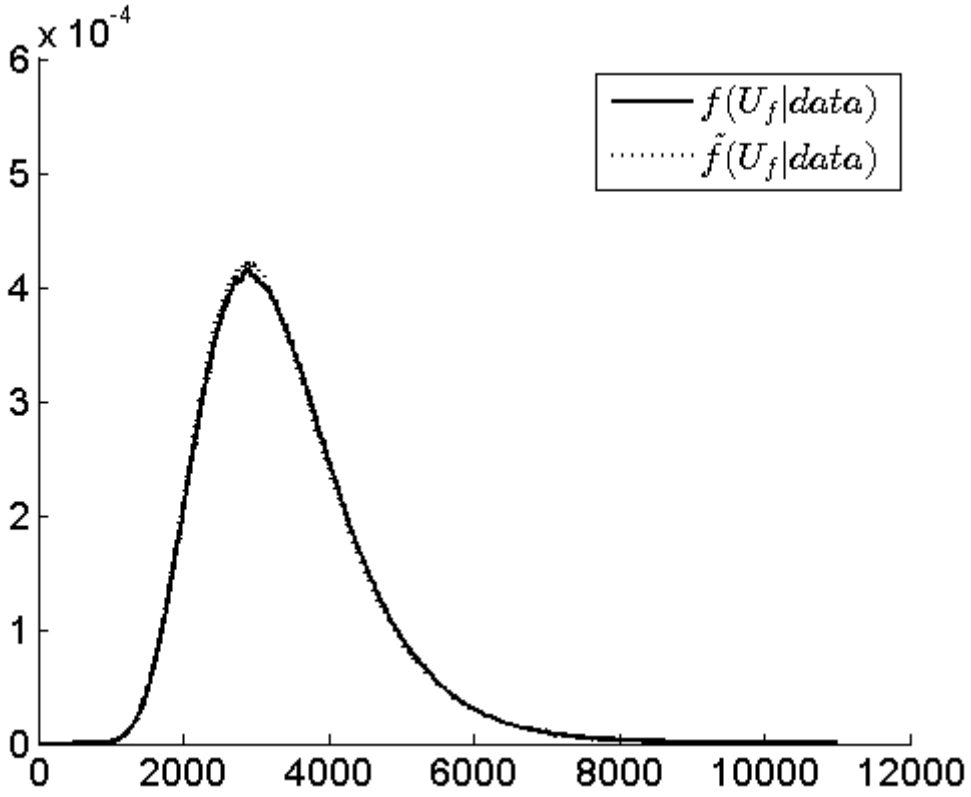
A comparison will be made between $f(U_f|data)$ and $\tilde{f}(U_f|data)$. The difference in the simulation procedure for these two density functions is that in the case of

$f(U_f|data)$ it is assumed that $0 < \mu < \infty$ which results in a posterior distribution of $p(\mu|data) = (n-1)\left\{\left(\frac{1}{\hat{\theta}}\right)^{n-1} - \left(\frac{1}{\bar{x}}\right)^{n-1}\right\}^{-1}(\bar{x}-\mu)^{-n}$ while for $\tilde{f}(U_f|data)$ it is assumed that $-\infty < \mu < \infty$ and the posterior distribution for μ is then $\tilde{p}(\mu|data) = (n-1)\left(\hat{\theta}\right)^{n-1}(\bar{x}-\mu)^{-n}$.

In the previous chapter a graph of the predictive density function $f(U_f|data)$ as well as descriptive statistics for the failure mileage data are given.

In the following figure comparisons are made between $f(U_f|data)$ and $\tilde{f}(U_f|data)$.

Figure 7.4.: Predictive Densities of $f(U_f|data)$ and $\tilde{f}(U_f|data)$



The descriptive statistics obtained from Figure 7.4 are presented in Table 7.6.

Table 7.6.: Descriptive Statistics of and $f\left(U_f|data\right)$ and $\tilde{f}\left(U_f|data\right)$

Descriptive Statistics	$f\left(U_f\|data\right)$	$\tilde{f}\left(U_f\|data\right)$
$Mean\left(U_f\right)$	3394.7	3406.6
$Median\left(U_f\right)$	3211.5	3225.8
$Mode(U_f)$	2900	2907
$Var\left(U_f\right)$	1.2317×10^6	1.2694×10^6
95% Equal-tail Interval	$(1736.5; 6027)$	$(1738; 6106)$
99.73% Equal-tail Interval	$(1249.05; 7973)$	$(1249.9; 8620)$

Also the exact mean and variance for $f\left(U_f|data\right)$ are

$$E\left(U_f|data\right) = \bar{x} + \tilde{K}L\left(aH - 1\right) = 3394.8$$

and

$$Var\left(U_f|data\right) = \frac{n^2}{m^2\left(n-1\right)\left(n-2\right)}\left\{J + \frac{H^2}{n-1}\right\}\tilde{K}M + \left(1-aH\right)^2\left\{\tilde{K}M - \tilde{K}^2L^2\right\} = 1.2439 \times 10^6$$

where

$$a = \frac{n}{m\left(n-1\right)},$$

$$H = 1 - \tilde{k}_2\left(m-1\right),$$

$$J = 1 + \tilde{k}_2^2\left(m-1\right)$$

and $\tilde{K}$, L and M defined as before.

The exact mean and variance for $\tilde{f}\left(U_f|data\right)$ are

$$\tilde{E}\left(U_f|data\right) = \bar{x} + \frac{n-1}{n-2}\left(aH-1\right)\hat{\theta} = 3415.0$$

and

$$\widetilde{Var}\left(U_f|data\right) = \frac{1}{(n-2)(n-3)}\left\{\frac{n^2}{m^2}\left(J + \frac{H^2}{n-1}\right) + \left(1-aH\right)^2\frac{(n-1)}{(n-2)}\right\}\left(\hat{\theta}\right)^2 = 1.2911\times10^6.$$

$\tilde{E}\left(U_f|data\right)$ and $\tilde{Var}\left(U_f|data\right)$ can be derived from $E\left(U_f|data\right)$ and $Var\left(U_f|data\right)$ by deleting the term $\left(\frac{1}{\bar{x}}\right)$ in $\tilde{K}$, L and M.

It seems that the predictive intervals for $\tilde{f}\left(U_f|data\right)$ are somewhat wider than in the case of $f\left(U_f|data\right)$.

In Table 7.7 comparisons are made between the run-lengths and expected run-lengths in the case of U_f for $-\infty < \mu < \infty$ and $0 < \mu < \infty$.

Table 7.7.: Descriptive Statistics for the Run-lengths and Expected Run-lengths in the Case of U_f; $-\infty < \mu < \infty$; $0 < \mu < \infty$and $\beta = 0.018$

Descriptive Statistics	$-\infty < \mu < \infty$		$0 < \mu < \infty$	
	$f\left(r|data\right)$ Equal Tail	Expected Run-Length Equal Tail	$f\left(r|data\right)$ Equal Tail	Expected Run-Length Equal Tail
Mean	444.95	444.95	418.68	419.68
Median	136.5	258.12	132.1	248.03
Variance	7.6243×10^5	2.8273×10^5	5.8236×10^5	2.0351×10^5
95% Interval	$(0; 1892.4)$		$(0; 1803.6)$	

It is clear from Table 10.1 that the corresponding statistics do not differ much. The mean, median, variance and 95% interval are however somewhat larger for $-\infty < \mu < \infty$.

7.8. Conclusion

This chapter develops a Bayesian control chart for monitoring the scale parameter, location parameter and upper tolerance limit of a two-parameter exponential distribution. In the Bayesian approach prior knowledge about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters. The information contained in the prior is combined with

the likelihood function to obtain the posterior distribution. By using the posterior distribution the predictive distributions of $\hat{\mu}_f$, $\hat{\theta}_f$ and U_f can be obtained.

The theory and results described in this paper have been applied to the failure mileages for military carriers analyzed by Grubbs (1971) and Krishnamoorthy and Mathew (2009). The example illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions and "run-lengths" for $\hat{\mu}_f$, $\hat{\theta}_f$ and U_f.

Results for $0 < \mu < \infty$ as presented in the previous chapter are compared to $-\infty < \mu < \infty$ in this chapter. It has been shown by changing the range of μ only the results of $\hat{\mu}_f$ are influenced. The results of $\hat{\theta}_f$ and U_f did not change due to a change in range of μ.

Mathematical Appendix to Chapter 7

Proof of Theorem 7.1

(a) The posterior distribution $p(\theta|data)$ is the same as the distribution of the pivotal quantity $G_\theta = \frac{2n\hat{\theta}}{\chi^2_{2n-2}}$.

Proof:

Let $Z \sim \chi^2_{2n-2}$

$\therefore f(z) = \frac{1}{2^{n-1}\Gamma(n-1)} z^{n-2} \exp\left\{-\frac{1}{2}z\right\}$

We are interested in the distribution of $\theta = \frac{2n\hat{\theta}}{Z}$.

Therefore $Z = \frac{2n\hat{\theta}}{\theta}$ and $\left|\frac{dZ}{d\theta}\right| = \frac{2n\hat{\theta}}{Z}$.

From this it follows that

$$\begin{aligned} f(G_\theta) &= f(\theta) = \frac{1}{2^{n-1}\Gamma(n-1)} \left(\frac{2n\hat{\theta}}{\theta}\right)^{n-2} \frac{2n\hat{\theta}}{\theta^2} \exp\left\{-\frac{n\hat{\theta}}{\theta}\right\} \\ &= \frac{n^{n-1}\left(\hat{\theta}\right)^{n-1}}{\Gamma(n-1)} \left(\frac{1}{\theta}\right)^n \exp\left\{-\frac{n\hat{\theta}}{\theta}\right\} = p(\theta|data) \end{aligned}$$

See Equation (6.3).

(b) The posterior distribution $p(\mu|data)$ is the same as the distribution of the pivotal quantity $G_\mu = \hat{\mu} - \frac{\chi_2^2}{\chi_{2n-2}^2}\hat{\theta}$.

Proof:

Let $F = \frac{\chi_2^2/2}{\chi_{2n-2}^2/(2n-2)} \sim F_{2,2n-2}$

$\therefore g(f) = \left(1 + \frac{1}{n-1}f\right)^{-n}$ where $0 < f < \infty$

We are interested in the distribution of $\mu = \hat{\mu} - \frac{2\hat{\theta}}{2n-2}F$ which means that $F = \frac{(n-1)}{\hat{\theta}}(\hat{\mu} - \mu)$ and $\left|\frac{dF}{d\mu}\right| = \frac{(n-1)}{\hat{\theta}}$.

Therefore

$$\begin{aligned} g(\mu) &= \left\{1 + \tfrac{1}{\hat{\theta}}(\hat{\mu} - \mu)\right\}^{-n} \tfrac{n-1}{\hat{\theta}} \\ &= (n-1)\hat{\theta}^{n-1}\left(\tfrac{1}{\bar{x}-\mu}\right)^n \text{ where } -\infty < \mu < \hat{\mu} \\ &= p(\mu|data) \end{aligned}$$

See Equation (7.5).

(c) The posterior distribution of $\mathrm{p}(\mu|\theta, data)$ is the same as the distribution of the pivotal quantity $G_{\mu|\theta} = \hat{\mu} - \frac{\chi_2^2}{2n}\theta$ (see Equation [7.1]).

Proof:

Let $\tilde{Z} \sim \chi_2^2$ then

$g(\tilde{z}) = \frac{1}{2}\exp\left\{-\frac{1}{2}\tilde{z}\right\}$.

Let $\mu = \hat{\mu} - \frac{\tilde{z}}{2n}\theta$, then $\tilde{z} = \frac{2n}{\theta}(\hat{\mu} - \mu)$ and $\left|\frac{d\tilde{z}}{d\mu}\right| = \frac{2n}{\theta}$.

Therefore

$$\begin{aligned} g(\mu|\theta) &= \tfrac{n}{\theta}\exp\left\{-\tfrac{n}{\theta}(\hat{\mu} - \mu)\right\} \qquad -\infty < \mu < \hat{\mu}_0 \\ &= p(\mu|\theta, data) \end{aligned}$$

See Equation (6.4).

Proof of Theorem 7.2

As before

$$f\left(\hat{\mu}_f|\mu,\theta\right)=\left(\frac{m}{\theta}\right)\exp\left\{-\frac{m}{\theta}\left(\hat{\mu}_f-\mu\right)\right\}\quad\hat{\mu}_f>\mu$$

and therefore

$$\tilde{f}\left(\hat{\mu}_f|\mu,data\right)=\int_0^\infty f\left(\hat{\mu}_f|\mu,\theta\right)\tilde{p}\left(\theta|\mu,data\right)d\theta.$$

Since

$$\tilde{p}\left(\theta|\mu,data\right)=\frac{\left\{n\left(\bar{x}-\mu\right)\right\}^n}{\Gamma\left(n\right)}\left(\frac{1}{\theta}\right)^{n+1}\exp\left\{-\frac{n}{\theta}\left(\bar{x}-\mu\right)\right\}$$

it follows that

$$\tilde{f}\left(\hat{\mu}_f|\mu,data\right)=\frac{n^{n+1}\left(\bar{x}-\mu\right)^n m}{\left[m\left(\hat{\mu}_f-\mu\right)+n\left(\bar{x}-\mu\right)\right]^{n+1}}\quad\hat{\mu}_f>\mu.$$

For $-\infty<\mu<x_{(1)}$,

$$\tilde{p}\left(\mu|data\right)=\left(n-1\right)\left(\hat{\theta}\right)^{n-1}\left(\frac{1}{\bar{x}-\mu}\right)^n$$

and

$$\begin{aligned}\tilde{f}\left(\hat{\mu}_f,\mu|data\right) &= \tilde{f}\left(\hat{\mu}_f|\mu,data\right)\tilde{p}\left(\mu|data\right)\\ &= \frac{n^{n+1}m(n-1)\left(\hat{\theta}\right)^{n-1}}{\left[m\left(\hat{\mu}_f-\mu\right)+n(\bar{x}-\mu)\right]^{n+1}}\end{aligned}.$$

Therefore

$$\begin{aligned} f\left(\hat{\mu}_f|data\right) &= \int_{-\infty}^{\hat{\mu}_f} f\left(\hat{\mu}_f, \mu|data\right) d\mu \quad -\infty < \hat{\mu}_f < x_{(1)} \\ &= \int_{-\infty}^{x_{(1)}} f\left(\hat{\mu}_f, \mu|data\right) d\mu \quad x_{(1)} < \hat{\mu}_f < \infty \\ &= \tilde{K}^* \left[\frac{1}{n\left(\bar{x}-\hat{\mu}_f\right)}\right]^n \quad -\infty < \hat{\mu}_f < x_{(1)} \\ &= \tilde{K}^* \left[\frac{1}{n\hat{\theta}+m\left(\hat{\mu}_f-x_{(1)}\right)}\right]^n \quad x_{(1)} < \hat{\mu}_f < \infty \end{aligned}$$

where

$$\tilde{K}^* = \frac{n^n\left(n-1\right)m}{\left(n+m\right)}\left(\hat{\theta}\right)^{n-1}.$$

Proof of Theorem 7.4

From Equation (6.1) it follows that

$$\hat{\theta}_f|\theta \sim \frac{\chi^2_{2m-2}}{2m}\theta$$

which means that

$$f\left(\hat{\theta}_f|\theta\right) = \left(\frac{m}{\theta}\right)^{m-1} \frac{\left(\hat{\theta}_f\right)^{m-2} \exp\left(-\frac{m}{\theta}\hat{\theta}_f\right)}{\Gamma\left(m-1\right)} \quad 0 < \hat{\theta}_f < \infty \tag{7.16}$$

The posterior distribution of θ (Equation [6.3]) is

$$\tilde{p}\left(\theta|data\right) = \frac{\left(n\hat{\theta}\right)^{n-1}}{\Gamma\left(n-1\right)}\left(\frac{1}{\theta}\right)^n \exp\left(-\frac{n}{\theta}\right) \quad 0 < \theta < \infty.$$

Therefore

$$\begin{aligned}
\tilde{f}\left(\hat{\theta}_f|data\right) &= \int_0^\infty f\left(\hat{\theta}_f|\theta\right)\tilde{p}\left(\theta|data\right)d\theta \\
&= \frac{m^{m-1}}{\Gamma(m-1)}\frac{\left(n\hat{\theta}\right)^{n-1}}{\Gamma(n-1)}\left(\hat{\theta}_f\right)^{m-2}\int_0^\infty\left(\frac{1}{\theta}\right)^{m+n-1}\exp\left\{-\frac{1}{\theta}\left(m\hat{\theta}_f+n\hat{\theta}\right)\right\}d\theta \\
&= \frac{\Gamma(m+n-2)m^{m-1}\left(n\hat{\theta}\right)^{n-1}\left(\hat{\theta}_f\right)^{m-2}}{\Gamma(m-1)\Gamma(n-1)\left(m\hat{\theta}_f+n\hat{\theta}\right)^{m+n-2}} \quad 0<\hat{\theta}_f<\infty.
\end{aligned}$$

Proof of Theorem 7.7

Let

$$\hat{\theta}=\sum_{i=1}^{m^*}\hat{\theta}_i.$$

As mentioned in Section 6 (see also Krishnamoorthy and Mathew (2009)) that it is well known that

$$\hat{\theta}_i\sim\frac{\theta}{2n}\chi^2_{2(n-1)}$$

which means that

$$\hat{\theta}\sim\frac{\theta}{2n}\chi^2_{2m^*(n-1)}.$$

Therefore

$$f\left(\hat{\theta}|\theta\right)=\left(\frac{n}{\theta}\right)^{m^*(n-1)}\frac{\left(\hat{\theta}\right)^{m^*(n-1)-1}\exp\left(-\frac{n\hat{\theta}}{\theta}\right)}{\Gamma\left[m^*\left(n-1\right)\right]}=L\left(\theta|\hat{\theta}\right)$$

i.e, the likelihood function.

As before we will use as prior $p\left(\theta\right)\propto\theta^{-1}$.

The posterior distribution

$$\begin{aligned} p\left(\theta|\hat{\theta}\right) &= p\left(\theta|data\right) \propto L\left(\theta|\hat{\theta}\right) p\left(\theta\right) \\ &= \frac{\left(n\hat{\theta}\right)^{m^*(n-1)}}{\Gamma[m^*(n-1)]} \left(\frac{1}{\theta}\right)^{m^*(n-1)+1} \exp\left(-\frac{n\hat{\theta}}{\theta}\right) \end{aligned}$$

An Inverse Gamma distribution.

Proof of Theorem 7.8

$$\hat{\theta}_f|\theta \sim \frac{\theta}{2n}\chi^2_{2(n-1)}.$$

Therefore

$$\begin{aligned} f\left(\hat{\theta}_f|data\right) &= \int_0^\infty f\left(\hat{\theta}_f|\theta\right) p\left(\theta|data\right) d\theta \\ &= \int_0^\infty \left(\frac{n}{\theta}\right)^{n-1} \frac{\left(\hat{\theta}_f\right)^{n-2} \exp\left(-\frac{n\hat{\theta}_f}{\theta}\right)}{\Gamma(n-1)} \times \\ &\quad \frac{\left(n\hat{\theta}\right)^{m^*(n-1)}}{\Gamma[m^*(n-1)]} \left(\frac{1}{\theta}\right)^{m^*(n-1)+1} \exp\left(-\frac{n\hat{\theta}}{\theta}\right) d\theta \\ &= \frac{(n)^{n-1}\left(\hat{\theta}_f\right)^{n-2}\left(n\hat{\theta}\right)^{m^*(n-1)}}{\Gamma(n-1)\Gamma[m^*(n-1)]} \int_0^\infty \left(\frac{1}{\theta}\right)^{m^*(n-1)+n} \exp\left\{-\frac{n}{\theta}\left[\hat{\theta}_f + \hat{\theta}\right]\right\} d\theta \\ &= \frac{\Gamma[m^*(n-1)+n-1]\left(\hat{\theta}\right)^{m^*(n-1)}}{\Gamma(n-1)\Gamma[m^*(n-1)]} \frac{\left(\hat{\theta}_f\right)^{n-2}}{\left(\hat{\theta}_f+\hat{\theta}\right)^{m^*(n-1)+n-1}} \quad \hat{\theta}_f > 0 \end{aligned}$$

From this it follows that

$$\hat{\theta}_f|data \sim \frac{\hat{\theta}}{m^*} F_{2(n-1);2m^*(n-1)}$$

where

$$\hat{\theta} = \sum_{i=1}^{m^*} \hat{\theta}_i.$$

Proof of Theorem 7.9

For given θ

$$\begin{aligned}
\psi(\theta) &= p\left(\hat{\theta}_f \leq \frac{\hat{\theta}}{m^*} F_{2(n-1);2m^*(n-1)}(\beta)\right) \\
&= p\left(\frac{\theta}{2n}\chi^2_{2(n-1)} \leq \frac{\hat{\theta}}{m^*} F_{2(n-1);2m^*(n-1)}(\beta)\right) \quad \text{given } \theta \\
&= p\left(\frac{2n\hat{\theta}}{\chi^2_{2m^*(n-1)}} \frac{\chi^2_{2(n-1)}}{2n} \leq \frac{\hat{\theta}}{m^*} F_{2(n-1);2m^*(n-1)}(\beta)\right) \quad \text{given } \chi^2_{2m^*(n-1)} \\
&= p\left(\chi^2_{2(n-1)} \leq \frac{\chi^2_{2m^*(n-1)}}{m^*} F_{2(n-1);2m^*(n-1)}(\beta)\right) \quad \text{given } \chi^2_{2m^*(n-1)} \\
&= \psi\left(\chi^2_{2m^*(n-1)}\right)
\end{aligned}$$

8. Piecewise Exponential Model

8.1. Introduction

The piecewise exponential model (PEXM) is one of the most popular and useful models in reliability and survival analysis.

The PEXM has been widely used to model time to event data in different contexts, such as in reliability engineering (Kim and Proschan (1991) and Gamerman (1994)), clinical situations such as kidney infections (Sahu, Dey, Aslanidu, and Sinha (1997), heart transplant data (Aitkin, Laird, and Francis (1983)), hospital mortality rate data (Clark and Ryan (2002)), economics (Bastos and Gamerman (2006)) and cancer studies including leukemia (Breslow (1974)). For further details see Demarqui, Loschi, Dey, and Colosimo (2012).

The PEXM assumes that times between failure are independent and exponentially distributed, but the mean is allowed to either increase or decrease with each failure. It can also be an appropriate model for repairable systems. According to Arab, Rigdon, and Basu (2012) there has been an increasing interest in developing Bayesian methods for repairable systems, due to the flexibility of these methods in accounting for parameter uncertainty (see for example Hulting and Robinson (1994), Pievatolo and Ruggeri (2004), Hamada, Wilson, Reese, and Martz (2008), Pan and Rigdon (2009) and Reese, Wilson, Guo, Hamada, and Johnson (2011)). In this chapter an objective Bayesian procedure will be applied for analyzing times between failures from multiple repairable systems.

8.2. The Piecewise Exponential Model

The model in its simplest form can be written as

$$f\left(x_j|\mu\delta\right) = \left(\frac{\delta}{\mu}j^{\delta-1}\right)^{-1} \exp\left\{-\frac{x_j}{\left(\frac{\delta}{\mu}j^{\delta-1}\right)}\right\} \quad x_j > 0.$$

The piecewise exponential model therefore assumes that the times between failures, $X_1, X_2, \ldots, X_J$ are independent exponential random variables with

$$E\left(X_j\right) = \frac{\delta}{\mu}j^{\delta-1}$$

where $\delta > 0$ and $\mu > 0$.

For example if $\delta = 0.71$ and $\mu = 0.0029$, then the expected time between the 9th and 10th failure is

$$E\left(X_{10}\right) = \frac{0.71}{0.0029}10^{0.71-1} = 125.56.$$

and the time between the 27th and 28th failure is

$$E\left(X_{28}\right) = \frac{0.71}{0.0029}28^{0.71-1} = 93.15.$$

In the PEXM model, μ is a scale parameter and δ is a shape parameter.

8.3. The Piecewise Exponential Model for Multiple Repairable Systems

Using the same notation as in Arab et al. (2012), let x_{ij} denotes the time between failures $(j-1)$ and j on system i for $j = 1, 2, \ldots, n_i$ and $i = 1, 2, \ldots, k$. The 0th failure occurs at time 0. Also let $N = \sum_{i=1}^{k} n_i$ denotes the total number of failures. Finally let $\underline{x}_i = \left[x_{i1,}x_{i2}, \ldots, x_{i,n_i}\right]'$ denote the times between failures for the ith system.

As in Arab et al. (2012) two cases for multiple systems will be considered.

8.4. Model 1: Identical Systems

Assume that all k systems are identical and that n_i failures on the ith system are observed. In other words $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ and $\delta_1 = \delta_2 = \cdots = \delta_k = \delta$. Since failures on separate systems are independent the likelihood function can be written as

$$\begin{aligned} L\left(\delta, \mu | \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k\right) &= L\left(\delta, \mu | data\right) \\ &= \prod_{i=1}^{k} \left\{ \prod_{j=1}^{n_i} \left(\frac{\delta}{\mu} j^{\delta-1}\right)^{-1} \exp\left[-\frac{x_{ij}}{\frac{\delta}{\mu} j^{\delta-1}}\right] \right\} \\ &= \left(\frac{\delta}{\mu}\right)^{-N} \left\{ \prod_{i=1}^{k} \prod_{j=1}^{n_i} j^{1-\delta} \right\} \exp\left\{ -\frac{\mu}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \right\}. \end{aligned}$$

For the Bayesian procedure a prior distribution that summarizes *a priori* uncertainty about the likely values of the parameters is needed. The prior distribution needs to be formulated based on prior knowledge. This is usually a difficult task because such prior knowledge may not be available. In such situations usually a "non-informative" prior distribution is used. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution of the parameters. The basic idea behind a "non-informative" prior is that it should be flat so that the likelihood plays a dominant role in the construction of the posterior distribution. If the form of the posterior distribution is complicated, numerical methods or Monte Carlo simulation procedures can be used to solve different complex problems. A "non-informative" prior may easily be obtained by applying Jeffreys' rule. Jeffreys' rule states that the prior distribution for a set of parameters is taken to be proportional to the square root of the determinant of the Fisher information matrix.

The following theorem can now be stated.

Theorem 8.1. *For the piecewise exponential model with identical systems, the Jeffreys' prior for the parameters μ and δ is given by $p_J(\mu, \delta) \propto \mu^{-1}$.*

Proof. The proof is given in the Mathematical Appendices to this chapter. □

8.5. The Joint Posterior Distribution - Identical Systems

$Posterior \propto Likelihood \times Prior$

Therefore

$$\begin{aligned} p(\mu,\delta|data) &\propto L(\mu,\delta|data)\, p_J(\mu,\delta) \\ &\propto \left(\tfrac{\delta}{\mu}\right)^{-N} \left\{\prod_{i=1}^{k}\prod_{j=1}^{n_i} j^{1-\delta}\right\} \exp\left\{-\tfrac{\mu}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right\} \mu^{-1}. \end{aligned}$$

From the joint posterior distribution the marginal posterior distribution can easily be obtained. Now

$$\begin{aligned} p(\delta|data) &= \int_0^\infty p(\mu,\delta|data)\, d\mu \\ &= \delta^{-N} \prod_{i=1}^{k}\prod_{j=1}^{n_i} j^{1-\delta} \left\{\int_0^\infty \mu^{N-1} \exp\left[-\tfrac{\mu}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right] d\mu\right\}. \end{aligned}$$

Since

$$\int_0^\infty \mu^{N-1} \exp\left\{-\frac{\mu}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right\} d\mu = \left(\frac{\delta}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}}\right)^{N} \Gamma(N)$$

it follows that

$$p(\delta|data) \propto \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right)^{-N} \prod_{i=1}^{k}\prod_{j=1}^{n_i} j^{1-\delta} \quad \delta > 0 \tag{8.1}$$

and

$$p(\mu|\delta,data) = \left(\frac{\delta}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}}\right)^{-N} \frac{1}{\Gamma(N)} \mu^{N-1} \exp\left\{-\frac{\mu}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right\} \quad \mu>0$$

(8.2)

- a Gamma density function.

Equation (8.2) follows from the fact that

$$p(\mu, \delta|data) = p(\delta|data)\, p(\mu|\delta, data).$$

8.6. Example (Arab et al. (2012))

Table 8.1.: Time Between Failures for Six Load-Haul-Dump (LHD) Machines

LHD1	LHD3	LHD9	LHD11	LHD17	LHD20
327	637	278	353	401	231
125	40	261	96	36	20
7	197	990	49	18	361
6	36	191	211	159	260
107	54	107	82	341	176
277	53	32	175	171	16
54	97	51	79	24	101
332	63	10	117	350	293
510	216	132	26	72	5
110	118	176	4	303	119
10	125	247	5	34	9
9	25	165	60	45	80
85	4	454	39	324	112
27	101	142	35	2	10
59	184	39	258	70	162
16	167	249	97	57	90
8	81	212	59	103	176
34	46	204	3	11	360
21	18	182	37	5	90
152	32	116	8	3	15
158	219	30	245	144	315
44	405	24	79	80	32
18	20	32	49	53	266
	248	38	31	84	
	140	10	259	218	
		311	283	122	
		61	150		
			24		

The data in Table 8.1 are failure data on load-haul-dump (LHD) machines given by Kumar and Klefsjö (1992) and reported in Hamada, Wilson, Reese, and Martz (2008, page 201).

In Figure 8.1 the posterior distribution of δ,

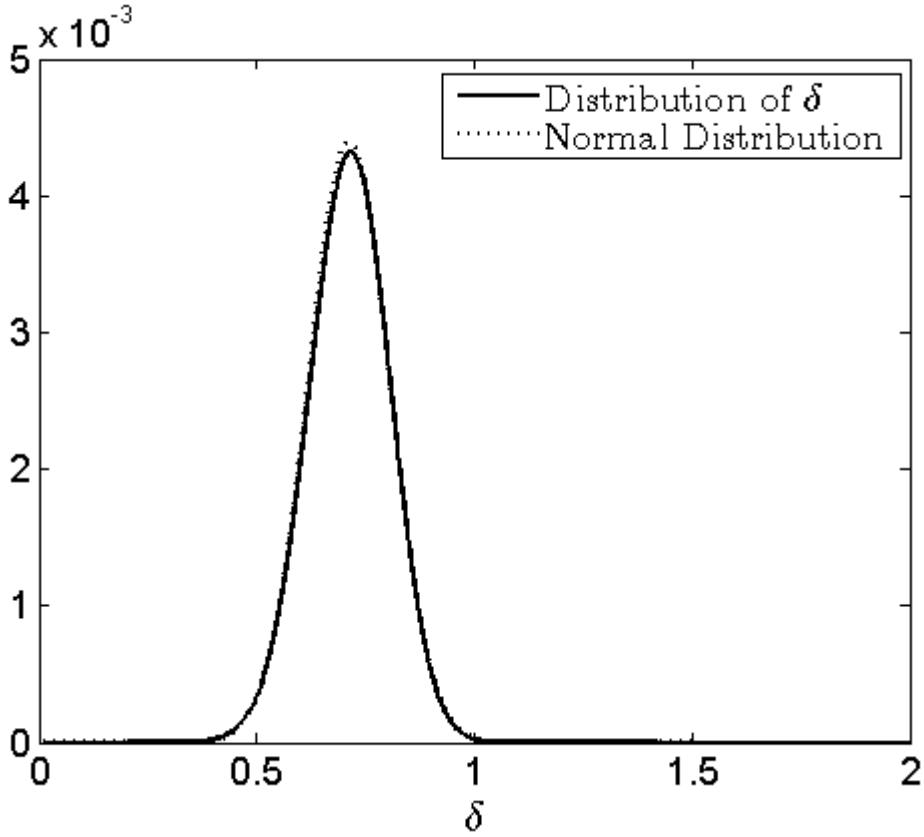
$$p\left(\delta|data\right) \propto \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right)^{-N} \prod_{i=1}^{k}\prod_{j=1}^{n_i} j^{1-\delta}$$

is illustrated for the data in Table 8.1 and in Figure 8.2 the posterior distribution of μ,

$$p\left(\mu|data\right) = \int_0^\infty p\left(\mu|\delta, data\right) p\left(\delta|data\right) d\delta$$

is displayed. The Gamma density function $p\left(\mu|\delta, data\right)$ is defined in Equation (6.6).

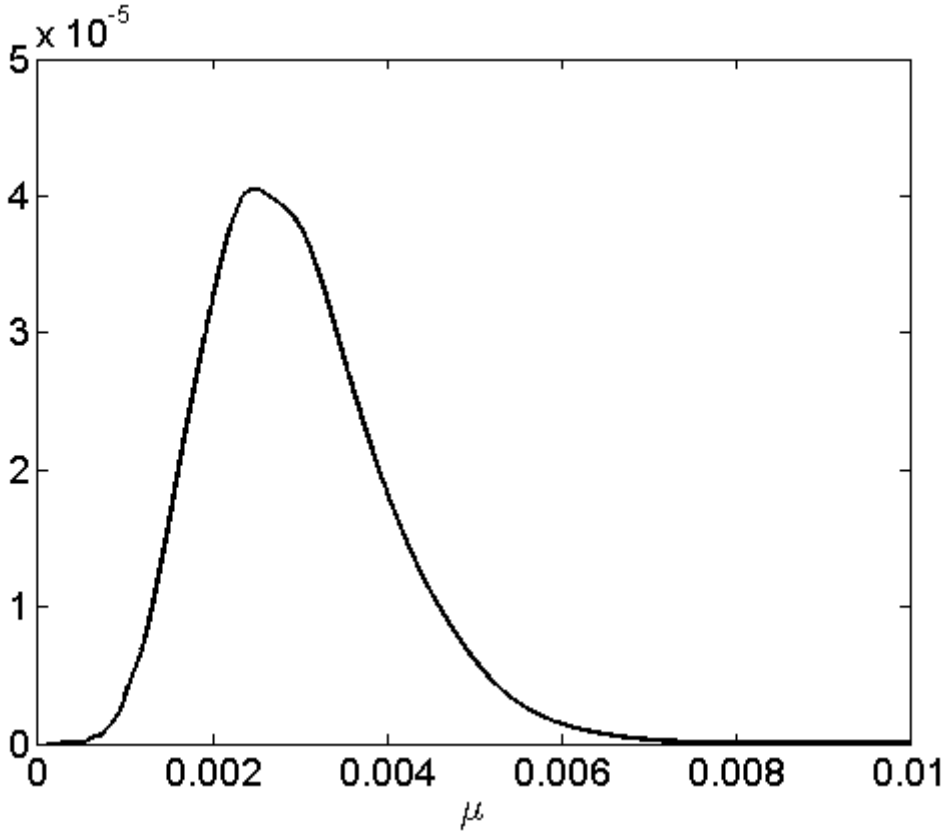
Figure 8.1.: Posterior Distribution of δ



$mean\left(\delta\right) = 0.7109,\ var\left(\delta\right) = 0.00856,\ 95\%\ HPD\ Interval\left(\delta\right) = (0.5296; 0.8922)$

The posterior distribution of δ fits almost perfectly to a normal distribution with the same mean and variance.

Figure 8.2.: Posterior Distribution of μ



$$mean\,(\mu) = 0.002982,\ var\,(\mu) = 1.1218e^{-6}$$
$$95\%\ HPD\ Interval\,(\mu) = (0.001165; 0.005120)$$
$$95\%\ Equal-tail\ Interval\,(\mu) = (0.00135; 0.00545)$$

Figure 8.2 is however obtained by using the following simulation procedure:

i. Simulate δ from $p\,(\delta|data)$.

ii. Substitute the simulated δ value in $p\,(\mu|\delta, data)$.

iii. Draw the Gamma density function $p\,(\mu|\delta, data)$.

Steps (i), (ii) and (iii) are repeated l times and by calculating the average of the l conditional Gamma density functions, the unconditional posterior distribution $p\,(\mu|data)$ is obtained. As mentioned before, this method is called the Rao-Blackwell method.

In Arab et al. (2012) a maximum likelihood procedure as well as a hierarchical Bayes method are discussed for estimating μ and δ. For the maximum likelihood procedure confidence intervals for the parameters are obtained using the delta method. Results from the hierarchical Bayes method were obtained using the Gamma prior

$$p(\mu|a,b) = \frac{b^a \mu^{a-1}}{\Gamma(a)} \exp(-b\mu) \quad \mu>0$$

and OpenBUGS.

In Table 8.2 the estimates of μ and δ as well as their confidence intervals are compared for the maximum likelihood, hierarchical Bayes and objective Bayes methods.

Table 8.2.: Point and Interval Estimates for the Parameters of the PEXM Model Assuming Identical Systems in the LHD Example

Method of Maximum Likelihood			
Parameters	**MLE**	**95% Confidence Interval**	
μ	0.002901	0.002794	0.003011
δ	0.716	0.5563	0.9215
Hierarchical Bayes Method $a = 0.1$, $b = 0.1$			
Parameters	**MLE**	**95% Confidence Interval**	
μ	0.002922	0.001264	0.005327
δ	0.705	0.5141	0.8838
Objective Bayes Method $p(\mu) \propto \mu^{-1}$			
Parameters	**MLE**	**95% Confidence Interval**	
μ	0.002982	0.00135	0.00545
δ	0.7109	0.5296	0.8922

From Table 8.2 it is clear that the point and interval estimates for the hierarchical Bayesian and objective Bayesian methods are very close to the maximum likelihood estimates and asymptotic confidence intervals obtained using the classical methods.

8.7. Simulation of PEXM Models Assuming Identical Systems and Proper Priors

To determine the capability (suitability) of the prior $p(\mu,\delta) \propto \mu^{-1}$ the following simulation study will be conducted. Ten thousand samples are drawn and each

sample represents data from six machines of sizes $n = \begin{bmatrix}23 & 25 & 27 & 28 & 26 & 23\end{bmatrix}$ where $\delta = 0.71$ and $\mu = 0.0029$. Each sample will therefore be similar to the dataset in Table 8.1. For each sample $10,000$ values are simulated from the posterior distributions of δ and μ and the means, variances and 95% intervals calculated. Table 8.3 gives the overall means, medians, modes, variances and coverage probabilities for the following priors:

- Improper Prior: $p(\mu, \delta) \propto \mu^{-1}$
- Prior (i): $p(\mu, \delta) \sim Gamma(a = 0.1, b = 0.1)$
- Prior (ii): $p(\mu, \delta) \sim Gamma(a = 1, b = 0.1)$
- Prior (iii): $p(\mu, \delta) \sim Gamma(a = 1, b = 1)$
- Prior (iv): $p(\mu, \delta) \sim Gamma(a = 2, b = 2)$
- Prior (v): $p(\mu, \delta) \sim Gamma(a = 1, b = 4)$

Table 8.3.: Simulation Study Comparing Different Priors

	Prior	Mean	Median	Mode (approx)	Var	HPD Interval		Equal-Tail Interval	
						Coverage	Length	Coverage	Length
δ	Imp.	0.7101	0.7084	0.6970	0.0099	0.9487	0.3890	0.9481	0.3896
	(i)	0.7137	0.7122	0.7150	0.0098	0.9543	0.3881	0.9533	0.3884
	(ii)	0.7474	0.7455	0.7500	0.0093	0.9256	0.3765	0.9253	0.3780
	(iii)	0.7463	0.7433	0.7320	0.0093	0.9263	0.3759	0.9273	0.3778
	(iv)	0.7793	0.7763	0.7750	0.0088	0.8765	0.3668	0.8776	0.3680
	(v)	0.7458	0.7451	0.7440	0.0093	0.9342	0.3761	0.9322	0.3777
μ	Imp.	0.0033	0.0031	0.0028	$1.67e^{-6}$	0.9432	0.0046	0.9475	0.0048
	(i)	0.0033	0.0031	0.0028	$1.68e^{-6}$	0.9494	0.0046	0.9548	0.0048
	(ii)	0.0037	0.0035	0.0032	$1.95e^{-6}$	0.9533	0.0050	0.9293	0.0052
	(iii)	0.0038	0.0035	0.0031	$1.94e^{-6}$	0.9531	0.0050	0.9320	0.0052
	(iv)	0.0042	0.0039	0.0032	$2.23e^{-6}$	0.9274	0.0054	0.8790	0.0055
	(v)	0.0037	0.0035	0.0031	$1.92e^{-6}$	0.9536	0.0050	0.9360	0.0051

From Table 8.3 it can be seen that the Jeffreys' prior $p(\mu, \delta) \propto \mu^{-1}$ gives the best estimates of μ and δ and also the best coverage; somewhat better than the Gamma prior with $a = 0.1$ and $b = 0.1$ used by Arab et al. (2012).

8.8. Objective Priors for the Mean

It might be of interest to make inferences about the mean of a piecewise exponential model. In doing so we will first derive (i) the reference prior and (ii) the probability-matching prior for the parameter

$$E\left(X_l\right) = \frac{\delta}{\mu} l^{\delta-1} = t\left(\mu, \delta\right) = t\left(\underline{\theta}\right).$$

8.8.1. Reference Prior

The reference prior is derived in such a way that it provides as little information as possible about the parameter. The idea of the reference prior approach is basically to choose the prior which, in a certain asymptotic sense, maximizes the information in the posterior distribution provided by the data.

Theorem 8.2. *The reference prior for the mean* $E\left(X_l\right) = \frac{\delta}{\mu} l^{\delta-1}$ *is* $p_R\left(\mu, \delta\right) \propto \mu^{-1}$.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

8.8.2. Probability-matching Prior

Datta and Ghosh (1995) derived the differential equation that 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 $0\left(n^{-1}\right)$ where n is the sample size.

The fact that the resulting Bayesian confidence interval of level $1-\alpha$ is also a good frequentist confidence interval at the same level is a very desirable situation.

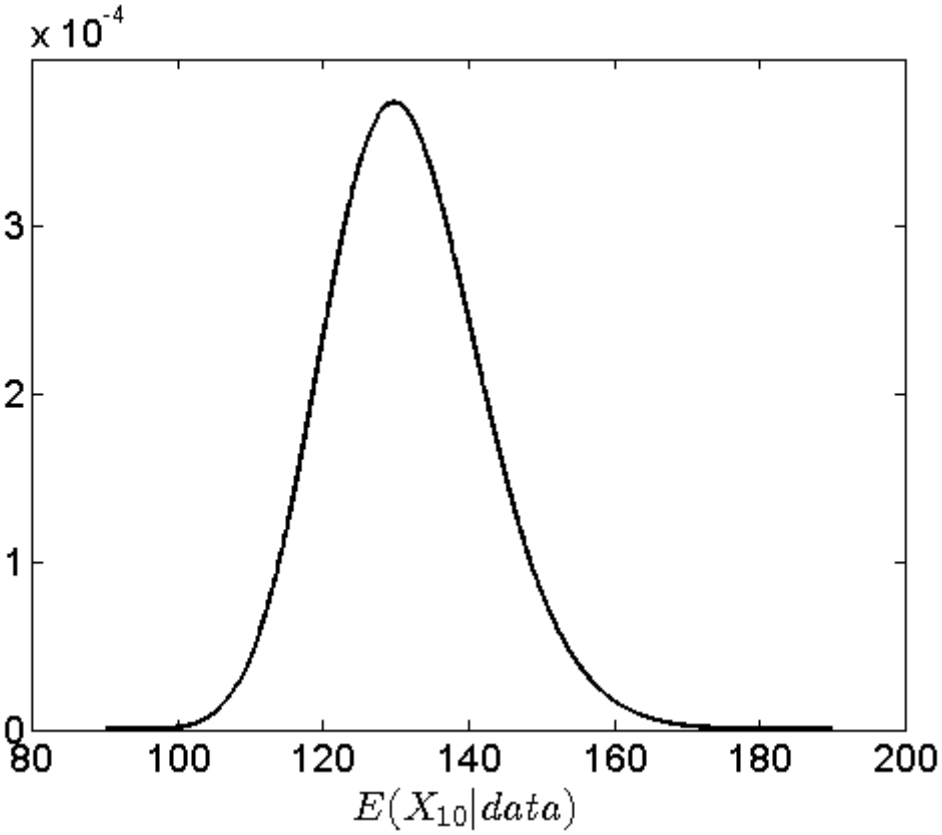
Theorem 8.3. *The probability-matching prior for the mean* $E\left(X_l\right) = \frac{\delta}{\mu} l^{\delta-1}$ *is* $p_M\left(\mu, \delta\right) \propto \mu^{-1}$.

Proof. The proof is given in the Mathematical Appendices to this chapter. □

8.9. Example: Posterior Distribution of the Mean $E(X_l) = \frac{\delta}{\mu} l^{\delta-1}$

Consider again the data in Table 8.1. The expected time between the 9th and 10th failure is $E(X_{10}) = \frac{\delta}{\mu} 10^{\delta-1}$. In Figure 8.3 the posterior distribution of $E(X_{10})$ is given.

Figure 8.3.: Posterior Distribution of Mean Time to Failure when $l = 10$

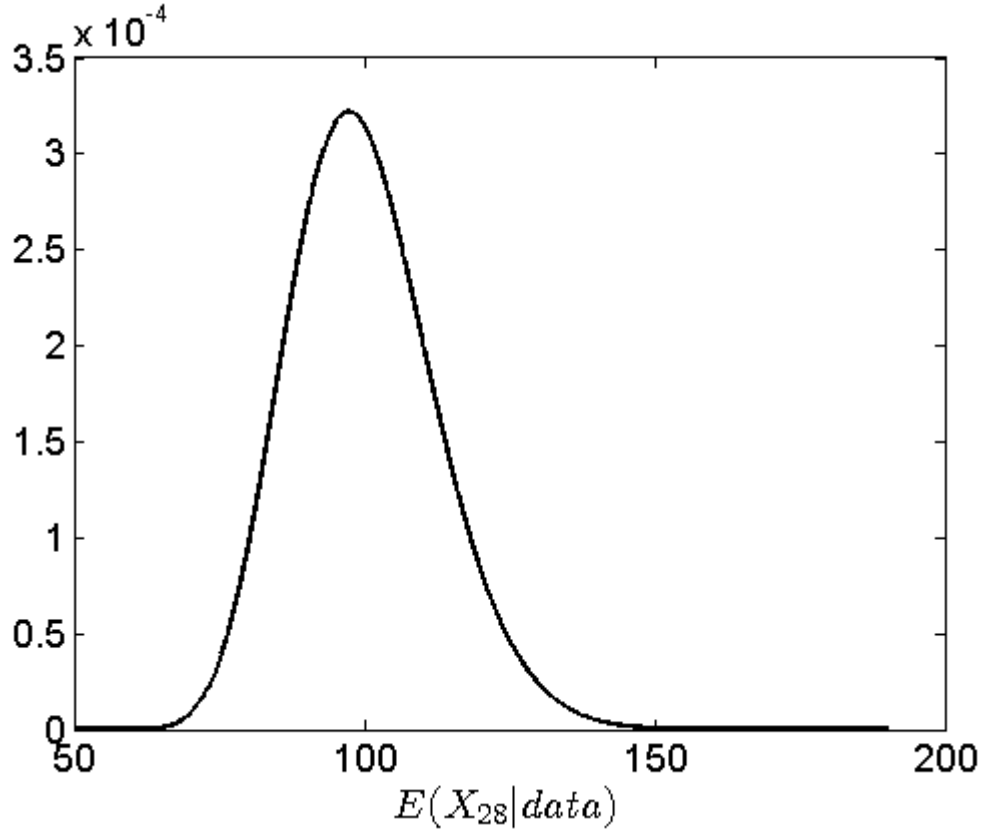


$mean = 129.598,\ median = 129.028,\ mode = 128.755,\ var = 112.329$
$95\%\ Equal-tail\ Interval = (110.42; 152.06)\ \ length = 41.66$
$95\%\ HPD\ Interval = (110.27; 151.89)\ \ length = 41.53$

It is clear from the figure that the posterior distribution is quite symmetrical.

In Figure 8.4, the posterior distribution of $E(X_{28})$ is illustrated.

Figure 8.4.: Posterior Distribution of Mean Time to Failure when $l = 28$



$mean = 97.811,\ median = 96.810,\ mode = 95.5,\ var = 149.544$
$95\%\ Equal - tail\ Interval = (76.78; 124.64)\ \ length = 47.84$
$95\%\ HPD\ Interval = (76.15; 123.32)\ \ length = 47.17$

Since $\mathrm{E}(\delta|data) = 0.7109 < 1$ the expected value $E(X_j)$ is a decreasing function of j, which corresponds to reliability deterioration. It is therefore obvious that $E(X_{28}|data) < E(X_{10}|data)$.

Figures 8.3 and 8.4 were obtained in the following way:

Let $y = \frac{\delta}{\mu} l^{\delta-1} = E(X_l)$.

We are interested in the distribution of $y = E(X_l)$. Now $\mu = \frac{\delta}{y} l^{\delta-1}$ and $\left|\frac{d\mu}{dy}\right| = \frac{1}{y^2} l^{\delta-1}$.

From the density function $p(\mu|\delta, data)$ it follows that

$$p(y|data, \delta) = \left(\frac{\delta}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} j^{1-\delta}}\right)^{-N} \frac{1}{\Gamma(N)} \left(\frac{1}{y}\delta l^{\delta-1}\right)^{N-1} \times$$

$$\exp\left\{\frac{-\delta l^{\delta-1}}{\delta y}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}\right\}\left(\frac{1}{y^2}\delta l^{\delta-1}\right).$$

$$\therefore p\left(y|data,\delta\right)=\left(l^{\delta-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}\right)^{N}\frac{1}{\Gamma\left(N\right)}\exp\left\{-\frac{1}{y}l^{\delta-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}\right\}.$$

An Inverse Gamma density function.

By using the Rao-Blackwell method $p\left(y|data\right)$ can be obtained.

8.10. Frequentist Properties of the Credibility Interval for $E\left(X_l|\mu,\delta\right)=\frac{\delta}{\mu}l^{\delta-1}$

To determine the frequentist properties (coverage probabilities) of the posterior distribution of $E\left(X_l|\mu,\delta\right)$, a simulation study as explained in Section 8.7 is done. In other words, 10,000 samples are drawn and each sample represents data from six machines of sizes $n=\begin{bmatrix}23 & 25 & 27 & 28 & 26 & 23\end{bmatrix}$ where $\delta=0.71$, $\mu=0.0029$ and $l=28$. Therefore

$$E\left(X_{28}|\mu,\delta\right)=93.1504.$$

In Table 8.4 the coverage percentage of the 95%credibility intervals are given

Table 8.4.: Coverage Percentage of the 95% Credibility Interval for $E\left(X_{28}|\mu,\delta\right)$ from 10,000 Simulated Samples

	% Coverage	Length
Equal-Tail Interval	95.55	48.51
HPD Interval	94.58	46.10

It is clear that the Bayesian credibility intervals have the correct frequentist coverage probabilities.

8.11. Predictive Distribution of a Future Observation X_f

The predictive distribution of a future observation X_f is

$$f\left(x_f|data\right) = \int_0^\infty \int_0^\infty f\left(x_f|\mu,\delta\right) p\left(\mu,\delta|data\right) d\mu d\delta$$

where

$$f\left(x_f|\mu,\delta\right) = \left(\frac{\delta}{\mu} f^{\delta-1}\right)^{-1} \exp\left\{-\frac{x_f}{\left(\frac{\delta}{\mu} f^{\delta-1}\right)}\right\} \tag{8.3}$$

and

$$p\left(\mu,\delta|data\right) = p\left(\mu|\delta,data\right) p\left(\delta|data\right)$$

is the joint posterior distribution.

The posterior distributions of δ and $\mu|\delta$ are given in Equations (6.6) and (6.7).

The unconditional predictive distribution of X_f can easily be obtained by using the simulation procedure described in Section 8.6:

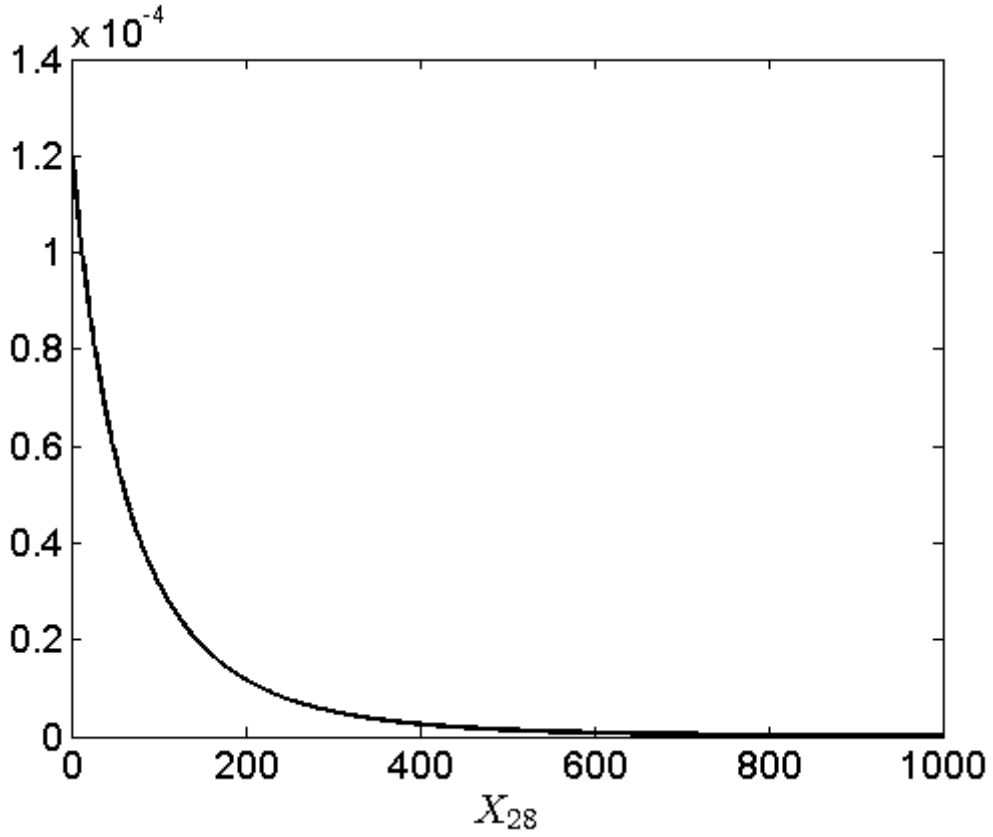
i. Obtain simulated values for δ and μ and substitute them in $f\left(x_f|\mu,\delta\right)$.

ii. Draw the exponential distribution $f\left(x_f|\mu,\delta\right)$.

iii. Repeat steps (i) and (ii) l times. The average of the l exponential distributions is $f\left(x_f|data\right)$, the unconditional predictive distribution of X_f. As mentioned before, this method is called the Rao-Blackwell procedure.

In Figure 8.5 the predictive density function of X_{28} is given for the six load-haul-dump machines.

Figure 8.5.: Predictive Density of X_{28}



$Mean = 97.96,\ Median = 68.97,\ Var = 9736.8$
$95\%\ HPD\ Interval = (0; 317.1)$
$95\%\ Equal - tail\ Interval = (2.44; 416.20)$
$99.73\%\ Equal - tail\ Interval = (0.1294; 666)$
$0.27\%\ Left - sided\ Interval = (0; 0.2605)$

8.12. Control Chart for $X_f = X_{28}$

By using the predictive distribution a Bayesian procedure will be developed in Phase II to obtain a control chart for $X_f = X_{28}$. Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length.

From Figure 8.5 it follows that for a 99.73% two-sided control chart the lower control limit is LCL=0.1294 and the upper control limit is UCL=666.

Let $R(\beta)$ represents the values of X_f that are smaller than LCL and larger than UCL. The run-length is defined as the number of future X_f values (r) until the

control chart signals for the first time (Note that r does not include the X_f value when the control chart signals). Given μ and δ and a stable Phase I process, the distribution of the run-length r is Geometric with parameter

$$\psi(\mu, \delta) = \int_{R(\beta)} f(x_f|\mu, \delta)\, dx_f$$

where $f(x_f|\mu, \delta)$ is given in Equation (8.3), i.e., the distribution of X_f given that μ and δ are known. The values of μ and δ are however unknown and the uncertainty of these parameter values are described by their joint posterior distribution $p(\mu, \delta|data)$. By simulating μ and δ from $p(\mu, \delta|data) = p(\mu|\delta, data)\, p(\delta|data)$ the probability density function of $f(x_f|\mu, \delta)$ as well as the parameter $\psi(\mu, \delta)$ can be obtained. This must be done for each future sample. In other words for each future sample μ and δ must first be simulated from $p(\mu, \delta|data)$ and then $\psi(\mu, \delta)$ calculated. Therefore, by simulating all possible combinations of μ and δ from their joint posterior distribution a large number of $\psi(\mu, \delta)$ values can be obtained. Also a large number of geometric distributions, i.e., a large number of run-length distributions each with a different parameter value $(\psi(\mu_1, \delta_1), \psi(\mu_2, \delta_2), \ldots, \psi(\mu_l, \delta_l))$ can be obtained.

As mentioned, the run-length r for given μ and δ is geometrically distributed with mean

$$E(r|\mu, \delta) = \frac{1 - \psi(\mu, \delta)}{\psi(\mu, \delta)}$$

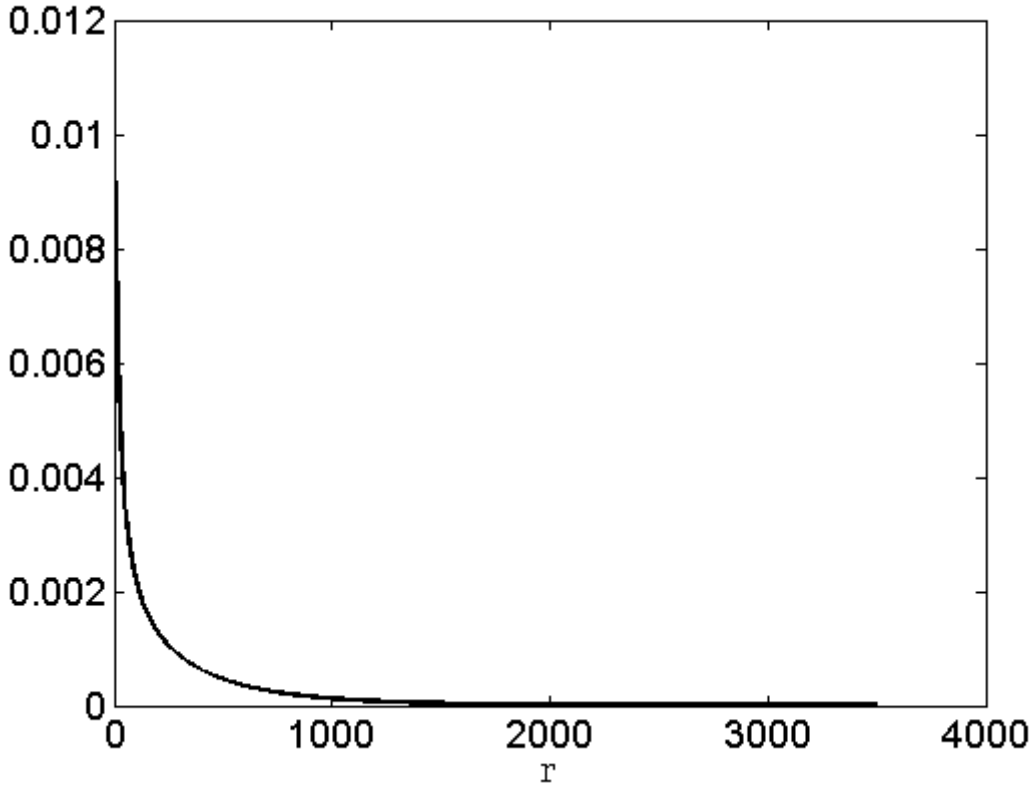
and

$$Var(r|\mu, \delta) = \frac{1 - \psi(\mu, \delta)}{\psi^2(\mu, \delta)}.$$

The unconditional moments $E(r|data)$, $E(r^2|data)$ and $Var(r|data)$ can therefore be obtained by simulation or numerical integration. For further details refer to Menzefricke (2002, 2007, 2010b,a).

By averaging the conditional distributions the unconditional distribution of the run-length can be obtained and is illustrated in Figure 8.6.

Figure 8.6.: Distribution of Run-length, $\beta = 0.0027$, Two-sided Interval

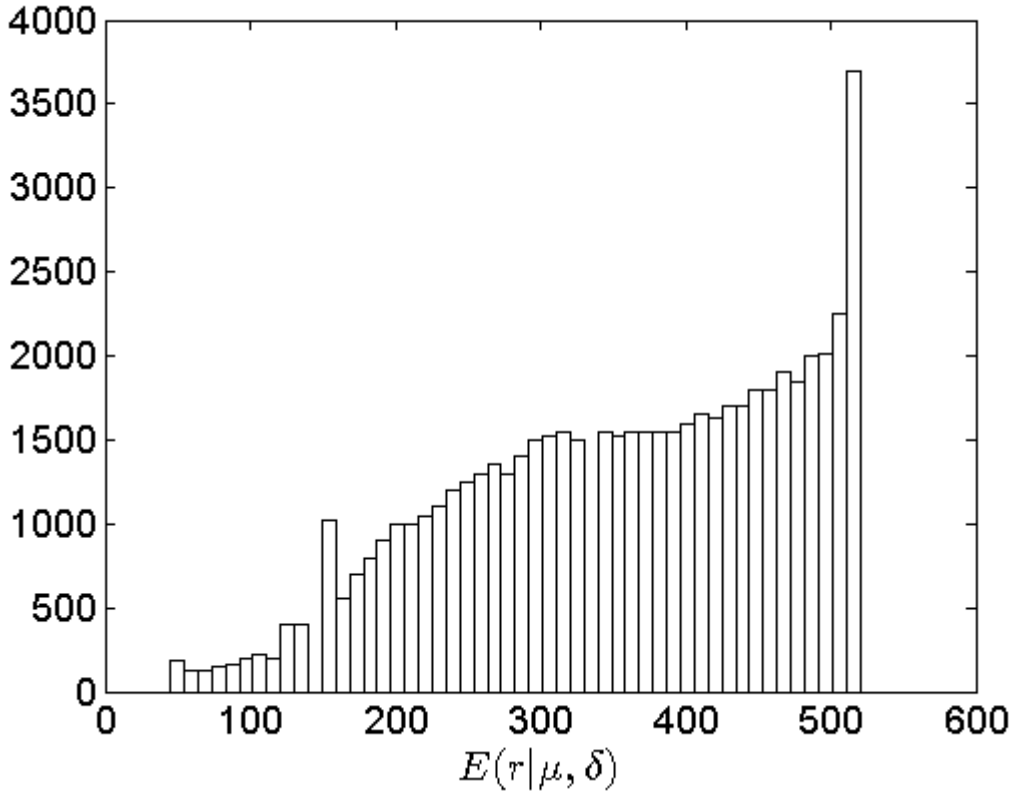


$$mean\,(r) = 369.09,\ median\,(r) = 239.75,\ var\,(r) = 1.5919e^5$$
$$95\%\ HPD\ Interval = (0; 1160.7)$$

The mean run-length of 369.09 corresponds to the value of $\frac{1}{\beta} = \frac{1}{0.0027} \approx 370$.

In Figure 8.7 the histogram of the expected run-length is given.

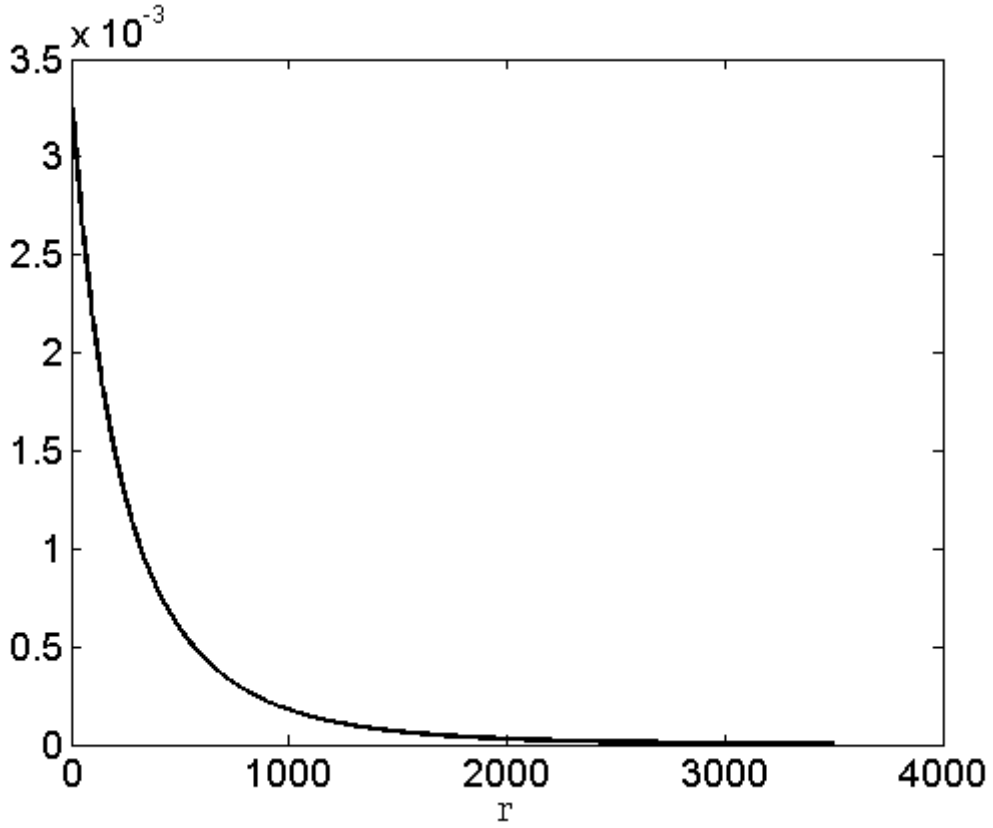
Figure 8.7.: Expected Run-length, $\beta = 0.0027$, Two-sided Interval



$mean\,(r) = 369.87,\ median\,(r) = 381.94,\ var\,(r) = 1.1407e^4$
$95\%\ HPD\ Interval = (42.35; 518.16)$

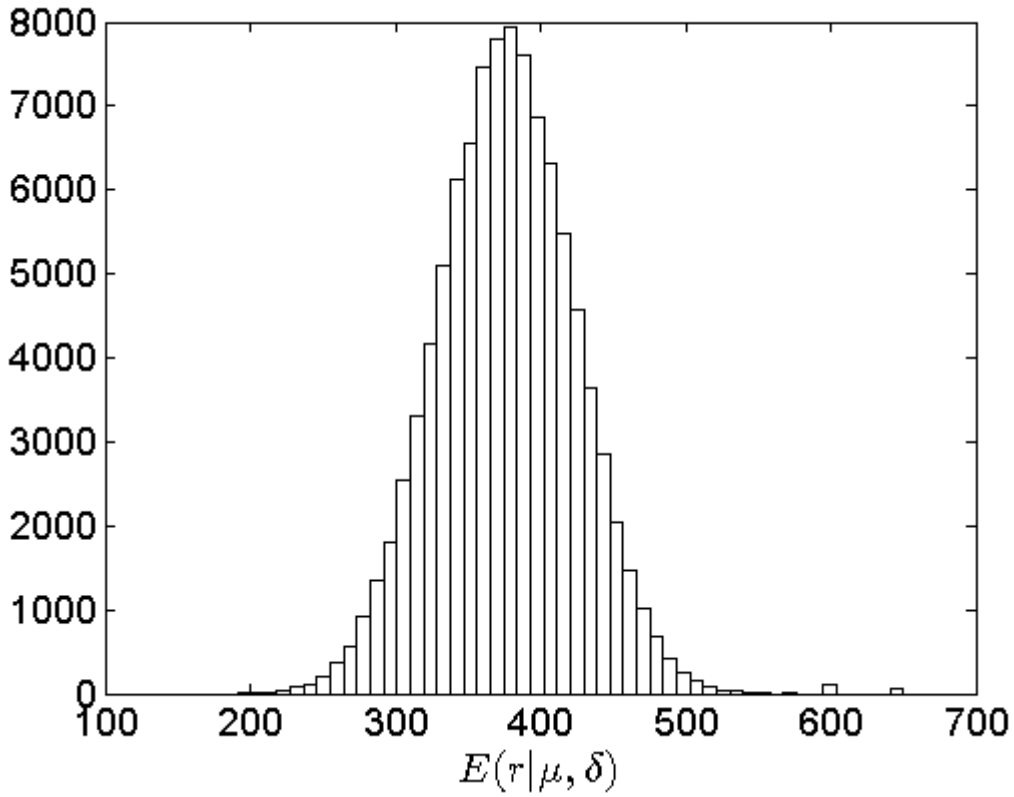
At the bottom of Figure 8.5 it is shown that a 0.27% left-sided interval = (0;0.2605). $R\,(\beta)$ therefore represents those values of X_f that are larger than 0 and smaller than 0.2605. The distribution of the run-length for this one-sided interval is displayed in Figure 8.8 and in Figure 8.9 the distribution of the expected run-length is given.

Figure 8.8.: Distribution of Run-length, $\beta = 0.0027$, One-sided Interval



$mean\,(r) = 376.30,\ median\,(r) = 256.9,\ var\,(r) = 1.4558e^5$
$95\%\ HPD\ Interval = (0; 1133.1)$

Figure 8.9.: Expected Run-length, $\beta = 0.0027$, One-sided Interval



$mean\,(r) = 376.30,\ median\,(r) = 372.58,\ var\,(r) = 2179.2$
$95\%\ HPD\ Interval = (296.14; 517.82)$

Because of the large sample size (N=152) $mean\,(r) = 376.30$ is approximately equal to $\frac{1}{\beta} = \frac{1}{0.0027} = 370$.

8.13. Frequentist Properties of the Predictive Distribution of a Future Observation $X_f = X_{28}$

It is also of interest to look at the coverage probability of the predictive distribution $f\,(x_f|data)$. The simulation study is explained in Sections 8.7 and 8.10. The only difference is that the simulated 28th observation of machine six will not form part of the data to obtain the posterior distribution $p\,(\mu, \delta|data)$. For each of the 10,000 datasets the 28th observation will therefore be different. An estimate of the coverage percentage will therefore be obtained from the number of times the predictive

interval contains the 28th observation.

In Table 8.5 the coverage percentage of the 95% prediction interval for X_{28} from 10,000 samples are given.

Table 8.5.: Coverage Percentage of 95% Prediction Interval

	% Coverage	Length
Equal-Tail Interval	95.52	351.92
HPD Interval	95.46	285.95

From Table 8.5 it is clear that the predictive interval has the correct frequentist coverage percentage.

8.14. Model 2: Systems with Different μ's but Common δ

As mentioned by Arab et al. (2012) it might happen that all systems wear out or improve at the same rate, but that the systems have different scale parameters. In the following theorem the Jeffreys' prior is derived for the case where δ is common to all systems but the μ_i's differ across systems.

Theorem 8.4. *For the piecewise exponential model with different μ's but comment δ, the Jeffreys' prior for the parameters $\mu_1, \mu_2, \ldots, \mu_k$ and δ is given by*

$$p_J(\mu_1, \mu_2, \ldots, \mu_k, \delta) \propto \prod_{i=1}^{k} \mu_i^{-1} \quad \mu_i > 0.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

8.15. The Joint Posterior Distribution of the Parameters in the Case of Model 2

The joint posterior distribution

$$\begin{aligned} p\left(\mu_1, \mu_2, \ldots, \mu_k, \delta | data\right) &\propto L\left(\mu_1, \mu_2, \ldots, \mu_k, \delta | data\right) \times p_J\left(\mu_1, \mu_2, \ldots, \mu_k, \delta\right) \\ &\propto \prod_{i=1}^{k} \left\{ \prod_{j=1}^{n_i} \left(\frac{\delta}{\mu_i} j^{\delta-1}\right)^{-1} \exp\left[-\frac{x_{ij}}{\left(\frac{\delta}{\mu_i}\right) j^{\delta-1}} \right] \right\} \prod_{i=1}^{k} \mu_i^{-1}. \end{aligned}$$

From this it follows that

$$p\left(\mu_i | data, \delta\right) = \left(\frac{\delta}{\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}} \right)^{-n_i} \frac{\mu_i^{n_i - 1}}{\Gamma\left(n_i\right)} \exp\left[-\frac{\mu_i}{\delta} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \right] \quad \mu_i > 0, i = 1, 2, \ldots, k \tag{8.4}$$

and

$$p\left(\delta | data\right) \propto \prod_{i=1}^{k} \left\{ \sum_{j=1}^{n_i} \left(x_{ij} j^{1-\delta}\right)^{-n_i} \left(\prod_{j=1}^{n_i} j^{1-\delta} \right) \right\} \quad \delta > 0 \tag{8.5}$$

The posterior distribution of δ for a piecewise exponential model with different scale parameters differs somewhat from the distribution of δ if $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ (given in Equation (6.6)).

In the figures below the posterior distributions of δ and $\mu_1, \mu_2, \ldots, \mu_k$ are displayed for the LHD example and in Table 8.6 the means variances and 95% credibility intervals are given for μ_i and δ.

Figure 8.10.: Posterior Density of δ: Model 2

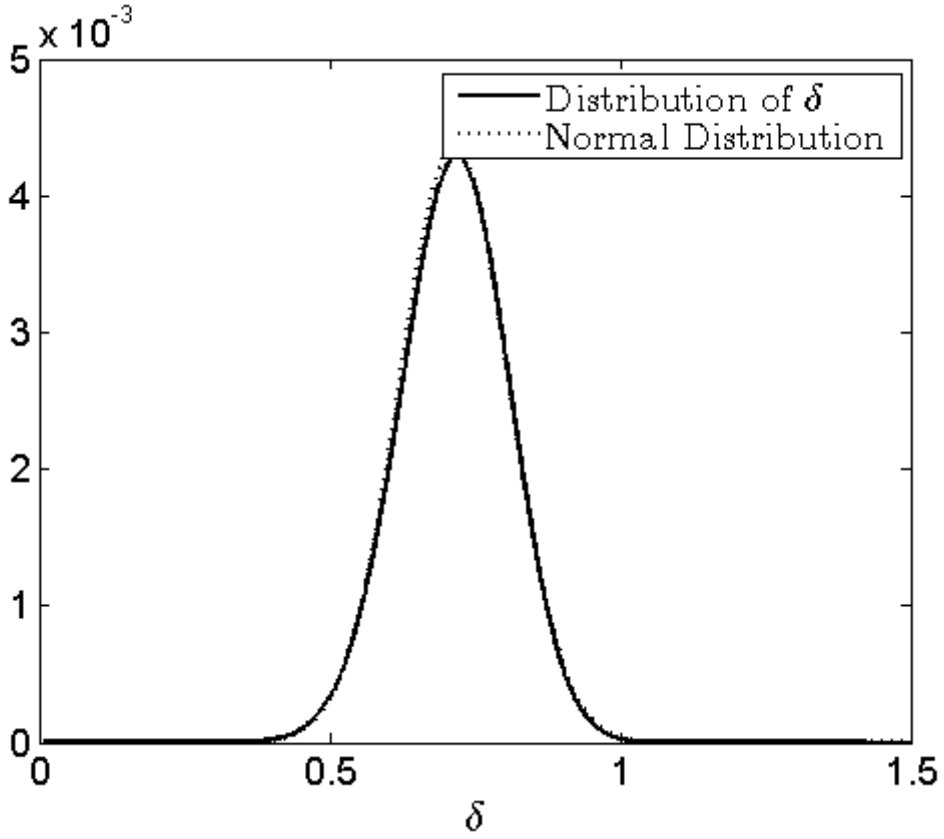


Figure 8.11.: Posterior Densities of $\mu_1, \mu_2, \ldots, \mu_6$: Model 2

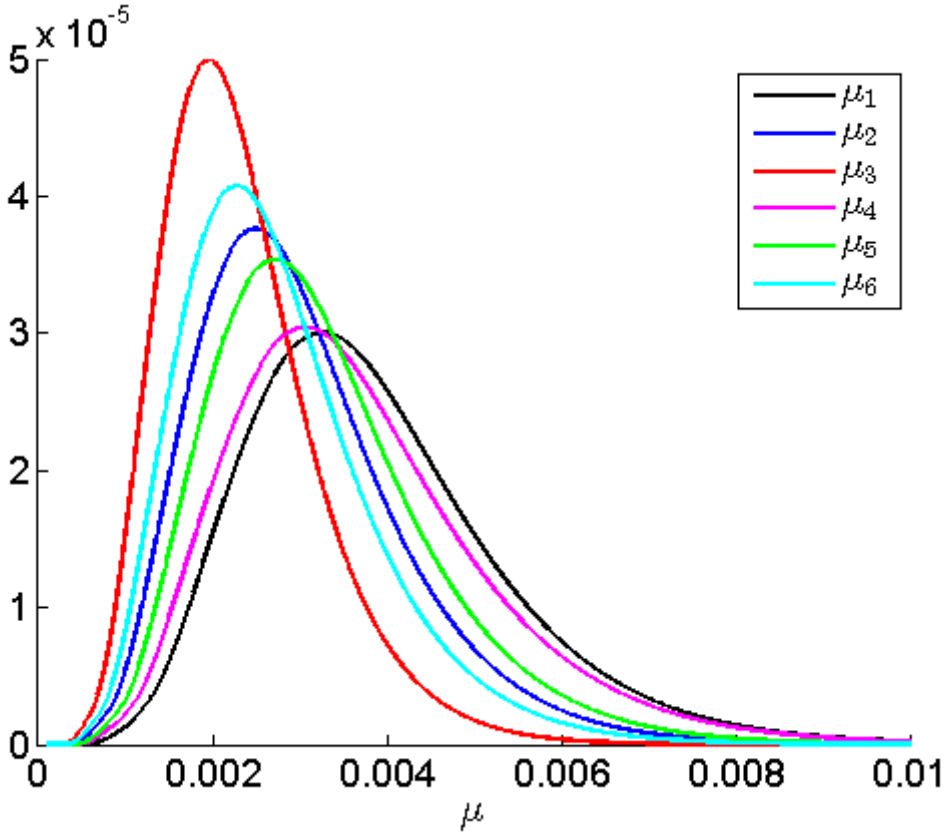


Table 8.6.: Point Estimates and Credibility Intervals for the Parameters of the PEXM Model in the Case of Systems with Different μ's and Common δ for the LHD Example

Parameter	Mean	Variance	95% HPD Interval	95% Equal-Tail Interval
μ_1	0.00387	$2.34e^{-6}$	0.00131-0.00692	0.001625-0.007515
μ_2	0.00298	$2.52e^{-6}$	0.00096-0.00545	0.001180-0.005935
μ_3	0.00231	$8.39e^{-7}$	0.00080-0.00415	0.000960-0.004490
μ_4	0.00366	$2.37e^{-6}$	0.00114-0.00674	0.001445-0.007380
μ_5	0.00323	$1.71e^{-6}$	0.00105-0.00583	0.001335-0.006340
μ_6	0.00275	$1.31e^{-6}$	0.00090-0.00506	0.001085-0.005460
δ	0.71355	0.00868	0.53076-0.89593	0.53077-0.89596

The point estimates and credibility intervals for the different μ's do not differ much.

8.16. Simulation Study of the Piecewise Exponential Model Assuming Systems with Different μ's and Common δ

In this simulation study $3,100$ samples are drawn and each sample is from six machines of size $n = \begin{bmatrix}23 & 25 & 27 & 28 & 26 & 23\end{bmatrix}$ with one δ and six μ's. For each sample $10,000$ values are drawn from the posterior distributions of δ and the μ's. The means, median, modes,variances and 95% credibility intervals are calculated. Table 8.10 gives the overall means, median, modes, variances and coverage probabilities. From the table it is clear that the point estimate of δ is for all practical purposes the same as the true value. The posterior means of the six μ's tend to be larger than the true values. The medians are better estimates of the μ's than the means. The 95% credibility intervals have the correct frequentist coverage.

Table 8.7.: Point Estimates and Credibility Intervals Obtained from a Simulation Study of the PEXM Model with Different μ's and Common δ

Parameter	True Value	Mean	Median	Mode (Approx)	Variance
δ	0.71	0.7094	0.7043	0.7000	0.0103
μ_1	0.0039	0.0046	0.0042	0.0034	$4.215e^{-6}$
μ_2	0.0030	0.0035	0.0032	0.0025	$2.637e^{-6}$
μ_3	0.0023	0.0027	0.0025	0.0020	$1.668e^{-6}$
μ_4	0.0037	0.0044	0.0039	0.0033	$4.007e^{-6}$
μ_5	0.0032	0.0037	0.0034	0.0029	$2.984e^{-6}$
μ_6	0.0028	0.0033	0.0029	0.0025	$2.272e^{-6}$

Parameter	Average 95% HPD Interval	HPD Interval		Equal-Tail	
		Cover	Length	Cover	Length
δ	(0.5205-0.9162)	0.9517	0.39856	0.9512	0.39964
μ_1	(0.0019-0.0093)	0.9517	0.00702	0.9471	0.00743
μ_2	(0.0013-0.0071)	0.9413	0.00547	0.9426	0.00581
μ_3	(0.0010-0.0056)	0.9468	0.00430	0.9504	0.00458
μ_4	(0.0016-0.0093)	0.9462	0.00680	0.9368	0.00721
μ_5	(0.0014-0.0073)	0.9491	0.00583	0.9591	0.00620
μ_6	(0.0012-0.0069)	0.9446	0.00507	0.9488	0.00538

8.17. Bayes Factors

As explained by Ando (2010) the Bayes factor is a quantity for competing models and for testing hypotheses in the Bayesian framework. It has played a major role in assessing the goodness of fit of competing models. It allows one to consider a pairwise comparison of models, say M_1 and M_2 based on the posterior probabilities. Suppose under model M_i, the data are related to parameters $\underline{\theta}_i$ by a distribution $f_i\left(\underline{y}|\theta_i\right)$ and the prior distribution is $\pi_i\left(\underline{\theta}_i\right)$, $i = 1, 2$. The posterior odds in favor of M_1 against M_2 can be written as

$$\frac{P\left(M_1|\underline{y}\right)}{P\left(M_2|\underline{y}\right)} = \frac{P\left(M_1\right)}{P\left(M_2\right)} \frac{q_1\left(\underline{y}\right)}{q_2\left(\underline{y}\right)} = \frac{P\left(M_1\right)}{P\left(M_2\right)} B\left(\underline{y}\right)$$

where $B\left(\underline{y}\right)$ is known as the Bayes factor (in favor of M_1 against M_2) and

$$q_i\left(\underline{y}\right) = \int \pi_i\left(\underline{\theta}_i\right) f_i\left(\underline{y}|\underline{\theta}_i\right) d\underline{\theta}_i$$

is the marginal likelihood (density) of $\underline{y}$ under M_i $(i = 1, 2)$. The Bayes factor chooses the model with the largest value of the marginal likelihood among a set of candidate models. The posterior odds on the other hand are the prior odds multiplied by the Bayes factor and as mentioned the Bayes factor can be seen as representing the weight of evidence in the data in favor of M_1 against M_2. If M_1 fits the data better than M_2, in the sense that $q_1\left(\underline{y}\right) > q_2\left(\underline{y}\right)$, then $B\left(\underline{y}\right) > 1$ and the posterior odds in favor of M_1 will be greater than the prior odds.

If improper priors $\pi_i\left(\underline{\theta}_i\right) = c_i h_i\left(\underline{\theta}_i\right)$, $i = 1, 2$ are used then the Bayes factor

$$B\left(\underline{y}\right) = \frac{c_1 \int h_1\left(\underline{\theta}_i\right) f_1\left(\underline{y}|\underline{\theta}_1\right) d\underline{\theta}_1}{c_2 \int h_2\left(\underline{\theta}_2\right) f_2\left(\underline{y}|\underline{\theta}_2\right) d\underline{\theta}_2}$$

depends on the ratio $\frac{c_1}{c_2}$ of two unspecified constants.

One approach to improper priors is to make use of a training sample. Berger and Pericchi (1996) proposed using all possible training samples and averaging the resulting Bayes factors. They call such an average an intrinsic Bayes factor. O'Hagan (1995) introduces an alternative to intrinsic Bayes factors that avoids the selection of

- and the subsequent averaging over training samples. His idea is to use a fraction b of the likelihood to make the improper prior, proper. This motivates the alternative definition of a Bayes factor

$$B_b\left(\underline{y}\right) = \frac{q_1\left(b, \underline{y}\right)}{q_2\left(b, \underline{y}\right)}$$

where

$$q_i\left(b, \underline{y}\right) = \frac{\int \pi_i\left(\underline{\theta}_i\right) f_i\left(\underline{y}|\underline{\theta}_i\right) d\underline{\theta}_i}{\int \pi_i\left(\underline{\theta}_i\right) f_i\left(\underline{y}|\underline{\theta}_i\right)^b d\underline{\theta}_i} \quad i = 1, 2$$

If $\pi_i\left(\underline{\theta}_i\right) = c_i h_i\left(\underline{\theta}_i\right)$ where $h_i\left(\underline{\theta}_i\right)$ is improper, the intermediate constant c_i cancels out, leaving

$$q_i\left(b, \underline{y}\right) = \frac{\int h_i\left(\underline{\theta}_i\right) f_i\left(\underline{y}|\underline{\theta}_i\right) d\underline{\theta}_i}{\int h_i\left(\underline{\theta}_i\right) f_i\left(\underline{y}|\underline{\theta}_i\right)^b d\underline{\theta}_i}.$$

$B_b\left(\underline{y}\right)$ will be referred to as a Fractional Bayes Factor (FBF).

Another way of writing $q_i\left(b, \underline{y}\right)$ is

$$q_i\left(b, \underline{y}\right) = \frac{m_i}{m_i\left(b\right)} \quad i = 1, 2$$

and

$$B_b\left(\underline{y}\right) = FBF_{12} = \frac{m_1 m_2\left(b\right)}{m_2 m_1\left(b\right)}.$$

8.18. Model Selection: Fractional Bayes Factor

In this section we will determine which one of the following models fits the LHD machine data, given in Section 8.6, the best.

Model 1: One δ and one μ

Model 2: One δ and k μ's

Marginal Likelihoods:

$$m_1 = \Gamma(N) \int_0^\infty \left\{ \prod_{i=1}^{k} \prod_{j=1}^{n_i} j^{1-\delta} \right\} \left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \right\}^{-N} d\delta$$

$$m_2 = \int_0^\infty \prod_{i=1}^{k} \left\{ \Gamma(n_i) \left(\prod_{j=1}^{n_i} j^{1-\delta} \right) \left(\sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \right)^{-n_i} \right\} d\delta$$

Fractional Marginal Likelihoods:

$$m_1(b) = \Gamma(bN) \int_0^\infty \left\{ \prod_{i=1}^{k} \prod_{j=1}^{n_i} j^{1-\delta} \right\}^b \left(b \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \right)^{-bN} d\delta$$

$$m_2(b) = \int_0^\infty \prod_{i=1}^{k} \left\{ \Gamma(bn_i) \left(\prod_{j=1}^{n_i} j^{1-\delta} \right)^b \left(b \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \right)^{-bn_i} \right\} d\delta$$

Fractional Bayes Factor:

$$FBF_{12} = \frac{m_1 m_2(b)}{m_2 m_1(b)}$$

For $b = 0.1$we have

$$\frac{m_1}{m_2} = 1.2251, \ \frac{m_2(b)}{m_1(b)} = 36.9168$$

and therefore

$$FBF_{12} = 45.2282.$$

Jeffreys (1939) recommended interpreting the Bayes factors as a scale of evidence. Table 8.8 gives Jeffreys' scale. Although the partitions seem to be somewhat arbitrary, it provides some descriptive statements. Kass and Raftery (1995) also give guidelines for interpreting the evidence from the Bayes factor.

Table 8.8.: Jeffreys' Scale of Evidence for Bayes Factor BF_{12}

Bayes Factor	Interpretation
$BF_{12} < 1$	Negative support for model M_1
$1 < BF_{12} < 3$	Barely worth mentioning evidence for M_1
$3 < BF_{12} < 10$	Substantial evidence for M_1
$10 < BF_{12} < 30$	Strong evidence for M_1
$30 < BF_{12} < 100$	Very strong evidence for M_1
$100 < BF_{12}$	Decisive evidence for M_1

Since $FBF_{12} = 45.2282$ there is very strong evidence for M_1, i.e., for the model with one δ and one μ.

Also $P\left(\text{Model } 1|data\right) = \left(1 + \frac{1}{FBF_{12}}\right)^{-1} = 0.9784$.

For further details see Ando (2010).

8.19. Conclusion

This chapter developed a Bayesian approach to model the piecewise exponential model using multiple priors and compared the results of these multiple priors. A Bayesian control chart for monitoring X_f have been implemented. Two models, one where all μ's across different systems are the same and another where μ is different for each system has been implemented and compared using Bayes factors. It has been shown that the model with less parameters are preferable.

Mathematical Appendix to Chapter 8

Proof of Theorem 8.1

To obtain the Jeffreys' prior, the Fisher information matrix must first be derived. By differentiating the log likelihood function, twice with respect to the unknown parameters and taking expected values the Fisher information matrix can be obtained:

$$l = \log_e L(\delta, \mu | data) = N \log_e \delta + N \log_e \mu + (1-\delta) \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log_e j - \frac{\mu}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta}.$$

Now

$$\frac{\partial l}{\partial \mu} = \frac{N}{\mu} - \frac{1}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta}$$

and

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{N}{\mu^2}.$$

Therefore

$$-E\left(\frac{\partial^2 l}{\partial \mu^2}\right) = \frac{N}{\mu^2}.$$

Also

$$\frac{\partial^2 l}{\partial \mu \partial \delta} = -\left\{ -\frac{1}{\delta^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} + \frac{1}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_i (-1) j^{-\delta} \log_e j \right\}.$$

This follows from the fact that $\frac{\partial}{\partial \delta}\left(j^{1-\delta}\right) = -j^{1-\delta} \log_e j$.

Therefore

$$\begin{aligned} -E\left(\tfrac{\partial^2 l}{\partial \mu \partial \delta}\right) &= \tfrac{-1}{\delta^2} \textstyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\tfrac{\delta}{\mu} j^{\delta-1}\right) j^{1-\delta} - \tfrac{1}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\tfrac{\delta}{\mu} j^{\delta-1}\right) j^{1-\delta} \log_e j \\ &= -\tfrac{1}{\mu} \left\{ \tfrac{N}{\delta} + \textstyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} \log_e j \right\}. \end{aligned}$$

Further

$$\begin{aligned} \tfrac{\partial l}{\partial \delta} &= \tfrac{-N}{\delta} - \textstyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} \log_e j - \mu \left\{ -\tfrac{1}{\delta^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} + \tfrac{1}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} (-1) j^{1-\delta} \log_e j \right\} \\ &= \tfrac{-N}{\delta} - \textstyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} \log_e j + \mu \left\{ \tfrac{1}{\delta^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} + \tfrac{1}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \log_e j \right\} \end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 l}{\partial \delta^2} &= \frac{N}{\delta^2}+\mu\{\frac{-2}{\delta^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}+\frac{1}{\delta^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}\left(-1\right)j^{1-\delta}\log_e j-\frac{1}{\delta^2}\sum_{i=1}^{j}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}\log_e\\ &\quad +\frac{1}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}\left(-1\right)j^{1-\delta}\left(\log_e j\right)^2\}\\ &= \frac{N}{\delta^2}-\mu\{\frac{2}{\delta^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}+\frac{2}{\delta^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}\log_e j\\ &\quad +\frac{1}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}j^{1-\delta}\left(\log_e j\right)^2\}\end{aligned}$$

Therefore

$$\begin{aligned}-E\left(\frac{\partial^2 l}{\partial \delta^2}\right) &= \frac{-N}{\delta^2}+\mu\{\frac{2}{\delta^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{\delta}{\mu}j^{\delta-1}\right)j^{1-\delta}+\frac{2}{\delta^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{\delta}{\mu}j^{\delta-1}\right)j^{1-\delta}\log_e j\\ &\quad +\frac{1}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{\delta}{\mu}j^{\delta-1}\right)j^{1-\delta}\left(\log_e j\right)^2\}\\ &= \frac{N}{\delta^2}+\frac{2}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2.\end{aligned}$$

The Fisher information matrix is therefore

$$F\left(\mu,\delta\right)=\begin{bmatrix}-E\left(\frac{\partial^2 l}{\partial \mu^2}\right) & -E\left(\frac{\partial^2 l}{\partial \mu \partial \delta}\right)\\ \\ -E\left(\frac{\partial^2 l}{\partial \delta \partial \mu}\right) & -E\left(\frac{\partial^2 l}{\partial \delta^2}\right)\end{bmatrix}$$

And therefore

$$F\left(\mu,\delta\right)=\begin{bmatrix}\frac{N}{\mu^2} & -\frac{1}{\mu}\left\{\frac{N}{\delta}+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right\}\\ \\ -\frac{1}{\mu}\left\{\frac{N}{\delta}+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right\} & \frac{N}{\delta^2}+\frac{2}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2\end{bmatrix}.$$

The Jeffreys' prior is

$$\begin{aligned}P_J\left(\mu,\delta\right) &\propto \left|F\left(\mu,\delta\right)\right|^{\frac{1}{2}}\\ &= \{\frac{N^2}{\mu^2\delta^2}+\frac{2N}{\mu^2\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j+\frac{N}{\mu^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2\\ &\quad -\frac{1}{\mu^2}\left(\frac{N}{\delta}+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right)^2\}^{\frac{1}{2}}\\ &= \mu^{-1}\left\{N\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2-\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right)^2\right\}\end{aligned}$$

Therefore

$$P_J(\mu, \delta) \propto \mu^{-1}.$$

Proof of Theorem 8.2

Define

$$A = \begin{bmatrix} \frac{\partial \mu}{\partial t(\underline{\theta})} & \frac{\partial \mu}{\partial \delta} \\ \\ \frac{\partial \delta}{\partial t(\underline{\theta})} & \frac{\partial \delta}{\partial \delta} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{-\delta}{t^2(\underline{\theta})} l^{\delta-1} & \frac{l^{\delta-1}}{t(\underline{\theta})}\left(1 + \delta \log_e l\right) \\ \\ 0 & 1 \end{bmatrix}.$$

The Fisher information matrix for $t(\underline{\theta})$ and δ is therefore

$$F\left(t\left(\underline{\theta}\right), \delta\right) = A' F(\mu, \delta) A = \begin{bmatrix} \tilde{F}_{11} & \tilde{F}_{12} \\ \\ \tilde{F}_{21} & \tilde{F}_{22} \end{bmatrix}$$

where

$$\tilde{F}_{11} = \frac{N}{t^2\left(\underline{\theta}\right)},$$

$$\tilde{F}_{12} = \frac{1}{t\left(\underline{\theta}\right)} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\log_e j - \log_e l\right)$$

and

$$\tilde{F}_{22} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\log_e j - \log_e l\right)^2 .$$

Now

$$\begin{aligned} P_R\left(t\left(\underline{\theta}\right)\right) &= \left(F_{11\cdot 2}\right)^{\frac{1}{2}} = \left\{\tilde{F}_{11} - \tilde{F}_{12}\tilde{F}_{22}^{-1}\tilde{F}_{21}\right\}^{\frac{1}{2}} \\ &= \tfrac{1}{t(\underline{\theta})}\left[N - \left\{\textstyle\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j - \log_e l\right)\right\}^2 \left\{\textstyle\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e l - \log_e j\right)^2\right\}^{-1}\right]^{\frac{1}{2}} . \end{aligned}$$

Therefore

$$P_R\left(t\left(\underline{\theta}\right)\right) \propto \frac{1}{t\left(\underline{\theta}\right)}.$$

Further

$$P_R\left(\delta | t\left(\underline{\theta}\right)\right) = \left(\tilde{F}_{22}\right)^{\frac{1}{2}} \propto \text{constant}$$

because $\tilde{F}_{22}$ does not contain δ.

For this it follows that

$$\begin{aligned} P_R\left(t\left(\underline{\theta}\right), \delta\right) &= P_R\left(t\left(\underline{\theta}\right)\right) P_R\left(\delta | t\left(\underline{\theta}\right)\right) \\ &\propto \tfrac{1}{t(\underline{\theta})}. \end{aligned}$$

The reference prior for the parameter space (μ, δ) is therefore

$$\begin{aligned} P_R\left(\mu, \delta\right) &\propto \tfrac{\mu}{\delta} l^{1-\delta} \left|\tfrac{dt(\underline{\theta})}{d\mu}\right| = \tfrac{\mu}{\delta} l^{1-\delta} \left|-\mu^{-2}\delta l^{\delta-1}\right| \\ &\propto \mu^{-1}. \end{aligned}$$

The reference prior is therefore exactly the same as the Jeffreys' prior.

Proof of Theorem 8.3

As before $E\left(X_l\right) = \frac{\delta}{\mu} l^{\delta-1} = t\left(\underline{\theta}\right)$.

$$F^{-1}\left(t\left(\underline{\theta}\right), \delta\right) = \begin{bmatrix} \tilde{F}^{11} & \tilde{F}^{12} \\ \\ \tilde{F}^{21} & \tilde{F}^{22} \end{bmatrix} = F^{-1}\left(\underline{\theta}\right)$$

and

$$\nabla_t'\left(\underline{\theta}\right) = \begin{bmatrix} \frac{\partial t(\underline{\theta})}{\partial t(\underline{\theta})} & \frac{\partial t(\underline{\theta})}{\partial \delta} \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Therefore

$$\nabla_t'\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right) = \begin{bmatrix} \tilde{F}^{11} & \tilde{F}^{12} \end{bmatrix}$$

and

$$\nabla_t'\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right) \nabla_t\left(\underline{\theta}\right) = \tilde{F}^{11}.$$

Define

$$\underline{\Upsilon}'\left(\underline{\theta}\right) = \frac{\nabla_t'\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right)}{\sqrt{\nabla_t'\left(\underline{\theta}\right) F^{-1}\left(\underline{\theta}\right) \nabla_t\left(\underline{\theta}\right)}} = \begin{bmatrix} \Upsilon_1\left(\underline{\theta}\right) & \Upsilon_2\left(\underline{\theta}\right) \end{bmatrix} = \begin{bmatrix} \sqrt{\tilde{F}^{11}} & \frac{\tilde{F}^{12}}{\sqrt{\tilde{F}^{11}}} \end{bmatrix}.$$

From this it follows that

$$\Upsilon_1\left(\underline{\theta}\right) = \sqrt{\tilde{F}^{11}} = \frac{t\left(\underline{\theta}\right)\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j - \log_e l\right)^2\right\}^{\frac{1}{2}}}{\sqrt{\tilde{A}}}$$

and

$$\Upsilon_2\left(\underline{\theta}\right) = \frac{\tilde{F}^{12}}{\sqrt{\tilde{F}^{11}}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e l - \log_e j\right)}{\sqrt{\tilde{} \left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\tilde{\log_e l - \log_e j}\right)^2 \tilde{A}\right\}}}$$

where

$$\tilde{A} = N \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\log_e l - \log_e j)^2 - \left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\log_e l - \log_e j) \right\}^2 .$$

As mentioned, Datta and Ghosh (1995) derived the differential equation that a prior must satisfy to be a probability matching prior.

$$P_M\left(t\left(\underline{\theta}\right), \delta\right) \propto \frac{1}{t\left(\underline{\theta}\right)}$$

is a probability-matching prior because it satisfies the differential equation

$$\frac{\partial}{\partial t\left(\underline{\theta}\right)} \left\{ \Upsilon_1\left(\underline{\theta}\right) P_M\left(t\left(\underline{\theta}\right), \delta\right) \right\} + \frac{\partial}{\partial \delta} \left\{ \Upsilon_2\left(\underline{\theta}\right) P_M\left(t\left(\underline{\theta}\right), \delta\right) \right\} = 0.$$

Similar to the reference prior it follows that $P_M(\mu, \delta) \propto \mu^{-1}$.

Proof of Theorem 8.4

The likelihood function for the case, δ common to all systems but the μ_i's differ across systems is

$$L(\delta, \mu_1, \mu_2, \ldots, \mu_k | data) = \prod_{i=1}^{k} \left\{ \prod_{j=1}^{n_i} \left(\frac{\delta}{\mu_i} j^{\delta - 1} \right)^{-1} \exp\left[- \frac{x_{ij}}{\left(\frac{\delta}{\mu_i} j^{\delta - 1} \right)} \right] \right\}$$

and the log likelihood function

$$\begin{aligned} l &= \log_e L(\delta, \mu_1, \mu_2, \ldots, \mu_k | data) \\ &= \textstyle\sum_{i=1}^{k} n_i \log_e \mu_i - N \log_e \delta + (1 - \delta) \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log_e j - \frac{1}{\delta} \sum_{i=1}^{k} \mu_i \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \end{aligned}$$

From this it follows that

$$-E\left(\frac{\partial^2 l}{(\partial \mu_i)^2} \right) = \frac{n_i}{\mu_i^2}, \quad i = 1, 2, \ldots, k,$$

$$-E\left(\frac{\partial^2 l}{\partial \mu_i \partial \delta}\right) = -\frac{1}{\mu_i}\left\{\frac{n_i}{\delta} + \sum_{j=1}^{n_i} \log_e j\right\}, \quad i = 1, 2, \ldots, k,$$

$$-E\left(\frac{\partial^2 l}{\partial \mu_i \partial \mu_l}\right) = 0, \quad i = 1, 2, \ldots, k, \ l = 1, 2, \ldots, k, \text{ and } i \neq l,$$

$$-E\left(\frac{\partial^2 l}{(\partial \delta)^2}\right) = \frac{N}{\delta^2} + \frac{2}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \log_e j + \sum_{i=1}^{k}\sum_{j=1}^{n_i} (\log_e j)^2$$

and the Fisher information matrix is given by

$$F(\mu_1, \mu_2, \ldots, \mu_k, \delta) = \begin{bmatrix} F_{11} & F_{12} \\ \\ F_{21} & F_{22} \end{bmatrix}$$

where

$$F_{11} = diag\left[\frac{n_1}{\mu_1^2} \quad \frac{n_2}{\mu_2^2} \quad \cdots \quad \frac{n_k}{\mu_k^2}\right],$$

$$F_{21} = -\left[\frac{1}{\mu_1}\left(\frac{n_1}{\delta^2} + \sum_{j=1}^{n_1} \log_e j\right) \quad \frac{1}{\mu_2}\left(\frac{n_2}{\delta^2} + \sum_{j=1}^{n_2} \log_e j\right) \quad \cdots \quad \frac{1}{\mu_k}\left(\frac{n_k}{\delta^2} + \sum_{j=1}^{n_k} \log_e j\right)\right] = F_{21}'$$

and

$$F_{22} = \frac{N}{\delta^2} + \frac{2}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \log_e j + \sum_{i=1}^{k}\sum_{j=1}^{n_i} (\log_e j)^2 .$$

The Jeffreys' prior is proportional to the square root of the determinant of the Fisher information matrix. Therefore since

$$\begin{aligned} |F(\mu_1, \mu_2, \ldots, \mu_k, \delta)| &= |F_{11}| \left| F_{22} - F_{21} F_{11}^{-1} F_{21} \right| \\ &= \left(\prod_{i=1}^{k} \frac{n_i}{\mu_i^2} \right) \left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\log_e j)^2 - \sum_{i=1}^{k} \frac{1}{n_1} \left(\sum_{j=1}^{n_i} \log_e j \right)^2 \right\} \end{aligned}$$

the Jeffreys' prior follows as

$$\begin{aligned} P_J(\mu_1, \mu_2, \ldots, \mu_k, \delta) &\propto |F(\mu_1, \mu_2, \ldots, \mu_k, \delta)|^{\frac{1}{2}} \\ &= \prod_{i=1}^{k} \mu_i^{-1} \quad \mu_i > 0; i = 1, 2, \ldots, k. \end{aligned}$$

9. Process Capability Indices C_{pl}, C_{pu} and C_{pk} and Control Chart for the C_{pk} Index

9.1. Introduction

Process capability indices have been widely used in the manufacturing industry. They measure the ability of a manufacturing process to produce items that meet certain specifications. A capability index relates the voice of the customer (specification limits) to the voice of the process. A large value of the index indicates that the current process is capable of producing items (parts, tablets) that will meet or exceed the customers' requirements. Capability indices are convenient because they reduce complex information about the process to a single number and measure relative variability similar to the coefficient of variation.

Application examples include the manufacturing of semiconductor products (Hoskins, Stuart, and Taylor (1988)), jet-turbine engine components (Hubele, Shahriari, and Cheng (1991)), wood products (Lyth and Rabiej (1995)), audio speaker drivers (Chen and Pearn (1997)) and many others.

There is a need to understand and interpret process capability indices. In the literature on statistical quality control there have been some attempts to study the inferential aspects of these indices. Most of the existing works in this area has been devoted to classical frequentist large sample theory.

As mentioned by Pearn and Wu (2005) a point estimate of the index is not very useful in making reliable decisions. An interval estimation approach is in fact more appropriate and widely accepted but the frequency distributions of these estimators are often very complicated which means that the calculation of exact confidence intervals will be difficult.

An alternative approach to the problem of making inference about capability indices is the Bayesian approach. As it is well known in the Bayesian approach the information contained in the prior is combined with the likelihood to obtain the posterior distribution of the parameters. Inferences about the unknown parameters are based on the posterior distribution.

9.2. Definitions and Notations

Four of the commonly used capability indices are:

$$C_p = \frac{u - l}{6\sigma},$$

$$C_{pu} = \frac{u - \mu}{3\sigma},$$

$$C_{pl} = \frac{\mu - l}{3\sigma}$$

and

$$C_{pk} = min\left(C_{pu}, C_{pl}\right).$$

C_{pk} is the normalized distance between the process mean and its closest specification limit. It can easily be verified that $C_{pk} = C_p\left(1 - w\right)$ where $w = \frac{2|m-\mu|}{u-l}$ and $m = \frac{u+l}{2}$ is the midpoint of the specification limits (u and l). Thus, C_{pk} modifies C_p by a standardized measure w of non-centrality of the process and $C_{pk} = C_p$ if and only if the process is centered at m.

The larger the value of C_{pk}, the more capable is the process. In general, if the value of a process capability index is greater than 1.0 the process is said to be capable. According to Niverthi and Dey (2000), the thrust these days in the manufacturing

industry is to achieve a C_{pk} value of at least 1.33. The definition of C_{pk} includes as special case those processes for which only one limit exists, by letting either $l \to -\infty$ or $u \to \infty$, in which case it reduces to the appropriate standardized measure.

Let $y_1, y_2, \ldots, y_n$ be an independent sample from a manufacturing process. In this chapter it will be assumed that the y_i $(i = 1, 2, \ldots, n)$ are independent, identically normally distributed random variables with mean μ and variance σ^2. Since both μ and σ^2 are unknown and no prior information is available, the conventional non-informative, Jeffreys' prior

$$p\left(\mu, \sigma^2\right) \propto \sigma^{-2} \tag{9.1}$$

will be specified for μ and σ^2 in this section. Using (9.1), it is well known (see for example Zellner (1971)) that the conditional posterior density function of μ is normal:

$$\mu|\sigma^2, \underline{y} \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right) \tag{9.2}$$

and in the case of the variance component σ^2, the posterior density function is given by

$$p\left(\sigma^2|\underline{y}\right) = K\left(\sigma^2\right)^{-\frac{1}{2}(n+1)} \exp\left\{-\frac{1}{2}\frac{(n-1)\,s^2}{\sigma^2}\right\} \quad \sigma^2 > 0 \tag{9.3}$$

an Inverted Gamma density function where $\underline{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}'$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance and

$$K = \left\{\frac{(n-1)\,s^2}{2}\right\}^{\frac{1}{2}(n-1)} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \tag{9.4}$$

is a normalizing constant.

From (9.3) it follows that

$$k = \frac{(n-1)\,s^2}{\sigma^2} \sim \chi^2_{n-1} = \chi^2_v \tag{9.5}$$

for a given s^2.

As mentioned these indices are used in process evaluation. From a Bayesian point of view the posterior distributions are of importance. One of the aims of this chapter is therefore to derive the exact and in some cases the conditional posterior distributions of the indices. The method proposed by Ganesh (2009) for multiple testing will be applied using a Bayesian procedure for C_{pl}, C_{pu} and C_{pk} to determine whether significant differences between four suppliers exist. A Bayesian control chart for C_{pk} will also be implemented.

An estimated index will be denoted by "hat" ($\widehat{\ }$). For example $\hat{C}_p = \frac{u-l}{6s}$, $\hat{C}_{pl} = \frac{\bar{y}-l}{3s}$, $\hat{C}_{pu} = \frac{u-\bar{y}}{3s}$ and $\hat{C}_{pk} = min\left(\hat{C}_{pu}, \hat{C}_{pl}\right)$.

The following theorems will now be proved.

9.3. The Posterior Distribution of the Lower Process Capability Index $C_{pl} = \frac{\mu-l}{3\sigma}$

Theorem 9.1. *The posterior distribution of $t = C_{pl}$ is given by*

$$p\left(t|\tilde{t}\right) = \frac{3\sqrt{n}\exp\left\{-\frac{9nt^2}{2}\right\}}{\Gamma\left(\frac{v}{2}\right)\sqrt{2\pi}} \sum_{j=0}^{\infty} \left(\frac{9nt\tilde{t}}{\sqrt{v}}\right)^j \frac{1}{j!} \frac{\Gamma\left(\frac{v+j}{2}\right) 2^{\frac{1}{2}j}}{\left(1+\frac{9n}{v}\tilde{t}^2\right)^{\frac{v+j}{2}}} \qquad -\infty < t < \infty \tag{9.6}$$

where

$$\tilde{t} = \frac{\bar{y}-l}{3s} = \hat{C}_{pl}$$

and

$v = n - 1.$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

Note:

Chou and Owen (1989) derived the distribution of $\tilde{t}$ which is given by

$$f\left(\tilde{t}|t\right) = \frac{3\sqrt{n}\exp\left\{-\frac{9nt^2}{2}\right\}}{\sqrt{v}\sqrt{2\pi}\Gamma\left(\frac{v}{2}\right)} \sum_{j=0}^{\infty} \left(\frac{9n\tilde{t}t}{\sqrt{v}}\right)^j \frac{1}{j!} \frac{\Gamma\left(\frac{v+j+1}{2}\right) 2^{\frac{1}{2}j}}{\left(1+\frac{9n}{v}\tilde{t}^2\right)^{\frac{v+j+1}{2}}}. \tag{9.7}$$

Equation (9.7) is a non-central t distribution with v degrees of freedom and non-centrality parameter δ where $\delta^2 = 9nt^2$.

9.4. The Posterior Distribution of $C_{pk} = min\left(C_{pl}, C_{pu}\right)$

When both specification limits are given, the C_p and C_{pk} indices can be used where

$$C_{pk} = min\left(C_{pl}, C_{pu}\right).$$

Unlike C_p, C_{pk} depends on both μ and σ. The C_{pk} index has been used in Japan and in the U.S. automotive companies (see Kane (1986)) and Chou and Owen (1989).

In Theorem 9.2 the posterior distribution of $c = C_{pk}$ will be derived.

Theorem 9.2. *The posterior distribution of $c = C_{pk}$ is given by*

$$\begin{aligned} p\left(c|y\right) \quad = \quad & \frac{3\sqrt{n}}{\sqrt{2\pi}} \int_{\frac{c^2 v}{\tilde{t}^2}}^{\infty} \left\{\exp\left(-\frac{9n}{2}\left[c - t^*\sqrt{\frac{k}{v}}\right]^2\right) + \exp\left(-\frac{9n}{2}\left[c - \tilde{t}\sqrt{\frac{k}{v}}\right]^2\right)\right\} \times \\ & \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} k^{\frac{v}{2}-1}\exp\left(-\frac{k}{2}\right) dk \end{aligned} \tag{9.8}$$

where

$$v = n - 1,$$

$$t^* = \hat{C}_{pu} = \frac{u - \bar{y}}{3s},$$

$$\tilde{t} = \hat{C}_{pl} = \frac{\bar{y} - l}{3s}$$

and

$$\tilde{b} = \hat{C}_p = \frac{u - l}{6s}.$$

Proof. The proof is given in the Mathematical Appendices to this chapter. □

9.5. Example: Piston Rings for Automotive Engines (Polansky (2006))

Consider a company with $N = 4$ contracted suppliers representing the four processes that produce piston rings for automobile engines studied by Chou (1994). The edge width of a piston ring after the preliminary disk grind is a very important quality characteristic in automobile engine manufacturing. The lower and upper specification limits of the quality characteristic are $l = 2.6795mm$ and $u = 2.7205mm$ respectively Four potential suppliers (Supplier 1 to Supplier 4) for such rings are under consideration by one quality control manager. Samples of size $n_1 = 50$, $n_2 = 75$, $n_3 = 70$ and $n_4 = 75$ are taken from the manufacturing processes of the suppliers. A summary of the results from the samples, $\hat{C}_{pl}$, $\hat{C}_{pu}$, $\hat{C}_{pk}$ values and other statistics are given in Table 9.1.

Table 9.1.: $\hat{C}_{pl}$, $\hat{C}_{pu}$, and $\hat{C}_{pk}$ Values for the Four Suppliers

Supplier (i)	1	2	3	4
Sample Size (n_i)	50	75	70	75
Estimated Mean ($\bar{y}_i$)	2.7048	2.7019	2.6979	2.6972
Estimated Standard Deviation (s_i)	0.0034	0.0055	0.0046	0.0038
$\hat{C}_{pl}^{(i)} = \frac{\bar{y}_i - l}{3s_i}$	2.4804	1.3576	1.3333	1.5526
$\hat{C}_{pu}^{(i)} = \frac{u - \bar{y}_i}{3s_i}$	1.5392	1.1273	1.6377	2.0439
$\hat{C}_{pk}^{(i)} = min\left(\hat{C}_{pl}^{(i)}, \hat{C}_{pu}^{(i)}\right)$	1.5392	1.1273	1.3333	1.5526

Looking at Table 9.1, it is clear that Suppliers 4 and 1 give the two largest values for $\hat{C}_{pl}$, $\hat{C}_{pu}$ and $\hat{C}_{pk}$, suggesting that they are the most capable. This may be because they seem to have the smallest variation within the specified range. They therefore represent the best two choices of suppliers. Suppliers 3 and 2 are not as capable as the former because of their greater variability. Because the estimated $\hat{C}_{pk}$ index for Supplier 1 is close to that of Supplier 4 we might feel that that the difference in capability of the processes between these suppliers is not significant. The same statement may hold true of Suppliers 2 and 3. Statistical methods for the comparison of the suppliers' process capability indices are required for the quality control manager to draw intelligent conclusion from this data. A Bayesian simulation procedure will be considered to determine which processes are significantly different from one another. The potential performance of the proposed method will be compared with the permutation approach by Polansky (2006).

Before discussion of the simulation procedure the posterior distributions of the capability indices will be looked at.

In the last part of this chapter control limits will be calculated for future capability indices. In Figure 9.1 the posterior distributions of C_{pk} are illustrated using Equation (9.8) and numerical integration.

Figure 9.1.: Posterior Distributions of C_{pk}

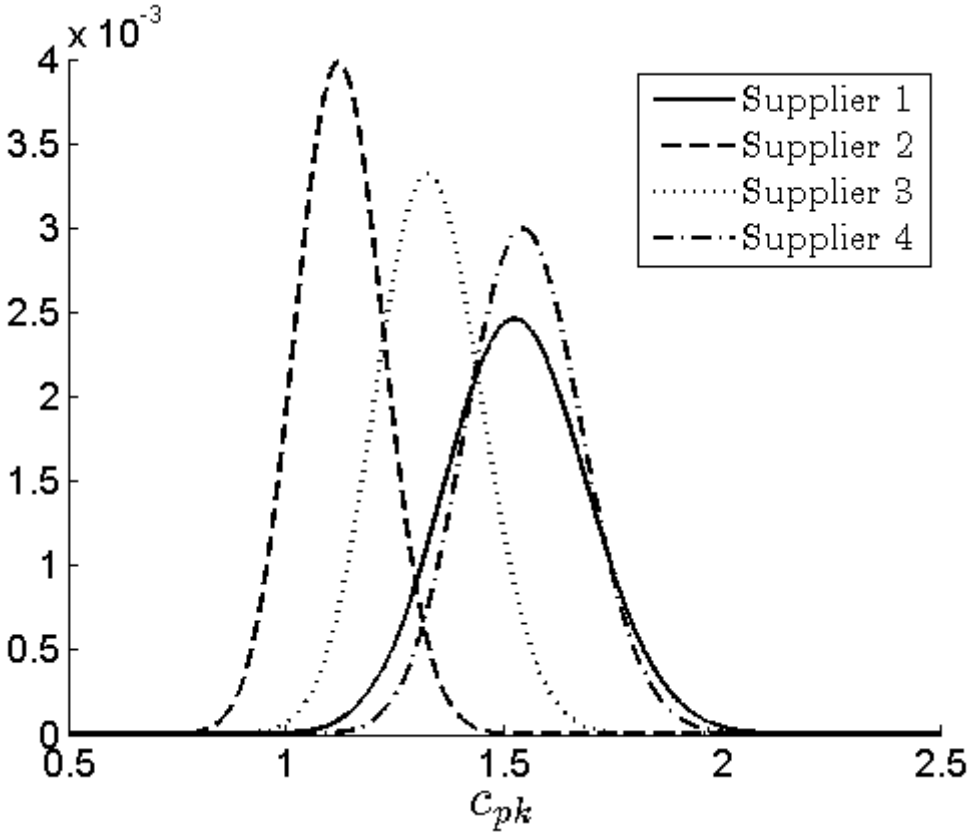


Table 9.2.: Posterior Means and Variances

	Supplier 1	Supplier 2	Supplier 3	Supplier 4
Posterior Mean	1.5314	1.1234	1.3285	1.5474
Posterior Variance	0.0263	0.0100	0.0144	0.0177

From Table 9.2 it can be seen that the posterior means are for all practical purposes the same as the $\hat{C}_{pk}$ values given in table 9.1. Further inspection of Figure 9.1 and Table 9.2 shows that Suppliers 4 and 1 have the largest posterior means, suggesting they are the most capable. In the next section a simple Bayesian solution to the problem of constructing simultaneous credibility intervals for the capability indices will be discussed.

9.6. Simultaneous Credibility Intervals

The method proposed by Ganesh (2009) can be compared to multiple testing, also referred to as the multiple comparison problem. In multiple testing, the objective is to control the family wise error rate. Similarly in his chapter, Ganesh control the simultaneous coverage rate. If the interest is in constructing simultaneous credibility intervals for all pairwise differences, a Bayesian version of Tukey's simultaneous confidence intervals can be used. Define

$$T^{(2)} = \underset{l}{max}\left\{C_{pk}^{(l)} - E\left(C_{pk}^{(l)}|\underline{y}\right)|\underline{y}\right\} - \underset{l}{min}\left\{C_{pk}^{(l)} - E\left(C_{pk}^{(l)}|\underline{y}\right)|\underline{y}\right\}$$

where $T_{\alpha}^{(2)}$ is the upper α percentage point of the distribution of $T^{(2)}$. Simultaneous $100\,(1-\alpha)\,\%$ credibility intervals for all pairwise differences are given by

$$E\left(C_{pk}^{(i)}|\underline{y}\right) - E\left(C_{pk}^{(j)}|\underline{y}\right) \pm T_{\alpha}^{(2)} \quad i = 1, 2, \ldots, 4;\ j = 1, 2, \ldots, 4;\ i \neq j$$

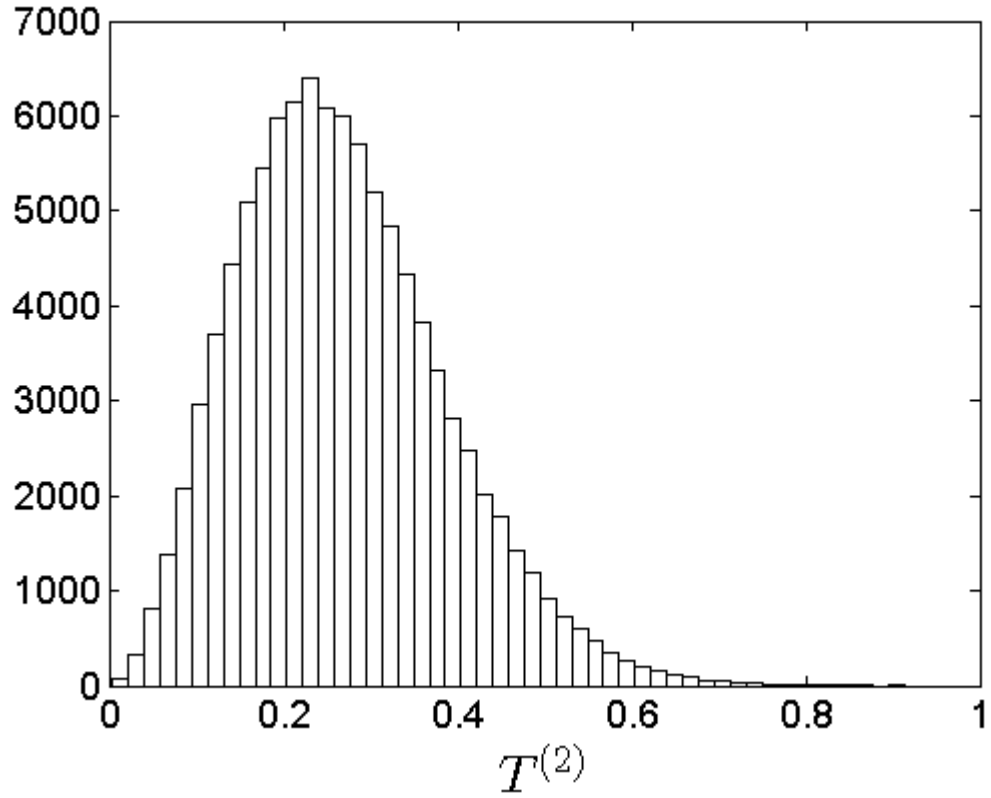
100,000 Monte Carlo simulations were used to calculate $E\left(C_{pk}^{(i)}|\underline{y}\right)$, $E\left(C_{pk}^{(j)}|\underline{y}\right)$ and $T_{\alpha}^{(2)}$.

The simulation procedure is as follows:

1. Simulate k from a χ_{n-1}^2 distribution.
2. Calculate $\sigma_i^{2*} = \frac{(n-1)s_i^2}{k}$ (Equation (9.5)) where (*) indicates a simulated value $(i = 1, 2, \ldots, 4)$.
3. $\sigma_i^* = \sqrt{\sigma_i^{2*}}$
4. By using the fact that $\mu_i|\sigma_i^2, \bar{y}_i \sim N\left(\bar{y}_i, \frac{\sigma_i^2}{n}\right)$ (Equation (9.2)) simulate μ_i^*.
5. From the definition of the capability index it follows that $C_{pk}^{(i)}$ can be simulated as $C_{pk}^{(i)*} = min\left(\frac{u-\mu_i^*}{3\sigma_i^*}, \frac{\mu_i^*-l}{3\sigma_i^*}\right)$.
6. Repeat steps 1 to 5 $\tilde{l}$ times. As mentioned, for this example $\tilde{l} = 100,000$.

In Figure 9.2 the posterior distribution of $T^{(2)}$ is given and in Table 9.3 credibility intervals for differences in C_{pk} are given using Ganesh's method.

Figure 9.2.: Distribution of $T^{(2)}$



$T^{(2)}_{0.05} = 0.4823$, $T^{(2)}_{0.1} = 0.4279$ and $T^{(2)}_{0.15} = 0.3915$

Table 9.3.: Credibility Intervals for Differences in C_{pk} - Ganesh Method

$E\left(C^{(i)}_{pk}\|\underline{y}\right) - E\left(C^{(j)}_{pk}\|\underline{y}\right)$	95% Interval	90% Interval	87.47% Interval
Supplier 1 - Supplier 2	(-0.0734;0.8915)	(-0.0187;0.8371)	**(0;0.8155)**
Supplier 1 - Supplier 3	(-0.2779;0.6867)	(-0.2234;0.6323)	(-0.2058;0.6097)
Supplier 1 - Supplier 4	(-0.4971;0.4675)	(-0.4427;0.4131)	(-0.4245;0.3910)
Supplier 2 - Supplier 3	(-0.6871;0.2775)	(-0.6326;0.2231)	(-0.6136;0.2019)
Supplier 2 - Supplier 4	(-0.9063;0.0583)	(-0.8519;0.0039)	**(-0.8323;-0.0168)**
Supplier 3 - Supplier 4	(-0.7016;0.2630)	(-0.6471;0.2086)	(-0.6265;0.1890)

For solving the supplier problem Polansky (2006) used multiple comparison techniques in conjunction with permutation tests. The multiple comparisons tests used were:

a. The Bonferonni method, which adjusts the significance levels of the pair wise tests.

b. The protected multiple comparison method, which requires that an omnibus test of equality between all of the process capability indices be rejected before pair-wise tests are performed and does not require adjustment of the significance level of the pair-wise tests.

Polansky (2006) came to the conclusion that at the 5% significance level suppliers 1,2 and 4 have process capabilities that are not significantly different. Similarly, suppliers 2 and 3 are not significantly different from one another, but supplier 2 is significantly different from Supplier 1 and 4.

According to Table 9.3 it is only at significance level of 12.5% that the Bayesian procedure shows a significant difference between Supplier 2 and Suppliers 1 and 4. To see if Ganesh (2009) version of Tukey's simultaneous confidence intervals is somewhat conservative, the following simulation study has been conducted to evaluate the coverage probability and power of the Bayesian hypothesis testing procedure.

I. a. Assume that $y \sim N(\mu_1, \sigma_1^2)$ where $\mu_1 = 2.7048$ and $\sigma_1^2 = (0.0034)^2$. The parameters μ_1 and σ_1^2 are obtained from the sample statistics of Supplier 1.

b. Simulate the sufficient statistics $\bar{y}_i \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$ and $(n_1 - 1)\, s_i^2 \sim \sigma_1^2 \chi^2_{n_1 - 1}$ to represent a data set for the four suppliers where $n_1 = 50$ and $i = 1, 2, 3, 4$.

c. By doing $\tilde{l} = 10000$ simulations $T^{(2)}_{0.05}$ can be calculated for our first dataset as well as the credibility intervals as described in Section 6.

d. If any one of the six credibility intervals do not contain zero, the null hypothesis

$$H_0 : C_{pk}^{(1)} = C_{pk}^{(2)} = C_{pk}^{(3)} = C_{pk}^{(4)}$$

will be rejected. Rejection of H_0 when it is true is called a Type I error.

e. Steps (a) - (d) are replicated $l = 20,000$ times with $\mu_1 = 2.7048$, $\sigma_1^2 = (0.0034)^2$ and $n_1 = 50$ and the estimated Type I error $= \frac{1008}{20000} = 0.0504$ which corresponds well with $\alpha = 0.05$. It means that for 1008 datasets one or more of the six credibility intervals did not contain zero.

II. Assume now that $y \sim N(\mu_2, \sigma_2^2)$ where $\mu_2 = 2.7019$, $\sigma_2^2 = (0.0054)^2$ and $n_2 = 75$. The parameter values are that of the sample statistics of the second supplier. Repeat steps I (a) - I (e) and also for Suppliers 3 and 4.

In Table 9.4 the estimated Type I errors for the four cases are given.

Table 9.4.: Estimated Type I Error for Different Parameter Combinations and Sample Sizes

n	μ	σ	Type I Error
50	2.7048	0.0034	0.0504
75	2.7019	0.0054	0.0483
70	2.6979	0.0046	0.0521
75	2.6972	0.0038	0.0507

The average Type I error = 0.0504 which as mentioned corresponds well with $\alpha = 0.05$. It therefore does not seem that Ganesh Bayesian version of Tukey's simultaneous confidence interval is too conservative.

9.7. Type II Error of Ganesh Bayesian Method

Acceptance of H_0 when it is false is called a Type II error. In Table 9.5 the sample statistics of Table 9.1 are used as parameter values.

Table 9.5.: C_{pk} Values for the Four Suppliers

Supplier (i)	1	2	3	4
Sample Size (n_i)	50	75	70	75
Mean (μ_i)	2.7048	2.7019	2.6979	2.6972
Standard Deviation (σ_i)	0.0034	0.0054	0.0046	0.0038
$\hat{C}_{pk}^{(i)}$	1.5392	1.1273	1.3333	1.5526

It is clear from Table 9.5 that the C_{pk} parameters values are all different. To get an estimate of the Type II error 10 000 data sets were generated with sample sizes as shown in Table 9.5. For each dataset $T_{\alpha}^{(2)}$ was calculated from 10,000 Monte Carlo simulations. The Type II error was estimated by observing the number of times that H_0 was accepted, i.e., the number of times that all six credibility intervals contain zero. The process was repeated 5 times and the following estimates of the Type II error were obtained: 0.4240, 0.4163, 0.4212 and 0.4182. The average Type II error estimate is therefore 0.42184 and is the result of 50,000 datasets. The power of the Bayesian procedure $= 1 - 0.42184 = 0.57816$ for this example.

9.8. Posterior Distributions of C_{pl} and C_{pu}

It might be of interest to also look at the posterior distributions of $C_{pl} = \frac{\mu - l}{3\sigma}$ and $C_{pu} = \frac{u - \mu}{3\sigma}$. The posterior distribution of C_{pl} is given in Equation (9.6) and can be used for illustration purposes. A much easier way to obtain the posterior distribution is to simulate a large number of conditional posterior distributions. The average of these conditional distributions is then the unconditional posterior distribution of C_{pl}. This procedure is called the Rao-Blackwell method.

In Figures 9.3 and 9.4 the posterior distributions of C_{pl} and C_{pu} are displayed.

Figure 9.3.: Posterior Distributions of C_{pl}

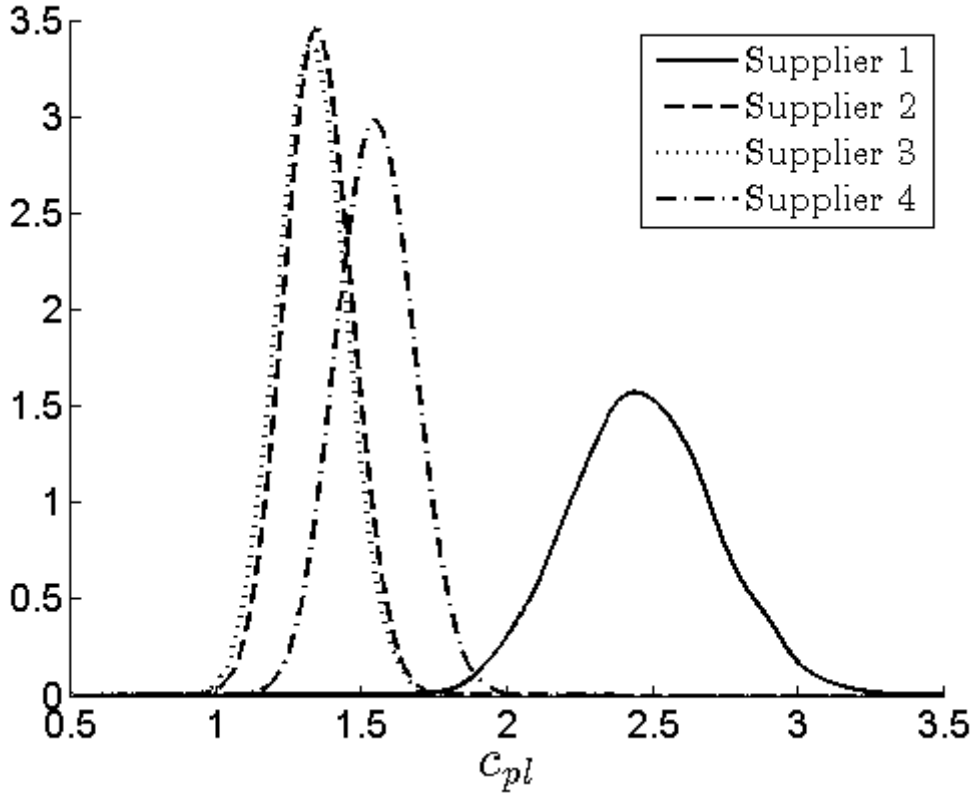
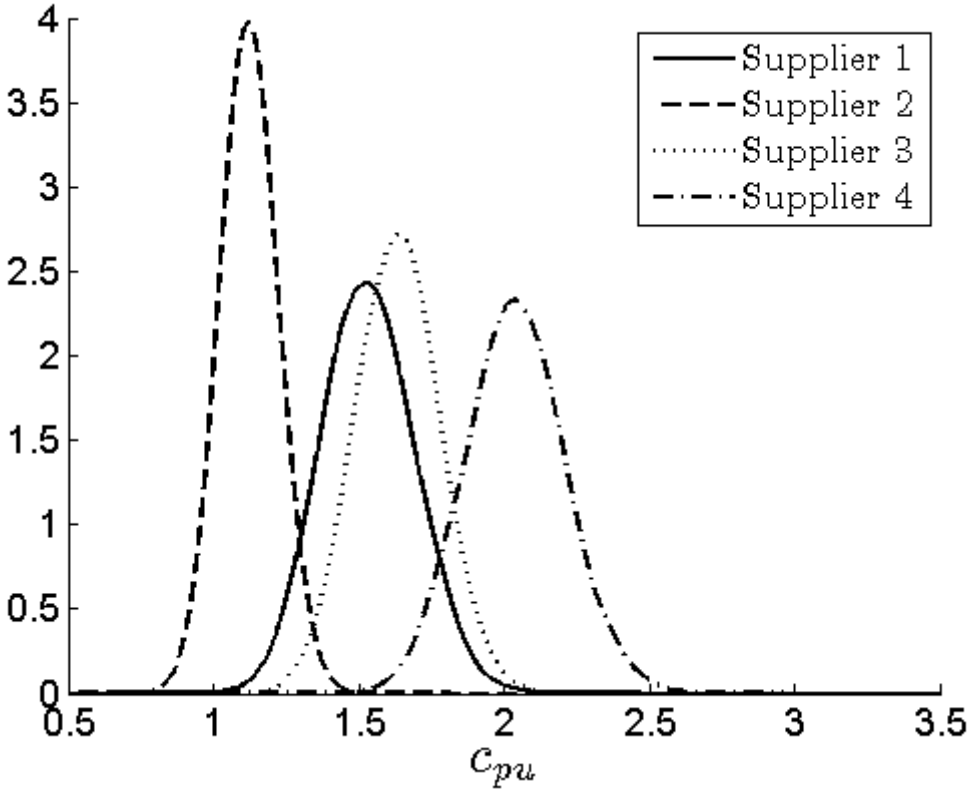


Figure 9.4.: Posterior Distributions of C_{pu}



In Table 9.6 the posterior means of C_{pl} and C_{pu} are given for the four suppliers and in Table 9.7 the 95% credibility intervals for the differences between suppliers are given using Ganesh method.

Table 9.6.: Posterior Means of C_{pl} and C_{pu}

Supplier	1	2	3	4
C_{pl}	2.4920	1.3615	1.3377	1.5585
C_{pu}	1.5460	1.1303	1.6431	2.0521

Table 9.7.: 95% Credibility Intervals for Differences between Suppliers

	C_{pl}	C_{pu}
Supplier 1 - Supplier 2	**(0.4910;1.7700)**	(-0.1281;0.9595)
Supplier 1 - Supplier 3	**(0.5148;1.7938)**	(-0.6408;0.4467)
Supplier 1 - Supplier 4	**(0.2940;1.5729)**	(-1.0498;0.0377)
Supplier 2 - Supplier 3	(-0.6157;0.6633)	(-1.0565;0.0310)
Supplier 2 - Supplier 4	(-0.8365;0.4424)	**(-1.4655;-0.3780)**
Supplier 3 - Supplier 4	(-0.8603;0.4187)	(-0.9527;0.1348)

According to the C_{pl} credibility interval Supplier 1 is significantly different from Suppliers 2, 3 and 4. The other suppliers do not differ significantly from each other. Inspection of the C_{pu} intervals shows that there is a significant difference between Suppliers 2 and 4.

9.9. The Predictive Distribution of a Future Sample Capability Index, $\hat{C}_{pk}^{(f)}$

To obtain a Bayesian control chart for the capability index C_{pk} the predictive distribution must first be derived.

Consider a future sample of m observations from the $N(\mu, \sigma^2)$ population, $y_{1f}, y_{2f}, \ldots, y_{mf}$. The future sample mean is defined as $\bar{y}_f = \frac{1}{m}\sum_{j=1}^{m} y_{jf}$ and a future sample variance by $s_f^2 = \frac{1}{m-1}\sum_{j=1}^{m}(y_{jf} - \bar{y}_f)^2$. A future sample capability index is therefore defined as

$$\hat{C}_{pk}^{(f)} = min\left(\hat{C}_{pu}^{(f)}, \hat{C}_{pl}^{(f)}\right)$$

where

$$\hat{C}_{pu}^{(f)} = \frac{u - \bar{y}_f}{3s_f}$$

and

$$\hat{C}_{pl}^{(f)} = \frac{\bar{y}_f - l}{3s_f}.$$

By using the results given in Smit and Chakraborti (2009) or by using similar theoretical derivations as in Theorem 9.2 it can be shown that the conditional predictive distribution of $\hat{C} = \hat{C}_{pk}^{(f)}$ is given by

$$f\left(\hat{C}|\mu,\sigma^2\right) = 3\sqrt{\frac{m}{m-1}}\int_0^{\frac{m-1}{C^2}C_p^2}\left\{\phi\left(3\sqrt{m}\left[C-2C_p+\sqrt{\frac{x}{m-1}}\hat{C}\right]\right)+\phi\left(3\sqrt{m}\left[C-\sqrt{\frac{x}{m-1}}\hat{C}\right]\right)\right\}\sqrt{x}f\left(x\right)dx \tag{9.9}$$

where

$$f\left(x\right) = \frac{1}{2^{\frac{m-1}{2}}\Gamma\left(\frac{m-1}{2}\right)}x^{\frac{1}{2}(m-1)-1}\exp\left(-\frac{x}{2}\right),$$

$$\phi\left(z\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right),$$

$$C = C_{pk} = min\left(\frac{u-\mu}{3\sigma},\frac{\mu-l}{3\sigma}\right)$$

and

$$C_p = \frac{u-l}{6\sigma}.$$

The unconditional predictive distribution $f\left(\hat{C}|data\right)$ can be obtained in the following way:

i. Simulate σ^2 and μ from their joint posterior distribution and calculate C and C_p. Since $\sigma^2|data \sim \frac{(n-1)S^2}{\chi^2_{n-1}}$ and $\mu|\sigma^2, data \sim N\left(\bar{y},\frac{\sigma^2}{n}\right)$, μ and σ^2 can easily be simulated. Let us call these simulated values μ_1 and σ_1^2.

ii. Substitute μ_1 and σ_1^2 in Equation (9.9) and do the numerical integration to obtain $f\left(\hat{C}|\mu_1,\sigma_1^2\right)$.

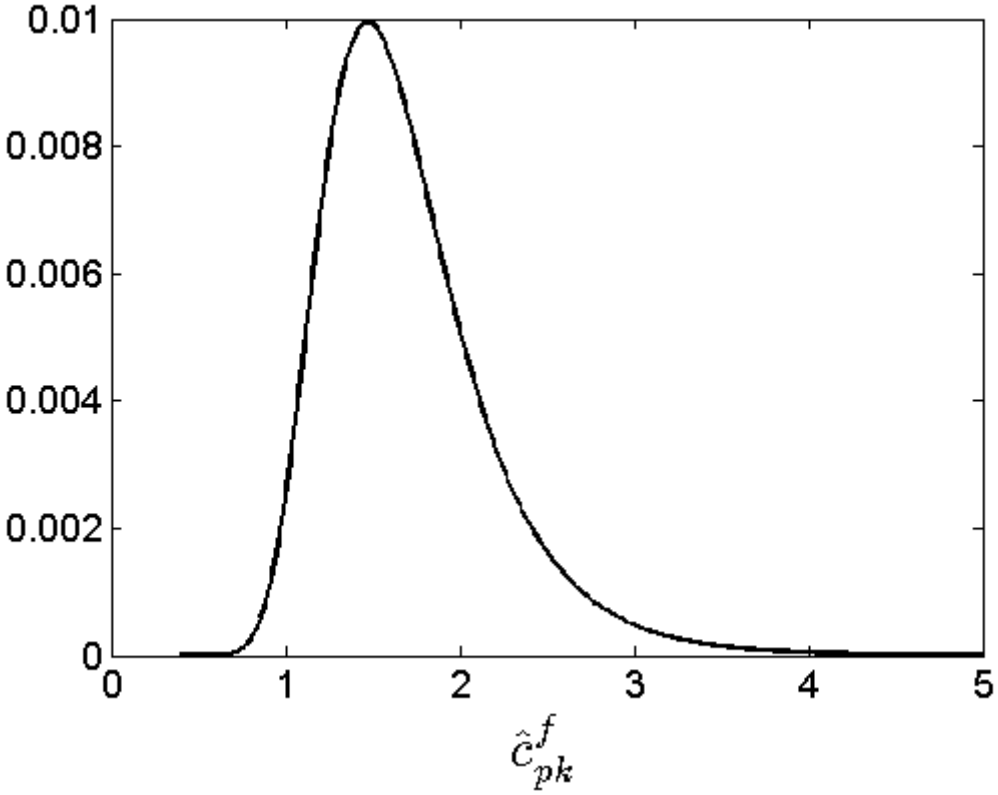
iii. Repeat (i) and (ii) l times to get $f\left(\hat{C}|\mu_1,\sigma_1^2\right), f\left(\hat{C}|\mu_2,\sigma_2^2\right),\ldots,f\left(\hat{C}|\mu_l,\sigma_l^2\right)$. The unconditional predictive distribution $f\left(\hat{C}|data\right)$ is the average of the

conditional predictive distributions (Rao-Blackwell method).

9.9.1. Example

Consider the following sample values: $n = 75$, $\bar{y} = 2.6972$ and $s = 0.0038$. These sample values are the statistics for Supplier 4. In Figure 9.5 the predictive distribution of $\hat{C} = \hat{C}_{pk}^{(f)}$ for $m = 10$ future observations are given.

Figure 9.5.: $f\left(\hat{C}_{pk}^{(f)}|data\right)$



$Mean\left(\hat{C}_{pk}^{(f)}\right) = 1.6870;\ Median\left(\hat{C}_{pk}^{(f)}\right) = 1.598;\ Mode\left(\hat{C}_{pk}^{(f)}\right) = 1.456;\ Var\left(\hat{C}_{pk}^{(f)}\right) = 0.2432$

$95\%\ Equal-tail\ interval = (0.9936; 2.8954),\ length = 1.9018$

$95\%\ HPD\ interval = (0.8790; 2.6360),\ length = 1.7570$

$p\left(\hat{C}_{pk}^{(f)} > 3.923\right) = 0.0027$

$p\left(\hat{C}_{pk}^{(f)} > 4.263\right) = 0.00135$

$p\left(\hat{C}_{pk}^{(f)} < 0.7905\right) = 0.00135$

9.10. Distribution of the Run-length and Average Run-length

Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the run-length and average run-length. From Figure 9.5 it follows that for a 99.73% two-sided control chart the lower control limit is $LCL = 0.7905$ and the upper control limit is $UCL = 4.263$. If a future capability index is smaller than 0.7905 or larger than 4.263, it falls in the rejection region and it is said that the control chart signals. The run-length is defined as the number of future $C_{pk}^{(f)}$ indices (r) until the control chart signals for the first time (Note that r does not include that $C_{pk}^{(f)}$ index when the control chart signals). Given μ and σ^2 and a stable Phase I process, the distribution of the run-length r is Geometric with parameter

$$\psi\left(\mu, \sigma^2\right) = \int_{R(\beta)} f\left(\hat{C}_{pk}^{(f)} | \mu, \sigma^2\right) d\hat{C}_{pk}^{(f)}$$

where $f\left(\hat{C}_{pk}^{(f)} | \mu, \sigma^2\right)$ is defined in Equation (9.9), i.e., the distribution of $\hat{C}_{pk}^{(f)}$ given that μ and σ^2 are known and $R(\beta)$ represents those values of $C_{pk}^{(f)}$ that are smaller than LCL and larger than UCL. The values of μ and σ^2 are however unknown and the uncertainty of these parameters are described by the joint posterior distribution $p(\mu, \sigma^2 | data) = p(\mu | \sigma^2, data)\, p(\sigma^2 | data)$ (Equations (9.2) and (9.3)).

By simulating μ and σ^2 from $p(\mu, \sigma^2 | data)$ the probability density function of $f\left(\hat{C}_{pk}^{(f)} | \mu, \sigma^2\right)$ as well as the parameter $\psi(\mu, \sigma^2)$ can be obtained. This must be done for each future sample. In other words, for each future sample μ and σ^2 must first be simulated from $p(\mu, \sigma^2 | data)$ and then $\psi(\mu, \sigma^2)$ calculated. Therefore by simulating all possible combinations of μ and σ^2 from their joint posterior distribution a large number of $\psi(\mu, \sigma^2)$ values can be obtained. Also a large number of Geometric distributions, i.e., a large number of run-length distributions each with a different parameter value $(\psi(\mu_1, \sigma_1^2), \psi(\mu_2, \sigma_2^2), \ldots, \psi(\mu_l, \sigma_l^2))$ can be obtained.

As mentioned the run-length r for given μ and σ^2 is geometrically distributed with mean

$$E\left(r | \mu, \sigma^2\right) = \frac{1 - \psi(\mu, \sigma^2)}{\psi(\mu, \sigma^2)}$$

and variance

$$Var\left(r|\mu,\sigma^2\right) = \frac{1-\psi\left(\mu,\sigma^2\right)}{\psi^2\left(\mu,\sigma^2\right)}.$$

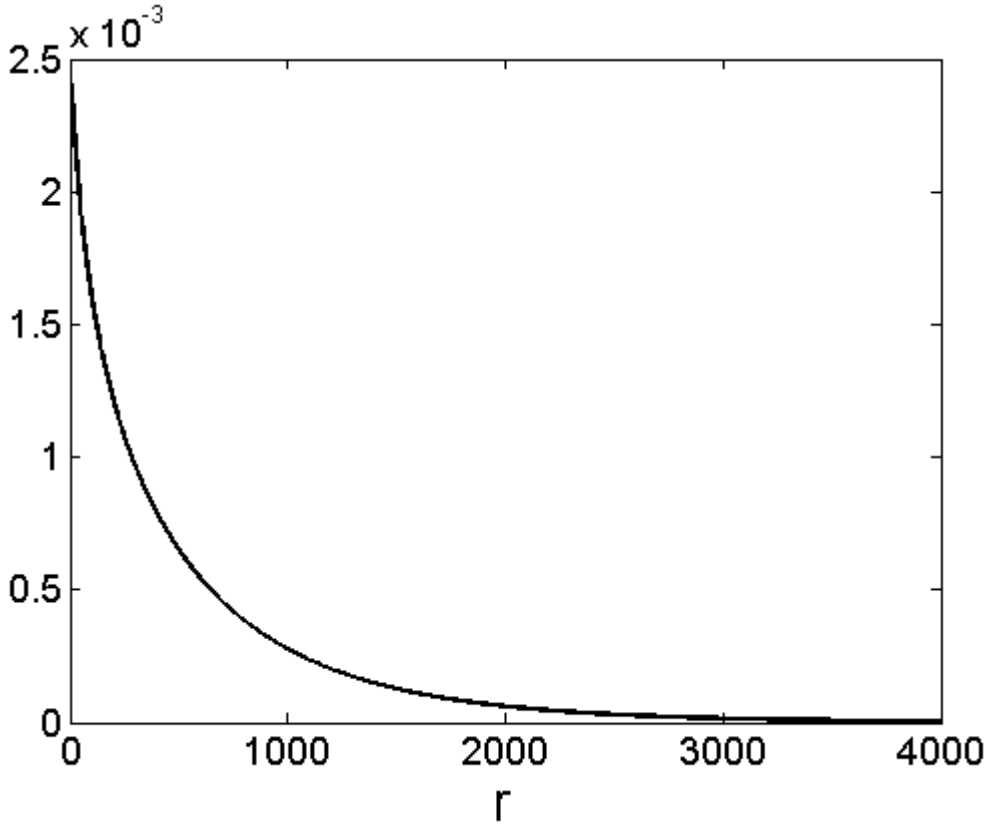
The unconditional moments $E\left(r|data\right)$, $E\left(r^2|data\right)$ and $Var\left(r|data\right)$ can therefore be obtained by simulation or numerical integration. For further details see Menzefricke (2002, 2007, 2010b,a) .

The mean of the predictive distribution of the run-length for the 99.73% two-sided control limits is $E\left(r|data\right) = 482.263$ somewhat larger than the 370 that one would have expected if $\beta = 0.0027$. The median on the other hand is less than 370, $Median\left(r\right) = 303.01$. For the 99.73% one-sided control chart $E\left(r|data\right) = 555.174$ and $Median\left(r\right) = 294.31$.

See Figures 9.6, 9.7, 9.8 and 9.9 for $m = 10$ future observations.

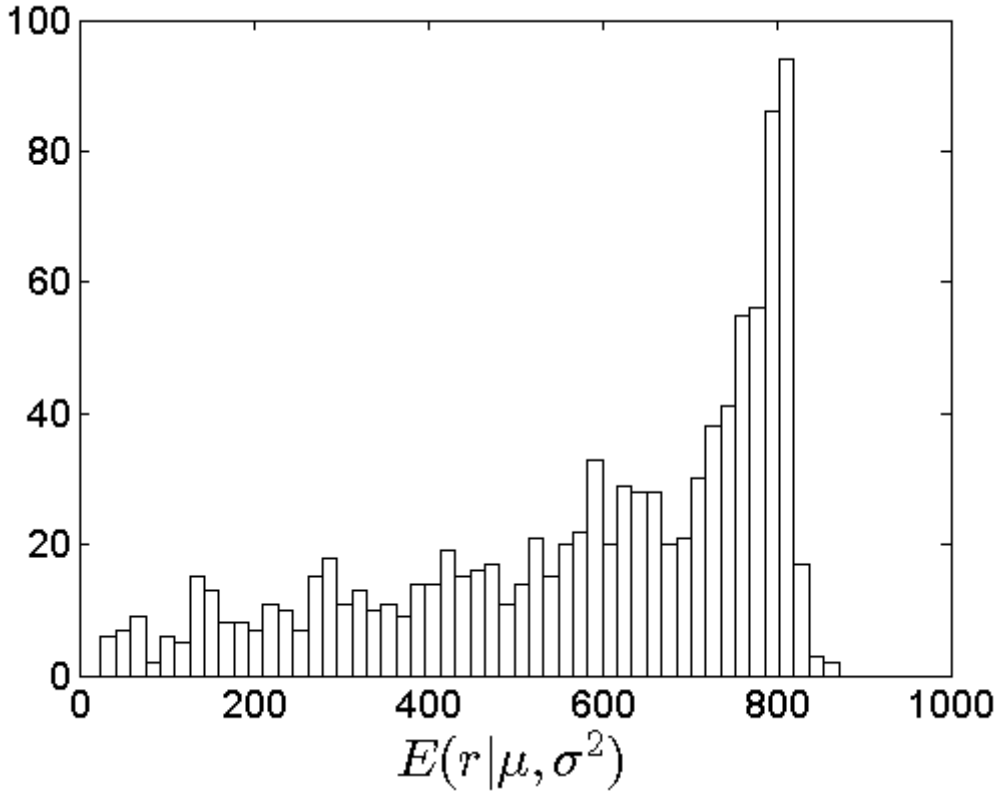
Define $\tilde{\psi}\left(\mu,\sigma^2\right) = \frac{1}{l}\sum_{i=1}^{l}\psi\left(\mu_i,\sigma_i^2\right)$. From Menzefricke (2002) it follows that if $l \to \infty$ then $\tilde{\psi} \to \beta$ and the harmonic mean of $r = \frac{1}{\beta}$. For $\beta = 0.0027$, the harmonic mean $= (0.0027)^{-1} = 370$.

Figure 9.6.: Distribution of Run-length - Two-sided Interval $\beta = 0.0027$



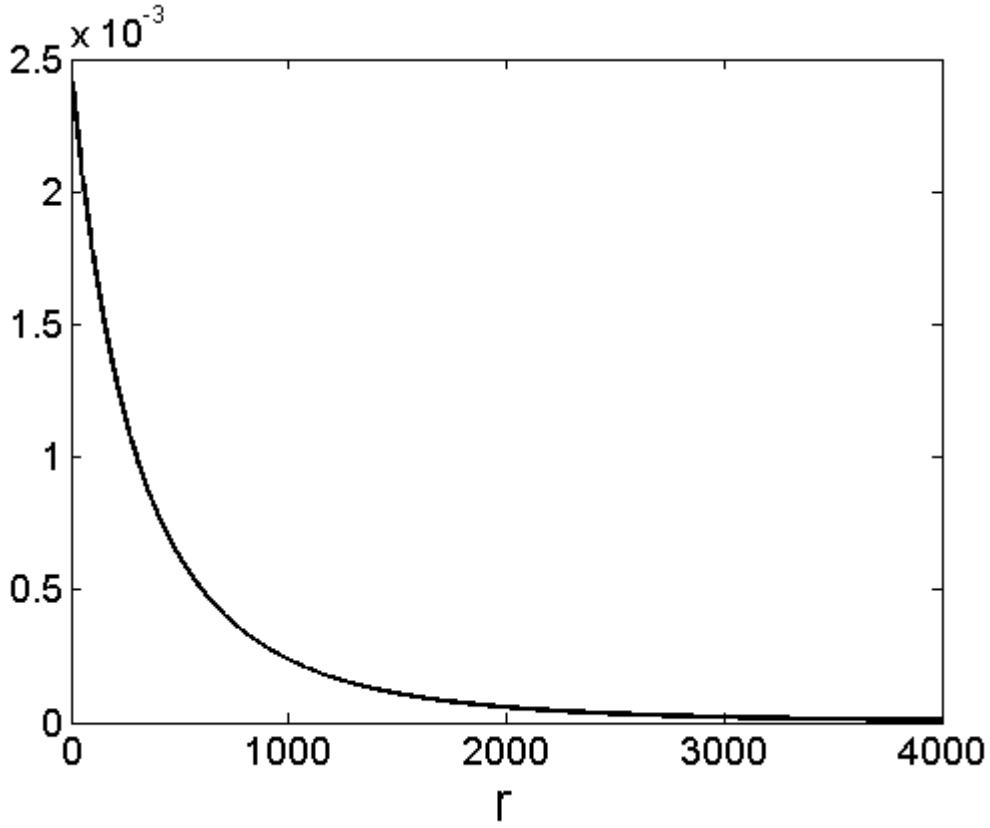
$Mean\,(r) = 482.263;\ Median\,(r) = 303.01;\ Var\,(r) = 2.8886 \times 10^5$
$95\%\ HPD\ interval = (0; 1554.5)$

Figure 9.7.: Distribution of Expected Run-length - Two-sided Interval $\beta = 0.0027$



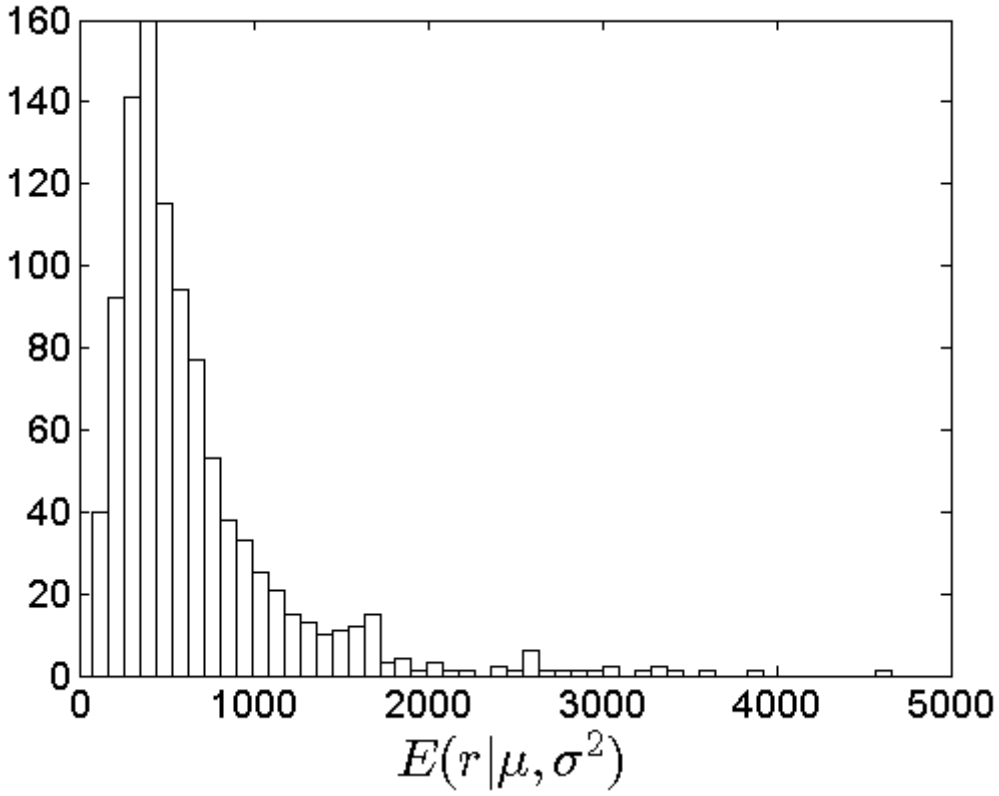
$Mean = 483.263;\ Median = 527.972,\ Var = 2.7907 \times 10^4$
$95\%\ HPD\ interval = (141.993; 689.429)$

Figure 9.8.: Distribution of Run-length - One-sided Interval $\beta = 0.0027$



$Mean\,(r) = 555.174;\ Median\,(r) = 294.31;\ Var\,(r) = 6.1818 \times 10^5$
$95\%\ HPD\ interval = (0; 1971.1)$

Figure 9.9.: Distribution of Expected Run-length - One-sided Interval $\beta = 0.0027$



$Mean = 563.578;\ Median = 436.528,\ Var = 2.0013 \times 10^5$
$95\%\ HPD\ interval = (0; 1384.5)$

9.11. Conclusion

This chapter developed a Bayesian method to analyze C_{pl}, C_{pu} and C_{pk}. Multiple testing strategies have been implemented on data representing four processes from four suppliers that produce piston rings for automobile engines studied by Chou (1994). The results have been compared to the frequentist results and it was shown that the Bayesian procedure proposed by Ganesh (2009) is somewhat more conservative. Control charts were further developed using the Bayesian approach for C_{pk}.

Mathematical Appendix to Chapter 9

Proof of Theorem 9.1

Since

$$\mu|\sigma^2, \underline{y} \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

and

$$k = \frac{vS^2}{\sigma^2} \sim \chi^2_v$$

it follows that

$$t|\tilde{t}, k \sim N\left(a\sqrt{k}, \frac{1}{9n}\right)$$

where

$$a = \frac{\tilde{t}}{\sqrt{v}}.$$

Therefore

$$\begin{aligned} p\left(t|\tilde{t}\right) &= \int_0^\infty f\left(t|\tilde{t}, k\right) f(k)\, dk \\ &= \frac{3\sqrt{n}}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)\sqrt{2\pi}} \int_0^\infty \exp\left[-\frac{9n}{2}\left(t - a\sqrt{k}\right)^2\right] k^{\frac{v}{2}-1} \exp\left[-\frac{k}{2}\right] dk \\ &= \frac{3\sqrt{n}\exp\left(-\frac{9nt^2}{2}\right)}{2^{\frac{v}{2}}\sqrt{2\pi}\Gamma\left(\frac{v}{2}\right)} \int_0^\infty k^{\frac{v}{2}-1} \left(\sum_{j=0}^\infty \frac{\left(9nta\sqrt{k}\right)^j}{j!} \exp\left[-\frac{k}{2}\left(1 + 9na^2\right)\right]\right) dk. \end{aligned}$$

Since

$$\int_0^\infty k^{\frac{1}{2}(v+j)-1} \exp\left[-\frac{k}{2}\left(1+9na^2\right)\right] dk = \frac{2^{\frac{1}{2}(v+j)}\Gamma\left(\frac{v+j}{2}\right)}{\left(1+9na^2\right)^{\frac{1}{2}(v+j)}}$$

and substituting $a = \frac{\tilde{t}}{\sqrt{v}}$, the posterior distribution of $t = \frac{\mu - l}{3\sigma} = C_{pl}$ follows as

$$p\left(t|\tilde{t}\right) = \frac{3\sqrt{n}\exp\left(-\frac{9nt^2}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{2\pi}} \sum_{j=0}^{\infty} \left(\frac{9nt\tilde{t}}{\sqrt{v}}\right)^j \frac{1}{j!} \frac{\Gamma\left(\frac{v+j}{2}\right) 2^{\frac{1}{2}j}}{\left(1+\frac{9n}{v}\tilde{t}^2\right)^{\frac{1}{2}(v+j)}} \quad -\infty < t < \infty.$$

Proof of Theorem 9.2

The C_{pk} index can also be written as

$$C = C_{pk} = \frac{u - l - 2\left|\mu - M\right|}{6\sigma}$$

where

$$M = \frac{u+l}{2}.$$

Since

$$\mu|\sigma^2, \underline{y} \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right),$$

it follows that

$$\mu - M \sim N\left(\zeta, \frac{\sigma^2}{n}\right)$$

where

$$\zeta = \bar{y} - M.$$

Let

$$w = \left|\mu - M\right|,$$

then

$$p\left(w|\sigma^2,\underline{y}\right)=\frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{n}{2}\frac{(w-\zeta)^2}{\sigma^2}\right\}+\frac{n}{\sigma\sqrt{2\pi}}\exp\left(-\frac{n}{2}\frac{(w+\zeta)^2}{\sigma^2}\right)$$

(See Kotz and Johnson (1993, page 26)).

Now $C=b-\tilde{a}w$, where $\tilde{a}=\frac{1}{3\sigma}$ and $b=C_p=\frac{u-l}{6\sigma}$.

Also $w=-\left(C-b\right)\frac{1}{\tilde{a}}$ and $\left|\frac{dw}{dc}\right|=\frac{1}{\tilde{a}}$.

From this it follows that

$$p\left(C|\sigma^2,\underline{y}\right)=\frac{\sqrt{n}}{\tilde{a}\sigma\sqrt{2\pi}}\left\{\exp\left(-\frac{n}{2\tilde{a}^2\sigma^2}\left[C-b+\tilde{a}\zeta\right]^2\right)+\exp\left(-\frac{n}{2\tilde{a}^2\sigma^2}\left[C-b-\tilde{a}\zeta\right]^2\right)\right\}\quad C<\tilde{b}<\frac{S}{\sigma}$$

where $\tilde{b}=\hat{C}_p=\frac{u-l}{6s}$.

Substituting for $\tilde{a}$, b and ζ and making use of the fact that $k=\frac{vS^2}{\sigma^2}\sim\chi^2_v$ it follows that

$$p\left(C|k,\underline{y}\right)=\frac{3\sqrt{n}}{\sqrt{2\pi}}\left\{\exp\left(-\frac{9n}{2}\left[C-t^*\sqrt{\frac{k}{v}}\right]^2\right)+\exp\left(-\frac{9n}{2}\left[C-\tilde{t}\sqrt{\frac{k}{v}}\right]^2\right)\right\}\quad C<\tilde{b}\sqrt{\frac{k}{v}}.$$

Therefore

$$p\left(C|\underline{y}\right)=\frac{3\sqrt{n}}{\sqrt{2\pi}}\int_{\frac{C^2v}{\tilde{b}^2}}^{\infty}\left\{\exp\left(-\frac{9n}{2}\left[C-t^*\sqrt{\frac{k}{v}}\right]^2\right)+\exp\left(-\frac{9n}{2}\left[C-\tilde{t}\sqrt{\frac{k}{v}}\right]^2\right)\right\}$$

$$\times\frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)}k^{\frac{v}{2}-1}\exp\left(-\frac{k}{2}\right)dk.$$

10. Conclusion

This chapter will conclude with a summary of the conclusions of the chapters in this thesis, possible shortcomings and possible future research in this area.

10.1. Summary and Conclusions

This thesis focused on Bayesian methods to implement Statistical Process Control in two phases (Phase I and Phase II), by looking at the following non-informative priors: reference priors, probability matching priors and Jeffreys' priors. Non-informative priors are used when no other information is available.

Chapter 2 and **Chapter 3** developed a Bayesian control chart for monitoring a common coefficient of variation and common standardized mean (respectively) across a range of sample values. In the Bayesian approach prior knowledge about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution. By using the posterior distribution the predictive distribution of a future coefficient of variation and standardized mean were obtained.

Determination of reasonable non-informative priors in multi-parameter problems is not an easy task. The Jeffreys' prior for example can have a bad effect on the posterior distribution. Reference and probability matching priors are therefore derived for a common coefficient of variation and standardized mean across a range of sample values. The theory and results are applied to a real problem of patients undergoing organ transplantation for which Cyclosporine is administered. This problem is discussed in detail by Kang, Lee, Seong, and Hawkins (2007). The 99.73% equal tail prediction interval of a future coefficient of variation is effectively identical to the lower and upper control chart limits calculated by Kang, Lee, Seong, and Hawkins

(2007). A simulation study shows that the 95% Bayesian confidence intervals for γ (coefficient of variation) and δ (standardized mean) has the correct frequentist coverage.

The example illustrated the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, prediction intervals and "run-lengths".

In **Chapter 4** Phase I and Phase II control chart limits have been constructed for the variance and generalized variance using Bayesian methodology. In this chapter we have seen that due to Monte Carlo simulation the construction of control chart limits using the Bayesian paradigm are handled with ease. Bayesian methods allow the use of any prior to construct control limits without any difficulty. It has been shown that the uncertainty in unknown parameters are handled with ease in using the predictive distribution in the determination of control chart limits. It has also been shown that an increase in number of samples m and the sample size n leads to a convergence in the "run-length" towards the expected value of 370 at $\beta = 0.0027$.

Chapter 5 developed a Bayesian control chart for monitoring an upper one-sided tolerance limit across a range of sample values.

Reference and probability matching priors are derived for the pth quantile of a normal distribution. The theory and results have been applied to air-lead level data analyzed by Krishnamoorthy and Mathew (2009) to illustrate the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, prediction intervals and run lengths.

The Bayesian procedure has also been extended to control charts of one-sided tolerance limits for a distribution of the difference between two independent normal variables.

Chapter 6 and **Chapter 7** developed a Bayesian control chart for monitoring the location parameter, scale parameter and upper tolerance limit of a two-parameter exponential distribution. By using a Bayes approach the posterior predictive distributions of $\hat{\mu}_f$ (future location observation), $\hat{\theta}_f$ (future scale observation) and U_f (future tolerance limit) were obtained.

The theory and results described in these chapters have been applied to the failure mileages for military carriers analyzed by Grubbs (1971) and Krishnamoorthy and Mathew (2009). The example illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions and "run-lengths" for $\hat{\mu}_f$, $\hat{\theta}_f$ and U_f.

Results for $0 < \mu < \infty$ as presented in Chapter 6 were compared against $-\infty < \mu < \infty$ in Chapter 7. It has been shown by changing the range of μ only the results of $\hat{\mu}_f$ are influenced. The results of $\hat{\theta}_f$ and U_f did not change due to a change in the range of μ.

Chapter 8 developed a Bayesian approach to model the piecewise exponential model using multiple priors and compared the results of these multiple priors. A Bayesian control chart for monitoring X_f (future failure time) have been implemented. Two models, one where all μ's (scale parameters) across different systems are the same and another where μ is different for each system has been implemented and compared using Bayes factors. It has been shown that the model with less parameters are preferable.

Chapter 9 developed a Bayesian method to analyze C_{pl}, C_{pu} and C_{pk}. Multiple testing strategies have been implemented on data representing four processes from four suppliers that produce piston rings for automobile engines studied by Chou (1994). The results have been compared to the frequentist results and it was shown that Ganesh (2009) Bayesian procedure method is somewhat more conservative. Control charts were further developed using the Bayesian approach for C_{pk}.

10.2. Possible Future Research

For future research it will also be of interest to consider the random effects model as the underlying model for the observed data. Let Y_{ij} denotes the jth measurement in the ith sample (group or level) assumed to follow the one-way random effects model:

$$Y_{ij} = \mu + r_i + \epsilon_{ij} \quad i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, n$$

where μ is an unknown general mean, r_i's represent random (sample) effects and ϵ_{ij}'s represent error terms. It is assumed that the r_i's and ϵ_{ij}'s are all independent having the distributions $\epsilon_{ij} \sim N(0, \sigma_1^2)$ and $r_i \sim N(0, \sigma_2^2)$.

It can easily be shown that the Jeffreys' independent prior for the random effects (variance component) model is

$$p\left(\mu, \sigma_1^2, \sigma_2^2\right) \propto \frac{1}{\sigma_1^2\left(\sigma_1^2 + n\sigma_2^2\right)} = \frac{1}{\sigma_1^2 \sigma_{12}^2}.$$

By multiplying the prior distribution with the likelihood function (the distribution of the data) the joint posterior distribution can be obtained. From the joint posterior distribution it follows that

$$\begin{aligned} \mu|data, \sigma_1^2, \sigma_2^2 &\sim N\left(\bar{Y}_{..}, \tfrac{\sigma_1^2+n\sigma_2^2}{nm}\right) \\ &= N\left(\bar{Y}_{..}, \tfrac{\sigma_{12}^2}{nm}\right) \end{aligned}$$

and

$$p\left(\sigma_1^2, \sigma_{12}^2|data\right) \propto \left(\sigma_1^2\right)^{-\frac{1}{2}(v_1+2)} \exp\left\{-\frac{1}{2}\frac{v_1 m_1}{\sigma_1^2}\right\} \left(\sigma_{12}^2\right)^{-\frac{1}{2}(v_2+2)} \exp\left\{-\frac{1}{2}\frac{v_2 m_2}{\sigma_{12}^2}\right\} \quad \sigma_1^2 > 0, \sigma_{12}^2 > 0$$

and $\sigma_{12}^2 > \sigma_1^2$

where

$\bar{Y}_{..}$ = the overall sample mean; $v_1 = m\,(n-1)$; $v_2 = m-1$; $v_1 m_1$ =residual sum of squares and $v_2 m_2$ =between samples sum of squares.

Let

$$\bar{Y}_f = \frac{1}{n}\sum_{j=1}^{n} Y_{jf}$$

be the sample mean of a future sample of n observations (i.e., the mean of a sample of n observations in Phase II). Since the sample size in Phase II is the same as the sample sizes in Phase I it follows that

$$\bar{Y}_f|\mu, \sigma_{12}^2, data \sim N\left(\mu, \frac{\sigma_{12}^2}{n}\right)$$

and

$$\bar{Y}_f|\sigma_{12}^2, data \sim N\left(\bar{Y}_{..}, \sigma_{12}^2\left(\frac{1}{n}+\frac{1}{nm}\right)\right).$$

Let

$$\frac{1}{n} + \frac{1}{nm} = a.$$

Since

$$\frac{v_2 m_2}{\sigma_{12}^2} | data \sim \chi_{v_2}^2$$

it follows that the unconditional predictive distribution of $\bar{Y}_f$ has a t-distribution:

$$\bar{Y}_f | data \sim t_{v_2} \left(\bar{Y}_{..}, \frac{v_2 m_2}{(v_2 - 2)} a \right).$$

The predictive distribution can be used to obtain the control limits in Phase II as well as the run-length and expected run-length.

If the sample size of a future sample is different from n say $\tilde{n}$, then the future sample mean is

$$\tilde{\bar{Y}}_f = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} Y_{jf}$$

and

$$\tilde{\bar{Y}}_f | \sigma_1^2, \sigma_2^2, data \sim N \left(\bar{Y}_{..}, \frac{\sigma_1^2 + \tilde{n}\sigma_2^2}{\tilde{n}} + \frac{\sigma_1^2 + n\sigma_2^2}{nm} \right). \tag{10.1}$$

Although the unconditional predictive distribution of $\tilde{\bar{Y}}_f$ does not follow a t-distribution if $\tilde{n} \neq n$, its distribution can easily be obtained using the following simulation procedure:

i. Simulate ζ from a $\chi_{v_1}^2$ distribution and calculate $\sigma_1^2 = \frac{v_1 m_1}{\zeta}$.

ii. Simulate δ from a $\chi_{v_2}^2$ distribution and calculate $\sigma_{12}^2 = \frac{v_2 m_2}{\delta}$.

iii. Calculate $\sigma_2^2 = \frac{\sigma_{12}^2 - \sigma_1^2}{n}$. It must be mentioned that even though σ_1^2 and σ_{12}^2 will always be positive, the estimated value of σ_2^2 can be negative. If a negative value of σ_2^2 is obtained then one possibility is to discard the negative value as

well as the corresponding value of σ_1^2. A better strategy in our opinion is to keep the simulated σ_1^2 value and take σ_2^2 as zero.

iv. Substitute the simulated values σ_1^2 and σ_2^2 values in Equation (10.1) and draw the normal density function.

v. Repeat steps (i) to (iv) a large number of times (say l times). The average of the l conditional densities gives $f\left(\tilde{\tilde{Y}}_f|data\right)$ the unconditional predictive distribution.

In the unbalanced case if the sample sizes are n_i $(i = 1, 2, \ldots, m)$ the problem becomes more difficult. In this case it is better to work with the parameters μ, σ_1^2 and $\tilde{r}$ where $\tilde{r} = \frac{\sigma_2^2}{\sigma_1^2}$. It is possible to derive reference and probability-matching priors for this parameter. The joint posterior distribution can be written in the form

$$p\left(\mu, \sigma_1^2, \tilde{r}|data\right) = p\left(\mu|\sigma_1^2, \tilde{r}, data\right) p\left(\sigma_1^2|\tilde{r}, data\right) p\left(\tilde{r}|data\right).$$

By using Monte Carlo simulation the unconditional predictive density function of a future sample mean can be obtained. As in the balanced data case the predictive distribution can be used to obtain control limits in Phase II.

The theory and results can easily be extended to the two-factor nested random effects model and cross classification models. The two-factor nested random effects model is given by

$$Y_{ijt} = \mu + d_i + p_{ij} + \epsilon_{ijt}$$

where $i = 1, 2, \ldots, b$; $j = 1, 2, \ldots, k$; $t = 1, 2, \ldots r$. Y_{ijt} represents the observations (for example percentage increase in length of synthetic yarn before breaking), μ is a common location parameter (the grand mean), d_i, p_{ij} and ϵ_{ijt} are the different kinds of random effects (for example days, packages within days and residual). It is further assumed that the random effects (d_i, p_{ij} and ϵ_{ijt}) are all independent and that $d_i \sim N\left(0, \sigma_d^2\right)$, $p_{ij} \sim N\left(0, \sigma_p^2\right)$ and $\epsilon_{ijt} \sim N\left(0, \sigma_\epsilon^2\right)$.

The non-informative prior that will be used is Jeffreys' independent prior (see also Box and Tiao (1973))

$$p\left(\mu, \sigma_\epsilon^2, \sigma_p^2, \sigma_d^2\right) \propto \sigma_\epsilon^{-2}\left(\sigma_\epsilon^2 + r\sigma_p^2\right)^{-1}\left(\sigma_\epsilon^2 + r\sigma_p^2 + kr\sigma_d^2\right)^{-1}.$$

It can be shown that the predictive density of the average of k^* future packages with r^* samples per package from a new (future) or unknown day given the variance components is normal with mean

$$E\left(\bar{Y}_f|data, \sigma_d^2, \sigma_p^2, \sigma_\epsilon^2\right) = \bar{Y}_{...}$$

and variance

$$Var\left(\bar{Y}_f|data, \sigma_d^2, \sigma_p^2, \sigma_\epsilon^2\right) = \frac{\sigma_\epsilon^2 + r^*\sigma_p^2 + k^*r^*\sigma_d^2}{k^*r^*} + \frac{\sigma_\epsilon^2 + r\sigma_p^2 + kr\sigma_d^2}{bkr}.$$

By simulating the variance components from their posterior distributions the unconditional predictive distribution of $\bar{Y}_f$ can be obtained.

By using similar simulation procedures, predictive distributions and control limits can be derived for more complicated linear models.

10.3. Shortcomings of this Thesis

The section on Future Research gives a good indication of the shortcomings of this thesis. The thesis should have started with a chapter on control charts for linear models (random effects models, two-factor nested random effects models and cross classification models). This should have been done firstly before reference and probability-matching priors, predictive distributions and Bayesian control charts were derived for more complicated problems like the coefficient of variation, tolerance intervals, piece-wise exponential models and capability indices.

A. MATLAB Code

A.1. MATLAB Code To Determine Coefficient of Variation from Sampling Distribution

```
%Sampling distribution of cv for given R
clear
tic
R=0.0751;
n=5; f=n-1;
Fw=[];
h=0.001;
for w=0:h:0.2;
c=n./R./((n+f*(w.^2)).^0.5);
Aw=(f^(f/2))*sqrt(n)*abs(w.^(f-1)).*exp(-n*f*(w.^2)./2./(R^2)
./(n+f*(w.^2))).(2^((f-2)/2))/gamma(f/2)/sqrt(2*pi);
syms q
```

```
F=inline((q.^f).*exp((-((q-c).^2))/2));
A=quad(F,10,30);
fw=Aw.*A./((n+f*(w.^2)).^((f+1)/2));
Fw=[Fw fw];
end
Fw=Fw/sum(Fw)/h;
w=0:h:0.2;
MEAN=w*Fw'/sum(Fw)
figure(1)
plot(w,Fw)
grid
CDF=cumsum(Fw)*h;
figure(2)
plot(w,CDF)
grid
toc
```

A.2. MATLAB Code to Simulate Coefficient of Variation from Berger Prior

```
clear
tic
X=[31.7  12.4
   37.7  15.3
```

```
40.6  9.1
50.5  4.6
52    10.5
57.6  6.2
58.3  6.6
58.9  8.4
61.2  8.1
64.3  7
64.5  8.8
65.6  4.1
68    3.7
71.8  6.2
72.1  8.4
78.4  6.8
78.4  4.6
79.5  5.7
83.2  10.5
85.1  4.8
85.6  5.4
86    10.1
87.3  7.9
89.1  10.3
95.4  6.2
101.9 4.8
```

```
105.4 5.6
107.2 2.2
108.2 3.3
112   8.7
112.3 5.7
113.5 9.4
114.3 3.5
116.8 6
117.8 5.7
120.3 5.8
143.7 5.6
148.6 5.5
149.1 3.1
149.9 2
151   4.4
153.6 6.6
172.2 7.2
179.8 7.9
185.3 7.6
192.1 5.3
193.8 5.9
195.1 11
195.2 5.1
195.4 9.4
```

```
196.4 5.6
199.6 6.8
204.4 3.7
207.8 12.4
219   7.6
222.9 4.8
225.1 5.7
227.6 6.5
240.5 3.8
241.1 8.4
252.2 8.3
262.2 5.8
277.9 8.7
278.3 6.2
303.4 8.8
309.7 3.9
323.9 4.1
328.7 4.1
341.2 6.5
347.3 4.9
361.4 8.3
361.5 13.4
361.8 6.1
374.6 5.8
```

```
376.3 2.8
382.3 5.8
401.7 7.3
415.2 15.1
428.8 4.5
442.1 9.9
450.1 7.4
496.5 4.8
499.7 10
504.6 8.4
523.1 5
531.7 8.5
556.4 11.8
571.4 5.9
584.1 8.3
597.6 4.2
606.2 8.2
609   9.7
635.4 5.6
672.2 7.2
695.9 2.7
696.4 10.6
721.3 9.8
752   4.2
```

```
    769.5 9.7
    772.7 9.6
    791.6 2
    799.9 11.4
    948.6 5.2
    971.8 11.1
    991.2 8.8];


n=5;
Sig=X(:,1).*X(:,2)/100;
T=[]; SIG=[];
for k=1:2
    lnD=0;
    for j=1:length(X)
        x=X(j,1);
        s2=(x*X(j,2)/100).^2;
        D2=(n-1)*s2/n+x^2;
        theta=0.02:0.00001:0.2;
        lnA=-n*(1-x^2/D2)/2./(theta.^2);
        lnB=-(n*D2/2).*((1./Sig(j)-x./D2./theta).^2);
        lnd=lnA+lnB;
        lnD=lnD+lnd;
    end
```

```
lnC=-log(theta)-0.5*log(theta+0.5);
lnF=lnC+lnD;
M=max(lnF);
lnF=lnF-M;
ftheta=exp(lnF)/sum(exp(lnF));
%plot(theta,ftheta)
%grid
%pause
ct=cumsum(ftheta);
rt=rand(1,1);
t=theta(min(find(ct>=rt)));
T=[T;t];

s=1:0.001:200;
Sig=[];
for j=1:length(X)
    x=X(j,1);
    s2=(x*X(j,2)/100).^2;
    D2=(n-1)*s2/n+x^2;
    lnfs=-(n+1)*log(s)-(n*D2/2).*((1./s-x./D2./t).^2);
    fsig=exp(lnfs)/sum(exp(lnfs));
    %plot(s,fsig)
    %grid
    %pause
```

```
        cs=cumsum(fsig);
        rc=rand(1,1);
        sig=s(min(find(cs>=rc)));
        Sig=[Sig;sig];
    end
    SIG=[SIG Sig];
end
toc
```

A.3. MATLAB Code to Determine Sampling and Predictive Densities for Coefficient of Variation

```
%SAMPLING AND PREDICTIVE DISTRIBUTION OF CV
clear
tic
%load posterior_data
%load postcv5
%load simul_second
load postr1r2
FFw=[]; RR=[];
k=100;
for i=1:k
```

```
   %i
R=R2(round(rand*length(R2)));
%R=0.0977;  %51;
n=5; f=n-1;
Fw=[];
for w=0:0.001:0.25;
c=n./R./((n+f*(w.^2)).^0.5);
Aw=(f^(f/2))*sqrt(n)*abs(w.^(f-1)).*exp(-n*f*(w.^2)./2./(R^2)
./(n+f*(w.^2))./(2^((f-2)/2))/gamma(f/2)/sqrt(2*pi);
h=0.002;
q=15:h:40;
F=(q.^f).*exp((-((q-c).^2))/2);
If=h*sum(F);
%plot(q,F)
%grid
%pause
fw=Aw.*If./((n+f*(w.^2)).^((f+1)/2));
Fw=[Fw fw];
end
FFw=[FFw;Fw];
RR=[RR;R];
end
w=0:0.001:0.25;
MFw=mean(FFw);
```

```
%MEAN=w*Fw'/sum(Fw)
plot(w,MFw)
grid
toc
```

A.4. MATLAB Code to Run Standardized Mean Simulations

```
clc, clear;

%Original Data;
data.orig(1,:) = 10:10:1050;
data.orig(2,:) = 0.75:0.75:78.75;
data.orig(3,:) = data.orig(2,:).^2;
data.n = 5;

%Set number of simulations;
datasim = 100;
nsim = 100000;

%Simulations
delta = zeros(datasim,nsim);
for i = 1:datasim
    data.sim(1,:) = normrnd(data.orig(1,:),sqrt((data.orig(3,:)/data.n)));
```

```
    data.sim(2,:) = ((data.orig(3,:).*chi2rnd(data.n-1,1,size(data.orig,2))))./(data.n-1)).^(0.5);

    %Calculate delta
    for j = 1:nsim
        delta(i,j) = (sum((data.sim(1,:)./data.sim(2,:)).*sqrt((chi2rnd(data.n-1,1,size(data.orig,2)))
./(data.n-1)) + ((normrnd(0,1,1,size(data.orig,2)))./(sqrt(data.n))))./size(data.orig,2);
    end
    delta(i,:) = sort(delta(i,:));
    statis.dmean(i) = mean(delta(i,:));
    statis.d(i) = (1/0.075) - statis.dmean(i);
    statis.low(i) = delta(i,0.025*nsim);
    statis.high(i) = delta(i,0.975*nsim);
    statis.cover(i) = (statis.low(i) <= (1/0.075))*(statis.high(i) >= (1/0.075));
    statis.right(i) = statis.low(i) > (1/0.075);
    statis.left(i) = statis.high(i) < (1/0.075);
end
coverp = sum(statis.cover)./datasim*100;
avgdiff = mean(statis.d);
rights = sum(statis.right);
lefts = sum(statis.left);
```

A.5. MATLAB Code to Determine Rejection Region for the Variance

```
clc, clear;

%Set values for n and m where m is the number of samples and n the size of
%each sample;
format longg;
n = 5;
m = 10;

%Determine degrees of freedom;
v = n-1;
k = m*v;

%Set alpha value;
alpha = 0.0027;

%Determine F value at k and v degrees of freedom at level alpha;
Fval = finv((1-alpha/2),v,k);
Fval2 = finv((alpha/2),v,k);

%Simulate Chi2 Values;
nsim = 30000;
simval = chi2rnd(k,1,nsim);
```

```
%Determine the test value to determine rejection region for Chi2 with v
%degrees of freedom;
critval = (1/m).*simval.*Fval;
critval2 = (1/m).*simval.*Fval2;

%Determine rejection region;
rejreg = 1-chi2cdf(critval,v)+chi2cdf(critval2,v);
q = 1-rejreg;

meanrl = mean(q./rejreg);
meanrl1 = mean(1./rejreg);
medianrl = median(q./rejreg);
medianrl1 = median(1./rejreg);
stdrl = std(q./rejreg);
stdrl1 = std(1./rejreg);

rl = sort(q./rejreg);
rllow = rl(0.025*nsim);
rlhigh = rl(0.975*nsim);

disp(meanrl)
disp(medianrl)
disp([rllow rlhigh])
```

```
%Determine runlength;
r = 0:1:3500;
rPDF = zeros(size(rejreg,2),size(r,2));
for i = 1:size(rejreg,2)
    rPDF(i,:) = geopdf(r+1,rejreg(i));
end
rpredPDF_ = mean(rPDF);
rpredPDF = rpredPDF_./sum(rpredPDF_);
rpredCDF = cumsum(rpredPDF);

rpredmean = sum(r.*rpredPDF);
r2mom = sum((r.^2).*rpredPDF);
var = r2mom - rpredmean^2;
std = sqrt(var);

medi = find(rpredCDF>=0.5,1,'first');
rmedian = r(medi);

low = find(rpredCDF>=0.025,1,'first');
rlow = r(low);

high = find(rpredCDF>=0.975,1,'first');
rhigh = r(high);
```

A.6. MATLAB Code to Determine Rejection Region for the Generalized Variance

```
function [meanrl,medianrl,rllow,rlhigh,rl] = varrlmp(n,m,p,alpha)
%Initial Values;
nsim = 30000;
rejregsim = 1000000;

%Vector of p's;
vecp = 1:p;

%F-values;
Fconst = ((n-vecp)./(m*(n-1)+1-vecp));

Fs = zeros(p,rejregsim);
for k = vecp
    Fs(k,:) = frnd((n-k),(m*(n-1)+1-k),1,rejregsim);
end
if p ~= 1
    Fcomb = prod(Fs);
else
    Fcomb = Fs;
end
```

```
Fcomb = sort(Fcomb);

Fcrit = Fcomb((1-alpha)*rejregsim);

Fval = prod(Fconst)*Fcrit;

%Simulate Chi2-values to determine the critical value and determine rejectionregion;
chivec = zeros(p,nsim);
rejchivec = zeros(p,rejregsim);
for i = vecp
    chivec(i,:) = 1./chi2rnd((m.*(n-1)+1-i),1,nsim);
    rejchivec(i,:) = chi2rnd(n-i,1,rejregsim);
end
if p ~= 1
    simval = 1./prod(chivec);
    rejchi = prod(rejchivec);
    critval = simval.*Fval;
else
    simval = 1./(chivec);
    rejchi = (rejchivec);
    critval = simval.*Fval;
end

rejreg1 = zeros(1,nsim);
```

```
for j = 1:nsim
    rejreg1(j) = (sum(rejchi>=critval(j))/rejregsim);
end

rejreg = rejreg1(rejreg1>0);

q = 1-rejreg;

meanrl = mean(q./rejreg);
medianrl = median(q./rejreg);

rl = sort(q./rejreg);
rllow = rl(0.025*nsim);
rlhigh = rl(0.975*nsim);
```

A.7. MATLAB Code to Determine Rejection Region of Tolerance Interval

```
%Tolerance Limit Simulation - RunLength - 6SEP2014
clc, clear
%Set Alpha Level;
alpha = 0.1;

%Airlead data
```

```
%data;
data.orig = [200 120 15 7 8 6 48 61 380 80 29 1000 350 1400 110];

%take the logarithm of the data;
data.log = log(data.orig);

%determine the mean and sample standard deviation of log data;
data.logmean = mean(data.log);
data.var = var(data.log);
data.sd = std(data.log);

%determine k1 from the non-central t distribution;
data.size = size(data.orig);
df = data.size(2) - 1;
ncparam = norminv(0.95,0,1)*sqrt(data.size(2));
k1 = nctinv((1-alpha),df,ncparam)/sqrt(data.size(2));

%Upper tolerance limit;
data.utl = data.logmean + k1*data.sd;

%Simulate future values;
nsim = 10000;

x = 1:0.1:16;
```

```
I = find(x==13.7,1,'first');

simmed.var = df*data.var./chi2rnd(df,1,nsim);
simmed.fmean = normrnd(data.logmean,sqrt(simmed.var./data.size(2)));
simmed.qvar = simmed.var./(df+1);

for i = 1:nsim
    simmed.V = chi2rnd(df,1,nsim);
    simmed.qmean = simmed.fmean + k1 * ((sqrt(simmed.var.*simmed.V))/(sqrt(df)));

    distq.pdf = zeros(nsim,size(x,2));

    for j = 1:1000
        distq.pdf(j,:) = normpdf(x,simmed.qmean(j),sqrt(simmed.qvar(j)));
    end;

    distq.finpdf = mean(distq.pdf);
    distq.finpdf1 = distq.finpdf/sum(distq.finpdf);
    distq.cdf = cumsum(distq.finpdf1);

    simmed.p(i) = 1 - distq.cdf(I);

    if simmed.p(i) < 0
```

```
        simmed.p(i) = 0;
    end

end
simmed.rl = (1-simmed.p)./simmed.p;

meanrl = mean(simmed.rl);
medianrl = median(simmed.rl);
sdrl = std(simmed.rl);

hist(simmed.rl,50) ,xlabel('E(r|$\mu$,$\sigma^2$)');
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k');
```

A.8. MATLAB Code to Determine Rejection Region of μ from the Two Parameter Exponential Distribution

```
clc, clear;

%Initial Values;
muhat = 162;
thhat = 835.21;
```

```
xbar = 997.21;
n = 19;
m = 19;
beta = 0.0027;
% LCL = 13.527;
% UCL = 489.52;
LCL = 40.623;
UCL = 352.077;

nsim = 100000;


%Simulate mu;
mu = 0.01:0.01:muhat;
mupdf = (n-1).*((((1/thhat).^(n-1))-((1/xbar).^(n-1)))^-1).*((xbar-mu).^(-n));
mupdfn = mupdf/sum(mupdf);
mucdf = cumsum(mupdfn);

musim = zeros(1,nsim);
for i = 1:nsim
    musim(i) = mu(find(mucdf>rand,1,'first'));
end

%Simulate Theta;
```

```
paralph = n;
parlam = n.*(xbar-musim);
thsim = 1./gamrnd(paralph,1./parlam);

% musim = [149.3 146.1 135.2 102.25 152.9];
% thsim = [856.6 618.0 1047.2 1054.1 704.4];

%Determine psi;
rlprob = zeros(1,nsim);
for k = 1:nsim
    parm = thsim(k)/m;
    c = max(LCL,musim(k));
    rlprob(k) = 1 - exp(m*musim(k)/thsim(k))*(expcdf(UCL,parm)-expcdf(c,parm));
end
rl = (1-rlprob)./rlprob;
rlplot = rl(rl<2500);

hist(rlplot,50),xlabel('E(r|\mu,\theta)','fontsize',16);
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k');
set(gca,'fontsize',14);
save('murl.mat')
```

A.9. MATLAB Code to Determine Rejection Region of θ from the Two Parameter Exponential Distribution

```
clc, clear;

%Initial Values;
muhat = 162;
thhat = 835.21;
xbar = 997.21;
n = 19;
m = 19;

LCL = 372.4;
UCL = 1837.7;

nsim = 10000;
csim = 100000;

%Simulate mu;
mu = 0.01:0.01:muhat;
mupdf = (n-1).*((((1/thhat).^(n-1))-((1/xbar).^(n-1)))^-1).*((xbar-mu).^(-n));
mupdfn = mupdf/sum(mupdf);
mucdf = cumsum(mupdfn);
```

```
musim = zeros(1,nsim);
for i = 1:nsim
    musim(i) = mu(find(mucdf>rand,1,'first'));
end

%Simulate Theta;
paralph = n;
parlam = n.*(xbar-musim);
thsim = 1./gamrnd(paralph,1./parlam);

%Determine psi;
rlprob = zeros(1,nsim);
for k = 1:nsim
    chi2nsim = chi2rnd((2*n-2),1,csim);
    thfsim = chi2nsim.*thsim(k)./(2*n);
    rlprob(k) = (sum(thfsim<LCL | thfsim > UCL))/csim;
    if mod(k,1000) == 0
        disp(k);
    end
end
rl = (1-rlprob)./rlprob;
rlplot = rl(rl<1500);
```

```
hist(rlplot,50),xlabel('E(r|\mu,\theta)','fontsize',16);
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k');
set(gca,'fontsize',14);
save('thrl.mat')
```

A.10. MATLAB Code to Determine Rejection Region of GPQ from the Two Parameter Exponential Distribution

```
clc, clear;

%Initial Values;
muhat = 162;
thhat = 835.21;
xbar = 997.21;
n = 19;
m = 19;
ktil = -3.6784;
LCL = 1524;
UCL = 6929;

nsim = 1500;
```

```
csim = 1500;

%Simulate mu;
mu = 0.01:0.01:muhat;
mupdf = (n-1).*((((1/thhat).^(n-1))-((1/xbar).^(n-1)))^-1).*((xbar-mu).^(-n));
mupdfn = mupdf/sum(mupdf);
mucdf = cumsum(mupdfn);

musim = zeros(1,nsim);
for i = 1:nsim
    musim(i) = mu(find(mucdf>rand,1,'first'));
end

% %Simulate Theta;
paralph = n;
parlam = n.*(xbar-musim);
thsim = 1./gamrnd(paralph,1./parlam);

thfsim = chi2rnd((2*m-2),1,csim)./(2*m);

uf = 0:1:11000;
rlprob = zeros(1,nsim);
for l = 1:nsim
    ufpdf = zeros(1,size(uf,2));
```

```
    eparm = thsim(l)/m;
    for k = 1:csim
        sparm = musim(l) - ktil*thfsim(k).*thsim(l);
        ufpdf = ufpdf + exp((1/eparm)*sparm).*exppdf(uf,eparm).*(uf > sparm);
    end
    ufpdfs = ufpdf./csim;
    ufpdfn = ufpdfs./sum(ufpdfs);
    ufcdf = cumsum(ufpdfn);
    if mod(l,10) == 0
        disp(l);
    end
    rlprob(l) = 1 - ufcdf(uf == UCL) + ufcdf(uf == LCL);
end

rl = (1-rlprob)./rlprob;
rlplot = rl(rl<2500);
% for z = 1:size(rlplot,2)
%     if rlplot(z) > 2500
%         rlplot(z) = 2500;
%     end
% end

hist(rlplot,50),xlabel('E(r|\mu,\theta)','fontsize',16);
h = findobj(gca,'Type','patch');
```

```
set(h,'FaceColor','w','EdgeColor','k');
set(gca,'fontsize',14);
save('gpqrl.mat')
```

A.11. MATLAB Code to Determine Distribution of μ from Two Parameter Exponential if $0 < \mu < \infty$

```
clc, clear;

%Initial Values;
muhat = 162;
thhat = 835.21;
xbar = 997.21;
n = 19;
m = 19;
alpha = 0.0258;

%Determine PDF;
muf = 0:0.001:800;

ksnum = (n^n)*(n-1)*m;
ksden = (n+m)*(((1/thhat)^(n-1))-((1/xbar)^(n-1)));
```

```
ks = ksnum/ksden;

t1 = (n .* (xbar-muf)).^(-n);
t2 = (m .* muf + n*xbar).^(-n);
t3 = (m .* (muf - muhat) + n*thhat).^(-n);

mupdf = ks.*(t1 - t2).*(muf <= muhat) + ks.*(t3-t2).*(muf > muhat);
mupdfn = mupdf./(sum(mupdf));
mucdf = cumsum(mupdfn);

LCL = muf(find(mucdf<alpha/2,1,'last'));
UCL = muf(find(mucdf>1-alpha/2,1,'first'));
set(0, 'defaultTextInterpreter', 'latex');
plot(muf,mupdfn,'color','k','linewidth',2),xlabel('$\hat{\mu}_f$','fontsize',20);
set(gca,'fontsize',14);

% mean = sum(muf.*mupdfn);
```

A.12. MATLAB Code to Determine Distribution of μ from Two Parameter Exponential if $-\infty < 0 < \infty$

```
clc, clear;
```

```
%Initial Values;
muhat = 162;
thhat = 835.21;
xbar = 997.21;
n = 19;
m = 19;
alpha = 0.0258;

%Determine PDF;
muf = -300:0.001:600;

ksnum = (n^n)*(n-1)*m*(thhat^(n-1));
ksden = (n+m);
ks = ksnum/ksden;

t1 = (n .* (xbar-muf)).^(-n);
%t2 = (m .* muf + n*xbar).^(-n);
t3 = (m .* (muf - muhat) + n*thhat).^(-n);

mupdf = ks.*(t1).*(muf <= muhat) + ks.*(t3).*(muf > muhat);
mupdfn = mupdf./(sum(mupdf));
mucdf = cumsum(mupdfn);
```

```
LCL = muf(find(mucdf<alpha/2,1,'last'));
UCL = muf(find(mucdf>1-alpha/2,1,'first'));
set(0, 'defaultTextInterpreter', 'latex');
plot(muf,mupdfn,'color','k','linewidth',2),xlabel('$\hat{\mu}_f$','fontsize',20);
set(gca,'fontsize',14);

% mean = sum(muf.*mupdfn);
```

A.13. MATLAB Code to Determine Distribution of θ from Two Parameter Exponential for both $0 < \mu < \infty$ and $-\infty < \mu < \infty$

```
clc, clear;

%Initial Values;
muhat = 162;
thhat = 835.21;
xbar = 997.21;
n = 19;
m = 19;
alpha = 0.018;

%Determine PDF;
```

```
thf = 0:0.001:3000;

term1 = m^(m-1);
term2 = n^(n-1);
term3 = gamma(m+n-2)/(gamma(m-1)*gamma(n-1));
term4 = thhat^(1-n);
term5 = xbar^(1-n);
term6 = thf.^(m-2);
term7 = (m.*thf+n*thhat).^(2-m-n);
term8 = (m.*thf+n*xbar).^(2-m-n);

tilterm1 = gamma(m+n-2).*(m^(m-1)).*((n*thhat)^(n-1)).*thf.^(m-2);
tilterm2 = gamma(m-1).*gamma(n-1).*((m.*thf+n.*thhat).^(m+n-2));


thpdf = term1.*term2.*term3.*((term4-term5).^-1).*term6.*(term7-term8);
thpdfn = thpdf./sum(thpdf);
thcdf = cumsum(thpdfn);

tilthpdf = tilterm1./tilterm2;
tilthpdfn = tilthpdf./sum(tilthpdf);

LCL = thf(find(thcdf<alpha/2,1,'last'));
UCL = thf(find(thcdf>1-alpha/2,1,'first'));
```

```
hold on;
set(0, 'defaultTextInterpreter', 'latex');
plot(thf,thpdfn,'k--','linewidth',2)
,xlabel('$\hat{\theta}_f$','fontsize',20);
set(gca,'fontsize',14);
set(0, 'defaultTextInterpreter', 'latex');
plot(thf,tilthpdfn,'k:','linewidth',2)
,xlabel('$\hat{\theta}_f$','fontsize',20), legend('$f({\theta}_f)$', '$\tilde{f}({\theta}_f)$');
set(gca,'fontsize',14);
hold off;
mean = sum(thf.*thpdfn);
```

A.14. MATLAB Code to Determine Distribution of U_f from Two Parameter Exponential for both $0 < \mu < \infty$ and $-\infty < \mu < \infty$

```
clc, clear;

%Initial Values;
muhat = 162;
thhat = 835.21;
xbar = 997.21;
```

```
n = 19;
m = 19;
ktil = -3.6784;
alpha = 0.018;

nsim = 1000;
csim = 1000;

%Simulate mu;
mu = 0.01:0.01:muhat;
mupdf = (n-1).*((((1/thhat).^(n-1))-((1/xbar).^(n-1)))^-1).*((xbar-mu).^(-n));
mutilpdf = (n-1).*((thhat).^(n-1)).*((xbar-mu).^(-n));
mupdfn = mupdf/sum(mupdf);
mutilpdfn = mutilpdf/sum(mutilpdf);
mucdf = cumsum(mupdfn);
mutilcdf = cumsum(mutilpdfn);

musim = zeros(1,nsim);
mutilsim = zeros(1,nsim);
for i = 1:nsim
    musim(i) = mu(find(mucdf>rand,1,'first'));
    mutilsim(i) = mu(find(mutilcdf>rand,1,'first'));
end
```

```
%Simulate Theta;
paralph = n;
parlam = n.*(xbar-musim);
parlamtil = n.*(xbar-mutilsim);
thsim = 1./gamrnd(paralph,1./parlam);
thtilsim = 1./gamrnd(paralph,1./parlamtil);

%Simulate Theta_f
% thfsim = chi2rnd((2*m-2),1,csim).*thsim./(2*m);

uf = 0:1:11000;
ufpdf = zeros(1,size(uf,2));
uftilpdf = zeros(1,size(uf,2));
for k = 1:nsim
    eparm = thsim(k)/m;
    eparmtil = thtilsim(k)/m;
    %Simulate Theta_f
    thfsim = chi2rnd((2*m-2),1,csim).*thsim(k)./(2*m);
    thftilsim = chi2rnd((2*m-2),1,csim).*thtilsim(k)./(2*m);
    for l = 1:csim
        sparm = musim(k) - ktil*thfsim(l);
        stilparm = mutilsim(k) - ktil*thftilsim(l);
        ufpdf = ufpdf + exp((1/eparm)*sparm).*exppdf(uf,eparm).*(uf > sparm);
        uftilpdf = uftilpdf + exp((1/eparmtil)*stilparm).*exppdf(uf,eparmtil).*(uf > stilparm);
```

```
    end
    if mod(k,10) == 0
        disp(k);
    end
end

ufpdfs = ufpdf./(nsim*csim);
uftilpdfs = uftilpdf./(nsim*csim);
ufpdfn = ufpdfs./sum(ufpdfs);
uftilpdfn = uftilpdfs./sum(uftilpdfs);
hold on;
set(0, 'defaultTextInterpreter', 'latex');
plot(uf,ufpdfs,'k-','linewidth',2)
,xlabel('$U_{f}$','fontsize',16);
set(gca,'fontsize',14);
plot(uf,uftilpdfs,'k:','linewidth',2)
,xlabel('$U_{f}$','fontsize',16),
legend('$f(U_{f}|data)$', '$\tilde{f}(U_{f}|data)$');
set(gca,'fontsize',14);
```

A.15. MATLAB Code To Determine Rejection Region of X_{28} from the Piecewise Exponential Model

```
clc, clear;

%Data and Initials;
data = xlsread('LHDData_orig.xls');
data(:,3) = data(:,3)/1000;
del = 0.01:0.001:2;
nsim = 2000;

%Determin N = number of observations;
N = size(data,1);

delpdf = zeros(1,size(del,2));
%Determine Distribution of delta
for i = 1:size(del,2);
    data(:,4) = data(:,2).^(1-del(i));
    data(:,5) = data(:,3) .* data(:,4);

    delpdf(i) = sum(data(:,5)).^(-1*N).*prod(data(:,4));
end;
```

```
delpdfN = delpdf./sum(delpdf);
delcdf = cumsum(delpdfN);

mu = 0.0001:0.0000001:0.01;
x = 50:0.01:190;
lpt = 28;
mudist = zeros(nsim,size(mu,2));
delsim = zeros(1,nsim);
for l = 1:nsim;
    delsim(l) = del(find(delcdf>rand,1,'first'));
    data(:,4) = data(:,2).^(1-delsim(l));
    data(:,5) = data(:,3) .* data(:,4)*1000;
    A = N;
    B = delsim(l)./sum(data(:,5));
    mudist(l,:) = gampdf(mu,A,B);
    if mod(l,10) == 0
        disp('l');
        disp(l);
    end
end;

mupdf = mean(mudist);
mupdfN = mupdf./sum(mupdf);
mucdf = cumsum(mupdfN);
```

```
musim = zeros(1,nsim);
for k = 1:nsim;
    musim(k) = mu(find(mucdf>rand,1,'first'));
    if mod(k,10) == 0
        disp('k');
        disp(k);
    end
end;

X28 = 0:0.01:1000;
epdf = zeros(nsim,size(X28,2));
%determine distribution of X_f f = 28;
for j = 1:nsim;
    eparm = delsim(j)./musim(j).*lpt.^(delsim(j)-1);
    epdf(j,:) = exppdf(X28,eparm);
    if mod(j,10) == 0
        disp('j');
        disp(j);
    end
end;
epdfm = mean(epdf);
epdfN = epdfm./sum(epdfm);
ecdf = cumsum(epdfN);
```

```
figure(1);
set(0, 'defaultTextInterpreter', 'latex');
plot(X28,epdfN,'k','linewidth',2),xlabel('$X_{28}$','fontsize',16);
set(gca,'fontsize',14);

LCL1 = 0.26;
LCL2 = 0.13;
UCL2 = 666;

rlprob1 = zeros(1,nsim);
rlprob2 = zeros(1,nsim);
for z = 1:nsim;
    dist = epdf(z,:);
    distN = dist./sum(dist);
    distcdf = cumsum(distN);
    rlprob2(z) = 1 - distcdf(X28 == UCL2) + distcdf(X28 == LCL2);
    rlprob1(z) = distcdf(X28 == LCL1);
    if mod(z,10) == 0
        disp('z');
        disp(z);
    end
end;
```

```
rl1 = (1-rlprob1)./rlprob1;
rl2 = (1-rlprob2)./rlprob2;

figure(2);
hist(rl1,50),xlabel('$E(r|\mu,\delta$)','fontsize',16);
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k');
set(gca,'fontsize',14);

figure(3);
hist(rl2,50),xlabel('$E(r|\mu,\delta$)','fontsize',16);
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k');
set(gca,'fontsize',14);

save('X28rl.mat','rl1','rl2', 'rlprob1', 'rlprob2');
```

A.16. MATLAB Code to Determine Rejection Region of C_{pk}

```
clc, clear;

%Limits
lim.U = 2.7205;
```

```
lim.L = 2.6795;

%Given Information
giv.n = 75;
giv.xbar = 2.6972;
giv.s = 0.0038;
giv.m = 10;
nsim = 1000;
giv.olim = 3.923;
giv.thlim = 4.263;
giv.tllim = 0.791;

alpha = 0.0027;

t = 0.4:0.001:5;
dist.parm1 = 3*(sqrt(giv.m/(giv.m-1)));

for j = 1:nsim;
    sigsim = ((giv.n-1)*giv.s^2)/chi2rnd(giv.n-1,1,1);
    musim = normrnd(giv.xbar,sqrt(sigsim/giv.n));
    calc.cpk = min(((lim.U-musim)/(3*sqrt(sigsim))),((musim-lim.L)/(3*sqrt(sigsim))));
    calc.cp = (lim.U-lim.L)/(6*sqrt(sigsim));
    for i = 1:size(t,2);
        ilimit = ((giv.m-1)/(t(i))^2)*(calc.cp^2);
```

```
            y = 0:0.01:ilimit;
            dist.parm2 = 3*sqrt(giv.m).*(calc.cpk - 2*calc.cp + sqrt(y./(giv.m-1))*t(i));
            dist.parm3 = 3*sqrt(giv.m).*(calc.cpk-sqrt(y./(giv.m-1)).*t(i));
            dist.parm4 = (normpdf(dist.parm2,0,1)+normpdf(dist.parm3,0,1)).*(sqrt(y)).*(chi2pdf(y,giv.m-1));
            dist.fcpk(i) = dist.parm1*trapz(y,dist.parm4);
        end
        dist.fcpk = dist.fcpk/sum(dist.fcpk);
        dist.Fcpk = cumsum(dist.fcpk);
        rl.pone(j) = 1 - dist.Fcpk(t==giv.olim);
        rl.ptwo(j) = 1 - dist.Fcpk(t==giv.thlim) + dist.Fcpk(t==giv.tllim);
        if mod(j,5) == 0
            disp(j);
        end
    end;

    rl.rlone = (1-rl.pone)./rl.pone;
    rl.rltwo = (1-rl.ptwo)./rl.ptwo;
    set(0, 'defaultTextInterpreter', 'latex');
    figure(1);
    hist(rl.rlone,50),xlabel('$E(r|\mu,\sigma^2)$','fontsize',20);
    h = findobj(gca,'Type','patch');
    set(h,'FaceColor','w','EdgeColor','k');
    set(gca,'fontsize',14);
    figure(2);
```

```
hist(rl.rltwo,50),xlabel('$E(r|\mu,\sigma^2)$','fontsize',20);
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k');
set(gca,'fontsize',14);

%save('rlres.mat');
```

A.17. MATLAB Code for Ganesh Simulation

```
clear
tic
n=[50;75;70;75];
x=[2.7048;2.7019;2.6979;2.6972];
s=[0.0034;0.0055;0.0046;0.0038];
%n=n(3); x=x(3); s=s(3);
LSL=2.6795; USL=2.7205;
k=4;
%Cphat=(USL-LSL)./6./s;
%Cpuhat=(USL-x)./3./s;
%Cplhat=(x-LSL)./3./s;
%Cpl=[]; Cpu=[];
Cpk=[];
N=100000;
```

```
for i=1:N
sig=s.*sqrt((n-1)./chi2rnd(n-1));
cpl=normrnd((x-LSL)./3./sig,1./3./sqrt(n));
cpu=normrnd((USL-x)./3./sig,1./3./sqrt(n));
%Cpl=[Cpl;cpl'];
%Cpu=[Cpu;cpu'];
cpk=min([cpl';cpu']);
Cpk=[Cpk;cpk];
end
M=mean(Cpk);
A=Cpk-ones(N,1)*M;
T2=[];
for j=1:N
t2=max(A(j,:))-min(A(j,:));
T2=[T2;t2];
end
toc

hist(T2,50);
set(0, 'defaultTextInterpreter', 'latex');
h = findobj(gca,'Type','patch'), xlabel('$T^{(2)}$','fontsize',20);
set(h,'FaceColor','w','EdgeColor','k');
set(gca,'fontsize',14);
```

A.18. MATLAB Code To Determine Continuous Distribution Function of the Run-Length

```
clc, clear;
load('matlab75.mat');
xg = 0:1:2000;
rlsim.p = rl.ptwo;
rlsim.p = rlsim.p(rlsim.p ~= 0);
nsim = size(rlsim.p,2);
prob = zeros(1,nsim);
for i = 1:nsim
    prob(i) = rlsim.p(i);
    sim.rl(i,:) = geopdf(xg,prob(i));
    if mod(i,100) == 0
        disp(i);
    end
end;
sim.rls = mean(sim.rl,1);
sim.rlspdf = sim.rls./sum(sim.rls);
sim.rlscdf = cumsum(sim.rlspdf);

rl.rlmean = sum(xg.*sim.rlspdf);
rl.rlvar = sum((xg.^2).*sim.rlspdf) - (rl.rlmean.^2);
```

```
rl.rlmedian = xg(find(sim.rlscdf>=0.5,1,'first'));
rl.elow = xg(find(sim.rlscdf >= 0.025,1,'first'));
rl.eup = xg(find(sim.rlscdf >= 0.975,1,'first'));

%hpd;
y = 0.01:0.01:0.05;
yi = 0.95:0.01:0.99;

diff1 = 10000000000;
hpdlow1 = zeros(1,size(y,2));
hpdhigh1 = zeros(1,size(y,2));
hpddiff1 = zeros(1,size(y,2));
for l = 1:size(y,2)
    hpdlow1(l) = xg(find(sim.rlscdf>=y(l),1,'first'));
    hpdhigh1(l) = xg(find(sim.rlscdf>=yi(l),1,'first'));
    hpddiff1(l) = hpdhigh1(l) - hpdlow1(l);
    if hpddiff1(l) < diff1
        rl.hpdlow = hpdlow1(l);
        rl.hpdhigh = hpdhigh1(l);
        rl.hpddiff = hpddiff1(l);
        diff1 = hpddiff1(l);
    end
end
```

```
figure(1)
plot(xg,sim.rlspdf,'color','k','linewidth',2),xlabel('r');

figure(2);
runl = (1-prob)./prob;
hist(runl,100);
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k');
```

Bibliography

Aitkin, M., Laird, N., Francis, B., 1983. A reanalysis of the stanford heart transplant data (with discussion). J Am Stat Assoc 78, 264–292.

Ando, T., 2010. Bayesian Model Selection and Statistical Modeling. Chapman & Hall/CRC.

Arab, A., Rigdon, S., Basu, A., 2012. Bayesian inference for the piecewise exponential model for the reliability of multiple repairable systems. Journal of Quality Technology 44 (1), 28–38.

Bastos, L. S., Gamerman, D., 2006. Dynamic survival models with spatial frailty. Lifetime Data Anal 12, 441–460.

Bayarri, M., Berger, J., 2004. The interplay of bayesian and frequentist analysis. Statistical Science 19 (1), 58–80.

Bayarri, M., García-Donato, G., 2005. A bayesian sequential look at u-control charts. Technometrics 47(2), 142–151.

Berger, J., Bernardo, J., 1992. On the development of reference priors. Bayesian Statistics 4, 35–60.

Berger, J. O., Liseo, B., Wolpert, R. L., 1999. Integrated likelihood methods for eliminating nuisance parameters. Statistical Science 14, 1–28.

Berger, J. O., Pericchi, L. R., 1996. The intrinsic bayes factor for model selection and prediction. Journal of the American Statistical Association 91 (433), 109–122.
URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476668

Bernardo, J., 1998. Bayesian Reference Analysis - A Postgraduate Tutorial Course. Departament d'Estadistica i l. O, Facultat de Matematiques 46100-Burjasson, Valenicia, Spain.

Blackwell, D., 1947. Conditional expectation and unbiased sequential estimation. Annals of Mathematical Statistics 18 (1), 105–110.

Box, G., Tiao, G., 1973. Bayesian Inference in Statistical Analysis. Wiley.

Breslow, N. E., 1974. Covariance analysis of censored survival data. Biometrics 30, 89–99.

Chakraborti, S., Human, S., Graham, M., 2008. Phase I statistical process control charts: An overview and some results. Quality Engineering 21 (1), 52–62.

Chen, K., Pearn, W., 1997. An application of non-normal process capability indices. Quality and Reliability Engineering International 13, 355–360.

Chou, Y.-M., 1994. Selecting a better supplier by testing process capability indices. Quality Engineering 6, 427–438.

Chou, Y.-M., Owen, D., 1989. On the distributions of the estimated process capability indices. Communications in Statistics - Theory and Methods 18 (12), 4549–4560.
URL http://dx.doi.org/10.1080/03610928908830174

Clark, D. E., Ryan, L. M., 2002. Concurrent prediction of hospital mortality and length of stay from risk factors on admission. Health Services Res 37, 631–645.

Datta, G., Ghosh, J., 1995. On priors providing frequentist validity for bayesian inference. Biometrika 82 (1), 37–45.

Demarqui, F., Loschi, R., Dey, D., Colosimo, E., 2012. A class of dynamic piecewise exponential models with random time grid. Journal of Statistical Planning and Inference 142, 728–742.

Duncan, A., 1965. Quality Control and Industrial Statistics. Vol. 3rd. Richard D. Irwin, Inc.

Gamerman, D., 1994. Bayes estimation of the piece-wise exponential distribution. IEEE Trans Reliab 43, 128–131.

Ganesh, N., 2009. Simultaneous credible intervals for small are estimation problems. Journal of Multivariate Analysis 100, 1610–1621.

Grubbs, F., 1971. Approximate fiducial bounds on reliability for the two parameter negative exponential distribution. Technometrics 13, 873–876.

Hahn, G., Meeker, W., 1991. Statistical Intervals: A Guide for Practitioners. John Wiley & Sons, Inc.

Hamada, M., January/February 2002. Bayesian tolerance interval control limits for attributes. Quality and Reliability Engineering International 18 (1), 45–52.

Hamada, M., Wilson, A., Reese, C., Martz, H., 2008. Bayesian Reliability. Springer.

Hoskins, J., Stuart, B., Taylor, J., 1988. A motorola commitment: A six sigma mandate. the motorola guide to statistical process control for continuous improvement towards six sigma quality.

Hubele, N., Shahriari, H., Cheng, C., 1991. A bivariate process capability vector. Marcel Dekker, 299–310.

Hugo, J., 2012. Bayesian tolerance intervals for variance component models. Ph.D. thesis, University of the Free State.

Hulting, F., Robinson, J., 1994. The reliability of a series of repairable systems. a bayesian apporach. Naval Research Logistics 41, 483–506.

Human, S., Chakraborti, S., Smit, C., 2010. Shewart-type control charts for variation in Phase I data analysis. Computational Statistics and Data Analysis 54, 863–874.

Iglewicz, B., 1967. Some properties of the coefficient of variation. Ph.D. thesis, Virginia Polytechnic Institute.

Jeffreys, H., 1939. Theory of Probability. Clarendon Press.

Johnson, N., Kotz, S., 1970. Distributions in statistics : continuous univariate distributions. New York ; Chichester, etc. : Wiley, includes bibliographical references.

Kane, V., 1986. Process capability indices. Journal of Quality Technology 18, 41–52.

Kang, C. W., Lee, M. S., Seong, Y. J., Hawkins, D. M., 2007. A control chart for the coefficient of variation. Journal of Quality Technology 39 (2), 151–158.

Kass, R., Raftery, A., 1995. Bayes factors. Journal of the American Statistical Association 90 (430), 773–795.

Kim, J., Proschan, F., 1991. Piecewise exponential estimator of the survival function. IEEE Trans Reliab 40, 134–139.

Kotz, S., Johnson, N., 1993. Process Capability Indices. Chapman & Hall.

Krishnamoorthy, K., Mathew, T., 2009. Statistical Tolerance Regions: Theory, Applications and Computation. Wiley Series in Probability and Statistics.

Kumar, U., Klefsjö, B., 1992. Reliability analysis of hydraulic systems of lhd machines using the power law process model. Reliability Engineering and System Safety 35, 217–224.

Lawless, J., 1982. Statistical models and methods for lifetime data. Statistics in Medicine 1 (3).

Lyth, D., Rabiej, R., 1995. Critical variables in wood manufacturing's process capability: Species, structure, and moisture content. Quality Engineering 8 (2), 275–281.

Menzefricke, U., 2002. On the evaluation of control chart limits based on predictive distributions. Communications in Statistics - Theory and Methods 31(8), 1423–1440.

Menzefricke, U., 2007. Control chart for the generalized variance based on its predictive distribution. Communications in Statistics - Theory and Methods 36(5), 1031–1038.

Menzefricke, U., 2010a. Control chart for the variance and the coefficient of variation based on their predictive distribution. Communications in Statistics - Theory and Methods 39(16), 2930–2941.

Menzefricke, U., 2010b. Multivariate exponentially weighted moving average chart for a mean based on its predictive distribution. Communications in Statistics - Theory and Methods 39(16), 2942–2960.

Mukherjee, A., McCracken, A., Chakraborti, S., 2014. Control charts for simultaneous monitoring of parameters of a shifted exponential distribution. Journal of Quality Technology, Accepted.

Niverthi, M., Dey, D., 2000. Multivariate process capability: A bayesian perspective. Communication in Statistics - Simulation and Computation 29 (2), 667–687.
URL http://dx.doi.org/10.1080/03610910008813634

O'Hagan, A., 1995. Fractional bayes factors for model comparison. Journal of the Royal Statistical Society. Series B (Methodological 57 (1), 99–138.

Pan, R., Rigdon, S., 2009. Bayes inference for general repairable systems. Journal of Quality Technology 41, 82–94.

Pearn, W., Wu, C., 2005. A bayesian approach for assessing process precision based on multiple samples. European Journal of Operational Research 165 (3), 685–695.

Pievatolo, A., Ruggeri, F., 2004. Bayesian reliability analysis of complex repairable systems. Applied Stochastic Models and Data Analysis 20, 253–264.

Polansky, A., 2006. Permutation methods for comparing process capabilities. Journal of Quality Technology 38 (3), 254–266.

Ramalhoto, M., Morais, M., 1999. Shewart control charts for the scale parameter of a weibull control variable with fixed and variable sampling intervals. Journal of Applied Statistics 26, 129–160.

Rao, C., 1945. Information and accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society 37 (3), 81–91.

Reese, C., Wilson, A., Guo, J., Hamada, M., Johnson, V., 2011. A bayesian model for integrating multiple sources of lifetime information in system reliability assessments. Journal of Quality Technology 43, 127–141.

Robert, C., 2001. The Bayesian Choice. New York, NY: Springer.

Roy, A., Mathew, T., 2005. A generalized confidence limit for the reliability function of a two-parameter exponential distribution. Journal of Statistical Planning and Inference 128, 509–517.

Rubin, D., 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. Ann. Statist. 12 (4), 1151–1172.

Sahu, S. K., Dey, D. K., Aslanidu, H., Sinha, D., 1997. A weibull regression model with gamma frailties for multivariate survival data. Lifetime Data Anal 3, 123–137.

Severine, T., Mukerjee, R., Ghosh, M., 2002. On an exact probability matching property of right-invariant priors. Biometrika 89 (4), 952–957.

Smit, C., Chakraborti, S., 2009. Sample capabilitiy indices: Didistribution properties and some applications. South African Statistical Journal 43, 117–125.

Sürücü, B., Sazak, H., 2009. Monitoring reliability for a three-parameter weibull distribution. Reliability Engineering and System Safety 94 (2), 503–508.

Tibshirani, R., 1989. Noninformative priors for one parameter of many. Biometrika 76 (3), 604–608.

Wolpert, R. L., 2004. A conversation with james o. berger. Statistical Science 19 (1), 205–218.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley.

Nomenclature

CV	coefficient of variation ($\gamma = \frac{\sigma}{\mu}$)
FAP	false alarm probability: the overall probability of at least one false alarm.
LCL	lower control limit
Phase I	Phase in statistical process control where the primary interest is to assess process stability.
Phase II	Phase in statistical process control where the primary interest is the online monitoring of the process.
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. Datta and Ghosh (1995) derived the differential equation which a prior must satisfy if the posterior probability of a one-sided credibility interval for a parameteric function and its frequentist probability agree up to $O\left(n^{-1}\right)$ where n is the sample size.
reference prior	The reference prior was introduced by Bernardo (1998) and Berger and Bernardo (1992). As mentioned by Pearn and 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.
run-length	The number of future r until the control chart signals for the first time.
SPC	statistical process control
standardized mean	$\delta = \frac{\mu}{\sigma}$
tolerance interval	An interval constructed in such a way that it will contain a specified proportion or more of the population with a certain degree of confidence.
UCL	upper control limit