A Bayesian Procedure for the Piecewise Exponential Model

R. van Zyl[1], A.J. van der Merwe[2]

[1]Quintiles International, ruaanvz@gmail.com

[2]University of the Free State

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

Abstract

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 report an objective Bayesian procedure will be applied for analyzing times between failures from multiple repairable systems.

Keywords: piecewise exponential, Bayes, Jeffreys' prior, control charts

1 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 PEXP model, μ is a scale parameter and δ is a shape parameter.

2 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 $\underline{x}_i = [x_{i1}, x_{i2}, \ldots, x_{i,n_i}]'$ denote the times between failures for the ith system.

As in Arab et al. (2012) three cases for multiple systems will be considered.

3 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

$$
\begin{aligned}
L\left(\delta, \mu | \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k\right) &= L\left(\delta, \mu | data\right) \\
&= \prod_{i=1}^{k} \left\{ \prod_{j=1}^{n_i} \left(\frac{\delta}{\mu} j^{\delta-1}\right)^{-1} \exp\left[-\frac{x_{ij}}{\frac{\delta}{\mu} j^{\delta-1}}\right] \right\} \\
&= \left(\frac{\delta}{\mu}\right)^{-N} \left\{ \prod_{i=1}^{k} \prod_{j=1}^{n_i} j^{1-\delta} \right\} \exp\left\{-\frac{\mu}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right\}.
\end{aligned}
$$

In the Bayesian approach 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 Jeffrey's rule. Jeffrey's 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 3.1. *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 Appendix A. □

4 The Joint Posterior Distribution - Identical Systems

$Posterior \propto Likelihood \times Prior$

Therefore

$$
\begin{aligned}
p\left(\mu, \delta | data\right) &\propto L\left(\mu, \delta | data\right) p_J\left(\mu, \delta\right) \\
&\propto \left(\frac{\delta}{\mu}\right)^{-N} \left\{ \prod_{i=1}^{k} \prod_{j=1}^{n_i} j^{1-\delta} \right\} \exp\left\{-\frac{\mu}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right\} \mu^{-1}.
\end{aligned}
$$

From the joint posterior distribution the marginal posterior distribution can easily be obtained. Now

$$\begin{aligned} p(\delta|data) &= \textstyle\int_0^\infty p(\mu,\delta|data)\,d\mu \\ &= \delta^{-N}\textstyle\prod_{i=1}^k\prod_{j=1}^{n_i} j^{1-\delta}\left\{\int_0^\infty \mu^{N-1}\exp\left[-\frac{\mu}{\delta}\sum_{i=1}^k\sum_{j=1}^{n_i} x_{ij}j^{1-\delta}\right]d\mu\right\}. \end{aligned}$$

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

$$p(\delta|data) \propto \left(\sum_{i=1}^k\sum_{j=1}^{n_i} x_{ij}j^{1-\delta}\right)^{-N}\prod_{i=1}^k\prod_{j=1}^{n_i} j^{1-\delta} \quad \delta > 0 \tag{4.1}$$

and

$$p(\mu|\delta,data) = \left(\frac{\delta}{\sum_{i=1}^k\sum_{j=1}^{n_i} x_{ij}j^{1-\delta}}\right)^{-N}\frac{1}{\Gamma(N)}\mu^{N-1}\exp\left\{-\frac{\mu}{\delta}\sum_{i=1}^k\sum_{j=1}^{n_i} x_{ij}j^{1-\delta}\right\} \quad \mu>0 \tag{4.2}$$

- a Gamma density function.

Equation (4.2) follows from the fact that

$$p(\mu,\delta|data) = p(\delta|data)\,p(\mu|\delta,data).$$

5 Example (Arab et al. (2012))

Table 5.1: 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 5.1 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 5.1 the posterior distribution of δ,

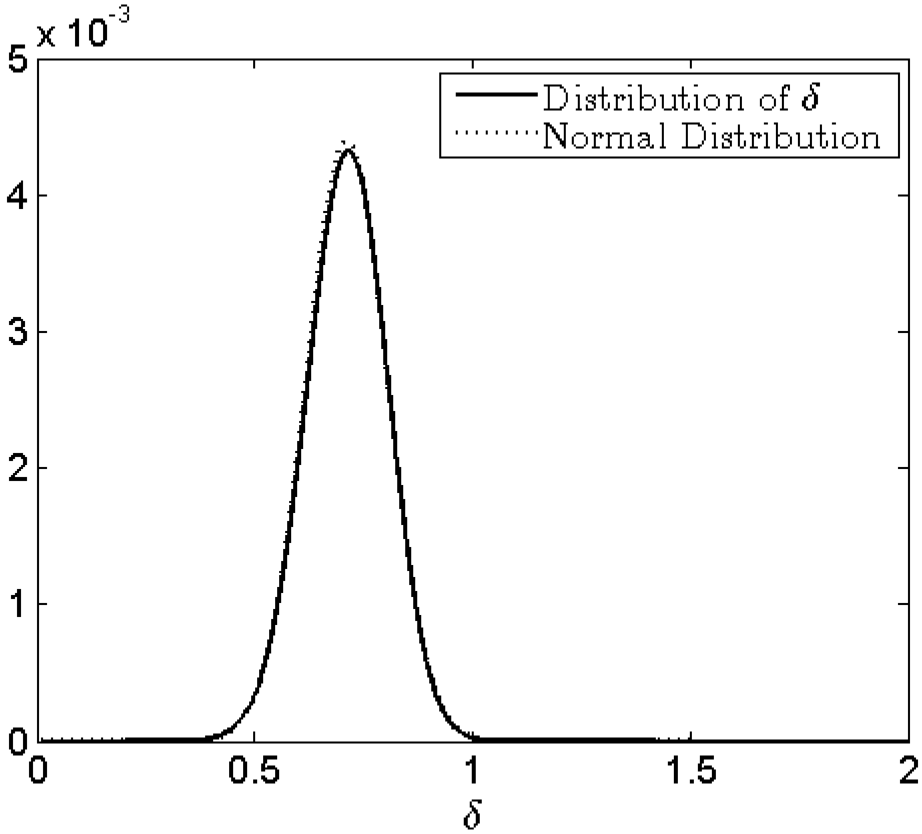
$$p\left(\delta|data\right) \propto \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right)^{-N} \prod_{i=1}^{k}\prod_{j=1}^{n_i} j^{1-\delta}$$

is illustrated for the data in Table 5.1 and in Figure 5.2 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 (4.1).

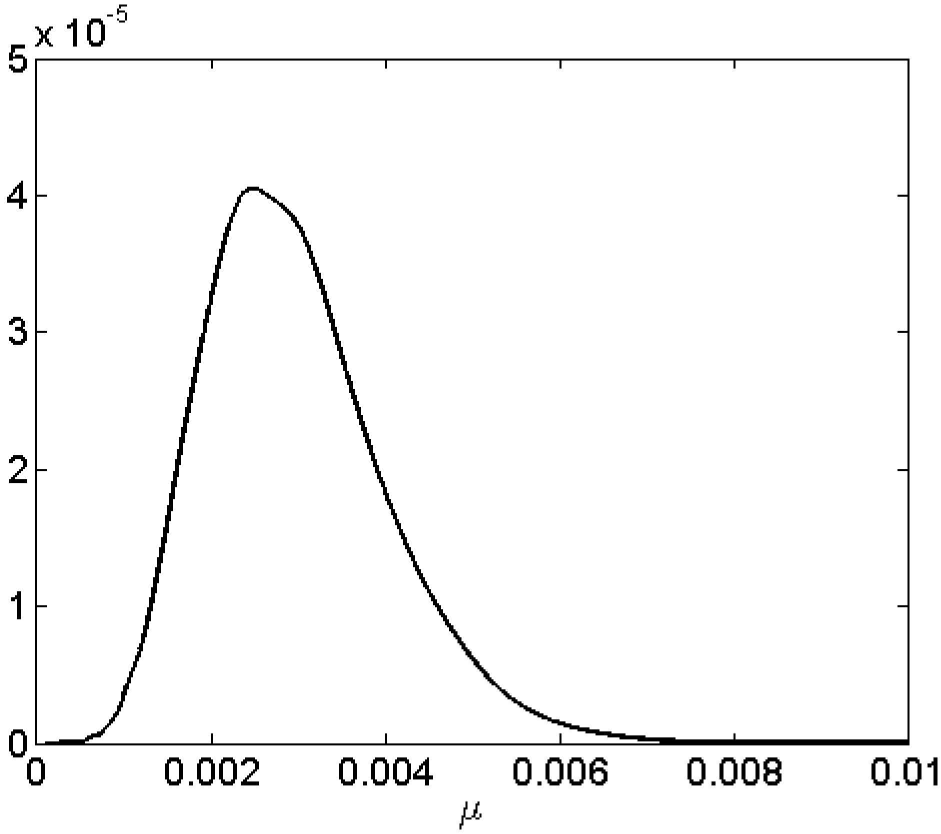
Figure 5.1: Posterior Distribution of δ



$mean\,(\delta) = 0.7109,\ var\,(\delta) = 0.00856,\ 95\%\ HPD\ Interval\,(\delta) = (0.5296; 0.8922)$

The posterior distribution of δ fits almost perfectly to a normal distribution with the same mean and variance.

Figure 5.2: 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 5.2 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 5.2 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 5.2: Point and Interval Estimates for the Parameters of the PEXP 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 5.2 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.

6 Simulation of PEXP 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 5.1. For each sample 10000 values are simulated from the posterior distributions of δ and μ and the means, variances and 95% intervals calculated. Table 6.1 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 6.1: Simulation Study Comparing Different Priors

	Prior	Mean	Median	Mode (approx)	Var	HPD Interval Coverage	HPD Interval Length	Equal-Tail Interval Coverage	Equal-Tail Interval 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 6.1 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).

7 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

$$E(X_l) = \frac{\delta}{\mu} l^{\delta-1} = t(\mu,\delta) = t(\underline{\theta}).$$

7.1 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 7.1. *The reference prior for the mean* $E(X_l) = \frac{\delta}{\mu} l^{\delta-1}$ *is* $p_R(\mu,\delta) \propto \mu^{-1}$.

Proof. The proof is given in Appendix B. □

7.2 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 $0(n^{-1})$ 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.

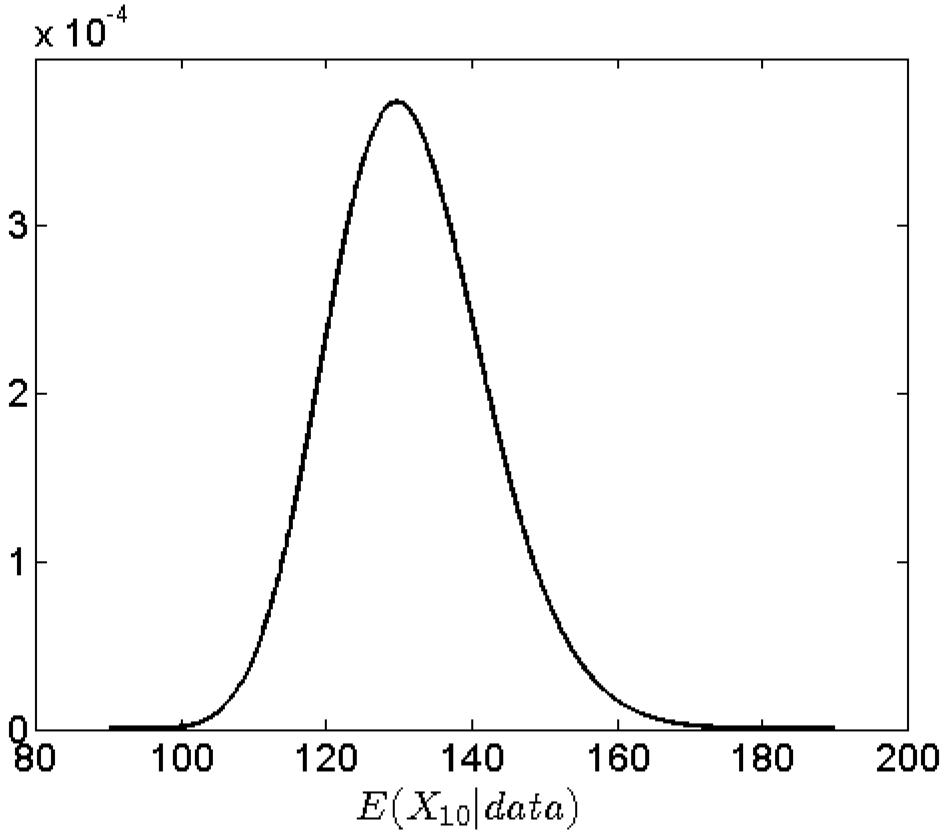
Theorem 7.2. *The probability matching prior for the mean* $E(X_l) = \frac{\delta}{\mu} l^{\delta-1}$ *is* $p_M(\mu,\delta) \propto \mu^{-1}$.

Proof. The proof is given in Appendix C. □

8 Example: Posterior Distribution of the Mean $E(X_l) = \frac{\delta}{\mu} l^{\delta-1}$

Consider again the data in Table 5.1. The expected time between the 9th and 10th failure is $E(X_{10}) = \frac{\delta}{\mu} 10^{\delta-1}$. In Figure 8.1 the posterior distribution of $E(X_{10})$ is given.

Figure 8.1: Posterior Distribution of Mean Time to Failure when $l = 10$

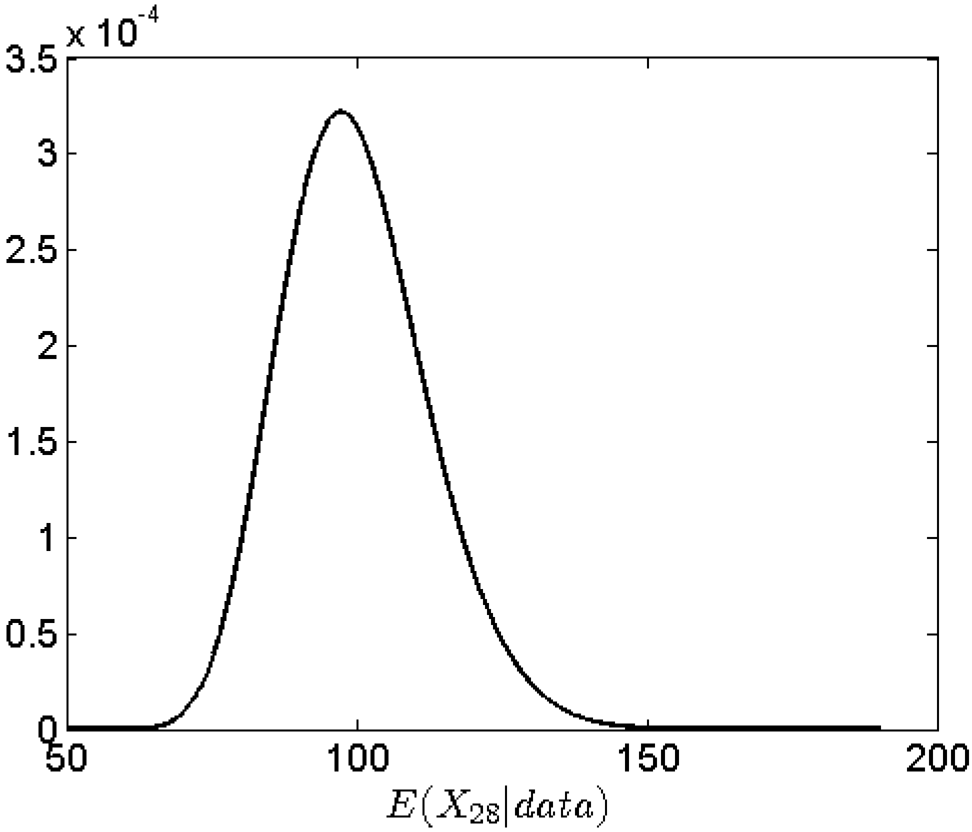


$$mean = 129.598,\ median = 129.028,\ mode = 128.755,\ var = 112.329$$
$$95\%\ Equal-tail\ Interval = (110.42; 152.06)\ \ length = 41.66$$
$$95\%\ HPD\ Interval = (110.27; 151.89)\ \ length = 41.53$$

It is clear from the figure that the posterior distribution is quite symmetrical.

In Figure 8.2, the posterior distribution of $E(X_{28})$ is illustrated.

Figure 8.2: 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 8.1 and 8.2 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

$$p(y|data,\delta) = \left(\frac{\delta}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} j^{1-\delta}}\right)^{-N} \frac{1}{\Gamma(N)} \left(\frac{1}{y}\delta l^{\delta-1}\right)^{N-1} \times$$

$$\exp\left\{\frac{-\delta l^{\delta-1}}{\delta y}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right\}\left(\frac{1}{y^2}\delta l^{\delta-1}\right).$$

$$\therefore p(y|data,\delta) = \left(l^{\delta-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right)^{N} \frac{1}{\Gamma(N)} \exp\left\{-\frac{1}{y} l^{\delta-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}\right\}.$$

An Inverse Gamma density function.

By using the Rao-Blackwell method $p(y|data)$ can be obtained.

9 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 6 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 9.1 the coverage percentage of the 95%credibility intervals are given

Table 9.1: Coverage Percentage of the 95% Credibility Interval for $E\left(X_{28}|\mu,\delta\right)$ from 100000 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.

10 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

$$f\left(x_f|\mu,\delta\right) = \left(\frac{\delta}{\mu}f^{\delta-1}\right)^{-1} \exp\left\{-\frac{x_f}{\left(\frac{\delta}{\mu}f^{\delta-1}\right)}\right\} \tag{10.1}$$

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 (4.1) and (4.2).

The unconditional predictive distribution of X_f can easily be obtained by using the simulation procedure described in Section 5:

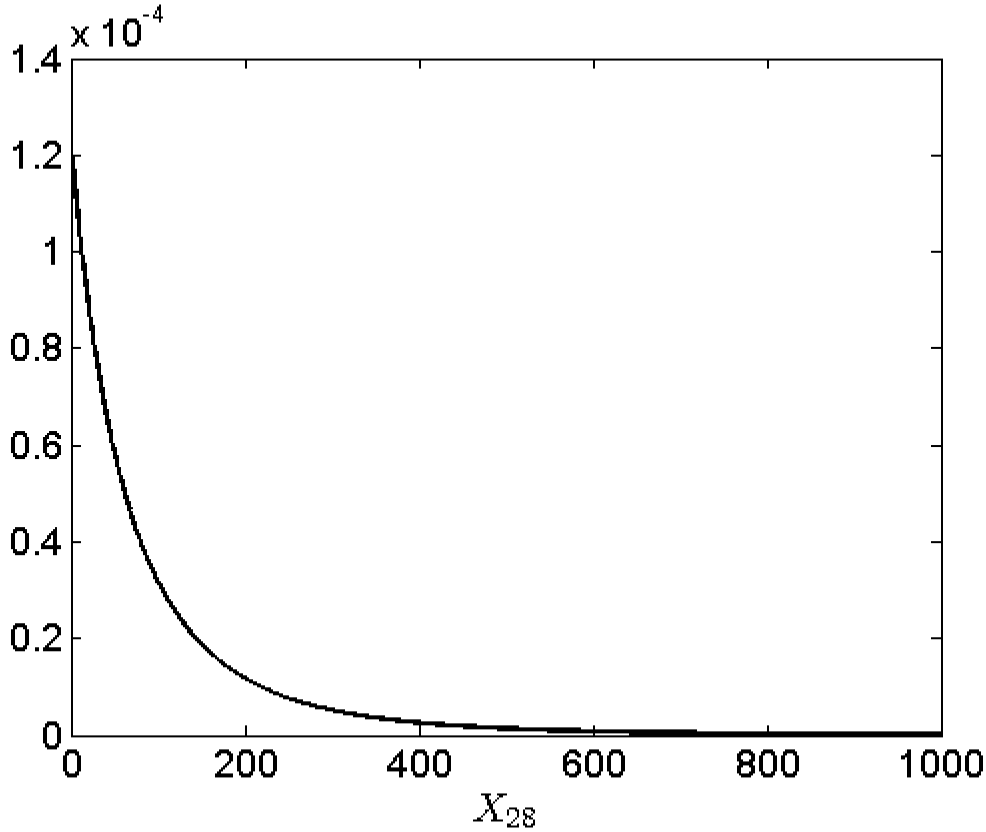
i. Obtain simulated values for δ and μ and substitute them in $f\left(x_f|\mu,\delta\right)$.

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 10.1 the predictive density function of X_{28} is given for the six load-haul-dump machines.

Figure 10.1: Predictive Density of X_{28}



$$Mean = 97.96,\ Median = 68.97,\ Var = 9736.8$$
$$95\%\ HPD\ Interval = (0; 317.1)$$
$$95\%\ Equal - tail\ Interval = (2.44; 416.20)$$
$$99.73\%\ Equal - tail\ Interval = (0.1294; 666)$$
$$0.27\%\ Left - sided\ Interval = (0; 0.2605)$$

11 Control Chart for $X_f = X_{28}$

It is well known that statistical quality 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 of the process in Phase II. 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 10.1 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 (10.1), 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

$$E(r|\mu,\delta) = \frac{1-\psi(\mu,\delta)}{\psi(\mu,\delta)}$$

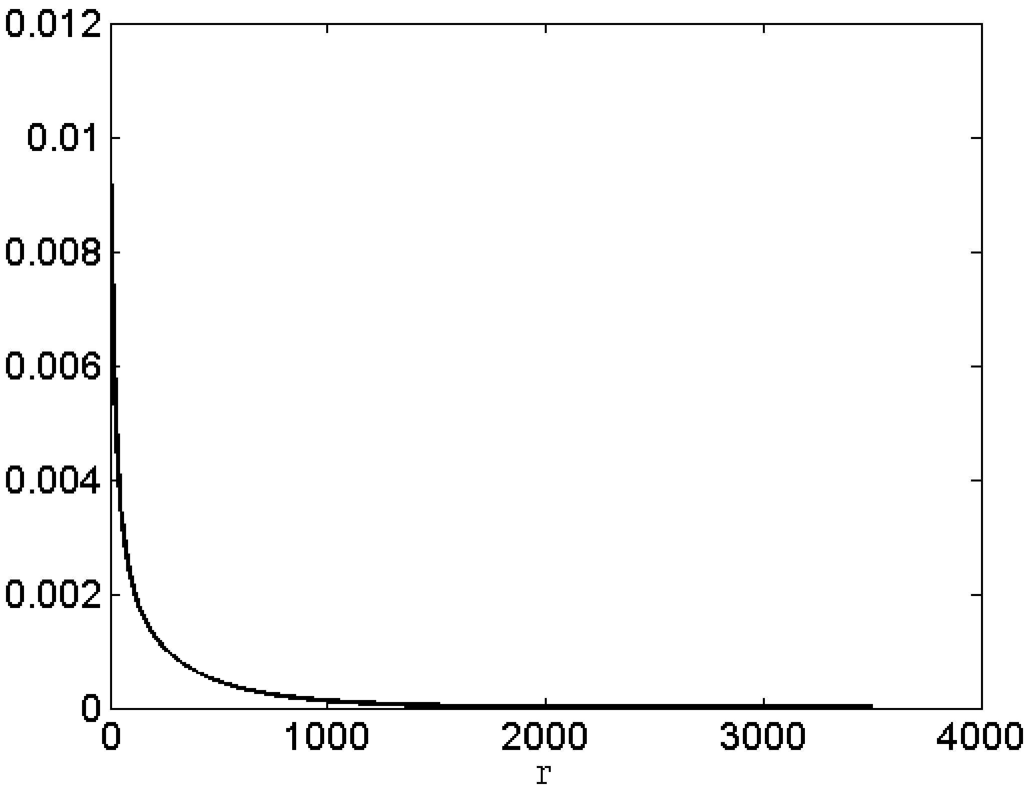
and

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

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

By averaging the conditional distributions the unconditional distribution of the run-length can be obtained and is illustrated in Figure 11.1.

Figure 11.1: Distribution of Run Length, $\beta = 0.0027$, Two-sided Interval

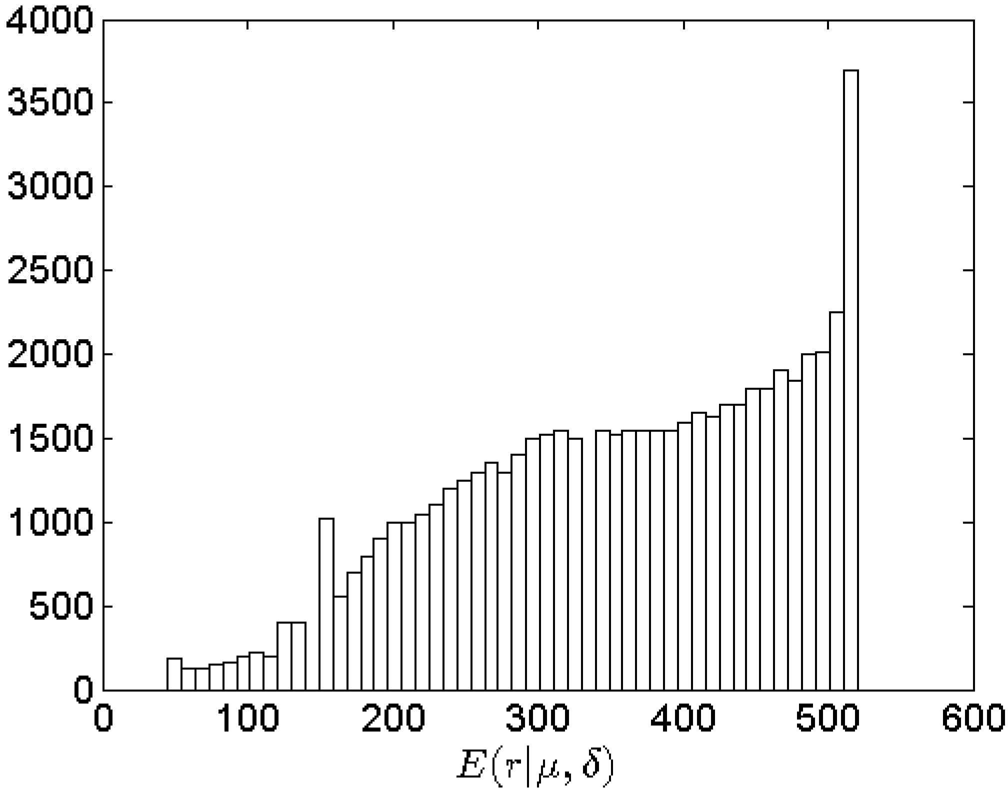


$$mean(r) = 369.09,\ median(r) = 239.75,\ var(r) = 1.5919e^5$$
$$95\%\ HPD\ Interval = (0; 1160.7)$$

The mean run-length of 369.09 corresponds to the value of $\frac{1}{\beta} = \frac{1}{0.0027} \approx 370$.

In Figure 11.2 the histogram of the expected run-length is given.

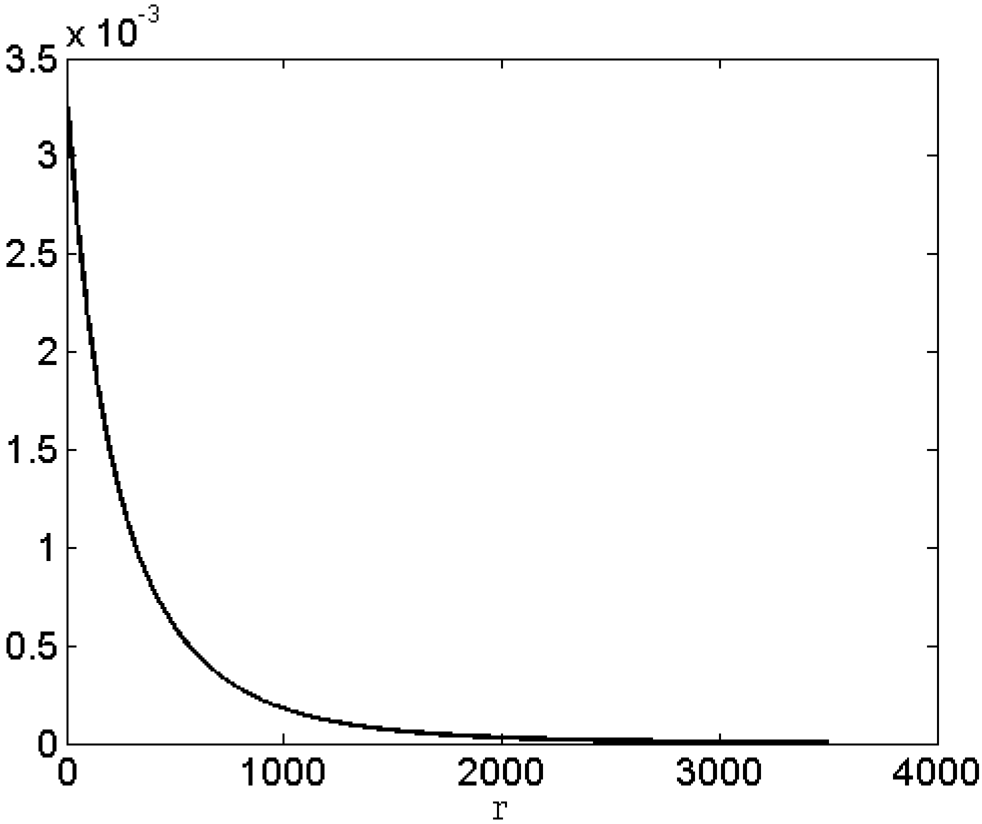
Figure 11.2: Expected Run Length, $\beta = 0.0027$, Two-sided Interval



$mean\,(r) = 369.87,\ median\,(r) = 381.94,\ var\,(r) = 1.1407e^4$
$95\%\ HPD\ Interval = (42.35; 518.16)$

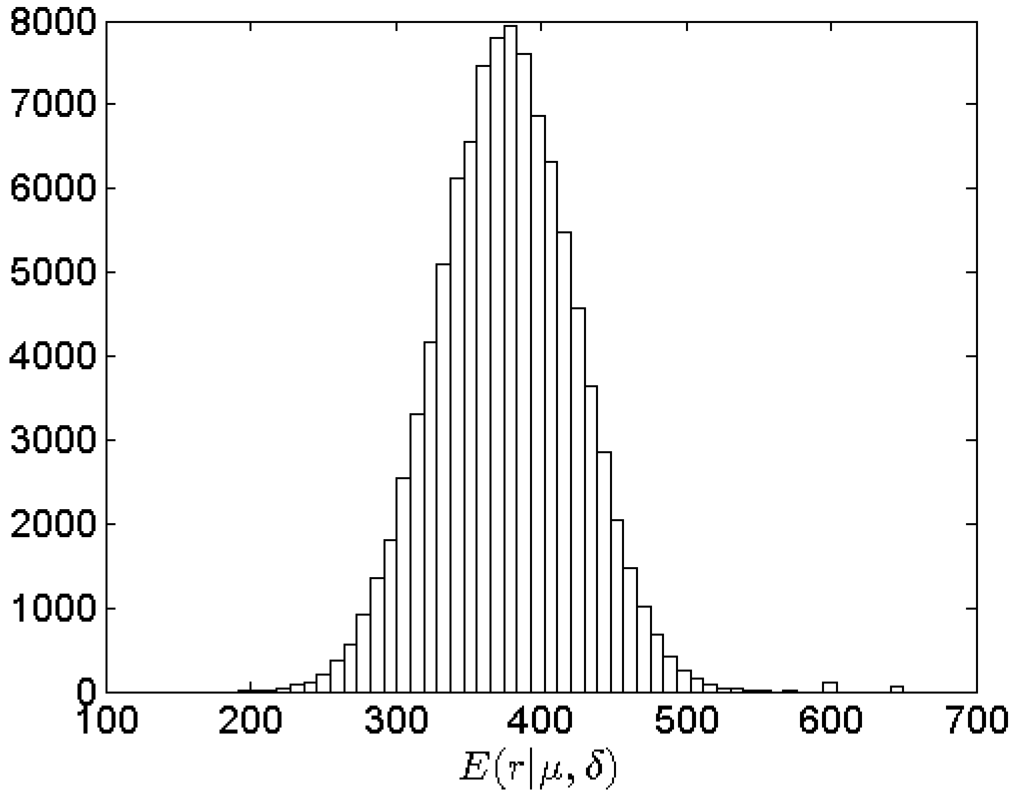
At the bottom of Figure 10.1 it is shown that a 0.27% left-sided interval = (0;0.2605). $R(\beta)$ therefore represents those values of X_f that are larger than 0 and smaller than 0.2605. The distribution of the run-length for this one-sided interval is displayed in Figure 11.3 and in Figure 11.4 the distribution of the expected run-length is given.

Figure 11.3: 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 11.4: Expected Run Length, $\beta = 0.0027$, One-sided Interval



$$mean\,(r) = 376.30,\ median\,(r) = 372.58,\ var\,(r) = 2179.2$$
$$95\%\ HPD\ Interval = (296.14; 517.82)$$

12 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 6 and 9. 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 10000 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 12.1 the coverage percentage of the 95% prediction interval for X_{28} from 10000 samples are given.

Table 12.1: Coverage Percentage of 95% Prediction Interval

	% Coverage	Length
Equal-Tail Interval	95.52	351.92
HPD Interval	95.46	285.95

From Table 12.1 it is clear that the predictive interval has the correct frequentist coverage percentage.

13 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 13.1. *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 Appendix D. □

14 The Joint Posterior Distribution of the Parameters in the Case of Model 2

The joint posterior distribution

$$\begin{aligned} p(\mu_1, \mu_2, \ldots, \mu_k, \delta | data) &\propto L(\mu_1, \mu_2, \ldots, \mu_k, \delta | data) \times p_J(\mu_1, \mu_2, \ldots, \mu_k, \delta) \\ &\propto \prod_{i=1}^{k} \left\{ \prod_{j=1}^{n_i} \left(\frac{\delta}{\mu_i} j^{\delta-1} \right)^{-1} \exp\left[-\frac{x_{ij}}{\left(\frac{\delta}{\mu_i}\right) j^{\delta-1}} \right] \right\} \prod_{i=1}^{k} \mu_i^{-1}. \end{aligned}$$

From this it follows that

$$p(\mu_i | data, \delta) = \left(\frac{\delta}{\sum_{j=1}^{n_i} x_{ij} j^{1-\delta}} \right)^{-n_i} \frac{\mu_i^{n_i - 1}}{\Gamma(n_i)} \exp\left[-\frac{\mu_i}{\delta} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \right] \quad \mu_i > 0, i = 1, 2, \ldots, k \tag{14.1}$$

and

$$p(\delta | data) \propto \prod_{i=1}^{k} \left\{ \sum_{j=1}^{n_i} \left(x_{ij} j^{1-\delta} \right)^{-n_i} \left(\prod_{j=1}^{n_i} j^{1-\delta} \right) \right\} \quad \delta > 0 \tag{14.2}$$

The posterior distribution of δ for a piecewise exponential model with different scale parameters differ somewhat from the distribution of δ if $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ (given in Equation (4.1)).

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 14.1 the means variances and 95% credibility intervals are given for μ_i and δ.

Figure 14.1: Posterior Density of δ: Model 2

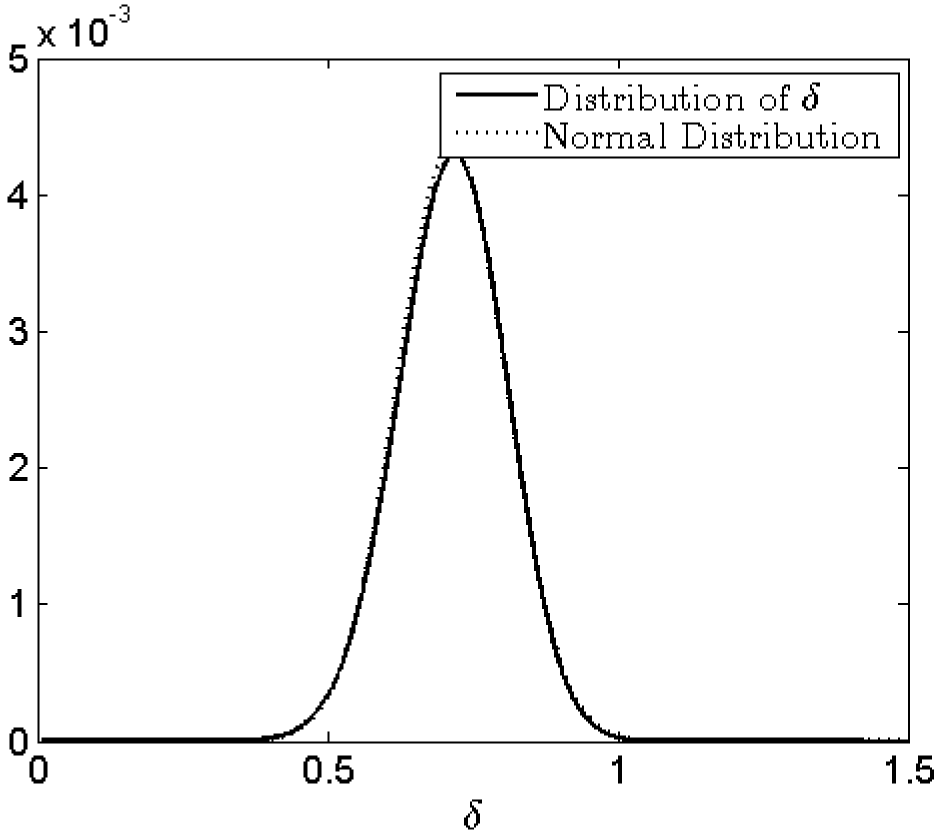


Figure 14.2: Posterior Densities of $\mu_1, \mu_2, \ldots, \mu_6$: Model 2

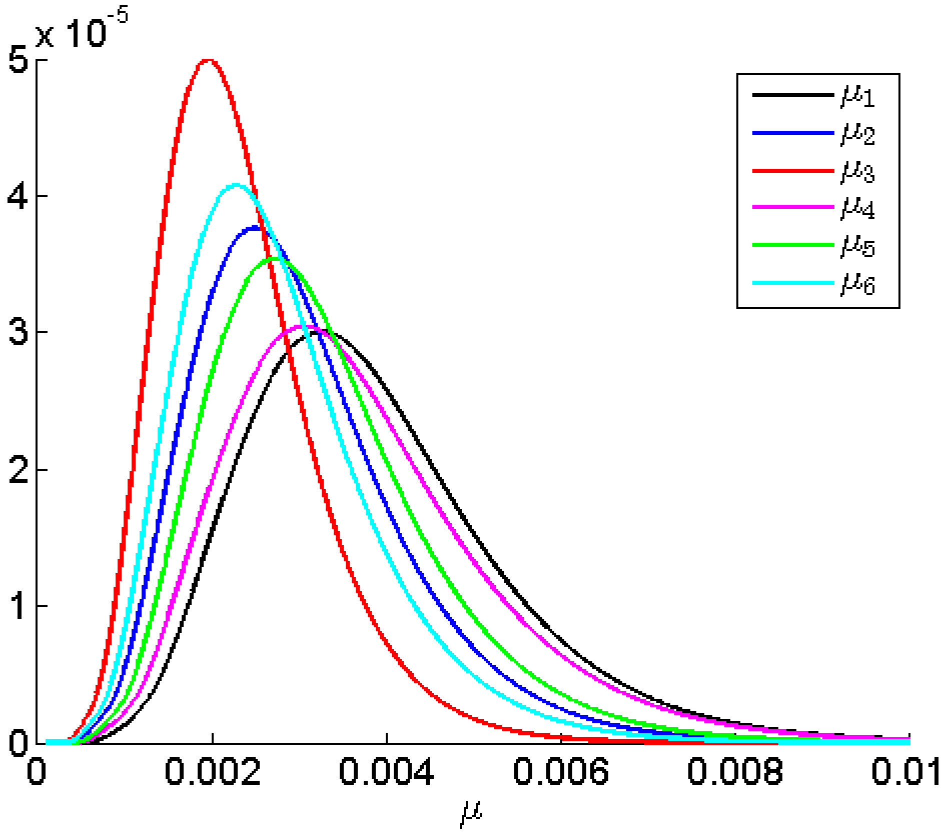


Table 14.1: Point Estimates and Credibility Intervals for the Parameters of the PEXP Model in the Case of Systems with Different μ's and Common δ for the LHD Example

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.

15 Simulation Study of the Piecewise Exponential Model Assuming Systems with Different μ's and Common δ

In this simulation study 3100 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 10000 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 14.1 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 15.1: Point Estimates and Credibility Intervals Obtained from a Simulation of the PEXP 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

16 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

$$q_i\left(b, \underline{y}\right) = \frac{\int \pi_i\left(\underline{\theta}_i\right) f_i\left(\underline{y}|\underline{\theta}_i\right) d\underline{\theta}_i}{\int \pi_i\left(\underline{\theta}_i\right) f_i\left(\underline{y}|\underline{\theta}_i\right)^b d\underline{\theta}_i} \qquad i = 1, 2$$

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

$$q_i\left(b, \underline{y}\right) = \frac{m_i}{m_i\left(b\right)} \qquad i = 1, 2$$

and

$$B_b\left(\underline{y}\right) = FBF_{12} = \frac{m_1 m_2\left(b\right)}{m_2 m_1\left(b\right)}.$$

17 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 5, 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 (1961) recommended interpreting the Bayes factors as a scale of evidence. Table 17.1 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 17.1: 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).

References

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.

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`

Breslow, N. E., 1974. Covariance analysis of censored survival data. Biometrics 30, 89–99.

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.

Gamerman, D., 1994. Bayes estimation of the piece-wise exponential distribution. IEEE Trans Reliab 43, 128–131.

Hamada, M., Wilson, A., Reese, C., Martz, H., 2008. Bayesian Reliability. Springer.

Hulting, F., Robinson, J., 1994. The reliability of a series of repairable systems. A bayesian apporach. Naval Research Logistics 41, 483–506.

Jeffreys, H., 1961. Theory of Probability. Clarendon Press.

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.

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.

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.

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.

Pievatolo, A., Ruggeri, F., 2004. Bayesian reliability analysis of complex repairable systems. Applied Stochastic Models and Data Analysis 20, 253–264.

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.

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.

Mathematical Appendices

A Proof of Theorem 4.1

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

$$\begin{aligned} \tfrac{\partial^2 l}{\partial \delta^2} &= \tfrac{N}{\delta^2} + \mu \{ \tfrac{-2}{\delta^3} \textstyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} + \tfrac{1}{\delta^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} (-1) j^{1-\delta} \log_e j - \tfrac{1}{\delta^2} \sum_{i=1}^{j} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \log_e j \\ &\quad + \tfrac{1}{\delta} \textstyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} (-1) j^{1-\delta} \left(\log_e j\right)^2 \} \\ &= \tfrac{N}{\delta^2} - \mu \left\{ \tfrac{2}{\delta^3} \textstyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} + \tfrac{2}{\delta^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \log_e j + \tfrac{1}{\delta} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} j^{1-\delta} \left(\log_e j\right)^2 \right\} \end{aligned}$$

Therefore

$$\begin{aligned}
-E\left(\frac{\partial^2 l}{\partial \delta^2}\right) &= \frac{-N}{\delta^2}+\mu\{\tfrac{2}{\delta^3}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\tfrac{\delta}{\mu}j^{\delta-1}\right)j^{1-\delta}+\tfrac{2}{\delta^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\tfrac{\delta}{\mu}j^{\delta-1}\right)j^{1-\delta}\log_e j \\
&\quad +\tfrac{1}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\tfrac{\delta}{\mu}j^{\delta-1}\right)j^{1-\delta}\left(\log_e j\right)^2\} \\
&= \frac{N}{\delta^2}+\frac{2}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2.
\end{aligned}$$

The Fisher information matrix is therefore

$$F(\mu,\delta)=\begin{bmatrix} -E\left(\frac{\partial^2 l}{\partial \mu^2}\right) & -E\left(\frac{\partial^2 l}{\partial \mu \partial \delta}\right) \\ -E\left(\frac{\partial^2 l}{\partial \delta \partial \mu}\right) & -E\left(\frac{\partial^2 l}{\partial \delta^2}\right) \end{bmatrix}$$

And therefore

$$F(\mu,\delta)=\begin{bmatrix} \frac{N}{\mu^2} & -\frac{1}{\mu}\left\{\frac{N}{\delta}+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right\} \\ -\frac{1}{\mu}\left\{\frac{N}{\delta}+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right\} & \frac{N}{\delta^2}+\frac{2}{\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2 \end{bmatrix}.$$

The Jeffreys' prior is

$$\begin{aligned}
P_J(\mu,\delta) &\propto \left|F(\mu,\delta)\right|^{\frac{1}{2}} \\
&= \left\{\frac{N^2}{\mu^2\delta^2}+\frac{2N}{\mu^2\delta}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j+\frac{N}{\mu^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2-\frac{1}{\mu^2}\left(\frac{N}{\delta}+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right)^2\right\}^{\frac{1}{2}} \\
&= \mu^{-1}\left\{N\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j\right)^2-\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log_e j\right)^2\right\}
\end{aligned}$$

Therefore

$$P_J(\mu,\delta)\propto \mu^{-1}.$$

B Proof of Theorem 8.1

Define

$$\begin{aligned}
A &= \begin{bmatrix} \frac{\partial \mu}{\partial t(\underline{\theta})} & \frac{\partial \mu}{\partial \delta} \\ \frac{\partial \delta}{\partial t(\underline{\theta})} & \frac{\partial \delta}{\partial \delta} \end{bmatrix} \\
&= \begin{bmatrix} \frac{-\delta}{t^2(\underline{\theta})}l^{\delta-1} & \frac{l^{\delta-1}}{t(\underline{\theta})}\left(1+\delta\log_e l\right) \\ 0 & 1 \end{bmatrix}.
\end{aligned}$$

The Fisher information matrix for $t(\underline{\theta})$ and δ is therefor

$$F\left(t\left(\underline{\theta}\right),\delta\right)=A'F\left(\mu,\delta\right)A=\begin{bmatrix}\tilde{F}_{11} & \tilde{F}_{12}\\ \\ \tilde{F}_{21} & \tilde{F}_{22}\end{bmatrix}$$

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

$$\begin{aligned}P_R\left(t\left(\underline{\theta}\right)\right) &= \left(F_{11\cdot 2}\right)^{\frac{1}{2}}=\left\{\tilde{F}_{11}-\tilde{F}_{12}\tilde{F}_{22}^{-1}\tilde{F}_{21}\right\}^{\frac{1}{2}}\\ &= \tfrac{1}{t(\underline{\theta})}\left[N-\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e j-\log_e l\right)\right\}^2\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\log_e l-\log_e j\right)^2\right\}^{-1}\right]^{\frac{1}{2}}.\end{aligned}$$

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.

C Proof of Theorem 8.2

As before $E(X_l) = \frac{\delta}{\mu} l^{\delta-1} = t(\underline{\theta})$.

$$F^{-1}(t(\underline{\theta}),\delta) = \begin{bmatrix} \tilde{F}^{11} & \tilde{F}^{12} \\ \tilde{F}^{21} & \tilde{F}^{22} \end{bmatrix} = F^{-1}(\underline{\theta})$$

and

$$\nabla_t'(\underline{\theta}) = \begin{bmatrix} \frac{\partial t(\underline{\theta})}{\partial t(\underline{\theta})} & \frac{\partial t(\underline{\theta})}{\partial \delta} \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Therefore

$$\nabla_t'(\underline{\theta}) F^{-1}(\underline{\theta}) = \begin{bmatrix} \tilde{F}^{11} & \tilde{F}^{12} \end{bmatrix}$$

and

$$\nabla_t'(\underline{\theta}) F^{-1}(\underline{\theta}) \nabla_t(\underline{\theta}) = \tilde{F}^{11}.$$

Define

$$\underline{\Upsilon}'(\underline{\theta}) = \frac{\nabla_t'(\underline{\theta}) F^{-1}(\underline{\theta})}{\sqrt{\nabla_t'(\underline{\theta}) F^{-1}(\underline{\theta}) \nabla_t(\underline{\theta})}} = \begin{bmatrix} \Upsilon_1(\underline{\theta}) & \Upsilon_2(\underline{\theta}) \end{bmatrix} = \begin{bmatrix} \sqrt{\tilde{F}^{11}} & \frac{\tilde{F}^{12}}{\sqrt{\tilde{F}^{11}}} \end{bmatrix}.$$

From this it follows that

$$\Upsilon_1(\underline{\theta}) = \sqrt{\tilde{F}^{11}} = \frac{t(\underline{\theta}) \left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\log_e j - \log_e l)^2 \right\}^{\frac{1}{2}}}{\sqrt{\tilde{A}}}$$

and

$$\Upsilon_2(\underline{\theta}) = \frac{\tilde{F}^{12}}{\sqrt{\tilde{F}^{11}}} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\log_e l - \log_e j)}{\sqrt{\tilde{} \left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\log_e l - \log_e j)^2 \tilde{A} \right\}}}$$

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.

$$P_M(t(\underline{\theta}),\delta) \propto \frac{1}{t(\underline{\theta})}$$

is a probability matching prior because it satisfies the differential equation

$$\frac{\partial}{\partial t(\underline{\theta})} \{ \Upsilon_1(\underline{\theta}) P_M(t(\underline{\theta}),\delta) \} + \frac{\partial}{\partial \delta} \{ \Upsilon_2(\underline{\theta}) P_M(t(\underline{\theta}),\delta) \} = 0.$$

Similar to the reference prior it follows that $P_M(\mu,\delta) \propto \mu^{-1}$.

D Proof of Theorem 14.1

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

$$F_{11} = diag \begin{bmatrix} \frac{n_1}{\mu_1^2} & \frac{n_2}{\mu_2^2} & \cdots & \frac{n_k}{\mu_k^2} \end{bmatrix},$$

$$F_{21} = -\left[\tfrac{1}{\mu_1} \left(\tfrac{n_1}{\delta^2} + \textstyle\sum_{j=1}^{n_1} \log_e j \right) \quad \tfrac{1}{\mu_2} \left(\tfrac{n_2}{\delta^2} + \textstyle\sum_{j=1}^{n_2} \log_e j \right) \quad \cdots \quad \tfrac{1}{\mu_k} \left(\tfrac{n_k}{\delta^2} + \textstyle\sum_{j=1}^{n_k} \log_e j \right) \right] = F_{21}'$$

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

$$
\begin{aligned}
\left|F\left(\mu_1, \mu_2, \ldots, \mu_k, \delta\right)\right| &= \left|F_{11}\right|\left|F_{22} - F_{21} F_{11}^{-1} F_{21}\right| \\
&= \left(\prod_{i=1}^{k} \frac{n_i}{\mu_i^2}\right)\left\{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\log_e j\right)^2 - \sum_{i=1}^{k} \frac{1}{n_1}\left(\sum_{j=1}^{n_i} \log_e j\right)^2\right\}
\end{aligned}
$$

The Jeffrey's prior follows as

$$
\begin{aligned}
P_J\left(\mu_1, \mu_2, \ldots, \mu_k, \delta\right) &\propto \left|F\left(\mu_1, \mu_2, \ldots, \mu_k, \delta\right)\right|^{\frac{1}{2}} \\
&= \prod_{i=1}^{k} \mu_i^{-1} \quad \mu_i > 0; i = 1, 2, \ldots, k.
\end{aligned}
$$