UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

RODRIGO CITTON PADILHA DOS REIS

ANÁLISE HIERÁRQUICA DE MÚLTIPLOS SISTEMAS REPARÁVEIS

Belo Horizonte
2014

RODRIGO CITTON PADILHA DOS REIS

ANÁLISE HIERÁRQUICA DE MÚLTIPLOS SISTEMAS REPARÁVEIS

Tese apresentada ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade Federal de Minas Gerais como parte dos requisitos para a obtenção do grau de Doutor em Estatística.

Orientador: Prof. Dr. Enrico A. Colosimo

Coorientador: Prof. Dr. Gustavo L. Gilardoni

Belo Horizonte
2014

Para Ayde e José Antônio,
por trinta anos de amor incondicional.
Para Laura,
pela energia de todos os dias.

Agradecimentos

Ao Professor Enrico A. Colosimo, orientador e amigo. Sou grato por todo incentivo e atenção que me dedicou desde a minha chegada em Belo Horizonte.

Ao Professor Gustavo L. Gilordoni, orientador. Seus ensinamentos foram fundamentais em minha formação. Sou grato também pela atenção dedicada nos momentos que estive em Brasília.

Aos professores do Programa de Pós-Graduação em Estatística da UFMG, pela valiosa contribuição em minha formação, em especial ao Bernardo, Denise, Luiz e Renato.

Aos funcionários do Departamento de Estatística da UFMG, em especial ao José Carlos, Kate, Maísa, Rogéria, Rose e Wilton.

Aos colegas do Programa de Pós-Graduação em Estatística da UFMG, pelo companheirismo, em especial ao Denis e a Maria Luiza.

Aos meus recentes colegas de trabalho do Projeto ELSA-Brasil, em especial a Sandhi e o Tom.

Aos meus amigos, que deram sentido a esta jornada: Fábio, Fernanda, Gustavo, José Luiz, Luís Gustavo, Márcia, Markus, Max, Paulo, Reinaldo e Wecsley.

Aos meus pais e irmãos, pelo apoio, carinho e torcida.

À Laura, o agradecimento mais especial, pelo carinho, amor, incentivo e por vibrar as energias. Sem a sua motivação diária, eu não teria chegado até aqui.

Ao apoio financeiro da FAPEMIG e CAPES.

À cidade de Belo Horizonte e ao povo mineiro.

Resumo

A tese tem como objetivo estudar determinados aspectos da modelagem hierárquica de dados de tempos de falha de múltiplos sistemas reparáveis. Mais especificamente, é abordado o modelo processo lei de potências hierárquico. O trabalho desenvolvido na tese está apresentado na forma de dois artigos, como segue.

Artigo 1: *Hierarchical modelling of power law processes for the analysis of repairable systems with different truncation times: An empirical Bayes approach*

Na análise de dados a partir de múltiplos sistemas reparáveis é usual observar tempos de truncamento diferentes e heterogeneidade entre os sistemas. Entre outras razões, a última é causada por diferentes linhas de fabricação e equipes de manutenção dos sistemas. Neste trabalho, um modelo hierárquico é proposto para a análise estatística dos múltiplos sistemas reparáveis sob diferentes tempos de truncamento. Uma reparametrização do processo de lei de potências é proposta a fim de se obter uma análise bayesiana semi conjugada. Uma abordagem *empirical Bayes* é utilizada para estimar os hiperparâmetros do modelo. A incerteza na estimativa destas quantidades é corrigida usando uma abordagem bootstrap paramétrica. Os resultados são ilustrados em um conjunto de dados reais de tempos de falha de transformadores de potência de uma empresa de energia elétrica no Brasil.

Palavras-chave: Amostragem por rejeição, Confiabilidade, Correção bootstrap, Máxima densidade a posteriori, Reparo mínimo.

Artigo 2: *Empirical Bayes and Jeffreys' prior for the hierarchical power law process*

Sã discutidos métodos alternativos para modelar o terceiro nível de um processo lei de potências hierárquico para modelar múltiplos sistemas reparáveis. É argumentado que a priori de Jeffreys tem vantagens com respeito a uma alternativa *empirical Bayes* e uma priori não informativa proposta na literatura. Mais especificamente, é mostrado por um estudo de simulação que as coberturas dos intervalos produzidos pelo método de Jeffreys são melhores que as coberturas dos intervalos produzidos pelos métodos *empirical Bayes* e não informativo. Os métodos também são ilustrados por análise de um conjunto de dados real.

Palavras-chave: Metropolis adaptativo, Modelo marginal, Múltiplos sistemas reparáveis,Reparo mínimo.

Abstract

The thesis has as objective to study certain aspects of the hierarchical modeling of data from several repairable systems. More specifically, the hierarchical power law process model is approached. The work developed in this thesis is presented in the form of two papers, as follows.

Paper 1: *Hierarchical modelling of power law processes for the analysis of repairable systems with different truncation times: An empirical Bayes approach*

In the data analysis from multiple repairable systems it is usual to observe both different truncation times and heterogeneity among the systems. Among other reasons, the latter is caused by different manufacturing lines and maintenance teams of the systems. In this paper, a hierarchical model is proposed for the statistical analysis of multiple repairable systems under different truncation times. A reparameterization of the power law process is proposed in order to obtain a quasi-conjugate bayesian analysis. An empirical Bayes approach is used to estimate model hyperparameters. The uncertainty in the estimate of these quantities are corrected by using a parametric bootstrap approach. The results are illustrated in a real data set of failure times of power transformers from an electric company in Brazil.

Keywords: Bootstrap correction, Maximum a posterior density, Minimal repair, Rejection sampling, Reliability.

Paper 2: *Empirical Bayes and Jeffreys' prior for the hierarchical power law process*

In this paper we discuss alternative methods to model the third stage of a hierarchical power law process for modelling of several repairable systems. We argue that the Jeffreys' prior has some advantages with respect to an empirical Bayes alternative or a noninformative prior proposed in the literature. More specifically, our simulations showed that the coverages of the intervals produced by the Jeffreys method are better than the interval coverages produced by empirical Bayes and noninformative methods. We also illustrate our methods with a real data set analysis.

Keywords: Adaptive Metropolis, Marginal model, Minimal repair, Multiple repairable systems.

Contents

Capítulo 1

Introdução

Seguindo a clássica definição de Ascher e Feingold (1984) um *sistema reparável* é um sistema que, quando uma falha ocorre, pode ser restaurado a uma condição operacional por algum processo de reparo que não seja a substituição de todo o sistema. Sendo assim, o que classifica um sistema como sendo reparável ou não reparável dependerá do custo da substituição. Por exemplo, dificilmente uma lâmpada será vista como um sistema reparável, já que o custo de sua substituição é relativamente baixo. Já o sistema de iluminação de um condomínio, ou de uma rua, pode ser visto como um sistema reparável, tendo em vista que o custo de substituição de todo o sistema é relativamente alto. No mundo real, diversos exemplos de sistemas reparáveis são encontrados, tais como carros, aviões, impressoras, sistemas de comunicação, dentre outros.

A ação de reparo realizada em um sistema reparável, quando este apresenta uma falha, possibilita a ocorrência de novas falhas. Tal estrutura de dados é conhecida na literatura estatística como *eventos recorrentes* ou *eventos repetidos*. Como observado por Peña (2006), eventos recorrentes não se restringem às áreas de confiabilidade e engenharia. Eles podem aparecer em diversas áreas, tais como saúde pública, biomedicina, economia, sociologia e política. Exemplos de eventos recorrentes nestas áreas podem ser: surto de uma doença, ocorrência de enxaqueca, queda de 200 pontos de um certo índice financeiro em um dia de negociação, superação de 70 pontos percentuais de aprovação de um governo, entre outros. Para uma discussão mais abrangente sobre eventos recorrentes veja Cook e Lawless (2007).

Provavelmente um dos objetivos mais básicos na análise de dados de sistemas reparáveis seja a distinção entre um "*sistema feliz*" e um "*sistema triste*", como já foi destacado no livro de Ascher e Feingold (1984). Os autores se referem a um sistema feliz como um sistema que ao envelhecer os tempos entre as falhas aumentam, de forma que o sistema parece estar melhorando. Já um sistema triste é visto como um sistema que ao envelhecer os tempos entre as falhas diminuem, e assim o sistema parece estar deteriorando. Ascher e Feingold (1984) ressaltam que o entendimento do processo de falhas não será alcançando através de uma análise que supõe que os tempos entre falhas sejam variáveis aleatórias independentes e identicamente distribuídas. Este entendimento só será possível com o uso de processos estocásticos.

Tipicamente, as suposições que fazemos sobre a maneira como um sistema envelhece e como ele é afetado por uma falha (e sua respectiva ação de reparo) irão conduzir a escolha de um modelo para dados de sistemas reparáveis. Um *reparo perfeito* (ou *de renovação*) significa que o reparo feito no sistema, faz com que este retorne a condição inicial de operação. Por outro lado, uma ação de *reparo mínimo* deixa o sistema na mesma condição à anterior a falha. Sendo assim, um sistema

que conforme envelhece apresenta maior número de falhas, ao receber uma ação de reparo perfeito, retorna a condição chamada de "tão bom quanto novo". Por sua vez, quando este mesmo sistema recebe uma ação de reparo mínimo, retorna a condição de "tão ruim quanto velho". Estas suposições acerca do tipo de reparo induzirão a distintos modelos, conforme será visto a seguir.

Outra questão relevante a ser considerada na modelagem de dados de sistemas reparáveis aparece quando analisamos dados de múltiplos sistemas. A heterogeneidade dos dados de diferentes sistemas é algo esperado de se observar. Tal variabilidade extra deve-se ao fato dos sistemas possuírem diferentes idades no momento inicial do acompanhamento, diferentes marcas de fabricação, diferentes equipes de manutenção, e por algumas vezes serem expostos a diferentes condições de operação. Esta característica deve ser incorporada ao modelo, seja por meio de covariáveis ou estruturas aleatórias que contemplem esta variabilidade extra.

Esta tese tem como objetivo estudar determinados aspectos da modelagem hierárquica de dados de tempos de falha de múltiplos sistemas reparáveis. O propósito das seguintes seções é introduzir de maneira sucinta alguns tópicos encontrados na análise de sistemas reparáveis e parte da teoria básica que embasou o conteúdo apresentado nos artigos.

1 Notação e conceitos básicos

Considere um sistema reparável em que o tempo usualmente parte de $t = 0$ e falhas ocorrem nos tempos $0 < T_1 < T_2 < \ldots$. Estes tempos de falha podem ser vistos como a realização de um processo pontual simples na reta real. O tempo não é necessariamente o tempo de calendário, mas pode ser o tempo de operação, número de ciclos, quilômetros rodados, etc. Quando uma falha ocorre, uma ação de reparo é feita para colocar o sistema de volta à operação. Nós fazemos a suposição de que os tempos de reparo são desprezíveis. Os tempos de falha geram um processo de contagem $\{N(t) : t \geq 0\}$, em que $N(t)$ conta as falhas no intervalo $(0, t]$. Uma terceira representação do mesmo processo pode ser dada pelos tempos entre falhas, definidos por $X_i = T_i - T_{i-1}$. Estes tempos entre falhas geralmente não são independentes nem identicamente distribuídos. A relação entre estas três representações do processo de contagem é mostrada na Figura 1.1.

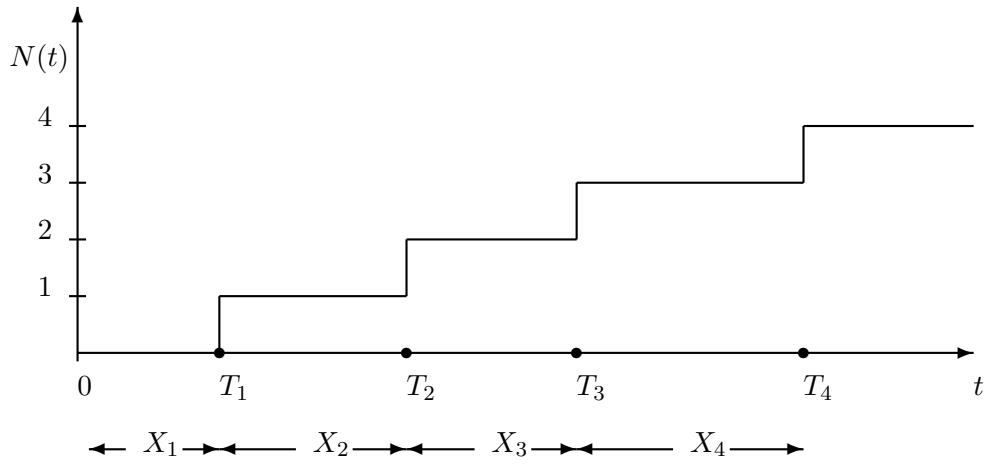

Figura 1.1: Relação entre as três representações do processo de contagem.

O seguinte teorema (ver Rigdon e Basu , 2000) formaliza a relação entre as três representações

do processo de contagem.

Teorema 1.1. *A função densidade de probabilidade conjunta de qualquer dos seguintes conjuntos de variáveis aleatórias determina a função densidade de probabilidade conjunta das outras.*

1. $N(u_1), N(u_2), \ldots, N(u_n)$ *para qualquer* n *e para qualquer* $u_1, u_2, \ldots, u_n$,
2. $T_1, T_2, \ldots, T_n$ *para qualquer* n,
3. $X_1, X_2, \ldots, X_n$ *para qualquer* n.

Se todo reparo é um reparo perfeito, então os tempos entre falhas são independentes e identicamente distribuídos (Rigdon e Basu , 2000). O *processo de renovação* é um modelo apropriado neste caso. A suposição de reparo mínimo induz ao *processo de Poisson não homogêneo* (PPNH), que é definido a seguir.

Definição 1.1 (Processo de Poisson)**.** *Um processo de contagem* $\{N(t) : t \geq 0\}$ *é dito ser um* processo de Poisson *se*

1. $N(0) = 0$.
2. *Para qualquer* $a < b \leq c < d$ *as variáveis aleatórias* $N(a, b]$ *e* $N(c, d]$ *são independentes.*
3. *Existe uma função* $\lambda(\cdot)$*, chamada a* função intensidade *do processo de Poisson, tal que*

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(N(t, t + \Delta t] = 1)}{\Delta t}.$$

4.

$$\lim_{\Delta t \to 0} \frac{\Pr(N(t, t + \Delta t] \geq 2)}{\Delta t} = 0.$$

Neste modelo a probabilidade de um sistema falhar durante $(t, t + \Delta t]$ é aproximadamente $\lambda(t)\Delta t$. Como um resultado da Definição 1.1, a variável aleatória $N(t)$ tem distribuição de Poisson com parâmetro $\Lambda(t) = \mathrm{E}\,(N(t)) = \int_0^t \lambda(u)du$, em que $\Lambda(t)$ é conhecida como a *função valor médio* do processo de Poisson. Note que um caso particular do PPNH é o processo de Poisson homogêneo (PPH), em que a função intensidade não depende de t, i.e., $\lambda(t) = \lambda$ é uma função constante. Neste caso, $\Lambda(t) = \lambda t$ e os tempos entre falhas são variáveis aleatórias independentes e identicamente distribuídas por uma distribuição exponencial de taxa λ. Note que a função intensidade do PPNH é uma *medida de confiabilidade* do sistema. Se esta é crescente, os tempos entre falhas estão diminuindo ao longo do tempo, e o sistema está deteriorando, e se esta é decrescente, os tempos entre falhas estão aumentando, e portanto, o sistema está melhorando.

Ao observarmos dados de um sistema reparável, dois esquemas de amostragem se destacam: *truncamento por falha* e *truncamento por tempo*. Quando o acompanhamento do sistema encerra após um número predeterminado de falhas, os dados são ditos serem truncados por falha. Por outro lado, os dados são ditos truncados por tempo quando o acompanhamento do sistema se encerra em um tempo τ preespecificado. Ao observarmos dados de um sistema truncado por tempo, tanto o número de falhas $N(\tau)$, quanto os tempos de falha $T_1, T_2, \ldots, T_{N(\tau)}$ são aleatórios. O seguinte teorema (ver Rigdon e Basu , 2000) é importante para especificarmos a distribuição conjunta de $N(\tau), T_1, T_2, \ldots, T_{N(\tau)}$.

Teorema 1.2 (Estatísticas de ordem)**.** *Se um PPNH com função intensidade $\lambda(t)$ é observado até o tempo τ, e se os tempos de falha são $T_1 < T_2 < \ldots < T_{N(\tau)}$, em que $N(\tau)$ é o número aleatório de falhas no intervalo $(0, \tau]$, então condicionado em $N(\tau) = n$, as variáveis aleatórias $T_1 < T_2 < \ldots < T_n$ são distribuídas como n estatísticas de ordem da distribuição com função distribuição acumulada*

$$G(t) = \frac{\Lambda(t)}{\Lambda(\tau)},\ 0 < t \leq \tau.$$

Assim, a distribuição conjunta do número de falhas e dos tempos de falha de um PPNH é obtida

$$\begin{aligned} p(n; t_1, \ldots, t_n) &= p(N(\tau) = n) p(t_1, \ldots, t_n | N(\tau) = n) \\ &= \frac{1}{n!} \left(\int_0^\tau \lambda(u) du \right)^n \exp\left\{ - \int_0^\tau \lambda(u) du \right\} \times n! \left\{ \prod_{j=1}^n G'(t_j) \right\} \\ &= \left\{ \prod_{j=1}^n \lambda(t_j) \right\} \exp\left\{ -\Lambda(\tau) \right\}, 0 < t_1 < t_2 < \ldots < t_n < \tau. \end{aligned} \tag{1.1}$$

Note que o PPNH é completamente especificado por sua função intensidade. Assim, quando formas paramétricas são especificadas para a função intensidade de um PPNH, estaremos interessados em fazer inferência a respeito dos parâmetros desta função. Uma das formas funcionais mais importantes e utilizadas é a intensidade *lei de potências*, que será discutida na Seção 2. Inferência baseada em verossimilhança no PPNH é feita utilizando a equação (1.1). Nesta tese consideramos apenas a inferência paramétrica. Para resultados recentes a respeito da inferência não paramétrica no PPNH veja Gámiz *et al.* (2011), Gilardoni e Colosimo (2011) e Gilardoni *et al.* (2013).

2 Processo lei de potências

Conforme comentado anteriormente, dentro da classe dos modelos PPNH, a forma paramétrica para função intensidade mais discutida na literatura é a intensidade *lei de potências*, dada por

$$\lambda(t) = \frac{\beta}{\theta} \left(\frac{t}{\theta} \right)^{\beta - 1},\ t > 0, \beta > 0, \theta > 0. \tag{1.2}$$

Um PPNH com função intensidade lei de potências é conhecido como um *processo lei de potências* (PLP). Na Equação (1.2), θ é um parâmetro de escala e β é um parâmetro de forma. Como pode ser visto na Figura 1.2, quando $\beta > 1$, a intensidade do processo é crescente no tempo, e quando $\beta < 1$, a intensidade é decrescente. Há ainda o caso em que $\beta = 1$, e assim $\lambda(t) = \theta^{-1}$, e temos o PPH como caso particular do PLP. Note que no PLP, a função do valor médio é $\Lambda(t) = (t/\theta)^\beta$.

A origem do PLP está ligada ao trabalho de Duane (1964), que através da análise de diversos sistemas, postulou que o número médio de falhas destes sistemas no intervalo $(0, t]$ deveria ser αt^β. Crow (1975) percebeu que o postulado de Duane poderia ser representado estocastimente por um PPNH com $\Lambda(t) = \alpha t^\beta$ e $\lambda(t) = \alpha\beta t^{\beta-1}$. Por muito tempo, o PLP foi referido como o *processo Weibull*, pois o tempo até a primeira falha do processo tem distribuição Weibull. Porém, o nome *processo lei de potências* é mais adequado, já que o leitor não faz confusão com o processo de renovação Weibull (tempos entre falhas independentes e identicamente distribuídos segundo a distribuição Weibull). A parametrização apresentada em 1.2, em que $\theta = \alpha^{-1/\beta}$, se deve a Finkelstein

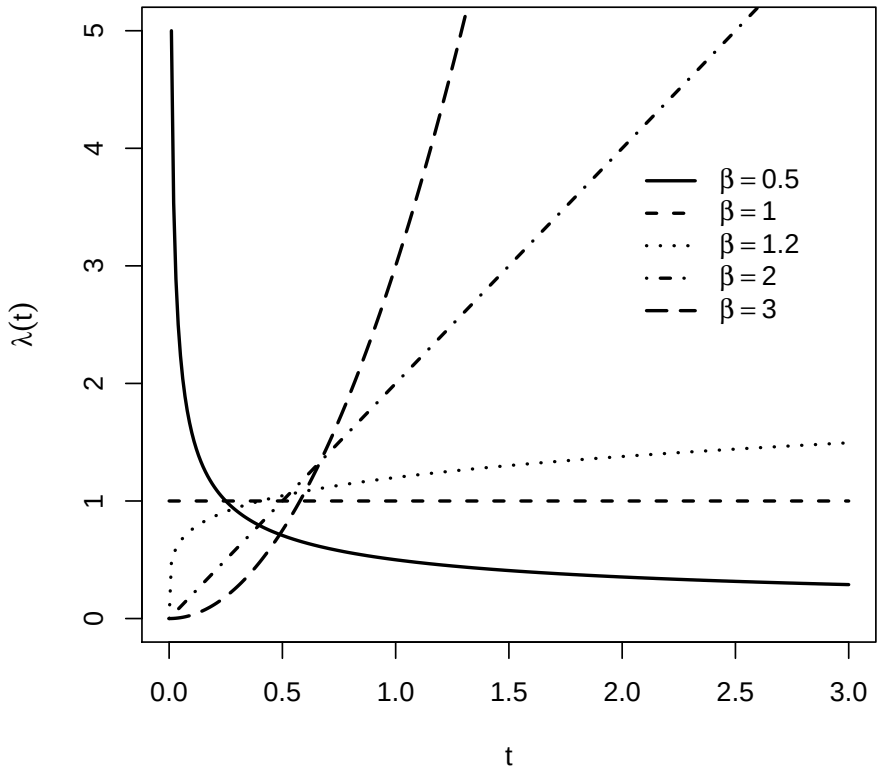

Figura 1.2: Função intensidade do PLP ($\theta = 1$ e diferentes valores de β).

(1976).

A simples formulação matemática e a riqueza em formas são provavelmente as principais razões para o vasto uso da intensidade lei de potências na análise de confiabilidade (veja, por exemplo, Kumar e Klefsjö (1992) e Coetzee (1997)). Ao substituir a expressão (1.2) em (1.1), obtemos a função de verossimilhança para (β, θ) como

$$p(n; t_1, \ldots, t_n|\beta, \theta) = \exp\left\{-(\tau/\theta)^\beta\right\} \frac{\beta^n}{\theta^{n\beta}} \prod_{j=1}^{n} t_j^{\beta-1}. \tag{1.3}$$

É fácil ver que, utilizando (1.3), os estimadores de máxima verossimilhança para (β, θ), são $\hat{\beta} = n/\sum_{j=1}^{n} \log(\tau/t_j)$ e $\hat{\theta} = \tau/n^{1/\hat{\beta}}$. Para uma ampla discussão acerca da inferência para o PLP veja Rigdon e Basu (2000). Alguns autores consideraram a abordagem bayesiana para fazer inferência nos parâmetros do PLP (veja, por exemplo Guida *et al.* , 1989; Bar-Lev *et al.* , 1992). Recentemente, Oliveira *et al.* (2012) consideraram uma interessante reparametrização do PLP em termos de β e $\eta = \Lambda(\tau) = (\tau/\theta)^\beta$. Como resultado desta reparametrização temos que a função de verossimilhança é proporcional ao produto de densidades gama. De forma mais direta, temos

$$p(n; t_1, \ldots, t_n|\beta, \eta) \propto \eta^n e^{-\eta} \times \beta^n e^{-\beta w},$$

em que $w = \sum_{j=1}^{n} \log(\tau/t_j)$. Oliveira *et al.* (2012) propuseram uma análise bayesiana conjugada para dados de um único sistema, e uma análise semi-conjugada no caso de analisar dados de múlti-

plos sistemas. Os autores ainda discutem a interpretação dos parâmetros e como eliciar informação a priori.

3 Modelagem hierárquica

Em problemas de confiabilidade, assim como em outras áreas, é comum observar um grupo de sistemas similares, porém não idênticos (Figura 1.3). Por exemplo, pode-se observar sistemas com diferentes intensidades de falha, apesar destes realizarem a mesma operação. Diferentes idades no início do acompanhamento, equipes de manutenção e locais de operação são possíveis razões para este comportamento heterogêneo dos sistemas. O problema estatístico está em combinar a informação dos diferentes sistemas para entender as mudanças no processo sob estudo. Se todos os sistemas observados são considerados ter a mesma intensidade, então a análise de múltiplos sistemas é semelhante a análise de um único sistema. Desta forma, os sistemas são considerados uma amostra aleatória do mesmo processo, e uma única intensidade é estimada. Caso haja evidência de que os sistemas realmente diferem, então uma análise individual de cada sistema será mais adequada. Uma formulação entre estas duas deve ser adequada para os casos intermediários (Gaver e O'Muircheartaigh , 1987). O cenário geral apresentado a seguir formaliza tal situação.

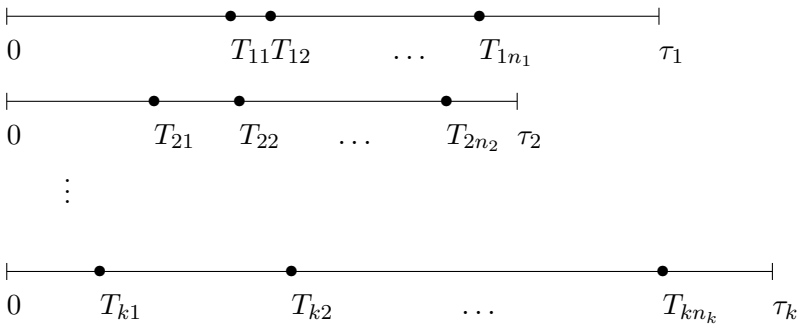

Figura 1.3: Histórico de falhas de um grupo de k sistemas reparáveis.

Considere um grupo de k sistemas que geram, de forma independente, falhas de acordo com um PLP de intensidade $\lambda_i(\cdot) = \lambda(\cdot|\mu_i)$ em que $\mu_i = (\beta_i, \theta_i)$ ou $\mu_i = (\beta_i, \eta_i)$, $i = 1, \ldots, k$. Para o i-ésimo sistema observamos n_i tempos de falha em um período τ_i de acompanhamento. Seja $D_i = (n_i; t_{i1}, \ldots, t_{i,n_i})$ o vetor que representa os dados de falha do i-ésimo sistema. A distribuição conjunta para $\mathbf{D} = (D_1, \ldots, D_k)$, os dados de falha dos k sistemas, é dada por

$$p(\mathbf{D}|\boldsymbol{\mu}) = \prod_{i=1}^{k} p(D_i|\mu_i),$$

em que $p(D_i|\mu_i)$ é a distribuição conjunta do número de falhas e dos tempos de falha do i-ésimo sistema dada pela expressão (1.3). Para descrever a variabilidade entre as intensidades, considere que $\lambda_i(\cdot)$ são realizações independentes de uma certa distribuição de probabilidade. De maneira mais específica, considere que os parâmetros μ_i da intensidade são distribuídos de acordo com $p(\boldsymbol{\mu}|\phi) = \prod_{i=1}^{k} p(\mu_i|\phi)$, em que $p(\mu_i|\phi)$ é uma densidade de probabilidade e ϕ é um vetor de

hiperparâmetros. Tal formulação é conhecida na literatura como modelo hierárquico. Com ϕ fixado, usando o teorema de Bayes, a distribuição a posteriori condicional é dada por

$$p(\boldsymbol{\mu}|\mathbf{D},\phi) = \frac{p(\mathbf{D}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\phi)}{\int p(\mathbf{D}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\phi)d\boldsymbol{\mu}} = \frac{p(\mathbf{D}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\phi)}{p(\mathbf{D}|\phi)}, \tag{1.4}$$

em que $p(\mathbf{D}|\phi)$ denota a distribuição marginal de $\mathbf{D}$.

Quando ϕ é desconhecido, duas abordagens concorrentes se destacam na literatura: a abordagem *empirical Bayes* e a *completamente bayesiana*. Na análise *empirical Bayes* se utiliza a distribuição marginal para estimar o vetor ϕ por $\hat{\phi}$, geralmente por máxima verossimilhança. Uma vez que o vetor ϕ foi estimado, procede-se a análise utilizando a distribuição a posteriori condicional (1.4) com $\phi = \hat{\phi}$. Já na análise *completamente bayesiana*, a incerteza com respeito a ϕ é descrita através de uma distribuição de probabilidade, e a distribuição posteriori é dada por

$$p(\boldsymbol{\mu}|\mathbf{D}) = \frac{\int p(\mathbf{D}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\phi)p(\phi)d\phi}{\int\int p(\mathbf{D}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\phi)p(\phi)d\boldsymbol{\mu}d\phi} = \int p(\boldsymbol{\mu}|\mathbf{D},\phi)p(\phi|\mathbf{D})d\phi. \tag{1.5}$$

Note que a distribuição a posteriori (1.5) é uma média da distribuição a posteriori condicional (1.4) com respeito a distribuição a posteriori marginal de ϕ. Ambas abordagens utilizam os dados observados para obter informação de ϕ e então "combinar evidência" (Carlin e Louis , 2000). Apesar das semelhanças (veja, por exemplo, Singpurwalla , 1989), cada abordagem possui suas próprias dificuldades. Se por um lado o *empirical Bayes* "evita" a especificação de um modelo a priori para ϕ, geralmente uma tarefa difícil, este falha ao ignorar a incerteza com respeito a estimação de ϕ. Sob o ponto de vista que o *empirical Bayes* é uma aproximação à abordagem *completamente bayesiana*, ajustes foram propostos na literatura (Laird e Louis , 1987; Kass e Steffey , 1989; Carlin e Gelfand , 1990).

Quando a função lei de potências é considerada na especificação da intensidade do PPNH, temos como resultado o *modelo PLP hierárquico*. O trabalho de Engelhardt e Bain (1987) parece ter sido o primeiro esforço neste sentido, e considera o parâmetro de escala variando entre os sistemas, enquanto o parâmetro de forma é o mesmo para todos os sistemas. Considerando uma classe mais geral de processos de contagem para modelagem de sistemas reparáveis, Lindqvist *et al.* (2003) introduziram um termo de *fragilidade* no modelo. Porém, no caso particular do PLP, isto é equivalente a considerar apenas o parâmetro de escala variando entre os sistemas.

Ryan *et al.* (2011) propuseram um modelo PLP hierárquico em que tanto o parâmetro de escala quanto o parâmetro de forma podem variar entre os sistemas. Uma abordagem completamente bayesiana foi adotada, e a reparametrização do PLP em termos de (β, η) foi adotada. Os autores utilizaram uma distribuição a priori produto de gamas para os parâmetros do PLP, que induziu uma distribuição a posteriori condicional conjugada. Por fim, uma distribuição a priori vaga, ou aproximadamente não informativa, para os hiperâmetros do modelo foi especificada.

4 Organização da tese

O Capítulo 2 apresenta o artigo *Hierarchical modelling of power law processes for the analysis of repairable systems with different truncation times: An empirical Bayes approach*. A modelagem hierárquica de múltiplos sistemas com tempos de truncamentos diferentes é discutida. A repara-

metrização do PLP em termos de (β, η) é utilizada. Como η depende do tempo de truncamento, não é razoável considerar os pares (β_i, η_i) intercambiáveis. A modelagem do segundo nível do modelo PLP hierárquico leva em consideração esta dificuldade. É proposto uma abordagem *empirical Bayes* com correção bootstrap para estimação de intervalos dos parâmetros do PLP. Uma análise semi conjugada é obtida e métodos eficientes de simulação da distribuição a posteriori condicional são discutidos. Os métodos são ilustrados em um conjunto de dados real. Este artigo foi submetido para o periódico *IIE Transactions*.

O Capítulo 3 apresenta o artigo *Empirical Bayes and Jeffreys' prior for the hierarchical power law process*. É discutida a modelagem dos hiperparâmetros do modelo PLP hierárquico. Alternativas à proposta de Ryan *et al.* (2011) são apresentadas, dentre elas, a derivação da priori de Jeffreys para este modelo. Métodos estocásticos para a aproximação da distribuição a posteriori resultante são discutidos. Os métodos são comparados através de um estudo de simulação e análise de um conjunto de dados real.

Referências

Ascher e Feingold (1984) H. Ascher e H. Feingold. *Repairable systems reliability: modelling, inference, misconceptions and their causes*. Marcel Dekker Inc, New York.

Bar-Lev *et al.* **(1992)** S. Bar-Lev, I. Lavi e B. Reiser. Bayesian inference for the power law process. *Annals of the Institute of Statistical Mathematics*, 44(4):623–639. URL http://EconPapers.repec.org/RePEc:spr:aistmt:v:44:y:1992:i:4:p:623-639.

Carlin e Gelfand (1990) B. P. Carlin e A. E. Gelfand. Approaches for the empirical Bayes confidence intervals. *Journal of the American Statistical Association*, 85:105–114. doi: 10.1080/01621459.1990.10475312.

Carlin e Louis (2000) B. P. Carlin e T. A. Louis. Empirical bayes: Past, present and future. *Journal of the American Statistical Association*, 95(452):1286–1289. doi: 10.1080/01621459.2000.10474331. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.2000.10474331.

Coetzee (1997) J. L. Coetzee. The role of NHPP models in the practical analysis of maintenance failure data. *Reliability Engineering and System Safety*, 56:161–168.

Cook e Lawless (2007) R. J. Cook e J. F. Lawless. *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health. Springer.

Crow (1975) L. H. Crow. Reliability analysis for complex repairable systems. Relatório técnico, US Army Materiel Systems Analysis Activity.

Duane (1964) J.T. Duane. Learning curve approach to reliability monitoring. *Aerospace, IEEE Transactions on*, 2(2):563–566. ISSN 0536-1516. doi: 10.1109/TA.1964.4319640.

Engelhardt e Bain (1987) M. Engelhardt e L.J. Bain. Statistical analysis of a compound power-law model for repairable systems. *Reliability, IEEE Transactions on*, R-36(4):392–396. ISSN 0018-9529. doi: 10.1109/TR.1987.5222421.

Finkelstein (1976) J. M. Finkelstein. Confidence bounds on the parameters of the weibull process. *Technometrics*, 18(1):115–117. doi: 10.1080/00401706.1976.10489408. URL http://www.tandfonline.com/doi/abs/10.1080/00401706.1976.10489408.

Gaver e O'Muircheartaigh (1987) D. P. Gaver e I. G. O'Muircheartaigh. Robust empirical Bayes analyses of event rates. *Technometrics*, 29:1–15. doi: 10.2307/1269878.

Gilardoni e Colosimo (2011) G. L. Gilardoni e E. A. Colosimo. On the superposition of overlapping Poisson processes and nonparametric estimation of their intensity function. *Journal of Statistical Planning and Inference*.

Gilardoni *et al.* **(2013)** G. L. Gilardoni, M. D. de Oliveira e E. A. Colosimo. Nonparametric estimation and bootstrap confidence intervals for the optimal maintenance time of a repairable system. *Computational Statistics & Data Analysis*, 63(0):113 – 124. ISSN 0167-9473. doi: http://dx.doi.org/10.1016/j.csda.2013.02.006. URL http://www.sciencedirect.com/science/article/pii/S0167947313000509.

Gámiz *et al.* **(2011)** M.L. Gámiz, Kulasekera, K.B., N. Limnios e B.H. Lindqvist. *Applied Nonparametric Statistics in Reliability*. Springer.

Guida *et al.* **(1989)** M. Guida, R. Calabria e G. Pulcini. Bayes inference for a non-homogeneous Poisson process with power intensity law. *IEEE Transactions on Reliability*, 38:603–609. doi: 10.1109/24.46489.

Kass e Steffey (1989) R. E. Kass e D. Steffey. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, 84:717–726. doi: 10.1080/01621459.1989.10478825.

Kumar e Klefsjö (1992) U. Kumar e B. Klefsjö. Reliability analysis of hidraulic systems of LHD machines using the power law process model. *Reliability Engineering and System Safety*, 35:217–224.

Laird e Louis (1987) N. M. Laird e T. A. Louis. Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association*, 82:739–750. doi: 10.2307/2288778.

Lindqvist *et al.* **(2003)** B. H. Lindqvist, G. Elvebakk e K. Heggland. The trend-renewal process for statistical analysis of repairable systems. *Technometrics*, 45:31–44.

Oliveira *et al.* **(2012)** M. D. Oliveira, E. A. Colosimo e G. L. Gilardoni. Bayesian inference for power law processes with applications in repairable systems. *Journal of Statistical Planning and Inference*, 142:1151–1160. doi: 10.1016/j.jspi.2011.11.016.

Peña (2006) E. A. Peña. Dynamic modeling and statistical analysis of event times. *Statistical Science*, 21(4):487–500. doi: 10.1214/088342306000000349. URL http://dx.doi.org/10.1214/088342306000000349.

Rigdon e Basu (2000) S. E. Rigdon e A. P. Basu. *Statistical Methods for the Reliability of Repairable Systems*. John Wiley & Sons.

Ryan *et al.* **(2011)** K. J. Ryan, M. S. Hamada e C. S. Reese. A Bayesian hierarchical power law process model for multiple repairable systems with an application to supercomputer reliability. *Journal of Quality Technology*, 43:209–223.

Singpurwalla (1989) N. Singpurwalla. A unifying perspective on statistical modeling. *SIAM Review*, 31:560–564.

Capítulo 2

Hierarchical modelling of power law processes for the analysis of repairable systems with different truncation times: An empirical Bayes approach

Rodrigo Citton P. dos Reis, Enrico A. Colosimo, and Gustavo L. Gilardoni

Abstract

In the data analysis from multiple repairable systems it is usual to observe both different truncation times and heterogeneity among the systems. Among other reasons, the latter is caused by different manufacturing lines and maintenance teams of the systems. In this paper, a hierarchical model is proposed for the statistical analysis of multiple repairable systems under different truncation times. A reparameterization of the power law process is proposed in order to obtain a quasi-conjugate bayesian analysis. An empirical Bayes approach is used to estimate model hyperparameters. The uncertainty in the estimate of these quantities are corrected by using a parametric bootstrap approach. The results are illustrated in a real data set of failure times of power transformers from an electric company in Brazil.

Keywords: *Bootstrap correction, Maximum a posterior density, Minimal repair, Multiple repairable systems, Rejection sampling, Reliability*

1 Introduction

An issue of interest to statisticians and engineers in the analysis of repairable systems data is how to model the changes in the performance of the system caused by the failure and/or maintenance

Submetido para *IIE Transactions* em 3 de outubro de 2014.

process. This involves usually a stochastic point process (Andersen et al., 1993; Cook and Lawless, 2007) and statistical analysis (Rigdon and Basu, 2000; Lindqvist, 2006). In the data from multiple repairable systems one observes usually different truncation times and heterogeneity among them. The latter is due to causes such as different locations, manufacturing lines and maintenance teams of the systems, among others. An interesting example of the joint presence of heterogeneity and different truncation times is provided by the power transformers of the electric company of Minas Gerais state in Brazil. These data were first reported and analyzed by Gilardoni and Colosimo (2007). Table 2.1 contains failure times from forty power transformers, recorded between January 1999 and July 2001. The data consist of the number of failures and failure and truncation times for the forty systems.

Table 2.1: Power transformers data.

System	Number of failures	Failure times (hours)		Trucation times	System	Number of failures	Failure times (hours)		Trucation times
1	2	8,839	17,057	21,887	17	1	15,524		21,886
2	2	9,280	16,442	21,887	18	0			21,440
3	1	10,445		13,533	19	0			369
4	0			7,902	20	2	11,664	17,031	21,857
5	0			8,414	21	0			7,544
6	0			13,331	22	0			6,039
7	1	17,156		21,887	23	1	2,168		6,698
8	1	16,305		21,887	24	1	18,840		21,879
9	1	16,802		21,887	25	0			2,288
10	0			4,881	26	0			2,499
11	0			16,625	27	1	10,668		16,838
12	2	7,396	7,541	19,590	28	1	15,550		21,887
13	0			2,121	29	0			1,616
14	2	15,821	19,746	19,877	30	1	14,041		20,004
15	0			1,927	31 - 40	0			21,888
16	1	15,813		21,886					

Power transformers are complex systems with a large number of components. These devices usually fail because of just one of these components. After this component is repaired, it is expected that the reliability of the transformer does not change. This type of repair is known as *minimal repair*. A failure process that undergoes minimal repair actions is modeled by a nonhomogeneous Poisson process (NHPP) (Baker, 1996). Succinctly, define $N(t)$ to be the number of failures in the interval $(0, t]$. A process $\{N(t) : t \geq 0\}$ having independent increments and starting at $N(0) = 0$ is said to be a Poisson process with intensity $\lambda(\cdot)$ if, for any t, the random variable $N(t)$ follows a Poisson distribution with mean $\Lambda(t) = \int_0^t \lambda(u)du$. The NHPP is a Poisson process with a nonconstant intensity function $\lambda(\cdot)$. In the repairable system literature, the most popular parametric form for λ is the power law process (PLP),

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}, \tag{2.1}$$

where β and θ are positive parameters. The corresponding mean function is

$$\Lambda(t) = \mathrm{E}\left[N(t)\right] = \int_0^t \lambda(u)\, du = \left(\frac{t}{\theta}\right)^{\beta}. \tag{2.2}$$

The popularity of the PLP model stems from both its mathematical simplicity and its flexibility, in the sense that (2.1) can accommodate situations where the systems either deteriorates ($\beta > 1$) or improves ($\beta < 1$) with time.

When observing data from a single system truncated at τ, the joint likelihood of the number of failures $n = N(\tau)$ and the failure times $0 < t_1 < \cdots < t_n < \tau$ is obtained after noting that $N(\tau)$ follows a Poisson distribution with mean $\Lambda(\tau)$ and, conditional on $N(\tau) = n$, the failure times have the same distribution as the order statistics of a sample of size n from the pdf $g(t) = [\lambda(t)/\Lambda(\tau)]\, I(0 < t < \tau)$, which in the PLP case becomes $g(t) = (\beta/\tau)(t/\tau)^{\beta}\, I(0 < t < \tau)$ (see, for instance, Rigdon and Basu, 2000). Therefore,

$$p(n; t_1, \ldots, t_n \,|\, \beta, \theta) = \exp\{-(\tau/\theta)^{\beta}\}\, \frac{\beta^n}{\theta^{n\beta}} \prod_{j=1}^{n} t_j^{\beta-1}\,. \tag{2.3}$$

(As usual, we assume here and throughout that empty sums and products are equal respectively to zero and one, so that (2.3) becomes $\exp\{-(\tau/\theta)^{\beta}\}$ when $n = 0$.) If we reparametrize the model in terms of β and $\eta = \mathrm{E}\,[N(\tau)] = (\tau/\theta)^{\beta}$, the likelihood (2.3) becomes

$$p(n; t_1, \ldots, t_n \,|\, \beta, \eta) \propto \gamma(\eta \,|\, n+1, 1)\, \gamma(\beta \,|\, n+1, w)\,, \tag{2.4}$$

where $w = \sum_{j=1}^{n} \log(\tau/t_j)$ and $\gamma(x \,|\, a, b) = b^a\, x^{a-1} e^{-bx}/\Gamma(a)$ is the density of the gamma distribution with mean a/b and variance a/b^2. The fact that β and η are orthogonal and the striking simplicity of (2.4) makes the (β, η) parameterization quite convenient. It has been used previously by Oliveira et al. (2012) in nonhierarchical modelling and Ryan et al. (2011) in the context of hierarchical models when all the truncation times are equal. Using either (2.3) or (2.4) it is easy to show then that the maximum likelihood estimates (MLEs) are $\hat{\eta} = n$ and, provided that $n > 0$, $\hat{\beta} = n/\sum_{j=1}^{n} \log(\tau/t_j) = n/w$ and $\hat{\theta} = \tau/n^{1/\hat{\beta}}$ (the MLEs of β and θ do not exist when $n = 0$). We note that, in the sequel, we will denote (2.3) by writing that $(n; t_1, \ldots, t_n) \sim PLP_{\tau}(n; t_1, \ldots, t_n \,|\, \beta, \theta)$.

An important aspect to consider regarding the power transformers data in Table 2.1 is the fact that these systems are located in different places along the Brazilian state of Minas Gerais. Thus, due to climate changes along this state, it is expected that they are exposed to different operating conditions. Therefore, rather than assuming that all 40 systems have the same (β, θ) parameters as in Oliveira et al (2012), an individual analysis of each system may be adequate. In other words, one may compute estimates $(\hat{\beta}_i, \hat{\theta}_i)$ for each of the 16 systems having $n_i > 0$. Figure 2.1 shows estimates for the intensity and mean functions (2.1) and (2.2) obtained by substituting the parameters by its MLEs. One can observe that the estimated intensities show quite different behavior (decreasing, concave increasing and convex increasing). While this may be because each system has its unique characteristics, it is more likely the consequence of the fact that the individual estimates are highly inaccurate because the number of observed failures for each system is very small. On the other hand, a hierarchical analysis which considers that the systems are somewhat *similar* may be more realistic and, at the same time, it would allow the estimates for the individual systems to *borrow strength* from the observed features of the other systems and hence help to improve accuracy.

The objective of this paper is to discuss a hierarchical model to analyze several repairable systems truncated at possible different times. More precisely, the first stage specifies a distribution

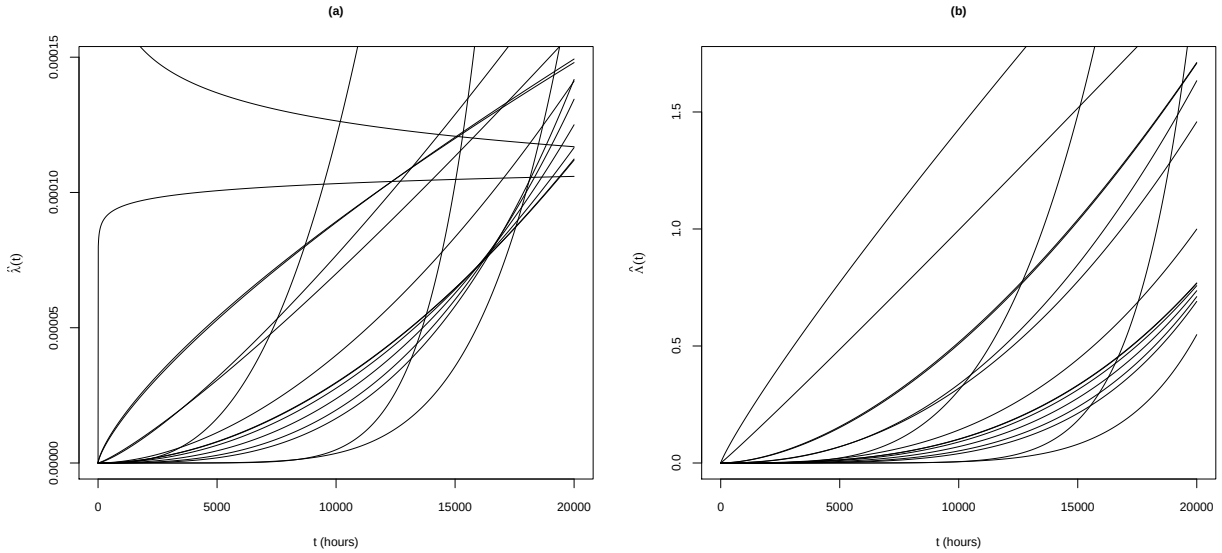

Figure 2.1: Maximum likelihood estimates of the intensity **(a)** and mean **(b)** functions for the sixteen transformers with $n_i > 0$.

for the failure times data conditional on the parameters of the PLP, while the second stage specifies a prior distribution for these parameters. Therefore, the specific features of each transformer are modeled in the first stage, while characteristics that are common to all transformers are taken into consideration in the second one. Although there has been some recent interest in the area of hierarchical modeling of repairable systems (see for instance Bhattacharjee et al., 2003; Pan and Rigdon, 2009; Ryan et al., 2011), statistical modeling and inference procedures for the case of multiple repairable systems with different truncation times are still under consideration in the literature. Lindqvist et al. (2003) considered the issue of heterogeneity between systems for more general counting process by introducing a frailty term in the model, although it affects only the scale parameter of the PLP intensity (see Lawless, 1987). Our model allows both the scale and shape parameters to vary among systems. Following Guida and Pulcini (2005), Giorgio et al. (2014) used a generalization of the prior proposed by Huang (2001) to model shape and scale parameter of the PLP intensity. The resulting prior depends upon five hyperparameters, one more than our prior model. Furthermore, their approach differs from ours in the sense that they estimate the five hyperparameters using the actual data to elicit an informative prior for a future analysis.

The rest of the paper is organized as follows. Section 2 describes the hierarchical model with special focus on the second stage distribution. More precisely, we argue that the (β, η) parameterization together with different truncation times implies that one cannot assume *exchangeability* and suggest a way to overcome this difficulty. Section 3 discusses an empirical Bayes strategy based on maximum posterior density or, equivalently, penalized likelihood estimation for the hyperparameters and, once that the hyperparameters have been estimated, an efficient rejection sampling strategy to obtain *iid* samples from the posterior distribution of the system-specific parameters . Section 3 also presents an implementation of a bootstrap procedure, suggested by (Laird and Louis, 1987), to correct for the underestimation of uncertainty inherent to the empirical Bayes approach. Section 4 contains an analysis of the power transformers data set, including estimation of the optimal maintenance period under a block maintenance policy. Finally, some conclusions are given in Section 5 and Appendix A describes how to obtain starting values for the penalized likelihood

maximization used to estimate the hyperparameters.

2 A hierarchical PLP model

As mentioned above, we follow Guida et al. (1989), Oliveira et al. (2012) and Ryan et al. (2011) and parametrize the PLPs in terms of β_i and $\eta_i = \Lambda_i(\tau_i) = (\tau_i/\theta_i)^{\beta_i}$, mainly in view of the simplifications that result from (2.4) and the consequent orthogonality. Of course, it is possible to go from one parameterization to the other provided that one multiplies both prior and posteriors by the appropriate jacobian.

Let $D_i = (n_i; t_{i1}, \ldots, t_{i,n_i})$, $\boldsymbol{D} = (D_1, \ldots, D_K)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)$. Assuming all throughout conditional independence across systems, the *data level* of the hierarchical model states that

$$p(\boldsymbol{D} \,|\, \boldsymbol{\beta}, \boldsymbol{\eta}) \propto \prod_{i=1}^{K} \gamma(\eta_i \,|\, n_i + 1, 1) \,\times\, \gamma(\beta_i \,|\, n_i + 1, w_i)\,, \tag{2.5}$$

where $w_i = \sum_{j=1}^{n_i} \log(\tau_i/t_{ij})$. In other words, data from the i-th system comes from a PLP with parameters β_i and $\theta_i = \tau_i\, \eta_i^{-1/\beta_i}$ observed up to time τ_i [cf. equations (2.3) and (2.4)]. To specify the *prior level* of the model we denote by $\boldsymbol{\phi} = (a_\beta, \beta_0, a_\eta, \theta_0)$ the set of hyperparameters and let

$$p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{\phi}) = \prod_{i=1}^{K} \gamma(\beta_i \,|\, a_\beta, a_\beta/\beta_0) \times \gamma(\eta_i \,|\, a_\eta, a_\eta(\theta_0/\tau_i)^{\beta_i})\,. \tag{2.6}$$

More specifically, we set β_i to follow a gamma distribution with mean β_0 and coefficient of variation $1/\sqrt{a_\beta}$ and, conditional on β_i, η_i follows also a gamma distribution with mean $(\tau_i/\theta_0)^{\beta_i}$ and coefficient of variation $1/\sqrt{a_\eta}$, so that β_0 and θ_0 can be thought off as prior guesses for the β_i's and the θ_i's and a_β and a_η are hyperparameters that control the precision of those prior guesses.

The rationale behind the prior distribution (2.6) can be explained as follows. We begin by noting that it follows from (2.4) that, in the case of a single system, the natural prior for the pair (β, η) is a product of gamma distributions of the form $\gamma(\beta \,|\, a_\beta, a_\beta/\beta_0) \times \gamma(\eta \,|\, a_\eta, a_\eta/\eta_0)$ (cf. Oliveira et al., 2012). Following this idea, Ryan et al. (2011) consider a hierarchical model for several PLPs all truncated at the same time $\tau_1 = \ldots = \tau_K = \tau$ and specify the prior level distribution also as a product of gamma distributions of the form $\prod_{i=1}^{K} \gamma(\beta_i \,|\, a_\beta, a_\beta/\beta_0) \times \gamma(\eta_i \,|\, a_\eta, a_\eta/\eta_0)$. However, this possibility does not seem appropriate when the systems have different truncation times, in the sense that it would imply that the pairs (β_i, η_i) $(i = 1, \ldots, K)$ are *exchangeable*, while one would expect larger values of $\eta_i = \mathrm{E}\,[N_i(\tau_i)]$ for those systems which are observed longer (i.e. which have large τ_i). Although assuming the η_i's to be exchangeable is not reasonable because their definition involves the τ_i's, which are different, it makes sense to assume that the θ_i's are exchangeable irrespective of the truncation times, because their definition (namely, θ_i is the time such that $\mathrm{E}\,[N_i(\theta_i)] = 1$) does not involve the τ_i's. Therefore, we want the prior level distribution $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{\phi})$ to be such that the pairs $(\beta_i, \theta_i = \tau_i\, \eta_i^{-1/\beta_i})$ are exchangeable. Now, it is straightforward to check that (2.6) implies that

$$p(\boldsymbol{\beta}, \boldsymbol{\theta} \,|\, \boldsymbol{\phi}) = \prod_{i=1}^{K} \gamma(\beta_i \,|\, a_\beta, a_\beta/\beta_0) \,\times\, \frac{a_\eta^{a_\eta}}{\Gamma(a_\eta)} \, \frac{\beta_i}{\theta_i} \left(\frac{\theta_0}{\theta_i}\right)^{a_\eta \beta_i} \exp\{-a_\eta(\theta_0/\theta_i)^{\beta_i}\}\,,$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$. Since the truncation times τ_i do not appear in the right hand side of this

last expression, this implies that the pairs (β_i, θ_i) are indeed exchangeable.

An alternative derivation of (2.6) is as follows. Write $p(\beta_i, \eta_i \,|\, \boldsymbol{\phi}) = p(\beta_i \,|\, \boldsymbol{\phi}) p(\eta_i | \beta_i, \boldsymbol{\phi})$ and suppose that one wants to set $\beta_i \,|\, \boldsymbol{\phi} \sim \text{Gamma}(a_\beta, a_\beta/\beta_0)$ and $\eta_i \,|\, \beta_i, \boldsymbol{\phi} \sim \text{Gamma}(a_\eta, b_\eta)$, where a_η and b_η could possibly depend on β_i and τ_i. Then the β_i's are exchangeable and a necessary condition for the pairs (β_i, θ_i) to be exchangeable is that $\text{E}\,[\theta_i^{-\beta_i} \,|\boldsymbol{\phi}]$ does not depend on the system i. Now, since $\theta_i^{-\beta_i} = \tau_i^{-\beta_i}\, \eta_i$,

$$\text{E}\,[\theta_i^{-\beta_i} \,|\boldsymbol{\phi}] = \text{E}\,\left[\text{E}\,[\tau_i^{-\beta_i}\, \eta_i \,|\, \beta_i, \boldsymbol{\phi}]\right] = \text{E}\,[\tau_i^{-\beta_i}\, (a_\eta/b_\eta) \,|\, \boldsymbol{\phi}]\,.$$

It is easy to see that for this not to depend on τ_i, it is necessary that there exists a function h such that $\text{E}\,[\tau_i^{-\beta_i}\, (a_\eta/b_\eta) \,|\, \boldsymbol{\phi}] = h(\beta_i)$. The prior $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{\phi})$ given in (2.6) corresponds to the choice $h(\beta_i) = \theta_0^{-\beta_i}$. In other words, the previous argument shows that for the prior (2.6) one has that $\text{E}\,[\theta_i^{-\beta_i} \,|\boldsymbol{\phi}] = \text{E}\,[\theta_0^{-\beta_i} | \boldsymbol{\phi}]$, showing again why θ_0 can be thought of as a prior guess for the θ_i's.

To complete the specification of the hierarchical model, we assume an independent prior distribution for the hyperparameters of the form

$$p(\boldsymbol{\phi}) = p(a_\beta) \times p(\beta_0) \times p(a_\eta) \times p(\theta_0) \propto \exp\{-\xi_1 a_\beta\}\, \exp\{-\xi_2 a_\eta\}\,, \tag{2.7}$$

i.e., we set both $p(\beta_0) \propto 1$ and $p(\theta_0) \propto 1$ and exponential densities with means ξ_1^{-1} and ξ_2^{-1} respectively for a_β and a_η. The exponential distribution is a common choice for the shape parameter of the Gamma-Poisson hierarchical model (see for example George et al. (1993), and related applications Pérez et al. (2006); Pesaran et al. (2006); Perkins et al. (2012)), that can be thought as a prototype for the PLP hierarchical model. In Section 3 we discuss the specification of ξ_1 and ξ_2.

In the rest of the paper we discuss an empirical Bayes procedure which estimates $\boldsymbol{\phi}$ from data by maximizing the posterior density $p(\boldsymbol{\phi}|\boldsymbol{D})$ or, equivalently, by maximizing a penalized likelihood (see Section 3 and Appendix A). Once that an estimate $\hat{\boldsymbol{\phi}}$ has been obtained, inferences about quantities specific to each system proceeds straightforward after noting from (2.5) and (2.6) that

$$p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi}) = \prod_{i=1}^{K} p(\eta_i \,|\, \beta_i, D_i, \boldsymbol{\phi}) \times p(\beta_i \,|\, D_i, \boldsymbol{\phi}), \tag{2.8}$$

where

$$p(\eta_i \,|\, \beta_i, D_i, \boldsymbol{\phi}) = \gamma(\eta_i \,|\, a_\eta + n_i, a_\eta\, (\theta_0/\tau_i)^{\beta_i} + 1), \tag{2.9}$$

and

$$p(\beta_i \,|\, D_i, \boldsymbol{\phi}) \propto \gamma(\beta_i \,|\, a_\beta + n_i, a_\beta/\beta_0 + w_i) \;\times\; \frac{[a_\eta(\theta_0/\tau_i)^{\beta_i}]^{a_\eta}}{[a_\eta(\theta_0/\tau_i)^{\beta_i} + 1]^{a_\eta + n_i}}. \tag{2.10}$$

3 Empirical Bayes inference for the hierarchical PLP model

To make inferences for the hierarchical PLP model we adopt a parametric empirical Bayes (PEB) approach. The PEB approach uses the observed data to estimate, usually by the maximum likelihood method, the hyperparameters $\boldsymbol{\phi} = (a_\beta, \beta_0, a_\eta, \theta_0)$. Then, one replaces $\boldsymbol{\phi}$ by its estimate $\hat{\boldsymbol{\phi}}$ in the conditional posterior (2.8)–(2.10) to make inference with respect to $(\boldsymbol{\beta}, \boldsymbol{\eta})$. For details about the PEB approach see, for instance, Morris (1983), Casella (1985) or, in the reliability literature, Gaver

and O'Muircheartaigh (1987).

This paper differs from the usual PEB approach in two ways. On one side, we estimate $\boldsymbol{\phi}$ by maximizing the marginal posterior $p(\boldsymbol{\phi}|\boldsymbol{D}) \propto p(\boldsymbol{D}|\boldsymbol{\phi}) \times p(\boldsymbol{\phi})$ rather than the likelihood $p(\boldsymbol{D}|\boldsymbol{\phi})$. On the other, we use a bootstrap approach introduced by Laird and Louis (1987) to correct for the underestimation of uncertainty due to ignoring the uncertainty in the estimation of $\boldsymbol{\phi}$, which is usually a drawback of the PEB approach. Hence, this section is divided into three subsections which discuss respectively (i) the maximum posterior density estimate for $\boldsymbol{\phi}$, (ii) a rejection sampling algorithm to sample from the conditional posterior $p(\boldsymbol{\beta}, \boldsymbol{\eta} \mid \boldsymbol{D}, \boldsymbol{\phi})$ and (iii) the parametric bootstrap strategy used to approximate the posterior marginal distribution $p(\boldsymbol{\phi}|\boldsymbol{D})$ which is then used to correct both standard errors of point estimates and credibility intervals for the system specific parameters.

3.1 Maximum posterior density estimate

From (2.5) and (3.4), the marginal likelihood for $\boldsymbol{\phi}$ is given by

$$\begin{aligned} p(\mathbf{D}|\boldsymbol{\phi}) &= \int_{\mathbb{R}_+^K} \int_{\mathbb{R}_+^K} p(\mathbf{D}|\boldsymbol{\beta}, \boldsymbol{\eta}) \times p(\boldsymbol{\beta}, \boldsymbol{\eta}|\boldsymbol{\phi}) d\boldsymbol{\eta} d\boldsymbol{\beta} \\ &= \prod_{i=1}^{K} \left(\frac{1}{t_{ij}}\right) \frac{\Gamma(a_\eta + n_i)}{\Gamma(a_\eta)\Gamma(a_\beta)} \left(\frac{a_\beta}{\beta_0}\right)^{a_\beta} \\ &\quad \times \int_0^\infty \left[\frac{a_\eta(\theta_0/\tau_i)^{\beta_i}}{a_\eta(\theta_0/\tau_i)^{\beta_i} + 1}\right]^{a_\eta} \left[\frac{1}{a_\eta(\theta_0/\tau_i)^{\beta_i} + 1}\right]^{n_i} \beta_i^{a_\beta + n_i - 1} e^{-\beta_i(a_\beta/\beta_0 + w_i)} d\beta_i. \end{aligned} \tag{2.11}$$

Note that the last integral in (2.11) has no closed form and it should have to be computed numerically in the maximization algorithm. Hence, the marginal posterior distribution of $\boldsymbol{\phi}$ is

$$p(\boldsymbol{\phi}|\mathbf{D}) \propto p(\mathbf{D}|\boldsymbol{\phi}) \times p(\boldsymbol{\phi}), \tag{2.12}$$

where $p(\boldsymbol{\phi})$ is given in (2.7). Note that maximizing (2.12) is equivalent to maximizing

$$\ell(\boldsymbol{\phi}) = \log p(\mathbf{D}|\boldsymbol{\phi}) - (\xi_1 a_\beta + \xi_2 a_\eta), \tag{2.13}$$

showing that one could think of the maximum posterior estimate of $\boldsymbol{\phi}$ as a penalized likelihood approach. Maximization of (2.13) is carried out numerically. Initial values to start the algorithm are discussed in Appendix A.

In order to evaluate the behavior of the estimators obtained from the maximization of (2.13), we conducted a Monte Carlo simulation study. The Monte Carlo scenarios were designed to generate data similar to the transformers example. Hence, we set the hyperparameters $\beta_0 = 2$, $\theta_0 = 10,000$, $a_\beta = 2, 10$, $a_\eta = 2, 10$, truncation times varying from 2,000 to 20,000 hours and $K = 10, 40, 70$ and 100 systems. We compared the mean and standard errors of the estimates $(\hat{a}_\beta, \hat{\beta}_0, \hat{a}_\eta, \hat{\theta}_0)$ of 500 Monte Carlo replicates using (i) maximization of the marginal likelihood, (ii) maximization of the marginal posterior of $\boldsymbol{\phi}$ with $\xi_1 = \xi_2 = 1$ and (iii) same as (ii) but with $\xi_1 = \xi_2 = 0.1$. All the results were obtained using the software R, version 3.0.1 (R Core Team, 2013).

The results are summarized in Figures 2.2–2.5. Briefly, the estimates for β_0 and θ_0 behave similar for the three methods. In other words, the introduction of a penalty of the form $\xi_1 a_\beta + \xi_2 a_\eta$ does not

impact much the estimates of β_0 and θ_0. On the other hand, the estimates of a_β and a_η obtained maximizing the marginal posterior performed better than the ones obtained by maximizing the marginal likelihood, in the sense that they have smaller bias and standard errors for small K. Of the two options $\xi_1 = \xi_2 = 1$ and $\xi_1 = \xi_2 = 0.1$, the latter seems to be slightly better. In terms of the prior distribution (2.7) for $\boldsymbol{\phi}$, this amounts to setting (improper) uniform priors for both β_0 and θ_0 and exponential distributions with mean and standard deviation $1/0.1 = 10$ for both a_β and a_η. We finally note that, as expected, as the amount of information grows (i.e., K grows), the three estimators seem to converge to the true values of $\boldsymbol{\phi}$.

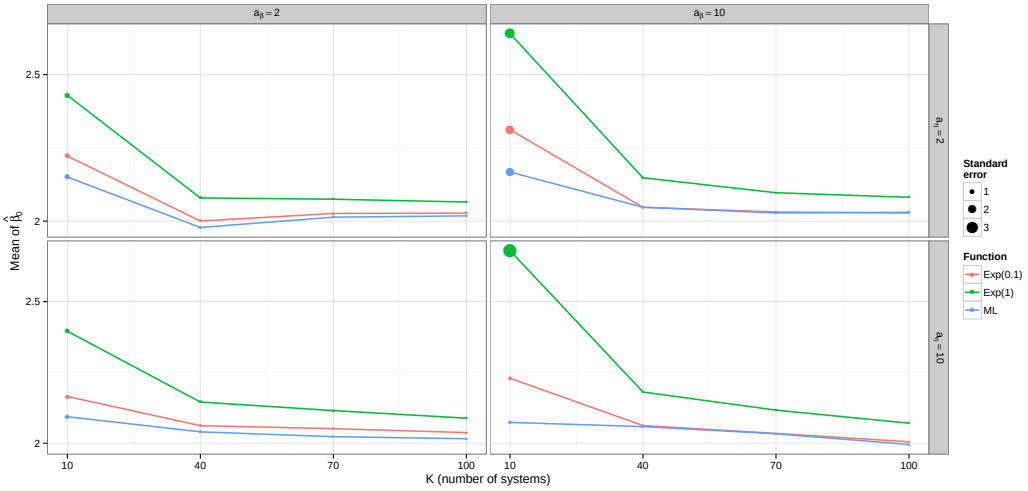

Figure 2.2: Mean value of the estimates of $\hat{\beta}_0$. Point sizes are proportional to the standard error of the estimates.

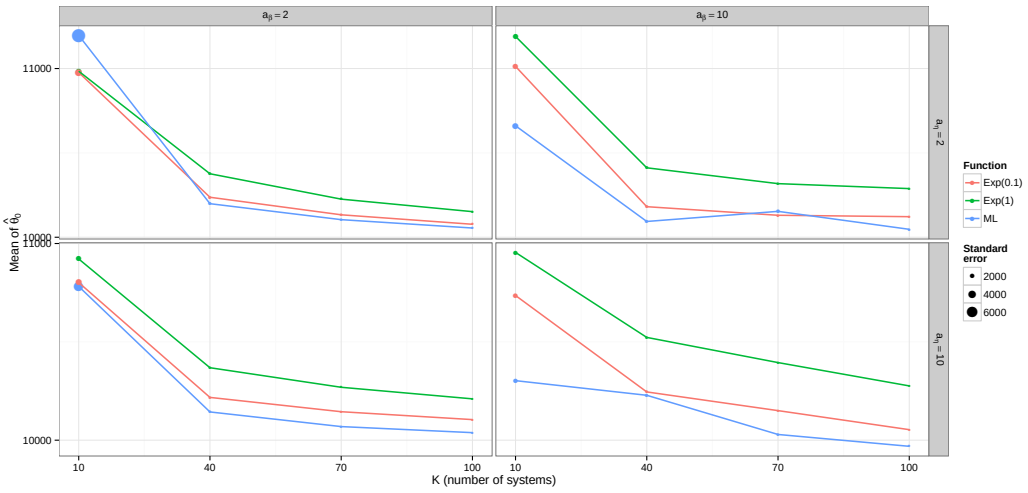

Figure 2.3: Mean value of the estimates of $\hat{\theta}_0$. Point sizes are proportional to the standard error of the estimates.

3.2 Empirical Bayes posterior analysis

For given $\boldsymbol{\phi}$, *iid* simulation from the conditional posterior distribution (2.8)–(2.10) is straightforward using the rejection sampling algorithm (see, for instance, Devroye, 1986; Gelman et al., 2003). Note first that (i) the pairs (β_i, η_i) are conditionally independent and (ii) given β_i, η_i follows a Gamma distribution. Hence, the only difficulty in order to sample from $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi})$ is how to sample from (2.10).

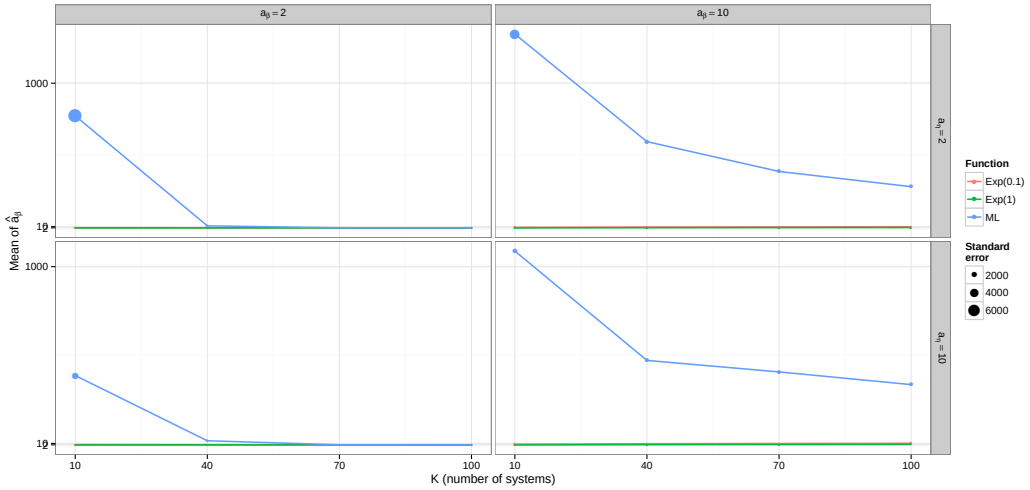

Figure 2.4: Mean value of the estimates of $\hat{a}_\beta$. Point sizes are proportional to the standard error of the estimates.

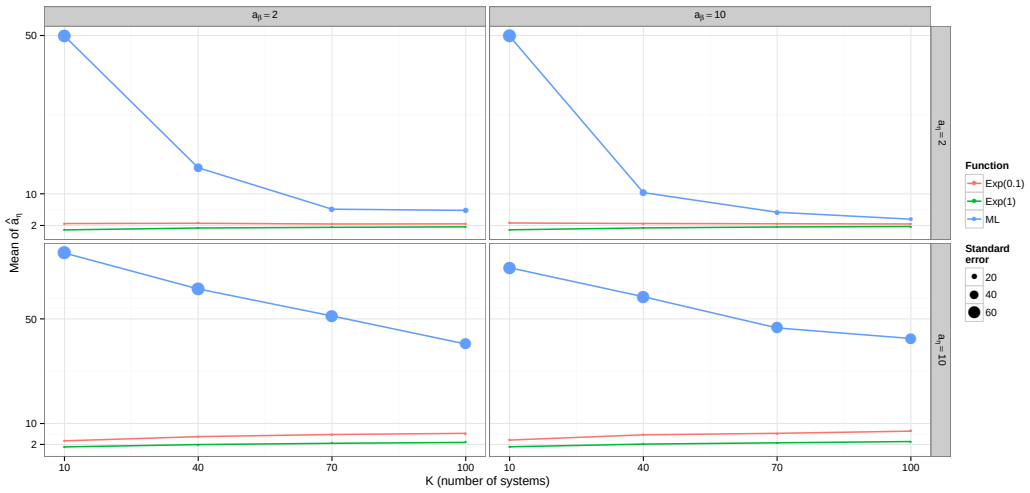

Figure 2.5: Mean value of the estimates of $\hat{a}_\eta$. Point sizes are proportional to the standard error of the estimates.

Let $F(\beta_i)$ be the last factor in the right hand side of (2.10), i.e.

$$F(\beta_i) = \frac{[a_\eta(\theta_0/\tau_i)^{\beta_i}]^{a_\eta}}{[a_\eta(\theta_0/\tau_i)^{\beta_i}+1]^{a_\eta+n_i}}.$$

Simple algebra shows that $F(\beta_i)$ is maximized when $\beta_i = \beta_i^* = \max\{0, -\log n/\log(\theta_0/\tau_i)\}$. Therefore, we can generate a random variable having the pdf (2.10) by

1. Generate $\beta_i^{(\text{cand})} \sim$ Gamma $(\beta_i | a_\beta + n_i, a_\beta/\beta_0 + w_i)$ and $u \sim$ Uniform(0,1).
2. Define $C_i = F(\beta_i^*)$. If $u\, C_i \leq F(\beta_i^{(\text{cand})})$, accept $\beta_i = \beta_i^{(\text{cand})}$. Otherwise, repeat step 1 until the acceptance condition is met.

Using the structure of the model we can then generate an observation from $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi})$ by running the previous algorithm K times to obtain $\beta_1, \ldots, \beta_K$ and then sampling $\eta_1, \ldots, \eta_K$ from the Gamma distributions (2.9). We then repeat this procedure M times to obtain an *iid* sample $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\eta}^{(1)}), \ldots, (\boldsymbol{\beta}^{(M)}, \boldsymbol{\eta}^{(M)})$ from $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi})$. Of course, in our application we set $\boldsymbol{\phi}$ to be the PEB estimate.

3.3 Parametric Bootstrap correction

From a Bayesian point of view, the PEB distribution $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \hat{\boldsymbol{\phi}})$ is an approximation to the marginal posterior distribution

$$p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}) = \int_{\mathbb{R}^4_+} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi}) p(\boldsymbol{\phi}|\mathbf{D}) d\boldsymbol{\phi}\,, \tag{2.14}$$

where $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi})$ is given by (2.8) and $p(\boldsymbol{\phi}|\mathbf{D})$ by (2.11)–(2.12). In other words, the PEB approach replaces $p(\boldsymbol{\phi}|\mathbf{D})$ by the Dirac measure $\delta_{\hat{\boldsymbol{\phi}}}$ to get

$$\tilde{p}_{\text{naive}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}) = \int_{\mathbb{R}^4_+} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi}) \delta_{\hat{\boldsymbol{\phi}}}(d\boldsymbol{\phi}) = p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \hat{\boldsymbol{\phi}}), \tag{2.15}$$

where $\hat{\boldsymbol{\phi}}$ is the maximum posterior density estimate of $\boldsymbol{\phi}$. This approximation is naive since it fails to take into account the uncertainty with respect to the estimation of $\boldsymbol{\phi}$. Consequently, posterior variances tend to be underestimated and credible intervals too narrow. Laird and Louis (1987) suggested that a more satisfactory solution would be to replace the posterior $p(\boldsymbol{\phi}|\mathbf{D})$ in (2.14) by the sampling distribution $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$ of $\hat{\boldsymbol{\phi}}$. When $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$ is not known or difficult to obtain, they propose to use a parametric bootstrap method to get a proxy for $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$. The bootstrap algorithm obtains bootstrap replications $\hat{\boldsymbol{\phi}}^{(b)}$ $(b = 1 \ldots, B)$ on which to base the approximation to $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$. Given $\hat{\boldsymbol{\phi}}$, the maximum posterior density estimate of $\boldsymbol{\phi}$ using the original data, we generate first $(\boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}^{(b)})$ from the prior distribution $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\hat{\boldsymbol{\phi}})$ and then $\boldsymbol{D}^{(b)}$ from $p(\boldsymbol{D}|\boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}^{(b)})$. Let $\hat{\boldsymbol{\phi}}^{(b)}$ be the maximum posterior density estimate of $\boldsymbol{\phi}$ using the simulated data $\mathbf{D}^{(b)}$, and $\hat{f}_B(\boldsymbol{\phi})$ be the discrete probability function that puts mass $1/B$ on $\hat{\boldsymbol{\phi}}^{(b)}$. The bootstrap corrected approximation to $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ is

$$\tilde{p}_{\text{boot}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}) = \int_{\mathbb{R}^4_+} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi}) \hat{f}_B(\boldsymbol{\phi}) d\boldsymbol{\phi} = \frac{1}{B} \sum_{b=1}^{B} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \hat{\boldsymbol{\phi}}^{(b)}). \tag{2.16}$$

An *iid* sample from the bootstrap corrected distribution $\tilde{p}_{\text{boot}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ is obtained by (i) drawing at random one of the bootstrap replications $\hat{\boldsymbol{\phi}}^{(b)}$ $(b = 1 \ldots, B)$ and (ii) generate a pair $(\boldsymbol{\beta}, \boldsymbol{\eta})$ from the conditional posterior $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\boldsymbol{D}, \hat{\boldsymbol{\phi}}^{(b)})$ using the drawed value of $\hat{\boldsymbol{\phi}}^{(b)}$ and the algorithm described in Section 3.2.

Carlin and Gelfand (1990) discuss other approaches to correct the PEB interval estimates. Kass and Steffey (1989) obtained first and second order approximations to $\text{Var}\,[h(\beta_i, \eta_i)|\mathbf{D}]$ in two-stage hierarchical models. We remark that obtaining these approximations is quite hard in the hierarchical PLP model, since they involve higher order derivatives of complicated expressions.

4 Application: Power transformers data set

We return now to the power transformers data in Table 2.1. Interest centers in estimation of some quantities associated to the reliability of each system. Among these we mention the β_i's, specifically to assess whether the systems are degrading $(\beta_i > 1)$ or improving $(\beta_i < 1)$, the scale parameters $\theta_i = \tau_i/\eta_i^{1/\beta_i}$, the probability that no failure occur in a period of time of length l_0 starting at s,

called the *reliability function* of the system (Hamada et al., 2008),

$$R_i(s, l_0) = \Pr(N_i(s, s + l_0) = 0|\beta_i, \theta_i) = \exp\left\{\left(\frac{s}{\theta_i}\right)^{\beta_i} - \left(\frac{s + l_0}{\theta_i}\right)^{\beta_i}\right\},$$

for given values of s and l_0 (e.g. $l_0 = 4,380$ and $8,760$ hours, corresponding respectively to 6 months and one year), and, finally, the *optimal maintenance check point* $t_{PM}^{*(i)}$ under a block policy (cf. Mazzuchi and Soyer, 1996), which we explain below.

4.1 Preventive maintenance policy

The optimal maintenance check point of the i-th system, $t_{PM}^{*(i)}$, is the value of t_{PM} that minimizes the expected cost

$$\mathrm{E}\left[C_i(t_{PM}, N_i(t_{PM}))\right] = \int \frac{C_{PM} + C_{MR}\eta_i(t_{PM}/\tau_i)^{\beta_i}}{t_{PM}} p(\beta_i, \eta_i|D_i)d\beta_i d\eta_i\,. \tag{2.17}$$

In order to compute an estimate of $t_{PM}^{*(i)}$ we use a sample $\{(\beta_i^{(m)}, \eta_i^{(m)}), m = 1, \ldots, M\}$ from the approximate posterior, either $\tilde{p}_{\text{naive}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ or $\tilde{p}_{\text{boot}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$, given in equations (2.15)–(2.16), and approximate the right hand side of (2.17) by $M^{-1}\sum_{m=1}^{M}[C_{PM} + C_{MR}\eta_i^{(m)}(t_{PM}/\tau_i)^{\beta_i^{(m)}}]/t_{PM}$. The estimate of the optimal maintenance check point is then obtained by a numerical minimization procedure.

The optimal maintenance checkpoint relates to the decision of whether to perform a *perfect* preventive maintenance on the system. (A perfect preventive maintenance leaves the system in *as good as new* condition and, hence, can also be thought of as the action of replacing the system by a new one.) One of the most common strategies of planned preventive maintenance is the block policy. This strategy consists in performing a preventive maintenance at the end of each time interval of length t_{PM}, regardless of the number of previous failures. Under the block policy, the cost per unit of time of the i-th system is

$$C_i(t_{PM}, N_i(t_{PM})) = \frac{C_{PM} + C_{MR}N_i(t_{PM})}{t_{PM}},$$

where $N_i(t_{PM})$ is the number of failures of the i-th system in the time interval of length t_{PM}, C_{PM} is the cost of the preventive maintenance, and C_{MR} is the cost of a minimal repair (unscheduled maintenance due to a failure). Since $N_i(t_{PM})$ is a random quantity, we obtain the conditional expected cost per time unit given (β_i, η_i) as

$$\mathrm{E}\left[C_i(t_{PM}, N_i(t_{PM}))|\beta_i, \eta_i\right] = \frac{C_{PM} + C_{MR}\Lambda_i(t_{PM})}{t_{PM}}\,. \tag{2.18}$$

A classical approach takes the optimal maintenance time to be the time that minimize (2.18) and compute an estimate replacing (β_i, η_i) by their estimates (see, for instance, Barlow and Hunter, 1960; Gilardoni and Colosimo, 2007, 2011; Oliveira et al., 2012; Gilardoni et al., 2013). Here, instead, we follow Mazzuchi and Soyer (1996) taking the optimal maintenance time $t_{PM}^{*(i)}$ as the value t_{PM} that minimizes the expected cost (2.17).

4.2 Results

The maximum posterior density estimates of the hyperparameters were obtained maximizing Equation (2.13) with $\xi_1 = \xi_2 = 0.1$. This gave $\hat{\boldsymbol{\phi}} = (\hat{a}_\beta, \hat{\beta}_0, \hat{a}_\eta, \hat{\theta}_0) = (7.02; 2.29; 4.71; 23,980)$. Using this estimates we then generated a sample of size $M = 10,000$ from both $\tilde{p}_{\text{naive}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ and $\tilde{p}_{\text{boot}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$, where for the latter it was used $B = 1,000$. Approximations to the estimates of the quantities of interests under squared error loss were then computed by taking the posterior sample averages of the corresponding functions. Likewise, approximate HPD intervals were computed taking the sampling quantiles, say a and $(1-b)$, so that $(1-a-b)$ gives the desired coverage (posterior probability) and the length of the interval is minimum.

Table 2.2 shows the maximum likelihood and PEB estimates of the β_i and η_i. Note that, unlike the ML approach, in the hierarchical approach estimates of β_i are obtained even for the systems that have no failures. Furthermore, note that the PEB estimates of β_i are a compromise between the ML estimates, which use only data from the i-th system, and the estimated prior mean of β_i, $\hat{\beta}_0$, which uses data from all systems. For the systems with $n_i = 0$, $\hat{\beta}_i$ is close to $\hat{\beta}_0$, since the individual likelihood has little or no information about β_i.

Table 2.2: Maximum likelihood (MLE), naive and bootstrap PEB estimates of (β_i, η_i) for the power transformers data.

	β_i					η_i				
		Naive		Bootstrap			Naive		Bootstrap	
System i	MLE	Mean	SD	Mean	SD	MLE	Mean	SD	Mean	SD
1	1.73	2.08	0.69	2.14	0.92	2	1.00	0.40	1.14	0.60
2	1.75	2.09	0.70	2.16	0.94	2	1.01	0.39	1.14	0.59
3	3.86	2.16	0.75	2.29	1.07	1	0.36	0.21	0.40	0.30
4	-	2.35	0.87	2.80	1.85	0	0.10	0.10	0.10	0.13
5	-	2.36	0.87	2.84	1.96	0	0.11	0.11	0.11	0.14
6	-	2.41	0.88	2.88	1.92	0	0.26	0.17	0.23	0.22
7	4.11	2.41	0.84	2.77	1.45	1	0.83	0.35	0.86	0.50
8	3.40	2.39	0.85	2.69	1.41	1	0.83	0.36	0.86	0.50
9	3.78	2.41	0.86	2.74	1.46	1	0.84	0.36	0.86	0.49
10	-	2.33	0.85	2.77	1.84	0	0.05	0.07	0.05	0.09
11	-	2.39	0.87	2.91	2.12	0	0.40	0.22	0.35	0.28
12	1.04	1.73	0.58	1.65	0.68	2	0.88	0.35	1.03	0.55
13	-	2.31	0.87	2.74	1.85	0	0.02	0.03	0.02	0.06
14	8.52	2.55	0.85	2.99	1.43	2	0.79	0.32	0.89	0.51
15	-	2.30	0.87	2.73	1.87	0	0.01	0.03	0.02	0.06
16	3.08	2.35	0.83	2.64	1.35	1	0.83	0.35	0.86	0.50
17	2.91	2.36	0.82	2.63	1.36	1	0.84	0.36	0.87	0.50
18	-	2.34	0.88	2.82	1.97	0	0.66	0.31	0.59	0.41
19	-	2.28	0.86	2.74	1.90	0	0.00	0.01	0.00	0.03
20	2.28	2.24	0.75	2.38	1.02	2	0.99	0.39	1.12	0.60
21	-	2.36	0.87	2.79	1.87	0	0.09	0.09	0.09	0.13
22	-	2.33	0.84	2.78	1.82	0	0.07	0.08	0.07	0.11
23	0.89	1.52	0.52	1.41	0.60	1	0.20	0.15	0.27	0.24
24	6.69	2.49	0.87	2.97	1.69	1	0.83	0.35	0.86	0.50
25	-	2.31	0.85	2.72	1.83	0	0.02	0.03	0.02	0.06
26	-	2.33	0.87	2.74	1.84	0	0.02	0.04	0.03	0.06
27	2.19	2.17	0.76	2.30	1.09	1	0.53	0.26	0.56	0.36
28	2.93	2.35	0.82	2.61	1.32	1	0.83	0.36	0.86	0.50
29	-	2.31	0.86	2.76	2.03	0	0.01	0.02	0.02	0.05
30	2.83	2.31	0.82	2.56	1.28	1	0.71	0.31	0.74	0.44
31 - 40	-	2.34	0.88	2.79	2.03	0	0.69	0.32	0.62	0.42

Table 2.3 presents PEB estimates for the quantities $\Pr(\beta_i > 1|\hat{\boldsymbol{\phi}})$ and $t_{PM}^{*(i)}$. If we look at the probability that a system is degrading, namely $\tilde{\Pr}(\beta_i > 1|D_i, \hat{\boldsymbol{\phi}})$, the smallest values are 0.742 and 0.845, respectively for systems 23 and 12, while all others are greater than 0.93, indicating strong evidence in the sense that the intensities are increasing and the transformers are degrading with time. This can be seen also in Figure 2.6, which shows the posterior means of the reliability function for the forty systems. Figure 2.6(a) shows, for instance, that a system that was followed-up for six months has probability of having no failure in the next six months varying from 0.832 to 0.942. Similarly, Figure 2.6(b) shows that if a system was followed-up to one year, the probability of observing no failures in the next year vary from 0.604 to 0.783. Note the distinct behavior of the reliability functions of systems 12 and 23. These two systems are the power transformers that presented the earliest failure times. The columns $t_{PM}^{*(i)}$ of Table 2.3 also show the optimal maintenance check points for each system. To compute this we followed Gilardoni and Colosimo (2007) and Oliveira et al. (2012), which consider that the cost of a minimal repair is fifteen times the cost of a preventive maintenance. The estimated optimal maintenance check points vary from 6,592 (system 20) to 9,348 hours (system 23). Using the same data, but considering that the forty power transformers are a sample of the same power law process (i.e. same β and θ for all systems), Gilardoni and Colosimo (2007) and Oliveira et al. (2012), using respectively ML and a Bayesian approach, arrived at an optimal time of about $6,420$ hours. The hierarchical approach has the advantage that each power transformer can be subject to its own optimal maintenance check point, allowing therefore a greater flexibility in the maintenance policy.

Table 2.3: PEB estimates for probability that a system is degrading ($\tilde{\Pr}(\beta_i > 1|D_i, \hat{\boldsymbol{\phi}})$) and optimal maintenance check points ($t_{PM}^{*(i)}$) for the power transformers data.

System	$\tilde{\Pr}(\beta_i > 1\|D_i, \hat{\boldsymbol{\phi}})$	$t_{PM}^{*(i)}$	System	$\tilde{\Pr}(\beta_i > 1\|D_i, \hat{\boldsymbol{\phi}})$	$t_{PM}^{*(i)}$
1	0.930	6,687	17	0.953	7,642
2	0.933	6,686	18	0.932	9,202
3	0.930	7,019	19	0.931	8,224
4	0.941	8,218	20	0.957	6,592
5	0.947	8,233	21	0.944	8,165
6	0.942	8,508	22	0.942	8,124
7	0.960	7,689	23	0.742	9,348
8	0.952	7,755	24	0.965	7,825
9	0.958	7,743	25	0.933	8,141
10	0.942	8,133	26	0.933	8,138
11	0.938	8,804	27	0.931	7,303
12	0.845	7,291	28	0.955	7,695
13	0.936	8,148	29	0.935	8,168
14	0.981	6,795	30	0.951	7,489
15	0.933	8,181	31-40	0.931	9,295
16	0.956	7,678			

An insight of the bootstrap correction can be seen from the histograms of the bootstrap sample of $\hat{\boldsymbol{\phi}}$ (Figure 2.7). Note that the sampling distribution of the estimates of the shape parameters a_β and a_η appear to be much more dispersed than those of β_0 and θ_0. The effect of the bootstrap correction can also be seen in Figure 2.8, which shows the HPD intervals for the β_i and θ_i computed using both the naive and the bootstrap corrected posterior. As expected, the bootstrap correction accounts for wider HPD intervals, which we believe reflects better the uncertainty in the data.

Finally, in order to understand the behavior of our model, Figure 2.9 shows the posterior means

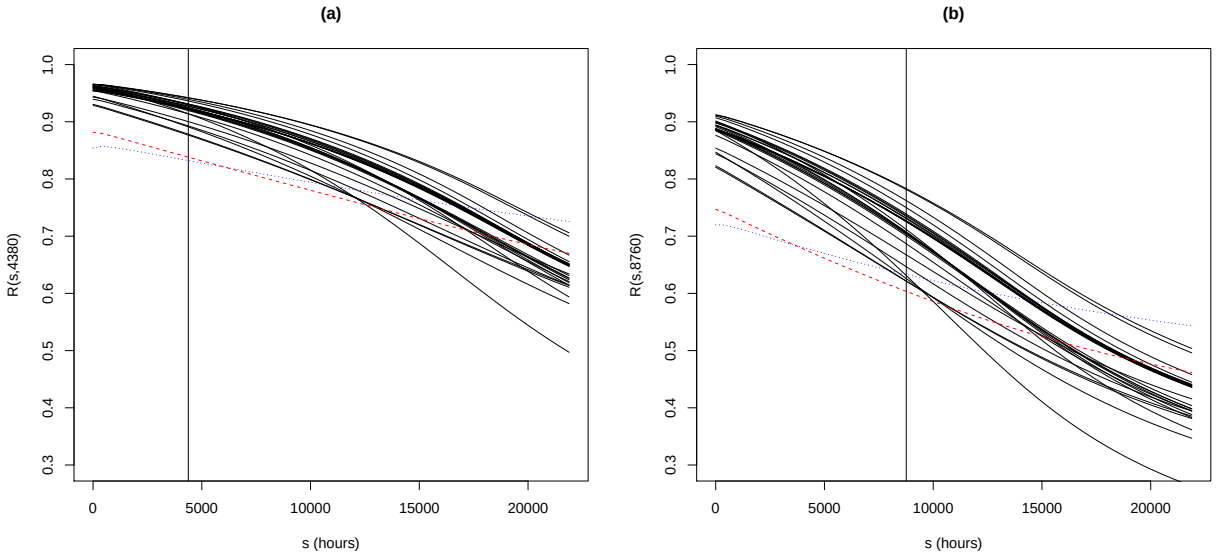

Figure 2.6: Posterior means of the reliability function of the forty power transformers when $l_0 = 4,380$ hours (6 months) **(a)** and $l_0 = 8,760$ hours (one year) **(b)**. The dashed red line and dotted blue line represent respectively systems 12 and 23. Vertical lines represtent $s = 4,380$ hours **(a)**, and $s = 8,760$ hours **(b)**.

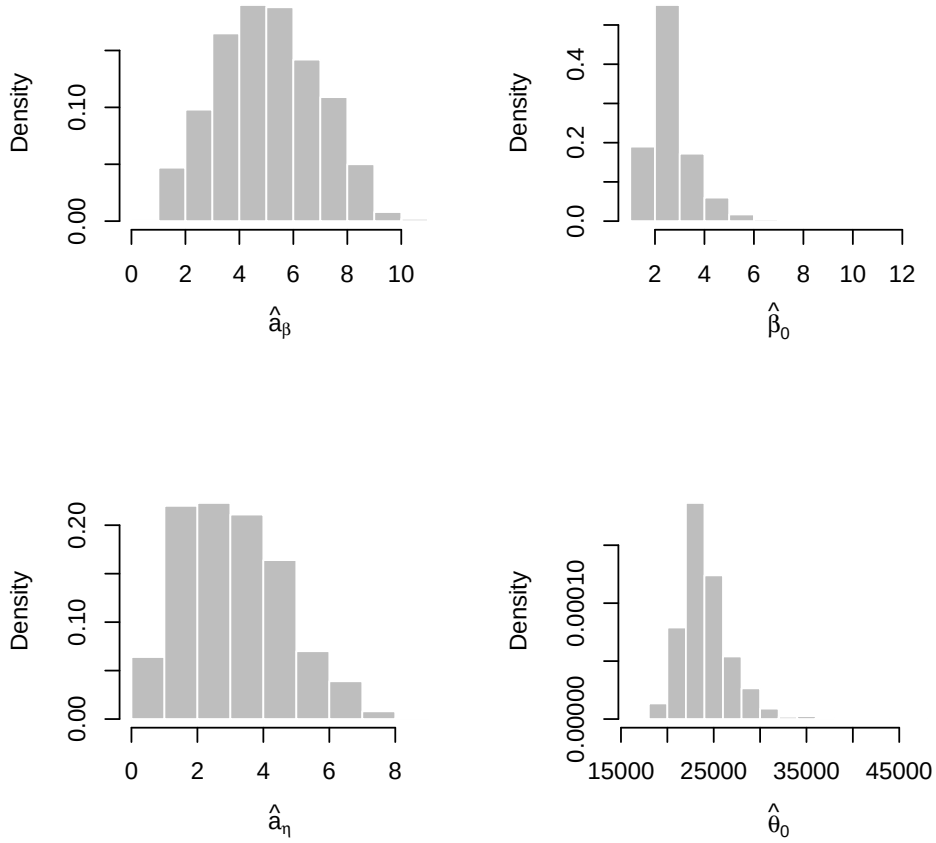

Figure 2.7: Bootstrap sample histograms of $\hat{\boldsymbol{\phi}}$ based on $B = 1,000$ for the power transformers data.

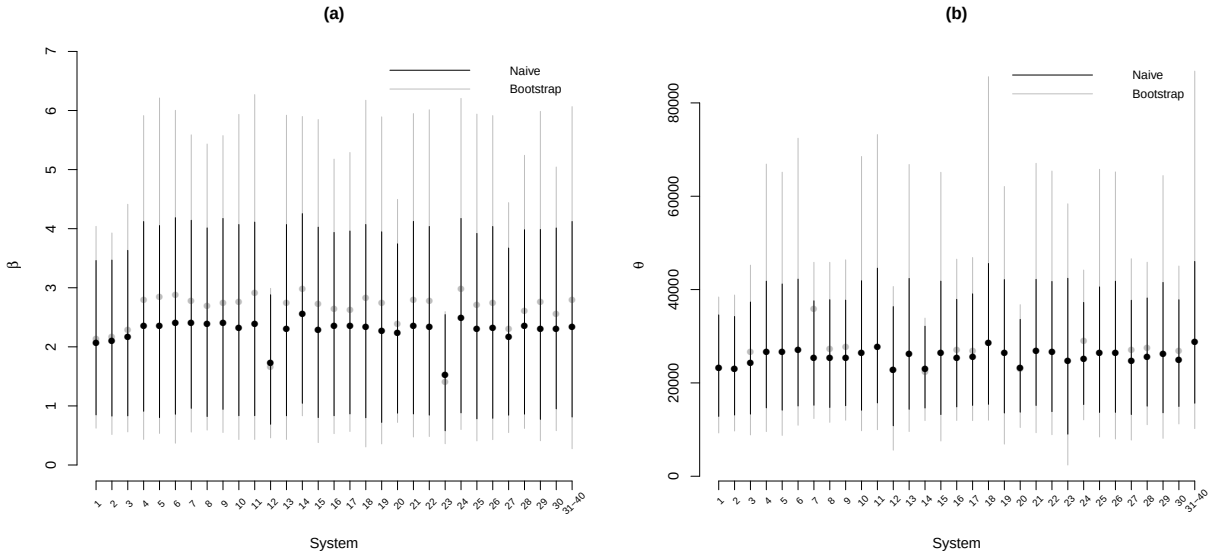

Figure 2.8: Naive and bootstrap PEB 95% HPD credible intervals of the parameters β_i **(a)**, and θ_i **(b)**. The points are posterior expectations.

$\tilde{\beta}_i$ for the parameter β_i, as a function of the prior standard deviation. As the standard deviation of β_i increases, the posterior mean of each β_i moves away in the direction of the ML estimate. On the other hand, as the standard deviation of β_i decreases to zero, the posterior mean of the β_i tend to the common value $\hat{\beta}_0$.

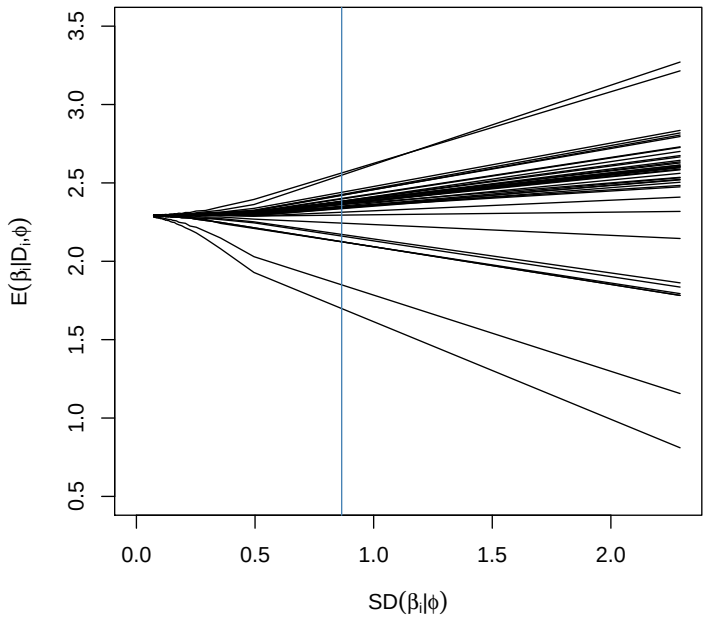

Figure 2.9: Posterior means of β_i for the power transformers data, as a function of the prior standard deviation $SD(\beta_i|\phi)$, conditionally on $\hat{a}_\eta = 5.28$, $\hat{\theta}_0 = 18,399.20$, $\hat{\beta}_0 = 2.29$ and a sequence of $a_\beta \in (1; 1,000)$. For each configuration value $(\hat{a}_\beta^{(b)}, \hat{\beta}_0^{(b)}, \hat{a}_\eta, \hat{\theta}_0)$, a sample of size $1,000$ of β_i was generated and the sample mean was computed. The blue vertical line is the observed prior standard deviation $\text{SD}(\beta_i|\hat{\boldsymbol{\phi}}) = 0.87$.

5 Conclusions

A hierarchical model was proposed for the analysis of multiple repairable systems with different truncation times. Scale and shape parameter of the power law intensity function of a nonhomogeneous Poisson proccess are allowed to vary among the systems. A suitable reparameterization was used to obtain a quasi-conjugate posterior analysis. This reparameterization introduced a difficulty in the sense that, when the truncation times are different, it is unreasonable to assume exchangeability in the second stage prior distribution. A parametric empirical Bayes approach was carried out in order to estimate the model parameters. The hyperpameter vector $\boldsymbol{\phi}$ was estimated by maximizing its posterior density, or equivalently, a marginal penalized likelihood function. Once that the hyperparameters were estimated, approximations to the estimates of the system specific parameters were obtained using an *iid* Monte Carlo sample from $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\boldsymbol{D}, \hat{\boldsymbol{\phi}})$. This Monte Carlo sample can be obtained using a simple and efficient rejection sampling algorithm. Furthermore, a parametric bootstrap method was used to correct the standard deviations of point estimates and the HPD intervals by taking into account the uncertainty in the estimate of the hyperparameters. These methods were used to analyze a real data set regarding failure times of 40 power transformers, including estimation of the optimal preventive maintenance time considering block policy.

Acknowledgments

This work was partially supported by CNPq, CAPES, FAPEMIG and FAPDF.

A Starting values for the maximum posterior estimation

The main ideia is to use the ML estimates of β_i and η_i as the true values in the second stage prior (3.4). Let $\hat{\boldsymbol{\beta}}_{ML}$ and $\hat{\boldsymbol{\eta}}_{ML}$ be the vectors of ML estimates for those systems with $n_i > 0$ (the ML estimate of β_i does not exist when $n_i > 0$). Taking logarithms in (3.4) and replacing the actual β_i and η_i by their ML estimates we obtain

$$\begin{aligned}\log\, p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}) &= \sum_{i:n_i>0} \Big\{ a_\eta[\log(a_\eta) + \hat{\beta}_i \log(\theta_0/\tau_i)] - \log\Gamma(a_\eta) + (a_\eta - 1)\log(\hat{\eta}_i) - \hat{\eta}_i a_\eta (\theta_0/\tau_i)^{\hat{\beta}_i} \\ &\quad + \; a_\beta \log(a_\beta/\beta_0) - \log\Gamma(a_\beta) + (a_\beta - 1)\log(\hat{\beta}_i) - \hat{\beta}_i(a_\beta/\beta_0) \Big\}\,. \end{aligned} \tag{2.19}$$

Hence, we take as starting values for $\boldsymbol{\phi}$ the solution of $\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))/\partial\boldsymbol{\phi} = \mathbf{0}$, that is

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial a_\beta} = \sum_{i:n_i>0} \left[\log(a_\beta/\beta_0) + 1 - \psi(a_\beta) + \log(\hat{\beta}_i) - \frac{\hat{\beta}_i}{\beta_0} \right] = 0, \tag{2.20}$$

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial \beta_0} = \frac{a_\beta}{\beta_0} \sum_{i:n_i>0} \left[\frac{\hat{\beta}_i}{\beta_0} - 1 \right] = 0, \tag{2.21}$$

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial a_\eta} = \sum_{i:n_i>0} \left[\log(a_\eta) + 1 + \hat{\beta}_i \log(\theta_0/\tau_i) - \psi(a_\eta) + \log(\hat{\eta}_i) - \hat{\eta}_i (\theta_0/\tau_i)^{\hat{\beta}_i} \right] = 0, \tag{2.22}$$

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial \theta_0} = \frac{a_\eta}{\theta_0} \sum_{i:n_i>0} \left[\hat{\beta}_i - \hat{\beta}_i \hat{\eta}_i \left(\frac{\theta_0}{\tau_i} \right)^{\hat{\beta}_i} \right] = 0. \tag{2.23}$$

Let K_* be the number of systems with $n_i > 0$. From Equation (2.21) we obtain that $\tilde{\beta}_0 = K_*^{-1} \sum_{i:n_i>0} \hat{\beta}_i$ and replacing β_0 by $\tilde{\beta}_0$ in Equation (2.20), we obtain $\tilde{a}_\beta$ as the solution of

$$\log(\tilde{a}_\beta) - \psi(\tilde{a}_\beta) - \log(\tilde{\beta}_0) - K_*^{-1} \sum_{i:n_i>0} \log(\hat{\beta}_i) = 0. \tag{2.24}$$

From Equation (2.23) we obtain that $\tilde{\theta}_0$ is the solution of

$$K_*^{-1} \sum_{i:n_i>0} \hat{\beta}_i - K_*^{-1} \sum_{i:n_i>0} \hat{\beta}_i \hat{\eta}_i (\tilde{\theta}_0/\tau_i)^{\hat{\beta}_i} = 0. \tag{2.25}$$

Finally, we replace θ_0 by $\tilde{\theta}_0$ in Equantion (2.22) to obtain $\tilde{a}_\eta$ as the solution of

$$\log(\tilde{a}_\eta) - \psi(\tilde{a}_\eta) - K_*^{-1} \sum_{i:n_i>0} \left[\hat{\beta}_i \log(\tilde{\theta}_0/\tau_i) + \log(\hat{\eta}_i) - \hat{\eta}_i (\tilde{\theta}_0/\tau_i)^{\hat{\beta}_i} \right] = 0. \tag{2.26}$$

We note that Equations (2.24) to (2.26) are all univariate and hence can be solved by simple numerical procedures. In the real data example analyzed in Section **??**, these starting values were close to the final estimates.

References

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer, 1993.

R. D. Baker. Some new tests of the power law process. *Technometrics*, 38:256–265, 1996.

R. Barlow and L. Hunter. Optimum preventive maintenance policies. *Operations Research*, 8: 90–100, 1960.

M. Bhattacharjee, E. Arjas, and U. Pulkkinen. Modelling heterogeneity in nuclear power plant valve failure data. In B. H. Lindqvist and K. A. Doksum, editors, *Mathematical and Statistical Methods for Reliability*, pages 341–353. World Scientific Publishing, 2003.

B. P. Carlin and A. E. Gelfand. Approaches for the empirical Bayes confidence intervals. *Journal of the American Statistical Association*, 85:105–114, 1990. doi: 10.1080/01621459.1990.10475312.

G. Casella. An introduction to empirical Bayes data analysis. *The American Statistician*, 39:83–87, 1985. doi: 10.2307/2682801.

R. J. Cook and J. F. Lawless. *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health. Springer, 2007.

L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

D. P. Gaver and I. G. O'Muircheartaigh. Robust empirical Bayes analyses of event rates. *Technometrics*, 29:1–15, 1987. doi: 10.2307/1269878.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, New York, 2003.

E. I. George, U. E. Makov, and A. F. M. Smith. Conjugate likelihood distributions. *Scandinavian Journal of Statistics*, 20:147–156, 1993.

G. L. Gilardoni and E. A. Colosimo. Optimal maintenance time for repairable systems. *Journal of Quality Technology*, 39:48–53, 2007.

G. L. Gilardoni and E. A. Colosimo. On the superposition of overlapping Poisson processes and non-parametric estimation of their intensity function. *Journal of Statistical Planning and Inference*, 2011.

G. L. Gilardoni, M. D. de Oliveira, and E. A. Colosimo. Nonparametric estimation and bootstrap confidence intervals for the optimal maintenance time of a repairable system. *Computational Statistics & Data Analysis*, 63(0):113 – 124, 2013. ISSN 0167-9473. doi: http://dx.doi.org/10.1016/j.csda.2013.02.006. URL http://www.sciencedirect.com/science/article/pii/S0167947313000509.

M. Giorgio, M. Guida, and G. Pulcini. Repairable system analysis in presence of covariates and random effects. *Reliability Engineering and System Safety*, 131:271–281, 2014. URL http://dx.doi.org/10.1016/j.ress.2014.04.009i.

M. Guida and G. Pulcini. Bayesian reliability assessment of repairable systems during multi-stage development programs. *IIE Transactions*, 37(11):1071 – 1081, 2005. ISSN 0740817X.

M. Guida, R. Calabria, and G. Pulcini. Bayes inference for a non-homogeneous Poisson process with power intensity law. *IEEE Transactions on Reliability*, 38:603–609, 1989. doi: 10.1109/24.46489.

M. S. Hamada, A. G. Wilson, C. S. Reese, and H. F. Martz. *Bayesian Reliability*. Springer Series in Statistics. Springer, 2008.

Y.-S. Huang. A decision model for deteriorating repairable systems. *IIE Transactions*, 33(6):479, 2001. ISSN 0740817X.

R. E. Kass and D. Steffey. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, 84: 717–726, 1989. doi: 10.1080/01621459.1989.10478825.

N. M. Laird and T. A. Louis. Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association*, 82:739–750, 1987. doi: 10.2307/2288778.

J. F. Lawless. Regression methods for poisson process data. *Journal of the American Statistical Association*, 82:808–815, 1987.

B. H. Lindqvist. On the statistical modeling and analysis of repairable systems. *Statistical Science*, 21:532–551, 2006. doi: 10.1214/088342306000000448.

B. H. Lindqvist, G. Elvebakk, and K. Heggland. The trend-renewal process for statistical analysis of repairable systems. *Technometrics*, 45:31–44, 2003.

T. A. Mazzuchi and R. Soyer. A Bayesian perspective on some replacement strategies. *Reliability Engineering and System Safety*, 51:295–303, 1996. doi: 10.1016/0951-8320(95)00077-1.

C. N. Morris. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78:47–55, 1983. doi: doi:10.2307/2287098.

M. D. Oliveira, E. A. Colosimo, and G. L. Gilardoni. Bayesian inference for power law processes with applications in repairable systems. *Journal of Statistical Planning and Inference*, 142:1151–1160, 2012. doi: 10.1016/j.jspi.2011.11.016.

R. Pan and S. E. Rigdon. Bayes inference for general repairable systems. *Journal of Quality Technology*, 41:82–94, 2009.

C. J. Pérez, J. Martín, and M. J. Rufo. Sensitivity estimations for Bayesian inference models solved by MCMC methods. *Reliability Engineering and System Safety*, 91:1310–1314, 2006. doi: 10.1016/j.ress.2005.11.029.

S. Perkins, M. Cohen, E. Rahme, and S. Bernatsky. Melanoma and rheumatoid arthritis (brief report). *Clinical Reumathology*, 31:1001–1003, 2012. doi: 10.1007/s10067-011-1908-x.

M. H. Pesaran, D. Pettenuzzo, and A. Timmermann. Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73:1057–1084, 2006.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

S. E. Rigdon and A. P. Basu. *Statistical Methods for the Reliability of Repairable Systems*. John Wiley & Sons, 2000.

K. J. Ryan, M. S. Hamada, and C. S. Reese. A Bayesian hierarchical power law process model for multiple repairable systems with an application to supercomputer reliability. *Journal of Quality Technology*, 43:209–223, 2011.

Capítulo 3

Empirical Bayes and Jeffreys' prior for the hierarchical power law process

Rodrigo Citton P. dos Reis, Enrico A. Colosimo, and Gustavo L. Gilardoni

Abstract

In this paper we discuss alternative methods to model the third stage of a hierarchical power law process for modelling of several repairable systems. We argue that the Jeffreys' prior has some advantages with respect to an empirical Bayes alternative or the noninformative prior proposed by Ryan et al. (2011). More specifically, our simulations showed that the coverages of the intervals produced by the Jeffreys method are better than the interval coverages produced by empirical Bayes and noninformative methods. We also illustrate our methods with a real data set analysis.
Keywords: *Adaptive Metropolis, Marginal model, Minimal repair, Multiple repairable systems*

1 Introduction

Following the classical definition of Ascher and Feingold (1984), a repairable system is a system which, after failing to perform one or more of its functions satisfactorily, can be restored to fully satisfactory performance by a method other than replacement of the entire system. Many real world systems, such as automobiles, airplanes, printers, communication systems and others are repaired and not replaced. When these systems are subjected to a customer use environment, it is often of great interest to determine the reliability and other performance measures under these conditions.

The failure process of a repairable system can be represented by a point process. Consider a single system that starts to operate at $t = 0$, and failures occur at times $T_1 < T_2 < \ldots$. When a failure occurs, some repair action is done to put the system back in operation. We assume that repairs times are negligible. Failure times generate a counting process $\{N(t) : t \geq 0\}$, where $N(t)$ counts the number of failures in the interval $(0, t]$. A counting process with independent increments

and $N(0) = 0$, is said to be a Poisson process with intensity $\lambda(\cdot)$ if, for any t, the random variable $N(t)$ follows a Poisson distribution with mean $\Lambda(t) = \mathrm{E}[N(t)] = \int_0^t \lambda(u)du$. The nonhomogeneous Poisson process (NHPP) is a Poisson process with a non-constant intensity. As Baker (1996) argued, this model can be justified when, for example, a system has many components that can fail so that replacing a failed component has little effect on future system reliability. It is also refered as the case of *minimal repair*, which turns the system in the "as-bad-as-old" condition, i.e., the repair done in the system leaves the system in exactly the same condition as it was just before the failure. For a good review of NHPP with aplications to repairable systems see Rigdon and Basu (2000). More general models for repairables systems are discussed in Lindqvist (2006).

The most popular intensity function within the class of NHPP models is the power law intensity, given by

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}, \ \theta > 0, \ \beta > 0, \tag{3.1}$$

where θ is a scale parameter and β is a shape one, and the resulting process is called the power law process (PLP). The PLP intensity is very flexible to model repairable systems that are deteriorating or improving. If $\beta > 1$, (3.1) increases with time, which happens when the system is deteriorating. When $\beta < 1$, the PLP intensity function is decreasing in t, which happens when the system is improving. Finally, for $\beta = 1$, then $\lambda(t) = 1/\theta$, and the PLP reduces to a homogeneous Poisson process. The corresponding mean function is given by $\Lambda(t) = (t/\theta)^{\beta}$.

If we observe a single system time truncated at τ, the joint likelihood of $N(\tau) = n$ and the failure times $0 < t_1 < \ldots < t_n < \tau$ is given by (Crow, 1977)

$$p(n; t_1, \ldots, t_n|\beta, \theta) = \exp\left\{-(\tau/\theta)^{\beta}\right\} \frac{\beta^n}{\theta^{n\beta}} \prod_{j=1}^{n} t_j^{\beta-1}. \tag{3.2}$$

Bain and Engelhardt (1980) discuss inference procedures for the PLP model from a frequentist perspective. Many works proposed a Bayesian approach to model and estimate the parameters of the power law process, (see, for example, Guida et al., 1989; Bar-Lev et al., 1992). Recently, Oliveira et al. (2012) used a reparametrization of the PLP model in terms of β and $\eta = \Lambda(\tau) = (\tau/\theta)^{\beta}$. With this parametrization, the likelihood function (3.2) becomes

$$p(n; t_1, \ldots, t_n|\beta, \eta) \propto (\eta^n e^{-\eta}) \times (\beta^n e^{-\beta w}), \tag{3.3}$$

where $w = \sum_{j=1}^{n} \log(\tau/t_j)$. In the sequel of this paper, we will denote (3.3) by writing that $n; t_1, \ldots, t_n \sim PLP_{\tau}(n; t_1, \ldots, t_n|\beta, \eta)$. Note that (3.3) is proportional to the product of gamma densities $\gamma(\eta|n+1, 1) \times \gamma(\beta|n+1, w)$, where $\gamma(x|a, b) = b^a x^{a-1} e^{-xb}/\Gamma(a)$. For a single repairable system, Oliveira et al. (2012) used this fact to obtain a conjugate Bayesian analysis with independent gamma distributions for (β, η). The authors also extended this idea for the case of data from several repairable systems from the same PLP. They obtained a quasi-conjugate Bayesian analysis, and provided strategies to elicit prior knowledge from engineering experts point of view.

When failure data come from several systems, heterogeneity is something that is expected to observe from the different systems. This is due to different age of the systems, manufacturing lines, and maintenance teams, among others. In this case, a hierarchical approach comes naturally to model heterogeneity. With this framework, it is assumed that systems are similar, but each system

can exihibit its own characteristics. Some recent works related with this area considering data from repairable systems are Bhattacharjee et al. (2003); Pan and Rigdon (2009); Giorgio et al. (2014). A hiearchical PLP model was proposed by Ryan et al. (2011). The authors used the same structure of Oliveira et al. (2012) for each system, and a diffuse independent gamma distributions to model the hyperparameters level, justifying this choice as being nearly noninformative.

The goal of this paper is twofold: (i) propose other approaches to built prior distributions for the model hyperparameters, namely, Jeffreys and empirical Bayes and (ii) compare these approachs in a simulation study and a real data set available in the literature. To our knowlegde, this type of hyperpriors has never been introduced in the literature for the hierarchical PLP model. The choice for the prior of the hyperparameters of the hierarchical PLP model could be compared with the issue of modeling of the variance components in a general, or generalized, linear mixed model (see,for instance, Natarajan and Kass, 2000; Arima et al., 2012).

The rest of the paper is organized as follows. In Section 2, we set notation and present the hierarchical PLP model for analysis of multiple repairable systems. The Section 3 presents a parametric empirical Bayes (PEB) procedure to dealing with the hyperparameters estimation, and we show how to adjust the variance of the PEB estimators. In Section 4, we develop the Jeffreys' prior for the hyperparameters of the model, and discuss how to simulate from the respective posterior distribution. In Section 4.1 we present a simulation study comparing the different approaches to dealing with the hyperparameters of the model. Section 6 the methods discussed in this paper are applied to real data sets. Finally, some conclusions are given in Section 7.

2 Hierarchical PLP model

Ryan et al. (2011) introduced a hierarchical PLP model for the analysis of multiple systems failure time data. Supose we observe k repairable systems until truncation time τ. Let $N_i(\tau) = n_i$ be the number of failures observed, and $(t_{i1}, \ldots, t_{i,n_i})$ are the observed failure times from the i-th system. $D_i = (n_i, t_{i1}, \ldots, t_{in_i})$ is the vector with all data from i-th system, $i = 1, \ldots, k$, and $\mathbf{D} = (D_1, \ldots, D_k)$ is the vector of data from all systems. The model can be described as follow:

S1. *Data model:* Conditionally on $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$, the systems are independent with distribution $p(\mathbf{D}|\boldsymbol{\beta}, \boldsymbol{\eta}) = \prod_{i=1}^{k} p(D_i|\beta_i, \eta_i)$, where $p(D_i|\beta_i, \eta_i)$ is given by (3.3). In other words, $(D_i|\beta_i, \eta_i) \overset{indep.}{\sim} PLP_\tau(D_i|\beta_i, \eta_i)$.

S2. *Parameters model:* Conditionally on $\phi = (a_\eta, b_\eta, a_\beta, b_\beta)$, the hyperparameters vector, $(\boldsymbol{\beta}, \boldsymbol{\eta})$ are assumed to be independent and identically distributed as

$$p(\beta_i, \eta_i|\phi) = \gamma(\eta_i|a_\eta, b_\eta) \times \gamma(\beta_i|a_\beta, b_\beta). \tag{3.4}$$

S3. *Hyperparameters model:* The hyperparameters $(a_\eta, b_\eta, a_\beta, b_\beta)$ are assumed to be independent and identically distributed as a gamma density:

$$p_{NI}(\phi) = \prod_{x \in \phi} \frac{0.021^{0.556}}{\Gamma(0.556)} x^{0.556-1} e^{-0.021x}, \tag{3.5}$$

where p_{NI} denotes a noninformative prior choice for ϕ.

The fact that (3.4) is the natural conjugate prior for the pair (β_i, η_i) (cf. Oliveira et al., 2012) justifies the choice of second stage. But Ryan et al. (2011) did not discuss in detail the choice of stage 3. The authors jutified their choice as a tempting to be nearly noninformative. Next, we present and discuss some properties of the hierarchical PLP model above.

2.1 The marginal model

The marginal distribution of data $\mathbf{D}$ is obtained by averaging the data model with respect to the distribution of the parameters. We denote by $p(\mathbf{D}|\phi)$ the marginal model

$$\begin{aligned} p(\mathbf{D}|\phi) &= \int_{\mathbb{R}_+^{2k}} p(\mathbf{D}|\boldsymbol{\eta}, \boldsymbol{\beta}) p(\boldsymbol{\eta}, \boldsymbol{\beta}|\phi) d\boldsymbol{\eta} d\boldsymbol{\beta} \\ &= \prod_{i=1}^{k} \int_{\mathbb{R}_+^2} p(D_i|\eta_i, \beta_i) p(\beta_i, \eta_i|\phi) d\eta_i d\beta_i = \prod_{i=1}^{k} p(D_i|\phi), \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} p(D_i|\phi) &= \frac{\Gamma(a_\eta + n_i)}{\Gamma(a_\eta) n_i!} \left(\frac{b_\eta}{b_\eta + 1}\right)^{a_\eta} \left(\frac{1}{b_\eta + 1}\right)^{n_i} \\ &\quad \times n_i! \left(\prod_{j=1}^{n_i} \frac{1}{t_{ij}}\right) \frac{\Gamma(a_\beta + n_i)}{\Gamma(a_\beta)} \left(\frac{b_\beta}{b_\beta + w_i}\right)^{a_\beta} \left(\frac{1}{b_\beta + w_i}\right)^{n_i}, i = 1, \ldots, k, \end{aligned} \tag{3.7}$$

and $w_i = \sum_{j=1}^{n_i} \log(\tau/t_{ij})$.

Due to the conjugacy of stages 1 and 2 of the model, the integrals in (3.6) have closed form solutions. Note that (3.7) is a joint model for the number of failures n_i, and failure times $\{t_{ij}, j = 1, \ldots, n_i\}$ obtained by the conditional model of $\{t_{ij}\}$ given n_i, and the marginal model of n_i. Furthermore, $N_i = N_i(\tau)$ is distributed as a negative binomial distribution, here denoted by $N_i \sim \text{NegBinom}(a_\eta; (b_\eta + 1)^{-1})$. We use (3.6) to derive empirical Bayes estimates and the Jeffreys' prior in Sections 3 and 4.

2.2 Condidional posterior distributions of β_i and η_i

From the formulation of stages 1 and 2 of the hierarchical PLP model, it is easy to see that the conditional posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\eta})$ given ϕ is $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \phi) = \prod_{i=1}^{k} p(\beta_i, \eta_i|D_i, \phi)$, where

$$p(\beta_i, \eta_i|D_i, \phi) = \gamma(\beta_i|a_\beta + n_i, b_\beta + w_i) \times \gamma(\eta_i|a_\eta + n_i, b_\eta + 1), \ i = 1, \ldots, k. \tag{3.8}$$

In other words, given data and ϕ, (β_i, η_i) are conditionally independent distributed as gamma densities. Note that (3.8) does not depend on the choice of the prior distribution for ϕ. Moreover, given the hyperparameters, the posterior means of β_i and η_i are given by

$$\text{E}[\beta_i|D_i, \phi] = \left(\frac{b_\beta}{b_\beta + w_i}\right) \text{E}[\beta_i|\phi] + \left(\frac{w_i}{b_\beta + w_i}\right) \hat{\beta}_i, \tag{3.9}$$

$$\text{E}[\eta_i|D_i, \phi] = \left(\frac{b_\eta}{b_\eta + 1}\right) \text{E}[\eta_i|\phi] + \left(\frac{1}{b_\eta + 1}\right) \hat{\eta}_i, \tag{3.10}$$

where $\hat{\beta}_i = n_i/w_i$ and $\hat{\eta}_i = n_i$ are the maximum likelihood estimates (MLE) of β_i and η_i, respectively, based on data from the i-th system. One can observe that as $b_\eta \to 0$ the conditional posterior mean $\mathrm{E}[\eta_i|D_i, \phi] \to \hat{\eta}_i$, meaning that the parameters η_i are more heterogeneous. On the other hand, if $b_\eta \to \infty \Rightarrow \mathrm{E}[\eta_i|D_i, \phi] \to \mathrm{E}[\eta_i|\phi]$, and the parameters η_i are all equal (as $\mathrm{Var}[\eta_i|\phi] = a_\eta/b_\eta^2 \to 0$). If we define $p_\eta = b_\eta/(b_\eta + 1)$ as a shrinkage parameter, we note that p_η controls the relative contribution of the prior mean to the posterior mean. Similar observiations could be done for the conditional posterior mean of β_i (3.9).

Finally, the conditional posterior variances of β_i and η_i given ϕ are as follow

$$\mathrm{Var}[\beta_i|D_i, \phi] = \frac{a_\beta + n_i}{(b_\beta + w_i)^2}, \tag{3.11}$$

$$\mathrm{Var}[\eta_i|D_i, \phi] = \frac{a_\eta + n_i}{(b_\eta + 1)^2}. \tag{3.12}$$

In Section 3 we show how to use (3.11) and (3.12) to produce empirical Bayes estimates.

2.3 Joint posterior distribution of β, η, ϕ

When looking to the joint posterior of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \phi)$, we note the following simplification

$$p(\boldsymbol{\beta}, \boldsymbol{\eta}, \phi|\mathbf{D}) = \left\{ \prod_{i=1}^{k} p(\beta_i, \eta_i|D_i, \phi) \right\} \times p(\phi|\mathbf{D}), \tag{3.13}$$

where $p(\beta_i, \eta_i|D_i, \phi)$ is given by (3.8), $p(\phi|\mathbf{D}) \propto p(\phi) \times p(\mathbf{D}|\phi)$ is the marginal posterior of ϕ, and $p(\phi)$ is a generic prior for ϕ. This result will be usefull to obtain simulations from $p(\boldsymbol{\eta}, \boldsymbol{\beta}, \phi|\mathbf{D})$ (cf. Section 4.1). In case $p(\phi)$ is the noninformative prior (3.5), there is a semi-conjugacy between stages 2 and 3 of the model

$$\begin{aligned} p(\phi|\boldsymbol{\beta}, \boldsymbol{\eta}) \propto & \left(b_\eta^{a_\eta k + 0.556} e^{-b_\eta(\sum_{i=1}^k \eta_i + 0.021)} \right) \times \left(\frac{\prod_{i=1}^k \left(\eta_i^{a_\eta - 1} \right)}{\Gamma(a_\eta)} a_\eta^{0.556} e^{-a_\eta 0.021} \right) \\ & \times \left(b_\beta^{a_\beta k + 0.556} e^{-b_\beta(\sum_{i=1}^k \beta_i + 0.021)} \right) \times \left(\frac{\prod_{i=1}^k \left(\beta_i^{a_\beta - 1} \right)}{\Gamma(a_\beta)} a_\beta^{0.556} e^{-a_\beta 0.021} \right). \end{aligned}$$

Ryan et al. (2011) used this fact to propose a Gibbs sampling algorithm to simulate from the posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \phi)$, where only two parameters need a Metropolis step to be simulated. The rest of the $2k + 2$ parameters are simulated directly from their respective full-conditional distributions (see Ryan et al., 2011).

3 The empirical Bayes estimates

The empirical Bayes method (here, we make reference to the parametric empirical Bayes in the sense of Morris (1983)) consists in replacing ϕ by $\hat{\phi} = (\hat{a}_\eta, \hat{b}_\eta, \hat{a}_\beta, \hat{b}_\beta)$, a estimate of ϕ (generally the maximum likelihood estimate) in the conditional posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\eta})$. Then, inference with respect $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ can be done by this approximate distribution

$$\tilde{p}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}) = p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \hat{\phi}), \tag{3.14}$$

where

$$\hat{\phi} = \arg\max_{\mathbb{R}^4_+} p(\mathbf{D}|\phi),$$

and $p(\mathbf{D}|\phi)$ is the marginal likelihood function given by (3.6).

The resulting conjugacy between first and second stages of the hierarchical PLP model, and the relative simple expression of the marginal model, makes the PEB method a first natural approach to make inference in this hierarchical set. For a discussion about the PEB approach see Morris (1983) or, to examples of the use of PEB in the reliability literature see, for instance, Gaver and O'Muircheartaigh (1987); Martz et al. (1999).

Note that the approximate posterior distribution (3.14) does not take into account the uncertainty with respect to the estimation of ϕ. To overcome this issue, we use first-order approximations for posterior means and variances of $g(\beta_i, \eta_i)$, where $g(\cdot)$ is a generic function. Here, we briefly describe this method for the hierarchical PLP model. The general approach to provide these approximations is given in Kass and Steffey (1989). First note that given ϕ, the pair (β_i, η_i) does not depend on data from j-th system D_j, for $j \neq i$, and then we have the following expressions for the posterior mean and variance

$$\begin{aligned}
\mathrm{E}\left[g(\beta_i,\eta_i)|\mathbf{D}\right] &= \mathrm{E}\left[\mathrm{E}\left[g(\beta_i,\eta_i)|D_i,\phi\right]\right], && (3.15)\\
\mathrm{Var}\left[g(\beta_i,\eta_i)|\mathbf{D}\right] &= \mathrm{E}\left[\mathrm{Var}\left[g(\beta_i,\eta_i)|D_i,\phi\right]\right] + \mathrm{Var}\left[\mathrm{E}\left[g(\beta_i,\eta_i)|D_i,\phi\right]\right]. && (3.16)
\end{aligned}$$

Now, if we replace the inner expectations in the right hand side of expressions above by their respective first-order Taylor expansions, and evaluate (3.15) and (3.16) in $g(\beta_i, \eta_i) = \beta_i$ and $g(\beta_i, \eta_i) = \eta_i$, we get the respective firt-order approximations to posterior means and variances of (β_i, η_i)

$$\begin{aligned}
\tilde{\mathrm{E}}\left[\beta_i|\mathbf{D}\right] &= \mathrm{E}\left[\beta_i|D_i,\hat{\phi}\right] && (3.17)\\
\tilde{\mathrm{Var}}\left[\beta_i|\mathbf{D}\right] &= \mathrm{Var}\left[\beta_i|D_i,\hat{\phi}\right] + \frac{1}{(\hat{b}_\beta + w_i)^2}\mathrm{Var}\left(\hat{a}_\beta\right) + \frac{(\hat{a}_\beta + n_i)^2}{(\hat{b}_\beta + w_i)^4}\mathrm{Var}\left(\hat{b}_\beta\right)\\
&\quad -2\frac{(\hat{a}_\beta + n_i)}{(\hat{b}_\beta + w_i)^3}\mathrm{Cov}(\hat{a}_\beta, \hat{b}_\beta), && (3.18)\\
\tilde{\mathrm{E}}\left[\eta_i|\mathbf{D}\right] &= \mathrm{E}\left[\eta_i|D_i,\hat{\phi}\right] && (3.19)\\
\tilde{\mathrm{Var}}\left[\eta_i|\mathbf{D}\right] &= \mathrm{Var}\left[\eta_i|D_i,\hat{\phi}\right] + \frac{1}{(\hat{b}_\eta + 1)^2}\mathrm{Var}\left(\hat{a}_\eta\right) + \frac{(\hat{a}_\eta + n_i)^2}{(\hat{b}_\eta + 1)^4}\mathrm{Var}\left(\hat{b}_\eta\right)\\
&\quad -2\frac{(\hat{a}_\eta + n_i)}{(\hat{b}_\eta + 1)^3}\mathrm{Cov}(\hat{a}_\eta, \hat{b}_\eta), && (3.20)
\end{aligned}$$

where $\mathrm{E}\left[\beta_i|D_i,\hat{\phi}\right]$, $\mathrm{E}\left[\eta_i|D_i,\hat{\phi}\right]$, $\mathrm{Var}\left[\beta_i|D_i,\hat{\phi}\right]$ $\mathrm{Var}\left[\eta_i|D_i,\hat{\phi}\right]$ and are given by expressions (3.9)-(3.12) replacing ϕ by $\hat{\phi}$. The variances and covariances terms in expressions (3.18) and (3.20) are given by the elements of the negative inverse Hessian matrix of $\log p(\mathbf{D}|\phi)$. All these terms are evalueted at the MLE $\hat{\phi}$.

To close this section, we show how the posterior mean and variance approximations of β_i and η_i obtained above could be used to make inferences for the PLP parameters. Given the data, β_i has an approximate gamma distribution

$$(\beta_i|\mathbf{D}) \dot{\sim} \mathrm{Gamma}\left(\left(\tilde{\mathrm{E}}\left[\beta_i|\mathbf{D}\right]\right)^2/\tilde{\mathrm{Var}}\left[\beta_i|\mathbf{D}\right],\ \tilde{\mathrm{E}}\left[\beta_i|\mathbf{D}\right]/\tilde{\mathrm{Var}}\left[\beta_i|\mathbf{D}\right]\right),\ i = 1, \ldots, k. \quad (3.21)$$

Similarly, given the data, η_i has an approximate gamma distribution

$$(\eta_i|\mathbf{D}) \dot{\sim} \text{Gamma}\left(\left(\tilde{\text{E}}\left[\eta_i|\mathbf{D}\right]\right)^2/\tilde{\text{Var}}\left[\eta_i|\mathbf{D}\right],\ \tilde{\text{E}}\left[\eta_i|\mathbf{D}\right]/\tilde{\text{Var}}\left[\eta_i|\mathbf{D}\right]\right),\ i=1,\ldots,k. \tag{3.22}$$

The use of the gamma distribution for these approximations is reasonable due to the fact that β_i and η_i are gamma distributed conditionally on the data and hyperparameters (3.8). The PEB point estimates of β_i and η_i are given by their respective approximate posterior means (3.17) and (3.19). Interval estimates are obtained by taking the quantiles of (3.21) and (3.22). When interest is center in some function of (β_i, η_i), for example, the parameter $\theta_i = \tau/\eta_i^{1/\beta_i}$, one can draw a sample of the distribution (3.22) and take Monte Carlo estimates of θ_i.

4 Jeffreys' prior for hierarchical PLP model

Jeffreys' prior is defined to be proportional to the square root of the determinant of the Fisher Information matrix of the model (see Kass and Wasserman, 1996). For the hierarchical PLP, the Jeffrey's prior can be obtained from the marginal model with respect to the hyperparameters ϕ. Note that the likelihood function of the marginal model (3.6) depends upon the observed data via the statistics

$$\begin{aligned} N_i = N_i(\tau) &\sim \text{NegBinom}\left(a_\eta; \frac{1}{b_\eta+1}\right), \\ W_i = \sum_{j=1}^{n_i} \log(\tau/T_{ij}) &\sim \text{Gamma}(n_i, \beta_i), \end{aligned}$$

for $i = 1,\ldots,k$, (see Crow, 1977). Averaging W_i with respect to the distribution of β_i, it follows that W_i has density function given by

$$p(w_i|n_i,\phi) = \frac{\Gamma(a_\beta+n_i)}{\Gamma(n_i)\Gamma(a_\beta)} \times \frac{w_i^{n_i-1} b_\beta^{a_\beta}}{(b_\beta+w_i)^{a_\beta+n_i}},\ n_i>0,\ i=1,\ldots,k,$$

that is the resulting distribution from a Gamma-Gamma model.

It is easy to see that the Fisher Information matrix $\mathbf{I}(\phi)$ is block diagnoal (Appendix A)

$$\mathbf{I}(\phi) = \begin{bmatrix} I_{a_\eta a_\eta} & I_{a_\eta b_\eta} & 0 & 0 \\ I_{a_\eta b_\eta} & I_{b_\eta b_\eta} & 0 & 0 \\ 0 & 0 & I_{a_\beta a_\beta} & I_{a_\beta b_\beta} \\ 0 & 0 & I_{a_\beta b_\beta} & I_{b_\beta b_\beta} \end{bmatrix}, \tag{3.23}$$

whose elements are given by

$$
\begin{aligned}
I_{a_\eta a_\eta} &= k(\psi'(a_\eta) - \mathrm{E}\left[\psi'(a_\eta + N_1)\right]), && (3.24)\\
I_{a_\eta b_\eta} &= -\frac{k}{b_\eta(b_\eta+1)}, && (3.25)\\
I_{b_\eta b_\eta} &= \frac{k a_\eta}{b_\eta^2(b_\eta+1)}, && (3.26)\\
I_{a_\beta a_\beta} &= k(\psi'(a_\beta) - \mathrm{E}\left[\psi'(a_\beta + N_1)\right]), && (3.27)\\
I_{a_\beta b_\beta} &= \frac{k}{b_\beta}\left\{a_\beta \mathrm{E}\left[(a_\beta + N_1)^{-1}\right] - 1\right\}, && (3.28)\\
I_{b_\beta b_\beta} &= \frac{k}{b_\beta^2}\left\{a_\beta\left[1 - (a_\beta+1)\mathrm{E}\left[(a_\beta + N_1 + 1)^{-1}\right]\right]\right\}, && (3.29)
\end{aligned}
$$

where $\psi'(\cdot)$ is the second derivative of the logarithm of the gamma function, also known as the *trigamma* function. The elements (3.24) and (3.27)-(3.29) have no closed form, but they depend upon expectations with respect the distribution of $N_1 \sim \mathrm{NegBinom}\left(a_\eta; (b_\eta+1)^{-1}\right)$, which are easily computable (see Appendix B). Finally, we denote the Jeffreys' prior of ϕ by $p_J(\phi)$, and this is proportional to

$$
\det[\mathbf{I}(\phi)]^{1/2} = [I_{a_\eta a_\eta} I_{b_\eta b_\eta} - I^2_{a_\eta b_\eta}]^{1/2} \times [I_{a_\beta a_\beta} I_{b_\beta b_\beta} - I^2_{a_\beta b_\beta}]^{1/2}.
$$

4.1 Posterior simulations

As showed by (3.13), the joint posterior density of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \phi)$ is a product of the conditional posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\eta})$ given ϕ and the marginal posterior distribution of ϕ. This enables us to propose the following steps to simulate from $p(\boldsymbol{\beta}, \boldsymbol{\eta}, \phi|\mathbf{D})$:

1. Draw a sample $\phi^{(m)}$ from $p(\phi|\mathbf{D}) \propto p_J(\phi) \times p(\mathbf{D}|\phi)$, $m = 1, \ldots, M$;
2. Draw a sample $\beta_i^{(m)}$ from $\gamma(\beta_i|D_i, \phi^{(m)})$, $m = 1, \ldots, M$, $i =, \ldots, k$;
3. Draw a sample $\eta_i^{(m)}$ from $\gamma(\eta_i|D_i, \phi^{(m)})$, $m = 1, \ldots, M$, $i =, \ldots, k$.

In order to efficiently sample from the marginal posterior distribution of ϕ, we used an adaptive Metropolis algorithm (see Haario et al., 2001; Roberts and Rosenthal, 2009). This method consists in taking the well known Metropolis algorithm to simulate a Markov chain with stationary distribution given by $p(\phi|\mathbf{D})$, but with an adaptive proposal distribution. Therefore, in the initial steps of the chain ($s \leq s_0$), candidates for ϕ are simulated from the distribution $Q_s(\phi, \cdot) = N(\phi^{(s-1)}, (0.1)^2 I_4/4)$, where I_4 is the 4×4 identity matrix. For the subsequent steps of the chain ($s > s_0$), candidates for ϕ are simulated from the distribution $Q_s(\phi, \cdot) = (1-\epsilon)N(\phi^{(s-1)}, (2.38)^2\Sigma_s/4) + \epsilon N(\phi^{(s-1)}, (0.1)^2 I_4/4)$, where Σ_s is the empirical covariance matrix of the $s-1$ previous realizations of ϕ, and ϵ is chosen to be small (generally $\epsilon = 0.5$). After the convergence of the chain, we keep a sample of size M from $p(\phi|\mathbf{D})$. The PLP parameters are simulated exactly from their conditional posterior distribution (3.8), replacing ϕ by $\phi^{(m)}$, $m = 1, \ldots, M$.

5 A simulation study

In this section we present a simulation study in order to compare the performance of the proposed priors distributions: PEB, Jeffreys' and noninformative prior proposed by Ryan et al. (2011). We follow the same simulation scheme in Ryan et al. (2011). Comparisons are made with respect to coverage probabilities and average lengths of intervals.

Data D_i were generated conditionally independent from the PLP model (3.3) given the pair (β_i, η_i). Truncation time of the k systems were fixed at $3,000$ hours. We studied the effect of sample size by varying the number of systems, $k \in \{20, 50, 80\}$ and also the expected number of failures. The pairs (β_i, η_i) were generated by model (3.4) given ϕ. The true values of the hyperparameters were taken to be $(a_\eta, b_\eta) \in \{(2.10, 0.42), (4.20, 0.42), (8.40, 0.42)\}$, and $(a_\beta, b_\beta) = (11.25, 7.5)$. These configurations were set to obtain the corresponding prior mean and variance for β_i, $\mathrm{E}[\beta_i|\phi] = 1.5$ and $\mathrm{Var}[\beta_i|\phi] = 0.20$, and for η_i, $\mathrm{E}[\eta_i|\phi] \in \{5, 10, 20\}$ and $\mathrm{Var}[\eta_i|\phi] \in \{11.90, 23.81, 47.61\}$. One can note that $\mathrm{E}[\eta_i|\phi] = \mathrm{E}[N_i|\phi]$, thus the above configurations allowed us to studied also the effect of failures number by system.

For each of the nine scenarios, we have generated 1,000 data sets. Inferences for the Bayesian methods were based on chains of size 20,000, discarding the first 10,000 as burn-in, resulting in samples of size 10,000 of the posterior distribution $p(\phi|\mathbf{D})$. Although a Gibbs sampling algorithm can be adopted in the presence of the noninformative prior in the model (see Ryan et al., 2011), we used the approach presented in Section 4.1 for both noninformative and Jeffreys' priors. We believe that in this way fair comparisons can be made between both of them.

The criteria used to compare the three approaches were the nominal coverage of 90% and 95% intervals from the 5-th and 95-th, and 2.5-th and 97.5-th sample percentiles, respectively, and the average interval length. Tables 3.1 to 3.3 present the coverage and average interval length averaged over the k systems for β_i, η_i estimated by the empirical Bayes (PEB), noninformative (NI) and Jeffreys' prior. All the results were obtained using the software R (R Development Core Team, 2013).

The results show that the Jeffreys' prior appears to dominate the noniformative prior and the PEB in terms of better coverage for β_i and η_i. One can see that as the number of systems increases the coverage of all three approaches becomes very close. However Jeffreys' prior performance seems to be better than the others. The intervals for β_i produced by PEB and noninformative prior are smaller than the intervals produced by the Jeffreys' prior. It can explain why the coverage from these two methods are smaller than the coverage from the Jeffreys' prior. Intervals for η_i produced by PEB are larger than the ones produced by the Jeffreys' and noninformative priors.

6 Valve failure data

In this section, we use a failure times data from a set of valves presented by Bhattacharjee et al. (2003). Bhattacharjee et al. (2003) presented a data set from nine years follow up of 104 systems. Here, we use a subset of 16 valves, those that presented at least one failure. Failures history from these 16 systems are displayed in Figure 3.1. It can be observed that all systems have the same truncation time $\tau = 3,286$ days. Systems present from one to eight failures per system, with different patterns.

Table 3.1: Coverage probabilities and average interval length for the PLP parameters ($\mathrm{E}[\eta_i|\phi] = 5$).

	β_i				η_i			
	0.90		0.95		0.90		0.95	
	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length
$k = 20$								
PEB	0.628	0.945	0.676	1.129	0.928	7.352	0.962	8.831
NI	0.845	1.211	0.915	1.502	0.842	5.452	0.911	6.556
Jeffreys	0.859	1.410	0.915	1.751	0.875	5.795	0.932	6.956
$k = 50$								
PEB	0.781	1.110	0.834	1.325	0.921	6.536	0.963	7.839
NI	0.844	1.170	0.914	1.436	0.886	5.723	0.941	6.862
Jeffreys	0.864	1.299	0.923	1.593	0.895	5.838	0.946	7.000
$k = 80$								
PEB	0.837	1.167	0.890	1.393	0.917	6.307	0.960	7.562
NI	0.854	1.171	0.919	1.428	0.893	5.782	0.945	6.931
Jeffreys	0.879	1.274	0.934	1.551	0.897	5.846	0.948	7.008

Table 3.2: Coverage probabilities and average interval length for the PLP parameters ($\mathrm{E}[\eta_i|\phi] = 10$).

	β_i				η_i			
	0.90		0.95		0.90		0.95	
	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length
$k = 20$								
PEB	0.760	0.946	0.815	1.130	0.974	12.846	0.990	15.390
NI	0.857	1.060	0.920	1.290	0.859	8.137	0.924	9.749
Jeffreys	0.874	1.150	0.932	1.400	0.873	8.434	0.930	10.094
$k = 50$								
PEB	0.856	1.038	0.914	1.239	0.948	10.518	0.979	12.577
NI	0.867	1.048	0.928	1.266	0.887	8.377	0.941	10.016
Jeffreys	0.882	1.101	0.938	1.328	0.895	8.514	0.947	10.177
$k = 80$								
PEB	0.872	1.053	0.928	1.257	0.934	9.768	0.971	11.677
NI	0.873	1.051	0.931	1.265	0.891	8.390	0.944	10.027
Jeffreys	0.885	1.089	0.939	1.310	0.895	8.471	0.947	10.123

Table 3.3: Coverage probabilities and average interval length for the PLP parameters ($\mathrm{E}[\eta_i|\phi] = 20$).

	β_i				η_i			
	0.90		0.95		0.90		0.95	
	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length	Cov. prob.	Aver. int. length
$k = 20$								
PEB	0.848	0.848	0.906	1.012	0.993	23.132	0.998	27.682
NI	0.874	0.884	0.933	1.062	0.872	11.762	0.931	14.071
Jeffreys	0.883	0.914	0.937	1.097	0.864	11.864	0.917	14.185
$k = 50$								
PEB	0.882	0.882	0.936	1.052	0.977	17.500	0.993	20.902
NI	0.884	0.888	0.940	1.063	0.890	11.997	0.943	14.325
Jeffreys	0.891	0.903	0.944	1.081	0.895	12.151	0.947	14.506
$k = 80$								
PEB	0.890	0.890	0.943	1.061	0.961	15.745	0.986	18.797
NI	0.891	0.892	0.945	1.066	0.892	12.058	0.944	14.391
Jeffreys	0.895	0.902	0.947	1.079	0.894	12.150	0.947	14.501

Chains of size 20,000, using the scheme presented in Section 4.1, were used to obtain the fitted values from our proposed approach. We discarded the first 10,000 values of chains as burn-in and the last 10,000 steps were used as a sample from the posterior distribution $p(\phi|\mathbf{D})$. In Figure 3.2 it is possible to compare the effect of the different prior choices for ϕ. One can see that the use of the noninformative prior results in a more diffuse posterior distribution for ϕ (Figure 3.2 (a) and (b)), when it is compared to the Jeffrey's prior (Figure 3.2 (c) and (d)). Furthermore, the posterior distribution of ϕ based on the Jeffreys' prior seems to be concentraed around the maximum likelihood estimates of ϕ, ($\hat{a}_\eta = 3.07, \hat{b}_\eta = 1.33, \hat{a}_\beta = 5.93, \hat{b}_\beta = 5.77$).

Finnaly, we compare the empirical Bayes (PEB), noninformative and Jeffreys estimates of the

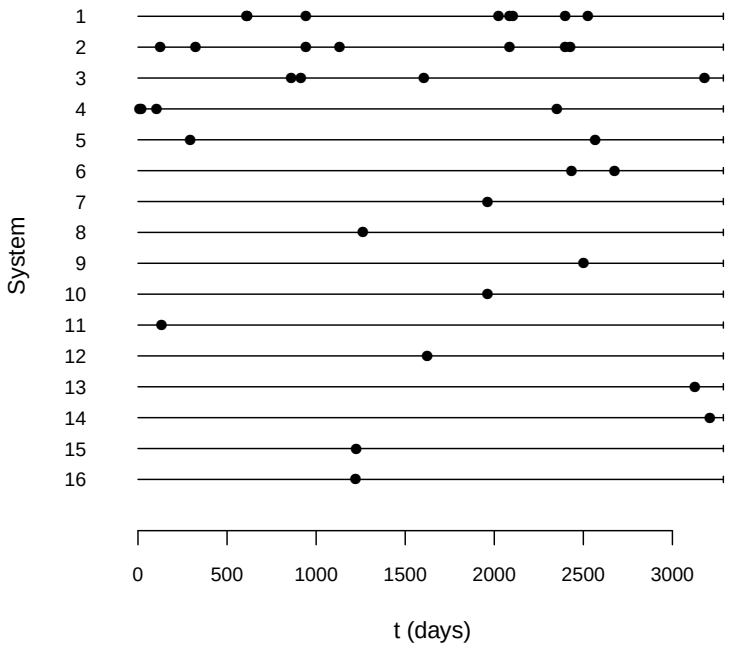

Figure 3.1: Event plot of 16 valves.

PLP parameters. The shape parameter β_i indicates whether the system is deteriorating ($\beta_i > 1$), or improving ($\beta_i < 1$), and the scale parameter $\eta_i = \mathrm{E}\left[N_i(\tau)\right]$ indicates the expect number of failures by the truncation time. Table 3.4 shows posterior means and 90% intervals for β_i and η_i. The posterior means were approximated by sample means, for example, $\mathrm{E}\left[\beta_i | D_i, \phi\right] \approx (10,000)^{-1} \sum_{m=1}^{10,000} \beta_i^{(m)}$, where $\{\beta_i^{(m)}\}$ is a sample obtained with the scheme presented in Section 4.1.

Table 3.4: Estimates of the PLP parameters for valve data: posterior means (90% intervals in parentheses).

Valve i	β_i			η_i		
	PEB	NI	Jeffreys	PEB	NI	Jeffreys
1	1.13 (0.68;1.67)	1.06 (0.65;1.62)	1.14 (0.65;1.79)	4.75 (2.09;8.29)	3.69 (1.97;6.58)	5.19 (2.63;8.65)
2	0.88 (0.52;1.31)	0.87 (0.52;1.27)	0.86 (0.48;1.34)	4.32 (1.87;7.60)	3.45 (1.90;6.09)	4.72 (2.40;7.93)
3	1.09 (0.59;1.71)	1.01 (0.57;1.63)	1.12 (0.54;1.94)	3.04 (1.21;5.54)	2.71 (1.48;4.52)	3.26 (1.49;5.74)
4	0.48 (0.26;0.76)	0.56 (0.27;0.88)	0.45 (0.19;0.75)	3.04 (1.21;5.54)	2.71 (1.48;4.52)	3.27 (1.50;5.79)
5	0.94 (0.46;1.55)	0.91 (0.46;1.50)	0.93 (0.38;1.69)	2.18 (0.77;4.16)	2.23 (1.06;3.64)	2.29 (0.84;4.29)
6	1.26 (0.62;2.10)	1.17 (0.58;2.19)	1.54 (0.59;3.24)	2.18 (0.77;4.16)	2.23 (1.07;3.62)	2.28 (0.83;4.36)
7	1.10 (0.51;1.88)	1.04 (0.50;1.89)	1.24 (0.43;2.56)	1.75 (0.56;3.47)	1.98 (0.78;3.32)	1.79 (0.50;3.61)
8	1.03 (0.48;1.75)	0.98 (0.47;1.71)	1.09 (0.38;2.19)	1.75 (0.56;3.47)	1.98 (0.79;3.31)	1.78 (0.49;3.52)
9	1.15 (0.53;1.95)	1.08 (0.52;1.96)	1.36 (0.46;2.99)	1.75 (0.56;3.47)	1.99 (0.80;3.29)	1.79 (0.49;3.60)
10	1.10 (0.51;1.88)	1.04 (0.36;1.28)	1.24 (0.43;2.58)	1.75 (0.56;3.47)	1.97 (0.80;3.26)	1.78 (0.50;3.55)
11	0.77 (0.36;1.31)	0.78 (0.51;1.88)	0.71 (0.23;1.33)	1.75 (0.56;3.47)	1.99 (0.81;3.33)	1.81 (0.50;3.62)
12	1.07 (0.50;1.82)	1.02 (0.49;1.81)	1.16 (0.43;2.33)	1.75 (0.56;3.47)	2.00 (0.78;3.33)	1.78 (0.51;3.56)
13	1.19 (0.55;2.03)	1.12 (0.52;2.12)	1.51 (0.50;3.47)	1.75 (0.56;3.47)	1.98 (0.77;3.27)	1.78 (0.50;3.60)
14	1.20 (0.55;2.04)	1.12 (0.53;2.12)	1.53 (0.50;3.52)	1.75 (0.56;3.47)	1.98 (0.78;3.25)	1.79 (0.49;3.56)
15	1.03 (0.48;1.74)	0.97 (0.47;1.68)	1.08 (0.38;2.13)	1.75 (0.56;3.47)	1.98 (0.81;3.29)	1.79 (0.50;3.60)
16	1.02 (0.48;1.74)	0.97 (0.47;1.70)	1.07 (0.38;2.12)	1.75 (0.56;3.47)	1.97 (0.76;3.28)	1.80 (0.52;3.62)

We have found similar results for all three approaches, specially for β_i. For all the 16 systems, the 90% intervals for parameters β_i contains the value 1, indicating that the systems follow an homogeneous Poisson process (constant intensity). Note that, in general, the intervals for β_i for systems with more failures are smaller than these for systems with only one failure. On the other

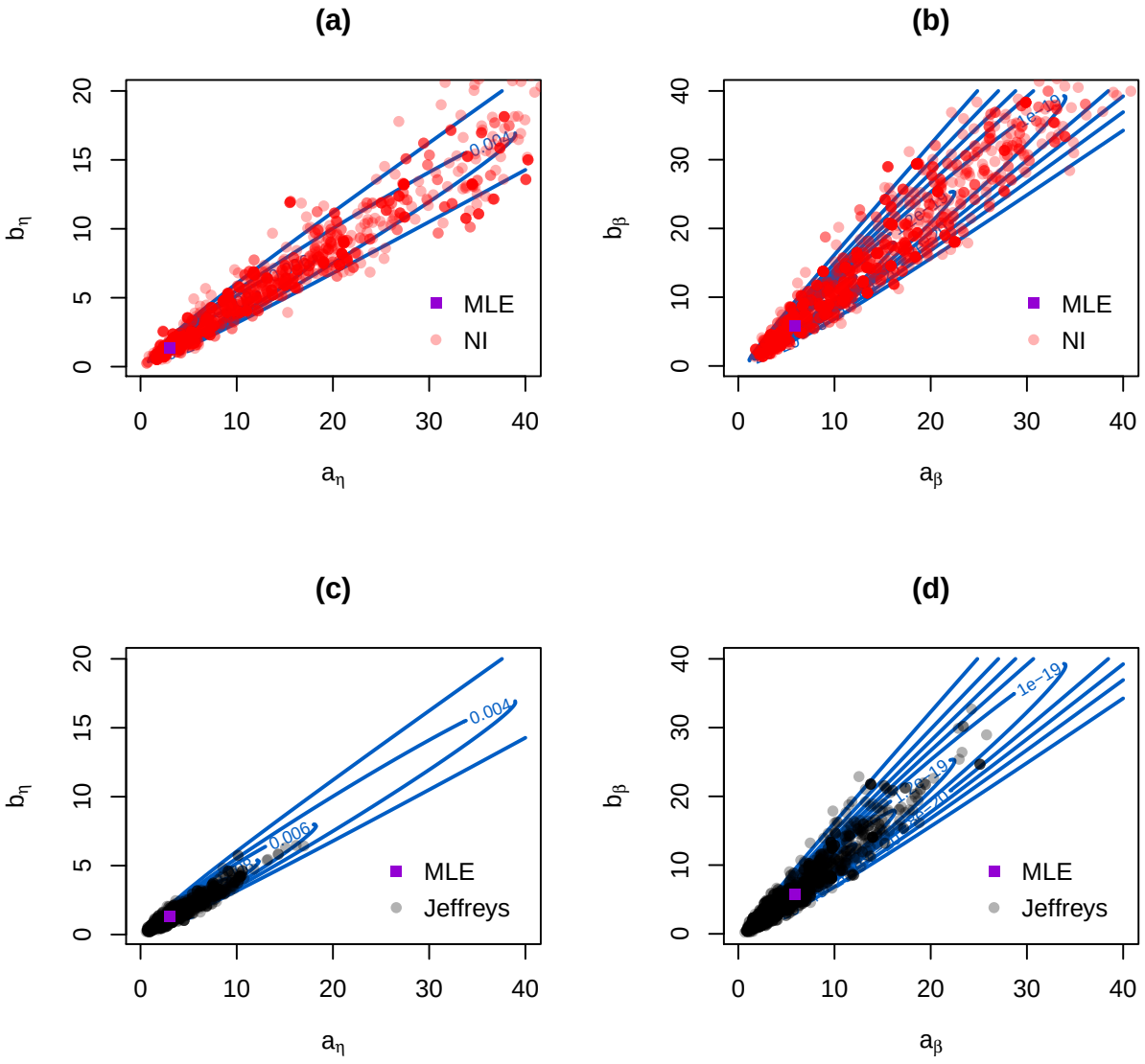

Figure 3.2: Contour plots of the marginal likelihood with posterior distributions of ϕ based on the noninformative (**(a)** and **(b)**) and Jeffreys' (**(c)** and **(d)**) priors.

hand, the estimates for η_i are indicating that we could expect different number of failures for each system. We can see that the intervals produced by the Jeffreys' prior are wider than the intervals produced by the noninformative prior. However, it is possible that the Jeffreys' prior are getting better results as indicated by the simulations presented in Section 5.

7 Conclusions

In this paper we discussed the choice of the prior distribution for the hyperparameters of a hieararchical PLP model. This model has interesting properties for the analysis of multiple repairable systems. However, the choice of hyperparameters prior is allways a difficult task. We developed alternatives approachs to the prior distribution proposed by Ryan et al. (2011). The first approach is a parametric empirical Bayes with adjusted posterior variances to take into account the uncertainty in estimating ϕ. The second approach is a Jeffreys' prior for ϕ. The resulting complex posterior dis-

tribution need Markov Chain Monte Carlo methods to approximate it. We showed how an adaptive Metropolis can be used to accomplish this task. Comparisons between the three approaches were made by simulations and a real data set illustration. Jeffreys' prior seems to be a promise method as a "default" (or objective) Bayesian analysis.

Acknowledgments

This work was partially supported by CNPq, CAPES, FAPEMIG and FAPDF.

A Derivation of the Fisher Information matrix $I(\phi)$

The contribution of the i-th system for the log-likelihood function of the marginal model (3.6) is

$$\begin{aligned}\log p(D_i|\phi) &= \log\Gamma(a_\eta + n_i) - \log\Gamma(a_\eta) + a_\eta \log b_\eta - (a_\eta + n_i)\log(b_\eta + 1) \\ &+ \log\Gamma(a_\beta + n_i) - \log\Gamma(a_\beta) + a_\beta \log b_\beta - (a_\beta + n_i)\log(b_\beta + w_i) \\ &- \sum_{j=1}^{n_i} \log t_{ij}. \end{aligned} \tag{3.30}$$

From (3.30) it is easy to note that

$$\frac{\partial^2}{\partial a_\eta \partial a_\beta}\log p(D_i|\phi) = \frac{\partial^2}{\partial a_\eta \partial b_\beta}\log p(D_i|\phi) = \frac{\partial^2}{\partial b_\eta \partial a_\beta}\log p(D_i|\phi) = \frac{\partial^2}{\partial b_\eta \partial b_\beta}\log p(D_i|\phi) = 0,$$

and it follows that the Fisher Information matrix for one system is

$$\begin{bmatrix} i_{a_\eta a_\eta} & i_{a_\eta b_\eta} & 0 & 0 \\ i_{a_\eta b_\eta} & i_{b_\eta b_\eta} & 0 & 0 \\ 0 & 0 & i_{a_\beta a_\beta} & i_{a_\beta b_\beta} \\ 0 & 0 & i_{a_\beta b_\beta} & i_{b_\beta b_\beta} \end{bmatrix},$$

whose elements are given by

$$\begin{aligned} i_{a_\eta a_\eta} &= \mathrm{E}\left[-\frac{\partial^2}{\partial a_\eta^2}\log p(D_i|\phi)\right] = \mathrm{E}\left[\psi'(a_\eta) - \psi'(a_\eta + N_i)\right], \\ i_{a_\eta b_\eta} &= \mathrm{E}\left[-\frac{\partial^2}{\partial a_\eta \partial b_\eta}\log p(D_i|\phi)\right] = -\frac{1}{b_\eta(b_\eta+1)}, \\ i_{b_\eta b_\eta} &= \mathrm{E}\left[-\frac{\partial^2}{\partial b_\eta^2}\log p(D_i|\phi)\right] = \mathrm{E}\left[\frac{a_\eta}{b_\eta^2} - \frac{a_\eta + N_i}{(b_\eta+1)^2}\right], \\ i_{a_\beta a_\beta} &= \mathrm{E}\left[-\frac{\partial^2}{\partial a_\beta^2}\log p(D_i|\phi)\right] = \mathrm{E}\left[\psi'(a_\beta) - \psi'(a_\beta + N_i)\right], \\ i_{a_\beta b_\beta} &= \mathrm{E}\left[-\frac{\partial^2}{\partial a_\beta \partial b_\beta}\log p(D_i|\phi)\right] = \mathrm{E}\left[\frac{1}{b_\beta + W_i} - \frac{1}{b_\beta}\right], \\ i_{b_\beta b_\beta} &= \mathrm{E}\left[-\frac{\partial^2}{\partial b_\beta^2}\log p(D_i|\phi)\right] = \mathrm{E}\left[\frac{a_\beta}{b_\beta^2} - \frac{a_\beta + N_i}{(b_\beta + W_i)^2}\right]. \end{aligned}$$

The elements $i_{a_\eta a_\eta}$ and $i_{a_\beta a_\beta}$ depend upon a expectation of the *trigamma* function, and have no closed form and their computation are detailed in Appendix B. Element $i_{b_\eta b_\eta}$ depends only by $\mathrm{E}[N_i] = a_\eta/b_\eta$, and is given by $a_\eta/(b_\eta^2(b_\eta+1))$. Finally, the elements $i_{a_\beta b_\beta}$ and $i_{b_\beta b_\beta}$ depend upon expectations with respect of the statistic W_i. Using the *law of iterated expectations*, we get that $i_{a_\beta b_\beta} = \mathrm{E}[\mathrm{E}[(b_\beta + W_i)|N_i]] - b_\beta^{-1}$ and $i_{b_\beta b_\beta} = a_\beta/b_\beta - \mathrm{E}[(a_\beta + N_i)\mathrm{E}[(b_\beta + W_i)^{-2}|N_i]]$. Using the fact that

$$\int_0^\infty p(w_i|n_i,\phi)dw_i = \int_0^\infty \frac{\Gamma(a_\beta+n_i)}{\Gamma(n_i)\Gamma(a_\beta)} \times \frac{w_i^{n_i-1}b_\beta^{a_\beta}}{(b_\beta+w_i)^{a_\beta+n_i}}dw_i = 1,$$

we obtain the following identity

$$\int_0^\infty \frac{w_i^{n_i-1}}{(b_\beta+w_i)^{a_\beta+n_i}}dw_i = \frac{\Gamma(n_i)\Gamma(a_\beta)}{\Gamma(a_\beta+n_i)}b_\beta^{-a_\beta}. \tag{3.31}$$

With this fact, it is possible to solve the inner expectations $\mathrm{E}[(b_\beta+W_i)^{-1}|N_i=n_i] = a_\beta/(b_\beta(a_\beta+n_i))$ and $\mathrm{E}[(b_\beta+W_i)^{-2}|N_i=n_i] = a_\beta(a_\beta+1)/(b_\beta^2(a_\beta+n_i)(a_\beta+n_i+1))$ using (3.31) with $a_\beta = a_\beta+1$ and $a_\beta = a_\beta+2$, respectively. Now, the elements $i_{a_\beta b_\beta}$ and $i_{b_\beta b_\beta}$ can be written as

$$i_{a_\beta b_\beta} = \frac{1}{b_\beta}\left\{a_\beta \mathrm{E}[(a_\beta+N_i)^{-1}] - 1\right\} \text{ and } i_{b_\beta b_\beta} = \frac{a_\beta}{b_\beta^2}\left\{1-(a_\beta+1)\mathrm{E}[(a_\beta+N_i+1)^{-1}]\right\}.$$

The Fisher Information matrix of the marginal model (3.6) is given by summing up in i each of the elements of the Fisher Information matrix for one system.

B Computation of the Jeffreys' prior $p_J(\phi)$

In order to compute the Jeffreys' prior for the hierarchical PLP model it is necessary to evaluate the elements of the Fisher Information matrix $I(\phi)$. Some of these elements, $I_{a_\eta a_\eta}, I_{a_\beta a_\beta}, I_{a_\beta b_\beta}, I_{b_\beta b_\beta}$, have no closed form, depending upon positive series. Here we show how to efficiently approximate these series. For example, note that element (3.24) is given by

$$k(\psi'(a_\eta) - \mathrm{E}[\psi'(a_\eta+N_1)]) = k\left\{\psi'(a_\eta) - \left[\psi'(a_\eta)\left(\frac{b_\eta}{b_\eta+1}\right)^{a_\eta} + \sum_{r=1}^{\infty}\psi'(a_\eta+r)p(N_1=r|\phi)\right]\right\},$$

where

$$p(N_1=r|\phi) = \frac{\Gamma(a_\eta+r)}{\Gamma(a_\eta)r!}\left(\frac{b_\eta}{b_\eta+1}\right)^{a_\eta}\left(\frac{1}{b_\eta+1}\right)^{r}.$$

Let $c_r = \psi'(a_\eta+r)p(N_1=r|\phi)$ be the general term of the series $S = \sum_{r=1}^\infty c_r$, then

$$\begin{aligned}\frac{c_{r+1}}{c_r} &= \frac{\psi'(a_\eta+r+1)}{\psi'(a_\eta+r)} \times \frac{p(N_1=r+1|\phi)}{p(N_1=r|\phi)} \\ &= \left\{1 - \frac{1}{[(a_\eta+r)\psi'(a_\eta+r)](a_\eta+r)}\right\} \times \left(1+\frac{a_\eta-1}{r+1}\right) \times \left(\frac{1}{b_\eta+1}\right).\end{aligned}$$

Note that $\psi'(a_\eta+r+1) = \psi'(a_\eta+r) - 1/(a_\eta+r)^2$, and $\lim_{r\to\infty}(a_\eta+r)\psi'(a_\eta+r) = 1$ (see Jeffreys and Jeffreys, 1950, pg. 466), then $\lim_{r\to\infty} c_{r+1}/c_r = (b_\eta+1)^{-1} = L < 1$. The series S can be bounded above and below by the sequences $L_r = S_r + c_r(L/(1-L))$ and $U_r = S_r + c_{r+1}/(1-c_{r+1}/c_r)$, respectively, where $S_r = \sum_{\ell=1}^r c_\ell$ is the r-th partial sum (see Braden, 1992). To approximate S with

error less than ε (for example, $\varepsilon = 10^{-6}$), we could use the value r with $U_r - L_r < 2\varepsilon$ and take the middlepoint of the interval (L_r, U_r). Approximations for the series in the elements (3.27)-(3.29) are given in a similar way.

References

S. Arima, G. S. Datta, and B. Liseo. Objective Bayesian analysis of a measurement error small area model. *Bayesian Analysis*, 7(2):363–384, 06 2012. doi: 10.1214/12-BA712.

H. Ascher and H. Feingold. *Repairable systems reliability: modelling, inference, misconceptions and their causes*. Marcel Dekker Inc, New York, 1984.

L. J. Bain and M. Engelhardt. Inferences on the parameters and current system reliability for a time truncated Weibull process. *Technometrics*, 22:421–426, 1980.

R. D. Baker. Some new tests of the power law process. *Technometrics*, 38:256–265, 1996.

S. Bar-Lev, I. Lavi, and B. Reiser. Bayesian inference for the power law process. *Annals of the Institute of Statistical Mathematics*, 44(4):623–639, 1992. URL http://EconPapers.repec.org/RePEc:spr:aistmt:v:44:y:1992:i:4:p:623-639.

M. Bhattacharjee, E. Arjas, and U. Pulkkinen. Modelling heterogeneity in nuclear power plant valve failure data. In B. H. Lindqvist and K. A. Doksum, editors, *Mathematical and Statistical Methods for Reliability*, pages 341–353. World Scientific Publishing, 2003.

B. Braden. Calculating sums of infinite series. *The American Mathematical Monthly*, 99:649–655, 1992.

L. H. Crow. Confidence interval procedures for reliability growth analysis. Technical report, US Army Materiel Systems Analysis Activity, 1977.

D. P. Gaver and I. G. O'Muircheartaigh. Robust empirical Bayes analyses of event rates. *Technometrics*, 29:1–15, 1987. doi: 10.2307/1269878.

M. Giorgio, M. Guida, and G. Pulcini. Repairable system analysis in presence of covariates and random effects. *Reliability Engineering and System Safety*, 131:271–281, 2014. URL http://dx.doi.org/10.1016/j.ress.2014.04.009i.

M. Guida, R. Calabria, and G. Pulcini. Bayes inference for a non-homogeneous Poisson process with power intensity law. *IEEE Transactions on Reliability*, 38:603–609, 1989. doi: 10.1109/24.46489.

H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

H. Jeffreys and B. S. Jeffreys. *Methods of Mathematical Physics*. Cambridge University Press, second edition, 1950.

R. E. Kass and D. Steffey. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, 84: 717–726, 1989. doi: 10.1080/01621459.1989.10478825.

R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91:1343–1370, 1996.

B. H. Lindqvist. On the statistical modeling and analysis of repairable systems. *Statistical Science*, 21:532–551, 2006. doi: 10.1214/088342306000000448.

H. F. Martz, R. L. Parker, and D. M. Rasmuson. Estimation of trends in the scram rate at nuclear power plants. *Technometrics*, 41:352–364, 1999.

C. N. Morris. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78:47–55, 1983. doi: doi:10.2307/2287098.

R. Natarajan and R. E. Kass. Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, 95(449):227–237, 2000. ISSN 01621459. doi: 10.2307/2669540. URL http://dx.doi.org/10.2307/2669540.

M. D. Oliveira, E. A. Colosimo, and G. L. Gilardoni. Bayesian inference for power law processes with applications in repairable systems. *Journal of Statistical Planning and Inference*, 142:1151–1160, 2012. doi: 10.1016/j.jspi.2011.11.016.

R. Pan and S. E. Rigdon. Bayes inference for general repairable systems. *Journal of Quality Technology*, 41:82–94, 2009.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/. ISBN 3-900051-07-0.

S. E. Rigdon and A. P. Basu. *Statistical Methods for the Reliability of Repairable Systems*. John Wiley & Sons, 2000.

G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18:349–367, 2009.

K. J. Ryan, M. S. Hamada, and C. S. Reese. A Bayesian hierarchical power law process model for multiple repairable systems with an application to supercomputer reliability. *Journal of Quality Technology*, 43:209–223, 2011.