Bayesian Inference for Power Law Processes with Applications in Repairable Systems

Maristela Dias de Oliveira[1], Enrico A. Colosimo[2] and Gustavo L. Gilardoni[3]

Universidade Federal da Bahia, Universidade Federal de Minas Gerais and Universidade de Brasília

Abstract. Statistical models for recurrent events are of great interest in repairable systems reliability and maintenance. The adopted model under minimal repair maintenance is frequently a Nonhomogeneous Poisson Process with the Power Law Process (PLP) intensity function. Although inference for the PLP is generally based on maximum likelihood theory, some advantages of the Bayesian approach have been reported in the literature. In this paper it is proposed that the PLP intensity be reparametrized in terms of (β, η), where β is the elasticity of the mean number of events with respect to time and η is the mean number of events for the period in which the sistem was actually observed. It is shown that β and η are orthogonal and that the likelihood becomes proportional to a product of Gamma densities. Therefore, the family of natural conjugate priors is also a product of Gammas. The idea is extended to the case that several realizations of the same PLP are observed along overlapping periods of time. The results are applied to a real problem concerning the determination of the optimal periodicity of preventive maintenance for a set of power transformers. Some Monte Carlo simulations are provided to study the frequentist behavior of the Bayesian estimates and to compare them with the maximum likelihood estimates.

Keywords. Conjugate prior, optimal maintenance, Poisson Process, posterior distribution, reference priors.

1   Introduction

Statistical models for recurrent events have been investigated in many papers in the literature. Such models are of great interest to study the reliability and maintenance policies for repairable systems (Ascher and Feingold (1984), Bain and Engelhardt (1991), Rigdon and Basu (2000)). Frequently, the adopted model under minimal repair maintenance is a Nonhomogeneous Poisson Process (NHPP), $\{N(t) : t \geq 0\}$, where $N(t)$ is the number of failures from the beginning of the follow-up until time t (Barlow and Hunter, 1960). A flexible parametric form for the intensity function of the NHPP is

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

with mean function

$$\Lambda(t) = EN(t) = \int_0^t \lambda(u)\, du = \left(\frac{t}{\theta} \right)^{\beta},$$

where $\theta > 0$ and $\beta > 0$. This model, known as the Power Law Process (PLP), was proposed by Crow (1974) and since then it has become the most popular parametric intensity in the repairable systems literature. The intensity function is increasing for $\beta > 1$, decreasing for $\beta < 1$ and constant (i.e. the NHPP is actually a Homogeneous Poisson Process) for $\beta = 1$. Adequacy of the PLP for a particular data set can be diagnosed graphically using either Duane plots (Duane, 1964) or some modified Total Time on Tests plots (Klefsjö and Kumar, 1992). More formal hypotheses tests are considered by Baker (1996) and Bhattacharjee *et al.* (2004).

Statistical inference for the PLP is generally based on the maximum likelihood estimator (MLE) and its asymptotic properties (Berman and Turner (1992), Zhao and Xie (1996)). However, some papers appeared in the literature showing some advantages of the Bayesian approach for the PLP model (Sen (2002), Guida *et al.* (1989)). The Bayesian approach deals with the uncertainty of the parameters in the model used to describe a recurrent system. A prior distribution is assumed to represent the uncertainty in the model parameters before the current data is observed. Reference prior distributions have been used in the Bayesian context by Guida *et al.* (1989), Sen (2002) and Yu *et al.* (2006), among others. On the other hand, identifying a family of conjugate prior distributions will often result in mathematical and computational simplifications. Moreover, a conjugate prior distribution can be interpreted as additional data, hence making prior elicitation easier (Raiffa and Schlaifer, 1961; Gelman *et al.*, 2003). Huang and Bier (1998), Huang (2001) and Kim *et al.* (2008) have proposed a conjugate prior distribution for the parameters of the PLP looking at the functional form of the likelihood function. More precisely, suppose that the process is observed in the fixed time interval $(0, T)$ yielding n events at moments $t_1 < \cdots < t_n$. The likelihood function is

$$L(\beta, \theta) = [\prod_{i=1}^n \lambda(t_i)]\, e^{-\Lambda(T)} = \beta^n \theta^{-n\beta}\, [\prod_{i=1}^n t_i]^{\beta-1}\, \exp\{-(T/\theta)^\beta\} \tag{2}$$

(Berman and Turner, 1992; Rigdon and Basu, 2000). Huang and Bier (1998) parametrize the intensity (1) in terms of β and $\lambda_0 = \theta^{-\beta}$ and, by analogy with the resulting likelihood, find the four parameters (i.e. m, c, α and y below) conjugate family

$$\pi(\beta, \lambda_0) \propto \lambda_0^{m-1} \beta^{m-1} [y^m e^{-\alpha}]^{\beta-1} \exp\{-\lambda_0 c y^\beta\}. \tag{3}$$

They go on by finding the normalizing constant and some moments and discussing properties of the resulting posterior. However, the whole approach becomes somewhat difficult because the parameter λ_0 lacks an operational interpretation and the distributions in the family (3) do not belong to any known class. Motivated by this, and also partly by results obtained by Sen and Khattree (1998) and Sen (2002) for the posterior analysis under noninformative priors of the form $(\theta \beta^\delta)^{-1}$, we propose here to parametrize the problem in terms of β and $\eta = \Lambda(T) = (T/\theta)^\beta$. On one side, β and η have simple operational definitions which will often make prior elicitation easier. On the other side, in the (β, η) parametrization the likelihood (2) becomes

$$L(\beta, \eta) = c \, [\beta^n e^{-n\beta/\hat{\beta}}] \, [\eta^n e^{-\eta}] \propto \gamma(\beta|\, n+1, n/\hat{\beta}) \, \gamma(\eta|\, n+1, 1), \tag{4}$$

where $c = \prod_{j=1}^n t_j^{-1}$, $\hat{\beta} = n/\sum_{j=1}^n \log(T/t_j)$ is the MLE of β and $\gamma(x|\, a, b) = b^a x^{a-1} e^{-bx}/\Gamma(a)$ $(x, a, b > 0)$ is the density of the Gamma distribution with shape and scale parameters equal to a and b, respectively. It follows then that β and η are orthogonal and the natural conjugate family has densities of the form

$$\pi(\beta, \eta) = \gamma(\beta|\, a_\beta, b_\beta) \times \gamma(\eta|\, a_\eta, b_\eta), \tag{5}$$

where the prior parameters a_β, b_β, a_η and b_η must all be positive if we want $\pi(\beta, \eta)$ to be proper, although non positive values can also be entertained as long as the posterior becomes proper. The posterior density is

$$\pi(\beta, \eta|\, t_1, \ldots, t_n, T) \propto L(\beta, \eta) \, \pi(\beta, \eta) \propto \gamma(\beta|\, a_\beta + n, b_\beta + n/\hat{\beta}) \times \gamma(\eta|\, a_\eta + n, b_\eta + 1), \tag{6}$$

so that both a priori and a posteriori β and η are independent, each following a Gamma distribution.

The parametrization (β, η) suggests rather easily how to treat the case when several realizations of the PLP are observed along overlapping time intervals. Although this case appears frequently in practice, because repairable systems are usually observed in different time intervals (truncation times), methodological developments have been somewhat lacking in the literature, especially in the Bayesian setting. More precisely, suppose that K realizations of the same PLP have been observed and let t_{ij} denote the j-th event time for the i-th realization ($j = 1, \ldots, n_i$ and $i = 1, \ldots, K$). Let T_i be the truncation time corresponding to the $i - th$ realization. Then we show in Section 3 that the parameters β and $\eta = \sum_{i=1}^K \Lambda(T_i) = \sum_{i=1}^K (T_i/\theta)^\beta$ are orthogonal and that, under the prior specification (5), the posterior has the same form (6) but with an additional factor which does not depend

2

on η and is proportional to $\exp\{KL[(\frac{T_1^{\hat{\beta}}}{\sum_{h=1}^{K} T_h^{\hat{\beta}}}, \cdots, \frac{T_K^{\hat{\beta}}}{\sum_{h=1}^{K} T_h^{\hat{\beta}}})||(\frac{T_1^{\beta}}{\sum_{h=1}^{K} T_h^{\beta}}, \cdots, \frac{T_K^{\beta}}{\sum_{h=1}^{K} T_h^{\beta}})]\}$, where $KL[\cdot||\cdot]$ is the Kullback-Leibler divergence. Hence, the form of the posterior lends itself to an easy i.i.d. simulation using for instance the rejection sampling algorithm.

Our interest in the case of many overlapping realizations stems mainly from a real application concerning the estimation of the optimal maintenance time for a set of power transformers, which we discuss in Section 5. In short, consider a repairable system modeled by a NHPP with an increasing intensity function subject to two types of repairs: either a *minimal repair* after a failure which restores the system (i.e. the intensity) to exactly the same level it was immediately before the failure or a *preventive maintenance* which restores the system to "as good as new" condition. If the preventive maintenances are performed every τ units of time, the expected cost per unit of time is

$$H(\tau) = [C_{PM} + C_{MR}EN(\tau)]/\tau = [C_{PM} + C_{MR}\Lambda(\tau)]/\tau , \tag{7}$$

where C_{MR} and C_{PM} are the expected costs associated to the two types of repair actions. It can be shown (Barlow and Hunter, 1960; Gilardoni and Colosimo, 2007) that the periodicity τ which minimizes $H(\tau)$ satisfies that $\tau\lambda(\tau) - \Lambda(\tau) = C_{PM}/C_{MR}$. In the special case of the PLP, τ becomes

$$\tau = \theta \left[\frac{C_{PM}}{(\beta - 1)C_{MR}} \right]^{1/\beta} . \tag{8}$$

However, inference about τ only makes sense when $\beta > 1$, leading to the necessity of truncating the prior density for β. This can be done preserving conjugacy by truncating the prior (5) to the set $\beta > 1$, because then the posterior density would be the same as (6) but also truncated for $\beta > 1$. However, because of the term $(\beta - 1)^{1/\beta}$ in the denominator of τ and the fact that the posterior density is non null near $\beta = 1$, the posterior expectation of τ will be infinite. Still, under the truncated prior, one can use for instance a *maximum a posteriori estimate* for the optimal time. An alternative, non-conjugate formulation, which puts less weight to values of β close to one and hence will make the posterior expectation of τ finite, is to consider a priori that $(\beta - 1)$ follows a Gamma distribution.

Besides Sections 3 and 5, which deal respectively with the many realizations setting and with the inference for the optimal periodicity for the power transformers data set, the rest of the paper is organized as follows. In Section 2 we make some additional considerations regarding inference for a single realization of the PLP. It also includes a discussion of reference and informative priors and some computational aspects when the interest is centered in a function of the parameters whose posterior expectation cannot be computed explicitly. Section 4 shows some Monte Carlo simulations that help to understand the frequentist behavior of the Bayes estimates under different prior specifications and to compare them to the MLE estimates in the case of several realizations. The simulation scenarios and prior distributions are motivated from the real case discussed in Section 5. Finally, some concluding remarks end the paper in Section 6.

2   A single PLP realization

As before, suppose that the process is observed in $(0, T)$, and let $\ell(\beta, \eta) = \log L(\beta, \eta) = \log c + n \log \beta - n\beta/\hat{\beta} + n \log \eta - \eta$ be the log-likelihood in the (β, η) parametrization. Since $\nabla \ell = (\partial \ell/\partial \beta, \partial \ell/\partial \eta)' = (n/\beta - n/\hat{\beta}, n/\eta - 1)'$, the maximum likelihood estimates (MLE) of β and η are $\hat{\beta}$ and n respectively. Hence, the MLE of $\theta = T \eta^{-1/\beta}$ is $\hat{\theta} = T \hat{\eta}^{-1/\hat{\beta}} = T n^{-1/\hat{\beta}}$. From the Fisher information matrix

$$I(\beta, \eta) = - \begin{pmatrix} E \frac{\partial^2 \ell}{\partial \beta^2} & E \frac{\partial^2 \ell}{\partial \beta \, \partial \eta} \\ E \frac{\partial^2 \ell}{\partial \beta \, \partial \eta} & E \frac{\partial^2 \ell}{\partial \eta^2} \end{pmatrix} = n \begin{pmatrix} \frac{1}{\beta^2} & 0 \\ 0 & \frac{1}{\eta^2} \end{pmatrix}, \tag{9}$$

it follows that the asymptotic covariance matrix of $(\hat{\beta}, \hat{\eta})$ is $\mathrm{Var}(\hat{\beta}, \hat{\eta}) \approx \mathrm{n}^{-1} \mathrm{Diag}(\beta^2, \eta^2)$.

2.1   Posterior Analysis

Let $a_\beta > -n$, $b_\beta > -n/\hat{\beta}$, $a_\eta > -n$ and $b_\eta > -1$ in (5) so that the posterior density (6) is proper. Suppose that the interest is centered in a function $\phi(\beta, \eta)$ such as $\theta = T/\eta^{1/\beta}$ or, perhaps, as in Sen (2002), the current intensity $\lambda(T) = \beta T^{\beta-1}/\theta^\beta = \beta \eta/T$. Under squared error loss, the Bayes estimate of ϕ is $E[\phi(\beta, \eta)|t_1, \ldots, t_n]$. For instance, the posterior expectation of the current intensity is

$$E[\lambda(T)|t_1, \ldots, t_n] = E[\frac{\beta\eta}{T}|t_1, \ldots, t_n] = \frac{1}{T} E[\beta|t_1, \ldots, t_n] E[\eta|t_1, \ldots, t_n] = \frac{1}{T} \frac{a_\beta + n}{b_\beta + n/\hat{\beta}} \frac{a_\eta + n}{b_\eta + 1}.$$

Credible intervals can be obtained from the posterior quantiles of ϕ. An alternative that we consider in Section 5 is to use *Maximum a Posteriori* estimates. In this case the mode of the posterior density (6) is attained for $\tilde{\beta} = (a_\beta + n - 1)/(b_\beta + n/\hat{\beta})$ and $\tilde{\eta} = (a_\eta + n - 1)/(b_\eta + 1)$. Hence, an alternative estimate for $\lambda(T) = \beta \eta/T$ is

$$\tilde{\lambda}(T) = \frac{1}{T} \tilde{\beta} \tilde{\eta} = \frac{1}{T} \frac{a_\beta + n - 1}{b_\beta + n/\hat{\beta}} \frac{a_\eta + n - 1}{b_\eta + 1}.$$

When integration of moments or quantiles of ϕ with respect to the posterior distribution (6) is difficult, one can easily generate Monte Carlo samples $(\beta_1, \eta_1), \ldots, (\beta_m, \eta_m)$ from the posterior and approximate, for instance, $E[\phi(\beta, \eta)|t_1, \ldots, t_n]$ by $m^{-1} \sum_{h=1}^m \phi(\beta_h, \eta_h)$.

2.2   Prior Elicitation

It follows from (9) that the noninformative Jeffrey's prior is

$$\pi(\beta, \eta) \propto [\det \mathrm{I}(\beta, \eta)]^{\frac{1}{2}} \propto (\beta\eta)^{-1}. \tag{10}$$

In the original (β, θ) parametrization this is equivalent to $\pi(\beta, \theta) \propto \theta^{-1}$ (see (11) below).

The improper reference priors $\pi(\beta, \theta) \propto (\theta \beta^\delta)^{-1}$ ($\delta < n$), considered by Bar-Lev *et al.* (1992) and Sen (2002), which generalize the noninformative priors $\pi(\beta, \theta) \propto \theta^{-1}$ and $\pi(\beta, \theta) \propto (\theta \beta)^{-1}$ (Lingham and Sivaganesan, 1997; Guida *et al.*, 1989; Box and Tiao, 1973), are special cases of (5) when $a_\eta = b_\eta = b_\beta = 0$ and $a_\beta = -\delta$. To see this, note that

$$
\begin{aligned}
\pi(\beta, \eta) &= \pi(\beta, \theta)|_{\theta = T/\eta^{1/\beta}} \times |J| \\
&\propto (\theta \beta^\delta)^{-1}\big|_{\theta = T/\eta^{1/\beta}} \times T \beta^{-1} \eta^{-1-1/\beta} \propto \beta^{-\delta-1} \eta^{-1},
\end{aligned} \tag{11}
$$

where $J = -T^\beta \beta^{-1} \eta^{-1-1/\beta}$ is the Jacobian of the transformation $(\beta, \theta) \mapsto (\beta, \eta)$.

To finish this section we note that the elicitation of proper informative priors in the (β, η) parametrization may be facilitated in view that both β and η have clear operational interpretations. In this sense, since

$$
\frac{d\Lambda(t)/\Lambda(t)}{dt/t} = t \frac{\Lambda'(t)}{\Lambda(t)} = t \frac{\lambda(t)}{\Lambda(t)} = t \frac{(\beta/\theta)(t/\theta)^{\beta-1}}{(t/\theta)^\beta} = \beta,
$$

β is the elasticity of the mean number of events $\Lambda(t)$ with respect to time, i.e. the relative change in Λ due to relative change in t. Indeed, the PLP is characterized by the fact that this elasticity is constant over time. On the other hand, $\eta = (T/\theta)^\beta = \Lambda(T) = EN(T)$ is the expected number of events during the period that the process has been observed.

3   Several overlapping realizations

The methods established in Section 2 can be easily extended to the case that K independent realizations of the same PLP, say $N_1(t), \ldots, N_K(t)$, are observed all up to the same time T. This follows from the well known fact that the superposition of NHPPs is also a NHPP whose intensity function is the sum of the individual intensities (Thompson, 1998). In other words, $N_+(t) = \sum_{i=1}^K N_i(t)$ has intensity $\lambda_+(t) = K\lambda(t) = K\beta t^{\beta-1}/\theta^\beta$ and hence is also a PLP with parameters $\beta_+ = \beta$ and $\theta_+ = \theta/K^{1/\beta}$. Therefore, one can use the ideas in Section 2 to draw inferences about β_+ and θ_+ and these are equivalent to inferences about the original parameters $\beta = \beta_+$ and $\theta = \theta_+ K^{1/\beta}$. However, it is not clear how to proceed when the K realizations have been observed along different time intervals.

3.1   Overlapping realizations of a PLP

Suppose that $N_1(t), \ldots, N_K(t)$ are independent realizations of the same PLP observed respectively up to times $T_1, \ldots, T_K$. Let t_{ij} be the j-th event time for the i-th realization, $i = 1, \cdots, K$; $j = 1, \cdots, n_i$. According to equation (2), the likelihood in the original (β, θ) parametrization is

$$
L(\beta, \theta) = \prod_{i=1}^K \left\{ e^{-(T_i/\theta)^\beta} \frac{\beta^{n_i}}{\theta^{n_i \beta}} \prod_{j=1}^{n_i} t_{ij}^{\beta-1} \right\} = \frac{\beta^n}{\theta^{n\beta}} [\prod_{i=1}^K \prod_{j=1}^{n_i} t_{ij}]^{\beta-1} \exp\{-\sum_{i=1}^K (T_i/\theta)^\beta\}, \tag{12}
$$

where $n = \sum_{i=1}^{K} n_i$ is the total number of events. If for some of the realizations no event has been observed, take the corresponding $n_i = 0$ and set in equation (12) empty sums and products equal to 0 and 1, respectively. Hence, the MLE satisfies that $\hat{\theta} = [\sum_{i=1}^{K} T_i^{\hat{\beta}}/n]^{1/\hat{\beta}}$ and

$$\frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \log t_{ij} = \frac{\sum_{i=1}^{K} T_i^{\hat{\beta}} \log T_i}{\sum_{i=1}^{K} T_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}}, \tag{13}$$

and must be obtained numerically (Rigdon and Basu, 2000).

If we reparametrize the problem in terms of β and $\eta = \sum_{i=1}^{K}(T_i/\theta)^\beta$, it follows after some algebra that the likelihood (12) becomes

$$L(\beta, \eta) = c \times \eta^n e^{-\eta} \times \beta^n e^{-n\beta/\hat{\beta}} \times e^{nF(\beta)} \propto \gamma(\eta| \, n+1, 1)\, \gamma(\beta| \, n+1, n/\hat{\beta})\, e^{nF(\beta)} \tag{14}$$

where now $c = \prod_{i=1}^{K} \prod_{j=1}^{n_i} t_{ij}^{-1}$, $\hat{\beta}$ satisfies (13) and

$$F(\beta) = \frac{\sum_{i=1}^{K} T_i^{\hat{\beta}} \log T_i}{\sum_{i=1}^{K} T_i^{\hat{\beta}}} \beta - \log \sum_{i=1}^{K} T_i^{\beta} \, .$$

Note that β and η are still orthogonal. The log-likelihood is $\ell(\beta, \eta) = \log c + n \log \eta - \eta + n \log \beta - n\beta/\hat{\beta} + nF(\beta)$. Therefore, the MLE are obtained solving $\partial\ell/\partial\beta = n/\beta - n/\hat{\beta} + nF'(\beta) = 0$ and $\partial\ell/\partial\eta = n/\eta - 1 = 0$, which gives $\hat{\eta} = n$ and, of course, $\hat{\beta}$ given by (13). In order to compute asymptotic variances note that $\frac{\partial^2\ell}{\partial\beta^2} = -n/\beta^2 + nF''(\beta)$, $\frac{\partial^2\ell}{\partial\eta^2} = -n/\eta^2$ and $\frac{\partial^2\ell}{\partial\beta\,\partial\eta} = 0$. Hence, the Fisher information matrix is $I(\beta, \eta) = n \, \mathrm{Diag}(\beta^{-2} - F''(\beta), \eta^{-2})$, where

$$F''(\beta) = -\sum_{i=1}^{K} \frac{T_i^{\beta}}{\sum_{h=1}^{K} T_h^{\beta}} [\log T_i]^2 + \left(\sum_{i=1}^{K} \frac{T_i^{\beta}}{\sum_{h=1}^{K} T_h^{\beta}} \log T_i\right)^2$$

is formally the same as minus the variance of a random variable taking values $\log T_i$ with probabilities proportional to T_i^{β} ($i = 1, \ldots, K$). The asymptotic covariance matrix of $(\hat{\beta}, \hat{\eta})$ is then $[I(\beta, \eta)]^{-1} = n^{-1} \mathrm{Diag}([1/\beta^2 - F''(\beta)]^{-1}, \eta^2)$. Asymptotic variances for functions of the parameters can be obtained using the Delta Method.

3.2 Posterior analysis

Under the prior specification (5), the posterior density becomes

$$\pi(\beta, \eta|D) \propto \gamma(\eta| \, a_\eta + n, b_\eta + 1) \times \gamma(\beta| \, a_\beta + n, b_\beta + n/\hat{\beta}) \times e^{nF(\beta)}, \tag{15}$$

where $D = \{t_{ij} : i = 1, \ldots, K; j = 1, \ldots, n_i\}$. It should be immediate from the comparison of (15) with (6) that the behavior of $F(\beta)$ is crucial to understand the difference between the one and the many realizations settings. Now, if $\hat{\beta}$ is the solution of (13), it follows that

$$F(\hat{\beta}) - F(\beta)$$

$$= \frac{\sum_{i=1}^{K} T_i^{\hat{\beta}} \log T_i}{\sum_{i=1}^{K} T_i^{\hat{\beta}}} \hat{\beta} - \log \sum_{i=1}^{K} T_i^{\hat{\beta}} - \frac{\sum_{i=1}^{K} T_i^{\hat{\beta}} \log T_i}{\sum_{i=1}^{K} T_i^{\hat{\beta}}} \beta + \log \sum_{i=1}^{K} T_i^{\beta}$$

$$= \frac{\sum_{i=1}^{K} T_i^{\hat{\beta}} \log(T_i^{\hat{\beta}}/T_i^{\beta})}{\sum_{i=1}^{K} T_i^{\hat{\beta}}} \beta + \log \frac{\sum_{i=1}^{K} T_i^{\beta}}{\sum_{i=1}^{K} T_i^{\hat{\beta}}}$$

$$= \sum_{i=1}^{K} \frac{T_i^{\hat{\beta}}}{\sum_{h=1}^{K} T_h^{\hat{\beta}}} \log \frac{T_i^{\hat{\beta}}/\sum_{h=1}^{K} T_h^{\hat{\beta}}}{T_i^{\beta}/\sum_{h=1}^{K} T_h^{\beta}}$$

$$= KL[(\frac{T_1^{\hat{\beta}}}{\sum_{h=1}^{K} T_h^{\hat{\beta}}}, \cdots, \frac{T_K^{\hat{\beta}}}{\sum_{h=1}^{K} T_h^{\hat{\beta}}}) || (\frac{T_1^{\beta}}{\sum_{h=1}^{K} T_h^{\beta}}, \cdots, \frac{T_K^{\beta}}{\sum_{h=1}^{K} T_h^{\beta}})] \geq 0 \,,$$

where $KL[(p_1, \ldots, p_K) || (q_1, \ldots, q_K)) = \sum_{i=1}^{K} p_i \log \frac{p_i}{q_i}$ is the Kullback-Leibler divergence. Hence, $F(\beta)$ attains a maximum when $\beta = \hat{\beta}$. Moreover, $F(\beta)$ is constant and equal to $F(\hat{\beta})$ if and only if $T_1 = T_2 = \cdots = T_K$.

In order to sample from the posterior distribution (15) we use the independence between η and β and obtain first $\eta_1, \ldots, \eta_m \overset{i.i.d.}{\sim} \text{Gamma}(a_\eta + n, b_\eta + 1)$. Simulation from the posterior distribution of β becomes easy by using, for instance, the rejection or importance sampling algorithms (see Gelman *et al.* (2003) or Devroye (1986)). For instance, the rejection algorithm produces an observation from the posterior of β by sampling repeatedly $\beta \sim \gamma(\beta | a_\beta + n, b_\beta + n/\hat{\beta})$ and $u \sim \text{Uniform}(0,1)$ until $u \leq \exp\{n[F(\beta) - F(\hat{\beta})]\}$. Repeating the rejection algorithm m times we obtain an i.i.d. sample $\beta_1, \ldots, \beta_m$. Once that an i.i.d. sample from the posterior $\pi(\beta, \eta | D)$ has been obtained we proceed essentially as in Section 2. In our practice the rejection sampling method has been quite efficient, in the sense that even for problems with few failures the rejection rate is below 10%.

The rejection algorithm can also be used when the prior for β is a Gamma distribution truncated to the right of $\beta = 1$. In this case, one just changes the proposal distribution above to be also a truncated Gamma. In other words, to obtain an observation from $\pi(\beta | D)$ one samples repeatedly $\beta \sim \gamma(\beta | a_\beta + n, b_\beta + n/\hat{\beta})$ and $u \sim \text{Uniform}(0,1)$ until both $\beta > 1$ and $u \leq \exp\{n[F(\beta) - F(\hat{\beta})]\}$. However, the rejection algorithm would need some major adaptation if one wants to consider a prior distribution for β which is restricted to have support in $(1, \infty)$ and which is not a truncated Gamma. For instance, we argue in Sections 4 and 5 that for the power transformers problem it may be better to consider a shifted Gamma prior, i.e. $\beta - 1 \sim \text{Gamma}(a_\beta, b_\beta)$. In this case one could use the Metropolis algorithm to obtain an approximate sample from the posterior of β. Briefly, we set a starting value β_0 (e.g. $\beta_0 = \hat{\beta}$) and proceed iteratively as follows. At step $(i+1)$ we generate $z \sim \text{Normal}(0,1)$ and $u \sim \text{Uniform}(0,1)$ and let $\beta_{cand} = \beta_i + Z$. Now if $u < \min\{\pi(\beta_{cand} | D)/\pi(\beta_i | D), 1\}$ we let $\beta_{i+1} = \beta_{cand}$, otherwise let $\beta_{i+1} = \beta_i$. In general the Metropolis algorithm produce correlated observations which may be unduly influenced by the starting value. If one wants to avoid

this, the algorithm can be run for $M = B + ml$ cycles, the first B observations discarded (the "burn-in") and then every other l of the remaining simulated values can be kept to end up with a size m approximate i.i.d. sample.

4   Monte Carlo Simulation

In this section we describe some Monte Carlo simulations in order to compare Bayes estimates under different prior specifications in the case of overlapping realizations of a PLP. The Bayes estimates were also compared to the ones obtained by maximum likelihood. As described in Section 1, the optimal preventive maintenance policy that minimizes expected cost per unit of time is the value τ defined by (8). τ was the quantity of interest in the simulations. The prior (and hence also the posterior) distribution must satisfy $Pr(\beta > 1) = 1$ since the intensity function of failures must be increasing as discussed in Section 1. This information has been incorporated in all of the prior specifications in simulations.

Different prior distributions for β and η were used in the simulations. The following notations and definitions were used in the simulation runs:

- MLE - Maximum likelihood estimate;

- BayesE$_1$ - Bayes estimator by considering a reference prior distribution (11) for $\delta = 1$ truncated at $\beta = 1$;

- BayesE$_2$ - Bayes estimator by considering Jeffrey's prior distribution (10) truncated at $\beta = 1$;

- BayesE$_3$ - Bayes estimator by considering gamma prior distributions truncated at $\beta = 1$. That is $\pi(\beta, \eta) \propto \gamma(\beta \,|\, a_\beta, b_\beta) \times \gamma(\eta \,|\, a_\eta, b_\eta)\, A_{(1,\infty)}(\beta)$, where $A_{(1,\infty)}(\beta) = 1,$ if $\beta \in (1, \infty)$ and 0, otherwise.

- BayesE$_4$ - Bayes estimator by considering a gamma prior distribution shifted to 1 for β and gamma for η. That means, $\beta - 1 \sim \gamma(a_\beta, b_\beta)$.;

- CP - Interval Coverage Percentage.

In the likelihood approach, asymptotic confidence intervals for τ were obtained by using the delta method (Gilardoni and Colosimo, 2007). In the Bayesian approach, we used the highest posterior density (HPD) intervals. The Bayesian estimates were the posterior distribution mode. That is, the value that maximize the posterior distribution of τ.

Throughout the Monte Carlo study we consider $\beta = 2$, $\theta = 24$ and $C_{MR}/C_{PM} = 16$, so that it follows from (8) that $\tau = 6$. The number of systems K and truncation times T_i's were set to study three different situations. The first two achieve a large number of failures by considering respectively many systems and large truncation times. More precisely, we have in situation 1 $K = 500$ systems all truncated at $T = 100$, so that the expected number of failures per system equals 17.361. Situation 2 considers $K = 50$ systems all truncated

at $T = 320$, resulting in 178 expected failures per system. Finally, the third situation has $K = 50$ systems truncated at T =30 and hence only 1.5625 expected failures per system, so that this situation is probably closer to the real example considered in the next section.

The results of the Monte Carlo simulations based on 3000 replicas are shown in Table 1. In the first two situations there were no significant differences among methods and prior distributions. Probably, sample sizes were large enough to overcome differences in the prior specifications. In the third situation, Bayes estimates have similar results. BayesE$_4$ has the worst interval covarage. In general, all estimates have a small bias, the MLE being the least biased.

Table 1: Summary of simulation results

		MLE	BayesE$_1$	BayesE$_2$	BayesE$_3$	BayesE$_4$
	Mean of $\hat{\tau}$	6.00	6.00	6.00	6.00	6.02
Situation 1	CP	94.9	94.3	94.4	94.4	95.0
	Mean length	0.475	0.475	0.474	0.474	0.482
	Mean of $\hat{\tau}$	6.00	6.00	6.00	6.00	6.03
Situation 2	CP	94.7	94.5	94.6	94.6	95.0
	Mean length	0.753	0.753	0.752	0.753	0.765
	Mean of $\hat{\tau}$	6.11	6.17	6.12	6.13	6.16
Situation 3	CP	95.4	95.2	95.9	95.1	93.1
	Mean length	2.125	2.149	2.125	2.124	1.971

5 Example: Maintenance of electrical power transformers

The data in Figure 1(a) shows the failure history of 40 electrical power transformers (Gilardoni and Colosimo, 2007). The usual nonparametric estimate of Λ (Meeker and Escobar, 1998) is shown in Figure 1(b). The convex form of this plot provides some evidence that the intensity function is increasing.

The same prior distributions used in Section 4 were considered here. According to the electrical power company, the ratio between minimal repair and preventive maintenance costs is $C_{MR}/C_{PM} = 15$. Table 4 presents the results. The interval based on the ML estimates is the shortest one. Point estimates are in agreement among Bayesian methods taking a value around 6400 hours, although the ML estimate is 6290 hours. Among the Bayesian intervals, those considering the Jeffrey's and the translated gamma prior are shorter. Posterior density function of τ appears to be slightly skewed to the right (see Figure 2 for the Jeffrey's prior case).

In addition to point and interval estimates for the optimal maintenance time τ, it is useful to gain an idea of the size of the difference between estimated and true expected costs
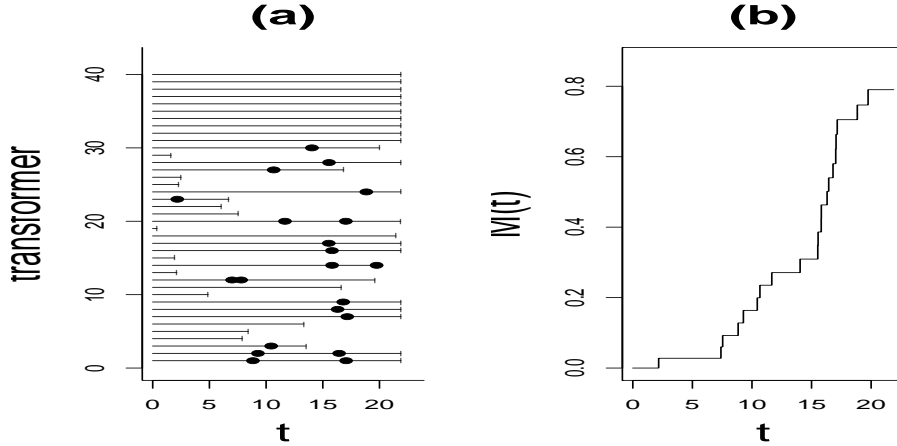
Figure 1: (a) Power transformer data. Each horizontal line represents a transformer and dots are observed failure times; (b) Mean Cumulative Failure (MCF) type estimate for Λ.

Table 2: τ estimates for the power transformers data (in 1000 hours).

	MLE	BayesE$_1$	BayesE$_2$	BayesE$_3$	BayesE$_4$
Estimate	6.29	6.44	6.41	6.41	6.56
Interval	[4.87;7.70]	[5.06;8.48]	[5.00;8.72]	[5.02;8.74]	[5.04;8.44]
Length	2.83	3.42	3.72	3.72	3.40

(Gilardoni and Colosimo, 2007). This difference can be obtained from (7) as

$$H(\hat{\tau}) - H(\tau) = \frac{1}{\hat{\tau}}\left[C_{PM} + C_{MR}\left(\frac{\hat{\tau}}{\theta}\right)^{\beta}\right] - \frac{1}{\tau}\left[C_{PM} + C_{MR}\left(\frac{\tau}{\theta}\right)^{\beta}\right]$$

$$= C_{PM}\left[\frac{1}{\hat{\tau}}\left(1 + \frac{C_{MR}}{C_{PM}}\frac{\hat{\tau}^{\beta}}{\sum_{i=1}^{K}T_i^{\beta}}\eta\right) - \frac{1}{\tau}\left(1 + \frac{C_{MR}}{C_{PM}}\frac{\tau^{\beta}}{\sum_{i=1}^{K}T_i^{\beta}}\eta\right)\right].$$

$H(\hat{\tau}) - H(\tau)$ measures the difference in the cost attained for the present state of information and that which could be attained if one had perfect information (i.e. if sampling could be continued forever). Observe that because of the definition of the optimal τ one must have that $H(\hat{\tau}) - H(\tau) \geq 0$. Hence, we usually compute a credible upper limit for the difference.

To compute a rao-blackwellized (**?**) approximation to the posterior density of the optimal τ, note that

$$\pi(\tau|D) = \int \pi(\tau|\beta, D)\pi(\beta|D)\, d\beta \approx \frac{1}{m}\sum_{j=1}^{m}\pi(\tau|\beta_j, D)\,,$$
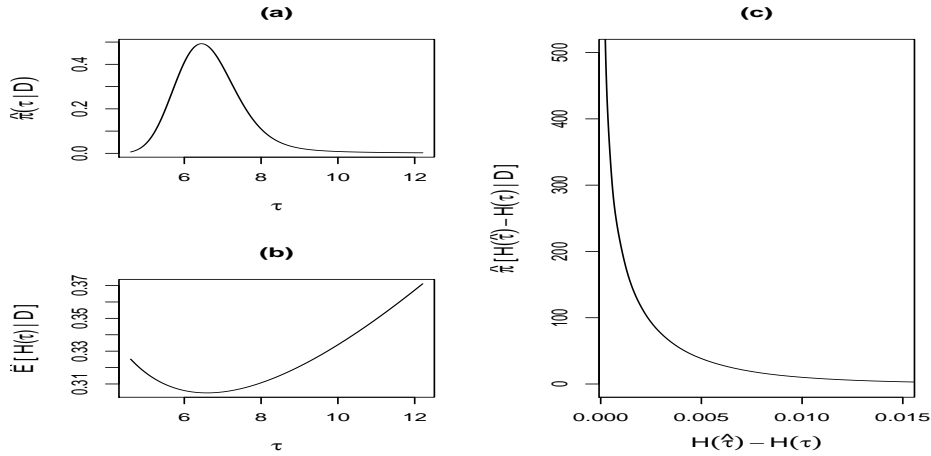
Figure 2: Approximate (a) posterior density of τ, (b) cost per unit of time and (c) posterior density of the difference between costs, all for the Jeffrey's prior.

where we use that $\psi = \frac{C_{MR}}{C_{PM}} \frac{(\beta-1)\tau^\beta}{\sum_{i=1}^K T_i^\beta}$ to obtain that

$$
\pi(\tau|\beta, D) = \pi(\psi|\beta, D) \left| \frac{\partial \psi}{\partial \tau} \right| = \gamma \left(\frac{C_{MR}}{C_{PM}} \frac{(\beta-1)\tau^\beta}{\sum_{i=1}^K T_i^\beta} \middle| a_\psi + n, b_\psi + 1 \right) \frac{C_{MR}}{C_{PM}} \frac{\beta(\beta-1)\tau^{\beta-1}}{\sum_{i=1}^K T_i^\beta} .
$$

Figure 2(c) presents the posterior density of the difference between costs.

6   Final Remarks

In this paper a conjugate prior distribution was derived for the PLP model in the one system case. The proposed conjugate prior is a product of gamma distributions for the parameters of the PLP in an alternative parametrization. The results are extended for overlapping realizations of the same PLP. Although in the many realization case the product of gamma prior is no longer conjugate, it was showed that posterior sampling is easy to implement.

Monte Carlo simulations are used in order to compare some proposed prior distributions in the context of a real application. Three different situations and four prior distributions are considered in the simulations. It can be observed no significant differences among prior distributions in the considered scenarios. They are also very close to the MLE results. In the real case analysis in Section 5, point estimates are similar among the methods. Bayesian intervals considering Jeffreys and gamma prior distributions are shorter than the one using prior of reference. Maximum likelihood confidence interval is the shortest one.

We considered just the time truncation situation in this paper. That is T is fixed by design. This is basically the way that most of the practical situations collect data from repairable systems. Another possible situation is the case in which data collection is ceased after a fixed number of failures. This sampling scheme is said to be failure truncated. Since

the likelihood is essentially the same as for the time truncated case, the Bayesian analysis takes the same form irrespective of the experimental design. However, a cautionary note regarding the transformation $\eta = (T/\theta)^\beta$ or $\eta = \sum_{i=1}^{K}(T_i/\theta)^\beta$ is in order when one or more of the realizations are failure truncated. In this case, one or more of the T_i's are random and η would depend on data. Hence, the prior density (5) depends indirectly on the observed data, which is not allowed from a strict Bayesian viewpoint. In our opinion this fact has little, if any, practical importance. Moreover, we stress that the problem appears only in the case of failure truncation and even then, disappears if one uses a non-informative prior $\pi(\eta) \propto \eta^{-1}$ (i.e. $a_\eta = b_\eta = 0$), because such a prior would be equivalent to $\pi(\theta) \propto \theta^{-1}$ (see equation (11)), which of course does not depend on data.

References

H. Ascher and H. W. Feingold, *Repairable Systems Reliability: Modeling, Inference, Misconception and their Causes*, Marcel Dekker: New York, 1984.

L. J. Bain and M. Engelhardt, *Statistical Analysis of Reliability and Life-Testing Models, Theory and Methods* Marcel Dekker: New York, 1991.

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

S. K. Bar-Lev, I. Lavi, and B. Reiser, Bayesian Inference for the Power Law Process. *Annals Institute Statistical Mathematics* 44, pp. 623-639, 1992.

R. E. Barlow and L. C. Hunter, Optimum Preventive Maintenance Policies. *Operations Research* 8, pp. 90-100, 1960.

M. Berman and T. R. Turner, Approximating Point Process Likelihoods with GLIM. *Applied Statistics* 41, pp. 31-38, 1992.

M. Bhattacharjee, J. V. Deshpande and U. V. Naik-Nimbalkar, Unconditional Tests of Goodness of Fit for the Intensity of Time-Truncated Nonhomogeneous Poisson Processes. *Technometrics* 46, pp. 330-338, 2004.

G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley: Reading, 1973.

L. R. Crow, Reliability Analysis for Complex Systems. *Reliability and Biometry*, F. Proschan and J. Serfling (Eds), pp. 379-410, 1974.

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

J. T. Duane, Learning Curve approach to Reliability Monitoring. *IEEE Transactions on Aerospace* 2, pp. 563-566, 1964.

A. E. Gelfand and A. F. M. Smith, Sampling-based Approaches to Calculating Marginal Densities. *Journal of the American Stasticial Association* 85, pp. 398-409, 1990.

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

I. B. Gerstack, *Models of Preventive Maintenance.* North Holand: Amsterdam, 1977.

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

M. Guida, R. Calabria, and G. Pulcini, Bayes Inference for a Non-Homogeneous Poisson Process with Power Intensity Law. *IEEE Transactions on Reliability* 38, pp. 603-609, 1989.

Y. S. Huang, A Decision Model for Deteriorating Repairable Systems. *IIE Transactions* 33, pp. 479-485, 2001.

Y. S. Huang and V. B. Bier, A Natural Conjugate Prior for the Non-Homogeneous Poisson Process with a Power Law Intensity Function. *Communications in Statistics - Simulation and Computation* 27, pp. 525-551, 1998.

T. S. Kim, C. S. Park and S. E. Ahn, Determinig the Optimal Maintenance Action for a Deteriorating Repairable System. *Probabilistic Engineering Mechanics* 23, pp. 95-101, 2008.

B. Klefsjö and U. Kumar, Goodness of Fit Tests for the Power-Law Process Based on the TTT-Plot. *IEEE Transactions on Reliability* 41, pp. 593-598, 1992.

R. T. Lingham and S. Sivaganesan, Testing Hypotheses About the Power Law Process Under Failure Truncation Using Intrinsic Bayes Factors. *Annals of the Institute of Statistical Mathematics*, 49, pp. 693-710.

W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data.* John Wiley: New York, 1998.

H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory.* Harvard University Press: Cambridge, Massachussets, 1961.

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

A. Sen, Bayesian Estimation and Prediction of the Intensity of the Power Law Process. *Journal of Statistical Computation and Simulation* 72, pp. 613-631, 2002.

A. Sen and R. Khattree, On estimating the current intensity of failure for the power-law process. *Journal of Statistical Planning and Inference*, 74, pp. 253-272, 1998.

W. A. Thompson, *Point Process Models with Applications to Safety and Reliability.* Chapman and Hall: London, 1988.

J. W. Yu, G. L. Tian and M. L. Tang, Predictive Analyses for Non-Homogeneuous Poisson Process with Power Law using Bayesian Approach. *Computational Statistics and Data Analysis* 51, 4254-4268, 2007.

M. Zhao and M. Xie, On Maximum Likelihood Estimation for a General Non-homogeneous Poisson Process. *Scandinavian Journal of Statistics* 23, 597-607, 1996.