Bayesian Estimation and Prediction of the Intensity of the Power Law Process

Ananda Sen

Oakland University, Rochester, MI, USA

Abstract

In analyzing failure data pertaining to a repairable system, perhaps the most widely used parametric model is a nonhomogeneous Poisson process with Weibull intensity, more commonly referred to as the Power Law Process (PLP) model. Investigations relating to inference of parameters of the PLP under a frequentist framework abound in the literature. The focus of this article is to supplement those findings from a Bayesian perspective, which has thus far been explored to a limited extent in this context. Main emphasis is on the inference of the intensity function of the PLP. Both estimation and future prediction are considered under traditional as well as more complex censoring schemes. Modern computational tools such as Markov Chain Monte Carlo are exploited efficiently to facilitate the numerical evaluation process. Results from the Bayesian inference are contrasted with the corresponding findings from a frequentist analysis, both from a qualitative and a quantitative viewpoint. The developed methodology is implemented in analyzing interval-censored failure data of equipments in a fleet of marine vessels.

Key Words: Repairable systems, Power law process, Intensity function, Bayesian inference, Markov chain Monte Carlo, Aggregated data.

1 Introduction

In analyzing failure data from a repairable system, possibly the single-most popular parametric model that has been used in practice is the Nonhomogeneous Poisson Process (NHPP) with Weibull intensity, more commonly referred to as the Power-Law Process (PLP) model. This model, introduced and investigated by Crow (1974), evolves as a stochastic formulation of a certain physical feature observed by Duane (1964) in his study of the failure process of various complex mechanical devices. Mathematically, the intensity function of the PLP assumes the form

$$\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1},\ \theta > 0,\ \beta > 0,\ t > 0.$$

The corresponding cumulative intensity function $\Lambda(t) = \int_0^t \lambda(s)\, ds = (t/\theta)^\beta$ is linear with time 't' on a log-log scale, conforming to Duane's (1964) observation. For this reason, PLP is often termed as the "Duane Model" by reliability engineers and statisticians. Since Crow's (1974) seminal work on PLP, numerous articles dealing with the inferential aspects of the model have appeared in the statistical and engineering literature. Mathematical tractability and well-documented inference procedures have attributed to the model's popularity among the practitioners. The model is quite flexible in that it incorporates both growth ($\beta > 1$)

and decay ($\beta < 1$) in reliability. The case of *no growth* (Homogeneous Poisson Process) corresponds to $\beta = 1$.

The essential ingredient of a NHPP is its intensity function, also referred to as its *rate of occurrence of failure* (ROCOF). For a general point process $N(t)$ denoting the number of failures in the time interval $[0, t]$, the intensity function $\lambda(t)$ is defined as

$$\lambda(t) = \lim_{\Delta t \longrightarrow 0} \frac{P\left[N(t + \Delta t) - N(t) \geq 1\right]}{\Delta t}. \tag{1}$$

The expression in (1) can be interpreted as the *probability of atleast one failure per unit time in an infinitesimal time interval* $(t, t + \Delta t]$. Under the assumption that ties (in failure times) occur only with probability zero, a premise we adopt in the current article,

$$\lambda(t) = \frac{d}{dt} E\left[N(t)\right],$$

$E\left[N(t)\right]$ denoting the mean number of failures in $[0, t]$.

The focus of this article is the inference concerning the intensity function of a special point process, namely, the PLP. Both estimation of *current intensity* and prediction of *future intensity* are considered in detail from a Bayesian viewpoint. The current intensity refers to the achieved value of the intensity function at the termination point of the developmental testing phase. This quantity can be thought of as an initial estimate of the ROCOF at the subsequent operational testing phase in which the system failures are assumed to be governed by a Homogeneous Poisson Process (at least initially) (c.f. Bain and Engelhardt, 1980, Crow, 1982). An objective evaluation of the *current intensity* is extremely crucial to reliability engineers in deciding the extent of effectiveness of a developmental program in achieving a planned reliability goal.

There has been considerable work relating to the estimation of current intensity of a PLP. Some of the earlier work based on the exact and approximate distributions of the maximum likelihood estimators (MLE) include Lee and Lee (1978), Bain and Engelhardt (1980), and Crow (1982). Higgins and Tsokos (1981) consider a quasi-Bayes estimation procedure, while Rigdon and Basu (1988, 1990) investigate the minimum mean-squared error estimation within the class of *scaled* MLE's. As Rigdon and Basu (1990) indicate, however, the scaling optimizer depends on the unknown parameter β. To circumvent this problem, they take $\beta = 1$, which they justify in view of the fact that it serves as the boundary for both the growth and decay scenarios. Sen and Khattree (1998) present a more rigorous development of the estimation problem based on a minimum risk criterion as well as the probabilistic Pitman-closeness criteria under a wide choice of loss functions. Calabria et al. (1988) and Qiao and Tsokos (1998) derive other estimators of the current intensity of a PLP.

Much of the investigation thus far on estimating the current intensity of a PLP has been undertaken under a frequentist framework. By contrast, in this article, we propose a Bayesian approach to this problem. The advantages of the Bayesian approach are manifold. Clearly, the Bayesian framework provides a natural, unified environment for carrying out the estimation and prediction process based on finite-sample calculations rather than the large-sample approximation often needed for the frequentist case. Moreover, it can readily incorporate any strong prior information in the inference process if one such is available.

Most importantly, in the specific case of inference for a PLP, the Bayesian approach alleviates certain problems encountered in the frequentist framework. In the context of observing the failure process of a repairable system, two types of inspection schemes are typically adopted in practice. The schemes, referred to as "time truncation" and "failure truncation", closely resemble "Type I" and "Type II" censoring, respectively, in that the process terminates either at a predetermined *time* or a *number of failures*. The inference for the two sampling schemes (time and failure-truncated) are intrinsically different in the frequentist case. In the time-truncated case, there is no clear-cut solution available for the *current intensity* estimation. The major obstacle lies in the lack of any pivotal quantity for inference purposes. To circumvent this, Bain and Engelhardt (1980) resort to an ad-hoc fiducial argument. For the failure-truncated case on the other hand, existence of suitable pivotals facilitate the inference of current intensity. But, since the *current intensity* is evaluated at the termination point, in the failure-truncated case it is a *random* quantity, and thus the premise of 'estimation' is ill-defined in the frequentist context. Happily, these problems do not surface in the Bayesian paradigm. A unified solution works for either sampling scenario and unlike much of the frequentist work, the Bayesian solution is not restricted to a *class* of estimators (e.g. class of scaled MLE's).

The article is organized as follows. In Section 2, the Bayesian formulation is put forth in a formal manner. Section 3 deals with the posterior analysis of *current intensity* under specific choices of informative as well as noninformative priors that have been widely used in practice. The problem of prediction is undertaken in Section 4. It is demonstrated how a simulation-based approach, coupled with appropriate data augmentation facilitates the analysis. In Section 5, we focus on the extension of the inference methods to interval-censored data arising from successive failures of a repairable system that are assumed to follow a PLP. In Section 6, we implement the developed methodology in analyzing equipment failures in a fleet of marine vessels where the data acquisition process is carried out in periodic intervals. We indicate in Section 7 how the prescribed methodology can be adapted to the inference problem that involves multiple copies of a repairable system.

2 Bayesian Formulation

For simplicity of exposition, we consider the failure process of a single system, under either time or failure-truncated sampling scheme. The methods and techniques we describe can easily be adopted to the situation dealing with multiple copies of a repairable system. Denoting by $0 < t_1 < t_2 < \cdots < t_n$, the n successive observed times to failure of the system, the likelihood function can be expressed as:

$$\begin{aligned} L(\theta, \beta) &= \prod_{i=1}^{n} \lambda(t_i) \exp\big[- \int_0^y \lambda(t)\, dt \big] \\ &= \left(\frac{\beta}{\theta}\right)^n \left[\prod_{i=1}^{n} \left(\frac{t_i}{\theta}\right)^{\beta-1} \right] \exp\left[- \left(\frac{y}{\theta}\right)^{\beta} \right], \end{aligned} \tag{2}$$

where

$$y = \begin{cases} t_n, & \text{if process is terminated at the } n^{\text{th}} \text{ failure.} \\ t_0, & \text{if process is truncated at a fixed time } t_0. \end{cases}$$

In our development, we tacitly assume that $n \geq 1$.

Several alternative prior models for the parameters of a PLP have been entertained in the literature. In this article, we shall consider a set of fully noninformative as well as a fully informative class of priors both of which are deemed quite reasonable in the context under consideration. The class of noninformative priors we consider is defined as

$$\pi(\theta, \beta) \propto (\theta\beta^{\delta})^{-1},\ \theta > 0,\ \beta > 0, \ldots, \tag{3}$$

where δ is a known quantity. Considering the distribution of the time to first failure of a PLP, the cases $\delta = 0$ and 1 in (3) can both be motivated as instances of the Jeffrey's class of priors for location-scale family of distributions (c.f. Bar-Lev et al., 1992). The cases corresponding to $\delta = 0$ and $\delta = 1$ have further been motivated by Lingham and Sivaganesan (1997) as reference priors for various hypothesis testing scenarios involving the growth parameter β. While we do not restrict δ solely to those two numbers, it turns out (to be seen in the next section) that we do need δ to be strictly less than the sample size n for the propriety of the posterior of $\lambda(y)$.

The class of informative priors we consider in this article have been introduced by Guida et al. (1989). They interpret the mean number of failures up to a predetermined time T, namely, $\Lambda(T) = \int_0^T \lambda(t)\, dt = (T/\theta)^{\beta}$ to be a parameter of interest, the prior information on which may be easily elicitated. A reasonable choice for the probability density function (pdf) for $\Lambda(T)$ is Gamma (a, b), conforming to

$$\pi\left(\Lambda(T)\right) \propto [\Lambda(T)]^{a-1} \exp[-b\Lambda(T)].$$

Note that in the above and in the sequel, a Gamma (α_1, α_2) refers to a Gamma random variable with pdf $f(x) \propto x^{\alpha_1 - 1} e^{-\alpha_2 x}, x > 0$. An independent prior pdf $g(\beta)$ of β then yields the joint prior of (θ, β) in the form

$$\pi(\theta, \beta) \propto \beta\theta^{-\beta a - 1} T^{\beta a} \exp[-b(T/\theta)^{\beta}] g(\beta). \tag{4}$$

Indeed (4) constitutes a generalization of the prior model used by Kuo and Yang (1996) in the sense that their model conforms to the special case of (4) when $T = 1$, a single unit of measured time. Choices for $g(\beta)$ may include any reasonable density with support on the positive real line. Guida et al. (1989) argue for a uniform prior for β over a positive support (β_1, β_2). Following Kyparisis and Singpurwalla (1985), one can also consider a more general Beta prior for β of the form

$$g(\beta) = \frac{\Gamma(k_1 + k_2)}{\Gamma(k_1)\Gamma(k_2)} \frac{(\beta - \beta_1)^{k_1 - 1}(\beta_2 - \beta)^{k_2 - 1}}{(\beta_2 - \beta_1)^{k_1 + k_2 - 1}},\ 0 \leq \beta_1 < \beta < \beta_2. \tag{5}$$

In what follows, we shall investigate the posterior inference of the intensity using both the noninformative and informative choice of priors.

3 Posterior Analysis of Current Intensity

In this section, we assimilate the results concerning the statistical inference of the *current intensity* $\lambda(y)$ evaluated at the truncation time y. We categorize the findings separately under the prior choices described in the previous section. This helps immensely in the comparison across different prior formulations as well as with the frequentist MLE–based results.

3.1 Analysis Under Noninformative Prior

Under the noninformative prior choice of (3), the posterior inference of current intensity $\lambda(y)$ of a PLP yields tractable, closed-form results. Further, the findings are remarkably similar to the frequentist results for the failure truncated case. The key step in the derivation, borrowed from Sen and Khattree (1998), is provided next. For facilitation of comparison, results in this section are expressed with the aid of $\widehat{\lambda}(y) = n^2 \left\{ y \sum_{i=1}^{n} \log(y/t_i) \right\}^{-1}$, the MLE of $\lambda(y)$.

Lemma 3.1 (Sen and Khattree, 1998): *A posteriori under the prior model of* (3)*,* $\lambda(y) \overset{d}{=} (\widehat{\lambda}(y)/4n^2)UV$*, where* U, V *are independent chi-square random variables with* $2n$ *and* $2(n-\delta)$ *degrees of freedom, respectively. Note that the posterior is improper for* $\delta \geq n$*.*

For $\delta = 1$, the result is identical to that for the failure-truncated case in the frequentist scenario. Note, however, that unlike the frequentist case, the result here continues to hold under time-truncated case and for any acceptable choice of δ. Several consequences of Lemma 3.1 are immediate. For instance, the Bayes estimator of $\lambda(y)$ under squared error loss is given by

$$\lambda_1^B(y) = \frac{n-\delta}{n} \widehat{\lambda}(y) \tag{6}$$

Clearly, for $\delta = 0$, $\lambda_1^B(y)$ matches the MLE $\widehat{\lambda}(y)$. Also, for $\delta = 1$, $\lambda_1^B(y)$ results in the estimator proposed by Higgins and Tsokos (1981) from a "heuristic" Bayesian viewpoint. By choosing other appropriate values of δ, alternative frequentist estimators proposed in the literature (c.f. Calabria et al., 1988; Qiao and Tsokos, 1998) are recovered from (6). The posterior median of $\lambda(y)$, namely,

$$\tilde{\lambda}_1^B(y) = (\widehat{\lambda}(y)/4n^2)\text{med}\,(UV),$$

corresponds to the Bayes estimator under absolute error loss. $\widetilde{\lambda}_1^B(y)$ is also the unique *Bayesian Posterior Pitman Closest estimator* (c.f. Ghosh and Sen, 1991) in the sense that it is closer to $\lambda(y)$ in comparison to any other estimator of $\lambda(y)$ with a posterior probability larger than 50%. It is also clear from Lemma 3.1 that a $100(1-\alpha)\%$ equal-tailed credible interval for $\lambda(y)$ is given by

$$\left[\frac{\widehat{\lambda}(y)}{4n^2} q_{\alpha/2},\ \frac{\widehat{\lambda}(y)}{4n^2} q_{1-\alpha/2} \right], \tag{7}$$

where q_γ is the $100\,\gamma^{\text{th}}$ percentile of the distribution of the product UV. It follows quite easily from Lemma 3.1 that asymptotically, as $n \to \infty$, $\sqrt{n}\left(\frac{\lambda(y)}{\widehat{\lambda}(y)} - 1\right) \xrightarrow{d} N(0,2)$ a posteriori, and thus, by contrast to (7), a $100(1-\alpha)\%$ large-sample Bayesian Credible interval for $\lambda(y)$ is given by:

$$\widehat{\lambda}(y)\left(1 \mp \sqrt{\frac{2}{n}} Z_{1-\alpha/2}\right),$$

where Z_γ is the $100\gamma^{\text{th}}$ percentile of standard normal distribution. This is identical to the classical large-sample result obtained by Crow (1982) for the failure-truncated case.

By choosing alternative loss functions, one can also derive as Bayesian estimators, other frequentist estimators suggested in the literature on the basis of somewhat ad-hoc considerations. Table 1 provides, under the present construct, a list of certain loss functions and the corresponding Bayesian estimators along with their frequentist analogs. In the table a denotes a generic parameter and $\hat{a}$ its estimator.

Loss Function	Bayes Estimator of $\lambda(y)$	Frequentist Analog
$L_1(a : \hat{a}) = (\hat{a} - a)^2$	$\lambda_1^B(y) = \frac{n-\delta}{n}\widehat{\lambda}(y)$	Higgins and Tsokos (1981) ($\delta = 1$) Calabria et al. (1988) ($\delta = 2, 3$) Qiao and Tsokos (1998) ($\delta = 5$)
$L_2(a; \hat{a}) = \left(\frac{\hat{a}}{a} - 1\right)^2$	$\lambda_2^B(y) = \frac{(n-2)(n-\delta-2)}{n^2}\widehat{\lambda}(y)$	Rigdon and Basu (1988) ($\delta = 1$)
$L_3(a; \hat{a}) = \frac{a}{\hat{a}} - \log\left(\frac{a}{\hat{a}}\right) - 1$	$\lambda_3^B(y) = \frac{n-\delta}{n}\widehat{\lambda}(y)$	Same as L_1.
$L_4(a; \hat{a}) = \frac{\hat{a}}{a} - \log\left(\frac{\hat{a}}{a}\right) - 1$	$\lambda_4^B(y) = \frac{(n-1)(n-\delta-1)}{n^2}\widehat{\lambda}(y)$	Rigdon and Basu (1988) ($\delta = 1$)

Table 1: Bayesian Estimators of $\lambda(y)$ Under Various Loss Functions

The loss functions presented in Table 1 have quite meaningful physical interpretations. L_2 corresponds to the standard weighted squared-error loss. Both L_3 and L_4 stem from the consideration of treating the deviations on either side asymmetrically. L_3 reflects a situation when one imposes heavier penalty on underestimation compared to overestimation. L_4 simply corresponds to the reverse scenario. Alternative motivations of L_3 as a LINEX-type loss (Varian, 1975; Zellner, 1986) or from consideration of entropy distance between probability distributions (Sinha and Ghosh, 1987) are also available in the literature. The most noteworthy point in this discussion is that unlike the frequentist estimators which are either (a) derived on ad-hoc or heuristic grounds, or (b) obtained as optimal in a *restricted* class of estimators, the corresponding Bayesian estimators are founded on firm, rigorous justification and are optimal over the class of all estimators under either sampling scenario.

3.1.1 Large-sample Comparisons

The remarkable similarity between the frequentist and the Bayesian inference with the non-informative prior motivates a study of the frequentist performance of the Bayesian estimators derived in this case. Towards that end, we present two results in this section. The first relates to the estimation of *current intensity* of the PLP. We also restrict attention to the 'failure-truncated' scenario, as the large-sample results in this scenario can be studied directly as a function of the growing sample size n, which is a non-random entity.

Theorem 3.1 *For the failure-truncated scheme under the sampling distribution, the quantity* $\sqrt{n}\left(\frac{\lambda_1^B(t_n)}{\lambda(t_n)} - 1\right)$ *converges to a normal random variable with mean zero and variance* 2*, as the sample size* n *grows large.*

Proof: By virtue of (6), note that

$$\sqrt{n}\left(\frac{\lambda_1^B(t_n)}{\lambda(t_n)} - 1\right) = \sqrt{n}\left(\frac{\widehat{\lambda}(t_n)}{\lambda(t_n)} - 1\right) - \frac{\delta}{\sqrt{n}}\frac{\widehat{\lambda}(t_n)}{\lambda(t_n)} \tag{8}$$

Using Crow's (1982) result, the first term on the right of (8) converges to a $N(0,2)$ random variable as $n \to \infty$. It also implies that the second term on the right hand side is $o_p(1)$. By an application of Slutsky's Theorem we have the result. □

Although the result is presented for λ_1^B, it is evident from Table 1 that the same large-sample behavior is exhibited by the other Bayes estimators as well. We now turn to the inference for the model parameters. Although it is not the main focus of the current article, the large-sample result presented nicely supplements a corresponding well-known result in the frequentist context. Further, it helps to identify and understand the extent of similarity between the frequentist and the Bayesian calculations in the present situation. The large-sample result is described for the parameterization $(\mu = \theta^{-\beta}, \beta)$ rather than in the original (θ, β) formulation. This parameterization has been adopted by various researchers, and is precisely the parameterization in which the corresponding MLE large-sample result has been derived. In the remaining part of this subsection, (μ^B, β^B) denote the Bayes estimators of (μ, β) under squared-error loss, and $(\widehat{\mu}, \widehat{\beta})$ indicate the corresponding MLE's.

Theorem 3.2 *For the failure-truncated scheme under the sampling distribution, as* $n \to \infty$, $\left(\sqrt{n}(\log n)^{-1}(\mu^B - \mu), \sqrt{n}(\beta^B - \beta)\right)'$ *converges in distribution to a (singular) bivariate normal with mean vector* $\mathbf{0}$ *and variance-covariance matrix*

$$\sum = \begin{bmatrix} \mu^2 & -\mu\beta \\ -\mu\beta & \beta^2 \end{bmatrix}$$

Proof: Under the noninformative prior (3), *a posteriori* β and μt_n^β are independently distributed as Gamma $\left(n-\delta, n/\widehat{\beta}\right)$ and Gamma $(n,1)$, respectively; a fact essentially inherent in the proof of Theorem 6.1 in Sen and Khattree (1998). Consequently, the Bayes estimator of β under squared-error loss is

$$\beta^B = \frac{n-\delta}{n}\,\widehat{\beta}$$

and thus

$$\sqrt{n}\left(\beta^B - \beta\right) = \sqrt{n}\left(\widehat{\beta}-\beta\right) - \frac{\delta}{\sqrt{n}}\,\widehat{\beta} \tag{9}$$

On the other hand, note that

$$\begin{aligned} \mu^B &= E_{\mu|\text{data}}[\mu] \\ &= E_{\beta|\text{data}}\left[t_n^{-\beta}\,E\left(\mu t_n^\beta|\beta,\text{data}\right)\right] \\ &= n\,E_{\beta|\text{data}}\left[t_n^{-\beta}\right] \\ &= n\left(1+\frac{\widehat{\beta}\log t_n}{n}\right)^{-(n-\delta)} \end{aligned}$$

which follows by the posterior distributional properties of β and μt_n^β. Consequently,

$$\begin{aligned} \log\mu^B &= \log n - (n-\delta)\,\log\left[1+\frac{\widehat{\beta}\log t_n}{n}\right] \\ &= \log n - \widehat{\beta}\log t_n + \frac{\delta}{n}\widehat{\beta}\log t_n + \frac{n-\delta}{n}\left(\textit{Higher powers of}\ \frac{\widehat{\beta}\log t_n}{n}\right) \\ &= \log\widehat{\mu} + \frac{\delta}{n}\widehat{\beta}\log t_n + \frac{n-\delta}{n}\left(\textit{Higher powers of}\ \frac{\widehat{\beta}\log t_n}{n}\right), \end{aligned} \tag{10}$$

where the last equality derives from the functional relationship $\widehat{\mu} = n t_n^{-\hat{\beta}}$ between the corresponding MLE's (c.f. Bhattacharyya and Ghosh, 1991). Under the sampling distribution, μt_n^β is distributed as a Gamma $(n,1)$ random variable, and thus, a simple application of Central Limit Theorem yields

$$\beta\log t_n = \log n - \log\mu + O_p(n^{-1/2}).$$

We now have

$$\frac{\widehat{\beta}\log t_n}{n} = \frac{\widehat{\beta}}{\beta}\left[\frac{\log n - \log\mu + O_p(n^{-1/2})}{n}\right] = O_p\left(\frac{\log n}{n}\right), \tag{11}$$

where we use the fact that $\widehat{\beta}/\beta \xrightarrow{P} 1$ as $n\to\infty$. Similar argument applies to the higher powers of $(\widehat{\beta}\log t_n)/n$, and so in view of (10) and (11), we arrive at

$$\begin{aligned} \frac{\sqrt{n}}{\log n}\left(\log\mu^B - \log\mu\right) &= \frac{\sqrt{n}}{\log n}\left(\log\widehat{\mu} - \log\mu\right) + o_p(1) \\ &= -\beta^{-1}\sqrt{n}(\widehat{\beta}-\beta) + o_p(1). \end{aligned} \tag{12}$$

The last equality is taken directly from Bhattacharyya and Ghosh (1991). As a consequence of delta method and (12),

$$\sqrt{n}(\log n)^{-1}(\mu^B - \mu) = \sqrt{n}(\log n)^{-1}\mu\left(\log \mu^B - \log \mu\right) + o_p(1) = -\mu\beta^{-1}\sqrt{n}(\widehat{\beta} - \beta) + o_p(1). \tag{13}$$

Since $2n\beta/\widehat{\beta} \sim \chi^2_{2(n-\delta)}$, it follows immediately that $\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \beta^2)$ as $n \to \infty$, and the required results follow in view of (9) and (13). □

Remark

The proof of Theorem 3.2 essentially demonstrates the asymptotic equivalence of the MLE's and the Bayesian estimators. The curious feature of the inference result is the singularity of the asymptotic variance-covariance matrix. The crucial observation underlying the singularity is given by the fact that

$$\sqrt{n}(\log n)^{-1}(\widehat{\mu} - \mu) + \mu\beta^{-1}\sqrt{n}(\widehat{\beta} - \beta) = o_p(1),$$

i.e. a linear combination of (non-uniformly) scaled MLE's converges to zero in probability. An identical relation holds when the MLE's are replaced by the Bayesian estimators. Fortunately, this does not pose any inferential problem in this case. The existence of pivotals facilitates exact joint inference for μ, β, and one does not have to take recourse to the large-sample results. Specifically, one can construct joint confidence region of the parameters based on the facts that the quantities $2n\beta/\widehat{\beta}$ and $2\mu t_n^\beta$ are independently distributed as chi-square random variables with $2(n-1)$ and $2n$ degrees of freedom, respectively. Theorem 6.1 of Sen and Khattree (1998) demonstrates that under the prior choice of (3), also *aposteriori* $2n\beta/\widehat{\beta}$ and $2\mu t_n^\beta$ are independent chi-square random variables with $2(n-\delta)$ and $2n$ as respective degrees of freedom. Thus, joint posterior credible region is also obtained without difficulty. The non-standard asymptotic result, nevertheless, presents interesting insights into the inference of an otherwise well-behaved model.

3.2 Analysis Under Informative Prior

We now consider the class of informative priors given in (4) and carry out the corresponding posterior inference for $\lambda(y)$. A result analogous to Lemma 3.1 is provided next.

Lemma 3.2 *Under the prior model of* (4), a posteriori, $\lambda(y) \stackrel{d}{=} y^{-1}U_1U_2$, *where conditionally* U_1 *given* U_2 *is distributed as Gamma* $\left(n + a, [1 + b(T/y)^{U_2}]\right)$, *and* U_2 *has a pdf* h *given by:*

$$h(u_2) \propto u_2^n \left\{\prod_{i=1}^n (t_i/T)\right\}^{u_2} [b + (y/T)^{u_2}]^{-(n+a)} g(u_2).$$

Proof: Using (2) and (4), the joint posterior of (θ, β) can be expressed as:

$$\pi(\theta, \beta|\text{data}) \propto \beta^{n+1} \left\{\left(\prod_{i=1}^n t_i\right) T^a\right\}^\beta g(\beta) \times \theta^{-\beta(n+a)-1} \exp\left\{-\left(\frac{y}{\theta}\right)^\beta \left[1 + b\left(\frac{T}{y}\right)^\beta\right]\right\}. \tag{14}$$

Using the change of variable $u_1 = (y/\theta)^\beta$ and $u_2 = \beta$, and noting that the jacobian of the transformation is $J = yu_2^{-1}u_1^{-u_2^{-1}-1}$, the joint posterior of (u_1, u_2) can be derived from (14) as

$$\pi(u_1, u_2|\text{data}) \propto u_2^n \left\{ \left(\prod_{i=1}^{n} (t_i/y) \right) \cdot (T/y)^a \right\}^{u_2} g(u_2)$$
$$\times\, u_1^{n+a-1} \exp\left\{-u_1[1 + b(T/y)^{u_2}]\right\}.$$

Upon observing that $\lambda(y) = y^{-1}u_1u_2$, the result follows. □

From Lemma 3.2, the Bayes estimator of $\lambda(y)$ under squared-error loss is obtained as

$$\lambda_*^B(y) = y^{-1} E_{U_2|\text{data}} \left[\frac{(n+a)U_2}{1 + b\left(\frac{T}{y}\right)^{U_2}} \right]. \tag{15}$$

Unlike the noninformative case, the expression in (15) has to be numerically evaluated and it does depend on the prior choice $g(\beta)$ of β. For the uniform prior of Guida et al. (1989) or the beta prior of Kyparisis and Singpurwalla (1985), (15) involves a one-dimensional numerical integration over a bounded interval. Alternatively, one can also adopt a simulation based approach to estimate the mean in (15) based on samples from the posterior pdf of β using standard rejection algorithms.

The case where T equals the truncation point y deserves a special mention as it provides an interesting analog to the noninformative case. Note that since the major role of T is in the prior construction, this scenario is only conceivable in the time-truncated situation. It is easy to see from Lemma 3.2 that in this case U_1 and U_2 in the lemma are distributed independently and (15) simplifies to

$$\lambda_*^B(y) = \{y(1+b)\}^{-1}(n+a)E[U_2|\text{data}],$$

namely, a multiplier of the marginal posterior mean of β. Further, if *a priori* β is distributed as a Gamma (α_1, α_2) random variable, we have U_1 and U_2 distributed *a posteriori* as independent Gamma $(n + a, 1 + b)$ and Gamma $(n + \alpha_1, \alpha_2 + \sum_{i=1}^n log(y/t_i))$, respectively. Consequently, $\lambda_*^B(y)$ reduces to

$$\lambda_*^B(y) = y^{-1} \frac{(n+a)(n+\alpha_1)}{\{\alpha_2 + \sum_{i=1}^n \log(y/t_i)\}(1+b)}.$$

This is in conformity with the noninformative scenario, where one can write $\lambda(y) \stackrel{d}{=} n\left\{y \sum_{i=1}^n \log(y/t_i)\right\}^{-1} \overline{U}_1\overline{U}_2$, $\overline{U}_1, \overline{U}_2$ being independent Gamma $(n, 1)$ and Gamma $(n - \delta, n)$ variables, respectively. Consequently, in this special case, the inference results for the informative and noninformative prior cases are virtually identical.

4 Prediction of Future Intensity

In this section, we shall provide the Bayesian predictive analysis of the intensity function of a PLP at a future failure. Several authors have investigated prediction issues for a PLP model. Kyparisis and Singpurwalla (1985) and Calabria et al. (1990) have focused on the predictive distribution of future failure times of a PLP. In addition, Bar-Lev et al. (1992) also discuss the prediction of number of failures in a future given time interval. Calabria and Pulcini (1996) and Beiser and Rigdon (1997) take up the prediction problem under the assumption that the failure process of the system follow a homogeneous Poisson process upon termination of the observation period. In this section we will develop the predictive inference of the intensity function at a future failure time under the assumption that a PLP prevails both for the observation period as well as in the future. The novelty in our solution lies in the application of a simulation-based Markov Chain Monte Carlo (MCMC) approach, popularized by Gelfand and Smith (1990), that in some cases simplifies the prediction problem substantially by avoiding high-dimensional numerical integration (or approximation), and in general presents a viable recourse for handling the problem of prediction.

The main objective here is to develop predictive inference for $\lambda(t_{n+m}, \theta, \beta) = (\beta/\theta)(t_{n+m}/\theta)^{\beta-1}$, the value of intensity function at the m^{th} future failure subsequent to the observation period, $m \geq 1$. The advantage of the MCMC approach is that since it estimates the joint predictive distribution of (t_{n+m}, θ, β), inference on any function of the triplet is available in virtually an automated way. So the solution we provide is more general than the prediction of a single quantity of interest.

As before, we integrate the failure and time-truncated schemes in the same framework. Note first that conditionally given the data and (θ, β), the pdf of t_{n+m} assumes the form

$$\begin{aligned} p(t_{n+m}|\text{data};\theta, \beta) &\propto \beta\theta^{-\beta} t_{n+m}^{\beta-1}[(t_{n+m}/\theta)^{\beta} - (y/\theta)^{\beta}]^{m-1} \\ &\quad \times \exp[-(t_{n+m}/\theta)^{\beta} + (y/\theta)^{\beta}], \quad t_{n+m} > y. \end{aligned} \tag{16}$$

Consequently, the joint predictive pdf of (t_{n+m}, θ, β) can be derived as

$$p(t_{n+m}, \theta, \beta|\text{data}) = p(t_{n+m}|\text{data}; \theta, \beta) \times p(\theta, \beta|\text{data}), \tag{17}$$

where $p(\theta, \beta|\text{data})$ refers to the posterior pdf of (θ, β). The distribution of any function of (t_{n+m}, θ, β) can then be obtained from (17) by an appropriate transformation followed by integration over a suitable space. Evidently, a tractable, closed-form solution is hard to obtain even for a simple prior choice. Below we document the findings from the MCMC approach applied to this problem in the context of the prior choices entertained in this article.

Noninformative Priors

A truncated pdf such as the one in (16) may be a potential cause for simulational problem for the MCMC approach. To circumvent this, we introduce the latent (unobserved) variables $T_{n+1}, T_{n+2}, \ldots, T_{n+m-1}$. Using the prior in (3), the joint density of the data and the unobservables is then

$$\begin{aligned} p(\text{data}, t_{n+1}, \ldots, t_{n+m}, \theta, \beta) \;&\propto\; \beta^{n+m-\delta}\theta^{-(m+n)\beta-1}\left(\prod_{i=1}^{n+m} t_i\right)^{\beta-1} \exp[-(t_{n+m}/\theta)^{\beta}], \\ & \qquad 0 < t_1 < \cdots < t_{n+m}; \theta > 0, \beta > 0. \end{aligned} \tag{18}$$

Note that the form in (18) is substantially simpler than (16) and facilitates the analysis to a large extent. For the MCMC approach, we need to generate samples from the appropriate full conditionals which are surprisingly simple in this case and are provided in Steps 1–2 below.

Step 1: Conditional of the Latent Variables

Given θ, β and the data, the random variables $X_1 = (T_{n+1}/\theta)^\beta - (y/\theta)^\beta$, $X_2 = (T_{n+2}/\theta)^\beta - (T_{n+1}/\theta)^\beta, \ldots, X_{n+m} = (T_{n+m}/\theta)^\beta - (T_{n+m-1}/\theta)^\beta$ are i.i.d. Exponential with mean 1. This is a direct consequence of the fact that for a general NHPP with mean function $\Lambda(t)$, the random variables $\Lambda(T_i) - \Lambda(T_{i-1}), i = 1, 2, \ldots$ constitute i.i.d. realizations from a standard Exponential. Also this property holds irrespective of whether we are dealing with a failure-truncated ($y = t_n$) or a time-truncated ($y = t_0$) case.

Step 2: Conditional for (θ, β)

Given the rest, the distribution of $W_1 = (T_{n+m}/\theta)^\beta$ and β are independent with W_1 having a Gamma$(n+m,\ 1)$ distribution and β following a Gamma $(n+m-\delta, \sum_{i=1}^{n+m} \log(t_{n+m}/t_i))$ distribution. This follows by appealing to (18) and applying a direct transformation to (θ, β).

MCMC iterations thus follow successive sample generations using the distributions in Step 1 and Step 2. In view of the known forms of the conditionals, the generation is fast and easy. For each triplet of generated (t_{n+m}, θ, β), $\lambda(t_{n+m}, \theta, \beta)$ (or any other function for that matter) can be calculated. All subsequent predictive calculations (e.g., distribution, mean, median, prediction interval) are based on a large MCMC output obtained after a sufficient burn-in.

Informative Prior

We now indicate the nuances of the predictive inference with the prior choice of (4)

Step 1: This is identical to Step 1 for the noninformative choice.

Step 2: The distribution of $W_1 = (T_{n+m}/\theta)^\beta$ and β are no longer independent. The joint conditional of W_1 and β given the rest is obtained as

$$p(w_1, \beta|\text{rest}) \propto \beta^{n+m} \left\{ \prod_{i=1}^{n+m} (t_i/t_{n+m}) \right\}^\beta (T/t_{n+m})^{a\beta} g(\beta) \times w_1^{m+n+a-1} \exp\{-w_1[1 + b(T/t_{n+m})^\beta]\}. \quad (19)$$

It is evident from (19) that the conditional distribution of W_1 given the rest is Gamma $(n+m+a,\ [1+b(T/t_{n+m})^\beta])$ analogous to Lemma 3.2.

Step 3: In this situation, additional steps are required to generate samples from the full conditional pdf of β given the rest. From (19), it follows that this conditional can be written as:

$$
\begin{aligned}
f(\beta|\text{rest}) \propto\ & \beta^{n+m}\left\{\prod_{i=1}^{n+m}(t_i/t_{n+m})\right\}^{\beta}(T/t_{n+m})^{a\beta} \\
& \times \exp[-w_1 b(T/t_{n+m})^{\beta}]g(\beta) \\
\equiv\ & k(\beta)g(\beta). \qquad (20)
\end{aligned}
$$

The pdf corresponding to (19) is non-standard and suitable rejection algorithms are required to sample from it. Note, however, that $k(\beta)$ is a positive, log-concave function of β thus having a unique positive maximum M which can be obtained by solving a nonlinear equation in β. One reasonable rejection method may then simply consist of the steps:

3.a Generate β from the prior $g(\beta)$.

3.b Generate u independently from a uniform $(0, 1)$.

3.c Accept β if $u \leq k(\beta)/M$, otherwise reject β.

Alternatively, a weighted bootstrap with weights proportional to $k(\beta_i)$, where β_i is a random sample from $g(\beta)$, can be used to generate random variates from a suitable approximation of f. In cases where the maximum M falls far off the prior support of $g(\beta)$, the bootstrap method, albeit approximate, may be a faster algorithm to employ. Both these methods are explained in detail in Smith and Gelfand (1992). Also note that a log-concave prior $g(\beta)$ renders $f(\beta|\cdot)$ to be log-concave also and in those cases an efficient algorithm such as Adaptive Rejection Sampling (ARS) (Gilks and Wild, 1992) can be employed.

5 Dealing with Aggregated Data

In engineering, biomedical and economic applications, one frequently encounters situations where the observation process is carried out at predetermined inspection times. In the past forty or so years numerous researchers have made methodological advances in the study of *aggregated, grouped* or *interval-censored* data which simply consists of counts of failed and censored observations within the inspection intervals. The focus of research in this area ranged from estimation and testing issues for specific life-distributions to distribution-free inference based on the theory of rank statistics. The traditional premise of aggregated data corresponds to the scenario where the data yielding the counts conform to a random sample from an unknown distribution. By contrast, in the study of interval-censored data for repairable systems, the context at hand, the counts in successive intervals correspond to failure times that are stochastically dependent. Triner and Phillips (1986) provide such an example of aggregated data consisting simply of number of failures of a system of generators on a marine vessel in nonoverlapping intervals of time (measured in years). Kyparisis and

Singpurwalla (1985) allude to an application in railroad engineering that necessitates analysis of interval censored data arising from a repairable system.

Kyparisis and Singpurwalla (1985) provide an approximate Bayesian inference for interval data arising from a repairable system whose failures are governed by a PLP. Let n_j denote the observed number of failures in the j^{th} time-interval $(l_{j1}, l_{j2}]$, $l_{j2} > l_{j1}$, and $j = 1, 2, \ldots, k$. Neither is it necessary to have the intervals of equal length nor is it assumed that the intervals are successive, i.e., $l_{j2} = l_{j+1,1}$. The likelihood function of (θ, β) under the PLP model is given by

$$L(\theta,\beta) \propto \prod_{j=1}^{k} \left[\left(\frac{l_{j2}}{\theta}\right)^{\beta} - \left(\frac{l_{j1}}{\theta}\right)^{\beta} \right]^{n_j} \times \exp\left\{ -\sum_{j=1}^{k} \left[\left(\frac{l_{j2}}{\theta}\right)^{\beta} - \left(\frac{l_{j1}}{\theta}\right)^{\beta} \right] \right\}. \tag{21}$$

A direct Bayesian analysis using (21) will generally be cumbersome and lead to intractable results. The solution provided by Kyparisis and Singpurwalla (1985) is based on numerical approximation. By contrast, our approach, based on MCMC, involves simulation and provides a general recourse to tackling interval data in the analysis of repairable systems. For the Bayesian implementation, we introduce the latent variables $T_{j1}, T_{j2}, \ldots, T_{jn_j}; j = 1, \ldots, k$, which represent the successive exact times to failure in the j^{th} inspection interval. Denoting by $\mathbf{t}$ the ensemble of T_{ji}'s, the augmented pdf assumes the form

$$\begin{aligned} p(\mathbf{t}; n_1, \ldots, n_m | \theta, \beta) &= \prod_{j=1}^{k} \prod_{i=1}^{n_j} \lambda(t_{ji}) \exp\left[\sum_{j=1}^{k} \int_{l_{j1}}^{l_{j2}} \lambda(t)\, dt \right] \\ &= (\beta/\theta^{\beta})^{\sum_{j=1}^{k} n_j} \left[\prod_{j=1}^{k} \prod_{i=1}^{n_j} t_{ji}^{\beta-1} \right] \exp\left\{ -\sum_{j=1}^{k} \left[\left(\frac{l_{j2}}{\theta}\right)^{\beta} - \left(\frac{l_{j1}}{\theta}\right)^{\beta} \right] \right\}. \end{aligned} \tag{22}$$

Equation (22) is in a more tractable form compared to (21) and is essentially obtained as a product of m time-truncated likelihoods within the given intervals. In this sense, (22) is a generalization of the expression in (2). In the simulation steps, additional care needs to be taken due to the presence of two sets of intrinsically different latent-variables, namely the T_{ji}'s; ones which were realized but were unobserved, and the T_{n+c}'s, $c = 1, \cdots, m$; the predicted ones. Here, as before $n = \sum_{j=1}^{k} n_j$ denotes the total number of observed failures. Step 1 of the conditional inference that is key to the MCMC algorithm now includes a new distributional result related to the latent variables T_{ji}'s.

New Step

Conditionally, given the data and the other parameters, the random variables $T_{j1}, T_{j2}, \cdots, T_{jn_j}$ are distributed as order statistics of size n_j from the pdf

$$\begin{aligned} f(t) &\propto \lambda(t) I(l_{ji} < t < l_{j2}) \\ &= \frac{\beta t^{\beta-1}}{l_{j2}^{\beta} - l_{j1}^{\beta}} I(l_{ji} < t < l_{j2}), \end{aligned}$$

$j = 1, \cdots, k$, where I denotes the indicator function.

Rest of the conditional inference is same as those of Step 1–Step 3 detailed in Section 4. Unlike the complete data case, statistical inference of *current intensity* for interval-censored data do not yield closed-form results. It can, however, be approached in a way identical to that for the predictive inference, by simply taking m=0 in the end-results. The simulation approach provides a nice, easy and efficient alternative to a direct analysis based on (20). Note that the existence of simple conditional inference in this case facilitates the efficiency of the simulation procedure in spite of the possibly large number of latent variable generation.

The latent variable approach in this context provides an alternative, interesting way to deal with agggregated data in general. For instance, in the case when one is interested in obtaining MLE's, the latent variables provide the groundwork for the EM algorithm (Dempster et al., 1977). Using the conditional distribution of T_{ji}'s given the counts n_j (incomplete data), for the PLP model, the EM algorithm reduces to a one-variable iterative step

$$n\left[\left(\beta^{(k+1)}\right)^{-1} - \frac{\sum_{j=1}^{m} l_{j2}^{\beta^{(k+1)}} \log l_{j2} - \sum_{j=1}^{m} l_{j1}^{\beta^{(k+1)}} \log l_{j1}}{\sum_{j=1}^{m} l_{j2}^{\beta^{(k+1)}} - \sum_{j=1}^{m} l_{j1}^{\beta^{(k+1)}}}\right]$$
$$= n\left(\beta^{(k)}\right)^{-1} - \sum_{j=1}^{m} n_j \frac{l_{j2}^{\beta^{(k)}} \log l_{j2} - l_{j1}^{\beta^{(k)}} \log l_{j1}}{l_{j2}^{\beta^{(k)}} - l_{j1}^{\beta^{(k)}}} \tag{23}$$

and a non-iterative step

$$\hat{\theta} = \left[n^{-1}\left(\sum_{j=1}^{m} l_{j2}^{\hat{\beta}} - \sum_{j=1}^{m} l_{j1}^{\hat{\beta}}\right)\right]^{1/\hat{\beta}}, \tag{24}$$

where $\beta^{(k)}$, $\beta^{(k+1)}$ are the estimates of β at the k^{th}, $(k+1)^{\text{st}}$ cycles, respectively, and $\hat{\beta}$ is the convergence point (MLE) of the sequence $\{\beta^{(k)}\}$ obtained from (23). Substantial simplification occurs in both the Bayesian and the frequentist calculations when the intervals are indeed successive and $l_{11} = 0$.

6 An Example

We implement the methodologies developed in this article in analyzing the failure data of a mechanical equipment fitted to a fleet of ships. The data originally described by Triner and Phillips (1986) and Sweeting and Phillips (1995), is reproduced in Table 2, and consists simply of the number of equipment failures over periodic yearly intervals. The mechanical equipment can be a set of two or more similar individual items per ship, for instance, the auxiliary generators. The failures are typically defined as noncompliance to operational guidelines or degradation above a threshold. A failed equipment is assumed to be repaired instantaneously in view of the short duration of repair times in comparison to the time between failures. Consequently, a *minimal repair* Nonhomogeneous Poisson Process model is deemed appropriate for this dataset. As evidenced from Table 2, the data collection process was initiated after a certain time has elapsed since the original installation of the equipment in question. This feature is not uncommon with failure history documentation in marine vessels presumably due to the fact that fleet owners often begin data acquisition subsequent to instituting a preventive maintenance scheme. Interval censoring in this case is clearly a result of unsystematic record keeping.

Inspection Interval (years)	Number of Failures
$1.5 - 2.5$	4
$2.5 - 3.5$	5
$3.5 - 4.5$	4
$4.5 - 5.5$	2
$5.5 - 6.5$	4
$6.5 - 7.5$	11
$7.5 - 8.5$	19
$8.5 - 9.5$	10
$9.5 - 10.33$	14

Table 2: Failures of mechanical equipment in a fleet of ships

Bayesian analysis is carried out for this example under the setting described in the article. We summarize here the results concerning the inference of the intensity function. Both informative and noninformative prior choices are considered. As indicated in Sweeting and Phillips (1995), the equipment under consideration is expected to exhibit a reliability decay over time. Incorporating this in the model, we entertain two distinct informative prior choices for $g(\beta)$. For the purposes of discussion, the noninformative and the two informative prior models are henceforth referred to as P1, P3, and P4, respectively. P3 assumes a Gamma $(0.01, 0.01)$ distribution for $\beta - 1$. The standard deviation of 10 makes the prior choice sufficiently diffused so that the posterior analysis is not driven too strongly by the prior model. The second informative prior model P4 assumes the Kyparisis and Singpurwalla (1985) choice of (5) for β on the support $(1, 5)$. This entails, $(\beta - 1)/4$ to have a standard Beta distribution with parameters k_1, k_2. For our analysis, we chose $k_1 = k_2 = 0.05$, translating into a mean of 0.5 and a variance of 0.238 for the standard Beta, once again rendering it to be a rather diffused choice. For illustration purposes, for both P3 and P4, T is chosen to be 1 year and $\Lambda(1)$ is assumed to have a Gamma $(0.01, 0.01)$ distribution. With a more systematic elicitation procedure, a stronger prior information on Λ can be obtained, thereby improving the accuracy of the analysis.

Using codes written in Fortran, for each of the prior models, fifty-one thousand MCMC samples of the triplet (θ, β, t_{n+m}) have been generated and five thousand of them are elected for the summary calculations with a burn-in of one thousand and a lag jump of ten. We used the ML estimates $\widehat{\theta} = 1.464, \widehat{\beta} = 2.203$ as the starting value of each chain. We have also used other starting points in order to verify stability and convergence of the generated chains. Table 3 exhibits posterior summary calculations for the current and predicted intensity $\lambda(\cdot)$ both noninformative (with $\delta = 1$) as well as the informative prior choices. Apart from the prior models mentioned above, the table also includes calculation for a specific prior

	Current Intensity			
Prior	Mean	Median	SD	95% Credible Interval
P1	18.540	18.363	3.089	[12.975, 25.264]
P2	18.667	18.426	3.120	[13.228, 25.427]
P3	23.394	23.181	3.418	[17.388, 30.731]
P4	23.792	23.570	3.567	[17.405, 31.540]
	m=5			
P1	18.938	18.758	3.139	[13.449, 25.632]
P2	18.851	18.628	3.168	[13.326, 25.540]
P3	24.184	23.970	3.431	[18.104, 31.381]
P4	24.528	24.334	3.470	[18.243, 31.722]
	m=10			
P1	19.699	19.481	3.382	[13.718, 26.818]
P2	19.757	19.548	3.296	[13.899, 26.934]
P3	25.362	25.116	3.586	[18.970, 32.930]
P4	25.549	25.292	3.659	[19.123, 33.308]

Table 3: Posterior summary calculations for current and future intensity functions for the fleet equipment failure data

model, denoted by P2, where β and $\Lambda(10.33)$ are assumed to be i.i.d. realizations from a Gamma $(0.01, 0.01)$ distribution. As indicated in Section 3.2, this special choice makes the posterior analysis analogous to the noninformative case. This observation is duly supported by the numerical calculations. The main distinction between the noninformative and the informative cases corresponding to models P3 and P4 is in the posterior mean/median estimates. The right shift in the central measures for the informative prior choices is evidently due to the added information for β. The precision of the estimates on the other hand, as measured by the standard deviation, does not change much across the prior choices as well as the prediction region.

The histograms in Figure 1 for the predicted intensity at the first and fifth future failure indicate a mild right-skew in the overall distribution. A distinctly unimodal feature of the distribution is apparent in each case. By contrast to the frequentist approach, the simulation based Bayesian route provides more complete and unified framework for estimation and prediction that does not require any large-sample approximation whatsoever.

7 Analysis of Multiple Systems

We have presented the entire development in this article in the context of observing failures for a single repairable system. It is quite common in reliability experiments to study multiple copies of the similar system simultaneously. The objective then is to draw inference on the underlying process via a judicious integration of the collective data. For example, one may be interested in predicting the performance of a new system yet to be installed on the basis of the past performance of similar systems. Let us now indicate how the inference methods detailed for a single system extend to the multiple systems case without much difficulty.

Assume that m identical systems are under investigation. Let t_{ij} denote the i^{th} observed failure time on the j^{th} system which has been observed until time y_j, where $y_j = t_{jn_j}$, if the process is terminated at the n_j^{th} failure; or $y_j = t_{0j}$, a predetermined time at which the process was time-truncated. Let n_j be the number of observed failures for the j^{th} system and $N = \sum_{j=1}^{m} n_j$ be the total number of failures over all systems. With the assumption of conditional independence of the system failure processes given the parameter values, the likelihood function under PLP is

$$\begin{aligned} L(\theta,\beta) &= \prod_{j=1}^{m} \left\{ \prod_{i=1}^{n_j} \lambda(t_{ij}) \right\} \exp[-\Lambda(y_j)] \\ &= (\beta/\theta^{\beta})^{N} \left(\prod_{j=1}^{m} \prod_{i=1}^{n_j} t_{ij} \right)^{\beta-1} \exp\left[-\sum_{j=1}^{m} (y_j/\theta)^{\beta} \right]. \qquad (25) \end{aligned}$$

Structurally, (25) is somewhat similar to (2) and thus, the inference methods detailed in this article work in an analogous, albeit more involved, manner. Indeed, several special cases and ramifications of (25) yield results that are immediate extensions of those for the single system case. For instance, consider the special case of a time-truncated scheme, when all systems are singly truncated at the same point y, i.e., $y_j = y$ for $j = 1, 2, \ldots, m$. Then, under the noninformative prior choice of (3), it follows that *a posteriori*, β and $(y/\theta)^{\beta}$ are distributed independently according to a Gamma $\left(N - \delta,\ \sum_j \sum_i \log(y/t_{ij})\right)$ and a Gamma (N, m), respectively. Consequently, a simple extension of Lemma 3.1 assumes the form:

$$\lambda(y)|\text{data} \stackrel{d}{=} \left(\frac{\widehat{\lambda}(y)}{4N^2} \right) \chi^2_{(2N)} \chi^2_{(2(N-\delta))}.$$

Here $\widehat{\lambda}(y)$, as before, denotes the MLE of $\lambda(y)$. Thus, the inference in this case, for the informative prior as well, is virtually identical to that for the case of a single system.

One can deal with nonidentical (but independent) systems in an analogous manner. In fact, if the difference between the systems is manifested through difference in the parameters, then the problem essentially reduces to that of single copies of different systems, and in spirit, is the same as that presented earlier in the article. A compromise, somewhat in the middle of dealing with completely identical and nonidentical systems, is to assume different scale parameters (θ_j) for the different systems while maintaining that the same growth parameter (β) acts on them. One can carry out the relevant Bayesian calculations after appropriately modifying (25). Rigdon and Basu (2000) present an Empirical Bayes analysis for nonidentical PLP models under this framework. One prior model, where tractable results are obtained, is given by a noninformative choice of the form

$$\pi(\theta_1, \ldots, \theta_m, \beta) \propto \beta^{-\delta} \prod_{j=1}^{m} \theta_j^{-1},$$

obtained by applying conditional independence of θ_j given β and adopting a noninformative choice both for β and θ_j. Analogous to the single system case with (3), here also, we have the result that a posteriori β and $(y_j/\theta_j)^\beta, j = 1, \ldots, m$ are mutually independent, each with an appropriate Gamma distribution.

8 Concluding Remarks

In this article, we have presented a simple sampling-based approach in dealing with the Bayesian prediction problem for a PLP. Kuo and Yang (1996) provide a Bayesian treatment for the NHPP in a generic manner. They concentrate on the characterization of a general NHPP and prescribe the use of MCMC approach for the Bayesian computation. We, however, focus on a particular model that is one of the most commonly used ones for hardware as well as software failures pertaining to a repairable system. Under quite reasonable and general prior choices, we provide detailed analysis for the prediction problem. Our analysis on one hand covers the previous studies undertaken in this context, and on the other, has made significant advances into dealing with more complex sampling schemes. The approach presented here is quite general in the sense of being adaptable to any prior choice as well as the inference of any process parameter.

We conclude our discussion with an allusion to a practical problem that constitutes the motivational premise for this work. Estimation of current intensity at the end of the developmental phase is useful from the viewpoint of obtaining a baseline for the operational phase. However, the main underlying assumption that the same NHPP or a HPP with a known intensity prevails in the operational stage, may be quite unrealistic for many systems. For example, it is well known in the defense industry that for environmental or other external reasons, various typical defense systems experience failure modes in the operational testing phase those have not surfaced in the controlled developmental phase. The standard analyses do not take into account this change in the failure process in any formal manner. A meaningful integration of the failure data obtained from developmental and operational phase and its subsequent analyses constitute an open research area that needs to be explored in detail.

References

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

[2] Bar-lev, S. K., Lavi, I., and Reiser, B. (1992). Bayesian inference for the power law process. *Annals of Institute of Statistical Mathematics*, **44**, 623–639.

[3] Beiser, J. A., and Rigdon, S. (1997). Bayes prediction for the number of failures of a repairable system. *IEEE Transactions on Reliability* **46**, 291–295.

[4] Bhattacharyya, G. K. and Ghosh, J. K. (1991) Asymptotic properties of estimators in a binomial reliability growth model and its continuous-time analog, Journal of Statistical Planning and Inference, 29, 43–53.

[5] Calabria, R., Guida, M., and Pulcini, G. (1988). Some modified Maximum Likelihood Estimators for the Weibull Process. *Reliability Engineering and System Safety*, **23**, 51–58.

[6] Calabria, R., Guida, M., and Pulcini, G. (1990). Bayes estimation of prediction intervals for a power law process. *Communications in Statistics–Theory and Methods*, **19**, 3023–3035.

[7] Calabria, R., and Pulcini, G. (1996). Maximum likelihood and Bayes prediction of current system lifetime. *Communications in Statistics–Theory and Methods*, **25**, 2297–2309.

[8] Crow, L. H. (1974). Reliability analysis for complex repairable systems. In: Proschan, F., Serfling, R. J. (Eds.), *Reliability and Biometry*, 379–410.

[9] Crow, L. H. (1982). Confidence interval procedures for the Weibull process with applications to reliability growth. *Technometrics*, **24**, 67–72.

[10] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, Series B*, **39**, 1–38.

[11] Duane, J. T. (1964). Learning curve approach to reliability monitoring. *IEEE Transactions on Aerospace*, **2**, 563–566.

[12] Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.

[13] Ghosh, M., and Sen, P. K. (1991). Bayesian Pitman closeness. *Communications in Statistics – Theory and Methods*, **20**, 3729–3750.

[14] Gilks, W. R. and Wild, P. (1992) Adaptive rejection sampling for Gibbs sampling, Applied Statistics, 41, 337–348.

[15] Guida, M., Calabria, R., and Pulcini, G. (1989). Bayes inference for a non-homogeneous Poisson process with power intensity law. *IEEE Transactions on Reliability*, **38**, 603–609.

[16] Higgins, J. J., and Tsokos, C. P. (1981). A quasi-Bayes estimate of the failure intensity of a reliability growth model. *IEEE Transactions on Reliability*, **R-30**, 471–475.

[17] Kuo, L., and Yang, T. Y. (1996). Bayesian computation for nonhomogeneous Poisson processes in software reliability. *Journal of the American Statistical Association*, **91**, 763–773.

[18] Kyparisis, J., and Singpurwalla, N. (1985). Bayesian inference for the Weibull process with applications to assessing software reliability growth and predicting software failures. In L. Billard (ed.), *Computer Science and Statistics: the Interface*, 57–64.

[19] Lee, L., and Lee, S. K. (1978). Some results on inference for the Weibull process. *Technometrics*, **20**, 41–45.

[20] Lingham, R. T., and Sivaganesan, S. (1997). Testing hypotheses about the power law process under failure truncation using intrinsic Bayes factors. *Annals of Institute of Statistical Mathematics*, **49**, 693–710.

[21] Qiao, H., and Tsokos, C. P. (1998). Best efficient estimates of the intensity function of the power law process. *Journal of Applied Statistics*, **25**, 111–120.

[22] Rigdon, S. E., and Basu, A. P. (1988). Estimating the intensity function of a Weibull process at the current ime: failure truncated case. *Journal of Statistical Computation and Simulation*, **30**, 17–38.

[23] Rigdon, S. E., and Basu, A. P. (1990). Estimating the intensity function of a power law process at the current time: time truncated case. *Communications in Statistics – Simulation*, **19**, 1079–1104.

[24] Rigdon, S., and Basu, A. P. (2000). *Statistical methods for the reliability of repairable systems*. John Wiley and Sons: New York.

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

[26] Sinha, B. K., and Ghosh, M. (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix and the generalized variance under entropy loss. *Statistics and Decisions*, **5**, 201–227.

[27] Smith, A. F. M., and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician*, **46**, 84–88.

[28] Sweeting, T. J., and Phillips, M. J. (1995). An application of nonstandard asymptotics to the analysis of repairable-systems data. *Technometrics*, **37**, 428–435.

[29] Triner, D. A., and Phillips, M. J. (1986). The Reliability of Equipment Fitted to a Fleet of Ships. In *Proceedings of the 9th Advances in Reliability Technology Symposium, Bradford*, Warrington, U.K.: Atomic Energy Authority.

[30] Varian, H. R. (1975). A Bayesian approach to real estate assessment. In: Fienberg, E., Zellner, A. (Eds.), *Studies in Bayesian Econometrics and Statistics*, in honor of L. J. Savage. North-Holland, Amsterdam, 195–208.

[31] Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, **81**, 446–451.

Figure 1: Predicted intensity at the first and the fifth future failures for ship-equipment failure data with prior choices P1 (top row), P3 (middle row), and P4 (bottom row)