Journal of the Korean Statistical Society (2005), **34**: *4, pp 331–344*

BAYESIAN INFERENCE FOR THE POWER LAW PROCESS WITH THE POWER PRIOR

HYUNSOO KIM[1], SANGA CHOI[2], AND SEONG W. KIM[3]

ABSTRACT

Inference on current data could be more reliable if there exist similar data based on previous studies. Ibrahim and Chen (2000) utilize these data to characterize the power prior. The power prior is constructed by raising the likelihood function of the historical data to the power a_0, where $0 \leq a_0 \leq 1$. The power prior is a useful informative prior in Bayesian inference. However, for model selection or model comparison problems, the propriety of the power prior is one of the critical issues. In this paper, we suggest two joint power priors for the power law process and show that they are proper under some conditions. We demonstrate our results with a real dataset and some simulated datasets.

AMS 2000 subject classifications. Primary 62G20; Secondary 62F15.
Keywords. Power prior, power law process, historical data, propriety.

1. INTRODUCTION

When we perform a Bayesian inference, prior elicitation plays a very important role. In principle, priors formally represent available information but in practice noninformative and improper priors are often used. Nevertheless, they cannot be used in some situations such as model selection or hypothesis testing. In these cases a proper prior on the parameters is needed making Bayesian inference plausible. Furthermore, noninformative priors may not reflect real prior information that one may need for a specific situation. Thus, when we have real prior information, it is possible to make posterior inference quite accurate. This often occurs when the current study is similar to the previous study in measuring

Received February 2005; accepted November 2005.
[1]Division of Applied Mathematics, Hanyang University, Ansan, 426-791, Korea
[2]Department of Mathematics, Ajou University, Suwon, 443-749, Korea
[3]Division of Applied Mathematics, Hanyang University, Ansan, 426-791, Korea (e-mail : seong@hanyang.ac.kr)

the response and covariates such as clinical trials, carcinogenicity studies, and other problems including reliability and survival data.

The data arising from previous studies are referred as 'historical data'. One method of constructing an informative prior based on the historical data is the power prior of Ibrahim and Chen (2000). Suppose that the historical data from a previous study is available for the current study. The power prior is defined by the likelihood function based on the historical data, raised to a power a_0, where a_0 $(0 \leq a_0 \leq 1)$ is a scalar parameter that controls the influence of the historical data on the current study. This idea has been discussed by several authors. Zellner (1988) was one of these authors, who proposed the idea of raising a likelihood to a power in the context of information and optimal processing rules.

The power prior is a useful informative prior for Bayesian inference such as model selection and model comparison because it inherently automates the informative prior specification for all possible models in the model space. Chen, Ibrahim, Shao and Weiss (1999) utilize the power prior for model selection in generalized linear mixed models. Ibrahim, Chen and Ryan (2000) use the power prior to analyze time series data.

Data truncation is commonly occurred in repairable systems, among which the power law process (PLP) is perhaps the most popular model. However, there are two sampling schemes in repairable systems, namely time truncation and failure truncation. In the former, the observation of the failure times is restricted to a pre-fixed interval $[0, t_0]$, and the failure times during this interval are recorded. On the other hand, in failure truncation, a pre-determined number, n, of successive failure times of the process are obtained. In this article we are only concerned with failure truncation. In particular, we use the power prior to estimate existing parameters in the PLP including the shape parameter, which determines the pattern of failure times.

Consider a nonhomogeneous Poisson processes (NHPP). A random variable of special interest is $N(t)$, the number of failures in the time interval $(0, t]$. The intensity function of a counting process $\{N(t), t \geq 0\}$ is defined as

$$\nu(t) = M'(t) = \frac{d}{dt}E[N(t)],$$

where $M(t)$ denotes the mean number of failures in the interval $(0, t]$, often called the *mean value function.* Let $X = (X_1, \cdots, X_n)$ be the first n failure times of the NHPP with observed values $x_1 < \cdots < x_n$. The joint density of X is then

$$L(X|\theta) = \prod_{i=1}^{n} \nu(x_i) \cdot \exp\{-M(x_n)\}. \tag{1.1}$$

Consider the PLP among nonhomogeneous Poisson processes. This model is motivated by Duane (1964) in order to fit the data in electricity-producing plants. Crow (1974) generalized the model with the following intensity function,

$$\nu(t) = \frac{\beta}{\eta} t^{\beta-1},\ \eta > 0,\ \beta > 0,\ t > 0,$$

where β is the shape parameter and η is the scale parameter. From (1.1) the likelihood function of the failure times is

$$L(\beta, \eta|\mathbf{x}) = \left(\frac{\beta}{\eta}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\beta-1} \exp\left\{-\frac{{x_n}^{\beta}}{\eta}\right\},\ 0 < x_1 < \cdots < x_n. \tag{1.2}$$

We note that the system is deteriorating when $\beta > 1$, and the system is improving over time when $\beta < 1$. Further, the process reduces to a homogeneous Poisson process when $\beta = 1$. Thus, the inference on the shape parameter is of interest.

There are several articles concerning Bayesian inference for the PLP. Kyparisis and Singpurwalla (1985) considered a Bayesian approach for making inference about the number of failures in a future time interval modelled by the PLP. Guida, Calabria and Pulcini (1989) present Bayesian procedures for the analysis of failure-truncated data from the PLP. Bar-Lev, Lavi and Reiser (1992) develop posterior distributions for the expected number of failures in the PLP. Lingham and Sivaganesan (1997) conducted a test for the shape parameter of the PLP. Kim and Sun (2000) considered a multiple test using an encompassing model. Kim, Kim and Kim (2003) dealt with model selection problems for the PLP.

The rest of this article is organized as follows. In Section 2, we overview the power prior in conjunction with the power law process. In Section 3, we present the proposed model and show the propriety of the joint power prior distributions. In Section 4, we analyze a real dataset and conduct a simulation study. We close the article with a brief discussion in Section 5.

2. THE POWER PRIOR

We consider the power prior of the PLP. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the failure time for the PLP from the current study. Suppose we have historical data $\mathbf{x_0} =$

$(x_{01}, \ldots, x_{0n_0})$. Let $L(\beta, \eta \mid \mathbf{x})$ denote the likelihood function for the current study, which is given by (1.2). Let $\pi_0(\beta, \eta \mid \cdot)$ denote the prior distribution for (β, η), which is called the initial prior. This initial prior is assumed before the historical data $\mathbf{x_0}$ are observed. From Ibrahim and Chen (2000) we define the joint power prior distribution of (β, η, a_0) for the current study as

$$\pi(\beta, \eta, a_0 \mid \mathbf{x_0}) \propto [L(\beta, \eta, \mid \mathbf{x_0})]^{a_0} \pi_0(\beta, \eta \mid c_0) \pi(a_0 \mid \gamma_0), \tag{2.1}$$

where c_0 is a specified hyperparameter for the initial prior and γ_0 is a specified hyperparameter for the prior distribution of a_0. The parameter c_0 controls the impact of the initial prior $\pi_0(\beta, \eta \mid c_0)$, and the parameter a_0 is a precision parameter. The parameter a_0 controls heaviness of the tails of the prior for (β, η). As a_0 becomes smaller, the tails of (2.1) become heavier. Such control may be important when there is heterogeneity between the previous and the current study or the sample sizes of two studies are quite different. It is reasonable that the range of a_0 is restricted to be between 0 and 1, and thus it is natural that the distribution for $\pi(a_0 \mid \gamma_0)$ is chosen to be a beta distribution. The beta prior for a_0 appears to be the most natural prior to use and leads to the most natural elicitation scheme. The prior in (2.1) does not have a closed form in general. However a desirable feature of (1.2) is that it creates heavier tails for the marginal prior of (β, η) than we assume a_0 is fixed. Thus (1.2) is more flexible in weighting the historical data. When we fix $a_0 = 1$, (2.1) can be the posterior distribution of (β, η, a_0) from the historical data. When $a_0 = 0$, the prior distribution does not depend on the historical data and (2.1) can be a usual prior.

The joint power prior in (2.1) can be generalized when multiple historical datasets are available. Suppose that there are N historical datasets, and let $\mathbf{x}_{0k} = (x_{0k_1}, \cdots, x_{0k_{n_{0k}}})$ be the historical data based on the kth study, $k = 1, \ldots, N$. In this case, it is desirable to define a precision parameter a_{0k} for each historical study, and take the distribution for a_{0k}'s to be i.i.d. beta distribution with parameters $\gamma_0 \equiv (\delta_0, \lambda_0), k = 1, \ldots, N$. For $\mathbf{a_0} = (a_{01}, \ldots, a_{0N})$, the joint power prior in (2.1) can be generalized as

$$\pi(\beta, \eta, \mathbf{a_0} \mid \mathbf{x}_{0k}) \propto \prod_{k=1}^{N} [L(\beta, \eta, \mid \mathbf{x}_{0k})]^{a_{0k}} \pi_0(\beta, \eta \mid c_0) \pi(\mathbf{a_0} \mid \gamma_0). \tag{2.2}$$

3. PROPRIETY OF THE POWER PRIOR

A proper prior distribution is essential in any informative Bayesian inference. Especially, it plays a crucial role in model selection or model comparison to compute Bayes factors and posterior probabilities. Chen, Ibrahim, and Shao (2000) found conditions for the propriety of the joint power prior distribution in the generalized linear models. In this paper, we derive them for the PLP.

Note that the likelihood function is given by (1.2). For the parameter a_0 we assume a beta prior with $\gamma_0 = (\delta_0, \lambda_0)$. If we impose two different initial priors for (β, η), then we come up with different results. They are described in Theorem 3.1 and Theorem 3.2.

THEOREM 3.1. *Suppose that the initial prior distribution for (β, η) is Jeffreys's prior, and a_0 has a beta distribution with hyperparameters (δ_0, λ_0). If $\delta_0 > 2$, then the joint power prior (2.1) for (β, η, a_0) is proper.*

PROOF. Note that Jeffreys's prior for (β, η) is $1/(\beta\eta)$. The joint power prior for (β, η, a_0) is then

$$\pi(\beta, \eta, a_0|\mathbf{x}_0) \propto \Big[\Big(\frac{\beta}{\eta}\Big)^{n_0}\Big(\prod_{i=1}^{n_0} x_{0i}\Big)^{\beta-1} \exp\Big\{-\frac{x_{0n_0}{}^{\beta}}{\eta}\Big\}\Big]^{a_0} \frac{1}{\beta\eta} a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}.$$

Let

$$h(a_0) = a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}.$$

Then we have

$$\begin{aligned}
&\int\int\int \pi(\beta, \eta, a_0|\mathbf{x}_0) d\eta d\beta da_0 \\
&\propto \int_0^1\int_0^\infty\int_0^\infty \Big[\Big(\frac{\beta}{\eta}\Big)^{n_0}\Big(\prod_{i=1}^{n_0} x_{0i}\Big)^{\beta-1} \exp\Big\{-\frac{x_{0n_0}{}^{\beta}}{\eta}\Big\}\Big]^{a_0} \frac{1}{\beta\eta}\, h(a_0) d\eta d\beta da_0 \\
&= \int_0^1 \Big(\prod_i x_{0i}\Big)^{-a_0} h(a_0)\, a_0{}^{-a_0 n_0}\, \Gamma(a_0 n_0) \int_0^\infty \beta^{a_0 n_0 - 1}\Big(\prod_i x_{0i}\Big)^{a_0\beta} x_{0_{n_0}}^{-a_0 n_0 \beta} d\beta da_0 \\
&= \int_0^1 \Big(\prod_i x_{0i}\Big)^{-a_0} h(a_0)\, a_0{}^{-a_0 n_0}\Big[\frac{\Gamma(a_0 n_0 + 1)}{a_0 n_0}\Big]^2 \\
&\quad \times \Big[a_0\Big(n_0 \log x_{0n_0} - \log\big(\prod_i x_{0i}\big)\Big)\Big]^{-a_0 n_0} da_0. \qquad (3.1)
\end{aligned}$$

Since

$$\Gamma(a_0 n_0) = \frac{\Gamma(a_0 n_0 + 1)}{a_0 n_0} \le \frac{\Gamma(n_0+1)}{a_0},$$

(3.1) is less than or equal to

$$\begin{aligned}
&\Gamma^2(n_0+1)\int_0^1 \Big(\prod_i x_{0i}\Big)^{-a_0} h(a_0)\, a_0{}^{-2a_0(n_0+1)} \\
&\qquad\qquad \times \Big(n_0 \log x_{0n_0} - \log\big(\prod_i x_{0i}\big)\Big)^{-a_0 n_0} da_0 \\
&\le \Gamma^2(n_0+1)\exp\Big\{\frac{2n_0}{e}\Big\}\int_0^1 h(a_0) a_0^{-2}\Big(\prod_i x_{0i}\Big)^{-a_0} \\
&\qquad\qquad\qquad \times \Big(n_0 \log x_{0n_0} - \log\big(\prod_i x_{0i}\big)\Big)^{-a_0 n_0} da_0 \\
&\le K\int_0^1 a_0^{\delta_0 - 3}(1-a_0)^{\lambda_0 - 1} da_0 < \infty,
\end{aligned}$$

where

$$\begin{aligned}
K &= \Gamma^2(n_0+1)\exp\{2n_0/e\}\Big(1 + \big(\prod_i x_{0i}\big)^{-1}\Big) \\
&\quad \times \Big[1 + \Big(n_0 \log x_{0n_0} - \log\big(\prod_i x_{0i}\big)\Big)^{-n_0}\Big].
\end{aligned}$$

This completes the proof. □

We need to define a truncated beta distribution which will be used to prove Theorem 3.2 below.

DEFINITION 3.1. *A truncated beta random variable X has the following probability density function:*

$$f(x) \propto x^{\alpha-1}(1-x)^{\beta-1}, 0 < a < x < b < 1.$$

We call a and b the lower bound and the upper bound respectively. And α and β are hyperparameters for a conventional beta distribution.

THEOREM 3.2. *Suppose that the initial prior distribution on (β, η) is a uniform prior, and the prior distribution for a_0 is a truncated beta distribution with hyperparameters (δ_0, λ_0). If $\delta_0 > 1$, then the joint power prior for (β, η, a_0) is proper.*

PROOF. The joint power prior for (β, η, a_0) is

$$\pi(\beta, \eta, a_0|\mathbf{x}_0) \propto \left[\left(\frac{\beta}{\eta}\right)^{n_0}\left(\prod_{i=1}^{n_0} x_{0i}\right)^{\beta-1} \exp\left\{-\frac{x_{0n_0}{}^{\beta}}{\eta}\right\}\right]^{a_0} h(a_0),$$

where a_0 is left truncated with lower bound

$$C_{(X)} = max\{\log x_{0_{n0}}/(n_0 \log x_{0_{n0}} - \log \prod_i x_{0i}), 1/(n_0 - 1)\},$$

provided $\log x_{0_{n0}}/(n_0 \log x_{0_{n0}} - \log \prod_i x_{0i}) < 1$. Then we have

$$\begin{aligned}
&\int\int\int \pi(\beta, \eta, a_0|\mathbf{x}_0) d\eta d\beta da_0 \\
&\propto \int_{C_{(X)}}^{1}\int_0^{\infty}\int_0^{\infty} \left[\left(\frac{\beta}{\eta}\right)^{n_0}\left(\prod_{i=1}^{n_0} x_{0i}\right)^{\beta-1} \exp\left\{-\frac{x_{0n_0}{}^{\beta}}{\eta}\right\}\right]^{a_0} h(a_0) d\eta d\beta da_0 \\
&= \int_{C_{(X)}}^{1} h(a_0)\Gamma(a_0 n_0 - 1)\int_0^{\infty} \beta^{a_0 n_0}\left(\prod_i x_{0i}\right)^{a_0(\beta-1)}\left(a_0 x_{0n_0}{}^{\beta}\right)^{1-a_0 n_0} d\beta da_0 \\
&= \int_{C_{(X)}}^{1} h(a_0)\Gamma(a_0 n_0 - 1)\Gamma(a_0 n_0 + 1) a_0{}^{1-a_0 n_0}\left(\prod_i x_{0i}\right)^{-a_0} \\
&\qquad \times \left\{\left(n_0 \log x_{0n_0} - \log \prod_i x_{0i}\right) a_0 - \log x_{0n_0}\right\}^{-(a_0 n_0+1)} da_0. \qquad (3.2)
\end{aligned}$$

Since

$$\Gamma(a_0 n_0 - 1) = \frac{\Gamma(a_0 n_0 + 1)}{(a_0 n_0 - 1)(a_0 n_0)} \leq \frac{\Gamma(n_0 + 1)}{a_0^2},$$

(3.2) is less than or equal to

$$\begin{aligned}
&\Gamma^2(n_0 + 1)\int_{C_{(X)}}^{1} h(a_0)\left(\prod_i x_{0i}\right)^{-a_0} a_0{}^{-(a_0 n_0+1)} \\
&\qquad \times \left\{\left(n_0 \log x_{0n_0} - \log \prod_i x_{0i}\right) a_0 - \log x_{0n_0}\right\}^{-(a_0 n_0+1)} da_0. \qquad (3.3)
\end{aligned}$$

Since

$$a_0{}^{-a_0 n_0} \leq e^{n_0/e},$$

(3.3) is less than or equal to

$$K\int_{C_{(X)}}^{1} h(a_0) a_0^{-1} da_0,$$

where $K = \Gamma^2(n_0+1)e^{n_0/e}C_{(X)}{}^{-2n_0}(1+(\prod_i x_{0i})^{-1})$. This completes the proof. $\square$

REMARK 3.1. *The condition on δ_0 in Theorem 1 is useful for choosing the hyperparameters (δ_0, λ_0) in (2.1).*

When multiple historical data sets are available the power prior in (2.1) can be generalized as (2.2). The conditions for the propriety in (2.2) are similar to those in (2.1).

THEOREM 3.3. *Suppose that the initial prior distribution for (β, η) is Jeffreys's prior, and $\mathbf{a}_0$ has a iid beta distribution with hyperparameters (δ_0, λ_0) for each of the historical data D_{0k}, $k = 1, \dots, M$. If $\delta_0 > 2$, then the joint prior distribution $\pi(\beta, \eta, \mathbf{a}_0 \mid D_{01}, \dots, D_{0M})$ given in (2.2) is proper.*

PROOF. Similar to Theorem 3.1, (2.2) is represented by

$$\int_{[0,1]^M}\int_0^\infty\int_0^\infty \prod_{k=1}^{M}\Big[\Big[\Big(\frac{\beta}{\eta}\Big)^{n_{0k}}\Big(\prod_{i=1}^{n_{0k}} x_{0_i}\Big)^{\beta-1}\exp\Big\{-\frac{x_{0_{n_{0k}}}^{\beta}}{\eta}\Big\}\Big]^{a_{0k}} \times a_{0k}^{\delta_0-1}(1-a_{0k})^{\lambda_0-1}\Big] \times \frac{1}{\eta\beta}d\eta d\beta d\mathbf{a}_0. \tag{3.3}$$

Integrating out η, (3.3) reduces to

$$\int_{[0,1]^M}\int_0^\infty \prod_k\Big[\Big(\prod_{i=1} x_{0i}\Big)^{a_{0k}(\beta-1)} a_{0k}^{\delta_0-1}(1-a_{0k})^{\lambda_0-1}\Big]\beta^{\sum_k a_{0k}n_{0k}-1} \times\Gamma\Big(\sum_k a_{0k}n_{0k}\Big)\Big(\sum_k a_{0k}x_{0n_{0k}}^{\beta}\Big)^{-\sum_k a_{0k}n_{0k}} d\beta d\mathbf{a}_0. \tag{3.4}$$

Since

$$\Big(\sum_{k=1}^{M} a_{0k}x_{0n_{0k}}^{\beta}\Big)^{-\sum_k a_{0k}n_{0k}} \le \prod_{k=1}^{M}\Big(a_{0k}x_{0n_{0k}}^{\beta}\Big)^{-a_{0k}n_{0k}} \quad \text{and} \quad a_{0k}^{-a_{0k}n_{0k}} \le \exp\Big(\frac{n_{0k}}{e}\Big),$$

(3.4) is less than or equal to

$$\exp\Big\{\sum n_{0k}/e\Big\}\int_{[0,1]^M}\int_0^\infty \prod_k\Big[\Big(\prod_{i=1} x_{0i}\Big)^{-a_{0k}} a_{0k}^{\delta_0-1}(1-a_{0k})^{\lambda_0-1}\Big] \times\Gamma\Big(\sum_k a_{0k}n_{0k}\Big)\beta^{\sum_k a_{0k}n_{0k}-1}\exp\Big\{-\beta\sum_k a_{0k}\sum_i \log\frac{x_{0n_{0k}}}{x_{0i}}\Big\}d\beta d\mathbf{a}_0. \tag{3.5}$$

Integrating out β, (3.5) reduces to

$$\exp\Big\{2\sum \frac{n_{0k}}{e}\Big\} \int_{[0,1]^M} \prod_k \Big[\Big(\prod_{i=1} x_{0i}\Big)^{-a_{0k}} \Big(\sum_i \log \frac{x_{0n_{0k}}}{x_{0i}}\Big)^{-a_{0k}n_{0k}} \times a_{0k}^{\delta_0-1}(1-a_{0k})^{\lambda_0-1}\Big] \Gamma^2\Big(\sum_k a_{0k}n_{0k}\Big) d\mathbf{a}_0. \tag{3.6}$$

By the inequality

$$\Gamma\Big(\sum_k a_{0k}n_{0k}\Big) = \frac{\Gamma(\sum_k a_{0k}n_{0k}+1)}{\sum_k a_{0k}n_{0k}} \le \frac{\Gamma(\sum_k n_{0k}+1)}{\prod_k a_{0k}},$$

(3.6) is less than or equal to

$$\exp\Big\{2\sum \frac{n_{0k}}{e}\Big\}\Gamma^2\Big(\sum_k n_{0k}+1\Big) \prod_{k=1}^{M} \Big\{\int_0^1 \Big(\prod_{i=1} x_{0i}\Big)^{-a_{0k}} \times \Big(\sum_i \log \frac{x_{0n_{0k}}}{x_{0i}}\Big)^{-a_{0k}n_{0k}} a_{0k}^{\delta_0-3}\,(1-a_{0k})^{\lambda_0-1}\, da_{0k}\Big\}. \tag{3.7}$$

Since

$$\Big(\prod_i x_{0i}\Big)^{-a_{0k}} \le 1 + \prod_i x_{0i}$$

and

$$\Big(\sum_i \log \frac{x_{0n_{0k}}}{x_{0i}}\Big)^{-a_{0k}n_{0k}} \le 1 + \Big(\sum_i \log \frac{x_{0n_{0k}}}{x_{0i}}\Big)^{-n_{0k}},$$

(3.7) is less than or equal to

$$K \prod_{k=1}^{M} \Big\{\int_0^1 a_{0k}^{\delta_0-3}\,(1-a_{0k})^{\lambda_0-1}\, da_{0k}\Big\} < \infty,$$

where

$$K = \exp\Big\{2\sum \frac{n_{0k}}{e}\Big\}\Gamma^2\Big(\sum_k n_{0k}+1\Big) \prod_{k=1}^{M} \Big[\Big(1+\prod_i x_{0i}\Big) \times \Big\{1 + \Big(\sum_i \log\Big(\frac{x_{0n_{0k}}}{x_{0i}}\Big)\Big)^{-n_{0k}}\Big\}\Big].$$

This completes the proof. □

TABLE 4.1 *Posterior estimates of parameters for the power law process. The numbers in parentheses are standard errors of the estimates.*

(δ_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	$E(a_0 \mid D, D_0)$ (SE.)	Para.	Post.Mean (SE.)
$(5, 5)$	$(0.500, 0.151)$	0.011 (0.005)	β	0.699 (0.068)
			η	15.365 (13.193)
$(20, 20)$	$(0.500, 0.078)$	0.041 (0.009)	β	0.700 (0.069)
			η	15.179 (12.731)
$(30, 30)$	$(0.500, 0.064)$	0.060 (0.011)	β	0.700 (0.066)
			η	15.077 (12.467)
$(50, 1)$	$(0.980, 0.019)$	0.106 (0.015)	β	0.701 (0.066)
			η	14.879 (11.900)
$(100, 1)$	$(0.990, 0.010)$	0.212 (0.021)	β	0.704 (0.062)
			η	13.582 (10.916)

REMARK 3.2. *Similar extension of Theorem 3.2 can be also obtained. In this case the condition on the hyperparameters, (δ_0, λ_0) for the propriety is $(\delta_0 + \lambda_0)/\delta_0 < N_0$. It is weaker than that for a single dataset. This implies that more information is incorporated into the analysis when multiple historical datasets are available.*

4. NUMERICAL EXAMPLES

EXAMPLE 4.1. We analyze the data of Maguire, Pearson and Wynn (1952) on the intervals in days between coal-mining disasters. The data are recorded between 6 December 1875 and 29 May 1951. However, according to Jarrett (1979) it turned out that more data are available starting 15 March 1851. Thus, we regard the data before 1875 as the historical data with the size of 80, and use the rest of the data as the current data with the size of 107. These data have been extensively used to fit several models including nonhomogeneous Poisson processes (*cf.* Barnard (1953); Cox and Lewis (1966)). First, we compute the MLEs for each sub-dataset. The MLE of (β, η) for the historical data is $(\hat{\beta}_{his}, \hat{\eta}_{his}) = (1.039, 16.317)$. On the other hand, the MLE for the current data is $(\hat{\beta}_{cur}, \hat{\eta}_{cur}) = (0.699, 11.910)$. It seems that the frequency of failures is an increasing function based on the current data. However, the historical data behave time independent in terms of the MLE $\hat{\beta}_{his}$. Second, we compute the posterior means of the parameters (a_0, β, η) with the joint power prior in (3.1), which requires a two-dimensional numerical integration. This can be done using trapezoidal rules.

TABLE 4.2 *Simulation results for the power law process with the same βs. The numbers in parentheses are standard errors of the estimates.*

		$\beta = (0.5, 0.5)$		
(δ_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	$E(a_0 \mid D, D_0)$	Para.	Post. Mean
(10,10)	(0.500,0.109)	0.633 (0.105)	β	0.512 (0.062)
			η	0.108 (0.021)
(20,20)	(0.500,0.078)	0.585 (0.065)	β	0.514 (0.071)
			η	0.109 (0.026)
(30,30)	(0.500,0.064)	0.560 (0.054)	β	0.517 (0.069)
			η	0.109 (0.025)
(50,10)	(0.833,0.048)	0.862 (0.020)	β	0.515 (0.069)
			η	0.108 (0.026)
(100,10)	(0.909,0.027)	0.919 (0.008)	β	0.505 (0.066)
			η	0.105 (0.024)
		$\beta = (2.0, 2.0)$		
(δ_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	$E(a_0 \mid D, D_0)$	Para.	Post. Mean
(10,10)	(0.500,0.109)	0.862 (0.007)	β	2.086 (0.268)
			η	0.108 (0.025)
(20,20)	(0.500,0.078)	0.768 (0.009)	β	2.081 (0.270)
			η	0.109 (0.023)
(30,30)	(0.500,0.064)	0.708 (0.008)	β	2.073 (0.264)
			η	0.111 (0.025)
(50,10)	(0.833,0.048)	0.913 (0.002)	β	2.065 (0.299)
			η	0.108 (0.026)
(100,10)	(0.909,0.027)	0.940 (0.001)	β	2.112 (0.290)
			η	0.112 (0.024)

Just as in Ibrahim and Chen (2000), we use 5 different choices of hyperparameters (δ_0, λ_0) of the beta prior distribution. Numerical results are reported in Table 4.1 We see that as the posterior mean of a_0 increases, so does the posterior mean of β. This is congruent with what we expect from the data. Moreover, the change is not so severe. This implies that the posterior mean of β is quite robust in terms of change of the hyperparameters.

EXAMPLE 4.2. We perform a simulation study. We fix $\eta = 0.1$ and use the same sample sizes 30 for both historical and current data. We compute the posterior means for various hyperparameters based on 200 replications. The estimates of parameters are reported in Table 4.2 and Table 4.3 In Table 4.2, we set the same values of β for the historical and the current data. In Table 4.3, we give the different values of β for two datasets. When the data are simulated with

TABLE 4.3 *Simulation results for the power law process with different βs. The numbers in parentheses are standard errors of the estimates.*

		$\beta = (0.5, 2.0)$		
(δ_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	$E(a_0 \mid D, D_0)$	Para.	Post. Mean
(10,10)	(0.500, 0.109)	0.559 (0.149)	β	0.701 (0.079)
			η	0.086 (0.009)
(20,20)	(0.500, 0.078)	0.532 (0.076)	β	0.701 (0.096)
			η	0.087 (0.009)
(30,30)	(0.500, 0.064)	0.522 (0.052)	β	0.704 (0.088)
			η	0.087 (0.009)
(50,10)	(0.833, 0.048)	0.855 (0.023)	β	0.663 (0.087)
			η	0.090 (0.011)
(100,10)	(0.909,0.027)	0.917 (0.008)	β	0.635 (0.078)
			η	0.089 (0.012)
		$\beta = (2.0, 0.5)$		
(δ_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	$E(a_0 \mid D, D_0)$	Para.	Post. Mean
(10,10)	(0.500, 0.109)	0.809 (0.015)	β	0.631 (0.085)
			η	0.093 (0.015)
(20,20)	(0.500, 0.078)	0.704 (0.015)	β	0.605 (0.078)
			η	0.094 (0.016)
(30,30)	(0.500, 0.064)	0.648 (0.013)	β	0.602 (0.081)
			η	0.095 (0.016)
(50,10)	(0.833,0.048)	0.896 (0.004)	β	0.627 (0.084)
			η	0.092 (0.014)
(100,10)	(0.909,0.027)	0.932 (0.002)	β	0.628 (0.085)
			η	0.095 (0.014)

the same βs, the posterior means of β are fairly stable regardless of the posterior means of a_0. Further, the estimates of β and η are quite close to corresponding true values. Based on the results in Table 4.3, it is consistent in the sense that the posterior mean of β shrinks toward the value of historical data as the posterior mean of a_0 increases. Even though we do not report several results for different sample sizes, similar results are obtained. That is, the sample size of the current data is large compared to the historical data, the impact of a_0 on β is not so big.

5. CONCLUDING REMARKS

In Bayesian analysis it is practical and desirable to use a power prior when historical data are available. It may be useful in many applications including model selection, hypothesis testing, and clinical trials. It also seems to be useful in making inference for truncated failure data. We proposed two different joint power priors in the power law process. We have showed that these power priors are proper under mild conditions. Our computational results turned out to be fairly consistent and robust both for real data and simulated datasets.

REFERENCES

BAR-LEV, S., LAVI, I. AND REISER, B., (1992). "Byesian inference for the power law process", *Annals of the Institute of Statistical Mathematics*, **44**(4), 623–639.

BARNARD, G. A., (1953). "Time intervals between accidents-a note on Maguire, Pearson and Wynn's paper", *Biometrika*, **40**, 212–213.

CHEN, M.-H., IBRAHIM, J. G. AND SHAO, Q.-M., (2000). "Power prior distributions for generalized linear models", *Journal of the Statistical Planning and Inference*, **84**, 121–137.

CHEN, M.-H., IBRAHIM, J. G., SHAO, Q.-M. AND WEISS, R. E., (1999) . "Prior elicitation for model selection and estimation in generalized linear mixed models", *Journal of the Statistical Planning and Inference*, **111**, 57–76.

COX, D. R. AND LEWIS, P. A. W., (1966). *The Statistical Analysis of Series of Events*, (London: Methuen).

CROW, L. H., (1974). "Reliability analysis of complex repairable systems, reliability and biometry, Proschan, F. and Serfling, R. J., Eds", *SIAM, Philadelphia*, 379–410.

DUANE, J. T., (1964). "Learning curve approach to reliability monitoring", *IEEE Transaction on Aerospace*, **2**, 563–566.

GUIDA, M., CALABRIA, R. AND PULCINI, G., (1989). "Bayesian inference for a nonhomogeneous Poisson process with power intensity law", *IEEE Transaction on Reliability*, **38**(5), 603–609.

IBRAHIM, J. G. AND CHEN, M.-H., (2000), "Power prior distributions for regression models", *Statistical Science*, **15**, 46–60.

IBRAHIM, J. G., CHEN, M.-H. AND RYAN, L. M., (2000). "Bayesian variable selection for time series count data", *Statistical Sinica*, **10**, 971–987.

JARRETT, R. G., (1979). "A note on the intervals between coal-mining disasters", *Biometrika*, **66**, 191–193.

KIM, S. W. AND SUN, D., (2000). "Intrinsic priors for model selection using an encompassing model with applications to censored failure time data", *Lifetime Data Analysis*, **6**, 251–269.

KIM, K., KIM, S. W. AND KIM, H., (2003). "Intrinsic Bayes factors for model selection in nonhomogeneous Poisson processes", *Far East Journal of Theoretical Statistics*, **11**(1), 15–30.

KYPARISIS, J. AND SINGPURWALLA, N., (1985). "Bayesian inference for the Weibull process with applications to assessing software reliability growth and predicting software failures", *Computer Science and Statistics: Proceeding of the 16th. Symposium on the interface of computer science and statistics*, 57–64.

LINGHAM, R. T. AND SIVAGANESAN, S., (1997). "Testing hypotheses about the power law process under failure truncation using intrinsic Bayes factors", *Annals of the Institute of Statistical Mathematics.*, **49**, 693–710.

MAGUIR, B. A., PEARSON, E. S. AND WYNN, A. H. A., (1952). "The time intervals between in dustrial accidents", *Biometrika*, **39**, 168–180.

ZELLNER, A., (1988). "Optimal information processing and Bayes's theorem (with discussion)", *The American Statistician*, **42**, 278–284.