Bayesian Nonparametric Inference for Nonhomogeneous Poisson Processes

LYNN KUO and SUJIT GHOSH
University of Connecticut and North Carolina State University
Institute of Statistics Mimeo Series No. 2530
March 2001

Abstract

Several classes of nonparametric priors are employed to model the rate of occurrence of failures of the nonhomogeneous Poisson process. The classes include the gamma process prior, the beta process prior, and the extended gamma process prior. We derive the posterior distribution for each process. Sampling-based methods are developed for Bayesian inference. Numerical comparisons among the three classes are performed on a real software failure data set.

Key words and phrases: Beta process prior, extended gamma process prior, gamma process prior, infinitely divisible distribution, Lévy process, sampling-based approach.

Abbreviated Title: Bayesian Nonparametric Poisson Process
1. Introduction

Nonhomogeneous Poisson point processes (NHPP) have been used extensively in modeling the number of failures in repairable systems and in software testing. Most Bayesian inference for NHPP assumes a parametric model for the rate of occurrence of failures (ROCOF) (also known as the intensity function), and proceeds by putting priors on the unknown parameters (cf. Kuo and Yang, 1996). The parametric models have the advantages of being simple to understand and to fit. However, it is often difficult to justify a specific parametric family. In order to overcome this difficulty, we take a nonparametric Bayesian approach in this paper. The nonparametric Bayesian approach allows us to incorporate our prior belief on the ROCOF. Moreover, it is more robust than the parametric approach against the misspecification of the distribution.

We model the ROCOF to be an unknown sample path of a stochastic process that potentially can have an infinite number of parameters. However, we assume specific prior distributions for the increments of the sample path. In particular, we assume two types of priors: one type is on the space of the cumulative ROCOFs and includes the gamma processes and the beta processes; the other is on the space of monotone ROCOFs and includes the gamma, beta, and the extended gamma processes.

All the processes have independent increments (or decrements). The gamma process assumes gamma distributions for the increments. The beta process proposed by Hjort (1990) assumes that the infinitesimal increments have beta distributions. The gamma and beta process priors for the cumulative ROCOF are flexible in the sense that they can be used to model non-monotone ROCOF’s. However, when the ROCOF is known to be monotone, for example, a decreasing function of time in the “happy” system or an increasing function in the “sad” system, the gamma and beta process priors on the cumulative ROCOF may be too general to be useful. In these cases, we can construct gamma processes, beta processes, and extended gamma processes (Dykstra and Laud, 1981) directly on the space of monotone ROCOFs, and not based on the cumulative ROCOFs.

On making Bayesian inference, we not only discuss how to obtain the posterior distributions analytically, but also develop sampling-based methods to fit the models. The sampling-based approach has the advantages that any features of the posterior distributions, for example, mean, variance, quantile, credible interval, and histogram can be obtained from
the random variates simulated from the posterior distribution. For each of the processes, we present two versions of the sampling-based algorithm. One version is what we call the "partitioning approximation" method where the updating is done by dividing the time interval into a large number of subintervals. The other version, called the "Lévy process generation", makes use of the Lévy representation in the posterior (or in the prior) and develop sampling techniques for generating random variates from the posterior distributions. We compare both versions for a particular data set. If the number of partitions is large enough in the "partitioning approximation" version, the difference between the two versions is negligible. Most readers may prefer the "partitioning approximation" version because it is easy to understand and easy to program.

For the gamma and the beta process priors on the cumulative ROCOFs, both algorithms sample replications of the random variates from the posterior distributions directly. In fact, for the gamma process prior that is a natural conjugate prior for the likelihood, neither version is needed in practice, because we can obtain the exact analytical expression for the posterior distribution. For the extended gamma process on the ROCOFs, both the Lévy process generation and the partitioning approximation versions use Markov chain Monte Carlo (MCMC) methods (cf. Gelfand and Smith 1990, and Tanner and Wong, 1987). Data augmentation with a class of multinomially distributed latent variables facilitates the specification of the conditional densities used in the MCMC algorithm. Therefore, we use the data augmentation technique in both versions.

The Bayesian computation schemes presented in this article relies on previous works in Damien, Laud, and Smith (1995, 1996), Laud, Smith and Damien (1996), and Walker and Damien (1998). However, they consider the case where the failure times subject to censoring are i.i.d.; we consider a different likelihood where i.i.d. assumptions do not apply. Therefore, we need to develop new algorithms for each of the nonparametric priors mentioned here to accommodate for the different likelihoods.

In addition to Bayesian model fitting, we also consider issues of model adequacy and model selection. We use model in a broad sense to include both the likelihood and the prior. The prediction intervals for the forecast (future failure times) can be constructed from the predictive density of the forecast given the past data by means of the sampling-based method. The model is judged to be adequate if at least $1 - \alpha$ of the actually observed failure epochs are contained in the $1 - \alpha$ prediction intervals. For model selection, we use a predictive
likelihood that selects the model with the largest predictive power.

In Section 2 we discuss how to construct the likelihood of the NHPP. Section 3 formulates the nonparametric priors. In Section 4 we obtain the analytical forms or their approximations for the posterior distributions. Section 5 exhibits the sampling-based approaches to simulate variates from the posterior distributions. Section 6 describes predictive inference for the future survival time and for the future mean time between failures. Section 7 discusses methods for checking model adequacy and for model selection. A numerical example is given in Section 8.

2. The Likelihood of NHPP

Let \( M(t) \) denote the number of failures in \((0, t]\). We assume that \( M(t) \) is a NHPP with mean function \( \Lambda(t) = E[M(t)] \). That is, the number of failures in disjoint intervals, say \((t_{i-1}, t_i)\) with \( t_{i-1} < t_i \), are independently distributed as Poisson random variables with mean \( \Lambda(t_i) - \Lambda(t_{i-1}) \). In our definition of the NHPP, we also consider “intervals” of the form \((t_{i-1}, t_i)\), \( \{t_i\} \), \((t_i, t_{i+1})\) etc. Let \( \lambda(t) = \frac{d}{dt} \Lambda(t) \) denote the RCOF of the NHPP, also called the intensity function; and \( \Lambda(t) \) the corresponding cumulative RCOF function. If \( \Lambda(t) \) has a jump at \( t_i \), then we define \( \lambda(t_i) = [\Lambda(t_i) - \Lambda^-(t_i)]\delta^\dagger(t_i) \), where \( \Lambda^-(t_i) = \lim_{s \to t_i^-} \Lambda(s) \), the left-hand limit, and \( \delta^\dagger \) is the Dirac delta function.

Given the time-truncated model (i.e., testing until time \( t_U \)), the ordered epochs of the observed \( n \) failure times are denoted by \( x_1, \ldots, x_n \), where \( x_1 < x_2 < \cdots < x_n \). We define the data set \( D_{t_U} \) to be \( \{n, x_1, x_2, \ldots, x_n; t_U\} \). Then the probability of observing no failures in \((0, x_1)\), one failure at \( x_1 \), no failures in \((x_1, x_2)\), and so on up to no failures in \((x_n, t_U)\) is given by

\[
L_{NHPP}(\Lambda|D_{t_U}) = \left( \prod_{i=1}^{n} e^{-[\Lambda^{-}(x_i) - \Lambda^{-}(x_{i-1})]} [\Lambda(x_i) - \Lambda^{-}(x_{i-1})] e^{-[\Lambda(x_{i-1}) - \Lambda^{-}(x_{i-1})]} \right) e^{-[\Lambda(t_U) - \Lambda(x_n)]},
\]

where \( x_0 = 0 \).

If we allow the possibilities of ties in the failure epochs, then we generalize the likelihood in (1) to

\[
L_{NHPPG}(\Lambda|D_{t_U}) = \left( \prod_{i=1}^{n} e^{-[\Lambda^{-}(x_i) - \Lambda^{-}(x_{i-1})]} [\Lambda(x_i) - \Lambda^{-}(x_{i-1})]^{d_i} e^{-[\Lambda(x_{i-1}) - \Lambda^{-}(x_{i-1})]} \right) e^{-[\Lambda(t_U) - \Lambda(x_n)]},
\]
where $d_i$ denotes the number of multiplicities of the failures observed at $x_i$. So the total number of failures observed in $(0, t_U]$ is $\sum_{i=1}^{n} d_i$.

For the failure truncated model (observed until the $n^{th}$ failure), we define the data set $D_{x_n}$ to be $\{x_1, x_2, \ldots, x_n\}$. Then, a similar expression to (1) can be obtained with $t_U$ replaced by $x_n$ for the likelihood.

All likelihood functions here suggest that if the prior on the cumulative ROCOF $\Lambda(t)$ has independent increments then the posterior of $\Lambda$ also has independent increments. In particular if the increments have gamma distributions in the prior process, then those increments also have gamma distributions in the posterior process.

In the following section, we define two general classes of nonparametric priors. The first class is constructed on the space of cumulative ROCOFs. In this class we focus on the gamma and the beta processes. The second class is constructed on the space of monotone ROCOFs and we focus on the extended gamma process.

### 3. Nonparametric Prior Processes

#### 3.1 Priors for the Cumulative ROCOF

Let $\Lambda_t = \Lambda(t)$ be a process with independent increments, non-decreasing almost surely, right continuous almost surely, $\lim_{t \to -\infty} \Lambda_t = 0$ almost surely, and $\lim_{t \to \infty} \Lambda_t = \infty$ almost surely. We allow the possibility that $\Lambda_t = \infty$ with positive probability for finite $t$. For such a process $\Lambda_t$, there exist at most countably many fixed points of discontinuity, say, at $t_1, t_2, \ldots$ with jumps of random size to be $S_1, S_2, \ldots$ with densities $f_1, f_2, \ldots$. If we remove these fixed jumps of discontinuity from the process $\Lambda_t$, then

$$Z_t = \Lambda_t - \sum_j S_j I_{[t_j, \infty)}(t)$$

is a process with independent increments with no fixed points of discontinuity. Then the moment generating function of such a process has the following Lévy representation

$$M_t(\theta) = E e^{-\theta Z_t} = \exp \left\{ -\theta b(t) + \int_0^{\infty} (e^{-\theta z} - 1) dN_t(z) \right\},$$

where $b$ is a non-decreasing and continuous function with $b(t) \to 0$ as $t \to -\infty$ and $N_t$ is a continuous Lévy measure satisfying some regularity conditions (cf. Ferguson, 1974, Ferguson and Phadia, 1978).
In the following construction, we assume \( b(t) \equiv 0 \). When the increments of the process \( \Lambda_t \) have gamma distributions, we call this process the gamma process. We use the notation \( \Gamma(\alpha, \beta) \) to denote a gamma distribution with mean \( \alpha/\beta \) throughout the paper. Let us assume \( \Lambda_t \) is a gamma process with \( \Lambda_t \sim \Gamma(c\Lambda_0(t), c) \). So the prior guess for the mean of \( \Lambda_t \) is \( E[\Lambda_t] = \Lambda_0(t) \) and the prior variance of \( \Lambda_t \) is \( \text{Var}(\Lambda_t) = \Lambda_0(t)/c \). So we choose \( \Lambda_0(t) \) to represent our prior belief on the mean cumulative ROCOF and \( c \) to represent the prior precision; the larger the \( c \) is, the more faith we have on our prior belief. If we do not have much prior knowledge, we can use a small value for \( c \) to depict our prior ignorance. The Lévy measure for this gamma process is given by

\[
dN_t(z) = c\Lambda_0(t)e^{-c}\frac{z^{-1}dz}{z(1-e^{-c})}.
\]

Similarly, we can consider the simple homogeneous process with the following Lévy measure,

\[
dN_t(z) = c\Lambda_0(t)e^{-c}\frac{(1-e^{-c})^{-1}dz}{z(1-e^{-c})},
\]

and a “Dirichlet” process (Ferguson, 1973), where a Dirichlet process \((\mathcal{D}(\alpha))\) is actually constructed on \( F(t) = 1 - e^{-\Lambda_t} \), with the following Lévy measure,

\[
dN_t(z) = \frac{e^{-\alpha(t)}(e^\alpha(t) - 1)}{z(1-e^{-c})}dz.
\]

Next we define the beta process proposed by Hjort (1990). Let \( A_0 \) denote our prior guess on the cumulative ROCOF with fixed points of jumps at \( t_1, t_2, \ldots, t_m \). Let \( c(\cdot) \) be a piecewise continuous, non-negative function on \([0, \infty)\). A beta process with parameters \( c(\cdot), A_0(\cdot) \), denoted by \( \Lambda \sim \text{beta}(c(\cdot), A_0(\cdot)) \), is a process with independent increments with the following Lévy representation,

\[
Ee^{-\theta\Lambda_t} = \left( \prod_{j: t_j \leq t} E(\exp(-\theta S_j)) \right) \exp \left\{ \int_0^t (e^{-\theta z} - 1)dN_t(z) \right\}, \tag{3}
\]

where \( S_j = \Lambda\{t_j\} \sim \text{beta}(c(t_j)A_0\{t_j\}, c(t_j)(1-A_0\{t_j\})) \), and the Lévy measure is given by

\[
dN_t(z) = \left[ \int_0^t c(s)z^{-1}(1-z)^{c(s)-1}dA_0(s) \right]dz
\]

for \( t \geq 0 \) and \( 0 < z < 1 \) with \( A_0(t) = A_0(t) - \sum_{t_j \leq t} A_0\{t_j\} \). Note \( \Lambda\{t_j\} \) denotes the size of the increment of \( \Lambda \) at \( t_j \). The notation \( A_0\{t_j\} \) is defined similarly. The mean and variance of the beta process are \( E\Lambda_t = A_0(t) \) and \( \text{Var}(\Lambda_t) = \int_0^t (c(s) + 1)^{-1}dA_0(s) \) for a continuous \( A_0 \). Therefore, we choose \( A_0(t) \) to reflect our prior belief on the mean of the cumulative ROCOF and \( c(\cdot) \) to satisfy the variance equation.
3.2 Priors for the Monotone ROCOF

When the ROCOF is known to be monotone, then modeling the ROCOF using a monotone function would be more effective than the priors given in the previous subsection (as the monotonicity of the ROCOF cannot be guaranteed). Instead of using the gamma or the beta processes for the cumulative ROCOF, we can directly use gamma, beta or extended gamma processes (cf. Dykstra and Laud, 1981, and Lo, 1982) based on the ROCOF itself. As the gamma process is a special case of the extended gamma process, we will only discuss the extended gamma process. First, we consider the “sad” system where the ROCOF, \( \lambda(t) \), is a non-decreasing function of \( t \). In order to formulate an extended gamma process, consider the gamma process \( Z(t) \sim \Gamma(\alpha(t), 1) \). Then we say \( \lambda(t) \) has an extended gamma process, denoted as \( G(\alpha(\cdot), \beta(\cdot)) \), if

\[
\lambda(t) = \int_{0,t} \beta(s) dZ(s),
\]

where \( \beta(s) \geq 0 \) is a functional parameter of the process and the integration is with respect to the sample path of the \( Z(t) \) process. It is straightforward (see Theorem 2.1 of Laud et al., 1996) to show that the moment generating function of such process has the following Lévy representation:

\[
M_t(\theta) = E e^{-\theta \lambda_t} = \exp \left\{ \int_0^\infty (e^{-\theta z} - 1) dN_t(z) \right\},
\]

where

\[
dN_t(z) = \frac{dz}{z} \int_{0,t} e^{-\theta z} d\alpha(s).
\]

Next we discuss some methods of eliciting the prior parameters \( \alpha(\cdot) \) and \( \beta(\cdot) \) to reflect our prior belief on the mean and variance of the cumulative ROCOF. Note we consider prior formulation in terms of the cumulative ROCOF as done in subsection 3.1, not ROCOF as considered by Dykstra and Laud. Therefore, we need to develop new guidelines for choosing prior parameters \( \alpha(\cdot) \) and \( \beta(\cdot) \). Observe that \( \Lambda(t) = \int_{[0,t]} \lambda(s) ds = \int_{[0,t]} (t - u) \beta(u) dZ(u) \) by the Fubini theorem. Let \( \mu_0(t) = E \Lambda(t) \) and \( \sigma_0^2(t) = Var(\Lambda(t)) \) represent our prior belief on the mean and the variance of the cumulative ROCOF. Then we find \( \alpha(t) \) and \( \beta(t) \) to satisfy the following equations:

\[
\mu_0(t) = \int_{[0,t]} (t - u) \beta(u) d\alpha(u),
\]

and

\[
\sigma_0^2(t) = \int_{[0,t]} (t - u)^2 \beta^2(u) d\alpha(u).
\]
Using the Leibniz’s rule for differentiation, we can show \( \frac{d^2 \mu_0(t)}{dt^2} = \beta(t)d\alpha(t)/dt \) and \( \frac{d^3 \sigma_0^2(t)}{dt^3} = 2\beta^2(t)d\alpha(t)/dt \). Therefore, we can choose \( \alpha(t) \) and \( \beta(t) \) to satisfy

\[
\beta(t) = \frac{d^3 \sigma_0^2(t)}{dt^3} / \left( \frac{d^2 \mu_0(t)}{dt^2} \right)
\]

and

\[
\frac{d\alpha(t)}{dt} = 2 \left[ \frac{d^2 \mu_0(t)}{dt^2} \right]^2 / \frac{d^3 \sigma_0^2(t)}{dt^3}.
\]

For the “happy” system where the ROCOF is a non-increasing function of \( t \), we model the ROCOF modifying the previous extended gamma process, as

\[
\lambda(t) = \int_{[t,t]} \beta(s)dZ(s).
\]

It is the same weighted gamma process as defined in the “sad” system except with a different domain of integration.

To elicit the prior parameters we now define,

\[
\overline{\lambda}(t) = \int_{[t,t]} \lambda(s)d(s) = \int_{[t,t]} (u-t)\beta(u)dZ(u).
\]

Further let,

\[
\overline{\mu}_0(t) = E\overline{\lambda}(t) = \int_{[t,t]} (u-t)\beta(u)d\alpha(u),
\]

and

\[
\overline{\sigma}_0^2(t) = \text{Var}\overline{\lambda}(t) = \int_{[t,t]} (u-t)^2\beta^2(u)d\alpha(u).
\]

Then we can solve the two equations similarly as before to obtain

\[
\beta(t) = \frac{d^3 \overline{\sigma}_0^2(t)}{dt^3} / \left( 2 \frac{d^2 \overline{\mu}_0(t)}{dt^2} \right)
\]

and

\[
\frac{d\alpha(t)}{dt} = -2 \left[ \frac{d^2 \overline{\mu}_0(t)}{dt^2} \right]^2 / \frac{d^3 \overline{\sigma}_0^2(t)}{dt^3}.
\]

The extended gamma process is suitable for monotone ROCOF. In fact, we can extend it to U-shaped ROCOF or unimodal symmetric ROCOF by considering

\[
\lambda(t) = \int k(t,s)\beta(s)dZ(s),
\]

with kernel \( k(t,s) = I_{\{|t-s| \leq a\}} \) for a known constant \( a \) (cf. Lo and Weng, 1989).
4. The Posterior Distribution

We use the general likelihood in (2) that allows for ties in the failure times to derive the posterior distribution for each of the nonparametric priors. In particular, we consider the gamma, beta, and extended gamma process priors as introduced in the previous section. For the other likelihood in (1), we just set \( d_i \) to be 1 in the following discussions. Similar results can be derived for the failure truncated likelihood. Although the results in this section are quite straightforward, they have not been presented before to the best of our knowledge.

4.1 Posterior for the Gamma Process Prior

Theorem 1. Let \( \Lambda_t \) be a sample path chosen from the gamma process prior with \( \Lambda_t \sim \Gamma(c \Lambda_0(t), c) \). Then the posterior distribution of \( \Lambda_t \) given the data \( D_{t_U} \) is also a gamma process with \( \Lambda_t | D_{t_U} \sim \Gamma(c \Lambda_0(t) + \sum_{i|t_i \leq t} d_i, 1 + c) \). Moreover, we have the following results for each \( i, i = 1, \ldots, n; \)

(a) If \( x_i \) is a prior fixed point of discontinuity, say \( t_j \), then \( x_i \) is also a point of discontinuity in the posterior with the size of the jump distributed as

\[
f^*_j(z|\text{data}) = \frac{z^{d_j}e^{-z}f_j(z)}{\int_0^{\infty} z^{d_j}e^{-z}f_j(z)dz}.
\]

In particular, if \( f_j \) (the prior density for the size of the jump at \( t_j \)) is a gamma density with parameters \( c \Lambda_0 \{x_i\} \) and \( c \), then \( f^*_j \) is also a gamma density with parameters \( c \Lambda_0 \{x_i\} + d_j \) and \( 1 + c \).

(b) If \( x_i \) is not a prior fixed point of discontinuity, then \( x_i \) is a fixed point of discontinuity in the posterior and the size of the jump at \( x_i \) is distributed as \( \Gamma(d_i, 1 + c) \).

(c) The posterior distribution of an increment \( (s, t] \) between \( x_i \) and \( x_{i+1} \) (or \( x_{i-1} \) and \( x_i \)) is distributed as \( \Gamma(c \Lambda_0(s, t), 1 + c) \), where \( x_0 = 0, x_{n+1} = t_U \), and \( \Lambda_0(s, t) = \Lambda_0(t) - \Lambda_0(s) \).

Proof. Part (a) is straightforward by examining the likelihood and prior. Part (b) follows from evaluating the posterior density of the increment of \( (x_i - \epsilon, x_i) \), then taking the limit as \( \epsilon \to 0 \). Part (c) follows because the posterior density of the increment is obtained from the prior density multiplying by \( e^{-(\Lambda(t) - \Lambda(s))} \) and re-normalizing. Finally, the result that the posterior of \( \Lambda \) is also a gamma process follows because the likelihood in (2) factors into independent increments of the \( \Lambda \) process, each increment having a posterior gamma distribution with the same scale parameter \( 1 + c \) and a shape parameter given by (a)-(c).
Next let us assume there are no fixed points of discontinuity in the prior; then we can derive the Lévy representation for the posterior process

\[ M_t(\theta|D_{t_U}) = \mathbb{E}(\exp(-\theta \Lambda(t))|D_{t_U}) = \left[ \prod_{i:x_i \leq t} \mathbb{E}(\exp(-\theta W_i)) \right] \exp \left\{ -\int_0^\infty (\exp(-\theta z) - 1) dN^*_i(z|D_{t_U}) \right\}, \]

where \( W_i \sim \Gamma(d_i, 1+e) \) and

\[ dN^*_i(z|D_{t_U}) = (c+1)\Lambda_0(t) e^{-(c+1)z} z^{-1} dz. \]

The above Lévy measure can be derived from the observation that \( M_t(\theta|D_{t_U}) = M_t(\theta + 1)/M_t(1) \) (because the density of the continuous part of the increment \( z \) is updated by multiplying it by \( e^{-z} \) and re-normalizing) or by direct evaluation of the posterior gamma densities.

4.2 Posterior for the Beta Process Prior

In this subsection, we first show how the posterior of \( \Lambda \) is obtained for \( n = 1 \). Then we generalize it to an arbitrary \( n \). Moreover, we show our results for the following general setup.

Let us assume \( \Lambda \) have fixed points of discontinuity at \( M = \{t_1, \ldots, t_m\} \) with jumps \( S_j = \Lambda\{t_j\} \) that have densities \( f_j(z) \) supported in \((0,1)\). The process \( \Lambda_c(t) = \Lambda(t) - \sum_{j \leq t} S_j \) is free of fixed points of discontinuity and has Lévy representation

\[ E \exp\{-\theta \Lambda_c(t)\} = \exp \left\{ \int_0^t (e^{-\theta z} - 1) dN_i(z) \right\}, \]

with

\[ dN_i(z) = \begin{cases} \int_0^t a(z,s) dH(s) dz, & \text{for } t \geq 0 \text{ and } z \in (0,1) \\ 0 & \text{if } z \geq 1. \end{cases} \]

Observe that we have used \( M, \{f_j\}_{j=1}^m, a, \) and \( H \) to describe the process. Let us call it the Lévy process. It is more general than the beta process. In the following, we use the notation \( \ast \) to describe the corresponding posterior elements.

Theorem 2. Suppose we observe \( d \) failures at \( x_1 \) and no other failures between \( 0 \) and \( t_U \); then the posterior of \( \Lambda \) defined above is still a Lévy process. Moreover, we have:
(a) If $x_1$ is a prior fixed point of discontinuity, say $t_j$, then the set of points of discontinuity in the posterior $M^* = M$, and $H^* = H$,

$$a^*(z, s) = e^{-z}a(z, s) \approx (1 - z)a(z, s),$$

$$f^*_i(z) = \begin{cases} 
\text{const. } z^d e^{-z} f_j(z) & \text{for } l = j, \\
\text{const. } e^{-z} f_i(z) & \text{for } l \neq j.
\end{cases}$$

(b) If $x_1$ is not a prior fixed point of discontinuity, then $\Lambda$ is again Lévy with $M^* = M \cup \{x_1\}$, $H^* = H$, and

$$a^*(z, s) = (1 - z)a(z, s),$$

$$f^*_i(z) = \text{const. } e^{-z} f_i(z) \text{ for all } l = 1, \ldots, m,$$

while the new jump at $x_1$ has size $z$ with density

$$f^*_{x_1}(z) = \text{const. } z^d e^{-z} a(z, x_1) \text{ for } z \in (0, 1).$$

Proof: (a) Observe the likelihood in this case is

$$e^{-\Lambda^-(x_1)} [\Lambda(x_1) - \Lambda^-(x_1)]^{d_1} e^{-[\Lambda(x_1) - \Lambda^-(x_1)]} e^{-[\Lambda(t_{j-1}) - \Lambda(x_1)]}.$$

Note the first part of the likelihood can be decomposed to

$$e^{-\Lambda^-(x_1)} = e^{-\Lambda^-(t_1) - \cdots - (\Lambda^-(t_{j-1}) - \Lambda(t_{j-1})) - \cdots - (\Lambda^-(x_1) - \Lambda(t_{j-1}))},$$

similarly, we can obtain the independent increment decomposition for the last part $e^{-[\Lambda(t_{j-1}) - \Lambda(x_1)]}$. Therefore, $f^*_i$ follows immediately from the usual prior and likelihood updating. Having removed the points of discontinuity from the process, we see any increment $z$ in the prior is updated by multiplying by $e^{-z}$ from the likelihood. If we make these increments infinitesimally small, they can be approximated by $1 - z$. This leads to $a^*$.

(b) The proof is similar to (a), except we need to evaluate the posterior density of $\Lambda(x_1) - \Lambda(x_1 - \epsilon)$, then take the limit as $\epsilon \to 0$. □

Now, we can generalize the above theorem for arbitrary $n$. If the prior has no fixed points of discontinuity, then the posterior has $n$ fixed points of discontinuity with $M^* = \{x_1, \ldots, x_n\}$ with the size of the jump $s_i$ at $x_i$ distributed with density $f^*_i(z) = \text{const. } z^d e^{-z} a(z, x_i)$ for $i = 1, \ldots, n$. Moreover, we have $H^* = H$ and $a^*(z, s) = (1 - z)a(z, s)$ for the continuous part of the posterior sample path.
Now consider a special case of the above Lévy process with $S_t \sim \beta(c(t_j) A_0 \{t_j\}, c(t_j)(1-A_0 \{t_j\}))$, and $a(z, s) = c(s) z^{-1} (1-z)^{c(s)}$ and $H(s) = A_0 c(s)$. Then this reduces to the beta process defined in Subsection 3.1. Therefore, the posterior density of the size of the jump at $x_i$ is given by

$$f^*_i(z) = \text{const. } z^{d_i+1-2c(x_i)+v} e^{-z},$$

when $x_i$ is a prior fixed point of discontinuity; if not, then the posterior jump at $x_i$ is given by

$$f^*_i(z) = \text{const. } z^{d_i-1} (1-z)^{c(x_i)-1} e^{-z}.$$  \hspace{1cm} (5)

Having removed the jumps at $\{x_i\}_{i=1}^n$, then the continuous part of the posterior sample path is a beta process with sample size $c(\cdot) + 1$. So the continuous part has the Lévy measure

$$dN^*_i(z) = \left[ \int_0^t (c(s) + 1) z^{-1} (1-z)^{d_i} e^{-z} \right] ds.$$  \hspace{1cm} (6)

In general, the posterior process is not a beta process anymore because of the factor $e^{-z}$ in the posterior densities for the jumps. However, if we assume the prior has no fixed points of discontinuity and approximate (5) by $f^*_i(z) = \text{const. } z^{d_i-1} (1-z)^{c(x_i)-1}$, then we can show the posterior is approximately a beta process with $c^*(\cdot) = c(\cdot) + dK(\cdot) + 1$ and $A^*_0 = \int_0^t [c(s) dA_0(s) + dK(s)]/[c(s) + dK(s) + 1]$, where $K$ is a discrete measure with $dK$ assigning mass $d_i$ at $x_i$. So the Lévy measure for the posterior process is

$$dN^*_i(z) = \left[ \int_0^t (c(s) + dK(s) + 1) z^{-1} (1-z)^{c(s)+dK(s)} e^{-z} \right] ds.$$  \hspace{1cm} (6)

### 4.3 Posterior for the Extended Gamma Process Prior

In this section, instead of modeling the cumulative ROCOF, we are modeling the monotone ROCOF with an extended gamma process prior. Therefore, it is more useful to consider the following approximation to the likelihood:

$$L_{NHPP}(\Lambda|D_{tu}) \approx \left( \prod_{i=1}^n e^{-[\Lambda(x_i) - \Lambda(x_i - \Delta)]} \right) e^{-[\Lambda(x_i) - \Lambda(x_i - \Delta)]} e^{-\int_0^{\tau(tu)} \lambda(t) dt}.$$  \hspace{1cm} (7)

Consequently, dividing the above equation by $\Delta^n$ and taking $\Delta \to 0$, we obtain the following likelihood

$$L(\lambda|D_{tu}) = \left( \prod_{i=1}^n \lambda^{d_i}(x_i) \right) e^{-\int_0^{\tau(tu)} \lambda(t) dt}.$$  \hspace{1cm} (8)
The posterior distribution for the extended gamma process with the likelihood in (8) with 
\( d_i = 1 \) for all \( i \) has been studied by Lo and Weng (1989). It is a mixture of extended gamma 
processes. In fact, they derive the result for a more general class than the extended gamma 
process. Next we apply their theorem 4.1 to the extended gamma process. The proof can 
be found in Sections 3 and 4 of Lo and Weng. In the following theorem, we use the notation 
\( \delta_u \) to denote a point mass measure of mass 1 at \( u \).

**Theorem 3.** Considering the likelihood in (8) with \( d_i = 1 \) and the prior of the extended 
gamma process for \( \lambda \), then the posterior of \( \lambda \) is a mixture of extended gamma processes:

\[
\lambda | D_{t_U} \sim \int G(\alpha + \sum \delta_{u_i}, \beta^*) dH(u|D_{t_U}),
\]

where the mixing measure \( H \) on the \( u = (u_1, \ldots, u_n) \) given \( D_{t_U} \) is given by

\[
dH(u|D_{t_U}) \propto \left( \prod_{i=1}^n \frac{\beta(u_i)I(u_i \leq x_i)}{1 + \beta(u_i)(x_i - u_i)^t} \right) \prod_{i=1}^n \left( \alpha + \sum_{1 \leq j \leq i-1} \delta_{u_j} \right) (du_i),
\]

where \( \beta^* \) can be defined by \( \beta^*(u) = \frac{\beta(u)}{1 + \beta(u)(t-u)^t} \) for an arbitrary \( t \) given in \( \lambda(t)|D_{t_U} \) in (9).

5. **Bayesian Computation**

5.1 **Computation for the Gamma Process Prior**

From Theorem 1, it follows that the posterior for \( \Lambda \) is distributed as 
\( \Gamma(c\Lambda_0(t) + \sum_{1 \leq i \leq t} d_i, 1 + c) \). Therefore, 
\( E(\Lambda|D_{t_U}) = [c\Lambda_0(t) + \sum_{1 \leq i \leq t} d_i]/[1 + c] \), and 
\( \text{Var}(\Lambda|D_{t_U}) = [c\Lambda_0(t) + \sum_{1 \leq i \leq t} d_i]/[1 + c]^2 \). The posterior 95% pointwise confidence interval for the cumulative ROCOF evaluated at \( t \) is computed from the 2.5 percentile and 97.5 percentile of the gamma distribution 
\( \Gamma(c\Lambda_0(t) + \sum_{1 \leq i \leq t} d_i, 1 + c) \).

The sampling-based method is not needed for this process unless we need plots of histograms of functionals of the \( \Lambda \) that have distributions not obvious to us. The sampling-based method consists of sampling 
\( Z_i \sim \Gamma(c\Lambda_0(x_i) - c\Lambda_0(x_{i-1}), 1 + c) \) for the continuous 
part of the increment between \( x_{i-1} \) and \( x_i \) and sampling 
\( W_i \sim \Gamma(d_i, 1 + c) \) for the discrete 
jump at \( x_i \) for \( i = 1, \ldots, n \); finally for the last increment between \( x_n \) and \( t_U \), we sample 
\( Z_{n+1} \sim \Gamma(c\Lambda_0(t_U) - c\Lambda_0(x_n), 1 + c) \). Then we add up these increments to obtain variates for 
\( \Lambda(x_i) \); for example, it can be estimated by \( \sum_{j=1}^{i} (Z_j + W_j) \). Alternatively, we can sample the cumulative \( Z \) directly: 
\( V_i \sim \Gamma(c\Lambda_0(x_i), 1 + c) \). Then we use 
\( V_i + \sum_{j=1}^{i} W_j \) to estimate \( \Lambda(x_i) \).
We replicate this sampling process $R$ times for $R$ large to obtain simulated variates from the posterior distribution.

Next we describe how to generate increments in the posterior distribution using the Lévy representation in (4). Although this is not needed because the above method is easier, it can be generalized to other nonparametric priors where the posterior densities are given in terms of the Lévy representation, for example, the homogeneous process and the “Dirichlet” process. Using the Lévy representation in (4) and the algorithm given in section 4.3 in Damien et al. (1995), we generate a continuous part of the increment $z_i$ in $x_{i-1}$ and $x_i$ for each $i, i = 1, \ldots, n + 1$ ($x_0 = 0$ and $x_{n+1} = t_U$) by the following steps (a)-(d) and a discrete jump of size $w_i$ at $x_i$ by step (e).

(a) Generate a sample $z_j^*$ by the rejection method or the Metropolis-Hastings method of size $m$ from a density

$$f(z) \sim \exp\{- (c + 1)z\}/(1 + z).$$

(b) Evaluate the normalizing constant $C = \int_0^\infty \exp\{- (c + 1)z\}/(1 + z)dz$.

(c) Simulate $m$ Poisson variates $y_j^*, j = 1, \ldots, m$ with parameter $\mu_j = [\Lambda_0^-(x_i) - \Lambda_0(x_{i-1})]C(1 + z_j^*)/mz_j^*$.

(d) Let $z_i = \sum_{j=1}^m z_j^*y_j^*$.

(e) We generate the discrete increment $w_i$ at $x_i$ by $\Gamma(d_i, c + 1)$ for $i = 1, \ldots, n$.

We choose the size $m$ large, say $m=1000$. Then we replicate the steps (a)-(e) for a large number of times, say $R = 5000$, to obtain the posterior variates for inference.

The above steps allow us to approximately compute $\hat{\Lambda}^-(x_i) = E(\Lambda^-(x_i)|D_i)$ and $\hat{\Lambda}(x_i) = E(\Lambda(x_i)|D_i)$ by adding the increment in step (e) for all $i$. Then we can use the following interpolation formula for each $i = 0, \ldots, n$ to obtain the posterior mean for any $t$ between the failure epochs ($x_0 = 0$ and $x_{n+1} = t_U$):

$$\hat{\Lambda}(t) = E(\Lambda(t)|D_i) = \hat{\Lambda}(x_i) + \frac{\Lambda_0(t) - \Lambda_0(x_i)}{\Lambda_0^-(x_{i+1}) - \Lambda_0^-(x_i)}[\hat{\Lambda}^-(x_{i+1}) - \hat{\Lambda}(x_i)], \text{ for } x_i \leq t < x_{i+1}. \quad (11)$$

Now we can generalize the algorithm to other nonparametric priors. For the simple homogeneous process (cf. Ferguson and Phadia, 1979), we obtain the posterior Lévy measure
\[ dN^*_t(z|D_{t_U}) \propto \exp\{- (c+1)z\}/(1 - \exp\{-z\})dz. \] Therefore, we modify the above procedure with (10) replaced by

\[ f(z) \sim z\exp\{- (c+1)z\}/[(1 + z)(1 - \exp\{-z\})], \]

with the same replacement for the normalizing constant in step (b). We replace step (e) with (e') We generate the discrete increment \( w_i \) at \( x_i \) with a density \( f_{W_i}(w) \propto w^{d_i}\exp\{- (c+1)w\}/(1 - \exp\{-w\}). \)

Similarly, we can apply the same idea to the "Dirichlet" process. The posterior Lévy measure is given by

\[ dN^*_t(z|D_{t_U}) \propto \exp\{- (\alpha(R) + 1)z\} \exp\{- (\alpha(t)z) - 1\}/[z(1 - \exp\{-z\})dz. \]

Therefore, we modify (10) with the following replacement

\[ f(z) \sim \exp\{- (\alpha(R) + 1)z\} (1 - \exp\{-[\alpha^{-}(x_i) - \alpha(x_{i-1})]\})/[(1 + z)(1 - \exp\{-z\})], \]

and we modify (e) with (e') We generate the discrete increment \( w_i \) at \( x_i \) with \( \Gamma(d_i + 1, \alpha(R) + 1) \) distribution.

Next let us describe the “partitioning approximation” method for the gamma process. We divide our interval \((0, t_U)\) into a large number \( J \) of disjoint subintervals. The prior consists of independent gamma distributed increments with the \( j^{th} \) increment \( z_j = \Lambda(t_j) - \Lambda(t_{j-1}) \) having \( \Gamma(c[\Lambda_0(t_j) - \Lambda_0(t_{j-1})], c) \) distribution. Suppose the failed epochs occur at the subintervals \( i_1, i_2, \ldots, i_n \). Then the likelihood can be approximated by \( \prod_{j=1}^{J} z_j^{\#_j} e^{-z_j} \) with \( \#_j = d_i \) if \( j = i_i \) and 0 elsewhere. Then the posterior can be obtained by sampling replication of independent variates with the \( \Gamma(c[\Lambda_0(t_j) - \Lambda_0(t_{j-1})] + \#_j, 1 + c) \) distribution for the \( j^{th} \) subinterval for all \( j = 1, \ldots, J \).

5.2 Posterior simulation for the Beta Process Prior

Now let us assume the beta prior has no fixed point of discontinuity, that is the Lévy representation is given in (3) without the product term. To evaluate the posterior distribution, we can either adopt the Damien et al. (1996) algorithm or the Walker-Damien (1998) algorithm. Both algorithms sample replicates of the increments directly from the posterior distribution. We will describe the Damien et al. method first and then the Walker-Damien method.

The Damien et al. (1995, 1996) method consists of sampling a continuous part of the increment \( z_i \) for \( \Delta_i = (x_{i-1}, x_i) \), and a discrete jump \( w_i \) at \( x_i \):
(a) Generate a sample $z_j^*$ by the rejection method of size $m$ from a density
\[
  f(z) \sim z dN_{\Delta_i}^*(z),
\]  
where $dN_{\Delta_i}^*(z)$ is $dN^*$ given in (6) restricted to $\Delta_i$.

(b) Evaluate the normalizing constant $C = \int_0^1 z dN_{\Delta_i}^*(z) \approx \sum_{j=1}^m z_j^*$.

(c) Simulate $m$ Poisson variates $y_j^*$, $j = 1, \ldots, m$ with parameter $\mu_j = C/(m z_j^*)$.

(d) Let $z_i = \sum_{j=1}^m z_j^* y_j^*$.

(e) We generate the discrete increment $w_i$ at $x_i$ by (5). It is a density $\propto w^{d_i-1}(1-w)^{(x_i)-1} e^{-w}$.

We choose the size $m$ large, say $m = 1000$. Then we replicate the steps (a)-(e) a large number of times, say $R = 5000$, to obtain the posterior variates for inference.

Note the above step (a) is accomplished by first generating a random variate $s^*$ from the density $\propto c(y) dA_{\alpha,c}(y)/[c(y) + 1]$ restricted to $\Delta_i$ by the rejection method. (If $c(y)$ is monotonically decreasing, then the rejection ratio $f/g$ is bounded above by $1/[1 + 1/c(x_i-1)]$ with $f(y) = c(y) dA_{\alpha,c}(y)/[c(y) + 1]$ and $g(y) = dA_{\alpha,c}(y)$.) Then $z^* \sim \text{beta}(1, c(s^*) + 1)$ is a random variable from $z dN_{\Delta_i}^*$.

Next we describe our adaptation of the Walker-Damien algorithm. Observe from (6) the posterior “approximately beta process” can be split into two parts: the fixed points of discontinuity at $x_i$ for $i = 1, \ldots, n$ with jumps distributed as $\text{beta}(d_i, c(x_i) + 1)$; the continuous part with Lévy measure
\[
  dN_i^*(z) = z^{-1} dz \int_0^t (c(s) + dK(s) + 1)(1-z)^{c(s)+dK(s)} \frac{c(s) dA_{\alpha,c}(s)}{c(s) + dK(s) + 1}.
\]  
So the posterior is sampled by sampling the jump $W_i$ at $x_i$ from the $\text{beta}(d_i, c(x_i) + 1)$ distribution first. To be more precise, this size of the jump should be sampled from a density that is proportional to $w^{d_i-1}(1-w)^{(x_i)-1} \exp(-w)$. Then we sample the continuous part of the posterior using (13) and the Walker-Damien method. Detailed discussion of the Walker-Damien method is given in Appendix A.

The above methods allow us to update the continuous part of the posterior distribution for the intervals $(x_{i-1}, x_i)$, for $i = 1, \ldots, n$, and for $(x_n, t_U)$. If we are interested in plotting the posterior cumulative hazard for every $t$, we can either evaluate the posterior density at
the finer mesh points, that is, we can take a partition not data dependent where all \( \Delta_j \) are small and equal length or use an interpolation formula as in (11) with \( \Lambda_0 \) replaced by \( A_0 \) to obtain the posterior curve.

Next let us describe the “partitioning approximation” method for the beta process. We divide our interval \((0, t_U)\) into a large number \( J \) of subintervals. If \( J \) is sufficiently large, we approximate the prior by independent beta distributions \( \text{beta}(c(j)\alpha_j(j), c(j)(1 - \alpha_j(j))) \) for the increment \( z_j = \Lambda(t_j) - \Lambda(t_{j-1}) \), where \( \alpha_j(j) = A_0(t_j) - A_0(t_{j-1}) \). Suppose the failed epochs occur at the subintervals \( i_1, i_2, \ldots, i_n \). Then the likelihood can be approximated by \( \prod_{j=1}^{J} z_j^{\#_j} e^{-z_j} \) with \( \#_j = d_i \) if \( j = i_l \) and 0 elsewhere. Then the posterior can be obtained by sampling replication of independent variates with density \( \propto z_j^{\#_j + (j)\alpha_j(j) - 1}(1 - z_j)^{c(j)(1 - \alpha_j(j)) - 1} e^{-z_j} \) for the \( j^{th} \) subintervals that can be done by the rejection method. Alternatively, we sample independent beta variates from the \( \text{beta}(\#_j + c(j)\alpha_j(j), c(j)(1 - \alpha_j(j)) + 1) \) distribution.

5.3 Posterior simulation for the Extended Gamma Process Prior

Because the posterior as shown in Theorem 3 is a complex mixture of extended gamma processes, some Bayesian computation techniques are needed to evaluate the posterior.

We first describe an algorithm that makes use of the infinitely divisible representation of the process. This algorithm consists of first generating a sequence of infinitely divisible (ID) variates from the prior using their Lévy representations for a given partition by the Walker-Damien (1998) method. Then we use a data augmentation idea similar to Laud et al. (1996) to update these ID variates. We will describe the data augmentation technique first and then backtrack it to describe the Walker-Damien method.

Suppose we have a “sad” system. Let \( 0 \leq t_0 < t_1 < \cdots < t_U \) denote a finite partition of \([0, t_U]\). Let \( \delta_j = \lambda(t_j) - \lambda(t_{j-1}) \). Let \( \delta \) denote \( (\delta_1, \ldots, \delta_J) \) and \( f_{\delta_j} \) denote the prior density of the increment \( \delta_j \). Observe \( \delta_j \) are independent in the prior with extended gamma distribution that can be generated by the Walker-Damien algorithm. Let \( \Delta_{ij} \) denote the interval \((t_{i-1}, t_i)\). We assume the failed epochs \( x_1, \ldots, x_n \) occur at the end of the intervals \( \Delta_{i1}, \ldots, \Delta_{in} \). Then the posterior distribution of \( \delta \) is

\[
\delta | D_{T_U} \propto \left( \sum_{j=1}^{i_1} \delta_j \right)^{d_1} \left( \sum_{j=1}^{i_2} \delta_j \right)^{d_2} \cdots \left( \sum_{j=1}^{i_n} \delta_j \right)^{d_n} e^{-\sum_{j=1}^{J-1} \delta_j(t_U - (t_{j-1} + t_j)/2) \prod_{j=1}^{J} f_{\delta_j}.}
\]

Next, we introduce a class of latent variables that are multinomially distributed. Let \( +k = i_1 + \cdots + i_k \) for each \( k = 1, \ldots, n \). Consider \( m_k = (m_{k1}, \ldots, m_{k+k}) \) for each \( k \), where
\( k = 1, \ldots, n \). Then we generate independent multinomial vectors for \( k = 1, \ldots, n \) with \( m_k =\) multinomial\((d_k, p_k, \ldots, p_{k+h})\), where \( p_k = \delta_j/(\sum_{j=1}^{k} \delta_j)\). Observe the posterior density of \( \delta \) given \( m = (m_1, \ldots, m_n) \) is proportional to \( \prod_{j=1}^{n} f^*_{\delta_j} \), where

\[
  f^*_{\delta_j} \propto \delta_j^{\sum_{i=1}^{m} m_{ij}} e^{-((u-(t_{j-1}+t_j)/2))\delta_j} f_{\delta_j}.
\]

Note in this expression, we set \( m_{ij} = 0 \) for all \( j \geq i_j \) for each \( i, i = 1, \ldots, n \). Therefore, we generate independent \( \delta_j, j = 1, \ldots, J \) from the density that is proportional to \( f^*_{\delta_j} \). It is done by first generating \( \delta_j \) from the prior \( f_{\delta_j} \) by the Walker-Damien algorithm and applying the rejection algorithm with the gamma kernel as the envelop function. In summary, this algorithm consists of iteratively generating multinomial latent variates and then generating \( \delta_j \) from \( f^*_{\delta_j} \) independently.

Now we describe the Walker-Damien method. Observe the Lévy measure for the extended gamma process is given by \( dN_t(z) = z^{-1} \int_{[0,\ell]} \exp\{-z/\beta(s)\} d\alpha(s) dz \). Suppose we are interested in generating \( \delta_j \) from \( f_{\delta_j} \). The Walker-Damien’s method consists of approximating \( \delta_j \) by \( X_\ell \), that is, a finite sum of the jumps of a Poisson process with intensity measure \( dN(.) \). Let \( \Delta = (t_{j-1}, t_j) \) and \( g_\ell(z) \propto z^{-1} I(z > \ell) \int_{\Delta} \exp\{-z/\beta(s)\} d\alpha(s) \). We define \( C_\ell = \int_{0}^{\infty} \int_{0}^{\infty} z^{-1} \exp\{-z/\beta(s)\} I(z > \ell, s \in \Delta) dz d\alpha(s) \). Then Walker-Damien method consists of taking \( \nu \sim \text{Poisson}(C_\ell) \) and taking \( z_1, \ldots, z_\nu \) i.i.d. from \( g_\ell \). Then let \( X_\ell = \sum_{k=1}^{\nu} z_k \).

The detail in sampling \( z \) from \( g_\ell \) is given in Appendix B.

Given that \( x_1, x_2, \ldots, x_n \) are the ordered epochs of the failure times, we let \( \Delta_j = (x_{j-1}, x_j) \), and apply the previous algorithm for \( \Delta = \Delta_j \) for each \( j = 1, \ldots, n \).

On the “partitioning approximation” method, we let \( \Delta_j = t_j - t_{j-1} \) go to 0. That is, the number of divisions has to be extremely large. Observe, infinitesimally, \( \delta_j \) has gamma distribution with shape parameter \( \alpha(t_j) - \alpha(t_{j-1}) \) and scale parameter \( \beta(t_j/2 + t_{j-1}/2) \). Therefore, we can take a partition where all \( \Delta_j \) are small and equal length. We follow the same data augmentation with latent variable \( m \) generated as in the previous paragraph. Then we generate \( \delta_j \) independently given \( m \) and the data by \( \delta_j \sim \Gamma(\alpha(t_j) - \alpha(t_{j-1}) + \sum_{i=1}^{n} m_{ij}, t_j - t_{j-1}/2 - t_j/2 + \beta(t_{j-1}/2 + t_j/2)) \).

Another “partitioning approximation” method can be obtained by writing the prior as a product of gamma densities with scale parameter 1. The likelihood is a product of sums of functions of the \( \beta s \). Let \( \Delta z_j = Z(t_j) - Z(t_{j-1}) \). If we take \( J \) large, then the posterior
distribution of the \((\Delta z_1, \ldots, \Delta z_J)\) can be approximated by
\[
\left( \sum_{j=1}^{i_1} \beta_j \Delta z_j \right) \cdots \left( \sum_{j=1}^{i_1+\cdots+i_n} \beta_j \Delta z_j \right) e^{-\sum_{j=1}^J \beta_j \Delta z_j (t_{u+1}/2-t_j/2)} \prod_{j=1}^J (\Delta z_j)^{n_{o_j}-1} e^{-\Delta z_j}.
\]
So we can introduce a similar data augmentation idea with the multinomial latent variables where the cell probabilities are functions of \(\beta\)s. Then we update the \(\Delta z_j\) variate independently given the latent variables by generating updated gamma variates.

For a “happy” system, we need to define \(\delta_j = \lambda(t_{j-1}) - \lambda(t_j)\). We also relabel from right to left the index set for the number of subintervals. Therefore, similar to (14), the posterior distribution on the decrements is
\[
\delta|D_{TU} \propto \left( \sum_{j=1}^{i_1} \delta_j \right) \cdots \left( \sum_{j=1}^{i_1+\cdots+i_n} \delta_j \right) \cdots \left( \sum_{j=1}^{i_3} \delta_j \right) e^{-\sum_{j=1}^J \delta_j (t_{u+1}/2-t_{j-1})} \prod_{j=1}^J f_{j}.
\]
Therefore, we can apply the same data augmentation idea to simulate from the posterior distribution.

We have focused on the extended gamma process for the monotone ROCOF. The above method can be easily modified to accommodate for other priors such as gamma, beta, and simple homogeneous on the space of monotone ROCOF. It only requires modifying the Walker-Damien algorithm to simulate \(\delta\) from the appropriate prior process and following the same procedure as before.

6. Posterior Predictive Inference

The predictive inference on the cumulative ROCOF or the ROCOF and their pointwise credible intervals can be obtained directly from the sampling-based methods discussed previously. Moreover, we are interested in predicting the future reliability function and the mean time between failures.

It is easier to consider the failure truncated model now, that is, testing until the \(n^{t}i\) failure. We can apply the same methodology as before to derive the posterior distributions with \(D_{TU}\) replaced by \(D_{x_n}\).

Prediction for the future survival function evaluated at \(t\) distance away from \(x_n\) can be obtained by
\[
E(S(t)|D_{x_n}) = E[E(P(X_{n+1} > x_n + t)|\Lambda, D_{x_n})|D_{x_n}] = E[\exp\{-\Lambda(x_n + t) + \Lambda(x_n)\}|D_{x_n}] .
\]
The last expectation is with respect to the posterior distribution of \( \Lambda \) given \( D_{x_n} \). In general, this expectation can be approximated by the sampling-based method using the output of MCMC type algorithms. More precisely, we sample the continuous part of the increment \( \delta^{(r)} = [\Lambda(x_n + t) - \Lambda(x_n)]^{(r)} \) in the \( r^{th} \) replication for the interval between \( x_n \) and \( x_n + t \), then we average out \( \exp(-\delta^{(r)}) \) over the replications.

For the gamma process, we can evaluate (15) precisely using the result of Theorem 1. It follows that,

\[
E(S(t)|D_{x_n}) = \left( \frac{1 + c}{2 + c} \right)^{\frac{\alpha_0(x_n + t) - \alpha_0(x_n)}{2}}
\]

independent of \( d = (d_1, \ldots, d_n) \). This result may be considered undesirable by some viewers, nevertheless it is expected in the construction. Although the posterior distribution of \( \Lambda \) depends on \( d \), it is reasonable to expect the distribution of the increment of \( \Lambda \) between \( x_n \) and \( x_n + t \) for any \( t \) to be independent of \( d \) because there are no data in the interval. A hierarchical Bayesian set up that putting prior on the parameters of the gamma process will alleviate this drawback. The posterior updating is more complex, nevertheless it is feasible.

Prediction of the future mean time between failures (MTBF) can be computed by

\[
E(T_{n+1}|D_{x_n}) = E[E(T_{n+1}|\Lambda, D_{x_n})|D_{x_n}]
\]

(16)

\[
= E[\int_0^\infty \exp\{-\Lambda(x_n + t) + \Lambda(x_n)\} dt | D_{x_n}]
\]

\[
= \int_0^\infty E(\exp\{-\Lambda(x_n + t) + \Lambda(x_n)\}) | D_{x_n}) dt.
\]

The expectation inside the integral is with respect to the posterior distribution of \( \Lambda \) given the data \( D_{x_n} \). Therefore, for gamma process prior it follows that,

\[
E(T_{n+1}|D_{x_n}) = \int_0^\infty \left( \frac{1 + c}{2 + c} \right)^{\frac{\alpha_0(x_n + t) - \alpha_0(x_n)}{2}} dt,
\]

(17)

If we take prior mean function to be \( \Lambda_0(t) = bt^a \), then (17) can be evaluated in closed form. In fact we can show that it is given by,

\[
\frac{\Gamma(\frac{1}{a})[1 - \text{ING}(-(\ln c^a)x_n^\alpha, \frac{1}{a})]}{a(c^a)^{x_n^\alpha}(-\ln c^a)^{\frac{1}{a}}},
\]

(18)

where

\[
c^a = \left( \frac{1 + c}{2 + c} \right)^b,
\]

and \( \text{ING} \) stands for the incomplete gamma function. When the integral in (17) cannot be evaluated in closed form, we can always employ numerical integration techniques.
7. Model Determination

For each $i$, $i = 1, \ldots, n$, the sequential conditional predictive density (Dawid 1984) for the future failure epoch $X_{i+1}$ given $D_{x_i} = (x_1, \ldots, x_i)$ can be computed by

$$ p(X_{i+1} \mid D_{x_i}) = \int p(X_{i+1} \mid \Lambda, D_{x_i})P(d\Lambda \mid D_{x_i}) $$

$$ = \int \lambda(X_{i+1}) \exp\{-\Lambda(X_{i+1}) + \Lambda(x_i)\}P(d\Lambda \mid D_{x_i}). \quad (19) $$

Therefore, we can construct a $1 - \alpha$ level prediction interval for $X_{i+1}$ from the Gibbs sample of $X_{i+1}$ using (19). For example, the 95% predictive interval for $X_{i+1}$ can be constructed from the 2.5% and 97.5% quantiles of the sample $\{x_{i+1}^r\}_{r=1}^R$ generated from (19). We construct intervals for each $X_i$ based on past data $D_{x_{i-1}}$ for $i = 2, \ldots, n$. We judge the model is adequate if at least 95% of the intervals contain the actual observed data $x_i$.

Now we elaborate on how to generate $X_{i+1}$ from the predictive density $p(X_{i+1} \mid D_{x_i})$ in (19). Observe the expression $\lambda(X_{i+1}) \exp\{-\Lambda(X_{i+1}) + \Lambda(x_i)\}$ is a density for a truncated variable with value greater than $x_i$. Therefore, we can simulate $X_{i+1}$ from the truncated density with parameters $\Lambda^r$ that are Gibbs variates from the posterior density $p(\Lambda \mid D_{x_i})$. Let $G$ denote the non-truncated distribution, where $G(x) = 1 - \exp\{-\Lambda(x)\}$. Then $x_{i+1}$ can be generated by setting $x_{i+1}^r = G^{-1}[G(x_i) + U^r(1 - G(x_i))]$, with $U^r$ being the independent uniform variates. Therefore, we can set $x_{i+1}^r = \Lambda^r^{-1}[\Lambda^r(x_i) - \ln(1 - U^r)]$, where $\Lambda^r$ is a random path sampled from the posterior distribution of $\Lambda \mid D_{x_i}$. This method is more efficient and therefore to be preferred to the rejection method for generating a truncated random variate. The generation of $x_{i+1}^r$ can be daunting because of the replication of the sample path. If $c$ is sufficiently large for the gamma process, then we can approximate the random sample path by the mean $E(\Lambda \mid D_{x_i})$. Therefore, the easy recipe $x_{i+1}^c \approx \Lambda_0^{-1}[\Lambda_0(x_i) - (1 + c)\ln(1 - U^r)/c]$ can be used to obtain predictive intervals for large $c$. Previously, we have discussed the sampling-based approach for the whole data $D_{x_n}$. We can easily change the program to handle $D_{x_i}$. Then we repeat this for each $i = 1, \ldots, n - 1$.

For model selection, we evaluate the predictive power for the “future” (latter part of the data) against the partial posterior distribution computed from the “past” (the earlier part of the data). We divide our data into two parts for a fixed $i$ (for example, $i = \lfloor n/2 \rfloor$): the first part consists of the first $i$ failure times; the second part consists of the remaining data. We use $D_{x_i}$ to denote the second part, that is, $(x_{i+1}, \ldots, x_n)$. Then we define the predictive
likelihood of the “future” $\overline{T}_{x_i}$ given the “past” $D_{x_i}$ for a model denoted by $M$:

$$p_M(\overline{T}_{x_i} | D_{x_i}) = \int p(\overline{T}_{x_i} | \Lambda, D_{x_i}) p_M(\Lambda | D_{x_i})$$

$$= \int \left( \prod_{i=1}^{n} \lambda(x_i) \right) e^{\Lambda(x_i)} - \Lambda(x_i) p_M(\Lambda | D_{x_i}).$$

(20)

The last equality follows from the Markovian property of the inter-arrival times for the NHPP. The modified posterior Bayes factor criterion prefers model $M_1$ to $M_2$ if the ratio $p_{M_2}(\overline{T}_{x_i} | D_{x_i}) / p_{M_1}(\overline{T}_{x_i} | D_{x_i}) < 1$. This ratio can easily be approximated by applying the Monte Carlo integration technique to $p_{M_1}(\overline{T}_{x_i} | D_{x_i})$ and $p_{M_2}(\overline{T}_{x_i} | D_{x_i})$ separately; this is, for $j = 1, 2$, the predictive likelihood $p_{M_j}(\overline{T}_{x_i} | D_{x_i})$ is estimated by averaging $p_{M_j}(\overline{T}_{x_i} | \Lambda_{M_j}, D_{x_i})$ over the $\Lambda_{M_j}$’s drawn in the replications of the Gibbs sampler based on the data $D_{x_i}$.

8. A Numerical Example

The data of 31 interfailure times $(t_1, \ldots, t_{31}) = (9, 12, 11, 4, 7, 2, 5, 8, 5, 7, 1, 6, 1, 9, 4, 1, 3, 3, 6, 1, 11, 33, 7, 91, 2, 1, 87, 47, 12, 9, 135)$ were based on the trouble report for one of the larger modules of the Naval Tactical Data System. The interfailure time $t_i$ is equal to $x_i - x_{i-1}$ for $i = 1, \ldots, 31$. We have considered three classes of prior processes to model the given data. We have used the parametric models as obtained in Kuo and Yang (1996) as our prior information. We have matched the prior processes using identical means and variances. For the gamma process prior we have used $c = 2$ and two choices for the mean $\Lambda_0(t)$: $1.24t^{0.544}$ (Weibull) and $14.37 \log(1 + t/67.96)$ (Pareto). Both mean functions are suitable for “happy” system where each $\lambda_0(t)$ is a decreasing function of $t$.

For the beta process prior we have used $A_0(t) = \Lambda_0(t)$ and $c(t) = c - 1$. Finally for the extended gamma process prior we have used $\mu_0(t) = \Lambda_0(t)$ and $\sigma^2(t) = \Lambda_0(t) / c$ and then used the equations for the “happy” system at the end of section 3.2 to obtain $\alpha(t)$ and $\beta(t)$. Here we may note the $E(\Lambda(t)) = \Lambda_0(t) = A_0(t) = \mu_0(t)$, so on the average we have the same prior information for the three prior processes. Also the above choices match the variances of the processes.

To obtain the posterior mean of the processes (and the associated 95% band) we have used two approaches: the “Lévy process generation” method and the “partitioning approximation” method. We have found that both approaches give almost identical results for the three processes except the Lévy process generation method yields tighter confidence bands.
than the partitioning approximation method for the gamma process. For the partitioning approximation method we have used $J = 1000$ for all three processes. Higher values of $J$ increase the computing time. However we found that the higher values of $J$ for our data hardly alter the posterior inference. Although the partitioning approximation method is comparatively easier to work with, this method usually takes longer computing time. Among the three prior processes, the gamma process is least computer intensive as we have explicit analytical results for the posterior process. Beta process took fairly long time and depends on the choice of $\epsilon$ as used in Appendix A. For our data we have used $\epsilon = 1/600 = 0.0017$. Other choices of $\epsilon$ are possible, but for our data we have found this choice works fine. Extended gamma process is the most computer intensive method. We ran the Gibbs sampler for over 20,000 iterations until convergence.

We have plotted the prior mean process along with the posterior mean (and associated 95% band) for each of the above processes. The figures indicate that the posterior distributions are almost identical for each of the three prior processes. This is not surprising because of our choices in matching the priors and in selecting the prior means for the “happy” system for the gamma and beta processes. We exhibit the results only for the gamma processes in this section for brevity. In Figure 1 we plot the prior mean represented by the dashed curve along with the posterior mean and the associated 95% band, represented by the three solid lines. We see that the prior guess used in this case has a good resemblance with the posterior mean. The actual data is plotted on the x-axis of each plot. In the first panel of Figure 1, we plotted the posterior (along with Weibull and Pareto prior) obtained by the “Lévy process representation”. In the second panel we present the posterior obtained by “partitioning approximation” methods. The results show that the two methods yield almost identical means and the partitioning approximation method yields wider confidence bands than the Lévy representation method. This is perhaps due to the sampling error of the partition. Notice that we have used tuning parameters $c = 2, m = 1000, R = 5000$, and $J = 1000$.

We use (18) to compute the predicted MTBFs and compare them with the observed ones. In Figure 2, we plot the predicted MTBFs versus the observed MTBF for different tuning parameters $c = .35, .9, 1.45, \text{ and } 2$. We plot the 45$^\circ$ line (dotted) in each plot. It is clear that we tend to over-predict the interfailure times for small values of $c$ and the prediction improves as we put more prior information (by increasing the value of $c$ and hence increasing the prior precision). Also the variation of the predicted MTBFs decreases (as anticipated)
as $c$ increases. The results for $c = 2$ are slightly more conservative than that of Goel and Okumoto (1979), that is the MTBFs are higher than that of Goel and Okumoto at the beginning of testing, and increases at a slower rate as testing continues.

Appendix A

We describe the Walker-Damien method for the beta process. Observe the Lévy measure for the posterior beta process is given by $dN_0^*(z) = z^{-1}dz \int_{[0,1]} (1 - z)^{c(s) + dK(s)}c(s) dA_{0,c}(s)$. Suppose we are interested in generating an increment $\delta_j$ of $\Lambda$ from the continuous part over the set $\Delta = (t_j-1, t_j)$. The Walker-Damien method consists of approximating $\delta_j$ by $X_\epsilon$ that is a finite sum of the jumps of a Poisson process with intensity measure $dN^*(.).$ Let $g_\epsilon(z) \sim z^{-1}I(z > \epsilon) \int_\Delta (1 - z)^{c(s) + dK(s)}c(s) dA_{0,c}(s)$. We define $C_\epsilon = \int_0^\infty \int_0^\infty z^{-1}(1 - z)^{c(s) + dK(s)}c(s) a_{0,c}(s) I(z > \epsilon, s \in \Delta)dzds$, where $a_{0,c}(s)ds = dA_{0,c}(s)$. Then Walker-Damien’s method consists of taking $\nu \sim \text{Poisson}(C_\epsilon)$ and taking $z_1, \ldots, z_\nu$ i.i.d. from $g_\epsilon$. Then let $X_\epsilon = \sum_{k=1}^\nu z_k$.

To sample $z$ from the density $g_\epsilon(z)$, we introduce latent variable $s$ and $u$ and consider the joint density of $z, s$ and $u$ by

$$f(z, s, u) \propto I(u < 1/z)(1 - z)^{c(s) + dK(s)}c(s) dA_{0,c}(s) I(z > \epsilon, s \in \Delta),$$

where $u$ is a random variable defined on the interval $(0, 1/\epsilon)$, $s \in \mathbb{R}^+$. So the marginal distribution of $z$ is $g_\epsilon(z)$. Let $\{z^{(l)}, s^{(l)}, u^{(l)}\}$ be the starting point for a hybrid Gibbs/Metropolis algorithm. Then we update the $z, s$, and $u$ variates by

(a) generate $z^{(l+1)}$ given $s^{(l)}$ and $u^{(l)}$ from a beta distribution with parameter $(1, c(s^{(l)}) + dK(s^{(l)}) + 1)$ restricted to the interval $(\epsilon, 1/u^{(l)})$, for $u^{(l)} < 1/\epsilon$. This can be done by setting

$$z^{(l+1)} = 1 - \{1 - (1 - z^{(l)}/c(s^{(l)}) + dK(s^{(l)}) + 1 - 1/u^{(l)}c(s^{(l)}) + dK(s^{(l)}) + 1)\}^{1/(c(s^{(l)}) + dK(s^{(l)}) + 1)},$$

where $U$ is a uniform variate on $(0, 1)$.

(b) generate $u^{(l+1)}$ given $z^{(l+1)}$ and $s^{(l)}$ from a uniform distribution on $(0, 1/z^{(l+1)})$.

(c) generate $s^{(l+1)}$ given $u^{(l+1)}$ and $z^{(l+1)}$ by a Metropolis (Metropolis et al., 1953) step: take $s^*$ from $A_{0,c}(\cdot)/A_{0,c}(\Delta)$ on $\Delta$ and $\xi$ from a uniform distribution on $(0, 1)$. If

$$\xi < \min\{1, (1 - z^{(l+1)}c(s^*) + dK(s^*)c(s^*)/[1 - z^{(l+1)}c(s^{(l)}) + dK(s^{(l)})c(s^{(l)})]\}$$

then $s^{(l+1)} = s^*$, else $s^{(l+1)} = s^{(l)}$. 

24
Next we define $h(z, s) = (1 - z)^{c(s)+dK(s)}c(s)dA_0,c(s)I(z > \epsilon, s \in \Delta)$. Therefore,

$$C_{\epsilon} = \int_0^\infty \int_0^\infty h(z, s)dzds,$$

Consequently, we can approximate $C_{\epsilon}$ by

$$C_{\epsilon} \approx \left( \frac{1}{L} \sum_{l=1}^L 1/z(l) \right) \int_0^\infty \int_0^\infty h(z, s)dzds,$$

$$\int_0^\infty \int_0^\infty h(z, s)dzds = \int_\Delta \int_\Delta (1 - z)^{c(s)+dK(s)}c(s)a_{0,c}(s)dzds$$

$$= \int_\Delta \int \frac{c(s)}{c(s) + dK(s) + 1} a_{0,c}(s)ds$$

$$\approx \int_\Delta \frac{c(s)}{c(s) + dK(s) + 1} a_{0,c}(s)ds - \epsilon \int_\Delta c(s)a_{0,c}(s)ds.$$

Because we need $\nu$ i.i.d. observations from $g_{\epsilon}(z)$, where $\nu \sim \text{Poisson}(C_{\epsilon})$, we will replicate the above chain $\nu$ times and take a simple sample $z(l)$ from each chain.

**Appendix B**

The Walker-Damien Method that generates an increment from the extended Gamma process prior is given here. To sample $z$ from the density $g_{\epsilon}(z)$, we introduce latent variable $s$ and $u$ and consider the joint density of $z$, $s$ and $u$ by

$$f(z, s, u) \propto I(u < 1/z)\exp\{-z/\beta(s)\}a(s)I(z > \epsilon, s \in \Delta),$$

where $u$ is a random variable defined on the interval $(0, 1/\epsilon)$, $s \in \mathbb{R}^\times$. So the marginal distribution of $z$ is $g_{\epsilon}(z)$. Let $\{z(l), s(l), u(l)\}$ be the starting point for a hybrid Gibbs/Metropolis algorithm. Then we update the $z$, $s$, and $u$ variates by

(a) generate $z^{(l+1)}$ given $s(l)$ and $u(l)$ from an exponential distribution with mean $\beta(s(l))$ restricted to the interval $(\epsilon, 1/u(l))$, for $u(l) < 1/\epsilon$. This can be done by setting $z^{(l+1)} = -\beta(s(l))\ln[\exp\{-\epsilon/\beta(s(l))\} - U(\exp\{-\epsilon/\beta(s(l))\} - \exp\{-1/(\beta(s(l))u(l))\})]]$, where $U$ is a uniform variate on $(0,1)$.

(b) generate $u^{(l+1)}$ given $z^{(l+1)}$ and $s(l)$ from a uniform distribution on $(0, 1/z^{(l+1)})$.

(c) generate $s^{(l+1)}$ given $u^{(l+1)}$ and $z^{(l+1)}$ by a Metropolis step: take $s^*$ from $\alpha(.)/\alpha(\Delta)$ on $\Delta$ and $\xi$ from a uniform distribution on $(0,1)$. If

$$\xi < \min\{1, \exp(-z^{(l+1)}[1/\beta(s^*) - 1/\beta(s(l))])\}$$

then $s^{(l+1)} = s^*$, else $s^{(l+1)} = s(l)$.
Next we define \( h(z, s) = \exp\{-z/\beta(s)\} d\alpha(s) I(z > \epsilon, s \in \Delta) \). Therefore,

\[
C_\epsilon = \int_0^\infty \int_0^\infty h(z, s) / z dz ds.
\]

Consequently, we can approximate \( C_\epsilon \) by

\[
C_\epsilon \approx \left( \frac{1}{T} \sum_{i=1}^T 1/z_i \right) \int_0^\infty \int_0^\infty h(z, s) / z dz ds,
\]

where

\[
\int_0^\infty \int_0^\infty h(z, s) / z dz ds = \int_\Delta \beta(s) \exp\{-\epsilon/\beta(s)\} d\alpha(s) \approx \int_\Delta \beta(s) d\alpha(s) - \epsilon \alpha(\Delta).
\]

Because we need \( \nu \) i.i.d. observations from \( g_\nu(z) \), where \( \nu \sim \text{Poisson}(C_\epsilon) \), we will replicate the above chain \( \nu \) times and take a simple sample \( z^{(E)} \) from each chain.

References


L. Kuo, Statistics Department, University of Connecticut, Storrs, CT 06269, USA.

S. Ghosh, Department of Statistics, NC State University, Raleigh, NC 27695-8203, USA.
Figure 1: Posterior estimates of the cumulative ROCOF for the gamma process prior.
Figure 2: Predicted MTBF versus observed interfailure times for the gamma process.