Financial and Statistical Methods

Bayesian Estimation of Renewal Function Based on Pareto-distributed Inter-arrival Times via an MCMC Algorithm

M.S. Aminzadeh[a], Min Deng[b]

Variance

The purpose of this article is to provide a computational tool via Maximum Likelihood (ML) and Markov Chain Mont Carlo (MCMC) methods for estimating the renewal function when the inter-arrival distribution of a renewal process is single-parameter Pareto (SPP). The proposed method has applications in a variety of applied fields such as insurance modeling and modeling self-similar network traffic, to name a few. It is shown that inter-arrivals of insured damages for floods and tornados during 2000-2020 in the USA have SPP distribution. It is also shown that inter-arrivals of recent hurricanes hitting New Orleans fit SPP distribution. For the Bayesian estimation of SPP parameters via the MCMC method, based on the Metropolis algorithm, gamma and shifted exponential distributions are used. Simulations confirm that the MCMC estimator of the renewal function outperforms maximum likelihood estimator (MLE) with regards to its accuracy when the sample size is relatively small. However, for large samples, the accuracies of ML and Bayes estimators for the renewal function are comparable.

1. INTRODUCTION

In the literature, we often see the application of a single-parameter Pareto (SPP) distribution as a model to describe income and wealth. In insurance applications, heavy-tailed distributions of the SPP sort are essential tools for modeling extreme losses, especially for more risky types of insurance, such as medical malpractice insurance. In the field of actuarial science, the SPP distribution is combined with another distribution to model the portion of data characterized by large claims with low frequencies, i.e., high-severity risks. For the other portion of data characterized by small claims with high frequencies, i.e., low-severity risks, other distributions such as the exponential or the log-normal are used. Such models are called composite models. Deng and Aminzadeh (2020) developed a Bayesian predictive Weibull-Pareto composite model, and Aminzadeh

and Deng (2019) developed a Bayesian predictive inverse gamma–Pareto composite model. Recent work in the field of network traffic has considered an SPP distribution as the interarrival distribution. Fras et al. (2010) considered the distribution for modeling self-similar network traffic. Harrison et al. (2014) used a network simulator to generate statistics on arrival times of traffic and concluded that an SPP distribution is more appropriate than the conventional exponential distribution in modeling interarrival times on a bursty network. The SPP distribution has also been used for insurance models: Hanafy (2020) used the distribution for interarrivals of claims on data from Egyptian fire insurance companies.

The Lomax distribution (also known as a Pareto Type II distribution) is a heavy-tailed probability distribution used in business, economics, actuarial science, queueing theory, and Internet traffic modeling. It is essentially the SPP dis-

a   Dr. Aminzadeh, 1985 Ph.D., Oregon State University. His research areas of interest include Time Series, Bayesian Inference, Predictive Inference, and Stochastic Processes. He has extensive teaching experience in Applied Probability, Stochastic Processes, and Theoretical and Applied Statistics courses at undergraduate/graduate levels at Temple University, Towson University, and Johns Hopkins Universities at APL. He served as the director of the Applied and Industrial Mathematics (APIM) graduate program at Towson University and reviewed articles for numerous journals such as IEEE Transactions on Reliability, Applied Statistics, Communications in Statistics, Theory, and Methods, IEEE Transactions on Automatic Control, Computational Statistics and Variance journal.

b   Dr. Min Deng, FSA, 1990 Ph.D., Pennsylvania State University. She has outstanding teaching experience, specializing in mathematics, statistics, and actuarial science. Dr. Deng has been widely recognized for her teaching at the University of Wisconsin-Stevens Point, Maryville University, and Towson University. She has published numerous articles. She is the author of the book "Paper: An Engineered Stochastic Structure". Her research interests include differential geometry in statistical inference, engineered stochastic structure, risk theory, and model construction. Professor Deng is the director of the Actuarial Science and Risk Management undergraduate program, and the Actuarial Science and Predictive Analytics Master's Program at Towson University.

tribution, but its support begins at zero and it has a probability density function (PDF) given as

$$f(x) = \frac{\alpha}{m}\left(1 + \frac{x}{m}\right)^{-(\alpha+1)}, \quad x > 0, \quad \alpha, m > 0$$

The above PDF can also be written as

$$f(x) = \frac{\alpha}{m}\left(\frac{x}{m}\right)^{-(\alpha+1)}, \quad x > m, \quad \alpha, m > 0.$$

which is the standard form of the PDF for the SPP and which is also used by Ramsay (2006). Garsva et al. (2014) considered a Pareto Type II distribution as a model for the packet interarrival time distribution in an academic computer network, and they concluded via a Kolmogorov-Smirnov test that it was an appropriate choice.

In this paper, we analyze flood and tornado insurance damage data sets (provided in the Emergency Events Database (EM-DAT); see appendix) in the United States from the year 2000 to the year 2020. We show that both data sets fit the SPP distribution well. The interarrivals between two insured damages are defined as the difference between the start date of an occurrence and the start date of the subsequent occurrence.

Our purpose in this paper is to use SPP as the interarrival distribution of a renewal process and provide the ML- and MCMC-based Bayes estimators of the renewal function. Martino (2018) provides a thorough review of MCMC methods and compares them with numerical methods. In the current paper, we use the Metropolis-Hastings algorithm—see Casella and Berger (2004)—for the MCMC method to estimate the renewal function.

In Section 2, we define the renewal function and give computational formulas for the renewal function based on Ramsay (2006). These formulas are used in the Mathematica code for simulations and estimation of the renewal function. Section 3 provides the ML estimation for the parameters of the SPP distribution. Section 4 provides the derivation of two conditional posterior distributions that are based on shifted-exponential and gamma priors. In Section 5 we outline the steps taken in the Mathematica code to use the MCMC algorithm and discuss a summary of the simulation results. Section 6 provides numerical examples to illustrate the computations, and Section 7 concludes. The Mathematica code used for the simulations and the estimation of the renewal function in this paper is available from the authors.

2. RENEWAL FUNCTION OF THE SPP DISTRIBUTION

Let $\{N(t), t \geq 0\}$ be a counting process, where $N(t)$ denotes the number of events that occur by time t. Let X_n denote the time between the $(n-1)$st and nth events of this process, $n \geq 1$. If $\{X_1, X_2, \dots\}$ is a sequence of independent identically distributed (iid) random variables, then the process $\{N(t), t \geq 0\}$ is called a renewal process.

Renewal theory has many applications in stochastic modeling, and as a result, statistical inference with respect to the renewal function is of interest in many applied fields. For example, in reliability theory, a renewal process can be used to model successive repairs of a failed machine. In

actuarial science, a renewal process can be used to model the successive occurrences of risks. Hardy (2006) provides many examples for risk measures with actuarial applications, such as in property and casualty (P&C) insurance, life insurance, and the banking industry. Necir, Rassoul, and Meraghni (2010) provide a semiparametric estimate of the renewal function for heavy-tailed risks. Woo (2011) provides two-sided bounds for the renewal function considering a variety of reliability classifications.

In this paper, we consider ML- and MCMC-based Bayes estimators of the renewal function. Actuaries should find the results useful in applications where events (risks) occur according to a counting process and the renewal random variable has an SPP distribution. One of the measures associated with a renewal process is the renewal function, $M(t) = E[N(t)]$. In some applications of renewal processes, one is interested in estimating the number of renewals by time t. Provided that the parameters associated with the interarrival distribution are known or estimated, the computation can be accomplished by using the integral equation

$$\begin{aligned} M(t) = E[N(t)] &= \int_0^\infty E[N(t)|X_1 = x]f(x)dx \\ &= F(t) + \int_0^\infty M(t-x)f(x)dx, \end{aligned} \tag{1}$$

where f and F are, respectively, the PDF and the cumulative distribution function (CDF) of the interarrival random variable X. For example: (a) for the uniform(0,1) interarrival distribution, $M(t) = e^t - 1, 0 \leq t \leq 1$, and (b) for the exponential interarrival distribution with the PDF $f(x) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}$, where λ is the scale parameter, $M(t) = \lambda t, t > 0$, the reason being that when the interarrival times have an exponential distribution, $N(t) \sim \text{Poisson}(\lambda t)$. However, the derivation of $M(t)$ in a closed form for many interarrival distributions is not straightforward. Let $S_n = \sum_{i=1}^n X_i$ denote the time of nth renewal. It can be shown that

$$M(t) = \sum_{n=1}^\infty F_{S_n}(t),$$

$$\partial M(t)/\partial t = M'(t) = \sum_{n=1}^\infty f_{S_n}(t), \tag{2}$$

where f_{S_n} and F_{S_n}, respectively, are the PDF and the CDF of S_n. Therefore, to find $M(t)$ in a closed form, not only must we identify the distribution of S_n, but we must also evaluate $\sum_{n=1}^\infty F_{S_n}(t)$, which may not be trivial.

Ross (2007) supplies us with a method to approximate the renewal function. The method computes $M_2(t), M_3(t), \dots, M_n(t)$ recursively to approximate the renewal function $M(t)$ using

$$M_r(t) = \frac{\sum_{k=1}^{r-1}(1 + M_{r-k}(t))E(X^k e^{-\delta_r X})(\delta_r^k/k!) + E(e^{-\delta_r X})}{1 - E(e^{-\delta_r X})},$$

$$M_1(t) = \frac{E(e^{-\delta_1 X})}{1 - E(e^{-\delta_1 X})},$$

where $\delta_r = r/t$ is used for the computation of $M_r, r = 1, 2, \dots$. Therefore, $M_1(t), M_2(t), M_3(t), \dots, M_n(t)$ are found consecutively to approximate the renewal function $M(t)$. Ross indicates that for a large r, $M(t)$ is approximated by $M_r(t)$ quite well. The method is useful when $E(X^k e^{-\delta_r X})$ and $E(e^{-\delta_r X})$ have explicit expressions for an

interarrival distribution. For example, if the interarrival distribution is gamma (2,1) (shape parameter = 2, and scale parameter = 1), then

$$E[X^k e^{-\frac{rX}{t}}] = t^2 (\frac{r+t}{t})^{-k} \frac{\Gamma(2+k)}{(r+t)^2}, \quad \frac{r}{t} > -1,$$

Then using (2) it can be shown that the exact expression for the renewal function is

$$M(t) = .5t - .25(1 - e^{-2t}), t > 0.$$

For $t = 10$, $M(10) = 4.75$ is the exact value, where its approximate value based on $r = 50$ is calculated in Ross (2007) as 4.75. As another example, for the interarrival random variable with PDF $f(x) = 1.4e^{-2x} + .3e^{-x}$ and CDF $F(x) = 1 - .3e^{-x} - .7^{-2x}$, $M(10)$ is approximated as 15.5089 (see Ross 2007). For some interarrival distributions, $M(t)$ cannot be derived in a closed form. For example, consider the inverse Gaussian (IG) distribution and let $X_i \sim \text{IG}(\phi, \lambda)$; it can be shown that $S_n \sim \text{IG}(\phi/n, n^2\lambda)$, but $\sum_{n=1}^{\infty} F_{S_n}(t)$ cannot be written in a closed form. See Aminzadeh (2011), where the author provides a Bayes estimator of the renewal function using IG distributed interarrivals. For $X \sim \text{IG}(\phi, \lambda)$, with given values for r and t and unknown values for ϕ and λ, Mathematica cannot provide a closed form for

$$\int_0^{\infty} x^k e^{-(r/t)x} f(x; \phi, \lambda) dx.$$

However, for example, with $\phi = 2, \lambda = 3, t = 20$, and $r = 5$, for the above expected value we get

$$2^k 3^{1/4+k/2} 5^{1/4+k/2} e^{1.5} \text{BesselK}[.5 - k, \sqrt{15}/2]/\sqrt{\pi}$$

via Mathematica with a long integration time, and the expression is quite complicated. SPP(m, α), which is considered in the current paper, is an even more difficult case, as

$$\int_m^{\infty} x^k e^{-(r/t)x} f(x; m, \alpha) dx$$

cannot be obtained in a closed formula, even for selected values of m, α, t, and r.

For the SPP distribution, the approximation Ross provides is not practical. The reason is that $E(X^k e^{-\delta_r X})$ cannot be written in a closed form and therefore an approximate expression for the expected values must be derived or evaluated numerically, which would depend on k. Even if a "good" expression for the expected value is found, different values of k would produce different expressions/approximations for the expected value. In addition, the recursive formula for $M_r(t)$ uses previous approximations for the expected value. Therefore even for a small value of r, $E(X^k e^{-\delta_r X})$ in the above sum has to be approximated for $k = 1, 2, \ldots (r-1)$. In addition to the $r - 1$ approximation errors, one has to take into account the accumulated approximation errors for the expected value that are involved in the computation of $M_{r-k}(t)$. Thus, due to the many accumulated approximation errors involved for approximating $E(X^k e^{-\delta_r X})$ and $M_r(t)$, Ross's method is not considered in this paper.

The PDF of the SPP(m, α) random variable X, with parameters α and m, is given by

$$f_X(x; \alpha, m) = \alpha x^{-(\alpha+1)} m^{\alpha}, \\ x > m, \text{ and} \quad (3) \\ \alpha, m > 0.$$

Consider a renewal process $\{N(t), t \geq 0\}$ and let X_n, the time between the $(n-1)$st and nth renewal, be distributed as SPP(α, m), $n = 1, 2, 3, \ldots$. Suppose from this process a random sample $X_1, \ldots, X_H$ is collected. The random sample will be used to find the ML and an approximate Bayes estimate for $M(t)$ via an MCMC algorithm.

Ramsay (2006) applied Laplace transforms and generalized exponential integrals to develop the following exact expressions for PDF f_{S_n} and CDF F_{S_n} of $S_n = \sum_{i=1}^{n} X_i$ under the assumption that $X_i, i = 1, 2, \ldots$ are iid from the SPP distribution and that α is a positive integer. Recall that S_n is the time of nth renewal. The following formulas are from Ramsay (2006):

$$f_{S_n}(s) = \frac{1}{nm} \int_0^{\infty} e^{-(1+\frac{s}{nm})v} \phi_{\alpha,n}(\frac{v}{n}) dv,$$

where,

$$\phi_{\alpha,n}(v) = (-1)^{n+1} \alpha^n \sum_{r=0}^{\lfloor (n-1)/2 \rfloor} (-\pi^2)^r \\ \binom{n}{2r+1} (G_{\alpha+1}(v))^{n-2r-1} (\frac{v^{\alpha}}{\alpha!})^{2r+1},$$

$$G_{\alpha}(v) = \frac{v^{\alpha-1}}{(\alpha-1)!} [\gamma + \ln(v) - \sum_{r=1}^{\alpha-1} \frac{1}{r}] \\ + \sum_{r=0, r \neq \alpha-1}^{\infty} \frac{v^r}{(r-\alpha+1)r!},$$

$$\gamma = .5772156649,$$

and

$$F_{S_n}(t) = \int_0^t f_{S_n}(s) ds \\ = \int_0^{\infty} \frac{1}{v} (1 - e^{-\frac{tv}{m}}) e^{-nv} \phi_{\alpha,n}(v) dv. \quad (4)$$

The approach for computing $F_{S_n}(t)$ in Ramsay is to use a series expansion and then use

$$e^{-nv} \phi_{\alpha,n}(v) = (-1)^{n+1} \alpha^n \sum_{r=0}^{\lfloor (n-1)/2 \rfloor} (-\pi^2)^r \\ \binom{n}{2r+1} (e^{-v} G_{\alpha+1}(v))^{n-2r-1} \\ (\frac{e^{-v} v^{\alpha}}{\alpha!})^{2r+1}$$

in (4) to compute $F_{S_n}(t)$. The above formulas are built into the Mathematica code of the current paper, and the NInte-grate function of Mathematica is used to compute (4) based on input values m, α, n, and t.

Although the results given in Ramsay assume that α is a positive integer, simulation studies confirm that for both ML and Bayes estimates of the renewal function, once the ML and Bayes estimates of the parameters α and m are obtained, the estimated values of α can be rounded to the nearest integer so that $F_{S_n}(t)$ and as a result an estimate for $M(t)$ can be obtained. The reason that in Ramsay α is assumed to be a positive integer is that to derive (4), for complex z and $\text{Re}(z) > 0$, it is shown that the Laplace transform of $f_n(t)$ is given by $f_n^*(z) = (f^*(z))^n$, where $f^*(z) = \alpha e^{mz} E_{\alpha+1}(mz)$, and $E_{\alpha}(z) = \int_1^{\infty} \frac{e^{-zt}}{t^{\alpha}} dt$, for $\alpha = 1, 2, 3, \ldots$, is the generalized exponential integral. And then $f_n^*(z)$ is inverted to get the Bromwich integral for $f_n(t)$, and consequently the CDF in (4) is derived. Simulations confirm that by rounding the estimated values of

α, the average squared error (ASE) of the ML and Bayes estimates of the renewal function is still quite small and that the MCMC-based Bayesian method outperforms the ML method with regard to accuracy when the sample size H is small. For large sample sizes, the ML and Bayes methods have comparable accuracies.

3. MAXIMUM LIKELIHOOD ESTIMATION OF M(T)

Based on a random sample $X_1, \ldots, X_H$ from $\text{SPP}(\alpha, m)$, and using the likelihood function

$$L(\alpha, m) = \alpha^H (\prod_{i=1}^{H} X_i)^{-(\alpha+1)} m^{H\alpha}, \tag{5}$$
$$X_i > m, \quad i = 1, 2, \ldots, H,$$

it can be shown that the ML estimates of α and m are

$$\hat{m} = X_{(1)}, \quad \hat{\alpha} = \frac{H}{T - H \ln\left(X_{(1)}\right)}, \quad T = \sum_{i=1}^{H} \ln(X_i),$$

and that $(X_{(1)}, T)$ is a minimal sufficient statistic for (m, α). $X_{(1)}$ is the first-order statistic. The goal is to estimate

$$M(t) = \int_0^t M'(s)ds = \sum_{n=1}^{\infty} F_{S_n}(t). \tag{6}$$

Therefore, for each value of n in (6), and given estimates of (α, m) (based on a random sample $X_1, \ldots, X_H$), we can compute $F_{S_n}(t)$ via (4). Computations confirm that for selected values α, m, and t, the CDF $F_n(t)$ in (4) is a decreasing function of n. The Mathematica code in a while loop searches for an appropriate upper limit k for the sum in (6). The loop starts with $k = 1$ and continues until $F_k(t) - F_{k-1}(t) > 0$, and, using the value of k, computes $\sum_{n=1}^{k-1} F_n(t)$ as an estimate for $M(t)$.

It is also noted that through an estimate of $M(\tau)$, where τ is a truncated time, the excess life of a renewal at time τ can be estimated. Let $Y(\tau)$ denote the excess life at τ and $\mu = E[X]$ be the mean of interarrival random variable X. We can use the well-known formula

$$M(\tau) = \frac{\tau}{\mu} - 1 + \frac{E[Y(\tau)]}{\mu},$$

to find the ML estimate of $E[Y(\tau)]$ via the ML estimate $\hat{\mu} = \frac{\hat{\alpha}\hat{m}}{\hat{\alpha}-1}$ and the ML estimate of $M(\tau)$. The same formula above can be used to get an approximate (using simple substitution) Bayes estimate of $E[Y(\tau)]$, once Bayes estimates of α, m, and $M(\tau)$ are found through the MCMC algorithm given in Section 5.

4. BAYESIAN ESTIMATION OF M(T)

In the literature, we find many instances where authors consider Bayesian inference for the parameters of the SPP distribution. Sun et al. (2021) adopt the Jeffreys prior to obtain Bayes estimates for m and α. Kim, Kang, and Lee (2009) also provide Bayes estimates via the non-informative Jeffreys prior for SPP. Noor and Aslam (2012) provide non-informative as well as informative Bayes estimates of SPP parameters; however, in their paper, gamma priors are used for both m and α.

The characteristics of the shifted exponential distribution with regard to its support (a, ∞) are similar to the support of SPP. This is the main reason that in the current paper we use the shifted exponential prior as opposed to another distribution with support $(0, \infty)$. The PDF of the shifted exponential prior is given as

$$\rho(m|a, b) = \frac{1}{b} e^{-\left(\frac{m-a}{b}\right)}, \quad m > a, \quad a, b > 0, \tag{7}$$

where a (location parameter) and b (scale parameter) are hyperparameters associated with the prior distribution.

For the conditional prior of $\alpha|m$, $\text{gamma}(c, d)$ is selected, with the PDF

$$\rho(\alpha|m) \propto \alpha^{c-1} e^{-\frac{\alpha}{d}}, \quad c, d > 0, \tag{8}$$

where c (shape parameter) and d (scale parameter) are hyperparameters associated with the conditional prior distribution. As mentioned earlier, although computation of the CDF $F_{S_n}(t)$ requires that α be an integer (as the formulas in (4) use only positive integers for α), using a gamma distribution for the conditional prior does not cause an issue with regard to the accuracy of the ML and Bayes estimates of the renewal function. For the current paper, in the Mathematica code for simulations, once a Bayes estimate for α is obtained, it is rounded to the nearest integer for computational purposes. From (7) and (8), we get the joint prior as

$$\rho(\alpha, m) \propto \alpha^{c-1} e^{-\left(\frac{\alpha}{d} + \frac{m}{b}\right)}, \quad m > a. \tag{10}$$

Now, using the likelihood function for a sample of size H, the joint posterior is given by

$$\pi(\alpha, m|\underline{x}) \propto \alpha^{H+c-1} e^{-\alpha(\frac{1}{d}+T)} m^{H\alpha} e^{-\frac{m}{b}}, \quad m > a. \tag{11}$$

To apply an MCMC algorithm, (11) is used to find the conditional posteriors as

$$\pi(\alpha|\underline{x}, m) \propto \alpha^{H+c-1} e^{-\alpha(\frac{1}{d}+T)} m^{H\alpha}, \tag{12}$$

and

$$\pi(m|\underline{x}, \alpha) \propto e^{-\frac{m}{b}} m^{(H\alpha+1)-1}, \quad m > a. \tag{13}$$

Simon and Adler (2021) provide a thorough reference on MCMC algorithms and compare several MCMC methods for estimating the parameters of the Pareto/negative binomial distribution (NBD) model and their performance. The *Handbook of Markov Chain Monte Carlo* (Brooks et al. 2011) covers MCMC foundations, methodology, and algorithms, as well as MCMC's use in a variety of practical applications in areas such as educational research, astrophysics, brain imaging, ecology, and sociology. Many authors have discussed the MCMC method via conditional posterior distributions and the Metropolis algorithm. In particular, the description and implementation of the Metropolis algorithm can be found in Casella and Berger (2004), among other books and articles.

Ji, Aminzadeh, and Deng (2020) propose an MCMC Bayesian approach using the Metropolis algorithm and conditional posteriors to obtain estimates of the parameters as well as an approximate predictive density via simulation, which can be used to compute life expectancy and other measures of interest in a Bayesian framework. Aminzadeh and Deng (2021) employ an MCMC method based on the Metropolis algorithm and conditional posterior distributions to estimate ruin probability based on nonhomogeneous Poisson process claim arrivals and inverse Gaussian–distributed claim aggregates. The MCMC algorithm we

use in this paper is based on the same logic as the aforementioned articles.

Considering the conditional posterior distributions in (12) and (13), we can see that the kernel of the PDF in (12) does not belong to the gamma$(H + c, \frac{d}{1+dT})$ distribution, due to the extra term $m^{H\alpha}$. Therefore, we use gamma(P_1, P_2) as a proposal distribution for the PDF in (12) within the MCMC algorithm. P_1 and P_2 are, respectively, the shape and scale parameters of the proposal distribution gamma(P_1, P_2). For selected "true"values of α and m, in simulation studies, values of P_1 and P_2 are selected in such a way that $P_1 P_2$ is close enough to α while P_2 is very small. This is because the expected value and variance of the proposal gamma(P_1, P_2) distribution are, respectively, $P_1 P_2$ and $P_1 P_2^2$.

As for the conditional posterior (13), we can see that it resembles the kernel of the generalized gamma(g, β, ξ, μ) distribution with the PDF

$$f(y|g, \beta, \xi, \mu) = \xi e^{-(\frac{y-\mu}{\beta})^{\xi}} (\frac{y-\mu}{\beta})^{g\xi-1}, y > \mu,$$

where $\beta = b, \mu = a, \xi = 1$, and $g = H\alpha + 1$. Note that μ is the location parameter and β is the scale parameter.

5. SIMULATION

For selected values of (α, m, t), the Mathematica code (available from the authors upon request) first computes the "true"value of the renewal function $M(t)$ via (4). The code then generates random samples $X_1, \ldots, X_H$ from the SPP(α, m) distribution using a built-in function in Mathematica. For the generated sample $i, (i = 1, 2, \ldots, N = 100)$, the following steps are taken:

1. ML estimates of α and m are computed via the formulas in Section 3.
2. For each value of n in the While loop, ML estimates from Step 1 are used to compute the CDF in (4). The while loop starts with $k = 1$ and computes $F_{S_1}(t), F_{S_2}(t), F_{S_3}(t), \ldots$ via (4). Then for each value of $k(k = 1, 2, 3 \ldots)$, $\sum_{n=1}^{k} F_{S_n}(t)$ is calculated. Computations confirm that for given values of α, m, and t, the CDF $F_{S_n}(t)$ in (4) is a decreasing function of n. The Mathematica code in the while loop searches for an appropriate upper limit k for the sum in (6). Computations continue until $F_k(t) - F_{k-1}(t) > 0$. And, using the value of k, the loop ends and $M_{i,ML}(t) = \sum_{n=1}^{k-1} F_{S_n}(t)$ is reported as the ML estimate of the renewal function based on sample i. It is noted that in this step, the ML estimates of α and m, which are based on a generated sample of size H from the SPP, are used in (4) to find $F_{S_n}(t)$ for $n = 1, 2, 3, \ldots$ and that S_n is a symbolic notation in the CDF $F_{S_n}(t)$. The computation of (4) does not depend on a value of S_n.
3. The mean and ASE of estimates $M_{i,ML}(t), (i = 1, 2, \ldots, N)$ of the renewal function based on N simulated samples are respectively computed as

$$\overline{\hat{M}}_{ML}(t) = \frac{\sum_{i=1}^{N} M_{i,ML}(t)}{N},$$
$$\xi_{ML}(\hat{M}(t)) = \frac{\sum_{i=1}^{N} (M_{i,ML}(t) - M(t))^2}{N}.$$

Note that the term MSE (mean squared error) refers to the expected value of an estimator under the squared-error loss function. In this paper, the term ASE refers to $\sum_{i=1}^{N} (\hat{\theta}_i - \theta)^2 / N$, where $\hat{\theta}_i$ is an estimate of the "true"parameter θ based on a generated sample $i, (i = 1, 2, \ldots N)$.

To compute the MCMC-based Bayes estimate of the renewal function, we use the conditional posterior distributions in (12) and (13). Let $f_1(\alpha|m_0) = \alpha^{H+c-1} e^{-\alpha(\frac{1}{d}+T)} m_0^{H\alpha}$ denote the conditional posterior PDF (RHS in (12)) of α given $m = m_0$, and $h_1(\alpha)$ is the PDF of the proposal distribution gamma (P_1, P_2) for α. As mentioned earlier, in simulation studies, for selected "true"values of α and m, the values of P_1 and P_2 are selected in such a way that $P_1 P_2$ is close enough to α while P_2 is very small. Also, the hyperparameters c and d are selected such that $cd \approx \alpha$, while d is small. Again, this is due to the expected value and variance of the gamma distribution. For example, if we select $\alpha = 3$ in simulation, then one possibility would be $c = 300, d = .01, P_1 = 10$, and $P_2 = .3$.

It is noted that the expected value and the variance of the generalized gamma$(H\alpha + 1, b, 1, a)$, which is used as the conditional posterior of $m|\alpha, \underline{x}$, are, respectively,

$$E[m|\alpha] = a + Hb\alpha, \quad \text{Var}(m|\alpha) = b^2(H\alpha + 1),$$

Therefore, the hyperparameters a and b are selected so that $m \approx a + Hb\alpha$. We choose a value for a so that it is close to $m(a < m)$ and then find the corresponding value for $b = \frac{m-a}{H\alpha}$. This way the Bayes estimates of m will be stable with a higher precision. For example, if the selected values are $m = 2.2, H = 30$, and $\alpha = 3.0$, then one possibility is $a = 2.1$ and $b = .0011$. In practice, given a sample from the SPP distribution, one does not have information on the "true"values of α and m. If P_1 and P_2 are selected such that their product is close to the ML estimate of α, while P_2 is small, the Bayes estimates of α and m would be stable with a small ASE, and as a result the renewal function can be estimated with a higher accuracy.

For example, if the ML estimates are $\hat{m} = 5.15, \hat{\alpha} = 2.15 \approx 2.0 = \alpha^*$, and $H = 30$, then we can choose hyperparameters as

$$c = 500, d = .001, P_1 = 10,$$
$$P_2 = .5, a = 2, b = \frac{\hat{m} - a}{H\alpha^*} = .0525$$

The same approach is used in Aminzadeh and Deng (2018) and Deng and Aminzadeh (2020) for the selection of hyperparameter values. Using ML estimates through a random sample helps to identify appropriate values for hyperparameters that are consistent with the sample information. Selection of hyperparameters using ML estimates is considered in the simulation studies (see Section 5 for a summary of results) as well as in the numerical example in Section 6.

4. The MCMC algorithm in this step is based on the Metropolis algorithm (see Casella and Berger 2004). The idea is to use a proposal distribution that is close enough to a conditional posterior distribution such as (12), which is not a recognizable distribution,

and implement the Metropolis algorithm through a very large number of iterations (we used $n_s = 20{,}000$) so that the Bayes estimates converge.

To get the MCMC-based estimate of the renewal function, for each generated sample $i, (i = 1, \ldots, N)$ from $SPP(\alpha, m)$, the following steps are taken:

(a) Generate an initial value α_0 from gamma(P_1, P_2) and set the first entry of the array $\alpha[1] = \alpha_0$. The method for choosing the shape parameter P_1 and the scale paremeter P_2 is outlined in the step 3.

(b) Generate a random observation W_1 from the generalized gamma $(H\alpha_0 + 1, b, 1, a)$, set $m_0 = W_1$, and let the first entry of the array $m[j], (j = 1, 2, \ldots, ns = 20{,}000)$ be $m[1] = m_0$, where ns is the number of iterations based on the MCMC method.

(c) For $j = 2, \ldots, ns$, repeat the following steps:

(i) Generate U_1 from uniform(0,1).

(ii) Generate V_1 from gamma(P_1, P_2).

(iii) Let $p = \text{Min}[\frac{f_1(V_1|m[j-1])h_1(\alpha[j-1])}{f_1(\alpha[j-1]|m[j-1])h_1(V_1)}, 1]$

(iv) If $U_1 < p, \alpha[j] = V_1$, otherwise $\alpha[j] = \alpha[j-1]$, set $\alpha_0 = \alpha[j]$

(v) Generate W_1 from the generalized gamma $(H\alpha_0 + 1, b, 1, a)$, set $m[j] = W_1$, and $m_0 = m[j]$

(vi) Let $j = j + 1$ and go to (i).

The first initial $\tau = 5{,}000$ estimates of α and m are discarded as burned-in values, and Bayes estimates of the parameters based on sample $i(i = 1, 2, \ldots, N)$ are found as

$$\hat{m}_{iB} = \frac{\sum_{j=\tau+1}^{ns} m[j]}{ns - \tau}, \quad \hat{\alpha}_{iB} = \frac{\sum_{j=\tau+1}^{ns} \alpha[j]}{ns - \tau}.$$

5. Given $\hat{m}_{iB}$ and $\hat{\alpha}_{iB}$, the code uses (4) and the same process as in step 2 above (except this time Bayes estimates of the parameters α and m are used in computations) to compute an approximate Bayes estimate $M_{(i,B)}(t)$ of the renewal function. The mean of N Bayes estimates for the renewal function and ASE are computed

$$\overline{M}_{Bayes}(t) = \frac{\sum_{i=1}^{N} M_{(i,B)}(t)}{N},$$

$$\xi_B(\hat{M}(t)) = \frac{\sum_{i=1}^{N} (M_{(B,i)}(t) - M(t))^2}{N}$$

The following tables provide a summary of the simulations. To compare the accuracy of the ML method versus the Bayesian method, the same sample $i(i = 1, 2, \ldots, N)$ is used to compute both $M_{(i,B)}(t)$ and $M_{i,ML}(t)$. The line over a parameter estimate such as $\hat{m}_{ML}$ denotes the average of estimated values in simulations.

Tables 1 and 3 reveal that as sample size H increases, the ASE of the ML and Bayes estimates for m, α, and $M(t)$ decrease. Also, we can see that for a small sample size ($H < 100$), the Bayes estimator of $M(t)$ outperforms the ML, as the ASE of the Bayes estimator is much smaller. Furthermore, for large values of H, the accuracies of the ML and Bayes estimators are comparable.

For the same generated samples used in Tables 1 and 3, we test the goodness of fit of the generated samples using Mathematica. Four distributions are considered: SPP, gamma, exponential, and inverse Gaussian. Given selected

values of parameters (m, α), 100 samples of size H are generated from the SPP, and the averages of the 100 p-values under the null hypotheses that the samples are from those distributions are listed in Tables 2 and 4. Tables 2 and 4 reveal that, in fact, the SPP is an appropriate distribution for the generated samples, as the p-values under the SPP distribution are large.

Table 5 shows a summary of the simulation studies when ML estimates are used to choose hyperparameter values (see step 3 in Section 5). The conclusion is the same as for Tables 1 and 3. That is, the Bayes estimator is more accurate than the ML estimator when the sample size H is small. For larger samples, they have a similar precision.

To confirm convergence of MCMC-based estimators for m and α, Figures 1 through 4, which are based on a generated sample of size H from the $SPP(m = 4.5, \alpha = 3)$, respectively provide an MCMC trace plot for the Bayes estimate of α ("true"value 3.0), a histogram for estimated values of α, an MCMC trace plot for the Bayes estimate of m ("true"value 4.5), and a histogram for estimated values of m in $n_s - \tau = 15{,}000$ iterations.

6. NUMERICAL EXAMPLES

EM-DAT provides worldwide data on dates and impacts of natural events from 1900 to the present. As an illustration of computations, we consider flood and tornado insurance damage in the United States from the year 2000 to the year 2020. The detailed flood and tornado data sets and examples of interarrival calculations are given in the appendix. Interarrivals between two insured damages are defined as the difference between the end date of an occurrence and the start date of the subsequent occurrence. To confirm that, in fact, both data sets fit $SPP(\alpha, m)$, we use the goodness-of-fit test from Gulati and Shapiro (2008), given below. A review of goodness-of-fit tests for the SPP distribution can be found in Chu, Dickin, and Nadarajah (2019). The data sets are sorted for computational convenience to find the ML estimate of m. Using notation from Chu, Dickin, and Nadarajah (2019), let

$$\Lambda_0 = \Lambda_1^2 + \Lambda_2^2$$

where

$$\Lambda_1 = \sqrt{\frac{12}{H-1}}(\overline{U} - .5),$$

$$\Lambda_2 = \sqrt{\frac{5}{4(H+2)(H-1)(H-2)}}$$

$$(H - 2 + 6H\overline{U} - 12 \sum_{i=1}^{H-1} \frac{iU_i}{H-1}),$$

$$\overline{U} = \sum_{i=1}^{H-1} \frac{U_i}{H-1}, \quad Y_i^* = \sum_{j=1}^{i} Y_i,$$

$$U_i = \frac{Y_i^*}{Y_n^*}, \quad Y_i = (H - i + 1)[X_{i:H} - X_{i-1,H}],$$

$$X_{0:H} = 0.$$

Under the null hypothesis H_0: the sample is from $SPP(\alpha, m)$, the test statistic is $\Lambda_0 \sim \chi_2^2$, and H_0 is rejected if $\Lambda_0 > \chi_{\nu,2}^2$, where ν is a significance level. The following

Table 1. Accuracy of ML and Bayes estimators for m, α, and $M(t)$

H	$\overline{\hat{m}_{ML}}$	$\xi_{ML}(\hat{m})$	$\overline{\hat{\alpha}_{ML}}$	$\xi_{ML}(\hat{\alpha})$	$\overline{\hat{M}(t)_{ML}}$	$\xi_{ML}(\hat{M}(t))$
20	4.54	.004	3.81	1.27	17.59	35.05
50	4.52	.001	2.93	.112	13.76	5.41
100	4.51	.0006	2.95	.057	14.06	.281
1000	4.50	4.54×10^{-6}	3.01	.011	14.12	.369
5000	4.50	1.08×10^{-7}	3.01	.001	14.12	.344
10000	4.50	4.37×10^{-8}	2.99	.0006	14.12	.344
H	$\overline{\hat{m}_B}$	$\xi_B(\hat{m})$	$\overline{\hat{\alpha}_B}$	$\xi_B(\hat{\alpha})$	$\overline{\hat{M}(t)_B}$	$\xi_B(\hat{M}(t))$
20	4.50	3.38×10^{-10}	3.00	.00008	14.12	.344
50	4.50	3.46×10^{-11}	2.99	.00006	14.12	.344
100	4.50	2.24×10^{-12}	2.98	.00003	14.12	.344
1000	4.50	1.95×10^{-12}	2.99	.00002	14.12	.344
5000	4.50	1.74×10^{-12}	2.98	.00001	14.12	.344
10000	4.50	1.650×10^{-12}	2.99	.00001	14.12	.344

Table 2. P-values for goodness of fit for SPP, gamma, exponential, and IG

H	SPP	Gamma	Exponential	IG
20	0.4598	0.3199	0.0004	0.3637
50	0.4763	0.0386	3.58×10^{-6}	0.0587
100	0.4926	0.0040	7.54×10^{-11}	0.0086
1000	0.4988	3.33×10^{-17}	5.07×10^{-16}	4.11×10^{-17}
5000	0.4692	7.28×10^{-16}	5.04×10^{-15}	5.06×10^{-16}
10000	0.4819	2.16×10^{-15}	9.92×10^{-15}	1.50×10^{-15}

($m = 4.5$, and $\alpha = 3.0$)

Table 3. Accuracy of ML and Bayes estimators for m, α, and $M(t)$

H	$\overline{\hat{m}_{ML}}$	$\xi_{ML}(\hat{m})$	$\overline{\hat{\alpha}_{ML}}$	$\xi_{ML}(\hat{\alpha})$	$\overline{\hat{M}(t)_{ML}}$	$\xi_{ML}(\hat{M}(t))$
20	2.25	.005	2.13	.165	23.55	35.70
50	2.22	.001	2.12	.087	25.47	5.55
100	2.20	.0002	1.99	.056	25.23	2.80
1000	2.20	8.46×10^{-7}	1.99	.004	24.95	.00009
5000	2.20	4.47×10^{-8}	1.99	.001	24.96	4.84×10^{-6}
10000	2.20	1.98×10^{-8}	2.00	.0005	24.96	2.14×10^{-6}
H	$\overline{\hat{m}_B}$	$\xi_B(\hat{m})$	$\overline{\hat{\alpha}_B}$	$\xi_B(\hat{\alpha})$	$\overline{\hat{M}(t)_B}$	$\xi_B(\hat{M}(t))$
20	2.20	6.04×10^{-10}	1.99	.00005	24.96	6.51×10^{-8}
50	2.20	1.25×10^{-10}	2.00	.00004	24.96	1.34×10^{-8}
100	2.20	1.0210^{-10}	1.99	.00003	24.96	1.11×10^{-8}
1000	2.20	1.56×10^{-11}	2.00	.00002	24.96	1.65×10^{-9}
5000	2.20	1.24×10^{-11}	2.00	.00001	24.96	1.34×10^{-9}
10000	2.20	1.23×10^{-11}	2.00	.00001	24.96	1.32×10^{-9}

($m = 2.2$, $\alpha = 2.0$, $t = 50$, $M(t) = 25.05$)
($a = 2.199$, $b = .001$, $c = 200$, $d = .01$, $P_1 = 1000$, $P_2 = .002$)

Mathematica code is used to find the p-value associated with a calculated value of Λ_0.

```
x: is the list for the sample observations
```

H is the sample size

```
Array[Y, H];
```

Table 4. P-values for GOF for SPP, Gamma, Exponential, IG

| H	SPP	Gamma	Exponential	IG
20	0.4633	0.2910	0.0042	0.3340
50	0.5023	0.0393	0.0004	0.0697
100	0.5393	0.0021	1.58×10^{-6}	0.0060
1000	0.5144	3.11×10^{-17}	1.64×10^{-16}	4.11×10^{-17}
5000	0.5012	1.10×10^{-15}	3.09×10^{-15}	6.86×10^{-16}
10000	0.5333	3.02×10^{-15}	6.21×10^{-15}	2.34×10^{-15}

($m = 2.2$, and $\alpha = 2$)

Table 5. Accuracy of Bayes estimators for m, α, and $M(t)$ when MLEs are used to choose hyper-parameters

H	$\overline{\hat{m}_{ML}}$	$\xi_{ML}(\hat{m})$	$\overline{\hat{\alpha}_{ML}}$	$\xi_{ML}(\hat{\alpha})$	$\overline{\hat{M}(t)_{ML}}$	$\xi_{ML}(\hat{M}(t))$
20	2.26	.009	2.12	.095	23.82	18.34
50	2.22	.0009	2.23	.155	26.65	13.81
100	2.21	.0002	2.11	.047	24.83	.029
1000	2.20	2.01×10^{-6}	1.99	.002	24.95	.0002
H	$\overline{\hat{m}_{B}}$	$\xi_{B}(\hat{m})$	$\overline{\hat{\alpha}_{B}}$	$\xi_{B}(\hat{\alpha})$	$\overline{\hat{M}(t)_{B}}$	$\xi_{B}(\hat{M}(t))$
20	2.20	.008	2.26	.074	24.27	.85
50	2.20	.0009	2.19	.048	24.74	.10
100	2.21	.0002	2.20	.041	24.83	.029
1000	2.20	1.99×10^{-6}	2.21	.001	24.95	.0002

($m = 2.2, \alpha = 2.0, t = 50, M(t) = 25.05$)
($a = \hat{m} - .001, b = \frac{\hat{m}-a}{H\hat{\alpha}}, c = 100\hat{m}, d = .01, P_1 = 1000, P_2 = \frac{\hat{m}}{1000}$)
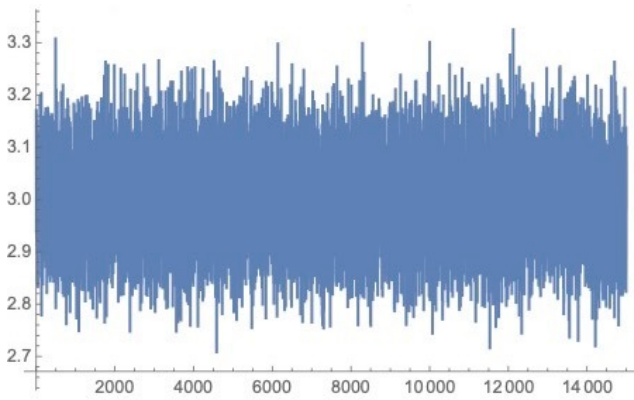
Figure 1. MCMC Chain PLot for Bayes Estimate of α
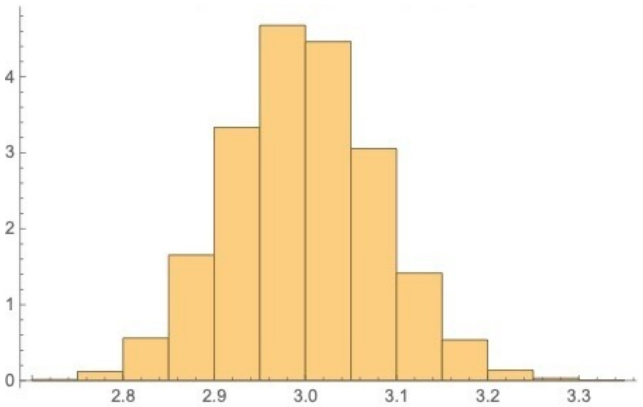
Figure 2. Histogram of α(Bayes)

```
Array[U, H];

Array[YS, H];

Y[1] = H*x[[1]];

YS[1] = Y[1];

j = 2;

While[j < H + 1,

Y[j]= (H - j + 1) (x[[j]] - x[[j - 1]]);

YS[j] = Sum[Y[i], {i, 1, j}];
```

```
j++;

];

i = 1;

While[i < H + 1,

U[i] = YS[i]/YS[H];

i++;

];

Ubar = Sum[U[i]/(H - 1), {i, 1, H - 1}];
```
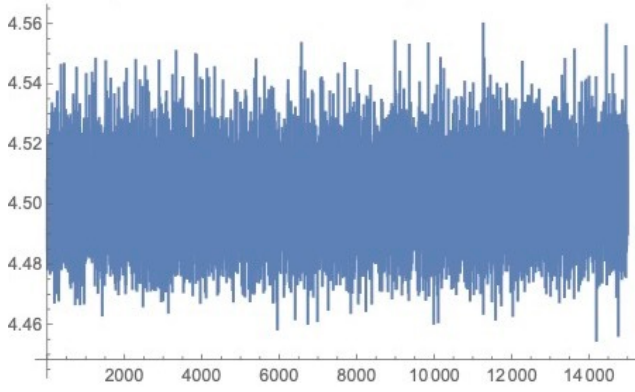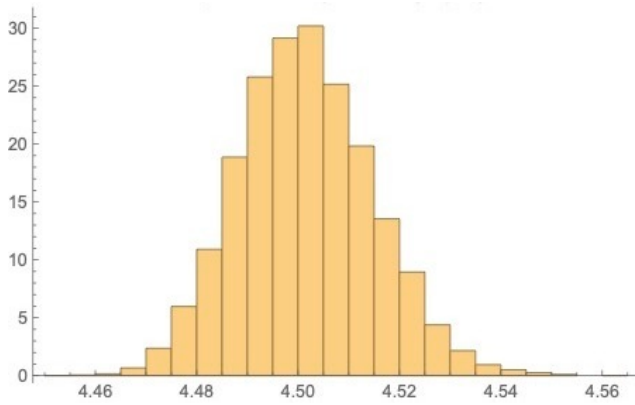
Figure 3. MCMC Chain Plot for Bayes Estimate of m

Figure 4. Histogram of m(Bayes)

```
Λ₁ = Sqrt[12/(H - 1)] (Ubar - .5);

Λ₂ = Sqrt[5/(4 (H + 2) (H - 1) (H - 2))]
  (H - 2 + 6 HUbar - 12 Sum[(i U[i])/(H - 1),
  {i, 1, H - 1}]);

Λ₀ = Λ₁² + Λ₂²

P-value = 1 - CDF[ChiSquareDistribution[2], Λ₀]
```

FLOOD DATA

The following denote $X_i, i = 1, 2, \ldots, 23$ values: 1.1, 1.2, 1.266, 2.066, 2.233, 2.433, 3.633, 5.6, 5.766, 6.5, 6.833, 7.5, 7.6, 8.666, 9.4, 10.733, 13.5, 18.9, 22.433, 23.628, 25.1, 30.996.

For the flood data, we get $\Lambda_0 = .0004$ with p-value = .999. Therefore, there is no significant evidence against the SPP distribution. We have $H = 23, \hat{m}_{ML} = 1.1, \hat{\alpha}_{ML} = .5959 \approx 1$. As mentioned earlier, we can use ML estimates to choose hyperparameters a, b, c, and d. We let $a = .57, c = 100, d = .01, P_1 = 1,000$, and $P2 = .001$, and b is calculated as 0.00024712. Using the above information, the Mathematica code based on $t = 30$ provides

$\hat{\alpha}_B = .99054, \quad \hat{m}_B = 1.490, \quad \hat{M}_B(30) = 7.28, \quad \hat{M}_{ML}(30) = 9.16$

Table 6 lists the p-values for goodness of fit of the flood data under four null hypotheses that data are from the four distributions (SPP, gamma, exponential, and inverse Gauss-

Table 6. P-values for GOF of Flood Data

SPP	Gamma	Exponential	IG
0.99	.255	.612	.095

Table 7. P-values for GOF of Tornado Data

SPP	Gamma	Exponential	IG
0.998	0.0062	0.0005	.00003

ian). We can see that there is no strong evidence against SPP and exponential; however, the p-value for SPP is the largest among the four p-values.

TORNADO DATA

The following denote $X_i, i = 1, 2, \ldots, 50$ values:

0.033, .1, 0.133, 0.166, 0.266, 0.3, 0.333, 0.366, 0.4, 0.4, 0.466, 0.566, 0.633, 0.766, 0.833, 0.9, 0.966, 1, 1.0333, 1.133, 1.2333, 1.266, 1.3, 1.466, 1.866, 2.066, 3.533, 4.233, 4.266, 5.133, 5.266, 5.333, 6.066, 7.366, 7.7, 8.066, 8.466, 8.466, 8.533, 8.9, 8.9, 9.533, 9.8, 10.8, 10.933, 13.066, 14.1, 16.23, 16.566, 24.133

For the tornado data, we get $\Lambda_0 = .002$ with p-value = .9989. Therefore, there is no significant evidence against the SPP distribution. We have $H = 50, \hat{m}_{ML} = .033$, and $\hat{\alpha}_{ML} = .2427 \approx 1$. As mentioned earlier, we can use ML estimates to choose the hyperparameters a, b, c, and d. We let $a = .03, c = 100, d = .01, P_1 = 1,000$, and $P2 = .001$, and b is calculated as 0.0386637. Using the above information, the Mathematica code based on $t = 2$ provides

$$\hat{\alpha}_B = 0.874733, \quad \hat{m}_B = 0.0410618,$$
$$\hat{M}_B(20) = 54.1, \quad \hat{M}_{ML}(20) = 55.6$$

Table 7 lists the p-values for goodness of fit of the tornado data under four null hypotheses that data are from the four distributions (SPP, gamma, exponential, and inverse Gaussian). We can see that there is no strong evidence against the SPP distribution, as the p-values for the other three distributions are very small.

It is noted that $M(t)$ estimates for the tornado data are larger than for the flood data. This is because the average interarrival for the flood data is 9.36 and for the tornado data, 4.90. Therefore, for a specified t, the number of floods by time t would be smaller than the number of tornados.

We also note that for both of the above data sets, the ML and Bayes estimates for α are less than 1. For computation of estimates of $M(t)$, as outlined in Section 2 using Ramsay (2006), we need α to be a positive integer. Therefore, in the Mathematica code both estimates are rounded up to 1 to obtain estimates for the renewal function. It is also noted that the expected value and the variance of the SPP random variable are defined, respectively, if $\alpha > 1$ and $\alpha > 2$. Hence, even though the first and the second moments are not defined when $0 < \alpha < 1$, the delta moment for SPP(m, α) is defined as

$$E[X^\delta] = \frac{\alpha m^\delta}{\alpha - \delta}, \quad 0 < \delta < \alpha.$$

Table 8. P-values for GOF of a generated sample from SPP

SPP	Gamma	Exponential	IG
0.952	0.0003	3.87×10^{-7}	.006

Table 9. Hurricane data for New Orleans 9/29/2021-1/18/2022

No. of Losses	Date of Loss	Interarrivals(days)
7	9/29/21	2
3	9/30/21	1
1	10/2/21	2
3	10/3/21	1
1	10/12/21	9
1	10/24/21	12
1	10/29/21	5
3	11/15/21	16
2	1/10/22	60
1	1/17/22	7
1	1/18/22	1

A GENERATED SAMPLE FROM SPP

The following data are generated from the SPP distribution using the built-in program in Mathematica, and the following denote $X_i, i = 1, 2, \ldots, 20$ values:

4.53749, 4.62575, 4.63825, 4.72415, 4.84672, 4.8955, 5.15921, 5.23818, 5.248, 5.39497, 5.54926, 5.62348, 5.69356, 5.72578, 5.87845, 6.42667, 6.51168, 8.86315, 9.56578, 16.6426.

For the above sample, we get $\Lambda_0 = .0967$ with p-value = .952. Therefore, as expected, the interarrivals have an SPP distribution. The p-values for goodness of fit of gamma, exponential, and inverse Gaussian, respectively, are .0003, 3.8×10^{-8}, and .006.

$H = 20, T = 35.6201, \hat{m}_{Ml} = 4.53749,$ and $\hat{\alpha}_{Ml} = 3.7226 \approx 4$. Based on ML estimates, we choose appropriate values for the hyperparameters, as follows:

$$a = 4.4, b = \frac{\hat{m}_{Ml} - a}{H\hat{\alpha}_{Ml}} = .00184, c = 400,$$
$$d = .01, , P_1 = 1,000, P_2 = .004.$$

Based on the above information, the Mathematica code computes the ML- and MCMC-based estimates for the renewal function at times $t = 30$ and $t = 50$ as

$$\hat{M}_B(30) = 20.75, \quad \hat{M}_{ML}(30) = 20.70,$$
$$\hat{M}_B(50) = 30.14, \quad \hat{M}_{ML}(50) = 30.11.$$

Table 9 shows data extracted from FEMA.gov (https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v1) for hurricanes hitting New Orleans from September 29, 2021, through January 18, 2022. Table 10 shows the p-values for goodness of fit of the SPP, gamma, exponential, and inverse Gaussian distributions for the third column, interarrivals (days).

Table 10. P-values for goodness of fit of hurricane data for New Orleans

SPP	Gamma	Exponential	IG
0.972	.671	.292	.818

Table 10 reveals that whereas all four distributions can be used for the data, the p-value for the SPP distribution is the largest.

7. SUMMARY

SPP distribution is considered as an inter-arrival distribution of a renewal process. Computational formulas from Ramsay (2006), for the sum of iid random variables from the SPP distribution, are used to provide ML- and MCMC-based Bayes estimates for the renewal function. The Bayes estimates of parameters (m, α) are based on shifted exponential and gamma priors. Two conditional posterior distributions $\pi(\alpha|\underline{x}, m)$ and $\pi(m|\underline{x}, \alpha)$ are derived. Since $\pi(\alpha|\underline{x}, m)$ is not a recognized PDF, an MCMC method based on the Metropolis algorithm is employed. The parameter estimates are then used to compute the ML estimate and an approximate Bayes estimate for the renewal function. Simulation studies confirm that if selected values for hyperparameters of the prior distributions of α and m are consistent with selected "true"values of α and m in simulations, then the Bayes estimator for the renewal function would have a much smaller ASE and it outperforms the MLE with regard to accuracy for small samples. Also, for large samples, simulations confirm that the accuracies of the ML and Bayes estimators are comparable. In practice, only one sample is available from the SPP distribution, and choosing

hyperparameters that are consistent with the ML estimates of α and m, as outlined in Section 5, would provide more stable Bayes estimates for the renewal function, as demonstrated in Table 5. For illustration purposes of the computations, three data sets are considered: two data sets from EM-DAT and one data set from FEMA.gov for hurricanes hitting New Orleans.

REFERENCES

Aminzadeh, M.S. 2011. "Bayesian Estimation of Renewal Function for Inverse-Gaussian Renewal Process." *Journal of Statistical Computation and Simulation* 81 (3): 331–41.

Aminzadeh, M.S., and M. Deng. 2018. "Bayesian Predictive Modeling for Exponential-Pareto Composite Distribution." *Variance Journal* 12 (1): 59–68.

———. 2019. "Bayesian Predictive Modeling for Inverse Gamma-Pareto Composite Distribution." *Communications in Statistics - Theory and Methods* 48 (8): 1938–54. https://doi.org/10.1080/03610926.2018.1440595.

———. 2021. "Bayesian Estimation of Ruin Probability Based on NHPP Claim Arrivals and Inverse-Gaussian Distributed Claim Aggregates." *Communications in Statistics-Theory and Methods* 50: 4096–4118.

Brooks, S., A. Gelman, G. Jones, and X.L. Meng. 2011. *Handbook of MCMC*. CRC Press.

Casella, G., and R. Berger. 2004. *Statistical Inference*. Cengage.

Chu, J., O. Dickin, and S. Nadarajah. 2019. "A Review of Goodness of Fit Tests for Pareto Distributions." *Journal of Computational and Applied Mathematics* 361: 13–41. https://doi.org/10.1016/j.cam.2019.04.018.

Deng, M., and M.S. Aminzadeh. 2020. "Bayesian Predictive Analysis for Weibull-Pareto Composite Model with an Application to Insurance Data." *Communications in Statistics - Simulation and Computation* 51 (5): 2683–2709. https://doi.org/10.1080/03610918.2019.1699572.

Fras, M., J. Mohorko, and Z Cuce. 2010. "Modeling of Measured Self-Similar Network Trafiic in Open Simulation Tool." *Informacije* 40 (3): 224–31.

Garsva, E., N. Paulauskas, G. Grazulevicius, and L. Gulbinovic. 2014. "Packet Inter-Arrival Time Distribution in Academic Computer Network." *Electronics and Electrical Engineering* 20 (3): 87–90. https://doi.org/10.5755/j01.eee.20.3.6683.

Gulati, S., and S. Shapiro. 2008. "Goodness-of-Fit Tests for Pareto Distribution." *Stat. Models Methods Biomed. Tech. Syst.*, 259–74.

Hanafy, Mohamed. 2020. "Modeling Bursts and Heavy Tails in Inter-Arrival Claims in Non-Life Insurance." *Journal of Financial Risk Management* 9 (3): 314–33. https://doi.org/10.4236/jfrm.2020.93017.

Hardy, M.R. 2006. *An Introduction to Risk Measures for Actuarial Applications*. Reproduced by the Casualty Actuarial Society and the Society of Actuaries.

Harrison, A.O., F.A. Oladeji, B.C. Benjamin, C.C. Okolie, and F.M. Okikiola. 2014. "The Desirability of Pareto Distribution for Modeling Modern Internet Traffic Characteristics." *International Journal of Novel Research in Engineering and Applied Science* 1 (1).

Ji, Min, Mostafa Aminzadeh, and Min Deng. 2020. "Predictive Modeling of Threshold Life Tables." *North American Actuarial Journal* 24 (2): 316–32. https://doi.org/10.1080/10920277.2019.1633929.

Kim, D.H., S.G. Kang, and W.D. Lee. 2009. "Non-Informative Priors for Pareto Distributions." *Journal of the Korean Data & Information Science Society* 20 (6): 1213–23.

Martino, Luca. 2018. "A Review of Multiple Try MCMC Algorithms for Signal Processing." *Digital Signal Processing* 75 (April): 134–52. https://doi.org/10.1016/j.dsp.2018.01.004.

Necir, Abdelhakim, Abdelaziz Rassoul, and Djamel Meraghni. 2010. "POT-Based Estimation of the Renewal Function of Interoccurrence Times of Heavy-Tailed Risks." *Journal of Probability and Statistics* Special issue: 1–14. https://doi.org/10.1155/2010/965672.

Noor, F., and M. Aslam. 2012. "Bayesian Analysis of Two-Parameter Pareto Mixture Using Censoring." *International Journal of Physical Sciences* 7 (44): 5878–91.

Ramsay, Colin M. 2006. "The Distribution of Sums of Certain I.I.D. Pareto Variates." *Communications in Statistics - Theory and Methods* 35 (3): 395–405. https://doi.org/10.1080/03610920500476325.

Ross, S.M. 2007. *Introduction to Probability Models*. 9th ed. Academic Press.

Simon, Lydia, and Jost Adler. 2021. "Worth the Effort? Comparison of Different MCMC Algorithms for Estimating the Pareto/NBD Model." *Journal of Business Economics* 92 (4): 707–33. https://doi.org/10.1007/s11573-021-01057-6.

Sun, Fupeng, Yueqi Cao, Shiqiang Zhang, and Huafei Sun. 2021. "The Bayesian Inference of Pareto Models Based on Information Geometry." *Entropy* 23 (1): 45. https://doi.org/10.3390/e23010045.

Woo, Jae-Kyung. 2011. "Refinements of Two-Sided Bounds for Renewal Equations." *Insurance: Mathematics and Economics* 48 (2): 189–96. https://doi.org/10.1016/j.insmatheco.2010.10.013.

Table 11. Date and Insured Damage Losses of Natural Event Flood in USA from 2000 to 2021

Start Year	StartMonth	StartDay	EndYear	EndMonth	EndDay	InsuredDamages ('000US$)
2002	6	30	2002	7	23	200500
2004	7	12	2004	7	14	370000
2005	2	17	2005	2	23	200000
2005	1	7	2005	1	11	200000
2005	10	8	2005	10	16	200000
2006	6	25	2006	7	1	401000
2006	5	11	2006	5	22	62500
2007	8	16	2007	8	27	450000
2008	6	9	2008	6	30	400000
2009	9	20	2009	9	21	200000
2011	4	18	2011	5	23	500000
2011	8	6	2011	8	9	86000
2013	9	12	2013	9	19	200000
2014	12	2	2014	12	5	50000
2014	8	11	2014	8	13	540000
2015	12	15	2016	1	6	200000
2015	5	23	2015	5	30	1500000
2016	4	16	2016	4	19	1000000
2016	5	27	2016	6	2	200000
2016	8	9	2016	8	16	2500000
2016	3	8	2016	3	13	333000
2017	4	28	2017	5	1	800000
2019	3	14	2019	3	31	2500000

APPENDIX

The following method is used to compute interarrival times. Let X_n, the interarrival time between the $(n-1)$st and nth renewal, be defined as the difference between the end date of an occurrence and the start date of the subsequent occurrence. $n = 1, 2, 3, \ldots$. Without loss of generally, we assume that there are 30 days in a month and 360 days in a year. Therefore, interarrival times are found as a fraction of a year.

$$\frac{30 \times \text{number of months} + \text{number of days}}{360}.$$

For example, in Table 3 above, the first flood began at 6/30/2002 and ended at 7/23/2002, and the second flood began at 7/12/2004 and ended at 7/14/2004. Therefore,

$$X_1 = \frac{30 \times 31 \text{ months}}{360} = 2.583 \text{ years.}$$

$2.583 \times 12 = 30.996$ months, which is the last number in the flood data.

$$X_2 = \frac{30 \times 24 \text{ months - 11 days}}{360} = 1.969 \text{ years,}$$

which is also converted to 23.628 months.

Table 12. Date and Insured Damage Losses of Natural Event Tornado in USA from 2000 to 2021

Start Year	StartMonth	StartDay	EndYear	EndMonth	EndDay	InsuredDamages ('000US$)
2000	1	2	2000	1	4	200500
2000	9	20	2000	9	20	62500
2001	2	24	2001	2	24	200500
2002	4	27	2002	5	3	2000500
2002	11	5	2002	11	10	450500
2003	3	18	2003	3	20	450000
2003	5	4	2003	5	10	3200000
2004	3	4	2004	3	7	200000
2006	3	11	2006	3	13	800000
2006	11	15	2006	11	16	62500
2006	12	25	2006	12	25	56000
2007	2	2	2007	2	2	200000
2007	3	1	2007	3	2	450000
2007	5	4	2007	5	8	200000
2007	2	2	2007	2	2	200000
2008	5	22	2008	5	26	1325000
2008	5	10	2008	5	12	450000
2008	2	5	2008	2	6	955000
2008	3	14	2008	3	14	450000
2008	4	9	2008	4	11	800000
2008	4	28	2008	4	28	62500
2009	4	9	2009	4	10	1150000
2009	6	10	2009	6	18	1100000
2009	2	10	2009	2	13	1350000
2009	5	7	2009	5	9	600000
2010	5	12	2010	5	16	2000000
2011	5	29	2011	6	1	450000
2011	4	22	2011	4	29	8000000
2011	4	14	2011	4	16	1500000
2011	5	20	2011	5	25	6900000
2012	3	2	2012	3	4	2500000
2012	2	28	2012	2	29	450000
2012	4	2	2012	4	3	800000
2012	4	14	2012	4	15	910000
2013	5	18	2013	5	22	1800000
2013	5	15	2013	5	17	200000
2013	11	16	2013	11	18	800000
2013	1	29	2013	1	30	200000
2013	5	26	2013	6	6	1425000
2015	4	24	2015	4	28	800000
2015	5	3	2015	5	5	200000
2015	4	7	2015	4	10	990000
2015	3	25	2015	3	26	450000
2015	12	26	2015	12	30	800000
2016	5	8	2016	5	11	800000
2017	8	5	2017	8	8	200000
2017	3	25	2017	3	28	2000000

2017	2	27	2017	3	2	1400000
2018	7	19	2018	7	22	400000
2019	3	3	2019	3	4	140000
2020	4	6	2020	4	9	2200000
2020	4	10	2020	4	14	2600000