BAYESIAN INFERENCE FOR DOUBLE PARETO LOGNORMAL QUEUES

BY PEPA RAMIREZ-COBO[1], ROSA E. LILLO,
SIMON WILSON AND MICHAEL P. WIPER

*CNRS France, Universidad Carlos III de Madrid, Trinity College Dublin and
Universidad Carlos III de Madrid*

In this article we describe a method for carrying out Bayesian estimation for the double Pareto lognormal (*dPlN*) distribution which has been proposed as a model for heavy-tailed phenomena. We apply our approach to estimate the $dPlN/M/1$ and $M/dPlN/1$ queueing systems. These systems cannot be analyzed using standard techniques due to the fact that the *dPlN* distribution does not possess a Laplace transform in closed form. This difficulty is overcome using some recent approximations for the Laplace transform of the interarrival distribution for the $Pareto/M/1$ system. Our procedure is illustrated with applications in internet traffic analysis and risk theory.

1. Introduction. Heavy-tailed distributions have been used to model a variety of phenomena in areas such as economics, finance, physical and biological problems; see Adler, Feldman and Taqqu (1999). In particular, a number of variables in teletraffic engineering, such as file sizes, packet arrivals, etc., have been shown to possess heavy-tailed distributions; this can be found, for example, in Paxson and Floyd (1995). Also, in an actuarial context, insurance claim sizes can often be very large and in such cases, may be modeled as long tailed; see, for example, Embrechts, Klüppelberg and Mikosch (1997). For a detailed review of heavy-tailed distributions, we refer the reader to Sigman (1999).

The Pareto distribution has often been applied to model the heavy-tail behavior of teletraffic variables [Resnick (1997)] and insurance claims [Philbrick (1985)]. In particular, in Ramirez, Lillo and Wiper (2008) a mixture of k Pareto distributions (*k-Par*) is used to model ethernet packets interarrival times. However, although the Pareto distribution often models the tails of a distribution well, it is unimodal and decreasing, which means that it will not model the body of the distribution correctly in many modeling situations as is shown in some of the examples in this paper.

Reed and Jorgensen (2004) recently introduced the double Pareto lognormal (*dPlN*) distribution as a versatile model for heavy-tailed data and considered various frequentist approaches to inference for this distribution. They did not recom-

---

mend the method of moments as an estimation method, and observed that the EM algorithm sometimes encounters convergence problems. In this work we focus on the Bayesian approach, which may be preferred for problems where the interest is not only in inference but also in prediction; see, for example, Robert (2001). The first objective of this paper is thus to develop an algorithm to implement Bayesian inference for the *dPlN* distribution.

The study of congestion in teletraffic systems and of ruin problems in insurance is directly related to the analysis of queueing systems, where the arrival or service process are defined by a heavy-tailed distribution. In this paper we consider the *dPlN*/*M*/1 and *M*/*dPlN*/1 queues, which, to our knowledge, have not been considered before in the literature.

The usual moment generating function approach to obtaining the equilibrium distribution of a queue [Gross and Harris (1998)] is difficult to implement because the *dPlN* distribution lacks a moment generating function in closed form. An alternative, which we shall apply, is based on a direct approximation of the nonanalytical Laplace transform using a variant of the transform approximation method (*TAM*); see Harris and Marchal (1998), Harris, Brill and Fischer (2000) and Shortle et al. (2004). The first version of the *TAM*, known as *Uniform TAM* or *U-TAM*, was implemented in Ramírez, Lillo and Wiper (2008), where estimation of the *k-Par*/*M*/1 queue was considered. In this paper we propose a variant of the *TAM* based on both the *Uniform* and *Geometrical TAM*s. By combining this variant of the *TAM* with the Bayesian inference method for the *dPlN* distribution, we can obtain estimates of queueing properties of interest such as the probability of congestion.

This paper is organized as follows. In Section 2 we review the definition and key properties of the *dPlN* distribution and present an approach to Bayesian inference for this distribution, illustrating our procedure with simulated and real data. In Section 3 we examine the *dPlN*/*M*/1 queueing system and show how the *TAM* approach can be used to approximate the Laplace transform of the *dPlN* distribution. Our results are then applied to a real example of internet traffic arrivals. In Section 4 we study the *M*/*dPlN*/1 queueing system and show how the waiting time distribution of this system can be estimated. We then apply our results to the estimation of the ruin probability given real insurance claims data. Conclusions and possible extensions to this work are considered in Section 5.

2. Bayesian inference for the double Pareto lognormal distribution.

2.1. *The double Pareto lognormal distribution.* A random variable Y is said to have a *Normal Laplace distribution* (*NL*), denoted $Y \sim NL(\alpha, \beta, \nu, \tau^2)$ if $Y = Z + W$, where $Z \sim N(\nu, \tau^2)$, and W is a skewed Laplace distributed variable with

density function

$$
f_W(w|\alpha, \beta) = \begin{cases} \dfrac{\alpha\beta}{\alpha + \beta} e^{\beta w}, & \text{if } w \le 0, \\[2mm] \dfrac{\alpha\beta}{\alpha + \beta} e^{-\alpha w}, & \text{if } w > 0, \end{cases}
$$

independent of Z, for $\alpha, \beta > 0$. The density function of Y is

$$
f_Y(y|\alpha, \beta, \nu, \tau^2) = \frac{\alpha\beta}{\alpha + \beta} \phi\left(\frac{y - \nu}{\tau}\right)
$$
$$
\times \left[R(\alpha\tau - (y - \nu)/\tau) + R(\beta\tau + (y - \nu)/\tau) \right],
$$

where $R(z)$ is the Mill's ratio defined by

$$
(2.1) \qquad\qquad R(z) = \Phi^c(z)/\phi(z),
$$

where $\Phi^c(z) = 1 - \Phi(z)$ and $\phi(z)$ and $\Phi(z)$ are the standard normal density and cumulative distributions respectively.

A random variable, X, is said to have a *dPlN* distribution with parameters $(\alpha, \beta, \nu, \tau^2)$ if $X = \exp(Y)$ where Y is Normal Laplace distributed.

The usual change of variable to the density of Y gives the density of X to be

$$
f_X(x|\alpha, \beta, \nu, \tau^2) = \frac{\alpha\beta}{\alpha + \beta} \left(\frac{1}{x}\right) \phi\left(\frac{\log x - \nu}{\tau}\right)
$$
$$
\times \left[R(\alpha\tau - (\log x - \nu)/\tau) + R(\beta\tau + (\log x - \nu)/\tau) \right].
$$

Also, Reed and Jorgensen (2004) show that the *dPlN*$(\alpha, \beta, \nu, \tau^2)$ can be represented as a mixture as

$$
f_X(x|\alpha, \beta, \nu, \tau^2) = \frac{\beta}{\alpha + \beta} f_1(x|\alpha, \nu, \tau^2) + \frac{\alpha}{\alpha + \beta} f_2(x|\beta, \nu, \tau^2),
$$

where the densities

$$
(2.2) \qquad f_1(x|\alpha, \nu, \tau^2) = \alpha x^{-\alpha-1} e^{\alpha\nu + \alpha^2\tau^2/2} \, \Phi\left(\frac{\log(x) - \nu - \alpha\tau^2}{\tau}\right),
$$

$$
(2.3) \qquad f_2(x|\beta, \nu, \tau^2) = \beta x^{\beta-1} e^{-\beta\nu + \beta^2\tau^2/2} \, \Phi\left(\frac{\log(x) - \nu + \beta\tau^2}{\tau}\right)
$$

are, respectively, the limiting forms (as $\beta \to \infty$ and $\alpha \to \infty$) of the *dPlN*$(\alpha, \beta, \nu, \tau^2)$ distribution.

Reed and Jorgensen (2004) illustrate the form of the *dPlN* density function for various different groups of parameter values. In particular, they show that it exhibits upper power-tail behavior in that $f_X(x) \to kx^{-\alpha-1}$ as $x \to \infty$. The *dPlN* distribution does not possess a moment generating function in closed form. However, if $r < \alpha$, the moment of order r exists:

$$
E(X^r|\alpha, \beta, \nu, \tau^2) = \frac{\alpha\beta}{(\alpha - r)(\beta + r)} e^{r\nu + r^2\tau^2/2}.
$$

Reed and Jorgensen (2004) also illustrate a procedure for frequentist inference for the *dPlN* distribution using the EM algorithm and note that under certain conditions, this approach suffers from problems of convergence. An alternative procedure which has not been examined thus far is to take a Bayesian approach, as we do here.

2.2. *Bayesian inference.* Given a random sample $\mathbf{x} = (x_1, \ldots, x_n)$ from the $dPlN(\alpha, \beta, \nu, \tau^2)$, the goal is to compute a posterior distribution $f(\alpha, \beta, \nu, \tau^2 | \mathbf{x})$. For ease of notation, we define $\boldsymbol{\theta} = (\alpha, \beta, \nu, \tau^2)$ in what follows. It is easier computationally to work with the normal Laplace, hence, we define $\mathbf{y} = (y_1, \ldots, y_n)$, where $y_i = \log(x_i)$, $i = 1, \ldots, n$, and compute the posterior density function $f(\boldsymbol{\theta} | \mathbf{y})$ using the normal Laplace likelihood.

The definition of a normal Laplace random variable $Y \sim NL(\alpha, \beta, \nu, \tau^2)$ suggests the use of a Gibbs sampler where one considers the two components of Y as auxiliary variables to be sampled along with $\boldsymbol{\theta}$ so that sampling $\boldsymbol{\theta}$ then reduces to sampling (α, β) and (ν, τ^2) from distributions with truncated skewed Laplace and Gaussian likelihoods respectively. The classical EM algorithm developed in Reed and Jorgensen (2004) was based on a similar idea, but, as noted earlier, this can show convergence problems.

The conditional distribution of $Z | Y = y, \alpha, \beta, \nu, \tau^2$ is a mixture of two truncated normal variables as stated in the following proposition.

PROPOSITION 1. *The conditional distribution of $Z | Y, \alpha, \beta, \nu, \tau^2$ is a weighted mixture of two truncated normal densities*:

(2.4)
$$f_{Z|y}(z|y, \alpha, \beta, \nu, \tau^2) = \left(R(y_\beta) \frac{\phi(z^\beta)}{\tau \Phi^c(y^\beta)} I_{z \geq y} + R(y_\alpha) \frac{\phi(z^\alpha)}{\tau \Phi^c(y^\alpha)} I_{z < y} \right) \Big/ (R(y_\alpha) + R(y_\beta)), \qquad z \in \mathbb{R},$$

where $R(\cdot)$ is given in (2.1), *and*

$$y_\alpha = \alpha \tau - (y - \nu)/\tau, \qquad y_\beta = \beta \tau + (y - \nu)/\tau,$$

$$y^\alpha = \frac{y - (\nu + \tau^2 \alpha)}{\tau}, \qquad y^\beta = \frac{y - (\nu - \tau^2 \beta)}{\tau},$$

$$z^\alpha = \frac{z - (\nu + \tau^2 \alpha)}{\tau}, \qquad z^\beta = \frac{z - (\nu - \tau^2 \beta)}{\tau}.$$

For a proof of Proposition 1 see Appendix A.

Note now that we can express the skewed Laplace distribution as the difference of two exponential variables, that is,

$$W = E_1 - E_2 \qquad \text{where } E_1 | \alpha \sim \mathcal{E}(\alpha) \text{ and } E_2 | \beta \sim \mathcal{E}(\beta).$$

The following proposition specifies the conditional distribution of $E_1 | W$.

PROPOSITION 2. *The distribution of $E_1|W, \alpha, \beta$ is a truncated exponential with support $[\max\{w, 0\}, \infty)$,*

$$(2.5) \qquad f_{E_1|W}(e_1|w, \alpha, \beta) = \frac{(\alpha + \beta)e^{-(\alpha+\beta)e_1}}{I_{w<0} + e^{-(\alpha+\beta)w}I_{w\geq 0}}$$

for $e_1 > \max\{w, 0\}$.

The proof of Proposition 2 can be found in Appendix B. Given a sample, $(y_1, \ldots, y_n)$ conditional on the parameters $(\alpha, \beta, \nu, \tau^2)$, then we can generate $(z_1, \ldots, z_n)$ from the formula in Equation (2.4). Also, we can define $\mathbf{w} = \mathbf{y} - \mathbf{z}$, $w_1 = y_1 - z_1, \ldots, w_n = y_n - z_n$ and then generate $\mathbf{e}_1 = (e_{1,1}, \ldots, e_{1,n})$ from the formula in Equation (2.5) and define $\mathbf{e}_2 = \mathbf{e}_1 - \mathbf{w}$. To undertake inference for ν, and τ^2, let us suppose that we use a normal, inverse gamma prior distribution

$$(2.6) \qquad \nu|\tau^2 \sim \mathcal{N}\left(m, \frac{\tau^2}{k}\right),$$

$$(2.7) \qquad \frac{1}{\tau^2} \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right).$$

Then, from standard Bayesian theory [see, e.g., Box and Tiao (1973)],

$$\nu|\tau^2, \mathbf{z} \sim \mathcal{N}\left(\frac{km + n\bar{z}}{k + n}, \frac{\tau^2}{k + n}\right),$$

$$\frac{1}{\tau^2}|\mathbf{z} \sim \mathcal{G}\left(\frac{a + n}{2}, \frac{b + (n - 1)s_z^2 + (kn/(k + n))(m - \bar{z})^2}{2}\right),$$

where $\bar{z} = \sum_{i=1}^{n} z_i/n$ and $s_z^2 = \sum_{i=1}^{n}\sum(z_i - \bar{z})^2/(n - 1)$. Also, given gamma priors $\alpha \sim \mathcal{G}(c_\alpha, d_\alpha)$, $\beta \sim \mathcal{G}(c_\beta, d_\beta)$, then

$$(2.8) \qquad \alpha|\mathbf{e}_1 \sim \mathcal{G}(c_\alpha + n, d_\alpha + n\bar{e}_1),$$

$$(2.9) \qquad \beta|\mathbf{e}_2 \sim \mathcal{G}(c_\beta + n, d_\beta + n\bar{e}_2).$$

Of course, many other prior structures are possible. In particular, it might be assumed that ν and τ^2 are independent a priori, or that there is some prior dependence between α, β and ν, τ^2. In the presence of real prior information, the use of such alternative structures could lead to more flexible modeling. However, the main disadvantage is that the semi conjugate structure implied given the proposed prior distributions is lost and more complex MCMC algorithms would have to be used to undertake inference.

Therefore, we can define the following Gibbs algorithm:

1. Set initial values $\alpha^{(0)}, \beta^{(0)}, \nu^{(0)}, \tau^{2(0)}$.
2. For $t = 1, \ldots, T$
   a. For $i = 1, \ldots, n$,

a1. Generate $z_i^{(t)}$ from $f(z|\alpha^{(t-1)}, \beta^{(t-1)}, \nu^{(t-1)}, \tau^{2^{(t-1)}}|y_i)$.

a2. Set $w_i^{(t)} = y_i - z_i^{(t)}$.

a3. Generate $e_{1,i}^{(t)}$ from $f(e_1|w_i, \alpha^{(t-1)}, \beta^{(t-1)})$.

a4. Set $e_{2,i}^{(t)} = e_{1,i}^{(t)} - w_i^{(t)}$.

b. Generate $\tau^{2^{(t)}} \sim f(\tau^2|\mathbf{z}^{(t)})$.

c. Generate $\nu^{(t)} \sim f(\nu|\mathbf{z}^{(t)}, \tau^{2^{(t)}})$.

d. Generate $\alpha^{(t)} \sim f(\alpha|\mathbf{e}_1^{(t)})$.

e. Generate $\beta^{(t)} \sim f(\beta|\mathbf{e}_2^{(t)})$.

In the presence of little prior information, it would appear natural to use a non-informative, improper prior distribution. However, it is easy to show that in this case, the posterior distribution is also improper.

PROPOSITION 3. *If an improper prior distribution for α and β is used in the sense that $\int_a^\infty f(\alpha|\beta)\, d\alpha$ diverges for all $a \geq 0$, $\beta > 0$ or $\int_b^\infty f(\beta|\alpha)\, d\beta$ is a divergent integral for any $b \geq 0$, $\alpha > 0$, then the posterior distribution is also improper.*

The proof of Proposition 3 can be found in Appendix C. This implies that in order to carry out Bayesian inference, it is fundamental to use a proper prior distribution for α, β.

2.3. *Illustration with simulated and real data sets.*

EXAMPLE 1. As an illustration of the proposed Gibbs sampler with simulated data, consider a sample of size 1000, generated from $dPlN(0.25, 0.5, 1, 1)$. The Gibbs algorithm was run for 500,000 iterations with initial values set to $\boldsymbol{\theta}^{(0)} = (0.2625, 0.5529, 1.1992, 0.8147)$, the maximum likelihood estimates. The hyperparameters were set to $m = 0$, $k = 4$ in (2.6), $a = b = 1$ in (2.7) and $c_\alpha = c_\beta = d_\alpha = d_\beta = 1$ in (2.8)–(2.9), and from now on these are the values used in the rest of the examples. In order to avoid high autocorrelation, we did thinning and took one sample out of 50. Gibbs sampler code was written in Matlab and, when run on Intel Core Duo at 2.4 GHz and 2 GB of DDR3 RAM, took approximately 19 minutes to perform 100,000 iterations. Figure 1 illustrates the mixing properties of the algorithm. We found $E(\boldsymbol{\theta}|\mathbf{y}) = (0.2578, 0.4995, 1.065, 1.1848)$ close to the maximum likelihood estimates. In addition, we computed credible intervals and correlations in the posterior as measures of precision of the estimates. Credible intervals (95%) for the parameters α β, ν and τ^2 were

$$C_\alpha = [0.2377, 0.2906], \qquad C_\beta = [0.4702, 0.6401],$$
$$C_\nu = [0.7044, 1.4409], \qquad C_{\tau^2} = [0.7178, 1.7352].$$
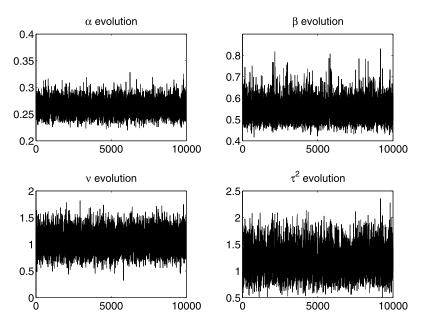
FIG. 1. *MCMC trace plots for Example* 1. *The Gibbs algorithm was applied to a sample of* 1000 *data generated from a dPlN distribution with parameters* $\boldsymbol{\theta} = (0.25, 0.5, 1, 1)$.

With respect to the posterior correlations, we found

$$
\begin{pmatrix}
 & \alpha & \beta & \nu & \tau^2 \\
\alpha & 1 & 0.1449 & 0.5525 & 0.5568 \\
\beta & & 1 & -0.4936 & 0.4534 \\
\nu & & & 1 & 0.2362 \\
\tau^2 & & & & 1
\end{pmatrix}.
$$

Notice that, for example, the parameters β and ν are negatively correlated a posteriori and α and ν positively, a consequence of the definition of a Normal-Laplace distribution as the sum of a normal and skewed Laplace variables.

In Figure 2 the fitted density function, estimated for the data (in log-scale), and almost undistinguishable from the theoretical one, is depicted. The fitted curve has been computed by simple averaging over the Gibbs sampled values, that is, $f_Y(y|\mathbf{y})$ has been estimated by

$$
\frac{1}{T} \sum_{t=1}^{T} f_Y\big(y|\alpha^{(t)}, \beta^{(t)}, \nu^{(t)}, \tau^{2(t)}\big).
$$

We should point out that, if instead of starting the MCMC from the maximum likelihood estimates, we start further from this point, the results are very similar to those obtained starting from the ML estimates, as long as the initial value of α is not very large. It has been observed that, if the starting value of α is large
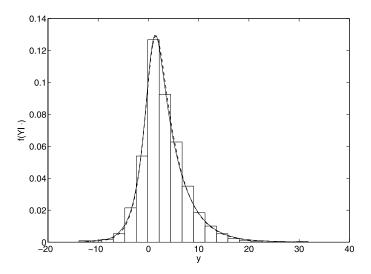
FIG. 2.   *Histogram, fitted (dotted line) and theoretical (solid line) pdf for the simulated data set of Example* 1.

and the sample has long tails (small α, as in this example), then convergence can be extremely slow and the Gibbs algorithm often remains stuck in the tail of the distribution for a long time. Because of this fact we suggest starting the MCMC algorithm with small values for α (not necessarily the ML estimates).

Finally, one may wonder how sensitive the method is to the hyperparameters. We performed several analyses and our experience is that if the real α or β are not very large, then the results are not affected by the choice of hyperparameters. For instance, in this example, we also set $m = 2$, $k = 4$, $a = b = 2$, $c_\alpha = c_\beta = 0.5$, and $d_\alpha = d_\beta = 0.2$, and found $E(\boldsymbol{\theta}|\mathbf{y}) = (0.2609, 0.5005, 1.1833, 1.0157)$ with credible intervals

$$C_\alpha = [0.2440, 0.2927], \qquad C_\beta = [0.4373, 0.5988],$$
$$C_\nu = [0.9188, 1.6039], \qquad C_{\tau^2} = [0.6210, 1.6226],$$

whose lengths are very similar to that found with the first choice of hyperparameters. Also, the fit to the data is almost the same as in Figure 2. The next example illustrates the performance of the method when α and/or β are large.

EXAMPLE 2.   Reed and Jorgensen (2004) state that if there is evidence in the analyzed data of heavy-tailed behavior just in one tail, then it is better to fit one of the limiting components f_1 (2.2) or f_2 (2.3); otherwise, a frequentist approach may result in the nonconvergence of the optimization algorithm. Here we apply the proposed Bayesian procedure to analyze simulated data from a $dPlN(\alpha, \beta, \nu, \tau^2)$ with large α, β. Specifically, we consider three data sets S1., S2. and S3., simulated from $dPlN(10, 0.5, 1, 1)$ (left heavy tail), $dPlN(0.5, 10, 1, 1)$ (right heavy

TABLE 1
Starting values (MLE) and posterior estimates for the considered simulated data S1., S2. *and* S3.
in Example 2, *where* α *or/and* β *take large values. Also, credible intervals for the large*
parameters are shown

	S1: *dPlN*(10, 0.5, 1, 1)	S2: *dPlN*(0.5, 10, 1, 1)	S3: *dPlN*(10, 10, 1, 1)
$\theta_0 = \theta_{MLE}$	(4.34, 0.50, 0.83, 1.05)	(0.56, 4.53, 1.28, 1.04)	(4.67, 5.53, 0.92, 0.91)
$E(\theta\|\mathbf{y})$	(22.74, 0.49, 1.09, 1.07)	(0.55, 3.81, 1.32, 1.02)	(40.81, 1.8921, 1.52, 0.81)
C_α	[1.5811, 30.4547]	—	[3.0217, 50.9491]
C_β	—	[1.7914, 9.7827]	[1.4307, 3.9651]

tail) and *dPlN*(10, 10, 1, 1) (similar to a Normal distribution but with heavier tails)
distributions, respectively. We assumed the same hyperparameters as in Example 1,
$(m, k, a, b, c_\alpha, c_\beta, d_\alpha, d_\beta) = (0, 4, 1, 1, 1, 1, 1, 1)$. Table 1 shows the starting val-
ues θ_0 (ML estimates), posterior estimates $E(\theta\|\mathbf{y})$ and 95% credible intervals for
the large parameters. We would like to point out the high variability in the in-
tervals, especially if α is large. However, as it can be seen in Figure 3, both the
frequentist and Bayesian approaches perform similarly when fitting the pdf to the
histogram of the data. This indicates that, as pointed out by Reed and Jorgensen
(2004), when α or β are large, the density function approaches to the three para-
meters limit case f_2 (2.3) or f_1 (2.2), and, thus, there is small difference in the
dPlN density function between multiple values of α or β.

To show the versatility of the *dPlN* model, we next consider two real data sets
from the insurance and internet context, respectively.

EXAMPLE 3. The first data set has been analyzed in Beirlant et al. (1998)
and Beirlant et al. (2004) and and can be found in http://lstat.kuleuven.be/Wiley/.
This contains 1668 claim sizes (expressed as a fraction of the sum insured) from
a fire insurance portfolio provided by the reinsurance brokers Boels & Bégaul Re
(AON). The data concern claim information from office buildings. Next to the
size of the claims, the sum insured per building was provided. The Gibbs sampler
was run under the same conditions as in the simulated-data example and posterior
estimates $E(\theta\|y) = (0.51, 4.99, 7.78, 0.76)$ were found. Note that the posterior
estimate for α indicates a clear long tail. Figure 4 shows the fit to the histogram
of the data in log-scale of the *dPlN* model (solid line) in comparison with the fit
provided by a mixture of Pareto distributions (dashed line), where the number of
the components in the mixture, k, may change at each iteration. Estimation for
the *k-Par* distribution was undertaken in Ramírez, Lillo and Wiper (2008), and
as it was commented in Section 1, here the Pareto (or mixture of Pareto) distri-
bution fails to capture the body of the distribution. In addition, the Bayesian ap-
proach considered in Ramírez, Lillo and Wiper (2008) is more time consuming
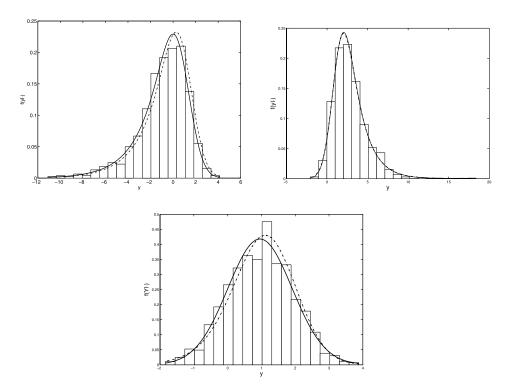
FIG. 3.   *Fitted pdfs using the ML values* (*solid line*) *and the posterior estimates from the Bayesian approach* (*dashed line*), *for data sets* S1., S2. *and* S3. *in Example* 2.

than the Gibbs sampler developed here. That algorithm was based on a Birth–Death MCMC method, where at each iteration a Metropolis–Hastings step is carried out. The Gibbs sampler has a number of well-known advantages over standard Metropolis–Hastings samplers. For example, the Gibbs sampler requires no tuning, which for Metropolis–Hastings algorithms can be time consuming—especially for long data sets where the algorithm takes longer to run.

EXAMPLE 4.   The second real example that we consider is from the tele-traffic context. It can be found in the *Internet Traffic Archive* (BC trace), http://www.sigcomm.org/ITA/, where 4 million packet traces of LAN and WAN traffic seen on an Ethernet at the Bellcore Morristown Research and Engineering facility are recorded. The considered trace, BC-pAug89, began at 11.25 on August 29, 1989, and ran about 3142 seconds (until 1 million packets had been captured). The measurement techniques in making the traces are described in Leland and Wilson (1991) and are a subset of those analyzed in Leland et al. (1994). The data set analyzed here consists of the measured transferred bytes/sec within the 3142 consecutive seconds.
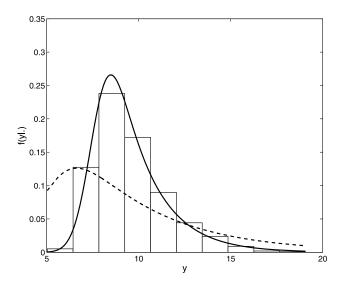
FIG. 4. *Histogram and fitted pdf in Example* 3, *for the Aon data set* (*claim sizes in a fire insurance portfolio*) *in log-scale, under the dPlN model* (*solid line*) *and mixture of Pareto components model* (*dashed line*).

We applied the Gibbs algorithm and found posterior estimates $E(\theta|y) = (8.59, 4.52, 11.83, 0.59)$. The mode of this data set is not close to zero, as can be observed in Figure 5, and, thus, the mixture of Pareto distributions shows a poor performance. Here again, the *dPlN* model performs well, not only capturing the tail but also the body of the set, as can be seen in the same figure.

Thus, from our experience the *dPlN* distribution has two advantages over the *k-Par* for fitting heavy-tailed data: first, it is able to capture both the tail and body of the distribution, and second, the estimation procedure for fitting the *dPlN* distribution is faster computationally than that proposed in Ramírez, Lillo and Wiper (2008), for the *k-Par* density.

3. Inference for the *dPlN*/*M*/1 queueing system. In this section we shall consider the *dPlN* distribution as a model for the arrival process in a single-server queueing system with independent, exponentially distributed service times. The next section reviews this queueing system, denoted as *dPlN*/*M*/1.

3.1. *The dPlN*/*M*/1 *queueing system.* The *dPlN*/*M*/1 system is an example of the *G*/*M*/1 queueing system, whose properties are well known [see Gross and Harris (1998)]. In particular, for the *dPlN*/*M*/1 system with parameters $\theta = (\alpha, \beta, \nu, \tau^2)$, standard results for *G*/*M*/1 queues imply that the mean inter-arrival time does not exist if $\alpha \leq 1$. In this case, the queueing system is automatically stable whatever the service rate μ [that is, $E(S) = 1/\mu$, where S denotes the
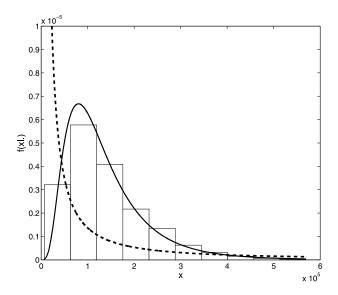
FIG. 5. *Histogram and fitted pdf in Example* 4, *for the teletraffic data set* (*number of bytes per second*), *under the dPlN model* (*solid line*) *and mixture of Pareto components model* (*dashed line*).

service time]. Otherwise, the traffic intensity is given by

$$(3.1) \qquad \rho = \frac{(\alpha - 1)(\beta + 1)}{\mu \alpha \beta e^{\nu + \tau^2/2}}.$$

If the system is stable ($\rho < 1$), then the steady-state probability for the number of customers Q in the system just before an arrival, the stationary time W_q spent queueing for service, and time W spent in the system are

$$P(Q = n) = (1 - r_0)r_0^n \qquad \text{for all } n \in \mathbb{N},$$

$$P(W_q \leq x) = 1 - r_0 e^{-\mu(1-r_0)x},$$

$$P(W \leq x) = 1 - e^{-\mu(1-r_0)x},$$

where $r_0 \in (0, 1)$ is the unique real root of the equation

$$(3.2) \qquad r_0 = f^*\big(\mu(1 - r_0)\big),$$

and $f^*(\cdot)$ is the Laplace–Stieltjes transform of the interarrival-time density function $f(\cdot)$ defined as

$$f^*(s) = \int_0^\infty e^{-sx} f(x) \, dx \qquad \text{for } \text{Re}(s) > 0.$$

However, the Laplace transform of the *dPlN* distribution is analytically intractable so that the standard techniques for finding the root of Equation (3.2) cannot be applied. Thus, an alternative approach to obtaining the steady state distributions is needed. The next section outlines such an approach.

3.2. *A variety of the transform approximation method.* The transform approximation method (*TAM*) was developed informally by Harris and Marchal (1998) and Harris, Brill and Fischer (2000) for the case of approximating the Laplace transform of the single parameter Pareto distribution and was later extended by Shortle et al. (2004). Here we describe the approach in the case of the *dPlN* distribution. To approximate the Laplace transform $f^*(s)$ of the distribution of a random variable X, the basic algorithm is as follows:

1. Pick a set of N probabilities, p_i, $0 < p_1 < \cdots < p_N < 1$.
2. Find the quantile t_i of order p_i, $P(X \le t_i) = p_i$.
3. Assign to each point t_i the probability

$$w_1 = \frac{p_1 + p_2}{2},$$

$$w_i = \frac{p_{i+1} - p_{i-1}}{2} \qquad \text{for } i = 2, \dots, N-1,$$

$$w_N = 1 - \frac{p_{N-1} + p_N}{2}.$$

4. Approximate the Laplace Transform $f^*(s)$ by $f_N^*(s) = \sum_{i=1}^{N} w_i e^{-st_i}$.

For the *dPlN* case and once the probabilities p_i have been selected, the quantiles in step 2 are approximated numerically by Newton–Raphson, with initial values obtained from the empirical distribution function of the data.

Harris, Brill and Fischer (2000) and Shortle et al. (2004) consider different alternatives for the defining probabilities p_i, although, as they point out, the choice of the optimal probabilities is an open question. The natural approach, known as uniform *TAM* or *U-TAM*, is to define uniform probabilities, $p_i = (i - 1)/N$. However, this approach leads to poor approximations in the tail of the distribution. An alternative algorithm applied in Shortle et al. (2004), which better captures heavy-tailed behavior, is the geometric or *G-TAM* algorithm which sets $p_i = 1 - q^i$, for $q \in (0, 1)$. But even when $q \to 0$, few quantiles are selected from the body of the distribution and a poor approximation of this part may be obtained with this approach.

We have found that a combination of both algorithms works better than applied separately. We used the *U-TAM* algorithm to obtain a proportion r of percentiles from the body of the distribution and the *G-TAM* algorithm is used to find the other $(1 - r)$ proportion of percentiles covering the heavy tail. We consider that the body of the distribution is defined by those percentiles t_i such that $P(X \le t_i) \le P(X \le E[X])$, in the case that $E[X]$ exists (otherwise, we use the median). Other alternatives (with larger quantiles) may be used, but in practice we have found that it makes little difference.

Formally, if r denotes the proportion of percentiles before $E[X]$, and q is the geometric rate, then $(r, q) \in \{r_{min}, \dots, r_{max}\} \times \{q_{min}, \dots, q_{max}\}$ form a grid where

the optimal value $(r^\star, q^\star)$ is chosen so that the *TAM* mean $(\sum_{i=1}^N w_i t_i)$ (or the *TAM* median: $t_a / \sum_{i=1}^a w_i \leq 0.5$ and $\sum_{i=1}^{a+1} w_i > 0.5$) matches the mean (or median) of the original distribution. In our examples we have found that a grid of size 8×17 is enough to get a distance less than 10^{-3} between the *TAM* mean/median and the theoretical mean/median. The proposed methodology satisfies the conditions of Theorem 1 in Shortle et al. (2004) so that convergence of $f_N^*(s)$ to $f^*(s)$ is assured as $N \to \infty$.

3.3. *Bayesian estimation of the dPlN/M/1 queueing system.*   Given the prior distributions and a sample of *dPlN* distributed interarrival data, we have seen that the Gibbs algorithm can be used to produce a sample of values $\boldsymbol{\theta}^{(t)} = (\alpha^{(t)}, \beta^{(t)}, \nu^{(t)}, \tau^{(t)})$ for $t = 1, \ldots, T$ from the posterior distribution of the *dPlN* parameters.

Supposing now that the service rate, μ, is known, then it is straightforward to estimate the probability that the system is stable,

$$(3.3) \qquad P(\rho < 1 | \mathbf{y}) = \frac{1}{T} \sum_{t=1}^T I(\rho^{(t)} < 1),$$

where $\rho^{(t)}$ is the value of ρ calculated from Equation (3.1) setting $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$ and $I(\cdot)$ is an indicator function. Given that this probability is high, then for each set $\boldsymbol{\theta}^{(t)}$ of generated parameters such that $\rho^{(t)} < 1$, the root $r_0^{(t)}$ can be generated using (3.2) and the *TAM* and, therefore, the conditional posterior distributions of queue size and waiting times, given stability, can be estimated by Rao Blackwellization, that is, by simply averaging over the parameters satisfying the stability condition. Thus, for example, the posterior distribution of queue size $P(Q = n | \mathbf{y})$ is estimated by

$$\frac{1}{S} \sum_{s=1}^S P(Q = n | \boldsymbol{\theta}^{(s)}, \mu),$$

where $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(S)}$ is the set of parameters satisfying the stability condition.

One point to note, however, is that, as commented in Wiper (1997), the means of the fitted equilibrium queue size and waiting time distributions do not exist. This is a typical feature for Bayesian inference in $G/M/\cdot$ or $M/G/\cdot$ queueing systems. Thus, if posterior summaries of these distributions are required, it is preferable to use the median and quantiles.

When the service parameter is unknown, then, given an independent sample of service time data, conjugate inference for the service rate can be carried out as in, for example, Armero and Bayarri (1994). For a Monte Carlo sample, $\mu^{(1)}, \ldots, \mu^{(T)}$ from the posterior distribution of the service rate, the traffic intensity may be estimated by calculating $\rho^{(t)}$ given $(\boldsymbol{\theta}^{(t)}, \mu^{(t)})$ and averaging as in (3.3). In order to condition on the existence of equilibrium, only those parameter sets $(\boldsymbol{\theta}^{(t)}, \mu^{(t)})$ such that $\rho^{(t)} < 1$ are retained.

3.4. *Application to internet traffic analysis.* Internet traffic data has lately become a wide field of study and numerous works have characterized it as having some unusual statistical properties such as self similarity and heavy tails; see, for example, Willinger, Paxson and Taqqu (1998). In particular, as shown in Paxson and Floyd (1995), internet arrival traffic cannot be well modeled by a Poisson process. As an alternative, heavy-tailed distributions can be considered.

Figure 6 shows the histogram of a set of interarrival times (in seconds) of a trace of 1 million ethernet packets, derived from BC-pAug89 in the Internet Traffic Archive (described in Example 3 of Section 2.3). The first (according to the outcome) 50,000 interarrival times (in sec) are analyzed here. Superimposed (in solid line) is the fitted *dPlN* density generated using the Bayesian algorithm described in Section 2. Also superimposed (dashed line) is the fitted Pareto density. In this example the Pareto distribution captures the tail of the distribution but has a poorer performance in the body of the distribution. It can be seen in Ramírez, Lillo and Wiper (2008) that a mixture of two Pareto components provides a good fit of this data set, however, the high computational cost of that algorithm makes this one based on the *dPlN* distribution preferable. The posterior mean parameter estimates for the *dPlN* model were $E(\boldsymbol{\theta}|x) = (2.15, 1.07, -6.00, 0.36)$.

Now we shall consider the queueing aspects. Given the *dPlN* arrival process, we shall assume that arrivals are processed by a single server with exponentially distributed service times with rate μ. Table 2 shows the posterior probability of equilibrium (third column) and the expected value for the traffic intensity (fourth
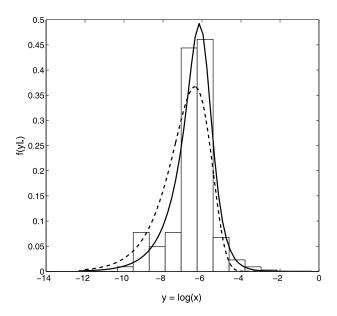

FIG. 6. *Histograms and fitted pdf for the internet data* (50,000 *real interarrival times*) *in log-scale, under the dPlN model* (*solid line*) *and Pareto model* (*dashed line*).

TABLE 2
Probability of equilibrium and traffic intensity. When μ is large (faster service on average), the probability of stability of the system increases

μ	$E(S)$	$\mathbb{P}(\rho < 1|\mathbf{y})$	$\mathbb{E}(\rho|\mathbf{y})$
1500	0.0006	1	0.2616
1000	0.001	1	0.3923
500	0.002	1	0.7844
400	0.0025	1	0.9798
395	0.002531	0.8257	0.9946
394	0.002538	0.7869	0.9969
393	0.002544	0.6115	0.9979
392	0.002510	0.4562	1.0008
391	0.002550	0.4284	1.0040
390	0.002564	0.2519	1.0065
385	0.002597	0	1.0194

column) for an assortment of values of μ [the expected service time is $E(S) = 1/\mu$]. From this table, it is clear that there is a high probability that the system is stable (that is, no congestion occurs) for values of μ greater than 394. Figure 7 depicts the fitted system waiting time W, and queue waiting time W_q, distributions for values of μ greater than 400. Table 3 illustrates the distribution of the number Q of clients in the system in equilibrium. We can see that as the service rate increases (i.e., the service is faster), then the median queueing and system waiting times and the number of clients in the system decrease, as would be expected.
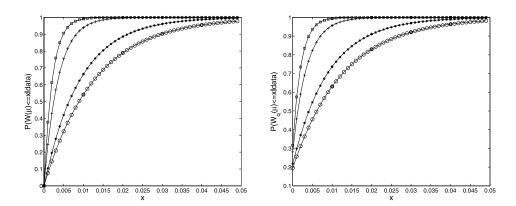

FIG. 7. *Predictive system and queue waiting times distributions for the internet data set for an assortment of service rates (○: $\mu = 400$, *: $\mu = 500$, +: $\mu = 1000$ and □: $\mu = 1500$). As expected, when the service is faster (μ increases), then the probability of waiting less than a short time is larger.*

TABLE 3
Predictive system size distribution just before an arrival for the internet data set, for an assortment of service rates μ. As expected, for faster services (large μ) the probability of an empty system is larger than for slower services

μ	$\mathbb{P}(Q = 0)$	$\mathbb{P}(Q = 1)$	$\mathbb{P}(Q = 2)$	$\mathbb{P}(Q = 3)$
1500	0.3167	0.2161	0.1475	0.1008
1000	0.2813	0.2019	0.1449	0.1042
500	0.2182	0.1703	0.1330	0.1039
400	0.1955	0.1570	0.1260	0.1014
395	0.1948	0.1569	0.1260	0.1013
394	0.1946	0.1565	0.1259	0.1013

In this example we have also compared the queueing results obtained with the *dPlN* model with those obtained from the queueing systems *Pareto*/M/1 and $M/M/1$. Different estimates of the system and queue waiting time distributions under the different queueing models were obtained. The fitted system size distribution just before an arrival among these different queues also varies, for example, the probability that the system size is larger than 2 or than 3, $P(Q > 2)$, $P(Q > 3)$ is larger with the *dPlN* model than with the Pareto or Exponential models. On the contrary, the values $P(Q > 0)$, $P(Q > 1)$ are smaller with the *dPlN* model than with the other ones.

4. The $M/dPlN/1$ queueing system and ruin probabilities. In this section we consider the $M/dPlN/1$ queueing system, with independent, exponentially distributed interarrival times and *dPlN* service times, and show how the Bayesian approach to estimate the *dPlN* can be used to estimate the probability of ruin from actuarial data.

4.1. *The $M/dPlN/1$ queueing system.* The general properties of the $M/G/1$ queueing system are well known; see, for example, Gross and Harris (1998). In particular, if the service time S is assumed to follow a *dPlN* distribution with $\boldsymbol{\theta} = (\alpha, \beta, \nu, \tau^2)$, then, if $\alpha \le 1$, $E(S) = \infty$ and the queueing system is never stable, whatever the interarrival rate λ. When $\alpha > 1$, the traffic intensity is given by

$$\rho = \frac{\lambda \alpha \beta e^{\nu + \tau^2/2}}{(\alpha - 1)(\beta + 1)}.$$

The Laplace transform $W_q^*(s)$ of the equilibrium waiting time in the queue is related to the Laplace transform $B^*(s)$ of the (*dPlN*) service time by

$$W_q^*(s) = \int_0^\infty e^{-st} \, dW_q(t) = \frac{(1 - \rho)s}{s - \lambda(1 - B^*(s))},$$

where $W_q(t)$ is the distribution function of the waiting time. In order to obtain the distribution function of the waiting time $W_q(t)$, we first apply the *TAM* to approximate $B^*(s)$ as earlier. Second, we can use a standard numerical approach to invert the Laplace transform, $W_q^*(s)$; see, for example, Shortle, Fischer and Brill (2007) for a review. In this case, we apply the recursion method by Fischer and Knepley (1977).

4.2. *Application to fire insurance claims.* In an insurance context, it is often assumed that claim sizes, C_i, are independent and identically distributed heavy-tailed random variables; see, for example, Rolski et al. (1999). Here, we shall assume that claim sizes can be modeled as *dPlN* random variables. Often, it is also supposed that the interclaim times, T_i, are independent, exponentially distributed variables with rate λ. Let u denote the initial reserve of an insurance company and let r be the rate at which premium accumulates. Then, the company's wealth, or risk portfolio at time t, is

$$R(t) = u + rt - \sum_{i=1}^{N(t)} C_i,$$

where $N(t) = \sup(n : \sum_{i=1}^{n} T_i \le t)$ is a Poisson counting process with rate λ.

Clearly, the insurance company will be interested in the probability that they may eventually be ruined, given their initial capital and premium rate, that is,

(4.1) $\psi(u, r) = P\big(R(t) < 0 \text{ for some } t \ge 0 \,|\, \text{initial capital } u, \text{ premium rate } r\big).$

If the mean claim size does not exist, then eventual ruin is certain. Otherwise, we can define the traffic intensity of this system as $\rho = \lambda E[C_i]/r$ and it is well known that ruin is certain if $\rho \ge 1$. In the case that $\rho < 1$, then in, for example, Prabhu (1998), it is shown that the ruin probability can be computed as the steady state probability that the waiting time exceeds u/r in a $M/G/1$ queueing system, where the interarrival time and service time distributions are the same as the distributions of T_i and C_i/r respectively. Table 4 shows this duality. Thus, estimating

TABLE 4
Duality between the probability of ruin in a risk theory context
and the $M/G/1$ queueing sytem with steady-state queue
waiting time distribution W_q

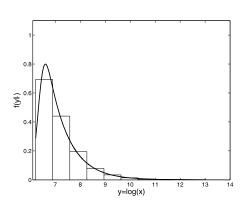| Queueing theory	Risk theory
Interarrival times	Interclaim times
Service times	Claim sizes
$\mathbb{P}(W_q > u)$ for a $M/G/1$	Probability of ruin with initial reserve u

the $M/dPlN/1$ queue allows us to estimate the probability of ruin where the claims sizes are assumed to follow a *dPlN* distribution.

Note that by scaling appropriately, it can be assumed without loss of generality that the premium rate, r, is equal to 1 and we shall do this from now on, writing $\psi(u)$ for the ruin probability of Equation (4.1).

Assuming the $M/dPlN/1$ model and given some initial reserve u and claim arrival rate λ and a sample of claim sizes, then the posterior parameter distribution of the *dPlN* claim size distribution can be estimated using the Bayesian approach as outlined in Section 2 and this can be combined with the *TAM* and recursion algorithms to estimate the ruin probability.

To illustrate this approach, we consider data treated in Beirlant and Goegebeur (2003) and Beirlant et al. (2004) representing 9181 fire claims values for the period 1972–1992 from a Norwegian insurance portfolio. Together with the year of occurrence, the values ($\times 1000$ Krone) of the claims are known. They can be found in http://ucs.kuleuven.be/Wiley/index.html. The left panel of Figure 8 shows the data in log-scale (values of the claims) and the Bayesian *dPlN* fit. The right panel of Figure 8 illustrates the log-transformed fitted Pareto (dotted line) and Exponential (dashed line) models to this data set. Again, the Pareto model does not capture the body of the distribution; the Exponential fit is even worse, it captures neither the body, nor the tail.

Assuming that the system is stable, we can now estimate the ruin probability for different interclaim rates and initial reserves. In this case, the expected claim size, conditional on this existing (i.e., that $\alpha > 1$), is approximately 2915, which implies that in order to avoid extremely high probabilities of ruin, we should typically consider plausible values of λ to be below 1/2915. Figure 9 depicts the posterior probability of ruin, $E(\psi(u)|\text{data})$, for a grid of values of different average interclaim times, $1/\lambda$, and various initial reserve levels, u. As would be expected,
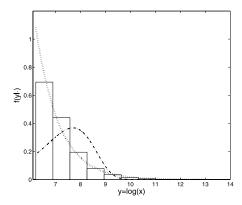

FIG. 8. *Histograms and fitted pdf for the Norwegian data* (*claim sizes*) *in log-scale*, *under the dPlN* (*left panel*, *solid line*), *the Pareto* (*right panel*, *dotted line*) *and Exponential* (*right panel*, *dashed line*) *models.*
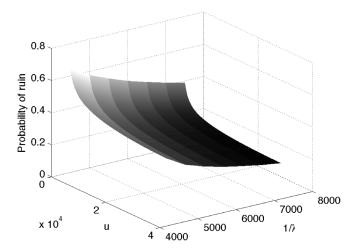
FIG. 9. *Probabilities of ruin (z-axis) for the Norwegian data insurance company, for an assortment of initial reserves u (values from 0 to 4×10^4) and mean interclaims times $1/\lambda$. As would be expected, when the initial reserve is low and the claims occur frequently on average (low values of $1/\lambda$), then the probability of ruin increases.*

when both the initial reserve u and the expected interclaim times $1/\lambda$ are low, then the ruin probability increases.

As we did for the $dPlN/M/1$ queueing system with the teletraffic data set, given theses claim sizes, we have also compared the performance of the $M/dPlN/1$ queue with the $M/Pareto/1$ and $M/M/1$ queueing system, assuming a rate $\lambda = 1/4000$. When fitting a Pareto distribution to the data with a Bayesian approach, it was found that a posteriori, the sampled parameters of the Pareto distribution led to a lack of moment of order one, indicating that, since $E(S|\mathbf{y}) = \infty$, then the corresponding $M/Pareto/1$ system is not stable, given the data. For the $M/M/1$ model something similar was found: $1 < \rho^{(t)} < \infty$ for most of the iterations, and, thus, the posterior probability that the system is stable was very low. Thus, we could not predict the probability of ruin, under these models. Finally, the same comments as in Section 3, concerning the estimation of the arrival rate λ (interclaim times rate) when it is considered as an unknown parameter, can be also applied here.

5. Conclusions. In this work we have developed Bayesian inference for the double Pareto lognormal distribution and have illustrated that this model can capture both the heavy-tail behavior and also the body of the distribution for real data examples. Bayesian inference was implemented with the Gibbs sampler, although, since $\boldsymbol{\theta}$ is only 4 dimensional, several alternatives exist and were attempted. The use of importance sampling was difficult because of a lack of good distributions for the initial sample that avoided degeneracy. A block Metropolis algorithm using a multivariate normal proposal, with covariance matrix estimated by maximum

likelihood, was also attempted but exhibited poor mixing for τ and slower computation time. This suggests that the Gibbs procedure should be preferred.

Second, we have combined this approach with techniques from the queueing literature in order to estimate posterior equilibrium distributions for the $dPlN/M/1$ and $M/dPlN/1$. To do this, we have adapted the transform approximation method, in order to estimate the Laplace transform of the $dPlN$ distribution and the waiting time distribution in the $M/dPlN/1$ system.

Finally, we have illustrated this methodology with real data sets, estimating first waiting times and congestion in internet and computing the probability of ruin in the insurance context, making use of the duality between queues and risk theory. Comparisons with the $M/M/1$, $Pareto/M/1$ and $M/Pareto/1$ have been also carried out. Differences among these queueing systems, especially when the service process is heavy-tailed, were found.

A number of extensions are possible. First, we could extend our results to the case of a multiple number of servers, that is, to the $dPlN/M/c$ and $M/dPlN/c$ queueing systems or to finite capacity systems. It would be also interesting to study the optimal control of the systems, that is, when to open or close the queue and which is the optimum number of servers, following the lines of Ausin, Lillo and Wiper (2007).

Also, in this article, we have just considered semi-Markovian queueing systems where either the service or interarrival times were exponential. An extension is to explore more general distributions, in particular the so-called phase-type distributions.

It would be interesting, too, to consider a nonparametric estimate of the Laplace transform from data, so that a parametric specification of the distribution entirely would be avoided. This has been suggested by one of the referees, and will be considered in future work.

Finally, in terms of the application to insurance, it would also be important to explore the estimation of transient or finite time ruin probabilities which are also of interest to insurers.

All Matlab codes and real data utilized in the examples are available in the supplemental material Ramirez et al. (2010).

APPENDIX A: PROOF OF PROPOSITION 1

For ease of notation, we write $z|y, \alpha, \beta, \nu, \tau^2$ as $z|y$ throughout this proof:

$$f_{Z|y}(z|y) = \frac{f_{Y,Z}(y,z)}{f_Y(y)}$$

$$= \frac{f_Z(z)f_W(y-z)}{f_Y(y)}$$

$$= \frac{1}{f_Y(y)}\frac{1}{\tau}\phi\left(\frac{z-\nu}{\tau}\right)\frac{\alpha\beta}{\alpha+\beta}$$

$$\times \left[\exp(\beta(y - z))I_{z \geq y} + \exp(-\alpha(y - z))I_{z < y} \right]$$

$$= \frac{e^{-v^2/(2\tau^2)}}{\tau f_Y(y)} \frac{\alpha\beta}{\alpha + \beta} \left[\exp\left(-\frac{1}{2\tau^2}[z^2 - 2z(v - \tau^2\beta) - 2\tau^2\beta y]\right) I_{z \geq y} \right.$$

$$\left. + \exp\left(-\frac{1}{2\tau^2}[z^2 - 2z(v + \tau^2\alpha) + 2\tau^2\alpha y]\right) I_{z < y} \right]$$

$$= \frac{e^{-v^2/(2\tau^2)}}{\tau f_Y(y)} \frac{\alpha\beta}{\alpha + \beta}$$

$$\times \left[\exp\left(-\frac{1}{2\tau^2}\left[(z - (v - \tau^2\beta))^2 - 2\tau^2\beta y - (v - \tau^2\beta)^2\right]\right) I_{z \geq y} \right.$$

$$\left. + \exp\left(-\frac{1}{2\tau^2}\left[(z - (v + \tau^2\alpha))^2 + 2\tau^2\alpha y - (v + \tau^2\alpha)^2\right]\right) I_{z < y} \right]$$

$$= \frac{e^{-v^2/(2\tau^2)}}{\tau f_Y(y)} \frac{\alpha\beta}{\alpha + \beta}$$

$$\times \left[e^{\beta y + (v - \tau^2\beta)^2/2\tau^2} \exp\left(-\frac{1}{2\tau^2}(z - (v - \tau^2\beta))^2\right) I_{z \geq y} \right.$$

$$\left. + e^{-\alpha y + (v + \tau^2\alpha)^2/2\tau^2} \exp\left(-\frac{1}{2\tau^2}(z - (v + \tau^2\alpha))^2\right) I_{z < y} \right]$$

$$= \frac{e^{-v^2/(2\tau^2)}}{f_Y(y)} \frac{\alpha\beta}{\alpha + \beta} \left[e^{\beta y + (v - \tau^2\beta)^2/2\tau^2} \Phi^c(y^\beta) \frac{\phi(z^\beta)}{\tau \Phi^c(y^\beta)} I_{z \geq y} \right.$$

$$\left. + e^{-\alpha y + (v + \tau^2\alpha)^2/2\tau^2} \Phi(y^\alpha) \frac{\phi(z^\alpha)}{\tau \Phi(y^\alpha)} I_{z < y} \right]$$

$$= \frac{1}{f_Y(y)} \frac{\alpha\beta}{\alpha + \beta} \left[e^{(2\beta y - 2v\beta + \tau^2\beta^2)/2} \Phi^c(y_\beta) \frac{\phi(z^\beta)}{\tau \Phi^c(y^\beta)} I_{z \geq y} \right.$$

$$\left. + e^{(-2\alpha y + 2v\alpha + \tau^2\alpha^2)/2} \Phi^c(y_\alpha) \frac{\phi(z^\alpha)}{\tau \Phi(y^\alpha)} I_{z < y} \right]$$

$$= \frac{1}{f_Y(y)} \frac{\alpha\beta}{\alpha + \beta} \phi\left(\frac{y - v}{\tau}\right)$$

$$\times \left[\frac{\Phi^c(y_\beta)}{\phi(y_\beta)} \frac{\phi(z^\beta)}{\tau \Phi^c(y^\beta)} I_{z \geq y} + \frac{\Phi^c(y_\alpha)}{\phi(y_\alpha)} \frac{\phi(z^\alpha)}{\tau \Phi^c(y^\alpha)} I_{z < y} \right],$$

which gives the conditional density

$$f_{Z|y}(z|y) = \left(R(y_\beta) \frac{\phi(z^\beta)}{\tau \Phi^c(y^\beta)} I_{z \geq y} + R(y_\alpha) \frac{\phi(z^\alpha)}{\tau \Phi^c(y^\alpha)} I_{z < y} \right) \Big/ (R(y_\alpha) + R(y_\beta)).$$

APPENDIX B: PROOF OF PROPOSITION 2

Since

$$W = E_1 - E_2 \qquad \text{where } E_1 \sim \mathcal{E}(\alpha) \text{ and } E_2 \sim \mathcal{E}(\beta),$$

then, the distribution of $E_1 | W$ is

$$
\begin{aligned}
f_{E_1|W}(e_1|w) &= \frac{f_{E_1,W}(e_1, w)}{f_W(w)} \\
&= \frac{f_{E_1,E_2}(e_1, e_1 - w)}{f_W(w)} \\
&= \frac{f_{E_1}(e_1) f_{E_2}(e_1 - w)}{f_W(w)} \\
&= \begin{cases} 0, & \text{if } e_1 \leq \max\{w, 0\}, \\ \dfrac{\alpha e^{-\alpha e_1} \beta e^{-\beta(e_1 - w)}}{(\alpha\beta/(\alpha+\beta))[e^{\beta w} I_{w<0} + e^{-\alpha w} I_{w \geq 0}]}, & \text{for } e_1 > \max\{w, 0\} \end{cases} \\
&= \frac{(\alpha+\beta) e^{-(\alpha+\beta) e_1}}{I_{w<0} + e^{-(\alpha+\beta)w} I_{w \geq 0}} \qquad \text{for } e_1 > \max\{w, 0\}.
\end{aligned}
$$

APPENDIX C: PROOF OF PROPOSITION 3

Note first that

$$
P(w_1 > 0, \ldots, w_n > 0 | \mathbf{y}, \boldsymbol{\theta}) = P(z_1 < y_1, \ldots, z_n < y_n | \mathbf{y}, \boldsymbol{\theta})
$$
$$
= \prod_{i=1}^{n} \Phi\left(\frac{y_i - \nu}{\tau}\right) > 0
$$

for any set $\mathbf{y}$ and where Φ is the standard normal cumulative distribution. Therefore,

$$
P(w_1 > 0, \ldots, w_n > 0 | \mathbf{y}) = \int P(w_1 > 0, \ldots, w_n > 0 | \mathbf{y}, \boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{y}) \, d\boldsymbol{\theta} > 0
$$

for any $\mathbf{y}$. Similarly, $P(w_1 < 0, \ldots, w_n < 0 | \mathbf{y}) > 0$ for any $\mathbf{y}$.

Now consider the posterior distribution of $\alpha, \beta | \mathbf{w}$,

$$
f(\alpha, \beta | \mathbf{w}) \propto f(\mathbf{w}|\alpha, \beta) f(\alpha, \beta)
$$
$$
\propto \left(\frac{\alpha\beta}{\alpha+\beta}\right)^n \exp\left(\beta \sum_{i=1}^{n} w_i I(w_i < 0)\right) \exp\left(-\alpha \sum_{i=1}^{n} w_i I(w_i > 0)\right)
$$
$$
\times f(\alpha, \beta).
$$

In the case that all $w_i < 0$, then when $\alpha \to \infty$, for any given β,

$$
f(\alpha | \beta, \mathbf{w}) \propto f(\alpha, \beta | \mathbf{w}) \to c(\beta) f(\alpha | \beta)
$$

for some $c(\beta) > 0$. Equally, if all $w_i > 0$, then when $\beta \to \infty$, for any given α,

$$f(\beta|\alpha, \mathbf{w}) \propto f(\alpha, \beta|\mathbf{w}) \to d(\alpha) f(\beta|\alpha)$$

for some $d(\alpha) > 0$. Therefore, if $\int_a^\infty f(\alpha|\beta)\,d\alpha$ is divergent for any $a \geq 0$, then we have immediately that when $\alpha \to \infty$, $f(\alpha|\mathbf{w}, \beta) \to c(\beta) f(\alpha|\beta)$, which implies that the posterior distribution of α is improper and similarly in the case of an improper prior for $\beta|\alpha$.

SUPPLEMENTARY MATERIAL

Supplement: Matlab Toolbox (DOI: 10.1214/10-AOAS336SUPP; .zip). The Matlab toolbox performs Bayesian estimation for the double Pareto Lognormal (*dPlN*) distribution, and for the queueing systems *dPlN*/*G*/1 and *M*/*dPlN*/1.

REFERENCES

ADLER, R., FELDMAN, R. and TAQQU, M. T. (1999). *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser, Boston. MR1652283

ARMERO, C. and BAYARRI, M. J. (1994). Bayesian prediction in *M*/*M*/1 queues. *Queueing Syst.* **15** 401–417. MR1266803

AUSÍN, M. C., LILLO, R. E. and WIPER, M. P. (2007). Bayesian control of the number of servers in a *GI*/*M*/*c* queueing system. *J. Statist. Plann. Inference* **137** 3043–3057. MR2364149

BEIRLANT, J., GOEGEBEUR, Y., VERLAAK, R. and VYNCKIER, P. (1998). Burr regression and portfolio segmentation. *Insurance Math. Econom.* **23** 231–250.

BEIRLANT, J. and GOEGEBEUR, Y. (2003). Regression with response distributions of Pareto-type. *Computat. Statist. Data Anal.* **42** 595–619. MR1967059

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications.* Wiley, New York. MR2108013

BOX, G. and TIAO, G. (1973). *Bayesian Inference in Statistical Analysis.* Wiley, New York.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg. MR1458613

FISCHER, M. and KNEPLEY, J. (1977). A numerical solution for some computational problems occurring in queueing theory. In *Algorithmic Methods in Probability, Studies in Management Science* 271–285. North-Holland, Amsterdam.

GROSS, D. and HARRIS, C. M. (1998). *Fundamentals of Queueing Theory*. Wiley, New York. MR1600527

HARRIS, C. M. and MARCHAL, W. G. (1998). Distribution estimation using Laplace transforms. *INFORMS J. Comput.* **10** 448–458. MR1656928

HARRIS, C. M., BRILL, P. H. and FISCHER, M. J. (2000). Internet-type queues with power-tailed interarrival times and computational methods for their analysis. *INFORMS J. Comput.* **12** 261–271.

LELAND, W. E. and WILSON, D. V. (1991). High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection. In *Proc. IEEE INFOCOM'91* 1360–1366. Bat Harbour, FL.

LELAND, W. E., TAQQU, M., WILLINGER, W. and WILSON, D. V. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking* **2** 1–15.

PAXSON, V. and FLOYD, S. (1995). Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking* **3** 236–244.

PHILBRICK, S. W. (1985). A practical guide to the single parameter Pareto distribution. In *Proceedings of the Casualty Actuarial Society* **LXXII** 44–123. Boca Raton, FL.

PRABHU, N. U. (1998). *Stochastic Storage Processes: Queues, Insurance Risk, Dams, and Data Communication*. Springer, Berlin. MR1492990

RAMÍREZ, P., LILLO, R. E. and WIPER, M. P. (2008). Bayesian analysis of a queueing system with a long-tailed arrival process. *Comm. Statist. Simulation Comput.* **4** 697–712.

RAMÍREZ, P., LILLO, R. E., WILSON, S. and WIPER, M. P. (2010). Supplement to "Bayesian inference for double Pareto Lognormal queues." DOI: 10.1214/10-AOAS336SUPP.

REED, W. J. and JORGENSEN, M. (2004). The double Pareto–lognormal distribution—A new parametric model for size distributions. *Comm. Statist. Theory Methods* **33** 1733–1753. MR2065171

RESNICK, S. I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.* **25** 1805–1848. MR1474072

ROBERT, C. P. (2001). *The Bayesian Choice*. Springer, New York. MR1835885

ROLSKI, T., SCHMIDLI, H., SCHMIDT, V. and TEUGELS, J. (1999). *Stochastic Processes for Insurance and Finance*. Wiley, Chichester. MR1680267

SHORTLE, J. F., BRILL, P. H., FISCHER, M. J., GROSS, D. and MASSI, D. M. B. (2004). An algorithm to compute the waiting time distribution for the $M/G/1$ queue. *INFORMS J. Comput.* **16** 52–161. MR2065995

SHORTLE, J. F., FISCHER, M. J. and BRILL, P. H. (2007). Waiting-time distribution of $M/D_N/1$ queues through numerical Laplace inversion. *INFORMS J. Comput.* **19** 112–120. MR2300590

SIGMAN, K. (1999). A primer on heavy-tailed distributions. *Queueing Syst.* **33** 261–275. MR1748646

WILLINGER, W., PAXSON, V. and TAQQU, M. S. (1998). Self-similarity and heavy tails: Structural modeling of network traffic. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. Adler, R. Feldman and M. S. Taqqu, eds.) 27–54. Birkhäuser, Boston. MR1652283

WIPER, M. P. (1997). Bayesian analysis of $Er/M/1$ and $Er/M/c$ queues. *J. Statist. Plann. Inference* **69** 65–79. MR1631145

P. RAMÍREZ-COBO
LABORATOIRE IMAGES ET SIGNAUX
  (UMR CNRS 5083)
ENSEIG, BP 46, 38402
SAINT-MARTIN D'HÈRES
FRANCE
E-MAIL: Pepa.Ramirezcobo@gipsa-lab.inpg.fr

R. E. LILLO
M. P. WIPER
DEPARTAMENTO DE ESTADÍSTICA
UNIVERSIDAD CARLOS III DE MADRID
SPAIN
E-MAIL: lillo@est-econ.uc3m.es
        mwiper@est-econ.uc3m.es

S. WILSON
DEPARTMENT OF STATISTICS
TRINITIY COLLEGE DUBLIN
IRELAND
E-MAIL: Simon.Wilson@tcd.ie