



Working Paper 08-04
Statistics and Econometrics Series 02
February 2008

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

INFERENCE FOR DOUBLE PARETO LOGNORMAL QUEUES WITH APPLICATIONS

Pepa Ramírez, Rosa E. Lillo, Michael P. Wiper¹
Simon P. Wilson²

Abstract

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

Keywords: Heavy tails, Bayesian inference, Queueing theory.

¹ Corresponding autor: Pepa Ramírez Cobo, Rosa E. Lillo and Michael P. Wiper, Universidad Carlos III, Departamento de Estadística, C/ Madrid 126. 28903 Getafe (Madrid), España, e-mail: jrcobo@est-econ.uc3m.es, lillo@est-econ.uc3m.es, mwiper@est-econ.uc3m.es

² Corresponding autor: Simon P. Wilson, Department of Statistics, Trinity College, Dublin 2, Ireland, e-mail: simon.wilson@tcd.ie

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 et al. (1998). In particular, a number of variables in teletraffic engineering such as file sizes, packet arrivals etc. have been shown to possess heavy tailed distributions, see e.g. Paxson & Floyd (1995). Also, in insurance problems, insurance claim sizes can often be very large and in such cases, may be modelled as long tailed, see e.g. Embrechts et al. (1997). For a detailed review of heavy tailed distributions, see 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). However, the Pareto distribution is often inappropriate for modelling the body of the distribution and in these cases, alternative models should be considered. Reed & Jorgensen (2004) recently introduced the double Pareto lognormal (*dPLN*) distribution as a model for heavy tailed data and considered various classical approaches to inference for this distribution. In many problems we are interested not just in inference but also in prediction and in such cases, a Bayesian approach may be preferred, see e.g. Robert (2001). Thus, the first objective of this paper is 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. Here we shall consider the *dPLN*/*M*/1 and *M*/*dPLN*/1 queueing systems where the inter-arrival and service times respectively are modelled using the *dPLN* distribution.

As the *dPLN* distribution does not possess a moment generating function or Laplace transform in closed form as most of heavy-tailed distributions, it is impossible to apply the usual queueing theory techniques to obtain the equilibrium distributions of the *dPLN*/*M*/1 and *M*/*dPLN*/1 queueing systems using standard techniques as in Gross & Harris (1998). An alternative, which we shall apply, is based on a direct approximation of the non-analytical Laplace transform using a variant of the transform approximation method (TAM), see Harris & Marchal (1998), Harris et al. (2000) and Shortle et al. (2004). By combining TAM with Bayesian inference for the *dPLN* distribution, we can obtain numerical predictions of queueing properties 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 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 $dPIN$ distribution. Our results are then applied to a real example of internet traffic arrivals. In Section 4, we study the $M/dPIN/1$ queuing 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 The Double Pareto Lognormal distribution and Bayesian inference.

Reed & Jorgensen (2004) introduced the $dPIN$ distribution as a model for heavy tailed data. A random variable, X is said to have a $dPIN$ distribution with parameters α, β, ν, τ if $X = \exp(Y)$ where $Y = Z + W$ is the sum of a normally distributed random variable, $Z \sim N(\nu, \tau^2)$, and an independent, skewed Laplace distributed variable, $W \sim SL(\alpha, \beta)$, with density function

$$f_W(w|\alpha, \beta) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} e^{\beta w} & \text{if } w \leq 0, \\ \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha w} & \text{if } w > 0, \end{cases}$$

where $\alpha, \beta > 0$. In this case, Y is said to have a *Normal Laplace distribution* (NL), denoted $Y \sim NL(\alpha, \beta, \nu, \tau)$ with density

$$f_Y(y|\alpha, \beta, \nu, \tau) = \frac{\alpha\beta}{\alpha+\beta} \phi\left(\frac{y-\nu}{\tau}\right) [R(p) + R(q)] \quad (1)$$

where $p = \alpha\tau - (y - \nu)/\tau$, $q = \beta\tau + (y - \nu)/\tau$ and $R(z) = \Phi^c(z)/\phi(z)$ is *Mill's ratio*, where $\Phi^c(z) = 1 - \Phi(z)$ and $\phi(z)$ and $\Phi(z)$ are the standard normal density and cumulative distributions respectively. The density function of X can then be derived via the usual change of variables formula.

Reed & Jorgensen (2004) demonstrate the properties of this distribution, in particular showing that it exhibits (both lower and) upper power-tail behavior in that $f_X(x) \rightarrow kx^{-\alpha-1}$ as $x \rightarrow \infty$ and $f_X(x) \rightarrow kx^{\beta-1}$ as $x \rightarrow -\infty$.

Reed & Jorgensen (2004) also illustrate a procedure for classical inference for the $dPIN$ 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.

Given a random sample $\mathbf{x} = (x_1, \dots, x_n)$ from the $dPIN(\alpha, \beta, \nu, \tau^2)$, the goal is to compute a posterior distribution $f(\alpha, \beta, \nu, \tau^2 | x_1, \dots, x_n)$. For ease of no-

tation, 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 $\mathbf{y} = (y_1, \dots, y_n)$, where $y_i = \log(x_i)$ for $i = 1, \dots, n$, and compute $f(\boldsymbol{\theta} | y_1, \dots, y_n)$ using the Normal Laplace likelihood.

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

It can be easily shown that the conditional distribution of $Z|Y = y$ is a weighted mixture of two truncated normal densities:

$$f_{Z|y}(z|y) = \frac{R(q) \frac{\frac{1}{\tau} \phi\left(\frac{z - (\nu - \tau^2 \beta)}{\tau}\right)}{\Phi^c\left(\frac{y - (\nu - \tau^2 \beta)}{\tau}\right)} I_{z \geq y} + R(p) \frac{\frac{1}{\tau} \phi\left(\frac{z - (\nu + \tau^2 \alpha)}{\tau}\right)}{\Phi^c\left(\frac{y - (\nu + \tau^2 \alpha)}{\tau}\right)} I_{z < y}}{R(p) + R(q)} \quad (2)$$

where $p = \alpha\tau - (y - \nu)/\tau$, $q = \beta\tau + (y - \nu)/\tau$.

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

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

Then, the distribution of $E_1|W = w$ is

$$f_{E_1|W}(e_1|w) = \begin{cases} 0 & \text{if } e_1 \leq \max\{w, 0\} \\ \frac{\alpha e^{-\alpha e_1} \beta^{-\beta(e_1 - w)}}{\frac{\alpha\beta}{\alpha+\beta} [e^{\beta w} I_{w < 0} + e^{-\alpha w} I_{w \geq 0}]} & \text{for } e_1 > \max\{w, 0\} \\ \frac{(\alpha + \beta) e^{-(\alpha + \beta)e_1}}{I_{w < 0} + e^{-(\alpha + \beta)w} I_{w \geq 0}} & \text{for } e_1 > \max\{w, 0\}. \end{cases} \quad (3)$$

This is just an exponential distribution truncated onto $[\max\{w, 0\}, \infty)$. Given a sample, $\mathbf{y} = y_1, \dots, y_n$, then conditional on the parameters α, β, ν, τ we can generate $\mathbf{z} = z_1, \dots, z_n$, from the formula in Equation 2. Also, we can define $\mathbf{w} = \mathbf{y} - \mathbf{z}$, and generate $\mathbf{e}_1 = e_{11}, \dots, e_{1n}$, from the formula in Equation 3 and define $\mathbf{e}_2 = \mathbf{e}_1 - \mathbf{w}$.

In order to undertake inference for ν, τ^2 , let's suppose we use a standard normal inverse gamma prior,

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

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

Then, by standard normal gamma theory,

$$\nu|\tau, \mathbf{z} \sim \mathcal{N}\left(\frac{am + n\bar{z}}{a + n}, \frac{\tau^2}{a + n}\right) \quad \text{where } \bar{z} = \sum_{i=1}^n z_i/n,$$

$$\frac{1}{\tau^2} | \mathbf{z} \sim \mathcal{G}\left(\frac{a + n}{2}, \frac{b + (n-1)s_z^2 + \frac{kn}{k+n}(m - \bar{z})^2}{2}\right) \quad \text{where } s_z^2 = \sum_{i=1}^n (z_i - \bar{z})^2/(n-1).$$

Also, given gamma priors $\alpha \sim \mathcal{G}(c_\alpha, d_\alpha)$, $\beta \sim \mathcal{G}(c_\beta, d_\beta)$, then it is easy to show that

$$\alpha|\mathbf{e}_1 \sim \mathcal{G}(c_\alpha + n, d_\alpha + n\bar{e}_1) \quad \text{where } \bar{e}_1 = \sum_{i=1}^n e_{1i}/n,$$

$$\beta|\mathbf{e}_2 \sim \mathcal{G}(c_\beta + n, d_\beta + n\bar{e}_2) \quad \text{where } \bar{e}_2 = \sum_{i=1}^n e_{2i}/n.$$

Thus, we can define the following Gibbs sampling algorithm,

1. Set initial values $\alpha^{(0)}, \beta^{(0)}, \nu^{(0)}, \tau^{(0)}$.
2. For $t = 1, \dots, T$
 - a. For $i = 1, \dots, n$,
 - a1. Generate $z_i^{(t)}$ from $f(z|\alpha^{(t-1)}, \beta^{(t-1)}, \nu^{(t-1)}, \tau^{(t-1)} | y_i)$.
 - 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 $\tau^{(t)} \sim f(\tau|\mathbf{z}^{(t)})$
 - c. Generate $\nu^{(t)} \sim f(\nu|\mathbf{z}^{(t)}, \tau^{(t)})$
 - d. Generate $\alpha^{(t)} \sim f(\alpha|\mathbf{e}_1^{(t)})$.
 - e. Generate $\beta^{(t)} \sim f(\beta|\mathbf{e}_2^{(t)})$.

If the usual non-informative prior distribution, $f(\alpha, \beta) \propto 1/(\alpha\beta)$ is used, it is interesting to note that it is necessary to define proper prior distributions. Then, given \mathbf{w} ,

$$\begin{aligned}
f(\alpha|\mathbf{w}, \beta) &\propto \frac{1}{\alpha} \left(\frac{\alpha\beta}{\alpha + \beta} \right)^n \left[e^{\beta \sum_{i=1}^n w_i I(w_i < 0)} + e^{-\alpha \sum_{i=1}^n w_i I(w_i > 0)} \right] \\
&\propto \frac{\alpha^{n-1}}{(\alpha + \beta)^n} \left[e^{\beta \sum_{i=1}^n w_i I(w_i < 0)} + e^{-\alpha \sum_{i=1}^n w_i I(w_i > 0)} \right]
\end{aligned} \tag{4}$$

and as long as some $w_i < 0$, then the integral of this density is divergent because

$$\int_0^\infty \frac{\alpha^{n-1}}{(\alpha + \beta)^n} d\alpha = \infty.$$

2.1 Illustration with simulated data

A sample of size 1000 was simulated from $dPIN(.25, .5, 1, 1)$ and the Gibbs algorithm was run for 1,000,000 iterations with initial values set to the maximum likelihood estimates, $\theta^{(0)} = (0.269, 0.456, 1.271, 0.980)$. The values generated from the simple Gibbs algorithm show quite high autocorrelation and therefore, we thinned the data and just selected every 100th sampled value. Figure

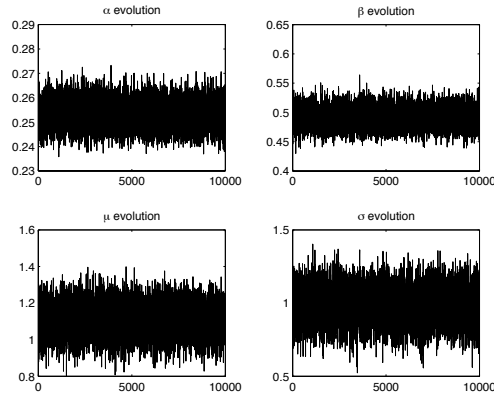


Fig. 1. Trace plots from the MCMC.

1 illustrates the mixing properties of the algorithm and Figure 2 depicts the evolution of the parameter means.

Figure 3 shows the predictive (dotted line) and theoretical (solid line) density function, estimated for the data. It is easy to see that the fitted results are close to the frequency histogram and theoretical density function. We also found $E(\theta|y) = (0.2585, 0.477, 1.1652, 0.973)$ close to the maximum likelihood estimates.

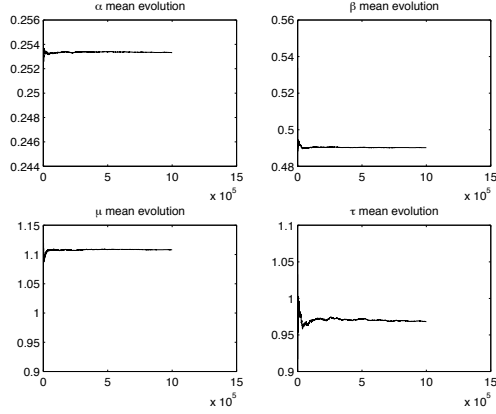


Fig. 2. Graph of the evolution of the parameters' sample averages.

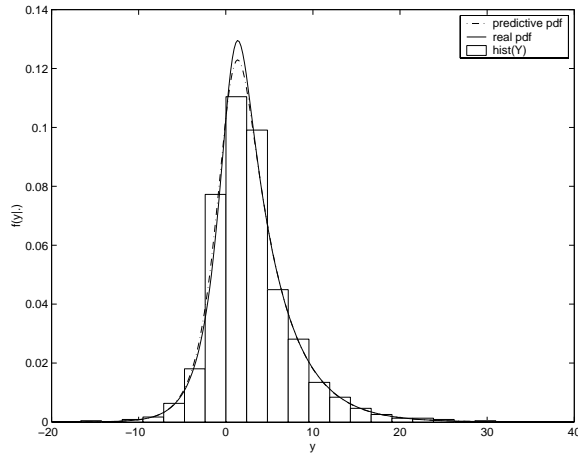


Fig. 3. Histogram, predictive (dotted line) and theoretical (solid line) pdf.

3 The $dPIN/M/1$ queueing system and Bayesian inference

In this section, we shall consider the $dPIN$ distribution as a model for the arrival process in a single-server queueing system with independent, exponentially distributed service times. This queueing system, denoted as $dPIN/M/1$ is an example of $G/M/1$ queueing system, whose properties are well known (see Gross & Harris (1998)).

For the $dPIN/M/1$ system with parameters $\theta = (\alpha, \beta, \nu, \tau)$, standard results for $G/M/1$ queues imply that the mean inter-arrival time does not exist if $\alpha < 1$. In this case, the queueing system is automatically stable whatever the service rate μ . Otherwise, the traffic intensity is given by

$$\rho = \frac{(\alpha - 1)(\beta + 1)}{\mu\alpha\beta e^{\mu + \tau^2/2}}. \quad (5)$$

If the system is stable ($\rho < 1$) then the steady-state probability for the number of customers Q in system just before an arrival is

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

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

$$r_0 = f^*(\mu(1 - r_0)), \quad (6)$$

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

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

However, the Laplace transform of the $dPIN$ distribution is analytically intractable so that the standard techniques for finding the root of Equation 6 cannot be applied. Thus, an alternative approach to obtaining the steady state distributions is needed. Next, we outline such approach.

The transform approximation method (TAM) was developed, informally by Harris & Marchal (1998) and Harris et al. (2000) to approximate 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 $dPIN$ distribution. To approximate the Laplace transform $f^*(s)$ of the distribution of a random variable X , the basic algorithm is:

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

$$\begin{aligned} w_1 &= \frac{p_1 + p_2}{2}, \\ w_i &= \frac{p_{i+1} - p_{i-1}}{2}, \text{ for } i = 2, \dots, N-1 \\ w_N &= 1 - \frac{p_{N-1} + p_N}{2} \end{aligned}$$

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

For the $dPIN$ case, the quantiles in step 2 are approximated numerically by Newton-Raphson, with initial values obtained from the empirical distribution function of the data.

Harris et al. (2000) and Shortle et al. (2004) consider different alternatives for the defining probabilities p_i . The easiest 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 and an alternative algorithm which better captures heavy tailed behavior, is the geometric or G-TAM algorithm which sets $p_i = 1 - q^i$, for $q \in (0, 1)$. In practice, we have found that a combination of both algorithms works well. Thus, the U-TAM algorithm is used to obtain percentiles from the body of the distribution and the G-TAM algorithm is used for the heavy tail. Formally, the proportion of percentiles before and after $E(X)$, and q , are chosen so that the TAM mean matches the mean of the original distribution (or median if the mean does not exist). This satisfies the conditions of Theorem 1 in Shurtle et al (2004) so that convergence of $f_N^*(s)$ to $f^*(s)$ is assured as $N \rightarrow \infty$.

Given an uninformative prior distribution and a sample of $dPlN$ distributed inter-arrival 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 $t = 1, \dots, l$ 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,

$$P(\rho < 1 | \text{data}) = \frac{1}{l} \sum_{t=1}^l I(\rho^{(t)} < 1) \quad (7)$$

where $\rho^{(t)}$ is the value of ρ calculated from (5) 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 (6) and therefore, the posterior predictive distributions of queue size etc. can be estimated by Rao Blackwellization.

One point to note however is that, as commented in Wiper (1997), it can be shown that the predictive means of the 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.

3.1 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 e.g. Willinger et al. (1998). In particular, as shown in Paxson & Floyd (1995), internet arrival traffic cannot be well modelled by a Poisson process. As an alternative, heavy tailed distributions can be considered.

Figure 4 shows a histogram of 50000 inter-arrival times, in seconds, of a trace

of 1 million ethernet packets taken from

<http://www.xtremes.de/xtremes/xtremes/download/download.htm>.

Superimposed is the predictive $dPIN$ density generated using the Bayesian algorithm described in Section 2. It can be seen that this model fits the data quite well. The posterior mean parameter estimates for this model were $E(\boldsymbol{\theta}|x) = (2.15, 1.07, -6.00, 0.36)$.

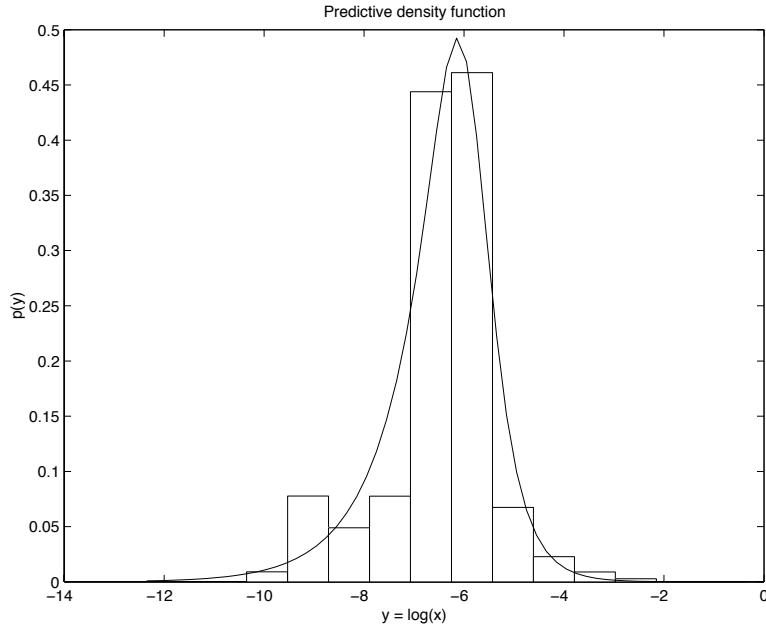


Fig. 4. Histograms and predictive pdf for the internet data.

Now we shall consider the queueing aspects. Given the $dPIN$ arrival process, we shall assume that arrivals are processed by a single server with exponentially distributed service times with rate μ . Table 1 shows the posterior probability of equilibrium and the expected value for the traffic intensity for an assortment of values of μ . From this table, it is clear that there is a high probability that the system is stable for values of μ greater than 394. Figure 5 and Table 2 illustrate the predictive system and queueing time distributions and the distribution of the number of clients in the system in equilibrium for values of μ greater than 400. We can see that as the service rate increases, then the median queueing and system waiting times and the number of clients in the system decrease, as would be expected.

We compared these results with those obtained from the queueing systems $Pareto/M/1$ and $M/M/1$. Figure 6 depicts the predictive systems and queue waiting times distributions when the inter-arrival time follows a $dPIN$ distribution (solid line), a Pareto distribution (dashed line) and an exponential distribution (dashdotted line) and when the service time is exponentially distributed with mean $1/500$. Table 3 compares the predictive system size distri-

μ	$E(S)$	$\mathbb{P}(\rho < 1 data)$	$\mathbb{E}(\rho)$
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

Table 1
Probability of equilibrium and traffic intensity.

bution just before an arrival among these different queues.

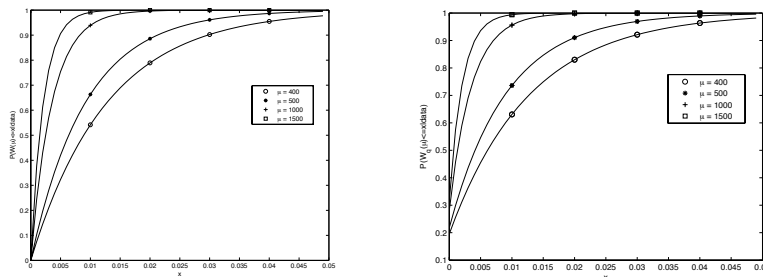


Fig. 5. Predictive system and queue waiting times distributions for the internet data set.

4 The $M/dPIN/1$ queueing system and ruin probabilities

In an insurance context, it is often assumed that claim sizes, C_i , are independent and identically distributed heavy-tailed random variables, see e.g. Rolski et al. (1999). Here, we shall assume that claim sizes can be modelled as $dPIN$ random variables. Often, it is also supposed that the inter-claim 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

μ	$\mathbb{P}(Q = 0)$	$\mathbb{P}(Q = 1)$	$\mathbb{P}(Q = 2)$	$\mathbb{P}(Q = 3)$
1500	.3167	.2161	.1475	.1008
1000	.2813	.2019	.1449	.1042
500	.2182	.1703	.1330	.1039
400	.1955	.1570	.1260	.1014
394	.1946	.1565	.1259	.1013
390	.1951	.1567	.1260	.1013

Table 2

Predictive system size distribution just before an arrival for the internet data set.

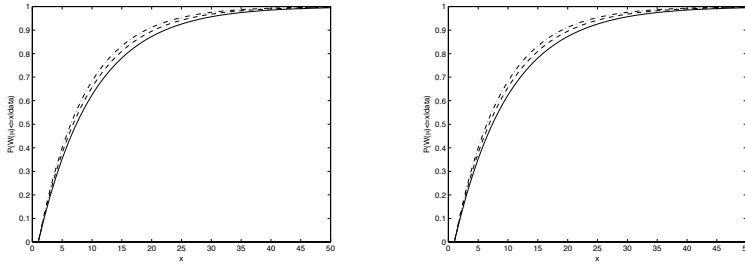


Fig. 6. Predictive system and queue waiting times distributions for the internet data set when $\mu = 500$.

<i>Arrival process</i>	$\mathbb{P}(Q = 0)$	$\mathbb{P}(Q = 1)$	$\mathbb{P}(Q = 2)$	$\mathbb{P}(Q = 3)$
<i>dPLN</i>	.2182	.1703	.1330	.1039
<i>Pareto</i>	.2383	.1813	.1051	.0800
<i>Exponential</i>	.2553	.1901	.1416	.1054

Table 3

Predictive system size distribution just before an arrival for the internet data set for different arrival processes when $\mu = 500$.

accumulates. Then, the company's wealth, or risk portfolio at time t is,

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

where $N(t) = \sup(n : \sum_{i=1}^n T_i \leq t)$ 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

$$\psi(a, r) = P(W(t) < 0 \text{ for some } t \geq 0 | \text{initial capital } u, \text{ premium rate } r). \quad (8)$$

Clearly, if the mean claim size does not exist, then eventual ruin is certain. Otherwise, we can define the traffic intensity of this system as $\rho = \lambda E[C_i]/r$ (this ρ will play the role of $\rho^{(t)}$ at each iteration in (7)) and it is well known that ruin is certain if $\rho \geq 1$. In the case that $\rho < 1$, then in e.g. 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 inter-arrival time and service time distributions are the same as the distributions of T_i and C_i/r respectively. 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 8.

From general properties of the $M/G/1$ queueing system, see e.g. Gross & Harris (1998), the Laplace transform $W_q^*(s)$ of the equilibrium waiting time in the queue is related to the Laplace transform $B^*(s)$ of the service time by:

$$W_q^*(s) = \int_0^\infty e^{-st} dW_q(t) = \frac{(1-\rho)s}{s - \lambda(1 - B^*(s))}, \quad (9)$$

where $W_q(t)$ is the distribution function of the waiting time.

In this section we consider the single-server queueing system, with independent, exponentially distributed inter-arrival times and $dPIN$ service times. For the $M/dPIN/1$ queue with arrival rate λ and $dPIN$ service times with $\theta = (\alpha, \beta, \nu, \tau)$ then, if $\alpha < 1$, the expected service time does not exist and the queueing system is never stable, whatever the inter-arrival rate λ . When $\alpha \geq 1$ the traffic intensity is given by

$$\rho = \frac{\lambda \alpha \beta e^{\nu + \tau^2/2}}{(\alpha - 1)(\beta + 1)}. \quad (10)$$

In order to obtain the distribution function of the waiting time $W_q(t)$, we firstly apply the TAM to approximate $B^*(s)$ as earlier. Secondly, we can use a standard numerical approach to invert the Laplace transform, $W_q^*(s)$, see e.g. Shortle et al. (2007) for a review. In this case, we apply the recursion method of Fischer & Knepley (1977) which is recommended in Shortle et al. (2007).

Assuming this 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 $dPIN$ 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.

4.1 Application to fire insurance claims

Here we consider data from e.g. Beirlant et al. (2004) representing 9181 Norwegian fire claims and available from

<http://ucs.kuleuven.be/Wiley/index.html>.

Figure 7 shows the data and the Bayesian *dPIN* fit. The posterior expected parameter values are $E(\boldsymbol{\theta}|x) = (1.24, 17.27, 6.39, .18)$.

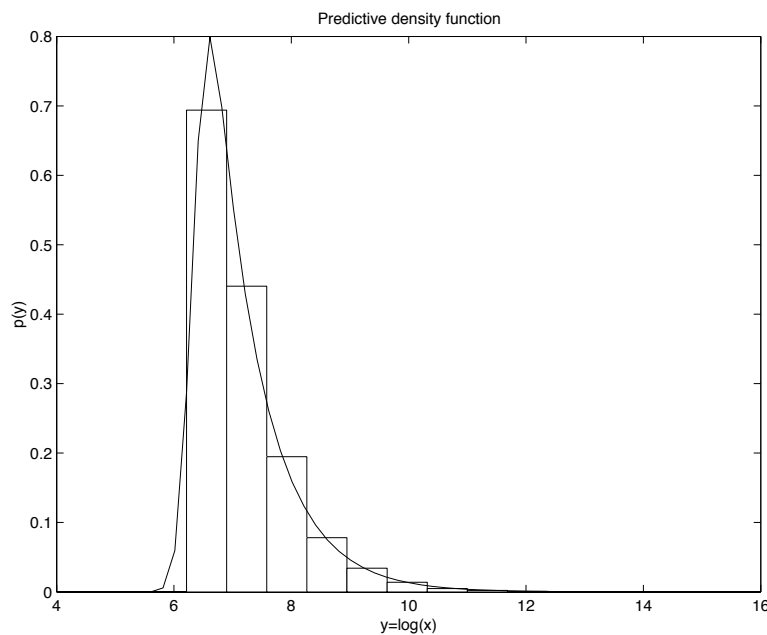


Fig. 7. Histograms and predictive pdf for the Norwegian data.

Assuming that the system is stable, we can now estimate the predictive ruin probability for different inter-claim rates and initial reserves. In this case, the expected claim size, conditional on this existing (i.e. that $\alpha \geq 1$) is approximately 2915, which implies that in order to avoid extremely high probabilities of ruin, we should typically consider values of λ below $1/2915$. Figure 8 depicts the predictive probability of ruin, $E[\psi(u)|\text{data}]$ for a grid of values of different inter-claim rates, λ and various initial reserve levels, u . We found that, given these claim sizes, both queuing systems $M/M/1$ and $M/Pareto/1$ are not stable whatever the arrival rates are. Apart from badly fitting the data, for both queuing systems the posterior expected service time did not exist and thus, these models would predict a certain ruin for the given data.

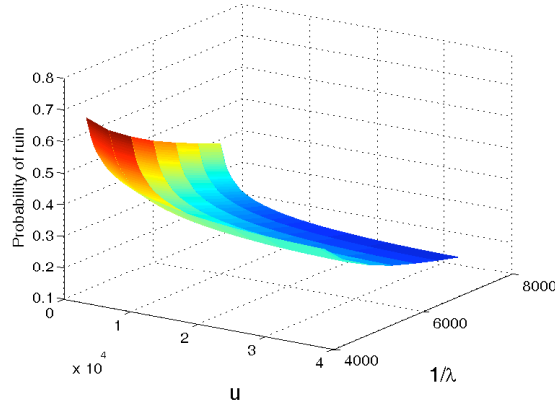


Fig. 8. Probabilities of ruin for the Norwegian data insurance company.

5 Conclusions and Extensions

In this work, we have developed Bayesian inference for the double Pareto log-normal 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. As θ is only 4-dimensional, a possibility of Bayesian inference is importance sampling, but we found it difficult to find good distributions for the initial sample. Another approach was a Metropolis algorithm, updating the four parameters as a block with independent proposal densities $q(\theta|\theta^{(t)}) = q(\alpha|\alpha^{(t)})q(\beta|\beta^{(t)})q(\nu|\nu^{(t)})q(\tau|\tau^{(t)})$, where for example, each q follows uniform distributions for α, β, τ and a normal distribution for ν . We also considered a multivariate proposal density, $q(\log \theta | \log \theta^{(t)}) \sim MN(\log \theta^{(t)}, \Sigma)$, a multivariate normal distribution with covariance matrix Σ , previously estimated using maximum likelihood or a Gibbs algorithm. Our experience with the Metropolis algorithm has been that this exhibits very slow and poor mixing for τ . This suggests that the Gibbs procedure should be preferred.

Secondly, we have combined this approach with techniques from the queueing literature in order to estimate predictive equilibrium distributions for the $dPIN/M/1$ and $M/dPIN/1$. To do this, we have adapted the Transform Approximation method, in order to estimate the Laplace transform of the $dPIN$ distribution and the waiting time distribution in the $M/dPIN/1$ 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 presented. Large differences among these queueing systems when the service process is heavy-tailed were found.

A number of extensions are possible. Firstly, we could extend our results to

the case of multiple number of servers, i.e. to the $dPIN/M/c$ and $M/dPIN/c$ 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 server, following the lines of Ausín et al. (2007).

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

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.

References

- Adler, R., Feldman, R., & Taqqu, M.T., A Practical Guide to Heavy Tails: Statistical Techniques and Applications, 1998. Birkhäuser: Boston.
- Ausín, M.C., Lillo, R.E & Wiper, M.P., Bayesian control of the number of servers in a GI/M/c queueing system, *Journal of Statistical Planning and Inference*, **137**, 3043–3057.
- Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J., *Statistics of Extremes: Theory and Applications*, 2004, Wiley: New York.
- Embrechts, P., Klüppelberg, C., & Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, 1997, Springer-Verlag: Heidelberg.
- Fischer, M., & Knepley, J., A numerical solution for some computational problems occurring in queueing theory. *Algorithmic Methods in Probability, Studies in Management Science*, 1977, Elsevier North-Holland: Amsterdam, pp. 271–285.
- Gross, D.& Harris, C.M., *Fundamentals of Queueing Theory*, 1998, Wiley: New York.
- Harris, C.M.& Marchal, W.G., Distribution estimation using Laplace transforms, *INFORMS Journal of Computing*, 1998, **10**, 448-458.
- Harris, C.M., Brill, P.H., & Fischer, M.J., Internet-type queues with power-tailed interarrival times and computational methods for their analysis, *INFORMS Journal of Computing*, 2000, **12**, 261-271.
- Paxson, V.& Floyd, S., Wide Area Traffic: The Failure of Poisson Modeling, *IEEE/ACM Transactions on Networking*, 1995, **3**, 236-244.
- Philbrick, S.W., A Practical Guide to the Single Parameter Pareto Distribution, CAS Proceedings 1985.
- Prabhu, N.U., *Stochastic Storage Processes: Queues, Insurance Risk, Dams, and Data Communication*, 1998, Springer: Berlin.
- Reed, W.J.& Jorgensen, M., The Double Pareto-Lognormal Distribution - A

- New Parametric Model for Size Distributions, *Communications in Statistics - Theory and Methods*, 2004, **33**, 1733-1753.
- Resnick, S.I., Heavy tail modeling and teletraffic data, *Annals of Statistics*, 1997, **25**, 1805-1848.
- Robert, C.P., *The Bayesian Choice*, 2001, Springer Verlag: New York.
- Rolski, T., Schmidli, H., Schmidt, V., & Teugels, J., *Stochastic Processes for Insurance and Finance*, 1999, Wiley: Chichester.
- Shortle, J.F., Brill, P.H., Fischer, M.J., Gross, D., & Massi, D.M.B., An Algorithm to Compute the Waiting Time Distribution for the M/G/1 Queue, *INFORMS Journal on Computing*, 2004, **16**, 152-161.
- Shortle, J.F., Fischer, M.J., & Brill, P.H., Waiting-Time Distribution of $M/D_N/1$ Queues Through Numerical Laplace Inversion, *INFORMS Journal on Computing*, 2007, **19**, 112-120.
- Sigman, K., A primer on heavy-tailed distributions. *Queueing Systems*, 1999, **33**, 261-275.
- Willinger, W., Paxson, V., & Taqqu, M.S., Self-similarity and Heavy Tails: Structural Modeling of Network Traffic, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, 1998, Adler, R., Feldman, R. & Taqqu, M.S., editors, Birkhauser: Boston, pp 27-54.
- Wiper, M.P., Bayesian analysis of Er/M/1 and Er/M/c Queues, *The Journal of Statistical Planning and Inference*, 1997, **69**, 65-79.