Statistics and Probability Letters 113 (2016) 103–107

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bayesian inference for extreme quantiles of heavy tailed distributions

Rafael B.A. Farias [a,*], Michel H. Montoril [b], José A.A. Andrade [a]

[a] Department of Statistics and Applied Mathematics, Federal University of Ceara, Brazil

[b] Department of Statistics, Federal University of Juiz de Fora, Brazil

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form 25 February 2016
Accepted 26 February 2016
Available online 12 March 2016

Keywords:
High quantile
HPD interval
Heavy tail

ABSTRACT

We propose a new method for estimating extremes quantiles of a wide class of heavy-tailed distributions. Our proposal makes Bayesian inference on extreme quantiles through High Posterior Density intervals. We evaluate the performance of the proposal by numerical results.

1. Introduction

The interest in predicting extreme events has grown in several areas of knowledge. For instance, according to Beirlant et al. (2004), industrial fire events can cause several side effects in loss of property, temporary unemployment and lost contracts. Some events occasionally include large claims, which put at risk the solvency of a portfolio or even a substantial part of the company. For this reason, it becomes fundamental for companies to gain some knowledge about high quantiles of claims. However, the sparsity of these data in the extremes may not provide estimates close to the real quantile of interest, since the estimators usually have high variability. Thus, interval estimation becomes a good option for inference in this case.

Some frequentist confidence intervals have been developed to predict high quantiles of data with heavy-tailed distributions. Peng and Qi (2006) proposed three methods, namely: the normal approximation, the likelihood ratio and the data tilting methods. These methods lead to the same asymptotic distribution, but the data tilting method provides better numerical results with respect to the amplitude and coverage probabilities of the confidence intervals for 99% and 99.9% quantiles.

The computational development in the last years has drawn attention of researchers to Bayesian methods, which have shown powerful results when used together with the Monte Carlo Markov Chains (MCMC) methods, mainly, when the posterior distribution is difficult to be calculated or when has an intractable expression. In extreme value theory, there is already a considerable amount of works using Bayesian methods (e.g., Coles and Tawn, 1996; Coles and Powell, 1996; Beirlant et al., 2004; Behrens et al., 2004; Cabras et al., 2011).

In this work we propose a new Bayesian alternative to estimate high quantiles of data modeled according to regularly varying distributions, which are heavy-tailed distributions, that is, we consider $1 - F(x) = \ell_F(x)x^{-\gamma}$, where F corresponds to the cumulative distribution function of the data and ℓ_F is a slowly varying function, i.e., $\lim_{x\to\infty} \ell_F(yx)/\ell_F(x) = 1, y > 0$. Our methodology, which can be based on non-informative prior distributions, can easily be implemented and has a low computational cost.

* Corresponding author.
E-mail address: rafael@dema.ufc.br (R.B.A. Farias).

http://dx.doi.org/10.1016/j.spl.2016.02.020

This paper is organized as follows. In Section 2 the Bayesian approach is presented. Some numerical results, based on Monte Carlo simulations, are reported in Section 3. In Section 4 we illustrate the methodology with an application to the Danish Fire Losses data set. Conclusions and additional remarks are presented in Section 5.

2. Bayesian estimator

Let $X_1, \ldots, X_n$ be an independent random sample from the distribution F, where $\bar{F}(x) \equiv 1 - F(x) = \ell_F(x)x^{-\gamma}$, $x > 0$, with $\gamma > 0$ called tail index. The function ℓ_F satisfies $\lim_{x\to\infty} \ell_F(yx)/\ell_F(x) = 1$, $y > 0$. Since ℓ_F is a slowly varying function, for some large $T > 0$, there exists a positive constant c such that $\bar{F}$ can be approximated by $\bar{F}(x) \approx cx^{-\gamma}$, $x > T$. In this case the $100(1-p)\%$ quantile, denoted here by x_p, could be approximated by $(p/c)^{-1/\gamma}$ when p is sufficiently small. Therefore, a good estimator for x_p would be $\hat{x}_p = (\hat{c}/p)^{1/\hat{\gamma}}$.

It is easy to show that the likelihood function for the censored data is $L(\boldsymbol{X}|c,\gamma) = \prod_{i=1}^{n}(c\gamma X_i^{-\gamma-1})^{\delta_i}(1-cT^{-\gamma})^{1-\delta_i}$, where $\boldsymbol{X} = (X_1, \ldots, X_n)$, and $\delta_i = \mathbb{1}(X_i > T)$ is an indicator function, $i = 1, \ldots, n$. Thus, denoting the prior distribution for the vector (c, γ) by $\pi(c,\gamma)$, the quantile x_p can be estimated by using the mean of the posterior distribution $\pi(c,\gamma|\boldsymbol{X})$, which is proportional to $\pi(c,\gamma)L(\boldsymbol{X}|c,\gamma)$.

Let us now consider the case where we do not have any prior information about the tail index γ. Since it is a positive parameter, we can assume that γ has a Gamma distribution with fixed shape and rate hyperparameters $s > 0$ and $r > 0$, respectively. Thus, $\pi(\gamma) = \Gamma(s)^{-1}r^s\gamma^{s-1}e^{-r\gamma}$, $\gamma > 0$. Furthermore, since $0 < \bar{F}(T) < 1$, it is reasonable to think that, given γ, $cT^{-\gamma}$ has a Beta distribution with hyperparameters $a > 0$ and $b > 0$. In this case, the prior distribution of $c|\gamma$ will be $\pi(c|\gamma) = B(a,b)^{-1}T^{-\gamma a}c^{-\gamma(a-1)}(1-cT^{-\gamma})^{b-1}$, $0 < c < T^{\gamma}$, where $B(\cdot,\cdot)$ corresponds to the Beta function.

Let $P_{X,\delta} = \prod_{i=1}^{n} X_i^{\delta_i}$ and $\lambda = \sum_{i=1}^{n}\delta_i$. The prior distributions above lead to the posterior distribution $\pi(c,\gamma|\boldsymbol{X}) \propto \gamma^{\lambda+s-1}c^{\lambda+a-1}(1-cT^{-\gamma})^{n-\lambda+b-1}\left(e^{r}TP_{X,\delta}\right)^{-\gamma}$, which is intractable. However, we can generate samples of $[x_p|\boldsymbol{X}]$ based on samples of $[(c,\gamma)|\boldsymbol{X}]$. This sample can be generated by using the Gibbs sampler through the following full conditional distributions $\pi(\gamma|c,\boldsymbol{X}) \propto \gamma^{\lambda+s-1}(1-cT^{-\gamma})^{n-\lambda+b-1}\left(e^{r}TP_{X,\delta}\right)^{-\gamma}$ and $\pi(c|\gamma,\boldsymbol{X}) \propto c^{\lambda+a-1}(1-cT^{-\gamma})^{n-\lambda+b-1}$, $0 < c < T^{\gamma}$.

The sampling is done through MCMC methods. Since $[c|\gamma,\boldsymbol{X}]$ belongs to the same class of distributions as $[c|\gamma]$, the Beta distribution can be used. A Metropolis–Hastings algorithm is considered within the Gibbs sampler, in order to draw data from the same distribution as $[\gamma|c,\boldsymbol{X}]$. The detailed MCMC algorithm to replicate the random vector $[(c,\gamma)|\boldsymbol{X}]$ is proposed below.

Algorithm for sampling from $[(c,\gamma)|X]$

1. Initialize the vector $(\gamma^{(0)}, c^{(0)})$ and the counter $j = 1$;
2. Generate a value $c^{(j)}$ from $[c|\gamma^{(j-1)},\boldsymbol{X}]$ according to the following steps:
   2.1. Draw c^* from a Beta$(\lambda + a, n - \lambda + b + 1)$;
   2.2. Calculate $c = c^*T^{\gamma^{(j-1)}}$.
3. Generate $\gamma^{(j)}$ from the distribution of $[\gamma|c^{(j)},\boldsymbol{X}]$ according to the following steps:
   3.1. Draw γ^* from a Gamma$(s^{(j-1)}, r^{(j-1)})$, where $s^{(j-1)} = (\gamma^{(j-1)})^2/\nu$, $r^{(j-1)} = \gamma^{(j-1)}/\nu$, with ν denoting a tuning parameter;
   3.2. Generate $U \sim$ Uniform$(0, 1)$ and compute $\rho = \min\left\{1, \frac{\pi(\gamma^*|c^{(j+1)},\boldsymbol{X})}{\pi(\gamma^{(j)}|c^{(j+1)},\boldsymbol{X})}\right\}$;
   3.3. Put $\gamma^{(j+1)} = \begin{cases}\gamma^*, & \text{if } U < \rho\\ \gamma^{(j)}, & \text{otherwise}\end{cases}$;
4. Move the counter from j to $j + 1$ and repeat Steps 2–4.

The Metropolis–Hastings algorithm used to generate the posterior sample from $\gamma|\boldsymbol{X}$ is equivalent to the random-walk Metropolis–Hastings algorithm with log-gamma distribution. The chain reaches the convergence after a *burn in* period of B iterations, which must be discarded. An adequate value for B can be checked by graphical tools (see, e.g., Plummer et al., 2006). Then, $\{(c^{(j)},\gamma^{(j)})\}_{j=B+1}^{M}$ can be considered a sample from the posterior distribution $[(c,\gamma)|\boldsymbol{X}]$. Furthermore, in order to reduce the autocorrelation of the sample, we can take a lagged subsample of size N, $\{(c^{(B+jl)},\gamma^{(B+jl)})\}_{j=1}^{N}$, for some lag l. For the sake of simplicity, let us denote the remaining data by $\{(c^{(l)},\gamma^{(l)})\}_{l=1}^{N}$. Thus, a sample of the posterior distribution of the high quantile $x_p|\boldsymbol{X}$ can be obtained by $\{x_{pn}^{(l)} = (c^{(l)}/p)^{1/\gamma^{(l)}}\}_{l=1}^{M}$, and Bayesian inferences such as HPD intervals are based on this generated sample.

Observe that ρ in Step 3.2 of the algorithm above determines the change of stage for $\gamma^{(j+1)}$, which happens when the generated value of γ^* is accepted in Step 3.3. The acceptance rate depends on the tuning parameter ν, which needs to be suitably chosen. A good choice for ν would be a value for which acceptance rate of γ^* is around 0.5, as recommended in the literature (Müller, 1991; Roberts et al., 1997).

Denote the order statistics of the random sample $X_1, \ldots, X_n$ by $X_{1,n} \leq \cdots \leq X_{n,n}$. In practice we can take $T = X_{n-k,n}$, for some appropriate k, which must be a function of n such that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$. When the threshold T is an order statistic as above, we have $\lambda = k$. It is important to emphasize that the choice of k is important for practical purposes, but beyond the scope of this paper.