

A Bayesian nonparametric approach for extreme values

Un approccio bayesiano non-parametrico per i valori estremi

Claudia Donatone¹

Dipartimento Statistica, Probabilità e
Statistiche Applicate
Università "La Sapienza" Roma
claudia.donatone@uniroma1.it

Pietro Muliere

Istituto Metodi Quantitativi
Università "L. Bocconi" Milano
pietro.muliere@uni-bocconi.it

Riassunto: In questo articolo viene presentato un approccio bayesiano non-parametrico all'analisi dei valori estremi utilizzando un processo di Poisson non-omogeneo. La funzione media del processo viene modellata a priori come un processo ad incrementi indipendenti e successivamente si ottengono le corrispondenti distribuzioni a posteriori utilizzando uno schema di campionamento di Gibbs.

Keywords: Extreme values, Bayesian nonparametrics approach, Extended gamma process, Lévy process, Log-beta process.

1. Introduction

Extreme values techniques are becoming increasingly popular in financial applications. The complex stochastic structure of financial markets does mean, however, that simple application of extreme value techniques can be misleading. For this reason, we have used a Bayesian nonparametric approach to analyze the behavior of these values, that allows to make inference and prediction on them.

We define an extreme value as an observation that exceeds a predefined threshold, u . The idea is modeling threshold exceedances using a nonhomogeneous Poisson process (NHPP) and to assign to the intensity of the process either log-beta process prior or extended gamma process prior. Posterior inferences are carried out via a Gibbs sampling scheme.

2. The model

We assume that threshold exceedances (threshold choice based on the analysis of the Poisson process approach used in the classical theory of the extreme values) follow a NHPP with the intensity, $\Lambda(\cdot)$, assumed to be a random measure.

The likelihood. In this section we present a detailed Bayesian analysis of NHPP, in order to make inference on the extreme values.

Following Kuo and Ghosh (1997), let us consider a time-truncated model where the process is observed up to a fixed time τ . We denote the ordered epochs of the n observed jumps by $0 = x_0 < x_1 < \dots < x_n \leq \tau$. If we allow ties in the observed jumps, then

¹Indirizzo per corrispondenza: Dipartimento Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza", P.le Aldo Moro 5, 00185, Roma, Italia. E-mail: claudia.donatone@uniroma1.it

the probability of observing no jumps in the interval $(0, x_1)$, d_1 jumps at x_1 , no jumps in (x_1, x_2) , and so on up to no jumps in (x_n, τ) , is given by

$$\ell(\Lambda|D, \tau) = \left\{ \prod_{i=1}^n [\Lambda(x_i) - \Lambda(x_i^-)]^{d_i} e^{-[\Lambda(x_i) - \Lambda(x_i^-)]} e^{-[\Lambda(x_i^-) - \Lambda(x_{i-1})]} \right\} e^{-[\Lambda(\tau) - \Lambda(x_n)]}.$$

Here, D denotes the data and d_i is the number of multiple jumps observed at time x_i . The total number of observed jumps in the interval $(0, \tau]$ is then $d = \sum_{i=1}^n d_i$.

The prior. We will assign a Lévy additive process prior, $L(\cdot)$ to $\Lambda(\cdot)$ (Muliere and Walker (2000)); in particular we used the *log-beta process* (Walker and Muliere (1997)), $\Lambda(\cdot) \sim \text{logBeP}\{\alpha(\cdot), \beta(\cdot)\}$ and the *extended gamma process* (Dykstra and Laud (1981)), $\Lambda(\cdot) \sim \text{EGaP}\{\alpha(\cdot), \beta(\cdot)\}$, defined exactly via Lévy process.

The prior can be characterized by

$$M = \{t_1, t_2, \dots\}, \{f_{t_1}, f_{t_2}, \dots\},$$

the set of fixed points of discontinuity together with the corresponding density functions for jumps, and $N_t(\cdot)$, the Lévy measure for the part of the process without fixed points of discontinuity. We assume the Lévy measure to be of the form

$$dN_t(\nu) = \int_{[0, t]} K(\nu, s) ds, \quad (1)$$

where $K(\nu, s)ds$ is the Lévy measure kernel. In particular, if

$$K(\nu, s)ds = 1 - \exp(-\nu)^{-1} \exp\{-\nu\beta(s)\} d\alpha(s) \quad (2)$$

then $L(\cdot)$ is a log-beta process (see Walker and Muliere (1997)).

Instead, if

$$K(\nu, s)ds = \nu^{-1} \exp\{-\nu\beta(s)\} d\alpha(s) \quad (3)$$

then $L(\cdot)$ is an extended gamma process (see Dykstra and Laud (1981)), where $\alpha(\cdot)$ is a continuous measure on the interval $[0, \infty)$ and $\beta(\cdot)$ is a piecewise continuous nonnegative function.

The posterior. Following Theorem 4.2 in Walker and Muliere (1997), it is possible to deduce the following:

1. If $\Lambda(\cdot)$ is an extended gamma process then the posterior distribution given data is also an extended gamma process;
2. If $\Lambda(\cdot)$ is a log-beta process then the posterior distribution given data is also a log-beta process.

In particular, if $\Lambda(\cdot) \sim \text{EGaP}\{\alpha(\cdot), \beta(\cdot)\}$, that is, if $\Lambda(\cdot)$ is an extended gamma process *without* prior fixed points of discontinuity, then the posterior distribution of $\Lambda(\cdot)$ is again an extended gamma process with parameters $\alpha^*(\cdot) = \alpha(\cdot)$ and $\beta^*(\cdot) = \beta(\cdot) + 1$, and *with* fixed points of discontinuity at $M^* = \{x_1, \dots, x_n\}$ with posterior distribution for the jumps $f_{x_i}^*(\nu) = \text{Ga}(\nu \mid d_i, \beta(x_i) + 1)$, $i = 1, \dots, n$ (we will use $(*)$ to denote the updated parameter/function).

Instead, if $\Lambda(\cdot) \sim \text{logBeP}\{\alpha(\cdot), \beta(\cdot)\}$, that is, if $\Lambda(\cdot)$ is a log-beta process *without* prior fixed points of discontinuity, then the posterior distribution of $\Lambda(\cdot)$ is again a Lévy process with a log-beta measure for the continuous part, with parameters $\alpha^*(\cdot) = \alpha(\cdot)$ and $\beta^*(\cdot) = \beta(\cdot) + 1$, and *with* fixed points of discontinuity at $M^* = \{x_1, \dots, x_n\}$ with posterior distribution for the jumps $f_{x_i}^*(\nu) \propto \nu^{d_i} e^{-\nu} k(\nu, x_i)$, $i = 1, \dots, n$, where $k(\nu, s)$ is given by (2).

3. Financial Application

In this section we describe how we applied the model discussed before in order to obtain inferences on the extreme values. We have used the daily time series of Coca Cola share price, over a 10-year period (1990-2000).

The aim of this analysis is to study the extreme values (threshold exceedances) of a financial series. Many empirical studies on series of this type have indicated that an approximation to stationarity can be obtained by taking logarithms of ratios of successive observations - the so called log-daily returns.

First of all, we have used the classical theory of the extreme values to choose the optimal threshold u . To do this we have compared the two most popular used distributions: the Generalized Pareto Distribution (GPD) and the Poisson Process characterization (Coles (2001)), and we have analyzed the two likelihoods and their parameters estimation. By this comparison, we have concluded that it is convenient to work with a threshold set at $u = 3.8$. Therefore, the data-set consists of the times in which the stock price exceeded this threshold. Note that the stock is considered risky when exceeding this critical level. Then, we have used the algorithm implemented by Gutiérrez-Peña and Nieto-Barajas (2003). We assigned to $\Lambda(\cdot)$ either the log-beta process prior or the extended gamma process prior. We centered the prior on a Weibull model, as discussed in Gutiérrez-Peña and Nieto-Barajas (2003). We ran the Gibbs sampler for 5000 iterations with a burn-in period of 1000 iterations. Convergence was based on plots of $\Lambda(t)$ for a few selected values of t . Figure 1(a)(b) shows the data (*), the posterior mean (continuous line), the median (dotted line) and a 95% credible band (dashed lines) for the mean function. Figure 1(a) shows that the posterior mean of the intensity measure is increasing as days go by: using the log-beta process prior, the posterior mean is not much sensible to the clusters of the observations; while Figure 1(b) gives more about the trend of the data.

The prediction distribution is been calculated as summarize of random draws from a Poisson distribution and it's represented by a nonparametric histogram density estimation of the number of exceedances in the 100 days after the observed period in Figure 2(a)(b).

It is obvious, comparing Figure 2(a) with 2(b), that the log-beta process prior gives us more information on the future riskness of the share. On the contrary, using the extended gamma process prior, the distribution presents a very asymmetrical trend that doesn't allow to get more information on the stability of the considered share.

Figure 1: Posterior mean, median and 95% band for $\Lambda(t)$ (logBeP (a) and EGaP (b)).

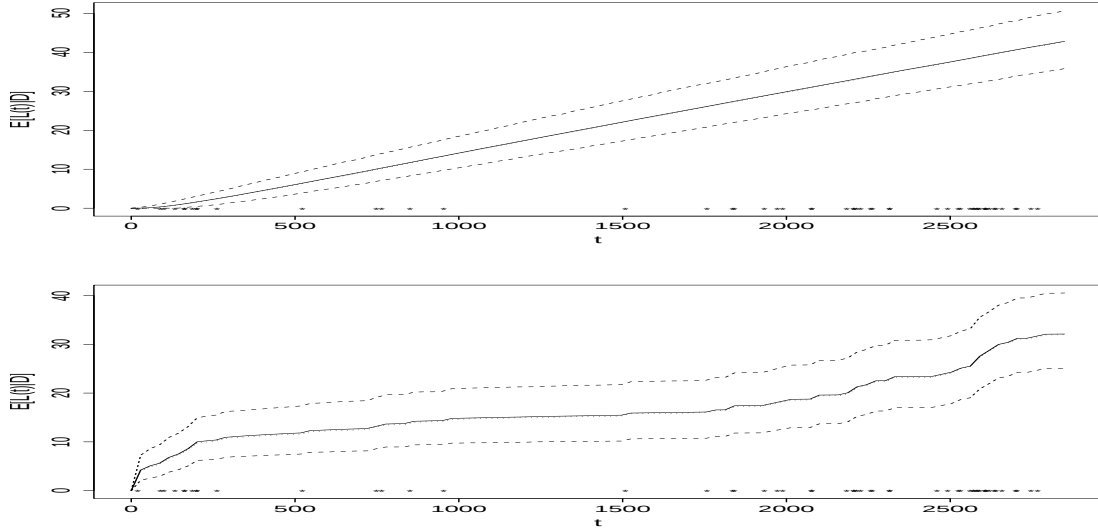
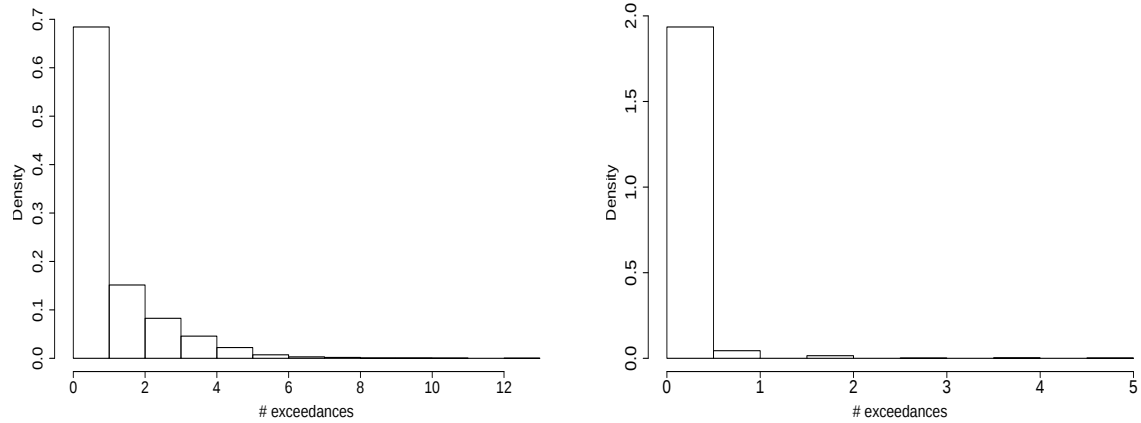


Figure 2: Predictive distribution of number of exceedances (logBeP (a) and EGaP (b)).



References

- Coles S. (2001) *An introduction to Statistical Modelling of extreme Values*, London: Springer.
- Dykstra R. and Laud P. (1981) A bayesian nonparametric approach to reliability, *Ann. Statist.*, 9, 356–367.
- Gutiérrez-Peña E. and Nieto-Barajas L.E. (2003) Bayesian nonparametric inference for mixed poisson processes, in: *Bayesian Statistics 7*.
- Kuo L. and Ghosh S.K. (1997) Bayesian nonparametric inference for nonhomogeneous poisson processes, Technical Report 9718, Department of Statistics. University of Connecticut, Storrs, USA.
- Muliere P. and Walker S. (2000) Neutral to the right processes from a predictive perspective: a review and new developments, *Metron*, LXVIII n.3-4, 13–30.
- Walker S. and Muliere P. (1997) Beta-stacy processes and a generalization of the Pólya-urn scheme, *Ann. Statist.*, 25, 1762–1780.