

Interval measurements of large volcanic eruptions

Misurazioni intervallari di eruzioni vulcaniche estreme

Claudia Furlan¹

Dipartimento di Scienze Statistiche,

Università degli Studi di Padova, e-mail: furlan@stat.unipd.it

Riassunto: Per la difficoltà nell'identificare la magnitudine precisa degli eventi vulcanici, le eruzioni vengono usualmente registrate tra un valore minimo e uno massimo e poi, usualmente, se ne cataloga la media. L'obiettivo è quello di capire quali differenze si possano ottenere, in fase di interpretazione e previsione delle eruzioni vulcaniche, se si considera l'informazione censurata originaria o le medie degli intervalli.

Keywords: Interval measurements, volcanic eruptions, extreme values, random effect models, Pareto distribution.

1. Introduction

We present a study on the dataset, found in Mason *et al.* (2004), of the most extreme eruptions on Earth, recorded in $M = \log_{10} m - 7$ where m is the estimated mass generated by the volcanic event, recorded in kg. The dataset is made up of the 47 largest volcanic eruptions in the Earth's history, of which only 3 are exactly recorded: indeed, most of magnitude measurements are usually taken into intervals and this leads to a problem of censored information, since eruptions are bounded between a maximum and a minimum value. Extreme value theory is the branch of statistics inferring extreme events in random processes and, of this field, the Pareto distribution attends to modelling the extreme values from a threshold.

2. Maximum likelihood versus Bayesian MCMC approach

In this section we approach the inference problem through the maximum likelihood method, for both the original information and the interval means, and through an MCMC algorithm in a Bayesian setting for the original measurements.

Maximum likelihood approach. *Original measurements:* Let X_i be the magnitude of the i -th eruption, $i = 1, \dots, 47$, let $u = 8$ be the threshold that indicates the lower bound of measurements and let $Y_i = X_i - u$ be modelled by a Generalized Pareto distribution, $Y_i \sim GPD(\sigma, \xi)$. Let A be the index subset corresponding to those eruptions exactly recorded, let $H(y)$ be the Pareto distribution function, let $H'(y)$ be the Pareto density, let Y_{mi} and Y_{Mi} be the minimum and maximum value for magnitude of the i -th eruption, then the *likelihood with the interval measurements* becomes:

$$L((y_{mi}, y_{Mi}), i = 1, \dots, 47) = \prod_{i \notin A} (H(y_{Mi}) - H(y_{mi})) \prod_{i \in A} H'(y_i). \quad (1)$$

¹Address of correspondence: Via Cesare Battisti 241, 35121 Padova

Interval means: Let $\bar{Y}_i = (Y_{mi} + Y_{Mi})/2$, $i = 1, \dots, 47$ be the mean of the i -th interval measurement, that is commonly reported as the measurement of the magnitude. Let $\bar{Y}_i \sim GPD(\sigma, \xi)$, then the *likelihood with the interval means* is:

$$L(\bar{y}_i, i = 1, \dots, 47) = \prod_{i=1}^{47} H'(\bar{y}_i). \quad (2)$$

Parameters σ and ξ have been estimated in both likelihood (1) and (2) with the maximum likelihood method.

Bayesian MCMC approach: a hierarchical model. Let Y_i be the real magnitude of an eruption and let $Y_i \sim GPD(\sigma, \xi)$, $i = 1, \dots, 47$. Consider an interval that contains Y_i , not necessarily symmetric with respect to Y_i , of width a . To simulate the real process of measuring the magnitude, let's suppose that $Z_i \sim U[Y_i - a/2, Y_i + a/2]$ would be the central value of the interval and, consequently, $Y_{mi} = Z_i - a/2$ and $Y_{Mi} = Z_i + a/2$ would be the lower and upper bound, respectively. The hierarchical model considered is:

$$\begin{aligned} Y_{mi}|Y_i &\sim U[Y_i - a, Y_i] & Y_{Mi}|Y_i &\sim U[Y_i, Y_i + a] \\ Y_i|\sigma, \xi &\sim GPD(\sigma, \xi) \\ \log \sigma &\sim N(0, 10^3) & \log \xi &\sim N(0, 10^3). \end{aligned} \quad (3)$$

Prior distributions are vague since no additional information is available at the moment. Parameters $\{Y_i\}$, σ and ξ in model (3) have been estimated through a Metropolis-Hastings MCMC algorithm (Gilks *et al.*, 1996). The algorithm was previously performed in a simulation study to check its reliability. When the algorithm was applied to data, the Markov chains of parameters preserved their stability with a good mixing, forgetting quickly the relative starting points.

Results and discussion. We applied the models to data (recorded in m) and all the 3 techniques gave approximatively the same answers about the estimate of σ and ξ : that is, considering the original measurements instead of the interval means does not lead to any significant improvements. In this case, since prior distributions of model (3) are vague the Bayesian approach is very similar *from a practical point of view* to the correspondent model (1) with the original measurement, based on the maximum likelihood method. Anyway, model (3) will be more helpful when additional information would be available. Usually only the interval means are available for statistical purposes, as in Furlan (2004), and it would be reasonable to raise a question about the reliability of the results obtained from analyses with this kind of censored data: in this work there is not evidence to conclude that relevant information is lost considering exclusively the interval means, rather than the original measurements, enabling to pass from models (1)-(3) to model (2) for the analysis of the magnitude of the most extreme volcanic events.

References

- Furlan C. (2004) Bayesian techniques for modelling volcanic processes, *Proceedings of the 19th International Workshop on Statistical Modelling*, 179–183.
- Gilks W., Richardson S. and Spiegelhalter D. (Eds.) (1996) *Markov chain Monte Carlo in practice*, Chapman & Hall, London.
- Mason B., Pyle D. and Oppenheimer C. (2004) The size and frequency of the largest explosive eruptions on earth, *Bulletin of Volcanology*, doi:10.1007/s00445–004–0335–9.