The Annals of Applied Statistics
2013, Vol. 7, No. 4, 1891–1894
DOI: 10.1214/13-AOAS614F
Main article DOI: 10.1214/12-AOAS614

DISCUSSION OF "ESTIMATING THE HISTORICAL AND FUTURE PROBABILITIES OF LARGE TERRORIST EVENTS" BY AARON CLAUSET AND RYAN WOODARD

BY QIURONG CUI, KARL ROHE AND ZHENGJUN ZHANG

University of Wisconsin-Madison

First and foremost, we commend the authors for their creative and original investigation. Although this comment will focus on methodological concerns, we note that the conclusions in Clauset and Woodard (2013) rely on several strong modeling assumptions. For example, the number of deaths in a terrorist incident is an independent draw from some unknown probability distribution that is fixed in space and time. Readers who are more interested in how Clauset and Woodard (2013) contribute to the discussion on foreign policy and/or national security should note that, in general, the advantage of statistical modeling is not necessarily that the solutions are precise, but rather that all assumptions are made explicit. Given the politically charged nature of this problem, we are wary of the assumptions (and thus conclusions) in the paper.

Many of the inferences in Clauset and Woodard (2013) (hereafter referred to as CW) rely on the bootstrap to measure the uncertainty in their statistical estimators. Similarly, other applied papers in the extreme value literature have relied on the bootstrap [e.g., Mohtadi and Murshid (2009)]. However, there are many scenarios in which the bootstrap can fail. Both Resnick (2007) and Hall and Weissman (1997) discuss some of these problems in the context of heavy-tailed distributions. In this discussion we provide a brief simulation that illustrates when the bootstrap succeeds and when it fails in the settings of CW.

This comment investigates the following question under three relevant models:

> If the bootstrap is used to create a (nominally) 90% confidence interval, will this interval actually cover the true parameter in 90% of experiments?

The first simulation model is the power law distribution with $\alpha = 2.4$ supported on $[10, \infty)$. The second simulation model comes from Clauset, Shalizi and Newman (2009), a paper that CW cite to justify their method of estimating $x_{\min}$. To sample a point X_i from this model (which we will refer to as the mixed power-law model), sample an observation Y_i uniformly at random from the RAND-MIPT data [MIPT, 2009] and sample an observation Z_i from power law with $\alpha = 2.4$ (corresponding to the estimate in CW) and supported on $[10, \infty)$. Then,

$$X_i = \begin{cases} Y_i, & \text{if } Y_i < x_{\min} = 10, \\ Z_i, & \text{o.w.} \end{cases}$$

Received August 2013; revised August 2013.

To investigate whether the bootstrap techniques in CW are sensitive to model misspecification error, the final simulation model is the Generalized Pareto Distribution (GPD) Coles (2001). The GPD distribution is specified by three parameters: location u, scale σ and shape ξ. The cumulative distribution function of the GDP distribution is

$$F_{u,\sigma,\xi}(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right), & \xi = 0. \end{cases} \tag{1}$$

When $\xi > 0$, the GPD distribution is regularly varying at ∞ with index $-1/\xi$ and thus tail equivalent to the power-law distribution with $\alpha = \xi^{-1} + 1$. Therefore, the GPD distribution can be considered as a perturbed power-law distribution. For simplicity, we set $u = 0, \sigma = 1$ in the following simulation study. For comparison purposes, ξ is set to be $1/(\alpha - 1) = 1/1.4$ for $\alpha = 2.4$.

In each of 1000 runs of the experiment, we sample $n = 1000$ data points from each of the simulation models, and we use the code from CW to (1) fit the power-law distribution and (2) compute the (nominally) 90% bootstrap confidence intervals for α and p, the probability of observing at least one catastrophic event. For each of the three models, we run the estimation two ways: First, with $x_{\min} = 10$ given to the algorithm and, second, where the algorithm estimates $x_{\min}$. Finally, each bootstrap confidence interval is inspected to see if it contains the true values of α and p (see next paragraph for a discussion on computing the true value of p). In total, this creates six different simulation setups. The simulation results are given in Table 1 and Figure 1.

To compute the true value of p, we follow the definition in CW:

$$p = 1 - E\left[P(X < \text{cat} \mid X > x_{\min})^{n_{\text{tail}}}\right], \tag{2}$$

where $n_{\text{tail}} \sim \text{Binomial}(n = 1000, p_{\text{tail}})$, and cat is the size of 9/11 (2749). In all six simulation setups, this value of p is approximately the probability that the maximum of $n = 1000$ draws is greater than cat.[1]

In the power-law and mixed power-law distributions $x_{\min} = 10$ is the lower end of the power-law distribution for the tail model. Using equation (2), $p = 0.3194$ under the power law and $p = 0.0244$ under the mixed power law. Under the GPD distribution an exact $x_{\min}$ is not available for computing p. We therefore applied an analogy of Kolmogorov–Smirnov criteria to minimize the maximum distance of the GPD distribution above $x_{\min}$ and power-law distribution, that is, we find $(x_{\min}, \alpha)$ that minimize

$$f(x_{\min}, \alpha) = \max_{x: x > x_{\min}} \left| \left(\frac{\sigma + \xi(x-u)}{\sigma + \xi(x_{\min} - u)}\right)^{-1/\xi} - \left(\frac{x}{x_{\min}}\right)^{1-\alpha} \right|.$$

[1] In fact, for $n = 1000$, they are equal in the first four decimal places.

TABLE 1

For each of the three simulation models, $x_{\min}$ is either given as 10 or estimated (est.) by the algorithm. The table presents (a) the estimated bias in coverage probabilities for the (nominally) 90% bootstrap confidence intervals and (b) the median width of the confidence intervals

Model	$x_{\min}$	Coverage bias (%)		Width of CI	
		α	p	α	p
PL, $\alpha = 2.4$	10	−1.0	−0.9	0.1	0.2
PL-Mix, $\alpha = 2.4$	10	−28.9	−25.3	0.7	0.03
GPD, $\xi = 1/1.4$	10	−0.6	2.4	0.6	0.1
Power Law, $\alpha = 2.4$	est.	1.5	0.9	0.2	0.3
PL-Mix, $\alpha = 2.4$	est.	0.5	0.4	0.8	0.1
GPD, $\xi = 1/1.4$	est.	−11.6	−18.4	0.5	0.2

Although the limiting behavior of $x_{\min}$ is left an undeveloped problem in CW, the solution of the above optimization problem could be a possible option heuristically. Numerical optimization (grid search in matlab) yields $\alpha = 2.31$ and $x_{\min} = 13.44$. With these values, the value of p in equation (2) is 0.0242.

Table 1 reports the results for the bootstrap confidence intervals for both the probability p of a catastrophic event and α, the power-law parameter. In brief, out of the six different setups, the confidence intervals failed for two of them. Under both (1) the mixed power-law distribution [from Clauset, Shalizi and Newman (2009)] with $x_{\min}$ given and (2) the GPD with $x_{\min}$ estimated, the nominally 90% bootstrap confidence intervals cover the true values of p with probabilities 0.64 and 0.71, respectively. One reason the bootstrap fails under the mixed power-law distribution is possibly due to an observation in CW, that a fixed choice of $x_{\min}$ underestimates the uncertainty in $\hat{p}$ due to the tail's unknown structure. The reason

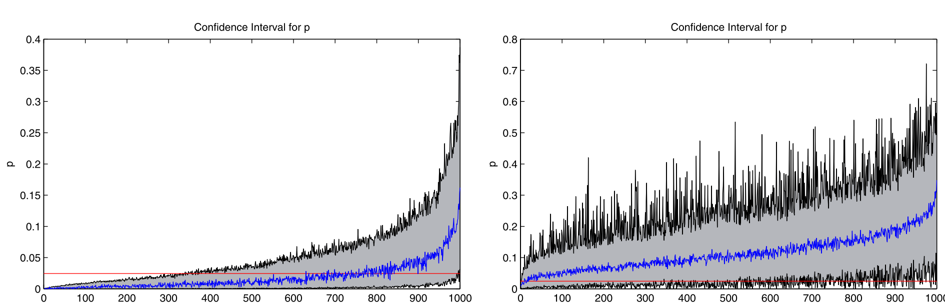

FIG. 1. *One thousand confidence intervals for p under the mixed power-law distribution with $x_{\min} = 10$ (on left) and GPD with $x_{\min}$ estimated (on right). These confidence intervals have poor coverage properties for the true value of p (represented as a red line). Each confidence interval is computed from a sample of $n = 1000$ data points and 1000 bootstrap samples. The order of the intervals is shuffled so that $\hat{p}$ (blue line) is increasing.*

the bootstrap fails under the GPD is that the algorithm tends to underestimate $x_{\min}$ and α, that is, it is inclined for heavier tails. This is consistent with the other discussants who suggest that $x_{\min}$ is potentially too small.

Although the bootstrap gives a straightforward path to computing confidence intervals, these simulations suggest that their coverage performance is sensitive to the data-generating model and whether or not $x_{\min}$ is estimated or known.

REFERENCES

CLAUSET, A., SHALIZI, C. R. and NEWMAN, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* **51** 661–703. MR2563829

CLAUSET, A. and WOODARD, R. (2013). Estimating the historical and future probabilities of large terrorist events. *Ann. Appl. Stat.* **7** 1838–1865.

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. MR1932132

HALL, P. and WEISSMAN, I. (1997). On the estimation of extreme tail probabilities. *Ann. Statist.* **25** 1311–1326. MR1447753

MIPT (2009). Terrorism knowledge base. Acessed 19 August 2013.

MOHTADI, H. and MURSHID, A. P. (2009). Risk of catastrophic terrorism: An extreme value approach. *J. Appl. Econometrics* **24** 537–559. MR2668785

RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York. MR2271424

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706
USA
E-MAIL: cui@stat.wisc.edu
karlrohe@stat.wisc.edu
zjz@stat.wisc.edu