Applications of exact extreme value theorem

Mikhail Makarov
EVMTech, Baarerstrasse 2, 6300 Zug, Switzerland; email: mmakarov@evmtech.com

In this paper we prove that it is sufficient to use Pareto distribution to approximate the tail of a slowly varying heavy-tailed distribution. Several applications of the results are considered.

INTRODUCTION

An extreme value theorem due to Gnedenko (1943), Balkema and de Haan (1974) and Pickands (1975) states that, for a broad class of the distributions, as the truncation point increases, the excess distribution converges to a generalized Pareto distribution (GPD). A result that seems to have been overlooked states that, for heavy-tailed distributions, the excess distribution converges to a specific type of GPD namely the shifted Pareto distribution. More precisely, the convergence results can be stated in two different ways.

(A) The non-exact theorem states that for heavy-tailed distributions, the distribution function is slowly varying if and only if the excess distribution converges to a *generalized Pareto* distribution.
(B) The exact theorem states that for heavy-tailed distributions, the distribution function is slowly varying if and only if the excess distribution converges to a *Pareto* distribution.

The non-exact theorem (A) is well known and is used to justify fitting distribution tails with GPD. Clearly, the exact result is stronger than the non-exact result and allows for replacing GPD with a much more simple Pareto distribution. In this paper we provide a simple proof of the convergence result (B) and show how, by using Pareto distribution instead of GPD, one can significantly simplify many computations.

Let us recall the main definitions and results for application of extreme value theory (EVT) to the excess distribution convergence (Embrechts (2000)).

For a random variable X with cumulative distribution function (cdf) F and a threshold u, the excess distribution function F_u is defined as follows:

$$F_u = \Pr[X - u \leqslant x \mid X > u] = \frac{F(x+u) - F(u)}{1 - F(u)}$$

To define the set of distributions for which EVT can be applied we introduce the following.

DEFINITION 1 A function L is slowly varying if for any $t > 0$:

$$\lim_{u \to \infty} \frac{L(tu)}{L(u)} = 1$$

DEFINITION 2 $F \in RV_\xi$ with $\xi > 0$ if $F(x) = 1 - L(x)x^{-1/\xi}$ for some slowly varying function L.

In order to formulate the convergence statement we introduce the notion of distance between distributions, namely for two cdfs F and G we define the distance as:

$$d(F, G) = \sup_x |F(x) - G(x)|$$

Finally recall that, for $\xi > 0$, the GPD is given by:

$$G_{\xi,\beta}(x) = 1 - (1 + \xi x/\beta)^{-1/\xi} \quad \text{for } x \geqslant 0$$

THEOREM 3 (Pickands–Balkema–de Haan) *For a cdf F, $F \in RV_\xi$ with $\xi > 0$ if and only if there exists $\beta(u)$ such that:*

$$\lim_{u \to \infty} d(F_u, G_{\xi,\beta(u)}) = 0$$

The above theorem is usually formulated for a wider class of distributions than that which we consider (RV_ξ with $\xi > 0$), but we will limit ourselves to the case of heavy-tailed distributions ($\xi > 0$). Although the theorem is formulated as an existence result, the function $\beta(u)$ seems to be unknown; in fact it is the following.

LEMMA 4 *For a cdf F, $F \in RV_\xi$ with $\xi > 0$ the function $\beta(u)$ can be selected as $\beta(u) = \xi u$.*

Interestingly enough, GPD with $\beta(u) = \xi u$ is equivalent to a Pareto distribution.

Let us recall that a cdf of a Pareto distribution is usually defined as:

$$P_{\alpha,u}(x) = 1 - \left(\frac{u}{x}\right)^\alpha \quad \text{for } x \geqslant u$$

A Pareto distribution, unlike an excess distribution, does not start from 0. However if one shifts it to 0 and replaces parameter α by $1/\xi$, one obtains:

$$\hat{P}_{\xi,u}(x) = 1 - \left(\frac{u}{x+u}\right)^{1/\xi} \quad \text{for } x \geqslant 0$$

Thus:

$$G_{\xi,\xi u}(x) = \hat{P}_{\xi,u}(x)$$

A Pareto distribution is much easier to use than GPD. We will give several practical applications of this result in the following sections.

EXACT EVT PROOF

In this section we provide a simple proof for the exact EVT theorem for heavy-tailed distributions.

THEOREM 5 (Exact EVT) *For a cdf F, $F \in RV_\xi$ with $\xi > 0$ if and only if:*

$$\lim_{u\to\infty} d(F_u, \hat{P}_{\xi,u}) = 0$$

Moreover, the choice $\beta(u) = \xi u$ is asymptotically unique ie, for any $\beta(u)$ with $\lim_{u\to\infty} d(F_u, G_{\xi,\beta(u)}) = 0$, it follows that $\lim_{u\to\infty} \beta(u)/(\xi u) = 1$.

PROOF Let us introduce an auxiliary distribution function:

$$\Phi_u(x) = \Pr\left[\frac{X-u}{u} > x \,\middle|\, X > u\right] = \frac{F(u(x+1)) - F(u)}{1 - F(u)}$$

Note that Φ_u has two convenient properties:

$$\begin{aligned} d(\Phi_u, \hat{P}_{\xi,1}) &= \sup_{0\leqslant x}\left|\frac{1 - F(u(x+1))}{1 - F(u)} - (x+1)^{-1/\xi}\right| \\ &= \sup_{0\leqslant y}\left|\frac{1 - F(y+u)}{1 - F(u)} - \left(\frac{y}{u} + 1\right)^{-1/\xi}\right| = d(F_u, \hat{P}_{\xi,u}) \end{aligned}$$

and, if $F(x) = 1 - L(x)\, x^{-1/\xi}$, then:

$$\begin{aligned} \Phi_u(x) &= 1 - (x+1)^{-1/\xi}\frac{L(u\,(x+1))}{L(u)} \\ &= \hat{P}_{\xi,1}(x) - (x+1)^{-1/\xi}\left(\frac{L(u\,(x+1))}{L(u)} - 1\right) \end{aligned}$$

($\Rightarrow$) Assume that $F \in RV_\xi$ with $\xi > 0$. Then from the second property of Φ_u and Definition 1 it follows that, for any fixed x:

$$\lim_{u\to\infty} \Phi_u(x) = \hat{P}_{\xi,1}(x)$$

As pointwise convergence to a continuous distribution implies uniform convergence, we conclude that:

$$\lim_{u\to\infty} d(\Phi_u, \hat{P}_{\xi,1}) = \lim_{u\to\infty} \sup_x |\Phi_u(x) - \hat{P}_{\xi,1}(x)| = 0$$

The theorem now follows from the first property of Φ_u.

($\Leftarrow$) The convergence:

$$\lim_{u\to\infty} d(F_u, \hat{P}_{\xi,u}) = 0$$

together with the first property of Φ_u implies that, for any x:

$$\lim_{u\to\infty} \Phi_u(x) = \hat{P}_{\xi,1}(x)$$

From the second property of Φ_u we conclude that, for any $x > 0$:

$$\lim_{u\to\infty} \frac{L(u(x+1))}{L(u)} = 1$$

Thus, for any $t > 1$:

$$\lim_{u\to\infty} \frac{L(tu)}{L(u)} = 1$$

As for $0 < t < 1$, $1/t > 1$ we obtain:

$$\lim_{u\to\infty} \frac{L(t\,u)}{L(u)} = \lim_{w\to\infty} \frac{L(w)}{L((1/t)w)} = 1$$

Therefore, by Definition 1, L is a slowly varying function and $F \in RV_\xi$.

To prove that the choice $\beta(u) = \xi u$ is asymptotically unique note that:

$$\begin{aligned} 2^{-1/\xi} - \left(1 + \frac{u\xi}{\beta}\right)^{-1/\xi} &= |G_{\xi,\beta(u)}(u) - \hat{P}_{\xi,u}(u)| \leqslant d(G_{\xi,\beta(u)},\, \hat{P}_{\xi,u}) \\ &\leqslant d(\hat{P}_{\xi,u},\, F_u) + d(F_u,\, G_{\xi,\beta(u)}) \end{aligned}$$

The right-hand side of the above inequality converges to 0. Thus, $\lim_{u\to\infty} \beta(u)/(\xi u) = 1$. □

HILL'S ESTIMATOR

As the first application of exact EVT, we will obtain Hill's estimate for the tail coefficient ξ. Assume that we have some losses from a RV_ξ distribution and we need to estimate the tail coefficient ξ. The tail coefficient can be estimated as follows:

1. Select a threshold $u > 0$.
2. Select all losses greater than or equal to u and subtract u from them.
3. Fit the shifted Pareto distribution to the selected losses using a maximum likelihood estimator. Use the maximum likelihood solution ξ as the estimator.

The maximum likelihood method when applied to a Pareto distribution leads to simple formulas and therefore to a simple estimate for ξ. Denote by $x_k, \ldots, x_n$ the losses greater than or equal to a selected threshold u. The probability distribution function for the shifted Pareto distribution is given by:

$$\hat{p}_{\xi,u}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\hat{P}_{\xi,u}(x) = \frac{\mathrm{d}}{\mathrm{d}x}(1 - (1 +\ x/u)^{-1/\xi}) = \frac{1}{\xi u}(1 +\ x/u)^{-1/\xi - 1}$$

The loglikelihood function is equal to:

$$\sum_{i=k}^{n} \log[\hat{p}_{\xi,u}(x_k - u)] = \sum_{i=k}^{n} [(-1/\xi - 1)\log[x_k/u] - \log[\xi u]]$$

The only parameter that needs to be estimated is ξ. Taking the derivative of the loglikelihood function with respect to ξ we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\sum_{i=k}^{n}\log[\hat{p}_{\xi,u}(x_k-u)]=\sum_{i=k}^{n}[(1/\xi^2)\log[x_k/u]-1/\xi]$$

Setting the derivative to be equal to zero, we obtain the estimator for the tail coefficient ξ:

$$\xi=\frac{1}{n-k+1}\sum_{i=k}^{n}\log[x_k/u].$$

The above estimate is exactly the Hill's estimator (Hill (1975)).

CONVERGENCE RATE

As the second application of the exact EVT we consider the speed of the EVT convergence as a function of the threshold. Determining the speed of convergence helps in selecting the threshold to obtain a desired precision and is clearly of large practical interest. Using the exact EVT theorem we will show that the rate of convergence can be very slow.

Let us consider a cdf:

$$F(x)=1-\frac{\log(x)}{x}$$

Clearly $F\in RV_1$ and therefore its tail is approximated by a Pareto distribution with $\xi=1$. To estimate the convergence rate we need to compute $d(F_u,\hat{P}_{1,u})$ where:

$$F_u(x)=1-\frac{\log(x+u)}{\log(u)}\frac{u}{x+u};\qquad \hat{P}_{1,u}(x)=1-\frac{u}{x+u}$$

To find the point with the largest distance between $F_u(x)$ and $P_{1,u}(x)$ we need to solve the equation:

$$\frac{\mathrm{d}}{\mathrm{d}x}[F_u(x)-P_{1,u}(x)]=0$$

One can easily check that the above equation has a unique solution $x^*=(e-1)u$. Thus:

$$d(F_u,\hat{P}_{1,u})=|F_u((e-1)u)-\hat{P}_{1,u}((e-1)u)|=\frac{1}{e\log(u)}$$

The above example demonstrates that the convergence rate can be very slow. For example, if the original threshold is 10,000, in order to double the precision of approximation one needs to increase the threshold to 100,000,000.

CONCLUSIONS

The non-exact version of the extreme value theorem is the main justification to use a GPD for distribution tail modeling (cf, Embrechts (2000) and Beirlant *et al* (2004)).

Based on the non-exact version of the theorem, a number of OpRisk models selected GPD to model the large losses. However, after careful analysis of EVT, one realizes that there are several fundamental problems in using EVT and GPD for tail modeling.

The first problem is that EVT cannot be used to estimate quantiles as it does not imply quantile convergence (cf, Makarov (2006)). That is, EVT implies convergence of GPD to the target distribution in the sense of Kolmogorov–Smirnov, but fails, in most cases, to imply the quantile convergence.

In this paper we have raised another problem in using GPD. It has been proved that, for the heavy-tailed distribution, the GPD is in fact the usual Pareto. In view of the result, a question arises: is there any reason to use GPD instead of Pareto distribution to model tails of heavy-tailed distributions?

REFERENCES

Balkema, A. A., and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability* **2**(5), 792–804.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes (Wiley Series In Probability and Statistics)*. Wiley, New York.

Embrechts, P. (2000). *Extremes and Integrated Risk Management*. Risk Books, London.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44**, 423–453.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**(5), 119–131.

Makarov, M. (2006). Extreme value theory and high quantile convergence. *Journal of Operational Risk* **2**, 51–57.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119–131.