ST-2: Extreme Events: Statistical Extreme Value Theory and Its Applications
2010 Casualty Loss Reserve Seminar

Presented by Robert A. Bear
Consulting Actuary and Arbitrator
RAB Actuarial Solutions, LLC
www.rabsolutions.net

Likelihood of Extreme Events

- We want to estimate the likelihood of extreme events because this is a key component of ERM work, among other reasons.
- The data is not sufficiently credible to permit one to directly specify a distribution and estimate its parameters.
- Recall that given a sequence of iidrv's, the Central Limit Theorem (CLT) gives us a limiting distribution for the sample mean that is used as an approximation for large sample sizes.
- Note that the approximating distribution given by the CLT for the sample mean $\bar{X}$ is normal regardless of the distribution of the underlying observations.
- EVT uses an analogous approach to estimate approximating distributions for sample extremes.

Distribution of Sample Maximum

- Given iidrv sequence $X_1,...,X_n$ with CDF F, we want to estimate the distribution of
$$M_n = \max\{X_1,...,X_n\}$$
- The observations usually represent values of a process measured at regular intervals, so that M_n represents the maximum of the process over n time units.
- We would like to apply the exact formula below:
$$\Pr\{M_n \le z\} = \{F(z)\}^n$$
- Since we don't know F, we look for approximations that can be estimated based upon extreme data only. This approach is an extreme value analog of the central limit theory.

Extremal Types Theorem

- If there exist sequences of constants such that

$$\Pr\{(M_n - b_n)/a_n \le z\} \to G(z)$$

$$as$$

$$n \to \infty$$

where G is a non-degenerate distribution function,

then G is a member of the Generalized Extreme Value (GEV) Family of Distributions.

Generalized Extreme Value (GEV) Family of Distributions

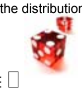
- The generalized extreme value distribution has cumulative distribution function

$$G(z;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

- where 1 + ξ(z − μ) / σ > 0,
- where $\mu \in \mathbb{R}$
 is the location parameter, σ > 0 is the scale parameter and
 ξ is the shape parameter.

Link to Fréchet, Weibull and Gumbel families

The shape parameter ξ governs the tail behavior of the distribution.



- Gumbel (Type I EV distribution): ξ → 0

$$G(z;\mu,\sigma) = e^{-e^{-(z-\mu)/\sigma}} \quad for\ z \in \mathbb{R}$$

- Fréchet (Type II EV distribution): ξ = 1/α > 0

$$G(z;\mu,\sigma,\alpha) = \begin{cases} 0 & z \le \mu \\ e^{-((z-\mu)/\sigma)^{-\alpha}} & z > \mu \end{cases}$$

- Reversed Weibull (Type III EV distribution): ξ = -1/α < 0

$$G(z;\mu,\sigma,\alpha) = \begin{cases} e^{-(-(z-\mu)/\sigma)^{\alpha}} & z < \mu \\ 1 & z \ge \mu \end{cases}$$

Formulas for Means and Variances

The means may be computed from

$$\begin{cases} \mu + \sigma \dfrac{\Gamma(1-\xi)-1}{\xi} & \text{if } \xi \neq 0, \xi < 1, \\ \mu + \sigma\gamma & \text{if } \xi = 0, \\ \text{not exists} & \text{if } \xi \geq 1, \end{cases}$$

where γ is Euler's constant.

The variances may be computed from

$$\begin{cases} \sigma^2 (g_2 - g_1^2)/\xi^2 & \text{if } \xi \neq 0, \xi < \dfrac{1}{2}, \\ \sigma^2 \dfrac{\pi^2}{6} & \text{if } \xi = 0, \\ \text{not exists} & \text{if } \xi \geq \dfrac{1}{2}, \end{cases}$$

where $g_k = \Gamma(1 - k\xi)$.

Approach to Modeling Extremes

(1) Block the data into sequences of observations of length n, generating a sequence of maxima $M_{n,1}, \ldots, M_{n,m}$ to which the GEV distribution can be fitted.

Often the blocks are chosen to correspond to a time period of length one year, so n is the number of observations in a year and the block maxima are annual maxima.

(2) One can easily calculate z_p for which $G(z_p) = 1 - p$

z_p is the return level associated with the return period 1/p, since this quantity is expected to be exceeded on average once every 1/p years.

z_p is exceeded by the annual maximum in any particular year with probability p.

Generalized Pareto Distribution

(1) Let X represent an arbitrary term in the iidrv sequence $X_1, X_2, \ldots$ with common CDF F, and assume that F satisfies the Extremal Types Theorem. Let $M_n = \max\{X_1, \ldots, X_n\}$

Then for large n, $\Pr\{M_n \leq z\} \approx G(z)$

where $G(z; \mu, \sigma, \xi)$ is a member of the Generalized Extreme Value (GEV) Family of Distributions.

(2) Then for large enough u, the distribution of Y=X-u is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y)}{\tilde{\sigma}}\right)^{-1/\xi}$$

and is defined on {y: y>0 and $(1 + \xi y/\tilde{\sigma}) > 0$} where

$\tilde{\sigma} = \sigma + \xi(u - \mu)$ Parameters are function of GEV parameters.

H(y) is known as Generalized Pareto family of distributions (GPD).

- Conclusion: If block maxima have approximate GEV distribution G, then threshold excesses have an approximate distribution within the Generalized Pareto family H with the same shape parameter ξ .

Generalized Pareto Properties

- The GPD is bounded only for negative values of ξ
- The GPD model for threshold excesses is equivalent to the familiar Shifted Pareto

$$H(y) = 1 - \left(\frac{\theta}{y+\theta}\right)^{\alpha} \quad \text{where } \theta = \tilde{\sigma}\alpha \text{ and } \alpha = 1/\xi$$

- The mean of a GPD distribution $H(y;\tilde{\sigma},\xi)$ is

$$E(Y) = \frac{\tilde{\sigma}}{1-\xi} \quad (\xi < 1)$$

- E(y) is a linear function of u. If the GPD is valid for excesses of a threshold u_0 , then it should be equally valid for all thresholds $u > u_0$ with adjustment to the scale parameter $\tilde{\sigma}$

  The plot of (u, average claim excess of u), called the **mean residual life plot**, should be linear in u above a threshold u_0 at which the GPD is a valid approximation to the excess distribution.

Threshold Selection for GPD

(1) Select the smallest threshold u_0 above which the graph of the mean residual plot is approximately linear.

(2) Given a sequence of iidrv's, fit GPD to losses excess of various relatively high thresholds. Let $\tilde{\sigma}_u$ represent the GPD scale parameter for a threshold $u > u_0$
It can be shown that

$$\tilde{\sigma}_u = \tilde{\sigma}_{u_0} + \xi(u - u_0)$$

and so estimates of $\sigma^* = \tilde{\sigma}_u - \xi u = \tilde{\sigma}_{u_0} - \xi u_0$ and ξ should be constant above u_0 if u_0 is a valid threshold for excesses modeled by the GPD.

This suggests plotting both estimates of σ^* and ξ against u, and selecting u_0 as the lowest value of u for which the estimates remain nearly constant.

Return Levels for the GPD

(1) Assume a GPD is a suitable model for the excess of a variable X above a threshold u. For x>u,

$$P\{X > x \mid X > u\} = \left(1 + \frac{\xi(x-u)}{\tilde{\sigma}}\right)^{-1/\xi}$$

Then $P\{X > x\} = \zeta_u \left(1 + \frac{\xi(x-u)}{\tilde{\sigma}}\right)^{-1/\xi}$ where $\zeta_u = P\{X > u\}$

(2) If X_m is the level that is exceeded on average once every m observations, then

$$x_m = u + \frac{\tilde{\sigma}}{\xi}[(m\zeta_u)^{\xi} - 1]$$

provided that m is sufficiently large so that $X_m > u$ and $\xi \neq 0$
(3)If $\xi = 0$ then $x_m = u + \tilde{\sigma}\log(m\zeta_u)$
(4)If there are n_y observations per year and you want the N-year return level, then compute the m-observation return level where $m = N * n_y$
(5) The sample proportion of observations exceeding u is the MLE for ζ_u

Expected Waiting Times Between Extremal Events

(1) Let $\{X_i\}$ represent an iidrv sequence with continuous CDF F.
Let u represent a given threshold and $p = P\{X > u\}$ = 1-F(u)

(2) The time of the first exceedance of u is a geometric rv with distribution

$$P\{L(u) = k\} = (1-p)^{k-1} p$$

The return period of the events $\{X_i > u\}$ is E{L(u)}=1/p

(3) The probability there will be at least one exceedance of u before time k (or within k observations) is

$$r_k = P\{L(u) \leq k]\} = 1-(1-p)^k$$

(4) The probability there will be an exceedance of u before the return period

$$P\{L(u) \leq E[L(u)]\} = 1-(1-p)^{[1/p]}$$

where [x] is the integer part of x.

(5) For high thresholds u (which implies p is very small), the probability that there will be an exceedance of u before the return period approaches 1-1/e = .63212. Thus, for high thresholds, the mean of L(u) is larger than its median. For example, if one is discussing a 1,000 year event, the probability that it will occur before 1,000 years is approximately 63%.

Order Statistics

(1) We have stated that the limiting distribution of block maxima, suitably rescaled, is GEV.

(2) It can be shown that if the kth largest order statistic in a block is normalized in the same way as the maximum, then its limiting distribution is a function of the parameters of the limiting GEV distribution of the block maximum.

(3) Let $X_{k,n}$ represent the kth upper order statistic from a finite sample of iidrv's, $X_1, \ldots, X_n$ with distribution F. Then

$$\mathrm{P}(X_{k,n} \leq x) = P(F_n(x) > 1 - k/n)$$

where $F_n(x)$ is the empirical cdf, the proportion of observations not exceeding x.

(4) If $U_{n,n} < \ldots < U_{1,n}$ are the order statistics from a sample of iidrv's uniformly distributed on (0,1), then a sample of order statistics $X_{n,n} < \ldots < X_{1,n}$ can be simulated by calculating the inverse of F at $U_{n,n}, \ldots, U_{1,n}$

Gumbel's Method of Exceedances

(1) Let $X_{n,n} < \ldots < X_{1,n}$ be the order statistics from an iidrv sequence with continuous distribution F. Let $S_r^n(k)$ represent the number of exceedances of the kth upper order statistic $X_{k,n}$ among the next r observations,

$X_{n+1}, \ldots, X_{n+r}$

Then $S_r^n(k)$ has a hypergeometric distribution.
It follows that the mean number of exceedances of $X_{k,n}$ is given by

$$E[S_r^n(k)] = \frac{rk}{n+1}$$

(2)Application 1: The probability of exceeding the kth order statistic from the last n observations in the next trial is $\frac{k}{n+1}$

(3) Application 2: The probability of a new record high in the next trial is $\frac{1}{n+1}$

Record Counting

(1) A record X_n occurs if it exceeds M_{n-1} (maximum of prior observations)

(2) Given a sequence of iidrv's with continuous distribution F, the expected number of records in n observations is given by

$E(N_n) = \sum_{k=1}^{n} \frac{1}{k}$ with variance $Var(N_n) = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k^2}\right)$

(3) For large sample sizes n, it can be shown that

$$E(N_n) \approx \log(n) + \gamma$$

where $\gamma \approx .5772$ is Euler's constant.

This formula may be used to test the iid hypothesis.
If for a relatively large sample size n, the observed number of records is close to the estimate given by this approximation formula, we would accept the iid hypothesis.

Conclusions

(1) Anytime something bad happens, you invariably hear that "it could always be worse." Now with the aid of Extreme Value Theory, you can go forth and calculate the odds that it will be worse next time or that it will be just less awful!

(2) The actuarial version of "it's 5:00 PM somewhere" is "it's Pareto if you go far enough out in the tail."