Report by Petro Borysov
April 2009

ESTIMATING THE TAILS OF LOSS SEVERITY DISTRIBUTIONS USING EXTREME VALUE THEORY

ALEXANDER J.MCNEIL

1. Introduction

Estimating loss severity distributions from historical data is an important actuarial activity in insurance. The main focus of this paper is estimation of the tails of loss of such distributions. Suggested modeling is based on the extreme value theory (EVT). One of the key results used is Pickands-Balkema-de Haan theorem, which essentially says that, for a wide class of distributions, losses which exceed high enough thresholds follow the Generalized Pareto distribution (GPD). Danish data on major fire insurance losses is analyzed as an illustration of the method.

2. Modeling Loss Severities

Suppose insurance losses are denoted by random variables $X_1, X_2, \ldots$. Assume that loses are identically distributed and independent. Denote their common distribution function by $F_X(x) = P\{X \leq x\}$ where $x > 0$.

Suppose we are interested in a high-excess loss layer with lower and upper attachment points r and R respectively, $r < R$. This means that payout Y_i on a loss X_i is given by

$$Y_i = \begin{cases} 0 & \text{if } 0 < X_i < r, \\ X_i - r & \text{if } r \leq X_i < R, \\ R - r & \text{if } R \leq X_i < \infty. \end{cases}$$

There are two related actuarial problems:

(1) The pricing problem. Given r and R what should this insurance layer cost to the customer?
(2) The optimal attachment point problem. How to choose r for payout greater than a specified amount to occur with at most a specified frequency?

Let N be the number of losses, then aggregate payout would be $Z = \sum_{i=1}^{N} Y_i$ and the pricing problem would be usually reduced to finding moments of Z. A common pricing formula is $Price = E[Z] + k \cdot Var[Z]$, $E[Z] = E[Y_i]E[N]$. The paper provides the calculation of $E[Y_i]$ among other things.

The attachment point problem can be formulated in the following way: how to choose r such that $P\{Z > 0\} < p$ for some fixed p. This problem comes down to the estimation of high quantile of the loss severity distribution $F_X(x)$.

Typically the data available is historical one on losses that exceed a certain amount δ known as displacement, where $\delta \ll r$. The d.f. of the truncated losses

can be defined by

$$F_{X^\delta}(x) = P\{X \leq x \mid X > \delta\} = \begin{cases} 0 & \text{if } x \leq \delta, \\ \frac{F_X(x)-F_X(\delta)}{1-F_X(\delta)} & \text{if } x > \delta. \end{cases}$$

The goal is to find an estimate $\widehat{F_{X^\delta}(x)}$ of the truncated severity distribution $F_{X^\delta}(x)$.

The Danish data analyzed in the paper comprises of 2157 losses over one million Danish Krone. If one would choose layer running from 50 to 200 then there will be only six observed losses. It is crucial for the insurance company to have a good estimate of the severity distribution in the tail.

3. Extreme value theory

Just as normal distribution proves to be limiting distribution for sample sums or averages, Generalized Extreme Value distribution (GEV) is the limiting distribution of sample extrema. Define the d.f. of the GEV by

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp(-(1+\xi(x-\mu)/\sigma)^{-1/\xi}) & \text{if } \xi \neq 0, \\ \exp(-e^{-(x-\mu)/\sigma}) & \text{if } \xi = 0 \end{cases}$$

where $1 + \xi(x - \mu)/\sigma > 0$ and ξ is known as shape parameter.

Generalized Pareto distribution (GPD) can be used to describe the behavior of large observations which exceed high thresholds and to model insurance losses. GPD is usually expressed as

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1+\xi(x-\mu)/\sigma)^{-1/\xi}) & \text{if } \xi \neq 0, \\ 1 - (-e^{-(x-\mu)/\sigma}) & \text{if } \xi = 0 \end{cases}$$

Consider a certain high threshold u which might be the lower attachment point r of a high-excess loss layer. We are interested in the excess above this threshold. Let $x_0 = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty$. Define the distribution function of the excess over the high threshold u by

$$F_u(x) = P\{X - u \leq x \mid X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)}$$

for $0 \leq x < x_0 - u$.

The theorem (Balkema and de Haan 1974m Picknds 1975) shows that under certain conditions the GPD is the limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint. It shows that there exist a positive measurable function $\sigma(u)$ such that

$$\lim_{u \to x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0,$$

The statistical relevance of this result is that it is possible to attempt to fit GPD do data which exceed high thresholds. The theorem gives theoretical grounds to expect that if we choose a high enough threshold, the data will show the GPD behavior.

If it is possible to fit the conditional distribution of the excess above a high threshold, it is also possible to fit it to the tail of the original distribution above the high threshold. For $x \geq u$, i.e. points in the tail of the distribution,

$$F(x) = P\{X \leq x\} = (1 - P\{X \leq u\})F_u(x - u) + P\{X \leq u\}.$$

Here $F_u(x-u)$ can be estimated by $G_{\xi,\sigma}(x-u)$ for u large, $P\{X \le u\}$ can be estimated from the data by $F_n(u)$. This means that for $x \ge u$

$$\widehat{F(x)} = (1 - F_n(u))G_{\xi,u,\sigma}(x) + F_n(u),$$

where ξ and σ are estimated by ML method. Estimated d.f. $\widehat{F(x)}$ is also GPD with same parameter ξ, but with $\tilde{\sigma} = \sigma(1-F_n(u))^{\xi}$ and $\tilde{\mu} = u - \tilde{\sigma}((1-F_n(u))^{-\xi} - 1)/\xi$.

4. Analysis of Danish Fire Loss Data

Exploratory data analysis was performed on Danish fire loss data. A truncated lognormal distribution was fitted using ML method and superimposed the resulting probability density on the histogram. Truncated lognormal appears to provide a reasonable fit but it is difficult to tell about the largest losses which are the main interest.

A concave departure from the ideal shape indicates heavy tail distribution in the QQ-plot against the exponential distribution. Another graphical tool, such as mean excess plot confirmed that data comes from a heavy tail distribution and in particular follows a GPD with positive shape parameter in the tail above threshold $u = 10$ or $u = 20$.

Overall fit of three distributions was compared: truncated lognormal, ordinary Pareto and GPD. The GPD seems to be quite good explanatory model for the highest losses. As mean excess plot suggested the GPD was fitted to those data points which exceed high thresholds of 10 or 20. It was described before that is possible to transform scale and location parameters by

$$\tilde{\sigma} = \sigma(1 - F_n(u))^{\xi} \text{ and } \tilde{\mu} = u - \tilde{\sigma}((1 - F_n(u))^{-\xi} - 1)/\xi$$

to estimate a GPD distribution function which fits the severity distribution itself in the tail area above the threshold.

$$F_{X^\delta}(x) = P\{X \le x \mid X > \delta\} = \begin{cases} 0 & \text{if } x \le \delta, \\ \frac{F_X(x) - F_X(\delta)}{1 - F_X(\delta)} & \text{if } x > \delta. \end{cases}$$

For the prising of layers or estimation of high quantiles using a GPD model the crucial parameter is ξ: the higher value of ξ - the heavier the tail and higher quantile estimates. For a three-parameter GPD model $G_{\xi,\mu,\sigma}$ the pth quantile is $\mu + \sigma/\xi((1-p)^{-\xi} - 1)$.

There is a bias-variance tradeoff in the choice of the optimal threshold. Since modeling approach is based on a limit theorem which applies above high thresholds, if threshold is chosen too low it is possible to get biased estimates because the theorem does not apply. On the other hand, if a threshold is set too high then only few data points will be available and estimates will be prone to high standard errors.

Using the model with a threshold at 10 the .999th quantile is estimated to be 95. But for the threshold $u = 4$ the quantile estimate goes up to 147. This underlines the fact that estimates of high quantiles are extremely model dependent.

To get an indication of the insurance layer prices from the model price can be calculated as $P = E[Y_i | X_i > \delta]$. For a layer (r, R), P is given by

$$P = \int_r^R (x - r) f_{X^\delta}(x) dx + (R - r)(1 - F_{X^\delta}(R))$$

where $f_{X^\delta}(x) = dF_{X^\delta}(x)/dx$. By picking a high threshold $u(< r)$ it is possible to estimate $\widehat{F_{X^\delta}(x)}$ and therefore the density $f_{X^\delta}(x)$. It is shown that the price depends on the choice of the high threshold, which in its turn depends on the use of it. In other insurance datasets the effect of varying the threshold may be different. Every dataset is unique and the estimation and pricing process should not be fully automated.