pfsl29.tex

Lecture 29. 13.12.2012 (half-hour: Problems)

VIII. EXTREME-VALUE THEORY

Usually in Statistics it is the *typical* reading in a sample that is of interest. Sometimes, however, it is the *largest*, or the *smallest*. For example:
the speed of a convoy is the speed of its slowest ship;
the strength of a chain is the strength of its weakest link;
it is the strongest gust of wind that blows the roof off a building;
it is the biggest claims that pose the greatest threat to the solvency of an insurance company, etc.
The area of Statistics dealing with such things is called *extreme-value theory (EVT)*.

We shall focus on the *sample maximum*

$$M_n := \max\{X_1, \ldots, X_n\}$$

of a (usually large) sample of size n, with X_i iid (for the minimum, work with $-X_i$).

The distribution function of M_n is the nth power F^n, as

$$P(M_n \leq x) = P(X_i \leq x, \ldots, x_n \leq x) = P(X_1 \leq x) \ldots P(X_n \leq x) = F(x)^n,$$

by independence.

This is reminiscent of the theory of *stable* distributions (III.4), where we worked instead with *sums* $S_n := X_1 + \ldots + X_n$. There, we took CFs, where the CF of the sum is the nth power ϕ^n of the CF ϕ of the X_i. Because of this, the mathematics here is similar but simpler. We obtain, as with stability, a *parametric* discription of the possible limit laws of $(M_n - a_n)/b_n$, after suitable centring and scaling. As with stability, we work to within type; we now obtain a *one-parameter* family of limit laws, rather than a *two-parameter* family as with stability – the *extremal* or *extreme-value* laws. The result is due to Fisher and Tippett (1928) (L. H. C. TIPPETT (1902-1985)) and B. V. GNEDENKO (1912-1995) in 1943.

Theorem (Fisher-Tippett theorem), 1928. To within type, the extremal laws are exactly the following:

$$\Phi_\alpha, \quad (\alpha > 0); \qquad \Psi_\alpha, \quad (\alpha > 0); \qquad \Lambda,$$

where

$$\Phi_\alpha := 0 \quad (x \leq 0), \quad \exp\{-x^{-\alpha}\} \quad (x \geq 0);$$

$$\Psi_\alpha := \exp\{-(-x)^\sigma\} \quad (x \le 0), \qquad 1 \quad (x \ge 0);$$
$$\Lambda(x) := \exp\{-e^{-x}\}.$$

These are known since as the Fréchet (heavy-tailed, Φ_α), Gumbel (light-tailed, Λ) and Weibull (bounded tail, Ψ_α), after Maurice FRÉCHET (1878-1973), French mathematician, in 1937, Emil Julius GUMBEL (1891-1966), German statistician, in 1935 and 1958, and Waloddi WEIBULL (1887-1979), Swedish engineer, in 1939 and 1951.

Particularly for statistical purposes, it is often better to combine these three into one parametric family, the *generalized extreme value (GEV)* laws. These have one extremal parameter α, real, plus two type parameters μ (real, location) and σ (> 0, scale):

$$G(x) := \exp\Big(-\Big[1 + \alpha\Big(\frac{x-\mu}{\sigma}\Big)\Big]^{-1/\alpha}\Big).$$

Here $\alpha > 0$ corresponds to the Fréchet Φ_α, $\alpha = 0$ to the Gumbel Λ (using $(1 + x/n)^n \to e^x$ as $n \to \infty$) and $\alpha < 0$ to the Weibull Ψ_α, and we restrict to the support of G in each case – the set where [...] above > 0. See Coles [Col] for a monograph treatment of the statistics of EVT.

As always, the bigger n, the better. But in EVT, there are two conflicting dangers. We are studying the extremes, and most readings are *not* extreme – so we exclude most readings. Exclude too many, and we have too little data left; exclude too few, and we bias things by including non-extremes. We can balance these two dangers by the *peaks over thresholds (POT)* method ([Col] Ch. 4). Here we select a high *threshold* u. Then the conditional distribution of $X - u | X > u$ is approximately the *generalized Pareto distribution (GPD)*

$$H(x) := 1 - \Big(1 + \frac{\alpha x}{\bar\sigma}\Big)^{-1/\alpha}, \quad \bar\sigma := \sigma + \alpha(u - \mu) \quad (x > 0,\ (1 + \alpha x/\bar\sigma)) > 0.$$

The name comes from the *Pareto distribution* (heavy tail – power-law tail of income (Vilfredo PARETO (1848-1923) in 1909).

Which F lead to which extremal law or GEV (the set of such F is called the *domain of attraction* of the limit law) can be answered. It is analogous to the corresponding domain-of-attraction problem for stable laws (III.4). Both involve *regular variation*, an important topic that we must omit here.

Point-process methods (as in $Ppp(\lambda)$, VI.2) are important in POT; see [Col] Ch. 7. The theory also extends to many (including infinitely many) dimensions. For these, other extensions, and applications to insurance, finance etc., see e.g. [Col]. NHB