Limit Theorems for Extremes with Random Sample Size

D.S. Silvestrov & J.L. Teugels
Umeå University K.U. Leuven

1. Introduction

The main object of our contribution is the derivation of a number of limit theorems for extremal processes constructed from samples with a random sample size. Extremal processes of such type naturally appear in various applications related to models with sample variables associated with stochastic flows.

2. Examples

In its simplest setting, consider a sequence $\{X_1, X_2, \ldots\}$ of i.i.d. random variables with common distribution F. Let further $\{N(t); t \geq 0\}$ be a counting process. Then we are interested in the limiting behaviour of

$$X^*_{N(t)} := max\left(X_1, X_2, \ldots, X_{N(t)}\right)$$

when $t \uparrow \infty$.

(*i*) Consider the $X-$ values as the intensities of eruptions of an active volcano while $N(t)$ is the number of such eruptions over the period $[0\,,\,t]$. The quantity $X^*_{N(t)}$ then refers to the heaviest eruption over the reference period.
It is well-known in volcanology that the processes $\{X_1, X_2, \ldots\}$ and $\{N(t); t \geq 0\}$ are not independent since long waiting times in between eruptions are often followed by intensive explosions.

(*ii*) Consider a specific but homogeneous insurance portfolio. The $X-$values are the claim sizes while the counting process is determined by the number of claims. If t is the time-length of one year, then $X^*_{N(t)}$ is the largest claim in the portfolio over the bookyear.
In practice it is often assumed that the processes $\{X_1, X_2, \ldots\}$ and $\{N(t); t \geq 0\}$ are independent. This asumption is mostly made out of convenience. However, in a fire portfolio, long draughts make such an assumption doubtful.

(*iii*) In the case of a chronic disease, let the $X-$values correspond to the intensities of the subsequent attacks while $N(t)$ counts the number of them.

(*iv*) Let the $X-$values come from an ordinary renewal process where the counting process $N(t)$ gives information on the number of renewals up to time t. Since

$$\{N(t) = n\} \Leftrightarrow \{X_1 + X_2 + \ldots + X_n \leq t < X_1 + X_2 + \ldots + X_n + X_{n+1}\}$$

there is a very strong dependence between the two processes.

3. Situations already studied

In the literature, there are two cases that have been studied in great detail.

(1.) The first covers the model where the sample and the sample size index are independent.

(2.) The second model deals with the case where the counting process might depend on the sample but converges in probability in the sense that $N(t)/t \to c \in (0\,,\,\infty)$. Because of the latter convergence, there is independence between the extremal process and the corresponding non-random sample.

4. Triangular arrays

We formulate our results in terms of triangular arrays. Assume that for each $\epsilon > 0$, the sequence of random variables $\{\xi_{\epsilon,n}; n = 1, 2, \ldots\}$ are i.i.d. and that n_ϵ is a non-random, positive integer-valued quantity tending to ∞ when $\epsilon \to 0$. In the classical scheme one is interested in the limit behaviour of

$$\xi_\epsilon(t) := \max_{k \le t n_\epsilon} \xi_{\epsilon,k}\ .$$

The fundamental convergence condition is

$$(A) \qquad\qquad \{\xi_\epsilon(t); t > 0\} \Rightarrow \{\xi_0(t); t > 0\}$$

when $\epsilon \to 0$. Equivalently, one has the *domain of attraction condition*

$$n_\epsilon P\{\xi_{\epsilon,1} > u\} \to g(u)$$

for an appropriate $g(u)$.

The limiting process is an *extremal process* characterized by the transition kernel

$$P\{\xi_0(s+t) \le u | \xi_0(s) = v\} = I(v \le u)\ \mathrm{e}^{-t\,g(u)}\ .$$

We now change the non-random sample size into a random quantity. Let μ_ϵ be a positive random variable and put $\nu_\epsilon := \mu_\epsilon / n_\epsilon$. In the case of a random sample triangular scheme we look for the limiting behaviour when $\epsilon \to 0$ of

$$\zeta_\epsilon(t) := \max_{k \le t\,\mu_\epsilon} \xi_{\epsilon,k} = \xi_\epsilon\left(t\,\frac{\mu_\epsilon}{n_\epsilon}\right) = \xi_\epsilon(t\,\nu_\epsilon)\ .$$

A natural condition on the sample size process is

$$(B) \qquad\qquad \nu_\epsilon \Rightarrow \nu_0$$

where the right hand side is a.s. positive.

Composition of the two processes suggests a combination of the two conditions (A) and (B) into the form

$$(C) \qquad\qquad \{(\xi_\epsilon(t), \nu_\epsilon); t > 0\} \Rightarrow \{(\xi_0(t), \nu_0); t > 0\}\ .$$

BASIC RESULT: Under (C), as $\epsilon \to 0$,

$$\{\zeta_\epsilon(t); t \in S\} \Rightarrow \{\zeta_0(t) = \xi_0(t\nu_0); t \in S\}$$

where S is determined by the moments of the jumps.

5. Extensions

Under appropriate conditions, this basic result is then extendable to more general situations.

1. The convergence $\{\zeta_\epsilon(t); t > 0\} \xrightarrow{J} \{\zeta_0(t) = \xi_0(t\nu_0); t > 0\}$ results from the same condition (C).

2. The counting variable ν_ϵ can be replaced by a *stopping-index process* $\{\nu_\epsilon(t); t > 0\}$ so that we are looking at the *randomly stopped extremal process* $\zeta_\epsilon(t) := \xi_\epsilon(\nu_\epsilon(t))$.

3. The corresponding J-convergence of the above.

4. Multivariate versions.

Special attention is paid to the following classical models.

(i) The sample $\{\xi_{\epsilon,n}; n = 1, 2, \ldots\}$ and the random sample size index μ_ϵ are independent. Among the contributors we mention S. Berman (1962), D.I. Thomas (1972), J. Galambos (1975, 92, 94), B.V. & D.B. Gnedenko (1982), J. Beirlant & J.L. Teugels (1992) and V.Yu. Korolev (1995).

(ii) The random sample size index μ_ϵ satisfies $\mu_\epsilon/n_\epsilon \xrightarrow{P} \nu_0$. Authors that have studied this situation are S. Berman (1962), O.E. Barndorff-Nielsen (1964), J. Mogyoródi (1967), P.K. Sen (1972) and J. Galambos (1973, 78, 92).

(iii) The process $\{\xi_0(t), t > 0\}$ and the limiting stopping-index process $\{\nu_0(t), t > 0\}$ are independent.

(iv) Renewal type extremal processes indexed by $\nu_\epsilon(t) = \sup(s : \tau_\epsilon(s))$ where $\tau_\epsilon(t) = \sum_{k \le tn_\epsilon} \kappa_{\epsilon,k}$ and $\{(\kappa_{\epsilon,n}, \xi_{\epsilon,n}); n = 1, 2, \ldots\}$ are i.i.d. random vectors. Related results can be found in Serfozo (1980), Rootzén (1988) and Leadbetter & Rootzén (1988).

References

- D.S. Silvestrov & J.L. Teugels, Limit theorems for extremes with random sample size, *Advances of Applied Probability*, **30**, 1998, to appear.
  Available at www.matstat.umu.se

- D.S. Silvestrov & J.L. Teugels, Asymptotics of multivariate extremes with random sample size, *Studia Scientiarum Mathematicarum Hungarica*, **34**, 1998, to appear.
  Available at www.matstat.umu.se