Extreme Value for Discrete Random Variables Applied to Avalanche Counts

Pascal Alain Sielenou

IRSTEA / LSCE / CEN (METEO FRANCE)

PostDoc (MOPERA and ECANA projects)

Supervised by Nicolas Eckert and Philippe Naveau

PEPER Workshop at Aussois
15-19 december 2013

Outline

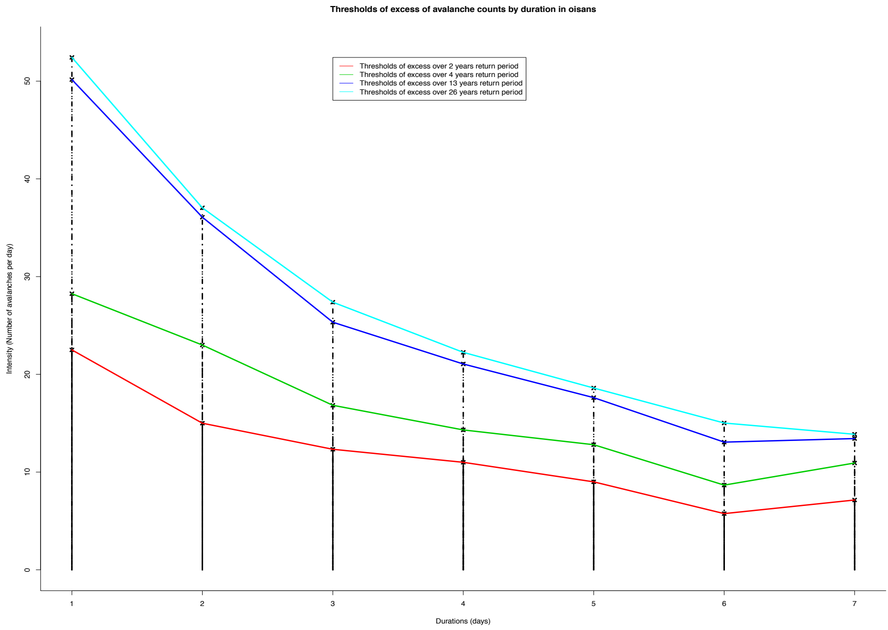
Introduction

Avalanche cycles are distinct period of time during which many natural avalanches occur at a given spatial scale: the mountain range, the district, etc. Loss of lives and economic damages from avalanche cycles have been recurrent in human history.

Thus, a more detailed description of the visible features of this phenomenon is needed in order to improve avalanche forecasting models so as to close ski resorts and evacuate the threatened mountain communities when a critical level is reached.

Intensity-Duration-Frequency (IDF) relationships (curves) can be used for this purpose.

IDF curve is a graph with duration plotted as abscissa, intensity as ordinate and a series of curves, one for each return period.


Thresholds of excess of avalanche counts by duration in oisans
Thresholds of excess over 2 years return period
Thresholds of excess over 4 years return period
Thresholds of excess over 13 years return period
Thresholds of excess over 26 years return period
Intensity (Number of avalanches per day)
0
10
20
30
40
50
1
2
3
4
5
6
7
Durations (days)

Data

Data

Daily observed avalanche events from the EPA (Enquête Permanente sur les Avalanches) over the period 1958-2009 in the 23 massifs of the French Alps.

Processed data

Series of exceedances above a threshold having two years return period for avalanche cycles of duration ranging from one to seven days.

Objective

Fit these seven series with suitable probabilistic distributions in order to construct the IDF curves for higher return periods.

Theorem (Feasible limit distribution for maxima)

Let $\{\xi_i\}$ *be a sequence of iid random variables with cumulative distribution function* $F(x)$*. The only family of distributions satisfying*

$$\lim_{n\to\infty} \left[F(a_n + b_n x)\right]^n = H(x) \quad \forall\ x,$$

where $H(x)$ *is not degenerate and* a_n *and* b_n *are constants, is*

$$H_\kappa(x;\lambda,\delta) = \exp\left\{-\left[1-\kappa\left(\frac{x-\lambda}{\delta}\right)\right]^{1/\kappa}\right\},\ 1-\kappa\left(\frac{x-\lambda}{\delta}\right) \geq 0,\ \kappa \neq 0,$$

where the support is $x \leq \lambda + \delta/\kappa$*, if* $\kappa > 0$*, or* $x \geq \lambda + \delta/\kappa$*, if* $\kappa < 0$*. The family of distributions for the case* $\kappa = 0$ *is obtained by taking the limit of (1) as* $\kappa \to 0$ *and getting*

$$H_0(x;\lambda,\delta) = \exp\left[-\exp\left(\frac{x-\lambda}{\delta}\right)\right], \quad -\infty < x < +\infty.$$

A few reviews on maxima of discrete distributions

For discrete random variables, the convergence of a linear normalization of maxima to non degenerate limit distributions is not always true (Anderson [1]).

Anderson [2] also provided a class of discrete distributions for which it is possible to obtain upper and lower bounds for the limiting distribution of maxima which are of Gumbel Type.

McCormick and Park [3] extended Anderson's results in the case of discrete stationary processes.

In this work, we provide sufficient conditions under which the maxima of discrete random variables can be (linearly) normalized to converge to non degenerate limit distributions.

Limit Theorem for Large Deviations of Sums

Definition (Condition (B_γ))

Let ξ be a random variable with mean $E\xi = 0$ and variance $\sigma^2 = E\xi^2$. We say that the random variables ξ satisfies condition (B_γ), if there exist $\gamma \geq 0$ and $K > 0$ such that

$$\left| E\xi^k \right| \leq (k!)^{1+\gamma} K^{k-2} \sigma^2, \quad k = 3, 4, \cdots \qquad (B_\gamma).$$

Theorem (Saulis et al. [4])

Let random variables ξ_j, $j = 1, 2, \cdots, n$, *satisfy condition* (B_γ). *Set*

$$S_n = \sum_{j=1}^{n} \xi_j, \qquad B_n^2 = ES_n^2, \qquad Z_n = \frac{S_n}{B_n}.$$

Then

$$\lim_{n\to\infty} \frac{1 - F_{Z_n}(x)}{1 - \Phi(x)} = 1,$$

for $x \geq 0$, $x = x_n = o\left(\Delta_{\gamma,n}^{1/3}\right)$ *as* $\Delta_{\gamma,n}^{1/3} \to \infty$, *where*

$$\Delta_{\gamma,n} = c_\gamma \Delta_n^{\frac{1}{1+2\gamma}}, \quad c_\gamma = \frac{1}{6}\left(\frac{\sqrt{2}}{6}\right)^{\frac{1}{1+2\gamma}},$$

$$\Delta_n = \frac{B_n}{K_n}, \quad K_n = 2\max\left\{K, \max_{1\leq j\leq n} \sigma_j\right\}.$$

Main Results

Theorem

For each positive integer n, let $\xi_{n,i}$, $i = 1, 2, \cdots, n$ denote iid integer valued random variables. Suppose that for each i,

$$\xi_{n,i} = \sum_{j=1}^{n} \xi_{n,i,j}$$

for some iid integer valued random variables $\xi_{n,i,j}$, $j = 1, 2, \cdots, n$ satisfying condition (B_γ) and whose distribution parameters satisfy

$$(\ln n)^{1/2} = o\left(\Delta_{\gamma,n}^{1/3}\right),$$

where $\Delta_{\gamma,n}$ is defined in the previous theorem.

Main Results

Continuation of Theorem

Then for every function $z(x) > 0$, the sequence $M_n = \max\{\xi_{n,1}, \xi_{n,2}, \cdots, \xi_{n,n}\}$ satisfies the equality

$$\lim_{n\to\infty} P\{M_n \le u_n(x)\} = e^{-z(x)},$$

for all $x \in Df(z)$ with $u_n(x) = \alpha_n\, y(x) + \beta_n$, where $y(x) = \ln\delta - \ln z(x)$ for any $\delta > 0$, and $\beta_n = 1/\alpha_n$ is given by

$$\beta_n = \left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2} - \frac{\ln\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)}{2\left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2}}.$$

Corollary (Poisson)

Consider independent random variables

$$\xi_{n,i} \rightsquigarrow \mathcal{P}(\lambda_n = cn), \quad i = 1, 2, \cdots, n,$$

where $c > 0$ *is a constant. Then for every function* $z(x) > 0$, *the sequence* $M_n = \max\{\xi_{n,1}, \xi_{n,2}, \cdots, \xi_{n,n}\}$ *satisfies the equality*

$$\lim_{n\to\infty} P\{M_n \le u_n(x)\} = e^{-z(x)},$$

for all $x \in Df(z)$ *with* $u_n(x) = \alpha_n\, y(x) + \beta_n$, *where* $y(x) = \ln\delta - \ln z(x)$ *for any* $\delta > 0$, *and* $\beta_n = 1/\alpha_n$ *is given by*

$$\beta_n = \left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2} - \frac{\ln\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)}{2\left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2}}.$$

Corollary (Negative Binomial)

Consider independent random variables

$$\xi_{n,i} \rightsquigarrow \mathcal{NB}\left(r_n = cn, p_n = \beta + \frac{1}{n+1}\right), \quad i = 1, 2, \cdots, n,$$

where $c > 0$ *and* $\beta \in (0, \frac{1}{2})$ *are constants. Then for every function* $z(x) > 0$, *the sequence* $M_n = \max\{\xi_{n,1}, \xi_{n,2}, \cdots, \xi_{n,n}\}$ *satisfies the equality*

$$\lim_{n\to\infty} P\{M_n \leq u_n(x)\} = e^{-z(x)},$$

for all $x \in Df(z)$ *with* $u_n(x) = \alpha_n\, y(x) + \beta_n$, *where* $y(x) = \ln\delta - \ln z(x)$ *for any* $\delta > 0$, *and* $\beta_n = 1/\alpha_n$ *is given by*

$$\beta_n = \left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2} - \frac{\ln\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)}{2\left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2}}.$$

Main Results

Definition

Let $\alpha \in (0, 1)$. A sequence $\{\xi_t\}$ is said to be a first order non negative integer valued process, denoted by $INAR_\alpha(1)$, if

$$\xi_t = \alpha \circ \xi_{t-1} + \zeta_t,$$

where

- $\alpha \circ \xi_{t-1} \rightsquigarrow Binomial(\xi_{t-1}, \alpha)$ and
- $\{\zeta_t\}$ is a sequence of iid integer valued random variables.

Theorem ($INAR_\alpha(1)$ with Poisson innovations)

If $\alpha \in (0, 1)$ *and* $\xi_{n,t} = \alpha \circ \xi_{n,t-1} + \zeta_{n,t}$ *is an* $INAR_\alpha(1)$ *process where*

$$\zeta_{n,t} \rightsquigarrow \mathcal{P}(\lambda_n = cn), \quad t = 1, 2, \cdots, n$$

with $c > 0$. *Then for every function* $z(x) > 0$, *the sequence* $M_n = \max\{\xi_{n,1}, \xi_{n,2}, \cdots, \xi_{n,n}\}$ *satisfies the equality*

$$\lim_{n\to\infty} P\{M_n \leq u_n(x)\} = e^{-z(x)},$$

for all $x \in Df(z)$ *with* $u_n(x) = \alpha_n y(x) + \beta_n$, *where* $y(x) = \ln\delta - \ln z(x)$ *for any* $\delta > 0$, *and* $\beta_n = 1/\alpha_n$ *is given by*

$$\beta_n = \left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2} - \frac{\ln\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)}{2\left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2}}.$$

Theorem ($INAR_\alpha(1)$ with Negative Binomial innovations)

If $\alpha \in (0,\,1)$ *and* $\xi_{n,t} = \alpha \circ \xi_{n,t-1} + \zeta_{n,t}$ *is an* $INAR_\alpha(1)$ *process where*

$$\zeta_{n,t} \rightsquigarrow \mathcal{NB}\left(r_n = cn, p_n = \beta + \frac{1}{n+1}\right), \quad t = 1, 2, \cdots, n$$

with $c > 0$ *and* $\beta \in (0,\,\frac{1}{2})$. *Then for every function* $z(x) > 0$, *the sequence* $M_n = \max\left\{\xi_{n,1}, \xi_{n,2}, \cdots, \xi_{n,n}\right\}$ *satisfies the equality*

$$\lim_{n\to\infty} P\left\{M_n \le u_n(x)\right\} = e^{-z(x)},$$

for all $x \in Df(z)$ *with* $u_n(x) = \alpha_n\, y(x) + \beta_n$, *where* $y(x) = \ln\delta - \ln z(x)$ *for any* $\delta > 0$, *and* $\beta_n = 1/\alpha_n$ *is given by*

$$\beta_n = \left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2} - \frac{\ln\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)}{2\left[\ln\left(\frac{n^2}{2\pi\,\delta^2}\right)\right]^{1/2}}.$$

Main Results

Corollary

Under the hypothesis of the previous theorems, if $\delta = 1$ *and* $z(x)$ *is either*

- $z(x) = e^{-x}$ *for all* $x \in \mathbb{R}$ *or*
- $z(x) = (1 - \kappa x)^{1/\kappa}$, $\kappa \neq 0$, *for small values of* x,

then, the maxima M_n *can be linearly normalized to converge to the generalized extreme value distributions.*

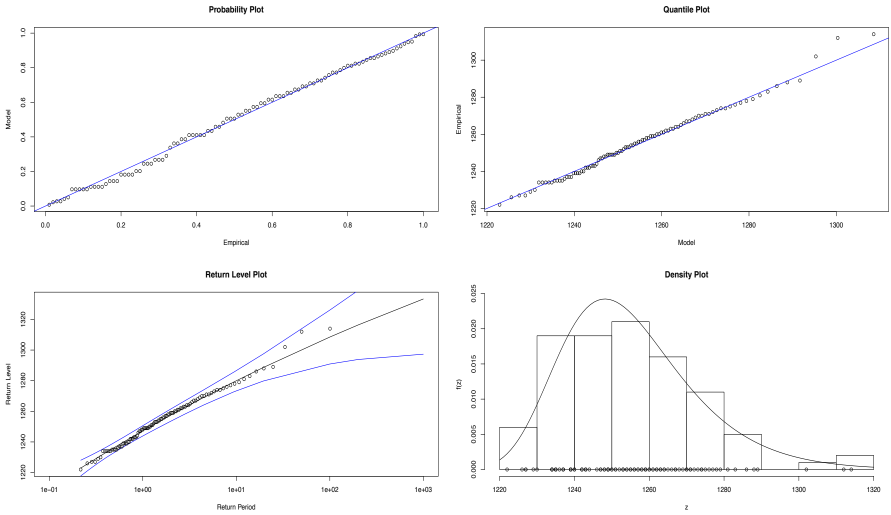


Figure: Diagnostic plots for GEV models to 12-length block maxima of a n-length sample of $Poisson(\lambda = n)$, $n = 1200$.

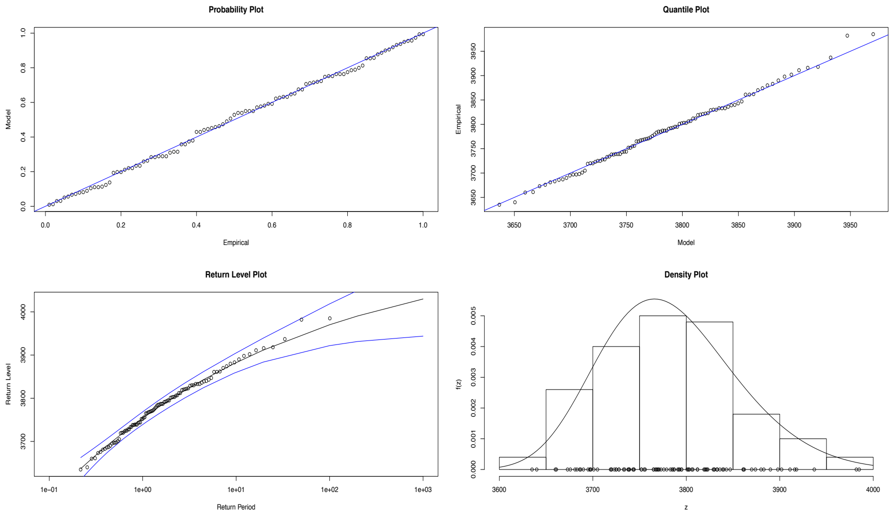


Figure: Diagnostic plots for GEV models to 12-length block maxima of a n-length sample of $NegativeBinomial\left(r = n, p = \beta + \frac{1}{n+1}\right)$, $\beta = 0.25$ and $n = 1200$.

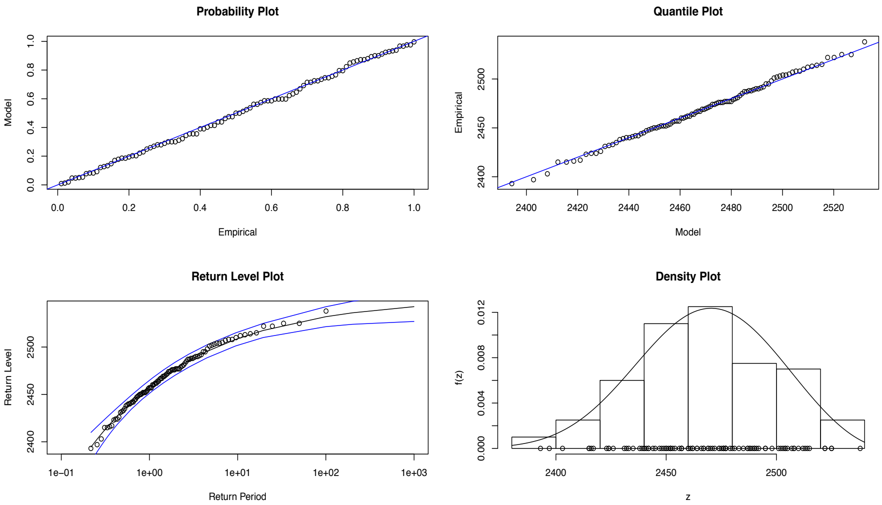


Figure: Diagnostic plots for GEV models to 12-length block maxima of n-length sample of $INAR_\alpha(1)$, $\alpha = 0.5$, with $Poisson(\lambda = n)$ innovations, $n = 1200$.

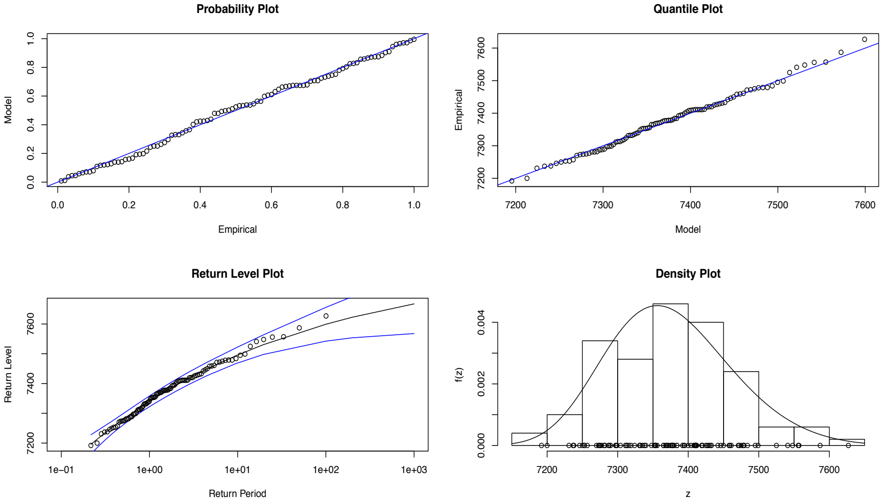


Figure: Diagnostic plots for GEV models to 12-length block maxima of a continuized n-length sample of $INAR_\alpha(1)$, $\alpha = 0.5$, with $NegativeBinomial\left(r = n, p = \beta + \frac{1}{n+1}\right)$ innovations, where $\beta = 0.25$, $n = 1200$.

Conclusions

In this work, we have given some conditions under which the maxima of discrete stationary processes can be normalized and converge to the classical generalized extreme value distributions.

These results have been illustrated with some simulated data.

Perspectives

Find among the 23 massifs of the French Alps the one whose avalanche data can modelized by discrete distributions whose parameters satisfy the hypothesis allowing the modelization of the extremes values with GEV or GP distributions.

Fit the extreme value of avalanche data of the retained massifs and construct the corresponding IDF curves. Then, use suitable translation vectors to deduce the IDF curve of avalanches for the other massifs.

References

- C. W. Anderson. Local limit theorems for the maxima of discrete random variables. *Math. Proc. Camb. Philos. Soc.* **88** 161-165, (1980).
- C. W. Anderson. Extreme value theory for a class of discrete distributions, with applications to some stochastic processes. *J. Appl. Probab.* **7**, 99-113, (1970).
- W. McCormick and Y. Park. Asymptotic analysis of extremes from autoregressive negative binomial processes. *J. Appl. Probab.* **29**, 904-920, (1992).
- L. Saulis and V. Statulevicius. *Limit Theorems for Large Deviations*. Kluwer Academic Publisher, Dordrecht, Boston, London, 1991.