Randomly Stopped k-th Order Statistics

Sandra Mendonça, Dinis Pestana and M. Ivette Gomes

Abstract Randomly stopped order statistics when the stopping rule is generated by a basic count distribution are investigated. Unified expressions in terms of the subordinator are presented, extending results from geometrically thinned sequences. Using the results on limit stable distributions for max-geometric laws, and Smirnov's techniques to deal with limit laws of extreme order statistics, some results on stability of Panjer subordinated randomly stopped order statistics are discussed.

1 Introduction

Risk analysis and extreme value theory walk naturally hand-in-hand. In this work inspired by Rachev and Resnick theory on stable limits of randomly stopped maxima with geometric subordinator — also called geo-max stability, (see [9]), we take this theory a step further, in a parallel way to classical extreme value theory.

After briefly stating, in Section 2, Rachev and Resnick results on limiting stable laws for geometrically thinned sequences of independent and identically distributed random variables, inspired on Smirnov [11] work on what he calls asymptotic results for variational series, we establish some basic results on other geometrically

Sandra Mendonça
Universidade da Madeira, CCEE, 9000-390 Funchal, Portugal
Universidade de Lisboa, CEAUL, Portugal
e-mail: smfm@uma.pt

Dinis Pestana
Universidade de Lisboa, CEAUL and DEIO–FCUL, Campo Grande 1749-016 Lisboa, Portugal
Instituto de Investigação Científica Bento da Rocha Cabral, Portugal
e-mail: dinis.pestana@fc.ul.pt

M. Ivette Gomes
Universidade de Lisboa, CEAUL and DEIO–FCUL, Campo Grande, 1749-016 Lisboa, Portugal
Instituto de Investigação Científica Bento da Rocha Cabral, Portugal
e-mail: ivette.gomes@fc.ul.pt

thinned order statistics. We further refer to e-bay auctions to claim that there are situations for which ordered statistics of geometrically thinned sequences count. In Section 3 we briefly discuss randomly stopped order statistics, and in Section 4 we describe the basic count distributions, the Panjer distributions ([7]) — Poisson, binomial and negative binomial, that in a sense are the yardstick with unitary dispersion coefficient, underdispersed and over dispersed count variables —, and Sundt's [13] extension (logarithmic and extended negative binomial), in order to present, in Section 5, unified expressions for the distribution of randomly stopped order statistics when the stopping rule is generated by a Panjer class subordinator. Finally, in Section 6, we use the ideas of Smirnov to investigate stability results for order statistics of thinned sequences.

2 Maxima and other Order Statistics of Geometrically Thinned Sequences

Poisson thinned sequences and geometrically thinned sequences have specially nice properties. In the sequel, we shall present the theory of Panjer recursion, and it will be obvious that Poisson thinning and geometric thinning are the simplest ways of Panjer random stopping because they use the $(0,p)$- and the $(p,0)$-Panjer subordinators. In this section we present some basic concepts that can guide us on extensions for the general case of randomly stopped order statistics with (a,b)-Panjer subordinators.

2.1 Stable limit laws for maxima of geometrically thinned sequences

For $p\in(0,1)$, define $q:=1-p$ and let $N(p)$ be a geometrically distributed random variable (rv), having probability of success p:

$$P[N(p)=k]=pq^{k-1},\ k=1,2,\ldots.$$

Let $\{X_i,\ i\in\mathbb{N}^+\}$ be a sequence of independent and identically distributed (iid) random variables, replicas of X, an absolute continuous rv, all independent from $N(p)$, with common distribution function F_X. Suppose that X belongs to the max-domain of the attraction of one of the possible max-stable types,

$$G_1(x)=\exp\left(-x^{-\alpha}\right)I_{(0,+\infty)}(x),$$
$$G_2(x)=\exp\left[-(-x)^{\alpha}\right]I_{(-\infty,0)}(x)+I_{[0,+\infty)}(x)\,(\alpha>0)\text{ and}$$
$$G_3(x)=\exp\left[-\exp(-x)\right],$$

where $\alpha > 0$, generally denoted $G(x)$. The limit laws $\mathscr{G}$, when $p \to 0$, of the maximum of the vector $\left(X_1, ..., X_{N(p)}\right)$, properly nomalized by functions of p, were described by Rachev and Resnick (cf. [9]). They are related to the above extreme value distributions through the expression

$$\mathscr{G}(x) = \frac{1}{1 - \ln G(x)}.$$

Therefore, the geo-max stable types are

$$\mathscr{G}_1(x) = \frac{1}{1 + x^{-\alpha}} I_{(0,+\infty)}(x) \text{ (loglogistic distribution),}$$
$$\mathscr{G}_2(x) = \frac{1}{1 + (-x)^{\alpha}} I_{(-\infty,0)}(x) + I_{[0,+\infty)}(x) \text{ (backward loglogistic distribution) and}$$
$$\mathscr{G}_3(x) = \frac{1}{1 + \exp(-x)} \text{ (logistic distribution).}$$

2.2 Order statistics of geometrically thinned sequences

Definition 1. Let the subordinator count variable $N(p)$ be independent of the independent replicas $\{X_1, X_2, ...\}$ of some rv X. We will designate $X_{N(p):N(p)}$ as the geometric maximum and $X_{1:N(p)}$ as the geometric minimum, both associated to the sequence $\{X_i,\ i \in \mathbb{N}^+\}$. More generally, when $N(p) \geq k$, we define the k-geometric order statistic as being the rv $X_{k:N(p)}$.

The distribution of the k-th geometrically randomly stopped order statistic is

$$\begin{aligned}
F_{X_{k:N(p)}|N(p)\geq k}(x) &= \frac{P\left[X_{k:N(p)} \leq x, N(p) \geq k\right]}{P[N(p) \geq k]} \\
&= \frac{1}{P[N(p) \geq k]} \sum_{j=k}^{+\infty} P\left[X_{k:N(p)} \leq x, N(p) \geq k | N(p) = j\right] P[N(p) = j] \\
&= \frac{1}{P[N(p) \geq k]} \sum_{j=k}^{+\infty} F_{X_{k:j}}(x) [N(p) = j]
\end{aligned}$$

and, although redundant, as it will be useful,

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x) = \frac{1}{P[N(p) \geq k]} \sum_{j=k}^{+\infty} F_{X_{j-k+1:j}}(x) P[N(p) = j].$$

Order statistics of geometrically thinned sequences inherit the relation $X_{k:n} \stackrel{d}{=} -\left[(-X)_{n-k+1:n}\right]$ of order statistics of independent random variables. In fact, using the well known expressions $1 - F_{X_{k:n}}(x) = \sum_{i=0}^{k-1} \binom{n}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i}$

(we can also write $F_{X_{n-k+1:n}}(x) = \sum_{i=0}^{k-1} \binom{n}{i} [1-F_X(x)]^i [F_X(x)]^{n-i}$), we obtain

$$\begin{aligned} F_{X_{k:N(p)}|N(p)\geq k}(x) &= \sum_{j=k}^{+\infty} \left\{ 1 - \sum_{i=0}^{k-1} \binom{j}{i} [F_X(x)]^i [1-F_X(x)]^{j-i} \right\} \frac{P[N(p)=j]}{P[N(p)\geq k]} \\ &= 1 - \sum_{j=k}^{+\infty} \sum_{i=0}^{k-1} \binom{j}{i} [F_X(x)]^i [1-F_X(x)]^{j-i} \frac{P[N(p)=j]}{P[N(p)\geq k]} \end{aligned}$$

and

$$\begin{aligned} F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x) &= \sum_{j=k}^{+\infty} \sum_{i=0}^{k-1} \binom{j}{i} [1-F_X(x)]^i [F_X(x)]^{j-i} \frac{P[N(p)=j]}{P[N(p)\geq k]} \qquad (1) \\ &= \sum_{j=k}^{+\infty} \sum_{i=0}^{k-1} \binom{j}{i} [F_{-X}(-x)]^i [1-F_{-X}(-x)]^{j-i} \frac{P[N(p)=j]}{P[N(p)\geq k]} \\ &= 1 - F_{(-X)_{k:N(p)|N(p)\geq k}}(-x) = F_{-\left[(-X)_{k:N(p)|N(p)\geq k}\right]}(x), \end{aligned}$$

which lead us to the equality

$$X_{k:N(p)}|N(p)\geq k \stackrel{d}{=} -\left[(-X)_{N(p)-k+1:N(p)}|N(p)\geq k\right].$$

In fact, from the proof it is obvious that the result is valid for the randomly stopped order statistics, whatever the count random subordinator N:

Proposition 1. *Given N a discrete rv with support $\mathbb{N}^+$ and $\{X_i,\ i\in\mathbb{N}^+\}$ a sequence of iid random variables, independent from N and equal in distribution to X, an absolute continuous rv, we have*

$$X_{k:N}|N\geq k \stackrel{d}{=} -\left[(-X)_{N-k+1:N}|N\geq k\right].$$

2.3 Who cares about geometrically thinned second maxima?

In auctions such as the eBay auctions the final price is a fixed increment of the second maximum biding. For interesting, fairly rare and expensive items many bidders wait until the very end before making their bid (which is called "sniping"). Since they wait to the very last moment, they do not know the others bids, bidding independently of each other. On the other hand, internet speed and communication jams preclude some bids to arrive before the auction is closed, and we shall accept that this thinning is geometric, and that for the same valuable item bids are iid.

Some seller(s) auction(s) similar items at different times, and sometimes under different identities, for instance we have seen dozens of times auctions of a rocking mother and child or of the king and queen, by Henry Moore, without a certificate of authenticity (COA), and numbered using some cypher from 1 to 9 (Moore only casted less than 10 of those miniatures). It is of course expected that the seller(s) will continue to put similar items on sale on the future, and that (s)he wants to maximise the selling price.

The seller can force bids to be greater than a "reserve price". Hence the problem is to determine the distribution of the geometrically thinned maximum, given that several previous thinned second maxima have been recorded. The seller can then choose an appropriate threshold as reserve price, depending obviously on his greed and on his need to make money, conflicting issues in determining the probability of selling he wants to attain.

As a curiosity: the price of such items is quite reasonable, since similar items with COA are auctioned at prices that range from 50 to 80 times the auction price of those fakes. So, as COA can also be faked, it is a matter of personal satisfaction if someone decides to acquire a nice replica at a sensible prices. Hence, someone interested in bidding to the next time one of those king and queen appears for auction, can also make and educated guess of his last moment bidding using the information of past second maxima bided.

3 Randomly Stopped Order Statistics

As before, let $p \in (0,1)$, $N(p)$ be a discrete rv having as support the set of the positive natural numbers, $\mathbb{N}^+$, $\{X_i,\ i \in \mathbb{N}^+\}$ be a sequence of iid random variables, replicas of an absolute continuous rv X, all independent from $N(p)$.

Definition 2. Using the variables above presented, when $N(p) \geq 1$, we will designate $X_{N(p):N(p)}$ as the $N(p)$ randomly stopped maximum and $X_{1:N(p)}$ as the $N(p)$ randomly stopped minimum, both associated to the sequence $\{X_i,\ i \in \mathbb{N}^+\}$. More generally, when $N(p) \geq k$, we define the k-th $N(p)$ ascending order statistic as being the rv $X_{k:N(p)}$.

Using the equality (1) it is not difficult to prove that

Theorem 1.

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x) = 1 - \sum_{i=k}^{+\infty} [1 - F_X(x)]^i \sum_{j=0}^{+\infty} \frac{P[N(p) = j+i]}{P[N(p) \geq k]} \times \binom{j+i}{i} [F_X(x)]^j . \tag{2}$$

In fact,

$$\sum_{j=k}^{+\infty} \frac{P[N(p)=j]}{P[N(p)\geq k]} \sum_{i=0}^{k-1} \binom{j}{i} [1-F_X(x)]^i [F_X(x)]^{j-i}$$
$$= \sum_{j=k}^{+\infty} \frac{P[N(p)=j]}{P[N(p)\geq k]} \left\{1 - \sum_{i=k}^{j} \binom{j}{i} [1-F_X(x)]^i [F_X(x)]^{j-i}\right\}$$
$$= 1 - \sum_{i=k}^{+\infty} \sum_{j=0}^{+\infty} \frac{P[N(p)=j+i]}{P[N(p)\geq k]} \binom{j+i}{i} [1-F_X(x)]^i [F_X(x)]^j .$$

4 Count Distributions

A discrete distribution whose support is $\mathbb{N}$ or $\mathbb{N}^+$, or an initial subsection of either $\mathbb{N}$ or $\mathbb{N}^+$, is a count distribution. Count distribution are the appropriate ones to use as subordinators of randomly stopped order statistics.

Three of the most used discrete models have an interesting property: a simple recursive relation for the successive probability atoms, several times rediscovered in different contexts (for instance, Katz [5] used it to organise a family of discrete models in the same spirit of the Pearson family), cf. [8]. An important breakthrough has been Panjer's [7] idea of using the above mentioned recursive expression to iteratively compute or approximate densities of the aggregate claim in risk theory, cf. [10].

Definition 1 *We say that a discrete rv $N_{a,b}$ belongs to Panjer family of order 0 if its mass distribution function satisfies the relation*

$$p_{n+1} = P\left[N_{a,b} = n+1\right] = \left(a + \frac{b}{n+1}\right) p_n,\ n \geq 0. \tag{3}$$

This family has only three non-degenerate members, cf. [7]:

Distribution	p_n	a	b
Poisson (p) $(p>0)$	$e^{-p}\frac{p^n}{n!},\ n\in\mathbf{N}$	0	p
Binomial (m,p) $(m\in\mathbf{N}_1,\ p\in(0,1))$	$\begin{cases}\binom{m}{n}p^n(1-p)^{m-n}, & n=0,\ldots,m\\ 0, & n=m+1,\ldots\end{cases}$	$-\frac{p}{1-p}$	$\frac{m+1}{1-p}\,p$
NegativeBinomial (r,p) $(r\in\mathbf{N}_1,\ p\in(0,1))$	$\binom{n+r-1}{n}p^n(1-p)^r,\ n=0,1,2,\ldots$	p	$(r-1)\,p.$

Hence the 0-Panjer distributions are exactly the discrete Morris natural exponential families whose variance is at most a quadratic function of the mean value (cf. [6]).

The NegativeBinomial$(1,p)$, usually referred to as Geometric(p) distribution, and the Poisson(p) are the most used subordinators of randomly stopped variables.

In fact, as their Panjer set of coefficients are respectively $(0,p)$ and $(p,0)$, expressions are simpler. Pestana and Velosa [8] analysis of the arithmetic behaviour of extended Panjer recursion exhibit how cumbersome expressions are for general Panjer coefficients (a,b) are, in contrast to the elegant expressions associated to the coefficient pairs $(0,p)$ and $(p,0)$ of the Poisson and of the geometric subordinator cases.

Remark 1. A non-recursive expression for the $0-$Panjer mass distributions would be:

$$\begin{aligned}
p_{n+1} &= P\left[N_{a,b}=n+1\right]=\left(a+\frac{b}{n+1}\right)p_n \\
&= \left(a+\frac{b}{n+1}\right)\left(a+\frac{b}{n}\right)\left(a+\frac{b}{n-1}\right)\ldots\left(a+\frac{b}{n-k}\right)\ldots(a+b)\,p_0 \\
&= \frac{(n+1)\,a+b}{n+1}\frac{na+b}{n}\frac{(n-1)\,a+b}{n-1}\ldots\frac{(n-k)\,a+b}{n-k}\ldots\frac{[n-(n-1)]\,a+b}{n-(n-1)}p_0 \\
&= \frac{p_0}{(n+1)!}\prod_{k=1}^{n+1}(ka+b)\,,\ n\geq 0.
\end{aligned}$$

In short,

$$p_n=\frac{p_0}{n!}\prod_{k=1}^{n}(ka+b)\,,n\geq 1.$$

For $a=0$ (in the Poisson case),

$$p_n=\frac{p_0}{n!}b^n,n\geq 1.$$

When in expression (3) we use instead of the condition $n\geq 0$ the condition $n\geq L$ (and take $p_n=0$, for $n<k$) we obtain an extension of the Panjer family that we identify by the name Panjer distribution of order L (cf. [12]).

For $L=1$, two new non-degenerate Panjer distributions do exist (cf. Sundt and Jewell [13]), the Logarithmic(p), defined for $n\geq 1$ by

$$p_n=-\frac{1}{\ln(1-p)}\frac{p^n}{n},p\in(0,1)$$

(observe that its Panjer coefficients are $(p,-p)$), and Engen's [1] Extended Negative Binomial (ENB) distribution, very cumbersome but useful in some ecological and population dynamics models, whose mass function is given by (cf. also [14]), for $\alpha\in(-1,0)$, $p\in(0,1]$ and $n\geq 1$,

$$p_n=\frac{\alpha\Gamma\,(n+\alpha)}{n!\,\Gamma\,(1+\alpha)}\frac{p^n\,(1-p)^{\alpha}}{1-(1-p)^{\alpha}}.$$

Observe that ENB distribution has Panjer coefficients $(p,-p\,(1-\alpha))$ and that making $\alpha\to 0$ we obtain the Logarithmic(p) distribution.

Although those two new Panjer families (Logarithmic and ENB) are not left truncated Panjer distributions of order 0, for $L \geq 2$ Hess *et al.* [4] have established that any L-Panjer distribution is the left endpoint truncation of an $(L-1)$-Panjer distribution. For that reason, they called the Binomial, the Poisson, the Negative Binomial, the Logarithmic and the Extended Negative Binomial distributions the basic count models.

5 Stopped Order Statistics with Panjer Subordinators

Randomly stopped order statistics with Panjer subordinators do have intrinsic interest, although results are cumbersome but for the Poisson, the Geometric and the Logarithmic cases. The expressions of their distribution functions have however some similarities:

5.1 Stopped order statistics with order 0 Panjer subordinator

Theorem 2. *If $N(p)$ is a member of Panjer family of order 0 then*

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x) = 1 - \frac{P[N(p^*) \geq k]}{P[N(p) \geq k]},$$

with $p^* = p[1-F_X(x)]$, if $N(p)$ is Poisson or binomial distributed, and $p^* = \frac{p[1-F_X(x)]}{1-pF_X(x)}$, if $N(p)$ is negative binomial distributed.

Proof. We will show the result separately for each family distribution of the Panjer family of order 0.

1. Consider $N(p)$ a rv with Poisson distribution with mean value $p > 0$, i.e.,

$$P[N(p) = i] = \exp(-p)\frac{p^i}{i!}, \text{ for } i = 0, 1, \ldots$$

From expression (2) we know that:

$$\begin{aligned}
&F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x) \\
&= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{+\infty} \frac{[1-F_X(x)]^i}{i!} \sum_{j=0}^{+\infty} \exp(-p)\frac{p^{j+i}}{(j+i)!}\frac{(j+i)!}{j!}[F_X(x)]^j \\
&= 1 - \frac{\exp(-p)}{P[N(p) \geq k]} \sum_{i=k}^{+\infty} \frac{[1-F_X(x)]^i p^i}{i!} \sum_{j=0}^{+\infty} \frac{1}{j!}[pF_X(x)]^j
\end{aligned}$$

$$= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{+\infty} \exp\{-p[1-F_X(x)]\} \frac{\{p[1-F_X(x)]\}^i}{i!}$$
$$= 1 - \frac{P[N(p[1-F_X(x)]) \geq k]}{P[N(p) \geq k]}.$$

2. Consider $N(p) \equiv N(m,p)$ a rv with binomial distribution with parameters $p \in (0,1)$ and $m \in \{1,2,...\}$, i.e.,

$$P[N(p) = k] = \binom{m}{k} p^k (1-p)^{m-k} \mathbf{I}_{\{0,1,...,m\}}(k).$$

Let $1 \leq k \leq m$. From expression (2) we know that:

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{m} \frac{[1-F_X(x)]^i}{i!} \sum_{j=0}^{m-i} \binom{m}{j+i} p^{j+i} (1-p)^{m-(j+i)} \frac{(j+i)!}{j!} [F_X(x)]^j$$
$$= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{m} \frac{\{p[1-F_X(x)]\}^i}{i!} \frac{m!}{(m-i)!} (1-p)^{m-i} \sum_{j=0}^{m-i} \binom{m-i}{j} \left[\frac{pF_X(x)}{1-p}\right]^j$$
$$= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{m} \{p[1-F_X(x)]\}^i \frac{m!}{(m-i)!i!} (1-p)^{m-i} \left(1 + \frac{pF_X(x)}{1-p}\right)^{m-i}.$$

Hence,

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{m} () \{p[1-F_X(x)]\}^i (1-p[1-F_X(x)])^{m-i}$$
$$= 1 - \frac{P[N(p[1-F_X(x)]) \geq k]}{P[N(p) \geq k]}.$$

3. Consider $N(p) \equiv N(r,p)$ a rv with negative binomial distribution with parameters $p \in (0,1)$ and $r \in \{1,2,...\}$, i.e.,

$$P[N(p) = k] = \binom{k+r-1}{k} p^k (1-p)^r \mathbf{I}_{\{0,1,2,...\}}(k)$$

Let $1 \leq k$. From expression (2) we know that:

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{+\infty} \frac{[1-F_X(x)]^i}{i!} \sum_{j=0}^{+\infty} \binom{j+i+r-1}{j+i} p^{j+i} (1-p)^r \frac{(j+i)!}{j!} [F_X(x)]^j$$
$$= 1 - \frac{1}{P[N(p) \geq k]} \sum_{i=k}^{+\infty} \frac{(1-p)^r \{p[1-F_X(x)]\}^i}{i!(r-1)!} \sum_{j=0}^{+\infty} \frac{(j+i+r-1)!}{j!} [pF_X(x)]^j.$$

Noting that, for $|y| < 1$,

$$\frac{(i-1)!}{(1-y)^i} = \sum_{j=0}^{+\infty} \frac{(j+i-1)!}{j!} y^j,$$

we obtain

$$\begin{aligned}
&F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x) \\
&= 1 - \frac{1}{P[N(p)\geq k]} \sum_{i=k}^{+\infty} \frac{(1-p)^r \{p[1-F_X(x)]\}^i}{i!\,(r-1)!} \frac{(i+r-1)!}{[1-pF_X(x)]^{i+r}} \\
&= 1 - \frac{1}{P[N(p)\geq k]} \sum_{i=k}^{+\infty} \binom{i+r-1}{i} \left\{\frac{p[1-F_X(x)]}{1-pF_X(x)}\right\}^i \left(\frac{1-p}{1-pF_X(x)}\right)^r \\
&= 1 - \frac{1}{P[N(p)\geq k]} \sum_{i=k}^{+\infty} \binom{i+r-1}{i} \left\{\frac{p[1-F_X(x)]}{1-pF_X(x)}\right\}^i \left(1-\frac{p[1-F_X(x)]}{1-pF_X(x)}\right)^r \\
&= 1 - \frac{P[N(p^*)\geq k]}{P[N(p)\geq k]},
\end{aligned}$$

with

$$p^* = \frac{p[1-F_X(x)]}{1-pF_X(x)}.$$

5.2 *Stopped order statistics with genuinely order 1 Panjer subordinator*

The Panjer families of different orders form an increasing chain. Besides the three nondegenerate distributions that belong to the Panjer family of order 0, the Panjer family of order 1, has as nondegenerate members the Logarithmic(p) distribution defined for $n \geq 1$ by

$$p_n = -\frac{1}{\ln(1-p)} \frac{p^n}{n} \;\; (p \in (0,1)),$$

(observe that its Panjer coefficients are $(p,-p)$) and the ENB distribution (cf. [14]), for $\alpha \in (-1,0)$, $p \in (0,1]$:

$$\begin{aligned}
p_n &= \frac{-\alpha\Gamma(n+\alpha)}{n!\,\Gamma(1+\alpha)} \frac{p^n}{1-(1-p)^{-\alpha}} \\
&= \frac{\alpha\Gamma(n+\alpha)}{n!\,\Gamma(1+\alpha)} \frac{p^n(1-p)^\alpha}{1-(1-p)^\alpha},\; n = 1,2,\ldots
\end{aligned}$$

In the first case, for $k = 1,2,\ldots$, from the equality (2),

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{1}{P[N(p)\geq k]}\sum_{i=k}^{+\infty}\frac{[1-F_X(x)]^i}{i!}\sum_{j=0}^{+\infty}\frac{-1}{\ln(1-p)}\frac{p^{j+i}}{j+i}\frac{(j+i)!}{j!}[F_X(x)]^j$$

$$= 1 - \frac{1}{P[N(p)\geq k]}\sum_{i=k}^{+\infty}\frac{\{p[1-F_X(x)]\}^i}{i!}\frac{-1}{\ln(1-p)}\sum_{j=0}^{+\infty}\frac{(j+i-1)!}{j!}[pF_X(x)]^j$$
$$= 1 - \frac{1}{P[N(p)\geq k]}\sum_{i=k}^{+\infty}\frac{\{p[1-F_X(x)]\}^i}{i!}\frac{-1}{\ln(1-p)}\frac{(i-1)!}{[1-pF_X(x)]^i}.$$

Taking $p^* = \frac{p[1-F_X(x)]}{1-pF_X(x)}$ we obtain

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{1}{P[N(p)\geq k]}\frac{\ln(1-p^*)}{\ln(1-p)}\sum_{i=k}^{+\infty}\frac{-1}{\ln(1-p^*)}\frac{(p^*)^i}{i}$$
$$= 1 - \frac{P[N(p^*)\geq k]}{P[N(p)\geq k]}\frac{\ln(1-p^*)}{\ln(1-p)}.$$

Since
$$p_1 = -\frac{p}{\ln(1-p)} \Rightarrow \ln(1-p) = \frac{-p}{P[N(p)=1]}$$

we obtain

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x) = 1 - \frac{P[N(p^*)\geq k]}{P[N(p)\geq k]}\frac{P[N(p)=1]}{P[N(p^*)=1]}\frac{p^*}{p}. \tag{4}$$

In the second case, again using (2), we obtain

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \sum_{i=k}^{+\infty}\frac{[1-F_X(x)]^i}{P[N(p)\geq k]\,i!}\sum_{j=0}^{+\infty}\frac{\alpha\Gamma(j+i+\alpha)}{(j+i)!\Gamma(1+\alpha)}\frac{p^{j+i}(1-p)^\alpha}{1-(1-p)^\alpha}\frac{(j+i)!}{j!}[F_X(x)]^j$$
$$= 1 - \frac{1}{P[N(p)\geq k]}\sum_{i=k}^{+\infty}\frac{\{p[1-F_X(x)]\}^i}{i!\Gamma(1+\alpha)}\frac{\alpha(1-p)^\alpha}{1-(1-p)^\alpha}\sum_{j=0}^{+\infty}\Gamma(j+i+\alpha)\frac{1}{j!}[pF_X(x)]^j.$$

Since
$$\sum_{j=0}^{+\infty}\Gamma(j+i+\alpha)\frac{1}{j!}[pF_X(x)]^j = \frac{\Gamma(\alpha+i)}{[1-pF_X(x)]^{\alpha+i}}$$

we obtain

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{1}{P[N(p)\geq k]}\sum_{i=k}^{+\infty}\frac{\{p[1-F_X(x)]\}^i}{i!\Gamma(1+\alpha)}\frac{\alpha(1-p)^\alpha}{1-(1-p)^\alpha}\frac{\Gamma(\alpha+i)}{[1-pF_X(x)]^{\alpha+i}}$$
$$= 1 - \frac{1}{P[N(p)\geq k]}\frac{(1-p)^\alpha}{1-(1-p)^\alpha}\frac{1}{[1-pF_X(x)]^\alpha}\frac{1-(1-p^*)^\alpha}{(1-p^*)^\alpha}\times$$
$$\times\sum_{i=k}^{+\infty}\frac{\alpha\Gamma(\alpha+i)}{i!\Gamma(1+\alpha)}(p^*)^i\frac{(1-p^*)^\alpha}{1-(1-p^*)^\alpha}$$

with $p^* = \frac{p[1-F_X(x)]}{1-pF_X(x)} = \frac{p-1+1-pF_X(x)}{1-pF_X(x)} = 1-\frac{1-p}{1-pF_X(x)}$. Hence,

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{P[N(p^*)\geq k]}{P[N(p)\geq k]}\frac{1}{[1-pF_X(x)]^\alpha}\frac{(1-p)^\alpha}{1-(1-p)^\alpha}\frac{1-(1-p^*)^\alpha}{(1-p^*)^\alpha}.$$

But

$$p_1 = \frac{\alpha p(1-p)^\alpha}{1-(1-p)^\alpha} \Rightarrow \frac{(1-p)^\alpha}{1-(1-p)^\alpha} = \frac{P[N(p)=1]}{\alpha p}$$

and hence

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)$$
$$= 1 - \frac{P[N(p^*)\geq k]}{P[N(p)\geq k]}\frac{1}{[1-pF_X(x)]^\alpha}\frac{P[N(p)=1]}{\alpha p}\frac{\alpha p^*}{P[N(p^*)=1]}$$
$$= 1 - \frac{P[N(p^*)\geq k]}{P[N(p)\geq k]}\frac{1}{[1-pF_X(x)]^\alpha}\frac{P[N(p)=1]}{P[N(p^*)=1]}\frac{p^*}{p}$$
$$= 1 - \frac{P[N(p^*)\geq k]}{P[N(p)\geq k]}\left(\frac{1-p^*}{1-p}\right)^\alpha\frac{P[N(p)=1]}{P[N(p^*)=1]}\frac{p^*}{p}.$$

6 Limit Theorems for Order Statistics of Geometrically Thinned Sequences

The geometric order statistics inherit some of the simple properties of order statistics in the traditional iid scheme. The extensions of Smirnov [11] asymptotic limit distributions for $k:N(p)$ geometrically randomly stopped order statistics that follow exhibit some of those similarities.

In this section we will study the limit behaviour (when $p\downarrow 0$) of the distributions of $X_{N(p)-k+1:N(p)|N(p)\geq k}$ and of $X_{k:N(p)|N(p)\geq k}$, when properly normalized. Let us start with $X_{N(p)-k+1:N(p)|N(p)\geq k}$ and call $\mathscr{G}^{(k)}$ the associated limit distribution. From expression (1) we know that:

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)=\sum_{j=k}^{+\infty}\sum_{i=0}^{k-1}\binom{j}{i}\left[1-F_X(x)\right]^i\left[F_X(x)\right]^{j-i}\frac{P\left[N(p)=j\right]}{P\left[N(p)\geq k\right]}.$$

Replacing the expression of the mass probability of $N(p)$ and noting that $1=\sum_{i=0}^{j}\binom{j}{i}\left[1-F_X(x)\right]^i\left[F_X(x)\right]^{j-i}$ we can write

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)=p\sum_{j=k}^{+\infty}\frac{(1-p)^{j-1}}{(1-p)^{k-1}}\left[1-\sum_{i=k}^{j}\binom{j}{i}\left[1-F_X(x)\right]^i\left[F_X(x)\right]^{j-i}\right].$$

After some simplifications and the exchange of the order of the two sums we obtain

$$\begin{aligned}&F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)\\&=1-\frac{p}{(1-p)^{k-1}}\sum_{i=k}^{+\infty}\left[1-F_X(x)\right]^i(1-p)^{i-1}\sum_{j=0}^{+\infty}\binom{j+i}{i}\left[(1-p)F_X(x)\right]^j.\end{aligned}$$

Finally, using the equality $\sum_{j=0}^{+\infty}\left[\binom{j+i}{i}y^j\right]=\frac{1}{(1-y)^{i+1}}$, $|y|<1$, and proceeding to some more simplifications, we obtain

$$F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x)=1-\left[1-\frac{pF_X(x)}{1-(1-p)F_X(x)}\right]^k. \tag{5}$$

Suppose that the rv $X_{N(p):N(p)}|N(p)\geq 1$, conveniently normalized, converges to a non-degenerate distribution function $\mathscr{G}$, i.e., that there exists α and $\beta>0$, functions of p, such that

$$\lim_{p\downarrow 0}F_{\frac{X_{N(p):N(p)}-\alpha(p)}{\beta(p)}}(x)=\lim_{p\downarrow 0}F_{X_{N(p):N(p)}}\left(\beta(p)x+\alpha(p)\right)=\mathscr{G}(x).$$

Let $x_p=\beta(p)x+\alpha(p)$. Hence, from (5) with $k=1$, we conclude that

$$\begin{aligned}\lim_{p\downarrow 0}F_{X_{N(p):N(p)}|N(p)\geq 1}(x_p)&=\lim_{p\downarrow 0}\left\{1-\left[1-\frac{p\,F_X(x_p)}{1-(1-p)F_X(x_p)}\right]\right\}\\&=\lim_{p\downarrow 0}\frac{p\,F_X(x_p)}{1-(1-p)F_X(x_p)}=\mathscr{G}(x).\end{aligned}$$

For general k we have then

$$\begin{aligned}\lim_{p\downarrow 0}F_{X_{N(p)-k+1:N(p)}|N(p)\geq k}(x_p)&=\lim_{p\downarrow 0}\left\{1-\left[1-\frac{p\,F_X(x_p)}{1-(1-p)F_X(x)}\right]^k\right\}\\&=1-\left[1-\mathscr{G}(x)\right]^k,\end{aligned}$$

proving the following

Theorem 3. *If $X_{N(p):N(p)}$ properly normalized weakly converges to a nondegenerate rv with distribution function $\mathscr{G}$, i.e., if there exists $\alpha(p)$ and $\beta(p) > 0$, such that*

$$\lim_{p\downarrow 0} F_{X_{N(p):N(p)}}\left(\beta(p)x + \alpha(p)\right) = \mathscr{G}(x),$$

then the $k-$geometric maximum, $X_{N(p)-k+1:N(p)}|N(p) \geq k$, properly normalized with the same functions, i.e., the rv

$$\frac{X_{N(p)-k+1:N(p)|N(p)\geq k} - \alpha(p)}{\beta(p)},$$

weakly converges to a nondegenerate rv with distribution function $\mathscr{G}^{(k)}$ given by

$$\mathscr{G}^{(k)}(x) = 1 - \left[1 - \mathscr{G}(x)\right]^k.$$

For the k-geometric order statistic a similar reasoning leads to the equality

$$F_{X_{k:N(p)}|N(p)\geq k}(x) = \left[\frac{F_X(x)}{p + (1-p)F_X(x)}\right]^k.$$

Taking $x_p = \beta(p)x + \alpha(p)$ and $p \downarrow 0$,suppose that the limit of $F_{X_{1:N(p)}|N(p)\geq 1}(x_p)$, when $p \downarrow 0$, exists and let $\mathscr{L}$ be such limit:

$$\lim_{p\downarrow 0} F_{X_{1:N(p)}|N(p)\geq 1}(x_p) = \mathscr{L}(x) = \lim_{p\downarrow 0} \frac{F_X(x_p)}{p + (1-p)F_X(x_p)}.$$

Then

$$\lim_{p\downarrow 0} F_{X_{k:N(p)}|N(p)\geq k}(x_p) = \lim_{p\downarrow 0}\left[\frac{F_X(x_p)}{p + (1-p)F_X(x_p)}\right]^k = \mathscr{L}^k(x),$$

which leads to the

Theorem 4. *If $X_{1:N(p)}$ properly normalized weakly converges to a nondegenerate rv with distribution function $\mathscr{L}$, i.e., if there exists $\alpha(p)$ and $\beta(p) > 0$, such that*

$$\lim_{p\downarrow 0} F_{X_{1:N(p)}}\left(\beta(p)x + \alpha(p)\right) = \mathscr{L}(x),$$

then the $k-$geometric order statistic, $X_{k:N(p)}|N(p) \geq k$, similarly normalized, i.e., the rv

$$\frac{X_{k:N(p)|N(p)\geq k} - \alpha(p)}{\beta(p)},$$

weakly converges to a nondegenerate random variable with distribution function $\mathscr{L}^{(k)}$ given by $\mathscr{L}^{(k)}(x) = \mathscr{L}^k(x)$.

Acknowledgements This research has been supported by National Funds through FCT — Fundação para a Ciência e a Tecnologia, project PEst-OE/MAT/UI0006/2011, and PTDC/FEDER.

References

1. Engen, S.: On species frequency models. Biometrika **61**, 263–270 (1974).
2. Galambos, J.: The Asymptotic Theory of Extreme Order Statistics. Wiley, (1987)
3. Gnedenko, B.V., Korolev, V.U. : Random Summation: Limit Theorems and Applications. CRC Press (1996)
4. Hess, K.T., Liewald, A., Schmidt, K.D.: An extension of Panjer's recursion. ASTIN Bull. **32** 283–297 (2002)
5. Katz, L.: Unified treatment of a broad class of discrete probability distributions. In Patil, G.P., ed., Classical and Contagious Discrete Distributions.Pergamon Press, Oxford, pp. 175–182 (1965)
6. Morris, C.L.: Natural exponential families with quadratic variance functions. Ann. Statist. **10** 65–80 (1982)
7. Panjer, H.H.: Recursive evaluation of a family of compound distributions. ASTIN Bull. **12** 22–26 (1981)
8. Pestana, D., Velosa, S.: Extensions of Katz-Panjer families of discrete distributions, REVS-TAT **2** 145–162 (2004)
9. Rachev, S.T., Resnick, S.: Max geometric infinite divisibility and stability. Communications in Statistics — Stochastic Models **7**, 191–218 (1991)
10. Rólski, T.;,Schmidli, H., Schmidt, V., Teugels, J.: Stochastic Processes for Insurance and Finance. Wiley, New York (1999)
11. Smirnov, N. V.: Limit distributions for the terms of a variational series. Trudy Mat. Inst. Steklov., Acad. Sci. USSR, Moscow-Leningrad **25** 3–60 (1949)
12. Sundt, B.: On some extensions of Panjer's class of counting distributions. ASTIN Bull. **22** 61–80 (1992)
13. Sundt, B., Jewell, W.: Further results on recursive evaluation of compound distributions, ASTIN Bull. **12** 27–39 (1981)
14. Willmot, G.E.: Sundt and Jewell's family of discrete distributions, ASTIN Bull. **18** 17–29 (1988)