
Extreme Value Results for Scan Statistics

M.V. Boutsikas, M.V. Koutras and F.S. Milienos

Department of Statistics and Insurance Science, University of Piraeus

Abstract: In the first part of this paper we focus on the classical scan and multiple scan statistic, defined on a sequence of independent and identically distributed (i.i.d.) binary trials and review a number of bounds and approximations for their distributions which have been developed by the aid of distance measures. Moreover, we discuss briefly a number of asymptotic results that have been established by setting up appropriate conditions guaranteeing the convergence (to zero) of the distance measures' upper bounds. In the second part, we study a multiple scan statistic enumerating variable by considering a general threshold-based framework, defined on i.i.d. continuous random variables. More specifically, we prove first a compound Poisson approximation for the total number of fixed length overlapping moving windows containing prespecified number of threshold exceedances. The classical scan and multiple scan statistic may be treated as a special case of this general model. Next we exploit the previous result to gain some new extreme value results for the scan enumerating statistic under the assumption that the continuous random variables belong to the maximum domain of attraction of one of the three extreme value distributions (Fréchet, reversed Weibull, Gumbel). Finally, we elucidate how the general results can be applied in a number of classical continuous distributions (Pareto, Uniform, Exponential and Normal).

Keywords and phrases: Scan and multiple scan statistic, Poisson and compound Poisson approximation, Erdős-Rényi statistic, extreme value theory, maximum domain of attraction, moving sums and exceedances.

1.1 Introduction

The discrete scan statistic $S_{n,k}$ in a sequence of n binary trials (1:success, 0:failure) has been defined as the maximum number of successes within any k consecutive trials (n and k are two positive integers with $k \leq n$). Due to its widespread

applicability in an abundance of research areas such as quality control, actuarial science, reliability theory, molecular biology etc. it has been an attractive subject of continuing research interest for the past few decades; see e.g. the monographs by Balakrishnan and Koutras (2002), Glaz, Naus and Wallenstein (2001) and the special issue edited by Glaz and Balakrishnan (1999).

An instance where $S_{n,k}$ arises in quite a natural way is in randomness tests when the null hypothesis of uniformity and independence of a sequence of binary observations $X_i, i = 1, 2, \dots, n$ is to be tested against the alternative hypothesis of clustering of 1's due to local dependence between $X_i, i = 1, 2, \dots, n$ or due to the existence of subsequences of consecutive X_i with $P(X_i = 1) > p$. As Glaz and Naus (1991) indicated, the generalized likelihood ratio test for checking the hypothesis of uniformity, rejects the null hypothesis of uniformity whenever $S_{n,k} \geq c$, with the value of c being determined by the significance level of the test. Apparently, the evaluation of c such that a prespecified significance level is achieved calls for the distribution of the test statistic $S_{n,k}$. Since randomness tests are frequently applied to large data sets, theoretical developments related to the asymptotic distribution of $S_{n,k}$ (as $n, k \rightarrow \infty$) will play a primary role in the analysis of the test.

In actuarial context, let us consider a portfolio with n daily claims and denote by $X_i, i = 1, 2, \dots, n$ the binary variable describing whether the i -th claim exceeds a threshold $u > 0$ ($X_i = 1$) or not ($X_i = 0$). Then $S_{n,k}$ will describe the maximum number of "large claims" (i.e. claims exceeding threshold u) in a period of k consecutive days. The primary interest in this situation is also focused on extremely long periods ($n, k \rightarrow \infty$) and therefore one should look at the asymptotic distribution of $S_{n,k}$.

Exact results for the distribution of the scan statistic were discussed in Fu (2001), Balakrishnan and Koutras (2002) and Fu and Lou (2003). Since the evaluation of the exact distribution is computationally intractable, especially for large values of the parameters n, k , several approximations and bounds have been developed during the last decades.

Another random variable closely related to $S_{n,k}$ is the number of occurrences of k consecutive trials which contain at least r successes among them. If we denote by $W_{n,k,r}$ the resulting (overlapping) enumerating variable when a sequence of n trials is realized, it is clear that the probability mass function of $W_{n,k,r}$ at 0 coincides to the quantity $P(S_{n,k} < r)$ i.e.

$$P(W_{n,k,r} = 0) = P(S_{n,k} < r).$$

The statistic $W_{n,k,r}$ is referred in the statistical literature under the name *multiple scan statistic*. Balakrishnan and Koutras (2002) have introduced two additional enumeration schemes for scan occurrence counting, and used the terminology "type III enumeration" for the overlapping scheme, while the terms "type I" and "type II" were reserved for the non-overlapping counting schemes.

In the present article we shall restrict our discussion to the overlapping scheme only.

Although, quite accurate approximations are available by now for the probability mass function of $W_{n,k,r}$ at 0 (for a review see Chen and Glaz (1999)) when the question comes to the whole distribution of $W_{n,k,r}$ the problem becomes extremely complex. Koutras and Alexandrou (1995) have described a method to obtain the exact distribution of $W_{n,k,r}$ by invoking a Markov chain embedding technique; however, this approach becomes unwieldy for k and r of moderate size, while its computational complexity for large k, r and n renders the whole procedure as non-feasible. Therefore, the development of asymptotic results for the distribution of $W_{n,k,r}$ is of special interest.

In the present article we are reviewing some asymptotic results on the scan and multiple scan statistic. In Section 1.2 we introduce all necessary notations and some notions that will be used in the subsequent sections. The scan and multiple scan statistic is introduced in a threshold-based framework and the classical binary statistics are viewed as special cases of the general model. An alternative realization of the multiple scan statistic in terms of order statistics is given as well.

Section 1.3 deals with the (binary) discrete scan statistic and multiple scan statistic. We review first (Subsection 1.3.1) a number of bounds and approximations which have been developed by the aid of distance measures between discrete distributions. It is worth stressing that we have confined ourselves to techniques offering error bounds, so that the establishment of convergence theorems be easily achieved (Subsection 1.3.2). Section 1.3 is completed by providing a number of extreme value results for the discrete scan statistic (Subsection 1.3.3).

In Section 1.4 we present several results for the multiple scan statistic under the threshold exceedance framework. A compound Poisson approximation is firstly established (Subsection 1.4.1); this result is subsequently used in Subsection 1.4.2 to derive extreme value results for the multiple scan statistic under the assumption that the random variables (whose threshold exceedances are studied) are belonging to the maximum domain of attraction of the three classical extreme value distributions (Fréchet, reversed Weibull, Gumbel). Finally, in Subsection 1.4.3, we present applications of the general results in a number of typical continuous distributions (Pareto, Uniform, Exponential, Normal and Weibull).

1.2 Definitions and notations

Let Y_1, Y_2, \dots, Y_n be a sequence of independent and identically distributed continuous random variables with cumulative distribution function F and denote by $X_i(u)$ the indicator variable

$$X_i(u) = I_{(u, \infty)}(Y_i) = \begin{cases} 1, & \text{if } Y_i > u \\ 0, & \text{if } Y_i \leq u \end{cases}, \quad i = 1, 2, \dots, n.$$

The quantity $u \in \Re$ is a fixed threshold which is exceeded by Y_i with probability

$$p = P(X_i(u) = 1) = E(X_i(u)) = P(Y_i > u) = \bar{F}(u)$$

where \bar{F} denotes the tail probability of Y_i . Considering all the moving windows of length k in the sequence Y_1, Y_2, \dots, Y_n , namely,

$$Y_i, Y_{i+1}, \dots, Y_{i+k-1}, \quad i = 1, 2, \dots, n - k + 1$$

we may introduce the k -scan exceedance process as follows

$$S_k^{(i)}(u) = \sum_{j=i}^{i+k-1} X_j(u) = \sum_{j=i}^{i+k-1} I_{(u, \infty)}(Y_j), \quad i = 1, 2, \dots, n - k + 1.$$

Manifestly, $S_k^{(i)}$ denotes the number of random variables, among $Y_i, Y_{i+1}, \dots, Y_{i+k-1}$, whose value exceeds threshold u , while

$$S_{n,k}(u) = \max_{1 \leq i \leq n-k+1} S_k^{(i)}(u) = \max_{1 \leq i \leq n-k+1} \sum_{j=i}^{i+k-1} X_j(u)$$

expresses the maximum number of exceedance occurrences among all possible moving windows of length k , in the sequence Y_1, Y_2, \dots, Y_n . A closely related random variable is

$$W_{n,k,r}(u) = \sum_{i=1}^{n-k+1} I_{[r, \infty)}(S_k^{(i)}(u))$$

which enumerates the total number of overlapping moving windows of length k in which the threshold exceedances are at least r . It is plain from the previous definitions that

$$P(W_{n,k,r}(u) = 0) = P(S_{n,k}(u) < r).$$

There is an interesting connection between the variables $W_{n,k,r}(u)$, $S_{n,k}(u)$ and moving order statistics of the original sample Y_1, Y_2, \dots, Y_n . To elucidate this, let us first arrange the observations of the i -th moving window (of length k)

$Y_i, Y_{i+1}, \dots, Y_{i+k-1}$ in descending order and denote by $Y_{r:k}^{(i)}$ the r -th larger observation, i.e.

$$Y_{1:k}^{(i)} \geq Y_{2:k}^{(i)} \geq \dots \geq Y_{k:k}^{(i)}.$$

For fixed r , consider next the random variables $Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}$, arrange them again in descending order and denote by $Y_{m:r:k}$ the m -th larger among them, that is

$$Y_{1:r:k} \geq Y_{2:r:k} \geq \dots \geq Y_{n-k+1:r:k}.$$

In particular, for $m = 1$ and $m = n - k + 1$ we may write

$$\begin{aligned} Y_{1:r:k} &= \max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}), \\ Y_{n-k+1:r:k} &= \min(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}). \end{aligned}$$

It should be mentioned that the parameter n has been suppressed, in the last notations. In case n is not fixed (e.g. if we wish to investigate the asymptotic behavior as $n \rightarrow \infty$) we shall use the notation $Y_{m:r:k}(n)$ instead of $Y_{m:r:k}$. The cumulative distribution function of $Y_{m:r:k}(n)$ can be readily expressed in terms of the distribution of $W_{n,k,r}(u)$ as follows

$$\begin{aligned} P(Y_{m:r:k}(n) \leq u) &= P(\text{at most } m-1 \text{ of } Y_{r:k}^{(1)}, \dots, Y_{r:k}^{(n-k+1)} \text{ exceed } u) \\ &= P\left(\sum_{i=1}^{n-k+1} I_{(u, \infty)}(Y_{r:k}^{(i)}) < m\right) \\ &= P\left(\sum_{i=1}^{n-k+1} I_{[r, \infty)}\left(\sum_{j=i}^{i+k-1} I_{(u, \infty)}(Y_j)\right) < m\right) \\ &= P\left(\sum_{i=1}^{n-k+1} I_{[r, \infty)}(S_k^{(i)}(u)) < m\right) \\ &= P(W_{n,k,r}(u) < m). \end{aligned} \tag{1.1}$$

As a consequence, the cumulative distribution function of the maximum of the moving window order statistics $Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}$ is given by the probability mass function of $W_{n,k,r}(u)$ at zero, namely

$$P(\max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}) \leq u) = P(W_{n,k,r}(u) = 0) \tag{1.2}$$

It should be stressed that the maximum appearing above is over a set of dependent variables, since $Y_{r:k}^{(i)}$ involve order statistics over overlapping sets of variables.

Note also that, the notation used here for the order statistics is slightly different from the one used in traditional order statistics books (see e.g. Arnold and Balakrishnan (1989) and David and Nagaraja (2003)); by our notation, the r -th order statistic refers to the r -th *largest* observation instead of the r -th *smallest*.

Let us next consider the special case where the random variables Y_i follow the uniform distribution on $(0, 1)$ and set $u = 1 - p$ where $0 < p < 1$. Then $X_i(u)$ becomes a Bernoulli variable, say X_i , with success probability p and $S_{n,k}(u)$ reduces to the binary discrete scan statistic $S_{n,k}$ mentioned in the introduction.

In the sequel we shall make use of the term *threshold based scan statistic* to refer to the general model and *binary scan statistic* to refer to the special case where X_i are i.i.d. Bernoulli variables with common success probabilities p . In the latter case, the following notations will be practiced

$$\begin{aligned} S_k^{(i)} &= \sum_{j=i}^{i+k-1} X_j, \quad i = 1, 2, \dots, n - k + 1, \\ S_{n,k} &= \max_{1 \leq i \leq n-k+1} S_k^{(i)}, \\ W_n &= W_{n,k,r} = \sum_{i=1}^{n-k+1} I_{[r, \infty)}(S_k^{(i)}). \end{aligned}$$

The notions $\sim, O(\cdot)$ will assume their usual meaning, i.e.

$$\begin{aligned} f(t) \sim g(t) \text{ as } t \rightarrow t_0 &\text{ if } \lim_{t \rightarrow t_0} \frac{f(t)}{g(t)} = 1, \\ f(t) = O(g(t)) &\text{ if } \frac{f(t)}{g(t)} \text{ is bounded.} \end{aligned}$$

1.3 The binary scan statistic

In this section we present several results for the binary discrete scan statistic $S_{n,k}$ and the binary multiple scan statistic $W_{n,k,r}$. In Subsection 1.3.1 we review several bounds and approximations that have been suggested for deriving accurate estimates of the distributions of the binary scan statistics when n is large. Most of them use an appropriate Poisson or compound Poisson approximation along with efficient error bounds for the distance between the exact and approximating distribution. Although another class of approximations which exploits product-type formula is available in the literature, it will not be covered

herein, since usually they do not offer error estimates and therefore no asymptotic (convergence) results can be achieved by them; the interested reader may refer to the monograph by Glaz, Naus and Wallenstein (2001) for a detailed presentation of product-type approximations.

A very popular method for establishing Poisson approximations for sums of (potentially) dependent Bernoulli random variables is the celebrated Chen-Stein method; see Arratia Goldstein and Gordon (1990) or the monograph by Barbour, Holst and Janson (1992). This method can be used to compute an upper bound for the total variation distance between the law $\mathcal{L}(W_{n,k,r})$ of the random variable $W_{n,k,r}$ and a Poisson distribution $Po(\lambda)$, i.e.

$$d_{TV}(\mathcal{L}(W_{n,k,r}), Po(\lambda)) = \sup_A |P(W_{n,k,r} \in A) - P(Z_\lambda \in A)|,$$

where Z_λ is a random variable following the Poisson distribution with $E(Z_\lambda) = \lambda > 0$ and A any subset of non-negative integers.

Since the scan enumerating variable $W_{n,k,r}$ takes into account clusters of consecutive trials with high concentration of successes and these events tend to occur in clumps, the compound Poisson distribution is an even more natural choice than the Poisson.

We recall that the term compound Poisson distribution with parameter λ and compounding distribution G (notation: $CP(\lambda, G)$) refers to the distribution of the random sum $U = \sum_{i=1}^N Z_i$ where N is a Poisson random variable with $\lambda = E(N)$ and Z_i are i.i.d. variables, independent of N , with cumulative distribution function G .

The probability generating function of U takes on the form

$$P_U(t) = E(t^U) = e^{-\lambda(1-E(t^Z))} = e^{-\lambda(1-P_Z(t))} \tag{1.3}$$

where $P_Z(t)$ is the probability generating function of Z_i , $i = 1, 2, \dots$. As a consequence one could evaluate the probability mass function of U by considering the power series expansion of $P_U(t)$ (provided that an explicit expression is available for the probability generating function $P_Z(t)$ of the compounding distribution).

1.3.1 Bounds and approximations

Denote by $b(x; l, p)$ and $B(x; l, p)$ the probability mass function and cumulative distribution function of a binomial random variable X , i.e.

$$b(x; l, p) = P(X = x) = \binom{l}{x} p^x (1-p)^{l-x}, \quad x = 0, 1, \dots, n,$$

$$B(x; l, p) = P(X \leq x) = \sum_{j=0}^{\lfloor x \rfloor} b(j; l, p), \quad x \in \mathfrak{R}.$$

where the symbol $[x]$ indicates the integer part of x . In the following sections we shall make use of the quantities

$$\begin{aligned} f(s; k, p) &= P(S_k^{(1)} < s, S_k^{(2)} < s, \dots, S_k^{(k)} < s, S_k^{(k+1)} \geq s) \\ G(s; k, p) &= P(S_k^{(1)} < s, S_k^{(2)} < s, \dots, S_k^{(k+1)} < s) \end{aligned}$$

which can be expressed in terms of $b(x; l, p)$ and $B(x; l, p)$ as follows (see, Glaz and Naus (1991))

$$\begin{aligned} f(s; k, p) &= \frac{p}{s} b(s-1; k-1, p) [s(1-p)b(s-1; k-1, p) \\ &\quad + (s-kp)B(s-2; k-1, p)], \quad (1.4) \\ G(s; k, p) &= B(s-1; k, p)^2 - b(s; k, p) [(s-1)B(s-2; k, p) \\ &\quad - kpB(s-3; k-1, p)], \end{aligned}$$

with $1 \leq s \leq k$ (if $s > k$ or $s < 0$ then we set $f(s; k, p) = 0$).

We shall start with a result pertaining to the binary scan statistic $S_{n,k}$ which was established by Arratia, Gordon and Waterman (1990) by the aid of the Chen-Stein method. Let us introduce first the random variables

$$C_i = \begin{cases} 1, & \text{if } \sum_{j=i}^{i+k-1} X_j = r \\ 0, & \text{else} \end{cases}, \quad i = 1, 2, \dots, n-k+1.$$

(convention: $C_i = 0$ for $i \leq 0$) and then define an auxiliary variable W by summing-up the quantities

$$D_i = C_i \prod_{j=1}^k (1 - C_{i-j}), \quad i = 1, 2, \dots, n-k+1$$

i.e.

$$W = \sum_{i=1}^{n-k+1} C_i \prod_{j=1}^k (1 - C_{i-j}).$$

Arratia, Gordon and Waterman (1990) proved the following interesting result.

Theorem 1.3.1 *If $p < r/k < 1$ then*

$$\left| P(S_{n,k} < r) - e^{-E(W)} \right| \leq 7kb(r; k, p) + (1 - B(r; k, p))$$

that is, $P(S_{n,k} < r)$ is bounded below and above by the quantities $e^{-E(W)} \pm UB$ where UB denotes the quantity in the RHS of the last inequality, and $E(W)$ equals $(n-k+1)E(D_i)$.

The same authors proved that $E(W)$ can be bounded as follows

$$\frac{r}{k} - p \leq \frac{E(W)}{(n - k + 1)b(r; k, p)} \leq \left(\frac{r}{k} - p\right) + 2\left(1 - \frac{r}{k}\right)(1 - B(r; k, p)) \quad (1.5)$$

while an alternative upper bound is given by

$$\frac{E(W)}{(n - k + 1)b(r; k, p)} \leq \left(\frac{r}{k} - p\right) + 2\left(1 - \frac{r}{k}\right)e^{-kH(r/k, p)}$$

where $H(\theta, p)$ denotes the Kullback-Leibler distance (or relative entropy)

$$H(\theta, p) = \theta \ln \left(\frac{\theta}{p}\right) + (1 - \theta) \ln \left(\frac{1 - \theta}{1 - p}\right) = \ln \frac{\theta^\theta (1 - \theta)^{1 - \theta}}{p^\theta (1 - p)^{1 - \theta}}, \quad 0 < p < \theta < 1 \quad (1.6)$$

For large values of k (with r/k being kept fixed) the second summand in the RHS of (1.5) becomes negligible and therefore the next simple approximation formula can be used

$$E(W) \approx \left(\frac{r}{k} - p\right)(n - k + 1)b(r; k, p).$$

Dembo and Karlin (1992), used the Chen-Stein method to establish upper bounds for the total variation distance between a scan process generated by a sequence of i.i.d. positive random variables (not necessarily binary) and a Poisson distribution. In the special case of i.i.d. binary variables, the next simpler results ensue.

Theorem 1.3.2 *If*

$$\lambda = (n - k + 1)B(r - 1; k, p), \mu = (n - k + 1)(1 - B(r - 1; k, p)).$$

then

$$\begin{aligned} \mathbf{a.} \quad d_{TV}(\mathcal{L}(W_{n,k,r}), Po(\mu)) &\leq (1 - e^{-\mu})[(2k - 1)(1 - B(r - 1; k, p)) \\ &\quad + 2 \sum_{i=1}^{k-1} P(S_k^{(i+1)} \geq r | S_k^{(1)} \geq r)] \end{aligned}$$

$$\begin{aligned} \mathbf{b.} \quad d_{TV}(\mathcal{L}(n - k + 1 - W_{n,k,r}), Po(\lambda)) &\leq (1 - e^{-\lambda})[(2k - 1)B(r - 1; k, p) \\ &\quad + 2 \sum_{i=1}^{k-1} B(r - 1; i, p)] \end{aligned}$$

The conditional probabilities $P(S_k^{(i+1)} \geq r | S_k^{(1)} \geq r)$ appearing in the first upper bound can be easily expressed in terms of binomial probabilities as follows

$$P(S_k^{(i+1)} \geq r | S_k^{(1)} \geq r) = \frac{1}{1 - B(r-1; k, p)} \cdot \sum_{s=0}^{k-i} (1 - B(r-s-1; i, p))^2 b(k-i; s, p)$$

with $B(x; l, p) = b(x; l, p) = 0$, for $x < 0$. Manifestly, the results of Theorem 1.3.2 can be exploited to bound the probability mass function or the cumulative distribution function of $W_{n,k,r}$ by an interval centered on the probability mass function or cumulative distribution function of a Poisson distribution respectively, whose length is two times the error bounds provided above.

As already mentioned, due to the fact that windows of high concentration of successes tend to occur in clumps, one should expect that the Poisson approximations described in Theorem 1.3.2 would naturally provide poor results. As a matter of fact the upper bound in (a) converges to zero only for $r = k$ and $p \rightarrow 0$ while the upper bound in (b) only for $r = 1$ and $p \rightarrow 1$

Motivated by a sequence matching problem where the need of an accurate approximation of the discrete scan statistic came into sight, Goldstein and Waterman (1992) used the declumping variables

$$E_i = I_{[r, \infty)}(S_k^{(i)}) \prod_{j=1}^{\min(s, i-1)} (1 - I_{[r, \infty)}(S_k^{(i-j)})), \quad i = 1, 2, \dots, n - k + 1$$

(where s is a fixed positive integer¹) to establish total variation bounds for the compound Poisson approximation of $W_{n,k,r}$. The binary variables E_i indicate the occurrence of a clump with starting point the variable X_i , while the distribution of the number C of occurrences of the event $S_k^{(j)} \geq r$ within the clump, is given by

$$P(C = c) = P\left(\sum_{j=i}^{\beta} I_{[r, \infty)}(S_k^{(j)}) = c | E_i = 1\right), \quad c = 1, 2, \dots$$

where

$$\beta = \min(\gamma \geq i : I_{[r, \infty)}(S_k^{(\gamma)}) = 1, I_{[r, \infty)}(S_k^{(\gamma+1)}) = 0, \dots, I_{[r, \infty)}(S_k^{(\gamma+s)}) = 0).$$

Considering $s = k$, the following result can be established by the aid of Chen-Stein method (cf. Goldstein and Waterman (1992)).

¹Products of the form $\prod_{i=i_1}^{i_2} f(i)$ with $i_1 > i_2$, are conventionally set equals 1.

Theorem 1.3.3 *The total variation distance between $W_{n,k,r}$ and a compound Poisson distribution $CP(\lambda, G)$ with $\lambda = (n - k + 1)(1 - B(r - 1; k, p))/E(C)$ and compounding distribution $G(x) = P(C \leq x)$ is bounded above as follows*

$$d_{TV}(\mathcal{L}(W_{n,k,r}), CP(\lambda, G)) \leq 6\lambda^2(1 + E(C))\frac{k}{n - k} + 2\lambda P(C > k).$$

Note that the above expression for λ was deduced upon ignoring the boundary effects. The exact distribution of C is quite intricate; to overcome this, Goldstein and Waterman (1992) have suggested an alternative simpler but less accurate approximation which will not be presented here. In case one is interested only in the distribution of $S_{n,k}$ and not the whole distribution of $W_{n,k,r}$ the following simple lower and upper bounds for $E(C)$ (given that $s = k$) may be proved useful

$$\frac{r}{k} - p \leq \frac{1}{E(C)} \leq \left(\frac{r}{k} - p\right) + 2\left(1 - \frac{r}{k}\right)e^{-kH(r/k, p)}.$$

A slightly different approach for deducing an upper bound for the distance between the distribution of $W_{n,k,r}$ and a compound Poisson distribution was practiced by Boutsikas and Koutras (2002). They used the "truncated" de-clumping variables

$$E'_i = (1 - I_{[r, \infty)}(S_k^{(i-1)})) \sum_{j=1}^k \prod_{l=i}^{i+j-1} (I_{[r, \infty)}(S_k^{(l)})), \quad i = 1, 2, \dots, n - k + 1$$

and exploiting a general result developed by the same authors (Boutsikas and Koutras (2001)), they proved the following theorem. For the presentation of the results up to the end of this section, the evaluation of $S_k^{(i)}$'s is carried out by assuming that the sequence of trials $X_i, i = 1, 2, \dots, n$ is extended for $i < 1$ and $i > n$. Note that the distance involved in this result is no more the total variation distance but the Kolmogorov distance which is defined as follows (U, V are any random variables with cumulative distributions functions F_U and F_V , respectively)

$$d_K(\mathcal{L}(U), \mathcal{L}(V)) = d_K(F_U, F_V) = \sup_{-\infty < x < \infty} |F_U(x) - F_V(x)|.$$

Theorem 1.3.4 *Let $\lambda = (n - k + 1)P(E'_1 > 0)$ and $G(x) = P(E'_1 \leq x | E'_1 > 0)$*

$x = 0, 1, \dots, k$. Then

$$\begin{aligned}
d_K(\mathcal{L}(W_{n,k,r}), CP(\lambda, G)) &\leq (\lambda + 1) \sum_{i=r}^k \binom{k}{i} p^i (1-p)^{k-i} + \\
&\quad (\lambda(3k-1) + k-1) \binom{k-1}{r-1} p^r (1-p)^{k-r+1} + \\
&\quad (n-k) \sum_{b=2}^{k-1} \sum_{i=\max\{0, r-k+b-1\}}^{\min\{b-2, r-2\}} \binom{k-b}{r-i-1} \binom{b-2}{i} \\
&\quad \binom{k-b}{r-i-2} p^{2r-i-1} (1-p)^{2k-b-2r+i+3}.
\end{aligned}$$

The advantage of this result over the one stated in Theorem 1.3.3 is that both the upper bound for the distance and the parameters λ, G of the compound Poisson distribution admit explicit expressions which are computationally tractable. It can be easily verified that $P(E'_1 > 0) = \binom{k-1}{r-1} p^r q^{k-r+1}$ and therefore

$$\lambda = (n-k+1)P(E'_1 > 0) = (n-k+1) \binom{k-1}{r-1} p^r q^{k-r+1}.$$

Moreover, the cumulative distribution function $G(x)$ of the compounding distribution has been computed by Boutsikas and Koutras (2002) as

$$\begin{aligned}
G(x) &= 1 - \sum_{j=\max\{0, r-x-1\}}^{\min\{k-x-1, r-1\}} \left(\frac{\binom{x}{x-r+j+1} \binom{k-x-1}{j}}{\binom{k-1}{r-1}} \right) \\
&\quad \cdot \left(\binom{x}{x-r+j+1} p^{r-j-1} (1-p)^{x-r+j+1} + \right. \\
&\quad \left. + \left(1 - \frac{(x+1)(1-p)}{x-r+j+2} \right) \left(\sum_{i=0}^{x-r+j} \binom{x}{i} (1-p)^i p^{x-i-1} \right) \right)
\end{aligned}$$

for $x = 1, 2, \dots, k-1, G(0) = 0, G(k) = 1$. The upper bound described in Theorem 1.3.4 is of order $O(p)$ (for $r < k$); therefore it will produce reasonable lower and upper bounds for the cumulative distribution function of $W_{n,k,r}$ when $p \rightarrow 0$. None the less, if p is fixed the quality of the bounds will not be as good and it is conceivable that no asymptotic results could be established by the aid of them for fixed p and $n, k \rightarrow \infty$.

In order to cover the last case, the following family of declumping variables may be used

$$E''_i = \left(\prod_{j=i-k}^{i-1} (1 - I_{[r, \infty)}(S_k^{(j)})) \right) I_{[r, \infty)}(S_k^{(i)}) \left(\sum_{l=i}^{i+k} I_{[r, \infty)}(S_k^{(l)}) \right), \quad i = 1, \dots$$

The last bracket enumerates the number of scanning windows of length k that begin at positions $i, i + 1, \dots, i + k$ and contain at least r successes each. On the other hand, the first bracket guarantees that in the previous k positions $i - k, i - k + 1, \dots, i - 1$ all scanning windows of length k contain fewer than r successes. As a matter of fact, it is the inclusion of this extra term that makes the construction of sharp bounds feasible. Exploiting the new family of declumping variables $E_i'', i = 1, 2, \dots$, Boutsikas and Koutras (2006) arrived at the next result.

Theorem 1.3.5 *Let $\lambda = (n - k + 1)P(E_1'' > 0) = (n - k + 1)f(r; k, p)$, and*

$$G(x) = P(E_1'' \leq x | E_1'' > 0)$$

$$= P \left(\sum_{l=k+1}^{2k+1} I_{[r, \infty)}(S_k^{(l)}) \leq x \mid I_{[r, \infty)}(S_k^{(l)}) = 0, l = 1, 2, \dots, k, I_{[r, \infty)}(S_k^{(k+1)}) = 1 \right),$$

$$x = 0, 1, \dots, k.$$

Then

$$d_K(\mathcal{L}(W_{n,k,r}), CP(\lambda, G)) \leq (2k - 1)\lambda p(1 - p)b(r - 1; k - 1, p)$$

$$+ 3\lambda k f(r; k, p) + (\lambda + 2)(1 - G(r; k, p))$$

where $f(r; k, p), G(r; k, p)$ are given in (1.4)

Although the evaluation of the compounding cumulative distribution function $G(x)$ is not easily accomplished, the above result is quite appealing for establishing the asymptotic behaviour of $W_{n,k,r}$ for fixed p and $n, k \rightarrow \infty$. Moreover, if the interest is focused on $S_{n,k}$ and not the whole distribution function of $W_{n,k,r}$ then the cumulative distribution function of $S_{n,k}$, i.e. $P(S_{n,k} < r) = P(W_{n,k,r} = 0)$ can be effortlessly assessed by the aid of the next simpler result.

Corollary 1.3.1 *The cumulative distribution function of $S_{n,k}$ can be approximated by $e^{-\lambda}$, $\lambda = (n - k + 1)f(r; k, p)$ with the error of approximation being bounded above as follows*

$$|P(S_{n,k} < r) - e^{-\lambda}| \leq (2k - 1)\lambda p(1 - p)b(r - 1; k - 1, p) + 3\lambda k f(r; k, p)$$

$$+ (\lambda + 2)(1 - G(r; k, p)).$$

1.3.2 Asymptotic results

In the present section we are presenting a number of asymptotic results pertaining to the Poisson and compound Poisson convergence of the discrete scan statistics $S_{n,k}$ and $W_{n,k,r}$. Although we are not going to present the technical details of the proofs of these results, alert readers may easily extract them by

the aid of the bounds (on the total variation or Kolmogorv distances) described in the previous section.

In the light of Theorem 1.3.1 and the discussion following it, we may state the next result, which was provided by Arratia, Gordon and Waterman (1990).

Corollary 1.3.2 *If n, k, r are positive integers with $p < r/k \neq 1$ and $\lambda = (n - k + 1) \binom{r}{k} - p) b(r; k, p)$ then*

$$P(S_{n,k} < r) \text{ can be approximated by } e^{-\lambda}$$

with the approximation error being of order $O(\frac{\ln n}{n})$.

In the next Corollary, which can be easily proved by the aid of Theorem 1.3.4 (see Boutsikas and Koutras (2002) for details), a compound Poisson convergence for $W_{n,k,r}$ is established.

Corollary 1.3.3 *Assume that k, r are kept fixed, while $n \rightarrow \infty, p_n \rightarrow 0$ so that $\lambda_n = (n - k + 1) \binom{k-1}{r-1} p_n^r (1 - p_n)^{k-r+1} \rightarrow \lambda \in (0, \infty)$. Then $W_{n,k,r}$ converges weakly to a compound Poisson distribution with parameter λ and compounding distribution*

$$G(x) = \begin{cases} 0, & x \leq 0 \\ 1 - \frac{\binom{k-x-1}{r-1}}{\binom{k-1}{r-1}}, & x = 1, 2, \dots, k - r, \\ 1, & x \geq k - r + 1. \end{cases}$$

Under the same assumptions for the cumulative distribution of $S_{n,k}$ we have

$$P(S_{n,k} < r) \approx e^{-\lambda}.$$

The probability mass function $g(x)$ of the compounding distribution $G(x)$ takes on the form

$$g(x) = G(x) - G(x - 1) = \frac{\binom{k-x-1}{r-2}}{\binom{k-1}{r-1}}, \quad x = 1, 2, \dots, k - r + 1.$$

It is of interest to note that, in the special case $r = 2 < k$, $g(x)$ reduces to the discrete uniform distribution on the integers $1, 2, \dots, k - 1$. Also, for $k = r$, the compounding distribution becomes degenerate (with all its mass being concentrated at 1) and the limiting compound Poisson law $CP(\lambda, G)$ reduces to an ordinary Poisson distribution.

The probability generating function of the compound Poisson distribution described in Corollary 1.3.3 reads (cf.(1.3))

$$\begin{aligned} P(t) &= E(t^{W_{n,k,r}}) = e^{-\lambda(1-E(t^Z))} = e^{-\lambda(1-\sum_{x=1}^{k-r+1} t^x g(x))} \\ &= \exp\left(-\lambda\left(1 - \sum_{x=1}^{k-r+1} t^x \frac{\binom{k-x-1}{r-2}}{\binom{k-1}{r-1}}\right)\right) \end{aligned}$$

and therefore we may easily compute the probabilities $P(W_{n,k,r} = i)$ considering the power series expansion of $P(t)$ and extracting the coefficients of i -th order term, $i = 0, 1, \dots$. This can be done numerically (for specific values of the parameters) or analytically for small order terms. For example,

$$\begin{aligned} P(W_{n,k,r} = 0) &= \frac{1}{0!} P(0) = e^{-\lambda}, \\ P(W_{n,k,r} = 1) &= \frac{1}{1!} \left. \frac{dP(t)}{dt} \right|_{t=0} = \lambda g(1) e^{-\lambda} = \lambda \frac{r-1}{k-1} e^{-\lambda} \\ P(W_{n,k,r} = 2) &= \frac{1}{2!} \left. \frac{d^2P(t)}{dt^2} \right|_{t=0} = \lambda \frac{(r-1)(k-r)}{(k-1)(k-2)} e^{-\lambda} + \frac{\lambda^2}{2!} \left(\frac{r-1}{k-1} \right)^2 e^{-\lambda} \end{aligned} \quad (1.7)$$

etc. Alternatively, one may resort to the next recursive scheme (see, Bowers *et al.* (1997))

$$\begin{aligned} P(W_{n,k,r} = 0) &= e^{-\lambda}, \\ P(W_{n,k,r} = i) &= \frac{\lambda k}{r i} \binom{k}{r}^{-1} \sum_{j=1}^{\min(k-r+1, i)} j \binom{k-j-1}{r-2} P(W_{n,k,r} = i-j), \\ & \qquad \qquad \qquad i = 1, 2, \dots \end{aligned}$$

For $r < k$, the rate of convergence ascertained by Corollary 1.3.3 for the approximation $P(S_{n,k} < r) \approx e^{-\lambda}$ is of order $O(p)$. As a consequence, it is not applicable for fixed p and $n, k \rightarrow \infty$. The last case is covered by the next result which can be inferred from Theorem 1.3.5; the interested reader is referred to Boutsikas and Koutras (2006) for the technical details.

We shall use the symbol $h(\theta, p)$ to denote the derivative of the Kullback-Leibler distance $H(\theta, p)$ (1.6) with respect to θ , i.e.

$$h(\theta, p) = \frac{d}{d\theta} H(\theta, p) = \ln \frac{\theta(1-p)}{p(1-\theta)}, \quad 0 < p < \theta < 1.$$

Corollary 1.3.4 *Let p be fixed, $\theta \in (p, 1)$, and k_n, r_n two sequences satisfying the condition*

$$\lim_{n \rightarrow \infty} \frac{r_n - \theta k_n}{\sqrt{k_n}} = 0.$$

If $\rho_n = r_n - \theta k_n$ and the sequence

$$l_n = n \frac{(\theta - p) e^{-k_n H(\theta, p) - \rho_n h(\theta, p)}}{\sqrt{2\pi\theta(1-\theta)k_n}}, \quad n = 1, 2, \dots$$

is bounded from above, then

$$P(S_{n,k} < r) \sim e^{-l_n}$$

with the rate of convergence being of order $O(\frac{\rho_n^2 + 1}{k_n})$.

Practically speaking, the last Corollary states that, for large values of n, k, r and $p < r/k \neq 1$ the cumulative distribution function of $S_{n,k}$ can be approximated by the aid of the formula (replace, r_n, k_n, θ by $r, k, r/k$ respectively in the formula of l_n)

$$P(S_{n,k} < r) \sim \exp\left(-n \frac{(r - kp)e^{-kH(r/k,p) - \rho h(r/k,p)}}{\sqrt{2\pi rk(k - r)}}\right).$$

1.3.3 Extreme value results

Extreme value results pertaining to moving sums of i.i.d. random variables (non necessarily binary) have been the subject of continuing interest for many decades. They are usually referred to as Erdős-Rényi laws and deal with the asymptotic distribution of

$$U_n = \max_{1 \leq i \leq n-k+1} \sum_{j=i}^{i+k-1} Z_j$$

where Z_1, Z_2, \dots is a sequence of i.i.d. random variables. The classical Erdős-Rényi (1970) theorem establishes the almost sure convergence to 1 of the sequence of random variables $U_n/(ak_n)$ for a large class of distributions for Z_i ($k = k_n = \lfloor c \ln(n) \rfloor$ for some positive constant c and $a > 0$ is a number depending on c and the distribution of Z_i). An extreme value theorem for U_n obtained by Deheuvels and Devroye (1987), states that under the assumption that Z_i follow a non-lattice distribution with zero mean,

$$\lim_{n \rightarrow \infty} P\left(\frac{U_n - b_n}{a_n} \leq x\right) = \Lambda(x)$$

where $\Lambda(x) = \exp(-e^{-x})$, $x \in \mathfrak{R}$ is the cumulative distribution of the Gumbel distribution, and $a_n, b_n \in \mathfrak{R}$ are appropriate sequences of normalizing constants.

We shall now present two extreme value results for the discrete scan statistic $S_{n,k}$; the fact that the original sequence of variables is binary (and therefore lattice), makes it possible to express the normalizing constants by explicit formulae.

The following theorem is a slight restatement of a result established by Arratia, Gordon and Waterman (1990).

Theorem 1.3.6 *Let $k > -(\ln p)^{-1} \ln n$, denote by $\theta = \theta(n, k, p) \in (p, 1)$ the unique solution² of the equation*

$$H(\theta, p) = \frac{\ln n}{k}$$

²Since $\frac{d}{d\theta} H(\theta, p) = h(\theta, p) > 0$, the quantity $H(\theta, p)$ varies monotonically from 0 to $-\ln p$; therefore the equation $H(\theta, p) = c$ admits a unique solution $\theta \in (p, 1)$ for $0 < c < -\ln p$

and define the normalizing constants b_n as follows

$$b_n = \theta \frac{\ln n}{H(\theta, p)} - \frac{1}{2h(\theta, p)} \ln(\ln n) - \frac{1}{2h(\theta, p)} \ln\left(\frac{2\pi\theta(1-\theta)}{H(\theta, p)}\right) + \frac{\ln(\theta-p)}{h(\theta, p)}.$$

Then for each $\varepsilon > 0$ such that $1 + \varepsilon \leq -(\ln n)^{-1} k \ln p \leq 1/\varepsilon$ the following result holds true, uniformly for $n, k \rightarrow \infty$

$$\sup_x |P(S_{n,k} - b_n < x) - \Lambda(h(\theta, p)x)| \rightarrow 0$$

The supremum is evaluated over all $x \in \mathfrak{R}$ such that $x + b_n$ is a positive integer.

Exploiting Corollary 1.3.4 it is not difficult to gain the following extreme value theorem which is also establishing the convergence to the Gumbel distribution of an appropriate normalized version of $S_{n,k}$ (cf. Boutsikas and Koutras (2006)).

Theorem 1.3.7 For fixed $p \in (0, 1)$, let θ to be a number in the interval $(p, 1)$ and define

$$\begin{aligned} k_n &= \lfloor \ln n / H(\theta, p) \rfloor, \\ b_n &= k_n \theta + \frac{1}{h(\theta, p)} \ln \frac{n(\theta-p)e^{-k_n H(\theta, p)}}{\sqrt{2\pi\theta(1-\theta)k_n}}, \\ \epsilon_n(y) &= \left(b_n + \frac{y}{h(\theta, p)} \right) - \left\lfloor b_n + \frac{y}{h(\theta, p)} \right\rfloor. \end{aligned}$$

Then

$$\lim_{n \rightarrow \infty} \left[P \left(\frac{S_{n,k} - b_n}{1/h(\theta, p)} < y \right) - \Lambda(y - \epsilon_n(y)h(\theta, p)) \right] = 0$$

with convergence rate of order $O\left(\frac{(\ln k_n)^2}{k_n}\right)$.

1.4 Scan statistics exceedances

In this section we shall present some new results for the multiple scan statistic under the threshold exceedance framework. More specifically, we consider a sequence Y_1, Y_2, \dots, Y_n of continuous i.i.d. random variables with cumulative distribution function F , and a threshold $u = u_n$ which varies with n . Exploiting Corollary 1.3.3 we establish first a compound Poisson convergence theorem for $W_{n,k,r}(u_{a_n})$ where a_n is an appropriate sequence of positive real numbers. Then we use this result to develop asymptotic results for $W_{n,k,r}(u_{a_n})$ or equivalently for the moving order statistic $Y_{m:r:k}(n)$ introduced in Section 1.2 under the assumption that the distribution of Y_i belongs to the domain of attraction of one of the three classical extreme type distributions (Weibull, Fréchet, Gumbel). Finally, application of the general results for typical continuous distributions, along with illustrative graphs exhibiting the quality of convergence are provided.

1.4.1 Compound Poisson approximation for $W_{n,k,r}(u)$

Let us assume that the threshold u_n varies with n so that the event $Y_i > u_n$ becomes a rare event, i.e. $P(Y_i > u_n) \rightarrow 0$ as $n \rightarrow \infty$. A standard condition that yields non degenerate results is the following

$$\lim_{n \rightarrow \infty} nP(Y_i > u_n) = \lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau \in (0, \infty).$$

Denoting by a_n the sequence $a_n = n^{1/r}$, $n = 1, 2, \dots$ it is clear that

$$\lim_{n \rightarrow \infty} n\bar{F}(u_{a_n})^r = \lim_{n \rightarrow \infty} (a_n\bar{F}(u_{a_n}))^r = \tau^r.$$

and viewing $X_i(u_{a_n}) = I_{(u_{a_n}, \infty)}(Y_i)$, $i = 1, 2, \dots$ as a sequence of binary trials with success probabilities $p_n = \bar{F}(u_{a_n})$ we may apply Corollary 1.3.3 with

$$\lambda_n = (n - k + 1) \binom{k-1}{r-1} p_n^r (1 - p_n)^{k-r+1} \rightarrow \binom{k-1}{r-1} \tau^r > 0$$

to conclude that $W_{n,k,r}(u_{a_n})$ converges to a compound Poisson distribution. Recalling formula (1.1) we arrive at the following interesting result.

Theorem 1.4.1 *If there exist a sequence u_n such that $\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau > 0$, then the distribution of the number $W_{n,k,r}(u_{a_n})$ of (overlapping) moving windows of length k that contain at least r exceedances of the threshold u_{a_n} with $a_n = n^{1/r}$ (k and r are fixed positive integers), converges to a compound Poisson distribution with parameter $\lambda = \binom{k-1}{r-1} \tau^r$ and compounding distribution with probability mass function*

$$g(x) = \frac{\binom{k-x-1}{r-2}}{\binom{k-1}{r-1}}, \quad x = 1, 2, \dots, k - r + 1. \quad (1.8)$$

Moreover, if we denote by f_{CP} the probability mass function of the compound Poisson distribution we may write

$$\lim_{n \rightarrow \infty} P(Y_{m:r:k}(n) \leq u_{a_n}) = \lim_{n \rightarrow \infty} P(W_{n,k,r}(u_{a_n}) < m) = \sum_{i=0}^{m-1} f_{CP}(i).$$

For the evaluation of the limiting distribution, see the discussion after Corollary 1.3.3.

Although u_{a_n} is typically defined over the set of positive integers n , hereafter we shall assume that such sequences can be extended in \mathfrak{R}^+ i.e. $u_x = u(x)$, $x \in \mathfrak{R}^+$. Under this assumption, it makes sense to write u_{a_n} where a_n is any sequence of (not necessarily integer) numbers, a notation that will be met frequently in the next Sections. It is worth stressing that this technique works in a nice way for all the examples illustrated in Section 1.4.3. If the aforementioned extension is not feasible, one may use $u_{\lfloor a_n \rfloor}$ in place of u_{a_n} and the results provided hereafter remain valid as well.

It is worthwhile to note that, for the determination of the limiting value $\binom{k-1}{r-1} \tau^r$ of λ_n , use was made of the fact that

$$(1 - p_n)^{k-r+1} = \left(1 - \frac{a_n \overline{F}(u_{a_n})}{a_n}\right)^{k-r+1} \sim \left(1 - \frac{\tau}{a_n}\right)^{k-r+1} \xrightarrow{n \rightarrow \infty} 1$$

However, the above convergence is quite slow, even for small values of r . For example, if $k = 4, r = 2, \tau = 2$ then $q_{100}^{k-r+1} \approx 0.512, q_{1000}^{k-r+1} \approx 0.822, q_{10000}^{k-r+1} \approx 0.941$. Things get much worse when r is larger. Therefore, although the asymptotic result in Theorem 1.4.1 is valid for $n \rightarrow \infty$, the distribution of $W_{r,k,n}(u_{a_n})$ may be approximated more accurately by a compound Poisson distribution $CP(\lambda_n^*, G)$ with parameter

$$\lambda_n^* = \binom{k-1}{r-1} \tau^r \left(1 - \frac{\tau}{a_n}\right)^{k-r+1} = \lambda \left(1 - \frac{\tau}{a_n}\right)^{k-r+1} \quad (1.9)$$

(instead of λ). This remark is exploited in the numeric experimentation carried out in Subsection 1.4.3.

A result of similar fashion with the one stated in Theorem 1.4.1 has been given by Dudkewicz (1998); however his result covers only the special case of moving minima and is applicable for different conditions on the parameters k, r .

In view of Theorem 1.4.1 and taking into account (1.2) we may write that

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}) \leq u_{a_n}) \\ = \lim_{n \rightarrow \infty} P(W_{n,k,r}(u_{a_n}) = 0) = e^{-\lambda}. \end{aligned}$$

Apparently the above formula displays the asymptotic behaviour of the maximum of a family of dependent variables, namely $Y_{r:k}^{(i)}$, $i = 1, 2, \dots, n - k + 1$,

whose marginal tails can be expressed through the tail of the binomial distribution with parameters k and $\overline{F}(u_{a_n})$, namely

$$nP(Y_{r:k}^{(i)} > u_{a_n}) = n \sum_{i=r}^k \binom{k}{i} \overline{F}(u_{a_n})^i F(u_{a_n})^{k-i} \sim \binom{k}{r} (a_n \overline{F}(u_{a_n}))^r \rightarrow \binom{k}{r} \tau^r.$$

If we assumed for the moment that $Y_{r:k}^{(i)}$ were independent, the limiting behaviour of their maximum would reduce to

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}) \leq u_{a_n}) \\ = \lim_{n \rightarrow \infty} \left(1 - \frac{nP(Y_{r:k}^{(i)} > u_{a_n})}{n} \right)^n = e^{-\lambda_{ind}}, \end{aligned}$$

where $\lambda_{ind} = \binom{k}{r} \tau^r$. The ratio

$$\frac{\lambda}{\lambda_{ind}} = \frac{\binom{k-1}{r-1} \tau^r}{\binom{k}{r} \tau^r} = \frac{r}{k}$$

characterizes the extremal dependence between $Y_{r:k}^{(i)}$, $i = 1, 2, \dots, n - k + 1$. This quantity has been termed as *extremal index* by Embrechts *et al.* (1997) and as stated therein, it provides information on the local dependence of the variables under study. According to the last result, the extremal index for the sequence $Y_{r:k}^{(i)}$, $i = 1, 2, \dots, n - k + 1$ decreases (and therefore the local dependence becomes stronger) as k increases or r decreases, a fact that can be easily interpreted intuitively.

1.4.2 Convergence of threshold based scan statistics under maximum domain of attraction assumptions

The probabilistic extreme value theory focuses on the stochastic behaviour of the maximum $M_n = \max(Z_1, Z_2, \dots, Z_n)$ (and the minimum) of sequences of i.i.d. random variables Z_1, Z_2, \dots, Z_n . The asymptotic properties of extremes (maxima or minima), intermediate order statistics and exceedances over (or below) prespecified thresholds are determined by the upper and lower tails of the underlying distribution.

Although extreme value theory seems to have originated mainly from the needs of astronomers in accepting or rejecting outlying observations, after the significant theoretical developments on it during 1920-1950, a substantial number of articles appeared which deal with practical applications of extreme value statistics in the stochastic analysis of meteorological phenomena (rainfalls, floods), strengths of materials, seismic activity, insurance and actuarial models, radioactive emissions etc. For a comprehensive bibliography of literature on extreme

value distributions and their applications, the interested reader is referred to the monographs by Kotz and Nadarajah (2000), Reiss and Thomas (2000), Coles (2001) and Embrechts *et al.* (1997).

In theory, the distribution of M_n can be derived exactly by the formula $P(M_n \leq x) = F(x)^n$; however this is not immediately helpful in practice, since the cumulative distribution function F of Z_i is usually unknown. A typical approach to overcome this, is to look for approximate (asymptotic) families of models for F^n which can be estimated on the basis of extreme data only. This is analogous to the classical practice of approximating the distribution of sample means by the normal distribution as justified by the central limit theorem.

Since $\lim_{n \rightarrow \infty} P(M_n \leq x) = \lim_{n \rightarrow \infty} F(x)^n = 0$ for all x such that $F(x) < 1$, the asymptotic distribution of M_n does not provide any valuable information. However, it is conceivable that more insight into the magnitude of maxima would be given by the centered and normalized maxima (a fact that parallels the normalization used in the classical limit theorem).

The fundamental Fisher-Tippett theorem, states that, if there exist constants $c_n > 0$ and $d_n \in \mathfrak{R}$ such that

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = H(x) \tag{1.10}$$

where H is a non-degenerate distribution function, then H belongs to the location-scale family of one of the following three distributions

$$\text{Frechet : } \quad \Phi_a(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-a}), & x > 0 \end{cases}, \quad a > 0$$

$$\text{(Reversed)Weibull : } \Psi_a(x) = \begin{cases} \exp(-(-x)^a), & x \leq 0 \\ 1, & x > 0 \end{cases}, \quad a > 0$$

$$\text{Gumbel : } \quad \Lambda(x) = \exp(-e^{-x}), \quad x \in \mathfrak{R}.$$

i.e. $H(x) = \Phi_a((x - \mu)/\sigma)$ or $H(x) = \Psi_a((x - \mu)/\sigma)$ or $H(x) = \Lambda((x - \mu)/\sigma)$ where $\mu \in \mathfrak{R}$ is a location parameter and $\sigma > 0$ a scale parameter. These families have been termed extreme value distributions, while the respective sequences c_n, d_n are called *norming constants*.

If (1.10) holds true for a cumulative distribution function F , then we shall say that \bar{F} belongs to the maximum domain of attraction of H (notation: $\bar{F} \in MDA(H)$). The next theorem characterizes the family of distributions that belong to a MDA (the reader may consult any of the monographs provided earlier for the proof).

Theorem 1.4.2 *The cumulative distribution function F belongs to the maximum domain of attraction of the extreme value distribution H with norming constants $c_n > 0, d_n \in \mathfrak{R}$ if and only if*

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n x + d_n) = -\ln H(x), \quad x \in \mathfrak{R}.$$

When $H(x) = 0$ the limit is interpreted as $+\infty$.

We shall now establish some non-degenerate convergence results for the threshold based multiple scan statistic $W_{n,k,r}$, under the assumption that the cumulative distribution function F of Y_i belongs to the MDA of Φ_a , Ψ_a or Λ .

Theorem 1.4.3 *If $\overline{F} \in MDA(H)$ with norming constants $c_n > 0, d_n \in \mathfrak{R}$ then*

$$\lim_{n \rightarrow \infty} P\left(\frac{Y_{m:r:k}(n) - d_{a_n}}{c_{a_n}} \leq x\right) = \lim_{n \rightarrow \infty} P(W_{n,k,r}(c_{a_n}x + d_{a_n}) < m) = \sum_{i=0}^{m-1} f_{CP}(x; i)$$

where $a_n = n^{1/r}$ and $f_{CP}(x; \cdot)$ is the probability mass function of a compound Poisson distribution with parameters

$$\lambda(x) = \binom{k-1}{r-1} (-\ln H(x))^r \quad (1.11)$$

and compounding probability mass function given by (1.8).

PROOF. Since $\overline{F} \in MDA(H)$, we conclude by virtue of Theorem 1.4.2, that

$$\lim_{n \rightarrow \infty} n\overline{F}(c_n x + d_n) = -\ln H(x), \quad x \in \mathfrak{R}.$$

Applying next Theorem 1.4.1 for $u_n = c_n x + d_n$ and $\tau = -\ln H(x)$ we deduce that $W_{n,k,r}(u_{a_n})$ converges in distribution to a compound Poisson distribution with parameter

$$\binom{k-1}{r-1} \tau^r = \binom{k-1}{r-1} (-\ln H(x))^r = \lambda(x) \quad (1.12)$$

and compounding probability mass function the one described in (1.8). The assertion for the asymptotic distribution of the moving window order statistic $Y_{m:r:k}$ follows immediately by exploiting formula (1.1). \blacksquare

According to Theorem 1.4.3, the asymptotic distribution of $W_{n,k,r}(c_{a_n}x + d_{a_n})$ can be approximated for large values of n by a compound Poisson distribution with parameter $\lambda(x)$ given by (1.12). In view of the comments following the proof of Theorem 1.4.1 one may improve the quality of approximation if he replaces the parameter $\lambda(x)$ by (cf. (1.9))

$$\lambda^*(x) = \lambda(x) \left(1 - \frac{\tau}{a_n}\right)^{k-r+1} = \binom{k-1}{r-1} (-\ln H(x))^r \left(1 + \frac{\ln H(x)}{n^{1/r}}\right)^{k-r+1}.$$

Specializing in the three classical extreme value distributions we mention in brief the following (we shall use in the sequel the notation x_F for the right end point of F i.e. $x_F = \sup\{x \in \mathfrak{R} : F(x) < 1\}$):

a. Maximum domain of attraction of Frechet

If $\bar{F} \in MDA(\Phi_a)$ then $x_F = \infty$ and a possible choice of c_n, d_n is $c_n = F^{-1}(1 - n^{-1})$ and $d_n = 0$ (where F^{-1} denotes the generalized inverse function of F). Since $-\ln H(x) = -\ln \Phi_a(x) = x^{-a}$ the parameters $\lambda(x), \lambda^*(x)$ reduce to

$$\lambda(x) = \binom{k-1}{r-1} x^{-ra}, \lambda^*(x) = \binom{k-1}{r-1} x^{-ra} \left(1 - \frac{x^{-a}}{n^{1/r}}\right)^{k-r+1}, \quad x > 0.$$

Typical members of this class are the classical heavy (right) tailed distributions, e.g. Cauchy, Pareto, Loggamma.

b. Maximum domain of attraction of (Reversed) Weibull

If $\bar{F} \in MDA(\Psi_a)$ then x_F is finite and a feasible choice for c_n, d_n is $c_n = x_F - F^{-1}(1 - n^{-1})$ and $d_n = x_F$. The parameters of the approximating compound Poisson distributions are now

$$\lambda(x) = \binom{k-1}{r-1} (-x)^{ra}, \tag{1.13}$$

$$\lambda^*(x) = \binom{k-1}{r-1} (-x)^{ra} \left(1 - \frac{(-x)^a}{n^{1/r}}\right)^{k-r+1}, \quad x \leq 0. \tag{1.14}$$

Typical members of this class are the Uniform and Beta distribution

c. Maximum domain of attraction of Gumbel

If $\bar{F} \in MDA(\Lambda)$ then F admits a representation of the form (see e.g. Embrechts et al. (1997))

$$\bar{F}(x) = c(x) e^{-\int_z^x \frac{g(t)}{a(t)} dt}, \quad z < x < x_F$$

where z is a real number with $z < x_F$ and c, g are (measurable) functions such that $c(x) \rightarrow c_0 > 0$, and $g(x) \rightarrow 1$ when $x \uparrow x_F$. The function $a(\cdot)$ is a positive, absolutely continuous function with density a' such that $a'(x) \rightarrow 0$ as $x \uparrow x_F$. A valid choice for the function a is

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt.$$

while the norming constants can be defined as $d_n = F^{-1}(1 - n^{-1})$ and $c_n = a(d_n)$. The parameters $\lambda(x), \lambda^*(x)$, of the approximating compound Poisson distributions read now

$$\lambda(x) = \binom{k-1}{r-1} e^{-rx}, \tag{1.15}$$

$$\lambda^*(x) = \binom{k-1}{r-1} e^{-rx} \left(1 - \frac{e^{-x}}{n^{1/r}}\right)^{k-r+1}. \tag{1.16}$$

Typical members of this class are the Normal, Exponential and Gamma distribution.

The special case $m = 1$ of Theorem 1.4.3 reveals the asymptotic distribution of the maximum of a specific order statistic evaluated on moving windows of fixed length k , namely

$$Y_{1:r:k} = \max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}).$$

In applications, Y_i usually represent values of a process measured on a regular time-scale, e.g. hourly measurements of sea level, daily claim sizes in a specific portfolio, monthly mean temperatures. Then the statistic $Y_{r:k}^{(i)}$ is a location measure for fixed length periods (e.g. $Y_{r:k}$ might be the median for the k observations obtained over k consecutive days) and therefore $Y_{1:r:k}$ will correspond to the maximum of the location measures. Such a scenario arises in quite a natural way if the collected data are erased after a certain period (due to storage restrictions) or can be scanned by a scanner of restricted range. Another instance where $Y_{1:r:k}$ might be used is in a similar fashion with the *MA* (moving average) charts, if we replace the average by a more robust location measure such as the median (or any other order statistic of the moving sample).

The asymptotic distribution of $Y_{1:r:k}$ is described in the next Corollary.

Corollary 1.4.1 *If $\bar{F} \in MDA(H)$ with norming constants $c_n > 0, d_n \in \mathfrak{R}$ then*

$$\lim_{n \rightarrow \infty} P \left(\frac{\max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}) - d_{a_n}}{c_{a_n}} \leq x \right) = e^{-\lambda(x)}$$

where $\lambda(x)$ is given by (1.11).

PROOF. Results immediately by virtue of the obvious formula

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left(\frac{\max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)}) - d_{a_n}}{c_{a_n}} \leq x \right) \\ = \lim_{n \rightarrow \infty} P(W_{n,k,r}(c_{a_n}x + d_{a_n}) < 1) = f_{CP}(x; 0) = e^{-\lambda(x)}. \end{aligned}$$

■

It is noteworthy that, the asymptotic distribution of the maximum of the moving window order statistics (properly centered and normalized) belongs to the same domain of attraction as the original distribution of Y_i ; the only parameters that are affected are the location parameter μ and the scale parameter σ . To establish a formal proof of this assertion it suffices to observe the following

a. Maximum domain of attraction of Frechet

If $H(x) = \Phi_a(x)$ then

$$\lambda(x) = \binom{k-1}{r-1} (-\ln \Phi_a(x))^{-r} = \binom{k-1}{r-1} (x)^{-ra} = \left(\frac{x-\mu}{\sigma}\right)^{-a'}, x > 0$$

where

$$a' = ar, \quad \mu = 0, \quad \sigma = \binom{k-1}{r-1}^{1/ra}$$

and therefore

$$e^{-\lambda(x)} = e^{-(\frac{x-\mu}{\sigma})^{-a'}} = \Phi_{a'}\left(\frac{x-\mu}{\sigma}\right).$$

b. Maximum domain of attraction of (Reversed) Weibull

If $H(x) = \Psi_a(x)$ then

$$\lambda(x) = \binom{k-1}{r-1} (-\ln \Psi_a(x))^{-r} = \binom{k-1}{r-1} (-x)^{ra} = \left(-\frac{x-\mu}{\sigma}\right)^{a'}, x \leq 0$$

where

$$a' = ar, \quad \mu = 0, \quad \sigma = \binom{k-1}{r-1}^{-1/ra}$$

and therefore

$$e^{-\lambda(x)} = e^{-(\frac{x-\mu}{\sigma})^{-a'}} = \Psi_{a'}\left(\frac{x-\mu}{\sigma}\right).$$

c. Maximum domain of attraction of Gumbel

If $H(x) = \Lambda(x)$ then

$$\lambda(x) = \binom{k-1}{r-1} (-\ln \Lambda(x))^{-r} = \binom{k-1}{r-1} e^{-rx} = e^{-\frac{x-\mu}{\sigma}}, x \in \mathfrak{R}$$

where

$$\mu = \frac{1}{r} \ln \binom{k-1}{r-1}, \sigma = \frac{1}{r}$$

and therefore

$$e^{-\lambda(x)} = e^{-e^{-\frac{x-\mu}{\sigma}}} = \Lambda\left(\frac{x-\mu}{\sigma}\right).$$

Finally, if Corollary 1.4.1 is to be used for approximating the distribution of $Y_{1:r;k}$ for small values of n , the quality of approximation will be improved by using $\lambda^*(x)$ instead of $\lambda(x)$.

1.4.3 Examples

In order to exemplify further the usefulness of extreme value theory as presented in the previous sections we consider some typical continuous distributions and illustrate the effectiveness of the approximations established by Theorem 1.4.3 and Corollary 1.4.1.

a. Pareto Distribution. Let us assume that Y_1, Y_2, \dots follow a typical Pareto distribution with cumulative distribution function

$$F(x) = 1 - \left(\frac{c}{x}\right)^a, x \geq c$$

where a and c are two positive parameters. This is perhaps the most popular heavy-tailed distribution with a lot of applications in socioeconomic and insurance/actuarial models, see e.g. Johnson *et al.* (1994) or the excellent text by Arnold (1985).

In this case, letting $c_n = F^{-1}(1 - n^{-1}) = cn^{1/a}$, $d_n = 0$, we get

$$n\bar{F}(c_n x + d_n) = n \left(\frac{c}{cn^{1/a}x}\right)^a = x^{-a}, x > 0,$$

which ascertains that $\bar{F} \in MDA(\Phi_a)$. Hence, by Theorem 1.4.3

$$\lim_{n \rightarrow \infty} P\left(\frac{Y_{m:r:k}(n)}{cn^{1/ra}} \leq x\right) = \lim_{n \rightarrow \infty} P(W_{n,k,r}(cn^{1/ra}x) < m) = \sum_{i=0}^{m-1} f_{CP}(x; i), x > 0$$

where $f_{CP}(x; \cdot)$ is the probability mass function of a compound Poisson distribution with parameter

$$\lambda(x) = \binom{k-1}{r-1} x^{-ra}, x > 0$$

and compounding distribution with the density (1.8). Moreover,

$$\lim_{n \rightarrow \infty} P\left(\frac{\max(Y_{r:k}^{(1)}, Y_{r:k}^{(2)}, \dots, Y_{r:k}^{(n-k+1)})}{cn^{1/ra}} \leq x\right) = e^{-\lambda(x)}, x > 0$$

while a better approximation could be achieved on using

$$\lambda^*(x) = \binom{k-1}{r-1} x^{-ra} \left(1 - \frac{x^{-a}}{n^{1/r}}\right)^{k-r+1}, x > 0.$$

in place of $\lambda(x)$.

By way of example we mention also that, should one be interested in the distribution of the second largest among the moving order statistics $Y_{r:k}^{(i)}$, $i =$

$1, 2, \dots, n - k + 1$, he could resort to the approximate formula (apply Theorem 1.4.3 for $m = 2$ and recall (1.7))

$$P(Y_{2:r:k}(n) \leq x) \approx e^{-\lambda^* \left(\frac{x}{cn^{1/ra}}\right)} \left(1 + \frac{r-1}{k-1} \lambda^* \left(\frac{x}{cn^{1/ra}}\right)\right).$$

In the following graphs we present the distribution of $Y_{m:r:k}(n)$ (adequately normalized) for $n = 50$ and $n = 500$ for 4 choices of the parameters m, k, r . The smooth curve displays the asymptotic distributions established above while the shaded histogram was obtained by the 100.000 simulated values of $Y_{m:r:k}(n)/(cn^{1/ra})$ (where $Y_i, i = 1, 2, \dots, n$ follow a Pareto distribution with $c = 1, a = 2$)

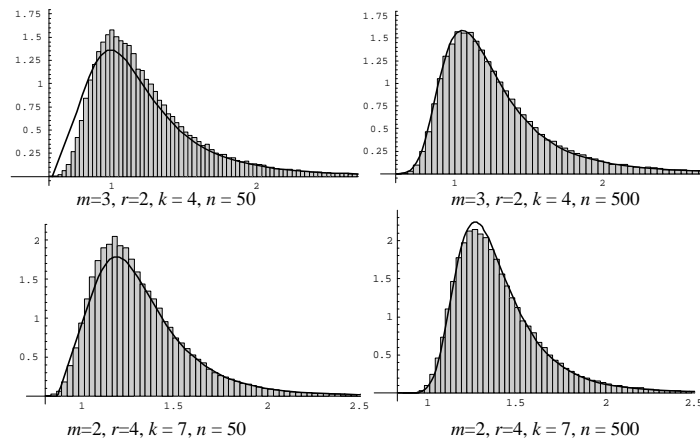


Figure 1.1: Exact (simulated) and approximate distribution for $Y_{m:r:k}$ for the Pareto distribution $F(x) = 1 - x^{-2}, x \geq 1$

b. Uniform Distribution. As a second example let us consider a sequence of random variables Y_1, Y_2, \dots which follow the uniform distribution

$$F(x) = x, \quad 0 < x < 1.$$

Since, $\bar{F} \in MDA(\Psi_a)$ with right end point $x_F = 1$ we may use the norming constants $c_n = x_F - F^{-1}(1 - n^{-1}) = n^{-1}, d_n = x_F = 1$ to gain the asymptotic results

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\frac{Y_{m:r:k}(n) - 1}{n^{-1/r}} \leq x\right) &= \lim_{n \rightarrow \infty} P(W_{n,k,r}(n^{-1/r}x + 1)) < m) \\ &= \sum_{i=0}^{m-1} f_{CP}(x; i), \quad x > 0 \end{aligned}$$

with the parameter $\lambda(x)$ of the compound Poisson distribution being given by (1.13).

In Figure 1.2 the exact distribution of $Y_{m:r:k}(n)$ (estimated by simulation) is compared to the approximate distribution gained by the last formula for the same set of choices for the parameters n, m, r, k as before (in order to achieve a better accuracy for the asymptotic results, formula (1.14) was used for the compound Poisson distribution parameters instead of (1.13)).

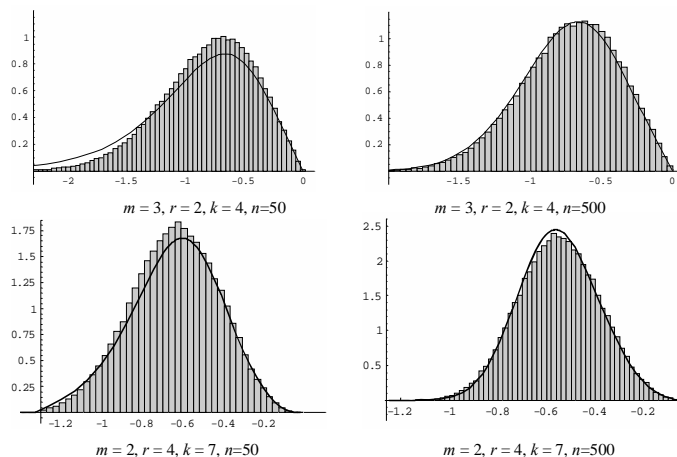


Figure 1.2: Exact (simulated) and approximate distribution for $Y_{m:r:k}$ for the Uniform distribution $F(x) = x, 0 < x < 1$

c. Normal and Exponential Distribution. Two typical examples of the distributions belonging to the maximum domain of attraction of the Gumbel distribution are the exponential with mean $1/\beta$ and the standard normal distribution. A set of appropriate norming constants is offered by (see e.g. Table 3.4.2 in Embrechts *et al.* (1997))

$$c_n = \beta^{-1}, \quad d_n = \beta^{-1} \ln n,$$

and

$$c_n = (2 \ln n)^{-1/2}, \quad d_n = (2 \ln n)^{1/2} - \frac{\ln 4\pi + \ln \ln n}{2(2 \ln n)^{1/2}}$$

respectively (the second pairs consists of a set of reasonable approximations for the norming constants). A direct application of Theorem 1.4.3 reveals that the asymptotic distribution of $Y_{m:r:k}(n)$, after carrying out a proper normalization, may be approximated by the aid of a compound Poisson with the parameter $\lambda(x)$ fetched from formula (1.15) (for a better approximation for finite n , we may use formula (1.16) instead the one given in (1.8)).

In Figures 1.3 and 1.4 a graphical comparison between the exact and asymptotic distributions of $Y_{m:r:k}(n)$ is carried out for exponential and normal sequences Y_1, Y_2, \dots respectively. The low quality of the approximation observed in Figure 1.4 should be attributed to the slow convergence of the distribution of the maximum of normal variables to the Gumbel distribution (which is due to the fact that the rate of convergence of $n\bar{\Phi}(c_n x + d_n)$ to e^{-x} is of order $O((\ln n)^{-1})$)

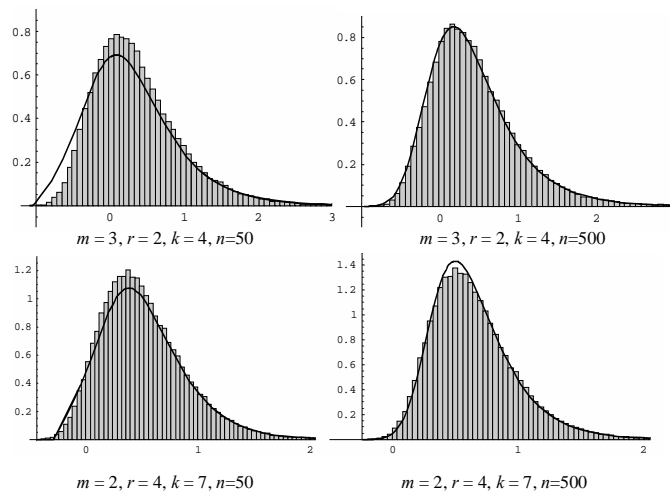


Figure 1.3: Exact (simulated) and approximate distribution for $Y_{m:r:k}$ for the Exponential distribution $F(x) = 1 - e^{-x}, x \geq 0$

In closing we mention that a series of results relating to minima (instead of maxima) and the asymptotic behaviour under the assumption that the underlying distribution belongs to a minimum domain of attraction of the three extreme type distributions could also be established. Since these outcomes follow immediately from the corresponding results established here by using $-Y_i$ in the place of Y_i we shall not pursue these topics here.

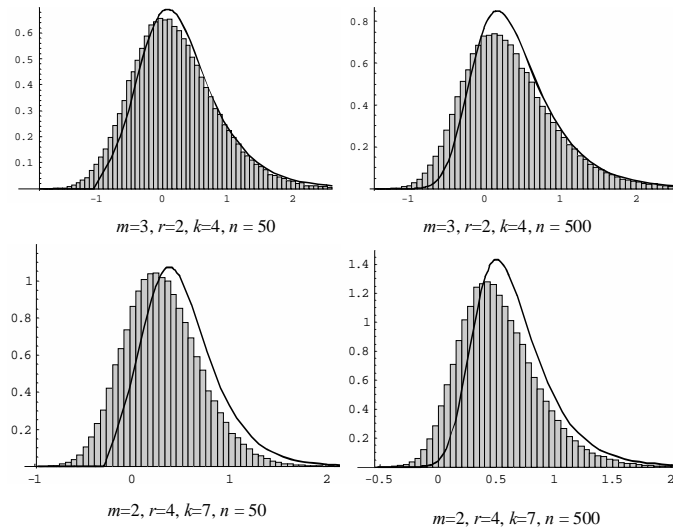


Figure 1.4: Exact (simulated) and approximate distribution for $Y_{m:r:k}$ for the (standard) Normal distribution $\Phi(x)$, $x \in \mathfrak{R}$

Acknowledgments Research of the last author was supported by the National Scholarship Foundation of Greece.

References

1. Arnold, B.C. (1985). Pareto distributions, In Encyclopedia of Statistical Sciences, S. Kotz, N.L. Johnson and C.B. Read (editors), 568-574. John Wiley & Sons, New York.
2. Arnold, B.C. and Balakrishnan, N. (1989). Relations, bounds, and approximations for order statistics. Springer, New York.
3. Arratia, R.L, Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. Statistical Science, 5, 403-423.

4. Arratia, R.L, Gordon, L. and Waterman, M. (1990). The Erdős-Rényi Law in distribution, for coin tossing and sequence matching. *Annals of Statistics*, 18, 539-570.
5. Balakrishnan, N. and Koutras, M.V. (2002). *Runs, scans and applications*. John Wiley & Sons, New York.
6. Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson approximation*. Clarendon Press, Oxford.
7. Boutsikas, M.V. and Koutras, M.V. (2001). Compound Poisson approximation for sums of dependent random variables. In Ch.A. Charalambides, M.V. Koutras, N. Balakrishnan (eds), *Probability and statistical models with applications*, 63-86. FL: Chapman & Hall, Boca Raton.
8. Boutsikas, M.V. and Koutras, M.V. (2002). Modeling claim exceedances over thresholds. *Insurance: Mathematics and Economics*, 30, 67-83.
9. Boutsikas, M.V. and Koutras, M.V. (2006). On the asymptotic distribution of the discrete scan statistic. *Journal of Applied Probability*, 43, 1137-1154.
10. Bowers, N.L., Gerber, H.U., Hickman, J., Jones, D.A. and Nesbitt, C.J. (1997). *Actuarial Mathematics*, 2nd Edition. The society of Actuaries, Illinois.
11. Chen, J. and Glaz, J. (1999). Approximations for the distribution and the moments of discrete scan statistics. In *Scan Statistics and Applications*, eds J. Glaz and N. Balakrishnan. Birkhäuser, Boston.
12. Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer-Verlag, London.
13. David, H.A. and Nagaraja, H.N. (2003). *Order Statistics* (third ed.). John Wiley & Sons, New York.
14. Dembo, A. and Karlin, S. (1992). Poisson approximations for r -scan processes. *Annals of Applied Probability*, 2, 329-357.
15. Deheuvels, P. and Devroye, L. (1987). Limit laws of Erdős-Rényi-Shepp type. *The Annals of Probability*, 15, 1363-1386.
16. Dudkiewicz, J. (1998). Compound Poisson approximation for extremes for moving minima in arrays of independent random variables. *Applicationes Mathematicae*, 25, 19-28.
17. Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modeling extremal events for insurance and finance*, Springer-Verlag, Berlin.

18. Erdős-Rényi (1970). On a new law of large numbers. *J. Analyse Math*, 23, 103-111.
19. Fu, J.C. (2001). Distribution of the scan statistic for a sequence of bistate trials. *Journal of Applied Probability*, 38, 908-916.
20. Fu, J.C. and Lou, W.Y.W. (2003). *Distribution theory of runs and patterns and its applications*. Word Scientific publishing, Singapore.
21. Glaz, J. and Balakrishnan, N. (eds.) (1999). *Scan Statistics and Applications*. Birkhäuser, Boston.
22. Glaz, J. and Naus, J. (1991). Tight bounds and approximations for scan statistic probabilities for discrete data. *The Annals of Applied Probability*, 1, 306-318.
23. Glaz, J., Naus, J. and Wallenstein, S. (2001). *Scan Statistics*. Springer Verlag, New York.
24. Goldstein, L. and Waterman, M. (1992). Poisson, compound Poisson and process approximations for testing statistical significance in sequence comparisons. *Bulletin of Mathematical Biology*, 54, 785-812.
25. Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions*, Vol. 1, John Wiley & Sons, New York.
26. Koutras, M.V. and Alexandrou, V.A. (1995). Runs, scans and urn model distributions: a unified Markov chain approach. *Ann. Inst. Statist. Math.*, 47, 743-766.
27. Kotz, S. and Nadarajah, S., (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.
28. Reiss, R.D. and Thomas, M. (1997). *Statistical analysis of extreme values*. Birkhäuser Verlag, Basel.