The Annals of Applied Probability
1992, Vol. 2, No. 2, 329–357

POISSON APPROXIMATIONS FOR r-SCAN PROCESSES

BY AMIR DEMBO AND SAMUEL KARLIN[1]

Stanford University

Let X_i be positive i.i.d. random variables (or more generally a uniformly mixing positive-valued ergodic stationary process). The r-scan process induced by $\{X_i\}$ is $R_i = \sum_{k=i}^{i+r-1} X_k$, $i = 1, 2, \ldots, n - r + 1$. Limiting distributions for the extremal order statistics among $\{R_i\}$ suitably normalized (and appropriate threshold values $a = a_n$ and $b = b_n$) are derived as a consequence of Poisson approximations to the Bernoulli sums $N^-(a) = \sum_{i=1}^{n+r-1} W_i^-(a)$ and $N^+(b) = \sum_{i=1}^{n-r+1} W_i^+(b)$, where $W_i^-(a)$ $[W_i^+(b)] = 1$ or 0 according as $R_i \le a$ $(R_i > b)$ occurs or not. Applications include limit theorems for r-spacings based on i.i.d. uniform $[0, 1]$ r.v.'s, for extremal r-spacings based on i.i.d. samples from a general density and for the r-scan process with a variable time horizon.

1. Introduction. The motivation of the paper stems from studies on inhomogeneities in long DNA sequences. The issues are relevant to the objectives of the human sequencing initiative (a multinational endeavor of much recent celebrity). Particular markers (in the language of DNA, e.g., restriction sites) are distributed along the length. It is of interest to evaluate the extent as distinguished from "chance" of extant segments along the length entailing excessive clumping, sparseness or regularity (very even spacings). A typical question might concern a long DNA sequence, say 66 million nucleotides long, containing $n = 1000$ markers and a mean distance between markers of about 66,000 nucleotides. A researcher looking through the whole sequence might observe $r = 5$ markers all falling within the same 19,800 nucleotides. Assuming that the positions of the 1000 markers are completely random, how likely is it that at least one such cluster would occur? Similar questions pertain to the sparseness of markers and other distributional properties. These questions can be approached in the following setting.

Consider a long length and a sequence of positive valued r.v.'s:

$$X_1, X_2, \ldots, X_n \tag{1.0}$$

corresponding to the successive distances between markers (X_1 is the distance to the first marker, X_2 between the first and second marker, etc.). In the simplest model X_i are i.i.d. but we shall also consider X_i generated in a Markov dependent fashion and even more generally as a strong ergodic

Received November 1990; revised February 1991.
[1]Supported in part by NSF Grant DMS-86-06244 and NIH Grants GM39907-02 and GM10452-26
AMS 1980 *subject classifications*. Primary 60F05, 60E20; secondary 60G50.
Key words and phrases. r-scans, r-spacings, extremal distributions, Poisson approximation.

stationary process. The sums

$$R_i = R_i^{(r)} = \sum_{j=i}^{i+r-1} X_j, \qquad i = 1, \ldots, n-r+1, \tag{1.1}$$

will be referred to as the *r-scan process*. In most applications r is fixed but for mathematical completeness, limit results are also developed with $r \uparrow \infty$ at an appropriately slower rate than n.

Consider the associated order statistics of $\{R_i\}_{i=1}^{n-r+1}$ denoted by

$$R_1^* \le R_2^* \le \cdots \le R_{n-r+1}^*$$

such that

$$m^{(r)} = R_1^* = \min_i \{R_i^{(r)}\} \quad \text{and} \quad M^{(r)} = R_{n-r+1}^* = \max_i \{R_i^{(r)}\}. \tag{1.2}$$

More generally, the r-scan kth minimum and maximum are, respectively,

$$m_k^{(r)} = R_k^*, \qquad M_k^{(r)} = R_{n-r-k+2}^*, \qquad k = 1, 2, 3, \ldots . \tag{1.3}$$

Asymptotic distributions for these extremal r.v.'s ($n \to \infty$) are of primary interest.

We need the following notation and terminology. The total variational distance between two random variables U and V with common state space is defined by

$$d(U,V) = \sup_A [\Pr\{U \in A\} - \Pr\{V \in A\}], \tag{1.4a}$$

and when U, V are nonnegative integer valued,

$$d(U,V) = \frac{1}{2} \sum_{j=0}^{\infty} |\Pr\{U = j\} - \Pr\{V = j\}|. \tag{1.4b}$$

The ascertainment of the limit distributions of the extremal variables of the r-scan process will ensue from analysis of the distributional properties of the count r.v.'s:

$$\begin{aligned} N^+(b) &= N_n^+(b) = \sum_{i=1}^{n-r+1} W_i^+(b), \\ N^-(a) &= N_n^-(a) = \sum_{i=1}^{n-r+1} W_i^-(a), \end{aligned} \tag{1.5}$$

where

$$W_i^-(a) = \begin{cases} 1, & \text{if } R_i \le a \\ 0, & \text{if } R_i > a \end{cases} \quad \text{and} \quad W_i^+(b) = \begin{cases} 1, & \text{if } R_i > b \\ 0, & \text{if } R_i \le b \end{cases}. \tag{1.6}$$

Obviously, when $N^-(a) = 0$ all r-scans exceed the level a and for $N^+(b) = 0$ no r-scans exceed the level b.

Poisson approximations to N^+ and N^- are as follows.

THEOREM 1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. with distribution function $F(x)$ and denote by $F_m(x)$ the distribution function of $\sum_{i=1}^m X_i$, the m-fold convolution of $F(x)$. Let Z_λ be a Poisson r.v. with parameter λ.*

Define

$$\lambda = (n - r + 1)F_r(a). \tag{1.7}$$

Then

$$\begin{aligned} d(N^-, Z_\lambda) &\le (1 - e^{-\lambda})\left[(2r - 1)F_r(a) + 2\sum_{m=1}^{r-1} F_m(a)\right] \\ &= (1 - e^{-\lambda})\delta(r, a), \end{aligned} \tag{1.8}$$

where $\delta(r, a)$ denotes the quantity in brackets.

The proof is given in Section 3.

Thus, for $F(x)$ continuous at $x = 0$ and r fixed with λ held fixed, obviously,

$$d(N^-, Z_\lambda) \to 0 \quad \text{as } n \to \infty \text{ with available rates of convergence.} \tag{1.9}$$

Because $\{X_i\}$ are positive random variables, $F_r(a) \le F_s(a)$ for $r > s$ and therefore the right-hand side of (1.8) is bounded by

$$4\sum_{m=1}^{r} F_m(a) \le 4M(a),$$

where $M(a) = \sum_{m=1}^{\infty} F_m(a)$ is the renewal function of the partial sum process generated by $S_m = \sum_{i=1}^m X_i$, $m = 1, 2, \ldots$. Thus the conclusion of (1.9) applies as long as $M(a) \to 0$ with n, a, r and λ related by (1.7). We discuss examples of this general formulation in Section 3.

With respect to $N^+(b)$, we have the following theorem.

THEOREM 2. *Under the conditions of Theorem 1, that is, $\{X_i\}$ are i.i.d. positive, define*

$$\mu = (n - r + 1)[1 - F_r(b)]. \tag{1.10}$$

Then

$$\begin{aligned} d(N^+, Z_\mu) &\le (1 - e^{-\mu})\left[(2r - 1)(1 - F_r(b)) \right.\\ &\qquad \left. + 2\sum_{m=1}^{r-1} \Pr\{R_{m+1} > b | R_1 > b\}\right] \\ &:= (1 - e^{-\mu})\varepsilon(r, b). \end{aligned} \tag{1.11}$$

For X_i bounded above by C, this quantity goes to 0 as b increases to C provided $F(\cdot)$ is continuous at C, the corresponding situation to Theorem 1.

It is not true that for all unbounded r.v.'s with continuous $F(x)$ the right-hand side in (1.11) goes to 0 as $b \uparrow \infty$. Distributions with heavy tails provide counterexamples; see Section 4. The fact that $d(N^+, Z_\mu) \to 0$ is sensitive to the tail behavior of $F(x)$. However, the following theorem applies.

THEOREM 3. *If for each fixed constant* $K > 0$,

$$\frac{1 - F(b - K)}{1 - F_2(b)} \to 0 \quad \text{as } b \to \infty, \tag{1.12}$$

then the right-hand side in (1.11) *goes to* 0 *for any fixed* r.

See Section 4 for an analysis of distribution models implying (1.12). The condition (1.12) is satisfied for the exponential density with any scale parameter and for any convolution of these but fails for $F(x) = x/(1 + x)$.

The dual variables of the extremal statistics (1.2) are of interest. Suppose an infinite sequence $X_1, X_2, \ldots$ is generated; see (1.0). Define, for fixed k,

$$T_k^- = \inf\{n;\, N_n^-(a) \geq k\} \tag{1.13}$$

and

$$T_k^+ = \inf\{n;\, N_n^+(b) \geq k\}. \tag{1.14}$$

From their definitions the identity of the following events are immediate:

$$\{T_k^- > n\} = \{N_n^-(a) < k\} = \{R_k^* > a\}, \tag{1.15a}$$

$$\{T_k^+ > n\} = \{N_n^+(b) < k\} = \{R_{n-r-k+2}^* \leq b\} \tag{1.15b}$$

and limit relations for T_k^- and T_k^+ ensue readily from those of N^- and N^+; see Sections 3 and 4.

The example of r-spacings sampled from a uniform distribution is of practical interest. Consider $(n - 1)$ i.i.d. samples from a uniform distribution on $[0, 1]$ and form the consecutive spacings $U_1, U_2, \ldots, U_n$ so that $U_i \geq 0$, $\sum_{i=1}^n U_i = 1$. It is well known [e.g., Karlin and Taylor (1981), Chapter 13, exercise 10, page 127] that the joint distributions

$$\{SU_1, SU_2, \ldots, SU_n\}, \qquad \{X_1, X_2, \ldots, X_n\} \tag{1.16}$$

are equivalent, where X_i are i.i.d. exponentially distributed and S independent of U_i is distributed as a gamma random variable (α, β) with $\alpha = n$ and unit scale parameter, $\beta = 1$. Obviously, the distributional properties of the r-spacings $\{U_i + \cdots + U_{i+r-1}\}_{i=1}^{n-r+1}$ can be reduced to those of the r-scan process based on $\{X_i\}$, where X_i are i.i.d. exponentially distributed. Using Berry–Esseen estimates of the normal approximation to a gamma$(n, 1)$ distribution and the results of Theorem 1 and 2, we deduce the following theorem.

THEOREM 4. *Let $W^-(a)$ be the count of r-spacings not exceeding a from a sample of $n-1$ i.i.d. random variables uniform $[0,1]$ and define*

$$\lambda = (n-r+1)\left[1 - e^{-na}\sum_{i=0}^{r-1}\frac{(an)^i}{i!}\right]. \tag{1.17}$$

Then

$$d(W^-(a), Z_\lambda) \le 4an(1-e^{-\lambda}) + O\left(\sqrt{\frac{\log n}{n}}\right).$$

A corresponding formula obtains for $d(W^+(b), Z_\mu)$; see Section 5.

Throughout the paper all log terms are natural logarithms.

Consider the question stated at the start. In this example, $r = 5$, $n = 1000$ leading to $a = 3 \times 10^{-4}$, $\lambda \simeq 0.02$ and $d(W^-(a), Z_\lambda) \le 0.025 + O(0.05)$. Then $P(W^-(a) \ge 1) \le 0.1$ [this estimate can be improved by sharpening the error term $O(\sqrt{\log n/n})$].

In concrete terms as a consequence of Theorem 4, we have for the kth minimal and maximal extremal r-spacings of uniform i.i.d. r.v.'s:

$$\lim_{n\to\infty} \Pr\left\{m_k^{(r)} \ge \frac{x}{n^{1+1/r}}\right\} = e^{-x^r/r!}\sum_{i=0}^{k-1}\left(\frac{x^r}{r!}\right)^i\frac{1}{i!} \tag{1.18}$$

and

$$\begin{aligned}\lim_{n\to\infty} \Pr\left\{M_k^{(r)} \le \frac{1}{n}[\ln n + (r-1)\ln\ln n + x]\right\} \\ = \exp\left\{\frac{-e^{-x}}{(r-1)!}\right\}\left(\sum_{i=0}^{k-1}\left(\frac{e^{-x}}{(r-1)!}\right)^i\frac{1}{i!}\right).\end{aligned} \tag{1.19}$$

Without error estimates, the result (1.18) for $k = 1$ is due to Cressie (1977) and (1.19) for $k = 1$ is due to Holst (1980). The above limit relations are generalized as follows. Consider $n-1$ independent samples from a continuous density function $f(x)$ on $[0,1]$. Let $U_1, U_2, \ldots, U_n$ be the corresponding spacings induced from the sample, and the sums

$$\tilde{R}_i^{(r)} = \sum_{k=i}^{r+i-1} U_k, \qquad i = 1, \ldots, n-r+1,$$

be the r-spacings process. Generalizing with $\tilde{m}^{(r)} = \min_i \tilde{R}_i^{(r)}$, we get

$$\lim_{n\to\infty} \Pr\left\{\tilde{m}^{(r)} \ge \frac{x}{n^{1+1/r}}\right\} = \exp\left\{-\frac{x^r}{r!}\int_0^1 [f(\xi)]^{r+1}\,d\xi\right\}. \tag{1.20}$$

The limiting maximum $\tilde{M}^{(r)}$ depends only on the neighborhoods of minimum points of the density $f(x)$; see Section 8.

The Poisson approximation laws of this paper rely on the powerful Chen–Stein method which also provides rates of convergence. We use the elegant version set forth in Arratia, Goldstein and Gordon (1989). Alternative formulations exploiting coupling arguments and other applications are presented in a series of papers by Barbour, Holst, Janson and Hall among others [see Barbour and Holst (1989) and references therein]. All these methods are beneficial in supplementary ways. For further applications pertinent to biomolecular sequences, see Arratia, Goldstein and Gordon (1990), Karlin and Leung (1991) and Karlin and Macken (1991).

The organization of the paper is as follows: Section 2 reviews preliminaries on the variational distance for sets of distributions. Section 3 presents the proof of Theorem 1 with a number of examples. Theorem 2 is proved in Section 4. Several ancillary results of independent interest comparing tail behavior of distributions and convolutions are set forth. Limit theorems for extremal r-spacings based on i.i.d. samples from a uniform distribution on $[0, 1]$ with rates of convergence are proved in Section 5. The corresponding r-spacings based on i.i.d. observations following a general density are presented in Section 8. Corresponding r-scan limit theorems in continuous time are considered in Section 6. Extremal statistics of single spacings ($r = 1$) based on samples from a general density are elaborated in Deheuvels (1986); see also Deheuvels and Devroye (1987) for some limit laws of iterated logarithm type.

Asymptotic statistics like our scan processes have led Godbole (1990) in the context of standard i.i.d. binomial events (N, p) to determine the limit distribution (Poisson λ) of the number of runs of length k (k fixed) where $N \to \infty$, $p \to 0$ maintaining $\lambda = Np^k$ constant. This result is a direct corollary of Theorem 1. Godbole also considers the case of Markov dependence between successive trials. This also follows as an example of our general treatment of Section 7.

A different class of scan statistics was investigated extensively by many authors including Naus, Glaz, Wallenstein and Neff [see the bibliographic compilation on this subject, Naus (1979)]. Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. observations from a uniform distribution on the unit interval. The scan statistic generally refers to the maximal number of events within a fixed size window $(t,t + w)$, t traversing 0 to $1 - w$. The scan statistic has also been formulated for i.i.d. Bernoulli variables and Poisson processes. Dual variables are based on the first time until k events happen. These considerations relate to our minimal k-scan statistics. Multiple coverage of the line, studied, for example, in Glaz and Naus (1979), corresponds to the distribution of $M_1^{(r)}$; see (1.19). Some studies have also been done in higher dimensions, for example, Melzak (1979) and Janson (1987).

Motivations of the scan statistics relate to characterizations of clusters of disease in time, generalized birthday proximities and the kth nearest-neighbor problems. Early work on the scan statistics is mostly based on the Karlin–McGregor theorems [Karlin and McGregor (1959)] of coincidence probabilities

in Markov processes [generalizations in Karlin (1988)]. Recent work tends to focus more on finding computationally tractable bounds and approximations, for example, Glaz (1989), Naus (1982) and Wallenstein and Neff (1987).

2. Preliminaries. The Chen–Stein method. The Chen–Stein method provides error estimates for the Poisson approximation of a sum of (dependent) Bernoulli random variables. These estimates typically involve only the first two moments of the sum in question. Here we adopt the Arratia, Goldstein and Gordon (1989) formulation of the Chen–Stein method [for other versions, see Chen (1975), Barbour and Hall (1984), Barbour and Holst (1989) and Holst and Janson (1990)].

LEMMA 2.1 [Arratia, Goldstein and Gordon (1989)]. *Let W_α be Bernoulli (p_α) random variables and $W = \sum_{\alpha \in I} W_\alpha$, where I is a finite or countable index set. Let $d(W, Z_\lambda)$ denote the total variation distance between the distribution of W and Z_λ:*

$$
\begin{aligned}
d(U,V) &= \sup_A \left(\Pr\{U \in A\} - \Pr\{V \in A\}\right) \\
&= \frac{1}{2} \sum_{k=0}^{\infty} \left|\Pr\{U = k\} - \Pr\{V = k\}\right|.
\end{aligned} \tag{2.1}
$$

Then

$$
d(W, Z_\lambda) \le (b_1 + b_2)\left(\frac{1 - e^{-\lambda}}{\lambda}\right) + b_3 \min\left(1, \frac{\sqrt{2}}{\sqrt{\lambda}}\right), \tag{2.2}
$$

where

$$
\begin{aligned}
\lambda &= \sum_{\alpha \in I} p_\alpha, \\
b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta, \\
b_2 &= \sum_{\alpha \in I} \sum_{\substack{\beta \in B_\alpha \\ \beta \neq \alpha}} E\left[W_\alpha W_\beta\right], \\
b_3 &= \sum_{\alpha \in I} E\left[\left|E(W_\alpha) - p_\alpha\right| \{W_\gamma\}_{\gamma \notin B_\alpha}\right],
\end{aligned} \tag{2.3}
$$

and $\{B_\alpha\}$ is any family of subsets indexed by I.

In our applications the sets B_α are usually specified such that W_α is independent of $\{W_\gamma\}_{\gamma \notin B_\alpha}$ and therefore $b_3 = 0$. Thus when B_α is sparse relative to I and $E[W_\beta | W_\alpha = 1]|_{\beta \neq \alpha}$ are small enough entailing both b_1 and b_2 small, then the distribution of W is approximately Poisson.

We summarize below several elementary properties of the total variation distance which are referenced in the sequel. For any three r.v.'s U, V, W,

$$d(U,W) \le d(U,V) + d(V,W). \tag{2.4}$$

From the definition (2.1), obviously,

$$|Pr\{U < k\} - \Pr\{V < k\}| \le d(U,V). \tag{2.5}$$

Coupling U and V in the same probability space, we have

$$d(U,V) \le \Pr\{U \ne V\} \le 1. \tag{2.6}$$

In particular, for two Poisson r.v.'s Z_λ and Z_μ of parameters λ and μ, respectively, since for $\lambda > \mu$, $Z_\lambda = Z_\mu + Z_{\lambda-\mu}$, Z_μ and $Z_{\lambda-\mu}$ independent, then (2.6) yields

$$d(Z_\lambda, Z_\mu) \le \Pr\{Z_\lambda \ne Z_\mu\} = \Pr\{Z_{|\lambda-\mu|} \ne 0\} = 1 - e^{-|\lambda-\mu|} \le |\lambda - \mu|. \tag{2.7}$$

It is elementary that

$$d(U,V) \le E_X[d((U|X),(V|X))] \tag{2.8}$$

for any three r.v.'s X, U, V. By induction on m, if $\{U_i\}_{i=1}^m$ are independent and $\{V_i\}_{i=1}^m$ are independent (straightforward for $m = 2$), then

$$d\left(\sum_{i=1}^m U_i, \sum_{i=1}^m V_i\right) \le \sum_{i=1}^m d(U_i, V_i). \tag{2.9}$$

3. Extremal values of minimal r-scans of i.i.d. positive r.v.'s. The formulation is given in Section 1. We use the notation of Theorem 1 and especially (1.5). Henceforth, in this section, we suppress the superscript on $W_\alpha^-(a)$ and write W_α.

PROOF OF THEOREM 1. We merely define B_α for the case at hand and calculate the constants b_1, b_2 and b_3 as required in Lemma 2.1. For each index α, let $B_\alpha = \{\beta, |\beta - \alpha| < r\}$. Certainly,

$$b_3 = 0 \quad \text{since } R_\alpha^{(r)} \text{ is plainly independent of } R_\gamma^{(r)} \text{ for } \gamma \notin B_\alpha. \tag{3.1}$$

Clearly,

$$\lambda = \sum_{\alpha=1}^{n-r+1} E[W_\alpha^-(a)] = (n - r + 1)F_r(a) \tag{3.2}$$

as prescribed in (1.7). Now, by stationarity,

$$\begin{aligned} b_1 &= \sum_{\alpha=1}^{n-r+1} |B_\alpha|[F_r(a)]^2 \le [F_r(a)]^2(2r-1)(n-r+1) \\ &= (2r-1)\lambda F_r(a). \end{aligned} \tag{3.3}$$

Observe that $E[W_\alpha W_\beta] = F_r(a)\Pr\{R_\beta \le a | R_\alpha \le a\}$. Therefore, by stationarity (i.i.d.),

$$b_2 = \sum_{\alpha=1}^{n-r+1} \sum_{\substack{\beta \ne \alpha \\ \beta \in B_\alpha}} E\big[W_\alpha W_\beta\big] \le (n-r+1)F_r(a)2\sum_{k=2}^{r} \Pr\{R_k \le a | R_1 \le a\}$$

$$= 2\lambda \sum_{k=2}^{r} \Pr\{R_k \le a | R_1 \le a\}, \tag{3.4}$$

where the inequality is due to end effects.

The conjunction of (3.1), (3.3) and (3.4) into (2.2) of Lemma 2.1 produces the estimate

$$d(N_n^-(a), Z_\lambda) \le (1-e^{-\lambda})\left[(2r-1)F_r(a) + 2\sum_{k=2}^{r} \Pr\{R_k \le a | R_1 \le a\}\right]. \tag{3.5}$$

But since $\{X_i\}$ are positive r.v.'s,

$$\Pr\{R_k \le a | R_1 \le a\} \le \Pr\left\{\sum_{i=r+1}^{r+k-1} X_i \le a\right\},$$

we can bound the right-hand side of (3.5) by

$$(1-e^{-\lambda})\delta(r,a), \tag{3.6}$$

$\delta(r,a)$ defined in (1.8). The proof of Theorem 1 is complete. □

REMARK. The r-scans of $\{X_i\}_1^n$ can be placed on a circle, treating therefore $R_\alpha = \sum_\alpha^{\alpha+r-1} X_i$, $\alpha = 1, \dots, n$ ($\alpha + r - 1$ reduced modulo n), and Theorem 1 manifestly applies with $\lambda = nF_r(a)$ provided $n > 2r$.

For the dual variable T_k^- [see (1.13) and (1.15a)], the following corollary ensues.

COROLLARY 3.1. *Under the conditions of Theorem* 1,

$$|\Pr\{R_k^* > a\} - \Pr\{Z_\lambda < k\}| < (1-e^{-\lambda})\delta(r,a) \tag{3.7}$$

and equivalently

$$|\Pr\{T_k^- > n\} - \Pr\{Z_\lambda < k\}| < (1-e^{-\lambda})\delta(r,a). \tag{3.8}$$

Because $\{X_i\}$ are positive i.i.d. r.v.'s, manifestly $F_k(a) \le F_l(a)$ for all $k > l$ and all $a > 0$. Consequently,

$$\delta(r,a) = (2r-1)F_r(a) + 2\sum_{m=1}^{r-1} F_m(a) \le 4\sum_{m=1}^{r} F_m(a) \le 4rF_1(a). \tag{3.9}$$

For r fixed and $F(x)$ continuous at 0,

$$\delta(r,a) \le 4rF_1(a) \to 0 \quad \text{as } a \to 0, \tag{3.10}$$

and the Poisson approximation holds with rate of convergence faster than $4rF(a)$. Observe

$$\delta(r,a) \le 4\sum_{m=1}^{r} F_m(a) \le 4M(a), \tag{3.11}$$

where $M(a)$ is the renewal function of the sum process $\{\sum_{i=1}^{m} X_i\}$. When $F(x)$ is continuous at 0, $M(a)\downarrow 0$ as $a \downarrow 0$ [see, e.g., Karlin and Taylor (1975), page 181].

COROLLARY 3.2. *If λ as determined in* (1.7), *increases to ∞ as $a \to 0$, then $(N^-(a) - \lambda)/\sqrt{\lambda}$ tends to an asymptotically standard normal distribution.*

This follows from the bound (3.11) coupled to the fact that Z_λ is asymptotically normal.

In the foregoing $N^-(a)$ ascertains the aggregate counts among all r-scans obeying $R_i \le a$. It is of interest (see the example at the close of this section) to confine the count to nonoverlapping r-scans at $i = 2$ and proceeding sequentially to $n - r + 1$, such that $W_i^-(a)$ [see (1.6)] is changed to 0 when $\sum_{j=1}^{r-1} W_{i-j}^-(a) > 0$ where $W_k^-(a) = 0$ for $k \le 0$. Let $\hat{W}_i(a)$ be the corresponding counts and define $\hat{N}^-(a) = \sum_{i=1}^{n-r+1} \hat{W}_i^-(a)$. The Chen–Stein method is not directly applicable to $\hat{N}^-(a)$ because of the long-range dependence among the $\hat{W}_i^-(a)$. However, with λ defined in (1.7) we apply Theorem 1, (2.4) and (2.6) to deduce

$$d\big(\hat{N}^-, Z_\lambda\big) \le d(N^-, Z_\lambda) + d(\hat{N}^-, N^-) \le (1 - e^{-\lambda})\delta(r,a) + \Pr\{\hat{N}^- \ne N^-\}.$$

Invoking the union of events bound and the inherent stationarity yields

$$\begin{aligned}\Pr\{\hat{N}^- \ne N^-\} &\le \sum_{i=1}^{n-r+1} \Pr\big\{W_i^-(a) \ne \hat{W}_i(a)\big\} \\ &\le 2(n - r + 1)\sum_{j=2}^{r} \Pr\{R_1 \le a,\, R_j \le a\} \\ &\le 2\lambda \sum_{j=1}^{r-1} F_j(a) \le 2\lambda M(a).\end{aligned}$$

We can summarize the foregoing analysis as follows.

COROLLARY 3.3. *For large n, the (nonoverlapping) count of r-scans $\hat{N}^-(a)$ defined above is distributed Poisson (parameter λ) with variational error*

$$d\big(\hat{N}^-, Z_\lambda\big) \le \big[4(1 - e^{-\lambda}) + 2\lambda\big]M(a).$$

Some concrete examples. If the density $f(x)$ of $F(x)$ is positive and continuous for an interval including the origin [say $f(0) = 1$], then $F_r(a)$ behaves like $a^r/r!$ for a small. The identity (1.7) yields asymptotically the equation

$$a \sim (r!)^{1/r}\left(\frac{\lambda}{n}\right)^{1/r}. \tag{3.12}$$

If $r \to \infty$, then $(r!)^{1/r} \sim re^{-1}$. Consider $\lambda \leq e^{cr}$ for some constant c. In this event $\lambda^{1/r}$ is bounded. Under these assumptions the order growth is

$$a = O\left(\frac{r}{n^{1/r}}\right).$$

Now for

$$r = O\left(\frac{\ln n}{\ln \ln n}\right) \tag{3.13}$$

then $r/n^{1/r} \sim 1/\ln \ln n \to 0$ and $a \to 0$. To summarize: If $\lambda \leq e^{cr}$ and $r \leq K \ln n/\ln \ln n$ and a determined as in (3.12), then

$$d(N^-(a), Z_\lambda) = O\left(\frac{1}{\ln \ln n}\right)$$

and if $\lambda \uparrow \infty$, then

$$\frac{N^-(a) - \lambda}{\sqrt{\lambda}} \to_{\text{law}} N(0,1)$$

(standard normal distribution).

For λ bounded $N^-(a)$ has a limiting Poisson (λ) distribution. Along these lines it is interesting that for $r = K \log n$ and λ bounded the limiting distribution $N^-(a)$ is not pure Poisson but compound Poisson [see Karlin and Ost (1987)]; actually for a success run example the maximal extremal r-scan is a Poisson random variable compounded with a geometric distribution.

For r bounded, the limit law $d(N^-(a), Z_\lambda) \to 0$ prevails as long as $\lambda/n \to 0$. For bounded r and λ, the rate of convergence is always $F_1(a)$. In general, the rate of convergence is $O(M(a))$.

Count of Bernoulli success runs of fixed length r. Consider $X_1, X_2, \ldots, X_n$ i.i.d. Bernoulli $(1-p)$ with $\lambda = np^r$ fixed as $n \to \infty$, $p \to 0$. Let $a = 1/2$. Then $N^-(a)$ counts the number of (overlapping) success runs of length r (here $X_i = 0$ indicates a success while $X_i = 1$ reflects a failure). For this model the number of *nonoverlapping* success runs $\hat{N}^-(a)$ was considered by Godbole (1990) whose main result is simply a special case of Corollary 3.3.

4. Significant maximal valued *r*-scans. In this section we describe cases for the Poisson approximation to $N^+(b) = \sum_{\alpha=1}^{n-r+1} W_\alpha^+(b)$ defined in (1.5). In the i.i.d. model the proof of Theorem 2 (see Section 1) paraphrases the analysis of Theorem 1, replacing $F_r(a)$ by $1 - F_r(b)$, mutatis mutandis. For

ready reference we restate the bound:

$$\mu = (n - r + 1)\big[1 - F_r(b)\big].$$

Then [cf. (1.11)]

$$d\big(N^+(b), Z_\mu\big) \le (1 - e^{-\mu})\Bigg[(2r-1)(1 - F_r(b)) + 2\sum_{m=1}^{r-1} \Pr\{R_{m+1} > b | R_1 > b\}\Bigg] = (1 - e^{-\mu})\varepsilon(r, b). \tag{4.1}$$

The analog of Corollary 3.1 is the following corollary.

COROLLARY 4.1.

$$|\Pr\{R^*_{n-r+2-k} \le b\} - \Pr\{Z_\mu < k\}| \le (1 - e^{-\mu})\varepsilon(r, b), \tag{4.2a}$$

$$|\Pr\{T_k^+ > n\} - \Pr\{Z_\mu < k\}| \le (1 - e^{-\mu})\varepsilon(r, b). \tag{4.2b}$$

REMARK 4.1. When the random variables X_i are bounded by $c < \infty$ [i.e., $F(c) = 1$], we define $\tilde{F}_m(x) = 1 - F_m(mc - x)$, the convolution distribution corresponding to the partial sums of $\tilde{X}_i = c - X_i$. Then with $\tilde{R}_i = rc - R_i$, $i = 1, \ldots, n - r + 1$, and setting $a = rc - b$, $\mu = \tilde{\lambda}$, where $N^+ = \tilde{N}^-$ and $\varepsilon(r, b) = \tilde{\delta}(r, a)$.

We investigate next conditions that ensure $\varepsilon(r, b) \to 0$ with $b \to \infty$ in the unbounded case, that is, $F(c) < 1$ for all $c > 0$. Parenthetically, when $F_1(x)$ has heavy tails, a Poisson approximation to $N^+(b)$ does not hold, as described below.

In estimating $\varepsilon(r, b)$ of (4.1) the following lemma is germane.

LEMMA 4.1. *Let $X_1, X_2, \ldots$ be i.i.d. Then*

$$\Pr\{X_{k+1} + \cdots + X_{r+k} > b | X_1 + \cdots + X_r > b\} \le \Pr\{X_k + \cdots + X_{r+k-1} > b | X_1 + \cdots + X_r > b\}. \tag{4.3}$$

That is, $c_k = \Pr\{X_k + \cdots + X_{r+k-1} > b | X_1 + \cdots + X_r > b\}$ is decreasing (not necessarily strictly) in k.

PROOF. The inequality (4.3) is equivalent to

$$\Pr\{X_{k+1} + \cdots + X_{r+k} > b,\ X_1 + \cdots + X_r > b\} \le \Pr\{X_k + \cdots + X_{r+k-1} > b,\ X_1 + \cdots + X_r > b\}. \tag{4.4}$$

The inequality (4.4) above, in more general form allowing a general nonnegative function ϕ, can be written in the form $d_k \geq d_{k+1}$, where for $k \leq r$,

$$d_k = E\big[\phi(X_k + \cdots + X_{r+k-1})\phi(X_1 + \cdots + X_r)\big]$$

$$= E_{X_k,\ldots,X_r}\big[E_{X_1,\ldots,X_{k-1}}\phi(X_1 + X_2 + \cdots + X_r)\big]^2$$

due to the i.i.d. character of $\{X_i\}$. Applying the Schwarz inequality on the variable X_k gives

$$d_k \geq E_{X_{k+1},\ldots,X_r}\big[E_{X_1,\ldots,X_k}\phi(X_1 + \cdots + X_r)\big]^2 = d_{k+1} \tag{4.5}$$

and the lemma is proved noting when $k > r$, c_k is the constant $1 - F_r(b)$. □

The inequality (4.5) is reminescent of a hierarchy of conditional variance inequalities developed in Karlin and Rinott (1982) in relation to jackknifing statistical estimation.

By virtue of Lemma 4.1 we have

$$\Pr\{R_m > b | R_1 > b\} \leq \Pr\{R_2 > b | R_1 > b\}, \qquad m \geq 2,$$

and the right-hand side of (4.1) is bounded above by

$$d\big(N^+(b), Z_\mu\big) \leq 4r \Pr\{R_2 > b | R_1 > b\}. \tag{4.6}$$

Conditioning on $Y = X_2 + \cdots + X_r$, we have

$$\Pr\{R_2 > b | R_1 > b\} = \frac{1 - F_{r-1}(b)}{1 - F_r(b)} + \frac{\int_0^b [1 - F(b - \xi)]^2 \, dF_{r-1}(\xi)}{1 - F_r(b)}. \tag{4.7}$$

Suppose for each K,

$$\lim_{b\to\infty} \frac{1 - F_{r-1}(b - K)}{1 - F_r(b)} = 0 \tag{4.8}$$

and r is fixed; then $d(N^+(b), Z_\mu) \to 0$ as asserted in Theorem 3 (see Section 1). Indeed, choose K large enough so that $4r[1 - F(K)] \leq \eta$ (η arbitrarily small). Then the right-hand side of (4.6) is estimated above by

$$4r\left[\frac{1 - F_{r-1}(b)}{1 - F_r(b)} + 1 - F(K) + \frac{F_{r-1}(b) - F_{r-1}(b - K)}{1 - F_r(b)}\right].$$

Under the force of (4.8) we have

$$\limsup_{b\to\infty} d\big(N^+(b), Z_\mu\big) \leq \eta, \tag{4.9}$$

and thereby Theorem 3 is proved under the condition (4.8). The full statement of Theorem 3 follows by Lemma 4.2 proved below.

Conditions for the validity of (4.8). We start with X_i distributed exponential (1) for which

$$r\left\{[1 - F(K)] + \frac{1 - F_{r-1}(b - K)}{1 - F_r(b)}\right\} = O\left(re^{-K} + \frac{r^2 e^K}{b}\right). \tag{4.10}$$

Thus, for r fixed, $K \uparrow \infty$ and $b \uparrow \infty$, such that $e^K/b \to 0$, $d(N^+, Z_\mu)$ has the error given by the right-hand side of (4.10). For any r determine $y = e^K > 0$ to minimize $r/y + r^2 y/b$, yielding $y = \sqrt{b/r}$ and error term $O(r^{3/2}/\sqrt{b})$. Alternatively, for the exponential (scale parameter 1) by a direct computation

$$\begin{aligned} \Pr(R_2 > b | R_1 > b) &= \frac{\sum_{i=0}^{r-2} \frac{b^i}{i!}\left[1 + (-1)^{r-i}\right] + (-1)^{r+1} e^{-b}}{\sum_{i=0}^{r-1} \frac{b^i}{i!}} \\ &\le e^{-b} + \frac{2r}{b}. \end{aligned} \tag{4.11}$$

Thus any scan length $r < o(\sqrt{b})$, $b \to \infty$, yields a Poisson approximation with error of $O(r^2/b)$ in the exponential case. If $b \uparrow \infty$ and $ne^{-b}b^{r-1}/(r-1)! \to \infty$, a central limit theorem applies to the random variables $N^+(b)$, $b \to \infty$.

By using estimates of large deviation theory, we can sharpen the error term of (4.11) for the exponential example. We have for any $K > 0$ and for $r > m$ the estimate

$$\Pr\{R_{m+1} > b | R_1 > b\} \le 1 - F_m(K) + \frac{1 - F_{r-m}(b - K)}{1 - F_r(b)}. \tag{4.12}$$

We specify $K = mt$, $b = r\tau$ with $t > 1$ and $\tau > 1$. By Chernoff's bound (1952),

$$1 - F_m(mt) = \Pr\left\{\sum_{i=1}^{m} X_i \ge mt\right\} \le e^{-mI(t)}$$

and $I(t) = \sup_{\theta > 0}\{\theta t - \log E[e^{\theta X_1}]\} = t - 1 - lnt$. It follows that

$$\sum_{m=1}^{r} [1 - F_m(mt)] \le \sum_{m=1}^{\infty} e^{-mI(t)} = \frac{1}{e^{I(t)} - 1}.$$

We also have

$$\frac{1 - F_{r-m}(b - K)}{1 - F_r(b)} = e^K \frac{\sum_{k=0}^{r-m-1} \frac{(b-K)^k}{k!}}{\sum_{k=0}^{r-1} \frac{b^k}{k!}} \le e^K \left(\frac{r}{b}\right)^m.$$

Therefore, for $\tau > t > 1$, $b = \tau r$ and for $e^t < \tau$,

$$\sum_{m=1}^{r-1} \frac{1 - F_{r-m}(r\tau - mt)}{1 - F_r(r\tau)} \leq \sum_{m=1}^{\infty} \left(\frac{e^t}{\tau}\right)^m = \frac{1}{\tau e^{-t} - 1}.$$

Thus in the exponential case, we have found

$$\varepsilon(r, r\tau) \leq \frac{4}{(e^t/et - 1)} + \frac{4}{\tau e^{-t} - 1}. \tag{4.13}$$

Choose t such that $e^{2t} = \tau et$, making the two terms on the right-hand side of (4.13) equal. This produces the bound

$$\varepsilon(r, r\tau) \approx O\left(\frac{e^t}{\tau}\right) = O\left(\sqrt{\frac{\ln \tau}{\tau}}\right), \qquad \tau = \frac{b}{r},$$

provided $b/r \uparrow \infty$.

Return now to the general distribution $F_1(x)$ of a positive variable.

LEMMA 4.2. *If for each fixed constant* $K \geq 0$,

$$\lim_{b\to\infty} \frac{1 - F_1(b - K)}{1 - F_2(b)} = 0, \tag{4.14}$$

then

$$\lim_{b\to\infty} \frac{1 - F_{r-1}(b - K)}{1 - F_r(b)} = 0. \tag{4.15}$$

PROOF. We advance the induction from r to $r + 1$. For this objective the fact $F_k(x) \geq F_{k+1}(x)$ for all $k = 1, 2, \ldots$ and $x \geq 0$ is relevant. Consider

$$\begin{aligned}\frac{1 - F_r(b - K)}{1 - F_{r+1}(b)} &= \frac{1 - F_{r-1}(b - K) + F_{r-1}(b - K) - F_r(b - K)}{1 - F_{r+1}(b)} \\ &\leq \frac{1 - F_{r-1}(b - K)}{1 - F_r(b)} + \frac{\int_0^{b-K} [1 - F_1(b - K - \xi)]\, dF_{r-1}(\xi)}{1 - F_{r+1}(b)}.\end{aligned}$$

The first term goes to 0 owing to the induction hypothesis.

The second term is estimated above by

$$\begin{aligned}&\frac{\int_0^{b-K-L} \dfrac{1 - F_1(b - K - \xi)}{1 - F_2(b - \xi)} [1 - F_2(b - \xi)]\, dF_{r-1}(\xi)}{\int_0^{b-K-L} [1 - F_2(b - \xi)]\, dF_{r-1}(\xi)} \\ &\quad + \frac{1 - F_{r-1}(b - K - L)}{1 - F_r(b)}.\end{aligned} \tag{4.16}$$

With K fixed, choose L and then b sufficiently large after L is fixed, such that

for all $\xi \le b - K - L$,

$$\frac{1 - F_1(b - K - \xi)}{1 - F_2(b - \xi)} \le \sup_{\tilde{b} \ge K + L} \frac{1 - F_1(\tilde{b} - K)}{1 - F_2(\tilde{b})} \le \varepsilon.$$

It follows by the induction hypothesis that the quantity (4.16) is bounded by ε as $b \to \infty$ and ε is arbitrarily small. Clearly, the induction is established. □

The same kind of manipulations proves the following theorem.

THEOREM 4.1. *If the condition* (4.14) *holds for the positive distributions* $G(x)$ *and* $H(x)$, *then it also holds for the convolution of* G *and* H, $F = G * H$.

COROLLARY 4.2. *Condition* (4.14) *holds for any distribution function* $F(x)$ *which is a finite or infinite convolution of exponentials of any scale parameters.*

PROOF. We examine

$$\frac{(H * G)(b) - (H * G)(b - K)}{1 - (H_2 * G_2)(b)}$$

$$= \frac{\left[\int_0^{b-L} + \int_{b-L}^{b}\right][G(b - \xi) - G(b - K - \xi)]\, dH(\xi)}{1 - (H_2 * G_2)(b)} = \gamma_1 + \gamma_2.$$

Clearly,

$$\gamma_2 \le \frac{H(b) - H(b - L)}{1 - H_2(b)}$$

because $(H_2 * G_2)(b) \le H_2(b)$ and paralleling the analysis of (4.16), we have

$$\gamma_1 \le \max_{c \ge L}\left[\frac{G(c) - G(c - K)}{1 - G_2(c)}\right].$$

Choose L large enough and fixed such that $\gamma_1 \le \varepsilon$, then send $b \to \infty$ to assure $\gamma_2 \to 0$. The proof of

$$\frac{1 - H * G(b)}{1 - H_2 * G_2(b)} \to 0$$

paraphrases the proof above. Accordingly, Theorem 4.1 is proved. □

THEOREM 4.2. *If the density* $f(x)$ *is log concave, then the condition* (4.14) *holds and the convergence is monotone.*

PROOF. The density $f(x)$ is log concave means that $f(x) = e^{-u(x)}$ and $u(x)$ is convex for $x > 0$. It is elementary [see Karlin (1968), Chapter 2] that $1 - F(x) = \int_x^\infty f(\xi)\, d\xi$ is also log concave and decreasing. Therefore,

$$1 - F(x) = e^{-\varphi(x)},$$

$\varphi(x)$ convex increasing for $x \geq 0$ and $\varphi(0) = 0$. Set $\varphi(x) = 0$ for $x < 0$ so that $\varphi(x)$ is convex for the whole real line.

By the iterated theorem of log concave densities [Karlin (1968), Chapter 3], we know that the convolution of the kernels $K_1(x) = 1 - F(x)$, $K_2(x) = \int_0^x [1 - F(x - \xi)] f(\xi)\, d\xi$ satisfies the determinant inequality

$$\begin{vmatrix} K_2(y) & K_2(x) \\ K_1(y) & K_1(x) \end{vmatrix} \geq 0 \quad \text{for } y > x,$$

or equivalently

$$\frac{K_2(y)}{K_1(y)} \geq \frac{K_2(x)}{K_1(x)},$$

that is,

$$\frac{F_1(y) - F_2(y)}{1 - F_1(y)} > \frac{F_1(x) - F_2(x)}{1 - F_1(x)},$$

which implies

$$\frac{1 - F_2(y)}{1 - F_1(y)} = 1 + \frac{F_1(y) - F_2(y)}{1 - F_1(y)} \geq \frac{1 - F_2(x)}{1 - F_1(x)}.$$

Thus $[1 - F_1(x)]/[1 - F_2(x)]$ is decreasing in x.

We prove next that

$$A(x) = \frac{F_1(x) - F_2(x)}{1 - F_1(x)} \uparrow \infty \quad \text{as } x \uparrow \infty.$$

To this end, we examine

$$A(x) = \frac{\int_0^x [1 - F(x - \xi)] f(\xi)\, d\xi}{1 - F(x)} = \int_0^x e^{-\varphi(x-\xi)+\varphi(x)-\varphi(\xi)} \varphi'(\xi)\, d\xi$$

and $\varphi'(\xi) \geq 0$ since φ is convex on $(0, \infty)$ with an infinite range. But $\varphi(x) - \varphi(\xi) - \varphi(x - \xi) + \varphi(0) > 0$ by convexity. Thus $A(x) \geq e^{-\varphi(0)}[\varphi(x) - \varphi(0)]$. But $\varphi(x)$ necessarily $\uparrow \infty$ since $f(x)$ is a positive density over the whole real line and therefore $A(x) \uparrow \infty$, entailing the result $[1 - F_1(b)]/[1 - F_2(b)] \downarrow 0$.

By similar arguments we can prove that

$$\frac{1 - F_2(b)}{1 - F_1(b - K)} \to \infty \tag{4.17}$$

as $b \to \infty$ for K fixed. Indeed, forming the corresponding integral, we get

$$\frac{F_1(b) - F_2(b)}{1 - F_1(b - K)} = \int_0^b e^{-\varphi(b-\xi)+\varphi(b-K)-\varphi(\xi)} \varphi'(\xi)\, d\xi$$

$$\geq e^{-\varphi(K)} \int_K^{b-K} e^{-\varphi(b-\xi)+\varphi(b-K)-\varphi(\xi)+\varphi(K)} \varphi'(\xi)\, d\xi$$

and since φ is convex on $(-\infty, \infty)$, the right-hand side goes to ∞ as before, establishing the result of (4.17). □

5. Extremal r-spacings of independent uniform [0, 1] variables. Let $V_1, V_2, \ldots, V_{n-1}$ be i.i.d. uniform [0, 1]. We form their order statistics $V_1^* \le V_2^* \le \cdots \le V_{n-1}^*$ and the associated spacings $U_i = V_i^* - V_{i-1}^*$, $i = 1, \ldots, n$, where $V_0^* = 0$, $V_n^* = 1$. The r-scans in this context consist of the sums $R_i = \sum_{k=i}^{i+r-1} U_k = V_{i+r-1}^* - V_{i-1}^*$, more aptly labeled here r-spacings. The order statistics of R_i are denoted by

$$R_1^* \le R_2^* \le \cdots \le R_{n-r+1}^*, \tag{5.1}$$

as in (1.2). The study of the extremal values of the sequence (5.1) can be effectively derived from the results of Theorems 1 and 2 by virtue of the following familiar distributional equivalence:

(5.2) $(S_n U_1, \ldots, S_n U_n)$ has the same joint distribution as $(X_1, X_2, \ldots, X_n)$,

where X_i are i.i.d. following an exponential (scale parameter 1) density and S_n is a gamma$(n, 1)$ r.v. independent of $\{U_i\}$. Therefore, $S_n R_i$ has the same joint distribution as $\tilde{R}_i = \sum_{k=i}^{i+r-1} X_k$, $i = 1, \ldots, n - r + 1$, and the count $\tilde{N}_n^-(a)$ of the indices satisfying $\tilde{R}_i \le a$, $i = 1, \ldots, n - r + 1$, has the same distribution as the count of indices $N_n^-(a)$ satisfying $S_n R_i \le a$. Thus

$$d(N_n^-(a), Z_\lambda) \le (1 - e^{-\lambda})\left[(2r - 1)F_r(a) + 2\sum_{m=1}^{r-1} F_m(a)\right], \tag{5.3}$$

where $\lambda = (n - r + 1)F_r(a)$ [$F_r(\cdot)$ is the r-fold convolution distribution of $F_1(x) = 1 - e^{-x}$].

Obviously,

$$N_n^-(a) = \sum_{i=1}^{n-r+1} I_i, \tag{5.4}$$

where $I_i = 1$ if $S_n R_i \le a$ and 0 otherwise. Let

$$N_{nU}^-(a) = \sum_{i=1}^{n-r+1} J_i, \tag{5.5}$$

where $J_i = 1$ if $nR_i \le a$ and 0 otherwise.

We prove the following theorem.

THEOREM 5.1. *For $\lambda = (n - r + 1)F_r(a) \approx n(a^r/r!)$, then*

$$d(N_{nU}^-(a), Z_\lambda) \le d(N_n^-(a), Z_\lambda) + O\left(\sqrt{\frac{\log n}{n}}\right), \tag{5.6}$$

which is of order $O(n^{-1/r})$ for $r > 2$.

PROOF. Consider two possibilities for S_n, namely,

$$\mathscr{E}_1\colon \left|\frac{S_n}{n} - 1\right| \le \sqrt{\frac{\log n}{n}} \quad \text{and} \quad \mathscr{E}_2\colon \left|\frac{S_n}{n} - 1\right| > \sqrt{\frac{\log n}{n}}\,.$$

Under the event $\mathscr{E}_1$ we estimate by a union bound, so

$$\Pr\{N_{nU}^-(a) \not\equiv N_n^-(a)\} \le \sum_{i=1}^{n-r+1} \Pr\left\{I_i \ne J_i \text{ and } \left|\frac{S_n}{n} - 1\right| \le \sqrt{\frac{\log n}{n}}\right\} + \Pr\left\{\left|\frac{S_n}{n} - 1\right| > \sqrt{\frac{\log n}{n}}\right\}.$$

But the outcome $I_i \ne J_i$ signifies either $S_n R_i \le a$ but $nR_i > a$ or $nR_i \le a$ but $S_n R_i > a$. In conjunction with

$$\left|\frac{S_n}{n} - 1\right| \le \sqrt{\frac{\log n}{n}}\,,$$

these conditions lead to the inequalities

$$\left(1 - \sqrt{\frac{\log n}{n}}\right)a \le \tilde{R}_i \le a \quad \text{or} \quad a \le \tilde{R}_i \le a\left(1 + \sqrt{\frac{\log n}{n}}\right),$$

whose probabilities are bounded by

$$O\left(\frac{a^r}{(r-1)!}\sqrt{\frac{\log n}{n}}\right).$$

Thus

$$\begin{aligned}\sum_{i=1}^{n-r+1} \Pr\left\{I_i \ne J_i \text{ and } \left|\frac{S_n}{n} - 1\right| \le \sqrt{\frac{\log n}{n}}\right\} &\le nO\left(\frac{a^r}{(r-1)!}\sqrt{\frac{\log n}{n}}\right) \\ &= \lambda r O\left(\sqrt{\frac{\log n}{n}}\right).\end{aligned} \tag{5.7}$$

The foregoing inequality implies

$$d(N_{nU}^-(a), N_n^-(a)) \le O\left(\sqrt{\frac{\log n}{n}}\right) + \Pr\left\{\left|\frac{S_n}{n} - 1\right| > \sqrt{\frac{\log n}{n}}\right\}. \tag{5.8}$$

The Berry–Esseen error estimate of the normal approximation to $(S_n - n)/\sqrt{n}$ yields

$$\Pr\left\{\left|\frac{S_n}{n} - 1\right| > \sqrt{\frac{\log n}{n}}\right\} = O\left(\frac{1}{\sqrt{n}}\right) + \frac{2}{\sqrt{2\pi}}\int_{\sqrt{\log n}}^{\infty} e^{-\xi^2/2}\,d\xi$$

$$= O\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{e^{-\log n/2}}{\sqrt{\log n}}\right) \tag{5.9}$$

$$= O\left(\frac{1}{\sqrt{n}}\right).$$

From the results above plus (2.6) and (2.4) of Section 2 we obtain

$$d(N_{nU}^{-}(a), Z_\lambda) \le d(N_n^{-}(a), Z_\lambda) + d(N_{nU}^{-}(a), N_n^{-}(a))$$

$$\le d(N_n^{-}(a), Z_\lambda) + O\left(\sqrt{\frac{\log n}{n}}\right),$$

as was to be proved in (5.6). □

An anonymous referee proposes that using by the coupling method the error term $O(\sqrt{\log n/n})$ can be deleted.

Focusing on the kth term of the Poisson approximation, we achieve the limiting formula (1.18).

As elaborated in Section 3 we can allow r to increase to ∞ at a sufficiently slow rate and retain the Poisson limit. For example, consider $r = (\ln n)^{1-\alpha}$, $\alpha > 0$, and determine a satisfying $(a^r/r!)n \approx \lambda$, $0 < \lambda < \infty$, then $a \approx (\ln n)^{1-\alpha}/e^{(\ln n)^\alpha}$ and

$$d(N_{nU^{(r)}}^{-}(a), Z_\lambda) \le O\left(\sqrt{\frac{\log n}{n}}\right) + O(a). \tag{5.10}$$

If $\ln \lambda$ goes to ∞ at a slower rate than r, then $N_{nU^{(r)}}^{-}(a)$ possesses a central limit theorem.

Maximum r-scan. The ascertainment of limit laws for N_{nU}^{+} parallels the analysis of N_{nU}^{-}. We reduce the problem to the i.i.d. case with the exponential density, yielding Poisson approximations. Explicitly,

$$d(N_n^{+}(b), Z_\mu) \le \varepsilon(r, b)$$

defined in (1.5), with $\mu \approx n[1 - F_r(b)]$. Consult Theorem 2 and Section 4.

In the case at hand we can achieve the error rate

$$\varepsilon(r, b) = O\left(\sqrt{\frac{\ln b/r}{b/r}}\right),$$

which goes to 0 provided $b/r \uparrow \infty$ or even the error rate $O(b^{-1})$ when r is fixed. The next step obtains the corresponding limit laws for $N_{nU}^{+}(b)$ which has the

same limit law as $N_n^+(b)$, where $N_{nU}^+(b)$ is the count of all $nR_i > b$. The equation

$$\mu = (n - r + 1)[1 - F_r(b)] = (n - r + 1)e^{-b}\sum_{i=0}^{r-1}\frac{b^i}{i!} \tag{5.11}$$

requires

$$b = \ln n + (r - 1)\ln\ln n - \ln \mu(r - 1)! + o(1). \tag{5.12}$$

The kth term of the Poisson approximation for the maximal r-scan uniform variable spacing has the limit law given in (1.19).

6. The r-scan process with a variable time horizon. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. positive random variables with distribution function $F(x)$. Let S_k be the partial sum process based on $\{X_i\}$. For $t > 0$ define n_t as the renewal count, that is, $n_t = k$ if and only if $S_k \le t < S_{k+1}$. We form the r-scan process $\{\hat{R}_i\}_1^{n_t - r+1}$ for the random number of r.v.'s $X_1, X_2, \ldots, X_{n_t}$ and let $\{R_i\}$ be the r-scan process for the deterministic number $\bar{n}_t$ = integer part of $E[n_t]$.

We denote the corresponding count variables of the r-scan $\{\hat{R}_i\}_1^{n_t - r+1}$ and $\{R_i\}_1^{\bar{n}_t - r+1}$ by $\hat{N}_t^-(a), \hat{N}_t^+(b)$ and $N_t^-(a), N_t^+(b)$, respectively. The following result holds.

THEOREM 6.1. *Let* $\lambda = (\bar{n}_t - r + 1)F_r(a)$ *and* $\mu = (\bar{n}_t - r + 1)[1 - F_r(b)]$. *Then*

$$d\left(\hat{N}_t^-, Z_\lambda\right) \le (1 - e^{-\lambda})\delta(r, a) + O\left(\sqrt{\frac{\log t}{t}}\right), \tag{6.1a}$$

$$d\left(\hat{N}_t^+, Z_\mu\right) \le (1 - e^{-\mu})\varepsilon(r, b) + O\left(\sqrt{\frac{\log t}{t}}\right). \tag{6.1b}$$

The proof adapts the methods of Section 5. A standard central limit theorem holds for

$$\frac{n_t - \dfrac{t}{\mu_x}}{\dfrac{\sigma}{\mu_x^{3/2}}\sqrt{t}}$$

with an error $O(1/\sqrt{t})$, where $\mu_x = E[X_1]$ and $\sigma^2 = \operatorname{Var} X_1$. Comparing $\hat{N}_t^-(a)$ with $N_t^-(a)$ under the two realizations $|n_t - \bar{n}_t| \le \sqrt{\log t/t}$ and $|n_t - \bar{n}_t| > \sqrt{\log t/t}$ and paraphrasing the analysis set forth in Section 5, the results of (6.1) are confirmed. We omit the details.

7. Extremal statistics for r-scans from stationary processes. Consider a finite stationary ergodic process $\mathscr{S}$ taking values from a finite set $S = \{1, 2, \ldots, s\}$. Conditioned on the realization $A_1, A_2, \ldots, A_n$ from $\mathscr{S}$ let $Y_1, \ldots, Y_n$ be independent positive r.v.'s and identically distributed for those A_i in the same state. For example, if $\mathscr{S}$ is a finite state stationary Markov chain, then with each state, say γ, there is a distribution function $F^{(\gamma)}(x)$ and i.i.d. random variables following $F^{(\gamma)}(x)$. The $\{Y_i\}$ for the realization $\{A_i\}$ accordingly consist of independent random variables, and the subsequence $\{Y_i^{(\gamma)}\}$ from $\{Y_i\}$ corresponding to visits to state γ are i.i.d. governed by the distribution $F^{(\gamma)}(x)$. We abbreviate the event $A_1 = l_1,\ A_2 = l_2, \ldots, A_n = l_n$ by $A_{\mathbf{n}} = l$ [$\mathbf{n} = (1, \ldots, n)$, $l = (l_1, \ldots, l_n)$].

Denote by R_i, $i = 1, \ldots, n - r + 1$, the associated r-scan process of $\{Y_i\}$ [see (1.1)], that is, $R_i = \sum_{k=i}^{i+r-1} Y_k$, $i = 1, \ldots, n - r + 1$, and $N^-(a)$ and $N^+(b)$ as in (1.5) the count of those $R_i \le a$ and $R_i > b$, respectively. By stationarity,

$$\begin{aligned} \lambda = E[N^-] &= (n - r + 1) \sum_l \Pr\{R_1 \le a | A_{\mathbf{n}} = l\} \Pr\{A_{\mathbf{n}} = l\} \\ &= (n - r + 1) \sum_l \Pr\{R_1 \le a,\ A_1 = l_1, \ldots, A_r = l_r\}. \end{aligned} \tag{7.1}$$

Also

$$\mu = E[N^+] = (n - r + 1) \sum_l \Pr\{R_1 > b,\ A_1 = l_1, \ldots, A_r = l_r\}. \tag{7.2}$$

The Poisson approximation to $N^-(N^+)$ by $Z_\lambda(Z_\mu)$ will be derived with error terms next.

Associated with a given realization $\mathscr{A} = \{A_1, \ldots, A_n\}$, we have the independent r.v. sequence $Y_1(A_1), \ldots, Y_n(A_n)$ and for each specific state $A_{i_1} = A_{i_2} = \cdots = A_{i_m} = \gamma$, the corresponding $Y_{i_r}(A_{i_r})$ are i.i.d governed by the distribution function $F^{(\gamma)}(x)$. We form the r-scan process conditioned on $\mathscr{A}$:

$$R_i(\mathscr{A}) = \sum_{k=i}^{i+r-1} Y_k(A_k), \qquad i = 1, \ldots, n - r + 1,$$

and parallel to (1.5) the associated counts

$$N_n^-(a) = \sum_{i=1}^{n-r+1} W_i^-(a), \qquad W_i^-(a) = \begin{cases} 1, & \text{if } R_i \le a \\ 0, & \text{otherwise.} \end{cases} \tag{7.3}$$

Clearly,

$$\lambda(\mathscr{A}) = E[N_n^-(a) | \mathscr{A}] = \sum_{i=1}^{n-r+1} \Pr\{R_i \le a | \mathscr{A}\}. \tag{7.4}$$

Paraphrasing the analysis of Theorem 1, we deduce

$$d\big((N^-(a)|\mathscr{A}), (Z_\lambda|\mathscr{A})\big) \le \big[b_1(\mathscr{A}) + b_2(\mathscr{A})\big]/\lambda(\mathscr{A}), \tag{7.5}$$

using the notation $(Z_\lambda|\mathscr{A})$ in place of the Poisson variable $Z_{\lambda(\mathscr{A})}$, where

$$b_1(\mathscr{A}) = (2r-1)\lambda(\mathscr{A}) \max_{\gamma} F_r^{\gamma}(a), \tag{7.6}$$

where F_r^{γ} is the r-fold convolution $F^{\gamma_1} * F^{\gamma_2} * \cdots * F^{\gamma_r}$, and $\gamma = (\gamma_1, \ldots, \gamma_r)$ traverses the set of all configurations of r-states, $\gamma_i \in \mathscr{S}$. The calculation of $b_2(\mathscr{A})$ paralleling (3.4) produces

$$\begin{aligned} &\sum_{i=1}^{n-r+1} \sum_{\substack{k=-r+1 \\ k\neq 0}}^{r-1} \Pr\{R_{i+k} \le a | R_i \le a, \mathscr{A}\} \Pr\{R_i \le a|\mathscr{A}\} \\ &\quad \le (2r-1)\lambda(\mathscr{A}) \max_{\gamma} F_1^{(\gamma)}(a). \end{aligned} \tag{7.7}$$

The combination (7.6)–(7.7) gives the bound (d is variational distance)

$$d\big((N^-(a)|\mathscr{A}), (Z_\lambda|\mathscr{A})\big) \le 4rG(a), \tag{7.8}$$

where $G(a) = \max_\gamma [F_1^{(\gamma)}(a)]$. Averaging over $\mathscr{A}$, we get

$$E_{\mathscr{A}}\big(d\big((N^-(a)|\mathscr{A}), (Z_\lambda|\mathscr{A})\big)\big) \le 4rG(a).$$

Appealing to (2.8), we deduce

$$d\big(N^-(a), Z_\lambda\big) \le E_{\mathscr{A}}\big[d\big((N^-(a)|\mathscr{A}), Z_\lambda\big)\big], \tag{7.9}$$

with Z_λ defined in (7.1). Invoking the triangle inequality (2.4) yields

$$\begin{aligned} d\big(N^-(a), Z_\lambda\big) &\le E_{\mathscr{A}}\big[d\big((N^-(a)|\mathscr{A}), (Z_\lambda|\mathscr{A})\big)\big] \\ &\quad + E_{\mathscr{A}}\big[d\big((Z_\lambda|\mathscr{A}), Z_\lambda\big)\big] \\ &\le 4rG(a) + E_{\mathscr{A}}\big[d\big(Z_\lambda(\mathscr{A}), Z_\lambda\big)\big]. \end{aligned} \tag{7.10}$$

The first term goes to 0 with $G(a) \to 0$ which is the case if $F_1^{(\gamma)}(a) \to 0$ for every state γ. We stipulated only finitely many states for simplicity but this could easily be generalized with appropriate technical requirements.

We will now prove that the ergodic nature of $\mathscr{S}$ compels the second term of (7.10) to go to 0 where the error term can be calculated. We consider all r-states of components from $\mathscr{S}$, $\gamma = (\gamma_1, \ldots, \gamma_r)$ totalling s^r possibilities. For the process $\mathscr{S}$ let $c_n(\gamma)$ be the expected number of consecutive r-states of type γ in the first n samples from the unconditional process $\mathscr{S}$ and let $c(\gamma; \mathscr{A})$ be the number of γ-states occurring in a particular realization $\mathscr{A} = (A_1, \ldots, A_n)$. The basic ergodic theorem entails that

$$\frac{c(\gamma, \mathscr{A})}{n} \to \pi(\gamma) \tag{7.11}$$

with probability 1 (for almost any realization) and of course $c_n(\gamma)/n \to \pi(\gamma)$, where $\pi(\gamma)$ is the unconditional probability of observing the r-state γ in any r consecutive terms of $\mathscr{S}$. Let $p_\gamma(a)$ be the probability that an r-scan induced by γ has value less than or equal to a, that is,

$$p_\gamma(a) = \Pr\{Y(\gamma_1) + \cdots + Y(\gamma_r) \le a\}. \tag{7.12}$$

The unconditional expected number of r-scan counts is

$$\lambda = \sum_{\gamma} c_n(\gamma) p_\gamma(a). \tag{7.13}$$

For each realization $\mathscr{A}$, the expected number of r-scan counts is

$$\lambda(\mathscr{A}) = \sum_{\gamma \in \mathscr{A}} p_\gamma(a) c(\gamma; \mathscr{A}).$$

We evaluate next, using the formulas for λ, $\lambda(\mathscr{A})$,

$$\begin{aligned} E_{\mathscr{A}}|\lambda - \lambda(\mathscr{A})| &\le \sum_{\gamma} c_n(\gamma) p_\gamma(a) E_{\mathscr{A}} \left| 1 - \frac{c(\gamma, \mathscr{A})}{c_n(\gamma)} \right| \\ &\le \lambda \max_{\gamma} E_{\mathscr{A}} \left| 1 - \frac{c(\gamma, \mathscr{A})}{c_n(\gamma)} \right|. \end{aligned} \tag{7.14}$$

Using the ergodic theorem the quantity above $\to 0$ since

$$\frac{c_n(\gamma)}{n} \to \pi(\gamma) \quad \text{and also} \quad \frac{c(\gamma, \mathscr{A})}{n} \to \pi(\gamma). \tag{7.15}$$

The convergence in (7.15) is geometrically fast for the Markov chain case. Let ε_n be the error in the rate of convergence of (7.15), providing an error estimate $O(\lambda \varepsilon_n)$ of the second term of (7.10). By similar means to (7.10), we derive

$$d(N^+(b), \mu) \le E_{\mathscr{A}}\big[d(N^+(b)|\mathscr{A}), (Z_\mu|\mathscr{A}) \big] + E_{\mathscr{A}}|\mu(\mathscr{A}) - \mu|, \tag{7.16}$$

where

$$\mu = \sum_{i=1}^{n-r+1} \Pr\{R_i > b|\mathscr{A}\}\Pr\{\mathscr{A}\} = \sum_{\gamma} c_n(\gamma) \Pr\{Y(\gamma_1) + \cdots + Y(\gamma_r) > b\}.$$

The second term in (7.16) is bounded by $O(\mu \varepsilon_n)$, by the corresponding analog of (7.14). The first term is bounded by

$$4r \max_{\gamma_1, \ldots, \gamma_{r+1}} \Pr\left\{ \sum_{i=2}^{r+1} Y_i(\gamma_i) > b \,\middle|\, \sum_{i=1}^{r} Y_i(\gamma_i) > b \right\}$$

[see (4.6)].

8. Extremal r-spacings for a general distribution. In Section 5 we ascertained the asymptotic distributions ($n \to \infty$) of various extremal r-spacings for n sampled i.i.d. uniform [0, 1] r.v.'s. We shall extend several of the results on extremal r-spacings to the context of samples from a general distribution. To this end, we concentrate first on a finite piecewise constant density. Thus let $V_1, \ldots, V_{n-1}$ be $(n-1)$ i.i.d. samples drawn from the density

$$f(x) = \begin{cases} p_j/d_j, & \Delta_j \le x < \Delta_j + d_j = \Delta_{j+1},\ j = 1, 2, \ldots, L, \\ 0, & \text{otherwise}, \end{cases} \tag{8.1}$$

where $\sum_{j=1}^{L} p_j = 1$, $p_j > 0$, and $d_j > 0$, $j = 1, 2, \ldots, L$. Consider the order statistics $V_1^* \le V_2^* \le \cdots \le V_{n-1}^*$ induced from $\{V_i\}$ and associated spacings $U_i = V_i^* - V_{i-1}^*$, $i = 1, \ldots, n$, where $V_0^* = \Delta_1$ and $V_n^* = \Delta_{L+1}$ and the interval $[\Delta_1, \Delta_{L+1}]$ subtends the support of the density f.

The r-scans $R_i = V_{i+r-1}^* - V_{i-1}^*$, $i = 1, \ldots, n - r + 1$, and their order statistics R_i^* are defined as in Section 5.

Let $N_n^-(a)$ be the number of r-scans among $R_i = \sum_{k=i}^{i+r-1} U_k$, $i = 1, \ldots, n - r + 1$, satisfying $R_i \le a$. We prove a Poisson approximation to $N^-(a) = N_n^-(a)$ with error estimates. For λ defined in (8.3) below, we prove

$$d\big(N_n^-(a), Z_\lambda\big) \le \varepsilon_n \tag{8.2}$$

with an explicit error term $\varepsilon_n \to 0$ as $n \to \infty$.

The appropriate λ for the Poisson approximation r-scan of $N^-(a)$ is specified to be

$$\lambda = \sum_{j=1}^{L} (np_j - r + 1) F_r\left(\frac{anp_j}{d_j}\right), \tag{8.3}$$

where $F_r(x)$ is the r-fold convolution of $F_1(x) = 1 - e^{-x}$ (see Section 3) and a is taken to be of order $0 < \gamma_1 \le an^{1+1/r} \le \gamma_2 < \infty$. Since $F_r(\delta)$ for δ small behaves as $\delta^r/r!$ with a as prescribed, it follows straightforwardly that λ is bounded away from 0 and ∞. In particular, for $an^{1+1/r} = x$,

$$\lambda \approx \frac{n}{r!} \sum_{j=1}^{L} p_j \frac{a^r n^r p_j^r}{d_j^r} = \frac{x^r}{r!} \sum_{j=1}^{L} \frac{p_j^{r+1}}{d_j^r} = \frac{x^r}{r!} \int [f(\xi)]^{r+1} d\xi.$$

Let $n_j - 1$, $n_j \ge 1$, denote the number of $\{V_i\}_{i=1}^{n-1}$ within $I_j = [\Delta_j, \Delta_{j+1})$ for $j = 1, 2, \ldots, L$ [so that $\sum_{j=1}^{L}(n_j - 1) = n - 1$]. For every realization of the random vector $\mathbf{n} = (n_1, \ldots, n_L)$, let $N_j^-(a)$, $j = 1, \ldots, L$, indicate the count of r-scans not exceeding level a based on the $n_j - 1$ values V_i contained within I_j augmented with the two boundary points Δ_j, Δ_{j+1}. Accordingly, $N_j^-(a)$ is the aggregate count of r-scans for $n_j - 1$ uniform random variables on the interval $[\Delta_j, \Delta_{j+1}]$. Let

$$\lambda_j = (n_j - r + 1) F_r\left(\frac{an_j}{d_j}\right). \tag{8.4}$$

Theorem 5.1 applies to the $\{V_i\}$ within $I_j = [\Delta_j, \Delta_{j+1}]$, entailing that

$$d\big(N_j^-(a), Z_{\lambda_j}\big) \le 4\frac{an_j}{d_j} + O\left(\sqrt{\frac{\log n_j}{n_j}}\right). \tag{8.5}$$

We may assume each $n_j \to \infty$ since the probability for the contrary event tails exponentially fast to 0 as $n \to \infty$. Moreover, for a realization $\{n_1, \ldots, n_L\}$ the random variables $\{N_j^-(a)\}_{j=1}^{L}$ juxtaposed with $\{Z_{\lambda_j}\}_{j=1}^{L}$ constitute independent

r.v.'s and (2.9) applies to give the variation distance estimate [take $\lambda^* = \lambda^*(\mathbf{n}) = \sum_{j=1}^{L}\lambda_j$]

$$d\left(\sum_{j=1}^{L} N_j^-(a), Z_{\lambda^*}\right) \le 4a\left(\sum_{j=1}^{L}\frac{n_j}{d_j}\right) + O\left(\sum_{j=1}^{L}\sqrt{\frac{\log n_j}{n_j}}\right).$$

Since $\sum_{j=1}^{L} N_j^-(a)$ and the r-scan count $N^-(a)$ differ by at most $(L-1)r$ of r-scans for those sums which involve contributions coming from adjacent intervals I_j and I_{j+1}, we may conclude that

$$\begin{aligned} &d(N^-(a), Z_{\lambda^*}) \\ &\quad \le d\left(N^-(a), \sum_{j=1}^{L} N_j^-(a)\right) + d\left(\sum_{j=1}^{L} N_j^-(a), Z_{\lambda^*}\right) \\ &\quad \le \max_{j=1,\dots,L-1}\left[\Pr\{\text{a boundary } r\text{-scan between } I_j \text{ and } I_{j+1} \le a\}\right] r(L-1) \\ &\qquad + 4a\left(\sum_{j=1}^{L}\frac{n_j}{d_j}\right) + O\left(\sum_{j=1}^{L}\sqrt{\frac{\log n_j}{n_j}}\right). \end{aligned} \tag{8.6}$$

For n large and all $n_j \to \infty$ we have the estimate (using the normal approximation; see Section 5)

$$\Pr\left\{\left|\frac{n_j - p_j n}{\sqrt{n}}\right| > \sqrt{\log n}\right\} \le O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore,

$$\Pr\left\{\max_{1\le j\le L}|n_j - p_j n| > \sqrt{n\log n}\right\} \le O\left(\frac{L}{\sqrt{n}}\right). \tag{8.7}$$

On the other hand, when $|n_j - p_j n| \le \sqrt{n\log n}$ for all j, we have [using the fact $F_r(\delta) \approx \delta^r/r!$ for δ small]

$$\begin{aligned} \left|\sum_{j=1}^{L}\lambda_j - \lambda\right| &= \left|\sum_{j=1}^{L}(n_j - r + 1)F_r\left(\frac{an_j}{d_j}\right) - \sum_{j=1}^{L}(np_j - r + 1)F_r\left(\frac{anp_j}{d_j}\right)\right| \\ &\le O\left(\sqrt{n\log n}\,a^r n^r\right) + \sum_{j=1}^{L}(np_j - r + 1)\left|F_r\left(\frac{anp_j}{d_j}\right) - F_r\left(\frac{an_j}{d_j}\right)\right| \\ &\le O\left(\sqrt{\log n}\,a^r n^{r+1/2}\right) + 2\sum_{j=1}^{L} np_j \frac{|np_j - n_j|}{\min_j d_j^r}\frac{a^r}{(r-1)!}n^{r-1} \\ &= O\left(\sqrt{n\log n}\,(an)^r\right) = O\left(\sqrt{\frac{\log n}{n}}\right), \end{aligned} \tag{8.8}$$

since r and L are finite, $an^{1+1/r}$ is bounded and d_j is positive for all j.

Observe next that

$$d(N^-(a), Z_\lambda) \le EE_{\mathbf{n}}\left[d\left(N^-(a), Z_{\lambda^*(\mathbf{n})}\right) + d\left(Z_{\lambda^*(\mathbf{n})}, Z_\lambda\right)\right]. \tag{8.9}$$

The inner expectation of the first term on the right-hand side of (8.9) by virtue of (8.6) is bounded by

$$O\left(\frac{1}{n}\right) + O(na) + E\left[\sum_{j=1}^{L} \sqrt{\frac{\log n_j}{n_j}}\,,\ n_j > \frac{p_j n}{2} \text{ for all } j\right] + \Pr\left\{\text{some } n_j < \frac{p_j n}{2}\right\}. \tag{8.10}$$

The last term of (8.9), due to the inequality (2.7), namely

$$d(Z_{\lambda^*}, Z_\lambda) \le \left|\sum_{j=1}^{L} \lambda_j - \lambda\right|,$$

is bounded by

$$E\left[\left|\sum_{j=1}^{L} \lambda_j - \lambda\right|\right]. \tag{8.11}$$

Using the estimates of (8.8) and (8.10) the terms of (8.10) and (8.11) are (for $r > 2$) of the order $O(na) = O(1/n^{1/r})$, since the other terms are of the order $O(\sqrt{\log n/n})$ or exponentially small in n.

When $f(v)$ is any continuous density with a bounded support, then by approximating with a finite piecewise constant density we achieve in a standard way, as $n \to \infty$, that $N^-(a)$ is approximately Poisson (λ) with

$$\lambda \simeq \frac{(an)^r}{r!} n \int [f(\xi)]^{r+1}\, d\xi. \tag{8.12}$$

Maximal r-spacings. Let $N_n^+(b)$ be the number of r-spacings satisfying $R_i > b$, where $V_1, \ldots, V_{n-1}$ are $n-1$ i.i.d. samples from the density (8.1). We assume that L and r are fixed while $n \to \infty$ and that $d_j p_j > 0$, $j = 1, \ldots, L$. Let $f^* = \min(p_j/d_j) > 0$ and assume for ease of exposition that the minimum is achieved uniquely. Let p^* be the probability that a sample V_i is drawn from the interval I_{j^*}, $j^* \in \{1, \ldots, L\}$ with the minimal density level f^*.

Fix $0 < \mu < \infty$ (μ is a free parameter) and determine

$$b = \frac{1}{nf^*}\left(\log\left(\frac{np^*}{\mu(r-1)!}\right) + (r-1)\log\log n\right). \tag{8.13}$$

Then as $n \to \infty$ we have a Poisson (μ) approximation for $N_n^+(b)$ with error term

$$d\left(Z_\mu, N^+(b)\right) = O\left(\frac{1}{\log n}\right). \tag{8.14}$$

The key step in establishing (8.14) is to prove that for b given in (8.13),

$$d\big(N^+(b), N_{j^*}^+(b)\big) \to 0 \quad \text{as } n \to \infty, \tag{8.15}$$

where $N_{j^*}^+(b)$ counts r-spacings exceeding the level b within I_{j^*}. Thus the counts of maximal r-spacings are essentially contributed by samples from the interval with smallest density value. By contrast [see (8.12)] the distribution of $N^-(a)$ depends on the entire density $f(\xi)$.

Apart from an event of probability $O(1/\sqrt{n})$, we have

$$\max_{j=1,\ldots,L} |n_j - np_j| \le \sqrt{n \log n}$$

and

$$\max_{j=1,\ldots,L} |S_{n_j} - np_j| \le \sqrt{n \log n}\,, \tag{8.16}$$

where n_j is the number of samples among $\{V_i\}$ in I_j and S_{n_j} evaluates a sum of n_j exponential (1) random variables. As previously (see Section 5), we can express the joint distribution of $V_i^* - V_{i-1}^*$ within the interval I_j in terms of the sums S_{n_j}.

When conditions (8.16) hold, we estimate, invoking the union of events bound, that

$$\begin{aligned} &\Pr\big\{N_n^+(b) \ne N_{j^*}^+(b)\big\} \\ &\le n\left(1 - F_r\left(bn \min_{j \ne j^*}\left(\frac{p_j}{d_j}\right)\right)\right)\left[1 + c\sqrt{\frac{(\log n)^3}{n}}\right]. \end{aligned} \tag{8.17}$$

Since $\min_{j \ne j^*}(p_j/d_j) > f^*$ and with b specified as in (8.13), the left-hand side of (8.17) is of the order $O(n^{-\alpha})$ for $\alpha = \min_{j \ne j^*}[p_j/(d_j f^*) - 1]$, and so

$$\lim_{n \to \infty} \Pr\big\{N_n^+(b) \ne N_{j^*}^+(b)\big\} = 0,$$

validating (8.15).

Since $|n_{j^*} - np^*| \le \sqrt{n \log n}$ [apart from a rare event of probability $O(1/\sqrt{n})$] the error bound of (8.14) ensues as a consequence of (4.11) [Poisson approximation with error $O(1/b)$ is established for the maximal r-scans of independent exponential (1) random variables].

Similar arguments apply when there are multiple intervals with the *same minimal density level*. Since (8.15) holds as long as the minimal density $f^* > 0$ is obtained on the union of several finite intervals, (8.14) and (8.13) also persist under these conditions. When $p_j = 0$ for some j, say j_0, such that the support of the underlying density is not connected, then clearly $N_n^+(d_{j_0}) \ge 1$ a.s., and the extremal maximal r-scans possess a rather trivial behavior. The case of a continuous sampling density $f(\xi)$ with connected support is interesting but not covered here.

REFERENCES

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25.

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen–Stein method (with discussion). *Statist. Sci.* **5** 403–434.

BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.

BARBOUR, A. D. and HOLST, L. (1989). Some applications of the Stein–Chen method for proving Poisson convergence. *Adv. in Appl. Probab.* **21** 74–90.

CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

CRESSIE, N. (1977). The minimum of higher order gaps. *Austral. J. Statist.* **19** 132–143.

DEHEUVELS, P. (1986). On the influence of the extremes of an i.i.d. sequence on the maximal spacings. *Ann. Probab.* **14** 194–208.

DEHEUVELS, P. and DEVROYE, L. (1987). Limit laws of Erdös–Rényi–Shepp type. *Ann. Probab.* **15** 1363–1386.

GLAZ, J. (1989). Approximations and bounds for the distribution of the scan statistic. *J. Amer. Statist. Assoc.* **84** 560–566.

GLAZ, J. and NAUS, J. (1979). Multiple convergence of the line. *Ann. Probab.* **7** 900–906.

GODBOLE, A. P. (1990). Poisson approximations for runs and patterns of rare events. Preprint.

HOLST, L. (1980). On multiple covering of a circle with random arcs. *J. Appl. Probab.* **17** 284–290.

HOLST, S. and JANSON, S. (1990). Poisson approximation using the Stein–Chen method and coupling: Number of exceedances of Gaussian random variables. *Ann. Probab.* **18** 713–723.

JANSON, S. (1987). Poisson convergence and Poisson processes with applications in random graphs. *Stochastic Process. Appl.* **26** 1–30.

KARLIN, S. (1968). *Total Positivity*. Stanford Univ. Press.

KARLIN, S. (1988). Coincident probabilities and applications in combinatorics. *J. Appl. Probab.* **25A** 185–200.

KARLIN, S. and LEUNG, M.-Y. (1991). Some limit theorems on distributional patterns of balls in urns. *Ann. Appl. Probab.* **1** 513–538.

KARLIN, S. and MACKEN, C. (1991). Some statistical problems in the assessment of inhomogeneities of DNA sequence data. *J. Amer. Statist. Assoc.* **86** 26–33.

KARLIN, S. and MCGREGOR, J. (1959). Coincidence probabilities. *Pacific J. Math.* **9** 1091–1108.

KARLIN, S. and OST, F. (1987). Counts of long aligned word matches among random letter sequences. *Adv. in Appl. Probab.* **19** 293–351.

KARLIN, S and RINOTT, Y. (1982). Applications of Anova type decompositions for comparisons of conditional variance statistics including jackknife estimates. *Ann. Statist.* **10** 485–501.

KARLIN, S. and TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*, 2nd ed. Academic, New York.

MELZAK, Z. A. (1979). Multi-indexing and multiple clustering. *Math. Proc. Cambridge Philos. Soc.* **86** 313–337.

NAUS, J. I. (1982). Approximations of distributions of scan statistics. *J. Amer. Statist. Assoc.* **77** 177–183.

NAUS, J. I. (1979). An indexed bibliography of clusters, clumps and coincidences. *Internat. Statist. Rev.* **47** 47–78.

WALLENSTEIN, S. and NEFF, N. (1987). An approximation for the distribution of the scan statistic. *Statistics in Medicine* **6** 197–207.

DEPARTMENTS OF MATHEMATICS
AND STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305