Record Sequences and their Applications

John Bunge and Charles M. Goldie

Cornell University and University of Sussex

(17 Apr 1999 8:31 p.m.)

Contents

1. Introduction

The probabilistic theory of extremes has been growing in impact in recent years and a state-of-the-art account can be found elsewhere. One corner of the field, the area known as *outstanding observations* or *record sequences*, or just *records*, has developed in a distinctive way and is the topic of this article. Suppose there is a sequence of independent identically distributed (i.i.d.) random variables X_1, X_2, $\dots$. These for instance could be the observations of one variable in an experiment that can be repeated as often as needed. Or they could be a stream of data values at reasonably long time intervals from a physical system that fluctuates, does not evolve, and has little dependence on past behaviour. (It could be that one postulates that the system is like that in order to test whether it is indeed so.) Knowing extreme-value theory, one might form the sequence of successive maxima $M_n := \max(X_1, X_2, \dots, X_n)$. After the first few values, M_n does not change very often. There will be long stretches where the graph of M_n against n stays flat. In fact, as we shall see below, the number of times M_n changes value between $n = 100$ and $n = 1000$, or, worse, between $n = 1000$ and $n = 10\,000$, is not likely to differ much from the number of times it changes between $n = 10$ and $n = 100$. In the circumstances it looks a good idea to throw away all the repeats and just consider the values taken by the M_n sequence. Call these R_1, R_2, R_3, $\dots$. So $R_1 := X_1$, R_2 is the first of X_2, X_3, $\dots$ to exceed R_1, and so on. These are the *record values*. Their distribution theory, both here-and-now and asymptotically, is *quite different* from that of the extreme values M_1, M_2, $\dots$, even though the R_n sequence is just the M_n sequence without repeats. The change in sampling frame, so to speak, alters everything.

There is some information left in the M_n sequence aside from the record values, namely the times at which its successive values are first taken. These are the *record* *times* L_1, L_2, L_3, $\ldots$. So $L_1 := 1$, L_2 is the first n such that $X_n > X_1$, L_3 is the first $n > L_2$ such that $X_n > X_{L_2}$, and so on. One sees that $R_k = X_{L_k}$ $(= M_{L_k})$.

A beautiful account of record sequences was given in Glick (1978). It is as worth reading now as then and in what follows we are not going to rewrite that article. Instead, we want to concentrate on clarifications of the basic structure of records that were found mainly in the ten years following Glick's article, and on developments and applications that have continued right up to now.

Ours is far from being the only view of the subject, and there are quite different accounts in the surveys of Nevzorov (1987), Nagaraja (1988) and Ahsanullah (1995). Substantial textbook treatments are in Resnick (1987) and Pfeifer (1989), and there are shorter accounts in Galambos (1987), Arnold *et al.* (1992), Resnick (1992) and Embrechts *et al.* (1997).

Everything contained in this discussion applies equally well to minima, hence *lower records*, just as much as to (upper) records. It is enough to apply results about upper records to the observations $-X_1$, $-X_2$, $\ldots$, whose upper record values are the lower record values of X_1, X_2, $\ldots$.

The above framework of record values and times arising from i.i.d. observations is what one might call the *classical record model*, which we develop in Sections 2–6 below. There are several ways to extend the classical model. One is to consider records from various sorts of *changing populations*. The models of interest are application-oriented and we discuss them in Section 7 together with certain extensions to the classical model and certain connections and applications in other fields, mainly combinatorics. The second main extension, which is the one that has progressed the furthest, is to let the original observations occur at *random times*. We devote our final five sections to these 'random record models'.

2. Partial records

As well as considering the set

$$\mathcal{R}_1 = \{R_1, R_2, R_3, \ldots\} \tag{2.1}$$

of observation values that are the largest yet seen when they appear, one can think about the set $\mathcal{R}_2$ of observation values that are the *second* largest on their appearance, and the set $\mathcal{R}_3$ of observations that are the *third* largest on their appearance, and so on. We call $\mathcal{R}_2$ the set of *2-records* or 2-record values, $\mathcal{R}_3$ the set of *3-records*, etc., and refer to them collectively as the sets of *partial records*. With this notation we can state what in our view is the most surprising fact about record values, as follows.

Theorem 2.1 (Ignatov's Theorem). $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, $\ldots$ *are independent and identically distributed random sets.*

We will say something about the name and origins of this theorem in the next section, after devoting this section to elucidating its content. First note that it is the *sets* that are independent and identically distributed; the *contents* of, say, $\mathcal{R}_1$, listed as in (2.1), are a strictly increasing sequence of r.v.s (random variables). The theorem says that the sets $\mathcal{R}_2$ and $\mathcal{R}_1$ are independent of each other, and the joint distribution of the 2-records is the same as that of the 1-records $R_1 < R_2 < R_3 < \cdots$; next, $\mathcal{R}_3$ is independent of $\mathcal{R}_1$ and $\mathcal{R}_2$ together, and the joint distribution of the 3-records is the same as that of the 1-records, and so on.

Before going further we ought to make precise the inequalities that implicitly are in use to define the records, 2-records and so on. We need to specify the inequalities carefully because, for good reasons as it turns out, we must allow for different observations having coincident values. Denote the number of elements of a set A by $\#A$. The *initial rank* of our n^{th} observation X_n is defined to be

$$\rho_n := \#\big\{i : i \in \{1, \ldots, n\},\ X_i \geq X_n\big\}. \tag{2.2}$$

So $\rho_1 = 1$, while for general n ρ_n is a random integer between 1 and n inclusive. If one plots the observations as points (n, X_n) in a plane, ρ_n is the number of points in the 'northwest corner' subtended by (n, X_n), edges included. In Figure 2.1, for instance, $\rho_1 = 1$, $\rho_2 = 2$ and $\rho_3 = 1$, while because $X_1 = X_4 = X_5$ we have $\rho_4 = 3$ and $\rho_5 = 4$.

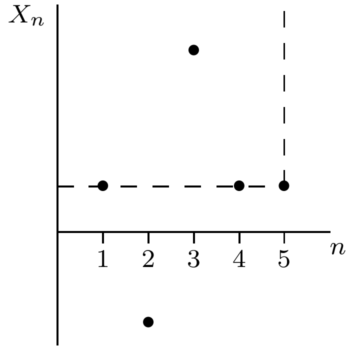

Figure 2.1. Initial ranks when observation values can coincide

Notice that if $X_m = X_n$ for some $m \neq n$ then ρ_m and ρ_n must be unequal. So for any chosen initial rank there can be no two coincident observations. Therefore we may unambiguously define

$$\mathcal{R}_k := \{X_n : \rho_n = k\}.$$

Theorem 2.1 holds in complete generality for the partial records thus defined. It need not be the case, however, that $\mathcal{R}_1$, or other $\mathcal{R}_k$, are infinite sets. To see this, let F be the common distribution function (d.f.) of the X_i and let $x_+ \in (-\infty, \infty]$ be its 'top':

$$x_+ := \sup\{x : F(x) < 1\}.$$

If x_+ is finite and $P(X_1 = x_+) = 1 - F(x_+ - 0) > 0$ then sooner or later there will be an observation with value x_+, and so $\mathcal{R}_1$ is a finite set with greatest element x_+. The same holds for $\mathcal{R}_2$, $\mathcal{R}_3$, $\ldots$. In all other cases it can be shown that each $\mathcal{R}_k$ is an infinite set that can be listed as a strictly increasing sequence that tends to x_+, as in (2.1).

The 'surprise' element in Ignatov's Theorem arises when one realizes that the constraints on the $(n+1)^{\text{st}}$ observation, if it is to be a 2-record, very much depend on the record values, as they happen to stand at time n. In a sense the theorem works

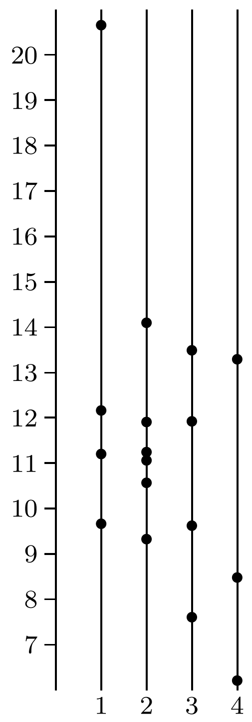

Figure 2.2. The 1-, 2-, 3- and 4-record values for Vancouver rainfall

because time has been wholly written out of its formulation. It looks only at the *traces* of the partial record values.

For illustration we display the 1-, 2-, 3- and 4-records from some actual data which may be assumed i.i.d. We take from Glick (1978), Data Set 1, the rainfall amounts in Vancouver in January, in inches for the years 1906 to 1966 inclusive. The partial records are shown in Figure 2.2. Were later observations included, more values might be added above the marks shown, but the traces of the 1- to 4-records up to level 13·28 are complete. If the original observations are indeed i.i.d., these are realizations of i.i.d. point processes (random sets).

The above illustration brings out the fact, known from theory, that even in quite long sequences of observations there are unlikely to be enough actual records to use for worthwhile statistical inference. By considering 1-, 2-, …, m-records, where m is chosen so that the expected number of partial records is as large as required, this problem of lack of data is overcome. The importance of Ignatov's Theorem is that models for the record values, which we will come to below, apply also to the partial record values and may thus be tested by inference based on all of them.

3. Record values

In this section we shall characterize the random set $\mathcal{R}_1$ of record values. Ignatov's Theorem 2.1 carries this characterization over to $\mathcal{R}_k$, the set of k-record values, for each k.

Recall that we let F denote the common d.f. of the observations X_i. It is convenient to extend F to set arguments, as in

$$F(x,\infty) = 1 - F(x) = P(X_1 > x),$$

$$F(a,b] = F(b) - F(a) = P(a < X_1 \le b),$$
$$F\{x\} = F(x) - F(x-0) = P(X_1 = x),$$

and so on. Let D be the set of points where F is discontinuous; then $F\{x\} = P(X_1 = x)$ is positive precisely when $x \in D$.

The simplest characterization of $\mathcal{R}_1$ is when F is discrete, so the X_i take values in the finite or countable set D. Suppose that each point $x \in D$ is owned by an individual demon, who decides with probability $p_x := F\{x\}/F[x,\infty)$ that x shall be an element of $\mathcal{R}_1$, and with probability $1 - p_x := F(x,\infty)/F[x,\infty)$ that it shall not. The demons do not communicate with each other, so their decisions are mutually independent. This describes $\mathcal{R}_1$ completely (as in this case it must obviously be a subset of D).

Before leaving the discrete case it is worth convincing oneself that, for any interval $(a,b]$,

$$P\big(\mathcal{R}_1 \cap (a,b] = \emptyset\big) = F(b,\infty)/F(a,\infty). \tag{3.1}$$

F does not change except at points of D, so for instance if $(a,b]$ contains just the points $x_1 < x_2 < x_3$ of D then

$$\begin{aligned} P\big(\mathcal{R}_1 \cap (a,b] = \emptyset\big) &= \frac{F(x_1,\infty)}{F[x_1,\infty)}\frac{F(x_2,\infty)}{F[x_2,\infty)}\frac{F(x_3,\infty)}{F[x_3,\infty)} \\ &= \frac{F[x_2,\infty)}{F(a,\infty)}\frac{F(x_2,\infty)}{F[x_2,\infty)}\frac{F(b,\infty)}{F(x_2,\infty)} = \frac{F(b,\infty)}{F(a,\infty)}, \end{aligned}$$

with a similar telescoping for more or fewer points of D. The argument extends even to cases when D is very thickly spread—it can be for instance be the set of rationals.

The other case where $\mathcal{R}_1$ is quite simple to characterize is when F is continuous, so D is empty. Here (3.1) is used as definition rather than conclusion. In fact, define a measure η on $\mathbb{R}$ by its 'distribution function'

$$\eta(-\infty,x] := -\ln F(x,\infty) \qquad (x \in \mathbb{R}), \tag{3.2}$$

and insist that for any (measurable) set $A \subseteq \mathbb{R}$,

$$P(\mathcal{R}_1 \cap A = \emptyset) = e^{-\eta(A)}. \tag{3.3}$$

In this, (3.1) is the special case when $A = (a,b]$. We call η the *avoidance measure* for $\mathcal{R}_1$, because through (3.3) it specifies the probability that $\mathcal{R}_1$ avoids any particular set. Because η, being a measure, is countably additive, if we fix disjoint sets A_1, A_2, … then (3.3) implies that the events that $\mathcal{R}_1$ has no elements in common with A_1, A_2, … are independent of each other (just as for the discrete case because of the independence of the demons). This suggests $\mathcal{R}_1$ might be a Poisson process, and indeed in this case it is. When F is continuous, $\mathcal{R}_1$ is the random set of points of an (inhomogeneous) Poisson process on $\mathbb{R}$ of intensity measure η. So $\mathcal{R}_1$ puts independent numbers of elements in disjoint subsets of $\mathbb{R}$, and the number of elements of $\mathcal{R}_1$ in any set A is Poisson-distributed with parameter $\eta(A)$.

The general case is nothing more than the two special cases put together. Thus define η by (3.2) and observe that the points x where $\eta(-\infty,x]$ is discontinuous, i.e.

the locations of the 'atoms' of the measure η, are precisely the points where F is discontinuous, that is, they are the points of the set D. The continuous part of η is the measure η_c given by

$$\eta_c(A) := \eta(A) - \sum_{x \in D \cap A} \eta\{x\}$$

on sets A with $\eta(A) < \infty$, and by

$$\eta_c(A) := \lim_{n \to \infty} \eta_c(A_n)$$

otherwise, where $A_n \uparrow A$ and $\eta(A_n) < \infty$ for all n.

Theorem 3.1 (Shorrock's Theorem). *$\mathcal{R}_1$ consists of*

1) *elements of D provided by independent demons: the demon who owns $x \in D$ decides with probability $1 - e^{-\eta\{x\}}$ that it shall be an element of $\mathcal{R}_1$, and with probability $e^{-\eta\{x\}}$ that it shall not;*
2) *the points of a Poisson process, independent of the demons, of intensity η_c.*

It is immediate from this theorem that $\mathcal{R}_1$ retains the independent increments property of each of its two components:

$\mathcal{R}_1$ has independent numbers of elements in disjoint subsets of $\mathbb{R}$.

The construction of $\mathcal{R}_1$ given by the above theorem is explicit, but in case it seems rather complicated we will give two concise characterizations which are equivalent to the construction. The first is as follows.

Theorem 3.2. *The distribution of $\mathcal{R}_1$ is the unique distribution of a random set such that (3.3) holds for all finite unions of intervals A, where the avoidance measure η is determined by (3.2).*

(By a 'random set' we mean, technically, a *simple* point process, i.e. one without multiple points; cf. Daley and Vere-Jones (1988).)

The above characterization brings out that even in the general case, η as determined by (3.2) still gives the *avoidance probabilities* for $\mathcal{R}_1$ via (3.3). Indeed, (3.3) holds for all Borel sets A, though the smaller class of sets mentioned in the theorem suffices for the characterization.

The second characterization is in terms of the function N_1, where $N_1(x)$ is defined to be the number of elements of $\mathcal{R}_1$ in the set $(-\infty, x]$. By a 'counting function' we mean a right-continuous function $f : \mathbb{R} \to \mathbb{R}$ which is flat except for jumps of height exactly $+1$, and for which $\lim_{x \downarrow -\infty} f(x) = 0$.

Theorem 3.3. *Define $H : \mathbb{R} \to \mathbb{R}$ by*

$$H(x) := \int_{(-\infty, x]} \frac{dF(y)}{F[y, \infty)} \qquad (x \in \mathbb{R}). \tag{3.4}$$

The distribution of N_1 is the unique distribution of a random counting function such that $M := N_1 - H$ is a martingale.

(If $x_+ < \infty$, read (3.4) as saying that H is constant in the interval $[x_+, \infty)$.)

For continuous-parameter martingales see e.g. Rogers and Williams (1994). In martingale terminology N_1 has a deterministic *compensator* which is given by H. In a terminology borrowed from reliability theory, H defined by (3.4) is the (cumulative) *hazard function* of F. If F has density f then (3.4) implies that $H(x) = \int_{-\infty}^{x} h(u)\, du$ where $h(x) := f(x)/(1 - F(x))$, that is, h is the *hazard rate* of F.

$H(x)$ is the expected number of elements of $\mathcal{R}_1$ in the set $(-\infty, x]$, and this suggests, on letting H denote also the measure determined by the function H, that $H(A) = \int_A dH(x)$ is the expected number of elements of $\mathcal{R}_1$ in the set A. This is indeed the case, so H is the *intensity* measure of the point process whose points are the elements of $\mathcal{R}_1$.

What is the relation between the intensity measure H and the avoidance measure η of (3.2) and (3.3)? Now H, like η, has atoms sited at the points of the set D, and a remaining continuous part H_c. In terms of these, for any Borel set $A \subseteq \mathbb{R}$,

$$H(A) = \eta_c(A) + \sum_{x \in D \cap A} (1 - e^{-\eta\{x\}}), \tag{3.5}$$

$$\eta(A) = H_c(A) + \sum_{x \in D \cap A} \big(-\ln(1 - H\{x\})\big). \tag{3.6}$$

So η_c and H_c are the same as each other, and hence when F is continuous the measures η and H are the same as each other.

From (3.3) and (3.6) we get

$$P(\mathcal{R}_1 \cap A = \emptyset) = e^{-H_c(A)} \prod_{x \in D \cap A} (1 - H\{x\}). \tag{3.7}$$

For each $n \in \mathbb{N}$, let $(E_j^{(n)})_{j=1,\ldots,n}$ be a partition of $\mathbb{R}$ into n disjoint intervals. Suppose that as $n \to \infty$ the maximum length of interval in the n^{th} partition tends to 0. Then it turns out that for sets A bounded away from x_+,

$$P(\mathcal{R}_1 \cap A = \emptyset) = \lim_{n \to \infty} \prod_{j=1}^{n} \big(1 - H(E_j^{(n)} \cap A)\big). \tag{3.8}$$

In fact it is not hard to show that the right-hand side of (3.8) reduces to that of (3.7). The limit procedure in (3.8) is rather like defining a Riemann integral, and, sure enough, the right-hand sides of (3.7–3.8) are alternative definitions of the *product integral* $\prod_A (1 - dH)$. Product integrals in this form have only recently been revived and codified (see Gill and Johansen (1990)). They lead to a nice extension of (3.7–3.8) for any measurable $\phi : \mathbb{R} \to [0, 1)$,

$$E \prod_{x \in \mathcal{R}_1} \big(1 - \phi(x)\big) = \prod (1 - \phi\, dH), \tag{3.9}$$

the right-hand side being a product integral while inside the expectation on the left-hand side there is just an infinite product. We can rewrite the product integral as a

D-indexed product times the exponential of an ordinary integral, similarly to (3.7), and deduce an inelegant formula for the Laplace functional of $\mathcal{R}_1$ (cf. Goldie and Rogers (1984), (2.3)).

Taking A in (3.7) to be $(-\infty, x]$, the left-hand side becomes $1 - F(x)$, so we have a formula inverting (3.4). When F is continuous this becomes simply $1 - F(x) = e^{-H(x)}$, inverting $H(x) = -\ln\bigl(1 - F(x)\bigr)$.

We close this section with history. The identification of $\mathcal{R}_1$, Theorem 3.1, is due to Shorrock (1972, 1974). The characterization in Theorem 3.2 relies on a standard result that simple point processes are characterized by their avoidance probabilities on a 'dissecting ring' of sets (see Daley and Vere-Jones, 1988, Theorem 7.3.II), and the characterization in Theorem 3.3 is from Goldie and Rogers (1984). In view of the different contributions to $\mathcal{R}_1$ made by the discrete and continuous parts of F it is not surprising that Theorem 2.1 took some while to reach its present formulation. The case with F continuous, when the $\mathcal{R}_k$ are i.i.d. Poisson processes, was formulated independently in Deheuvels (1983) and Ignatov (1978). Various different proofs appeared (see Mathematical Reviews 87j:60074 for a list). The general case of the result was given in Goldie and Rogers (1984). Finally, W. Vervaat realized that the case of F discrete (so only the Bernoulli 'demons' are present) has an easier proof than other cases, and this insight together with the (subtle) deduction of the general case from the discrete case appeared in Engelen *et al.* (1988). The proof of the general case thus given, with a minor compression by Rogers (1989), remains the shortest known. Subsequently, two further proofs have appeared in Samuels (1992) and Yao (1997), possessing aspects of interest.

4. Sojourns

In Section 1 we defined the record times but time has been rigorously excluded from the subsequent two sections. Here we will bring it back into the picture.

The first task, though, is to fix a positive integer m and arrange the elements of the sets $\mathcal{R}_1, \ldots, \mathcal{R}_m$ in one list in increasing order, obtaining a sequence which we denote $Q^{(m)}$, with elements $Q_j^{(m)}$:

$$Q^{(m)} := \bigl(Q_j^{(m)}\bigr)_{j=1,2,\ldots} = (Q_1^{(m)}, Q_2^{(m)}, \ldots),$$

where $Q_j^{(m)} \le Q_{j+1}^{(m)}$ for all j. Here we are notating $Q^{(m)}$ as an infinite sequence, which is the case so long as F has no *atom at the supremum of its support.* In the contrary case, when F has an atom at the supremum of its support, i.e. $x_+ < \infty$ and $F(x_+) < 1$, $Q^{(m)}$ is with probability 1 a finite sequence and minor notational changes are needed. Note that the sequence $Q^{(m)}$ can have repeated values, for if the same value appears in several of the $\mathcal{R}_i$ then it goes in the list as many times as it appears. However if F is continuous $Q^{(m)}$ has no repeats and is the sequence of points of a Poisson process on $\mathbb{R}$ of intensity $m\eta$, where the avoidance measure η comes from (3.2). In the general case $Q^{(m)}$ is a little more complicated but its distribution is known from Ignatov's and Shorrock's Theorems 2.1 and 3.1, and one can still do calculations with it without real difficulty.

Denote the observations $X_1, \ldots, X_n$ rearranged in increasing order by $X_n^{(n)} \le \cdots \le X_n^{(2)} \le X_n^{(1)}$. These are the order statistics and we want to follow the progress of the m^{th} order statistic $X_n^{(m)}$ as the sample size n increases through values m, $m+1$, $\ldots$. The values then taken by $X_n^{(m)}$ are precisely the elements of the sequence $Q^{(m)}$, in order but with more repeats. In fact the signal for the m^{th} order statistic to take a step through the sequence $Q^{(m)}$ is when an observation of initial rank at most m appears. These times are, in terms of the initial ranks ρ_n defined in (2.2),

$$\begin{aligned} L_1^{(m)} &:= m, \\ L_{j+1}^{(m)} &:= \inf\{n : n > L_j^{(m)},\ \rho_n \le m\} \qquad (j = 1, 2, \ldots), \end{aligned} \tag{4.1}$$

where we make the convention $\inf \emptyset := +\infty$. The m^{th} order statistic $X_n^{(m)}$ has value $Q_j^{(m)}$ for all n satisfying $L_j^{(m)} \le n < L_{j+1}^{(m)}$. The time that the m^{th} order statistic spends at the value $Q_j^{(m)}$, its 'sojourn' in that value, is thus

$$\Delta_j^{(m)} := L_{j+1}^{(m)} - L_j^{(m)} \qquad (j = 1, 2, \ldots). \tag{4.2}$$

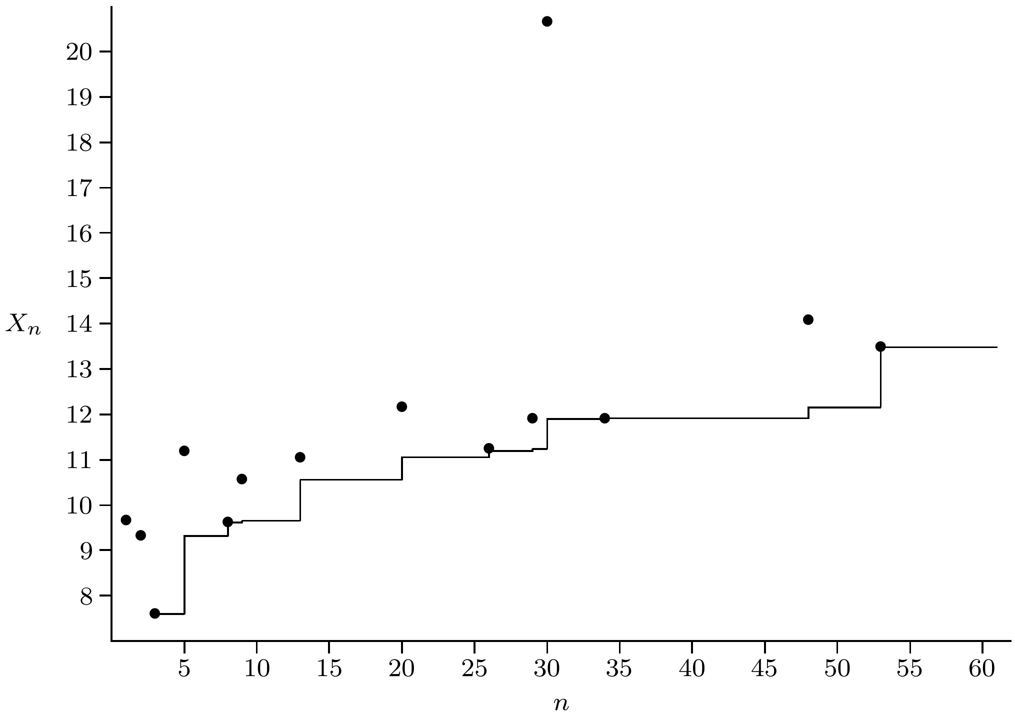

Figure 4.1. The 3rd order statistic for Vancouver rainfall

In Figure 4.1 we plot the 3rd order statistic ($m = 3$) for the data used for Figure 2.2. The observations having initial rank at most 3 are plotted as (n, X_n). Each such X_n eventually becomes a value taken by the third order statistic, with a sojourn that is shown as a horizontal line segment at the same height as X_n. There are no coincident values here, so to illustrate this effect we use in Figure 4.2 the precipitation data for Vancouver in July in *reverse time order* (1966 back to 1906) and again display the 3rd order statistic. The times $n = 1$ through to $n = 61$ correspond to years 1966 down to 1906. One can see that the two values of 3·36, which occurred at times $n = 1$ and

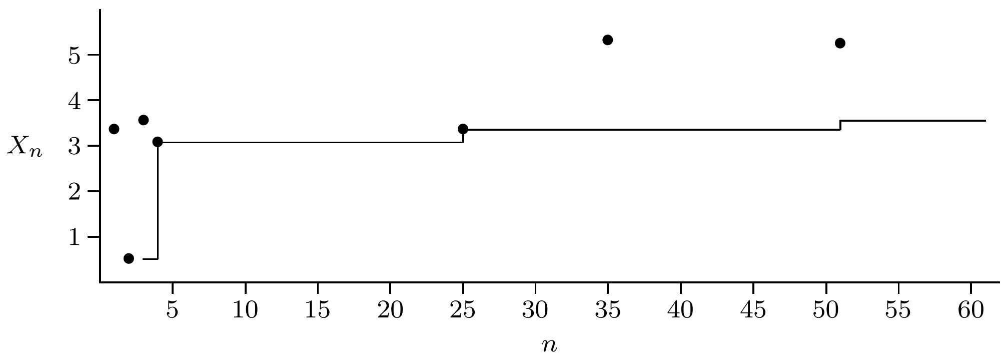

Figure 4.2. Vancouver in July in reverse time: 3rd order statistic

$n = 25$, give rise to end-to-end sojourns from $n = 25$ to $n = 35$ and from $n = 35$ to $n = 51$, at the same height.

The point of all the definitions above is to state the following.

Theorem 4.1. *The sojourns $\Delta_1^{(m)}$, $\Delta_2^{(m)}$, $\ldots$ are conditionally independent, given $Q^{(m)}$, with geometric distributions*

$$P(\Delta_j^{(m)} = l | Q^{(m)}) = \big(F(Q_j^{(m)})\big)^{l-1}\big(1 - F(Q_j^{(m)})\big) \qquad (l = 1, 2, \ldots). \tag{4.3}$$

What this means is that one can evaluate the probability of any event involving the $\Delta_j^{(m)}$, conditionally on the whole sequence $Q^{(m)}$. For instance

$$P(\Delta_j^{(m)} > k,\ \Delta_{j+1}^{(m)} > l | Q^{(m)}) = \big(F(Q_j^{(m)})\big)^k \big(F(Q_{j+1}^{(m)})\big)^l. \tag{4.4}$$

We do not have to condition on the whole sequence $Q^{(m)}$, for we can throw away from the conditioning anything that is not involved in the right-hand side ('redundant conditioning'). Thus $P(\Delta_j^{(m)} > k,\ \Delta_{j+1}^{(m)} > l | Q_j^{(m)}, Q_{j+1}^{(m)})$ also evaluates as the right-hand side of (4.4). With the information in Theorems 2.1 and 3.1 one thus has the joint distribution, for m fixed, of the $\Delta_j^{(m)}$ and the $Q_j^{(m)}$, for all $j = 1, 2, \ldots$.

Recall from Section 3 that the random set $\mathcal{R}_1$ has independent numbers of elements in any chosen disjoint sets. By Ignatov's Theorem 2.1, the same holds for every $\mathcal{R}_k$ and their independence makes the sequence $Q^{(m)}$ inherit an analogous property. Theorem 4.1 then implies that *the sojourns of the m^{th} order statistic in any chosen disjoint sets are independent random variables*. So for instance the number of n for which $2 \le X_n^{(m)} < 3$ is independent of the number of n for which $3 \le X_n^{(m)} < 4$. What we are really talking about here is the inverse of the m^{th} order-statistic process. So let us define $X^{(m)\leftarrow}(\cdot)$ to be a version of the inverse:

$$X^{(m)\leftarrow}(x) := \inf\{n : n \ge m,\ X_n^{(m)} \ge x\} \qquad (x \le x_+),$$

This defines $X^{(m)\leftarrow}(\cdot)$ to have *left-continuous* paths, ensuring the convenient relationship

$$X^{(m)\leftarrow}(x) \le n \quad \text{iff} \quad x \le X_n^{(m)}.$$

It is natural to consider increments of $X^{(m)\leftarrow}(\cdot)$ of the form

$$X^{(m)\leftarrow}[x,y) := X^{(m)\leftarrow}(y) - X^{(m)\leftarrow}(x) \qquad (x < y \le x_+),$$

as $X^{(m)\leftarrow}[x,y)$ is then *the amount of time the m^{th} order statistic spends in the interval* $[x,y)$:

$$X^{(m)\leftarrow}[x,y) = \#\{n : X_n^{(m)} \in [x,y)\}.$$

We may define $X^{(m)\leftarrow}I$ for other forms of interval I by taking right-hand limits. Thus for instance

$$X^{(m)\leftarrow}\{x\} = X^{(m)\leftarrow}(x+) - X^{(m)\leftarrow}(x) = \#\{n : X_n^{(m)} = x\}.$$

Theorem 4.2. *Fix $m \in \mathbb{N}$. The process $X^{(m)\leftarrow}$ has independent increments: for any disjoint intervals $I_1, \ldots, I_k$ in $(-\infty, x_+]$,*

$$P(X^{(m)\leftarrow}I_1 = n_1, \ldots, X^{(m)\leftarrow}I_k = n_k) = \prod_{j=1}^{k} P(X^{(m)\leftarrow}I_j = n_j) \qquad (n_1, \ldots, n_k \in \mathbb{N}_0),$$

where $\mathbb{N}_0 := \{0,1,2,\ldots\}$, and

$$P(X^{(m)\leftarrow}[x,y) = n) = \\ = \begin{cases} \left(\frac{F[y,\infty)}{F[x,\infty)}\right)^m, & \text{for } n = 0, \\ \left(\frac{F[y,\infty)}{F[x,\infty)}\right)^m \sum_{k=1}^{m\wedge n} \binom{m}{k}\binom{n-1}{k-1} F[x,y)^k F(-\infty,y)^{n-k}, & \text{for } n = 1,\ 2,\ \ldots. \end{cases} \tag{4.5}$$

The notation $m \wedge n$ stands for $\min(m,n)$. We emphasize that this Theorem holds for all F, without continuity assumptions. In view of the complicated final formula it is useful to understand how it can be directly derived:

Proof of (4.5). For $n = 0$, notice that $X^{(m)\leftarrow}[x,y) = 0$ if and only if all of the first m of the X_i that are at least x are also at least y. Now the observations that are at least x have conditional d.f. $G(u) := F(u)\mathbf{1}_{u \ge x}/F[x,\infty)$, so each has probability $G[y,\infty) = F[y,\infty)/F[x,\infty)$ of being at least y, and (4.5) for $n = 0$ follows.

For $n \ge 1$, consider the time when the m^{th} observation of value at least x occurs. Out of these m observations of value at least x, we need k to be less than y, where $1 \le k \le m$. This has probability

$$\binom{m}{k}\left(\frac{F[x,y)}{F[x,\infty)}\right)^k \left(\frac{F[y,\infty)}{F[x,\infty)}\right)^{m-k},$$

and accounts for 1 time unit spent by the m^{th} order statistic in the interval $[x,y)$. For it to have sojourn time n there, (a) of the next $n-1$ observations, exactly $n-k$ must be less than y, and (b) the next observation after those $n-1$ must be at least y. Event (a) has probability $\binom{n-1}{k-1}F[y,\infty)^{k-1}F(-\infty,y)^{n-k}$ and necessitates $k \le n$, while event (b) has probability $F[y,\infty)$. Assembling these contributions together we find

$$P(X^{(m)\leftarrow}[x,y) = n) = \\ = \sum_{k=1}^{m\wedge n} \binom{m}{k}\left(\frac{F[x,y)}{F[x,\infty)}\right)^k \left(\frac{F[y,\infty)}{F[x,\infty)}\right)^{m-k} \binom{n-1}{k-1} F[y,\infty)^{k-1} F(-\infty,y)^{n-k} F[y,\infty),$$

which tidies up to give the claimed formula. $\square$

The m^{th} order statistic spends a sojourn in each element of the sets $\mathcal{R}_1, \ldots, \mathcal{R}_m$, and we may attach these sojourn times as marks to the respective elements of the sets. The resulting marked point processes *retain* the independence that Ignatov's Theorem 2.1 establishes for $\mathcal{R}_1, \ldots, \mathcal{R}_m$. When F is continuous the probabilistic structure of the marked point processes can be simply described, and this is best done by splitting each of $\mathcal{R}_1, \ldots, \mathcal{R}_m$ into separate (unmarked) point processes according to the values of the marks. We are led to the following result, due to Goldie and Rogers (1984) and Stam (1985).

Theorem 4.3. *For $i = 1, \ldots, m$ let ${}^m\mathcal{R}_i$ be the marked point process obtained from $\mathcal{R}_i$ by marking each element with the sojourn that the m^{th} order statistic spends in that element. Then ${}^m\mathcal{R}_1, \ldots, {}^m\mathcal{R}_m$ are independent and identically distributed.*

For $l = 1, 2, \ldots$ let ${}^m\mathcal{R}_i^{(l)}$ be the point process consisting of all elements of ${}^m\mathcal{R}_i$ whose marks have value l. If F is continuous then ${}^m\mathcal{R}_i^{(l)}$ are independent Poisson processes, for $l \in \mathbb{N}$ and $i = 1, \ldots, m$. The (cumulative) intensity function of ${}^m\mathcal{R}_i^{(l)}$, i.e. the expected number of points it has in the interval $(-\infty, x]$, is

$$E\big({}^m\mathcal{R}_i^{(l)}(-\infty, x]\big) = \big(F(x)\big)^l/l \qquad (x \in \mathbb{R}).$$

To understand where the latter formula comes from, recall that by Shorrock's Theorem 3.1, $\mathcal{R}_i$ is identically distributed to $\mathcal{R}_1$, and that by Theorem 3.3 it is, when F is continuous, a Poisson process with intensity $dH(x) = dF(x)/F[x, \infty)$. By Theorem 4.1, therefore, ${}^m\mathcal{R}_i^{(l)}$ is a Poisson process with intensity

$$\big(F(x)\big)^{l-1}\big(1 - F(x)\big)dH(x) = \big(F(x)\big)^{l-1}dF(x),$$

and so

$$E\big({}^m\mathcal{R}_i^{(l)}(-\infty, x]\big) = \int_{-\infty}^{x} \big(F(u)\big)^{l-1}dF(u) = \big(F(x)\big)^l/l.$$

5. Record times

The times $L_j^{(m)}$, defined in (4.1), can be expressed by (4.2) as

$$L_n^{(m)} = m + \sum_{i=1}^{n-1} \Delta_i^{(m)} \qquad (m, n = 1, 2, \ldots), \tag{5.1}$$

where we interpret a sum such as $\sum_{i=1}^{0}$ as zero. This gives in principle a means to model the $L_j^{(m)}$, since by Theorem 4.1 we have the distribution of the $\Delta_j^{(m)}$ in terms of the partial-record-value sequence $Q^{(m)}$. When F is continuous we can cut through this complicated structure, since the times become distribution-free, as we shall see. Specifying F continuous still allows the supremum x_+ of its support to be finite, but the continuity means in particular that F can have no jump there, so $\lim_{x\uparrow x_+} F(x) = 1$ in all cases. The basis of all the simplifications in the F-continuous case is the following celebrated insight of Dwass (1960, Theorem 1) and Rényi (1962, p. 56) concerning the initial ranks ρ_n defined in (2.2).

Lemma 5.1 (Dwass-Rényi Lemma). *If F is continuous then $\rho_1, \rho_2, \rho_3, \ldots$ are independent and for each n ρ_n is uniformly distributed on $\{1,\ldots,n\}$.*

Uniform distribution here is of course *discrete uniform*, with $P(\rho_n = j) = 1/n$ for $j = 1, \ldots, n$.

When F is continuous the times $L_n^{(m)}$ defined in (4.1) become exactly the times when the m^{th} order statistic first exists and then changes value. Several characterizations/representations/constructions of the whole sequence $L^{(m)} := (L_n^{(m)})_{n\in\mathbb{N}}$ then flow from the Dwass-Rényi Lemma 5.1. First, define the m^{th} *record counting sequence* $N^{(m)} = (N_n^{(m)})_{n\geq m}$ by

$$N_n^{(m)} := \sum_{k=m}^{n} \delta_k^{(m)} \qquad (n = m, m+1, \ldots), \tag{5.2}$$

where $\delta_k^{(m)} := \mathbf{1}\{\rho_k \leq m\}$; then we have a representation as an inverse of $N^{(m)}$:

$$L_n^{(m)} = \inf\{j : j \geq m,\ N_j^{(m)} \geq n\} \qquad (n = 1, 2, \ldots).$$

The point is that $N^{(m)}$ has a very simple structure that does not depend on F, so long as F is continuous: the summands $\delta_m^{(m)}$, $\delta_{m+1}^{(m)}$, $\ldots$ in (5.2) are independent Bernoulli r.v.s with $P(\delta_k^{(m)} = 1) = m/k$.

The second characterization of $L^{(m)}$ is as a Markov chain.

Theorem 5.2. *Fix $m \in \mathbb{N}$ and assume F is continuous. Then $L^{(m)} = (L_n^{(m)})_{n\in\mathbb{N}}$ is a Markov chain on $\{m, m+1, \ldots\}$ with initial state $L_1^{(m)} = m$ and transition probabilities $p_{jk} = P(L_{n+1}^{(m)} = k | L_n^{(m)} = j)$ given by*

$$p_{jk} = m\frac{\prod_{l=0}^{m-1}(j-l)}{\prod_{l=0}^{m}(k-l)} \qquad (k = j+1, j+2, \ldots).$$

The case $m = 1$ merits a separate treatment, since the times $L_n^{(1)}$ are the record times L_n as defined in Section 1.

Corollary 5.3. *When F is continuous the sequence $(L_n)_{n\in\mathbb{N}}$ of record times is a Markov chain on $\mathbb{N}$ with initial state $L_1 = 1$ and transition probabilities $p_{jk} = P(L_{n+1} = k | L_n = j)$ given by*

$$p_{jk} = \frac{j}{(k-1)k} \qquad (k = j+1, j+2, \ldots). \tag{5.3}$$

We shall call the times $L_n^{(m)}$ the m^{th} *record times*. This is the usual nomenclature, although the label is occasionally assigned by some authors to the times $T_n^{(m)}$ when observations of initial rank *exactly* k occur. The latter times are worth some consideration. Thus put

$$\begin{aligned} T_1^{(m)} &:= \inf\{n : n \geq 1,\ \rho_n = m\}, \\ T_{j+1}^{(m)} &:= \inf\{n : n > T_j^{(m)},\ \rho_n = m\}, \qquad (j = 1, 2, \ldots), \end{aligned}$$

and note that $L_j^{(1)} = L_j = T_j^{(1)}$. From the Dwass-Rényi Lemma 5.1 it is clear that when $m > 1$ the sequence $(T_n^{(m)})_{n\in\mathbb{N}}$ is a Markov chain with the same transition probabilities as the record-time sequence $(L_n)_{n\in\mathbb{N}}$, but with a random initial state:

Proposition 5.4. *Fix $m \in \{2, 3, \ldots\}$. When F is continuous the sequence $(T_n^{(m)})_{n\in\mathbb{N}}$ is a Markov chain on $\{m, m+1, \ldots\}$ with initial distribution*

$$P(T_1^{(m)} = j) = \frac{m-1}{(j-1)j} \qquad (j = m, m+1, \ldots) \tag{5.4}$$

and transition probabilities $p_{jk} = P(T_{n+1}^{(m)} = k | T_n^{(m)} = j)$ given by (5.3).

For the next representation, due to Williams (1973), we employ the 'ceiling' function

$$\lceil x \rceil := \inf\{n : n \in \mathbb{Z},\ n \geq x\} \qquad (x \in \mathbb{R}).$$

Theorem 5.5. *Assume F continuous. Let E_2, E_3, ... be independent unit exponential r.v.s, and define successively*

$$\begin{aligned} \tilde{L}_1 &:= 1, \\ \tilde{L}_{n+1} &:= \lceil \tilde{L}_n e^{E_{n+1}} \rceil \qquad (n = 1, 2, \ldots). \end{aligned}$$

Then the sequence $(L_n)_{n\in\mathbb{N}}$ has the same distribution as the sequence $(\tilde{L}_n)_{n\in\mathbb{N}}$.

It is important to understand that this gives much more than just each L_n having the distribution of $\tilde{L}_n$. All *joint* distributions of the various L_n are those of the corresponding $\tilde{L}_n$.

We can easily carry this representation over to the times $T_j^{(m)}$ when observations of initial rank m occur. For put $T_0^{(m)} := m - 1$ and observe that the transition matrix (p_{jk}) of (5.3) then gives the correct distribution (5.4) for $T_1^{(m)}$. Thus we can consider $(T_n^{(m)})_{n\in\mathbb{N}_0}$ (where $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$) as a Markov chain with initial state $m-1$ and transition probabilities p_{jk}, and are led to the following.

Corollary 5.6. *Assume F continuous. Fix $m \in \{2, 3, \ldots\}$. Let E_1, E_2, ... be independent unit exponential r.v.s, and define successively*

$$\begin{aligned} \tilde{T}_0^{(m)} &:= m - 1, \\ \tilde{T}_{n+1}^{(m)} &:= \lceil \tilde{T}_n^{(m)} e^{E_{n+1}} \rceil \qquad (n = 0, 1, 2, \ldots). \end{aligned}$$

Then the sequence $(T_n)_{n\in\mathbb{N}_0}$ has the same distribution as the sequence $(\tilde{T}_n)_{n\in\mathbb{N}_0}$.

A similar representation to the above can be given for the m^{th} record times $L_n^{(m)}$ (Goldie, 1983, Theorem II.6.1) but it is somewhat more complicated.

The final representation that we select for inclusion here is one for record values and times together, but the representation that it gives just for record times is still valuable on its own. As remarked in Section 4, when F is continuous the sequence $Q^{(m)}$, consisting of all record values of orders up to m, is the sequence of points of a Poisson process of intensity $m\eta$ where η is the avoidance measure of (3.2). We can generate these from the points of a unit Poisson process by a form of 'probability integral transform'. Then given these Poisson points we can generate conditionally independent sojourns as in Theorem 4.1 by secondary randomizations, and finally add up the sojourns to give the m^{th}-record times by (5.1). The trick that does the secondary randomization is the following easily checked result.

Lemma 5.7. *Fix $0 < p < 1$, let E have a unit exponential distribution and set*

$$Y := \left\lceil \frac{E}{-\ln(1-p)} \right\rceil .$$

Then Y has the geometric distribution

$$P(Y = y) = (1-p)^{y-1} p \qquad (y = 1, 2, \ldots).$$

We are led to the following representation, due to Vervaat (1977) and Deheuvels (1984).

Theorem 5.8. *Fix $m \in \mathbb{N}$. Assume F continuous. Let $\varepsilon_1, \varepsilon_2, \ldots$ and $E_1, E_2, \ldots$ be independent unit exponential r.v.s, and put $S_n := \sum_{j=1}^{n} \varepsilon_j$ for $n = 1, 2, \ldots$. Define a function g (an inverse to the cumulative avoidance measure $\eta(-\infty, x]$) by*

$$g(y) := \inf\{x : -\ln F(x, \infty) \geq y\} \qquad (0 < y < \infty).$$

Set

$$\tilde{Q}_n^{(m)} := g(S_n/m),$$
$$\tilde{L}_n^{(m)} := m + \sum_{i=1}^{n-1} \left\lceil \frac{E_i}{-\ln(1 - e^{-S_i/m})} \right\rceil ,$$

both for $n = 1, 2, \ldots$. Then the double sequence $(Q_n^{(m)}, L_n^{(m)})_{n \in \mathbb{N}}$ has the same distribution as $(\tilde{Q}_n^{(m)}, \tilde{L}_n^{(m)})_{n \in \mathbb{N}}$.

6. Limit theory and strong approximation

We first summarize the distributional limit theory, discovered by Resnick in 1973, for the sequence (R_n) of record values; see Resnick (1987, §4.2) for a full account. By Ignatov's Theorem 2.1, each set $\mathcal{R}_k$ is probabilistically identical to the set $\mathcal{R}_1$ of records, whose elements form the sequence (R_n). Thus limit theory for 1–records applies equally to k–records (observations of initial rank k), for each k. Further, if F is continuous then the sequence $Q^{(m)}$ that we met in Section 4, of observations of initial rank *at most* m, is distributed as the sequence of record values from observations with d.f. F^m. Hence it too has the same distributional limit theory, modulo adjustments to formulae for domains of attraction and norming constants.

The idea is to find all limits in distribution of $(R_n - b_n)/a_n$ for suitable constants $b_n \in \mathbb{R}$ and $a_n > 0$. First note that if $G(x)$ is such a limit d.f., that is, $(R_n - b_n)/a_n \Rightarrow G$ where $\Rightarrow$ denotes weak convergence, then $G(ax + b)$ is another limit d.f., attained by replacing the constants a_n and b_n by aa_n and $b_n + ba_n$ respectively, where $a > 0$ and $b \in \mathbb{R}$. Thus the limit distributions are determined only to within *type*: they are equivalence classes under the operation of positive affine transformations $x \mapsto ax + b$ where $a > 0$. The possible limit types for record values are as follows, expressed in terms of the standard Normal d.f. $\Phi(x) := \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-u^2/2}\, du$.

Theorem 6.1 *(Resnick, 1973). Assume F continuous. Then the possible non-degenerate limit d.f.s for $(R_n - b_n)/a_n$, as $n \to \infty$, are those within the type of one of*

1) $\tilde{\Phi}_\alpha(x) := \begin{cases} 0, & \text{for } x \le 0, \\ \Phi(\ln x^\alpha), & \text{for } x \ge 0; \end{cases}$

2) $\tilde{\Psi}_\alpha(x) := \begin{cases} \Phi\big(\ln(-x)^{-\alpha}\big), & \text{for } x \le 0, \\ 1, & \text{for } x \ge 0; \end{cases}$

3) $\Phi(x)$,

where $\alpha > 0$ is a constant.

The limit distributions are closely associated with the three classical *extreme-value distributions*, which are the possible limits for normed and translated maxima of i.i.d. r.v.s; see Resnick (1987) for details. Note that $\tilde{\Phi}_1$ is the lognormal distribution.

Letting G stand for one of the limit d.f.s $\tilde{\Phi}_\alpha$, $\tilde{\Psi}_\alpha$ and Φ, we say that F is in the *domain of attraction for records of* G, notation $F \in \mathrm{DR}(G)$, if constants $a_n > 0$ and $b_n \in \mathbb{R}$ exist such that $(R_n - b_n)/a_n \Rightarrow G$. (As usual, (R_n) is the sequence of record values from i.i.d. observations with d.f. F.) There are several ways to express the domains of attraction and suitable scaling and location constants, but the following are the most explicit. Recall from Section 3 that when F is continuous the hazard function H is given by $H(x) = -\ln\big(1 - F(x)\big)$. Its inverse $H^{\leftarrow}$ may be defined by $H^{\leftarrow}(u) := \inf\{x : H(x) \ge u\}$ where $\inf \emptyset := +\infty$.

Theorem 6.2 *(Resnick, 1973). Let F be continuous.*

1) *$F \in \mathrm{DR}(\tilde{\Phi}_\alpha)$ if and only if $x_+ = \infty$ and $\surd H(x) = c(x) + \int_1^x d(t)t^{-1}\,dt$ where $c(x) \to c \in \mathbb{R}$ and $d(x) \to \frac{1}{2}\alpha$ as $x \to \infty$. In that case, suitable scaling and location constants are $a_n := H^{\leftarrow}(n)$, $b_n := 0$.*
2) *$F \in \mathrm{DR}(\tilde{\Psi}_\alpha)$ if and only if $x_+ < \infty$ and $\surd H(x_+ - u) = c(u) + \int_u^1 d(t)t^{-1}\,dt$ for $u > 0$, where $c(u) \to c \in \mathbb{R}$ and $d(u) \to \frac{1}{2}\alpha$ as $u \downarrow 0$. In that case, suitable constants are $a_n := x_+ - H^{\leftarrow}(n)$, $b_n := x_+$.*
3) *$F \in \mathrm{DR}(\Phi)$ if and only if*

$$\surd H(x) = c(x) + \int_{-\infty}^{x} \frac{dt}{f(t)} \qquad (x < x_+),$$

where $f(\cdot) > 0$, f' exists and, as $x \uparrow x_+$, $c(x) \to c \in \mathbb{R}$ and $f'(x) \to 0$. In that case, suitable constants are $a_n := H^{\leftarrow}(n + \surd n) - H^{\leftarrow}(n)$, $b_n := H^{\leftarrow}(n)$.

Turning from weak to almost-sure limit theory, the best approach has turned out to be to develop strong approximations in the sense of Komlós, Major and Tusnády (see, e.g., Csörgő and Révész, 1981), from which limit theorems may be readily extracted at will. Pfeifer and Zhang (1989) give a valuable survey. The representations of Theorems 5.5 and 5.8 are starting points for strong approximations. In Theorem 5.5, for large n the $\tilde{L}_n$ will be large and the ceiling operator $\lceil\cdot\rceil$ has a very small effect. Therefore $\tilde{L}_n/\tilde{L}_{n+1}$ is approximately $e^{-E_{n+1}}$. In Pfeifer (1987) this approximation is made explicit, being represented through a uniform r.v. defined on a further augmentation of the probability space. Thus the author sets up a sequence (W_n) of independent

Uniform$(0,1)$ r.v.s, which is also independent of (L_n), such that

$$e^{-E_{n+1}} = (1 - W_{n+1})\frac{L_n}{L_{n+1}} + W_{n+1}\frac{L_n}{L_{n+1} - 1} \qquad (n = 1, 2, \ldots).$$

This device, which is a particular case of a useful general technique for Markov chains, leads to the following strong approximation. We write Δ_n for $\Delta_n^{(1)} = L_{n+1} - L_n$, and a.s. for 'almost-surely'.

Theorem 6.3 *(Pfeifer, 1987). Assume F continuous. On a suitably augmented probability space there exist independent unit exponential r.v.s E_n, and a non-negative r.v. Z, such that with $S_n := \sum_{j=2}^{n} E_j$,*

1) $\ln L_n = Z + S_n + \mathrm{o}(1)$ *a.s. and* $\ln \Delta_n = Z + S_{n+1} + \ln(1 - e^{-E_{n+1}}) + \mathrm{o}(1)$ *a.s., as* $n \to \infty$;
2) Z *and* S_n *are asymptotically independent;*
3) $EZ = 1 - \gamma$ *where* $\gamma \doteq 0{\cdot}577216$ *is Euler's constant;*
4) $E(Z^k) < \infty$ *for all* $k \in \mathbb{N}$.

Note that the r.v.s $-\ln(1 - e^{-E_{n+1}})$ *appearing in 1) are independent* Uniform$(0,1)$.

One may further approximate the sequence (S_n) in this theorem by a Wiener process. By that means or by application of standard limit theory one can immediately deduce the following classical results for a sequence of r.v.s Z_n which can be either $\ln L_n$ or $\ln \Delta_n$ (or even $H(R_n)$):

$$\frac{Z_n}{n} \to 1 \text{ a.s.}, \qquad \frac{Z_n - n}{\sqrt{n}} \Rightarrow \mathrm{N}(0,1), \tag{6.1}$$

as $n \to \infty$, and the lim sup and lim inf of $(Z_n - n)\big/\sqrt{2n \ln\ln n}$ are respectively ± 1 a.s.

By applying the above to the sequence $(T_n^{(m)})$, which is governed by the same Markov chain as (L_n) (see Corollary 5.6), it is clear that the above theorem's results also hold for $(T_n^{(m)})$. This and further strong approximations of the above type, and their consequences, may be found in Deheuvels (1988).

We briefly discuss one further strong-approximation technique, due to Resnick, which uses the important construct of an F-*extremal process*, and also provides insight into the r.v. Z that appears in Pfeifer's Theorem 6.3. Consider a half-plane $(0,\infty) \times \mathbb{R}$ whose points (t,x) represent time $t > 0$ and space $x \in \mathbb{R}$. Assume F continuous, and let N be a planar Poisson process on this space, with mean measure defined by

$$EN\big((0,t] \times (x,\infty)\big) = t\big(-\ln F(x)\big).$$

The idea is that in any Borel set, the number of points of N is Poisson-distributed with mean the value of the mean measure for that set, and the numbers of points of N in disjoint sets are independent r.v.s. Notice that the value of the mean measure for any semi-infinite rectangle $(0,t] \times (x,\infty)$ is finite. So the number of points of N in the rectangle is finite a.s. However the mean measure of the infinite strip $(0,t] \times \mathbb{R}$ is infinite and so there will certainly be points of N in the strip. Therefore, with probability one, there will exist a point of N of maximal height. This is the value of the F-extremal process Y at time t. Formally,

$$Y_t := \sup\{x : N\big((0,t] \times (x,\infty)\big) > 0\}.$$

This defines the stochastic process $(Y_t)_{t>0}$ to have paths that change only by upward jumps, and have $\lim_{t\downarrow 0} Y_t = x_-$ and $\lim_{t\uparrow\infty} Y_t = x_+$, a.s. Properties of the process may be found in Resnick (1987, §4.3), but for present purposes two are important:

1) the jump times of (Y_t) form a logarithmic Poisson process on $(0,\infty)$, i.e. the mean number of jumps in any time interval $(s,t]$ is $\ln(t/s)$;
2) with (X_n) i.i.d. with d.f. F, the F-extremal process 'interpolates' the sequence of partial maxima of the X_i, in that

$$\big(\max(X_1,\ldots,X_n)\big)_{n\in\mathbb{N}} =_d (Y_n)_{n\in\mathbb{N}},$$

where $=_d$ means equality in distribution. The right-hand side is just the sequence of values of (Y_t) at *integer* times, and the equality in distribution here refers to the *whole sequences* on either side.

In Figure 6.1 we plot a simulated F-extremal process, with F the standard Normal distribution.

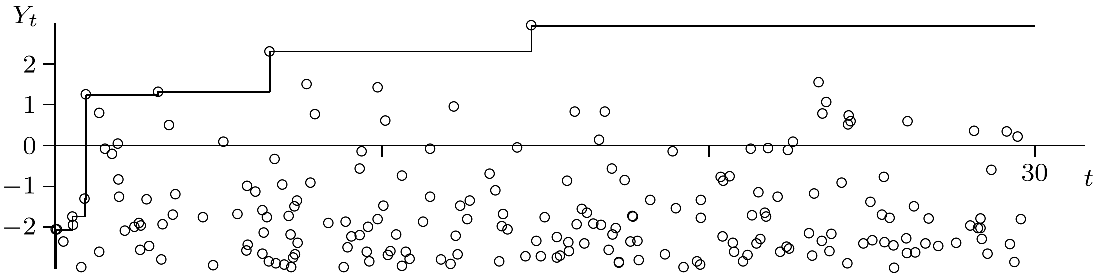

Figure 6.1. A simulated N(0,1)-extremal process, for $0 < t \leq 30$

For purposes of strong approximation, property 2) means that the jump-times of the sequence of maxima, i.e. the record times of the X_i, are shadowed, more and more closely as time increases, by the jump-times of (Y_t), while 1) means that the jump-times of (Y_t) form a sequence with a very simple probability distribution. The F-extremal process has an *excess* of jumps from time 1 onwards, compared to the number of record values, and the total excess S in the limit as $t \to \infty$ is a.s. finite and has mean $E(Z)$. The strong approximation is expressed, in terms of the jump times τ_1, τ_2, $\ldots$ of the F-extremal process from time 1 onwards, by

$$\ln L_n = \ln \tau_{n+S} + \mathrm{o}(1) = \ln \tau_n + \mathrm{O}(\ln n) \text{ a.s.}$$

as $n \to \infty$. The sequence $(\ln \tau_n)$ forms a unit (homogeneous) Poisson process on $(0,\infty)$, so its large-n behaviour can be read off, and conclusions such as (6.1), for $Z_n := \ln L_n$, derived anew.

7. Extensions, connections and applications

A number of variations on the classical i.i.d. record model have been proposed, involving various combinations of non- stationarity and dependence among the observations. The oldest appears to be the so-called F^α *model* (Yang, 1975). In this setup $(X_n)_{n\in\mathbb{N}}$ is a sequence of independent but *non*-identically distributed r.v.s with $X_n \sim F^{\alpha_n}$ for some fixed d.f. F and sequence $(\alpha_n)_{n\in\mathbb{N}}$ with $\alpha_n > 0$. Yang's original idea was that the α_n represented the population size and the observations X_n were themselves the maxima from increasingly large populations; he used this to model Olympic record-breaking. This model has subsequently been studied by several authors, especially Nevzorov (see Nevzorov, 1995, and references therein). Alternatively we may suppose that the distribution of the X_n changes only after a *record* event, as proposed by Pfeifer (see Pfeifer, 1984, and references therein); this may be useful if the system producing the observations is altered or 'repaired' after a record occurs. A still different non-stationary setup was studied by Ballerini and Resnick (1985), who supposed that $X_n = Y_n + cn$ for some stationary sequence (Y_n) and constant c. This model gave a good fit to a certain sports dataset (see also Feuerverger and Hall, 1996). On the other hand, it seems to be more difficult to deal with stochastic dependence among the observations: here we note that Ballerini (1994) considered a dependent F^α setup, and that Haiman and Puri (1990) studied m-records over stationary sequences with a certain dependence structure.

Interest in Olympic records has naturally led to statistical investigations of two questions: model stationarity and prediction of future records. There has long been a thread of significant papers on these topics: see Robinson and Tawn (1995) and references therein, and Smith (1997).

The distribution-free nature of the Dwass-Rényi Lemma 5.1 suggests a connection between records and combinatorics, and indeed there is a fruitful link with, in particular, the theory of *random permutations*. A permutation of the integers $1, \ldots, n$ is a bijective mapping from the set $\{1, \ldots, n\}$ to itself. There are $n!$ such mappings, and a *random* permutation generally means that all $n!$ are equally likely, so that each has probability $1/n!$ of being chosen. Much is known about random permutations and some of it translates into valuable results about records. We describe how one classical formula carries over in two distinct ways into the records context.

A consequence of the Dwass-Rényi Lemma is that if δ_k is the 'indicator' r.v. that is 1 if X_k is a record value, and 0 otherwise, then provided F is continuous the δ_k are independent Bernoulli r.v.s with $P(\delta_k = 1) = 1/k$. Knowledge of the δ_k is equivalent to being given the record times (L_n). The number of records among $X_1, \ldots, X_n$ is

$$N_n = \sum_{k=1}^{n} \delta_k \tag{7.1}$$

(the $m = 1$ case of (5.2)). Each record among $X_1, \ldots, X_n$ gives rise to a time interval starting with its time of occurrence L_n and ending just before the next record time L_{n+1}, or at time n, whichever comes first. Thus a set of 'inter-record time gaps' $L_2 - L_1$, $L_3 - L_2, \ldots, L_{N_n} - L_{N_n-1}$, $n + 1 - L_{N_n}$ is defined, where if $N_n = 1$ there is just one time gap $n + 1 - L_1 = n + 1 - 1 = n$. In all cases the time gaps add up to n. Now a celebrated formula about random permutations, translated over to records, says that the probability that there are a_j time gaps of length j, for $j = 1, \ldots, n$, is

$$p_n(a_1, \ldots, a_n) = \prod_{j=1}^{n} \frac{1}{j^{a_j} a_j!} \mathbf{1}\left\{ \sum_{j=1}^{n} j a_j = n \right\}. \tag{7.2}$$

This formula, due to Goncharov, is in fact a special case of the famous *Ewens sampling formula* of population genetics. In the general case, permutations are not equally likely but have probabilities proportional to θ^k where k is the number of cycles in the permutation and θ is a positive parameter. Our special case, when $\theta = 1$, is the only one to have been found to have meaning in records terms. However, even for that case, many limit theorems and large-sample approximations flow from (7.2): see, e.g., Arratia *et al.* (1992) and Arratia and Tavaré (1992) for details, as well as for the general case of (7.2).

The basis of the above link between records and permutations is the sequence of independent Bernoulli r.v.s appearing in (7.1), which may be used to build a random permutation through its cycle representation. There are several ways to construct the link (Goldie, 1989), and we describe one other realization of (7.2) which arises from one of them. Recall our notation $X_n^{(1)} \geq \cdots \geq X_n^{(n)}$ for the order statistics of X_1, $\ldots$, X_n. The indicator r.v.s $\mathbf{1}\{X_n^{(k)} \in \mathcal{R}_1\}$ are independent, for $k = 1, \ldots, n$, with $P(X_n^{(k)} \in \mathcal{R}_1) = 1/k$ (Goldie, 1989), and this fact suggests defining on them something analogous to the inter-record time gaps above. Thus let $M_1 < M_2 < \cdots < M_{N_n}$ be the list of indices i for which $X_n^{(i)} \in \mathcal{R}_1$, or formally

$$\begin{aligned} &M_1 := 1, \\ &M_j := \min\{k : k > M_{j-1},\ X_n^{(k)} \in \mathcal{R}_1\} \qquad (j = 2, \ldots, N_n). \end{aligned}$$

Define gaps $M_2 - M_1$, $M_3 - M_2$, $\ldots$, $M_{N_n} - M_{N_n - 1}$, $n+1-M_{N_n}$. Then the probability that there are a_j of these gaps of length j, for $j = 1, \ldots, n$, is given by (7.2).

The probability distribution of the number of records N_n up to time n, known as the *Karamata-Stirling law*, is best defined through the representation (7.1). An explicit formula for the probabilities involves Stirling numbers (David and Barton, 1962, pp. 181–2), but as Feller (1968, XI.2(c)) observes, the probability generating function is simpler:

$$E(s^{N_n}) = (-1)^n \binom{-s}{n}.$$

The Karamata-Stirling law occurs also in summability theory (Bingham, 1988).

While most of the theory of records depends critically on the order properties of the real line, some aspects permit generalization to partially ordered sets (Goldie and Resnick, 1989). For instance, with several of the possible extensions of the notion of record, criteria can be given for there being finitely or infinitely many records in chosen regions, including the region that is the whole space. When the space is the Euclidean plane it is typical that there are only finitely many bivariate records (records in both components simultaneously), that being the case for example when the observations are drawn from any bivariate Normal distribution with correlation not equal to +1 (Gnedin, 1994). However one can estimate (Goldie and Resnick, 1995) the small probability that there will be a large number of bivariate records in chosen regions in $\mathbb{R}^2$, and find where they will fall. A complementary result is in Deuschel and Zeitouni (1995), and it is interesting that the authors of this article establish a new connection with random permutations, different from those discussed above, namely an extension of a celebrated theorem of Vershik and Kerov on longest increasing subsequences.

8. Random record processes

Imagine now that the observations X_1, X_2, ... occur at *random times* T_1, T_2, ... In this case the process of successive maxima of the X_n is called a *random record process*, 'random' referring here to the arrival times of the X_n not their values. The sequence T_1, T_2, ... is sometimes called the 'pacing process' and we speak, for example, of 'Poisson paced records.' Much less work has been done on this process than on the classical record process discussed above; the first paper appeared only in 1971 (Pickands, 1971). Nevertheless the random record process has been shown to have deep and sometimes surprising connections to other parts of probability theory. This is the perspective we continue to take in this survey: we discuss known results for three classes of pacing processes (Poisson, pure birth, and renewal), where possible with a view toward more general theory.

More formally, then, let $(T_n, X_n)_{n\in\mathbb{N}}$ denote a bivariate random sequence. We suppose that $0 < T_1 < T_2 < \ldots$ so that marginally $(T_n)_{n\in\mathbb{N}}$ forms a point process on $[0, \infty)$, but for the moment we make no further assumptions. Then the random record process is the bivariate sequence of record times and values, that is $(T_{L_n}, X_{L_n})_{n\in\mathbb{N}} = (T_{L_n}, R_n)_{n\in\mathbb{N}}$. This is the most general formulation under which the random record process retains its particular identity: from this perspective it can be regarded as a *thinning* of a *marked point process* where the points T_n are marked by the observations X_n. There is now a well-developed martingale-based theory of thinned marked point processes (Last and Brandt, 1995), so it is reasonable to ask what this theory can tell us about random record processes. However, the resources of that theory have not yet been fully brought to bear on the random record problem, for two main reasons. The first is historical: martingale-based point processes and random record processes come from different streams in probability and the two have not yet completely converged. This is exacerbated by the second reason: although the mark sequence (X_n) is independent of the time sequence (T_n) in the models we will consider, and this is the canonical situation for marked point processes, extracting the sequence of record times (T_{L_n}) from (T_n, X_n) introduces a further marked point process in which the marks depend on the points in a rather non-standard way. To see this, let δ_n denote the indicator of the event $\{X_n$ is a record$\} = \{$a record occurs at $T_n\}$. Then the Dwass-Rényi Lemma says that $(\delta_n)_{n\in\mathbb{N}}$ is sequence of independent Bernoulli random variables with $P(\delta_n = 1) = 1/n$, and (δ_n) is independent of (T_n) since (X_n) is. The random record *times* can then be regarded either as those points of the marked point process (T_n, δ_n) for which $\delta_n = 1$, or simply as the process $(T_n\delta_n)$. In other words, the record arrival process is indeed a thinning of (T_n) — but the thinning probability depends on the state of (T_n), that is, on n. The standard theory of thinned marked point processes does allow for variable thinning probabilities but the probability typically depends on the location of the point or on some other process that is independent of (T_n). The random record process should fall within the purview of martingale- based theory of thinned marked point processes, but the matter has apparently not yet been explored in depth.

Therefore we turn now to results obtained by (more or less) classical methods in specific cases. We will take the following as our basic setup: $(T_n)_{n\in\mathbb{N}}$ is a simple, nonexplosive point process on $[0,\infty)$ (no multiple points and no finite accumulation points), and the observations or marks $(X_n)_{n\in\mathbb{N}}$ form an i.i.d. sequence which is independent of (T_n). Here we will take the d.f. F of the X_n to be continuous and in some cases even absolutely continuous; furthermore, our random record process (T_{L_n}, R_n) only looks at *first* records. Variations on this theme have occasionally been considered and we will mention these as appropriate. We see immediately that, since (X_n) is independent of (T_n), *marginally* the record sequence (R_n) behaves according to the classical model. Thus there are two main areas of interest here: the joint behaviour of (T_{L_n}, R_n) and the marginal behaviour of (T_{L_n}), whether in finite time or asymptotically as time goes to infinity. Record times have attracted the lion's share of attention as we will see. Note: here and throughout we define $N(t) := \#\{T_n : T_n \le t, n \ge 1\}$ and $L(t) := \#\{T_{L_n} : T_{L_n} \le t, n \ge 1\}$ for $t \ge 0$.

9. Poisson paced records.

We begin with the Poisson pacing process, since it is the most familiar conceptually and was the first to be studied in this context (Pickands, 1971). This model appears to have originally been motivated by an insurance problem in which claims of random size arrive at random times, and we are concerned about the arrival times and values of claims of record size. Let $(T_n)_{n\in\mathbb{N}}$ be the points of an inhomogeneous Poisson process with rate $\lambda(t)$ for $t \ge 0$, so that $E\big(N(t)\big) = \Lambda(t) := \int_{[0,t]} \lambda(u)\,du$ for $t \ge 0$. We suppose that the distribution F of the X_n is absolutely continuous with density f. Then it is straightforward to write the joint density of $(T_{L_n}, R_n)_{n\in\mathbb{N}}$:

$$f_{T_{L_1},\ldots,T_{L_n},R_1,\ldots,R_n}(t_1,\ldots,t_n,r_1,\ldots,r_n) = \\ = e^{-\Lambda(t_1)}\lambda(t_1)f(r_1)\prod_{i=2}^{n} e^{-(\Lambda(t_i)-\Lambda(t_{i-1}))(1-F(r_{i-1}))}\lambda(t_i)f(r_i), \tag{9.1}$$

for $0 < t_1 < \ldots < t_n$ and $r_1 < \ldots < r_n$. This simply says that at time t_i the record value r_i occurs, and then from time t_i to t_{i+1} no observation X exceeds r_i. One can readily integrate (9.1) with respect to t_1, …, t_n to verify that marginally the record values are distributed as in the classical model. It is less easy to derive the marginal distribution of the record times, but this can be achieved by other methods. The result, found via extremal process methods by Deheuvels (1982) and via characteristic functions by Bunge and Nagaraja (1992b), is the following.

Theorem 9.1. *Let ε_1, ε_2, … and E_1, E_2, … be i.i.d. unit exponential r.v.s, and put $S_0 := 0$ and $S_n := \sum_{j=1}^{n} \varepsilon_j$ for $n = 1,\ 2,\ \ldots$. In the Poisson-paced record process as defined above,*

$$(T_{L_1}, T_{L_2}, \ldots, T_{L_n}) =_d \left(\Lambda^{\leftarrow}(E_1 e^{S_0}), \Lambda^{\leftarrow}(E_1 e^{S_0} + E_2 e^{S_1}), \ldots, \Lambda^{\leftarrow}\Big(\sum_{i=1}^{n} E_i e^{S_{i-1}}\Big)\right) \tag{9.2}$$

for all $n \in \mathbb{N}$.

In particular if (T_n) is a *homogeneous* Poisson process with rate λ then the action of the function $\Lambda^{\leftarrow}$ may be replaced in (9.2) by multiplication by $1/\lambda$.

It is worth pausing here to compare this result with the representation of record times in the classical model given by Theorem 5.8. Taking $m = 1$ in Theorem 5.8 we have

$$L_n =_d 1 + \sum_{i=1}^{n-1} \left\lceil \frac{E_i}{-\ln(1 - e^{-S_i})} \right\rceil. \tag{9.3}$$

But $e^x \sim -1/\ln(1 - e^{-x})$ as $x \to \infty$, so since $\lim_{n\to\infty} S_n = \infty$ almost surely we are justified in regarding (9.2) with $\lambda(t) \equiv 1$ as a continuous-time version of (9.3) in the homogeneous case.

One reason for studying specifically inhomogeneous Poisson paced records is as follows. In real life we frequently get the impression that records are being broken 'too often' in financial data, weather, sports, and so forth. According to the theory of the classical model, this means that there must be some inhomogeneity or non-stationarity in the process. In the classical model non-stationarity can enter only via changing distributions of the X_n, but in the random record model it may be that observations are occurring at an increasing rate: in the Poisson case this corresponds to an increasing rate function $\lambda(t)$. Thus we are led to study the asymptotic ($t \to \infty$) behaviour of the inhomogeneous Poisson-paced record process. The basic result is as follows (due to Gaver and Jacobs (1978) for $\lambda(t) = e^{\lambda_0 t}$, and Bunge and Nagaraja (1992a) for the generalization below). Denote the inter-record times by $U_{L_n} = T_{L_n} - T_{L_{n-1}}$ for $n \geq 1$, with $T_{L_0} := 0$.

Theorem 9.2. *In the Poisson-paced record model, suppose that*

$$\lim_{t\to\infty} \frac{\lambda(t)}{\Lambda(t)} = \lambda_0 \in (0, \infty).$$

Then

$$(U_{L_{N+1}}, \ldots, U_{L_{N+n}}) \Rightarrow \frac{1}{\lambda_0}(E_1, \ldots, E_n) \qquad (N \to \infty),$$

and

$$(U_{L_1(T)}, \ldots, U_{L_n(T)}) \Rightarrow \frac{1}{\lambda_0}(E_1, \ldots, E_n) \qquad (T \to \infty),$$

for $n = 1, 2, \ldots,$ where $U_{L_{N+j}}$ is the j^{th} inter-record time after the N^{th} record occurrence and $U_{L_j(T)}$ is the j^{th} inter-record time after time T.

In other words, when $\lambda(t)/\Lambda(t) \to \lambda_0$ so that $\lambda(t)$ has essentially exponential growth, the record arrival process converges to a homogeneous Poisson process with rate λ_0, whether we consider the record arrivals after a large number of records N or after a long time T.

Thus one way to get regularly occurring records in the time-limit is to have the underlying observations 'speeding up exponentially'. But a homogeneous Poisson process of record arrivals can also be constructed directly, for all $t \in [0, \infty)$, via the F^α model. Define the interarrival times $U_n := T_n - T_{n-1}$, where $T_0 := 0$, and let $(U_n)_{n\in\mathbb{N}}$

be a delayed renewal process with $U_1 \sim$ exponential(λ) and $U_n \sim$ exponential(λ/p) for $n \geq 2$, for some fixed $p \in (0,1)$ and $\lambda > 0$. Now let $\alpha_1 := 1$ and $\alpha_n := p/(1-p)^{n-1}$ for $n \geq 2$. Then it is readily shown that $P(X_n \text{ is a record}) \equiv p$ for all $n \geq 2$, and from this it follows that the record arrival process is homogeneous Poisson with rate λ (Bunge and Nagaraja, 1992b). Thus regular record breaking may be caused by exponentially increasing arrival rates of the observations or by geometrically increasing sizes of the underlying populations, and there are other possible causes as we will see.

We conclude this section with two further comments on known results. First, a slightly modified version of (9.1) can be used to compute the limiting joint density of inter-record times and normalized record values when $\lambda(t)/\Lambda(t) \to \lambda_0$, for the first n records either after the N^{th} record time ($N \to \infty$) or after a long time T ($T \to \infty$) (Bunge and Nagaraja, 1992a). But while the inter-record times converge marginally to i.i.d. exponential in either case, it is interesting to note that the limiting distributions of the normalized record *values* differ depending on whether the normalization is based on N or on T. Second, we note that Hofmann and Nagaraja (1997) have recently studied F^α random record models in some detail, obtaining (among other results) generalizations of Theorems 9.1 and 9.2 for the F^α case.

10. Birth-process paced records.

We now consider records over observations occurring at the points of a pure birth process. This makes sense as a model for real phenomena if we imagine, for example, a population of micro-organisms reproducing in an environment with unlimited food, and we are interested in the successive maximum level of some measurement on each new organism. Let $(T_n)_{n\in\mathbb{N}}$ be the jump points of a pure birth process with $m+1$ 'progenitors' for some fixed non-negative integer m, so that $P\big(N(t+h) - N(t) = 1 \;\big|\; N(t) = n\big) = (1+m+n)h + o(h)$ as $h \downarrow 0$, for each $t > 0$. More specifically $(T_n)_{n\in\mathbb{N}}$ is a continuous-time Markov process with q-matrix given by

$$q_{n,n+1} = -q_{n,n} = 1 + m + n \qquad (n = 0, 1, \ldots),$$

and $q_{i,j} = 0$ otherwise; in this case U_1, U_2, $\ldots$ are independent r.v.s with $U_n \sim$ exponential($n+m$) for $n \geq 1$. We then have the following result (see Bruss and Rogers (1991) for $m = 0$, Bunge and Nagaraja (1992b) for $m = 0, 1$, and Browne and Bunge (1995) for arbitrary m).

Theorem 10.1. *Suppose that (T_n) is a birth process as described above. Then the record arrival process $(T_{L_n})_{n\in\mathbb{N}}$ is equal in distribution to the arrival time sequence $(\tau_n)_{n\in\mathbb{N}}$ of a* Markov renewal process $(\tau_n, \theta_n)_{n\in\mathbb{N}}$. *Here $\theta_n \in \{1, 2, \ldots, 1+m\}$ and*

$$P(\tau_1 \leq t) = 1 - e^{-(1+m)t},$$
$$P(\tau_{n+1} - \tau_n \leq t, \theta_{n+1} = j | \theta_n = i) = \pi_{ij}(1 - e^{-jt}) \qquad (n \geq 1),$$

for $t \geq 0 and 1 \leq i \leq j \leq 1+m$, where $\theta_1 = 1 + m$ with probability 1 and

$$(\pi_{ij}) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{1+m} & \frac{1}{1+m} & \frac{1}{1+m} & \cdots & \frac{1}{1+m} \end{pmatrix} \qquad (1 \leq i \leq j \leq 1+m).$$

Theorem 10.1 says that the inter-record times are conditionally exponentially distributed: they start out with mean $1/(1+m)$ and the mean changes at each record occurrence according to the Markov chain (θ_n), eventually settling down to mean 1 (the absorbing state of (θ_n)), that is, a homogeneous Poisson record arrival process. Here, then, is another model that produces regular record arrivals; in fact for $m = 0$ the record arrival process is homogeneous Poisson for all $t \in [0,\infty)$, and even if $m > 0$ it attains this behaviour after some random time that may be large but is finite with probability 1. Bruss and Rogers (1991) studied the case $m = 0$ (possibly with a deterministic time-change) for k-records, finding that the k-record arrival processes are i.i.d. Poisson (after some delay if $k > 1$).

We can compare Theorem 10.1 to Theorem 9.2 a little more formally, as follows. Note that in the birth process discussed above we have $E\big(N(t)\big) = (1+m)(e^t - 1)$ so that

$$\frac{\frac{d}{dt}E\big(N(t)\big)}{E\big(N(t)\big)} \to 1 \qquad (t \to \infty), \tag{10.1}$$

which is exactly the condition of Theorem 9.2 if we take the (unimportant) constant λ_0 to be 1. It seems reasonable to envision a 'meta-theorem' stating that in general (10.1) implies a homogeneous Poisson weak limit for the record arrival process. But the proofs of Theorems 10.1 and 9.2 seem to be quite case-specific, so a unified formulation and proof remain to be found. Below we will see another version of this result, for renewal-paced records, derived by a still different approach.

11. Renewal paced records.

Finally we turn to the richest—and most difficult—random record model: the case where $(T_n)_{n\in\mathbb{N}}$ is a renewal process. By this we mean an ordinary or 'undelayed' renewal process, so that $(U_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of positive r.v.s; the assumption that (T_n) is simple implies that $P(U_n = 0) = 0$. Only one exact result is known in this case (Browne and Bunge, 1995).

Theorem 11.1. *Suppose that (T_n) is a renewal process with*

$$(U_n)_{n\in\mathbb{N}} =_d \left(\frac{1}{\lambda} E_n^{1/\alpha} Z_n\right)_{n\in\mathbb{N}},$$

where $(E_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of unit exponential r.v.s and $(Z_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of r.v.s, independent of (E_n), with Laplace- Stieltjes transform (LST) e^{-s^α} for some $\alpha \in (0,1]$, so that Z_n is positive α-stable. Then

$$(U_{L_1},\ldots,U_{L_n}) =_d \frac{1}{\lambda}\big(U_1 e^{S_0/\alpha},\ldots,U_n e^{S_{n-1}/\alpha}\big) \tag{11.1}$$

for all $n \in \mathbb{N}$, where $S_n := \sum_{i=1}^n E_i$ and $S_0 := 0$.

Theorem 11.1 directly generalizes Theorem 9.1, because when $\alpha = 1$ (T_n) is a renewal process with exponential(λ) interarrival distribution, i.e., it is a homogeneous Poisson process with rate λ. The presence of the homogeneous Poisson process (S_n) in the exponent on the right-hand side of (11.1) is noteworthy, and we may ask whether any other random record processes admit such a 'logarithmic Poisson' representation. We return to this question at the end of this section.

Lacking an exact distributional representation, what can we say about a general renewal-paced record process? One way to answer this question is to look at the tails of the inter- record time distributions: from this perspective the theory of regular variation holds the promise of fairly general results. This approach originated with Westcott (1977, 1979) and was followed up by Embrechts and Omey (1983) and Yakymiv (1986). Here we give the main result of Embrechts and Omey (1983), who studied regular variation of the tails of general *subordinated distributions*, that is, distributions of random sums of random variables, about which more below. Let RV_α^∞ denote the set of regularly varying functions (at ∞) with index α, i.e., $f \in RV_\alpha^\infty$ if $f(x) \sim x^\alpha H(x)$ for some slowly varying H (see Bingham *et al.*, 1989, for a full account).

Theorem 11.2 *Let $(T_n)_{n\in\mathbb{N}}$ be a renewal process with interarrival times $(U_n)_{n\in\mathbb{N}}$ and suppose that U_n has d.f. G. Then for any fixed $\alpha \in [0,1)$, the following are equivalent:*

$$\begin{aligned} 1 - G(x) &\in RV_{-\alpha}^\infty, \\ P(U_{L_n} > x) &\in RV_{-\alpha}^\infty, \\ P(T_{L_n} > x) &\in RV_{-\alpha}^\infty. \end{aligned}$$

Moreover, as $x \to \infty$,

$$P(U_{L_n} > x) \sim P(T_{L_n} > x) \sim \frac{1 - G(x)}{(n-1)!}\left(\ln\left(\frac{1}{1 - G(x)}\right)\right)^{n-1}.$$

One remarkable aspect of this result is its assertion of the *same* tail behaviour of record and inter-record times, which reaffirms the fact that the last inter-record time dominates them all.

A completely different approach to the approximation of renewal-paced record times was taken by Ennadifi (1995a, 1995b). She considered *strong* approximation of the record arrival process via Wiener processes defined along with the random record process on a suitably enlarged probability space. We present two examples of Ennadifi's results here, while noting that their implications have not yet been fully explored. Assume that U_n has its moment generating function existing in some nondegenerate neighborhood of zero, so that in particular it has finite mean and variance.

Theorem 11.3 *It is possible to define $L(t)$ and a Wiener process $W(t)$ on the same probability space such that for each $\varepsilon > 0$,*

$$\sup_{\varepsilon \le t \le T} |L(t) - \ln t - W(\ln t)| = O(\ln \ln T),$$

almost surely as $T \to \infty$.

Based on similar considerations she obtained the following.

Theorem 11.4 *The sequence of record times of the process $\left(L(e^t)\right)_{t\geq 0}$ is almost surely asymptotically close to the arrival times of a homogeneous Poisson process of unit rate.*

Theorem 11.4 serves as yet another version of Theorem 9.2 and the asymptotic part of Theorem 10.1.

Finally we consider questions related to random sums of random variables, or *random sums* for short. Here we will consider only the second inter-record time in a renewal paced record process. Observe that this is

$$U_{L_2} = T_{L_2} - T_{L_1} =_d \sum_{i=1}^{N} U_i,$$

where N is a random variable independent of $(U_n)_{n\in\mathbb{N}}$. The distribution of N is easy to calculate because if $X_1 = R_1$ were fixed at, say, r, then N would be the first time such that $X_n > r$, so that $P(N = n) = F^{n-1}(r)(1 - F(r))$ for $n = 1, 2, \ldots$. But we can integrate this geometric probability with respect to r to find that unconditionally

$$P(N = n) = \frac{1}{n(n+1)} \qquad (n = 1, 2, \ldots), \tag{11.2}$$

retrieving the distribution of L_2 in Corollary 5.3. There is a vast literature on random sums of various kinds and in various applications (cf., e.g., Gnedenko and Korolev, 1996). Just as with the martingale-based theory of marked point processes, though, the application of the existing theory to random record processes, and in particular sums of the type (11.2), has not yet been clarified. Note in particular that $E(N) = \infty$ in (11.2), which puts the random record problem outside the purview of a good part of random sum theory. Nonetheless we can envision applying this theory (or a suitable extension of it) to random record processes along the following lines. The equation

$$\sum_{i=1}^{N} U_i =_d \frac{1}{p} U_1 \tag{11.3}$$

has been well-studied in the case where $N \sim$ geometric with success probability $p \in (0, 1)$; (11.3) is known as *geometric stability* (Kalashnikov, 1997). If we now take p to be a uniform$(0, 1)$ r.v., say ξ, we obtain

$$U_{L_2} =_d \sum_{i=1}^{N} U_i =_d \frac{1}{\xi} U_1. \tag{11.4}$$

Recalling that $e^E =_d 1/\xi$ when $E \sim$ exponential(1), we see that (11.4) applies exactly to the homogeneous Poisson-paced record process with $\lambda = 1$ (compare Theorem 9.1), and in fact (11.4) can be generalized to the renewal process of Theorem 11.1. As mentioned above, then, the question is: to what extent does the 'log-Poisson' representation of Theorem 11.1 characterize the specific renewal process used there? That is, to what extent does (11.4) characterize U_1? The following partial result is based on analysis of functional equations derived from the 'random stability' equation (11.4) (Browne and Bunge, 1995).

Theorem 11.5 *Suppose that $(T_n)_{n\in\mathbb{N}}$ is a renewal process with interarrival times $(U_n)_{n\in\mathbb{N}}$, and (11.1) holds with S_n as defined in Theorem 11.1 and $\alpha > 0$. Then*

$$U_n =_d \frac{1}{\lambda} E_n^{1/\alpha} Z_n \qquad (n \in \mathbb{N}),$$

where E_n and Z_n are as defined in Theorem 11.1, and $\alpha \in (0, 1]$.

Efforts to obtain more explicit characterizations of renewal processes from their random record process representations have so far foundered on analytical difficulties, especially the fact that here $E(N) = \infty$.

12. Extensions and applications.

Not many variations on the basic random record model have yet been studied in depth. Westcott (1977) looked at some specific models incorporating dependence or non-stationarity but found that they gave rise to intractable calculations. As noted above, Hofmann and Nagaraja (1997) were able to extend some known results for random record models to the F^α setting. In terms of applications, there has been considerable research on the use of random record models in best-choice or 'secretary' problems, notably due to Bruss (see Bruss, 1998, and references therein). In this application one supposes that 'options' or 'choices' appear at random times, and the objective is to choose the 'best' according to some optimality criterion; the random record models that admit closed-form representations often also admit elegant solutions to such problems.

References

Ahsanullah, M. (1995). Record values. In: N. Balakrishnan and A. P. Basu, eds., *The Exponential Distribution: Theory, Methods and Applications.* Gordon & Breach, Amsterdam, 279–296.

Arnold, B. C., N. Balakrishnan and H. N. Nagaraja (1992). *A First Course in Order Statistics.* Wiley, New York.

Arratia, R., A. D. Barbour and S. Tavaré (1992). Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.* **2**, 519–535.

Arratia, R. and S. Tavaré (1992). The cycle structure of random permutations. *Ann. Probab.* **20**, 1567–1591.

Ballerini, R. (1994). A dependent F^α-scheme. *Statist. Probab. Lett.* **21**, 21–25.

Bingham, N. H. (1988). Tauberian theorems for Jakimovski and Karamata-Stirling methods. *Mathematika* **35**, 216–224.

Bingham, N. H., C. M. Goldie and J. L. Teugels (1989). *Regular Variation*, corrected pbk. ed. Cambridge Univ. Press, Cambridge, UK.

Browne, S. and J. Bunge (1995). Random record processes and state dependent thinning. *Stochastic Process. Appl.* **55**, 131–142.

Bruss, F. T. (1998). Quick solutions for general best choice problems in continuous time. *Comm. Statist.—Stoch. Models* **14**, 241–264.

Bruss, F. T. and L. C. G. Rogers (1991). Pascal processes and their characterization. Stochastic Process. Appl. **37**, 331–338.

Bunge, J. and H. N. Nagaraja (1992a). Dependence structure of Poisson-paced records. *J. Appl. Probab.* **29**, 587–596.

Bunge, J. and H. N. Nagaraja (1992b). Exact distribution theory for some point process record models. *Adv. in Appl. Probab.* **24**, 24–44.

Csörgő, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Daley, D. J. and D. Vere-Jones (1988). *An Introduction to the Theory of Point Processes.* Springer, New York.

David, F. N. and D. Barton (1962). *Combinatorial Chance.* Hafner, New York.

Deheuvels, P. (1982). Spacings, record times and extremal processes. In: G. Koch and F. Spizzichino, eds., *Exchangeability in Probability and Statistics.* North-Holland, Amsterdam, 233–243.

Deheuvels, P. (1983). The strong approximation of extremal processes, II. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **62**, 7–15.

Deheuvels, P. (1984). On record times associated with kth extremes. In: Proceedings of the 3rd Pannonian Symposium on Mathematical Statistics (Visegrad, Hungary, 1982). Akadémiai Kiadó, Budapest, 43–51.

Deheuvels, P. (1988). Strong approximations of k-th records and k-th record times by Wiener processes. *Probab. Theory Related Fields* **77**, 195–209.

Deuschel, J.-D. and O. Zeitouni (1995). Limiting curves for i.i.d. records. *Ann. Probab.* **23**, 852–878.

Dwass, M. (1960). Some k-sample rank order tests. In: I. Olkin et al., eds., *Contributions to Probability and Statistics: Essays in Honor of H. Hotelling.* Stanford Studies in Math. & Stat. **2**, Stanford U. Press, Stanford, 198–202.

Embrechts, P., C. Klüppelberg and T. Mikosch (1997). *Modelling Extremal Events: for Insurance and Finance.* Springer, Berlin.

Embrechts, P. and E. Omey (1983). On subordinated distributions and random record processes. *Math. Proc. Camb. Phil. Soc.* **93**, 339-353.

Engelen, R., P. Tommassen and W. Vervaat (1988). Ignatov's theorem: a new and short proof. *J. Appl. Probab.* **25A**, 229–236.

Ennadifi, G. (1995a). Strong approximation of the number of renewal paced record times. *J. Statist. Plann. Inference* **45**, 113-132.

Ennadifi, G. (1995b). Asymptotic behaviour of record times for renewal paced records. *Statist. Decisions* **13**, 379–397.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, vol. I, 3rd ed. Wiley, New York.

Feuerverger, A. and P. Hall. (1996). On distribution-free inference for record-value data with trend. *Ann. Statist.* **24**, 2655–2678.

Galambos, J. (1987). *Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Krieger, Malabar, Florida.

Gaver, D. and P. A. Jacobs (1978). Non-homogeneously paced random records and associated extremal processes. *J. Appl. Probab.* **15**, 552–559.

Gill, R. D. and S. Johansen (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18**, 1501–1555.

Glick, N. (1978). Breaking records and breaking boards. *Amer. Math. Monthly* **85**, 2–26.

Gnedenko, B. V. and V. Yu. Korolev (1996). *Random Summation.* CRC Press, Boca Raton, Florida.

Gnedin, A. V. (1994). Multiple maxima of a normal sample. Report, Universität Göttingen, Institut für Mathematische Stochastik

Goldie, C. M. (1983). *On Records and Related Topics in Probability Theory.* Report, School of Mathematical Sciences, University of Sussex.

Goldie, C. M. and S. I. Resnick (1989). Records in a partially ordered set. *Ann. Probab.* **17**, 678–699.

Goldie, C. M. and S. I. Resnick (1995). Many multivariate records. *Stochastic Process. Appl.* **59**, 185–216.

Goldie, C. M. and L. C. G. Rogers (1984). The k-record processes are i.i.d. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **67**, 197–211.

Haiman, G. and M. L. Puri (1990). A strong invariance principle concerning the J-upper order statistics for stationary m-dependent sequences. *J. Statist. Plann. Inference* **25**, 43–51.

Hofmann, G. and H. N. Nagaraja (1997). Random and point process record models in the F^α setup. Preprint, Dept. Statistics, Ohio State University.

Ignatov, Z. (1978). Point processes generated by order statistics and their applications. In: P. Bártfai and J. Tomkó, eds., *Point Processes and Queueing Problems, Keszthely (Hungary).* Coll. Mat. Soc. János Bolyai **24**, North-Holland, Amsterdam, 109–116.

Kalashnikov, V. (1997) *Geometric Sums: Bounds for Rare Events with Applications.* Kluwer, Dordrecht.

Last, G. and A. Brandt. (1995). *Marked Point Processes on the Real Line.* Springer-Verlag, New York.

Nagaraja, H. N. (1988). Record values and related statistics—a review. *Comm. Statist.—Theory Meth.* **17**, 2223–2238.

Nevzorov, V. B. (1995). Asymptotic distributions of records in nonstationary schemes. *J. Statist. Plann. Inference* **45**, 261–273.

Nevzorov, V. B. (1987). Records. *Theory Prob. Appl.* **32**, 201–228. (Transl. of Russian original in *Teor. Veroyatnost. i Primenen.* **32**(1987), 219–251.)

Pfeifer, D. (1982). Characterizations of exponential distributions by independent nonstationary record increments. *J. Appl. Probab.* **19**, 127–135.

Pfeifer, D. (1987). On a joint strong approximation theorem for record and inter-record times. *Probab. Theory Related Fields* **75**, 212–221.

Pfeifer, D. (1989). *Einführung in die Extremwertstatistik.* Teubner, Stuttgart.

Pfeifer, D. and Y.-S. Zhang (1989). A survey on strong approximation techniques in connection with records. In: J. Hüsler and R.-D. Reiss, eds., *Extreme Value Theory: Proceedings, Oberwolfach, 1987.* Lecture Notes in Statistics **51**, Springer, New York, 50–58.

Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *J. Appl. Probab.* **8**, 745–756.

Rényi, A. On the extreme elements of observations. MTA III. Oszt. Közl. **12**, 105–121. Reprinted in *Selected Papers*, vol. 3, ed. P. Turán. Akadémiai Kiadó, Budapest (1976), 50–65.

Resnick, S. I. (1973). Limit laws for record values. *Stochastic Process. Appl.* **1**, 67–82.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, New York.

Resnick, S. I. (1992). *Adventures in Stochastic Processes.* Birkhäuser, Boston.

Robinson, M. E. and J. A. Tawn (1995). Statistics for exceptional athletics records. *J. Roy. Statist. Soc. C, Appl. Statist.* **44**, 499–511.

Rogers, L. C. G. (1989). Ignatov's theorem: an abbreviation of the proof of Engelen, Tommassen and Vervaat. *Adv. in Appl. Probab.* **21**, 933–934.

Rogers, L. C. G. and D. Williams (1994). *Diffusions, Markov Processes and Martingales, I: Foundations*, 2nd ed. Wiley, Chichester.

Samuels, S. M. (1992). An all-at-once proof of Ignatov's theorem. In: F. T. Bruss, T. S. Ferguson and S. M. Samuels, eds., *Strategies for Sequential Search and Selection in Real Time.* Amer. Math. Soc., Providence, 231–237.

Shorrock, R. W. (1972). On record values and record times. *J. Appl. Probab.* **9**, 316–326.

Shorrock, R. W. (1974). On discrete-time extremal processes. *Adv. in Appl. Probab.* **6**, 580–592.

Smith, R. L. (1997). Letter to the Editors, on 'Statistics for exceptional athletics records' by Robinson and Tawn, with authors' response. *J. Roy. Statist. Soc. C, Appl. Statist.* **46**, 123–128.

Stam, A. J. (1985). Independent Poisson processes generated by record values and inter-record times. *Stochastic Process. Appl.* **19**, 315–325.

Vervaat, W. (1977). On records, maxima and a stochastic difference equation. Report 7702, Mathematisch Instituut, Katholieke Universiteit Nijmegen.

Westcott, M. (1977). The random record model. *Proc. Roy. Soc. Lond. A* **356**, 529–547.

Westcott, M. (1979). On the tail behaviour of record-time distributions in a random record process. *Ann. Probab.* **7**, 868–873.

Williams, D. (1973). On Rényi's 'record' problem and Engel's series. *Bull. London Math. Soc.* **5**, 235–237.

Yakymiv, A. L. (1986). Asymptotic properties of the times the states change in a random record process. *Theory Probab. Appl.* **31**, 508–512.

Yang, M. (1975). On the distribution of the inter-record times in an increasing population. *J. Appl. Probab.* **12**, 148–154.

Yao, Y.-C. (1997). On independence of k-record processes: Ignatov's theorem revisited. *Ann. Appl. Probab.* **7**, 815–821.