Statistica Sinica **7**(1997), 893-906

PREDICTION OF RECORD-BREAKINGS

Masaaki Sibuya and Kazuo Nishimura

Keio University and Komazawa University

Abstract: For analyzing record statistics from a sequence of non-i.i.d. random variables, a model with one or two parameters, controlling the occurrences of record-breakings, is proposed. Under the model, the number of record-breakings within n steps has a probability function including the Stirling-Carlitz polynomial of the first kind. Its expected number is still $O(\log n)$ and saturates in the long run. Waiting time to the sth occurrence also has a probability function of similar form. Under this model, methods for predicting future record-breakings are proposed, and applied to some practical data sets of weather and sports.

Key words and phrases: Nevzorov model, random walk, record value theory, Stirling-Carlitz polynomial, Stirling family of discrete distributions.

1. Introduction and Preliminaries

1.1. Upper new records

Let $(Z_k)_{k=1}^{\infty}$ be a sequence of random variables (r.v.), in which Z_k is an upper new record (u.n.r.), or a record breaking, if $Z_j < Z_k$, $1 \le j \le k-1$. Following the usual convention Z_1 is a u.n.r. Let X_n be the number of u.n.r.'s in $(Z_k)_{k=1}^{n}$, and let W_s be the time, or sequential index, when the sth u.n.r. occurs. The theory of these 'record statistics' for a sequence of independent and identically distributed (i.i.d.) or exchangeable r.v.'s is well established, and was surveyed by Glick (1978) and Galambos (1978). In the usual applications like environment, sports and economy, the i.i.d. or exchageable assumption is too restrictive, and some generalizations have been proposed. See, for example, Nevzorov (1987) and Nagaraja (1988) for extensive surveys up to the date. In this paper, a simple approach, which is distribution-free but parametric, is proposed for the prediction of future u.n.r.'s.

1.2. Nevzorov's model

Nevzorov (1984, 1987) considered the following model. Let $(Z_k)_{k=1}^{\infty}$ be a sequence of independent but nonidentically distributed r.v.'s, and let Z_k itself be the maximum of a random sample of varying size α_k from a common continuous distribution function (d.f.) $H(\cdot)$, that is, Z_k has the d.f. $H^{\alpha_k}(\cdot)$. Let $Y_k, k = 2, 3, \ldots$ be binary r.v.'s such that if Z_k is a u.n.r. $Y_k = 1$, else $Y_k = 0$. Then $(Y_k)_{k=2}^{\infty}$ are independent, and

$$\Pr\{Y_k = 1\} = \alpha_k/(\alpha_1 + \cdots + \alpha_k) =: \lambda_k. \tag{1.1}$$

The sequence $(\alpha_k)_{k=1}^{\infty}$ can consist of any positive numbers, and determined conversely from $(\lambda_k)_{k=2}^{\infty}$ by $\alpha_k = (\alpha_1 + \cdots + \alpha_{k-1})\lambda_k/(1-\lambda_k)$, $k = 2, 3, \ldots$ starting from any α_1.

Nevzorov's model is useful for studying limit theorems, but not for statistical inference based on a single time series data set or its u.n.r.'s. In the analysis of annual records, for example, we cannot expect a long data set, and we restrict the sequence (α_k) or (λ_k) to a specific one, which has one or two parameters, includes the i.i.d. case and leads to distributions of nice properties. To prepare for the discussion, the simplest case of Nevzorov's model is reviewed now.

1.3. The i.i.d. case

If α_k is a constant, $(Z_k)_{k=1}^{\infty}$ is an i.i.d. sequence, $\lambda_k = 1/k$, and Y_k has the probability generating function (p.g.f.)

$$pgf(z; Y_k) = (z + k - 1)/k, \quad k = 1, 2, \ldots \tag{1.2}$$

From (1.2) the p.g.f. of $X_n = Y_1 + \cdots + Y_n$ is

$$pgf(z; X_n) = z^{[n]}/n!, \tag{1.3}$$

(in terms $z^{[n]}$ in (1.6)) which shows the probability function (p.f.)

$$\Pr\{X_n = x\} = \begin{bmatrix} n \\ x \end{bmatrix} \frac{1}{n!}, \qquad 1 \le x \le n, \tag{1.4}$$

and further, the p.f. of the waiting time W_s defined in Subsection 1.1 is

$$\Pr\{W_s = w\} = \begin{bmatrix} w-1 \\ s-1 \end{bmatrix} \frac{1}{w!}, \qquad s \le w. \tag{1.5}$$

In the above expressions the brackets denote 'unsigned' Stirling numbers of the first kind, which are defined by the polynomial identity

$$z^{[n]} = z(z+1)\cdots(z+n-1) = \sum_{m=1}^{n} \begin{bmatrix} n \\ m \end{bmatrix} z^m. \tag{1.6}$$

Hence, the Stirling numbers are positive integers if $n = 1, 2, \ldots$ and $1 \le m \le n$, and zero otherwise for nonnegative integer n and integer m except for the case $n = m = 0$. (About Stirling numbers see, for example, Graham et al. (1989)).

In the next Section 2, a polynomial generalizing Stirling number of the first kind, Carlitz (1980a, b), is introduced. Based on this polynomial, a two-parameter family of distributions $\textsc{Str1F}(n, \theta, \tau)$ for the number X_n of u.n.r.'s, and another $\textsc{Str1W}(s, \theta, \tau)$ for the waiting time W_s until the sth u.n.r. are defined. They reduce, if $\tau = 0$ and $\theta = 1$, to (1.4) and (1.5), respectively, and

include the distributions of the number of future u.n.r.'s and conditional waiting time.

In Section 3, the predictions of the number of u.n.r.'s are discussed. The point and interval estimations in the case τ is known, and the point estimation in the case τ is unknown, are treated. In Section 4, the predictions of the waiting time are discussed.

In the final Section 5, applications to weather, Vancouver monthly precipitation and Tokyo August temperature, and to sports, Nippon Derby and Olympic Games, are shown.

2. Two-Parameter Family of Distributions for Record Breakings

2.1. Stirling-Carlitz polynomial

The Stirling number of the first kind is extended in many directions. The following is one of the most natural extensions. Define $R_1(n, m; t)$ by the polynomial identity

$$(z + t)^{[n]} = \sum_{m=0}^{n} R_1(n, m; t) z^m. \tag{2.1}$$

The coefficient $R_1(n, m; t)$ is a polynomial of t of order $n - m$ with nonnegative integer coefficients. Expanding the left side of (2.1) in the form of (1.6) and comparing it with the right side of (2.1), we find

$$R_1(n, m; t) = \sum_k \binom{n}{k} \begin{bmatrix} k \\ m \end{bmatrix} t^{[n-k]}.$$

Note that

$$R_1(n, m; 0) = \begin{bmatrix} n \\ m \end{bmatrix}, \qquad R_1(n, m; 1) = \begin{bmatrix} n+1 \\ m+1 \end{bmatrix},$$

and

$$R_1(n + 1, m; t) = R_1(n, m + 1; t) + (n + t) R_1(n, m; t). \tag{2.2}$$

The polynomial $R_1(n, m; t)$, as well as its $R_2(n, m; t)$ extending the Stirling number of the second kind, was introduced by Carlitz (1980a, b) and discussed further by Koutras (1982), Broder (1984), Shanguman (1984) and Neuman (1987).

2.2. The distribution of the number of u.n.r.'s

The definition (2.1) of R_1 introduces the family STR1F(n, θ, τ) of X_n with p.f.,

$$p_F(x; n, \theta, \tau) = \Pr\{X_n\} = R_1(n, x, \tau) \theta^x / (\theta + \tau)^{[n]}, \quad x = 0, 1, \ldots, n, \tag{2.3}$$

$$n = 1, 2, \ldots, \quad 0 < \theta < \infty, \quad 0 \le \tau < \infty,$$

with the p.g.f.

$$pgf(z; X_n) = (\theta z + \tau)^{[n]}/(\theta + \tau)^{[n]} \ . \tag{2.4}$$

The distribution STR1F(n, θ, τ) is actually outside the Nevzorov model, since (2.4) corresponds to $\lambda_k = \theta/(\theta + \tau + k - 1), k = 1, 2, \ldots,$ and $\lambda_1 < 1$ for $\tau > 0$. The necessity of leaving the model will be clear in the next section.

From (2.3) or (2.4),

$$\Pr\{X_n = 0\} = \tau^{[n]}/(\theta + \tau)^{[n]}, \tag{2.5}$$

$$\Pr\{X_n = 1\} = \begin{cases} \left(\theta\tau^{[n]}/(\theta + \tau)^{[n]}\right) \displaystyle\sum_{k=1}^{n} (\tau + k - 1)^{-1}, & \text{for } \tau > 0, \\ (n-1)!/(\theta + 1)^{[n-1]}, & \text{for } \tau = 0, \end{cases} \tag{2.6}$$

$$E(X_n) = \theta \sum_{k=1}^{n} (\theta + \tau + k - 1)^{-1} = \theta(\log n - \log(\theta + \tau)) + O(1/n), \tag{2.7}$$

and

$$\mathrm{Var}(X_n) = \theta \sum_{k=1}^{n} (\theta + \tau + k - 1)^{-1} - \theta^2 \sum_{k=1}^{n} (\theta + \tau + k - 1)^{-2}. \tag{2.8}$$

To see the role of the parameter τ of the distribution, the graphs of $(\mathrm{Var}(X_n), E(X_n))$ are plotted in Figure 1 for $n = 10$, $\tau = 0$, 0.001, 0.01, 0.1, 1, 10 and ∞. The mean is an increasing function of θ. As $\theta, \tau \to \infty$ such that $\theta/(\theta + \tau) \to \rho$, X_n approaches the binomial distribution with parameter (n, ρ).

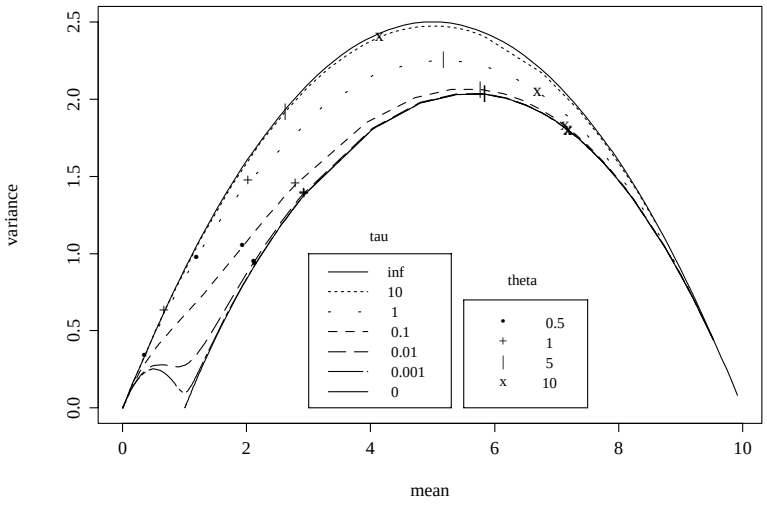

Figure 1. Relation between mean and variance for $0 \le \tau \le \infty$ and $\theta = 0.5$, 1, 5, 10.

2.3. The distribution of the waiting time

Consider the random walk of a particle on $\mathcal{N}_0 = \{0, 1, 2, \ldots\}$, starting from 0 at time 0, moving one step to the right with the time-dependent probability $\theta/(\theta+\tau+k-1)$ at time $k\,(=1,2,\ldots)$, and staying motionless with the probability $(\tau+k-1)/(\theta+\tau+k-1)$. X_n is the position of the particle at time n. If W_s now denotes the time when the particle arrives first at s, the family STR1W(s,θ,τ) of the waiting time is

$$p_W(w;s,\theta,\tau) := \Pr\{W_s = w\} = R_1(w-1, s-1;\tau)\theta^s/(\theta+\tau)^{[w]}, \tag{2.9}$$

$$w = s, s+1, \ldots; \quad s = 1, 2, \ldots, \quad 0 < \theta < \infty, \quad 0 \le \tau < \infty.$$

It is shown that p_W is a proper distribution, namely $\sum_{w=s}^{\infty} p_W(w;s,\theta,\tau) = 1$, and from this fact

$$E((\theta+\tau+W_s-1)^{(r)}) = (\theta/(\theta-r))^s\,(\theta+\tau-1)^{(r)}, \quad \theta > r, \tag{2.10}$$

where $z^{(r)} = z(z-1)\cdots(z-r+1)$, and

$$\mu_W(s,\theta,\tau) := E(W_s) \tag{2.11}$$
$$= \begin{cases} (\theta^s(\theta-1)^{-s} - 1)\,(\theta+\tau-1), & \text{if } \theta > 1,\ s = 1, 2, \ldots; \\ \infty\,, & \text{if } 0 < \theta \le 1,\ s = 2, 3, \ldots, \end{cases}$$

for $\tau \ge 0$. Trivially $E(W_1)$=1 for any $\theta > 0$ if $\tau = 0$.

The above results are obtained by considering the random walk. They can also be obtained from the general results on the Makov chain (see Fu (1996)).

The distributions STR1F, STR1W and related ones belong to the 'extended Stirling family of probability distributions', which have p.f. including explicitly the Stirling-Carlitz polynomial of the first or the second kind as a factor. They extend the Stirling family of probability distributions (Sibuya (1986, 1988)). The properties of the new extended family, as well as the above mentioned Markov chain approach, was reported elsewhere (Nishimura and Sibuya (1997)).

3. Prediction of the Number of u.n.r.'s

3.1. Future events

Let X_n be the random variable with STR1F(n,θ,τ). The problem is to predict

$$X_{n+m} - X_n = \sum_{k=1}^{m} Y_{n+k} =: X_{n,m}, \tag{3.1}$$

which has, because of (2.4), the distribution STR1F$(m,\theta,\tau+n)$ and is independent of X_n. That is, $X_{n,m}$ depends on the fact that the highest record of the

past n observations is known, and does not depend on the number X_n of u.n.r.'s in the past.

Even in the i.i.d. case, namely $\tau = 0$ and $\theta = 1$, $X_{n,m}$ has STR1F$(m, 1, n)$ and we need STR1F(n, θ, τ), $\tau > 0$. Some known results are obtained more systematically in terms of the new family. For example, using (2.5),

$$\Pr\{X_{n,m} > 0;\ \theta = 1, \tau = 0\} = 1 - n^{[m]}/(n+1)^{[m]} = m/(n+m). \tag{3.2}$$

As noted in Subsection 2.2, the assumption $\tau > 0$ contradicts the convention $X_1 = 1$ (since we regard Z_1 as a u.n.r.). In developing the theory of inference on STR1F(n, θ, τ), however, there is no problem assuming $\tau > 0$. In the analysis of actual data, it is likely the model $\tau > 0$ to be useful provided that $\Pr\{X_1 = 0\} = \tau/(\theta + \tau)$ is not large. This point will be discussed in Subsection 3.4. If τ is known (whether it is 0 or not), X_n is a sufficient statistics for the observations $(Y_1, \ldots, Y_n)$, and we discuss this case first, in Subsections 3.2 and 3.3.

If τ is known, for testing $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$, the rule to reject H_0 if $X_n \ge c$, where c is a constant determined by the significance level, is the uniformly most powerful. A special case, $\tau = 0$ and $H_0 : \theta = 1$, is the hypothesis of randomness of $(Z_k)_{k=1}^n$. The power of the randomness test against $H_1 : \theta > 1$ was compared with Kendall's rank correlation test by Iiyama et al. (1995).

3.2. Known τ, point prediction

Write (2.7) as

$$\mu_F(n, \theta, \tau) := E(X_n) = \theta \sum_{k=1}^{n} (\theta + \tau + k - 1)^{-1}, \tag{3.3}$$

and X_n is the uniformly minimum variance unbiased estimator of $\mu_F(n, \theta, \tau)$. To predict $X_{n,m}$ or to estimate $\mu_F(m, \theta, \tau + n)$, a simple method is to use

$$\hat{X}_{n,m} = \mu_F(m, \hat{\theta}, \tau + n), \tag{3.4}$$

where $\hat{\theta}$ is the solution of $\mu_F(n, \hat{\theta}, \tau) = X_n$ and is the maximum likelihood (m.l.) estimator. To solve this nonlinear equation of $\hat{\theta}$, we start from $\hat{\theta}_0 \ll 1$ and repeat the Newton-Raphson process for $\hat{\theta}_j$ while it is increasing, or start from an over-estimate $\hat{\theta}_0 = (\tau + (n-1)/2)/(n/X_n - 1)$.

If τ is known, the Fisher information of STR1F(n, θ, τ) is

$$I_F(\theta; n, \tau) = \mu_F(n, \theta, \tau)/\theta^2 - \sum_{k=1}^{n} (\theta + \tau + k - 1)^{-2} = \mathrm{Var}(X_n)/\theta^2. \tag{3.5}$$

The last equality holds because of (2.8). Since $(\partial/\partial\theta)\mu_F(n,\theta,\tau) = \mathrm{Var}(X_{n,m})/\theta$, the asymptotic variance of the estimate $\hat{X}_{n,m}$ is $\mathrm{Var}(X_{n,m})/\mathrm{Var}(X_n)$.

3.3. Known τ, interval prediction

The relationship (3.1) leads to the conditional distribution of $(X_n, X_{n,m})$ given $X_n + X_{n,m} = u$,

$$\begin{aligned} p_C(x;u,n,m) &:= \Pr\{(X_n, X_{n,m}) = (x, u-x) \mid X_n + X_{n,m} = u\} \\ &= R_1(n,x;\tau)R_1(m,u-x;\tau+m)/R_1(n+m,u;\tau), \end{aligned} \tag{3.6}$$

which is independent of θ. Takeuchi (1975) showed that there exists, for such a case, a randomized unbiased prediction interval with the given confidence coefficient of the uniformly shortest expected length. Since the randomized selection of the end points is necessary for the exact confidence coefficient, this optimal interval is not practical.

Because of the log-concavity of p_F's of X_n and $X_{n,m}$, the p.f. (3.6) is also shown to be log-concave, and hence it is unimodal. For a given u, let $[a^*(u), b^*(u)]$ be the 'modal interval' of p_C, depending on n, m, and confidence coefficient α. It is the shortest interval such that

$$\begin{gathered} p_C(x;u,n,m) \geq p_C(y;u,n,m) \quad \text{for any } x \in [a^*(u), b^*(u)],\ y \notin [a^*(u), b^*(u)], \\ \text{and} \quad \Pr\{X_n \in [a^*(u), b^*(u)] \mid X_n + X_{n,m} = u\} \geq \alpha. \end{gathered}$$

Define the confidence belt

$$B_0(\alpha) = \{(x,u) : x \in [a^*(u), b^*(u)],\ 0 \leq u \leq m+n\};$$

then the interval estimator of $X_{n,m}$ given $X_n = x$ is

$$B(x,\alpha) = \{y = u - x : (x,u) \in B_0(\alpha)\} =: [a(x), b(x)].$$

This interval guarantees $\Pr\{X_{n,m} \in B(X_n, \alpha)\} \geq \alpha$.

3.4. Unknown τ

If τ is unknown, we use X_{n_0} and X_n, $n_0 < n$, to estimate θ and τ and to predict $X_{n,m}$. That is, the estimating equation is

$$\mu_F(n_0, \hat{\theta}, \hat{\tau}) = X_{n_0} \quad \text{and} \quad \mu_F(n, \hat{\theta}, \hat{\tau}) = X_n, \tag{3.7}$$

where the latter equation can be replaced by

$$\mu_F(n - n_0, \hat{\theta}, \hat{\tau} + n_0) = X_n - X_{n_0}. \tag{3.8}$$

Unfortunately, the estimate is not stable with respect to the choice of n_0. Fortunately, however, if n is large compared with m the estimate of $X_{n,m}$ is not affected much by this instability. Under the same condition, the convention $X_1 = 1$ seems to have little effect on the estimate $X_{n,m}$. The application to the Nippon Derby data in Section 5 confirms this fact.

If τ is unknown and the observation $(y_1, \dots, y_n)$ of $(Y_1, \dots, Y_n)$ is available, we maximize the likelihood

$$l(\theta, \tau) = \sum_{k=1}^{n} (y_k \log \theta + (1 - y_k) \log(\tau + k - 1) - \log(\theta + \tau + k - 1)) \tag{3.9}$$

with respect to θ and τ to obtain the m.l. equations

$$x_n/\theta = \sum_{y_k=0} (\tau + k - 1)^{-1} = \sum_{k=1}^{n} (\theta + \tau + k - 1)^{-1}, \tag{3.10}$$

where $x_n = \sum_{k=1}^{n} y_k$ is the observation of X_n. These equations can be solved by the Newton-Raphson method starting from the moment estimates, or from the m.l. estimates of θ for some preassigned values of τ.

Testing the hypothesis $\tau = 0$ and interval prediction of $X_{n,m}$ for unknown τ are open problems.

4. Prediction of Waiting Time

4.1. Future events

Let us turn to the prediction of W_{s+t} based on the given value $W_s = w$. From the relation

$$\begin{aligned} &\Pr\{W_{s+t} = u\} \\ &= \sum_{w+v=u} \Pr\{W_s = w\} \Pr\{W_{s+t} - W_s = v \mid W_s = w\} \\ &= \sum_{w+v=u} R_1(w-1, s-1; \tau) \frac{\theta^s}{(\theta + \tau)^{[w]}} R_1(v - 1, t - 1; \tau + s) \frac{\theta^t}{(\theta + \tau + s)^{[v]}}, \end{aligned} \tag{4.1}$$

we find

$$\Pr\{W_{s+t} = w + v \mid W_s = w\} = R_1(v - 1, t - 1; \tau + s) \frac{\theta^t}{(\theta + \tau + s)^{[t]}}, \tag{4.2}$$

$$\begin{aligned} &v = t, t + 1, \dots; \quad t = 1, 2, \dots; \\ &w = s, s + 1, \dots; s = 1, 2, \dots; \quad 0 < \theta < \infty; \quad 0 \le \tau < \infty. \end{aligned}$$

That is, $W_{s+t}\,|W_s$ has the distribution $\text{STR1W}(t, \theta, \tau + s)$, and it depends on the fact the sth u.n.r. was observed, but does not on when it occurred. The point prediction of $W_{s+t}\,|W_s$ or the point estimation of

$$E(W_{s+t} \mid W_s = w) = \mu_W(t, \theta, \tau + s), \tag{4.3}$$

is possible when the expectation is finite, namely, only if $\theta > 1$.

4.2. Point estimation

We discuss only the known τ case. The score function of $p_W(\cdot\,; s, \theta, \tau)$ is $s\theta^{-1} - \sum_{k=1}^{W_s}(\theta + \tau + k - 1)^{-1}$ and the m.l. equation of θ is expressed in terms of μ_F (not μ_W), $s = \mu_F(W_s, \hat{\theta}, \tau)$. Compare it with the m.l. equation in (3.2), $X_n = \mu_F(n, \hat{\theta}, \tau)$. The same computer function, with the arguments of the modified role, is utilized for solving the equation.

The moment estimating equation of θ is $W_s = \mu_W(s; \hat{\theta}, \tau)$, whose solution is more complicated than that of the m.l. equation.

A point estimate of $W_{s+t} - W_s$, using one of the above estimates $\hat{\theta}$, is $\mu_W(t, \hat{\theta}, \tau + s)$. The Fisher information of p_W is

$$I_W(\theta; s, \tau) = s\theta^{-2} \sum_{w=0}^{\infty} p_W(w; s, \theta, \tau) \sum_{k=0}^{w} (\theta + \tau + k - 1)^{-2},$$

where the summations start from $w = k = 1$ if $\tau = 0$. Compare I_W with I_F in (3.5). The estimation is difficult unless θ is large enough and many u.n.r.'s are occurring.

5. Applications

5.1. Vancouver precipitation

Glick (1978) listed monthly total precipitations during 1900–1973, except for 1904 and 1905, and monthly total hours of bright sunshine during 1909–1973, in Vancouver, B.C. We predict the number of u.n.r.'s of monthly total precipitation, for each month, in the last 26 years from that in the first 39 years (the 3/5 parts), assuming $\tau = 0$.

The data and actual results are shown in Table 1. The predictors of θ, $X_{n,m}$ and the estimates of their standard deviations are computed following Subsection 3.2, and the interval prediction of $X_{n,m}$ is computed following Subsection 3.3. All results are summarized in Table 2. Observing Table 2, we can accept the i.i.d. assumption, $\tau = 0$ and $\theta = 1$. The randomness tests based on the total 65 years are not significant, and the predictors look reasonable.

Table 1. Vancouver monthly precipitation. Number of upper new records during the first n=39 years and the last m=26 years.

month	Ja	Feb	Mar	Ap	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
X_n	4	4	2	3	4	3	5	2	6	3	3	5
$X_{n,m}$	0	1	0	0	2	1	0	0	0	1	1	0

Table 2. Point and interval prediction of the future record breakings based on the past in Table 1. $[a, b]$ is the interval covering $X_{n,m}$ with confidence coefficient 0.9017.

X_n	$\hat{\theta}$	SD($\hat{\theta}$)	$\hat{X}_{n,m}$	SD($\hat{X}_{nm}$)	a	b	actual $X_{n,m}$
1	0.0000	0.0000	0.0000	0.0000	0	1	–
2	0.2640	0.0758	0.1839	0.1829	0	1	0,0
3	0.5744	0.1919	0.3976	0.3930	0	1	0,1,1,1
4	0.9325	0.3566	0.6408	0.6289	0	2	0,1,2
5	1.3400	0.5802	0.9134	0.8893	0	2	0,0
6	1.7994	0.8752	1.2156	1.1730	0	3	0
7	2.3140	1.2575	1.5477	1.4786	0	3	–
8	2.8878	1.7472	1.9103	1.8052	0	3	–
9	3.5258	2.3696	2.3044	2.1516	0	4	–

Actually, there were dry years, especially dry winters, in the beginning, and if we use the lower new records the predicted values $\hat{\theta}$ and $\hat{X}_{n,m}$ are larger and the actual values of $X_{n,m}$ are smaller. It is an open problem to predict upper and lower records simultaneously.

5.2. Tokyo summer temperature and horse race

Iiyama et al. (1995) listed Tokyo August mean temperatures during 1950–1994, and the winning times of Nippon Derby (Tokyo thoroughbred race) during 1932–1994, except for 1945 and 1946. The mean temperature is the monthly arithmetic mean of 24 hourly measurements of everyday. We study first the u.n.r.'s of the temperature and the lower new records of the winning time, applying the predictors of Subsection 3.2. The times (years) when a new record occurred are listed in Table 3.

Table 3. Data for Figs. 2 and 3. The time when a new record occurred. From Iiyama et al. (1995).

Fig.	Dataset	n	the years when a new record occurred
2	Tokyo Aug. T.	45	1, 2, 3, 5, 8, 13, 24, 29, 45
3	Nippon Derby	61	1, 2, 6, 7, 11, 12, 18, 24, 27, 28, 30, 39, 40, 41, 46, 49, 55, 57

Using the data of N years (N =45 or 60) we predict the last $N-n$ years from the first n years, changing $n = 22, \ldots, 40$ for August temperature and $n = 30, \ldots, 56$ for Nippon Derby. The results are summarized in Figures 2 and 3. In Figures 2a and 3a for θ, $\hat{\theta}$'s are plotted by dots against n, and $\hat{\theta} \pm 2SD(\hat{\theta})$ are plotted by minuses. In Figures 2b and 3b for $X_{n,m}$, $\hat{X}_{n,m}$ are plotted by dots, $\hat{X}_{n,m} + 2SD(\hat{X}_{n,m})$ by minuses, and actual values of $X_{n,m}$ by pluses. The values of $\hat{X}_{n,m} - 2SD(\hat{X}_{n,m})$ are negative and not shown.

Fig. 2a

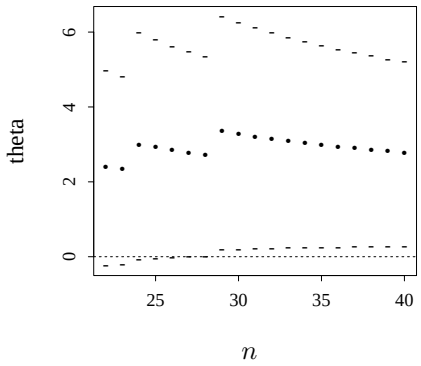

Fig. 2b

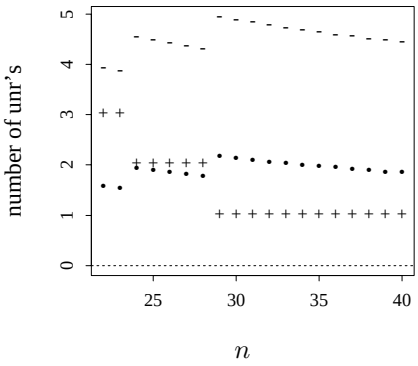

Figure 2. Tokyo August temperature (45 years).

Fig. 3a

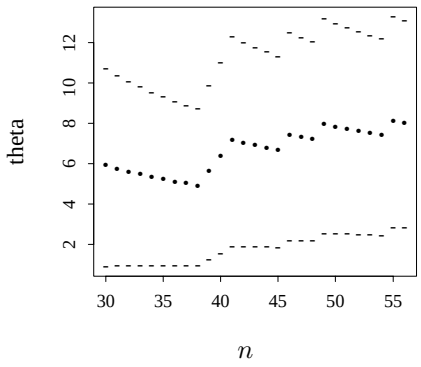

Fig. 3b

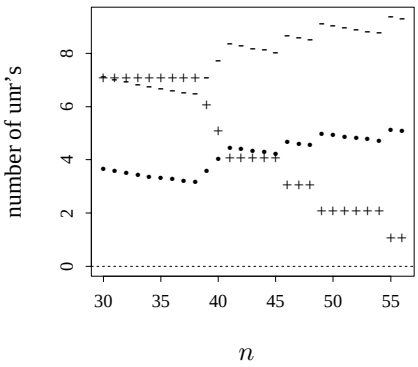

Figure 3. Nippon Derby (61 years).

In both datasets, θ is larger than 1, and the hypothesis of randomness is rejected in both cases. After 1950, the effect of the increasing energy consumption and the green-house effect is apparent in Tokyo temperature. The horse-race speed is steadly up (from 2 min. 45 sec. to 2 min. 25 sec.) in these years due to improvement of race field, training of riders and horses, and breeding.

In Fig. 2b, the actual observations of $X_{n,m}$ are not far from the predicted values, and the assumption $\tau = 0$ seems to work. The estimates $\hat{\theta}$ are rather stable. The point predicion of the waiting time of Subsection 4.2 is applied to this case, with $W_9 = 45$ regarded as the given data. The results in Table 4 are comparable to the point estimate $\hat{X}_{45,5} = 0.31$ and $\hat{X}_{45,10} = 0.59$.

In Fig. 3b, the gaps between the actual observations $X_{n,m}$ and the predicted values are larger, and the estimates $\hat{\theta}$ are not stable. Hence, the prediction method of Subsection 3.4, using the estimating equation (3.7) and the m.l. equation (3.10), is applied to get the results in Table 5. Compared with the case where $\tau = 0$ is assumed, the two-parameter model gives a more reasonable predictor, although estimates vary by the prediction methods. In the table, to check the effect of disregarding the convention $X_1 = 1$, the estimate from the data neglecting the first occurrence is computed, with $(n_0 = 30, n = 61)$. The effect is comparable with that of changing the prediction method. The likelihood function of this example has a steep edge and the maximization is difficult. This fact may correspond to the instability in the moment method.

Table 4. Tokyo August temperature (Fig. 2 and Table 3). Point estimation of W_{s+t} from the last occurrence $s = 9$ and $W_s = 45$. $\hat{\theta} = 3.10$.

t	1	2	3	4	5
$\hat{W}_{s+t}$	5.28	13.07	24.56	41.52	66.53

Table 5. Nippon Derby (Fig. 3 and Table 3). Point prediction (unknown τ) of the future record breakings.

methods	$\hat{\theta}$	$\hat{\tau}$	$\hat{X}_{61,5}$	$\hat{X}_{61,10}$
τ=0	8.25	–	0.58	1.12
$n_0 = 20$ (1/3)	34.18	54.32	1.79	3.32
$n_0 = 30$ (1/2)	18.23	18.48	2.09	3.57
$n_0 = 40$ (2/3)	28.24	40.69	1.86	3.39
1st negl. (1/2)	23.41	29.80	1.95	3.46
max. likel.	21.34	25.22	1.99	3.50

5.3. Olympic Games (22 times, 1896-1992)

Modern Olympic Games started 1896. Among 25 planned sessions, three were cancelled by the First and Second World Wars, and there are at most 22 golden medalist records for each game (Watanabe (1989)). The marathon course distance was not fixed in the early era. Hence, the times in the era are modified according to the present 42.195km. Many record-breakings are still occurring in traditional sports, except for some field sports and races. Since the data are not long, the two parameter model is not adequate, and we assume $\tau = 0$.

Table 6 shows the interval prediction, with confidence coefficient more than 0.9, of the number of record-breakings in future five games.

Table 6. Interval prediction $[a, b]$ of the number of new records in future 5 Olympic Games.

category	sports	n	X_n	$\hat{\theta}$	a	b
field (men)	long jump	22	9	5.17	0	2
	triple jump	22	13	12.46	0	4
	shot put	22	16	24.89	0	5
	discus	22	17	32.46	1	5
track (men)	100m race	22	8	4.07	0	2
	400m race	22	10	6.49	0	3
	800m race	22	14	15.53	0	4
	1500m race	22	14	15.53	0	4
marathon (men)		22	12	10.04	0	3
swimming (men)	100m free	19	16	45.01	1	5
	400m free	19	18	158.78	3	5
	1500m free	19	14	22.46	0	4
	800m relay	19	17	73.39	2	5
swimming (women)	100m free	16	11	14.27	0	4
	400m free	16	15	109.78	2	5
	400m relay	16	15	109.78	2	5

confidence coefficients of $[a, b]$	n=22	α =0.9016
	19	0.9004
	16	0.9038

In the long jump, 100m and 400m races, and the marathon, all for men, the record-breakings are saturated. In other field and track sports, the record-breakings are starting to saturate. In swimming, specially in the 400m free and 800m relay for men, and the 400m free and 400m relay for women, new record-breakings are promising.

Acknowledgements

This paper is based on an invited report at the 3rd ICSA Statistical Conference, August 1995, Beijing. The authors have benefitted from valuable discussions at the conference.

References

Broder, A. Z. (1984). The r-Stirling numbers. *Discrete Math.* **49**, 241-259.

Carlitz, L. (1980a). Weighted Stirling numbers of the first and second kind—I. *Fibonacci Quart.* **18**, 147-162.

Carlitz, L. (1980b). Weighted Stirling numbers of the first and second kind—II. *Fibonacci Quart.* **18**, 242-257.

Fu, J. C. (1996). Distribution theory of runs and patterns associated with a sequence of multi-state trials. *Statist. Sinica* **6**, 957-974.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Wiley, New York. (2nd ed. 1987, Krieger, Malabar, FL.)

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

Graham, R. L., D. E. Knuth and O. Patashnik (1989). *Concrete Mathematics.* Addison-Wesley, Reading, Mass.

Iiyama, Y., Nishimura, K. and Sibuya, M. (1995). Power of record-breaking test. *Japan. J. Appl. Statist.* **24**, 13-26. (in Japanese)

Koutras, M. (1982). Non-central Stirling numbers and some applications. *Discrete Math.* **42**, 73-89.

Nagaraja, H. N. (1988). Record values and related statistics—A review. *Commun. Statist. Theory Methods* **17**, 2223-2238.

Neuman, E. (1987). Stirling polynomials of the second kind. *J. Combinat. Math. Combinat. Comput.* **1**, 175-180.

Nevzorov, V. B. (1984). Record moments in the case of nonidentically distributed random variables. *Theory Probab. Appl.* **29**, 845-846.

Nevzorov, V. B. (1987). Records. *Theory Probab. Appl.* **32**, 201-228.

Nishimura, K. and Sibuya, M. (1997). Extended Stirling family of discrete probability distributions. *Commun. Statist. Theory Methods* **26**, 1727-1744.

Shanmugam, R. (1984). On central versus factorial moments. *South African Statist. J.* **18**, 97-110.

Sibuya, M. (1986). Stirling family of probability distributions. *Japan. J. Appl. Statist.* **15**, 131-146. (in Japanese)

Sibuya, M. (1988). Log-concavity of Stirling numbers and unimodality of Stirling distributions. *Ann. Inst. Statist. Math.* **40**, 693-714.

Takeuchi, K. (1975). *Tokei-teki Yosoku Ron* (*Theory of Statistical Prediction*). Baifu-kan, Tokyo. (in Japanese)

Watanabe, S. (ed.) (1989). *Encyclopedia Nipponica 2001.* Shogakukan, Tokyo. (in Japanese)

Department of Mathematics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223 Japan.

E-mail: sibuyam@takachiho.ac.jp

Department of Business Administration, Komazawa University, 1 Komazawa, Setagaya-ku, Tokyo, 154 Japan.

E-mail: nishimura@math.keio.ac.jp

(Received January 1996; accepted November 1996)