Adv. Appl. Prob. **47,** 1175–1189 (2015)
Printed in Northern Ireland

RECORDS FROM STATIONARY OBSERVATIONS SUBJECT TO A RANDOM TREND

RAÚL GOUET,* *Universidad de Chile*
F. JAVIER LÓPEZ ** *** AND
GERARDO SANZ,** **** *Universidad de Zaragoza*

Abstract

We prove strong convergence and asymptotic normality for the record and the weak record rate of observations of the form $Y_n = X_n + T_n$, $n \geq 1$, where $(X_n)_{n\in\mathbb{Z}}$ is a stationary ergodic sequence of random variables and $(T_n)_{n\geq 1}$ is a stochastic trend process with stationary ergodic increments. The strong convergence result follows from the Dubins–Freedman law of large numbers and Birkhoff's ergodic theorem. For the asymptotic normality we rely on the approach of Ballerini and Resnick (1987), coupled with a moment bound for stationary sequences, which is used to deal with the random trend process. Examples of applications are provided. In particular, we obtain strong convergence and asymptotic normality for the number of ladder epochs in a random walk with stationary ergodic increments.

Keywords: Record; stationary process; ergodic theorem; strong convergence; random trend; asymptotic normality

2010 Mathematics Subject Classification: Primary 60G10; 60G70
Secondary 60F05; 60F15

1. Introduction

Records capture attention as they arise in diverse domains such as economics or meteorology and, of course, sport. The mathematical theory has been developed over decades and reached a fair level of maturity, which can be appreciated in [1] and [21]; see also [14] for recent results on record counts from independent and identically distributed (i.i.d.) observations.

The literature on statistical analysis of record data reveal that records occur more often than predicted by the standard i.i.d. theory. This was pointed out in [9], where a model with linear deterministic trend was considered. Later, a power model which partly retains the theoretical simplicity of the i.i.d. case was introduced in [27].

The theory of records from observations with linear trend was initiated by Ballerini and Resnick in [2]. They obtained strong convergence and asymptotic normality for the record rate from observations of the form $Y_n = X_n + cn$, where the X_n are integrable, i.i.d. with continuous common distribution, and c is a positive constant. These results were later extended to stationary X_n in [3] with applications to athletic data. For additional theoretical results on the

Received 10 June 2014; revision received 24 November 2014.
* Postal address: Dpto. Ingeniería Matemática and Centro de Modelamiento Matemático (UMI 2807, CNRS), Universidad de Chile, Beauchef 851, 8370456 Santiago, Chile. Email address: rgouet@dim.uchile.cl
** Postal address: Dpto. Métodos Estadísticos and BIFI, Facultad de Ciencias, Universidad de Zaragoza. C/ Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
*** Email address: javier.lopez@unizar.es
**** Email address: gerardo.sanz@unizar.es

model with deterministic trend, linear or not, see [5] and [6]. Also, interesting distribution-free inference methods were developed in [8].

The study of record events has attracted the interest of scientists beyond the probability-statistics community in recent years. In particular, a fresh look at the problem of records from observations with linear trend can be found in the physics literature, for example, [10], [19], [20], and [25]. See also [23] and [26] for applications of the model with deterministic trend to the analysis of climate change.

The main results of this paper are the strong convergence (to a positive constant) and a central limit theorem for the record and the weak record rate in a model consisting of stationary ergodic observations, subject to a stochastic trend process, whose increments are stationary ergodic. These results generalize those of [3] for stationary observations with a deterministic linear trend.

The proof of the strong convergence of the record rate relies on a result of [7] concerning the almost sure (a.s.) convergence of the ratio of the sum of indicators to the sum of their conditional expectations with respect to an increasing family of sigma algebras. We show that the process of conditional expectations couples with a stationary process and we then apply Birkhoff's pointwise ergodic theorem to obtain the strong convergence of the record rate, unlike in [3], where the proof is based on Kingman's subadditive ergodic theorem. For the central limit theorem we consider first a martingale approach, which leads to asymptotic normality with a random centering process. Then we follow the strategy in [3] to obtain a central limit theorem with deterministic centering, where, as can be expected, extra moment and mixing conditions are needed due to the presence of a stochastic trend process.

We provide various examples of applications of our results. In particular, we analyze the case of random walks with stationary increments. This problem has been studied in the literature when the increments are independent [20] and [24]; our results are more general since they include the case of correlated increments.

2. Definitions and preliminaries

Let the base process $(W_n)_{n\in\mathbb{Z}}$ with $W_n = (X_n, \tau_{n+1})$ be defined as a bivariate, (strictly) stationary, and ergodic random sequence such that $\mathbb{E}[X_0^+] < \infty$ and $0 < c := \mathbb{E}[\tau_0] < \infty$, where $\mathbb{Z}$ denotes the set of integers, $x^+ := x \vee 0$, and $u \vee v := \max\{u, v\}$. The base process is taken as double-ended stationary for convenience, since any stationary single-ended sequence can be extended to a double-ended one. Also, ergodicity is assumed for ease since, otherwise, the asymptotic record rate has to be expressed as an expectation, conditional on the σ-algebra of invariant events of $(W_n)_{n\in\mathbb{Z}}$. We use the notation W_m^n for $(W_m, \dots, W_n)$ with $-\infty \le m \le n \le \infty$.

Let $(Y_n)_{n\ge 1}$ be the sequence defined by

$$Y_n = X_n + T_n,$$

where $T_n := \sum_{k=1}^n \tau_k, n \ge 1$, denotes the random trend or drift process. The first observation Y_1 is conventionally taken as a record and for $n \ge 2$, Y_n is said to be a (upper) record if $Y_n > M_{n-1}$, where $M_{n-1} := \max\{Y_1, \dots, Y_{n-1}\}$ (also denoted by $\bigvee_{i=1}^{n-1} Y_i$). The record indicators are then given by $I_1 = 1$ and $I_n = \mathbf{1}_{\{Y_n > M_{n-1}\}}, n \ge 2$. Finally, the counting process of records is defined by the sums of indicators $N_n = \sum_{k=1}^n I_k$ and the record rate by $N_n/n, n \ge 1$.

Remark 1. Note that the random drift T_n can be described as positive and linear in expectation because $\mathbb{E}[T_n] = nc > 0$. Observe also that Y_n can be decomposed as $Y_n = X_n' + nc$

https://doi.org/10.1239/aap/1449859805 Published online by Cambridge University Press

with $X'_n := X_n + T_n - nc$. Such representation apparently implies that the random drift can be reduced to a linear deterministic drift. However, this is not so because X'_n is not stationary in general and so the type of sequence Y_n studied in this paper generalizes those previously considered in the literature. On the other hand, we point out that both sequences (X_n) and (τ_n) are allowed to be dependent (correlated), also possibly mutually dependent, but must yet have finite expectation.

Lemma 1. (i) *It holds that $M_n \to \infty$ and $N_n \to \infty$ a.s.*

(ii) *The sequence $Z_n := M_{n-1} - T_n$, $n \geq 2$, satisfies the recurrence relation*

$$V_{n+1} = (V_n \vee X_n) - \tau_{n+1}. \tag{1}$$

Proof of Lemma 1(i). As M_n is increasing, it converges to a finite limit or diverges to ∞ a.s. On the other hand, for all $a \in \mathbb{R}$, we have

$$\mathbb{P}[M_n > a] \geq \mathbb{P}[X_n > a - T_n] \geq \mathbb{P}\left[X_n > a - \frac{nc}{2}, T_n \geq \frac{nc}{2}\right] \to 1,$$

since $\mathbb{P}[X_n > a - nc/2] = \mathbb{P}[X_0 > a - nc/2] \to 1$ and $\mathbb{P}[T_n \geq nc/2] \to 1$, by Birkhoff's theorem. Hence, $M_n \to \infty$, which clearly implies $N_n \to \infty$.

Proof of Lemma 1(ii). By direct substitution into (1).

We show next that (1) has a stationary solution, which couples with Z_n. Stochastic recursions appear in many areas of applied probability; see [11] for results related to the (max, +) algebra.

Proposition 1. *Let $Z^*_n = \bigvee_{k\geq 1}\{X_{n-k} - \sum_{j=n-k+1}^{n} \tau_j\}$, $n \in \mathbb{Z}$. Then*

(i) *Z^*_n is a proper stationary solution of (1) and*

(ii) *$Z^*_n = Z_n$ a.s. for sufficiently large n.*

Proof of Proposition 1(i). Note that Z^*_n is a measurable function of $W^{n-1}_{-\infty}$, so Z^*_n is stationary. Also, substitution into (1) shows that Z^*_n solves the recurrence.

We verify that Z^*_n is proper, that is, $\mathbb{P}[Z^*_n \in \mathbb{R}] = 1$. Due to stationarity it suffices to show that $\mathbb{P}[X_{-k} > \sum_{j=-k+1}^{0} \tau_j, \text{i.o.}] = 0$ (i.o. stands for 'infinitely often'). Birkhoff's theorem implies that $\mathbb{P}[\sum_{j=-k+1}^{0} \tau_j \leq kc/2, \text{i.o.}] = 0$; hence,

$$\mathbb{P}\left[X_{-k} > \sum_{j=-k+1}^{0} \tau_j, \text{i.o.}\right] \leq \mathbb{P}\left[X_{-k} > \frac{kc}{2}, \text{i.o.}\right].$$

Furthermore, by stationarity $\mathbb{P}[X_{-k} > kc/2] = \mathbb{P}[X_0 > kc/2]$ for $k \geq 1$, and $\sum_{k=1}^{\infty}\mathbb{P}[X_0 > kc/2] < \infty$ because $\mathbb{E}[X_0^+] < \infty$. The conclusion then follows from the Borel–Cantelli lemma.

Proof of Proposition 1(ii). We iterate (1) with starting value Z_2 to obtain

$$Z_n = \bigvee_{k=1}^{n-2}\left(X_{n-k} - \sum_{j=n-k+1}^{n} \tau_j\right) \vee \left(Z_2 - \sum_{j=3}^{n} \tau_j\right), \qquad n \geq 2.$$

We claim that for large enough n, $Z_n = \bigvee_{k=1}^{n-2}(X_{n-k} - \sum_{j=n-k+1}^{n} \tau_j)$ a.s. because $Z_2 - \sum_{j=3}^{n}\tau_j \to -\infty$ a.s., by Birkhoff's theorem. To prove the claim let us assume on the contrary that $\mathbb{P}[Z_n = Z_2 - \sum_{j=3}^{n}\tau_j,\ \text{i.o.}] > 0$, which, from the definition of Z_n in Lemma 1(ii), is equivalent to $\mathbb{P}[M_{n-1} = M_1,\ \text{i.o.}] > 0$. As this contradicts Lemma 1(i), the claim is proven.

On the other hand, by iterating (1) with starting value Z_2^*, we obtain

$$Z_n^* = \bigvee_{k=1}^{n-2}\left(X_{n-k} - \sum_{j=n-k+1}^{n}\tau_j\right) \vee \left(Z_2^* - \sum_{j=3}^{n}\tau_j\right), \qquad n \geq 2.$$

So from the previous claim, we have $Z_n^* = Z_n \vee (Z_2^* - \sum_{j=3}^{n}\tau_j)$ a.s. for large enough n. Finally, we obtain $Z_n^* = Z_n$ for large enough n because

$$\mathbb{P}\left[Z_2^* - \sum_{j=3}^{n}\tau_j > Z_n,\ \text{i.o.}\right] = \mathbb{P}[M_{n-1} < Z_2^* + T_2,\ \text{i.o.}] = 0,$$

by Lemma 1(i).

Definition 1. Let $(\mathcal{F}_n)_{n\in\mathbb{Z}}$ be the increasing family of σ-algebras given by

$$\mathcal{F}_n = \sigma\{X_k, \tau_{k+1}, k \leq n\} = \sigma\{W_{-\infty}^{n}\}, \qquad n \in \mathbb{Z}.$$

Also, let $G_{n-1}(x) = \mathbb{P}[X_n > x \mid \mathcal{F}_{n-1}]$ and $G_{n-1}^{\mathrm{w}}(x) = \mathbb{P}[X_n \geq x \mid \mathcal{F}_{n-1}]$ for $x \in \mathbb{R}, n \in \mathbb{Z}$, be the (regular, conditional on $\mathcal{F}_{n-1}$) survival function and the weak survival function of X_n, respectively.

Proposition 2. *Let $G_{n-1}(x)$ and $G_{n-1}^{\mathrm{w}}(x)$ from Definition 1 and let Z_n^* be as defined in Proposition 1. Then $G_{n-1}(Z_n^*)$, $G_{n-1}^{\mathrm{w}}(Z_n^*)$, $n \in \mathbb{Z}$, are stationary and ergodic.*

Proof. First note that Z_n^* is $\mathcal{F}_{n-1}$-measurable; hence, $G_{n-1}(Z_n^*) = \mathbb{P}[X_n > Z_n^* \mid W_{-\infty}^{n-1}]$, $n \in \mathbb{Z}$. From the definition of conditional expectation, there exists a measurable function $f_0 \colon \mathbb{R}^\infty \to \mathbb{R}$ such that $G_0(Z_1^*) = \mathbb{E}[\mathbf{1}_{\{X_1 > Z_1^*\}} \mid W_{-\infty}^{0}] = f_0(W_{-\infty}^{0})$ and

$$\mathbb{E}[f_0(W_{-\infty}^{0})g(W_{-\infty}^{0})] = \mathbb{E}[\mathbf{1}_{\{X_1>Z_1^*\}}g(W_{-\infty}^{0})]$$

for any bounded and measurable function $g \colon \mathbb{R}^\infty \to \mathbb{R}$. We claim that $f_0(W_{-\infty}^{n-1})$ is a version of $\mathbb{E}[\mathbf{1}_{\{X_n>Z_n^*\}} \mid W_{-\infty}^{n-1}]$ and, therefore, that $G_{n-1}(Z_n^*)$ is stationary and ergodic. The claim follows at once from the stationarity (and ergodicity) of W_n since, for $n \in \mathbb{Z}$,

$$\mathbb{E}[f_0(W_{-\infty}^{n-1})g(W_{-\infty}^{n-1})] = \mathbb{E}[f_0(W_{-\infty}^{0})g(W_{-\infty}^{0})]$$

and

$$\mathbb{E}[\mathbf{1}_{\{X_1>Z_1^*\}}g(W_{-\infty}^{0})] = \mathbb{E}[\mathbf{1}_{\{X_n>Z_n^*\}}g(W_{-\infty}^{n-1})].$$

The argument for $G_{n-1}^{\mathrm{w}}(Z_n^*)$ is identical.

3. Main results

3.1. Strong convergence of the record and the weak record rate

The strong convergence of the record rate for stationary observations with random drift, is contained in the following theorem.

Theorem 1. *It holds that*

$$\frac{N_n}{n} \to p := \mathbb{E}[G_0(Z_1^*)] = \mathbb{P}\left[X_1 > \bigvee_{k\geq 1}\left\{X_{1-k} - \sum_{j=2-k}^{1}\tau_j\right\}\right] \quad \textit{a.s.}$$

https://doi.org/10.1239/aap/1449859805 Published online by Cambridge University Press

Proof. Let $G_{n-1}(x)$ and Z_n be as described in Definition 1 and Lemma 1(ii). We invoke Proposition 6 (see Appendix A) with $U_n = I_n$ and $\mathcal{G}_n = \mathcal{F}_n$, recalling that, from Lemma 1(i), $N_n \to \infty$ a.s. Hence, $\sum_{n\geq 1} U_n = \infty$ a.s. and (10) holds. Furthermore, the conditional expectation of I_n is easily calculated as

$$\begin{aligned}\mathbb{E}[I_n \mid \mathcal{F}_{n-1}] &= \mathbb{P}[Y_n > M_{n-1} \mid \mathcal{F}_{n-1}] \\ &= \mathbb{P}[X_n > M_{n-1} - T_n \mid \mathcal{F}_{n-1}] \\ &= G_{n-1}(Z_n), \qquad n \geq 2,\end{aligned} \tag{2}$$

so, from (2) and (10), we obtain

$$\frac{N_n}{\sum_{k=2}^{n} G_{k-1}(Z_k)} \to 1 \quad \text{a.s.} \tag{3}$$

On the other hand, by Proposition 2, $G_{n-1}(Z_n^*)$ is stationary ergodic and so Birkhoff's theorem yields that

$$\frac{1}{n}\sum_{k=1}^{n} G_{k-1}(Z_k^*) \to \mathbb{E}[G_0(Z_1^*)] \quad \text{a.s.} \tag{4}$$

Furthermore, from Proposition 1(ii) we know that Z_n and Z_n^* couple; hence, (4) also holds with Z_n replacing Z_n^*, that is,

$$\frac{1}{n}\sum_{k=1}^{n} G_{k-1}(Z_k) \to \mathbb{E}[G_0(Z_1^*)] \quad \text{a.s.}$$

and the conclusion follows from (3).

A weak record is an observation which is greater than or equal to the current maximum. We define the indicators of weak records by $I_1^{\mathrm{w}} = 1$ and $I_k^{\mathrm{w}} = \mathbf{1}_{\{Y_k \geq M_{k-1}\}}$, $k \geq 2$; the counting process and the rate by $N_n^{\mathrm{w}} = \sum_{k=1}^{n} I_k^{\mathrm{w}}$ and N_n^{w}/n, respectively. Of course, records and weak records coincide unless the distribution of the observations has discontinuities; see [12] and [13] for results in the i.i.d. case. Observe that $N_n^{\mathrm{w}} \geq N_n \to \infty$, by Lemma 1(i). We now state the analog of Theorem 1 for weak records.

Theorem 2. *It holds that*

$$\frac{N_n^{\mathrm{w}}}{n} \to p^{\mathrm{w}} := \mathbb{E}[G_0^{\mathrm{w}}(Z_1^*)] = \mathbb{P}\Bigg[X_1 \geq \bigvee_{k\geq 1}\Bigg\{X_{1-k} - \sum_{j=2-k}^{1} \tau_j\Bigg\}\Bigg] \quad a.s. \tag{5}$$

Proof. As that of Theorem 1, mutatis mutandis.

We have the following result concerning the positivity of the limits in Theorems 1 and 2. Observe that the integrability hypothesis of X_0 is crucial.

Proposition 3. *Let p, p^{w} be as defined in Theorems 1 and 2. Then $p^{\mathrm{w}} \geq p > 0$.*

Proof. Clearly, since records are also weak records, we have $p^{\mathrm{w}} \geq p$ and so it suffices to prove that $p > 0$. Observe that $p = \mathbb{E}[G_0(Z_1^*)] = 0$ implies that $G_0(Z_1^*) = 0$ a.s. and so, by stationarity, $G_{n-1}(Z_n^*) = 0$ for all $n \in \mathbb{Z}$ a.s. Now, since Z_n^* and Z_n couple, the series $\sum_{n\geq 1} G_{n-1}(Z_n)$ converges. Therefore, by (9), $\sum_{n\geq 1} I_n < \infty$, thus contradicting Lemma 1.

https://doi.org/10.1239/aap/1449859805 Published online by Cambridge University Press

Remark 2. It is easy to find an example with $p = 1$ (see after Proposition 4). In this case all observations are records except for a finite number. Indeed, we consider the indicators of not being a record, that is, $1 - I_n$. Then, by (10), the total number of no-records $\sum_{n\geq 1}(1 - I_n)$ is finite if and only if $\sum_{n\geq 1}(1 - G_{n-1}(Z_n))$ is finite. The last sum converges because $p = 1$ implies that $G_0(Z_1^*) = 1$ a.s.

3.2. Asymptotic normality

The asymptotic normality of N_n was first investigated in [2] in the context of a base process W_n, where the X_n are i.i.d. with continuous distribution F, and the drift process is deterministic, that is, $\tau_n = c$. The result was later extended in [3] to stationary, strongly mixing, and square-integrable X_n, always under deterministic drift. Their method of proof relies on the approximation of the indicators I_n by stationary ones.

We consider first a different approach based on the conditional centering of N_n. It is clear that $N_n - \sum_{k=1}^{n}\mathbb{E}[I_k \mid \mathcal{F}_{k-1}]$, $n \geq 1$, is a martingale with bounded increments. So the martingale central limit theorem can be applied; see, for example, [17, Corollary 3.1]. To that end observe that the Lindeberg-type condition is satisfied and, letting $\xi_k = I_k - \mathbb{E}[I_k \mid \mathcal{F}_{k-1}]$, we have $\mathbb{E}[\xi_k^2 \mid \mathcal{F}_{k-1}] = G_{k-1}(Z_k)(1 - G_{k-1}(Z_k))$. Hence, by Proposition 1(ii), Proposition 2, and Birkhoff's theorem, we have

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\xi_k^2 \mid \mathcal{F}_{k-1}] \to \mathbb{E}[G_0(Z_1^*)(1 - G_0(Z_1^*))] \quad \text{a.s.}$$

We have thus proven the following proposition.

Proposition 4. *If* $\sigma_M^2 := \mathbb{E}[G_0(Z_1^*)(1 - G_0(Z_1^*))] > 0$ *then the following convergence holds:*

$$\frac{1}{\sqrt{n}}\left(N_n - \sum_{k=1}^{n} G_{k-1}(Z_k)\right) \xrightarrow{\mathrm{D}} N(0, \sigma_M^2),$$

where '$\xrightarrow{\mathrm{D}}$*' denotes convergence in distribution.*

Examples with $\sigma_M = 0$ are easy to construct. Take $(X_n)_{n\in\mathbb{Z}}$ i.i.d. uniform in $[0, 1]$ and $\tau_n = 3$, then $Y_n \in [3n, 3n + 1]$ and $G_{n-1}(Z_n) = 1$ for $n \geq 1$. So all observations are records and there is no asymptotic normality for N_n.

Proposition 5. *If* $\sigma_M = 0$*, the martingale* $N_n - \sum_{k=1}^{n} G_{k-1}(Z_k)$ *converges a.s.*

Proof. The argument is similar to that used in the proof of Proposition 3: $\sigma_M = 0$ implies $G_{k-1}(Z_k^*)(1 - G_{k-1}(Z_k^*)) = 0$, $k \in \mathbb{Z}$, and because of the coupling of Z_k and Z_k^*, the series $\sum_{k\geq 1}\mathbb{E}[\xi_k^2 \mid \mathcal{F}_{k-1}] = \sum_{k\geq 1} G_{k-1}(Z_k)(1 - G_{k-1}(Z_k))$ converges and so does the martingale.

Remark 3. In the above proof observe that the random variables $G_{k-1}(Z_k^*)$ take values in $\{0, 1\}$ and the same is true for $G_{k-1}(Z_k)$ for large enough k. Since $\sum_{k=1}^{\infty}(I_k - G_{k-1}(Z_k)) < \infty$, we have $I_k = G_{k-1}(Z_k)$ for large enough k. Suppose that, for example, X_n i.i.d. with distribution F and $\tau_n = c$, then $G_0(Z_1^*) = \mathbb{P}[X_1 > Z_1^* \mid \mathcal{F}_0] \geq \mathbb{P}[X_1 > X_0 - c \mid X_0] = 1 - F(X_0 - c) > 0$. So $\mathbb{P}[G_0(Z_1^*) = 0] = 0$, which implies that $G_{k-1}(Z_k^*) = 1$ for all k and $G_{k-1}(Z_k) = 1$ for large enough k. In other words, in this model $\sigma_M = 0$ entails $p = 1$ and all observations are records save a finite number. This is not true in the general case of stationary observations.

The result of Proposition 4 does not depend on any mixing condition on the base process, but it is not satisfactory because the centering sequence is random and there seems to be no simple way of replacing it by a deterministic one. We present below a second central limit theorem for N_n with deterministic centering, requiring the strong mixing of the base process W_n plus some moment conditions on X_n and τ_n. The proof follows closely that of [3, Theorem 2], but needs extra conditions for handling the tail probabilities in the presence of a random trend. In fact we rely on a bound for moments of stationary mixing sequences from [28]; see Lemma 2. We recall the definition of the α-mixing coefficients. Let $\mathcal{F}_{-\infty}^{0} = \sigma\{W_{-\infty}^{0}\}$, $\mathcal{F}_{n}^{\infty} = \sigma\{W_{n}^{\infty}\}$, and

$$\alpha(n) := \sup_{A\in\mathcal{F}_{-\infty}^{0},\, B\in\mathcal{F}_{n}^{\infty}} |\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]|.$$

Let also $\mathcal{F}_{-\infty}^{0,\tau} = \sigma\{\dots, \tau_{-1}, \tau_0\}$, $\mathcal{F}_{n}^{\infty,\tau} = \sigma\{\tau_n, \tau_{n+1}, \dots\}$, and

$$\alpha^{\tau}(n) := \sup_{A\in\mathcal{F}_{-\infty}^{0,\tau},\, B\in\mathcal{F}_{n}^{\infty,\tau}} |\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]|. \tag{6}$$

Theorem 3. *Suppose that* $\sum_{n\ge 1}\alpha(n) < \infty$, $\mathbb{E}[X_0^2] < \infty$, $\mathbb{E}[|\tau_0|^{r+a}] < \infty$, *and that for some* $r > 4$, $a > 0$, $\sum_{n\ge 1} n^{r/2+1}[\alpha^{\tau}(n)]^{a/(r+a)} < \infty$. *Then the following convergence holds:*

$$\frac{N_n - np}{\sqrt{n}} \xrightarrow{\mathrm{D}} N(0,\sigma^2) \quad \textit{if } \sigma > 0, \tag{7}$$

where $p = \mathbb{E}[I_0^*] = \mathbb{P}[X_0 > Z_0^*]$ *and* $\sigma^2 = p(1-p) + 2\sum_{n\ge 1}\gamma(n)$ *with* $\gamma(n) = \operatorname{cov}(I_0^*, I_n^*)$ *and* $I_n^* = \mathbf{1}_{\{X_n > Z_n^*\}}$ *for* $n \in \mathbb{Z}$.

Proof. We follow the strategy of [3], which consists of proving a central limit theorem for a sequence of strongly mixing indicators and then transferring the result to N_n. Let $Z_n^k = \bigvee_{i=1}^{k}\{X_{n-i} - \sum_{j=n-i+1}^{n}\tau_j\}$ for $k \ge 1$, $n \in \mathbb{Z}$, and recall that $Z_n^* = \bigvee_{i\ge 1}\{X_{n-i} - \sum_{j=n-i+1}^{n}\tau_j\}$. Let also $I_n^k = \mathbf{1}_{\{X_n > Z_n^k\}}$ for $n \in \mathbb{Z}$, and $N_n^k = \sum_{i=1}^{n} I_i^k$, $N_n^* = \sum_{i=1}^{n} I_i^*$ for $n \ge 1$.

As in [3, p. 807], we note that I_n^k, $n \in \mathbb{Z}$, is stationary with mixing coefficients $\alpha_k(n)$ such that $\alpha_k(n) \le 1$ for $n \le k$, and $\alpha_k(n) \le \alpha(n-k)$ for $n > k$. Since, by hypothesis, the mixing coefficients are summable, [18, Theorem 18.5.4] can be applied to yield the following result. Let $p_k = \mathbb{E}[I_0^k]$ and $\sigma_k^2 = \gamma_k(0) + 2\sum_{n\ge 1}\gamma_k(n)$ with $\gamma_k(n) = \operatorname{cov}(I_0^k, I_n^k)$. Then $\sigma_k^2 < \infty$ and, if $\sigma_k > 0$ the following convergence holds:

$$\frac{N_n^k - np_k}{\sqrt{n}} \xrightarrow{\mathrm{D}} N(0,\sigma_k^2). \tag{8}$$

The next step is to apply [4, Theorem 4.2] to obtain the asymptotic normality of N_n^* by letting $k \to \infty$ in (8). To that end, we first show (in Lemmas 3 and 4) that $p_k \to p$ and $\sigma_k \to \sigma$ as $k \to \infty$. Finally, in Lemma 5, we verify that

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}[|(N_n^k - np_k) - (N_n^* - np)| > \varepsilon\sqrt{n}] \to 0 \quad \text{for all } \varepsilon > 0.$$

The conclusion (7) follows because the coupling of Z_n and Z_n^* implies that $(N_n - N_n^*)/\sqrt{n} \to 0$ a.s.

Remark 4. If the increments τ_n of the trend process are bounded then the condition on $\alpha^{\tau}(n)$ in Theorem 3 can be relaxed to $\sum_{n\ge 1}(n+1)^{r/2+1}\alpha^{\tau}(n) < \infty$; see [28, Theorem 2].

Remark 5. Theorem 3 and the corresponding lemmas can be easily adapted to weak records, yielding

$$\frac{N_n^{\mathrm{w}} - np^{\mathrm{w}}}{\sqrt{n}} \xrightarrow{\mathrm{D}} N(0, \sigma_{\mathrm{w}}^2),$$

where $\sigma_{\mathrm{w}}^2 = p^{\mathrm{w}}(1 - p^{\mathrm{w}}) + 2\sum_{n\geq 1} \operatorname{cov}(I_0^{\mathrm{w}*}, I_n^{\mathrm{w}*})$ with $I_n^{\mathrm{w}*} = \mathbf{1}_{\{X_n \geq Z_n^*\}}$ for $n \in \mathbb{Z}$.

4. Examples

Example 1. Let $(X_n)_{n\in\mathbb{Z}}$ be i.i.d. with common distribution function F (not necessarily continuous).

(a) Let $\tau_n = c > 0$, $n \in \mathbb{Z}$. Then, from Theorems 1 and 2,

$$\frac{N_n}{n} \to p = \int \prod_{k\geq 1} \tilde{F}(x + kc) F(\mathrm{d}x), \qquad \frac{N_n^{\mathrm{w}}}{n} \to p^{\mathrm{w}} = \int \prod_{k\geq 1} F(x + kc) F(\mathrm{d}x),$$

where $\tilde{F}(x) = \mathbb{P}[X_1 < x]$. Also, Theorem 3 can be applied to obtain the asymptotic normality of N_n and N_n^{w}.
For the Gumbel distribution $F(x) = \exp(-\mathrm{e}^{-x})$, the explicit result $p = 1 - \mathrm{e}^{-c}$ is easily obtained; see [2]. This particular case is interesting in its own right because the sequence Y_n can be seen as an F^α-scheme, that is, the Y_n are independent with respective distribution functions $F_n(x) = F(x)^{\alpha_n}$, where $\alpha_n = \mathrm{e}^{nc}$. Therefore, the record indicators I_n are independent and so strong convergence and asymptotic normality follow; see [21] for information on the F^α-scheme. Also, the variance in Theorem 3, whose exact evaluation is in general out of reach, is given by $\sigma^2 = p(1 - p)$.

(b) Let $(\tau_n)_{n\in\mathbb{Z}}$ be i.i.d., independent of $(X_n)_{n\in\mathbb{Z}}$. Then, from Theorem 1,

$$\frac{N_n}{n} \to p = \int \mathbb{E}\bigg[\prod_{k\geq 1} \tilde{F}(x + T_k)\bigg] F(\mathrm{d}x).$$

For $F(x) = \exp(-\mathrm{e}^{-x})$, we have

$$p = \mathbb{E}\bigg[\int_0^1 u^{\sum_{k=1}^{\infty} \exp(-T_k)}\, \mathrm{d}u\bigg] = \mathbb{E}\bigg[\bigg(1 + \sum_{k=1}^{\infty} \mathrm{e}^{-T_k}\bigg)^{-1}\bigg].$$

Example 2. (*Ladder variables.*) Let $(\eta_n)_{n\geq 1}$ be a stationary ergodic sequence with $\mathbb{E}[\eta_1] > 0$ and let $S_n = \sum_{j=1}^n \eta_j, n \geq 1$, $S_0 = 0$. We are interested in the asymptotic record rate, denoted by λ, of the random walk (with positive drift) S_n. In this context, record times and record values are referred to as (ascending) ladder epochs and heights, respectively. To that end, we define a base process $(W_n)_{n\in\mathbb{Z}}$ with $X_n = 0$ for all n, and $(\tau_n)_{n\in\mathbb{Z}}$ the stationary ergodic double-ended extension of $(\eta_n)_{n\geq 1}$ with $\tau_n = \eta_n$ for $n \geq 1$. Given that the number of ladder epochs of S_n is equal to N_n, from Theorem 1, we obtain

$$\lambda = \mathbb{P}\bigg[\bigwedge_{k\geq 1}\bigg\{\sum_{j=1-k}^{0} \tau_j\bigg\} > 0\bigg],$$

where $\bigwedge$ denotes the min operator. Observe that λ depends on the auxiliary random variables τ_n, $n \leq 0$, instead of depending only on the original increments η_n. However, due to stationarity, we have

$$\lambda = \lim_{n\to\infty} \mathbb{P}[\eta_n > 0, \eta_n + \eta_{n-1} > 0, \dots, \eta_n + \cdots + \eta_1 > 0].$$

Observe also that $\lambda = \lim_{n\to\infty} \mathbb{P}[V_n > 0]$, where $V_n := \min\{\eta_n, \eta_n + \eta_{n-1}, \dots, \eta_n + \cdots + \eta_1\}$ satisfies $V_{n+1} = (V_n + \eta_{n+1}) \wedge \eta_{n+1}$, $n \geq 1$, with $V_1 = \eta_1$. This representation can be useful when $(\eta_n)_{n\geq 1}$ is a Markov chain since then $(V_n, \eta_n)_{n\geq 1}$ is also a Markov chain and λ can be obtained in terms of its stationary distribution.

On the other hand, if the increments are reversible, in the sense of $(\eta_1, \dots, \eta_n)$ and $(\eta_n, \dots, \eta_1)$ being equally distributed for all $n \geq 1$, then λ is simply $\mathbb{P}[\bigwedge_{k\geq 1} S_k > 0]$, the probability that the random walk stays strictly positive. Reversibility occurs, for instance, when η_n is a time-reversible Markov chain.

In the case of i.i.d. increments η_n, the limit above is well known since N_n can be seen as the counting function of a renewal process and, therefore, $N_n/n \to 1/\mathbb{E}[L_1]$, where L_1 is the first ladder epoch. See [16, Chapters 2 and 3] for more details on ladder variables. To the best of the authors' knowledge, the result in the general stationary case appears to be new.

We also consider weak records in the random walk S_n, $n \geq 0$, corresponding to weak ladder variables. From Theorem 2, the asymptotic rate of weak ladder epochs is given by

$$\lambda^{\mathrm{w}} = \mathbb{P}\bigg[\bigwedge_{k\geq 1}\bigg\{\sum_{j=1-k}^{0} \tau_j\bigg\} \geq 0\bigg].$$

In this case $\lambda^{\mathrm{w}} = \lim_{n\to\infty} \mathbb{P}[V_n' \geq 0]$, where $V_n' := V_n \wedge 0$, $n \geq 1$, satisfies $V_{n+1}' = (V_n' + \eta_{n+1}) \wedge 0$, $n \geq 0$, with $V_0' = 0$; that is, V_n' is a random walk taking values in $(-\infty, 0]$ with increments $(\eta_n)_{n\geq 1}$ and reflecting barrier at 0.

Example 3. (*Range of a Bernoulli random walk.*) Let $(\eta_n)_{n\geq 1}$ be a stationary ergodic sequence with $\eta_n \in \{-1, 1\}$, $\mathbb{P}[\eta_1 = 1] = \rho > \frac{1}{2}$, and let $S_n = \sum_{j=1}^{n} \eta_j$, $n \geq 1$, $S_0 = 0$. We consider R_n, the range of the walk up to time n, defined as the number of distinct values in $(S_1, S_2, \dots, S_n)$. As in Example 2, we define a base process $(W_n)_{n\in\mathbb{Z}}$ with $X_n = 0$, $n \in \mathbb{Z}$, and $(\tau_n)_{n\in\mathbb{Z}}$ the stationary ergodic double-ended extension of $(\eta_n)_{n\geq 1}$ with $\tau_n = \eta_n$ for $n \geq 1$. Note that, due to the nature of η_n, R_n and N_n are asymptotically equivalent in the sense that $R_n - N_n$ converges a.s. as $n \to \infty$. Hence, from Theorem 1, $R_n/n \to \lambda$, where λ is defined in Example 2. This result is a particular instance of the Kesten–Spitzer–Whitman theorem; see [22, p. 38].

Also, from Theorem 3, $(R_n - \lambda n)/\sqrt{n} \xrightarrow{\mathrm{D}} N(0, \sigma^2)$. For illustration we explicitly calculate below λ and σ in the case of i.i.d. increments η_n.

From the gambler's ruin problem, we have $\lambda = 2\rho - 1$. Now $\operatorname{cov}(I_0^*, I_n^*) = \mathbb{E}[I_0^* I_n^*] - \lambda^2$ and because of the independence of the τ_n, we obtain

$$\begin{aligned}
\mathbb{E}[I_0^* I_n^*] &= \mathbb{P}\bigg[\sum_{j=-k+1}^{0} \tau_j > 0, \sum_{j=n-k+1}^{n} \tau_j > 0 \text{ for all } k \geq 1\bigg] \\
&= \mathbb{P}\bigg[\sum_{j=0}^{k-1} \tau_j > 0, \sum_{j=n}^{n+k-1} \tau_j > 0 \text{ for all } k \geq 1\bigg]
\end{aligned}$$

https://doi.org/10.1239/aap/1449859805 Published online by Cambridge University Press

$$= \mathbb{P}\left[\sum_{j=0}^{k-1} \tau_j > 0,\ 1 \le k \le n\right] \mathbb{P}\left[\sum_{j=n}^{n+k-1} \tau_j > 0 \text{ for all } k \ge 1\right]$$

$$= \mathbb{P}\left[\sum_{j=0}^{k-1} \tau_j > 0,\ 1 \le k \le n\right] \lambda.$$

So $\operatorname{cov}(I_0^*, I_n^*) = \lambda \mathbb{P}[T_k > 0, 1 \le k \le n] - \lambda^2$. Note also that

$$\mathbb{P}[T_k > 0, 1 \le k \le n] - \lambda = \mathbb{P}[n + 1 \le H_0 < \infty],$$

where $H_0 = \min\{k \ge 1 \colon T_k \le 0\}$ (hitting time of $\{\dots, -1, 0\}$). Therefore,

$$\sum_{n \ge 1} \operatorname{cov}(I_0^*, I_n^*) = \lambda \sum_{n \ge 1} \mathbb{P}[n + 1 \le H_0 < \infty] = \lambda \mathbb{E}[(H_0 - 1)\mathbf{1}_{\{H_0 < \infty\}}].$$

We have $\mathbb{P}[H_0 = 1] = 1 - \rho$ and, from the hitting time theorem (see [15, p. 79]),

$$\mathbb{P}[H_0 = n] = \frac{\rho}{n-1} \mathbb{P}[T_{n-1} = -1], \qquad n \ge 2.$$

Hence,

$$\mathbb{E}[(H_0 - 1)\mathbf{1}_{\{H_0 < \infty\}}] = \rho \sum_{n \ge 1} \mathbb{P}[T_{n-1} = -1] = \rho \sum_{m \ge 0} \binom{2m+1}{m} \rho^m (1-\rho)^{m+1} = \frac{1-\rho}{2\rho - 1}$$

and, thus, $\sigma^2 = 4\rho(1 - \rho)$.

Remark 6. Records in random walks, as considered in Examples 2 and 3, have received much attention in the physics literature in recent years. The unbiased case (zero drift) was analyzed in [19] using a theorem of Sparre Andersen; for this model, assuming the independence of the increments, $\mathbb{E}[N_n] \sim 2\sqrt{n/\pi}$, regardless of the (symmetric and continuous) distribution of the increments.

The biased case, assuming i.i.d. increments, with density symmetric around $c > 0$ was studied in [20]. In that paper, no restriction on the moments of the increments was imposed. In particular, it is shown that $\mathbb{E}[N_n]$ grows as a power of n when the distribution of the increments has no expectation. Also, when the increments have no variance, the distribution of N_n approaches a non-Gaussian limit. Our results do not cover those situations since we need finite expectation of the increments to obtain the linear record rate (Theorem 1) and another moment condition implying the existence of variance, required for the Gaussian limit law of N_n (Theorem 3). A differential feature of our results is that we do not impose the independence of the increments, neither the continuity, or symmetry of their distribution. In that sense, our results in Theorems 1 and 3 reveal a kind of universality principle for random walks with correlated increments and positive drift: under some moment restrictions, the number of records grows linearly and fluctuations are Gaussian when n is large.

Appendix A

To make this paper self-contained we present in this appendix a key result used in our proofs. We also collect technical lemmas related to the proof of Theorem 3.

Proposition 6. (Dubins–Freedman strong law.) *Let $(U_n)_{n\geq 1}$ be a sequence of nonnegative and bounded random variables, adapted to the increasing family of σ-algebras $(\mathcal{G}_n)_{n\geq 0}$. Then*

$$\left\{\sum_{n\geq 1} U_n = \infty\right\} = \left\{\sum_{n\geq 1} \mathbb{E}[U_n \mid \mathcal{G}_{n-1}] = \infty\right\} \quad a.s. \tag{9}$$

and

$$\frac{\sum_{k=1}^{n} U_k}{\sum_{k=1}^{n} \mathbb{E}[U_k \mid \mathcal{G}_{k-1}]} \to 1 \quad on \quad \left\{\sum_{n\geq 1} \mathbb{E}[U_n \mid \mathcal{G}_{n-1}] = \infty\right\} \quad a.s. \tag{10}$$

Proof. See [7] or [17].

Lemma 2. *Let $(\tau_n)_{n\in\mathbb{Z}}$ be the stationary process of Section 2 with $\mathbb{E}[\tau_0] = c \neq 0$. Suppose that $\mathbb{E}[|\tau_0|^{r+a}] < \infty$ and $\sum_{n\geq 1} n^{r/2+1}[\alpha^\tau(n)]^{a/(r+a)} < \infty$ for some $r > 2, a > 0$, where $\alpha^\tau(n)$ is defined in (6). Then*

$$\mathbb{P}\left[\frac{1}{nc}\sum_{j=1}^{n} \tau_j \leq \frac{1}{2}\right] \leq Kn^{-r/2}.$$

Proof. Let $S_n = \sum_{j=1}^{n}(\tau_j - c)$ and $c > 0$. Then, from Markov's inequality and [28, Theorem 1],

$$\mathbb{P}\left[\sum_{j=1}^{n} \tau_j \leq \frac{nc}{2}\right] = \mathbb{P}\left[S_n \leq -\frac{nc}{2}\right] \leq \mathbb{P}\left[|S_n|^r > \left(\frac{nc}{2}\right)^r\right] \leq \frac{\mathbb{E}[|S_n|^r]}{(nc/2)^r} \leq Kn^{-r/2},$$

where $K > 0$ is a constant. The argument for $c < 0$ is identical.

Lemma 3. *Under the hypotheses and with the notation of Theorem 3.*

(i) *There exists a random variable K_n such that $Z_n^* = Z_n^{K_n}$ for all $n \in \mathbb{Z}$, and*

(ii) *$p_k \to p$ and $\gamma_k(n) \to \gamma(n)$ as $k \to \infty$.*

Proof of Lemma 3(i). By stationarity we take $n = 0$. Clearly $Z_0^k \uparrow Z_0^*$ a.s. and the result follows if we show that only finitely many terms $X_{-i} - \sum_{j=-i+1}^{0}\tau_j$, $i \geq 1$, are greater than the first a.s. That is, if $\mathbb{P}[X_{-i} - \sum_{j=-i+1}^{0}\tau_j > X_{-1} - \tau_0,\ i \geq 1,\ \text{i.o.}] = 0$. Using the same argument as in the proof of Proposition 1, this probability is bounded above by $\mathbb{P}[X_{-i-1} - X_{-1} > ic/2,\ i \geq 1,\ \text{i.o.}]$. Observe that

$$\mathbb{P}\left[X_{-i-1} - X_{-1} > \frac{ic}{2}\right] \leq \mathbb{P}\left[|X_{-i-1}| + |X_{-1}| > \frac{ic}{2}\right] \leq 2\mathbb{P}\left[|X_0| > \frac{ic}{4}\right],$$

so $\sum_{i\geq 1}\mathbb{P}[|X_0| > ic/4] < \infty$ because $\mathbb{E}[|X_0|] < \infty$ and the conclusion follows.

Proof of Lemma 3(ii). From the proof of Lemma 3(i), we have $I_0^* = I_0^{K_0}$ a.s. and so $I_0^k \to I_0^*$ as $k \to \infty$, which yields $p_k \to p$ and $\gamma_k(n) \to \gamma(n)$ by the dominated convergence theorem.

Lemma 4. *Under the hypotheses and with the notation of Theorem 3.*

(i) *There exists a summable sequence $\bar{\gamma}(n)$ such that $|\gamma_k(n)| \leq \bar{\gamma}(n)$, and*

(ii) *$\sigma_k^2 \to \sigma^2$ as $k \to \infty$.*

Proof of Lemma 4(i). As in [3, p. 808], if $k \leq a_n := \lfloor n/2 \rfloor$, we have $|\gamma_k(n)| \leq \alpha(a_n)$. When $k > a_n$ we bound $|\gamma_k(n)|$, but ci is replaced by the corresponding random trend. Observe that if $l \leq k$ then $I_n^l \geq I_n^k$ and $I_n^k = I_n^k I_n^l$. So $I_0^k I_n^k = I_0^k I_n^{a_n} - I_0^k I_n^{a_n}(1 - I_n^k)$ and, hence,

$$|\gamma_k(n)| = |\mathbb{E}[I_0^k I_n^k] - p_k^2| \leq |\mathbb{E}[I_0^k I_n^{a_n}] - p_k^2| + \mathbb{E}[I_0^k I_n^{a_n}(1 - I_n^k)]. \tag{11}$$

Let A'_{nk} and B'_{nk} be the first and second summands in the right-hand side of (11), respectively. Then

$$\begin{aligned}
B'_{nk} &\leq \mathbb{P}\bigg[Z_n^{a_n} < X_n \leq \bigvee_{i=a_n+1}^{k} \bigg\{X_{n-i} - \sum_{j=n-i+1}^{n} \tau_j\bigg\}, X_0 > Z_0^k\bigg] \\
&\leq \mathbb{P}\bigg[\bigcup_{i=a_n+1}^{k} \bigg\{X_{n-i} - X_n \geq \sum_{j=n-i+1}^{n} \tau_j\bigg\}\bigg] \\
&\leq \sum_{i>a_n} \mathbb{P}\bigg[X_{n-i} - X_n \geq \sum_{j=n-i+1}^{n} \tau_j\bigg] \\
&\leq \sum_{i>a_n} \mathbb{P}\bigg[|X_{n-i}| + |X_n| \geq \sum_{j=n-i+1}^{n} \tau_j\bigg] \\
&\leq \sum_{i>a_n} \mathbb{P}\bigg[|X_0| \geq \frac{1}{2}\sum_{j=1}^{i} \tau_j\bigg] + \sum_{i>a_n} \mathbb{P}\bigg[|X_0| \geq \frac{1}{2}\sum_{j=-i+1}^{0} \tau_j\bigg].
\end{aligned} \tag{12}$$

Observe that both probabilities above can be bounded by $\mathbb{P}[|X_0| \geq ci/4] + \mathbb{P}[\sum_{j=1}^{i} \tau_j \leq ci/2]$ and that $\sum_{i>a_n} \mathbb{P}[|X_0| \geq ci/4]$ is well defined and summable (with respect to n) because X_0 is square-integrable. Also, by Lemma 2, $\mathbb{P}[\sum_{j=1}^{i} \tau_j \leq ci/2] \leq K i^{-r/2}$ for some constant K. So, B'_{nk} is summable because, from the inequalities above,

$$B'_{nk} \leq B'_n := 2\sum_{i>a_n} \bigg(\mathbb{P}\bigg[|X_0| \geq \frac{ci}{4}\bigg] + K i^{-r/2}\bigg).$$

On the other hand,

$$\begin{aligned}
A'_{nk} &\leq |\mathbb{E}[I_0^k I_n^{a_n}] - p_{a_n} p_k| + p_k(p_{a_n} - p_k) \\
&\leq \alpha(a_n) + p_{a_n} - p_k \\
&\leq \alpha(a_n) + \mathbb{P}\bigg[Z_n^{a_n} < X_n \leq \bigvee_{i=a_n+1}^{k} \bigg\{X_{n-i} - \sum_{j=n-i+1}^{n} \tau_j\bigg\}\bigg]
\end{aligned} \tag{13}$$

and we see that the probability in (13) is bounded as in (12); hence, $A'_{nk} \leq A'_n := \alpha(a_n) + B'_n$.

Proof of Lemma 4(ii). By Lemma 3(ii), $\gamma_k(n) \to \gamma(n)$ and, by Lemma 4(i), $\gamma_k(n)$ is dominated by $\bar{\gamma}(n) := A'_n + B'_n$, which is summable and, consequently, the result follows from the dominated convergence theorem applied to $\sum_{n\geq 1} \gamma_k(n)$.

Lemma 5. *Under the hypotheses and with the notation of Theorem 3,*

$$\limsup_{n\to\infty} \mathbb{P}[|(N_n^k - np_k) - (N_n^* - np)| > \varepsilon\sqrt{n}] \to 0 \quad \textit{for all } \varepsilon > 0 \textit{ as } k \to \infty. \tag{14}$$

https://doi.org/10.1239/aap/1449859805 Published online by Cambridge University Press

Proof. We estimate the variance

$$\operatorname{var}\left(\frac{N_n^k - N_n^*}{\sqrt{n}}\right) = \frac{1}{n}\operatorname{var}\left(\sum_{l=1}^{n}(J_l^k - q_k)\right) = \operatorname{var}(J_0^k) + \frac{2}{n}\sum_{l=1}^{n-1}(n-l)\operatorname{cov}(J_0^k, J_l^k), \quad (15)$$

where $J_l^k = I_l^k - I_l^* = \mathbf{1}_{\{Z_l^k < X_l \le Z_l^*\}} = \mathbf{1}_{\{Z_l^k < X_l \le \bigvee_{i=k+1}^{\infty}\{X_{l-i} - \sum_{j=l-i+1}^{l}\tau_j\}\}}$ and $q_k = p_k - p = \mathbb{P}[Z_0^k < X_0 \le \bigvee_{i>k}\{X_{-i} - \sum_{j=-i+1}^{0}\tau_j\}]$. For the variance term, we have, by Lemma 3(ii),

$$\operatorname{var}(J_0^k) = q_k(1-q_k) \le q_k \to 0. \quad (16)$$

For the covariances assume that $n > 2k$, then as $k \to \infty$,

$$\begin{aligned}
\frac{2}{n}\sum_{l=1}^{2k}(n-l)\operatorname{cov}(J_0^k, J_l^k) &\le \frac{2}{n}\sum_{l=1}^{2k}(n-l)\operatorname{var}(J_0^k) \\
&\le 4kq_k \\
&\le 4k\mathbb{P}\left[\bigcup_{i>k}\left\{X_{-i} - X_0 \ge \sum_{j=-i+1}^{0}\tau_j\right\}\right] \\
&\le 4k\sum_{i>k}\left(\mathbb{P}\left[|X_{-i}| \ge \frac{1}{2}\sum_{j=-i+1}^{0}\tau_j\right] + \mathbb{P}\left[|X_0| \ge \frac{1}{2}\sum_{j=-i+1}^{0}\tau_j\right]\right) \\
&\le 8k\sum_{i>k}\left(\mathbb{P}\left[|X_0| \ge \frac{ci}{4}\right] + \mathbb{P}\left[\sum_{j=1}^{i}\tau_j \le \frac{ci}{2}\right]\right) \\
&\to 0. \qquad (17)
\end{aligned}$$

We bound $\operatorname{cov}(J_0^k, J_l^k)$ for $l > 2k$. Let $L_l^k = \mathbf{1}_{\{X_l > \bigvee_{i=k+1}^{\lfloor l/2\rfloor}\{X_{l-i} - \sum_{j=l-i+1}^{l}\tau_j\}\}}$ then

$$\operatorname{cov}(J_0^k, J_l^k) = \operatorname{cov}(J_0^k, J_l^k(1-L_l^k)) + \operatorname{cov}(J_0^k, J_l^k L_l^k) \quad (18)$$

and

$$\begin{aligned}
|\operatorname{cov}(J_0^k, J_l^k L_l^k)| &\le \mathbb{E}[J_l^k L_l^k] \\
&\le \mathbb{P}\left[Z_l^k < X_l \le \bigvee_{i>\lfloor l/2\rfloor}\left\{X_{l-i} - \sum_{j=l-i+1}^{l}\tau_j\right\}\right] \\
&\le \mathbb{P}\left[X_l \le \bigvee_{i>\lfloor l/2\rfloor}\left\{X_{l-i} - \sum_{j=l-i+1}^{l}\tau_j\right\}\right] \\
&\le \sum_{i>\lfloor l/2\rfloor}\mathbb{P}\left[X_{l-i} - X_l \ge \sum_{j=l-i+1}^{l}\tau_j\right] \\
&\le 2\sum_{i>\lfloor l/2\rfloor}\left(\mathbb{P}\left[|X_0| \ge \frac{ci}{4}\right] + \mathbb{P}\left[\sum_{j=1}^{i}\tau_j \le \frac{ci}{2}\right]\right).
\end{aligned}$$

https://doi.org/10.1239/aap/1449859805 Published online by Cambridge University Press

Hence,

$$\left|\frac{2}{n}\sum_{l=2k+1}^{n-1}(n-l)\operatorname{cov}(J_0^k, J_l^k L_l^k)\right| \le 4\sum_{l>2k}\sum_{i>\lfloor l/2\rfloor}\left(\mathbb{P}\left[|X_0| \ge \frac{ci}{4}\right] + \mathbb{P}\left[\sum_{j=1}^{i}\tau_j \le \frac{ci}{2}\right]\right)$$
$$\to 0 \quad \text{as } k \to \infty. \tag{19}$$

For the first term in the right-hand side of (18) it suffices to see that its absolute value is bounded above by $\alpha(\lfloor l/2\rfloor)$. Indeed, $J_l^k(1-L_l^k)$ is the indicator of

$$\left\{Z_l^k < X_l \le Z_l^*,\, X_l \le \bigvee_{i=k+1}^{\lfloor l/2\rfloor}\left\{X_{l-i} - \sum_{j=l-i+1}^{l}\tau_j\right\}\right\}$$
$$= \left\{Z_l^k < X_l \le \bigvee_{i=k+1}^{\lfloor l/2\rfloor}\left\{X_{l-i} - \sum_{j=l-i+1}^{l}\tau_j\right\}\right\} \in \mathcal{F}_{l-\lfloor l/2\rfloor}^{\infty}.$$

Therefore,

$$\left|\frac{2}{n}\sum_{l=2k+1}^{n-1}(n-l)\operatorname{cov}(J_0^k, J_l^k(1-L_l^k))\right| \le 2\sum_{l>2k}\alpha\left(\left\lfloor\frac{l}{2}\right\rfloor\right) \to 0 \quad \text{as } k\to\infty. \tag{20}$$

The conclusion follows from (16), (17), (19), (20), and Tchebychev's inequality applied to (14).

Acknowledgements

The authors are grateful to the editor and an anonymous referee for their useful comments. Support from project PFB-03-CMM, Fondecyt grant 1120408, and project MTM2010-15972 of MINECO is gratefully acknowledged. The authors are members of the research group Modelos Estocásticos, supported by DGA and the European Social Fund.

References

[1] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*. John Wiley, New York.
[2] Ballerini, R. and Resnick, S. (1985). Records from improving populations. *J. Appl. Prob.* **22,** 487–502.
[3] Ballerini, R. and Resnick, S. I. (1987). Records in the presence of a linear trend. *Adv. Appl. Prob.* **19,** 801–828.
[4] Billingsley P. (1968). *Convergence of Probability Measures.* John Wiley, New York.
[5] Borovkov, K. (1999). On records and related processes for sequences with trends. *J. Appl. Prob.* **36,** 668–681.
[6] De Haan, L. and Verkade, E. (1987). On extreme-value theory in the presence of a trend. *J. Appl. Prob.* **24,** 62–76.
[7] Dubins, L. E. and Freedman, D. A. (1965). A sharper form of the Borel–Cantelli lemma and the strong law. *Ann. Math. Statist.* **36,** 800–807.
[8] Feuerverger, A. and Hall, P. (1996). On distribution-free inference for record-value data with trend. *Ann. Statist.* **24,** 2655–2678.
[9] Foster, F. G. and Stuart, A. (1954). Distribution-free tests in time-series based on the breaking of records, with discussion. *J. R. Statist. Soc.* B **16,** 1–22.
[10] Franke J., Wergen G. and Krug J. (2010). Records and sequences of records from random variables with a linear trend. *J. Statist. Mech. Theory Exp.* **2010,** P10013.
[11] Glasserman, P. and Yao, D. D. (1995). Stochastic vector difference equations with stationary coefficients. *J. Appl. Prob.* **32,** 851–866.
[12] Gouet, R., López, F. J. and Sanz, G. (2007). Asymptotic normality for the counting process of weak records and δ-records in discrete models. *Bernoulli* **13,** 754–781.
[13] Gouet, R., López, F. J. and Sanz, G. (2008). Laws of large numbers for the number of weak records. *Statist. Prob. Lett.* **78,** 2010–2017.

[14] GOUET, R., LÓPEZ, F. J. AND SANZ, G. (2011). Asymptotic normality for the number of records from general distributions. *Adv. Appl. Prob.* **43,** 422–436.
[15] GRIMMETT, G. AND STIRZAKER, D. R. (2001). *Probability and Random Processes*, 3rd edn. Oxford University Press.
[16] GUT, A. (2009). *Stopped Random Walks. Limit Theorems and Applications*, 2nd edn. Springer, New York.
[17] HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.
[18] IBRAGIMOV, I. A. AND LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.
[19] MAJUMDAR, S. N. AND ZIFF, R. M. (2008). Universal record statistics of random walks and Lévy flights. *Phys. Rev. Lett.* **101,** 050601.
[20] MAJUMDAR, S. N., SCHEHR, G. AND WERGEN, G. (2012). Record statistics and persistence for a random walk with a drift. *J. Phys. A* **45,** 1–42, 355002.
[21] NEVZOROV, V. B. (2001). *Records: Mathematical Theory.* (Trans. Math. Monogr. **194**). American Mathematical Society, Providence, RI.
[22] SPITZER, F. (1964). *Principles of Random Walk.* Van Nostrand, Princeton, NJ.
[23] WERGEN, G. AND KRUG, J. (2010). Record-breaking temperatures reveal a warming climate. *Europhys. Lett.* **92,** 30008.
[24] WERGEN, G., BOGNER, M. AND KRUG, J. (2011). Record statistics for biased random walks, with an application to financial data. *Phys. Rev. E* **83,** 051109.
[25] WERGEN, G., FRANKE, J. AND KRUG, J. (2011). Correlations between record events in sequences of random variables with a linear trend. *J. Statist. Phys.* **144,** 1206–1222. (Erratum: **145** (2011), 1405–1406.)
[26] WERGEN, G., HENSE A. AND KRUG J. (2014). Record occurrence and record values in daily and monthly temperatures. *Climate Dynamics* **42,** 1275–1289.
[27] YANG, M. C. K. (1975). On the distribution of the inter-record times in an increasing population. *J. Appl. Prob.* **12,** 148–154.
[28] YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrscheinlichkeitsth.* **52,** 45–57.