

CENTRAL LIMIT THEOREMS FOR THE NUMBER OF RECORDS IN DISCRETE MODELS

RAUL GOUET,* *Universidad de Chile*

F. JAVIER LÓPEZ** *** AND

GERARDO SANZ,** **** *Universidad de Zaragoza*

Abstract

Consider a sequence (X_n) of independent and identically distributed random variables taking nonnegative integer values, and call X_n a record if $X_n > \max\{X_1, \dots, X_{n-1}\}$. By means of martingale arguments it is shown that the counting process of records among the first n observations, suitably centered and scaled, is asymptotically normally distributed.

Keywords: Extreme; martingale; record; central limit theorem

2000 Mathematics Subject Classification: Primary 60G70

Secondary 60G42

1. Introduction

Let (X_n) be a sequence of nonnegative, independent, and identically distributed (nonnegative i.i.d.) random variables (RVs) with common distribution function F , and let $M_n = \max\{X_1, \dots, X_n\}$, for $n \geq 1$, be the sequence of partial maxima; conventionally we write $M_0 = -1$. We say that X_n is a (strict upper) record if $X_n > M_{n-1}$. The indicator of a record is denoted by $I_n = \mathbf{1}_{\{X_n > M_{n-1}\}}$ and the associated counting process by $N_n = \sum_{k=1}^n I_k$. See [1], [2], and [9] for references and information on the theory and applications of records.

We are interested here in the asymptotic normality of N_n , suitably centered and scaled, when the underlying distribution F is concentrated on the nonnegative integers.

A beautiful and well-known result of Rényi [12] states that the indicators I_n are independent, with $P[I_n = 1] = 1/n$, when F is continuous. Therefore, the central limit theorem (CLT)

$$\frac{N_n - \log n}{\sqrt{\log n}} \xrightarrow{w} N(0, 1),$$

is readily obtained. (Here, ' \xrightarrow{w} ' denotes weak convergence.)

When F is discontinuous, the indicators I_n are not independent and their distributions depend on F . Therefore, this case is somewhat more complicated and results are rather scarce. Vervaat [13], in a remarkable pioneering paper, obtained a variety of functional CLTs for records of nonnegative, integer-valued random variables. In particular, his work contains the asymptotic normality of N_n for the geometric distribution, which happens to be an especially motivating

Received 29 September 2003; revision received 24 December 2004.

* Postal address: Departamento de Ingeniería Matemática y Centro de Modelamiento Matemático, Universidad de Chile, UMI-CNRS-2807, Casilla 170-3, Correo 3, Santiago, Chile. Email address: rgouet@dim.uchile.cl

** Postal address: Departamento de Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain.

*** Email address: javierl@unizar.es

**** Email address: gerardo@unizar.es

distribution, given its connection with the asymptotic analysis of search costs in data structures known as *skip lists* (see [11]). In the context of the geometric distribution, Prodingler [10] obtained asymptotic expressions for the mean and variance of N_n , and Bai *et al.* [3] established asymptotic normality with rates. Notice that computer science specialists use the term *left-to-right maxima* when referring to upper records.

A martingale approach, combined with asymptotic results for sums of partial minima, was used in [7] to derive strong convergence results for N_n from quite general discrete models, including the geometric and Poisson random variables. Our aim here is to further exploit the connection between records and martingales, obtaining the corresponding CLTs.

In this paper, we establish CLTs for a wide range of discrete distributions, identifying the centering and scaling sequences (Theorem 1(a)). We also show that, for a certain class of distributions, no such CLT holds (Theorem 1(b)). While, to the best of the authors' knowledge, only the case of the geometric distribution has previously appeared in the literature, the results of this paper cover a rather general class of discrete distributions; see Remark 1.

We conclude this introduction with additional definitions and notation. Let (X_n) denote a sequence of i.i.d. RVs, with common right-continuous distribution function F , such that $P[X_n = k] = p_k > 0$ for $k \in \mathbb{Z}_+ = \{0, 1, \dots\}$ and $n \geq 1$. Clearly, $F(x) < 1$ for all x and, hence, $N_n \nearrow \infty$ almost surely.

For $k \in \mathbb{Z}_+$ and $t \geq 0$, let $y_k = 1 - F(k) = \sum_{i>k} p_i$ be the discrete survival function (with $y_{-1} = 1$) and let $m(t) = \min\{j \in \mathbb{Z}_+ : y_j < 1/t\}$ be the quantile function. The discrete failure or hazard rate r_k is defined as

$$r_k = \frac{P[X_1 = k]}{P[X_1 \geq k]} = \frac{p_k}{y_{k-1}}$$

and is easily seen to satisfy

$$r_k = 1 - \frac{y_k}{y_{k-1}} \quad \text{and} \quad y_k = \prod_{i=0}^k (1 - r_i).$$

Let $\theta(k) = \sum_{i=0}^k r_i$ be the cumulative hazard function and let $\Theta(t) = \max\{k : \theta(k) \leq t\}$ for $t \geq p_0$. Then $\theta(\Theta(t)) \leq t < \theta(\Theta(t) + 1)$ and $P[\theta(X_n) > t] = P[X_n > \Theta(t)] = y_{\Theta(t)}$, for $n \geq 1$ and $t \geq p_0$.

Martingales are taken relative to the *natural* filtration (\mathcal{F}_n) , with $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ for $n \geq 1$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Almost-sure convergence, convergence in probability, and weak convergence will be denoted respectively by the arrows ' $\xrightarrow{\text{a.s.}}$ ', ' $\xrightarrow{\text{P}}$ ', and (recall) ' $\xrightarrow{\text{w}}$ '. For increasing sequences (a_n) and (b_n) , we write $a_n \sim b_n$ if either both (a_n) and (b_n) converge to a finite limit or they both grow to infinity with $\lim a_n/b_n = 1$.

A word about the organization of the paper. The main result and several examples are presented in Section 2. The proof, developed in Section 3, is rather long and has been split into several lemmas and propositions, distributed in subsections. Section 4 contains the proofs of particular cases, shown as examples in Section 2. Finally, in Section 5, we present our concluding remarks and some ideas for further research.

2. Main result and examples

Our main result is the asymptotic normality of the counting process of records N_n , suitably centered and scaled, which is applicable to a wide spectrum of discrete models. We use a

martingale approach that connects the CLT with convergence results from the theory of sums of partial minima of i.i.d. RVs, as developed by Deheuvels in [4].

Theorem 1. Let $z_k = \sum_{i>k} r_i y_i$ and $b_n^2 = \sum_{k=0}^{m(n)} z_k r_k / y_k$, for $k, n \in \mathbb{Z}_+$.

(a) Assume that $\sum_{k=0}^\infty (1 - r_k) = \infty$. If $\limsup_{k \rightarrow \infty} r_k < 1$ or $\liminf_{k \rightarrow \infty} r_k > 0$ then

$$\frac{N_n - \theta(m(n))}{b_n} \xrightarrow{w} N(0, 1). \tag{1}$$

(b) If $\sum_{k=0}^\infty (1 - r_k) < \infty$ then $N_n - m(n)$ is tight. In particular, there are no sequences $(a_n), (b_n) \nearrow \infty$ such that $(N_n - a_n)/b_n$ converges in distribution to a nondegenerate random variable.

Remark 1. Theorem 1 gives a rather complete picture of the asymptotic normality of the number of records for discrete distributions. In fact, any sequence $(r_k), 0 < r_k < 1, k \geq 0$, with $\sum_{k=0}^\infty r_k = \infty$ is the failure rate sequence of a distribution on the nonnegative integers. Only the very special case of distributions whose failure rates (r_k) satisfy both $\liminf_{k \rightarrow \infty} r_k = 0$ and $\limsup_{k \rightarrow \infty} r_k = 1$ is not covered by Theorem 1.

Before going into the proof of Theorem 1, we present some examples whose details are worked out in Section 4.

Example 1. (Geometric with parameter p .)

$$(\log n)^{-1/2} \left(N_n + \frac{p \log n}{\log(1 - p)} \right) \xrightarrow{w} N \left(0, -\frac{p(1 - p)}{\log(1 - p)} \right). \tag{2}$$

The convergence in (2) was previously obtained in [13] and [3] using completely different methods. To the best of the authors' knowledge, the cases covered by the next examples are new.

Example 2. (Converging failure rates $r_k \rightarrow r, 0 < r < 1$, with $\sum_{i=1}^n |r_i - r|/\sqrt{n} \rightarrow 0$.)

$$(\log n)^{-1/2} \left(N_n + \frac{r \log n}{\log(1 - r)} \right) \xrightarrow{w} N \left(0, -\frac{r(1 - r)}{\log(1 - r)} \right). \tag{3}$$

A concrete random variable with converging r_k s satisfying the hypothesis of this example is the negative binomial, with $p_k = (-1)^k \binom{-a}{k} p^a (1 - p)^k$ for $k \geq 0, 0 < p < 1$, and $a > 1$. In this case, (3) holds with $r = p$.

Example 3. (Alternating geometric with parameters p and q .) Here, we mean $r_{2k} = p$ and $r_{2k+1} = q$, where $0 < p < q < 1$ and $k \geq 0$. This random variable can be seen as the number of tails in independent throws of alternating coins, with respective head probabilities p and q , until the first head appears. In this case,

$$(\log n)^{-1/2} \left(N_n + \frac{(p + q) \log n}{\log(1 - p)(1 - q)} \right) \xrightarrow{w} N \left(0, -\frac{p(1 - p) + q(1 - q)}{\log(1 - p)(1 - q)} \right). \tag{4}$$

Example 4. (Converging failure rates $r_k \rightarrow 0$, with $\sum_{k=1}^\infty r_k^2 < \infty$.)

$$(\log n)^{-1/2} (N_n - \log n) \xrightarrow{w} N(0, 1). \tag{5}$$

For a concrete RV X , consider $y_k = (k + 1)^{-d}, k \geq 0, d > 0$. Then $r_k = d/(k + 1) + O(k^{-2})$ and (5) applies.

Example 5. (Converging failure rates $r_k \rightarrow 1$, such that $\sum_{k=1}^\infty (1 - r_k) = \infty$.) Let $r_k = 1 - ak^{-\alpha} + \delta_k$, for $k \geq 1$, $a > 0$, and $0 < \alpha \leq 1$, with (δ_k) such that $\sum_{k=1}^\infty |\delta_k| < \infty$. Then

$$(\log m(n))^{-1/2}(N_n - m(n) + a \log m(n)) \xrightarrow{w} N(0, a) \tag{6}$$

for $\alpha = 1$ and

$$(m(n))^{-(1-\alpha)/2} \left(N_n - m(n) + \frac{a}{1-\alpha} (m(n))^{1-\alpha} \right) \xrightarrow{w} N\left(0, \frac{a}{1-\alpha}\right) \tag{7}$$

for $\alpha < 1$. Also, $m(n) \sim \log n / \alpha \log \log n$.

In the particular case of the Poisson distribution with parameter λ , we have $a = \lambda$ and $\alpha = 1$. From (6) we obtain

$$(\log \log n)^{-1/2}(N_n - m(n) + \lambda \log(m(n))) \xrightarrow{w} N(0, \lambda),$$

with $m(n) \sim \log n / \log \log n$.

Remark 2. Notice the differences between continuous and discrete distributions. For continuous distributions, the number of records is always asymptotically normal, with the variance growing as $\log n$ regardless of the parent distribution F . For discrete distributions, the asymptotic normality of the number of records depends on the distribution F via the failure rates r_k : for distributions with very light tails (those with $\sum_{k=0}^\infty (1 - r_k) < \infty$), the number of records is not asymptotically normal; moreover, when a CLT holds, the scaling sequence grows at a speed that depends on the r_k s.

3. Proof of main result

3.1. Sums of partial minima

As stated in the introduction, the martingale approach we use depends on asymptotic results for sums of partial minima of i.i.d. RVs. The behavior of sums of minima is fairly well known and the following weak law of large numbers [4] is particularly useful for our purposes.

Let (Z_n) be a sequence of nonnegative i.i.d. RVs, with common distribution function G such that $G(z) > 0$ for all $z > 0$, and let $S_n = \sum_{i=1}^n \min\{Z_1, \dots, Z_i\}$. In addition, let $\tilde{G}(t) = \inf\{z \geq 0: G(z) \geq t\}$ for $0 \leq t < 1$, and let

$$H(x) = \int_1^{e^x} \tilde{G}(1/u) du \quad \text{for } x \geq 0.$$

Theorem 2. If $\lim_{x \rightarrow \infty} H(x)$ is finite then S_n grows almost surely to a finite limit. Otherwise, if there is a sequence (x_n) , increasing to ∞ , such that

$$\frac{H(x_n + \log n)}{H(\log n)} \rightarrow 1 \tag{8}$$

and

$$\frac{\sum_{k=1}^n k \tilde{G}(1/k)^2}{(\sum_{k=1}^n \tilde{G}(1/k))^2} \rightarrow 0, \tag{9}$$

then

$$\frac{S_n}{H(\log n)} \xrightarrow{p} 1.$$

For a proof of Theorem 2, see [4, Chapter II, Theorem 7 and Corollary 4].

3.2. A martingale CLT

Martingale CLTs generally require the convergence in probability of a quadratic process and a Lindeberg-type condition of asymptotic negligibility. Here we have chosen to work with the following version, whose conditions are relatively simple to verify; see [8, p. 58].

Theorem 3. Consider a sequence (ξ_i) such that $E[\xi_i | \mathcal{F}_{i-1}] = 0$ and $E[\xi_i^2] < \infty$ for all $i \geq 1$. If, for a sequence $(b_n) \nearrow \infty$, we have

$$\frac{1}{b_n^2} \sum_{i=1}^n E[\xi_i^2 | \mathcal{F}_{i-1}] \xrightarrow{p} 1 \tag{10}$$

and

$$\frac{1}{b_n^2} \sum_{i=1}^n E[\xi_i^2 \mathbf{1}_{\{|\xi_i| > \varepsilon b_n\}} | \mathcal{F}_{i-1}] \xrightarrow{p} 0 \tag{11}$$

for all $\varepsilon > 0$, then $\sum_{i=1}^n \xi_i/b_n \xrightarrow{w} N(0, 1)$.

As we will see later, it turns out that the processes involved in conditions (10) and (11) can be written in terms of sums of partial minima of RVs.

3.3. The fundamental martingale

The CLT for records of various discrete models is based on a single fundamental martingale, presented below. The idea comes from the easily verifiable fact that $N_n - pM_n$ is a martingale when the underlying RVs are geometric with parameter p .

Proposition 1. (a) The process

$$N_n - \theta(M_n) = N_n - \sum_{k=0}^{M_n} r_k, \tag{12}$$

for $n \geq 1$, is a square-integrable martingale.

(b) Let $\xi_k = I_k - [\theta(M_k) - \theta(M_{k-1})]$ be the increments of martingale (12) and let $E[\xi_k^2 | \mathcal{F}_{k-1}]$ be the increments of the processes of conditional variances in (10). Then

$$E[\xi_k^2 | \mathcal{F}_{k-1}] = \sum_{i > M_{k-1}} p_i(1 - r_i) = \sum_{i > M_{k-1}} r_i y_i.$$

Proof. (a) First, it is easy to see that

$$E[I_k | \mathcal{F}_{k-1}] = P[X_k > M_{k-1} | \mathcal{F}_{k-1}] = 1 - F(M_{k-1}) = y_{M_{k-1}}.$$

Let $\Delta\theta(M_k) = \theta(M_k) - \theta(M_{k-1})$. Then

$$\begin{aligned} E[\Delta\theta(M_k) | \mathcal{F}_{k-1}] &= E\left[\left(\sum_{i=0}^{M_k} r_i - \sum_{i=0}^{M_{k-1}} r_i\right) \middle| \mathcal{F}_{k-1}\right] \\ &= \sum_{i=1}^{\infty} \left(\sum_{j=1}^i r_{M_{k-1}+j}\right) P[X_k = M_{k-1} + i | \mathcal{F}_{k-1}] \\ &= \sum_{j=1}^{\infty} r_{M_{k-1}+j} P[X_k > M_{k-1} + j - 1 | \mathcal{F}_{k-1}] = \sum_{j=1}^{\infty} p_{M_{k-1}+j} = y_{M_{k-1}}. \end{aligned}$$

Therefore, $N_n - \theta(M_n)$ is a martingale. To verify the square integrability of (12), it suffices to check that $\theta(X_n)$ is indeed square integrable:

$$\begin{aligned} E[\theta(X_n)^2] &= \sum_{k=0}^{\infty} \left(\sum_{i=0}^k r_i \right)^2 p_k = \sum_{k=0}^{\infty} \left(\sum_{i=0}^k r_i^2 + 2 \sum_{i=0}^{k-1} \sum_{j=i+1}^k r_i r_j \right) p_k \\ &= \sum_{i=0}^{\infty} r_i^2 \sum_{k=i}^{\infty} p_k + 2 \sum_{i=0}^{\infty} r_i \sum_{j=i+1}^{\infty} r_j \sum_{k=j}^{\infty} p_k \\ &\leq \sum_{i=0}^{\infty} r_i y_{i-1} + 2 \sum_{i=0}^{\infty} r_i y_i \\ &= 1 + 2 \sum_{i=0}^{\infty} p_i (1 - r_i) \leq 3. \end{aligned}$$

(b) From the martingale property, and noting that $I_k \Delta\theta(M_k) = \Delta\theta(M_k)$, we have

$$\begin{aligned} E[\xi_k^2 \mid \mathcal{F}_{k-1}] &= E[I_k \mid \mathcal{F}_{k-1}] - 2 E[\Delta\theta(M_k) \mid \mathcal{F}_{k-1}] + E[(\Delta\theta(M_k))^2 \mid \mathcal{F}_{k-1}] \\ &= E[(\Delta\theta(M_k))^2 \mid \mathcal{F}_{k-1}] - y_{M_{k-1}}. \end{aligned}$$

On the other hand, writing m for M_{k-1} , we have

$$\begin{aligned} E[(\Delta\theta(M_k))^2 \mid \mathcal{F}_{k-1}] &= E[((\theta(X_k) - \theta(M_{k-1}))^+)^2 \mid \mathcal{F}_{k-1}] \\ &= \sum_{i=1}^{\infty} \left(\sum_{j=1}^i r_{m+j} \right)^2 P[X_k = m + i \mid \mathcal{F}_{k-1}] \\ &= \sum_{i=1}^{\infty} \left(\sum_{j=1}^i r_{m+j}^2 + 2 \sum_{1 \leq j_1 < j_2 \leq i} r_{m+j_1} r_{m+j_2} \right) p_{m+i} \\ &= \sum_{j=1}^{\infty} r_{m+j}^2 \sum_{i=j}^{\infty} p_{m+i} + 2 \sum_{1 \leq j_1 < j_2 < \infty} r_{m+j_1} r_{m+j_2} \sum_{i=j_2}^{\infty} p_{m+i} \\ &= \sum_{j>m} r_j p_j + 2 \sum_{1 \leq j_1 < j_2 < \infty} r_{m+j_1} p_{m+j_2} \\ &= \sum_{j>m} r_j p_j + 2 \sum_{j>m} r_j y_j \\ &= y_m + \sum_{j>m} r_j y_j, \end{aligned}$$

and the conclusion follows.

Remark 3. The martingale (12) reveals an interesting and apparently new property of discrete extremes that could be useful beyond our result on asymptotic normality for records. On the other hand, when F is continuous, it can be easily checked that the martingale analogous to (12) is given by $N_n + \log(1 - F(M_n))$.

Remark 4. Another martingale related to N_n is readily constructed by subtracting the predictable process $\sum_{k=1}^n y_{M_{k-1}}$ from N_n . This process is useful in characterizing the strong asymptotic behavior of N_n (see [7]), but is unsuitable for weak limits.

3.4. Asymptotic conditional variance

It is important to notice that the process of conditional variances in (10) behaves as a sum of partial minima of i.i.d. RVs. This is because $u(M) = \sum_{i>M} r_i y_i$ is a decreasing function of M and, therefore,

$$E[\xi_k^2 \mid \mathcal{F}_{k-1}] = u(M_{k-1}) = \min\{u(X_1), \dots, u(X_{k-1})\}$$

for $k \geq 2$. In what follows, we consider (10) and apply Theorem 2 to obtain the scaling sequence (b_n) of $N_n - \theta(M_n)$.

From Proposition 1, we have

$$\sum_{k=2}^n E[\xi_k^2 \mid \mathcal{F}_{k-1}] = \sum_{k=2}^n \min\{Z_1, \dots, Z_{k-1}\} = \sum_{k=2}^n z_{M_{k-1}},$$

where

$$Z_k = \sum_{i>X_k} r_i y_i = \sum_{i>X_k} p_i (1 - r_i), \quad k \geq 1.$$

These random variables are i.i.d., take the values

$$z_j = \sum_{i>j} r_i y_i = \sum_{i>j} p_i (1 - r_i)$$

with respective probabilities p_j , and have common distribution function G given by

$$G(z) = \sum_{i \geq j} p_i = y_{j-1}, \quad z_j \leq z < z_{j-1}.$$

The inverse of G is easily seen to be

$$\tilde{G}(t) = z_j, \quad y_j < t \leq y_{j-1}.$$

Equivalently, $\tilde{G}(1/t) = z_{m(t)}$, where $m(t) = \min\{j \in \mathbb{Z}_+ : y_j < 1/t\}$, for $t > 1$.

Proposition 2. *Let (Z_n) be the sequence of i.i.d. RVs defined above and let*

$$b_n^2 = \sum_{k=0}^{m(n)} \frac{z_k r_k}{y_k}. \tag{13}$$

(a) *Assume that $\sum_{k=0}^\infty (1 - r_k) = \infty$. If $\limsup_{k \rightarrow \infty} r_k < 1$ or $\liminf_{k \rightarrow \infty} r_k > 0$ then*

$$\frac{1}{b_n^2} \sum_{k=1}^n \min\{Z_1, \dots, Z_k\} \xrightarrow{P} 1.$$

(b) *If $\sum_{k=0}^\infty (1 - r_k) < \infty$ then*

$$\sum_{k=1}^n \min\{Z_1, \dots, Z_k\} \xrightarrow{\text{a.s.}} Z,$$

where Z is a finite random variable.

Proof. We first determine the normalizing sequence $H(\log n)$. Let $t > 1$. Then

$$\begin{aligned}
 H(\log t) &= \int_1^t \tilde{G}(1/u) \, du = \int_1^t z_{m(u)} \, du \\
 &= \sum_{j=0}^{m(t)} \int_{y_{j-1}^{-1}}^{y_j^{-1}} z_{m(u)} \, du - \int_t^{y_{m(t)}^{-1}} z_{m(u)} \, du \\
 &= \sum_{j=0}^{m(t)} z_j(y_j^{-1} - y_{j-1}^{-1}) - z_{m(t)}(y_{m(t)}^{-1} - t) \\
 &= \sum_{j=0}^{m(t)} \frac{z_j r_j}{y_j} - \rho(t),
 \end{aligned} \tag{14}$$

where $\rho(t) = z_{m(t)}(y_{m(t)}^{-1} - t)$. Clearly $z_j \leq y_j$; therefore,

$$\rho(t) \leq z_{m(t)}(y_{m(t)}^{-1} - y_{m(t)-1}^{-1}) = z_{m(t)}r_{m(t)}/y_{m(t)} \leq 1$$

and we have $H(\log n) \sim b_n^2$.

In order to prove part (a), we check hypotheses (8) and (9) of Theorem 2. Suppose first that $\limsup_{k \rightarrow \infty} r_k < 1$ and note that $z_k = \tilde{G}(1/t)$ for $y_k < 1/t \leq y_{k-1}$; that is, $m(t) = k$. Then

$$\frac{z_{m(t)}}{y_{m(t)-1}} \leq t\tilde{G}(1/t) < \frac{z_{m(t)}}{y_{m(t)}}.$$

On the other hand, it is easy to see that $z_k/y_{k-1} \geq \inf_{j \geq k} (1 - r_j)^2$. Hence, we have

$$\liminf_{t \rightarrow \infty} t\tilde{G}(1/t) > 0$$

and there exist constants A and B , $0 < A \leq B \leq 1$, such that

$$A/t \leq \tilde{G}(1/t) \leq B/t \tag{15}$$

for all $t > 1$. Now, from the definition of H and (15) we find that

$$0 \leq \frac{H(x_n + \log n) - H(\log n)}{H(\log n)} \leq \frac{Bx_n}{A \log n}$$

for $n \geq 1$, and (8) follows upon choosing, for instance, $x_n = \log(\log n + 3)$. Equation (9) also follows immediately from (15), since

$$\frac{\sum_{k=1}^n k(\tilde{G}(1/k))^2}{(\sum_{k=1}^n \tilde{G}(1/k))^2} \leq \frac{B^2 \sum_{k=1}^n 1/k}{A^2(\sum_{k=1}^n 1/k)^2} \rightarrow 0.$$

Now take $\liminf_{k \rightarrow \infty} r_k > 0$. In this case, there exists a δ , $0 < \delta < 1$, such that $1 - r_k < \delta$ for all $k \geq 0$. Noting that $y_k = p_{k+1}/r_{k+1}$, we have $z_k = \sum_{i>k} y_i r_i = \sum_{i>k} p_{i+1} r_i / r_{i+1}$. Then, since

$$(1 - \delta)y_{k+1} = (1 - \delta) \sum_{i>k} p_{i+1} < \sum_{i>k} \frac{p_{i+1} r_i}{r_{i+1}} < \frac{1}{1 - \delta} \sum_{i>k} p_{i+1} = \frac{1}{1 - \delta} y_{k+1},$$

we have $(1 - \delta)y_{k+1} < z_k < (1/(1 - \delta))y_{k+1}$ for all $k \geq 0$. Also,

$$(1 - \delta)^2(1 - r_{k+1}) < \frac{z_k r_k}{y_k} = \frac{z_k(1 - r_{k+1})r_k}{y_{k+1}} < \frac{1}{1 - \delta}(1 - r_{k+1}) \tag{16}$$

and, consequently, there exist constants A and B , $0 < A < B$, such that

$$A \sum_{k=1}^{m(n)} (1 - r_k) \leq H(\log n) \leq B \sum_{k=1}^{m(n)} (1 - r_k). \tag{17}$$

From (14) and (16), it is clear that (8) holds if and only if there exists a sequence $(x_n) \nearrow \infty$, with $x_1 \geq 1$, such that

$$\frac{\sum_{i=m(n)+1}^{m(nx_n)} (1 - r_i)}{\sum_{i=1}^{m(n)} (1 - r_i)} \rightarrow 0. \tag{18}$$

In order to prove (18), recall that

$$\frac{1}{nx_n} \leq y_{m(nx_n)-1} = y_{m(n)} \prod_{i=m(n)+1}^{m(nx_n)-1} (1 - r_i) < \frac{1}{n} \delta^{m(nx_n)-m(n)-1}$$

for all n such that $m(nx_n) - m(n) \geq 1$, and, therefore, that $m(nx_n) - m(n) - 1 < C \log x_n$ for some $C > 0$ and every $n \geq 1$. If we now define $x_n = (\sum_{i=1}^{m(n)} (1 - r_i)) \vee 1$, which increases to ∞ with n , we have

$$\frac{\sum_{i=m(n)+1}^{m(nx_n)} (1 - r_i)}{\sum_{i=1}^{m(n)} (1 - r_i)} \leq \frac{C \log x_n + 1}{x_n} \rightarrow 0.$$

To prove (9), note that

$$\sum_{i=1}^n i \tilde{G}(1/i)^2 = \sum_{k=1}^{m(n)} \sum_{i \leq n, m(i)=k} i \tilde{G}(1/i)^2 \leq \sum_{k=1}^{m(n)} \sum_{m(i)=k} i z_k^2 = \sum_{k=1}^{m(n)} z_k^2 h(k),$$

where $h(k) = \sum_{m(i)=k} i \sim (y_k^{-2} - y_{k-1}^{-2})/2$ as $k \rightarrow \infty$, since $y_k/y_{k-1} < \delta$ for all $k \geq 1$. Hence,

$$\begin{aligned} \sum_{k=1}^{m(n)} z_k^2 h(k) &\sim \sum_{k=1}^{m(n)} \frac{z_k^2 (y_k^{-2} - y_{k-1}^{-2})}{2} \\ &< \frac{1}{(1 - \delta)^2} \sum_{k=1}^{m(n)} \frac{y_{k+1}^2 (y_k^{-2} - y_{k-1}^{-2})}{2} \\ &= \frac{1}{(1 - \delta)^2} \sum_{k=1}^{m(n)} \frac{(1 - r_{k+1})^2 r_k (2 - r_k)}{2} < C \sum_{k=1}^{m(n)} (1 - r_k)^2. \end{aligned}$$

Also, from the definition of H and a simple change of variable, we obtain

$$H(\log t) = \int_1^t \tilde{G}(1/x) dx$$

and, since $\tilde{G}(1/k)$ decreases to 0, we have $\sum_{k=1}^n \tilde{G}(1/k) \sim H(\log n)$. Now, since $0 < 1 - r_k < 1$ and $\sum_{k=1}^{m(n)} (1 - r_k)$ diverges, for some $C' > 0$ we have

$$\frac{\sum_{i=1}^n i \tilde{G}(1/i)^2}{(H(\log n))^2} \leq C' \frac{\sum_{k=1}^{m(n)} (1 - r_k)^2}{(\sum_{k=1}^{m(n)} (1 - r_k))^2} \leq C' \frac{\sum_{k=1}^{m(n)} (1 - r_k)}{(\sum_{k=1}^{m(n)} (1 - r_k))^2} \rightarrow 0.$$

We end the proof by noting that assertion (b) follows at once from Theorem 2 since, by (17), $\sum_{j=1}^\infty (1 - r_j) < \infty$ implies that $\lim_{t \rightarrow \infty} H(t) < \infty$ and, hence, $\sum_{k=1}^n \min\{Z_1, \dots, Z_k\}$ grows almost surely to a finite limit.

3.5. The Lindeberg-type condition

In Subsection 3.4 we determined the normalizing constant for the martingale $N_n - \theta(M_n)$. Here, we deal with the Lindeberg-type condition (11) and prove the CLT. We first give two technical lemmas that will be used in the sequel.

Lemma 1. *If $(b_n) \nearrow \infty$ satisfies the condition*

$$\sum_{k=2}^n Y_{\Theta(\varepsilon b_n + \theta(M_{k-1}))} \xrightarrow{P} 0 \tag{19}$$

for all $\varepsilon > 0$, then (11) holds.

Proof. Let $\varepsilon > 0$ and N be such that $b_N \geq 2/\varepsilon$. If $|\xi_k| > \varepsilon b_n$, then it easily follows that

$$\Delta\theta(M_k) \geq |\xi_k| - 1 \geq \varepsilon b_n - 1 \geq \varepsilon b_n / 2$$

and

$$\xi_k^2 \leq I_k + (\Delta\theta(M_k))^2 + 2I_k \Delta\theta(M_k) \leq 4(\Delta\theta(M_k))^2,$$

for $n \geq N$ and $k \geq 2$. Therefore,

$$E[\xi_k^2 \mathbf{1}_{\{|\xi_k| > \varepsilon b_n\}} \mid \mathcal{F}_{k-1}] \leq 4 E[(\Delta\theta(M_k))^2 \mathbf{1}_{\{\Delta\theta(M_k) > \varepsilon b_n / 2\}} \mid \mathcal{F}_{k-1}].$$

This inequality means that (11) is satisfied if we establish that

$$\frac{1}{b_n^2} \sum_{k=2}^n E[(\Delta\theta(M_k))^2 \mathbf{1}_{\{\Delta\theta(M_k) > \varepsilon b_n\}} \mid \mathcal{F}_{k-1}] \xrightarrow{P} 0. \tag{20}$$

Noting that $\Theta(\varepsilon b_n + \theta(M_{k-1})) \geq M_{k-1}$, we compute

$$\begin{aligned} & E[(\Delta\theta(M_k))^2 \mathbf{1}_{\{\Delta\theta(M_k) > \varepsilon b_n\}} \mid \mathcal{F}_{k-1}] \\ &= E[(\theta(X_k) - \theta(M_{k-1}))^2 \mathbf{1}_{\{X_k > \Theta(\varepsilon b_n + \theta(M_{k-1}))\}} \mid \mathcal{F}_{k-1}] \\ &= \sum_{j=\Theta(\varepsilon b_n + \theta(M_{k-1})) + 1}^\infty \left(\sum_{i=M_{k-1} + 1}^j r_i \right)^2 p_j. \end{aligned}$$

Writing m for M_{k-1} and l for $\Theta(\theta(M_{k-1}) + \varepsilon b_n)$, we have

$$\begin{aligned} \sum_{j=l+1}^{\infty} \left(\sum_{i=m+1}^j r_i \right)^2 p_j &= \sum_{j=l+1}^{\infty} p_j \sum_{i=m+1}^j r_i^2 + 2 \sum_{j=l+1}^{\infty} p_j \sum_{i_1=m+1}^j r_{i_1} \sum_{i_2=i_1+1}^j r_{i_2} \\ &= \sum_{i=m+1}^l r_i^2 \sum_{j=l+1}^{\infty} p_j + \sum_{i=l+1}^{\infty} r_i^2 \sum_{j=i}^{\infty} p_j + 2 \sum_{i_1=m+1}^l r_{i_1} \sum_{i_2=i_1+1}^l r_{i_2} \sum_{j=l+1}^{\infty} p_j \\ &\quad + 2 \sum_{i_1=m+1}^l r_{i_1} \sum_{i_2=l+1}^{\infty} r_{i_2} \sum_{j=i_2}^{\infty} p_j + 2 \sum_{i_1=l+1}^{\infty} r_{i_1} \sum_{i_2=i_1+1}^{\infty} r_{i_2} \sum_{j=i_2}^{\infty} p_j \\ &= y_l \sum_{i=m+1}^l r_i^2 + \sum_{i=l+1}^{\infty} r_i^2 y_{i-1} + 2y_l \sum_{i_1=m+1}^l r_{i_1} \sum_{i_2=i_1+1}^l r_{i_2} \\ &\quad + 2 \sum_{i_1=m+1}^l r_{i_1} \sum_{i_2=l+1}^{\infty} r_{i_2} y_{i_2-1} + 2 \sum_{i_1=l+1}^{\infty} r_{i_1} \sum_{i_2=i_1+1}^{\infty} r_{i_2} y_{i_2-1} \\ &= y_l \left(\sum_{i=m+1}^l r_i^2 + 2 \sum_{i_1=m+1}^l r_{i_1} \sum_{i_2=i_1+1}^l r_{i_2} + 2 \sum_{i=m+1}^l r_i \right) \\ &\quad + \sum_{i=l+1}^{\infty} r_i p_i + 2 \sum_{i=l+1}^{\infty} r_i y_i \\ &\leq y_l \left(\left(\sum_{i=m+1}^l r_i \right)^2 + 2 \sum_{i=m+1}^l r_i + 3 \right) \\ &= y_l [(\theta(l) - \theta(m))^2 + 2(\theta(l) - \theta(m)) + 3], \end{aligned}$$

where the inequality follows because $\sum_{i>l} r_i p_i$ and $\sum_{i>l} r_i y_i$ are bounded above by y_l . On the other hand, since $\theta(\Theta(t)) \leq t$ for all $t \geq p_0$, we have $\theta(\Theta(\varepsilon b_n + \theta(M_{k-1}))) \leq \varepsilon b_n + \theta(M_{k-1})$, meaning that $\theta(l) - \theta(m) \leq \varepsilon b_n$ and

$$E[(\Delta\theta(M_k))^2 \mathbf{1}_{\{\Delta\theta(M_k) > \varepsilon b_n\}} \mid \mathcal{F}_{k-1}] \leq y_{\Theta(\varepsilon b_n + \theta(M_{k-1}))} (\varepsilon^2 b_n^2 + 2\varepsilon b_n + 3).$$

Finally, the convergence in (20) (which implies (11)) follows from hypothesis (19) and the previous inequality.

Lemma 2. *Let $k \in \mathbb{Z}_+$. Then, for $a > 0$,*

$$y_{\Theta(\theta(k)+a)} \leq e^{1-a} y_k \tag{21}$$

and, for $a < 0$ such that $\theta(k) + a \geq p_0$,

$$y_{\Theta(\theta(k)+a)} \geq e^{-a} y_k. \tag{22}$$

Proof. Recall that $\Theta(t) = \max\{k : \theta(k) \leq t\}$ and $\theta(\Theta(t)) \leq t < \theta(\Theta(t) + 1)$, for $t \geq p_0$. Then, letting $T = \theta(k) + a$, with $a > 0$, we obtain

$$a < \theta(\Theta(T) + 1) - \theta(k) = \sum_{i=k+1}^{\Theta(T)+1} r_i.$$

Thus, from the elementary inequality $1 - t \leq e^{-t}$, we obtain the first bound (21), since

$$\frac{y_{\Theta(T)}}{y_k} = \prod_{i=k+1}^{\Theta(T)} (1 - r_i) \leq \exp\left(-\sum_{i=k+1}^{\Theta(T)+1} r_i + r_{\Theta(T)+1}\right) \leq e^{-a+1}.$$

For $a < 0$, again letting $T = \theta(k) + a \geq p_0$, we have $\theta(\Theta(T)) \leq T$. This yields

$$\sum_{i=\Theta(T)+1}^k r_i \geq -a$$

and (22) follows at once, since

$$\frac{y_k}{y_{\Theta(T)}} = \prod_{i=\Theta(T)+1}^k (1 - r_i) \leq \exp\left(-\sum_{i=\Theta(T)+1}^k r_i\right) \leq e^a.$$

3.6. Proof of main result

We now state and prove the main result of this section.

Theorem 4. *Assume that $\sum_{k=0}^{\infty} (1 - r_k) = \infty$. If $\limsup_{k \rightarrow \infty} r_k < 1$ or $\liminf_{k \rightarrow \infty} r_k > 0$ then*

$$\frac{N_n - \theta(M_n)}{b_n} \xrightarrow{w} N(0, 1), \tag{23}$$

where (b_n) is as defined in (13).

If $\sum_{k=0}^{\infty} (1 - r_k) < \infty$ then $N_n - \theta(M_n)$ converges almost surely to a finite limit.

Proof. Let $\sum_{k=0}^{\infty} (1 - r_k) = \infty$. We check hypotheses (10) and (11), beginning with the case $\limsup_{k \rightarrow \infty} r_k < 1$. Clearly, (10) follows from Proposition 2. To check (11), recall that, for $\limsup_{k \rightarrow \infty} r_k < 1$, we have $z_k = \sum_{i>k} p_i (1 - r_i) \geq \inf_{i>0} (1 - r_i) y_k$ and, hence, $y_k \leq C z_k$ for a positive constant C and all $k \geq 0$. Now, by (21), $y_{\Theta(\varepsilon b_n + \theta(M_{k-1}))} \leq e^{1 - \varepsilon b_n} y_{M_{k-1}}$ and, therefore,

$$\begin{aligned} \sum_{k=2}^n y_{\Theta(\varepsilon b_n + \theta(M_{k-1}))} &\leq e^{1 - \varepsilon b_n} \sum_{k=2}^n y_{M_{k-1}} \\ &\leq C e^{1 - \varepsilon b_n} \sum_{k=2}^n z_{M_{k-1}} \\ &= C b_n^2 e^{1 - \varepsilon b_n} \frac{\sum_{k=2}^n z_{M_{k-1}}}{b_n^2} \xrightarrow{p} 0, \end{aligned}$$

by Proposition 2. Hence, (11) follows from Lemma 1.

We now consider the case $\liminf_{k \rightarrow \infty} r_k > 0$. As above, (10) follows from Proposition 2. To check (11), observe that there exists a $\delta, 0 < \delta < 1$, such that $1 - r_i \leq \delta$ for all $i \geq 0$. Now, since $z_k = \sum_{i>k} y_i r_i = \sum_{i>k} p_{i+1} r_i / r_{i+1}$, we have $z_k \geq (1 - \delta) y_{k+1}$ for all $k \geq 0$. On the other hand, from the definitions of θ and Θ , we have

$$\theta(\lfloor \varepsilon b_n + M_{k-1} \rfloor) = \sum_{i=M_{k-1}+1}^{\lfloor \varepsilon b_n + M_{k-1} \rfloor} r_i + \theta(M_{k-1}) \leq \varepsilon b_n + \theta(M_{k-1}),$$

where $\lfloor \cdot \rfloor$ denotes the largest integer less than or equal to its argument. Hence,

$$\lfloor \varepsilon b_n + M_{k-1} \rfloor = \Theta(\theta(\lfloor \varepsilon b_n + M_{k-1} \rfloor)) \leq \Theta(\varepsilon b_n + \theta(M_{k-1}))$$

and

$$y_{\Theta(\varepsilon b_n + \theta(M_{k-1}))} \leq y_{\lfloor \varepsilon b_n + M_{k-1} \rfloor} = y_{M_{k-1}+1} \prod_{i=M_{k-1}+2}^{\lfloor \varepsilon b_n + M_{k-1} \rfloor} (1 - r_i) \leq C z_{M_{k-1}} \delta^{\varepsilon b_n},$$

for a positive constant C and any n and k with $n \geq k \geq 2$ and $\varepsilon b_n \geq 1$. Thus, from Proposition 2,

$$\sum_{k=2}^n y_{\Theta(\varepsilon b_n + \theta(M_{k-1}))} \leq C \delta^{\varepsilon b_n} b_n^2 \frac{\sum_{k=2}^n z_{M_{k-1}}}{b_n^2} \xrightarrow{P} 0.$$

Finally, if $\sum_{k=0}^{\infty} (1 - r_k) < \infty$ then, by Proposition 2, the process of conditional variances converges and, therefore, the martingale converges almost surely.

3.7. Centering sequences

Here we consider the final step towards Theorem 1, namely the substitution of $\theta(M_n)$ by a deterministic sequence (a_n) in (23). This amounts to showing that

$$\frac{\theta(M_n) - a_n}{b_n} \xrightarrow{P} 0,$$

where (b_n) is as defined in (13). To this end, we use the following elementary result from extreme value theory; see [5, p. 116, Proposition 3.1.1].

Lemma 3. For a τ , $0 \leq \tau \leq \infty$, and a sequence (u_n) , the convergence $n(1 - F(u_n)) \rightarrow \tau$ is equivalent to the convergence $P[M_n \leq u_n] \rightarrow e^{-\tau}$.

Proposition 3. Assume that $\sum_{k=0}^{\infty} (1 - r_k) = \infty$. If $\limsup_{k \rightarrow \infty} r_k < 1$ or $\liminf_{k \rightarrow \infty} r_k > 0$ then

$$\frac{\theta(M_n) - \theta(m(n))}{b_n} \xrightarrow{P} 0. \tag{24}$$

Proof. The convergence in (24) is equivalent to $P[\theta(M_n) \leq \varepsilon b_n + \theta(m(n))] \rightarrow 1$ and $P[\theta(M_n) \leq -\varepsilon b_n + \theta(m(n))] \rightarrow 0$, for all $\varepsilon > 0$. According to Lemma 3, these conditions are respectively equivalent to

$$n P[\theta(X_n) > \varepsilon b_n + \theta(m(n))] = n y_{\Theta(\varepsilon b_n + \theta(m(n)))} \rightarrow 0 \tag{25}$$

and

$$n P[\theta(X_n) > -\varepsilon b_n + \theta(m(n))] = n y_{\Theta(-\varepsilon b_n + \theta(m(n)))} \rightarrow \infty. \tag{26}$$

Since $y_{m(n)} < 1/n$, from (21) we have

$$n y_{\Theta(\varepsilon b_n + \theta(m(n)))} \leq n e^{1-\varepsilon b_n} y_{m(n)} \leq e^{1-\varepsilon b_n} \rightarrow 0$$

and (25) follows.

On the other hand, since $y_{m(n)-1} \geq 1/n$, when $\limsup_{k \rightarrow \infty} r_k < 1$, (22) yields

$$n y_{\Theta(-\varepsilon b_n + \theta(m(n)))} \geq n e^{\varepsilon b_n} y_{m(n)} \geq e^{\varepsilon b_n} (1 - r_{m(n)}) \rightarrow \infty$$

and (26) follows. However, in order to apply (22) we require that $\theta(m(n)) - \varepsilon b_n \geq p_0$. This is the case since $z_k = \sum_{j>k} p_j(1 - r_j) \leq y_k$ and, therefore,

$$b_n^2 = \sum_{k=0}^{m(n)} \frac{z_k r_k}{y_k} \leq \sum_{k=0}^{m(n)} r_k = \theta(m(n)).$$

Then, for n sufficiently large, as $\theta(m(n)) \rightarrow \infty$ there exists an n such that

$$\theta(m(n)) \geq (\sqrt{p_0} + \varepsilon)^2 \quad \text{and} \quad \theta(m(n)) - \varepsilon b_n \geq \sqrt{\theta(m(n))}(\sqrt{\theta(m(n))} - \varepsilon) \geq p_0.$$

Now assume that $\liminf_{k \rightarrow \infty} r_k > 0$. Note first that $|\theta(M_n) - \theta(m(n))| \leq |M_n - m(n)|$; hence, it is enough to prove that

$$\frac{M_n - m(n)}{b_n} \xrightarrow{P} 0$$

or, equivalently, that

$$ny_{\lfloor \varepsilon b_n + m(n) \rfloor} \rightarrow 0 \quad \text{and} \quad ny_{\lfloor -\varepsilon b_n + m(n) \rfloor} \rightarrow \infty.$$

Consider a $\delta, 0 < \delta < 1$, such that $1 - r_i \leq \delta$ for all $i \geq 0$, and an n such that $\varepsilon b_n > 1$; for the first limit, we have

$$ny_{\lfloor \varepsilon b_n + m(n) \rfloor} = ny_{m(n)} \prod_{i=m(n)+1}^{\lfloor m(n) + \varepsilon b_n \rfloor} (1 - r_i) \leq ny_{m(n)} \delta^{\lfloor \varepsilon b_n \rfloor} \leq \delta^{\lfloor \varepsilon b_n \rfloor} \rightarrow 0$$

and, for the second, we have

$$ny_{\lfloor -\varepsilon b_n + m(n) \rfloor} = ny_{m(n)-1} \prod_{i=\lfloor m(n) - \varepsilon b_n \rfloor + 1}^{m(n)-1} (1 - r_i)^{-1} \geq \left(\frac{1}{\delta}\right)^{\lfloor \varepsilon b_n - 1 \rfloor} \rightarrow \infty.$$

3.8. Proof of Theorem 1

Part (a) of Theorem 1 follows immediately from Theorem 4 and Proposition 3. To prove part (b), note that the tightness of $N_n - m(n)$ is equivalent to

$$\frac{N_n - m(n)}{c_n} \xrightarrow{P} 0,$$

for every $(c_n) \nearrow \infty$. Write $N_n - m(n)$ as $(N_n - \theta(M_n)) + (\theta(M_n) - M_n) + (M_n - m(n))$ and let $(c_n) \nearrow \infty$. The convergence of $\sum_{k=0}^{\infty} (1 - r_k)$ yields, from Theorem 4, the convergence of the martingale and, consequently, $(N_n - \theta(M_n))/c_n \xrightarrow{\text{a.s.}} 0$. Also,

$$M_n - \theta(M_n) = \sum_{i=0}^{M_n} (1 - r_i)$$

converges and, so,

$$\frac{\theta(M_n) - M_n}{c_n} \xrightarrow{\text{a.s.}} 0.$$

Finally, for the case $\sum_{k=0}^{\infty} (1 - r_k) = \infty$, the proof of Proposition 3 implies that

$$\frac{M_n - m(n)}{c_n} \xrightarrow{P} 0.$$

4. Examples

In this section, we give detailed proofs of results presented in Examples 1–5, following Theorem 1. We begin with a proposition showing simplified expressions for the normalizing sequences in the case of converging failure rates.

Proposition 4. *Let (b_n) be as defined in (13) and let $r_k \rightarrow r \in [0, 1]$.*

- (a) *If $r = 0$ then $b_n^2 \sim \log n$.*
- (b) *If $r \in (0, 1)$ then $b_n^2 \sim -[r(1 - r)/\log(1 - r)] \log n$.*
- (c) *If $r = 1$ then $b_n^2 \sim \sum_{i=0}^{m(n)} (1 - r_i)$.*

Proof. (a) It is clear that $z_k = \sum_{i>k} p_i(1 - r_i) \sim \sum_{i>k} p_i = y_k$. Therefore,

$$\sum_{j=0}^{m(n)} \frac{z_j r_j}{y_j} \sim \sum_{j=0}^{m(n)} r_j.$$

On the other hand, from the definition of $m(t)$, we have $y_{m(n)} < 1/n \leq y_{m(n)-1}$. Next, recalling the identity $y_k = \prod_{i=0}^k (1 - r_i)$, we obtain

$$-\sum_{k=0}^{m(n)-1} \log(1 - r_k) \leq \log n < -\sum_{k=0}^{m(n)} \log(1 - r_k). \tag{27}$$

The above inequalities clearly imply that $-\sum_{k=0}^{m(n)} \log(1 - r_k) \sim \log n$ and, since $r_k \rightarrow 0$, we obtain $b_n^2 \sim \log n$.

(b) Here it is easily seen that $z_k/y_k \rightarrow 1 - r$, and we have

$$b_n^2 \sim (1 - r) \sum_{j=0}^{m(n)} r_j \sim r(1 - r)m(n).$$

Moreover, dividing (27) by $m(n)$ and taking limits, we find that $\log n/m(n) \rightarrow -\log(1 - r)$ and conclude that

$$b_n^2 \sim -\frac{r(1 - r)}{\log(1 - r)} \log n.$$

(c) The convergence $r_k \rightarrow 1$ implies that $z_k \sim y_{k+1}$ and, hence, that $z_k r_k/y_k \sim z_k/y_k \sim y_{k+1}/y_k = 1 - r_{k+1}$. From this, we finally obtain

$$b_n^2 \sim \sum_{j=0}^{m(n)} (1 - r_j).$$

4.1. Proofs of examples

4.1.1. *Proof of Example 1.* In this case, $r_k = p, k \geq 0, \theta(k) = p(k + 1)$, and

$$b_n^2 \sim -p(1 - p) \log n / \log(1 - p).$$

Also, $m(n) = \lfloor -\log n / \log(1 - p) \rfloor$. The convergence follows from Theorem 1.

4.1.2. *Proof of Example 2.* Here $b_n^2 \sim -r(1-r)\log n/\log(1-r)$. Moreover, $m(n) \sim -\log n/\log(1-r)$ and $\theta(m(n)) \sim -r\log n/\log(1-r)$. The result follows if

$$\frac{\theta(m(n)) + r\log n/\log(1-r)}{\sqrt{\log n}} \rightarrow 0.$$

A readily checked sufficient condition is that $\sum_{i=1}^n |\delta_i|/\sqrt{n} \rightarrow 0$, with $\delta_i = r_i - r$. This follows directly from the definition of $m(n)$. Indeed, since $y_{m(n)} < 1/n \leq y_{m(n)-1}$, we have

$$\begin{aligned} & -\frac{1}{\sqrt{m(n)}} - \frac{\sum_{i=0}^{m(n)} \log(1 - \delta_i/(1-r))}{\sqrt{m(n)} \log(1-r)} \\ & < \frac{m(n) + \log n/\log(1-r)}{\sqrt{m(n)}} \\ & \leq \frac{-\sum_{i=0}^{m(n)} \log(1 - \delta_i/(1-r))}{\sqrt{m(n)} \log(1-r)} + \frac{\log(1 - \delta_{m(n)}/(1-r))}{\sqrt{m(n)} \log(1-r)}. \end{aligned}$$

It is clear that the left- and right-hand sides of this expression tend to 0 when $\sum_{i=1}^n \delta_i/\sqrt{n} \rightarrow 0$. For the negative binomial distribution, it can be shown (see [13, p. 323, Example 3.1]) that

$$p - \frac{(a-1)(1-p)}{k} \leq r_k \leq p$$

and the condition above is satisfied with $r = p$.

4.1.3. *Proof of Example 3.* Let $r_{2k} = p$ and $r_{2k+1} = q$ for $k \geq 0$, where $0 < p < q < 1$. Theorem 1 applies since $\limsup_{k \rightarrow \infty} r_k = q < 1$.

For the centering and scaling sequences, first note that $|\theta(m(n)) - m(n)(p+q)/2| \leq 1$. On the other hand, from $y_{m(n)} < 1/n \leq y_{m(n)-1}$, we obtain

$$\left| m(n) + \frac{\log n}{\log \sqrt{(1-p)(1-q)}} \right| \leq C$$

for all n , where C is a positive constant. Therefore, there exists a constant C' such that

$$\left| \theta(m(n)) + \frac{p+q}{2} \frac{\log n}{\log \sqrt{(1-p)(1-q)}} \right| \leq C' \tag{28}$$

for all n . This means that the CLT holds with $\theta(m(n))$ replaced by the sequence in (28).

In order to obtain the scaling sequence b_n^2 , note that $y_{2k} = (1-p)A^k$ and $y_{2k+1} = A^{k+1}$, with $A = (1-p)(1-q)$. Furthermore,

$$z_{2k} = \frac{A^{k+1}}{1-A}(p(1-p) + q) \quad \text{and} \quad z_{2k+1} = \frac{A^{k+1}}{1-A}(1-p)(p+q(1-q)).$$

Finally, after some algebraic manipulation, we find that

$$\lim_{n \rightarrow \infty} \frac{b_n^2}{m(n)} = \lim_{n \rightarrow \infty} \frac{1}{m(n)} \sum_{k=0}^{m(n)} \frac{z_k r_k}{y_k} = \frac{1}{2}(p(1-p) + q(1-q)),$$

and (4) follows.

4.1.4. *Proof of Example 4.* From Proposition 4, we have $b_n^2 \sim \log n$. Under the condition $\sum_{k=1}^\infty r_k^2 < \infty$, it is easy to check that $|\theta(m(n)) - \log n| \leq C$ for a positive constant C , and the result follows.

4.1.5. *Proof of Example 5.* In this case, from Theorem 1 and Proposition 4, we have

$$\frac{N_n - m(n) + \sum_{i=1}^{m(n)} (1 - r_i)}{\sqrt{\sum_{i=1}^{m(n)} (1 - r_i)}} \xrightarrow{w} N(0, 1),$$

from which (6) and (7) easily follow.

Convergence for the Poisson distribution with parameter λ is a consequence of the following approximation for the failure rates r_k ; see [13, p. 328]:

$$\frac{\lambda}{k + 1} - \left(\frac{\lambda}{k + 1}\right)^2 \leq 1 - r_k \leq \frac{\lambda}{k + 1}.$$

5. Concluding remarks

Martingale methods are powerful and elegant tools for the asymptotic analysis of processes such as N_n but, in some cases, the technical details become rather involved. As suggested by a referee, an interesting alternative would be to explore the applicability of results on the limit behavior of sums of $\{0, 1\}$ -valued dependent RVs.

Another question of some interest is how the normalizing sequences (a_n) and (b_n) are related to the mean and variance of N_n , respectively. Some progress can be made for the centering sequence by using the martingale property and bounding the difference $E[\theta(M_n)] - \theta(m(n))$, as shown below.

Proposition 5. *If $\limsup_{k \rightarrow \infty} r_k < 1$ or $\liminf_{k \rightarrow \infty} r_k > 0$ then*

$$\frac{E[N_n]}{\theta(m(n))} \rightarrow 1. \tag{29}$$

Proof. First, from the martingale property, we have $E[N_n] = E[\theta(M_n)]$. On the other hand,

$$\begin{aligned} E[\theta(M_n)] &= \sum_{k=0}^\infty \theta(k) P[M_n = k] \\ &= \sum_{i=0}^\infty r_i (1 - (1 - y_{i-1})^n) \\ &= \sum_{i=0}^{m(n)} r_i (1 - (1 - y_{i-1})^n) + \sum_{i=m(n)+1}^\infty r_i (1 - (1 - y_{i-1})^n) \\ &= \theta(m(n)) - T_n + \sum_{i=m(n)+1}^\infty r_i (1 - (1 - y_{i-1})^n), \end{aligned}$$

with $T_n = \sum_{i=0}^{m(n)} r_i (1 - y_{i-1})^n$. We use the elementary inequality $1 - (1 - x)^n \leq nx$, for $0 < x < 1$ and $n \in \mathbb{Z}_+$, to obtain

$$\sum_{i=m(n)+1}^\infty r_i (1 - (1 - y_{i-1})^n) \leq n \sum_{i=m(n)+1}^\infty r_i y_{i-1} = n y_{m(n)} < 1.$$

Therefore,

$$-T_n \leq E[N_n] - \theta(m(n)) \leq 1,$$

and (29) follows if we show that

$$\frac{T_n}{\theta(m(n))} \rightarrow 0. \tag{30}$$

To this end, assume that there exists a sequence (x_n) tending to 0 such that $1/n \leq x_n \leq 1$ and

$$\frac{\theta(m(n)) - \theta(m(nx_n))}{\theta(m(n))} \rightarrow 0.$$

Then, since $i \leq m(n)$ is equivalent to $y_{i-1} \geq 1/n$, we have

$$\begin{aligned} T_n &= \sum_{i=0}^{m(nx_n)} r_i (1 - y_{i-1})^n + \sum_{i=m(nx_n)+1}^{m(n)} r_i (1 - y_{i-1})^n \\ &\leq \left(1 - \frac{1}{nx_n}\right)^n \theta(m(nx_n)) + \left(1 - \frac{1}{n}\right)^n (\theta(m(n)) - \theta(m(nx_n))), \end{aligned} \tag{31}$$

which clearly implies (30).

We finally show that such a sequence exists when $\liminf_{k \rightarrow \infty} r_k > 0$ or $\limsup_{k \rightarrow \infty} r_k < 1$. Suppose first that $\liminf_{k \rightarrow \infty} r_k > 0$, and take $x_n = [1/(m(n) + 1)] \vee (1/n)$. Consider a δ , $0 < \delta < 1$, such that $1 - r_k < \delta$ for all $k \geq 0$. Then, for all n such that $m(n) - m(nx_n) \geq 1$,

$$\frac{1}{n} \leq y_{m(n)-1} = y_{m(nx_n)} \prod_{i=m(nx_n)+1}^{m(n)-1} (1 - r_i) < \frac{1}{nx_n} \delta^{m(n)-m(nx_n)-1},$$

meaning that $m(n) - m(nx_n) - 1 < C \log(1/x_n)$ for some $C > 0$ and every $n \geq 1$. Therefore,

$$\begin{aligned} 0 \leq \frac{\theta(m(n)) - \theta(m(nx_n))}{\theta(m(n))} &= \frac{\sum_{i=m(nx_n)+1}^{m(n)} r_i}{\sum_{i=0}^{m(n)} r_i} \\ &\leq \frac{m(n) - m(nx_n)}{(1 - \delta)m(n)} \leq \frac{C \log(1/x_n) + 1}{(1 - \delta)m(n)} \rightarrow 0. \end{aligned}$$

In the case $\limsup_{k \rightarrow \infty} r_k < 1$, take $x_n = 1/(\log n + 1)$. Recall the definition of z_j . Then there exists an $a > 0$ such that $a \leq z_j/y_j \leq 1$ for all $j \geq 0$. Therefore,

$$\begin{aligned} 0 \leq \frac{\sum_{i=m(nx_n)+1}^{m(n)} r_i}{\sum_{i=0}^{m(n)} r_i} &\leq \frac{(1/a) \sum_{i=m(nx_n)+1}^{m(n)} r_i z_i / y_i}{\sum_{i=0}^{m(n)} r_i z_i / y_i} \\ &\leq \frac{H(\log n) - H(\log(nx_n)) + 1}{aH(\log n)} \leq C \frac{\log n - \log(nx_n) + 1}{\log n} \rightarrow 0, \end{aligned}$$

where the third inequality follows from (14) and the last one from the definition of H and (15).

Proposition 6 shows that $\theta(m(n))$ can be replaced by $E[N_n]$ in (1) when $\limsup_{k \rightarrow \infty} r_k < 1$.

Proposition 6. *If $\limsup_{k \rightarrow \infty} r_k < 1$ then*

$$\frac{E[N_n] - \theta(m(n))}{b_n} \rightarrow 0.$$

Proof. It suffices to show that

$$\frac{T_n}{b_n} = \sum_{i=0}^{m(n)} \frac{r_i(1 - y_{i-1})^n}{b_n} \rightarrow 0,$$

where $b_n^2 = \sum_{k=0}^{m(n)} z_k r_k / y_k$. From (31), with $x_n = 1/(\log n + 1)$, we obtain

$$0 \leq \frac{T_n}{b_n} \leq \left(1 - \frac{\log n + 1}{n}\right)^n \frac{\theta(m(n))}{b_n} + \left(1 - \frac{1}{n}\right)^n \frac{\theta(m(n)) - \theta(m(nx_n))}{b_n}. \tag{32}$$

Since $\limsup_{k \rightarrow \infty} r_k < 1$, from (14) and (15) there exist constants $A, B > 0$ such that $A \log n \leq b_n^2 \leq B \log n$ for all $n \geq 2$. Also, since there exists an $a > 0$ such that $a \leq z_j / y_j \leq 1$ for all $j \geq 0$, we have

$$C \log n \leq \theta(m(n)) \leq D \log n$$

for some $C, D > 0$ and all $n \geq 2$. Therefore, the right-hand side of (32) can be bounded above by

$$C' \left[\left(1 - \frac{\log n + 1}{n}\right)^n \sqrt{\log n} + \left(1 - \frac{1}{n}\right)^n \frac{\log \log n}{\sqrt{\log n}} \right] \rightarrow 0.$$

Unfortunately, it appears that the above arguments do not carry over to the case in which $\liminf_{k \rightarrow \infty} r_k > 0$.

The information obtained from Propositions 5 and 6 on the asymptotic behavior of $E[N_n]$ is still limited. We hope to carry out a more detailed analysis using, for instance, Rice’s method (see [6] and [10]), since $E[N_n]$ can be written as a binomial alternating sum

$$\sum_{k=1}^n \binom{n}{k} (-1)^{k+1} \varphi(k),$$

with $\varphi(k) = \sum_{i=0}^{\infty} r_i y_{i-1}^k$.

Acknowledgements

We thank an anonymous referee for a careful reading of the manuscript and useful comments which helped us improve the presentation of the paper; in particular, a much shorter proof of part (b) of Proposition 1. Financial support is gratefully acknowledged from the FONDAP Project in Applied Mathematics, FONDECYT grants 1020836 and 7020836, project 2001-2449 of MCYT, and CONSI+D project P119/2001 of DGA. F. J. López and G. Sanz are members of the research group on Stochastic Models (DGA).

References

- [1] AHSANULLAH, M. (1995). *Record Statistics*. Nova Science Publishers, Commack, NY.
- [2] ARNOLD, B. C., BALAKRISHNAN, N. AND NAGARAJA, H. N. (1998). *Records*. John Wiley, New York.
- [3] BAI, Z., HWANG, H. AND LIANG, W. (1998). Normal approximation of the number of records in geometrically distributed random variables. *Random Structures Algorithms* **13**, 319–334.
- [4] DEHEUVELS, P. (1974). Valeurs extrémales d’échantillons croissants d’une variable aléatoire réelle. *Ann. Inst. H. Poincaré Sect. B* **10**, 89–114.
- [5] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997). *Modelling Extremal Events*. Springer, Heidelberg.
- [6] FLAJOLET, P. AND SEDGEWICK, R. (1995). Mellin transforms and asymptotics: finite differences and Rice’s integrals. *Theoret. Comput. Sci.* **144**, 101–124.

- [7] GOUET, R., LÓPEZ, F. J. AND SAN MIGUEL, M. (2001). A martingale approach to strong convergence of the number of records. *Adv. Appl. Prob.* **33**, 864–873.
- [8] HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.
- [9] NEVZOROV, V. B. (2001). *Records: Mathematical Theory* (Transl. Math. Monogr. **194**). American Mathematical Society, Providence, RI.
- [10] PRODINGER, H. (1996). Combinatorics of geometrically distributed random variables: left-to-right maxima. *Discrete Math.* **153**, 253–270.
- [11] PUGH, W. (1990). Skip lists: a probabilistic alternative to balanced trees. *Commun. ACM* **32**, 668–676.
- [12] RÉNYI, A. (1962). Théorie des éléments saillants d'une suite d'observations. *Ann. Fac. Sci. Univ. Clermont-Ferrand* **8**, 7–13.
- [13] VERVAAT, W. (1973). Limit theorems for records from discrete distributions. *Stoch. Process. Appl.* **1**, 317–334.