

A law of the iterated logarithm for the number of occupied boxes in the Bernoulli sieve

Alexander Iksanov*, Wissem Jedidi † and Fethi Bouzeffour ‡

March 19, 2017

Abstract

The Bernoulli sieve is an infinite occupancy scheme obtained by allocating the points of a uniform $[0, 1]$ sample over an infinite collection of intervals made up by successive positions of a multiplicative random walk independent of the uniform sample. We prove a law of the iterated logarithm for the number of non-empty (occupied) intervals as the size of the uniform sample becomes large.

AMS 2010 subject classifications: primary 60F15; secondary 60K05

Keywords: Bernoulli sieve, infinite occupancy, law of iterated logarithm, perturbed random walk, renewal theory

1 Introduction

Let $R := (R_k)_{k \in \mathbb{N}_0}$ be a multiplicative random walk defined by

$$R_0 := 1, \quad R_k := \prod_{i=1}^k W_i, \quad k \in \mathbb{N}$$

where $(W_k)_{k \in \mathbb{N}}$ are independent copies of a random variable W taking values in the open interval $(0, 1)$. Also, let $(U_j)_{j \in \mathbb{N}}$ be independent random variables which are independent of R and have the uniform distribution on $[0, 1]$. A random occupancy scheme in which ‘balls’ U_1, U_2 , etc. are allocated over an infinite array of ‘boxes’ $(R_k, R_{k-1}]$, $k \in \mathbb{N}$ is called *Bernoulli sieve*. The Bernoulli sieve was introduced in [4] and further investigated in numerous articles which can be traced via the references given in the recent work [1]. We also refer to [1] for more details concerning the Bernoulli sieve including the origin of this term.

Since a particular ball falls into the box $(R_k, R_{k-1}]$ with random probability

$$p_k^* := R_{k-1} - R_k = W_1 W_2 \cdots W_{k-1} (1 - W_k), \quad (1.1)$$

*Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine, e-mail: iksan@univ.kiev.ua

†Department of Statistics & OR, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia and Université de Tunis El Manar, Faculté des Sciences de Tunis, LR11ES11 Laboratoire d’Analyse Mathématiques et Applications, 2092, Tunis, Tunisia, e-mail: wissem-jedidi@yahoo.fr

‡Department of Mathematics, College of Sciences, King Saud University, Riyadh 11451, Saudi Arabia, e-mail: fbouzeffour@ksu.edu.sa

the Bernoulli sieve is also the classical infinite occupancy scheme with the *random probabilities* $(p_k^*)_{k \in \mathbb{N}}$. In this setting, given the random probabilities (p_k^*) , the balls are allocated over the boxes $(R_1, R_0], (R_2, R_1], \dots$ independently with probability p_j^* of hitting box j . Assuming that the number of balls equals n , denote by K_n^* the number of non-empty boxes.

Under the condition $\sigma^2 := \text{Var}|\log W| \in (0, \infty)$ (which implies that $\mu := \mathbb{E}|\log W| < \infty$) it was shown in Corollary 1.1 of [5] that, as $n \rightarrow \infty$,

$$\frac{K_{[e^n]}^* - \mu^{-1} \int_0^n \mathbb{P}\{|\log(1 - W)| \leq y\} dy}{\sqrt{\sigma^2 \mu^{-3} n}}$$

converges in distribution to the standard normal law. The same conclusion can also be derived from a functional limit theorem obtained recently in [1]. The purpose of the present article is to obtain a law of the iterated logarithm that corresponds to the aforementioned central limit theorem.

For a family or a sequence (x_t) denote by $C((x_t))$ the set of its limit points.

Theorem 1.1. *Assume that $\sigma^2 \in (0, \infty)$ and that*

$$\mathbb{E}|\log(1 - W)|^a < \infty \tag{1.2}$$

for some $a > 0$. Then

$$C\left(\left(\frac{K_{[e^n]}^* - \mu^{-1} \int_0^n \mathbb{P}\{|\log(1 - W)| \leq y\} dy}{\sqrt{2\sigma^2 \mu^{-3} n \log \log n}} : n \geq 3\right)\right) = [-1, 1] \quad \text{a.s.}$$

In particular,

$$\limsup (\liminf)_{n \rightarrow \infty} \frac{K_{[e^n]}^* - \mu^{-1} \int_0^n \mathbb{P}\{|\log(1 - W)| \leq y\} dy}{\sqrt{2n \log \log n}} = +(-)\sigma \mu^{-3/2} \quad \text{a.s.}$$

Remark 1.2. Below we shall show that Theorem 1.1 is a consequence of Corollary 2.2 and Proposition 2.3. Condition (1.2) is only needed to ensure the applicability of Proposition 2.3. We do not know whether Proposition 2.3, hence Theorem 1.1, still hold true without condition (1.2).

The proof of Theorem 1.1 given in Section 2 relies upon a number of auxiliary results that are stated and proved in Section 3.

2 Proof of Theorem 1.1

Let $(\xi_k, \eta_k)_{k \in \mathbb{N}}$ be a sequence of i.i.d. two-dimensional random vectors with generic copy (ξ, η) where both ξ and η are positive. No condition is imposed on the dependence structure between ξ and η . Set

$$N(x) := \sum_{k \geq 0} \mathbb{1}_{\{S_k + \eta_{k+1} \leq x\}}, \quad x \geq 0$$

where $(S_n)_{n \in \mathbb{N}_0}$ is the zero-delayed ordinary random walk with increments ξ_n for $n \in \mathbb{N}$, i.e., $S_0 = 0$ and $S_n = \xi_1 + \dots + \xi_n$, $n \in \mathbb{N}$.

Put $\rho^*(x) := \sum_{k \geq 0} \mathbb{1}_{\{p_k^* \geq 1/x\}}$ for $x > 0$. It is natural to call $\rho^*(x)$ the number of ‘large boxes’ in the Bernoulli sieve. Relevance of $N(x)$ to the present context is justified by the equality

$$\rho^*(x) = N^*(\log x) \tag{2.1}$$

where the random variable $N^*(x)$ corresponds to $\xi_k = |\log W_k|$ and $\eta_k = |\log(1 - W_k)|$ for $k \in \mathbb{N}$.

Our strategy is as follows. First, we show in Corollary 2.2 that the number of occupied boxes K_n^* is well-approximated in the a.s. sense by $\rho^*(n)$. A similar approximation in the sense of distributional convergence was established in [5] and [1]. We would like to stress that proving the a.s. approximation is more delicate and calls for an additional argument. Second, we prove in Proposition 2.3 a law of the iterated logarithm for $N(x)$ defined in terms of arbitrary perturbed random walk. In view of (2.1) these two results are sufficient to complete the proof of Theorem 1.1.

We consider the infinite occupancy scheme in which balls are allocated independently with probability p_k of hitting box k . Denote by K_n be the number of occupied boxes in the scheme when n balls have been thrown. For $n \in \mathbb{N}$ set $\Theta_n := \sum_{k \geq 1} e^{-np_k} \mathbb{1}_{\{np_k \geq 1\}}$ and $\Delta_n := n \sum_{k \geq 1} p_k \mathbb{1}_{\{np_k < 1\}}$.

Lemma 2.1. *Suppose that*

$$\sum_{n \geq 1} n^{-2} \Theta_{[e^n]}^4 < \infty \quad \text{and} \quad \sum_{n \geq 1} n^{-2} \Delta_{[e^n]}^4 < \infty. \quad (2.2)$$

Then

$$\lim_{n \rightarrow \infty} n^{-1/2} \left(K_{[e^n]} - \sum_{k \geq 1} \mathbb{1}_{\{[e^n] p_k \geq 1\}} \right) = 0 \quad \text{a.s.} \quad (2.3)$$

Proof. We shall use a representation $K_n = \sum_{k \geq 1} \mathbb{1}_{\{Z_{n,k} \geq 1\}}$ where $Z_{n,k}$ is the number of balls that fall in box k . Observe that the random variable $Z_{n,k}$ has the binomial distribution with parameters n and p_k . With this at hand we can write

$$\left| K_n - \sum_{k \geq 1} \mathbb{1}_{\{np_k \geq 1\}} \right| \leq \sum_{k \geq 1} \mathbb{1}_{\{Z_{n,k} = 0\}} \mathbb{1}_{\{np_k \geq 1\}} + \sum_{k \geq 1} \mathbb{1}_{\{Z_{n,k} \geq 1\}} \mathbb{1}_{\{np_k < 1\}}.$$

Let $(A_k)_{k \in \mathbb{N}}$ be a sequence of sets which satisfy $\sum_{k \geq 1} \mathbb{1}_{A_k} < \infty$. The multinomial theorem tells us that

$$\begin{aligned} \left(\sum_{k \geq 1} \mathbb{1}_{A_k} \right)^4 &= \sum_{k \geq 1} \mathbb{1}_{A_k} + 14 \sum_{1 \leq j < i} \mathbb{1}_{A_j} \mathbb{1}_{A_i} \\ &+ 36 \sum_{1 \leq j < i < l} \mathbb{1}_{A_j} \mathbb{1}_{A_i} \mathbb{1}_{A_l} + 24 \sum_{1 \leq j < i < l < m} \mathbb{1}_{A_j} \mathbb{1}_{A_i} \mathbb{1}_{A_l} \mathbb{1}_{A_m}. \end{aligned} \quad (2.4)$$

Further, for $m \in \mathbb{N}$ and distinct $i_1, \dots, i_m \in \mathbb{N}$

$$\mathbb{P}\{Z_{n,i_1} = 0, \dots, Z_{n,i_m} = 0\} = (1 - p_{i_1} - \dots - p_{i_m})^n \leq e^{-n(p_{i_1} + \dots + p_{i_m})}. \quad (2.5)$$

Even though there is a precise formula

$$\begin{aligned} \mathbb{P}\{Z_{n,i_1} \geq 1, \dots, Z_{n,i_m} \geq 1\} &= 1 - \sum_{k=1}^m (1 - p_{i_k})^n + \sum_{1 \leq k < l \leq m} (1 - p_{i_k} - p_{i_l})^n \\ &+ \dots + (-1)^m (1 - p_{i_1} - \dots - p_{i_m})^n, \end{aligned}$$

a crude upper bound is of greater use for our needs:

$$\mathbb{P}\{Z_{n,i_1} \geq 1, \dots, Z_{n,i_m} \geq 1\} \leq n(n-1) \dots (n-m+1) p_{i_1} \dots p_{i_m} \leq n^m p_{i_1} \dots p_{i_m}.$$

While the product $p_{i_1} \cdots p_{i_m}$ is the probability of the event that the boxes i_1, \dots, i_m turn out occupied when throwing m balls, the product $n(n-1) \cdots (n-m+1) = m!(\binom{n}{m})$ is the number of ways to allocate m balls out of n into the boxes i_1, \dots, i_m .

Using (2.4) and then (2.5) we obtain

$$\begin{aligned}
& \mathbb{E} \left(\sum_{k \geq 1} \mathbb{1}_{\{Z_{n,k}=0\}} \mathbb{1}_{\{np_k \geq 1\}} \right)^4 \\
& \leq \sum_{k \geq 1} e^{-np_k} \mathbb{1}_{\{np_k \geq 1\}} + 14 \sum_{1 \leq j < i} e^{-np_j} \mathbb{1}_{\{np_j \geq 1\}} e^{-np_i} \mathbb{1}_{\{np_i \geq 1\}} \\
& + 36 \sum_{1 \leq j < i < l} e^{-np_j} \mathbb{1}_{\{np_j \geq 1\}} e^{-np_i} \mathbb{1}_{\{np_i \geq 1\}} e^{-np_l} \mathbb{1}_{\{np_l \geq 1\}} \\
& + 24 \sum_{1 \leq j < i < l < m} e^{-np_j} \mathbb{1}_{\{np_j \geq 1\}} e^{-np_i} \mathbb{1}_{\{np_i \geq 1\}} e^{-np_l} \mathbb{1}_{\{np_l \geq 1\}} e^{-np_m} \mathbb{1}_{\{np_m \geq 1\}} \\
& \leq \Theta_n + 7\Theta_n^2 + 6\Theta_n^3 + \Theta_n^4.
\end{aligned}$$

This in combination with (2.2) entails $\lim_{n \rightarrow \infty} n^{-1/2} \sum_{k \geq 1} \mathbb{1}_{\{Z_{[e^n],k}=0\}} \mathbb{1}_{\{[e^n]p_k \geq 1\}} = 0$ a.s. by the Borel-Cantelli lemma.

Arguing similarly we infer

$$\begin{aligned}
& \mathbb{E} \left(\sum_{k \geq 1} \mathbb{1}_{\{Z_{n,k} \geq 1\}} \mathbb{1}_{\{np_k < 1\}} \right)^4 \\
& \leq n \sum_{k \geq 1} p_k \mathbb{1}_{\{np_k < 1\}} + 14n^2 \sum_{1 \leq j < i} p_j \mathbb{1}_{\{np_j < 1\}} p_i \mathbb{1}_{\{np_i < 1\}} \\
& + 36n^3 \sum_{1 \leq j < i < l} p_j \mathbb{1}_{\{np_j < 1\}} p_i \mathbb{1}_{\{np_i < 1\}} p_l \mathbb{1}_{\{np_l < 1\}} \\
& + 24n^4 \sum_{1 \leq j < i < l < m} p_j \mathbb{1}_{\{np_j < 1\}} p_i \mathbb{1}_{\{np_i < 1\}} p_l \mathbb{1}_{\{np_l < 1\}} p_m \mathbb{1}_{\{np_m < 1\}} \\
& \leq \Delta_n + 7\Delta_n^2 + 6\Delta_n^3 + \Delta_n^4
\end{aligned}$$

which in combination with (2.2) proves $\lim_{n \rightarrow \infty} n^{-1/2} \sum_{k \geq 1} \mathbb{1}_{\{Z_{[e^n],k} \geq 1\}} \mathbb{1}_{\{[e^n]p_k < 1\}} = 0$ a.s. by another appeal to the Borel-Cantelli lemma. \square

Corollary 2.2.

$$\lim_{n \rightarrow \infty} n^{-1/2} (K_{[e^n]}^* - \rho^*(e^n)) = 0 \quad \text{a.s.}$$

Proof. Recalling (2.1) we have

$$0 \leq \rho^*(e^n) - \rho^*([e^n]) = N^*(n) - N(\log[e^n]) \leq N^*(n) - N^*(n-1)$$

for large enough n . By Lemma 3.4(b), the right-hand side divided by $n^{1/2}$ converges to zero a.s. Hence, it suffices to prove that

$$\lim_{n \rightarrow \infty} n^{-1/2} (K_{[e^n]}^* - \rho^*([e^n])) = 0 \quad \text{a.s.} \tag{2.6}$$

We write

$$\begin{aligned}
\Delta_n^* & := n \sum_{k \geq 1} p_k^* \mathbb{1}_{\{np_k^* < 1\}} = n \int_{(n, \infty)} x^{-1} d\rho^*(x) = \int_{(1, \infty)} x^{-1} d(\rho^*(nx) - \rho^*(n)) \\
& = \int_1^\infty x^{-2} (\rho^*(nx) - \rho^*(n)) dx
\end{aligned}$$

having utilized integration by parts and the asymptotics $\rho^*(x) = O(\log x)$ as $x \rightarrow \infty$ a.s. (see Lemma 3.4(a)) for the last step. Further, using convexity of $x \mapsto x^4$, $x > 0$ and Corollary 3.3 yields

$$\begin{aligned} \mathbb{E}(\Delta_n^*)^4 &= \mathbb{E}\left(\sum_{k \geq 2} \int_{k-1}^k x^{-2}(\rho^*(nx) - \rho^*(n))dx\right)^4 \\ &\leq \mathbb{E}\left(\sum_{k \geq 2} (\rho^*(nk) - \rho^*(n))((k-1)k)^{-1}\right)^4 \\ &\leq \sum_{k \geq 2} \mathbb{E}(\rho^*(nk) - \rho^*(n))^4((k-1)k)^{-1} \leq C \sum_{k \geq 2} (\log k)^4((k-1)k)^{-1} < \infty \end{aligned}$$

which proves

$$\sum_{n \geq 1} n^{-2} \mathbb{E}(\Delta_{[e^n]}^*)^4 < \infty.$$

Arguing similarly we obtain

$$\Theta_n^* := \sum_{k \geq 1} e^{-np_k^*} \mathbb{1}_{\{np_k^* \geq 1\}} = \int_{[1, n]} e^{-n/x} d\rho^*(x) = \int_1^n e^{-x}(\rho^*(n) - \rho^*(n/x))dx$$

and

$$\begin{aligned} &\mathbb{E}(\Theta_n^*)^4 \\ &= \mathbb{E}\left(\sum_{k=2}^n \int_{k-1}^k e^{-x}(\rho^*(n) - \rho^*(n/x))dx\right)^4 \\ &\leq (e^{-1} - e^{-n})^4 \mathbb{E}\left(\sum_{k \geq 2} (\rho^*(n) - \rho^*(n/k))(e^{-k+1} - e^{-k})(e^{-1} - e^{-n})^{-1}\right)^4 \\ &\leq (e-1)(e^{-1} - e^{-n})^3 \sum_{k \geq 2} \mathbb{E}(\rho^*(n) - \rho^*(n/k))^4 e^{-k} \leq C \sum_{k \geq 2} (\log k)^4 e^{-k} < \infty. \end{aligned}$$

Thus,

$$\sum_{n \geq 1} n^{-2} \mathbb{E}(\Theta_{[e^n]}^*)^4 < \infty.$$

Invoking now Lemma 2.1 enables us to conclude that (2.6) holds conditionally on $(p_k^*)_{k \in \mathbb{N}}$, hence also unconditionally. The proof of Corollary 2.2 is complete. \square

Proposition 2.3. *Suppose that $\mathbf{s}^2 := \text{Var } \xi \in (0, \infty)$ and $\mathbb{E}\eta^a < \infty$ for some $a > 0$. Then*

$$C\left(\left(\frac{N(n) - \mathbf{m}^{-1} \int_0^n F(y)dy}{\sqrt{2\mathbf{s}^2 \mathbf{m}^{-3} n \log \log n}} : n \geq 3\right)\right) = [-1, 1] \quad \text{a.s.}$$

where $\mathbf{m} := \mathbb{E}\eta < \infty$ and $F(y) := \mathbb{P}\{\eta \leq y\}$ for $y \geq 0$.

Proof. Put

$$\nu(x) := \sum_{k \geq 0} \mathbb{1}_{\{S_k \leq x\}}, \quad x \geq 0.$$

It is known (see the proof of Theorem 3.2 in [1]) that

$$\lim_{n \rightarrow \infty} n^{-1/2} \left(N(n) - \int_{[0, n]} F(n-y) d\nu(y) \right) = 0 \quad \text{a.s.}$$

whenever $\mathbb{E}\eta^a < \infty$ for some $a > 0$ (the finiteness of $\text{Var } \xi$ is not needed). Thus, it remains to prove that

$$C\left(\left(\frac{\int_{[0,n]} F(n-y)d(\nu(y) - \mathfrak{m}^{-1}y)}{\sqrt{2\mathfrak{s}^2\mathfrak{m}^{-3}n \log \log n}} : n \geq 3\right)\right) = [-1, 1] \quad \text{a.s.} \quad (2.7)$$

Put $a(t) := \sqrt{2\mathfrak{s}^2\mathfrak{m}^{-3}t \log \log t}$ for $t \geq 3$. Integrating by parts yields

$$\begin{aligned} \frac{\int_{[0,n]} F(n-y)d(\nu(y) - \mathfrak{m}^{-1}y) - \mathbb{P}\{\eta = n\}}{a(n)} &= \int_{[0,n]} \frac{\nu(n-y) - \mathfrak{m}^{-1}(n-y)}{a(n)} dF(y) \\ &= \int_{[0,\delta]} \frac{\nu(n-y) - \mathfrak{m}^{-1}(n-y)}{a(n)} dF(y) \\ &\quad + \int_{(\delta,n)} \frac{\nu(n-y) - \mathfrak{m}^{-1}(n-y)}{a(n)} dF(y) \\ &=: Z_1(n) + Z_2(n) \end{aligned}$$

for any fixed $\delta \in (0, n]$. We have a.s.

$$\begin{aligned} \frac{\nu(n) - \mathfrak{m}^{-1}n}{a(n)} F(\delta) - \frac{\nu(n) - \nu(n-\delta)}{a(n)} F(\delta) \\ \leq Z_1(n) \\ \leq \frac{\nu(n) - \mathfrak{m}^{-1}n}{a(n)} F(\delta) + \frac{\mathfrak{m}^{-1}\delta}{a(n)} F(\delta). \end{aligned}$$

Fix any $x_0 \in [-1, 1]$. According to (3.6), there exists a sequence (n_k) satisfying $\lim_{k \rightarrow \infty} n_k = \infty$ a.s. and $\lim_{k \rightarrow \infty} (\nu(n_k) - \mathfrak{m}^{-1}n_k)/a(n_k) = x_0$ a.s. By Lemma 3.4(b), $\lim_{k \rightarrow \infty} (\nu(n_k) - \nu(n_k - \delta))/a(n_k) = 0$ a.s. Therefore, $\lim_{\delta \rightarrow \infty} \lim_{k \rightarrow \infty} Z_1(n_k) = x_0$ a.s. Further,

$$\begin{aligned} & - \frac{\sup_{y \in [0,n]} |\nu(y) - \mathfrak{m}^{-1}y|}{a(n)} (F(n-) - F(\delta)) \\ & \leq \frac{\inf_{y \in [0, n-\delta]} (\nu(y) - \mathfrak{m}^{-1}y)}{a(n)} (F(n-) - F(\delta)) \\ & \leq Z_2(n) \\ & \leq \frac{\sup_{y \in [0,n]} |\nu(y) - \mathfrak{m}^{-1}y|}{a(n)} (F(n-) - F(\delta)). \end{aligned}$$

Using (3.7) we conclude that

$$\lim_{\delta \rightarrow \infty} \limsup_{k \rightarrow \infty} Z_2(n_k) = \lim_{\delta \rightarrow \infty} \liminf_{k \rightarrow \infty} Z_2(n_k) = 0 \quad \text{a.s.}$$

The proof of (2.7) is complete. \square

Now Theorem 1.1 follows from Corollary 2.2 in combination with a specialization of Proposition 2.3 for $\rho^*(e^n) = N^*(n)$ which reads

$$C\left(\left(\frac{\rho^*(e^n) - \mu^{-1} \int_0^n \mathbb{P}\{\log |1 - W| \leq y\} dy}{\sqrt{2\sigma^2 \mu^{-3} n \log \log n}} : n \geq 3\right)\right) = [-1, 1] \quad \text{a.s.}$$

3 Auxiliary results

The following result can be found in the proof of Lemma A.3 in [1].

Lemma 3.1. *Let $G : [0, \infty) \rightarrow [0, \infty)$ be a locally bounded function. Then, for any $l \in \mathbb{N}$*

$$\mathbb{E} \left(\sum_{k \geq 0} G(t - S_k) \mathbb{1}_{\{S_k \leq t\}} \right)^l \leq \left(\sum_{j=0}^{\lfloor t \rfloor} \sup_{y \in [j, j+1)} G(y) \right)^l \mathbb{E}(\nu(1))^l, \quad t \geq 0. \quad (3.1)$$

Lemma 3.2. *For $0 \leq y < x$ with $x - y > 1$ $\mathbb{E}(N(x) - N(y))^4 \leq C(x - y)^4$ for a positive constant C which does not depend on x and y .*

Proof. Throughout the proof we assume that x and y satisfy the assumptions of the lemma.

We start with

$$\mathbb{E}(N(x) - N(y))^4 \leq 8(\mathbb{E}(X(x, y))^4 + \mathbb{E}(Y(x, y))^4)$$

where

$$\begin{aligned} X(x, y) &:= \sum_{j \geq 0} ((\mathbb{1}_{\{S_j + \eta_{j+1} \leq x\}} - F(x - S_j) \mathbb{1}_{\{S_j \leq x\}}) \\ &\quad - (\mathbb{1}_{\{S_j + \eta_{j+1} \leq y\}} - F(y - S_j) \mathbb{1}_{\{S_j \leq y\}})); \end{aligned}$$

$$\begin{aligned} Y(x, y) &:= \sum_{j \geq 0} (F(x - S_j) \mathbb{1}_{\{S_j \leq x\}} - F(y - S_j) \mathbb{1}_{\{S_j \leq y\}}) \\ &= \int_{[0, y]} (\nu(x - z) - \nu(y - z)) dF(z) + \int_{(y, x]} \nu(x - z) dF(z); \end{aligned}$$

$F(z) = \mathbb{P}\{\eta \leq z\}$ is the distribution function of η and $\nu(z) = \sum_{k \geq 0} \mathbb{1}_{\{S_k \leq z\}}$ for $z \geq 0$.

We intend to show that $\mathbb{E}(X(x, y))^4 \leq C(x - y)^2$. With $x, y \geq 0$ fixed, $X(x, y)$ equals the terminal value of the martingale $(R(k), \mathcal{F}_k)_{k \in \mathbb{N}_0}$ where $R(0) := 0$,

$$\begin{aligned} R(k) &:= \sum_{j=0}^{k-1} ((\mathbb{1}_{\{S_j + \eta_{j+1} \leq x\}} - F(x - S_j) \mathbb{1}_{\{S_j \leq x\}}) \\ &\quad - (\mathbb{1}_{\{S_j + \eta_{j+1} \leq y\}} - F(y - S_j) \mathbb{1}_{\{S_j \leq y\}})), \end{aligned}$$

$\mathcal{F}_0 := \{\Omega, \emptyset\}$ and $\mathcal{F}_k := \sigma((\xi_j, \eta_j) : 1 \leq j \leq k)$. We use the Burkholder-Davis-Gundy inequality (Theorem 11.3.2 in [3]) to obtain, for any $l \in \mathbb{N}$

$$\begin{aligned} &\mathbb{E}(X(x, y))^{2l} \\ &\leq C_l \left(\mathbb{E} \left(\sum_{k \geq 0} \mathbb{E}((R(k+1) - R(k))^2 | \mathcal{F}_k) \right)^l + \sum_{k \geq 0} \mathbb{E}(R(k+1) - R(k))^{2l} \right) \\ &=: C_l(I_1 + I_2) \end{aligned}$$

for a positive constant C_l . We shall show that

$$I_1 \leq 2^l \mathbb{E}(\nu(1))^l (b(x - y))^l \quad (3.2)$$

where $b(t) := \sum_{k=0}^{\lfloor t \rfloor + 1} (1 - F(k))$ for $t \geq 0$ and that

$$I_2 \leq 2^{2l} \mathbb{E} \nu(1) b(x - y). \quad (3.3)$$

These estimates serve our needs because $b(t) \leq \lfloor t \rfloor + 2 \leq 3t$ whenever $t > 1$.

PROOF OF (3.2). We first observe that

$$\begin{aligned} & \sum_{k \geq 0} \mathbb{E} ((R(k+1) - R(k))^2 | \mathcal{F}_k) \\ &= \int_{(y, x]} F(x-z)(1-F(x-z)) d\nu(z) \\ &+ \int_{[0, y]} (F(x-z) - F(y-z))(1-F(x-z) + F(y-z)) d\nu(z) \\ &\leq \int_{(y, x]} (1-F(x-z)) d\nu(z) + \int_{[0, y]} (F(x-z) - F(y-z)) d\nu(z) \end{aligned}$$

whence

$$I_1 \leq 2^{l-1} \left(\mathbb{E} \left(\int_{(y, x]} (1-F(x-z)) d\nu(z) \right)^l + \mathbb{E} \left(\int_{[0, y]} (F(x-z) - F(y-z)) d\nu(z) \right)^l \right)$$

having utilized $(u+v)^l \leq 2^{l-1}(u^l + v^l)$ for nonnegative u and v . Using Lemma 3.1 with $G(z) = (1-F(z)) \mathbb{1}_{[0, x-y]}(z)$ and $G(z) = F(x-y+z) - F(z)$, respectively, we obtain

$$\begin{aligned} & \mathbb{E} \left(\int_{(y, x]} (1-F(x-z)) d\nu(z) \right)^l \\ &= \mathbb{E} \left(\int_{[0, x]} (1-F(x-z)) \mathbb{1}_{[0, x-y]}(x-z) d\nu(z) \right)^l \\ &\leq \mathbb{E}(\nu(1))^l \left(\sum_{n=0}^{\lfloor x \rfloor} \sup_{y \in [n, n+1)} ((1-F(z)) \mathbb{1}_{[0, x-y]}(z)) \right)^l \\ &\leq \mathbb{E}(\nu(1))^l \left(\sum_{n=0}^{\lfloor x-y \rfloor} (1-F(n)) \right)^l \leq \mathbb{E}(\nu(1))^l (b(x-y))^l \end{aligned} \quad (3.4)$$

and

$$\begin{aligned} & \mathbb{E} \left(\int_{[0, y]} (F(x-z) - F(y-z)) d\nu(z) \right)^l \\ &\leq \mathbb{E}(\nu(1))^l \left(\sum_{n=0}^{\lfloor y \rfloor} \sup_{z \in [n, n+1)} (F(x-y+z) - F(z)) \right)^l \\ &\leq \mathbb{E}(\nu(1))^l \left(\sum_{n=0}^{\lfloor y \rfloor} (1-F(n)) - \sum_{n=0}^{\lfloor y \rfloor} (1-F(x-y+n+1)) \right)^l \\ &\leq \mathbb{E}(\nu(1))^l \left(\sum_{n=0}^{\lfloor y \rfloor} (1-F(n)) - \sum_{n=0}^{\lfloor y \rfloor + 2} (1-F(n)) + \sum_{n=0}^{\lfloor x-y \rfloor + 1} (1-F(n)) \right)^l \\ &\leq \mathbb{E}(\nu(1))^l (b(x-y))^l. \end{aligned} \quad (3.5)$$

Combining (3.4) and (3.5) for $l = 2$ yields (3.2).

PROOF OF (3.3). Let us calculate

$$\begin{aligned}
& \mathbb{E}((R(k+1) - R(k))^4 | \mathcal{F}_k) \\
& \leq 8((1 - F(x - S_k))^4 F(x - S_k) + (F(x - S_k))^4 (1 - F(x - S_k))) \mathbb{1}_{\{y < S_k \leq x\}} \\
& + ((1 - F(x - S_k) + F(y - S_k))^4 (F(x - S_k) - F(y - S_k))) \\
& + (F(x - S_k) - F(y - S_k))^4 (1 - F(x - S_k) + F(y - S_k)) \mathbb{1}_{\{S_k \leq y\}} \\
& \leq 8((1 - F(x - S_k)) \mathbb{1}_{\{y < S_k \leq x\}} + (F(x - S_k) - F(y - S_k)) \mathbb{1}_{\{S_k \leq y\}}).
\end{aligned}$$

Therefore,

$$I_2 \leq 8 \left(\mathbb{E} \int_{(x,y]} (1 - F(x - z)) d\nu(z) + \mathbb{E} \int_{[0,y]} (F(x - z) - F(y - z)) d\nu(z) \right).$$

Using now formulae (3.4) and (3.5) with $l = 1$ yields (3.3).

Passing to $Y(x, y)$ we have

$$\mathbb{E}(Y(x, y))^4 \leq 8 \left(\mathbb{E} \left(\int_{[0,y]} (\nu(x - z) - \nu(y - z)) dF(z) \right)^4 + \mathbb{E}(\nu(x - y))^4 \right).$$

Using the fact that $z^{-4} \mathbb{E}(\nu(z))^4$ converges as $z \rightarrow \infty$ to a nonnegative constant (see Theorem 5.1 on p. 57 in [6]) we infer $\mathbb{E}(\nu(x - y))^4 \leq C(x - y)^4$ (recall that $x - y > 1$). Finally,

$$\begin{aligned}
& \mathbb{E} \left(\int_{[0,y]} (\nu(x - z) - \nu(y - z)) dF(z) \right)^4 \\
& \leq (F([y] + 1))^4 \mathbb{E} \left(\sum_{k=0}^{[y]} (\nu(x - k) - \nu(y - k - 1)) \frac{F(k + 1) - F(k)}{F([y] + 1)} \right)^4 \\
& \leq \sum_{k=0}^{[y]} \mathbb{E}(\nu(x - k) - \nu(y - k - 1))^4 (F(k + 1) - F(k)) \\
& \leq \mathbb{E}(\nu(x - y + 1))^4 \leq C(x - y)^4
\end{aligned}$$

where we have used distributional subadditivity of $\nu(z)$ (see formula (5.7) on p. 58 in [6]) for the penultimate inequality. \square

In view of (2.1) the next result is an immediate consequence of Lemma 3.2.

Corollary 3.3. $\mathbb{E}(\rho^*(x) - \rho^*(y))^4 \leq C(\log(x/y))^4$ for a positive constant C which does not depend on x and y .

Lemma 3.4. (a) $N(x) = O(x)$ a.s. as $x \rightarrow \infty$;
(b) For any $c > 0$ and any fixed $\delta > 0$ $\lim_{n \rightarrow \infty} n^{-c}(N(n) - N(n - \delta)) = 0$ a.s. and $\lim_{n \rightarrow \infty} n^{-c}(\nu(n) - \nu(n - \delta)) = 0$ a.s.

Proof. (a) Since the η_k is a.s. positive, it follows that $N(x) \leq \nu(x)$ a.s. It remains to note that $\nu(x) = O(x)$ a.s. as $x \rightarrow \infty$ by the strong law of large numbers for the renewal processes, see Theorem 5.1 on p. 57 in [6].

(b) The limit relation that involves N can be found in the proof of Proposition 3.3 in [1]. Setting $\eta \equiv 1$ immediately gives the second limit relation. \square

Proposition 3.5. *Suppose that $s^2 = \text{Var } \xi \in (0, \infty)$. Then*

$$C\left(\left(\frac{\nu(t) - m^{-1}t}{\sqrt{2s^2m^{-3}t \log \log t}}\right) : t \geq 3\right) = [-1, 1] \quad \text{a.s.} \quad (3.6)$$

where $m = \mathbb{E}\xi < \infty$, and

$$\limsup_{n \rightarrow \infty} \frac{\sup_{0 \leq y \leq n} |\nu(y) - m^{-1}y|}{\sqrt{2n \log \log n}} = sm^{-3/2} \quad \text{a.s.} \quad (3.7)$$

Remark 3.6. While formula (3.6) was known before, see, for instance, Theorem 11.1 on p. 108 in [6], we have not been able to locate formula (3.7) in the literature. We derive both (3.6) and (3.7) from a functional law of the iterated logarithm. The proof of (3.6), other than that mentioned on p. 108 in [6], is included, for it requires no extra work in the given framework.

Proof of Proposition 3.5. Denote by D the Skorokhod space of right-continuous real-valued functions which are defined on $[0, \infty)$ and have finite limits from the left at each positive point. We shall need the commonly used J_1 -topology on D , see [2, 8].

For integer $n \geq 3$, set

$$X_n(t) := \frac{\nu(nt) - m^{-1}nt}{\sqrt{2s^2m^{-3}n \log \log n}}, \quad t \geq 0.$$

We shall write (X_n) for $(X_n(t))_{t \geq 0}$. Let K denote the set of real-valued absolutely continuous functions g on $[0, \infty)$ such that $g(0) = 0$ and $\int_0^\infty (g'(t))^2 dt \leq 1$. The set K is called the *Strassen set*. It is known (see p. 44 in [7] or Theorem 7.3 on p. 173 in [6]) that the sequence $(X_n)_{n \geq 3}$ is, with probability one, relatively compact in the J_1 -topology, and the set of its limit points coincides with K . The evaluation and the supremum functionals $h_1, h_2 : D \rightarrow \mathbb{R}$ defined by $h_1(x) := x(1)$ and $h_2(x) := \sup_{t \in [0, 1]} |x(t)|$, respectively, are continuous in the J_1 -topology at each $x \in K$. Hence, for $i = 1, 2$, by the continuous mapping theorem $(h_i(X_n))_{n \geq 3}$ are, with probability one, relatively compact in the J_1 -topology, and the sets of their limit points coincide with $h_i(K)$.

PROOF OF (3.6). We first show that (3.6) holds with an integer argument replacing a continuous argument. To this end, it remains to prove that $h_1(K) = [-1, 1]$ which is a consequence of two facts: (I) $g(1) \in [-1, 1]$ for each $g \in K$; (II) each point of $[-1, 1]$ is a possible value of $g(1)$ for some $g \in K$.

Let $g \in K$ and $t \in (0, 1]$. From

$$(g(t))^2 = \left(\int_0^t g'(y) dy\right)^2 \leq \int_0^t (g'(y))^2 dy \int_0^t dy \leq t \quad (3.8)$$

it follows that $g(1) \in [-1, 1]$. To prove (II), set $g_a^\pm(t) := \pm \min(t, a)$, for each $a \in [0, 1]$. Then $g_a^\pm \in K$ and $g_a^\pm(1) = \pm a$.

Recall the notation $a(t) = \sqrt{2s^2m^{-3}t \log \log t}$ for $t \geq 3$. To pass in (3.6) from an integer argument to a continuous argument it is enough to check that if $\lim_{k \rightarrow \infty} (\nu(t_k) - m^{-1}t_k)/a(t_k) = b$ a.s. for some sequence (t_k) of *real numbers* and some $b \in \mathbb{R} \cup \{\pm\infty\}$, then $\lim_{k \rightarrow \infty} (\nu(n_k) - m^{-1}n_k)/a(n_k) = b$ a.s. for some sequence (n_k) of *integers*. Writing

$$\begin{aligned} \frac{\nu([t_k]) - m^{-1}[t_k]}{a([t_k] + 1)} - \frac{m^{-1}}{a([t_k] + 1)} &\leq \frac{\nu(t_k) - m^{-1}t_k}{a(t_k)} \leq \frac{\nu([t_k]) - m^{-1}[t_k]}{a([t_k])} \\ &+ \frac{\nu([t_k] + 1) - \nu([t_k])}{a([t_k])}, \end{aligned}$$

where $[x]$ denotes the integer part of x , and noting that $\lim_{t \rightarrow \infty} (\nu(t+1) - \nu(t))/a(t) = 0$ a.s. by Lemma 3.4(b) and $\lim_{t \rightarrow \infty} a(t+1)/a(t) = 1$ we conclude that the implication above does indeed hold with $n_k := [t_k]$.

PROOF OF (3.7). From what has been proved above it follows that the left-hand side of (3.7) equals $\sup_{g \in K} (\sup_{t \in [0,1]} |g(t)|)$ a.s. In view of (3.8) the last expression does not exceed one. Since $\sup_{t \in [0,1]} |g_1^+(t)| = 1$ (recall that $g_1^+(t) = \min(t, 1)$), we infer $\sup_{g \in K} (\sup_{t \in [0,1]} |g(t)|) = 1$ which completes the proof of (3.7). \square

Acknowledgements The second and third authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for funding their Research group No. (RG-1437-020).

References

- [1] ALSMEYER, G., IKSANOV, A. AND MARYNYCH, A. (2017). Functional limit theorems for the number of occupied boxes in the Bernoulli sieve. *Stoch. Proc. Appl.* **127**, 995–1017.
- [2] BILLINGSLEY, P. (1968). *Convergence of probability measures*. Wiley: New York.
- [3] CHOW, Y.S. AND TEICHER, H. (2003). *Probability theory: independence, interchangeability, martingales*, 3rd edition. Springer: New York.
- [4] GNEDIN, A. (2004). The Bernoulli sieve. *Bernoulli*. **10**, 79–96.
- [5] GNEDIN, A., IKSANOV, A. AND MARYNYCH, A. (2010). Limit theorems for the number of occupied boxes in the Bernoulli sieve. *Theory Stoch. Proc.* **16(32)**, 44–57.
- [6] GUT, A. (2009). *Stopped random walks. Limit theorems and applications*, 2nd edition. Springer-Verlag: New York.
- [7] VERVAAT, W. (1972). *Success epochs in Bernoulli trials (with applications in number theory)*, Mathematical Centre Tracts, **42**. Mathematisch Centrum: Amsterdam.
- [8] WHITT, W. (2002). *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. Springer-Verlag: New York.