SHARP PROBABILITY ESTIMATES FOR RANDOM WALKS WITH BARRIERS

KEVIN FORD

ABSTRACT. We give sharp, uniform estimates for the probability that a random walk of n steps on the reals avoids a half-line $[y,\infty)$ given that it ends at the point x. The estimates hold for general continuous or lattice distributions provided the 4th moment is finite.

1. INTRODUCTION

Let $X_1, X_2, \dots$ be independent, identically distributed random variables with mean $\mathbf{E}X_1 = 0$ and variance $\mathbf{E}X_1^2 = 1$. Let $S_0 = T_0 = 0$ and for $n \ge 1$ define

$$S_n = X_1 + \cdots + X_n$$

and

$$T_n = \max(0, S_1, \dots, S_n).$$

The estimation of the distribution of S_n for general random variables has a long and rich history (see e.g. [10]).

The distribution of T_n was found more recently. In 1946, Erdős and Kac [5] showed that

$$\lim_{n\to\infty} \mathbf{P}[T_n \le x\sqrt{n}] = 2\Phi(x) - 1$$

uniformly in $x \ge 0$, where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

is the distribution function for the normal distribution. Several estimates for the error term have been proved on the assumption that $\mathbf{E}|X_1|^3 < \infty$, the best uniform bound (and best possible uniform bound) being the result of Nagaev [13]

$$\mathbf{P}[T_n \le x\sqrt{n}] = 2\Phi(x) - 1 + O(1/\sqrt{n}),$$

uniformly in $x \ge 0$ (the constant implied by the $O-$symbol depends only on $\mathbf{E}|X_1|^3$). Sharper error terms are possible when $|x| \ge 1$, see e.g. Arak [3] and Chapter 4 of [2].

We are interested here in approximations of the conditional probability

$$R_n(x,y) = \mathbf{P}[T_{n-1} < y | S_n = x]$$

which are sharp for a wide range of x, y. By the invariance principle, we expect

$$R_n(u\sqrt{n}, v\sqrt{n}) \to 1 - e^{-2v(v-u)} \qquad (n \to \infty)$$

Date: 11 July 2008.
The author was supported by NSF grants DMS-0301083 and DMS-0555367.
2000 Mathematics Subject Classification: Primary 60G50.
Key words and phrases : random walk, barrier, ballot theorems.

for u, v fixed, $u \le v$ and $v \ge 0$, since this holds for the case of Bernoulli random variables (see (2.1) below).

Before stating our results, we motivate the study of $R_n(x, y)$ with three examples, two of which are connected with empirical processes.

2. THREE EXAMPLES

The example which is easiest to analyze is the case of a simple random walk with Bernoulli steps. Let $X_1, X_2, \dots$ satisfy $\mathbf{P}[X_i = 1] = \mathbf{P}[X_i = -1] = \frac{1}{2}$. By the reflection principle, the number of walks of n steps for which $T_n \ge y$ and $S_n = x$ is equal to the number of walks of n steps with $S_n = 2y - x$ (by inverting $X_{k+1}, \dots, X_n$, where k is the smallest index with $S_k = y$). Thus, if n and x have the same parity, then

$$R_n(x, y) = 1 - \frac{\binom{n}{\frac{n+x-2y}{2}}}{\binom{n}{\frac{n-x}{2}}}.$$

This includes as a special case a version of the classical Bertrand ballot theorem from 1887. Two candidates in an election receive p and q votes, respectively, with $p \ge q$. If the votes are counted in random order, the probability that the first candidate never trails in the counting is

$$R_{p+q}(q - p, 1) = \frac{p - q + 1}{p + 1}.$$

More generally, suppose $1 \le y \le n/2$, $-n/2 \le x < y$ and $2y - x \le n/2$. Writing $\beta = (2y - x)/n$ and $\alpha = x/n$, so that $\beta > \alpha > 0$, we obtain by Stirling's formula,

$$\begin{aligned} R_n(x, y) &= 1 - \frac{\binom{n}{\frac{n}{2}(1+\beta)}}{\binom{n}{\frac{n}{2}(1+\alpha)}} \\ &= 1 - (1 + O(1/n))\sqrt{\frac{1-\alpha^2}{1-\beta^2}} \left(\frac{(1+\alpha)^{1+\alpha}(1-\alpha)^{1-\alpha}}{(1+\beta)^{1+\beta}(1-\beta)^{1-\beta}} \right)^{n/2} \\ &= 1 - (1 + O(1/n))\sqrt{\frac{1-\alpha^2}{1-\beta^2}} \exp\left\{ \frac{n}{2} \left(\alpha^2 - \beta^2 + O\left(\alpha^4 + \beta^4\right)\right)\right\}. \end{aligned}$$

If $x = O(\sqrt{n})$ and $y - x = O(\sqrt{n})$, then $\alpha = O(n^{-1/2})$ and $\beta = O(n^{-1/2})$ and we have

$$R_n(x, y) = 1 - (1 + O(1/n)) \exp\{\tfrac{n}{2}(\alpha^2 - \beta^2)\} = 1 - e^{-2y(y-x)/n} + O(1/n). \tag{2.1}$$

Two special cases are connected with empirical processes. Let $U_1, \dots, U_n$ be independent random variables with uniform distribution in $[0, 1]$, suppose $F_n(t) = \frac{1}{n}\sum_{U_i \le t} 1$ is their empirical distribution function and $0 \le \xi_1 \le \dots \le \xi_n \le 1$ are their order statistics.

In his seminal 1933 paper [11] on the distribution of the statistic

$$D_n = \sqrt{n} \sup_{0 \le t \le 1} |F_n(t) - t|,$$

Kolmogorov related the problem to a similar conditional probability for a random walk. Specifically, let $X_1, X_2, \dots, X_n$ be independent random variables with discrete distribution

$$\mathbf{P}[X_j = r-1] = \frac{e^{-1}}{r!} \qquad (r = 0, 1, 2, \dots) \tag{2.2}$$

Kolmogorov proved that for integers $u \ge 1$,

$$\begin{aligned}\mathbf{P}(\sup_{0\le t\le 1} |F_n(t)-t| \le u/n) &= \frac{n!e^n}{n^n}\mathbf{P}\left(\max_{0\le j\le n-1}|S_j| < u, S_n = 0\right)\\ &= \mathbf{P}\left(\max_{0\le j\le n-1}|S_j| < u \,\Big|\, S_n = 0\right).\end{aligned}$$

Consider next

$$Q_n(u,v) = \mathbf{P}[\xi_i \ge \tfrac{i-u}{v}\ (1\le i\le n)] = \mathbf{P}\left(F_n(t) \le \frac{vt+u}{n}\ \ (0\le t\le 1)\right)$$

for $u \ge 0, v > 0$. Smirnov in 1939 proved the asymptotic $Q_n(\lambda\sqrt{n}, n) \to 1 - e^{-2\lambda^2}$ as $n\to\infty$ for fixed λ. Small modifications to Kolmogorov's proof yield, for *integers* $u \ge 1$ and for $n \ge 2$, that

$$Q_n(u,n) = R_n(0,u)$$

for the variables X_j given by (2.2). When $v \ne n$, however, it does not seem possible to express $Q_n(u,v)$ in terms of these variables X_j.

In [8], new bounds on $Q_n(u,v)$ were proved and applied to a problem of the distribution of divisors of integers (see also articles [6], [7] for more about this application). A more precise uniform estimate was proved in [9], namely

$$Q_n(u,v) = 1 - e^{-2uw/n} + O\left(\frac{u+w}{n}\right) \qquad (n\ge 1, u\ge 0, w\ge 0), \tag{2.3}$$

where $w = u+v-n$ and the constant implied by the $O-$symbol is independent of u, v and n. This was accomplished using $X_j = 1 - Y_j$, where $Y_1, Y_2, \dots$ are independent random variables with exponential distribution, i.e. with density function $f(x) = e^{-x}$ for $x \ge 0$, $f(x) = 0$ for $x < 0$. Letting $W_k = Y_1 + \dots + Y_k$, Rényi [16] whowed that

$$(\xi_1, \xi_2, \cdots, \xi_n) \text{ and } \left(\frac{W_1}{W_{n+1}}, \frac{W_2}{W_{n+1}}, \cdots, \frac{W_n}{W_{n+1}}\right)$$

have the same distribution. An easy consequence is

$$Q_n(u,v) = \mathbf{P}\big[W_j - j \ge -u\ (1\le j\le n) \mid W_{n+1} = v\big] = R_{n+1}(n+1-v, u).$$

3. Statement of the main results

Our aim in this paper is to prove a result analogous to (2.1) and (2.3) for sums of very general random variables X_1. We will restrict ourselves to random variables with either a continuous or lattice distribution, to maintain control of the density function of S_n. Let F be the distribution function of X_1 and let F_n the distribution function of S_n for $n \ge 1$. Let $\phi(t) = \mathbf{E}e^{itX_1}$ be the characteristic function of X_1.

We henceforth assume that either

$$\text{(C)} \qquad \begin{cases} X_1 \text{ has a continuous distribution and} \\ \exists n_0 : \int |\phi(t)|^{n_0}\, dt < \infty \end{cases}$$

or that

$$\text{(L)} \qquad X_1 \text{ has a lattice distribution.}$$

If (L), let $f(x) = \mathbf{P}(X_1 = x)$, $f_n(x) = \mathbf{P}(S_n = x)$ and $n_0 = 1$. We also suppose the support of f is contained in the lattice $\mathscr{L} = \{\gamma + m\lambda : m \in \mathbb{Z}\}$, where λ is the maximal span of the distribution (the support of f is not contained in any lattice $\{\gamma' + m\lambda' : m \in \mathbb{Z}\}$ with $\lambda' > \lambda$). The support of f_n is then contained in the lattice $\mathscr{L}_n = \{n\gamma + m\lambda : m \in \mathbb{Z}\}$. If (C), let f be the density function of X_1, let f_n the density function of S_n, define $\mathscr{L} = \mathbb{R}$ and $\mathscr{L}_n = \mathbb{R}$.

Define the moments

$$\alpha_u = \mathbf{E}X_1^u, \qquad \beta_u = \mathbf{E}|X_1|^u.$$

In what follows, the notation $f = O(g)$ for functions f, g means that for some constant $c > 0$, $|f| \le cg$ for all values of the domain of f, which will usually be given explicitly. Unless otherwise specified, c may depend only on the distribution of X_1, but not on any other parameter. Sometimes we use the Vinogradov notation $f \ll g$ which means $f = O(g)$. As $R_n(x, y)$ is only defined when $f_n(x) > 0$, when $f_n(x) = 0$ we define $R_n(x, y) = 1$.

Theorem 1. *Assume (C) or (L), $\beta_u < \infty$ for some $u > 3$, and let $M > 0$. Uniformly in $n \ge 1$, $0 \le y \le M\sqrt{n}$, $0 \le z \le M\sqrt{n}$ with $y \in \mathscr{L}_n$, $y - z \in \mathscr{L}_n$ and $f_n(y - z) > 0$,*

$$R_n(y - z, y) = 1 - e^{-2yz/n} + O\left(\frac{y + z + 1}{n} + \frac{1}{n^{\frac{u-2}{2}}} \right).$$

Here the constant implied by the $O-$symbol depends on the distribution of X_1, u and also on M, but not on n, y or z.

Corollary 1. *Assume (C) or (L) and $\beta_u < \infty$ for some $u > 3$. For $w \le v$ and $v \ge 0$,*

$$R_n(w\sqrt{n}, v\sqrt{n}) = 1 - e^{-2v(v-w)} + O(n^{-1/2}),$$

the constant implied by the $O-$symbol depending on $\max(v, v - w)$ and on the distribution of X_1.

Corollary 2. *Assume (C) or (L), and $\beta_4 < \infty$. If y and z satisfy $y \to \infty$, $y = o(\sqrt{n})$, $z \to \infty$, and $z = o(\sqrt{n})$ as $n \to \infty$, then*

$$\lim_{n \to \infty} \frac{R_n(y - z, y)}{2yz/n} = 1.$$

All three examples given in section 2 staisfy the hypotheses of Theorem 1 and the two corollaries. Indeed, for these examples all moments of X_1 exist.

Using "almost sure invariance" principles or "strong approximation" theorems (see e.g. [4], [15]), one can approximate the walk $(S_n)_{n \ge 0}$ with a Wiener process $W(n)$. Assuming that $\beta_4 < \infty$ and no higher moments exist, one has $S_n - W(n) = o(n^{1/4})$ almost surely, the exponent $1/4$ being best possible (cf. [4], Theorems 2.6.3, 2.6.4). This rate of approximation is, however, far too weak to prove results as strong as Theorem 1.

In section 4, we list some required estimates for $f_n(x)$. Section 5 contains two recursion formulas for $R_n(x, y)$. Although our main interest is in the case when $y \ge x$, we shall need estimates when $y < x$ in order to prove Theorem 1. This is accomplished in §6. Finally, in §7, we prove Theorem 1. It is critical to our analysis that the densities $f_n(x)$ have regular behavior, and the hypotheses (C), (L) and $\beta_u < \infty$ ensures that this is the case for $|x| = O(\sqrt{n})$. Extending the range of validity of the asymptotic for $R_n(x, y)$ beyond the range of (x, y) covered by Theorem 1 would require that we have more precise estimates for $f_n(x)$ for $|x|$ of larger order than $\sqrt{n}$. In specific cases, such as the exponential distribution, normal distribution or binomial distribution, exact expressions for $f_n(x)$ make it possible to achieve this goal (see e.g. (2.3)).

It is of some interest to know if $\beta_4 < \infty$ really is a necessary condition for Theorem 1 to hold. Recently, Addario-Berry and Reed [1] showed (as a special case of their Theorem 1), for an arbitrary lattice random variable X_1 with zero mean and finite variance, that

$$\frac{yz}{n} \ll R_n(y-z, y) \ll \frac{yz}{n} \qquad (1 \le y, z \le \sqrt{n}, n \ge n_0), \tag{3.1}$$

the constants implied by the $\ll$-symbols and n_0 depending on the distribution of X_1. The same proof gives (3.1) under hypotheses (C) and $\beta_2 < \infty$; see (a) below (for non-lattice variables, the authors prove analogous bounds for the probability that $T_n < y$ given that $y - z - c \le S_n \le y - z$, for a fixed $c > 0$). When $y = 1$, the upper bound in (3.1) is the same as the conclusion as Theorem 1, but is proved under a weaker hypothesis. When y is larger, however, the error term in the conclusion of Theorem 1 can be of much lower order than the main term, and a hypothesis stronger than $\beta_2 < \infty$ should be required. Addario-Berry and Reed also construct examples of variables X_1 where $\mathbf{E}X_1^2 = \infty$ or $k/\sqrt{n} \to \infty$, while $R_n(-k, 1)$ is not of order k/n.

4. ESTIMATES FOR DENSITY FUNCTIONS

At the core of our arguments are approximations of the density function $f_n(x)$. This is the only part of the proof which uses the hypothesis on $\phi(t)$ from (C).

Lemma 4.1. *Assumer (C) or (L), and $\beta_2 = 1$. Then, uniformly for $n \ge n_0$ and all x,*

$$f_n(x) \ll \frac{1}{\sqrt{n}}. \tag{4.1}$$

Assume $3 \le u \le 4$, $\beta_u < \infty$, and (C) or (L). Then, uniformly for $n \ge n_0$ and $x \in \mathscr{L}_n$,

$$\begin{aligned} f_n(x) &= \frac{e^{-x^2/2n}}{\sqrt{2\pi n}} \left[1 + O\left(\frac{|x|}{n} + \frac{|x|^3}{n^2} \right) \right] + O(n^{(1-u)/2}) \\ &= \frac{e^{-x^2/2n}}{\sqrt{2\pi n}} + O\left(\frac{|x|}{n^{3/2}(1 + x^2/n)} + n^{(1-u)/2} \right). \end{aligned}$$

Proof. We apply results from [10], §46, §47 and §51. Assume (C). By the proof of Theorem 1 in §46, we may replace conditions 1), 2) of §46, Theorem 1 and the theorem in §47 with the hypothesis that n_0 exists. Note that these theorems are only stated with the hypothesis that β_u exists for intergal u, but straightforward modification of the proofs yields the above inequalities for real $u \in [3, 4]$: Start with the inequality $e^{it} = 1 + it - \frac{1}{2}t^2 - \frac{i}{6}t^3 + O(|t|^u)$, which follows from Taylor's formula for $|t| \le 1$ and the triangle inequality for $|t| > 1$. Consequently,

$$\phi(t) = 1 - \frac{1}{2}t^2 - \frac{i\alpha_3}{6}t^3 + O(|t|^u)$$

and hence, for $|t|$ small enough,

$$\begin{aligned}\phi^n(t) &= \exp\left[-\frac{nt^2}{2} - \frac{i\alpha_3 n}{6}t^3 + O(n|t|^u)\right] \\ &= e^{-nt^2/2}\left[1 - \frac{i\alpha_3 n}{6}t^3 + O\left(t^6 n^2 e^{O(|t|^3 n)} + |t|^u n e^{O(|t|^u n)}\right)\right].\end{aligned}$$

Here we used the inequalities $|e^v - 1| \le |v|e^{|v|}$ and $|e^v - 1 - v| \le |v|^2 e^{|v|}$. Therefore,

$$\left|\phi^n(t) - e^{-nt^2/2}\left(1 - \frac{i\alpha_3 n}{6}t^3\right)\right| \ll (t^6 n^2 + |t|^u n)e^{-nt^2/4} \quad (|t| \le c) \tag{4.2}$$

for some $c > 0$. In the proofs in §46, §47 and §51, use (4.2) in place of Theorem 1 of §41. □

5. Recursion formulas

It is convenient to work with the density function

$$\widetilde{R}_n(x,y) = f_n(x)R_n(x,y) = \mathbf{P}[T_{n-1} < y, S_n = x].$$

The last expression stands for $\frac{d}{dx}\mathbf{P}[T_{n-1} < y, S_n \le x]$ when (C) holds. Notice that if $f_n(x) = 0$, then $\widetilde{R}_n(x,y) = 0$ by our convention.

Lemma 5.1. *Assume (C). Then, for $n \ge 2$, $y \ge 0$ and $s \ge 0$,*

$$\widetilde{R}_n(y+s,y) = \int_0^\infty f(s+t)\widetilde{R}_{n-1}(y-t,y)\,dt.$$

If (L), then for $n \ge 2$, $y > 0$, $s \ge 0$ and $y + s \in \mathscr{L}_n$,

$$\widetilde{R}_n(y+s,y) = \sum_{\substack{s+t\in\mathscr{L}_n \\ t>0}} f(s+t)\widetilde{R}_{n-1}(y-t,y).$$

Proof. If $S_n = y + s$ and $T_{n-1} < y$, then $X_n = s + t$ where $t > 0$. □

Lemma 5.1 expresses $\widetilde{R}_n(x,y)$ with $x \ge y$ in terms of $\widetilde{R}_{n-1}(x,y)$ with $x \le y$. The next lemma works the other direction, and is motivated by the reflection principle: a walk that crosses the point y and ends up at $S_n = x$ should be about as likely as a walk that ends up at $S_n = 2y - x$ (by inverting the part of the walk past the first crossing of y). We thus expect that for $x < y$,

$$\widetilde{R}_n(x,y) \approx f_n(x) - f_n(2y-x).$$

Lemma 5.2. *Assume $n \ge 2$, $y > 0$ and $a \ge 0$. If (C), then for any x*

$$\begin{aligned}\widetilde{R}_n(x,y) = f_n(x) - f_n(y+a) + \widetilde{R}_n(y+a,y) \\ + \int_0^\infty \sum_{k=1}^{n-1} \widetilde{R}_k(y+\xi,y)\left(f_{n-k}(a-\xi) - f_{n-k}(x-y-\xi)\right)\,d\xi.\end{aligned}$$

If (L), then for $x, y+a \in \mathscr{L}_n$,

$$\begin{aligned}\widetilde{R}_n(x,y) &= f_n(x) - f_n(y+a) + \widetilde{R}_n(y+a,y) \\ &\quad + \sum_{k=1}^{n-1} \sum_{\substack{y+\xi\in\mathscr{L}_k \\ \xi\geq 0}} \widetilde{R}_k(y+\xi,y)\left(f_{n-k}(a-\xi) - f_{n-k}(x-y-\xi)\right).\end{aligned}$$

Proof. First, we have

$$\begin{aligned}\widetilde{R}_n(x,y) &= f_n(x) - \mathbf{P}[T_{n-1} \geq y, S_n = x] \\ &= f_n(x) - f_n(y+a) + f_n(y+a) - \mathbf{P}[T_{n-1} \geq y, S_n = x].\end{aligned}$$

If $T_j \geq y$, then there is a unique k, $1 \leq k \leq j$, for which $T_{k-1} < y$ and $S_k \geq y$. Thus,

$$f_n(y+a) = \sum_{k=1}^{n} \mathbf{P}[T_{k-1} < y, S_k \geq y, S_n = y+a].$$

If (C) then

$$\begin{aligned}f_n(y+a) &= \sum_{k=1}^{n-1} \int_0^\infty \mathbf{P}[T_{k-1} < y, S_k = y+\xi, S_n = y+a]\, d\xi + \mathbf{P}[T_{n-1} < y, S_n = y+a] \\ &= \sum_{k=1}^{n-1} \int_0^\infty \widetilde{R}_k(y+\xi,y) f_{n-k}(a-\xi)\, d\xi + \widetilde{R}_n(y+a,y).\end{aligned}$$

Likewise, if (L) then

$$f_n(y+a) = \widetilde{R}_n(y+a,y) + \sum_{k=1}^{n-1} \sum_{\substack{y+\xi\in\mathscr{L}_k \\ \xi\geq 0}} \widetilde{R}_k(y+\xi,y) f_{n-k}(a-\xi).$$

In the same way

$$\begin{aligned}\mathbf{P}[T_{n-1} \geq y, S_n = x] &= \sum_{k=1}^{n-1} \mathbf{P}[T_{k-1} < y, S_k \geq y, S_n = x) \\ &= \sum_{k=1}^{n-1} \begin{cases} \int_0^\infty \widetilde{R}_k(y+\xi,y) f_{n-k}(x-y-\xi)\, d\xi & \text{if (C)} \\ \sum_{\substack{y+\xi\in\mathscr{L}_k \\ \xi\geq 0}} \widetilde{R}_k(y+\xi,y) f_{n-k}(x-y-\xi) & \text{if (L).} \end{cases}\end{aligned}$$

□

Motivated by the reflection principle, we will apply Lemma 5.2 with a close to $y-x$. The integral/sum over ξ is then expected to be small, since $f_{n-k}(y-x-\xi) - f_{n-k}(x-y-\xi)$ should be small when ξ is small (by Lemma 4.1) and $\widetilde{R}_k(y+\xi,y)$ should be small when ξ is large. This last fact is crucial to our argument, and we develop the necessary bounds in the next section.

6. ROUGH ESTIMATES

Roughly speaking, if $T_{n-1} < y$ and $S_n = y + s$ with $s \ge 0$, then S_{n-1} will be close to y with high probability. The probability that $T_{n-1} < y$ is about $\min(1, y/\sqrt{n})$ (see Lemma 6.1 below) On the other hand, if $y/\sqrt{n}$ is large, then $S_{n-1} \approx y$ is a rare event. Therefore, as a function of y, $\widetilde{R}_n(y+s, y)$ should increase linearly in y for $1 \le y \le \sqrt{n}$, then decrease for larger y.

We begin with a lemma concerning the distribution of T_n. Part (1) is due to Kozlov ([12], Theorem A, (13)) and part (2) was proved by Pemantle and Peres ([14], Lemma 3.3 (ii)). In [14], (2) is stated only for $h = 0$, but the same proof gives the more general inequality.

Lemma 6.1. *Assume X_1 is any random variable with $\beta_2 = 1$. Then*

(1) $\mathbf{P}\{T_n \le h\} \ll (h+1)/\sqrt{n}$.
(2) $\mathbf{E}\{S_n^2 | T_n \le h\} \ll n$, *uniformly in $h \ge 0$.*

Theorem 2. *Assume (C) or (L), $\beta_2 = 1$ and $n \ge 1$. For all $y \ge 0$, $z \ge 0$, we have*

$$\widetilde{R}_n(y-z, y) \ll \frac{\min(y+1, \sqrt{n}) \min(z+1, \sqrt{n})}{n^{3/2}}. \tag{a}$$

If $n \ge 3n_0$, $y \ge \sqrt{n}$ and $0 \le z \le y/2$, then

$$\widetilde{R}_n(y-z, y) \ll \frac{\min(z+1, \sqrt{n})}{y^2}. \tag{b}$$

Proof. The proof of (a) follows the upper bound proof of Theorem 1 from [1]. The idea is to consider simultaneously the random walk $0, S_1, S_2, \ldots$ and the "reverse" walk $0, \widetilde{S}_1, \widetilde{S}_2, \ldots$, where $\widetilde{S}_k = -(X_n + X_{n-1} + \cdots + X_{n-k+1})$. Let $\widetilde{T}_n = \max(0, \widetilde{S}_1, \ldots, \widetilde{S}_n)$. Note that $T_n \le y$ and $S_n = y - z$ imply $\widetilde{T}_n \le z$.

Inequality (a) is trivial for $1 \le n < 3n_0$. Let $n \ge 3n_0$, put $a = \lfloor n/3 \rfloor$ and $b = n - a$. Then $\widetilde{R}_n(y-z, y) \le \mathbf{P}(E_1, E_2, E_3)$, where $E_1 = \{T_a \le y\}$, $E_2 = \{\widetilde{T}_a \le z\}$ and $E_3 = \{S_n = y - z\}$. Think of the random walk $0, S_1, \ldots, S_n$ as the union of three independent subwalks: one consisting of the first a steps, one consisting of steps numbered $a+1$ to b, and one consiting of the last a steps reversed. Note that $E_3 = \{S_b - S_a = y - z - S_a + \widetilde{S}_a\}$. Since $S_b - S_a$ is independent of S_a, $\widetilde{S}_a$ and of events E_1 and E_2, we have by (4.1)

$$\mathbf{P}(E_3 | E_1, E_2) \le \sup_w f_{b-a}(w) \ll n^{-1/2}.$$

As E_1 and E_2 are independent, we have by Lemma 6.1 part (1)

$$\widetilde{R}_n(y-z, y) \le \mathbf{P}E_1 \, \mathbf{P}E_2 \, \mathbf{P}\{E_3 | E_1, E_2\} \ll \frac{\min(y+1, \sqrt{n}) \min(z+1, \sqrt{n})}{n^{3/2}}.$$

To prove (b), we observe that $S_n = y - z \ge y/2$. Thus, $S_a \ge y/6$, $S_b - S_a \ge y/6$ or $\widetilde{S}_a \le -y/6$. Suppose first that $S_a \ge y/6$. Replace E_1 by $E_1' = \{S_a \ge y/6\}$ in the above argument and note that

$$\mathbf{P}\{S_a \ge y/6\} \le \frac{\mathbf{E}S_a^2}{(y/6)^2} = \frac{36a}{y^2} \ll \frac{n}{y^2}.$$

Arguing as in the proof of (a), we find that

$$\mathbf{P}\{T_n \le y, S_n = y - z, S_a \ge y/6\} \ll \binom{n}{y^2} \frac{\min(z+1,\sqrt{n})}{n} \ll \frac{\min(z+1,\sqrt{n})}{y^2}.$$

Next, suppose that $S_b - S_a \ge y/6$. In the above argument, replace E_1 with $E_1'' = \{S_b - S_a \ge y/6\}$. Then $E_3 = \{S_a = y - z - (S_b - S_a) + \widetilde{S}_a\}$. Again, $\mathbf{P}\{E_3|E_1'', E_2\} \le \sup_w f_a(w) \ll n^{-1/2}$ and we obtain

$$\mathbf{P}\{T_n \le y, S_n = y - z, S_b - S_a \ge y/6\} \ll \frac{\min(z+1,\sqrt{n})}{y^2}.$$

Finally, suppose $\widetilde{S}_a \le -y/6$. Replace E_2 with $E_2' = \{\widetilde{S}_a \le -y/6, \widetilde{T}_a \le z\}$. Here we use the trivial bound $\mathbf{P}E_1 \le 1$ and deduce

$$\mathbf{P}\{T_n \le y, S_n = y - z, \widetilde{S}_a \le -y/6\} \le \mathbf{P}E_1\, \mathbf{P}E_2'\, \mathbf{P}\{E_3|E_1, E_2'\} \ll n^{-1/2}\mathbf{P}E_2'.$$

By Markov's inequality and Lemma 6.1 parts (1) and (2),

$$\mathbf{P}E_2' \le \mathbf{P}\{\widetilde{T}_a \le z\}\mathbf{P}\{\widetilde{S}_a \ge y/6|\widetilde{T}_a \le z\} \le \mathbf{P}\{\widetilde{T}_a \le z\}\frac{\mathbf{E}\{\widetilde{S}_a^2|\widetilde{T}_a \le z\}}{(y/6)^2} \ll \frac{\min(z+1,\sqrt{n})}{\sqrt{n}} \cdot \frac{n}{y^2}.$$

This completes the proof of (b). ☐

Combining Theorem 2 with Lemma 5.1 gives us useful bounds on $\widetilde{R}_n(y+\xi, y)$ when $\xi \ge 0$.

Theorem 3. *Assume (C) or (L), and $\beta_2 = 1$. Suppose $y \ge 0$ and $\xi \ge 0$. Then*

$$\widetilde{R}_n(y+\xi, y) \ll \frac{y+1}{n^{3/2}} \int_0^\infty (t+1) f(\xi+t)\, dt \ll \frac{y+1}{n^{3/2}}.$$

If $n \ge 3n_0 + 1$ and $y > \sqrt{n}$, then

$$\widetilde{R}_n(y+\xi, y) \ll \frac{1}{y^2} \int_0^\infty (t+1) f(\xi+t)\, dt + \frac{1 - F(\xi + y/2)}{n^{1/2}}.$$

Proof. Apply Lemma 5.1 and Theorem 2 (a) for the first part, and observe that the integral is $\le \mathbf{E}|X_1|$. For the second part, use Theorem 2 (b) for $t \le y/2$, and $\widetilde{R}_{n-1}(y-t, y) \ll n^{-1/2}$ for $t > y/2$. ☐

7. Proof of Theorem 1

We begin by proving a lemma which is of independent interest.

Lemma 7.1. *Assume $\beta_u < \infty$ for some $u \ge 2$, and $y \ge 0$. If (C) then*

$$\sum_{n=1}^\infty \int_0^\infty \xi^{u-2} \widetilde{R}_n(y+\xi, y)\, d\xi = O(1).$$

If (L) then

$$\sum_{n=1}^\infty \sum_{\substack{y+\xi \in \mathscr{L}_n \\ \xi \ge 0}} \xi^{u-2} \widetilde{R}_n(y+\xi, y) = O(1).$$

Proof. Assume (C). First,

$$\sum_{n=1}^{3n_0}\int_0^\infty \xi^{u-2}\widetilde{R}_n(y+\xi,y)\,d\xi \le \sum_{n=1}^{3n_0}\int_0^\infty \xi^{u-2}f_n(y+\xi)\,d\xi \ll \sum_{n=1}^{3n_0}\mathbf{E}|S_n|^{u-1}\ll 1.$$

By Theorem 3,

$$\begin{aligned}\sum_{n\ge 3n_0+1}\int_0^\infty \xi^{u-2}\widetilde{R}_n(y+\xi,y)\,d\xi &\ll \Bigg(\sum_{3n_0+1\le n\le y^2+1}\frac{1}{y^2}+\sum_{n>y^2+1}\frac{y+1}{n^{3/2}}\Bigg)\\ &\qquad\times\int_0^\infty (t+1)\int_0^\infty \xi^{u-2}f(\xi+t)\,d\xi\,dt\\ &\quad+\sum_{n\le y^2+1}\frac{1}{n^{1/2}}\int_0^\infty \xi^{u-2}\int_\xi^\infty f(v+y/2)\,dv\,d\xi\\ &\ll \mathbf{E}(|X_1|^u+|X_1|^{u-1})+(y+1)\int_0^\infty v^{u-1}f(v+y/2)\,dv\\ &\ll 1+\mathbf{E}|X_1|^u\ll 1.\end{aligned}$$

The proof when (L) holds is similar. □

Remark. A random walk $S_0, S_1, \dots$ with $\beta_2 = 1$ crosses the point y with probability 1. There is a unique n for which $T_{n-1} < y$ and $S_n \ge y$, and Lemma 7.1 states that $\mathbf{E}(S_n - y)^{u-2} = O(1)$.

We now prove Theorem 1 (again showing the details only for the case of (C) holding). It suffices to assume that n is sufficiently large. Let $n \ge 10n_0$ and put $x = y - z$. By Lemma 5.2 with $a = z$,

$$\begin{aligned}\widetilde{R}_n(x,y) &= f_n(x)-f_n(y+z)+\widetilde{R}_n(y+z,y)\\ &\quad+\int_0^\infty \sum_{k=1}^{n-1}\widetilde{R}_{n-k}(y+\xi,y)(f_k(z-\xi)-f_k(-z-\xi))\,d\xi.\end{aligned}\tag{7.1}$$

If β_u exists, where $3 < u \le 4$, then

$$\int_0^\infty (t+1)f(\xi+t)\,dt = \mathbf{P}\{X_1\ge\xi\}+\int_0^\infty \mathbf{P}\{X_1\ge \xi+t\}\,dt \ll \frac{1}{(\xi+1)^{u-1}}.$$

Therefore, by Theorem 3,

$$\widetilde{R}_n(y+\xi,y)\ll \frac{y+1}{n^{3/2}(1+\xi)^{u-1}}.\tag{7.2}$$

Let V_1 be the contribution to the integral in (7.1) from $1 \le k \le n_0$, let V_2 be the contribution from $n_0 + 1 \le k \le n/2$ and V_3 is the contribution from $n/2 < k \le n - 1$. By (7.2),

$$V_1 \ll \frac{y+1}{n^{3/2}}\sum_{k=1}^{n_0}\int_0^\infty f_k(z-\xi)+f_k(-z-\xi)\,d\xi = 2n_0\frac{y+1}{n^{3/2}}.\tag{7.3}$$

When $k \geq n_0 + 1$, Lemma 4.1 implies that

$$
\begin{aligned}
f_k(z-\xi) - f_k(-z-\xi) &= \frac{e^{-\frac{1}{2k}(z-\xi)^2}}{\sqrt{2\pi k}}\left(1 - e^{-2\xi z/k}\right) + O\left(\frac{1}{k^{(u-1)/2}}\right) \\
&\quad + O\left[\left(\frac{|z-\xi|}{k^{3/2}} + \frac{|z-\xi|^3}{k^{5/2}}\right)e^{-(z-\xi)^2/2k} + \left(\frac{z+\xi}{k^{3/2}} + \frac{(z+\xi)^3}{k^{5/2}}\right)e^{-(z+\xi)^2/2k}\right] \\
&\ll \frac{1}{k^{(u-1)/2}} + \frac{(z+1)(\xi+1)}{k^{3/2}}e^{-(z-\xi)^2/2k}.
\end{aligned} \tag{7.4}
$$

By (7.2), we have

$$
\begin{aligned}
V_2 &\ll \frac{y+1}{n^{3/2}} \sum_{n_0+1 \leq k \leq n/2} \int_0^\infty \frac{1}{k^{(u-1)/2}(\xi+1)^{u-1}} + \frac{z+1}{k^{3/2}(\xi+1)^{u-2}} e^{-(z-\xi)^2/2k}\, d\xi \\
&\ll \frac{y+1}{n^{3/2}}\left[1 + (z+1)\sum_{k \leq n/2} \frac{1}{k^{3/2}} \int_0^\infty \frac{1}{(\xi+1)^{u-2}} e^{-(z-\xi)^2/2k}\, d\xi\right].
\end{aligned}
$$

The integral on the right side is

$$
\begin{aligned}
&\leq e^{-z^2/8k} \int_0^{z/2} \frac{d\xi}{(\xi+1)^{u-2}} + \int_{-z/2}^\infty \frac{e^{-w^2/2k}}{(z+w)^{u-1}}\, dw \\
&\ll e^{-z^2/8k} + \min\left(\frac{1}{(z+1)^{u-3}}, \frac{k^{1/2}}{(z+1)^{u-2}}\right).
\end{aligned}
$$

Hence

$$
\begin{aligned}
V_2 &\ll \frac{y+1}{n^{3/2}}(z+1)\sum_{k=1}^\infty k^{-3/2}\left(e^{-z^2/8k} + \min\left(\frac{1}{(z+1)^{u-3}}, \frac{k^{1/2}}{(z+1)^{u-2}}\right)\right) \\
&\ll \frac{y+1}{n^{3/2}}(z+1)\left[\frac{1}{z+1} + \frac{1}{(z+1)^{u-2}}\sum_{k \leq z^2} \frac{1}{k}\right] \ll \frac{y+1}{n^{3/2}}.
\end{aligned} \tag{7.5}
$$

By Lemma 7.1 and (7.4),

$$
V_3 \ll \left(\frac{z+1}{n^{3/2}} + \frac{1}{n^{(u-1)/2}}\right)\sum_{j=1}^\infty \int_0^\infty (\xi+1)\widetilde{R}_j(y+\xi, y)\, d\xi \ll \frac{z+1}{n^{3/2}} + \frac{1}{n^{(u-1)/2}}. \tag{7.6}
$$

Putting together (7.1), (7.2), (7.3), (7.5) and (7.6), we arrive at

$$
\widetilde{R}_n(x,y) = f_n(x) - f_n(y+z) + O\left(\frac{y+z+1}{n^{3/2}} + \frac{1}{n^{(u-1)/2}}\right).
$$

Since $|x| \leq M\sqrt{n}$, Lemma 4.1 implies $f_n(x) \gg n^{-1/2}$ for sufficiently large n, the implied constant depending on the distribution of X_1 and also on M. Hence

$$
R_n(x,y) = 1 - \frac{f_n(y+z)}{f_n(x)} + O\left(\frac{y+z+1}{n} + \frac{1}{n^{(u-2)/2}}\right).
$$

Finally, by Lemma 4.1 again,

$$\begin{aligned}\frac{f_n(y+z)}{f_n(x)} &= e^{-\frac{1}{2n}((y+z)^2-x^2)} + O\left(\frac{|x|+y+z+1}{n} + \frac{1}{n^{(u-2)/2}}\right) \\ &= e^{-2yz/n} + O\left(\frac{y+z+1}{n} + \frac{1}{n^{(u-2)/2}}\right).\end{aligned}$$

Again the implied constant depends on M. This concludes the proof of Theorem 1.

Acknowledements. The author thanks Valery Nevzorov for suggesting to utilize the reflection principle in a form similar to that in Lemma 5.2. The author is grateful to the referees for carefully reading the paper and for several small corrections and suggestions.

REFERENCES

1. L. Addario-Berry and B. Reed, *Ballot theorems for random walks with finite variance*, preprint July, 2008. ArXiv math.PR/0802.2491.
2. M. Ahsanullah and V. B. Nevzorov, *Ordered random variables*, Nova Science Publishers Inc., Huntington, NY, 2001. MR MR2019107 (2004j:62002)
3. T. V. Arak, *The distribution of the maximum of the successive sums of independent random variables*, Teor. Verojatnost. i Primenen. **19** (1974), 257–277, Russian. English summary. English translation in Theor. Probability Appl. **19** (1974), 245–266.
4. M. Csörgő and P. Révész, *Strong approximations in probability and statistics*, Academic Press, 1981.
5. P. Erdős and M. Kac, *On certain limit theorems of the theory of probability*, Bull. Amer. Math. Soc. **52** (1946), 292–302. MR MR0015705 (7,459b)
6. K. Ford, *Du théorème de Kolmogorov sur les distributions empiriques à la théorie des nombres*, L'héritage de Kolmogorov en mathématiques, Editions Belin, Paris, 2004, (French), pp. 111–120.
7. ______, *From Kolmogorov's theorem on empirical distribution to number theory*, Kolmogorov's legacy in mathematics, Editions Belin / Springer-Verlag, Paris, 2007, (English), pp. 97–108.
8. ______, *The distribution of integers with a divisor in a given interval*, Ann. Math. **168** (2008), 367–433.
9. ______, *Sharp probability estimates for generalized Smirnov statistics*, Monatshefte Math. **153** (2008), 205–216.
10. B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Translated from the original 1949 Russian edition, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont., 1968. MR MR0233400 (38 #1722)
11. A. N. Kolmogorov, *Sulla determinazione empirica di una legge di distribuzione (on the empirical determination of a distribution law)*, Giorn. Ist. Ital. Attuar. **4** (1933), 83–91.
12. M. Kozlov, *On the asymptotic probability of nonextinction for a critical branching process in a random environment*, Teor. Verojatnost. i Primen. **21** (1976), 813–825, Russian. English summary. English translation in Theor. Probability Appl. **21** (1976), 791–804.
13. S. V. Nagaev, *The rate of convergence of the distribution of the maximum of sums of independent random variables*, Teor. Verojatnost. i Primenen **15** (1970), 320–326, Russian. English Translation in Theor. Probability Appl. **15** (1970), 309–314.
14. R. Pemantle and Y. Peres, *Critical random walk in random environment on trees*, Ann. Prob. **23 (1)** (1995), 105–140.
15. W. Philipp, *Invariance principles for independent and weakly dependent random variables*, Dependence in Probability and Statistics (Oberwolfach, 1985), Progr. Probab. Statist. **11**, Birkhäuser, Boston, MA, 1986, pp. 225–268.
16. A. Rényi, *On the theory of order statistics*, Acta Math. Acad. Sci. Hung. **4** (1953), 191–232.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, IL 61801
E-mail address: ford@math.uiuc.edu