

MONTE CARLO ALGORITHMS FOR FINDING THE MAXIMUM OF A RANDOM WALK WITH NEGATIVE DRIFT

LUDWIG BARINGHAUS * ** AND
RUDOLF GRÜBEL, * *** *Universität Hannover*

Abstract

We discuss two Monte Carlo algorithms for finding the global maximum of a simple random walk with negative drift. This problem can be used to connect the analysis of random input Monte Carlo algorithms with ideas and principles from mathematical statistics.

Keywords: Brownian motion with drift; conditioning; efficiency; fast Fourier transform; ladder variable; Markov chain; randomization

2000 Mathematics Subject Classification: Primary 60G50; 68Q25

1. Introduction

Let X_i , $i \in \mathbb{N}$, be a sequence of independent random variables on some probability space (Ω, \mathcal{A}, P) , with

$$P(X_i = 1) = 1 - P(X_i = -1) = p$$

for all $i \in \mathbb{N}$. The sequence $(S_k)_{k \in \mathbb{N}_0}$ with $S_0 = 0$ and $S_k = \sum_{i=1}^k X_i$ for all $k \in \mathbb{N}$ is a simple random walk on the integers, with parameter p . Let

$$M_n := \max_{0 \leq k \leq n} S_k \quad \text{and} \quad M := \sup_{k \in \mathbb{N}_0} S_k$$

be the maximum of the initial segment of length $n \in \mathbb{N}_0$ and the global supremum, respectively. The random walk is said to be symmetric if $p = \frac{1}{2}$ and asymmetric or biased if $p \neq \frac{1}{2}$; we will assume throughout that $0 < p < \frac{1}{2}$. The strong law of large numbers then implies that $S_k \rightarrow -\infty$ almost surely as $k \rightarrow \infty$. In particular, M is then finite with probability 1 and there is an \mathbb{N}_0 -valued random variable τ such that

$$M(\omega) = S_{\tau(\omega)}(\omega) \quad \text{and} \quad S_k(\omega) < M(\omega) \quad \text{for } k < \tau(\omega), \quad (1)$$

for P -almost all $\omega \in \Omega$.

In this paper, we consider the problem of finding M for a given path $k \mapsto S_k(\omega)$. The authors believe that this problem is interesting in its own right; it is related, via the comparison of strings, to a practical problem from molecular biology. Furthermore, it can be used to connect the probabilistic theory of algorithms to ideas and concepts from mathematical statistics such as conditioning, randomization, and efficiency.

Received 19 March 2004.

* Postal address: Institut für Mathematische Stochastik, Universität Hannover, Postfach 60 09, D-30060 Hannover, Germany.

** Email address: lbaring@stochastik.uni-hannover.de

*** Email address: rgrubel@stochastik.uni-hannover.de

In the symmetric case, Odlyzko (1995) considered the paths of some fixed length n as an input model for an unknown environment, and investigated the number of S_k , $1 \leq k \leq n$, that have to be examined in order to determine M_n . In the worst case this number grows as $n/2$ as $n \rightarrow \infty$, but there is an algorithm that needs only about $c_0\sqrt{n}$ such probes on average. Chassaing (1999) obtained a similar result for the biased case, in which about $c_1 \log n$ values turn out to be sufficient on average. Recently, Chassaing *et al.* (2003) obtained an associated asymptotic optimality result in distribution; Odlyzko (1995) and Chassaing (1999) gave such results for the asymptotic average case behaviour.

The main interest in the papers mentioned in the previous paragraph is in algorithms that determine with certainty the maximum, M_n , of an initial segment of length n of the path. For the global maximum, M , this is not possible and we have to content ourselves with a Monte Carlo algorithm, i.e. a procedure that returns the correct value only with some probability $1 - \alpha$. The error margin, α , is greater than 0 but can be chosen to be arbitrarily small at the cost of an increase in running time. Odlyzko (1995) also considered Monte Carlo algorithms for M_n in the symmetric case; he showed that a further reduction from $n^{1/2}$ to $(\log n)^2$ is possible if an error of magnitude n^{-10} is accepted.

In the next section, we describe the two algorithms and compare them from a theoretical point of view. Section 3 contains the main results, and the proofs are presented in Section 4. In Section 5, we collect some comments on possible extensions of the results and variants of the proofs. Some numerical aspects are discussed in an appendix.

2. The algorithms

We now describe the two algorithms that we will analyse and compare in the sequel. As we regard p as known and α as given, a very simple Monte Carlo algorithm consists of examining the first N_1 values, where N_1 is the $(1 - \alpha)$ th quantile of the distribution of τ , with τ as in (1), i.e.

$$N_1 \equiv N_1(\alpha, p) := \inf\{k \in \mathbb{N}_0 : P(\tau \leq k) \geq 1 - \alpha\}.$$

We will refer to this as algorithm (A1) below. Algorithm (A1) returns M_{N_1} as the prospective global maximum.

For the second method, algorithm (A2), we step through the values of the random walk one by one and keep track of the drop,

$$Z_k := \max\{S_0, \dots, S_k\} - S_k, \quad k = 0, 1, \dots, \tag{2}$$

of the random walk below its current maximum. Let $T_C^Z := \inf\{k \in \mathbb{N}_0 : Z_k = C\}$ be the time of the first visit to C of the process Z . Suppose now that we stop at T_C^Z and return $M_{T_C^Z}$ as the likely global maximum. For any stopping time, T , with respect to the natural filtration associated with $(S_k)_{k \in \mathbb{N}_0}$, the post- T process $(S_{T+k} - S_T)_{k \in \mathbb{N}_0}$ is again a random walk with parameter p . This method therefore gives the correct result with probability $P(M \leq C)$. It is known that the global maximum, M , of a simple random walk with parameter p has a geometric distribution with parameter $1 - p/q$, where $q := 1 - p$, i.e.

$$P(M = k) = \left(\frac{p}{q}\right)^k \left(1 - \frac{p}{q}\right) \quad \text{for all } k \in \mathbb{N}_0$$

(see also the proof of Theorem 1). Hence, if we choose

$$C \equiv C(\alpha, p) := \left\lceil \frac{\log \alpha}{\log(p/q)} \right\rceil - 1, \tag{3}$$

then we arrive at an algorithm with error probability not exceeding α . (We write $\lfloor x \rfloor$ for the greatest integer less than or equal to x and $\lceil x \rceil$ for the least integer greater than or equal to x .) This takes a random number, $N_2 \equiv N_2(\alpha, p) = T_{C(\alpha, p)}^Z$, of steps.

According to Korf *et al.* (2003), algorithm (A2) was used in an early version of BLAST, the basic local alignment search tool in molecular biology. To evaluate the similarity between two strings from the alphabet $\Sigma = \{A, G, C, T\}$, we scan the two sequences from left to right, increasing a score by 1 if the letters in corresponding positions coincide, and subtracting 1 otherwise. In the basic model that is used to assess the significance of similarities, it is assumed that the individual strings are obtained by sampling uniformly from the alphabet Σ . This fits into the above framework, with $p = \frac{1}{4}$, if we wish to find the prefix with maximum score and if we consider the sequences to be infinitely long.

The traditional connection between the theory of algorithms and probability theory, as exemplified by Knuth (2000), is the average case analysis of a given algorithm under the assumption of random input. A standard reference for Monte Carlo and Las Vegas algorithms, and randomized algorithms in general, is the book by Motwani and Raghavan (1995). The study of Monte Carlo algorithms with random input seems to be less developed, but offers some interesting connections to various areas of probability theory and statistics. For example, with random input and an algorithm that gives the correct answer without certainty, but only with some probability (of at least) $1 - \alpha$, it makes sense to consider the conditional probability of a correct answer given the ‘data’, i.e. the part of the input of the algorithm on which the output is based. In the case of algorithm (A2), the argument given above for the selection of $C \equiv C(\alpha, p)$ shows that the conditional probability of a correct answer is at least $1 - \alpha$, regardless of the actual path-segment observed, whereas this is not the case with algorithm (A1) (or with Odlyzko’s Monte Carlo algorithm in the symmetric case). Consider, for example, a situation in which algorithm (A1) is used on an initial segment of length n , and where M_n happens to be equal to S_n : the algorithm would return M_n despite the fact that the conditional probability of a correct answer, i.e. of M_n being equal to the global maximum M , is much smaller than $1 - \alpha$. We see here a possible relationship with the conditionality principle from theoretical statistics; see, e.g. Section 2.3(iii), pp. 38–39, of Cox and Hinkley (1974).

While such basic issues, and similarly optimality results, are certainly most interesting, we will not pursue these any further, but will concentrate on the quantitative analysis of the algorithms described above and their quantitative comparison. (An exception is Remark 1, below, where we briefly return to a more abstract point of view.) We will see, for example, that two different notions of efficiency both lead to the conclusion that algorithm (A2) is about four times better than algorithm (A1) (in terms of the expected length of the segment of the random walk expected) if α is small and p is close to $\frac{1}{2}$.

3. Main results

We first consider some explicit results for given α and p , and then turn to asymptotics. For algorithm (A1), the distribution of τ , as defined in (1), is important.

Theorem 1. *The probability-generating function and the probability mass function of τ are given by*

$$g_\tau(z) := \sum_{k=0}^{\infty} \mathbb{P}(\tau = k) z^k = \frac{2(q-p)z}{2qz - 1 + \sqrt{1 - 4pqz^2}} \quad (4)$$

and

$$\begin{aligned}
 P(\tau = 0) &= 1 - \frac{p}{q}, & P(\tau = 1) &= p \left(1 - \frac{p}{q}\right), \\
 P(\tau = k) &= \left(1 - \frac{p}{q}\right) \left\{ p - \sum_{j=1}^{\lfloor k/2 \rfloor} \binom{2(j-1)}{j-1} \frac{1}{j} (pq)^j \right\}, & k &= 2, 3, \dots,
 \end{aligned}
 \tag{5}$$

respectively.

In contrast to the situation in algorithm (A1), the length of the initial segment of the random walk needed for algorithm (A2) is now a random variable. Recall that $N_2(\alpha, p)$ is equal to an entry time of the Z -process.

Theorem 2. Let g_C be the generating function associated with T_C^Z . Then, with $d(z) := \sqrt{1 - 4pqz^2}$,

$$g_C(z) = \frac{2(2qz)^C d(z)}{(d(z) + 1 - 2pz)(1 + d(z))^C + (d(z) - 1 + 2pz)(1 - d(z))^C}.
 \tag{6}$$

We can use (4) or (5) to calculate the quantiles needed for algorithm (A1); see the appendix. In particular, for $p = \frac{1}{4}$ we obtain

$$P(\tau \leq 22) = 0.998\,796\dots, \quad P(\tau \leq 23) = 0.999\,023\dots$$

Hence, $\alpha = 0.001$ would lead to $N_1 = 23$, but the true error probability would then be smaller than 0.001. A similar situation arises in connection with statistical tests, where we start with a significance level α as an upper bound for the type-I error of wrongly rejecting the hypothesis. Translating the randomization device familiar in the latter context to the present situation would lead us to choose N_1 at random, independently of $(S_k)_{k \in \mathbb{N}_0}$, with

$$P(N_1 = 22) = 1 - P(N_1 = 23) = \frac{0.001 - P(\tau > 23)}{P(\tau = 23)} = 0.102\,164\dots$$

This results in a procedure that fully exploits the allowed error probability, 0.001.

Similarly, we can use (6) to obtain the probability mass function for T_C^Z . With $p = \frac{1}{4}$ and $\alpha = 0.001$, (3) gives $C(\alpha, p) = 6$, but then the true error probability would be

$$P(M > 6) = \left(\frac{p}{q}\right)^7 = \frac{1}{3^7} = 0.000\,457\dots$$

With $C = 5$ we obtain $0.001\,371\dots$, so a randomized variant exploiting the full error margin would have $C = 5$ with probability β and $C = 6$ with probability $1 - \beta$, where

$$\beta = \frac{3^7 \cdot 10^{-3} - 1}{2} = 0.5935.$$

Randomization also makes it possible to compare the algorithms on an equal basis, as the nonrandomized versions may differ substantially in their actual errors. In our numerical example, algorithm (A1) uses about 98% of α (in the sense that the actual error probability is only $1 - 0.999\,023\dots$, rather than 0.001), and algorithm (A2) only about 46%; for the randomized algorithms we obtain

$$E[N_1] = 22.8978\dots, \quad E[N_2] = 9.816, \quad \text{and} \quad P(N_2 > E[N_1]) = 0.021\,685\dots$$

In order to arrive at more general conclusions, we consider the behaviour of N_1 and N_2 for values of $p = \frac{1}{2} - \varepsilon$, $\varepsilon > 0$, close to the critical case, $p = \frac{1}{2}$, and for values of the error bound, α , close to 0.

Theorem 3. *Let $N_1(\alpha, p)$ be the number of steps required by algorithm (A1) for a random walk with parameter p if α is the prescribed upper bound for the error probability. Then*

$$\lim_{\alpha \downarrow 0} \frac{N_1(\alpha, p)}{\log \alpha} = \frac{1}{\log(2\sqrt{pq})} \quad \text{for all } p \in (0, \frac{1}{2}) \tag{7}$$

and

$$\lim_{\varepsilon \downarrow 0} \varepsilon^2 N_1(\alpha, \frac{1}{2} - \varepsilon) = q_{1-\alpha} \quad \text{for all } \alpha \in (0, 1), \tag{8}$$

where $q_{1-\alpha}$ denotes the $(1 - \alpha)$ th quantile of the distribution with Laplace transform

$$\phi(t) = \frac{4}{2 + \sqrt{2t + 4}} \quad \text{for all } t \geq 0 \tag{9}$$

and density

$$f(x) = \sqrt{\frac{8}{\pi x}} \exp(-2x) \left(1 - 2\sqrt{x} \exp(2x) \int_{2\sqrt{x}}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy \right), \quad x > 0. \tag{10}$$

The function

$$z \mapsto \exp\left(\frac{z^2}{2}\right) \int_z^{\infty} \exp\left(-\frac{y^2}{2}\right) dy$$

is known as Mills' ratio; see, e.g. Stuart and Ord (1987, Section 5.38). By using the classical expansions for the tail of the standard normal distribution, such as given in Problem 1, p. 193, of Feller (1968), we can easily derive approximations to f , the corresponding distribution function F , and the quantile function $q_{1-\alpha} = F^{-1}(1 - \alpha)$.

Theorem 4. *Let $N_2(\alpha, p)$ be the number of steps required by algorithm (A2) for a random walk with parameter p if α is the prescribed upper bound for the error probability and $C \equiv C(\alpha, p)$ is chosen according to (3). Then*

$$E[N_2(\alpha, p)] = \frac{C - (p/q)^C}{q - p} - \frac{p - q(p/q)^C}{(q - p)^2}.$$

In particular,

$$\lim_{\alpha \downarrow 0} \frac{E[N_2(\alpha, p)]}{\log \alpha} = \frac{1}{(1 - 2p) \log(p/q)} \quad \text{for all } p \in (0, \frac{1}{2}) \tag{11}$$

and

$$\lim_{\varepsilon \downarrow 0} \varepsilon^2 E[N_2(\alpha, \frac{1}{2} - \varepsilon)] = \frac{-\log \alpha - (1 - \alpha)}{8} \quad \text{for all } \alpha \in (0, 1). \tag{12}$$

Furthermore, $\varepsilon^2 N_2(\alpha, \frac{1}{2} - \varepsilon)$ converges in distribution, as $\varepsilon \downarrow 0$, to a probability measure with Laplace transform

$$\psi_\alpha(t) = \frac{2\eta(t)}{(2 + \eta(t))\alpha^{1/2-\eta(t)/4} - (2 - \eta(t))\alpha^{1/2+\eta(t)/4}}, \quad \eta(t) := \sqrt{4 + 2t}. \tag{13}$$

On the basis of Theorems 3 and 4, we can compare the behaviours of the algorithms for small error probabilities and, separately, for parameters of the underlying random walk that are close to the critical value $\frac{1}{2}$. Again, this can be related to a statistical situation. Suppose that we wish to test a simple hypothesis, H_0 , that $\theta = \theta_0$, against a simple alternative, H_1 , that $\theta = \theta_1$, where the parameter θ determines the distribution of the data. Let α be the significance level. Suppose that we assign a fixed upper bound, $\beta \in (0, 1)$, for the type-II error probability. Then, the sample size required by a given procedure is a function of the significance level and the parameters, implying that two such procedures can be compared using the ratio of the corresponding sample sizes. In typical cases, the limit of this ratio exists as $\alpha \downarrow 0$, and does not depend on β . The limit is called the Bahadur efficiency of the two competing tests. For a fixed α and as $\theta_1 \rightarrow \theta_0$, in which limit it becomes more and more difficult to distinguish the hypothesis from the alternative, the limit of the ratio of the sample sizes again exists in typical cases. This limit is called the Pitman efficiency of the two competing tests. For details on these efficiency concepts, we refer the reader to Chapter 14, pp. 201–205, of van der Vaart (1998).

By translating significance levels into error bounds and sample size into running time, which here is the length of the initial segment of the random walk needed by the respective algorithm, we can similarly introduce efficiency concepts for Monte Carlo algorithms with random input. In the present situation, from (11) and (7) we obtain

$$\text{Eff}_B(A1, A2)(p) := \lim_{\alpha \downarrow 0} \frac{E[N_2(\alpha, p)]}{N_1(\alpha, p)} = \frac{\log(2\sqrt{pq})}{(1 - 2p) \log(p/q)}$$

for the ‘Bahadur efficiency’ of our algorithms, and, from (12) and (8),

$$\text{Eff}_P(A1, A2)(\alpha) := \lim_{p \uparrow 1/2} \frac{E[N_2(\alpha, p)]}{N_1(\alpha, p)} = \frac{-\log \alpha - (1 - \alpha)}{8q_{1-\alpha}}$$

for the ‘Pitman efficiency’ of our algorithms. The graphs of these two functions are given in Figure 1. It appears that algorithm (A2) performs better. In particular, for small error probabilities, algorithm (A2) needs only about one-half as many observations on average as algorithm (A1) for p close to 0, and about one-quarter as many for p close to $\frac{1}{2}$. Note, however, that there is no upper bound on the running time of algorithm (A2), in contrast to algorithm (A1). Indeed, according to the limit distribution in Theorem 4, it follows that $P(N_2(\alpha, p) > N_1(\alpha, p))$ converges to a positive limit as $p \uparrow \frac{1}{2}$. (To see that the limit cannot be 0, we note that the Laplace transform of a distribution with bounded support has to be an entire function.)

A simple calculation gives

$$\lim_{p \uparrow 1/2} \text{Eff}_B(A1, A2)(p) = \frac{1}{4}.$$

To obtain the limit of $\text{Eff}_P(A1, A2)$ as $\alpha \downarrow 0$, we use the classical expansion for Mills’ ratio mentioned above to derive

$$f(x) \sim \frac{1}{\sqrt{2\pi}} x^{-3/2} \exp(-2x), \quad x \rightarrow \infty,$$

from (10). The tail of the distribution satisfies

$$\int_x^\infty f(t) dt \sim \frac{1}{\sqrt{8\pi}} x^{-3/2} \exp(-2x) \quad \text{as } x \rightarrow \infty.$$

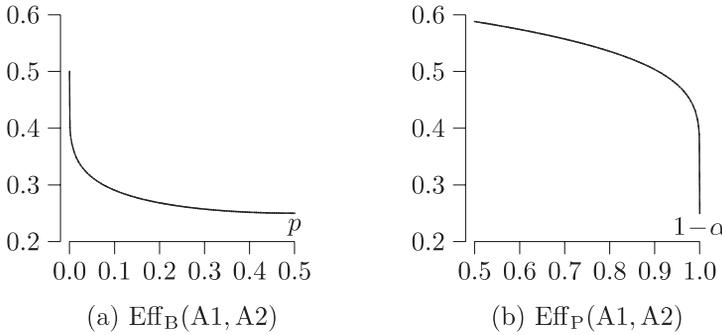


FIGURE 1: The efficiencies of algorithms (A1) and (A2).

From this we deduce that both $q_{1-\alpha} \sim -\frac{1}{2} \log \alpha$ and

$$\lim_{\alpha \downarrow 0} \text{Eff}_P(A1, A2)(\alpha) = \frac{1}{4}.$$

It seems quite remarkable that the two limits coincide, as they rely on the different concepts of tail behaviour and distributional limits (see also Remark 2, below). Again, such an equality of limiting efficiencies is well known in mathematical statistics: for typical testing problems, it can be verified that the limiting approximate Bahadur efficiency, as $\theta_1 \rightarrow \theta_0$, and the limiting Pitman efficiency, as $\alpha \downarrow 0$, coincide; see, e.g. Kallenberg and Koning (1995).

4. Proofs

Proof of Theorem 1. As in Chassaing (1999), we use the classical ladder method. Let $T = T_1^S$ be the first strictly ascending ladder time, i.e. the time needed by the random walk $(S_n)_{n \in \mathbb{N}_0}$ to arrive at 1; since $p < \frac{1}{2}$ this may never happen. The probability-generating function associated with the defective distribution of T is known to be

$$g(z) = \frac{1 - \sqrt{1 - 4pqz^2}}{2qz}; \tag{14}$$

see, e.g. Feller (1968, p. 272). With $(T, 1)$ as the new origin we obtain a random walk again, identical in distribution to the process we started with and independent of the segment up to T of the latter. The first maximum position can therefore be regarded as a geometric random sum of summands that are independent and equal in distribution to T , conditional on $T < \infty$.

From (14) it follows that $P(T < \infty) = P(\tau > 0) = g(1) = p/q$. We now use the familiar fact that the generating function of a random sum is the composition of the generating function of the distribution of the number of terms and the generating function associated with the distribution of the individual summands.

In order to obtain the associated probability mass function, we note that the generating function can be written as

$$g_\tau(z) = \left(1 - \frac{p}{q}\right) \frac{1 - 2qz + \sqrt{1 - 4pqz^2}}{2(1 - z)}, \quad z \in (0, 1).$$

By expanding the term on the right-hand side in powers of z and using

$$\begin{aligned} (1 - 4pqz^2)^{1/2} &= \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} (-1)^k 2^{2k} (pq)^k z^{2k} \\ &= 1 - 2 \sum_{k=1}^{\infty} \binom{2(k-1)}{k-1} \frac{1}{k} (pq)^k z^{2k}, \end{aligned}$$

we arrive at

$$\begin{aligned} g_{\tau}(z) &= \left(1 - \frac{p}{q}\right) \left\{ 1 - qz - \sum_{k=1}^{\infty} \binom{2(k-1)}{k-1} \frac{1}{k} (pq)^k z^{2k} \right\} (1-z)^{-1} \\ &= \left(1 - \frac{p}{q}\right) \left\{ 1 + pz + \sum_{k=2}^{\infty} \left[p - \sum_{j=1}^{\lfloor k/2 \rfloor} \binom{2(j-1)}{j-1} \frac{1}{j} (pq)^j \right] z^k \right\}, \end{aligned}$$

which proves the assertion.

Proof of Theorem 2. The process $Z = (Z_k)_{k \in \mathbb{N}_0}$, defined in (2), is easily seen to be a Markov chain with state space \mathbb{N}_0 , starting point 0, and transition probabilities

$$\begin{aligned} P(Z_{k+1} = 0 \mid Z_k = 0) &= p, \\ P(Z_{k+1} = i - 1 \mid Z_k = i) &= p \quad \text{for all } i \in \mathbb{N}, \\ P(Z_{k+1} = i + 1 \mid Z_k = i) &= q \quad \text{for all } i \in \mathbb{N}_0. \end{aligned} \tag{15}$$

The associated transition matrix P defines a linear operator that acts on sequences $(a(i))_{i \in \mathbb{N}_0}$ of real numbers as follows: $Pa = (b(i))_{i \in \mathbb{N}_0}$, with

$$b(0) = pa(0) + qa(1) \quad \text{and} \quad b(i) = pb(i-1) + qb(i+1) \quad \text{for all } i \in \mathbb{N}.$$

We now fix some $\lambda \geq 1$ and consider the equation $Pb = \lambda b$, where we assume that $b(0) = 1$. This leads to a simple difference equation that is uniquely solved by

$$b_{\lambda}(i) = \left(\frac{1}{2} + \frac{\lambda - 2p}{2c(\lambda)}\right) \left(\frac{\lambda + c(\lambda)}{2q}\right)^i + \left(\frac{1}{2} - \frac{\lambda - 2p}{2c(\lambda)}\right) \left(\frac{\lambda - c(\lambda)}{2q}\right)^i,$$

where $c(\lambda) = \sqrt{\lambda^2 - 4pq}$. Since $\lambda \geq 1$ and $p < \frac{1}{2}$ we always have $c(\lambda) > 0$. Dropping λ temporarily from the notation, we let $f(n, i) := \lambda^{-n} b_{\lambda}(i)$ and $Y_n := f(n, Z_n)$. With $\mathcal{F}_n := \sigma\{X_0, \dots, X_n\}$, we then obtain

$$\begin{aligned} E[Y_{n+1} \mid \mathcal{F}_n] &= \lambda^{-n-1} E[b_{\lambda}(Z_{n+1}) \mid Z_n] \\ &= \lambda^{-n-1} (Pb_{\lambda})(Z_n) \\ &= \lambda^{-n} b_{\lambda}(Z_n) \\ &= Y_n, \end{aligned}$$

meaning that $(Y_n, \mathcal{F}_n)_{n \in \mathbb{N}_0}$ is a martingale. Now let $T = T_C^Z$ be the time of the first visit to $C \in \mathbb{N}$ of the process Z . The optional sampling theorem (see, e.g. Theorem 10.10(a)(ii), p. 100, of Williams (1991)), gives $E[Y_T] = E[Y_0] = 1$; clearly T is finite and $Z_T \equiv C$. Together these imply that $g_C(1/\lambda) = E[\lambda^{-T_C}] = 1/b_{\lambda}(C)$.

Proof of Theorem 3. To prove the first assertion, we introduce the independent random variables N, T_1, T_2, \dots , where N has the geometric distribution

$$P(N = n) = \left(1 - \frac{p}{q}\right) \left(\frac{p}{q}\right)^n, \quad n \in \mathbb{N}_0,$$

and the distribution of T_r is the first passage distribution through r of the random walk with parameter p , conditional on the event that a passage through r exists. In terms of ladder variables, T_r is equal in distribution to the r th strictly ascending ladder epoch, provided that this is finite, and N is equal in distribution to the total number of strictly ascending ladder variables, which in the present context is equal to the global supremum. In particular, $P(\tau > n) = P(T_N > n)$ for each $n \in \mathbb{N}_0$. The distribution of $T_r, k \geq r$, is given by

$$\begin{aligned} P(T_r = 2k - r) &= \left(\frac{1}{2p}\right)^r \frac{r}{2k - r} \binom{2k - r}{k} 2^r (pq)^k \\ &= \left(\frac{1}{2p}\right)^r \frac{r}{2k - r} \frac{k(k - 1) \cdots (k - r + 1)}{2k(2k - 1) \cdots (2k - r + 1)} \binom{2k}{k} 2^{-2k} 2^r (4pq)^k; \end{aligned}$$

see, e.g. Feller (1968, pp. 274–275). By Stirling’s formula, there exists a positive constant c_0 such that

$$P(T_1 > n) > c_0 n^{-3/2} (2\sqrt{pq})^n \quad \text{for all } n \in \mathbb{N}.$$

Setting $c_1 = c_0(1 - p/q)(p/q)$, we see that

$$P(T_N > n) \geq P(N = 1, T_1 > n) > c_1 n^{-3/2} (2\sqrt{pq})^n \quad \text{for all } n \in \mathbb{N}. \tag{16}$$

Because

$$\frac{k(k - 1) \cdots (k - r + 1)}{2k(2k - 1) \cdots (2k - r + 1)} \leq 2^{-r},$$

it follows that

$$\begin{aligned} P(T_r > n) &= \sum_{k > (n+r)/2} P(T = 2k - r) \\ &\leq \sum_{k > (n+r)/2} \left(\frac{1}{2p}\right)^r \frac{r}{2k - r} \binom{2k}{k} 2^{-2k} (4pq)^k \\ &\leq \frac{r}{n} \left(\frac{1}{2p}\right)^r \sum_{k > (n+r)/2} \binom{2k}{k} 2^{-2k} (4pq)^k. \end{aligned}$$

By applying Stirling’s formula again, we obtain a constant c_2 such that

$$\begin{aligned} P(T_r > n) &\leq c_2 r \left(\frac{1}{2p}\right)^r (2\sqrt{pq})^r n^{-3/2} (2\sqrt{pq})^n \\ &= c_2 r \left(\sqrt{\frac{q}{p}}\right)^r n^{-3/2} (2\sqrt{pq})^n. \end{aligned}$$

This implies that

$$\begin{aligned} P(T_N > n) &= \sum_{r=1}^n P(N = r, T_r > n) + P(N > n) \\ &\leq c_2 \left(1 - \frac{p}{q}\right) \sum_{r=1}^n r \left(\sqrt{\frac{p}{q}}\right)^r n^{-3/2} (2\sqrt{pq})^n + \left(\frac{p}{q}\right)^n. \end{aligned} \tag{17}$$

Because $p/q \leq 2\sqrt{pq}$, we therefore have

$$P(T_N > n) \leq c_3(2\sqrt{pq})^n \quad \text{for all } n \in \mathbb{N},$$

for some positive constant c_3 . Taken together, (16) and (17) imply that

$$\frac{\log \alpha - \log c_1}{\log(2\sqrt{pq})} \leq N_1(\alpha, p) \leq \frac{\log \alpha - \log c_3}{\log(2\sqrt{pq})},$$

which proves the first assertion.

To prove the second statement, we use Theorem 1 to obtain the Laplace transform, ϕ_ε , associated with $\varepsilon^2\tau_\varepsilon$, where τ_ε denotes the first time that the random walk with parameter $\frac{1}{2} - \varepsilon$ attains its global maximum. We find that

$$\phi_\varepsilon(t) = g_{\tau_\varepsilon}(e^{-\varepsilon^2 t}) = \frac{4\varepsilon e^{-\varepsilon^2 t}}{(1 + 2\varepsilon)e^{-\varepsilon^2 t} - 1 + \sqrt{1 - (1 - 4\varepsilon^2)e^{-2\varepsilon^2 t}}}.$$

Some straightforward calculations show that $\lim_{\varepsilon \downarrow 0} \phi_\varepsilon(t) = \phi(t)$ for all $t \geq 0$. By the continuity theorem for Laplace transforms (see, e.g. Feller (1971, p. 431)), this implies that $\varepsilon^2\tau_\varepsilon$ converges in distribution, as $\varepsilon \downarrow 0$, to a random variable Z with $\phi(t) = E[e^{-tZ}]$ for all $t \geq 0$.

To obtain the density of the distribution with Laplace transform given by (9), we note that ϕ can alternatively be written as

$$\phi(t) = \begin{cases} 1, & t = 0, \\ 2 \frac{\sqrt{t/2 + 1} - 1}{t/2}, & t > 0. \end{cases}$$

By using the identity

$$\int_0^\infty \exp(-tx) \frac{1}{\sqrt{2x}} \exp\left(-\frac{x}{2}\right) D_{-2}(\sqrt{2x}) \, dx = \sqrt{\frac{\pi}{2}} \frac{\sqrt{t+1} - 1}{t},$$

valid for $t > 0$, where

$$D_{-2}(z) = \exp\left(-\frac{z^2}{4}\right) \left(1 - z \exp\left(\frac{z^2}{2}\right) \int_z^\infty \exp\left(-\frac{y^2}{2}\right) \, dy\right)$$

is a parabolic cylinder function (see Gradstein and Ryshik (1981, p. 276, Equation 7.725.1), we see that f as given in (10) is the density of the distribution with Laplace transform ϕ .

By using the connection with Mills' ratio mentioned in Section 3, we can show that this density is strictly greater than 0 on the whole range, $(0, \infty)$, of x . Hence, the convergence in distribution implies the convergence of the quantiles. The second formula in the theorem now follows by recalling the definition of N_1 as a quantile of the distribution of τ_ε .

Proof of Theorem 4. We could of course use Theorem 2 together with the relation

$$E[N_2(\alpha, p)] = g'_C(1),$$

but we give a direct argument instead, avoiding a tedious calculation (or recourse to computer algebra).

As in the proof of Theorem 2, we consider $Z = (Z_k)_{k \in \mathbb{N}_0}$ to be a Markov chain with transition matrix P ; see (15). We write $E_j[\cdot]$ for expectation with respect to starting point $j \in \mathbb{N}_0$. Let $D = \{0, 1, \dots, C - 1\}$ and let $T = T_C^Z$ be the hitting time of $D^c = \{C, C + 1, \dots\}$. Define a function h on the state space \mathbb{N}_0 by $h(j) = E_j[T]$ for $j \in \mathbb{N}_0$. We are interested in $E[N_2]$, which is equal to $E_0[T]$. Note that h is finite and that $h(j) = 0$ for each $j \geq C$. By applying the maximum principle (see, e.g. Brémaud (1999, p. 181)), we deduce that h is the unique solution to

$$h = \begin{cases} Ph + 1, & \text{on } D, \\ 0, & \text{on } D^c. \end{cases}$$

This is a system of difference equations,

$$\begin{aligned} ph(0) + qh(1) + 1 &= h(0), \\ ph(i - 1) + qh(i + 1) + 1 &= h(i), \quad i = 1, \dots, C - 2, \\ ph(C - 2) + 1 &= h(C - 1). \end{aligned}$$

By setting $u_j = h(j) - h(j - 1)$, $j \geq 1$, we obtain

$$u_{j+1} = \frac{p}{q}u_j - \frac{1}{q},$$

which, after iteration and using $u_1 = -1/q$, leads to

$$u_j = -\frac{1 - (p/q)^j}{q - p} \quad \text{for all } j \in \mathbb{N}.$$

Because $u_C = -h(C - 1)$, $h(0) = h(C - 1) - (u_1 + \dots + u_{C-1})$, and

$$\begin{aligned} -(u_1 + \dots + u_{C-1}) &= \frac{C - 1}{q - p} - \frac{1}{q - p} \sum_{j=1}^{C-1} \left(\frac{p}{q}\right)^j \\ &= \frac{C - 1}{q - p} - \left(\frac{1}{q - p}\right)^2 \left(p - q\left(\frac{p}{q}\right)^C\right), \end{aligned}$$

we obtain

$$E[N_2] = h(0) = \frac{C - (p/q)^C}{q - p} - \frac{p - q(p/q)^C}{(q - p)^2}.$$

The second and the third assertions follow by straightforward calculations, using

$$\lim_{\alpha \downarrow 0} \frac{C(\alpha, p)}{\log \alpha} = \frac{1}{\log p/q}, \quad \lim_{\varepsilon \downarrow 0} \varepsilon C(\alpha, \frac{1}{2} - \varepsilon) = -\frac{\log \alpha}{4}.$$

Finally, to prove (13) we proceed as in the corresponding part of the proof of Theorem 3, now using (6) instead of (4).

5. Final remarks

Remark 1. We could of course combine algorithm (A1) with the ideas of Odlyzko (1995) and Chassaing (1999), probing only a subset of the first N_1 values. For algorithm (A2) such a reduction seems not to be possible. On the other hand, if we insist on a procedure that satisfies

the error bound conditionally on the input, as explained at the end of Section 2, and if we consider the values S_0, S_1, S_2, \dots as arriving successively, then algorithm (A2) is easily seen to be optimal within the class of strategies that can be realized as stopping times with respect to the filtration $(\sigma\{S_0, \dots, S_n\})_{n \in \mathbb{N}_0}$.

Remark 2. The explicit formula (4) in Theorem 1 made it possible to obtain the distributional asymptotics of $\varepsilon^2 \tau_\varepsilon$, as $\varepsilon \downarrow 0$, in a direct manner in the second half of the proof of Theorem 3. Alternatively, we could invoke the convergence of the (appropriately rescaled) random walk with parameter $\frac{1}{2} - \varepsilon$ to a Brownian motion with drift parameter -2 . (Note, however, that convergence in distribution does not imply convergence of the respective first moments.) This provides an interpretation of Z as the time at which the limit process attains its global maximum; see also Borodin and Salminen (2002, p. 65). Similarly, we can interpret the limit in Theorem 4 as the distribution of the first hitting time at $-(\log \alpha)/4$ of a Brownian motion with drift parameter 2 and reflecting boundary at 0.

Remark 3. The problem of finding the global maximum of a random walk from a finite initial segment also appears in insurance mathematics in connection with the simulation of ruin functions; see, e.g. Chan *et al.* (2003), but the step distribution is not concentrated on $\{-1, 1\}$ in this context.

Appendix A.

The numerical values mentioned after Theorem 2 can be obtained as `taucdf [24]` and `taucdf [23]` respectively after running the following simple and self-explanatory R program:

```
1  n <- 128
2  z <- exp(complex(real=0*(0:(n-1)), imag=-2*pi*(0:(n-1))/n))
3  p <- 0.25
4  g <- function(z) {(2*(1-2*p)*z)/(2*(1-p)*z-1+sqrt(1-4*p*(1-p)*z^2))}
5  y <- fft(g(z), inv=TRUE)/n
6  taucdf <- cumsum(Re(y))
```

To prevent wraparound effects the power of 2 in line one has to be made sufficiently large. Use of the fast Fourier transform algorithm makes the calculation of high quantiles feasible, even if p is close to $\frac{1}{2}$. Note that `taucdf [k]` is the numerical approximation to $P(\tau \leq k - 1)$, $k = 1, \dots, n$. For details on the language or, more appropriately, the statistical environment R, we refer the reader to the URL <http://cran.r-project.org>. Other examples of the use of fast Fourier transforms in computational probability, together with a discussion of associated errors and ways to handle them, can be found in Grübel (1989) and Grübel and Hermesmeier (1999). For the distribution of τ we could of course use (5) instead, but for the corresponding calculation using algorithm (A2) we only have (6); we do not know an explicit formula for the corresponding mass function.

References

- BORODIN, A. N. AND SALMINEN, P. (2002). *Handbook of Brownian Motion—Facts and Formulae*, 2nd edn. Birkhäuser, Basel.
- BRÉMAUD, P. (1999). *Markov Chains. Gibbs Fields, Monte Carlo Simulation and Queues*. Springer, New York.
- CHAN, W.-S., YANG, H. AND ZHANG, L. (2003). Some results on ruin probabilities in a two-dimensional risk model. *Insurance Math. Econom.* **32**, 345–358.

- CHASSAING, P. (1999). How many probes are needed to compute the maximum of a random walk? *Stoch. Process. Appl.* **81**, 129–153.
- CHASSAING, P., MARCKERT, J. F. AND YOR, M. (2003). A stochastically quasi-optimal search algorithm for the maximum of the simple random walk. *Ann. Appl. Prob.* **13**, 1264–1295.
- COX, D. R. AND HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.
- FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd edn. John Wiley, New York.
- FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn. John Wiley, New York.
- GRADSTEIN, I. S. AND RYSHIK, I. M. (1981). *Tables of Series, Products and Integrals*, Vol. 2. Deutsch, Frankfurt.
- GRÜBEL, R. (1989). The fast Fourier transform algorithm in applied probability theory. *Nieuw Arch. Wisk.* **7**, 289–300.
- GRÜBEL, R. AND HERMESMEIER, R. (1999). Computation of compound distributions. I. Aliasing errors and exponential tilting. *Astin Bull.* **29**, 197–214.
- KALLENBERG, W. C. M. AND KONING, A. J. (1995). On Wieand's theorem. *Statist. Prob. Lett.* **25**, 121–132.
- KNUTH, D. E. (2000). *Selected Papers on Analysis of Algorithms*. CSLI Publications, Stanford, CA.
- KORF, I., YANDELL, M. AND BEDELL, J. (2003). *BLAST. An Essential Guide to the Basic Local Alignment Search Tool*. O'Reilly, Sebastopol.
- MOTWANI, R. AND RAGHAVAN, P. (1995). *Randomized Algorithms*. Cambridge University Press.
- ODLYZKO, A. M. (1995). Search for the maximum of a random walk. *Random Structures Algorithms* **6**, 275–295.
- STUART, A. AND ORD, K. (1987). *Kendall's Advanced Theory of Statistics*, Vol. 1, *Distribution Theory*, 5th edn. Griffin, London.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge University Press.