On the Number of Times where a simple Random Walk reaches its Maximum

Katzenbeisser, Walter; Panny, Wolfgang

DOI:
10.57938/9f8b2c6c-bfdb-4209-a183-00457831a90c

Published: 01/01/1990

Document Version:
Publisher's PDF, also known as Version of record

On the Number of Times where a Simple Random Walk reaches its Maximum

Walter Katzenbeisser, Wolfgang Panny

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 2
1990

ON THE NUMBER OF TIMES WHERE A SIMPLE RANDOM WALK REACHES ITS MAXIMUM

W. KATZENBEISSER and W. PANNY

University of Economics, Vienna

Abstract. Let $\mathbf{Q}_n$ denote the number of times where a simple random walk reaches its maximum, where the random walk starts at the origin and returns to the origin after $2n$ steps. Such random walks play an important rôle in probability and statistics. In this paper the distribution and the moments of $\mathbf{Q}_n$ are considered and their asymptotic behavior is studied.

Introduction. Let $\mathbf{X}_k$, $k = 1, 2, \ldots$, be independent and identically distributed random variables with

$$\mathbf{P}(\mathbf{X}_k = 1) = \mathbf{P}(\mathbf{X}_k = -1) = 1/2.$$

Consider the simple random walk

$$\mathbf{S}_m = \sum_{k=1}^{m} \mathbf{X}_k,\ m = 1, 2, \ldots, 2n \quad \text{with} \quad \mathbf{S}_0 = 0 \quad \text{and} \quad \mathbf{S}_{2n} = 0,$$

i.e. a simple random walk starting at 0 and leading to 0 after $2n$ steps. Such random walks play an important rôle in probability and statistics. Consider e.g. the random variable $\mathbf{D}_n^+$, the maximal (one-sided) height,

$$\mathbf{D}_n^+ = \max_{0 \le m \le 2n} \mathbf{S}_m,$$

which essentially constitutes the teststatistic of the one-sided Kolmogorov-Smirnov two-sample test with equal sample sizes. Another random variable closely related to $\mathbf{D}_n^+$ is $\mathbf{Q}_n$, where

$$\mathbf{Q}_n = [\text{number of times where the random walk reaches its maximum}].$$

A classical paper on the subject, containing among others the distribution of $\mathbf{D}_n^+$, $\mathbf{Q}_n$ and the joint distribution of $\mathbf{D}_n^+$ and $\mathbf{Q}_n$ is Dwass [1967]. Results on $\mathbf{D}_n^+$, emphasizing asymptotic properties can be found e.g. in Kemperman [1959], Katzenbeisser and Panny [1984a], Panny [1984]. A slightly more general problem is considered in Katzenbeisser and Panny [1984b], alternative notions of height are discussed in

Panny and Prodinger [1985]. In a recent paper Katzenbeisser and Panny [1990] the conditional distribution of $\mathbf{D}_n^+$ given $\mathbf{Q}_n$ is considered.

In this paper we deal with the distribution and the moments of $\mathbf{Q}_n$, where our main concern is the analysis of the asymptotic behavior.

Probability distribution. From Dwass [1967, p.1047]

$$\mathbf{P}(\mathbf{Q}_n \geq r) = \frac{\binom{2n-r+1}{n-r+1}}{\binom{2n}{n}}$$

one easily obtains

$$\mathbf{P}(\mathbf{Q}_n = r) = \frac{\binom{2n-r}{n-1}}{\binom{2n}{n}}, \tag{1}$$

where $r = 1, 2, \ldots, n+1$. To analyze the asymptotic behavior of (1) as $n \to \infty$, we use generalized trinomial coefficients (GTC's), defined by (cf. Panny [1984]):

$$\binom{n; \alpha, \beta, \gamma}{k} = [v^k](\alpha v^2 + \beta v + \gamma)^n,$$

where $[v^k]Q(v)$ denotes the coefficient of v^k in the power series $Q(v)$. GTC's are related to ordinary binomial coefficients by

$$\binom{n; 1/2, 0, 1/2}{2m} = \binom{n}{m} 2^{-n}.$$

By means of GTC's, (1) can be rewritten as:

$$\mathbf{P}(\mathbf{Q}_n = r) = 2^{-r} \frac{\binom{2n-r; 1/2, 0, 1/2}{(2n-r)+(r-2)}}{\binom{2n; 1/2, 0, 1/2}{2n}}.$$

Hence the asymptotic approximation of the GTC's (cf. Theorem 3 from Panny [1984, p.15]) yields after some manipulations

$$\mathbf{P}(\mathbf{Q}_n = r) = 2^{-r} \sqrt{\frac{2n}{2n-r}} \exp\left(-\frac{(r-2)^2}{2(2n-r)}\right) + O(n^{-1}), \qquad n \to \infty. \tag{2}$$

The O-term in eq.(2) holds uniformly for all r. For $r = O(n^{1/2})$ the following weaker result can be derived:

$$\mathbf{P}(\mathbf{Q}_n = r) = 2^{-r} \exp\left(-\frac{r^2}{4n}\right) + O(n^{-1/2}), \qquad n \to \infty. \tag{3}$$

For r fixed, the O-term improves to $O(n^{-1})$. Moreover, for $r = O(n^{\epsilon}), \epsilon > 0$, (1) as well as (3) becomes exponentially small and consequently the error term in (3) is of the same order of magnitude for this case. This simpler version will prove useful for the asymptotics of the moments. The following table illustrates the usefulness of approximations (2) and (3) even for smaller values of n.

	$n = 10$			$n = 100$		
r	exact	app.(2)	app.(3)	exact	app.(2)	app.(3)
1	0.5000	0.4997	0.4877	0.50000	0.50000	0.49875
2	0.2632	0.2635	0.2262	0.25126	0.25126	0.24751
3	0.1316	0.1317	0.0998	0.12563	0.12563	0.12222
4	0.0619	0.0617	0.0419	0.06250	0.06249	0.06005
5	0.0271	0.0267	0.0167	0.03093	0.03093	0.02936
6	0.0108	0.0106	0.0064	0.01523	0.01522	0.01428
7	0.0039	0.0037	0.0023	0.00746	0.00745	0.00691
8	0.0012	0.0011	0.0008	0.00363	0.00363	0.00333
9	0.0003	0.0003	0.0003	0.00176	0.00176	0.00160
10	0.0000	0.0000	0.0000	0.00085	0.00085	0.00076
11	0.0000	0.0000	0.0000	0.00041	0.00041	0.00036

Moments. From (1) we have by definition

$$\mathbf{E}(\mathbf{Q}_n{}^s) = \sum_{r=0}^{n+1} r^s \frac{\binom{2n-r}{n-1}}{\binom{2n}{n}}, \tag{4}$$

where $s \geq 1$. Converting from powers to factorials by means of Stirling numbers of the second kind $\left\{ {n \atop k} \right\}$, viz.

$$r^s = \sum_{k=0}^{s} \left\{ {s \atop k} \right\} [r]_k,$$

where $[r]_k = r(r-1)\cdots(r-k+1)$. Hence (4) can be written as

$$\sum_{0 \leq r} \sum_{0 \leq k \leq s} \left\{ {s \atop k} \right\} [r]_k \frac{\binom{2n-r}{n-1}}{\binom{2n}{n}}.$$

But

$$\sum_{k \leq r \leq n+1} [r]_k \binom{2n-r}{n-1} = k! \binom{2n+1}{n+k},$$

by a well-known binomial identity (cf. Graham, Knuth, Patashnik [1989, (5.26), p. 169]). Thus we have derived the following expression for the s-th moment of $\mathbf{Q}_n$:

$$\mathbf{E}(\mathbf{Q}_n{}^s) = \sum_{k=0}^{s} \left\{ {s \atop k} \right\} k! \frac{\binom{2n+1}{n+k}}{\binom{2n}{n}}, \quad s \geq 1. \tag{5}$$

Specializing on $s = 1$ and $s = 2$ a straightforward computation yields

$$\mathbf{E}(\mathbf{Q}_n) = \frac{2n+1}{n+1}, \qquad \mathbf{var}(\mathbf{Q}_n) = \frac{n^2\,(2n+1)}{(n+1)^2\,(n+2)}.$$

Since $\mathbf{D}_n^+$ and $\mathbf{Q}_n$ are closely related to each other it might be interesting to consider their covariance, too. It is well known (cf. Katzenbeisser and Panny [1984a, p.169]) that

$$\mathbf{E}(\mathbf{D}_n^+) = \frac{1}{2}\left(\frac{2^{2n}}{\binom{2n}{n}} - 1\right), \qquad \mathbf{var}(\mathbf{D}_n^+) = \frac{1}{4}\left(4n + 1 - \frac{2^{4n}}{\binom{2n}{n}^2}\right).$$

From Dwass's result on the joint distribution of $\mathbf{D}_n^+$ and $\mathbf{Q}_n$ (cf. Dwass [1967, VIII(a), p.1047]) one obtains after some manipulations the following closed form expression:

$$\mathbf{E}(\mathbf{D}_n^+\mathbf{Q}_n) = \frac{2^{2n}}{\binom{2n}{n}} - \frac{2n+1}{n+1}.$$

Hence

$$\mathbf{cov}(\mathbf{D}_n^+, \mathbf{Q}_n) = \frac{1}{2(n+1)}\left(\frac{2^{2n}}{\binom{2n}{n}} - 2n - 1\right)$$

and

$$\mathbf{corr}(\mathbf{D}_n^+, \mathbf{Q}_n) = -\frac{1}{n}\sqrt{\frac{n+2}{2n+1}}\,\frac{(2n+1)\binom{2n}{n}2^{-2n} - 1}{\sqrt{(4n+1)\binom{2n}{n}^2 2^{-4n} - 1}} \sim -\sqrt{\frac{2}{4-\pi}}\,n^{-1/2},$$

which shows that $\mathbf{D}_n^+$ and $\mathbf{Q}_n$ are negatively correlated, as was to be expected. Moreover, the last formula reveals that both random variables are asymptotically uncorrelated, as $n \to \infty$.

Unfortunately, (5) may become rather tedious to compute for larger values of n. Moreover, it does not help too much to estimate the order of magnitude of $\mathbf{E}(\mathbf{Q}_n{}^s)$. The following asymptotic considerations will show how to overcome these difficulties.

Neglecting for the moment the error term in (3), for large values of n the s-th moment of $\mathbf{Q}_n$ can be expressed as well by

$$\mathbf{E}(\mathbf{Q}_n{}^s) = \sum_{r=0}^{n+1} r^s 2^{-r} \exp(-\frac{r^2}{4n}). \tag{6}$$

Dissecting the range of summation we obtain

$$S_1 = \sum_{r=0}^{n^\epsilon} r^s 2^{-r} \exp(-\frac{r^2}{4n}) \quad \text{and} \quad S_2 = \sum_{n^\epsilon}^{n+1} r^s 2^{-r} \exp(-\frac{r^2}{4n}),$$

where $\epsilon > 0$. $S_2 = O(\exp(-n^\epsilon)$, which can be shown e.g. by use of the incomplete Γ-function. Hence, extending the range of summation to infinity results only in an exponentially small error. Expanding the exponential function yields:

$$\sum_{r=0}^{\infty} r^s 2^{-r}\Big(1 - \frac{r^2}{4n} + \frac{r^4}{32n^2} - \dots\Big) + O(n^{-m})$$

for all $m > 0$, as $n \to \infty$. Therefore we introduce the functions $g(s)$,

$$g(s) = \sum_{r=0}^{\infty} r^s 2^{-r}, \tag{7}$$

which allows to rewrite sum (6) as:

$$g(s) - \frac{g(s+2)}{4n} + \frac{g(s+4)}{32n^2} - \dots + O(n^{-m}) \;=\; g(s) + O(n^{-1}), \quad n \to \infty.$$

Let us now consider the error term in (3) neglected so far. Its contribution to S_1 is $O(n^{-1/2+\delta})$, for all $\delta > 0$, as $n \to \infty$. Its contribution to S_2 is exponentially small. Accordingly, the asymptotic behavior of the moments can be described by:

$$\mathbf{E}(\mathbf{Q}_n{}^s) = g(s) + O(n^{-1/2+\delta}), \tag{8}$$

for all $\delta > 0$, as $n \to \infty$. Eq.(8) shows that $\mathbf{E}(\mathbf{Q}_n{}^s)$ tends to the limit $g(s)$, as $n \to \infty$, which implies that the moments are essentially independent of n.

It only remains to study the function $g(s)$ in more detail. It is well-known (cf. Graham, Knuth, Patashnik [1989, (7.46), p.337]) that:

$$\sum_{r=0}^{\infty} r^s z^r = \sum_{k=0}^{s} \left\{ {s \atop k} \right\} \frac{k!\, z^k}{(1-z)^{k+1}}.$$

Hence

$$g(s) = 2 \sum_{k=0}^{s} \left\{ {s \atop k} \right\} k!. \tag{9}$$

Incidentally, since $\left\{ {s \atop k} \right\}$ equals the number of partitions of the set $\{1, 2, \dots, s\}$ into exactly k nonempty subsets, $g(s)/2$ has the following combinatorial interpretation: Consider all mappings f from $\{1, 2, \dots, s\}$ into $\{1, 2, \dots, s\}$. Then $g(s)/2$ counts the number of these mappings with the additional property that f is surjective onto $\{1, 2, \dots, k\}, 1 \le k \le s$.

Representation (9) for the $g(s)$-functions shows that expression (8) could also have been obtained directly from (5). Expanding the quotient of the binomial coefficients in (5), viz.

$$\frac{\binom{2n+1}{n+k}}{\binom{2n}{n}} = 2 - \frac{2k^2 - 2k + 1}{n} + O(n^{-2}), \quad n \to \infty,$$

results in

$$2\sum_{k=0}^{s}\left\{{s \atop k}\right\}k! - \frac{1}{n}\sum_{k=0}^{s}\left\{{s \atop k}\right\}k!(2k^2-2k+1)+O(n^{-2}).$$

Hence the O-term in (8) actually can be improved to $O(n^{-1})$.

Both approaches can be refined to get further terms in the asymptotic expansion (8). In principle, they could be extended to achieve an O-term as small as we please. Considering only the next term one obtains e.g.:

$$\mathbf{E}(\mathbf{Q}_n{}^s) = g(s) - \frac{1}{4n}\big(g(s+2)-5g(s+1)+4g(s)\big)+O(n^{-2}), \qquad n\to\infty. \quad (10)$$

The following table compares exact and approximate values (computed by means of (10)) for $\mathbf{E}(\mathbf{Q}_n{}^s)$.

	$\mathbf{E}(\mathbf{Q}_n{}^1)$		$\mathbf{E}(\mathbf{Q}_n{}^2)$		$\mathbf{E}(\mathbf{Q}_n{}^3)$		$\mathbf{E}(\mathbf{Q}_n{}^4)$	
n	ex.	app.	ex.	app.	ex.	app.	ex.	app.
10	1.91	1.90	5.09	4.90	18.06	15.10	78.94	36.10
50	1.98	1.98	5.79	5.78	23.97	23.82	129.58	127.22
100	1.99	1.99	5.89	5.89	24.95	24.91	139.23	138.61
250	2.00	2.00	5.96	5.96	25.57	25.56	145.55	145.44
500	2.00	2.00	5.98	5.98	25.78	25.78	147.75	147.72
1000	2.00	2.00	5.99	5.99	25.89	25.89	148.87	148.86

From a practical point of view eq. (9) constitutes no essential improvement over the exact expression (5). Therefore, in the remainder of this paper we shall investigate more closely the behavior of the $g(s)$-functions. A series expansion of $g(s)$ will be derived, whose essential term, viz. $s!/(\ln 2)^{s+1}$, will turn out to be an excellent approximation. Consider

$$f(x) = \frac{1}{1-e^x} = 1 + e^x + e^{2x} + e^{3x} + \dots, \qquad x<0.$$

The s-th derivative of $f(x)$ is

$$f^{(s)}(x) = 1^s e^x + 2^s e^{2x} + 3^s e^{3x} + \dots, \qquad x<0,$$

which entails

$$g(s) = f^{(s)}(-\ln 2).$$

Since

$$\coth\frac{x}{2} = \frac{e^x+1}{e^x-1},$$

$f(x)$ may be expressed as

$$f(x) = \frac{1}{2}\Big(1-\coth\frac{x}{2}\Big).$$

The Laurent series of $\coth x$ is

$$\coth x = \frac{1}{x} - \sum_{k\geq 1} \frac{2^{2k} B_{2k}}{(2k)!} x^{2k-1}, \quad |x| < \pi,$$

where B_{2k} denotes the $2k$-th Bernoulli number. Therefore $f^{(s)}(x)$ has the Laurent expansion

$$f^{(s)}(x) = (-1)^{s+1} \frac{s!}{x^{s+1}} + \frac{1}{2}\delta_{0,s} - \sum_{2k\geq s+1} \frac{B_{2k}[2k-1]_s}{(2k)!} x^{2k-s-1}, \quad |x| < 2\pi,$$

where $\delta_{i,j}$ is Kronecker's delta. So, we finally arrive at the following series expansion for $g(s)$:

$$g(s) = \frac{1}{2}\delta_{0,s} + \frac{s!}{(\ln 2)^{s+1}} + \frac{(-1)^s}{(\ln 2)^{s+1}} \sum_{2k\geq s+1} \frac{B_{2k}[2k-1]_s(\ln 2)^{2k}}{(2k)!}.$$

Above series is a covergent series which could as well be obtained by an application of Euler's summation formula to eq.(7). Confining to the first term only in this expansion and rounding to the nearest integer, i.e. $[s!/(\ln 2)^{s+1}]$, furnishes the correct value for $1 \leq s \leq 15$. For larger values of s the relative error of this approximation is smaller than 10^{-16}. The following table shows some values for small s:

s	$g(s)$	$s!/(\ln 2)^{s+1}$
1	2	2.0814
2	6	6.0056
3	26	25.9926
4	150	149.9975
5	1082	1082.0030
6	9366	9366.0025
7	94586	94585.9975
8	1091670	1091669.9958
9	14174522	14174522.0032
10	204495126	204495126.0105

References

DWASS, M. [1967] Simple random walk and rank order statistics, *Ann. Math. Statist.* **38**, 1042–1053.

GRAHAM, R. L., KNUTH, D. E., PATASHNIK, O. [1989] *Concrete Mathematics*, Addison-Wesley, Reading/Mass.

KATZENBEISSER, W., PANNY, W. [1984a] The moments of the teststatistics $D_{n,n}$ and $D^+_{n,n}$ of the Kolmogorov-Smirnov test and their asymptotic behavior, in: *Proc. of the 4th Pannonian Symp. on Math. Stat.*, W. Grossmann, G. Pflug, I. Vince, E. Wertz, eds., 167–183; Akadémiai Kiadó, Budapest.

KATZENBEISSER, W., PANNY, W. [1984b] Asymptotic results on the maximal deviation of simple random walks, *Stochastic Processes Appl.* **18**, 263–275.

KATZENBEISSER, W., PANNY, W. [1990] Some further results on the height of lattice paths, to appear in: *J. Stat. Plann. Inference.*

KEMPERMAN, J. H. B. [1959] Asymptotic expansions for the Smirnov test and for the range of cumulative sums, *Ann. Math. Statist.* **30**, 448–462.

PANNY, W. [1984] *The Maximal Deviation of Lattice Paths*, Athenäum/Hain/Hanstein, Königstein/Ts.

PANNY, W., PRODINGER, H. [1985] The expected height of paths for several notions of height, *Stud. Sci. Math. Hung.* **20**, 119–132.

Institut für Statistik
Wirtschaftsuniversität Wien

Augasse 2–6
A–1090 Vienna
Austria