Ultrahigh Moments for a Brownian Excursion

Joel H. Spencer
Courant Institute
spencer@cs.nyu.edu

June 7, 2002

Note: This is a somewhat speculative report, describing approaches to a problem which have not been put on a rigorous foundation.

1 The Exact Problem

Let $X_1, \ldots, X_n$ be i.i.d., each with distribution minus one plus a Poisson with mean one. Set $S_0 = 1$, $S_i = S_{i-1} + X_i$, the result of a walk with steps X_i beginning at 1. We condition on $S_n = 0$ and $S_i > 0$ for $i < n$, that the excursion first hits 0 at time n. Set $M = \sum_{i=1}^{n}(S_i - 1)$. We seek an asymptotic formula for $E[\binom{M}{k}]$ where $k = k(n)$.

The application is to graph theory. Let $c(n,k)$ denote the number of connected labeled graphs with n vertices and $n-1+k$ edges. Then [2] $E[\binom{M}{k}] = n^{n-2}c(n,k)$. An asymptotic formula for $c(n,k)$ was found by Bender, Canfield and McKay [1] in 1990. Hopefully one can get an alternate (simpler?) proof of this formula from the straight probability problem and also finding where the "weight" of $E[\binom{M}{k}]$ comes from gives insight into the nature of the random connected graph.

If $k > (\frac{1}{2}+\epsilon)n\ln n$ then a classic result of Erdös and Rényi gives that almost all graphs on n vertices, k edges are connected so that $E[\binom{M}{k}] \sim n^{2-n}\binom{N}{k}$ with $N = \binom{n}{2}$. Hence we restrict ourselves to $k < (\frac{1}{2}+\epsilon)n\ln n$.

We shall consider the asymptotics of the k-th moment, $E[M^k]$. This is asymptotic to $E[(M)_k]$ for $k = o(n)$ and differs from $E[(M)_k]$ by a calculatable constant when $k = \Theta(n)$. For $E[M^k]$ there is no longer a natural upper bound for k.

There are four basic regions: k constant, $k \to \infty$ but $k = o(n)$, $k = \Theta(n)$, $n \ll k$.

2 The region k constant

This was done in [2]. We scale time by n and distance by $n^{1/2}$ getting a brownian excursion with $f(t) = S_{nt}n^{-1/2}$. Then $M \sim Ln^{3/2}$ where $L = \int_0^1 f(t)dt$ is the mean distance from the origin in a brownian excursion. Then $E[M^k] \sim n^{3k/2}E[L^k]$ where the $E[L^k]$ are the moments of L which have been calculated by G. Louchard in 1984. This matches a known 1977 paper of E.M. Wright in which the asymptotic number of connected graphs with n vertices, $n-1+k$ edges was found.

3 The region $k \to \infty$ but $k = o(n)$

Now scale time by n and distance by $n^{1/2}k^{1/2}$ so that an excursion is associated with the function $f(t) = S_{nt}n^{-1/2}k^{-1/2}$. Now moving from $f(t)$ to $f(t+dt) = f(t)+f'(t)dt$ corresponds to the original walk moving $f'(t)(nk)^{1/2}dt$ in $n \cdot dt$ steps. With $k = o(n)$ this is not a very large deviation and the step size can be considered just to have mean zero and standard deviation one so that the probability is $\exp[-(f'(t)^2k/2) \cdot dt]$ for the walk to go this distance. Letting $\Pr[f]$ denote (nonrigorously!) the probability that the excursion follows path f we have $\Pr[f] \sim \exp[-k\int_0^1 f'(t)^2/2 \cdot dt]$. Such paths have value $n^{3k/2}k^{k/2}\exp[k\ln[\int_0^1 f(t)dt]]$. Ignoring the scaling terms, taking logs, and dividing by k we set

$$\Psi(f) = \ln[\int_0^1 f(t)dt] - \int_0^1 f'(t)^2/2 \cdot dt$$

so that the bigger $\Psi(f)$ is the larger the contribution to $E[M^k]$ of excursions of shape f.

This leads to a calculus of variations problem. Fixing $\int_0^1 f(t)dt$ we want to minimize $\int_0^1 f'(t)^2/2 \cdot dt$. The solution is a parabola $f(t) = at(1-t)$. Such f have $\Psi(f) = \ln(a/6) - \frac{a^2}{6}$ which is maximized at $a = \sqrt{3}$. Plugging back in this f gives that $E[M^k]$ is roughly $n^{3k/2}k^{k/2}(1/12e)^{k/2}$.

Suppose now $k \to \infty$ slowly. We outline an argument to give an asymptotic formula for $E[M^k]$ and thus an asymptotic formula for $c(n,k)$. However, making this argument rigorous presents a daunting technical challenge and we should note that we were guided by the already calculated value of $c(n,k)$.

We shall calculate probabilities for the unrestricted random walk with step size Poisson of mean one minus one and introduce the conditioning at the end. Split the excursion of time n into s equal parts. Define σ by $n/s = \sigma^2$ for convenience. Set $a_0 = 1, a_s = 0$ and, for $0 < i < s$, $a_i = \sqrt{3}\sqrt{k}\sqrt{n}\frac{i}{s}(1-\frac{i}{s})$ and consider those walks that at time iN/s are at position a_i. (That is, the walk follows the parabola given by the Calculus of Variations solution. We ignore integrality here and throughout this outline.) Now consider in general a walk of length M with each step of distribution Poisson of mean one, minus one. For a wide range of m the probability that the total distance is m is asymptotic to $(2\pi M)^{-1/2}e^{-m^2/2M}$. This is natural from the approximation by Brownian motion but also can be computed directly as the total distance is Poisson of mean M minus M. Then the probability of the walk passing through these points is asymptotically $(2\pi\sigma^2)^{-s}\exp[\sum_{0<i\leq s} -(a_i - a_{i-1})^2/2\sigma^2]$.

Of course, the paths don't have to go precisely through the $a_i's$. Set $SUM = \sum_{i=0}^{s} a_i$. We consider those paths such that the sum of their values at the iN/s is *precisely* SUM. We parametrize them by considering the walks that at time iN/s are at position $a_i + z_i$ and requiring $z_0 = z_s = 0$ (so the path begins and ends at the right place) and $\sum_{i=0}^{s} z_i = 0$ (so that SUM remains the same. These probabilities are as above except that $\exp[-(a_i - a_{i-1})/2\sigma^2]$ is replaced by $\exp[-(a_i - a_{i-1} + z_i - z_{i-1})^2/2\sigma^2$. Now in the cross terms z_i will have a coefficient of $-a_{i+1} - 2a_i + a_{i-1}$. As the a_i give a parabola this coefficient is constant (i.e., independent of i) and so with $\sum_i z_i = 0$ the total contribution of the cross terms is a factor of one! [This is not serendipitous but rather reflects the parabola being the solution of the Calculus of Variations problem.] One is left with an additional factor of $\exp[\sum_i -(z_i - z_{i-1})^2/2\sigma^2]$. To calculate this set $b_i := z_i - z_{i-1}$ for $1 \le i \le s$ so that the factor is $\exp[\sum_i -b_i^2/2\sigma^2]$. As an unrestricted sum over all possible integers $b_1, \ldots, b_s$ this would split into s identical products, each of which is asymptotically $\sigma\sqrt{2\pi}$ to give $(2\pi\sigma^2)^s$, which conveniently cancels the factor when all $z_i = 0$. But the sum is now restricted to $\sum_i b_i = 0$ (so that $z_s = 0$) and $\sum_i ib_i = 0$ (so that $\sum_i z_i = 0$). We may think of the b_i as weighted with a normal distribution with variance σ^2. Then $\sum_i b_i$ has variance $s\sigma^2 = n$ and is precisely zero with weight $(2\pi n)^{-1/2}$. The variable $\sum_i (i - \frac{s+1}{2})b_i$ is then orthogonal and has variance $\sigma^2 \sum_i (i - \frac{s+1}{2})^2 \sim \sigma^2 s^3/12 = s^2 n/12$ so that it is precisely zero with weight $(2\pi n s^2/12)^{-1/2}$. Together, the total probability of all paths running through these points is $(2\pi n)^{-1} s^{-1}\sqrt{12}$ times $\exp[\sum_i -(a_i - a_{i-1})^2/2\sigma^2]$.

A path going through such points is likely to have M close to $SUM\frac{N}{s}$, effectively approximating the integral (as M is the sum over all values) by the trapezoidal rule. We'll make the assumption (which we do not justify rigorously) that we can asymptotically replace M by $SUM \cdot \frac{N}{s}$. Now we are in the Brownian calculation and the contribution is $n^{3k/2}(k/12e)^{k/2}$ (the main term) times the $(2\pi n)^{-1} s^{-1}\sqrt{12}$ factor.

The sum of the values at the iN/s need not, of course, be precisely SUM and this gives another factor. Suppose SUM is replaced by $SUM(\sqrt{3}+\epsilon)/\sqrt{3}$. This changes the parabola by replacing $\sqrt{3}$ by $\alpha := \sqrt{3}+\epsilon$. The main factor has a term $\alpha^k e^{-k\alpha^2/6}$ (the remaining terms independent of α) which is maximized at $\alpha = \sqrt{3}$. The logarithm divided by k is then $\ln\alpha - \alpha^2/6 \sim c_0 - \frac{1}{3}\epsilon^2$ by Taylor Series, with c_0 the value at the maximum. When SUM is multiplied by $(\sqrt{3}+\epsilon)/\sqrt{3}$ the contribution is then multiplied by $\exp[-k\epsilon^2/3]$. Setting $\gamma := \epsilon(2k/3)^{1/2}$ we have that when SUM has $\gamma n^{1/2} s(3/2)^{1/2}/6$ added to it the contribution is multiplied by $\exp[-\gamma^2/2]$. This would give an extra factor of $(2\pi)^{1/2}$ but with the scaling factor the extra factor is $(2\pi)^{1/2} n^{1/2} s(3/2)^{1/2}/6 = s(\pi n/12)^{1/2}$. Note this cancels the previous s^{-1} factor and now the total contribution is $n^{3k/2}(k/12e)^{k/2}$ (the main term) times $(n\pi)^{-1/2}/2$.

While we have required our paths to begin at one and end at zero we have not yet introduced the requirement that they not otherwise touch the X-axis. This factor comes in at the beginning and at the end. At the beginning we have conditioned essentially on slope $\epsilon := \sqrt{3}\sqrt{k}n^{-1/2}$ so it is as if each step was Poisson with mean $(1+\epsilon)$ minus one. The probability that such a walk never hits the origin is [not an easy problem!] asymptotically 2ϵ so this gives an additional factor of $2\sqrt{3k/n}$. At the end, looking backwards, we start at zero and each step is one minus Poisson with mean $1-\epsilon$. Here the probability that such a walk never returns to the origin (i.e., goes positive and stays positive forever) is asymptotic to ϵ giving an additional factor of $\sqrt{3k/n}$ – so the total additional factors are $6k/n$ giving now a total contribution of $n^{3k/2}(k/12e)^{k/2}$ times $3k\pi^{-1/2}n^{-3/2}$.

Finally, the actual expectation is in the space conditional on the walk being an excursion so we must divide by the probability that the unrestricted walk really is an excursion. Remarkably, this has an exact value in relatively simple form. First, the total distance is Poisson with mean n minus n and so the probability this is precisely -1 is the probability that Poisson of mean n has value $n-1$ which is $e^{-n}n^{n-1}/(n-1)!$. Now we claim that given the walk ends at the origin the probability that it is an excursion (i.e., hadn't hit the origin before) is precisely $1/n$. We may think of balls labelled $1,\ldots,n-1$ each being independently and uniformly places on one of the positions $1,\ldots,n$ and letting X_i be the number of balls in position i. We set $W_0=1$ and $W_i=W_{i-1}+X_i-1$. There are precisely n^{n-2} cases when this is an excursion as they are in bijective correspondence with labelled trees T on $0,1,\ldots,n-1$ as follows: apply breadth-first search to T starting at 0, adding new vertices in numerical order. When vertex j is discovered at "time" i place ball j into position j. As there are n^{n-1} possible placements of the balls the probability is $1/n$ as claimed. Thus the exact probability that the unrestricted walk is an excursion is $e^{-n}n^{n-1}/n!$. This is asymptotic to $(2\pi)^{-1/2}n^{-3/2}$ by Stirling's formula.

Dividing by this final term, $E[M^k]\sim n^{3k/2}(k/12e)^{k/2}\cdot 3\sqrt{2}k$.

This yields Corollary 2 of [1] (noting their $w_k\to 1$ and that their $c(n,n+k)$ is our $c(n,k+1)$) as

$$c(n,k)=n^{n-2}E[\binom{M}{k}]\sim n^{n-2}n^{3k/2}(e/12k)^{k/2}\cdot 3k^{1/2}\pi^{-1/2}$$

4 A Parking Approach

Here we examine a somewhat different approach which in some rough sense attempts to move from a Brownian Bridge to a Brownian Excursion. Place balls $1,\ldots,n-1$ independently and uniformly into boxes $1,\ldots,n$. Let Y_j, $1\le j\le n-1$, be the position of the j-th ball. Let X_i, $1\le i\le n$, the number of balls in the i-th box minus one, the number of j with $Y_j=i$ minus one. Set $S_0=1$ and $S_i=1+\sum_{l=1}^{i}X_l$. Let $PARK$ be the event that $S_i>0$ for all $1\le i<n$ - that for each such i there are at least i balls in the first i boxes. Note that the joint distribution of the X_i conditional on $PARK$ is identical to the distribution defined at the top of this paper. [Generally, if $W_1,\ldots,W_a$ are independent Poissons of mean one and we condition on $W_1+\ldots+W_a=b$ it is equivalent to throwing b balls into a boxes and letting W_i be the number of

balls in the i-th box.] We set $M=\sum_{i=1}^{n}(S_i-1)$ as before. We observe that

$$M=\sum_{j=1}^{n}\left(\frac{n}{2}-Y_j\right)$$

(Both formulae give $M=0$ when there is one ball in each of the first $n-1$ boxes. Both formulae go down (up) by one when a single ball is moved one space to the left (right). Hence both formulae are always equal.)

Our original problem is then reformulated as that of estimating $E[\binom{M}{k}|PARK]$. For $PARK$ to hold there can be no ball in the final, n-th, box. Hence we may think of the $n-1$ balls being placed independently and uniformly in the first $n-1$ boxes, so that the Y_j are independent and uniform on $1,\ldots,n-1$. Observe that M now is the sum of independent identically distributed distributions which have a simple form and are of zero mean. Thus the calculation of the factorial moments of M is attackable by standard techniques. The S_i form a bridge, with $S_0=S_{n-1}=1$. The event $PARK$, that all intermediate $S_j\geq 1$, turns this into an excursion. What is the affect on these moments of the conditioning by $PARK$?

`Conjecture:` Let $k=k(n)$ satisfy $k(n)\to\infty$ and $k(n)=o(n)$. Then

$$E\left[\binom{M}{k}\chi(PARK)\right]\sim E\left[\binom{M}{k}\right](2\epsilon)(e\epsilon)$$

where $\epsilon=(3k)^{1/2}n^{-1}$. (Here $\binom{M}{k}$ is understood to be zero if M is negative.)

Here is the motivation. The calculus of variation discussed previously gives that the contribution to the k-th moment of M is concentrated around the excursion $S_j=(3k)^{1/2}j(n-j)n^{-2}$. In the region around zero this curve has slope ϵ. The number $1+X_l$ of balls in the l-th box is averaging $1+\epsilon$. It is "as if" the $1+X_l$ were independent Poisson distributions with mean $1+\epsilon$. As such the escape probability (in an infinite process with step size X_l (which is Poisson mean $1+\epsilon$ minus one) where you start at one and die if you hit zero) would be $\sim 2\epsilon$. Around $n-1$ (the right hand end) the curve has slope $-\epsilon$ and it is "as if" the $1+X_l$ had independent Poisson distributions with mean $1-\epsilon$. As such the escape probability (in an infinite process with step size $-X_l$ (which is one minus Poisson mean $1-\epsilon$) where you start at one and die if you go hit or cross zero – looking at the process in reverse time) would be $\sim e\epsilon$. This gives the two extra factors.

5 The region $k=\Theta(n)$

We set $k=cn$ and consider c a positive constant, $n\to\infty$.

We scale time by n and distance by n so that an excursion is associated with the function $f(t)=S_{nt}/n$. (Note $f(t)\leq 1$ tautologically as $X_i\geq -1$ and $S_n=0$.) Now moving from $f(t)$ to $f(t+dt)=f(t)+f'(t)dt$ corresponds to the original walk moving $f'(t)n\cdot dt$ in $n\cdot dt$ steps. This is in the realm of large deviations and we can no longer think of the steps X_i as simply having mean zero and standard deviation one. Letting $U_m = Y_1 + \ldots + Y_m$ with the Y_i i.i.d. with distribution minus one plus Poisson of mean one, standard methods give $\Pr[U_m \sim \alpha m] = \exp[(h(\alpha) + o(1))m]$ where $h(\alpha) = \alpha - (\alpha+1)\ln(\alpha+1)$ and the domain of definition is $\alpha \geq -1$. In our case this is $\exp[nh(f'(t))dt]$ so that $\Pr[f] \sim \exp[n \int_0^1 h(f'(t))dt]$. Such paths have value $n^{2cn} \exp[cn \ln[\int_0^1 f(t)dt]]$. Ignoring the scaling terms, taking logs, and dividing by n we set

$$\Psi(f) = c \ln[\int_0^1 f(t)dt] - \int_0^1 f'(t) - (f'(t)+1)\ln(f'(t)+1)dt$$

This leads to a calculus of variations problem. Fixing $\int_0^1 f(t)dt$ we want to minimize $\int_0^1 f'(t) - (f'(t)+1)\ln(f'(t)+1)dt$. The solution is a function of the form

$$f(t) = \frac{1-e^{-bt}}{1-e^{-b}} - t$$

where b is positive. (One can check that as $b \to 0$ this curve approaches a parabola symmetric about $t = \frac{1}{2}$ so that this solution meshes with the $k = o(n)$ solution.) This gives

$$\Psi(f) = c \ln\left[\frac{1}{1-e^{-b}} - \frac{1}{2} - \frac{1}{b}\right] - \ln\left[\frac{b}{1-e^{-b}}\right] + 1 - \frac{be^{-b}}{1-e^{-b}}$$

We select $b = b(c)$ to maximize the right hand side, it does not appear to have a closed form, and let $z = z(c)$ denote the maximal value. Then $E[M^{cn}]$ is roughly $n^{2cn}e^{nz(c)}$.

6 The region $n \ll k$

Here $E[M^k] = [\frac{n^2}{2}(1+o(1))]^k$. We note $M \leq \frac{n^2}{2}$ tautologically, since the best M can do is jump to $n-1$ on the first step and slide back down to zero one step at a time. Conversely, for any fixed s the probability of having $\sim n/(s+1)$ steps s followed by $ns/(s+1)$ steps -1 is β^{-n} for some calculatable $\beta = \beta(s)$ but then its k-th root is negligible and $M = \frac{n^2}{2}\frac{s}{s+1}$ which is within an arbitrary factor of $\frac{n^2}{2}$.

This region meshes with the $k = cn$ region as when $c \to \infty$ $b = b(c) \sim 2c \to \infty$ and f approaches the spike function $f(t) = 1 - t$, the original walk jumping from $S_0 = 1$ to $S_i \sim n$ with $i = o(n)$.

7 Two Questions

```
1.  Can the above be made rigorous?
2.  Can the estimates be improved to give an asymptotic formula?
```

References

[1] Ed Bender, E. Rodney Canfield and B.D. McKay, The Asymptotic Number of Labeled Connected Graphs with a Given Number of Vertices and Edges, Random Structures & Algorithms 2 (1990), 127-169

[2] J. Spencer, Enumerating Graphs and Brownian Motion, Communications on Pure and Appl. Math. 50 (1997), 293-296