Lecture 7: Asymptotics of the Bernoulli Random Walk

Scribe: Jeff Aristoff

Department of Mathematics, MIT

February 24, 2005

The general method of steepest-descent asymptotics was described in Lecture 6 and applied to obtain a globally valid approximation the PDF of the position of a random walk. In this lecture, we will illustrate the method for the case of a symmetric Bernoulli random walk on the integers, where each step displacement is ± 1 with probability $1/2$. First, we will derive the necessary transform. Then, we will use the saddle point method (a special case of steepest-descent asymptotics) to obtain a globally valid asymptotic expansion as N, the number of steps taken by the walker, goes to infinity. Finally, we will compare this result with our previous result from the Gram-Charlier expansion, as well as provide numerical results to justify our conclusions.

1 The Bernoulli Random Walk - Integral Representation

(The following derivation of the integral representation of the Bernoulli random walk was given in 2003 and scribed by Kevin Chu. It provides a natural starting point for saddle-point asymptotics. See also Hughes Ch. 2.)

In the Bernoulli random walk, the individual steps are i.i.d. with probability density function given by

$$p(m) = \frac{1}{2} \left(\delta_{m,1} + \delta_{m,-1} \right), \tag{1}$$

where $\delta_{i,j}$ is the Kronecker delta and $m \in Z$. The probability density function after N steps can be written using Bachelier's equation in its discrete form:

$$P_N(m) = \sum_{j=-\infty}^{\infty} p(j)P_{N-1}(m - j). \tag{2}$$

Because this problem is discrete, we proceed with the analysis by using Fourier series (or equivalently, the z-transform) instead of Fourier transforms:

$$\hat{p}(k) = \sum_{m=-\infty}^{\infty} e^{ikm}p(m) \tag{3}$$

$$p(m) = \int_{-\pi}^{\pi} e^{-ikm}\hat{p}(k)\frac{dk}{2\pi}, \tag{4}$$

where k is a continuous variable and m is discrete. For convenience, we will continue to refer to $\hat{p}(k)$ as the transform of $p(m)$.

For the Bernoulli random walk, the characteristic function, $\hat{p}(k)$, is given by

$$\hat{p}(k) = \cos(k). \tag{5}$$

As in the continuous case, the transform of a convolution (in this case discrete) is the product of the transforms. Applying this result, we arrive at the usual expression for $\hat{P}_N(k)$:

$$\hat{P}_N(k) = \cos^N(k), \tag{6}$$

which immediately leads to the result

$$P_N(m) = \int_{-\pi}^{\pi} e^{-ikm} \cos^N(k) \frac{dk}{2\pi}. \tag{7}$$

This integral can be evaluated exactly using contour integration (see Supplementary Note on Lecture 16 in the course reader, 2003) to obtain

$$P_N(m) = \left[1 + (-1)^{N+m}\right] 2^{-N-1} \frac{N!}{\Gamma\left(1 + \frac{N-|m|}{2}\right) \Gamma\left(1 + \frac{N+|m|}{2}\right)}, \quad |m| < N. \tag{8}$$

While it is possible to examine the large n limit by using Stirling's formula to approximate the factorial and gamma functions in equation (8), it is instructive to apply the saddle-point method directly to the integral (7) to find the asymptotic limit for $P_N(m)$.

We begin the analysis by making a few observations about the integrand, $f(k) = e^{-ikm} \cos^N(k)$, in equation (7):

- $f(k)$ is 2π-periodic

- $f(k + \pi) = (-1)^{N+m} f(k)$.

Using these observations, we can manipulate the integral for $P_N(m)$ into an integral over the interval $(-\pi/2, \pi/2)$:

$$
\begin{aligned}
\int_{-\pi}^{\pi} e^{-ikm} \cos^N(k) \frac{dk}{2\pi} &= \int_{-\pi/2}^{3\pi/2} e^{-ikm} \cos^N(k) \frac{dk}{2\pi} \\
&= \left(\int_{-\pi/2}^{\pi/2} + \int_{\pi/2}^{3\pi/2} \right) e^{-ikm} \cos^N(k) \frac{dk}{2\pi} \\
&= \left[1 + (-1)^{N+m}\right] \int_{-\pi/2}^{\pi/2} e^{-ikm} \cos^N(k) \frac{dk}{2\pi},
\end{aligned}
$$

where we used the 2π-periodicity of $f(k)$ in the first step to shift the domain of integration and the behavior of $f(k)$ under a shift of π to change the integral over $(\pi/2, 3\pi/2)$ into an integral over $(-\pi/2, \pi/2)$. Thus, we find that

$$P_N(m) = \left[1 + (-1)^{N+m}\right] \int_{-\pi/2}^{\pi/2} e^{-ikm} \cos^N(k) \frac{dk}{2\pi}. \tag{9}$$

We are now ready to use the saddle-point method. For convenience, we write $P_N(m)$ as

$$P_N(m) = \frac{\left[1 + (-1)^{N+m}\right]}{2\pi} I_N(m), \tag{10}$$

where $I_N(m)$ is defined by

$$I_N(m) = \int_{-\pi/2}^{\pi/2} e^{-ikm} \cos^N(k) \frac{dk}{2\pi}. \tag{11}$$

In order to evaluate this integral for large N, we will invoke the method of steepest descent, also known as saddle point asymptotics.

2 Saddle Point Asymptotics of the Bernoulli Random Walk

As an illustration of the steepest-descent method given in the previous lecture, we will proceed to deform the contour of integration in Equation 11 in such a way to account for the dominant contributions of the integrand. This will allow us to find a globally valid asymptotic expansion as the number of random steps goes to infinity. First, we write

$$I_N(m) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-Nf(k,\xi)} dk \tag{12}$$

where $f(k, \xi) = ik\xi - \log(\cos(k))$ and $\xi = \frac{m}{N} = O(1)$, (ξ fixed). This scaling will therefore give a valid result for $m = O(N)$, which means the approximation, once found, will be valid for all m[1].

To evaluate the above integral as $N \to \infty$, we must find its dominant contributions. These will occur at either the endpoints of the integral, at saddle points of the modulus of the integrand (if any exist), or near singularities, such as branch cuts, where the integrand is not analytic. In our case, the endpoints do not contribute as $f(k = \pm\frac{\pi}{2}, \xi) \to \infty$ and thus the integrand approaches zero exponentially quickly near the endpoints[2]. The saddle points, denoted k_s, are found where $f(k, \xi)$ achieves its minimum, where

$$f' = i\xi + \tan(k_s) = 0 \tag{13}$$

or equivalently,

$$k_s(\xi) = -\tan^{-1}(i\xi). \tag{14}$$

This form however, will not allow us to easily determine the saddle point contribution. Instead, write

$$\tan(k_s) = \frac{\sin(k_s)}{\cos(k_s)} = \frac{e^{ik_s} - e^{-ik_s}}{e^{ik_s} + e^{-ik_s}} \left(\frac{1}{i}\right) = -i\xi \tag{15}$$

and substitute $q = e^{ik_s}$ to find $q = \pm\sqrt{\frac{1+\xi}{1-\xi}}$ and hence

$$k_s = -i \log q = -\frac{i}{2} \log \frac{1+\xi}{1-\xi} + \pi l \tag{16}$$

where l is an integer[3]. Also, since $|\xi| \le 1$, the principal branch of the logarithm will be bounded for all ξ except $|\xi| = 1$. However, when $|\xi| = 1$, $P_N(N) = \frac{1}{2}^N$. Therefore, we can handle this case separately and consider only when $|\xi| < 1$.

---

[1]Recall that the position, m, can never be greater than the total number of steps taken, N.

[2]For this reason, it is much easier to evaluate $I_N(m)$ than $P_N(m)$ using the saddle point method.

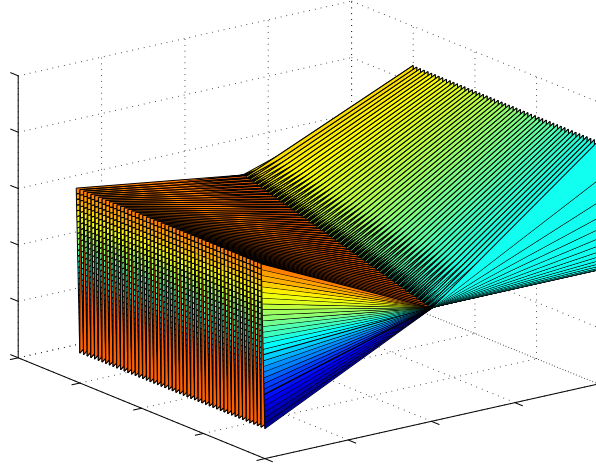[3]For $z = re^{i\theta} = re^{i(\theta+2\pi)}$, $\log(z)$ is defined as $\log(r) + i(\theta + 2\pi)$.

Figure 1: The real part of $f(k, \xi)$ showing the saddle points (which are also saddle points of $|e^{f(k,\xi)}| = e^{\mathrm{Re}f(k,\xi)}$).

Finally, we require the value of $f(k_s, \xi)$ as well as $f''(k_s, \xi)$ along the principal branch. Plugging k_s into the definition of $f(k, \xi)$, we obtain

$$
\begin{aligned}
f(k_s, \xi) &= i\xi k_s - \log(\cos k_s) \\
&= i\xi k_s - \log\left(\frac{e^{ik_s} + e^{-ik_s}}{2}\right) \\
&= \frac{-i^2\xi}{2}\log\left(\frac{1-\xi}{1+\xi}\right) - \log\sqrt{\frac{1}{1-\xi^2}} \\
&= \frac{\xi}{2}\log\left(\frac{1+\xi}{1-\xi}\right) + \frac{1}{2}\log\left(1-\xi^2\right)
\end{aligned}
\tag{17}
$$

and

$$
f''(k_s, \xi) = -\sec^2 k_s = \frac{-1}{\cos^2(k_s)} = 1 - \xi^2.
\tag{18}
$$

Near k_s, what does f look like? Taylor expanding f at k_s and recalling that $f'(k_s, \xi) = 0$ yields

$$
f(k, \xi) \sim f(k_s, \xi) + \frac{f''(k_s, \xi)}{2}(k - k_s)^2.
\tag{19}
$$

The direction of steepest descent will be where the real part of the integral is growing. Therefore, this path will be along the real direction.

We now have everything needed to use the results from last lecture, where we found that the saddle point contribution is

$$
e^{-Nf(k_s,\xi)}\sqrt{\frac{2\pi}{N \mid f''(k_s, \xi) \mid}}.
\tag{20}
$$

Therefore, combining equations 10, 17, 18, and 20 gives

$$
P_N(m) = P_N(N\xi) \sim \frac{1 + (-1)^{N+m}}{\sqrt{2\pi N(1-\xi^2)}}\left(\frac{1+\xi}{1-\xi}\right)^{-\frac{N\xi}{2}}
\tag{21}
$$

which is a globally valid solution with $\xi = \frac{m}{N}$ for $\mid \xi \mid < 1$.

3   Performance of the Saddle Point Approximation

The *exact* solution for the position of the Bernoulli random walk can be expressed combinatorially as the following:

$$P_N(m) = \left(\frac{1}{2}\right)^N \binom{N}{\frac{N+m}{2}} \tag{22}$$

for $(N + m)$ even, otherwise $P_N(m) = 0$. In addition, we found previously that the Gram-Charlier expansion for the Bernoulli random walk was the following:

$$\phi_N(z) = \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}} \left(1 - \frac{H_4(z)}{12N} + \dots\right) \tag{23}$$

where $\phi_N(z)$ is locally valid for $z = \frac{m}{\sqrt{N}} = O(1)$. In comparison, Equation 21 is valid for $z = \frac{m}{N} = O(1)$. Indeed, there arise times when one needs to know what is happening in the tail regions of the Bernoulli random walk, and thus the Gram-Charlier expansion will not be sufficient[4].

    The central region is well approximated by $\phi_N(z)$ given by the Gram-Charlier expansion. How well does the saddle point expansion perform in the central region as $\xi \to 0$? This question can be answered by Taylor expanding the components of Equation 21 near $\xi = 0$ which gives

$$f(k_s, \xi) \sim \frac{\xi^2}{2} + \frac{\xi^4}{12} \tag{24}$$

and

$$(1 - \xi^2)^{-\frac{1}{2}} \sim 1 + \frac{\xi^2}{2} + \frac{3\xi^4}{8}. \tag{25}$$

Therefore, the saddle point approximation from Equation 21 gives

$$P_N(m) \sim \frac{1 + (-1)^{N+m}}{\sqrt{2\pi N}} e^{-N\frac{\xi^2}{2}} e^{-N\frac{\xi^4}{12}} \left(1 + \frac{\xi^2}{2} + \frac{3\xi^4}{8}\right) \tag{26}$$

Let $z = \xi\sqrt{N}$. Then we find

$$P_N(m) \sim \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi N}} \left(1 - \frac{1}{12N}(z^4 - 6z^2) + O\left(\frac{1}{N^2}\right)\right) \tag{27}$$

which can be written in terms of the fourth Hermite Polynomial (with weight function $e^{\frac{-z^2}{2}}$) as

$$P_N(m) \sim \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi N}} \left(1 - \frac{1}{12N}(H_4(z) - 3) + O\left(\frac{1}{N^2}\right)\right). \tag{28}$$

For small N, this is nearly the Gram-Charlier result with a shift of 3 at the origin. This shift is needed because the saddle-point approximation coincides with the leading-order CLT approximation *exactly* at $z = 0$, while the Gram-Charlier correction terms improve the approximation in the central region. This helps us understand the poor approximation by the saddle-point method for small N. The figures in the Section 4 will illustrate this result, as well as provide a qualitative comparison between the various approximations.

---

[4]For example, see the percentile order statistics question on Problem Set 2, 2005.

4   The Bernoulli Random Walk - Numerical Results

(The figures and first draft of this section were provided by C. H. Rycroft.)

Up to this point we have considered several approximations to the Bernoulli walk, as well as the exact solution. In this section, we will present numerical results which illustrate their advantages and disadvantages. First, let us summarize our analytic results, in terms of the appropriately scaled variable, z:

- The **Exact Probability Distribution** after N steps is given by

$$P_N(X = x) = \left(\frac{1}{2}\right)^N \binom{N}{\frac{N+x}{2}} \qquad \text{(for } x = -N, -N+2, \ldots, N.)$$

- The **Central Limit Theorem** tells us that in the central region $z = O(1)$, $\phi_N(z)$ tends to the Gaussian distribution

$$\phi_N(z) \sim \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}.$$

- The **first order Gram-Charlier correction** is given by

$$\phi_N(z) \sim \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \left(1 - \frac{2}{4!N} H_4(z)\right),$$

where the fourth Hermite polynomial is $H_4(z) = z^4 - 6z^2 + 3$. As described in the lectures, we expect this correction to become important for $|z| > (24/|\lambda_4|)^{1/4} N^{1/4} = 1.86 N^{1/4}$.

- The **globally valid asymptotic approximation** to $\phi_N(z)$, that is valid for any z in the limit $N \to \infty$, can be expressed as

$$\phi_N(z) \sim \frac{1}{\sqrt{2\pi \left(1 - \frac{z^2}{N}\right)}} \exp\left(-\frac{z\sqrt{N}}{2} \log\left(\frac{\sqrt{N} + z}{\sqrt{N} - z}\right) - \frac{N}{2} \log\left(1 - \frac{z^2}{N}\right)\right).$$

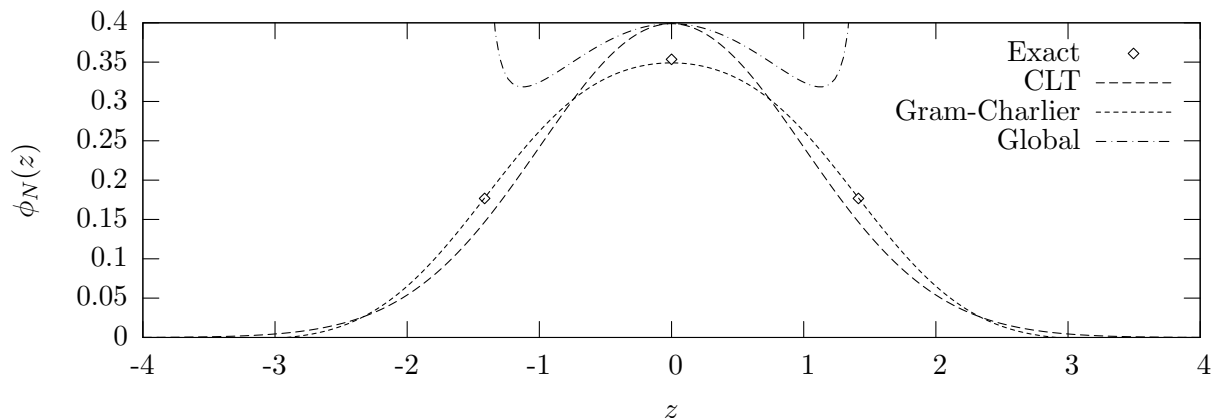The following four figures illustrate the numerical results for a given number of steps N.

Figure 1: Even at $N = 2$, the CLT approximation is quite reasonable, while the Gram-Charlier expansion with only two terms is remarkably accurate, and by far the best. The saddle-point approximation is uniformly poor for this small value of N: It blows up in the tails and does not even describe the central region very well. Note that it also always coincides with the CLT approximation
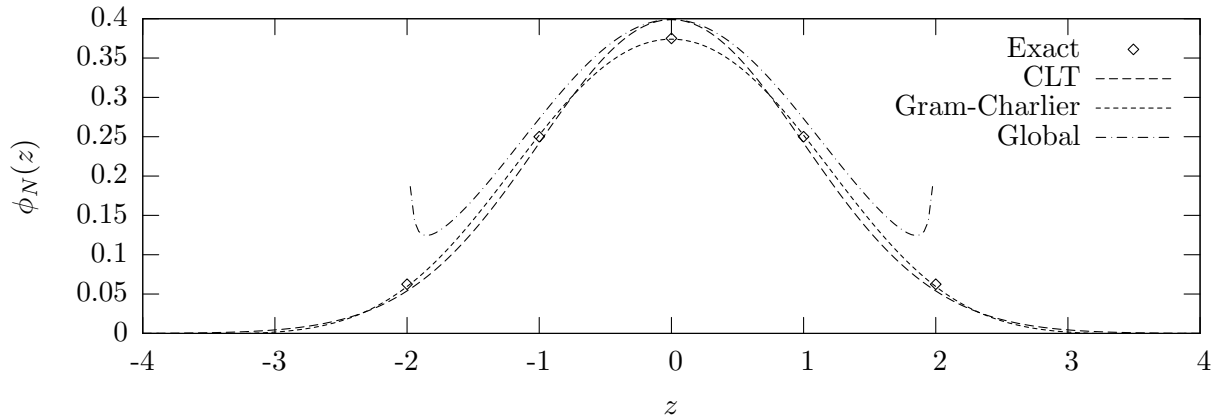
Figure 2: At $N = 4$, all three distributions match the exact solution, with the Gram-Charlier approximation being again the most accurate. The globally valid approximation blows up at $x = \pm N$.
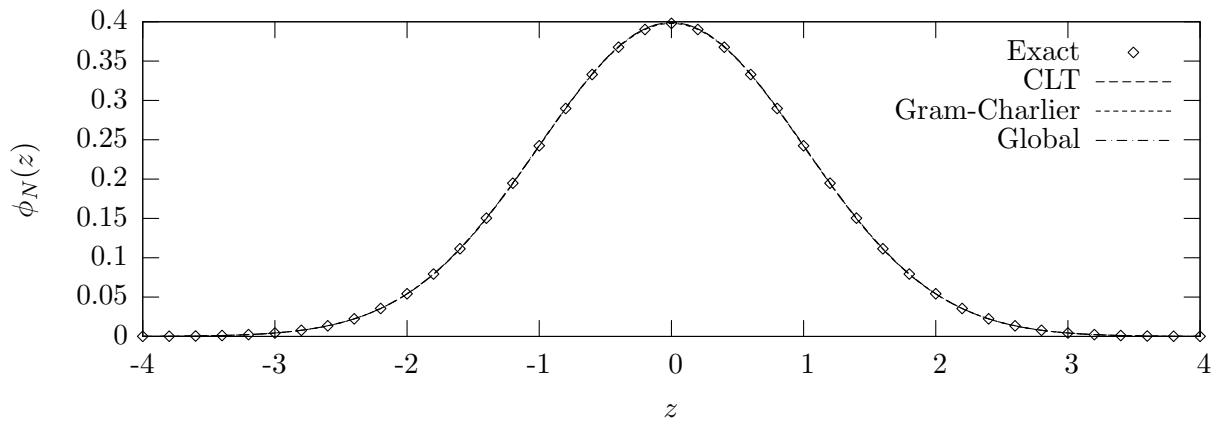

Figure 3: At $N = 100$ the three curves are indistinguishable from the exact solution in the whole of the central region.
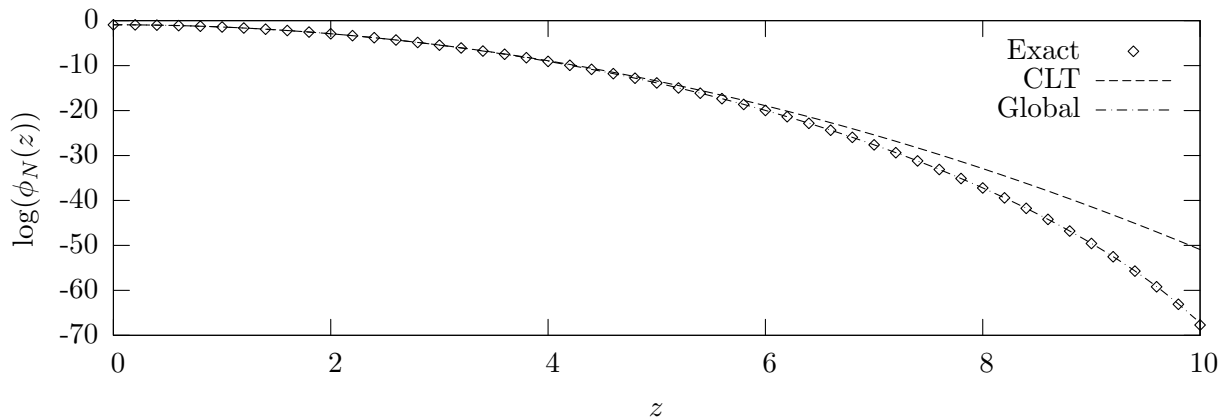

Figure 4: A logarithmic plot of the tails of the distribution for $N = 100$ highlights the difference between the approximations. The CLT matches the exact solution in the central region but significantly deviates for large z. The globally valid approximation, however, matches the exact solution very accurately for the entire range. The Gram-Charlier approximation becomes negative in the tails and cannot even be plotted there.