

Lecture 3: Central Limit Theorem

Scribe: Jacy Bird
(Division of Engineering and Applied Sciences, Harvard)

February 8, 2003

The goal of today's lecture is to investigate the asymptotic behavior of $P_N(\vec{x})$ for large N . We use Laplace's Method to show that a well-behaved random variable tends to a multivariate normal distribution. This lecture also introduces normalized cumulants and briefly describes how these affect the distribution shape.

General Distribution

As we saw in previous lecture, the probability density function for the random-walk position after N independent steps, \mathbf{X}_N , is given by

$$P_N(\mathbf{x}) = \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x}} \left(\prod_{n=1}^N \hat{p}_n(\mathbf{k}) \right) \frac{d^d\mathbf{k}}{(2\pi)^d}. \quad (1)$$

In the case where $p(\mathbf{x})$ is iid, Equation (1) simplifies to

$$P_N(\mathbf{x}) = \int_{-\infty}^{\infty} e^{i\mathbf{k}\cdot\mathbf{x}} (\hat{p}(\mathbf{k}))^N \frac{d^d\mathbf{k}}{(2\pi)^d}. \quad (2)$$

Here we have defined $p(\mathbf{x})$ to be the PDF for each displacement and $P_N(\mathbf{x})$ to be the PDF for the position, \mathbf{X}_N , after N (IID) steps. Thus it follows that there are two moment generating functions (MGF) of interest: $\hat{p}(\mathbf{k})$ is the MGF for \mathbf{x}_n and $\hat{P}_N(\mathbf{k})$ is the MGF for \mathbf{X}_N . Likewise, we can define a cumulant generating function (CGF) for steps \mathbf{x}_n and \mathbf{X}_N . As we saw in Lecture #2, the CGF for steps \mathbf{x}_n is defined as

$$\begin{aligned} \psi(\mathbf{k}) &= \log \hat{p}(\mathbf{k}) \\ &= -i\underline{c}_1 \cdot \mathbf{k} - \frac{1}{2} \mathbf{k} \cdot \underline{c}_2 \cdot \mathbf{k} + \dots \end{aligned} \quad (3)$$

Since cumulants for IID random variables are additive, the cumulant generating function for \mathbf{X}_N is denoted $\psi_N = (\mathbf{k}) \log \hat{P}_N(\mathbf{k}) = N\psi(\mathbf{k})$. As shown in Lecture #2, the variance of the distribution, σ^2_N , equals $C_{2,N} = Nc_2$ which is equivalent to $N\sigma^2$. Thus the width of the distribution, σ_N , scales like \sqrt{N} .

Shape of the Distribution for Large N

The exact expression for $P_N(\mathbf{x})$ becomes increasingly complex as N increases; however, for sufficiently large N the distribution tends to a general shape. In the last section (and last lecture),

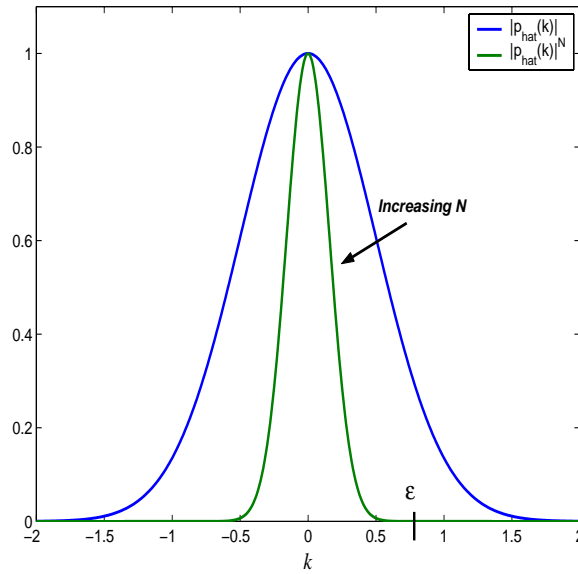


Figure 1: The dominant contribution of the integral occurs within a region ε away from the maximum point as $N \rightarrow \infty$.

it was showed that the width of this distribution scales as \sqrt{N} . In this section we are going to investigate the \sqrt{N} scaling more formally and use Laplace’s Method to derive the asymptotic distribution shape. For a review of Asymptotic Methods, the reader is referred to Bender and Orszag (1978)¹. Laplace’s Method takes advantage of the fact that the dominant contribution of an integral $\int f(k)e^{N\phi(x,k)}dk$ for large N occurs a small distance, ε , from the $\max \phi(k)$ (Fig. 1).

We run into potential problems if we let \mathbf{x} and N both go to infinity, as will be discussed in a later lecture. Therefore for the time being, we will only look at the ”central region” of the distribution.

Let’s begin by rewriting Equation (2) in terms of the cumulant generating function,

$$P_N(\mathbf{x}) = \int_{all \mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}+N\psi(\mathbf{k})} \frac{d^d \mathbf{k}}{(2\pi)^d} \tag{4}$$

For large N , the dominant contribution of the integral is near the origin, so we can collapse the limits of integration to a small region around the origin (this approximation introduces exponentially small error). Thus for large N ,

$$P_N(\mathbf{x}) \int_{|\mathbf{k}|<\varepsilon} e^{i\mathbf{k}\cdot\mathbf{x}} e^{N\psi(\mathbf{k})} \frac{d^d \mathbf{k}}{(2\pi)^d} \tag{5}$$

The next step is to Taylor expand $\psi(\mathbf{k})$ about the origin which as was done in Equation (3). This leads to

$$P_N(\mathbf{x}) \int_{|\mathbf{k}|<\varepsilon} e^{i\mathbf{k}\cdot\mathbf{x}} e^{-iN\underline{c}_1\cdot\mathbf{k}-\frac{N}{2}\mathbf{k}\cdot\underline{c}_2\cdot\mathbf{k}+\dots} \frac{d^d \mathbf{k}}{(2\pi)^d} \tag{6}$$

¹See Bender and Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, Sec. 6.4 (Springer, 1978), for a detailed discussion on Laplace’s Method.

For sufficiently large N we can truncate the series after the $\mathbf{k} \cdot \underline{\underline{c}}_2 \cdot \mathbf{k}$ term. Now we expand the limits of integration back to $-\infty$ to ∞ (again, this only introduces exponentially small errors). Thus for sufficiently large N ,

$$P_N(\mathbf{x}) \sim \int_{all \mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}} e^{-iN\underline{\underline{c}}_1 \cdot \mathbf{k} - \frac{N}{2} \mathbf{k} \cdot \underline{\underline{c}}_2 \cdot \mathbf{k}} \frac{d^d \mathbf{k}}{(2\pi)^d} \quad (7)$$

$$\sim \int e^{i\mathbf{k} \cdot (\mathbf{x} - N\underline{\underline{c}}_1)} e^{-\frac{1}{2} \mathbf{k} \cdot (N\underline{\underline{c}}_2) \cdot \mathbf{k}} \frac{d^d \mathbf{k}}{(2\pi)^d} \quad (8)$$

The next step is to transform this expression into a Gaussian integral. More specifically we want to transform $e^{-\frac{1}{2} \mathbf{k} \cdot (N\underline{\underline{c}}_2) \cdot \mathbf{k}}$ to $e^{-\mathbf{w}^2/2}$. Provided that $\underline{\underline{c}}_2$ is positive definite and symmetric, we define

$$\mathbf{w} \equiv \sqrt{N\underline{\underline{c}}_2} \cdot \mathbf{k} \quad (9)$$

Note 1: It's worth reviewing what is meant by the square root of a matrix (or 2nd order tensor). From linear algebra we know that \mathbf{M} can be decomposed into $\mathbf{Q}\mathbf{A}\mathbf{Q}^T$, where \mathbf{Q} is orthogonal ($\mathbf{Q}^{-1} = \mathbf{Q}^T$, $|\mathbf{Q}| = 1$) and \mathbf{A} is diagonal ($\mathbf{A}_{ij} = a_i \delta_{ij}$ with $a_i > 0$ being the eigenvalues of \mathbf{M}). In general, we can define \mathbf{M} raised to the α power as $\mathbf{M}^\alpha \equiv \mathbf{Q}\mathbf{A}^\alpha\mathbf{Q}^T$. Here \mathbf{A}^α is diagonal with $(\mathbf{A}^\alpha)_{ij} = a_i^\alpha \delta_{ij}$. Hence, the square root of \mathbf{M} is defined above with $\alpha = \frac{1}{2}$.

Returning to the distribution, we make the change of variable

$$\mathbf{w} = \sqrt{N\underline{\underline{c}}_2} \mathbf{k} \quad (10)$$

$$d^d \mathbf{w} = \left| \sqrt{N\underline{\underline{c}}_2} \right| d^d \mathbf{k} \quad (11)$$

$$= N^{d/2} \left| \sqrt{\underline{\underline{c}}_2} \right| d^d \mathbf{k} \quad (12)$$

In these equations $\left| \sqrt{\underline{\underline{c}}_2} \right|$ is the determinant of $\sqrt{\underline{\underline{c}}_2}$ and equals $\prod_{i=1}^d \sqrt{a_i}$ where a_i are the eigenvalues of $\underline{\underline{c}}_2$. With the substitution, Equation (8) is rewritten as

$$P_N(\mathbf{x}) \sim \int e^{i(\mathbf{x} - N\underline{\underline{c}}_1) \cdot \left(N\underline{\underline{c}}_2 \right)^{-\frac{1}{2}} \cdot \mathbf{w}} e^{-\frac{\mathbf{w}^2}{2}} \frac{d^d \mathbf{w}}{(2\pi)^d N^{\frac{d}{2}} \left| \sqrt{\underline{\underline{c}}_2} \right|} \quad (13)$$

Equation (13) can be further simplified by defining a new variable, $\mathbf{z} \equiv \frac{\mathbf{x} - N\underline{\underline{c}}_1}{\sqrt{\underline{\underline{c}}_2}}$

$$P_N(\mathbf{x}) \sim \frac{1}{(2\pi)^d N^{\frac{d}{2}} \left| \sqrt{\underline{\underline{c}}_2} \right|} \int e^{i\mathbf{z} \cdot \mathbf{w}} e^{-\frac{\mathbf{w}^2}{2}} d^d \mathbf{w} \quad (14)$$

The above equation [Equation (14)] is expressed in terms of a multivariate Gaussian distribution. However, this distribution can also be interpreted as a product of d 1-D Gaussian distributions taken separately. Thus Equation (14) can be rewritten as

$$P_N(\mathbf{x}) \sim \frac{1}{N^{\frac{d}{2}} \left| \sqrt{\underline{\underline{c}}_2} \right|} \left(\int_{-\infty}^{\infty} e^{iz_1} e^{-\frac{w_1^2}{2}} \frac{dw_1}{2\pi} \right) \cdot \left(\int_{-\infty}^{\infty} e^{iz_2} e^{-\frac{w_2^2}{2}} \frac{dw_2}{2\pi} \right) \dots \quad (15)$$

Noting that integrals in the above equation are inverse Fourier Transforms of the Gaussian $e^{-\frac{w^2}{2}}$, we simplify the expression to

$$P_N(\mathbf{x}) \sim \frac{1}{N^{\frac{d}{2}} \sqrt{|\underline{c}_2|}} \left(\frac{e^{-\frac{z_1^2}{2}}}{\sqrt{2\pi}} \right) \cdot \left(\frac{e^{-\frac{z_2^2}{2}}}{\sqrt{2\pi}} \right) \cdots \quad (16)$$

$$P_N(\mathbf{x}) \sim \frac{e^{-\frac{\mathbf{z}^2}{2}}}{(2\pi N)^{\frac{d}{2}} \sqrt{|\underline{c}_2|}} \quad (17)$$

The transform to \mathbf{z} can also be viewed as the definition of a new random variable. In the last few steps, we have actually shown that the scaled random variable

$$\mathbf{Z}_N = (\mathbf{X} - N\underline{c}_1)(N\underline{c}_2)^{-\frac{1}{2}} = (\mathbf{X}_N - \langle \mathbf{X}_N \rangle) \cdot \underline{c}_{2,N}^{-1/2} \quad (18)$$

has a PDF, $\phi_N(\mathbf{z})$, that tends to a multivariate standard normal distribution

$$\phi_N(\mathbf{z}) \sim \frac{e^{-\mathbf{z}^2/2}}{(2\pi)^{d/2}}. \quad (19)$$

This result is known as the Multidimensional Central Limit Theorem (CLT).

Note 2: $P_N(\mathbf{x}) d^d \mathbf{x} = \phi_N(\mathbf{z}) d^d \mathbf{z}$

Note 3: For $d = 1$, $Z_N = \frac{X - Nc_1}{\sigma\sqrt{N}}$ since $c_2 = \sigma^2$. The CLT only applies to the "central region" of $P_N(\mathbf{x})$ where $\mathbf{z} = O(1)$. This implies that in multiple dimensions $vcX_N = N\underline{c}_1 + O(\sqrt{N\underline{c}_2})$ which reduces to $X_N = Nc_1 + O(\sigma\sqrt{N})$ when $d = 1$.

Note 4: Note: An explicit assumption throughout has been the existence of \underline{c}_1 and \underline{c}_2 . If \underline{c}_2 is not finite, the CLT does not hold.

Note 5: For a more rigorous derivation of the CLT, the reader is referred to W. Feller, *An Introduction to Theory of Probability*. Vol. I,II (1970). Gnedenko, Kolmogorov (1968).

Note 6: Equation (19) is an asymptotic solution and it appears that the mean, \underline{c}_1 , and variance, \underline{c}_2 , of the distribution are of importance. However as we will see in the next section and in subsequent lectures, higher cumulants become important when we either look at smaller N or if we investigate the tails of the distribution.

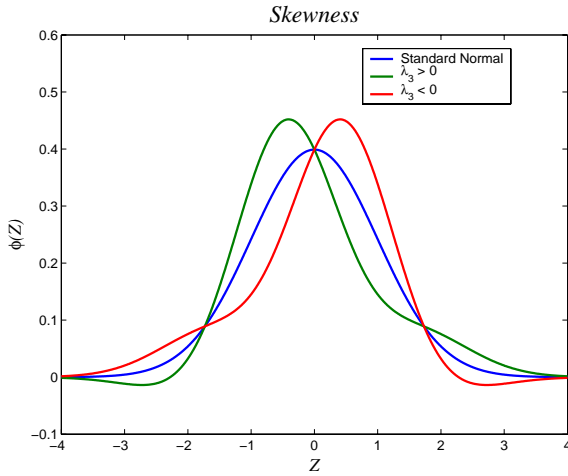


Figure 2: Effect of skewness on distribution shape.

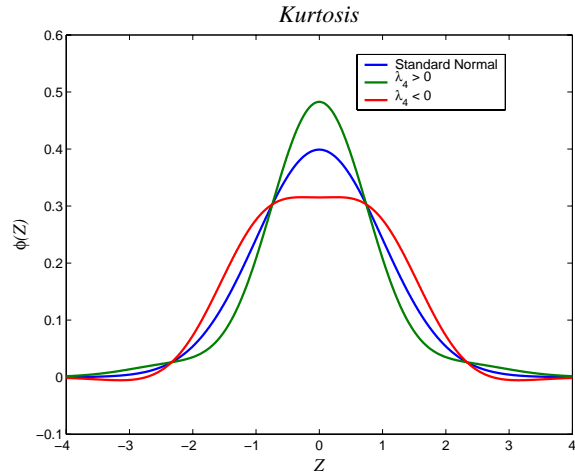


Figure 3: Effect of kurtosis on distribution shape.

Convergence in the Central Region

For simplicity, let's consider $d=1$, though it is not difficult to generalize to higher dimensions. We begin again with the distribution

$$P_N(x) = \int_{-\infty}^{\infty} e^{ikx} e^{N\psi(k)} \frac{dk}{2\pi}$$

We apply Laplace's Method to higher orders of asymptotic expansion as $N \rightarrow \infty$ (Z_N remains fixed). The first step is to note that the dominate contribution to the integral occurs near $k = 0$, so we Taylor expand $\psi(k)$ about the origin producing

$$P_N(x) = \int_{-\epsilon}^{\epsilon} e^{ikx} e^{N[-ic_1k - \frac{1}{2}c_2k^2 + \frac{i}{3!}c_3k^3 + \frac{1}{4!}c_4k^4 + \dots]} \frac{dk}{2\pi} \tag{20}$$

where the imaginary numbers result from the definition of the cumulants. Substituting $w = \sigma\sqrt{N}k$ and $z = \frac{x - Nc_1}{\sigma\sqrt{N}}$, the above equation reduces to

$$\phi_N(z) \sim \int_{-\epsilon}^{\epsilon} e^{-i\omega z} e^{-\frac{\omega^2}{2}} e^{\frac{\lambda_3}{3!} \frac{(i\omega)^3}{\sqrt{N}} + \frac{\lambda_4}{4!} \frac{(i\omega)^4}{N} + \dots} \frac{d\omega}{2\pi}, \tag{21}$$

where $\lambda_3 = \frac{c_3}{\sigma^3}$, $\lambda_4 = \frac{c_4}{\sigma^4}$, ... $\lambda_m = \frac{c_m}{\sigma^m}$ are the normalized cumulants of $p(x)$. The normalized cumulants play a large role in determining the distribution shape (Fig. 2 and Fig. 3). For sufficiently large N , we can truncate the Taylor series and then expand the limits of integration back to $-\infty$ to ∞ . That is, as $N \rightarrow \infty$

$$\phi_N(z) \sim \int_{-\infty}^{\infty} e^{i\omega z} e^{-\frac{\omega^2}{2}} \left[1 + \frac{\lambda_3}{3!} \frac{(i\omega)^3}{\sqrt{N}} + \frac{\lambda_4}{4!} \frac{(i\omega)^4}{N} + \frac{1}{2} \left(\frac{\lambda_3}{3!} \right)^2 \frac{(i\omega)^6}{N} + O\left(\frac{\omega^5}{N\sqrt{N}} \right) \right] \frac{d\omega}{2\pi}. \tag{22}$$

The topics in the last part of this lecture where presented more thoroughly in 2003. To the readers benefit, I have included part of the 2003 notes by David Vener.

Noticing that

$$\int_{-\infty}^{\infty} (i\omega)^n e^{-i\omega z} e^{-\frac{\omega^2}{2}} = (-1)^n \frac{d^n}{dz^n} \int_{-\infty}^{\infty} e^{-i\omega z} e^{-\frac{\omega^2}{2}}$$

allows us to rewrite Equation (22) as

$$\phi_N(z) \sim \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \left[1 + \frac{\lambda_3}{3!} \frac{1}{\sqrt{N}} H_3(z) + \frac{1}{N} \left\{ \frac{\lambda_4}{4!} H_4(z) + \frac{1}{2} \left(\frac{\lambda_3}{3!} \right)^2 H_6(z) \right\} + O\left(\frac{z^5}{N\sqrt{N}}\right) \right], \quad (23)$$

where $H_n(z)$ is the n^{th} Hermite polynomial defined by

$$\frac{d^n}{dz^n} e^{-\frac{z^2}{2}} \equiv (-1)^n H_n(z) e^{-\frac{z^2}{2}}.$$

Note 7: The Hermite polynomials come up in many different places in mathematical physics, including, perhaps most prominently, in the eigenfunctions for the quantum harmonic oscillator. The first few Hermite polynomials are

$$\begin{aligned} H_1 &= z, \\ H_2 &= z^2 - 1, \\ H_3 &= z^3 - 3z, \\ H_4 &= z^4 - 6z^2 + 3, \\ H_5 &= z^5 - 10z^3 + 15z, \\ H_6 &= z^6 - 15z^4 + 15z^2 - 15. \end{aligned}$$

Therefore, as $N \rightarrow \infty$ we can write Equation (23) as

$$\phi_N(z) \sim \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \left[1 + \frac{h_3(z)}{\sqrt{N}} + \frac{h_4(z)}{N} + \dots \right], \quad (24)$$

where

$$h_3(z) = \frac{\lambda_3}{3!} H_3(z), \quad (25)$$

$$h_4(z) = \frac{\lambda_4}{4!} H_4(z) + \frac{1}{2} \left(\frac{\lambda_3}{3!} \right)^2 H_6(z). \quad (26)$$

Note 8: In 1 dimension, this general expansion related to the CLT is often called the ‘Edgeworth expansion’ in the mathematical literature. (See, e.g. W. Feller, *An Introduction to Probability*, Vol. 2.) In more dimensions when related to random flights, expansions of this type are sometimes called ‘Gram-Charlier Expansions’. (See, e.g. B. Hughes, *Random Walks and Random Environments*.)