Chapter 15

Probability and Moment Approximations Using Limit Theorems

15.1 Introduction

So far we have described the methods for determining the exact probability of events using probability mass functions (PMFs) for discrete random variables and probability density functions (PDFs) for continuous random variables. Also of importance were the methods to determine the moments of these random variables. The procedures employed were all based on knowledge of the PMF/PDF and the implementation of its summation/integration. In many practical situations the PMF/PDF may be unknown or the summation/integration may not be easily carried out. It would be of great utility, therefore, to be able to approximate the desired quantities using much simpler methods. For random variables that are the sum of a *large* number of independent and identically distributed random variables this can be done. In this chapter we focus our discussions on two very powerful theorems in probability—*the law of large numbers* and *the central limit theorem.* The first theorem asserts that the sample mean random variable, which is the average of IID random variables and which was introduced in Chapter 14, converges to the *expected value*, a number, of each random variable in the average. The law of large numbers is also known colloquially as the *law of averages.* Another reason for its importance is that it provides a justification for the relative frequency interpretation of probability. The second theorem asserts that a properly normalized sum of IID random variables converges to a Gaussian *random variable.*

The theorems are actually the simplest forms of much more general results. For example, the theorems can be formulated to handle sums of nonidentically distributed random variables [Rao 1973] and dependent random variables [Brockwell and Davis 1987].

15.2 Summary

The Bernoulli law of large number is introduced in Section 15.4 as a prelude to the more general law of large numbers. The latter is summarized in Theorem 15.4.1 and asserts that the sample mean random variable of IID random variables will converge to the expected value of a single random variable. The central limit theorem is described in Section 15.5 where it is demonstrated that the repeated convolution of PDFs produces a Gaussian PDF. For continuous random variables the central limit theorem, which asserts that the sum of a large number of IID random variables has a Gaussian PDF, is summarized in Theorem 15.5.1. The precise statement is given by (15.6). For the sum of a large number of IID discrete random variables it is the CDF that converges to a Gaussian CDF. Theorem 15.5.2 is the central limit theorem for discrete random variables. The precise statement is given by (15.9). The concept of confidence intervals is introduced in Section 15.6. A 95% confidence interval for the sample mean estimate of the parameter p of a $\text{Ber}(p)$ random variable is given by (15.14). It is then applied to the real-world problem of opinion polling.

15.3 Convergence and Approximation of a Sum

Since we will be dealing with the sum of a large number of random variables, it is worthwhile first to review some concepts of convergence. In particular, we need to understand the role that convergence plays in approximating the behavior of a sum of terms. As an illustrative example, consider the determination of the value of the sum

$$s_N = \sum_{i=1}^{N} \frac{(1+a^i)}{N}$$

for some large value of N. We have purposedly chosen a sum that may be evaluated in closed form to allow a comparison to its approximation. The exact value can be found as

$$\begin{aligned} s_N &= \frac{1}{N}\sum_{i=1}^{N} 1 + \frac{1}{N}\sum_{i=1}^{N} a^i \\ &= 1 + \frac{1}{N}\frac{a - a^{N+1}}{1-a}. \end{aligned}$$

Examples of s_N versus N are shown in Figure 15.1. The values of s_N have been connected by straight lines for easier viewing. It should be clear that as $N \to \infty$, $s_N \to 1$ if $|a| < 1$. This means that if N is sufficiently large, then s_N will differ from 1 by a very small amount. This small amount, which is the error in the approximation of s_N by 1, is given by

$$|s_N - 1| = \frac{1}{N}\frac{a - a^{N+1}}{1-a}$$

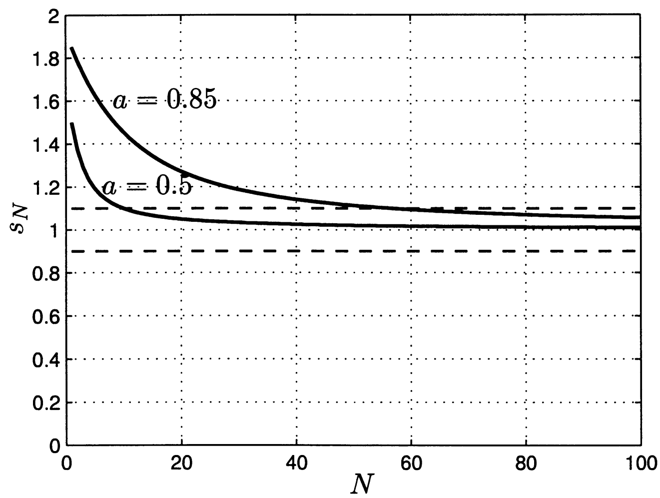


Figure 15.1: Convergence of sum to 1.

and will depend on a as well as N. For example, if we wish to claim that the error is less than 0.1, then N would have to be 10 for $a = 0.5$ but N would need to be 57 for $a = 0.85$, as seen in Figure 15.1. Thus, in general the error of the approximation will depend upon the particular sequence (value of a here). We can assert, without actually knowing the value of a as long as $|a| < 1$ and hence the sum converges, that s_N will eventually become close to 1. The error can be quite large for a fixed value of N (consider what would happen if $a = 0.999$). Such are the advantages (sum will be close to 1 for *all* $|a| < 1$) and disadvantages (how large does N have to be?) of limit theorems. We next describe the law of large numbers.

15.4 Law of Large Numbers

When we began our study of probability, we argued that if a fair coin is tossed N times in succession, then the relative frequency of heads, i.e., the number of heads observed divided by the number of coin tosses, should be close to 1/2. This was why we intuitively accepted the assignment of a probability of 1/2 to the event that the outcome of a fair coin toss would be a head. If we continue to toss the coin, then as $N \to \infty$, we expect the relative frequency to approach 1/2. We can now prove that this is indeed the case under certain assumptions. First we model the repeated coin toss experiment as a sequence of N Bernoulli subexperiments (see also Section 4.6.2). The result of the ith subexperiment is denoted by the discrete random variable X_i, where

$$X_i = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails.} \end{cases}$$

We can then model the overall experimental output by the random vector $\mathbf{X} = [X_1\, X_2 \ldots X_N]^T$. We next assume that the discrete random variables X_i are IID with marginal PMF

$$p_X[k] = \begin{cases} \frac{1}{2} & k=0 \\ \frac{1}{2} & k=1 \end{cases}$$

or the experiment is a sequence of *independent* and identical Bernoulli subexperiments. Finally, the relative frequency is given by the sample mean random variable

$$\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{15.1}$$

which was introduced in Chapter 14, although there it was used for the average of *continuous* random variables. We subscript the sample mean random variable by N to remind us that N coin toss outcomes are used in its computation. Now consider what happens to the mean and variance of $\bar{X}_N$ as $N \to \infty$. The mean is

$$\begin{aligned} E_{\mathbf{X}}[\bar{X}_N] &= \frac{1}{N}\sum_{i=1}^{N} E_{\mathbf{X}}[X_i] \\ &= \frac{1}{N}\sum_{i=1}^{N} E_{X_i}[X_i] \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2} \\ &= \frac{1}{2} \quad \text{for all } N. \end{aligned}$$

The variance is

$$\begin{aligned} \operatorname{var}(\bar{X}_N) &= \operatorname{var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{N} \operatorname{var}(X_i) \quad (X_i\text{'s are independent} \Rightarrow \text{uncorrelated}) \\ &= \frac{\operatorname{var}(X_i)}{N} \quad (X_i\text{'s are identically distributed} \\ &\qquad\qquad \Rightarrow \text{have same variance}). \end{aligned}$$

But for a Bernoulli random variable, $X_i \sim \operatorname{Ber}(p)$, the variance is $\operatorname{var}(X_i) = p(1-p)$. Since $p = 1/2$ for a fair coin,

$$\begin{aligned} \operatorname{var}(\bar{X}_N) &= \frac{p(1-p)}{N} \\ &= \frac{1}{4N} \to 0 \quad \text{as } N \to \infty. \end{aligned}$$

Therefore the width of the PMF of $\bar{X}_N$ must decrease as N increases and eventually go to zero. Since the variance is defined as

$$\operatorname{var}(\bar{X}_N) = E_{\mathbf{X}}\left[\left(\bar{X}_N - E_{\mathbf{X}}[\bar{X}_N]\right)^2\right]$$

we must have that as $N \to \infty$, $\bar{X}_N \to E_{\mathbf{X}}[\bar{X}_N] = 1/2$. In effect the random variable $\bar{X}_N$ becomes not random at all but a constant. It is called a *degenerate* random variable. To further verify that the PMF becomes concentrated about its mean, which is 1/2, we note that the sum of N IID Bernoulli random variables is a binomial random variable. Thus,

$$S_N = \sum_{i=1}^{N} X_i \sim \operatorname{bin}\left(N, \frac{1}{2}\right)$$

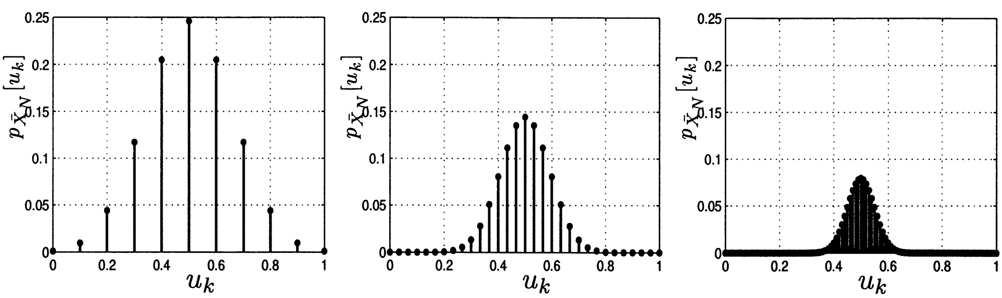
and therefore the PMF is

$$p_{S_N}[k] = \binom{N}{k}\left(\frac{1}{2}\right)^N \qquad k = 0, 1, \ldots, N.$$

To find the PMF of $\bar{X}_N$ we let $\bar{X}_N = (1/N)\sum_{i=1}^{N} X_i = S_N/N$ and note that $\bar{X}_N$ can take on values $u_k = k/N$ for $k = 0, 1, \ldots, N$. Therefore, using the formula for the transformation of a discrete random variable, the PMF becomes

$$p_{\bar{X}_N}[u_k] = \binom{N}{Nu_k}\left(\frac{1}{2}\right)^N \qquad u_k = k/N;\ k = 0, 1, \ldots, N \tag{15.2}$$

which is plotted in Figure 15.2 for various values of N. Because as N increases $\bar{X}_N$



(a) $N = 10$ (b) $N = 30$ (c) $N = 100$

Figure 15.2: PMF for sample mean random variable of N IID Bernoulli random variables with $p = 1/2$. It models the relative frequency of heads obtained for N fair coin tosses.

takes on values more densely in the interval $[0, 1]$, we do not obtain a PMF with all its mass concentrated at 0.5, as we might expect. Nonetheless, the probability that the sample mean random variable will be concentrated about 1/2 increases. As an example, the probability of being within the interval $[0.45, 0.55]$ is 0.2461 for $N = 10$, 0.4153 for $N = 30$, and 0.7287 for $N = 100$, as can be verified by summing the values of the PMF over this interval. Usually it is better to plot the CDF since as $N \to \infty$, it can be shown to converge to the unit step beginning at $u = 0.5$ (see Problem 15.1). Also, it is interesting to note that the PMF appears Gaussian, although it changes in amplitude and width for each N. This is an observation that we will focus on later when we discuss the central limit theorem. The preceding results say that for large enough N the sample mean random variable will *always* yield a *number*, which in this case is 1/2. By "always" we mean that every time we perform a repeated Bernoulli experiment consisting of N independent and fair coin tosses, we will obtain a sample mean of 1/2, *for N large enough.* As an example, we have plotted in Figure 15.3 five realizations of the sample mean random variable or $\bar{x}_N$ versus N. The values of $\bar{x}_N$ have been connected by straight lines for easier viewing. We see that

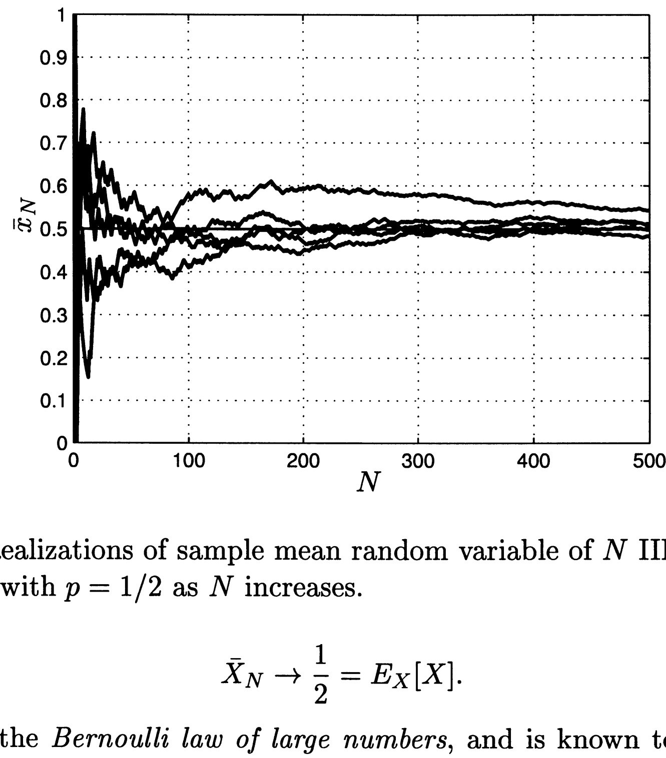


Figure 15.3: Realizations of sample mean random variable of N IID Bernoulli random variables with $p = 1/2$ as N increases.

$$\bar{X}_N \to \frac{1}{2} = E_X[X]. \tag{15.3}$$

This is called the *Bernoulli law of large numbers*, and is known to the layman as *the law of averages.* More generally for a Bernoulli subexperiment with probability p, we have that

$$\bar{X}_N \to p = E_X[X].$$

The sample mean random variable converges to the expected value of a single random variable. Note that since $\bar{X}_N$ is the relative frequency of heads and p is the probability of heads, we have shown that *the probability of a head in a single coin toss can be interpreted as the value obtained as the relative frequency of heads in a large number of independent and identical coin tosses.* This observation also justifies our use of the sample mean random variable as an estimator of a moment since

$$\widehat{E_X[X]} = \frac{1}{N}\sum_{i=1}^{N} X_i \to E_X[X] \qquad \text{as } N \to \infty$$

and more generally, justifies our use of $(1/N)\sum_{i=1}^{N} X_i^n$ as an estimate of the nth moment $E[X^n]$ (see also Problem 15.6).

A more general law of large numbers is summarized in the following theorem. It is valid for the sample mean of IID random variables, either discrete, continuous, or mixed.

Theorem 15.4.1 (Law of Large Numbers) *If $X_1, X_2, \ldots, X_N$ are IID random variables with mean $E_X[X]$ and $\mathrm{var}(X) = \sigma^2 < \infty$, then $\lim_{N\to\infty} \bar{X}_N = E_X[X]$.*

Proof:
Consider the probability of the sample mean random variable deviating from the expected value by more than ϵ, where ϵ is a small positive number. This probability is given by

$$P\left[|\bar{X}_N - E_X[X]| > \epsilon\right] = P\left[|\bar{X}_N - E_{\mathbf{X}}[\bar{X}_N]| > \epsilon\right].$$

Since $\mathrm{var}(\bar{X}_N) = \sigma^2/N$, we have upon using Chebyshev's inequality (see Section 11.8)

$$P\left[|\bar{X}_N - E_X[X]| > \epsilon\right] \le \frac{\mathrm{var}(\bar{X}_N)}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2}$$

and taking the limit of both sides yields

$$\lim_{N\to\infty} P\left[|\bar{X}_N - E_X[X]| > \epsilon\right] \le \lim_{N\to\infty} \frac{\sigma^2}{N\epsilon^2} = 0.$$

Since a probability must be greater than or equal to zero, we have finally that

$$\lim_{N\to\infty} P\left[|\bar{X}_N - E_X[X]| > \epsilon\right] = 0 \qquad (15.4)$$

which is the mathematical statement that the sample mean random variable converges to the expected value of a single random variable.

□

The limit in (15.4) says that for large enough N, the *probability of the error* in the approximation of $\bar{X}_N$ by $E_X[X]$ exceeding ϵ (which can be chosen as small as desired) will be exceedingly small. It is said that $\bar{X}_N \to E_X[X]$ *in probability* [Grimmett and Stirzaker 2001].

Convergence in probability does not mean all realizations will converge.

Referring to Figure 15.3 it is seen that for all realizations except the top one, the error is small. The statement of (15.4) does allow some realizations to have an error greater than ϵ for a given large N. However, the probability of this happening becomes very small but not zero as N increases. For all practical purposes, then, we can ignore this occurrence. Hence, convergence in probability is somewhat different than what one may be familiar with in dealing with convergence of *deterministic sequences.* For deterministic sequences, all sequences (since there is only one) will have an error less than ϵ for all $N \geq N_\epsilon$, where N_ϵ will depend on ϵ (see Figure 15.1). The interested reader should consult [Grimmett and Stirzaker 2001] for further details. See also Problem 15.8 for an example.

We conclude our discussion with an example and some further comments.

Example 15.1 – Sample mean for IID Gaussian random variables

Recall from the real-world example in Chapter 14 that when a signal is present we have

$$X_{s+W_i} \sim \mathcal{N}(A, \sigma^2) \quad i = 1, 2, \ldots, N.$$

Since the random variables are IID, we have by the law of large numbers that

$$\bar{X}_N \to E_X[X] = A.$$

Thus, the upper curve shown in Figure 14.6 must approach $A = 0.5$ (with high probability) as $N \to \infty$.

◇

In applying the law of large numbers *we do not need to know the marginal PDF.* If in the previous example, we had $X_{s+W_i} \sim \mathcal{U}(0, 2A)$, then we also conclude that $\bar{X}_N \to A$. As long as the random variables are IID with mean A and a finite variance, $\bar{X}_N \to A$ (although the error in the approximation will depend upon the marginal PDF—see Problem 15.3).

15.5 Central Limit Theorem

By the law of large numbers the PMF/PDF of the sample mean random variable decreases in width until all the probability is concentrated about the mean. The theorem, however, does not say much about the PMF/PDF itself. However, by considering a slightly modified sample mean random variable, we can make some more definitive assertions about its probability distribution. To illustrate the necessity of doing so we consider the PDF of a continuous random variable that is the sum of N continuous IID random variables. A particularly illustrative example is for $X_i \sim \mathcal{U}(-1/2, 1/2)$.

Example 15.2 – PDF for sum of IID $\mathcal{U}(-1/2, 1/2)$ random variables

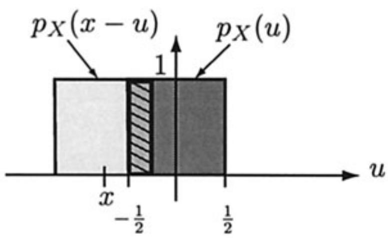
Consider the sum

$$S_N = \sum_{i=1}^{N} X_i$$

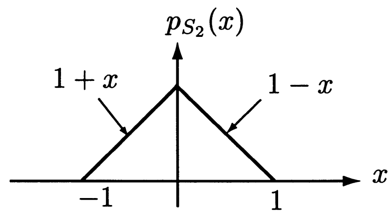
where the X_i's are IID random variables with $X_i \sim \mathcal{U}(-1/2, 1/2)$. If $N = 2$, then $S_2 = X_1 + X_2$ and the PDF of S_2 is easily found using a convolution integral as described in Section 12.6. Therefore,

$$p_{S_2}(x) = p_X(x) \star p_X(x) = \int_{-\infty}^{\infty} p_X(u) p_X(x-u) du$$

where $\star$ denotes convolution. The evaluation of the convolution integral is most easily done by plotting $p_X(u)$ and $p_X(x-u)$ versus u as shown in Figure 15.4a. This is necessary to determine the regions over which the product of $p_X(u)$ and $p_X(x-u)$ is nonzero and so contributes to the integral. The reader should be able to show, based upon Figure 15.4a, that the PDF of S_2 is that shown in Figure 15.4b. More generally, we have from (14.22) that



(a) Cross-hatched region contributes to integral



(b) Result of convolution

Figure 15.4: Determining the PDF for the sum of two independent uniform random variables using a convolution integral evaluation.

$$\begin{aligned} p_{S_N}(x) &= \int_{-\infty}^{\infty} \phi_X^N(\omega) \exp(-j\omega x) \frac{d\omega}{2\pi} \\ &= \underbrace{p_X(x) \star p_X(x) \star \cdots \star p_X(x)}_{(N-1) \text{ convolutions}}. \end{aligned}$$

Hence to find $p_{S_3}(x)$ we must convolve $p_{S_2}(x)$ with $p_X(x)$ to yield $p_X(x) \star p_X(x) \star p_X(x)$ since $p_{S_2}(x) = p_X(x) \star p_X(x)$. This is

$$p_{S_3}(x) = \int_{-\infty}^{\infty} p_{S_2}(u) p_X(x-u) du$$

but since $p_X(-x) = p_X(x)$, we can express this in the more convenient form as

$$p_{S_3}(x) = \int_{-\infty}^{\infty} p_{S_2}(u) p_X(u-x)\, du.$$

The integrand may be determined by plotting $p_{S_2}(u)$ and the right-shifted version $p_X(u-x)$ and multiplying these two functions. The different regions that must be considered are shown in Figure 15.5. Hence, referring to Figure 15.5 we have

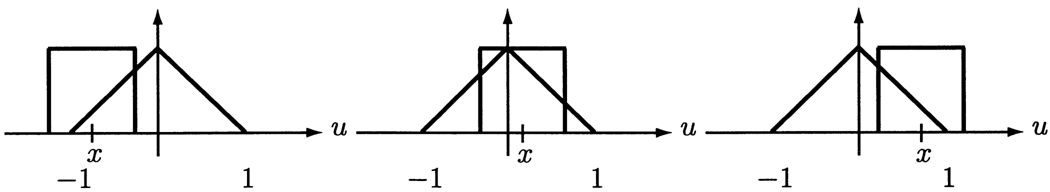


Figure 15.5: Determination of limits for convolution integral.

$$\begin{aligned}
p_{S_3}(x) &= \int_{-1}^{x+1/2} p_{S_2}(u) \cdot 1 du \\
&= \frac{1}{2}x^2 + \frac{3}{2}x + \frac{9}{8} & -\frac{3}{2} \leq x \leq -\frac{1}{2} \\
p_{S_3}(x) &= \int_{x-1/2}^{x+/2} p_{S_2}(u) \cdot 1 du \\
&= -x^2 + \frac{3}{4} & -\frac{1}{2} < x \leq \frac{1}{2} \\
p_{S_3}(x) &= \int_{x-1/2}^{1} p_{S_2}(u) \cdot 1 du \\
&= \frac{1}{2}x^2 - \frac{3}{2}x + \frac{9}{8} & \frac{1}{2} < x \leq \frac{3}{2}
\end{aligned}$$

and $p_{S_3}(x) = 0$ otherwise. This is plotted in Figure 15.6 versus the PDF of a $\mathcal{N}(0, 3/12)$ random variable. Note the close agreement. We have chosen the mean and variance of the Gaussian approximation to *match* that of $p_{S_3}(x)$ (recall that $\text{var}(X) = (b-a)^2/12$ for $X \sim \mathcal{U}(a,b)$ and hence $\text{var}(X_i) = 1/12$). If we continue the convolution process, the mean will remain at zero but the variance of S_N will be $N/12$.

◇

A MATLAB program that implements a repeated convolution for a PDF that is nonzero over the interval $(0,1)$ is given in Appendix 15A. It can be used to verify

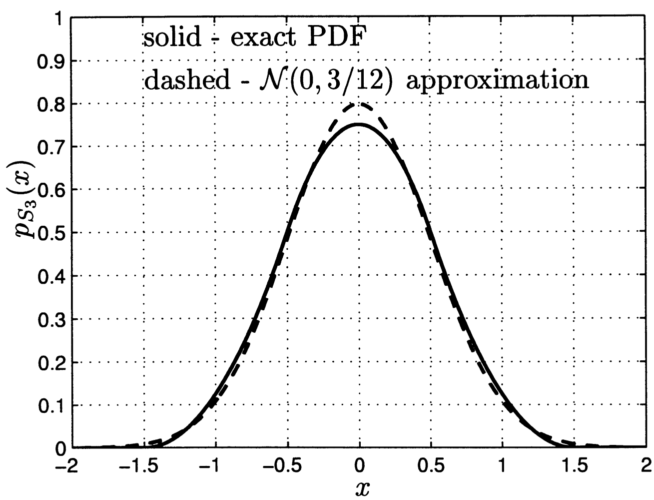
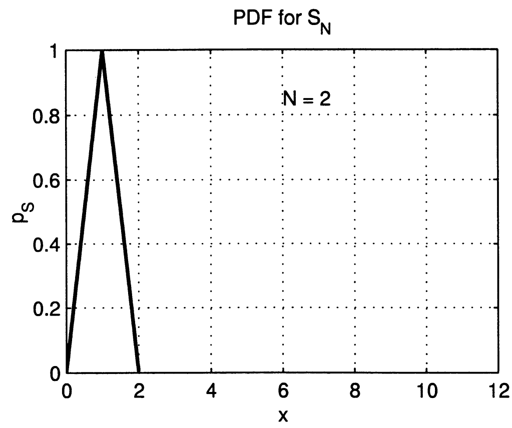


Figure 15.6: PDF for sum of 3 IID $\mathcal{U}(-1/2,1/2)$ random variables and Gaussian approximation.

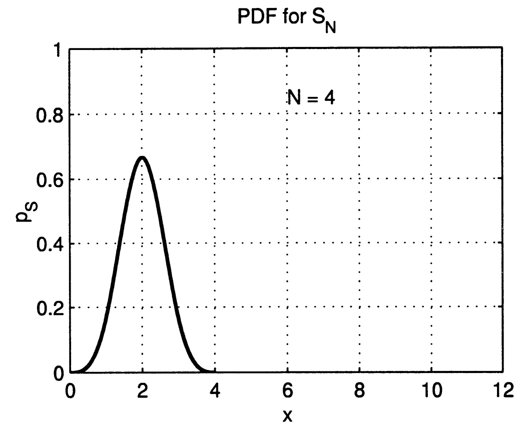
analytical results and also to try out other PDFs. An example of its use is shown in Figure 15.7 for the repeated convolution of a $\mathcal{U}(0,1)$ PDF. Note that as N increases the PDF moves to the right since $E[S_N] = NE_X[X] = N/2$ and the variance also increases since $\text{var}(S_N) = N\text{var}(X) = N/12$. Because of this behavior it is not possible to state that the PDF converges to *any* PDF. To circumvent this problem it is necessary to *normalize* the sum so that its mean and variance are fixed as N increases. It is convenient, therefore, to have the mean fixed at 0 and the variance fixed at 1, resulting in a *standardized sum.* Recall from Section 7.9 that this is easily accomplished by forming

$$\frac{S_N - E[S_N]}{\sqrt{\text{var}(S_N)}} = \frac{S_N - NE_X[X]}{\sqrt{N\text{var}(X)}}. \tag{15.5}$$

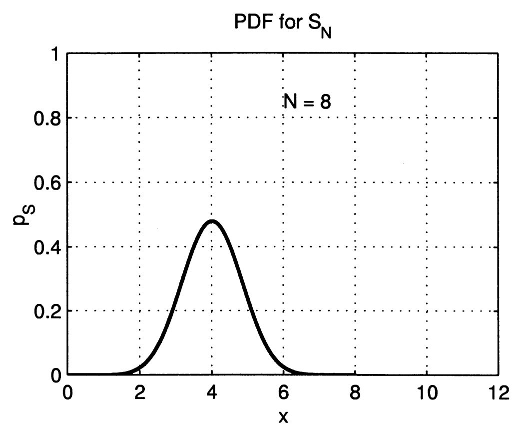
By doing so, we can now assert that this standardized random variable will converge to a $\mathcal{N}(0,1)$ random variable. An example is shown in Figure 15.8 for $X_i \sim \mathcal{U}(0,1)$ and for $N = 2,3,4$. This is the famous *central limit theorem*, which says that *the PDF of the standardized sum of a large number of continuous IID random variables will converge to a Gaussian PDF.* Its great importance is that in many practical situations one can model a random variable as having arisen from the contributions of many small and similar physical effects. By making the IID assumption we can assert that the PDF is Gaussian. There is no need to know the PDF of each random variable or even if it is known, to determine the exact PDF of the sum, which may not be possible. Some application areas are:



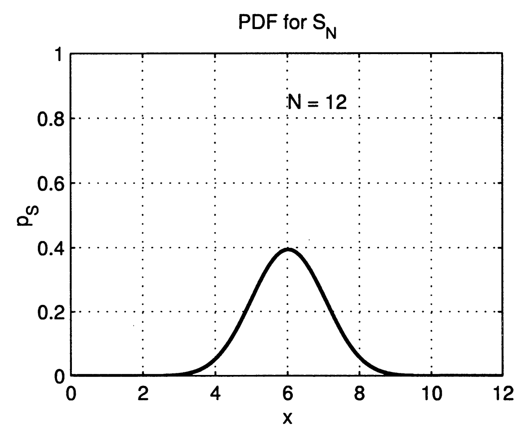
(a) p_X



(b) p_{S_4}



(c) p_{S_8}



(d) $p_{S_{12}}$

Figure 15.7: PDF of sum of N IID $\mathcal{U}(0,1)$ random variables. The plots were obtained using `clt_demo.m` listed in Appendix 15A.

1. Polling (see Section 15.6) [Weisburg, Krosnick, Bowen 1996]

2. Noise characterization [Middleton 1960]

3. Scattering effects modeling [Urick 1975]

4. Kinetic theory of gases [Reif 1965]

5. Economic modeling [Harvey 1989]

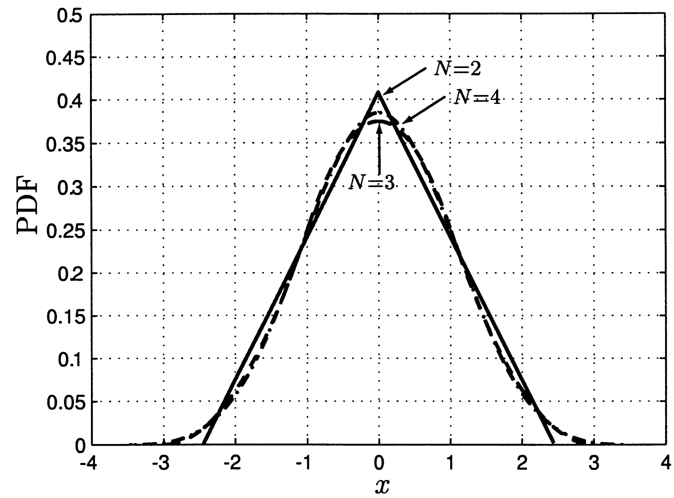


Figure 15.8: PDF of standardized sum of N IID $\mathcal{U}(0,1)$ random variables.

and many more.

We now state the theorem for continuous random variables.

Theorem 15.5.1 (Central limit theorem for continuous random variables) *If $X_1, X_2, \ldots, X_N$ are continuous IID random variables, each with mean $E_X[X]$ and variance* $\text{var}(X)$, *and $S_N = \sum_{i=1}^N X_i$, then as $N \to \infty$*

$$\frac{S_N - E[S_N]}{\sqrt{\text{var}(S_N)}} = \frac{\sum_{i=1}^N X_i - NE_X[X]}{\sqrt{N\text{var}(X)}} \to \mathcal{N}(0,1). \tag{15.6}$$

Equivalently, the CDF of the standardized sum converges to $\Phi(x)$ or

$$P\left[\frac{S_N - E[S_N]}{\sqrt{\text{var}(S_N)}} \le x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt = \Phi(x). \tag{15.7}$$

The proof is given in Appendix 15B and is based on the properties of characteristic functions and the continuity theorem. An example follows.

Example 15.3 – PDF of sum of squares of independent $\mathcal{N}(0,1)$ random variables

Let $X_i \sim \mathcal{N}(0,1)$ for $i = 1, 2, \ldots, N$ and assume that the X_i's are independent. We wish to determine the approximate PDF of $Y_N = \sum_{i=1}^N X_i^2$ as N becomes large. Note that the exact PDF for Y_N is a χ_N^2 PDF so that we will equivalently find an approximation to the PDF of the standardized χ_N^2 random variable. To apply the central limit theorem we first note that since the X_i's are IID so are the X_i^2's (why?). Then as $N \to \infty$ we have from (15.6)

$$\frac{\sum_{i=1}^N X_i^2 - NE_X[X^2]}{\sqrt{N\text{var}(X^2)}} \to \mathcal{N}(0,1).$$

But $X^2 \sim \chi_1^2$ so that $E_X[X^2] = 1$ and $\text{var}(X^2) = 2$ (see Section 10.5.6 and Table 11.1 for a $\chi_N^2 = \Gamma(N/2, 1/2)$ PDF) and therefore

$$\frac{\sum_{i=1}^N X_i^2 - N}{\sqrt{2N}} \to \mathcal{N}(0,1).$$

Noting that for finite N this result can be viewed as an approximation, we can use the approximate result

$$Y_N = \sum_{i=1}^N X_i^2 \sim \mathcal{N}(N, 2N)$$

in making probability calculations. The error in the approximation is shown in Figure 15.9, where the approximate PDF (shown as the solid curve) of Y_N, which is a $\mathcal{N}(N, 2N)$, is compared to the exact PDF, which is a χ_N^2 (shown as the dashed curve). It is seen that the approximation becomes better as N increases.

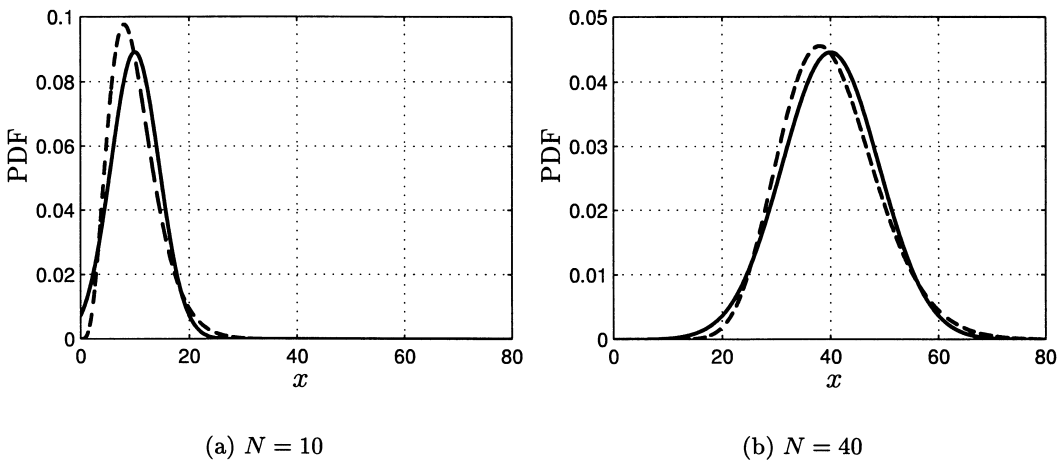


Figure 15.9: χ_N^2 PDF (dashed curve) and Gaussian PDF approximation of $\mathcal{N}(N, 2N)$ (solid curve).

◇

For the previous example it can be shown directly that the characteristic function of the standardized χ_N^2 random variable converges to that of the standardized Gaussian random variable, and hence so do their PDFs by the continuity theorem (see Section 11.7 for third property of characteristic function and also Problem 15.17). We next give an example that quantifies the numerical error of the central limit theorem approximation.

Example 15.4 – Central limit theorem and computation of probabilities—numerical results

Recall that the Erlang PDF is the PDF of the sum of N IID exponential random variables, where $X_i \sim \exp(\lambda)$ for $i = 1, 2, \ldots, N$ (see Section 10.5.6). Hence, letting $Y_N = \sum_{i=1}^{N} X_i$ the Erlang PDF is

$$p_{Y_N}(y) = \begin{cases} \frac{\lambda^N}{(N-1)!} y^{N-1} \exp(-\lambda y) & y \geq 0 \\ 0 & y < 0. \end{cases} \tag{15.8}$$

Its mean is N/λ and its variance is N/λ^2 since the mean and variance of an $\exp(\lambda)$ random variable is $1/\lambda$ and $1/\lambda^2$, respectively. If we wish to determine $P[Y_N > 10]$, then from (15.8) we can find the exact value for $\lambda = 1$ as

$$P[Y_N > 10] = \int_{10}^{\infty} \frac{1}{(N-1)!} y^{N-1} \exp(-y) dy.$$

But using

$$\int y^n \exp(-y) dy = -n! \exp(-y) \sum_{k=0}^{n} \frac{y^k}{k!}$$

[Gradshteyn and Ryzhik 1994], we have

$$\begin{aligned} P[Y_N > 10] &= \frac{1}{(N-1)!} \left[-(N-1)! \exp(-y) \sum_{k=0}^{N-1} \frac{y^k}{k!} \Bigg|_{10}^{\infty} \right] \\ &= \exp(-10) \sum_{k=0}^{N-1} \frac{10^k}{k!}. \end{aligned}$$

A central limit theorem approximation would yield $Y_N \sim \mathcal{N}(N/\lambda, N/\lambda^2) = \mathcal{N}(N, N)$ so that

$$\hat{P}[Y_N > 10] = Q\left(\frac{10 - N}{\sqrt{N}}\right)$$

where the $\hat{P}$ denotes the approximation of P. The true and approximate values for this probability are shown in Figure 15.10. The probability values have been connected by straight lines for easier viewing.

◇

For the sum of IID *discrete* random variables the situation changes markedly. Consider the sum of N IID Ber(p) random variables. We already know that the PMF is binomial so that

$$p_{S_N}[k] = \binom{N}{k} p^k (1-p)^{N-k} \qquad k = 0, 1, \ldots, N-1.$$

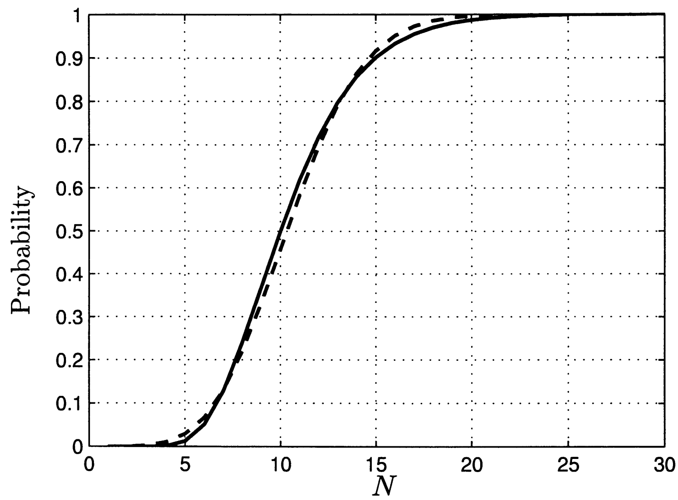
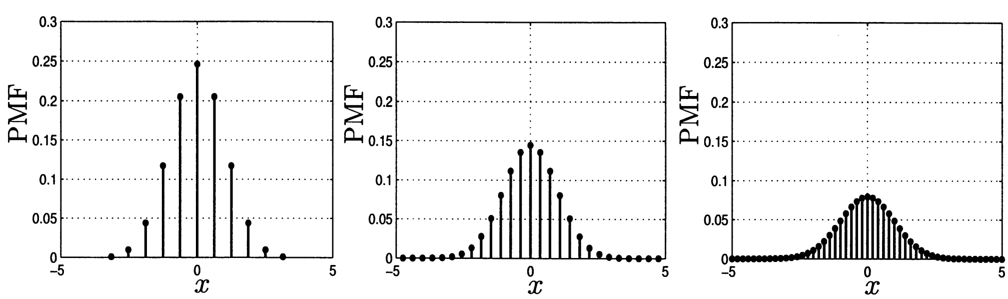


Figure 15.10: Exact and approximate calculation of probability that $Y_N > 10$ for Y_N an Erlang PDF. Exact value shown as dashed curve and Gaussian approximation as solid curve.

Hence, this example will allow us to compare the true PMF against any approximation. For reasons already explained we need to consider the PMF of the standardized sum or

$$\frac{S_N - E[S_N]}{\sqrt{\text{var}(S_N)}} = \frac{S_N - Np}{\sqrt{Np(1-p)}}.$$

The PMF of the standardized binomial random variable PMF with $p = 1/2$ is shown in Figure 15.11 for various values on N. Note that it does not converge to any given



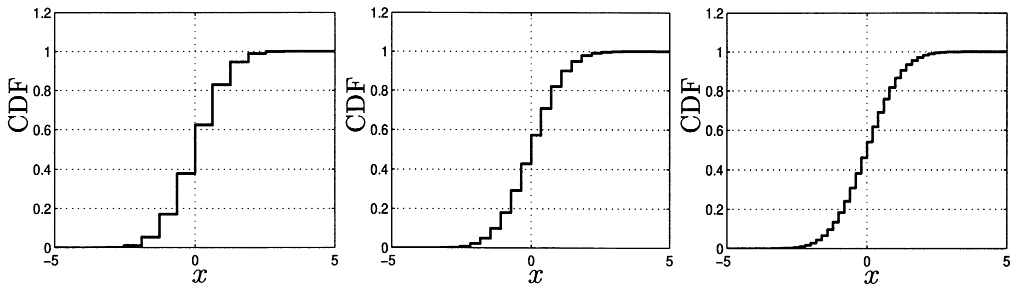
(a) $N = 10$ (b) $N = 30$ (c) $N = 100$

Figure 15.11: PMF for standardized binomial random variable with $p = 1/2$.

PMF, although the "envelope", whose amplitude decreases as N increases, appears to be Gaussian. The lack of convergence is because the sample space or values that the standardized random variable can take on changes with N. The possible values are

$$x_k = \frac{k - Np}{\sqrt{Np(1-p)}} = \frac{k - N/2}{\sqrt{N/4}} \qquad k = 0, 1, \dots, N$$

which become more dense as N increases. However, what does converge is the CDF as shown in Figure 15.12. Now as $N \to \infty$ we can assert that the CDF converges,



(a) $N = 10$ (b) $N = 30$ (c) $N = 100$

Figure 15.12: CDF for standardized binomial random variable with $p = 1/2$.

and furthermore it converges to the CDF of a $\mathcal{N}(0, 1)$ random variable. Hence, the central limit theorem for discrete random variables is stated in terms of its CDF. It says that as $N \to \infty$

$$P\left[\frac{S_N - E[S_N]}{\sqrt{\text{var}(S_N)}} \le x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt = \Phi(x)$$

and is also known as the *DeMoivre-Laplace theorem.* We summarize the central limit theorem for discrete random variables next.

Theorem 15.5.2 (Central limit theorem for discrete random variables) *If $X_1, X_2, \dots, X_N$ are IID discrete random variables, each with mean $E_X[X]$ and variance* $\text{var}(X)$, *and $S_N = \sum_{i=1}^{N} X_i$, then as $N \to \infty$*

$$P\left[\frac{S_N - E[S_N]}{\sqrt{\text{var}(S_N)}} \le x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt = \Phi(x) \qquad (15.9)$$

An example follows.

Example 15.5 – Computation of binomial probability

Assume that $Y_N \sim \text{bin}(N, 1/2)$, which may be viewed as the PMF for the number of heads obtained in N fair coin tosses, and consider the probability $P[k_1 \le Y_N \le k_2]$.

Then the exact probability is

$$P[k_1 \leq Y_N \leq k_2] = \sum_{k=k_1}^{k_2} \binom{N}{k} \left(\frac{1}{2}\right)^N . \tag{15.10}$$

A central limit theorem approximation yields

$$\begin{aligned} P[k_1 \leq Y_N \leq k_2] &= P\left[\frac{k_1 - N/2}{\sqrt{N/4}} \leq \frac{Y_N - N/2}{\sqrt{N/4}} \leq \frac{k_2 - N/2}{\sqrt{N/4}}\right] \\ &\approx \Phi\left(\frac{k_2 - N/2}{\sqrt{N/4}}\right) - \Phi\left(\frac{k_1 - N/2}{\sqrt{N/4}}\right) \qquad \text{(from (10.25))} \end{aligned}$$

since

$$\frac{Y_N - Np}{\sqrt{Np(1-p)}} = \frac{Y_N - N/2}{\sqrt{N/4}}$$

is the standardized random variable for $p = 1/2$. For example, if we wish to compute the probability of between 490 and 510 heads out of $N = 1000$ tosses, then

$$\begin{aligned} P[490 \leq Y_N \leq 510] &\approx \Phi\left(\frac{510 - 500}{\sqrt{250}}\right) - \Phi\left(\frac{490 - 500}{\sqrt{250}}\right) \\ &= 1 - Q\left(\frac{10}{\sqrt{250}}\right) - \left(1 - Q\left(\frac{-10}{\sqrt{250}}\right)\right) \\ &= 1 - 2Q\left(\frac{10}{\sqrt{250}}\right) = 0.4729. \end{aligned}$$

The exact value, however, is from (15.10)

$$P[490 \leq Y_N \leq 510] = \sum_{k=490}^{510} \binom{N}{k} \left(\frac{1}{2}\right)^N = 0.4933 \tag{15.11}$$

(see Problem 15.24 on how this was computed). A slightly better approximation using the central limit theorem can be obtained by replacing $P[490 \leq Y \leq 510]$ with $P[489.5 \leq Y \leq 510.5]$, which will more closely approximate the discrete random variable CDF by the continuous Gaussian CDF. This is because the binomial CDF has jumps at the integers as can be seen by referring to Figure 15.12. By taking a slighter larger interval to be used with the Gaussian approximation, the area under the Gaussian CDF more closely approximates these jumps at the endpoints of the interval. With this approximation we have

$$\begin{aligned} P[489.5 \leq Y \leq 510.5] &\approx Q\left(\frac{489.5 - 500}{\sqrt{250}}\right) - Q\left(\frac{510.5 - 500}{\sqrt{250}}\right) \\ &= 0.4934 \end{aligned}$$

which is quite close to the true value!

◊

15.6 Real-World Example – Opinion Polling

A frequent news topic of interest is the opinion of people on a major issue. For example, during the year of a presidential election in the United States, we hear almost on a daily basis the percentage of people who would vote for candidate A, with the remaining percentage voting for candidate B. It may be reported that 75% of the population would vote for candidate A and 25% would vote for candidate B. Upon reflection, it does not seem reasonable that a news organization would contact the entire population of the United States, almost 294,000,000 people, to determine their voter preferences. And indeed it is unreasonable! A more typical number of people contacted is only about 1000. How then can the news organization report that 75% of the population would vote for candidate A? The answer lies in the polling *error* – the results are actually stated as 75% with *a margin of error of* ±3%. Hence, it is not claimed that exactly 75% of the population would vote for candidate A, but between 72% and 78% would vote for candidate A. Even so, this seems like a lot of information to be gleaned from a very small sample of the population.

An analogous problem may help to unravel the mystery. Let's say we have a coin with an unknown probability of heads p. We wish to estimate p by tossing the coin N times. As we have already discussed, the law of large numbers asserts that we can determine p without error if we toss the coin an infinite number of times and use as our estimate the relative frequency of heads. However, in practice we are limited to only N coin tosses. How much will our estimate be in error? Or more precisely, how much can the true value deviate from our estimate? We know that the number of heads observed in N independent coin tosses can be anywhere from 0 to N. Hence, our estimate of p for $N = 1000$ can take on the possible values

$$\hat{p} = 0, \frac{1}{1000}, \frac{2}{1000}, \ldots, 1.$$

Of course, most of these estimates are not very probable. The probability that the estimate will take on these values is

$$P[\hat{p} = k/1000] = \binom{1000}{k} p^k (1-p)^{1000-k} \quad k = 0, 1, \ldots, 1000$$

which is shown in Figure 15.13 for $p = 0.75$. The probabilities for $\hat{p}$ outside the interval shown are approximately zero. Note that the maximum probability is for the true value $p = 0.75$. To assess the error in the estimate of p we can determine the *interval* over which say 95% of the $\hat{p}$'s will lie. The interval is chosen to be centered about $\hat{p} = 0.75$. In Figure 15.13 it is shown as the interval contained within the dashed vertical lines and is found by solving

$$\sum_{k=k_1}^{k_2} \underbrace{\binom{1000}{k} (0.75)^k (0.25)^{1000-k}}_{P[k \text{ heads}]} = 0.95 \tag{15.12}$$

yielding $k_1 = 724$ and $k_2 = 776$, which results in $\hat{p}_1 = 0.724$ and $\hat{p}_2 = 0.776$. Hence,

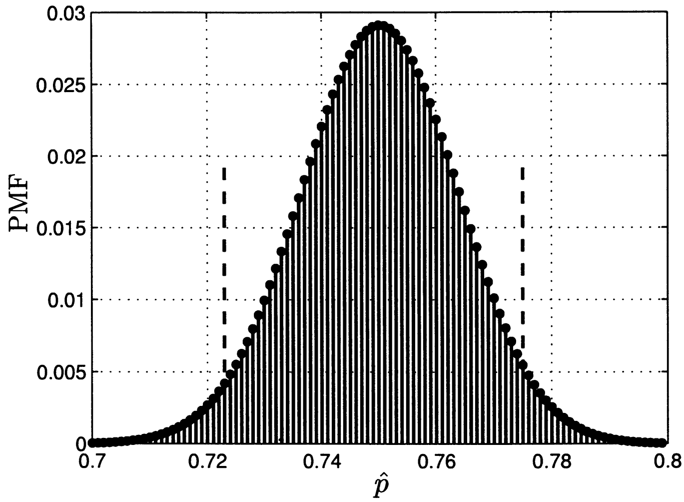


Figure 15.13: PMF for estimate of p for a binomial random variable. Also, shown as the dashed vertical lines are the boundaries of the interval within which 95% of the estimates will lie.

for $p = 0.75$ we see that 95% of the time (if we kept repeating the 1000 coin toss experiment), the value of $\hat{p}$ would be in the interval $[0.724, 0.776]$. We can assert that we are 95% confident that for $p = 0.75$

$$p - 0.026 \le \hat{p} \le p + 0.026$$

or

$$-p + 0.026 \ge -\hat{p} \ge -p - 0.026$$

or finally

$$\hat{p} - 0.026 \le p \le \hat{p} + 0.026.$$

The interval $[\hat{p}-0.026, \hat{p}+0.026]$ is called the 95% *confidence interval.* It is a *random interval* that covers the true value of $p = 0.75$ for 95% of the time. As an example a MATLAB simulation is shown in Figure 15.14. For each of 50 trials the estimate of p is shown by the dot while the confidence interval is indicated by a vertical line. Note that only 3 of the intervals fail to cover the true value of $p = 0.75$. With 50 trials and a probability of 0.95 we expect 2.5 intervals not to cover the true value.

Instead of having to compute k_1 and k_2 using (15.12), it is easier in practice to use the central limit theorem. Since $\hat{p} = \sum_{i=1}^{N}(X_i/N)$, with $X_i \sim \text{Ber}(p)$, is a sum of IID random variables we can assert from Theorem 15.5.2 that

$$P\left[-b \le \frac{\hat{p} - E[\hat{p}]}{\sqrt{\text{var}(\hat{p})}} \le b\right] \approx \Phi(b) - \Phi(-b).$$

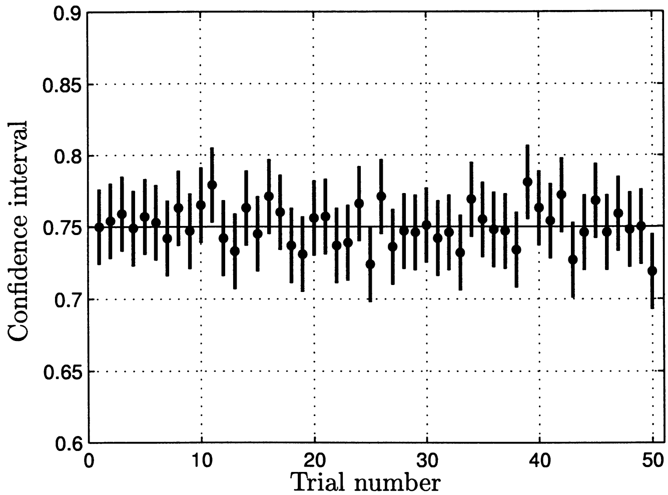


Figure 15.14: 95% confidence interval for estimate of $p = 0.75$ for a binomial random variable. The estimates are shown as dots.

Noting that $X_i \sim \text{Ber}(p)$, $E[\hat{p}] = E[\sum_{i=1}^{N} X_i/N] = Np/N = p$ and $\text{var}(\hat{p}) = \text{var}(\sum_{i=1}^{N} X_i/N) = Np(1-p)/N^2 = p(1-p)/N$, we have

$$P\left[-b \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/N}} \leq b\right] \approx \Phi(b) - \Phi(-b).$$

For a 95% confidence interval or $\Phi(b) - \Phi(-b) = 0.95$, we have $b = 1.96$, as may be easily verified. Hence, we can use the approximation

$$-1.96 \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/N}} \leq 1.96$$

which after the same manipulation as before yields the confidence interval

$$\hat{p} - 1.96\sqrt{\frac{p(1-p)}{N}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{p(1-p)}{N}}. \tag{15.13}$$

The only difficulty in applying this result is that we don't know the value of p, which arose from the variance of $\hat{p}$. To circumvent this there are two approaches. We can replace p by its estimate to yield the confidence interval

$$\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}. \tag{15.14}$$

A more conservative approach is to note that $p(1-p)$ is maximum for $p = 1/2$. Using this number yields a larger interval than necessary. However, it allows us to determine before the experiment is performed and the value of $\hat{p}$ revealed, the length of the confidence interval. This is useful in planning how large N must be in order to have a confidence interval not exceeding a given length (see Problem 15.25). If we adopt the latter approach then the confidence interval becomes

$$\hat{p} \pm 1.96\sqrt{\frac{p(1-p)}{N}} = \hat{p} \pm 1.96\sqrt{\frac{1/4}{N}} \approx \hat{p} \pm \frac{1}{\sqrt{N}}.$$

In summary, if we toss a coin with a probability p of heads N times, then the interval $[\hat{p} - 1/\sqrt{N}, \hat{p} + 1/\sqrt{N}]$ will contain the true value of p more than 95% of the time. It is said that the error in our estimate of p is $\pm 1/\sqrt{N}$.

Finally, returning to our polling problem we ask N people if they will vote for candidate A. The probability that a person chosen at random will say "yes" is p, because the proportion of people in the population who will vote for candidate A is p. We liken this to tossing a single coin and noting if it comes up a head (vote "yes") or a tail (vote "no"). Then we continue to record the responses of N people (continue to toss the coin N times). Assume, for example, 750 people out of 1000 say "yes". Then $\hat{p} = 750/1000 = 0.75$ and the *margin of error* is $\pm 1/\sqrt{N} \approx 3\%$. Hence, we report the results as 75% of the population would vote for candidate A with a margin of error of 3%. (Probabilistically speaking, if we continue to poll groups of 1000 voters, estimating p for each group, then about 95 out of 100 groups would cover the true value of $100p\,\%$ by their estimated interval $[100\hat{p} - 3, 100\hat{p} + 3]\,\%$.) We needn't poll 294,000,000 people since we assume that the *percentage of the 1000 people polled who would vote for candidate A is representative of the percentage of the entire population.* Is this true? Certainly not if the 1000 people were all relatives of candidate A. Pollsters make their living by ensuring that their sample (1000 people polled) is a representative cross-section of the entire United States population.

References

Brockwell, P.J., R.A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.

Gradshteyn, I.S., I.M. Ryzhik, *Tables of Integrals, Series, and Products*, Fifth Ed., Academic Press, New York, 1994.

Grimmett, G., D. Stirzaker, *Probability and Random Processes*, Third Ed., Oxford University Press, New York, 2001.

Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, New York, 1989.

Middleton, D., *An Introduction to Statistical Communication Theory*, McGraw-Hill, New York, 1960.

Rao, C.R., *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York, 1973.

Reif, F., *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York, 1965.

Urick, R.J., *Principles of Underwater Sound*, McGraw-Hill, New York, 1975.

Weisburg, H.F., J.A. Krosnick, B.D. Bowen, *An Introduction to Survey Research, Polling and Data Analysis*, Sage Pubs., CA, 1996.

Problems

15.1 (f) For the PMF given by (15.2) plot the CDF for $N = 10$, $N = 30$, and $N = 100$. What function does the CDF appear to converge to?

15.2 (c) If $X_i \sim \mathcal{N}(1,1)$ for $i = 1, 2 \ldots, N$ are IID random variables, plot a realization of the sample mean random variable versus N. Should the realization converge and if so to what value?

15.3 (w,c) Let $X_{1_i} \sim \mathcal{U}(0,2)$ for $i = 1, 2 \ldots, N$ be IID random variables and let $X_{2_i} \sim \mathcal{N}(1,4)$ for $i = 1, 2 \ldots, N$ be another set of IID random variables. If the sample mean random variable is formed for each set of IID random variables, which one should converge faster? Implement a computer simulation to check your results.

15.4 (⌣̈) (w) Consider the weighted sum of N IID random variables $Y_N = \sum_{i=1}^{N} \alpha_i X_i$. If $E_X[X] = 0$ and $\text{var}(X) = 1$, under what conditions will the sum converge to a number? Can you give an example, other than $\alpha_i = 1/N$, of a set of α_i's which will result in convergence?

15.5 (w) A random walk is defined as $X_N = X_{N-1} + U_N$ for $N = 2, 3, \ldots$ and $X_1 = U_1$, where the U_i's are IID random variables with $P[U_i = -1] = P[U_i = +1] = 1/2$. Will X_N converge to anything as $N \to \infty$?

15.6 (w) To estimate the second moment of a random variable it is proposed to use $(1/N)\sum_{i=1}^{N} X_i^2$. Under what conditions will the estimate converge to the true value?

15.7 (⌣̈) (w) If X_i for $i = 1, 2 \ldots, N$ are IID random variables, will the random variable $(1/\sqrt{N})\sum_{i=1}^{N} X_i$ converge to a number?

15.8 (t,c) In this problem we attempt to demonstrate that convergence in probability is different than standard convergence of a sequence of real numbers.

Consider the sequence of random variables

$$Y_N = \frac{X_N}{\sqrt{N}} + u\left(\frac{X_N}{\sqrt{N}} - 0.1\right)$$

where the X_N's are IID, each with PDF $X_N \sim \mathcal{N}(0,1)$ and $u(x)$ is the unit step function. Prove that $P[|Y_N| > \epsilon] \to 0$ as $N \to \infty$ by using the law of total probability as

$$\begin{aligned} P[|Y_N| > \epsilon] &= P[|Y_N| > \epsilon | X_N/\sqrt{N} > 0.1] P[X_N/\sqrt{N} > 0.1] \\ &\quad + P[|Y_N| > \epsilon | X_N/\sqrt{N} \le 0.1] P[X_N/\sqrt{N} \le 0.1]. \end{aligned}$$

This says that $Y_N \to 0$ in probability. Next simulate this sequence on the computer for $N = 1, 2, \ldots, 200$ to generate 4 realizations of $\{Y_1, Y_2, \ldots, Y_{200}\}$. Examine whether for a given N all realizations lie within the "convergence band" of $[-0.2, 0.2]$. Next generate an additional 6 realizations and overlay all 10 realizations. What can you say about the convergence of any one realization?

15.9 (w) There are 1000 resistors in a bin labeled 10 ohms. Due to manufacturing tolerances, however, the resistance of the resistors are somewhat different. Assume that the resistance can be modeled as a random variable with a mean of 10 ohms and a variance of 2 ohms^2. If 100 resistors are chosen from the bin and connected in series (so the resistances add together), what is the approximate probability that the total resistance will exceed 1030 ohms?

15.10 (w) Consider a sequence of random variables $X_1, X_1, X_2, X_2, X_3, X_3, \ldots$, where $X_1, X_2, X_3 \ldots$ are IID random variables. Does the law of large numbers hold? How about the central limit theorem?

15.11 (w) Consider an Erlang random variable with parameter N. If N increases, does the PDF become Gaussian? Hint: Compare the characteristic functions of the exponential random variable and the $\Gamma(N, \lambda)$ random variable in Table 11.1.

15.12 (f) Find the approximate PDF of $Y = \sum_{i=1}^{100} X_i^2$, if the X_i's are IID with $X_i \sim \mathcal{N}(-4, 8)$.

15.13 (⌣) (f) Find the approximate PDF of $Y = \sum_{i=1}^{1000} X_i$, if the X_i's are IID with $X_i \sim \mathcal{U}(1, 3)$.

15.14 (f) Find the approximate probability that $Y = \sum_{i=1}^{10} X_i$ will exceed 7, if the X_i's are IID with the PDF

$$p_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

15.15 (c) Modify the computer program `clt_demo.m` listed in Appendix 15A to display the repeated convolution of the PDF

$$p_X(x) = \begin{cases} \frac{\pi}{2}\sin(\pi x) & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

and examine the results.

15.16 (c) Use the computer program `clt_demo.m` listed in Appendix 15A to display the repeated convolution of the PDF $\mathcal{U}(0,1)$. Next modify the program to display the repeated convolution of the PDF

$$p_X(x) = \begin{cases} |2-4x| & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Which PDF results in a faster convergence to a Gaussian PDF and why?

15.17 (t) In this problem we prove that the PDF of a standardized χ^2_N random variable converges to a Gaussian PDF as $N \to \infty$. To do so let $Y_N \sim \chi^2_N$ and show that the characteristic function is

$$\phi_{Y_N}(\omega) = \frac{1}{(1-2j\omega)^{N/2}}$$

by using Table 11.1. Next define the standardized random variable

$$Z_N = \frac{Y_N - E[Y_N]}{\sqrt{\text{var}(Y_N)}}$$

and note that the mean and variance of a χ^2_N random variable is N and $2N$, respectively. Show the characteristic function of Z_N is

$$\phi_{Z_N}(\omega) = \frac{\exp(-j\omega\sqrt{N/2})}{(1-j\omega\sqrt{2/N})^{N/2}}.$$

Finally, take the natural logarithm of $\phi_{Z_N}(\omega)$ and note that for a complex variable x with $|x| \ll 1$, we have that $\ln(1-x) \approx -x - x^2/2$. You should be able to show that as $N \to \infty$, $\ln \phi_{Z_N}(\omega) \to -\omega^2/2$.

15.18 (w) A particle undergoes collisions with other particles. Each collision causes its horizontal velocity to change according to a $\mathcal{N}(0, 0.1)$ cm/sec random variable. After 100 independent collisions what is the probability that the particle's velocity will exceed 5 cm/sec if it is initially at rest? Is this result exact or approximate?

15.19 (⌣̈) (f) The sample mean random variable of N IID random variables with $X_i \sim \mathcal{U}(0,1)$ will converge to $1/2$. How many random variables need to be averaged before we can assert that the approximate probability of an error of not more than 0.01 in magnitude is 0.99?

15.20 (⌣) **(w)** An orange grove produces oranges whose weights are uniformly distributed between 3 and 7 ozs. If a truck can hold 4000 lbs. of oranges, what is the approximate probability that it can carry 15,000 oranges?

15.21 (w) A sleeping pill is effective for 75% of the population. If in a hospital 160 patients are given a sleeping pill, what is the approximate probability that 125 or more of them will sleep better?

15.22 (⌣) **(w)** For which PDF will a sum of IID random variables when added together have a PDF that converges to a Gaussian PDF the fastest?

15.23 (⌣) **(w)** A coin is tossed 1000 times, producing 750 heads. Is this a fair coin?

15.24 (f,c) To compute the probability of (15.11) we can use the following approach to compute each term in the summation. Each term can be written as

$$p_{Y_N}[k] = \binom{N}{k}\left(\frac{1}{2}\right)^N = \frac{N(N-1)\cdots(N-k+1)}{1(2)(3)\cdots(k)}\left(\frac{1}{2}\right)^N.$$

Taking the natural logarithm produces

$$\ln p_{Y_N}[k] = \sum_{i=N-k+1}^{N} \ln(i) - \sum_{i=1}^{k} \ln(i) - N\ln(2)$$

which is easily done on a computer. Next, exponentiate to find $p_{Y_N}[k]$ and add each of the terms together to finally implement the summation. Carry this out to verify the result given in (15.11). What happens if you try to compute each term directly?

15.25 (f) In a poll of candidate preferences for two candidates, we wish to report that the margin of error is only $\pm 1\%$. What is the maximum number of people whom we will need to poll?

15.26 (⌣) **(w)** A clinical trial is performed to determine if a particular drug is effective. A group of 100 people is split into two equal groups at random. The drug is administered to group 1 while group 2 is given a placebo. As a result of the study, 40 people in group 1 show a marked improvement while only 30 people in group 2 do so. Is the drug effective? Hint: Find the confidence intervals (using (15.14)) for the percentage of the people in each group who show an improvement.

Appendix 15A

MATLAB Program to Compute Repeated Convolution of PDFs

```
%  This program demonstrates the central limit theorem.  It determines
%  the PDF for the sum S_N of N IID random variables. Each marginal PDF
%  is assumed to be nonzero over the interval (0,1). The repeated
%  convolution integral is implemented using a discrete convolution. The
%  plots of the PDF of S_N as N increases are shown successively
%  (press carriage return for next plot).
%
% clt_demo.m
clear all
delu=0.005;
u=[0:delu:1-delu]'; % p_X defined on interval [0,1)
p_X=ones(length(u),1); % try p_X=abs(2-4*u) for really strange PDF
x=[u;u+1]; % increase abcissa values since repeated
           % convolution increases nonzero width of output
p_S=zeros(length(x),1);
N=12; % number of random variables summed
for j=1:length(x) % start discrete convolution approximation
                  % to continuous convolution
  for i=1:length(u)
    if j-i>0&j-i<=length(p_X)
      p_S(j)=p_S(j)+p_X(i)*p_X(j-i)*delu;
    end
  end
end
plot(x,p_S) % plot results for N=2
grid
```

```
axis([0 N 0 1]) % set axes lengths for plotting
xlabel('x')
ylabel('p_S')
title('PDF for S_N')
text(0.75*N,0.85,'N = 2') % label plot with the
                          % number of convolutions
for n=3:N
  pause
  x=[x;u+n-1]; % increase abcissa values since
               % repeated convolution increases
               % nonzero width of output
  p_S=[p_S;zeros(length(u),1)];
  g=zeros(length(p_S),1);
  for j=1:length(x) % start discrete convolution
  for i=1:length(u)
    if j-i>0
      g(j,1)=g(j,1)+p_X(i)*p_S(j-i)*delu;
    end
  end
end
p_S=g; % plot results for N=3,4,...,12
plot(x,p_S)
grid
axis([0 N 0 1])
xlabel('x')
ylabel('p_S')
title('PDF for S_N')
text(0.75*N,0.85,['N = ' num2str(n)])
end
```

Appendix 15B

Proof of Central Limit Theorem

In this appendix we prove the central limit theorem for continuous random variables. Consider the characteristic function of the standardized continuous random variable

$$Z_N = \frac{S_N - NE_X[X]}{\sqrt{N\text{var}(X)}}$$

where $S_N = \sum_{i=1}^{N} X_i$ and the X_i's are IID. By definition of Z_N the characteristic function becomes

$$\begin{aligned}
\phi_{Z_N}(\omega) &= E_{Z_N}[\exp(j\omega Z_N)] \\
&= E_{\mathbf{X}}\left[\exp\left(j\omega\frac{\sum_{i=1}^{N} X_i - NE_X[X]}{\sqrt{N\text{var}(X)}}\right)\right] \\
&= E_{\mathbf{X}}\left[\prod_{i=1}^{N}\exp\left(j\omega\frac{X_i - E_X[X]}{\sqrt{N\text{var}(X)}}\right)\right] \\
&= \prod_{i=1}^{N} E_{X_i}\left[\exp\left(j\omega\frac{X_i - E_X[X]}{\sqrt{N\text{var}(X)}}\right)\right] \qquad \text{(independence of } X_i\text{'s)} \\
&= \left[E_X\left[\exp\left(j\omega\frac{X - E_X[X]}{\sqrt{N\text{var}(X)}}\right)\right]\right]^N \qquad \text{(identically distributed } X_i\text{'s).}
\end{aligned}$$

But for a complex variable ξ we can write its exponential as a Taylor series yielding

$$\exp(\xi) = \sum_{k=0}^{\infty} \frac{\xi^k}{k!} \qquad \text{(see Problem 5.22).}$$

Thus,

$$E_X\left[\exp\left(j\omega\frac{X - E_X[X]}{\sqrt{N\text{var}(X)}}\right)\right]$$

$$\begin{aligned}
&= E_X\left[\sum_{k=0}^{\infty}\frac{(j\omega)^k}{k!}\left(\frac{X-E_X[X]}{\sqrt{N\mathrm{var}(X)}}\right)^k\right] \\
&= \sum_{k=0}^{\infty}\frac{(j\omega)^k}{k!}E_X\left[\left(\frac{X-E_X[X]}{\sqrt{N\mathrm{var}(X)}}\right)^k\right] \qquad \text{(assume interchange valid)} \\
&= 1+j\omega E_X\left[\frac{X-E_X[X]}{\sqrt{N\mathrm{var}(X)}}\right]+\frac{1}{2}(j\omega)^2E_X\left[\left(\frac{X-E_X[X]}{\sqrt{N\mathrm{var}(X)}}\right)^2\right]+E_X[R(X)]
\end{aligned}$$

where $R(X)$ is the third-order and higher terms of the Taylor expansion. But

$$\begin{aligned}
E_X\left[\frac{X-E_X[X]}{\sqrt{N\mathrm{var}(X)}}\right] &= \frac{E_X[X]-E_X[X]}{\sqrt{N\mathrm{var}(X)}}=0 \\
E_X\left[\left(\frac{X-E_X[X]}{\sqrt{N\mathrm{var}(X)}}\right)^2\right] &= \frac{E_X[(X-E_X[X])^2]}{N\mathrm{var}(X)}=\frac{1}{N}
\end{aligned}$$

and so

$$\phi_{Z_N}(\omega)=\left[1-\frac{\omega^2}{2N}+E_X[R(X)]\right]^N.$$

The terms comprising $R(X)$ are

$$\begin{aligned}
R(X) &= \frac{(j\omega)^3}{3!}E_X\left[\left(\frac{X-E_X[X]}{\sqrt{N\mathrm{var}(X)}}\right)^3\right]+\cdots \\
&= \frac{1}{N^{3/2}}\frac{(j\omega)^3}{3!}E_X\left[\left(\frac{X-E_X[X]}{\sqrt{\mathrm{var}(X)}}\right)^3\right]+\cdots
\end{aligned}$$

which can be shown to be small, due to the division of the successive terms by $N^{3/2}, N^2, \ldots$, relative to the $-\omega^2/(2N)$ term. Hence as $N \to \infty$, they do not contribute to $\phi_{Z_N}(\omega)$ and therefore

$$\begin{aligned}
\phi_{Z_N}(\omega) &\to \left(1-\frac{\omega^2}{2N}\right)^N \\
&\to \exp\left(-\frac{1}{2}\omega^2\right)=\phi_Z(\omega) \qquad \text{(see Problem 5.15)}
\end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$. Since the characteristic function of Z_N converges to the characteristic function of Z, we have by the continuity theorem (see Section 11.7) that the PDF of Z_N must converge to the PDF of Z. Therefore, we have finally that as $N \to \infty$

$$p_{Z_N}(z) \to p_Z(z)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right).$$