Review Example

Two sensors A, and B provide the following measurement vectors:

\[ X = [3, 4, -4, 4] \]
\[ Y = [-4, -2, 1, 6] \]

a) What is \( E[X] \), \( E[Y] \)?
b) What is \( \text{COV}[X, Y] \)?
c) What is \( \rho_{XY} \)?
d) Are the measurement vectors orthogonal?

Section 9.5

Central Limit Theorem

Figure 9.1

The PMF of the \( X \), the number of heads in \( n \) coin flips for \( n = 5, 10, 20 \). As \( n \) increases, the PMF more closely resembles a bell-shaped curve.
**Theorem 9.12**  Central Limit Theorem

Given $X_1, X_2, \ldots$, a sequence of iid random variables with expected value $\mu_X$ and variance $\sigma_X^2$, the CDF of $Z_n = (\sum_{i=1}^{n} X_i - n\mu_X)/\sqrt{n\sigma_X^2}$ has the property

$$\lim_{n \to \infty} F_{Z_n}(z) = \Phi(z).$$

**Figure 9.2**

(a) $n = 1$
(b) $n = 2$
(c) $n = 3$
(d) $n = 4$

The PDF of $W_n$, the sum of $n$ uniform (0,1) random variables, and the corresponding central limit theorem approximation for $n = 1, 2, 3, 4$. The solid --- line denotes the PDF $f_{W_n}(w)$, and the broken --- line denotes the Gaussian approximation.

**Figure 9.3**

$n = 2, p = 1/2$
$n = 4, p = 1/2$
$n = 8, p = 1/2$
$n = 16, p = 1/2$

The binomial $(n, p)$ CDF and the corresponding central limit theorem approximation for $n = 4, 8, 16, 32$, and $p = 1/2$.

**Definition 9.2**  Approximation

Let $W_n = X_1 + \cdots + X_n$ be the sum of $n$ iid random variables, each with $E[X] = \mu_X$ and $\text{Var}[X] = \sigma_X^2$. The central limit theorem approximation to the CDF of $W_n$ is

$$F_{W_n}(w) \approx \Phi\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right).$$
Example 9.10

To gain some intuition into the central limit theorem, consider a sequence of iid continuous random variables \( X_i \), where each random variable is uniform \((0,1)\). Let

\[
W_n = X_1 + \cdots + X_n.
\]

(1)

Recall that \( \mathbb{E}[X] = 0.5 \) and \( \text{Var}[X] = 1/12 \). Therefore, \( W_n \) has expected value \( \mathbb{E}[W_n] = n/2 \) and variance \( n/12 \). The central limit theorem says that the CDF of \( W_n \) should approach a Gaussian CDF with the same expected value and variance. Moreover, since \( W_n \) is a continuous random variable, we would also expect that the PDF of \( W_n \) would converge to a Gaussian PDF. In Figure 9.2, we compare the PDF of \( W_n \) to the PDF of a Gaussian random variable with the same expected value and variance. First, \( W_2 \) is a uniform random variable with the rectangular PDF shown in Figure 9.2(a). This figure also shows the PDF of \( W_2 \), a Gaussian random variable with expected value \( \mu = 0.5 \) and variance \( \sigma^2 = 1/12 \). Here the PDFs are very dissimilar. When we consider \( n = 2 \), we have the situation in Figure 9.2(b). The PDF of \( W_2 \) is a triangle with expected value \( 1 \) and variance \( 2/12 \). The figure shows the corresponding Gaussian PDF. The following figures show the PDFs of \( W_3, \ldots, W_6 \). The convergence to a bell shape is apparent.

Example 9.11

Now suppose \( W_n = X_1 + \cdots + X_n \) is a sum of independent Bernoulli \((p)\) random variables. We know that \( W_n \) has the binomial PMF

\[
P_{W_n}(w) = \binom{n}{w} p^w (1-p)^{n-w}.
\]

(1)

No matter how large \( n \) becomes, \( W_n \) is always a discrete random variable and would have a PDF consisting of impulses. However, the central limit theorem says that the CDF of \( W_n \) converges to a Gaussian CDF. Figure 9.3 demonstrates the convergence of the sequence of binomial CDFs to a Gaussian CDF for \( p = 1/2 \) and four values of \( n \), the number of Bernoulli random variables that are added to produce a binomial random variable. For \( n \geq 32 \), Figure 9.3 suggests that approximations based on the Gaussian distribution are very accurate.

Example 9.12 Problem

A compact disc (CD) contains digitized samples of an acoustic waveform. In a CD player with a “one bit digital to analog converter,” each digital sample is represented to an accuracy of \( \pm 0.5 \) mV. The CD player oversamples the waveform by making eight independent measurements corresponding to each sample. The CD player obtains a waveform sample by calculating the average (sample mean) of the eight measurements. What is the probability that the error in the waveform sample is greater than \( 0.1 \) mV?

Example 9.12 Solution

The measurements \( X_1, X_2, \ldots, X_8 \) all have a uniform distribution between \( v - 0.5 \) mV and \( v + 0.5 \) mV, where \( v \) mV is the exact value of the waveform sample. The compact disk player produces the output \( U = W_8/8 \), where

\[
W_8 = \sum_{i=1}^{8} X_i.
\]

(1)

To find \( P[|U - v| > 0.1] \) exactly, we would have to find an exact probability model for \( W_8 \), either by computing an eighthfold convolution of the uniform PDF of \( X_i \) or by using the moment generating function. Either way, the process is extremely complex. Alternatively, we can use the central limit theorem to model \( W_8 \) as a Gaussian random variable with \( \mathbb{E}[W_8] = 8\mu_X = 8v \) mV and variance \( \text{Var}[W_8] = 8\text{Var}[X] = 8/12 \). Therefore, \( U \) is approximately Gaussian with \( \mathbb{E}[U] = \mathbb{E}[W_8]/8 = v \) and variance \( \text{Var}[W_8]/64 = 1/96 \). Finally, the error, \( U - v \) in the output waveform sample is approximately Gaussian with expected value \( 0 \) and variance \( 1/96 \). It follows that

\[
P[|U - v| > 0.1] = 2 \left[ 1 - \Phi \left( 0.1 / \sqrt{1/96} \right) \right] = 0.3272.
\]

(2)
Example 9.13 Problem

A modem transmits one million bits. Each bit is 0 or 1 independently with equal probability. Estimate the probability of at least 502,000 ones.

Example 9.13 Solution

Let $X_i$ be the value of bit $i$ (either 0 or 1). The number of ones in one million bits is $W = \sum_{i=1}^{10^6} X_i$. Because $X_i$ is a Bernoulli (0.5) random variable, $E[X_i] = 0.5$ and $\text{Var}[X_i] = 0.25$ for all $i$. Note that $E[W] = 10^6 E[X_i] = 500,000$, and $\text{Var}[W] = 10^6 \text{Var}[X_i] = 250,000$. Therefore, $\sigma_W = 500$. By the central limit theorem approximation,

$$P \left[ W \geq 502,000 \right] = 1 - P \left[ W \leq 502,000 \right]$$

$$\approx 1 - \Phi \left( \frac{502,000 - 500,000}{500} \right) = 1 - \Phi(4). \quad (1)$$

Using Table 4.2, we observe that $1 - \Phi(4) = Q(4) = 3.17 \times 10^{-5}$.

Example 9.14 Problem

Transmit one million bits. Let $A$ denote the event that there are at least 499,000 ones but no more than 501,000 ones. What is $P[A]$?

Example 9.14 Solution

As in Example 9.13, $E[W] = 500,000$ and $\sigma_W = 500$. By the central limit theorem approximation,

$$P \left[ A \right] = P \left[ W \leq 501,000 \right] - P \left[ W < 499,000 \right]$$

$$\approx \Phi \left( \frac{501,000 - 500,000}{500} \right) - \Phi \left( \frac{499,000 - 500,000}{500} \right)$$

$$= \Phi(2) - \Phi(-2) = 0.9544. \quad (1)$$
Definition 9.3 De Moivre–Laplace Formula

For a binomial \((n, p)\) random variable \(K\),
\[
P[k_1 \leq K \leq k_2] \approx \Phi\left(\frac{k_2 + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 0.5 - np}{\sqrt{np(1-p)}}\right).
\]

Example 9.15 Problem

Let \(K\) be a binomial \((n = 20, p = 0.4)\) random variable. What is \(P[K = 8]\)?

Example 9.15 Solution

Since \(E[K] = np = 8\) and \(\text{Var}[K] = np(1-p) = 4.8\), the central limit theorem approximation to \(K\) is a Gaussian random variable \(X\) with \(E[X] = 8\) and \(\text{Var}[X] = 4.8\). Because \(X\) is a continuous random variable, \(P[X = 8] = 0\), a useless approximation to \(P[K = 8]\). On the other hand, the De Moivre–Laplace formula produces
\[
P[8 \leq K \leq 8] \approx P[7.5 \leq X \leq 8.5]
= \Phi\left(\frac{0.5}{\sqrt{4.8}}\right) - \Phi\left(\frac{-0.5}{\sqrt{4.8}}\right) = 0.1803. \tag{1}
\]

The exact value is \(\binom{20}{8}(0.4)^8(1-0.4)^{12} = 0.1797\).

Example 9.16 Problem

\(K\) is the number of heads in 100 flips of a fair coin. What is \(P[50 \leq K \leq 51]\)?
Example 9.16 Solution

Since $K$ is a binomial ($n = 100, p = 0.5$) random variable,
\[
P[50 \leq K \leq 51] = P_K(50) + P_K(51)
= \binom{100}{50} \left( \frac{1}{2} \right)^{100} + \binom{100}{51} \left( \frac{1}{2} \right)^{100} = 0.1576. \quad (1)
\]

Since $E[K] = 50$ and $\sigma_K = \sqrt{np(1-p)} = 5$, the ordinary central limit theorem approximation produces
\[
P[50 \leq K \leq 51] \approx \Phi \left( \frac{51 - 50}{5} \right) - \Phi \left( \frac{50 - 50}{5} \right) = 0.0793. \quad (2)
\]

This approximation error of roughly 50% occurs because the ordinary central limit theorem approximation ignores the fact that the discrete random variable $K$ has two probability masses in an interval of length 1. As we see next, the De Moivre–Laplace approximation is far more accurate.
\[
P[50 \leq K \leq 51] \approx \Phi \left( \frac{51 + 0.5 - 50}{5} \right) - \Phi \left( \frac{50 - 0.5 - 50}{5} \right)
= \Phi(0.3) - \Phi(-0.1) = 0.1577. \quad (3)
\]

Quiz 9.5

(a) The expected access time is
\[
E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx = \int_0^{12} x \frac{x}{12} \, dx = 6 \text{ ms}. \quad (1)
\]

(b) The second moment of the access time is
\[
E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) \, dx = \int_0^{12} x^2 \frac{x}{12} \, dx = 48. \quad (2)
\]

The variance of the access time is $\text{Var}[X] = E[X^2] - (E[X])^2 = 12$.

(c) Using $X_i$ to denote the access time of block $i$, we can write
\[
A = X_1 + X_2 + \cdots + X_{12} \quad (3)
\]

Since the expectation of the sum equals the sum of the expectations,
\[
E[A] = E[X_1] + \cdots + E[X_{12}] = 12 E[X] = 72 \text{ ms}. \quad (4)
\]

Quiz 9.5 Solution (Continued)

(d) Since the $X_i$ are independent,
\[
\text{Var}[A] = \text{Var}[X_1] + \cdots + \text{Var}[X_{12}] = 12 \text{Var}[X] = 144. \quad (5)
\]

Thus $A$ has standard deviation $\sigma_A = 12$.

(e) To use the central limit theorem, we use Table 4.2 to evaluate
\[
P[A \leq 75] = P \left[ \frac{A - E[A]}{\sigma_A} \leq \frac{75 - E[A]}{\sigma_A} \right]
\approx \Phi \left( \frac{75 - 72}{12} \right) = 0.5987. \quad (6)
\]

Then $P[A > 75] = 1 - P[A \leq 75] = 0.4013$.

(f) Once again, we use the central limit theorem and Table 4.2 to estimate
\[
P[A < 48] = P \left[ \frac{A - E[A]}{\sigma_A} < \frac{48 - E[A]}{\sigma_A} \right]
\approx \Phi \left( \frac{48 - 72}{12} \right) = 0.0227. \quad (7)
\]
Section 10.1

Sample Mean: Expected Value and Variance

Sample Mean ≠ Expected Value

- The first thing to notice is that $M_n(X)$ is a function of the random variables $X_1, \ldots, X_n$ and is therefore a random variable itself.

- It is important to distinguish the sample mean, $M_n(X)$, from $E[X]$, which we sometimes refer to as the mean value of random variable $X$.

- While $M_n(X)$ is a random variable, $E[X]$ is a number.

- To avoid confusion when studying the sample mean, it is advisable to refer to $E[X]$ as the expected value of $X$, rather than the mean of $X$.

- The sample mean of $X$ and the expected value of $X$ are closely related.

- A major purpose of this chapter is to explore the fact that as $n$ increases without bound, $M_n(X)$ predictably approaches $E[X]$.

- In everyday conversation, this phenomenon is often called the law of averages.

Definition 10.1 Sample Mean

For iid random variables $X_1, \ldots, X_n$ with PDF $f_X(x)$, the sample mean of $X$ is the random variable

$$M_n(X) = \frac{X_1 + \ldots + X_n}{n}.$$ 

Theorem 10.1

The sample mean $M_n(X)$ has expected value and variance

$$E[M_n(X)] = E[X], \quad \text{Var}[M_n(X)] = \frac{\text{Var}[X]}{n}.$$
Proof: Theorem 10.1

From Definition 10.1, Theorem 9.1, and the fact that $E[X_i] = E[X]$ for all $i$,

$$E[M_n(X)] = \frac{1}{n}(E[X] + \cdots + E[X]) = \frac{1}{n}(E[X] + \cdots + E[X]) = E[X].$$  \hspace{1cm} (1)

Because $\text{Var}[aY] = a^2 \text{Var}[Y]$ for any random variable $Y$ (Theorem 3.15), $\text{Var}[M_n(X)] = \text{Var}[X_1 + \cdots + X_n]/n^2$. Since the $X_i$ are iid, we can use Theorem 9.3 to show

$$\text{Var}[X_1 + \cdots + X_n] = \text{Var}[X_1] + \cdots + \text{Var}[X_n] = n \text{Var}[X].$$ \hspace{1cm} (2)

Thus $\text{Var}[M_n(X)] = n \text{Var}[X]/n^2 = \text{Var}[X]/n$.

Comment: Theorem 10.1

- Theorem 10.1 demonstrates that as $n$ increases without bound, the variance of $M_n(X)$ goes to zero.
- When we first met the variance, and its square root the standard deviation, we said that they indicate how far a random variable is likely to be from its expected value.
- Theorem 10.1 suggests that as $n$ approaches infinity, it becomes highly likely that $M_n(X)$ is arbitrarily close to its expected value, $E[X]$.
- In other words, the sample mean $M_n(X)$ converges to the expected value $E[X]$ as the number of samples $n$ goes to infinity.

Quiz 10.1

$X$ is the exponential (1) random variable; $M_n(X)$ is the sample mean of $n$ independent samples of $X$. How many samples $n$ are needed to guarantee that the variance of the sample mean $M_n(X)$ is no more than 0.01?

Quiz 10.1 Solution

An exponential random variable with expected value 1 also has variance 1. By Theorem 10.1, $M_n(X)$ has variance $\text{Var}[M_n(X)] = 1/n$. Hence, we need $n = 100$ samples.
Section 10.2

Deviation of a Random Variable from the Expected Value

Proof: Theorem 10.2

Since $X$ is nonnegative, $f_X(x) = 0$ for $x < 0$ and

$$E[X] = \int_0^{c^2} x f_X(x) \, dx + \int_{c^2}^{\infty} x f_X(x) \, dx \geq \int_{c^2}^{\infty} x f_X(x) \, dx. \quad (1)$$

Since $x \geq c^2$ in the remaining integral,

$$E[X] \geq c^2 \int_{c^2}^{\infty} f_X(x) \, dx = c^2 P[X \geq c^2]. \quad (2)$$

Theorem 10.2: Markov Inequality

For a random variable $X$, such that $P[X < 0] = 0$, and a constant $c$,

$$P[X \geq c^2] \leq \frac{E[X]}{c^2}.$$ 

Example 10.1

Let $X$ represent the height (in feet) of a storm surge following a hurricane. If the expected height is $E[X] = 5.5$, then the Markov inequality states that an upper bound on the probability of a storm surge at least 11 feet high is

$$P[X \geq 11] \leq \frac{5.5}{11} = \frac{1}{2}. \quad (1)$$
**Example 10.2**

Suppose random variable $Y$ takes on the value $c^2$ with probability $p$ and the value 0 otherwise. In this case, $E[Y] = pc^2$, and the Markov inequality states

$$P[Y \geq c^2] \leq E[Y]/c^2 = p. \quad (1)$$

Since $P[Y \geq c^2] = p$, we observe that the Markov inequality is in fact an equality in this instance.

**Theorem 10.3  Chebyshev Inequality**

For an arbitrary random variable $Y$ and constant $c > 0$,

$$P[|Y - \mu_Y| \geq c] \leq \frac{\text{Var}[Y]}{c^2}.$$