Section 10.1

Sample Mean: Expected Value and Variance

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*

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

Sample Mean $\neq$ 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 $\mathrm{E}[X]$, which we sometimes refer to as the *mean value* of random variable X.
- While $M_n(X)$ is a random variable, $\mathrm{E}[X]$ is a number.
- To avoid confusion when studying the sample mean, it is advisable to refer to $\mathrm{E}[X]$ 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 $\mathrm{E}[X]$.
- In everyday conversation, this phenomenon is often called the *law of averages*.

Theorem 10.1

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

$$\mathsf{E}\left[M_n(X)\right] = \mathsf{E}\left[X\right], \qquad \mathsf{Var}[M_n(X)] = \frac{\mathsf{Var}[X]}{n}.$$

Proof: Theorem 10.1

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

$$\mathsf{E}\left[M_n(X)\right] = \frac{1}{n}\left(\mathsf{E}\left[X_1\right] + \cdots + \mathsf{E}\left[X_n\right]\right) = \frac{1}{n}\left(\mathsf{E}\left[X\right] + \cdots + \mathsf{E}\left[X\right]\right) = \mathsf{E}\left[X\right]. \tag{10.1}$$

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

$$\mathsf{Var}[X_1 + \cdots + X_n] = \mathsf{Var}[X_1] + \cdots + \mathsf{Var}[X_n] = n \,\mathsf{Var}[X]. \tag{10.2}$$

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

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, $\mathsf{E}[X]$.

- In other words, the sample mean $M_n(X)$ converges to the expected value $\mathsf{E}[X]$ 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

Theorem 10.2 Markov Inequality

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

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

Proof: Theorem 10.2

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

$$\mathsf{E}[X] = \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^{\infty} x f_X(x)\,dx \geq \int_{c^2}^{\infty} x f_X(x)\,dx. \tag{10.3}$$

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

$$\mathsf{E}[X] \geq c^2 \int_{c^2}^{\infty} f_X(x)\,dx = c^2\,\mathsf{P}\left[X \geq c^2\right]. \tag{10.4}$$

Example 10.1

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

$$\mathrm{P}[X \geq 11] \leq 5.5/11 = 1/2. \tag{10.5}$$

Example 10.2

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

$$\mathsf{P}\left[Y \geq c^2\right] \leq \mathsf{E}\left[Y\right]/c^2 = p. \tag{10.6}$$

Since $\mathsf{P}[Y \geq c^2] = p$, 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 $c > 0$,

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

Proof: Theorem 10.3

In the Markov inequality, Theorem 10.2, let $X = (Y - \mu_Y)^2$. The inequality states

$$\mathsf{P}\left[X \geq c^2\right] = \mathsf{P}\left[(Y - \mu_Y)^2 \geq c^2\right] \leq \frac{\mathsf{E}\left[(Y - \mu_Y)^2\right]}{c^2} = \frac{\mathsf{Var}[Y]}{c^2}. \tag{10.7}$$

The theorem follows from the fact that $\{(Y - \mu_Y)^2 \geq c^2\} = \{|Y - \mu_Y| \geq c\}$.

Example 10.3 Problem

If the height X of a storm surge following a hurricane has expected value $\mathsf{E}[X] = 5.5$ feet and standard deviation $\sigma_X = 1$ foot, use the Chebyshev inequality to to find an upper bound on $\mathsf{P}[X \geq 11]$.

Example 10.3 Solution

Since a height X is nonnegative, the probability that $X \geq 11$ can be written as

$$\text{P}\left[X \geq 11\right] = \text{P}\left[X - \mu_X \geq 11 - \mu_X\right] = \text{P}\left[|X - \mu_X| \geq 5.5\right]. \tag{10.8}$$

Now we use the Chebyshev inequality to obtain

$$\text{P}\left[X \geq 11\right] = \text{P}\left[|X - \mu_X| \geq 5.5\right] \leq \text{Var}[X]/(5.5)^2 = 0.033 \approx 1/30. \tag{10.9}$$

Although this bound is better than the Markov bound, it is also loose. $\text{P}[X \geq 11]$ is seven orders of magnitude lower than 1/30.

Theorem 10.4 Chernoff Bound

For an arbitrary random variable X and a constant c,

$$\mathsf{P}[X \geq c] \leq \min_{s \geq 0} e^{-sc} \phi_X(s).$$

Proof: Theorem 10.4

In terms of the unit step function, $u(x)$, we observe that

$$\mathrm{P}[X \geq c] = \int_c^\infty f_X(x)\ dx = \int_{-\infty}^\infty u(x-c) f_X(x)\ dx. \tag{10.10}$$

For all $s \geq 0$, $u(x-c) \leq e^{s(x-c)}$. This implies

$$\mathrm{P}[X \geq c] \leq \int_{-\infty}^\infty e^{s(x-c)} f_X(x)\ dx = e^{-sc} \int_{-\infty}^\infty e^{sx} f_X(x)\ dx = e^{-sc}\phi_X(s). \tag{10.11}$$

This inequality is true for any $s \geq 0$. Hence the upper bound must hold when we choose s to minimize $e^{-sc}\phi_X(s)$.

Example 10.4 Problem

If the probability model of the height X, measured in feet, of a storm surge following a hurricane at a certain location is the Gaussian (5.5, 1) random variable, use the Chernoff bound to find an upper bound on $\mathrm{P}[X \geq 11]$.

Example 10.4 Solution

In Table 9.1 the MGF of X is

$$\phi_X(s) = e^{(11s+s^2)/2}. \qquad (10.12)$$

Thus the Chernoff bound is

$$\text{P}[X \geq 11] \leq \min_{s \geq 0} e^{-11s} e^{(11s+s^2)/2} = \min_{s \geq 0} e^{(s^2-11s)/2}. \qquad (10.13)$$

To find the minimizing s, it is sufficient to choose s to minimize $h(s) = s^2 - 11s$. Setting the derivative $dh(s)/ds = 2s - 11 = 0$ yields $s = 5.5$. Applying $s = 5.5$ to the bound yields

$$\text{P}[X \geq 11] \leq e^{(s^2-11s)/2}\Big|_{s=5.5} = e^{-(5.5)^2/2} = 2.7 \times 10^{-7}. \qquad (10.14)$$

Quiz 10.2

In a subway station, there are exactly enough customers on the platform to fill three trains. The arrival time of the nth train is $X_1 + \cdots + X_n$ where $X_1, X_2, \ldots$ are iid exponential random variables with $E[X_i] = 2$ minutes. Let W equal the time required to serve the waiting customers. For $P[W > 20]$, the probability that W is over twenty minutes,

(a) Use the central limit theorem to find an estimate.

(b) Use the Markov inequality to find an upper bound.

(c) Use the Chebyshev inequality to find an upper bound.

(d) Use the Chernoff bound to find an upper bound.

(e) Use Theorem 4.11 for an exact calculation.

Quiz 10.2 Solution

The train interarrival times X_1, X_2, X_3 are iid exponential (λ) random variables. The arrival time of the third train is

$$W = X_1 + X_2 + X_3. \tag{1}$$

In Theorem 9.9, we found that the sum of three iid exponential (λ) random variables is an Erlang ($n = 3, \lambda$) random variable. From Appendix A, we find that W has expected value and variance

$$\mathrm{E}[W] = 3/\lambda = 6, \tag{2}$$

$$\mathrm{Var}[W] = 3/\lambda^2 = 12. \tag{3}$$

(a) By the Central Limit Theorem,

$$\begin{aligned} \mathrm{P}[W > 20] &= \mathrm{P}\left[\frac{W-6}{\sqrt{12}} > \frac{20-6}{\sqrt{12}}\right] \\ &\approx Q\left(\frac{7}{\sqrt{3}}\right) = 2.66 \times 10^{-5}. \end{aligned}$$

(b) From the Markov inequality, we know that

$$\mathrm{P}[W > 20] \leq \frac{\mathrm{E}[W]}{20} = \frac{6}{20} = 0.3. \tag{4}$$

[Continued]

(c) To use the Chebyshev inequality, we observe that $\mathsf{E}[W] = 6$ and W nonnegative imply

$$\mathsf{P}\left[|W - \mathsf{E}[W]| \geq 14\right] = \mathsf{P}[W - 6 \geq 14] + \underbrace{\mathsf{P}[W - 6 \leq -14]}_{=0}$$
$$= \mathsf{P}[W \geq 20]. \quad (5)$$

Thus

$$\mathsf{P}[W \geq 20] = \mathsf{P}\left[|W - \mathsf{E}[W]| \geq 14\right] \quad (6)$$
$$\leq \frac{\mathsf{Var}[W]}{14^2} = \frac{3}{49} = 0.061. \quad (7)$$

(d) For the Chernoff bound, we note that the MGF of W is

$$\phi_W(s) = \left(\frac{\lambda}{\lambda - s}\right)^3 = \frac{1}{(1 - 2s)^3}. \quad (8)$$

The Chernoff bound states that

$$\mathsf{P}[W > 20] \leq \min_{s \geq 0} e^{-20s}\phi_X(s) = \min_{s \geq 0} \frac{e^{-20s}}{(1 - 2s)^3}. \quad (9)$$

[Continued]

To minimize $h(s) = e^{-20s}/(1-2s)^3$, we set the derivative of $h(s)$ to zero:

$$\frac{dh(s)}{ds} = \frac{e^{-20s}(40s-14)}{(1-2s)^4} = 0. \tag{10}$$

This implies $s = 7/20$. Applying $s = 7/20$ into the Chernoff bound yields

$$\mathrm{P}[W > 20] \le \left. \frac{e^{-20s}}{(1-2s)^3} \right|_{s=\frac{7}{20}} = 0.0338.$$

(c) Theorem 4.11 says that for any $w > 0$, the CDF of the Erlang $(3, \lambda)$ random variable W satisfies

$$F_W(w) = 1 - \sum_{k=0}^{2} \frac{(\lambda w)^k e^{-\lambda w}}{k!} \tag{11}$$

Equivalently, for $\lambda = 1/2$ and $w = 20$,

$$\begin{aligned} \mathrm{P}[W > 20] &= 1 - F_W(20) \\ &= e^{-10}\left(1 + \frac{10}{1!} + \frac{10^2}{2!}\right) \\ &= 61e^{-10} = 0.0028. \end{aligned} \tag{12}$$

Although the Chernoff bound is weak in that it overestimates the probability by a factor of 12, it is a valid bound. By contrast, the Central Limit Theorem approximation grossly underestimates the true probability.

Section 10.3

Laws of Large Numbers

Theorem 10.5 Weak Law of Large Numbers (Finite Samples)

For any constant $c > 0$,

(a) $\mathrm{P}[|M_n(X) - \mu_X| \geq c] \leq \dfrac{\mathrm{Var}[X]}{nc^2}$,

(b) $\mathrm{P}[|M_n(X) - \mu_X| < c] \geq 1 - \dfrac{\mathrm{Var}[X]}{nc^2}$.

Proof: Theorem 10.5

Let $Y = M_n(X)$. Theorem 10.1 states that

$$\mathrm{E}[Y] = \mathrm{E}[M_n(X)] = \mu_X \qquad \mathrm{Var}[Y] = \mathrm{Var}[M_n(X)] = \mathrm{Var}[X]/n. \tag{10.15}$$

Theorem 10.5(a) follows by applying the Chebyshev inequality (Theorem 10.3) to $Y = M_n(X)$. Theorem 10.5(b) is just a restatement of Theorem 10.5(a), since

$$\mathrm{P}[|M_n(X) - \mu_X| \geq c] = 1 - \mathrm{P}[|M_n(X) - \mu_X| < c]. \tag{10.16}$$

Theorem 10.6 Weak Law of Large Numbers (Infinite Samples)

If X has finite variance, then for any constant $c > 0$,

(a) $\lim_{n\to\infty} \mathsf{P}[|M_n(X) - \mu_X| \geq c] = 0$,

(b) $\lim_{n\to\infty} \mathsf{P}[|M_n(X) - \mu_X| < c] = 1$.

The Weak Law

- In words, Theorem 10.6(b) says that the probability that the sample mean is within $\pm c$ units of E$[X]$ goes to one as the number of samples approaches infinity.

- Since c can be arbitrarily small (e.g., 10^{-2000}), both Theorem 10.5(a) and Theorem 10.6(b) can be interpreted as saying that the sample mean converges to E$[X]$ as the number of samples increases without bound.

- The weak law of large numbers is a very general result because it holds for all random variables X with finite variance.

- Moreover, we do not need to know any of the parameters, such as the expected value or variance, of random variable X.

Definition 10.2 Convergence in Probability

The random sequence Y_n *converges in probability to a constant* y *if for any* $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}\left[|Y_n - y| \geq \epsilon\right] = 0.$$

Theorem 10.7

As $n \to \infty$, the relative frequency $\hat{P}_n(A)$ converges to $\mathsf{P}[A]$; for any constant $c > 0$,

$$\lim_{n\to\infty} \mathsf{P}\left[\left|\hat{P}_n(A) - \mathsf{P}\left[A\right]\right| \geq c\right] = 0.$$

Proof: Theorem 10.7

The proof follows from Theorem 10.6 since $\widehat{P}_n(A) = M_n(X_A)$ is the sample mean of the indicator X_A, which has expected value $\mathsf{E}[X_A] = \mathsf{P}[A]$ and variance $\mathsf{Var}[X_A] = \mathsf{P}[A](1 - \mathsf{P}[A])$.

Quiz 10.3

$X_1, \ldots, X_n$ are n iid samples of the Bernoulli ($p = 0.8$) random variable X.

(a) Find $\mathsf{E}[X]$ and $\mathrm{Var}[X]$.

(b) What is $\mathrm{Var}[M_{100}(X)]$?

(c) Use Theorem 10.5 to find α such that

$$\mathsf{P}\left[|M_{100}(X) - p| \geq 0.05\right] \leq \alpha.$$

(d) How many samples n are needed to guarantee

$$\mathsf{P}\left[|M_n(X) - p| \geq 0.1\right] \leq 0.05.$$

Quiz 10.3 Solution

(a) Since X is a Bernoulli random variable with parameter $p = 0.8$, we can look up in Appendix A to find that $\mathrm{E}[X] = p = 0.8$ and variance

$$\mathrm{Var}[X] = p(1-p) = (0.8)(0.2) = 0.16. \tag{1}$$

(b) By Theorem 10.1,

$$\mathrm{Var}[M_{100}(X)] = \frac{\mathrm{Var}[X]}{100} = 0.0016. \tag{2}$$

(c) Theorem 10.5 uses the Chebyshev inequality to show that the sample mean satisfies

$$\mathrm{P}\left[|M_n(X) - \mathrm{E}\left[X\right]| \geq c\right] \leq \frac{\mathrm{Var}[X]}{nc^2}. \tag{3}$$

Note that $\mathrm{E}[X] = P_X(1) = p$. To meet the specified requirement, we choose $c = 0.05$ and $n = 100$. Since $\mathrm{Var}[X] = 0.16$, we must have

$$\frac{0.16}{100(0.05)^2} = \alpha \tag{4}$$

This reduces to $\alpha = 16/25 = 0.64$.

(d) Again we use Equation (3). To meet the specified requirement, we choose $c = 0.1$. Since $\mathrm{Var}[X] = 0.16$, we must have

$$\frac{0.16}{n(0.1)^2} \leq 0.05 \tag{5}$$

The smallest value that meets the requirement is $n = 320$.

Point Estimates of Model Parameters

Model Parameters

- The general problem is estimation of a *parameter* of a probability model.
- A parameter is any number that can be calculated from the probability model.
- For example, for an arbitrary event A, $\mathsf{P}[A]$ is a model parameter.

Estimates of Model Parameters

- Consider an experiment that produces observations of sample values of the random variable X.
- The observations are sample values of the random variables $X_1, X_2, \ldots$, all with the same probability model as X.
- Assume that r is a parameter of the probability model.
- We use the observations $X_1, X_2, \ldots$ to produce a sequence of estimates of r.
- The estimates $\hat{R}_1, \hat{R}_2, \ldots$ are all random variables.
- $\hat{R}_1$ is a function of X_1.
- $\hat{R}_2$ is a function of X_1 and X_2, and in general $\hat{R}_n$ is a function of $X_1, X_2, \ldots, X_n$.

Definition 10.3 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

Definition 10.4 Unbiased Estimator

An estimate, $\widehat{R}$*, of parameter* r *is* unbiased *if* $\text{E}[\widehat{R}] = r$*; otherwise,* $\widehat{R}$ *is* biased.

Asymptotically Unbiased

Definition 10.5 Estimator

The sequence of estimators $\widehat{R}_n$ *of parameter* r *is* asymptotically unbiased *if*

$$\lim_{n\to\infty} \mathsf{E}[\widehat{R}_n] = r.$$

Definition 10.6 Mean Square Error

The mean square error *of estimator* $\hat{R}$ *of parameter* r *is*

$$e = \mathrm{E}\left[(\hat{R} - r)^2\right].$$

Theorem 10.8

If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has mean square error $e_n = \text{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0$, then the sequence $\hat{R}_n$ is consistent.

Proof: Theorem 10.8

Since $\mathsf{E}[\hat{R}_n] = r$, we apply the Chebyshev inequality to $\hat{R}_n$. For any constant $\epsilon > 0$,

$$\mathsf{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \frac{\mathsf{Var}[\hat{R}_n]}{\epsilon^2}. \tag{10.20}$$

In the limit of large n, we have

$$\lim_{n\to\infty} \mathsf{P}\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \lim_{n\to\infty} \frac{\mathsf{Var}[\hat{R}_n]}{\epsilon^2} = 0. \tag{10.21}$$

Example 10.5 Problem

In any interval of k seconds, the number N_k of packets passing through an Internet router is a Poisson random variable with expected value $\mathsf{E}[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of the parameter r packets/second. Is each estimate $\hat{R}_k$ an unbiased estimate of r? What is the mean square error e_k of the estimate $\hat{R}_k$? Is the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ consistent?

Example 10.5 Solution

First, we observe that $\hat{R}_k$ is an unbiased estimator since

$$E[\hat{R}_k] = E[N_k/k] = E[N_k]/k = r. \tag{10.22}$$

Next, we recall that since N_k is Poisson, $\text{Var}[N_k] = kr$. This implies

$$\text{Var}[\hat{R}_k] = \text{Var}\left[\frac{N_k}{k}\right] = \frac{\text{Var}[N_k]}{k^2} = \frac{r}{k}. \tag{10.23}$$

Because $\hat{R}_k$ is unbiased, the mean square error of the estimate is the same as its variance: $e_k = r/k$. In addition, since $\lim_{k\to\infty} \text{Var}[\hat{R}_k] = 0$, the sequence of estimators $\hat{R}_k$ is consistent by Theorem 10.8.

Theorem 10.9

The sample mean $M_n(X)$ is an unbiased estimate of $\mathsf{E}[X]$.

Theorem 10.10

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = \mathsf{E}\left[(M_n(X) - \mathsf{E}[X])^2\right] = \mathsf{Var}[M_n(X)] = \frac{\mathsf{Var}[X]}{n}.$$

Standard Error

- In the terminology of statistical inference, $\sqrt{e_n}$, the standard deviation of the sample mean, is referred to as the *standard error* of the estimate.

- The standard error gives an indication of how far we should expect the sample mean to deviate from the expected value.

- In particular, when X is a Gaussian random variable (and $M_n(X)$ is also Gaussian), Problem 10.4.1 asks you to show that

$$\mathrm{P}\left[\mathrm{E}[X] - \sqrt{e_n} \leq M_n(X) \leq \mathrm{E}[X] + \sqrt{e_n}\right] = 2\Phi(1) - 1 \approx 0.68. \tag{10.24}$$

  In words, Equation (10.24) says there is roughly a two-thirds probability that the sample mean is within one standard error of the expected value.

- This same conclusion is approximately true when n is large and the central limit theorem says that $M_n(X)$ is approximately Gaussian.

Example 10.6 Problem

How many independent trials n are needed to guarantee that $\hat{P}_n(A)$, the relative frequency estimate of $\mathrm{P}[A]$, has standard error ≤ 0.1?

Example 10.6 Solution

Since the indicator X_A has variance $\text{Var}[X_A] = \text{P}[A](1 - \text{P}[A])$, Theorem 10.10 implies that the mean square error of $M_n(X_A)$ is

$$e_n = \frac{\text{Var}[X]}{n} = \frac{\text{P}[A](1 - \text{P}[A])}{n}. \tag{10.25}$$

We need to choose n large enough to guarantee $\sqrt{e_n} \le 0.1$ ($e_n \le= 0.01$) even though we don't know $\text{P}[A]$. We use the fact that $p(1-p) \le 0.25$ for all $0 \le p \le 1$. Thus, $e_n \le 0.25/n$. To guarantee $e_n \le 0.01$, we choose $n = 0.25/0.01 = 25$ trials.

Theorem 10.11

If X has finite variance, then the sample mean $M_n(X)$ is a sequence of consistent estimates of $\mathsf{E}[X]$.

Proof: Theorem 10.11

By Theorem 10.10, the mean square error of $M_n(X)$ satisfies

$$\lim_{n\to\infty} \text{Var}[M_n(X)] = \lim_{n\to\infty} \frac{\text{Var}[X]}{n} = 0. \tag{10.26}$$

By Theorem 10.8, the sequence $M_n(X)$ is consistent.

Estimating the Variance

- When $\mathsf{E}[X]$ is a known quantity μ_X, we know $\mathsf{Var}[X] = \mathsf{E}[(X - \mu_X)^2]$.

- In this case, we can use the sample mean of $W = (X - \mu_X)^2$ to estimate Var[X].,

$$M_n(W) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_X)^2 . \tag{10.28}$$

  If Var[W] exists, $M_n(W)$ is a consistent, unbiased estimate of Var[X].

- When the expected value μ_X is unknown, the situation is more complicated because the variance of X depends on μ_X.

- We cannot use Equation (10.28) if μ_X is unknown.

- In this case, we replace the expected value μ_X by the sample mean $M_n(X)$.

Definition 10.7 Sample Variance

The sample variance of n independent observations of random variable X is

$$V_n(X) = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - M_n(X)\right)^2 .$$

Theorem 10.12

$$\mathsf{E}\left[V_n(X)\right] = \frac{n-1}{n}\mathsf{Var}[X].$$

Proof: Theorem 10.12

Substituting Definition 10.1 of the sample mean $M_n(X)$ into Definition 10.7 of sample variance and expanding the sums, we derive

$$V_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j. \tag{10.29}$$

Because the X_i are iid, $\mathsf{E}[X_i^2] = \mathsf{E}[X^2]$ for all i, and $\mathsf{E}[X_i]\,\mathsf{E}[X_j] = \mu_X^2$. By Theorem 5.16(a),

$$\mathsf{E}[X_i X_j] = \mathsf{Cov}[X_i, X_j] + \mathsf{E}[X_i]\,\mathsf{E}[X_j] = \mathsf{Cov}[X_i, X_j] + \mu_X^2. \tag{10.30}$$

Combining these facts, the expected value of V_n in Equation (10.29) is

$$\begin{aligned}\mathsf{E}[V_n] &= \mathsf{E}\left[X^2\right] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathsf{Cov}[X_i, X_j] + \mu_X^2\right) \\ &= \mathsf{Var}[X] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}[X_i, X_j]. \end{aligned} \tag{10.31}$$

Since the double sum has n^2 terms, $\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_X^2 = n^2\mu_X^2$. Of the n^2 covariance terms, there are n terms of the form $\mathsf{Cov}[X_i, X_i] = \mathsf{Var}[X]$, while the remaining covariance terms are all 0 because X_i and X_j are independent for $i \neq j$. This implies

$$\mathsf{E}[V_n] = \mathsf{Var}[X] - \frac{1}{n^2}(n\,\mathsf{Var}[X]) = \frac{n-1}{n}\mathsf{Var}[X]. \tag{10.32}$$

Theorem 10.13

The estimate

$$V_n'(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M_n(X))^2$$

is an unbiased estimate of $\mathrm{Var}[X]$.

Proof: Theorem 10.13

Using Definition 10.7, we have

$$V_n'(X) = \frac{n}{n-1} V_n(X), \tag{10.34}$$

and

$$\mathsf{E}\left[V_n'(X)\right] = \frac{n}{n-1}\,\mathsf{E}\left[V_n(X)\right] = \mathsf{Var}[X]. \tag{10.35}$$

Quiz 10.4

X is the continuous uniform $(-1, 1)$ random variable. Find the mean square error, $\mathsf{E}[(\mathsf{Var}[X] - V_{100}(X))^2]$, of the sample variance estimate of $\mathsf{Var}[X]$, based on 100 independent observations of X.

Quiz 10.4 Solution

Define the random variable $W = (X - \mu_X)^2$. Observe that $V_{100}(X) = M_{100}(W)$. By Theorem 10.10, the mean square error is

$$\mathsf{E}\left[(M_{100}(W) - \mu_W)^2\right] = \frac{\mathsf{Var}[W]}{100}. \tag{1}$$

Observe that $\mu_X = 0$ so that $W = X^2$. Thus,

$$\mu_W = \mathsf{E}\left[X^2\right] = \int_{-1}^{1} x^2 f_X(x)\, dx = 1/3, \tag{2}$$

$$\mathsf{E}\left[W^2\right] = \mathsf{E}\left[X^4\right] = \int_{-1}^{1} x^4 f_X(x)\, dx = 1/5. \tag{3}$$

Therefore $\mathsf{Var}[W] = \mathsf{E}[W^2] - \mu_W^2 = 1/5 - (1/3)^2 = 4/45$ and the mean square error is $4/4500 = 0.0009$.

Section 10.5

Matlab

Sample Mean Traces

- In particular, for a random variable X, we can view a set of iid samples $X_1, \ldots, X_n$ as a random vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix}'$.

- This vector of iid samples yields a vector of sample mean values

$$\mathbf{M}(\mathbf{X}) = \begin{bmatrix} M_1(X) & M_2(X) & \cdots & M_n(X) \end{bmatrix}'$$

  where

$$M_k(X) = \frac{X_1 + \cdots + X_k}{k} \tag{10.36}$$

- We call a graph of the sequence $M_k(X)$ versus k a *sample mean trace*.

- By graphing the sample mean trace as a function of n we can observe the convergence of the point estimate $M_k(X)$ to $\mathrm{E}[X]$.

Example 10.7 Problem

Write a Matlab function `bernoullitraces(n,m,p)` to generate m sample mean traces, each of length n, for the sample mean of a Bernoulli (p) random variable.

Example 10.7 Solution

```
function MN=bernoullitraces(n,m,p);
x=reshape(bernoullirv(p,m*n),n,m);
nn=(1:n)'*ones(1,m);
MN=cumsum(x)./nn;
stderr=sqrt(p*(1-p))./sqrt((1:n)');
plot(1:n,0.5+stderr,...
    1:n,0.5-stderr,1:n,MN);
```

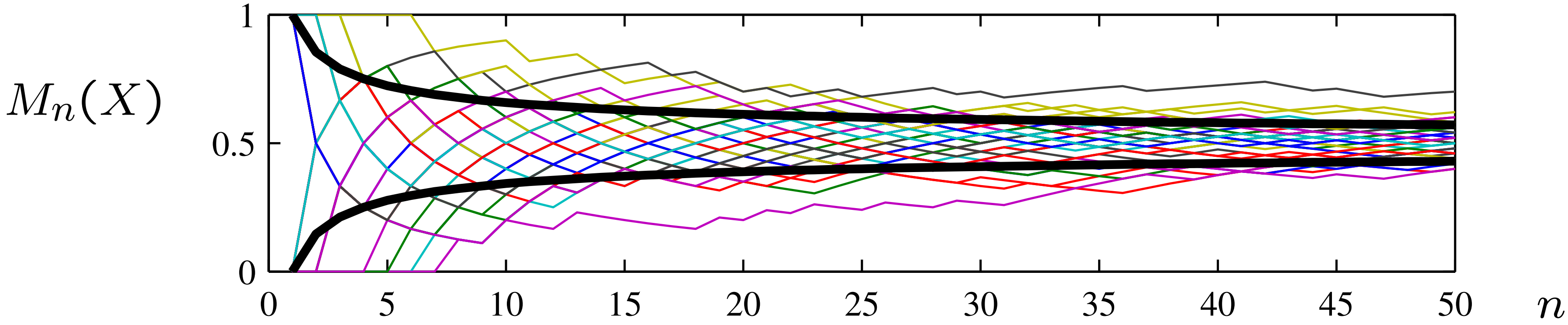
In `bernoullitraces`, each column of `x` is an instance of a random vector $\mathbf{X}$ with iid Bernoulli (p) components. Similarly, each column of `MN` is an instance of the vector $\mathbf{M}(\mathbf{X})$.

The output graphs each column of `MN` as a function of the number of trials n. In addition, we calculate the standard error $\sqrt{e_k}$ and overlay graphs of $p - \sqrt{e_k}$ and $p + \sqrt{e_k}$. Equation (10.24) says that at each step k, we should expect to see roughly two-thirds of the sample mean traces in the range

$$p - \sqrt{e_k} \le M_k(X) \le p + \sqrt{e_k}. \tag{10.37}$$

A sample graph of `bernoullitraces(50,40,0.5)` is shown in Figure 10.1. The figure shows how at any given step, approximately two thirds of the sample mean traces are within one standard error of the expected value.

Figure 10.1



Sample output of **`bernoullitraces.m`**, including the deterministic standard error graphs. The graph shows how at any given step, about two thirds of the sample means are within one standard error of the true mean.

Quiz 10.5

Generate $m = 1000$ traces (each of length $n = 100$) of the sample mean of a Bernoulli (p) random variable. At each step k, calculate M_k and the number of traces, such that M_k is within one standard error of the expected value p. Graph $T_k = M_k/m$ as a function of k. Explain your results.

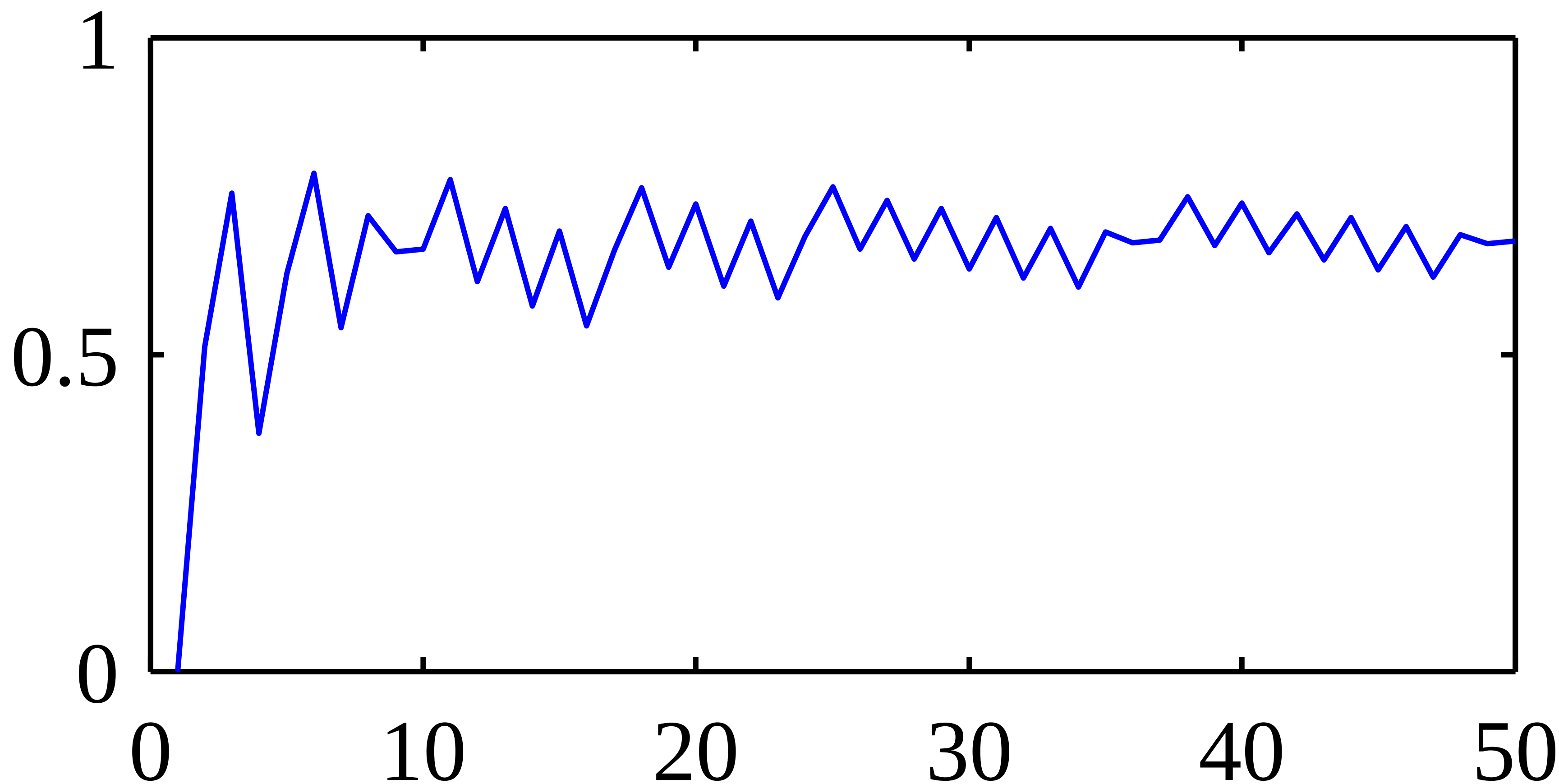
Quiz 10.5 Solution

Following the `bernoullitraces.m` approach, we generate $m = 1000$ sample paths, each sample path having $n = 100$ Bernoulli traces. at time k, `OK(k)` counts the fraction of sample paths that have sample mean within one standard error of p. The program `bernoullisample.m` generates graphs the number of traces within one standard error as a function of the time, i.e. the number of trials in each trace.

```
function OK=bernoullisample(n,m,p);
x=reshape(bernoullirv(p,m*n),n,m);
nn=(1:n)'*ones(1,m);
MN=cumsum(x)./nn;
stderr=sqrt(p*(1-p))./sqrt((1:n)');
stderrmat=stderr*ones(1,m);
OK=sum(abs(MN-p)<stderrmat,2)/m;
plot(1:n,OK);
```

[Continued]

The following graph was generated by `bernoullisample(50,5000,0.5)`:



As we would expect, as m gets large, the fraction of traces within one standard error approaches $2\Phi(1)-1 \approx 0.68$. The unusual sawtooth pattern, though perhaps unexpected, is examined in Problem 10.5.1.