Section 10.1

Sample Mean: Expected Value and Variance

Definition 10.1 Sample Mean

For iid random variables $X_1, \dots, X_n$ with PDF $f_X(x)$, the $\boxed{\text{sample mean}}$ *of X is the random variable*

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

Sample mean is not an expected value (or expectation)

$$\text{Random Variable} \leftarrow M_n(X) \neq E[X] \rightarrow \text{A constant/number}$$

As n increases, $M_n(X) \rightarrow E[X]$

Theorem 10.1

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

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

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

$M_n(X)$ is the "average" or "sum times a factor"

Recall the property of expected value of sum

For any set of random variables $X_1, \ldots, X_n$, the sum $W_n = X_1 + \cdots + X_n$ has expected value

$$\mathrm{E}[W_n] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n].$$

$$\begin{aligned}
\mathrm{E}[M_n(X)] &= \frac{1}{n}(\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]) \\
&= \frac{1}{n}(\mathrm{E}[X] + \cdots + \mathrm{E}[X]) \\
&= \mathrm{E}[X]
\end{aligned}$$

Theorem 10.1

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

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

$$\begin{aligned}\mathsf{Var}[M_n(X)] &= \mathsf{Var}[\frac{X_1 + \cdots + X_n}{n}] \\ &= \mathsf{Var}[X_1 + \cdots + X_n]/n^2\end{aligned}$$

When $X_1, \ldots, X_n$ are uncorrelated,

$$\mathsf{Var}[W_n] = \mathsf{Var}[X_1] + \cdots + \mathsf{Var}[X_n].$$

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

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

Theorem 10.1

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

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

As n increases to infinite,

the sample mean goes to the expected value

the variance of sample mean goes to zero.

the expected value of sample mean is always expected value.

Recall the physical meaning of variance:

how far a random variable is likely to be from its expected value.

Useful Inequalities in Probability

- Markov Inequality
- Chebyshev Inequality

Theorem 10.2 Markov Inequality

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

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

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

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

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

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

Or given condition c>0,

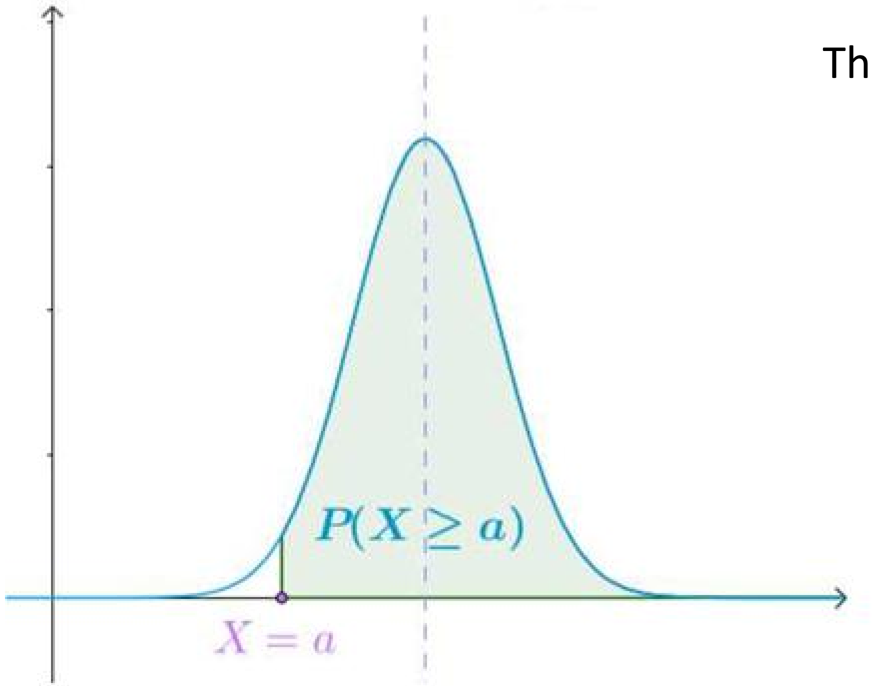
$$\mathsf{P}\left[X \geq c\ \right] \leq \frac{\mathsf{E}[X]}{c}$$

Theorem 10.2 Markov Inequality

positive

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

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



The average net worth of citizens in Banana Republic is $ 51,350

$$P(X \geq 1000000) \leq \frac{51350}{1000000} \approx 5.14\%$$

So, Markov estimates that less than 5 out of 100 persons are millionaires in Banana Republic .

But, can we do better?

Theorem 10.3 Chebyshev Inequality

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

$$P\left[X \geq c^2\right] = P\left[(Y - \mu_Y)^2 \geq c^2\right] \leq \frac{E\left[(Y - \mu_Y)^2\right]}{c^2} = \frac{\text{Var}[Y]}{c^2}. \tag{1}$$

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

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

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

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 $\mathsf{P}[X \geq 11]$.

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

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

$$\mathsf{P}\left[X \geq 11\right] = \mathsf{P}\left[X - \mu_X \geq 11 - \mu_X\right] = \mathsf{P}\left[|X - \mu_X| \geq 5.5\right]. \tag{1}$$

Now we use the Chebyshev inequality to obtain

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

Theorem 10.2 Markov Inequality

positive

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

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

Average net worth all Banana Republicans \$ 51,350

$$P(X \geq 1000000) \leq \frac{51350}{1000000} \approx 5.14\%$$

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

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

Average net worth all Banana Republicans \$ 51,350,
Standard deviation is \$44,000

$$P(X \geq 1000000) = P(|X - 51350| \geq 1000000 - 51350) \leq \frac{44000^2}{948650^2} \approx 0.2\%$$

2 out of 1000 individuals is a more reasonable result.

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 random variables. Let $W = X_1 + X_2 + X_3$ equal the time required to serve the waiting customers. $E[W] = 6$. $\text{Var}[W] = 12$. 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.

Quiz 10.2 Solution

(a) By the Central Limit Theorem,

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

(b) From the Markov inequality, we know that

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

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

$$\begin{aligned} P[W \geq 20] &= P[|W - E[W]| \geq 14] \\ &\leq \frac{\text{Var}[W]}{14^2} = \frac{3}{49} = 0.061. \end{aligned}$$

Section 10.3

Laws of Large Numbers

Theorem 10.5 Weak Law of Large Numbers (Finite Samples)

For any constant $c > 0$,

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

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

Describe the distance between sample mean and expected value

Put sample mean and Chebyshev inequality together

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

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

$\mathsf{E}\left[Y\right] = \mathsf{E}\left[M_n(X)\right] = \mu_X \qquad \mathsf{Var}[Y] = \mathsf{Var}[M_n(X)] = \mathsf{Var}[X]/n.$

Theorem 10.6 Weak Law of Large Numbers (Infinite Samples)

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

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

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

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

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

- The probability that the sample mean is within $2c$ units of expected value goes to one as the number of samples approaches infinity.
- This law holds for all random variables X with finite variance

Definition 10.2 Convergence in Probability

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

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

- The probability that $Y_n \neq y$ goes to zero as the number of samples approaches infinity.

Theorem 10.7

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

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

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 $\mathsf{Var}[X]$.

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

$$\boxed{\mathsf{P}[|M_n(X) - \mu_X| \geq c] \leq \frac{\mathsf{Var}[X]}{nc^2}}$$

(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 $\mathsf{E}[X] = p = 0.8$ 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

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

Note that $\mathsf{E}[X] = P_X(1) = p$. 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, $\mathrm{P}[A]$ 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, $\hat{R}$, *of parameter* r *is* unbiased *if* $\mathrm{E}[\hat{R}] = r$; *otherwise,* $\hat{R}$ *is* biased.

Definition 10.5 Estimator

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

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

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

$$P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2}. \tag{1}$$

In the limit of large n, we have

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

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 $\mathrm{E}[N_k] = kr$ 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

$$\mathsf{E}[\hat{R}_k] = \mathsf{E}\left[N_k/k\right] = \mathsf{E}\left[N_k\right]/k = r. \tag{1}$$

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

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

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 $\lim_{k\to\infty} \mathsf{Var}[\hat{R}_k] = 0$, 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

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

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

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

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

Proof: Theorem 10.11

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

$$\lim_{n\to\infty} \mathsf{Var}[M_n(X)] = \lim_{n\to\infty} \frac{\mathsf{Var}[X]}{n} = 0. \tag{1}$$

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

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

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

- 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

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

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

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

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), $\mathsf{E}[X_iX_j] = \mathsf{Cov}[X_i, X_j] + \mathsf{E}[X_i]\,\mathsf{E}[X_j]$. Thus, $\mathsf{E}[X_iX_j] = \mathsf{Cov}[X_i, X_j] + \mu_X^2$. Combining these facts, the expected value of V_n in Equation (10.29) is

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

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

$$\mathsf{E}\left[V_n\right] = \mathsf{Var}[X] - \frac{1}{n^2}\left(n\,\mathsf{Var}[X]\right) = \frac{n-1}{n}\mathsf{Var}[X]. \tag{3}$$

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

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

and

$$\mathsf{E}\left[V'_n(X)\right] = \frac{n}{n-1}\,\mathsf{E}\left[V_n(X)\right] = \mathsf{Var}[X]. \tag{2}$$

Quiz 10.4

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

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

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

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

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

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