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* $\boxed{\text{sample mean}}$ *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)

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

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

Theorem 10.1

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

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

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

$\rightarrow$ $M_n(X)$ 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,

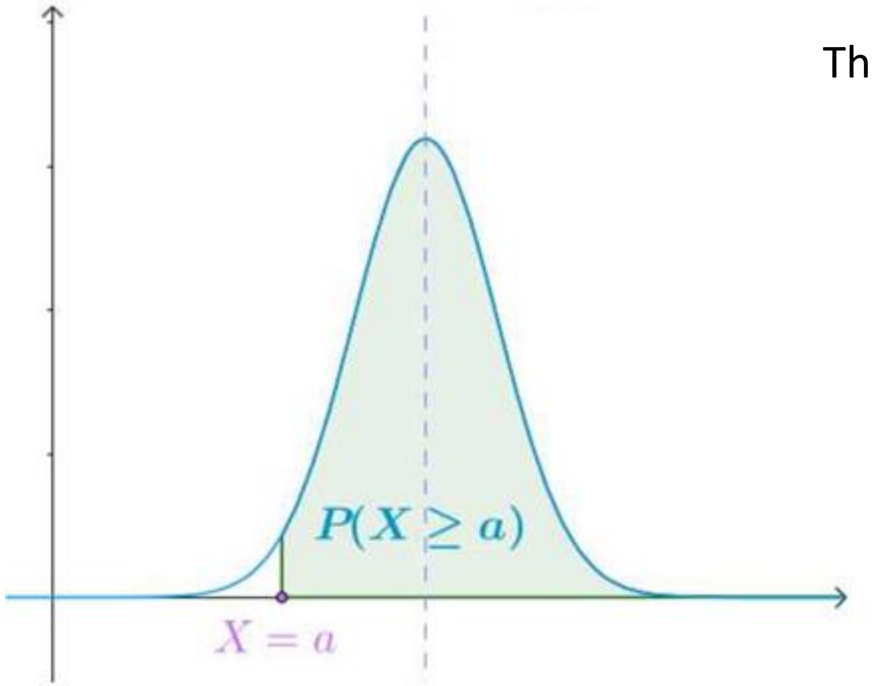
$$\mathsf{P}\left[X \geq c\right] \leq \frac{\mathsf{E}[X]}{c}$$

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

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

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 $c > 0$,

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

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

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

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

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 $c > 0$,

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

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,

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

(b) From the Markov inequality, we know that

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

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

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

Section 10.3

Laws of Large Numbers

Theorem 10.5 Weak Law of Large Numbers (Finite Samples)

For any constant $c > 0$,

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

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

Describe the distance between sample mean and expected value

Put sample mean and Chebyshev inequality together

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

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

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

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 $\epsilon > 0$,

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

- 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 $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.

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

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

(c) Use Theorem 10.5 to find α such that

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

(d) How many samples n are needed to guarantee

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

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

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

(b) By Theorem 10.1,

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

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

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

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

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

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