STAT 3375Q: Introduction to Mathematical Statistics I

Lecture 5: Discrete Random Variables; Expected Value

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

January 31, 2024

Outline

1. Previously...
   - Random Variables
   - Probability Distribution of Discrete Random Variables

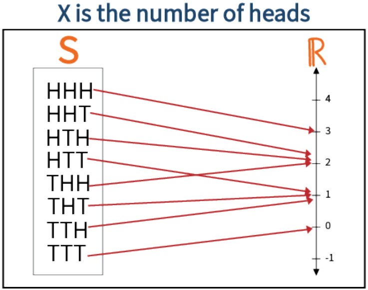
2. Expected Value of Discrete Random Variables

3. Variance of Discrete Random Variables

Previously...

Random Variables

- Random Variable: is a function that maps outcomes into real numbers. Example: Three fair coins are flipped. Let the random variable X be the number of heads obtained.

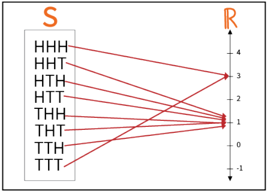


- Discrete random variables: random variables whose values take only a finite or countably infinite number of possible values.
- Continuous random variables: are random variables that can take on an infinite number of possible values.

Probability Distribution of Discrete Random Variables

- Some notations:
  - random variables: uppercase letter.
    Example: Let Y be the absolute difference between the number of heads and the number of tails.
  - particular value of the random variable (result): lowercase letter.
    Example: $y = 3$, $y = 1$.
- Goal: Compute $P(Y = y)$.
  Read as: *Probability that the (random variable) Y takes on the value y.*

Y is the absolute difference between num. of heads and tails



Definition 3.2

The probability that Y takes on the value y, denoted $P(Y = y)$, is defined as the sum of the probabilities of all sample points in S that are assigned the value y.

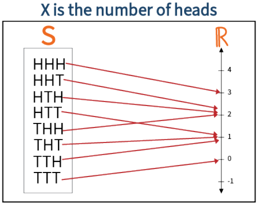
Note: $P(Y = y)$ is also often written as $p(y)$, which is read as "p of y".

Probability Distribution of Discrete Random Variables

Definition 3.3

The probability distribution for a discrete variable Y can be represented by a formula, a table, or a graph that provides $p(y) = P(Y = y)$ for all y.

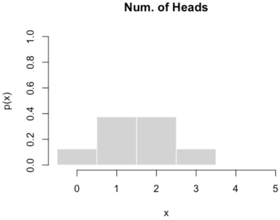
- Example:



- Table of the probability distribution:

x	$p(x)$
0	1/8
1	3/8
2	3/8
3	1/8

- Histogram of the probability distribution:



Probability Distribution of Discrete Random Variables

A basketball player takes 4 independent free throws with a probability of 0.7 of getting a basket on each shot. Let $Y =$ the number of baskets he gets. Write out the full probability distribution for Y.

Answer:

y	$p(y)$
0	$C_0^4(1-0.7)(1-0.7)(1-0.7)(1-0.7) = 0.0081$
1	$C_1^4(0.7)(1-0.7)(1-0.7)(1-0.7) = 0.0756$
2	$C_2^4(0.7)(0.7)(1-0.7)(1-0.7) = 0.2646$
3	$C_3^4(0.7)(0.7)(0.7)(1-0.7) = 0.4116$
4	$C_4^4(0.7)(0.7)(0.7)(0.7) = 0.2401$

Note: $y = 1$ corresponds to the following event:
$\{FT_1^{\text{made}}, FT_2^{\text{miss}}, FT_3^{\text{miss}}, FT_4^{\text{miss}}\} \cup \{FT_1^{\text{miss}}, FT_2^{\text{made}}, FT_3^{\text{miss}}, FT_4^{\text{miss}}\} \cup \{FT_1^{\text{miss}}, FT_2^{\text{miss}}, FT_3^{\text{made}}, FT_4^{\text{miss}}\} \cup \{FT_1^{\text{miss}}, FT_2^{\text{miss}}, FT_3^{\text{miss}}, FT_4^{\text{made}}\}$.

This is why we need to multiply C_1^4 since there are C_1^4 ways to choose which free throw to make.

Expected Value of Discrete Random Variables

Expected Value of Discrete Random Variables

Definition 3.4: Expected Value

Let Y be a discrete random variable with the probability function $p(y)$. Then the *expected value* of Y, denoted $E(Y)$, is defined to be

$$E(Y) = \sum_{y} y p(y).$$

Remarks:

- Expected value of Y $\Leftrightarrow$ Expectation of Y
- The expected value of Y is also very often called the mean of Y and is denoted by μ_Y or μ.
- The expected value, or mean, of Y gives a single value that acts as a representative or average of the values of Y, and for this reason it is often called a measure of central tendency.
- Expected value can be seen as an average of the values that the random variable Y can take but weighted by their corresponding probabilities.
- The expected value gives an idea of the average value attained by the random variable Y when the experiment is repeated many times.

Expected Value of Discrete Random Variables

Applications:

- to determine which of the outcomes is most likely to happen
- to determine the average payoff or loss in a game of chance
- to determine premiums on insurance policies

Remarks: You can think of the expected value as the number that is representative of the values that the random variable can take...

Expected Value of Discrete Random Variables

Example:
What is the expected value of random variable X if its probability distribution is the following:

x	$p(x)$
0	1/5
1	1/5
2	1/5
3	1/5
4	1/5

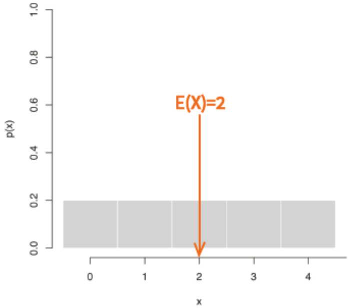
Expected Value of Discrete Random Variables

Strategy: Use the formula for computing expected value: $E(X) = \sum_x xp(x)$.

Example:

What is the expected value of random variable X if its probability distribution is the following:

x	$p(x)$	$xp(x)$
0	1/5	$0 \times 1/5 = 0$
1	1/5	$1 \times 1/5 = 1/5$
2	1/5	$2 \times 1/5 = 2/5$
3	1/5	$3 \times 1/5 = 3/5$
4	1/5	$4 \times 1/5 = 4/5$



Answer:

$$E(X) = \sum_x xp(x) = 0 + 1/5 + 2/5 + 3/5 + 4/5 = 2.$$

Note that we can arrive at the same answer by simply computing the average of the x values: $\frac{0+1+2+3+4}{5} = \frac{10}{5} = 2$.

Expected Value of Discrete Random Variables

Example:

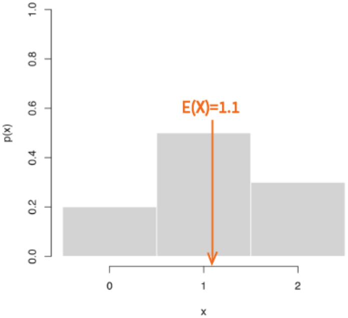
A men's soccer team plays soccer zero, one, or two days a week. The probability that they play zero days is 0.2, the probability that they play one day is 0.5, and the probability that they play two days is 0.3. Find the expected value of the number of days per week the men's soccer team plays soccer.

Answer: Let X be the number of days the men's soccer team plays soccer per week.

x	$p(x)$	$xp(x)$
0	0.2	$0 \times 0.2 = 0$
1	0.5	$1 \times 0.5 = 0.5$
2	0.3	$2 \times 0.3 = 0.6$

$$E(X) = \sum_x xp(x) = 0+0.5+0.6 = 1.1.$$

The men's soccer team would, on the average, expect to play soccer 1.1 days per week.



Expected Value of Discrete Random Variables

Example:

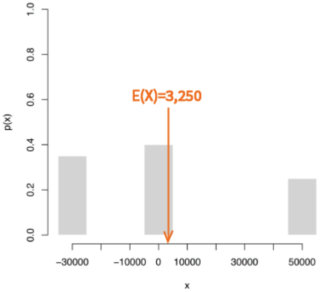
Your company plans to invest in a particular project. There is a 35% chance that you will lose \$30,000, a 40% chance that you will break even, and a 25% chance that you will make \$55,000. Based solely on this information, what should you do?

Answer: Let X be the return on investment of the project.

x	$p(x)$	$xp(x)$
-30,000	0.35	$-30,000 \times 0.35 = -10,500$
0	0.40	$0 \times 0.40 = 0$
55,000	0.25	$55,000 \times 0.25 = 13,750$

$$E(X) = \sum_x xp(x) = -10,500 + 0 + 13,750 = 3,250.$$

The expected value of the return on investment is 3,250. Since the return is positive, you should proceed with the project.



Expected Value of Discrete Random Variables

Example: *Computing Auto Insurance Premiums*

An automobile insurance company has determined the probabilities for various claim amounts for drivers ages 16 through 21 as shown in the table. Calculate the expected value and describe what this means in practical terms.

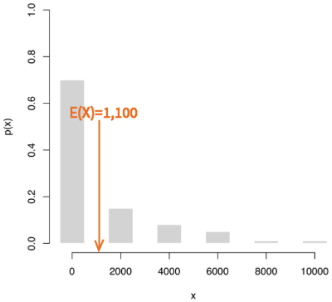
Amount of Claim	Probability
\$0	0.70
\$2,000	0.15
\$4,000	0.08
\$6,000	0.05
\$8,000	0.01
\$10,000	0.01

Expected Value of Discrete Random Variables

Example: *Computing Auto Insurance Premiums*

An automobile insurance company has determined the probabilities for various claim amounts for drivers ages 16 through 21 as shown in the table. Calculate the expected value and describe what this means in practical terms.

Amount of Claim x	Probability $p(x)$	$xp(x)$
\$0	0.70	$0 \times 0.70 = 0$
\$2,000	0.15	$2{,}000 \times 0.15 = 300$
\$4,000	0.08	$4{,}000 \times 0.08 = 320$
\$6,000	0.05	$6{,}000 \times 0.05 = 300$
\$8,000	0.01	$8{,}000 \times 0.01 = 80$
\$10,000	0.01	$10{,}000 \times 0.01 = 100$



Answer: Let X be the insurance claim.

$$E(X) = \sum_x xp(x) = 0 + 300 + 320 + 300 + 80 + 100 = 1,100.$$

This means that in the long run, the insurance company can expect to pay each customer \$1,100. The company needs to charge its customers at least \$1,100 to break even.

Expected Value of Discrete Random Variables

Example: *Computing Life Insurance Premiums*

A 40-year-old man in the U.S. has a 0.242% risk of dying during the next year. An insurance company charges $275 for a life-insurance policy that pays a $100,000 death benefit. What is the expected value for the person buying the insurance?

Expected Value of Discrete Random Variables

Example: *Computing Life Insurance Premiums*

A 40-year-old man in the U.S. has a 0.242% risk of dying during the next year. An insurance company charges \$275 for a life-insurance policy that pays a \$100,000 death benefit. What is the expected value for the person buying the insurance?

Answer: Let X be the insurance payoff to the customer.

Payoff x	Probability $p(x)$	$xp(x)$
\$100,000 - \$275 = \$99,725	0.00242	\$99,725 × 0.00242 = \$241.33
−\$275	(1 - 0.00242) = 0.99758	−\$275 × 0.99758 = −\$274.33

$$E(X) = \sum_x xp(x) = \$241.33 - \$274.33 = -\$33.$$

Expected Value of Discrete Random Variables

Example:

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and does not move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog. What is the probability distribution and expectation of X?

Answer: Let L be the event of left step, R of right step, and N of no step.

Event	x	$p(x)$
LL	-2	p_2^2
NL ∪ LN	-1	$2p_2p_3$
NN ∪ LR ∪ RL	0	$p_3^2 + 2p_1p_2$
NR ∪ RN	1	$2p_1p_3$
RR	2	p_1^2

Check that the probabilities sum to 1 for the table above to be a valid probability distribution: $p_2^2 + 2p_2p_3 + p_3^2 + 2p_1p_2 + 2p_1p_3 + p_1^2 = (p_1 + p_2 + p_3)^2 = 1^2 = 1$ since we know from the given that $p_1 + p_2 + p_3 = 1$.

$\mu = E(X) = \sum_x xp(x) = -2p_2^2 - 2p_2p_3 + 0(p_3^2 + 2p_1p_2) + 2p_1p_3 + 2p_1^2 = 2(p_1^2 - p_2^2)$.

Expected Value of Discrete Random Variables

Theorem 3.2: Expected Value of Functions of Random Variables

Let Y be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of Y. Then the expected value of $g(Y)$, denoted $E\{g(Y)\}$, is given by

$$E\{g(Y)\} = \sum_{y} g(y)p(y).$$

Expected Value of Discrete Random Variables

Example:

X is a random variable with the following probability distribution:

x	$p(x)$
-3	1/6
6	1/2
9	1/3

Suppose that $g(X) = X^2 + 2$, find $E\{g(X)\}$.

Expected Value of Discrete Random Variables

Strategy: Use the formula for computing expected value: $E\{g(X)\} = \sum_x g(x)p(x)$.

Example:

X is a random variable with the following probability distribution:

x	$p(x)$	$g(x)$	$g(x)p(x)$
-3	1/6	$(-3)^2 + 2 = 11$	$11/6$
6	1/2	$6^2 + 2 = 38$	$38/2$
9	1/3	$9^2 + 2 = 83$	$83/3$

Suppose that $g(X) = X^2 + 2$, find $E\{g(X)\}$.

Answer:

$$E\{g(X)\} = \sum_x g(x)p(x) = 11/6 + 38/2 + 83/3 = 48.5.$$

Expected Value of Discrete Random Variables

Theorem 3.3: Expected Value of a Constant

Let Y be a discrete random variable with probability function $p(y)$ and c be a constant. Then $E(c) = c$.

Proof of Theorem 3.3:

Suppose $g(Y) = c$. By Theorem 3.2, we know that

$$\begin{aligned} E\{g(Y)\} &= \sum_y g(y)p(y) \\ E(c) &= \sum_y cp(y) \\ &= c\sum_y p(y) \\ &= c(1) \quad \text{since } \sum_y p(y) = 1 \text{ (Theorem 3.1)} \\ &= c. \end{aligned}$$

Expected Value of Discrete Random Variables

Theorem 3.4: Expected Value of a Scaled Random Variable

Let Y be a discrete random variable with probability function $p(y)$, $g(Y)$ be a function of Y, and c be a constant. Then

$$E\{cg(Y)\} = cE\{g(Y)\}$$

Proof of Theorem 3.4:

By Theorem 3.2, we know that

$$\begin{aligned}
E\{g(Y)\} &= \sum_y g(y)p(y) \\
E\{cg(Y)\} &= \sum_y cg(y)p(y) \\
&= c\sum_y g(y)p(y) \\
&= cE\{g(Y)\}. \quad \text{definition of expected value}
\end{aligned}$$

Theorem 3.5: Expected Value of a Sum of Random Variables

Let Y be a discrete random variable with probability function $p(y)$, $g_1(Y), g_2(Y), \ldots, g_k(Y)$ be k functions of Y. Then

$$E\{g_1(Y)+g_2(Y)+\ldots+g_k(Y)\} = E\{g_1(Y)\}+E\{g_2(Y)\}+\ldots+E\{g_k(Y)\}.$$

Proof of Theorem 3.5:

Suppose $k=2$. By Theorem 3.2, we know that

$$\begin{aligned} E\{g(Y)\} &= \sum_y g(y)p(y) \\ E\{g_1(Y)+g_2(Y)\} &= \sum_y \{g_1(y)+g_2(y)\}p(y) \\ &= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y) \\ &= E\{g_1(Y)\} + E\{g_2(Y)\}. \quad \text{definition of expected value} \end{aligned}$$

Variance of Discrete Random Variables

Variance of Discrete Random Variables

Definition 3.5: Variance

If Y is a random variable with mean $E(Y) = \mu$, the *variance* of a random variable Y, denoted $V(Y)$ or σ^2, is defined to be the expected value of $(Y - \mu)^2$. That is,

$$V(Y) = E\{(Y - \mu)^2\}.$$

The *standard deviation* of Y, denoted $SD(Y)$ or $sd(Y)$ or σ, is the positive square root of $V(Y)$.

Remarks:

- Variance is a measure of the dispersion, or scatter, of the values of the random variable about the mean μ.
- If the values tend to be concentrated near the mean, the variance is small.
- If the values tend to be far from the mean, the variance is large.
- In finance, standard deviation is also often termed as volatility.
- Variance is also commonly used to measure uncertainty of the outcome or prediction.

Variance of Discrete Random Variables

Arrange the figures based on their variance, from smallest to largest.

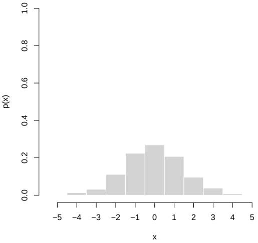


Figure: A

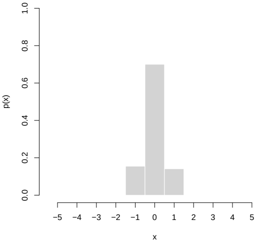


Figure: B

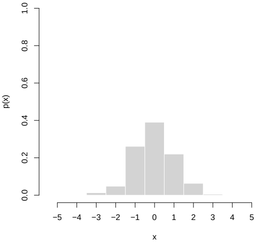


Figure: C

Arrange the figures based on their variance, from smallest to largest.

Answer:

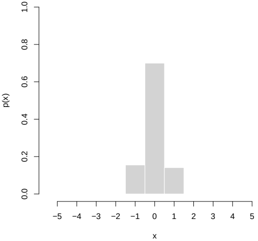


Figure: B

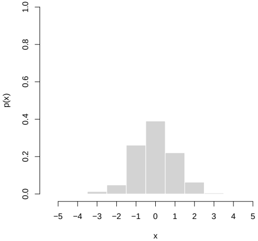


Figure: C

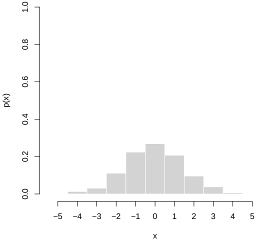
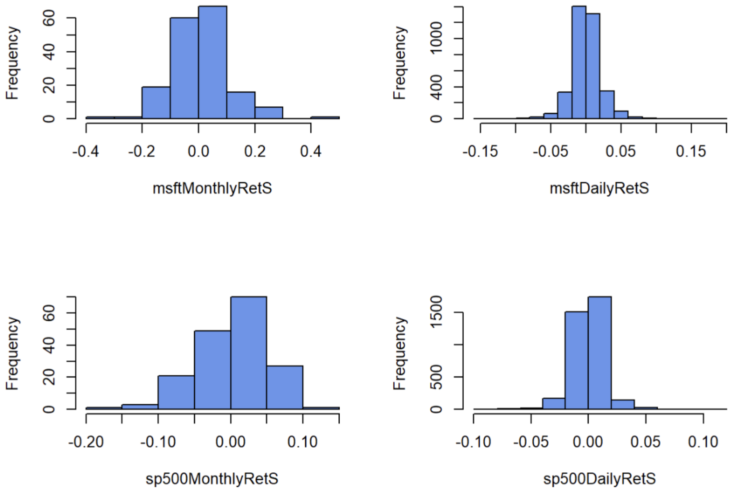


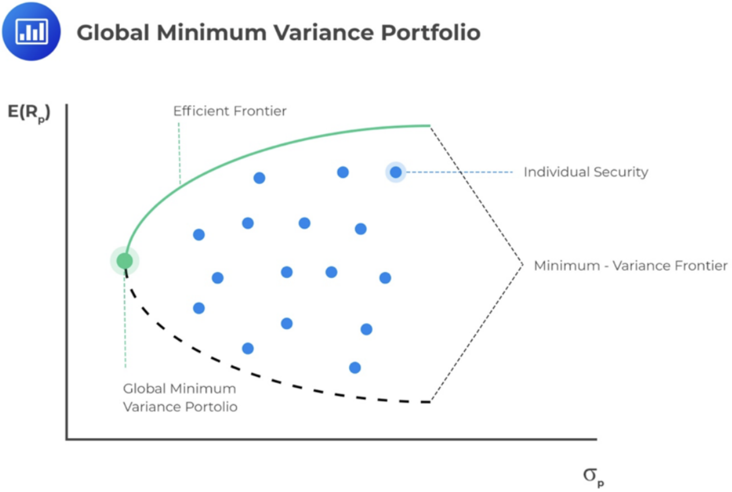
Figure: A

Variance: Application



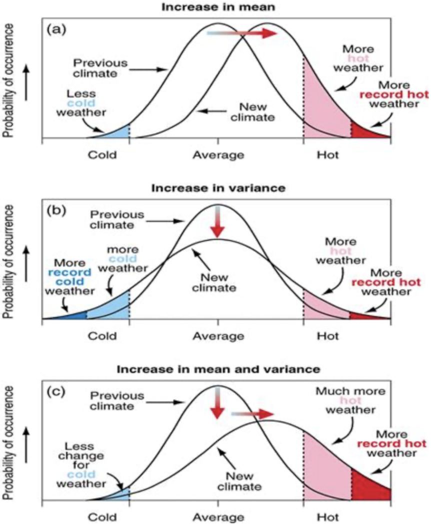
Source: https://bookdown.org/compfinezbook/introcompfinr/Univariate-Descriptive-Statistic.html

Variance: Application



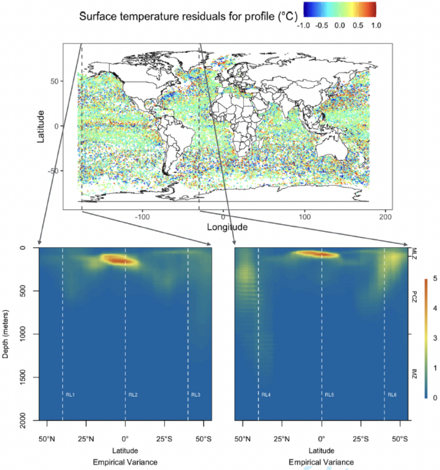
Source: https://analystprep.com/cfa-level-1-exam/portfolio-management/minimum-variance-portfolios/

Variance: Application



Source: https://freakonometrics.hypotheses.org/tag/temperature

Variance: Application



Source: Salvaña, M. L. O. & Jun, M. (2022) 3D bivariate spatial modelling of Argo ocean temperature and salinity profiles. In preparation. https://arxiv.org/pdf/2210.11611.pdf.

Variance: Application

Jayson Tatum and the variance problem

Jayson Tatum is a high variance superstar, but is that a good thing?

By Michael Spooner | Mar 18, 2023, 9:00am EDT | 25 Comments / 25 New

SHARE

THE LINES

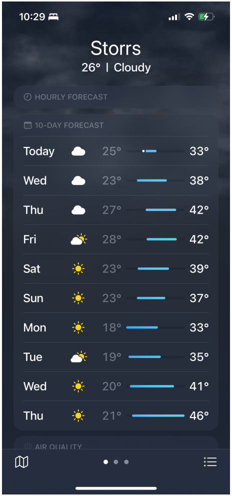
College Football Conference Odds: Using Variance To Pick Better Longshots In Futures & Win Totals

College Football | Featured | News | Odds

Written By Brett Gibbons | Last Updated June 28, 2023

Conference	Variance Score
Pac-12	109.34
Big Ten	103.63
SEC	65.23
ACC	56.01
American	45.00
Mountain West	39.28
Big 12	39.24
Sun Belt	27.64
Conference USA	20.42
MAC	17.41

Variance: Application



Variance of Discrete Random Variables

Theorem 3.6: Variance

Let Y be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$. Then the

$$V(Y) = \sigma^2 = E\{(Y-\mu)^2\} = E(Y^2) - \mu^2.$$

Proof of Theorem 3.6:

$$\begin{aligned}
V(Y) = \sigma^2 &= E\{(Y-\mu)^2\} && \text{definition of variance} \\
&= E\{Y^2 - 2\mu Y + \mu^2\} && \text{squaring} \\
&= E(Y^2) - E(2\mu Y) + E(\mu^2) && \text{expected value of sums (Thm. 3.5)} \\
&= E(Y^2) - E(2\mu Y) + \mu^2 && \text{expected value of constant is itself} \\
&= E(Y^2) - 2\mu E(Y) + \mu^2 && \text{expected value of scaled random variables} \\
&= E(Y^2) - 2\mu(\mu) + \mu^2 && E(Y) = \mu \\
&= E(Y^2) - 2\mu^2 + \mu^2 \\
&= E(Y^2) - \mu^2.
\end{aligned}$$

Variance of Discrete Random Variables

Example:

What is the variance and the standard deviation of random variable X if its probability distribution is the following:

x	$p(x)$
0	$1/5$
1	$1/5$
2	$1/5$
3	$1/5$
4	$1/5$

Variance of Discrete Random Variables

Strategy: Use the formula for computing variance: $V(X) = E\{(X-\mu)^2\} = E(X^2) - \mu^2$.

Example:

What is the variance and the standard deviation of random variable X if its probability distribution is the following:

x	$p(x)$	$x^2 p(x)$
0	$1/5$	$0^2 \times 1/5 = 0$
1	$1/5$	$1^2 \times 1/5 = 1/5$
2	$1/5$	$2^2 \times 1/5 = 4/5$
3	$1/5$	$3^2 \times 1/5 = 9/5$
4	$1/5$	$4^2 \times 1/5 = 16/5$

Answer:

$$V(X) = \sum_x x^2 p(x) - \mu^2 = (0 + 1/5 + 4/5 + 9/5 + 16/5) - 2^2 = 6 - 4 = 2.$$

$$\sigma = \sqrt{V(X)} = \sqrt{2}.$$

Variance of Discrete Random Variables

Example:
I roll a fair die and let X be the resulting number. Find $E(X)$, $V(X)$, and σ.
Answer:

x	$p(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$$\mu = E(X) = \sum_x xp(x) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{21}{6} = \frac{7}{2}.$$

$$\sigma^2 = V(X) = \sum_x x^2 p(x) - \mu^2$$
$$= \left\{1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right)\right\} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

$$\sigma = \sqrt{V(X)} = \sqrt{\frac{35}{12}}.$$

Questions?

Homework Exercises: 3.7, 3.19, 3.27, 3.31, 3.33

Solutions will be discussed this Friday by the TA.