STAT 22000 Lecture Slides
Random Variables

Yibi Huang
Department of Statistics
University of Chicago

Outline

Coverage: Section 2.4 in the text.

- Random Variables
- Expected Value
- Standard Deviation
- Linear combinations of random variables

Random variables

Random Variables

A *random variable* is a numeric quantity whose value depends on the outcome of a random event

- We use a capital letter, like X, to denote a random variable
- The values of a random variable are denoted with a lowercase letter, in this case x, e.g., $P(X = x)$

Example 1. Let X be the number of heads in 2 tosses of a coin. The possible outcomes are $\{HH, HT, TH, TT\}$ and

$$X(HH) = 2, \quad X(HT) = 1, \quad X(TH) = 1, \quad X(TT) = 0$$

Example 2. Let Y be the number of tosses required to get a head. The possible outcomes are $S = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$

$$Y(H) = 1, \; Y(TH) = 2, \; Y(TTH) = 3, \; Y(TTTH) = 4, \ldots$$

Discrete and Continuous Random Variable

There are two types of random variables:

- *Discrete random variables* often take only integer values
  - Example: Number of credit hours, Difference in number of credit hours this term vs last
- *Continuous random variables* take real (decimal) values
  - Example: Cost of books this term, lifetime of a battery

Probability Distribution

A **probability distribution** of a discrete random variable is a list of its possible values and the probabilities that it takes on those values.

Value of X	x_1	x_2	x_3	$\dots$
Probability	p_1	p_2	p_3	$\dots$

A probability distribution must satisfy two conditions:

- $0 \leq p_i \leq 1$ for all i
- $p_1 + p_2 + \cdots = 1$

Example: A Card Game

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw. What's the probability distribution of your earning?

Event	X	$P(X)$
Heart (not ace)	1	12/52
Ace	5	4/52
King of spades	10	1/52
All else	0	35/52
Total		1

Expected Value (Mean)

The *expected value* (or the *mean*) of a random variable is a weighted average of the possible values of the random variable.

$$\mu = E(X) = \sum_i x_i \, P(X = x_i)$$

Example. For the card game described earlier, what is the expected payout in a game?

Event	X	$P(X)$	$X \cdot P(X)$
Heart (not ace)	1	$12/52$	$1 \times (12/52) = 12/52$
Ace	5	$4/52$	$5 \times (4/52) = 20/52$
King of spades	10	$1/52$	$10 \times (1/52) = 10/52$
All else	0	$35/52$	$0 \times (35/52) = 0$
Total			$E(X) = 42/52 \approx 0.81$

Interpretation of the Expected Value

If we play the card game a huge number of time, what's the average earning per game?

Let X_i be the earning in the ith game. Since the earning can only be 0, 1, 5, or 10, the average amount of earning per game is

$$\frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\$0 \cdot \binom{\text{\# of games}}{\text{earning \$0}}}{n} + \frac{\$1 \cdot \binom{\text{\# of games}}{\text{earning \$1}}}{n} + \frac{\$5 \cdot \binom{\text{\# of games}}{\text{earning \$5}}}{n} + \frac{\$10 \cdot \binom{\text{\# of games}}{\text{earning \$10}}}{n}$$

By the law of large numbers, $\frac{\binom{\text{\# of games}}{\text{earning \$0}}}{n} \to P(X = 0)$ as n gets big, and likewise for the rest. So the *long run average earning in a game* is just the *expected value*.

$$0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 5 \cdot P(X = 5) + 10 \cdot P(X = 10) = E(X)$$

Example: A Card Game (Cont'd)

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw. If it charges a certain amount of money each time to play the game, what is the maximum amount you would be willing to pay? Explain your reasoning.

At most $E(X) = 42/52 \approx 0.81$. *If it charges more than that each time to play the game, the gambler will lose money in the long-run.*

Fair game

A *fair game* is defined as a game that costs as much as its expected payout, i.e. expected profit is 0.

Do you think casino games in Vegas cost more or less than their expected payouts?

If those games cost less than their expected payouts, it would mean that the casinos would be losing money on average, and hence they wouldn't be able to pay for all this:

Image by Moyan_Brenn on Flickr http://www.flickr.com/photos/aigle_dore/5951714693.

Variance & Standard Deviation of a Random Variable

We are also often interested in the variability in the values of a random variable.

The *variance* of a random variable X, denoted as σ_X^2 or $V(X)$ is defined as

$$\sigma_X^2 = V(X) = \sum_{i=1}(x_i - \mu)^2 P(X = x_i).$$

The *standard deviation* of a random variable X, denoted as σ_X or $SD(X)$ is simply the square root of the variance:

$$\sigma_X = SD(X) = \sqrt{V(X)}$$

For the previous card game example, how much would you expect the winnings to vary from game to game?

X	$P(X)$	$X\,P(X)$	$P(X)\,(X - E(X))^2$
1	12/52	$1 \times \frac{12}{52} = \frac{12}{52}$	$\frac{12}{52}(1 - \frac{42}{52})^2 \approx 0.00853$
5	4/52	$5 \times \frac{4}{52} = \frac{20}{52}$	$\frac{4}{52}(5 - \frac{42}{52})^2 \approx 1.3505$
10	1/52	$10 \times \frac{1}{52} = \frac{10}{52}$	$\frac{1}{52}(10 - \frac{42}{52})^2 \approx 1.6242$
0	35/52	$0 \times \frac{35}{52} = 0$	$\frac{35}{52}(0 - \frac{42}{52})^2 \approx 0.4416$
Total		$E(X) = \frac{42}{52}$	$V(X) \approx 3.4246$
			$SD(X) \approx \sqrt{3.4246} \approx 1.85$

Interpretation of Standard Deviation

Which of the following is the best interpretation of the standard deviation \$1.85 found in the previous slide?

1. Players of the game receive an average of \$1.85 of payout.
2. About 68% of the players of the card game receive a payout of \$1.85 more or less from the expected payout \$0.81.

   About 90% of the players receive a payout within $0.81 \pm 1.85 = (-1.03, 2.67)$ since out of the possible values: 0,1, 5, 10, only 0 and 1 is in this range, and $P(X = 0) + P(X = 1) = (35/52) + (12/52) = 47/52 \approx 90\%$.
3. The payout received by players of the game differed from the expected payout by about \$1.85 on average. . . . . Answer
4. The payout received by a player of the game differed from the payout received by another player by about \$1.85 on average.

Properties of the Expected Value and the SD

Suppose X is a random variable and c is a fixed number. Then

- $E(X + c) = E(X) + c$,
- $E(cX) = cE(X)$
- $\mathrm{SD}(X + c) = \mathrm{SD}(X)$
- $\mathrm{SD}(cX) = |c|\mathrm{SD}(X)$

Example 1. The payout X of the card game described earlier has expected value $E(X) \approx \$0.81$ and $\mathrm{SD}(X) \approx 1.85$. Suppose it charges \$1 to play the game. What are the expected value and the SD of the net profit $Y = X - 1$ to play the card game once?

$$E(Y) = E(X - 1) = E(X) - 1 \approx 0.81 - 1 = -0.19$$

$$\mathrm{SD}(Y) = \mathrm{SD}(X - 1) = \mathrm{SD}(X) \approx 1.85$$

Properties of Mean and Variance II

Example 2. For the casino that offers the card game, he will collect \$1 from the player and give player the payout according to the outcome. So his net profit from a game is $Z = 1 - X = -Y$. What are the expected value and the SD of the casino's net profit from a game?

$$E(Z) = E(-Y) = -E(Y) \approx -(-0.19) = 0.19$$
$$\mathrm{SD}(Z) = \mathrm{SD}(-Y) = |-1|\mathrm{SD}(Y) \approx 1.85$$

Properties of Mean and Variance II

Suppose X and Y are random variables. Then

- $E(X + Y) = E(X) + E(Y)$ (always valid)
- $V(X + Y) = V(X) + V(Y)$ when X and Y are independent

Example 3. If a player played the card game twice, each time the card drawn is placed back in the deck before the next game, what are the expected value and the SD of his net profit from the 2 games?

- Total net profit $= Y_1 + Y_2$, where $Y_i =$ net profit from the ith game.
- Y_1 and Y_2 are indep as the card drawn is placed back each time, and $E(Y_i) = -0.19$, $V(Y_i) = (\mathrm{SD}(Y_i))^2 = 1.85^2$.

$$E(Y_1 + Y_2) = E(Y_1) + E(Y_2) \approx -0.19 \times 3 = -0.38$$

$$V(Y_1 + Y_2) = V(Y_1) + V(Y_2) \approx 1.85^2 \times 2$$

$$SD(Y_1 + Y_2) = \sqrt{V(Y_1 + Y_2)} \approx \sqrt{1.85^2 \times 2} \approx 2.62$$

Properties of Mean and Variance II

In general, if $X_1, \ldots, X_n$ are random variables, then

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) \qquad \text{always valid}$$
$$V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n) \text{ when } X_i\text{'s are indep.}$$

Example 4. If a player played the card game 200 times and each time the card drawn is placed back in the deck, what are the expected value and the SD of his net profit from the 200 games?

- Total net profit $= Y_1 + \cdots + Y_{200}$, where $Y_i =$ net profit from the ith game.
- Y_1, Y_2, and Y_3 are indep as the card drawn is placed back each time, and $E(Y_i) = -0.19$, $V(Y_i) = (\text{SD}(Y_i))^2 = 1.85^2$.

$$E(Y_1 + \cdots + Y_{200}) = E(Y_1) + \cdots + E(Y_{200}) \approx -0.19 \times 200 \approx -38$$
$$V(Y_1 + \cdots + Y_{200}) = V(Y_1) + \cdots + V(Y_{200}) \approx 1.85^2 \times 200$$
$$SD(Y_1 + \cdots + Y_{200}) = \sqrt{V(Y_1) + \cdots + V(Y_{200})} \approx \sqrt{1.85^2 \times 200} \approx 26.2$$

Example – American Roulette

The game of American roulette involves spinning a wheel with 38 slots: 18 red, 18 black, and 2 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money. Suppose you bet \$1 on red. What's the expected value and standard deviation of your winnings?

Outcome	Red	Black or Green
Profit X	1	-1
$P(X)$	18/38	20/38

$$E(X) = 1 \cdot \frac{18}{38} + (-1) \cdot \frac{20}{38} = -\frac{2}{38} = -\frac{1}{19}$$

$$SD(X) = \sqrt{\left(1 - (-\frac{1}{19})\right)^2 \cdot \frac{18}{38} + \left(-1 - (-\frac{1}{19})\right)^2 \cdot \frac{20}{38}} = \sqrt{\frac{360}{361}} \approx 0.9986$$

Example – American Roulette (Cont'd)

A gambler has \$500 at hand. Consider the following 3 strategies.

(a) betting \$500 on red in a single spin

(b) betting \$300 on red in the first spin, and then \$200 on red in a second spin

(c) betting \$1 on red in 500 different spins

Let X_i be one's earning for every \$1 bet on red in the ith round.

The total winning of the three strategies are

(a) $500X_1$

(b) $300X_1 + 200X_2$

(c) $X_1 + X_2 + \cdots + X_{500}$

Example – American Roulette (Cont'd)

Recall the X_i is one's net profit per dollar bet on red in the ith round. X_i's are independent with mean $E(X_i) = -\$1/19$

For the three gambling strategies, the expected winnings are

(a) $E(500X_1) = 500E(X_1) = 500 \cdot (-\$1/19) = -\$500/19 \approx -\$26.3$

(b) $E(300X_1 + 200X_2) = 300E(X_1) + 200E(X_2) = 300 \cdot \frac{-\$1}{19} + 200 \cdot \frac{-\$1}{19} = -\$500/19$

(c) $E(X_1 + X_2 + \cdots + X_{500}) = E(X_1) + E(X_2) + \cdots + E(X_{500}) = -\$500/19$

So the three strategies have identical expected net profit.

Example – American Roulette (Cont'd)

Recall the X_i is one's earning for betting \$1 on red in the ith round. So X_i's are independent with mean $E(X_i) = -\$1/19$ and variance $V(X_i) = \frac{360}{361}$.

For the three gambling strategies, the variance and standard deviation of the winnings are

(a) $SD(500X_1) = 500SD(X_1) = 500 \cdot \sqrt{\frac{360}{361}} \approx \$499.3$

(b) $V(300X_1 + 200X_2) = 300^2 V(X_1) + 200^2 V(X_2) = (300^2 + 200^2) \cdot \frac{360}{361}$

$SD(300X_1 + 200X_2) = \sqrt{(300^2 + 200^2) \cdot \frac{360}{361}} \approx \$360.05$

(c) $V(X_1 + X_2 + \cdots + X_{500}) = V(X_1) + V(X_2) + \cdots + V(X_{500}) = 500 \cdot \frac{360}{361}$

$SD(X_1 + X_2 + \cdots + X_{500}) = \sqrt{500 \cdot \frac{360}{361}} = \$22.33$

The last strategy has the least variability.