Chapter 4: Random Variables

Random Variable: A random variable is a variable which takes on a value based on the outcome of a random phenomenon or experiment. We usually denote random variables by capital letters, like X, Y, Z, etc. We denote observed or realized variables with lower case letters, x, y, z, etc.

Example 1: Toss 2 fair coins. Let X be the number of heads on the two tosses. Then X is a random variable. Its possible values are 0,1,2. Since there are only a finite number of possible values, X is said to be a discrete random variable.

Example 2: Pick a person at random and let Y be the person's height in inches. Since, in theory, the height could be any real number up to 80 inches or so, Y is said to be a continuous random variable.

Probability Distributions: In Example 1, we could list the possible values of X and their probabilities. This is called the *probability distribution* or *probability model* for X. To get the distribution of X, we first list the outcomes of the experiment:

Outcome	Probability	X
HH	¼	2
HT	¼	1
TH	¼	1
TT	¼	0

Therefore, the probability distribution of X is

Number of heads x	Probability $P(X = x)$
0	1/4
1	1/2
2	1/4

Note that this is still just a model for the distribution of X. The reality of the number of heads on tosses of two coins might deviate from this model (why?), though we might expect the deviation to be small if you did a good job of tossing.

Probability distributions can also be represented by probability functions. We will see these later.

Expected Value: The expected value of a random variable is the mean of the random variable. It's the average value of X we would observe over the long run. We can think of it this way: out of 4 trials or experiments, we would expect to observe 0 heads once, 1 head twice and 2 heads once. The mean of these four results would be

$$\mu = E(X) = \frac{0(1)+1(2)+2(1)}{4} = \frac{4}{4} = 1$$

We can see that

$$E(X) = 0(1/4) + 1(1/2) + 2(1/4) = \sum xP(X = x)$$

We use the Greek letter μ, and not $\bar{y}$, for the mean of a random variable because it is not the mean of a few outcomes from the random process, but the true long-run mean of X. We can think of E(X) as being the *weighted average* of all possible events, where the weights are the probabilities.

Variance: The variance of a random variable is the average (or expected) value of the squared deviations from the mean:

$$\sigma^2 = Var(X) = \sum (x-\mu)^2 P(X = x)$$

For the coin tosses,

$$Var(X) = (0-1)^2(1/4) + (1-1)^2(1/2) + (2-1)^2(1/4) = ¼ + ¼ = ½$$

The standard deviation of X is

$$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{1/2} = 0.707$$

The standard deviation is a measure of the spread of the probability distribution of X. So while we expect, on average, to get one head on two tosses, the actual results can deviate from this expected value. Roughly speaking, we expect the number of heads we observe to be about .71 from the mean (sometimes more, sometimes less).

Example 3: Roulette. A roulette wheel in the U.S. has 18 red slots, 18 black slots and 2 green slots (0 and 00). You can bet \$1 on either red or black. If you win, you get your \$1 back plus an additional \$1. If you lose, you lose your \$1 bet. What are your expected winnings for this bet on one spin of the wheel? Remember that this represents how much you would expect to win on average per bet in the long run.

What are the variance and standard deviation of your winnings for one bet?

Properties of expected values and variances

- E(X ± *c*) = E(X) ± *c*
- Var(X ± *c*) = Var(X)

where *c* is a constant.

Example: Suppose I paid you 50 cents to play roulette once. What are the expected value and standard deviation of your net winnings for one bet?

More properties:

- $E(aX) = aE(X)$
- $Var(aX) = a^2 Var(X)$

Example: Suppose you bet $2 on a single bet on red on the roulette wheel so that you win or lose $2 instead of $1. What are the expected value and standard deviation of your net winnings?

If X and Y are random variables, then

- $E(X + Y) = E(X) + E(Y)$.
- $E(X - Y) = E(X) - E(Y)$

Note: X and Y need not be independent.
Suppose you make $1 bets on red on two successive spins of a roulette wheel. What are your expected total net winnings? How does this compare to making one $2 bet on one spin of the wheel?

If X and Y are <u>independent</u> random variables, then

- $Var(X + Y) = Var(X) + Var(Y)$
- $Var(X - Y) = Var(X) + Var(Y)$ (Note: the variance of the difference of two independent random variables is the <u>sum</u> of the variances)

What are the variance and standard deviation of your total net winnings for two successive $1 bets on the roulette wheel? How does this compare to making one $2 bet on one spin of the wheel?

Suppose you make 100 successive $1 roulette bets. What are the expected value and standard deviation of your total net winnings?

The rules for expected values and variance can be extended to linear combinations of more than two random variables. A linear combination of the random variables $X_1, X_2, \ldots, X_n$ is $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ where $a_1, \ldots, a_n$ are any constants (positive or negative). The result is that

$$E(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n).$$

In addition, if $X_1, X_2, \ldots, X_n$ are independent, then

$$\text{Var}(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1^2\text{Var}(X_1) + a_2^2\text{Var}(X_2) + \ldots + a_n^2\text{Var}(X_n^2)$$

Note: you can see from this formula why Var(X – Y) = Var(X) + Var(Y). Write X – Y as (1)X + (-1)Y and use the last formula above to write Var((1)X + (-1) Y) = 1^2Var(X) + $(-1)^2$Var(Y) = Var(X) + Var(Y).

Example: You will roll three dice: a red one, a green one, and a blue one. You will win, in dollars, 3 times the number on the red die, plus 2 times the number on the green die, minus the number on the blue die. What are the expected value and standard deviation of your winnings? (Hint: first calculate the expected value and variance of the outcomes for a single die roll: 3.5, 2.917, 1.708).

What would be a fair price to pay to play this game?

IMPORTANT: Variances of independent random variables can be added, but NOT standard deviations! If you want the standard deviation of the sum of independent random variables, you must first figure out the variance of the sum and then take the square root. You can't sum the standard deviations.