












## 1. RANDOM VARIABLES & PROBABILITY DISTRIBUTIONS (4.1 & handout 3.1-3.3)

Definition – RANDOM VARIABLE usually denoted by a capital letter, usually  $X$ , and it is the NUMERICAL outcome of a random process. Basically a random variable is the assignment of a number to each basic outcome in a random process.

Review examples of random processes: coin toss, casino game of craps, sales made by a telemarketer in a day, children born to a family, product failure, a commute to work

Review basic outcomes: Heads/Tails, rolling a pair of 2 spots, 0 to ? sales, Boy-Girl-Girl etc.

Example: The game of Craps is a good example, here is the language at a craps table, and then the conversion of the game to a random variable:

Craps Outcomes (Outcomes of a random process)	As understood by the gamblers at the Craps Table	Probability Distribution	
		A Random Variable $X$ takes values of	$P(X=x_i)$ or $P(x_i)$
	"Craps", "two aces," "snake eyes," "twice in the rice,"	2	1/36
	"Craps," "ace-deuce," "ace caught a deuce," "winner on the dark side,"	3	2/36
	"Little Joe," (2-2) "ace trey, the country way." (3-1)	4	3/36
	"After five, the field's alive," "thirty-two juice roll", "fiver, fiver, racetrack driver,"	5	4/36
	"Big Red", "catch'em in the corner," (3-3) "like a blue chip stock," "pair-o-treys, waiter's roll," (3-3)	6	5/36
	"five two, you're all through," "six ace, end of the race," ""seven's a bruiser, the front line's a loser,"	7	6/36
	"A square pair", "like mom and dad," "Ozzie and Harriet," "the windows,"	8	5/36
	"Center field," "center of the garden," "ocean liner niner,"	9	4/36
	"Puppy paws," "pair-a-roses," "the big one on the end."	10	3/36
	"Yo leven," "yo levine the dancing queen," "six five, no jive."	11	2/36
	"Craps," "boxcars," "all the spots we got," "triple dipple, in the lucky ducky,"	12	1/36

A random variable is just a function or a rule that takes each basic outcome in the sample space ( $S$ ) and assigns it a real number. A random variable is only random in the sense that the actual value the variable takes is not determined until an action (a trial) has been completed (the dice rolled). Note that the value can change from trial to trial. A random variable makes random processes easier to work with because it is a numeric description of a non-numeric process.

## 2. Discrete Random Variables (4.1 and handout 3.1-3.3)

A discrete random variable is countable and finite. Recall some variables are naturally discrete such as you can't roll a 3.1 on a die or you can't have 351.7 employees at a firm. Discrete variables jump from one whole integer to the next. So one can "list" the outcomes of the known possible values a random variable  $X$  can take. And one can list the probabilities (or percentage of occurrences, or chance) of each possible outcome (value).

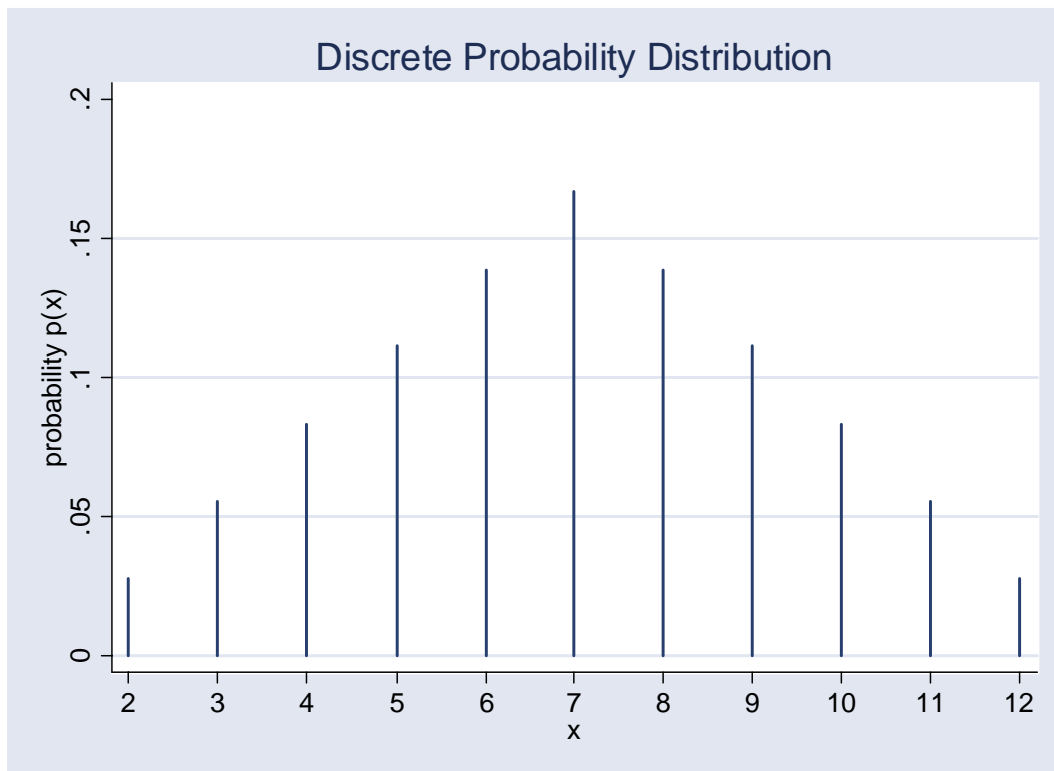
Discrete Random variables have probability distributions (or probability functions AKA probability mass functions). Probability distributions are a way of organizing outcomes and probabilities and representing them in a "shorthand" fashion using math notation or a table or a graphic. More formally, we write  $f(x_i) = P(X = x_i)$

There are 2 properties:

- the probabilities must be greater than or equal to zero i.e.  $0 \leq P(X = x_i) \leq 1$
- the sum of the probabilities of the  $k$  different outcomes must be exactly 1.

$$\sum_{i=1}^k P(X = x_i) = 1$$

The customer ratings, the number of car accidents in day, the number of heads in 10 tosses, the number of times a "7" is rolled in 25 rolls of a pair of dice are all discrete examples.



This is a graph of the probability distribution for the game of craps. Craps takes on a limited number of distinct values and is called discrete random variable.

A cumulative probability distribution (or a cumulative distribution function) is the probability that  $X \leq x_i$ , this gives the probability, for example, of "seeing a 6 or less" or "4 or less".

#### 4. Random Variables have means & variances too...

##### A. THE EXPECTED VALUE or THE POPULATION MEAN

The mean of a list of numbers like 1, 2, 3, 4, 5 or like 1,1,2,2,3,3,4,4,5,5, gives every value in the list equal weight (.20 for each). The mean of a much longer list of numbers could be tabled, like this (look familiar?):

thumbs up	Freq.	Percent	Cum.
0	1,000	10.00	10.00
1	3,000	30.00	40.00
2	3,500	35.00	75.00
3	2,500	25.00	100.00
Total	10,000	100.00	

We can calculate the mean using:  
 $(0 \cdot .1) + (1 \cdot .3) + (2 \cdot .35) + (3 \cdot .25)$   
 $= 1.75$

The mean of a random variable is like the list of numbers but more like the tabled example. The random variable assigns probabilities to the outcomes (values the variable can take) and these outcomes do not need to occur with equal frequency. Given the frequencies, we can find the percentage and recall calculating the mean of a variable from its relative frequency distribution ( $f/n$ ) (p 34 of the textbook). This extends to random variables, we can find the mean of a random variable:

$$E(X) \equiv \mu \equiv \sum_{\text{all } X} x_i p(x_i)$$

Generally, the mean of random variables are written  $\mu_x$  pronounced "mu sub x" to represent the mean of any random variable  $X$ . What would the symbol  $\mu_y$  mean to you? Another way it might be written is  $E(X)$  or the "expected value" or the "expectation". It is also known as the population mean because probabilities are understood to be the limit of the relative frequencies.

The expected value (or expectation or mean) of a random variable is the sum of the probability of each possible outcome of the random process (experiment) multiplied by its value. Thus, it represents the average value one "expects" to see if the process is repeated many times with stable (identical) probabilities. Note that the actual value itself may not be expected in the general sense, it may be unlikely or even impossible. For example, the expected value of the roll of a single die is 3.5, but you cannot actually roll a 3.5.

Example: Suppose it is known that a salesman typically makes 3 phone calls a year to each home in his region and that his chance of making 3 sales (one for each call) is 5/100 (or 5% or .05), 2 sales is 15/100 (or 15% or .15), 1 sale is 30/100 (or 30% or .30) and no sale is 50/100 (or 50% or .50). If he earns \$200 per sale, how much can he EXPECT to earn in a year (or what would the MEAN of this random variable be)?

We know that in 3 phone calls, he faces 4 possible outcomes, but what is "most likely to happen?" or "what is EXPECTED to happen?"

To find the mean of this random variable  $X$ , multiply each possible outcome by its probability and add up the products:

The formula expanded (see page 114)  $\mu_x = x_1 p_1 + x_2 p_2 + x_3 p_3 + \dots + x_n p_n$

So for the example above,  $\mu_x$  is  $(0 \cdot .50) + (1 \cdot .30) + (2 \cdot .15) + (3 \cdot .05) = .75$  sales

Note several things

1. The outcomes are "listable", there are 4 outcomes (0, 1, 2, 3)

2. The probabilities are values between 0 and 1, their sum is 1.0
  - a.  $.50 + .30 + .15 + .05 = 1.0$
  - b.  $.50$  is 50% or 50/100,  $.30$  is 30% or 30/100 etc.
3. The mean is  $.75$  sales, this suggests that if he makes 3 phone calls, he can expect to make less than one sale. In other words, if he make 300 phone calls, he would expect to make 75 sales (that's  $.75 * 100$  because  $3 * 100 = 300$ ). If the company is paying him \$200 per sale, he can expect to make \$15,000 in a year.

Question: What's his median sales?

## B. THE (POPULATION) VARIANCE OF A DISCRETE RANDOM VARIABLE

Recall that when we describe a variable, we talk about "center" (or what is typical) and "spread" (dispersion). Variance is a measure of dispersion based on the pattern of outcomes around the mean and gives us a sense of what is "typical" in terms of deviations. We are often interested in variance because we know that not all values of a variable are the same.

Just like the mean of a discrete random variable weights each outcome by their respective probabilities, the squared deviations are weighted by their probabilities. For a discrete random variable  $X$ , the variance is:

$$\sigma^2 = \sum_{all X} (x_i - \mu)^2 p(x_i)$$

The standard deviation is the square root of the variance.

Example: A salesman makes 3 calls per household over the course of one year, how many sales can he expect to make?

$f(x_i)$	0	1	2	3
$p(x_i)$	.50	.30	.15	.05

The expected value or mean was  $.75$  sales (from the previous section)

The variance then is:

$$(0 - .75)^2 * .50 + (1 - .75)^2 * .30 + (2 - .75)^2 * .15 + (3 - .75)^2 * .05 = .7875 \text{ sales squared}$$

The standard deviation is a little easier to interpret

Square root of  $.7875 = .8874$  so  $.8874$  sales

Interpreting these two will be a little easier in coming days. For now, interpret them like the mean and standard deviation you are familiar with, just keep in mind that these are the means and standard deviations of random variables.

Note: your textbook gives an alternate formula for the variance

$$\sigma^2 = \sum_{all X} (x_i)^2 p(x_i) - \mu^2$$