L5 Monday, Sept. 7, 2020

HW: page 90: #'s 2, 3 and 6
page 97: #'s 12, 13, 14

Chapter 3: Discrete Random Variables

Section 3.1; Basic Definition
Section 3.2; Probability Distribution for a Discrete Random Variable
Section 3.3; The Expected Value of a Random Variable

Section 2.12; Random Sampling

In a statistical experiment, the observations make up a *sample* which were selected from a *population*. The sample is used to make inferences about the characteristics of the population.

Def: Let N be the number of elements in the population and n be the number of elements in the sample, $N \geq n$. If the sample is collected in such a way that each of the ${}_N C_n$ samples has an equal probability of being selected, the sampling is said to be *random*, and the result is said to be a *random sample.*

Note: It is hard to come up with a random sample. You can make a career out of just making random samples. There are tables of random digits used to generate random numbers (ex. Table 12, Appendix 3). However, these tables are usually computer generated from some sort of algorithm. These "random number" are considered pseudo-random but may be the best way at the time.

Sections 3.1; Basic Definition
Def 3.1

A random variable Y is said to be discrete if it can assume only a countable number (finite or infinite) of values.

Section 3.2; Probability Distribution for a Discrete Random Variable

Notation:

We use an uppercase letter, such as Y or X, to denote a (unknown) random variable. Use lower case letters, such as y or x, for known or particular values.

Ex.

Roll a die, define P(Y) be the probability that a certain value is shown on the die. So $P(Y=2)=\frac{1}{6}$. In general $P(Y=y)=\frac{1}{6}$ or written as $p(y)=\frac{1}{6}$

The expression ($Y = y$) can be read as the set of all points in S assigned the value y by the random variable Y.

Def 3.3:

The *probability distribution* for a discrete variable, Y, can be represented by either a formula, a table, or a graph that provides $p(y) = P(Y = y)$ for all y.
(This is a list of all possible outcomes along with their probabilities. It is how the probabilities are distributed about the outcomes.)

3 ways

Ex.

A supervisor in a plant has three men and three women working for him. He wants to choose two workers for a special job. He does not want to show any bias in the selection so he select two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y.

Formula $p(i) = \dfrac{\binom{3}{i}\binom{3}{2-i}}{\binom{6}{2}}$

$p(0) = \dfrac{\binom{3}{0}\binom{3}{2}}{\binom{6}{2}}$

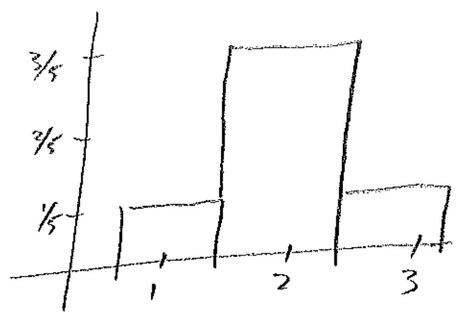
$p(1) = \dfrac{\binom{3}{1}\binom{3}{1}}{\binom{6}{2}}$

$p(2) = \dfrac{\binom{3}{2}\binom{3}{0}}{\binom{6}{2}}$

Table

y	$p(y)$
0	1/5
1	3/5
2	1/5

Graph



Ex.

Roll a pair of fair die. Define a probability distribution for the sum of the face values of the dice.

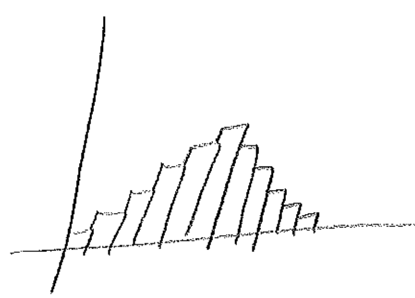
+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table

y	2	3	4	5	6	7	8	9	10	11	12
$p(y)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Formula

$$p(y) = \begin{cases} \dfrac{y-1}{36} & \text{if } y \le 7 \\ \dfrac{13-y}{36} & \text{if } y > 7 \end{cases}$$



Theorem 3.1

For any discrete distribution, the following must be true:

1. $0 \le p(y) \le 1$ for all y.
2. $\sum_y p(y) = 1$

Section 3.3; The Expected Value of a Random Variable

Def 3.4:

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

$$E(Y) = \sum_y y \cdot p(y)$$

In this case the expected value of Y is also the mean of the population i.e.

$\mu = E(Y) = \sum_y y \cdot p(y)$ (use all $y \in Y$)

Expected value can be extended to include functions. Suppose Y is a discrete random variable with the probability function $p(y)$. Define $g(Y)$ as a real valued function of Y. Then the expected value of $g(Y)$ is given by

$$E[g(Y)] = \sum_y g(y) \cdot p(y)$$

Ex.

Below is the probability distribution for some variable Y.

y	$p(y)$
0	1/8
1	1/4
2	3/8
3	1/4

Find $\mu = E(Y)$ and $E(Y^2)$

$E(Y) = 0(\frac{1}{8}) + 1(\frac{1}{4}) + 2(\frac{3}{8}) + 3(\frac{1}{4})$
$= 0 + \frac{1}{4} + \frac{3}{4} + \frac{3}{4}$
$= \frac{7}{4}$ $(1\frac{3}{4})$

$E(Y^2) = 0^2 \cdot \frac{1}{8} + 1^2(\frac{1}{4}) + 2^2(\frac{3}{8}) + 3^2(\frac{1}{4})$
$= 0 + \frac{1}{4} + \frac{6}{4} + \frac{9}{4}$
$= \frac{16}{4} = 4$

Ex.

Here is a game. A person draws a single card from an ordinary 52-card playing deck. A person is wins \$15 for drawing a jack or a queen and wins \$5 for drawing a king or an ace. A person who draws any other card losses \$4. What is the expected payout for this game?

payout	Y
+15	J, Q
5	K A
−4	other

$E(\text{payout}) = 15(\frac{2}{13}) + 5(\frac{2}{13}) - 4(\frac{9}{13})$
$= \frac{30 + 10 - 36}{13} = \frac{4}{13}$ $(\approx \$.31)$

(you want to play this alot.

Def 3.5: *(note: this is not the easiest way of finding variance by hand.)*

Let Y be a discrete random variable with mean $\mu = E(Y)$, the variance of a random variable Y is defined to be the expected value of $(Y-\mu)^2$. That is,

$$V(Y) = E[(Y-\mu)^2] = \sum_y (y-\mu)^2 \cdot p(y)$$

In this case the variance Y is the variance of the population, i.e. $\sigma^2 = V(Y)$, with the standard deviation $\sigma = \sqrt{\sigma^2} = \sqrt{V(Y)}$.

(Notice: variance is the expected value of distance from the mean square of the variable.)

Three little properties of expected value:

i. $E(c) = c$

$$E(c) = \sum_y c \cdot p(y) = c \underbrace{\sum p(y)}_{=1} = c$$

ii. $E(c \cdot g(Y)) = c \cdot E[g(Y)]$

$$E(c \cdot g(Y)) = \sum_Y c\, g(y_i) p(y_i) = c \sum_Y g(y_i) p(y_i) = c\, E(g(Y))$$

iii. $E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = E[(g_1(Y)] + \cdots + E[g_k(Y)]$

$$E[g_1(Y) + \cdots + g_k(Y)] = \sum_Y (g_1(y_i) + \cdots + g_k(y_i)]\, p(y_i)$$

$$= \sum_Y g_1(y_i) p(y_i + g_2(y_i) p(y_i) + \cdots + g_k(y_i) p(y_i))$$

$$= \sum g_1(y_i) p(y_i) + \cdots + \sum g_k(y_i)\, p(y_i)$$

$$= E(g_1(Y)) + \cdots + E[g_k(Y)]$$

Theorem 3.6 (a simpler way of finding the variance.)

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

Then

$$\sigma^2 = V(Y) = E[(y-\mu)^2] = E(Y^2) - \mu^2 .$$

Proof:

$$V(Y) = E((Y-\mu)^2) \qquad \text{mult out}$$

$$= E(Y^2 - 2\mu Y + \mu^2)$$

$$= E(Y^2) - E(2\mu Y) + E(\mu^2) \qquad \text{Now } \mu \text{ is a constant}$$

$$= E(Y^2) - 2\mu \underbrace{E(Y)}_{\mu} + \mu^2$$

$$= E(Y^2) - 2\mu^2 + \mu^2$$

$$= E(Y^2) - \mu^2$$