Discrete Random Variables I (3.1 - 3.3)

1. Random Variables

   (a) **Random Variable:** A variable whose value is a chance or random event associated with the outcome of an experiment. A random variable is any rule that associates a number with each outcome in the sample space, S, of an experiment. There are two types of random variables, discrete and continuous.

   (b) **Discrete Random Variable:** A random variable whose possible values constitute either a finite set or a countably infinite sequence is called a discrete random variable.

   (c) **Continuous Random Variable:** A random variable whose values are not countable, but consist of an entire interval on a number line is called a continuous random variable.

   (d) Notation: $X(s) = x$ Here x is the numerical value of random variable X associated with outcome s.

2. Discrete Random Variables

   (a) Bernoulli Random Variable: A discrete random variable whose only possible values are 0 and 1. This is the simplest of all discrete random variables. The sample space required is:

$$S = \{S, F\} \qquad X = \{0, 1\}$$

   (b) Defining Discrete Random Variables from Sample Space Outcomes

      i. Example: Each time an electrical switch is tested, the trial is either a success (S) or failure (F). Suppose switches are tested repeatedly until a success occurs on three consecutive tials. Let Y denote the number of trials necessary to achieve this.

         A. List all outcomes corresponding to the 5 smallest possible values of Y, and state which values are associated with each one.

$Y \equiv \{\text{number of trials needed to get SSS}\}$

Y values	Possible Outcomes
$Y = 3$	SSS
$Y = 4$	FSSS
$Y = 5$	FFSSS SFSSS
$Y = 6$	FFFSSS SFFSSS FSFSSS SFFSSS SSFSSS
$Y = 7$	FFFFSSS SFFFSSS SSFFSSS SFSFSSS FFSFSSS
	FSFFSSS FSSFSSS

         B. Five smallest possible values of Y : $Y_{S5} \equiv \{3, 4, 5, 6, 7\}$

(c) Example: The number of gas pumps in use at both a 6-pump station and a 4-pump station is to be determined.

	0	1	2	3	4	5	6
0	(0,0)	(0,1)	(0,2)	(0,3)	(0,4)	(0,5)	(0,6)
1	(1,0)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,0)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,0)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,0)	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)

List the possible values for each of the following random variables.

i. $T = \{$ the total number of pumps in use at a particular time$\}$
for outcome $s = (2, 3)$, $T(2, 3) = 2 + 3 = 5$

$$T = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

ii. $X = \{$ the difference between the number of pumps in use at stations 1(6-pumps) and 2(4-pumps)$\}$.
for outcome $s = (2, 3)$, $X(2, 3) = 3 - 2 = 1$

$$X = \{-4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6\}$$

iii. $U = \{$ the maximum number of pumps in use at either station$\}$
for outcome $s = (2, 3)$, $U(2, 3) = max(2, 3) = 3$

$$U = \{0, 1, 2, 3, 4, 5, 6\}$$

iv. $Z = \{$ number of stations having exactly 2 pumps in use at a particular time $\}$
for outcome $s = (2, 3)$, $Z(2, 3) = 1$

$$Z = \{0, 1, 2\}$$

3. Probability Distributions for Discrete Random Variables

(a) For every possible value, x, that random variable X can take on, the Probability Mass Function (pmf) specifies the probability of observing that value when the experiment is performed.

(b) Let X be a discrete random variable that can take on the following values:

$$X : \quad \{-2, 0, 1, 3, 4, 7, 8\}$$

(c) The Probability Mass Function (pmf) assigns a probability to each possible value of X. These probabilities are always positive and individually range in value between a minimum of 0 and a maximum of 1.

(d) The sum of these probabilities, when summed over all possible values X can take on, is equal to 1.

x	-2	0	1	3	4	7	8
$P(X = x)$	.13	.15	.17	.20	.15	.11	.09

- $P(X = 1) = .17$
- $P(X < 1) = .13 + .15 = .28$
- $P(X \leq 1) = .13 + .15 + .17 = .45$
- $P(0 \leq X \leq 4) = .15 + .17 + .20 + .15 = .67$
- $P(0 < X \leq 4) = .17 + .20 + .15 = .52$

(e) Example: Consider a group of 5 potential blood donors (A,B,C,D,and E) of whom only A and B have type O+ blood. Five blood samples, one from each donor, will be typed in random order until an individual with O+ blood. If random variable Y is identified as:
$Y \equiv \{\text{Number of typings needed to identify 0+ individual}\}$
Determine the pmf for random variable Y

Donor	Blood Type
A	0+
B	0+
C	not 0+
D	not 0+
E	not 0+

Let $O \equiv$ 0+ and $N \equiv$ $not0+$
$P(Y = 1) = P(O) = \frac{2}{5} = 0.4$
$P(Y = 2) = P(NO) = \frac{P_{1,2}P_{1,3}}{P_{2,5}} = \frac{2(3)}{20} = 0.3$
$P(Y = 3) = P(NNO) = \frac{P_{1,2}P_{2,3}}{P_{3,5}} = \frac{2(6)}{60} = 0.2$
$P(Y = 4) = P(NNNO) = \frac{P_{1,2} \cdot P_{3,3}}{P_{4,5}} = \frac{2 \cdot 6}{120} = 0.1$

recall that: $P_{k,n} = \frac{n!}{(n-k)!}$; $P_{2,5} = \frac{5 \cdot 4 \cdot 3!}{3!} = 20$
$P_{1,2} = \frac{2}{1} = 2$; $P_{1,3} = \frac{3(2)}{2} = 3$
$P_{2,3} = \frac{3(2)}{1} = 6$; $P_{3,5} = \frac{5 \cdot 4 \cdot 3 \cdot 2}{2} = 60$
$P_{3,3} = \frac{3 \cdot 2}{1} = 6$; $P_{4,5} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1} = 120$

The probability mass function (pmf) for random variable Y is as shown below:

y	1	2	3	4
$P(Y = y)$	0.4	0.3	0.2	0.1

Note that, for now, we will use $p(y) \equiv P(Y = y)$

(f) The **Cumulative Distribution Fnction (cdf)**, denoted $F(x)$, is the probability that the observed value of X will be at most x. $F(x)$ ranges in value from a minimum of 0 to a maximum of 1. Consider the following pmf for random variable X.

x	-2	0	1	3	4	7	8
$P(X = x)$	.13	.15	.17	.20	.15	.11	.09

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

- $F(-2) = P(X \leq -2) = p(-2) = .13$
- $F(0) = P(X \leq 0) = p(-2) + p(0) = .13 + .15 = .28$
- $F(1) = P(X \leq 1) = p(-2) + p(0) + p(1) = .13 + .15 + .17 = .45$
- $F(3) = P(X \leq 3) = p(-2) + p(0) + p(1) + p(3) = .45 + .20 = .65$
- $F(4) = P(X \leq 4) = p(-2) + p(0) + p(1) + p(3) + p(4) = .45 + .20 + .15 = .80$
- $F(7) = P(X \leq 7) = p(-2)+p(0)+p(1)+p(3)+p(4)+p(7) = .45+.20+.15+.11 = .91$
- $F(8) = P(X \leq 8) = p(-2) + p(0) + p(1) + p(3) + p(4) + p(7) + p(8) = .80 + .11 + .09 = 1.00$
- Tabulated cdf values for X are:

x	-2	0	1	3	4	7	8
$F(x)$	.13	.28	.45	.65	.80	.91	1.00

(g) The Cumulative Distribution Function(cdf) can be used to determine a wide variety of probabilities.

(h) In general, $P(a \leq x \leq b) = F(b) - F(a-)$
where $a- \equiv$ Largest possible X value that is less than a.

(i) When X values are integers

$$P(a \leq x \leq b) = F(b) - F(a - 1)$$

(j) For example, using the cdf above with $a = 1$ and $b = 7$

$$P(1 \leq x \leq 7) = F(7) - F(0) = .65 - .13 = .52$$

here $a - \quad = 0$. Also,

(k) $P(0 \leq X \leq 4) = F(4) - F(-2) = .80 - .13 = .67$

(l) $P(X \geq 4) = 1 - F(3) = 1 - (.15 + .11 + .09) = .65$

(m) pmf values can also be computed from the difference of cdf values:

$$F(3) = p(-2) + p(0) + p(1) + p(3)$$
$$F(1) = p(-2) + p(0) + p(1)$$

$$F(3) - F(1) = p(3)$$

4. Example: Some parts of California are particularly earthquake prone. Suppose that in one such area, 30% of all homeowners are insured against earthquake damage. Four homeowners are to be selected at random; let X be the number among the four homeowners who have earthquake insurance(EQI).

   (a) Determine the probability mass function (pmf) for X.

   - Let $S \equiv$ {Owner has EQI}; $F \equiv$ {Owner does not have EQI}
   - $P(S) = 0.30$; $P(F) = 0.70$
   - $X \equiv$ {No. S's out of four}
   - $X \equiv \{0, 1, 2, 3, 4\}$

X	Outcomes	$p(x)$
0	FFFF	$0.7(0.7)(0.7)(0.7) = 0.2401$
1	SFFF, FSFF, FFSF, FFFS	$4[0.3(0.7)(0.7)(0.7)] = 0.4116$
2	FFSS, FSFS, SFFS, FSSF, SSFF, SFSF	$6[(0.3)^2 \cdot (0.7)^2] = 0.2646$
3	FSSS, SFSS, SSFS, SSSF	$4[(0.3)^3 \cdot (0.7)] = 0.0756$
4	SSSS	$0.3^4 = 0.0081$

   - The pmf for X is:

x	0	1	2	3	4
$p(x)$	.2401	.4116	.2646	.0756	.0081

5. Expected Values of Discrete Random Variables

   (a) The **Expected Value of** X, denoted $E(X)$, is a measure of the center (or mean value) of the distribution of X values. $E(X)$ generally not equal one of the values of X, but is a value calculated from them.

$$E(X) = \mu_X = \sum_x x \cdot p(x)$$

x	-2	0	1	3	4	7	8
$P(X = x)$	.13	.15	.17	.20	.15	.11	.09

$$E(X) = \mu_X = \sum_x x \cdot p(x)$$

- $E(X) = -2(.13) + 0(.15) + 1(.17) + 3(.20) + 4(.15) + 7(.11) + 8(.09) = 2.60$

(b) Example: Determine the expected number of credit cards a Saint Mary's student will possess based on the pmf data below.

$$X \equiv \{ \text{ Number of Credit Cards possessed by Student} \}$$

x	0	1	2	3	4	5	6
$P(X = x)$	.08	.28	.38	.16	.06	.03	.01

$$E(X) = \mu_X = \sum_x x \cdot p(x)$$

- $E(X) = 0(.08) + 1(.28) + 2(.38) + 3(.16) + 4(.06) + 5(.03) + 6(.01) = 1.96$

6. The **Expected Value of a Function,** $h(X)$, denoted $E[h(X)]$, can be computed for any function $h(X)$ from the pmf of X values.

$$E[h(X)] = \mu_{h(X)} = \sum_x h(x) \cdot p(x)$$

7. Example: Let X be a discrete random variable with pmf:

x	1	2	3	4
$P(X = x)$	.2	.4	.3	.1

(a) Let $h(x)$ be the following function of random variable x:

$$h(x) = 5 + 2x + 3x^2$$

(b) Determine the expected value of $h(x)$, $E[h(x)]$:

$$\begin{array}{ll} X = 1: & h(1) = 5 + 2(1) + 3(1)^2 = 10 \\ X = 2: & h(2) = 5 + 2(2) + 3(2)^2 = 21 \\ X = 3: & h(3) = 5 + 2(3) + 3(3)^2 = 38 \\ X = 4: & h(4) = 5 + 2(4) + 3(4)^2 = 61 \end{array}$$

x	1	2	3	4
$h(x)$	10	21	38	61
$P(X = x)$	.2	.4	.3	.1

- $E[(h(X)] = \sum h(x) \cdot p(x) = 10(.2) + 21(.4) + 38(.3) + 61(.1) = 27.9$

8. The **Variance of** X, denoted $Var(X)$, is a measure of the variability (spread, dispersion) in the distribution of X. The variance is defined as follows:

$$Var(X) = \sigma_X^2 = \sum_{i=1}^{n}[x_i - E(X)]^2 p(x_i)$$

- $Var(X) = (-2-2.6)^2(.13) + (0-2.6)^2(.15) + (1-2.6)^2(.17) + (3-2.6)^2(.20) + (4-2.6)^2(.15) + (7-2.6)^2(.11) + (8-2.6)^2(.09) = 9.28$

- Shortcut Formula for the Variance

$$Var(X) = \sigma_X^2 = E(X^2) - [E(X)]^2 = [\sum_{i=1}^{n} x_i^2 p(x_i)] - [E(X)]^2$$

- $Var(X) = [(-2)^2(.13) + (0)^2(.15) + (1)^2(.17) + (3)^2(.20) + (4)^2(.15)$ $+(7)^2(.11) + (8)^2(.09)] - (2.60)^2 = 9.28$

- The **Standard Deviation of** X, denoted $SD(X)$, is equal to the positive square root of the variance of X, so $SD(X)$ always has a positive value.

$$SD(X) = \sigma_X = \sqrt{\sum_{i=1}^{n}[x_i - E(X)]^2 p(x_i)}$$

  - $SD(X) = \sqrt{Var(X)} = \sqrt{9.28} = 3.046$

- Example: An auto service facility that specializes in engine tuneups has been in business at the same location for the last 10 years. Company records over this period show that 50% of the cars coming in for tuneups had 4 cylinder engines, 30% had 6 cylinder engines, and 20% had 8 cylinder engines. No cars coming in for a tuneup had other than a 4, 6, or 8 cylinder engine.

  (a) Define random variable X as the number of cylinders in the engine of the next car coming in for a tuneup. So X is a discrete random variable that can only take on values of 4, 6, or 8.

$$X: \quad \{4, 6, 8\}$$

  (b) The Probability Mass Function (pmf) is

x	4	6	8
$P(X = x)$	.50	.30	.20

  (c) The Cumulative Distribution Function (cdf) is

x	4	6	8
$F(x)$	.50	.80	1.0

  (d) The Expected Value, $E(X)$, is

$$E(X) = \mu_X = \sum_x x * p(x) = 4(.5) + 6(.3) + 8(.2) = 5.4$$

(e) The Variance, $Var(X)$, is

$$Var(X) = \sum_{i=1}^{n}[x_i - E(X)]^2 p(x_i) = .5(4-5.4)^2 + .3(6-5.4)^2 + .2(8-5.4)^2 = 2.44$$

(f) The Standard Deviation, $SD(X)$, is

$$SD(X) = \sqrt{Var(X)} = \sqrt{2.44} = 1.562$$