Chapter 5_Part_A Probability Distributions

Ram C Kafle, Ph.D.
Assistant Professor of Statistics
Department of Mathematics and Statistics
Sam Houston State University

Random Variables

A quantitative variable x is a **random variable** if the value that x takes on in a given experiment or observation is a chance or random outcome.

A **discrete random variable** can take on only a finite number of values or a countable number of values.

A **continuous random variable** can take on any of the countless number of values in a line interval.

- The distinction between discrete and continuous random variables is important because of the different mathematical techniques associated with the two kinds of random variables.

Random Variables

A **random variable** X associates a numerical value with each outcome of an experiment.

- Example: Number of heads as random Variable

Suppose we are interested in the number of heads (X) observed when we toss two coins. Then

1. Write down the Sample Space

The Sample Space is: S = {HH, HT, TH, TT}

2. List all the numerical values of the random variable X.

$X = \{2, 1, 1, 0\}$.

Number of heads as random variable in three tosses of a coin

- Let X be the number of heads obtained in three tosses of a coin.
- First table below list the numerical values of X.
- Second table on right, identify the events that correspond to the distinct values of X

Outcome	Value of X
HHH	3
HHT	2
HTH	2
HTT	1
THH	2
THT	1
TTH	1
TTT	0

Numerical Value of X as an Event	Composition of the Event
$[X = 0]$ =	{TTT}
$[X = 1]$ =	{HTT, THT, TTH}
$[X = 2]$ =	{HHT, HTH, THH}
$[X = 3]$ =	{HHH}

Probability Distribution of A Discrete

The **probability distribution** or, simply the **distribution,** of a discrete random variable X is a list of the distinct numerical values of X along with their associated probabilities.

Often, a formula can be used in place of a detailed list.

Note:

- The probability of each outcome (random variable) should be greater than or equal to zero but less than or equal to 1.
  (That is, $0 \leq P(x) \leq 1, for\ all\ x$).
- The sum of the probabilities should add up to one.

(That is, $\sum P(x) = 1, for\ all\ x$).

For example, lets consider the number of heads observed when we toss two coins. Construct a probability distribution for this experiment.

X	**0**	**1**	**2**
P(X)	1/4	1/2	1/4

The Probability Distribution for Tossing a fair coin: If X represents the number of heads, then find the probability distribution of X.

TABLE 1 The Probability Distribution of X, the Number of Heads in Three Tosses of a Coin

Value of X	Probability
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
Total	1

TABLE 2 Form of a Discrete Probability Distribution

Value of x	Probability $f(x)$
x_1	$f(x_1)$
x_2	$f(x_2)$
.	.
.	.
.	.
x_k	$f(x_k)$
Total	1

Form of a Discrete Probability Distribution.

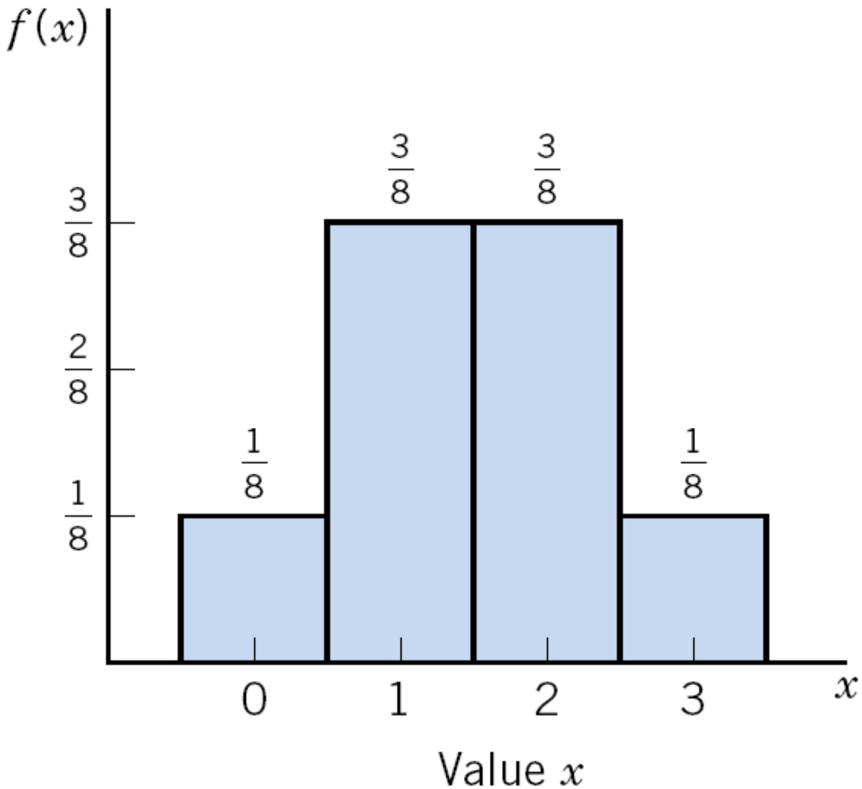
The **probability distribution** of a discrete random variable X is described as the function

$$f(x_i) = P[X = x_i]$$

which gives the probability for each value and satisfies:

1. $f(x_i) \geq 0$ for each value x_i of X
2. $\sum_{i=1}^{k} f(x_i) = 1$

Graphical Representation of a Probability Distribution: Probability Histogram



The probability histogram of *X*, the number of heads in three tosses of a coin

Probability Distribution for News Source Preference

Suppose 60% of the students at a large university prefer getting their daily news from the Internet as opposed to television. These are the only two choices. Four students are randomly selected. Let X be the number of students sampled who prefer news from the Internet. Obtain the probability distribution of X and plot the probability histogram.

$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$
TTTT	TTTI	TTII	TIII	IIII
	TTIT	TITI	ITII	
	TITT	TIIT	IITI	
	ITTT	ITTI	IIIT	
		ITIT		
		IITT		

The Probability Distribution of X i

x	$f(x)$
0	.0256
1	.1536
2	.3456
3	.3456
4	.1296
Total	1.0000

The Probability Distribution of X .

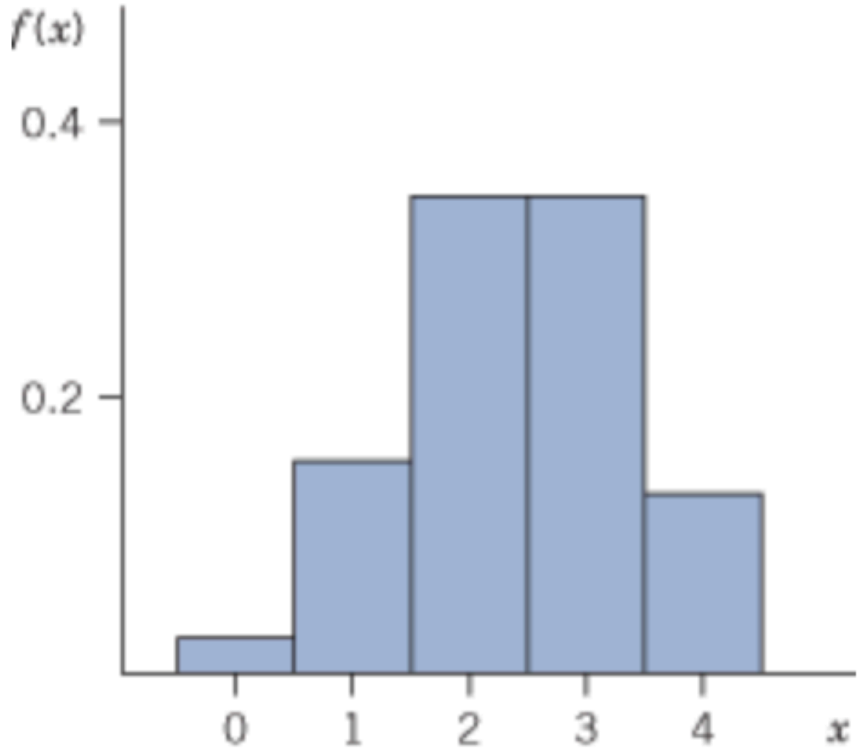


Figure 2 Probability histogram for news source preference.

Probability histogram.

Example: Determining the Probability

Table below describes the number of homework assignments due next week for a randomly selected set of students taking at least 14 credits.

TABLE 5 A Probability Distribution for Number of Homework Assignments Due Next Week

Value x	Probability $f(x)$
0	.02
1	.23
2	.40
3	.25
4	.10

A) Determine the probability that X is equal to or larger than 2.

$P(X \geq 2) = f(2) + f(3) + f(4)$

$= 0.40 + 0.25 + 0.10 = 0.75$

B) What is the probability that X is less than or equal to 2.

$P(X \leq 2) = f(0) + f(1) + f(2) = 0.02 + 0.23 + 0.40 = 0.65$

Example 1 book – *Discrete probability distribution*

•Dr. Mendoza developed a test to measure boredom tolerance. He administered it to a group of 20,000 adults between the ages of 25 and 35.

•The possible scores were 0, 1, 2, 3, 4, 5, and 6, with 6 indicating the highest tolerance for boredom. The test results for this group are shown in Table 5-1.

Score	Number of Subjects
0	1400
1	2600
2	3600
3	6000
4	4400
5	1600
6	400

Boredom Tolerance Test Scores for 20,000 Subjects

Table 5-1

Example 1 – *Discrete probability distribution* cont' d

- **a.** If a subject is chosen at random from this group, the probability that he or she will have a score of 3 is 6000/20,000, or 0.30. In a similar way, we can use relative frequencies to compute the probabilities for the other scores (Table 5-2).

Score x	Probability $P(x)$
0	0.07
1	0.13
2	0.18
3	0.30
4	0.22
5	0.08
6	0.02
	$\Sigma P(x) = 1$

Probability Distribution of Scores on Boredom Tolerance Test

Table 5-2

Example 1 – *Discrete probability distribution* cont' d

•**b.** The graph of this distribution is simply a relative-frequency histogram (see Figure 5-1) in which the height of the bar over a score represents the probability of that score.

Graph of the Prot

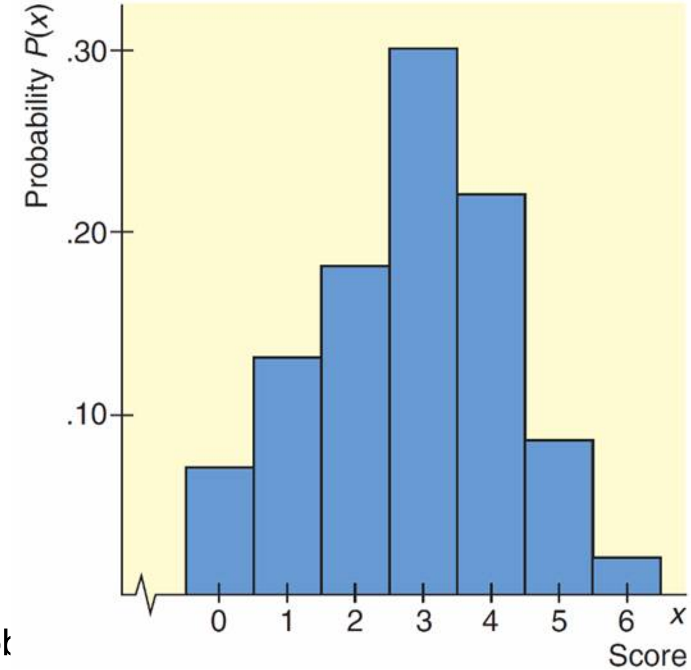


Figure 5-1

Mean(Expected Value) and Standard Deviation of a Probability Distribution

- The expected value of a random variable X is given by

$$\mu = \sum xP(x)$$

- The Standard Deviation of a random variable is given by

$$\sigma = \sqrt{\sum x^2 P(x) - \mu^2}$$

Example: Calculating the Population mean (Expected) number of heads:

With X denoting the number of heads in three tosses of a fair coin, Calculate the mean (Expected) Value of X.

The expected value of X = μ=1.5

Note: P(x)= f(x)

TABLE 6 Mean of the Distribution of Table 1

x	$f(x)$	$xf(x)$
0	$\frac{1}{8}$	0
1	$\frac{3}{8}$	$\frac{3}{8}$
2	$\frac{3}{8}$	$\frac{6}{8}$
3	$\frac{1}{8}$	$\frac{3}{8}$
Total	1	$\frac{12}{8} = 1.5 = \mu$

Example: Expected Value

A trip insurance policy pays $3000 to the customer in case of a loss due to theft or damage on a five-day trip. If the risk of such a loss is assessed to be 1 in 200, what is the expected cost, per customer, to cover?

Let, X be the payment per customer. The probability that the company will be liable to pay $3000 to a customer is 1/200= 0.005. Hence, the probability distribution of X, the payment per customer, is as given in the table.

Payment x	Probability $f(x)$
$ 0	.995
$3000	.005

We calculate the expected value

E(X)= 0*0.995+3000*0.005 = $15.00

Read: NO Casino Game Has a Positive Expected Profit

TABLE 8 Calculation of Variance by the Alternative Formula

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	.1	.0	.0
1	.2	.2	.2
2	.4	.8	1.6
3	.2	.6	1.8
4	.1	.4	1.6
Total	1.0	2.0 $= \mu$	5.2 $= \sum x^2 f(x)$

By using the formula in the previous slide, Variance = $5.2 - (2)^2$
$= 1.2$

And, the **Standard Deviation (S.D.)** = $\sqrt{(1.2)} = 1.095$

Example

A construction company submits bids for two projects. Listed here are the profit and the probability of winning each project. Assume that the outcomes of the two bids are independent.

	Profit	Chance of Winning Bid
Project *A*	\$175,000	.50
Project *B*	\$220,000	.65

(a) List the possible outcomes (win/not win) for the two projects and find their probabilities.

(b) Let X denote the company's total profit out of the two contracts. Determine the probability distribution of X.

(c) If it costs the company \$2000 for preparatory surveys and paperwork for the two bids, what is the expected net profit?

(a) We denote "win" by W and "not win" by N, and attach subscripts A or B to identify the project. Listed here are the possible outcomes and calculation of the corresponding probabilities. For instance,

$$\begin{aligned} P(W_A N_B) &= P(W_A)P(N_B), \text{ by independence} \\ &= 0.50 \times 0.35 = 0.175 \end{aligned}$$

Outcome	Probability
$W_A W_B$	$0.50 \times 0.65 = 0.325$
$W_A W_B$	$0.50 \times 0.35 = 0.175$
$N_A W_B$	$0.50 \times 0.65 = 0.325$
$N_A N_B$	$0.50 \times 0.35 = 0.175$

(b) & (c) The amounts of profit (X) for the various outcomes are listed below.

Outcome	Profit $(\$)$ X
$W_A W_B$	$175{,}000 + 220{,}000 = 395{,}000$
$W_A N_B$	175,000
$N_A W_B$	220,000
$N_A N_B$	0

In the next table, we present the probability distribution of X and calculate $E(X)$.

x	$f(x)$	$xf(x)$
0	0.175	0
175,000	0.175	30,625
220,000	0.325	71,500
395,000	0.325	128,375
Total		$230{,}500 = E(X)$

Expected net profit $= E(X) - \text{cost} = \$230{,}500 - \$2{,}000 = \$228{,}500$.

A surprise quiz contains three multiple-choice questions: Question 1 has four suggested answers, Question 2 has three, and Question 3 has two. A completely unprepared student decides to choose the answers at random. Let X denote the number of questions the student answers correctly.

(a) List the possible values of X.

(b) Find the probability distribution of X.

(c) Find P [At least 1 correct] $= P[X \geq 1]$.

(d) Plot the probability histogram.

Example:
According to a survey, 60% of all students at a large university suffer from math anxiety. Two students are randomly selected from this university. Let *x* denote the number of students in this sample who suffer from math anxiety.

Develop a table of the probability distribution of *x*.

Find the probability that at least one student suffer from math anxiety.

An instant lottery ticket costs \$2. Out of a total of 10,000 tickets printed for this lottery, 1000 tickets contain a prize of \$5 each, 100 tickets have a prize of \$10 each, 5 tickets have a prize of \$1000 each, and one ticket has a prize of \$5000. Let X be the random variable that denotes the net amount a player wins by playing this lottery. Write the probability distribution of X. Determine its mean.

Note that the price of the ticket (\$2) must be deducted from the amount won. For example, the \$5 prize results in a net gain of \$5 – \$2 = \$3.

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
–2	.8894	–1.7788	4	3.5576
3	.1000	.3000	9	.9000
8	.0100	.0800	64	.6400
998	.0005	.4990	996,004	498.0020
4998	.0001	.4998	24,980,004	2498.0004
		$\Sigma xP(x) = -.3991$		$\Sigma x^2P(x) = 3001.1000$

$\mu = \sum xP(x) = -\$0.3991 \approx -\$0.40$

$\sigma = \sqrt{\sum x^2P(x) - \mu^2} = \sqrt{3001.10 - (-.3991)^2} = \$54.78$

On average, the players who play this game are expected to lose \$0.40 per ticket with a standard deviation of \$54.78.