MAT 271E: PROBABILITY AND STATISTICS

PROF. DR. CANAN SARICAM

WEEK 7

RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

EXPECTED VALUE

CHAPTER 5: RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

- When the outcomes of a random trial are quantitative, we associate them with a random variable.

- A random variable is a variable whose numerical value is determined by the outcome of a random trial.

- A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment.

- Ex: Getting 1 when we toss a dice,

random trial is tossing a dice; random variable is 1;

CHAPTER 5: RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

A random variable is denoted by an uppercase letter, such as X and corresponding lowercase letter such as x, is used to denote a possible value of X (x symbolizes the value that is actually observed, often called the realization of X, after the experiment has been performed).

The expression $p(x) = P(X = x)$ is read as

«The probability that X will assume the value x»

Ex: The probability of getting 1 when we toss a dice: $P(X=1)=$

CHAPTER 5: RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

- A random variable that may assume only distinct values on a scale is called discrete random variables.

Ex: The number of defective units in a pair is a discrete random variable; it may assume one of three distinct values (0, 1, 2) but no intermediate values.

- A random variable that may assume any value on a continuum is called a continuous random variable.

Ex: The temp in a warehouse may be any value on the temperature continuum between say -40°C and 45°C.

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

- The probability distribution for a discrete random variable X associates a probability P(X=x) with each distinct outcome x.
- Probability distribution functions are also known as probability mass function.

Ex: Develop the probability distribution function for tossing a dice:

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

All probability distribution functions satisfy two basic properties.

$$p(x) \geq 0$$

$$\sum_{all\ x} p(x) = 1$$

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

- Associated with each probability distribution function, $p(x)$, there is a cumulative distribution function $F(x)$ which gives the probability that the random variable X will assume a value less than or equal to a stipulated value x.

$$F(x)=P(X \leq x)$$

- So, the cumulative probability distribution for a discrete random variable X associates a cumulative probability $P(X<x)$ with every possible outcome x

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Ex: Develop the cumulative probability distribution function for tossing a dice:

x	1	2	3	4	5	6
$F(x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Number of visits x	P(x)
0	0.37
1	0.40
2	0.15
3	0.05
4	0.03

A health system analyst obtained the probability distribution in the table. For the annual number of visits by families to a clinic.

Find the cumulative probability distribution function.

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

- Two important properties of any cumulative probability distribution re:

- All cumulative probabilities $P(X \leq x)$ lie between 0 and 1
- The cumulative probability $P(X \leq x)$ never decreases as x increases

Two random variable x and Y are independent if:

$P\ (X=x \cap Y=y)= P(X=x)P(Y=y)$ for all x and y

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Example 5.2 Formulate and state the probability distribution of the random variable Y, the sum of the faces of two fair dice in single roll of the dice.

**Anything that is a function of one or more random variables is a random variable itself.

Example 5.2 - solution

x_2 \ x_1	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

Value of $y = x_1 + x_2$

Example 5.2 - solution

y	2	3	4	5	6	7	8	9	10	11	12
$p(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

- **Example 5.5**

A carton contains 6 fuses: 3 are 10-amp fuses, 1 is a 15-amp fuse, and 2 are 20-amp fuses. Two fuses are drawn at random and without replacement, from the carton.

Let X be the random variable defined by the sum of the amperages of the 2 fuses that are drawn. What is the probability distribution of X?

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Example 5.5 - solution

Sample space for drawing 2 fuses without replacement

	2nd fuse drawn					
1st fuse drawn	10	10	10	15	20	20
10	—	10, 10	10, 10	10, 15	10, 20	10, 20
10	10, 10	—	10, 10	10, 15	10, 20	10, 20
10	10, 10	10, 10	—	10, 15	10, 20	10, 20
15	15, 10	15, 10	15, 10	—	15, 20	15, 20
20	20, 10	20, 10	20, 10	20, 15	—	20, 20
20	20, 10	20, 10	20, 10	20, 15	20, 20	—

DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Example 5.5 - solution

The dashes along the major diagonal indicate that the same fuse cannot be drawn twice. Note that, if order is considered unimportant, there would be only ${}_6C_2 = 15$ equally likely outcomes, enumerated by the nonempty cells either above or below the diagonal in the table.

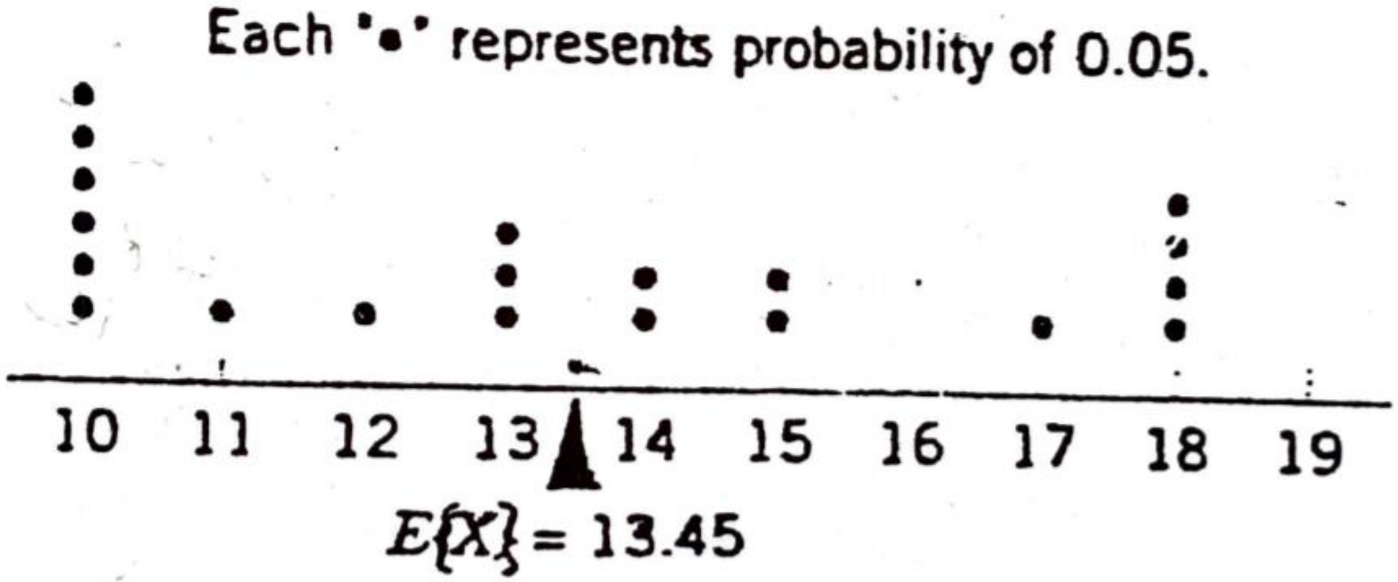
Counting the number of outcomes that yields the same value x and summing their probabilities [i.e., 6 mutually exclusive outcomes generate a sum of 20; $P(X = 20) = \frac{6}{30}$] yield the following probability distribution function for X.

x	20	25	30	35	40
$p(x)$	0.2	0.2	0.4	0.133	0.067

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

If X is a random variable and the random experiment that determines the value of X is repeated many times, a sequence of values is obtained for X.

A summary of these values such as the average (mean) of the values could be used to dentify a central value of the random variable.



EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

The mean or expected value of the discrete random variable X denoted as μ_x or E{ X } is

$$\mu_x = E\{ X \} = \sum_x x \quad P(x)$$

Where the summation is over all outcomes x. The notation E{ } is read as expectation of

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.6

EXAMPLE 5.6.

Consider the following distribution for a random variable X.

x	10	11	12	13	14	15	16	17	18	19
$p(x)$	0.3	0.05	0.05	0.15	0.1	0.1	0	0.05	0.2	0

E{x}=10*0.3+11*0.05+12*0.05+13*0.15+14*0.1+15*0.1+16*0+17*0.05+18*0.2+19*0=13.45

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.7

EXAMPLE 5.7:

We see that a day's production of 850 manufactured parts contains 50 parts that do not conform to customer requirements. Two parts are selected at random, without replacement from the batch. Let the random variable X equal the number of nonconforming parts in the sample. What is the cumulative distribution function of X?

What is the expected value of X?

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.7 - Solution

The question can be answered by first finding the probability mass function of X.

$$P(X = 0) = (800/850)(799/849) = 0.886$$
$$P(X = 1) = 2(800/850)(50/849) = 0.111$$
$$P(X = 2) = (50/850)(49/849) = 0.003$$

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.7 - Solution

$$F(0) = P(X \leq 0) = 0.886$$
$$F(1) = P(X \leq 1) = 0.886 + 0.111 = 0.997$$
$$F(2) = P(X \leq 2) = 1$$

x	0	1	2
$p(x)$	0.886	0.111	0.003

Therefore,

$$\begin{aligned} \mu_X = E\{X\} &= 0p(0) + 1p(1) + 2p(2) \\ &= 0(0.886) + 1(0.111) + 2(0.003) \\ &= 0.117 \end{aligned}$$

Notice that X never assumes the value 0.117, but the weighted average of the possible values of X is 0.117.

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.8

$$p(x) = \frac{2!}{x!(2-x)!} 0.8^x 0.2^{2-x}, \text{ for } x = 0, 1, 2. \text{ Then,}$$

$$\mu_X =$$

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.8

$$\begin{aligned}\mu_X = E(X) &= 0f(0) + 1f(1) + 2f(2)\\ &= 0 + 1(0.32) + 2(0.64)\\ &= 1.6\end{aligned}$$

Notice that X never assumes the value 1.60, but the weighted average of the possible values of X is 1.60.

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.9

Test coverage in semiconductor testing is assumed to be 80% effective. That is, the probability that a defective chip fails the test is 0.8 Three defective chips are to be tested. Assume the failure of each defective chip is independent of the other tests. Let the random variable X be the number of defective chips that fail the test. What is the expected value of X?

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.9 - solution

Outcomes	x	Probability
(fff)	3	0.512
(ffp)	2	0.128
(fpf)	2	0.128
(fpp)	1	0.032
(pff)	2	0.128
(pfp)	1	0.032
(ppf)	1	0.032
(ppp)	0	0.008

The possible outcomes of the experiment and the corresponding values of X are shown below.

The probability of each outcome is found by using the independence assumption.

In the table below, p denotes that a chip passes the test and f denotes that a chip fails the test.

Therefore, the outcome that the first chip passes and the remaining chips fail is denoted as pff.

Furthermore, P(pff) = 0.2 X 0.8 X 0.8 ..:.. O. 128.

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.9 - solution

From these results, the probability mass function of X is found.

x	0	1	2	3
$p(x)$	0.008	0.096	0.384	0.512

Therefore,

$$E\{X\} = 0(0.008) + 1(0.096) + 2(0.384) + 3(0.512) = 2.40$$

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Variance of Random Variable

The variance of a discrete random variable X is denoted by $\sigma^2\{X\}$ and defined:

$$\boxed{\sigma^2\{X\} = \sum_x (x - E\{X\})^2 P(x)}$$

where the summation is over all outcomes x. The notation $\sigma^2\{\ \}$ is read "variance of."

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

- **Standard deviation of a random variable**

The variance $\sigma^2\{X\}$ is expressed in the squared units of X. When we take the positive square root of $\sigma^2\{X\}$, we return to the original units and obtain the standard deviation of X.

The positive square root of the variance of X is called the standard deviation of X and is denoted by $\sigma\{X\}$:

$$\sigma\{X\} = \sqrt{\sigma^2\{X\}}$$

The notation $\sigma\{\ \}$ is read "standard deviation of."

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.10

Two new product designs are to be compared on the basis of revenue potential Marketing feels that the revenue from Design A can be predicted quite accurately to be $3 million.

The revenue potential of Design B is more difficult to assess. Marketing concludes that there is a probability of 0.3 that the revenue from Design B will be $7 million, but there is a 0.7 probability that the revenue will be only $2 million. Which design do you prefer?

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.10 - solution

Let X denote the revenue from Design A. Because there is no uncertainty in the revenue from Design A, we can model the distribution of the random variable X as \$3 million with probability one. Therefore, E{X} = \$3 million.

Let Y denote the revenue from Design B. The expected value of Y in millions of dollars is

$$E\{Y\} = \$7(0.3) + \$2(0.7) = \$3.5$$

Because E{Y} exceeds E{X}, we might prefer Design B. However,

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Let’s Calculate the variance and standard deviations for Design A an B?

EXPECTED VALUE AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Example 5.12-5.13

EXAMPLE 5.12

In Example 5.10

3.625 + 1.575

$$\sigma_Y^2 = (7 - 3.5)^2(0.3) + (2 - 3.5)^2(0.7)$$
$$= 5.25 \text{ millions of dollars squared}$$

$\sqrt{5.25} = 2.29$

Because the units of the variables in this example are millions of dollars, and because the variance of a random variable squares the deviations from the mean, the units of σ_Y^2 are millions of dollars squared. These units make interpretation difficult.

EXAMPLE 5.13

In Examples 5.10 and 5.12, $\sigma_Y = 2.29$. Because the units of standard deviation are the same as the units of the random variable, the standard deviation is easier to interpret. In this example, we can summarize our results as "the average deviation of Y from its mean is \$2.29 million."

FUNCTIONS OF RANDOM VARIABLES: LINEAR FNS

- Often an interest centers on a random variable W that is a linear function of X, **W=a+bx,** where a and b are constants.

- To find the expected value of a linear function of X, we could obtain the probability distribution of **W=a+bX** from that of X and then use the definiton of expected value to find **E{W}**

FUNCTIONS OF RANDOM VARIABLES: LINEAR FNS

When the expected value of X is already known, there is no need to find the probability distribution of W first. The following theorem can be used directly.

$$E\{a + bX\} = a + bE\{X\} \qquad (5.4)$$

Since $W = a + bX$, this theorem says that $E\{W\}$ is the same linear function of $E\{X\}$ as W is of X. There are three important special cases:

$$E\{a\} = a$$
$$E\{bX\} = bE\{X\}$$
$$E\{a + X\} = a + E\{X\}$$

FUNCTIONS OF RANDOM VARIABLES: LINEAR FNS

Example 5.14

EXAMPLE 5.14 Production Run

Let X denote the size of a production run in a shop, and let W denote the total cost of the production run. It is known that $W = 200 + 4X$, where $a = \$200$ is the fixed setup cost of a run and $b = \$4$ is the cost per unit produced. If the size of the production run is $x = 40$, for instance, the total cost is $w = 200 + 4(40) = \$360$.

In the production run example, $\boxed{\text{it is known that } E\{X\} = 50}$. Hence, by using theorem, we obtain for $W = 200 + 4X$:

$$E\{W\} = 200 + 4(50) = \$400$$

Thus, the mean cost per production run is \$400.

FUNCTIONS OF RANDOM VARIABLES:LINEAR FNS

- To find the variance of W (**W = a + bX**), which is the linear function of X when the variance of X is already known can be calculated:

$$\sigma^2\{a + bX\} = b^2\sigma^2\{X\}$$

Two important special cases are:

$$\sigma^2\{a + X\} = \sigma^2\{X\}$$

$$\sigma^2\{bX\} = b^2\sigma^2\{X\}$$

FUNCTIONS OF RANDOM VARIABLES:LINEAR FNS

Example 5.14 continued..

Let X denote the size of a production run in a shop, and let W denote the total cost of the production run. It is known that $W = 200 + 4X$, where $a = \$200$ is the fixed setup cost of a run and $b = \$4$ is the cost per unit produced. If the size of the production run is $x = 40$, for instance, the total cost is $w = 200 + 4(40) = \$360$. □

In the production run example, it is known that $\sigma^2\{X\} = 300$. Hence, using (5.5), we obtain for $W = 200 + 4X$:

FUNCTIONS OF RANDOM VARIABLES:LINEAR FNS

Example 5.14 continued..

In the production run example, it is known that $\sigma^2\{X\} = 300$. Hence, using (5.5), we obtain for $W = 200 + 4X$:

b

$$\sigma^2\{W\} = (4)^2(300) = 4800$$

so $\sigma\{W\} = \sqrt{4800} = \boxed{\$69.}$

Note that the constant setup cost $a = \$200$ had no effect on the variance of the total cost W. This is intuitively reasonable since the fixed setup cost does not vary from one production run to another. The fixed setup cost only shifts the position of the probability distribution but does not affect the variability of the distribution. □

FUNCTIONS OF RANDOM VARIABLES: Sum or difference of two independent random variables

(5.6)

The expected value and the variance of the sum of two independent random variables X and Y are as follows:

Expected Value	$E\{X+Y\} = E\{X\} + E\{Y\}$	(5.6a)
Variance	$\sigma^2\{X+Y\} = \sigma^2\{X\} + \sigma^2\{Y\}$	(5.6b)

Note that the expected value of a sum of two independent random variables is simply the sum of the expected values of each of the two random variables, and similarly for the variance.

The expected value and the variance of a *difference* of two independent random variables are presented next.

(5.7)

The expected value and the variance of the difference of two independent random variables X and Y are as follows:

Expected Value	$E\{X-Y\} = E\{X\} - E\{Y\}$	(5.7a)
Variance	$\sigma^2\{X-Y\} = \sigma^2\{X\} + \sigma^2\{Y\}$	(5.7b)

FUNCTIONS OF RANDOM VARIABLES: Sum or difference of two independent random variables

Example 5.15

EXAMPLE 5.15

Let X denote the bonus amount received by salesperson Anderson, and let Y denote the bonus amount received by salesperson Brody. Then $T = X + Y$ represents the total bonus amount received by the two salespeople. □

FUNCTIONS OF RANDOM VARIABLES: Sum or difference of two independent random variables

Example 5.15

EXAMPLES

In the sales bonus example, $E\{X\} = E\{Y\} = \$200$, and $\sigma^2\{X\} = \sigma^2\{Y\} = 60{,}000$ (calculations not shown). Hence, for the total bonus amount $T = X + Y$,

we obtain by (6.7):

$$E\{T\} = 200 + 200 = \$400$$

$$\sigma^2\{T\} = 60{,}000 + 60{,}000 = 120{,}000$$

$\sigma\{X\} = \sigma\{Y\} = \$\ 244.95$

$\sigma\{T\} = \$\ 346.4$

These are the same results, of course, as those obtained from the probability distribution of T.

FUNCTIONS OF RANDOM VARIABLES: Sum or difference of two independent random variables

Example 5.16

Gross Profit ☐ EXAMPLE 5.16

Let X denote quarterly sales revenues and Y quarterly direct costs. Then $W = X - Y$ represents quarterly gross profit.

FUNCTIONS OF RANDOM VARIABLES: Sum or difference of two independent random variables

Example 5.16

In the gross profit example, it is known that $E\{X\} = \$100{,}000$, $E\{Y\} = \$70{,}000$, $\sigma^2\{X\} = 8{,}000{,}000$, $\sigma^2\{Y\} = 4{,}000{,}000$, and X and Y are independent. For the gross profit $W = X - Y$, we obtain by (6.8):

$$E\{W\} = 100{,}000 - 70{,}000 = \$30{,}000$$

$$\sigma^2\{W\} = 8{,}000{,}000 + 4{,}000{,}000 = 12{,}000{,}000$$

The standard deviation of W is $\sigma\{W\} = \sqrt{12{,}000{,}000} = \$3464$. □

($\sigma\{X\} = \$2828.4$, $\sigma\{Y\} = \$\ 2000$)