Chapter 04

Discrete Probability Distributions

Biostatistics For the Health Sciences

Dr. Moustafa Omar Ahmed Abu-Shawiesh
Professor of Statistics

4.1 Introduction

Chapter 3 defined probability and introduced some basic tools used in working with probabilities. By assessing the probabilities of certain events from actual past data, we can consider specific probability models (probability distributions) that fit our problems. This chapter introduces the general concept of a discrete random variable and describes the binomial distribution in depth which is considered an important application of this type of random variables.

4.2 Random Variables

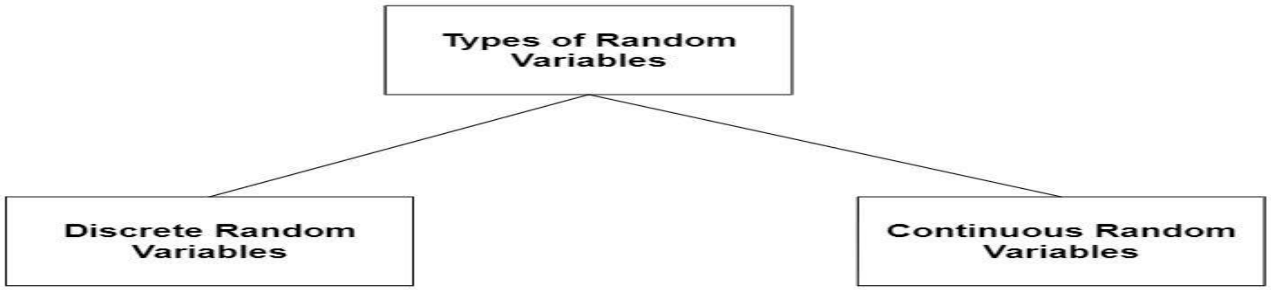
In this section we will consider the concept of a random variable (r. v.) as follows:

DEFINITION 4.1 A random variable is a function that assigns numeric values to different outcomes in a sample space (S) and can be defined as follows:.

Definition: Random Variable

A random variable, say X, is a real-valued function whose domain is the sample space (S) of a random experiment and range is a subset (A) from the set of real numbers ($\mathbb{R}$), that is: $X: S \rightarrow A \subset \mathbb{R}$

Two types of random variables are discussed in this text: discrete and continuous.



Notation: We usually denote random variables by capital letters such as X , Y , Z , …

DEFINITION 4.2 A random variable for which there exists a discrete set of numeric values is a **discrete random variable.**

Notation: A discrete r. v. can take on only a finite number or a countable infinite number of values.

Example

Here are some examples of a discrete random variable:

- The number of arrivals at an emergency room between midnight and 6:00 A.M.
- The number of boys in a three-child family.
- The number of defective light bulbs in a box of 100 bulbs.
- The number of clerical errors on a medical report.
- The number of new cases of influenza in KAUH next month.
- The number of episodes of otitis media in the first 2 years of life.

DEFINITION 4.3 A random variable whose possible values cannot be enumerated is a **continuous random variable.**

Notation: The outcomes for a continuous random variable are measured and it can take on an infinite number of possibilities in an interval.

Example

Here are some examples of a continuous random variable:

- The time between patients entering an examination room at a Otolaryngology clinic in KAUH.
- The temperature of a cup of coffee served at a hospital for patients.
- The average amount of money spent on treatments in hospitals each year by households.
- The amount of rain recorded at KAUH one day.
- The average weight of newborn babies born in a particular Jordanian hospital one month.
- The distance an ambulance car is driven each day in a certain hospital.

Notation

The value of the discrete random variable should be always an integer value, while the value of the continuous random variable can be integer and non-integer.

4.3 The Probability Mass Function for a Discrete Random Variable

The values taken by a discrete random variable, say X, and its associated probabilities $P(X = r)$ can be expressed by a rule or a relationship called a probability mass function (pmf).

DEFINITION 4.4 A **probability-mass function** is a mathematical relationship, or rule, that assigns to any possible value r of a discrete random variable X the probability $Pr(X = r)$. This assignment is made for all values r that have positive probability. The probability-mass function is sometimes also called a **probability distribution.**

Notation

For any probability mass function (pmf), two conditions should be satisfied as follows:

(1) The probability of any particular value must be between 0 and 1, that is:

$$0 \leq P(X = r) \leq 1 \quad \text{for all values of } r$$

(2) The sum of the probabilities of all values must exactly equal 1, that is:

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

Notation

The probability mass function (pmf) for a discrete random variable (X) can be displayed in a table giving the values and their associated probabilities, or it can be expressed as a mathematical formula giving the probabilities of all possible values.

Examples of Probability Mass Functions

Example (1)

A university medical research center finds out that treatment of skin cancer by the use of chemotherapy has a success rate of 70%. The probability mass function (probability distribution) of X successful cures of five patients treated with chemotherapy is given in the table below:

r	0	1	2	3	4	5
$P(X = r)$	0.002	0.029	0.132	0.309	0.360	0.168

or it can be written as follows:

$$f(r) = P(X = r) = \begin{cases} 0.002 & , \quad r = 0 \\ 0.029 & , \quad r = 1 \\ 0.132 & , \quad r = 2 \\ 0.309 & , \quad r = 3 \\ 0.360 & , \quad r = 4 \\ 0.168 & , \quad r = 5 \end{cases}$$

Example (2)

Suppose that the probability mass function (pmf) for the discrete random variable X which represents the number of women who will be cured of breast cancer after giving them doses of chemotherapy in a random sample of 4 women selected from the King Hussein Cancer Center (KHCC) in Jordan is given as follows:

$$f(r) = P(X = r) = \begin{cases} \dfrac{2r+1}{25} & , \quad r = 0, 1, 2, 3, 4 \\ 0 & , \quad Otherwise\ (Elsewhere) \end{cases}$$

which can be expressed in the form of a table as follows:

r	0	1	2	3	4
$P(X = r)$	0.04	0.12	0.20	0.28	0.36

or it can be written as follows:

$$f(r) = P(X = r) = \begin{cases} 0.04 & , \quad r = 0 \\ 0.12 & , \quad r = 1 \\ 0.20 & , \quad r = 2 \\ 0.28 & , \quad r = 3 \\ 0.36 & , \quad r = 4 \end{cases}$$

Example (3)

Determine which one of the following tables is or is not a probability mass function (probability distribution) for the discrete random variable X:

(I)

(A)

r	0	1	2	3
$P(X = r)$	0.18	0.34	0.36	0.13

(C)

r	0	1	2	3
$P(X = r)$	0.10	0.42	0.03	0.44

(B)

r	0	1	2	3
$P(X = r)$	0.12	0.13	0.61	0.14

(D)

r	0	1	2	3
$P(X = r)$	0.26	0.44	− 0.15	0.45

(II)

(A)

r	$P(X = r)$
0	0.15
1	0.25
2	0.10
3	0.25
4	0.30

(C)

r	$P(X = r)$
0	0.15
1	0.20
2	0.30
3	0.10

(B)

r	$P(X = r)$
0	0.15
1	- 0.20
2	0.30
3	0.20
4	0.15

(D)

r	$P(X = r)$
-1	0.15
0	0.30
1	0.20
2	0.15
3	0.10
4	0.10

Example (4)

Hypertension Many new drugs have been introduced in the past several decades to bring hypertension under control—that is, to reduce high blood pressure to normotensive levels. Suppose a physician agrees to use a new antihypertensive drug on a trial basis on the first four untreated hypertensives she encounters in her practice, before deciding whether to adopt the drug for routine use. Let X = the number of patients of 4 who are brought under control. Then X is a discrete random variable, which takes on the values 0, 1, 2, 3, 4. Also, suppose that from the previous experience with the drug, the drug company expects that for any clinical practice the probability that 0 patients of 4 will be brought under control is 0.008, 1 patient of 4 is 0.076, 2 patients of 4 is 0.265, 3 patients of 4 is 0.411, and all 4 patients is 0.240. This probability mass function (pmf) or probability distribution is displayed in Table 4.1 given below:

Table 4.1 Probability mass function (pmf) for the hypertension control

r	0	1	2	3	4
$P(X = r)$	0.008	0.076	0.265	0.411	0.240

Use this table to answer the following questions:

(a) What is the probability that exactly 2 patients will be brought under control?

Answer

$P(X = 2) = 0.265$

(b) What is the probability that less than 2 patients will be brought under control?

Answer

P(less than 2 patients will be brought under control)

$= P(X < 2)$

$= P(X = 0 \text{ or } 1)$

$= P(X = 0) + P(X = 1)$

$= 0.008 + 0.076$

$= 0.084$

(c) What is the probability that 2 or fewer patients will be brought under control?

Answer

P(2 or fewer patients will be brought under control)

$= P(X \leq 2)$

$= P(X = 0) + P(X = 1) + P(X = 2)$

$= 0.008 + 0.076 + 0.265$

$= 0.349$

(d) What is the probability that more than 2 patients will be brought under control?

Answer

P(more than 2 patients will be brought under control)

= P(X > 2)

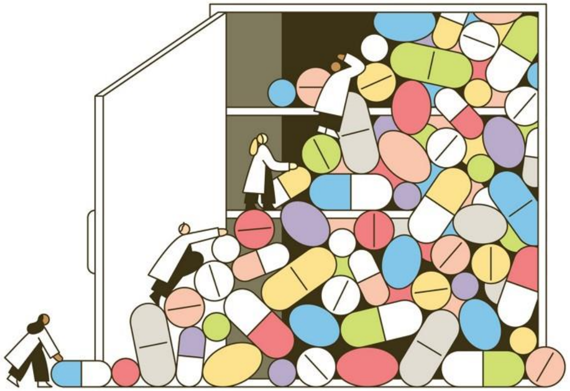
= P(X = 3 or 4)

= P(X = 3) + P(X = 4)

= 0.411 + 0.240 = 0.651

OR

P(X > 2) = 1 - P(X ≤ 2) = 1 – 0.349 = 0.651



Rule: For a constant k we have P(X > k) + P(X ≤ k) = 1 .

(e) What is the probability that more than 1 and less than or equal to 3 patients will be brought under control?

Answer

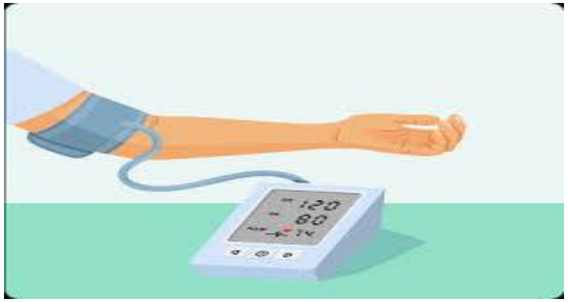
P(more than 1 and less than or equal to 3 patients will be brought under control)

= P(1 < X ≤ 3)

= P(X = 2) + P(X = 3)

= 0.265 + 0.411

= 0.676



4.4 The Expected Value of a Discrete Random Variable

If a random variable has a large number of values with positive probability, then the probability mass function (pmf) is not a useful summary measure. Indeed, we face the same problem as in trying to summarize a sample by enumerating each data value. Measures of location and spread can be developed for a random variable in much the same way as they were developed for samples. The analog to the arithmetic mean $\overline{X}$ is called the expected value of a random variable, or population mean, and is denoted by E(X) or μ. The expected value represents the "average" value of the random variable. It is obtained by multiplying each possible value (r) by its respective probability ($P(X = r)$)and summing these products over all the values that have positive (that is, nonzero, $P(X = r) > 0$) probability.

DEFINITION 4.5

The expected value of a discrete random variable X, denoted by μ or E(X), is defined as follows:

$$\mu \equiv E(X) = \sum_{i=1}^{k} r_i\, P(X = r_i)$$
$$= r_1\ P(X = r_1) + r_2\, P(X = r_2) + \ldots + r_k\, P(X = r_k)$$

where the r_i's are the values the random variable assumes with positive probability. Note that k may be either finite or infinite. In either case, the individual values must be distinct from each other.

4.5 The Variance and Standard Deviation of a Discrete Random Variable

The variance of the random variable, X , or population variance, is denoted by Var(X) or σ^2. The variance represents the spread, relative to the expected value, of all values that have positive probability. In particular, the variance is obtained over all values that have positive probability and can be defined as follows:

DEFINITION 4.6

The variance of a discrete random variable X, denoted by Var(X) or σ^2, is defined as follows:

$$\begin{aligned}\sigma^2 \equiv \mathrm{Var(X)} &= \mathrm{E}(X-\mu)^2\\ &= \textstyle\sum_{i=1}^{k}(r_i-\mu)^2 P(X=r_i)\\ &= \textstyle\sum_{i=1}^{k} r_i^2 P(X=r_i) - \left(\sum_{i=1}^{k} r_i\, P(X=r_i)\right)^2\\ &= E(X^2) - (E(X))^2\\ &= E(X^2) - (\mu)^2\end{aligned}$$

which the random variable takes on positive probability. The standard deviation of a random variable X, denoted by σ or sd(X), is defined by the square root of its variance, that is:

$$\mathrm{sd(X)} = \sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var(X)}}$$

Example (5)

Hypertension The probability mass function (pmf) or probability distribution for X = the number of patients of 4 who are brought under control displayed in Table 4.1 is given as follows:

r	0	1	2	3	4
$P(X = r)$	0.008	0.076	0.265	0.411	0.240

(a) Find the expected value for the discrete random variable (X)?

(b) Find the variance and standard deviation for the discrete random variable (X)?

Solution

r	$P(X = r)$	r^2	$r * P(X = r)$	$r^2 * P(X = r)$
0	0.008	0	0	0
1	0.076	1	0.076	0.076
2	0.265	4	0.530	1.060
3	0.411	9	1.233	3.699
4	0.240	16	0.960	3.840
Sum	**1**	**---**	**2.799**	**8.675**

(a) Expected Value (Mean)

$$\mu \equiv E(X) = \sum_{i=1}^{k=5} r_i \, P(X = r_i) = 2.799 \approx 2.80$$

Conclusion: Thus, on the average about 2.80 hypertensives would be expected to be brought under control for every 4 who are treated.

(b) Variance

$$\sigma^2 \equiv \text{Var}(X) = E(X^2) - (\mu)^2$$
$$= 8.675 - (2.799)^2$$
$$= 8.675 - 7.834401$$
$$= 0.840599$$

(c) Standard deviation

$$\sigma = \text{sd}(X) = \sqrt{\sigma^2} = \sqrt{0.840599}$$
$$= 0.92$$

Conclusion: Thus, on the average about 2.80 hypertensives would be expected to be brought under control for every 4 who are treated, and this value(2.80) will be increased or decreased on the average by 0.92.

Example (6)

Otolaryngology Consider the discrete random variable X which representing the number of episodes of otitis media in the first 2 years of life. Suppose this random variable has a probability mass function (pmf) as given in Table 4.3 shown below:

Table 4.3 The PMF for number of episodes of otitis media in the first 2 years of life

r	0	1	2	3	4	5	6
$Pr(X = r)$	.129	.264	.271	.185	.095	.039	.017

Answer the following:

(a) Find the expected value for the discrete random variable (X)? $\mu = E(X)$ =2.038

(b) Find the variance and standard deviation for the discrete random variable (X)?
$E(X^2)$= 6.12 , $\sigma^2 = \text{Var}(X) = 1.967$, $\sigma = \text{sd}(X)$ =1.402

4.8 The Binomial Distribution

All discrete random variables follow a binomial distribution have a common structure as follows:

- A sample of n independent trials ($n \geq 1$).
- Each one of the n trials have only two possible outcomes.
- These possible outcomes are denoted as: "success" and "failure."
- The probability of a success at each trial is assumed to be some constant p.
- The probability of a failure at each trial is 1 – p = q implies that p + q = 1.

Notice that the term "success" is used in a general way, without any specific contextual meaning.

DEFINITION 4.9

Factorial: The special symbol (n!) is called the n factorial and is defined as follows:

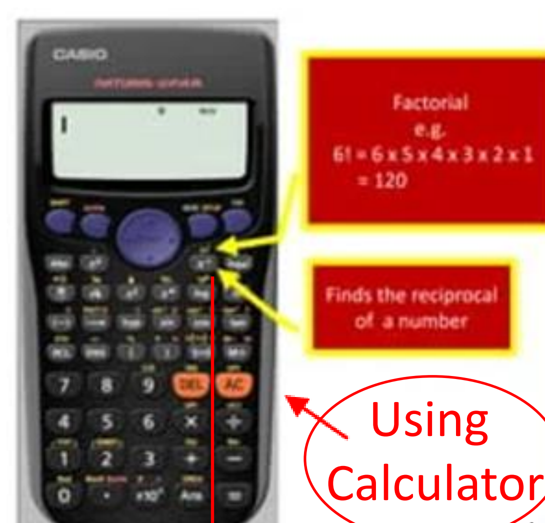
$$n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$$

Special Cases: $0! = 1$ and $1! = 1$

EXAMPLE 4.18 Evaluate 5 factorial. (n)

(n!) $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

n SHIFT $x!$ = ...
5 SHIFT 5! = 120



Using Calculator

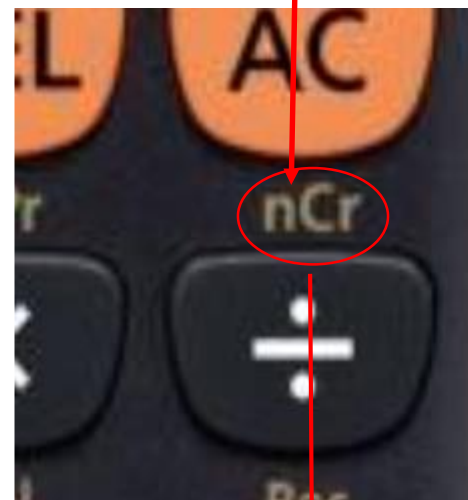
DEFINITION 4.11

Binomial Coefficient: For any non-negative integers n, k, where n ≥ k, the number of combinations of n things taken k at a time is known as the binomial coefficient denoted by ${}_nC_k$ or $\binom{n}{k}$ and can be calculated as follows:

$$ {}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!} $$

Henceforth, for consistency we will always use the more common notation $\binom{n}{k}$ for combinations. In words, this is expressed as "n choose k."

Using Calculator



n **SHIFT nCr** r ➡ 7 **SHIFT nCr** 3 = 35

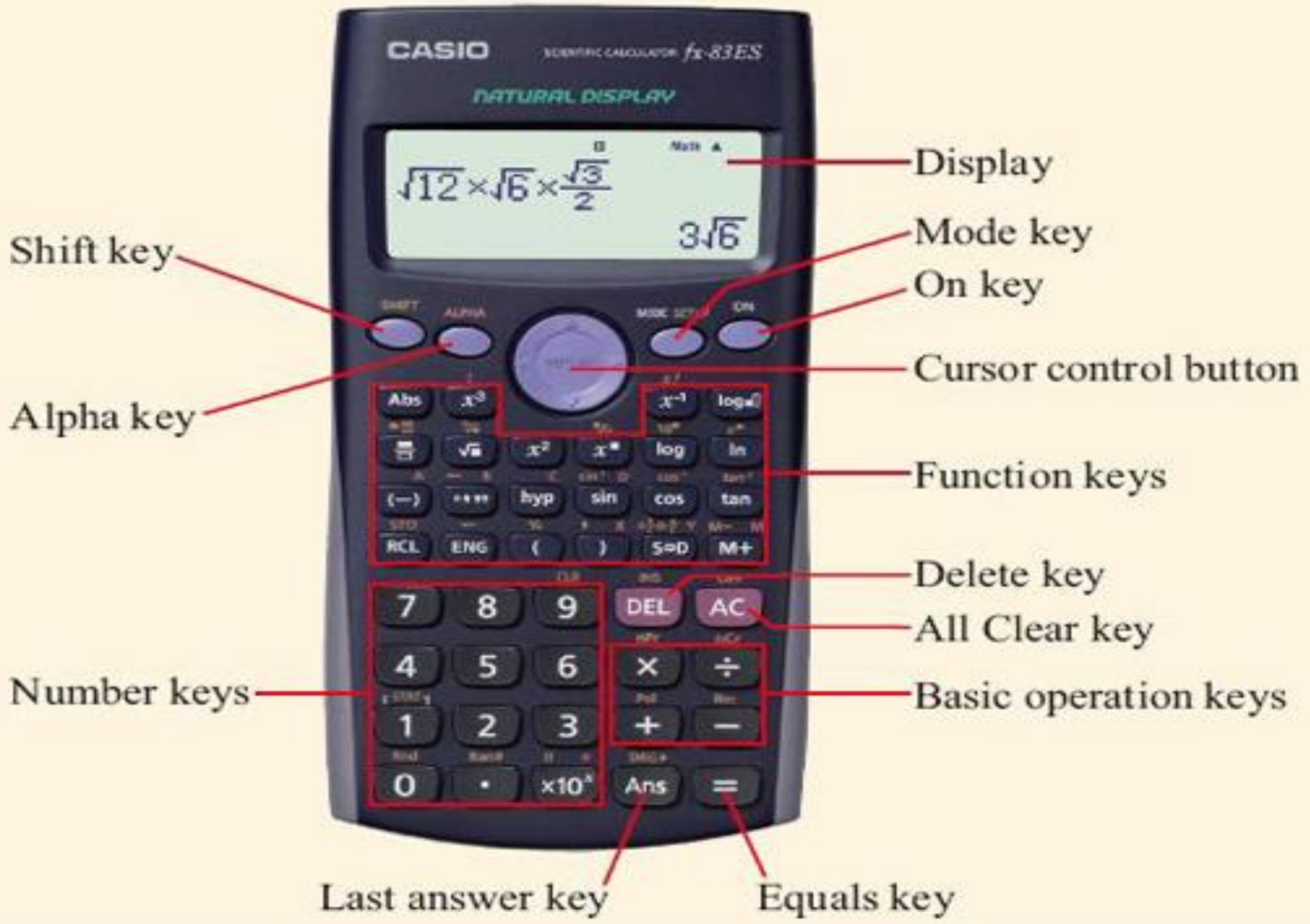
EXAMPLE 4.21

Evaluate ${}_7C_3$?

Solution

$$ {}_7C_3 = \binom{7}{3} = \frac{7!}{3!(7-3)!} = \frac{7!}{3!\cdot 4!} = \frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{3\cdot 2\cdot 1\cdot 4\cdot 3\cdot 2\cdot 1} = \frac{5040}{144} = 35 $$

${}_nC_k$ $\binom{n}{k}$ $\frac{n!}{k!(n-k)!}$



CASIO
fx-83ES
NATURAL DISPLAY
Display
Mode key
On key
Shift key
Cursor control button
Alpha key
Function keys
Delete key
All Clear key
Number keys
Basic operation keys
Last answer key
Equals key

EQUATION 4.5 Binomial Probability Mass Function

If the discrete random variable X is defined to be as "the number of "successes" in the n independent Bernoulli trials", then X is said to have a Binomial distribution denoted by $X \sim B(n\,,p)$ where n is the number of trials and p is the probability of success. Let $X \sim B(n, p)$, then the probability mass function (pmf) or probability distribution for X denoted by $\mathrm{f(k) = P(X = k)}$ is given as follows:

$$f(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & , \quad k = 0, 1, 2, \dots, n \\ 0 & , \quad Otherwise \end{cases}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

4.9 Expected Value and Variance of the Binomial Distribution

The expected value (mean), the variance and the standard deviation of the binomial distribution are important both in terms of our general knowledge about the binomial distribution and for our later work on estimation and hypothesis testing.

EQUATION 4.7

The expected value and the variance of a binomial distribution are obtained as follows:

Expected Value (Mean): $\mu = E(X) = np$

Variance: $Var(X)$: $\sigma^2 = np(1-p) = npq$

Standard Deviation: $\sigma = sd(X) = \sqrt{Var(X)} = \sqrt{np(1-p)} = \sqrt{npq}$

Special Case

If $(n = 1)$, then X takes two values, 0 and 1, with probabilities p and $q = 1 - p$. In this case the Binomial distribution is called the Bernoulli distribution denoted by $X \sim B(1, p)$ or $X \sim Ber(p)$, and the probability distribution is given as follows:

Examples of a Binomial Random Variable

- Number of workers suffering from COVID-19 disease in a random sample of 6 workers.
- Number of patients recover from a COVID-19 disease in a random sample of 9 patients suffering from this disease.

Example

Suppose it is known that a new drug is successful in curing a COVID-19 in 22% of the cases. If it is tried on a random sample of 5 patients, then answer the following:

(a) Find the probability that:

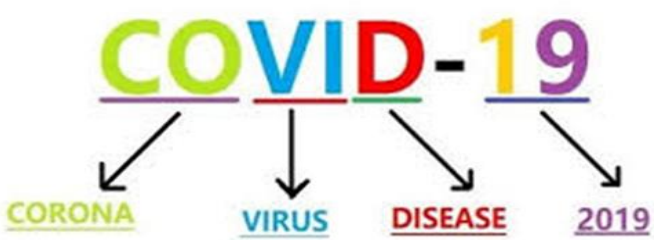
(i) no one of patients will be cured?

(ii) exactly 3 patients will be cured?

(iii) at least 2 patients will be cured?

(iv) at most 4 patients will be cured?

(v) from 1 to 4 patients will be cured?



(b) Find the mean and the standard deviation for the number of patients who are cured?

Solution

The success is the cure of patient and by assuming that the results for individual patients are independent, we note that the Binomial distribution with $n = 5$ and $p = 0.22$ is appropriate for $X =$ number of patients who are cured in the random sample of size 5 patients, that is, $X \sim B(5, 0.22)$, with a probability mass function (pmf) given as follows:

$$f(k) = P(X = k) = \begin{cases} \binom{5}{k}(0.22)^k(1 - 0.22)^{5-k} & , \quad k = 0, 1, 2, 3, 4, 5 \\ 0 & , \quad Otherwise \end{cases}$$

where $\binom{5}{k} = \frac{5!}{k!(5-k)!}$

By using the above equation, we get the probability mass function (probability distribution) table for the number of patients who are cured, X, as follows:

$$f(0) = P(X = 0) = \binom{5}{0}(0.22)^0(0.78)^{5-0} = \left(\frac{5!}{0!\,5!}\right)(1)(0.78)^5 = (1)(1)(0.2887) = 0.289$$

$$f(1) = P(X = 1) = \binom{5}{1}(0.22)^1(0.78)^{5-1} = \left(\frac{5!}{1!\,4!}\right)(0.22)(0.78)^4 = (5)(0.22)(0.3702) = 0.407$$

$$f(2) = P(X = 2) = \binom{5}{2}(0.22)^2(0.78)^{5-2} = \left(\frac{5!}{2!\,3!}\right)(0.22)^2(0.78)^3 = (10)(0.0484)(0.4746) = 0.230$$

$$f(3) = P(X = 3) = \binom{5}{3}(0.22)^3(0.78)^{5-3} = \left(\frac{5!}{3!\,2!}\right)(0.22)^3(0.78)^2 = (10)(0.0106)(0.6084) = 0.065$$

$$f(4) = P(X = 4) = \binom{5}{4}(0.22)^4(0.78)^{5-4} = \left(\frac{5!}{4!\,1!}\right)(0.22)^4(0.78)^1 = (5)(0.0023)(0.78) = 0.009$$

$$f(5) = P(X = 5) = \binom{5}{5}(0.22)^5(0.78)^{5-5} = \left(\frac{5!}{5!\,0!}\right)(0.22)^5(0.78)^0 = (1)(0.0005)(1) = 0.001$$

The Probability Mass Function (Probability Distribution) for X

k	0	1	2	3	4	5
$P(X = k)$	0.289	0.407	0.230	0.009	0.065	0.001

(a) Find the probability that:

(i) Probability that no one of patients will be cured
= P(X = 0) = 0.289

(ii) Probability that exactly 3 patients will be cured
= P(X = 3) = 0.065

(iii) Probability that at least 2 patients will be cured
= P(X ≥ 2)
= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)
= 0.229 + 0.065 + 0.009 + 0.001
= 0.304

OR

(iii) Probability that at least 2 patients will be cured
= P(X ≥ 2)
= 1 – P(X < 2)
= 1 – { P(X = 0) + P(X = 1) }
= 1 – { 0.289 + 0.407 } = 1 – 0.696 = 0.304

(iv) Probability that at most 4 patients will be cured

$= P(X \leq 4)$

$= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$= 0.289 + 0.407 + 0.229 + 0.065 + 0.009$

$= 0.999$

OR

(iv) Probability that at most 4 patients will be cured

$= P(X \leq 4)$

$= 1 - P(X > 4)$

$= 1 - \{ P(X = 5)\} = 1 - \{ 0.001\} = 0.999$

(v) Probability that from 1 to 4 patients will be cured

$= P(1 \leq X \leq 4)$

$= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$= 0.407 + 0.229 + 0.065 + 0.009$

$= 0.71$

(b) The mean and the standard deviation for the number of patients who are cured can be calculated as follows:

(i) The Mean: $\mu = E(X) = np = (5)(0.22) = 1.1\ patient.$

(ii) The Standard Deviation: $\sigma = SD(X) = \sqrt{\sigma^2} = \sqrt{np(1-p)}$

$$= \sqrt{(5)(0.22)(0.78)}$$

$$= 0.926\ patient.$$

Using Binomial Tables

Often a number of binomial probabilities need to be evaluated for the same n and p, which would be tedious if each probability had to be calculated from the Equation. Instead, for small n ($n \le 20$) and selected values of p, refer to Table 1 page 867 in the Appendix, where individual binomial probabilities are calculated. In this table, the number of trials (n) is provided in the first column, the number of successes (k) out of the n trials is given in the second column, and the probability of success for an individual trial (p) is given in the first row. Binomial probabilities are provided for n = 2, 3, . . . , 20 ; p = 0.05, 0.10, . . . , 0.50.

TABLE 1 Exact binomial probabilities $Pr(X = k) = \binom{n}{k} p^k q^{n-k}$

n	*k*	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0313
	1	.2036	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1563
	2	.0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0000	.0004	.0022	.0064	.0146	.0283	.0488	.0768	.1128	.1563
	5	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0313

EXAMPLE 4.28

Infectious Disease Evaluate the probability of 2 lymphocytes out of 10 white blood cells if the probability of any one cell being a lymphocyte is .2.

Solution: Refer to Table 1 with $n = 10$, $k = 2$, $p = .20$. The appropriate probability, given in the $k = 2$ row and $p = .20$ column under $n = 10$, is .3020.

EXAMPLE 4.29

Pulmonary Disease An investigator notices that children develop chronic bronchitis in the first year of life in 3 of 20 households in which both parents have chronic bronchitis, as compared with the national incidence of chronic bronchitis, which is 5% in the first year of life. Is this difference "real," or can it be attributed to chance? Specifically, how likely are infants in at least 3 of 20 households to develop chronic bronchitis if the probability of developing disease in any one household is .05?

Solution: Suppose the underlying rate of disease in the offspring is .05. Under this assumption, the number of households in which the infants develop chronic bronchitis will follow a binomial distribution with parameters $n = 20$, $p = .05$. Thus, among 20 households the probability of observing k with bronchitic children is given by

$$\binom{20}{k}(.05)^k(.95)^{20-k},\ k = 0, 1, \ldots, 20$$

The question is: What is the probability of observing at least 3 households with a bronchitic child? The answer is

$$Pr(X \geq 3) = \sum_{k=3}^{20}\binom{20}{k}(.05)^k(.95)^{20-k} = 1 - \sum_{k=0}^{2}\binom{20}{k}(.05)^k(.95)^{20-k}$$

These three probabilities in the latter sum can be evaluated using the binomial table (Table 1). Refer to $n = 20$, $p = .05$, and note that $Pr(X = 0) = .3585$, $Pr(X = 1) = .3774$, $Pr(X = 2) = .1887$. Thus,

$$Pr(X \geq 3) = 1 - (.3585 + .3774 + .1887) = .0754$$

Problems

4.9
4.10
4.14
4.17
4.33 - 4.37

Exercises-Binomial Distribution

Exercise (1)

The probability that a patient in KAUH recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, then answer the following:

(a) Find the probability that:

(i) at least 10 patients survive? Ans: 0.0338.

(ii) exactly 5 patients survive? Ans: 0.1859.

(iii) from 3 to 8 patients survive? Ans: 0.8779.

(b) Find the mean and standard deviation for the number of people that survive?

Ans: $\mu = 6$ people and $\sigma = 1.897$ people.

Notation: *Use Binomial Table*

Exercise (2)

It is known that 60% of children in Jordan inoculated with the Influenza vaccine are protected from flu in Winter. If 5 children are inoculated, then answer the following:

(a) Find the probability that:

(i) none contracts the disease? Ans: 0.0778 .

(ii) fewer than 2 contracts the disease? Ans: 0.3370 .

(iii) more than 3 contracts the disease? Ans: 0.0870 .

(b) Find the mean and standard deviation for the number of children contracts the disease? Ans: $\mu = 3$ children and $\sigma = 1.095$ children.