Probability and Statstics

Dr. Faisal Bukhari
Associate Professor
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab

Textbooks

- **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

- **Elementary Statistics: Picturing the World,** 6^{th} Edition, Ron Larson and Betsy Farber

- **Elementary Statistics**, 13^{th} Edition, Mario F. Triola

Reference books

- **Probability Demystified**, Allan G. Bluman
- **Schaum's Outline of Probability and Statistics**
- **MATLAB Primer**, Seventh Edition
- **MATLAB Demystified** by McMahon, David

Reference books

- **Probability Demystified**, Allan G. Bluman
- **Schaum's Outline of Probability and Statistics**
- **MATLAB Primer**, Seventh Edition
- **MATLAB Demystified** by McMahon, David

References

Readings for these lecture notes:

- **Probability & Statistics for Engineers & Scientists**, Ninth edition, Ronald E. Walpole, Raymond H. Myer

- **Elementary Statistics**, 10th Edition, Mario F. Triola

- **Probability Demystified**, Allan G. Bluman

These notes contain material from the above three books.

“A goal is a dream with a deadline.”

— Napoleon Hill

Find the binomial probability distribution, when n = 6 and p = 0.6. Also find its mean and variance. Verify the results.

Solution:

$$\mathbf{b(x; n, p) = c_x^n p^x q^{n-x}}, \quad \mathbf{x = 0, 1, 2, \ldots, n}$$

where $\mathbf{c_x^n = \frac{n!}{x!(n-x)!}}$

p = 0.60.

q = 1- p = 0.4

X = 0, 1 , 2, 3, 4, 5, 6

Probability Distribution

x	P(x)	xP(x)	x^2P(x)
0	0.0041	0	0
1	0.0369	0.0369	0.0369
2	0.1382	0.2764	0.5528
3	0.2765	0.8295	2.4885
4	0.3110	1.2440	4.9760
5	0.1866	0.9330	4.6650
6	0.0467	0.2802	1.6812
	$\sum_{i=1}^{6} P_i =$ **1.0000**	$\sum xP(x) = 3.6$	$\sum x^2 P(x) =$ 14.4004

Note:

1. $0 \leq P(x) \leq 1$

2. $\sum P(x) = 1$

$\mu = E(x) = \sum xP(x) = 3.6000$

$E(x^2) = \sum x^2 P(x) = 14.4$

$\sigma^2 = E(x^2) - [E(x)]^2 = 14.4004 \quad - (3.6000)^2$

$\sigma^2 = 1.4400$

Verification:

$\mu = np = (6)(0.6) = 3.6$

$\sigma^2 = npq = (6)(0.6)(0.4) = 1.4400$

Hypergeometric Distribution

- **Hypergeometric Distribution** If we sample from a small finite population **without replacement**, the binomial distribution should not be used because the events are **not independent.**

- If sampling is done **without replacement** and the outcomes belong to **one of two types**, we can use the **hypergeometric distribution**

Hypergeometric Distribution

- The simplest way to view the **distinction** between the binomial distribution and the hypergeometric distribution is to note the **way the sampling is done**.

- The types of applications for the **hypergeometric** are very similar to those for the binomial distribution. We are interested in computing probabilities for the number of observations that **fall into a particular category**.

Applications

- Applications for the hypergeometric distribution are found in many areas, with heavy use in **acceptance sampling, electronic testing, and quality assurance**. Obviously, in many of these fields, **testing is done at the expense of the item being tested**.

- That is, the item is **destroyed** and hence **cannot be replaced** in the sample

Hypergeometric Distribution [1]

A **hypergeometric experiment** has the following properties:

1. Each trial of an experiment results in **an outcome** that can be classified into one of the two categories **success or failure**.
2. The successive trials are **dependent.**
3. The probability of success **changes** from trial to trial.
4. The experiment is repeated **a fixed number** of times.

Hypergeometric Distribution [2]

- The number **X of successes** of a hypergeometric experiment is called a **hypergeometric random variable**.

- Accordingly, the probability distribution of the hypergeometric variable is called the **hypergeometric distribution**, and its values are denoted by $h(x; N, n, k)$, since they depend on the **number of successes k** in the set N from which we **select n items**.

Hypergeometric Distribution [3]

This distribution is the case of sampling **without replacement**. The formula to calculate probabilities is given by

$$P(X = x) = h(x;\ N,\ n,\ k)$$

$$= ({}_{k}C_{x})\ ({}_{N-k}C_{n-x})\ /\ ({}_{N}C_{n}),$$

$$\max\{0,\ n-(N-k)\} \leq x \leq \min\{n,\ k\}$$

OR

$h(x; N, n, k) = \dfrac{({}_{k}C_{x})\ ({}_{N-k}C_{n-x})}{{}_{N}C_{n}}$, $\max\{0, n - (N-k)\} \leq x \leq \min\{n, k\}$

Hypergeometric Distribution [4]

- It has **three** parameters i.e., N, n, and k
- **N**: The number of items in the **population**
- **k**: The number of items in the **population** that are classified as **successes.**
- **n**: The number of items in the sample
- **x**: The number of items in the **sample** that are **classified as successes**.

Hypergeometric Distribution [5]

Example1: Suppose we randomly select **5** cards **without replacement** from an ordinary deck of playing cards. What is the probability of getting exactly **2 red cards**?

Solution: This is a hypergeometric experiment in which we know the following:

N = 52; since there are 52 cards in a deck.

k = 26; since there are 26 red cards in a deck.

n = 5; since we randomly select 5 cards from the deck.

x = 2; since 2 of the cards we select are red.

$h(x; N, n, k) = ({}_{k}C_{x})({}_{N-k}C_{n-x}) / {}_{N}C_{n}$

$h(2; 52, 5, 26) = ({}_{26}C_{2})({}_{26}C_{3}) / ({}_{52}C_{5})$

$h(2; 52, 5, 26) = (325)\ (2600) / (2,598,960\ ) = 0.32513$

Hypergeometric Distribution [6]

Example: A committee of **4** people is selected at random without replacement from a group of **6** men and **4** women. Find the probability that the committee consists of **2** men and **2** women.

Solution:

$P(X = x) = h(x; N, n, k) = ({}_kC_x)({}_{N-k}C_{n-x})/({}_NC_n)$, $\max\{0, n-(N-k)\} \le x \le \min\{n, k\}$

$N = 10$; $k = 6$; $n = 4$; $x = 2$ (let x denotes number of men)

$h(x; N, n, k) = ({}_6C_x)({}_4C_{4-x})/({}_{10}C_4)$

$h(2; 10, 4, 6) = ({}_6C_2)({}_4C_2)/({}_{10}C_4)$

$h(2; 10, 4, 6) = (15)(6)/(210) = 0.429$

Hypergeometric Distribution [6]

Example: A lot of **12 oxygen tanks** contains **3** defective ones. If **4 tanks** are randomly selected and tested, find the probability that exactly **one will be defective**.

Solution:

$P(X = x) = h(x; N, n, k) = ({}_kC_x)({}_{N-k}C_{n-x})/({}_NC_n)$, $\max\{0, n-(N-k)\} \leq x \leq \min\{n, k\}$

OR

$h(x; N, n, k) = \dfrac{({}_kC_x)\,({}_{N-k}C_{n-x})}{{}_NC_n}$, $\max\{0, n - (N-k)\} \leq x \leq \min\{n, k\}$

$N = 12$; $k = 3$; $n = 4$; $x = 1$ (let x denotes defective tanks)

$P(X = 1) = ({}_3C_1)({}_9C_3)/({}_{12}C_4)$

$P(X = 1) = (3)(84)/(495) = 0.509$

Hypergeometric Distribution [1]

Example: In a box of **12** shirts there are **5** defective ones. If **5** shirts are sold at random, find the probability that exactly two are defective.

Solution:

$h(x; N, n, k) = \frac{({}_kC_x)\ ({}_{N-k}C_{n-x})}{{}_NC_n}$, $\max\{0, n-(N-k)\} \leq x \leq \min\{n, k\}$

Defective shirts	Non-detective shirts	Total
5	**7**	**12**

N = 12, k = 5, n = 5, and x = 2

Let **X** denotes the number of **defective shirts**

$P(X = 2) = ({}_5C_2)({}_7C_3)/{}_{12}C_5 = 0.442$

Hypergeometric Distribution [2]

Example: In a fitness club of **18** members, **10** prefer the exercise bicycle and **8** prefer the aerobic stepper. If **6** members are selected at random, find the probability that exactly **3** use the bicycle.

Solution:

$$h(x; N, n, k) = \frac{({}_kC_x)\,({}_{N-k}C_{n-x})}{{}_NC_n},\ \max\{0, n-(N-k)\} \le x \le \min\{n, k\}$$

Exercise Bicycle	Aerobic Stepper	Total
10	**8**	**18**

N = 18, k = 10, n = 6, and x = 3

Let **X** denotes the number of **bicycles**

$P(X=3) = ({}_{10}C_3)({}_8C_3)/{}_{18}C_6 = 0.362$

Hypergeometric Distribution [3]

Example: In a shipment of **10** lawn chairs, **6** are brown and **4** are blue. If **3** chairs are sold at random, find the probability that all are **brown**.

Solution:

$$h(x; N, n, k) = \frac{({}_kC_x)\,({}_{N-k}C_{n-x})}{{}_NC_n},\ \max\{0, n-(N-k)\} \le x \le \min\{n, k\}$$

Brown	Blue	Total
6	4	10

N = 10, k = **6**, n = 3, and x = **3**

Let **X** denotes the number of **brown** chairs

$P(X = 3) = ({}_6C_3)({}_4C_0)/{}_{10}C_3 = 0.167$

Hypergeometric Distribution [4]

Example: A class consists of 5 women and 4 men. If a committee of 3 people is selected at random, find the probability that all 3 are women.

Solution:

$$h(x; N, n, k) = \frac{({}_kC_x)\ ({}_{N-k}C_{n-x})}{{}_NC_n},\ \max\{0, n-(N-k)\} \leq x \leq \min\{n, k\}$$

Men	Women	Total
4	5	9

N = 9, k = 5, n = 3, and x = 3

Let **X** denotes the number of **women**

$P(X=3) = ({}_5C_3)({}_4C_0)/{}_9C_3 = 0.119$

Hypergeometric Distribution [5]

Example: A box contains **3 red** balls and **3 white balls**. If **two balls** are selected at random without replacement, find the probability that both are **red**.

Solution:

$$h(x; N, n, k) = \frac{({}_kC_x)\,({}_{N-k}C_{n-x})}{{}_NC_n},\ \max\{0, n-(N-k)\} \leq x \leq \min\{n, k\}$$

Red	White	Total
3	3	6

N = 6, k = 3, n = 2, and x = 2

Let **X** denotes the number of **red balls**

$P(X = 2) = ({}_3C_2)({}_3C_0) / {}_6C_2 = 0.2$

Hypergeometric Distribution

Example: Suppose that a shipment contains 5 defective items and 10 non defective items. If 7 items are selected at random without replacement, what is the probability that at least 3 defective items will be obtained?

Solution:

Defective	Non defective	Total
5	10	15

Here N = 15, n =7

k = 5 (defective items in the population)

Let X denotes number of defective items

$P(X \geq 3) = 1 - P(X < 3) = 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\}$

$$h(x; N, n, k) = \frac{({}_{k}C_{x})\,({}_{N-k}C_{n-x})}{{}_{N}C_{n}},$$

$$\max\{0, n - (N-k)\} \leq x \leq \min\{n, k\}$$

$$\therefore \max\{0, n - (N-k)\} = \max\{0, 7 - (15 - 5)\} = 0$$

$$P(0) = \frac{({}^{5}_{0}c)({}^{10}_{7}c)}{({}^{15}_{7}c)} = 0.0186$$

$$P(1) = \frac{({}^{5}_{1}c)({}^{10}_{6}c)}{({}^{15}_{7}c)} = 0.1631$$

$$P(2) = \frac{({}^{5}_{2}c)({}^{10}_{5}c)}{({}^{15}_{7}c)} = 0.3916$$

$P(X \geq 3) = 1 – P(X<3) = 1 – (0.0186 + 0.1631 + 0.3916)$

$= 0.4267$