# MAT 3103: Computational Statistics and Probability

## Chapter 5: Discrete Distribution


**Probability Distribution:**

A probability distribution is a list of all of the possible outcomes of a random variable along with their corresponding probability values.

Things happen all the time: dice are rolled, it rains, buses arrive. The fact is, the specific outcomes are certain: the dice came up 3 and 4, there was half an inch of rain today, the bus took 3 minutes to arrive. Before, we can only talk about how likely the outcomes are. Probability distributions describe what we think the probability of each outcome is, which is sometimes more interesting to know than simply which single outcome is most likely. A probability distribution is a summary of probabilities for the values of a random variable.

**Discrete probability distribution:**

It describes the probability of occurrence of each value of a discrete random variable. Discrete probability distributions are used in machine learning, most notably in the modeling of binary and multi-class classification problems, but also in evaluating the performance for binary classification models, such as the calculation of confidence intervals, and in the modeling of the distribution of words in text for natural language processing. Knowledge of discrete probability distributions is also required in the choice of activation functions in the output layer of deep learning neural networks for classification tasks and selecting an appropriate loss function.

Under discrete probability distribution, we will learn

**(a) Binomial distribution**

**(b) Poisson distribution**

**(c) Geometric distribution**

**Binomial distribution:**

The binomial distribution is the discrete probability distribution of the number of successes in a sequence of *n* independent yes/no experiments, each of which yields success with probability *p*. Binomial process is a random process in which:

- The process is performed under the same conditions for fixed and finite number of trials.
- Each trial is independent of other trials.
- Each trial has two mutually exclusive outcomes.
- The probability of success, *p*, remains constant from trial to trial (so the probability of failure, *q*, where, $q = 1-p$).

In general, if the random variable *X* follows the binomial distribution with parameters *n* and *p*, we write $X \sim B(n, p)$. The probability of getting exactly *x* successes in *n* trials is given by the probability mass function:

$$P(X = x) = {}^{n}C_{x} p^{x} (1-p)^{n-x}; \; x = 0, 1, 2, \ldots, n.$$

Where, $p$ = the probability of a success.

$n$ = the number of trials.

$x$ = the number of success in n trials.

Binomial distribution is often used in quality control of items manufactured by a production line when each item is classified as either defective or non-defective. To estimate the chance of pass or fail in the examination as well as death or survive of the patients, Binomial distribution is widely used.

**Note**: For binomial distribution: $E(X) = np$, $V(X) = npq$; mean > variance. The terms 'success' and 'failure' are label and might be misleading. For example, counting the number of defective items produced by a machine might be thought of as counting successes if you are looking for defective items!

**Example 5.1:** A fair coin is tossed 8 times. Find the probability that exactly 3 tails will befall.

**Solution:** Let, *X* = number of tails. Here, $p = q = 0.5$, and $n = 8$. Then,

$P(X = x) = n_{c_x}\, p^x\, q^{n-x} = 8_{c_x}(0.5)^x(0.5)^{8-x}$; $x = 0, 1, 2, \ldots, 8$. Therefore,

$P(X = 3) = 8_{c_3}(0.5)^3(0.5)^{8-3} = 0.219$.

**Example 5.2:** A student randomly guesses 5 questions; each question has 5 possible choices. Find the probability that he guesses exactly 3 correctly.

**Solution:** Let, $X$ = number of correct guess. Here, $p = \frac{1}{5} = 0.2$, $q = 1 - p = 0.8$, and $n = 5$. Then,

$P(X = x) = n_{c_x}\ p^x\ q^{n-x} = 5_{c_x}(0.2)^x(0.8)^{5-x}$; $x = 0, 1, 2, \ldots, 5$. Therefore,

$P(X = 3) = 5_{c_3}(0.2)^3(0.8)^{5-3} = 0.05$.

**Example 5.3:** Apollo hospital records show that 75% of patients suffering from kidney disease die of it. What is the probability that out of 6 randomly selected patients, 4 will recover?

**Solution:** Let, $X$ = number of recovered patients. Here, $p = 0.25$, $q = 1 - p = 0.75$, and $n = 6$. Then, $P(X = x) = n_{c_x}\ p^x\ q^{n-x} = 6_{c_x}(0.25)^x(0.75)^{6-x}$; $x = 0, 1, 2, \ldots, 6$. Therefore,

$P(X = 4) = 6_{c_4}(0.25)^4(0.75)^{6-4} = 0.033$.

**Example 5.4:** A survey found that 30% AIUB students earn money from part-time job. If 5 students are selected at random, find the probability that at least 3 of them have part-time job.

**Solution:** Let, $X$ = no. of students having part-time job, and $p = 0.3$, $q = 1 - p = 0.7$, and $n = 5$. Then, $P(X = x) = n_{c_x}\ p^x\ q^{n-x} = 5_{c_x}(0.3)^x(0.7)^{5-x}$; $x = 0, 1, 2, \ldots, 5$. Therefore,

$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$

$= 5_{c_3}(0.3)^3(0.7)^{5-3} + 5_{c_4}(0.3)^4(0.7)^{5-4} + 5_{c_5}(0.3)^5(0.7)^{5-5}$

$= 0.162$.

**Example 5.5:** A group of students at AIUB were equally trained to develop a program. All of them had 50% chance to develop the program successfully. Ten randomly selected students were asked to develop the program separately. Find the probability that out of 10, (a) none becomes successful, (b) 5 become successful, (c) at best 1 becomes successful, and (d) at least 2 become successful.

**Solution:** Let, $X$ = number of successful students. Here, $p = q = 0.5$, and $n = 10$. Then,

$P(X = x) = n_{c_x}\ p^x\ q^{n-x} = 10_{c_x}(0.5)^x(0.5)^{10-x} = 10_{c_x}(0.5)^{10}$; $x = 0, 1, 2, \ldots, 10$. So,

(a) $P(X = 0) = 10_{c_0}\ (0.5)^{10} = 0.00098$.

(b) $P(X = 5) = 10_{c_5}\ (0.5)^{10} = 0.2461$.

(c) $P(X \leq 1) = [P(X = 0) + P(X = 1)] = [10_{c_0}\ (0.5)^{10} + 10_{c_1}\ (0.5)^{10}] = 0.0107$.

(d) $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - 0.0107 = 0.9893$.

**MATLAB code**

1. Compute the pdf of the Binomial distribution with 10 trials and the probability of success 0.5.

```
x = 0:10;
y = binopdf (x,10,0.5);
```

2. Eighty five percent devices of a workshop work properly. One day 20 devices are selected at random. Find the probability that, out of 20 devices 5 work properly

```
binopdf (5,20,0.85)
```

**Poisson distribution:**

A Poisson distribution is used to estimate how likely it is that something will happen "$X$" number of times within a specified period of time. A Poisson random variable is the number of successes that result from a Poisson experiment. The probability distribution of a Poisson random variable is called a Poisson distribution, defined as:

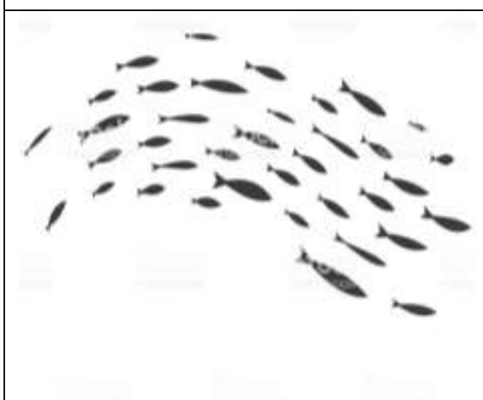
$$P(X = x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!};\ x = 0, 1, 2, \dots ;\ \lambda > 0.$$

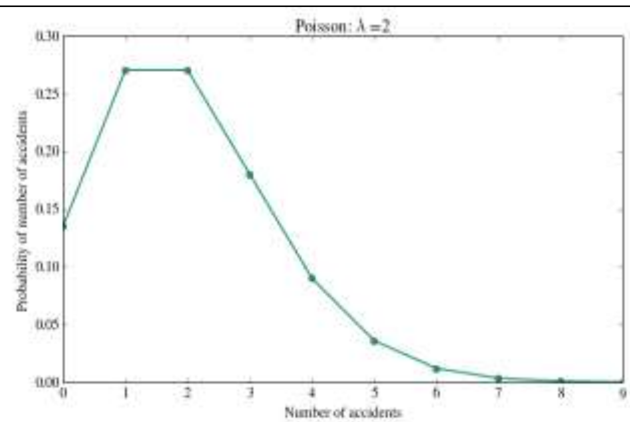
Here, $\lambda$ = The mean number of successes that occur in a specified region.

$X$ = The actual number of successes that occur in a specified region.

The mean = variance = $\lambda$ of the distribution. It is used, when certain conditions are met, as a probability distribution, where the possible number of discrete occurrences is much larger than the average number of occurrences in a given interval of time or space.

French word "Poisson" means "fish", French mathematician Siméon-Denis Poisson developed Poisson distribution.





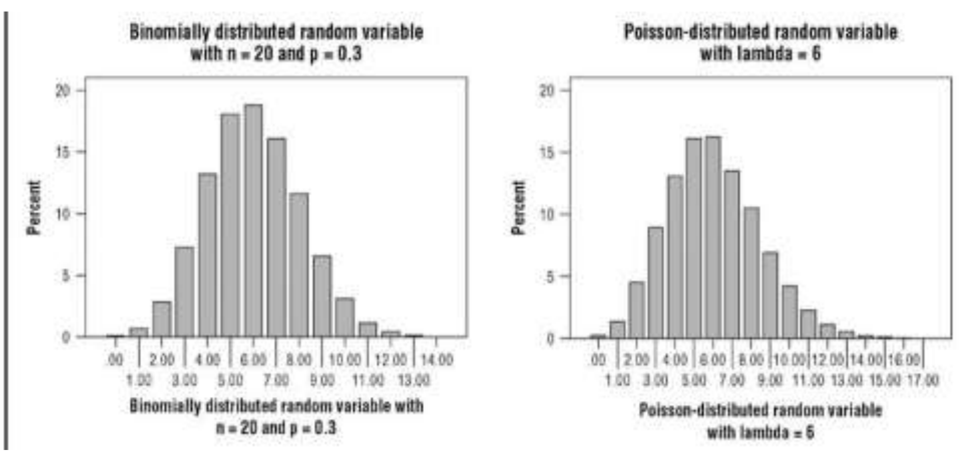
**Historical background of Poisson distribution:**

Like many statistical tools and probability metrics, the Poisson Distribution was originally applied to the world of gambling. In 1830, French mathematician Siméon Denis Poisson developed the distribution to indicate the low to high spread of the probable number of times that a gambler would win at a gambling game – such as baccarat – within a large number of times that the game was played.

The wide range of possible applications of Poisson's statistical tool became evident several years later, during World War II, when a British statistician used it to analyze bomb hits in the city of London. R.D. Clarke refined the Poisson Distribution as a statistical model and worked to reassure the British government that the German bombs fell randomly, or purely by chance, and that its enemies lacked sufficient information to be targeting certain areas of the city. Since then, the Poisson Distribution's been applied across a wide range of fields of study, including medicine, astronomy, business, and sports.

**Poisson distribution vs Binomial distribution:**

Poisson distribution is a limiting case of the binomial distribution under the terms: (a) $n \to \propto$ (very large), (b) $p \to 0$ (very small), and (c) $np = \lambda$; $\lambda > 0$. Poisson distribution has only one parameter $\lambda$ whereas binomial distribution has two parameters, *n* and *p*. Especially for very large values of *n* and very small values of *p* Binomial variable transform to Poisson variable.




The applications of Poisson distribution in real life is very long. The distribution is suitable in observing

i) The number of bankruptcies that are filed in a month,

ii) The number of arrivals at a car wash in one hour,

iii) The number of network failures per day, etc.

**Example 5.6:** Vehicles pass through Kuril junction at an average rate of 300 per hour. Find the probability that no vehicle passes in a given minute.

**Solution:** Let, $X$ = no. of vehicles pass in a given minute. Given, $\lambda$ = 300/60 = 5. Then,

$$P(X = x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} = \frac{e^{-5}\,5^{x}}{x!};\ x \geq 0. \text{ Hence, } P(X = 0) = \frac{e^{-5}\,5^{0}}{0!} = e^{-5}.$$

**Example 5.7:** RFL assembles electric motors. The probability that a motor is defective is 0.01. What is the probability that a random sample of 200 motors will contain exactly 4 defective motors?

**Solution:** Let, $X$ = number of defective motors. Given, $\lambda = 200 \times 0.01 = 2$. Then,

$$\mathrm{P}(X = \mathrm{x}) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} = \frac{e^{-2}\,2^{x}}{x!};\ x \geq 0. \text{ Hence, } \mathrm{P}(X = 4) = \frac{e^{-2}\,2^{4}}{4!} = 0.090.$$

**Example 5.8:** Electricity fails according to Poisson distribution with average of 4 failures per 20 weeks in AIUB. Find the probability that there will be no electricity failure in a specific week.

**Solution:** Let, $X$ = no. of electricity failure in a specific week. Given, λ = 4/20 = 0.20. Then,

$$P(X = x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} = \frac{e^{-0.20}(0.20)^{x}}{x!};\ x \geq 0. \text{ Hence, } P(X = 0) = \frac{e^{-0.20}(0.20)^{0}}{0!} = e^{-0.20}.$$

**Example 5.9:** The average number of signals sent from Kamalapur railway station, not reaching properly to Noakhali railway station, is 3 per day. Find the probability that on a particular day, the number of signals not reaching properly is (a) at best 2, and (b) at least 3.

**Solution:** Let, $X$ = no. of signals not reaching properly. Given, λ = 3. Then,

$$P(X = \mathrm{x}) = \frac{e^{-\lambda}\,\lambda^{x}}{x!} = \frac{e^{-3}\,3^{x}}{x!};\ x \geq 0.$$

(a) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \mathrm{e}^{-3}\left[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right] = 0.42319$

(b) $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - 0.42319 = 0.57681$

---

**MATLAB code**

1. Compute the pdf of the Poisson distribution with λ = 4.

```
x = 0:10;
y = poisspdf (x, 4);
```

2. In the computer hard disk manufacturing process, flaws occur randomly. Assuming that on average a 4 GB hard disk has two flaws, compute the probability that a disk has no flaws.

```
poisspdf (0, 2)
```

---

**Geometric Distribution:**

In case of binomial distribution, we observe x successes in a random experiment when the experiment is repeated n times under homogeneous conditions. In such an experiment the researcher may be interested in observing the incidence of first success ($X = x$, $x = 1, 2, \ldots,$) or the number of failures before the first success ($Y = X - 1 = y$; $y = 0,1, 2, \ldots$). It is not known that how many trials are needed to observe the first success, or after how many failures will occur before the first success. The probability of this type of uncertainty can evaluated by probability model known as **geometric probability distribution.** The probability function of the above-mentioned random variable is given

$$P\,(X = k) = p\,(1 - p)^{k-1}\,,\ k = 1,2,3,\ldots \quad \text{or}$$

$$P\,(X = k) = p\,(1 - p)^{k}\,,\ k = 0,1,2,3,\ldots$$

In both cases $p$ is the probability of success, ($0 < p < 1$).

The mean and variance of $X$ are $E\,(X) = \frac{q}{p}$ and $V\,(X) = \frac{q}{p^2}$, respectively.

The distribution is suitable in observing (i) Number of defective items in a lot, (ii) Number of faded signals in a communication system, (iii) Number of missed calls in a telephone system.

**Example 5.10:** In a telephone system 80% are voice calls. Find the probability that the 6th call will be a data call.

**Solution:** Let $X$ be the number of data call. Given $p = P$ [data call] = 0.2, $q = 1 - p = 0.8$.

$P\,(X = 6) = p\,(1 - p\,)^{6-1} = 0.2\,(0.8)^5 = 0.0655$.

**Example 5.11:** A programmer in 70% cases becomes successful in developing program. Find the probability that in developing program at 8th attempt he / she fails.

**Solution:** Let $X$ be the number of times a programmer fails. Given $p = P$ [Failure] = 0.3.

$P\,(X = 8) = p\,(1 - p\,)^{8-1} = 0.3\,(0.7)^7 = 0.0247$.

---

**MATLAB code**

1. Compute the pdf of the geometric distribution with the probability of success 0.25.

```
x = 0:20;
y = geopdf (x, 0.25);
```

2. Suppose you toss a fair coin repeatedly, and a "success" occurs when the coin lands with heads facing up. What is the probability of observing exactly three tails ("failures") before tossing a head? [ The probability of success (tossing a heads) p in any given trial is 0.5 ]

geopdf(3, 0.5)

## **Exercise 5**

5.1. Suppose $X$ is a binomial variate with parameters $n = 4$ and $p = 0.45$. Find (a) $P[X = 2]$,

(b) $P[X > 3]$, (c) $P[X < 2]$.

5.2. For a binomial distribution, mean is 2 and variance is 1. Estimate $n$, $p$ and $q$.

5.3. 80% electric circuits in AIUB work properly. One day 10 circuits were selected randomly. What was the expected number of devices working properly along its variance?

5.4. In a class 40% are female students. Find the probability that the 9th student enter in the class will be a male.

5.5. 70% of all business startups in the IT industry report that they generate a profit in their first year. If a sample of 10 new IT business startups is selected, find the probability that exactly 7 will generate a profit in their first year.

5.6. Two percent handsets produced by Walton are usually found defective. Walton produces 300 sets per day. Find the probability that there will be (a) 4 defective sets (b) at best 3 defective sets (c) 2 to 5 defective sets in a day's production.

5.7. A box of candies has many different colors in it. There is a 15% chance of getting a pink candy. What is the probability that at least 2 candies in a box are pink out of 8?

5.8. In a typical El Clasico game against arch-rival Barcelona, Real Madrid coach Zidane can expect 2 injuries on average. Find the probability that Madrid will have at most 1 injury in the game.

5.9. Let $X$ be the number of typos on a printed page with a mean of 4 typos per page. What is the probability that a randomly selected page has a) at least two typo b) at best 3 typo c) exactly 5 typos on it?

5.10. The number of constructions related accidents per working week follows Poisson distribution with mean 1. Find the probability that in a particular week, there will be (a) less than 2 accidents, and (b) more than 1 accidents.

5.11. 20% devices in a laboratory do not work properly. One day 10 devices are selected randomly. Find the probability that, out of the 10 devices, (a) at least 3 work properly, (b) none work properly and (c) 2 to 5 work properly. Find the expected number of devices which work properly.

5.12. 2.5% signals sent from a server do not reach to its goal properly. A station sends 150 signal per week. Find the probability that, (a) at least 1 do not reach properly, and (b) at best 2 do not reach properly.

5.13. 30% students in AIUB got $A^+$ grade in Math course last semester. Five students are randomly selected. Find the probability that out of the 5 students, (a) 2 got $A^+$, (b) all got $A^+$ and (c) at least 1 got $A^+$. Find the variance of no of

**Sample MCQs**

1. Seventy percent patients of general practitioner are children. Find the probability that one day the 7th patient is a child.

a) 0.0005 b) 0.0050 c) 0.0450 d) 0.0003

2. For a binomial distribution, mean is 2 and variance is 1. Estimate the value of *p* and *n*.

a) 0.45, 5 b) 0.50, 4 c) 0.04, 4 d) 0.30, 6

3. Eighty percent devices of a workshop work properly. One day 10 devices are selected at random. What is the probability that at best 2 works properly?

a) $8.7\times 10^{-5}$ b) $8.7\times 10^{-3}$ c) $7.8\times 10^{-3}$ d) $8.7\times 10^{-5}$

4. Two percent mobile sets produced by a company are usually found defective. The company produces 200 sets per day. Find the probability that in a day's production there will be 4 defective sets.

a) 0.7505 b) 0.1350 c) 0.19537 d) 0.2343

5. 60% streetlights in Dhaka city work properly. One day 25 streetlights were selected randomly. What was the expected number of streetlights working properly along its variance?

a) 45, 5 b) 15, 6 c) 20, 4 d) 30, 6

6. The average number of signals sent from a station is 3 per day which do not reach properly to another station. Find the probability that no signal which are sent in a day but not reached properly.

a) 0.7008 b) 0.0351 c) 0.0498 d) 0.1343