Topic 9.3
Binomial Distribution, Poisson Distribution

TMA1201 Discrete Structures & Probability
Faculty of Computing & Informatics
Multimedia University

What you will learn in this lecture:

- The Bernoulli distribution
- The Binomial distribution
  - Parameters to define Binomial
  - Mean and variance of Binomial
  - Probabilities of Binomial experiments
- The Poisson distribution
  - Poisson distribution
  - Poisson distribution as a limit
  - Poisson Distribution respective to time interval t

Bernoulli Distribution

- Arises when investigating proportions.
  - E.g., proportion of TMA1201 students who wear glasses.
  - Each student either wears or does not wear glasses (binary outcome).
- Let Y be the random variable for an individual outcome of a person in the population that we are interested in.
- There are two outcomes: $Y = 1$ and $Y = 0$. Generally speaking, these refer to "Success" and "Failure" respectively.
- The parameter p represents the unknown proportion of 1's occurring.

Bernoulli Distribution

- The probability distribution of Y is:

$Y = y$	1 (Success)	0 (Failure)
$P(Y = y)$	p	$1 - p$

Mean and Variance of Bernoulli

- We can find the mean of the Bernoulli distribution.
- The mean is colloquially defined as what is the average outcome of an event.

Mean = $1 \times p + 0 \times (1 - p) = p$

- We can also find the variance of the Bernoulli distribution.

Variance = $(1 - p)^2 \times p + (0 - p)^2 \times (1 - p) = p(1 - p)$

Binomial Distribution

- Suppose we take a sample of size *n* from the underlying population and look at the distribution of the number of successes.
- Total number of successes:

$$X = Y_1 + Y_2 + Y_3 + \ldots + Y_n$$

where all the Yi 's are independent of each other.

Mean and Variance of Binomial

- Combining random variables gives the mean of *X* as

$$\mu_X = \mu_{Y_1} + \mu_{Y_2} + \cdots + \mu_{Y_n}$$

- The variance of *X* is given as

$$\sigma_X^2 = \sigma_{Y_1}^2 + \sigma_{Y_2}^2 + \cdots + \sigma_{Y_n}^2$$

- Since all the *Y*'s come from the same population then

$$\mu_{Y_1} = \mu_{Y_2} = \cdots = \mu_{Y_n} = \mathrm{p}$$

$$\sigma_{Y_1}^2 = \sigma_{Y_2}^2 = \cdots = \sigma_{Y_n}^2 = \mathrm{p}(1\text{-}\mathrm{p})$$

Mean and Variance of Binomial

- Hence the mean of X is

$$\mu_X = \mu_{Y_1} + \mu_{Y_2} + \cdots + \mu_{Y_n} = p + p + \cdots + p = n\mathrm{p}$$

- The variance of X is

$$\sigma_X^2 = \sigma_{Y_1}^2 + \sigma_{Y_2}^2 + \cdots + \sigma_{Y_n}^2 = \sigma^2 + \sigma^2 + \cdots + \sigma^2 = n\sigma^2$$
$$\sigma_X^2 = np(1 - p)$$

Three Conditions Associated with Binomial Distribution

- Outcome is binary.
- We have n independent trials.

  Probability of success p must stay constant.

Note About These Conditions

Outcome is binary

- There may be more than two possible outcomes, as long as the outcomes can be combined into two subsets.
- One subset is success, the other is failure.
- Examples
  - Rolling a 4 is a “success”; any other numbers is a “failure”.
  - Having blue eyes is a “success”, any other eye colors is a “failure”.

Note About These Conditions

Probability of success must stay constant

- Sampling without replacement from a small population does not produce a binomial random variable.

Example

Suppose a class consists of 10 boys and 10 girls. Five students are randomly selected to be in a play. Let X = the number of girls selected.

This is not binomial because each time someone is removed from the sample, the probability that a girl is selected changes.

Probability of *x* Successes

- The probability of *x* successes, where *x* takes the values 0 to *n*, is given by the formula:

$$\boxed{\begin{aligned} \mathrm{b}(\mathrm{x};\ \mathrm{n}, p) = \ \mathrm{P}(\mathrm{X} = x) = \ {}^{\mathrm{n}}\mathrm{C}_x p^x (1-p)^{n-x}, \\ x = 0, 1, 2, \dots, n \end{aligned}}$$

A Quick Recap on Binomial Distribution

- Binomial distribution is a commonly used discrete probability distribution since many statistical problems deal with the situations referred to as repeated trials.
- Those experiments that possess the following properties are called binomial experiments:

1) The experiment consists of a sequence of n identical trials.
2) Two outcomes, success and failure, are possible on each trial.
3) The probability of a success, denoted by π, does not change from trial to trial.
4) The trials are independent.

Example 1

- A software company ABC is concerned about a low retention rate for employees. On the basis of past experience, management has seen a turnover of 10% of the hourly employees annually. Thus, for any hourly employees chosen at random, management estimates a probability of 0.1 that the person will not be with the company next year. Choosing 3 hourly employees at random, what is the probability that one of them will leave the company this year?
- Solution

  Let X be the number among the 3 chosen hourly employees that will leave the company. Hence $X \sim Bin(n = 3, p = 0.1)$ and $P(X = 1) = {}^3C_1 (0.1)^1 (0.9)^{3-1} = \cdots = 0.243$

Example 2

- The probability that a patient recovers from SARS is 0.4. If 15 people are known to have contracted this disease, what is the probability that:

  (1) at least 13 survive? (2) at most 3 survive?

  (3) at most 13 die?

- Solution:

  Let X be the number among 15 SARS patients recover from the disease.

$$X \sim Bin(n = 15, p = 0.4).$$

1) $P(X \geq 13) = P(X = 13) + P(X = 14) + P(X = 15)$

$= {}^{15}C_{13}(0.4)^{13}(0.6)^{2} + {}^{15}C_{14}(0.4)^{14}(0.6)^{1} + {}^{15}C_{15}(0.4)^{15}(0.6)^{0}$

$= 0.0003$

Example 2 (Cont.)

- The probability that a patient recovers from SARS is 0.4. If 15 people are known to have contracted this disease, what is the probability that:

  (1) at least 13 survive? (2) at most 3 survive?

  (3) at most 13 die?

- Solution:

  2) $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0.0905$

  3) At most 13 die is equivalent to at least 2 survive

  $P(X \geq 2) = P(X = 2) + P(X = 3) + \ldots + P(X = 15) = 0.9948$, or

  $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 0.9948$

Poisson Distribution

- The Poisson distribution is a discrete probability distribution.
- Applies to occurrences of some event over a specific interval.
- The random variable X is the number of occurrences of the event in an interval.
- The interval can be time, distance, area, volume, or some similar unit.

Poisson Distribution

- The probability distribution of the Poisson random variable X, representing the number of outcomes occurring over an interval is

$$\boxed{\mathrm{p}(\mathrm{x};\ \mu) = \mathrm{P}(\mathrm{X} = x) = \frac{\mathrm{e}^{-\mu}\mu^{x}}{x!};\quad x = 0,1,2,\cdots;\mu > 0}$$

μ is the average number of occurrence in a given interval

- A short notation to designate that X has the Poisson distribution with parameter μ is $X \sim Poi(\mu)$

- If $X \sim Poi(\mu)$, then $\mu_X = \mu$ and $\sigma_X^2 = \mu$

Example 3

- Patients arrive at the emergency room of Putrajaya Hospital at the average rate of 6 per hour on weekend evenings. What is the probability of 4 arrivals in an hour during a weekend evening?
- Solution

  Let X be number of arrivals to the emergency room of Putrajaya Hospital in a weekend evening during one hour interval.

$$X \sim Poi(\mu = 6)$$

$$P(X = 4) = \frac{e^{-6} 6^4}{4!} = 0.1339$$

The Poisson Distribution as a Limit

- The rationale for using the Poisson distribution in many situations is provided by the following proposition.

- Suppose that in the binomial pmf b(x; n, p), we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then
$$b(x;\ n, p)\ \to\ p(x; \mu)$$

- According to this proposition, in any binomial experiment in which n is large and p is small, $b(x;\ n, p)\ \approx\ p(x; \mu)$, where . $\mu\ = np$.
- As a rule of thumb, this approximation can safely be applied if $n > 50$ and $np < 5$.

Example 4

If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is .005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors? At most three pages with errors?

With S denoting a page containing at least one error and F an error-free page, the number X of pages containing at least one error is a binomial rv with $n = 400$ and $p = .005$, so $np = 2$. We wish

$$P(X = 1) = b(1; 400, .005) \approx p(1; 2) = \frac{e^{-2}(2)^1}{1!} = .270671$$

The binomial value is $b(1; 400, .005) = .270669$, so the approximation is very good.

Similarly,

$$\begin{aligned} P(X \leq 3) &\approx \sum_{x=0}^{3} p(x, 2) = \sum_{x=0}^{3} e^{-2}\frac{2^x}{x!} \\ &= .135335 + .270671 + .270671 + .180447 \\ &= .8571 \end{aligned}$$

and this again is quite close to the binomial value $P(X \leq 3) = .8576$.

The Poisson Process

Poisson process is process satisfy the following properties:

1. The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region.
2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

Poisson Distribution respective to time interval t

- The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$p(x; \mu = \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \cdots$$

where λ is the average number of outcomes per unit time, distance, area, or volume and e =2.71828....

- Both the mean and the variance of the Poisson distribution p(x; λt) are λt.

Example 4

- Customers arrive randomly at a department store at an average rate of 3.4 per minute. Find the probability that

  1) No customer arrives in any particular minute.

  2) Two or more customers arrive in any particular minute.

  3) One or more customers arrive in any 30-second period.

- Solution

  Let X be number of customers arrived at a department store during a period of 1 minute. $X \sim Poi(\mu = 3.4)$

  1) $P(X = 0) = \frac{e^{-3.4} 3.4^0}{0!}$ $= 0.0334$

  2) $P(X \geq 2) = 1 - P(X \leq 1) =$ $= 0.8516$

  3) Let Y be number of customers arrived at a department store during a period of 30-second. $Y \sim Poi(\mu = \frac{3.4}{2} = 1.7)$,

  $P(Y \geq 1) = 1 - P(Y=0) =$ $= 0.8173$

Summary

- Binomial distribution is defined by two parameters n and p.
- The mean of the Binomial is np.
- The variance of the Binomial is np(1-p).
- Three conditions for Binomial
  1) Outcome is binary.
  2) We have n independent trials.
  3) Probability of success must stay constant.
- Poisson distribution is defined by the parameter μ.
- Both mean and variance of the Poisson is μ.
- The Poisson is used to find the probability of certain number of outcomes occurring over an interval.

Exercise 1

A report suggests that 75% of Asian children under 20 live with both parents. A random sample of 25 Asian children under 20 is selected, and X is the binomial random variable for the number of these 25 who live with both parents.

1) Define the parameters of the distribution of X.

2) Find $P(X = 15)$.

3) Find the probability that 11 or fewer live with both parents.

Exercise 2

A study indicates that 4% of American teenagers have tattoos. A sample of 30 teenagers is randomly selected. Let X be the number of teenagers who have tattoos.

(a) Define the parameters of the distribution of X.

(b) What values do X take on?

(c) Find the probability that exactly 3 teenagers have tattoos.

(d) Find the mean and variance of X.

Exercise 3

A clothing store has determined that 30% of the people who enter the store will make a purchase. Eight people enter the store during a one-hour period. Find

(a) the probability that exactly four people will make a purchase.

(b) the probability that at least one person will make a purchase.

(c) the mean and variance.

Exercise 4

The number of calls coming per minute into a hotel reservation counter is Poisson random variable with mean 3. Find

(a) the probability that no calls come in a given 1 minute period.

(b) the probability that no more than 8 calls come in any 2-minute period.

(c) the probability that exactly 2 calls in any 30-second period.

(d) the mean and variance.