Statistics and Probability CT:01

Last edited time	@January 16, 2024 11:43 PM
Created	@October 14, 2023 11:28 AM
Tags	NSTU Year 2 Term 2
Created by	B Borhan

Resources:

- **Random variable, Probability Mass function & Density function :** https://www.youtube.com/watch?v=kkmCFyxRdlI
- **Binomial & Poisson distribution :** https://www.youtube.com/watch?v=gbqB35_yoo0
- **Bernoulli Distribution :** https://www.youtube.com/watch?v=wwuhjBoaT_g
- **Moments in Statistics: Mean, Variance, Skewness & Kurtosis - Importance and Applications :** https://www.youtube.com/watch?v=T_1IJZxFuxw
- **Kurtosis :** https://www.youtube.com/watch?v=B4o3fl3JnG4 , https://www.youtube.com/watch?v=xc05vFDkw88
- **CDF :** https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/distribution-functions/cumulative-distribution-function.html#:~:text=The cumulative distribution function F(x) lies within 0 and,F (x) %3D 1.

Statistics

Basically use it for result finding.

Data Type:

- Ungroup Data
- Group Data

Random Experiment & Random Value:

কোন এক্সপেরিমেন্ট করার টাইমে প্রত্যেক ট্রায়ালে যদি ভিন্ন ভিন্ন আউটকাম আসে, সেইটা একটা র‍্যান্ডম এক্সপেরিমেন্ট। যেমন একটা কয়েন টসে কখন হেড, কখন টেইল পাওয়া যাবে, তাই এইটা একটা র‍্যান্ডম এক্সপেরিমেন্ট। এইরকম একটা র‍্যান্ডম এক্সপেরিমেন্টের প্রত্যেকটা আউটকামকে একটা নিউমেরিক ভ্যালু দিয়ে প্রকাশ করা যায়। এই প্রকাশের জন্য র‍্যান্ডম ভ্যারিয়েবল ব্যাবহার করা হয়। [Reference]

A random variable is a rule that assigns a numerical value to each outcome in a sample space. [Reference]

Types of Random Variable:

- **Discrete Random Variable :**

  A discrete random variable is one which may take an only a countable number of distinct values. [Ref]

  র‍্যান্ডম ভ্যারিয়েবলের পসিবল ভ্যালুগুলো গণনা করা গেলে, সেইটা discrete random variable বলে। [Ref]

  A coin is flipped twice and the random variable
  X is the number of heads. Then sample space
  S={HH, HT, TH, TT}
  X={0, 1, 2}

  - **Probability Mass Function (PMF) / Probability Function / Frequency Function/ Discrete Probability Function**

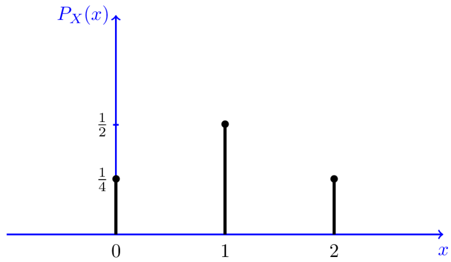
    Random Variable : $X = \{x_1, x_2, ...\}$

PMF. $f(x) = P(X = x)$

$f(x)$ satisfies the following properties:

- $f(x) >= 0$
- $\sum f(x) = 1$
- $f(x) = P(x)$

[Ref]



[Ref] PMF Graph

- **Continuous Random Variable**

  যদি কোন র‍্যান্ডম ভ্যারিয়েবলের ভ্যালু একটা রেঞ্জের মধ্যে যেকোনো একটা ভ্যালু হতে পারে, গণনা করা যায়না, তাইলে সেইটা হবে একটা continuous random variable। [Ref]

  A **continuous random variable** is one which takes an infinite number of possible values. Example : height, weight, the amount of sugar in an orange [Ref]

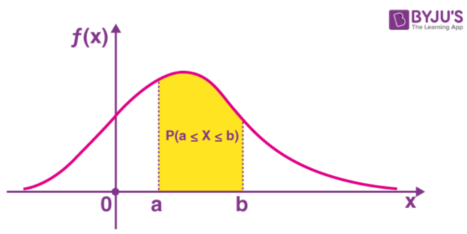
  - **Probability Density Function (PDF)**

    Calculate within an interval.

    $$f(x) = \begin{cases} \frac{2}{27}(1+x) & \text{if } x_i < x < x_j \\ 0 & \text{otherwise} \end{cases}$$

    $f(x)$ **is PDF, which satisfies these conditions:**

    - $f(x) \geq 0$ for $-\infty < x < \infty$
    - $\int_{-\infty}^{\infty} f(x) = 1$
    - $P(a \leq x \leq b) = \int_a^b f(x)$



[Ref], PDF Graph

Binomial Distribution

Binomial Experiment: A binomial experiment is an experiment using a fixed number of independent with only two outcomes.

$$b(x,n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0,1,2,3,....n \\ 0 & \text{otherwise} \end{cases}$$

$x =$ success

$n =$number of event

$p =$probability of success

$1-p =$ probability of unsuccess

$P(Success) + P(Unsuccess) = 1$

Poisson Distribution

When you should use it:

- When the probability is **too small,** such as 0.001, 0.01, 0.0002..

$\mu = np$

$P(x,\mu) = \frac{e^{-\mu}\mu^x}{x!}$

Bernoulli Distribution

A Bernoulli Distribution is a discrete probability which has only two possible outcomes.

Possible Outcomes:

Success - 1

Failure - 0

Example:

- Coin : Head - Success, Tail - Failure
- Rolling a dice : 6 - Success, 1,2,3,4,5 - Failure

$P(Success = 1) = P$

$P(Failure = 0) = 1 - P$

$PMF, f(x) = P^x(1-P)^{1-x}$

Expected value,

$E(x) = \sum xf(x)$
$= O(1-P) + 1(P) = P$

Variance of discrete probability distribution,

$V(x) = E(x^2) + [E(x)]^2 = P - P^2 = P(1-P)$

Moment

- Moments are a set of statistical parameters to measure a distribution.

- Moments are helpful to describe the distribution.
- To describe statistical properties of set of points
- To compare different sets of points, or to measure the similarity between two sets of points
- Moments can be useful tool for understanding the relationship between different sets of data

Classifications:

- **Raw Moment**

  $\mu_r' = \frac{f(x_i-A)^r}{n}$, $A =$ any arbitrary value
- **Central Moment**

  $\mu_r = \frac{f(x_i-\bar{X})^r}{n}$, $\bar{X} =$ mean

- **Moment about origin**

  $\gamma_r = \frac{f(x_i)^r}{n}$

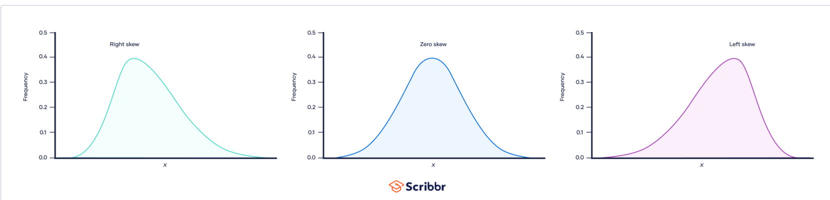
$r = n - th\ moment$

Variance (σ)

- $\mu_2 = \sigma^2$
- Standard Variance, $\sigma = \sqrt{u_2}$

Types:

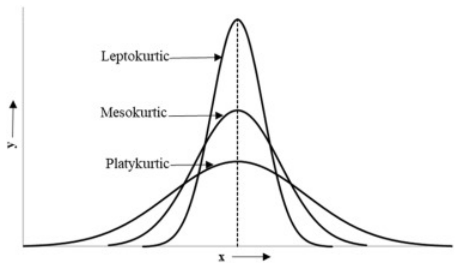
- **Mean**
  - $Mean(\bar{X}) = \frac{(x_1+x_2+x_3+...)}{n}$
- **Variance**
  - Variance measures the dispersion or **spread of the points around the mean**
  - $Variance, V(X)\ or\ \sigma^2 = \frac{((x_1-\bar{x})^2+(x_2-\bar{x})^2+...+(x_n-\bar{x})^2)}{n}$
  - The square root of the variance is called the standard derivation
- **Skewness**
  - It measures how asymmetric the distribution is about it's mean.
  - $Skewness = \frac{(x_1-\bar{x})^3+...+(x_n-\bar{x})^n}{n}$
  - Symmetrical : if the values to the left and right of the mean are roughly the same
  - $\beta_2 = \frac{\mu_3^2}{\mu_2^3}$
  - **Types:**



[Ref]

  - Symmetrical Distribution (No Skewness)
    - It's left and right sides are mirror images

            - mean = median
        - Positively
            - longer on the right side of its peak than on it’s left
            - mean > median
        - Negatively
            - longer on the left side of it’s peak than on its right
            - mean < median
- **Kurtosis**
    - Kurtosis refers to the degree of flatness or peakedness in the region of a curve.
    - **Types:**



        - **Platykurtic**
            - A frequency distribution is said to be platykurtic, when it is flatter than the normal curve.
            - Kurtosis: >0.263
            - $\beta_2 = \frac{\mu_4}{\mu_2^2},\ \gamma_2 = \beta_2 - 3$
              $\beta_2 < 0, \gamma_2 < 0$
        - **Leptokurtic**
            - A frequency distribution is said to be leptokurtic, when it is more peaked than the normal curve.
            - Kurtosis: < 0.263
            - $\beta_2 > 3, \gamma_2 > 0$
            - **Properties:**
                - Positive excess kurtosis
                - Slender in shape
                - Fatter tails
        - **Mesokurtic**
            - A frequency distribution is said to be Mesokurtic, when it almost resembles to the normal curve
            - Kurtosis : 0.263
            - $\beta_2 = 3, \gamma_2 = 0$

<u>Question Analysis:</u>

- **Difference Between Binomial and Poisson Distribution.**

Binomial	Poisson
It is **biparametric** (Has 2 parameters)	**Uniparametric**
The number of **attempts are fixed**	The number of **attempts are unlimited**
The probability of success is constant	The probability of success is extremely small

Binomial	Poisson
There are only **two possible outcomes**.	There are **unlimited** possible outcomes.
Mean **>** Variance	Mean **=** Variance

- **A and B play a game in which their chances of winning are in the ratio 5:2. Find A's chance of winning at least three games out of six games played.**

$P(A) = \frac{5}{7} = 0.71 \quad P(A') = 0.29$

At least 3 games win for A =

${}^6C_3 \times (0.71)^3 \times (0.29)^3 +$
${}^6C_4 \times (0.71)^4 \times (0.29)^2 +$
${}^6C_5 \times (0.71)^5 \times (0.29)^1 +$
${}^6C_6 \times (0.71)^6 \times (0.29)^0$
$= \boxed{}$

- **A random variable X has the following probability distribution :**

Value of x	0	1	2	3	4	5	6
P(x)	a	3a	5a	7a	9a	11a	13a

(i) **Determine the value of a**

$\sum P(x) = 1$

$\Rightarrow a + 3a + 5a + \dots + 17a = 1$
$\Rightarrow a = \frac{1}{81}$

(ii) **Also find** $P(0 < x < 5)$

$P(0<x<5) = P(1) + P(2) + P(3) + P(4)$
$= 3a + 5a + 7a + 9a$
$= 24 \times \frac{1}{81} \quad \left[a = \frac{1}{81}\right]$
$= \frac{8}{27}$

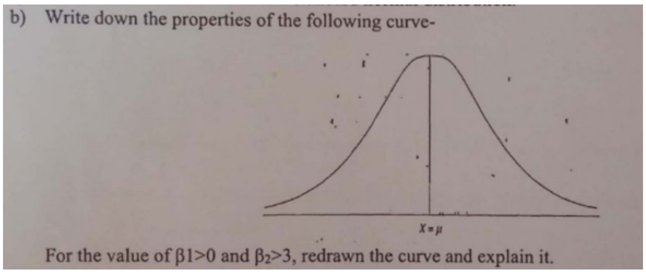
- **Define Normal Distribution and Standard Normal Distribution.**

  The Normal Distribution is defined by the probability density function for a continuous random variable in a system. Let us say, $f(x)$ is the probability density function and X is the random variable. Hence, it defines a function which is integrated between the range or interval $(x$ to $x + dx)$, giving the probability of random variable X, by considering the values between x and $x + dx$.

$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

The standard normal distribution (z-distribution) is a special normal distribution, where the **mean is 0** and **the standard derivation** (σ) **is 1.**

$z = \frac{x-\mu}{\sigma}, \ \mu =$ mean



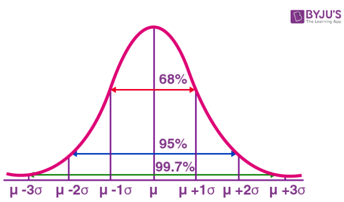
It's a leptokurtic curve. A frequency distribution is said to be leptokurtic, when it is more peaked than the normal curve.

Properties:

- Excessive positive kurtosis
- Slender in shape
- Flatter tails
- Kurtosis < 0.263

- **Describe the area property/properties of normal distribution.**
  - mean = median = mode
  - The total area under the curve should be equal to 1.
  - The normally distributed curve should be symmetric at the centre.
  - There should be half of the values are to the left of the center and exactly half of the the right.
  - The normal distribution curve must have only one peak.
  - The curve approaches the x-axis, but it never touches.

- Approximately 68% of the data falls within σ
- Approximately 95% of the data falls within 2σ
- Approximately 99.7% of the data falls within 3σ



- **Difference between a discrete random variable and a continuous random variable?**

Discrete random variable	Continuous random variable
Has finite number of possible values	could have any value within a certain range
Takes countable set of values	Takes on an uncountable set of values
Described by PMF	Described PDF
P(a ≤ X ≤ b) is meaningful	P(a ≤ X ≤ b) represents the probability of a range

Discrete random variable	Continuous random variable
Number of students in a class	Height of individuals in a population

Let x be a discrete random variable with the following probability mass function

$$P_X(k) = \begin{cases} 0.1 \; for \; k = 0 \\ 0.5 \; for \; k = 1 \\ 0.4 \; for \; k = 2 \end{cases}$$

Find EX and Var(X).

$$EX = \sum_{x_k \in R_X} x_k P_X(x_k)$$

$$= 0 \times 0.1 + 0.5 \times 1 + 0.4 \times 2$$

$$= 1.2$$

$$Var(X) = EX^2 - (EX)^2$$

$$EX^2 = 0^2 \times 0.1 + 1^2 \times 0.5 + 2^2 \times (0.4)$$

$$= 2.1$$

$$Var(X) = 2.1 - (1.2)^2$$

$$= 0.66$$

- **Write down the probability density function of continuous random variable.**

Let X be continuous random variable. Then a probability distribution or **probability density function (pdf)** of X is a function $f(X)$ such that for any two numbers a and b with $a \leq b$, we have

$P(a \leq X \leq b) = \int_a^b f(x)dx$

- **Importance of Normal Distribution**
  - The normal distribution is easy to work with mathematically.
  - Many things are actually normally distributed, or very close to it . Such as, height and intelligent are approximately normally distributed; measurement errors also often have a normal distribution.

Let X be a continuous random variable with the following PDF

$F_X(x)=ce^{-x}$ $x \geq 0$

Otherwise $f_X(x)=0$

where c is a positive constant.

a. Find c.
b. Find the CDF of X, $F_X(x)$.
c. Find $P(1<X<3)$

$$\int_0^\infty c e^{-x} \, dx = 1$$

$$\Rightarrow -c \left[e^{-\infty} - e^{-0} \right] = 1$$

$$\Rightarrow c = 1$$

b)

$$F_x(x) = e^{-x}, \quad x \geq 0$$
$$0, \quad \text{otherwise}$$

$$x < 0, \int_{-\infty}^{x} F_x(x)\, dx = \int_{-\infty}^{x} 0 \, dx = 0$$

$x \geq 0$,

$$\int_{-\infty}^{x} f(x)\, dx = \int_{-\infty}^{x} e^{-x}\, dx$$

$$= \int_{-\infty}^{0} 0\, dx + \int_{0}^{x} e^{-x}\, dx$$

$$= -e^{-x} + 1$$

$$F(x) = 1 - e^{-x}$$

c)

$F(x) = 1 - e^{-x}$
$P = F(2) - F(1) = -e^{-2} + e^{-1}$
$= 0.23254415793$

4.	Let X be a continuous random variable with PDF $f(x)=3x^2$ for $0 \leq x \leq 3$. Find mean, variance, β1 and β2.	10

$$E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$= 3 \int_{0}^{3} x^3\, dx$$

$$= \frac{3}{4} \left[x^4 \right]_0^3$$

$$= \frac{3^5}{4}$$

$$V(x) = Ex^2 - (Ex)^2$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$
$$= \int_0^3 x^2 \cdot 3x^2\,dx$$
$$= 3\left[\frac{x^5}{5}\right]_0^3$$
$$= \frac{3^6}{5}$$

$$V(x) = \frac{3^6}{5} - \frac{3^5}{4}$$

$$u_4 = u_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1^4$$

$$E(x^4) = \int_0^3 x^4 \cdot 3x^2 = \frac{3^8}{7}$$
$$E(x^3) = \int_0^3 x^3 \cdot 3x^2 = \frac{3^7}{6}$$
$$E(x^2) = \int_0^3 x^2 \cdot 3x^2 = \frac{3^6}{5}$$
$$E(x) = \int_0^3 x \cdot 3x^2 = \frac{3^5}{4}$$

$$\mu_2 = E(x^2) - [E(x)]^2$$
$$\mu_4 = E(x^4) - 4E(x^3)E(x) + 6E(x^2)[E(x)]^2 - 3[E(x)]^4$$
$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$E(x^4) = \int_0^3 x^4 \cdot 3x^2 = \frac{3^8}{7}$$
$$E(x^3) = \int_0^3 x^3 \cdot 3x^2 = \frac{3^7}{6}$$
$$E(x^2) = \int_0^3 x^2 \cdot 3x^2 = \frac{3^6}{5}$$
$$E(x) = \int_0^3 x \cdot 3x^2 = \frac{3^5}{4}$$

$$\mu_3 = E(x^3) - 3E(x^2)E(x) + 2[E(x)]^3$$
$$\mu_2 = E(x^2) - [E(x)]^2$$
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \boxed{}$$