Statistics Exam Prep

Created by	Borhan
Last edited time	@April 9, 2024 3:50 PM
Tag	

▼ Topic Lists from Nawaj

Random variable সংজ্ঞা, Example.

Discrete " " " , "

Probability Mass function কী? condition .

Step function. Example

Discrete-এর অসুবিধা,

Continuous Random variable সংজ্ঞা, Example

Probability Density function, Graph-এ ব্যাখ্যা, condition

moment (μ) এর সংজ্ঞা, about origin, about point. Example 5.3.

5.8, 5.4, 5.5, 5.6, 5.7

Kurtosis figure, math, সংজ্ঞা, formula

Bernoulli, Binomial সংজ্ঞা eq^n সহ, condition, Example (সহ),

Discrete-এ curve-fit 7.1, 7.2, 7.4

(most) Poison Distribution, condition, অসুবিধা, Application, Moments (সহ ছবি)

1st ~~2nd~~ degree recursion 7.24, 7.25, 7.26

এ curve fitting

Chapter: 9 Theory. 704-এর Equation দেও হবে, অনুরূপ

9.1, Exp... curve fitting (Math

অনুশীলনী

Chapter: 10 সংজ্ঞা Eq^n of correlation, 760 দুইটির σ^1_x, σ^1_y ও দিয়ে

প্রমাণ, 10.3.1 (main eq^n), প্রমাণ, 10.7

(765p)

chapter: 8 Gaussian normal (583) distribution-এর definition, eq^n, প্রমাণ

nor. Dis. property < Area / Characteristic অন্যান্য

587-এ Graph

Importance, 8.11, 8.12, 601.

Mean yield

Nawaj

▼ Random Variable

Random Variable: A random variable is a rule that assigns a numerical value to each outcome in a sample space.

Example: If the random variable Y is the number of heads we get from tossing two coins, then Y could be 0, 1, or 2.

▼ Types of Random variables

BASIS	DISCRETE RANDOM VARIABLE	CONTINUOUS RANDOM VARIABLE
Meaning	A discrete random variable is one which may take an only a **countable** number of **distinct** values.	A continuous random variable is one which takes an **infinite** number of possible values.
Range of specified number	Complete	Incomplete
Values	Values are obtained by **counting**.	Values are obtained by **measuring**.
Classification	Non-overlapping	**Overlapping**
Assumes	Distinct or separate values.	Any value between the two values.
Represented by	**Isolated** points	**Connected** points
Described by	Probability Mass Function	Probability Density Function
Example	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}	Height, weight, the amount of sugar in an orange.

▼ Probability Mass Function and Probability Density Function

Probability Mass Function (PMF)	Probability Density Function (PDF)
Gives the probability of **discrete random variables.**	Gives the probability density of **continuous random variables**.
Applicable to discrete random variables.	Applicable to continuous random variables.
Takes on **specific values with non-negative probabilities.**	Defines **probabilities within a range**, not specific values.
$f(x)$ **is PMF, satisfies these properties:** • $f(x) >= 0$ • $\sum f(x) = 1$	$f(x)$ **is PDF, which satisfies these conditions:** • $f(x) \geq 0$ for $-\infty < x < \infty$ • $\int_{-\infty}^{\infty} f(x) = 1$

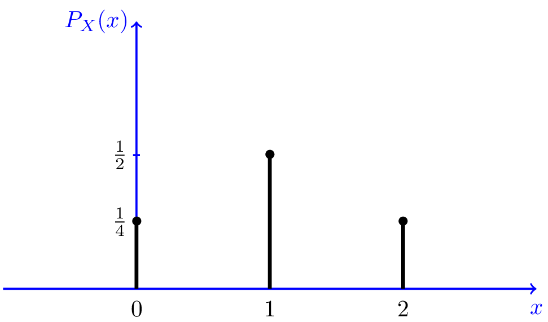
- $f(x) = P(x)$

Represents the likelihood of a particular outcome.

Example: Given a probability mass function $f(x) = bx^3$ for x = 1, 2, 3. Find the value of b.

Solution: According to the properties of probability mass function,

$\sum_{x=1}^{3} f(x) = ... = b = \frac{1}{36}$



- $P(a \le x \le b) = \int_a^b f(x)$

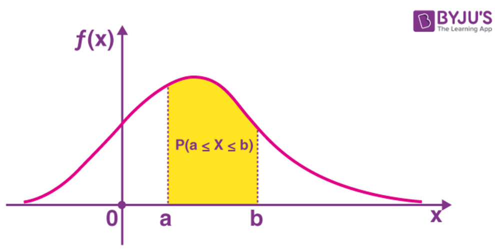
Represents the likelihood of an outcome occurring within a range.

Example : If the probability density function is given as:

$$f(x) = \begin{cases} x(x-1) & \text{if } 0 \le x < 3 \\ x & \text{x } \ge 3 \end{cases}$$

Find P(1< X < 2).

Ans: $P(1 < x < 2) = \int_1^2 f(x) = ... = \frac{5}{6}$



CDF

The cumulative distribution function (CDF) of random variable X is defined as FX(x) = P(X ≤ x), for all x ∈ R. Note that the subscript X indicates that this is the CDF of the random variable X. Also, note that the CDF is defined for all x ∈ R.

B

Binomial Distribution and Bernoulli Distribution

Binomial Experiment

Consists of a **fixed** number of **identical**, **independent trials**.

Bernoulli Experiment

Involves a **single trial** or experiment .

Each trial has **two possible outcomes**: success or failure.	Only **two possible outcomes**, often labeled as success (1) or failure (0).
The probability of success (p) remains **constant** for each trial.	The probability of **success (p) remains constant.**
The probability of failure (q) is complementary to the probability of success. (q = 1 - p)	The probability of failure (q) is complementary to the probability of success. (q = 1 - p)
The **probability mass function** is given by the binomial distribution formula.	Essentially, **a special case of the binomial distribution** with only one trial.
Tossing a coin multiple times, counting successes.	Flipping a coin once, where heads might be considered a success.
$b(x, n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, .n \\ 0 & \text{otherwise} \end{cases}$ $x =$ success $n =$number of event $p =$probability of success $1 - p =$ probability of failure	$P(\text{Success}) + P(\text{Failure}) = 1$

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
There are only **two possible outcomes**.	There are **unlimited** possible outcomes.

Binomial	Poisson
Mean **>** Variance	Mean **=** Variance

Poisson Distribution

Conditions

- An event can occur any **number of times during a time period**
- Events occur **independently**
- The rate of **occurrence** is **constant**
- The probability of an event is proportional to the length of the time period

$P(X) = \frac{\lambda^x e^{-\lambda}}{x!}$, $\lambda = \text{parameter of distribution=average or mean}$

Example: Number of suicides reported in a particular day.

Expected Value/Mean and Variance

The expected value is the arithmetic mean of the possible values a random variable can take, weighted by the probability of those outcomes.

Variance is the expected value of the **squared deviation** from the mean of a random variable.

Expected value for PMF, $EX = \sum xf(x)$

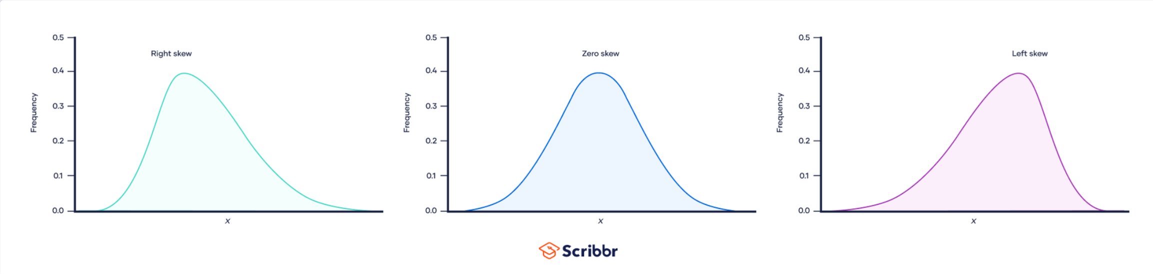
Expected value for PDF, $EX = \int_{-\infty}^{\infty} xf(x)dx$

Variance, $Var(x) = EX^2 - (EX)^2$

$EX^2 = \sum x^2 f(x) = \int_{-\infty}^{\infty} x^2 f(x)dx$

Skewness

- It measures how asymmetric the distribution is about it’s mean.



- **Types:**
    - **For Symmetrical/Zero Skew, $\beta_1 = 0$**
        - It's left and right sides are mirror images
        - mean = median = mode
    - **Left Skew/Negative Skew**
        - longer on the right side of its peak than on it's left
        - Mean < Median < Mode
    - **Right Skew/Positive Skew**
        - longer on the left side of it's peak than on its right
        - Mode < Median < Mean

    Left and right skew can't be considered from β_1

    $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, $\beta_1 \geq 0$, $\gamma_1 = \pm\sqrt{\beta_1} = \frac{\mu_3}{\sigma^2}$

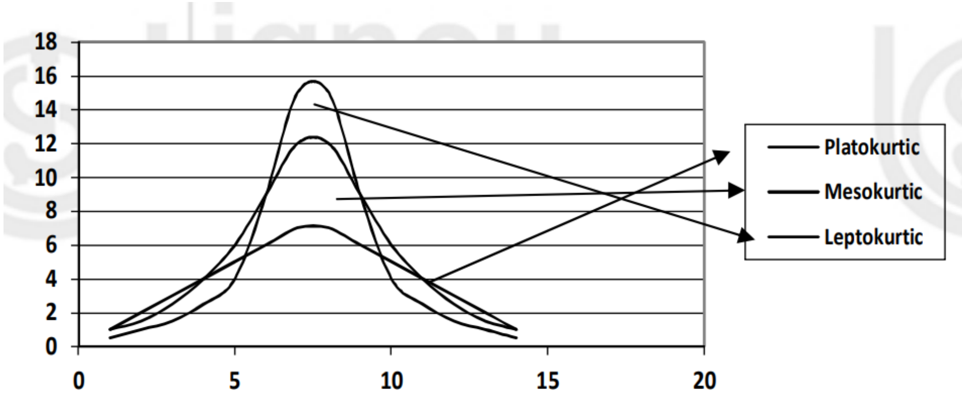
    Then the sign of skewness would depend upon the value of μ_3 whether it is positive or negative.

    Formula for calculating skewness,

    $S_k = \frac{\text{Mean}-\text{mode}}{\sigma}$

▼ Kurtosis

- Kurtosis refers to the **degree of flatness or peakedness** in the region of a curve.



	Mesokurtic	Platykurtic	Leptokurtic
Tailedness	Medium	Thin	Fat
Outlier Frequency	Medium	Low	High
Kurtosis (beta_2)	Moderate (=3)	Low (<3)	High (>3)
Excess kurtosis	0	Negative	Positive
Example Distribution	Normal	Uniform	Laplace

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$, $\gamma_2 = \beta_2 - 3$

▼ Step Function (Skipped)

▼ Moment

- Moments are a set of **statistical parameters to measure a distribution**.
- Moments can be useful tool for understanding the relationship between different sets of data

Classifications:

If $x_1, x_2, ..., x_n$ be the values of a variable x with corresponding frequencies $f_1, f_2, ..., f_n$ respectively the $r-th$ moment about

- **Raw Moment/Moments about the origin:**

  $\mu_r' = E(X^r) = \frac{\sum f_i x_i^r}{N}$

- **Central Moment:**

  $\mu_r = E|(x - \mu x)^r|$

- **Moments about a point**

$\mu_r = \frac{f_i(x_i - A)^r}{N}$, $\text{A} = \text{Arbitrary value}$

Types:

- First → Mean
- Second → Variance
- Third → Skewness
- Fourth → Kurtosis

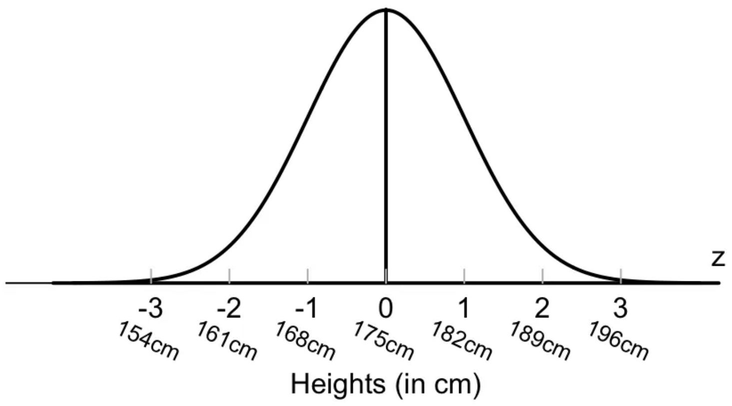
$\mu_3 = E(X^3) - 3E(X^2)E(X) + 2[E(X)]^3$

$\mu_4 = E(X^4) - 4E(X^3)E(X) + 6E(X^2)[E(X)]^2 - 3[EX]^4$

▼ Normal Distribution

- Bell Curve
- mean=median=mode
- symmetric about mean
- A probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean,.
- Known as Gaussian distribution
- Mesokurtic
- Mean as a a standard point
- The probability of first half and last half is 0.5 individually
- There's only one peak, unimodal
- The curve never touch the x axis

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5}(\frac{x-\mu}{\sigma})^2$



Probability of ≤ 175 cm or ≤ 0 : 0.5

- **Importance of Normal Distribution**
    - The **discrete probability distributions** can be approximated to normal distribution
    - **All small sample distribution** approximated to normal distribution for large values
    - The entire **small sample theory** is based on the fundamental assumption that the parent population from which the sample is drawn assumed to be normal.
    - In conduction of **large sample tests**, the sample distribution of sample means **sample variances, sample proportions** tends to normal distribution.
    - Entire large sample theory depends on the area's property.
    - **Since error function** follow normal distribution. It has a fundamental importance of the theory
    - **Central Limit Theorem** follows normal distribution.

▼ Standard Normal Distribution

- A special normal deviation which
  - mean is 0
  - the standard deviation is 1
- Also known as z-distribution