Date:

LECTURE #13

Poisson Distribution

(keyword : per year or anything)

Formula:

$$P(x; \lambda t) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

Using Matlab: Y = poisspdf (x, LAMBDA)

Approximation to Binomial

- $n \geq 20$ and $p \leq 0.05$
- $n \geq 100$ and $np \leq 10$

Geometric Distribution

(Discrete probability distribution of random variable X)

Formula:

$g(x; p) = p\, q^{x-1}$, $x = 1, 2, 3 \ldots$

$g(x; p) = p\, q^{x}$, $x = 0, 1, 2, \ldots$

Conditions

- Trial is repeated until success occur
- Repeated Trial are independent of other.
- Probability of success is same for each trial.

X represents the number of trials in first success occur.

Mean of G.D = $1/p$

Variance of G.D = q/p^2

For larger value of x, $P(x)$ gets closer to zero.

Using Matlab:

Y = geopdf (x, p).

Discrete Uniform Distribution (Equally Probable)

Formula = $\frac{1}{k}$; $P(x; k) = \frac{1}{k}$

Using Matlab: Y = unidpdf (X, N)

Date:

LECTURE # 14

Negative Binomial Distribution

(Same as Binomial with exception that the trial will be repeated until a fixed number of successes occur)

Formula:

$$b^*(x; k, p) = {}_{x-1}C_{k-1}\, p^k q^{x-k}, \quad x = k, k+1, \dots$$

Constitute a geometric progression, denoted as $g(x; p)$

{ Trials are independent }

LECTURE # 16
(PART-1)

A (finite) sequence of experiments in which each experiment has a finite number of outcomes with given probability

(Finite) Stochastic Process

A convenient way of describing such a process and computing the probability of any event

Tree Diagram

Date:

LECTURE # 16
(PART - II)

- **Mutually Exclusive or Disjoint**

  (if $A \cap B = \{\}$ or ϕ)

- **Addition Rule I**

  (Two events are mutually exclusive,

  $P(A \text{ or } B) = P(A) + P(B)$ or $P(A \cup B) = P(A) + P(B)$)

- **Rule of elimination** or Total Probability Theorem.

  $$P(B) = \Sigma_{i=1}^{n} (A_i \cap B) = \Sigma_{i=1}^{n} P(A_i) P(B \mid A_i)$$

- **Conditional Probability of A_i given B,**

  $$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

So by substituting it in rule of elimination :

$$P(A_i \mid B) = \frac{P(A_i) P(B \mid A_i)}{P(A_1) P(B \mid A_1) + P(A_2) P(B \mid A_2) + \ldots + P(A_n) P(B \mid A_n)}$$

or $$P(A_i \mid B) = \frac{P(A_i) P(B \mid A_i)}{\Sigma_{i=1}^{n} P(A_i) P(B \mid A_i)}$$

Conditional Probability share **Properties** of ordinary Probabilities

(i) $P(A \mid B) \geq 0$

(ii) $P(S \mid B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

(iii) $P(B \mid B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

Date:

LECTURE # 17

Uniform Distribution

(The density function of continuous uniform random variable X on the distribution interval [A, B] is $f(x; A, B) = \frac{1}{B-A}$, $A \le x \le B$, 0 otherwise)

Mean = $\frac{A+B}{2}$ and **Variance** = $\frac{(B-A)^2}{12}$

Normal Distribution

(The density of normal random variable X, with mean μ and variance σ^2)

Formula:

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

Properties of Normal Distribution:

- It is bell-shaped.
- Maximum occurrence at $x = \mu$
- Mean, Median, Mode are at center of Distribution.
- Curve is symmetric about vertical axis through μ.
- Concave downwards if $\mu - \sigma < X < \mu + \sigma$ otherwise concave upward.
- It is symmetric about mean.
- It is continuous.
- Never touches the x-axis
- Total Area under curve is 100% or 1.

with one σ = 68% area under curve, with two σ = 95% and with three σ = 100% area falls.

Date:

Points of Inflection

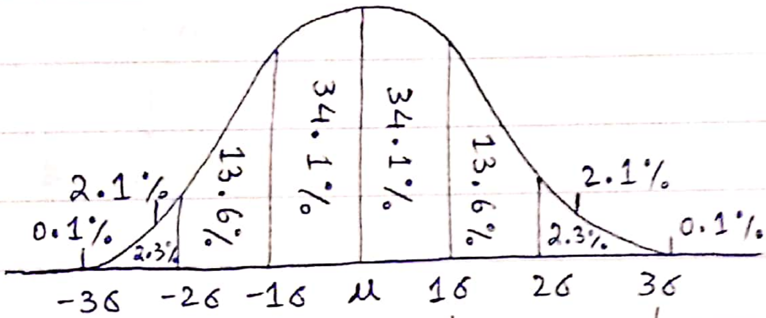
(Points at which curve changes from curving upward to curving downward are called inflection points)

Mean of Normal Distribution $= \mu$

variance of Normal Distribution $= \sigma^2$

LECTURE #18

Normal Distribution



$\mu - 1\sigma \le X \le \mu + 1\sigma$

68%

$\mu - 2\sigma \le X \le \mu + 2\sigma$

95%

$\mu - 3\sigma \le X \le \mu + 3\sigma$

99% ≈ Almost 100%

Standard Normal Distribution

- All same properties of normal distribution
- with mean zero and standard deviation one.

Value can be transformed from normal distribution to standard normal distribution.

Formula : $Z = \dfrac{\text{Value} - \text{Mean}}{\text{Standard Deviation}}$ as $Z = \dfrac{X - \mu}{\sigma}$

Values of standard normal are called z values or z-score

Flowcharts

To compute Area

$Z = \dfrac{x - \mu}{\sigma}$

↓

Area

To compute x or z score

Area

↓

$x = \mu + \sigma z$

LECTURE # 19

Percentiles

(A marker on a normal curve such that the marker is greater than or equal to that percentage of results).

Deciles

(Values in the order statistic that divides the data into ten (10) equal parts).

LECTURE # 20

Chebyshev's Theorem

(The probability that any random variable X will assume a value within k standard deviations of mean is atleast $1 - 1/k^2$)

Formula: $P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$

Measures of Location:

- **Mean**: Sample Mean is denoted by $\bar{X}$.

  $\bar{X} = \Sigma X / n$

- **Sample Median**:

  Arrange in ascending order

  $$x = \begin{cases} x_{(n+1)/2} & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & \text{if } n \text{ is even} \end{cases}$$

Trimmed Mean

- Firstly, arrange them in ascending order
- Certain percent * n shows how many numbers you need to remove from dataset (from both ends)
- Compute the mean of remaining values

Measures of Variability

- **Sample range** = $X_{max} - X_{min}$
- **Sample Variance** = $s^2 = \Sigma(X - \bar{X})^2 / n-1$
- **Sample Standard Deviation** = $s = \sqrt{s^2}$

Date:

LECTURE # 21

- **Population** (Totality of observations)
- **Sample** (Subset of Population)
- **Bias** (Consistently overestimated or underestimated characteristics)
- **Overestimate** (To form too high an estimate of)
- **Underestimate** (Estimate something smaller or less important)
- **Statistical Inference** (Drawing conclusions about population characteristic)
- **Population Parameters** (Descriptive measure for an population which are population mean and population variance)
- **Statistic** (Any function of random variables consituting random sample)
- **Sampling Distribution** (The probability distribution of statistic)

Indication :

N : Population Size

μ : Population Mean $= \Sigma X / n$

σ^2 : Population Variance

$$\sigma^2 = \frac{\Sigma (X - \bar{X})^2}{N}$$

π : Population Proportion $= \frac{k}{N}$

n : Sample Size

$\bar{X}$: Sample Mean $= \Sigma X / n$

S^2 : Sample Variance

$$S^2 = \frac{\Sigma (X - \bar{X})^2}{n} \quad , \quad \frac{\Sigma (X - \bar{X})^2}{n-1}$$

Biased : n Unbiased : $n-1$

p : Sample Proportion $= \frac{x}{n}$

Central Limit Theorem

(If $\bar{X}$ is the mean of a random sample of size n taken from a population with mean μ and variance σ^2, then limiting form of distribution of $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$)

Scanned with CamScanner

Date:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}$$

(Statistical Inference is the act of generalizing from sample to population with calculated degree of certainty.)

FORMS OF STATISTICAL INFERENCE

- **Estimation**
  - **Point Estimate**
    - Single value estimate for a population parameter
  - **Interval Estimate**
    - Interval or range of values to estimate population parameter
- **Hypothesis Testing**

{ To form an interval estimate, use the point estimate as center of interval and then add and subtract margin of error. }

Level of Confidence: c or $1 - \alpha$
(The probability that interval estimate contains the population parameter).

Critical Values
(The number on borderline separating sample statistics that are likely to occur from those that are unlikely to occur)

Date

Margin of Error : E

(The greatest possible distance between point estimate and the value of parameter)

$$E = Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$ Margin of error for μ (σ known)

Conditions:

- Sample is random.
- Any one is true:
  - Population is normally distributed
  - $n \geq 30$

Unbiased Estimator

(A statistic $\hat{\theta}$ is said to be unbiased estimator of parameter θ if $\mu_{\hat{\theta}} = E(\hat{\theta}) = \theta$)

- $E(\bar{X}) = \mu$, So $\bar{X}$ is unbiased estimator of μ.
- $E(S^2) = \sigma^2$, So S^2 is unbiased estimator of σ^2.

LECTURE #22

Sampling Distribution of Means and Central Limit Theorem.

$$\mu_{\bar{x}} = \frac{1}{n}(\mu + \mu + \mu + \dots + \mu) \text{ (n terms)}$$

$$= \frac{n\mu}{n} = \mu \quad \text{So} \quad \mu_{\bar{x}} = \mu$$

Variance : $$\sigma_{\bar{x}}^2 = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \dots + \sigma^2) \text{ (n terms)}$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \quad \text{So} \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Date:

Standard Deviation = $\sqrt{\text{Variance}}$

$$\sqrt{\sigma_{\bar{x}}^2} = \sqrt{\frac{\sigma^2}{n}}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Sampling Distribution of Difference b/w Two means:

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$Z = \frac{\bar{X} - \mu}{\sigma}, \text{ so, } Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Types of Error

Type I error

Rejection of H_0 when it is true.

Type II error

Acceptance of H_0 when it is false.

{The probability of committing Type I error is also called level of significance, as α }

Scanned with CamScanner

Date:

Normal Population, σ^2 unknown, n small

Confidence Interval for mean μ :

$$\bar{X} - t_{(v,\alpha/2)} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{(v,\alpha/2)} \frac{s}{\sqrt{n}}$$

$v = n - 1$ (For one ~~mean~~ sample)

If two samples given then use following formula for v,

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{((s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1))}$$

Any Population, σ^2 unknown / known, n Large.

Confidence Interval for mean μ :

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

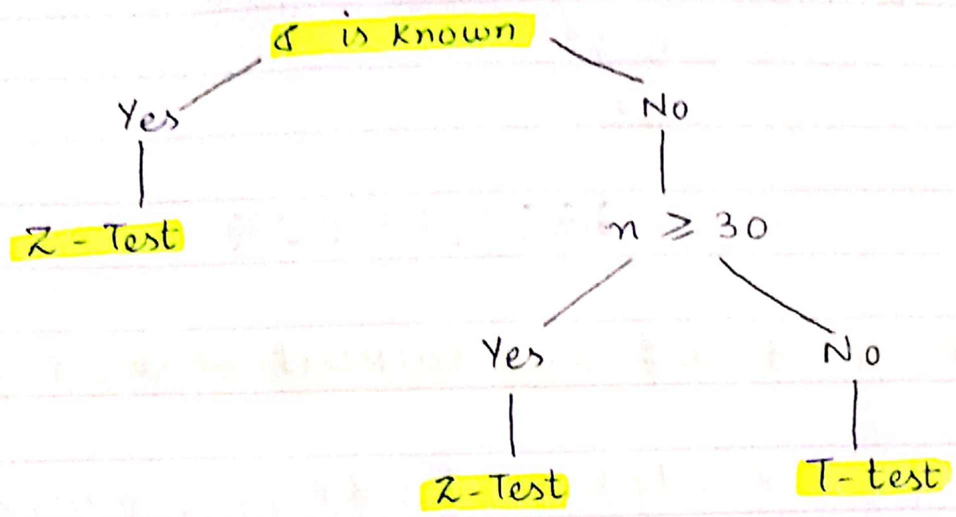
Sample Size for population mean :

$$n = \left(\frac{\sigma Z_{\alpha/2}}{e}\right)^2$$

(If n is in fractional form, we always round it to next whole number)

Date:

LECTURE # 23



Confidence Interval, CI

(Interval estimate is a range of values used to estimate the true value of population parameter)

Confidence Level, $1-\alpha$

(Proportion of times that confidence Interval actually does contain population parameter. Also known as Degree of confidence or confidence coefficient)

Margin of Error : E or e

$$E = Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$ (Margin of error for proportion p)

Confidence Interval :

$$\hat{p} - E < p < \hat{p} + E$$ (For proportion p)

Date:

Sample size for population proportion :

$$n = \frac{\hat{p}\hat{q}\, z_{\alpha/2}^2}{E^2}$$

LECTURE #24

One - sided confidence Interval on μ, σ^2 unknown :

Upper one-sided bound : $\bar{X} + t_{(\alpha, n-1)}\, s/\sqrt{n}$

Lower one-sided bound : $\bar{X} - t_{(\alpha, n-1)}\, s/\sqrt{n}$

Two samples :

(Estimating The Difference b/w Two Proportions)

C.I :

$$(\hat{p}_1 - \hat{p}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

(Estimating The Difference b/w Two Means)

C.I: ($\mu_1 - \mu_2$, σ_1^2 and σ_2^2 known)

$$(\bar{X}_1 - \bar{X}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

(Estimating The Difference b/w Two Means)

C.I: ($\mu_1 - \mu_2$, σ_1^2 and σ_2^2 unknown but equal)

Pooled Estimate of variance = $S_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

$$(\bar{X}_1 - \bar{X}_2) - t_{(\alpha/2,\, n_1+n_2-2)}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{(\alpha/2,\, n_1+n_2-2)}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

(Estimating The Difference b/w Two Means)
C.I ($\mu_1 - \mu_2$, σ_1^2 and σ_2^2 unknown but unequal)

$$(\bar{X}_1 - \bar{X}_2) - t_{(\frac{\alpha}{2}, v)}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{(\frac{\alpha}{2}, v)}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

LECTURE # 25

Hypothesis
(A claim or statement about property of population)

Hypothesis Test (Test of significance)
(A standard procedure for testing a claim)

Approach to Hypothesis Testing with Fixed Probability of Type-I error:

- State null and alternative hypotheses.
- Choose the fixed significance level α.
- Test statistic
- Calculations
- Critical Region
- Conclusions

Pairs of Hypotheses:

$H_0: \mu \leq k$	$H_0: \mu \geq k$	$H_0: \mu = k$
$H_1: \mu > k$	$H_1: \mu < k$	$H_1: \mu \neq k$

{Try to use equal symbol in H_0, its more professional}

Date:

Rejection of H_0

$Z < Z_{\alpha}$ (True)	$Z > Z_{\alpha}$ (True)	$Z < Z_{\alpha/2}$ or (True) $Z > Z_{\alpha/2}$ (True)
$T < T_{v,\alpha}$ (True)	$T > T_{v,\alpha}$ (True)	$T < T_{v,\alpha/2}$ or (True) $T > T_{v,\alpha/2}$ (True)

{ if false Then we do not reject H_0 so we accept H_0 and reject H_1 }

Two Tailed : $Z_{\alpha/2}$ (Equality Symbol)

One. Tailed : Z_{α} (lesser or Greater)

Left: $\mu < a$

Right: $\mu > a$

LECTURE # 26

Pairs of hypotheses :

$H_0 : p \leq p_0$	$H_0 : p \geq p_0$	$H_0 : p = p_0$
$H_1 : p > p_0$	$H_1 : p < p_0$	$H_1 : p \neq p_0$

Rejection of H_0

$Z > Z_{\alpha}$	$Z < Z_{\alpha}$	$Z < Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$
$T > T_{v,\alpha}$	$T < T_{\alpha}$	$T < T_{\alpha/2}$ or $T > T_{\alpha/2}$

Date

$$Z_{cal} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

Two samples :
(Tests on Two Proportions)

$$Z_{cal} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_c q_c \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $p_c = \frac{x_1 + x_2}{n_1 + n_2}$ and $q_c = 1 - p_c$.