PROBABILITY AND STATISTICS

FIRST YEAR NOTES

-AKSHANSH CHAUDHARY

Probability and Statistics Notes, First Edition

Presented by: Akshansh Chaudhary
Graduate of BITS Pilani, Dubai Campus
Batch of 2011

Course content by: Mr. R. Mutharasan
Then Faculty, BITS Pilani, Dubai Campus

Layout design by: AC Creations © 2013

The course content was prepared during Spring, 2012.
More content available at: www.Akshansh.weebly.com

Chapter - 1.

Introduction to Probability & Counting.

* Popul^n: Collection of objects on which we are making a decision is called popul^n.
* Sample: A subset of popul^n i.e. used to draw conclusion about popul^n is called a sample.

Note

1. Probability of an event is expressed in terms of %age from 0 to 100.
2. If the probability of an event = 1, it is called a certain event.
3. An event is known as impossible event, if it never occurs. It is denoted by ϕ. & $P(\phi) = 0$.

* Random Experiment: An expt. is known as a random expt. if its results are unpredictable, even though it is conducted under identical cond^{ns}. eg. tossing a coin.

* Sample space: Sample space of a random expt. is the result of all possible outcomes of that expt. eg: when 2 coins are thrown,

$$S = \{HH, HT, TH, TT\}$$

* ~~Each elmt. of~~ or subset of sample space is called an Event.
* Each elmt. of a sample space is called Sample Point.

* Mutually exclusive event : 2 events A & B are said to be mutually exclusive if the occurence of 1 event excludes the possibility of occurence of other event. eg: In throwing a die, A be the event of getting odd nos. & B be event of getting even nos, then,

$A = \{1, 3, 5\}$ & $B = \{2, 4, 6\}$.

Here, A & B are ~~mui~~ mutually exclusive ($A \cap B = \phi$).

* Equally likely : 2 events (or more) are said to be equally likely if they have equal chance of occurence.

* Independent events : 2 events A & B are said to be independent if the occurence or non occurence of 1 event does not depend upon the occurence & non occurence of others.

eg: Let A be the event of appearing head on the first toss of a coin. & B be the event of appearing tail on the 2nd toss of a coin. Then, A & B are said to be independent.

* Methods to find the probability
(i) Personal approach : Assigning a probability based on personal experience.
(ii) Frequency approach : Based on no. of times it has been occuring. It can't be conducted in some cases since it maybe very expensive.
(iii) Classical approach : It is defined as the no. of

favourable cases by total no. of cases.
Applicable only if events are equally likely.

* **Permutation**

Arrangement of objects in a definite order.
eg: 2 letters are taken from set $A = \{a, b, c\}$
$\therefore$ various arrays are $(a,b)(b,a)(b,c)(c,b)(a,c)$ & (c,a)
Denoted by ${}^3P_2 = 6$.

* $${}^nP_r = \frac{n!}{(n-r)!}$$

* **Combination**

Selection of objects without regard to order.
eg: 2 letters chosen from the set $A = \{a, b, c\}$.
$\therefore$ various combinⁿˢ are $(a,b)(b,c)$ & (c,a).
Denoted by ${}^3C_2 = 3$

* $${}^nC_r = \frac{n!}{r!\,(n-r)!}$$

* **Factorial notation** :-

Let n be a +ve integer. The product
$n(n-1)(n-2) \ldots 3.2.1 = n!$ or $\underline{n|}$
Note :- $0! = 1$.

* **Multiplication principle** :

If an expt. can take place ~~in case~~ in k ~~changes~~ stages & n_i denotes the no. of ways in which stage i occurs $(i = 1, 2, 3, \ldots k)$, then altogether, the expt. can occur $n_1 . n_2 . n_3 \ldots n_k$ ways.

eg: how many diff^t combin^{ns} can be made for a briefcase, that has 3 dial lock? Each dial has nos. 0-9.

Ans = $10 \times 10 \times 10 = 1000$.

eg: how many license plates can be made if the 1st 3 entries must be letters followed by 3 nos; $26 \times 26 \times 26 \times 1000$.

eg: suppose a multiple choice exam has 10 ques. & each ques. has 5 ans. How many diff^t ways ~~wo~~ could exam be answered

Ans :- 5^{10}.

* Permut^n of indistinguishable objects.

If $\exists$ n objects of which n_1 are of kind 1, n_2 are of kind 2, ~~etc~~, ... n_k are of kind k, then, no. of arrays of these objects is

$$\frac{n!}{n_1!\, n_2!\, n_3! \cdots n_k!} \quad ; \text{ where } n_1 + n_2 + \cdots + n_k = n$$

eg: 1.3.8

Exercise (Pg-17)

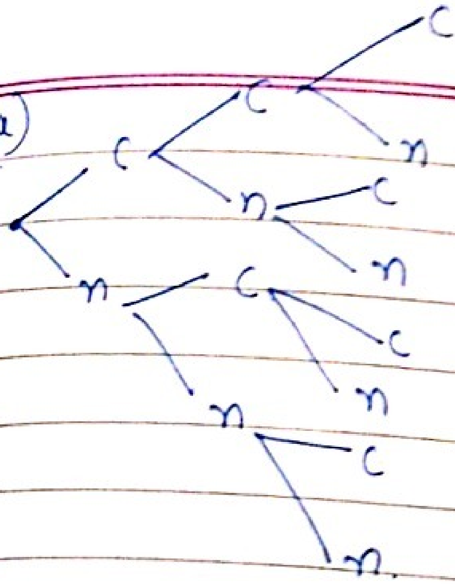
Q.1 Reqd. prob. = $\frac{3}{10}$; Frequency approach (not equally likely)

Q 2) " = $\frac{30}{75}$; Frequency approach

Q.3) $S = \{DD, \bar{D}\bar{D}, D\bar{D}, \bar{D}D\}$

Reqd $P = \frac{1}{4}$; Classical approach.

Q.5) (a)



(b) $S = \{ccc, ccn, cnc, cnn, ncc, ncn, nnc, nnn\}$

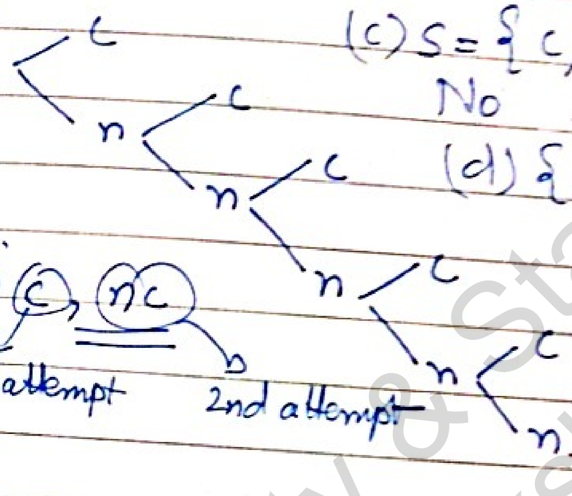
(c) $S \rightarrow A_1, A_2 = \{cc\}\ A_3\{nnn\}$

(d) No, yes, yes, No

(e) No, ∵ not equally likely

Q.7)

(a)



(b) No

(c) $S = \{c, nc, nnc, nnnc, \_\_\_ \}$
No

(d) $\{c, nc, nnc, nnnc\} = A$

(e) c, nc
1st attempt 2nd attempt

Q.10) (a) $4 \times 3 = 12$ (b) $12 \times 5 = 60$ (c) $60 \times 6 = 360$

Q.13) (a) 2^5 (b) 2^4 (c) $1 \times 1 \times 1 \times 2 \times 2$ (d) 1

Q.15) (a) $5!$ (b) $4!(\times 2)$

Q.18) (a) ${}^8C_3$ (b) $\frac{{}^1C_1 \times {}^7C_2}{{}^8C_3}$ or $\frac{3}{8}$

Q.20) (a) ${}^{2000}C_{120}$ (c) $\frac{{}^2C_2 \times {}^{1998}C_{118}}{{}^{2000}C_{120}}$ (b) ${}^{1998}C_{118} \times {}^2C_2$

Q.22) $1 - \frac{{}^3C_0 \cdot {}^{17}C_5}{{}^{20}C_5}$

Q.21) (a) ${}^{10}C_3 \cdot {}^8C_2 \cdot {}^4C_2 \cdot {}^3C_1$

(b) ${}^{10}C_3 \cdot {}^8C_2 \cdot ({}^1C_1 \cdot {}^3C_1) \cdot {}^3C_1$

Q.26) (a) 3^6 (b) $1/3^6$ (c) ${}^6C_2 \cdot {}^4C_2 \cdot {}^2C_2$. 3

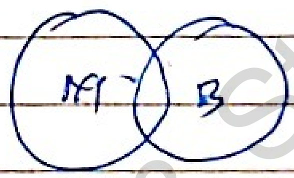
Chapter - 2

PROBABILITY LAWS

1) $12/13$ 2) (a) $\frac{5}{35} + \frac{4}{35} + \frac{1}{35} = \frac{2}{7}$ (b) $\frac{24}{35}$

3) 0.45, 0.13, 0.46.

4)



$0.95 - 0.76 = 0.19$

$P(\text{main}) = 0.95$, $P(\text{back}) = 0.8$

~~P(A∪B)~~ $P(M \cup B) = 0.99$

$\Rightarrow 0.99 = 0.95 + 0.8 - P(M \cap B)$

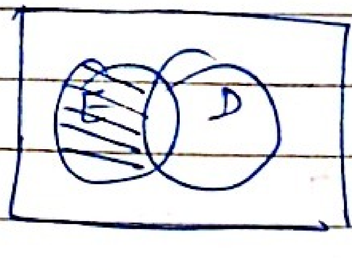
(a) $\Rightarrow P(M \cap B) = 0.76$

(b) 0.19

(c) 0.01

0.95, .99, 0.76

Q.5) B, E, D



$E \cap B = 0.3$

$D \to 0.4$

$E \cup D = 0.68$

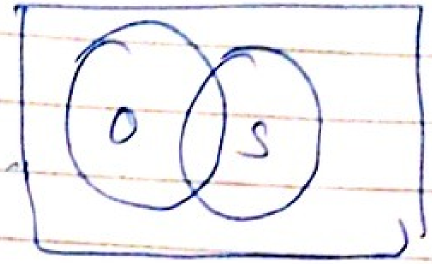
E?

(a) $0.68 = 0.4 + E - 0.3$

$0.58 - 0.3 =$ (b) 0.28 $0.58 = E$

(c) 0.1

Q.6) $O = 0.75$
$S = 0.15$
$O \cup S = 0.85$
(a) $O \cap S = 0.05$
(b) 0.1



$0.15 - 0.05$
$\cancel{0.0}$

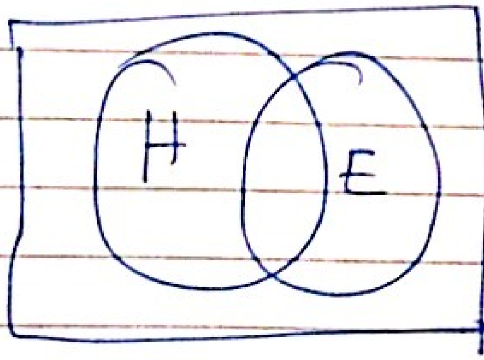
Q.7) $T = 0.6$
$R = 0.3$
$T \cap R = 0.2$
$T \cup R = 0.7$

0.1

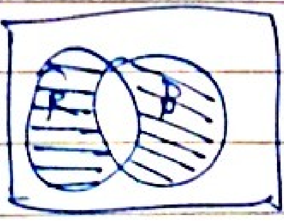
Q.8) $S = 0.25$
$L = 0.5$
$S\,L - (S \cap L) = 0.35$
$(S \cap L) = ?$ $\quad 0.5 - 0.35 = 0.15$
$1 - (S \cup L) = ?$ $\quad 0.4$

Q.10) $H = 0.8$
$E = 0.4$
$H \cap E = 0.35$

0.45



Q.9) $R = 0.6$
$B = 0.7$
$R - (R \cap B) = 0.18$
(a) $(R \cap B) = 0.42$
(b) $0.18 + 0.28 = 0.46$



$A - (A \cap B)$
$+$
$B - (B \cap A)$

$P\left(\frac{A_2}{A_1}\right) = \frac{P(A_1 \cap A_2)}{P(A_1)}$

$P(A_2) = P(A_1 \cap A_2)/P(A_1) \rightarrow$ Independent events

1) $C\cancel{S} = 0.01$

$S = 0.005$

$C \cup S = 0.014$

$S \sim (C \cap S) = 0.005 - 0.001$

$= \underline{0.004}$

3) $E \cap D = 0.3$

$D = 0.4$

$E \cup D = 0.68$

$E = 0.58$

(a) $P\left(\frac{D}{E}\right) = \frac{P(E \cap D)}{P(E)} = \frac{0.3}{0.58} = \frac{30}{58} = \frac{15}{29}$

(b) $P\left(\frac{\bar{D}}{E}\right) = \frac{P(E \cap \bar{D})}{P(E)} = \frac{0.28}{0.58} = \frac{14}{29}$

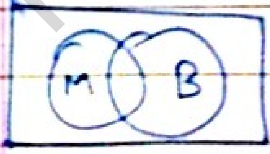
(c) Cond^nal

(d) $P\left(\frac{D}{\bar{E}}\right) = \frac{P(D \cap \bar{E})}{P(\bar{E})} = \frac{0.1}{0.42} = \frac{5}{21}$

(e) $P(D) \neq P\left(\frac{D}{E}\right)$

4) $P(M) = 0.95 \quad P(B) = 0.8$

$P(M \cup B) = 0.99 \quad P(M \cap B) = 0.76$



(a) $P\left(\frac{B}{\bar{M}}\right) = \frac{P(B \cap \bar{M})}{P(\bar{M})}$

$= \frac{0.04}{0.05} = \frac{4}{5}$

(b) No.

Q.15) $P(C \cap T) = 0.05$

$P(C) = 0.4$

$P(T) = 0.35$

(a) $P\left(\frac{C}{T}\right) = \frac{P(C \cap T)}{P(T)} = \frac{0.05}{0.35} = \frac{1}{7}$

(b) $P\left(\frac{T}{C}\right) = \frac{P(T \cap C)}{P(C)} = \frac{0.35}{0.4} = \frac{7}{8}$

Q.16) (a) $10 \times 10 = 100$

(b) $\cancel{1}20$

(c) 10

(d) 1

$P\left(\frac{9}{2}\right) = \frac{P(2 \cap 9)}{P(2)}$

$= \frac{1}{10}$

Q.17) $P(T) = 0.05$

$P(L) = 0.8$

$P(T \cap L) = 0.01$

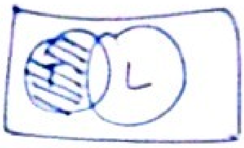
(a) $P\left(\frac{L}{T}\right) = \frac{0.01}{0.05} = \frac{1}{5}$

(b) $P\left(\frac{T}{L}\right) = \frac{0.01}{0.8} = \frac{1}{80}$

(c) $P\left(\frac{T}{\bar{L}}\right) = \frac{P(T \cap \bar{L})}{P(\bar{L})} = \frac{0.04}{0.2} = \frac{1}{5}$

$\cancel{(d)}$ 2 $\cancel{b}$ P(c) 0.04

(d) 0.84



Q.18) $P(A_1) = 1/2$ 0.35

$P(A_2) = 7/10$

Q.19) No

Q.20) $P(R_H\text{-ve}) = 0.39$

$$P\left(\frac{R_H\text{-ve}}{2\text{-ve}}\right) = \frac{P(R_H\text{-ve} \cap 2\text{-ve})}{P(2\text{-ve})}$$

P(−ve from one parent) $= \frac{1}{2}$

So, $P(2\text{-ve}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Q.21) $\cancel{P(R_{H}\text{-ve} \cap 2\text{-ve}) = P(R_H) \cap P(R_H)}$ $= (0.39)^2$

Q.21) $P(\text{Parent } 1 \cap P2) = 0.39 \times 0.39 = (0.39)^2$

Q.22)

Q.23) $P(G) = 0.3$

$P(M) = 0.23$

$$P\left(\frac{M}{G}\right) = 0.7 = \frac{P(G \cap M)}{P(G)}$$

(a) $\Rightarrow P(G \cap M) = 0.21$

(b) $P\left(\frac{G}{M}\right) = \frac{0.21}{0.23} = \frac{21}{23}$.

Q.27) $P(P) = 0.67$,

$P\left(\frac{C}{P}\right) = 0.0144$ $P(C) = ?$

$P\left(\frac{C}{\bar{P}}\right) = 0.0012$

Baye's form

For n mutually exclusive events & considering any event A_j,

$$P\left(\frac{A_j}{B}\right) = \frac{P(B/A_j)\cdot P(A_j)}{\sum_{i=1}^{n} P(A_i)\,P(B/A_i)}$$

$\frac{C \cap P}{0.67} = 0.0144$

$\Rightarrow P(C \cap P) = 0.0144 \times 0.67$

$\frac{P(C) - (0.0144 \times 0.67)}{0.33} = 0.0012$

$\Rightarrow P(C) = (0.0012 \times 0.33) + (0.0144 \times 0.67)$

Q.24) $P(B) = 0.35$

$P(A) = 0.1$

$P\left(\frac{B}{A}\right) = 0.4$

Ans $= 0.04$ (0.4×0.1)

Q.25)

$P(F) = 0.5$

$P(D) = 0.33$

$P\left(\frac{D}{F}\right) = 0.17$ $P(F \cap D) = 0.17 \times 0.5$

Q.26) $\cancel{P(Y) = (0.17)\left(\frac{1}{2}\right)}$

$\left(\frac{1}{2}\right)$ $P\left(\frac{Y}{1}\right) = 0.17$, $P\left(\frac{Y}{2}\right) = 0.03$

P

Q.35) $\cancel{P\left(\frac{Y}{A}\right) = 0.85}$

$P(A) = 0.1$, $P(\bar{A}) = 0.9$

$P\left(\frac{C}{A}\right) = 0.85$

$P\left(\frac{C}{\bar{A}}\right) = 0.04$.

$$P\left(\frac{A}{C}\right) = \frac{P(C/A) \cdot P(A)}{P(C/A) \cdot P(A) + P(C/\bar{A}) \cdot P(\bar{A})}$$

$$P\left(\frac{A}{C}\right) = \frac{(0.85)(0.1)}{(0.85)(0.1) + (0.9)(0.04)}$$ Ans

$$= \frac{85}{85+36} = \frac{85}{121}$$

Q.36) $P\left(\frac{D}{P}\right) = 0.5$, $\cancel{P(\frac{S}{M})}$ $P(S) = 0.01$

$P(P) = 0.05$.

$P\left(\frac{S}{D}\right) = ?$

$$P\left(\frac{S}{D}\right) = \frac{P(D/S) \cdot P(S)}{P(D/S) \cdot P(S) + P(D/P) \cdot P(P)}$$

$$= \frac{0.5 \times 0.01}{}$$

Q.41) $P(A) = 0.6$ $\quad P\left(\frac{J}{A}\right) = 0.01$

$P(B) = 0.3$ $\quad P\left(\frac{J}{B}\right) = 0.05$

$P(C) = 0.1$ $\quad P\left(\frac{J}{C}\right) = 0.04$

$P\left(\frac{A}{J}\right)$? $\quad P\left(\frac{B}{J}\right) = ?$ $\quad P\left(\frac{C}{J}\right) = ?$

(a) $$P\left(\frac{A}{J}\right) = \frac{P(J/A) \cdot P(A)}{P(J/A) \cdot P(A) + P(J/B) \cdot P(B) + P(J/C) \cdot P(C)}$$

$$= \frac{(0.01)(0.6)}{(0.01)(0.6) + (0.05)(0.3) + (0.04)(0.1)}$$

$$P\left(\frac{A}{J}\right) = \frac{6}{25}$$

(b) $P\left(\frac{B}{J}\right) = \cancel{P\ 0.05}\ \frac{15}{25} = \frac{3}{5}$

(c) $P\left(\frac{C}{J}\right) = \frac{4}{25}$ $\quad$ ~~P(overload 2 or more) = 0.05~~

Q.40) $P\left(\frac{Bl}{A}\right) = 0.01$, $\quad P\left(\frac{O}{A}\right) \cancel{P\left(\frac{A}{O}\right)} = 0.6$ $\quad P\left(\frac{Bl}{Ov}\right) = 0.05$

$P\left(\frac{Bl}{B}\right) = 0.02$ $\quad P\left(\frac{O}{B}\right) \cancel{P\left(\frac{B}{O}\right)} = 0.2$ $\quad P(O) = 0.05$

$P\left(\frac{Bl}{C}\right) = 0.03$ $\quad P\left(\frac{O}{C}\right) \cancel{P\left(\frac{C}{O}\right)} = 0.15$

(a) $P\left(\frac{A}{O}\right)$? $P\left(\frac{B}{O}\right)$? $P\left(\frac{C}{O}\right)$.

$\cancel{P\left(\frac{Bl}{H}\right) = 1 - P\left(\frac{A}{Bl/H}\right)}$

Date 16/2/12
Page

Chapter - 3
Discrete distributions

In statistics, we deal with random variables, whose observed value is determined by chance. They are of 2 categories:
1. Discrete
2. Cts.

1. Discrete: A random variable is said to discrete, if, it can take, either a finite or a countable infinite no. of possible values.
2. Cts:- A random variable is cts if, it can assume, any value in some interval, or intervals of real nos. & the probability, that it assumes any specific value = 0.

* The behaviour of a random variable in terms of probabilities is accomplished by means of 2 f^{ns} :- 1) Density f^n or Probability f^n or Probability mass f^n or Probability density f^n → pdf

Denoted by $P(x)$ or $f(x)$.

2) Cumulative distribu^n f^n :- Denoted by $F(x)$
The word cumulative suggest the role of f^n. ~~that~~ It accumulates the probability found by means of density.

Puffin Date 19/2/12 Page

* Discrete density :-

The f^n $f(x)$ is said to be the pdf of a discrete random variable x, if

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

The necessary & sufficient cond^{ns} for a f^n to be a discrete density are

(i) $f(x) \geq 0$ for every x.

2) $\sum_{\text{all } x} f(x) = 1$.

* Cumulative distrib^n f^n : (cdf)

Let X be a discrete random variable with density f. The cdf for X, denoted by F is defined by

$$F(x) = P(X \leq x).$$

~~eg:-~~

Pg - (82)

Q.9) X: no. of sys. operable

$\Rightarrow X \in \{0, 1, 2, 3\}$.

Let $f(x)$ be the probability density f^n of x.

Given, probability for each sys. operable $= 0.9$.

$\therefore$ The " " " " inoperable $= 0.1$.

$f(0) = 0.1 \times 0.1 \times 0.1 = (0.1)^3 = 0.001$

$f(1) = (0.1)^2 (0.9) \times 3 = 0.027$

$f(2) = (0.1)(0.9)^2 \times 3 = 0.243$

$f(3) = (0.9)^3 = 0.729$

(a) Density table

x	0	1	2	3
$f(x)$	0.001	0.027	0.243	0.729

(b) $f(x) = k_{(x)} (0.9)^x (0.1)^{3-x}$

Put $x = 0$

$f(0) = k_{(0)} (0.9)^0 (0.1)^{3-0}$

$\Rightarrow k_{(0)} = 1 = \binom{3}{x} = {}^3C_x \; , x = 0$

Put $x = 1$

$\Rightarrow f(1) = k_{(1)} (0.9)(0.1)^2$

$\Rightarrow (0.1)^2 (0.9)(3) = k_{(1)}$

$\Rightarrow k_{(1)} = 3 = {}^3C_1 = {}^3C_x \; , x = 1$.

Put $x = 2$

$f(2) = k_{(2)} (0.9)^2 (0.1)$

$\Rightarrow \cancel{(0.1)}\cancel{(0.9)^2}(3) = k_{(2)} \cancel{(0.9)^2}\cancel{(0.1)}$

$\Rightarrow k_{(2)} = {}^3C_2 = {}^3C_x \; , x = 2$

So, in general

$k_{(x)} = {}^3C_x \; , \; x = 0, 1, 2, 3$.

Ans

(c) $F(x) = ?$

x	0	1	2	3
$F(x)$	0.001	0.028	~~0.270~~ 0.271	1.000

(d) $P(X \geq 1) = 1 - P(X < 1)$

$= 1 - F(0)$

$= 0.999$

(e) $P(X \leq 1) = F(1) = 0.028$

Puffin
Date
Page

§ EXPECTATION & DISTRIBUTION PARAMETERS.

The 3 important parameters in statistics are
1. mean, μ.
2. variance, σ^2
3. Standard deviation, σ

, the values of these parameters are approximated using various ~~stastical~~ statistical techniques.

- **Expectation:**

Expectⁿ of a discrete random variable, X is denoted by E(X), which is defined by

$$E(X) = \sum_{\text{all } x} x \cdot f(x)$$

Density fⁿ

random variable

eg: When 1 die is thrown,
$x \in [1, 6]$, $f(x) = 1/6$.

* Properties of expectation :-

1) $E(c) = c$; c is a constt.
2) $E(cX) = c\,E(X)$
3) $E(X+Y) = E(X) + E(Y)$

* Variance

It is a statistical parameter that reflects consistency.
If variance is less, then, each value of the populⁿ is very close ~~tt~~ to the mean &, if variance is high, ∃ high deviations of each & every value from the mean.

Puffin
Date
Page

Let X be a random variable with mean μ. The variance of X, denoted by

$$Var\ X \text{ or } \sigma^2,$$

given by

$$Var\ X = \sigma^2 = E[(X-\mu)^2]$$

Computational formula

$$Var\ X = E(X^2) - [E(X)]^2$$

- Note :-

$$\sigma^2 = E[(X-\mu)^2]$$
$$= E(X^2 - 2X\mu + \mu^2)$$
$$= E(X^2) - 2\mu E(X) + E(\mu^2)$$
$$= E(X^2) - 2\mu E(X) + \mu^2$$
$$= E(X^2) - 2\mu\mu + \mu^2$$
$$\Rightarrow \sigma^2 = E(X^2) - [E(X)]^2 \quad ; \ \boxed{E(X) = \mu}$$

* Standard deviation (σ) (S.D.)

Let X be a random variable with var, σ^2. The S.D of X, denoted by σ is given by

$$\sigma = \sqrt{Var}.$$

Properties for variance :-

P1. $Var\ c = 0$; c : constt

P2. $Var\ cX = c^2\ Var\ X$.

P3. If X & Y are independent, then,

$$var(X+Y) = var(X) + var(Y)$$

Puffin
Date
Page

Pg-83

Q.14 (a) $E(X) = 0(0.7) + 1(0.2) + 2(0.05) + (0.03)3 + 4(0.01) + 5(0.01)$

$= 0.2 + 0.1 + 0.09 + 0.04 + 0.05$

$= 0.48$

(b) $\mu_x = E(X) = 0.48$

(c) $E(X^2) = \sum x^2 f(x) = 0\cdot(0.7) + 1(0.2) + 4(0.05) + 9(0.03) + 16(0.01) + 25(0.01)$

$= 0.2 + 0.2 + 0.18 + 0.16 + 0.25$

$= 1.08$

(d) $\text{Var } X = E(X^2) - [E(X)]^2$

$= 1.08 - (0.48)^2$

$= \cancel{0.7696}\ 0.8496$

0.921
$0\ \sqrt{0.8496}$

(e) $\sigma^2_x = 0.8496$

(f) $= \sqrt{\sigma^2} = 0.921$

(g) Unit: no. of graftings.

Pg-84) Q.21) $E(X) = 3$
$E(X^2) = 25$
$E(Y) = 10$
$E(Y^2) = 164$

(a) $9 + 10 - 8 = 11$

(b) $6 - 30 + 7 = -31$

(c) $25 - 9 = 16$

(d) 4

(e) 64

(f) 8

(g) $\text{var}(3X + Y - 8)$

$= \text{var}(3X) + \text{var}(Y) + \text{var}(-8)$

$= 3^2 \text{var}(X) + 64 + 0 \Rightarrow 9(16) + 64 = 208$

(h) $var(2X-3Y+7)$

$= var(2X) + var(-3Y) + var(7)$

$= 4(16) + 9(64) + 0$

$= 640$

(i) $E\left[\frac{(X-3)}{4}\right] \quad \frac{1}{4}(E(X)-3) = \frac{3-3}{4} = 0$

$var\left[\frac{(X-3)}{4}\right] = \frac{1}{16}\left[var(X) - 0\right] = \frac{16}{16} = 1$

(j) $E\left[\left(\frac{Y-10}{8}\right)\right] = \frac{1}{8}(10-10) = 0$

$var\left[\left(\frac{Y-10}{8}\right)\right] = \frac{1}{64}(64-0) = 1.$

(k) No.

* Mean & Variance are equal $\not\Rightarrow$ distribⁿ are equal.

* Geometric distribution:

probability for success

The distribⁿ of a discrete random variable X is called geometric distribⁿ, with parameter p $(0<p<1)$. If its density fⁿ (pdf) is defined by $\boxed{f(x) = p(1-p)^{x-1}}$; $x \in (1, 2, 3, \ldots)$

$p(q)^{x-1}$

* Mean $= \mu = \frac{1}{p} = \frac{d}{dt}(M_x(t))$

* $Var\ X = \frac{q}{p^2} = \frac{1-p}{p^2}$.

* MOMENTS: Let X be a random variable. The k^{th} ordinary moment for X is defined as $E(X^k)$, $k = 1, 2, 3 \ldots$

For discrete distrib^{ns},

$$\boxed{E(X^k) = \sum_{\text{all } k} x^k \cdot f(x)}$$

Q. If the probability that an applicant for a driver's license will pass the road test on any given time is 0.8. What is the probability that he will finally pass the test?

(a) On 4th trial. $(0.2)(0.2)(0.2)(0.8) = 0.0064$

(b) In fewer than 4 trials $(0.8) + (0.2)(0.8) + (0.2)(0.2)(0.8)$ $= 0.992$

Let X denote the no. of trials req^d to achieve the 1st success. Then, X follows geometric distrib^n, given by :-

$$P(X=x) = p\, q^{x-1} \quad (x = 1, 2, 3, \ldots)$$

$$p = 0.8,\ q = 0.2$$

(a) $\Rightarrow P(X=4) = (0.8)(0.2)^{4-1} = (0.2)^3(0.8)$

(b) $P(X<4) = (0.8)(0.2)^{1-1} + (0.8)(0.2)^{2-1} + (0.8)(0.2)^{3-1}$

* Note :- $E(X) = \mu$ is called 1st ordinary moment for ~~x~~ X.

$E(X^2)$ is called 2nd ordinary moment & so on.

* Moment Generating f^n (MGF)

It enables us to find these moments with less effort.

Let X be a random variable, with $S \longrightarrow$ 'f'.

The MGF for X, denoted ~~by~~ & defined by.

$$M_X(t) = E(e^{tX})$$

Note:- For discrete distrib^{ns} :-

$$\text{MGF} = \sum_{\text{all } x} e^{tx} \cdot f(x) = M_X(t) = E(e^{tx})$$

For geometric distrib^{ns}

$$M_X(t) = \frac{pe^t}{1-qe^t} \qquad (q = 1-p)$$

Differentiating $M_x(t)$, we have

$$\frac{d}{dt}(M_x(t)) = \frac{d}{dt}\left(\frac{pe^t}{1-qe^t}\right)$$

$$= \frac{(1-qe^t)(pe^t) - pe^t(-qe^t)}{(1-qe^t)^2}$$

$$= \frac{pe^t - \cancel{pqe^{2t}} + \cancel{pqe^{2t}}}{(1-qe^t)^2}$$

$$\Rightarrow \frac{d}{dt}M_x(t) = \frac{pe^t}{(1-qe^t)^2} = \frac{M_x(t)}{1-qe^t}.$$

$t=0$

$$\Rightarrow \left.\frac{d}{dt}M_{(x)}(t)\right|_{t=0} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

$$\Rightarrow \text{Mean}(\mu') = \frac{1}{p}.$$

IIly, we have $E(X^2)$, $E(X^3)$ ... from MGF

Q.25 (a) P(success) = 0.05 , p = 0.05

(b) $f(x) = pq^{x-1}, \quad x = 1, 2, 3, \ldots$

$= (0.05)(0.95)^{x-1}$

(c) $M_x(t) = \dfrac{pe^t}{1-qe^t} = \dfrac{(0.05)e^t}{1-(0.95)e^t}$

(d) $E(X) = \dfrac{d}{dt}(M_x(t))\Big|_{t=0} = \dfrac{d}{dt}\left(\dfrac{0.05e^t}{1-0.95e^t}\right)\Big|_{t=0}$

$= \dfrac{1}{0.05} = 20$

$E(X^2) = \dfrac{d^2}{dt^2} M_x(t)\Big|_{t=0}$

$\dfrac{(1-qe^t)^2 pe^t - pe^t(2)(1-qe^t)(-q)}{(1-qe^t)^4}$

$= \dfrac{p^2(p) - p(2)(p)(-q)}{p^2 \times p^2}$

$= \dfrac{p+2q}{p^2} = 1.95 \times 20 \times 20$

$= 780$

$\begin{array}{r} 1.95 \\ \times 4 \\ \hline 780 \end{array}$

$Var\,X = (E(X^2)) - [E(X)]^2$

$= 400 - 780 = -380$

$\sigma = \sqrt{Var\,X} = 19.4936$.

(e) $P(X \geq 3) = 1 - P(X < 3)$

$= 1 - [f(1) + f(2)]$

$= 1 - (0.05 + 0.05 \times 0.95)$

$= 0.9025$.

Pg 37

Q.31 $f(x) = \frac{(x-3)^2}{5}, \quad x = 3, 4, 5$

(a) Verify : f is a density f^n

Clearly, $f(x) \geq 0, \ \forall x \geq 3 \ \& \leq 5$

$$\sum_{x=3}^{5} f(x) = \sum_{x=3}^{5} \left(\frac{(x-3)^2}{5}\right) = 0 + \frac{1}{5} + \frac{4}{5} = 1$$

So, it is a density

Density table :-

x	3	4	5
$f(x)$	0	$1/5$	$4/5$

(b) $E(X) = \sum_{x=3}^{5} x f(x) = 3(0) + 4\left(\frac{1}{5}\right) + 5\left(\frac{4}{5}\right)$

$$= \frac{4}{5} + \frac{20}{5} = \frac{24}{5} = 4.8$$

(c) $mgf = e^{3t}(f(3)) + e^{4t}(f(4)) + e^{5t} f(5)$

$$= e^{3t}(0) + e^{4t}\left(\frac{1}{5}\right) + e^{5t}\left(\frac{4}{5}\right)$$

$$MGF = \frac{e^{4t} + 4e^{5t}}{5}.$$

$$\frac{d}{dt}(M_x(t)) = E(X) = \frac{4e^{4t} + 20e^{5t}}{5}$$

At $t = 0$, $E(X)\Big|_{t=0} = 4.8$, same as above

(e) $E(X^2) = \sum_{x=3}^{5} x^2 f(x) = 9(0) + 16\left(\frac{1}{5}\right) + 25\left(\frac{4}{5}\right)$

$$= \frac{16 + 100}{5} = \frac{116}{5}.$$

$E(x) = \sum x f(x)$

MGF $= E(e^{tx}) = \sum e^{tx} f(x)$

Puffin Date ______ Page ______

(1) $E(X^2) = \frac{d^2}{dt^2} M_x(t) = \frac{d}{dt}\left(\frac{4e^{4t} + 20e^{5t}}{5}\right)$

$= \frac{16e^{4t} + 100e^{5t}}{5}$

At $t=0$, $E(X^2) = \frac{116}{5}$ Ans

(2) $\sigma^2 = E(X^2) - [E(X)]^2$

$= \frac{116 \times 5}{25} - \frac{24 \times 24}{25}$

$= \frac{580 - 576}{25} = \frac{4}{25} = 0.16$

$\sigma = \sqrt{\sigma^2} = 0.4$

Pg-88

Q 35) $f(x) = c e^{-x}$, $x \in Z^+$

(a) (i) $c \geq 0$

(ii) $\sum_{\text{all } x} f(x) = 1 \Rightarrow ce^{-1} + ce^{-2} + \dots = 1$

$\Rightarrow c\,(e^{-1} + e^{-2} + e^{-3} + \dots) = 1$

$= c\left(\frac{1/e}{1 - 1/e}\right) = 1$

$\Rightarrow c\left(\frac{1}{e-1}\right) = 1 \Rightarrow \boxed{c = e - 1}$

$\Rightarrow f(x) = (e-1)e^{-x}$

(b) $\sum_{\text{all } x} e^{tx} f(x) = M_x(t)$

$\sum_{\text{all } x} e^{tx}(e-1)e^{-x} = \sum_{\text{all } x} e^{x(t-1)}(e-1)$

$= (e-1)\left[e^{(t-1)} + e^{2(t-1)} + e^{3(t-1)} + \dots\right]$

$= (e-1)\left(\frac{e^{t-1}}{1 - e^{t-1}}\right)$

$(e-1)\frac{\frac{e^t}{\cancel{e}}}{\frac{e - e^t}{\cancel{e}}} = \frac{(e-1)e^t}{e - e^t}$

(c) $E(X) = \frac{d}{dt}\left(M_x(t)\right)\Big|_{t=0}$

$$= \frac{d}{dt}\,\frac{(e-1)e^{t-1}}{1-e^{t-1}}$$

$$= (e-1)\left[\frac{(1-e^{t-1})(e^{t-1}) \mp e^{t-1}(\cancel{1}-e^{t-1})}{(1-e^{t-1})^2}\right]$$

$$= \frac{(e-1)\left(e^{t-1} - \cancel{(e^{t-1})^2} + \cancel{e^{(t-1)2}}\right)}{(1-e^{t-1})^2}\Bigg|_{t=0}$$

$$= \cancel{(e-1)}$$

$$(e-1)\left[\frac{e^{-1}}{(1-e^{-1})^2}\right]$$

$$= (e-1)\left[\frac{\frac{1}{e}}{\left(\frac{e-1}{e}\right)^2}\right] = (e-1)\left[\frac{e-2}{(e-1)^2}\right]$$

$$= \frac{(e-2)}{(e-1)}$$

$\Rightarrow E(X) = \frac{e}{e-1}$

(c) $E(X) = \frac{d}{dt} M_x(t)$

$$= \frac{d}{dt}\,\frac{(e-1)e^t}{(e-e^t)}$$

$$= (e-1)\left[\frac{(e-e^t)(e^t) - e^t(-e^t)}{(e-e^t)^2}\right]$$

$$= (e-1)\left[\frac{e^{t+1} - \cancel{e^{2t}} + \cancel{e^{2t}}}{(e-e^t)^2}\right] = \frac{e}{e-1}$$

BINOMIAL DISTRIBUTION :-

The distribⁿ of a discrete random variable x, is called binomial distribⁿ with parameters n, p, if its density fⁿ is defined by

$$f(x) = {}^nC_x\, p^x\, q^{n-x} \quad (x = 0, 1, 2, \ldots n)$$

Note :-

$$(x+a)^n = \sum_{r=1}^{n} {}^nC_r\, x^{n-r} a^r$$

Results :-

(i) MGF $= m_x(t) = (q + pe^t)^n$; $q = 1-p$

(ii) Mean $= \mu = E(X) = np$

(iii) Variance $= \sigma^2 = npq$

(iv) Std. deviation $= \sigma = \sqrt{npq}$

(v) X is a $B(n, p)$ variate

Binomial distribⁿ with 2 parameters n & p.

ex:- Out of 800 families with 4 children each, how many families would be expected to have atleast 1 boy

$n = 4$, $p = 1/2$, $q = 1/2$

Reqᵈ P. $= P(X \geq 1)$ or $1 - P(X=0) = 15/16$

Ans $= P_{\text{1 family}} \times 800 = 750$

(Pg-88) Q 36 , $n = 15$, $p = 0.2$

$X \rightarrow B(15, 0.2)$

Puffin
Date $E(X)$
Page

Pg. 88 Q.36) $n = 15,\ p = 0.2$

(a) $f(x) = {}^nC_x p^x q^{n-x}$

$= {}^{15}C_x (0.2)^x (0.8)^{15-x}, \quad x \in [0, 15]$

(b) $M_x(t) = (q + pe^t)^n = [0.8 + (0.2)e^t]^{15}$

(c) $E(X) = np = 3$

$Var.(X) = npq = 2.4$

(d) $E(X) = \left.\frac{d}{dt} M_x(t)\right|_{t=0}$

$= \left. n(q+pe^t)^{n-1}[pe^t]\right|_{t=0}$

$= n[q+p]^{n-1}[p]$

$= np = 3$

$E(X^2) = \frac{d^2 (M_x(t))}{dt^2}$

$= \frac{d}{dt}\left[n(q+pe^t)^{n-1}(pe^t)\right]$

$= np\left[\frac{d}{dt}(q+pe^t)^{n-1}(e^t)\right]$

$= np\left[(n-1)(q+pe^t)^{n-2}e^t + e^t(q+pe^t)^{n-1}\right]$

$= 11.4$

$Var(X) = E(X^2) - [E(X)]^2$

$= 2.4$

(e) $P(X \leq 1) = f(0) + f(1)$

$= {}^{15}C_0 (0.8)^{15} + {}^{15}C_1 (0.2)(0.8)^{14}$

$= (0.8)^{15} + 15(0.2)(0.8)^{14}$

$= (0.8)^{14}(0.8 + 3)$

$= (3.8)(0.8)^{14} = 0.167$

Q.46) $np = 5, \quad npq = 4$

$\Rightarrow q = \frac{4}{5}, \; p = \frac{1}{5}, \; n = 25.$

Q.4) $n = 10, \; p = 0.8$.

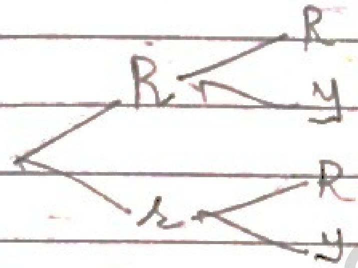
(a) $P(X \geq 9) = P(X = 10) + P(X = 9)$

$= {}^{10}C_{10} (0.8)^{10} (0.2)^{10-10}$

$+ {}^{10}C_{9} (0.8)^{9} (0.2)^{1} = 0.3758$

(b) $(A)^5 = (0.3758)^5$.

Q.39) (a) $S = \{RR, RY, rR, rY\}$



Reqd P $= 1/4$.

(b) $n = 5, \; p = 0.25$

$f(x) = {}^{n}C_{x} \, p^{x} \, q^{n-x} = {}^{5}C_{x} (0.25)^{x} (0.75)^{5-x}$

$E(X) = np = 5(0.25) = 1.25$

$P(X >= 3) = \sum_{x=3}^{5} {}^{5}C_{x} (0.25)^{x} (0.75)^{5-x}$

$= 0.1035$

☆ HYPERGEOMETRIC Distribⁿ :-

The distribⁿ of a discrete random variable X is called hypergeometric distribⁿ with parameters (N, n, r), if its density fⁿ is defined by :-

$$f(x) = \frac{{}^{r}C_{x} \; {}^{N-r}C_{n-x}}{{}^{N}C_{n}},$$

where, $\max[0, n-(N-r)] \leq x \leq \min(n, r)$ & $N, n, r \in \mathbb{Z}^{+}$.

- r & x are related
- N & n are related.

Pg-92 Q.59 $N = 3000$, $r = 600$, $n = 20$, $x \in [0, 20]$

(a) $f(x) = \frac{{}^{r}C_x \; {}^{N-r}C_{n-x}}{{}^{N}C_n}$

$f(x) = \frac{{}^{600}C_x \; {}^{2400}C_{20-x}}{{}^{3000}C_{20}}$; $x \in [0, 20]$

(b) $E(X) = \sum x f(x)$

$= \sum_{x=0}^{20} x \left[\frac{{}^{600}C_x \cdot {}^{2400}C_{20-x}}{{}^{3000}C_{20}} \right] = \frac{nr}{N} = \frac{20 \times \cancel{600}^{2}}{\cancel{3000}} = \boxed{4}$

$Var(X) = \frac{nr(N-r)(N-n)}{N^2(N-1)} = \frac{20(600)(2400)(2980)}{(3000)^2(2999)}$

$= 3.179$

(c) $P(X <= 3) = \sum_{x=0}^{3} f(x) = \sum_{x=0}^{3} \frac{{}^{600}C_x \; {}^{2400}C_{20-x}}{{}^{3000}C_{20}}$

(d)

*Note:-

Mean $= \cancel{M} \cdot \mu = \frac{nr}{N}$

$Var X = \sigma^2 = \frac{nr(N-r)(N-n)}{N^2(N-1)}$

Pg-91 Q.54 $E(X) = \frac{nr}{N} = 4.25$

Q. A panel of 7 judges is to decide which of the 2 final contestants A & B will be declared the winner

Puffin
Date ____
Page ____

4 judges $\xrightarrow{\text{vote}}$ A, other 3 $\rightarrow$ B. If we randomly select 3 of judges & seek their verdict, what is 'P' that a majority of them will favour A.

$N = 7,\ r = 4,\ n = 3$ (judges)

${}^{7}C_{3}$ find $f(x)$

Find $P(X \geq 2) = f(X=2) + f(X=3) = 22/35$

* Binomial Approximation to Hypergeometric distribⁿ.

Let x follows hypergeometric distribⁿ with parameters N, n & r. It is approximated by binomial distribⁿ with the parameters of binomial distribⁿ

$\underset{(\text{binomial})}{n} = \underset{(\text{hypergeometric})}{n}$; $p = \frac{r}{N}$; if $\frac{n}{N} \leq 0.05$, for the approximⁿ to be good

(pg. 92)

Q.59 (d) Limiting case :- $N = 3000,\ n = 20,\ r = 600$

$\frac{n}{N} = \frac{20}{3000} \leq 0.05$

$\therefore$ The approximⁿ is good.

Hence, for binomial distribⁿ,

$n = 20,\ p = \frac{r}{N} = \frac{600}{3000} = 0.2$

$\therefore$ Density fⁿ $= f(x) = {}^{n}C_{x}\, p^{x} q^{n-x}$

$\Rightarrow f(x) = {}^{20}C_{x} (0.2)^{x} (0.8)^{20-x}$

$P(X \leq 3) = f(0) + f(1) + f(2) + f(3) = \sum_{x=0}^{3} f(x)$

$\Rightarrow P(X \leq 3) = 0.4114$ (Pg-691)

Pg-75 Ex 3.7.3 $N = 1000$, $n = 20$, $k = 100$

$P(X \geq 3) = 1 - P(X<3)$

$= 1 -$

$\frac{n}{N} = \frac{20}{1000} = \frac{1}{50} < 0.05$, So, approxmⁿ is good.

So, $n_{binomial} = 20$, $p = \frac{k}{N} = 0.1$

$\Rightarrow f(x) = {}^{20}C_x (0.1)^x (0.9)^{20-x}$.

$\Rightarrow P(X \geq 3) = 1 - [f(0) + f(1) + f(2)]$

$= 1 - 0.6769$

$= 0.3231$

$$\begin{array}{r} 1.0000 \\ -\,0.6769 \\ \hline 0.3231 \end{array}$$

★ POISSON Distribⁿ :-

A random variable X is said to have Poisson's distribⁿ with parameter k, if $f(x)$ is given by

$$f(x) = \frac{e^{-k} k^x}{x!} \quad \left\{ \begin{array}{l} x = 0, 1, 2 \ldots \\ k > 0 \end{array} \right\}.$$

Note :- $E(X) = k$

$Var(X) = k$

$m.g.f = M_X(t) = e^{k(e^t - 1)}$

ex: The no. of monthly breakdowns of a comp. is a random variable having Poison's distribⁿ with mean $= 1.8$. Find P(this comp. fⁿ for a month)

(i) without a breakdown

(ii) atleast 1 breakdown.

Puffin
Date
Page

X : no. of breakdowns of comp ($k = 1.8$, mean)

$f(x) = \frac{(e^{-k}) k^x}{x!}$

$= \frac{(e^{-1.8})(1.8)^x}{x!}$ $x = 0$

(i) $f(x=0) = e^{-1.8} = 0.1653$
(ii) $P(X \geq 1) = 1 - P(X=0) = 1 - 0.1653 = 0.8347$

Q. ~~X~~ $var(X) = 1 = k$
(a) Mean = 1
(b) $P(X \leq 2) = f(0) + f(1) + f(2)$
$= \frac{1}{e}\left[1 + 1 + \frac{1}{2}\right] = \frac{5}{2e}$

Pg - 72)
Q.61) ✓

* Applied Poisson Problems:
The steps used in the solⁿ of an applied Poisson problems are:
S1. Determine the basic unit of measurement being used.
S2. Find the avg. no. of occurences of the event per unit. This No. is denoted by λ.
S3) Identify the length / size of observⁿ period denoted by s.
S4) The random variable X, the no. of occurences of the event in the given period follows a Poisson's distribⁿ with a parameter

$\boxed{k = \lambda s.}$

Pg-93 Q.63 $s = 2$

unit: no/cm^2

$\lambda = 5$

$k = 10$

$P(X \leq 3)$

$= f(0) + f(1) + f(2) + f(3)$

$f(x) = \left(\frac{e^{-k} \; k^{x}}{x!}\right) = \left(\frac{e^{-10} \times 10^{x}}{x!}\right)$

$\cancel{(x=0)+(x=1)+(x=2)+(x=3)}$

Q65) $X = 2$, $\lambda = 10$, $s = 3$

Date: 5/3/12

Chapter: 4

CONTINUOUS DISTRIB^n : Replace Σ for discrete by $\int$ for cts

★ A random variable is cts if it can assume any value in some interval, or intervals of real nos. & the probability that it assumes any specific value = 0.

A f^n 'f' is called a density f^n for x if (i) $f(x) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Note: $P(a \leq x \leq b) = \int_a^b f(x)\,dx$.

★ Cumulative distrib^n f^n (c.d.f)

Let X be a cts random variable with density f. Then, c.d.f for X, denoted by F is defined by

$$F(x) = P(X \leq x)$$

- To compute $F(x) = P(X \leq x)$
$$= \int_{-\infty}^{x} f(x)\,dx$$
- To obtain f from F :-
$$f(x) = F'(x).$$
- Mean : $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$.

* In cts. distrib^n, x cannot take an exact value. It always take range of values.

- Variance : $Var(X) = \sigma^2 = E(X^2) - [E(X)]^2$.

Pg-138 Q.1) $f(x) = kx$, $x \in [2, 4]$

(a) $\int_{-\infty}^{\infty} f(x) = 1$

$\Rightarrow \int_{2}^{4} kx\,dx = 1$

$\Rightarrow k\int_{2}^{4} x\,dx = 1$

$\Rightarrow k\left(\frac{x^2}{2}\right)_2^4 \Rightarrow k(8-2) = 1$

$\Rightarrow k = \frac{1}{6}$

(b) $P(2.5 \leq x \leq 3)$

$\int_{2.5}^{3} f(x)\,dx = \int_{2.5}^{3} kx\,dx = \frac{1}{6}\left[\frac{x^2}{2}\right]_{2.5}^{3}$

(b) $= \frac{1}{6}\left[\frac{9}{2} - \frac{6.25}{2}\right] = \frac{2.75}{12}$

(c) $P(X = 2.5) = 0$

(d) $\int_{2.5}^{3} f(x) = \frac{2.75}{12} = 0.2291$

Q.3 $f(x) = \left(\frac{1}{10}\right) e^{-x/10}$, $x > 0$.

(a) ✓ $\int_{0}^{\infty} \left(\frac{1}{10}\right) e^{-x/10}\,dx$ $\quad e^{-x/10} = t \Rightarrow -\frac{1}{10} e^{-x/10}\,dx = dt$

$= 10\int_{1/10}^{0} -dt = -\left[0 - \frac{1}{10}\right] = +\frac{10}{10} = 1$

So its a density $1 - P(X \leq 6)$

& $P(X \leq 7)$, $P(X \geq 7)$, $P(X = 7)$ — for cts random variable it can only take range of values.

$\int_0^7 f(x)dx$ $\quad \int_7^{\infty} f(x)dx$ $\quad \to 0$ $\quad \left(\int_7^7 f(x)dx = 0\right)$

(c) $P(1 \leq x \leq 2) = \int_1^2 \left(\frac{1}{10}\right) e^{-x/10} dx$

$= \frac{1}{10} \left(\frac{e^{-x/10}}{-1/10} \right)$

$= \left[(-1) e^{-x/10} \right]_1^2$

$= -e^{-2/10} + e^{-1/10}$

$= \frac{1}{e^{1/10}} - \frac{1}{e^{2/10}}$

Unusual :- ∵ P is very less.

$= \frac{e^{2/10} - e^{1/10}}{e^{3/10}} = 0.086$

Pg-140

Q 9) (a) $f(x) = \frac{x}{6}$, $2 \leq x \leq 4$.

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{12}(x^2-4) & 2 \leq x \leq 4 \\ 1 & x > 4 \end{cases}$$

★

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

$= \frac{1}{6}\int_2^{x} x\,dx$

$= \frac{1}{12}(x^2 - 4)$

$= 1$

(b) $P(2.5 \leq x \leq 3)$

$= F(3) - F(2.5)$

M1 $= \frac{1}{12}(9-4) - \frac{1}{12}(6.25-4)$

$= \frac{5 - 2.25}{12} = \frac{2.75}{12} \to 0.2292$

Ans : it is same as earlier

M2 $\int_{2.5}^{3} f(x)dx$.

(c) $F'(x)$ or $\Rightarrow f(x) = \begin{cases} \frac{x}{6}, & x \in [2,4] \\ 0 & x < 2 \ \& \ x > 4 \end{cases}$

Pg-141

Q.15) $f(x) = \frac{x}{6}$, $x \in [2,4]$.

(a) $E(X) = \int_2^4 x f(x)\, dx$ (c) $\sigma^2 = (b) - [(a)]^2$

(d) $\sqrt{(c)} = \sigma$ Ans

(b) $E(X^2) = \int_2^4 x^2 f(x)\, dx$ $= 10$

✲ UNIFORM DISTRIBⁿ.

The distribⁿ of a cts. random variable X is said to follow uniform distribⁿ with parameters α, β, if its density fⁿ is defined by

- $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & , \alpha < x < \beta \\ 0 & , \text{otherwise} \end{cases}$

→ constt (∵ it is uniform distribⁿ)

- Mean $= \mu = \frac{\alpha+\beta}{2}$

- Variance $(\sigma^2) = \frac{(\beta-\alpha)^2}{12}$ ✲ m.g.f $= M_x(t) = \frac{e^{t\beta} - e^{ta}}{t(\beta-a)}$

ex:- Buses arrive at a specified stop at 15 min intervals starting at 7 a.m. If a passenger arrives at the stop at a random time that is uniformly distributed b/w 7 & 7:30, find P that he waits

(a) < 5 mins. for bus

(b) atleast 12 min.

$\beta = 30$, $\alpha = 0$

$f(x) = \begin{cases} \frac{1}{30} & , x \in (0,30) \\ 0 & , \text{else} \end{cases}$

(a) Reqd. $P = P(10 < X < 15) = F(15) - F(10)$
or $P(25 < X < 30)$

$$P = \int_{10}^{15} \frac{1}{30}\,dx + \int_{25}^{30} \frac{1}{30}\,dx = \frac{1}{3}$$

(b) $$P(X \geq 12) = \int_{0}^{3} \frac{1}{30}\,dx + \int_{15}^{18} \frac{1}{30}\,dx = \frac{1}{5}$$

§ GAMMA FUNCTION

The f^n is denoted by Γ and defined by

$$\boxed{\Gamma_\alpha = \int_0^\infty z^{\alpha-1} e^{-z}\,dz} \quad , \quad \alpha > 0.$$

Properties :-

1) $\Gamma_1 = 1$
2) For $\alpha > 1$, $\Gamma_\alpha = (\alpha - 1)\,\Gamma_{(\alpha-1)}$
3) $\Gamma_{1/2} = \sqrt{\pi}$
4) $\Gamma_{\alpha+1} = \alpha!$ eg $\Gamma_5 = 4\Gamma_4 = 4.3\Gamma_3 = 4.3.2\Gamma_2$
$= 5!$

eg :- $$\int_0^\infty x^3 e^{-x/2}\,dx$$

Put $z = x/2 \Rightarrow x^3 = 8z^3$
$dx = 2dz$

$$= \int_0^\infty 8z^3 e^{-z}(2dz) = 16\int_0^\infty z^3 e^{-z}\,dz$$

$\alpha = 4$.

So, $$16\int_0^\infty z^3 e^{-z}\,dx = 16 \times \Gamma_4 = 16(3!) = 96$$

(a) Reqd. P = P(10 < X < 15) = F(15) − F(10)
or P(25 < X < 30)

$$P = \int_{10}^{15} \frac{1}{30}\, dx + \int_{25}^{30} \frac{1}{30}\, dx = \frac{1}{3}$$

(b) $P(X \geq 12) = \int_{0}^{3} \frac{1}{30}\, dx + \int_{15}^{18} \frac{1}{30}\, dx = \frac{1}{5}$

§ GAMMA FUNCTION

The f^n is denoted by Γ and defined by

$$\boxed{\Gamma_\alpha = \int_0^\infty z^{\alpha-1} e^{-z}\, dz}, \quad \alpha > 0.$$

Properties :-

1) $\Gamma 1 = 1$
2) For $\alpha > 1$, $\Gamma \alpha = (\alpha - 1)\, \Gamma(\alpha - 1)$
3) $\Gamma \frac{1}{2} = \sqrt{\pi}$
4) $\Gamma(\alpha + 1) = \alpha !$ eg $\Gamma 5 = 4 \Gamma 4 = 4.3 \Gamma 3 = 4.3.2 \Gamma 2$
$= 5!$

eg:- $\int_0^\infty x^3 e^{-x/2}\, dx$

Put $z = x/2 \Rightarrow x^3 = 8z^3$
$dx = 2dz$.

$$= \int_0^\infty 8z^3 e^{-z} (2dz) = 16 \int_0^\infty z^3 e^{-z}\, dz.$$

$\alpha = 4$.

So, $16 \int_0^\infty z^3 e^{-z}\, dx = 16 \times \Gamma 4 = 16(3!) = 96$

Q $\int_0^\infty z^3 e^{-z} dz = \Gamma 4 = 3! = 6$

Q $\int_0^\infty \frac{1}{54} x^2 e^{-x/3} dx = \frac{1}{2} \int_0^\infty z^2 e^{-z} dz$
$= \frac{1}{2} \Gamma 3 = 1$

$\frac{x}{3} = z \Rightarrow (3dz)$

✲ GAMMA DISTRIB^n

A random variable X with density

$$\boxed{f(x) = \frac{1}{\Gamma\alpha \, \beta^\alpha} x^{\alpha-1} e^{-x/\beta}} \; ; \; x>0, \quad \alpha, \beta > 0$$

is said to have Gamma distrib^n with parameters α & β.

- Mean :- $\mu = \alpha\beta$

Variance :- $\alpha\beta^2$

m.g.f $= M_x(t) = (1-\beta t)^{-\alpha}$; $t < \frac{1}{\beta}$.

Pg-144 Q.29) $\alpha = 3, \beta = 4$; (a)✓ (b)✓ (c)✓

✲ EXPONENTIAL DISTRIB^n :-

The distrib^n of a cts. random variable X is exponential with parameter β if

$$\boxed{f(x) = \frac{1}{\beta} e^{-x/\beta}} \; ; \; \beta > 0, \; x > 0$$

↳ (Put $\alpha = 1$ in Gamma distrib^n, you get Exponential distrib^n)

Mean : $\mu = \beta$

Variance : β^2

m.g.f $= M_x(t) = (1-\beta t)^{-1}$; $t < \frac{1}{\beta}$

} Put $\alpha = 1$ $\forall$ formulas in Gamma Distrib^n.

Date
Page

* Note:
Consider a Poisson process with parameter λ.
Let W denote the time of occurence of 1st event. Then, W has exponential distribⁿ with

$\boxed{\beta = \frac{1}{\lambda}} = k$ (Taken previously).

Pg-144) Q.35) Let X denote time to be waited to see 1st lightning strike $\Rightarrow k = 2$

$\beta = \frac{1}{k} \Rightarrow \beta = 0.5$

So. p.d.f $= f(x) = \frac{1}{\beta} e^{-x/\beta}$

$\Rightarrow f(x) = \frac{1}{0.5} e^{-x/0.5} = \begin{cases} 2e^{-2x} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$

Reqd. $P = P(X \leq 1)$

$= \int_0^1 f(x)\,dx = \int_0^1 2e^{-2x}\,dx$

$= 2\left[\frac{e^{-2x}}{-2}\right]_0^1 = -e^{-2} + e^0 = 1 - e^{-2}$
$= 0.867$

Q.37) $k = \lambda = 1$

$\beta = 1/1 = 1$

X denotes time for 1st destructive earthquake to happen

So. $f(x) = \frac{1}{1} e^{-x/1} = e^{-x}$

$(0.25 = 3/12)$

$P(X \geq 0.25) = 1 - P(X < 0.25)$

$= 1 - \int_0^{0.25} e^{-x}\,dx = 1 + \left[e^{-x}\right]_0^{0.25}$

$= 1 + e^{-0.25} - 1$

$= \frac{1}{e^{0.25}} = 0.778$

Puffin
Date
Page

Q. 36) $k = 1$, $\beta = 1/3$

$f(x) = \frac{1}{1/3} e^{-x/1/3} = 3e^{-3x}$

$P(X \geq \frac{30}{60}) = \int_{1/2}^{\infty} 3e^{-3x} dx$

☆ Chi Squared Distribⁿ

Let X be a gamma variate with $\beta = 2$ & $\alpha = \frac{\gamma}{2}$

γ: +ve $\mathbb{Z}$.

Then, X has chi squared distribⁿ with γ degrees of freedom.

χ^2_γ : symbol for this random variable.

p.d.f of χ^2_γ for γ degree of freedom is given by

$$f(\chi^2) = \begin{cases} \frac{1}{2^{\gamma/2}\Gamma\frac{\gamma}{2}} (\chi^2) e^{-\frac{\chi^2}{2}} & ; \chi^2 > 0 \\ 0 & ; \chi^2 \leq 0 \end{cases}$$

Note:-

Mean $= \mu = E(\chi^2) = \gamma$

Variance $(\sigma^2) = 2\gamma$

$M_x(t) = (1-2t)^{-\gamma/2}$

(Pg-145)

Q.38 $\gamma = 15$

(a) Mean $= \gamma = 15$, Variance $= 2\gamma = 30$.

(b) $f(\chi^2_{15}) = \begin{cases} \frac{1}{2^{\gamma/2}\Gamma(\gamma/2)} (\chi^2) e^{-\chi^2/2} & ; \chi^2 > 0 \\ 0 & ; \chi^2 \leq 0 \end{cases}$

$\gamma/2 \to 7.5$

(c) $M_x(t) = (1-2t)^{-\gamma/2} = (1-2t)^{-15/2}$

Q.17

(Pg.141) $f(x) = \frac{1}{10} e^{-x/10}$: exponential; $\beta = 10$.

(a) $M_x(t) = (1-\beta t)^{-1}$

$= (1-10t)^{-1}; t < \frac{1}{10}$

$E(X) = \frac{d}{dt}(M_x(t))$

$= (-1)(1-10t)^{-2}(-10)$

$= (-1)(1)(-10)$

$\Rightarrow E(X) = 10$

$var(X) = \beta^2 = 100$

So, std. deviation $= \sqrt{\beta^2} = 10$

✯ Normal Distribution

The distrib^n of a random variable X is known as normal distrib^n with parameters μ & σ if its p.d.f ($f(x)$) is given by.

$\mu \to$ Mean, $\sigma \to$ Std. deviation

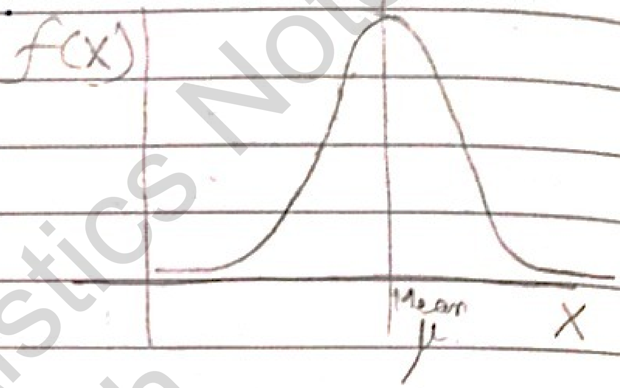
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left(\frac{(X-\mu)^2}{2\sigma^2}\right)}; \quad \sigma > 0, x, \mu \in (-\infty, \infty)$$

We say X is a $N(\mu, \sigma^2)$ variate.

Note :- $Var(X) = \sigma^2$.

① $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$

② The graph of the density of a normal random variable is symmetric, bell shaped & centered at its mean.



★ Standard Normal Distribⁿ

done because table V (Pg-697) has values of only Std. Normal Distribⁿ

The normal distribⁿ with $\mu = 0$ & $\sigma = 1$ is called Std. Normal Distribⁿ. Its denoted by the letter 'Z'.

If X is a normal variate with parameters μ & σ, then, corresponding std. variate is given by

$$z = \frac{x-\mu}{\sigma}.$$

★ Normal Probability Rule :

Let X be normally distributed with parameters μ & σ, then,

$$P(-\sigma < X-\mu < \sigma) = 0.68$$
$$P(-2\sigma < X-\mu < 2\sigma) = 0.95$$
$$P(-3\sigma < X-\mu < 3\sigma) = 0.997$$

Symbol :- $P(Z \geq Z_r) = r$.

r : Probability ; Z_r : Pt. associated with a std. normal random variable

Pg. 145

Q39) (a) $P(Z \le 1.57) = 0.9418$

(b) $P(Z < 1.57) = 0.9418$ (equality & inequality are same for cts distribⁿ)

(c) $P(Z = 1.57) = 0$

(d) $P(Z > 1.57) = 1 - P(Z \le 1.57) = 1 - 0.9418 = 0.0582$

(e) $P(-1.25 \le Z \le 1.75) = P(Z \le 1.75) - P(Z \le -1.25)$
$= 0.9599 - 0.1056 = 0.8543$

(g) $P(Z \geq Z_{0.9}) = 0.9$ ($\because P(Z \geq Z_r) = r$)

$P(Z \le Z_{0.9}) = 1 - P(Z > Z_{0.9})$
$= 1 - 0.9 = 0.1$

$0.1000 \approx 0.1003 = P(Z \le -1.28)$ (from table)

So, $P(Z \le Z_{0.9}) \approx P(Z \le -1.28)$

So, $Z_{0.9} = -1.28$

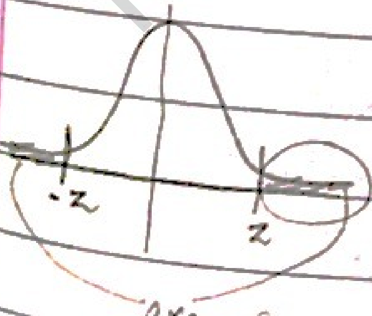
(f) $Z_{0.10}$? We know $Z_{1.0} = 0$.

$\Rightarrow Z_{0.10} = Z_1 - Z_{0.9} = 0 - (-1.28) = 1.28$

(h) Find pt z s.t. $P(-z \le Z \le z) = 0.95$

~~From table~~

$P(-z \le Z \le z) = F(z) - F(-z)$
$= F(z) - (1 - F(z)) = 0.95$
$\Rightarrow F(z) = \cancel{\frac{0.95}{2}} \; \frac{1.95}{2} = 0.975$



area is same

$\Rightarrow F(z) = F(1.96)$
$\Rightarrow z = 1.96$.

So, $F(-z) = 1 - F(z)$

$\rightarrow$ total area

Puffin
Date
Page

Q 43) X : normally distributed

$\mu = 153$
$S.D. = \sigma = 25$

(a) $P(128 \leq x \leq 178) = P\left(\frac{128-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{178-\mu}{\sigma}\right)$

$= P\left(\frac{128-153}{25} \leq Z \leq \frac{178-153}{25}\right)$

$= P(-1 \leq Z \leq 1)$

$= F(1) - F(-1)$

$= 0.8413 - 0.1587 = 0.6826$

So, reqd %age = 68.26%

(b) $P(X > 228) = P\left(\frac{X-\mu}{\sigma} > \frac{228-153}{25}\right)$

$= P(Z > 3)$

$= 1 - P(Z < 3)$

$= 1 - 0.9987 = 0.0013 = 0.13\%$

* Calculation of Probability for normal variance X with mean μ & variance σ^2.

RESULTS

1) $P(X <= a) = F\left(\frac{a-\mu}{\sigma}\right)$

2) $P(X > b) = 1 - P(X < b)$
$= 1 - F\left(\frac{b-\mu}{\sigma}\right)$

3) $P(a \leq X \leq b) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$

4) $P(|X| > a) = 2P(X < -a) = 2F\left(\frac{-a-\mu}{\sigma}\right)$

5) $P(|X| < a) = 1 - P(|X| > a) = 1 - 2F\left(\frac{-a-\mu}{\sigma}\right)$

Puffin
Date
Page

Pg-145) Q 41) Given $X \to N(\mu, \sigma)$
$\to N(0, 100)$

(a) Find $P(|X| < 200) = 1 - 2F\left(\frac{-200-0}{100}\right)$
for distance
$= 1 - 2(0.0228)$
$= 0.9544$

(b) $P(|X| > 250) = 2F\left(\frac{-a-\mu}{\sigma}\right) = \cancel{0.0456}$
$2F(-2.5)$
$= 0.0124 = 1.24\%$

(c) $P(|X| \geq x_1) = 2F\left(\frac{-x_1 - 0}{100}\right) = 20\%$
$= 0.2$
$\Rightarrow F\left(\frac{-x_1}{100}\right) = 0.1 = F(-1.28)$
$\Rightarrow \frac{-x_1}{100} = -1.28$
$\Rightarrow x_1 = 128$ parsecs

(d) $e^{0.t + \sigma^2 t^2/2} = e^{10^4 t^2/2}$

Q. 43) Pg-146) $\mu = 10, \sigma = 3$
(a) $P(X < 15) = 0.9525$ (Table)
(b) $P(X < hrs) = 0.05 = \Phi$
$\Rightarrow t = 5.065$ hrs

✳ NORMAL APPROXIM^n to BINOMIAL DISTRIB^n :
Let X be binomial, with parameters n, p ; for large n , X is approximately normal, with mean np & variance npq. The approxim^n is acceptable only if $p \leq 0.5$ & $np > 5$ or
$p > 0.5$ & $n(1-p) > 5$.

$P(X \geq 5) = P(Y \geq 4.5) = 1 - P(y < 4.5)$
$= 1 - P(Z < \frac{4.5-\mu}{\sigma})$
$= 1 - F(\frac{4.5-\mu}{\sigma})$

Pg. 148

Q.153) X has a normal distrib^n:-
$n = 150$; $p = 0.05$
Verify:- ✓
Using normal approxim^{n}:-

Note: When ~~Binom~~ Binomial variate is approxim-ated by normal variate ~~Y~~ y, then,
$X \longrightarrow y$
$P(x > a) = P(y \geq a + 0.5)$
$P(X \geq a) = P(Y \geq a - 0.5)$
$P(X < a) = P(Y \leq a - 0.5)$
$P(X \leq a) = P(Y \leq a + 0.5)$

Let Y be the corresponding normal variate with parameters μ & σ.
$\mu = np = 7.5$
$\sigma = \sqrt{npq} = 2.67$
Reqd prob. $= P(X \geq 1)$
$= P(Y \geq a - 0.5)$
$= P(Y \geq 1 - 0.5)$
$= P(Y \geq 0.5)$ (normal variate)
$= P\left(\frac{Y-\mu}{\sigma} \geq \frac{0.5-\mu}{\sigma}\right)$
$= P\left(Z \geq \frac{-7}{2.67}\right)$
$= P(Z \geq -2.62)$
$= 1 - F(-2.62)$
$= 1 - 0.0044 = 0.9956$

(b) $P(X \leq 3) = P(Y \leq 3+0.5)$

$= P(Y \leq 3.5)$

$= F\left(\frac{3.5-\mu}{\sigma}\right) = F\left(\frac{-4}{2.67}\right) = \cancel{0.096}\ 0.0681$

Pg. 146) Q.44) $\mu = 1876$ ft

$\sigma = 6$ inches $= 0.5$ ft.

(a) $P(X \leq 1875) = P\left(\frac{X-\mu}{\sigma} \leq \frac{1875-\mu}{\sigma}\right) = P\left(Z \leq \frac{1875-1876}{0.5}\right)$

$= 2.28\%$

(b) $P(X \geq 1878) = P(Z \geq 4)$ Unusual

$= F(-4) = 0$

☆ CHEBYSHEV's inequality:

Let X be a random variable with mean μ & S.D $\rightarrow \sigma$. Then, for any +ve no. 'k',

$P[|X-\mu| < k\sigma] \geq 1 - \frac{1}{k^2}$ → +ve no.

ex. A fair die is tossed 720 times. Use Chebyshev's inequality to find a lower bound for the probability of getting 100-140 6's.

$n = 720,\ p = \frac{1}{6},\ \mu = np = 120,\ \sigma = \sqrt{npq} = 10$

By inequality

$P[|X-120| < 10k] \geq 1 - \frac{1}{k^2}$

i.e

$P(-10k \leq X-120 \leq 10k) \geq 1 - \frac{1}{k^2}$

$\Rightarrow P(120-10k \leq X \leq 10k+120) \geq 1 - \frac{1}{k^2}$

no of sixes b/w 100 & 140

So, $120 - 10k = 100$ & $10k + 120 = 140$

$\Rightarrow k = 2$ & $k = 2$

$\Rightarrow P(100 \leq X \leq 140) \geq 1 - \frac{1}{2^2} \geq 0.75$

So $P = 0.75$

Puffin
Date
Page

ex. Let X be a random variable, $\mu = 4$ & $\sigma = 1$. Use Chebyshev's inequality to find min. value for $P(2 < X < 6)$

$$P[\,|X-\mu| < k\sigma\,] \geq 1 - \frac{1}{k^2}$$

$$\Rightarrow P(-k \leq x-4 \leq k) \geq 1 - \frac{1}{k^2} \quad k=2$$

$$\Rightarrow P(2 \leq x \leq 6) \geq 1 - \frac{1}{4} = 0.75 \text{ ans}$$

* WEIBULL DISTRIB^n :

A random variable X is said to have Weibull distrib^n with parameters α & β if its density f^n is given by

$$f(x) = \alpha\beta x^{\beta-1} \cdot e^{-\alpha x^\beta} \quad ; x>0 ; \alpha, \beta > 0.$$

* To prove :- f is a density

$$\int_0^\infty f(x)\,dx = \int_0^\infty \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}\,dx$$

Let $\alpha x^\beta = t \Rightarrow \alpha\beta x^{\beta-1}dx = dt$

$$\Rightarrow \int_0^\infty e^{-t}\,dt = -\left[e^{-t}\right]_0^\infty = -[0-1] = 1 \quad \text{H.P.}$$

* Mean $= \mu = \alpha^{-1/\beta}\,\Gamma\left(1+\frac{1}{\beta}\right)$

* Var, $\sigma^2 = \alpha^{-2/\beta}\,\Gamma\left(1+\frac{2}{\beta}\right) - \mu^2$.

Pg-149) Q.58) (a) $f(x) = (0.01).2x^1 e^{-0.01x^2}$

$$= \begin{cases} 0.02\,x\,e^{-x^2/100} & , x \geq 0 \\ 0 & x < 0 \end{cases}$$

(b) $\mu = (0.01)^{-1/2}\,\Gamma\,3/2 = 10\left(\frac{1}{2}\right)\Gamma(1/2) = 5\sqrt{\pi}$

$\sigma^2 = (0.01)^{-1}\,\Gamma 2 - 25\pi = 100 - 25\pi = 25(4-\pi)$

14-15 Q 37)(e) $H(X) = X^2 + 3X + 2$

$E(H(X)) = ?$ $\alpha = 2, \beta = 1$

$var = \underbrace{E(X^2)}_{?} - [E(X)]^2$

μ

$E(H(X)) = E(X^2 + 3X + 2) = E(X^2) + 3E(X) + E(2)$

3μ 2

Puffin
Date 20/3/12
Page

Chapter-5
JOINT DISTRIBUTIONS

- 2D random variables X, Y.
  Discrete — Cts.

• **Discrete Joint Density:**
Let X, Y be 2 discrete random variables. The joint density for (X, Y) is defined by

$$f_{XY}(x, y) = P(X = x \ \& \ Y = y)$$

(for 1D, $f(x) = P(X = x)$)

* The cond^ns for a f^n to be discrete joint density is
1. $f_{XY}(x, y) \geq 0$

2. $\sum_{\text{all } x} \sum_{\text{all } y} f_{XY}(x, y) = 1$.

* **Marginal Density (discrete)**
The marginal densities are individual densities of X & Y denoted by
f_X & f_Y resp. Mathematically,

$$\boxed{f_X(x) = \sum_{\text{all } y} f_{XY}(x, y), \quad f_Y(y) = \sum_{\text{all } x} f_{XY}(x, y)}$$

Pg-180) Q.1) X : no. of defective welding
Y : no. of ~~+~~ improperly tightening bolts

(a) $f(2,1) = 0.005$
$P(X = 2, Y = 1) = 0.005$

(b) $P(X \geq 1, Y \geq 1) = f(1,1) + f(1,2) + f(1,3) = 0.0$
$+ f(2,1) + f(2,2) + f(2,3)$

$= f_X(0) + f_X(1)$

(c) $P(X \leq 1) = 0.3(10 + 0.050 + 0.020 + 0.01 + 0.060 + 0.01 + 0.008 + 0.003) = 0.78$

(d) $= 0.32 + 0.13 = 0.45 = P(Y \geq 2)$

Pg-181

Q.5) X : Syntax error

Y : logic error

(a) 0.4

(b) $P(X \geq 1, Y \leq 1) = 0.429$

(c) Shown in Pg-182.

(f) $f(0,0) = f_X(0) \cdot f_Y(0)$

$\Rightarrow 0.4 \neq 0.525 \times 0.762 \Rightarrow 0.4 \neq 0.40005$

So, not independent

(d) $P(X \geq 2) = 0.123$

(e) $P(Y = 1) + P(Y = 2) = 0.22$

★ Note :- X & Y are said to be independent, if,

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

- Continuous Joint Density :

The its joint density of 2 D random variable X & Y is $f_{XY}(x,y)$ with following properties:

1. $f_{XY}(x,y) \geq 0$ for $-\infty < x < \infty$, $-\infty < y < \infty$

2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\, dy = 1$

3) $P(a \leq x \leq b, c \leq y \leq d) = \int_a^b \int_c^d f_{XY}(x,y)\, dx\, dy$

Puffin
Date
Page

* Marginal Density (Continuous)

Marginal density for X & Y are defined by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy.$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

ex: Verify $f_{XY}(x,y) = \frac{x^3y^3}{16}$, $x \in [0,2]$, $y \in [0,2]$ is a joint density

$$f_X(x) = \int_0^2 f_{XY}(x,y)\,dy$$

$$= \frac{x^3}{16}\int_0^2 y^3\,dy$$

$$= \frac{x^3}{16}\left[\frac{y^4}{4}\right]_0^2 = \frac{x^3}{4}, \quad 0 \le x \le 2$$

$$f_Y(y) = \frac{y^3}{16}\int_0^2 \frac{x^3}{16}\,dx = \frac{y^3}{4} ; \quad 0 \le y \le 2$$

To verify X & Y are independent

So. $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

$$\Rightarrow \frac{x^3y^3}{16} = \frac{x^3}{4}\cdot\frac{y^3}{4} = \frac{x^3y^3}{16}$$

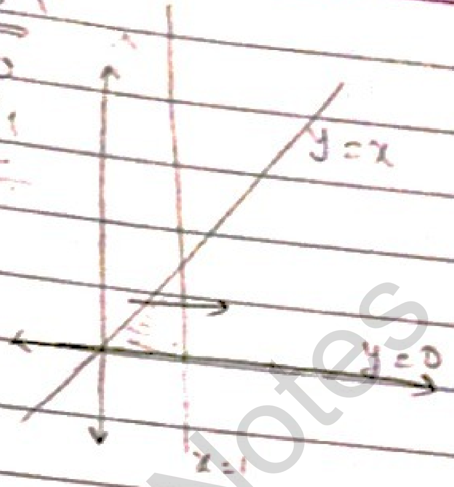
So, x & y are independent

ex:- Verify $f_{XY}(x,y)$ is a density (joint)

$f_{XY}(x,y) = \frac{x^3y^3}{16}$ i.e. $\iint f_{XY}(x,y)\,dx\,dy = 1$

Q.9 $f_{XY}(x,y) = \frac{1}{x}$, $0 < y < x < 1$

(a) $f_{XY}(x,y) = \frac{1}{x} > 0 \ \forall \ 0 < y < x < 1$



$\iint_R f_{XY}(x,y)\,dx\,dy = 1$

$= \iint_R \frac{1}{x}\,dx\,dy$

$= \int_{y=0}^{y=1} \left\{ \int_{x=y}^{x=1} \frac{1}{x}\,dx \right\} dy = \int_{y=0}^{1} [\ln x]_y^1\,dy$

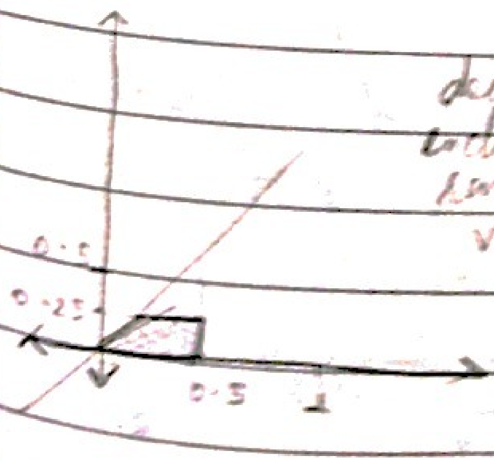
$= \int_{y=0}^{1} [-\ln y]\,dy$

$= [-y \ln y + y]$

$= [y[1 - \ln y]]_0^1$

$= 1$ So its joint density

(b) $P(X \leq 0.5 \ \& \ Y \leq 0.25) = \iint f_{XY}(x,y)\,dx\,dy$

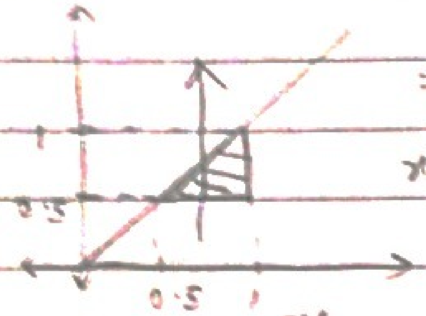


doesn't include range of values $\leftarrow$ { Area of shaded region $= \frac{1}{2} \times (0.25)(0.25 + 0.5)$

Or $\int_{y=0}^{0.25} \left\{ \int_{x=y}^{x=0.5} \left(\frac{1}{x}\,dx \right) \right\} dy = 0.4233$

(c) $P(X > .5 \text{ or } Y > .25) = 1 - \text{Part (b)}$
$= 1 - 0.423$ (See graph)

(d) $P(X > 0.5 \;\&\; Y > 0.5)$



$$= \int_{x=0.5}^{1} \int_{y=0.5}^{x} dy \left\{ \frac{1}{x} dx \right. = 0.153$$

$0.5 < y < x < 1$
varies from

(e) $f_X(x) = \int_0^x f_{XY}(x,y)\, dy = \frac{1}{x}\int_0^x dy = \frac{x}{x} = 1$, $0 < x < 1$

$f_Y(y) = \int_y^1 f_{XY}(x,y)\, dx = [\ln x]_y^1 = -\ln y$, $0 < y < 1$

(f) $P(X < 0.5) = \int_0^{0.5} (1)\, dx = 0.5$

→ Marginal density for X

(g) $P(Y < 0.25) = \int_0^{0.25} (-\ln y)\, dy = -(y \ln y - y)_0^{0.25} = 0.597$

(h) $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

$\Rightarrow \frac{1}{x} \neq (1)(-\ln y)$

So, not independent.

☆ EXPECTATION & COVARIANCE

Let X, Y be a 2D random variable with a joint density f_{XY} & $H(x,y)$ is any f^n of X & Y. Then,

☆ $E(H(X,Y)) = \sum_{\text{all } x} \sum_{\text{all } y} H(x,y) f_{X,Y}(x,y) \rightarrow$ discrete

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y) f_{XY}(x,y)\, dy\, dx \rightarrow$ cts

$E(x) = \int x f(x)\, dx$

$E(X) = \iint x\, f_{xy}(x,y)\, dx\, dy$

Univariate mean through Joint density

If the joint density for (X,Y) is known, then, avg. value of X & Y can be found as follows

$$\left.\begin{aligned} E(X) &= \sum_{\text{all } x}\sum_{\text{all } y} x \cdot f_{XY}(x,y) \\ E(Y) &= \sum_{\text{all } x}\sum_{\text{all } y} y \cdot f_{XY}(x,y) \end{aligned}\right\} \text{DISCRETE}$$

$$\left.\begin{aligned} E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot f_{XY}(x,y)\, dx\, dy \\ E(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{XY}(x,y)\, dx\, dy \end{aligned}\right\} \text{CONTINUOUS}$$

IIly for $E(XY) = \sum\sum xy\, f_{XY}(x,y)$ or $\iint xy\, f_{XY}(x,y)\, dx\, dy$

- To find $\text{cov}(X,Y) = E(XY) - E(X) \cdot E(Y)$

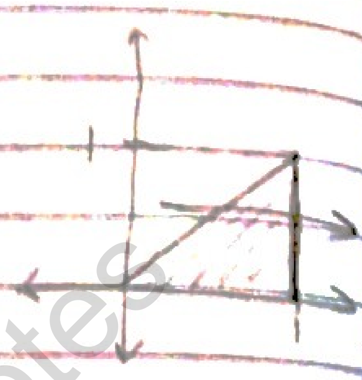
Conditional Densities

Conditional Densities: Let (X,Y) be a 2D random variable with a joint density f_{XY} & marginal densities f_X & f_Y. Then, the cond^{nal} density for X, given $Y=y$ is denoted by

$$f_{X/Y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} \qquad = P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

IIly, $$f_{Y/X}(y) = \frac{f_{XY}(x,y)}{f_X(x)}$$

Pg-185 Q.21) $f_{XY}(x,y) = \frac{1}{x}$; $0<y<x<1$



$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{XY}(x,y)\,dx\,dy$$

$$= \int_0^1 \left(\int_y^1 x \cdot \frac{1}{x}\,dx\right) dy$$

$$= \int_0^1 (1-y)\,dy = \left(y - \frac{y^2}{2}\right)_0^1 = 1 - \frac{1}{2} = \frac{1}{2}$$

$$E(Y) = \int_0^1\int_y^1 y\, f_{XY}(x,y)\,dx\,dy$$

$$= \int_0^1 y\left\{\int_y^1 \frac{1}{x}\,dx\right\} dy = \frac{1}{4}$$

$-y \ln y$

$= \frac{1}{4}$

$$E(XY) = \int_0^1\int_y^1 xy\,(dx\,dy)\, f_{xy}(x,y)$$

$$= \int_0^1 y\left\{\int_1^y dx\right\} dy$$

$$= \int_0^1 (y^2 - y)\,dy = \left(\frac{y^3}{3} - \frac{y^2}{2}\right)_0^1 = \frac{1}{3} - \frac{1}{2} = \frac{-1}{6}$$

$$Cov(X,Y) = E(XY) - E(X).E(Y)$$

$$= \frac{1}{6} - \frac{1}{8} = \frac{1}{24}$$

126) Q 29) $f_{X/Y}(x,y) = \frac{c}{x}$; $27 \le y \le x \le 33$

; $c = \frac{1}{6-27\ln 33/27}$

(a) $E(\cancel{X})$ Condnal density

$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy$

$= \int_{27}^{x} \frac{c}{x}\, dy = \frac{c}{x}(x-27) = c - \frac{27c}{x}$

$27 < x \le 33$

b $f_Y(y) = \int_{y}^{33} f_{XY}(x,y)\, dx$

$= c(\ln 33 - \ln y)$; $27 \le y \le 33$

$f_{X/Y=31}(x) = \frac{f_{XY}(x,31)}{f_Y(31)} = \frac{\cancel{c}/x}{\cancel{c}(\ln 33 - \ln 31)}$

$= \frac{1}{x(\ln 33 - \ln 31)}$ $31 \le x \le 33$

$f_{Y/X=30}(y) = \frac{f_{XY}(30,y)}{f_X(30)} = \frac{\cancel{c}/30}{\cancel{c}(1-\frac{27}{30})} = \frac{1}{3}$

$y \in [27, 30]$

$E\left(\frac{X}{Y=31}\right) = \int_{-\infty}^{\infty} x f_{X/Y=31}(x)\, dx$

$= \int_{31}^{33} x \frac{1}{x(\ln 33 - \ln 31)}\, dx$

$E\left(\frac{X}{Y=31}\right) = \frac{2}{\ln(33/31)}$

Puffin
Date
Page

$$E\left(\frac{y}{x=30}\right) = \int_{-\infty}^{\infty} y f_{y/x=30}(y)\,dy$$

$$= \int_{27}^{30} y\left(\frac{1}{3}\right)dy$$

$$= \frac{1}{6}\left(30^2-27^2\right) = \frac{1}{\cancel{6}}\left(\frac{\cancel{3}}{2}\right)(57) = \cancel{\frac{57}{2}}\; 28.5$$

Pg-181)
Pg-184) Q.2.15 Smaller value of X associated with
(a) larger value of Y. So, -ve covariance can come.

(b) $E(X) = \sum_{x=0}^{3}\sum_{y=0}^{4} x f_{XY}(x,y)$

$= [0.f(0,0) + 0.f(0,1) + \dots 0f(0,4)]$
$+[1.f(1,0) + 1.f(1,1) + \dots 1f(1,4)]$
$+ \dots [3.f(3,0) + \dots 3.f(3,4)]$

$= 0.\frac{1}{35} + 1.\frac{12}{35} + 2\times\frac{18}{35} + 3.\frac{4}{35}$

$= \cancel{\frac{60}{35}}\; \frac{12}{7}$

$E(Y) = \sum_{x=0}^{3}\sum_{y=0}^{4} y f_{xy}(x,y)$ — $\frac{4}{35} + \frac{36}{35} + \frac{36}{35} + \frac{4}{35}$
$= 80/35 = 16/7$

$E(XY) = \sum_{x=0}^{3}\sum_{y=0}^{4} xy f_{xy}(x,y)$

$= 1.1(0) + 1.2.(0) + 1.(3)(12/35)$
$+ 2(2)(18/35) + 3(1).4/35$

$= \frac{36}{35} + \frac{72}{35} + \frac{12}{35} = \cancel{\frac{120}{35}}\; \cancel{\frac{24}{35}}\; 24/7$

$cov(X, Y) = E(XY) - E(X) \cdot E(Y)$

$$= \frac{12}{3}\left[\frac{34\times7}{49} - \frac{13\times12}{49}\right]$$

$$= \frac{12[14 - 16]}{49} = \frac{-24}{49}$$

☆ SIMULATION

→ SIMULATING A DISCRETE DISTRIB^n

Random nos. are used for simulating a discrete random variable & cts random variable. We follow mathematical procedure to simulate such random variables.

ex: ~~Let~~ let X be a random variable with distrib^n →

x	0	1	2	3	4
f(x)	0.2	0.4	0.1	0.2	0.1

(a) Simulate 10 values of x, (b) find μ, (c) use following set of random nos. to simulate X : 5, 9, 7, 5, 0; 1, 7, 8, 2, 1

There are 10 chances out of 10 chance (choose from 0 – 9)

x	f(x)	Cumulative Prob.	Random nos.	
0	0.2	0.2	0 – 1	Assign 2 nos. out of 10
1	0.4	0.6	2 – 5	next 4 nos. out of 10
2	0.1	0.7	6	one no. out of 10
3	0.2	0.9	7 – 8	next 2 nos. out of 10
4	0.1	1	9	

(c) 5, 9, 7, 5, 0 ; 1, 7, 8, 2, 1

Simulated values: 1, 4, 3, 1, 0 ; 0, 3, 3, 1, 0

(b) $\mu = \frac{1+4+3+1+0+0+3+3+1+0}{10} = \frac{16}{10} = 1.6$

Pg. 94) Q.72) $X \longrightarrow B(5, 0.1)$

$$f(x) = {}^nC_x\, p^x q^{n-x}$$

$$\Rightarrow f(x) = {}^5C_x (0.1)^x (0.9)^{5-x}$$

X	f(x)	Cum. Prob	Random nos
0	0.59049	0.59049	0 – 59048
1	0.32805	0.91854	59049 – 91853
2	0.0729	0.99144	91854 – 99143
3	0.0081	0.99954	99144 – 99953
4	0.00045	0.99999	99954 – 99998
5	0.00001	1	99999

Choose the nos :- 70977, 49626, 88474, 48237, 77233, 77452, 59638, 31278, 23216, 42698

So, simulated values are 1, 0, 1, 0, 1, 1, 1, 0, 0, 0

$$\mu = \frac{1+0+1+0+1+1+1+0+0+0}{10} = 0.5 = np = (5 \times 0.1)$$

So, simulation is real simulⁿ

↳ 4.9. SIMULATING FOR A CTS. RANDOM VARIABLE

S1) Let X be a cts random variable with c.d.f $F(x)$

S2) Generate a random no. y in $0 < y < 1$

S3) Take $y = F(x)$

S4) Find $x = g(y)$

This X is a ~~simul~~ simulated fⁿ observⁿ.

Puffin
Date
Page

Q3)
Q 2) $\beta = 4, \alpha = 0$ → uniform cts distrib^n

$$f(x) = \frac{1}{\beta - \alpha} = \begin{cases} \frac{1}{4} & 0 < x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

S1) $$F(x) = \begin{cases} 0 & x < 0 \\ x/4 & 0 \le x \le 4 \\ 1 & x > 4 \end{cases}$$ (Integrating)

S2) Let $y = 0.4$

Let $y = F(x)$

$\Rightarrow y = x/4 \Rightarrow x = 4y \Rightarrow x = 4(0.4) = 1.6$

$\therefore$ One simulated observ^n for x is $x = 4(0.4) = 1.6$

||ly, 19 others

Hw

Q. Simulate 5 observ^{ns} of an exponential variate with parameter 0.5

Parameter $= \beta = \frac{1}{2}$ $\qquad f(x) = \frac{1}{\beta} e^{-x/\beta}$

$\Rightarrow f(x) = 2e^{-2x}$

$F(x) = 1 - e^{-2x}$ (Integrate)

So, $1 - e^{-2x} = y$

$\Rightarrow e^{-2x} = 1 - y$

$\Rightarrow x = -\frac{1}{2} \ln(1-y)$

↳ Given values of y, find x ✓

Chapter → 7
ESTIMATION

- The process of finding the value of a popul^n parameter is called estim^n. The paramter is denoted by θ & the statistics is called an ~~est~~ ESTIMATOR of θ & is denoted by $\hat{\theta}$.

If θ is estimated by a single value, then, its called POINT ESTIMATION.

Let $x_1, x_2, \ldots, x_n$ be a random sample from the distrib^n of X. The statistic $\sum$.

$$\sum_{i=1}^{n} \frac{x_i}{n} \text{ is called the SAMPLE MEAN.}$$

& is denoted by $\bar{X}$.

If μ is the POPULATION MEAN, then, the sample mean $\bar{X}$ is called a POINT ESTIMATOR OF μ.

The 2 imp. properties of a Point Estimator are: P1) $\hat{\theta}$ is unbiased for θ. → have value close to mean

P2) $\hat{\theta}$ is to have small variance for large sample sizes.

An estimator, $\hat{\theta}$ of θ is said to be an unbiased estimator of θ, if $E(\hat{\theta}) = \theta$.

☆ **Results** 1) Let $\bar{X}$ be the sample mean for a random sample of size 'n', from a distrib^n with mean μ & variance σ^2. Then,

☆☆ $Var(\bar{X}) = \frac{\sigma^2}{n}$; $Mean(\bar{X}) = \mu$, $SD = \frac{\sigma}{\sqrt{n}}$

Puffin
Date
Page

* Notation : μ : popul^n mean (p.m)
σ^2 : popul^n variance (p.v)
$\bar{X}$: sample mean (s.m)
S^2 : sample variance (s.v)

$$\bar{X} = \sum_{i=1}^{n} \frac{x_i}{n} \; ; \; S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Pg-247) Q.13)(c) Estimator for mean $= \bar{X} = \sum \frac{x_i}{n}$

$= \frac{\text{Sum of values}}{25}$
$= 169.04$

$\hat{\mu} = 169.04$

$$S^2 = \frac{1}{24} \sum_{i=1}^{25} (X_i - 169.04)^2$$

Put all 25 values given

$= 81.54$

$\therefore$ An estimate for σ^2 is $\hat{\sigma}^2 = 81.54$

Pg-244) Q.1 $\mu = 8$, $\sigma^2 = 5$, $n = 20$
$\cancel{\bar{X}}$ Mean of $\bar{X} = \mu = 8$
$Var(\bar{X}) = \sigma^2/n = 5/\cancel{8}\,20 = 0.25$

Pg-245) Q.5) $n = 10$, $p = ?$ X: binomial distrib^n
(a) $\mu = E(X) = \bar{x} = np = 10p$
$\hat{p} = \bar{x}/10$.
(b) $\bar{x} = \sum \frac{x_i}{n} = \frac{3+4+4+5+6}{5} = \frac{22}{5} = 4.4$

$$\hat{p} = \frac{\bar{x}}{n} = \frac{4.4}{10} = 0.44$$

* Method of Moments :

This is one of the methods to derive the estimator for a popul^n parameter. We have.

$$M_K = \sum_{i=1}^{n} \frac{(x_i)^k}{n};$$

where, M_K is the k^{th} moment of x.

Put $k=1$, we have, $M_1 = \bar{X}$ $\left(\frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1+x_2+\ldots x_n}{n} = \bar{X}\right)$

$\Rightarrow \hat{\mu} = M_1$

$k=2$, we have $M_2 = \sum_{i=1}^{n} \frac{x_i^2}{n}$

$Var\ \sigma^2 = E(X^2) - [E(X)]^2$

$\Rightarrow \hat{\sigma}^2 = M_2 - M_1^2$

Q.15) Pg 248 $X \to B(20, p)$

$M_1 = E(X) = np \Rightarrow \bar{X} = 20p$

$p = \frac{\bar{X}}{20}$

So, $\hat{p} = \frac{\bar{X}}{20}$

$\bar{X} = \sum \frac{x_i}{n} = \frac{13+15+17+12+10}{5}$

$\Rightarrow \bar{X} = 13.4$

So, $\hat{p} = \frac{13.4}{20} = 0.67$.

Pg-249) Q.21) $f(x) = \frac{1}{\beta} e^{-x/\beta}$ (exponential), Mean $= \mu = \beta$

$\Rightarrow \hat{\mu} = \hat{\beta}$ (estim^n for mean = estim^n of β) → (1)

Moment estimator for mean is $\hat{\mu} = M_1$ → (2)

So, $M_1 = \hat{\beta} = 169.04$.

Moment estimator for σ^2, $\hat{\sigma}^2 = M_2 - M_1^2$

$$= \sum_{i=1}^{n} \frac{x_i^2}{n} - (169.04)^2$$

$$= \frac{(160)^2 + (176)^2 + \dots (161)^2}{25} - (169.04)^2$$

$$= 78.28$$

Note: Note value of $\hat{\sigma}^2$ in both parts are diff^t
$\because$ Moment estimator for $\hat{\sigma}^2$ uses simple arithmetic mean

Q.18) Pg. 248) λ, $s = 2$

$$\bar{X} = \sum \frac{x_i}{n} = \frac{2+5+1+\dots+5}{10} = 3.1$$

$$\bar{X} = \lambda s = 3.1 \Rightarrow (\hat{\lambda s}) = 3.1$$

$$\hat{\lambda} = \frac{3.1}{s} = \frac{3.1}{2} = 1.55$$

Q.19 Find the moment estimates for α & β, which are the parameters of Γ distrib^n.

for Γ distribution, mean $= \mu = \alpha\beta$, variance $= \alpha\beta^2$

$$\hat{\mu} = (\hat{\alpha}\hat{\beta}) \Rightarrow M_1 = \hat{\alpha}\hat{\beta} \rightarrow ①$$

also, $\hat{\sigma}^2 = (\hat{\alpha}\hat{\beta}^2) \Rightarrow M_2 - M_1^2 = \hat{\alpha}(\hat{\beta}^2) \rightarrow ②$

$\Rightarrow$ ② $\div$ ①

$$\Rightarrow \frac{\cancel{\hat{\alpha}}(\hat{\beta}^{\cancel{2}})}{\cancel{\hat{\alpha}}\cancel{\hat{\beta}}} = \frac{M_2 - M_1^2}{M_1} \Rightarrow \boxed{\frac{M_2 - M_1^2}{M_1} = \hat{\beta}} \rightarrow ③$$

using ③ in ①

$$\Rightarrow M_1 = \hat{\alpha}\left(\frac{M_2 - M_1^2}{M_1}\right) \Rightarrow \boxed{\hat{\alpha} = \frac{M_1^2}{M_2 - M_1^2}}$$

Puffin
Date
Page

* Notation:- $\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdot a_3 \cdots\cdots a_n$

* Maximum likelihood estimator.

Method of maximum likelihood

S1. Get a sample $x_1, x_2, \ldots, x_n$ of size n from a popul^n of a random variable X, with density f and ~~+~~ associated parameter, θ.

S2. Define a f^n

$$L(\theta) = \prod_{i=1}^{n} f(x_i).$$

This f^n is called likelihood function.

S3. Find the expression for θ, that maximises $L(\theta)$. Replace θ by $\hat{\theta}$. $\hat{\theta}$ is the maximum likelihood estimator for θ.

S4. To find the value of θ, find $\hat{\theta}$ for a given sample.

page 250 Q.31) (exponential distribution); parameter: β.

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \; ; x > 0.$$

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{1}{\beta} e^{-x_i/\beta}\right) = \frac{1}{\beta} e^{-x_1/\beta} \cdot \frac{1}{\beta} e^{-x_2/\beta} \cdots \frac{1}{\beta} e^{-x_n/\beta}$$

$$= \left(\frac{1}{\beta}\right)^n \cdot e^{-\frac{(x_1 + x_2 + \cdots + x_n)}{\beta}}$$

Now, $\bar{X} = \frac{x_1 + x_2 + \cdots + x_n}{n} \Rightarrow x_1 + x_2 + \cdots + x_n = n\bar{X}$

$$\Rightarrow L(\beta) = \left(\frac{1}{\beta}\right)^n e^{-n\bar{X}/\beta}$$

$$\Rightarrow \ln L(\beta) = n \ln\left(\frac{1}{\beta}\right) - \frac{n\bar{X}}{\beta}$$

* Maximum likelihood estimate for a parameter of a distribution is the mean of that distribution ($\bar{x}$)

$$\Rightarrow \frac{1}{L(\beta)} \cdot L'(\beta) = \frac{n}{1/\beta}\left(\frac{-1}{\beta^2}\right) + \frac{n\bar{x}}{\beta^2}$$

$$\Rightarrow \frac{L'(\beta)}{L(\beta)} = \frac{-n}{\beta} + \frac{n\bar{x}}{\beta^2} = \frac{-n\beta + n\bar{x}}{\beta^2} = 0$$

$$\Rightarrow \boxed{\bar{x} = \beta}$$

Q. Estimate β for the sample 1, 2, 1, 7, 5

$$\bar{x} = \frac{\sum x_i}{n} = \frac{16}{5} = 3.2 = \hat{\beta}$$

$$\Rightarrow \hat{\beta} = 3.2.$$

h.w page 250, Q. 33 & page 251 Q. 35

page 234,

* Theorem 7.3.1, 7.3.2, 7.3.3, 7.3.4 → read them.

page 251, Q. 37

(a) $m_x(t) = e^{2t + 9t^2/2} \equiv m_x(t) = e^{\mu t + \sigma^2 t^2/2}$

$\Rightarrow \mu = 2, \sigma^2 = 9.$

So, its normal variate with mean, $\mu = 2$, variance, $\sigma^2 = 9$.

(c) $m_x(t) = .25e^t/(1 - .75e^t) \equiv pe^t/(1 - qe^t)$

$\Rightarrow p = 0.25$ & $q = 0.75$

So, its a geometric distribution, $p = 0.25$.

Puffin

★ Interval Estimation

• **Confidence interval:**
A confidence interval is a particular kind of interval estimate of a population parameter.

• A $100(1-\alpha)\%$ cofidence interval for a parameter θ is a random interval $[L_1, L_2]$, s.t.,

$$P[L_1 \leq \theta \leq L_2] = 1-\alpha$$

• **Central limit theorem:**
Let $x_1, x_2, \ldots, x_n$ be a sample of size, n, taken from a popul^n, having distrib^n with mean, μ & variance, σ^2. Then, for large n, $\bar{X}$ is approx. normal with mean μ & variance σ^2/n. Also,

$$\boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}} \text{ is approx. Standard normal.}$$

This approxim^n is good for $\boxed{n \geq 25}$

For normal population with mean, μ & variance, σ^2,
A $100(1-\alpha)\%$ confidence interval on μ is given by

$$\left[\bar{X} - Z_{\alpha/2} \cdot \sigma/\sqrt{n},\ \bar{X} + Z_{\alpha/2} \cdot \sigma/\sqrt{n}\right]$$

—x—

page 254, Q.49.

(a) X cannot be normally distributed ∵ its discrete random variable.

(b) $\hat{\mu} = \bar{X} = \sum \frac{x_i}{n} = \frac{112}{40} = 2.8$

Puffin

Date

Page

(c) $\sigma = 0.5$, $n = 40$ (sample size). [>25)
↳ central limit thm. applicable.

So, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \simeq$ standard normal.

Now, $100(1-\alpha)\% = 99\%$

$\Rightarrow 1 - \frac{99}{100} = \frac{1}{100} = 0.01 = \alpha$

$\Rightarrow \alpha/2 = 0.005$

$\therefore$ 99% C.I (confidence interval) is

$[\bar{X} - Z_{\alpha/2} \cdot \sigma/\sqrt{n}, \bar{X} + Z_{\alpha/2} \cdot \sigma/\sqrt{n}]$

$= [2.8 - Z_{0.005} \cdot 0.5/\sqrt{40}, 2.8 + Z_{0.005} \cdot 0.5/\sqrt{40}]$
↳ table V (page - 697) = 2.575

$= [2.5964, 3.0036]$

(d) $\mu = 3.0$.
($\because 3 \in$ confidence interval obtained above $\Rightarrow$ it does not refute).

Page 256, Q.56; $\sigma = 5$, $n = 36$

(a) $\hat{\mu} = \bar{X} = \frac{\sum x_i}{n} = \frac{257}{36} = 7.139$

(b) $\bar{X}$ is normal, with mean $= 7.139$ & variance $\sigma^2/n = 25/36$

$\because n > 25$, so, approxim^n can be applied

Hence, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim$ standard normal

$100 \cdot (1-\alpha)\% = 98\%$

$\Rightarrow 1 - \frac{98}{100} = 0.02 = \alpha$

$\Rightarrow \alpha/2 = 0.01$.

C.I $= [\bar{X} - Z_{\alpha/2} \cdot \sigma/\sqrt{n}, \bar{X} + Z_{\alpha/2} \cdot \sigma/\sqrt{n}]$ ✓ $= [5.2015, 9.0765]$

Puffin
Date 2.1.2012
Page

(d) as it (10) goes outside interval, the statement is surprising.

Chapter · 8

Inference on mean & variance of a distrib^n:

* Interval estim^n of variance σ^2.

Theorem : 8.1.1 : page 260 from book.

* CI for σ^2 : $\left[\dfrac{(n-1)s^2}{\chi^2_{\alpha/2}}, \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]$

; χ^2 has $(n-1)$ degrees of freedom.

page 288 Q.1 (b) $\hat{\sigma}^2 = s^2 = \frac{1}{29}\sum(X_i - \bar{X}) = 0.0129$.

(c) $100(1-\alpha)\% = 95$

$\Rightarrow \alpha = 1 - 0.95 = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$.

$\therefore CI = \left[\dfrac{(30-1)(0.0129)}{\chi^2_{(0.025)}}, \dfrac{(30-1)(0.0129)}{\chi^2_{(1-0.025)}}\right]$; value for 0.025 = 16

; table IV.
(see value of 1 − 0.025 on top row & value of $\gamma = n-1$) = 45.7

$\Rightarrow CI = [0.0082, 0.0234]$

(d) CI on $\sigma = \sqrt{[0.0082, 0.0234]} = [0.0906, 0.153]$

(e) 0.2 goes outside the interval, so, its surprising

. [0.0906, 0.153].

Puffin
Date
Day

* T-DISTRIBUTION (or Student-T distribⁿ)

Defnⁿ: Z : std. normal distribⁿ, say.
χ^2_ν : for ν degrees of freedom, then,

$$\boxed{T = \frac{Z}{\sqrt{\frac{\chi^2_\nu}{\nu}}}}$$

→ ν : degree of freedom

Theorem :- For random variables $X_1, X_2, \ldots, X_n$ with normal distribution of mean μ & variance σ^2, then, the random variable

$$\boxed{\frac{\bar{X} - \mu}{S/\sqrt{n}}}$$ follows T distribⁿ.

; $\nu = n-1$

* C.I. on μ (for normal distribⁿ: μ, σ^2) is given by

$$\left[\bar{X} - t_{\alpha/2} \cdot S/\sqrt{n}, \bar{X} + t_{\alpha/2} \cdot S/\sqrt{n}\right]$$

Page 291. Q.11.

(a) $\bar{x} = \frac{1}{n}\sum x_i = 1.2896$; $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$

$s = \sqrt{s^2} = 0.0035$

$= \frac{1}{20-1}\sum(x_i - 1.2896)^2$

$= 0.00001225$

(b) $100(\alpha - 1)\% = 95\% \Rightarrow \frac{\alpha}{2} = 0.025$

So, $CI\Big|_\mu = \left[1.2896 - t_{0.025} \cdot 0.0035/\sqrt{20}, 1.2896 + t_{0.025} \cdot \frac{0.0035}{\sqrt{20}}\right]$

Table VI Table VI

Take value of $1 - 0.025$
& $\nu = 19$. $= 2.093$

$\mu = 1.29$. It belongs to our interval (CI). So, it doesn't suspect us.

Sample size req^d to estimate μ:

The min. sample size req^d to estimate μ to within 'd' units with $(1-\alpha)100\%$ CI is given by

(i) $$n = \frac{Z^2_{\alpha/2}\,\sigma^2}{d^2} \quad (\sigma : \text{known})$$

(ii) $$n = \frac{Z^2_{\alpha/2}\,\hat{\sigma}^2}{d^2} \quad (\sigma : \text{unknown})$$

$\hat{\sigma}$ → only an estimation

94. Q.19.(b) $\sigma = 500 \Rightarrow \hat{\sigma}^2 = 500^2$; $d = 50$, confidence: 95%

$$n = Z^2_{\alpha/2} \frac{\hat{\sigma}^2}{d^2}$$

$100(1-\alpha)\% = 95\% \Rightarrow \alpha = 0.025$.

$$\therefore n = Z^2_{0.025} \frac{500^2}{50^2}$$

→ search body of table I for 0.975 to get 1.96

Section - 8.3

HYPOTHESIS TESTING.

Null hypothesis (H_0) Alternative hypothesis (H_1)

page 268 : guidelines for hypothesis testing.

figure 8.7.

Decision	H_0-true	H_0-false
Accepted H_0	Correct decision	Type II error (β)
Rejected H_0	Type-I error (α)	Correct decision

Date
Page

* Critical region (C)
Set of all values which leads to ~~see~~ reject H_0.

* Region of ~~r~~ acceptance:
The region/values outside critical region.

* ~~Er~~ Errors & its probabilities:
α: probability of making type-I error (also called level of significance).
β: probability of making type-II error.

Result: $\beta + \text{power} = 1$ (power → numerical value)
$\Rightarrow \text{power} = 1 - \beta$

page 295 (Q.21) (a) $H_0: \mu_0 = 0.08$
$H_1: \mu_1 < 0.08$ → reduced from 8%
(b) If type-I error is committed ⇒ avg. metal waste is reduced from 8%, when infact, its not reduced
(c) If type-II error is committed ⇒ avg metal waste is not reduced, but, infact, it is reduced
(d) P(H_0 is rejected given H_0 is true) = 0.05.

—x—

* ONE-TAILED & TWO-TAILED TESTS.

$\theta \rightarrow \theta_0$ → numerical value

$H_0: \theta = \theta_0$: null hypothesis.

Any one of the following:
$H_1: \theta \neq \theta_0$ $(\theta > \theta_0 \text{ or } \theta < \theta_0)$: Two tailed tests.
$H_1: \theta > \theta_0$: One tailed tests (or ~~Left~~ right tailed tests).
$H_1: \theta < \theta_0$: One tailed tests (or ~~Right~~ left tailed tests).

$P[X \geq x_1 / p = p_0] \geq (1-\alpha)$ RT

page 296 Q.26 $H_0 : p \geq 0.7$

$H_1 : p < 0.7$

n: sample size : 10 (n, p given ⇒ binomial)

$\alpha = 0.05$

for left tailed test

$c = \{0, 1, 2, \ldots, x_1\}$; x_1 : max value of x, s.t. $P[X \leq x_1 / p = p_0] \leq \alpha$,

Now,

$P[X \leq 4 / p = 0.7] = 0.047$

$\Rightarrow P[X \leq x_1 / p = 0.7] \leq 0.05$

∴ critical region (c) = $\{0, 1, 2, 3, 4\}$

with $\alpha = 0.047$

(b) if $x = 5$, H_0 will be accepted ∵ $5 \notin c$

Type II error is committed.

Q.27 (a) $H_0 : p_0 \leq 0.5$

$H_1 : p_1 > 0.5$

(b) $n = 15$, $p = 0.5$

$E(X) = np = 7.5$

(c) $c = \{11, 12, 13, 14, 15\}$

If α : level of significance,

$P[X \geq 11 \mid p = 0.5] = \alpha$

But, $P[X \geq 11 \mid p = 0.5] = 1 - P[X \leq 10 \mid p = 0.5]$

$= 1 - 0.9408$ (table) $= 0.0592$

(d) $\beta = P[X \leq 10 \mid p = 0.6] = 0.7827$ (table: $n = 15$, $p = 0.6$)

(e) power $= 1 - \beta = 0.2173$

(f) $X = 12 \in C$. ∴ H_0 will be rejected at 5.92% level of significance. Type-I error is committed

(g) If $X = 10$, H_0 is accepted. So, type II error is committed

* SIGNIFICANCE TESTING

Another method in which we accept or reject H_0.

○ Test statistics for hypothesis testing concerning mean:

$$H_0 : \mu = \mu_0$$

→ Normal distribution

1. $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, σ : known

2. $T = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$, σ : unknown $(n-1 \text{ d.f})$

→ T distribution degrees of freedom

○ Critical regions concerning mean:

• σ is known : normal distribution

$C = \{Z > Z_\alpha\}$ if $H_1 : \mu > \mu_0$

$C = \{Z < -Z_\alpha\}$ if $H_1 : \mu < \mu_0$

$C = \{|Z| > Z_\alpha\}$ if $H_1 : \mu \neq \mu_0$.

• σ is unknown : T distribution

$C = \{t > t_\alpha\}$ if $H_1 : \mu > \mu_0$

$C = \{t < -t_\alpha\}$ if $H_1 : \mu < \mu_0$

$C = \{|t| > t_{\alpha/2}\}$ if $H_1 : \mu \neq \mu_0$

Page 298. (Q.31) σ known, given in part (c).

(a) $H_0 : \mu < 0.05$

$H_1 : \mu > 0.05$ (right tail test)

(b) type-I error means titanium %age is assumed to increase above 5%, while it is not actually ↑. Type II error means, it is not ↑ when, in fact, the %age ↑

(c) $n=100$, $\bar{X}=0.051$; $\sigma=0.008$

Test statistic is : $z = \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \dfrac{0.051-0.05}{0.008/\sqrt{100}} = 1.25$

$P[Z>1.25] = 1-P[Z\leq 1.25] = 1-F(1.25)$

$= 1-0.8944 = 0.1056$

$\therefore$ p value equals 0.1056. Its not too small.

$\therefore$ We can accept H_0

page 299 Q.35). (a) $n=25$, $\alpha=0.05$.

Critical pt. for left-tailed test is $-t_{n-1,\alpha}$

i.e. $-t_{24,\,0.05} = -1.711$

Table VI

$\therefore$ critical region is $t<-1.711$

(c) $n=20$, $\alpha=0.025$

for right-tailed test is, $+t_{n-1,\alpha}$

C.P. $= +t_{19,\,0.025} = 2.093$

$\therefore$ C.R. $= t > 2.093$

(e) $n=20$, $\alpha=0.10$

C.P. $= \pm t_{n-1,\alpha/2} = \pm t_{19,0.025} = \pm 1.729$

C.R. $= |t| > 1.729$

* Hypothesis test on variance.

Three forms :

1. $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$ (RT)

C.R. $= \chi^2_{n-1} > \chi^2_{n-1,\alpha}$ right tail

Puffin
Date
Page

2. $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$ (LT)

C.R. : $\chi^2_{n-1} < \chi^2_{n-1, 1-\alpha}$ left tail

3. $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ (TT)

C.R. $= \{ \chi^2_{n-1} < \chi^2_{n-1, 1-\alpha/2}$ two tail

or $\chi^2_{n-1} > \chi^2_{n-1, \alpha/2} \}$

Test statistic : $\chi^2_{n-1} = \dfrac{(n-1)s^2}{\sigma_0^2}$; $(n-1)$: degrees of freedom.

page 303. Q.47.(a) $H_0 : \mu = 1$ & $H_1 : \mu \neq 1$; $\alpha = 0.05$ level

$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{1.0084 - 1}{0.00282/\sqrt{15}} = 1.154.$

total no. of readings

C.R. $= \{ t < -t_{\alpha/2}$ or $t > t_{\alpha/2} \}$

$= \{ t < -t_{0.025}$ or $t > t_{0.025} \}$

$= \{ t < -2.145$ or $t > 2.145 \}$

rejected accepted rejected

-2.145 2.145

$t = 1.154$ comes here. (at region of acceptance)

H_0 is accepted $\Leftarrow$

$\because t = 1.154 \notin$ C.R.

$\therefore H_0$ is accepted.

(b) $H_0 : \sigma = 0.0025$ & $H_1 : \sigma > 0.0025$; $\alpha = 0.05$ level

$\bar{x} = 1.00084$, $s = .00282$

$\chi^2_{14} = \dfrac{(n-1)s^2}{\sigma_0^2} = 17.813376$

accepted region. $\chi^2 = 17.81$ lies here

C.R : $\chi^2_{14} > \chi^2_{14, 0.05} \Rightarrow \chi^2_{14} > 23.7$

23.7

$\chi^2 = 17.81 < 23.7$. So, H_0 is accepted.

Chapter - 9

INFERENCE ON PROPORTIONS

- If, for 'n' trials, an event occurs x times, then, the proportion of occurence of this event is

$$p = x/n$$

ex: Let p denotes the proportion of defectives in a population. Consider a sample of size 150, in which the no. of defectives = 20. Then,

estimation of proportion $\hat{p} = \frac{20}{150} = \frac{2}{15}$.

* **Point Estimator of p:**

$$\hat{p} = \frac{x}{n} = \frac{\text{no. of items with a property}}{\text{sample size}}$$

* $(1-\alpha)100\%$ confidence interval of p:

$$\left[\hat{p} - Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n},\ \hat{p} + Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\right]$$

* **Minimum sample size required to get 'p' within d units:**

$$n = \frac{Z^2_{\alpha/2}\,\hat{p}(1-\hat{p})}{d^2}$$ (when prior estimates of p is available.)

$$n = \frac{Z^2_{\alpha/2}}{4d^2}$$ (no prior estimates of p is available.)

page 326 Q.1. $p = \frac{x}{n} = \frac{45}{50} = \frac{9}{10} = 0.9$

(a) $\hat{p} = 0.9$

(b) $(1-\alpha)100\% = 90\%$

$\Rightarrow \alpha = 0.1 \Rightarrow \alpha/2 = 0.05$

C.I : $\left[0.9 - Z_{0.05}\sqrt{\frac{0.9(0.1)}{50}}, 0.9 + Z_{0.05}\sqrt{\frac{0.9(0.1)}{50}}\right]$

Table V (= 1.65)

$= [0.8302, 0.9698]$.

(c) $d = 0.02$ (given),

$\alpha/2 = 0.05$

$\Rightarrow n = \frac{Z^2_{\alpha/2}\,\hat{p}(1-\hat{p})}{d^2}$

$\Rightarrow n = \frac{(1.65)^2 \, 0.9(0.1)}{(0.02)^2} = 608.85 \simeq 609.$

* Testing hypothesis for proportions

- 1. $H_0 : p = p_0$ vs $H_1 : p < p_0$ (LT)
  2. $H_0 : p = p_0$ vs $H_1 : p > p_0$ (RT)
  3. $H_0 : p = p_0$ vs $H_1 : p \neq p_0$ (TT)

- C.R. : For LT : C.R. : $Z < -Z_\alpha$
  RT : C.R. : $Z > Z_\alpha$
  TT : C.R. : $Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$

- Test statistic :

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \quad \text{or} \quad \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

statement to be checked

page 328 Q.11: (a) $H_0 : p \le 0.99$, $H_1 : p > 0.99$

(b) $n = 300$, $X = 298$

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{298 - 300(0.99)}{\sqrt{300(0.99)(0.01)}} = 0.58025$$

observed value $\Rightarrow Z_{\text{observed value}} = 0.58$

p_{value} : $P[Z > 0.58]$ → RT test ($\because H_1 : p > 0.99$)

$= 1 - F(0.58)$ → table

$= 1 - 0.7190$

$= 0.281$ (its large $\Rightarrow H_0$ can't be rejected)

* **Comparison of two proportions:**

$\exists$ 2 populations and the same trait is studied on these 2 populations. Let p_1 & p_2 be the proportions of trait happening in 2 popul^ns. Sometimes, we may compare p_1 & p_2. So, the inference may be on $p_1 - p_2$

* **Point estimate for $p_1 - p_2$**

If the event occurs x_1 times in a sample of size n_1, taken from the first popul^n & x_2 times, in a sample of size n_2, taken from the second popul^n, then,

$$\star \quad \boxed{\widehat{p_1 - p_2} = \hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}}$$

Note: For large sample, $\hat{p}_1 - \hat{p}_2$ is approx. normal with $\mu = p_1 - p_2$ & $\sigma^2 = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$.

Puffin
Date ______
Page ______

* C.I. on $p_1 - p_2$:

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

page 329 Q.17. $n_1 : 500$, $x_1 = 178$; $> 72{,}000$ per yr.
$n_2 : 450$, $x_2 = 220$

$p_1 = \frac{x_1}{n_1}$, $p_2 = \frac{x_2}{n_2}$
$\Rightarrow 0.356$ $\Rightarrow 0.489$

(a) $\hat{p}_1$ ✓, $\hat{p}_2$, $\widehat{p_1 - p_2} = \hat{p}_1 - \hat{p}_2 = -0.133$

(b) $(1-\alpha)100\% = 95\% \Rightarrow \alpha = 0.05$
$\Rightarrow$ C.I. $= (-0.133) \pm Z_{0.025}\sqrt{\frac{0.356(1-0.356)}{500} + \frac{0.489(1-0.489)}{450}}$ ✓
$Z_{0.025} \to 1.96$ (Table V)

i.e. c.i $= [-0.195, -0.071]$
$\Rightarrow$ the probability for $p_1 - p_2$ to be in this interval is 95%.

(c) for them to have same, $\hat{p}_1 - \hat{p}_2 = 0$.
the value of $\hat{p}_1 - \hat{p}_2 = 0 \notin [-0.195, -0.071]$ i.e. it has the probability of only 5% which is very less.

—x—

∴ Hypothesis testing on $p1 - p2$:

case 1: $H_0 : p1 - p2 = \delta$ vs
$H_1 : p_1 - p_2 < \delta$ LT
$H_1 : p_1 - p_2 > \delta$ RT
$H_1 : p_1 - p_2 \neq \delta$ TT

Test statistic is approximately std. normal &

$$Z = (\hat{p}_1 - \hat{p}_2) - \delta \Big/ \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Puffin
Date
Page

case 2: $H_0 : p_1 - p_2 = 0$ vs $H_1 : p_1 - p_2 < 0 \ (p_1 < p_2)$ LT
$(\text{or } p_1 = p_2)$ $\qquad p_1 - p_2 > 0 \ (p_1 > p_2)$ RT
$p_1 - p_2 \neq 0 \ (p_1 \neq p_2)$ TT

Test statistic is

$$z = \hat{p}_1 - \hat{p}_2 \Big/ \sqrt{\hat{p}_1(1-\hat{p}_1)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$; \text{where, } \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2} = \frac{X_1 + X_2}{n_1+n_2}$$

Critical region:

RT : $\{Z > Z_\alpha\}$
LT : $\{Z < -Z_\alpha\}$
TT : $\{Z < -Z_\alpha \text{ or } Z > Z_\alpha\}$

page 331 (Q.26)(2) $H_0 : p_1 - p_2 = 0.02$
$H_1 : p_1 - p_2 > 0.02$

(i) Critical pt :
$\alpha = 0.05$

Critical pt., $Z_\alpha = Z_{0.05} = 1.645$

(ii) $p_1 = \frac{x_1}{n_1} = \frac{5}{100}$, $p_2 = \frac{x_2}{n_2} = \frac{1}{100}$, $\widehat{p_1 - p_2} = 0.04$.

Test statistic $\quad 0.05 \quad .0.01 \quad , 0.02$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - \delta}{\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}} = 0.835$$

$\quad , 100 \quad , 100$

C.R. : $\{Z > 1.645\}$ $\qquad H_1$

$0.835 \notin$ C.R. So, H_0 is accepted. Hence, claim is rejected. (no switch over to laser heating).

page 326 Q.2: (a) $\hat{p} = \frac{x}{n} = \frac{75}{193} = 0.3886$.

(b) $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$$c.i = \left[\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$\hat{p} \to 0.3886$, $Z_{\alpha/2} \to 1.96$

$c.i = [0.319831, 0.45737]$

(c) $n = ?$, $d = 0.03$

$$n = \frac{Z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{d^2} = 1014.14 \simeq 1015$$

*always take larger limit

Q.2 $x_1 = 600$, $x_2 = 900$, $n = 1000$.

$\hat{p}_1 = 0.6$, $\hat{p}_2 = 0.9$

(a) $\alpha = 0.1$

$$c.i = \left[\hat{p}_1 \mp Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}}\right]$$

$\hat{p}_1 \to 0.6$, $Z_{\alpha/2} \to 1.65$

$c.i = [0.5744, 0.62556]$

(b) $c.i = [0.8814, 0.91859]$; $\hat{p} = \hat{p}_2$, $Z_{\alpha/2} = 1.96$.

page 328 Q.12) (a) $H_0 : p \geq 0.08$

$H_1 : p < 0.08$ (LT)

(b) $n = 64$, $x_1 = 4$, $p_0 = 0.08$.

Test statistic: $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \simeq -0.516$

$\Rightarrow Z_{observed} = -0.516$

$\therefore$ p-value $= P[Z \leq -0.516] = F[-0.516] = 0.3029$

$\because$ p value is not small, H_0 cannot be rejected.

(c) H_0 is accepted

page 329 Q.19. (a) $\hat{p}_1 = \frac{x_1}{n_1} = 0.31$ $\hat{p}_2 = \frac{x_2}{n_2} = 0.4$

$\widehat{p_1 - p_2} = -0.09$

(b) $\alpha = 0.1$

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$= (-0.09) \pm \underbrace{(1.65)\sqrt{\frac{0.31(0.69)}{200} + \frac{0.4(0.6)}{190}}}_{0.07969}$$

$\Rightarrow c.i = [-0.16969, -0.0103]$

(c) $p_1 - p_2 \Rightarrow p_1 - p_2 = 0$. Only 10% of the value of $p_1 - p_2$ don't lie in c.i. $0 \notin$ c.i So, its surprising.

$$\rho = \frac{cov(X,Y)}{\sqrt{var(X)\, var(Y)}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})/n}{\sqrt{\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n}\sum_{i=1}^{n}\frac{(y_i-\bar{y})^2}{n}}}$$

Puffin Date 22/4/12 Page

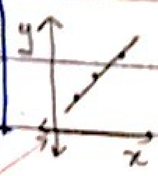
Chapter 11

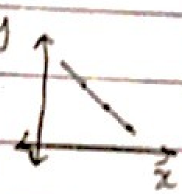
CORRELATION

* Let ~~x~~ X & ~~y~~ Y be 2 random variables. Then, correlation is the study of the strength of the linear relationship that exists between ~~x~~ X & Y.

It's a measure, denoted by 'ρ', known as PEARSON COEFFICIENT OF CORRELATION, given by

$$\boxed{\rho = \frac{cov(X,Y)}{\sqrt{var(X).var(Y)}}}$$





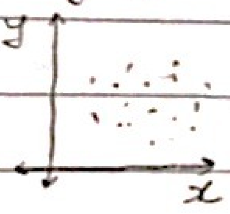
→ $-1 \leq \rho \leq 1$

case 1 → $\rho = 1$: Perfect Positive Relationship

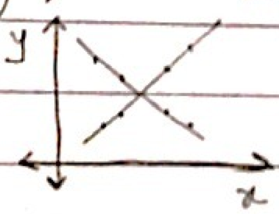
case 2 → $\rho = -1$: Perfect Negative Relationship

case 3 → $\rho = 0$: No Relationship, or, Relationship exists, but not linear.

Values scattered around an area on graph



graph comes out as



* COMPUTATIONAL FORMULA

$$\hat{\rho} = r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

eg: compute corrⁿ coeff. b/w X & Y using following data:

X	1	3	5	7	8	10
Y	8	12	15	17	18	20

	X_i	Y_i	X_i^2	Y_i^2	X_iY_i
	1	8	1	64	8
	3	12	9	144	36
	5	15	25	225	75
	7	17	49	289	119
	8	18	64	324	144
	10	20	100	400	200
Σ	34	90	248	1446	582

$$\Rightarrow \hat{\rho} = r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$\overset{6}{=} \frac{n(582) - 34\times 90}{\sqrt{[n(248) - 1156][n(1446) - 8100]}}$$

$$\Rightarrow r = 0.9879$$

↳ almost perfect +ve. ($r \simeq 1$).

—— × ——

* $(1-\alpha)100\%$ c.i on R :

Lower bound : $\dfrac{(1+R) - (1-R)\exp\left(2Z_{\alpha/2}/\sqrt{n-3}\right)}{(1+R) + (1-R)\exp\left(2Z_{\alpha/2}/\sqrt{n-3}\right)}$

Upper bound : $\dfrac{(1+R) - (1-R)\exp\left(-2Z_{\alpha/2}/\sqrt{n-3}\right)}{(1+R) + (1-R)\exp\left(-2Z_{\alpha/2}/\sqrt{n-3}\right)}$

* Test statistic for testing :

$H_0 : \rho = 0$

T distribution $T_{n-2} = \dfrac{R\sqrt{n-2}}{\sqrt{1-R^2}}$; C.R. $\begin{cases} t_{n-2} > t_{n-2,\alpha} & \text{(RT)} \\ t_{n-2} < -t_{n-2,\alpha} & \text{(LT)} \\ |t| > t_{n-2,\alpha/2} & \text{(TT)} \end{cases}$

* Test statistic for testing :

$H_0 : \rho = \rho_0$

Std. normal $Z = \dfrac{\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right) - \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{n-3}}}$

; C.R. $\begin{cases} Z > Z_\alpha & \text{(RT)} \\ Z < -Z_\alpha & \text{(LT)} \\ |Z| > Z_{\alpha/2} & \text{(TT)} \end{cases}$

—×—

page 433, (Q.47)

(h) Do just as in ↖ example.

(Q.48) $H_0 : \rho = 0.$

$H_1 : \rho \neq 0$

$\alpha = 0.1$, $\alpha/2 = 0.05$, $n = 10.$

$\because \rho = 0 \Rightarrow$ Test statistic for T distribⁿ.

$T_{n-2} = \dfrac{R\sqrt{n-2}}{\sqrt{1-R^2}}$; (R = 0.887, comes from table)

$\Rightarrow 5.433$.

T_8 (observed value) = 5.433

C.R. : $\{ t_8 < -t_{0.05}$ or $t_8 > t_{0.05} \}$

: $\{ t_8 < -1.86$ or $t_8 > 1.86 \}$

∴ H_0 is rejected.

∴ X & Y are uncorrelated is not true.

page 436 Q.49) $\alpha = 0.1$

$\alpha/2 = 0.05$, R = 0.887

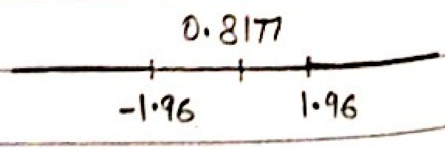
90% c.i = [0.656095, 0.9660488]

Q.50) $H_0 : \rho = 0.8$; $\alpha = 0.05$

$H_1 : \rho \neq 0.8$; $\alpha/2 = 0.025$

(nothing given, like an inequality ⇒ TT)

Z = 0.81771 $\qquad Z_{\alpha/2} = 1.96$



For TT, C.R :- $|Z| > Z_{\alpha/2}$

$Z < -1.96$ or $Z > 1.96$

So, $Z \notin$ C.R.

∴ H_0 is accepted.

24/4/12

Regression:

Suppose, we want to measure the water temp., at diff^t water depth; even though, the depth is same, the temp. at diff^t places will be diff^t due to geographical reason.

So, for a given depth x, we'll get diff^t random temps denoted by Y/X.

The avg. of these temps has some mean, denoted by $\mu_{Y/X}$. A graph of $\mu_{Y/X}$ for given values of x, is called a REGRESSION CURVE of Y on X.

For various values of x, find mean of Y

Illy, we can define a regression curve of X on Y.

A REGRESSION LINE of Y on X is the line,

$$\mu_{Y/X} = \beta_0 + \beta_1 x$$

; here, β_0 & β_1 are real consts. to be estimated.

- Estimators of β_0 & β_1 are denoted by b_0, b_1.
  When β_0 & β_1 are evaluated, we denote them by b_0 & b_1

Least square estimation:

The parameters, β_0 & β_1 can be estimated by the method of least squares.

cov(x, y)

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = b_1$$

var(X)

$$\hat{\beta}_0 = b_0 = \bar{y} - b_1\bar{x}$$

$\bar{x} = \frac{\sum x}{n}$

$\bar{y} = \frac{\sum y}{n}$

page 427 Q.7 (b) $\beta_1 = \dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$

$\Rightarrow \beta_0 = 0.0957334454 = b_0$

$\beta_0 = \bar{y} - 5.557326456$

$\Rightarrow \beta_0 = 0.2176735444 = b_1$

$\Rightarrow \mu_{y/x} = b_0 + b_1 x$

$= 0.21767 + (0.09573)x$

(c) $x = 50$

$\mu_{y/50} = 5.00417$

(d) $2 b_1 = 0.19146 \times 10^8$ (increases)

(e) $2 b_1 = 0.19146 \times 10^8$ (decreases)

$\Big\}$ same value.

CHAPTER - 12.
Multiple Linear Regression Model

In regression, the variable X is a predictor variable & Y is a response variable. The 2 models available are
(i) Polynomial model
(ii) Multiple linear regression model. (m.l.r.m)

(i) • In polynomial model, Y is expressed as a polynomial of X, of degree 'p', i.e.,

only one variable (degree : p)

$$\mu_{Y/X} = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots\ldots + \beta_p x^p$$

(ii) • In m.l.r.m, we express mean of Y, as a linear f^n of more than 1 predicted variable.

i.e., $\mu_{Y/X_1, X_2, X_3, \ldots, X_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k$

k variables (degree : 1)

The estimates of $\beta_0, \beta_1, \beta_2 \ldots\ldots$ can be obtained by solving the normal eq^{ns} by least square method.

* Normal eq^{ns} for Polynomial model:
For 2^{nd} degree polynomial model,

$$\mu_{Y/X} = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots\ldots,$$

normal eq^{ns} are :-
$$n b_0 + b_1 \sum x + b_2 \sum x^2 = \sum y$$
$$b_0 \sum x + b_1 \sum x^2 + b_2 \sum x^3 = \sum xy$$
$$b_0 \sum x^2 + b_1 \sum x^3 + b_2 \sum x^4 = \sum x^2 y$$

example. the following are the data on drying time of a paint & amt. of additive i.e. intended to reduce drying time.

Amt. of additive (gms.) (x)	0	1	2	3	4	5	6	7	8
Drying time (hrs) (y)	12	10.5	10	8	7	8	7.5	8.5	9

i. find a 2nd degree parabola, by least squares.

polynomial model

normal eq^{ns}

x	y	xy	x^2	x^3	x^4	x^2y
0	12	0	0	0	0	0
1	10.5	10.5	1	1	1	10.5
2	10	20	4	8	16	40
3	8	24	9	27	81	72
4	7	28	16	64	256	112
5	8	40	25	125	625	200
6	7.5	45	36	216	1296	270
7	8.5	59.5	49	343	2401	416.5
8	9.0	72	64	512	4096	576
Σ 36	80.5	299	204	1296	8772	1697.0

Substituting,
$$9b_0 + 36b_1 + 204b_2 = 80.5$$
$$36b_0 + 204b_1 + 1296b_2 = 299$$
$$204b_0 + 1296b_1 + 8772b_2 = 1697.0$$

$b_0 = 12.18$

$b_1 = -1.846$

$b_2 = 0.1829$

$\Rightarrow \mu_{Y/x} = 12.18 - 1.846x + 0.1829x^2$

Ans

* **Multiple l.r.m:**

$$\mu_{Y/x_1, x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Normal eq^{ns}: $b_0 n + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} = \sum y_i$

$$b_0 \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} = \sum x_{1i} y_i$$

$$b_0 \sum x_{2i} + b_1 \sum x_{2i} x_{1i} + b_2 \sum x_{2i}^2 = \sum x_{2i} y_i$$

—x—

example: Following are the data on the no. of twists req^d to break a certain kind of alloy bar & %age of 2 alloy elements present in the metal. Estimate β_0, β_1 & β_2 by considering the model

$$\mu_{Y/x_1, x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

No. of twists (Y)	41	49	69	65	40	50	58	57	31	36	44	57	19	31	33	43
% of elemt A	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
% of element B	5	5	5	5	10	10	10	10	15	15	15	15	20	20	20	20

$\sum x_{1i} = 40$ $\quad \sum x_{2i} = 200$ $\quad \sum x_{1i}^2 = 120$ $\quad \sum x_{2i}^2 = 3000$

$\sum x_{1i} x_{2i} = 500$ $\quad \sum y_i = 723$ $\quad \sum x_{1i} y_i = 1963$ $\quad \sum x_{2i} y_i = 8210$

$\Rightarrow 16 b_0 + 40 b_1 + 200 b_2 = 723$

$40 b_0 + 120 b_1 + 500 b_2 = 1963$

$200 b_0 + 500 b_1 + 3000 b_2 = 8210$

$\rightarrow$ solving : $b_0 = 46.4$

$b_1 = 7.78$

$b_2 = -1.65$

$\therefore \mu_{Y/x_1, x_2} = 46.2 + 7.78x - 1.65 x_2$

* Matrix approach to Least Squares

$$\hat{\beta} = b = (X'X)^{-1} (X'Y)$$

example : Use matrix method to fit a st. line

$$\mu_{Y/x} = b_0 + b_1 x$$

for the following data

x	0	1	2	3	4
y	8	9	4	3	1

$\rightarrow$ 1st column is 1 always

$\rightarrow$ 2nd column is the elements

Matrix X : $\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}_{5\times2}$

$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}_{2\times5}$

Y : $\begin{bmatrix} 8 \\ 9 \\ 4 \\ 3 \\ 1 \end{bmatrix}_{5\times1}$

$X'X = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}$

$[X'X]^{-1} = \frac{1}{50} \begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix}$

$$X'Y = \begin{bmatrix} 25 \\ 30 \end{bmatrix}_{2\times 1}$$

$$\Rightarrow \hat{\beta} = b = \frac{1}{50}\begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix}\begin{bmatrix} 25 \\ 30 \end{bmatrix}$$

$$= \frac{1}{50}\begin{bmatrix} 750-300 \\ -250+150 \end{bmatrix}$$

$$= \frac{1}{50}\begin{bmatrix} 450 \\ -100 \end{bmatrix}$$

$$\Rightarrow \hat{\beta} = b = \begin{bmatrix} 9 \\ -2 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

$$\Rightarrow b_0 = 9 \quad , b_1 = -2$$

$$\mu_{Y/X} = b_0 + b_1 x = 9 - 2x.$$

example: Fit a 2nd degree polynomial

$$\mu_{Y/X} = b_0 + b_1 x + b_2 x^2$$

x	5	5	10	10	15	15	20	20	25	25
y	14	12.5	7	5	2.1	1.8	6.2	4.9	13.2	14.6

$$X = \begin{bmatrix} 1 & 5 & 25 \\ 1 & 5 & 25 \\ 1 & 10 & 100 \\ 1 & 10 & 100 \\ 1 & 15 & 225 \\ 1 & 15 & 225 \\ 1 & 20 & 400 \\ 1 & 20 & 400 \\ 1 & 25 & 625 \\ 1 & 25 & 625 \end{bmatrix} \qquad Y = \begin{bmatrix} 14 \\ 12.5 \\ 7 \\ 5 \\ 2.1 \\ 1.8 \\ 6.2 \\ 4.9 \\ 13.2 \\ 14.6 \end{bmatrix}$$

Ans:

$$\mu_{Y/X} = 27.3 - 3.313x + 0.111x^2$$

Two different Regression models:

1 Exponential model: $y = ae^{bx} \rightarrow ①$

taking ln : $\ln y = \ln a + bx \rightarrow ②$

let : $\ln y = Y$; $b_0 = \ln a$; $b_1 = b$

$\Rightarrow$ ② becomes : $Y = b_0 + b_1 x$

- $b_1 = \dfrac{n\sum xY - (\sum x)(\sum Y)}{n\sum x^2 - (\sum x)^2}$
- $b_0 = \bar{Y} - b_1 \bar{X}$

Now, $b_0 = \ln a \Rightarrow a = e^{b_0}$

$b = b_1$

ex. fit an exponential curve in the form $y = ae^{bx}$ to following data :

						Σ
x	2	4	6	8	10	30
y	4.077	11.084	30.128	81.897	222.62	349.806
$X = x^2$	4	16	36	64	100	220
$Y = \ln y$	1.405	2.405	3.405	4.405	5.405	17.027
XY	2.81	9.62	20.43	35.24	54.05	122.152

$b_1 = \dfrac{5(122.152) - 30(17.027)}{5(220) - 900} = 0.5$

$b_0 = 0.4054$

$a = e^{b_0} = 1.499$

$b = b_1 = 0.5$

$\Rightarrow y = ae^{bx} = (1.499)e^{(0.5)x}$

2. Power model :

$$y = ax^b$$

taking ln : $\ln y = \ln a + b \ln x$

$\quad\quad\quad\quad\quad\quad Y \quad\quad b_0 \quad b_1 \quad X$

So, $Y = \ln y$ $\quad \Rightarrow Y = b_0 + b_1 X$

$X = \ln x$ $\quad ; b_0 = \dfrac{n\sum xy - (\sum x)(\sum Y)}{n\sum x^2 - (\sum x)^2}$

$b_0 = \ln a$

$b_1 = b$ $\quad b_1 = \bar{Y} - b_1 \bar{X}$.

$$a = e^{b_0}$$
$$b = b_1.$$

ex. fit a curve of the form $y = ax^b$ to following data :

x	y	$\ln x (=X)$	$\ln y (=Y)$	X^2	XY
2	43	0.693	3.761	0.48	2.606
4	25	1.386	3.218	1.92	4.460
7	18	1.9459	2.89	3.786	5.623
10	13	2.3025	2.5649	5.301	5.905
20	8	2.9957	2.079	8.974	6.228
40	5	3.688	1.609	13.601	5.933
60	3	4.094	1.098	16.76	4.495
80	2	4.382	0.693	19.201	3.036
Σ		21.487	17.9129	70.023	38.286

$\Rightarrow b_1 = -0.798$

$b_0 = 4.3824$

$a = 80.0311$, $b = -0.798$

$Y = (80.0311) e^{-0.798}$ Ans