Module 2.1

1/17/20,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________

(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

Continuous Distributions

1) **Uniform Distribution:-**

→ Gives Probability that Observation will occur within a Particular interval when Probability of occurrence within that interval is directly Proportional to interval length.

→ Easiest distribution to generate. It serves as a basis for other random variables in simulation.

→ E.g. used to generate random values.

→ $U(0,1)$ is a special case of beta distribution.

→ $P(x_1 < X < x_2) = F(x_2) - F(x_1) = \dfrac{x_2 - x_1}{b-a}$

Applications:- It is used as a "first" model for a Process that is randomly varying between "a" and "b".

→ The $U(0,1)$ is used to generate uniformly distributed random numbers between 0 and 1.

PDF :- $f(x) = \dfrac{1}{b-a}$ if $a \le x \le b$

0 , otherwise

CDF

0 if $x < a$

$F(x) = \dfrac{x-a}{b-a}$ if $a \le x \le b$

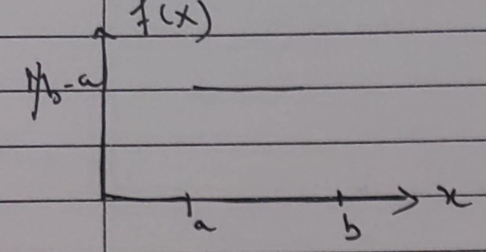
1 if $x > b$

Parameters "a" and "b" are real nos with $a < b$.

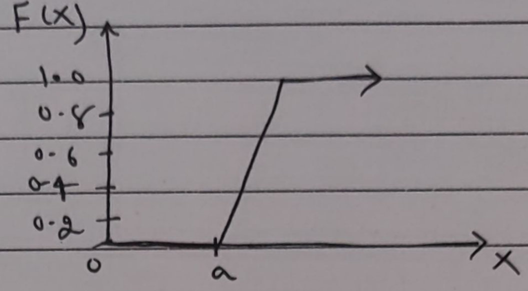
Mean = $E(X) = \frac{a+b}{2}$.

Variance = $V(X) = \frac{(b-a)^2}{12}$.

Mode does not uniquely exist.



PDF



CDF

Example:- Local Trains arrive at the "Ambarnath" railway station at every 15 minutes beginning at 5.00 a.m. A passenger arrives at railway station randomly which is uniformly distributed between 10.00 a.m and 10.30 a.m. Find the probability that the passenger has to wait for the train for

i) less than 6 minutes.

ii) more than 10 minutes.

⟹ Let X be the random variable that denotes the number of minutes pass 10.00 a.m. that the passenger arrives at railway station. Hence X is uniformly distributed over a period (0, 30).

i) The passenger has to wait less than 6 minutes if he arrives between 10:09 a.m and 10.15 a.m. or between 10.24 a.m and 10.30 a.m.

→ So the required probability is.

$P(9 < X < 15) + P(24 < X < 30)$.

$= F(15) - F(9) + F(30) - F(24)$.

$= \frac{15-9}{30} + \frac{30-24}{30} = \frac{6}{30} + \frac{6}{30} = \frac{12}{30} = \frac{2}{5}$.

ii) The passenger has to wait for more than 10 minute if he arrives between 10.00 a.m. to 10.05 a.m. or between 10.15 a.m. to 10.20.

$\therefore P(0<X<5) + P(15<X<20)$

$= F(5) - F(0) + F(20) - F(15)$

$= \frac{5-0}{30} + \frac{20-15}{30} = \frac{5}{30} + \frac{5}{30} = \frac{10}{30} = \underline{1/3}.$

2) Exponential Distribution :- $\exp(\lambda)$.

→ Gives distribution of time between independent events occuring at a constant rate.

→ E.g. used to model the lifetime of a component that fails instantaneously (light bulb).

→ Special case of both gamma and weibull distribution.

→ Only continuous distribution with memoryless property.

i.e for all $s \geq 0$ and $t \geq 0$, $P(X>s+t \mid X>s) = P(X>t)$.

→ it means that component does not "remember" that it has already been use for a time 's'. A used component is as good as new.

→ The above property can be proved by using conditional property as follows.

$P(X>s+t \mid X>s) = \frac{P(X>s+t)}{P(X>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$

$= P(X>t)$

Application → Probability distribution of life, presuming constant conditional failure (or hazard) rate since distribution has long tail.

→ Consequently applicable in many, but not all reliability situations.

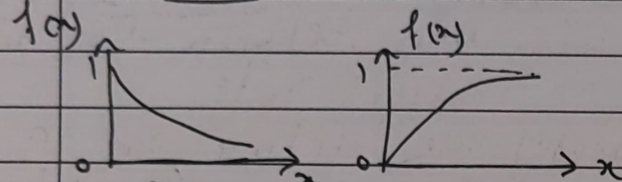
→ Interarrival time of customer to a system when arrivals occur completely random at a constant rate.

→ Service times which are highly variable.

PDF $f(x) = \lambda e^{-\lambda x}$; if $x \geq 0$.
$= 0$; otherwise.

CDF
$$F(x) = \begin{cases} 0 & ; \text{ if } x < 0. \\ \int_0^x \lambda e^{-\lambda t}\,dt = 1 - e^{-\lambda x} & ; \text{ if } x \geq 0. \end{cases}$$

→ **Range** = $[0, \infty]$.
Mean = $E(X) = 1/\lambda$.
Variance = $V(X) = 1/\lambda^2$
Mode = 0



PDF   CDF

Example 1 – A component of a system whose time to failure is exponentially distributed with failure rate $\lambda = 1/6$. If 6 such components are installed in different systems what is the probability that atleast 2 are still working at the end of 9 years?

⇒ Let X be a random variable that denotes the failure time of a component.

→ The probability that a component is working at the end of 9 years is same as component will fail after 9 years and is given by

$$P(X > 9) = 1 - P(X \leq 9)$$
$$= 1 - F(9) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x} = e^{-\frac{1}{6} \cdot 9}$$
$$= e^{-3/2} = \underline{0.223}$$

→ Let Y be a random variable that denotes the number of component in working condition.

∴ Probability that at least two component will work $= P(Y \geq 2) = 1 - P(Y < 2)$

1/17/20,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

$= 1 - [P(0) + P(1)]$

but $P(X=k) = \binom{n}{k} p^k q^{n-k}$

$\therefore P(X \geq 2) = 1 - \left[\binom{6}{0} p^0 q^6 + \binom{6}{1} p^1 q^5\right]$

$= 1 - \left[\frac{6!}{0!\cdot 6!}(1)\cdot q^6 + \frac{6!}{1!\,5!} p^1 q^5\right]$

$= 1 - [q^6 + 6pq^5]$

but $P = 0.223$ which is the probability of component failing after 9 years.

$\Rightarrow q = 1 - 0.223 = 0.777$

$\therefore P(X \geq 2) = 1 - [(0.777)^6 + 6(0.223)(0.777)^5]$

$= 1 - 0.599 = 0.401$

3) Gamma Distribution (Gamma(β, θ))

→ A basic distribution of statistics for variables bounded at one side - for example x greater than or equal to zero.

→ used to model distribution of time required for exactly k independent events to occur, assuming events take place at a constant rate.

→ Gamma function is used to define this distribution.

It is defined for all $\beta > 0$ as.

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

→ if β is an integer ⇒ $\Gamma(1) = 1$ ⇒ $\Gamma(\beta) = (\beta-1)!$

→ Hence we say that gamma function is a generalization of factorial notion.

→ Gamma distribution is less variable than exponential distribution with same mean.

→ expo(λ) and gamma $(1, \theta)$ are same.

Application :- Distribution of time between re-calibrations of instrument that needs recalibration after k uses, time between inventory restocking, time to failure for a system with standby components.

PDF

$$f(x) = \begin{cases} \dfrac{\beta\theta}{\Gamma(\beta)} (\beta\theta x)^{\beta-1} e^{-\beta\theta x} & ; \text{ if } x > 0. \\ 0 & ; \text{ otherwise.} \end{cases}$$

CDF

$$F(x) = \begin{cases} 1 - \displaystyle\int_x^\infty \frac{\beta\theta}{\Gamma(\beta)} (\beta\theta t)^{\beta-1} e^{-\beta\theta t} \, dt & \text{ if } x \geq 0 \\ 0 & ; \text{ otherwise.} \end{cases}$$

Parameters Shape parameter $\beta > 0$ and scale $\theta > 0$.

Mean $E(x) = 1/\theta$.

Variance $V(x) = 1/\beta\theta^2$

Mode $\frac{1}{\theta}\left(\frac{1}{\beta} - 1\right)$; if $(1/\beta) \geq 1$

0 ; if $(1/\beta) < 1$.

Example :- Lead time is gamma distributed in 100s of unit with a shape parameter of 3 and a scale parameter of 1. what is the Probability that the lead time exceeds 2 (hundred) units during an upcoming cycle?

$\Rightarrow$ Given that shape parameter, $\beta = 3$ and scale parameter, $\theta = 1$.

Let x be the random variable that denotes lead time.

$\rightarrow$ The Probability that lead time exceeds 2.

$= P(x > 2) = 1 - P(x \leq 2)$

$= 1 - F(2)$

$$= 1 - \left[1 - \int_2^{\infty} \frac{(3)(1)}{\Gamma(3)} \left((3)(1)(t)\right)^2 \cdot e^{-(3)(1)t} \cdot dt\right.$$

$$= \int_2^{\infty} \frac{3}{2!} \cdot 9t^2 \cdot e^{-3t} \, dt$$

$$= \frac{27}{2} \int_2^{\infty} t^2 e^{-3t} \, dt$$

$$= \frac{27}{2} \left[(t^2)\left(\frac{e^{-3t}}{-3}\right) - (2t)\left(\frac{e^{-3t}}{9}\right) + 2\left(\frac{e^{-3t}}{-27}\right)\right]_2^{\infty}$$

$$= \frac{27}{2} \left[0 - \left(\frac{-4}{3} e^{-6} - \frac{4}{9} e^{-6} - \frac{2}{27} e^{-6}\right)\right]$$

$$= \frac{27}{2} \left[e^{-6} \left(4/3 + 4/9 + 2/27\right)\right]$$

$$= 25 e^{-6} = 0.062$$

7) Erlang Distribution :- Erlang(θ)

→ Let there be k stages of services arranged in series each having an exponentially distributed service time with the same mean $1/k\theta$. When an entity requires a service, it passes through all k stages with a random service time in each stage. The passage from one stage to next occurs without loss of time.

→ The next entity cannot enter the first stage until the entity in process has negotiated all stages. It can be shown that the distribution of overall service time is an Erlang distribution of kth order with a mean service time $1/\theta$.

→ We know that, the expected value of sum of random variables is the sum of expected value of each random variable.

$$E(X) = E(X_1) + E(X_2) + \cdots E(X_k).$$

but expected value of exponentially distributed x_i is $1/k\theta$.

$$\therefore E(x) = \frac{1}{k\theta} + \frac{1}{k\theta} + \cdots \frac{1}{k\theta} = \frac{1}{\theta}.$$

→ used in communication network.

$$\text{PDF} = f(x) = \frac{\beta\theta}{\Gamma(\beta)} (\beta\theta x)^{\beta-1} e^{-\beta\theta x} \quad \text{if } x>0,\ \beta = k \text{ an integer.}$$

$$= 0 \quad \text{otherwise.}$$

$$\text{CDF} = F(x) = 1 - \sum_{i=0}^{k-1} \frac{e^{-k\theta x}(k\theta x)^i}{i!} \quad \text{if } x>0$$

$$= 0 \quad ; \quad \text{otherwise}$$

1/17/20,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

→ Mean = $E(X) = 1/\theta$

→ Variance $= V(X) = 1/k\theta^2$

→ Mode $= k-1/k\theta$.

Example :- The time intervals between dial up connection of an internet service provider are exponentially distributed with a mean of 15 seconds. Find the probability that third dial-up connection occur after 30 seconds have elapsed.

⇒ This problem is defining the probability of combinations of exponential arrival processes hence the trunk usage in a telephone n/w is modeled by Erlang distribution.

here

$k\theta = 1/15$, and $x = 30$.

∴ The probability that third connection occur within 30 seconds is

$$F(30) = 1 - \sum_{i=0}^{2} \frac{e^{-\frac{1}{15}30} \cdot \left(\frac{1}{15}\cdot 30\right)^i}{i!} = 0.323$$

hence the probability that the third connection occurs after 30 seconds is $(1-0.323) = 0.677$.

5) **Normal Distribution** $= N(\mu, \sigma^2)$

→ $X \sim N(\mu, \sigma^2)$ indicates that the random variable X is normally distributed with mean μ and variance σ^2.

→ $f(\mu - x) = f(\mu + x)$ i.e PDF is symmetric about μ.

→ The maximum value of PDF occurs at $x = \mu$.. i.e mean and mode are equal.

→ Neither the CDF nor its inverse can be expressed in terms of simple mathematical function. Numerical methods are available but integration of each pair (μ, σ^2) must be evaluated. To overcome this transformation, $Z = (t - \mu)/\sigma$ allows the evaluation independent of μ and σ.

→ if $X \sim N(\mu, \sigma^2)$ and $Z = (X - \mu)/\sigma$,

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

$$= P(Z \leq (x-\mu)/\sigma) = \int_{-\infty}^{x-\mu/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \phi(z) \, dz = \phi\left(\frac{x-\mu}{\sigma}\right)$$

→ The value $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz$ where $-\infty < z < \infty$ is the PDF of normal distribution with mean 0 and variance 1.

→ Thus $Z \sim N(0,1)$ and Z has std normal distribution.

→ The CDF is given by $\phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$

E.g. Distribution of physical measurements on living organisms, intelligence test scores, product dimensions, average temperatures and so on.

→ Many methods of statistical analysis presume normal distribution.

PDF $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ if $-\infty < x < \infty$

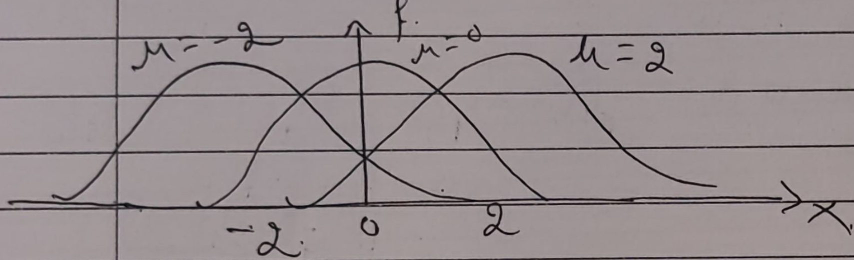
$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $-\infty < z < \infty$.

CDF $F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$.

Mean $= E(X) = \mu$.

Variance $= V(X) = \sigma^2$

Mode $= \mu$.



Example :- A certain passenger has to go Sahar international airport from hotel Taj. There are two routes from Hotel Taj to Sahar international airport. By Route A, the travel time in minutes from Hotel Taj to airport is normally distributed with $\mu = 27$ and $\sigma = 5$. But by route B it is normally distributed with $\mu = 30$, $\sigma = 2$. Which route is better choice by passenger if (i) ~~has~~ one has 30 min ii) 34 minutes ?

⇒ Given that

Route A - $\mu = 27$ and $\sigma = 5$

Route B - $\mu = 30$ and $\sigma = 2$.

Let X be a random variable that represents the travel time.

i) For Route A :- $P(X>30) = 1 - P(X \leq 30)$

$= 1 - F(30)$

$= 1 - \phi\left(\frac{30-27}{5}\right)$

$= 1 - \phi(0.6) = 1 - 0.7257$

$= 0.2743$

The value of $\phi(z)$ is computed by using table of cumulative normal distribution.

→ For Route B

$F(X>30) = 1 - P(X \leq 30)$

$= 1 - F(30) = 1 - \phi\left(\frac{30-30}{2}\right)$

$= 1 - \phi(0) = 1 - 0.5\text{?}$

$= 0.52$

Here, route A is better choice for passenger.

ii) For Route A :- $P(X>34) = 1 - P(X \leq 34)$

$= 1 - F(34)$

$= 1 - \phi\left(\frac{34-27}{5}\right)$

$= 1 - \phi(1.4) = 1 - 0.9192$

$= 0.0808$

→ For Route B

$P(X>34) = 1 - P(X \leq 34)$

$= 1 - F(34)$

$= 1 - \phi\left(\frac{34-30}{2}\right)$

$= 1 - \phi(2) = 1 - 0.9773$

$= 0.0227$

here, route B is better choice for passenger.

1/17/20,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

6) Weibull Distribution :- Weibull (β, α).

→ The weibull distribution is often used to model "time until failure"

→ It is applied in actuarial science and in engineering work.

→ It is also an appropriate distribution for describing data corresponding to resonance behaviour such as variation with energy of the cross section of a nuclear reaction or variation with velocity of absorption of radiation in the Mossbauer effect.

E.g. Life distribution for some capacitors, ball bearings, relays and so on.

Application :- General time to failure distribution due to wide diversity of hazard rate curves, and extreme value distribution for minimum of N values from distribution bounded at left.

→ Time to complete some task.

$$PDF = f(x) = \frac{\beta}{\alpha}\left(\frac{x-\nu}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-\nu}{\alpha}\right)^{\beta}} \quad ; \text{ if } x \geq \nu.$$

$$0 \quad ; \text{ otherwise}$$

When $\nu = 0$, PDF becomes

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

Put $\beta = 1$.

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

the above eqn is similar to PDF of exponential distribution with $\lambda = 1/\alpha$.

if $\alpha = 2$, in case of weibull distribution then it is called Rayleigh distribution which is special case.

$$\therefore f(x) = \frac{\beta}{2}\left(\frac{x-\gamma}{2}\right)^{\beta-1} e^{\left[-\left(\frac{x-\gamma}{2}\right)^{\beta}\right]} \quad ; \text{ if } x \geq \gamma.$$

$$= 0 \quad ; \text{ otherwise.}$$

→ Rayleigh distribution is used to model multipath fading, radiation, wind speeds.

CDF $F(x) = \begin{cases} 0 & ; \text{ if } x < \gamma. \\ 1 - e^{-(x-\gamma/\alpha)^{\beta}} & ; \text{ if } x \geq \gamma. \end{cases}$

Mean $= E(x) = \gamma + \alpha \Gamma\left(\frac{1}{\beta} + 1\right)$.

Variance $= V(x) = \alpha^2 \left[\Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right)\right]^2\right]$

Mode $= \alpha\left(\frac{\beta-1}{\beta}\right)^{1/\beta}$; if $\beta \geq 1$.

$= 0$; $\beta < 1$.

Example :- The time to failure of a nickle cadmium battery is weibull distributed with parameters $\gamma = 0$, $\beta = 1/4$ and $\alpha = 1/2$ years.

i) Find the fraction of batteries that are expected to fail on or before 1.5 yrs.

ii) What fraction of batteries are expected to last longer than the mean life?

iii) What fraction of batteries are expected to fail between 1.5 and 2.5 years?

$\Rightarrow$ Let X be a random variable that denotes the time to failure of nickle cadmium battery. hence X is Weibull distributed with parameters $\gamma = 0$, $\beta = 1/4$; $\alpha = 1/2$ and the CDF is.

$$F(x) = 1 - e^{\left[-\left(x-\gamma/\alpha\right)^{\beta}\right]}$$

i) $P(X \leq 1.5) = F(1.5) = 1 - e^{\left[-\left(\frac{1.5-0}{1/2}\right)^{1/4}\right]}$

$= 1 - 0.268 = \underline{0.732}$

ii). $E(X) = \gamma + \alpha\Gamma\left(1/\beta + 1\right)$.

$= 0 + \frac{1}{2}\Gamma\left(\frac{1}{1/4} + 1\right)$.

$= \frac{1}{2}\Gamma(5) = \frac{1}{2}(4!) = 12$.

$\therefore$ Probability that battery will last longer than 12 years is.

$P(X > 12) = 1 - P(X \leq 12)$.

$= 1 - F(12)$.

$= 1 - \left(1 - e^{-(12-0/1/2)^{1/4}}\right)$.

$= e^{-(24)^{1/4}} = \underline{0.109}$

iii) $P(1.5 \leq X \leq 2.5) = F(2.5) - F(1.5)$.

$= 1 - e^{\{-(2.5-0/1/2)^{1/4}\}} - 0.732$.

$= 1 - 0.224 - 0.732$.

$= \underline{0.044}$

7) Triangular Distribution :- Triang(a, b, c).

→ Used as a rough model in the absence of data.

→ Mode is more often used than the mean to characterize the triangular distribution.

PDF

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} ; & \text{if } a \le x \le b. \\ \dfrac{2(c-x)}{(c-b)(c-a)} ; & \text{if } b < x \le c. \\ 0 ; & \text{otherwise.} \end{cases}$$

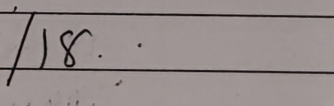
CDF

$$F(x) = \begin{cases} 0 ; & \text{if } x \le a. \\ \dfrac{(x-a)^2}{(b-a)(c-a)} ; & \text{if } a < x \le b. \\ 1 - \dfrac{(c-x)^2}{(c-b)(c-a)} ; & \text{if } b < x \le c. \\ 1 ; & \text{if } x > c. \end{cases}$$

Mean $E(X) = a+b+c/3$

$V(X) = (a^2+b^2+c^2-ac-ab-bc)/18$.

Mode $= 3E(X) - (a+c) = b$.



Example - Demand for electricity at John Schems in month of April has triangular distribution with $a = 1000$ kwh and $c = 1800$ kwh. The median is 1425 kwh. Determine the model value of kwh for the April month.

⇒ Let X is random variable that denotes the demand.

Assume that $1000 = a < \text{median} = 1425 < b = \text{mode}$.

∴ Probability demand is less than or equal to 1425 kw is

$P(X \le 1425) = F(1425)$

(Median is Point at which $F(X) = 0.5$)

1/17/20,000

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

$$\therefore 0.5 = \frac{(1425 - 1000)^2}{(b-1000)(1800-1000)}$$

$$0.5 = \frac{(425)^2}{(b-1000)(800)}$$

$$\underline{b} = \underline{1451.56 \text{ kWh}}$$

As per assumption $1451.56 > 1425$, hence
<u>mode</u> = <u>1451.56 kWh</u>.

8) Log - Normal Distribution :- $\underline{LN(\mu, \sigma^2)}$

→ Permits representation of random variable whose logarithm follows normal distribution.

→ Model for a process arising from many small multiplicative errors. Appropriate when the value of an observed variable is a random proportion of previously observed value.

→ In the case where the data are lognormally distributed, the geometric mean acts as a better data descriptor than the mean.

→ The more closely the data follow a lognormal distribution, the closer the geometric mean is to the median and hence log expression produces a symmetrical distribution.

→ E.g. Time to perform some task.

→ The ratio of two log-normally distributed variables is lognormal.

Applications :- Distribution of sizes from a breakage process distribution of income size in heritances and bank deposits, distribution of various biological phenomena, life distribution of some transistor types.

PDF $f(x) = \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2/2\sigma^2}$ if $x > 0$.

$= 0$; otherwise.

Mean :- $E(X) = e^{\mu + \frac{\sigma^2}{2}}$

Variance $= V(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$

Mode :- $e^{\mu - \sigma^2}$.

Let $Y \sim N(\mu, \sigma^2)$ then $X = e^Y$ has a lognormal distribution with parameters μ and σ^2.

→ if μ_L and σ_L^2 are mean and variance of lognormal distribution then.

$$\mu = \ln\left\{\frac{\mu_L^2}{\sqrt{\mu_L^2 + \sigma_L^2}}\right\} \text{ and } \sigma^2 = \ln\left\{\frac{\mu_L^2 + \sigma_L^2}{\mu_L^2}\right\}.$$

9) Empirical Distribution

→ It is used when a random variable has no known distribution.

→ The observed data is used to specify directly a distribution from which random values are generated during the simulation.

Continuous data

→ For continuous random variables the type of empirical distribution that can be defined on the basis of

i) Actual values of individual original observn $X_1, X_2 \ldots X_n$.

ii) The number of X_i's that fall into each of several specified intervals, which is called as grouped data or data in form of histogram.

Original data

→ If original data are available then define a continuous distribution F by first sorting the x_i's into increasing order.

→ Let $X_{(i)}$ denote the ith smallest of X_i's such that $X_{(1)} \leq X_{(2)} \ldots \leq X_{(n)}$.

Then

CDF $F(x)$ is given by

$$F(x) = \begin{cases} 0 & ; \text{ if } x < x_1 \\ \dfrac{i-1}{n} + \dfrac{X - x_{(i-1)}}{n(x_i - x_{i-1})} & ; \text{ if } x_{i-1} \leq x \leq x_i \\ = 1 & ; \text{ if } x_{(n)} \leq x \end{cases}$$

→ Note $F(x)$ rises most rapidly where X_i's are most densely distributed.

→ For each i, $F(x_{(i)}) = \frac{i-1}{n}$ which represents the approximate number of x_j's that are less than $X_{(i)}$.

Disadvantage 1) Random values generated from it during a simulation run can never be less than $x_{(1)}$ or greater than $X_{(n)}$.

2) The mean of $F(x)$ is not equal to sample mean $\bar{X}(n)$ of the X_i's

Discrete data

→ It is very easy to define empirical distribution for discrete data provided that the original data values $X_1, X_2 \cdots X_n$ are available.

→ For each possible value x, an empirical mass function $P(x)$ can be defined which is to be the proportion of X_i's that are equal to x.

(12) 1/17/20,000 Module 2.1

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

1

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ______________________

Date : ______________ 20

Examination : ______________ Branch/Semester ______________

Subject : ______________________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

Statistical Models in Simulation

→ Probability theory is the study of random phenomena which are charaterized by the fact that their future behavior is not predictable in a deterministic fashion. Such phenomena are usually represented in terms of mathematical description by constructing an ideal probabilistic model of the real world simulation. This model consists of all possible outcomes and corresponding probabilities. The probability theory allows us to predict patterns of future outcomes.

→ As we know that a model is an abstract represen' of real world problem so predictions based on model must be validated against actual results collected from the real phenomena.

→ A validation process may suggest modification to the original model. The statistical theory is useful with inductive process of drawing inference about the model and its parameters based on the limited information avaiable. Hence the statistical theory is helpful in validation process.

→ The role of probability theory is to analyze the behaviour of system as per the given distribution and corresponding probability. whereas statistical theory is helpful in choosing these probability and validating the model assumption

Why Probability and Statistics for Simulation?

1) Understand how to model a Probabilistic system.
2) Validate the simulation model.
3) Choose the input Probability distribution.
4) Generate random samples from these distributions.
5) Perform statistical analysis of simulation o/p data.
6) Design the simulation experiments.
7) Evaluate and compare alternatives.

Terminology / concepts in Probability and Statistics

1) Experiment :-
It is process whose outcome is not known with certainty. Thus it is an action or an operation which can produce any result. It is also called as trial and outcomes are called as event. e.g Rolling of a die is an experiment and getting 4 is an event.

2) Sample Space, S :-
It is set of all possible outcomes of an experiment.

3) Sample Points :-
It is possible outcomes (values) in a sample space. e.g in case of rolling a die, all the outcomes 1, 2, 3, 4, 5 and 6 together constitutes a sample space and getting anyone of the faces upwards is called a sample point.

4) Random Variables (RV) :-
It is function / rule that assign a real number to each point in the sample space S. The set of values which the random variable X takes is called the spectrum of random variable.

5) Notation:-
uppercase letters (X, Y) for RVs, lowercase for the values.

6) Discrete Random Variable:-
→ A random variable X is said to be discrete if it can take on at most countable no of values say $x_1, x_2 \dots$
→ The sample space of discrete R.V may be discrete, continuous or a mixture of discrete and continuous points but the co-domain is only discrete. e.g Marks obtained by student in mid-term test. The possible values are 0, 10, 15, 20, 25 or 30
→ The probability that discrete random variable X takes on the value x_i is given by $P(x_i) = P(X = x_i)$, such that $P(x_i) \geq 0$ and $\sum_{i=1}^{\infty} P(x_i) = 1$
→ The pair $(x_i, P(x_i))$ $i = 1, 2, \dots$ is called the probability distribution of X and $P(x_i)$ is called the probability mass function (pmf) of X.

Discrete Distributions.
Bernoulli :- Bernoulli (P).
i) A Bernoulli (P) random variable X can be thought of as the outcome of an experiment that either "fails" or "succeeds"
ii) $X_j = 1$ if jth experiment succeeds and $X_j = 0$ if jth experiment fails.
iii) The n Bernoulli trial are called Bernoulli Process if -
→ The result of each trial may be either a sucess or a failure.
→ The probability p of sucess is the same in every trial.

→ The trials are independent :- The outcome of one trial has no influence on later outcomes.

$\Longrightarrow P(x_1, x_2 \cdots x_n) = P_1(x_1)\, P_2(x_2) \cdots P_n(x_n)$.

E.g. toss a coin or a die

Applications :- i) Random events with two possible outcomes.
ii) used to generate discrete random variable such as binomial and geometric.

PMF :-

$$P_j(x_j) = P(x_j) = \begin{cases} p & \text{if } x_j = 1,\ j = 1, 2 \cdots n \\ 1-p = q & \text{if } x_j = 0. \\ 0 & \text{otherwise.} \end{cases}$$

CDF :-

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p = q & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x \end{cases}$$

Parameters $p \in (0, 1)$.

Range :- $\{0, 1\}$.

Mean :- $E(X_j) = 0 \cdot q + 1 \cdot p = p$

Variance :- $V(X_j) = [(0^2 \cdot q) + (1^2 \cdot p)] - p^2$

$= p - p^2 = p(1-p) = pq$

Mode :-

0	if $p < 1/2$
0 and 1	if $p = 1/2$
1	if $p > 1/2$.

1/17/20,000 (2)

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

2) Binomial Distribution :-

Binomial :- Bin(n, P).

i) Gives Probability of number of successes in n independent trials when Probability of success P on single trial is a constant.

ii) To determine the Probability of a particular outcome with all the success.

iii) E.g what is the Probability of 8 or more tails" in 10 tosses of a fair coin?

iv) can be sometimes approximated by normal or by Poisson distribution.

Applications :- i) Used frequently in quality control, reliability, survey sampling and other industrial problems.

→ To.

i) classify defective or non defective items in a batch of size n.

ii) find out demand (No of items) placed by a customer in case of inventory problem.

PME :- $P(x) = \binom{n}{x} p^x q^{n-x}$ if $x \in \{0, 1, \ldots n\}$ where $\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$

$= 0$ otherwise.

CDF :-

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sum_{i=0}^{x} \binom{n}{x} p^i q^{n-i} & \text{if } 0 \leq x \leq n \\ 1 & \text{if } n < x. \end{cases}$$

Parameters :- n is +ve integer, $p \in (0,1)$.

Range :- $\{0, 1, \dots n\}$.

Mean :- $E(X) = p + p + \dots + p = np$.

Variance :- $V(X) = pq + pq + \dots + pq = npq$.

Mode :- $= p(n+1) - 1$ and $p(n+1)$. if $p(n+1)$ is an integer.

$= \lfloor p(n+1) \rfloor$ otherwise.

Example :-

The probability of a computer chip failure is 0.05. Everyday a random sample of size 14 is taken. What is the probability that

i) at most 3 will fail.

ii) at least 3 will fail.

$\Rightarrow$ Given that sample size $= n = 14$.

here $p = 0.05 \Rightarrow q = 0.95$.

The total no of failure chips in the sample. X, would be modeled by binomial distribution and it is given by

$$P(x) = \begin{cases} \binom{14}{x} (0.05)^x (0.95)^{14-x} & , x = 0, 1, 2, \dots 14 \\ 0 & , \text{otherwise}. \end{cases}$$

i) P(at most 3 chips will fail) = $P(X \leq 3)$ =
$P(0) + P(1) + P(2) + P(3)$.

$$= \sum_{x=0}^{3} \binom{14}{x} (0.05)^x (0.95)^{14-x}$$

$$= \frac{14!}{0!\cdot 14!} (0.05)^0 (0.95)^{14} + \frac{14!}{1!\cdot 13!} (0.05)^1 (0.95)^{13} + \frac{14!}{2!\cdot 12!} (0.05)^2 (0.95)^{12} + \frac{14!}{3!\cdot 11!} (0.05)^3 (0.95)^{11}.$$

$$= (0.95)^{14} + 14(0.05)(0.95)^{13} + 91(0.05)^2 (0.95)^{12} + 364(0.05)^3 (0.95)^{11}$$

$$= 0.488 + 0.359 + 0.123 + 0.021.$$
$$= 0.996$$

ii) P(at least 3 chips will fail) = $P(X \geq 3)$.
$= 1 - P(X < 3)$.

$= 1 - (P(0) + P(1) + P(2))$
$= 1 - (0.488 + 0.359 + 0.123)$.
$= 1 - 0.97$
$= 0.03$

Negative Binomial Distribution :- is the distribution of the number of the trials until the kth success, for $k = 1, 2 \ldots$ If Y has a -ve binomial distribution with parameters p and k, then the distribution of Y is given by.

$$P(y) = \begin{cases} \binom{y-1}{k-1} q^{y-k} p^k, & y = k, k+1, k+2 \ldots \\ 0, & \text{otherwise.} \end{cases}$$

Becoz we can think of the -ve binomial random variable X as the sum of k independent geometric random variables it is easy to see that $E(X) = k/p$ and $V(X) = kq/p^2$

Example :- Forty percent of assembled ink-jet printers are rejected at the inspection station. Find the probability that the first acceptable ink-jet printer is the third one inspected.

Considering each inspection as a Bernoulli trial with $q = 0.4$ and $P = 0.6$ yields.

$$P(3) = 0.4^2 (0.6) = 0.096.$$

→ Thus in only about 10% of the cases is the first acceptable printer the third one from any arbitary starting point. To determine the probability that the third printer inspected is the second acceptable printer, we use the negative binomial distribution,

$$P(3) = \binom{3-1}{2-1} 0.4^{3-2} (0.6)^2$$

$$= \binom{2}{1} (0.4) (0.6)^2$$

$$= 0.288$$

1/17/20,000

(3)

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

3) Geometric Distribution

→ Geometric : geom(P).

→ Gives Probability of requiring exactly x Bernoulli trials before the first success is achieved.

→ E.g. Determination of Probability of requiring exactly five tests firings before first success is achieved.

→ Only discrete distribution with memory less Property.

→ If an event has not occured during first 'm' repetitions of an experiment, then the probability that it will not occur during the next 'n' repetitions is same as the Probability that it will not occur during the first 'n' repetitions.

Applications :- → Used in quality control, reliability and other industrial situations.

→ To find out no of items inspected before the first defective item is encountered. It has important application in queuing theory (no of units which are being served or waiting to be served at any given time).

→ PMF $= P(x) = q^{x-1} \cdot p$, if $x \in \{1, 2 \cdots\}$

$= 0$, otherwise.

→ CDF $F(x) = 1 - q^{x}$, if $x \geq 1$

$= 0$, otherwise.

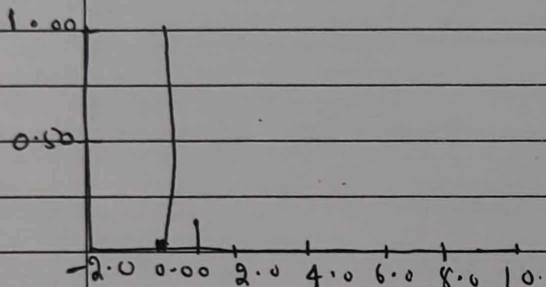
→ Parameters - $p \in (0,1)$.

→ Range $= \{1, 2, \cdots\}$.

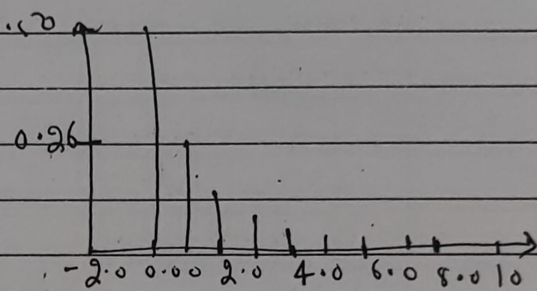
→ Mean $=$ $E(X) = 1/p$

→ Variance $= V(X) = q/p^2$

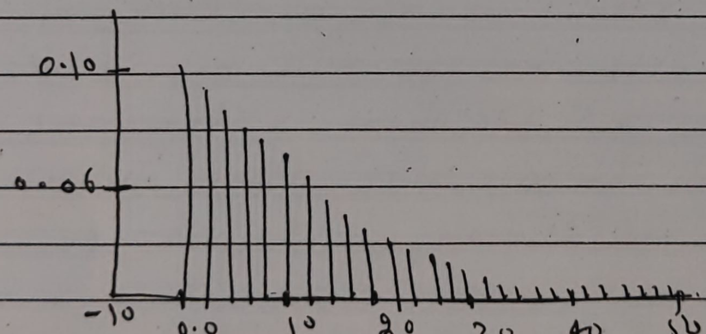
→ Mode $= 0$



Geometric (0.9)



Geometric (0.5)



Geometric (0.1)

PMF of Geometric distribution [40].

* A computer lab assistant types 2 documents erroneously for every 100 documents. What is the probability that the tenth document typed is the first erroneous document?

⇒ Consider typing of each document as a Bernoulli trial with $2/100 = \underline{0.02}$ and $q = \underline{0.98}$.

→ By geometric distribution,
P(the tenth document typed is first erroneous document)

$$= P(10) = 0.02(0.98)^9 = \underline{0.0167}.$$

4) Poisson Distribution

Poisson (α).

→ Gives probability of exactly x independent occurrences during a given period of time if events take place independently and a constant rate.

→ May also represent no of occurrences over constant areas or volumes.

→ E.g. Used to represent distribution of no of defects in a piece of material, customer arrivals, insurance claims, incoming telephone calls, alpha particles emitted and so on.

→ Frequently used as approximation to binomial distribution.

Application :- Used frequently in quality control, reliability, queuing theory and so on...

PMF $P(x) = \frac{e^{-\alpha}\alpha^{x}}{x!}$ if $x \in 0,1,2\ldots$

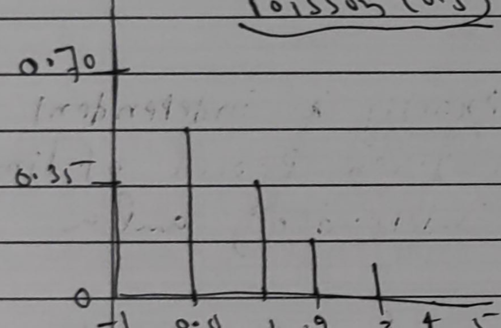
$= 0$ otherwise.

CDF $F(X) = 0$ if $x < 0$

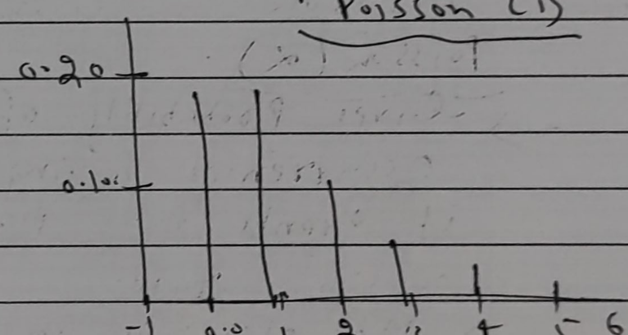
$= \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^{i}}{i!}$ if $x \geq 0$.

Parameters	$\alpha > 0$.
Range	$\{0, 1, \ldots\}$.
Mean	$E(X) = \alpha$
Variance	$V(X) = \alpha$.
Mode	$\alpha - 1$ and α if α is an integer.
	$\lfloor \alpha \rfloor$ otherwise.

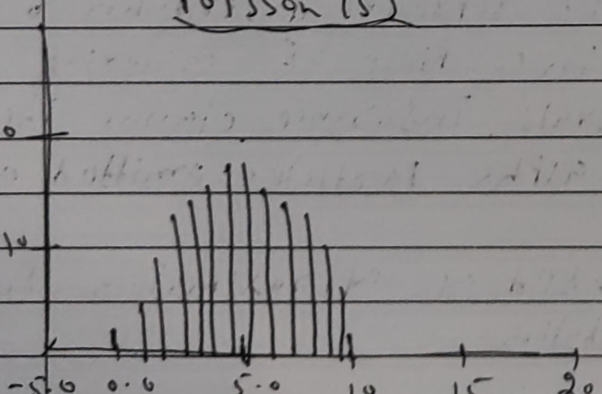
Poisson (0.5)



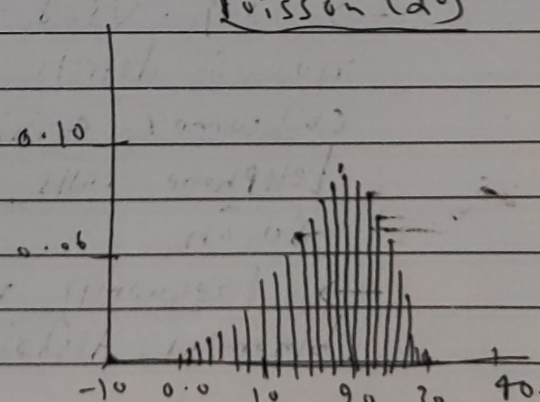
Poisson (1)



Poisson (5)



Poisson (20)



The PMF of Poisson distribution [40].

1/17/20,000 (4)

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________ (In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

The number of accidents in a year to taxi driver in Mumbai follows a Poisson distribution with mean equal to 3. Out of 100 taxi drivers, find approximately the no of drivers with:-

i) no accident in a year.

ii) more than 3 accidents in a year.

⇒ Given that mean $\alpha = 3$

Let x be a random variable which represents the no of accidents in a year.

i) The Probability that zero / no accident in a year $= P(0) = \frac{e^{-\alpha}\alpha^0}{0!} = \frac{e^{-3}\cdot(3)^0}{1}$

$= 0.0498$

∴ Out of 100 taxi drivers the no of taxi drivers with no accidents in a year is $0.0498 \times 100 = 4.98 \simeq 5$.

2) Probability that more than 3 accidents in a year

$= P(X>3) = 1 - P(X \leq 3)$

$= 1 - F(3)$

$= 1 - \sum_{i=0}^{3} e^{-3}\cdot(3)^i / i!$

$= 1 - e^{-3}(1+3+4.5+4.5)$

$= 1 - 0.6474 = 0.3526$

$\therefore$ out of 100 taxi drivers, the no of taxi drivers with more than 3 accidents in a year is $0.3526 \times 100 = 35.26 \simeq \underline{\underline{35}}$

Example 1) A Production process manufactures alternators for outboard engines used in recreational boating. On the average 1% of the alternators will not perform up to the required standards when ~~used~~ tested at the engine assembly plant. When a large shipment of alternators is received at the plant, 100 are tested and if more than two are non conforming the shipment is returned to the alternator manufacturer. What is the probability of returning a shipment.

$\Rightarrow$ Let X be defined as no of defectives in the sample. Then X is binomial $(n = 100, P = 0.01)$ with the probability mass function.

$\rightarrow$ $$P(x) = \binom{100}{x} (0.01)^x (0.99)^{100-x}, \quad x = 0, 1, \ldots 100.$$

$\rightarrow$ The probability of returning the shipment is.

$$P(X > 2) = 1 - P(X \leq 2).$$

$$1 - \binom{100}{0} (0.99)^{100} - \binom{100}{1} (0.01)(0.99)^{99} - \binom{100}{2} (0.01)^2 (0.99)^{98} = \cancel{0.794}$$

$$= \underline{\underline{0.0795}}$$

1 - 0.3(6 - 0.3697 - 0.1848

2). ~~An industrial chemical that will retard the spread of fire in paint has been developed~~. The local sales representative has estimated from past experience that 48% of sales calls will result in an order

a) What is the probability that first order will come on the fourth sales call of the day?

b) If eight sales calls are made in a day, what is Probability of receiving exactly six orders?

c) If four sales calls are made before lunch, what is the Probability that one or fewer results in an order?

⇒ Let x be defined as the number of calls received until an order is Placed. Then, x is geometric (P=0.48) with the Probability mass function

$$P(x) = (0.52)^{x-1}(0.48), \quad x = 0, 1, 2 \ldots$$

a) The Probability that the first order will come on the fourth call is

$$P(4) = 0.0675.$$

b) The number of orders y in eight calls is binomial (n=8, P=0.48) with Probability mass function.

$$P(y) = \binom{8}{y}(0.48)^{y}(0.52)^{8-y}, \quad y = 0, 1, \ldots 8.$$

The Probability of receiving exactly six orders in eight calls is

$$P(6) = 0.0926.$$

c) The number of orders x in four calls is binomial (n=4, P=0.48) with Probability mass function

$$P(x) = \binom{4}{x}(0.48)^{x}(0.52)^{8-x}, \quad x = 0, 1, 2, 3, 4.$$

Example

3) The Hawks are currently winning 0.55 of their games. There are 5 games in the next two weeks. What is the probability that they will win more games than they lose?

⇒ Let x be defined as no of games won in the next two weeks. The random variable x is described by binomial distribution.

$$P(x) = \binom{5}{x}(0.55)^x (0.45)^{5-x}$$

$$P(3 \le x \le 5) = P(3) + P(4) + P(5)$$

$$= \binom{5}{3}(0.55)^3(0.45)^2 + \binom{5}{4}(0.55)^4(0.45) + \binom{5}{5}(0.45)^0(0.55)^5$$

$$= 0.337 + 0.206 + 0.0500 \approx \underline{0.593}$$

4). The number of hurricanes hitting the coast of Florida annually has a Poisson distribution with a mean of 0.8.

a) What is probability that more than two hurricanes will hit the Florida coast in a year?

b) What is the probability that exactly one hurricane will hit the cost of Florida in a year?

⇒ The no of hurricanes per year, x, is a Poisson ($\alpha = 0.8$) with PMF

$$P(x) = e^{-0.8}(0.8)^x / x!, \quad x = 0, 1, \ldots$$

a) $P(X > 2) = 1 - P(x \le 2)$

$$= 1 - e^{-0.8} - e^{-0.8}(0.8) - e^{-0.8}(0.8^2/2)$$

$$= \underline{0.0474}$$

b) $\underline{P(1)} = \underline{0.3595}$

1/17/20,000

(5)

K. J. SOMAIYA COLLEGE OF ENGINEERING

(AFFILIATED TO THE UNIVERSITY OF MUMBAI)

Candidate Roll No. ____________
(In figures)

Test Exam.

Name : ____________

Date : ____________ 20

Examination : ____________ Branch/Semester ____________

Subject : ____________

Junior Supervisor's full Signature with Date

Question No.	1	2	3	4	5	6	7	8	9	10	11	12	Total
Marks Obtained													

Example:- A college Professor of electrical engg is leaving home for the summer, but would like to have a light burning at all times to discourage burglars. The professor rigs up a device that will hold two light bulbs. The device will switch the current to second bulb if first bulb fails. The box in which the light bulbs are packaged says. "Average life 1000 hrs, exponentially distributed." The Professor will be gone 90 days (2160 hrs). What is the Probability that a light will be burning when the summer is over and Professor returns?

⇒ The probability that system will operate at least x hrs is called the reliability function R(x).

$$R(x) = 1 - F(x).$$

⇒ In this case, the total system lifetime is given by ESh with $\beta = k = 2$ bulbs and $k\theta = 1/1000$ per hr, so $\theta = 1/2000$ per hr. Thus $F(2160)$ can be determined as follows.

$$F(2160) = 1 - \sum_{i=0}^{1} \frac{e^{-2(1/2000)(2160)}\,[(2)(1/2000)(2160)]^{i}}{i!}$$

$$= 1 - e^{-2.16} \sum_{i=0}^{1} \frac{(2.16)^{i}}{i!} = 0.636$$

∴ The chances are about 36% that a light will be burning when the professor returns.

Example: A medical examination is given in three stages by a physician. Each stage is exponentially distributed with a mean service time of 20 minutes. Find the probability that the exam will take 50 minute or less. Also, compute the expected length of the exam.

⟹ In this case $k = 3$ stages and $k\theta = 1/20$, so that $\theta = 1/60$ per min. Thus $F(50)$ is

$$F(50) = 1 - \sum_{i=0}^{2} \frac{e^{-(3)(1/60)(50)} [(3)(1/60)(50)]^i}{i!}$$

$$= 1 - \sum \frac{e^{-5/2} (5/2)^i}{i!}$$

→ The cumulative Poisson distribution shown in table A4 can be used to calculate that.

$$F(50) = 1 - 0.543 = 0.457.$$

→ The probability is 0.457 that the exam will take 50 minutes or less. The expected length of exam is found from:

$$E(X) = 1/\theta = 1/(1/60) = \underline{60 \text{ min}}$$

→ The variance of X is $V(X) = 1/\beta\theta^2 = 1200 \text{ min}^2$ – incidentally the mode of the Erlang distribution is.

Mode $= k-1/k\theta$.

Thus the model value in this example is.

$$\text{Mode} = \frac{3-1}{3(1/60)} = \underline{40 \text{ min}}.$$

Example :- The time to failure for a component is known to have a Weibull distribution with $\gamma = 0$, $\alpha = 200$ hrs. and $\beta = 1/3$. The mean time to failure is given by

$\Rightarrow$ $E(X) = 200\,\Gamma(3+1) = 200(3!) = 1200$ hrs.

$\rightarrow$ The Probability that unit fails before 2000 hrs is computed as.

$$F(2000) = 1 - \exp\left[-\left(\frac{2000}{200}\right)^{1/3}\right]$$

$$= 1 - e^{-\sqrt[3]{10}} = 1 - e^{-2.15} = 0.884.$$

Example :- Records indicate that 1.8% of the entering students at a large state university drop out of school by midterm. What is the probability that three or fewer students will drop out of a random group of 200 entering students?

$\Rightarrow$ Using Poisson approximation with the mean α, given by

$$\alpha = np = 200 ~~(0.018)~~ (0.018) = 3.6$$

$\rightarrow$ The Probability that $0 \le x \le 3$ students will drop out of school is given by.

$$F(3) = \sum_{x=0}^{3} \frac{e^{-\alpha}\alpha^{x}}{x!} = 0.5148$$