(d) What amount of time ...
within $t \leq 90\%$ of the time?

Reliability and Failure rates:

Let the random variable X be the lifetime or the time to failure of a component.
The probability that the component survives until some time t is called the reliability R(t) of the component.
Thus $R(t) = P(X > t) = 1 - F(t)$ where F is the distribution function of the component lifetime X. The component is normally assumed to be working properly at time $t=0$ $[R(0)=1]$ & no component can work forever without failure $\lim_{t \to +\infty} R(t) = 0$
R(t) is a monotone decreasing function of t.
For t less than 0, reliability has no meaning, but we let $R(t) = 1$ for $t<0$. F(t) will often be called the unreliability

Consider a fixed number of identical components N_0, under test.
After time t, $N_f(t)$ components have failed and $N_s(t)$ components have survived with $N_f(t) + N_s(t) = N_0$.
The estimated probability of survival may be written (using the frequency interpretation of probability) as

$$\hat{P}\text{ (survival)} = \frac{N_s(t)}{N_0}$$

In the limit as $N_0 \to \infty$, we expect $\hat{P}$ (survival) to approach R(t)
As the test progresses, $N_s(t)$ gets smaller and R(t) decreases.

$$R(t) \simeq \frac{N_s(t)}{N_0}$$

$$= \frac{N_0 - N_f(t)}{N_0}$$

$$= 1 - \frac{N_f(t)}{N_0}$$

The total number of components N_0 is constant, while the number of failed components N_f increases with time.
Taking derivatives on both sides of the preceding eq^n we get

$$R'(t) = -\frac{1}{N_0} N_f'(t)$$

In this equation $N_f'(t)$ is the rate at which components fail.
Therefore as $N_0 \to \infty$, the right-hand side may be interpreted as the negative of the failure density function $f_X(t)$:

$$R'(t) = -f_X(t)$$

$f(t)\Delta t$ is the (unconditional) probability that a component will fail in the interval $(t, t+\Delta t)$.
However if we have observed the component functioning upto some time t, we expect the (conditional) probability of its failure to be different from $f(t)\Delta t$.
This leads us to the notion of instantaneous failure rate as follows.

The conditional probability that the component does not survive for an (additional) interval of duration x given that it has survived until time t can be written as

$$G_Y(x|t) = \frac{P(t < X \leq t+x)}{P(X > t)} = \frac{F(t+x) - F(t)}{R(t)}$$

Definition (Instantaneous Failure Rate):

The instantaneous failure rate $h(t)$ at time t is defined to be

$$h(t) = \lim_{x\to 0} \frac{1}{x} \frac{F(t+x)-F(t)}{R(t)}$$

$$= \lim_{x\to 0} \frac{R(t)-R(t+x)}{xR(t)}$$ so that

$$h(t) = \frac{f(t)}{R(t)}$$

Thus $h(t)\Delta t$ represents the conditional probability that a component having survived to age t will fail in the interval $(t, t+\Delta t)$.

$h(t)$ is called hazard rate, force of mortality, intensity rate, conditional failure rate, or simply failure rate.

It should be noted that the exponential distribution is characterized by a constant failure rate, since

$$h(t) = \frac{f(t)}{h(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

By integrating both sides of equation, we get

$$\int_0^t h(x)dx = \int_0^t \frac{f(x)}{R(x)} dx$$

$$= \int_0^t -\frac{R'(x)}{R(x)} dx$$

$$= -\int_{R(0)}^{R(t)} \frac{dR}{R'}$$

$$\int_0^t h(x)dx = -\ln R(t)$$

using the boundary condition $R(0)=1$

$$\therefore R(t) = \exp\left[-\int_0^t h(x)dx\right]$$

This formula holds even when the distribution of the time to failure is not exponential.

The cumulative failure rate $H(t) = \int_0^t h(x)dx$ is referred to as the cumulative hazard.

Reliability interms of cumulative hazard is

$$R(t) = e^{-H(t)}$$

1) The failure rate of a certain component is $h(t) = \lambda_0 t$ when $\lambda_0 > 0$ is a given constant. Determine the reliability $R(t)$ of the component. Repeat for $h(t) = \lambda_0 t^{1/2}$

Given that $h(t) = \lambda_0 t \Rightarrow h(x) = \lambda_0 x$

$$H(t) = \int_0^t h(x)dx$$

$$= \int_0^t \lambda_0 x \, dx$$

$$= \lambda_0 \frac{x^2}{2}\Big|_0^t$$

$$= \lambda_0 \frac{t^2}{2}$$

$$R(t) = e^{-H(t)}$$

$$= e^{-\frac{\lambda_0 t^2}{2}}$$

$h(t) = \lambda_0 t^{1/2} \Rightarrow h(x) = \lambda_0 x^{1/2}$

$$H(t) = \int_0^t h(x)dx = \int_0^t \lambda_0 x^{1/2}$$

$$= \lambda_0 \frac{x^{\frac{1}{2}+1}}{1/2+1}\Big|_0^t$$

$$= \lambda_0 \frac{x^{3/2}}{3/2}\Big|_0^t = \frac{2\lambda_0 x^{3/2}}{3}\Big|_0^t$$

$$= \frac{2\lambda_0 t^{3/2}}{3}$$

$$R(t) = e^{-H(t)} = e^{-\frac{2\lambda_0 t^{3/2}}{3}}$$

⇒ The failure rate of a computer system for onboard control of a space vehicle is estimated to be the following function of time:

$$h(t) = \alpha\mu t^{\alpha-1} + \beta\gamma t^{\beta-1}$$

Derive an expression for the reliability $R(t)$ of the system. Plot $h(t)$ and $R(t)$ as functions of time with parameter values $\alpha = \frac{1}{4}$, $\beta = \frac{1}{7}$, $\mu = 0.0004$ and $\gamma = 0.0007$

$$H(t) = \int_0^t h(x)dx$$

$$= \int_0^t \alpha\mu x^{\alpha-1} + \beta\gamma x^{\beta-1}$$

$$= \left[\alpha\mu \frac{x^{\alpha-1+1}}{\alpha-1+1} + \beta\gamma \frac{x^{\beta-1+1}}{\beta-1+1}\right]_0^t = \frac{\alpha\mu x^{\alpha}}{\alpha} + \frac{\beta\gamma x^{\beta}}{\beta}$$

$$H(t) = \mu t^{\alpha} + \gamma t^{\beta}$$

$R(t) = e^{-H(t)}$

$= e^{-(ut^{\alpha} + rt^{\beta})}$

$t=0$, $h(0) = \alpha u(0)^{\alpha-1} + \beta r(0)^{\beta-1}$

$= \frac{1}{4}(0.0004)(0) + \frac{1}{7}(0.0007)(0)$

$= 0$

$t=1$, $h(1) = \frac{1}{4}(0.0004)(1)^{\frac{1}{4}-1} + \frac{1}{7}(0.0007)(1)^{\frac{1}{7}-1}$

$= (0.0001)(1)^{-3/4} + (0.0001)(1)^{-6/7}$

$= \underline{\underline{0.0002}}$

$1e^{-5} = 1\times10^{-5}$

$5e^{-5} = 5\times10^{-5}$

$t=2$, $h(2) = (0.0001)(2)^{-3/4} + (0.0001)(2)^{-6/7}$

$= 0.000059 + 0.000055 = 0.00011$

$t=3$, $h(3) = (0.0001)(3)^{-3/4} + (0.0001)(3)^{-6/7}$

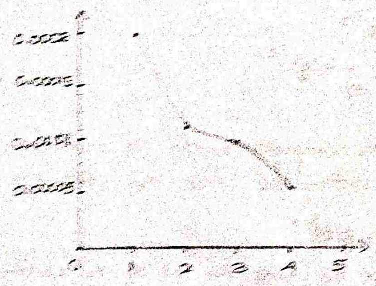
$= 3.28\times10^{-5}$

$= 0.0000823$

$t=4$, $h(4) = (0.0001)(4)^{-3/4} + (0.0001)(4)^{-6/7}$

$= 0.0000658$

$t=5$, $h(5) = 0.000055$



$R(t) = e^{-(ut^{\alpha} + rt^{\beta})}$

$t=0$, $R(0) = e^{-(u(0)^{\alpha} + r(0)^{\beta})}$

$= 1$

$t=1$, $R(1) = e^{-(0.0004(1)^{1/4} + 0.0007(1)^{1/7})}$

$= e^{-(0.0011)}$

$= 0.9989$

$t=2$, $R(2) = e^{-(0.0004(2)^{1/4} + 0.0007(2)^{1/7})}$

$= e$

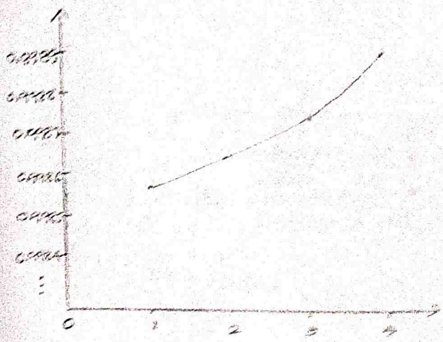
$= 0.99875$

$t=3$, $R(3) = e^{-(0.0004(3)^{1/4} + 0.0007(3)^{1/7})}$

$= 0.99865$

$t=4$, $R(4) = e^{-(0.0004(4)^{1/4} + 0.0007(4)^{1/7})}$

$= 0.99858$



Hyperexponential Distribution:

Many processes in nature can be divided into sequential phases. If the time the process spends in each phase is independent and exponentially distributed, then it can be shown that the overall time is hyperexponentially distributed.

Eg: Service time for input-output operations in a computer system often possess this distribution.

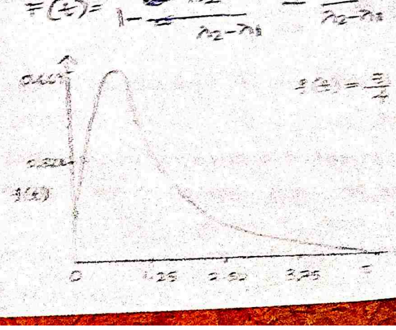
The distribution has 2 parameters, one for each of its distinct phases. A two stage hyperexponential r.v. X with parameters λ_1 and λ_2 $(\lambda_1 \neq \lambda_2)$ will be denoted by $X \sim Hypo(\lambda_1, \lambda_2)$ and its pdf is given by

$$f(t) = \frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right), \quad t>0 \quad \text{—(1)}$$

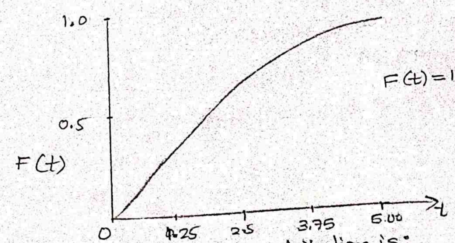
The dist^n function is

$$F(t) = 1 - \frac{\lambda_2 e^{-\lambda_1 t}}{\lambda_2-\lambda_1} + \frac{\lambda_1}{\lambda_2-\lambda_1} e^{-\lambda_2 t} \quad t\geq 0 \quad \text{—(2)}$$

$f(t) = \frac{3}{4}\left(e^{-t} - e^{-5t}\right)$



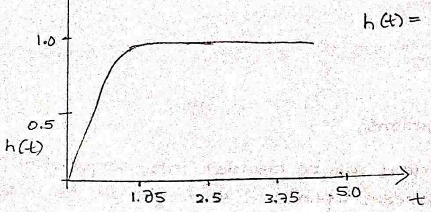
(2) Fig. CDF of Hypo



$$F(t) = 1 - \left(\frac{5}{4}\right)e^{-t} - \left(\frac{1}{4}\right)e^{-5t}$$

The hazard rate of this distribution is:

$$h(t) = \frac{\lambda_1\lambda_2\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right)}{\lambda_2 e^{-\lambda_1 t} - \lambda_1 e^{-\lambda_2 t}} \quad \text{—} (3)$$



$$h(t) = \frac{5\left(e^{-t} - e^{-5t}\right)}{5e^{-t} - e^{-5t}}$$

Erlang and Gamma distribution:

When r sequential phases have identical exponential distribution, then the resulting density is known as r-stage (or r-phase) Erlang and is given by:

$$f(t) = \frac{\lambda^r t^{r-1} e^{-\lambda t}}{(r-1)!}, \quad t>0, \lambda>0, r=1,2,3\ldots \quad \text{—} (4)$$

The CDF is

$$F(t) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t\geq 0, \lambda>0, r=1,2,3.. \quad (5)$$

Also,

$$h(t) = \frac{\lambda^r t^{r-1}}{(r-1)! \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!}}, \quad t>0, \lambda>0, r=1,2,3,\ldots \quad (6)$$

The exponential distribution is special case of the Erlang distribution with $r=1$.

Consider a component subjected to an environment so that N_t, the number of peak stresses in the interval $(0,t]$ is poisson distr. with parameter λt.

Suppose further that the component can withstand $(r-1)$ peak stress and the rth occurrance of a peak stress causes a failure.

Then the component lifetime X is related to N_t, so that the following two events are equivalent:

$$[X>t] = [N_t < r]$$

Thus,

$$R(t) = P(X>t) = P(N_t<r) = \sum_{k=0}^{r-1} P(N_t=k) = e^{-\lambda t}\sum_{k=0}^{r-1}\frac{(\lambda t)^k}{k!}$$

Then $F(t) = 1 - R(t)$. We conclude that the component lifetime has an r stage Erlang distribution.

If we let r (call it α) take nonintegral values, then we get the gamma density

$$f(t) = \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}, \quad \alpha>0, t>0 \quad \text{—} (8)$$

where the gamma function is defined by the following integral

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \quad \alpha>0 \quad \text{—} (9)$$

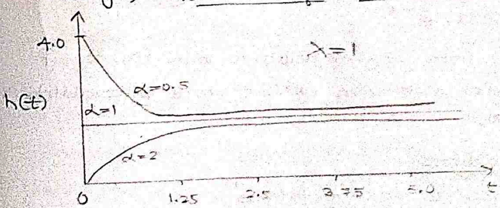
A random variable X with pdf (8) will be denoted by $X \sim GAM(\lambda, \alpha)$. This distr has 2 parameters

α = shape parameter, since as α increases, the density becomes more peaked.

λ = scale parameter i.e the distribution depends on λ only through the product λt.

The Gamma distr. is DFR for $0<\alpha<1$ & IFR for $\alpha>1$

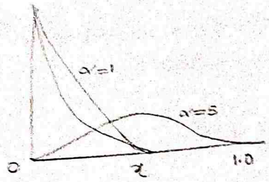
(Fig 3) Failure rate of Gamma distribution:



For $\alpha=1$, the distr. degenerates to the exponential distr i.e $Exp(\lambda) = GAM(\lambda, 1)$

The Gamma distribution is the continuous counterpart of the negative binomial distr.

The chi-square distr is a special case of the gamma distr with $\alpha = n/2$ (n is the integer) & $\lambda = \frac{1}{2}$. Thus if $X \sim GAM(\frac{1}{2}, n/2)$ then it is said to have a chisquare distr with n degrees of freedom.



Gamma pdf

Hyperexponential Distribution:

A process with sequential phases give rise to a hypoexponential or an Erlang distribution, depending on whether the phases have idential distribution.

Instead, if a process consist of alternate phases. i.e during any single experiment, the process experience one and only one of the many alternate phases. and these phases have exponential distribution, then the overall distribution is hyperexponential.

The density function of a k-phase hyperexponential random variable is

$$f(t) = \sum_{i=1}^{k} \alpha_i \lambda_i e^{-\lambda_i t} \quad t>0,\ \lambda_i>0,\ \alpha_i>0\ \sum_{i=1}^{k}\alpha_i = 1 \quad —(10)$$

and the CDF is

$$F(t) = \sum_i \alpha_i (1-e^{-\lambda_i t}),\ t\geq 0 \quad —(11)$$

The failure rate is,

$$h(t) = \frac{\sum \alpha_i \lambda_i e^{-\lambda_i t}}{\sum \alpha_i e^{-\lambda_i t}},\ t\geq 0 \quad —(12)$$

has a DFR from $\sum \alpha_i \lambda_i$ down to min $\{\lambda_1, \lambda_2, \dots\}$

Hyperexponential distribution exhibits more variability then the exponential.

Weibull distribution:

It is been used to describe fatigue[1].

Electronic component failure and ball becoming failure[2].

[1] extreme redness resulting from mental or physical illness

[2] due to improper/Inadequate lubrication, contamination, overload, improper handling and installation

It is most widely used parametric family of failure distribution.

The density function is:

$$f(t) = \lambda\alpha t^{\alpha-1} e^{-\lambda t^{\alpha}} \quad —(13)$$

The CDF is

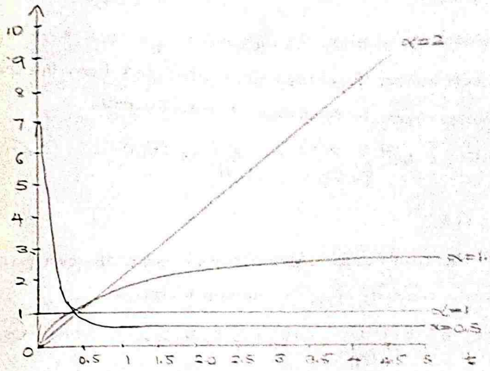
$$F(t) = 1-e^{-\lambda t^{\alpha}} \quad —(14)$$

and the failure rate is

$$h(t) = \lambda\alpha t^{\alpha-1} \quad —(15)$$

and the cummulative hazard is a power function

$H(t) = \lambda t^{\alpha}$ for all (13), (14), (15) $t\geq 0,\ \lambda\geq 0$



Failure rate of weibull distribution with various values of α and λ=1

Conditional probability:- P(A|B) for two events A and B, we can define the conditional probability $P(A|X=x)$ of two event A, given that the event $[X=x]$ has occured as

$$P(A|X=x) = \frac{P(A \text{ occurs and } X=x)}{P(X=x)} \quad —(1)$$

2 Dimensional random variable:

Joint distributions: conditional pmf:

Let X and Y be discrete random variables having a joint pmf $p(x,y)$. The conditional pmf of Y given X is

$$P_{Y|X}(y|x) = P(Y=y|X=x)$$

$$= \frac{P(Y=y, X=x)}{P(X=x)} \quad \rightarrow (2)$$

$$= \frac{P(x,y)}{P_X(x)} \quad \text{if } P_X(x)\neq 0$$

$$P(x,y) = P_X(x)\,P_{Y|X}(y|x) = P_Y(y)\,P_{X|Y}(x|y) \quad (3)$$

If X and Y are Independent, then $P_{Y|X}(y|x) = P_Y(y)$ (4)

$$P_Y(y) = \sum_{\text{all } x} P(x,y) = \sum_{\text{all } x} P_{Y|X}(y|x)\,P_X(x) \quad (5)$$

This is another form of the theorem of total probability.

conditional distribution fn ~~CDF~~ is given by

$$F_{Y|X}(y|x) = P(Y \le y \mid X=x) = \frac{P(Y \le y \text{ and } X=x)}{P(X=x)} \quad (6)$$

for all values of y and for all values of x such that $P(X=x)>0$

Conditional distribution function can be obtained from the conditional pmf (assuming that both X and Y are discrete):

$$F_{Y|X}(y|x) = \frac{\sum_{t \le y} P(x,t)}{P_X(x)} = \sum_{t \le y} P_{Y|X}(t|x) \quad (7)$$

Conditional pdf:

Let X and Y be continuous random variables with joint pdf $f(x,y)$

The conditional density $f_{Y|X}$ is defined by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \text{ if } 0 < f_X(x) < \infty \quad (8)$$

It follows from the definition of conditional density that

$$f(x,y) = f_X(x)\,f_{Y|X}(y|x) = f_Y(y)\,f_{X|Y}(x|y) \quad (9)$$

If X & Y are independent, then

$f(x,y) = f_X(x)\,f_Y(y)$ which implies that

$$f_{Y|X}(y|x) = f_Y(y) \quad (10)$$

It is necessary & sufficient for the marginal density of Y in terms of conditional density by integration

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

$$= \int_{-\infty}^{\infty} f_X(x)\,f_{Y|X}(y|x)\,dx \quad (11)$$

This is the continuous analog of the theorem of total probability.

Further in the definition of conditional density we can reverse the role of X and Y to define (whenever $f_Y(y) > 0$):

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Using the expression (11) for $f_Y(y)$ and noting that $f(x,y) = f_X(x)\,f_{Y|X}(y|x)$,

$$f_{X|Y}(x|y) = \frac{f_X(x)\,f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x)\,f_{Y|X}(y|x)\,dx} \quad (12)$$

This is continuous analog of Baye's rule.

The conditional pdf can be used to obtain the conditional probability

$$P(a \le Y \le b \mid X=x) = \int_a^b f_{Y|X}(y|x)\,dy, \quad a \le b \quad (13)$$

In particular the conditional distribution function $F_{Y|X}(y|x)$ is defined, analogous to (6) as

$$F_{Y|X}(y|x) = P(Y \le y \mid X=x) = \frac{\int_{-\infty}^{y} f(x,t)\,dt}{f_X(x)}$$

$$= \int_{-\infty}^{y} f_{Y|X}(t|x)\,dt \quad (14)$$

Exercise on two dimensional random variables:

1) The joint probability mass function (pmf) of X & Y is given the table below

	1	2	3	4	5	6	$P_X(x)=f(x)$
0	0 P(0,1)	0 P(0,2)	1/32 P(0,3)	2/32 P(0,4)	2/32 P(0,5)	3/32 P(0,6)	P(X=0) = 8/32
1	1/16 P(1,1)	1/16 P(1,2)	1/8 P(1,3)	1/8 P(1,4)	1/8 P(1,5)	1/8 P(1,6)	P(X=1) = 20/32
2	1/32 P(2,1)	1/32 P(2,2)	1/64 P(2,3)	1/64 P(2,4)	0 P(2,5)	2/64 P(2,6)	P(X=2) = 4/32
$P_Y(y) = f(y)$	P(Y=1) = 3/32	P(Y=2) = 3/32	P(Y=3) = 11/64	P(Y=4) = 13/64	P(Y=5) = 6/32	P(Y=6) = 16/64	1

Compute the following:

(a) $P(X \le 1)$
(b) $P(Y \le 3)$
(c) $P(X \le 1, Y \le 3)$
(d) $P(X \le 1 \mid Y \le 3)$
(e) $P(Y \le 3 \mid X \le 1)$
(f) $P(X + Y \le 4)$

(a) $P(X \le 1) = P(X=0) + P(X=1)$

$$= \frac{8}{32} + \frac{20}{32} = \frac{28}{32} = \frac{7}{8}$$

(b) $P(Y \le 3) = P(Y=1) + P(Y=2) + P(Y=3)$

$$= \frac{3}{32} + \frac{3}{32} + \frac{11}{64} = \frac{6+6+11}{64} = \frac{23}{64}$$

(c) $P(X \le 1, Y \le 3) = P(0,1) + P(0,2) + P(0,3) + P(1,1) + P(1,2) + P(1,3)$

$$= 0 + 0 + \frac{1}{32} + \frac{1}{16} + \frac{1}{16} + \frac{1}{8}$$

$$= \frac{1+2+2+4}{32} = \frac{9}{32}$$

d) $P(X \le 1 \mid Y \le 3)$

$$P(X \le 1 \mid Y \le 3) = \frac{P(X \le 1, Y \le 3)}{P(Y \le 3)}$$

$$= \frac{9/32}{23/64}$$

$$= \frac{18}{23}$$

e) $P(Y \le 3 \mid X \le 1) = \frac{P(X \le 1, Y \le 3)}{P(X \le 1)}$

$$= \frac{9/32}{7/8}$$

$$= \frac{9}{28}$$

f) $P(X+Y \le 4) = P(0,1) + P(0,2) + P(0,3) + P(0,4) + P(1,1) + P(1,2) + P(1,3) + P(2,1) + P(2,2)$

$$= 0 + 0 + \frac{1}{32} + \frac{2}{32} + \frac{1}{6} + \frac{1}{6} + \frac{1}{8} + \frac{1}{32} + \frac{1}{32}$$

$$= \frac{5}{32} + \frac{2}{6} + \frac{1}{8}$$

$$= \frac{15+32+12}{96}$$

$$= \frac{69}{96} = \frac{23}{32}$$

2) The following table represent the joint probability distribution of the discrete random variable (X, Y)

	1	2	3	$P(y)$	
1	$1/12$	$1/6$	0	$1/4$	$P(Y=1)$
2	0	$1/9$	$1/5$	$14/45$	$P(Y=2)(y) = \sum_x f_{X,Y}$
3	$1/18$	$1/4$	$2/15$	$79/180$	$P(Y=3)\ f_X(x) = \sum_y f_{X,Y}$
$P(x)$	$\frac{5}{36}$	$\frac{19}{36}$	$\frac{1}{3}$	1	
	$P(X=1)$	$P(X=2)$	$P(X=3)$		

a) Evaluate the marginal distribution of X and Y

(b) Find the conditional distribution of X given $Y=2$

(c) Find the conditional distribution of Y given $X=3$

(b) $P(X=2 \mid Y=2) = \frac{P(X=2, Y=2)}{P(Y=2)}$

$$= \frac{P(2,2)}{P(Y=2)} = \frac{1/9}{14/45}$$

(c) Conditional distribution of y given $x=3$

$$P(Y=3 \mid X=3) = \frac{P(X=3, Y=3)}{P(X=3)} = \frac{P(3,3)}{P(X=3)}$$

$$= \frac{2/15}{1/3}$$

$$= \frac{2}{5}$$

3) The Joint pdf of two random variables X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{8}x(x-y) & 0<x<2 \ ; -x<y<x \\ 0 & \text{otherwise.} \end{cases}$$

Find $f(y/x)$.

From the definition of the conditional probability density function of Y

$$f_{XY}(y|x) = \frac{f(x,y)}{f_X(x)}$$ where $f_X(x)$ is the marginal density function of x

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

$$= \int_{-\infty}^{\infty} \frac{1}{8}x(x-y)\,dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{8}x(x-y)\,dy = \int_{-\infty}^{\infty} \frac{1}{8}(x^2 - xy)\,dy$$

$$= \frac{1}{8}\left[x^2 y - \frac{xy^2}{2}\right]_{-x}^{x}$$

$$= \frac{1}{8}\left(x^3 - \frac{x^3}{2} + x^3 - \frac{x^3}{2}\right) = \frac{x^3}{8}\left(1+1-\frac{1}{2}-\frac{1}{2}\right)$$

$$= \frac{x^3}{8}, \quad 0<x<2$$

$$f_{XY}(y|x) = \frac{\frac{1}{8}x(x-y)}{x^3/8} \quad 0<x<2 \ ; \ -x<y<x$$

$$f(y|x) = \begin{cases} \frac{x-y}{x^3} & -x<y<x \\ & \text{otherwise} \end{cases}$$

4) The joint pdf of the random variable (X, Y) is given by

$f(x,y) = Kxy\,e^{-(x^2+y^2)}$; $x>0$; $y>0$

Find the value of K and prove that X and Y are Independent

Here the range space is the entire first quadrant of the xy plane. By the property of the joint pdf

$$\int_{x>0}\int_{y>0} kxy e^{-(x^2+y^2)} dx dy = 1$$

$$K \int_0^{\infty}\int_0^{\infty} y e^{-y^2} . x e^{-x^2} dx dy = 1$$

put $x^2 = t$ $\quad 2x dx = dt \Rightarrow x dx = \frac{dt}{2}$

$$\frac{K}{2}\int_0^{\infty} y e^{-y^2} \int_0^{\infty} e^{-t} dt dy = 1$$

$$\frac{K}{2}\int_0^{\infty} y e^{-y^2} \left(\frac{e^{-t}}{-1}\right)_0^{\infty} dy = 1$$

put $y^2 = v$ $\quad 2y dy = dv$

$$\frac{K}{2} \cdot \frac{1}{2} \int_0^{\infty} e^{-v} dv = 1$$

$$\frac{K}{4}\left(\frac{e^{-v}}{-1}\right)_0^{\infty} = 1$$

$$\frac{K}{4}(1) = 1$$

$$K = 4$$

Similarly

$$f_Y(y) = \int_0^{\infty} f(x,y) dx$$

$$= 4y \int_0^{\infty} y e^{-(x^2+y^2)} dx$$

$$= 4y \int_0^{\infty} y e^{-x^2} . e^{-y^2} dx$$

$$= 4y^2 e^{-y^2} \int_0^{\infty} x e^{-x^2} dx$$

$x^2 = t$

$2x dx = dt$

$$= 4y^2 e^{-y^2} \int_0^{\infty} e^{-t} \frac{dt}{2}$$

$$= 2y^2 e^{-y^2} \left[\frac{e^{-t}}{-1}\right]_0^{\infty}$$

$$= 2y e^{-y^2} \quad y > 0$$

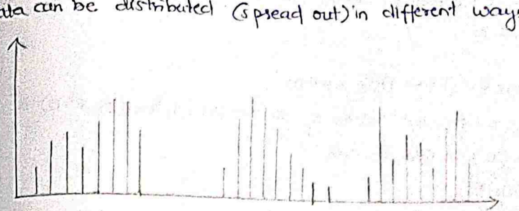
Now $f_X(x) . f_Y(y) = 2x e^{-x^2} . 2y e^{-y^2} = 4xy e^{-(x^2+y^2)}$

$= f_{XY}(x,y)$

Hence X & Y are independent.

Normal Distribution: (Gaussian)

Data can be distributed (spread out) in different ways.



(a) (b) (c)

(a) Spread out more on right

(b) Spread out more on left

(c) Jumbled up.

(d)

(d) Data tends to be around a central value with no bias left or right and it gets close to a "normal distribution."

Ex: That follows normal distribution are

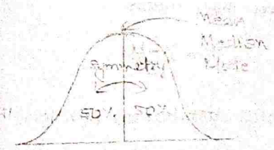
i) Heights of people

ii) errors in measurements

iii) Blood pressures

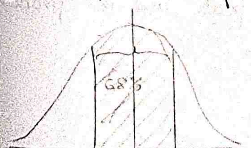
iv) Marks in a test.

50% of the values are < mean and 50% of -11- > mean



Standard Deviation:

S.D is a measure of how spread out numbers are



68% of values are within 1SD of the mean

or

98% - 2SD

99.7% - 3SD

95% of students at school are b/w 1.1m and 1.7m tall. Assume the data is normally distributed, calculate the data is normally mean & S.D

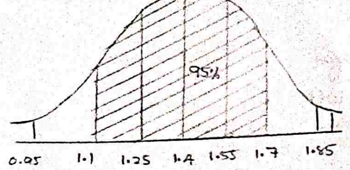
Soln:

Mean = $(1.1+1.7)/2 = 1.4m$

95% of 2 SD either side of the mean (a total of 4SD) so

1 SD = $(1.7-1.1)/4 = 0.15m$

As a result:



We say that ~~sa~~ any value is

* likely to be within 1SD (68 out of 100 should be)
* very likely -"- 2SD (95/100)
* almost certainly -"- 3SD (997/100)

The no. of S.D from the mean is also called the "Standard score" "Sigma" or "Z score"

Ex: In that same school one of your friend is 1.85m tall

1.85m is 3SD from the mean 1.4

So your friend height has a z score of 3.0

or

$1.85 - 1.4 = 0.45$ m from the mean

$1SD = 0.15$

$0.45/0.15 = 3.0$ SD

$$Z = \frac{X - \mu}{\sigma}$$

The N.D is a probability distribution function that describes how the values of a variable are distributed. It is a symmetric dist. where most of the observations cluster around the central peak and the probabilities for values further away.