—/—/—

Statistical Analysis in Physics

☆ Randomness and Random Variable:-

→ When you take any data, you take out central tendencies → mean, mode, median.

→ When you perform any experiment, for eg. finding out the value of acceleraⁿ due to gravity, $g = 9.8\ m/s^2$ or $10\ m/s^2$

or

Time period of a simple pendulum.

L.C = 0.01s
(stopwatch)

$T = 2\pi\sqrt{\frac{l}{g}}$

4 people are taking observatⁿ

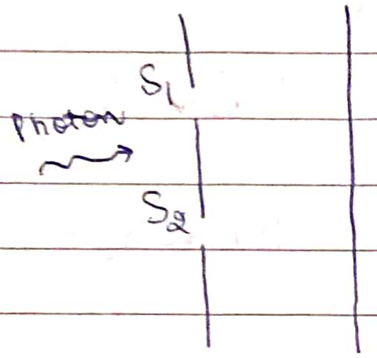
	t_{20} (20 observatⁿs)
S_1	24.13
S_2	24.22
S_3	23.98
S_4	?

→ we cannot tell exactly

This phenomenon is called randomness.

☆ **Random experiment:-** Random experiment is one in which you cannot predict exactly what your next observation would be even if you have previous data (but we know possible outcomes).

E.g. ② Double-slit experiment with photons



we can't predict from which slit photon would go through?

③ Radioactivity → Spontaneous random experiment

④ Tossing a coin

⑤ Taking a card out of deck of cards.

→ Non Random Experiment :-

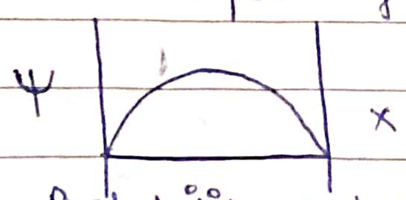
Digital Electronics Experiments

S	R	Q
0	1	0
1	0	1

Deterministic

⑥ Quantum Mechanics :-

Particle in a box → undeterministic (pos^n of particle)



Probability → $|\psi|^2$

★ Randomness is the fundamental property in Quantum mechanics.

(Just a function which allots a real no./value which is associated with each sample point)

★ **Random Variable:-** (X) $\overset{\text{(Real no.)}}{(x} \rightarrow f(x))$ (probability distribut^n)

$$X : S \rightarrow R$$

S ↓ Sample space; R (Real no.)

Eg:- Expt: Tossing 2 coins

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

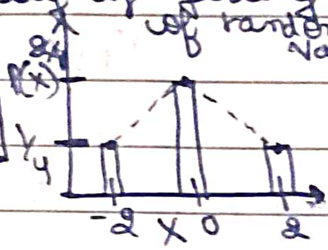
$X \Rightarrow$ no. of heads − no. of tails

$X = \{2, -2, 0, 0\}$

Find probability distribut^n (probability of getting each of random variable)

X	$P(X)$	X^2	$X^2 P(X)$
2	1/4	4	1
0	2/4	0	0
-2	1/4	4	1
	4		2

$\left[\sum P(X) = 1,\ 1 \geq P(X) \geq 0\right]$



$$E(X) = \sum_{i=1}^{3} X\,P(X) \quad \text{(discrete distribut}^n\text{)}$$

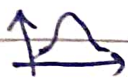
$$\mu = E(X) = 2 \times \tfrac{1}{4} + 0 \times \tfrac{2}{4} - 2 \times \tfrac{1}{4} = \boxed{0}$$

$$\text{Variance} = E(X^2) - \mu^2$$
$$= 2 - 0 = \boxed{2}$$

⇒ Mathematically,

Sample space (set of all possible outcomes) $\xrightarrow{X = \text{func}^n}$ Number — Real space

★ **Types of Probability distributions:-**



- Discrete
  - Binomial
  - Poisson

  $f(x_k) \geq 0$

  $\sum f(x_k) = 1$

- Continuous
  - Beta
  - Gamma
  - Normal

  $f(x_i) \geq 0$

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

★ Expectation Values –

$$\underset{(\text{mean})}{\mu} = E(X) = \sum_{x} x_x f(x_x) \quad (\text{Discrete})$$

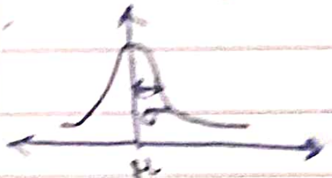
$$E(X^2) = \sum_{x} x_x^2 f(x_x)$$

$$E(x) = \int x f(x)\,dx \quad (\text{continuous})$$

$$E(x^2) = \int x^2 f(x)\,dx$$

★ Normal Distribution :– (Gaussian Distribution)

$$\left[N(x;\ \underset{\text{mean}}{\mu},\ \underset{\text{Variance}}{\sigma^2}) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \right]$$



$$X \sim N(X;\ 1.5,\ 2^2)$$

$$P(X \geq 2) \rightarrow \int_2^{\infty} f(x)\,dx$$

[95% credible interval = $\mu - 1.96\sigma,\ \mu + 1.96\sigma$]

→ Standard Normal variate, $Z = \frac{X-\mu}{\sigma}$

Q→ Find mean and variance of Beta and Gamma functions respectively.

→ $$\text{Beta}(\alpha, \beta) = \frac{\Gamma_{\alpha+\beta}}{\Gamma_\alpha \Gamma_\beta} x^{\alpha-1}(1-x)^{\beta-1} \quad (x \text{ range} \rightarrow 0 \text{ to } 1)\ [0,1]$$

$$\text{Gamma}(x;\ r, \nu) = \frac{\nu^r}{\Gamma_r} x^{\alpha-1} e^{-\nu x} \quad (x \rightarrow [0,\infty))$$

For Beta →

$$\mu = E(x) = \int_0^1 x f(x)\,dx$$

we know $\int_0^1 f(x)\,dx = 1$

$$= \int_0^1 x \cdot \frac{\Gamma_{\alpha+\beta}}{\Gamma_\alpha \Gamma_\beta} x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$= \frac{\Gamma_{\alpha+\beta}}{\Gamma_\alpha \Gamma_\beta} \int_0^1 x^{\alpha}(1-x)^{\beta-1}\,dx$$

$$= \frac{\Gamma\alpha+\beta}{\Gamma\alpha\,\Gamma\beta}\int_0^1 x^{(\alpha+1)-1}(1-x)^{\beta-1} \times \left(\frac{\Gamma\alpha+\beta+1}{\Gamma\alpha+\beta+1}\right) \times \left(\frac{\Gamma\alpha+1}{\Gamma\alpha+1}\right) dx$$

$$= \frac{\Gamma\alpha+\beta+1}{\Gamma\alpha} \times \frac{\Gamma\alpha+\beta}{\Gamma\beta \times \Gamma\alpha+\beta+1} \times \frac{\Gamma\alpha+1}{\Gamma\alpha+1} \times \int_0^1 x^{(\alpha+1)-1}(1-x)^{\beta-1}\,dx$$

$$\int_0^1 f(x)\,dx = 1$$

$$= \cancel{(\alpha+\beta+1)}\;\cancel{\Gamma\alpha+\beta}\;\cancel{(\alpha+1)}\;\cancel{\Gamma\alpha}$$

$$= \frac{\Gamma\alpha+\beta \times \Gamma\alpha+1}{\Gamma\alpha \times \Gamma\alpha+\beta+1} = \frac{\cancel{\Gamma\alpha+\beta} \times \alpha\,\cancel{\Gamma\alpha}}{\cancel{\Gamma\alpha} \times (\alpha+\beta)\,\cancel{\Gamma\alpha+\beta}}$$

Mean (Beta) $= \boxed{\dfrac{\alpha}{\alpha+\beta}}$

Similarly, $E(X^2) = \int_0^1 x^2 f(x)\,dx$

$$= \int_0^1 x^2 \frac{\Gamma\alpha+\beta}{\Gamma\alpha\,\Gamma\beta}\, x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$= \int_0^1 \frac{\Gamma\alpha+\beta}{\Gamma\alpha\,\Gamma\beta} \times \frac{\Gamma\alpha+\beta+2}{\Gamma\alpha+\beta+2} \times \frac{\Gamma\alpha+2}{\Gamma\alpha+2} \times x^{(\alpha+2)-1}(1-x)^{\beta-1}\,dx \to 1$$

$$= \frac{\Gamma\alpha+\beta\;\;\Gamma\alpha+2}{\Gamma\alpha\;\;\Gamma\alpha+\beta+2} = \frac{\cancel{\Gamma\alpha+\beta}\;\alpha(\alpha+1)\,\cancel{\Gamma\alpha}}{\cancel{\Gamma\alpha}\,(\alpha+\beta+1)(\alpha+\beta)\,\cancel{\Gamma\alpha+\beta}}$$

$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

Variance $= E(X^2) - (E(X))^2$

$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2}$$

$$= \frac{(\alpha^2+\alpha)(\alpha+\beta+1) - \alpha^2(\alpha+\beta)}{(\alpha+\beta)}$$

$$= \frac{\cancel{\alpha^3} + \cancel{\alpha^2\beta} + \alpha^2 + \alpha^2 + \alpha\beta + \alpha - \cancel{\alpha^3} - \cancel{\alpha^2\beta}}{(\alpha+\beta)}$$

$$\text{Variance}_{(\text{Beta})} = \frac{\alpha(2\alpha+\beta+1)}{(\alpha+\beta)} \quad \text{or} \quad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

⇒ for Gamma :-

$$E(x) = \int_0^\infty x\, f(x)\, dx$$

$$= \int_0^\infty x\left(\frac{v^r}{\Gamma_r} x^{r-1} e^{-xv}\right) dx$$

$$= \frac{v^r}{\Gamma_r} \int_0^\infty x^{r-1+1} e^{-xv}\, dx$$

$(r+1) = r'$

$$= \frac{v^r}{\Gamma_r} \int_0^\infty \frac{\Gamma_{r'}}{\Gamma_{r'}} x^{r'-1} e^{-xv} \times \frac{\cancel{e^{-v}}\, v^{r'}}{\cancel{e^{-v}}\, v^{r+1}} dx \to 1$$

$$= \frac{v^r}{\Gamma_r} \times \frac{\Gamma_{r+1}}{v^{r+1}} \times \frac{\cancel{e^{-v}}}{\cancel{e^{-(r+1)v}}}$$

$$= \frac{v^r}{v^{r+1}} \times \frac{\Gamma_{r+1}}{\Gamma_r} \times \frac{\cancel{e^{-v}}}{\cancel{e^{-(r+1)v}}}$$

$$= \frac{\cancel{v^r}}{\cancel{v^r}\cdot v} \times \frac{r\,\cancel{\Gamma_r}}{\cancel{\Gamma_r}} \times \frac{e^{-rv}}{\cancel{e^{-rv}}}$$

$$\mu = \text{Mean}_{(\text{Gamma})} = \frac{r}{v}$$

Similarly, $\text{Variance}_{(\text{Gamma})} = E(X^2) - \mu^2$

$$E(x^2) = \int_0^\infty x^2 \frac{v^r}{\Gamma_r} x^{r-1} e^{-vx}\, dx$$

→ 1 (since $\int_0^\infty f(x)\, dx = 1$)

$$= \frac{v^r}{\Gamma_r} \int_0^\infty x^{(r+2)-1} e^{-vx} \frac{v^{r+2}}{v^{r+2}} \frac{\Gamma_{r+2}}{\Gamma_{r+2}}\, dx$$

$$= \frac{\cancel{v^r}}{\cancel{\Gamma_r}} \times \frac{(r+1)\, r\, \cancel{\Gamma_r}}{\cancel{v^r}\cdot v^2} = \frac{r(r+1)}{v^2}$$

__/__/__

$$\text{Variance}(\sigma^2) = \frac{r(r+1)}{\gamma^2} - \frac{r^2}{\gamma^2}$$
(gamma)

$$= \frac{\cancel{r^2} + r - \cancel{r^2}}{\gamma^2} = \boxed{\frac{r}{\gamma^2}}$$

Numericals:-

(i) Conditional Probability

1) A dice is rolled. Find the probability of getting even given that roll always gives more than 2.

=> $S = \{1, 2, 3, 4, 5, 6\}$ B → (Sample Space reduce) = $\{3, 4, 5, 6\}$

A: Getting even $A \cap B \Rightarrow \{4, 6\}$

B: more than 2

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{2/6}{4/6} = \frac{1}{2}$$

2) A bag contains 5 white and 4 red balls. Find probability of drawing a red ball in second draw given that white ball is drawn in first draw.

=> | 05W |
| 04R |

$P(WR/W\_) = ?$

$$\frac{P(W \cap R)}{P(WW \text{ or } WR)} = \frac{5/9 \times 4/8}{5/9 \times 4/8 + 5/9 \times 4/8}$$

$$= \frac{1}{2}$$

3) | 10 White |
| 015 Blue |

$P(WB)$ (without replacement) = ?

$$P(WB) = \frac{\cancel{10}^{2}}{\cancel{25}_{5}} \times \frac{\cancel{15}^{1}}{\cancel{24}} = \frac{1}{4}$$

4) A couple has two children. Find probability that

Both are boys if it is known that one child is a boy.

$\Rightarrow$ $S = \{BB, GG, GB, BG\}$

A: Both are boys : $\{BB\}$

B: One child is a boy $\rightarrow$ $\{BB, BG, GB\}$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

→ Total Probability

5:)

$P(B_1) = 1/2$ → Bag 1: 4 Red + 3 Blue

$P(B_2) = 1/2$ → Bag 2: 2 Red + 4 Blue

One ball from one bag (choosing one bag and then one ball)

$$P(Red) = P(B_1)\,P(R/B_1) + P(B_2) * P(R/B_2)$$

$$= \frac{1}{2} \times \frac{4}{7} + \frac{1}{2} \times \frac{2}{6}$$

$$= \frac{2}{7} + \frac{1}{6} = \frac{12+7}{42} = \boxed{\frac{19}{42}}$$

6:)

Machine A: Produc^n → 25%, % of defective bolts produced → 5%

Machine B: Produc^n → 35%, % of defective bolts produced → 4%

Machine C: Produc^n → 40%, % of defective bolts produced → 2%

$\therefore$ $P(mA) = \frac{25}{100} = 1/4$ $\quad P(D/mA) = \frac{5}{100} = 1/20$

$P(mB) = \frac{35}{100} = \frac{7}{20}$ $\quad P(D/mB) = \frac{4}{100} = \frac{1}{25}$

$P(mC) = \frac{40}{100} = 2/5$ $\quad P(D/mC) = \frac{2}{100} = \frac{1}{50}$

$$P(D) = \frac{1}{4} \times \frac{1}{20} + \frac{7}{20} \times \frac{1}{25} + \frac{2}{5} \times \frac{1}{50} = \boxed{0.0345}$$

—/—/—

⇒ **Baye's Rule** :-

7) In Q6, if bolt is found to be defective, find the probability that it is manufactured from machine A?

⇒ A ⇒ Machine A manufactured

B ⇒ Defective

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)} = \frac{1/20 \times 1/4}{0.0345} = 0.362$$

8) Lab test → 99% effective (disease detecⁿ)

$P(+/D) = 0.99$

$P(D) = 0.005$

(i) $P(+) = ?$

(ii) $P(D/+) = ?$

⇒ (i) $P(+) = 0.005 \times 0.99 + 0.01 \times 0.995$

$= \boxed{0.0149}$

(ii) $$P(D/+) = \frac{P(+/D)\,P(D)}{P(+)} = \frac{0.99 \times 0.005}{0.0149}$$

$$= \boxed{0.3322}$$

☆ Continuous Probability Distributⁿ Funcⁿ → PDF (Probability Density Funcⁿ)

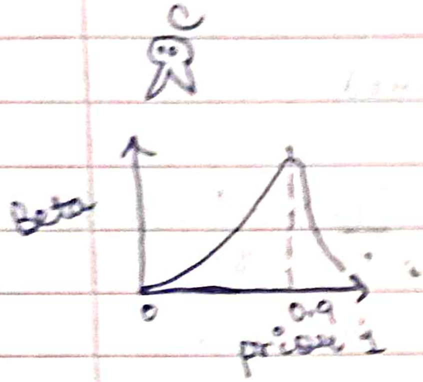
Discrete → PMF (Probability Mass Funcⁿ)

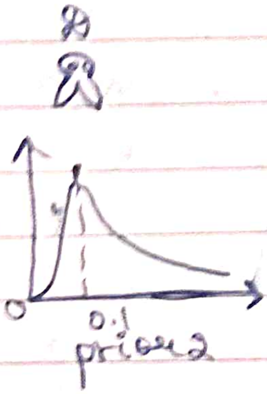
☆ Bayesian Statistics :-

[Posterior Distribution ∝ Prior (Personal Bias Distributⁿ) × Likelihood Real data (For eg. survey)]

For eg. there are 2 political parties A and B. Person C believes strongly that A would win because she talked to people of this region and also prior

admire A strongly (personal bias). Person D believe A would not win because he doesn't like it (personal bias).
prior





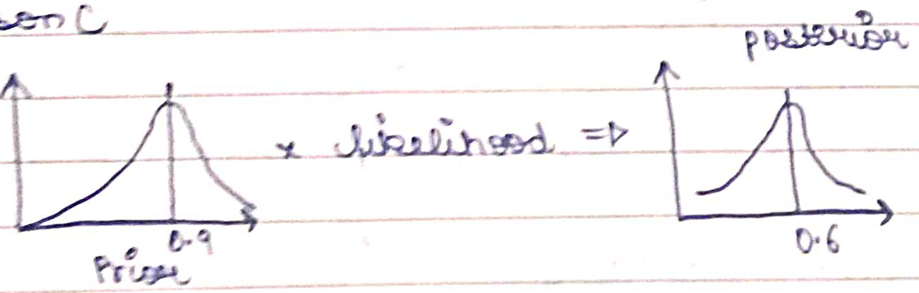
we took a survey of 10 people regarding it in Yes or No.
Out of 10 people, 3 people said yes to A winning (likelihood). ($n=10$ (Binomial) $y=3$)

$$\left[Bin(n, \theta) = {}^{n}C_{y}\, \theta^{y} (1-\theta)^{n-y} \right]$$

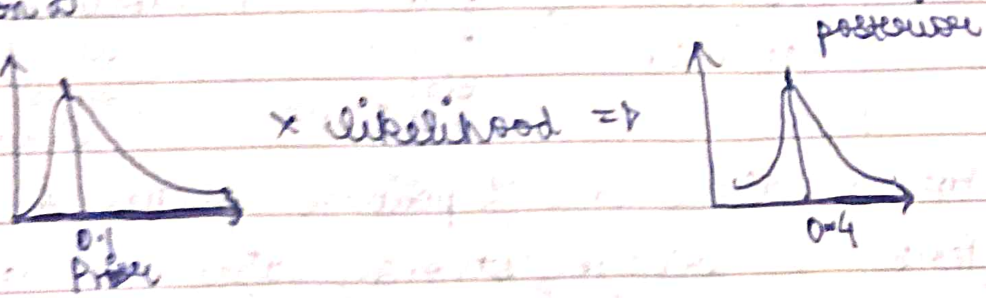
posterior $\propto$ prior $\times$ likelihood

$$\propto Beta(\ \ ) \times {}^{10}C_{3}\, \theta^{3} (1-\theta)^{10-3}$$

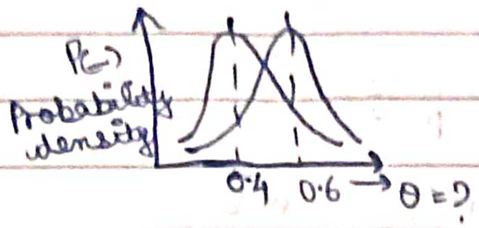
→ Person C



→ Person D



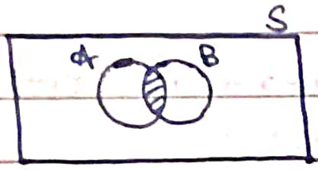
Joining these two graphs



Now,	Prior	Likelihood
	Beta (α, β)	~~Beta~~ Binomial (n, θ)
	Gamma (r, v)	~~Gamma~~ Poisson (λ)
	Normal (μ, σ^2)	Normal (μ, σ^2)

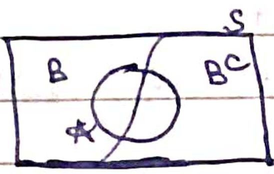
☆ Different Formulas :-

1.)



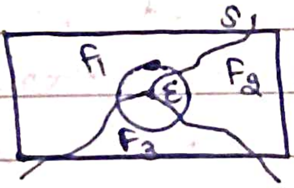
$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$ (conditional Probability)

2.)



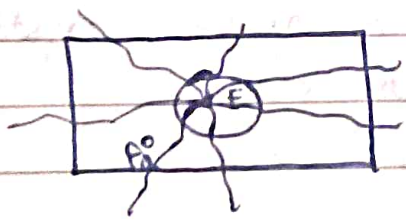
$$P(A) = P(A \cap B) + P(A \cap B^c)$$
(Total Probability)

3.)



$$P(E) = P(E \cap F_1) + P(E \cap F_2) + P(E \cap F_3)$$
$$P(E) = P(E/F_1)P(F_1) + P(E/F_2)P(F_2) + P(E/F_3)P(F_3)$$
(Total probability)

4.)



$$P(E) = \sum_{i=1}^{n} P(E/F_i)\, P(F_i)$$

$$P(A/D) = \frac{P(A \cap D)}{P(D)} = \cancel{P(A/D)}\ \frac{P(D/A)\,P(A)}{P(D)}$$
(Bayes' Theorem)

_/_/__

$$\Rightarrow \boxed{P(A/B)\,P(B) = P(B/A)\,P(A)}$$

★ Frequentist

$$P(F_i/E) = \frac{P(F_i \cap E)}{P(E)}$$

$$= \frac{P(E/F_i)\,P(F_i)}{\sum_{i=1}^{n} P(E/F_i)\,P(F_i)}$$

★ Bayesian Approach ⇒

$$\underbrace{g(\theta/y)}_{\text{posterior}} = \frac{\overset{\text{likelihood}}{f(y/\theta)} \times \overset{\text{prior}}{g(\theta)}}{\int f(y/\theta)\,g(\theta)\,d\theta}$$

(Probability density) $g(\theta/y)$: $\theta \to F_i$, $y \to E$

→ Generally, θ can be parameter (Probability/ mean/ variance/ regress^n coefficients)

→ Let's take θ as probability.

→ Eg. Suppose θ is probability of getting heads on tossing coin.

A tosses coin 5 times and he gets heads all the time. (so, for A, θ = 1) (Frequentist Approach)

But, if B says one more time and 6th time A gets tail, so θ becomes 5/6. (Frequentist)

Frequentist Approach ⇒ For a given system, θ is fixed but unknown to me.

Coin

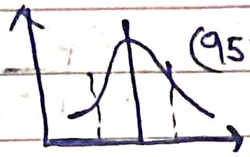
θ = constt. (unknown)

5 times	6 times	- - - - -	10000
θ = 1	θ = 5/6		$\theta = \frac{6520}{1000}$

[More no. of trials means more closer value to the true value of θ]

__/__/____

=> **Bayesian Approach:-** θ = not const,
θ = unknown
↓
(Random variable)
=> Probability density



(95% Credible)

$\theta \rightarrow \theta$ mean

$$\mu = E(\theta) = \int \theta \, g(\ ) \, d\theta$$

Q→ θ = ??
given, prior = Beta (3,2)
Likelihood $(n=10, y=6) \rightarrow$ Binomial (n,θ)

$$g(\theta/y) = \frac{Beta(3,2) \times Bin(10,\theta)}{\int_0 Beta(3,2)\ Bin(10,\theta)\, d\theta}$$

$$Beta(3,2) = \frac{\Gamma_5}{\Gamma_3 \Gamma_2} \theta^{3-1}(1-\theta)^{2-1}$$

$$Bin(10,\theta) = {}^{10}C_6\ \theta^6 (1-\theta)^{10-6}$$

$$\therefore g(\theta/y) = \frac{\cancel{\frac{\Gamma_5}{\Gamma_3 \Gamma_2}} \times \theta^{3-1}(1-\theta)^{2-1} \times \cancel{{}^{10}C_6} \times \theta^6 \times (1-\theta)^{10-6}}{\int_0^1 \cancel{\frac{\Gamma_5}{\Gamma_3 \Gamma_2}}\ \theta^{3-1}(1-\theta)^{2-1} \times \cancel{{}^{10}C_6} \times \theta^6 \times (1-\theta)^{10-6} d\theta}$$

$$= \frac{\theta^{9-1} \times (1-\theta)^{6-1} \times \frac{\Gamma_{9+6}}{\Gamma_{9+6}} \times \frac{\Gamma_9 \times \Gamma_6}{\Gamma_9 \ \Gamma_6}}{\int_0 \theta^{9-1} \times (1-\theta)^{6-1}}$$

$\boxed{1}$

$g(\theta/y) = Beta(9,6)$

__/__/__

$$\mu = \int \theta \, g(\theta/y) \, d\theta = \frac{\alpha'}{\alpha'+\beta'}$$

$$\theta_{mean} = \frac{9}{9+6} = \frac{9}{15} = \boxed{0.6}$$

$$\sigma^2 = \frac{\alpha'\beta'}{(\alpha'+\beta')^2(\alpha+\beta+1)}$$

Q→ In a study of water supply from a stream, $n = 176$ samples were taken from different sites. Out of these $y = 24$ had high bacteria level. Let θ be true probability that a sample of water from this stream has a high bacteria level.

(a) Use beta distributⁿ beta(1,10) as prior for θ, calculate posterior distributⁿ.

⇒ prior ⇒ Beta (1, 10)

$$= \boxed{\frac{\Gamma 11}{\Gamma 1 \, \Gamma 10} \times \theta^{1-1} (1-\theta)^{10-1}}$$

Likelihood ⇒ Bin (176, θ)

$$= \boxed{{}^{176}C_{24} \, \theta^{24} (1-\theta)^{152}}$$

Posterior = prior × likelihood

$$= \frac{\cancel{\frac{\Gamma 11}{\Gamma 1 \, \Gamma 10}} \times \theta^0 (1-\theta)^9 \times \cancel{{}^{176}C_{24}} \, \theta^{24}(1-\theta)^{152}}{\int_0^1 \cancel{\frac{\Gamma 11}{\Gamma 1 \, \Gamma 10}} \times \theta^0 \times (1-\theta)^9 \times \cancel{{}^{176}C_{24}} \, \theta^{24}(1-\theta)^{152}}$$

$$= \frac{\theta^{25-1} (1-\theta)^{162-1} \times \frac{\Gamma 25+162}{\Gamma 25+162} \times \frac{\Gamma 25}{\Gamma 25} \times \frac{\Gamma 162}{\Gamma 162}}{\int_0^1 \theta^{25-1} (1-\theta)^{162-1}}$$

$1 \left(\because \int_0^1 f(x)\,dx = 1 \text{ (Beta)}\right)$

posterior = Beta(25, 162)

(b) Find posterior mean and variance.

→ $\text{Mean} = \frac{\alpha}{\alpha+\beta} = \frac{25}{25+162} = 0.133$

$\text{Variance} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{25\times162}{(25+162)^2(25+162+1)}$

$= \boxed{0.000616}$

(c) Approximate Posterior distributⁿ to Normal distributⁿ using posterior μ & σ^2 as normal μ and σ^2 and find 95% credible interval.

⟹ $\mu = 0.133$

$\sigma^2 = 0.000616 \Rightarrow \sigma = 0.0248$

95% Credible interval (Normal) $\Rightarrow (\mu - 1.96\sigma, \mu + 1.96\sigma)$

$= (0.133 - 1.96\times0.0248, 0.133 + 1.96\times0.0248)$

$= (0.084392, 0.181608)$

★ Poisson Distribution :- No.' of count in given time interval ('λ')

or

limiting case of Binomial distributⁿ in which $n \to \infty$ and $\theta \to 0$ (probability). (A very rare occurrence). For eg. accidents happening on the road.

Frequentist ⇒ λ fixed (const.)
(unknown)

Bayesian ⇒ λ ⇒ Random Variable
(not fixed)
(unknown)

For eg. count of nuclei decaying per hour,

5, 2, 0, 23, 10

y_1, y_2, y_3, y_4, y_5 --- $f(y_1, y_2, y_3, \ldots, y_n / \lambda)$

$$\Rightarrow \frac{\lambda^{y_1} e^{-\lambda}}{y_1!}, \frac{\lambda^{y_2} e^{-\lambda}}{y_2!}, \ldots$$

$$\Downarrow$$

$$\boxed{\frac{\lambda^{\sum y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!}}$$

⇒ Prior ⇒ gamma

$$g(\lambda) = \text{gamma}(\lambda; r, v) = \frac{v^r}{\Gamma_r} \lambda^{r-1} e^{-v\lambda}$$

→ posterior

$$= \frac{\dfrac{\lambda^{\sum y_i} e^{-n\lambda}}{\cancel{\prod y_i!}} \times \cancel{\dfrac{v^r}{\Gamma_r}} \lambda^{r-1} e^{-v\lambda}}{\int_0^\infty \cancel{\dfrac{v^r}{\Gamma_r}} \lambda^{r-1} e^{-v\lambda} \dfrac{\lambda^{\sum y_i} e^{-n\lambda}}{\cancel{\prod_{i=1}^{n} y_i!}} \, d\lambda}$$

$$= \frac{\lambda^{\overbrace{(r+\sum y_i)}^{r'}-1} e^{-(v+n)\lambda} \times \dfrac{(v+n)^{r+\sum y_i}}{(v+n)^{r+\sum y_i}} \times \dfrac{\Gamma_{r+\sum y_i}}{\cancel{\Gamma_{r+\sum y_i}}}}{\int_0^\infty \lambda^{(r+\sum y_i)-1} e^{-(v+n)\lambda} \, d\lambda}$$

$$= \text{gamma}(\lambda; (r+\sum y_i), (v+n))$$

$\left[\text{Mode (beta)} = \frac{\alpha - 1}{\alpha + \beta - 2}\right]$

Gamma $\rightarrow$ Mean $= \frac{r}{\nu}$

Variance $= \frac{r}{\nu^2}$

Modal value $= \frac{r-1}{\nu}$

★ Binomial $\rightarrow$ mean $= np$ $\left\{ \begin{matrix} {}^nC_y\, p^y q^{n-y} \\ \text{or} \\ {}^nC_y\, \theta^y (1-\theta)^{n-y} \end{matrix} \right\}$

variance $= npq$

Poisson $\rightarrow$ mean $= \lambda$

variance $= \lambda$

Q $\rightarrow$ Find mean and variance of binomial and poisson.

$\Rightarrow$ Binomial $= {}^nC_x\, \theta^x (1-\theta)^{n-x} = f(x)$

For mean, $E(x) = \sum_{x=0}^{n} x f(x)$

$$= \sum_{x=0}^{n} x\, {}^nC_x\, \theta^x (1-\theta)^{n-x}$$

$$= 0 + \sum_{x=1}^{n} x \times \frac{n!\, \theta^x}{(n-x)!\, x!} (1-\theta)^{n-x}$$

$$= 0 + \sum_{x=1}^{n} \cancel{x} \times \frac{n!\, \theta^x}{(n-x)!\, \cancel{x}\,(x-1)!} \times (1-\theta)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n\,(n-1)!}{((n-1)-(x-1))!\,(x-1)!} \times \theta^x \times \frac{\theta}{\theta} \times \frac{(1-\theta)^{(n-1-(x-1))} \times (1-\theta)}{(1-\theta)}$$

$$= \boxed{n\theta} \quad \left(\text{since } \sum_{x=0}^{n} f(x) = 1 \right)$$

for variance,

Variance $= E(x^2) - [E(x)]^2 = [E(x^2) - \mu^2]$

$E(x^2) = E(x^2 - x + x)$

$= E(x(x-1) + x) = E(x(x-1)) + E(x)$

$= E(x(x-1)) + \mu$

$E(x^2) = E(x(x-1)) + \mu$

$$E(x(x-1)) = \sum_{x=0}^{n} x(x-1)\, {}^nC_x\, \theta^x (1-\theta)^{n-x}$$

$$= 0 + \sum_{x=1}^{n} \cancel{x(x-1)} \frac{n!}{(n-x)!\, \cancel{x(x-1)}(x-2)!} \theta^x (1-\theta)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n(n-1)(n-2)!}{((n-2)-(x-2))!\,(x-2)!} \times \theta^x \times \frac{\theta^2}{\theta^2} \times (1-\theta)^{(n-2)-(x-2)}$$

$$= n(n-1)\theta^2 = \boxed{n^2\theta^2 - n\theta^2}$$

$$\text{Variance} = \cancel{n^2\theta^2} - n\theta^2 - \underset{(\mu^2)}{\cancel{n^2\theta^2}} + \mu$$

$$= n\theta - n\theta^2$$

$$= \boxed{n\theta(1-\theta)} = \boxed{npq}$$

★ **Poisson** :- $f(x) = \dfrac{\lambda^x \cdot e^{-\lambda}}{x!}$

$$\mu = E(x) = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = 0 + \sum_{x=0}^{\infty} \cancel{x} \frac{\lambda^x e^{-\lambda}}{\cancel{x}(x-1)!}$$

$$= \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!} \times \frac{\lambda}{\lambda}$$

$$= \boxed{\lambda}$$

__/__/__

$Var = E(x^2) - \mu^2$

$E(x^2) = E(x(x-1)) + \mu$

$$E(x(x-1)) = \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x e^{-\lambda}}{x(x-1)(x-2)!} \times \frac{\lambda^2}{\lambda^2}$$

$= \lambda^2$

$E(x^2) = \lambda^2 + \lambda(\mu)$

$Var = \cancel{\lambda^2} + \boxed{\lambda} - \cancel{\lambda^2}$

likelihood	parameter	Prior	posterior
Binomial	$Bin(n, \theta)$	$Beta(\alpha, \beta)$	$Beta(\alpha', \beta')$
poisson	$Pois(\lambda)$	$gamma(r, v)$	$Gamma(r', v')$
normal	$N(\mu, \sigma^2)$	$normal(\mu, \sigma^2)$	$Normal(\mu', \sigma'^2)$

★ Normal → likelihood → $N(m, s^2)$

Normal → Prior → $N(\mu, \sigma^2)$ → given
??

→ likelihood → $y_1, y_2, y_3, \ldots y_n$ $(\sim N(\mu, \sigma^2))$

$\frac{1}{\sqrt{2\pi}\sigma} e^{-(y_1-\mu)^2/2\sigma^2}$ (similarly for all data points)

likelihood $= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_{i=1}^{n}(y_i-\mu)^2/2\sigma^2}$

→ Prior, $g(\mu) \sim N(m, s^2) \rightarrow \frac{1}{\sqrt{2\pi}\sigma} e^{-(\mu-m)^2/2s^2}$

what to find (whatever is written here is variable)

Posterior →
$$\frac{\frac{1}{\sqrt{2\pi}s} e^{-(\mu-m)^2/2s^2} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2\sigma^2}}}{\int_{-\infty}^{\infty} \_\_\_\_ \, d\mu}$$

$$= e^{-\frac{n}{2\sigma^2}[\mu-\bar{y}]^2} \cdot e^{-(\mu-m)^2/2s^2} \quad \left(\bar{y} = \frac{\sum y_i}{n}\right)$$

$$\int \_\_\_ \, d\mu$$

$$= e^{-\left[\frac{n}{2\sigma^2}[\mu-\bar{y}]^2 + 1/2s^2[\mu-m]^2\right]}$$

$$= e^{-\left[n/2\sigma^2(\mu^2+\bar{y}^2-2\mu\bar{y}) + 1/2s^2[\mu^2+m^2-2\mu m]\right]}$$

$$= e^{-\left[\left(\frac{n}{2\sigma^2}+\frac{1}{2s^2}\right)\mu^2 + \left(n/2\sigma^2\,\bar{y}^2 + \frac{m^2}{2s^2}\right) - 2\mu\left(\frac{n}{2\sigma^2}\bar{y} + \frac{m}{2s^2}\right)\right]}$$

const.

$$= e^{-\left[\frac{n}{2\sigma^2}+\frac{1}{2s^2}\right]\left[\mu^2 - 2\mu\left(\frac{\frac{n}{2\sigma^2}\bar{y} + m/2s^2}{n/2\sigma^2 + \frac{1}{2s^2}}\right)\right]}$$

$$= e^{-\left[\frac{n}{2\sigma^2}+\frac{1}{2s^2}\right]\left(\mu^2 - 2\mu\left(\frac{\frac{n}{\sigma^2}\bar{y} + m/s^2}{n/\sigma^2 + 1/s^2}\right) + \left(\frac{m/s^2 + \frac{n\bar{y}}{\sigma^2}}{1/s^2+\frac{n}{\sigma^2}}\right)^2 - \left(\frac{m/s^2 + \frac{n\bar{y}}{\sigma^2}}{1/s^2+\frac{n}{\sigma^2}}\right)^2\right)}$$

const

$$= e^{-1/2\left(\frac{n}{\sigma^2}+1/s^2\right)\left(\mu - \left(\frac{m/s^2 + \frac{n\bar{y}}{\sigma^2}}{1/s^2 + n/\sigma^2}\right)\right)^2} \times C$$

$$\text{Prior} = e^{-1/2\left(n/\sigma^2 + 1/s^2\right)^{-1}\left(\mu - \left(\frac{m/s^2 + \frac{n\bar{y}}{\sigma^2}}{1/s^2 + n/\sigma^2}\right)\right)^2}$$

$$\sigma'^2 = \left(\frac{1}{s^2} + \frac{n}{\sigma^2}\right)^{-1}$$

$$\boxed{\frac{1}{\sigma'^2} = \frac{1}{s^2} + n/\sigma^2}$$

$$\mu' = \cancel{m/s^2}\; m\left(\frac{1/s^2}{1/s^2+\frac{n}{\sigma^2}}\right) + \left(\frac{\frac{n}{\sigma^2}}{1/s^2+\frac{n}{\sigma^2}}\right)\bar{y}$$

$y \to$ Data (given)

Likelihood $\sim N(\mu, \sigma^2)$ — μ_{like}, σ^2_{like}; σ^2 (given)

prior $\sim N\left(\mu^{Prior}_{(m)}, \sigma^2_{(s^2)}{}^{Prior}\right)$

$$\text{posterior} \sim N(\overset{\mu_{post}}{\mu'}, \overset{\sigma^2_{post}}{\sigma'^2})$$

$$\left[\frac{1}{\sigma^2_{Post}} = \frac{1}{\sigma^2_{Prior}} + \frac{n}{\sigma^2_{likelihood}}\right]$$

$$\left[\mu_{Post} = \frac{1/\sigma^2_{prior}}{1/\sigma^2_{Prior} + \frac{n}{\sigma^2_{like}}} \times \mu_{Prior} + \frac{\frac{n}{\sigma^2_{like}}}{1/\sigma^2_{Pr} + \frac{n}{\sigma^2_{like}}} \bar{y}\right]$$

Mean of data set

★ **Probability Mass func^n:-** Let X be a Discrete Random Variable such that $P(X=x)=P_i$, then P_i is said to be Probability Mass func^n if it satisfies following conditions:-

(i) $P_i \geq 0$
(ii) $\sum P_i = 1$

For pdf (probability density func^n), (i) $f(x) \geq 0$, $-\infty < x < \infty$
(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

★ **Cumulative Distribut^n Func^n:-** $F(x)$

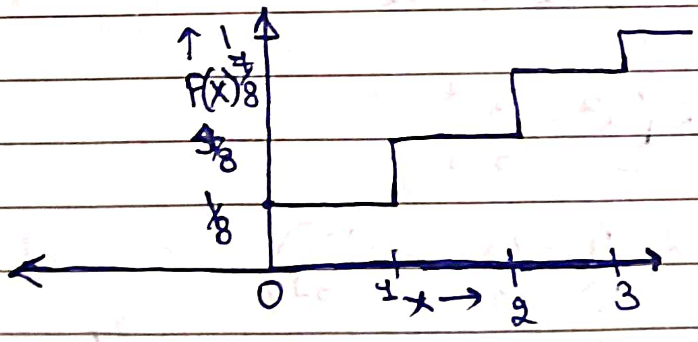
Let X be a Discrete Random Variable such that its cdf is defined as $f(x) = \sum P_i = P(x \leq x_i)$.

For eg.

$$F(x) = \begin{cases} 1/8 & x \leq 0 \\ 4/8 & x \leq 1 \\ 7/8 & x \leq 2 \\ 1 & x \leq 3 \end{cases}$$

For eg.

x (No. of heads)	0	1	2	3
P	1/8	3/8	3/8	1/8



(Plotting CDF)

Q→ If x is a continuous Random variable with the following pdf

$$f(x) = \begin{cases} \alpha(2x-x^2) & 0<x<2 \\ 0 & o/w \end{cases}$$

(i) find α

(ii) $P(x>1)$

⇒ (i) By def^n of pdf, $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx = 1$$

$$= \int_{0}^{2} \alpha(2x-x^2)\,dx = 1$$

$$= \alpha\left[\cancel{2}\frac{x^2}{\cancel{2}} - \frac{x^3}{3}\right]_0^2 = 1$$

$$= \alpha\left[4 - \frac{8}{3}\right] = 1$$

$$\alpha = 3/4 \quad —(1)$$

(ii) $P(x>1)$

$$= \int_{1}^{2} f(x)\,dx + \int_{2}^{\infty} \cancel{f(x)}^{\,0}\,dx$$

$$= \alpha\int_{1}^{2}(2x-x^2)\,dx$$

$$= \alpha\left[x^2 - \frac{x^3}{3}\right]_1^2$$

$$= \alpha\left(\left(4-\frac{8}{3}\right) - \left(1-\frac{1}{3}\right)\right)$$

$$= \alpha\left(4 - 8/3 - 1 + 1/3\right) = \alpha\left(3 - 7/3\right) = \alpha\left(\frac{2}{3}\right)$$

$= \frac{3}{4} \times \frac{2}{3} = \boxed{0.5}$ Ans

★ Cumulative Distribut^n Func^n (cdf) :- (Continuous R.V.)

Let X be continuous random variable having pdf $f(x)$ then $F_x(x)$ will be a continuous distribut^n func^n of x if

$$F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$

cdf

→ Rel^n b/w Cumulative Distribut^n func^n (cdf) and probability density func^n (pdf) :-

$$\boxed{\frac{d}{dx} F(x) = f(x)}$$

Q→ Probability ~~dist~~ density func^n of Random Variable X if

$$f(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 2-x, & 1 < x < 2 \\ 0, & o/w \end{cases}$$

① find $P(X \geq 1.5)$

② find cdf (cumulative distribut^n func^n).

⇒ ① $P(X \geq 1.5) = \int_{1.5}^{\infty} f(x)\,dx$

$$= \int_{1.5}^{2} f(x)dx + \int_{2}^{\infty} \cancel{f(x)}dx$$

$$= \int_{1.5}^{2} (2-x)dx$$

$$= 2\left[x - \frac{x^2}{2}\right]_{1.5}^{2} = 2\left((2-2) - \left(1.5 - \frac{2.25}{2}\right)\right)$$

$$= \boxed{0.125}$$

② cdf

$x \le 0$

$$P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

$0 < x < 1$

$$F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx$$

$$= 0 + \int_{0}^{x} x\,dx = \frac{x^2}{2}$$

$1 < x \le 2$

$$f(x) = \int_{-\infty}^{0} \cancel{f(x)}^{\,0}\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{x} f(x)\,dx$$

$$= \int_{0}^{1} x\,dx + \int_{1}^{x} (2-x)\,dx$$

$$= -1 + 2x - x^2/2$$

$x > 2$

$$f(x) = \int_{-\infty}^{0} \cancel{f(x)}^{\,0}\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{\infty} \cancel{f(x)\,dx}^{\,0}$$

$$F_X(x) = \begin{cases} 0, & -\infty < x < 0 \\ x^2/2, & x \le 1 \\ -1+2x-\frac{x^2}{2}, & x \le 2 \\ \end{cases}$$

Find $P(X \ge 1.5) \Rightarrow -1 + 2(1.5) - \frac{1.5^2}{2}$

$= 0.125$

$\star$ Variance $= E(x-\bar{x})^2$
$= E(x^2 - 2x\bar{x} + \bar{x}^2)$
$= E(x^2) - 2\bar{x}E(x) + \bar{x}^2$
$= E(x^2) - 2\bar{x}\bar{x} + \bar{x}^2$
$= \boxed{E(x^2) - \bar{x}^2}$

$\rightarrow \left[\int_0^\infty x^{n-1} e^{-ax} dx = \frac{\Gamma n}{a^n}\right]$

$\star$ Bivariate Random Variable:-
Let S be a sample space associated with random experiment. Let X(t), Y(t) be two random func^n's each assigning a real no. to each outcome $t \in S$.
Then (X, Y) is called a Bivariate or two dimensional Random Variable.

$\star$ Joint Probability Mass Func^n of (X, Y):-
If (X, Y) is a two-dimensional discrete Random Variable, where $X = x_i$ & $Y = y_j$, $i \& j = 1, 2, 3$ ...
Then $P(X = x_i, Y = y_j) = P_{ij}$ is called PMF of (X, Y) provided
① $P_{ij} \geq 0$
② $\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} = 1$
The set of triple (x_i, y_j, P_{ij}) is called Joint Probability of (X, Y)

$\Rightarrow$ Marginal Probability mass func^n of x:-
$P(X = x_i) = \sum_j P_{ij} = P_{i1} + P_{i2} + \ldots$
• collec^n of pair (x_i, P_{i*}) → Marginal PMF of x.
Similarly, $P(Y = y_j) = \sum_i P_{ij} = P_{1j} + P_{2j} \ldots$

Pair $(y_j, P_j) \rightarrow$ Marginal Probability of y

If X & Y are independent,
if $P_{XY}(x, y) = P_X(x) P_Y(y)$

Q $\rightarrow$ Two balls are selected at random from a box containing three red, two green and four white balls. If X and Y are no. of red balls & green balls respectively, find

① Joint Probability of X & Y

② Marginal Probabilities of X & Y

③ Conditional distribut^n of X given Y=1
~~(CDF)~~

$\Rightarrow$ ①

X $\rightarrow$

X	P(X)	Y	P(Y)
0	$^6C_2/^9C_2$	0	$^7C_2/^9C_2$
1	$\frac{^3C_1 \times ^6C_1}{^9C_2}$	1	$\frac{^2C_1 \times ^7C_1}{^9C_2}$
2	$^3C_2/^9C_2$	2	$^2C_2/^9C_2$
3			

$\rightarrow$ Don't do like this

$P(0,0) = {^4C_2}/{^9C_2}$

X \ Y	0	1	2	Total
0	1/6	2/9	1/36	15/36
1	1/3	1/6	0	18/36
2	1/12	0	0	3/36
Total	21/36	14/36	1/36	1

$\Rightarrow$ Joint Probability of X and Y (make this table)

② Marginal Prob. of x :-

x	$P_X(x)$
0	15/36
1	18/36
2	3/36

Marginal Prob. of y :-

y	$P_Y(y)$
0	21/36
1	14/36
2	1/36

$$P(X/Y=1) = \frac{P(X \cap Y_1)}{P(Y=1) \to \text{Total}}$$

③ Conditional distⁿ of X given Y=1

X	P(X/Y=1)
0	$\frac{2/9}{14/36} = 4/7$
1	6/14
2	0

Q → Joint PMF of (X,Y) is given by

$$P(x,y) = K(2x+3y)$$

$x = 0, 1, 2$, $Y = 1, 2, 3$.

find ① K ② Marginal Prob of x and Y

③ Conditional Prob of X given Y=1

④ " " " Y " X=2

⑤ Prob. distributⁿ of X+Y

=> ①

X \ Y	1	2	3	Total
0	3K	6K	9K	18K
1	5K	8K	11K	24K
2	7K	10K	13K	30K
Total	15K	24K	33K	72K

$\boxed{K = 1/72}$

②

X	$P_X(x)$
0	18/72
1	24/72
2	30/72

Y =>

Y	$P_Y(y)$
1	15/72
2	24/72
3	33/72

③, ④ → same as previous quesⁿ

⑤

Z(X+Y)	P(X+Y)
1	3/72
2	11/72
3	[illegible]/72
4	21/72
5	13/72

Q→ Are X and Y independent?

marginal Prob.

X	$P_X(x)$
0	18/72
1	29/72
2	30/72

Y	$P_Y(y)$
1	15/72
2	24/72
3	33/72

By def^n of independence,

$$P_{XY}(x,y) = P_X(x)P_Y(y)$$

LHS

$$P_{XY}(0,1) = P_X(0)P_Y(1)$$

$$P_{XY}(0,1) = 3/72 \text{ (from joint prob. table)}$$

$$P_X(0) = 18/72, \; P_Y(1) = 15/72$$

$$P_X(0) \times P_Y(1) = 5/96 \neq P_{XY}(0,1)$$

(Not independent)
(X & Y)

Q→ Two random variables X and Y have following pdf:-

$$f(x,y) = \begin{cases} 2-x-y, & 0\leq x\leq 1, 0\leq y\leq 1 \\ 0, & \text{otherwise} \end{cases}$$

→ find marginal dist^n of x & y.

⇒ For x,

$$f_X(x) = \int_0^1 (2-x-y)\,dy = 3/2 - x, \; 0<x<1$$

For y, $f_Y(y) = \int_0^1 (2-x-y)\,dx = 2 - 1/2 - y$

$$= 3/2 - y, \; 0<y<1$$

Q→ $f(x,y) = \begin{cases} 2e^{-2x}e^{-y}, & \text{if } x \geq 0, y \geq 0 \\ 0, & \text{o/w} \end{cases}$

→ find probability that $0 \leq x \leq 1,\ 1 \leq y \leq 2$.

Solⁿ → $\displaystyle\int_0^1\int_1^2 2e^{-2x}e^{-y}\,dx\,dy$

~~f(x,~~

$$= 2\int_0^1 e^{-2x}\left[\frac{e^{-y}}{-1}\right]_1^2 dx$$

$$= 2\left[\frac{e^{-2x}}{-2}\right]_0^1 \times \left[\frac{e^{-y}}{-1}\right]_1^2$$

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

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

★ **Properties:-**

- Joint CDF : $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$

  Joint PDF : $f_{X,Y}(x,y) = \dfrac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y)$

- Valid joint PDF

  $f_{X,Y}(x,y) \geq 0$

  $\displaystyle\iint_{(x,y)} f_{X,Y}(x,y)\,dx\,dy = 1$

- Marginal PDF

  $X, Y \Rightarrow f_{X,Y}(x,y)$

  - $f_X(x) = \displaystyle\int_y f_{X,Y}(x,y)\,dy$
  - $f_Y(y) = \displaystyle\int_x f_{X,Y}(x,y)\,dx$

— / — / —

- Conditional Probability

$$f(x/y) = \frac{\overbrace{f(x,y)}^{\text{joint probability of } x \text{ and } y}}{\underbrace{f(y)}_{\text{Marginal dist}^n \text{ of } y}}$$

- Independence of R.V.s

$$f(x,y) = f(x) \cdot f(y)$$

Joint Probability = product of marginal probabilities

or

$$f(x/y) = f(x)$$

$$f(y/x) = f(y)$$

★ Properties of Variance of Random Variable

→ If X is a random variable & a, b → constts, then $\boxed{var(aX+b) = a^2 var(X)} + \cancel{var(b)}^{0}$

→ If X_1 and X_2 are independent variables, then

(a) $var(X_1 + X_2) = var(X_1) + var(X_2)$

(b) $var(X_1 - X_2) = var(X_1) - var(X_2)$

★ Moment for discrete random variables :-

→ Moment about origin :- $\mu'_r = E[x-0]^r$

$\mu'_1 = E[x]$, $\mu'_2 = E[x^2]$, $\mu'_3 = E[x^3]$, $\mu'_4 = E[x^4]$

mean

→ Moment about any other pt. a :- $\mu''_r = E[x-a]^r$

$\mu''_1 = E[x-a]$ $\quad \mu''_3 = E[x-a]^3$

$\mu''_2 = E[x-a]^2$ $\quad \mu''_4 = E[x-a]^4$

__/__/__

→ Moment about mean :- $\mu_r = E[x-\bar{x}]^r$

$\mu_0 = 1$

$\mu_1 = E[x-\bar{x}] = E[x] - E[\bar{x}] = E[x] - \bar{x}$
$= \bar{x} - \bar{x} = 0$

↑ first moment about mean is zero.

$\mu_2 = E[x-\bar{x}]^2 = \sigma^2 =$ Variance

Note → Mean is first moment about origin and variance is second moment about mean.

→ $\mu = E[x^2 + \bar{x}^2 - 2x\bar{x}]$

$= E[x^2] + E(\bar{x}^2) - E(2x\bar{x})$

$= E(x^2) + \bar{x}^2 - 2\bar{x}^2$

$= E(x^2) - (E(x))^2$

→ $\mu_2 = \mu_2' - (\mu_1')^2$

→ $\mu_3 = E(x-\bar{x})^3 := E(x^3 - 3x^2\bar{x} + 3x\bar{x}^2 - \bar{x}^3)$

$= E(x^3) - 3\bar{x}E(x^2) + 3\bar{x}^2 E(x) - \bar{x}^3$

$= E(x^3) - 3\bar{x}E(x^2) + 2\bar{x}^3$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$

→ $\mu_4 = E(x-\bar{x})^4$

$= E(x^4 - {}^4C_1 x^3\bar{x} + {}^4C_2 x^2\bar{x}^2 - {}^4C_3 x\bar{x}^3 + {}^4C_4 \bar{x}^4)$

$= E(x^4) - 4\bar{x}E(x^3) + 6\bar{x}^2E(x^2) - 4\bar{x}^3E(x) + \bar{x}^4$

$= E(x^4) - 4\bar{x}E(x^3) + 6\bar{x}^2E(x^2) - 3\bar{x}^4$

$[\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3(\mu_1')^4]$

★ Moment for continuous R.V.

(i) Moment about a point 'a'

$$\mu_r' = \int_{-\infty}^{\infty} (x-a)^r f(x)\,dx$$

(ii) Moment about mean

$$\mu_r = \int_{-\infty}^{\infty} (x-\bar{x})^r f(x)\,dx$$

(iii) Moment about origin

$$\mu_r' = \int_{-\infty}^{\infty} (x-0)^r f(x)\,dx$$

Q→ Calculate first four moments about mean for following distributⁿ

X	6	7	8	9	10	11	12
Y	3	6	9	13	8	5	4

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{6\times3 + 7\times6 + 8\times9 + 9\times13 + 10\times8 + 11\times5 + 12\times4}{3+6+9+13+8+5+4}$$

$$= \boxed{9}$$

→ $\mu_1 = E(x-\bar{x}) = 0$

Imp.

$$\mu_2 = E(x-\bar{x})^2 = \frac{\sum (x-9)^2 \cdot f}{\sum f} = \frac{18}{48} = 0.37$$

(measurement of joint variance of X and Y)

★ **Covariance**:- If X and Y → random variables ($E(X^2) < \infty$ → finite second moments)

Then,

$$Cov(X,Y) = E[\{X-E(X)\}\{Y-E(Y)\}]$$
$$= E(XY) - E(X)E(Y)$$

If X and Y are independent, then

$$E(XY) = E(X)E(Y)$$

∴

$$Cov(X,Y) = E(X)E(Y) - E(X)E(Y) = \boxed{0}$$

NO linear relation b/w X and Y

(correlⁿ coefficient $= \dfrac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$)

___/___/___

★ Covariance Matrix = $\begin{bmatrix} Var(X_1) & - & - & Cov(x_n, x_1) \\ \cancel{Cov(x_n, x_1)} & Var(X_2) & & | \\ Covar(x_n, x_1) & - & - & Var(X_n) \end{bmatrix}$

★ Correlation matrix :- Table showing correlatⁿ coefficients b/w different sets of variables.
(Values of correlatⁿ coefficient range from -1 to 1)
(Corr(X,X)=1) · (Cov(X,Y) = Cov (Y,X))
For eg.

Correlatⁿ	Height	Weight	Age
Height	1	0.87	0.65
Weight	0.87 ~~0.63~~	1	0.62
Age	0.65	0.62	1

★ Properties of Covariance :-
Cov (X,Y) = Cov (Y,X)

★ Linearity:-
Cov (aX + b, Y) = a. Cov(X,Y)

★ Cov (X + Z, Y) = Cov (X,Y) + Cov (Z, Y)

★ Cov (X, X) = Var(X)

★ Cov (X, Y) = 0 (X & Y are independent)

★ Cov (aX, bY) = ab. Cov (X,Y)

★ Properties of correlatⁿ coefficient :- (r)

1.) $-1 \leq r \leq 1$

2.) $r(X,Y) = r(Y,X)$

3.) $r(aX+b, cY+d) = r(X,Y)$

__/__/__

★ Eigenvalue Decomposit^n of Covariance Matrix

For eg. $A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$

$\det|A - \lambda I| = 0 \rightarrow$ (Eigenvalue)

$(4-\lambda)(3-\lambda) - 2 = 0$

$12 - 4\lambda - 3\lambda + \lambda^2 - 2 = 0$

$\cancel{4\lambda^2 - 14\lambda + 10 = 0}$

$\cancel{2\lambda^2 - 7\lambda + 5 = 0}$

$\cancel{2\lambda^2 - 5\lambda - 2\lambda + 5 = 0}$

$\cancel{\lambda(2\lambda-5) - (2\lambda-5) = 0}$

$\cancel{\lambda = 1, 5/2}$

$\lambda^2 - 7\lambda + 10 = 0$

$\lambda_1 = 5$ & $\lambda = 2$ $\rightarrow$ Eigenvalues

→ For Eigen vector :

$\underbrace{AX = \lambda X}_{\text{Eigenvectors}}$

Let eigenvectors be a & b

→ For $\lambda_1 = 5$,

$$\begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = 5\begin{bmatrix} a \\ b \end{bmatrix}$$

$4a + b = 5a$

$a = b$

$2a + 3b = 5b$

$2a = 2b \Rightarrow a = b$

→ $X_1 = k\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

→ For $\lambda_2 = 2$

$$\begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 2 \begin{bmatrix} a \\ b \end{bmatrix}$$

$4a + b = 2a$

$2a + b = 0$

$2a = -b$

$2a + 3b = 2b$

$2a = -b$

$$X_2 = k \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

★ So, we can decompose covar matrix like this and we can write it as:-

$$\underbrace{\Sigma}_{\text{(Covariance Matrix)}} = D \Lambda D^{-1} \text{ or } D \Lambda D^{T} \quad (\text{since } D^{-1} = D^{T} \text{ orthogonal matrix})$$

Λ → Diagonal Matrix with Diagonal elements as eigenvalues

[D → Orthogonal matrix with eigenvectors as columns.]

★ Quantiles:- Suppose we have a probability distributn.

Quantiles divide dataset into equal-sized intervals.

→ Q_1 (25th Percentile) (0.25 or 25% of total area covered)

→ Q_2 (50th ") → (Median)

→ Q_3 (75th ") → (75% of area)

—/—/—

★ **Point Estimation :-**

→ Provides a single best guess for unknown parameter.

eg. Suppose we survey 100 students & find their average height → 168 cm
point estimate of populatⁿ mean height → 168 cm

(Estimating the populatⁿ mean (μ) using sample mean ($\bar{x}$))

★ **Interval Estimatⁿ :- (Precision)**

→ Provides range of values (confidence interval) within which a parameter is likely to lie.

→ More reliable than point estimatⁿ bcoz it accounts for uncertainity.

eg. From above eg., instead of saying 'avg. height is 168 cm', we say;
Avg height is b/w 165 cm and 171 cm with 95% confidence.

★ **Central Limit Theorem**

⇒ Statement :- It states that regardless of the shape of popuⁿ distributⁿ, the sampling distributⁿ of sample mean approaches normal distributⁿ as sample size (n) increases, provided samples are independent and identically distributed.

Mathematically, if $X_1, X_2, \ldots, X_n$ are independent R.V. with mean μ and variance σ^2, then standardized normal mean or standard normal variate:

$$\boxed{Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}$$

converges to a std. normal distributⁿ $N(0,1)$ as $n \to \infty$.

★ **Consequences :-**

1.) Justifies the use of Normal approximatⁿ (Sample mean follows " distributⁿ even if original data is not " " . Allows Normal based inferences in statistics)

2.) Enables Practical Statistical Analysis (Real world datasets → not normal ↓ Helps in converting them into normal)

3.) Improves Precisⁿ with Larger Samples

4.) Supports Inferential Statistics (t-tests, regressⁿ, ANOVA etc.)

★ **Limitations :-**

1.) High Sample Size Requirement (for non-normal or large error observⁿs)

2.) Indepence (of Samples) Assumptⁿ

3.) Finite Variance Requirement (Cauchy-Lorentz distributⁿ → CLT ⊗)