Chaitanya Kulkarni
B00814455

MACHINE LEARNING [CS-580L-01]
HOMEWORK-2.
(THEORY).

Q1.] Derive maximum likelihood estimators for

① Parameter p, Bernoulli (p) sample of size n

→ Let X be random variable having parameter p.

∴ we have sample $X = \{x_1, x_2, \ldots, x_n\}$

Probability density function is given by -

$$f(x) = p^x (1-p)^{1-x} \quad \text{---} \quad x = 0, 1.$$

∴ Likelihood

$$L = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$$

∴ Taking log, we get

$$\left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln (1-p)$$

To find max likelihood estimator, differentiate the eq^n w.r.t p

$$\therefore \frac{d}{dp}\left[\left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln (1-p)\right] = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

$$\therefore \frac{\sum_{i=1}^{n} x_i}{p} = \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\frac{1-p}{p} = \frac{n - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i}$$

$$\therefore p = \frac{\sum_{i=1}^{n} x_i}{n}.$$

$\therefore$ MLE for $p \Rightarrow \boxed{p = \frac{\sum_{i=1}^{n} x_i}{n}}$

2. Parameter p based on Binomial (N, p) sample of size n. Compute your estimators if the observed sample is $(3, 6, 2, 0, 0, 3)$ and $N = 10$.

→ Let X be random variable with parameters N and p of size n.

$\therefore$ Sample space $= X = \{x_1, x_2, \cdots, x_n\}$

Now Probability Density function of distribution (N,P) is given as -

$$f(x) = \binom{N}{P} p^x (1-p)^{n-x} \quad \cdots \quad x = 0,1,\ldots,n$$

$\therefore$ Likelihood

$$L = \prod_{i=1}^{n} \binom{N}{x} p^{x_i} (1-p)^{n-x_i}$$

taking log, we get

$$\sum_{i=1}^{n} \ln \binom{N}{x_i} + \left(\sum_{i=1}^{n} x_i\right) \ln(p) + \left(nN - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

now, differentiating w.r.t to p.

$$\frac{d}{dp}\left[\sum_{i=1}^{n} \ln \binom{N}{x_i} + \left(\sum_{i=1}^{n} x_i\right) \ln(p) + \left(nN - \sum_{i=1}^{n} x_i\right) \ln(1-p)\right] = 0$$

$$\frac{0 + \sum_{i=1}^{n} x_i}{p} - \frac{nN - \sum_{i=1}^{n} x_i}{1-p} = 0$$

$$\frac{1-p}{p} = \frac{nN - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i}$$

$$\therefore p = \frac{\sum_{i=1}^{n} x_i}{nN}$$

$\therefore$ MLE of $P \Rightarrow P = \dfrac{\sum_{i=1}^{n} x_i}{nN}$

Now, Given Sample = $\{3, 6, 2, 0, 0, 3\}$

$N = 10.$

$\therefore P = \dfrac{3+6+2+0+0+3}{10} = \dfrac{14}{60} = \dfrac{7}{30}$

$\boxed{P = 0.234}$

(3.) Parameters a and b based on a Uniform (a,b) sample of size n.

→ Like above, let sample be

$X = \{x_1, x_2, x_3, ----, x_n\}$

~~Prob~~ Probability density function for uniform Distribution

$$f(x) = \begin{cases} \left(\frac{1}{b-a}\right)^n & --- \quad a \le x_1 ... x_n \le b \\ 0 & \text{otherwise} \end{cases}$$

$\therefore$ Likelihood

$$L(X|a,b) = \left(\frac{1}{b-a}\right)^n$$

To find MLE, minimize the denominator

$\therefore$ It is monotonically increasing in 'a' and decreasing in 'b'.

$\therefore$ It is maximized at the Largest value of a and smallest value of b where density is not 0.

$\therefore$ MLE for a and b,

$\boxed{a = \min(x_i)}$ $\boxed{b = \max(x_i)}$

④ Parameter μ based on a Normal (μ, σ^2) sample of size n with known variance σ^2 and unknown mean μ.

→ let X be sample Normal random variable, such that $(x_1, ... x_n) \in X$ for size n.

$\therefore$ Probability Density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$\therefore$ likelihood,

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Taking log, we get,

$$\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

Now differentiate w.r.t μ

$$\frac{d}{d\mu}\left[\log L(x_1 \ldots x_n|\mu)\right] = \frac{d}{d\mu}\left[\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$\therefore \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \mu$$

$$n\hat{\mu} = \sum_{i=1}^{n} x_i$$

$$\therefore \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\therefore \text{MLE} \quad \boxed{\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

5. Parameter σ based on a Normal (μ, σ^2) sample of size n with known mean μ and unknown variance σ^2.

→ let sample $X = \{x_1, \ldots x_n\} \in X$.

Prob density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$\therefore$ MLE for $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$

Now, likelihood for Normal (μ, σ^2)

$$L(x_i|\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Taking log, we get

$$\log L(x_i|\sigma^2) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

Now differentiate wrt σ^2

$$\frac{d}{d\sigma^2}\left[\log L(x_i|\sigma^2)\right] = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^4}$$

$$\therefore -n\sigma^2 + \sum_{i=1}^{n} (x_i-\mu)^2$$

$$\therefore n\sigma^2 = \sum_{i=1}^{n} (x_i-\mu)^2$$

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

$$\therefore \text{MLE} = \boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i-\mu)^2}{n}}}$$

(6.) Parameters (μ, σ^2) based on a Normal (μ, σ^2) sample of size n with unknown mean μ and variance σ^2.

→ Like above, we have prob. density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$$

∴ Likelihood,

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \cdot e^{\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

Taking log, we get,

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

Taking partial derivative w.r.t μ

$$-2\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} = 0$$

$$\sum_{i=1}^{n}\frac{(x_i-\mu)}{2\sigma^2} = 0, \quad \sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\therefore \sum_{i=1}^{n}\mu = \sum_{i=1}^{n}x_i \quad \therefore n\mu = \sum_{i=1}^{n}x_i$$

$$\therefore \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

$$\therefore \text{MLE} \ \boxed{\hat{\mu} = x}.$$

Now MLE for σ^2, take partial derivative of eq (1) w.r.t σ^2.

$$\therefore \frac{d}{d\sigma^2}\left[\log L(\mu, \sigma^2)\right] = \frac{-n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2(\sigma^2)^2} = 0$$

$$\therefore -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

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

Now substitute μ in above eqn.

$$\therefore \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-x)^2}{n}.$$

$$\therefore \text{MLE} = \boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-x)^2}{n}}}$$

Q2.] Given, α_H (Coins) $= 60$ $\quad$ α_T (coins) $= 40$
α_H (Thumbtack) $= 70$ $\quad$ α_T (Thumbtack) $= 30$

① MLE and MAP Estimates –

Coin

$$MLE_{(Heads)} = \frac{\alpha_H}{\alpha_H + \alpha_T} = \frac{60}{60+40} = \underline{\underline{0.6}}.$$

$$MLE\ (Tails) = \frac{\alpha_T}{\alpha_H + \alpha_T} = \frac{40}{60+40} = \underline{\underline{0.4}}$$

Thumbtack

$$MLE\ (Heads) = \frac{\alpha_H}{\alpha_H + \alpha_T} = \frac{70}{70+30} = \underline{\underline{0.7}}$$

$$MLE\ (Tails) = \frac{\alpha_T}{\alpha_H + \alpha_T} = \frac{30}{70+30} = \underline{\underline{0.3}}$$

* Maximum a Posteriori Estimation :

① For $B(1,1)$

Coin

$$MAP_{(Heads)} = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2} = \frac{60+1-1}{60+1+40+1-2} = \underline{\underline{0.6}}$$

$$MAP_{(T)} = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2} = \frac{40+1-1}{60+1+40+1-2} = \underline{\underline{0.4}}$$

Thumbtack

$$MAP\ (Heads) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2} = \frac{70-1+1}{70+1+30+1-2} = \underline{\underline{0.7}}$$

$$MAP_{(T)} = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2} = \frac{30-1+1}{70+1+30+1-2} = \underline{\underline{0.3}}$$

(2.) For Beta (40, 60)

Coin

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{60 + 40 - 1}{60 + 40 + 40 + 60 - 2}$$

$$= \underline{\underline{0.5}}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{40 + 60 - 1}{60 + 40 + 40 + 60 - 2}$$

$$= \underline{\underline{0.5}}$$

Thumbtack.

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{70 + 40 - 1}{70 + 40 + 30 + 60 - 2}$$

$$= \underline{0.551}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{30 + 60 - 1}{70 + 40 + 30 + 60 - 2}$$

$$= \underline{\underline{0.449}}$$

(3.) For Beta (30, 70)

Coin

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{60 + 30 - 1}{60 + 30 + 40 + 70 - 2}$$

$$= \underline{\underline{0.449}}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{40 + 70 - 1}{60 + 30 + 40 + 70 - 2}$$

$$= \underline{\underline{0.551}}$$

Thumbtack.

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{70 + 30 - 1}{70 + 30 + 30 + 70 - 2}$$

$$= \underline{\underline{0.5}}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{30 + 70 - 1}{70 + 30 + 30 + 70 - 2}$$

$$= \underline{\underline{0.5}}$$

④ For Beta (100, 100)

Coin

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{60 + 100 - 1}{60 + 100 + 40 + 100 - 2}$$

$$= \underline{0.5335}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{40 + 100 - 1}{60 + 100 + 40 + 100 - 2}$$

$$= \underline{0.4665}$$

Thumbtack

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{70 + 100 - 1}{70 + 100 + 30 + 100 - 2}$$

$$= \underline{0.567}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{30 + 100 - 1}{70 + 100 + 30 + 100 - 2}$$

$$= \underline{0.4328}$$

⑤ For Beta (1000, 1000)

Coin

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{60 + 1000 - 1}{60 + 1000 + 40 + 1000 - 2}$$

$$= \underline{0.505}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{40 + 1000 - 1}{60 + 1000 + 40 + 1000 - 2}$$

$$= \underline{0.495}$$

Thumbtack

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{70 + 1000 - 1}{70 + 1000 + 30 + 1000 - 2}$$

$$= \underline{0.509}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{30 + 1000 - 1}{70 + 1000 + 30 + 1000 - 2}$$

$$= \underline{0.491}$$

⑥ For Beta (100000, 100000)

Coin

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{60 + 100000 - 1}{60 + 100000 + 40 + 100000 - 2}$$

$$= \underline{\underline{0.50005}}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{40 + 100000 - 1}{60 + 100000 + 40 + 100000 - 2}$$

$$= \underline{\underline{0.49995}}$$

Thumbtack

$$MAP(H) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{70 + 100000 - 1}{70 + 100000 + 30 + 100000 - 2}$$

$$= \underline{\underline{0.50009}}$$

$$MAP(T) = \frac{\alpha_T + \beta_T - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{30 + 100000 - 1}{70 + 100000 + 30 + 100000 - 2}$$

$$= \underline{\underline{0.49991}}$$

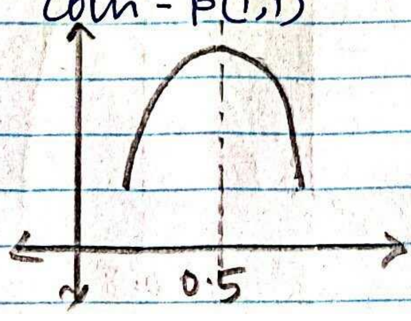
ii. Curve for each scenario –

① For Beta (1,1)

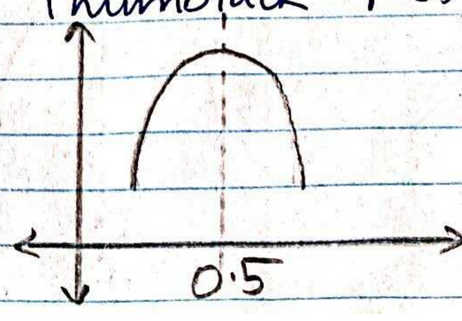
- Since the prior Beta β(1,1) is very small ~~as~~ as compared to given α_H and α_T. Curve will tilt towards α_H and α_T values rather than the minimal Prior Beta.

∴ Graph for β(1,1)

Coin - β(1,1)

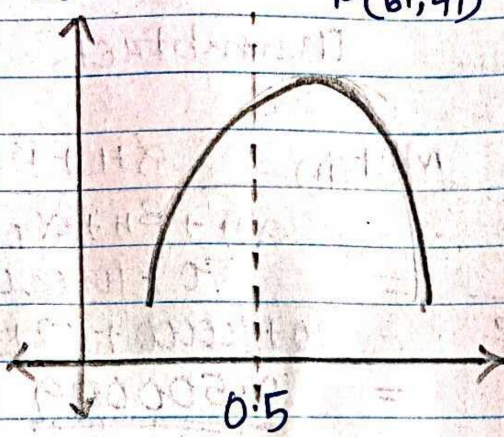


Thumbtack β(1,1)

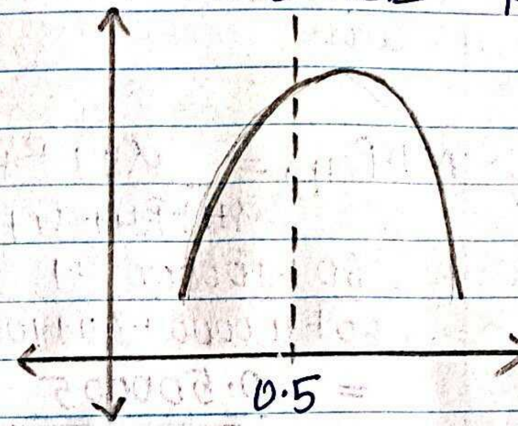


now, adding α values

Coin $\beta(61,41)$



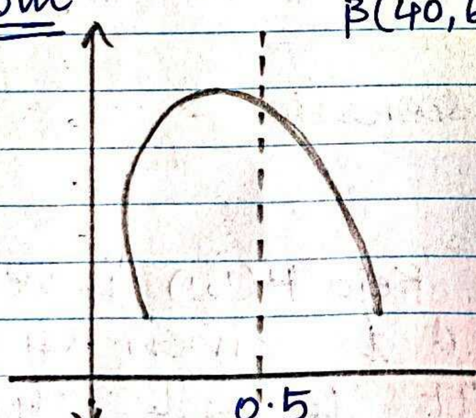
Thumbtack $\beta(71,31)$



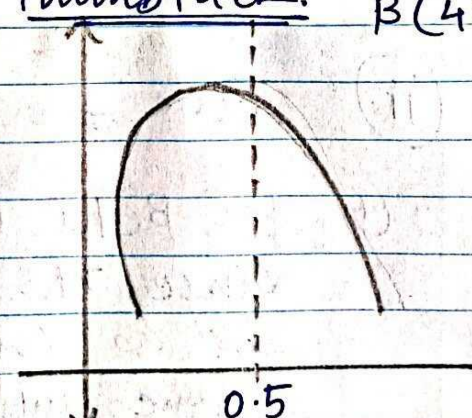
(2) For Beta (40, 60)

- Since $\beta_H < \beta_T$, curve tilts left and after adding α values, the curve fits to the mean since $\beta_H = \beta_T$.

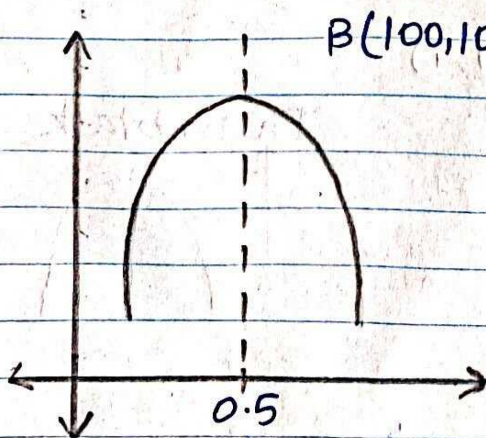
<u>Coin</u> $\beta(40,60)$



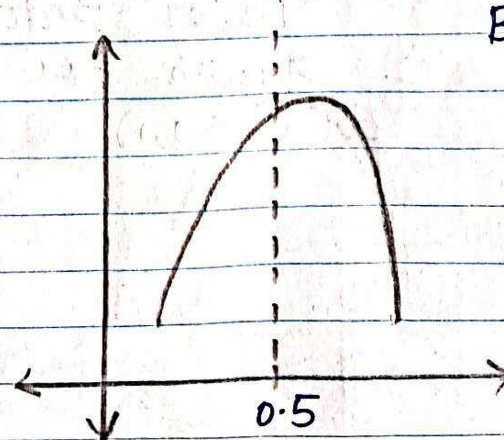
<u>Thumbtack.</u> $\beta(40,60)$



$B(100,100)$



$B(110,90)$

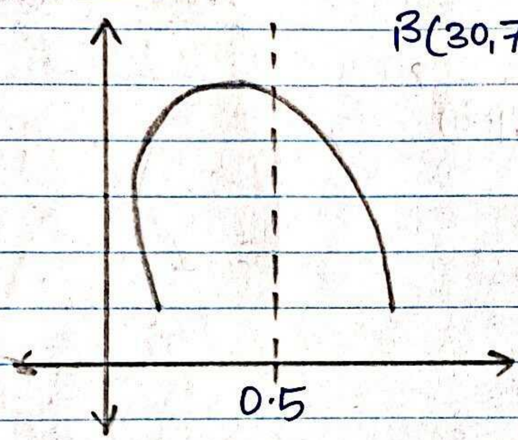


③ For Beta (30,70)

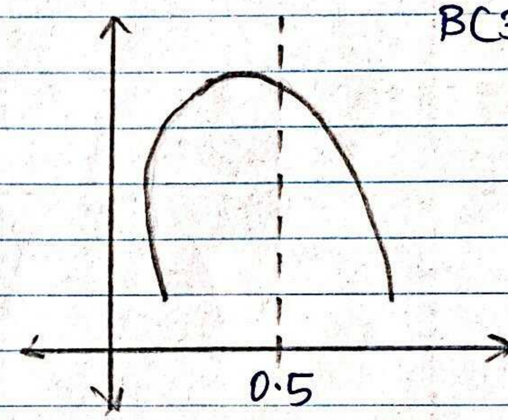
- Since $\beta_H < \beta_T$, again curve tilts towards left.
- After adding α values, curve will tilt towards left for coin and
- Curve remain fit for Thumbtack

<u>Coin</u> <u>Thumbtack.</u>

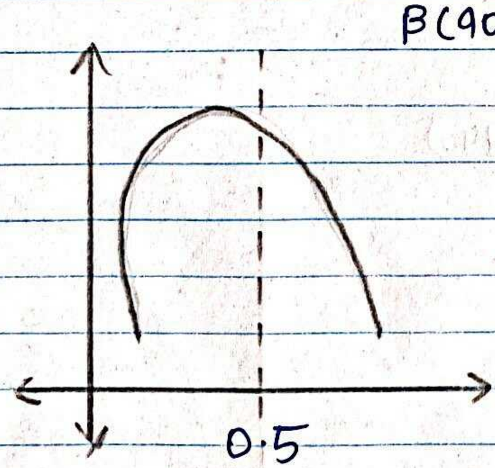
$\beta(30,70)$



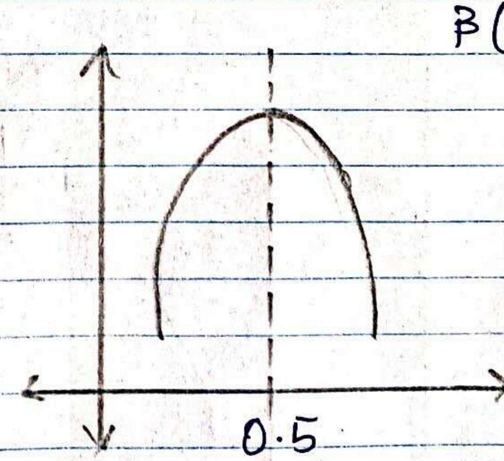
$B(30,70)$



$\beta(90,110)$



$\beta(100,100)$

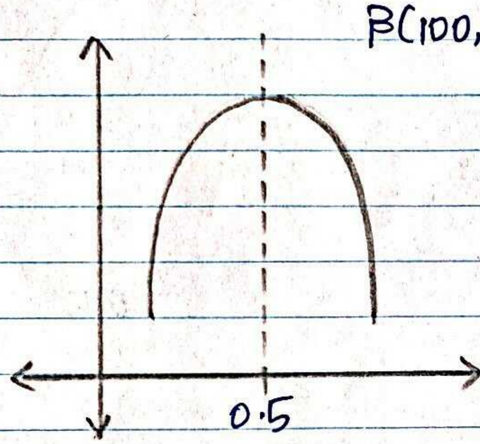


(4) For Beta (100,100)

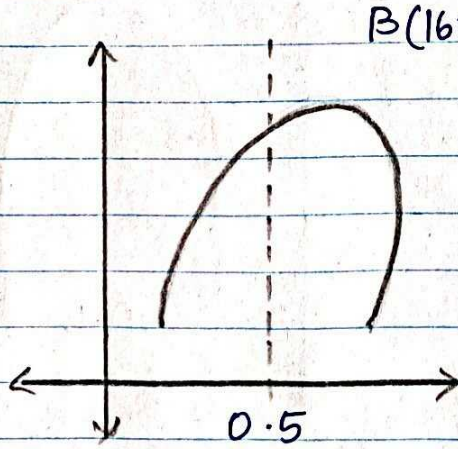
- Since $\beta_H = \beta_T$, the curve fits the mean graph.
- After adding α values, $\beta_H > \beta_T$. Hence the graph will tilt towards right.

Coins

$\beta(100,100)$

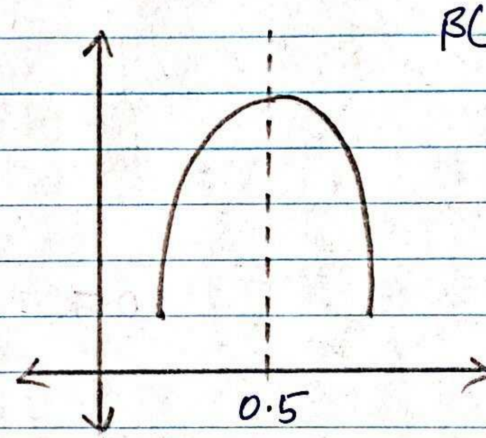


$\beta(160,140)$

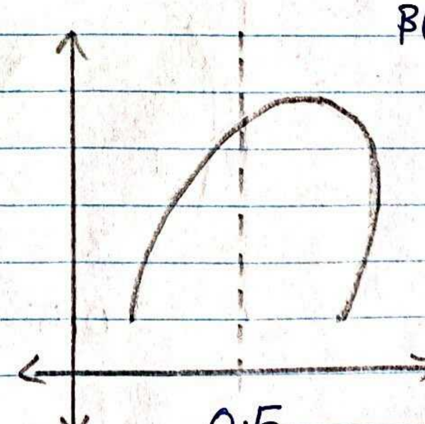


Thumbtack

$\beta(100,100)$



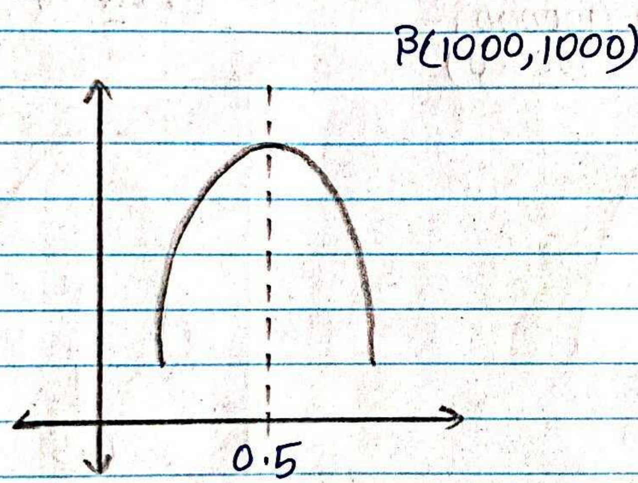
$\beta(170,130)$

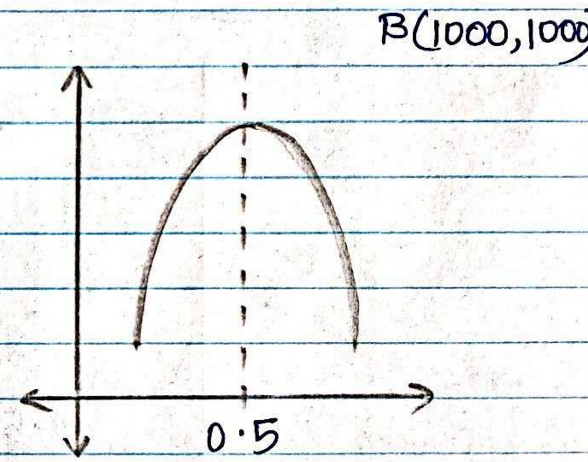


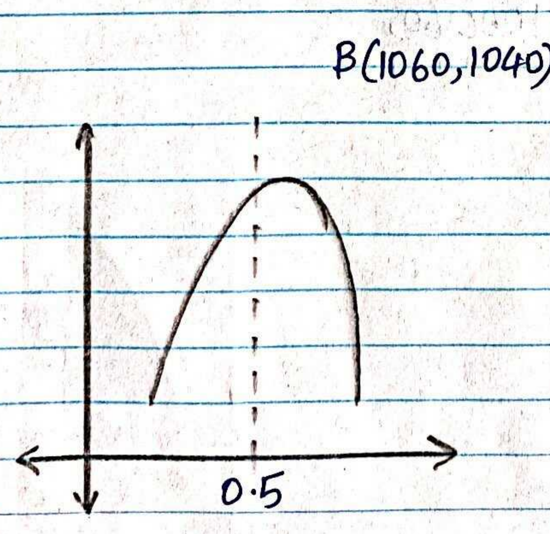
(5) For Beta (1000, 1000)

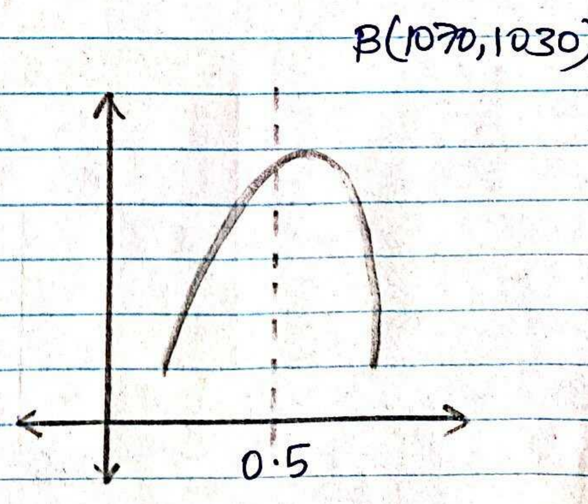
- Since $\beta_H = \beta_T$, the curve will ~~b~~ rem-
  -ain in the centre.
- After adding α values, ~~the~~ since the β values are much larger than α, the β will dominate the curve, and the curve will slightly tilt towards right.

Coin | Thumbtack.







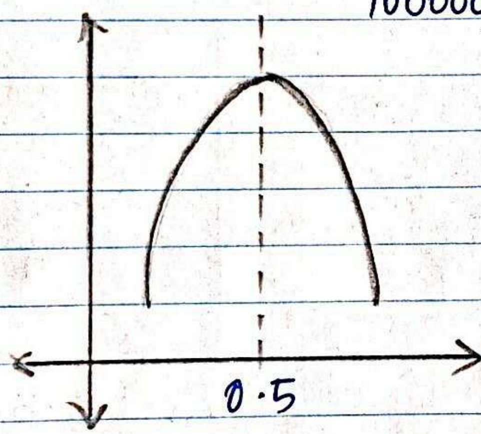


6. For Beta (100000, 100000)

- Since $\beta_H = \beta_T$, the curve will remain in the centre.
- Also, since the β values are very large, β will dominate the curve even after adding the α values.
- Hence the graph will slightly tilt towards right.

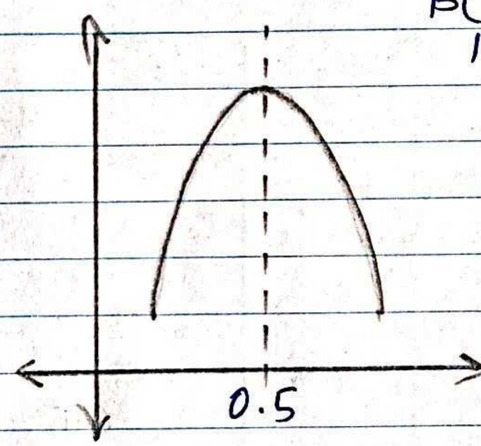
Coin

$\beta(100000, 100000)$

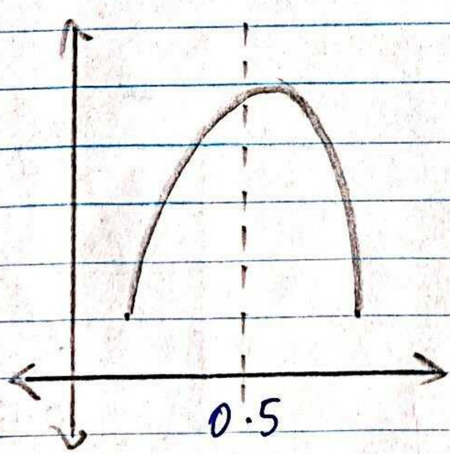


Thumbtack

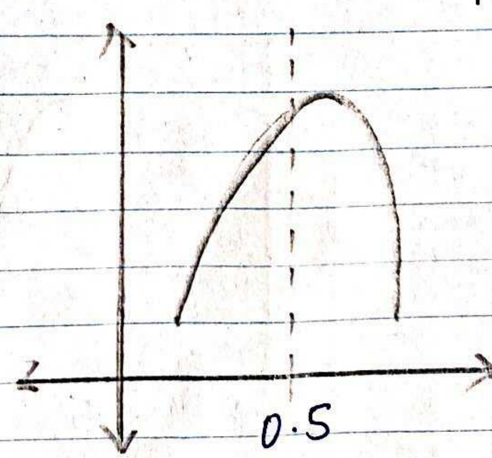
$\beta(100000, 100000)$



$\beta(100060, 100040)$



$\beta(100070, 100030)$



(iii.) TRUE:

- As we collect more data instances, the estimation would tilt towards the α values. The maximum likelihood Estimation (MLE) would dominate the Prior Beta values since it would be constant if we increase the data instances.
- Hence, the data instances increase the MLE estimate and would approach MAP estimation.

(iv.) TRUE.

- When the Prior value is very large, like in the example given, $\beta(100000, 100000)$ or larger, then the MAP estimate will dominate the α values since α values are comparatively small.
- Hence the estimation graph will be tilted towards the prior Beta values.
- Therefore the MAP estimate would approach the same value when we use a larger prior.