Q1. Maximum likelihood estimators for-

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

Let X be a Bernoulli random variable having parameter p.
We are given sample size = n, hence random sample of X = {X_1, X_2, ... X_n)

Probability density function for any f(x) with parameter (p) is given by-
$f(x) = p^x (1-p)^{1-x}$, $x = 0, 1$

∴ Likelihood of sample is given by the product of all the random sample values (X_i) from 1 to n.

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

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

Taking log-likelihood,

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

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

Maximum likelihood estimator of $f(x)$ can be found by differentiating with respect to p and equating the result to 0.

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

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

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

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

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

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

1.2) parameter p based on a Binomial(N, p) sample of size n. Compute estimators for observed sample $(3, 0, 2, 0, 0, 3)$ and $N = 10$.

Let X be a Binomial random variable having parameters N and P.
Sample size is n.
So, $X = \{x_1, x_2 \dots x_n\}$

Probability density function for the Binomial distribution (N, P) —

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

Likelihood function of the given sample is given by the product of all the random values (x_i) from 1 to n.

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

Taking the log-likelihood,

$$\Rightarrow \ln \left[\prod_{i=1}^{n} \binom{N}{x} p^{x_i} (1-p)^{N-x_i} \right]$$

Scanned with CamScanner

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

Maximum likelihood of $f(x)$ can be found by differentiating with respect to P and equating the result to 0.

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

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

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

$$\Rightarrow \frac{1}{P} - \cancel{\frac{P}{P}} = \frac{nN}{\sum_{i=1}^{n} x_i} - \cancel{\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i}}$$

$$P = \frac{\sum_{i=1}^{n} x_i}{nN} \quad \left(\text{MLE of Binomial } (N, P)\right)$$

Sample $(3, 6, 2, 0, 0, 3)$, $N = 10$, $n = 6$

$$P = \frac{3+6+2+0+0+3}{6 \times 10} \Rightarrow \frac{14}{60} = \frac{7}{30}$$

$$\therefore P = \frac{7}{30}$$

1.3) parameters a, and b based on a uniform (a, b) sample of size n

Let x be any uniform random variable having parameters a and b.
Sample size is n
So, $X = \{x_1, x_2 \dots x_n\}$

Probability density function for the uniform distribution is —

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{else} \end{cases}$$

Likelihood function of the given sample is —

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

For MLE of L, we need to minimize the denominator —

$$\min(b-a)$$

Even after minimizing the difference between a and b, we need to keep all the sample of x $\{x_1 .. x_n\}$ in the range of a, b.

Maximum likelihood estimator for a and b would be -

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

1. 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 a random variable.
Sample size is n.
So, $x = \{x_1, x_2, \ldots x_n\}$

Probability density function for normal distribution is –

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

Likelihood function for the given sample is –

$$L(x_1 \ldots x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Taking the log-likelihood,

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

Maximum likelihood of f(x) can be found by differentiating with respect to μ and equating the result to 0.

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

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

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

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

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

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

Let x be a Normal random variable.
Sample size is n.
So, $d = \{x_1, x_2 \ldots \ldots x_n\}$

Probability density function for normal distribution is –

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

Likelihood function for the given sample is –

$$L(x_1, x_2 \ldots x_n | \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(x-\mu)^2}{2\sigma^2}$$

Taking the log-likelihood,

$$\log L(x_1, x_2 \ldots x_n | \sigma^2) = \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x-\mu)^2}{2\sigma^2}$$

Maximum likelihood estimator can be found by differentiating with respect to σ^2 and equating the result with 0.

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

Scanned with CamScanner

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

Multiplying both side by $2(\sigma^2)^2$ we get

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

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

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

If μ is known we can substitute $\mu = x$.

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

Scanned with CamScanner

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

Probability density function for normal distribution is -

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

(Here, $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$)

Likelihood function for the given sample is -

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

$$= \prod_{i=1}^{n} (\sigma^2)^{-n/2} (2\pi)^{-n/2} \exp\left[\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right]$$

$\llcorner$ (a)

Taking the Log Likelihood -

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

Taking partial derivative with respect to μ and equating the result to 0 -

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

Scanned with CamScanner

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

Multiplying both sides with σ^2, we get

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

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

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

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} \quad - (b)$$

MLE for unknown μ.

For MLE of σ^2, we take the partial derivative of equation (a) with respect to σ^2 and equating the result to 0.

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

Multiplying both side by $2(\sigma^2)^2$ we get,

$$0 = \left[\frac{-n}{2\sigma^2} + \frac{\Sigma(x_i - \mu)^2}{2(\sigma^2)^2}\right] \times 2(\sigma^2)^2$$

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

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

Substituting $\mu = x$ from equation (b) —

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

This is MLE for σ^2.

Q2.

1. Coin Toss – 60 Heads, 40 Tails
Thumbstack – 70 Heads, 30 Tails
Beta priors – Beta(1,1), Beta(40,60), Beta(30,70), Beta(100,100), Beta(1000,1000) and Beta(100000,100000)

MLE and MAP estimates for coin and thumbstack

MLE for the coin

$$\hat{\theta}_{coin} = \frac{\beta_H - 1}{\beta_H + \beta_T - 2} = \frac{60-1}{60+40-2} = \frac{59}{98}$$

$$\hat{\theta}_{coin} = 0.60$$

Maximum likelihood estimator for thumbstack

$$\hat{\theta}_{MLE} = \frac{\beta_H - 1}{\beta_H + \beta_T - 2} = \frac{70-1}{70+30-2} = \frac{69}{98}$$

$$\hat{\theta}_{thumbstack} = 0.70$$

MLE for coin = 0.60
MLE for thumbstack = 0.70

MAP for coin, prior Beta(1,1) and likelihood priors 60, 40

$$\hat{\theta}_{MAP} = \arg\max P(\theta | D)$$

Scanned with CamScanner

$$= \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

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

$$= 0.60$$

Prior Beta (40, 60) and likelihood prior 60, 40

$$\hat{\theta}_{MAP} = \arg\max P(\theta / D)$$

$$= \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{60 + 40 - 1}{60 + 40 + 40 + 60 - 2} = \frac{99}{198}$$

$$= 0.5$$

Prior Beta (30, 70) and likelihood prior 60, 40

$$\hat{\theta}_{MAP} = \arg\max P(\theta / D)$$

$$= \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

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

$$= \frac{89}{198} = 0.44$$

Thumbstack

Prior Beta(100, 100) and likelihood priors 70, 30

$$\hat{\theta}_{MAP} = \arg\max P(\theta / D)$$

$$= \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

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

$$= \frac{169}{298} = 0.56$$

Prior Beta (1000, 1000) and likelihood priors 70, 30

$$\hat{\theta}_{MAP} = \arg\max P(\theta / D)$$

$$= \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

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

$$= \frac{1069}{2098}$$

$$= 0.509$$

Prior Beta (100,000, 10,000) and likelihood priors 70, 30

$$\hat{\theta}_{MAP} = \arg\max P(\theta | D)$$

$$= \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

$$= \frac{70 + 100,000}{70 + 100,000 + 30 + 100,000 - 2}$$

$$= \frac{100,069}{200,098}$$

$$= 0.500$$

Q2.2 Coin : 60H 40T

$\beta(1,1)$, $\beta(40,60)$, $\beta(30,70)$

$\alpha_H = 60$, $\alpha_T = 40$

First

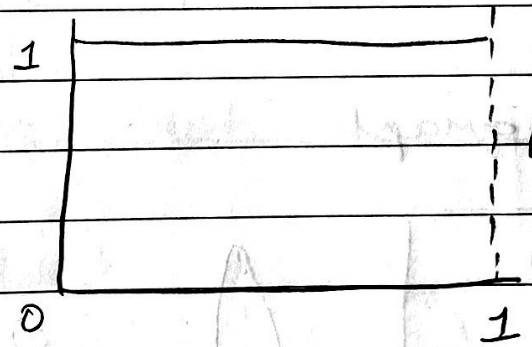
$\beta_H = 1$ $\beta_T = 1$

$\alpha_T + \beta_T = 41$

$\alpha_H + \beta_H = 61$

Prior graph -

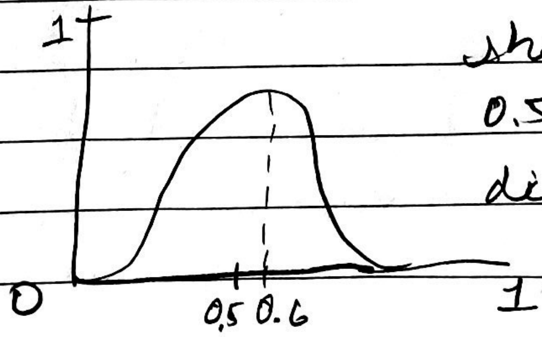
$\beta(1,1)$



Parameter values upto 1 uniform distribution

Posterior graph -

$\beta(61,41)$



Peak slightly shifted to right of 0.5 since there is difference in parameter values w.r.t posterior values.

Second

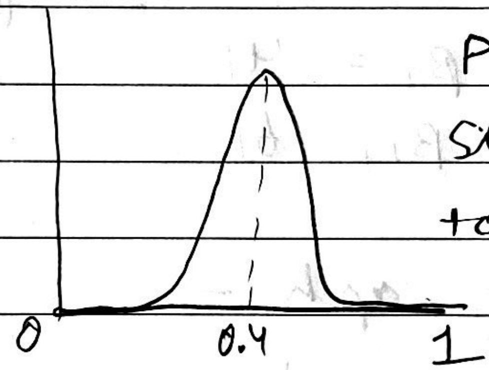
$\beta_H = 40$ $\beta_T = 60$

$\alpha_T + \beta_T = 100$

$\alpha_H + \beta_H = 100$

Prior graph

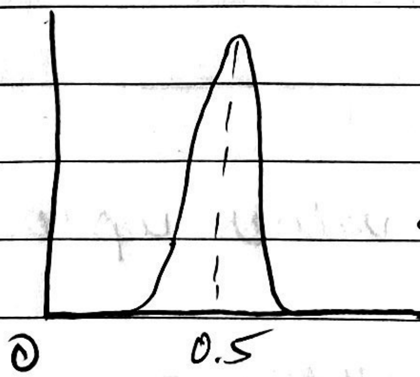
$\beta(40, 60)$



Peak on 0.4. Slightly shifted to the left of 0.5.

Posterior graph

$\beta(100, 100)$



Peak on 0.5. Slightly wide graph but not as wide as first case

Scanned with CamScanner

Third

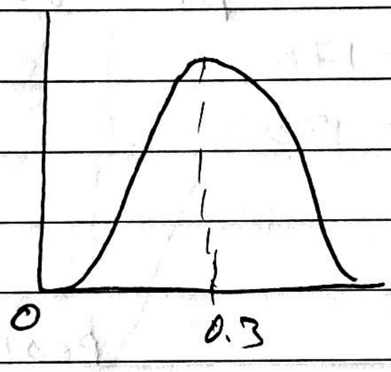
$\beta_H = 30$ $\beta_T = 70$

$\alpha_T + \beta_T = 110$

$\alpha_H + \beta_H = 90$

Prior graph

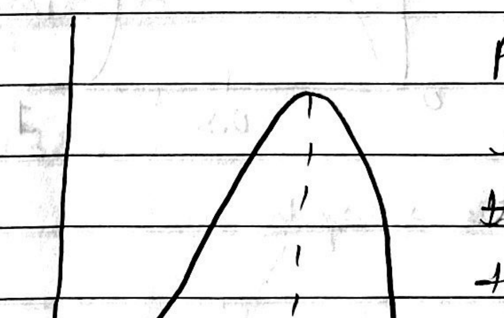
$\beta(30, 70)$



Peak on 0.3. Shifted left of 0.5.

Posterior graph

$\beta(90, 110)$



Peak slightly shifted from 0.5 to left since there is difference in parameter wrt posterior values.

Thumbstack : 70 H 30 T

$\beta(100,100)$, $\beta(1000,1000)$, $\beta(100000, 100000)$

$\alpha_H = 70$ $\alpha_T = 30$

First

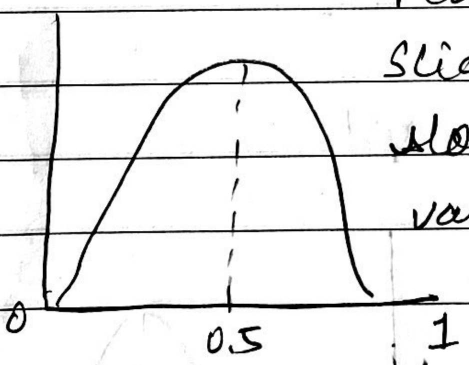
$\beta_H = 100$ $\beta_T = 100$

$\alpha_T + \beta_T = 170$

$\alpha_H + \beta_H = 130$

Prior graph

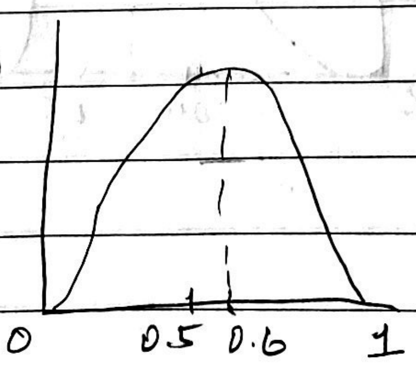
$\beta(100,100)$



Peak on 0.5
slightly wide graph
slope since prior
value are not too
high.

Posterior graph

$\beta(170,130)$



Peak slightly
shifted to the
right of 0.5.

Second

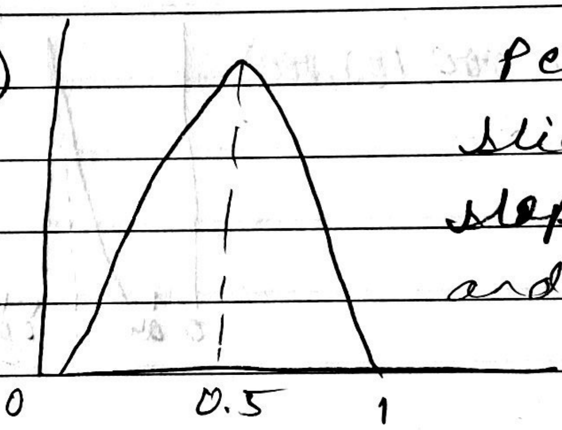
$\beta_H = 1000$, $\beta_T = 1000$

$\alpha_T + \beta_T = 1070$

$\alpha_H + \beta_H = 1030$

Prior graph

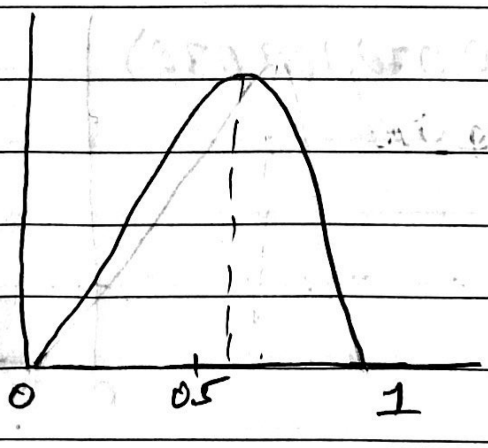
$\beta(1000, 1000)$



Peak on 0.5. slightly steep slope as rising and descending near 0.5.

Posterior graph

$\beta(1070, 1030)$



Peak slightly shifted to the right of 0.5.

Third

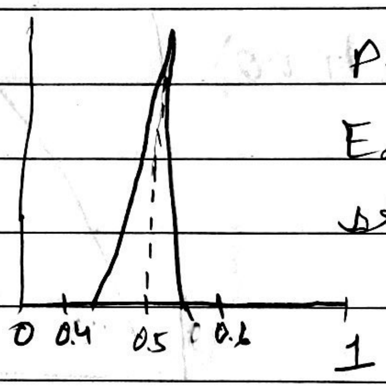
$\beta_H = 100,000$ $\beta_T = 100,000$

$\alpha_T + \beta_T = 100,070$

$\alpha_H + \beta_H = 100,030$

Prior graph

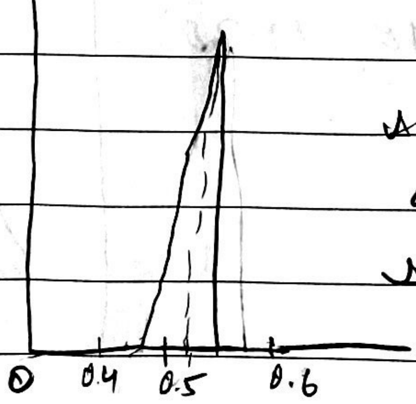
$\beta(100,000, 100,000)$



Peak on 0.5. Extremely steep slope as rising and descending near 0.5

Posterior graph

$\beta(100,070, 100,030)$



Peak slightly shifted from 0.5. Extremely steep slope as rising and descending near 0.5.

Q2.3. TRUE

If Beta prior is uniform then the MLE estimate will approach the MAP estimate no matter how many data instances we collect.

If Beta prior is non uniform, given infinite data points, MLE estimate approaches the MAP estimate. In case of Bernoulli Trial with a Beta prior (alpha, beta), the MAP estimate (for heads) is given by:

$$\frac{(\text{number of heads} + \alpha - 1)}{(\text{total number of trials} + \beta + \alpha - 2)}$$

If we divide the numerator and the denominator by number of trails, we will see that, since the number of data points are infinite, $\frac{\beta + \alpha - 2}{\text{number of trials}}$ will tend to 0, and $\frac{\alpha - 1}{\text{number of trials}}$ will also tend to 0. So, MAP estimate becomes $\frac{\text{number of heads}}{\text{number of trials}}$, and it approaches the MLE estimate, Both are same.

Q2.4. FALSE

When the MLE of coin and thumbstack are different the MAP estimate doesnot approach the same value. It is equal only if the Beta prior is sampled from uniform distribution.