I have done this assignment completely on my own. I have not copied it, nor have I given my solution to anyone else. I understand that if I am involved in plagiarism or cheating I will have to sign an official form that I have cheated and that this form will be stored in my official university record. I also understand that I will receive a grade of 0 for the involved assignment for my first offense and that I will receive a grade of “F” for the course for any additional offense.

Bhuvan Chadha

MACHINE LEARNING
ASSIGNMEN-2

Submitted By -
BHUVAN CHADHA
B00815783

Point Estimation

Q1) 1. Maximum Likelihood Estimator for parameter p, Bernaulli (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 $X = \{X_1, X_2 \ldots X_n\}$

Now, we know, Probability Density function for any $f(x)$ with parameter (p) is given by -

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

Hence, likelihood of ~~$f(x)$~~ Sample is given by the product of all the random Sample values (X_i) from 1 to n.

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

The above equation can be written as,

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

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 w.r.t p and equating the result to 0.

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

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

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

Solving for p,

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

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

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

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

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

This is the Maximum Likelihood ~~Function~~ Estimator (MLE) for p.

2. Maximum likelihood Estimator for parameter p based on a Binomial (N, p) Sample of size n.

Assuming X as any Binomial Random variable with parameters N and p. for size n, let $X = \{X_1, X_2, \dots X_n\}$ be a random sample.

Probability Density function for the Binomial distribution (N, p) is given—

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

$$(x = 0, 1, \dots n)$$

Hence, likelihood function of the given sample is given by the product of all the random sample 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,

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

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

We can now find the maximum likelihood by differentiating the above expression w.r.t p and equating it to 0.

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

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

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

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

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

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

This is the maximum likelihood estimator (MLE) for Binomial (N, p) for p.

Now, we are given a sample $(3, 6, 2, 0, 0, 3)$ and $N = 10$. Here, sample size, $n = 6$.

Substituting the values in the above derived equation for p,

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

$$\therefore \boxed{p = \frac{7}{30}}$$

3. Maximum likelihood Estimator for parameters a and b based on a Uniform (a,b) sample of size n.

Assuming X as any Uniform Random Variable with parameters a and b.
For size n, let $x = \{x_1, x_2, \ldots x_n\}$ be a random sample.

Probability Density function for the Uniform Distribution for a and b is given by -

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

Hence, the likelihood function of the given sample is given by L.

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

Now, to ~~ma~~ find the Maximum likelihood function for L, we need to minimize the denominator of the function, i.e.,

$$\min(b-a)$$

Even after minimizing the difference between the parameters a and b, we need to keep all the samples of ~~x~~ $X = \{X_1, \ldots, X_n\}$ in the range a, b.

Maximum likelihood Estimator for a and b would be –

$$\hat{a} = \min(X_i)$$

$$\text{and, } \hat{b} = \max(X_i)$$

These values give us the minimum length as it is the smallest interval to include all sample points of X.

4. Maximum likelihood Estimators for parameter μ based on a Normal (μ, σ^2) Sample of size n with known σ^2 and unknown μ

Let X be any sample ~~space~~ Normal random variable, such that $(X_1, \ldots X_n) \in X$ for n-size.

Probability Density function for Normal distribution with parameters μ and σ^2

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

Hence, likelihood function for Normal (μ, σ^2) is given by -

$$L(X_1, \ldots X_n | \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

taking log-likelihood on both sides,

$$\log L(X_1, \ldots X_n | \mu) = \log \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right]$$

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

Now, to get Maximum likelihood estimator, we need to differentiate L w.r.t μ and equating the result to zero.

$$\therefore \frac{d}{d\mu}\left[\log L(X_1, \ldots X_n | \mu)\right] = \frac{d}{d\mu}\left[\sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

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

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

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

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

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

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

6. Maximum Likelihood Estimators for 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 parameters μ and σ^2 is given by –

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

Here, $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$ holds true.

Now, likelihood function can be written as—

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

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

Taking the log-likelihood on both sides,

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

$\longrightarrow$ ①

Now, taking partial derivative w.r.t μ and equating the result to 0, to get MLE,

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

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

Multiplying both sides by σ^2

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

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

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

$$\Rightarrow \boxed{\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}} \Rightarrow x \longrightarrow (2)$$

This is the maximum likelihood estimator for unknown μ.

To find the MLE for σ^2, we take the partial derivative of equation (1) w.r.t σ^2 and equating the result with 0.

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

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

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

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

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

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

Substituting $\mu = x$ from eq^n ②,

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

This is the Maximum Likelihood Estimator for σ^2.

5. Maximum Likelihood Estimator for parameter σ based on a Normal (μ, σ^2) Sample of size n with known μ and unknown σ^2.

Let X be any Normal Random Variable. For size n, we have $\{X_1, \dots X_n\} \in X$

Probability Density Function for Normal distribution for parameters μ and σ^2 is -

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

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

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

Substituting $\mu = x$ from eq^n ②,

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

This is the Maximum Likelihood Estimator for σ^2.

5. Maximum Likelihood Estimator for parameter σ based on a Normal (μ, σ^2) Sample of size n with known μ and unknown σ^2.

Let X be any Normal Random Variable. For size n, we have $\{X_1, \ldots X_n\} \in X$

Probability Density Function for Normal distribution for parameters μ and σ^2 is -

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

$\therefore$ Likelihood function for Normal (μ, σ^2) is given by -

$$L(X_1, \dots X_n | \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x-\mu)^2}{2\sigma^2}$$

Taking Log-likelihood on both sides,

$$\log L(X_1, \dots X_n | \sigma^2) = \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x-\mu)^2}{2\sigma^2}$$

For Maximum likelihood Estimator, we differentiate L w.r.t σ^2 and equating the result with 0.

$$\frac{d}{d\sigma^2}\left[\log L(X_1 \dots X_n | \sigma^2)\right] = \cancel{\frac{d}{d\sigma}}\left[\cancel{\log A}\right.$$

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

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

Multiplying both sides by $2(\sigma^2)^2$,

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

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

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

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

Now, if μ is known, we can substitute $\mu = x$ in above equation.

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

This is the maximum Likelihood Estimator for parameter σ^2.

Q 2)

(1) We are given –

A Coin – 60 Heads, 40 Tails

A Thumbtack – 60 Heads, 40 Tails

(α_H) (α_T)

Also, Beta priors given are –

Beta (1000, 1000), Beta (100, 100),

Beta (1, 1).

Now,

Maximum Likelihood Estimator for the coin and the thumbtack is –

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

$$\Rightarrow \frac{60 - 1}{60 + 40 - 2} \Rightarrow \frac{59}{98}$$

$$\therefore \hat{\theta}_{MLE} = \underline{\underline{0.60}}$$

MLE for both coin and thumbtack is same as the data given is same for both.

MAP: 'Maximum A Posteriori' Approximation uses the most likely parameter and the priors.

$\therefore$ For Prior ~~Beta 100~~ Beta(1000, 1000), and likelihood priors 60 and 40,

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

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

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

$$\Rightarrow \frac{1059}{2098} \Rightarrow \underline{0.50}$$

For prior values Beta (100, 100), using the above formula,

$$\hat{\theta}_{MAP} = \frac{60 + 100 - 1}{60 + 100 + 40 + 100 - 2}$$

$$\Rightarrow \frac{159}{298} \Rightarrow \underline{0.53}$$

Again, for Prior Beta(1,1), using the same formula,

$$\hat{\theta}_{MAP} = \frac{60+1-1}{60+1+40+1-2}$$

$$\Rightarrow \frac{60}{100} \Rightarrow \underline{\underline{0.60}}$$

(2.) We now know that the value of Maximum Likelihood Estimator is increasing when Beta(β_H, β_T) decreases.

Scenario 1,

Beta$(1000, 1000)$.

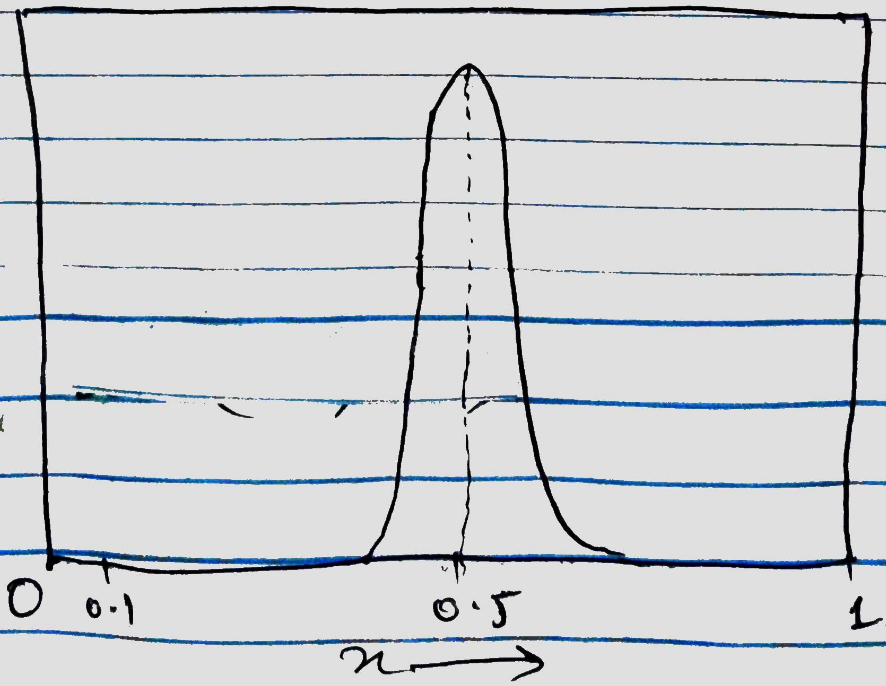
Priors $\rightarrow$ $\beta_H = 1000$, $\beta_T = 1000$

Data $\rightarrow$ $\alpha_H = 60$, $\alpha_T = 40$.

Posteriors $\rightarrow$ Beta$(1060, 1040)$

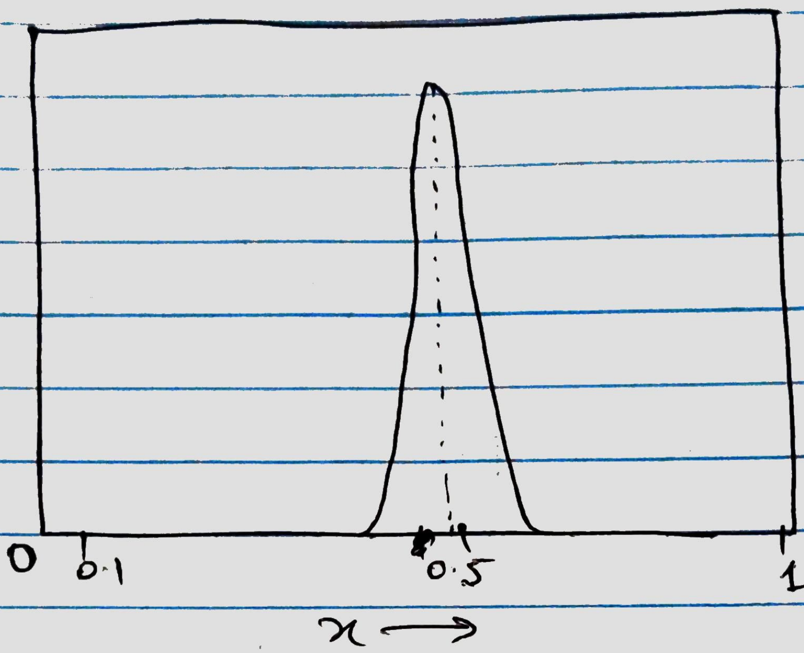
Creating a graph for the scenario,

Prior Beta$(1000, 1000)$



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

Posterior Beta (1060, 1040)



$\alpha_H + \beta_H = 1060$
$\alpha_T + \beta_T = 1040$

- Here, the Prior graph is a steep slope rising and descending very close to 0.5, because the parameter value [0, 1] (x) is very small as compared to the Beta (1000, 1000) priors.

Similarly, in posterior graph, since there has not been significant increase or decrease in the Beta prior, which is Beta (1060, 1040), the graph remains approximately similar to the prior.

- And since the β_H and β_T and $\alpha_H + \beta_H$ and $\alpha_T + \beta_T$ values are equal or very close, the graph is centered.

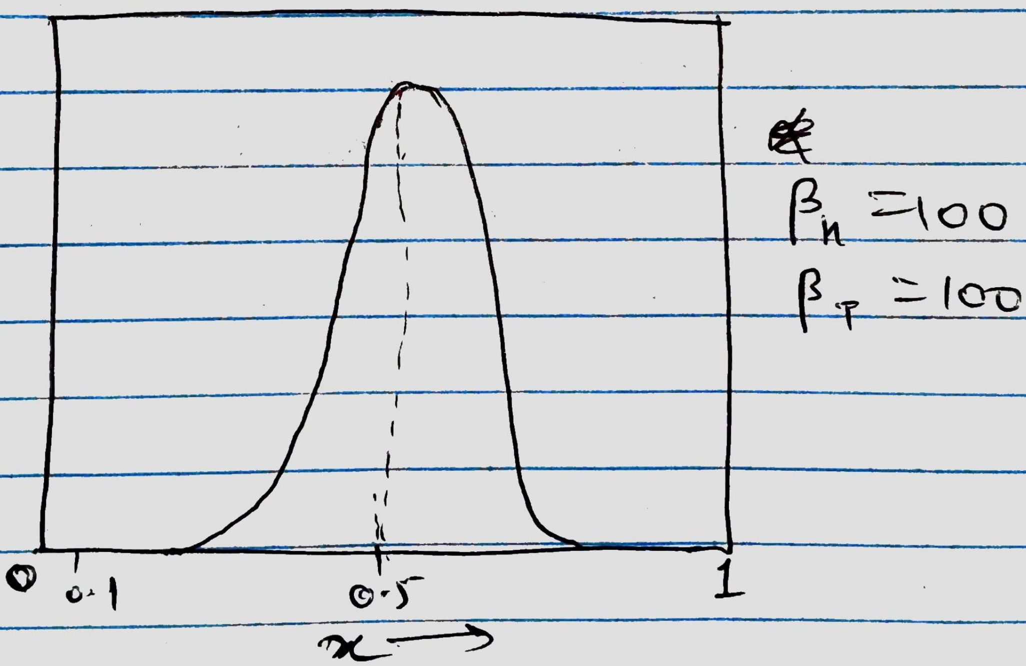
Scenario 2

Beta (100,100)

Priors → $\beta_H = 100$, $\beta_T = 100$
Data → $\alpha_H = 60$, $\alpha_T = 40$.
Posteriors → Beta (160, 140)

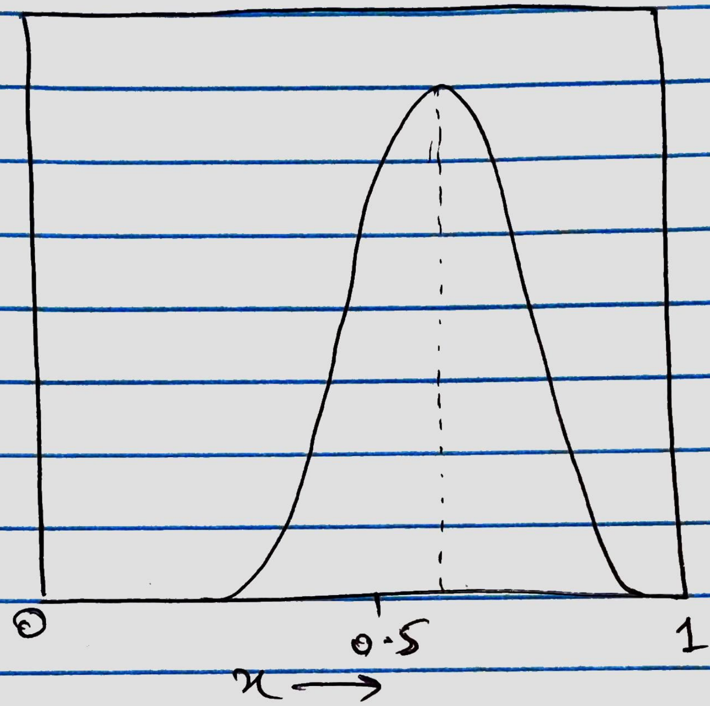
Creating a graph for this Scenario,

Prior Beta(100,100)



This prior graph is a little wide than the previous scenario. This is because the ~~β~~ Beta prior value has reduced from (1000, 1000) to (100, 100).

Posterior Beta(160, 140)



$\alpha_H + \beta_H = 160$
$\alpha_T + \beta_T = 140$

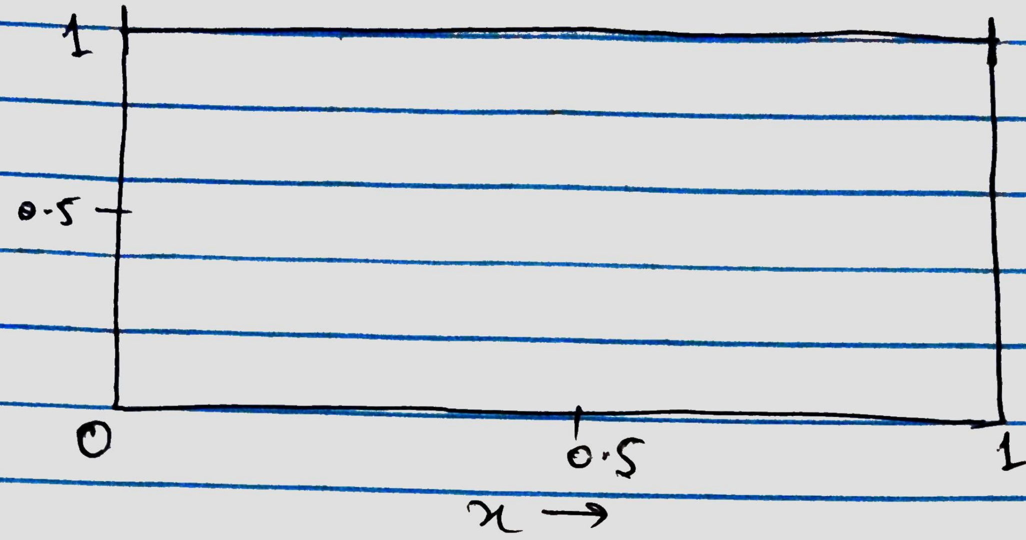
Here, the graph is similar to the prior graph, just a minor shift ~~from~~ of the peak from 0.5 to 0.53.
This is because, there is a difference between the posterior values as compared to the parameter values (x)

Scenario 3

Beta(1,1)

Priors $\longrightarrow$ $\beta_H = 1$, $\beta_T = 1$
Data $\longrightarrow$ $\alpha_H = 60$, $\alpha_T = 40$.
Posteriors $\longrightarrow$ Beta(61, 41).

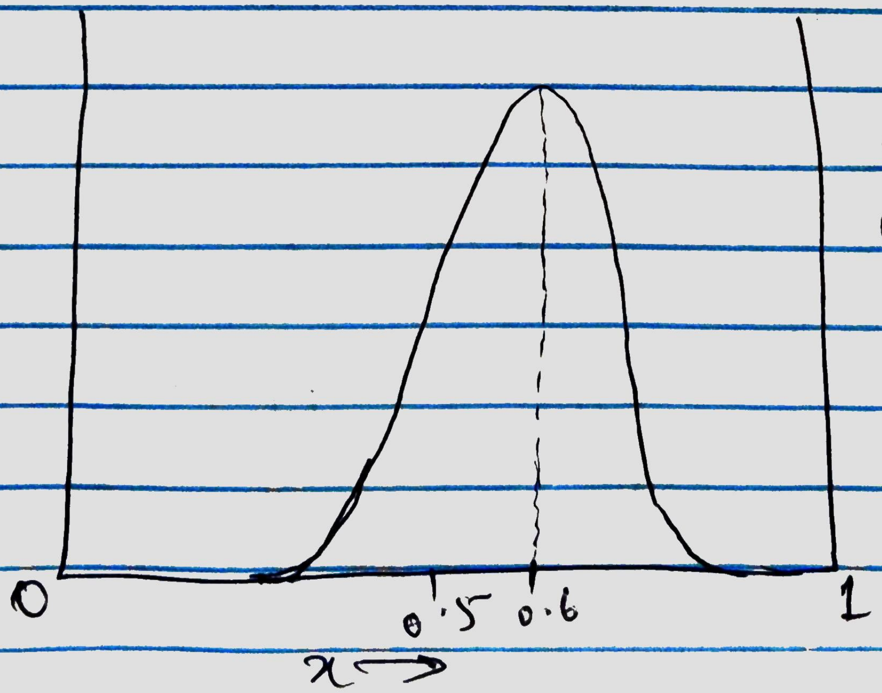
Graph — Beta (1,1) Prior.



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

In this prior graph, the parameter value is upto 1 and the priors given are also (1, 1).
Hence, the graph is a straight line with no interaction at x-axis and no slope. It ~~stats~~ starts from $y=1$ and ends at $x=1$.

Posterior Beta (61, 41)



$\alpha_H + \beta_T = 61$
$\alpha_T + \beta_T = 41$

The posterior graph for Beta(61, 41) has the peak over 0.6 instead of 0.5. The graph has shifted towards 1.
This is because the priors Beta(1,1) have decreased and the posteriors ~~(60, 40)~~ (61, 41) has considerable difference.

Hence, as the prior value decreases, the parameter values are affected.

(3) The MLE estimate of both coin and the thumbtack is same as well as MAP values are same for both coin and thumbtack.
This is because, the Beta priors are equal in all the scenarios. The graphs are same because the tosses have same results (60, 40) for both the coin and the thumbtack.
Hence, the answer is FALSE.

(4.) The MAP estimate for both coin and thumbtack are equal. This is because the toss results as well as Beta prior values are same for both coin and thumbtack.
Hence, the answer is FALSE.