Naman Garg
HW2

Answers to the point estimation problems in the homework 2

2. Point Estimation

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

Likelihood is the probability mass function for a random Variable x

$$f(x;P) = p^{x}(1-p)^{1-x}$$

Likelihood $f(x_1 --- x_n; P) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

Log Likelihood $\ln f(x_1 --- x_n; P) = \sum_{i=1}^{n} (x_i \ln p + (1-x_i)\ln(1-p))$

Take derivative and solve for p

$$\frac{\delta \ln f}{\delta p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n}(1-x_i)}{1-p} = 0$$

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

$$= \frac{\sum^{n} x_i}{p(1-p)} = \frac{n}{1-p}$$

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

2. parameter p, Binomial (N, p) sample size n

$$f(x) = \left(\frac{N!}{x!(N-x)!}\right) p^x (1-p)^{N-x}$$

$$f(x_1, \ldots, x_n; p) = \prod_{i=1}^{n} \left(\left(\frac{N!}{x_i!(N-x_i)!}\right) p^{x_i} (1-p)^{N-x_i} \right)$$

$$\ln f(x_1 \ldots x_n; p) = \ln\left(\prod_{i=1}^{n} \left(\frac{N!}{x_i!(N-x_i)!}\right)\right) + \sum_{i=1}^{n} x_i \ln p + \left(Nn - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

$$\frac{d \ln f}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(N.n - \sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

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

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

$$\boxed{\hat{p} = \frac{\sum_{i=1}^{n} x_i}{Nn}}$$

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

Sample size $= n = 6$

$$\sum_{i=1}^{6} x_i = 3 + 6 + 2 + 0 + 0 + 3$$

$$= 14$$

$$N \times n = 10 \times 6 = 60$$

$$\boxed{\hat{p} = \frac{\sum_{i=1}^{n} x_i}{Nn} = \frac{14}{60} = 0.2333}$$

3. parameters a, b ; Uniform (a, b) size n

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

$$f(x_1 \ldots x_n; a, b) = \prod_{i=1}^{n} \left(\frac{1}{b-a}\right) = \frac{1}{(b-a)^n}$$

$$\ln f(x_1 \ldots x_n; a, b) = n \ln \frac{1}{b-a} = -n \ln(b-a)$$

$$\frac{\delta \ln f}{\delta a} = \frac{n}{b-a} \rightarrow \text{strictly increasing}$$

$$\frac{\delta \ln f}{\delta b} = -\frac{n}{b-a} \rightarrow \text{strictly decreasing}$$

$$\hat{a} = \min(x_1 \ldots x_n), \quad \hat{b} = \max(x_1 \ldots x_n)$$

4. parameter μ, Normal (μ, σ^2), sample size n

unkown σ^2, unknown μ

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

$$f(x_1 \ldots x_n; \mu) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$$

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

$$\ln f(x_1 \ldots x_n; \mu) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

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

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

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

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

5. parameter σ, Normal (μ, σ^2)

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

$$f(x_1 -- x_n;\sigma) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

$$\ln(f(x_i -- x_n;\sigma)) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{d\ln f}{d\sigma} = -\frac{n}{\sigma} - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2}(-2\sigma^{-3}) = 0$$

$$\frac{n}{\sigma} = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3}$$

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

6 parameters (μ, σ^2) Normal (μ, σ^2) sample size n

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

$$f(x_1 -- x_n;\mu,\sigma^2) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\ln f(x_i -- x_n;\mu,\sigma^2) = \sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

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

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

$$\frac{\delta \ln f}{\delta \sigma^2} = \sum_{i=1}^{n}\left(-\frac{1}{2}\times\frac{1}{\sigma^2} - \frac{(x_i-\mu)^2(-1)}{2\sigma^4}\right) = 0$$

$$\frac{n}{2\sigma^2} = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^4}$$

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

(2) Coin and Thumtack, Coin - 60 head, 40 tails
thumtack - 70 head, 30 tails

priors Beta(1,1), Beta(40,60), Beta(30,70)
Beta(100,100), Beta(1000,1000), Beta(100,000, 100,000)

1) MLE and MAP for coin and thumbtack

Let θ be parameter to maximize probobity for D

$P(\text{Heads}) = \theta$, $P(\text{tails}) = 1-\theta \rightarrow \theta \in [0,1]$

α_H Heads, α_T tails, Data → data of sample obserbations

Likelihood $P(D|\theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T}$

Log likelihood $\ln P(D|\theta) = \alpha_H \ln\theta + \alpha_T \ln(1-\theta)$

$$\frac{d \ln P(D|\theta)}{d\theta} = \frac{\alpha_H}{\theta} - \frac{\alpha_T}{1-\theta} = 0 \Rightarrow \alpha_H - \theta\alpha_H = \theta\alpha_T$$

$$\boxed{\hat{\theta}_{MLE} = \frac{\alpha_H}{\alpha_H + \alpha_T}}$$

$\hat{\theta}_{MLE}\ \text{Coin} = \frac{60}{100} = 0.6$

$\hat{\theta}_{MLE}\ \text{thumtack} = \frac{70}{100} = 0.7$

MAP → $P(\theta|D)$ $\quad P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$ Bayes theorem

$\hat{\theta}_{MAP} \rightarrow \arg\max_{\theta} P(\theta|D)$

$= \arg\max_{\theta} P(D|\theta)P(\theta)$ $\quad$ P(D) doesn't depend on θ

We assume prior $P(\theta)$ is Beta Distribution as $\theta \in [0,1]$

$$P(\theta) = \frac{\theta^{\beta_H - 1}(1-\theta)^{\beta_T - 1}}{B(\beta_H, \beta_T)} \sim \text{Beta}(\beta_H, \beta_T)$$

$$P(D|\theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T}$$

$$P(\theta|D) = \theta^{\alpha_H}(1-\theta)^{\alpha_T} \times \frac{\theta^{\beta_H-1}(1-\theta)^{\beta_T-1}}{B(\beta_H, \beta_T)}$$

$$\ln P(\theta|D) = \alpha_H \ln\theta + \alpha_T \ln(1-\theta) + (\beta_H - 1)\ln\theta + (\beta_T - 1)\ln(1-\theta) + \ln B(\beta_H, \beta_T)$$

$$\frac{d \ln P(\theta|D)}{d\theta} = \frac{\alpha_H}{\theta} + \frac{\alpha_T}{1-\theta} + \frac{\beta_H - 1}{\theta} + \frac{\beta_T - 1}{1-\theta} = 0$$

$$\alpha_H - \theta\alpha_H + \theta\alpha_T + \beta_H - 1 + \theta\beta_H - \theta + \theta\beta_T - \theta = 0$$

$$\boxed{\hat{\theta}_{MAP} = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}}$$

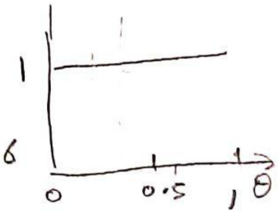
2) Coin Beta (1,1)

Head = 60 $\hat{\Theta}_{MLE} = 0.6$

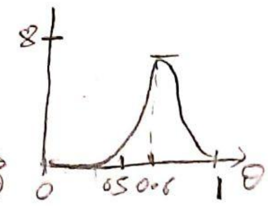
Tails = 40 $\hat{\Theta}_{MAP} = \frac{60+1-1}{100+2-2} = 0.6$

$\hat{\Theta}_{MLE} = \frac{\alpha_H}{\alpha_H + \alpha_T}$ $\hat{\Theta}_{MAP} = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$

beta(1,1)



beta(61,41)



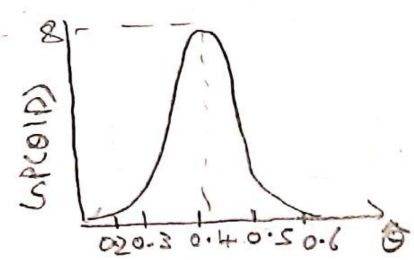
Beta (40,60)

$\hat{\Theta}_{MLE} = 0.6$

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

$= \frac{98}{198} = 0.5$

beta(40,60)

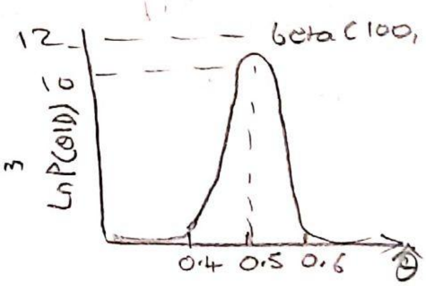


for B(1,1)

$\hat{\Theta}_{MLE} = \hat{\Theta}_{MAP}$

MLE is special case of MAP when prior is uniform

beta(100,100)



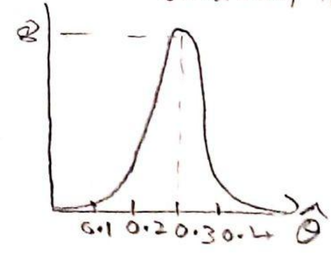
Beta (30,70)

$\hat{\Theta}_{MLE} = 0.6$

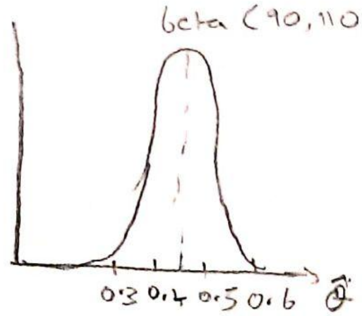
$\hat{\Theta}_{MAP} = \frac{60+30-1}{100+100-2}$

$= \frac{89}{198} = 0.45$

beta(30,70)



beta(90,110)



The stronger or larger the prior is the stronger the effect on curve due to the parameter

Thumbtack

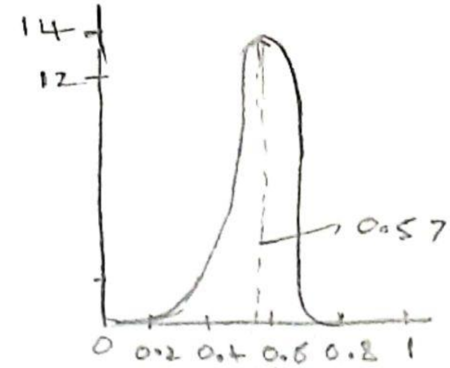
Heads = 70

Tails = 30

beta (100,100)

$\theta_{MLE} = 0.7$

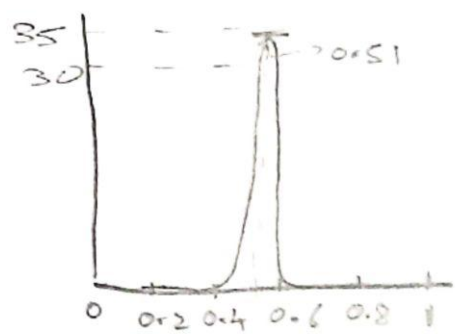
$$\theta_{MAP} = \frac{70+100-1}{100+200-2} = \frac{169}{298} = 0.57$$



beta (1000, 1000)

$\theta_{MLE} = 0.7$

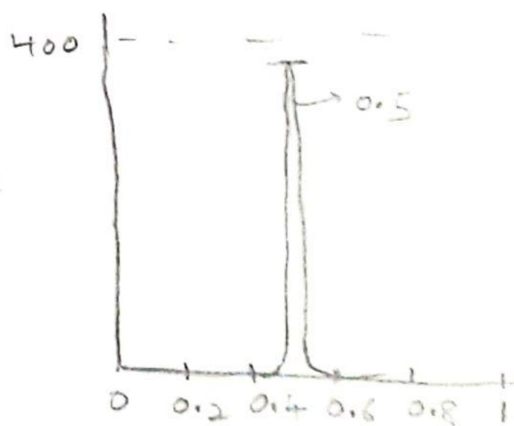
$$\theta_{MAP} = \frac{1000+70-1}{2000+100-2} = \frac{1069}{2098} = 0.51$$



beta (100,000, 100,000)

$\theta_{MLE} = 0.7$

$$\theta_{MAP} = \frac{100{,}000+70-1}{200{,}000+100-2} = \frac{100069}{200098} = 0.5$$



The answer to the above problem is that the larger the prior will be the more it will effect the value of the parameter. As seen in the above curves as the prior reaches the value beta(100,000, 100,000) the parameter ap

3) False, the MLE estimate will approach the MAP estimate just once when for $B(\alpha, \beta)$ $\alpha = 1$, $\beta = 1$.

4) True, when we use a Stronger prior the effect of the prior will be weak for Small sample size

eg Thumb~~s~~tack and coin the θ_{MAP} value for prior $B(100{,}000, 100{,}000)$ is same

coin $\theta_{MLE} = 60$ $\hat{\theta}_{MAP} = 0.5$

Thumblack $\theta_{MLE} = 70$ $\hat{\theta}_{MAP} = 0.5$