CS 436/580L Homework 2

NAME : SHWETA SHARAD MESTRY

BU NUMBER: B00815342

2. Point Estimation

Derive maximum likelihood estimators for

1.

Let X be the Bernoulli random variable with parameter p.

Let $x_1, x_2, x_3 \dots x_n$ be the independent random ~~variables~~ sample of X.

We want to determine MLE for the parameter p in Bernoulli(p) model with sample of size n.

P.D.F for Bernoulli distribution with parameter p is

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

$\therefore$ Therefor the likelihood function of sample is

$$f(x_1, x_2, x_3 \dots x_n, p) = \prod_{i=1}^{n} f(x_i, p)$$

($\because$ x_i are independent)

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

Scanned by CamScanner

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

Taking the natural Logarithm on both sides,

$$L(x_1, x_2, \ldots x_n, p) = \sum x_i \log(p) + (n - \sum x_i) \log(1-p)$$

As we have asked to find MLE of parameter p, we will take the partial derivative of the function with respect to p and setting it to 0,

$$\frac{\partial \log L(p)}{\partial p} = \frac{\sum x_i}{p} - \left(\frac{n - \sum x_i}{1-p}\right) = 0$$

Multiplying through by $p(1-p)$,

$$\sum x_i (1-p) - (n - \sum x_i) p = 0$$

$$\sum x_i - p\cancel{\sum x_i} - np + p\cancel{\sum x_i} = 0$$

$$\sum x_i - np = 0$$

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

2. Given, datapoint/random Binomial variables from observed sample is $(3, 6, 2, 0, 0, 3)$ and $N = 10$.

P.D.F for Binomial distribution with parameter p is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{where } x = 0, \ldots n$$

~~The~~ ∴ the likelihood function of the sample is

$$f(x_1 \dots x_m, p) = \prod_{i=1}^{m} f(x_i, p) \quad (x_i \text{ are independent})$$

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

Taking the natural logarithm on both sides,

$$\log L(x_1 \dots x_m, p) = \sum_{i=1}^{m} \log \binom{n}{x_i} + \sum_{i=1}^{m} x_i \log(p) + \left(mn - \sum_{i=1}^{m} x_i\right) \cdot \log(1-p)$$

The maximum value can be found by taking the partial derivative of the above function with p and setting it to 0. So,

$$\frac{\partial L(x_1 \dots x_m, p)}{\partial p} = 0 + \frac{\sum_{i=1}^{m} x_i}{p} - \frac{mn - \sum_{i=1}^{m} x_i}{1-p} = 0$$

Multiplying through by $p(1-p)$,

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

$$\sum_{i=1}^{m} x_i - \cancel{p \sum_{i=1}^{m} x_i} - pmn + \cancel{p \sum_{i=1}^{m} x_i} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{m} x_i}{mn}$$

Given observed sample variables are 3, 6, 2, 0, 0, 3 and $m = 6$, $N = 10$

$$\hat{p} = \frac{3+6+2+0+0+3}{6 \times 10} = \frac{14}{60} = 0.2333$$

$=$

3. Let X be the Uniform random variable with parameter a and b,
Let $x_1, \ldots x_n$ be the independent random sample of X drawn from a uniform distribution $U(a,b)$ of sample size n.

P.D.F for uniform distribution with parameter a and b. is

$$f(x) = \frac{1}{b-a} \quad \text{on } [a,b].$$

$\therefore$ The likelihood function of the sample is

$$f(x_1, \ldots x_n; a, b) = \begin{cases} \left(\frac{1}{b-a}\right)^n & \text{if all } x_i \text{ are in the interval } [a,b] \\ 0 & \text{otherwise} \end{cases}$$

This can be maximized by making b-a as small as possible with restriction that the interval $[a,b]$ must include all the data.

$\therefore$ MLE for the pair (a,b) is

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

4. P.D.f for Normal Distribution for known variance σ^2 and unknown mean μ with sample of size n. is:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

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

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

Taking the natural logarithm on the both the sides,

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

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

The maximum value can be found by taking the partial derivative of the above function with μ and setting it to 0, So,

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

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

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

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

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

5. P.D.F. for Normal Distribution for know mean μ and unknow variance σ^2 with sample of size n is:

$$f(x_i \mid \mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

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

Taking the natural logarithm on the both the sides,

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

The maximum value can be found by taking the partial derivative of above function with respect to σ^2 and setting it to 0.

$$\frac{\partial L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \left(\Sigma \frac{(x_i-\mu)^2}{2}\right)\frac{1}{(\sigma^2)^2} = 0$$

$\sigma^2 \rightarrow L(\mu, \sigma^2)$ is maximized when

$$\sigma^2 = \Sigma \frac{(x_i-\bar{x})^2}{n} = \overline{(x_i-\bar{x})^2}$$

{for $\mu = \bar{x}$}

$$\hat{\sigma} = \sqrt{\overline{(x_i-\bar{x})^2}}$$

6. Let data $x_1, x_2 \ldots x_n$ are drawn from a $N(\mu, \sigma^2)$ distribution, where μ and σ are unknown.

P.D.F for normal distribution is

$$f(x_i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$\because$ x_i are independent their joint pdf is the product of the individual pdf's is

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

Taking log on both sides,

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

$\because$ $L(\mu, \sigma^2)$ is a function of the two unknown variables μ, σ.

To find maximum value of $\hat{\mu}$, we will take partial derivative of eq. (I) w.r.t. μ and setting it 0. So

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

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

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} = \text{the mean of the data}$$

To find maximum value of $\hat{\sigma}^2$, we will take partial derivative of eq. (I) w.r.t. σ and setting it 0. So.

$$\frac{\partial L(\mu, \sigma^2)}{\partial \sigma} = \frac{-n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^3} = 0$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = \text{the variance of the data}$$

Coin and thumbtack problem

1. MLE for the thumtack :

α_H = # of Heads in 100 trials = 60

α_T = # of tails in 100 trials. = 40

$P(Heads) = \theta_1$, $P(Tails) = 1-\theta_1$

Tossing a thumbtack follows Binomial Distribution.

Let assume 'D' be the observed set of α_H Heads and α_T tails.

P.D.F of Binomial Distribution,

$$P(D|\theta_1) = \theta_1^{\alpha_H}(1-\theta_1)^{\alpha_T}$$

Choose θ_1 to maximize probability of D.

$$\hat{\theta}_1 = \arg\max_{\theta_1} P(D|\theta_1)$$

$$= \arg\max_{\theta_1} \log P(D|\theta_1)$$

$$= \arg\max_{\theta_1} \log\left(\theta^{\alpha_H}(1-\theta_1)^{\alpha_T}\right)$$

To find maximum value of θ_1 we will take partial derivative of above equation w.r.t. θ_1 and setting it to 0. So,

$$\frac{\partial L \cdot P(D|\theta_1)}{\partial \theta_1} = \frac{\partial}{\partial \theta_1}\left[\log \theta_1^{\alpha_H}(1-\theta_1)^{\alpha_T}\right] = 0$$

$$= \frac{\partial}{\partial \theta_1}\left[\alpha_H \log \theta_1 + \alpha_T \log(1-\theta_1)\right] = 0$$

$$= \alpha_H \frac{\partial}{\partial \theta_1} \log \theta_1 + \alpha_T \frac{\partial}{\partial \theta_1} \log(1-\theta_1) = 0$$

$$= \frac{\alpha_H}{\theta_1} - \frac{\alpha_T}{1-\theta_1} = 0$$

Multiplying by $\theta_1(1-\theta_1)$ on both sides,

$$(1-\theta_1)\alpha_H - \theta_1 \alpha_T = 0$$

$$\alpha_H - \theta_1 \alpha_H - \theta_1 \alpha_T = 0$$

$$\alpha_H - \theta_1(\alpha_H + \alpha_T) = 0$$

$$\hat{\theta}_1 = \frac{\alpha_H}{\alpha_H + \alpha_T}$$

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

$$\hat{\theta}_1 = \frac{60}{60+40} = \frac{60}{100} = \frac{3}{5} = 0.6$$

<u>MLE for the coin:</u>

r_H = # of Heads in 100 trials when coin is flipped = 60

r_T = # of tails in 100 trials when coin is flipped = 40

$P(\text{Heads}) = \theta_2$, $P(\text{Trials}) = 1 - \theta_2$

Flipping a coin is a binomial Distribution.
Let assume D_2 be observed set of r_H Heads and r_T tails,

P.D.F of binomial distribution,

$$P(D_2|\theta_2) = \theta_2^{r_H}(1-\theta_2)^{r_T}$$

Maximum probability of D

$$\arg\max_{\theta_2} P(D_2|\theta_2) = \arg\max_{\theta_2} \theta_2^{r_H}(1-\theta_2)^{r_T}$$

Taking the natural Logarithm on both the sides,

$$\arg\max_{\theta_2} \log P(D_2|\theta_2) = \arg\max_{\theta_2} \log \theta_2^{r_H}(1-\theta_2)^{r_T}$$

to find maximum value of θ_2, we will take partial derivative of above equation w.r.t. θ_2 and setting it to 0. So,

$$\frac{\partial LP(D_2|\theta_2)}{\partial\theta_2} = \frac{\partial}{\partial\theta_2} \log \theta_2^{r_H}(1-\theta_2)^{r_T} = 0$$

$$= \frac{\partial}{\partial\theta_2}\left[r_H \log\theta_2 + r_T \log(1-\theta_2) \right] = 0$$

$$= \frac{r_H}{\theta_2} - \frac{r_T}{(1-\theta_2)} = 0$$

Multiplying by $\theta_2(1-\theta_2)$ on both sides,

$$(1-\theta_2) r_H - \theta_2 r_T = 0$$

$$\hat{\theta}_2 = \frac{r_H}{r_H + r_T}$$

Given $r_H = 60$, $r_T = 40$

$$\hat{\theta}_2 = \frac{60}{60+40} = \frac{3}{5} = 0.6$$

MAP estimate for thumbtack:

Consider Beta prior Beta(1,1).

MAP uses most likely parameter:

Beta prior equivalent to extrae thumbtack flips.

$$\theta_{MAP} = \frac{\alpha_H + 1}{\alpha_H + \alpha_T + 2} = \frac{60+1}{100+2} = 0.598.$$

MAP estimate for Coin:

Consider Beta prior Beta(1,1).

$$\hat{\theta} = \frac{r_H + \beta_H - 1}{r_H + \beta_H + r_T + \beta_T - 2} = \frac{60+1-1}{60+1+40+1-2}$$

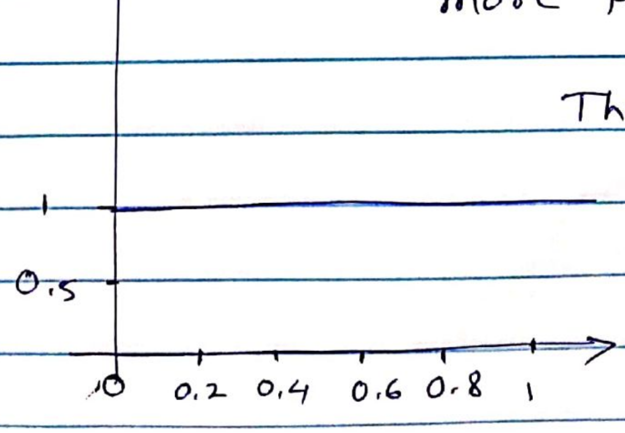
$$= \frac{60}{100} = 0.6$$

As $N \to \infty$, prior is "forgotten"

But, for small sample size, prior is important.

2. Beta(1,1) $\rightarrow$ equivalent to the uniform (0,1) distribution. The distribution becomes more peaked as α and β increases.

The modal is symmetric



3. True:

MLE estimate of bothe coin and thumbtack is the same but the MAP estimate is not same.

We have calculated MLE of coin ~~by~~ with assumption that coin is biased. Because we got 60 Heads in 100 flips. However we cannot estimate that the coin is biased if we do this experient more than 100 (or large N).

Therefore, we had to consider prior probability while calculating MAP.

~~example~~.

4. True: