Date:

MLE
MAP
Assignment

Chinmaya Udupa
SC19B081

MLE

1 A a) uniform $p(x|a) = \begin{cases} \frac{1}{a} & [x \in [0,a]] \\ 0 & \text{otherwise} \end{cases}$

$$\hat{a}_{ML} = \underset{a}{\operatorname{argmax}} \prod_{i=1}^{n} p(x_i|a)$$

$$\Rightarrow L(a|x) = \left(\frac{1}{a}\right)^n$$

$$l(a|x) = -n \ln(a)$$

$$\Rightarrow \boxed{\hat{a}_{ML} = \max(x_1, x_2 \ldots x_n)}$$

b) exponential $p(x|\eta) = \frac{1}{\eta} \exp\left(-\frac{x}{\eta}\right) \quad \eta > 0$

$$L(\eta|x_i) = \prod_{i=1}^{n} \frac{1}{\eta} \exp\left(\frac{-x_i}{\eta}\right)$$

$$l(\eta|x_i) = -n \ln(\eta) - \frac{1}{\eta} \sum_{i=1}^{n} x_i$$

$$\frac{\partial l}{\partial \eta} = \frac{-n}{\eta} + \frac{1}{\eta^2} \sum_{i=1}^{n} x_i = 0$$

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

$$\boxed{\hat{\eta}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

c) Gaussian

$$P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

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

substituting $\nu = \frac{1}{\sigma^2}$ $\quad \frac{\sqrt{\nu}}{\sqrt{2\pi}} \exp\left(-\frac{\nu}{2}(x-\mu)^2\right)$

$$l(\mu,\sigma|x_i) = \frac{n}{2}\ln\left(\frac{\nu}{2\pi}\right) - \frac{\nu}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

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

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

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

$$\frac{\partial l}{\partial \nu} = \frac{n}{\nu} - \sum_{i=1}^{n}(x_i-\hat{\mu})^2 = 0$$

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

B $\text{Bias} = E_\theta(\hat{\theta}) - \theta$

a) $\hat{a}_{ML} = \max(x)$

$\theta = a$

$\left[\mathbb{E}(\max(x)) = \max(x), \text{ as it's a single value}\right]$

it is obvious that $\max(x) \neq a$ (not necessary to occur all the time)

b) $\hat{\eta}_{ML} = \sum_{i=1}^{n} \frac{x_i}{n}$; $\theta = \eta$

$$\mathbb{E}\left(\sum_{i=1}^{n} \frac{x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(x_i) = \frac{1}{n}\sum_{i=1}^{n} \eta = \frac{\cancel{n}\eta}{\cancel{n}}$$

$\Rightarrow \mathbb{E}(\hat{\eta}) = \eta \Rightarrow \text{Bias} = 0$

c) $\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\theta = \mu$

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(x_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

$\Rightarrow \mathbb{E}(\hat{\mu}_{ML}) = \mu \Rightarrow \text{Bias} = 0$

$$\mathbb{E}\left(\hat{\sigma}^2_{MLE}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2\right)$$

$$= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2x_i\hat{\mu} + \hat{\mu}^2\right)$$

$$= -2\hat{\mu}^2$$

$$= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 + \hat{\mu}^2 - \sum_{i=1}^{n} 2x_i\hat{\mu}\right)$$

$$= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \hat{\mu}^2\right)$$

$$= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(x_i^2) - \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{E}(x_i^2) + \sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\mathbb{E}(x_i x_j)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(x_i^2) - \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{E}(x_i^2) + n(n-1)\mu^2\right)$$

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

$$= \frac{n-1}{n}\mathbb{E}(x_i^2) - \frac{n-1}{n}\mu^2$$

$$= \left[\sigma^2 = \mathbb{E}(x_i^2) - \mu^2\right]$$

$$= \frac{n-1}{n}\left(\sigma^2 + \mu^2 - \mu^2\right) \Rightarrow \boxed{\mathbb{E}\left(\hat{\sigma}^2_{MLE}\right) = \frac{n-1}{n}\sigma^2}$$

M T W T F S S

Page No.:

Date:

YOUVA

$\Rightarrow \text{Bias}(\hat{\sigma}^2_{MLE}) = \frac{(n-1)}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \Rightarrow$ biased estimator

multiplying $\hat{\sigma}^2_{MLE}$ with $\frac{n}{n-1} \Rightarrow \text{Bias} = 0$ $\left[\text{as } \mathbb{E}\left(\frac{n}{n-1}\hat{\sigma}^2_{MLE}\right) = \frac{n}{n-1}\mathbb{E}\left(\hat{\sigma}^2_{MLE}\right)\right.$

M	T	W	T	F	S	S

Page No.:
Date:
YOUVA

D $B(\hat{\theta}) = E(\hat{\theta}) - \theta \qquad V(\hat{\theta}) = E\left(\left(\hat{\theta} - E(\hat{\theta})\right)^2\right)$

$$MSE(\hat{\theta}) = E\left((\hat{\theta}-\theta)^2\right)$$

$$= E\left(\left((\hat{\theta}-E(\hat{\theta})) + (E(\hat{\theta})-\theta)\right)^2\right)$$

$$= E\left((\hat{\theta}-E(\hat{\theta}))^2\right) + (E(\hat{\theta})-\theta)^2 + 2E\left((\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\hat{\theta})\right)$$

$$= V(\hat{\theta}) + B^2(\hat{\theta}) + \underbrace{2E\left((\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\hat{\theta})\right)}_{0}$$

$\Rightarrow$ $\underline{MSE(\hat{\theta}) = B^2(\hat{\theta}) + V(\hat{\theta})}$

$$MSE(\hat{\sigma}^2_{ML}) = E\left((\hat{\sigma}^2_{ML} - \sigma^2)^2\right) \text{ or } B^2(\hat{\theta}) + V(\hat{\theta})$$

$$\boxed{B(\hat{\sigma}^2_{ML}) = \frac{-1}{n}\sigma^2}$$ [as shown earlier]

$$\boxed{V(\hat{\sigma}^2_{ML}) = E\left((\hat{\sigma}^2 - E(\hat{\sigma}^2))^2\right)}$$

$$= E\left(\left(\hat{\sigma}^2 - \frac{n-1}{n}\sigma^2\right)^2\right)$$

$$= E\left(\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 - \frac{(n-1)}{n}\sigma^2\right)^2\right) \boxed{= \frac{2\sigma^4(n-1)}{n^2}}$$

$$\boxed{V(\hat{\sigma}^2_{n-1}) = E\left(\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 - \sigma^2\right)^2\right)} \boxed{= \frac{2\sigma^4}{n-1}}$$

$$\boxed{B(\hat{\sigma}^2_{n-1}) = 0}$$

$$MSE(\hat{\sigma}^2_{ML}) = \left(-\frac{1}{n}\sigma^2\right)^2 + \frac{2\sigma^4(n-1)}{n^2} = \frac{\sigma^4}{n^2} + \frac{2\sigma^4(n-1)}{n^2}$$

$$= \frac{\sigma^4(2n-1)}{n^2}.$$

$$MSE(\hat{\sigma}^2_{n-1}) = 0^2 + \frac{2\sigma^4}{n-1} = \frac{2\sigma^4}{n-1}$$

$$MSE(\hat{\sigma}^2_{n-1}) - MSE(\hat{\sigma}^2_{ML}) = \frac{2\sigma^4}{n-1} - \frac{\sigma^4(2n-1)}{n^2}$$

$$\cancel{= \frac{\sigma^4((2n^2-2n+1)}{n^2(n-1)}}$$

$$\cancel{= 2\sigma^4}$$

$$= \frac{\sigma^4}{n^2(n-1)}\left[2n^2 - (2n-1)(n-1)\right]$$

$$= \frac{\sigma^4}{n(n-1)}$$

as $\sigma, n > 0$

$$\Rightarrow \boxed{MSE(\hat{\sigma}^2_{n-1}) > MSE(\hat{\sigma}^2_{ML})}$$

MAP

1. Given likelihood function, we have to find ML

$$L(\theta|D) = \theta^{n_H}(1-\theta)^{n-n_H}$$

taking log on both sides

$$l(\theta|D) = n_H \ln(\theta) + (n-n_H)\ln(1-\theta)$$

$$\frac{\partial l}{\partial \theta} = \frac{n_H}{\theta} - \frac{n-n_H}{1-\theta} = 0$$

$$\Rightarrow (1-\theta)n_H = \theta(n-n_H)$$

$$\Rightarrow \boxed{\hat{\theta}_{ML} = \frac{n_H}{n}}$$

2 $$p'(\theta) = \begin{cases} 0.5 & \theta = 0.5 \\ 0.5 & \theta = 0.4 \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\arg\max}\; p(D|\theta)\,p'(\theta)$$

$$p(D|\theta) = \theta^{n_H}(1-\theta)^{n-n_H}$$

$$\Rightarrow p(\theta|D) \propto \begin{cases} 0.5\left(\frac{1}{2}\right)^{n_H}\left(\frac{1}{2}\right)^{n-n_H} & @\,\theta = 0.5 \\ 0.5\left(\frac{2}{5}\right)^{n_H}\left(\frac{3}{5}\right)^{n-n_H} & @\,\theta = 0.4 \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow p(\theta|D) \propto \begin{cases} 0.5 \times \left(\frac{1}{2}\right)^{n} \\ \frac{0.5 \times}{} \left(\frac{2}{3}\right)^{n_H} \left(\frac{3}{5}\right)^{n} \end{cases}$$

$\underset{\theta}{\operatorname{argmax}} (p(\theta|D)) = 0.5$ if

$$\left(\frac{1}{2}\right)^{n} > \left(\frac{2}{3}\right)^{n_H} \left(\frac{3}{5}\right)^{n}$$

$$\left(\frac{5}{6}\right)^{n} > \left(\frac{2}{3}\right)^{n_H}$$

$$n \ln\left(\frac{5}{6}\right) > n_H \ln\left(\frac{2}{3}\right)$$

inequality flips because $\ln(5/6)$ & $\ln(2/3)$ are negative

$$\text{if } \frac{n_H}{n} > \frac{\ln(5/6)}{\ln(2/3)} = 0.449 \quad \hat{\theta}^{1}_{MAP} = 0.5$$

$$\text{else } \frac{n_H}{n} < \frac{\ln(5/6)}{\ln(2/3)} = 0.449 \text{ then } \hat{\theta}^{1}_{MAP} = 0.4$$

3 $p(\theta|D) \propto \theta^{n_H} (1-\theta)^{n-n_H} \times \theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{n_H+\alpha-1} (1-\theta)^{n-n_H+\beta-1}$

taking ln on both sides and maximizing it

$$l(\theta|D) = (n_H + \alpha - 1) \ln\theta + (n - n_H + \beta - 1) \ln(1-\theta)$$

$$\frac{\partial l}{\partial \theta} = \frac{n_H + \alpha - 1}{\theta} = \frac{n - n_H + \beta - 1}{1-\theta}$$

$$\Rightarrow \hat{\theta}^{2}_{MAP} = \frac{n_H + \alpha - 1}{n + \alpha + \beta - 2}$$

M T W T F S S

Page No.:

Date:

YOUVA

as n_H & n increase, α, β & constants ~~do~~ won't change ~~your~~ ~~pos~~ estimate significantly

as $n_H, n \to \infty$ $\quad \hat{\theta}_{MAP} \simeq \hat{\theta}_{ML}$

4. if $\theta = 0.41$

then according to $\hat{\theta}'_{MAP}$, $\quad \hat{\theta}'_{MAP} = 0.4$

$\hat{\theta}'_{MAP}$ will learn faster but will not converge to the right answer, $\hat{\theta}_{ML}$ learns slower but with sufficient data, will reach θ_{actual}

Hence $\hat{\theta}_{ML}$ is a better estimator (given enough data

MLE and MAP on Gaussian

1. Given $f(x|\theta) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$

where σ^2 is known, likelihood function is defined by

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$l(\theta|x) = \sum_{i=1}^{n} \ln\left(f(x_i|\theta)\right)$$

$$l(\theta|x) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right)$$

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

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

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

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

Q. $f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\sigma - \frac{1}{2\sigma^2}(x-\mu)^2\right)$

we are interested in maximizing $f(\mu|D)$

i.e, $f(D|\mu)\, f(\mu)$ should be maximize

given $f(\mu) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{1}{2\beta^2}(\mu - \nu)^2\right)$

$\cancel{\Rightarrow f(\mu|x) \propto \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 - \frac{1}{2\beta^2}(\mu-\nu)^2\right)}$

where $f(D|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)$

$$\Rightarrow f(D|\mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$\Rightarrow f(\mu|D) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\beta^2}(\mu-\nu)^2\right)$$

maximizing $f(\mu|D)$ over μ gives us $\hat{\mu}_{MAP}$

$$\boxed{\text{arg max}(\exp(f(x))) \equiv \text{arg} \exp(\text{arg max}(f(x)))}$$

as exponential is monotonic

$$\underset{\mu}{\operatorname{argmax}}\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\beta^2}(\mu-\nu)^2\right)$$

$$\equiv \underset{\mu}{\operatorname{argmax}}\left(-\left(\beta^2\sum_{i=1}^{n}(x_i-\mu)^2 + \sigma^{2}(\mu-\nu)^2\right)\right)$$

given that the above term is equal to

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

differentiating the above term WRT μ and equate it to 0

$$2\left[\mu\sqrt{n\beta^2+\sigma^2} - \frac{\sigma^2\nu+\beta^2\sum_{i=1}^{n}x_i}{\sqrt{n\beta^2+\sigma^2}}\right]\sqrt{n\beta^2+\sigma^2} = 0$$

$$\Rightarrow \mu\sqrt{n\beta^2+\sigma^2} = \frac{\sigma^2\nu+\beta^2\sum_{i=1}x_i}{\sqrt{n\beta^2+\sigma^2}}$$

$$\Rightarrow \hat{\mu}_{MAP} = \frac{\sigma^2\nu+\beta^2\sum_{i=1}x_i}{n\beta^2+\sigma^2}$$

$$= \boxed{\frac{\sigma^2\nu+n\beta^2\hat{\mu}_{ML}}{n\beta^2+\sigma^2}}$$

where $\hat{\mu}_{ML}$ as derived earlier $= \frac{\sum_{i=1}^{n}x_i}{n}$

Date:

3. as $n \to \infty$

$$\lim_{n \to \infty} \hat{\mu}_{MAP} = \lim_{n \to \infty} \left(\frac{\sigma^2 \nu + n\beta^2 \hat{\mu}_{ML}}{n\beta^2 + \sigma^2} \right)$$

$$= \cancel{\hat{\mu}_{MAP}} \; \frac{\cancel{n\beta^2}\, \hat{\mu}_{ML}}{\cancel{n\beta^2}}$$

$$\boxed{\lim_{n \to \infty} \hat{\mu}_{MAP} = \hat{\mu}_{ML}}$$

Parameter Estimation

1. a) $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

$$L(\lambda|x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$l(\lambda|x_i) = \sum_{i=1}^{n} x_i \ln(\lambda) - \lambda - \ln(x_i!)$$

$$l(\lambda|x_i) = \ln(\lambda)\sum_{i=1}^{n}(x_i) - \lambda n - \sum_{i=1}^{n}\ln(x_i!)$$

$$\frac{\partial l}{\partial \lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0$$

$$\rightarrow \boxed{\hat{\lambda}_{ML} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

b) $\mathbb{E}(\hat{\lambda}_{ML}) = \mathbb{E}\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(x_i) \qquad \mathbb{E}(x_i) = \lambda$$

$$= \frac{1}{n}\sum_{i=1}^{n}\lambda = \frac{n\lambda}{n} = \underline{\underline{\lambda}}$$

as $\mathbb{E}(\hat{\lambda}_{ML}) = \lambda \Rightarrow$ unbiased estimator

2 & 3 $P(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$ is the given prior with α & β known to us

to find $\hat{\lambda}_{MAP}$

we need to maximize ~~$P(\lambda|D) P(\lambda)$~~ $p(\lambda|D)\,p(\lambda)$

where we have calculated $p(\lambda|D)$ to be equal to

$$\frac{\lambda^{\sum x_i} e^{-\lambda n}}{\prod_{i=1}^{n} x_i!}$$

$$\Rightarrow p(D|\lambda) \propto \lambda^{\sum x_i} e^{-\lambda n} \times \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\Rightarrow p(D|\lambda) \propto \lambda^{\sum x_i + \alpha - 1} e^{-\lambda(\beta+n)}$$

maximizing the log of this function

$$l(D|\lambda) \propto \left(\sum x_i + \alpha - 1\right) \ln(\lambda) - \lambda(\beta+n)$$

$$\frac{\partial l}{\partial \lambda} = \frac{\sum x_i + \alpha - 1}{\lambda} - (\beta+n) = 0$$

$$\hat{\lambda}_{MAP} = \frac{\sum x_i + \alpha - 1}{\beta + n}$$

M T W T F S S

Page No.:

Date:

YOUVA

4) Given N samples & known variance, $\hat{\mu}_{MLE}$ as estimated previously question 1

c) $\hat{\mu}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$

5) ~~see~~ again, we have seen this in question 2

a) $\hat{\mu}_{MAP} = \frac{\sigma^2 \mu_p + \sigma_p^2 \sum x_i}{\sigma^2 + n \sigma_p^2}$

~~6)~~