Assignment #5

ABHINAV KUMAR
180050003

BHASKAR GUPTA
180050022

Q.1

MLE estimate $\hat{\mu}_{MLE} = \Sigma x_i / n$

i) Gaussian prior

$$P(\mu|data) = \frac{P(data|\mu)P(\mu; 10.5, 1)}{\int P(data|\mu)P(\mu; 10.5, 1)\,d\mu}$$

MAP Estimate: $\frac{\partial}{\partial\mu} P(\mu|data) = 0 \equiv \frac{\partial}{\partial\mu}\log[P(\mu|data)] = 0$

$$\log P(\mu|data) \propto \cancel{\sum_i e^{-(x_i-\mu)^2/2\sigma^2}} \; \cancel{+}$$

$$\sum_i \frac{(x_i-\mu)^2}{2\sigma_{true}^2} + \frac{(\mu-10.5)^2}{2(1)}$$

$$\frac{\partial}{\partial\mu}\log P(\mu|data) = \frac{\Sigma(x_i-\mu)}{\sigma_{true}^2} + \frac{(10.5-\mu)}{(1)} = 0$$

$$\Rightarrow \frac{n\mu}{\sigma_{true}^2} + \mu = 10.5 + \frac{\Sigma x_i}{\sigma_{true}^2}$$

$$\mu = \frac{10.5 + \frac{\Sigma x_i}{\sigma_{true}^2}}{\frac{n}{\sigma_{true}^2} + 1}$$

ii) Uniform prior

$$P(\mu|data) = \frac{P(data|\mu) \times (1/2)}{\int_\mu P(data|\mu) \times 1/2\, d\mu} \rightarrow e^{-\Sigma(x_i-\mu)^2/2\sigma_{true}^2}$$

$$\Rightarrow \frac{\partial}{\partial\mu}\log P(\mu|data) = 0 \Rightarrow \mu = \Sigma x_i/n$$

But since μ has a uniform prior in $(9.5, 11.5)$, and Likelihood function is parabolic,
if $\bar{x} < 9.5$, MAP estimate in $(9.5, 11.5)$ is 9.5
if $\bar{x} > 11.5$, MAP estimate in $(9.5, 11.5)$ is 11.5
else MAP estimate is $\bar{x}$.

b) As N increases, the relative error for all the three estimates of μ converges to that of MLE estimate. And, Gaussian prior boxplot has the minimum variance amongst all three estimate as the posterior mean which is product of two Gaussians,

has σ^{2*} (variance) $= \dfrac{(\sigma^2)(\sigma_0^2/n)}{\sigma^2 + \sigma_0^2/n} < \left(\dfrac{\sigma_0^2}{n}\right)$ [ML estimator Variance]

Hence, Gaussian prior is preferable.

Q2) $y = -\frac{1}{\lambda} \log x.$

$g(x) = -\frac{1}{\lambda} \log x.$

$x = e^{-\lambda y}$

$\therefore g^{-1}(y) = e^{-\lambda y}$

$p(y) = \left| q(g^{-1}(y)) \frac{d}{dy} (g^{-1}(y)) \right|$

$p(y) = \lambda e^{-\lambda y}$

$P(Y/\lambda) = \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i}$

To find $\hat{\lambda}_{ML}$ we will differentiate $P(Y/\lambda)$ w.r.t. λ.

$\frac{d}{d\lambda} P(Y|\lambda) = 0.$

$N\lambda^{N-1} e^{-\lambda \sum y_i} - (\sum y_i) \lambda^N e^{-\lambda \sum y_i} = 0.$

$\hat{\lambda}_{ML} = \frac{N}{\sum y_i}$

Scanned by CamScanner

$$P(\lambda|Y) = \frac{P(Y|\lambda)P(\lambda)}{\int_0^\infty P(Y|\lambda)P(\lambda)d\lambda}$$

$$= \frac{\lambda^N e^{-\lambda\sum y_i} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_0^\infty \lambda^N e^{-\lambda\sum y_i} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda}$$ (ignored the constant as they will be cancelled)

$$= \frac{\lambda^{N+\alpha-1} e^{-\lambda(\beta+\sum y_i)}}{\frac{\Gamma(N+\alpha)}{(\sum y_i+\beta)^{N+\alpha}}}$$

$$E(\lambda) = \int_0^\infty \frac{\lambda . \lambda^{N+\alpha-1} e^{-\lambda(\beta+\sum y_i)}}{\frac{\Gamma(N+\alpha)}{(\sum y_i+\beta)^{N+\alpha}}} d\lambda$$

$$= \frac{N+\alpha}{\sum y_i+\beta}$$ We can find it by using the property $\Gamma(x+1) = x\Gamma(x)$

$$\boxed{\begin{aligned} \lambda_{\text{posterior mean}} &= \frac{N+\alpha}{\sum y_i+\beta} . \\ \hat{\lambda}_{ML} &= \frac{N}{\sum y_i} \end{aligned}}$$

Posterior mean for small N. M.L.E for larger N

The posterior mean estimate is more accurate for small N. As N increase MLE becomes more efficient.

P.M.E. Tends to M.L.E as $N \to \infty$ and has less variance.

Posterior is more reliable for limited data, but as N increases weight of the biased prior causes M.L.E. to be better.

DATE / / PAGE

(Q3) Points are of the form $(r_0 \cos\theta, r_0 \sin\theta)$.

$$E[r_0\cos\theta, r_0\sin\theta] = [0,0].$$

$$C = E\left(\begin{bmatrix} r_0\cos\theta \\ r_0\sin\theta \end{bmatrix} [r_0\cos\theta \;\; r_0\sin\theta]\right)$$

$$= E\begin{bmatrix} r_0^2\cos^2\theta & r_0^2\sin\theta\cos\theta \\ r_0^2\sin\theta\cos\theta & r_0^2\sin^2\theta \end{bmatrix}$$

$$C = \begin{bmatrix} r_0^2/2 & 0 \\ 0 & r_0^2/2 \end{bmatrix}$$

The theoritical and empirical mean and covariance matrix are closely correlated.

The modal point is theoritically $(0,0)$ but no data point is present there.

We can even find the modal radius:

Scanned by CamScanner

$$P(r,\theta) = \frac{1}{2\pi\left(\frac{r_0^2}{2}\right)} e^{\frac{-r^2}{r_0^2}} r\,dr\,d\theta$$

$$P(r) = \frac{1}{\left(\frac{r_0^2}{2}\right)} e^{-r^2/r_0^2} r\,dr$$

$$\frac{dP(r)}{dr} = 0$$

$$\Rightarrow \cancel{e^{-r^2/r_0^2}} \quad 1 - \frac{2r^2}{r_0^2} \cancel{e^{-r^2/r_0^2}} = 0$$

$$r = \frac{r_0}{\sqrt{2}} \text{ (modal radius)}$$

The Gaussian model is not satisfactory as The modal radius is at $1/\sqrt{2}$ times the distance of the data

Scanned by CamScanner

$$\text{Likelihood } (L) = \prod_{i=1}^{N} p(x_i ; u ; C)$$

$$L = \frac{1}{\left(2\pi\sqrt{|C|}\right)^N} e^{-\frac{1}{2}\sum_i \left((x_i - u)^T C^{-1} (x-u)\right)}$$

~~logL~~ Taking log and differentiating w.r.t. u we get.

$$-\frac{1}{2} \sum_i \frac{d}{du} \left[(x_i - u)^T C^{-1} (x_i - u) \right] = 0.$$

$$\sum_i C^{-1} (x_i - u) = 0.$$

$$\hat{u} = \frac{\sum_i x_i}{N}$$

$$\frac{\partial}{\partial C} \left(\log |C| + \frac{1}{2} \sum_i (x_i - u)^T C^{-1} (x_i - u) \right) = 0.$$

$$\hat{C} = \frac{\sum_i (x_i - \hat{u})(x_i - \hat{u})^T}{N}$$

$$\hat{C} = \frac{\sum_i x_i x_i^T}{N} - \hat{u}\hat{u}^T.$$

Q.3 c) Yes. They match the theoretical values

$\hat{Cov} = \begin{bmatrix} .5 & -.0007 \\ -.0007 & .5 \end{bmatrix}$ $C_{the} = \begin{bmatrix} .5 & 0 \\ 0 & .5 \end{bmatrix}$

$\hat{\mu} = \begin{bmatrix} .0016 \\ .0008 \end{bmatrix}$ $\mu_{theo} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

N = 100000

Q.4

a) $\hat{\Theta}_{ML}$

~~1~~ likelihood function $f(x;\Theta) = \frac{1}{\Theta^n} \rightarrow$ decreasing function,

$\Theta \geq \max\{x_1 .. x_n\}$

$\Rightarrow$ Max-likelihood estimate is $\boxed{\hat{\Theta}_{ML} = \max\{x_1 - - x_n\}}$

$P(\Theta|x) \propto P(x|\Theta) \times P(\Theta)$

$P(x|\Theta) \propto \frac{1}{\Theta^n}$, $\Theta \geq \max\{x_1 - - x_n\}$

otherwise 0.

$P(\Theta) \propto \left(\frac{\Theta_m}{\Theta}\right)^\alpha$ for $\Theta \geq \Theta_m$

0 otherwise

$\Rightarrow P(x|\Theta) P(\Theta) \propto \frac{1}{\Theta^{n+\alpha}}$, $\Theta \geq \max(\max\{x_1 - - x_n\}, \Theta)$

0 otherwise

$\Rightarrow$ Posterior is also Pareto with parameters $(n+\alpha, \max\{x_1 - - x_n, \Theta\})$

MAP estimate $\boxed{\hat{\Theta}_{MP} = \max\{x_1 - - x_n, \Theta\}}$

[Same reasoning as MLE case, $P(\Theta|x)$ is decreasing].

b) $\hat{\theta}_{ML} = \max\{x_1 \ldots x_n\}$

$\hat{\theta}_{MP} = \max\{x_1 \ldots x_n, \theta_m\}$

As sample size tends to infinity & by the prior we know that $\hat{\theta}_1 > \theta_m$, there will be x_i's $> \theta_m$ (more chances of finding) & hence

$\hat{\theta}_{ML} = \max\{x_1 \ldots x_n\} = \max\{x_1 \ldots x_n, \theta_m\}$

(since, $x_i > \theta_m$, both equalities are the same).

$\Rightarrow \hat{\theta}_{MAP} \rightarrow \hat{\theta}_{ML}$ or rather $\hat{\theta}_{ML} = \hat{\theta}_{MAP}$ if samples are perfectly random.

And above is desirable as MLE estimate has asymptotically lowest MSE.

c) Posterior $P(\theta|x)$ is also proportional to $\text{Pareto}(n+\alpha, \underbrace{\max\{x_1 \ldots x_n, \theta\}}_{\theta_m'})$

$$\Rightarrow E_{P(\theta|x)}[\hat{\theta}] \propto \int_{\theta_m'} \frac{(\theta_m')^{n+\alpha}}{(\theta)^{n+\alpha}} \theta \, d\theta$$

$$\propto (\theta_m')^{n+\alpha} \cdot \frac{1}{(n+\alpha-2)(\theta_m')^{n+\alpha-2}} \propto \frac{(\theta_m')^2}{n+\alpha-2}$$

More rigorous →

$$P(\theta|x) = c \cdot \frac{\frac{1}{\theta^n} \times \left(\frac{\theta_m}{\theta}\right)^\alpha}{\int \frac{1}{\theta^n} \times \frac{(\theta_m)^\alpha}{\theta^\alpha} \, d\theta} \qquad \theta \geq \underbrace{\max\{x_1 \ldots x_n, \theta\}}_{\theta_m'}$$

$$= \frac{c}{\theta^{n+\alpha}} \times \frac{1}{\int_{\theta_m'}^{\infty} \frac{d\theta}{\theta^{n+\alpha}}} = \frac{c/\theta^{n+\alpha}}{\frac{1}{(n+\alpha-1)(\theta_m')^{n+\alpha-1}}}$$

$$\Rightarrow E_{p(\theta|x)}[\hat{\theta}] = \int_{\theta_m'} \frac{c}{\theta^{n+\alpha}} \times \frac{(\theta_m')^{n+\alpha-1}}{\cancel{\theta}} \times (n+\alpha-1)\, \theta \, d\theta$$

$$= \frac{(c)(\theta_m')^{n+\alpha-1}\,(n+\alpha-1)}{(\theta_m')^{n+\alpha-2}\,(n+\alpha-2)}$$

$$= \boxed{(c)(\theta_m')\frac{(n+\alpha-1)}{(n+\alpha-2)}}$$

$$\theta_m' = \max\{\theta_m, x_1 \dots x_n\}$$

d) a

$$\lim_{n\to\infty} (c)\,\theta_m' \frac{(n+\alpha-1)}{(n+\alpha-2)} = c\,\theta_m'$$

The exact value of $\hat{\theta}_{PM}$ depends on coefficient in Pareto but asymptotically, $\hat{\theta}_m'$ also tends to $\hat{\theta}_{ML}$ by the same reasoning as in (b). This is desirable as we want our estimate to have minimum MSE & MLE has minimum estimate.