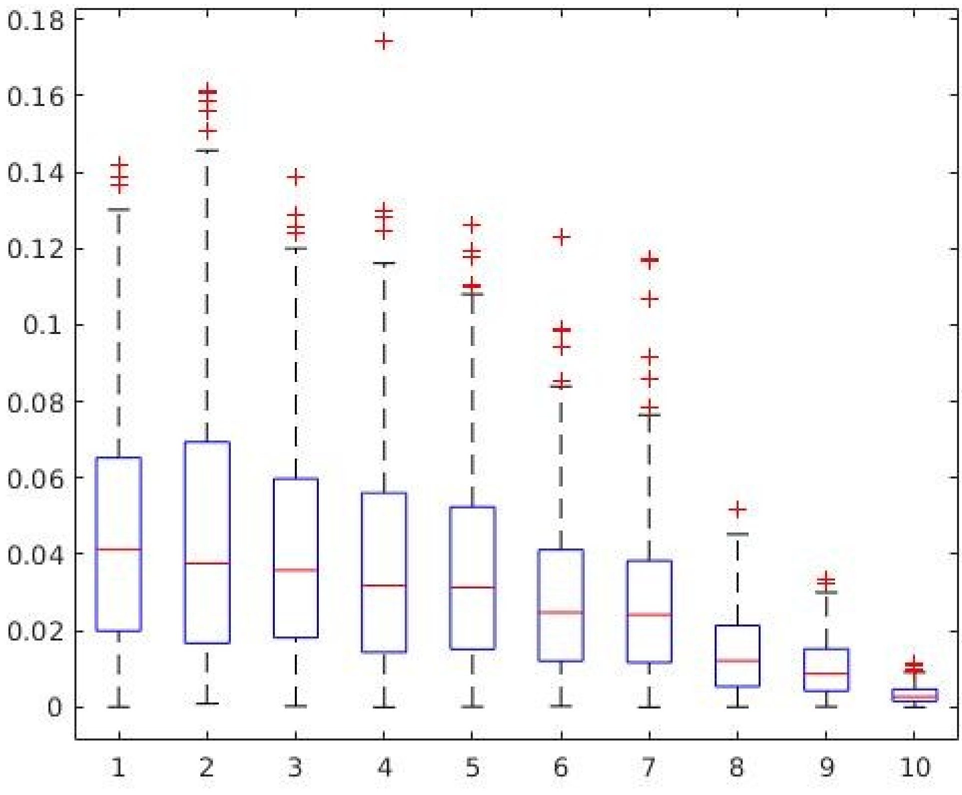
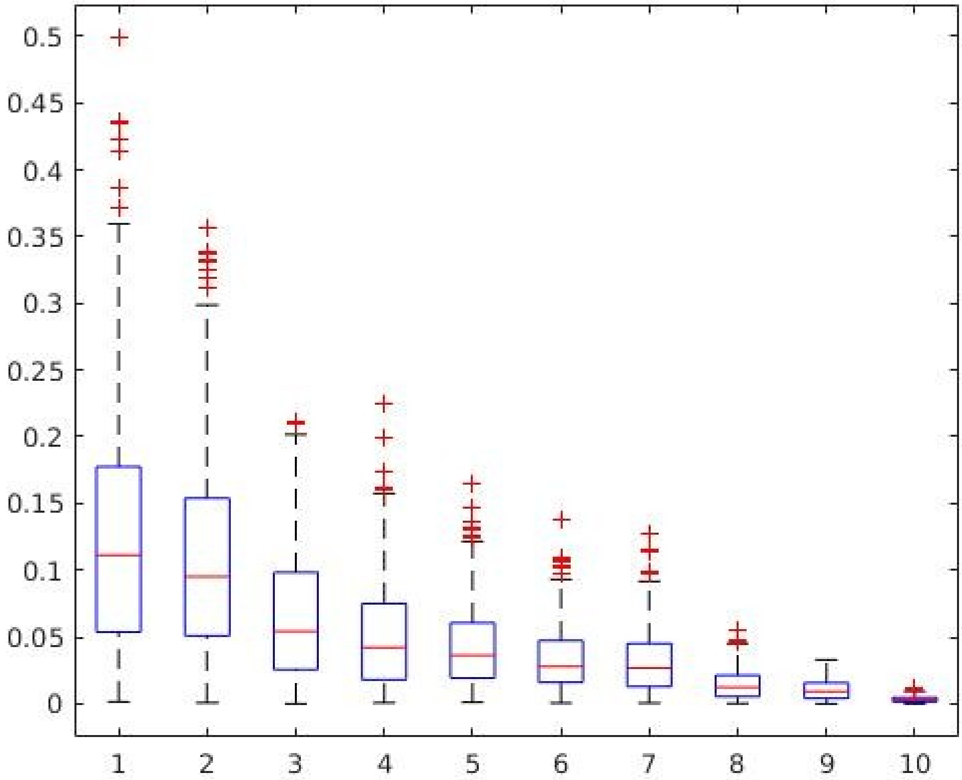
ASSIGNMENT-3

170050021 Abhijeet Singh Yadav
170050029 Devansh Garg

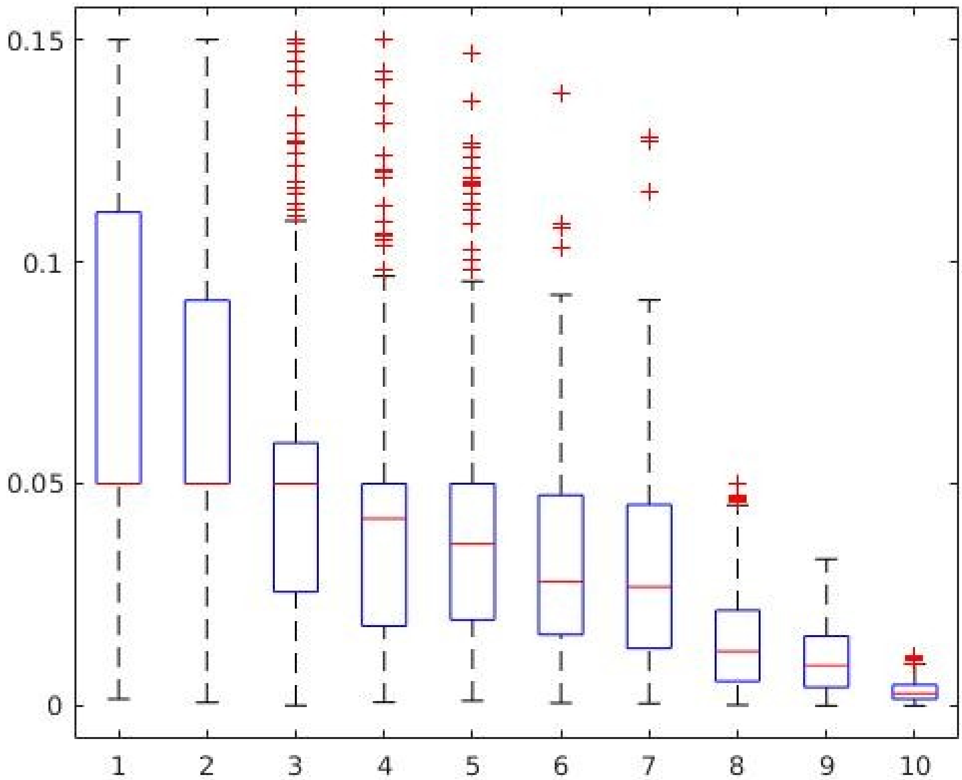
Q1.



Relative error with MAP-1



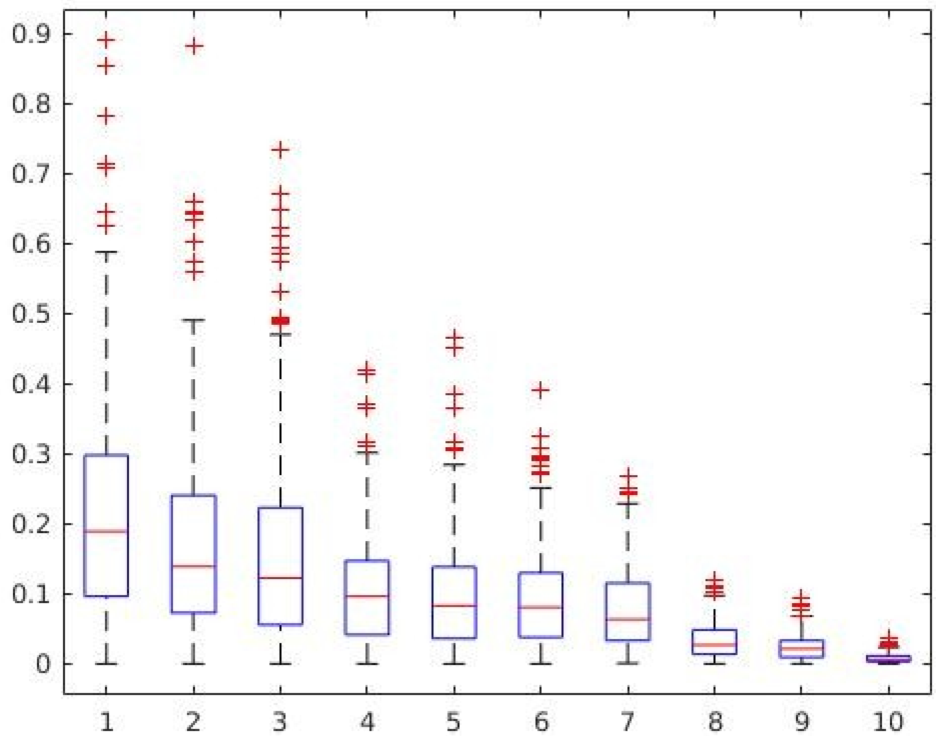
Relative error with ML



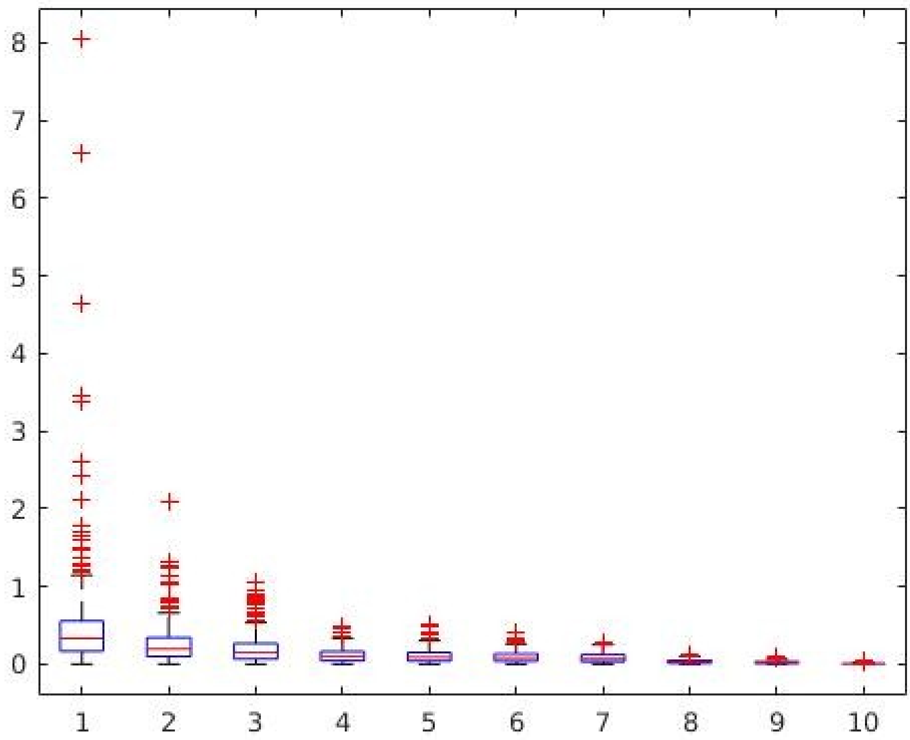
Relative error with MAP-2

Interpretation:
1. As N tends to infinity the Error becomes zero for all three estimators which shows their consistency
2. Gaussian Prior is preferred for small values of N because its error values are centered around zero and less scattered as it has small boxes
And for large values MLE is preffered as it has lowest mean squared error , consistency and aymptotic normality.

Q2.



Relative error when posterior mean estimator is used



Relative error with MLE

Interpretation:

1. As N tends to infinity the Error becomes zero for both the estimators which shows both estimators are consistent.
2. Posterior mean estimator is preferred for small values of N because its error values are centered around zero and less scattered as it has small boxes

Q.

$$\text{Prior} = P(\lambda) = \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}$$

now, for the CDF of 'y'

$$F_Y(y) = P(Y \leq y)$$

$$= P\left(\frac{-1}{\lambda}\log x \leq y\right)$$

$$= P(\log x \geq -\lambda y)$$

$$= P(x \geq e^{-\lambda y})$$

$$= 1 - P(x < e^{-\lambda y})$$

$$= 1 - e^{-\lambda y} \quad \{\text{As 'x' is of the form } (0,1)\}$$

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

$$P(data/\lambda) = \text{liklihood} = \lambda^n e^{-\lambda\sum_{i=1}^{n} y_i} \quad \left\{\prod_{i=1}^{n} \lambda e^{-\lambda y_i}\right\}$$

for M.L.E.

$$\frac{\partial(\text{liklihood})}{\partial\lambda} = 0$$

$$\Rightarrow n\lambda^{n-1}e^{-\lambda\sum y_i} - \sum y_i \lambda^n e^{-\lambda\sum y_i} = 0$$

$$\Rightarrow n = \sum y_i \lambda$$

$$\Rightarrow \hat{\lambda}^{ML} = \frac{n}{\sum_{i=1}^{n} y_i}$$

Probability of data can be calculated as

$$P(\text{data}) = \int_0^\infty P(\text{data}|\lambda)\, P(\lambda)\, d\lambda$$

$$= \int_0^\infty \lambda^n e^{-\lambda \Sigma y_i} \cdot \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}\, d\lambda$$

then,

$$\hat{\lambda}^{\text{Posterior mean}} = E_{P(\theta|\text{data})}[\theta] \quad \theta : \text{parameter}$$

$$= \int_0^\infty \lambda \frac{P(\text{data}|\lambda) P(\lambda)}{P(\text{data})}\, d\lambda$$

$$= \frac{\int_0^\infty \lambda \cdot \lambda^n e^{-\lambda\Sigma y_i} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}\, d\lambda}{\int_0^\infty \lambda^n e^{-\lambda\Sigma y_i} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}\, d\lambda}$$

As $P(\text{data})$ is constant w.r.t. λ so is β^α and $\Gamma\alpha$

$$= \frac{\int_0^\infty \lambda^{n+\alpha} e^{-\lambda(\beta+\Sigma y_i)}\, d\lambda}{\int_0^\infty \lambda^{n+\alpha-1} e^{-\lambda(\beta+\Sigma y_i)}\, d\lambda}$$

Apply integration by parts

$$\hat{\lambda}^{\text{Posterior}}_{\text{mean}} = \frac{\left(\left[\frac{e^{-\lambda(\Sigma y_i+\beta)}}{-(\Sigma y_i+\beta)} \cdot \lambda^{n+\alpha} \cancel{d\lambda} \right]_0^\infty - \int_0^\infty \frac{(n+\alpha)\,\lambda^{n+\alpha-1} e^{-\lambda(\Sigma y_i+\beta)}}{-(\Sigma y_i+\beta)}\, d\lambda \right)}{\int_0^\infty \lambda^{n+\alpha-1} e^{-\lambda(\Sigma y_i+\beta)}\, d\lambda}$$

The first Term in numerator goes to zero

So,

$$\hat{\lambda}^{\text{Posterior}}_{\text{mean}} = \frac{(n+\alpha)}{\Sigma y_i+\beta} \cdot \frac{\cancel{\int_0^\infty \lambda^{n+\alpha-1} e^{-\lambda(\Sigma y_i+\beta)} d\lambda}}{\cancel{\int_0^\infty \lambda^{n+\alpha-1} e^{-\lambda(\Sigma y_i+\beta)} d\lambda}}$$

$$= \frac{n+\alpha}{n\bar{y}+\beta}$$

Q3.

Q3:

$$P(x;\mu,C) = \frac{1}{\sqrt{2\pi|C|}} e^{\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right)}$$

The likelihood is

$$f(X|\mu,C) = \frac{1}{(\sqrt{2\pi|C|})^n} e^{-\left(-\frac{1}{2}\sum_i (x_i-\mu)^T C^{-1}(x_i-\mu)\right)}$$

the log likelihood

$$\log(f(X|\mu,C)) = -\frac{n}{2}\log(2\pi|C|) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T C^{-1}(x_i-\mu)$$

$\therefore$ $\frac{\partial \omega^T A \omega}{\partial \omega} = 2A\omega$ if 'ω' dosen't depend on A ($A \equiv$ symmetric)

which is the case here as Covariance matrix C is symmetric

So,

$$\frac{\partial \log(f(X|\mu,C))}{\partial \mu} = \frac{-1}{2}\sum_{i=1}^{n}(-1).2.\,C^{-1}(x_i-\mu)$$

$$= \frac{1}{2}\sum_{i=1}^{n}\left(2C^{-1}(x_i-\mu)\right)$$ For MLE it goes to zero

$\therefore C \neq 0$

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

$$\Rightarrow \left[\hat{\mu}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}\right]$$

as we draw value from a ring as x being $\begin{bmatrix}\mu\cos\theta \\ \mu\sin\theta\end{bmatrix}$

$$\hat{\mu} = \frac{\sum_{i=1}^{n}\begin{bmatrix}\mu\cos\theta \\ \mu\sin\theta\end{bmatrix}}{n} = 0$$

as for a large data set every θ will have such θ' that

$(\theta = \theta + \pi)$ $\cos\theta = -\cos\theta'$, $\sin\theta = -\sin\theta'$ and they become zero

$\therefore$ $(x_i-\mu)^T C^{-1}(x_i-\mu)$ will be a scalar it can be equated to the trace of a 1×1 matrix having $(x_i-\mu)^T C^{-1}(x_i-\mu)$ as element

$$(x_i-\mu)^T C^{-1}(x_i-\mu) = \text{trace}\left((x_i-\mu)^T C^{-1}(x_i-\mu)\right)$$

so,

$$\frac{\partial \log f(X|\mu,C)}{\partial C^{-1}} = \frac{\partial}{\partial C^{-1}}\left[-\frac{n}{2}\log 2\pi - \frac{n}{2}\log|C| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T C^{-1}(x_i-\mu)\right]$$

$$= \frac{\partial}{\partial C^{-1}}\left(\frac{n}{2}\log|C^{-1}| - \frac{1}{2}\sum_{i=1}^{n}\underbrace{(x_i-\mu)^T C^{-1}(x_i-\mu)}_{\text{equivalent to trace}\left((x_i-\mu)^T C^{-1}(x_i-\mu)\right)}\right] \quad \text{—(1)}$$

now,

$$\therefore \frac{\partial \log|A|}{\partial A} = A^{-T}$$

and, $x^TAx = \text{trace}(x^TAx) = \text{trace}(xx^TA)$ [Property of cyclic trace]

so,

$$\frac{\partial}{\partial A}\text{tr}[AB] = B^T$$

so $\frac{\partial}{\partial A}\text{tr}(C^{-1}xx^T) = x^+x$

hence eqn (1) reduces to

$$= \frac{nC}{2} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T \quad \text{(using cyclic trace)}$$

equating to zero,

$$|\hat{C}| = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$$

$$= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}\mu\cos\theta \\ \mu\sin\theta\end{bmatrix}[\mu\cos\theta \ \ \mu\sin\theta]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}\mu^2\cos^2\theta & \mu^2\sin\theta\cos\theta \\ \mu^2\sin\theta\cos\theta & \mu^2\sin^2\theta\end{bmatrix}$$

$avg(\cos^2\theta)$ over $[0, 2\pi] = 1/2$

$avg(\sin\theta\cos\theta)$ over $[0, 2\pi] = 0$

so, $C = \frac{1}{2}\begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix}$ as $\cancel{N} \to \infty$ ~~all the~~

(b) In a gaussian

mode = mean

so here, mode = (0,0)

but all the points are $(r\cos\theta, r\sin\theta)$ which ~~together~~ can't be $(0,0)$ as $r \neq 0$

→ This doesn't fit the given data well as the mode value must be something on the boundary point

→ The applied model is not correct hence the model is not good. Also points ~~all~~ are equidistant from mean so they must have the same probability so gaussian model is of no use here.

① In a unit radius circle for 10^5 points

mean = $(0.00013, -0.00060)$ theoretical value = $(0,0)$

(Almost equal)

Covariance matrix = $\begin{bmatrix} 0.5001 & -0.0036 \\ 0.0019 & 0.4909 \end{bmatrix}$ theoretical value = $\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$

(Almost matches)

Q4.

Q.4.

$$\therefore \quad X \sim U(0,\theta)$$

$$f_X(x) = \frac{1}{\theta} \quad 0<x<\theta$$

$$0 \quad \text{otherwise}$$

$$f_X(x_1,x_2,\dots x_n;\theta) = \frac{1}{\theta^n} \quad 0<x_1,x_2\dots,x_n<\theta$$

$$0 \quad \text{otherwise}$$

and for that $\theta \geq \max(x_1,x_2\dots x_n)$

$$\Rightarrow \theta_{min} = \max(x_1,x_2,\dots x_n)$$

$$\hat{\theta}_{ML} = \max(x_1,x_2\dots x_n)$$

for ~~Posterior~~ MAP estimate

$$P(\theta|data) = \frac{P(data|\theta)\,P(\theta)}{P(data)}$$

$$= \begin{cases} C\left(\frac{1}{\theta^n}\right)\left(\frac{\theta m}{\theta}\right)^{\alpha} & \theta \geq x_1,x_2,\dots x_n,\theta m \\ 0 & \text{otherwise} \end{cases}$$

using the desired Pareto-distribution - Prior

so,

$$\hat{\theta}_{MAP} = \max(x_1,x_2\dots x_n,\theta m)$$

$$P(\text{data}) = \int_{\hat{\theta}_{MAP}}^{\infty} P(\text{data}|\theta)\,P(\theta)\,d\theta$$

$$= \int_{\hat{\theta}_{MAP}}^{\infty} C \left(\frac{1}{\theta}\right)^{n} \left(\frac{\theta m}{\theta}\right)^{\alpha} d\theta$$

$$\hat{\theta}_{\substack{\text{Posterior}\\ \text{mean}}} = E_{P(\theta|\text{data})}[\theta]$$

$$= \int_{\hat{\theta}_{MAP}}^{\infty} \theta \, \frac{P(\text{data}|\theta)P(\theta)}{P(\text{data})} \, d\theta$$

$$= \frac{\displaystyle\int_{\hat{\theta}_{MAP}}^{\infty} \theta \, \cancel{C} \left(\frac{1}{\theta}\right)^{n} \left(\frac{\theta m}{\theta}\right)^{\alpha} d\theta}{\displaystyle\int_{\hat{\theta}_{MAP}}^{\infty} \cancel{C} \left(\frac{1}{\theta}\right)^{n} \frac{\theta m^{\alpha}}{\theta^{\alpha}} \, d\theta}$$

$$= \frac{\displaystyle\int_{\hat{\theta}_{MAP}}^{\infty} \theta^{1-n-\alpha} \, d\theta}{\displaystyle\int_{\hat{\theta}_{MAP}}^{\infty} \theta^{-n-\alpha} \, d\theta}$$

$$= \frac{\left. \dfrac{\theta^{2-n-\alpha}}{2-n-\alpha} \right|_{\hat{\theta}_{MAP}}^{\infty}}{\left. \dfrac{\theta^{1-n-\alpha}}{1-n-\alpha} \right|_{\hat{\theta}_{MAP}}^{\infty}} = \frac{\left(\hat{\theta}_{MAP}\right)^{2-n-\alpha} (1-n-\alpha)}{\left(\hat{\theta}_{MAP}\right)^{1-n-\alpha} (2-n-\alpha)}$$

$$= \left(\hat{\theta}_{MAP}\right) \left(\frac{n+\alpha-1}{n+\alpha-2}\right)$$

Part (b) and (d).

(b) Let θ^{true} be the true value of θ

If $\theta^{true} > \theta_m$ then $\hat{\theta}_{MAP} \to \hat{\theta}_{MLE}$ as $n \to \infty$

because X_i will cover whole range of $[0, \theta]$ and there will be some X_i in $(\theta_m, \theta^{true})$

$$\Rightarrow \hat{\theta}_{MAP} \to \hat{\theta}_{MLE}$$ [Desirable] good estimator for large n

otherwise, if $\theta^{true} < \theta_m$

$$\Rightarrow \theta_m > \max(X_1, X_2 \dots X_n)$$

$$\Rightarrow \hat{\theta}_{MAP} = \theta_m$$

and $\hat{\theta}_{MAP} \not\to \hat{\theta}_{MLE}$ as it will always be more than the former

[Not desirable]

(d)
$$\hat{\theta}_{\substack{posterior\\mean}} = \hat{\theta}_{MAP}\left(\frac{n+\alpha-1}{n+\alpha-2}\right)$$

as $n \to \infty$

$$\hat{\theta}_{\substack{posterior\\mean}} \to \hat{\theta}_{MAP}$$

again like in Part (b) due to similar reason

if $\theta_m < \theta^{true}$ $\quad \hat{\theta}_{P.m.} \to \hat{\theta}_{MLE}$ $\quad$ desirable

$\theta_m > \theta^{true}$ $\quad \hat{\theta}_{P.m.} \not\to \hat{\theta}_{MLE}$ $\quad$ undesirable.