REPORT FOR QUESTION 1

1 COMPARING ESTIMATES

1.1 ML Estimate

Let X denote the vector of N samples. For a given mean μ, the likelihood is:

$$P(X|\mu) = \prod_{i=1}^{N} P(x_i|\mu) = \prod_{i=1}^{N} G(x_i, \mu, \sigma^2)$$

This product is proportional to $G(\mu, \bar{x} = \frac{1}{N}\sum_i x_i, \frac{\sigma^2}{N})$, and hence is maximised at the mean.

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

1.2 MAP Estimate

The posterior distribution is given as:

$$P(\mu|X) = \frac{G(\mu, \bar{x}, \frac{\sigma^2}{N})P(\mu)}{\int G(\mu, \bar{x}, \frac{\sigma^2}{N})P(\mu)d\mu}$$

For the MAP estimate, denominator is irrelevant as it does not depend on μ. We need to equate derivative of numerator (with respect to μ) to 0.

1.2.1 Gaussian Prior

We have $P(\mu) = G(\mu, \mu_0, \sigma_0) = G(\mu, 10.5, 1)$. Product of two Gaussians is also a Gaussian. The mean of the resulting Gaussian (which is also the MAP Estimate) is given as:

$$\mu = \frac{\bar{x}\sigma_0^2 + \mu_0\frac{\sigma^2}{N}}{\sigma_0^2 + \frac{\sigma^2}{N}}$$

1.2.2 Uniform Prior

We have

$$P(\mu) = \begin{cases} 0.5 & 9.5 \leq \mu \leq 11.5 \\ 0 & otherwise \end{cases}$$

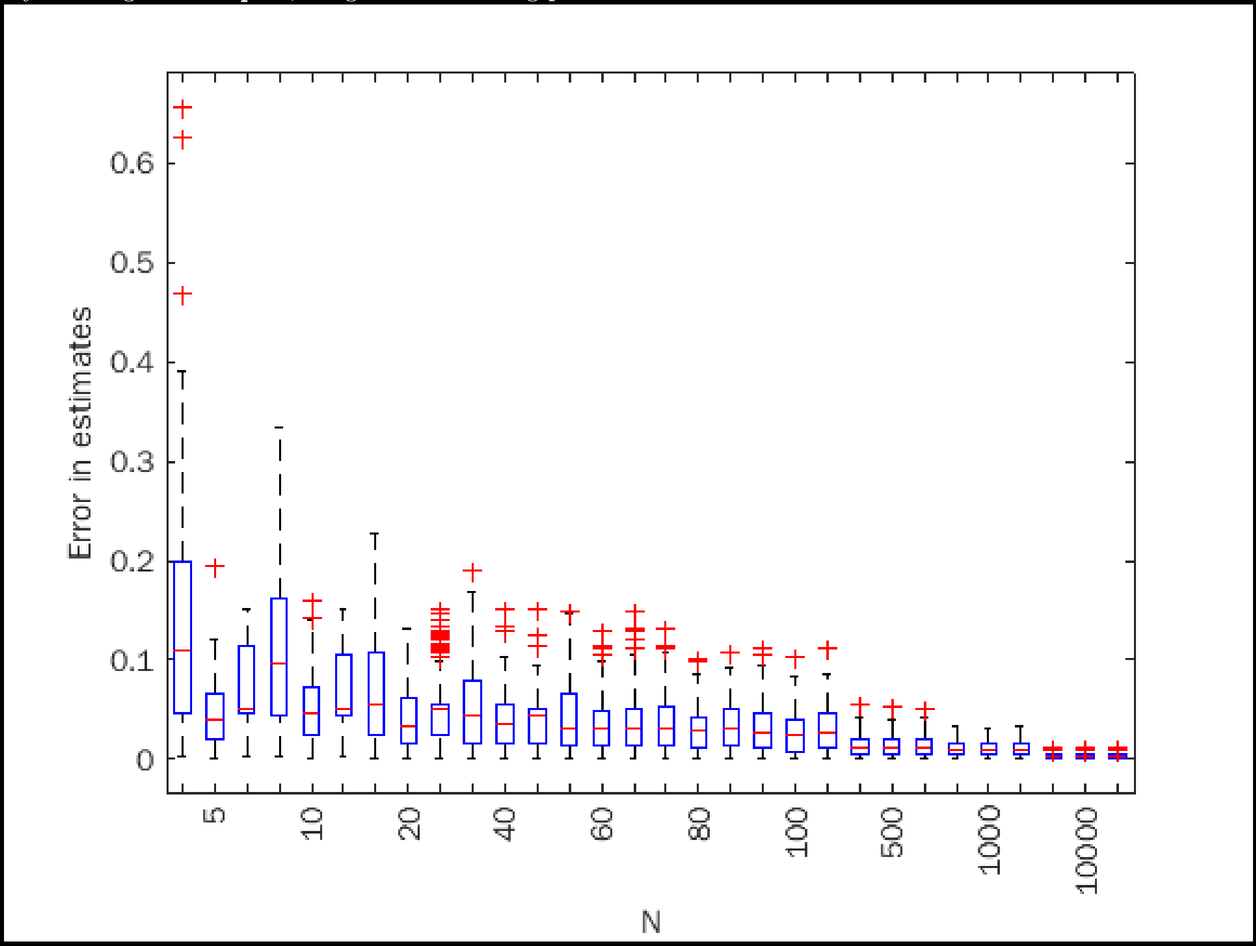
This case is almost similar to ML Estimate, since the prior distribution is constant. The difference is that the MAP Estimate will not go beyond the range where the prior is non-zero. Hence we will set the closest value as the estimate in such cases, since that will lead to maximum value.

$$\mu = \begin{cases} 9.5 & \bar{x} < 9.5 \\ \bar{x} & 9.5 \leq \bar{x} \leq 11.5 \\ 11.5 & 11.5 < \bar{x} \end{cases}$$

.

1.3 Error Plot

By running the file `q1.m`, we got the following plot:



1.4 Interpretation

The general observation is that the MAP estimate is more accurate than ML estimate. For lower sample size, the prior information helps in reducing the variation in the possible values that can be taken by the estimate. Although as N increases, the error nearly vanishes for all 3 estimates and they converge to the true value.

Among the 3, the estimate having the Gaussian prior (central plot for each value of N) gives the best estimate. The Uniform prior estimate (right plot for each N) is slightly better than the ML estimate (left plot for each N). Due to uneven (non-uniform weights) distribution given by the Gaussian prior, the error in the second estimate is relatively the least. Due to Uniform prior giving equal importance to the whole range, it carries lesser information of the actual value.

Q 2 (Report)

1) Analytical form of Transformed data.

so in question given $\boxed{y = \left(-\frac{1}{\lambda}\right) \times \log(x)}$

given the initial RV $X \sim U[0,1]$ generated by unifrnd( ) function in matlab
Now we know that when a RV is transformed to Y the distribution $q(Y)$ is given by.

$$\boxed{q(y) = p(f^{-1}(y)) \left| \frac{d}{dy} f^{-1}(y) \right|} \quad \text{——— } (1)$$

solving it further $\quad Y = f(x)$

hence $\quad Y = \frac{-1}{\lambda} \log(x)$

$$-\lambda Y = \log(x)$$

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

hence

$$f^{-1}(Y) = X$$

$\therefore \quad \boxed{f^{-1}(y) = e^{-\lambda y}}$

putting value back in eq (1) the analytical form $q(y) =$

ANS = $\boxed{q(y) = \lambda e^{-\lambda y}}$

2) MLE ($\hat{\lambda}_{ML}$)

We are given sample with varying size N, so in general the likelihood function for N data points $(x_1, x_2, x_3, x_4, \ldots x_N)$ is given by

$$\boxed{P(x_1, x_2, x_3, x_4 \ldots x_N \mid \lambda) = \lambda^N e^{\left(-\lambda \sum_i x_i\right)}} \quad \text{——— } (2)$$

To simplify the expression define $\sum_i x_i = W$, Now differentiating the log of the likelihood function to MLE

$$\frac{d}{d\lambda}\left(\log\left(\lambda^{N}\cdot e^{(-\lambda W)}\right)\right)\Big|_{\lambda=\hat{\lambda}^{MLE}} = 0$$

$$\frac{d}{d\lambda}\left(N\log(\lambda) + (-\lambda W)\right)\Big|_{\lambda=\hat{\lambda}^{MLE}} = 0$$

$$\Rightarrow \left(\frac{N}{\lambda^{MLE}} - W\right) = 0$$

hence
$$\boxed{\hat{\lambda}^{MLE} = \frac{N}{W} = \frac{N}{\sum_i x_i}}$$

In Code directly using the result in line No 38.

3) Posterior Mean ($\hat{\lambda}^{Posterior\,M}$)

Given we are with a prior distribution of gamma (PDF) for λ

$$\boxed{f_{Prior}(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}} \quad \text{——— (3)}$$

we know that posterior distribution $P(\lambda/X) = \frac{\text{likelihood} \times \text{Prior}}{\text{Data}}$

$$P(\lambda/X) = \frac{P(X/\lambda)\,P(\lambda)}{\int_\lambda P(X/\lambda)P(\lambda)\,d\lambda}$$

putting in the values from eq (2) & (3)

$$P(\lambda/X) = \frac{\frac{\beta^{\alpha}}{\Gamma(\alpha)} \times \lambda^{\alpha-1}e^{-\beta\lambda} \times \lambda^{N} \times e^{-\lambda W}}{\int_\lambda \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\lambda^{N}e^{-\lambda W}}$$

$$\boxed{P(\lambda/X) = \frac{\lambda^{N+\alpha-1}e^{-\lambda(\beta+W)}}{\int_\lambda \lambda^{N+\alpha-1}e^{-\lambda(\beta+W)}}}$$

We already know that the posterior mean

$$\hat{\lambda}^{Posterior\,Mean} = \int_0^\infty \lambda\,P(\lambda/X)\,d\lambda$$

$$\hat{\lambda}_{\text{Posterior Mean}} = \frac{\int_0^\infty \lambda^{N+\alpha-1} e^{-\lambda(\beta+w)}}{\int_0^\infty \lambda^{N+\alpha-1} e^{-\lambda(\beta+w)}}$$

Using the Result that

$$\boxed{I = \int_0^\infty \lambda^a e^{(-b\lambda)} d\lambda = \frac{\Gamma(a+1)}{b^{(a+1)}}}$$

which can be proven by substitution $u = \lambda b$

$$\boxed{I = \int_0^\infty \left(\frac{u}{b}\right)^a e^{(-u)} \frac{du}{b} = \frac{1}{b^{a+1}} \int_0^\infty u^a e^{(-u)} du = \frac{\Gamma(a+1)}{b^{a+1}}}$$

Using the above result we get.

$$\hat{\lambda}_{\text{Posterior Mean}} = \frac{\Gamma(N+\alpha+1)}{\Gamma(N+\alpha)} \times \frac{(w+\beta)^{N+\alpha}}{(w+\beta)^{N+\alpha+1}}$$

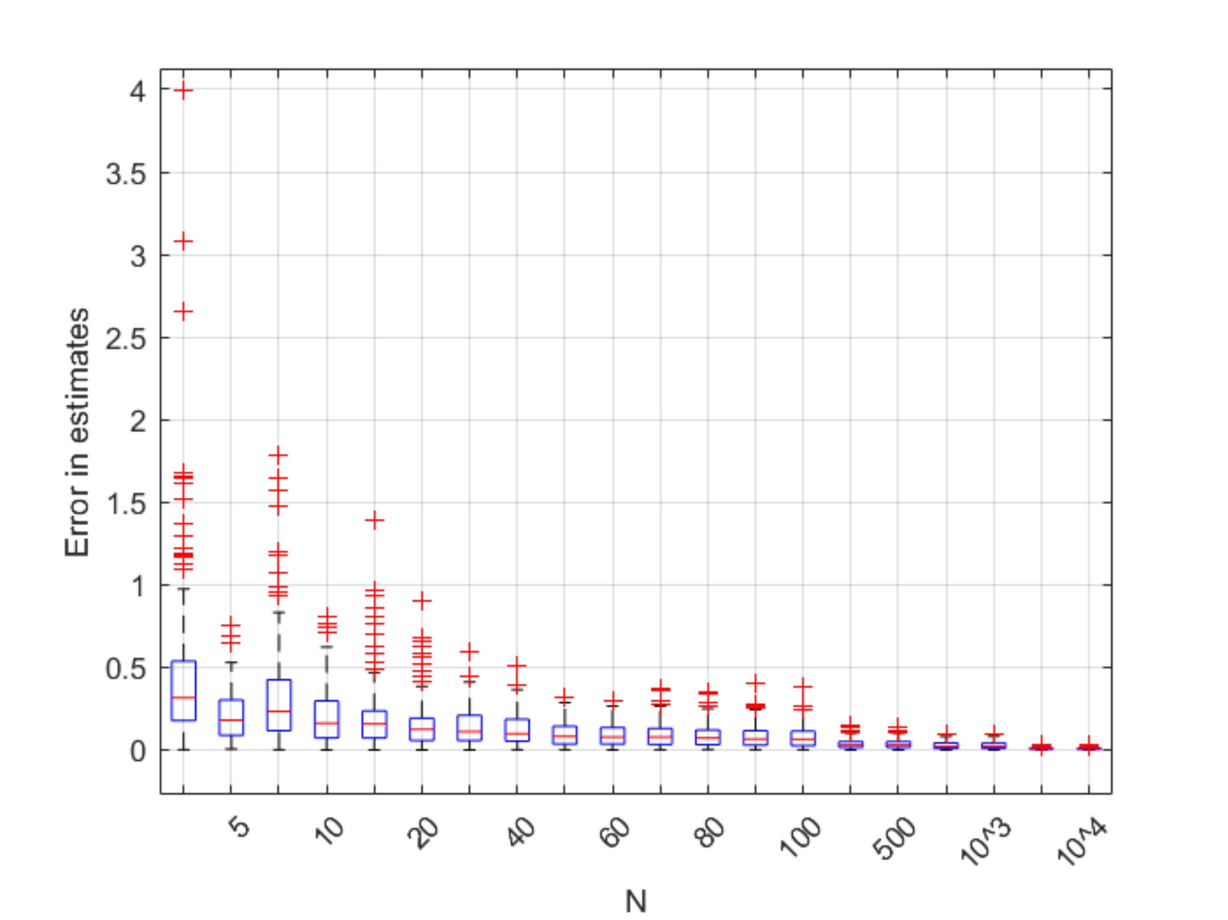
$$= \frac{N+\alpha}{w+\beta}$$

hence

$$\boxed{\hat{\lambda}_{\text{Posterior Mean}} = \frac{N+\alpha}{w+\beta}}$$

Using the derived result in code at line No 31

4) The Graphs.



Error in estimates
4
3.5
3
2.5
2
1.5
1
0.5
0
5
10
20
40
60
80
100
500
10^3
10^4
N

4] Interpretation

As its clearly observable from the graph that the Posterior Means are much more accurate this is a direct consequence of having a prior knowledge of where and what is the behaviour of parameter here in our case the prior information was given in terms of α & β.

1] Also after observing the graphs as N increases the relative error decreases and consequently because of which the parameter tends to its true value.

2] Out of the two our first choice should be Bayesian estimate since after careful analysis of the graphs the relative error for posterior mean for smaller sizes of data set is small as compared to the relative error observed in same datasets in MLE.

Since in labs and in scientific settings in Real world we can perform experiments only limited number of times, considering the fact that the dataset here generated would be quite small in size we would prefer to use Bayesian estimation since it produces less relative error for smaller datasets.

REPORT FOR QUESTION 3

1 UNIFORM RV & PARETO PRIOR

1.1 ML Estimate

Take N samples of X. The likelihood for X is:

$$P(X|\theta) = \begin{cases} \frac{1}{\theta^N} & 0 \leq x_i \leq \theta \\ 0 & otherwise \end{cases}$$

To maximise this, we will choose the minimum permissible value of θ, which is θ_m.

$$\widehat{\theta}^{ML} = \theta_m$$

1.2 MAP Estimate

We have the prior

$$P(\theta) = \begin{cases} k(\theta_m/\theta)^\alpha & \theta \geq \theta_m \\ 0 & otherwise \end{cases}$$

where k is proportionality constant. The posterior distribution is given as:

$$\begin{aligned} P(\theta|X) &= \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)P(\theta)d\theta} \\ &= \frac{(1/\theta^N)k(\theta_m/\theta)^\alpha}{\int_{\theta=\theta_m}^{\infty}(1/\theta^N)k(\theta_m/\theta)^\alpha d\theta} \end{aligned}$$

Evaluating integral in denominator:

$$\int_{\theta=\theta_m}^{\infty}(1/\theta^N)k(\theta_m/\theta)^\alpha d\theta = k\theta_m^\alpha \int_{\theta=\theta_m}^{\infty} \theta^{-\alpha-N} d\theta = k\theta_m^\alpha \left(\frac{\theta^{-\alpha-N+1}}{-\alpha-N+1}\right)_{\theta_m}^{\infty} = \frac{k}{\theta_m^{N-1}(N+\alpha-1)}$$

Hence the posterior distribution is:

$$P(\theta|X) = \begin{cases} \frac{(N+\alpha-1)}{\theta_m}\left(\frac{\theta_m}{\theta}\right)^{\alpha+N} & \theta \geq \theta_m \\ 0 & otherwise \end{cases}$$

To maximise this, we will choose the minimum permissible value of θ, which is θ_m.

$$\widehat{\theta}^{MAP} = \theta_m$$

1.3 Comparison of ML Estimate and MAP Estimate

The ML and MAP estimates tend to the same value, irrespective of the sample size. As N increases, both ML and MAP estimates become more accurate and converge to the true value. The result is desirable, and MAP estimate is slightly more accurate due to the prior shape parameter.

1.4 Estimator for Posterior Mean

We consider only the non-zero probability density region for the integration:

$$\begin{aligned}
E_{P(\theta|X)}[\Theta] &= \int_{\theta=\theta_m}^{\infty} \theta P(\theta|X) d\theta \\
&= (N+\alpha-1)\theta_m^{N+\alpha-1} \int_{\theta=\theta_m}^{\infty} \theta^{-\alpha-N+1} d\theta \\
&= \left(\frac{N+\alpha-1}{N+\alpha-2}\right) \theta_m^{N+\alpha-1} \left(-\theta^{-\alpha-N+2}\right)_{\theta_m}^{\infty} \\
&= \left(\frac{N+\alpha-1}{N+\alpha-2}\right) \theta_m
\end{aligned}$$

Thus the estimate of the posterior mean is:

$$\widehat{\theta}^{PosteriorMean} = \left(\frac{N+\alpha-1}{N+\alpha-2}\right) \theta_m$$

1.5 Comparison of ML Estimate and Posterior Mean

As the sample size N becomes larger, the posterior mean and the ML estimate both converge to the true value. This is desirable, since this means the prior data is also reliable for estimation, and the posterior mean will keep improving (getting closer to the true value) as we repeat the estimation.