2. Given $X \sim U(0,1)$

$$P(x) = \begin{cases} 1 & \text{if } 1 \geq x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We consider the transformation given below.

$$Y = \frac{-1}{\lambda} \log(X)$$

Hence,

$$\boxed{X = \exp(-\lambda Y)}$$

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

Using the notion of transformation of variables and the formula to determine the probability density of function of Y, we get that

$$\text{pdf of } Y,\ q(y) = P(g^{-1}(y)) \cdot \left| \frac{d}{dy}(g^{-1}(y)) \right|$$

$$\Rightarrow \text{Hence, } q(y) = 1 \cdot \lambda \exp(-\lambda y)$$

$$\Rightarrow q(y) = \lambda \exp(-\lambda y)$$

Thus, we have that

$$q(y|\lambda) = \lambda \exp(-\lambda y)$$

The likelihood is given by,

$$\text{Likelihood}, L = P(\{Y_i\}|\lambda) = \lambda^n \cdot \exp(-\lambda \Sigma_i y_i)$$

For the likelihood to be maximised,

$$\log(\text{Likelihood}) = n \log\lambda - \lambda \Sigma_i y_i$$

$$\Rightarrow \frac{d}{d\lambda}(\log(L)) = \frac{n}{\lambda} - \Sigma_i y_i$$

For $\log(L)$ to be maximum,

$$\frac{d}{d\lambda}(\log(L)) = 0.$$

$$\Rightarrow \frac{n}{\Sigma_i y_i} = \hat{\lambda} \quad \text{———— (i)}$$

Hence, the maximum likelihood estimate of λ is,

$$\hat{\lambda} = \frac{n}{\Sigma_i y_i}$$

Given that,

$$\text{Prior} \equiv \text{Gamma Distribution}$$
$$= \Gamma(\alpha = 5.5, \beta = 1).$$

Hence,

$$P(\lambda) = \frac{\beta^\alpha \cdot \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \quad \text{———— (ii)}$$

Date:

Thus,

$$\text{Posterior} = P(\lambda|_{\{y_i\}}) \propto P(x) \cdot P(\{y_i\}|_\lambda)$$

$$\Rightarrow \boxed{P(\lambda|_{\{x_i\}}) \propto \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{\alpha-1} \cdot e^{-\beta\lambda} \cdot \lambda^n e^{-\lambda \Sigma y_i}}$$

Thus,

$$\boxed{P(\lambda|_{\{x_i\}}) \propto \lambda^{n+\alpha-1} \cdot \exp(-\lambda(\Sigma y_i + \beta))}$$

$\Rightarrow$ the posterior is of the form, $\Gamma(n+\alpha, \beta + \Sigma y_i)$ i.e. a Gamma Distribution

Now, to derive the posterior mean, for any Gamma distribution

$$E(\lambda) = \int_\lambda \lambda \cdot \frac{\beta^\alpha \cdot \lambda^{\alpha-1} \cdot e^{-\beta\lambda}}{\Gamma(\alpha)} \cdot d\lambda$$

$$\Rightarrow E(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_\lambda \lambda^\alpha \cdot \exp(-\beta\lambda) \cdot d\lambda$$

Let $\lambda = t/\beta$

$$\Rightarrow E(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \int t^\alpha \cdot \exp(-t) \cdot \frac{dt}{\beta}$$

$$\Rightarrow E(\lambda) = \frac{1}{\Gamma(\alpha) \cdot \beta} \Gamma(\alpha+1)$$

$$\Rightarrow \boxed{E(\lambda) = \frac{\alpha}{\beta}}$$

Hence, for our posterior,

$$\boxed{\text{Posterior Mean} = \frac{\alpha + n}{\beta + \sum_i Y_i}}$$

- Interpretation:

  (i) The error is smaller for the Posterior in case of smaller sample sizes, as compared to the MLE.

  As the sample size increases, both the errors decrease, but the rate of drop in the error of MLE is much better than that of the Posterior.

  At extremely large data sets, the posterior mean, MLE mean converges.

  (ii) From this, it can be concluded that the Posterior has a better performance on smaller datasets, while as $N \to \infty$, MLE should be preferred.

  This can be justified by saying that the Posterior has a pre initiated bias due to the hyper parameters and hence performs well on smaller datasets.

  Also, it has a smaller variance, and hence the rate of decrease of error is lower as compared to MLE at higher value.