STAT7301 – Assignment 4
Joel Thomas 44793203

Q6.21

$$f(x;\theta) = \frac{\theta^x e^{-\theta}}{x!\,(1-e^{-\theta})}, \qquad x \in \{1,2,\dots\}, \qquad \theta > 0$$

Likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta), \qquad \left(X_i \sim^{iid}\right)$$

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

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

Log-likelihood function:

$$\log\big(L(\theta)\big) = \log\left(\frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{(1-e^{-\theta})^n \prod_{i=1}^{n} x_i!}\right)$$

$$\log\big(L(\theta)\big) = \log\left(e^{-n\theta}\right) + \log\left(\theta^{\sum_{i=1}^{n} x_i}\right) - \log\left(\left(1-e^{-\theta}\right)^n\right) - \log\left(\prod_{i=1}^{n} x_i!\right)$$

$$\log\big(L(\theta)\big) = -n\theta + \log(\theta)\sum_{i=1}^{n} x_i - n\log\left(1-e^{-\theta}\right) - \log\left(\prod_{i=1}^{n} x_i!\right)$$

Given the data $\boldsymbol{x} = [5,5,4,4,4,4,3,3,3,3,2,2,2,2,1,1]$ and $n = 16$, we obtain the plots in figure 1 below. See attached code $Q6\_21.m$ – using Grid Search on the likelihood function $L(\theta)$ (and verifying the solution by performing it again on the log-likelihood function $\log\big(L(\theta)\big)$), the MLE of θ given by $\hat{\theta}$ is approximately 2.8214.

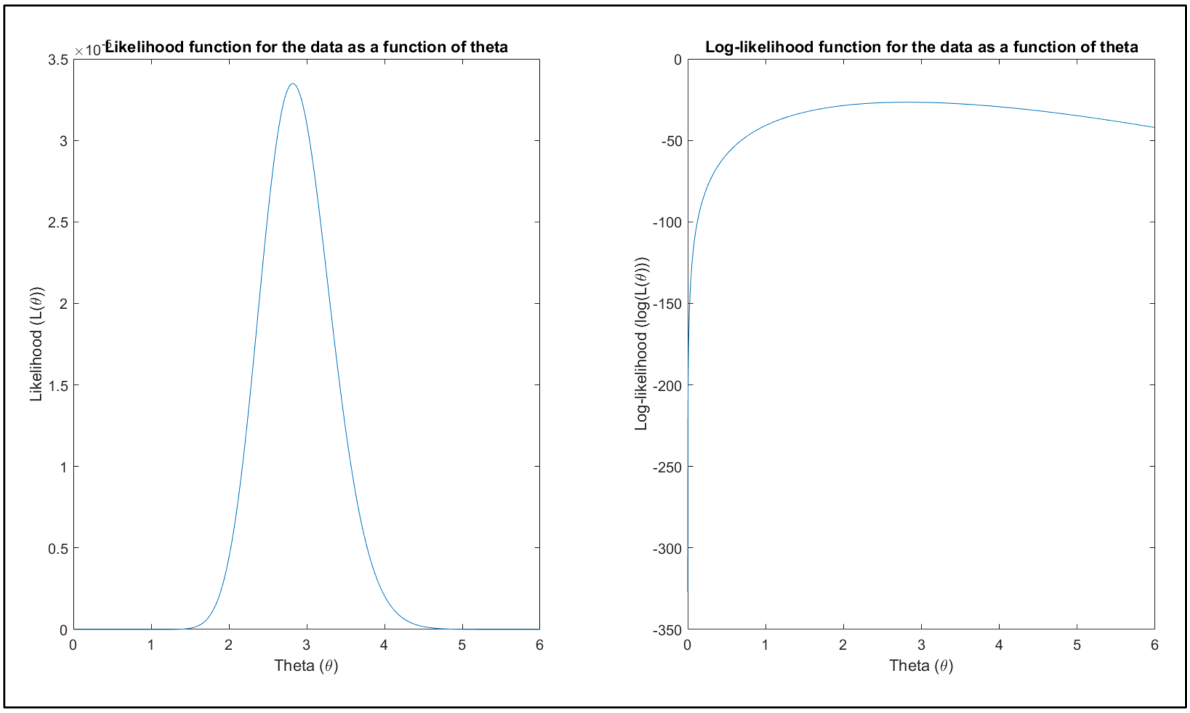

Figure 1: Plots of the likelihood $L(\theta)$ and log-likelihood $log(L(\theta))$ functions of the data.

Q6.25

a)

From the question, we know that:

$$\log(L(\theta)) = x_1 \log(\theta + 2) + (x_2 + x_3)\log(1 - \theta) + x_4 \log(\theta) + c, \qquad c = \text{arbitrary constant}$$

Score function:

$$S(\theta) = \frac{\partial \log(L(\theta))}{\partial \theta}$$

$$S(\theta) = \frac{\partial}{\partial \theta}(x_1 \log(\theta + 2) + (x_2 + x_3)\log(1 - \theta) + x_4 \log(\theta) + c)$$

$$S(\theta) = \frac{x_1}{\theta + 2} - \frac{x_2 + x_3}{1 - \theta} + \frac{x_4}{\theta}$$

$$x_1 = 125, \qquad x_2 = 18, \qquad x_3 = 20, \qquad x_4 = 34$$

$$\therefore S(\theta) = \frac{125}{\theta + 2} - \frac{18 + 20}{1 - \theta} + \frac{34}{\theta}$$

$$S(\theta) = \frac{34}{\theta} + \frac{125}{\theta + 2} - \frac{38}{1 - \theta}$$

Hessian function:

$$H(\theta) = \frac{\partial^2 \log(L(\theta))}{\partial \theta^2}$$

$$H(\theta) = \frac{\partial}{\partial \theta}\left(\frac{\partial \log(L(\theta))}{\partial \theta}\right)$$

$$H(\theta) = \frac{\partial}{\partial \theta}(S(\theta))$$

$$H(\theta) = \frac{\partial}{\partial \theta}\left(\frac{34}{\theta} + \frac{125}{\theta + 2} - \frac{38}{1 - \theta}\right)$$

$$H(\theta) = -1 \cdot 34 \cdot \theta^{-2} - 1 \cdot 1 \cdot 125 \cdot (\theta + 2)^{-2} - -1 \cdot -1 \cdot 38 \cdot (1 - \theta)^2$$

$$H(\theta) = -\frac{34}{\theta^2} - \frac{125}{(\theta + 2)^2} - \frac{38}{(1 - \theta)^2}$$

b)

Newton-Raphson procedure:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

For the purposes of this question, $x_{n+1} = \hat{\theta}_{n+1}, x_n = \hat{\theta}_n, f(x_n) = S(\hat{\theta}_n)$ and $f'(x_n) = H(\hat{\theta}_n)$ since the hessian function for θ is simply the derivative of the score function for θ. We set the initial point to be $\hat{\theta}_0 = 0.1$. See attached code $Q6\_25.m$ – using a Newton-Raphson procedure, the MLE of θ given by $\hat{\theta}$ is approximately equal to 0.6268.

c)

See attached code $Q6\_25.m$ – using Grid Search on the log-likelihood function $\log(L(\theta))$, the MLE of θ given by $\hat{\theta}$ is approximately 0.6268.

d)

The EM approach from example 6.21 yields the MLE of θ given by $\hat{\theta}$ to also be approximately 0.6268. Thus, the Newton-Raphson, Grid Search and EM approaches all provide the same ML estimate for θ.

Q7.3

X has cdf $F_X(x) = F_N = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(x_i \leq x)$. Hence it has support for $X = X_1, \dots, X_N$ (discrete random variable) where each X_i has probability mass $\frac{1}{N}$ associated with it.

$$\rightarrow \mathbb{E}[X] = \sum_{i=0}^{\infty} x_i f_X(x_i)$$

$$\mathbb{E}[X] = \sum_{i=1}^{N} x_i \cdot \mathbb{P}(X = x_i)$$

$$\mathbb{E}[X] = \sum_{i=1}^{N} x_i \cdot \frac{1}{N}$$

$$\therefore \mathbb{E}[X] = \frac{1}{N}\sum_{i=1}^{N} x_i = \bar{x}$$

$$\rightarrow \mathbb{E}[X]^2 = \bar{x}^2$$

$$\mathbb{E}[X^2] = \sum_{i=1}^{N} x_i^2 \cdot \mathbb{P}(X = x_i)$$

$$\mathbb{E}[X^2] = \sum_{i=1}^{N} x_i^2 \cdot \frac{1}{N}$$

$$\mathbb{E}[X^2] = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$\therefore Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2$$

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2\bar{x}^2 + \bar{x}^2$$

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2\frac{\bar{x}}{N}\sum_{i=1}^{N} x_i + \frac{1}{N}\sum_{i=1}^{N} \bar{x}^2$$

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2x_i\bar{x} + \bar{x}^2$$

$$\therefore Var(X) = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

Q7.8

a)

If $X \sim Exp(\lambda) \rightarrow f_X(x) = \lambda e^{-\lambda x}$. Since the exponential distribution only has support for $x > 0$:

$$\int_{-\infty}^{m} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{m} f(x)\,dx$$

$$\int_{-\infty}^{m} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{m} \lambda e^{-\lambda x}\,dx$$

$$\int_{-\infty}^{m} f(x)\,dx = 0 + \int_{0}^{m} \lambda e^{-\lambda x}\,dx$$

$$\int_{-\infty}^{m} f(x)\,dx = \left[-e^{-\lambda x}\right]_{x=0}^{x=m}$$

$$\int_{-\infty}^{m} f(x)\,dx = -e^{-\lambda m} + e^{0}$$

$$\int_{-\infty}^{m} f(x)\,dx = 1 - e^{-\lambda m} = \frac{1}{2}$$

$$\therefore e^{-\lambda m} = 1 - \frac{1}{2}$$

$$-\lambda m = \log\left(\frac{1}{2}\right)$$

$$-\lambda m = \log(2^{-1})$$

$$-\lambda m = -\log(2)$$

$$\therefore m = \frac{\log(2)}{\lambda}$$

b)

Likelihood function:

$$L(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda)\,, \qquad \left(X_i \sim^{iid} Exp(\lambda)\right)$$

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

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

Log-likelihood function:

$$\log\left(L(\lambda)\right) = \log\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$$\log\left(L(\lambda)\right) = \log(\lambda^n) + \log\left(e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$$\log\left(L(\lambda)\right) = n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

Solving Likelihood equation to obtain MLE for λ given by $\hat{\lambda}$:

$$\frac{\partial \log\left(L(\lambda)\right)}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left(n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i\right)$$

$$\frac{\partial \log\left(L(\lambda)\right)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

See attached code $Q7\_8.m$ – the estimate for λ given by $T = \log(2)\setminus\tilde{x}$ is approximately equal to 0.4925 whereas the MLE for λ given by $\hat{\lambda} = \frac{1}{\bar{x}}$ is approximately equal to 0.2773.

c)

See attached code $Q7\_8.m$ – after carrying out a bootstrap analysis of both estimators using $K = 10000$ resamples, two KDE plots were generated to compare their accuracies as shown in figure 2 below. The major difference between the two estimators is that the T estimator appears to follow a bimodal distribution (two peaks) and there is a lot more uncertainty inherent in this estimator as evident from the width of its distribution relative to the distribution of the MLE estimator.

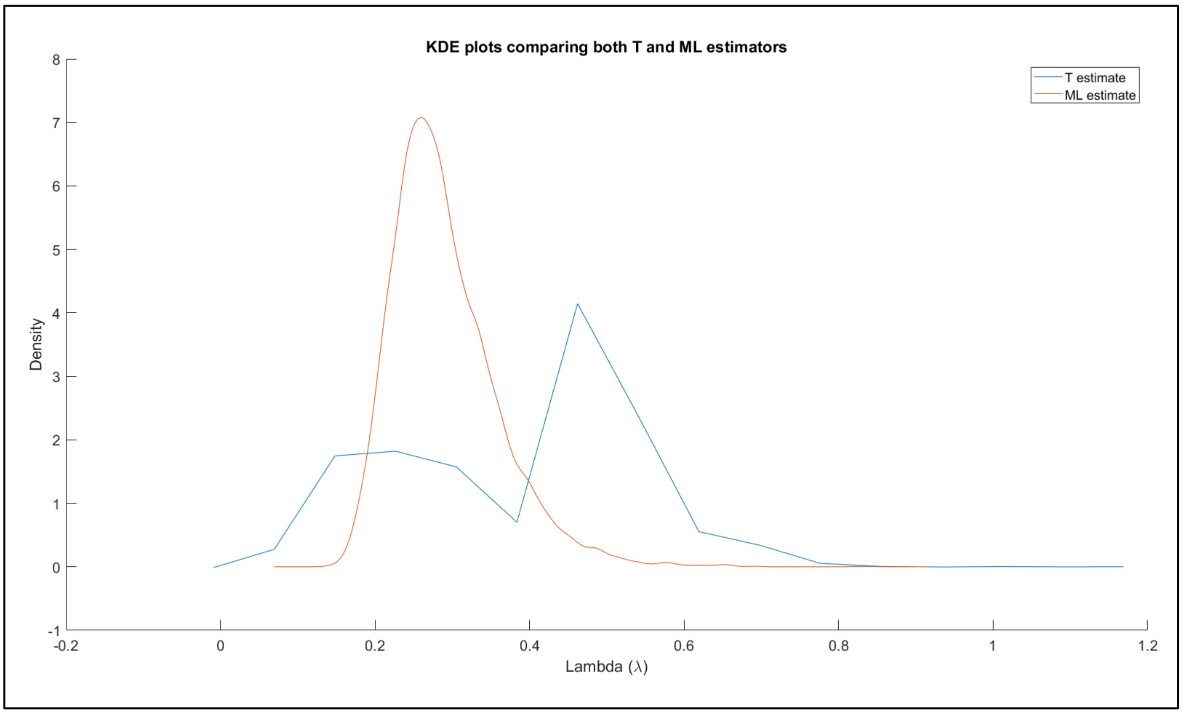

Figure 2: KDE plots after performing bootstrap analysis of both estimators using $K = 10000$ resamples.

Q7.14

See attached code $Q7\_14.m$ – based on figure 3 below, a rough estimate for the burn-in period is $B \approx 1000$. This estimate is justified because the target distribution is $N(10,2)$ so we expect the random walk to centre around 10 (with some volatility due to the variance) after the burn-in period and this is evident in figure 3. The 20 95% confidence intervals, each based on the mean of 100 values for $\log(X_i^2)$ (which in turn is based on the mean of all values of a random walk between the burn-in period B and the sample size N), can be viewed by viewing the CIs variable in the attached code. Based on these confidence intervals, the true value for l given by 4.58453 is only contained in these intervals 50% of the time (current run) which is a probability much smaller than 95%. Thus, the combination of a burn-in size of $B = 1000$ and a sample size of $N = 5000$ is inadequate to provide an accurate estimate for l.

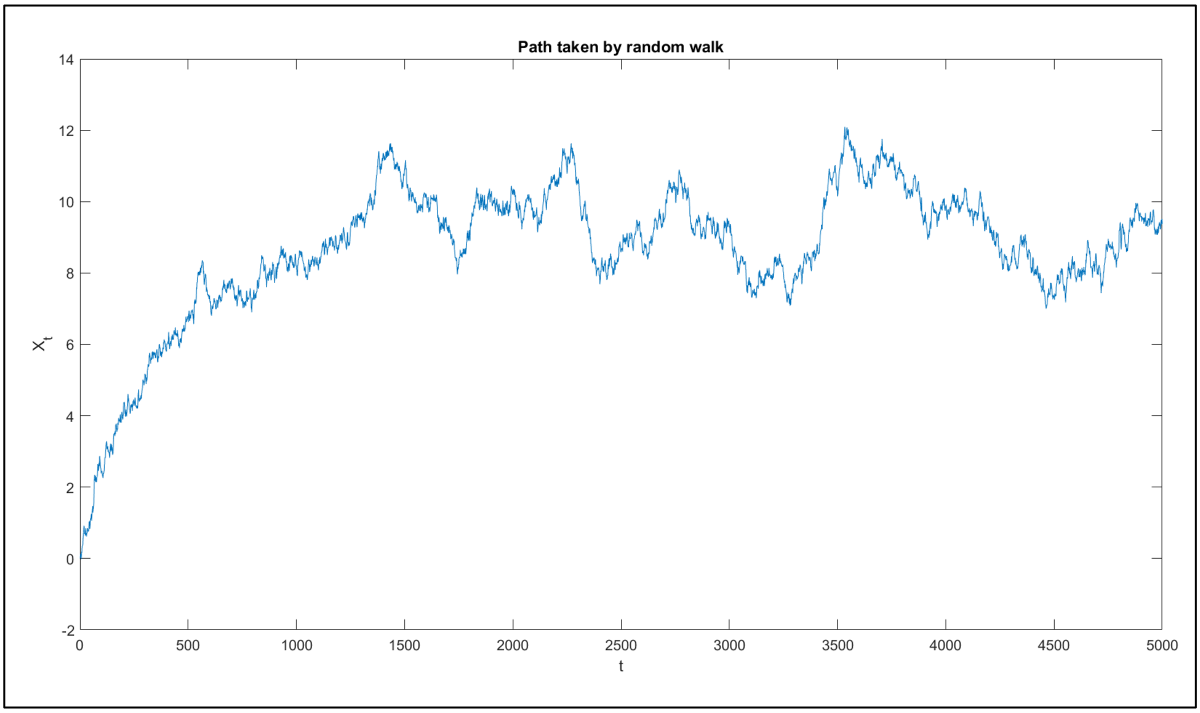

Figure 3: Plot of the path taken by the random walk (will be different for each re-run).

Q7.19

a)

See attached code $Q7\_19.m$.

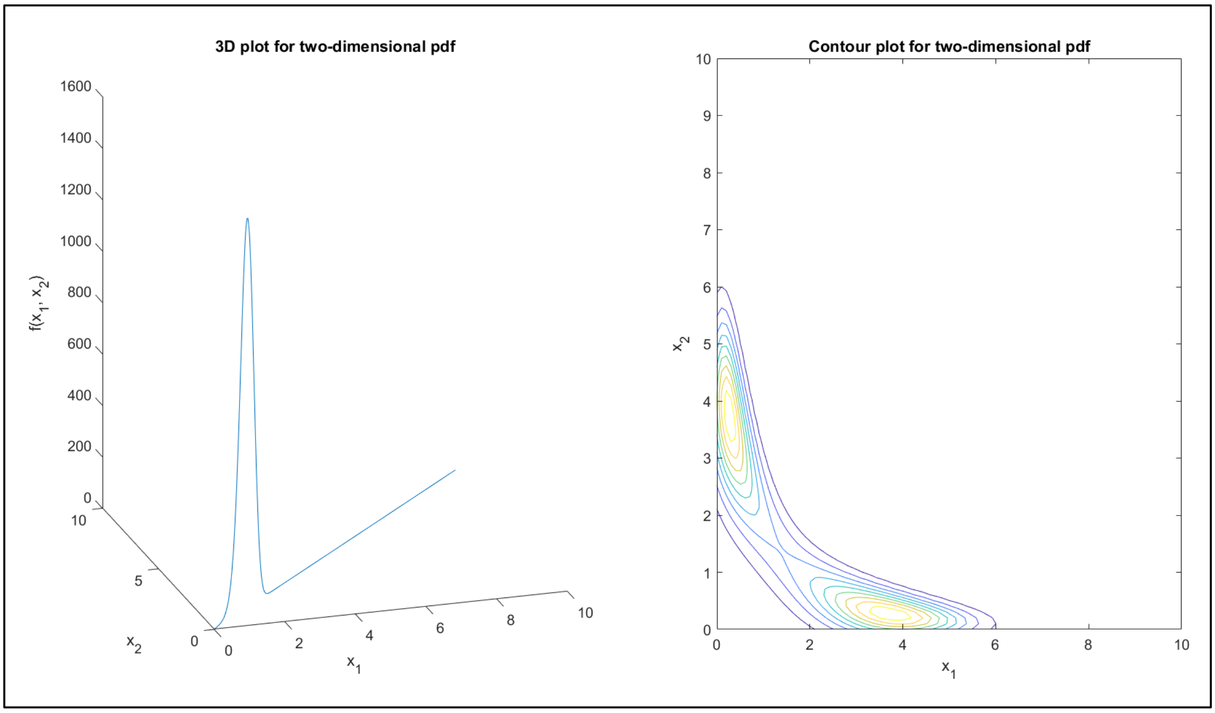

Figure 4: 3D and contour plots for two-dimensional pdf $f(x)$ ignoring c.

b)

Implemented random walk sampler viewable in attached code $Q7\_19.m$.

c)

Since $\boldsymbol{Z} \sim N(\boldsymbol{0}, I)$:

$$\sigma\boldsymbol{Z} \sim N(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma & 0\\0 & \sigma\end{bmatrix}$$

Hence the parameter σ affects the variances for each component of the Markov chain leaving the mean for and covariance between each component unchanged (independent components/random variables). A higher variance indicates higher volatility inherent in the path generated by the random walk and this is the only difference evident in figure 5 below.

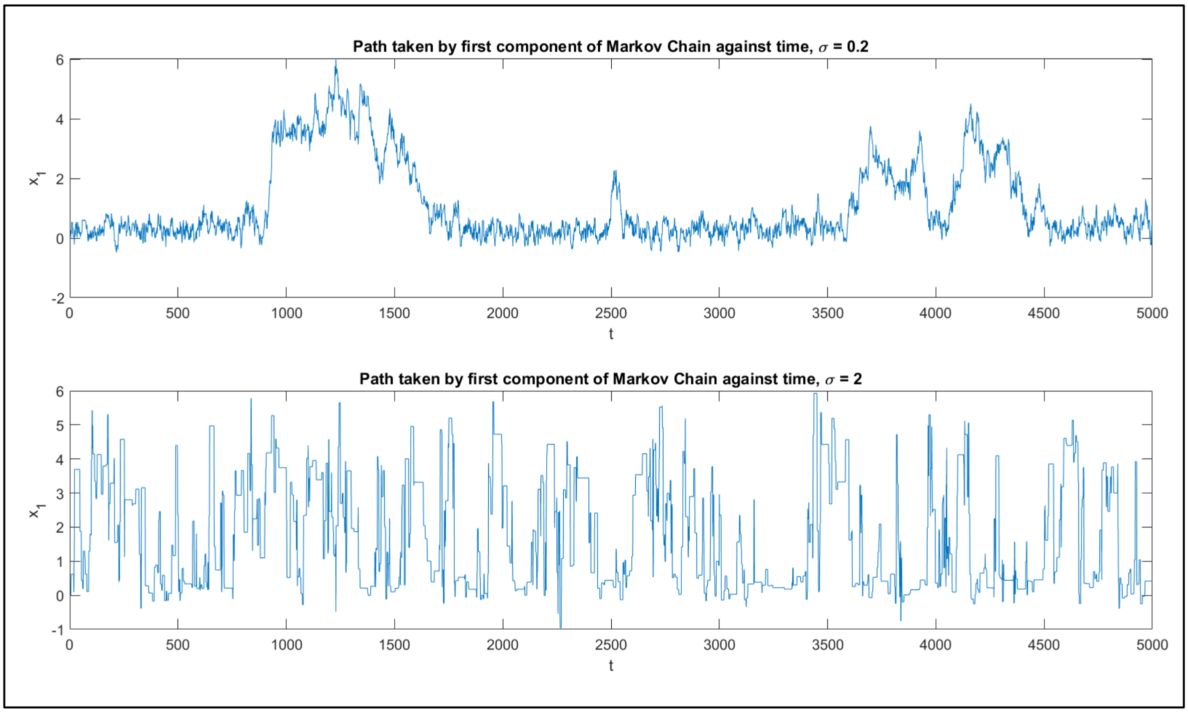

Figure 5: Progression of the first component of the Markov chain against time, for $\sigma = 0.2$ and $\sigma = 2$.

d)

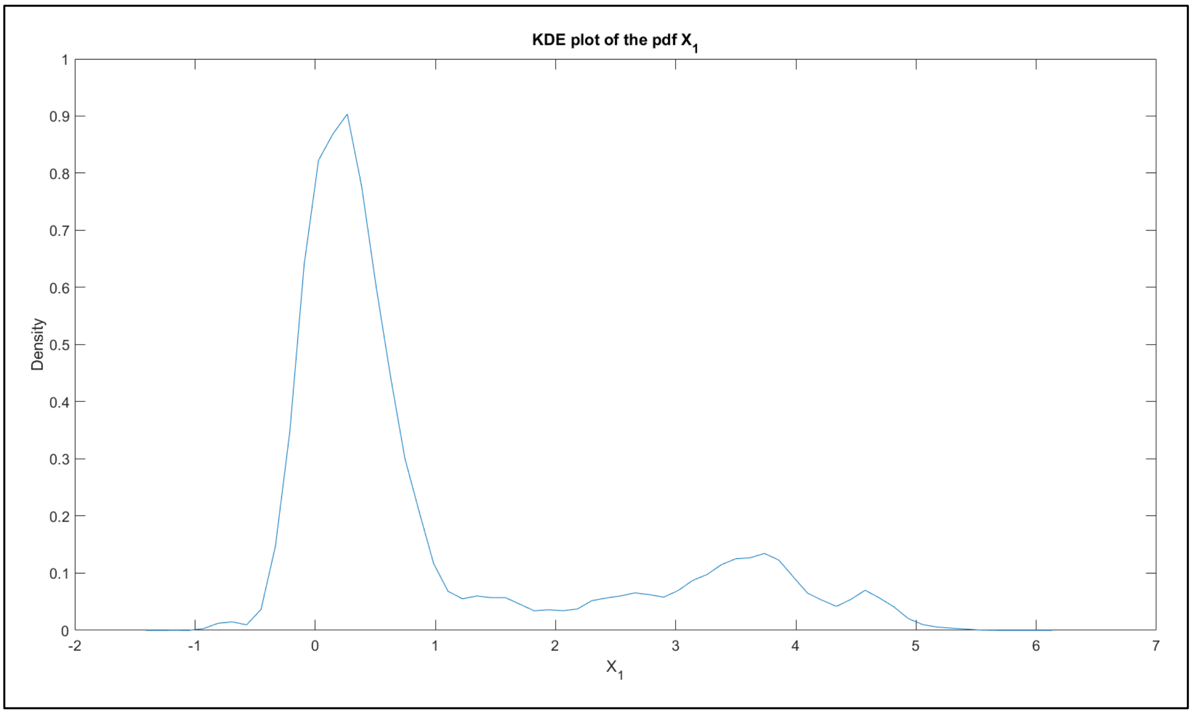

Figure 6: KDE plot of the pdf X_1 if $\boldsymbol{X} = (X_1, X_2) \sim f$.

Q8.7

a)

$$posterior \propto likelihood \times prior$$

$$prior = f(p)$$

$$f(p) = \frac{1}{1-0}, \qquad \big(p \sim U(0,1)\big)$$

$$f(p) = 1$$

$$likelihood = f(x|p)$$

$$f(x|p) = (1-p)^{x-1}p, \qquad \big((X|p) \sim Geom(p)\big)$$

$\therefore posterior = f(p|x)$

$f(p|x) \propto f(x|p)f(p)$

$f(p|x) \propto (1-p)^{x-1}p \cdot 1$

$f(p|x) \propto p(1-p)^{x-1}$

$f(p|x) \propto p^{2-1}(1-p)^{x-1}$

$\rightarrow (p|x) \sim \mathrm{B}(\alpha, \beta) = \mathrm{B}(2, x), \qquad \alpha, \beta > 0$

b)

The mode for a $\mathrm{B}(\alpha, \beta)$ is given by $\frac{\alpha-1}{\alpha+\beta-2}$. Hence, the mode for the posterior which is distributed $\mathrm{B}(2, x)$ is given by:

$$= \frac{2-1}{2+x-2}$$

$$= \frac{1}{x} \in (0,1], \qquad x \in \mathbb{N} \text{ since } (X|p) \sim Geom(p) \text{ takes values } x \in \mathbb{N}$$

c)

The expectation for a $\mathrm{B}(\alpha, \beta)$ is given by $\frac{\alpha}{\alpha+\beta}$. Hence, the expectation for the posterior which is distributed $\mathrm{B}(2, x)$ is given by:

$$= \frac{2}{2+x}$$

$$= \frac{2}{x+2} \in \left(0, \frac{2}{3}\right], \qquad x \in \mathbb{N} \text{ since } (X|p) \sim Geom(p) \text{ takes values } x \in \mathbb{N}$$

Q8.8

a)

$$posterior \propto likelihood \times prior$$

$$prior = f(\theta) \propto c, \qquad c = \text{constant}$$

$$likelihood = f(\boldsymbol{x}|\theta)$$

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i;\theta), \qquad \left(X_i \sim^{iid} Exp\left(\frac{1}{\theta}\right)\right)$$

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{1}{\theta}x_i}$$

$$f(\boldsymbol{x}|\theta) = \left(\frac{1}{\theta}\right)^n e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}$$

$$f(\boldsymbol{x}|\theta) = \theta^{-n} e^{-n\bar{x}\theta^{-1}}$$

$$\therefore posterior = f(\theta|\boldsymbol{x})$$

$$f(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta) f(\theta)$$

$$f(\theta|\boldsymbol{x}) \propto \theta^{-n} e^{-n\bar{x}\theta^{-1}}$$

$$f(\theta|\boldsymbol{x}) \propto \theta^{-(n+1)-1} e^{-n\bar{x}\theta^{-1}}$$

$$\rightarrow (\theta|\boldsymbol{x}) \sim \text{Inverse-Gamma}(\alpha,\beta) = \text{Inverse-Gamma}(n+1, n\bar{x}), \qquad \alpha,\beta > 0$$

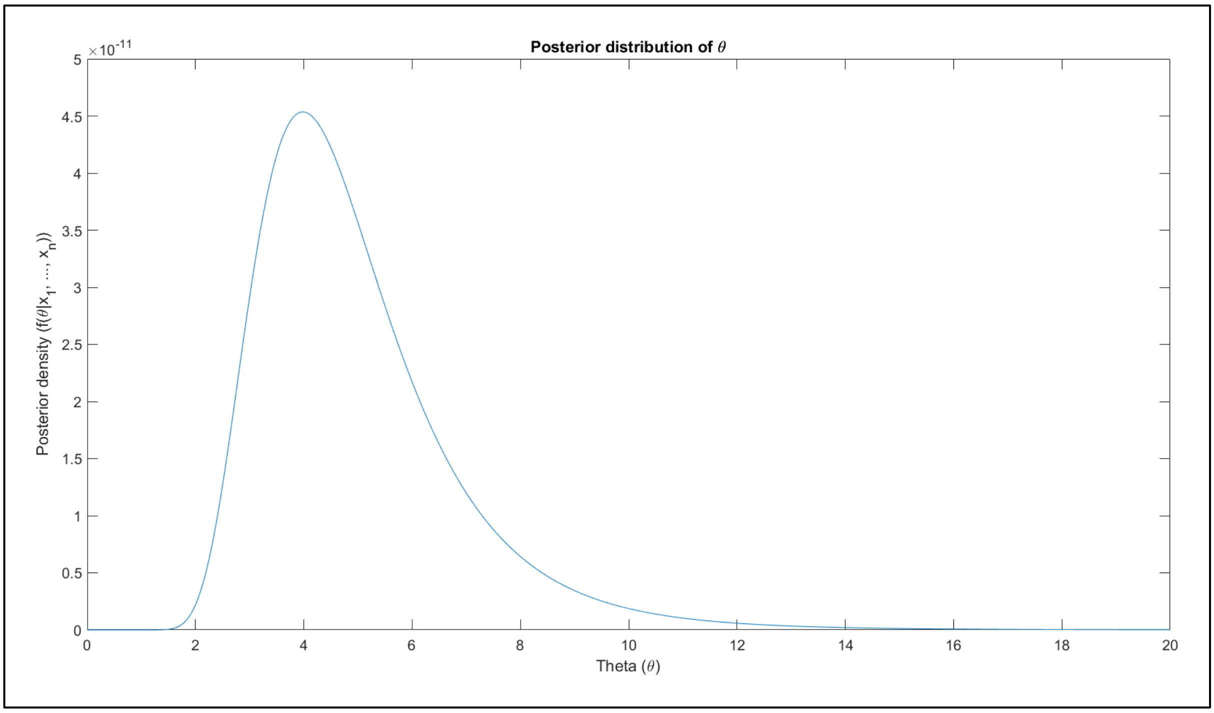

Figure 7: Plot of the unnormalized posterior distribution of θ given by $f(\theta|\boldsymbol{x})$.

b)

See attached code $Q8\_8.m$ – running an independent sampler using the Metropolis-Hastings MCMC algorithm with $N = 10^5$ sample and dumping the first $B = 1000$ samples as burn-in, the 2.5% lower and 97.5% upper quantiles from the posterior distribution are approximately 2.1620 and 7.2544 respectively.

Q6.17

If $X \sim \text{Gamma}(\alpha, \lambda)$ then:

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

Likelihood function:

$$L(\alpha, \lambda; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \alpha, \lambda), \qquad \left(X_i \sim^{iid} \text{Gamma}(\alpha, \lambda)\right)$$

$$L(\alpha, \lambda; \boldsymbol{x}) = \prod_{i=1}^{n} \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}$$

$$L(\alpha, \lambda; \boldsymbol{x}) = \frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} e^{-\lambda \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i^{\alpha-1}$$

Log-likelihood function:

$$\log\left(L(\alpha,\lambda;\boldsymbol{x})\right) = \log\left(\frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} e^{-\lambda\sum_{i=1}^n x_i} \prod_{i=1}^n x_i^{\alpha-1}\right)$$

$$\log\left(L(\alpha,\lambda;\boldsymbol{x})\right) = \log(\lambda^{n\alpha}) + \log\left(e^{-\lambda\sum_{i=1}^n x_i}\right) + \log\left(\prod_{i=1}^n x_i^{\alpha-1}\right) - \log(\Gamma(\alpha)^n)$$

$$\log\left(L(\alpha,\lambda;\boldsymbol{x})\right) = n\alpha\log(\lambda) - \lambda\sum_{i=1}^n x_i + (\alpha-1)\sum_{i=1}^n \log(x_i) - n\log\left(\Gamma(\alpha)\right)$$

Score function:

$$\boldsymbol{S}(\alpha,\lambda) = \begin{bmatrix} \dfrac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha} \\ \dfrac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda} \end{bmatrix}, \qquad \text{(by definition)}$$

$$\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha} = \frac{\partial}{\partial\alpha}\left(n\alpha\log(\lambda) - \lambda\sum_{i=1}^n x_i + (\alpha-1)\sum_{i=1}^n \log(x_i) - n\log\left(\Gamma(\alpha)\right)\right)$$

$$\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha} = n\log(\lambda) + \sum_{i=1}^n \log(x_i) - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha} = n\left(\log(\lambda) - \psi(\alpha)\right) + \sum_{i=1}^n \log(x_i)$$

$$\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda} = \frac{\partial}{\partial\lambda}\left(n\alpha\log(\lambda) - \lambda\sum_{i=1}^n x_i + (\alpha-1)\sum_{i=1}^n \log(x_i) - n\log\left(\Gamma(\alpha)\right)\right)$$

$$\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^n x_i$$

$$\therefore \boldsymbol{S}(\alpha,\lambda) = \begin{bmatrix} n\left(\log(\lambda) - \psi(\alpha)\right) + \sum_{i=1}^n \log(x_i) \\ \dfrac{n\alpha}{\lambda} - \sum_{i=1}^n x_i \end{bmatrix}$$

Verify expectation of score function is a vector of zeros:

$$\mathbb{E}[\boldsymbol{S}(\alpha,\lambda)] = \begin{bmatrix} \mathbb{E}\left[n\big(\log(\lambda) - \psi(\alpha)\big) + \sum_{i=1}^{n} \log(X_i)\right] \\ \mathbb{E}\left[\frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i\right] \end{bmatrix}$$

$$\mathbb{E}\left[n\big(\log(\lambda) - \psi(\alpha)\big) + \sum_{i=1}^{n} \log(X_i)\right] = n\big(\log(\lambda) - \psi(\alpha)\big) + \sum_{i=1}^{n} \mathbb{E}[\log(X_i)], \qquad \text{(constants, linearity of } \mathbb{E}[\cdot])$$

$$\mathbb{E}\left[n\big(\log(\lambda) - \psi(\alpha)\big) + \sum_{i=1}^{n} \log(X_i)\right] = n\big(\log(\lambda) - \psi(\alpha)\big) + n\mathbb{E}[\log(X_i)], \qquad \big(X_i \sim^{iid} \text{Gamma}(\alpha,\lambda)\big)$$

Following from the last line, we know that:

$X_i \sim^{iid} \text{Gamma}(\alpha,\lambda) \rightarrow \log(X_i) \sim^{iid} \text{log-Gamma}(\alpha,\lambda)$

We use the well-known result that the expectation of a log-Gamma random variable is given by $\psi(\alpha) - \log(\lambda)$. Continuing from the last line:

$$\mathbb{E}\left[n\big(\log(\lambda) - \psi(\alpha)\big) + \sum_{i=1}^{n} \log(X_i)\right] = n\big(\log(\lambda) - \psi(\alpha)\big) + n(\psi(\alpha) - \log(\lambda))$$

$$\mathbb{E}\left[n\big(\log(\lambda) - \psi(\alpha)\big) + \sum_{i=1}^{n} \log(X_i)\right] = 0$$

$$\mathbb{E}\left[\frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i\right] = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} \mathbb{E}[X_i], \qquad \text{(constants, linearity of } \mathbb{E}[\cdot])$$

$$\mathbb{E}\left[\frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i\right] = \frac{n\alpha}{\lambda} - n \cdot \frac{\alpha}{\lambda}, \qquad \left(\mathbb{E}[X] = \frac{\alpha}{\lambda} \text{ if } X \sim \text{Gamma}(\alpha,\lambda)\right)$$

$$\mathbb{E}\left[\frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i\right] = 0$$

$$\rightarrow \mathbb{E}[\boldsymbol{S}(\alpha,\lambda)] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\therefore I(\alpha,\lambda) = \begin{bmatrix} \mathbb{E}\left[\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\right)^2\right] & \mathbb{E}\left[\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\cdot\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda}\right] \\ \mathbb{E}\left[\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda}\cdot\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\right] & \mathbb{E}\left[\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda}\right)^2\right] \end{bmatrix}, \qquad \text{(by definition)}$$

$$I(\alpha,\lambda) = \begin{bmatrix} Var\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\right) & Cov\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha},\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda}\right) \\ Cov\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda},\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\right) & Var\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda}\right) \end{bmatrix}$$

since $Var(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \mathbb{E}[Z^2]$ if $\mathbb{E}[Z] = \mathbb{E}[Z]^2 = 0$ (we have proved earlier that $\mathbb{E}[\boldsymbol{S}(\alpha,\lambda)] = \begin{bmatrix}0\\0\end{bmatrix}$).

Continuing by finding each of the elements of $I(\alpha,\lambda)$:

$$Var\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\right) = Var\left(n\left(\log(\lambda) - \psi(\alpha)\right) + \sum_{i=1}^{n}\log(X_i)\right)$$

$$Var\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\right) = \sum_{i=1}^{n} Var(\log(X_i)), \qquad \left(\text{constants}, X_i \sim^{iid} \text{Gamma}(\alpha,\lambda)\right)$$

Using the well-known result that the variance of a log-Gamma random variable is given by $\psi'(\alpha)$.

$$Var\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha}\right) = n\psi'(\alpha)$$

$$Cov\left(\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\alpha},\frac{\partial \log\left(L(\alpha,\lambda;\boldsymbol{x})\right)}{\partial\lambda}\right)$$

$$= Cov\left(n\left(\log(\lambda) - \psi(\alpha)\right) + \sum_{i=1}^{n}\log(X_i), \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i\right)$$

$$= Cov\left(\sum_{i=1}^{n}\log(X_i), -\sum_{i=1}^{n} X_i\right), \qquad \text{(constants)}$$

$$= \sum_{i=1}^{n}\sum_{i=1}^{n} Cov(\log(X_i), X_i), \qquad \left(\text{covariance property for sums}, X_i \sim^{iid} \text{Gamma}(\alpha,\lambda)\right)$$

$$= \sum_{i=1}^{n}\sum_{i=1}^{n} \mathbb{E}[X_i \log(X_i)] - \mathbb{E}[X_i]\mathbb{E}[\log(X_i)]$$

$$= n^2(\mathbb{E}[X_i \log(X_i)] - \mathbb{E}[X_i]\mathbb{E}[\log(X_i)])$$

$$\mathbb{E}[X\log(X)] = \int_0^\infty x\log(x)\cdot\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\,dx$$

$$\mathbb{E}[X\log(X)] = \int_0^\infty \log(x)\cdot\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha}e^{-\lambda x}\,dx$$

$$\mathbb{E}[X\log(X)] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}\int_0^\infty \log(x)\cdot\frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}x^{(\alpha+1)-1}e^{-\lambda x}\,dx$$

$$\mathbb{E}[X\log(X)] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}\cdot\mathbb{E}[\log(Z)], \qquad \big(Z\sim\text{Gamma}(\alpha+1,\lambda)\big)$$

$$\mathbb{E}[X\log(X)] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}(\psi(\alpha+1)-\log(\lambda)), \qquad (\mathbb{E}[\log(Z)] = \psi(\alpha+1)-\log(\lambda))$$

$$\mathbb{E}[X]\mathbb{E}[\log(X)] = \frac{\alpha}{\lambda}(\psi(\alpha)-\log(\lambda))$$

$$\therefore\ \mathbb{E}[X\log(X)] - \mathbb{E}[X]\mathbb{E}[\log(X)] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}(\psi(\alpha+1)-\log(\lambda)) - \frac{\alpha}{\lambda}(\psi(\alpha)-\log(\lambda))$$

$$\mathbb{E}[X\log(X)] - \mathbb{E}[X]\mathbb{E}[\log(X)] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}\left(\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)}-\log(\lambda)\right) - \frac{\alpha}{\lambda}(\psi(\alpha)-\log(\lambda))$$

$$\mathbb{E}[X\log(X)] - \mathbb{E}[X]\mathbb{E}[\log(X)] = \frac{\Gamma'(\alpha+1)}{\Gamma(\alpha)\lambda} - \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}\log(\lambda) - \frac{\alpha}{\lambda}\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \frac{\alpha}{\lambda}\log(\lambda)$$

$$\mathbb{E}[X\log(X)] - \mathbb{E}[X]\mathbb{E}[\log(X)] = \frac{\Gamma'(\alpha+1)}{\Gamma(\alpha)\lambda} - \frac{\alpha}{\lambda}\log(\lambda) - \frac{\alpha}{\lambda}\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \frac{\alpha}{\lambda}\log(\lambda)\,, \qquad \left(\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda} = \frac{\alpha!}{(\alpha-1)!} = \alpha\right)$$

$$\mathbb{E}[X\log(X)] - \mathbb{E}[X]\mathbb{E}[\log(X)] = \frac{1}{\lambda\Gamma(\alpha)}\big(\Gamma'(\alpha+1)-\alpha\Gamma'(\alpha)\big)$$

$$\mathbb{E}[X\log(X)] - \mathbb{E}[X]\mathbb{E}[\log(X)] = -\frac{1}{n\lambda}$$

$$\therefore\ Cov\left(\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\alpha},\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\lambda}\right) = n^2(\mathbb{E}[X_i\log(X_i)] - \mathbb{E}[X_i]\mathbb{E}[\log(X_i)])$$

$$Cov\left(\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\alpha},\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\lambda}\right) = n^2\cdot -\frac{1}{n\lambda} = -\frac{n}{\lambda}$$

$$\rightarrow Cov\left(\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\lambda},\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\alpha}\right) = Cov\left(\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\alpha},\frac{\partial\log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial\lambda}\right) = -\frac{n}{\lambda}$$

$$Var\left(\frac{\partial \log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial \lambda}\right) = Var\left(\frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i\right)$$

$$Var\left(\frac{\partial \log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial \lambda}\right) = \sum_{i=1}^{n} Var(X_i), \qquad \left(\text{constants}, X_i \sim^{iid} \text{Gamma}(\alpha,\lambda)\right)$$

$$Var\left(\frac{\partial \log\big(L(\alpha,\lambda;\boldsymbol{x})\big)}{\partial \lambda}\right) = \frac{n\alpha}{\lambda^2}, \qquad \left(Var(X_i) = \frac{\alpha}{\lambda^2}\right)$$

$$\therefore I(\alpha,\lambda) = \begin{bmatrix} n\psi'(\alpha) & -\frac{n}{\lambda} \\ -\frac{n}{\lambda} & \frac{n\alpha}{\lambda^2} \end{bmatrix}$$

$$I(\alpha,\lambda) = n\begin{bmatrix} \psi'(\alpha) & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \frac{\alpha}{\lambda^2} \end{bmatrix}$$

Q6.23

See attached code $Q6\_23.m$ – after plotting the coverage probability for both intervals over a range of different n, it is evident that the score interval in Example 6.16 on page 175 indeed has better coverage behaviour than the standard confidence interval in Problem 5.22 on page 158 over the complete range of p. This is especially noticeable for small values of n e.g. $n = 10$ as shown in the first subplot in figure 8 below.

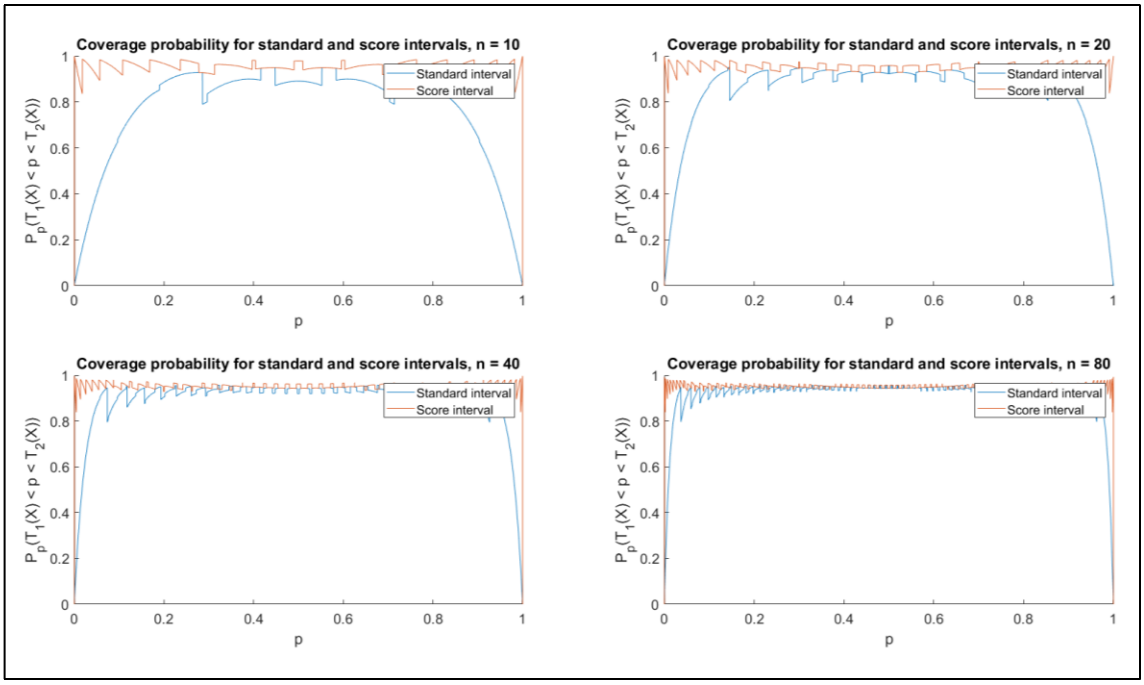

Figure 8: Plot of the exact coverage probability for the standard and score intervals as functions for p for $n \in \{10, 20, 40, 80\}$.