IB Statistics
Example Sheet 1

Samuel Lee

Question 1

"Ask your supervisor to test you on the sheet of common distributions handed out in lectures."

Question 2

If $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$ are independent, and $Z = \min(X, Y)$, we have

$$F_Z(z) = 1 - \Pr(Z > z) = 1 - \Pr(X > z, Y > z) = 1 - [1 - F_X(z)]\,[1 - F_Y(z)] = 1 - e^{-(\lambda+\mu)z}$$

which means $Z \sim \text{Exp}(\lambda + \mu)$.

If $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$, their moment generating functions are

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha} \text{ for } t < \lambda$$
$$M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\beta} \text{ for } t < \lambda$$

Since X and Y are independent, the moment generating function for $X + Y$ is the product of the two above:

$$M_{X+Y}(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha+\beta} \quad \text{for } t < \lambda$$

which means $X + Y$ has a $\Gamma(\alpha + \beta, \lambda)$ distribution.

Letting $Z = \frac{X}{X+Y}$, we have

$$\begin{aligned}
F_Z(z) &= \Pr\left(\frac{X}{X+Y} \le z\right) \\
&= \Pr\left(X \le \frac{z}{1-z}Y\right) \\
&= \int_0^\infty \Pr\left(X \le \frac{z}{1-z}Y \,\middle|\, Y = y\right) f_Y(y)\mathrm{d}y \\
&= \int_0^\infty \int_0^{\frac{z}{1-z}y} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\mathrm{d}x \; \frac{\lambda^\beta}{\Gamma(\beta)} y^{\beta-1} e^{-\lambda y}\mathrm{d}y \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \int_0^{\frac{z}{1-z}y} x^{\alpha-1} e^{-\lambda x}\mathrm{d}x \; y^{\beta-1} e^{-\lambda y}\mathrm{d}y
\end{aligned}$$

Using Leibniz's integral rule, we have

$$
\begin{aligned}
f_Z(z) &= F_Z'(z) \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \frac{y}{(1-z)^2} \left(\frac{z}{1-z}y\right)^{\alpha-1} e^{-\frac{z}{1-z}\lambda y} y^{\beta-1} e^{-\lambda y} \mathrm{d}y
\end{aligned}
$$

since $F_Z(z)$ only depends on z through the inner integral, and the inner integral only depends on z through the upper limit. Simplifying, we have

$$
\begin{aligned}
f_Z(z) &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty y^{\alpha+\beta-1} z^{\alpha-1} (1-z)^{-\alpha-1} e^{-\frac{\lambda y}{1-z}} \mathrm{d}y \\
&= \frac{\lambda^{\alpha+\beta} z^{\alpha-1} (1-z)^{-\alpha-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty y^{\alpha+\beta-1} e^{-\frac{\lambda y}{1-z}} \mathrm{d}y
\end{aligned}
$$

Letting $u = \frac{\lambda y}{1-z}$,

$$
\begin{aligned}
f_Z(z) &= \frac{\lambda^{\alpha+\beta} z^{\alpha-1} (1-z)^{-\alpha-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \left(\frac{1-z}{\lambda}u\right)^{\alpha+\beta-1} e^{-u} \frac{1-z}{\lambda} \mathrm{d}u \\
&= \frac{z^{\alpha-1}(1-z)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty u^{\alpha+\beta-1} e^{-u} \mathrm{d}u \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1} (1-z)^{\beta-1}
\end{aligned}
$$

which means $Z = \frac{X}{X+Y}$ has a $\mathrm{Beta}(\alpha, \beta)$ distribution.

Question 3

(a)

For a sample $\mathbf{X} = X_1, \ldots, X_n$ with realised values $\mathbf{x} = x_1, \ldots, x_n$ and some given $\theta > 0$, where X_i is independent and Poisson-distributed with parameter $i\theta$, the likelihood is

$$
L(\theta|\mathbf{X}) = f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^n f_{X_i}(x_i|\theta) = \prod_{i=1}^n e^{-i\theta} \frac{(i\theta)^{x_i}}{x_i!}
$$

We can factorise the above as such:

$$
\begin{aligned}
L(\theta|\mathbf{X}) &= \prod_{i=1}^n e^{-i\theta} \frac{(i\theta)^{x_i}}{x_i!} \\
&= \prod_{i=1}^n \left(e^{-\theta}\right)^i \frac{i^{x_i}}{x_i!} \theta^{x_i} \\
&= \underbrace{e^{-\frac{n(n+1)}{2}\theta} \theta^{\sum x_i}}_{g(T(\mathbf{x})|\theta)} \underbrace{\prod_{i=1}^n \frac{i^{x_i}}{x_i!}}_{h(\mathbf{x})}
\end{aligned}
$$

where $T(\mathbf{X}) = \sum X_i$. The above indicates that $T(\mathbf{X})$, the sum of all observations, is a sufficient statistic for θ due to the factorisation theorem.

A Poisson random variable X_i with parameter $i\theta$ has the following probability generating function:

$$G_{X_i}(z) = \mathrm{E}\left[z^{X_i}\right] = \sum_{k=1}^{\infty} z^k e^{-i\theta} \frac{(i\theta)^k}{k!} = e^{-i\theta} \sum_{k=1}^{\infty} \frac{(zi\theta)^k}{k!} = e^{-(1-z)i\theta}$$

We know that realisations of X_i are independent, so the probability generating function for $T(\mathbf{X})$ is

$$G_{T(\mathbf{X})}(z) = \mathrm{E}\left[z^{\sum X_i}\right] = \mathrm{E}\left[\prod_{i=1}^{n} z^{X_i}\right] = \prod_{i=1}^{n} G_{X_i}(z) = \prod_{i=1}^{n} e^{-(1-z)i\theta} = e^{-(1-z)\frac{n(n+1)}{2}\theta}$$

which means $T(\mathbf{X})$ is Poisson-distributed with parameter $\frac{n(n+1)}{2}\theta$.

Assuming the likelihood function admits a local maximum, the first-order condition must be satisfied to get the maximum-likelihood estimator:

$$\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\hat{\theta}} = -\frac{n(n+1)}{2} L(\hat{\theta}|\mathbf{X}) + T(\mathbf{x}) \frac{L(\hat{\theta}|\mathbf{X})}{\hat{\theta}} = 0$$

The feasible solution satisfies

$$\frac{T(\mathbf{x})}{\hat{\theta}} - \frac{n(n+1)}{2} = 0$$

which implies $\hat{\theta} = \frac{2}{n(n+1)} T(\mathbf{x})$. The expected value of θ_{MLE} is

$$\mathrm{E}[\hat{\theta}] = \mathrm{E}\left[\frac{2}{n(n+1)} \sum_{i=1}^{n} X_i\right] = \frac{2}{n(n+1)} \sum_{i=1}^{n} \mathrm{E}[X_i] = \frac{2}{n(n+1)} \sum_{i=1}^{n} i\theta = \theta$$

which means the maximum-likelihood estimator is unbiased.

(b)

We now have $\mathbf{X} = X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Exp}(\theta)$. This means the likelihood function given $\mathbf{X} = \mathbf{x}$ is

$$L(\theta|\mathbf{X}) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum x_i}$$

which indicates that $T(\mathbf{X}) = \sum X_i$ is a sufficient statistic for θ since we can factorise the likelihood function into $T(\mathbf{x}|\theta)h(\mathbf{x})$ where $T(\mathbf{x}|\theta) = \theta^n e^{-\theta \sum x_i}$ and $h(\mathbf{x}) = 1$.

To find the distribution of $T(\mathbf{X})$, we use moment generating function of X_i:

$$M_{X_i}(t) = E\left[e^{tX_i}\right] = \int_0^{\infty} e^{tx} \theta e^{-\theta x} \mathrm{d}x = \theta \int_0^{\infty} e^{(t-\theta)x} \mathrm{d}x = \frac{\theta}{\theta - t} \text{ for } t < \theta$$

Since the X_i are i.i.d., the moment generating function of $T(\mathbf{X})$ is

$$M_{T(\mathbf{X})}(t) = M_{X_i}(t)^n = \left(\frac{\theta}{\theta - t}\right)^n$$

which means $T(\mathbf{X})$ follows a $\Gamma(n, \theta)$ distribution.

To find the maximum likelihood estimator, we impose the first-order condition again:

$$\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\hat{\theta}} = \frac{n}{\hat{\theta}} L(\hat{\theta}|\mathbf{X}) - T(\mathbf{x}) L(\hat{\theta}|\mathbf{X}) = 0 \implies \hat{\theta} = \frac{n}{T(\mathbf{x})}$$

which means the maximum-likelihood estimator is the reciprocal of the sample mean. Letting $T(\mathbf{X}) = T$, its expected value is

$$\begin{aligned}
\mathrm{E}[\hat{\theta}] = \mathrm{E}\left[\frac{n}{T}\right] &= \int_0^\infty \frac{n}{T} \frac{\theta^n}{\Gamma(n)} T^{n-1} e^{-\theta T} \mathrm{d}T \\
&= \frac{n\theta^n}{\Gamma(n)} \int_0^\infty T^{n-2} e^{-\theta T} \mathrm{d}T \\
&= \frac{n\theta^n}{\Gamma(n)} \int_0^\infty \left(\frac{u}{\theta}\right)^{n-2} e^{-u} \frac{1}{\theta} \mathrm{d}u && (u = \theta T) \\
&= \frac{n\theta}{\Gamma(n)} \int_0^\infty u^{n-2} e^{-u} \mathrm{d}u \\
&= n\theta \frac{\Gamma(n-1)}{\Gamma(n)} \\
&= \frac{n}{n-1}\theta && (\Gamma(n+1) = n\Gamma(n))
\end{aligned}$$

which means the maximum-likelihood estimator is biased. However, the estimator is consistent since $\lim_{n\to\infty} \frac{n}{n-1} = \lim_{n\to\infty} \left(1 + \frac{1}{n-1}\right) = 1$.

The biasedness of $\hat{\theta}$ is chiefly because $\mathrm{E}\left[\frac{1}{\bar{x}}\right]$ is not generally equal to $\frac{1}{\mathrm{E}[\bar{x}]}$ due to Jensen's inequality. So a good guess for ψ might be $\psi = h(\theta) = \frac{1}{\theta}$. And because h is injective and $\hat{\theta}$ maximises the likelihood function, $\hat{\psi} = h(\hat{\theta})$ also maximises the likelihood function across all values of ψ. In this case $\hat{\psi} = \frac{1}{\hat{\theta}} = \frac{T(\mathbf{x})}{n} = \bar{x}$, and $\mathrm{E}[\bar{x}] = \frac{1}{\theta} = \psi$ which means $\hat{\theta}$ is unbiased.

Question 4

With $\mathbf{X} = X_1, \ldots, X_n \overset{\text{iid}}{\sim} U[\theta, 2\theta]$, given some realised values $\mathbf{x}$, the likelihood function is

$$L(\theta|\mathbf{x}) = \mathbb{1}\left(\theta \leq \min\{\mathbf{x}\}\right) \mathbb{1}\left(2\theta \geq \max\{\mathbf{x}\}\right) \theta^{-n}$$

For the proposed estimator $\tilde{\theta}$, we have

$$\mathrm{E}[\tilde{\theta}] = \mathrm{E}\left[\frac{2}{3}X_1\right] = \frac{2}{3}\int_\theta^{2\theta} \frac{x}{\theta}\mathrm{d}x = \frac{2}{3}\left[\frac{x^2}{2\theta}\right]_\theta^{2\theta} = \frac{2}{3}\frac{3}{2}\theta = \theta$$

From the likelihood function we derived, $T(\mathbf{X}) = (\min\{\mathbf{X}\}, \max\{\mathbf{X}\})$ is a minimal sufficient statistic for θ. This is true since, given any two sample points $\mathbf{x}$ and $\mathbf{y}$, the likelihood ratio

$$\frac{L(\theta|\mathbf{X} = \mathbf{x})}{L(\theta|\mathbf{X} = \mathbf{y})} = \frac{\mathbb{1}\left(\theta \leq \min\{\mathbf{x}\}\right) \mathbb{1}\left(2\theta \geq \max\{\mathbf{x}\}\right)}{\mathbb{1}\left(\theta \leq \min\{\mathbf{y}\}\right) \mathbb{1}\left(2\theta \geq \max\{\mathbf{y}\}\right)}$$

is only constant as a function of θ if $T(\mathbf{x}) = T(\mathbf{y})$, which implies that $T(\mathbf{X}) = (\min\{\mathbf{X}\}, \max\{\mathbf{X}\})$ is a minimal sufficient statistic for θ.

The Rao-Blackwell theorem implies that $\hat{\theta} = \mathrm{E}[\tilde{\theta}|T(\mathbf{X})]$ is a uniformly better estimator of θ. This is equal to

$$\begin{aligned}
\mathrm{E}\left[\frac{2}{3}X_1 \,\middle|\, T(\mathbf{X})\right] &= \frac{2}{3}\mathrm{E}[X_1|\min\{\mathbf{X}\} = x_{min}, \max\{\mathbf{X}\} = x_{max}] \\
&= \frac{2}{3}\Big\{\mathrm{E}[X_1|\min\{\mathbf{X}\} = x_{min}, \max\{\mathbf{X}\} = x_{max}, X_1 = x_{min}] \times \Pr(X_1 = x_{min}) \\
&\quad + \mathrm{E}[X_1|\min\{\mathbf{X}\} = x_{min}, \max\{\mathbf{X}\} = x_{max}, X_1 = x_{max}] \times \Pr(X_1 = x_{max}) \\
&\quad + \mathrm{E}[X_1|\min\{\mathbf{X}\} = x_{min}, \max\{\mathbf{X}\} = x_{max}, x_{min} < X_1 < x_{max}] \\
&\qquad \times \Pr(x_{min} < X_1 < x_{max})\Big\} \\
&= \frac{2}{3}\left\{x_{min} \times \frac{1}{n} + x_{max} \times \frac{1}{n} + \frac{x_{min} + x_{max}}{2} \times \left(1 - \frac{2}{n}\right)\right\} \\
&= \frac{2}{3}\frac{2x_{min} + 2x_{max} + (n-2)(x_{min} + x_{max})}{2n} \\
&= \frac{x_{min} + x_{max}}{3}
\end{aligned}$$

Question 5

With $\mathbf{X} = X_1, \ldots, X_n \overset{\text{iid}}{\sim} U[0, \theta]$, the likelihood function is

$$L(\theta|\mathbf{X}) = \mathbb{1}(\theta \geq \max\{\mathbf{x}\})\theta^{-n}$$

which is maximised at $\hat{\theta} = \max\{\mathbf{x}\}$.

The distribution of $Y = \frac{\hat{\theta}}{\theta}$ follows

$$\begin{aligned}
F_Y(y) &= \Pr\left(\frac{\max\{\mathbf{X}\}}{\theta} \leq y\right) \\
&= \Pr(\max\{\mathbf{X}\} \leq \theta y) \\
&= \prod_{i=1}^{n} \Pr(X_i \leq \theta y) \\
&= \left(\frac{\theta y}{\theta}\right)^n \\
&= y^n
\end{aligned}$$

for $y \in [0, 1]$. For a $100(1-\alpha)\%$ confidence interval for θ, we might consider $C(\mathbf{X}) = [C_L, C_U]$ such that

$$\Pr(Y \in C(\mathbf{X})) = \Pr(C_L \leq Y \leq C_U) = C_U^n - C_L^n = 1 - \alpha$$

The above equation is just-identified if we fix $C_U = 1$, which is justified since Y is at most 1. We then have $C_L = \alpha^{\frac{1}{n}}$, and

$$1 - \alpha = \Pr\left(\alpha^{\frac{1}{n}} \leq \frac{\hat{\theta}}{\theta} \leq 1\right) = \Pr\left(\theta \leq \frac{\hat{\theta}}{\alpha^{\frac{1}{n}}} \leq \frac{\theta}{\alpha^{\frac{1}{n}}}\right) = \Pr\left(\hat{\theta} \leq \theta \leq \frac{\hat{\theta}}{\alpha^{\frac{1}{n}}}\right)$$

where the last equality uses $\Pr(\hat{\theta} \leq \theta) = 1$ and $\Pr\left(\theta \leq \frac{\theta}{\alpha^{\frac{1}{n}}}\right) = 1$ since $\alpha^{\frac{1}{n}} \leq 1$. We thus have a one-sided $100(1-\alpha)\%$ confidence interval for θ based on $\hat{\theta}$: $C(\mathbf{X}) = \left[\hat{\theta}, \frac{\hat{\theta}}{\alpha^{\frac{1}{n}}}\right]$.

Question 6

For S and C to be 95% confidence sets, we must have $\Pr((\theta_1, \theta_2) \in S) = \Pr((\theta_1, \theta_2) \in C) = 0.95$. Checking this,

$$\begin{aligned}
\Pr((\theta_1, \theta_2) \in S) &= \Pr(|\theta_1 - X_1| \leq 2.236) \times \Pr(|\theta_2 - X_2| \leq 2.236) \qquad \text{(Independence)} \\
&= \Pr(-2.236 \leq X_1 - \theta_1 \leq 2.236) \times \Pr(-2.236 \leq X_2 - \theta_2 \leq 2.236) \\
&= [\Phi(2.236) - \Phi(-2.236)] \times [\Phi(2.236) - \Phi(-2.236)] \\
&= [\Phi(2.236) - (1 - \Phi(2.236))]^2 \\
&= \left[2 \times \frac{1 + \sqrt{0.95}}{2} - 1\right]^2 \\
&= 0.95 \\
\Pr((\theta_1, \theta_2 \in C)) &= \Pr\left((\theta_1 - X_1)^2 + (\theta_2 - X_2)^2 \leq 5.991\right) \\
&= F_{\chi^2(2)}(5.991) \qquad = 0.95
\end{aligned}$$

since a squared $N(0,1)$ variable has a χ^2 distribution with 1 degree of freedom, and sums of two independent χ^2 variables with m and n degrees of freedom have a χ^2 distribution with $m + n$ degrees of freedom.

A sensible criterion for choosing between different confidence sets might be that we pick the set which is most conservative in that it covers the smallest partition of the parameter space. S is a square with length 2×2.236 and therefore has an area of $(2 \times 2.236)^2 \approx 20.00$, while C has radius $\sqrt{5.991}$ and therefore has an area of $\pi\sqrt{5.991}^2 \approx 18.82$. Therefore C is the more conservative confidence set.

Question 7

By Bayes's theorem,

$$\begin{aligned}
\pi_{\lambda|X}(1|x) &= \frac{\pi_{X|\lambda}(x|1)\pi_\lambda(1)}{\pi_X(x)} \\
&= \frac{\pi_{X|\lambda}(x|1)\pi_\lambda(1)}{\pi_{X|\lambda}(x|1)\pi_\lambda(1) + \pi_{X|\lambda}(x|1.5)\pi_\lambda(1.5)} \\
&= \frac{0.4 \times \prod_{i=1}^{5} e^{-1}\frac{1^{x_i}}{x_i!}}{0.4 \times \prod_{i=1}^{5} e^{-1}\frac{1^{x_i}}{x_i!} + 0.6 \times \prod_{i=1}^{5} e^{-1.5}\frac{1.5^{x_i}}{x_i!}} \\
&= \frac{e^{-5}\frac{0.4}{3!\times1!\times4!\times6!\times2!}}{e^{-5}\frac{0.4}{3!\times1!\times4!\times6!\times2!} + e^{-7.5}\frac{0.6\times1.5^{3+1+4+6+2}}{3!\times1!\times4!\times6!\times2!}} \\
&\approx 0.01221
\end{aligned}$$

Going through the same steps, we have

$$\pi_{\lambda|X}(1.5|x) = \frac{e^{-7.5}\frac{0.6\times1.5^{3+1+4+6+2}}{3!\times1!\times4!\times6!\times2!}}{e^{-5}\frac{0.4}{3!\times1!\times4!\times6!\times2!} + e^{-7.5}\frac{0.6\times1.5^{3+1+4+6+2}}{3!\times1!\times4!\times6!\times2!}} \approx 0.9878$$

Question 8

(a)

If $T(X)$ is a sufficient statistic for θ, we can factorise $f_X(\cdot;\theta)$ as such:

$$f_X(x;\theta) = g(T(x);\theta)h(x)$$

The maximum-likelihood estimator must then satisfy

$$\hat{\theta}_{\text{MLE}} = \arg\max_\theta g(T(x);\theta)h(x) = \arg\max_\theta g(T(x);\theta)$$

where the second equality is true since $f_X \geq 0$, and $\hat{\theta}_{\text{MLE}}$ is unique so we can neglect cases where $h(x) = 0$ (if $h(x) = 0$ all values of θ have the same likelihood and are observationally equivalent). Therefore $\hat{\theta}_{\text{MLE}}$ depends on the sample only through $T(X)$.

(b)

By Bayes's theorem, the posterior density function is

$$\pi_{\theta|X}(\theta|x) = \frac{\pi_{X|\theta}(x|\theta)\pi_\theta(\theta)}{\int_\Theta \pi_{X|\theta}(x|\theta)\pi_\theta(\theta)\mathrm{d}\theta} = \frac{g(T(x)|\theta)h(x)\pi_\theta(\theta)}{\int_\Theta g(T(x)|\theta)h(x)\pi_\theta(\theta)\mathrm{d}\theta} = \frac{g(T(x)|\theta)\pi_\theta(\theta)}{\int_\Theta g(T(x)|\theta)\pi_\theta(\theta)\mathrm{d}\theta}$$

where Θ is the parameter space. The above also depends on the sample only through $T(X)$, so any unique minimiser of the expected value of any loss function under the posterior distribution will depend on the sample only through $T(X)$.

Question 9

Again, Bayes's theorem gives us

$$f_{\theta|\mathrm{X}}(\theta|\mathbf{x}) = \frac{f_{\mathrm{X}|\theta}(\mathbf{x}|\theta)f(\theta)}{\int_0^\infty f_{\mathrm{X}|\theta}(\mathbf{x}|\theta)f(\theta)\mathrm{d}\theta} = \frac{\left(\prod_{i=1}^n \mathbb{1}_{\{0\leq x_i\leq 1\}}\theta x_i^{\theta-1}\right)\frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta}}{\int_0^\infty \left(\prod_{i=1}^n \mathbb{1}_{\{0\leq x_i\leq 1\}}\theta x_i^{\theta-1}\right)\frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta}\mathrm{d}\theta}$$
$$= \frac{\left(\prod_{i=1}^n x_i\right)^{\theta-1}\theta^{n+\alpha-1}e^{-\lambda\theta}}{\int_0^\infty \left(\prod_{i=1}^n x_i\right)^{\theta-1}\theta^{n+\alpha-1}e^{-\lambda\theta}\mathrm{d}\theta}$$

provided $0 \leq x_{min} \leq x_{max} \leq 1$. Under the posterior distribution, the expected value of the quadratic loss function given some estimator $\hat{\theta}$ is

$$\mathrm{E}[L(\hat{\theta})] = \frac{1}{\int_0^\infty \left(\prod_{i=1}^n x_i\right)^{\theta-1}\theta^{n+\alpha-1}e^{-\lambda\theta}\mathrm{d}\theta}\int_0^\infty \left(\prod_{i=1}^n x_i\right)^{\theta-1}\theta^{n+\alpha-1}e^{-\lambda\theta}\left(\hat{\theta}-\theta\right)^2\mathrm{d}\theta$$

The term outside the integral is always positive, so we can focus on minimising the integral. Also, the endpoints of the integral are independent of $\hat{\theta}$, so differentiation carries through the integral by the Leibniz integral rule. Assuming the integral is convex in $\hat{\theta}$, we can find a $\hat{\theta}_{\text{Bayes}}$ that satisfies the first-order condition:

$$\int_0^\infty \left(\prod_{i=1}^n x_i\right)^{\theta-1}\theta^{n+\alpha-1}e^{-\lambda\theta}\left(\hat{\theta}_{\text{Bayes}}-\theta\right)\mathrm{d}\theta = 0 \implies \hat{\theta}_{\text{Bayes}} = \frac{\int_0^\infty \left(\prod_{i=1}^n x_i\right)^{\theta-1}\theta^{n+\alpha}e^{-\lambda\theta}\mathrm{d}\theta}{\int_0^\infty \left(\prod_{i=1}^n x_i\right)^{\theta-1}\theta^{n+\alpha-1}e^{-\lambda\theta}\mathrm{d}\theta}$$

If nothing went wrong above, and if this is the most the problem can be simplified, then we might have to solve for this numerically.

Question 10

The probability generating function for X_{ni} is

$$G_{X_{ni}}(z) = \mathrm{E}\left[z^{X_{ni}}\right] = p_n z + 1 - p_n = 1 + p_n(z-1)$$

which means the probability generating function for S_n is

$$G_{S_n}(z) = [1 + p_n(z-1)]^n$$

Taking the limit as n goes to infinity, we have

$$\lim_{n\to\infty} G_{S_n}(z) = \lim_{n\to\infty}[1 + p_n(z-1)]^n = \lim_{n\to\infty}\left[1 + \frac{np_n(z-1)}{n}\right]^n = e^{\lambda(z-1)}$$

which would show that S_n tends to a Poisson distribution as $n \to \infty$ (though I'm not sure how to justify the last equality).

Question 11

X follows a covariance-stationary first-order autoregressive structure: the mean and variance of X_i are 0 and 1 for all i. It seems like a good guess that a sufficient statistic will include the sample autocovariance at lag 1. We denote by $\mathbf{X}_{(k)}$ the sample from X_1 to X_k. From the definition of conditional probability, we can express $f_{\mathbf{X}_{(k)}}(\mathbf{x}_{(k)}) = f_{\mathbf{X}_{(k-1)}}(\mathbf{x}_{(k-1)}) \times f_{X_k|\mathbf{X}_{(k-1)}}(x_k|\mathbf{x}_{(k-1)})$. Also, the distribution of X_k only depends on $\mathbf{X}_{(k-1)}$ through X_{k-1}, so $f_{X_k|\mathbf{X}_{(k-1)}}(x_k|\mathbf{x}_{(k-1)}) = f_{X_k|X_{k-1}}(x_k|x_{k-1})$. Applying this recursively gives us

$$
\begin{aligned}
L(\theta|\mathbf{X}) = f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{X}_{(n-1)}}(\mathbf{x}_{(n-1)}) \times f_{X_n|X_{n-1}}(x_n|x_{n-1}) \\
&= f_{\mathbf{X}_{(n-2)}}(\mathbf{x}_{(n-2)}) \times f_{X_{n-1}|X_{n-2}}(x_{n-1}|x_{n-2}) \times f_{X_n|X_{n-1}}(x_n|x_{n-1}) \\
&\ \ \vdots \\
&= f_{X_1}(x_1) \times f_{X_2|X_1}(x_2|x_1) \times \ldots \times f_{X_n|X_{n-1}}(X_n|X_{n-1})
\end{aligned}
$$

Since $X_1 = \varepsilon_1$, we have $f_{X_1}(x_1) = \phi(x_1)$. Also, given $X_{k-1} = x_{k-1}$, the event $X_k = x_k$ is equivalent to the event $Y_k = x_k - \theta x_{k-1}$, where $Y_k = \sqrt{1-\theta^2} \times \varepsilon_k \sim N(0, 1-\theta^2)$. So we have

$$
\begin{aligned}
L(\theta|\mathbf{X}) &= \phi(x_1) \times f_Y(x_2 - \theta x_1) \times \ldots \times f_Y(x_n - \theta x_{n-1}) \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \times \frac{1}{\sqrt{2\pi\,(1-\theta^2)}} e^{-\frac{(x_2-\theta x_1)^2}{2(1-\theta^2)}} \times \ldots \times \frac{1}{\sqrt{2\pi(1-\theta^2)}} e^{-\frac{(x_n-\theta x_{n-1})^2}{2(1-\theta^2)}} \\
&= (2\pi)^{-\frac{n}{2}} (1-\theta^2)^{-\frac{n-1}{2}} \exp\left(-\frac{x_1^2}{2} - \sum_{i=2}^{n} \frac{(x_i - \theta x_{i-1})^2}{2(1-\theta^2)}\right) \\
&= (2\pi)^{-\frac{n}{2}} (1-\theta^2)^{-\frac{n-1}{2}} \exp\left(-\frac{x_1^2(1-\theta^2) + \sum_{i=2}^{n} (x_i - \theta x_{i-1})^2}{2(1-\theta^2)}\right) \\
&= (2\pi)^{-\frac{n}{2}} (1-\theta^2)^{-\frac{n-1}{2}} \exp\left(-\frac{x_1^2(1-\theta^2) + \sum_{i=2}^{n} \left(x_i^2 - 2\theta x_i x_{i-1} + \theta^2 x_{i-1}^2\right)}{2(1-\theta^2)}\right) \\
&= (2\pi)^{-\frac{n}{2}} (1-\theta^2)^{-\frac{n-1}{2}} \exp\left(-\frac{\sum_{i=1}^{n} x_i^2 - 2\theta \sum_{i=2}^{n} x_i x_{i-1} + \theta^2 \left(\sum_{i=2}^{n} x_{i-1}^2 - x_1^2\right)}{2(1-\theta^2)}\right) \\
&= (2\pi)^{-\frac{n}{2}} (1-\theta^2)^{-\frac{n-1}{2}} \exp\left(-\frac{\sum_{i=1}^{n} x_i^2 - 2\theta \sum_{i=2}^{n} x_i x_{i-1} + \theta^2 \sum_{i=3}^{n} x_{i-1}^2}{2(1-\theta^2)}\right)
\end{aligned}
$$

which indicates $T(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=2}^{n} X_i X_{i-1}, \sum_{i=3}^{n} X_{i-1}^2\right)$ is a sufficient statistic for θ by the factorisation theorem.

Question 12

If $\hat{\theta}$ and U are correlated, then we can have an estimator $\tilde{\theta} = \hat{\theta} + \alpha U$, and we have

$$
\operatorname{Var}[\tilde{\theta}] = \operatorname{Var}[\hat{\theta}] + \alpha^2 \operatorname{Var}[U] + 2\alpha \operatorname{Cov}[\hat{\theta}, U]
$$

which is smaller than $\mathrm{Var}[\hat{\theta}]$ if we choose α to satisfy

$$\alpha^2 \mathrm{Var}[U] + 2\alpha \mathrm{Cov}[\hat{\theta}, U] < 0 \implies \begin{cases} \alpha \in \left(-\frac{2\mathrm{Cov}[\hat{\theta},U]}{\mathrm{Var}[U]}, 0\right) & \text{if } \mathrm{Cov}[\hat{\theta}, U] > 0 \\ \alpha \in \left(0, -\frac{2\mathrm{Cov}[\hat{\theta},U]}{\mathrm{Var}[U]}\right) & \text{if } \mathrm{Cov}[\hat{\theta}, U] < 0 \end{cases}$$

so uncorrelatedness is necessary.

For sufficiency, if $\hat{\theta}$ is uncorrelated with any U, then for any $\tilde{\theta}$ where $\mathrm{E}[\tilde{\theta}] = \theta$,

$$\mathrm{Var}[\tilde{\theta}] = \mathrm{Var}[\hat{\theta} + \underbrace{\tilde{\theta} - \hat{\theta}}_{\mathrm{E}[\tilde{\theta}-\hat{\theta}]=0}] = \mathrm{Var}[\hat{\theta}] + \mathrm{Var}[\tilde{\theta} - \hat{\theta}] + 2\,\mathrm{Cov}[\hat{\theta}, \tilde{\theta} - \hat{\theta}] \geq \mathrm{Var}[\hat{\theta}]$$

which implies $\hat{\theta}$ is a UMVU estimator.