Stats 743b - Assignment 1

Tommy Flynn (400121496)

May 3, 2023

Question 1:

Fix $n \in \mathbb{N}$, $\mathcal{I} = 1, 2, \ldots, n$, and let $\{X_{k:n}\}_{k \in \mathcal{I}}$ be the sequence of order statistics from a random sample of size n from an absolutely continuous distribution with cdf $F(x)$ and pdf $f(x)$. For this to be a Markov chain we need to confirm it has the Markov property:

$$\forall k \in \mathcal{I}\backslash\{1\} \quad P(X_{k:n} \leq x_k | X_{1:n} \leq x_1, X_{2:n} \leq x_2, \ldots, X_{k-1:n} \leq x_{k-1}) = P(X_{k:n} \leq x_k | X_{k-1:n} \leq x_{k-1})$$

We will establish this property by showing the underlying conditional densities are the same. Fix $k \in \mathcal{I} \setminus \{1\}$ then we have:

$$f_{k|1,\ldots,k-1:n}(x_k | x_1, \ldots, x_{k-1})$$

$$= \frac{f_{1,\ldots,k:n}(x_1, \ldots, x_k)}{f_{1,\ldots,k-1:n}(x_1, \ldots, x_{k-1})}$$

$$= \frac{\frac{n!}{(n-k)!}\cancel{f(x_1)\cdots} f(x_{k-1}) f(x_k)[1 - F(x_k)]^{n-k}}{\frac{n!}{(n-k+1)!}\cancel{f(x_1)\cdots} f(x_{k-1})[1 - F(x_{k-1})]^{n-k+1}}$$

$$= \frac{\frac{n!}{(n-k)!} f(x_{k-1}) f(x_k)[1 - F(x_k)]^{n-k}}{\frac{n!}{(n-k+1)!} f(x_{k-1})[1 - F(x_{k-1})]^{n-k+1}} \cdot \frac{\frac{[F(x_{k-1})]^{k-2}}{(k-2)!}}{\frac{[F(x_{k-1})]^{k-2}}{(k-2)!}}$$

$$= \frac{\frac{n!}{(k-2)!(n-k)!} f(x_{k-1}) f(x_k)[1 - F(x_k)]^{n-k}[F(x_{k-1})]^{k-2}}{\frac{n!}{(k-2)!(n-k+1)!} f(x_{k-1})[1 - F(x_{k-1})]^{n-k+1}[F(x_{k-1})]^{k-2}}$$

$$= \frac{f_{k-1,k:n}(x_{k-1}, x_k)}{f_{k-1:n}(x_{k-1})}$$

$$= f_{k|k-1:n}(x_k | x_{k-1})$$

$\implies \forall n \in \mathbb{N}, \forall k \in \mathcal{I} \setminus \{1\}$
$P(X_{k:n} \leq x_k | X_{1:n} \leq x_1, X_{2:n} \leq x_2, \ldots, X_{k-1:n} \leq x_{k-1}) = P(X_{k:n} \leq x_k | X_{k-1:n} \leq x_{k-1})$

Therefore, the sequence of order statistics from an absolutely continuous distribution forms a Markov chain.

Question 2:

Consider the scenario from Question 1 with two-point distribution taking on values $a < b$ with probabilities π and $1 - \pi$ respectively. Suppose we have the first $k - 1$ order statistics in the sequence and want to know the value of $X_{k:n}$. Because we have a two-point distribution, it is either the case that $X_{k-1:n} = a$ or $X_{k-1:n} = b$. If $X_{k-1:n} = a$ then all previous order statistics must have a value of a. But then the previous values are irrelevant for $X_{k:n}$ since the next value will be either a or b with the usual probabilities regardless. So we have:

Case $X_{k-1:n} = a$:
$P(X_{k:n} = a | X_{1:n} = a, \ldots, X_{k-1:n} = a) = P(X_{k:n} = a | X_{k-1:n} = a) = \pi$
$P(X_{k:n} = b | X_{1:n} = a, \ldots, X_{k-1:n} = a) = P(X_{k:n} = b | X_{k-1:n} = a) = 1 - \pi$

If $X_{k-1:n} = b$ then there is a sequence (possible empty) of a's followed by a sequence of at least one b. But then the previous values are irrelevant for $X_{k:n}$ since the next value will have to be b. So we have:

Case $X_{k-1:n} = b$:
$P(X_{k:n} = a | X_{1:n} = a, \ldots, X_{i:n} = a, X_{i+1:n} = b, \ldots, X_{k-1:n} = b) = P(X_{k:n} = a | X_{k-1:n} = b) = 0$
$P(X_{k:n} = b | X_{1:n} = a, \ldots, X_{i:n} = a, X_{i+1:n} = b, \ldots, X_{k-1:n} = b) = P(X_{k:n} = b | X_{k-1:n} = b) = 1$

Therefore, the sequence of order statistics for a two-point distribution forms a Markov chain.

Question 3:

Consider the scenario from Question 1 and 2 with a discrete distribution taking on at least 3 values. Let $x, y, z \in \mathcal{S}$ with $x < y < z$ and $k \in \mathcal{I} \setminus \{1\}$. Then we have:

$$P(X_{k+1:n} = z | X_{k-1:n} = x, X_{k:n} = y)$$

$$= \frac{f_{k-1,k,k+1:n}(x, y, z)}{f_{k-1,k:n}(x, y, z)}$$

$$= \frac{\frac{n!}{(n-k-1)!(k-2)!} \cancel{\int_{F(x-)}^{F(x)} (u_{k-1})^{k-2}} \, du_{k-1} \int_{F(y-)}^{F(y)} \, du_k \int_{F(z-)}^{F(z)} (1 - u_{k+1})^{n-k-1} \, du_{k+1}}{\frac{n!}{(n-k)!(k-2)!} \cancel{\int_{F(x-)}^{F(x)} (u_{k-1})^{k-2}} \, du_{k-1} \int_{F(y-)}^{F(y)} (1 - u_k)^{n-k} \, du_k}$$

Multiplying by $\dfrac{\frac{1}{k-1} \int_{F(y-)}^{F(y)} (u_k)^{k-1} \, du_k \int_{F(y-)}^{F(y)} (u_k)^{k-1} (1 - u_k)^{n-k} \, du_k}{\frac{1}{k-1} \int_{F(y-)}^{F(y)} (u_k)^{k-1} \, du_k \int_{F(y-)}^{F(y)} (u_k)^{k-1} (1 - u_k)^{n-k} \, du_k}$ gives us:

$$= P(X_{k+1:n} = z | X_{k:n} = y) \frac{[F(y) - F(y-)] \int_{F(y-)}^{F(y)} (u_k)^{k-1} (1 - u_k)^{n-k} \, du_k}{\int_{F(y-)}^{F(y)} (u_k)^{k-1} \, du_k \int_{F(y-)}^{F(y)} (1 - u_k)^{n-k} \, du_k}$$

$= P(X_{k+1:n} = z | X_{k:n} = y) \cdot m$ where $m \in (0, 1)$ since the numerator is strictly less than the denominator (proved in general below $*$). Therefore, for discrete distributions with at least 3 points we have

$$P(X_{k+1:n} = z | X_{k-1:n} = x, X_{k:n} = y) = mP(X_{k+1:n} = z | X_{k:n} = y) \quad m \in (0, 1)$$
$$\implies P(X_{k+1:n} = z | X_{k-1:n} = x, X_{k:n} = y) < P(X_{k+1:n} = z | X_{k:n} = y)$$
$$\implies \text{ the sequence of order statistics fails to form a Markov chain}$$

Proof of $*$: $m \in (0, 1)$

Let $a < b \in [0, 1]$ and $r < s \in \mathbb{N}$. Given $(b - a) \int_a^b u^r (1 - u)^{s-r} \, du - \int_a^b u^r \, du \int_a^b (1 - u)^{s-r} \, du$ by the mean value theorem there exists $c \in (a, b)$ such that $\int_a^b u^r \, du = (b - a)c^r$. Thus, we have

$$\begin{aligned}
&(b - a) \int_a^b u^r (1 - u)^{s-r} \, du - \int_a^b u^r \, du \int_a^b (1 - u)^{s-r} \, du \\
&= (b - a) \int_a^b u^r (1 - u)^{s-r} \, du - (b - a)c^r \int_a^b (1 - u)^{s-r} \, du \\
&= (b - a) \int_a^b (u^r - c^r)(1 - u)^{s-r} \, du \\
&< (b - a) \int_a^b (u^r - c^r)(1 - c)^{s-r} \, du \qquad \text{(since } (1 - u)^{s-r} \text{ is monotonically decreasing with } 0 < c < 1\text{)} \\
&= (b - a)(1 - c)^{s-r}[(b - a)c^r - (b - a)c^r] \\
&= 0 \\
&\implies (b - a) \int_a^b u^r (1 - u)^{s-r} \, du < \int_a^b u^r \, du \int_a^b (1 - u)^{s-r} \, du \implies m \in (0, 1)
\end{aligned}$$

Question 4:

Let $U_{1:n} < \cdots < U_{n:n}$ be the order statistics from a random sample of size n from Uniform$(0,1)$.

a) Let $1 \le r < s \le n$, $V_1 = \frac{U_{r:n}}{U_{s:n}}, V_2 = U_{s:n}$. We have

$$f_{s:n}(x_s) = \text{Beta}(s, n-s+1) = \frac{1}{B(s,n-s+1)} x_s^{s-1}(1-x_s)^{n-s}$$

and

$$f_{r,s:n}(x_r, x_s) = \text{BivBeta}(r, s-r, n-s+1) = \frac{1}{B(r,s-r,n-s+1)} x_r^{r-1}(x_s - x_r)^{s-r-1}(1-x_s)^{n-s}$$

The transformation is $U_{r:n} = V_1V_2, U_{s:n} = V_2 \implies J = \left| \begin{pmatrix} v_2 & v_1 \\ 0 & 1 \end{pmatrix} \right| = v_2$

$\implies f_{V_1,V_2}(v_1, v_2) = \frac{1}{B(r,s-r,n-s+1)}(v_1v_2)^{r-1}(v_2 - v_1v_2)^{s-r-1}(1-v_2)^{n-s} \cdot v_2$

$= \frac{1}{B(r,s-r,n-s+1)} \cdot \frac{(s-1)!}{(s-1)!} \cdot \left[v_1^{r-1}(1-v_1)^{s-r-1}\right] \left[v_2 v_2^{r-1} v_2^{s-r-1}(1-v_2)^{n-s}\right]$

$= \frac{1}{B(r,s-r)} \left[v_1^{r-1}(1-v_1)^{s-r-1}\right] \cdot \frac{1}{B(s,n-s+1)} \left[v_2^{s-1}(1-v_2)^{n-s}\right]$

We see that the joint distribution V_1 and V_2 is the product of the marginal distributions. This means that V_1 and V_2 are independent and that $V_1 \sim \text{Beta}(r, s-r)$ and $V_2 \sim \text{Beta}(s, n-s+1)$

b) $V_1 \sim \text{Beta}(r, s-r)$ and $V_2 \sim \text{Beta}(s, n-s+1)$.

c) We will calculate the covariance using the independence of V_1, V_2 and their Beta distributions.

$\text{Cov}[U_{r:n}, U_{s:n}] = \text{Cov}[V_1V_2, V_2]$

$= \text{E}[V_1V_2^2] - \text{E}[V_1V_2]\text{E}[V_2]$

$= \text{E}[V_1]\text{E}[V_2^2] - \text{E}[V_1]\text{E}[V_2]^2$

$= \text{E}[V_1]\text{Var}[V_2]$

$= \frac{r}{s} \cdot \frac{s(n-s+1)}{(n+1)^2(n+2)}$

$= \frac{r}{n+1} \frac{(n+1-s)}{n+1} \frac{1}{n+2}$

$= \frac{r}{n+1}(1 - \frac{s}{n+1})\frac{1}{n+2}$

Question 5:

First, define the variables $V_1 = \frac{U_{1:n}}{U_{2:n}}, \dots, V_{n-1} = \frac{U_{n-1:n}}{U_{n:n}}, V_n = U_{n:n}$. Proceeding as in Question 4, we have the transformation $U_{1:n} = V_1 V_2 \cdots V_n, U_{2:n} = V_2 \cdots V_n, \dots, U_{n:n} = V_n$ with Jacobian:

$$J = \left| \begin{pmatrix} V_2 \cdots V_n & & \cdots & & \\ 0 & V_3 \cdots V_n & & \cdots & \\ 0 & 0 & V_4 \cdots V_n & & \cdots \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & 0 & V_n \end{pmatrix} \right| = V_2 V_3^2 V_4^3 \cdots V_n^{n-1}$$

Thus, the joint distribution is given by:

$f_{V_1,\dots,V_n}(v_1, \dots, v_n) = n! f(V_1 \cdots V_n) \cdots f(V_n) \cdot V_2 V_3^2 V_4^3 \cdots V_n^{n-1}$

$= (1) \cdot (2V_2) \cdot (3V_3^2) \cdots (nV_n^{n-1})$

We see that the joint distribution of $V_1, \dots V_n$ is the product of the marginal distributions. Therefore, $V_1, \dots, V_n$ are independent with distributions $\text{Beta}(1,1), \dots, \text{Beta}(n,1)$ respectively.

Question 6:

Let $X_1, X_2, \ldots, X_{n+1}$ be independent standard exponential random variables with $S_{n+1} = \sum_{i=1}^{n+1} X_i$.

a) Define the variables $V_1 = \frac{X_1}{S_{n+1}}, \ldots, V_n = \frac{X_1+X_2+\cdots+X_n}{S_{n+1}}$. Then we have the transformation $X_1 = V_1 V_{n+1}, X_2 = (V_2 - V_1)V_{n+1}, \ldots, X_n = (V_n - V_{n-1})S_{n+1}, X_{n+1} = V_{n+1} \implies J = V_{n+1}^n$.

$$f_{V_1,\ldots,V_{n+1}}(v_1, \ldots, v_n) = e^{-v_{n+1}(v_1+(v_2-v_1)+(v_3-v_2)+\cdots+1)} v_{n+1}^n = e^{-v_{n+1}} v_{n+1}^n$$

Now the joint density of $(\frac{X_1}{S_{n+1}}, \ldots, \frac{X_1+\cdots X_n}{S_{n+1}})$ is

$$f_{V_1,\ldots,V_n}(v_1, \ldots, v_n) = \int_0^\infty e^{-v_{n+1}} v_{n+1}^n \, dv_{n+1} = \Gamma(n+1) = n! \quad 0 \leq v_1 \leq v_2 \cdots \leq v_n \leq 1$$

b) We see that the joint density is also that of the order statistics from a random sample from a Uniform$(0,1)$ distribution. More specifically we have $\frac{S_i}{S_{n+1}} \stackrel{d}{=} U_{i:n}$.

Question 7:

Let $X_1, X_2, \cdots, X_n$ be a random sample from a power function distribution with density function $f(x) = \frac{2x}{\theta^2}$, $0 < x < \theta$. Let $X_{1:n} < \cdots < X_{n:n}$ be the corresponding order statistics.

a) The joint density of $(X_{1:n}, \cdots, X_{n:n})$ is given by: $f_{1,\ldots,n:n}(x_1, \cdots, x_n) = n!\frac{2^n x_1 \cdots x_n}{\theta^{2n}}$.

b) Let $X_{n:n}$ be our statistic, then $p(\mathbf{X}|\theta) = \frac{2^n x_1 \cdots x_n}{\theta^{2n}}$ and $q(t|\theta) = n\frac{2t}{\theta^2}\left[\frac{t^2}{\theta^2}\right]^{n-1} = \frac{2nt^{2n-1}}{\theta^{2n}} \implies \frac{p(\mathbf{X}|\theta)}{q(t|\theta)}$ is free of θ since the factors of $\left(\frac{1}{\theta^{2n}}\right)$ cancel $\implies X_{n:n}$ is a sufficient statistic.

c) We have $\mathbb{E}[X_{n:n}] = \int_0^\theta u \cdot 2n\frac{u^{2n-1}}{\theta^{2n}}\,du = \theta\frac{2n}{2n+1} \implies \hat{\theta} = \frac{2n+1}{2n}X_{n:n}$ is an unbiased estimator of θ in terms of $X_{n:n}$.

d) Let $g(\cdot)$ be any measurable function, then $\mathbb{E}_\theta[g(X_{n:n})] = 0 \iff \int_0^\theta g(u)2n\frac{u^{2n-1}}{\theta^2}\,du = 0 \iff \int_0^\theta g(u)u^{2n-1}\,du = 0$ applying Libeniz rule $\implies g(\theta)\theta^{2n-1} = 0 \implies g(\theta) = 0, \forall\theta > 0 \implies g = 0$ almost everywhere $\implies P_\theta(g(X_{n:n}) = 0) = 1 \implies X_{n:n}$ is a complete statistic. This means that $X_{n:n}$ is a complete sufficient statistic and we have unbiased estimator $\hat{\theta}$ which is a function of $X_{n:n}$ so using the Lehmann–Scheffe theorem we can say that it is the uniformly minimum variance unbiased estimator of θ.

Question 8:

In continuation of Question 7, suppose we have $P_n = \frac{X_{n:n}}{\theta}$

a) P_n is a function of the sample and the parameter so we must check that its distribution is free of θ for it to be a pivot. Let $Y = \frac{X_{n:n}}{\theta} \iff \theta Y = X_{n:n} \implies J = \theta \implies g(y) = \frac{2n(\theta y)^{2n-1}}{\theta^{2n}} \cdot \theta = 2ny^{2n-1} \sim \text{Beta}(2n, 1)$. Since P_n is $\text{Beta}(2n, 1)$ it can serve as a pivot.

b) To construct a $100(1-\alpha)\%$ confidence interval for θ we take the $\alpha/2$ and $1-\alpha/2$ critical values of $\text{Beta}(2n, 1)$ as a and b and calculate the following

$1 - \alpha = P(a \leq P_n \leq b)$

$= P(a \leq \frac{X_{n:n}}{\theta} \leq b)$

$= P(\frac{1}{b} \leq \frac{\theta}{X_{n:n}} \leq \frac{1}{a})$

$= P(\frac{X_{n:n}}{b} \leq \theta \leq \frac{X_{n:n}}{a})$

c) The mean and variance of $X_{n:n}$ are given by

$\mathbb{E}[X_{n:n}] = \theta \frac{2n}{2n+1}$

$\mathbb{E}[X_{n:n}^2] = \int_0^\theta u^2 \cdot 2n \frac{u^{2n-1}}{\theta^{2n}} \, du = \theta^2 \frac{n}{n+1}$

$\text{Var}[X_{n:n}] = \theta^2 \frac{n}{n+1} - \theta^2 \frac{4n^2}{(2n+1)^2} = \theta^2 \frac{n}{(n+1)(2n+1)^2}$

with $\lim_{n\to\infty} \mathbb{E}[X_{n:n}] \to 0$ and $\text{Var}[X_{n:n}] \to 0$. Let $Z_n = n(\theta - X_{n:n})$, then

$P(Z_n \leq z)$

$= P(n(\theta - X_{n:n}) \leq z)$

$= P(\theta - X_{n:n} \leq \frac{z}{n})$

$= P(X_{n:n} \geq \theta - \frac{z}{n})$

$= 1 - (\frac{\theta - z/n}{\theta})^{2n}$ using cdf $(\frac{t}{\theta})^{2n}$

$= 1 - ((1 - \frac{z/\theta}{n})^n)^2$

Therefore, $\lim_{n\to\infty} F_{Z_n}(z) = 1 - e^{-2z/\theta} \implies \frac{Z_n}{\theta} = n(1 - P_n) \stackrel{d}{=} \text{Exponential}(\lambda = 2)$.

Question 9:

Take two independent random draws from Uniform$(0, 1)$ to create interval (a, b) and another two indepdent random draws from Uniform$(0, 1)$ to create interval (c, d).

(a/b) Denote the four draws by u_1, u_2, u_3, u_4 and the corresponding order statistics $u_{1:4}, u_{2:4}, u_{3:4}, u_{4:4}$. There are a total of 4! mappings from u_i to $u_{i:n}$. The intervals do not intersect if they lie beside each other i.e. if u_1, u_2 gets mapped to $u_{1:4}, u_{2:4}$ or $u_{3:4}, u_{4:4}$ while u_3, u_4 gets mapped to $u_{3:4}, u_{4:4}$ or $u_{1:4}, u_{2:4}$ respectively. So there is a factor of 2×2 for the individual arrangements of u_1, u_2 and u_2, u_3 and a factor of 2 for the groupings lying left or right. Therefore, the probability that the intervals do no intersect is $\frac{8}{24} = \frac{1}{3}$. Since they are complements, the probability that the intervals overlap is $1 - \frac{1}{3} = \frac{2}{3}$.

Question 10:

Let $X_1, \ldots, X_n$ be a random sample from Uniform$(-\theta, \theta)$ for $\theta > 0$ and let $X_{1:n} < \cdots < X_{n:n}$ be the corresponding order statistics.

a) The joint density of $(X_1, \cdots, X_n)$ is given by

$$f_{1,\ldots,n}(x_1, \cdots, x_n) = \frac{1}{(2\theta)^n}, -\theta \leq x_i \leq \theta$$

and the joint density of $(X_{1:n}, X_{n:n})$ is

$$f_{1,n:n}(x_1, x_n) = n(n-1)\frac{1}{4\theta^2} \cdot (\frac{x_n - x_1}{2\theta})^{n-2} = n(n-1)\frac{(x_n - x_1)^{n-2}}{(2\theta)^n} - \theta \leq x_1 \leq x_n \leq \theta$$

Thus, the ratio $\frac{p(\mathbf{X}|\theta)}{q(t|\theta)}$ is free of θ since the factors of $(2\theta)^n$ cancel each other $\implies$ $(X_{1:n}, X_{n:n})$ is jointly sufficient for θ.

b) The log likelihood function is given by $\ell = \log\left(\frac{1}{(2\theta)^n}\right) = -n[\log(2) + \log(\theta)] \implies \frac{\partial \ell}{\partial \theta} = \frac{-n}{\theta} \implies$ monotonically decreasing so MLE of θ is $\hat{\theta} = \max\{-X_{1:n}, X_{n:n}\} = \max_{1 \leq i \leq n}\{|X_i|\} = Y$.

c) By independence we have $F_Y(y) = P(Y \leq y) = [P(-y \leq X_1 \leq y)]^n = \left[\frac{y+\theta}{2\theta} - \frac{-y+\theta}{2\theta}\right]^n = (\frac{y}{\theta})^n$. So the exact distribution of $\hat{\theta}$ is

$$f_Y(y) = \frac{ny^{n-1}}{\theta^n}, \quad 0 \leq y \leq \theta$$

Question 11:

a) Continuing from Question 10, we have the mean and variance of $\hat{\theta}$ as

$\mathbb{E}[\hat{\theta}] = \mathbb{E}[Y] = \frac{n}{\theta^n}\int_0^\theta y^n\, dy = \frac{n}{\theta^n}\left[\frac{y^{n+1}}{n+1}\right]_0^\theta = \theta\frac{n}{n+1}$

$\mathbb{E}[Y^2] = \frac{n}{\theta^n}\int_0^\theta y^{n+1}\, dy = \frac{n}{\theta^n}\left[\frac{y^{n+2}}{n+2}\right]_0^\theta = \theta^2\frac{n}{n+2}$

$\mathrm{Var}[\hat{\theta}] = \theta^2\frac{n}{n+2} - (\theta\frac{n}{n+1})^2 = \theta^2\frac{n}{(n+2)(n+1)^2}$

We see that as $n \to \infty, \mathbb{E}[\hat{\theta}] \to \theta$ and $\mathrm{Var}[\hat{\theta}] \to 0$.

b) Define measurable function $g(x_{1:n}, x_{n:n}) = x_n + x_1$ then

$\mathbb{E}_\theta[g(x_1, x_n)] = \int_{-\theta}^\theta \int_{-\theta}^{x_n} (x_n + x_1) \cdot n(n-1)\frac{(x_n-x_1)^{n-2}}{(2\theta)^n}\, dx_1 dx_n$

$= \int_{-\theta}^\theta \frac{x_n(x_n+\theta)^{n-1}}{n-1} - \frac{\theta(x_n+\theta)^{n-1}}{n-1} + \frac{(x_n+\theta)^n}{n(n-1)}\, dx_n$

$= \frac{\theta(2\theta)^n}{n(n-1)} - \frac{(2\theta)^n+1}{n(n-1)(n+1)} - \frac{\theta(2\theta)^n}{n(n-1)} + \frac{(2\theta)^n+1}{n(n-1)(n+1)}$

$= 0 \quad \forall\theta > 0$

Note that $P_\theta(g(x_1, x_n) = 0) = P_\theta(x_n + x_1 = 0) = 0, \forall\theta > 0$ almost surely. Thus, $(X_{1:n}, X_{n:n})$ cannot be complete by definition.

Question 12:

Let $X_1, \dots, X_n$ be a random sample from Poisson(λ) with $n > 3$.

a) We have that $T = \sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$. For measurable g, $\mathbb{E}_\lambda[g(T)] = 0 \iff e^{-n\lambda}\sum_{t=0}^\infty g(t)(n\lambda)^t/t! = 0 \iff g(0)n^0/0! + g(1)n^1/1! + \cdots + \cdots = 0 \iff g(0) = g(1) = \cdots = 0\,\forall\lambda \in (0,\infty) \implies P_\lambda[g(T) = 0] = 1$ so T is a complete statistic for λ. We also have $\frac{p(\mathbf{X}|\lambda)}{q(t|\lambda)} = \frac{e^{-n\lambda}\lambda^t \prod_{i=1}^n 1/x_i!}{e^{-n\lambda}n^t\lambda^t/t!}$ which is free of λ showing T is a complete sufficient statistic.

Now consider $a \in \mathbb{N}$ with

$\mathbb{E}\left[\left(1 - \frac{a}{n}\right)^T\right] = \sum_{k=0}^\infty \left(1 - \frac{a}{n}\right)^k e^{-n\lambda}\frac{(n\lambda)^k}{k!}$

$$= e^{-n\lambda}\sum_{k=0}^\infty \frac{\left(\left(1 - \frac{a}{n}\right)n\lambda\right)^k}{k!}$$

$= e^{-n\lambda}e^{n\lambda(1-a/n)}$

$= e^{-a\lambda}$

We are interested in $\theta = e^{-\lambda}(1 - e^{-\lambda})^2 = e^{-\lambda} - 2e^{-2\lambda} + e^{-3\lambda}$. Define the estimator

$$\hat{\theta} = \left(1 - \frac{1}{n}\right)^T - 2\left(1 - \frac{2}{n}\right)^T + \left(1 - \frac{3}{n}\right)^T$$

Then from above we have

$$\mathbb{E}[\hat{\theta}] = \theta$$

Since $\hat{\theta}$ is an unbiased estimator of θ and is a function of the complete and sufficient statistic T, by the Lehmann-Schefe theorem it is the uniformly minimum variance unbiased estimator of θ.

b) Using the same argument we have

$\text{Var}\left[\left(1 - \frac{a}{n}\right)^T\right] = \mathbb{E}\left[\left(1 - \frac{a}{n}\right)^{2T}\right] - \mathbb{E}\left[\left(1 - \frac{a}{n}\right)^T\right]^2 = e^{-n\lambda}e^{n\lambda(1-a/n)^2} - e^{-2a\lambda} = e^{-2a\lambda}(e^{\frac{a^2\lambda}{n}} - 1)$. By independence $\text{Var}[\hat{\theta}] = e^{-2\lambda}(e^{\lambda/n} - 1) + 4e^{-4\lambda}(e^{4\lambda/n} - 1) + e^{-3\lambda}(e^{9\lambda/n} - 1)$. As $n \to \infty$ we have $\text{Var}[\hat{\theta}] \to 0$.

Question 13:

Let $X_1, \ldots, X_n$ be the lifetimes of n components from Exponential(θ) distribution, with $n > 2$. Further, let S denote the lifetime of a series system of 2 such components and P denote the lifetime of a parallel system of 2 such components.

a) Let the series system lifetime be given by $S = X_i + X_j, i \neq j$. Since we have a sum of two iid exponential random variables we know $S \sim$ Gamma$(2, \theta)$. Then

$P(S > t)$

$= \int_t^\infty \frac{1}{\theta^2} s e^{-s/\theta}\, ds$

$= -[\frac{1}{\theta} s e^{-s/\theta} + e^{-s/\theta}]\Big|_t^\infty$

$= \frac{1}{\theta} t e^{-t/\theta} + e^{-t/\theta} = R_S$

Take $\hat{R}_S = I(S > t)$ as an unbiased estimator for the quantity above and condition it on the complete and sufficient statistic $Z_n = \sum_{i=1}^n X_i \sim$ Gamma(n, θ) as follows

$R_S^* = \mathbb{E}[\hat{R}_S | Z_n = z] = \frac{P(S>t, Z_n=z)}{f_{Z_n}(z)}$

$= \frac{\int_t^z \frac{1}{\theta^2} s e^{-s/\theta} \frac{1}{\theta^{n-2}\Gamma(n-2)} e^{-(z-s)/\theta} (z-s)^{n-3}\, ds}{\frac{1}{\theta^n \Gamma(n)} e^{-z/\theta} z^{n-1}}$

$= -\frac{(n-1)(n-2)}{z^{n-1}} \left[\frac{s(z-s)^{n-2}}{n-2} + \frac{(z-s)^{n-1}}{(n-1)(n-2)} \right] \Big|_t^z$

$= \frac{(n-1)t}{z}(1 - \frac{t}{z})^{n-2} + (1 - \frac{t}{z})^{n-1}$

Since R_S^* is a unbiased estimator of R_S and is a function of the complete and sufficient statistic Z_n, by the Lehmann-Schefe theorem it is the uniformly minimum variance unbiased estimator of R_S.

b) Let the parallel system lifetime be given by $P = \min(X_i, X_j), i \neq j$. Since we have the minimum of two iid exponential random variables we know $P \sim$ Exponential(2θ). Then

$P(P \leq t^*)$

$= 1 - e^{-2t^*/\theta}$

Take $\hat{R}_P = I(P \leq t)$ as an unbiased estimator for the quantity above and condition it on Z_n

$R_P^* = \mathbb{E}[\hat{R}_P | Z_n = z] = 1 - \frac{P(P>t^*, Z_n=z)}{f_{Z_n}(z)}$

$= 1 - \frac{\int_{t^*}^z \frac{1}{2\theta} e^{-2p/\theta} \frac{1}{\theta^{n-1}\Gamma(n-1)} e^{-(z-2p)/\theta} (z-2p)^{n-2}\, dp}{\frac{1}{\theta^n \Gamma(n)} e^{-z/\theta} z^{n-1}}$

$$= 1 + \frac{(n-1)}{z^{n-1}} \left[\frac{(z-2p)^{n-1}}{n-1} \right] \Big|_{t^*}^{z}$$

$$= 1 - (1 - \tfrac{2t^*}{z})^{n-1}$$

Since R_P^* is a unbiased estimator of R_P and is a function of the complete and sufficient statistic Z_n it is the uniformly minimum variance unbiased estimator of R_P.

c) $P(R_P \leq r)$

$$= P(1 - (1 - \tfrac{2t^*}{z})^{n-1} \leq r)$$

$$= P(1 - 2t^*/z \geq (1-r)^{1/(n-1)})$$

$$= P(z \leq \tfrac{2t^*}{1-(1-r)^{1/(n-1)}}) \sim \text{Gamma}(n, \tfrac{2t^*}{1-(1-r)^{1/(n-1)}})$$

Question 14:

Let X_1, X_2, X_3 be iid random variables from a Normal$(\mu, 1)$ distribution.

a) The Fisher information in X_i about μ is given by:

$$\mathcal{I}(\mu) = -\mathbb{E}\left[\frac{\partial^2}{\partial \mu^2} \log f(X_i|\mu)\right], \quad i = 1, 2, 3$$

We have: $\log(f(x_i|\mu) = \frac{-1}{2}\log(2\pi) - \frac{1}{2}(x_i - \mu)^2$

$\implies \frac{\partial}{\partial \mu} \log f(x_i|\mu) = x_i - \mu$

$\implies \frac{\partial^2}{\partial \mu^2} \log f(x_i|\mu) = -1$

$\implies \mathcal{I}_i(\mu) = 1, \quad i = 1, 2, 3$

b) For $(X_i, X_j), i = 1, 2, 3, i \neq j$ we have: $\log(f(x_i, x_j|\mu) = -\log(2\pi) - \frac{1}{2}[(x_i - \mu)^2 + (x_j - \mu)^2]$

$\implies \frac{\partial}{\partial \mu} \log f(x_i, x_j|\mu) = x_i + x_j - 2\mu$

$\implies \frac{\partial^2}{\partial \mu^2} \log f(x_i, x_j|\mu) = -2$

$\implies \mathcal{I}_{i,j}(\mu) = 2, \quad i = 1, 2, 3, i \neq j$

c) For (X_1, X_2, X_3) we have: $\log(f(x_1, x_2, x_3|\mu) = -\frac{3}{2}\log(2\pi) - \frac{1}{2}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2]$

$\implies \frac{\partial}{\partial \mu} \log f(x_1, x_2, x_3|\mu) = x_1 + x_2 + x_3 - 3\mu$

$\implies \frac{\partial^2}{\partial \mu^2} \log f(x_1, x_2, x_3|\mu) = -3$

$\implies \mathcal{I}_{1,2,3}(\mu) = 3$

From these calculations, we can conjecture that for $X = (X_1, X_2, \ldots, X_n)$ iid Normal$(\mu, 1)$, the Fisher information in X about μ is $\mathcal{I}_X(\mu) = n\mathcal{I}_1(\mu) = n$. That is, each new observation contributes the same amount of information.

Question 15:

Continuing from Question 14, let $X_{1:3} < X_{2:3} < X_{3:3}$ be the order statistics from X_1, X_2, X_3.

a) For convenience let Φ and ϕ denote the standard normal cdf and pdf. Then the Fisher information in $X_{i:n}$ about $\mu, i = 1, 2, 3$ is as follows:

$$f_{i:n}(x_i;\mu) = \tfrac{n!}{(n-i)!(i-1)!}[\Phi(x_i-\mu)]^{i-1}[1-\Phi(x_i-\mu)]^{n-i}\phi(x_i-\mu)$$

$$\implies \log f_{i:n}(x_i;\mu) = \log(c) + (i-1)\log\Phi(x_i-\mu) + (n-i)\log[1-\Phi(x_i-\mu)] + \log\phi(x_i-\mu)$$

$$\implies \tfrac{\partial}{\partial\mu} f_{i:n}(x_i;\mu) = \tfrac{-(i-1)}{\Phi(x_i-\mu)}\phi(x_i-\mu) + \tfrac{n-i}{1-\Phi(x_i-\mu)}\phi(x_i-\mu) + (x_i-\mu)$$

$$\implies \tfrac{\partial^2}{\partial\mu^2} f_{i:n} = -(i-1)\left[\tfrac{(x_i-\mu)\phi(x_i-\mu)}{\Phi(x_i-\mu)} + \left(\tfrac{\phi(x_i-\mu)}{\Phi(x_i-\mu)}\right)^2\right] + (n-i)\left[\tfrac{(x_i-\mu)\phi(x_i-\mu)}{1-\Phi(x_i-\mu)} - \left(\tfrac{\phi(x_i-\mu)}{1-\Phi(x_i-\mu)}\right)^2\right] - 1$$

$$\implies \mathcal{I}_{i:n}(\mu) = \mathbb{E}\left[1 + (i-1)\left[\tfrac{(x_i-\mu)\phi(x_i-\mu)}{\Phi(x_i-\mu)} + \left(\tfrac{\phi(x_i-\mu)}{\Phi(x_i-\mu)}\right)^2\right] - (n-i)\left[\tfrac{(x_i-\mu)\phi(x_i-\mu)}{1-\Phi(x_i-\mu)} - \left(\tfrac{\phi(x_i-\mu)}{1-\Phi(x_i-\mu)}\right)^2\right]\right]$$

b) To calculate the expectations in R you can use simulation or numerical integration. I decided to use simulation by generated 10^5 realizations of the expression and took the column means returning:

$$\mathcal{I}_{1:3}(\mu) \approx 1.826589, \quad \mathcal{I}_{2:3}(\mu) \approx 2.229209, \quad \mathcal{I}_{3:3}(\mu) \approx 1.827742$$

From this we can conjecture that for Normal$(\mu, 1)$ the Fisher information of individual order statistics is symmetric about the median and hits a maximum at the median. We also note that an individual order statistic contains more information than a random observation.

c) For $(X_{i:n}, X_{j:n}), 1 \le i < j \le 3, (z_i = x_i - \mu)$:

$$f_{i,j:n}(x_i, x_j;\mu) = c[\Phi(z_i)]^{i-1}[\Phi(z_j)-\Phi(z_i)]^{j-i-1}[1-\Phi(z_j)]^{n-j}\phi(z_i)\phi(z_j)$$

$$\implies \tfrac{\partial}{\partial\mu}\log f_{i,j:n} = -(i-1)\tfrac{\phi(z_i)}{\Phi(z_i)} - (j-i-1)\tfrac{\phi(z_j)-\phi(z_i)}{\Phi(z_j)-\Phi(z_i)} + (n-j)\tfrac{\phi(z_j)}{1-\Phi(z_j)} + x_i + x_j - 2\mu$$

$$\implies \tfrac{\partial^2}{\partial\mu^2} = -(i-1)\left[\tfrac{z_i\phi(z_i)}{\Phi(z_i)} + \left(\tfrac{\phi(z_i)}{\Phi(z_i)}\right)^2\right] - (j-i-1)\left[\tfrac{z_j\phi(z_j)-z_i\phi(z_i)}{\Phi(z_j)-\Phi(z_i)} - \left(\tfrac{\phi(z_j)-\phi(z_i)}{\Phi(z_j)-\Phi(z_i)}\right)^2\right] + (n-j)\left[\tfrac{z_j\phi(z_j)}{1-\Phi(z_j)} - \left(\tfrac{\phi(z_j)}{1-\Phi(z_j)}\right)^2\right] - 2$$

$$\implies \mathcal{I}_{i,j:n}(\mu) = \mathbb{E}\Bigg[2 + (i-1)\left[\frac{z_i\phi(z_i)}{\Phi(z_i)} + \left(\frac{\phi(z_i)}{\Phi(z_i)}\right)^2\right] - (j-i-1)\left[\frac{z_j\phi(z_j)-z_i\phi(z_i)}{\Phi(z_j)-\Phi(z_i)} - \left(\frac{\phi(z_j)-\phi(z_i)}{\Phi(z_j)-\Phi(z_i)}\right)^2\right] + (n-j)\left[\frac{z_j\phi(z_j)}{1-\Phi(z_j)} - \left(\frac{\phi(z_j)}{1-\Phi(z_j)}\right)^2\right]\Bigg]$$

d) After running the simulations described above we get:

$$\mathcal{I}_{1,2:3}(\mu) = 2.614908 \quad \mathcal{I}_{1,3:3}(\mu) = 2.331147 \quad \mathcal{I}_{2,3:3}(\mu) = 2.614301$$

Again we see symmetry about the median and more information is gained by k order statistics than k random observations. What is notable here is that the majority of the information is stored in the first half (or last half) of the data.

e) The Fisher information in $(X_{1:3}, X_{2:3}, X_{3:3})$ about μ is given by $\mathcal{I}_{1,2,3:n}(\mu) = \mathcal{I}_{1,2,3}(\mu) = 3$. That is, the information contained in all of the order statistics is exactly the information contained in the original sample. This is intuitively true as they are the exact same observations but one could quickly check the details by realizing the joint pdf of all order statistics is $n!$ times the original joint pdf as it is an $n!$ to 1 transformation. Thus, it is clear that the Fisher information is identical.

Question 16:

Let $X_1, \ldots, X_n$ be a random sample from Normal(μ, σ^2), z_p the upper pth quantile of standard normal and $\xi_p = \mu + \sigma z_p$ the upper pth quantile of Normal(μ, σ^2).

a) One possible pivot in terms of $\bar{X}$ and S^2 is

$$T = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma - \sqrt{n} z_p}{S/\sigma} = \frac{\sqrt{n}}{S}\left[\bar{X} - \xi_p\right] \sim t(n-1, \lambda)$$

Note that if Z is standard normal and V is chi-squared with ν degrees of freedom, then $T = \frac{Z+\lambda}{\sqrt{V/\nu}}$ follows a non-central t-distribution with ν degrees of freedom and non-central parameter λ. With this in mind, we see that our pivot follows a non-central t-distribution with $\nu = n - 1$ and $\lambda = -\sqrt{n} z_p$.

b) To calculate a $100(1-\alpha)\%$ confidence interval for ξ_p using the pivot, we first calculate the critical values $t_1 = t_{\alpha/2, n-1, \lambda}$ and $t_2 = t_{1-\alpha/2, n-1, \lambda}$. Then we have:

$1 - \alpha = P[t_1 \leq T \leq t_2]$

$= P[t_1 \leq \frac{\sqrt{n}}{S}\left[\bar{X} - \xi_p\right] \leq t_2]$

$= P[\bar{X} - t_2 \frac{S}{\sqrt{n}} \leq \xi_p \leq \bar{X} - t_1 \frac{S}{\sqrt{n}}]$

c) To determine the percentage points of the involved pivot we must calculate t_1 and t_2. Thus, we need either a table or software to compute the quantiles $1 - \alpha/2$ and $\alpha/2$ for the non-central t-distribution based on $n - 1$ degrees of freedom and non-central parameter $-\sqrt{n} z_p$.

d) Let $n = 20$, then we have the following percentage points using R:

$$\alpha = 0.1,\ \ p = 0.1 \implies t_{0.05,19,5.73} = 3.84, \quad t_{0.95,19,5.73} = 8.61$$
$$\alpha = 0.05,\ \ p = 0.1 \implies t_{0.025,19,5.73} = 3.51, \quad t_{0.975,19,5.73} = 9.30$$

$$\alpha = 0.1,\ \ p = 0.25 \implies t_{0.05,19,3.02} = 1.35, \quad t_{0.95,19,3.02} = 5.22$$
$$\alpha = 0.05,\ \ p = 0.25 \implies t_{0.025,19,3.02} = 1.04 \quad t_{0.975,19,3.02} = 5.72$$

Question 17:

Let $X_1, \ldots, X_n$ be a random sample from an absolutely continuous distribution with cdf $F(x)$ and pdf $f(x)$, and let ξ_p be the upper pth quantile. Using the order statistics $X_{1:n} < \cdots < X_{n:n}$ let the $100(1-\alpha)\%$ confidence interval for ξ_p be $(X_{r:n}, X_{s:n})$ for $1 \leq r < s \leq n$.

a) Given sample size n, the interval $(X_{r:n}, X_{s:n})$ is a valid $100(1-\alpha)\%$ confidence interval for ξ_p if we would choose $r < s$ such that $P(X_{r:n} \leq \xi_p \leq X_{s:n}) = \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} \approx 1 - \alpha$. This is an approximation due to the discrete nature of the binomial distribution. If multiple choices satisfy this property, then we choose among the minimum difference $s - r$.

b) Let $n = 20$. I wrote R code to check

$$(1-\alpha) - c \leq \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} \leq (1-\alpha) + c$$

for small value c.

For $p = 0.1$, the largest probability was $\approx 88\%$ achieved by the pairs $\{(r,s) | r = 1, s = 7, \ldots, 20\}$. Thus, 88% confidence interval is as close as we can get to 90% and 95% for $p = 0.1$.

For $p = 0.25, \alpha = 0.1$, the pairs $\{(r,s) | r = 3, s = 10, \ldots, 20\} \cup \{(1,8)\}$ satisfy $90\% \pm 1\%$ confidence intervals. For $p = 0.25, \alpha = 0.05$, the pairs $\{(1,9), (2,9), (2,10)\}$ satisfy $95\% \pm 2\%$ confidence intervals.

c) The selection criterion will be minimum width while prioritizing percentages over the requested value.

In the case of $p = 0.1$, we are forced to choose $(X_{1:20}, X_{7:20})$ for both $\alpha = 0.1, 0.05$ as it caps at 88%. For $p = 0.25, \alpha = 0.1$, we choose $(X_{3:20}, X_{11:20})$ as it is the smallest $s - r$ in that set whose probability is larger than 90%. Likewise, for $p = 0.25, \alpha = 0.05$, we choose $(X_{1:20}, X_{9:20})$.

Question 18:

After running $N = 1000$ repetitions of a random sample of size $n = 20$ from a standard normal distribution, we have the following averages for $p = 0.1$, $\alpha = 0.05$. Note that $z_{0.1} = -1.281552$ for checking coverage probability.

Parametric average width: $= 1.29$
Parametric coverage probability: $= 96\%$
Non-parametric average width: $= 1.45$
Non-parametric coverage probability: $= 90\%$

The Parametric confidence interval is superior with a smaller average width while having a larger coverage probability. From Question 17, we expect the non-parametric coverage probability to be 88% which explains why it is much lower.

Question 19:

Repeating the comparison in Question 18 for Logistc, Laplace, and Normal Outlier gives the following:

a)
Parametric average width: = 3.21833
Parametric coverage probability: = 21%
Non-parametric average width: = 2.84
Non-parametric coverage probability: = 87%

b)
Parametric average width: = 2.03
Parametric coverage probability: = 66%
Non-parametric average width: = 2.47
Non-parametric coverage probability: = 87%

c)
Parametric average width: = 1.70
Parametric coverage probability: = 78%
Non-parametric average width: = 2.24
Non-parametric coverage probability: = 94%

d) It seems that the parametric method works well the more normal the data is while the non-parametric method is useful for general distributions especially without much computation.

Question 20:

a) Assuming Normal(μ, σ^2) for the Darwin data, we have estimates

$$\bar{x} = 20.93333 \quad s^2 = 1424.638$$

b) In R, we sort the data to get the observed order statistics and apply the normal cdf using the estimates in part a) to get the empirical probabilities. Plotting these against the theoretical probabilities $\frac{1}{16}, \ldots, \frac{15}{16}$ gives us the p-p plot below. Next we apply the the standard normal quantile function to the theoretical probabilities to get the theoretical quantiles and plot the order statistics agains them resulting in the q-q plot below. Both plots follow the reference lines except two points on the left which lie below. This indicates the underlying distribution is likely normal but there is some left skewness perhaps caused by outliers.

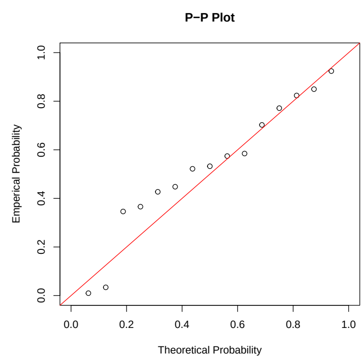

Figure 1: P-P Plot

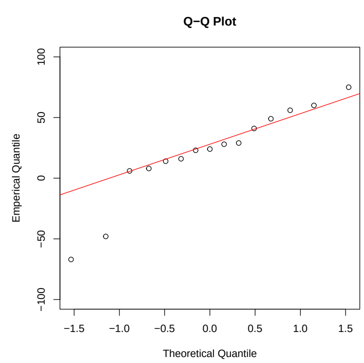

Figure 2: Q-Q Plot

c) The 10% trimmed mean is $\bar{x}_{0.10} = 26.72727$. Those two particular low values have been removed and thus the mean has moved up quite a bit from 20.9 to 26.7. The p-p and q-q plots are much more well behaved suggesting that normality is valid and those points were outliers.

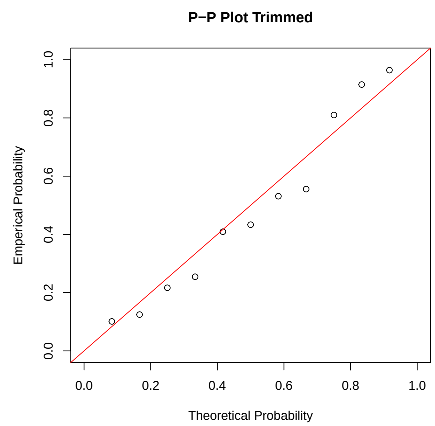

Figure 3: P-P Plot Trimmed

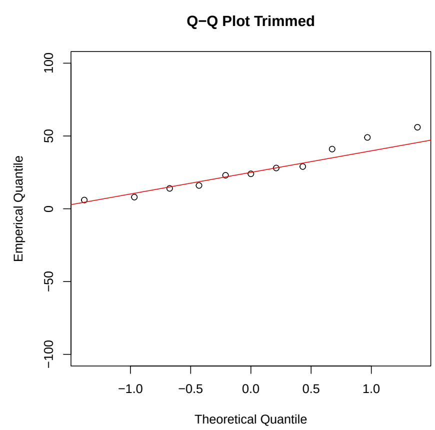

Figure 4: Q-Q Plot Trimmed