THE UNIVERSITY OF NEW SOUTH WALES

DEPARTMENT OF STATISTICS

Solutions to selected exercises for MATH5905, Statistical Inference

Part two: Data reduction. Sufficient statistics. Classical estimation

Question 1: a) Denoting $T = \sum_{i=1}^n X_i$, you can factorise $L(\mathbf{X}, \mu)$ with

$$h(\mathbf{X}) = \exp(-\frac{1}{2}\sum_{i=1}^n X_i^2),$$

$$g(T, \mu) = \exp(-\frac{n}{2}\mu^2)\exp(T\mu)\frac{1}{(\sqrt{2\pi})^n}.$$

b) Denoting $T = \sum_{i=1}^n X_i^2$, you can factorise $L(\mathbf{X}, \sigma^2)$ with

$$h(\mathbf{X}) = 1, \qquad g(T, \sigma^2) = \exp(-\frac{1}{2\sigma^2}T)\frac{1}{(\sqrt{2\pi}\sigma)^n}.$$

c) For a point x and a set A, we use the notation

$$I_A(x) = I(x \in A) = \begin{cases} 1 \text{ if } x \text{ is in A,} \\ 0 \text{ if } x \text{ is not in } A \end{cases}$$

Then

$$L(\mathbf{X}, \theta) = \prod_{i=1}^n I_{(\theta,\theta+1)}(x_i) = I_{(\theta,\theta+1)}(x_{(n)})I_{(\theta,\theta+1)}(x_{(1)}) = I_{(x_{(n)}-1,\infty)}I_{(-\infty,x_{(1)})}(\theta).$$

Hence $T = \left\{ \begin{array}{c} X_{(1)} \\ X_{(n)} \end{array} \right\}$ can be taken as sufficient vector-statistic.

d) Denoting $T = \sum_{i=1}^n X_i$, you can factorise $L(\mathbf{X}, \lambda)$ with

$$g(T, \lambda) = \exp(-n\lambda)\lambda^T \qquad \text{and} \qquad h(\mathbf{X}) = \frac{1}{\prod_{i=1}^n X_i!}.$$

According to the factorisation criterion, T is sufficient.

Now, using the definition **and** noting that $T = \sum_{i=1}^n X_i \sim \text{Po}(n\lambda)$ we have:

$$P(\mathbf{X} = \mathbf{x}|T = t) = \frac{P(\mathbf{X} = \mathbf{x} \bigcap T = t)}{P(T = t)} = \begin{cases} 0 & \text{if} \quad \sum_{i=1}^n x_i \neq t \\ \frac{P(\mathbf{X}=\mathbf{x})}{P(\sum_{i=1}^n X_i = t)} & \text{if} \quad \sum_{i=1}^n x_i = t \end{cases}$$

Since $\sum_{i=1}^n X_i \sim \text{Po}(n\lambda)$, the latter expression on the right can be shown to be equal to

$$\frac{t!}{n^t \prod_{i=1}^n x_i!}$$

and it does not depend on λ. Hence $T = \sum_{i=1}^n X_i$ is sufficient according to the original definition of sufficiency.

Question 2: For $S = X_1 + X_2 + X_3$ we already know ($n = 3$ is a special case of the general case considered at the lecture.) To show that $T = X_1X_2 + X_3$ is **not** sufficient, it suffices to show that, say, $f_{(X_1,X_2,X_3|T=1)}(0,0,1|1)$ **does** depend on p. You can see that

$$\begin{aligned} f_{(X_1,X_2,X_3|T=1}(0,0,1|1)) &= \frac{P(X_1 = 0 \bigcap X_2 = 0 \bigcap X_3 = 1 \bigcap T = 1)}{P(T = 1)} \\ &= \frac{(1-p)^2 p}{3p^2(1-p) + p(1-p)^2} \\ &= \frac{1-p}{1+2p} \end{aligned}$$

Hence $T = X_1X_2 + X_3$ is not sufficient for p.

Question 3: We will show that $T_1 = X_1 + X_2$ is sufficient but $T_2 = X_1X_2$ is **not** sufficient. By a direct check we have

$$\begin{aligned} &P(X_1 = 0 \cap X_2 = 0|X_1 + X_2 = 0) = 1, \\ &P(X_1 = 1 \cap X_2 = 0|X_1 + X_2 = 0) = P(X_1 = 1 \cap X_2 = 1|X_1 + X_2 = 0) = P(X_1 = 0 \cap X_2 = 1|X_1 + X_2 = 0) = 0 \\ &P(X_1 = 1 \cap X_2 = 0|X_1 + X_2 = 1) = \frac{\theta(4-\theta)/12}{\theta(4-\theta)/6} = \frac{1}{2} = P(X_1 = 0 \cap X_2 = 1|X_1 + X_2 = 1) \\ &P(X_1 = 0 \cap X_2 = 0|X_1 + X_2 = 1) = 0 = P(X_1 = 1 \cap X_2 = 1|X_1 + X_2 = 0) \\ &P(X_1 = 1 \cap X_2 = 1|X_1 + X_2 = 2) = \frac{\theta(\theta-1)/12}{\theta(\theta-1)/12} = 1 \\ &P(X_1 = 0 \cap X_2 = 1|X_1 + X_2 = 2) = P(X_1 = 1 \cap X_2 = 0|X_1 + X_2 = 2) = 0 \\ &P(X_1 = 0 \cap X_2 = 0|X_1 + X_2 = 2) = 0 \end{aligned}$$

and we see that in all possible cases the conditional distribution does not involve the parameter θ. However, for $T_2 = X_1X_2$ we can see by following the same pattern, that

$$P(X_1 = 1 \cap X_2 = 0|X_1X_2 = 0) = \frac{4\theta - \theta^2}{\theta - \theta^2 + 12}.$$

This clearly depends on θ hence T_2 is not sufficient.

Question 4: The conditional probability $P(\mathbf{X} = \mathbf{x}|X_1 = x_1)$ is the probability $P(X_2 = x_2 \cap \cdots \cap X_n = x_n)$ and it depends on p since for each i we have

$$P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}.$$

Question 5: We need to show that at least in some cases there is explicit dependence of the conditional distribution of the vector $\left\{ \begin{array}{c} X_1 \\ X_2 \end{array} \right\}$ given the statistic $T = X_1 + X_2$. We note that possible realisations of T are $t = 2, 3, \ldots, 2\theta$. We examine $P(\left\{ \begin{array}{c} X_1 \\ X_2 \end{array} \right\} = \left\{ \begin{array}{c} x_1 \\ x_2 \end{array} \right\} |X_1 + X_2 = x)$. Of course, if $x_1 + x_2 \neq x$, this conditional probability is zero and does not involve θ. Let us now study the case when $x_1 + x_2 = x$. We have two scenarios:
First scenario: $2 \le x \le \theta$. Then

$$P\Big(\left\{ \begin{array}{c} X_1 \\ X_2 \end{array} \right\} = \left\{ \begin{array}{c} x_1 \\ x_2 \end{array} \right\} |X_1 + X_2 = x\Big) = \frac{P(X_1 = x_1 \bigcap X_2 = x - x_1)}{\sum_{i=1}^{x-1} P(X_1 = i \bigcap X_2 = x - i)} = \frac{(1/\theta)^2}{(x-1)(1/\theta)^2} = \frac{1}{x-1}$$

which does not involve θ.
Second scenario: $\theta < x \le 2\theta$. Then:

$$P\Big(\left\{ \begin{array}{c} X_1 \\ X_2 \end{array} \right\} = \left\{ \begin{array}{c} x_1 \\ x_2 \end{array} \right\} |X_1 + X_2 = x\Big) = \frac{P(X_1 = x_1 \bigcap X_2 = x - x_1)}{\sum_{i=x-\theta}^{\theta} P(X_1 = i \bigcap X_2 = x - i)} = \frac{(1/\theta)^2}{(-x + 2\theta + 1)(1/\theta)^2} = \frac{1}{2\theta - x + 1}$$

In the second case, the conditional distribution explicitly involves θ hence $T = X_1 + X_2$ can not be sufficient for θ.

Question 6: Similar solution to **Question 4** above and we leave this as an exercise for you.

Question 7: a) The ratio takes the form

$$\frac{L(\mathbf{x},\lambda)}{L(\mathbf{y},\lambda)} = \lambda^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} \frac{\prod_{i=1}^n (y_i)!}{\prod_{i=1}^n (x_i)!}$$

and this would not depend on λ if and only if $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Hence $T = \sum_{i=1}^n X_i$ is minimal sufficient.

b)The ratio takes the form

$$\frac{L(\mathbf{x},\sigma^2)}{L(\mathbf{y},\sigma^2)} = \exp(-\frac{1}{2\sigma^2}(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2)).$$

This would not depend on σ^2 if and only if $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$. Hence $T(\mathbf{X}) = \sum_{i=1}^n X_i^2$ is minimal sufficient.

c) Similarly, $T = \prod_{i=1}^n X_i$ is minimal sufficient. We can also take $\tilde{T} = \sum_{i=1}^n logX_i$ as minimal sufficient.

d) We have

$$\frac{L(\mathbf{x},\theta)}{L(\mathbf{y},\theta)} = \frac{I_{(x_{(n)},\infty)}(\theta)}{I_{(y_{(n)},\infty)}(\theta)}.$$

This has to be considered as a function of θ for fixed $x_{(n)}$ and $y_{(n)}$. Assume that $x_{(n)} \neq y_{(n)}$ and, to be specific, let $x_{(n)} > y_{(n)}$ first. Then the ratio $\frac{L(\mathbf{x},\theta)}{L(\mathbf{y},\theta)}$ is:

- not defined if $\theta \leq y_{(n)}$,
- equal to zero when $\theta \in [y_{(n)}, x_{(n)})$.
- equal to one when $\theta > x_{(n)}$.

In other words, the ratio's value depends on the position of θ on the real axis, that is, it is a function of θ. Similar conclusion will be reached if we had $x_{(n)} < y_{(n)}$ (do it yourself). Hence, if and only if $x_{(n)} = y_{(n)}$ will the ratio not depend on θ. This implies that $T = X_{(n)}$ is minimal sufficient.

e) $T = (X_{(1)}, X_{(n)})$ is minimal sufficient. We know from 1c) that $L(\mathbf{x},\theta)$ depends on the sample via $x_{(n)}$ and $y_{(n)}$ only. If $\mathbf{x} = (x_1,\ldots,x_n)$ and $\mathbf{y} = (y_1,\ldots,y_n)$ are such that either $x_{(1)} \neq y_{(1)}$ or $x_{(n)} \neq y_{(n)}$ or both then $\frac{L(\mathbf{x},\theta)}{L(\mathbf{y},\theta)}$ will have different values in different intervals, that is, will depend on θ. For this **not** to happen, $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$ must hold.

f) Similar to e). $T = (x_{(1)}, x_{(n)})$ is minimal sufficient.

Question 8: a) Since

$$L(\mathbf{x},\theta) = \theta^n (\prod_{i=1}^n x_i)^{\theta-1}$$

we see by the factorisation criterion that $T = \prod_{i=1}^n x_i$ is sufficient. Note that $\tilde{T} = \sum_{i=1}^n \log x_i$ is also sufficient since it is an 1-to-1 transformation of T.

b) Since

$$L(\mathbf{x},\theta) = \frac{1}{(6\theta^4)^n}(\prod_{i=1}^n x_i^3) e^{-(\sum_{i=1}^n x_i)/\theta}.$$

We can factorise with $h(\mathbf{x}) = \prod_{i=1}^n x_i^3$, $g(t,\theta) = \frac{1}{(6\theta^4)^n} e^{-t/\theta}$, where $t = \sum_{i=1}^n x_i$.

Question 9, 10: Left for you as exercises. I have treated the location case for the Cauchy family in the lectures, the scale case is along the same lines.

Question 11: Parts (a) to (d) we went through during the lectures. For part (e) look at the score representation.

Question 12: Take $\hat{\tau} = I_{\{X_1=0 \bigcap X_2=0\}}(\mathbf{X})$. Then we have that $\mathbb{E}(\hat{\tau}) = e^{-2\lambda}$ (that is, $\hat{\tau}$ is unbiased for $\tau(\lambda) = e^{-2\lambda}$). Then the UMVUE would be

$$\mathbb{E}\Big(\hat{\tau}\Big|\sum_{i=1}^{n} X_i = t\Big) = 1 \cdot P\Big(\hat{\tau} = 1 | \sum_{i=1}^{n} X_i = t\Big).$$

We know that $\sum_{i=1}^{n} X_i \sim \text{Po}(n\lambda)$. The unbiased estimate is

$$\begin{aligned} a(t) &= \frac{P(X_1 = 0 \bigcap X_2 = 0 \bigcap \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} \\ &= \frac{P(X_1 = 0 \bigcap X_2 = 0 \bigcap \sum_{i=3}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} \\ &= \frac{(n-2)^t}{n^t} \\ &= (1 - \frac{2}{n})^t. \end{aligned}$$

We can check directly that this estimator is unbiased for $\tau(\lambda)$ (*although this is not necessary: we have stated a general theorem that Rao-Blackwellization preserves the unbiasedness property. I have included the calculations below just as an additional exercise:*

$$\mathbb{E}\big[a(T)\big] = \sum_{t=0}^{\infty}(1 - \frac{2}{n})^t \frac{e^{-n\lambda}(n\lambda)^t}{t!} = e^{-n\lambda} \sum_{t=0}^{\infty} \frac{[\lambda(n-2)]^t}{t!} = e^{-2\lambda}.$$

The variance given by the Cramer-Rao lower bound is:

$$\frac{(\tau'(\lambda))^2}{nI_{X_1}(\lambda)} = \frac{\lambda(-2e^{-2\lambda})^2}{n} = \frac{4\lambda e^{-4\lambda}}{n}$$

For the variance of the unbiased estimator, we have:

$$\begin{aligned} \text{Var}(a(T)) &= \sum_{t=0}^{\infty}(1 - \frac{2}{n})^{2t} \frac{e^{-n\lambda}(n\lambda)^t}{t!} - (e^{-2\lambda})^2 \\ &= e^{-n\lambda} \sum_{t=0}^{\infty} \frac{(n-2)^{2t}\lambda^t}{n^t t!} - e^{-4\lambda} \\ &= e^{-n\lambda} e^{(n-4+\frac{4}{n})\lambda} - e^{-4\lambda} \\ &= e^{-4\lambda}[e^{4\lambda/n} - 1] > 0. \end{aligned}$$

The latter value is strictly larger than the bound:

$$e^{-4\lambda}\Big[e^{4\lambda/n} - 1 - \frac{4\lambda}{n}\Big] = e^{-4\lambda}(\frac{1}{2!}(\frac{4\lambda}{n})^2 + \frac{1}{3!}(\frac{4\lambda}{n})^3 + \dots) > 0.$$

Question 13: This is again just to refresh some required, useful technical skills.

$$f_X(x) = \int_0^x 8xydy = 4x^3 \text{ if } x \text{ in (0,1) (and zero else)}$$

$$f_Y(y) = \int_y^1 8xydx = 4y - 4y^3 \text{ if } y \text{ in (0,1) (and zero else)}$$

$$f_{Y|X}(y|x) = \frac{8xy}{4x^3} = \frac{2y}{x^2} \text{ if } 0 < y < x, 0 < x < 1 \text{ (and zero else)}$$

$$a(x) = \mathbb{E}(Y|X = x) = \int_0^x y f_{Y|X}(y|x)dy = \frac{2x}{3}, 0 < x < 1$$

$$\mathbb{E}(a(X)) = \int_0^1 a(x) f_X(x) dx = \int_0^1 \frac{2x}{3} 4x^3 dx = \frac{8}{15}$$

$$\mathbb{E}(Y) = 4\int_0^1 y(y - y^3) dy = \frac{8}{15}$$

Similarly $Ea^2(x) = \frac{8}{27}, \text{Var}(a(X)) = \frac{8}{27} - (\frac{8}{15})^2 = \frac{8}{675}$

$$\mathbb{E}(Y^2) = \frac{1}{3}, \text{Var}(Y) = \frac{11}{225}$$

and we see directly that indeed $\text{Var}(a(X)) < \text{Var}(Y)$ holds.
Again note that the fact that by conditioning we reduce the variance was proved quite generally in the lectures. In this problem we are just checking that indeed $Var(a(X)) < Var(Y)$ on a particular example.

Question 14: Steps:

a) $T = \sum_{i=1}^n X_i$ is complete and sufficient for θ.
b) If $\hat{\tau} = X_1 X_2$ then $E\hat{\tau} = \theta^2$ (that is, $\hat{\tau}$ is unbiased for θ^2).
c) $a(t) = E(\hat{\tau}|T = t) = \cdots = \frac{t(t-1)}{n(n-1)}$ which is the UMVUE.

We can also check directly the unbiasedness of this estimator:

$$\begin{aligned}
\mathbb{E}(a(T)) &= \mathbb{E}[\bar{X}(\frac{n}{n-1}\bar{X} - \frac{1}{n-1})] \\
&= \frac{n}{n-1}\mathbb{E}(\bar{X})^2 - \frac{\mathbb{E}(\bar{X})}{n-1} \\
&= \frac{n}{n-1}[\text{Var}(\bar{X}) + (\mathbb{E}(\bar{X}))^2] - \frac{\theta}{n-1} \\
&= \frac{n}{n-1}(\frac{\theta(1-\theta)}{n} + \theta^2) - \frac{\theta}{n-1} \\
&= \theta^2.
\end{aligned}$$

Question 15: Since $f(x;\theta)$ is an **one-parameter exponential** family, with $d(x) = x$. Using our general statement from the lecture, we can claim that $T = \sum_{i=1}^n X_i$ is **complete and minimal sufficient** for θ. We also know that for this distribution $\mathbb{E}(X_1) = \theta, \text{Var}(X_1) = \theta^2$ holds. Let us calculate:

$$\mathbb{E}(\bar{X}^2) = \text{Var}(\bar{X}) + (\mathbb{E}(\bar{X}))^2 = \frac{\text{Var}(X_1)}{n} + (\mathbb{E}X_1)^2 = \frac{n+1}{n}\theta^2 \neq \theta^2.$$

After bias-correction, by Lehmann-Scheffe's theorem:

$$\frac{n(\bar{X})^2}{n+1} = \frac{T^2}{n(n+1)}$$

is unbiased for θ **and** since T is complete and sufficient, we conclude that $\frac{T^2}{n(n+1)}$ is UMVUE for θ^2.

Question 16: a) $T = X_{(n)}$ is complete and sufficient for θ, with

$$f_T(t) = \frac{nt^{n-1}}{\theta^n}, 0 < t < \theta.$$

Hence $\mathbb{E}(T^2) = \frac{n}{n+2}\theta^2$. Hence $T_1 = \frac{n+2}{n}T^2$ is unbiased estimator of θ^2. By Lehmann-Scheffe,

$$\frac{n+2}{n}T^2$$

is the UMVUE.

Its variance:

$$\begin{aligned}\mathbb{E}(\frac{n+2}{n}T^2)^2 - \theta^4 &= (\frac{n+2}{n})^2\mathbb{E}T^4 - \theta^4 \\ &= (\frac{n+2}{n})^2 n\int_0^\theta \frac{t^{n+3}}{\theta^n}dt - \theta^4 \\ &= \theta^4[\frac{(n+2)^2}{n}\frac{1}{n+4} - 1] \\ &= \frac{4\theta^4}{n(n+4)}.\end{aligned}$$

b) Similar to a). $\frac{n-1}{n}\frac{1}{T}$ is the UMVUE; its variance is $\frac{1}{n(n-2)\theta^2}$.

Question 17: This is a *more difficult (*) question.* It is meant to challenge the interested students.

a) The density $f(t;\theta)$ in 7a) is also called $Gamma(n,\theta)$ **density**. To show the result, we could use convolution. Reminder: the **convolution formula** for the density of the sum of two independent random variables X, Y :

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t-x)\mathrm{d}x$$

In particular, if the random variables are non-negative, the above formula simplifies to:

$$f_{X+Y}(t) = \int_0^t f_X(x)f_Y(t-x)\mathrm{d}x, \quad \text{if } t>0 \text{ (and 0 elsewhere).}$$

Applying it for the two non-negative random variables in our case, we get:

$$f_{X_1+X_2}(t) = \theta^2\int_0^t e^{-\theta x}e^{-t\theta+\theta x}dx = \theta^2 e^{-t\theta}\int_0^t dx = \theta^2 te^{-t\theta}.$$

which means that for $n=2$ the claim is proved (note that $\Gamma(2)=1$.) We apply **induction** to show the general case. Assume that for $T=\sum_{i=1}^k X_i$, the formula is also true and we want to show that then it is true for $k+1$. We apply for $\sum_{i=1}^{k+1} X_i = \sum_{i=1}^k X_i + X_{k+1}$ the convolution formula and we get:

$$f_{\sum_{i=1}^{k+1} X_i}(t) = \frac{t^k\theta^{k+1}e^{-\theta t}}{\Gamma(k+1)},$$

that is, the claim is true for $k+1$.

Note: It is possible to give an alternative proof by using the moment generating functions approach. Try it if you feel familiar enough with moment generating functions.

b) Consider the estimator $\hat{\tau} = I_{\{X_1>k\}}(\mathbf{X})$. Then, $\mathbb{E}(\hat{\tau}) = 1\cdot P(X_1>k) = \int_k^\infty \theta e^{-\theta x}dx = e^{-k\theta}$.

c) Let $T=\sum_{i=1}^n X_i$. Consider for small enough Δx_1 :

$$\begin{aligned} f_{X_1|T}(x_1|t)\Delta x_1 &= \frac{f_{X_1,T}(x_1,t)\Delta x_1\Delta t}{f_T(t)\Delta t} \\ &\approx \frac{P[x_1<X_1<x_1+\Delta x_1; t<\sum_{i=1}^n X_i<t+\Delta t]}{\frac{1}{\Gamma(n)}\theta^n t^{n-1}e^{-\theta t}\Delta t} \\ &\approx \frac{P[x_1<X_1<x_1+\Delta x_1; t-x_1<\sum_{i=2}^n X_i<t-x_1+\Delta t]}{\frac{1}{\Gamma(n)}\theta^n t^{n-1}e^{-\theta t}\Delta t} \\ &\approx \frac{P(x_1<X_1<x_1+\Delta x_1)P(t-x_1<\sum_{i=2}^n X_i<t-x_1+\Delta t)}{\frac{1}{\Gamma(n)}\theta^n t^{n-1}e^{-\theta t}\Delta t} \\ &\approx \frac{\theta e^{-\theta x_1}\frac{1}{\Gamma(n-1)}\theta^{n-1}(t-x_1)^{n-2}e^{-\theta(t-x_1)}\Delta x_1\Delta t}{\frac{1}{\Gamma(n)}\theta^n t^{n-1}e^{-\theta t}\Delta t} = (n-1)\frac{(t-x_1)^{n-2}}{t^{n-1}}\Delta x_1.\end{aligned}$$

Going to the limit as Δx_1 tends to zero, we get

$$f_{X_1|T}(x_1|t) = \frac{n-1}{t}(1-\frac{x_1}{t})^{n-2}, \quad 0 < x_1 < t < \infty.$$

Now we can find the UMVUE. It will be:

$$\mathbb{E}(I_{(k,\infty)}(X_1)|T=t) = \int_k^\infty f_{X_1|T}(x_1|t)dx_1 = \int_k^t \frac{n-1}{t^{n-1}}(t-x_1)^{n-2}dx_1 = \Big(\frac{t-k}{t}\Big)^{n-1}.$$

That is,

$$\Big(\frac{T-k}{T}\Big)^{n-1} I_{(k,\infty)}(T)$$

with $T = \sum_{i=1}^n X_i$ is the UMVUE of $e^{-k\theta}$.

Question 18: The restriction $\theta \in (0, 1/5)$ makes sure that the probabilities calculated as a function of θ indeed belong to $[0,1]$. Let $\mathbb{E}_\theta h(X) = 0$ for all $\theta \in (0, 1/5)$. This means:

$$h(0)2\theta^2 + h(1)(\theta - 2\theta^3) + h(2)\theta^2 + h(3)(1 + 2\theta^3 - 3\theta^2 - \theta) = 0.$$

We rewrite the above relationship as follows:

$$[2h(3) - 2h(1)]\theta^3 + [2h(0) + h(2) - 3h(3)]\theta^2 + [h(1) - h(3)]\theta + h(3) = 0$$

for all $\theta \in (0, 1/5)$. The main theorem of algebra implies then that the coefficients in front of each power of the 3rd order polynomial in θ must be equal to zero. Hence $h(3) = 0 \Longrightarrow h(1) - h(3) = 0 \Longrightarrow h(1) = 0 \Longrightarrow 2h(0) + h(2) = 0$. The latter relationship **does not** necessarily imply that both $h(0) = 0, h(2) = 0$ must hold. Hence the family of distributions is **not** complete.

Question 19: Parts 19a), 19b), 19c) were treated in lecture and are complete. We consider 19d) here. We have to show that $T = X_{(n)}$ is complete. We know that the density of T is

$$f_T(t) = \frac{nt^{n-1}}{\theta^n}, \quad 0 < t < \theta \text{ (and 0 else)}.$$

Let $\mathbb{E}_\theta g(T) = 0$ for all $\theta > 0$. This implies:

$$\int_0^\theta g(t)\frac{nt^{n-1}}{\theta^n}dt = 0 = \frac{1}{\theta^n}\int_0^\theta g(t)nt^{n-1}dt$$

for all $\theta > 0$ must hold. Since $\frac{1}{\theta^n} \neq 0$ we get $\int_0^\theta g(t)nt^{n-1}dt = 0$ for all $\theta > 0$. Differentiating both sides with respect to θ we get

$$ng(\theta)\theta^{n-1} = 0$$

for all $\theta > 0$. This implies $g(\theta) = 0$ for all $\theta > 0$. This also means $P_\theta(g(T) = 0) = 1$. In particular, this result implies that $S = \frac{n+1}{n}X_{(n)}$ is the UMVUE of $\tau(\theta) = \theta$ in this model since $\mathbb{E}_\theta S = \theta$ holds (see previous lectures) **and** S is a function of sufficient **and** complete statistic.

Question 20: The likelihood is

$$L(\mathbf{X},\mathbf{Y}; \mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = \frac{1}{(\sqrt{2\pi})^n \sigma_1^{n_1}\sigma_2^{n_2}} e^{\{-\frac{1}{2}\sum_{i=1}^{n_1}\frac{(x_i-\mu_1)^2}{\sigma_1^2} - \frac{1}{2}\sum_{i=1}^{n_2}\frac{(y_i-\mu_2)^2}{\sigma_2^2}\}}$$

and log-likelihood is

$$\ln L = -n\ln(\sqrt{2\pi}) - n_1\sigma_1 - n_2\sigma_2 - \frac{1}{2}\sum_{i=1}^{n_1}\frac{(x_i-\mu_1)^2}{\sigma_1^2} - \frac{1}{2}\sum_{i=1}^{n_2}\frac{(y_i-\mu_2)^2}{\sigma_2^2}$$

Solving the equation system

$$\frac{\partial}{\partial\mu_1}\ln L = 0 \qquad \text{and} \qquad \frac{\partial}{\partial\mu_2}\ln L = 0$$

delivers

$$\hat{\mu_1} = \bar{X}_{n_1} \qquad \text{and} \qquad \hat{\mu_2} = \bar{Y}_{n_2}$$

for the MLE. Using the transformation invariance property, we get $\hat{\theta} = \bar{X}_{n_1} - \bar{Y}_{n_2}$ for the maximum likelihood estimator of θ. Further:

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\bar{X}_{n_1}) + \mathrm{Var}(\bar{Y}_{n_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n - n_1} = f(n_1).$$

To find the minimum, we set the derivative with respect to n_1 to be equal to zero and solve the resulting equation. This gives: $\frac{\sigma_1}{\sigma_2} = \frac{n1}{n2}$. With other words, the sample sizes must be proportional to the standard deviations. In particular, if n is fixed, we get $n_1 = \frac{\sigma_1}{\sigma_1+\sigma_2}n$.

Question 21: i) The likelihood is

$$L(\mathbf{X};\theta) = \theta^n \prod_{i=1}^n x_i^{-2} I_{[\theta,\infty)}(x_{(1)}).$$

We consider L as a function of theta after the sample has been substituted. When θ moves on the positive half-axis, this function first grows monotonically (when θ moves between 0 and $x_{(1)}$) and then drops to zero onward since the indicator becomes equal to zero. Hence L is a discontinuous function of θ and its maximum is attained at $x_{(1)}$. This means that $\hat{\theta}_{mle} = X_{(1)}$.
ii) Using the factorisation criterion, we see that $X_{(1)}$ is sufficient. It is also minimal sufficient due to dimension considerations. The minimal sufficiency can also be shown by directly examining the ratio $\frac{L(\mathbf{X};\theta)}{L(\mathbf{Y};\theta)}$.

Question 22: a) The likelihood is

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

with log-likelihood

$$\ln L(\mathbf{X};\theta) = n\ln\theta + (\theta-1)\sum_{i=1}^n \ln x_i.$$

The score:

$$\frac{\partial}{\partial\theta}\ln L = \frac{n}{\theta} + \sum_{i=1}^n \ln x_i = 0$$

which gives the root

$$\hat{\theta} = \hat{\theta}_{mle} = \frac{-n}{\sum_{i=1}^n \ln x_i}.$$

Then, using the translation invariance property, we get

$$\hat{\tau(\theta)} = \frac{\hat{\theta}}{\hat{\theta}+1}.$$

b) We have that

$$\sqrt{n}(\hat{\theta} - \theta \to^d N(0, \frac{1}{I_{X_1}(\theta)}).$$

We need to find $I_{X_1}(\theta)$. To this end, we take:

$$\ln f(x;\theta) = ln\theta + (\theta-1)\ln x;$$
$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = \frac{1}{\theta} + \ln x;$$
$$\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta) = -\frac{1}{\theta^2}.$$

This means that $I_{X_1}(\theta) = \frac{1}{\theta^2}$ and

$$\sqrt{n}(\hat{\theta} - \theta) \to^d N(0, \theta^2).$$

Since $\tau(\theta) = \frac{\theta}{\theta+1}$, by applying the delta method we get

$$\sqrt{n}(\hat{\tau} - \tau) \to^d N\Big(0, \frac{\theta^2}{(1+\theta)^4}\Big).$$

c) According to the factorisation criterion, $\prod_{i=1}^n X_i$ is sufficient (also, $\sum_{i=1}^n \ln X_i$ is sufficient). Since the density belongs to an one-parameter exponential we do have completeness, as well. The statistic $T = \sum_{i=1}^n X_i$ is **not** sufficient. Consider for example $0 < t < 1, n = 2, T = X_1 + X_2$. Using the convolution formula (see previous tutorial sheet) we have:

$$f_{X_1+X_2}(t) = \theta^2 \int_0^t x^{\theta-1}(t-x)^{\theta-1}dx.$$

Changing the variables: $x = ty, dx = tdy$, we can continue to obtain:

$$f_{X_1+X_2}(t) = t^{2\theta-1}\theta^2 \int_0^1 y^{\theta-1}(1-y)^{\theta-1}dy = t^{2\theta-1}\theta^2 B(\theta, \theta).$$

Then the conditional density becomes:

$$f_{(X_1,X_2)|T}(x_1, x_2|t) = \frac{\theta^2(x_1x_2)^{\theta-1}}{t^{2\theta-1}\theta^2 B(\theta,\theta)}$$

(if $x_1 + x + 2 = t$, and, of course, zero elsewhere). Hence the conditional density of the sample given the value of the statistic does depend on the parameter.

d)Looking at

$$\frac{\partial}{\partial\theta}\ln L = -n\Big(\frac{-\sum_{i=1}^n \ln x_i}{n} - \frac{1}{\theta}\Big)$$

we see that for $\frac{1}{\theta}$ the CRLB will be attained. This means that $\frac{1}{\theta}$ can be estimated by the UMVUE

$$T = -\frac{\sum_{i=1}^n lnX_i}{n}.$$

The attainable bound is easily seen to be

$$\frac{1}{n\theta^2}.$$

Question 23: a) The density of a single observation is

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where only σ^2 is assumed unknown. Then

$$\ln L(\mathbf{X}; \sigma^2) = -n\ln((\sqrt{2\pi}) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2}$$

Then the equation

$$\frac{\partial}{\partial\sigma^2}\ln L = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^4} = 0$$

has a root

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$$

which is also the MLE.

Further,

$$\ln f(x; \mu, \sigma^2) = -\ln((\sqrt{2\pi}) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2},$$

$$\frac{\partial}{\partial\sigma^2}\ln f = -\frac{1}{2\sigma^2} + \frac{1}{2}\frac{(x-\mu)^2}{\sigma^4},$$
$$\frac{\partial^2}{\partial\sigma^2\partial\sigma^2}\ln f = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}.$$
Taking $-E(\dots)$ in the last equation gives $I_{X_1}(\sigma^2) = \frac{1}{2\sigma^4}$. Hence:
$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \longrightarrow^d N(0, 2\sigma^4).$$

b) We apply the delta method. First, we notice that
$$\hat{\sigma}_{mle} = \hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}$$
is the MLE (due to the transformation invariance property). Now:
$$\sqrt{n}(\hat{\sigma} - \sigma) \longrightarrow^d N(0, ((\frac{\partial}{\partial\sigma^2}h)^2 2\sigma^4)$$
where $h(\sigma^2) = \sqrt{\sigma^2}$. Hence $\frac{\partial}{\partial\sigma^2}h(\sigma^2) = \frac{1}{2\sigma}$ and we get, after substitution:
$$\sqrt{n}(\hat{\sigma} - \sigma) \longrightarrow^d N(0, \sigma^2/2).$$

Question 24 a) i) The MLE of λ is $\bar{X}$ hence of $\tau(\lambda) = \frac{1}{\lambda}$ would be $\hat{\tau} = \frac{1}{\bar{X}}$.
ii) Since $P(\bar{X} = 0) > 0$, we get that even the first moment is infinite (not to mention the second) and there is no finite variance.
iii) The delta method gives us:
$$\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\lambda}) \longrightarrow^d N(0, \frac{1}{\lambda^4}I^{-1}_{X_1}(\lambda)))$$
(since in our case $h(\lambda) = \frac{1}{\lambda}$, $\frac{\partial}{\partial\lambda}h(\lambda) = -\frac{1}{\lambda^2}$.) But, as you can easily see (and we discussed at lectures), for $Po(\lambda)$, we have $I_{X_1}(\lambda) = \frac{1}{\lambda}$, therefore
$$\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\lambda}) \longrightarrow^d N(0, \frac{1}{\lambda^3}).$$
(Comparing the outcomes in (ii) and (iii) we see that although the finite variance does not exist, the asymptotic variance is well defined ($= \frac{1}{\lambda^3}$.))

b) i) $\sqrt{\bar{X}}$ is the MLE and, using the delta method, we get
$$\sqrt{n}(\sqrt{\bar{X}} - \sqrt{\lambda}) \longrightarrow^d N(0, (\frac{1}{2\sqrt{\lambda}})^2\lambda) = N(0, \frac{1}{4}).$$
(Since the asymptotic variance becomes constant ($= \frac{1}{4}$) and does not depend on the parameter, we call the transformation $h(\lambda) = \sqrt{\lambda}$ a **variance stabilising transformation**).
ii) $\sqrt{\bar{X}} \pm \frac{z_{\alpha/2}}{2\sqrt{n}}$ would be the confidence interval for $\sqrt{\lambda}$ and
$$((\sqrt{\bar{X}} - \frac{z_{\alpha/2}}{2\sqrt{n}})^2, (\sqrt{\bar{X}} + \frac{z_{\alpha/2}}{2\sqrt{n}})^2)$$
would be the confidence interval for λ.