Foundations of Statistical Inference, BS2a
Exercises 2: brief solutions

1.
$$L(\theta;\mathbf{x}) = 1 \times I\left[\theta - \frac{1}{2} < X_1, \dots, X_n < \theta + \frac{1}{2}\right]$$

Since
$$\theta \geq \max(X_i) - \frac{1}{2}$$
and
$$\theta \leq \min(X_i) + \frac{1}{2}$$
, we have that any value of $\widehat{\theta}$ in the interval
$$[\max(X_i) - \frac{1}{2}, \min(X_i) + \frac{1}{2}]$$
satisfies the condition that the likelihood is 1.That is the MLE is not unique. A natural method of moments estimator is $\bar{X}$, with $\mathbb{E}[\bar{X}] = \theta$.

2.
$$P(Y \leq y) = \left(\sum_{j=1}^{y} \theta^{-1}\right)^n = (y/\theta)^n$$

The distribution of Y is
$$P(Y = y) = P(Y \leq y) - P(Y \leq y-1) = \left(y^n - (y-1)^n\right)/\theta^n$$

For sufficiency
$$P(X_1 = x_1, \dots, X_n = x_n \mid Y = y) = \frac{1}{y^n - (y-1)^n},$$

where $\max(x_i) = y$. The distribution does not depend on θ. For completeness note that $\mathbb{E}[h(Y)] = 0$ for all θ implies that
$$\sum_{j=1}^{\theta} h(j)\left(j^n - (j-1)^n\right) = 0$$

for all θ. Take $\theta = 1, \theta = 2, \dots$ to show that this implies $h(j) = 0$ for all $j = 1, 2, \dots$. For unbiassedness
$$\mathbb{E}\left[\frac{Y^{n+1} - (Y-1)^{n+1}}{Y^n - (Y-1)^n}\right] = \frac{1}{\theta^n}\sum_{r=1}^{\theta}[r^{n+1} - (r-1)^{n+1}] = \frac{\theta^{n+1}}{\theta^n} = \theta$$

If there is a MVUE then an unbiassed complete sufficient statistic attains it. So is there an MVUE?

Lehman-Scheffe Theorem tells us that if T is a complete sufficient statistic for θ and $\hat{\theta}$ is an unbiased estimator for θ then $\hat{\theta}_T$ is an MVUE. As

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

is unbiased and Y is complete sufficient, we have that $\hat{\theta}_Y = \hat{\theta}$ is MVUE.

3. m fixed trials, R successes, N extra trials until s successes have occurred.

$$P(R=r, N=n) = \binom{m}{r} p^r (1-p)^{m-r} \binom{n-1}{s-1} p^s (1-p)^{n-s}$$

By the factorization theorem (r,n) is sufficient for p. In calculations we use the notation that

$$a_{[b]} = a(a-1)\dots(a-b+1)$$

Then clearly $\mathbb{E}[R/m] = p$ and

$$\begin{aligned}
\mathbb{E}\left[\frac{S-1}{N-1}\right] &= \sum_{n=s}^{\infty} \frac{s-1}{n-1}\binom{n-1}{s-1} p^s (1-p)^{n-s} \\
&= \sum_{n=s}^{\infty} \frac{(n-1)_{[s-1]}}{(s-2)!}(n-1)^{-1} p^s (1-p)^{n-s} \\
&= \sum_{n=s}^{\infty} \frac{(n-2)_{[s-2]}}{(s-2)!} p^s (1-p)^{n-s} \\
&= p\sum_{n=s}^{\infty} \frac{(n-2)_{[s-2]}}{(s-2)!} p^{s-1} (1-p)^{n-1-(s-1)} \\
&= p \times 1
\end{aligned}$$

Thus $\mathbb{E}[U(R,N)] = p - p = 0$, however $U(R,N) \neq 0$, so (R,N) is sufficient for p, but not complete.

4. (i) Let $x = x_1, ..., x_n)$. The likelihood for θ is $L(\theta; x_1, ..., x_n) = \theta^n \left[\prod_{i=1}^n x_i\right]^{\theta-1}$ so $T(x) = \prod_{i=1}^n x_i$ is sufficient for θ, by the factorization theorem, using $K_1(t,\theta) = L$ and $K_2(x) = 1$.

Alternatively, recognize an exponential family $L(\theta; x_1, ..., x_n) = \theta^n \exp\left\{(\theta-1)\sum_{i=1}^n \log x_i\right\}$ so $T(x) = \sum_{i=1}^n \log x_i$

(ii) Rao-Blackwell-ize. Suppose $T(X) = t$. Let

$$\hat{\theta}_T = \mathbb{E}(-\log(X_1)|T(X) = t)$$

This has variance less than or equal the variance of the original estimator. We need to compute $\hat{\theta}_T$ as a function of t. Notice that if $Y_i = -\log(X_i)$ for $i = 1, ..., n$ then $Y_i \sim \text{Exp}(\theta)$. Let $S(y) = \sum_i y_i$. If $T(x) = t$ then $S(y) = s$ with $s = -\log(t)$. If $S = s$ then y_1 is at most s.

$$\begin{aligned}
\hat{\theta}_T &= \mathbb{E}\left(Y_1 | \sum_{i=1}^{n} Y_i = -\log(t)\right) \\
&= \int_0^s y_1 f_{Y_1|S}(y_1|S(y) = s) dy_1.
\end{aligned}$$

Now *either* observe that $Y_i, i = 1, 2..., n$ are exponential intervals summing to s, so Y_1 has the same distribution as $U_{(1)}$, if $U_1, U_2, ..., U_{n-1}$ are iid $U(0, s)$, yielding (see lecture example on moment estimation)

$$f_{Y_1|S}(y_1|S(y) = s) = \frac{(n-1)}{s}\left(\frac{s - y_1}{s}\right)^{n-2},$$

or use the property that $S \sim \Gamma(n, \theta)$ (since it is a sum of iid $\text{Exp}(\theta)$'s), and $S - Y_1 \sim \Gamma(n - 1, \theta)$:

$$\begin{aligned}
f_{Y_1|S}(y_1|S(y) = s) &= \frac{f_{Y_1,S}(y_1, s)}{f_S(s)} \\
&= \frac{f_{Y_1,S-Y_1}(y_1, s - y_1)}{f_S(s)} \\
&= \frac{f_{Y_1}(y_1) f_{S-Y_1}(s - y_1)}{f_S(s)} \\
&= \theta e^{-y_1\theta} \frac{\frac{\theta^{n-1}}{\Gamma(n-1)}(s - y_1)^{n-2} \exp(-(s - y_1)\theta)}{\frac{\theta^n}{\Gamma(n)} s^{n-1} \exp(-s\theta)} \\
&= \frac{(n-1)}{s}\left(\frac{s - y_1}{s}\right)^{n-2}.
\end{aligned}$$

Drop this into the expression for the expectation.

$$\begin{aligned}
\hat{\theta}_T &= \int_0^s y_1 \frac{(n-1)}{s}\left(\frac{s-y_1}{s}\right)^{n-2} dy_1 \\
&= (n-1)s\int_t^1 z(1-z)^{n-2}dz \\
&= (n-1)s\frac{\Gamma(2)\Gamma(n-1)}{\Gamma(n+1)} \\
&= -\log(t)/n
\end{aligned}$$

This is just $\sum_i y_i/n$ as the estimator for the mean of an exponential. You might have guessed this. Its variance is $\mathrm{var}(\hat{\theta}_T) = n^{-1}/\theta^2$ which is smaller than $\mathrm{var}(-\log(X_1)) = 1/\theta^2$.

5. We know that

$$f_X(x;\theta) = f_{X|T}(x \mid t;\theta) f_T(t;\theta)$$

(a)

$$\begin{aligned}
i_X(\theta) &= \mathbb{E}\left[-\frac{\partial^2 \log f_X(x;\theta)}{\partial\theta^2}\right] \\
&= \mathbb{E}\left[-\frac{\partial^2 \log f(x \mid t;\theta)}{\partial\theta^2} - \frac{\partial^2 \log f_T(t;\theta)}{\partial\theta^2}\right] \\
&= i_{X|T}(\theta) + i_T(\theta)
\end{aligned}$$

(b) $i_{X|T}(\theta)$ is the Fisher information for θ in the distribution of $X \mid T$, so is non-negative and thus

$$i_X(\theta) \geq i_T(\theta).$$

(c) $T = \{X_{(1)}, \ldots, X_{(r)}\}$. The distribution of $X_{(1)}, \ldots, X_{(r)}$ is

$$\frac{n!}{(n-r)!} f(x_{(1)}) \cdots f(x_{(r)})[1 - F(x_{(r)})]^{n-r}$$

where $0 < x_{(1)} < \cdots < x_{(r)} < \infty$. That is

$$\frac{n!}{(n-r)!}\theta^r e^{-\theta\sum_1^r x_{(i)}} e^{-\theta(n-r)x_{(r)}}$$

with loglikelihood

$$l(\theta) = \text{const} - \theta\sum_1^r x_{(i)} - \theta(n-r)x_{(r)} + r\log\theta$$

Clearly $I_\theta = r/\theta^2$.

6. Since S is sufficient for θ,

$$\mathbb{E}(T(X)|S(x)) = g(S(x))$$

for some function g, since $\mathbb{E}(T(X)|S(x))$ is a function of S alone, by the Rao-Blackwell theorem. But

$$\mathbb{E}[\mathbb{E}(T(X)|S(X))] = \mathbb{E}(T(X)),$$

by the partition theorem for expectation, so

$$\mathbb{E}[T(X) - g(S(X))] = 0.$$

However, $S(x)$ is minimal sufficient for θ, so $S(x) = f(T(x))$ for some function f, and hence

$$\mathbb{E}(T(X) - g(f(T(X)))) = 0.$$

If T is complete, this implies

$$T(x) - g(f(T(x))) = 0$$

and hence

$$T(x) = g(S(x)).$$

Since T can be written as a function of the minimal sufficient statistic S, it is itself minimal sufficient.