1. Estimation in the exponential distribution.

   (a)
   $$\mathbf{E}(\tilde{\theta}^2) = \mathbf{E}\{X_1^2/2\} = \frac{1}{2\theta}\int x^2 e^{-x/\theta} = \frac{1}{2\theta}\theta^2\Gamma(3) = \theta.$$

   (b) The joint density of $X = (X_1, \ldots, X_n)$ is
   $$f(x;\theta) = \frac{1}{\theta^n}\prod_i e^{-x_i/\theta} = \frac{1}{\theta^n}e^{-\sum_i x_i/\theta},$$
   and Neyman's factorization theorem yields the result.

   (c) Let $T = t(X) = \sum_i X_i$. Using the hint provided we find that
   $$\begin{aligned}\mathbf{E}\{X_1^2/2 \mid T = t\} &= \frac{t^2}{2}\mathbf{E}\{(X_1/T)^2 \mid T = t\} = \frac{t^2}{2}\int_0^1 \frac{b^2(1-b)^{n-2}}{B(1,n-1)}\,db \\ &= \frac{t^2}{2}\frac{B(3,n-1)}{B(1,n-1)},\end{aligned}$$
   where
   $$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$
   Inserting this into the last expression yields the Rao-Blackwellized estimator
   $$\check{\theta}^2 = E\{X_1^2/2 \mid T\} = \frac{T^2}{2}\frac{\Gamma(3)\Gamma(n-1)\Gamma(n)}{\Gamma(n+2)\Gamma(1)\Gamma(n-1)} = \frac{(\sum X_i)^2}{n(n+1)}.$$

   (d) The Fisher information about θ is obtained as follows.
   $$S(\theta) = \frac{\partial}{\partial\theta}\log f(X;\theta) = \frac{n}{\theta} - \frac{\sum_i X_i}{\theta^2}, \quad -\frac{\partial^2}{\partial\theta^2}\log f(X;\theta) = \frac{-n}{\theta^2} + \frac{2\sum_i X_i}{\theta^3}.$$
   Taking expectations yields
   $$i(\theta) = \frac{-n}{\theta^2} + \frac{2n\theta}{\theta^3} = \frac{n}{\theta^2},$$
   so the Cramér–Rao bound for unbiased estimation of θ^2 becomes
   $$(2\theta)^2\frac{\theta^2}{n} = \frac{4\theta^4}{n}.$$
   Let $Y_i = X_i/\theta$; then $U = \sum Y_i$ is Gamma distributed with parameters $(n, 1)$ so
   $$\begin{aligned}\mathbf{V}(U^2) &= \mathbf{E}(U^4) - \{\mathbf{E}(U^2)\}^2 = \frac{\Gamma(n+4)}{\Gamma(n)} - \left(\frac{\Gamma(n+2)}{\Gamma(n)}\right)^2 \\ &= n(n+1)(n+2)(n+3) - n^2(n+1)^2 = 2n(n+1)(n+3).\end{aligned}$$

Thus

$$\mathbf{V}(\check{\theta}^2) = \theta^4 \frac{2n(n+1)(n+3)}{n^2(n+1)^2} = \theta^4 \frac{2(2n+3)}{n(n+1)},$$

and the Bahadur efficiency becomes

$$\text{beff}(\check{\theta}^2) = \frac{2n+2}{2n+3} < 1,$$

so the bound is not attained and the estimator is not Bahadur efficient (although close).

(e) The MLE is determined by $S(\hat{\theta}) = 0$ so we get $\hat{\theta} = \bar{X}$ and $\hat{\theta}^2 = \bar{X}^2$. This has expectation and variance equal to

$$\mathbf{E}(\hat{\theta}^2) = \frac{\theta^2}{n^2}\mathbf{E}(U^2) = \theta^2 \frac{n+1}{n}, \quad \mathbf{V}(\hat{\theta}^2) = \theta^4 \frac{2n(n+1)(2n+3)}{n^4}.$$

Thus we have $\mathbf{V}(\hat{\theta}^2) > \mathbf{V}(\check{\theta}^2)$ and since the last is unbiased this implies in particular

$$\text{mse}(\check{\theta}^2) < \text{mse}(\hat{\theta}^2).$$

Note that $\check{\theta}$ is the bias-corrected version of $\hat{\theta}$.

2. Properties of Fisher information

(a) We have

$$-\frac{\partial^2}{\partial \theta^2} \log f(X;\theta) = -\frac{\partial^2}{\partial \theta^2} \log f(X_1;\theta) - \cdots - \frac{\partial^2}{\partial \theta^2} \log f(X_n;\theta).$$

Taking expectations on both sides yields

$$i(\theta) = i_1(\theta) + \cdots + i_n(\theta).$$

(b) We have that

$$f(x;\theta) = f(x \mid t(x);\theta) f\{t(x);\theta\}$$

Now, let $Y = t(X)$, take logaritms and second derivatives to get

$$-\frac{\partial^2}{\partial \theta^2} \log f(X;\theta) = -\frac{\partial^2}{\partial \theta^2} \log f(X \mid Y;\theta) - \frac{\partial^2}{\partial \theta^2} \log f(Y;\theta)$$

Next, take expectations to find

$$i(\theta) = -\mathbf{E}\left\{\frac{\partial^2}{\partial \theta^2} \log f(X \mid Y;\theta)\right\} + i_Y(\theta).$$

Iterate expectations in the first term to yield

$$i(\theta) = \mathbf{E}\left\{-\mathbf{E}\left(\frac{\partial^2}{\partial \theta^2} \log f(X \mid Y;\theta) \mid Y\right)\right\} = i_{X \mid Y}(\theta).$$

(c) The loss of information $i_{X\,|\,Y}(\theta)$ is always non-negative as it is the expectation of the variance of the conditional score statistic. This yields the inequality

$$i(\theta) \geq i_Y(\theta).$$

Equality can only happen when the conditional score statistic is identically equal to 0 and therefore

$$f(x\,|\,t(x);\theta) = h(x)$$

is independent of θ, i.e. that $Y = t(X)$ is sufficient.

3. Estimation in a uniform position-scale model.

(a) The likelihood function is

$$L(\mu,\delta) = \begin{cases} (2\delta)^{-n} & \text{if } \mu-\delta > x_{(1)} \text{ and } \mu+\delta < x_{(n)} \\ 0 & \text{otherwise} \end{cases}$$

so $(X^{(1)}, X^{(n)})$ is clearly sufficient. But since $(X^{(1)}, X^{(n)})$ also can be inferred from the likelihood function it is also minimal sufficient.

(b) We have

$$\mathbf{E}(X) = \mu, \quad \mathbf{V}(X) = \delta^2/3$$

so

$$\tilde{\mu} = \bar{X}, \quad \tilde{\delta} = \sqrt{\frac{3\,SSD}{n}}.$$

(c) The likelihood function is maximized when δ is minimized, subject to the constraint that the interval $(\mu-\delta, \mu+\delta)$ must contain $X_{(1)}$ and $X_{(n)}$. This minimum is attained when

$$\hat{\mu} = (X_{(1)} + X_{(n)})/2, \quad \hat{\delta} = (X_{(n)} - X_{(1)})/2.$$

(Note there was a small typo here in the original problem sheet).

(d) Let $Y_i = X_i - \mu$. Then Y has a symmetric distribution so

$$\mathbf{E}(Y_{(1)}) = -\mathbf{E}(Y_{(n)})$$

and thus

$$\mathbf{E}(\hat{\mu}) = \mathbf{E}\{(X_{(1)} + X_{(n)})/2\} =) = \mu + \mathbf{E}\{(Y_{(1)} + Y_{(n)})/2\} = \mu.$$

(e) The variance of $\tilde{\mu}$ is $\delta^2/(3n)$.

To find the variance of the MLE we introduce $Z_i = (X_i - \mu)/\delta$ which are clearly uniform on $(-1,1)$, and $(U,V) = (Z_{(1)}, Z_{(n)})$. Realising that

$W = V + 1$ is uniform on $(0, 2)$ and using a result from the previous problem sheet we get

$$\mathbf{V}(U) = \mathbf{V}(V) = \mathbf{V}(W) = \frac{4n}{(n+2)(n+1)^2}.$$

To find the covariance of U and V we first find their joint density. We get for $-1 < u < v < 1$ that

$$P(U > u, V \leq v) = \frac{(v-u)^n}{2^n}$$

and differentiation w.r.t. u and v yields the density

$$f(u,v) = \begin{cases} n(n-1)(v-u)^{n-2}/2^n & \text{if } -1 < u < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Now the variance of $V - U$ is found by direct integration to

$$\mathbf{V}(V-U) = \frac{n(n-1)}{2^n(n+1)(n+2)}$$

so using

$$\mathbf{V}(V-U) = \mathbf{V}(U) + \mathbf{V}(V) - 2\operatorname{Cov}(U,V), \mathbf{V}(U+V) = \mathbf{V}(U) + \mathbf{V}(V) + 2\operatorname{Cov}(U,V)$$

we get

$$\mathbf{V}\hat{\mu} = \delta^2\left\{\frac{1}{4}\mathbf{V}(U+V)\right\} = \delta^2\left\{\frac{n}{(n+2)(n+1)^2} - \frac{n(n-1)}{2^{n+2}(n+1)(n+2)}\right\},$$

which goes to 0 at the rate of n^{-2}.

(f) When $\delta = 1$, we have that the distribution of A does not depend on μ since

$$X_{(n)} - X_{(1)} = (X_{(n)} - \mu) - (X_{(1)} - \mu).$$

(g) Let $B = (U+V)/2$ and $A = (V-U)/2$. This linear transformation has determinant $1/2$ so the joint density of (a, b) is

$$f(a,b) = \begin{cases} 2n(n-1)a^{n-2} & \text{if } -1 < b-a < b+a < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The conditional density of B given $A = a$ is then found by keeping a constant so

$$f(b \mid a) \propto \begin{cases} 1 & \text{if } -1 < b-a < b+a < 1 \\ 0 & \text{otherwise.} \end{cases}$$

i.e. the MLE is uniformly distributed on the appropriate interval:

$$f(\hat{\mu} \mid a) \propto \begin{cases} 1 & \text{if } -1 + a + \mu < b < 1 + \mu - a \\ 0 & \text{otherwise.} \end{cases}$$