MATH96046/MATH97073 Statistical Theory: coursework solutions

1. (a) Neither of the first two distributions are exponential families since both their supports depend on the parameters. For the uniform this is $[0, \theta]$, while for the Pareto this is $[x_0, \infty)$.
In the second case, the support no longer depends on the parameter (which is fixed), hence it can be written in exponential family form

$$\exp\{-(\alpha+1)\log y + \log \alpha + \alpha \log x_0\}$$

with natural statistics $\log y$.

(b) The posterior distribution is proportional to

$$\begin{aligned}\pi(\theta|x) \propto L_n(\theta)\pi(\theta) &\propto \left(\prod_{i=1}^n \frac{1}{\theta} 1\{0 \le x_i \le \theta\}\right) \theta^{-\alpha-1} 1\{\theta \ge k\} \\ &\propto \theta^{-n-\alpha-1} 1\{\min_{1 \le i \le n} x_i \ge 0\} 1\{\max_{0 \le i \le n} x_i \le \theta\} \\ &\propto \theta^{-n-\alpha-1} 1\{\max_{0 \le i \le n} x_i \le \theta\},\end{aligned}$$

i.e. $\theta|X_1, \ldots, X_n \sim Par(\alpha+n, \max_{0 \le i \le n} X_i)$ [note $x_0 = k$ from the prior is included in the maximum].
From this, we see that the prior can be interpreted as contributing α observations whose maximal value is x_0.
The Bayes estimator under squared error loss is the posterior mean. For $Y \sim Par(\alpha, x_0)$,

$$EY = \int_{x_0}^\infty y \alpha x_0^\alpha y^{-\alpha-1} dy = \alpha x_0^\alpha \left[\frac{y^{1-\alpha}}{1-\alpha}\right]_{x_0}^\infty = \frac{\alpha}{\alpha-1} x_0$$

as long as $\alpha > 1$. Substituting in the posterior parameters, the posterior mean equals

$$E^\pi[\theta|x] = \frac{\alpha+n}{\alpha+n-1} \max_{0 \le i \le n} x_i = \frac{\alpha+n}{\alpha+n-1} M_n.$$

The posterior mean is unique and hence it is admissible (Proposition 5.3).

(c) The log-likelihood based on $X_1 > 0$ is

$$\ell_1(\theta; X_1) = -\log\theta - \infty 1\{\theta < X_1\}$$

(it's fine if this is not written as precisely as long as there is some recognition that the log-likelihood takes value $-\infty$ if $X_1 > \theta$. It's also fine if an extra $-\infty 1\{X_1 \ge 0\}$ term is included).
The Fisher information is defined as $I(\theta) = E_\theta[\ell_1'(\theta; X_1)^2] = -E_\theta[\ell_1''(\theta; X_1)]$. While $\theta \mapsto \ell_1(\theta; X_1)$ is not technically differentiable at $\theta = X_1$, X_1 takes this value with probability zero under the $U[0, \theta]$ distribution. Thus it does not effect the expectation defining the Fisher information. The derivative of the log-likelihood can thus be written as

$$\ell_1'(\theta; X_1) = \frac{1}{\theta} 1\{X_1 \le \theta\},$$

since it is considered zero for $\theta < X_1$, when $\ell_1'(\theta; X_1) = -\infty$. The Fisher information is then

$$I_1(\theta) = \frac{1}{\theta^2} E_\theta 1\{X_1 \le \theta\} = \frac{1}{\theta^2}$$

This gives Jeffreys prior

$$\pi(\theta) \propto \sqrt{I_1(\theta)} \propto \theta^{-1}$$

for $\theta > 0$.

Looking at the form of the posterior distribution, this is the same as using a '$Par(0,0)$' prior, and hence the posterior is $Par(n, \max_{1\le i\le n} X_i)$.

(d) Since $E_\theta \max_i X_i = \frac{n}{n+1}\theta$, the estimator $\tilde{\theta}_n$ is unbiased. Hence its mean squared error equals its variance:

$$\mathrm{MSE}_\theta(\tilde{\theta}_n) = \mathrm{Var}_\theta(\tilde{\theta}_n) = \left(\frac{n+1}{n}\right)^2 \frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{1}{n(n+2)}\theta^2.$$

Comparing the MSEs for our two estimators, we want to compare when

$$\mathrm{MSE}_\theta(\tilde{\theta}_n) = \frac{1}{n(n+2)}\theta^2 \le \frac{2\theta^2}{(n+1)(n+2)} = \mathrm{MSE}_\theta(\hat{\theta}_n).$$

Rearranging, this is equivalent to $n \ge 1$ with strict inequality as soon as $n \ge 2$. Thus the MLE is strictly dominated for all $\theta > 0$ and is hence inadmissible.

(e) Using the bias-variance decomposition,

$$\begin{aligned}
\mathrm{MSE}_\theta(T_\delta) &= \mathrm{Bias}_\theta(T_\delta)^2 + \mathrm{Var}_\theta(T_\delta) \\
&= \left(\delta\frac{n}{n+1}\theta - \theta\right)^2 + \delta^2 \mathrm{Var}_\theta\left(\max_{1\le i\le n} X_i\right) \\
&= \delta^2\left(\frac{n}{n+1}\right)^2\theta^2 - 2\delta\frac{n}{n+1}\theta^2 + \theta^2 + \delta^2\frac{n}{(n+1)^2(n+2)}\theta^2 \\
&= \delta^2\frac{n}{(n+1)^2}\left[n + \frac{1}{n+2}\right] - 2\delta\frac{n}{n+1}\theta^2 + \theta^2 \\
&= \left\{\delta^2\frac{n}{n+2} - 2\delta\frac{n}{n+1} + 1\right\}\theta^2
\end{aligned}$$

since $n + \frac{n}{n+2} = \frac{(n+1)^2}{n+2}$. This is a quadratic in δ and is hence minimized at its stationary point. Setting the derivative with respect to δ equal to zero and solving for δ gives $\delta_* = \frac{n+2}{n+1}$. Since the expression for $\mathrm{MSE}_\theta(T_\delta)$ factorizes in θ, it is minimized for all $\theta > 0$ by setting $\delta = \delta_* = \frac{n+2}{n+1}$. Since $\frac{n+1}{n} \ne \delta_*$, the estimator $\tilde{\theta}_n$ in (d) is dominated by T_{δ_*} for all $\theta > 0$ and is hence itself inadmissible.

2. (a) The log-likelihood equals

$$\ell_n(\theta) = \log\left(\prod_{i=1}^n \frac{1}{\theta}e^{-x_i/\theta}\right) = -n\log\theta - \frac{1}{\theta}\sum_{i=1}^n x_i$$

with derivative

$$\ell_n'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n x_i.$$

By rearranging the inequality, one sees that $\ell_n'(\theta) \geq 0$ if and only if $\theta \leq \bar{x}_n$ with equality only at $\theta = \bar{x}_n$. Thus $\theta \mapsto \ell_n(\theta)$ is increasing on $(0, \bar{x}_n)$ and decreasing on $(\bar{x}_n, \infty)$. Thus the MLE is at $\bar{X}_n$ if $\bar{X}_n \leq \delta$, otherwise it lies outside the parameter space. If $\bar{X}_n > \delta$, then the likelihood is increasing on $(0, \delta]$ and so the MLE is δ. In conclusion,

$$\hat{\theta}_n = \min(\bar{X}_n, \delta) = \begin{cases} \bar{X}_n & \text{if } \bar{X}_n \leq \delta, \\ \delta & \text{if } \bar{X}_n > \delta. \end{cases}$$

(b) By the central limit theorem, we have $\sqrt{n}(\bar{X}_n - \theta) \to^d N(0, \theta^2)$ since $E_\theta X_1 = \theta$ and $\mathrm{Var}_\theta(X_1) = \theta^2$ [follows from standard properties of exponential random variables, but can also be worked out explicitly]. If $\theta \in (0, \delta)$,

$$P_\theta(\sqrt{n}(\hat{\theta}_n - \theta) \leq x) = P_\theta(\hat{\theta}_n \leq \theta + x/\sqrt{n}) = P_\theta(\bar{X}_n \leq \theta + x/\sqrt{n})$$

if $\theta + x/\sqrt{n} \leq \delta$. This will be true for any x and large enough n since $\theta < \delta$. Therefore,

$$\lim_{n\to\infty} P_\theta(\sqrt{n}(\hat{\theta}_n - \theta) \leq x) = \lim_{n\to\infty} P_\theta(\sqrt{n}(\bar{X}_n - \theta) \leq x) = P(N(0, \theta^2) \leq x),$$

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

One can alternatively apply the general asymptotic normality result for the MLE

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

and compute the Fisher information $I_{X_1}(\theta) = \theta^{-2}$.
If $\theta = \delta$, by the CLT $\sqrt{n}(\bar{X}_n - \delta) \to^d N(0, \delta^2)$, so

$$\sqrt{n}(\hat{\theta}_n - \delta) = \sqrt{n}(\min(\bar{X}_n, \delta) - \delta) = \sqrt{n}\min(\bar{X}_n - \delta, 0) \to^d \min(N(0, \delta^2), 0)$$

using the continuous mapping theorem.

(c) Let $\varphi = g(\theta) = 1/\theta$. By the invariance of MLE,

$$\hat{\varphi}_n = g(\hat{\theta}_n) = \frac{1}{\hat{\theta}_n} = \frac{1}{\min(\bar{X}_n, \delta)} = \max(1/\bar{X}_n, 1/\delta).$$

If $\theta \in (0, \delta)$ or equivalently $\varphi \in (1/\delta, \infty)$, using the delta method

$$\sqrt{n}(\hat{\varphi}_n - \varphi) \to^d N(0, g'(\theta)^2\theta^2) = N(0, 1/\theta^2) = N(0, \varphi^2)$$

since $g'(\theta) = -1/\theta^2$.
If $\theta = \delta$ or equivalently $\varphi = 1/\delta$, again by the delta method,

$$\begin{aligned} \sqrt{n}(\hat{\varphi}_n - \varphi) \to^d g'(\delta)\min(N(0, \delta^2), 0) &= -\frac{1}{\delta^2}\min(N(0, \delta^2), 0) \\ &= -\min(N(0, 1/\delta^2), 0) \\ &= \max(N(0, 1/\delta^2), 0) \end{aligned}$$

using the symmetry of the mean-zero normal distribution.