Statistical Theory - Problem Sheet 4

Spring 2023. Email corrections to Kolyan Ray: kolyan.ray@ic.ac.uk

1. (a) The posterior distribution is proportional to

$$\binom{n}{X}\theta^X(1-\theta)^{n-X}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \propto \theta^{a+X-1}(1-\theta)^{b+n-X-1},$$

which is the form of a Beta$(a+X, b+n-X)$ distribution. The posterior mean is $\bar{\theta}_n(X) = E[\theta|X] = \frac{X+a}{n+a+b}$.

(b) We have $\theta \sim U[0,1] = \text{Beta}(1,1)$. The posterior mean is thus $\frac{X+1}{n+2}$, which is different from the MLE X/n.

(c) We have

$$\begin{aligned}\sqrt{n}(\bar{\theta}_n(X)-\theta_0) = \sqrt{n}\left(\frac{X+a}{n+a+b}-\theta_0\right) &= \sqrt{n}\frac{a-(a+b)\theta_0}{n+a+b} + \frac{n}{n+a+b}\sqrt{n}\left(\frac{X}{n}-\theta_0\right)\\ &\to^d N(0,\theta_0(1-\theta_0)),\end{aligned}$$

where we used Slutsky's lemma, that $\sqrt{n}(X/n-\theta_0) \to^d N(0,\theta_0(1-\theta_0))$ by the CLT (write $X = \sum_{i=1}^n Y_i$ for $Y_1,\ldots,Y_n \sim^{iid}$ Bernoulli(θ_0)) and the first term converges to zero in probability.

2. The posterior satisfies

$$\begin{aligned}\Pi(\theta = k|X_1,\ldots,X_n) &\propto \prod_{i=1}^n 1\{k < X_i < k+1\}\frac{1}{m}1_{\{1,\ldots,m\}}(k)\\ &\propto 1\{\min X_i > k\}1\{\max X_i < k+1\}1_{\{1,\ldots,m\}}(k)\end{aligned}$$

where the proportionality is as a function of k. The above expression is non-zero if and only if $k \in \{1,\ldots,m\}$ and $\max X_i - 1 < k < \min X_i$. Since $0 \leq \max X_i - \min X_i < 1$, there is only one integer which satisfies this, namely $\lfloor \min X_i \rfloor$. Thus the posterior puts all its mass at this single point, i.e. $\Pi(\theta = \lfloor \min X_i \rfloor | X_1,\ldots,X_n) = 1$. The Bayes estimator under squared error loss is the posterior mean, which is $\hat{\theta}_{Bayes} = \lfloor \min X_i \rfloor$. Since it is unique, it is admissible Q4(a).

3. Write $\theta = (\mu,\sigma^2)$ and $\rho(a,b,\theta) = R(\hat{\theta}_{a,b},\theta)$ for the risk, which equals

$$\begin{aligned}\rho(a,b,\theta) = E_\theta[(aX+b-\mu)^2] &= E_\theta[(a(X-\mu)+b+(a-1)\mu)^2]\\ &= a^2\text{Var}_\theta(X) + (b+(a-1)\mu)^2\\ &= a^2\sigma^2 + (b+(a-1)\mu)^2\end{aligned}$$

[note: this is just the bias-variance decomposition of the mean-squared error].

(i) If $a > 1$,

$$\rho(a,b,\theta) \geq a^2\sigma^2 > \sigma^2 = \rho(1,0,\theta),$$

so $aX+b$ is dominated by X.

(ii) If $a < 0$, then $a - 1 < -1$ and so $(a-1)^2 > 1$,

$$\rho(a,b,\theta) \geq (b + (a-1)\mu)^2 = (a-1)^2 \left(\mu + \frac{b}{a-1}\right)^2$$
$$> \left(\mu + \frac{b}{a-1}\right)^2 = \rho(0, -\frac{b}{a-1}, \theta),$$

so $aX + b$ is dominated by the constant estimator $-\frac{b}{a-1}$.

(iii) If $a = 1$ and $b \neq 1$,

$$\rho(1, b, \theta) = \sigma^2 + b^2 > \sigma^2 = \rho(1, 0, \theta),$$

so $X + b$ is dominated by X.

4. (a) Assume that δ is a unique Bayes rule that is not admissible. Since δ is not admissible, there exists another decision rule δ' with $R(\delta', \theta) \leq R(\delta, \theta)$ for all θ and $R(\delta', \theta) < R(\delta, \theta)$ for some θ. Then

$$\int R(\delta', \theta)\pi(\theta)d\theta \leq \int R(\delta, \theta)\pi(\theta)d\theta$$

and so δ' is a Bayes rule, which contradicts the uniqueness of δ (replace the integrals by sums if the prior is discrete).

(b) Assume that δ has constant risk $R(\delta, \theta) \equiv R(\delta)$, is admissible but not minimax. Since δ is not minimax there exists δ' such that $\sup_\theta R(\delta', \theta) < R(\delta)$. Then for all θ we have $R(\delta', \theta) < R(\delta)$, which contradicts that δ is admissible.

5. By Proposition 5.1 in the notes, we want to find δ minimizing the posterior risk, namely $E[L(\delta(x), \theta)|X = x]$.

(a) Let θ be distributed according to the posterior $\theta|X = x$. A median m is any value such that we have $P(\theta \leq m) \geq 1/2$ and $P(\theta \geq m) \geq 1/2$. This is equivalent to $P(\theta < m) \leq 1/2$ and $P(\theta > m) \leq 1/2$. Let m be a median and $c > m$, then

$$|\theta - c| - |\theta - m| = \begin{cases} c - m & \text{if } \theta \leq m, \\ m + c - 2\theta & \text{if } m \leq \theta \leq c, \\ m - c & \text{if } \theta \geq c, \end{cases}$$

[draw a picture]. Taking expectations and splitting the integral into the above 3 regions,

$$\begin{aligned} E|\theta - c| - E[\theta - m| &= E[(|\theta - c| - |\theta - m|)(1_{(-\infty,m]}(\theta) + 1_{(m,c)}(\theta) + 1_{[c,\infty)}(\theta)] \\ &= (c-m)P(\theta \leq m) + (m+c)P(m < \theta < c) \\ &\quad - \int_{m<\theta<c} 2\theta \, dP(\theta) + (m-c)P(\theta \geq c) \\ &= (c-m)[P(\theta \leq m) - P(\theta > m)] + 2\int_{m<\theta<c} (c-\theta)dP(\theta) \\ &\geq 0, \end{aligned}$$

i.e. $E|\theta - c| \geq E|\theta - m|$. If the last inequality is an equality, then $P(\theta \leq m) = P(\theta > m) = 1/2$ and $P(m < \theta < c) = 0$, so that

$$P(\theta \geq c) = P(\theta > m) - P(c > \theta > m) = 1/2 \quad \text{and} \quad P(\theta \leq c) \geq P(\theta \leq m) = 1/2$$

and thus c is also a median. For $c < m$, we similarly write

$$E|\theta - c| - E|\theta - m| = (m-c)(P(\theta \geq m) - P(\theta < m)) + 2\int_{c<\theta<m} (\theta - c)\dot{\mathrm{P}}(\theta) \geq 0$$

and likewise obtain in the case of equality

$$P(\theta \leq c) = P(\theta < m) - P(c < \theta < m) = 1/2 \quad \text{and} \quad P(\theta \geq c) \geq P(\theta \geq m) = 1/2,$$

so c is also a median.

(b) Let $\delta = E^\pi[w(\theta)\theta|X]/E^\pi[w(\theta)|X]$. For λ another decision rule, the difference in posterior risks is

$$\begin{aligned}
&E^\pi[w(\theta)(\lambda-\theta)^2|X] - E^\pi[w(\theta)(\delta-\theta)^2|X] \\
&= E^\pi[w(\theta)(2\theta(\delta-\lambda) + \lambda^2 - \delta^2)|X] \\
&= 2(\delta-\lambda)E^\pi[w(\theta)\theta|X] + (\lambda^2-\delta^2)E^\pi[w(\theta)|X] \\
&= 2(\delta-\lambda)\delta E^\pi[w(\theta)|X] - (\delta^2-\lambda^2)E^\pi[w(\theta)|X] \\
&= (\delta-\lambda)E^\pi[w(\theta)|X](2\delta - (\delta+\lambda)) \\
&= (\delta-\lambda)^2 E^\pi[w(\theta)|X] \geq 0.
\end{aligned}$$

The uniqueness holds since $E^\pi[w(\theta)|X] > 0$.

6. (a) The risk equals

$$\begin{aligned}
R(\bar\theta_n,\theta) = E_\theta\left[\left(\frac{X+a}{n+a+b} - \theta\right)^2\right] &= E_\theta\left[\left(\frac{(X-n\theta)+a(1-\theta)-b\theta}{n+a+b}\right)^2\right] \\
&= \frac{n\theta(1-\theta) + (a(1-\theta)-b\theta)^2}{(n+a+b)^2} \\
&= \frac{((a+b)^2-n)\theta^2 + (n-2a(a+b))\theta + a^2}{(n+a+b)^2}.
\end{aligned}$$

This is constant in $\theta \in [0,1]$ if and only if $(a+b)^2 = n$ and $2a(a+b) = n$ with $a, b > 0$. These equations are solved by $a = b = \sqrt{n}/2$. Thus $\tilde\theta_n = (X + \sqrt{n}/2)/(n+\sqrt{n})$ is a Bayes rule with constant risk, and is hence minimax. Since the posterior mean is the unique Bayes solution for this problem, we conclude that it is the unique minimax estimator.

(b) The constant risk in (a) is $\frac{1}{4(\sqrt{n}+1)^2}$. The MLE $\hat\theta_n = X/n$ is unbiased and so has risk $R(\hat\theta_n,\theta) = \mathrm{MSE}_\theta(\hat\theta_n) = \mathrm{Var}_\theta(\hat\theta_n) = \frac{\theta(1-\theta)}{n}$, which is maximized at $\sup_\theta R(\hat\theta,\theta) = \frac{1}{4n} > \frac{1}{4(\sqrt{n}+1)^2}$. Hence the MLE is not minimax.

(c) We have

$$\lim_{n\to\infty} \frac{\sup_\theta R(\hat\theta_n,\theta)}{\sup_\theta R(\tilde\theta_n,\theta)} = \lim_{n\to\infty} \frac{4(1+\sqrt{n})^2}{4n} = 1$$

and

$$\lim_{n\to\infty} \frac{R(\hat\theta_n,\theta)}{R(\tilde\theta_n,\theta)} = \lim_{n\to\infty} \frac{\theta(1-\theta)4(1+\sqrt{n})^2}{n} = 4\theta(1-\theta),$$

which is less than one for all $\theta \neq 1/2$.

7. Since (b), (c), (d) and (e) are exponential families, they have complete sufficient statistics by Proposition 5.4 (completness) and an Example on p10 (sufficiency) in the notes. Writing these pdfs in exponential family form,

$$\begin{aligned}
f_\theta^b(x) &= \exp\{-\theta\log(1+x) + \log\theta - \log(1+x)\}, \\
f_\theta^c(x) &= \exp\{x\log\theta + \log\log\theta - \log(\theta-1)\}, \\
f_\theta^d(x) &= \exp\{e^\theta e^{-x} + \theta - x\}, \\
f_\theta^e(x) &= \exp\left\{x\log\left(\tfrac{\theta}{1-\theta}\right) + 2\log(1-\theta) + \log\binom{2}{x}\right\},
\end{aligned}$$

from which we read off (b) $\sum_i \log(1+X_i)$, (c) $\sum_i X_i$, (d) $\sum_i e^{-X_i}$ and (e) $\sum_i X_i$.

For (a), $\prod_i f_\theta(x_i) = (2/\theta^2)^n \prod_i x_i 1\{\max x_i \le \theta\}$, so that $Y = \max X_i$ is sufficient by the factorization criterion. By the usual computation for order statistics, we have $P(Y \le y) = (y/\theta)^{2n}$ and pdf $f_Y(y) = \frac{2n}{\theta^{2n}} y^{2n-1} 1_{(0,\theta)}(y)$. We now check completeness. For a function g, suppose

$$E_\theta g(Y) = \frac{2n}{\theta^{2n}} \int_0^\theta g(y) y^{2n-1} dy = 0, \qquad \forall \theta > 0.$$

Differentiating with respect to θ,

$$-(2n)^2 \theta^{-2n-1} \int_0^\theta g(y) y^{2n-1} dy + \frac{2n}{\theta^{2n}} g(\theta)\theta^{2n-1} = \frac{2n}{\theta} g(\theta) = 0$$

for all $\theta > 0$. Hence $g \equiv 0$ and Y is complete.

8. The likelihood equals

$$\prod_{i=1}^n e^{-(x_i-\theta)} 1\{x_i > \theta\} = e^{n\theta - \sum x_i} 1\{\min x_i > \theta\},$$

so by the factorization criterion, $T = \min x_i$ is sufficient for θ. We saw in Q5 on PS3 that it has pdf $f_T(t) = ne^{-n(t-\theta)} 1_{(\theta,\infty)}(t)$ and suppose for some function g that

$$E_\theta g(T) = n \int_\theta^\infty g(t) e^{-n(t-\theta)} dt = 0 \qquad \forall \theta > 0.$$

Differentiating with respect to θ (similar to Q8(a)) implies $g(\theta) = 0$ for all $\theta > 0$. Hence $T = \min X_i$ is complete.

Since $\{f_\theta : \theta > 0\}$ is a location family (i.e. $f_\theta(x) = f(x-\theta)$ for some f), we can write $X_i =^d \theta + Z_i$ where $Z_1, \dots, Z_n \sim^{iid} f$ (in this case Exp(1) random variables). Then

$$S^2 =^d \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z}_n)^2,$$

whose distribution does not depend on θ. Hence S^2 is ancillary for θ and so is independent of $\min X_i$ by Basu's theorem (Theorem 5.2).

We saw in Q5 on PS3 that $E_\theta \min X_i = \theta + 1/n$, so $\min X_i - 1/n$ is an unbiased estimator of θ that is a function of a complete sufficient statistic, and hence is UMVUE by the Lehmann-Scheffe theorem.

9. To find the UMVUE, we look for a complete sufficient statistic, which must be minimal sufficient. Consider the ratio

$$\frac{\prod_{i=1}^n \frac{1}{2\theta} 1\{|x_i| < \theta\}}{\prod_{i=1}^n \frac{1}{2\theta} 1\{|x_i'| < \theta\}} = \frac{1\{\max |x_i| < \theta\}}{1\{\max |x_i'| < \theta\}}.$$

Hence $T = \max |X_i|$ is minimal sufficient and so a candidate. Since $|X_i| \sim^{iid} U[0,\theta)$, we have seen that the pdf for T is $f_T(t) = nt^{n-1}/\theta^n 1_{[0,\theta)}(t)$ and that T is complete in an example in the notes (p. 43). To compute the UMVUE, we want an unbiased estimator that is a function of T. Note that

$$E_\theta T = n \int_0^\theta \frac{t^n}{\theta^n} dt = \frac{n}{n+1}\theta.$$

Hence $\frac{n+1}{n} T$ is an unbiased estimator of θ that is a function of T, and hence is the UMVUE by the Lehmann-Scheffe theorem.

10. Since $E\sum_i a_i X_i = \sum_i a_i E X_i = \mu \sum_i a_i$, the estimator is unbiased if $\sum_i a_i = 1$. The variance of such an estimator equals $\mathrm{Var}(\sum a_i X_i) = \sum a_i^2 \mathrm{Var}(X_i) = \sigma^2 \sum a_i^2$. Thus we want to minimize $\sum a_i^2$ subject to $\sum a_i = 1$. Using the constraint,

$$\begin{aligned}\sum a_i^2 = \sum (a_i - 1/n + 1/n)^2 &= \sum (a_i - 1/n)^2 + \frac{2}{n}\sum (a_i - 1/n) + \sum 1/n^2 \\ &= \sum (a_i - 1/n)^2 + 1/n.\end{aligned}$$

Hence $\sum a_i^2$ is minimized by choosing $a_i = 1/n$. Thus $\bar{X}_n = \frac{1}{n}\sum X_i$ has the minimum variance σ^2/n among all linear unbiased estimators.

[One can do the constrained minimization in other ways, e.g. using Lagrange multipliers.]

11. Using the exponential family form, $\sum X_i$ is a complete and sufficient statistic for θ (Proposition 5.4 in the notes). Since

$$E_\theta\left(\bar{X}_n^2 - 1/n\right) = \mathrm{Var}_\theta(\bar{X}_n) + (E_\theta \bar{X}_n)^2 - 1/n = 1/n + \theta^2 - 1/n = \theta^2$$

is unbiased for θ^2 and is a function of $\sum X_i$, it is the UMVUE for θ^2 by the Lehmann-Scheffe theorem.

Since $\bar{X}_n \sim N(\theta, 1/n)$, its variance equals

$$\begin{aligned}\mathrm{Var}_\theta(\bar{X}_n^2 - 1/n) = \mathrm{Var}_\theta(\bar{X}_n^2) &= E_\theta(\bar{X}_n^4) - (E_\theta \bar{X}_n^2)^2 \\ &= \theta^4 + 6\theta^2/n + 3/n^2 - (\theta^2 + 1/n)^2 = 4\theta^2/n + 2/n^2.\end{aligned}$$

From Problem Sheet 1 Q1, we have Fisher information $I_{X_1}(\theta) = 1$ in this model. Hence the Cramer-Rao bound for $g(\theta) = \theta^2$ is $\frac{g'(\theta)^2}{n I_{X_1}(\theta)} = \frac{4\theta^2}{n}$. Thus the UMVUE $\bar{X}_n^2 - 1/n$ of θ^2 does not attain the CR bound and so no unbiased estimator does.

Note the ratio of the variance to the lower bound $\to 1$ as $n \to \infty$, so it asymptotically attains the CR bound. This is consistent since the MLE for θ^2 is $\bar{X}_n^2$ by the invariance of MLE, which has the same variance as $\bar{X}_n^2 - 1/n$, while the asymptotic normality result says the MLE will have limiting variance equal to the CR bound.