Solutions 3

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

1. Let $\epsilon > 0$. By continuity of S_n and by $\hat{\theta}_n$ being the only zero of S_n we have

$$P(S_n(\theta_0 \pm \epsilon) < 0 < S_n(\theta_0 \mp \epsilon)) \leq P(\theta_0 - \epsilon < \hat{\theta}_n < \theta_0 + \epsilon).$$

The left probability converges to one (see below) since $S(\theta_0 \pm \epsilon) < 0 < S(\theta_0 \mp \epsilon)$ and $S_n(\theta_0 \pm \epsilon) \to^P S(\theta_0 \pm \epsilon)$ as $n \to \infty$. Consequently, for all $\epsilon > 0$ we have $P(|\hat{\theta}_n - \theta_0| < \epsilon) \to 1$ as $n \to \infty$. We conclude that $\hat{\theta}_n \to^P \theta_0$ as $n \to \infty$.
To rigorously show the left probability converges to one, note that by the convergence in probability, $P(|S_n(\theta_0 \pm \epsilon) - S(\theta_0 \pm \epsilon)| \leq \delta) \to 1$ for any $\delta > 0$ as $n \to \infty$. On the event $\{|S_n(\theta_0 \pm \epsilon) - S(\theta_0 \pm \epsilon)| \leq \delta\}$, the event in the left-hand side of the last display equals

$$\begin{aligned}&\{S_n(\theta_0 \pm \epsilon) - S(\theta_0 \pm \epsilon) + S(\theta_0 \pm \epsilon) < 0 < S_n(\theta_0 \mp \epsilon) - S(\theta_0 \mp \epsilon) + S(\theta_0 \mp \epsilon)\} \\ &\quad \supseteq \{S(\theta_0 \pm \epsilon) + \delta < 0 < S(\theta_0 \mp \epsilon) - \delta\}.\end{aligned}$$

Taking $\delta > 0$ small enough (e.g. $\delta = \frac{1}{2}\min\{S(\theta_0 \pm \epsilon), S(\theta_0 \mp \epsilon)\}$), the last event holds true. Thus for such $\delta > 0$,

$$P(S_n(\theta_0 \pm \epsilon) < 0 < S_n(\theta_0 \mp \epsilon)) \geq P(|S_n(\theta_0 \pm \epsilon) - S(\theta_0 \pm \epsilon)| \leq \delta) \to 1$$

as required.

2. The log-likelihood is given by

$$\ell_n(\theta) = n \log \theta + (\theta - 1)\sum_{j=1}^n \log x_j - \sum_{j=1}^n \exp(\theta \log x_j)$$

and its derivatives are

$$\begin{aligned}\ell_n'(\theta) &= \frac{n}{\theta} + \sum_{j=1}^n \log x_j - \sum_{j=1}^n (\log x_j)\exp(\theta \log x_j) \\ &= \frac{n}{\theta} + \sum_{j=1}^n (\log x_j)(1 - \exp(\theta \log x_j)), \\ \ell_n''(\theta) &= -\frac{n}{\theta^2} - \sum_{j=1}^n (\log x_j)^2 \exp(\theta \log x_j) < 0.\end{aligned}$$

So ℓ_n' is decreasing and $\lim_{\theta \to 0} \ell_n'(\theta) = \infty$. With probability one, there is some $X_j \neq 1$, and by separately looking at the cases $X_j < 1$ and $X_j > 1$, we see that $\ell_n'(\theta)$ is negative for large θ. We conclude that with probability one, there is a unique solution to $\ell_n'(\theta) = 0$, and thus a unique MLE exists.

For the consistency at θ_0, define

$$\begin{aligned}S_n(\theta) &:= \frac{1}{n}\ell_n'(\theta) = \frac{1}{\theta} + \frac{1}{n}\sum_{j=1}^n \log x_j - \frac{1}{n}\sum_{j=1}^n (\log x_j)\exp(\theta \log x_j), \\ S(\theta) &:= E_{\theta_0}[\ell_1'(\theta)] = \frac{1}{\theta} + E_{\theta_0}\left[\log X - (\log X)\exp(\theta \log X)\right].\end{aligned}$$

By interchanging differentiation $d/\mathrm{d}\theta$ and dx-integration,

$$\begin{aligned} S(\theta_0) = E_{\theta_0}[\ell_1'(\theta_0)] &= \int_0^\infty \left.\frac{d \log f_\theta(x)}{d\theta}\right|_{\theta=\theta_0} f_{\theta_0}(x) dx \\ &= \int_0^\infty \left.\frac{d f_\theta(x)}{d\theta}\right|_{\theta=\theta_0} dx = \frac{d}{d\theta} \int_0^\infty f_\theta(x) dx \Big|_{\theta=\theta_0} = 0. \end{aligned}$$

Further we have $S'(\theta) < 0$ for all $\theta \in \Theta$. By the LLN. $S_n(\theta) \to^p S(\theta)$ for all $\theta \in \Theta$ under P_{θ_0}. We observe that the statement of Question 1 also holds if for every $\epsilon > 0$ small enough, $S(\theta_0 - \epsilon) > 0 > S(\theta_0 + \epsilon)$ and if with probability one S_n has exactly one zero $\hat\theta_n$ for every $n \in \mathbb{N}$. We conclude that the MLE is consistent by this statement.

3. We have $E_\theta[X^4] = 3\theta^2 =: g(\theta)$. The MLE of θ is $\hat\theta_{ML} = \overline{X^2} = \frac{1}{n}\sum_{j=1}^n X_j^2$ (Q1 on Problem Sheet 2), and so an MLE for ϕ is $\hat\phi = g(\hat\theta_{ML}) = 3\left(\overline{X^2}\right)^2$. By the CLT, it holds that $\sqrt{n}(\overline{X^2} - \theta) \to^d N(0, 2\theta^2)$. We calculate $g'(\theta) = 6\theta$ and derive by the delta method $\sqrt{n}(\hat\phi - \phi) \to^d N(0, 72\theta^4)$.

4. We recall $\hat\theta = \max(\overline{X}, 0)$, where $\overline{X} = \frac{1}{n}\sum_{j=1}^n X_j$ (Q7 Problem Sheet 2). We have

$$P_\theta(\sqrt{n}(\hat\theta - \theta) \le x) = P_\theta(\hat\theta \le \theta + x/\sqrt{n}) = P_\theta(\overline{X} \le \theta + x/\sqrt{n})$$

if $\theta + x/\sqrt{n} \ge 0$. For $\theta > 0$ this is fulfilled for each x if n is large enough. Consequently,

$$\lim_{n\to\infty} P_\theta(\sqrt{n}(\hat\theta - \theta) \le x) = \lim_{n\to\infty} P_\theta(\sqrt{n}(\overline{X} - \theta) \le x) = \phi(x)$$

since $\sqrt{n}(\overline{X} - \theta) \to^d N(0,1)$ (actually the $\to^d$ is $=^d$). Thus for $\theta > 0$, $\sqrt{n}(\hat\theta - \theta) \to^d N(0,1)$. If $\theta = 0$, then

$$\begin{aligned} \sqrt{n}(\hat\theta - \theta) &= \sqrt{n}\max(\overline{X}, 0) \\ &= \sqrt{n}\max(\overline{X} - \theta, 0) \\ &= \max(\sqrt{n}(\overline{X} - \theta), 0) \\ &\to^d \max(Z, 0), \end{aligned}$$

for $Z \sim N(0,1)$ by the continuous mapping theorem. The asymptotic normality result holds for points in the interior of Θ. We see that for points on the boundary, there might not be asymptotic normality of the MLE.

5. (i) The likelihood equals

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

which is maximized by taking $\hat\theta = \min_i x_i$. The density function follows directly from that for order statistics:

$$f_{X_{(1)}}(t) = n[1 - F(t)]^{n-1} f(t) = n[e^{-(t-\theta)}]^{n-1} e^{-(t-\theta)} 1_{[\theta,\infty)}(t) = n e^{-n(t-\theta)} 1_{[\theta,\infty)}(t),$$

which implies $E_\theta \hat\theta = \int_\theta^\infty t n e^{-n(t-\theta)} dt = \theta + 1/n$. For the consistency, for any $\epsilon > 0$,

$$P_\theta(|\hat\theta - \theta| > \epsilon) = P_\theta(\hat\theta - \theta > \epsilon) = \int_{\theta+\varepsilon}^\infty n e^{-n(t-\theta)} dt = e^{-n\epsilon} \to 0.$$

(ii) By integration by parts, $E_\theta \hat{\theta}^2 = \int_\theta^\infty t^2 n e^{-n(t-\theta)} dt = \theta^2 + 2\theta/n + 2/n^2$. Thus

$$\mathrm{Var}_\theta(\tilde{\theta}) = \mathrm{Var}_\theta(\hat{\theta}) = E_\theta \hat{\theta}^2 - (E_\theta \hat{\theta})^2 = 1/n^2.$$

The Cramer-Rao lower bound does not apply since the range of the pdf depends on the parameter θ.

(iii) We note that $X_i =^d \theta + Z_i$, for $Z_1, \ldots, Z_n \sim^{iid} \mathrm{Exp}(1)$. Thus $n(\hat{\theta} - \theta) =^d n \min_i Z_i$, and one can work out directly that $P(n \min_i Z_i \geq x) = P(Z_1 \geq x/n)^n = e^{-x}$, i.e. $n(\hat{\theta} - \theta) \to^d \mathrm{Exp}(1)$. Finally, $n(\tilde{\theta} - \theta) = n(\hat{\theta} - \theta) - 1 \to^d -1 + \mathrm{Exp}(1)$ by Slutsky's lemma.

6. The posterior distribution is proportional to

$$\pi(\theta) \prod_{i=1}^n f_\theta(X_i) \propto \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)} \prod_{i=1}^n e^{-\theta} \frac{\theta^{X_i}}{X_i!} \propto \theta^{\alpha + \sum X_i - 1} e^{-(\beta+n)\theta},$$

which is the form of a $\Gamma(\alpha + \sum_i X_i, \beta + n)$ distribution.

7. (i) We have

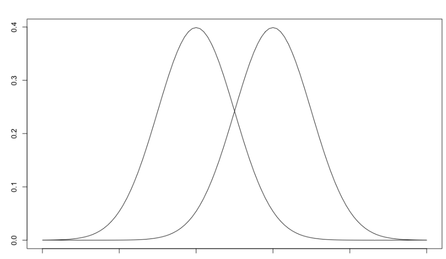
$$f_\theta(x) \propto \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), \qquad \pi(\theta) \propto \exp\left(-\frac{(\theta-\mu)^2}{2v^2}\right).$$

By the formula for the density of the posterior distribution, we have

$$\begin{aligned}
\pi(\theta|X_1, \ldots, X_n) &\propto \pi(\theta) \prod_{i=1}^n f(\theta, X_i) \\
&\propto \exp\left(-\frac{(\theta-\mu)^2}{2v^2} - \sum_{i=1}^n \frac{(X_i - \theta)^2}{2\sigma^2}\right) \\
&\propto \exp\left(-\frac{\theta^2}{2v^2} - \frac{n\theta^2}{2\sigma^2} + \theta\left(\frac{2\mu}{2v^2} + \sum_{i=1}^n \frac{2X_i}{2\sigma^2}\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{1}{v^2} + \frac{n}{\sigma^2}\right)\left(\theta - \frac{\frac{\mu}{v^2} + \frac{n}{\sigma^2}\bar{X}}{\frac{1}{v^2} + \frac{n}{\sigma^2}}\right)^2\right),
\end{aligned}$$

which shows that $\theta|X_1, \ldots, X_n$ has the claimed normal distribution.

(ii) Without loss of generality let $\theta_1 < \theta_2$. Since the normal distributions have the same variance, the pdf for θ_2 is larger than that for θ_1 if and only if $x \geq \delta = (\theta_1 + \theta_2)/2$, see the picture below.



Using also the symmetry about δ,

$$\begin{aligned}
\|N(\theta_1,\tau^2)-N(\theta_2,\tau^2)\|_{TV} &= \frac{1}{\sqrt{2\pi\tau^2}}\int_{\mathbb{R}}\left|e^{-(x-\theta_1)^2/(2\tau^2)}-e^{-(x-\theta_2)^2/(2\tau^2)}\right|dx\\
&= \frac{2}{\sqrt{2\pi\tau^2}}\int_{-\infty}^{\delta} e^{-(x-\theta_1)^2/(2\tau^2)}-e^{-(x-\theta_2)^2/(2\tau^2)}dx\\
&= 2[P(N(\theta_1,\tau^2)\le\delta)-P(N(\theta_2,\tau^2)\le\delta)]\\
&= 2[P(N(0,1)\le(\delta-\theta_1)/\tau)-P(N(0,1)\le(\delta-\theta_2)/\tau)]\\
&= 2\left[\Phi\left(\frac{\theta_2-\theta_1}{2\tau}\right)-\Phi\left(\frac{\theta_1-\theta_2}{2\tau}\right)\right].
\end{aligned}$$

In conclusion,

$$\|N(\theta_1,\tau^2)-N(\theta_2,\tau^2)\|_{TV} = 2\left[2\Phi\left(\frac{|\theta_2-\theta_1|}{2\tau}\right)-1\right].$$

(iii) Since $\theta|X_1,\dots,X_n \sim N(\frac{n}{n+\frac{1}{v^2}}\bar{X}_n, \frac{1}{n+\frac{1}{v^2}})$ and $\hat{\theta}_{ML}=\bar{X}_n$, we have

$$\sqrt{n}(\theta-\hat{\theta}_{ML})|X_1,\dots,X_n \sim N\left(-\frac{\sqrt{n}}{nv^2+1}\bar{X}_n, \frac{n}{n+1/v^2}\right).$$

Using (ii), the total variation between the two distributions equals

$$4\Phi\left(\frac{1}{2v\sqrt{nv^2+1}}|\bar{X}_n|\right)-2$$

Under the frequentist assumption, $\bar{X}_n \to^p \theta_0$ by the weak law of large numbers, and so $|\bar{X}_n| \to^p |\theta_0|$ by the continuous mapping theorem. Applying Slutsky's theorem then gives $\frac{1}{2v\sqrt{nv^2+1}}|\bar{X}_n| \to^d 0$. Since $\Phi(x)$ is continuous, by the continuous mapping theorem, the total variation distance we computed above converges in distribution to $4\Phi(0)-2=0$. Looking at the (non-random) variance of the posterior distribution, $\frac{n}{n+1/v^2}\to 1$ as $n\to\infty$. Hence the posterior distribution will asymptotically look like a $N(0,1)$ distribution, which equals the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{ML}-\theta_0)\to^d N(0, I_{X_1}(\theta_0)^{-1}) = N(0,1)$ in this model (Q10(b) on Problem Sheet 1).

[We have proved that for large n, the posterior asymptotically looks like $\theta|X_1,\dots,X_n \approx^d N(\hat{\theta}_{ML}, n^{-1}I_{X_1}(\theta_0)^{-1})$, i.e. the posterior is normal, centered at the MLE and has variance equal to the inverse Fisher information. This result is called the Bernstein-von Mises theorem and actually holds more generally, but its proof is beyond the scope of this module. One can actually show the posterior converges in total variation also by showing $\|N(0,\frac{n}{n+1/v^2})-N(0,1)\|_{TV}\to 0$ and using the triangle inequality for the total variation norm - you might want to have a go at this.]

8. (i) The Fisher information is $I(\theta)=1/\theta$ (Q1 Problem Sheet 2), and hence the Jeffreys prior is $\pi(\theta)\propto\sqrt{1/\theta}$ for $\theta>0$. The posterior is proportional to

$$\pi(\theta)f_\theta(X)\propto\theta^{-1/2}\frac{\theta^X}{X!}e^{-\theta}\propto\theta^{X-1/2}e^{-\theta}.$$

This is the form of a Gamma$(X+1/2,1)$ distribution, which is a proper probability distribution since it can be normalized as $\pi(\theta|X)=\frac{1}{\Gamma(X+1/2)}\theta^{X-1/2}e^{-\theta}$.

(ii) We have log-likelihood

$$\ell(\theta) = \log(e^{-\theta y}(1-e^{-\theta})^{1-y}) = -\theta y + (1-y)\log(1-e^{-\theta}),$$
$$\ell'(\theta) = -y + (1-y)\frac{e^{-\theta}}{1-e^{-\theta}}$$
$$\ell''(\theta) = -(1-y)\frac{e^{-\theta}}{(1-e^{-\theta})^2},$$

so that $I(\theta) = -E_\theta[l''(\theta)] = \frac{e^{-\theta}}{1-e^{-\theta}}$. Thus the Jeffreys priors satisfies $\pi(\theta) \propto (\frac{e^{-\theta}}{1-e^{-\theta}})^{1/2}$. The posterior is proportional to

$$\pi(\theta)\tilde{f}_\theta(Y) \propto e^{-\theta y}(1-e^{-\theta})^{1-y}\left(\frac{e^{-\theta}}{1-e^{-\theta}}\right)^{1/2}$$

To show it yields a proper probability distribution, we only need to show that the integral is finite, in which case we can normalize the density. Since $|e^{-\theta}| \le 1$, $|1-e^{-\theta}| \le 1$ for $\theta \ge 0$ and using the hint,

$$\int_0^\infty e^{-\theta y}(1-e^{-\theta})^{1-y}\left(\frac{e^{-\theta}}{1-e^{-\theta}}\right)^{1/2} d\theta \le \int_0^\infty \left(\frac{e^{-\theta}}{1-e^{-\theta}}\right)^{1/2} d\theta = \pi < \infty.$$

(iii) If $X = 0$ and $Y = 1$, we have likelihoods $f_\theta(0) = e^{-\theta}$ and $\tilde{f}_\theta(1) = e^{-\theta}$, which are proportional. The likelihood principle states that we should make the same inference based on these two likelihoods. However, the posteriors based on the Jeffreys priors are proportional to (i) $\theta^{-1/2}e^{-\theta}$ and (ii) $(e^{-\theta})^{3/2}/\sqrt{1-e^{-\theta}}$, respectively. Since these are not equal, the posteriors will be different and we make different inference based on them, i.e. Jeffreys priors do not satisfy the likelihood principle.

9. We have

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

The posterior distribution is proportional to

$$\pi(\mu,\sigma^2|X_1,\ldots,X_n) \propto (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\right)\sigma^{-2}$$
$$\propto \sigma^{-(n+2)}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\right).$$

Conditioning on σ yields

$$\pi(\mu|\sigma^2, X_1,\ldots,X_n) \propto \exp\left(-\frac{n}{2\sigma^2}\left(\mu-\bar{X}\right)^2\right)$$

so that the distribution of $\mu|\sigma^2, X_1,\ldots,X_n$ is $N(\bar{X}, \sigma^2/n)$.