MA40092 PROBLEM SHEET 6 - SOLUTIONS

Bookwork question: (§3.1)
To prove that maximum likelihood estimates are functionally invariant, that is if $\hat{\theta}$ is the MLE of θ, then the MLE of $g(\theta)$ is $g(\hat{\theta})$, let $\phi = g(\theta)$. Denote the likelihood function for θ by $L(\theta)$ and the likelihood function for ϕ by $\tilde{L}(\phi)$. Although the question only asks you to consider invertible $g()$, we will also consider the case where g is not invertible separately:

g **is invertible** In this case the likelihood $\tilde{L}(\phi)$ is easy to define since we know exactly which value of θ corresponds to any ϕ:

$$\tilde{L}(\phi) = L(\theta) \text{ where } \theta = g^{-1}(\phi)$$

As a result, since $L(\hat{\theta}) \geq L(\theta)$ by the definition of maximum likelihood

$$\begin{aligned} \tilde{L}(g(\hat{\theta})) = L(\hat{\theta}) &\geq L(\theta) = \tilde{L}(g(\theta)) \\ \text{that is } \tilde{L}(g(\hat{\theta})) &\geq \tilde{L}(\phi)\ \forall \phi \end{aligned}$$

so $g(\hat{\theta})$ maximises the likelihood $\tilde{L}$ and so is the maximum likelihood estimate of ϕ.

g **is not invertible** Note: this part of the proof is not examinable, it is purely for those who are interested. In this case there is not a unique θ corresponding to each ϕ and to define $\tilde{L}$ we need to make a choice as to which θ to use for each ϕ. Define

$$\tilde{L}(\phi) = \max_{\theta : g(\theta) = \phi} L(\theta)$$

then again we can see that the largest value of $\tilde{L}$ occurs at $g(\hat{\theta})$ since this $\hat{\theta}$ maximises L.

Example 2: Maximum likelihood estimation (§1.4, §3.1)

$$\begin{aligned}
L(\theta_1, \theta_2) &= \frac{n!}{n_0! n_1! n_2! n_3!} (\theta_1\theta_2)^{n_0} (\theta_1(1-\theta_2))^{n_1} ((1-\theta_1)\theta_2)^{n_2} ((1-\theta_1)(1-\theta_2))^{n_3} \\
&= \frac{n!}{n_0! n_1! n_2! n_3!} \theta_1^{n_0+n_1} \theta_2^{n_0+n_2} (1-\theta_1)^{n_2+n_3} (1-\theta_2)^{n_1+n_3} \\
\ell(\theta_1, \theta_2) &= constant + (n_0+n_1)\ln\theta_1 + (n_0+n_2)\ln\theta_2 + \\
&\quad (n_2+n_3)\ln(1-\theta_1) + (n_1+n_3)\ln(1-\theta_2) \\
\frac{\partial \ell}{\partial \theta_1} &= \frac{n_0+n_1}{\theta_1} - \frac{n_2+n_3}{1-\theta_1} \\
\frac{\partial \ell}{\partial \theta_2} &= \frac{n_0+n_2}{\theta_2} - \frac{n_1+n_3}{1-\theta_2} \\
\frac{\partial^2 \ell}{\partial \theta_1^2} &= -\frac{n_0+n_1}{\theta_1^2} - \frac{n_2+n_3}{(1-\theta_1)^2} \\
&< 0 \\
\frac{\partial^2 \ell}{\partial \theta_2^2} &= -\frac{n_0+n_2}{\theta_2^2} - \frac{n_1+n_3}{(1-\theta_2)^2} \\
&< 0 \\
\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} &= 0
\end{aligned}$$

Solving $\frac{\partial \ell}{\partial \theta_1} = 0$ and $\frac{\partial \ell}{\partial \theta_2} = 0$ simultaneously:

$$\begin{aligned}
0 &= \frac{n_0+n_1}{\hat{\theta}_1} - \frac{n_2+n_3}{1-\hat{\theta}_1} \\
&= (1-\hat{\theta}_1)(n_0+n_1) - \hat{\theta}_1(n_2+n_3) \\
\hat{\theta}_1 &= \frac{n_0+n_1}{n_0+n_1+n_2+n_3} \\
&= \frac{n_0+n_1}{n} \\
0 &= \frac{n_0+n_2}{\hat{\theta}_2} - \frac{n_1+n_3}{1-\hat{\theta}_2} \\
\hat{\theta}_2 &= \frac{n_0+n_2}{n}
\end{aligned}$$

Asymptotically $\hat{\theta} \sim N(\theta, I_n(\theta_1, \theta_2)^{-1})$ where the information matrix is defined as

$$\begin{aligned}
I_n(\theta_1,\theta_2) &= -\mathsf{E}\begin{bmatrix} \frac{\partial^2\ell}{\partial\theta_1^2} & \frac{\partial^2\ell}{\partial\theta_1\partial\theta_2} \\ \frac{\partial^2\ell}{\partial\theta_1\partial\theta_2} & \frac{\partial^2\ell}{\partial\theta_2^2} \end{bmatrix} \\
\text{Note that } \mathsf{E}(n_0) &= n\mathsf{P}(X=0) \\
&= n\theta_1\theta_2 \\
\text{so } \mathsf{E}(n_0+n_1) &= n\theta_1\theta_2 + n\theta_1(1-\theta_2) \\
\mathsf{E}(n_2+n_3) &= n(1-\theta_1) \text{ etc} \\
\Rightarrow I_n(\theta_1,\theta_2) &= \begin{bmatrix} \frac{n}{\theta_1(1-\theta_1)} & 0 \\ 0 & \frac{n}{\theta_2(1-\theta_2)} \end{bmatrix} \\
\Rightarrow I_n(\theta_1,\theta_2)^{-1} &= \begin{bmatrix} \frac{\theta_1(1-\theta_1)}{n} & 0 \\ 0 & \frac{\theta_2(1-\theta_2)}{n} \end{bmatrix}
\end{aligned}$$

So, asymptotically, $\hat{\theta}_1 \sim N(\theta_1, \frac{\theta_1(1-\theta_1)}{n})$ and $\hat{\theta}_2 \sim N(\theta_2, \frac{\theta_2(1-\theta_2)}{n})$ independently.

Example 3: Maximum likelihood estimation (§1.4, §2.3, §3.1)
EITHER $X_1, \ldots, X_n$ are independent Poisson random variables with mean ϕ^2.

$$\begin{aligned}
f_\phi(\mathbf{x}) &= \prod_{i=1}^n \frac{e^{-\phi^2}(\phi^2)^{x_i}}{x_i!} \\
&= \frac{e^{-n\phi^2}\phi^{2\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \\
\ell(\phi) &= -n\phi^2 + (2\sum_{i=1}^n x_i)\ln\phi + constant \\
\frac{d\ell}{d\phi} &= -2n\phi + \frac{2\sum_{i=1}^n x_i}{\phi} \\
\frac{d^2\ell}{d\phi^2} &= -2n - \frac{2\sum_{i=1}^n x_i}{\phi^2} \quad < 0 \\
\text{Solving } 0 &= -2n\hat{\phi} + \frac{2\sum_{i=1}^n x_i}{\hat{\phi}} \text{ gives}
\end{aligned}$$

$$\hat{\phi}^2 = \frac{\sum_{i=1}^n x_i}{n}$$

So the maximum likelihood estimator of ϕ is $\sqrt{\frac{\sum_{i=1}^n x_i}{n}}$ since ϕ may be taken to be positive.

$$\begin{aligned} I_n(\phi) &= -\mathsf{E}\left[\frac{d^2\ell}{d\phi^2}\right] \\ &= 2n + \frac{2\sum_{i=1}^n \mathsf{E}(X_i)}{\phi^2} \\ &= 2n + \frac{2n\phi^2}{\phi^2} \\ &= 4n \end{aligned}$$

and therefore, asymptotically, $\hat{\phi} \sim N(\phi, \frac{1}{4n})$.
OR, working in terms of the original parameterisation, $X_1, \ldots, X_n$ are independent Poisson random variables with mean θ

$$\begin{aligned} f_\theta(\mathbf{x}) &= \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\ &= \frac{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \\ \ell(\theta) &= -n\theta + (\sum_{i=1}^n x_i)\ln\theta + constant \\ \frac{d\ell}{d\theta} &= -n + \frac{\sum_{i=1}^n x_i}{\theta} \\ \frac{d^2\ell}{d\theta^2} &= -\frac{\sum_{i=1}^n x_i}{\theta^2} \quad < 0 \\ \text{Solving } 0 &= -n + \frac{\sum_{i=1}^n x_i}{\theta} \text{ gives} \\ \hat{\theta} &= \frac{\sum_{i=1}^n x_i}{n} \\ \text{with } I_n(\theta) &= -\mathsf{E}\left(-\frac{\sum_{i=1}^n X_i}{\theta^2}\right) \\ &= \frac{n}{\theta} \end{aligned}$$

So, noting that $\phi = \sqrt{\theta}$ the maximum likelihood estimator of ϕ is $\sqrt{\frac{\sum_{i=1}^n x_i}{n}}$ since ϕ may be taken to be positive. We also know by the Cramer-Rao lower bound that the asymptotic variance is

$$\begin{aligned} \text{CRLB} &= \frac{\left(\frac{d\phi}{d\theta}\right)^2}{I_n(\theta)} \\ &= \frac{\left(\frac{1}{2\theta^{1/2}}\right)^2}{\frac{n}{\theta}} \\ &= \frac{1}{4n} \end{aligned}$$

To see why the CRLB calculated for ϕ as a function of θ is the same as it would be if calculated directly as $I_n(\phi)^{-1}$, recall the notation of the proof of invariance of MLEs:

$$\begin{aligned}
I_n(\theta) &= \mathsf{E}\left[\left(\frac{d\ln f_\theta(\mathbf{x})}{d\theta}\right)^2\right] \\
&= \mathsf{E}\left[\left(\frac{d\ln L(\theta)}{d\theta}\right)^2\right] \\
&= \mathsf{E}\left[\left(\frac{d\ln \tilde{L}(\phi)}{d\theta}\right)^2\right] \text{ where } \phi = \phi(\theta), \text{ in this case } \phi(\theta) = \sqrt{\theta} \\
&= \mathsf{E}\left[\left(\frac{d\ln \tilde{L}(\phi)}{d\phi} \times \frac{d\phi}{d\theta}\right)^2\right] \text{ by the chain rule} \\
&= \mathsf{E}\left[\left(\frac{d\ln \tilde{L}(\phi)}{d\phi}\right)^2\right]\left(\frac{d\phi}{d\theta}\right)^2 \text{ since the derivative is a constant} \\
&= I_n(\phi)(\phi'(\theta))^2 \\
\text{i.e. } \frac{1}{I_n(\phi)} &= \frac{(\phi'(\theta))^2}{I_n(\theta)} \text{ as we expect from the proof of the CRLB}
\end{aligned}$$

Example 4: Maximum likelihood estimation (§1.4, §2.3, §3.1, §3.2)

$$\begin{aligned}
f_{\lambda,\eta}(x) &= \lambda \exp\left(-\lambda(x-\eta)\right), \; x > \eta \\
L(\lambda) &= \prod_{i=1}^{n} \lambda \exp\left(-\lambda(x_i - \eta)\right) \\
&= \lambda^n \exp(-\lambda(\sum_{i=1}^{n}(x_i - \eta))) \\
\ell(\lambda) &= n\ln\lambda - \lambda(\sum_{i=1}^{n} x_i - n\eta) \\
\frac{d\ell}{d\lambda} &= \frac{n}{\lambda} - (\sum_{i=1}^{n} x_i - n\eta) \\
\frac{d^2\ell}{d\lambda^2} &= -\frac{n}{\lambda^2}
\end{aligned}$$

Solving $\frac{d\ell}{d\lambda} = 0 \Rightarrow \hat{\lambda} = (\sum_{i=1}^{n} x_i/n - \eta)^{-1}$ (at which point the second derivative is negative so this is indeed the MLE).

$$\begin{aligned}
I_n(\lambda) &= -\mathsf{E}\left(\frac{d^2\ell}{d\lambda^2}\right) \\
&= \frac{n}{\lambda^2} \\
\Rightarrow \hat{\lambda} &\sim N(\lambda, \frac{\lambda^2}{n}) \text{ asymptotically}
\end{aligned}$$

$$\text{i.e. } \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda^2/n}} \quad \sim \quad N(0,1)$$

and so a $(1-\alpha)100\%$ confidence interval for λ when n is assumed to be large is

$$\hat{\lambda} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\frac{\lambda^2}{n}}$$

which can be approximated (as λ is unknown) by

$$\hat{\lambda} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\frac{\hat{\lambda}^2}{n}}.$$