M3S3/M4S3
ASSESSED COURSEWORK 2

SOLUTIONS

(a) Using the estimator of $I(\theta)$ denoted $\widehat{I}_n\left(\widetilde{\theta}_n\right)$, where

$$\begin{aligned}
\widehat{I}_n\left(\widetilde{\theta}_n\right) &= -\frac{1}{n}\sum_{i=1}^{n}\Psi\left(X_i,\widetilde{\theta}_n\right) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f_X\left(X_i,\theta\right)\Big|_{\theta=\widetilde{\theta}_n} \\
&= -\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n}\log f_X\left(X_i,\theta\right)\Bigg|_{\theta=\widetilde{\theta}_n} \\
&= -\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\, l_n(\theta)\Big|_{\theta=\widetilde{\theta}_n} \\
&= -\frac{1}{n}\ddot{l}_n\left(\widetilde{\theta}_n\right)
\end{aligned}$$

we have

$$W_n = n\left(\widetilde{\theta}_n-\theta_0\right)^T\widehat{I}_n\left(\widetilde{\theta}_n\right)\left(\widetilde{\theta}_n-\theta_0\right) = -\left(\widetilde{\theta}_n-\theta_0\right)^2\ddot{l}_n\left(\widetilde{\theta}_n\right)$$

as $\left(\widetilde{\theta}_n-\theta_0\right)$ is a scalar quantity.

[2 MARKS]

Similarly, for the Rao statistic, we may use

$$\widehat{I}_n(\theta_0) = -\frac{1}{n}\sum_{i=1}^{n}\Psi\left(X_i,\theta_0\right) = -\frac{1}{n}\ddot{l}_n(\theta_0)$$

as an estimator/estimate of $I(\theta_0)$, the single datum or unit information matrix Then

$$\begin{aligned}
Z_n \equiv Z_n(\theta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S\left(X_i,\theta_0\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_X\left(X_i,\theta\right)\Big|_{\theta=\theta_0} \\
&= \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f_X\left(X_i,\theta\right)\Bigg|_{\theta=\theta_0} \\
&= \frac{1}{\sqrt{n}}\dot{l}_n(\theta_0)
\end{aligned}$$

and thus, as all quantities are scalars

$$R_n = Z_n(\theta_0)^T\left[\widehat{I}_n(\theta_0)\right]^{-1}Z_n(\theta_0) = \frac{\left\{Z_n(\theta_0)\right\}^2}{\widehat{I}_n(\theta_0)} = \frac{\left\{\dfrac{1}{\sqrt{n}}\dot{l}_n(\theta_0)\right\}^2}{-\dfrac{1}{n}\ddot{l}_n(\theta_0)} = -\left\{\dot{l}_n(\theta_0)\right\}^2\left\{\ddot{l}_n(\theta_0)\right\}^{-1}$$

[2 MARKS]

For the Rao statistic it is more common and more straightforward to use $\widehat{I}_n(\theta_0)$ rather than $\widehat{I}_n\left(\widetilde{\theta}_n\right)$ as the estimate of the Fisher information, although under the null hypothesis the asymptotic distribution is the same in both cases - using θ_0 is obviously more straightforward as we do not need to compute $\widetilde{\theta}_n$.

(b) For the Poisson case, for $\lambda > 0$

$$f_X(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots$$

and so if $s_n = \sum\limits_{i=1}^{n} x_i$

$$l_n(\lambda) = -n\lambda + s_n \log\lambda - \sum_{i=1}^{n} \log x_i!$$

and so

$$\dot{l}_n(\lambda) = -n + \frac{s_n}{\lambda} \qquad \ddot{l}_n(\lambda) = -\frac{s_n}{\lambda^2}$$

and hence the MLE, from $\dot{l}_n\left(\widehat{\lambda}_n\right) = 0$, is $\widehat{\lambda}_n = s_n/n = \overline{x}$, with estimator $S_n/n = \overline{X}$. Thus

- **Wald Statistic:** using the formula above

$$W_n = -\left(\widetilde{\theta}_n - \theta_0\right)^2 \ddot{l}_n\left(\widetilde{\theta}_n\right) = -\left(\overline{X} - \lambda_0\right)^2 \left(\frac{-S_n}{\left(\overline{X}\right)^2}\right) = n\frac{\left(\overline{X} - \lambda_0\right)^2}{\overline{X}}.$$

[3 MARKS]

- **Rao Statistic:** using the formula above

$$R_n = -\left\{\dot{l}_n(\theta_0)\right\}^2 \left\{\ddot{l}_n(\theta_0)\right\}^{-1} = \frac{-\left(\dfrac{S_n}{\lambda_0} - n\right)^2}{-\dfrac{S_n}{\lambda_0^2}} = \frac{(S_n - n\lambda_0)^2}{S_n} = \frac{n\left(\overline{X} - \lambda_0\right)^2}{\overline{X}}$$

  that is, identical to Wald.

[3 MARKS]

Note: *in this case, we can compute the Fisher Information $I(\lambda_0)$ exactly - we have*

$$I(\lambda_0) = E_{X|\lambda_0}\left[-\Psi(X,\lambda_0)\right] = E_{X|\lambda_0}\left[\frac{X}{\lambda_0^2}\right] = \frac{1}{\lambda_0^2}E_{X|\lambda_0}[X] = \frac{\lambda_0}{\lambda_0^2} = \frac{1}{\lambda_0}$$

so a perhaps preferable version of the Rao statistic is

$$R_n = \frac{\{Z_n(\theta_0)\}^2}{I(\theta_0)} = \frac{\left(\dfrac{1}{\sqrt{n}}\left(\dfrac{S_n}{\lambda_0} - n\right)\right)^2}{\dfrac{1}{\lambda_0}} = \frac{\lambda_0}{n}\left(\frac{S_n}{\lambda_0} - n\right)^2 = \frac{n\left(\overline{X} - \lambda_0\right)^2}{\lambda_0}$$

As a general rule, if the Fisher Information can be computed exactly, then the exact version should be used for the Rao/Score statistic rather than an estimated version.

- **Likelihood Ratio Statistic**: by definition, using the notation $\widetilde{\Lambda}_n$ (... sorry ...)

$$\widetilde{\Lambda}_n = \frac{L_n\left(\widehat{\lambda}_n\right)}{L_n(\lambda_0)} = \frac{e^{-n\widehat{\lambda}_n}\widehat{\lambda}_n^{S_n}}{e^{-n\lambda_0}\lambda_0^{S_n}} = \exp\left\{-n\left(\widehat{\lambda}_n - \lambda_0\right) + S_n\left(\log\widehat{\lambda}_n - \log\lambda_0\right)\right\}$$

  or equivalently

$$2\log\widetilde{\Lambda}_n = -2n\left(\widehat{\lambda}_n - \lambda_0\right) + 2S_n\left(\log\widehat{\lambda}_n - \log\lambda_0\right)$$

[3 MARKS]

(c) Under the normal model, the likelihood is

$$L_n(\mu,\sigma) = f_{X|\mu,\sigma}\left(x;\mu,\sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

and thus, in terms of the random variables, for general X,

$$l(X;\theta) = \log f_{X|\mu,\sigma}\left(X;\mu,\sigma^2\right) = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(X-\mu)^2$$

and, for μ

$$\frac{\partial}{\partial\mu} l(X;\theta) = \frac{1}{\sigma^2}(X-\mu) \qquad\qquad \frac{\partial^2}{\partial\mu^2}\{l(X;\theta)\} = -\frac{1}{\sigma^2}$$

whereas for σ^2

$$\frac{\partial}{\partial\sigma^2}\{l(X;\theta)\} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(X-\mu)^2 \qquad\qquad \frac{\partial^2}{\partial\left(\sigma^2\right)^2}\{l(X;\theta)\} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(X-\mu)^2$$

and

$$\frac{\partial^2}{\partial\mu\partial\sigma^2}\{l(X;\theta)\} = -\frac{1}{\sigma^4}(X-\mu)$$

(here taking σ^2 as the variable with which we differentiating with respect to). Now

$$E_{f_{X|\mu,\sigma}}\left[(X-\mu)\right] = 0 \qquad\qquad E_{f_{X|\mu,\sigma}}\left[(X-\mu)^2\right] = \sigma^2$$

we have for the Fisher Information for $\left(\mu,\sigma^2\right)$ from a single datum as

$$I\left(\mu,\sigma^2\right) = -\begin{bmatrix} E\left[-\frac{1}{\sigma^2}\right] & E\left[-\frac{1}{\sigma^4}(X-\mu)\right] \\ E\left[-\frac{1}{\sigma^4}(X_1-\mu)\right] & E\left[\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(X-\mu)^2\right] \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

say, and $I_n\left(\mu,\sigma^2\right) = nI\left(\mu,\sigma^2\right)$.

(i) The Wald Statistic in this multiparameter setting is, from notes

$$W_n = n\left(\widetilde{\theta}_{n1} - \theta_{10}\right)^T \left[\widehat{I}_n^{11}\left(\widetilde{\theta}_n\right)\right]^{-1} \left(\widetilde{\theta}_{n1} - \theta_{10}\right).$$

Here, σ^2 is **estimated under** $\mathbf{H}_1$ as given in notes, so

$$\widetilde{\theta}_{n1} = \overline{X} \qquad\qquad \theta_{10} = 0 \qquad\qquad \left[\widehat{I}_n^{11}\left(\widetilde{\theta}_n\right)\right]^{-1} = \widehat{I}_{11} - \widehat{I}_{12}\widehat{I}_{22}^{-1}\widehat{I}_{21} = \widehat{I}_{11} = \frac{1}{\widehat{\sigma}^2} = \frac{1}{S^2}$$

$$\Longrightarrow W_n = n\left(\overline{X}\right)^T \left[\frac{1}{S^2}\right]\left(\overline{X}\right) = \frac{n\left(\overline{X}\right)^2}{S^2}$$

[4 MARKS]

(ii) Under H_0, the μ and σ^2 are completely specified, whereas under H_1, the MLEs of μ and σ^2 are

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

Hence the Wald Statistic is

$$W_n = n\left(\widetilde{\theta}_n - \theta_0\right)^T \left[\widehat{I}_n\left(\widetilde{\theta}_n\right)\right]\left(\widetilde{\theta}_n - \theta_0\right) = \begin{bmatrix} \sqrt{n}\left(\overline{X} - 0\right) \\ \sqrt{n}\left(S^2 - \sigma_0^2\right) \end{bmatrix}^T \begin{bmatrix} \frac{1}{S^2} & 0 \\ 0 & \frac{1}{2S^4} \end{bmatrix} \begin{bmatrix} \sqrt{n}\left(\overline{X} - 0\right) \\ \sqrt{n}\left(S^2 - \sigma_0^2\right) \end{bmatrix}$$

$$= \frac{n\left(\overline{X}\right)^2}{S^2} + \frac{n\left(S^2 - \sigma_0^2\right)^2}{2S^4}$$

[3 MARKS]

To clarify notation, if f_X, l, S *and* Ψ *denote the density, its log, the score (first partial derivative of* l *wrt* θ*) and the second partial derivative*

$$l(\theta) = \log f_X(X;\theta)$$

$$S(\theta) \equiv S(X;\theta) = \frac{\partial}{\partial\theta}\{l(\theta)\} \qquad \text{a } k \times 1 \text{ vector}$$

$$\Psi(\theta) \equiv \Psi(X;\theta) = \frac{\partial^2}{\partial\theta^2}\{l(\theta)\} \qquad \text{a } k \times k \text{ matrix}$$

with the "full-likelihood" versions

$$l_n(\theta) = \sum_{i=1}^{n} \log f_X(X;\theta) \qquad S_n(\theta) \equiv S_n(X,\theta) \equiv \frac{\partial}{\partial\theta}\{l_n(\theta)\} \qquad \Psi_n(\theta) \equiv \Psi_n(X,\theta) = \frac{\partial^2}{\partial\theta^2}\{l_n(\theta)\}$$

- **UNIT INFORMATION MATRIX** (with scalar X)

$$I(\theta) = E_{X|\theta}\left[S(X;\theta)\, S(X;\theta)^T\right] = -E_{X|\theta}\left[\Psi(X;\theta)\right]$$

- **FULL LIKELIHOOD INFORMATION MATRIX** (with vector $X = (X_1, ..., X_n)$)

$$I_n(\theta) = E_{X|\theta}\left[S_n(X;\theta)\, S_n(X;\theta)^T\right] = -E_{X|\theta}\left[\Psi_n(X;\theta)\right] = nI(\theta)$$

- **ESTIMATORS**

$$\widehat{I}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} S(X_i;\theta)\, S(X_i;\theta)^T = -\frac{1}{n}\sum_{i=1}^{n} \Psi\left(X_i, \widetilde{\theta}_n\right) \qquad \text{estimator of } I(\theta)$$

$$\widehat{I}_n^n(\theta) = n\widehat{I}_n(\theta) = \sum_{i=1}^{n} S(X_i;\theta)\, S(X_i;\theta)^T = -\sum_{i=1}^{n} \Psi(X_i, \theta) \qquad \text{estimator of } I_n(\theta)$$

- **ESTIMATES (OBSERVED INFORMATION)** (with observed data)

$$\widehat{\mathcal{I}}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} S(x_i;\theta)\, S(x_i;\theta)^T = -\frac{1}{n}\sum_{i=1}^{n} \Psi(x_i, \theta) \qquad \text{estimate of } I(\theta)$$

$$\widehat{\mathcal{I}}_n^n(\theta) = n\widehat{\mathcal{I}}_n(\theta) \qquad \text{estimate of } I_n(\theta)$$