M3S3/M4S3 : SOLUTIONS 3

1. (a) Using the hint given; we know, by properties of vector random variables,

$$Var[Y] = Var\left[\sum_{i=1}^{k} a_i X_i\right] = Var\left[\boldsymbol{a}^\mathsf{T}\boldsymbol{X}\right] = \boldsymbol{a}^\mathsf{T}\Sigma\boldsymbol{a}$$

where variances taken with respect to the distribution of Y and $\boldsymbol{X}$ on the left and right hand sides respectively. But Y is a scalar random variable that is non degenerate, provided $a_i \neq 0$ for $i = 1, \ldots, k$. Thus $Var[Y] > 0$, and hence $\boldsymbol{a}^\mathsf{T}\Sigma\boldsymbol{a} > 0$. *Note that this solution assumes at least one X_i is non degenerate (with variance > 0).*

(b) As $\Sigma\,\Pi = \mathbf{1}_k$, the $k \times k$ identity, we have by multiplying out the block matrices

$$\begin{aligned}
\Sigma_{11}\Pi_{11} + \Sigma_{12}\Pi_{21} &= \mathbf{1}_d && (1)\\
\Sigma_{11}\Pi_{12} + \Sigma_{12}\Pi_{22} &= \mathbf{0} && (2)\\
\Sigma_{21}\Pi_{11} + \Sigma_{22}\Pi_{21} &= \mathbf{0} && (3)\\
\Sigma_{21}\Pi_{12} + \Sigma_{22}\Pi_{22} &= \mathbf{1}_{k-d} && (4)
\end{aligned}$$

From equation (2), premultiplying by Σ_{11}^{-1} and rearranging, we have

$$\Pi_{12} = -\Sigma_{11}^{-1}\Sigma_{12}\Pi_{22} \tag{5}$$

and thus from equation (4) we have

$$\Sigma_{21}(-\Sigma_{11}^{-1}\Sigma_{12}\Pi_{22}) + \Sigma_{22}\Pi_{22} = \mathbf{1}_{k-d} \qquad \therefore \qquad (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\,\Pi_{22} = \mathbf{1}_{k-d}$$

and hence

$$\Pi_{22} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}. \tag{6}$$

Substituting back into equation (5) yields

$$\Pi_{12} = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}. \tag{7}$$

Now, by symmetry of form, we can exchange the roles of the indices and deduce immediately that

$$\begin{aligned}
\Pi_{11} &= (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} && (8)\\
\Pi_{21} &= -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}. && (9)
\end{aligned}$$

Thus we have Σ^{-1} in terms of the blocks of Σ.

2. As I is presumed positive definite and hence non-singular, we have immediately that

$$\det I \equiv |I| = I_{11}I_{22} - I_{12}I_{21} > 0.$$

Using the above formulae (or the ones from lectures), we know that in this scalar case

$$I^{11} = \left(I_{11} - \frac{I_{12}I_{21}}{I_{22}}\right)^{-1} = \frac{I_{22}}{I_{11}I_{22} - I_{12}I_{21}}$$

so

$$(I_{11})^{-1} < I^{11} \qquad \Longleftrightarrow \qquad \frac{1}{I_{11}} < \frac{I_{22}}{I_{11}I_{22} - I_{12}I_{21}} \qquad \Longleftrightarrow \qquad I_{11}I_{22} - I_{12}I_{21} < I_{11}I_{22}.$$

as I_{11} and $I_{11}I_{22} - I_{12}I_{21}$ are positive. This leaves the inequality $I_{12}I_{21} > 0$; but in this scalar case, by symmetry of I, we know that $I_{21} = I_{12}$, so it is **always** true that $I_{12}I_{21} = I_{12}^2 > 0$ unless $I_{12} = 0$, in which case the parameters are orthogonal.

3. We have, by the quadratic approximation,

$$\boldsymbol{l}_n(\boldsymbol{\theta}) = \boldsymbol{l}_n(\widehat{\boldsymbol{\theta}}_n) + \dot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^{\mathsf{T}}\ddot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)$$

But $\widehat{\boldsymbol{\theta}}_n$ is the MLE, so

$$\dot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}$$

so, in fact,

$$\boldsymbol{l}_n(\boldsymbol{\theta}) = \boldsymbol{l}_n(\widehat{\boldsymbol{\theta}}_n) + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^{\mathsf{T}}\ddot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) \tag{10}$$

and as $\boldsymbol{l}_n(\widehat{\boldsymbol{\theta}}_n)$ is a constant, the right hand side has a functional dependence on $\boldsymbol{\theta}$ only through the quadratic form. This form explains the role of the *curvature*, or second partial derivative matrix

$$-\Psi(\boldsymbol{\theta}; X) = \ddot{l}(\boldsymbol{\theta}; X)$$

as

$$\ddot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n) = \sum_{i=1}^{n} \ddot{l}(\widehat{\boldsymbol{\theta}}_n; X_i) = -\sum_{i=1}^{n} \Psi(\widehat{\boldsymbol{\theta}}_n; X_i)$$

At $\widehat{\boldsymbol{\theta}}_n$, the log-likelihood curves *downwards* at a rate determined by $\ddot{\boldsymbol{l}}_n(\widehat{\boldsymbol{\theta}}_n)$.

(a) If $X_i \sim \mathit{Poisson}(\lambda)$, let $s_n = \sum_{i=1}^{n} x_i$. Then

$$\begin{aligned} l_n(\lambda) &= \mathit{constant} + s_n \log\lambda - n\lambda \\ \dot{l}_n(\lambda) &= s_n/\lambda - n \\ \ddot{l}_n(\lambda) &= -s_n/\lambda^2 \end{aligned}$$

and as the MLE is $\widehat{\lambda}_n = \bar{x}$, we have from equation (10) the likelihood approximation

$$\boldsymbol{l}_n(\lambda) = \boldsymbol{l}_n(\widehat{\lambda}_n) - \frac{1}{2}\frac{s_n}{\widehat{\lambda}_n^2}(\lambda - \widehat{\lambda}_n)^2 = \boldsymbol{l}_n(\bar{x}) - \frac{n(\lambda - \bar{x})^2}{2\bar{x}}$$

(b) If $X_i \sim N(0, \sigma^2) \equiv N(0, \theta)$, say, where $\theta = \sigma^2$. Then, if $q_n = \sum_{i=1}^{n} x_i^2$, we have

$$\begin{aligned} l_n(\theta) &= \mathit{constant} - \frac{n}{2}\log\theta - \frac{q_n}{2\theta} \\ \dot{l}_n(\theta) &= -\frac{n}{2\theta} + \frac{q_n}{2\theta^2} \\ \ddot{l}_n(\theta) &= \frac{n}{2\theta^2} - \frac{q_n}{\theta^3} \end{aligned}$$

The MLE is $\widehat{\theta}_n = q_n/n$, and thus

$$\ddot{l}_n(\widehat{\theta}_n) = \frac{n}{2\widehat{\theta}_n^2} - \frac{q_n}{\widehat{\theta}_n^3} = -\frac{n^3}{2q_n^2}$$

we have from equation (10) the likelihood approximation

$$\boldsymbol{l}_n(\theta) = \boldsymbol{l}_n(\widehat{\theta}_n) - \frac{1}{2}\frac{n^3}{2q_n^2}(\theta - \widehat{\theta}_n)^2 = \boldsymbol{l}_n(q_n/n) - \frac{n^3(\theta - q_n/n)^2}{4q_n^2}.$$

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

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

we have

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

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

Similarly, for the Rao statistic, we may use

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

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

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

and thus, as all quantities are scalars

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

For the Rao statistic it is more common and more straightforward to use $\widehat{I}_n(\theta_0)$ rather than $\widehat{I}_n(\widetilde{\theta}_n)$ 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 $\lambda > 0$

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

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

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

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 $\dot{l}_n(\widehat{\lambda}_n) = 0$, is $\widehat{\lambda}_n = s_n/n = \overline{x}$, with estimator $S_n/n = \overline{X}$. Thus

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

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

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

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

  that is, identical to Wald.

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

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

so a perhaps preferable version of the Rao statistic is

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

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 Λ_n here

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

  or equivalently

$$2 \log \Lambda_n = -2n(\widehat{\lambda}_n - \lambda_0) + 2S_n(\log \widehat{\lambda}_n - \log \lambda_0)$$

(c) Under the normal model, the likelihood is

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

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

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

and, for μ

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

whereas for σ^2

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

and

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

(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 (μ, σ^2) from a single datum as

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

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

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

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

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

$$\widetilde{\theta}_{n1} = \overline{X} \qquad\qquad \theta_{10} = 0 \qquad\qquad \left[\widehat{I}_n^{11}(\widetilde{\theta}_n)\right]^{-1} = \widehat{I}_{n11} - \widehat{I}_{n12}\widehat{I}_{n22}^{-1}\widehat{I}_{n21} = \widehat{I}_{n11} = \frac{1}{\widehat{\sigma}^2} = \frac{1}{S^2}$$

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

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

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

Hence the Wald Statistic is

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