APTS

Statistical Asymptotics 2013

Solutions to example sheet questions

Please send details of any errors to:
Andrew.Wood@nottingham.ac.uk

Solution to Question 1

(a) Sample $y_1, \ldots, y_n \overset{\text{IID}}{\sim}$ Poisson(λ).

$$L(\lambda) = \prod_{i=1}^{n} \left\{ \frac{\lambda^{y_i} \mathrm{e}^{-\lambda}}{y_i!} \right\} = \exp\left\{ \left(\sum_{i=1}^{n} y_i \right) \log \lambda - n\lambda - \log \left(\prod_{i=1}^{n} y_i! \right) \right\}.$$

This is a $(1,1)$ exponential family with natural parameter $\theta = \log \lambda$ and natural statistic $\sum_{i=1}^{n} y_i$.

(b) Sample $y_1, \ldots, y_n \overset{\text{IID}}{\sim}$ Binomial(n, p).

$$\begin{aligned} L(p) &= \prod_{i=1}^{n} \left\{ \binom{m}{y_i} p^{y_i} (1-p)^{m-y_i} \right\} \\ &= \exp\left\{ \left(\sum_{i=1}^{n} y_i \right) \log \left(\frac{p}{1-p} \right) + nm \log(1-p) + \log \left(\prod_{i=1}^{n} \binom{m}{y_i} \right) \right\}. \end{aligned}$$

This is a $(1,1)$ exponential family with natural parameter $\theta = \log(p/(1-p))$ and natural statistic $\sum_{i=1}^{n} y_i$.

(c) Sample $y_1, \ldots, y_n \overset{\text{IID}}{\sim}$ Geometric(p).

$$L(p) = \prod_{i=1}^{n} (1-p) p^{y_i} = \exp\left\{ n \log(1-p) + \left(\sum_{i=1}^{n} y_i \right) \log p \right\}.$$

This is a $(1,1)$ exponential family with natural parameter $\theta = \log p$ and natural statistic $\sum_{i=1}^{n} y_i$.

(d) Sample $y_1, \ldots, y_n \overset{\text{IID}}{\sim}$ Gamma(α, β). α known.

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \left\{ \frac{\beta^{\alpha}}{\Gamma(\alpha)} y_i^{\alpha-1} \mathrm{e}^{-\beta y_i} \right\} \\ &= \exp\left\{ -\beta \sum_{i=1}^{n} y_i + (\alpha - 1) \sum_{i=1}^{n} \log y_i + n \log \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)} \right) \right\}. \end{aligned}$$

This is a $(1,1)$ exponential family with natural parameter $\theta = -\beta$ and natural statistic $\sum_{i=1}^{n} y_i$.

(e) Sample $y_1, \ldots, y_n \overset{\text{IID}}{\sim}$ Gamma(α, β). α, β unknown.

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^{n} \left\{ \frac{\beta^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} \mathrm{e}^{-\beta y_i} \right\} \\ &= \exp\left\{ -\beta \sum_{i=1}^{n} y_i + (\alpha - 1) \sum_{i=1}^{n} \log y_i + n \log\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right) \right\}. \end{aligned}$$

This is a $(2, 2)$ exponential family with natural parameter $\theta = (\theta_1, \theta_2)^\top$ where $\theta_1 = -\beta$ and $\theta_2 = \alpha - 1$, and natural statistic $t = (t_1, t_2)^\top$ where

$$t_1 = \sum_{i=1}^{n} y_i \quad \text{and} \quad t_2 = \sum_{i=1}^{n} \log y_i.$$

(f) The negative binomial probability mass function is

$$\begin{aligned} P(Y = y) &= \int_{\lambda=0}^{\infty} \frac{\lambda^y \mathrm{e}^{-\lambda}}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda} \, \mathrm{d}\lambda \\ &= \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)\Gamma(y+1)} \frac{\beta^\alpha}{(\beta + 1)^{\alpha+y}}. \end{aligned}$$

This is *not* exponential family when α is unknown because $\Gamma(\alpha + y)$, which depends on both the parameter α and the observation y, cannot be factorised in the required fashion.

Solution to Question 2

Sample $y_1, \ldots, y_n \overset{\text{IID}}{\sim} N(\mu, \mu^2)$.

So

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\mu^2}} \exp\left(-\frac{1}{2} \frac{(y_i - \mu)^2}{\mu^2} \right) \right\} \\ &= \exp\left\{ \frac{1}{\mu} \sum_{i=1}^{n} y_i - \frac{1}{2\mu^2} \sum_{i=1}^{n} y_i^2 - \frac{n}{2} - \frac{n}{2} \log(2\pi\mu^2) \right\}. \end{aligned}$$

This is $(2,1)$ exponential family with natural statistic $t = (t_1, t_2)^\top$, where

$$t_1 = \sum_{i=1}^{n} y_i \quad \text{and} \quad t_2 = \sum_{i=1}^{n} y_i^2$$

and the natural parameter $\theta = (\theta_1, \theta_2)^\top$ is a function of the real-valued parameter, μ, and is given by

$$\theta_1 \equiv \theta_1(\mu) = \frac{1}{\mu} \quad \text{and} \quad \theta_2 \equiv \theta_2(\mu) = -\frac{1}{2\mu^2}.$$

Consequently, this is a $(2,1)$ exponential family.

Solution to Question 3

Let Θ denote the parameter space and let $g \colon \Theta \to \Theta$ denote a function which is $1:1$ and onto.

Suppose $L(\theta)$ is the likelihood for θ. Then, since g is $1:1$ and onto, there exists an $\tilde{L} \colon \Theta \to \mathbb{R}$ such that $\tilde{L}(g(\theta))$ is the likelihood for $g(\theta)$ and, in particular, $\tilde{L}(g(\theta)) \equiv L(\theta)$ for all $\theta \in \Theta$.

If $\hat{\theta}$ is the MLE for θ, then by definition,

$$L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta).$$

But $L(\hat{\theta}) = \tilde{L}(g(\hat{\theta}))$ and

$$\begin{aligned} &\sup_{\theta \in \Theta} L(\theta) = \sup_{\theta \in \Theta} \tilde{L}(g(\theta)) \\ &\implies \tilde{L}(g(\hat{\theta})) = \sup_{\theta \in \Theta} \tilde{L}(g(\theta)) = \sup_{g(\theta) \in \Theta} \tilde{L}(g(\theta)). \end{aligned}$$

Therefore $g(\hat{\theta})$ is the MLE of $g(\theta)$, so the MLE is equivariant.

Solution to Question 4

$y_1 \sim N(\mu, \tau(1-\rho^2))$

$$y_j = \mu + \rho(y_{j-1} - \mu) + \epsilon_j, \quad j = 2, \ldots, n, \qquad \epsilon_i \overset{\text{IID}}{\sim} N(0, \tau).$$

Then by direct calculation or from standard results in time series analysis,

$$E(y_i) = \mu, \quad i = 1, \ldots, n,$$

$$\begin{aligned} V = \operatorname{cov}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} &= \frac{\tau}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & & & \\ \rho^2 & & \ddots & & \\ \vdots & & & \ddots & \\ \rho^{n-1} & \ldots & \ldots & \ldots & 1 \end{pmatrix} \\ &= \frac{\tau}{1-\rho^2} R. \end{aligned}$$

From standard theory for an AR(1) process,

$$R^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & \ldots & \ldots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \ldots & 0 \\ 0 & -\rho & 1+\rho^2 & & & 0 \\ \vdots & 0 & \vdots & \ddots & 1+\rho^2 & -\rho \\ \vdots & \vdots & & & & \\ 0 & 0 & \ldots & 0 & -\rho & 1 \end{pmatrix}$$

and

$$\det(R^{-1}) = (1-\rho^2)^{-n+1}.$$

Therefore

$$\begin{aligned} V^{-1} &= \tau^{-1}(1-\rho^2)R^{-1} \\ &= \frac{1}{\tau}\begin{pmatrix} 1 & -\rho & 0 & \ldots & 0 \\ -\rho & 1+\rho^2 & & & \\ 0 & & \ddots & & \\ \vdots & & & 1+\rho^2 & -\rho \\ 0 & \ldots & 0 & -\rho & 1 \end{pmatrix}. \end{aligned}$$

Consequently, the log-likelihood is given by

$$l(\mu, \tau, \rho) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\tau + \frac{1}{2}\log(1-\rho^2)$$

$$-\frac{1}{2\tau}\{(y_1-\mu)^2+(y_n-\mu)^2\}-\frac{1}{2\tau}\sum_{i=2}^{n-1}(y_i-\mu)^2(1+\rho^2)+\frac{\rho}{\tau}\sum_{i=2}^{n}(y_i-\mu)(y_{i-1}-\mu).$$

When $\mu=0$ this reduces to

$$l(\tau,\rho)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\tau+\frac{1}{2}\log(1-\rho^2)+\theta_1(\tau,\rho)S_1(y)+\theta_2(\tau,\rho)S_2(y)+\theta_3(\tau,\rho)S_3(y)$$

where

$$\theta_1(\tau,\rho)=-\frac{1}{2\tau},\quad \theta_2(\tau,\rho)=-\frac{(1+\rho^2)}{2\tau},\quad \theta_3(\tau,\rho)=\frac{\rho}{\tau},$$

$$S_1(y)=y_1^2+y_n^2,\quad S_2(y)=\sum_{i=2}^{n-1}y_i^2,\quad S_3(y)=\sum_{i=2}^{n}y_iy_{i-1}.$$

This is a (3.2) exponential family with natural statistics $\theta_1,\theta_2,\theta_3$, which depend on ρ and τ.

Solution to Question 5

Sample $y_1,\dots,y_n \overset{\text{IID}}{\sim} \text{Poisson}(\theta)$. New parametrisation: $\psi=\mathrm{e}^{-\theta}$.

The log-likelihood for θ is

$$\begin{aligned}
l(\theta)&=\sum_{i=1}^{n}\log\left\{\frac{\theta^{y_i}\mathrm{e}^{-\theta}}{y_i!}\right\}\\
&=\left(\sum_{i=1}^{n}y_i\right)\log\theta-n\theta-\sum_{i=1}^{n}\log(y_i!)\\
S(\theta)&=\frac{\partial l}{\partial\theta}(\theta)=\theta^{-1}\left(\sum_{i=1}^{n}y_i\right)-n\\
j(\theta)&=-\frac{\partial^2 l}{\partial\theta^2}(\theta)=\theta^{-2}\left(\sum_{i=1}^{n}y_i\right)\\
i(\theta)&=E_\theta[j(\theta)]=\theta^{-2}\times n\theta=n/\theta.
\end{aligned}$$

In the new parametrisation: $\theta=\log\left(\frac{1}{\psi}\right)$.

Define $\tilde{l}(\psi)=l\{\theta(\psi)\}$. Then

$$\tilde{l}(\psi)=\left(\sum_{i=1}^{n}y_i\right)\log\log\left(\frac{1}{\psi}\right)-n\log\left(\frac{1}{\psi}\right)-\sum_{i=1}^{n}\log(y_i!).$$

Noting that

$$\frac{\partial}{\partial\psi}\log\log\left(\frac{1}{\psi}\right)=\frac{1}{\log\left(\frac{1}{\psi}\right)}\frac{\partial}{\partial\psi}\log\left(\frac{1}{\psi}\right)=-\frac{1}{\psi\log\left(\frac{1}{\psi}\right)}$$

and

$$\frac{\partial^2}{\partial\psi^2}\log\log\left(\frac{1}{\psi}\right)=-\frac{\partial}{\partial\psi}\frac{1}{\psi\log\left(\frac{1}{\psi}\right)}=\frac{1}{\psi^2\log\left(\frac{1}{\psi}\right)}-\frac{1}{\psi^2\left\{\log\left(\frac{1}{\psi}\right)\right\}^2},$$

it is seen that

$$\tilde{S}(\psi)=\frac{\partial\tilde{l}}{\partial\psi}(\psi)=-\left(\sum_{i=1}^n y_i\right)\times\frac{1}{\psi\log\left(\frac{1}{\psi}\right)}+\frac{n}{\psi},$$

$$\tilde{j}(\psi)=-\frac{\partial^2\tilde{l}}{\partial\psi^2}(\psi)=\left(\sum_{i=1}^n y_i\right)\times\frac{1}{\psi^2\left\{\log\left(\frac{1}{\psi}\right)\right\}^2}-\frac{\sum_{i=1}^n y_i}{\psi^2\log\left(\frac{1}{\psi}\right)}+\frac{n}{\psi^2}$$

and

$$\begin{aligned}\tilde{i}(\psi)=E_\psi[\tilde{j}(\psi)]&=\frac{n\log\left(\frac{1}{\psi}\right)}{\psi^2\left\{\log\left(\frac{1}{\psi}\right)\right\}^2}-\frac{n\log\left(\frac{1}{\psi}\right)}{\psi^2\log\left(\frac{1}{\psi}\right)}+\frac{n}{\psi^2}\\&=\frac{n}{\psi^2\log\left(\frac{1}{\psi}\right)}.\end{aligned}$$

Note that

$S(\theta)$ and $\tilde{S}(\psi)$ are different
$j(\theta)$ and $\tilde{j}(\psi)$ are different
$i(\theta)$ and $\tilde{i}(\psi)$ are different

but

$$\tilde{S}(\psi)=S\{\theta(\psi)\}\theta'(\psi)$$

and

$$\tilde{i}(\psi)=i(\theta(\psi))\{\theta'(\psi)\}^2.$$

Also, setting

$S(\hat{\theta})=0$ gives $\hat{\theta}=n^{-1}\sum_{i=1}^n y_i$

$\tilde{S}(\hat{\psi})=0$ gives $\hat{\psi}=\mathrm{e}^{-n^{-1}\sum_{i=1}^n y_i}=\mathrm{e}^{-\hat{\theta}}$,

so the MLE is equivariant; see Question 3.

Solution to Question 6

Let the cell totals be n_1, n_2, n_3 and n_4 and write $n = n_1 + n_2 + n_3 + n_4$. The pmf is given by

$$f(n_1, n_2, n_3, n_4 \mid \theta) = \binom{n}{n_1\, n_2\, n_3\, n_4} \pi_1(\theta)^{n_1} \pi_2(\theta)^{n_2} \pi_3(\theta)^{n_3} \pi_4(\theta)^{n_4}.$$

A statistic $S = S(n_1, n_2, n_3, n_4)$ is minimal sufficient if

$$\frac{f_S(S(n_1, n_2, n_3, n_4) = s)}{f_S(S(m_1, m_2, m_3, m_4) = s)} \text{ independent of } \theta \Leftrightarrow S(n_1, n_2, n_3, n_4) = S(m_1, m_2, m_3, m_4).$$

Let us see whether the full sample (n_1, n_2, n_3, n_4) is minimal sufficient.

Clearly we have $\Leftarrow$ because $m_1 = n_1, \ldots, m_4 = n_4$ implies that the likelihood ratio is independent of θ.

Conversely, for the likelihood ratio to be independent of θ we must have

$$(1-\theta)^{m_1 - n_1} (1+\theta)^{m_2 - n_2} (2-\theta)^{m_3 - n_3} (2+\theta)^{m_4 - n_4}$$

independent of θ. But if any of $m_i - n_i$ is non-zero, then the likelihood ratio is a non-constant rational function of θ, and so cannot be independent of θ. Therefore, the minimal sufficient statistic is the full sample (n_1, n_2, n_3, n_4).

Solution to Question 7

Write $l(\psi,\chi)$ for the log-likelihood in the (ψ,χ) parametrisation. Then

$$\psi \text{ and } \chi \text{ orthogonal} \Leftrightarrow E_{\psi,\chi}\left[\frac{\partial^2 l}{\partial\psi\partial\chi}(\psi,\chi)\right] = 0.$$

Suppose now we transform to $\psi = g(\alpha)$ and $\chi = h(\beta)$, where g and h are $1:1$ and smooth.

As these transformations are $1:1$ and smooth, g' and h' are finite and non-zero for all α, β respectively.

Define

$$\tilde{l}(\alpha,\beta) = l(g(\alpha), h(\beta)).$$

Then

$$\frac{\partial \tilde{l}(\alpha,\beta)}{\partial\alpha} = g'(\alpha)\frac{\partial l}{\partial\psi}(g(\alpha), h(\beta))$$

and

$$\begin{aligned}\frac{\partial^2 \tilde{l}(\alpha,\beta)}{\partial\alpha\partial\beta} &= g'(\alpha)h'(\beta)\frac{\partial l}{\partial\psi\partial\chi}(g(\alpha), h(\beta)) \\ &= g'(\alpha)h'(\beta)l_{\psi,\chi}(\psi,\chi).\end{aligned}$$

Therefore, since $g'(\alpha)$ and $h'(\beta)$ are finite and non-zero,

$$E_{\psi,\chi}\left[\frac{\partial^2 l(\psi,\chi)}{\partial\psi\partial\chi}\right] = 0 \Leftrightarrow E_{\alpha,\beta}\left[\frac{\partial^2 \tilde{l}(\alpha,\beta)}{\partial\alpha\partial\beta}\right] = 0.$$

So if ψ and χ are orthogonal, so are $\alpha = g^{-1}(\psi)$ and $\beta = h^{-1}(\chi)$.

Solution to Question 8

$\theta = (\psi, \lambda)$. Switch to parametrisation (ψ, ϕ) where $\lambda = \lambda(\psi, \phi)$.

Then

$$\begin{aligned}\tilde{l}(\psi, \phi) &= l(\psi, \lambda(\psi, \phi)) \\ &= s_1(y)^\top c_1(\psi) + s_2(y)^\top c_2(\psi, \lambda(\psi, \phi)) - k(\psi, \lambda(\psi, \phi)).\end{aligned}$$

$$\frac{\partial \tilde{l}}{\partial \phi} = s_2(y)^\top \frac{\partial c_2}{\partial \lambda^\top} \frac{\partial \lambda}{\partial \phi^\top} - \frac{\partial k}{\partial \lambda^\top} \frac{\partial \lambda}{\partial \phi^\top}.$$

So

$$E_{\psi,\phi}\left[\frac{\partial \tilde{l}}{\partial \phi}\right] = 0 \Rightarrow \phi^\top \frac{\partial c_2}{\partial \lambda^\top} \frac{\partial \lambda}{\partial \phi^\top} = \frac{\partial k}{\partial \lambda^\top} \frac{\partial \lambda}{\partial \phi^\top} \tag{$*$}$$

since $E(s_2(y)] = \phi$ by definition.

Now differentiate $(*)$ with respect to ψ, noting that ψ and ϕ are functionally independent in the new parametrisation, to obtain

$$\phi^\top \frac{\partial}{\partial \psi}\left[\frac{\partial l_2}{\partial \lambda^\top} \frac{\partial \lambda}{\partial \phi^\top}\right] = \frac{\partial}{\partial \psi}\left[\frac{\partial k}{\partial \lambda^\top} \frac{\partial \lambda}{\partial \phi^\top}\right]. \tag{$**$}$$

But

$$\frac{\partial^2 \tilde{l}}{\partial \psi \partial \phi^\top} = s_2(y)^\top \frac{\partial}{\partial \psi}\left[\frac{\partial c_2}{\partial \lambda^\top} \frac{\partial \lambda^\top}{\partial \phi}\right] - \frac{\partial}{\partial \psi}\left[\frac{\partial k}{\partial \lambda^\top} \frac{\partial \lambda^\top}{\partial \phi}\right]$$

so

$$\begin{aligned}E_{\psi,\phi}\left[\frac{\partial^2 \hat{l}}{\partial \psi \partial \phi^\top}\right] &= \phi^\top \frac{\partial}{\partial \psi}\left[\frac{\partial c_2}{\partial \lambda^\top} \frac{\partial \lambda^\top}{\partial \phi}\right] - \frac{\partial}{\partial \psi}\left[\frac{\partial k}{\partial \lambda^\top} \frac{\partial \lambda^\top}{\partial \phi}\right] \\ &= 0,\end{aligned}$$

due to $(**)$. So ψ and ϕ are orthogonal.

– – –

We may write

$$f(y; \psi, \phi) = \exp\left\{(\psi - 1)\log y - \frac{1}{\phi} y - \psi \log \phi - \log \Gamma(\psi)\right\}.$$

The mean of y is $\psi\phi$ and $\psi - 1$, or ψ, is the natural parameter for $\log y$.

So from the first part of the question, ψ and $\psi\phi$ are orthogonal.

Solution to Question 9

$f(y|\lambda, \gamma) = a(\lambda, y) \exp\{\lambda t(y; \gamma)\} \quad \lambda \in \mathbb{R},\ \gamma \in \mathbb{R}^k$

$$l(\lambda, \gamma) = \log\{a(\lambda, y)\} + \lambda t(y; \gamma)$$

$$\frac{\partial l}{\partial \lambda} = t(y; \gamma), \quad \frac{\partial l}{\partial \gamma} = \lambda \frac{\partial t}{\partial \gamma}, \quad \frac{\partial^2 l}{\partial \lambda \partial \gamma} = \frac{\partial t}{\partial \gamma}$$

We know that $E\left(\frac{\partial l}{\partial \lambda}\right) = 0$, $E\left(\frac{\partial l}{\partial \gamma}\right) = 0$, so, assuming $\lambda \neq 0$, it must also be the case that

$$E\left(\frac{\partial^2 l}{\partial \lambda \partial \gamma}\right) = \frac{1}{\lambda} E\left(\frac{\partial l}{\partial \gamma}\right) = 0.$$

This implies that λ and γ are orthogonal.

—

$$\begin{aligned}
& t(y; \gamma) = \gamma^\top y - k(\gamma) \\
\Longrightarrow \ & f(y|\lambda, \gamma) = a(\lambda, y) \exp\{\lambda \gamma y - \lambda k(\gamma)\} \\
\Longrightarrow \ & \int a(\lambda, y) \mathrm{e}^{\lambda \gamma y} \, \mathrm{d}y = \mathrm{e}^{\lambda k(\gamma)}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\int a(\lambda, y) \mathrm{e}^{\lambda \gamma y} \mathrm{e}^{\theta y} \, \mathrm{d}y &= \int a(\lambda, y) \mathrm{e}^{\lambda\left(\gamma + \frac{\theta}{\lambda}\right) y} \, \mathrm{d}y \\
&= \mathrm{e}^{\lambda k\left(\gamma + \frac{\theta}{\lambda}\right)}
\end{aligned}$$

from which it follows that

$$\int f(y|\lambda, \gamma) \mathrm{e}^{\theta y} \, \mathrm{d}y = \mathrm{e}^{\lambda k\left(\gamma + \frac{\theta}{\lambda}\right) - \lambda k(\gamma)}$$

which implies that the cumulant generating function of y is

$$\lambda \left\{ k\left(\gamma + \frac{\theta}{\lambda}\right) - k(\gamma) \right\}.$$

The mean of y is

$$E(y) = \frac{\partial}{\partial \theta} \lambda \left\{ k\left(\gamma + \frac{\theta}{\lambda}\right) - k(\gamma) \right\}\Bigg|_{\theta=0} = \lambda . \frac{1}{\lambda} \frac{\partial k}{\partial \gamma} \left(\gamma + \frac{\theta}{\lambda}\right)\Bigg|_{\theta=0} = \frac{\partial k}{\partial \gamma} = \mu(\gamma).$$

The variance of y, $V(\mu)$, is

$$\mathrm{Var}(y) = V(\mu) = \frac{\partial^2}{\partial \theta^2} \lambda \left\{ k\left(\gamma + \frac{\theta}{\lambda}\right) - k(\gamma) \right\} = \frac{1}{\lambda} \frac{\partial^2 k}{\partial \gamma^2}\Bigg|_{\gamma = \gamma(\mu)}$$

$$P(y;\phi,\lambda) = \frac{\sqrt{\lambda}}{\sqrt{2\pi}} y^{-\frac{3}{2}} \mathrm{e}^{\sqrt{\lambda\phi}} \exp\left\{-\frac{1}{2}\left(\frac{\lambda}{y} + \phi y\right)\right\}.$$

Put $\phi = -2\lambda\psi$. Then

$$\int \sqrt{\frac{\lambda}{2\pi}} y^{-\frac{3}{2}} \exp\left\{-\frac{\lambda}{2y}\right\} \exp\{\lambda\psi y\} = \mathrm{e}^{-\lambda\sqrt{-2\psi}}.$$

So

$$\int \sqrt{\frac{\lambda}{2\pi}} y^{-\frac{3}{2}} \exp\left\{-\frac{\lambda}{2y}\right\} \exp\{\lambda\psi y\} \exp(\theta y)\ \mathrm{d}y = \exp\{-\lambda\sqrt{-2(\psi+\theta/\lambda)}\}$$

$$\implies E(\mathrm{e}^{\theta y}) = \exp\left[-\lambda\sqrt{-2(\psi+\theta/\lambda)} + \lambda\sqrt{-2\psi}\right].$$

Therefore $k(\psi) = -\sqrt{-2\psi}$ and $a(\lambda, y) = \sqrt{\frac{\lambda}{2\pi}} y^{-\frac{3}{2}} \mathrm{e}^{-\lambda/(2y)}$.

Moreover

$$E(y) = \frac{\sqrt{2}}{2}\frac{1}{\sqrt{-2\psi}} = \frac{1}{\sqrt{-2\psi}} = \mu = \sqrt{\frac{\lambda}{\phi}},$$

$$\mathrm{Var}(y) = \frac{1}{\lambda}\frac{1}{\sqrt{2}}\frac{1}{2}\frac{1}{(\sqrt{-\psi})^3} = \frac{1}{\lambda}\mu^3 = \sigma^2 V(\mu) = \frac{\sqrt{\lambda}}{(\sqrt{\phi})^3}$$

where $\sigma^2 = 1/\lambda$ and $V(\mu) = \mu^3$.

y_i has pdf

$$f_i(y|\gamma,\lambda) = a(\lambda w_i, y_i)\mathrm{e}^{\lambda w_i(\gamma y_i - k(\gamma))}$$

so joint pdf of $y_1, \ldots, y_n$ is

$$\prod_{i=1}^{n} a(\lambda w_i, y_i) \exp\left\{\lambda\left(\gamma\sum_{i=1}^{n} w_i y_i - w_+ k(\gamma)\right)\right\} = \left\{\prod_{i=1}^{n} a(\lambda w_i, y_i)\right\} \exp\{\lambda w_+(\gamma t - k(\gamma))\}$$

where $t = \sum_{i=1}^{n} w_i y_i / w_+$.

The marginal pdf of t is given by

$$\int \left(\prod_{i=1}^{n} a(\lambda w_i, y_i)\right) \exp\{\lambda w_+(\gamma t - k(\gamma))\}\ \mathrm{d}y_1\, \mathrm{d}y_n$$

$$\{y_1, \ldots, y_n \colon \sum w_i y_i / w_+ = t\}$$

$$= a_+(\lambda, t) \exp\{\lambda w_+(\gamma t - k(\gamma))\}$$

where

$$a_+(\lambda, t) = \int_{(y_1, \ldots t_n)^\top \in A_t} \prod_{i=1}^{n} a(\lambda w_i, y_i) \, \mathrm{d}y_1 \ldots \mathrm{d}y_n,$$

and $A_t = \{(y_1 \ldots y_n)^\top : \sum w_i y_i / w_+ = t\}$. Finally, choose $w_i = 1$; then $w_+ = n$. We have

$$y_i \sim ED(\mu, \sigma^2 V(\mu))$$

and so

$$\bar{y} \sim ED\left(\mu, \frac{\sigma^2}{n} V(\mu)\right).$$

From above,

$$\mu = \sqrt{\frac{\lambda}{\phi}} = \sqrt{\frac{n\lambda}{n\phi}}$$

and

$$\frac{\sigma^2}{n} V(\mu) = \frac{1}{n} . \frac{1}{\lambda} \left(\frac{\lambda}{\phi}\right)^{\frac{3}{2}} = \frac{\sqrt{n\lambda}}{(\sqrt{n\phi})^{\frac{3}{2}}}$$

$$\implies \phi \mapsto n\phi, \quad \lambda \mapsto n\lambda.$$

Consequently,

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i \sim IG(n\phi, n\lambda).$$

Solution to Question 10

$y_i \sim \text{Poisson}\{\exp(\lambda + \psi x_i)\}, \quad i = 1, \ldots, n; \quad y_i$ independent.

Joint pmf (by independence) is

$$p_{\boldsymbol{Y}}(y_1, \ldots, y_n) = \prod_{i=1}^{n} \frac{\{\exp(\lambda + \psi x_i)\}^{y_i} \exp(-\mathrm{e}^{\lambda + \psi x_i})}{y_i!}.$$

Pmf of sum $S = \sum_{j=1}^{n} Y_j$ is

$$p_S(s) = \frac{\left(\sum_{i=1}^{n} \exp(\lambda + \psi x_i)\right)^s \exp\left(-\sum_{i=1}^{n} \mathrm{e}^{\lambda + \psi x_i}\right)}{s!}.$$

Conditional distribution of $\boldsymbol{Y} = (Y_1, \ldots, Y_n) \mid S = s$ is

$$\begin{aligned}
\frac{p_{\boldsymbol{Y}}(y_1, \ldots, y_n)}{p_{\boldsymbol{S}}(s)} &= \frac{\prod_{i=1}^{n} \left[\frac{\{\exp(\lambda + \psi x_i)\}^{y_i} \exp(-\mathrm{e}^{\lambda + \psi x_i})}{y_i!}\right]}{\frac{\left(\sum_{i=1}^{n} \mathrm{e}^{\lambda + \psi x_i}\right)^s \exp\left(-\sum_{i=1}^{n} \mathrm{e}^{\lambda + \psi x_i}\right)}{s!}} \\
&= \frac{s!}{\prod y_i!} \prod_{i=1}^{n} \left(\frac{\exp(\lambda + \psi x_i)}{\sum_{j=1}^{n} \exp(\lambda + \psi x_j)}\right)^{y_i} = \frac{s!}{\prod_{i=1}^{n} y_i!} \prod_{i=1}^{n} \frac{\mathrm{e}^{\psi x_i y_i}}{\left(\sum_{i=1}^{n} \mathrm{e}^{\psi x_j}\right)^{y_i}}.
\end{aligned}$$

This is independent of λ.

The conditional log-likelihood $l_c(\psi|s)$ is given by

$$l_c(\psi|s) = \left(\sum_{i=1}^{n} \psi x_i y_i\right) - s \log\left(\sum_{j=1}^{n} \mathrm{e}^{\psi x_j}\right) + \text{const.}$$

– – –

To find the profile log likelihood $l_P(\psi)$, first define

$$\begin{aligned}
l(\psi, \lambda) &= \log p_{\boldsymbol{Y}}(y_1, \ldots, y_n) \\
&= \sum_{i=1}^{n} [y_i(\lambda + \psi x_i) - \mathrm{e}^{\lambda + \psi x_i}] + \text{const.}
\end{aligned}$$

Then

$$\frac{\partial l}{\partial \lambda} = s - \mathrm{e}^{\lambda} \sum_{i=1}^{n} \mathrm{e}^{\psi x_i}$$

and further work shows that the MLE of λ for fixed ψ satisfies

$$\frac{s}{\sum_{i=1}^n \mathrm{e}^{\psi x_i}} = \mathrm{e}^{\hat{\lambda}_\psi}, \quad \hat{\lambda}_\psi = \log s - \log\left(\sum_{i=1}^n \mathrm{e}^{\psi x_i}\right).$$

Substituting,

$$l(\psi, \hat{\lambda}_\psi) = \sum_{i=1}^n y_i x_i \psi - s \log\left(\sum_{i=1}^n \mathrm{e}^{\psi x_i}\right) + \text{const.}$$

which agrees with the conditional log-likelihood.

Solution to Question 11

$$l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2$$

$$l_P(\mu) = \sup_{\sigma^2 > 0} l(\mu, \sigma^2) = l(\mu, \hat{\sigma}^2_\mu)$$

where $\hat{\sigma}^2_\mu$ maximises the log-likelihood for μ. Straightforward calculation shows that $\hat{\sigma}^2_\mu = n^{-1}\sum_{i=1}^n (y_i - \mu)^2$ and therefore

$$l_P(\mu) = -\frac{n}{2}\log\left\{\sum_{i=1}^n (y_i - \mu)^2\right\} + \text{const.}$$

An asymptotically correct confidence interval for μ based on the profile log-likelihood $l_P(\mu)$ will be of the form

$$\{\mu \colon 2[l_P(\hat{\mu}) - l_P(\mu)] \le c_{1-\alpha}\},$$

where $c_{1-\alpha}$ is such that $P[\chi^2_1 \le c_{1-\alpha}] = 1 - \alpha$. Since $N(0,1)^2 \overset{d}{=} \chi^2_1$, it follows that $c_{1-\alpha} = z^2_{1-\alpha/2}$ where $z_{1-\alpha/2}$ is such that $P[N(0,1) \le z_{1-\alpha/2}] = 1 - \alpha/2$. Noting that $\hat{\mu} = \bar{y} = n^{-1}\sum_{i=1}^n y_i$ and writing $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (y_i - \bar{y})^2$ for the full MLE of σ^2,

$$\begin{aligned} 2[l_P(\hat{\mu}) - l_P(\mu)] &= n \log\left\{\frac{\sum_{i=1}^n (y_i - \mu)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right\} \\ &= n \log\left\{\frac{\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right\} \\ &= n \log\left\{1 + \frac{1}{n}\frac{n(\bar{y} - \mu)^2}{\hat{\sigma}^2}\right\}. \end{aligned}$$

Therefore

$$
\begin{aligned}
&2[l_P(\hat{\mu}) - l_P(\mu)] \leq z^2_{1=\alpha/2} \\
\iff\ & 1 + \frac{(\bar{y}-\mu)^2}{\hat{\sigma}^2} \leq \mathrm{e}^{z^2_{1-\alpha/2}/n} \\
\iff\ & \frac{|\bar{y}-\mu|}{\hat{\sigma}} \leq \sqrt{\mathrm{e}^{z^2_{1-\alpha/2}/n} - 1} \\
\iff\ & \mu \in \left(\bar{y} - \hat{\sigma}\sqrt{\mathrm{e}^{z^2_{1-\alpha/2}/n} - 1}, \bar{y} + \hat{\sigma}\sqrt{\mathrm{e}^{z^2_{1-\alpha/2}/n} - 1} \right).
\end{aligned}
$$

Note that this is not the same as the standard t-interval for μ,

$$
\left(\bar{y} - \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1,1-\alpha/2}, \bar{y} + \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1,1-\alpha/2} \right).
$$

However, as $n \to \infty$,

$$
\begin{aligned}
\sqrt{\mathrm{e}^{z^2_{1-\alpha/2}/n} - 1} &= (1 + z^2_{1-\alpha/2}/n - 1 + O(n^{-2}))^{\frac{1}{2}} \\
&= \frac{z_{1-\alpha/2}}{\sqrt{n}}.
\end{aligned}
$$

So as $n \to \infty$, the confidence interval for μ converges to

$$
\left(\bar{y} - \frac{\hat{\sigma} z_{1=\alpha/2}}{\sqrt{n}}, \bar{y} + \frac{\hat{\sigma}}{\sqrt{n}} z_{1-\alpha/2} \right).
$$

This interval is asymptotically correct as $n \to \infty$ but it does not fully account for the extra uncertainty due to the estimation of σ^2 (unlike the t-interval, which does fully account for this uncertainty).

Solution to Question 12

Recall from the slides (see Part III on Edgeworth expansions) that, by definition,

$$H_r(y) = (-1)^r \left. \frac{\mathrm{d}^r \phi(y)}{\mathrm{d}y^r} \middle/ \phi(y) \right.$$

or, equivalently,

$$H_r(y)\phi(y) = (-1)^r \frac{\mathrm{d}^r \phi(y)}{\mathrm{d}y^r}.$$

The identity follows from repeated integration by parts (use induction to make it fully rigorous). In particular, consider

$$\int_{-\infty}^{\infty} \mathrm{e}^{ty} H_r(y)\phi(y)\ \mathrm{d}y = t^r \mathrm{e}^{\frac{1}{2}t^2}.$$

Integrating the LHS by parts (integrate e^{ty}, differentiate $\phi(y)H_r(y)$), we obtain

$$\begin{aligned}\left[\frac{1}{t}\,\mathrm{e}^{ty} H_r(y)\phi(y)\right]_{-\infty}^{\infty} &- \int \frac{1}{t}\mathrm{e}^{ty}\frac{\mathrm{d}}{\mathrm{d}y}(H_r(y)\phi(y))\ \mathrm{d}y \\ &= 0 - \frac{1}{t}\int_{-\infty}^{\infty} \mathrm{e}^{ty}\frac{\mathrm{d}}{\mathrm{d}y}(-1)^r \frac{\mathrm{d}^r\phi(y)}{\mathrm{d}y^r}\ \mathrm{d}y \\ &= \frac{1}{t}\int_{-\infty}^{\infty} \mathrm{e}^{ty}(-1)^{r+1}\frac{\mathrm{d}^{r+1}}{\mathrm{d}y^{r+1}}\phi(y)\ \mathrm{d}y = \frac{1}{t}\int_{-\infty}^{\infty} \mathrm{e}^{ty} H_{r+1}(y)\phi(y)\ \mathrm{d}y.\end{aligned}$$

So

$$t^{-1}\int_{-\infty}^{\infty} \mathrm{e}^{ty} H_{r+1}(y)\phi(y)\ \mathrm{d}y = t^r \mathrm{e}^{\frac{1}{2}t^2},$$

and, multiplying both sides by t, the identity follows.

– – –

Recall that

$$S_n^* = \frac{S_n - n\mu}{n^{\frac{1}{2}}\sigma},$$

where $S_n = \sum_{i=1}^n Y_i$ and $E(Y_i) = \mu$, $\mathrm{Var}(Y_i) = \sigma^2$.

Let $\kappa_1(=\mu)$, $\kappa_2(=\sigma^2), \kappa_3, \ldots$ denote the cumulants of Y_1. Then

$$\mathrm{cum}_1(S_n^*) \equiv E(S_n^*) = 0,$$

$$
\begin{aligned}
&\mathrm{cum}_2(S_n^*) \equiv \mathrm{Var}(S_n^*) = 1,\\
&\mathrm{cum}_3(S_n^*) = \frac{n\kappa_3}{(n^{\frac{1}{2}}\sigma)^3} = n^{-\frac{1}{2}}\frac{\kappa_3}{\sigma^3} = n^{-\frac{1}{2}}\rho_3
\end{aligned}
$$

and

$$
\mathrm{cum}(S_n^*) = \frac{n\kappa_j}{(n^{\frac{1}{2}}\sigma)^j} = n^{1-j/2}\rho_j,
$$

where ρ_j is the jth standardised cumulant of Y_1.

It follows from the definition of the CGF that

$$
K_{S_n^*}(t) = \tfrac{1}{2}t^2 + \tfrac{1}{6}n^{-\frac{1}{2}}\rho_3 t^3 + \frac{n^{-1}}{24}\rho_4 t^4 + O(n^{-\frac{3}{2}}).
$$

Therefore, from the relationship between MGF and CGF,

$$
\begin{aligned}
M_{S_n^*}(t) &= \exp\{K_n^*(t)\}\\
&= \exp\left\{\tfrac{1}{2}t^2 + \frac{n^{-\frac{1}{2}}}{6}\rho_3 t^3 + \frac{n^{-1}}{24}\rho_4 t^4 + O(n^{-\frac{3}{2}})\right\}\\
&= \mathrm{e}^{\frac{1}{2}t^2}\exp\left\{\frac{n^{-\frac{1}{2}}}{6}\rho_3 t^3 + \frac{n^{-1}}{24}\rho_4 t^4 + O(n^{-\frac{3}{2}})\right\}\\
&= \mathrm{e}^{\frac{1}{2}t^2}\left[1 + \frac{n^{-\frac{1}{2}}}{6}\rho_3 t^3 + \frac{n^{-1}}{24}\rho_4 t^4 + \frac{n^{-1}}{72}\rho_3^2 t^6 + O(n^{-\frac{3}{2}})\right] \qquad (*)
\end{aligned}
$$

as required. Note that in the final step, we used a second-order Taylor expansion, assuming that n is large.

Note that if we define

$$
f_{S_n^*}(y) = \phi(y)\left[1 + \frac{n^{-\frac{1}{2}}}{6}\rho_3 H_3(y) + \frac{n^{-1}}{24}\rho_4 H_4(y) + \frac{n^{-1}}{72}\rho_3^2 H_6(y) + O(n^{-\frac{3}{2}})\right]
$$

then

$$
M_{S_n^*}(t) = \int_{-\infty}^{\infty} \mathrm{e}^{ty}\, f_{S_n^*}(y)\,\mathrm{d}y = \text{RHS of } (*).
$$

Solution to Question 13

The result will follow if we can show that, for all integers $r \geq 1$,

$$\int_{-\infty}^{y} H_r(x)\phi(x)\ \mathrm{d}x = -\phi(y)H_{r-1}(y).$$

From the defining property of Hermite polynomials,

$$\phi(x)H_r(x) = (-1)^r \frac{\mathrm{d}^r}{\mathrm{d}x^r}\phi(x).$$

So from the fundamental theorem of calculus,

$$\begin{aligned}\int_{-\infty}^{y} H_r(x)\phi(x)\ \mathrm{d}x &= \int_{-\infty}^{y} (-1)^r \frac{\mathrm{d}^r}{\mathrm{d}x^r}\phi(x)\ \mathrm{d}x \\ &= (-1)\left[(-1)^{r-1}\frac{\mathrm{d}^{r-1}}{\mathrm{d}x^{r-1}}\phi(x)\right]_{-\infty}^{y} \\ &= -\phi(y)H_{r-1}(y), \quad \text{as required.}\end{aligned}$$

Solution to Question 14

Here, $K_{S_n}(t) = n\mu t + \frac{n}{2}\sigma^2 t^2$ and

$$\hat{f}_S(s) = \frac{1}{\sqrt{2\pi K''_{S_n}(\hat{t})}}\, \mathrm{e}^{K_{S_n}(\hat{t}) - \hat{t}s} \qquad (*)$$

where $\hat{t}$ solves $K'_{S_n}(\hat{t}) = s$.

Now $K'_{S_n}(t) = n\mu + n\sigma^2 t$, so

$$s = n\mu + n\sigma^2\hat{t} \Rightarrow \hat{t} = \frac{S - n\mu}{n\sigma^2}.$$

Also, $K''_{S_n}(t) = n\sigma^2$,

$$\begin{aligned}K_{S_n}(\hat{t})s - \hat{t}(s) &= n\mu\left(\frac{S-n\mu}{n\sigma^2}\right) + \frac{n\sigma^2}{2}\left(\frac{S-n\mu}{n\sigma^2}\right)^2 - \left(\frac{S-n\mu}{n\sigma^2}\right)s \\ &= \frac{(S-n\mu)^2}{2n\sigma^2} = \frac{(S-n\mu)^2}{n\sigma^2} = -\frac{1}{2}\frac{(S-n\mu)^2}{n\sigma^2}.\end{aligned}$$

Substituting into $(*)$, we obtain

$$\hat{f}_{S_n}(s) = \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(S-n\mu)^2}{n\sigma^2}\right\}$$

which is exactly equal to the pdf of $N(n\mu, n\sigma^2)$.

Solution to Question 15

$$M_{Y_1}(t) = E(\mathrm{e}^{tY_1}) = (1-t)^{-1},$$

so $M_{S_n}(t) = (1-t)^{-n} = \exp\{K_{S_n}(t)\}$ where $K_{S_n}(t) = -n\log(1-t)$. So

$$K'_{S_n}(t) = \frac{n}{1-t}, \quad K''_{S_n}(t) = \frac{n}{(1-t)^2}$$

$$K'_{S_n}(\hat{t}) = s \Rightarrow \hat{t} - 1 - \frac{n}{s}.$$

The saddlepoint approximation at $S_n = s$ is given by

$$\begin{aligned}
\hat{f}_{S_n}(s) &= \frac{1}{\sqrt{2\pi K''_{S_n}(\hat{t})}} \exp\{K_{S_n}(\hat{t}) - \hat{t}s\} \\
&= \frac{1}{\sqrt{2\pi s^2/n}} \exp\left\{-n\log\left(\frac{n}{s}\right) - s - n\right\} \\
&= \sqrt{\frac{n}{2\pi}} \times \frac{\mathrm{e}^{-n}}{n^n} s^{n-1}\mathrm{e}^{-s} \\
&= \frac{1}{\hat{\Gamma}(n)} s^{n-1}\mathrm{e}^{-s}
\end{aligned}$$

where $\hat{\Gamma}(n)$ is Stirling's application to $\Gamma(n)$.

But the true pdf of S_n is Gamma with index n and scale parameter 1, i.e.

$$f(y) = \frac{1}{\Gamma(n)} s^{n-1}\mathrm{e}^{-y}.$$

So $\hat{f}_{S_n}(y)$ is exact up to the normalising constant.

Solution to Question 16

This is covered in considerable detail in the preliminary notes and the slides (see part III).

Solution to Question 17

Consider the integral

$$I = \int_a^b \mathrm{e}^{-\lambda g(x)}\ \mathrm{d}x,$$

where $g(x)$ has a unique stationary minimum at $x = \hat{x} \in (a, b)$. Put $z = \lambda^{\frac{1}{2}}(x - \hat{x})$. Then

$$I = \int_a^b \mathrm{e}^{-\lambda g(x)}\ \mathrm{d}x = \lambda^{-\frac{1}{2}} \int_{\lambda^{\frac{1}{2}}(a-\hat{x})}^{\lambda^{\frac{1}{2}}(b-\hat{x})} \mathrm{e}^{-\lambda g(\hat{x}+\lambda^{-\frac{1}{2}}z)}\ \mathrm{d}z.$$

Now

$$g(\hat{x} + \lambda^{-\frac{1}{2}}z) = g(\hat{x}) + 0 + \frac{1}{2!}g''(\hat{x})\frac{z^2}{\lambda} + \frac{1}{3!}g^{(3)}(\hat{x})\frac{z^3}{\lambda^{\frac{3}{2}}} + \frac{1}{4!}g^{(4)}(\hat{x})\frac{z^4}{\lambda^2} + O(\lambda^{-\frac{5}{2}}).$$

Consequently, writing $\hat{g} = g(\hat{x})$, $\hat{g}^{(j)} = g^{(j)}(\hat{x})$ etc,

$$\begin{aligned} I &= \lambda^{-\frac{1}{2}} \int_{\lambda^{\frac{1}{2}}(a-\hat{x})}^{\lambda^{\frac{1}{2}}(b-\hat{x})} \exp\left\{-\lambda\hat{g} - \tfrac{1}{2}\hat{g}^{(2)}z^2 - \frac{\lambda^{-\frac{1}{2}}}{6}g^{(3)}z^3 - \frac{\lambda^{-1}}{24}\hat{g}^{(4)} + O(\lambda^{-\frac{3}{2}})\right\}\ \mathrm{d}z \\ &\sim \frac{\mathrm{e}^{-\lambda\hat{g}}}{\lambda^{\frac{1}{2}}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\hat{g}^{(2)}z^2}\left[1 - \frac{\lambda^{-\frac{1}{2}}}{6}\hat{g}^{(3)}z^3 - \frac{\lambda^{-1}}{24}g^{(4)}z^4 + \frac{\lambda^{-1}}{72}(\hat{g}^{(3)})^2z^6 + O(\lambda^{-\frac{3}{2}})\right]\ \mathrm{d}z. \end{aligned}$$

By symmetry,

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}az^2}z^3\ \mathrm{d}z = 0$$

and also

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}az^2}z^{2j}\ \mathrm{d}z = O(1)$$

for any fixed integer $j \geq 1$. Consequently

$$I = \sqrt{\frac{2\pi}{\lambda\hat{g}^{(2)}}}\mathrm{e}^{-\lambda\hat{g}}[1 + O(\lambda^{-1})],$$

using the Gaussian integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\hat{g}^{(2)}z^2}\ \mathrm{d}z = \sqrt{\frac{2\pi}{\hat{g}^{(2)}}}.$$

Solution to Question 18

$$y_1, \ldots, y_n \overset{\text{IID}}{\sim} \exp(1/\mu)$$

$$\Rightarrow l(\mu) = -n \log \mu - \sum_{i=1}^{n} y_i/\mu$$

$$l'(\mu) = -\frac{n}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{n} y_i$$

$$l'(\hat{\mu}) = 0 \Rightarrow \hat{\mu} = \sum_{i=1}^{n} y_i/n$$

So

$$l(\mu) = -n \log \mu - n\hat{\mu}/\mu$$

$$l(\hat{\mu}) = -n \log \hat{\mu} - n$$

$$\Rightarrow l(\mu) - l(\hat{\mu}) = n \log \hat{\mu}/\mu - n\hat{\mu}/\mu + n$$

$$j(\mu) = -l''(\mu) = -\frac{n}{\mu^2} + \frac{2n}{\mu^3}\hat{\mu}.$$

So

$$|j(\hat{\mu})|^{\frac{1}{2}} = \frac{n^{\frac{1}{2}}}{\hat{\mu}},$$

and

$$p^*(\hat{\mu}) \propto |\hat{j}|^{\frac{1}{2}} \, \mathrm{e}^{l(\mu)-l(\hat{\mu})} \propto \left(\frac{\hat{\mu}}{\mu}\right)^n \frac{1}{\hat{\mu}} \, \mathrm{e}^{-n\hat{\mu}/\mu}$$

$$\Rightarrow p^*(\hat{\mu}) = \frac{1}{\Gamma(n)} \left(\frac{n}{\mu}\right)^n \hat{\mu}^{n-1} \, \mathrm{e}^{-n\hat{\mu}/\mu}.$$

Note that

$$\sum_{i=1}^{n} y_i \sim \text{Gamma}(n, 1/\mu)$$

and so

$$\frac{1}{n} \sum_{i=1}^{n} y_i \sim \text{Gamma}(n, n/\mu).$$

Therefore the p^* formula is *exact* in this case.

Solution to Question 19

$$x_1, \ldots, x_n \overset{\text{IID}}{\sim} \exp(\lambda) \qquad y_1, \ldots, y_n \overset{\text{IID}}{\sim} \exp(\psi\lambda)$$

$$l(\psi, \lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i + n \log(\lambda\psi) - \lambda\psi \sum_{i=1}^{n} y_i$$

$$\frac{\partial l}{\partial \psi} = \frac{n}{\psi} - \lambda \sum y_i, \qquad \frac{\partial l}{\partial \lambda} = \frac{2n}{\lambda} - \sum_{i=1}^{n} x_i - \psi \sum_{i=1}^{n} y_i,$$

$$\frac{\partial l}{\partial \psi} = 0 \text{ and } \frac{\partial l}{\partial \lambda} = 0 \Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i},\ \hat{\psi} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}.$$

Therefore we may write

$$l(\psi, \lambda) = 2n \log \lambda + n \log \psi - n\lambda\hat{\lambda}^{-1} - n\psi\lambda\hat{\psi}^{-1}\hat{\lambda}^{-1}$$

$$-\frac{\partial^2 l}{\partial \psi^2} = \frac{n}{\psi^2}, \quad -\frac{\partial^2 l}{\partial \lambda^2} = \frac{2n}{\lambda^2}, \quad -\frac{\partial^2 l}{\partial \psi \partial \lambda} = n\hat{\psi}^{-1}\hat{\lambda}^{-1}.$$

Therefore

$$\hat{j} = \begin{bmatrix} \frac{n}{\hat{\psi}^2} & \frac{n}{\hat{\psi}\hat{\lambda}} \\ \frac{n}{\hat{\psi}\hat{\lambda}} & \frac{2n}{\hat{\lambda}^2} \end{bmatrix} \quad \text{and} \quad |\hat{j}|^{\frac{1}{2}} = \frac{n}{\hat{\psi}\hat{\lambda}}.$$

So

$$\begin{aligned} p^*(\hat{\psi}, \hat{\lambda}) &\propto |\hat{j}|^{\frac{1}{2}} \exp\{l(\psi, \lambda) - l(\hat{\psi}, \hat{\lambda})\} \\ &= \frac{n\,\mathrm{e}^{2n}}{\hat{\psi}\hat{\lambda}} \left(\frac{\lambda}{\hat{\lambda}}\right)^{2n} \left(\frac{\psi}{\hat{\psi}}\right)^{n} \exp\left\{-\frac{n\lambda}{\hat{\lambda}}\left(1 + \frac{\psi}{\hat{\psi}}\right)\right\}. \end{aligned}$$

Let us evaluate

$$I = \int_0^\infty \int_0^\infty \left(\frac{\lambda}{\hat{\lambda}}\right)^{2n} \left(\frac{\psi}{\hat{\psi}}\right)^{n} \frac{1}{\hat{\psi}\hat{\lambda}} \exp\left\{-\frac{n\lambda}{\hat{\psi}}\left(1 + \frac{\psi}{\hat{\psi}}\right)\right\}\, \mathrm{d}\hat{\lambda}\,\mathrm{d}\hat{\psi}.$$

Put $u = \lambda/\hat{\lambda}$, so $u^{-1}\mathrm{d}u = -\hat{\lambda}^{-1}\mathrm{d}\hat{\lambda}$

$$\begin{aligned} \Rightarrow I &= \int_0^\infty \left(\frac{\psi}{\hat{\psi}}\right)^{n} \frac{1}{\hat{\psi}} \left[\int_0^\infty u^{2n-1}\,\mathrm{e}^{-nu\left(1+\frac{\psi}{\hat{\psi}}\right)}\,\mathrm{d}u\right]\,\mathrm{d}\hat{\psi} \\ &= \int_0^\infty \left(\frac{\psi}{\hat{\psi}}\right)^{n} \frac{1}{\hat{\psi}} \frac{\Gamma(2n)}{n^{2n}} \left(1 + \frac{\psi}{\hat{\psi}}\right)^{2n}\,\mathrm{d}\hat{\psi} \\ &= \frac{\Gamma(2n)}{n^{2n}} \int_0^\infty \frac{1}{\psi} \left(\frac{\hat{\psi}}{\psi}\right)^{n-1} \left(1 + \frac{\hat{\psi}}{\psi}\right)^{-2n}\,\mathrm{d}\hat{\psi} \\ &= \frac{\Gamma(2n)}{n^{2n}} \frac{\Gamma(n)}{\Gamma(2n)} = \frac{\Gamma(n)}{n^{2n}}, \end{aligned}$$

using the normalising constant for the F distribution given in the question. Therefore the p^* approximation to the pdf of $(\hat{\psi}, \hat{\lambda})$ is

$$p^*(\hat{\psi}, \hat{\lambda}) = \frac{n^{2n}}{\Gamma(n)} \frac{1}{\hat{\psi}\hat{\lambda}} \left(\frac{\lambda}{\hat{\lambda}}\right)^{2n} \left(\frac{\psi}{\hat{\psi}}\right)^{n} \exp\left\{-n\frac{\lambda}{\hat{\lambda}}\left(1+\frac{\psi}{\hat{\psi}}\right)\right\}$$

and repeating the first step in the calculation of I above,

$$p^*(\hat{\psi}) = \frac{n^{2n}}{\Gamma(n)} \frac{\Gamma(2n)}{n^{2n}} \frac{1}{\psi} \left(\frac{\hat{\psi}}{\psi}\right)^{n-1} \left(1+\frac{\hat{\psi}}{\psi}\right)^{-2n}$$

which is exact, with the correct normalising constant.

Solution to Question 20

$$y_1, \ldots, y_n \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$$

$$l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)$$

$$\frac{\partial l}{\partial \mu} = 0 \Rightarrow \hat{\mu} = n^{-1}\sum_{i=1}^{n} y_i$$

$$\Rightarrow l_P(\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\mu})^2$$

$$\frac{\partial l_P}{\partial \sigma^2}(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \hat{\mu})^2$$

$$E\left[\frac{\partial l_P}{\partial \sigma^2}(\sigma^2)\right] = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}E\left[\sum_{i=1}^{n}(y_i - \hat{\mu})^2\right]$$

But

$$E\left[\sum_{i=1}^{n}(y_i - \hat{\mu})^2\right] = (n-1)\sigma^2$$

$$\Rightarrow E\left[\frac{\partial l_P}{\partial \sigma^2}(\sigma^2)\right] = -\frac{n}{2\sigma^2} + \frac{(n-1)\sigma^2}{2\sigma^4} = -\frac{1}{2\sigma^2} \neq 0$$

so the profile score is biased in this case.

The modified profile likelihood is given by

$$\tilde{L}_P(\psi) = L_P(\psi)M(\psi).$$

First of all note that

$$\begin{aligned}\sum_{i=1}^{n}(y_i-\mu)^2 &= \sum_{i=1}^{n}(y_i-\hat{\mu})^2 + n(\hat{\mu}-\mu)^2 \\ &= n\hat{\sigma}^2 + n(\hat{\mu}-\mu)^2,\end{aligned}$$

and $(\hat{\mu}, \hat{\sigma}^2)$ is minimal sufficient for the data $y_1, \ldots, y_n$. Therefore we may write

$$\begin{aligned}l(\mu,\sigma^2) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2 \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(n\hat{\sigma}^2 + n(\hat{\mu}-\mu)^2) \\ l_{\mu;\hat{\mu}} &= \frac{\partial^2 l}{\partial\mu\partial\hat{\mu}} = \frac{\partial}{\partial\mu}\left[-\frac{n(\hat{\mu}-\mu)}{\sigma^2}\right] = \frac{n}{\sigma^2} \\ j_{\mu\mu} &= -\frac{\partial^2 l}{\partial\mu^2} = \frac{n}{\sigma^2}.\end{aligned}$$

Recall from the slides that $M(\psi) = |l_{\chi;\hat{\chi}}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi})|^{-1} \times |j_{\chi\chi}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi})|^{\frac{1}{2}}$.

Here, $\psi = \sigma^2$ and $\chi = \mu$, so

$$M(\sigma^2) = \sigma^2/n/(\sigma^2/n)^{\frac{1}{2}} = (\sigma^2/n)^{\frac{1}{2}}.$$

Also, $\hat{\mu}_{\sigma^2} = \hat{\mu}$, so

$$\begin{aligned}\tilde{L}_P(\sigma^2) &= L_P(\sigma^2)M(\sigma^2) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \mathrm{e}^{-n\hat{\sigma}^2/(2\sigma^2)}(\sigma^2/n)^{\frac{1}{2}} \\ &= \left(\frac{1}{2\pi}\right)^{n/2}\left(\frac{1}{\sigma^2}\right)^{(n-1)/2} \mathrm{e}^{-n\hat{\sigma}^2/(2\sigma^2)} \\ \tilde{l}_P(\sigma^2) &= \log\tilde{L}_P(\sigma^2) \\ &= -\frac{(n-1)}{2}\log\sigma^2 - \frac{n\hat{\sigma}^2}{2\sigma^2} + \text{constant}.\end{aligned}$$

So

$$\frac{\partial\tilde{l}_P}{\partial\sigma^2}(\sigma^2) = -\frac{(n-1)}{2\sigma^2} + \frac{n\hat{\sigma}^2}{2\sigma^4}.$$

But

$$E[n\hat{\sigma}^2] = E\left[\sum_{i=1}^{n}(y_i - \hat{\mu})^2\right] = (n-1)\sigma^2$$

so

$$E\left[\frac{\partial \tilde{l}_P}{\partial \sigma^2}(\sigma^2)\right] = -\frac{(n-1)}{2\sigma^2} + \frac{(n-1)\sigma^2}{2\sigma^4} = 0.$$

Consequently, we may conclude that the modified profile score is unbiased.

Solution to Question 21

y_i has pdf $\mu_i^{-1}\,\mathrm{e}^{-y_i/\mu_i}$, $\mu_i = \lambda \mathrm{e}^{\psi x_i}$

$\implies$ log-likelihood $l(\psi, \lambda)$ is

$$l(\psi, \lambda) = -n\log\lambda - \psi\sum_{i=1}^{n} x_i - \lambda^{-1}\sum_{i=1}^{n} y_i \mathrm{e}^{-\psi x_i}$$

$$\frac{\partial l}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n} y_i \mathrm{e}^{-\psi x_i}$$

$\implies$ MLE of λ for fixed ψ is

$$\hat{\lambda}_\psi = \frac{1}{n}\sum_{i=1}^{n} y_i \mathrm{e}^{-\psi x_i}$$

$$\implies \; l_p(\psi) = -n\log\left(\frac{1}{n}\sum_{i=1}^{n} y_i \mathrm{e}^{-\psi x_i}\right) - \psi\sum_{i=1}^{n} x_i - n$$

and

$$L_P(\psi) = \left(\frac{n}{\sum_{i=1}^{n} y_i \mathrm{e}^{-\psi x_i}}\right)^n \mathrm{e}^{-\psi\sum_{i=1}^{n} x_i - n}.$$

Modified profile likelihood is

$$L_{MP}(\psi) = L_P(\psi) M(\psi)$$

where

$$M(\psi) = |l_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi; \hat{\psi}, \hat{\lambda})|^{-1}| - l_{\lambda\lambda}(\psi, \hat{\lambda}_\psi; \hat{\psi}, \hat{\lambda})|^{\frac{1}{2}}$$

where $\hat{\psi}$ and $\hat{\lambda}$ are the full MLEs.

As $\hat{\psi}, \hat{\lambda}$ are not sufficient for the data $y_1, \ldots, y_n$ we need to consider an ancillary statistic:

$$a = (a_1, \ldots, a_n) \quad \text{where} \quad a_i = \log y_i - \log \hat{\lambda} - \hat{\psi} x_i.$$

Consider transforming from $y_1, \ldots, y_n$ to $a_1, \ldots, a_{n-2}, \hat{\psi}, \hat{\lambda}$. [Later we shall see that it makes no difference which subset of $n-2$ a_i's we choose.] Then

$$y_i = \mathrm{e}^{a_j} \hat{\lambda} \mathrm{e}^{\hat{\psi} x_i}, \quad i = 1, \ldots, n-2$$

and

$$y_{n-1} = y_{n-1}(a_1, \ldots, a_{n-2}, \hat{\psi}, \hat{\lambda})$$

and

$$y_n = y_n(a_1, \ldots, a_{n-2}, \hat{\psi}, \hat{\lambda}).$$

From the score equations we have

$$\frac{\partial l}{\partial \lambda} = 0 \Longrightarrow -\frac{n}{\hat{\lambda}} + \frac{1}{\hat{\lambda}^2} \sum_{i=1}^{n-2} \mathrm{e}^{a_i} \hat{\lambda} \mathrm{e}^{\hat{\psi} x_i} \mathrm{e}^{-\hat{\psi} x_i} + \hat{\lambda}^{-2} y_{n-1} \mathrm{e}^{-\hat{\psi} x_{n-1}} + \hat{\lambda}^{-2} y_n \mathrm{e}^{-\hat{\psi} x_n}$$

and

$$\frac{\partial l}{\partial \psi} = 0 \Longrightarrow -\sum_{i=1}^{n} x_i + \hat{\lambda}^{-1} \sum_{i=1}^{n-2} \mathrm{e}^{a_i} \hat{\lambda} \mathrm{e}^{\hat{\psi} x_i} \mathrm{e}^{-\hat{\psi} x_i} x_i + \hat{\lambda}^{-1} x_{n-1} y_{n-1} \mathrm{e}^{-\hat{\psi} x_{n-1}} + \hat{\lambda}^{-1} x_n y_n \mathrm{e}^{-\hat{\psi} x_n}.$$

Thus we have simultaneous equations in y_{n-1} and y_n. It is easily checked that the solution is of the form

$$y_{n-1} = \hat{\lambda} g_{n-1}(a_1, \ldots, a_{n-2}, \hat{\psi})$$
$$y_n = \hat{\lambda} g_n(a_1, \ldots, a_{n-2}, \hat{\psi}),$$

where g_{n-1} and g_n do not depend on $\hat{\lambda}$. Thus

$$\frac{\partial y_{n-1}}{\partial \hat{\lambda}} = y_{n-1}/\hat{\lambda} \quad \text{and} \quad \frac{\partial y_n}{\partial \hat{\lambda}} = y_n/\hat{\lambda}.$$

Consequently,

$$l_{\lambda;\hat{\lambda}}(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = \frac{\partial l}{\partial \hat{\lambda}} \left(-\frac{n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^{n} y_i \mathrm{e}^{-\psi x_i} \right)$$

$$= \frac{1}{\lambda^2}\sum_{i=1}^{n}\frac{\partial y_i}{\partial\hat{\lambda}}\,\mathrm{e}^{-\psi x_i} = \frac{1}{\lambda^2}\frac{1}{\hat{\lambda}}\sum_{i=1}^{n} y_i\,\mathrm{e}^{-\psi x_i}$$

$$\implies l_{\lambda;\hat{\lambda}}(\psi,\hat{\lambda}_\psi;\hat{\psi},\hat{\lambda}) = \frac{n}{\hat{\lambda}\hat{\lambda}_\psi}.$$

Also,

$$-l_{\lambda\lambda} = -\frac{n}{\lambda^2} + \frac{2}{\lambda^3}\sum_{i=1}^{n} y_i\,\mathrm{e}^{-\psi x_i}$$

$$\implies -l_{\lambda\lambda}(\psi,\hat{\lambda}_\psi;\hat{\psi},\hat{\lambda}) = \frac{2n}{\hat{\lambda}_\psi^2} - \frac{n}{\hat{\lambda}_\psi^2} = \frac{n}{\hat{\lambda}_\psi^2}.$$

So

$$M(\psi) = \left(\frac{\hat{\lambda}}{n}\hat{\lambda}_\psi\right)\sqrt{\frac{n}{\hat{\lambda}_\psi^2}} = n^{-\frac{1}{2}}\hat{\lambda}.$$

Note that in this example, $M(\psi)$ does not depend on ψ.

Solution to Question 22

Let

$$l(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

denote the log-likelihood. From standard calculations, the maximised log-likelihood under H_0 is given by

$$l(\mu_0,\hat{\sigma}_0^2) = -\frac{n}{2}\log(2\pi\hat{\sigma}_0^2) - \frac{n}{2},$$

where

$$\hat{\sigma}_0^2 = n^{-1}\sum_{i=1}^{n}(y_i-\mu_0)^2;$$

and the maximised log-likelihood under the alternative is given by

$$l(\hat{\mu},\hat{\sigma}^2) = -\frac{n}{2}\log(2\pi\hat{\sigma}^2) - \frac{n}{2},$$

where

$$\hat{\mu} = n^{-1}\sum_{i=1}^{n} y_i \quad \text{and} \quad \hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu})^2.$$

So twice the log of the ratio of maximised likelihoods is given by

$$\begin{aligned} w &= 2[l(\hat{\mu}, \hat{\sigma}^2) - l(\mu_0, \hat{\sigma}_0^2)] \\ &= n\log\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right). \end{aligned}$$

But

$$\begin{aligned} \hat{\sigma}_0^2 = n^{-1}\sum_{i=1}^{n}(y_i - \mu_0)^2 &= n^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu} + \hat{\mu} - \mu_0)^2 \\ &= n^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu})^2 + (\hat{\mu} - \mu_0)^2 \\ &= \hat{\sigma}^2 + (\hat{\mu} - \mu_0)^2. \end{aligned}$$

So

$$w = n\log\left(\frac{\hat{\sigma}^2 + (\hat{\mu} - \mu)}{\hat{\sigma}^2}\right) = n\log\left(1 + \frac{t^2}{n-1}\right),$$

where $t^2 \equiv (\hat{\mu} - \mu_0)^2/\{\hat{\sigma}^2/(n-1)\}$ is the t-statistic. For n large,

$$\log\left(1 + \frac{t^2}{n-1}\right) = \frac{t^2}{n-1} - \frac{1}{2}\frac{t^4}{(n-1)^2} + O(n^{-3}).$$

Also, from standard results for the t-distribution with $n-1$ degrees of freedom, which you can try to derive or look up,

$$E[T^2] = \frac{n-1}{n-3} = 1 + \frac{2}{n} + O(n^{-2})$$

and

$$E[T^4] = \frac{3(n-1)^2}{(n-3)(n-5)} = 3 + O(n^{-1}).$$

Putting these results together,

$$\begin{aligned} E\left[n\log\left(1 + \frac{T^2}{n-1}\right)\right] &= \left[1 + \frac{2}{n} + O(n^{-2})\right](1 - n^{-1})^{-1} - \frac{1}{2}(3 + O(n^{-1}))\frac{1}{n}(1 - n^{-1})^{-2} \\ &= 1 + \frac{3}{n} + O(n^{-2}) - \frac{3}{2n} + O(n^{-2}) = 1 + \frac{3}{2n} + O(n^{-2}). \end{aligned}$$

So $b \equiv \frac{3}{2}$ in this case.

The Bartlett correction generally improves the χ^2 approximation.

Solution to Question 23

The details are similar to those given for the logistic regression example in Part III of the slides. In particular, writing $\beta_p = \gamma$, the approximation to the marginal posterior of γ is given by

$$\hat{\pi}(\gamma|y) = \frac{L(\hat{\beta}_\gamma)}{(2\pi)^{\frac{1}{2}} L(\hat{\beta})} \left\{ \frac{|j(\hat{\beta})|}{|j_{p-1}(\hat{\beta}_\gamma)|} \right\}^{\frac{1}{2}},$$

where

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{\mathrm{e}^{y_i \beta^\top x_i} \exp(-\mathrm{e}^{\beta^\top x_i})}{y_i!} \right\}$$

is the likelihood for the model;

$\hat{\beta}_\gamma$ is the MLE of β under $H_0\colon \beta_p = \gamma$, $\beta_1, \ldots, \beta_{p-1}$ unrestricted;

$\hat{\beta}$ is the MLE under the general alternative $\beta_1, \ldots, \beta_p$ all unrestricted;

$$j(\beta) = -\frac{\partial^2 l}{\partial \beta \partial \beta^\top} = \sum_{i=1}^{n} x_i x_i^\top \, \mathrm{e}^{\beta^\top x_i}$$

is the observed information matrix for the full model;

$$j_{p-1}(\beta) = \left\{ -\frac{\partial^2 l}{\partial \beta_i \partial \beta_j} \right\}_{i,j=1}^{p-1}$$

is the observed information under the model H, obtained by remaining the pth row and pth column of $j(\beta)$; and $|\cdot|$ denotes determinant.

Solution to Question 24

$$l(\mu_1,\ldots,\mu_n,\sigma^2) = -n\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{(x_i-\mu_i)^2+(y_i-\mu_i)^2\}$$

$$\frac{\partial l}{\partial \mu_i} = \frac{1}{\sigma^2}[(x_i-\mu_i)+(y_i-\mu_i)]$$
$$\frac{\partial l}{\partial \mu_i} = 0 \Rightarrow \hat{\mu}_i = \frac{x_i+y_i}{2}.$$

So

$$l(\hat{\mu}_1,\ldots,\hat{\mu}_n,\sigma^2) = -n\log(2\pi\sigma^2) - \frac{1}{4\sigma^2}\sum_{i=1}^{n}(x_i-y_i)^2$$

and

$$\frac{\partial l}{\partial \sigma^2}(\hat{\mu}_1,\ldots,\hat{\mu}_n,\sigma^2) = -\frac{n}{\sigma^2} + \frac{1}{4\sigma^4}\sum_{i=1}^{n}(x_i-y_i)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{4n}\sum_{i=1}^{n}(x_i-y_i)^2.$$

But $E(x_i-y_i)^2 = 2\sigma^2$, so

$$E(\hat{\sigma}^2) = \frac{\sigma^2}{2}.$$

By the weak law of large numbers,

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2/2 \neq \sigma^2,$$

so $\hat{\sigma}^2$ is *not* a consistent estimator of σ^2.

Use

$$m(\psi) = |l_{\chi;\hat{\chi}}(\psi,\hat{\chi}_\psi;\hat{\psi},\hat{\chi})|^{-1}|j_{\chi\chi}(\psi,\hat{\chi}_\psi;\hat{\psi},\hat{\chi})|^{\frac{1}{2}},$$

where $\chi = (\mu_1,\ldots,\mu_n)^\top$ and $\psi = \sigma^2$.

Now

$$\begin{aligned}(x_i-\mu_i)^2+(y_i-\mu_i)^2 &= (x_i-\hat{\mu}_i+\hat{\mu}_i-\mu_i)^2+(y_i-\hat{\mu}_i+\hat{\mu}_i-\mu_i)^2\\ &= (x_i-\hat{\mu}_i)^2+(y_i-\hat{\mu}_i)^2+2(\hat{\mu}_i-\mu_i)^2.\end{aligned}$$

So

$$
\begin{aligned}
l(\mu_1,\ldots,\mu_n,\sigma^2) &= -n\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{(x_i-\mu_i)^2+(y_i-\mu_i)^2\} \\
&= -n\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{(x_1-\hat{\mu}_i)^2+(y_2-\hat{\mu}_i)^2\} - \frac{n}{\sigma^2}\sum_{i=1}^{n}(\hat{\mu}_i-\mu_i)^2 \\
&= -n\log(2\pi\sigma^2) - \frac{n\hat{\sigma}^2}{\sigma^2} - \frac{n}{\sigma^2}\sum_{i=1}^{n}(\hat{\mu}_i-\mu_i)^2.
\end{aligned}
$$

Now

$$
\frac{\partial^2 l}{\partial\mu_i\partial\hat{\mu}_i} = \frac{2n}{\sigma^2}, \qquad \frac{\partial^2 l}{\partial\mu_i\partial\hat{\mu}_k} = 0 \quad (i \neq k)
$$

$$
j_{\mu_p\mu_q} = -\frac{\partial^2 l}{\partial\mu_p\partial\mu_q} = \begin{cases} 2n/\sigma^2 & p = q \\ 0 & p \neq q \end{cases}
$$

so

$$
|l_{\mu_j\hat{\mu}}(\sigma^2,\hat{\mu}_{\sigma^2};\hat{\sigma}^2,\hat{\mu})|^{-1} = \prod_{i=1}^{n}\frac{\sigma^2}{2n} = \frac{\sigma^{2n}}{(2n)^n}
$$

and $|\hat{j}|^{\frac{1}{2}} = \left(\frac{2n}{\sigma^2}\right)^{n/2}$.

Therefore

$$
M(\sigma^2) = \frac{\sigma^{2n}}{(2n)^n}\Big/\frac{\sigma^n}{(2n)^{n/2}} = c(\sigma^2)^{n/2},
$$

where c is a constant.

Consequently, the log modified profile likelihood is given by

$$
\begin{aligned}
\tilde{l}_P(\sigma^2) &= \log L_P(\sigma^2) + \log M(\sigma^2) \\
&= -n\log\sigma^2 - \frac{n\hat{\sigma}^2}{\sigma^2} - \frac{n}{\sigma^2}\sum_{i=1}^{n}(\hat{\mu}_i-\hat{\mu}_i)^2 + \frac{n}{2}\log\sigma^2 + \text{constant} \\
&= -\frac{n}{2}\log\sigma^2 - \frac{n\hat{\sigma}^2}{\sigma^2} + \text{constant},
\end{aligned}
$$

and

$$
\frac{\partial\tilde{l}_P}{\partial\sigma^2}(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{n\hat{\sigma}^2}{\sigma^4}.
$$

$$
\begin{aligned}
\frac{\partial\tilde{l}_P}{\partial\sigma^2}(\hat{\sigma}_p^2) = 0 &\Rightarrow \hat{\sigma}_p^2 = 2\hat{\sigma}^2 \\
&\Rightarrow \hat{\sigma}_p^2 \text{ is unbiased and consistent.}
\end{aligned}
$$

Distribution of S is

$$2\sigma^2\chi^2_n \sim \mathrm{Gamma}\left(\frac{n}{2}, \frac{1}{4\sigma^2}\right).$$

So the marginal log-likelihood $l_M(\sigma^2)$ for σ^2 based on S is

$$l_M(\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{S}{4\sigma^2}.$$

But

$$n\hat{\sigma}^2 = \frac{S}{4} \Rightarrow l_m(\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{n\hat{\sigma}^2}{\sigma^2}.$$

So l_M agrees with the modified profile log-likelihood up to an additive constant.