Solution 1

(a) The geometric density on $\mathcal{Y}$ is $\log f(Y;\theta) = Y\log(1-\theta)+\log\theta$, and as $\mathrm{E}_g(Y)=\lambda$ we have $\mathrm{E}_g\{\log f(Y;\theta)\} = \lambda\log(1-\theta)+\log\theta$. Differentiation of this with respect to θ yields $-\lambda/(1-\theta)+1/\theta$, and setting this equal to zero yields $\theta_g = (1+\lambda)^{-1}$; note that the second derivative is negative for all θ. The mean of the geometric distribution with support $\mathcal{Y}$ is $1/\theta_g - 1 = \lambda$, so the mis-specified geometric model matches the mean of g.

The log likelihood for an exponential family is $s(y)^{\mathrm{T}}\varphi - k(\varphi) + \log m(y)$, and its expected value under another exponential family model would be

$$\mathrm{E}_g\{s(Y)\}^{\mathrm{T}}\varphi - k(\varphi) + \mathrm{E}_g\{\log m(Y)\},$$

which is to be maximised with respect to φ. This implies that $\mathrm{E}_g\{s(Y)\} = \nabla k(\varphi)$, so φ is chosen so that the mean $\nabla k(\varphi)$ under the candidate model equals the mean of $s(Y)$ under the true model. So this is a general property.

(b) The log likelihood derivatives for a single observation are $\theta^{-1} - Y/(1-\theta)$ and $-\{1/\theta^2 + Y/(1-\theta)^2\}$, and these give $\hbar_1(\theta_g) = 1/\{\theta_g(1-\theta_g)\}$ and $\imath_1(\theta_g) = 1/\{\theta_g^2(1-\theta_g)\}$, so the sandwich variance for a sample of size n is $\hbar(\theta_g)/\{n\imath_1(\theta_g)^2\} = \theta_g^3(1-\theta_g)/n$, or $\lambda/\{(1+\lambda)^4 n\}$.

(c) The log likelihood is $\ell(\theta) = n\overline{Y}\log(1-\theta)+n\log\theta$, which gives $\widehat{\theta} = h(\overline{Y})$, where $h(x) = 1/(1+x)$ for $x>0$. The delta method variance is $h'\{\mathrm{E}(Y)\}^2\mathrm{var}_g(\overline{Y}) = \{-1/(1+\lambda)^2\}^2\lambda/n = \theta_g^4(1-\theta_g/(\theta_g n) = \theta_g^3(1-\theta_g)/n$.

This is not really a surprise, but it is reassuring that the sandwich formula gives the natural variance computed directly by applying the delta method to the formula for the MLE.

Solution 2

(a) The density function is $(\lambda/\psi)(1-y/\psi)_+^{\lambda-1}$, so the log likelihood for a random sample can be written as

$$\ell(\psi,\lambda) = n\log\lambda - n\log\psi + (\lambda-1)\sum_{j=1}^n \log(1-y_j/\psi)_+,$$

and this implies that $\widehat{\lambda}_\psi = \mathrm{argmax}_\lambda \ell(\psi,\lambda) = n/s_\psi$. Hence the profile log likelihood is

$$\ell_{\mathrm{p}}(\psi) \equiv s_\psi - n\log\psi - n\log s_\psi, \quad \psi > \max(y_1,\dots,y_n),$$

because clearly the upper bound to the support of the density, ψ, exceeds all the y_j.

(b) We write the joint density as

$$f(y_1,\dots,y_n;\psi,\lambda) = (\lambda/\psi)^n\exp\{-(\lambda-1)s_\psi\}, \quad \lambda>0,$$

and note that if ψ is fixed then this is a (1,1)-exponential family with canonical parameter $-\lambda$ and canonical statistic s_ψ, which must therefore be minimal sufficient for λ. Now

$$\mathrm{P}(Z>z) = \mathrm{P}\{-\log(1-Y/\psi)>z\} = \mathrm{P}(1-Y/\psi<e^{-z}) = \mathrm{P}\{Y>\psi(1-e^{-z})\} = e^{-\lambda z}, \quad z>0,$$

so $S_\psi = \sum_{j=1}^n Z_j$ is the sum of n independent exponential variables and therefore has a gamma (n,λ) distribution, and density

$$f_{S_\psi}(s_\psi;\psi,\lambda) = \frac{\lambda^n s_\psi^{n-1}}{\Gamma(n)}e^{-\lambda s_\psi}, \quad s_\psi>0.$$

Thus the conditional density of $Y_1,\dots,Y_n$ given $S_\psi = s_\psi$ is of the stated form.

(c) The conditional log likelihood from (b) is

$$\ell_c(\psi) = \log f(y \mid s_\psi; \psi) \equiv s_\psi - n\log\psi - (n-1)\log s_\psi, \quad \psi > \max(y_1, \dots, y_n),$$

so the only difference is the replacement of n by $n-1$ in the last term. The latter is preferable, because it is based on a true density, so it should have slightly better behaviour in small samples.

Solution 3

(a) Differentiation to obtain the first Bartlett identity gives

$$0 = f(y;\psi,\lambda)|_{y=\psi} + \int_0^\psi \frac{\partial f(y;\psi,\lambda)}{\partial\psi}\,\mathrm{d}y = f(y;\psi,\lambda)|_{y=\psi} + \int_0^\psi \frac{\partial \log f(y;\psi,\lambda)}{\partial\psi} f(y;\psi,\lambda)\,\mathrm{d}y,$$

whose first term is zero only if $\lambda > 1$; otherwise the identity fails. As we can write

$$\frac{\partial \log f(y;\psi,\lambda)}{\partial\psi} = \frac{\lambda-1}{\psi(1-y/\psi)} - \frac{\lambda}{\psi},$$

we see that differentiation of the identity

$$\int_0^\psi \frac{\partial \log f(y;\psi,\lambda)}{\partial\psi} f(y;\psi,\lambda)\,\mathrm{d}y = 0$$

will result in an expression

$$\left.\frac{\partial \log f(y;\psi,\lambda)}{\partial\psi} f(y;\psi,\lambda)\right|_{y=\psi}$$

that is finite only if $\lambda > 2$. Thus the standard regularity conditions fail for $\lambda \le 2$.

(b) The sample maximum satisfies

$$\mathrm{P}(M_n \le x) = \mathrm{P}(X \le x)^n = \left\{1 - (1 - x/\psi)^\lambda\right\}^n, \quad 0 < x < \psi,$$

so provided $w \in (0, \psi n^{1/\lambda})$, we can write

$$\begin{aligned}
\mathrm{P}\left\{n^{1/\lambda}(\psi - M_n) \le w\right\} &= \mathrm{P}\left(M_n \ge \psi - wn^{-1/\lambda}\right) \\
&= 1 - \left[1 - \left\{1 - (\psi - wn^{-1/\lambda})/\psi\right\}^\lambda\right]^n \\
&= 1 - \left\{1 - \frac{(w/\psi)^\lambda}{n}\right\}^n \\
&\to 1 - \exp\{-(w/\psi)^\lambda\}, \quad n \to \infty, \\
&= \mathrm{P}(W \le w),
\end{aligned}$$

and this is valid for all $w > 0$ in the limit, because $\psi n^{1/\lambda} \to \infty$. This proves the desired convergence; W has the *Weibull distribution.*

When $\lambda \le 2$, the standardised variable $n^{1/\lambda}(\psi - M_n)$ converges to its limit at least as fast as $n^{1/2}$, so in this case inference based on M_n would preferable to maximum likelihood estimation (even if the latter was regular). But when $\lambda > 2$ the Bartlett identities are satisfied, and maximum likelihood estimation, with its rate of $n^{1/2}$, will be (asymptotically) preferable.

Solution 4

(a) If the inverses exist and we write

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix},$$

then

$$\begin{aligned} A^{11} &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, & A^{12} &= -A_{11}^{-1}A_{12}A^{22}, \\ A^{21} &= -A_{22}^{-1}A_{21}A^{11}, & A^{22} &= (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}, \end{aligned}$$

from which the given formula follows at once.

(b) Differentiation gives

$$\frac{\partial \ell_{\mathrm{p}}(\psi)}{\partial \psi} = \ell_{\psi}(\psi, \widehat{\lambda}_{\psi}) + \frac{\partial \widehat{\lambda}_{\psi}^{\mathrm{T}}}{\partial \psi}\ell_{\lambda}(\psi, \widehat{\lambda}_{\psi}) = \ell_{\psi}(\psi, \widehat{\lambda}_{\psi})$$

because $\widehat{\lambda}_{\psi}$ maximises $\ell(\psi, \lambda)$ in the λ-direction and thus $\ell_{\lambda}(\psi, \widehat{\lambda}_{\psi}) = 0$. Hence

$$\frac{\partial^2 \ell_{\mathrm{p}}(\psi)}{\partial \psi \partial \psi^{\mathrm{T}}} = \ell_{\psi\psi}(\psi, \widehat{\lambda}_{\psi}) + \ell_{\psi\lambda}(\psi, \widehat{\lambda}_{\psi})\frac{\partial \widehat{\lambda}_{\psi}}{\partial \psi^{\mathrm{T}}}.$$

The final expression here is obtained by differentiation of the equation $\ell_{\lambda}(\psi, \widehat{\lambda}_{\psi}) = 0$, giving

$$0 = \ell_{\lambda\psi}(\psi, \widehat{\lambda}_{\psi}) + \ell_{\lambda\lambda}(\psi, \widehat{\lambda}_{\psi})\frac{\partial \widehat{\lambda}_{\psi}}{\partial \psi^{\mathrm{T}}},$$

and thus $\partial \widehat{\lambda}_{\psi}/\partial \psi^{\mathrm{T}} = -\ell_{\lambda\lambda}(\psi, \widehat{\lambda}_{\psi})^{-1}\ell_{\lambda\psi}(\psi, \widehat{\lambda}_{\psi})$, resulting in

$$\frac{\partial^2 \ell_{\mathrm{p}}(\psi)}{\partial \psi \partial \psi^{\mathrm{T}}} = \ell_{\psi\psi}(\psi, \widehat{\lambda}_{\psi}) - \ell_{\psi\lambda}(\psi, \widehat{\lambda}_{\psi})\ell_{\lambda\lambda}(\psi, \widehat{\lambda}_{\psi})^{-1}\ell_{\lambda\psi}(\psi, \widehat{\lambda}_{\psi}).$$

On multiplying by -1 and using the notation $-\ell_{\psi\psi}(\psi, \widehat{\lambda}_{\psi}) = \tilde{\jmath}_{\psi\psi}$, etc., we obtain the expression for $\tilde{\jmath}_{\mathrm{p}}$, which becomes $\widehat{\jmath}_{\mathrm{p}} = \widehat{\jmath}_{\psi\psi} - \widehat{\jmath}_{\psi\lambda}\widehat{\jmath}_{\lambda\lambda}^{-1}\widehat{\jmath}_{\lambda\psi}$ when $\psi = \widehat{\psi}$, giving the required distributional result for $\widehat{\psi}$.