Solution 1

For the expected information we note that the log likelihood

$$\ell(\theta) = \log f(y;\theta) = (\alpha+1)\log(\theta+y) - \log\alpha - \alpha\log\theta, \quad \theta > 0,$$

has second derivative with respect to θ

$$-\frac{\alpha+1}{(\theta+y)^2} + \frac{\alpha}{\theta^2},$$

and as

$$\mathrm{E}\left\{(\theta+Y)^{-2}\right\} = \int_0^\infty \frac{1}{(\theta+y)^2}\frac{\alpha\theta^\alpha}{(\theta+y)^{\alpha+1}}\,\mathrm{d}y = \frac{\alpha}{(\alpha+2)\theta^2}\int_0^\infty \frac{(\alpha+2)\theta^{\alpha+2}}{(\theta+y)^{\alpha+2+1}}\,\mathrm{d}y$$

and the final integral here equals unity, the expected information for a single observation is

$$\imath_1(\theta) = \frac{\alpha}{\theta^2} - (\alpha+1)\frac{\alpha}{(\alpha+2)\theta^2} = \frac{\alpha}{(\alpha+2)\theta^2}.$$

Provided that differentiation with respect to θ and integration with respect to y commute, which is the case here, the Cramèr–Rao lower bound says that no unbiased estimator can have a lower asymptotic variance than $1/\{n\imath_1(\theta)\}$. This applies to the method-of-moments estimator $\tilde\theta$ because we saw previously that $\mathrm{var}(\tilde\theta) = \alpha\theta^2/\{n(\alpha-2)\}$, provided $\alpha > 2$, and therefore

$$\frac{\mathrm{var}(\tilde\theta)}{1/\{n\imath_1(\theta)\}} = \frac{\alpha^2}{(\alpha+2)(\alpha-2)} = \frac{\alpha^2}{\alpha^2-4} > 1,$$

with the ratio exploding as $\alpha \downarrow 2$, i.e., $\tilde\theta$ is increasingly inefficient as $\alpha \downarrow 2$, and actually has infinite variance when $0 < \alpha < 2$.

Solution 2

(a) The variable $Z = \lfloor Y/\delta \rfloor$ is discrete, taking values in $\{0, 1, \ldots\}$, and $\mathrm{P}(Z = z)$ equals

$$\mathrm{P}(\lfloor Y/\delta\rfloor = z) = \mathrm{P}\{Y < (z+1)\delta\} - \mathrm{P}(Y \le z\delta) = 1 - e^{-(z+1)\lambda\delta} - (1 - e^{-z\lambda\delta}) = \left(1 - e^{-\lambda\delta}\right)e^{-z\lambda\delta}.$$

As the Z_j are independent, their likelihood can be expressed as

$$L(\lambda) = f(z_1,\ldots,z_n;\lambda) = \prod_{j=1}^n \left(1 - e^{-\lambda\delta}\right)e^{-z_j\lambda\delta}, \quad \lambda > 0,$$

and the log likelihood is

$$\ell(\lambda) = n\log\left(1 - e^{-\lambda\delta}\right) - \lambda\delta\sum_{j=1}^n z_j.$$

On differentiating this twice with respect to λ we obtain the given result; note that the z_j drop out, so $\mathrm{E}(Z_j)$ is not required.

(b) The limit is obtained because $1 - e^{-\lambda\delta} = \lambda\delta - (\lambda\delta)^2/2 + o(\delta^3)$. It is easy to check that the expected information based on the Y_j is n/λ^2, so the ratio of information quantities (aka the relative efficiency) is

$$\frac{\imath(\delta)}{\imath(0)} = \frac{(\lambda\delta)^2\exp(-\lambda\delta)}{\{1 - \exp(-\lambda\delta)\}^2},$$

which equals $e^{-1}/(1 - e^{-1})^2 = 0.920$ when $\lambda\delta = 1$.

(c) The relative efficiency is 99.92% in this case. Hence there is essentially no loss of information on λ if the data are rounded to the nearest 0.1, when the mean is 1. This is slightly misleading,, because in fact we should use the likelihood based on the Z_j, whereas in practice we would usually use the Z_j in the likelihood based on the Y_j.

(d) If we replace Y with $Z\delta$ in the usual log likelihood, we obtain average log likelihood

$$\overline{\ell}(\lambda') = n^{-1}\left(n\log\lambda' - \lambda'\sum_{j=1}^{n}\delta Z_j\right), \quad \lambda' > 0,$$

and as $n\to\infty$ this converges to $\log\lambda' - \lambda'\delta\mathrm{E}(Z)$, which implies that the maximum likelihood estimator converges to $1/\{\delta\mathrm{E}(Z)\}$. Now $1+Z$ has a geometric distribution with success probability $1-e^{-\lambda\delta}$, so $\mathrm{E}(Z) = (1-e^{-\lambda\delta})^{-1}-1 = (e^{\lambda\delta}-1)^{-1}$, implying that the maximum likelihood estimator based on the Zs converges to $(e^{\lambda\delta}-1)/\delta = \lambda + \delta\lambda^2/2 + \delta^2\lambda^3/6 + \cdots$; i.e., λ will be overestimated by $O(\delta)$.

Solution 3

(a) The mean and variance of $\overline{Y}$ are respectively $\theta/2$ and $\theta^2/(12n)$, so $2\overline{Y}$ is an unbiased estimator of θ. If $\stackrel{\mathrm{D}}{=}$ denotes equality in distribution then we can write $Y_j \stackrel{\mathrm{D}}{=} \theta U_j$, where $U_j \sim U(0,1)$ are independent for each j, and therefore

$$Q' = \overline{Y}/\theta \stackrel{\mathrm{D}}{=} \theta\overline{U}/\theta = \overline{U}$$

is a function of the data and of θ with a distribution that is (in principle) known. Thus Q is a pivot. Its distribution could be estimated to arbitrary accuracy by simulation of $U_1,\ldots,U_n$ for any fixed n or we can note that it is symmetric about 1/2 with variance $1/(12n)$, so a central limit theorem gives

$$\sqrt{12n}(Q-1/2) \,\dot\sim\, \mathcal{N}(0,1).$$

This approximation will be good enough for most practical purposes when $n\geq 12$: taking $Z = \sum_{j=1}^{12} U_j - 1/2 \,\dot\sim\, \mathcal{N}(0,1)$ is an old fudge to get standard normal variables from $U(0,1)$ ones.

(b) We have

$$\mathrm{P}(z_{\alpha/2} \leq \sqrt{12n}(Q-1/2) \leq z_{1-\alpha/2}) = 1-\alpha,$$

and replacing Q by $\overline{Y}/\theta$ and inverting the pivot yields an interval with limits

$$L = \frac{\overline{Y}}{\frac{1}{2} + z_{1-\alpha/2}/(12n)^{1/2}}, \quad U = \frac{\overline{Y}}{\frac{1}{2} + z_{\alpha/2}/(12n)^{1/2}}.$$

Computing this with $n=16$ and $\overline{y} = 320869$ gives the interval $(500225.9, 894902.8) \doteq (500226, 894903)$.

(c) The interval computed using the maximum is $(524136.7, 659001.1) \doteq (524137, 659001)$. This is clearly better than that in (b), because (i) it is much shorter but both are 95% intervals, (ii) it lies wholly to the right of the observed maximum and therefore does not include values we already know to be impossible, and (iii) it is exact for any n. (The last reason is less important, because as mentioned above, the normal approximation to the distribution of $\overline{Y}$ is excellent for $n=16$.)

Solution 4

(a) The negative log likelihood function is

$$\begin{aligned}
-\ell(\psi,\lambda) &\equiv \frac{1}{2}\sum_{j=1}^{n}\begin{pmatrix} y_j-\psi & x_j-\lambda\end{pmatrix}\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix}^{-1}\begin{pmatrix} y_j-\psi\\ x_j-\lambda\end{pmatrix} + \frac{1}{2\sigma_2^2}\sum_{j=n+1}^{n+m}(x_j-\lambda)^2,\\
&= \frac{1}{1-\rho^2}\sum_{j=1}^{n}\left\{\frac{(y_j-\psi)^2}{2\sigma_1^2} - \frac{\rho(y_j-\psi)(x_j-\lambda)}{\sigma_1\sigma_2} + \frac{(x_j-\lambda)^2}{2\sigma_2^2}\right\} + \sum_{j=n+1}^{n+m}\frac{(x_j-\lambda)^2}{2\sigma_2^2}, \quad \psi,\lambda\in\mathbb{R},
\end{aligned}$$

and its first derivatives are

$$
\begin{aligned}
-\ell_\psi(\psi,\lambda) &= \frac{1}{1-\rho^2}\sum_{j=1}^n\left\{\frac{\psi-y_j}{\sigma_1^2}+\frac{\rho(x_j-\lambda)}{\sigma_1\sigma_2}\right\} \propto \psi-\overline{y}_n+\sigma_1\rho(\overline{x}_n-\lambda)/\sigma_2,\\
-\ell_\lambda(\psi,\lambda) &= \frac{1}{1-\rho^2}\sum_{j=1}^n\left\{\frac{\rho(y_j-\psi)}{\sigma_1\sigma_2}+\frac{\lambda-x_j}{\sigma_2^2}\right\}+\sum_{j=n+1}^{n+m}\frac{\lambda-x_j}{\sigma_2^2}.
\end{aligned}
$$

The first of these implies that if $\ell_\psi(\widehat{\psi},\widehat{\lambda})=0$ then $\widehat{\psi}=\overline{y}_n-\sigma_1\rho(\overline{x}_n-\widehat{\lambda})/\sigma_2$. If we substitute this expression into the second equation, setting $\ell_\lambda(\widehat{\psi},\widehat{\lambda})=0$, and simplify, we obtain $\widehat{\lambda}=(n+m)^{-1}\sum_{j=1}^{n+m}x_j$ after a little struggle, which gives the desired formula for $\widehat{\psi}$ on replacing the observations with the corresponding random variables.

(b) Linearity of the expectation operator implies that $\mathrm{E}(\overline{Y}_n)=\mathrm{E}(Y)=\psi$, $\mathrm{E}(\overline{X}_n)=\mathrm{E}(\overline{X}_{n+m})=\mathrm{E}(X)=\lambda$ and thus $\mathrm{E}(\widehat{\psi})=\mathrm{E}(\overline{Y}_n)+\rho\sigma_1\mathrm{E}(\overline{X}_{n+m}-\overline{X}_n)/\sigma_2=\psi+\rho\sigma_1(\lambda-\lambda)/\sigma_2=\psi$. Clearly $\mathrm{var}(\overline{Y}_n)=\sigma_1^2/n$.

The Fisher information matrix is

$$
\imath(\psi,\lambda)=\frac{1}{1-\rho^2}\begin{pmatrix}\frac{n}{\sigma_1^2} & -\frac{n\rho}{\sigma_1\sigma_2}\\ -\frac{n\rho}{\sigma_1\sigma_2} & \frac{n}{\sigma_2^2}+\frac{m(1-\rho^2)}{\sigma_2^2}\end{pmatrix},
$$

and the (ψ,ψ) corner of $\imath(\psi,\lambda)^{-1}$ is $\sigma_1^2\{1-m\rho^2/(n+m)\}/n$, which is the asymptotic variance of $\widehat{\psi}$; in fact here this variance is exact, as a direct computation of $\mathrm{var}(\widehat{\psi})$ shows. Hence the relative efficiency is

$$
\frac{\mathrm{var}(\overline{Y}_n)}{\mathrm{var}(\widehat{\psi})}=\frac{1}{1-m\rho^2/(n+m)}.
$$

(i) When $\rho=0$, the relative efficiency is one, because X is then independent of Y and the auxiliary data give no information about ψ.

(ii) When $m\to\infty$, λ is known exactly from $\overline{X}_{n+m}$, so $\widehat{\psi}=\overline{Y}_n+\rho\sigma_1(\lambda-\overline{X}_n)/\sigma_2$ and the relative efficiency becomes $1/(1-\rho^2)$.

(iii) When $\rho\to\pm1$ the auxiliary observations $X_{n+1},\ldots,X_{n+m}$ are perfectly informative about ψ, so the sample size is effectively $n+m$, and the relative efficiency is $(n+m)/n$.