Solution 1

(a) In this case θ is the population variance and the standard calculation (check it if unsure)

$$\mathrm{E}\left\{\sum_{j=1}^n (Y_j-\overline{Y})^2\right\} = \mathrm{E}\left[\sum_{j=1}^n \{(Y_j-\mu)-(\overline{Y}-\mu)\}^2\right] = \mathrm{E}\left\{\sum_{j=1}^n (Y_j-\mu)^2\right\} - n\mathrm{E}\left\{(\overline{Y}-\mu)^2\right\} = n\theta - n\theta/n$$

where $\mu = \mathrm{E}(Y_j)$ gives

$$\mathrm{E}(T) = \mathrm{E}\left\{n^{-1}\sum (Y_j-\overline{Y})^2\right\} = \frac{(n-1)}{n}\theta = \theta - \theta/n,$$

so $\gamma = -1/n$. Note that the only assumptions about the Y_j are they are independent with mean μ and variance θ.

Now $T^* = n^{-1}\sum (Y_j^* - \overline{Y^*})^2$, where the $Y_j^* \overset{\text{iid}}{\sim} \{y_1,\ldots,y_n\}$ with probabilities $1/n$, and this distribution has mean $\overline{y}$ and variance

$$\frac{1}{n}\sum y_j^2 - \overline{y}^2 = \frac{1}{n}\sum (y_j-\overline{y})^2 = t.$$

We can apply the computation above to see that T^* is also downwardly-biased as an estimate of its population variance t, with mean

$$\mathrm{E}^*(T^*) = \frac{(n-1)}{n}t,$$

so $C = \mathrm{E}^*(T^*)/t - 1 = -1/n = \gamma$, as stated.

(b) The estimator of C would be $C^* = R^{-1}\sum_{r=1}^R T_r^*/t - 1$, which has variance $R^{-1}\mathrm{var}^*(T^*)/t^2$, because each of the T_r^* has variance $\mathrm{var}^*(T^*)$ and they are independent.

Solution 2

(a) By definition, the median of $Y_1^*,\ldots,Y_n^*$ when $n = 2m+1$ is $Y_{(m+1)}^*$. Hence

$$T^* > y_{(l)} \quad\Longleftrightarrow\quad Y_{(m+1)}^* > y_{(l)} \quad\Longleftrightarrow\quad Y_{(n)}^*,\ldots,Y_{(m+1)}^* > y_{(l)},$$

which is true if and only at most m of the Y^* are less than or equal to $y_{(l)}$. The probability that a single Y^* is less than or equal to $y_{(l)}$ is $p = l/n$, and as the Y^* are independent, this gives the stated binomial probability, because if we let I_j be the indicator of the event $Y_j^* \le y_{(l)}$, then

$$\mathrm{P}^*(T^* > y_{(l)}) = \mathrm{P}^*\left(\sum_{j=1}^n I_j \le m\right) = \sum_{j=0}^m \binom{n}{j} p^j (1-p)^{n-j},$$

as required.

(b) It is easy to check that $\mathrm{P}^*(T^* > y_{(8)}) \doteq 1 - \mathrm{P}^*(T^* > y_{(3)}) \doteq 0.05$, so $\mathrm{P}^*(y_{(4)} \le T^* \le y_{(8)}) \doteq 0.9$.

For the bootstrap confidence interval, note that

$$0.9 \doteq \mathrm{P}^*(y_{(4)} \le T^* \le y_{(8)}) = \mathrm{P}^*(y_{(4)} - t \le T^* - t \le y_{(8)} - t),$$

where t is the observed median $y_{(6)}$, so the basic bootstrap argument gives (approximate) 90% confidence interval $(2y_{(6)} - y_{(8)}, 2y_{(6)} - y_{(4)})$.

Solution 3

(a) We have

$$\begin{aligned} 0 &= \int a\left\{x; t(G_\varepsilon)\right\} \, \mathrm{d}G_\varepsilon(x) \\ &= (1-\varepsilon)\int a\left\{x; t(G_\varepsilon)\right\} \, \mathrm{d}G(x) + \varepsilon \int a\left\{x; t(G_\varepsilon)\right\} \, \mathrm{d}H_y(x) \\ &= (1-\varepsilon)\int a\left\{x; t(G_\varepsilon)\right\} \, \mathrm{d}G(x) + \varepsilon a\left\{y; t(G_\varepsilon)\right\}, \end{aligned}$$

and differentiation using the chain rule gives

$$0 = a\left\{y; t(G_\varepsilon)\right\} - \int a\left\{x; t(G_\varepsilon)\right\} \, \mathrm{d}G(x) + \varepsilon a_\theta\left\{y; t(G_\varepsilon)\right\} \frac{\partial t(G_\varepsilon)}{\partial \varepsilon} + (1-\varepsilon)\int a_\theta\left\{x; t(G_\varepsilon)\right\} \frac{\partial t(G_\varepsilon)}{\partial \varepsilon} \, \mathrm{d}G(x),$$

which reduces to

$$0 = a\left\{y; t(G)\right\} + \int a_\theta\left\{x; t(G)\right\} \, \mathrm{d}G(x) \left.\frac{\partial t(G)}{\partial \varepsilon}\right|_{\varepsilon=0}$$

on setting $\varepsilon = 0$. This yields the specified formula for the influence function, even if $a(y;\theta)$ is a $d \times 1$ vector.

(b) In the case of a random sample $y_1, \ldots, y_n$, the EDF $\widehat{G}$ puts masses $1/n$ on each of the y_j and $\widehat{\theta} = t(\widehat{G})$, so the empirical influence function reduces to the given formula.

(c) The formula $a(x;\theta) = x - \theta$ leads to $\theta = \int x \, \mathrm{d}G(x)$, i.e., θ is the population mean, and we saw in the lectures that $l_j = y_j - \overline{y}$.

To apply the formulation here, note that $\int a(x;\theta) \, \mathrm{d}\widehat{G}(x) = 0$ implies that $\widehat{\theta} = \overline{y}$, and $a_\theta(x;\theta) = -1$, so we again find

$$l_j = L_t(y_j; \widehat{G}) = \frac{a(y_j; \widehat{\theta})}{-n^{-1}\sum_{k=1}^n \partial a(y_k; \widehat{\theta})/\partial\theta} = \frac{y_j - \overline{y}}{-n^{-1}\sum_{j=1}^n (-1)} = y_j - \overline{y}.$$

(d) In the case of a maximum likelihood estimator of the vector θ we have

$$a(x;\theta) = \frac{\partial \log f(x;\theta)}{\partial \theta}, \quad a_\theta(x;\theta) = \frac{\partial^2 \log f(x;\theta)}{\partial\theta\partial\theta^{\mathrm{T}}},$$

corresponding to the score and (minus) the observation information contributions. Hence

$$l_j = L_t(y_j; \widehat{G}) = \left\{-n^{-1}\sum_{k=1}^n a_\theta(y_k; \widehat{\theta})\right\}^{-1} a(y_j; \widehat{\theta}) = \left(n^{-1}\widehat{\jmath}\right)^{-1} S_j,$$

as required, and therefore the covariance matrix of $\widehat{\theta}$ is

$$v_L = \frac{1}{n^2}\sum_{j=1}^n (n\widehat{\jmath}^{-1}S_j)(n\widehat{\jmath}^{-1}S_j)^{\mathrm{T}} = \jmath^{-1}\left(\sum_{j=1}^n S_j S_j^{\mathrm{T}}\right)\jmath^{-1}.$$

(e) In this case $\log f(y;\theta) = -\log\theta - y/\theta$, so $a(y;\theta) = -1/\theta + y/\theta^2$ and $a_\theta(y;\theta) = 1/\theta^2 - 2y/\theta^3$. One can easily check that $\widehat{\theta} = \overline{y}$ and $\widehat{\jmath} = n\widehat{\theta}^2$, so the sandwich variance is

$$(n\widehat{\theta}^2)^{-1} \times \sum_{j=1}^n (-1/\widehat{\theta} + y_j/\widehat{\theta}^2)^2 \times (n\widehat{\theta}^2)^{-1} = \frac{1}{n^2}\sum_{j=1}^n (y_j - \widehat{\theta})^2 = \frac{1}{n^2}\sum_{j=1}^n (y_j - \overline{y})^2$$

Now if $n \to \infty$ under the exponential model, then $n^{-1}\sum_{j=1}^n (y_j - \overline{y})^2 \to \theta^2$, so this expression is roughly θ^2/n for large n, and this is also true for $\widehat{\jmath}^{-1}$. But if the exponential model is not true, then the sandwich is valid anyway, because the given formula is an (almost unbiased) estimator for any distribution with a finite variance. The downside to using the sandwich estimator when the exponential model is correct is that it is less efficient, in the sense that the resulting confidence intervals will have worse properties (they are longer and more variable).