Problem 1 Suppose that estimator T has expectation equal to $\theta(1+\gamma)$, so that the bias is $\theta\gamma$. The bias factor γ can be estimated by $C = \mathrm{E}^*(T^*)/T - 1$, where E^* denotes expectation over bootstrap sampling and T^* is based on the bootstrap samples.

(a) Show that in the case of the variance estimate $T = n^{-1}\sum(Y_j - \overline{Y})^2$, C is exactly equal to γ.

(b) If C were approximated from R resamples by C^*, what would be the simulation variance of C^*?

Problem 2

Let T be the median of a random sample of size $n = 2m+1$ with ordered values $y_{(1)} < \cdots < y_{(n)}$; the observed value of T is therefore $t = y_{(m+1)}$.

(a) Show that $T^* > y_{(l)}$ if and only if fewer than $m+1$ of the Y_j^* are less than or equal to $y_{(l)}$, and deduce that

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

This specifies the exact resampling distribution of the sample median, and can be used to prove that the bootstrap estimate of $\mathrm{var}(T)$ is consistent as $n \to \infty$.

(b) Use the resampling distribution in (a) to show that for $n = 11$,

$$\mathrm{P}^*(T^* \le y_{(3)}) = \mathrm{P}^*(T^* \ge y_{(9)}) = 0.051,$$

and deduce that an approximate basic bootstrap 90% confidence interval for the population median is $(2y_{(6)} - y_{(8)}, 2y_{(6)} - y_{(4)})$.

Problem 3 A parameter $\theta = t(G)$ is determined implicitly through the estimating equation

$$\int a(x;\theta)\,\mathrm{d}G(x) = \int a\{x; t(G)\}\,\mathrm{d}G(x) = 0.$$

(a) Replace G by $G_\varepsilon = (1-\varepsilon)G + \varepsilon H_y$, where H_y is the Heaviside function putting unit mass at y, and show by differentiation with respect to ε that the influence function for $t(\cdot)$ is

$$L_t(y;G) = \left.\frac{\partial t(G_\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = \left\{-\int a_\theta(x;\theta)\,\mathrm{d}G(x)\right\}^{-1} a(y;\theta), \quad \text{where} \quad a_\theta(x;\theta) = \frac{\partial a(x;\theta)}{\partial \theta^{\mathrm{T}}}.$$

(b) Hence show that with a random sample $y_1, \ldots, y_n$ and $\widehat{\theta} = t(\widehat{G})$ the jth empirical influence value is

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

(c) Check that this gives the result anticipated when $a(x;\theta) = x - \theta$.

(d) Let $\widehat{\theta}$ be the maximum likelihood estimator of the parameter of a regular parametric model $f(y;\theta)$ based on a random sample $y_1, \ldots, y_n$. Show that the jth empirical influence value for $\widehat{\theta}$ at y_j may be written as $n\widehat{\jmath}^{-1} S_j$, where

$$\widehat{\jmath} = -\sum_{j=1}^{n} \frac{\partial^2 \log f(y_j; \widehat{\theta})}{\partial \theta \partial \theta^{\mathrm{T}}}, \quad S_j = \frac{\partial \log f(y_j; \widehat{\theta})}{\partial \theta},$$

and deduce that the nonparametric delta method variance estimate for $\widehat{\theta}$ has the sandwich form

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

(e) When $y_1, \ldots, y_n$ is a random sample from the exponential distribution with mean θ compare the result in (b) to the usual parametric approximation.