STA 243 Spring 2017 **Assignment 6** Due date: June 9, 2017 (5pm)

For this assignment you can do it in a group of maximum two people. Objectives of this assignment:

- To gain experience in applying the bootstrap methodology.
- Last chance for Thomas to torture you.

Altogether there are 3+1 questions.

1. Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$. The MLE is

$$\hat{\theta} = X_{\max} = \max(X_1, \ldots, X_n).$$

   (a) Find the distribution of $\hat{\theta}$ (in terms of θ and n).

   (b) Derive the analytic expression for the variance of $\hat{\theta}$. Call it $\text{Var}_{F_\theta}(\hat{\theta})$.

   (c) Generate a data set of size $n = 50$ and $\theta = 3$. Then generate $B = 5000$ bootstrap samples using parametric bootstrap. Use the bootstrap samples to approximate $\text{Var}_{F_\theta}(\hat{\theta})$. Compare your answer to (b).

   (d) With the same data set: repeat (c) with nonparametric bootstrap.

   (e) With the same data set: plot the histograms of $\hat{\theta}^*$ obtained from the parametric and nonparametric bootstraps.

   (f) Compare the true distribution of $\hat{\theta}$ to those histograms obtained in (e).

   Note: this is an example where the nonparametric bootstrap fails. Can you guess why?

2. Generate two regression data sets (x_i, y_i)'s with $n = 512$, one from the test function in Assignment 2 and the other from

$$f(x) = (4x - 2) + 2\exp\{-16(4x - 2)^2\},$$

   where $x_i = (i - 1)/n$. Set the noise variance as $\sigma^2 = (\|f\|/5)^2$.

   (a) Obtain the regression curve estimates for both test functions using the genetic algorithm you implemented for Assignment 2.

   (b) Construct 95% pointwise confidence bands for both curve estimates using the bootstrap. Use both "bootstrapping residuals" and "bootstrapping pairs" approaches. For the test function from Assignment 2, comment on the shape of the confidence bands near jump points.

   (c) Describe how you would obtain a confidence interval for the location of a jump point using the bootstrap. You do not need to do any programming for this part.

3. Let $X_1, \ldots X_{25} \sim t_3$. Let $\theta = (q_{0.75} - q_{0.25})/1.34$, where q_p denotes the p^{th} quantile. Do a simulation to compare the coverages and lengths of the following confidence intervals for θ:

   (a) normal theory interval with standard error from the jackknife,

   (b) normal theory interval with standard error from the bootstrap, and

   (c) bootstrap percentile interval.

   Note: the jackknife does not give a consistent estimator for the variance of a quantile.

4. (Note: This question is optional, and will not be graded.) In this exercise we will follow the necessary steps for deriving the bias corrected and accelerated (BC_a) percentile confidence interval. Let $X_1, \ldots, X_n \sim F$, where F is a distribution parametrized by $\theta \in \mathbb{R}$. Let $\hat{\theta}_n = \theta(\hat{F}_n)$ be an estimator for θ, where $\hat{F}_n$ is the empirical distribution function. Now suppose there exists a monotonic transformation $g(\cdot)$ and constants a and b such that $\sigma_{g(\theta)} = 1 + ag(\theta)$ and

$$P_F\left(\frac{g(\hat{\theta}_n) - g(\theta)}{\sigma_{g(\theta)}} + b \leq x\right) = G(x),$$

where $G(x)$ is a known continuous distribution function.

(a) Assume that a and b are known. Show that a $100\%(1-2\alpha)$ confidence interval for θ is given by

$$[L_n,\ U_n] \overset{\text{def}}{=} \left[g^{-1}\left(\frac{g(\hat{\theta}_n) - [G^{-1}(1-\alpha) - b]}{1 + a[G^{-1}(1-\alpha) - b]}\right), g^{-1}\left(\frac{g(\hat{\theta}_n) - [G^{-1}(\alpha) - b]}{1 + a[G^{-1}(\alpha) - b]}\right)\right].$$

(b) Let $\hat{\theta}_n^*$ be the bootstrap version of $\hat{\theta}_n$ and $\sigma_{g(\hat{\theta}_n)} = 1 + ag(\hat{\theta}_n)$. The bootstrap argument asserts that $\frac{g(\hat{\theta}_n) - g(\theta)}{\sigma_{g(\theta)}}$ and $\frac{g(\hat{\theta}_n^*) - g(\hat{\theta}_n)}{\sigma_{g(\hat{\theta}_n)}} | \hat{F}_n$ have approximately the same distribution.

Denote now $H(x) = P_{\hat{F}_n}(\hat{\theta}_n^* \leq x)$, the distribution function of the bootstrap sample of $\hat{\theta}_n^*$. Show that

$$H(L_n) \approx G\left(b - \frac{G^{-1}(1-\alpha) - b}{1 + a\left[G^{-1}(1-\alpha) - b\right]}\right), \quad H(U_n) \approx G\left(b - \frac{G^{-1}(\alpha) - b}{1 + a\left[G^{-1}(\alpha) - b\right]}\right).$$

(c) Now assume that $G(x)$ is the standard normal distribution function. Show that L_n and U_n can be approximated by

$$\left[H^{-1}(\alpha_1),\ H^{-1}(\alpha_2)\right] = \left[\hat{\theta}_n^{*(\alpha_1)},\ \hat{\theta}_n^{*(\alpha_2)}\right],$$

where $\alpha_1 = \Phi\left(b + \frac{b + z^{(\alpha)}}{1 - a(b + z^{(\alpha)})}\right)$ and $\alpha_2 = \Phi\left(b + \frac{b + z^{(1-\alpha)}}{1 - a(b + z^{(1-\alpha)})}\right)$.

(d) Give an estimate of b by showing that $H(\hat{\theta}_n) \approx G(b)$.

(e) (This part is optional. That is, you can still get a perfect score for this assignment without completing this part. However, you will not get any hints or help from your instructor or TA.) If a and b are small, it can be shown that (by Taylor's expansion)

$$\alpha_1 \approx \alpha + [2b + a(z^{(\alpha)})^2]\phi(z^{(\alpha)}),\ \alpha_2 \approx 1 - \alpha + [2b + a(z^{(1-\alpha)})^2]\phi(z^{(1-\alpha)}),$$

where ϕ is the standard normal density. On the other hand, if g is asymptotically linear, the best α_1 and α_2 (due to Edgeworth and Cornell-Fisher expansions) are

$$\alpha_1 = \alpha + \left[\frac{1}{3}\gamma + \frac{1}{6}\gamma(z^{(\alpha)})^2\right]\phi(z^{(\alpha)}),\ \alpha_2 = 1 - \alpha + \left[\frac{1}{3}\gamma + \frac{1}{6}\gamma(z^{(1-\alpha)})^2\right]\phi(z^{(1-\alpha)}),$$

where γ is the skewness of $\hat{\theta}$. By comparing the expressions of α_1 or α_2, we obtain $a = \gamma$. Therefore, a can be estimated by estimating the skewness of $\hat{\theta}$.

Note: The actual argument of approximating a is technical and unintuitive. You can treat it as a way to better approximate $\hat{\theta}_n$ when the distribution of $\hat{\theta}_n$ is not symmetric. This argument also reveals that the optimal a and b are the same.

—— End of Assignment 6 ——