Large Sample Theory

Homework 1: Bootstrap Method, CLT

Due Date: October 3rd, 2004

1. Suppose that someone collects a random sample of size 4 of a particular measurement. The observed values are $\{2, 4, 9, 12\}$.
(a) Find the bootstrap mean and variance of the above sample.
(b) Find the relationship between sample mean and bootstrap mean.
(c) Find the relationship between sample variance and bootstrap variance.
(d) Consider a general problem in which we have a random sample $\{x_1, \ldots, x_n\}$. Establish the relationships you found in (b) and (c) under this setting.

2. Let $X_1, \ldots, X_n$ be i.i.d. random variables and $\hat{\theta} = \bar{X}^2$.
(a) Derive the asymptotic distribution of $\sqrt{n}[\hat{\theta} - E^2(X)]$.
(b) Show that the bootstrap variance estimator based on i.i.d. X_i^*'s from F_n is equal to
$$\bar{V}_B = \frac{4\bar{X}^2\hat{c}_2}{n} + \frac{4\bar{X}\hat{c}_3}{n^2} + \frac{\hat{c}_4}{n^3},$$
where $\hat{c}_j$'s are the sample central moments defined by $n^{-1}\sum_{i=1}^n (X_i - \bar{X})^j$.
(c) Is $\bar{V}_B$ a consistent estimate of the asymptotic variance derived in (a)?
Hint: For (a), you can use Slutsky's theorem.

3. Let $X_1, \ldots, X_n$ be i.i.d. from a pdf $\sigma^{-1}f((x-\mu)/\sigma)$ on R, where f is known. Let
$$H_n(t) = P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{S} \leq t\right)$$
and
$$\hat{H}_B(t) = P\left(\frac{\sqrt{n}(\bar{X}^* - \bar{X})}{S^*} \leq t \,\middle|\, X_1, \ldots, X_n\right)$$
be the bootstrap estimator of H_n, where S^2 is the sample variance, X_i^*'s are i.i.d. from $s^{-1}f((x-\bar{x})/s)$, given $\bar{X} = \bar{x}$ and $S = s$, and S^* is the bootstrap analogue of S. Show that $H_n(t) = \hat{H}_B(t)$.

4. Suppose $Y_i = \alpha + \beta x_i + \epsilon_i$ for $i = 1, 2, \ldots$, where the x_i are known numbers not all equal and the ϵ_j are independent random variables with mean 0 and common variances σ^2.
(a) Show that the least-squares estimate, $\hat{\beta}_n$, of β was consistent provided $\sum_{i=1}^n (x_i - \bar{x}_n)^2 \to \infty$.
(b) Show that $\hat{\beta}_n$ is asymptotically normal when the ϵ_j are identically distributed and
$$\max_{1 \leq j \leq n} \frac{(x_j - \bar{x}_n)^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \to 0 \quad \text{as } n \to \infty.$$

5. Suppose we would like to estimate mean (θ) of a population and a random sample $\{x_1, \ldots, x_n\}$ is being taken. Consider $\hat{\theta} = \bar{X}$.
(a) Find the jackknife estimate of θ.
(b) Find the jackknife variance estimate of θ.

6. Let $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_n})^{\mathbf{T}}$ be an $n \times 1$ vector of random variables and let $\mathbf{A}$ be an $n \times n$ symmetric matrix. If $E(\mathbf{X}) = \theta$ and $Var(\mathbf{X}) = \mathbf{\Sigma} = (\sigma_{\mathbf{ij}})_{\mathbf{n \times n}}$, show that

$$E[\mathbf{X^T A X}] = \mathbf{tr}[\mathbf{A\Sigma}] + \theta^{\mathbf{T}}\mathbf{A}\theta$$

and

$$Var[\mathbf{X^T A X}] = (\mu_{\mathbf{4}} - \mathbf{3}\mu_{\mathbf{2}}^{\mathbf{2}})\mathbf{a^T a} + \mathbf{2}\mu_{\mathbf{2}}^{\mathbf{2}}\mathbf{tr}(\mathbf{A^2}) + \mathbf{4}\mu_{\mathbf{2}}\theta^{\mathbf{T}}\mathbf{A^2}\theta + \mathbf{4}\mu_{\mathbf{3}}\theta^{\mathbf{T}}\mathbf{Aa},$$

where $\mu_r = E[(X_i - \theta_i)^r]$ and $\mathbf{a}$ is the column vector of the diagonal elements of $\mathbf{A}$.

7. Let X_i be iid and real-valued for $i = 1, \ldots, n$; 4th moments exist. The sample variance is s^2.
(a) Find the mean and variance of s^2.
(b) How would you estimate $Var(s^2)$ from the sample by the method of moments?
(c) How would you estimate $Var(s^2)$ from the sample by the jackknife?
(d) How would you estimate $Var(s^2)$ if the parent distribution was normal? How does the normal-theory estimator behave if the parent distribution is not normal?
Hint: You can solve (a) using the results derived in last question.

8. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. Let $V = \sum_{j=1}^{n}(X_j - \bar{X})^2$.
(a) For what constants $c_1(n)$ depending on n is $c_1(n)V$ an unbiased estimator of σ^2?
(b) For what constants $c_2(n)$ depending on n is the mean-square error $E_{\sigma^2}[(c_1(n)V - \sigma^2)^2]$ minimized for all $\sigma > 0$?

9. Find the jackknife estimate of the variance of the median. Is this estimate a good one? You can use either a simulation or an analytic method to address this question by assuming that the unknown population distribution is the uniform distribution on the interval $[0, 1]$.

10. A random variable $\hat{\theta}_n$, based on a random sample of size $n > 1$, is said to be an *unbiased estimator* of a parameter θ if $E(\hat{\theta}_n) = \theta$ for all $\theta \in \Theta$, the parameter space. Suppose that $\hat{\theta}_n$ is unbiased for θ and that $V(\hat{\theta}_n) = O(n^{-1})$.
(a) If $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, show that no unbiased estimator of e^{μ} can be found based on the sample mean $\bar{X}$.
(c) Show that $E(\exp(\bar{X})) = \exp(\mu + \sigma^2/2n)$ and $E(e^{S^2/2n}) = \exp(\sigma^2/2n) + o(n^{-1})$, where $S^2 = \sum_{i=1}^{n}(X_i - \mu)^2/n$. Hence find an estimator of $\exp(\mu)$ that is unbiased apart from terms of smaller order than n^{-1}.