Math 6070-1: Spring 2013
Problem set 5

Due date: April 24, 2013

1. (Normal-data bootstrapping) Simulate 1,000 i.i.d. samples from N$(0\,,1)$, and let us call this sample $X_1,\ldots,X_{1000}$.

   (a) Take N bootstrap subsamples in order to estimate

   $$\theta = \mathrm{P}\left\{\cosh(\bar{X}) \le 3\right\},$$

   for $N = 1,000$, $5,000$, and $10,000$.

   (b) Use $50,000$ simulations [each of size 1,000, of couse!] in order to find an accurate evaluation of θ.

   (c) Discuss how well your bootstrap method is working.

2. (Introduction to time series) Download the atmospheric CO_2 concentration data by Uhse, Schmidt, and Levin from

   `http://cdiac.ornl.gov/ftp/trends/co2/westerland.co2`.

   (a) This data is a time series of the form $(t\,,x_t)$, where $t = 1,2,\ldots$. Plot the time series [that is, plot t versus x_t for all times t].

   (b) Define a new data set, starting with $y_1 := x_1$, and then

   $$y_t := x_t - x_{t-1} \qquad \text{for all } t \ge 2.$$

   Plot t versus y_t. Justify the phrase, "we have obtained the y_t's by de-trending the x_t's."

   (c) In order to understand more deeply what you have done, posit a statistical model of the type

   $$x_t = \beta_0 + \beta_1 t + w_t \qquad \text{for all } t \ge 1,$$

   where w_t is an i.i.d. sequence of random variables with mean zero and unknown variance σ^2; this is called a *linear regression model*. The constants β_0 and β_1 are unknown statistical parameters.

      i. Compute y_t explicitly in terms of w_t.

ii. Are the w_t's independent? Perform a test, and describe both your methods, and the outcome.

iii. Describe an unbiased estimator for β_0, using only the y_t's. Prove that your estimator is unbiased. You may not use, without proof, any facts from regression analysis.

(d) Let $X_1, \ldots, X_n$ be i.i.d. $\mathrm{N}(\theta\,, 1)$, where θ is the unknown parameter. Define T_n to be the *Hodges estimator* of θ; that is,

$$T_n := \begin{cases} \frac{1}{2}\bar{X} & \text{if } |\bar{X}| \leq n^{-1/4}, \\ \bar{X} & \text{if } |\bar{X}| > n^{-1/4}. \end{cases}$$

i. Is T_n biased? Analyze carefully.

ii. Prove that T_n is consistent for θ by proving the even more general statement that

$$\sqrt{n}\,(T_n - \theta) \xrightarrow{d} \mathrm{N}(0\,, \sigma^2(\theta)) \qquad \text{as } n \to \infty,$$

where

$$\sigma^2(\theta) := \begin{cases} 1/4 & \text{if } \theta = 0, \\ 1 & \text{if } \theta \neq 0. \end{cases}$$

iii. Let F_n^θ denote the cdf of $\sqrt{n}(T_n - \theta)$. That is,

$$F_n^\theta(x) := \mathrm{P}\left\{\sqrt{n}\,(T_n - \theta) \leq x\right\}.$$

The previous exercise can be restated as follows: $F_n^\theta(x) \to \Phi(x/\sigma(\theta))$ for all $x \in \mathbf{R}$ as $n \to \infty$, where Φ denotes the standard normal cdf. Prove the stronger assertion that

$$\max_{x \in \mathbf{R}} |F_n^\theta(x) - \Phi(x/\sigma(\theta))| \to 0 \qquad \text{as } n \to \infty.$$

iv. (Hard/extra credit) The parametric bootstrap estimator of $F_n^\theta(x)$ is, of course, $F_n^{\bar{X}}(x)$. Prove that

$$\max_{x \in \mathbf{R}} \left|F_n^{\bar{X}}(x) - F_n^\theta(x)\right| \xrightarrow{\mathrm{P}} 0 \qquad \text{as } n \to \infty,$$

if and only if $\theta = 0$. In other words, the bootstrap fails generically in this case, unless θ happens to be zero.