Homework Assignment 4

Due via Canvas, Apr 21 by midnight

SDS 384-11 Theoretical Statistics

1. Let $\mathcal{P}$ be the set of all distributions on the real line with finite first moment. Show that there does not exist a function $f(x)$ such that $Ef(X) = \mu^2$ for all $P \in \mathcal{P}$ where μ is the mean of P, and X is a random variable with distribution P.

2. A continuous distribution with CDF $F(x)$, on the real line is symmetric about the origin if, and only if, $1 - F(x) = F(-x)$ for all real x. This suggests using the parameter,

$$\theta(F) = \int (1 - F(x) - F(-x))^2 dF(x) \tag{1}$$

$$= \int ((1 - F(-x))^2 dF(x) - 2\int (1 - F(-x))F(x)dF(x) + \int F(x)^2 dF(x) \tag{2}$$

as a nonparametric measure of how asymmetric the distribution is. Find a kernel h, of degree 3, such that $E_F h(X_1, X_2, X_3) = \theta(F)$ for all continuous F. Find the corresponding U statistic.

3. Look at the seminar paper "Probability Inequalities for Sums of Bounded Random Variables" by Wassily Hoeffding. It should be available via `lib.utexas.edu`. You can assume that n is a multiple of m (the degree of the kernel). Assume that the kernel is bounded, i.e. $|h(X_1, \ldots, X_m) - \theta| \leq b$, where $\theta = E[h(X_1, \ldots, X_m)]$.

   (a) Read and reproduce the proof of equation 5.7 for large sample deviation of order m U statistics.

   (b) Also prove Bernstein's inequality (see below) for U statistics. This is buried in the paper, you will have to find the bits and pieces and put them together. The Bernstein inequality is given by:

$$P(|U_n - \theta| \geq \epsilon) \leq a \exp\left(-\frac{n\epsilon^2/m}{c_1\sigma^2 + c_2\epsilon}\right),$$

   where $\sigma^2 = \text{var}(h(X_1, \ldots, X_m))$ and c_1, c_2 are universal constants.

4. (VC dimension) Compute the VC dimension of the following function classes

   (a) Circles in $\mathbb{R}^2$

   (b) Axis aligned rectangles in $\mathbb{R}^2$

   (c) Axis aligned squares in $\mathbb{R}^2$