Math 6070-1, Spring 2014; Assignment #6

Due: Wednesday April 2, 2014

1. Complete reading Sections 1 and 2 of the module on empirical processes at `http://www.math.utah.edu/~davar/math6070/2014/Kolmogorov-Smirnov.pdf`. Also, read Section 3.1.

2. Let $X_1, \ldots, X_n$ be a random sample from a population with a continuous strictly increasing CDF F, and define $\hat{F}_n$ to be the empirical CDF. Suppose F is continuously differentiable with derivative f, which is of course the PDF of the X_j's. Define

$$\bar{D}_n := \int_{-\infty}^{\infty} \left[\hat{F}_n(x) - F(x)\right]^2 f(x)\,\mathrm{d}x$$

   (a) Prove that $\bar{D}_n$ is finite, as well as distribution free.

   (b) In the case that the X_i's are Unif$(0\,,1)$, express $\bar{D}_n$ explicitly in terms of the order statistics $X_{1:n}, \ldots, X_{n:n}$.

   (c) Prove that $\bar{D}_n \stackrel{\mathrm{P}}{\to} 0$ as $n \to \infty$ in two different ways:

      i. Do this by appealing to the Glivenko–Cantelli theorem.

      ii. Do this by first computing the variance of $\bar{D}_n$. [You may not use the Glivenko–Cantelli theorem for this part.]

3. Let $X_1, \ldots, X_n$ be a random sample from a CDF F and $Y_1, \ldots, Y_m$ an independent random sample from a CDF G. We wish to test $H_0 : F = G$ versus the two-sided alternative, $H_1 : F \neq G$. Let $\hat{F}_n$ and $\hat{G}_n$ denote the respective empirical CDFs of the X_i's and the Y_j's.

   (a) Describe a condition on F and/or G under which

$$\Delta_n := \max_x |\hat{F}_n(x) - \hat{G}_n(x)|$$

   is distribution free; you need to justify your assertions.

   (b) Compute, numerically, $\mathrm{P}\{\Delta_{40} \leq x\}$ for $x = 0.09, 0.1, 0.11, 0.12$. Describe your algorithm and justify why it works. Make detailed comments on how accurate your computations are.