Math 6070-1: Spring 2013
Problem set 4

Due date: March 20, 2013

Recall that if $X_1, \ldots, X_n$ are i.i.d. with a continuous distribution function F, then the distribution of D_n does not depend on F, where D_n denotes the Kolmogorov–Smirnov statistic,

$$D_n := \max_{-\infty < x < \infty} \left| \hat{F}_n(x) - F(x) \right|,$$

and $\hat{F}_n(x) := \sum_{j=1}^n \mathbf{I}\{X_j \le x\}$ denotes the empirical distribution function.

1. Use simulation in order to answer the following for every integer $n = 10, 20, 30, \ldots, 100$, and $n = 1,000$: Compute δ such that

$$\mathrm{P}\{D_n > \delta\} = \alpha,$$

for every $\alpha = 0.01, 0.05, 0.1$. Report your findings in a carefully-drafted table. Explain your simulation/statistical methods in detail.

2. In the lecture notes, the following approximation is recommended when n is large [or reasonably large] and $x \ge 1$:

$$\mathrm{P}\left\{ D_n > \frac{x}{\sqrt{n}} \right\} \approx 2\exp(-2x^2).$$

In particular, when $n = 1,000$ and $x \approx 1.358$, the preceding suggests that

$$\mathrm{P}\left\{ D_{1000} > 0.0429 \right\} \approx 0.05.$$

Is the preceding approximation effective when $n = 1,000$?

3. Recall the data set http://www.math.utah.edu/~davar/math6070/2013/Homework_NEW/wind.txt from Assignment 2. Apply the Kolmogorov–Smirnov test, and your simulation table, in order to perform an exact 95% test for normality of the data. You will need to apply plug-in estimates for μ and σ.

4. Suppose $X_1, \ldots, X_n$ are i.i.d. with unknown distribution function F and pdf $f := F'$.

(a) Show that if $\varphi(x) \geq 0$ and $\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x < \infty$, then

$$M_n := \int_{-\infty}^{\infty} \left| \hat{F}_n(x) - F(x) \right| \varphi(x)\,\mathrm{d}x < \infty.$$

(b) Find a function φ that makes M_n distribution free.

(c) For your choice of φ, describe how you would simulate a table of probabilities for M_n.

(d) For your choice of φ, describe how you would construct an exact 95% statistical test for checking whether or not a given pdf f is the underlying pdf.