Math 6070-1: Spring 2013
Problem set 3

Due date: March 6, 2013

1. Let $X_1, X_2, \ldots$ be an i.i.d. sample from a density function f. We assume that f is differentiable in an open neighborhood V of a fixed point x, and $B := \max_{z\in V} |f'(z)| < \infty$.

   (a) Prove that for all $\lambda > 0$, $m \geq 1$, and all $x \in \mathbf{R}$,
   $$\mathrm{P}\left\{\min_{1\leq j\leq m} |X_j - x| \geq \lambda\right\} = \left[1 - \int_{x-\lambda}^{x+\lambda} f(z)\,\mathrm{d}z\right]^m.$$

   (b) Prove that for all $\epsilon > 0$ small enough,
   $$\max_{z\in[x-\epsilon\,,x+\epsilon]} \left|f(x) - f(z)\right| \leq 2B\epsilon.$$
   Use this to estimate $|\int_{x-\epsilon}^{x+\epsilon} f(z)\,\mathrm{d}z - 2\epsilon f(x)|$.

   (c) Suppose that as $m \to \infty$, $\lambda_m \to \infty$ and $\lambda_m^2/m \to 0$. Then, prove that
   $$\lim_{m\to\infty} \frac{-1}{2\lambda_m} \ln \mathrm{P}\left\{\min_{1\leq j\leq m} |X_j - x| \geq \frac{\lambda_m}{m}\right\} = f(x).$$

   (d) Devise an estimator of $f(x)$ based on the previous steps.

2. Download and save the following 1,000 simulated data:
   http://www.math.utah.edu/~davar/math6070/2013/Homework_NEW/wind.txt.

   (a) Plot your data in several ways.

   (b) Suggest 3–4 density estimates that use different kernels and/or bandwidths, carry out the procedures, and estimate the number of the modes of the data, using Parzen's method.

   (c) Test to see if the data is normal, using a hybrid of a χ^2-squared test with plug-in parametric estimates of μ and σ. Report your results carefully.

   (d) Compare your test with the results of `qq-plot`.