In Class Workup 1-29-14
Form into working groups. Each person should be sure to understand how to solve each question. Keep your record on your own paper and study it as part of preparing for exam 1.

Your group leader will submit one paper (not their own) with all participating students names on it. You may solicit my help or the help of anyone in class.

1-9 Kernel density smoothing and bootstrap.

1. Express the kernel shown below mathematically as a function \( f \) on real line \( \mathbb{R} \) and also as a function \( f \) defined on \( \mathbb{R} \) in computer language \( \mathbb{R} \).

![Kernel Density](image)

2. Refer to (1). Calculate the variance \( \int_{-1}^{1} x^2 f(x) \, dx \) (twice \( \int_{0}^{1} x^2 f(x) \, dx \)) by hand.

3. Using facts proven in this course (do not prove them) determine \( h > 0 \) for which the function \( g(x) = \frac{1}{h} f(\frac{x}{h}) \) is a probability density having mean 0 and variance 1.

4. For \( g \) determined in (2) and data \( X_i = 3.1, 3.6 \) plot the weighted kernels \( \frac{1}{n} g(x - X_i) \) and plot their sum. The sum is the kernel density smoothing of the data.
5. If $f$ is any kernel (i.e. pr density having mean 0 and variance 1) the bandwidth $h$ define a density smoothing of data $X_i$ by kernel $f$ by $p(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} f\left(\frac{x-X_i}{h}\right)$.

5a. What is the variance of density $p$? Do not prove it, just use properties we covered.

5b. If $f$ is the standard normal density how many derivatives has density $p$ at every $x$?

6. We smooth an iid sample sam from a population pop. Our bandwidth is $h = 0.03$.

6a. Is our sample density estimating the population density? If not, what is it estimating?

6b. If the variance of the population is 200 what is the variance of the population smoothed by the same method we used on our sample?

6c. What is the approximate variance of the sample density?

7. 

7a. On a line use hashmarks $|$ to represent data for which *variable bandwidth* density smoothing is strongly indicated.

7b. In (7a) indicate an $x$ location at which small bandwidth $h$ is indicated and another location where larger bandwidth is indicated.

7c. Briefly explain the reasoning behind variable bandwidth density estimation.
8a. List all four iid bootstrap samples of n = 2.

8b. For each BS samples of 2 rough-sketch the sample density for a N[0, 1] kernel.

9. Label everything below.
9a. Sketch a hypothetical population density on [0, 10]. Use the right half page.

9b. In (9a) overlay a hypothetical sketch of many density estimates of (9a), each obtained from a different iid sample of (appreciable) size n. This plot shows the variations inherent in estimating population density by sample density.

9c. Butterfly a sketch left of (9ab) illustrating how we might hope bootstrap replicates in which sam plays the role of pop would reveal the approximate form of (9ab) but using only the one sample of n from pop (this is reliant on the bootstrap succeeding).
10-11 Random sampling from prescribed distributions using $U \sim \text{uniform}[0,1]$.

10. Sketch a density shape with the x axis labeled. The area need not equal one.
10a. Use the accept-reject method to obtain a r.v. $X$ having the probability distribution proportional to your curve. Clearly indicate how you are using the given uniform r.v., which ones result in rejection, and what the numerical value of the selected $X$ is.

\begin{center}
\begin{tabular}{cccccc}
U1 & 0.85198744 & 0.99541745 & 0.01840702 & 0.54963336 & 0.49330030 \\
U2 & 0.75961255 & 0.22250290 & 0.40361242 & 0.89962819 & 0.18632295 \\
\end{tabular}
\end{center}

10b. Suppose a density is proportional to a function $J(x) \leq C < \infty$. In terms of $J(x)$, $C$, U1, U2 express the condition under which we may accept $X = U1$ as a sample from the probability density proportional to function $J$. 
10c. Recollect the idea of looking down on a bunch of people placed in linear columns running perpendicular to a chalk line scale, indicated with their blood pressure scores. Selecting a person at random we get a blood pressure score having the distribution of blood pressures in the group. Tossing a cork down on the group at random in the box they occupy, if it hits someone, selects that individual at random. So we get accept-reject method. Illustrate and annotate this in such a way as to describe it.

11. Sketch a CDF.

11a. Use the random uniform sample $U = 0.65$ to select a random sample possessing the CDF you sketched. Place scales in the plot and illustrate what you are doing.

11b. Why is it working? Give the simple proof if $F$ is strictly increasing and continuous.
12. Importance Sampling

12. Probability density \( p(x) = 4 \) on \( 0 < x < 1 \), zero elsewhere. Probability density \( q(x) = \frac{x}{8}, 0 < x < 4 \), zero elsewhere.

12a. Use a uniform r.v. \( U \) to produce a sample from \( p \).

12b. Verify \( p \) is a probability density.

12c. Using mathematical expressions weight your \( X \) sample from \( p \) obtained in (12a) multiplying it by a weight \( W(X) \) to achieve \( \mathbb{E}_p W(X) X = \mathbb{E}_Q X \) (the Q-mean).

12d. For two passes at (12c) suppose \( U_1 = 0.32 \) and \( U_2 = 4.1 \). What is your estimate for the Q-mean using importance sampling? Give the numerical expression and reduce it to a number. Of course \( n = 2 \) is far too small for the resulting estimate to be reliable.

12e. If you had independently repeated (12d) for \( n = 100 \) iid uniforms what would you do to prepare a z-CI for the Qmean? Explain.

12f. Use calculus to discover the Q-mean and give its numerical value.

12g. If you had a large number \( n \) of iid \( p \)-samples what precisely could you plot to estimate the Q-density (not just estimate Q-mean but estimate the Q-density utilizing weights applied to a Psample).