Risk and Portfolio Management with Econometrics, Courant Institute, Spring 2009
http://www.math.nyu.edu/faculty/goodman/teaching/RPME09/index.html
Always check the class bboard on the blackboard site from home.nyu.edu (click on academics, then on the course name) before doing any work on the assignment.

Assignment 4, due February 18

Corrections: (none yet)

1. We are interested in estimating the large n bias of maximum likelihood estimation. We do this in two steps. The first step is a general mathematical fact that underlies Ito's lemma.

   (a) Suppose $X \sim \mathcal{N}(x_o, \epsilon)$. Suppose that $f(x)$ is a smooth and bounded function of x. Show that

   $$E\left[f(X)\right] \;=\; f(x_0) \,+\, \frac{\epsilon}{2} f''(x_0) \,+\, O\left(\epsilon^2\right) \;.$$

   (b) Show that the bias in maximum likelihood estimation satisfies, for large n,

   $$E\left[\widehat{\theta}\right] - \theta_* \approx \frac{c}{n} \;,$$

   and give a formula for c that has something in common with the formula for I. Hint: start as in class with

   $$0 \;=\; M(\widehat{\theta}) \;\approx\; M(\theta_*) \,+\, M'(\theta_*)\left(\widehat{\theta} - \theta_*\right) \,+\, \frac{1}{2} M''(\theta_*)\left(\widehat{\theta} - \theta_*\right)^2 \;.$$

   Assume that $M(\theta_*) \sim \mathcal{N}(0, I/n)$ and that M' and M'' have their expected values.

2. A *mixture model* is a model that says that a given sample is chosen from population 1 with probability p and from population 2 with probability $1 - p$. Suppose population j has probability density $f_j(x)$. For each sample, we first choose $j = 1$ or $j = 2$ with $Pr(j = 1) = p$. Then we choose $X \sim f_j$. All choices are independent. In this exercise, we suppose the distributions f_1 and f_2 are known and that we only want to estimate the mixture parameter, p.

   (a) Write an expression for the probability density $f(x, p)$, where p is the probability parameter above. Show that the information in estimating p is given by

   $$I \;=\; E\left[\frac{\left(f_1(X) - f_2(X)\right)^2}{f(x,p)^2}\right] \;=\; \int \frac{\left(f_1(X) - f_2(X)\right)^2}{f(x,p)}\, dx \;. \qquad (1)$$

(b) Suppose $f_1 = \mathcal{N}(0,1)$ and $f_2 = \mathcal{N}(0,4)$. Use Matlab to make a plot of the integrand (1) and use some integration method (trapezoid rule, Simpson's rule, some built in Matlab integrator, whatever) to compute the integral to within 1%. The reason to plot the integrand before integrating is that the plot will tell you what range of integration to use. Use $p = .2$.

(c) Generate a set of L independent samples of the distribution of part (b) (with $p = .2$) and use the method of maximum likelihood to estimate p. Take $L = 50$. The challenge of this part is that you have to find a way to solve the equations $M(\vec{X}, \widehat{p}) = 0$, given your artificial data set $\vec{X} = (X_1, \ldots, X_L)$. This amounts to finding a way to solve one nonlinear equation in one unknown. You can use bisection or Newton's method or find a Matlab function that does it.

(d) Repeat the computation of part (c) n times and get n independent estimates $\widehat{p}_k$ for $k = 1, \ldots, n$. Each $\widehat{p}$ comes from a new and independent dataset $\vec{X}$. This requires nL independent samples of f in all. Compute the sample mean and sample variance of the $\widehat{p}$. Do these satisfy our theoretical large n theory of mean and variance of $\widehat{p}$?

(e) Make a histogram of the samples of $\widehat{p}$ and compare it to a normal with the correct mean and variance. Take $n = 10^6$.

(f) Repeat part (e) with $L = 10$ and $L = 200$. For $L = 10$, you should clearly see departure of the distribution of $\widehat{p}$ from normality. Note that the mean and variance of $\widehat{p}$ depend on L.

3. Consider a statistical model $Y = a + bX + R$, where R has a two sided exponential distribution with density

$$f(r) = \frac{\lambda}{2} e^{-\lambda |r|} . \tag{2}$$

This is a non-gaussian distribution that has thin tails, though fatter than gaussian tails. Suppose we have L pairs (X_k, Y_k) and we want to estimate a and b using maximum likelihood using the probability model

$$Y_k = a + bX_k + R_k , \tag{3}$$

where the R_k are independent samples of the density (2).

(a) Show that (2) represents a probability density by checking that the integral is one.

(b) Show that maximum likelihood estimation of a and b leads to the condition that a be the median of the numbers $Y_k - bX_k$, and that $\sum \pm X_k = 0$, where the sign multiplying X_k depends on the sign of the residual for point k.

(c) What is the maximum likelihood estimator of λ?

(d) Let $\widehat{a}_n$ and $\widehat{b}_n$ be the estimates of a and b using n data points. Show that there is an absolute bound C (which depends on $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, but not on the new data point, so that

$$\left| \widehat{a}_{n+1} - \widehat{a}_n \right| \leq C \quad , \qquad \left| \widehat{b}_{n+1} - \widehat{b}_n \right| \leq C \; .$$

This C is independent of the new data point, (X_{n+1}, Y_{n+1}), no matter how large or strange it may be. Show that "linear" regression (based on least squares, which is maximum likelihood with a gaussian error model) does not have that property. On the contrary, a new data point can make arbitrarily large changes to the regression coefficients. This shows that using the double exponential error model leads to more *robust* estimates of a and b. Accidental corruption of a small number of data points does not completely ruin the answer.