Statistics 221 – Assignment 1

Due: Wednesday, March 3, 2004

As some of you may not have a copy of Lange yet, I have included the text of the questions that have been taken from Lange. For the problems that require programming, you may use any package or language that you like (though S-Plus/R or Matlab are suggested), but as part of your answer, please include documented code indicating what each section of the code does.

1. Lange 5.13

   Newton's method can be used to extract roots. Consider the function $g(x) = x^m - c$ for some integer $m > 0$ and $c > 0$. Show that Newton's method is defined by

   $$x_n = x_{n-1}\left(1 - \frac{1}{m} + \frac{c}{m x_{n-1}^m}\right)$$

   Prove that $x_n \geq c^{\frac{1}{m}}$ for all $x_{n-1} > 0$ and that $x_n \leq x_{n-1}$ whenever $x_{n-1} \geq c^{\frac{1}{m}}$. Thus, if $x_0 \geq c^{\frac{1}{m}}$, then x_n monotonely decreases to $c^{\frac{1}{m}}$. If $0 < x_0 < c^{\frac{1}{m}}$, then $x_1 > c^{\frac{1}{m}}$, but thereafter, x_n monotonely decreases to $c^{\frac{1}{m}}$.

2. For each of the following equations, determine a function $g(x)$ and an interval $[a, b]$ on which fixed-point iteration will converge to a positive solution of the equation.

   (a) $3x^2 - e^x = 0$

   (b) $x - \cos x = 0$

3. For many adult onset diseases, such as Alzheimer's or Schizophrenia, information about lifetime risk of disease in relatives can be used to infer information about possible genetic models for the disease of interest. For these diseases, there is variable age of onset, so a person who is susceptible for the disease may be unaffected since they haven't reached their age of onset yet.

   Assume that you have n independent samples, where

   $$z_i = \begin{cases} 1 & \text{if subject } i \text{ is affected} \\ 0 & \text{if subject } i \text{ is not affected} \end{cases}$$

   and

   $$t_i = \begin{cases} \text{Age of onset for subject } i & \text{if } z_i = 1 \\ \text{Censoring age for subject } i & \text{if } z_i = 0 \end{cases}$$

   Then the likelihood function has the form

   $$\prod_{i=1}^{n} \{p f(t_i) z_i + (1 - pF(t_i))(1 - z_i)\}$$

where $f(t)$ and $F(t)$ are the density and cumulative distribution function of the age of onset distribution respectively.

Use both Newton-Raphson and bisection to find the maximum likelihood estimate of p, the lifetime risk for the dataset `risk.dat`, which is available on the Assignments page of the course web site. When estimating p, assume that the age of onset distribution is normal with mean $\mu = 40$ and standard deviation $\sigma = 15$. Also find the the standard error of $\hat{p}$ based on the information for p.

(Note: the dataset is a modification of the dataset collected by Winokur et al. (1969) on Major Affective Illness)

4. Lange 6.11

   Prove that the series $B_n = \sum_{k=0}^{n} \frac{A^k}{k!}$ converges. Its limit is the matrix exponential e^A.

5. Lange 6.14

   Relative to any induced matrix norm, show that $\text{cond}(A) \geq 1$ and that $\text{cond}(A^{-1}) = \text{cond}(A)$ and $\text{cond}(cA) = \text{cond}(A)$ for any scalar $c \neq 0$. Also verify that if U is orthogonal, then $\text{cond}_2(U) = 1$ and

$$\text{cond}_2(A) = \text{cond}_2(AU) = \text{cond}_2(UA).$$