STAT 243   ASSIGNMENT 3   FALL, 2009

1. Cholesky Decomposition

Write and test a function for the Cholesky decomposition of a symmetric positive definite matrix. Test it on an arbitrary symmetric positive definite matrix, and verify that it works through multiplication or using R or matlab.

2. Simulation

In a one-way ANOVA, we test the null hypothesis that the means of several different groups are all equal to each other. Let x_{ij} represent the jth observation in the ith group, with $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$. Then a suitable test statistic for the null hypothesis is:

$$F = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2/(k-1)}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2/(N-k)}$$

where

$$\bar{x}_i = (1/n_i) \sum_{j=1}^{n_i} x_{ij}, \qquad \bar{x} = (1/N) \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}, \qquad \text{and } N = \sum_{i=1}^{k} n_i$$

Under the null hypothesis of the means being equal for all k groups, the statistic F follows the F-distribution with $k-1$ and $N-k$ degrees of freedom.

(a) Generate 5 groups of data, each with 6 uncorrelated observations from a normal distribution with mean 0 and variance 1. (This means that in the formulas above, $k = 5$, and $n_i = 6$ for $i = 1, \ldots, 5$.) Use your programs from assignment 2 to generate the random numbers. Test to see if the F-statistic described above follows an F-distribution by comparing the upper 90th, 95th and 99th percentiles of your observed F-statistics with their theoretical values found either in tables or from the appropriate `dcdflib` routine. (You'll need to sort your data to find the percentiles; either use routines from a subroutine library such as the GNU Scientific Library (`libgsl`), or "borrow" the function in sort.c from the s243 class account – i.e. - `~s243/samples/sort.c`)

(b) Repeat part a), but instead of using 6 uncorrelated observations in each of the 5 groups, generate 6 correlated variables with $\text{corr}(x_i, x_j) = .7^{|i-j|}$. What does this tell you about the behavior of the F-statistic in a one-way ANOVA when the data is correlated?

3. Gram-Schmidt orthogonalization

Write and test a function for the Gram-Schmidt orthogonalization of an arbitrary $n \times p$ matrix. Test it on a matrix of your choice, and verify that it works either through multiplication or using R or matlab.

4. Regression using Gram-Schmidt orthogonalization

Write a program which takes as input a matrix of X and y values and then performs a regression using the function for Gram-Schmidt orthogonalization that you wrote in part 3. The output of the program should include the parameter estimates, the standard errors of the parameters, the estimated value of σ^2 and the residual value for each observation.

Hints: Let X be $n \times p$ and y be $n \times 1$. If you use the augmented matrix $X : y$ instead of just X, and orthogonalize only the first p columns, then the last column of the orthogonalized matrix will be the residuals. To get the standard errors of the parameter estimates, you may need to invert an upper triangular matrix. You can use the following algorithm. Let T be a $p \times p$ upper triangular matrix, and let U be it's inverse.

$$
\begin{aligned}
&\text{for } j = \ p \text{ to } 1 \text{ by } -1 \\
&\quad u_{jj} = 1/t_{jj} \\
&\quad \text{for } k = j - 1 \text{ to } 1 \text{ by } -1 \\
&\qquad u_{kj} = -(\textstyle\sum_{i=k+1}^{j} t_{ki} u_{ij})/t_{kk} \\
&\quad \text{end} \\
&\text{end}
\end{aligned}
$$