STAT/BIOSTAT 534 Statistical Computing
Homework 7

Adrian Dobra
adobra@uw.edu

Problem 1 (100 points)

Please write a C/C++ program that performs the following functions.

1. Load the file "erdata.txt". This is the same data file you used in previous assignments. It has $n = 158$ rows and $p = 51$ variables. Use the GSL function `gsl_stat_covariance`
`http://www.gnu.org/software/gsl/manual/html_node/Covariance.html`
to calculate the $p \times p$ sample covariance matrix Σ of these p variables.

2. Write a function that draws independent samples from the multivariate normal distribution $N_p(0, \Sigma)$. To construct the sampler, you need to obtain the Cholesky decomposition of Σ, i.e.

$$\Sigma = \Psi \Psi^T,$$

where Ψ is a $p \times p$ lower triangular matrix. The relevant GSL function for performing the Cholesky decomposition of Σ is `gsl_linalg_cholesky_decomp`:
`http://www.gnu.org/software/gsl/manual/html_node/Cholesky-Decomposition.html#index-gsl_005flinalg_005fcholesky_005fdecomp-1343`

Next you should generate p independent $N(0, 1)$ random numbers $Z_1, Z_2, \ldots, Z_p$ using the GSL function `gsl_ran_ugaussian`:
`http://www.gnu.org/software/gsl/manual/html_node/The-Gaussian-Distribution.html#index-gsl_005fran_005fugaussian-1684`

Arrange $Z_1, Z_2, \ldots, Z_p$ in a $p \times 1$ column matrix Z. Your random draw from $N_p(0, \Sigma)$ is given by $X = \Psi Z$. The most efficient way to calculate the matrix product between Ψ and Z is to use the GSL BLAS function `gsl_blas_dgemm`. While I would encourage you to make use of this function, you are not required to do so to successfully complete your assignment.

3. Use the sampler you wrote to simulate 10000 random draws from $N_p(0, \Sigma)$. The output of your program should be the sample covariance matrix of the 10000 random draws. Check your code by comparing the elements of this matrix with the elements of Σ.