Homework #2—Bardsley: MATH 495
Due Wednesday, October 22, 2008

1. Note that the probability density function for $\mathbf{y} \sim N(\mathbf{0}, \mathbf{C})$, where $\mathbf{0}$ is the $n \times 1$ zero vector and $\mathbf{C}$ is an $n \times n$ covariance matrix, has the form

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{\sqrt{\det(\mathbf{C})(2\pi)^n}} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{C}^{-1}\mathbf{y}\right),$$

where "det" denotes determinant. If $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$, where $\mathbf{I}$ is the $n \times n$ identity matrix, show that

$$p_{\mathbf{y}}(\mathbf{y}) \propto p_{\mathbf{x}}((\mathbf{R}')^{-1}\mathbf{y}),$$

where $\mathbf{C} = \mathbf{R}'\mathbf{R}$ is the Cholesky factorization of $\mathbf{C}$.

This motivates the use of the following MATLAB commands (see problem 1(b), homework #1—Bardsley) for computing random draws from $\mathbf{y}$ when the covariance matrix $\mathbf{C}$ is 2×2:

```
R = chol(C);
x = randn(2,n);
y = R'*x;
```

Note that `randn(2,n)` generates n random draws from $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$.

2. In order to see that using a numerical ODE solver will give effectively the same results as an analytic solution, return to the aids data and logistic ODE (problem 3 in HW #1) and compute the parameters for the optimal fit (`b_opt`) using: (i) the analytic solution of the logistic ODE (what your code already does); and (ii) a numerical solution of the logistic ODE using `ode23`. Report the values of `b_opt` in both cases, using the `format long` command to obtain more decimal places in the numbers you report. Email me the function evaluation code that you've modified.

Note: no sampling is needed in this problem.

3. Modify `LynxHare.m` so that it instead implements the parameter estimation/sampling approach on the following British boarding school data (British Medical Journal, March 4, 1978, p. 587), where the infected are taken to be the boys confined to bed and the total number of individuals is $N = 763$:

days	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Infected	1	3	7	25	72	222	282	256	233	189	123	70	25	11	4

Use the SIR ODE model to fit the data:

$$\begin{aligned} \frac{dS}{dt} &= -aIS, \quad S(0) = S_0, \\ \frac{dI}{dt} &= aIS - bI, \quad I(0) = I_0, \end{aligned}$$

where S denotes susceptible and I infected individuals. Report your parameter estimates as well as 95% confidence intervals from the normal based theory, and the quantile approach for both the Monte Carlo and bootstrapping sampling methods. Email me your codes for this problem.