Homework #1—Bardsley: MATH 495
Due Wednesday, October 15, 2008

Sampling from Normal Distributions:

1. (a) Write a MATLAB m-file that does the following: (i) generates 100,000 samples from $x \sim \mathrm{N}(0,1)$ (`samples=randn(100000,1)`); (ii) computes a normalized histogram of the samples of x using `nhist_fn.m` from the Bardsley 495 web site; (iii) uses MATLAB's `quantile` function to compute nonparametric 95% confidence interval enpoints (`quantile(samples,[0.025,0.975])`); (iv) plots the normalized histogram, the quantile points, and the normal pdf (`normpdf`) with mean 0 and variance 1 in the same plot. In the figure, make sure each of the above is visible, and give the plot an appropriate legend using the `legend` function. Compare the quantile confidence interval endpoints with theoretical values, which are approximately -1.96 and 1.96.

   (b) Write an m-file that computes 10,000 samples from

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \right)$$

   (`samples=R'*randn(2,10000)`, where `R` is the Cholesky factorization (square root) of the covariance). Plot the samples of x_1 versus x_2 and then repeat parts (ii)-(iv) in problem 1(a) for the samples of **both** x_1 and x_2. Compare the quantile confidence interval endpoints with theoretical values, which are approximately $-1.96\sigma_i$ and $1.96\sigma_i$, where σ_i is the std of x_i, for $i = 1, 2$.

   (c) Perform the χ^2 test for normality for the samples from x in (a) and both x_1 and x_2 in (b). This is accomplished for the samples from x in (a) via `[h,p]=chi2gof(samples)`. Report whether the test passes or fails and interpret the p values. There may be a better normality test to use here. If you think so, I'd like to hear about the method and why you think it's better.

$\star$ Email me a listing of your code for each part (one m-file is okay). Label the file appropriately, e.g. `Prob1HW1.m`. Your file should generate all of the figures and output asked for when I run it (you'll loose points if your code doesn't run!). Write accompanying commentary in a separate document.

Sampling from Non-Normal Distributions:

First, recall that in the m-file `NonlinDemo_MonteCarlo.m` we sampled (in a Monte Carlo fashion) from the non-normal distribution of the least squares parameter estimate $\hat{\boldsymbol{\beta}}$, and that we generated confidence intervals and figures in a fashion very similar to those obtained in problem 1 above.

2. The distribution of $\hat{\boldsymbol{\beta}}$ can also be sampled using the method of **bootstrapping**. Bootstrapping of the residuals for the test case in `NonlinDemo_MonteCarlo.m` is straightforward to implement. Simply modifying the code so that random samples from a Gaussian (see line 55) are replaced by random samples *with replacement* from the residual vector `r`, which is an outputted from `nlinfit` on line 36. (Use the fact that random samples with replacement from $\{1,2,3,4,5\}$ are obtained via `ceil(5*rand(5,1))`.) Compare the estimates and the nonparametric confidence intervals with those obtained using the Monte Carlo method and with the intervals obtained using the normal based theory.

$\star$ Email me a listing of your code. Label the file appropriately, e.g. `Prob2HW1.m`. Your file should generate all of the figures generated by `NonlinDemo_MonteCarlo.m` Write accompanying commentary in a separate document.

3. **Logistic Growth/AIDS Data:**

   (a) Modify the code `NonlinDemo_MonteCarlo.m` so that the first 12 years of the AIDS data contained in `aids.mat` on the web site is used and the nonlinear model is the logistic equation

$$X(X_0, M, k, t) = \frac{M X_0\, e^{kt}}{M + X_0(e^{kt} - 1)}.$$

The data and model vectors are then given by

$$\mathbf{y} = \begin{bmatrix} 260 \\ 992 \\ 2717 \\ 5341 \\ 8224 \\ 13195 \\ 21355 \\ 32196 \\ 35230 \\ 43352 \\ 45524 \\ 47572 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} X_0 \\ M \\ k \end{pmatrix}, \quad \mathbf{X}(\boldsymbol{\beta}) = \begin{bmatrix} X(X_0, M, k, 0) \\ X(X_0, M, k, 1) \\ X(X_0, M, k, 2) \\ X(X_0, M, k, 3) \\ X(X_0, M, k, 4) \\ X(X_0, M, k, 5) \\ X(X_0, M, k, 6) \\ X(X_0, M, k, 7) \\ X(X_0, M, k, 8) \\ X(X_0, M, k, 9) \\ X(X_0, M, k, 10) \\ X(X_0, M, k, 11) \end{bmatrix}$$

Use the initial guess `b_0 = [260, 47572, 1]'` for `nlinfit`. Also, you will have to modify figures 2, 3 and 4 to reflect the fact that you are estimating 3, rather than 2, parameters. Make your modifications so that the number of figures does not increase: in figure 2, you should plot the M samples versus both the k and X_0 samples, and the k samples versus the X_0 samples (use the `subplot` function); and figures 3 and 4 should contain 3, rather than 2, subplots. Finally, the output to the screen should contain the confidence interval information for all three parameters. Report these results together with those obtained using `nlparci`?

⋆ Email me a listing of your code for each part. Label the file appropriately, e.g. `Prob3aHW1.m`. Your file should generate all of the figures and output asked for. Write accompanying commentary in a separate document.

(b) Modify your code in 3(a) so that bootstrapping of the residuals is used instead. Report the confidence intervals obtained using the nonparametric approach. How do they compare with those obtained in 3(a)? Email me a listing of your code. Label the file appropriately, e.g. `Prob3bHW1.m`. Write accompanying commentary in a separate document.

Extra Credit: Fit a logistic model to *all* of the aids data found in `aids.mat` on the web site. Do this by breaking the data into two groups— one for the first 12 years (as in the previous parts) and the other for the remaining years—and by fitting a logistic curve with different parameter values for each group. Use bootstrapping to obtain confidence intervals for your parameter estimates. Email me a listing of your code for each part. Label the file appropriately, e.g. `ExCrHW1.m`. Your file should generate figures analogous to those in the previous example. Write accompanying commentary in a separate document.

Optimization and ODEs:

4. Compute (by hand) the Jacobian of $\mathbf{r}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{X}(\boldsymbol{\beta})$ for $\mathbf{X}(\boldsymbol{\beta})$, defined as in 3(a), with respect to $\boldsymbol{\beta}$. Use your result to fill in the lines of the m-file `logistic.m`, then try running `GaussNewton.m` which is on the web page. You should find that it doesn't work, even with a good initial guess. This motivates the Levenburg-Marquardt modification used by `nlinfit`. Hand in your hand calculations of $J(\boldsymbol{\beta})$ and email me a copy of `logistic.m`.

5. (a) Show that the logistic function defined in part (a) is a solution of the differential equation

$$\frac{dX}{dt} = kX\left(1 - \frac{X}{M}\right), \quad X(0) = X_0.$$

(b) Show that Weibull function (Problem 2, Homework 2, Graham)

$$y(t) = \alpha\left\{1 - \exp[-(t/\sigma)^\gamma]\right\}$$

is the solution of the of the differential equation

$$\frac{dy}{dt} = \frac{\gamma}{\sigma^\gamma} t^{\gamma-1}(\alpha - y), \quad y(0) = 0.$$