M647, Spring 2011, Assignment 5, due Friday Feb. 25

1. [10 pts] The general single species population model is

$$\frac{dy}{dt} = \frac{r}{a} y (1 - (\frac{y}{K})^a); \quad y(0) = y_0,$$

which can be solved exactly as

$$y(t) = \frac{K y_0}{\Big(y_0^a + (K^a - y_0^a) e^{-rt} \Big)^{1/a}}.$$

It's easy to see that if $a = 1$ this is simply the logistic model, and it's straightforward to show, using L'Hospital's rule, that the Gompertz model is obtained in the limit as $a \to 0$. Fit this model to the U. S. population data in *uspop.m* (available on the course web site), and use your results to argue that the Gompertz model is the best model from this family for fitting U.S. population growth.

Note. Physically a should be a positive parameter. Use MATLAB's documentation on *lsqcurvefit* to find out how to incorporate a lower bound on your parameter values.

2. [10 pts] Fit the data stored in *nlregdata2.mat* (available on the course web site) to the relation

$$y = p_1 x_1^{p2} x_2^{p_3}.$$

Compute our usual standard deviation estimate for your fit, and also find 95.45% confidence intervals for your parameter values.

Note. For this data, you have $q = N - m = 441 - 3 = 438$ degrees of freedom, and for values this large the student's t distribution becomes problematic to work with directly. (If you're not sure why, try computing *gamma(439/2)* in MATLAB.) On the other hand, for $q = 438$ the student's t distribution is essentially identical to the standard normal Gaussian distribution (Gaussian with mean 0 and variance 1). In light of this, you can find 95.45% confidence intervals by using exactly two standard deviations for the standard normal Gaussian: i.e., take $l = 2$.

3. [10 pts] Fit the data stored in *linearsystemregressiondata.mat* (available on the course web site) to the system

$$\begin{aligned} y_1 &= p_1 + p_2 x_1 + p_3 x_2 \\ y_2 &= p_4 + p_5 x_1 + p_2 x_2. \end{aligned}$$

(The last term, $p_2 x_2$ is not a typo; I want at least one parameter to appear in both equations.) In particular, do this by defining an appropriate design matrix F. Compute our usual standard deviation for your fit, and find 95% confidence intervals for the parameters. I simulated this data by taking a particular choice of parameter values. Can you guess what the choice was?

4. [10 pts] Fit the data stored in *systemregressiondata2.mat* (available on the course web site) to the nonlinear system

$$\begin{aligned} y_1 &= p_1 x_1^{p_2} + p_3 e^{p_4 x_2} \\ y_2 &= p_5 e^{p_4 x_1} + p_6 x_2^{p_2}. \end{aligned}$$

Compute our usual standard deviation estimate for your fit, and (this is the hard part) find 95% confidence intervals on your parameter estimates.

5. [10 pts] For linear regression, we said in class that our MLE estimator for $\vec{p}$ is

$$\hat{p} = (F^{tr} F)^{-1} F^{tr} \vec{Y},$$

where each component Y_k of $\vec{Y}$ is a Gaussian $N(\mu_k, \sigma^2)$ random variable. We claimed that for each $j = 1, 2, \ldots, m$,

$$\begin{aligned} E[\hat{p}_j] &= p_j \\ \mathrm{Var}[\hat{p}_j] &= V_{jj} \sigma^2, \end{aligned}$$

where the V_{jj} are diagonal elements of the *curvature* matrix $V = (F^{tr} F)^{-1}$. Derive these two equalities.