PSYC 6273 – Computer programming for experimental psychology
Winter 2011

Lecture 9 workshop

1. In the lecture demonstration code for bootstrapping, we used a statistical model to generate resampled data: we assumed that observer's responses on each trial were Bernoulli random variables, and we sampled the data from a binomial distribution.

An alternative method is to resample the original data. Generate some initial data like this:

```
x = -10 + 20*rand(1,50);
y = 0.05*(x-3).^2 + 5 + randn(size(x));
```

This gives us 50 (x,y) pairs, which would normally come from an experiment rather than being generated artificially. Generate each set of bootstrap data by picking 50 of these (x,y) pairs randomly, with replacement. Make a sum-of-squares fit of the following function to the initial dataset and to 100 resampled datasets:

```
fitfn = @(x,p) p(1)*(x-p(2)).^2 + p(3);
```

Report the sum-of-squares fit parameters for the initial dataset and their bootstrapped standard errors.

2. A single outlier can have a large effect on the usual method of calculating the standard deviation. That is, the usual method is not *robust*. This can be problematic when finding the standard deviation of bootstrapped parameter estimates.

One way of calculating a robust standard deviation is to (1) discard the 5% highest and 5% lowest data points, (2) use the usual formula, (3) multiply by a correction factor to compensate for throwing away the most extreme data points. Write a function `rstd.m` to calculate a robust standard deviation. Generate several datasets and test your function to make sure it gives the right answers.

Hint: read the help on `quantile` to learn how to find the 0.05 and 0.95 quantiles of a dataset. Discard the data values less than the 0.05 and greater than the 0.95 quantiles.

Hint: what should the correction factor be? Find the factor by trial and error so that it gives the right answer for a large sample with a known standard deviation, e.g., `randn(100000,1)`.