P318 Computational Modeling
Spring 2014
Week 9 Homework
Due: Monday March 17th

This assignment will ask you to explore Bayesian modeling using the simplest case of the binomial distribution.

Following the discussion in class, use the case of the coin flipping scenario, where you observe *x* heads out of *N* flips. The goal will be either to estimate the value *p*, the probability of observing a head, or to test the hypothesis that the coin is fair or unfair.

Recall that the likelihood of observing *x* heads out of *N* flips given that the probability of observing a head is *p* is given by the binomial distribution:

$$P(x \mid p) = Binomial(x \mid p) = \binom{N}{x} p^x (1-p)^{N-x}$$

Assume that the prior probability of *p* is given by the beta distribution:

$$P(p) = Beta(p \mid a,b) = \frac{1}{Be(a,b)} p^{a-1}(1-p)^{b-1}$$

While the parameters (*a*,*b*) of the beta distribution can be any positive real value, we can interpret *a* and *b* as the prior number of observed heads versus tails, respectively, if *a* and *b* are constrained to be integer values. A noninformative prior, a uniform distribution of values of p, is a special case when *a*=1 and *b*=1.

Recall that the posterior probability of *p* given that you observe *x* out of *N* heads assuming a binomial likelihood function and a beta prior is also a beta distribution:

$$P(p \mid x) = Beta(p \mid x+a, N-x+b)$$

Note that the posterior probability is only a beta distribution because the prior was assumed to be a beta distribution (the conjugate prior). If the prior was chosen to be another distribution, then the posterior would be a different distribution.

- Create a function that plots the prior, likelihood, and posterior. Use the subplot command to plot these vertically so that they are aligned with one another. Of course, your function should take as parameters all the necessary values to generate these plots. Feel free to use built-in pdf functions in Matlab for the binomial and beta distributions.

- Explore how the prior influences the posterior. You will need to decide on values of the data and setting on the prior to illustrate each of these:
  - First, using a uniform prior.
  - Second, using data that swamps the prior. In other words, set up a situation where there might be a strong prior, a strong *a priori* reason for *p* to have a particular value. But after seeing the data, the effect of the prior on the posterior is relatively weak.

- Third, using a prior that swamps the data. This is the opposite of the above. Set up a situation where there is a strong prior on p, data that suggests a different value of p, but that data is not strong enough to override the strong prior.

- Explore how the prior influences Bayesian model selection. Use the full equation for testing whether a coin is fair given in the class slides.

  $$\frac{P(M_{fair} \mid x)}{P(M_{unfair} \mid x)} = \exp\left(x\log(p_0) + (N-x)\log(1-p_0) - \log(Be(a,b)) - \log(Be(x+a, N-x+b))\right)$$

  For each, plot the Bayes factor as a function of different possible values of x given the particular values of N, a, and b you chose.
  - First, using a uniform prior.
  - Second, using data that swamps the prior (that the coin is fair).
  - Third, using a prior (that the coin is fair) that swamps the data.