Homework 11: Batting Averages

due Thursday, November 21 @ 9am

The batting average for hitter *i* is defined as his total number of hits, x_i, divided by the number of at bats, N_i

Your Task: Estimate a hitter's true batting average, μ_i, (for the entire season) from the first month of data.

In the file laa_2011_april.txt you will find the batting statistics for 13 players from the month of April. Use that info to predict their full-year batting averages (laa_2011_full.txt)

Assume the Following Statistical Model for batting:

$$x_i \mid \mu_i \sim \text{Binomial}(\mu_i, N_i)$$

N_i = number of at bats (AB), x_i = number of hits (H)

```
xi = pymc.Binomial('xi', n=N[i], p=mui, value=num_hits[i])
```

(a) Find the maximum likelihood estimate (MLE) of μ_i for each player from the April data.

(b) Draw a sample from the posterior (of size > 1000) assuming a Beta prior for each μ_i :

$\mu_i \sim$ Beta(alpha, beta)

The Beta prior encodes our prior knowledge about baseball batting averages. The hyperparameters alpha and beta should be chosen to reflect this belief.

For this assignment, we will choose alpha and beta based on the league-wide averages from 2010.

In 2010 the mean batting average was 0.255 and the variance between players was 0.0011.

Choose alpha and beta to satisfy these prior beliefs!

(c) Check convergence of your MCMC sampler by looking at the trace plots for at least 3 of the μ_i

(d) Compute the posterior mean and posterior 95% CI for each μ_i

For how many of the 13 players does the full-season batting average fall within the 95% CI?

(e) Make the following plots:

1. The full-season batting average of each player versus the MLE from (a)
2. The full-season batting average of each player versus the posterior mean from (d) (Include error bars to show the 95% CI).