

Homework 9

Due Friday, March 11, 2016

Batting Averages

Going to <http://espn.go.com/mlb/stats/batting> you can get batting averages for 146 players in the MLB in 2014. These are defined as the number of hits divided by at bats, rounded to 3 decimal digits.

The tab `Splits` allows you to subset the data. Select April. You are going to use the data on batting during the month of april for the same 146 players to predict their batting average for the entire season.

Specify a hierarchical model that may be meaningful in this setting. Implement a Gibbs Sampler to carry out the computations you might need and provide your predictions. How well did you do?

A testing problem

Let $X_i|\mu_i$ independent and $\sim \mathcal{N}(\mu_i, \sigma^2)$ with σ^2 known and $i = 1, \dots, n$. And let the prior distribution on μ be such that $\mu_i \sim \pi\mathcal{N}(0, \tau^2) + (1 - \pi)\delta_0$, with τ, π known. Calculate the posterior distribution and describe how you would choose between the hypotheses $H_0^i : \mu_i = 0$ and $H_1^i : \mu_i \neq 0$. Given that there are n such hypotheses is there a role for multiple comparison correction?

Examples of exchangeable parameters

List at least 3 examples, from different scientific/industry domains, where researchers are interested in a large number of parameters and it makes sense to assume a priori that the parameters are exchangeable. For each example, identify a scientific publication that uses a hierarchical Bayes approach.