Stat 471: Lecture 24
Bayesian Inference.

Moo K. Chung
mchung@stat.wisc.edu

November 5, 2003

1. For model $Y_i = g(x_i; \beta) + \epsilon_i$, we studied two methods for estimating parameters β and doing statistical inference. In the parametric regression, the distribution of ϵ_i is specified, i.e. $\epsilon_i \sim (0, \sigma^2)$ with unknown variance σ^2. In the bootstrap approach, we did not assume any distribution form for ϵ_i. In both parametric and nonparametric bootstrap approach, we assumed β to be fixed unknown scalar values to be estimated.

   In the *Bayesian* framework, parameters are assumed to be random and described by a probability distribution called the *prior distribution* $\pi(\beta)$. This is formulated before the data are observed based on a certain prior belief. Let $f(Y|\beta)$ be the sampling distribution of a sample Y_i. The *posterior distribution* of β is given by *Bayes formula*

   $$\pi(\beta|Y) = \frac{f(Y|\beta)\pi(\beta)}{\int f(Y|\beta)\pi(\beta)\, d\beta}.$$

   Note that

   $$\pi(\beta|Y) \propto f(Y|\beta)\pi(\beta) = \pi(\beta)\Pi_{i=1}^n f(y_i|\beta).$$

   The product term is the *likelihood*. The Bayes estimate $\hat{\beta}$ would be the mean of the posterior distribution, i.e.

   $$\hat{\beta} = \mathbb{E}(\beta|Y) = \frac{\int \beta f(Y|\beta)\pi(\beta)\, d\beta}{\int f(Y|\beta)\pi(\beta)\, d\beta}.$$

   In general one need to compute $\mathbb{E}(h(\beta)|Y)$ for Bayesian inference but this may be hard and Monte-Carlo integration based on MCMC might be one answer.

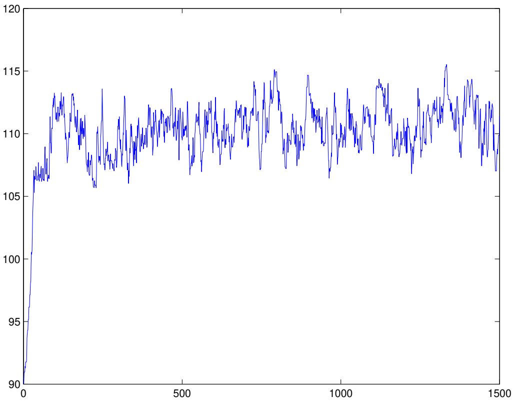
2. Consider a simple linear model

   $$Y_i = \beta_0 + \epsilon_i,\ i = 1, \cdots, n$$

   where $\epsilon_i \sim N(0, \sigma^2)$ with known σ^2. This can be written as $Y_i|\beta_0 \sim N(\beta_0, \sigma^2)$. Let the prior be $N(\mu, \tau^2)$. Then the posterior is

   $$\pi(\beta_0|y) \propto \exp(-\frac{(\beta_0 - \mu)^2}{2\tau}) \exp(-\frac{\sum_{i=1}^n (\beta_0 - y_i)^2}{2\sigma^2}).$$

   We can find, for instance, the Bayes estimate $\hat{\beta}$ analytically but let us estimate via MCMC. Let y_i be `grip` from `strength.data`. Based on the random-walk Metropolis algorithm, we simulate the posterior $\pi(\beta_0|Y)$ based on the symmetric propositional density $q(y|x) \sim N(x, 1)$.



```
sigma=23.6; mu=110.2; tau=5
n=1500; b(1)=90;
for i=2:n
  y=b(i-1) + normrnd(0,1);
  pi_n=normpdf(y,mu,tau)*prod(
         normpdf(grip,y,sigma));
  pi_d=normpdf(b(i-1),mu,tau)*prod(
        normpdf(grip,b(i-1),sigma));
  alpha=min([1,pi_n/pi_d]);
  u=rand;
  if u <= alpha
    b(i)=y;
  else
    b(i)=b(i-1)
  end
end
>> mean(b(500:1500))
ans =
  110.7409
```

3. Homework 5. Fit model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ using the 2D Gibbs sampler from lecture 23. You need to simulate 2D Markov chain (β_0^i, β_1^i) whose invariant distribution is the posterior $\pi(\beta_0, \beta_1|Y)$. All these examples are somewhat boring use of MCMC. Use it for complicated nonlinear model fitting when it is hard to compute $\mathbb{E}(h(\beta)|Y)$.

4. Lecture 25-27 will deal with Monte-Carlo optimization and the EM algorithm. Please read the textbook.