Monte Carlo Approximations to Posterior Distributions
September 21, 2009

Readings Hoff Chapter 4
Posterior Integration

Last time we talked about finding the posterior distribution of some function of $\theta$ (monotone functions)

- For expectations of $\phi = g(\theta)$ use “Theorem of the Unconscious Statistician”

$$\int_{g(\Theta)} \phi p(\phi \mid Y) d\phi = \int_{\Theta} g(\theta) p(\theta \mid Y) d\theta$$

- What if we do not know how to compute the integral?

- Common problem as we move in to higher dimensional parameters $(\theta_1, \theta_2, \ldots, \theta_p)$

Appeal to Law of Large Numbers!
Simulation from Distributions

Suppose we could take a sample of $S$ values from the posterior distribution of $\theta$

$$\theta^{(1)}, \ldots, \theta^{(S)} \overset{iid}{\sim} p(\theta \mid Y)$$

for large $S$.

- Law of Large Numbers

$$\frac{1}{S} \sum \theta^{(i)} \rightarrow \mathbb{E}[\theta \mid Y]$$

$$\frac{1}{S} \sum g(\theta)^{(i)} \rightarrow \mathbb{E}[g(\theta) \mid Y]$$
Distributions

\[ \theta^{(1)}, \ldots, \theta^{(S)} \overset{iid}{\sim} p(\theta \mid Y) \]

- Cumulative ordered values approximate \( F(\theta \mid Y) \) (cdf)
- Empirical distribution of the sample \( \theta^{(1)}, \ldots, \theta^{(S)} \) approximates \( p(\theta \mid Y) \) (use histogram or kernel density estimator)
- Probability \( g(\theta) > c \) approximated by proportion of samples where event \( g(\theta_i) > c \) occurs
- Sample moments/quantiles/functions approximate true moments/quantiles/functions

Extends easily to higher dimensional parameters
Giardia Example

Posterior $\pi \mid Y \sim \text{Beta}(7, 82)$

- Exact posterior mean $7/(7 + 82) = 0.0786$
- Quantiles: $0.03258, 0.0755, 0.1425$
- $\text{qbeta(c(.025, .5, .975), 7, 82)}$

Simulation based:

```r
> th = rbeta(10000, 7, 82)
> mean(th)
[1] 0.07818332
> quantile(th, c(.025, .5, .975))
  2.5%  50%  97.5%
0.03206155 0.07492002 0.14238109
```
Using functions in the CODA package

Download the CODA package and load into R

```r
> library(coda)
# Coerce the vector into a MCMC object
> theta.post = as.mcmc(th)
# find HPD interval using the CODA function
> HPDinterval(theta.post)

   lower  upper
var1  0.02927370 0.1370069
```

From `solve.HPD.beta` code

95% HPD interval (0.0277, 0.1358)
Odds

```r
> odds.post = as.mcmc(th/(1 - th))
> HPDinterval(odds.post)

   lower     upper
var1 0.02550343 0.1532560
```

Note: HPD regions not invariant under transformations! If $(\Theta_H)$ is a $1-\alpha$ 100% HPD region for $\theta$ and we are interested in $g(\theta)$ then

- $g(\Theta_H)$ is a $(1-\alpha)$100% probability region for $g(\theta)$
- $g(\Theta_H)$ is NOT a $(1-\alpha)$100% HPD region for $g(\theta)$

Why?
Exact Distribution of Odds

For the “energetic student”, starting with posterior distribution for $\theta$, use a change of variables to find the posterior density for the odds $o = \frac{\theta}{1 - \theta}$.
Comparing Distributions

Data from VA Hospitals.

- For each year observe $n$ patients and $y$, the number of cases (really failures).

- Observed data $Y = \{y_1, n_1; y_2, n_2\}$ for hospital 21:
  - In 1992, $y_1 = 306$, $n_1 = 651$
  - In 1993, $y_2 = 300$, $n_2 = 705$

First Model: Independent binomial outcomes in each year with probabilities $\theta_1$ and $\theta_2$
Question of Interest

Has the true probability of success/failure changed between 1992 and 1993?

- Independent continuous uniform priors →
  independent posteriors:

  - $\theta_1 \mid Y \sim \text{Beta}(307, 346)$ and
  - $\theta_2 \mid Y \sim \text{Beta}(301, 406)$ (independent of $\theta_1$)
  - $\theta_i$ independent and $y_i \mid \theta_i$ independent imply $\theta_i$

Graph posteriors – is there “Overlap?”
New parameter $\delta = \theta_2 - \theta_1$ measures difference.

- **Immediately:** $E(\delta \mid Y) = E(\theta_2 \mid Y) - E(\theta_1 \mid Y) = 0.426 - 0.470 = -0.044$.

- Is this significantly different from 0? Is it really negative? (improvement in care)

- **Immediately:**
  
  $V(\delta \mid Y) = V(\theta_2 \mid Y) + V(\theta_1 \mid Y) = 0.0275^2, \; \text{sd} = 0.0275$

- mean $\pm 2 \; \text{sd} = (-0.044 \pm 2 \cdot 0.0275)$ includes zero (rough)

Can compute $p(\delta \mid Y)$ by transformation – but messy. Use Simulation!
Posterior simulation

Large sample of $S$ values for $\theta_1$, similar for $\theta_2$ and then compute $\delta$

```r
> y1 = 306; y2 = 300; n1 = 651; n2 = 705;
> S=5000
> t1 = rbeta(S,y2+1,n2-y2+1)
> t2 = rbeta(S,y1+1,n1-y1+1)
> d = t1 - t2
> hist(d, nclass=30,prob=T)
> sum(d<0)/S
[1] 0.9494
```

About a 95% posterior probability that $\delta < 0$
The difference (of $\delta$ from 0) is “statistically significant at the 5% level” (one-sided).

For the VA, $\delta < 0$ represents an improvement in quality of care, so the data indicates a (very) likely improvement between 92 and 93.

A 90% (equal-tails) posterior interval for $\delta$ is

quantile(d, prob=c(0.05, 0.95))

A 90% HPD interval for $\delta$ is

HPDinterval(as.mcmc(d), prob=.90)

90% CI: $(-0.089, 0.000120)$ versus 90% HPD interval: $(-.094, -.0054)$
WinBUGS (Win = Windows, BUGS = Bayes Using Gibbs Sampling) is a scripting language for specifying Bayesian models and sampling from priors and posterior distributions.

model {
    y1 ~ dbin(theta1, 651)
    y2 ~ dbin(theta2, 705)
    theta1 ~ dbeta(1,1)
    theta2 ~ dbeta(1,1)
    delta <- theta2 - theta1
}

See WinBUGS the Movie!