Monte Carlo Approximations to Posterior Distributions
September 29, 2010

Readings Hoff Chapter 4

September 28, 2010
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”

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

- 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!
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

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

Sample means converge to their expectations
$\theta^{(1)}, \ldots, \theta^{(S)} \overset{iid}{\sim} p(\theta \mid Y)$

- Cumulative ordered values approximate $F(\theta \mid Y)$ (empirical 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: $0.03258, 0.0755, 0.1425$
- $q\text{beta}(c(0.025, 0.5, 0.975), 7, 82)$

Simulation based:

```r
> th = rbeta(10000, 7, 82)
> mean(th)
[1] 0.07818332
> quantile(th, c(0.025, 0.5, 0.975))
 2.5% 50% 97.5%
0.03206155 0.07492002 0.14238109
```

Sample size $S$ determines the accuracy
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

\[ \text{odds.post} = \text{as.mcmc}(\text{th}/(1 - \text{th})) \]
\[ \text{HPDinterval(odds.post)} \]

\begin{tabular}{lc}
\text{lower} & \text{upper} \\
\hline
\text{var1} & 0.02550343 0.1532560 \\
\end{tabular}

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?
Distributions

Draws from Posterior

- kernel density
- exact

Draws from Posterior

- kernel density
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 $o = \theta/(1 - \theta)$. 
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 $\theta_2$
Has the probability changed between 1992 and 1993?

- Independent continuous uniform priors $\rightarrow$ independent posteriors:
  - $\theta_1 | Y \sim \text{Beta}(307, 346)$ and
  - $\theta_2 | Y \sim \text{Beta}(301, 406)$ (independent of $\theta_1$)
  - $\theta_i$ independent and $y_i | \theta_i$ independent imply $\theta_i$ independent a posterior (Factorization Theorem)

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

- Immediately:
  \[
  E(\delta | Y) = E(\theta_2 | Y) - E(\theta_1 | Y) = 0.426 - 0.470 = -0.044.
  \]
- Is this significantly different from 0? Is it really negative? (improvement in care)
  - Immediately: $V(\delta | Y) = V(\theta_2 | Y) + V(\theta_1 | Y) = 0.0275^2$, sd = 0.0275
  - mean $\pm$ 2 sd = ($-0.044 \pm 2 \cdot 0.0275$) includes zero (rough)

Can compute $p(\delta | Y)$ by transformation – but messy.

Use Monte Carlo Simulation!
Posterior simulation

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

\[
\begin{align*}
&> y1 = 306; y2 = 300; n1 = 651; n2 = 705; \\
&> S=5000 \\
&> t1 = \text{rbeta}(S, y2+1, n2-y2+1) \\
&> t2 = \text{rbeta}(S, y1+1, n1-y1+1) \\
&> d = t1 - t2 \\
&> \text{hist}(d, \text{nclass}=30, \text{prob}=T) \\
&> \text{sum}(d<0)/S \\
&[1] 0.9494 \\
\end{align*}
\]

About a 95% posterior probability that \( \delta < 0 \)
(similar results with Jeffreys’ prior \( B(1/2, 1/2) \))
Posterior Densities

Uniform Priors (left) versus Jeffreys’ Priors (right)
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 highly probable improvement between 92 and 93

A 90% (equal-tails) posterior interval for $\delta$ is $\text{quantile}(d, \text{prob} = c(0.05, 0.95))$

A 90% HPD interval for $\delta$ is $\text{HPDinterval}(\text{as.mcmc}(d), \text{prob} = .90)$

90% CI: $(-0.089, 0.000120)$ versus 90% HPD interval: $(-0.094, -0.0054)$

HPD region has the shortest length of all credible intervals

Monte Carlo variability?
Predictive Distribution

Given data from 1993, what would be predictive distribution for failure for the next case that year?

- \( Y^* \mid \theta_2 \sim \text{Ber}(1, \theta_2) \)
- Plug in estimate (posterior mean) of \( \theta_2 \) in above
- Underestimates uncertainty
- Better: Find distribution of \( Y^* \mid Y_2 \)

Posterior Predictive Distribution
Exact Method

\[ p(Y^* \mid Y_2) = \int \frac{p(y^* \mid \theta_2) p(y_2 \mid \theta_2) p(\theta)}{p(Y_2)} d\theta \]

or

\[ p(Y^* \mid Y_2) = \int p(y^* \mid \theta_2) p(\theta_2 \mid Y_2) d\theta \]

Equivalent!

\[ \begin{align*}
E[Y^* \mid Y_2] &= E_{\theta_2 \mid Y_2}[E[Y^* \mid \theta_2]] \\
\text{Var}[Y^* \mid Y_2] &= \text{Var}[E[Y^* \mid \theta_2]] + E[\text{Var}[Y^* \mid \theta_2]]
\end{align*} \]
Simulation Method

For $s = 1, \ldots, S$

1. Generate $\theta_2 \mid Y_2 \sim B(y_2 + 1, n_2 - y_2 + 1)$

2. Generate $Y^* \mid \theta_2 \sim \text{Ber}(\theta_2)$

Histogram of $Y^*$ is estimate of posterior predictive density
Recap

1. sample $Y_1, \ldots, Y_n$ from a population model $p(y \mid \theta)$
2. sample means estimate expectations $\bar{Y}$ is an estimate of $E[Y]$
3. Bayes Theorem: Posterior distribution
   \[ p(\theta \mid Y_1, \ldots, Y_n) \propto L(\theta)p(\theta) \]
   - Calculate posterior expectations/probabilities analytically
   - Take a sample $\theta^{(1)}, \ldots, \theta^{(S)}$ from the “population” distribution
     \[ p(\theta \mid Y_1, \ldots, Y_n) \]
   - Sample means of $\theta$ estimate the “population” values: $\bar{\theta}$ is a point estimate of $E[\theta \mid Y_1, \ldots, Y_n]$