Monte Carlo Approximation

Yingbo Li

Clemson University

MATH 9810

Monte Carlo Approximation

Suppose we want to summarize posterior distribution of function of θ, say $\phi = g(\theta)$. For example, we might want to compute the expectation, $E(\phi \mid Y)$.

- We have

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

- 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 simulation and the Law of Large Numbers.

Simulation as Approximation

Suppose we can sample S values from the posterior distribution of θ, so that

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

for large S.

- Law of Large Numbers

$$\begin{aligned} \mathsf{E}[\theta \mid Y] &\approx \frac{1}{S} \sum \theta^{(i)} \\ \mathsf{E}[g(\theta) \mid Y] &\approx \frac{1}{S} \sum g(\theta^{(i)}) \end{aligned}$$

Sample means converge to their expectations

Simulated Distributions

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

- Cumulative ordered values approximate $F(\theta \mid Y)$ (empirical cdf)

$$P(\theta < c \mid Y) \approx \frac{\#(\theta^{(s)} < c)}{S}$$

- Empirical distribution of the sample $\theta^{(1)}, \ldots, \theta^{(S)}$ approximates $p(\theta \mid Y)$. Visualize with histogram or density estimator.
- Sample moments/quantiles/functions approximate true moments/quantiles/functions
- For example, proportion of samples where event $g(\theta^{(i)}) > c$ approximates $p(g(\theta) > c \mid Y)$

Extends to higher dimensional parameters

Television Example

Posterior with uniform prior: $\theta \mid Y \sim \text{Beta}(693, 357)$

- Exact posterior mean $693/(693 + 357) = 0.66$
- 95% central interval, i.e., quantile-based CI, $(0.631, 0.668)$
- This interval from qbeta(c(.025, .975), 693, 357)

Simulation based:

```
> th = rbeta(10000, 693, 357);
> mean(th);
[1] 0.6601443
> median(th);
[1] 0.6602355
> quantile(th, c(0.025, 0.975));
     2.5%     97.5%
0.6313407 0.6883350
```

S determines the accuracy. Make large when practical; usually 1000 is enough.

Functions of θ

Simulate posterior distribution of odds of having a television in room

$$o = \frac{\theta}{1-\theta} \Longrightarrow \theta = \frac{o}{1+o}, \quad \frac{d\theta}{do} = \frac{1}{(1+o)^2}$$

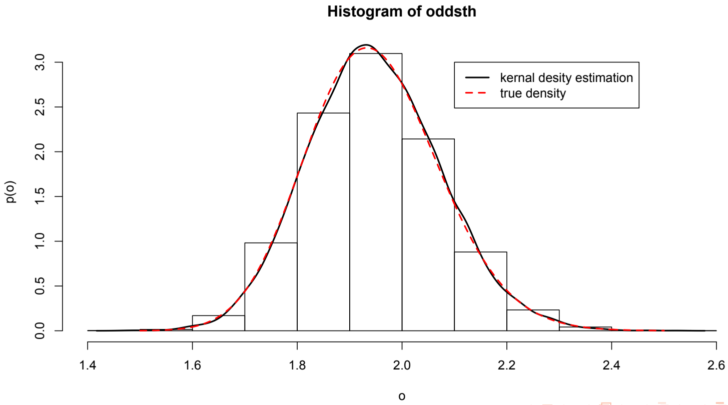
Change of variable $\theta \mid Y \sim \text{Beta}(a, b)$:

$$p(o \mid Y) = \frac{1}{B(a,b)} \cdot \frac{o^{a-1}}{(1+o)^{a+b}}$$

Monte Carlo approximation: draw independent samples from $p(o \mid Y)$:

```
> th = rbeta(10000, 693, 357)
> oddsth = th / (1 - th)
```

```
> hist(oddsth, prob = TRUE, xlab = 'o', ylab = 'p(o)');
> lines(density(oddsth), lwd = 2);
> tmp = seq(1.5, 2.5, by = 0.001);
> ptmp = exp(lgamma(693 + 357) - lgamma(693) - lgamma(357)
+ (693 - 1) * log(tmp) - (693 + 357) * log(1 + tmp));
> lines(tmp, ptmp, col = 2, lty = 2, lwd = 2);
```



Monte Carlo Approximation about o

Monte Carlo approximation: draw independent samples from $p(o \mid Y)$:

```
> oddsth = th / (1 - th);
> mean(oddsth);
[1] 1.947789
> median(oddsth);
[1] 1.943216
> quantile(oddsth, c(0.025, 0.975));
    2.5%    97.5%
1.712531 2.208573
```

Note that 95% central credible interval is invariant to (monotone) transformation.

$$\frac{0.6313}{1-0.6313} = 1.7122, \quad \frac{0.6883}{1-0.6883} = 2.2082$$

HPD interval from the CODA package

```
> library(coda)
# Coerce the vector into a MCMC object
> theta.mcmc = as.mcmc(th)
# find HPD interval using the CODA function
> HPDinterval(theta.mcmc)
         lower     upper
var1 0.6327825 0.6896061
```

From `solve.HPD.beta` code
95% HPD interval $(0.631, 0.689)$

HPD Regions Not Invariant

```
> library(coda)
> oddsth.mcmc = as.mcmc(oddsth)
> HPDinterval(oddsth.mcmc)
        lower   upper
var1 1.695732 2.18715
```

If (Θ_H) is a $100(1-\alpha)\%$ HPD region for θ, and we are interested in $g(\theta)$, then

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

$$\frac{0.6328}{1-0.6328} = 1.7233, \quad \frac{0.6896}{1-0.6896} = 2.2216$$

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 θ_1 and θ_2.
Question of Interest: has the probability changed between 1992 and 1993?

- Independent continuous uniform priors $\rightarrow$ independent posteriors:
- $\theta_1 \mid Y \sim \text{Beta}(307, 346)$ and
- $\theta_2 \mid Y \sim \text{Beta}(301, 406)$ (independent of θ_1)
- θ_i independent and $y_i \mid \theta_i$ independent imply θ_i independent *a posterior*

Difference

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: $V(\delta \mid Y) = V(\theta_2 \mid Y) + V(\theta_1 \mid Y) = 0.0275^2$, sd $= 0.0275$
- mean $\pm$ 2 sd $= (-.044 \pm 2 \times 0.0275)$ includes zero (rough)

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

Use Monte Carlo Simulation!

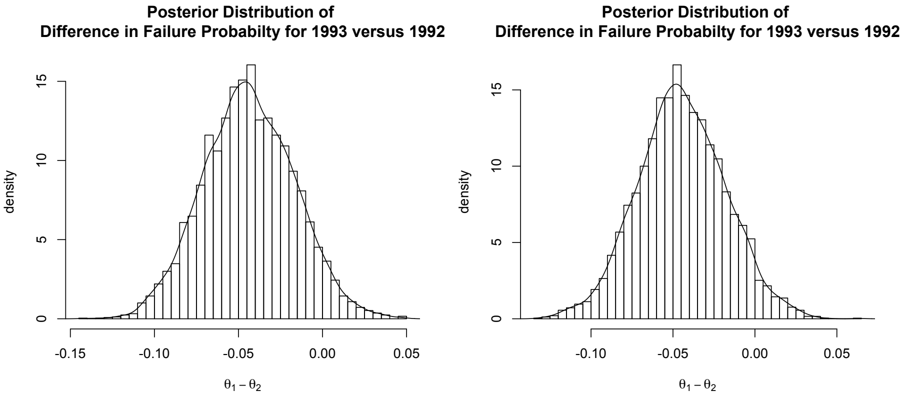
Posterior Simulation

Large sample of S values for θ_1, similar for θ_2 and then compute δ

```
> 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 $\delta < 0$
(similar results with Jeffreys' prior $B(1/2, 1/2)$)

Posterior Densities



Uniform Priors (left) versus Jeffreys' Priors (right)

Posterior Predictive Distributions

Given the traffic accident waiting time data, what is probability that the duration Y^* between the last and next accident exceeds 30 days?

- Model for eruption times: $Y^* \mid \beta \sim Exp(\beta)$
- Could plug in estimate (posterior mean) of β in above, but... this underestimates uncertainty
- Better: Find distribution of $Y^* \mid Y$, where Y is the data for the $n = 10$ measurements

Posterior Predictive Distributions

$$
\begin{aligned}
p(Y^* \mid Y) &= \int p(y^*, \beta \mid Y) d\beta \\
&= \int p(y^* \mid \beta, Y) p(\beta \mid Y) d\beta
\end{aligned}
$$

Can solve analytically, but easier to simulate this distribution. For $i = 1, \ldots, S$

1. Draw $\beta^{(i)}$ from $p(\beta \mid Y)$, which is $G(10, \text{rate} = 336)$
2. Generate Y^* from $p(Y \mid \beta^{(i)})$, which is $Exp(\beta^{(i)})$

Histogram of Y^* is estimate of posterior predictive density

Simulation in Traffic Accident Example

Simulate $S = 10000$ new waiting times

```
> beta = rgamma(10000, 10, rate = 336);
> newy = rexp(10000, beta)
```

Can summarize $p(Y^* \mid Y)$ via histograms and quantiles.

```
# posterior predictive probability that Y* > 30
> mean(newy > 30)
[1] 0.4292
# 95% probability for waiting time:
> quantile(newy, c(0.025, 0.975));
      2.5%        97.5%
  0.8908783 152.2844839
```

Analytic Solution

For exponential pdf and $\beta \mid y \sim G(10, 336)$, we have

$$
\begin{aligned}
p(Y^* \mid Y) &= \int p(y^* \mid \beta, Y) p(\beta \mid Y) d\beta \\
&= \int \beta \exp(-\beta y^*) \frac{336^{10}}{\Gamma(10)} \beta^{10-1} \exp(-336\beta) d\beta \\
&= \frac{336^{10}}{\Gamma(10)} \int \beta^{11-1} \exp(-(336 + y^*)\beta) d\beta
\end{aligned}
$$

Inside integral is kernel of Gamma distribution, $G(11, 336 + y^*)$. So we multiply by appropriate constant to make integral equal one, resulting in

$$
p(Y^* \mid Y) = \left(\frac{336^{10}}{\Gamma(10)} \right) \left(\frac{\Gamma(11)}{(336 + y^*)^{11}} \right)
$$

Inverse CDF Method for Sampling PDFs

We can find the cdf by integrating

$$\begin{aligned} F(w) = Pr(Y^* < w \mid Y) &= \int_0^w \frac{(10)336^{10}}{(336+y^*)^{11}} dy^* \\ &= -\frac{336^{10}}{(336+y^*)^{10}} \Big|_0^w \\ &= 1 - \frac{336^{10}}{(336+w)^{10}} \end{aligned}$$

To sample a value Y^* from $f(Y^* \mid Y)$,

1. Draw u from uniform on $(0,1)$.
2. Find w such that $F(w) = u$, i.e., $w = F^{-1}(u)$.
3. Value of w is a draw of Y^*

Posterior Predictive Model Checks

Can use posterior predictive distribution for model checking.

1. Specify a statistic $t(Y)$ that is useful for model checking, e.g., means, variances, functions of residuals.
2. Generate large number of new datasets from $p(Y^* \mid Y)$
3. Compute $t(Y^*)$ in each simulated dataset
4. Compute $p(t(Y^*) > t(Y)|Y)$ and $1 - p(t(Y^*) > t(Y)|Y)$
5. If either is small, suggests model does not describe Y very well

Posterior predictive checks cannot validate that your model is correct, only reveal if it does not describe the data well.

Generating Many Datasets

Simulate $S = 1000$ new datasets

```
> S = 1000;
> newy = matrix(nrow = S, ncol = 10);
> for(i in 1:S){
+   beta = rgamma(1, 10, rate = 336);
+   newy[i, ] = rexp(10, beta);
+ }
```

Note that I define “newy” as a maatrix beforehand, and populate its rows with the simulated data. R will give you errors if you do not define matrices or vectors before you populate them.

PPM Checks for Volcano Data

Reasonable test statistics include quantities like,

- $t_1(Y) = mean(Y)/sd(Y)$
- $t_2(Y) = max(Y)$
- $t_3(Y) = min(Y)$

Use the apply command for fast computations

```
> t1sims = apply(newy, 1, mean) / apply(newy, 1, sd);
> mean(t1sims > mean(y) / sd(y));
[1] 0.688
>
> t2sims = apply(newy, 1, max);
> mean(t2sims > max(y));
[1] 0.383
>
> t3sims = apply(newy, 1, min);
> mean(t3sims > min(y));
[1] 0.641
```