Gibbs Sampling

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

04/20/2020
Overview of posterior simulation methods

- Direct sampling
- Non-iterative indirect sampling:
  - Importance sampling
  - Rejection sampling
- Markov chain Monte Carlo sampling:
  - Metropolis-Hastings algorithm
  - Gibbs sampling
- And many more!
Let $\theta = (\theta_1, \ldots, \theta_k)$, with density $p$.

Recall: Draw direct samples $\theta^{(t)}$ from $p(\theta_1, \ldots, \theta_k)$ as follows:

- Draw $\theta_1^{(t)}$ from $p(\theta_1)$,
- Draw $\theta_2^{(t)}$ from $p(\theta_2 | \theta_1^{(t)})$,
- $\ldots$,
- Draw $\theta_k^{(t)}$ from $p(\theta_k | \theta_{k-1}^{(t)}, \ldots, \theta_1^{(t)})$.

Marginalizing over other parameters, as for $p(\theta_1)$, can be difficult.

Sampling from full conditionals, such as $p(\theta_k | \theta_{k-1}^{(t)}, \ldots, \theta_1^{(t)})$, is easier.

This motivates Gibbs sampling!
Gibbs sampling

- Wish to draw \( \theta^{(1)} \), \( \theta^{(2)} \), \ldots from distribution \( p \)
  - \( \theta^{(t)} = (\theta_1^{(t)}, \ldots, \theta_k^{(t)}) \)

- Specify initial vector \( \theta^{(0)} \)

- For \( t = 1, \ldots, T \):
  - Draw \( \theta_1^{(t)} \) from \( p(\theta_1 | \theta_2^{(t-1)}, \ldots, \theta_k^{(t-1)}) \)
  - Draw \( \theta_2^{(t)} \) from \( p(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_k^{(t-1)}) \)
    
    : 

  - Draw \( \theta_k^{(t)} \) from \( p(\theta_k | \theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{k-1}^{(t)}) \)

- Each draw from a full conditional using previous iteration
Gibbs sampling

- $\theta(t)$ converges in distribution to a draw from $p$ as $t \to \infty$
  - In practice, $p$ is a multivariate posterior density

- Gibbs sampling is a special case of Metropolis-Hastings
  - Proposal distributions are full conditionals, giving acceptance probability 1
Gibbs sampling is the standard tool for estimating large multivariate posteriors

As in MH sampling, convergence is quicker if initialization is in area of high posterior concentration

As in MH sampling, need to consider burn-in and auto-correlation
  - But no rejecting of proposals

With low autocorrelation, we say the parameters “mix well”

Gibbs sampling assumes we can obtain direct samples from full conditionals
  - Also “hybrid” versions, in which another method (such as MH) is used to sample from intractable full conditionals.
Example: IQ (cont.)

- Human IQs have variance 225 and are to be centered at 100
- Infer the “calibration” $\mu$ and variance $\sigma^2$ of a certain IQ test
- Sample of 20 individuals, each takes the test $n_i$ times.
  - $n = \sum_{i=1}^{m} n_i$

**Model:**

- $y_{ij} \sim \text{Normal}(\theta_i, \sigma^2)$ for $i = 1, \ldots, m$, $j = 1, \ldots, n_i$
- $\theta_i \sim \text{Normal}(\mu, 225)$ for $i = 1, \ldots, m$

- Use independent Jeffries priors for $\mu$, $\sigma^2$
  - $p(\mu, \sigma^2) = \frac{1}{\sigma^2}$
Example: IQ (cont.)

- The full joint density $p(y, \theta, \sigma^2, \mu)$ is

$$p(\mu, \sigma^2) \prod_{i=1}^{m} \text{Normal}(\theta_i \mid \mu, \tau^2) \prod_{j=1}^{n_i} \text{Normal}(y_{ij} \mid \theta_i, \sigma^2)$$

- Derive the full conditionals:
  - $p(\theta_i \mid y, \sigma^2, \mu) = \text{Normal} \left( \frac{\sigma^2 \mu + n_i \tau^2 \bar{y}_i}{n_i \tau^2 + \sigma^2}, \frac{\sigma^2 \tau^2}{n_i \tau^2 + \sigma^2} \right)$ for $i = 1, \ldots, 20$
  - $p(\sigma^2 \mid y, \theta, \mu) = \text{IG}(n/2, \sum_{i,j} (y_{ij} - \theta_i)^2 / 2)$
  - $p(\mu \mid y, \theta, \sigma^2) = \text{Normal}(\bar{\theta}, 225/m)$
Example: IQ (cont.)

- Use Gibbs to draw samples $(\theta^{(t)}, \sigma^2(t), \mu^{(t)})$ from
  \[
p(\theta, \sigma^2, \mu | y)
  \]

- Initial values: $\sigma^2 = 64, \mu = 100$
  - Begin sampling with $\theta$

- Each iteration draws from $p(\theta | y, \sigma^2, \mu)$, $p(\sigma^2 | y, \theta, \mu)$ and $p(\mu | y, \theta, \sigma^2)$

- Run 10000 iterations, using the first 2000 as burn-in

http://www.ericfrazierlock.com/Gibbs_Sampling_Rcode1.r
Gibbs draws $\mu^{(1)}, \mu^{(2)}, \ldots$:

Autocorrelation of draws $r = 0.058$
Estimated marginal posterior density for $\mu$: 

Histogram of posterior draws, $\mu$
Example: IQ (cont.)

- Gibbs draws $\sigma^2(1), \sigma^2(2), \ldots$:

- Autocorrelation of draws $r = 0.348$
Estimated marginal posterior density for $\sigma^2$:
Gibbs draws $\theta_i^{(1)}, \theta_i^{(2)}, \ldots$ for three individuals $i$:
Example: IQ (cont.)

- IQs should be centered at 100

- Check if the test is calibrated too high
  \[ P(\mu > 100 \mid y) \approx \frac{\sum_{i=1}^{N} \mathbb{I}_{\mu^{(i)}>100}}{N} = 0.996 \]

- Can use draws to construct 95% credible interval for IQ of each individual
  
  - Use the empirical 2.5% and 97.5% quantiles of simulated draws

http://www.ericfrazerlock.com/Gibbs_Sampling_Rcode1.r
Mean and 95% CI for each individual: