Gibbs Sampling

PUBH 8442: Bayes Decision Theory and Data Analysis

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

04/14/2021

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!

Multivariate posterior

- 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 $p(\theta_2 \mid \theta_1^{(t)})$,
  - ..., draw $\theta_k^{(t)}$ from $p(\theta_k \mid \theta_{k-1}^{(t)}, \ldots, \theta_1^{(t)})$.
- Marginalizing over other parameters, as for $p(\theta_1)$, can be difficult $p(\theta_1) = \int p(\theta_1, \ldots, \theta_k)\, d\theta_2 \cdots d\theta_k$
- Sampling from *full conditionals*, such as $p(\theta_k \mid \theta_{k-1}^{(t)}, \ldots, \theta_1^{(t)})$, is easier $p(\theta_k \mid \theta_1, \ldots, \theta_{k-1}) \propto p(\theta_1, \ldots, \theta_k)$
  - This motivates Gibbs sampling! $= \dfrac{p(\theta_1, \ldots, \theta_k)}{\int p(\theta_1, \ldots, \theta_k)\, d\theta_k}$

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 $p(\theta_1 \mid \theta_2^{(t-1)}, \ldots, \theta_k^{(t-1)})$
  - Draw $\theta_2^{(t)}$ from $p(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_k^{(t-1)})$

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

In MH, $r = \frac{h(\theta^*)\, q(\theta^{(t-1)} | \theta^*)}{h(\theta^{(t-1)})\, q(\theta^* | \theta^{(t-1)})}$

For Gibbs, $h(\vec{\theta}) \propto p(\theta_1, \ldots, \theta_K)$,

$\theta^* = (\theta_1^*, \theta_2^{(t-1)}, \ldots, \theta_K^{(t-1)})$ where $\theta_1^* \sim p(\theta_1 | \theta_2^{(t-1)}, \ldots, \theta_K^{(t-1)})$

$\Rightarrow r = \frac{p(\theta_1^*, \theta_2^{(t-1)}, \ldots, \theta_K^{(t-1)})}{p(\theta_1^{(t-1)}, \theta_2^{(t-1)}, \ldots, \theta_K^{(t-1)})} \cdot \frac{p(\theta_1^{(t-1)} | \theta_2^{(t-1)}, \ldots, \theta_K^{(t-1)})}{p(\theta_1^* | \theta_2^{(t-1)}, \ldots, \theta_K^{(t-1)})} \cdot \frac{p(\theta_2^{(t-1)}, \ldots}{p(\theta_2^{(t-1)}, \ldots}$

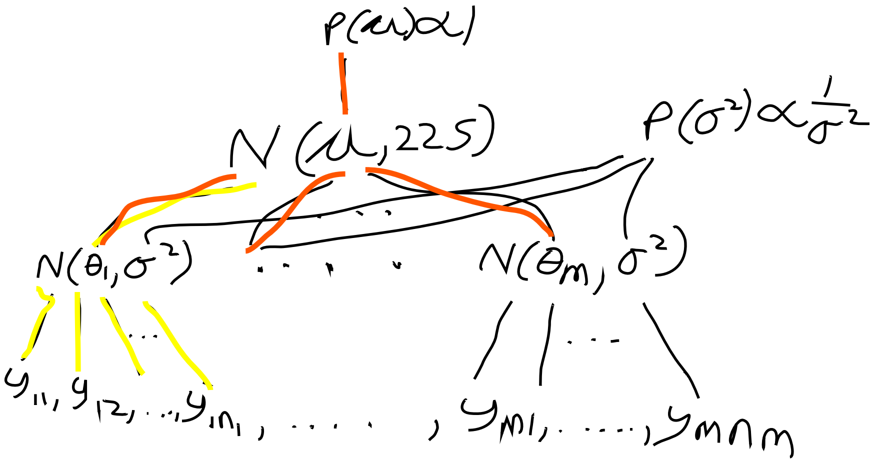
$= 1$

Comments

- 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" μ and variance σ^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 μ, σ^2
  - $p(\mu, \sigma^2) = \frac{1}{\sigma^2}$



p(μ) ∝ 1
N(μ, 225)
p(σ²) ∝ 1/σ²
N(θ₁, σ²)
N(θm, σ²)
y11, y12, ..., y1n1, ..., ym1, ..., ymnm

Example: IQ (cont.)

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

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

$p(\mu, \sigma^2, \vec\theta)$ $\quad$ $p(\vec y \mid \mu, \sigma^2, \vec\theta)$

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

$$\rho^2 = 225$$

$$p(\theta_i \mid \mu, \sigma^2, \vec{y}) \propto N(\theta_i \mid \mu, \rho^2) \prod_{j=1}^{n_i} N(y_{ij} \mid \theta_i, \sigma^2)$$

$$p(\sigma^2 \mid \vec{y}, \theta, \mu) \propto \frac{1}{\sigma^2} \prod_{i=1}^{m} \prod_{j=1}^{n_i} N(y_{ij} \mid \theta_i, \sigma^2)$$

$$= \frac{1}{\sigma^2} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i,j} (y_{ij} - \theta_i)^2}$$

$$\propto IG\left(\frac{n}{2}, \frac{1}{2} \sum_{i,j} (y_{ij} - \theta_i) \right)$$

Example: IQ (cont.)

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

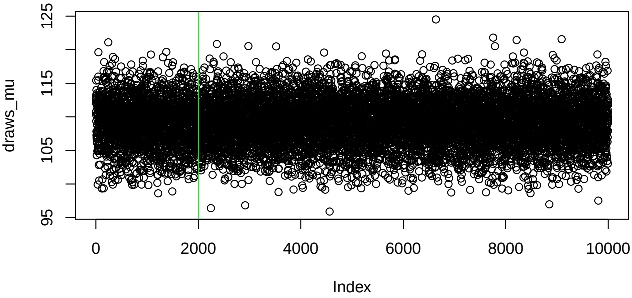
$$p(\theta, \sigma^2, \mu \mid \mathbf{y})$$

- Initial values: $\sigma^2 = 64, \mu = 100$
  - Begin sampling with θ
- Each iteration draws from $p(\theta \mid \mathbf{y}, \sigma^2, \mu)$, $p(\sigma^2 \mid \mathbf{y}, \theta, \mu)$ and $p(\mu \mid \mathbf{y}, \theta, \sigma^2)$
- Run 10000 iterations, using the first 2000 as burn-in

http://www.ericfrazerlock.com/Gibbs_Sampling_Rcode1.r

Example: IQ (cont.)

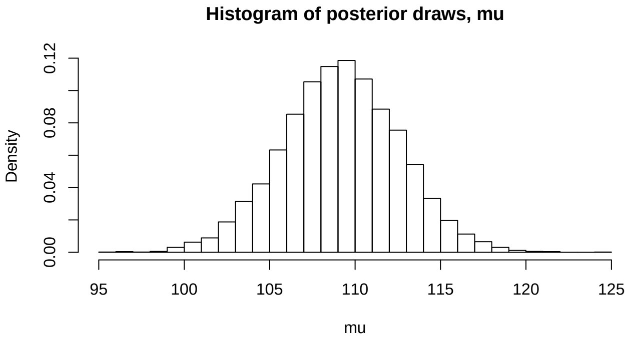
- Gibbs draws $\mu^{(1)}, \mu^{(2)}, \ldots$:



- Autocorrelation of draws $r = 0.058$

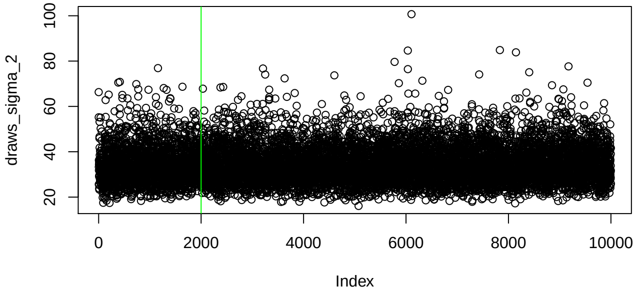
Example: IQ (cont.)

- Estimated marginal posterior density for μ:



Example: IQ (cont.)

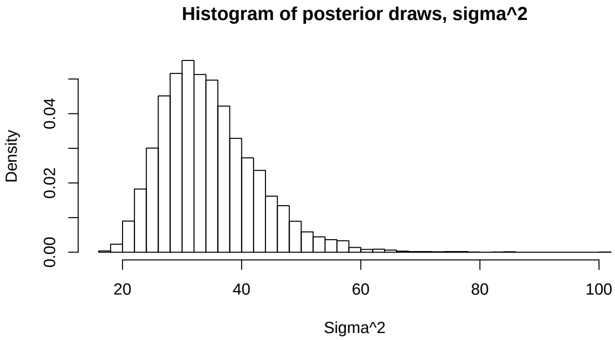
- Gibbs draws $\sigma^{2(1)}, \sigma^{2(2)}, \ldots$:



- Autocorrelation of draws $r = 0.348$

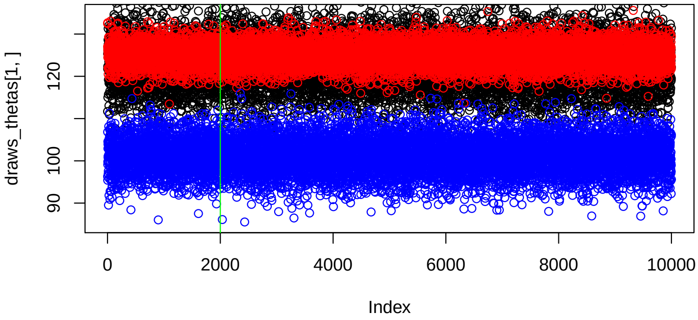
Example: IQ (cont.)

- Estimated marginal posterior density for σ^2:



Example: IQ (cont.)

- 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

$$P(\mu > 100 \mid \mathbf{y}) \approx \frac{\sum_{i=1}^{N} \mathbb{1}_{\mu^{(i)} > 100}}{N} = 0.996$$

- 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

Example: IQ (cont.)

- Mean and 95% CI for each individual: