# More on Posterior Computation

PUBH 8442: Bayes Decision Theory and Data Analysis

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

04/16/2024

# Gibbs Sampling Techniques

- Several techniques exist to reduce autocorrelation and improve convergence in Gibbs sampling.
- *Augmented* sampling introduces auxiliary variables to condition on during sampling.
  - E.g: To compute posterior for variables $A$ and $B$, introduce $C$ and sample from $P(C \mid A, B)$, $P(A \mid B, C)$ and $P(B \mid A, C)$.
- *Collapsed* sampling marginalizes over some variables.
  - E.g: Rather than drawing from $P(A \mid B, C)$, marginalize over $C$ and sample from $P(A \mid B)$.
- *Blocked* sampling draws from the joint distribution of variable sets rather than their separate full conditionals.
  - E.g.: Draw from $P(A, B \mid C)$ rather than $P(A \mid B, C)$ and $P(B \mid C, A)$.

# Example: Normal-Inverse-Gamma mixture sampling

- Draw $z_i \in \{1, \ldots, K\}$, to indicate the component that generated $y_i$, from $p(z_i = k \mid \mathbf{y}, q, \theta)$
  - Augmenting with $Z_i$'s to facilitate sampling other parameters
- Draw weights $q$ from $p(q \mid \mathbf{y}, \theta, \mathbf{z})$.
- Draw $\theta_k = (\mu_k, \sigma_k^2)$ from normal-inverse gamma posterior for $\mathbf{y}_k = \{y_i : z_i = k\}$:

$$\text{Draw } \sigma_k^2 \mid \mathbf{z}, \mathbf{y}, q \sim IG(a + n_k/2 - 1/2, b + 1/2||\mathbf{y}_k - \bar{y}_k||^2)$$

$$\text{Draw } \mu_k \mid \sigma_k^2, \mathbf{z}, \mathbf{y}, q \sim \text{Normal}\left(\frac{\sigma_k^2\mu_0 + n_k\tau^2\bar{y}_k}{n\tau^2 + \sigma_k^2}, \frac{\sigma_k^2\tau^2}{n\tau^2 + \sigma_k^2}\right)$$

  - Blocking $\mu_k, \sigma_k^2$
  - Draw for $\sigma_k^2$ collapses over $\mu_k$

# Hamiltonian Monte Carlo

- *Hamiltonian* Monte Carlo is an enhanced varient of Metropolis-Hastings sampling
- Proposals are motivated by Hamiltonian dynamics
  - Borrowing ideas of "potential" and "kinetic" energy in physics
- Informed by first-order gradient of posterior at each step
  - Allows draws to move more quickly to regions of higher posterior probability
- Often improves on Gibbs or MH sampling for complex posteriors
- Default estimation approach for the STAN software
- For an overview see this chapter by RM Neal

# Approximate Bayesian Computing (ABC)

- *Approximate Bayesian computing* simulates new data $\mathbf{y}^*$ and assesses how well the simulated data fits the observed data $\mathbf{y}$.
- Let $\mathbf{y}$ be discrete, with pmf specified by $\theta$: $P(\mathbf{y} \mid \theta)$
- To draw from $p(\theta \mid \mathbf{y})$, consider the following algorithm:
  - Propose $\theta^*$ from prior $p(\theta)$
  - Draw $\mathbf{y}^*$ from $P(\mathbf{y} \mid \theta^*)$
  - Accept $\theta^*$ if $\mathbf{y}^* = \mathbf{y}$
- The accepted $\theta^*$ represent direct draws from $p(\theta \mid \mathbf{y})$.

# Approximate Bayesian Computing (ABC)

- If $\mathbf{y}$ is continuous, accept $\theta^*$ if $\mathbf{y}^*$ is "close" to $\mathbf{y}$
- E.g., if $\mathbf{y} \in \mathbb{R}^n$, accept if

$$||\mathbf{y}^* - \mathbf{y}||^2 < \epsilon$$

  - Generally accepted $\theta^*$ yield draws from $p(\theta \mid \mathbf{y})$ as $\epsilon \to 0$
- Can be very inefficient – high rejection rates
  - Adaptive methods like this approach can help
- Only requires a way to generate samples $\mathbf{y}^*$ given $\theta$.
  - No need to fully specify the likelihood function $p(\mathbf{y} \mid \theta)$ or $P(\mathbf{y} \mid \theta)$
  - "likelihood-free"

# Example: socks

- Given $m$ socks are picked from laundry, and $y$ are unique, how many pairs of socks are in the load?[1]


[1] Example motivated by: http://www.sumsar.net/blog/2014/10/tiny-data-and-the-socks-of-karl-broman/

# Example: socks

- Let $n_{\text{pairs}}$ be the total number of pairs of socks
- Uniform prior on $n_{\text{pairs}} \in \{6, \ldots, 50\}$
- Observe $y = 9$ unique socks in first $m = 12$ taken out
- Would like to find the posterior $p(n_{\text{pairs}} \mid y)$
- Avoid specifying the likelihood $P(y \mid n_{\text{pairs}})$ directly

# Example: socks

- ABC approach:
  - Draw $n^*_{\text{pairs}}$ uniformly from $\{6, \ldots, 50\}$
  - Simulate a sample of $m = 12$ socks without replacement from $n^*_{\text{pairs}}$ pairs ($2 \times n^*_{\text{pairs}}$ total)
  - Compute number of unique socks $y^*$
  - If $y^* = 9$, consider $n^*_{\text{pairs}}$ a draw from $p(n_{\text{pairs}} \mid y)$
- Repeat above algorithm for $50,000$ proposed draws

http://www.ericfrazerlock.com/More_on_Posterior_Computation_Rcode1.r

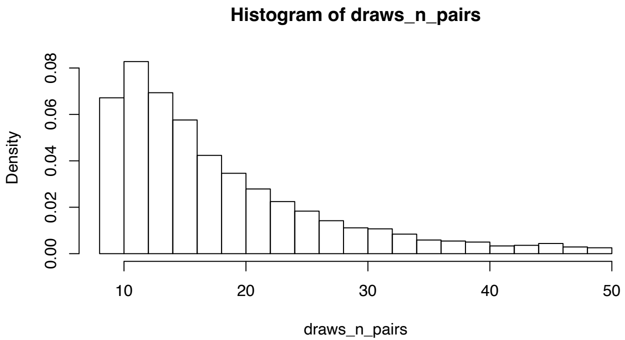
# R CODE

```
m=12
y=9
T=50000
draws_n_pairs = c()
for(i in 1:T){
  n_pairs_star = sample(c(6:50),1)
  ## vector with two 1's, two 2's, etc.
  ##(each number represents a sock pair):
  sim_socks = rep(c(1:n_pairs_star),2)
  sock_sample = sample(sim_socks,m)
  y_star = length(unique(sock_sample))
  if(y_star==y) draws_n_pairs = c(draws_n_pairs,n_pairs_star)
}

hist(draws_n_pairs, freq=FALSE, breaks=20)

AcceptanceRate = length(draws_n_pairs)/T
```

# Example: socks




- Acceptance rate: 0.11

# Variational Bayes

- A *variational Bayes* algorithm approximates the posterior $p(\theta \mid \mathbf{y})$ by a *variational distribution* $q(\theta)$
- Where $q(\theta)$, or $q(\theta \mid \nu)$, represents a family of distributions, usually of simpler form than $p(\theta \mid \mathbf{y})$
- The approximation is determined by minimizing a distance $D(q, p)$
- In *mean-field variational Bayes* this distance is Kullback-Leibler divergence:

$$KL(q||p) = \int q(\theta) \log \frac{q(\theta)}{p(\theta \mid \mathbf{y})d\theta}$$

- For *expectation propagation*, the KL-divergence is reversed: $D = KL(p||q)$

# Mean-field variational Bayes

- For mean-field variational Bayes we often assume the posterior can be factorized into $M$ chunks:

$$q(\theta) = \prod_{i=1}^{M} q_i(\theta_i)$$

- Iteratively update $q_i$'s to maximize $KL(q||p)$
  - Uses the mean-field approximation and calculus of variations
  - Analogous to the EM-algorithm
- Converges to a local optimum
- For example, could optimize over

$$q(\theta_i) = N(\mu_i, \sigma_i^2)$$

# Variational Bayes: Comments

- The accuracy of variational approximation $q$ can depend on:
  - Appropriateness of simplifying assumptions on $q$
  - Convergence properties of the iterative algorithm - may converge to a local mode that is not close to optimal
- Variational techniques tend to underestimate posterior uncertainty.
- Could use variational techniques to initialize MCMC sampling
- For more information see this tutorial by C. Fox and S. Roberts