Bayesian Computation

Andrew Gelman
Department of Statistics and Department of Political Science
Columbia University

Class 5, 6 Oct 2011
For Metropolis, compute \( \min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \right) \)

For Metropolis-Hastings, compute \( \min \left( 1, \frac{p(\theta^*|y) J(\theta|\theta^*)}{p(\theta|y) J(\theta^*|\theta)} \right) \)

Always compute log-density, never the density
Thus compute \( \exp \left( \log p(\theta^*|y) - \log p(\theta|y) \right) \), etc.

Don’t say \( \alpha = 0.65253 \) (unless the standard error is really 0.00002)
Review of homework 5

- For Metropolis, compute \( \min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \right) \)

- For Metropolis-Hastings, compute \( \min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \frac{J(\theta|\theta^*)}{J(\theta^*|\theta)} \right) \)

- Always compute log-density, never the density. Thus compute \( \exp(\log p(\theta^*|y) - \log p(\theta|y)) \), etc.

- Don’t say \( \alpha = 0.65253 \) (unless the standard error is really 0.00002)
Review of homework 5

- For Metropolis, compute 
  \[ \min \left( 1, \frac{p(\theta^* | y)}{p(\theta | y)} \right) \]

- For Metropolis-Hastings, compute 
  \[ \min \left( 1, \frac{p(\theta^* | y)}{p(\theta | y)} \frac{J(\theta | \theta^*)}{J(\theta^* | \theta)} \right) \]

- Always compute log-density, never the density
  Thus compute \( \exp \left( \log p(\theta^* | y) - \log p(\theta | y) \right) \), etc.

- Don’t say \( \alpha = 0.65253 \) (unless the standard error is really 0.00002)
Review of homework 5

- For Metropolis, compute \( \min \left( 1, \frac{p(\theta^* | y)}{p(\theta | y)} \right) \)

- For Metropolis-Hastings, compute \( \min \left( 1, \frac{p(\theta^* | y)}{p(\theta | y)} \frac{J(\theta | \theta^*)}{J(\theta^* | \theta)} \right) \)

- Always compute log-density, never the density. Thus compute \( \exp \left( \log p(\theta^* | y) - \log p(\theta | y) \right) \), etc.

- Don’t say \( \alpha = 0.65253 \) (unless the standard error is really 0.00002)

Andrew Gelman
Bayesian Computation
Review of homework 5

- For Metropolis, compute \( \min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \right) \)

- For Metropolis-Hastings, compute \( \min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \frac{J(\theta|\theta^*)}{J(\theta^*|\theta)} \right) \)

- Always compute log-density, never the density
  Thus compute \( \exp \left( \log p(\theta^*|y) - \log p(\theta|y) \right) \), etc.

- Don’t say \( \alpha = 0.65253 \) (unless the standard error is really 0.00002)
Your guesses

- Zach: $\delta = 0.93$, $\Sigma = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 20 \end{pmatrix}$
- Wei: $\delta = 2.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Michael: $\delta = 1.2$, $\Sigma = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}$
- Kristen: $\delta = 1.0$, $\Sigma = \begin{pmatrix} 2.0 & 0.5 \\ 0 & 2.3 \end{pmatrix}$
- Gustavo: $\delta = 0.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Vince: $\delta = 1.05$, $\Sigma = \begin{pmatrix} 2.0 & 6.4 \\ 6.4 & 47.0 \end{pmatrix}$
Your guesses

- **Zach**: $\delta = 0.93$, $\Sigma = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 20 \end{pmatrix}$

- **Wei**: $\delta = 2.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

- **Michael**: $\delta = 1.2$, $\Sigma = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}$

- **Kristen**: $\delta = 1.0$, $\Sigma = \begin{pmatrix} 2.0 & 0.5 \\ 0 & 2.3 \end{pmatrix}$

- **Gustavo**: $\delta = 0.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

- **Vince**: $\delta = 1.05$, $\Sigma = \begin{pmatrix} 2.0 & 6.4 \\ 6.4 & 47.0 \end{pmatrix}$
Your guesses

- **Zach**: $\delta = 0.93, \Sigma = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 20 \end{pmatrix}$
- **Wei**: $\delta = 2.5, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- **Michael**: $\delta = 1.2, \Sigma = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}$
- **Kristen**: $\delta = 1.0, \Sigma = \begin{pmatrix} 2.0 & 0.5 \\ 0 & 2.3 \end{pmatrix}$
- **Gustavo**: $\delta = 0.5, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- **Vince**: $\delta = 1.05, \Sigma = \begin{pmatrix} 2.0 & 6.4 \\ 6.4 & 47.0 \end{pmatrix}$
Your guesses

- Zach: $\delta = 0.93$, $\Sigma = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 20 \end{pmatrix}$
- Wei: $\delta = 2.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Michael: $\delta = 1.2$, $\Sigma = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}$
- Kristen: $\delta = 1.0$, $\Sigma = \begin{pmatrix} 2.0 & 0.5 \\ 0 & 2.3 \end{pmatrix}$
- Gustavo: $\delta = 0.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Vince: $\delta = 1.05$, $\Sigma = \begin{pmatrix} 2.0 & 6.4 \\ 6.4 & 47.0 \end{pmatrix}$
Your guesses

- Zach: $\delta = 0.93$, $\Sigma = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 20 \end{pmatrix}$
- Wei: $\delta = 2.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Michael: $\delta = 1.2$, $\Sigma = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}$
- Kristen: $\delta = 1.0$, $\Sigma = \begin{pmatrix} 2.0 & 0.5 \\ 0 & 2.3 \end{pmatrix}$
- Gustavo: $\delta = 0.5$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Vince: $\delta = 1.05$, $\Sigma = \begin{pmatrix} 2.0 & 6.4 \\ 6.4 & 47.0 \end{pmatrix}$
Your guesses

- Zach: $\delta = 0.93, \Sigma = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 20 \end{pmatrix}$
- Wei: $\delta = 2.5, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Michael: $\delta = 1.2, \Sigma = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}$
- Kristen: $\delta = 1.0, \Sigma = \begin{pmatrix} 2.0 & 0.5 \\ 0 & 2.3 \end{pmatrix}$
- Gustavo: $\delta = 0.5, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Vince: $\delta = 1.05, \Sigma = \begin{pmatrix} 2.0 & 6.4 \\ 6.4 & 47.0 \end{pmatrix}$
Your guesses

- Zach: $\delta = 0.93, \Sigma = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 20 \end{pmatrix}$
- Wei: $\delta = 2.5, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Michael: $\delta = 1.2, \Sigma = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.7 \end{pmatrix}$
- Kristen: $\delta = 1.0, \Sigma = \begin{pmatrix} 2.0 & 0.5 \\ 0 & 2.3 \end{pmatrix}$
- Gustavo: $\delta = 0.5, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Vince: $\delta = 1.05, \Sigma = \begin{pmatrix} 2.0 & 6.4 \\ 6.4 & 47.0 \end{pmatrix}$
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$.
    That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$.
    From a glance at the posterior distribution,
    $\text{sd}(\alpha|y) \approx 1, \text{sd}(\beta|y) \approx 3, \text{corr}(\alpha, \beta|y) \approx 0.7$.
    So try $\Sigma_{\text{jump}} \approx 1.7^2 \left( \begin{array}{ccc} 1^2 & 1 \cdot 3 \cdot 0.7 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 & 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3 \cdot 0.7 & 0.7^2 \end{array} \right) = \left( \begin{array}{ccc} 3 & 6 & 26 \\ 6 & 26 \end{array} \right)$.
- Guess at approx optimal shift:
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1$, $sd(\beta|y) \approx 3$, $corr(\alpha, \beta|y) \approx 0.7$

  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1 & 3 \\ 3 & 0.7 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$

- Guess at approx optimal shift:
My guess

- Look at posterior distribution

- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1$, $sd(\beta|y) \approx 3$, $corr(\alpha, \beta|y) \approx 0.7$
  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$
  - Guess at approx optimal shift:
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1, \; sd(\beta|y) \approx 3, \; corr(\alpha, \beta|y) \approx 0.7$
  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$
- Guess at approx optimal shift:
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1$, $sd(\beta|y) \approx 3$, $corr(\alpha, \beta|y) \approx 0.7$
  - So try $\Sigma_{jump} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$
- Guess at approx optimal shift:
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    \[ sd(\alpha|y) \approx 1, \quad sd(\beta|y) \approx 3, \quad corr(\alpha, \beta|y) \approx 0.7 \]
  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$
- Guess at approx optimal shift:
  - From 1 sd to the center
    - Guess $\delta \approx 3$
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1$, $sd(\beta|y) \approx 3$, $corr(\alpha, \beta|y) \approx 0.7$
  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$

- Guess at approx optimal shift:
  - From 1 sd to the center
  - Guess $\delta \approx 3$
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1, \ sd(\beta|y) \approx 3, \ corr(\alpha, \beta|y) \approx 0.7$
  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$
- Guess at approx optimal shift:
  - From 1 sd to the center
  - Guess $\delta \approx 3$
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1, \ sd(\beta|y) \approx 3, \ corr(\alpha, \beta|y) \approx 0.7$
  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$
- Guess at approx optimal shift:
  - From 1 sd to the center
  - Guess $\delta \approx 3$
My guess

- Look at posterior distribution
- (Approx) optimal Metropolis:
  - Take the posterior covariance matrix and multiply it by $2.4^2/d$
  - That is, scale the posterior ellipse by $2.4/\sqrt{d} = 2.4/\sqrt{2} = 1.7$
  - From a glance at the posterior distribution,
    $sd(\alpha|y) \approx 1$, $sd(\beta|y) \approx 3$, $corr(\alpha, \beta|y) \approx 0.7$
  - So try $\Sigma_{\text{jump}} \approx 1.7^2 \begin{pmatrix} 1^2 & 1 \cdot 3 \cdot 0.7 \\ 1 \cdot 3 \cdot 0.7 & 3^2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 26 \end{pmatrix}$
- Guess at approx optimal shift:
  - From 1 sd to the center
  - Guess $\delta \approx 3$
Building, understanding, and checking the model

- Binomial model for \(#\text{deaths given } \#\text{rats}\)
- Logistic model for \(\Pr(\text{death})\)
- Prior distribution for the logistic regression coefficients
- Discuss extensions to the model
Building, understanding, and checking the model

- Binomial model for \#deaths given \#rats
- Logistic model for Pr(death)
- Prior distribution for the logistic regression coefficients
- Discuss extensions to the model
Building, understanding, and checking the model

- Binomial model for \#deaths given \#rats
- Logistic model for Pr(death)
  - Prior distribution for the logistic regression coefficients
  - Discuss extensions to the model
Building, understanding, and checking the model

- Binomial model for \#deaths given \#rats
- Logistic model for $Pr(\text{death})$
- Prior distribution for the logistic regression coefficients
- Discuss extensions to the model
Building, understanding, and checking the model

- Binomial model for \#deaths given \#rats
- Logistic model for \Pr(\text{death})
- Prior distribution for the logistic regression coefficients
- Discuss extensions to the model
Weakly informative priors

- Logistic regression coefficients
- Variance parameters in hierarchical models
- Mixture models
- Toxicology example
- General principles and theory
Weakly informative priors

- Logistic regression coefficients
- Variance parameters in hierarchical models
- Mixture models
- Toxicology example
- General principles and theory
Weakly informative priors

- Logistic regression coefficients
- Variance parameters in hierarchical models
- Mixture models
- Toxicology example
- General principles and theory
Weakly informative priors

- Logistic regression coefficients
- Variance parameters in hierarchical models
- Mixture models
- Toxicology example
- General principles and theory
Weakly informative priors

- Logistic regression coefficients
- Variance parameters in hierarchical models
- Mixture models
- Toxicology example
- General principles and theory
Weakly informative priors

- Logistic regression coefficients
- Variance parameters in hierarchical models
- Mixture models
- Toxicology example
- General principles and theory
For next week’s class

- Homework 6 due 5pm Tues
- All course material is at http://www.stat.columbia.edu/~gelman/bayescomputation
- Next class:
For next week’s class

- Homework 6 due 5pm Tues
- All course material is at http://www.stat.columbia.edu/~gelman/bayescomputation
- Next class:
  - A student presents slice sampling
  - A student presents multilevel regression and poststratification
For next week’s class

- Homework 6 due 5pm Tues
- All course material is at http://www.stat.columbia.edu/~gelman/bayescomputation
- Next class:
  - A student presents slice sampling
  - A student presents multilevel regression and poststratification
For next week’s class

- Homework 6 due 5pm Tues
- All course material is at http://www.stat.columbia.edu/~gelman/bayescomputation
- Next class:
  - A student presents slice sampling
  - A student presents multilevel regression and poststratification
For next week’s class

- Homework 6 due 5pm Tues
- All course material is at http://www.stat.columbia.edu/~gelman/bayescomputation
- Next class:
  - A student presents slice sampling
  - A student presents multilevel regression and poststratification
For next week’s class

- Homework 6 due 5pm Tues
- All course material is at http://www.stat.columbia.edu/~gelman/bayescomputation
- Next class:
  - A student presents slice sampling
  - A student presents multilevel regression and poststratification