Astrostatistics: Sat 24 Feb 2017

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics

- Example Class 2: Friday, Mar 2, **2:00pm**, MR 5
- Fitting Statistical Models to Astronomical Data
  - Markov Chain Monte Carlo
  - Refs: Not much in Ivezic, Ch 5, F&B Ch 3
  - Gelman et al. Bayesian Data Analysis (Ch 11 & 12)
  - Givens & Hoeting. “Computational Statistics” (Ch 7 & 8)
  - Roberts & Casella. “Monte Carlo Statistical Methods” (theory) (Ch 6 & 7)
  - Hogg & DFM, 2017 “Data analysis recipes: Using Markov Chain Monte Carlo.” https://arxiv.org/abs/1710.06068

Today

- Metropolis: Tuning the Proposal Scale
- Metropolis-Hastings algorithm
- Gibbs Sampling
- Metropolis-within-Gibbs
- Theoretical Justifications

What if you can't directly sample the posterior: $\theta_i \sim P(\theta \mid D)$?

$$\mathbb{E}[f(\boldsymbol{\theta})|\, D] = \int f(\boldsymbol{\theta}) P(\boldsymbol{\theta}|\, D)\, d\boldsymbol{\theta} \approx \frac{1}{m}\sum_{i=1}^{m} f(\boldsymbol{\theta}_i)$$

- Posterior simulation - Markov Chain Monte Carlo:

- Generate a correlated sequence (chain) of random variates (Monte Carlo) that (in a long run limit) are draws from the posterior distribution. The next value in the sequence only depends on the current values (Markov).

- Many degrees of freedom for user: choose to most efficiently generate *independent* samples

d-dim Metropolis Algorithm: Posterior $P(\theta|D)$, Symmetric Proposal/Jump dist'n $J(\theta^*|\theta) = J(\theta|\theta^*)$

1. Choose a random starting point θ_0

2. At step i = 1…N, propose a new parameter value $\theta^* \sim N(\theta_{i-1}, \mathbf{\Sigma_p})$. The proposal distr'n is $J(\theta^*|\theta_{i-1}) = N(\theta^*|\theta_{i-1}, \mathbf{\Sigma_p})$

3. Evaluate ratio of posteriors at proposed vs current values. $r = P(\theta^* \mid \mathbf{y}) / P(\theta_{i-1} \mid \mathbf{y})$.

4. Accept θ^* with probability min(r,1): $\theta_i = \theta^*$. If not accept, stay at same value $\theta_i = \theta_{i-1}$ for the next step & include in chain.

5. Repeat steps 2-4 until reach some measure of convergence (G-R) and gather enough independent samples to compute your inference (reduce Monte Carlo error)

Tuning d-dim Metropolis

- $\theta^* \sim N(\theta_i, \mathbf{\Sigma_p})$: if proposal scale $\mathbf{\Sigma_p}$ is too large, will get too many rejections and not go anywhere. If proposal scale too small, you will accept very many small moves: inefficient random walk

- Laplace Approximation: $$P(\boldsymbol{\theta}|\boldsymbol{D}) \approx N(\boldsymbol{\theta}|\,\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$$
  $$\hat{\boldsymbol{\theta}} = \text{ posterior mode} \qquad (\boldsymbol{\Sigma}^{-1})_{ij} = \frac{\partial^2 \log P(\boldsymbol{\theta}|\boldsymbol{D})}{\partial\theta_i\,\partial\theta_j}\bigg|_{\hat{\boldsymbol{\theta}}}$$

- Choose $\mathbf{\Sigma_p} = c^2\,\mathbf{\Sigma}$: $$c \approx 2.4/\sqrt{d}$$

- Aim for an acceptance ratio of 44% in 1D, 23% in d > 5

Metropolis-Hastings Algorithm: More General Jumping Rule: $J(\theta^*|\theta_i)$ (Need not be symmetric)

1. Choose a random starting point θ_0

2. At step i = 1…N, propose a new parameter value: $\theta^* \sim J(\theta^*|\theta_{i-1})$

3. Evaluate ratio of posteriors at proposed vs current values.
   $r = [P(\theta^* \mid \mathbf{y}) / J(\theta^*|\theta_{i-1})\ ] / [P(\theta_{i-1} \mid \mathbf{y}) / J(\theta_{i-1}|\theta^*)\ ]$

4. Accept θ^* with probability min(r,1): $\theta_i = \theta^*$. If not accept, stay at same value $\theta_i = \theta_{i-1}$ & include in chain.

5. Repeat steps 2-4 until reach some measure of convergence and gather enough samples to compute your inference

Gibbs Sampling

- Multi-dimensional sampling, when you can utilise the set of conditional posterior distributions.
- If joint posterior is $P(\theta, \phi | \boldsymbol{D})$
- And you can solve for tractable conditionals:

$$P(\theta | \phi, \boldsymbol{D})$$

$$P(\phi | \theta, \boldsymbol{D})$$

- Jump along one parameter-dimension at a time

d-dim Gibbs Sampler

$$\boldsymbol{\theta} = (\theta_1 \ldots \theta_d) \qquad \boldsymbol{\theta}_{-j} \equiv (\theta_1, \ldots \theta_{j-1}, \theta_{j+1} \ldots \theta_d)$$

1. Choose a random starting point θ_0

2. At step i = 1…N, cycle through the d-parameters:
   For each j = 1… d, move jth parameter to $\theta_j \sim P(\theta_j | \boldsymbol{\theta}^{i-1}_{-j}, \boldsymbol{D})$
   and always accept.

   The proposal distr'n is $J(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{i-1}) = P(\theta_j | \, \boldsymbol{\theta}_{-j}, \boldsymbol{D})$

3. The Metropolis-Hastings ratio is always 1 (don't need to compute)

4. Always accept θ_j^*.

5. Repeat steps 2-4 until reach some measure of convergence (G-R) and gather enough independent samples to compute your inference (reduce Monte Carlo error)

Gibbs Sampling: Example (Gelman BDA Section 11.1)

Likelihood: $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$ ρ known

Priors: $P(\theta_1) = P(\theta_2) \propto 1$

Posterior: $\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \Big| \boldsymbol{y} \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$

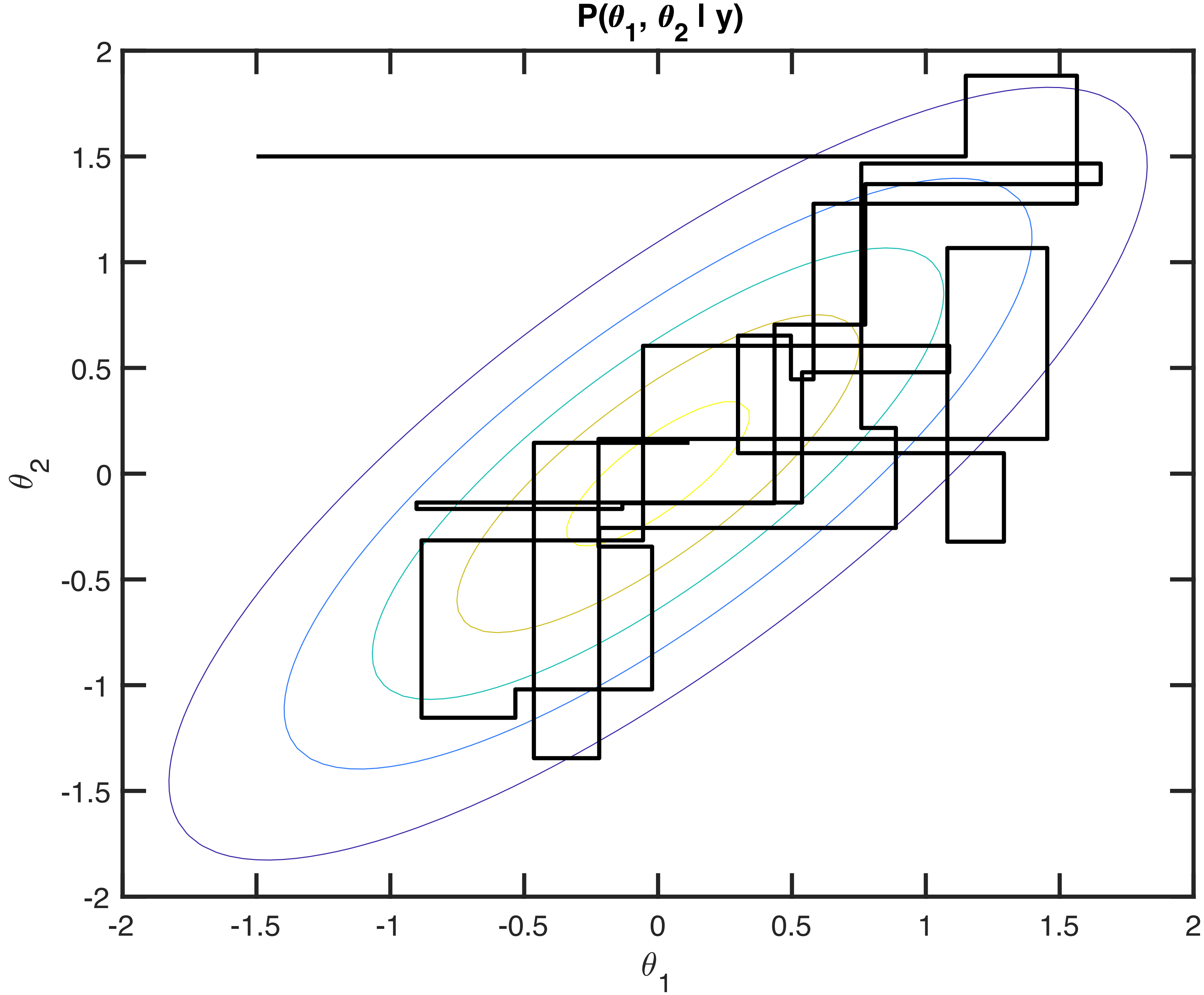
$$P(\boldsymbol{\theta}|y) = P(\theta_1|\theta_2, y)P(\theta_2|y) = P(\theta_2|\theta_1, y)P(\theta_1|y)$$

Conditional Posteriors:

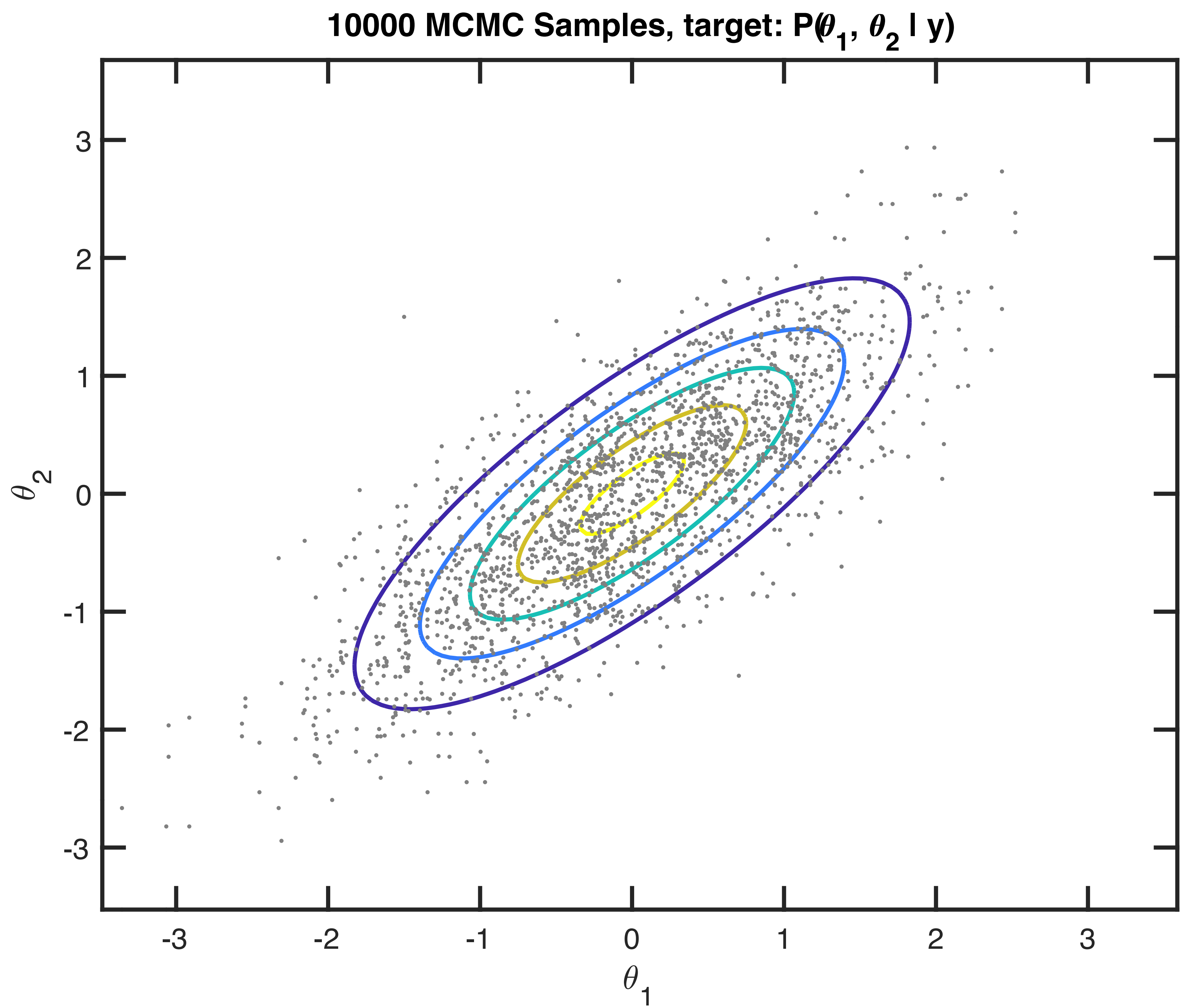
$$\theta_1|\theta_2, \boldsymbol{y} \sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$

$$\theta_2|\theta_1, \boldsymbol{y} \sim N(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2)$$

Gibbs Sampling: demo
gibbs_example.m
2D Trace Paths for 50 iterations



Gibbs Sampling: demo
gibbs_example.m
Joint Posterior Densities



Gibbs Sampling: demo
gibbs_example.m
Marginal Posterior Densities