Bayesian Statistics

Markov Chain Monte Carlo: Gibbs sampler

Nan Lin
Department of Mathematics
Washington University in St. Louis

Posterior inference by simulation

- Approach 1: Independence sampling
  - Simulate independent samples from the posterior distributions
  - Draw $\theta^{(1)}, \ldots, \theta^{(M)}$ i.i.d. from the posterior distribution $f(\theta|y)$
  - This is what we have been doing
- Approach 2: Markov Chain Monte Carlo (MCMC)
  - Draw $\theta^{(i+1)}$ from $g\left(\theta^{(i+1)}\middle|\theta^{(i)}\right)$ such that

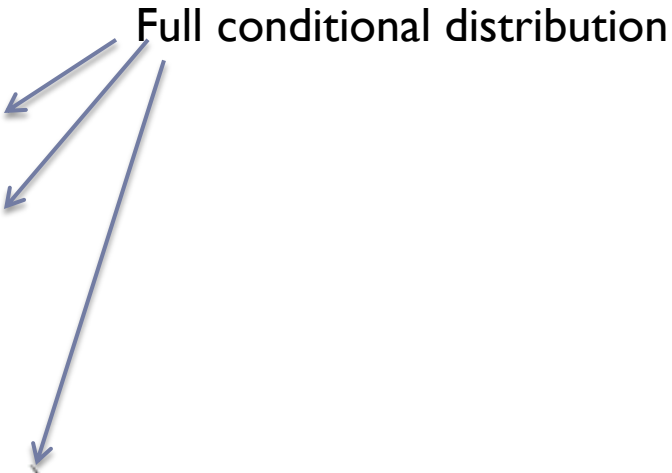
$$f\left(\theta^{(i+1)}\right) \to f(\theta|y)$$

  - Gibbs sampling
  - Metropolis-Hastings

Gibbs sampler

- Used for multiparameter models
- Parameter: $\theta = (\theta_1, \ldots, \theta_k)$

- An iterative algorithm
  - draw θ_1 from $p(\theta_1|\theta_2, \theta_3, \ldots, \theta_k, y)$
  - draw θ_2 from $p(\theta_2|\theta_1, \theta_3, \ldots, \theta_k, y)$
  - $\cdots$
  - draw θ_k from $p(\theta_k|\theta_1, \theta_2, \ldots, \theta_{k-1}, y)$



Gibbs sampler (cont)

- Full conditional distribution $p(\theta_j|\theta_{-j}, y)$
  - Where $\theta_{-j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)$
- In iteration t, draw $\theta_j^t \sim p(\theta_j|\theta_{-j}^t, y)$,
  - where $\theta_{-j}^t = (\theta_1^t, \ldots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \ldots, \theta_k^t)$
  - Each θ_j is updated conditional on the "latest" values of θ

Example: Simulate from a bivariate normal distribution

- Joint distribution

$$\mathbf{Z} = (X, Y)' \sim N(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix})$$

- Full conditional distribution

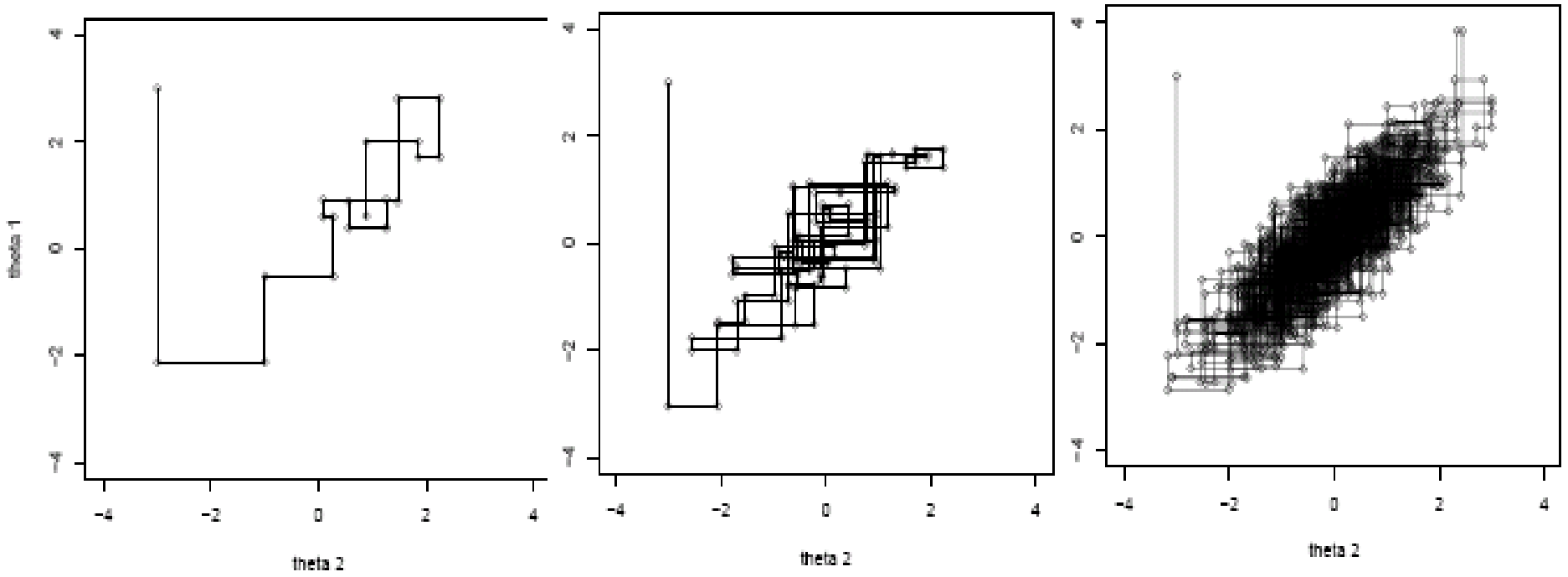
$$X|Y = y \sim N(\rho y, 1-\rho^2) \sim \rho y + \sqrt{1-\rho^2} N(0,1),$$

$$Y|X = x \sim N(\rho x, 1-\rho^2) \sim \rho x + \sqrt{1-\rho^2} N(0,1).$$

R code

```
x=y=c()
x[1]=0
y[1]=0
rho=0.9
c=sqrt(1-rho*rho)
for (i in 2:6000){
        x[i]=rho*y[i-1]+ c*rnorm(1)
        y[i]=rho*x[i]+ c*rnorm(1)
}

par(mfrow=c(2,1))
plot(x[1:50],y[1:50]);
plot(x[4000:6000],y[4000:6000],pch='.');
```


theta 1
theta 2
theta 2
theta 2

Example: a normal model with a semi-conjugate prior

- Recall our previous discussion on normal model with a conjugate prior
- Data: $y_i|\theta,\sigma^2 \sim N(\theta,\sigma^2)$ i.i.d.
  - θ,σ^2 are both unknown
- Conjugate prior
  - $\theta|\sigma^2 \sim N(\mu_0,\frac{\sigma^2}{\kappa_0})$
  - $\sigma^2 \sim Inv-\chi^2(\nu_0,\sigma_0^2)$
- Posterior distribution
  - $\theta|\sigma^2,y \sim N(\mu_n,\frac{\sigma_n^2}{\kappa_n})$
  - $\sigma^2|y \sim Inv-\chi^2(\nu_n,\sigma_n^2)$
    - $\mu_n=\frac{\kappa_0}{\kappa_0+n}\mu_0+\frac{n}{\kappa_0+n}\bar{y}$
    - $\kappa_n=\kappa_0+n$
    - $\nu_n=\nu_0+n$
    - $\nu_n\sigma_n^2=\nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2$

Example: a normal model with a semi-conjugate prior

- This conjugate prior distribution relates the prior variance of θ to the sampling variance of our data in such a way that μ_0 can be thought of as κ_0 prior samples from the population. In some situations this makes sense, but in others we may want to specify our uncertainty about as being independent of σ^2, so that $p(\theta, \sigma^2) = p(\theta) \times p(\sigma^2)$
  - $\theta|\sigma^2 \sim N(\mu_0, \tau_0^2)$
  - $\sigma^2 \sim Inv - \chi^2(\nu_0, \sigma_0^2)$
- We knew that
  - $\theta|\sigma^2, y \sim N(\mu_n, \tau_n^2)$

$$\mu_n = \frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau_0^2 + n/\sigma^2} \quad \text{and} \quad \tau_n^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

Example: a normal model with a semi-conjugate prior

- Let $\tilde{\sigma}^2 = 1/\sigma^2$ be the precision parameter

$$\begin{aligned} p(\tilde{\sigma}^2|\theta, y_1, \ldots, y_n) &\propto p(y_1, \ldots, y_n, \theta, \tilde{\sigma}^2) \\ &= p(y_1, \ldots, y_n|\theta, \tilde{\sigma}^2)p(\theta|\tilde{\sigma}^2)p(\tilde{\sigma}^2)\,. \end{aligned}$$

- Under prior independence, $p(\theta|\tilde{\sigma}^2) = p(\theta)$

$$\begin{aligned} p(\tilde{\sigma}^2|\theta, y_1, \ldots, y_n) &\propto p(y_1, \ldots, y_n|\theta, \tilde{\sigma}^2)p(\tilde{\sigma}^2) \\ &\propto \left((\tilde{\sigma}^2)^{n/2} \exp\{-\tilde{\sigma}^2 \sum_{i=1}^{n}(y_i - \theta)^2/2\}\right) \times \\ &\quad \left((\tilde{\sigma}^2)^{\nu_0/2-1} \exp\{-\tilde{\sigma}^2\nu_0\sigma_0^2/2\}\right) \\ &= (\tilde{\sigma}^2)^{(\nu_0+n)/2-1} \times \exp\{-\tilde{\sigma}^2 \times [\nu_0\sigma_0^2 + \sum(y_i - \theta)^2]/2\}. \end{aligned}$$

- $\sigma^2|\theta, y_1, \ldots, y_n \sim Inv - gamma(\frac{\nu_n}{2}, \frac{\nu_n\sigma_n^2(\theta)}{2})$
  - i.e. $Inv - \chi^2(\nu_n, \sigma_n^2(\theta))$

$$\nu_n = \nu_0 + n\,, \quad \sigma_n^2(\theta) = \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + ns_n^2(\theta)\right], \text{ and } s_n^2(\theta) = \sum(y_i - \theta)^2/n$$

Example: a normal model with a semi-conjugate prior

- Gibbs sampler
  - Given a current state of the parameters $\phi^{(s)} = \{\theta^{(s)}, \tilde{\sigma}^{2(s)}\}$, generate a new state as follows

    1. sample $\theta^{(s+1)} \sim p(\theta|\tilde{\sigma}^{2(s)}, y_1, \ldots, y_n)$;
    2. sample $\tilde{\sigma}^{2(s+1)} \sim p(\tilde{\sigma}^2|\theta^{(s+1)}, y_1, \ldots, y_n)$;
    3. let $\phi^{(s+1)} = \{\theta^{(s+1)}, \tilde{\sigma}^{2(s+1)}\}$.

  - Output: a dependent sequence $\{\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(S)}\}$.

```
### data
mean.y<-mean(y) ; var.y<-var(y) ; n<-length(y)
###

### starting values
S<-1000
PHI<-matrix(nrow=S,ncol=2)
PHI[1,]<-phi<-c( mean.y, 1/var.y)
###

### Gibbs sampling
set.seed(1)
for(s in 2:S) {

# generate a new theta value from its full conditional
mun<- ( mu0/t20 + n*mean.y*phi[2] ) / ( 1/t20 + n*phi[2] )
t2n<- 1/( 1/t20 + n*phi[2] )
phi[1]<-rnorm(1, mun, sqrt(t2n) )

# generate a new 1/sigma^2 value from its full conditional
nun<- nu0+n
s2n<- (nu0*s20 + (n-1)*var.y + n*(mean.y-phi[1])^2 ) /nun
phi[2]<- rgamma(1, nun/2, nun*s2n/2)

PHI[s,]<-phi          }
###
```

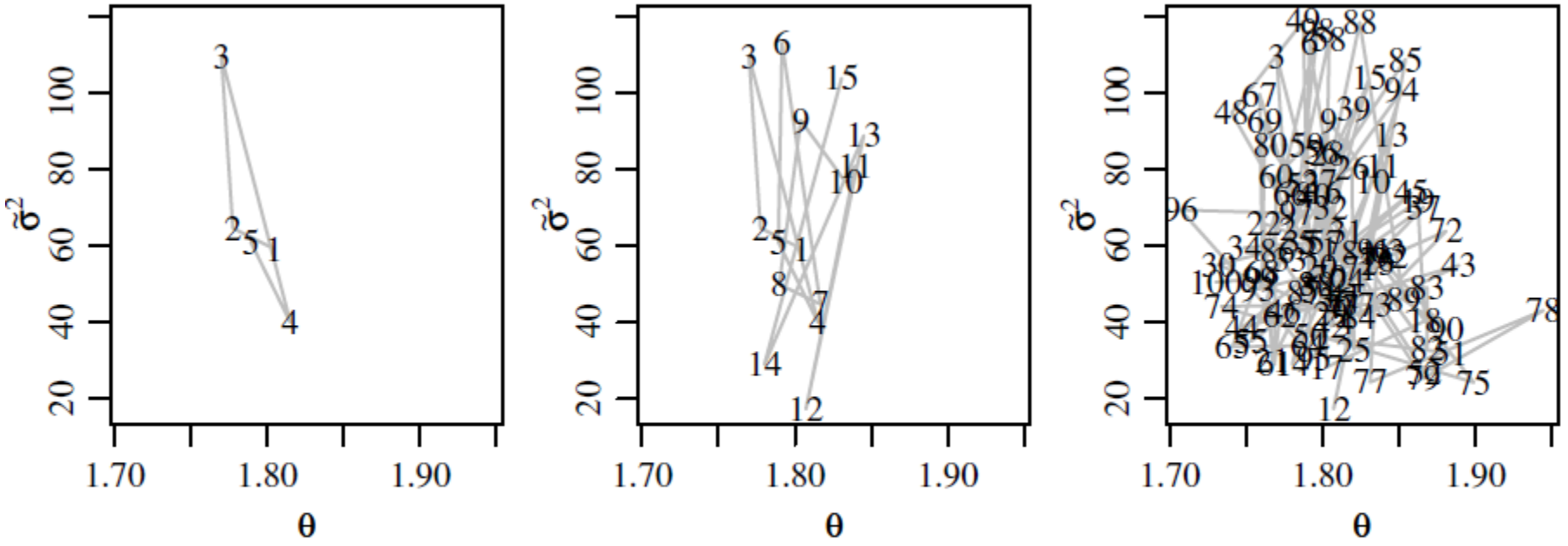


Fig. 6.2. The first 5, 15 and 100 iterations of a Gibbs sampler.

Posterior credible interval

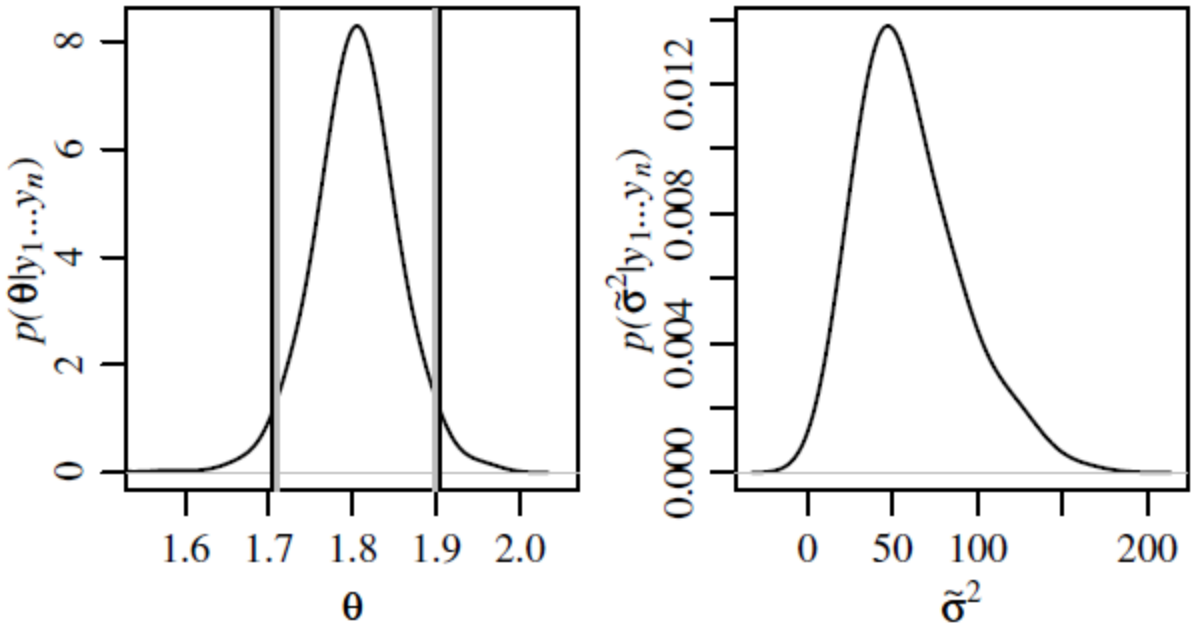
```
### CI for population mean
> quantile(PHI[,1],c(.025,.5,.975))
    2.5%      50%    97.5%
1.707282 1.804348 1.901129

### CI for population precision
> quantile(PHI[,2],c(.025,.5, .975))
     2.5%       50%     97.5%
 17.48020  53.62511 129.20020


### CI for population standard deviation
> quantile(1/sqrt(PHI[,2]),c(.025,.5, .975))
      2.5%        50%       97.5%
0.08797701 0.13655763 0.23918408
```

Marginal posterior density

- Nonparametric density estimate based on the simulated values from the Gibbs sampler



General property of Gibbs sampler

- Parameters $\phi = \{\phi_1, \ldots, \phi_p\}$
- Target distribution $p(\phi) = p(\phi_1, \ldots, \phi_p)$.
- In the previous example, $\phi = \{\theta, \sigma^2\}$

$$p(\phi) = p(\theta, \sigma^2 | y_1, \ldots, y_n).$$

- Gibbs sampling

1. sample $\phi_1^{(s)} \sim p(\phi_1 | \phi_2^{(s-1)}, \phi_3^{(s-1)}, \ldots, \phi_p^{(s-1)})$
2. sample $\phi_2^{(s)} \sim p(\phi_2 | \phi_1^{(s)}, \phi_3^{(s-1)}, \ldots, \phi_p^{(s-1)})$

$\vdots$

p. sample $\phi_p^{(s)} \sim p(\phi_p | \phi_1^{(s)}, \phi_2^{(s)}, \ldots, \phi_{p-1}^{(s)})$.

General property of Gibbs sampler

- Output: a dependent sequence

$$\begin{aligned}\phi^{(1)} &= \{\phi_1^{(1)}, \ldots, \phi_p^{(1)}\} \\ \phi^{(2)} &= \{\phi_1^{(2)}, \ldots, \phi_p^{(2)}\} \\ &\vdots \\ \phi^{(S)} &= \{\phi_1^{(S)}, \ldots, \phi_p^{(S)}\}.\end{aligned}$$

  - $\phi^{(s)}$ depends on $\phi^{(0)}, \ldots, \phi^{(s-1)}$ only through $\phi^{(s-1)}$
  - $\phi^{(s)}$ is conditionally independent of $\phi^{(0)}, \ldots, \phi^{(s-2)}$ given $\phi^{(s-1)}$
- This called a Markov property, and the sequence is called a *Markov chain*.
- For the models we discuss in this class, the sampling distribution of $\phi^{(s)}$ approaches the target distribution as $s \to \infty$, no matter what the starting value $\phi^{(0)}$ is.

$$\Pr(\phi^{(s)} \in A) \to \int_A p(\phi)\, d\phi \quad \text{as } s \to \infty.$$

General property of Gibbs sampler

- More importantly, for most functions g of interest,

$$\frac{1}{S}\sum_{s=1}^{S} g(\phi^{(s)}) \rightarrow \mathrm{E}[g(\phi)] = \int g(\phi)p(\phi)\, d\phi \quad \text{as } S \rightarrow \infty.$$

  - One can approximate $E[g(\phi)]$ with the sample average of $\{g(\phi^{(1)}), \ldots, g(\phi^{(S)})\}$. This is the *Monte Carlo* part.
- Hence, we call this type method Markov chain Monte Carlo (MCMC) method.
- In the previous example, based on 1000 simulated values from the Gibbs sampler, we have the following approximation

$$\mathrm{E}[\theta|y_1, \ldots, y_n] \approx \frac{1}{1000}\sum_{s=1}^{1000} \theta^{(s)} = 1.804, \text{ and}$$

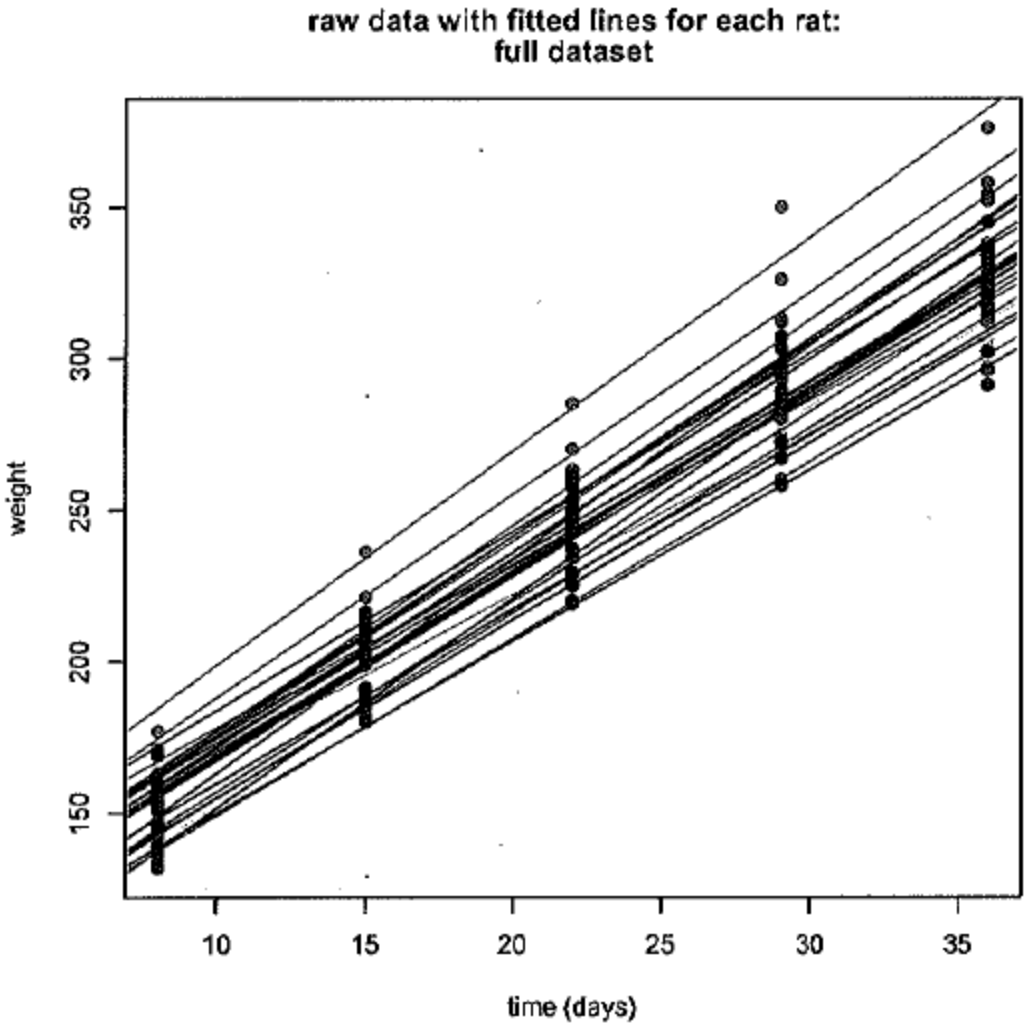
$$\Pr(\theta \in [1.71, 1.90]|y_1, \ldots, y_n) \approx 0.95.$$

Example: a longitudinal data growth curve model

Rat Population growth data (Gelfand et al. 1990, JASA 1990)

	Weights Y_{ij} of rat i on day x_{ij}				
	$x_1 = 8$	$x_2 = 15$	$x3 = 22$	$x_4 = 29$	$x_5 = 36$
rat 1	151	199	246	283	320
rat 2	145	199	249	293	354
...					
rat 30	153	200	244	286	324

- Y_{ij} weight of the ith rat at measurement point j
- x_{ij} denotes the rat's age in days at time point j


raw data with fitted lines for each rat:
full dataset
weight
150
200
250
300
350
10
15
20
25
30
35
time (days)

Model

- **Stage I: Sampling Distribution**

$$Y_{ij} \sim N(\alpha_i + \beta_i x_{ij}, \sigma^2),\ i = 1, \ldots, k = 30\ j = 1, \ldots, n = 5$$

- **Stage II: Prior**

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim N\left(\begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix}, \Sigma \right),\ i = 1, \ldots, k$$

- **Stage III: Hyperprior**

$$\sigma^2 \sim IG(a, b)$$

$$\begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \sim N\left(\begin{bmatrix} \eta_0 \\ \eta_1 \end{bmatrix}, C \right)$$

$$\Sigma^{-1} \sim W((\rho R)^{-1}, \rho),\ E(\Sigma^{-1}) = R^{-1},\ var(\Sigma) \propto \rho^{-1}$$

- we seek the marginal posterior distribution for α_0, β_0 given the observed data and predictive intervals for the individual future growth given the first-week measurement
- the number of unknown parameters is 66: (30 α_is + 30 β_is + $\alpha_0 + \beta_0 + \sigma^2$ + 3 unique component of Σ).
- let's re-write the sampling distribution as:

$$\boldsymbol{y}_i \sim N(X_i\boldsymbol{\theta}_i, \sigma^2 I_n),\ i = 1, \ldots, k = 30\ j = 1, \ldots, n = 5$$

where:

$$\boldsymbol{y}_i^t = (y_{i1}, \ldots, y_{in_i}),\ X_i = \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix},\ \boldsymbol{\theta}_i^t = (\alpha_i, \beta_i)$$

- find the full conditional distributions

$$\begin{aligned}
\boldsymbol{\theta}_i \mid \boldsymbol{y}, \boldsymbol{\theta}_0, \Sigma^{-1}, \sigma^2 &\sim N\left(D_i\left[\sigma^{-2} X_i^t \boldsymbol{y}_i + \Sigma^{-1}\boldsymbol{\theta}_0\right], D_i\right) \\
&\quad \text{ind } i = 1, \ldots, k \\
\boldsymbol{\theta}_0 \mid \boldsymbol{y}, \{\boldsymbol{\theta}_i\}, \Sigma^{-1}, \sigma^2 &\sim N\left(V\left[k\Sigma^{-1}\bar{\boldsymbol{\theta}} + C^{-1}\boldsymbol{\eta}\right], V\right) \\
\Sigma^{-1} \mid \boldsymbol{y}, \{\boldsymbol{\theta}_i\}, \boldsymbol{\theta}_0, \sigma^2 &\sim W\left(\left[\textstyle\sum_{i=1}^k (\boldsymbol{\theta}_i - \boldsymbol{\theta}_0)^t(\boldsymbol{\theta}_i - \boldsymbol{\theta}_0) + \rho R\right]^{-1}, k + \rho\right) \\
\sigma^2 \mid \boldsymbol{y}, \{\boldsymbol{\theta}_i\}, \boldsymbol{\theta}_0, \Sigma &\sim IG\left(\tfrac{kn}{2} + a, \left[\tfrac{1}{2}\textstyle\sum_{i=1}^k (\boldsymbol{y}_i - X_i\boldsymbol{\theta}_i)^t(\boldsymbol{y}_i - X_i\boldsymbol{\theta}_i) + b^{-1}\right]^{-1}\right)
\end{aligned}$$

where

- $D_i^{-1} = \sigma^{-2} X_i^t X_i + \Sigma^{-1}$, $\boldsymbol{\theta}_0^t = (\alpha_0, \beta_0)^t$
- $V = (k\Sigma^{-1} + C^{-1})^{-1}$, $\bar{\boldsymbol{\theta}} = \frac{1}{k}\sum_{i=1}^k \boldsymbol{\theta}_i$

Distinguishing parameter estimation from posterior approximation

- Bayesian data analysis using Monte Carlo methods
  - Data analysis: the statistical part
  - Numerical approximation: the Monte Carlo part
- Ingredients of Bayesian data analysis
  - Model specification
  - Prior specification
  - Posterior summary
- When the posterior distribution is complicated, a useful way to "look at" the posterior distribution is by studying Monte Carlo samples from the posterior distribution

Distinguishing parameter estimation from posterior approximation

- Monte Carlo and MCMC sampling algorithms
  - are not models,
  - they do not generate "more information" than is in y and $p(\phi)$,
  - they are simply "ways of looking at" $p(\phi|y)$.
- For example, if we have Monte Carlo samples $\phi^{(1)}, \ldots, \phi^{(S)}$ from $p(\phi|y)$, then these samples help describe $p(\phi|y)$,

$$\frac{1}{S}\sum \phi^{(s)} \approx \int \phi p(\phi|\boldsymbol{y})\, d\phi$$
$$\frac{1}{S}\sum 1(\phi^{(s)} \leq c) \approx \Pr(\phi \leq c|\boldsymbol{y}) = \int_{-\infty}^{c} p(\phi|\boldsymbol{y})\, d\phi.$$

- "*Estimation*": how we use $p(\phi|y)$ to make inference about ϕ
- "*Approximation*": the use of Monte Carlo procedures to approximate integrals.

Independent MC simulation vs MCMC

- Independent MC sample $\{\phi^{(1)}, \ldots, \phi^{(S)}\}$ from target distribution $p(\phi)$
  - $P\left(\phi^{(s)} \in A\right) = \int_A p(\phi) d\phi$ for any set A and all $s = 1, \ldots, S$
- This property is not true for MCMC samples. What is true is that

$$\lim_{s \to \infty} \Pr(\phi^{(s)} \in A) = \int_A p(\phi)\, d\phi.$$

Example: a three-component normal mixture

- Target distribution: joint distribution of two random variables
  - A discrete random variable $\delta = \{1,2,3\}$
  - A continuous random variable $\theta \in R$
  - $\{P(\delta = 1), P(\delta = 2), P(\delta = 3)\} = \{.45, .10, .45\}$
  - $\theta|\delta \sim N(\mu_\delta, \sigma_\delta^2)$
    - $\{\mu_1, \mu_2, \mu_3\} = (-3,0,3)$
    - $\{\sigma_1^2, \sigma_2^2, \sigma_3^2\} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$
  - Marginal distribution of θ is a three-component normal mixture

Independent MC

Repeat

- Draw δ^* from the marginal distribution of δ
- Draw θ^* from the conditional distribution $N(\mu_{\delta^*}, \sigma^2_{\delta^*})$

- (θ^*, δ^*) represents a sample from the joint distribution $p(\theta, \delta)$

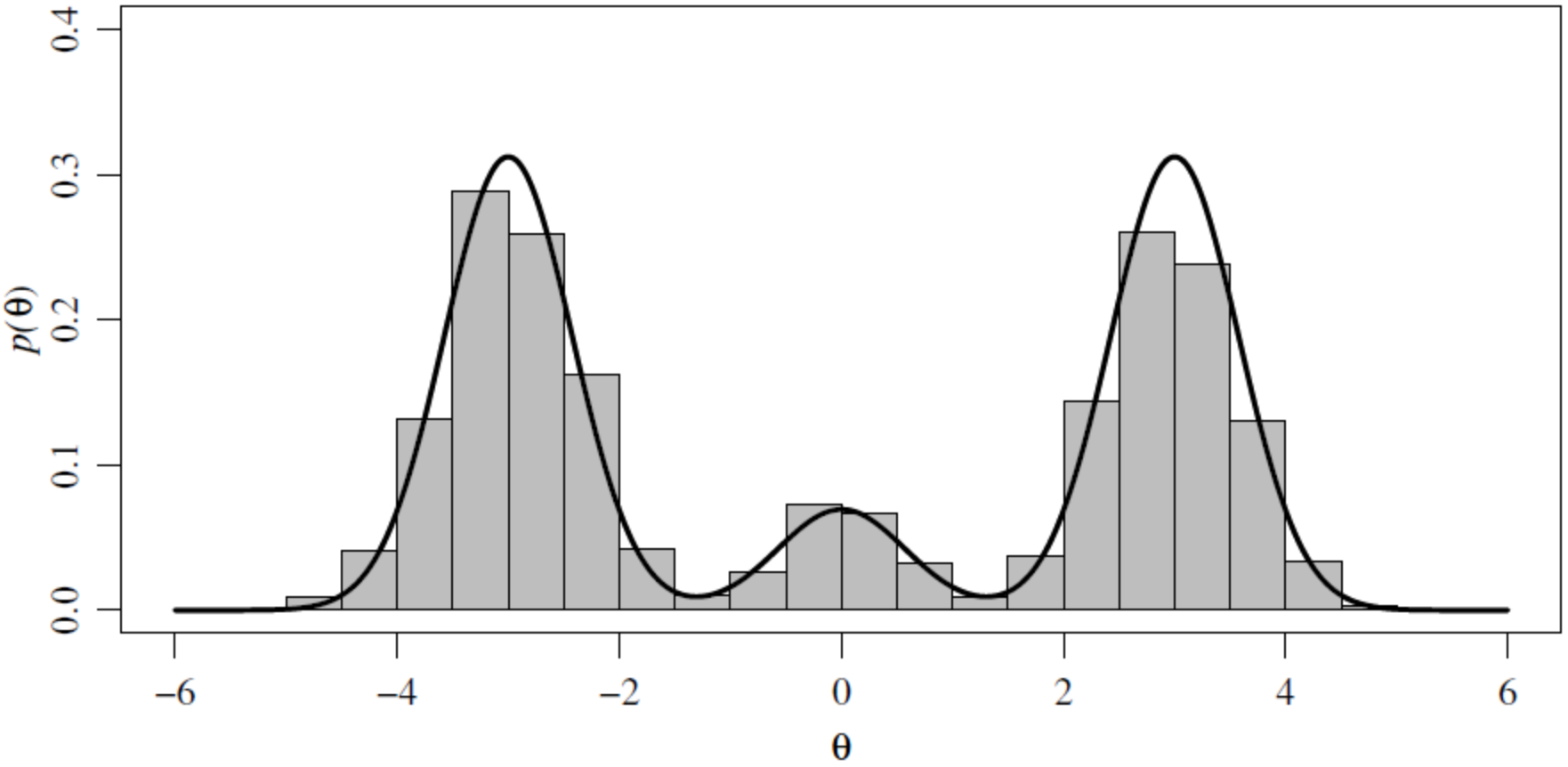


Fig. 6.4. A mixture of normal densities and a Monte Carlo approximation.

Gibbs sampler

- It is given that $\theta|\delta \sim N(\mu_\delta, \sigma_\delta^2)$
- Using Bayes' theorem, it is easy to derive that

$$\Pr(\delta = d|\theta) = \frac{\Pr(\delta = d) \times \text{dnorm}(\theta, \mu_d, \sigma_d)}{\sum_{d=1}^{3} \Pr(\delta = d) \times \text{dnorm}(\theta, \mu_d, \sigma_d)}, \text{ for } d \in \{1, 2, 3\}.$$

- The figure on the next slide shows a histogram of 1,000 MCMC values of θ generated with the Gibbs sampler. Notice that the empirical distribution of the MCMC samples gives a poor approximation to $p(\theta)$.
  - Autocorrelation is high

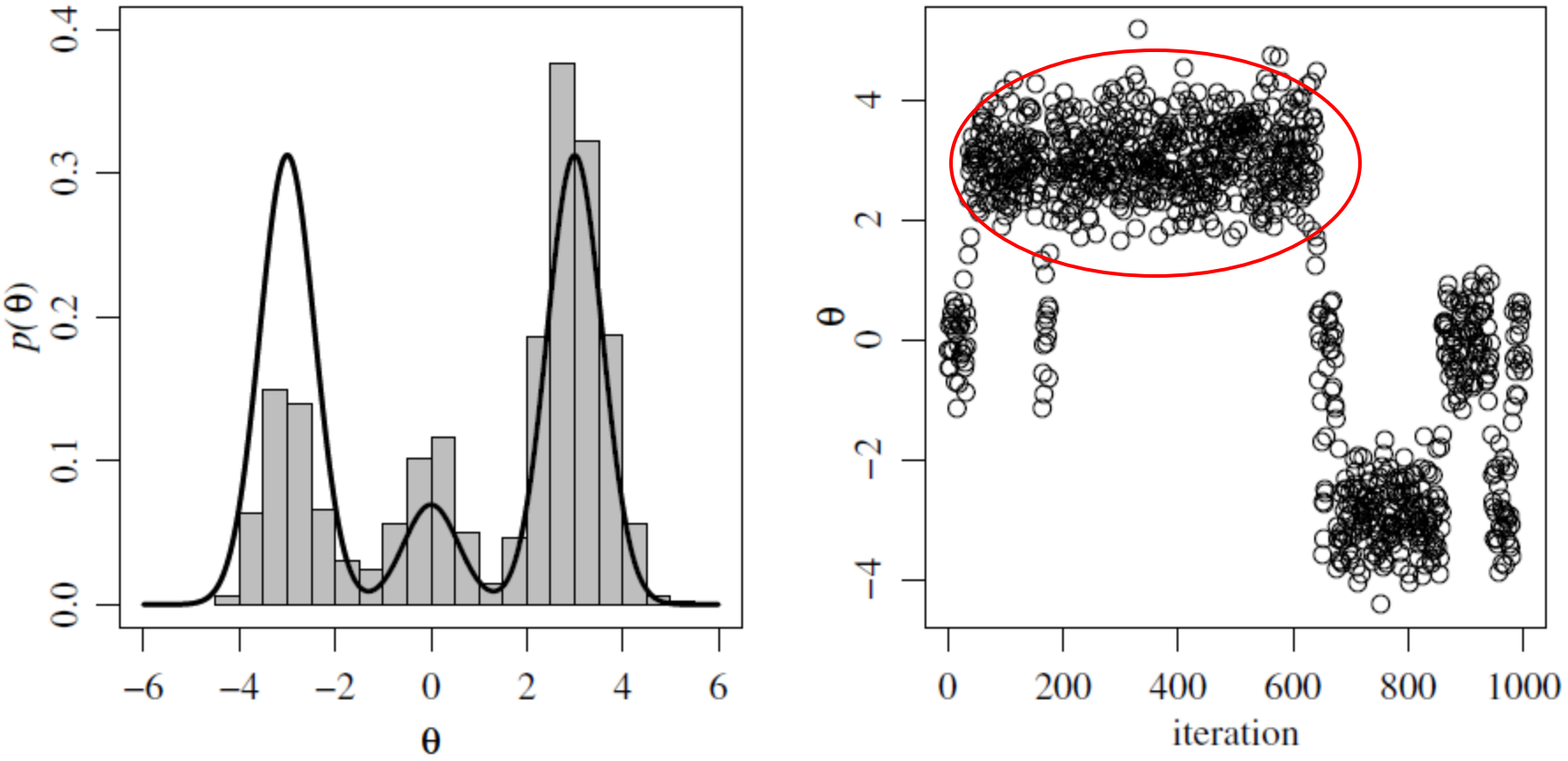


Fig. 6.5. Histogram and traceplot of 1,000 Gibbs samples.

Values of θ near -3 are underrepresented, whereas values near zero and +3 are overrepresented

- The chain got stuck in these regions

A longer simulation

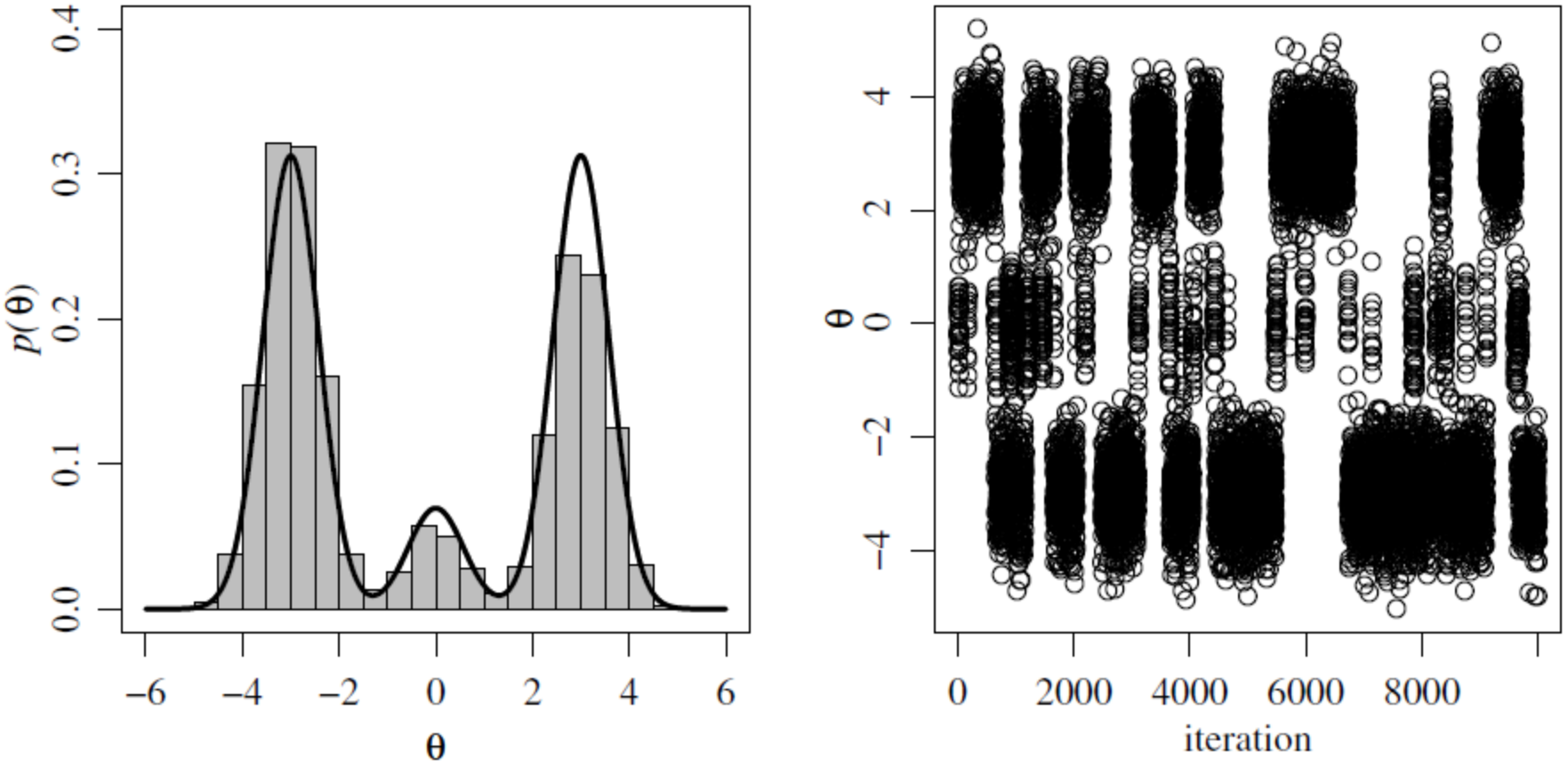


Fig. 6.6. Histogram and traceplot of 10,000 Gibbs samples.

Convergence diagnostics

- The MCMC theory guarantees that the Gibbs sampler "eventually" will provide a good approximation to the target distribution .
- But "eventually" can be a very long time in some situations

- Let's think the simulated MCMC sample $\{\phi^{(1)}, \ldots, \phi^{(S)}\}$ as the trajectory of a particle ϕ moving around the parameter space
  - The amount of time the particle spends in a given set A should be proportional to the target probability $\int_A p(\phi) d\phi$

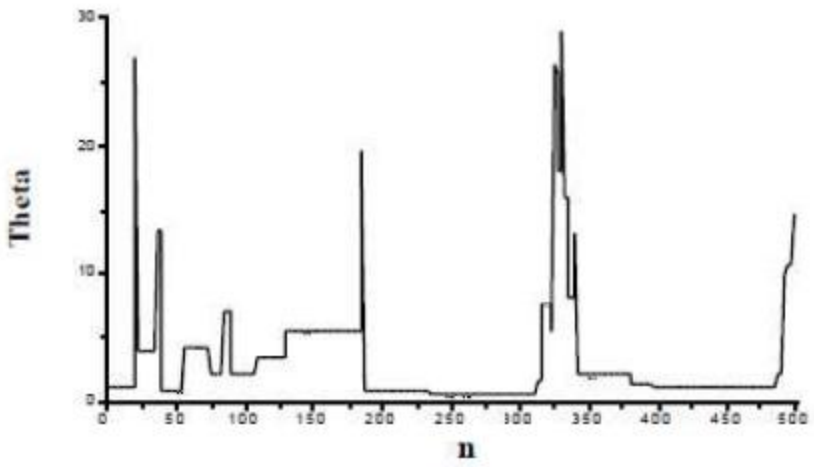
Convergence diagnostics

- Consider three disjoint sets A_1, A_2, A_3
  - For example, regions near the three modes
  - If $P(A_2) < P(A_1) \approx P(A_3)$, the particle should spend little time in A_2, and about the same amount of time in A_1 and A_3
- If we start the chain in A_2, the number of iterations S should be large enough to let the particle have a chance to
  1. move out of A_2 and into higher probability regions
  2. move between A_1 and A_3, and any other sets of high probability.

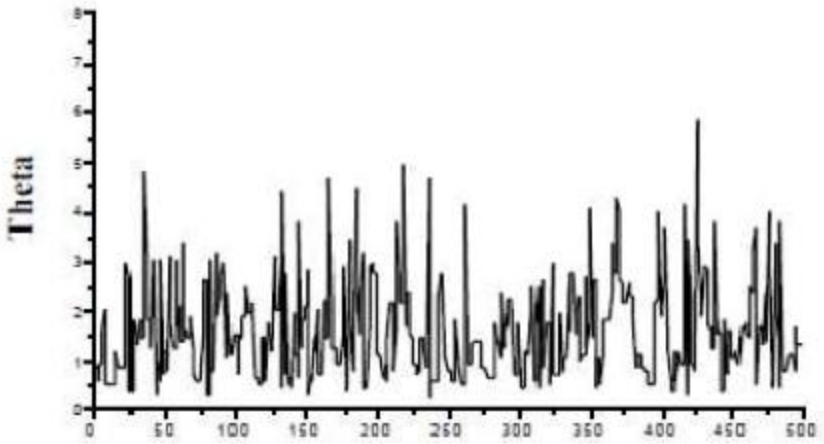
Convergence diagnostics

- Item 1 relates to whether the Markov chain achieved stationarity (*convergence*)
  - Samples taken in one part of the chain have a similar distribution to samples taken in other parts.
  - In general, convergence is faster if the chain is started from a high probability region
- Item 2 relates to how quickly the particle moves around the parameter space (the speed of *mixing*).
  - Independent MC has perfect mixing
    - zero autocorrelation
    - Jump between different regions in one step
  - MCMC may have poor mixing when
    - autocorrelation is high
    - it takes a long time between jumps to different parts of the parameter space

- Poor mixing



- Well mixing



Impact of autocorrelation in MCMC

- Suppose we want to approximate $E[\phi] = \int \phi p(\phi) d\phi$ based on the empirical distribution of $\{\phi^{(1)}, \ldots, \phi^{(S)}\}$ using $\bar{\phi} = S^{-1} \sum_s \phi^{(s)}$
- If using independent MC,

$$\mathrm{Var}_{\mathrm{MC}}[\bar{\phi}] = \mathrm{E}[(\bar{\phi} - \phi_0)^2] = \frac{\mathrm{Var}[\phi]}{S},$$

- If using MCMC,
  - assuming stationarity is achieved

$$\mathrm{Var}_{\mathrm{MCMC}}[\bar{\phi}] = \mathrm{Var}_{\mathrm{MC}}[\bar{\phi}] + \frac{1}{S^2} \sum_{s \neq t} \mathrm{E}[(\phi^{(s)} - \phi_0)(\phi^{(t)} - \phi_0)].$$

The correlation is often positive

$$
\begin{aligned}
\mathrm{Var}_{\mathrm{MCMC}}[\bar{\phi}] &= \mathrm{E}[(\bar{\phi} - \phi_0)^2] \\
&= \mathrm{E}[\{\frac{1}{S}\sum(\phi^{(s)} - \phi_0)\}^2] \\
&= \frac{1}{S^2}\mathrm{E}[\sum_{s=1}^{S}(\phi^{(s)} - \phi_0)^2 + \sum_{s\neq t}(\phi^{(s)} - \phi_0)(\phi^{(t)} - \phi_0)] \\
&= \frac{1}{S^2}\sum_{s=1}^{S}\mathrm{E}[(\phi^{(s)} - \phi_0)^2] + \frac{1}{S^2}\sum_{s\neq t}\mathrm{E}[(\phi^{(s)} - \phi_0)(\phi^{(t)} - \phi_0)] \\
&= \mathrm{Var}_{\mathrm{MC}}[\bar{\phi}] + \frac{1}{S^2}\sum_{s\neq t}\mathrm{E}[(\phi^{(s)} - \phi_0)(\phi^{(t)} - \phi_0)].
\end{aligned}
$$

Assessing autocorrelation

- Lag-t sample autocorrelation function

$$\text{acf}_t(\phi) = \frac{\frac{1}{S-t}\sum_{s=1}^{S-t}(\phi_s - \bar{\phi})(\phi_{s+t} - \bar{\phi})}{\frac{1}{S-1}\sum_{s=1}^{S}(\phi_s - \bar{\phi})^2},$$

  - Use R function `acf()`
  - Lower acf is better
- *Effect sample size*: the number of independent Monte Carlo samples necessary to give the same precision as the MCMC samples

$$\text{Var}_{\text{MCMC}}[\bar{\phi}] = \frac{\text{Var}[\phi]}{S_{\text{eff}}},$$

  - Use the R-command `effectivSize` in the `coda` package

Convergence diagnostics for the semiconjugate normal analysis

- Traceplot

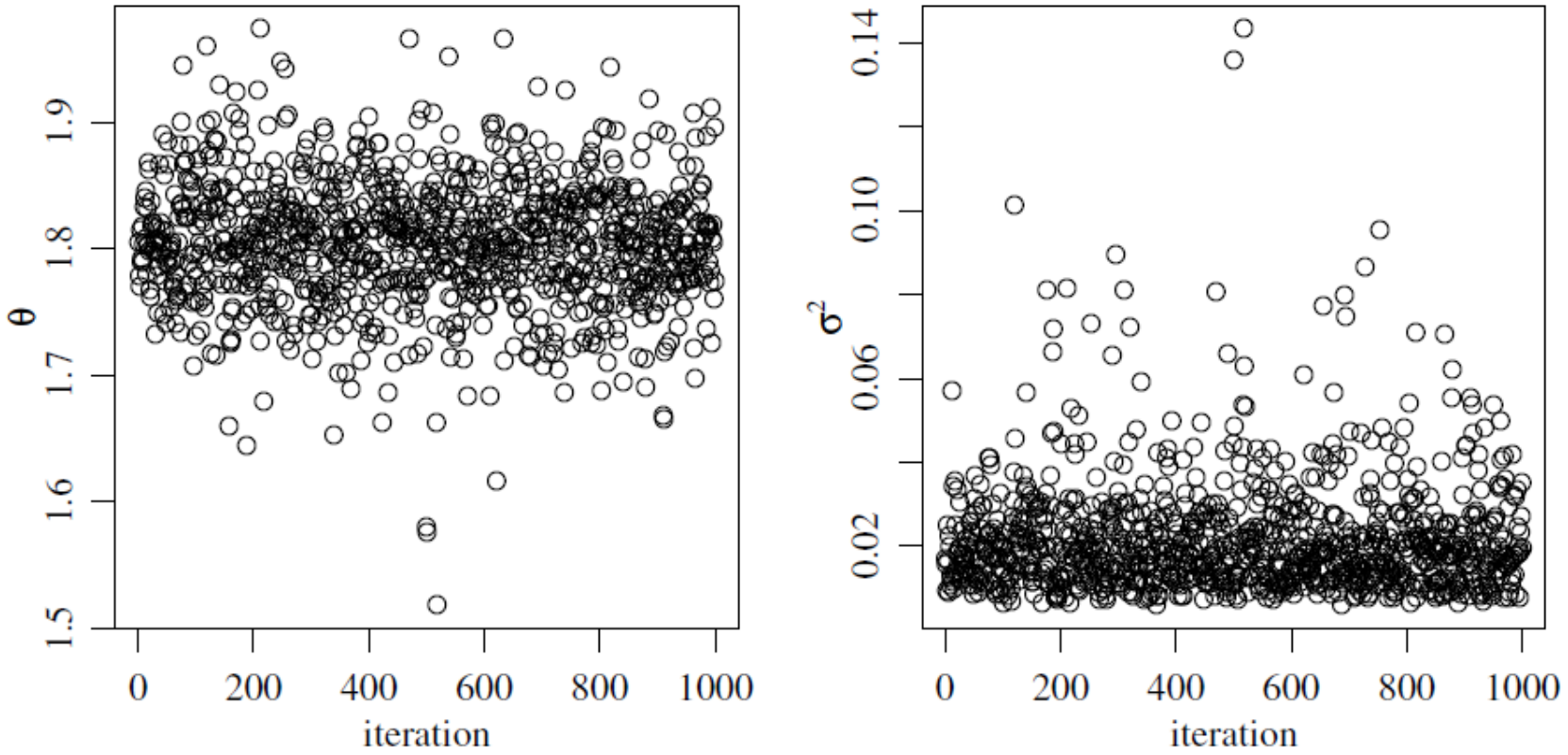


Fig. 6.7. Traceplots for θ and σ^2.

Convergence diagnostics for the semiconjugate normal analysis

- For θ,
  - Lag-1 autocorrelation = 0.031
  - Effective sample size =1000
- For σ^2,
  - Lag-1 autocorrelation = 0.147
  - Effective sample size =742

Some additional issues

- Burn-in
- Reduce autocorrelation by thinning
- Sample size inflation factor (SSIF)
  - Based on an AR(1) model,

$$\theta_t = \mu + \alpha(\theta_{t-1} - \mu) + \epsilon \qquad \epsilon \sim N(0, \sigma^2)$$

  - The standard error of $\bar{\theta} = \frac{1}{n}\sum_{t=1}^{n} \theta_t$ is

$$\mathrm{SE}(\bar{\theta}) = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{1+\rho}{1-\rho}}$$

  - SSIF = $\sqrt{(1+\rho)/(1-\rho)}$,
    - e.g., for $\rho = 0.95$, $\mathrm{SSIF} = 39$ →Roughly forty times as many points are required for the same precision as with an independent sequence