



## Normal Model with Semi Natural Conjugate Prior

- prior

$$\mu \sim N(\gamma_0, \phi_0^2)$$

$$\sigma \sim \chi^{-1}(\kappa_0^2, \nu_0)$$

$$h(\mu, \sigma) = h(\mu)h(\sigma)$$

$$\propto \frac{1}{\sqrt{\phi_0^2}} \exp\left(-\frac{1}{2\phi_0^2}(\mu - \gamma_0)^2\right) \times \sigma^{-\nu_0-1} \exp\left(-\frac{\kappa_0^2}{2\sigma^2}\right)$$

- posterior

$$h(\mu, \sigma | \mathbf{y}) \propto f(\mathbf{y} | \mu, \sigma) h(\mu, \sigma)$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{\text{RSS}_y + \kappa_0^2}{\sigma^2} + \frac{N(\gamma_0 - \bar{y})^2}{\sigma^2 + N\phi_0^2} + \left(\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}\right)\left(\mu - \frac{\gamma_0\sigma^2 + N\phi_0^2\bar{y}}{\sigma^2 + N\phi_0^2}\right)^2\right)\right]$$

$$\times \sigma^{-N-\nu_0-1} \phi_0^{-1}$$

cannot be identified.

5

## Normal Model with Semi Natural Conjugate Prior

- full conditional posterior

$$h(\mu, | \sigma, \mathbf{y}) \propto \exp\left[-\frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}\right)\left(\mu - \frac{\gamma_0\sigma^2 + N\phi_0^2\bar{y}}{\sigma^2 + N\phi_0^2}\right)^2\right]$$

$$\mu | \sigma, \mathbf{y} \sim N\left[\frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}}\left(\frac{1}{\phi_0^2}\gamma_0 + \frac{N}{\sigma^2}\bar{y}\right), \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}}\right]$$

$$h(\sigma | \mu, \mathbf{y}) \propto \sigma^{-(N+\nu_0)-1} \times \exp\left[-\frac{1}{2}\left(\frac{\text{RSS}_y + N(\mu - \bar{y})^2 + \kappa_0^2}{\sigma^2}\right)\right]$$

$$\sigma | \mu, \mathbf{y} \sim \chi^{-1}(N + \nu_0, N(\mu - \bar{y})^2 + \text{RSS}_y + \kappa_0^2)$$

6

## Example of Gibbs Sampler

- prior

$$\gamma_0 = 50, \phi_0 = 10, \nu_0 = 2, \kappa_0 = 12.5 \text{ (prior mean } \sigma = 5)$$

- data

$$N=20, \text{ mean}=100, \text{ std}=10$$

7

## Example of Gibbs Sampler

- generate X:μ given Y:σ

$$\mu | \sigma, \mathbf{y} \sim N\left[\frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}}\left(\frac{1}{\phi_0^2}\gamma_0 + \frac{N}{\sigma^2}\bar{y}\right), \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}}\right]$$

- generate Y:σ given X:μ

$$\sigma | \mu, \mathbf{y} \sim \chi^{-1}(N + \nu_0, N(\mu - \bar{y})^2 + \text{RSS}_y + \kappa_0^2)$$

8

# Example of Gibbs Sampler

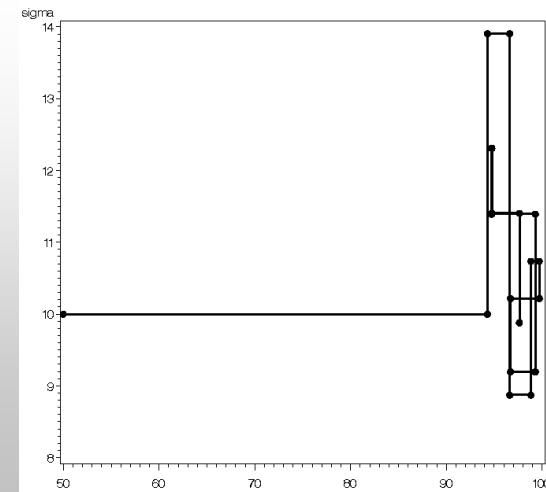
- generated random numbers

rep	mu	sigma
1	50.0000	10.0000
2	94.3247	10.0000
2	94.3247	13.9047
3	96.6447	13.9047
3	96.6447	8.8730
4	98.8580	8.8730
4	98.8580	10.7365
5	99.7417	10.7365
5	99.7417	10.2181
6	96.7341	10.2181
6	96.7341	9.1961
7	99.3172	9.1961
7	99.3172	11.3930
8	94.7472	11.3930
8	94.7472	12.3099
9	94.7861	12.3099
9	94.7861	11.4048
10	97.6591	11.4048
10	97.6591	9.8787

9

# Example of Gibbs Sampler

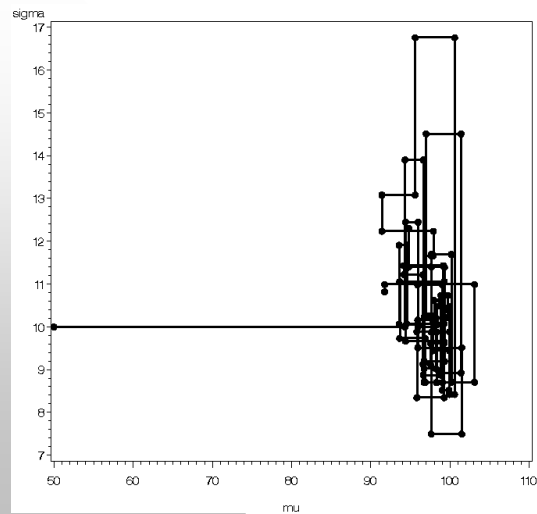
- bivariate history



10

# Example of Gibbs Sampler

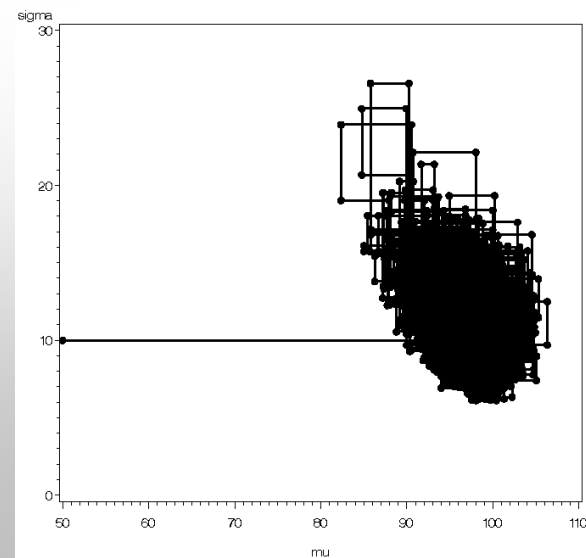
- bivariate history



11

# Example of Gibbs Sampler

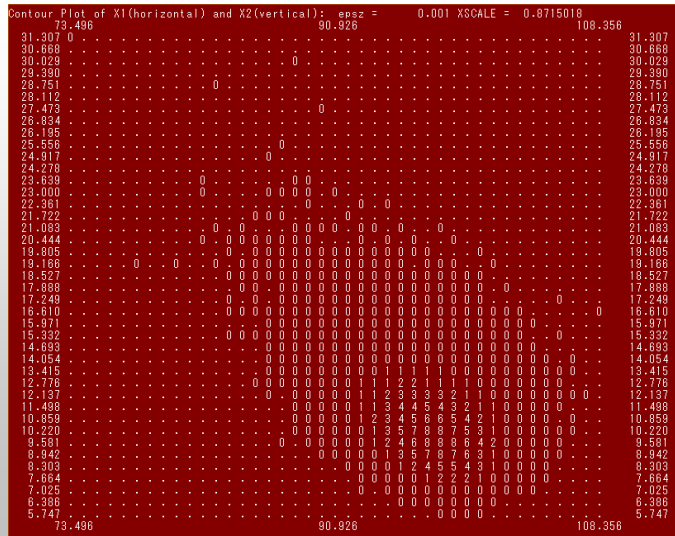
- bivariate history



12

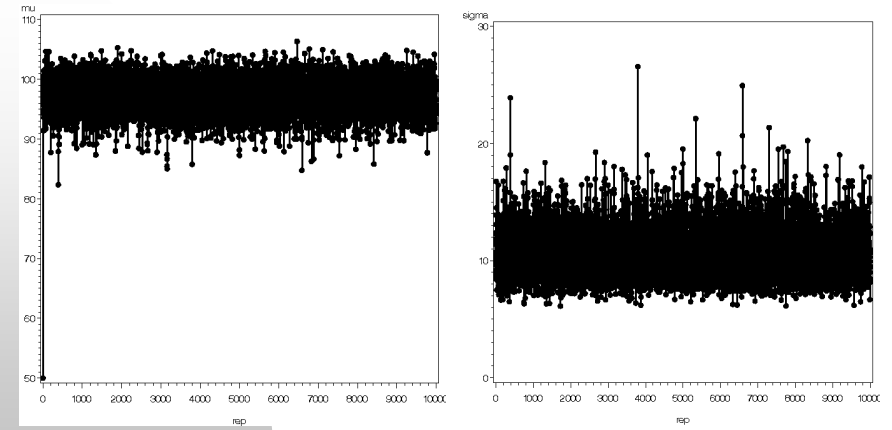
# Example of Gibbs Sampler

- joint distribution



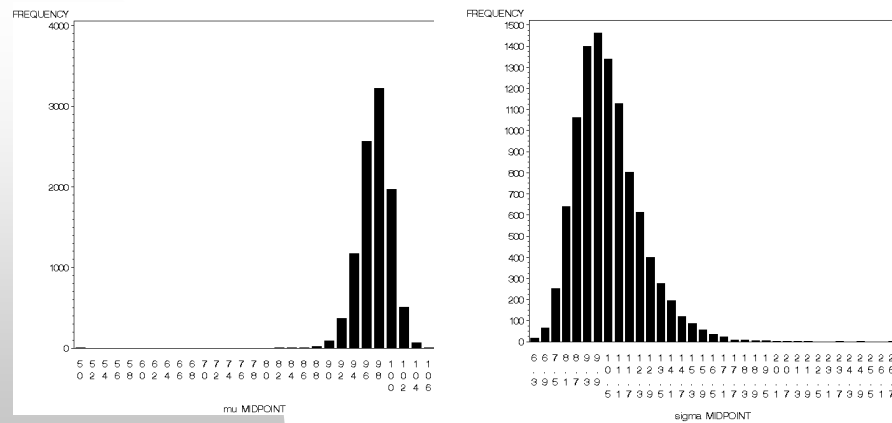
# Example of Gibbs Sampler

- marginal history plot



# Example of Gibbs Sampler

- marginal distribution of random numbers



# Bivariate Normal Example

- Why does it work?

Joint (target)

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

Conditional

$$X|Y \sim N(\rho Y, 1 - \rho^2)$$

$$Y|X \sim N(\rho X, 1 - \rho^2)$$

## Bivariate Normal Example

$$1.1) \quad X^{(1)} \mid y^{(0)} \sim N(\rho y^{(0)}, 1 - \rho^2)$$

$$1.2) \quad Y^{(1)} \mid x^{(1)} \sim N(\rho x^{(1)}, 1 - \rho^2)$$

$$Y^{(1)} \mid y^{(0)} \sim N(\rho^2 y^{(0)}, \rho^2(1 - \rho^2) + 1 - \rho^2)$$

$$\sim N(\rho^2 y^{(0)}, (1 - \rho^4))$$

$$2.1) \quad X^{(2)} \mid y^{(1)} \sim N(\rho y^{(1)}, 1 - \rho^2)$$

$$X^{(2)} \mid y^{(0)} \sim N(\rho^3 y^{(0)}, \rho^2(1 - \rho^4) + 1 - \rho^2)$$

$$\sim N(\rho^3 y^{(0)}, (1 - \rho^6))$$

$$2.2) \quad Y^{(2)} \mid x^{(2)} \sim N(\rho x^{(2)}, 1 - \rho^2)$$

$$Y^{(2)} \mid y^{(0)} \sim N(\rho^4 y^{(0)}, \rho^2(1 - \rho^6) + 1 - \rho^2)$$

$$\sim N(\rho^4 y^{(0)}, (1 - \rho^8))$$

Use the Theorem on p19.

17

## Bivariate Normal Example

$$\text{cov}(X^{(1)}, Y^{(1)} \mid y^{(0)}) = \rho(1 - \rho^2)$$

$$\text{cov}(X^{(2)}, Y^{(2)} \mid y^{(0)}) = \rho(1 - \rho^6)$$

■ In general

$$X^{(t)} \mid y^{(0)} \sim N(\rho^{2t-1} y^{(0)}, (1 - \rho^{4(t-1)+2}))$$

$$Y^{(t)} \mid y^{(0)} \sim N(\rho^{2t} y^{(0)}, (1 - \rho^{4t}))$$

$$\text{cov}(X^{(t)}, Y^{(t)} \mid y^{(0)}) = \rho(1 - \rho^{4(t-1)+2})$$

Converges to  $N(0, \rho)$  regardless of  $y_0$ .

18

## Bivariate Normal Example

■ Theorem

If

$$A \mid B \sim N(d + c B, v_1)$$

and

$$B \sim N(m_2, v_2)$$

then

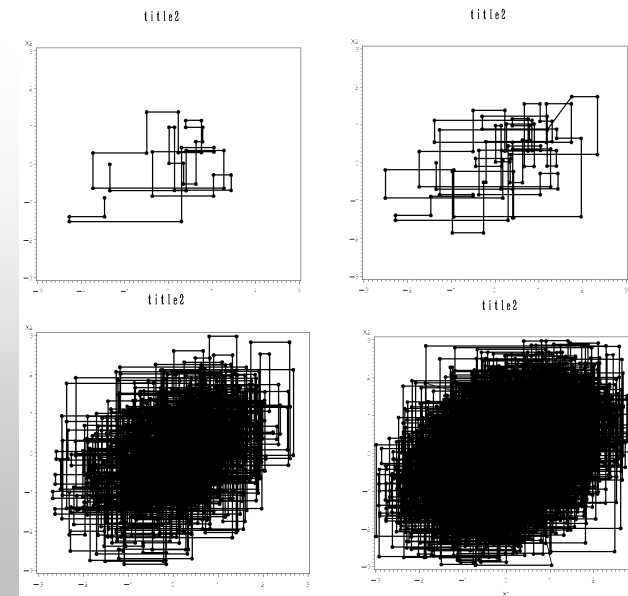
$$A \sim N(d + c m_2, c^2 v_2 + v_1)$$

and

$$\text{cov}(A, B) = c v_2$$

19

## Bivariate Normal Example



20

## Hammersley-Clifford Theorem

- Joint pdf can be recovered from the full conditionals.

$$f(x_1, x_2) \propto \frac{f(x_1|\xi_2)f(x_2|x_1)}{f(\xi_1|\xi_2)f(\xi_2|x_1)} \propto \frac{f(x_1|\xi_2)f(x_2|x_1)}{f(\xi_2|x_1)}$$

$$f(x_1, x_2) = \frac{f(x_2|x_1)}{\int_{X_2} \frac{f(x_2|x_1)}{f(x_1|x_2)} dx_2}$$

21

## Gibbs Sampler

- Gibbs Sampler

1 : Two Variate Gibbs Sampler

1. Estimate  $f^0(Y)$
2. Draw  $Y^{(0)} = y^{(0)}$  according to  $f^0(Y)$ .
3. Draw  $X^{(1)} = x^{(1)}$  according to  $f_{X|Y}(X|Y = y^{(0)})$
4. Draw  $Y^{(1)} = y^{(1)}$  according to  $f_{Y|X}(Y|X = x^{(1)})$
5. Draw  $X^{(2)} = x^{(2)}$  according to  $f_{X|Y}(X|Y = y^{(1)})$
6. Draw  $Y^{(2)} = y^{(2)}$  according to  $f_{Y|X}(Y|X = x^{(2)})$
- ⋮
7. Draw  $X^{(t+1)} = x^{(t+1)}$  according to  $f_{X|Y}(X|Y = y^{(t)})$
8. Draw  $Y^{(t+1)} = y^{(t+1)}$  according to  $f_{Y|X}(Y|X = x^{(t+1)})$
- ⋮

22

## Gibbs Sampler

- values at  $t$  depend only on the values at  $t-1$ .

$$X^{(t)} \sim f_{X|Y}(x^{(t)}|y^{(t-1)})$$

$$Y^{(t)} \sim f_{Y|X}(y^{(t)}|x^{(t)})$$

- the conditional probabilities do not depend on  $t$ .

$$f_{X|Y}(x^{(t)}|y^{(t-1)}) = f_{X|Y}(x^{(t-1)}|y^{(t-2)})$$

$$f_{Y|X}(y^{(t)}|x^{(t)}) = f_{Y|X}(y^{(t-1)}|x^{(t-1)})$$

### discrete case

$$\Pr(X^{(t)} = x_i | Y^{(t-1)} = y_j) = \Pr(X^{(t-1)} = x_i | Y^{(t-2)} = y_j)$$

$$\Pr(Y^{(t)} = y_j | X^{(t)} = x_i) = \Pr(Y^{(t-1)} = y_j | X^{(t-1)} = x_i)$$

23

## Gibbs Sampler : Discrete Case

- discrete domain of  $X$  and  $Y$

$$(x_1, x_2, \dots, x_{n_x})$$

$$(y_1, y_2, \dots, y_{n_y})$$

- $(X = x_i, Y = y_j) \rightarrow (xy)_{ij}$

- Transition Kernel  
from  $(xy)_{ij}$  to  $(xy)_{IJ}$

$$K_{XY}[(xy)_{ij}, (xy)_{IJ}] \equiv \Pr(X^{(t+1)} = x_I, Y^{(t+1)} = y_J | X^{(t)} = x_i, Y^{(t)} = y_j)$$

from, to

to | from

24

# Gibbs Sampler : Discrete Case

- Transition Kernel of the Gibbs Sampler

$$\begin{aligned}
 K_{XY}[(xy)_{ij}, (xy)_{IJ}] &= \\
 &\Pr(X^{(t+1)} = x_I, Y^{(t+1)} = y_J | X^{(t)} = x_i, Y^{(t)} = y_j) \\
 &= \Pr(X^{(t+1)} = x_I | Y^{(t)} = y_j) \times \\
 &\quad \Pr(Y^{(t+1)} = y_J | X^{(t+1)} = x_I) \\
 &= f_{X|Y}(x_I | y_j) f_{Y|X}(y_J | x_I) \\
 &\sum_{i=1}^{n_x} f_{X|Y}(x_i | y_j) = 1 \quad \sum_{i=1}^{n_y} f_{Y|X}(y_i | x_j) = 1
 \end{aligned}$$

# Gibbs Sampler : Discrete Case

- Example of Gibbs transition kernel

$$n_x = 3, n_y = 2 \quad K_{XY}[(xy)_{ij}, (xy)_{IJ}] \equiv \Pr(X^{(t+1)} = x_I, Y^{(t+1)} = y_J | X^{(t)} = x_i, Y^{(t)} = y_j)$$

	(I:J) 1:1	(I:J) 2:1	(I:J) 3:1
(i:j) 1:1	$f_{X Y}(x_1 y_1) f_{Y X}(y_1 x_1)$	$f_{X Y}(x_2 y_1) f_{Y X}(y_1 x_2)$	$f_{X Y}(x_3 y_1) f_{Y X}(y_1 x_3)$
(i:j) 2:1	$f_{X Y}(x_1 y_1) f_{Y X}(y_1 x_1)$	$f_{X Y}(x_2 y_1) f_{Y X}(y_1 x_2)$	$f_{X Y}(x_3 y_1) f_{Y X}(y_1 x_3)$
(i:j) 3:1	$f_{X Y}(x_1 y_1) f_{Y X}(y_1 x_1)$	$f_{X Y}(x_2 y_1) f_{Y X}(y_1 x_2)$	$f_{X Y}(x_3 y_1) f_{Y X}(y_1 x_3)$
(i:j) 1:2	$f_{X Y}(x_1 y_2) f_{Y X}(y_1 x_1)$	$f_{X Y}(x_2 y_2) f_{Y X}(y_1 x_2)$	$f_{X Y}(x_3 y_2) f_{Y X}(y_1 x_3)$
(i:j) 2:2	$f_{X Y}(x_1 y_2) f_{Y X}(y_1 x_1)$	$f_{X Y}(x_2 y_2) f_{Y X}(y_1 x_2)$	$f_{X Y}(x_3 y_2) f_{Y X}(y_1 x_3)$
(i:j) 3:2	$f_{X Y}(x_1 y_2) f_{Y X}(y_1 x_1)$	$f_{X Y}(x_2 y_2) f_{Y X}(y_1 x_2)$	$f_{X Y}(x_3 y_2) f_{Y X}(y_1 x_3)$

	(I:J) 1:2	(I:J) 2:2	(I:J) 3:2
(i:j) 1:1	$f_{X Y}(x_1 y_1) f_{Y X}(y_2 x_1)$	$f_{X Y}(x_2 y_1) f_{Y X}(y_2 x_2)$	$f_{X Y}(x_3 y_1) f_{Y X}(y_2 x_3)$
(i:j) 2:1	$f_{X Y}(x_1 y_1) f_{Y X}(y_2 x_1)$	$f_{X Y}(x_2 y_1) f_{Y X}(y_2 x_2)$	$f_{X Y}(x_3 y_1) f_{Y X}(y_2 x_3)$
(i:j) 3:1	$f_{X Y}(x_1 y_1) f_{Y X}(y_2 x_1)$	$f_{X Y}(x_2 y_1) f_{Y X}(y_2 x_2)$	$f_{X Y}(x_3 y_1) f_{Y X}(y_2 x_3)$
(i:j) 1:2	$f_{X Y}(x_1 y_2) f_{Y X}(y_2 x_1)$	$f_{X Y}(x_2 y_2) f_{Y X}(y_2 x_2)$	$f_{X Y}(x_3 y_2) f_{Y X}(y_2 x_3)$
(i:j) 2:2	$f_{X Y}(x_1 y_2) f_{Y X}(y_2 x_1)$	$f_{X Y}(x_2 y_2) f_{Y X}(y_2 x_2)$	$f_{X Y}(x_3 y_2) f_{Y X}(y_2 x_3)$
(i:j) 3:2	$f_{X Y}(x_1 y_2) f_{Y X}(y_2 x_1)$	$f_{X Y}(x_2 y_2) f_{Y X}(y_2 x_2)$	$f_{X Y}(x_3 y_2) f_{Y X}(y_2 x_3)$

# Gibbs Sampler : Discrete Case

- numerical example:

$$f_{X|Y}(x_i | y_j)$$

	X			
	1	2	3	4
Y 1	0.167	0.333	0.333	0.167
Y 2	0.250	0.250	0.333	0.167
Y 3	0.143	0.286	0.429	0.143

$$f_{Y|X}(y_i | x_j)$$

	Y		
	1	2	3
X 1	0.200	0.600	0.200
X 2	0.286	0.429	0.286
X 3	0.222	0.444	0.333
X 4	0.250	0.500	0.250

# Gibbs Sampler : Discrete Case

- transition kernel

$$K_{XY} =$$

	1:1	2:1	3:1	4:1	1:2	2:2	3:2	4:2	1:3	2:3	3:3	4:3
1:1	0.033	0.095	0.074	0.042	0.100	0.143	0.148	0.083	0.033	0.095	0.111	0.042
2:1	0.033	0.095	0.074	0.042	0.100	0.143	0.148	0.083	0.033	0.095	0.111	0.042
3:1	0.033	0.095	0.074	0.042	0.100	0.143	0.148	0.083	0.033	0.095	0.111	0.042
4:1	0.033	0.095	0.074	0.042	0.100	0.143	0.148	0.083	0.033	0.095	0.111	0.042
1:2	0.050	0.071	0.074	0.042	0.150	0.107	0.148	0.083	0.050	0.071	0.111	0.042
2:2	0.050	0.071	0.074	0.042	0.150	0.107	0.148	0.083	0.050	0.071	0.111	0.042
3:2	0.050	0.071	0.074	0.042	0.150	0.107	0.148	0.083	0.050	0.071	0.111	0.042
4:2	0.050	0.071	0.074	0.042	0.150	0.107	0.148	0.083	0.050	0.071	0.111	0.042
1:3	0.029	0.082	0.095	0.036	0.086	0.122	0.190	0.071	0.029	0.082	0.143	0.036
2:3	0.029	0.082	0.095	0.036	0.086	0.122	0.190	0.071	0.029	0.082	0.143	0.036
3:3	0.029	0.082	0.095	0.036	0.086	0.122	0.190	0.071	0.029	0.082	0.143	0.036
4:3	0.029	0.082	0.095	0.036	0.086	0.122	0.190	0.071	0.029	0.082	0.143	0.036

## Gibbs Sampler : Discrete Case

- marginal probability of XY at t+1

$$f_{XY}^{(t+1)}(x_I, y_J) = \sum_{ij} f_{XY}^{(t)}(x_i, y_j) K_{XY}[(xy)_{ij}, (xy)_{IJ}]$$

marginal at t x conditional

$$\Pr(A) = \sum_{j=1}^{n_B} \Pr(B_j) \Pr(A|B_j)$$

- Global balance condition

$$f_{XY}(x_I, y_J) = \sum_{ii} f_{XY}(x_i, y_j) K_{XY}[(xy)_{ij}, (xy)_{IJ}]$$

where  $f_{XY}(x_i, y_j)$  is the **stationary distribution** associated with **K**.

29

## Gibbs Sampler : Discrete Case

- If we construct the transition kernel  $K_{XY}[(xy)_{ij}, (xy)_{IJ}]$  from the full conditional distribution of the target joint distribution  $f_{XY}(x_i, y_j)$ , the target joint distribution is the stationary distribution.

$$K_{XY}[(xy)_{ij}, (xy)_{IJ}] = f_{X|Y}(x_I|y_j) f_{Y|X}(y_J|x_I)$$

$$f_{X|Y}(x_i|y_j) = \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)}$$

$$f_{Y|X}(y_j|x_i) = \frac{f_{XY}(x_i, y_j)}{f_X(x_i)}$$

$$f_X(x_i) = \sum_{j=1}^{n_Y} f_{XY}(x_i, y_j), \quad f_Y(y_j) = \sum_{i=1}^{n_X} f_{XY}(x_i, y_j)$$

30

## Gibbs Sampler : Discrete Case

- Suppose that  $f_{XY}^{(t)}(x_i, y_j) = f_{XY}(x_i, y_j)$ , then

$$\begin{aligned} f_{XY}^{(t+1)}(x_I, y_J) &= \sum_{ij} f_{XY}(x_i, y_j) K_{XY}[(xy)_{ij}, (xy)_{IJ}] \\ &= \sum_{ij} f_{XY}(x_i, y_j) f_{X|Y}(x_I|y_j) f_{Y|X}(y_J|x_I) \\ &= \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} f_{XY}(x_i, y_j) f_{X|Y}(x_I|y_j) f_{Y|X}(y_J|x_I) \\ &= \sum_{j=1}^{n_Y} \left( \sum_{i=1}^{n_X} f_{XY}(x_i, y_j) \right) f_{X|Y}(x_I|y_j) f_{Y|X}(y_J|x_I) \\ &= \sum_{j=1}^{n_Y} f_Y(y_j) f_{X|Y}(x_I|y_j) f_{Y|X}(y_J|x_I) \\ &= \sum_{j=1}^{n_Y} f_Y(y_j) \frac{f_{XY}(x_I, y_j)}{f_Y(y_j)} f_{Y|X}(y_J|x_I) \\ &= \sum_{j=1}^{n_Y} f_{XY}(x_I, y_j) f_{Y|X}(y_J|x_I) \\ &= f_X(x_I) f_{Y|X}(y_J|x_I) = f_X(x_I) \frac{f_{XY}(x_I, y_J)}{f_X(x_I)} \\ &= f_{XY}(x_I, y_J) \quad \text{global balance condition} \end{aligned}$$

31