26/04/2022. LECTURE 19: NOTES MADE ON THE VISUALISER.

Gibbs sampler can be viewed as a special case of the Metropolis-Hastings algorithm, each step t consists of d M-H steps each with an acceptance probability of 1.

EXAMPLE.

Suppose that the joint distribution of $\theta = (\theta_1, \theta_2)$ is given by:

$$f(\theta_1, \theta_2) \propto \binom{n}{\theta_1} \theta_2^{\theta_1 + \alpha - 1} (1-\theta_2)^{n - \theta_1 + \beta - 1}$$

for known n, α, β and $\theta_1 \in \{0, 1, \dots, n\}$ and $0 < \theta_2 < 1$

[θ_1 is discrete and θ_2 continuous]

Suppose we consider using the Gibbs sampler to sample from this distribution.

$$f(\theta_1 | \theta_2) = \frac{f(\theta_1, \theta_2)}{f(\theta_2)} \propto f(\theta_1, \theta_2)$$

function of θ_1

viewed as a function of θ_1

[For a posterior, $f(\theta_1 | \theta_2, x) \propto f(\theta_1, \theta_2 | x)$]

$$\propto \binom{n}{\theta_1} \theta_2^{\theta_1} (1-\theta_2)^{n-\theta_1}$$

i.e. $\theta_1 | \theta_2 \sim Bin(n, \theta_2)$

$$f(\theta_2 \mid \theta_1) \propto \frac{f(\theta_1, \theta_2)}{f(\theta_1)} \propto f(\theta_1, \theta_2)$$

function of θ_2

viewed as a function of θ_2.

$$\propto \theta_2^{\theta_1 + \alpha - 1} (1 - \theta_2)^{n - \theta_1 + \beta - 1}$$

i.e. $\theta_2 \mid \theta_1 \sim \text{Beta}(\theta_1 + \alpha, n - \theta_1 + \beta)$

We now generate a Gibbs sequence of RVs.

1). Initial value $\theta_1^{(0)} \in \{0, 1, \ldots, n\}$, $0 < \theta_2^{(0)} < 1$

(or sample from $\text{Beta}(\theta_1^{(0)} + \alpha, n - \theta_1^{(0)} + \beta)$.)

2). At time t, $[\theta^{(t-1)} = (\theta_1^{(t-1)}, \theta_2^{(t-1)})]$

- Obtain $\theta_1^{(t)}$ by sampling from $\text{Bin}(n, \theta_2^{(t-1)})$, the distribution of $\theta_1 \mid \theta_2^{(t-1)}$

[M-H Step: move from $\theta_{[1]}^{(t-1)} = (\theta_1^{(t-1)}, \theta_2^{(t-1)})$ to $\theta_{[1]}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t-1)})$ using the proposal $f(\theta_1 \mid \theta_2^{(t-1)})$

Thus, $q(\theta_{[1]}^{(t)} \mid \theta_{[1]}^{(t-1)}) = f(\theta_1^{(t)} \mid \theta_2^{(t-1)})$

$q(\theta_{[1]}^{(t-1)} \mid \theta_{[1]}^{(t)}) = f(\theta_1^{(t-1)} \mid \theta_2^{(t-1)})$

$$f(\theta_{[1]}^{(t)}) = f(\theta_1^{(t)}, \theta_2^{(t-1)}) = f(\theta_1^{(t)} \mid \theta_2^{(t-1)}) f(\theta_2^{(t-1)})$$

$$f(\theta_{[1]}^{(t-1)}) = f(\theta_1^{(t-1)}, \theta_2^{(t-1)}) = f(\theta_1^{(t-1)} \mid \theta_2^{(t-1)}) f(\theta_2^{(t-1)})$$

(stationary distribution is $f(\theta_1, \theta_2)$)

Thus,

$$\frac{f(\theta_{[1]}^{(t)})}{f(\theta_{[1]}^{(t-1)})} = \frac{f(\theta_1^{(t)} \mid \theta_2^{(t-1)})}{f(\theta_1^{(t-1)} \mid \theta_2^{(t-1)})} = \frac{q(\theta_{[1]}^{(t)} \mid \theta_{[1]}^{(t-1)})}{q(\theta_{[1]}^{(t-1)} \mid \theta_{[1]}^{(t)})}$$

So that, $$\frac{f(\theta_{[1]}^{(t)})\, q(\theta_{[1]}^{(t-1)} \mid \theta_{[1]}^{(t)})}{f(\theta_{[1]}^{(t-1)})\, q(\theta_{[1]}^{(t)} \mid \theta_{[1]}^{(t-1)})} = 1$$

In the corresponding M-H algorithm, $\alpha(\theta_{[1]}^{(t-1)}, \theta_{[1]}^{(t)}) = \min(1, 1) = 1$ i.e. Accept the move. (proposed move from $\theta_{[1]}^{(t-1)}$ to $\theta^{+} = \theta_{[1]}^{(t)}$ is accepted).]

- Obtain $\theta_2^{(t)}$ by sampling from $\text{Beta}(\theta_1^{(t)} + \alpha, n - \theta_1^{(t)} + \beta)$, the distribution of $\theta_2 \mid \theta_1^{(t)}$.

[M-H move from $\theta_{[2]}^{(t-1)} = (\theta_1^{(t)}, \theta_2^{(t-1)})$ to $\theta_{[2]}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)})$ using the proposed $f(\theta_2 \mid \theta_1^{(t)})$ with an acceptance probability of 1]

3. Repeat step 2.

The sequence $\theta_1^{(0)}, \theta_2^{(0)}, \theta_1^{(1)}, \theta_2^{(1)}, \ldots$ is such that the distribution of $(\theta_1^{(k)}, \theta_2^{(k)})$ tends to $\pi(\theta_1, \theta_2) = f(\theta_1, \theta_2)$ as $k \to \infty$

[so the distribution of $\theta_1^{(k)}$ tends to $\pi(\theta_1) = f(\theta_1)$ and that of $\theta_2^{(k)}$ to $\pi(\theta_2) = f(\theta_2)$ as $k \to \infty$].

We'll illustrate this in R. To do so, we'll find the two marginal distributions.

$$f(\theta_1) = \int_0^1 f(\theta_1, \theta_2)\, d\theta_2 \propto \int_0^1 \binom{n}{\theta_1} \underbrace{\theta_2^{\theta_1+\alpha-1}(1-\theta_2)^{n-\theta_1+\beta-1}}_{\text{kernel of Beta}(\theta_1+\alpha,\, n-\theta_1+\beta)} d\theta_2$$

$$= \binom{n}{\theta_1} B(\theta_1+\alpha, n-\theta_1+\beta)$$

Binomial - Beta distribution, values n, α, β.

$$f(\theta_2) \propto \sum_{\theta_1=0}^{n} \binom{n}{\theta_1} \theta_2^{\theta_1+\alpha-1}(1-\theta_2)^{n-\theta_1+\beta-1} = \theta_2^{\alpha-1}(1-\theta_2)^{\beta-1}$$

$\theta_2 \sim \text{Beta}(\alpha, \beta)$

R Example.

1). $n=10$, $\alpha=1$, $\beta=1$ so that θ_1 is the discrete uniform and θ_2 is the uniform on $(0,1)$.

Figure 1: 50 observations, start with $\theta_1^{(0)}=5$, $\theta_2^{(0)}=0.5$ (function gibbs2).

Figure 2: 1500 observations, start with $\theta_1^{(0)}=5$, $\theta_2^{(0)}=0.5$

2). $n=10$, $\alpha=2$, $\beta=3$

Figure 3: 50 observations, Figure 4: 1500 observations.

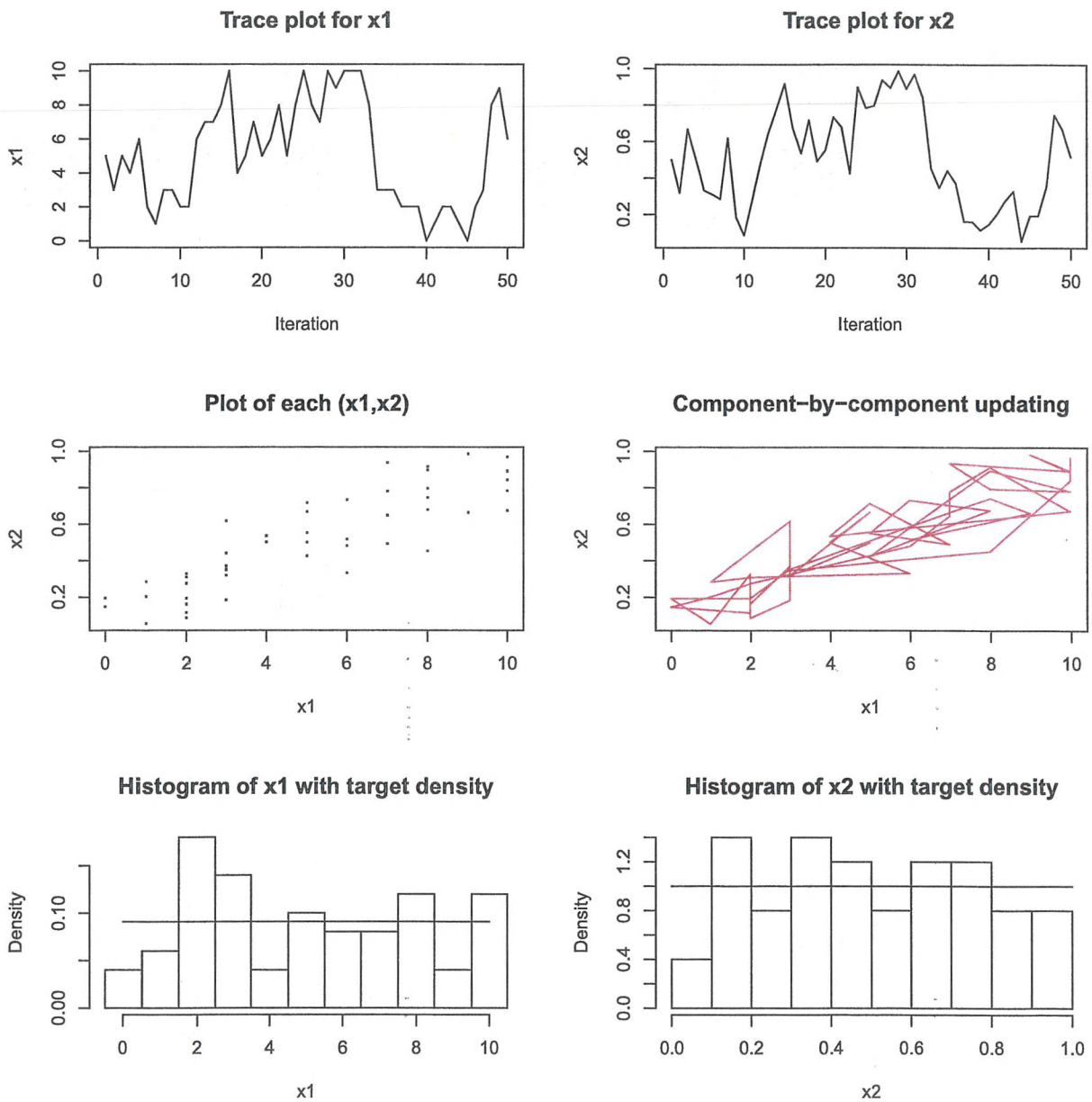

Figure 1: Fifty iterations from a Gibbs sampler for $x_1 \,|\, x_2 \sim Bin(n, x_2)$, $x_2 \,|\; x_1 \sim Beta(x_1 + \alpha, n - x_1 + \beta)$ where $n = 10$ and $\alpha = \beta = 1$. The marginal distributions are $x_1 \sim Beta - binomial(n, \alpha, \beta)$ and $x_2 \sim Beta(\alpha, \beta)$. For $\alpha = \beta = 1$, x_1 is the discrete uniform on $\{0, 1, \ldots, n\}$ and $x_2 \sim U(0, 1)$.

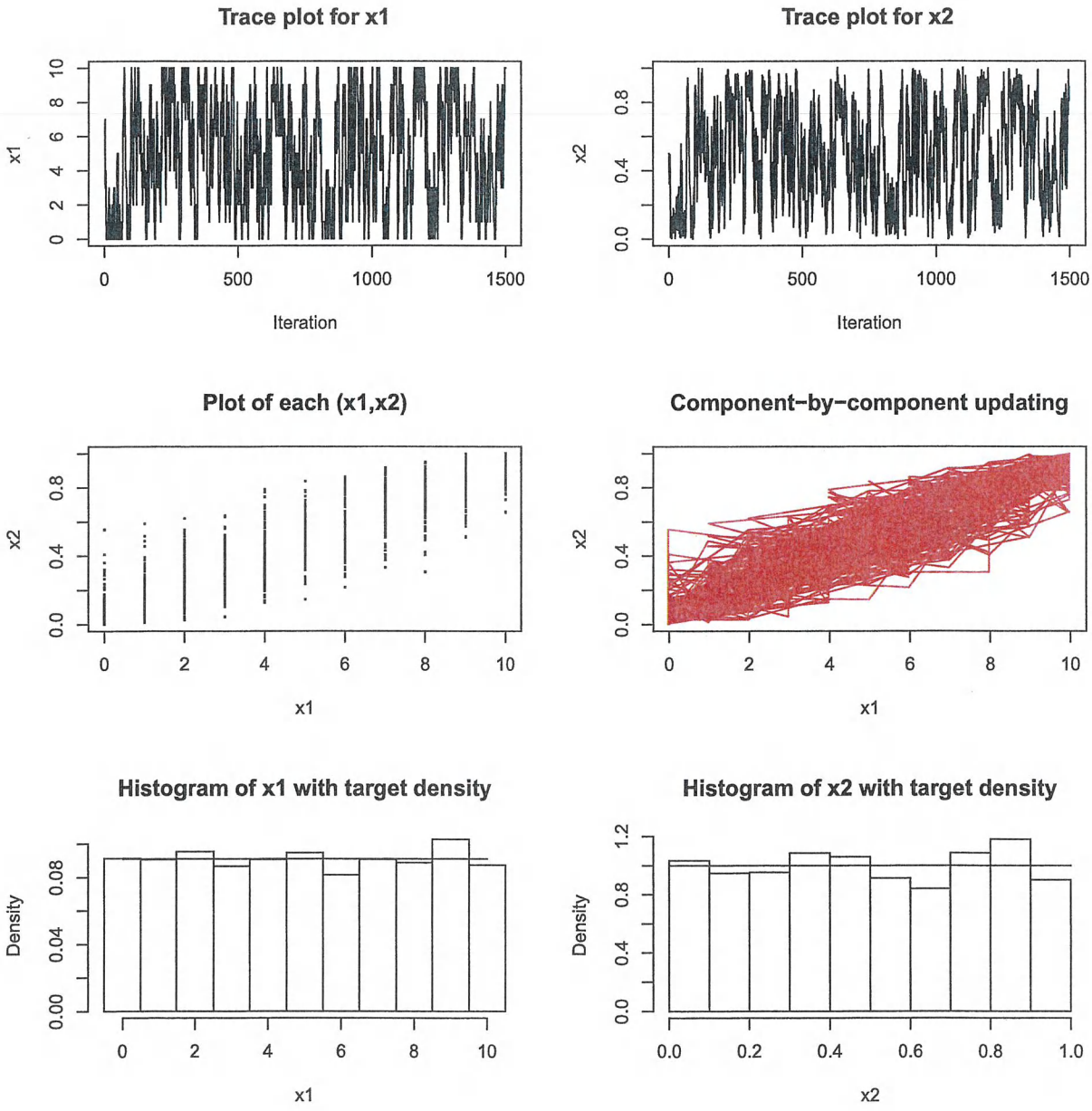

Figure 2: 1500 iterations from a Gibbs sampler for $x_1 \,|\, x_2 \sim Bin(n, x_2)$, $x_2 \,|\, x_1 \sim Beta(x_1 + \alpha, n - x_1 + \beta)$ where $n = 10$ and $\alpha = \beta = 1$. The marginal distributions are $x_1 \sim Beta - binomial(n, \alpha, \beta)$ and $x_2 \sim Beta(\alpha, \beta)$. For $\alpha = \beta = 1$, x_1 is the discrete uniform on $\{0, 1, \ldots, n\}$ and $x_2 \sim U(0, 1)$.

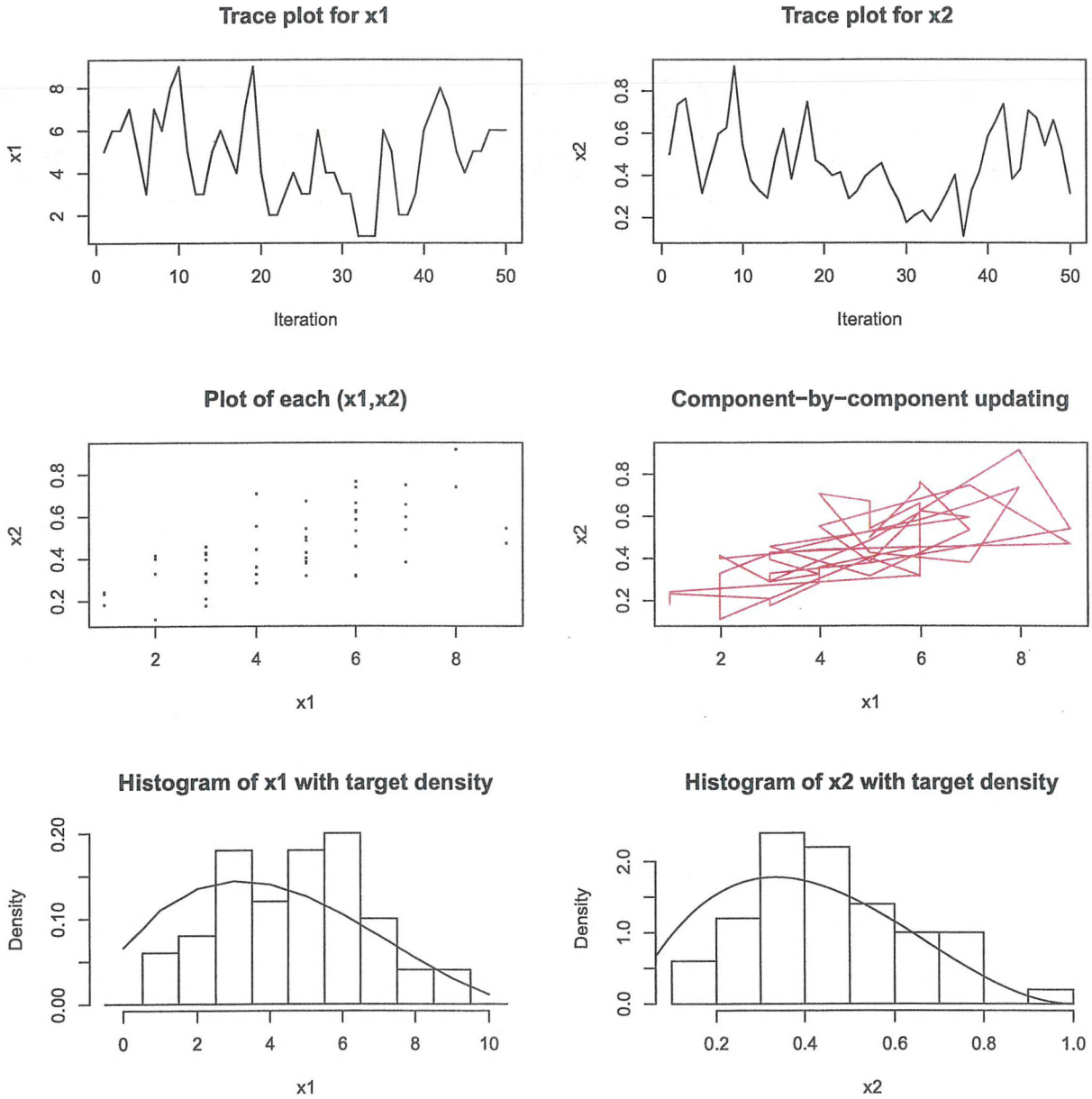

Figure 3: Fifty iterations from a Gibbs sampler for $x_1 \mid x_2 \sim Bin(n, x_2)$, $x_2 \mid\ x_1 \sim Beta(x_1 + \alpha, n - x_1 + \beta)$ where $n = 10$, $\alpha = 2$ and $\beta = 3$. The marginal distributions are $x_1 \sim Beta - binomial(n, \alpha, \beta)$ and $x_2 \sim Beta(\alpha, \beta)$.

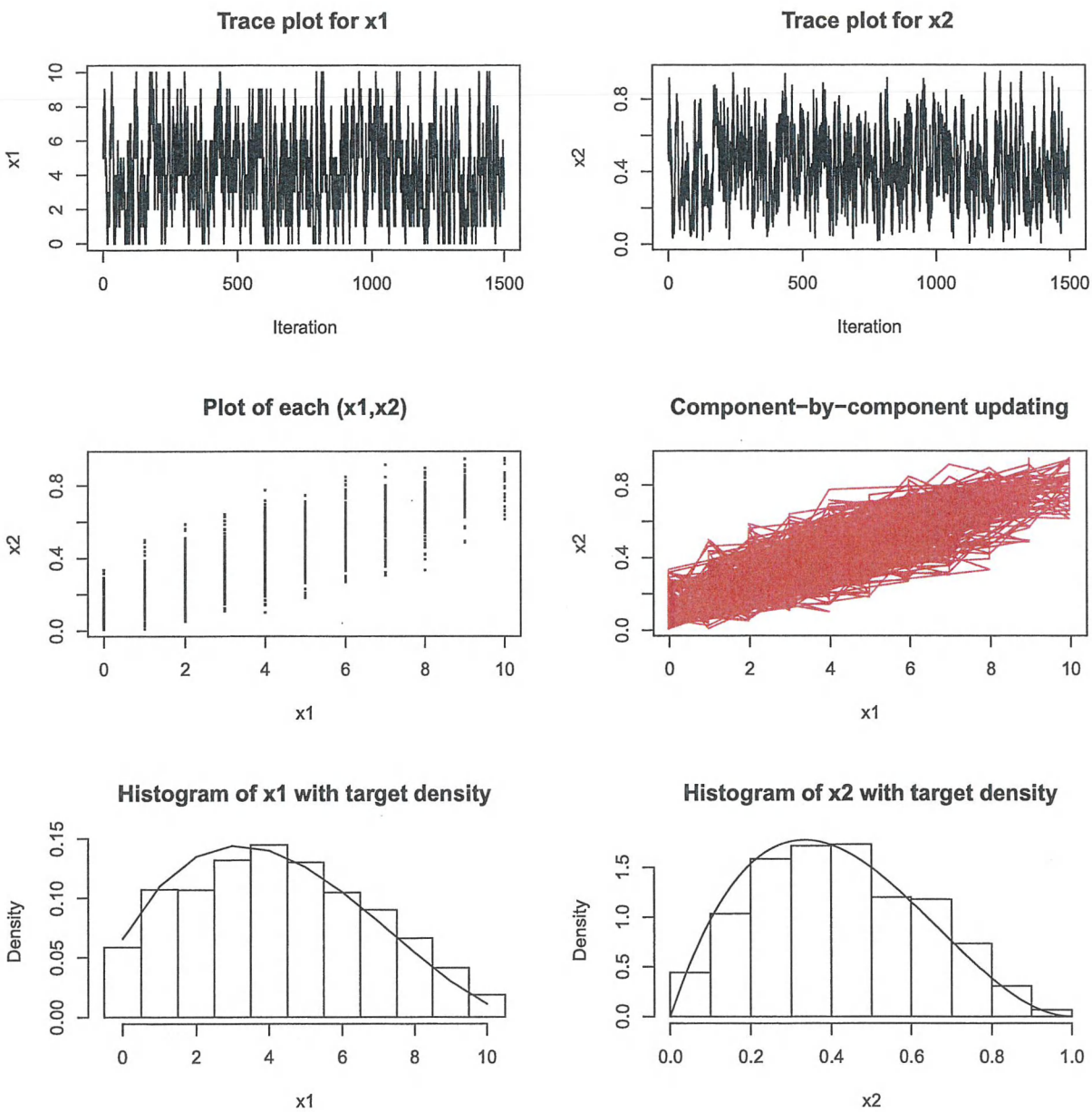

Figure 4: 1500 iterations from a Gibbs sampler for $x_1 \mid x_2 \sim Bin(n, x_2)$, $x_2 \mid\ x_1 \sim Beta(x_1 + \alpha, n - x_1 + \beta)$ where $n = 10$, $\alpha = 2$ and $\beta = 3$. The marginal distributions are $x_1 \sim Beta - binomial(n, \alpha, \beta)$ and $x_2 \sim Beta(\alpha, \beta)$.