Computational Statistics
Gibbs sampling

J. Tang

University of Birmingham

From Metropolis–Hastings to Gibbs sampling

Metropolis–Hastings constructs a Markov chain by proposing a new state and accepting or rejecting it.

Gibbs sampling uses a particular proposal: draw one coordinate from its full conditional distribution.

Today we will

- define full conditionals and construct one-coordinate proposals;
- implement a complete Gibbs sweep using the latest coordinate values.

We will then work through a Gaussian example and combine Gibbs updates with MH steps.

The target is a joint distribution

Let

$$x = (x_1, \ldots, x_d) \in \mathbb{R}^d, \qquad x \sim \pi(x_1, \ldots, x_d).$$

We want a Markov chain with invariant distribution π.

- x^t is the state after iteration t; superscript t is an iteration index.
- x_j^t is coordinate j of that state.
- The target may be a posterior, but the algorithm applies to general joint targets.

For example, retained states can estimate

$$\mathcal{I} = \mathbb{E}_\pi[f(X)], \qquad \widehat{\mathcal{I}}_N = \frac{1}{N} \sum_{t=B+1}^{B+N} f(x^t).$$

The full conditional distribution

Remove coordinate j from the vector:

$$\boxed{x_{-j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d).}$$

The full conditional of x_j, given every other coordinate, is

$$\pi(x_j \mid x_{-j}) = \frac{\pi(x_j, x_{-j})}{\displaystyle\int_{\mathbb{R}} \pi(u, x_{-j})\, du},$$

whenever the denominator is positive.

When can we use Gibbs?

We must be able to draw from each required full conditional. A familiar closed-form distribution is convenient, but is not a mathematical requirement.

Deriving a full conditional from an unnormalised target

Suppose $\pi(x) = g(x)/Z$, where Z is unknown. For fixed x_{-j},

$$\pi(x_j \mid x_{-j}) \propto g(x_j, x_{-j}).$$

To identify the conditional:

1. Hold every coordinate except x_j fixed.
2. Retain the factors that depend on x_j.
3. Recognise or normalise the resulting one-dimensional density.

Evaluating a density is not the same as sampling from it

Knowing the conditional up to a constant is enough for an MH update. An exact Gibbs update additionally needs a way to sample from that conditional.

One Gibbs update changes one coordinate

At the current state x^t, select a coordinate j and draw

$$x_j^* \sim \pi(\cdot \mid x_{-j}^t).$$

Construct the candidate

$$\boxed{x^* = (x_1^t, \ldots, x_{j-1}^t, x_j^*, x_{j+1}^t, \ldots, x_d^t).}$$

For a three-dimensional state and $j = 2$:

$$(a, b, c) \quad \longrightarrow \quad (a, b^*, c), \qquad b^* \sim \pi(x_2 \mid x_1 = a, x_3 = c).$$

All other coordinates are unchanged: $x_{-j}^* = x_{-j}^t$.

The Gibbs proposal

The one-coordinate proposal can be written as

$$q_j(x^* \mid x^t) = \begin{cases} \pi(x_j^* \mid x_{-j}^t), & x_{-j}^* = x_{-j}^t, \\ 0, & \text{otherwise.} \end{cases} \tag{7.3}$$

Interpretation:

- randomly draw the selected coordinate from its full conditional;
- keep the other coordinates fixed exactly.

For continuous state spaces

This is shorthand for a conditional density in coordinate j and point masses in the fixed coordinates, rather than an ordinary density over all of $\mathbb{R}^d$.

Gibbs is a special case of Metropolis–Hastings

Recall the MH acceptance probability:

$$a(x^*, x^t) = \min\left\{1, \frac{\pi(x^*) q_j(x^t \mid x^*)}{\pi(x^t) q_j(x^* \mid x^t)}\right\}.$$

For the Gibbs proposal,

$$q_j(x^* \mid x^t) = \pi(x_j^* \mid x_{-j}^t), \qquad q_j(x^t \mid x^*) = \pi(x_j^t \mid x_{-j}^t).$$

The reverse proposal uses the same conditioning vector because

$$x_{-j}^* = x_{-j}^t.$$

Why every Gibbs proposal is accepted

Write the common fixed coordinates as $z = x^t_{-j} = x^*_{-j}$. The target factorises as

$$\pi(x^*) = \pi(x^*_j \mid z)\pi_{-j}(z), \qquad \pi(x^t) = \pi(x^t_j \mid z)\pi_{-j}(z).$$

Consequently, on the relevant support,

$$\begin{aligned} r &= \frac{\pi(x^*)}{\pi(x^t)} \frac{q_j(x^t \mid x^*)}{q_j(x^* \mid x^t)} \\ &= \frac{\pi(x^*_j \mid z)\pi_{-j}(z)}{\pi(x^t_j \mid z)\pi_{-j}(z)} \frac{\pi(x^t_j \mid z)}{\pi(x^*_j \mid z)} = 1. \end{aligned}$$

$$\boxed{a(x^*, x^t) = 1.}$$

Acceptance one: what it does and does not mean

What follows

- No accept/reject test is needed for an exact Gibbs draw.
- There is no random-walk step size to tune for this conditional draw.
- The conditional proposal already supplies the MH correction.

What does not follow

- Successive states need not be independent.
- The chain need not explore the target rapidly.
- The selected coordinate need not change, for example with a discrete conditional.

A complete sequential Gibbs sweep

Choose an initial state x^0 in the target support. For $t = 0, \dots, T-1$:

$$
\begin{aligned}
x_1^{t+1} &\sim \pi(x_1 \mid x_2^t, \dots, x_d^t), \\
x_2^{t+1} &\sim \pi(x_2 \mid x_1^{t+1}, x_3^t, \dots, x_d^t), \\
&\vdots \\
x_j^{t+1} &\sim \pi(x_j \mid x_1^{t+1}, \dots, x_{j-1}^{t+1}, x_{j+1}^t, \dots, x_d^t), \\
&\vdots \\
x_d^{t+1} &\sim \pi(x_d \mid x_1^{t+1}, \dots, x_{d-1}^{t+1}).
\end{aligned}
$$

Use the newest available values. Store x^{t+1} after the full sweep.

Reading the indices: a three-coordinate sweep

Suppose the old state is $x^t = (a, b, c)$.

Sub-step	Conditional draw	Current working state
Start	—	(a, b, c)
Update 1	$u \sim \pi(x_1 \mid b, c)$	(u, b, c)
Update 2	$v \sim \pi(x_2 \mid u, c)$	(u, v, c)
Update 3	$w \sim \pi(x_3 \mid u, v)$	(u, v, w)

The new state is $x^{t+1} = (u, v, w)$.

Two different indices

t counts full sweeps; j selects the coordinate being updated within a sweep.

Sequential and simultaneous updates

The values used in each conditional determine the transition of the chain.

Sequential Gibbs

After drawing a coordinate, immediately store its new value. Use that updated value in every subsequent conditional within the sweep.

- **Sequential updates** condition on the latest values after each sub-step.
- **Simultaneous updates** draw every coordinate using only the old state x^t. This generally does *not* preserve the desired joint target.

Initialising a working vector with $z = x^t$ is convenient. What matters is that each conditional uses the current z_{-j} as the sweep proceeds.

One-coordinate updates and full sweeps

Each coordinate-update kernel K_j leaves the target invariant:

$$\pi K_j = \pi.$$

A full systematic sweep composes these kernels:

$$K = K_1 K_2 \cdots K_d, \qquad \pi K = ((\pi K_1) K_2) \cdots K_d = \pi.$$

- A fixed order may be changed, provided the sequential updates remain valid.
- A random coordinate choice with fixed, state-independent probabilities also preserves π.
- Updating one coordinate forever need not explore the full joint target.

Invariance alone does not establish convergence from an arbitrary starting point; suitable irreducibility and ergodicity conditions are still needed.

Gibbs sampling: key steps

1. Define x_{-j} and obtain a sampler for $\pi(x_j \mid x_{-j})$.
2. Update coordinate j, leaving every other coordinate fixed: Eq. (7.3).
3. Substitute the conditional proposal into the MH ratio.
4. Use joint $=$ conditional $\times$ marginal to obtain $a = 1$.
5. Cycle through the coordinates using their latest values.

What to retain

Gibbs sampling replaces a difficult joint draw by a sequence of conditional draws. The result is a dependent Markov chain with the desired invariant distribution.

A bivariate Gaussian example

Consider the joint Gaussian target:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \qquad |\rho| < 1.$$

Its density, up to a constant, is

$$\pi(x_1, x_2) \propto \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)} \right].$$

The two full conditionals are

$$X_1 \mid X_2 = x_2 \sim N(\rho x_2, 1 - \rho^2),$$

$$X_2 \mid X_1 = x_1 \sim N(\rho x_1, 1 - \rho^2).$$

Deriving the first Gaussian conditional

Hold x_2 fixed and complete the square:

$$x_1^2 - 2\rho x_1 x_2 + x_2^2 = (x_1 - \rho x_2)^2 + (1 - \rho^2)x_2^2.$$

The second term does not depend on x_1, so

$$\pi(x_1 \mid x_2) \propto \exp\left[-\frac{(x_1 - \rho x_2)^2}{2(1 - \rho^2)}\right].$$

Hence the conditional mean is ρx_2 and conditional variance is $1 - \rho^2$. A draw can be generated as

$$X_1 = \rho x_2 + \sqrt{1 - \rho^2}\, Z, \qquad Z \sim N(0, 1).$$

One numerical Gibbs sweep

Take $\rho = 0.8$, so the conditional standard deviation is $\sqrt{1-\rho^2} = 0.6$. Start at $x^0 = (2, -1)$. For illustration, use normal variates $z_1 = 0.5$ and $z_2 = 0.25$.

$$x_1^1 = 0.8 x_2^0 + 0.6 z_1 = 0.8(-1) + 0.6(0.5) = -0.5,$$

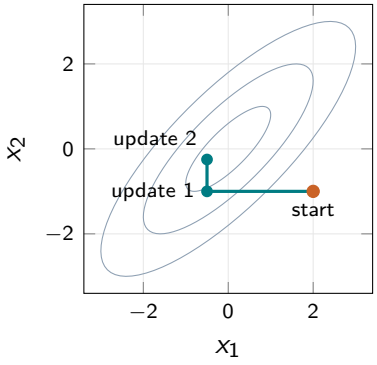
$$x_2^1 = 0.8 x_1^1 + 0.6 z_2 = 0.8(-0.5) + 0.6(0.25) = -0.25.$$

The working states are

$$(2, -1) \;\longrightarrow\; (-0.5, -1) \;\longrightarrow\; (-0.5, -0.25).$$

Both updates are accepted. Record $x^1 = (-0.5, -0.25)$ after the sweep.

The geometry of coordinate updates



Contours show the Gaussian target with $\rho = 0.8$.

- Update x_1: move horizontally while x_2 stays fixed.
- Update x_2: move vertically while x_1 stays fixed.
- The two segments together form one full sweep.

The displayed path is the worked sweep from the previous slide.

Why simultaneous updates can be wrong

Suppose (X_1, X_2) already has the Gaussian target, with covariance ρ. Consider simultaneous updates using fresh $Z_1, Z_2 \sim N(0,1)$, independent of each other and of the current state:

$$X_1' = \rho X_2 + \sqrt{1-\rho^2} Z_1, \qquad X_2' = \rho X_1 + \sqrt{1-\rho^2} Z_2.$$

Both marginal variances remain one, but

$$\mathrm{Cov}(X_1', X_2') = \rho^2 \,\mathrm{Cov}(X_2, X_1) = \rho^3.$$

For $\rho = 0.8$, this is 0.512, instead of the required 0.8.

Correct marginals do not ensure a correct joint distribution

This simultaneous kernel fails to preserve the target. Sequential Gibbs instead uses X_1' when updating X_2'.

Exercise: using the latest coordinate values

Use the same Gaussian model, now with $\rho = 0.6$ and initial state $x^0 = (1, -1)$. For this worked calculation, set $z_1 = z_2 = 0$.

1. Find the conditional standard deviation.
2. Update x_1 and then x_2 in one sequential Gibbs sweep.
3. What would an all-old update produce for the second coordinate?
4. Is an independent uniform draw needed for acceptance?

Record the intermediate state

$(1, -1) \longrightarrow (\,?\,, -1) \longrightarrow (\,?\,,\,?\,)$.

Exercise solution and mixing

The conditional standard deviation is $\sqrt{1 - 0.6^2} = 0.8$.

$$x_1^1 = 0.6(-1) + 0.8(0) = -0.6, \qquad x_2^1 = 0.6(-0.6) + 0.8(0) = -0.36.$$

The sequential result is $(-0.6, -0.36)$. An all-old update would use $x_1^0 = 1$ and give $x_2^1 = 0.6$ instead. No acceptance uniform is needed: both exact Gibbs draws have $a = 1$.

Acceptance is not a mixing diagnostic

As $|\rho|$ approaches one, the conditional variance $1 - \rho^2$ becomes small. Coordinate updates can explore a narrow diagonal target slowly, despite accepting every draw.

Metropolis-within-Gibbs

If some full conditionals cannot be sampled directly:

1. Use an exact Gibbs draw for coordinates with available conditional samplers.
2. For another coordinate j, hold all other coordinates fixed at their latest values z.
3. Perform an MH step targeting $\pi(x_j \mid z)$.
4. Continue to the next coordinate, then store the completed sweep.

The common principle

Each sub-update must preserve its conditional target. One MH step can be enough for a valid sub-update; it need not be an independent conditional draw.

The component-wise MH ratio

Let v be the current coordinate, z the fixed remaining coordinates, and propose $u \sim q_j(\cdot \mid v; z)$.

$$a_j = \min\left\{1, \frac{\pi(u \mid z) q_j(v \mid u; z)}{\pi(v \mid z) q_j(u \mid v; z)}\right\} = \min\left\{1, \frac{\pi(u, z) q_j(v \mid u; z)}{\pi(v, z) q_j(u \mid v; z)}\right\}.$$

- The normalising constant of the conditional cancels.
- If accepted, change only coordinate j; otherwise retain its current value.
- Exact Gibbs is the special choice $q_j(u \mid v; z) = \pi(u \mid z)$, giving $a_j = 1$.

All remaining coordinates stay fixed during this sub-step.

Summary

- A full conditional describes one coordinate given all the others.
- Gibbs proposes from the exact full conditional, so the MH ratio is one.
- A sweep updates each coordinate using the latest available values.
- Acceptance one does not imply independent draws or rapid mixing.
- Metropolis-within-Gibbs uses MH steps when direct conditional draws are unavailable.

The central idea

Construct a sampler for a joint distribution by composing updates that preserve its full conditionals.