Gibbs Sampling

Gibbs Sampling - 2D Case

Gibbs sampling is a Markov chain Monte Carlo (MCMC) algorithm used to sample from a multivariate probability distribution when direct sampling is challenging. It does so by iteratively sampling from the conditional distributions of each variable given the others. Assume we are interested in sampling from the joint distribution $\pi(x_1, x_2), x \in \mathbb{R}^2$

Systematic Scan Algorithm: Let $\left(X_1^{(1)}, X_2^{(1)}\right)$ be the initial state then iterate for $t = 2, 3, \ldots$

1. Sample $X_1^{(t)} \sim \pi_{X_1|X_2}\left(\cdot|X_2^{(t-1)}\right)$
2. Sample $X_2^{(t)} \sim \pi_{X_2|X_1}\left(\cdot|X_1^{(t)}\right)$

Questions?

Looking at the algorithm it is not immediately obvious that the target distribution π is indeed the stationary distribution of the Markov chain defined by the Gibbs sampler.

- Is the joint distribution uniquely specified by the conditional distributions?
- Does the Gibbs sampler provide a Markov chain with the correct stationary distribution?

The Hammersley-Clifford Theorem

Under the positivity condition, that is Support of joint $=$ Cartesian product of marginal supports, then the full conditional distributions *uniquely* determine the joint distribution. Then for any (z_1, z_2) in the support of π (meaning $\pi(z_1, z_2) > 0$), we have for $d = 2$:

$$\pi(x_1, x_2) \propto \frac{\pi_{X_1|X_2}(x_1|z_2)}{\pi_{X_1|X_2}(z_1|z_2)} \cdot \frac{\pi_{X_2|X_1}(x_2|x_1)}{\pi_{X_2|X_1}(z_2|x_1)}$$

Connection to Gibbs Sampling:

1. **Validates the method:** Alternating sampling from full conditionals targets the correct joint distribution
2. **Guarantees uniqueness:** When positivity holds, we know *which* distribution we're sampling from
3. **Ensures convergence:** The Markov chain has π as its stationary distribution

Transition Kernel of Gibbs Sampler

The transition kernel for a 2D systematic Gibbs sampler from state $x^{(t-1)}$ to state $x^{(t)}$ is:

$$K(x^{(t-1)}, x^{(t)}) = \pi_{X_1|X_2}(x_1^{(t)}|x_2^{(t-1)}) \cdot \pi_{X_2|X_1}(x_2^{(t)}|x_1^{(t)})$$

This represents the composition of two steps:

1. transition from $(x_1^{(t-1)}, x_2^{(t-1)})$ to $(x_1^{(t)}, x_2^{(t-1)})$
2. and then from $(x_1^{(t)}, x_2^{(t-1)})$ to $(x_1^{(t)}, x_2^{(t)})$

The kernel is the product of these conditional probabilities since the updates are performed sequentially within each iteration.

Invariance of the Target Distribution

Proof: $d = 2$ for points $x = (x_1, x_2)$ and $y = (y_1, y_2)$.

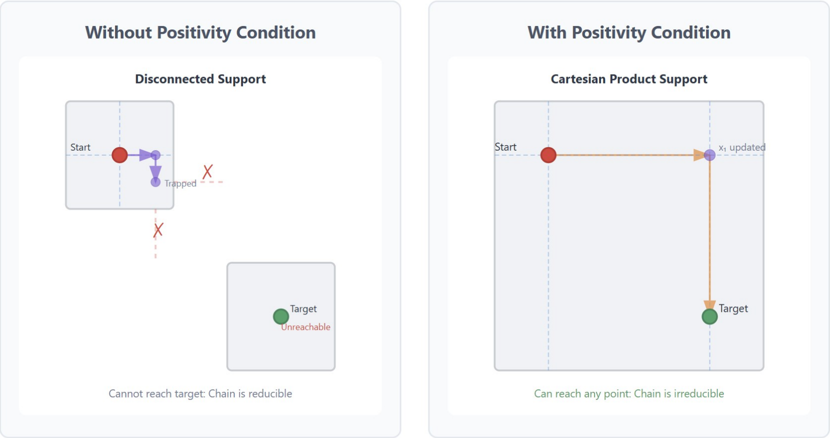
$$
\begin{aligned}
\int K(x,y)\pi(x)dx &= \int \pi(y_2|y_1)\pi(y_1|x_2)\pi(x_1, x_1)dx_1 dx_2 && (1)\\
&= \int \pi(y_2|y_1)\pi(y_1|x_2)\pi(x_2)dx_2 && (2)\\
&= \pi(y_2|y_1)\int \pi(y_1|x_2)\pi(x_2)dx_2 && (3)\\
&= \pi(y_2|y_1)\pi(y_1) && (4)\\
&= \pi(y_1, y_2) = \pi(y) && (5)
\end{aligned}
$$

1) Used definition of $K(x, y)$, 2) Marginalized out x_1, 3) Took out terms not depending on x_2, 4) Marginalized out x_2, 5) Used definition of conditional probability again

π-irreducible - visual argument

Gibbs Sampler Irreducibility under the Positivity Condition

Visual proof that π satisfying the positivity condition ensures Markov chain irreducibility



π-irreducible

Assume π satisfies the positivity condition, then the Gibbs sampler yields a π-irreducible Markov chain.

Irreducibility Write K for the transition kernel. We need to show that for any set $A \subset \mathbb{X}$ such that $\pi(A) > 0$, we have $K(x, A) > 0$ for any $x \in \mathbb{X}$. We have:

$$K(x, A) = \int_A K(x, y)\, dy = \int_A \pi_{X_1|X_2}(y_1 \mid x_2) \times \pi_{X_2|X_1}(y_2 \mid y_1)\, dy_1 dy_2$$

Proof by contradiction: Suppose that $K(x, A) = 0$ for some A with $\pi(A) > 0$. Then we must have:

$$\pi_{X_1|X_2}(y_1 \mid x_2) \times \pi_{X_2|X_1}(y_2 \mid y_1) = 0$$

for almost all $y = (y_1, y_2) \in A$.

π-irreducible cont and Recurrence

By the Hammersley-Clifford theorem, the joint distribution satisfies:

$$\pi(y_1, y_2) \propto \frac{\pi_{X_1|X_2}(y_1 \mid z_2)}{\pi_{X_1|X_2}(z_1 \mid z_2)} \times \frac{\pi_{X_2|X_1}(y_2 \mid y_1)}{\pi_{X_2|X_1}(z_2 \mid y_1)} = 0$$

for almost all $y = (y_1, y_2) \in A$. and hence implies $\pi(A) = 0$, which **contradicts** our assumption that $\pi(A) > 0$.

Recurrence: follows from irreducibility and the fact that π is invariant (see Meyn and Tweedie, Proposition 10.1.1.)

Convergence

Assume the Markov chain generated by the systematic scan Gibbs sampler is π-irreducible and recurrent (both conditions hold when the positivity condition is satisfied) then we have for any π-integrable function $\phi : \mathbb{X} \to \mathbb{R}$:

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \phi\left(X^{(i)}\right) = \int_{\mathbb{X}} \phi(x)\,\pi(x)\,dx$$

for π-almost all starting value $X^{(1)}$.