Gibbs Sampling

Gibbs Sampling with Systematic Scan Algorithm - 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

$$\pi(x) = \pi(x_1, x_2), \quad x \in \mathbb{R}^2.$$

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?
- If yes, does the Markov chain converge towards this invariant distribution?

The Hammersley-Clifford Theorem (1970)

Key Result: Under the positivity condition, the full conditional distributions *uniquely* determine the joint distribution.

For $d = 2$: $\pi(x_1, x_2) \propto \frac{\pi_{X_1|X_2}(x_1|x_2)}{\pi_{X_1|X_2}(z_1|x_2)} \cdot \frac{\pi_{X_2|X_1}(x_2|x_1)}{\pi_{X_2|X_1}(z_2|x_1)}$

Positivity Condition:

- Support of joint = Cartesian product of marginal supports
- If $\pi_{X_i}(x_i) > 0$ for all i, then $\pi(x_1, \ldots, x_d) > 0$

Warning: Not all conditionals are compatible!

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^{t-1}, y^{t-1}) to (x^t, y^{t-1})
2. transition from (x^t, y^{t-1}) to (x^t, y^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_1, y_1) and (x_2, y_2).

$$
\begin{aligned}
\int K((x_1, y_1) \to (x_2, y_2))\pi(x_1, y_1)dx_1 dy_1 &= \int \pi(x_2|y_1)\pi(y_2|x_2)\pi(x_1, y_1)dx_1 dy_1 && (1)\\
&= \int \pi(x_2|y_1)\pi(y_2|x_2)\pi(y_1)dy_1 && (2)\\
&= \pi(y_2|x_2) \int \pi(x_2|y_1)\pi(y_1)dy_1 && (3)\\
&= \pi(y_2|x_2)\pi(x_2) && (4)\\
&= \pi(x_2, y_2) && (5)
\end{aligned}
$$

π-irreducible

Assume $\pi(x_1, x_2)$ satisfies the positivity condition, then the Gibbs sampler yields a π-irreducible Markov chain.

Irreducibility Write K for the Gibbs sampler 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:

$$K(x, A) = \int_A K(x, y)\, dy = \int_A \pi_{X_1|X_2}(y_1 \mid x_2) \times \pi_{X_2|X_1}(y_2 \mid y_1)\, dy_1 dy_2 \tag{6}$$

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

$$\pi_{X_1|X_2}(y_1 \mid x_2) \times \pi_{X_2|X_1}(y_2 \mid y_1) = 0 \tag{7}$$

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

π-irreducible cont, Recurrence, and Convergence

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

$$\pi(y_1, y_2) \propto \frac{\pi_{X_1|X_2}(y_1 \mid x_2)}{\pi_{X_1|X_2}(\cdot \mid \cdot)} \times \frac{\pi_{X_2|X_1}(y_2 \mid y_1)}{\pi_{X_2|X_1}(\cdot \mid \cdot)} = 0 \tag{8}$$

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.)

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 integrable function $\phi : \mathbb{X} \to \mathbb{R}$:

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