VE414 Lecture 15

Jing Liu

UM-SJTU Joint Institute

October 29, 2019

Theorem

The joint distribution $f_{\mathbf{Y}}$ is the invariant distribution of the Markov Chain

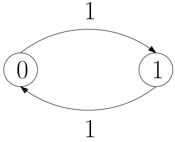
$$\{\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}, \dots\}$$

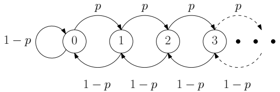
generated by the Gibbs sampling scheme, it is invariant in the sense that

$$\mathbf{Y}^{(t)} \sim f_{\mathbf{Y}} \qquad \text{whenever} \qquad \mathbf{Y}^{(t-1)} \sim f_{\mathbf{Y}}$$

Proof

- Note the above theorem does not guarantee a sample generated by Gibbs follows the joint, it merely states the joint is the invariant distribution.
- Recall the following Markov chains again, they do not converge





- We know being irreducible, recurrent and aperiodic guarantee convergence.

Theorem

Suppose the joint probability density function is positive

$$f_{\{Y_1,\ldots,Y_p\}}\left(y_1,\ldots y_p\right)>0$$

for all $y_1,\ldots,y_p$ when the marginal probability density functions are positive

$$f_{Y_i}\left(y_i\right)>0$$

then the sequence

$$\{f_{\mathbf{Y}^{(1)}},f_{\mathbf{Y}^{(2)}},\ldots\}$$

corresponding to the Gibbs sampling converges to $f_{\mathbf{Y}}$ for every $\mathbf{y}_0\in\mathcal{D}$, and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}h\left(\mathbf{Y}^{(t)}\right)\to\mathbb{E}\left[h\left(\mathbf{Y}\right)\right]$$

provided the transition kernel $\kappa\left(\mathbf{y}^{(t-1)},\mathbf{y}^{(t)}\right)$ is absolutely continuous.

- Unfortunately, proving the last theorem is beyond the scope of this course.
- However, a lot of our understanding on Markov Chains with discrete state space can be extended so that we can have a roughly idea why Gibbs works.

Definition

Given a distribution ϕ on the state space $\mathcal{D}$, then the term ϕ-irreducible is used to describe a Markov chain for which every state with positive probability under ϕ

$$\operatorname{supp}(\phi) = \{x \in \mathcal{D} \colon \phi(x) > 0\}$$

communicates with every other state. In terms of a continuous state space $\mathcal{D}$, it means for all $x \in \mathcal{D}$ and all measurable sets $\mathcal{A}$ such that $\phi(\mathcal{A}) > 0$, for which

$$\int_{\mathcal{A}} \kappa^t(x, y)\, dy > 0 \qquad \text{for some } t$$

If it holds for $t = 1$, then the chain is said to be strongly ϕ-irreducible.

Theorem

If the joint probability density function is positive

$$f_{\{Y_1,\ldots,Y_p\}}\left(y_1,\ldots y_p\right)>0$$

for all $y_1,\ldots,y_p$ when the marginal probability density functions are positive

$$f_{Y_i}\left(y_i\right)>0$$

then the Markov chain defined by Gibbs sampling is strongly f-irreducible.

- The idea of periodicity becomes really technical in a continuous state space, but the essential idea remains the same, roughly speaking, it is aperiodic if the chain cannot get stuck in a loop of subsets of $\mathcal{D}$ indefinitely.
- It can be shown using the updating formula the Markov chain generated by Gibbs sampling is aperiodic if it is irreducible.

- It can also be shown a continuous state space Markov chain is recurrent if

1. The chain is f-irreducible for some distribution f.
2. For every measurable set $\mathcal{A} \in \mathcal{D}$ such that

$$\int_{\mathcal{A}} f(\mathbf{y})\, d\mathbf{y} > 0$$

   the expected number of times to reach $\mathcal{A}$ is infinite for every $x \in \mathcal{A}$.

- Since f is invariant, it is clear that we will visit $\mathcal{A}$ infinitely many time.

Q: Given the above discussion on the convergence of Gibbs sampling, can you think of an example that Gibbs sampling will fail sample from $f_{\mathbf{Y}}$?

$$\begin{aligned}
&\mathbf{Y} \sim \mathrm{Uniform}\left(\mathcal{C}_1 \cup \mathcal{C}_2\right) \\
&\mathcal{C}_1 = \{(x_1, x_2) \mid (x_1 - 1)^2 + (x_2 - 1)^2 \leq 1\} \\
&\mathcal{C}_2 = \{(x_1, x_2) \mid (x_1 + 1)^2 + (x_2 + 1)^2 \leq 1\}
\end{aligned}$$

- Given the above convergence result, how can we obtain an estimate of

$$\mathbb{E}\left[h\left(\mathbf{Y}\right)\right]$$

where $h\colon \mathcal{D} \to \mathbb{R}$ is integrable, using samples from Gibbs sampling.

- We could take n samples after many Gibbs iterations, say m, and expect

$$\mathbb{E}\left[h\left(\mathbf{Y}\right)\right] \approx \frac{1}{n}\sum_{t=m}^{m+n} h\left(\mathbf{Y}^{(t)}\right)$$

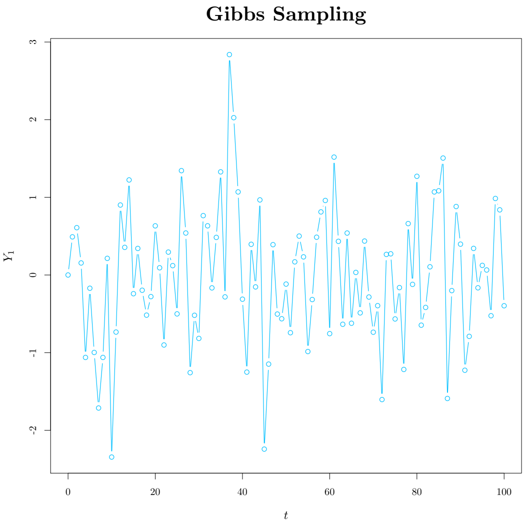
- Alternatively, we could construct n Markov Chains using Gibbs sampling,

$$\mathbb{E}\left[h\left(\mathbf{Y}\right)\right] \approx \frac{1}{n}\sum_{j=1}^{n} h\left(\mathbf{Y}_j^{(k_j)}\right)$$

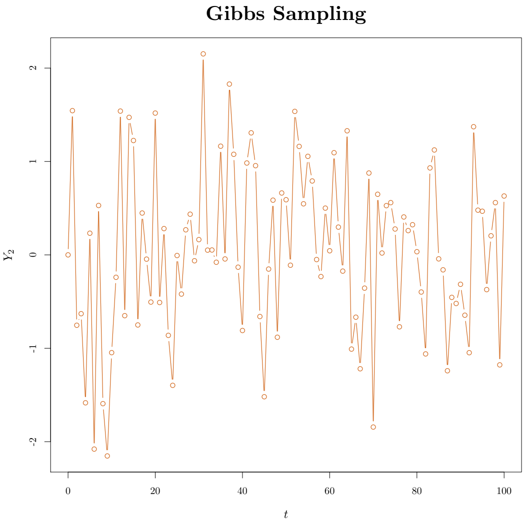
where only the last value $\mathbf{Y}_j^{(k_j)}$ of each chain is used.

Q: How to perform Gibbs sampling on the Bivariate normal?

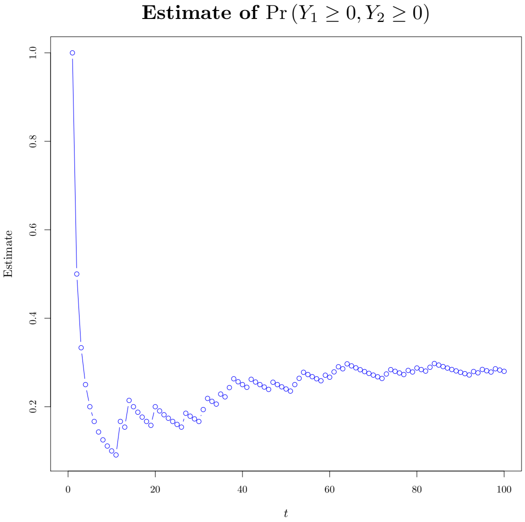
Q: How can we determine whether we have reached the invariant distribution?



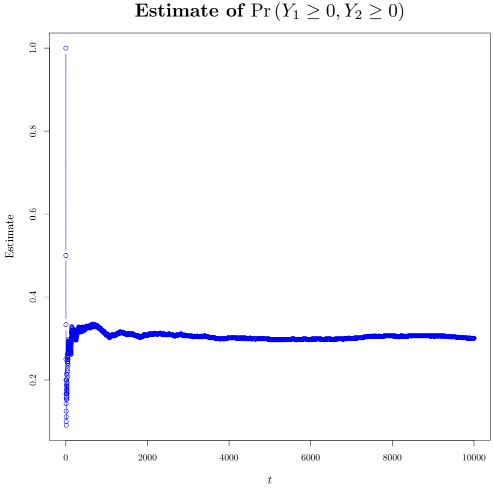
- Various plots based on the sample are usually the way to check.



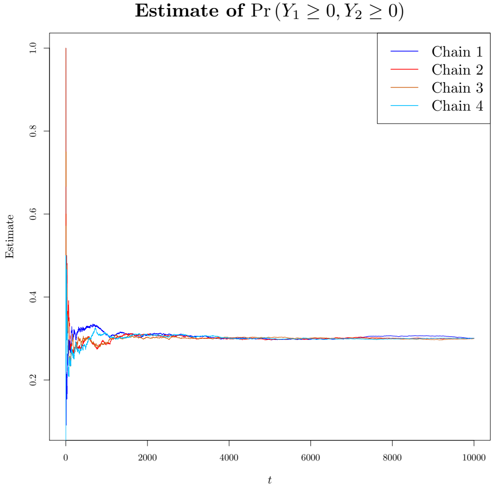
Q: How to estimate the probability $\Pr(Y_1 \geq 0, Y_2 \geq 0)$ base on the sample?



- The last plot suggests the chain is yet to converge, we need a bigger n.



- Of course, we can generate multiple chains using Gibbs sampling.



- In practice, a few chains are run, and each took a certain burn-in period.