Gibbs Sampling

Carlo Hämäläinen
carlo.hamalainen@gmail.com

November 18, 2009

This file and related Sage source code is available at http://carlo-hamalainen.net/stuff/gibbs

Suppose that X and Y are two random variables. The Gibbs sampling algorithm gives us a way to sample from $f(x)$ by sampling from the conditional distributions $f(x \mid y)$ and $f(y \mid x)$. In statistical models we often know the conditional distribution while the marginal distribution $f(x)$ is difficult to compute.

In Gibbs sampling we generate a series of random variables

$$Y_0', X_0', Y_1', X_1', Y_2', X_2', \dots Y_k', X_k'$$

where $Y_0' = y_0$ is fixed and the rest of the variables are sampled according to

$$\begin{aligned} X_j' &\sim f(x \mid Y_j' = y_j') \\ Y_{j+1}' &\sim f(y \mid X_j' = x_j'). \end{aligned}$$

If k is large enough then X_k' will effectively be a sample from the marginal distribution $f(x)$.

We can work through the two variable case with an explicit example, following [1]. Let X and Y each be (marginally) Bernoulli random variables with joint distribution as follows:

	$X = 0$	$X = 1$
$Y = 0$	p_1	p_2
$Y = 1$	p_3	p_4

where $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. Using this table we can calculate the conditional distributions. For example,

$$P(X = 1 \mid Y = 1) = \frac{P(X = 1 \text{ and } Y = 1)}{P(Y = 1)} = p_4/(p_3 + p_4).$$

The matrix $A_{y|x}$ gives the conditional probability of Y given $X = x$:

$$A_{y|x} = \begin{pmatrix} \frac{p_1}{(p_1+p_3)} & \frac{p_3}{(p_1+p_3)} \\ \frac{p_2}{(p_2+p_4)} & \frac{p_4}{(p_2+p_4)} \end{pmatrix}$$

The matrix $A_{x|y}$ gives the conditional probability of X given $Y = y$:

$$A_{x|y} = \begin{pmatrix} \frac{p_1}{(p_1+p_2)} & \frac{p_2}{(p_1+p_2)} \\ \frac{p_3}{(p_3+p_4)} & \frac{p_4}{(p_3+p_4)} \end{pmatrix}$$

The transition $X_0' \to Y_1' \to X_1'$ has probability

$$P(X_1' = x_1 \mid X_0' = x_0) = \sum_y P(X_1' = x_1 \mid Y_1' = y)P(Y_1' = y \mid X_0' = x_0).$$

So the matrix $A_{x|x}$ describing the transitions $X_0' \to X_1'$ is

$$A = A_{x|x} = A_{y|x}A_{x|y} = \begin{pmatrix} \frac{p_3^2}{(p_3+p_4)(p_1+p_3)} + \frac{p_1^2}{(p_1+p_3)(p_1+p_2)} & \frac{p_3p_4}{(p_3+p_4)(p_1+p_3)} + \frac{p_1p_2}{(p_1+p_3)(p_1+p_2)} \\ \frac{p_3p_4}{(p_3+p_4)(p_2+p_4)} + \frac{p_1p_2}{(p_2+p_4)(p_1+p_2)} & \frac{p_4^2}{(p_3+p_4)(p_2+p_4)} + \frac{p_2^2}{(p_2+p_4)(p_1+p_2)}. \end{pmatrix}$$

We now have a Markov chain with two states, 0 and 1, and transition probabilities given by the matrix A. Note that $A_{0,0} + A_{0,1} = 1$ and $A_{1,0} + A_{1,1} = 1$ since A is a stochastic matrix.

The stationary distribution is given by the normalised eigenvector f, where $fA = f$. For our the 2×2 case the f vector is:

$$f = \begin{pmatrix} \frac{(\frac{(p_2+p_4)p_1}{(p_1+p_2+p_3+p_4)} + \frac{(p_2+p_4)p_3}{(p_1+p_2+p_3+p_4)})}{(p_2+p_4)} & \frac{(p_2+p_4)}{(p_1+p_2+p_3+p_4)} \end{pmatrix}$$

Observe that f is undefined if $p_2 + p_4 = 0$. In terms of the original joint distribution, this means that there is zero probability of reaching a state $(1, Y)$ for any Y. So this case is in some sense degenerate.

For an explicit example, set

$$p_1 = 0.26275562241164158$$
$$p_2 = 0.6960509654605056$$
$$p_3 = 0.025834046671036285$$
$$p_4 = 0.015359365456816687$$

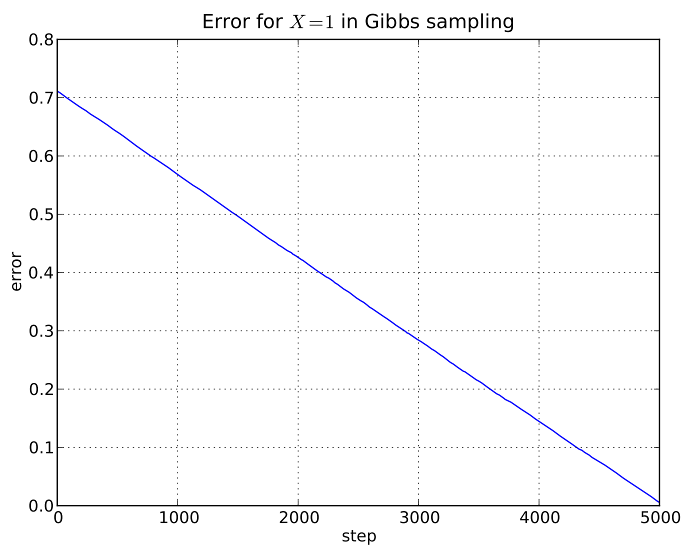
Then

$$A = \begin{pmatrix} 0.305652971862979 & 0.694347028137022 \\ 0.281667794759494 & 0.718332205240506 \end{pmatrix}$$

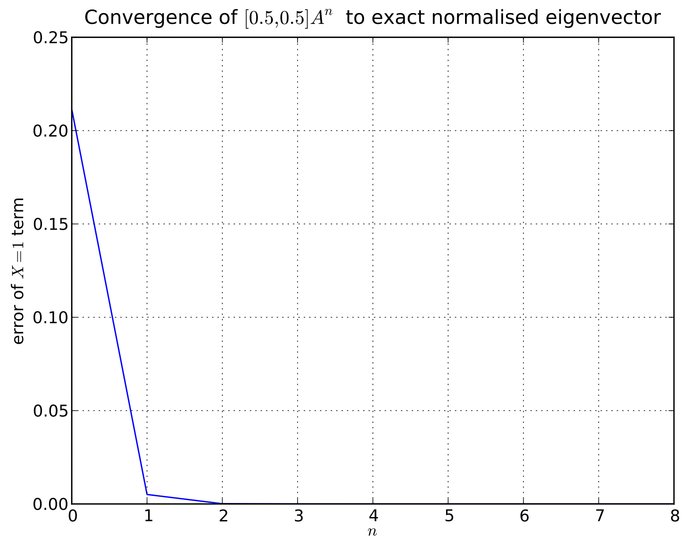
and

$$f = \begin{pmatrix} 0.288589669082678 & 0.711410330917322 \end{pmatrix}$$

The error in the stationary distribution of $X = 1$ is given in the following plot:



With the same matrices, we can also compare the error in the $X = 1$ stationary probability, starting from an initial distribution of $f_0 = [0.5 \;\; 0.5]$:



References

[1] George Casella and Edward I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992. (document)