Explanatory Notes 6 for Bayesian Inference

The Gibbs sampler

Takis Konstantopoulos, Spring 2009

This is a method for simulating from a probability on a product space. To explain it, suppose that S_1, S_2 be discrete sets and let π be a probability on $S_1 \times S_2$. In other words, we are given a collection of numbers

$$\pi(x_1, x_2) \geq 0, \qquad \text{with} \sum_{x_1 \in S_1} \sum_{x_2 \in S_2} \pi(x_1, x_2) = 1.$$

We denote the marginal laws as follows:

$$\pi_1(x_1) = \sum_{x_2 \in S_2} \pi(x_1, x_2), \quad \pi_2(x_2) = \sum_{x_1 \in S_1} \pi(x_1, x_2).$$

We denote the conditional laws as follows:

$$\pi_{12}(x_1|x_2) = \frac{\pi(x_1, x_2)}{\pi_2(x_2)}, \quad \pi_{21}(x_2|x_1) = \frac{\pi(x_2, x_1)}{\pi_1(x_1)}.$$

We define a stochastic process

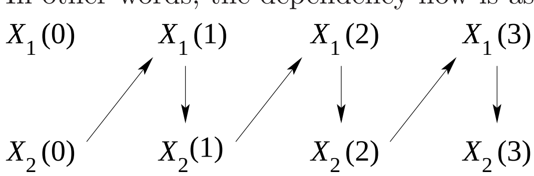
$$\boldsymbol{X}(n) = (X_1(n), X_2(n)), \quad n = 0, 1, \ldots,$$

with values in $S_1 \times S_2$, by starting with a fairly arbitrary $\boldsymbol{X}(0) = (X_1(0), X_2(0))$, and, by updating, at each step $n \geq 0$, as follows:

$$P(X_1(n+1) = x_1 | X_2(n) = x_2, X_1(n), \boldsymbol{X}(n-1), \ldots, \boldsymbol{X}(0)) = \pi_{12}(x_1|x_2)$$

$$P(X_2(n+1) = x_2 | X_1(n+1) = x_1, \boldsymbol{X}(n), \ldots, \boldsymbol{X}(0)) = \pi_{21}(x_2|x_1).$$

In other words, the dependency flow is as in the scheme



in the sense that a random variable at the tip of the arrow depends only on the one at the beginning of the same arrow and not on the ones before.

We can easily show the following:

Lemma 1. *If $\boldsymbol{X}(0)$ has law π then $\boldsymbol{X}(n)$ has law π for all $n \geq 1$.*

Proof. Use induction. Assume that $\boldsymbol{X}(n)$ has law π and prove that so does $\boldsymbol{X}(n+1)$. We have

$$\begin{aligned}
P(X_1(n+1) = x_1, &X_2(n+1) = x_2) \\
&= P(X_2(n+1) = x_2 | X_1(n+1) = x_1) P(X_1(n+1) = x_1) \\
&= \pi_{21}(x_2|x_1) \sum_{x_2' \in S} P(X_1(n+1) = x_1 | X_2(n) = x_2') P(X_2(n) = x_2') \\
&= \pi_{21}(x_2|x_1) \sum_{x_2' \in S} \pi_{12}(x_1|x_2') \pi_2(x_2') \\
&= \pi_{21}(x_2|x_1) \sum_{x_2' \in S} \pi(x_1, x_2') \\
&= \pi_{21}(x_2|x_1) \pi_1(x_1) \\
&= \pi(x_1, x_2).
\end{aligned}$$

□

Thus, the process $(\boldsymbol{X}(n), n \geq 0)$ is stationary if we choose $\boldsymbol{X}(0)$ to have law π.

Now, under irreducibility and aperiodicity assumptions we can show that

$$\lim_{n \to \infty} P(\boldsymbol{X}(n) = (x_1, x_2)) = \pi(x_1, x_2),$$

and so we can obtain a sample of π, approximately, by simulating the process long enough.

A pseudo-code for this is as follows:

```
Compute the conditional distribution p12(x1|x2)
Compute the conditional distribution p21(x2|x1)
N = 50000   # number of iterations
x1 = c(1:N)
x2 = c(1:N)
x1[1] = some reasonable value
x2[1] = some reasonable value
for(i in 1:(N-1)){
Generate x1[i+1] from p12(\cdot | x2[i])
Generate x2[i+1] from p21(\cdot | x1[i+1])
}
Return (x1[N], x1[N]) as an approximate sample from pi
```

Note that, since we expect, usually, to approximate *pi* long before the number N, we can use the whole trajectory as approximate data from π (throwing away the first small part of transient behaviour which, usually, won't hurt us anyway).

The method works for continuous distributions as well, with obvious modifications as regards conditional densities instead of conditional probabilities.

We explain it better in

An example. Suppose that we want to obtain a sample from a bivariate normal random variable (X, Y). To make things interesting assume that $\operatorname{cov}(X, Y) \neq 0$. For example, assume that

$$X = U + V, \quad Y = 3U + V,$$

where U, V are i.i.d. standard normals. Thus

$$EX = 0, \quad EY = 0$$
$$\operatorname{var} X = EX^2 = 2, \quad \operatorname{var} Y = EY^2 = 10, \quad \operatorname{cov}(X, Y) = EXY = 4$$

We compute the conditional densities of X given Y and of Y given X. (See Explanatory Notes 1.) We have that X given Y is normal with mean $E(X|Y)$ and deterministic variance equal to $\operatorname{var}(X - E(X|Y))$. To find $E(X|Y)$ we write

$$E(X|Y) = \lambda Y,$$

for some constant λ, which is comuted by the fact that $X - \lambda Y$ and Y are uncorrelated:

$$0 = E(X - \lambda Y)Y = EXY - \lambda EY^2 = 4 - 10\lambda, \quad \lambda = 4/10 = 2/5.$$

We then have

$$\operatorname{var}(X|Y) = E(X - E(X|Y))^2 = E(X - \tfrac{2}{5}Y)^2$$
$$= E((U + V) - \tfrac{2}{5}(3U + V))^2 = E(-\tfrac{1}{5}U + \tfrac{3}{5}V) = \frac{1}{25} + \frac{9}{25} = \frac{10}{25} = \frac{2}{5}.$$

We thus have

$$(X|Y) \text{ has law } N((2/5)Y, 2/5).$$

To find $E(Y|X)$ we write

$$E(Y|X) = \lambda' X,$$

for some constant λ', computed by

$$0 = E(Y - \lambda' X)X = 4 - 2\lambda', \quad \lambda' = 2.$$

Thus $E(Y|X) = 2X$. Also,

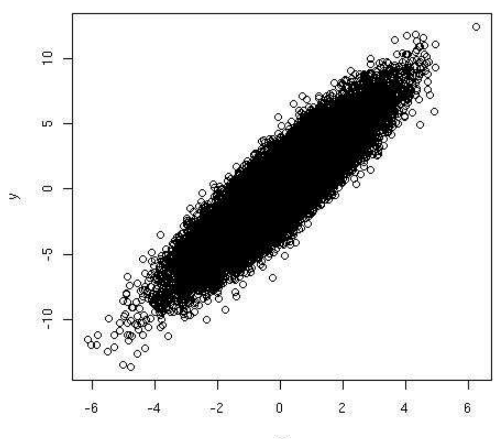
$$\mathrm{var}(Y|X) = E(Y - 2X)^2 = EY^2 + 4EX^2 - 4EXY = 10 + 8 - 16 = 2.$$

Thus

$$(Y|X) \text{ has law } N(2X, 2).$$

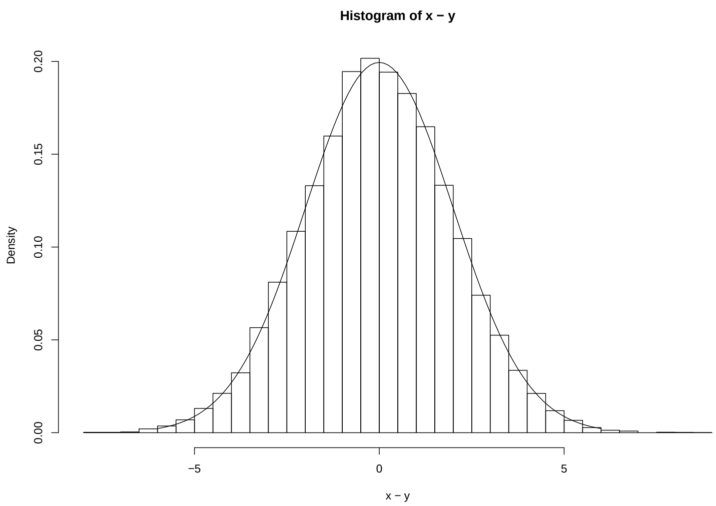
```
N=30000
x=c(1:N)
y=c(1:N)
x[1]=0
y[1]=0
for(i in 1:(N-1)){
y[i+1] = rnorm(1, 2*x[i], sqrt(2))
x[i+1] = rnorm(1, (2/5)*y[i+1], sqrt(2/5))
}
```

We check what we found
mean(x) = 0.01679832
var(x) = 2.005077
mean(y) = 0.04314908
var(y) = 10.01576
cov(x,y) = 4.007448
all in close approximation with the needed moments. Scatter Plot:



Actually, the scatter plot shows the positions of the simulated Markov chain on the plane. Notice how the clustering of the points represents a typical slice of the theoretical normal density.

Another test: We know that $X-Y$ has mean 0 and variance $E(-2U)^2 = 4$. Let's then do the histogram of $X-Y$, together with the normal density of $N(0,4)$:



This is a good fit.