Computational statistics - Homework 5

Markov Chain Monte Carlo methods

Matteo Silvestri

Professor Maarten Jansen

2024

Table des matières

1 Problem

The MATLAB routine available at the Github repository implements a Gibbs sampler for a vector of two random variables X_1, X_2 for which it is known that the conditional probabilities have a bounded Poisson law given by

$$\mathbb{P}(X_1 = x | X_2 = u) = \mathbb{P}(X_2 = x | X_1 = u) = \frac{e^{-(u+a)}(u+a)^x}{x!}$$

for $x \in \{0, 1, ..., b-1\}$ and

$$\mathbb{P}(X_1 = b | X_2 = u) = \mathbb{P}(X_2 = b | X_1 = u) = 1 - \sum_{k=0}^{b-1} \frac{e^{-(u+a)}(u+a)^k}{k!} .$$

The routine plots the empirical probability mass function $\widehat{p}_{X_1}(x)$ and (graphically) checks that the empirical cumulative distribution functions $\widehat{F}_{X_1}(x)$ and $\widehat{F}_{X_2}(x)$ are almost the same.

(a) When a = 0, the Gibbs sampler fails. Explain what happens, using for instance plot(xx1).

(b) When $b \to \infty$, the joint and marginal probability mass functions degenerate. Illustrate and explain, using plot(xx1). More precisely, prove that the expected values $\mathbb{E}(X_1)$ and $\mathbb{E}(X_2)$ must be infinite (hint : use the law of iterated expected values).

2 Solution

Figure 2.1 shows that the empirical CDFs for the two variables are almost numerically equal. While Figure 2.2, 2.3 represents the probability mass functions, which are also very similar because of the simmetry and indeed we obtain CDF using the MATLAB routine : `FX1hat = cumsum(pX1hat);` For these plots, we used the parameters $a = 0.5$ and $b = 100$.

(a) If a = 0, this implies that X_1, X_2 are exactly bounded Poisson with parameter u. In particular, we have that if $X_2 = 0$, then $\mathbb{P}(X_1 = x|X_2 = u) = 0$ with probability 1. This is not a big problem for the model, as it may happen that for some value the conditional probability degenerates. In this case, however, it becomes a catastrophe for the Gibbs sampler, because we generate X_2 in the same way using X_1. This means that as soon as X_2 takes value of 0 then X_1 takes value of 0 and both variables remain null all the time. We represent the variables values during the sampling process in figure 2.4; we can see how after 9 samples we get stuck on 0 values.

(b) If $b \to \infty$, then

$$\mathbb{E}\left[X_1|X_2 = \mu\right] \sim \text{Poisson}(\mu + a)$$

and it is the same for $\mathbb{E}\left[X_2|X_1 = \mu\right]$, so we have two Poisson distributions. This seems to be the perfect situation in which we can apply Hammersley-Clifford theorem, i.e. we can compute joint distribution from the conditional ones and obtain from those the marginal distributions. But we can see in the Figure 2.5 that something it is not working, and this is related to the fact that the marginal distributions of X_1, X_2 does not exist in this case, because they are degenerating. Indeed, if we apply the law of iterated expected values, it must be true that

$$\begin{cases} \mathbb{E}[X_1] = \mathbb{E}[\mathbb{E}[X_1|X_2]] = \mathbb{E}[X_2 + a] = \mathbb{E}[X_2] + a \\ \mathbb{E}[X_2] = \mathbb{E}[\mathbb{E}[X_2|X_1]] = \mathbb{E}[X_1 + a] = \mathbb{E}[X_1] + a \end{cases}$$

and this is satisfied if and only if $\mathbb{E}[X_1], \mathbb{E}[X_2]$ are infinite. This result is confirmed numerically by the Figure 2.6.

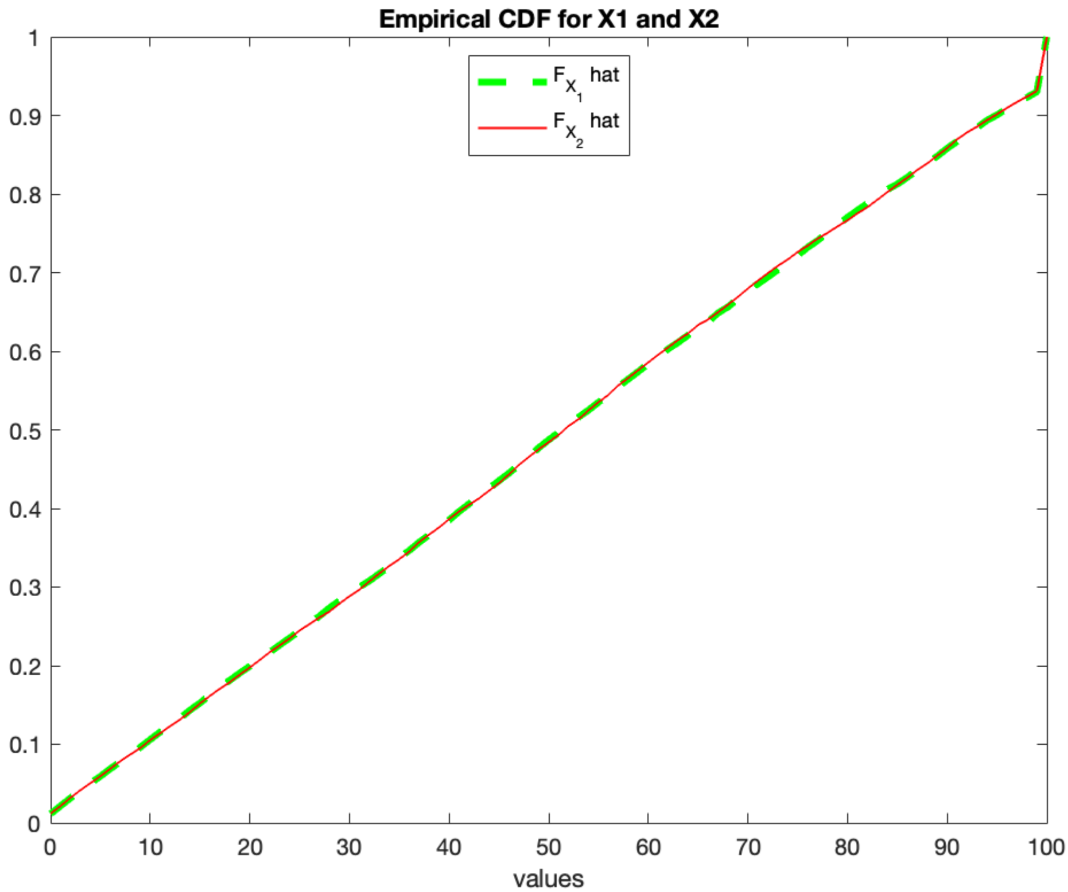


FIGURE 2.1 – Plot of the CDFs

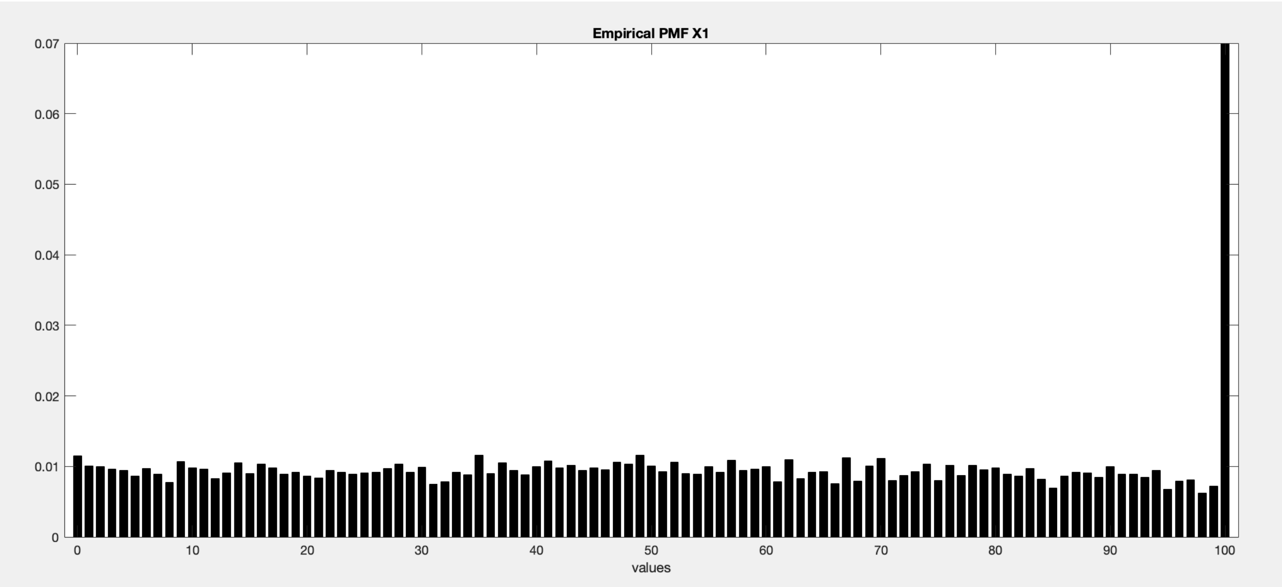


FIGURE 2.2 – Plot of the Probability Mass Functions for X_1

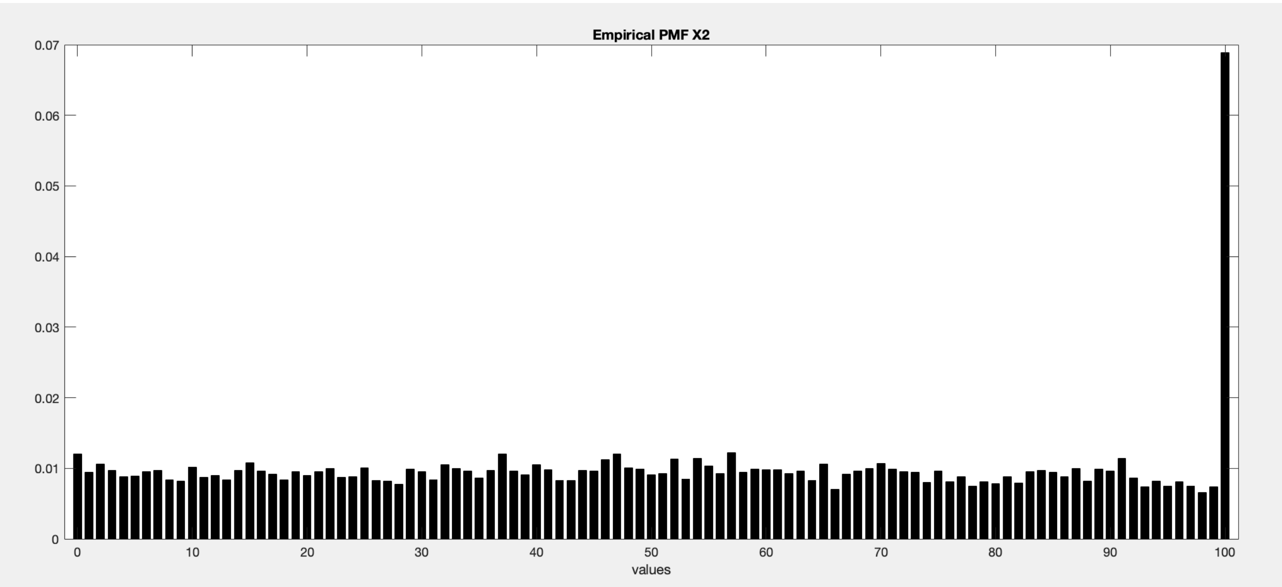
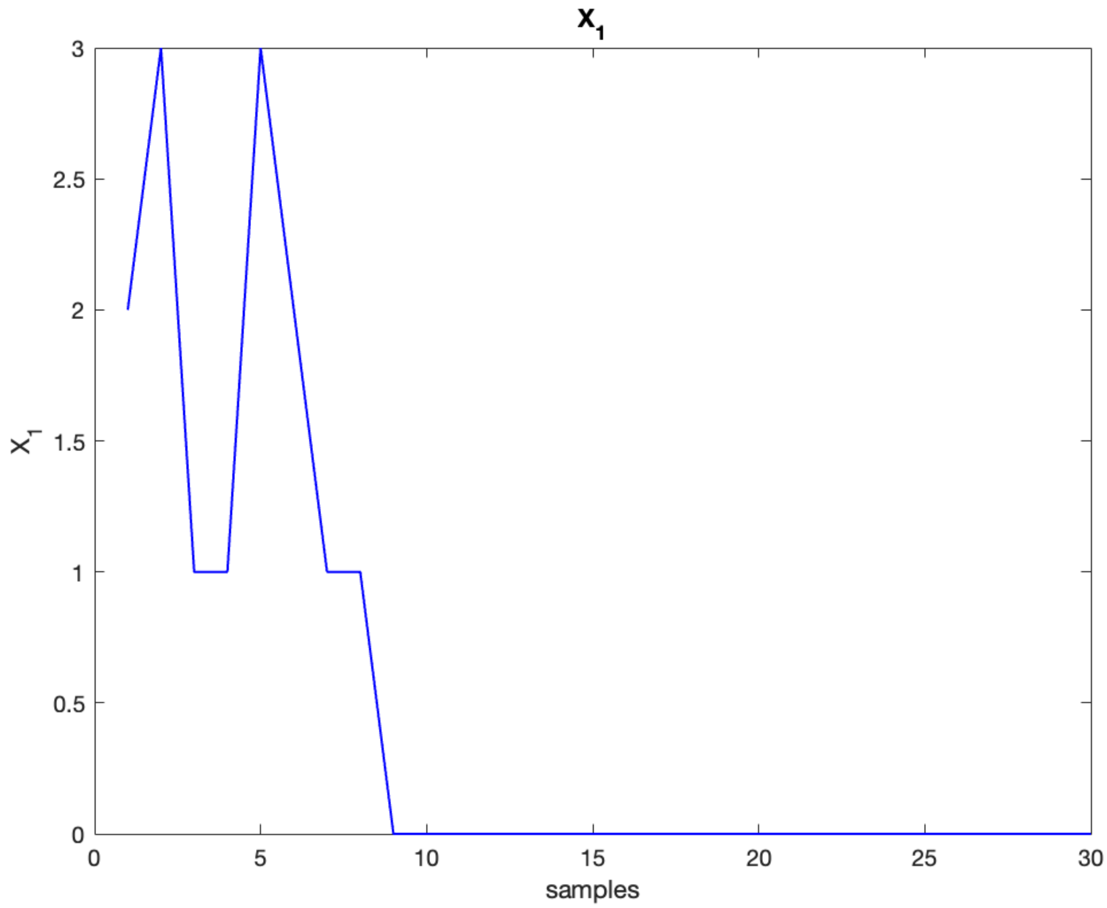


FIGURE 2.3 – Plot of the Probability Mass Functions for X_2



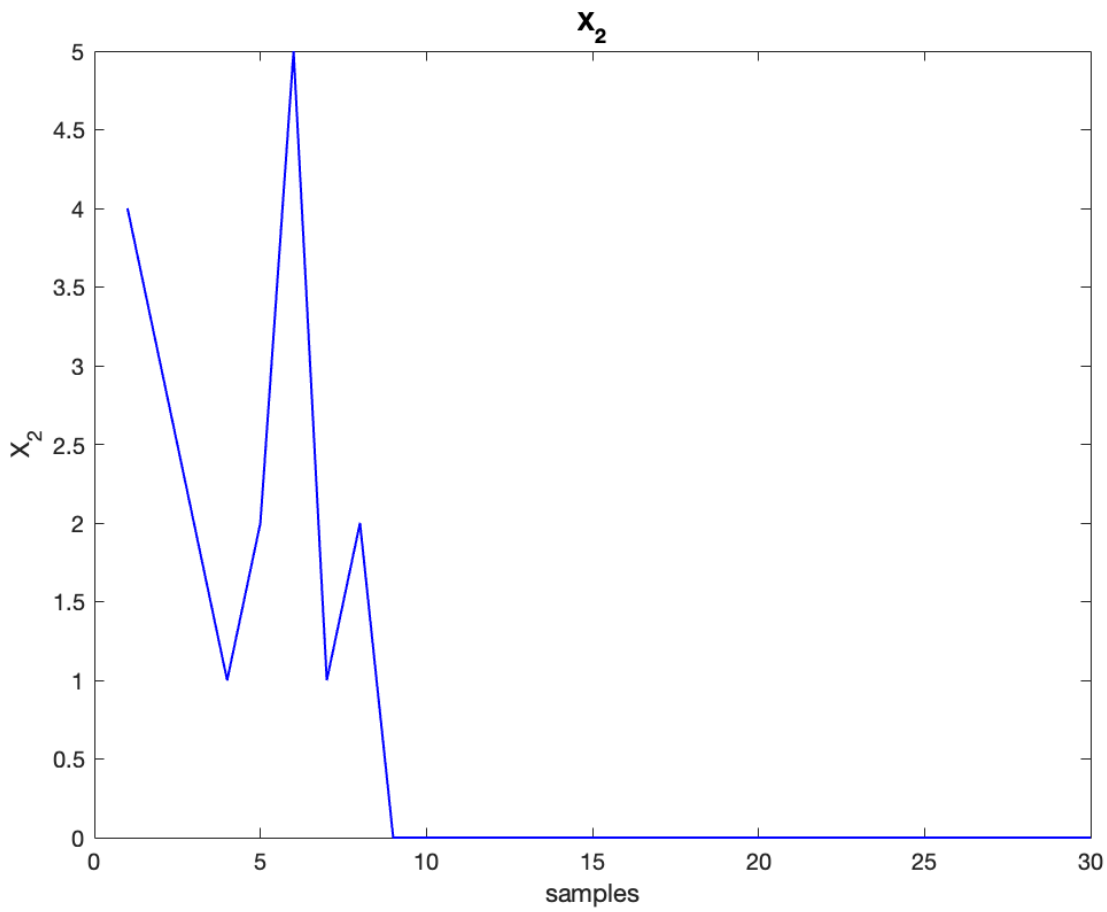


FIGURE 2.4 – X_1, X_2 values for the first 30 samples with a = 0

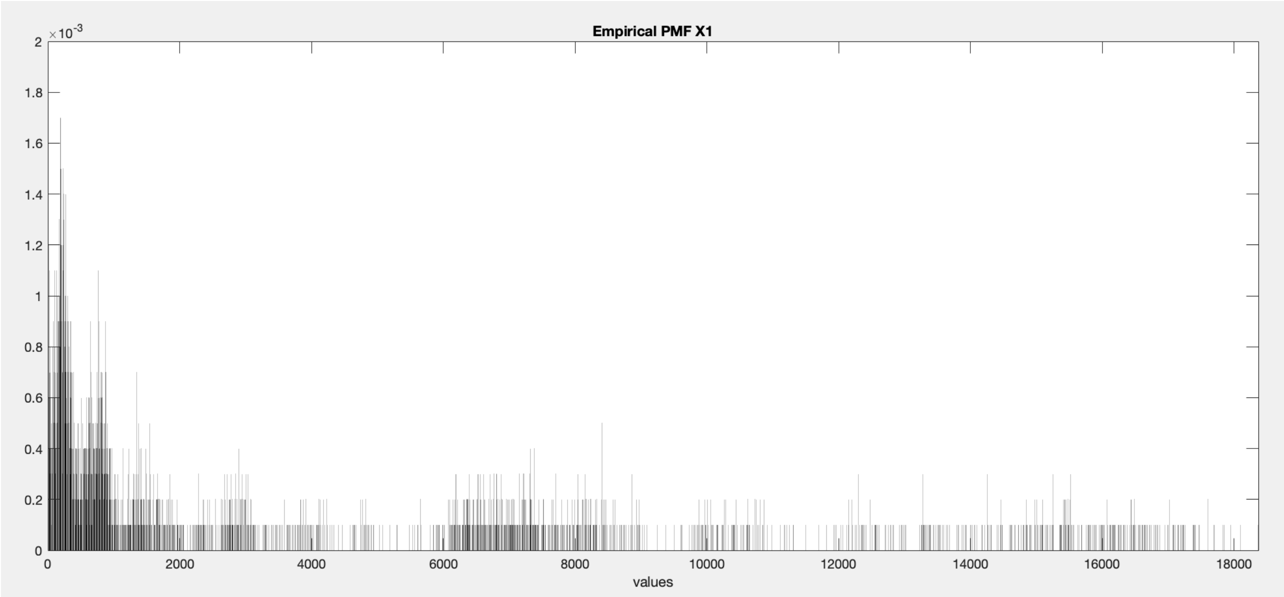


FIGURE 2.5 – Plot of the Probability Mass Functions for X_1 with $b = \infty$

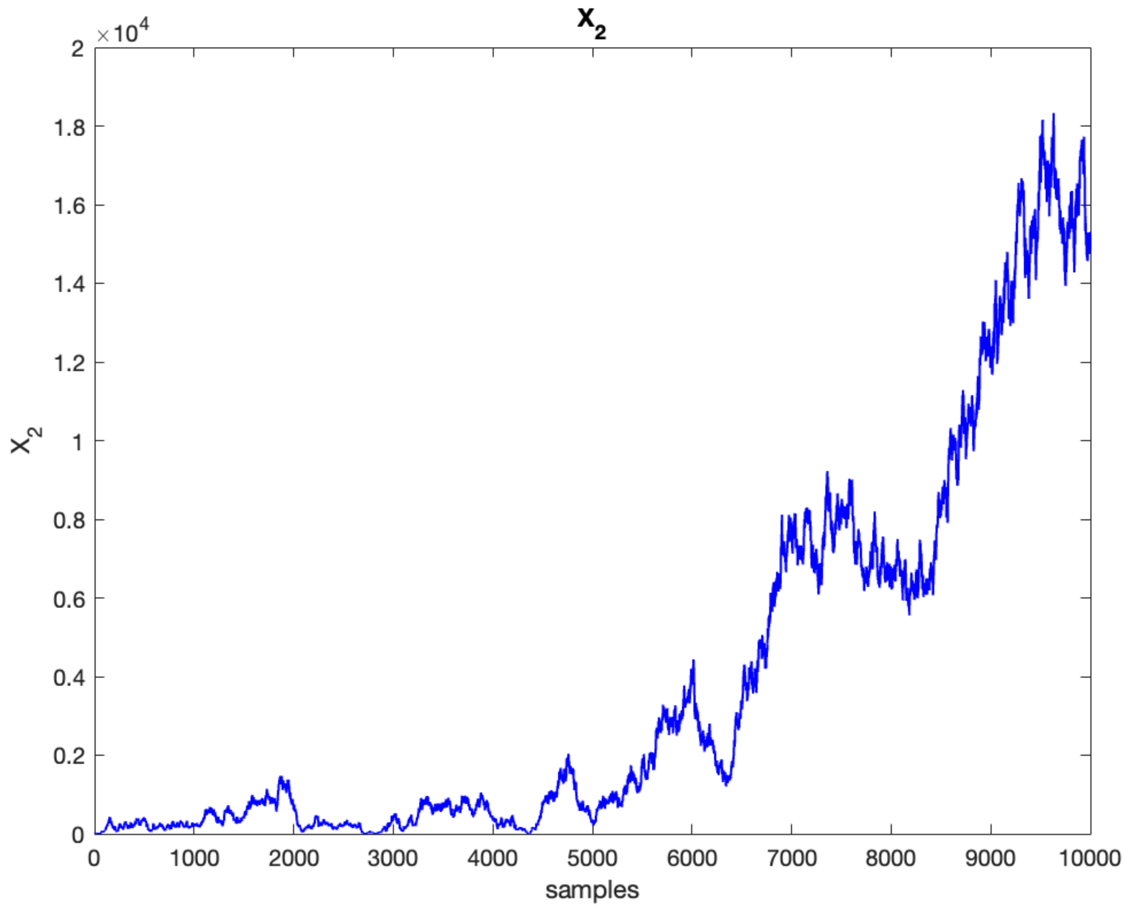


FIGURE 2.6 – Plot of theX_2 values with $b = \infty$