The Gibbs Sampler

Peadar Coyle

1. Introduction

We know that, using importance sampling, we can approximate an expectation $\mathbb{E}_f(h(X))$ without having to sample directly from f. However, finding an instrumental distribution which allows us to *efficiently* estimate $\mathbb{E}_f(h(X))$ can be difficult, especially in large dimensions. In this chapter and the following chapters we will use a somewhat different approach. We will discuss methods that alow obtaining an *approximate* sample from f without having to sample f directly. More mathematically speaking, we will discuss methods that generate a Markov chain whose *stationary distribution* is the distribution of interest f. Such methods are often called MCMC methods. Let us state a few definitions before continuning.

Definition The prior distribution is a key part of Bayesian inference and represents the information about an uncertain parameter θ that is combined with the probability distribution of new data to yield the *posterior distribution*.

Poisson change point model . Assume the following Poisson model of two regimes for n random variables $Y_1, \cdots, Y_n$[1]

$$Y_i \sim \texttt{Poi}(\lambda_1) \; for \; i = 1, \cdots, M$$

$$Y_i \sim \texttt{Poi}(\lambda_2) \; for \; i \; = \; + 1, \cdots, n$$

A suitable (conjugate) prior distibution for λ_j is the $\texttt{Gamma}(\alpha_j, \beta_j)$ distribution with density

$$f(\lambda_j) = \frac{1}{\Gamma(\alpha_j)} \lambda_j^{\alpha_j - 1} \beta_j^{\alpha_j} \exp(-\beta_j \lambda_j)$$

The joint distribution of $Y_1, \cdots, Y_n$, λ_1, λ_2, and M is

$$f(y_1, \cdots, y_n, \lambda_1, \lambda_2, M) = \left(\prod_{i=1}^{M} \frac{\exp(-\lambda_1) \lambda_1^{y_i}}{y_i!} \right) \cdot \left(\prod_{i=1}^{M} \frac{\exp(-\lambda_1) \lambda_1^{y_i}}{y_i!} \right)$$
$$\cdot \frac{1}{\Gamma(\alpha_1)} \lambda_1^{\alpha_1 - 1} \beta_1^{\alpha_1} \exp(-\beta_1 \lambda_1) \cdot \frac{1}{\Gamma(\alpha_2)} \lambda_2^{\alpha_2 - 1} \beta_2^{\alpha_2} \exp(-\beta_2 \lambda_2)$$

[1] The probability distribution function of the $\texttt{Poi}(\lambda)$ distribution is p(y) = $\frac{\exp(-\lambda)\lambda^y}{y!}$

If M is known, the *posterior distribution* of λ_1 has the density

$$f(\lambda_1|Y_1,\cdots,Y_n,M) \propto \lambda_1^{\alpha_1-1+\sum i=1^M y_i} \exp(-\beta_1+M)\lambda_1),$$

so

$$\lambda_1|Y_1,\cdots,Y_n,M \sim \texttt{Gamma}\left(\alpha_1+\sum_{i=1}^{M} y_i, \beta_1+M\right) \tag{1}$$

$$\lambda_2|Y_1,\cdots,Y_n,M \sim \texttt{Gamma}\left(\alpha_2+\sum_{i=M+1}^{n} y_i, \beta_2+n-M\right) \tag{2}$$

Now assume that we do not know the change point M and that we assume a uniform prior on the set $\{1,\cdots,M-1\}$. It is easy to compute the distribution of M given the observations $Y_1,\cdots,Y_n$, and λ_1 and λ_2. It is a discrete distribution with probability density function proportional to

$$p(M|Y_1,\cdots,Y_n,\lambda_1,\lambda_2) \propto \lambda_1^{\sum_{i=1}^{M} y_i} \cdot \lambda_2^{\sum_{i=M+1}^{n} y_i} \cdot \exp((\lambda_2-\lambda_1)\cdot M) \tag{3}$$

The conditional distributions in (4.1) to (4.3) are all easy to sample from. It is however rather difficult to sample from the joint posterior of (λ_1,λ_2,M). $\triangleleft$
The example above suggests the strategy of alternately sampling from the (full) conditional distributions(1 to 3 in the example). Ths tentative strategy however raises some questions.

- Is the joint distribution uniquely specified by the conditional distributions?
- Sampling alternately from the conditional distributions yields a Markov chain: the newly proposed values only depend on the present values, not the past values. Will this approach yield a Markov chain with the correct invariant distribution? Will the Markov chain converge to the invariant distribution?

The answer to both questions will turn out to be yes - under certain conditions. The next section will however state the Gibbs sampling algorithm.