

# Elliptical Slice Sampling

*Francesca Panero and Xuewen Yu*

*October 20, 2017*

## 1. Introduction

## 2. Methodology

### 2.1 Slice sampling

Slice sampling was introduced by R. M. Neal in 2003 in the homonym paper published in the Annals of Statistics. Neal's idea was to present a new approach to sample from a probability density function, that tried to overcome some drawbacks of the two major Monte Carlo Markov Chain methods of sampling: the Gibbs sampler and the Metropolis Hastings method.

The aim of these three methods is to sample from a probability density function from which we are not able to sample directly (very often, from posterior distributions in Bayesian models). Their basic idea is to construct a Markov chain that have as stationary distribution the target one we care about, and samples from this chain will eventually come from the desired distribution. Gibbs sampler exploits the possibility of sampling from the full conditional density functions of each parameter given all the others (that are usually simpler to deal with than the original joint distribution), since the complete vector of parameters will eventually be sampled from the joint distribution of these. No tuning of parameters of the chain is required, but we must be able to sample from these conditional distributions in order to implement the method, and sometimes this could not be the case. Nevertheless, some algorithms to overcome this problem have been proposed, in particular ARS (adaptive rejection sampling) and MARS (adaptive rejection Metropolis sampling). Metropolis Hasting, instead, initializes the chain with a value sampled from a proposal distribution from which we are able to sample from and that attains some nice links with the target distribution. This sample can be accepted or rejected according to a probability that depends on the likelihood of the new candidate point and the previous one. This method presents some scale parameters for which there does not exist a certain rule of decision; moreover, the proposal distribution does not always come from a straightforward reasoning.

Slice sampling tries to overcome these flaws in the previous methods. The basic idea lies in the fact that if we want to sample  $x$  from a distribution  $p(x)$  that is proportional to a certain function  $f(x)$ , it would be sufficient just to sample uniformly from the area below  $f(x)$ . By defining an auxiliary random variable  $y$ , we exploit Gibbs sampler to sample from the two full conditional distributions. In particular,  $y|x$  is distributed as an *Uniform*(0,  $f(x)$ ) and  $x|y$  is distributed as an uniform on the so called slice  $S = \{x : y < f(x)\}$ . The joint density of  $x$  and  $y$  is, then

$$p(x, y) = \begin{cases} 1/Z & \text{for } 0 < y < f(x) \\ 0 & \text{otherwise} \end{cases}$$

where  $Z = \int f(x)dx$ . From this it can easily be seen that

$$p(x) = \int_0^{f(x)} p(x, y)dy = \frac{1}{Z}f(x)$$

as desired. Sampling on the slice can be difficult, and it is sometimes substituted with some update for  $x$  which leaves invariant the uniform distribution.

## 2.2 Elliptical Slice Sampling

Elliptical Slice Sampling is a particular case of Monte Carlo markov chain method that avoids the tuning of parameters, simpler and often faster than other methods to sample from the posterior distribution of models with multivariate Gaussian prior.

Let  $\mathbf{f}$  be the vector of latent variables and a Gaussian distribution with zero vector mean and covariance matrix  $\Sigma$ :

$$\mathbf{f} \sim \mathcal{N}(\mathbf{f}; \mathbf{0}, \Sigma) = |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{f}^T \Sigma^{-1} \mathbf{f}\right)$$

Let

$$L(\mathbf{f}) = p(\text{data}|\mathbf{f})$$

be the likelihood function. Our target distribution is the posterior of this model:

$$p^*(\mathbf{f}) \propto \mathcal{N}(\mathbf{f}; \mathbf{0}, \Sigma) L(\mathbf{f})$$

Neal, in 1999, introduced the idea of a Metropolis-Hastings algorithm for this type of problems, by proposing as new state

$$\mathbf{f}' = \sqrt{1 - \epsilon^2} \mathbf{f} + \epsilon \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

where  $\epsilon \in [-1, 1]$  is a the step-size parameter, and its varying defines half of an ellipse, having  $\boldsymbol{\nu}$  fixed. When  $\epsilon = 0$  we are resampling the same value. The probability of accepting the move is

$$p(\text{accept}) = \min(1, L(\mathbf{f}')/L(\mathbf{f}))$$

otherwise the next state is still  $\mathbf{f}$ . A drawback of this algorithm is that  $\epsilon$  needs to be tuned so that the Markov chain can mix efficiently. A representation of the whole ellipse gives a richer choice of updates, so we'll exploit:

$$\mathbf{f}' = \boldsymbol{\nu} \sin \theta + \mathbf{f} \cos \theta$$

that represents the ellipse centered in the origin of the axes and passing through  $\mathbf{f}$  and  $\boldsymbol{\nu}$ , and  $\theta$  is a new step-size parameter.

To avoid the tuning of the parameter, the idea is to augment the prior in order that  $\theta$  becomes a random variable and we can sample it exploiting the slice sampling. The augmented model becomes

$$\begin{aligned} \boldsymbol{\nu}_0 &\sim \mathcal{N}(\mathbf{0}, \Sigma) \\ \boldsymbol{\nu}_1 &\sim \mathcal{N}(\mathbf{0}, \Sigma) \\ \theta &\sim \text{Uniform}[0, 2\pi] \\ \mathbf{f}' &= \boldsymbol{\nu}_0 \sin \theta + \boldsymbol{\nu}_1 \cos \theta \end{aligned}$$

which assures that the distribution of  $\mathbf{f}$  is always  $\mathcal{N}(\mathbf{0}, \Sigma)$ . The new target distribution is

$$p^*(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \mathbf{f}) \propto \mathcal{N}(\boldsymbol{\nu}_0; \mathbf{0}, \Sigma) \mathcal{N}(\boldsymbol{\nu}_1; \mathbf{0}, \Sigma) L(\mathbf{f}(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \theta))$$

The slice sampling exploits the following algorithm: 1. Sample from  $p(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \theta | \boldsymbol{\nu}_0 \sin \theta + \boldsymbol{\nu}_1 \cos \theta = \mathbf{f})$ :

$$\begin{aligned} \theta &\sim \text{Uniform}[0, 2\pi] \\ \boldsymbol{\nu} &\sim \mathcal{N}(\mathbf{0}, \Sigma) \\ \boldsymbol{\nu}_0 &\leftarrow \mathbf{f} \sin \theta + \boldsymbol{\nu} \cos \theta \\ \boldsymbol{\nu}_1 &\leftarrow \mathbf{f} \cos \theta - \boldsymbol{\nu} \sin \theta \end{aligned}$$

2. Sample  $\theta$  using slice sampling on  $p^*(\theta | \boldsymbol{\nu}_0, \boldsymbol{\nu}_1) \propto L(\boldsymbol{\nu}_0 \sin \theta + \boldsymbol{\nu}_1 \cos \theta)$
3. Return  $\mathbf{f}' = \boldsymbol{\nu}_0 \sin \theta + \boldsymbol{\nu}_1 \cos \theta$

### 3. Experiment

In this section, we will validate the elliptical slice sampling algorithm on 2 models: Gaussian regression and Log Gaussian Cox process.

#### 3.1 Model Description

##### 3.1.1 Gaussian Regression

Observations  $y_n$  are drawn from Normal distribution with mean  $f_n$  and variance  $\sigma_n^2$ , for  $n = 1, \dots, N$ . Let  $N$  denote the sample size,  $D$  denote the number of dimensions.  $\mathbf{f} = (f_1, \dots, f_N) \sim N(0, \Sigma)$ . To simulate  $\mathbf{f}$ , we define the covariance matrix as

$$\Sigma_{i,j} = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D (x_{d,i} - x_{d,j})^2 / l^2\right) \quad (1)$$

Covariance matrix  $\Sigma$  is computed by inputting  $\mathbf{X}$ , which is a  $D \times N$  matrix.  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$ , where the column vector  $\mathbf{x}_n$ ,  $n = 1, \dots, N$ , is the ‘feature’ vector with  $D$  dimensions. We will draw  $\mathbf{x}_n$  from a  $D$ -dimensional unit hypercube for all  $n$ . We can simulate observations  $y_n$  after generating  $f_n$  and fixing  $\sigma_n^2$  and  $\sigma_f^2$ .  $y_1, \dots, y_N$  are *i.i.d* normal variables, so the likelihood function as a function of  $\mathbf{f}$  is:

$$L_r(\mathbf{f}) = \prod_{n=1}^N N(y_n; f_n, \sigma_n^2) \quad (2)$$

##### 3.1.2 Log Gaussian Cox process

Cox process is a “doubly stochastic” Poisson process with a stochastic intensity measure [?]. Log Gaussian Cox process is introduced by Moller [?] as the Cox process where the logarithm of the intensity function is a Gaussian process. Mathematically, let  $y_n$  denote the observations. Then  $y_n \sim \text{Poisson}(\lambda_n)$  with mean  $\lambda_n$ . The intensity function can be estimated given the log Gaussian Cox process observation within a bounded subset. This means we can partition the space finitely into  $N$  bins and  $y_n$  is the number of events in bin  $n$  for all  $n = 1, \dots, N$ . We assume that every bin has a constant intensity function  $\lambda_n$ . Let  $m$  be the offset to the log mean  $\lambda_n$ , and define it as the sum of the mean log-intensity of the Poisson process and the log of the bin size [?].

$$y_n | f_n \sim \text{Poisson}(\exp(f_n + m)) \quad (3)$$

$$\mathbf{f} \sim \mathbf{N}(\mathbf{0}, \Sigma) \quad (4)$$

$\Sigma$  is defined in Equation (1). Then the likelihood of  $\mathbf{y}$  is:

$$L_p(\mathbf{f}) = \prod_{n=1}^N \frac{\lambda_n^{y_n} \exp(-\lambda_n)}{y_n!}, \lambda_n = e^{f_n + m} \quad (5)$$

## 3.2 Implementation and Results

### 3.2.1 Gaussian Regression

Let  $l = 1$ ,  $\sigma_f^2 = 1$ ,  $\sigma_n^2 = 0.3^2$ . Firstly, we validate the Elliptical Slice Sampling algorithm on Gaussian regression model when  $\mathbf{f}$  is bivariate Normal variable. In this case, let  $N = 2$  and  $D = 1$  such that  $\{\mathbf{x}_n\}_{n=1}^2$  is one dimensional. Since both prior distribution and likelihood function are Gaussian, the posterior distribution of  $\mathbf{f}$  should be Gaussian. We perform Henze-Zirkler's test to assess whether the outputs follow Bivariate Normal distribution. This is based on the measure of distance, which is nonnegative, between the characteristic function of the multivariate normality and the empirical characteristic function. The distribution of the test statistic is approximately log normal. We achieved a p-value much larger than 0.05, which indicates that there is no strong evidence to reject the null hypothesis that the outputs of the algorithm follows multivariate normal distribtuion. We can visualise the distribution of outputs in Figure 1.