Recovering the posterior doubly-intractable distribution of a parameter

Simulation and Monte Carlo methods project presentation

Lucas Degeorge, Baptiste Dupré, Aymeric Tiberghien

ENSAE Paris

April 25, 2023

Table of contents

1. The Ising model
2. Exchange algorithms
3. Russian Roulette

Introduction and theoretical framework

We consider a $N \times N$ grid of spins : $\mathbf{y} = (y_1, \ldots, y_{N^2})$ where each $y_i \in \{-1, +1\}$.

+1	-1	-1	-1	-1
-1	-1	+1	-1	+1
-1	+1	-1	-1	+1
-1	+1	+1	-1	+1
+1	-1	+1	+1	-1

Table: Example of a 5×5 grid.

From now, we will only consider grids for $N = 10$.

Probabilities in the Ising model

The likelihood of the Ising model is defined by :

$$p(\mathbf{y};\alpha,\beta) = \frac{1}{\mathcal{Z}(\alpha,\beta)} \exp\left(\alpha \sum_{i=1}^{N^2} y_i + \beta \sum_{(i,j)\in V} y_i y_j\right)$$

where V is the set of nearest neighbour pairs and $\mathcal{Z}(\alpha,\beta)$ is the normalizing constant.

Two spins are considered as nearest neighbours in this situation :

+1	-1	-1
-1	+1	+1
-1	+1	-1

Table: The nearest neighbours (in green) of a spin (in red)

Remark on neighbours

We consider periodic boundary conditions. It means:

+1	-1	-1	+1
-1	+1	+1	-1
-1	+1	-1	+1
+1	-1	-1	-1

Table: The nearest neighbours for a spin in a corner

+1	-1	-1	+1
-1	+1	+1	+1
+1	-1	-1	-1
-1	+1	-1	+1

Table: The nearest neighbours for a spin on an edge

Issues with our distribution

The normalizing constant $\mathcal{Z}(\alpha, \beta)$ is defined by :

$$\mathcal{Z}(\alpha, \beta) = \sum_{\mathcal{Y}} \left(\alpha \sum_{i=1}^{N^2} y_i + \beta \sum_{(i,j) \in V} y_i y_j \right)$$

where $\mathcal{Y}$ denotes the set of all possible grids : $\mathcal{Y} = \{-1, +1\}^{N^2}$.

It can't be computed as it requires summing over the 2^{N^2} possible grids.

Conditional probabilities

Starting from the joint probability :

$$p(\mathbf{y};\alpha,\beta) \propto \exp\left(\alpha\sum_{i=1}^{N^2} y_i + \beta\sum_{(i,j)\in V} y_i y_j\right)$$

the conditional distribution of component $k \in \{1,\ldots,N^2\}$ given $y_{-k} = (y_1,\ldots,y_{k-1},y_{k+1},\ldots,y_{N^2})$ is

$$p(y_k|y_{-k},\alpha,\beta) \propto p(\mathbf{y};\alpha,\beta)$$

We only consider the terms involving y_k in $p(\mathbf{y};\alpha,\beta)$:

$$p(y_k|y_{-k},\alpha,\beta) \propto \exp\left(\alpha y_k + \beta\sum_{i\sim k} y_i y_k\right)$$

Conditional probabilities

For $x \in \{-1, +1\}$,

$$p(Y_k = x | y_{-k}, \alpha, \beta) \propto \exp\left(\alpha x + \beta x \sum_{i \sim k} y_i\right)$$

Finally,

$$\boxed{p(Y_k = x | y_{-k}, \alpha, \beta) = \frac{\exp\left(2\alpha x + 2\beta x \sum_{i \sim k} y_i\right)}{1 + \exp\left(2\alpha x + 2\beta x \sum_{i \sim k} y_i\right)}}$$

We recognize the logistic function :

$$p(Y_k = x | y_{-k}, \alpha, \beta) = logistic(2\alpha x + 2\beta x \sum_{i \sim k} y_i)$$

Gibbs sampling

Algorithm 1 Gibbs sampling for the Ising model

Require: $N_{gs} = 0$
Require: $(y_{1,0}, \ldots, y_{N^2,0}) \in \{-1, +1\}^{N^2}$
1: **for** $i = 0$ to $N_{gs} - 1$ **do**
2: Generate $y_{1,i+1} \sim p(\cdot|(y_{2,i}, \ldots, y_{N^2,i}), \alpha, \beta)$
3: Generate $y_{2,i+1} \sim p(\cdot|(y_{1,i+1}, y_{2,i} \ldots, y_{N^2,i}), \alpha, \beta)$
$\vdots$
4: Generate $y_{k,i+1} \sim p(\cdot|(y_{1,i+1}, \ldots, y_{k-1,i+1}, y_{k+1,i}, \ldots y_{N^2,i}), \alpha, \beta)$
$\vdots$
5: Generate $y_{N^2,i+1} \sim p(\cdot|(y_{1,i+1}, \ldots, y_{N^2-1,i+1}), \alpha, \beta)$
6: **end for**

Objective and issues

Given a grid of spins $\mathbf{y} = (y_1, \ldots, y_{N^2})$, we want to recover the parameters α and β which led to the generation of **y**

From now, we state that $\alpha = 0$ and our objective is to estimate the theoretical value of β : β_{th}

The posterior probability we want to infer is

$$p(\beta|\mathbf{y}) = \frac{p(\mathbf{y}|\beta) \times p(\beta)}{p(\mathbf{y})} = \frac{f(\mathbf{y}|\beta)}{\mathcal{Z}(\beta)} \times p(\beta) \times \frac{1}{p(\mathbf{y})}$$

$\mathcal{Z}(\beta)$ and $p(\mathbf{y})$ are intractable. $p(\beta|\mathbf{y})$ is a **doubly-intractable** distribution

Let's try Metropolis-Hasting's algorithm

Algorithm 2 Metropolis-Hastings algorithm

Require: T, initial β, a proposal $q(\cdot|\beta, \mathbf{y})$

```
for t = 1 to T do
    Propose β' ~ q(·|β, y)
    Compute a = [p(β'|y)·q(β|β',y)] / [p(β|y)·q(β'|β,y)]
    Draw r ~ U_[0,1]
    if r < a then
        β ← β'
    end if
end for
```

$$a = \frac{p(\beta'|\mathbf{y}) \cdot q(\beta|\beta', \mathbf{y})}{p(\beta|\mathbf{y}) \cdot q(\beta'|\beta, \mathbf{y})}, \quad r \sim \mathcal{U}_{[0,1]}$$

Let's try Metropolis-Hastings algorithm

One can notice that, as

$$p(\beta|\mathbf{y}) = \frac{f(\mathbf{y}|\beta)}{\mathcal{Z}(\beta)} \times p(\beta) \times \frac{1}{p(\mathbf{y})}$$

a satisfies :

$$a = \frac{f(\mathbf{y}|\beta') \cdot p(\beta') \cdot q(\beta|\beta', \mathbf{y})}{f(\mathbf{y}|\beta) \cdot p(\beta) \cdot q(\beta'|\beta, \mathbf{y}} \times \frac{\mathcal{Z}(\beta)}{\mathcal{Z}(\beta')}$$

Let's try Metropolis-Hastings algorithm

One can notice that, as

$$p(\beta|\mathbf{y}) = \frac{f(\mathbf{y}|\beta)}{\mathcal{Z}(\beta)} \times p(\beta) \times \frac{1}{p(\mathbf{y})}$$

a satisfies :

$$a = \frac{f(\mathbf{y}|\beta') \cdot p(\beta') \cdot q(\beta|\beta', \mathbf{y})}{f(\mathbf{y}|\beta) \cdot p(\beta) \cdot q(\beta'|\beta, \mathbf{y}} \times \frac{\mathcal{Z}(\beta)}{\mathcal{Z}(\beta')}$$

The red terms are not tractable.

We need to find an alternative to Metropolis-Hastings algorithm

Single-variable Exchange algorithm

Algorithm 3 Single-variable Exchange algorithm

Require: T, initial β, a proposal $q(\cdot|\beta, \mathbf{y})$

1: **for** $t = 1$ to T **do**
2: Propose $\beta' \sim q(\cdot|\beta, \mathbf{y})$
3: Generate an auxiliary variable $\mathbf{w} \sim \frac{f(\cdot|\beta')}{\mathcal{Z}(\beta')}$
4: Compute $a = \frac{q(\beta|\beta', \mathbf{y})}{q(\beta'|\beta, \mathbf{y})} \cdot \frac{p(\beta')}{p(\beta)} \cdot \frac{f(\mathbf{y}|\beta') \cdot f(\mathbf{w}|\beta)}{f(\mathbf{y}|\beta) \cdot f(\mathbf{w}|\beta')}$
5: Draw $r \sim \mathcal{U}_{[0,1]}$
6: **if** $r < a$ **then**
7: $\beta \leftarrow \beta'$
8: **end if**
9: **end for**

Single-variable Exchange algorithm

For our simulations, we used:

- for the prior over β: $\mathcal{U}_{[0,1]}$. So $\frac{p(\beta')}{p(\beta)} = 1$

Single-variable Exchange algorithm

For our simulations, we used:

- for the prior over β: $\mathcal{U}_{[0,1]}$. So $\frac{p(\beta')}{p(\beta)} = 1$
- for the proposal: $q(\cdot|\beta, \mathbf{y}) = \mathcal{N}(\beta, \sigma)$ where σ is to be tuned

Single-variable Exchange algorithm

For our simulations, we used:

- for the prior over β: $\mathcal{U}_{[0,1]}$. So $\frac{p(\beta')}{p(\beta)} = 1$
- for the proposal: $q(\cdot|\beta, \mathbf{y}) = \mathcal{N}(\beta, \sigma)$ where σ is to be tuned
- for the likelihood $f(\cdot|\beta)$: $\mathbf{y} \mapsto \exp\left(\beta \sum_{(i,j)\in V} y_i y_j\right)$ from the Ising model.

Single-variable Exchange algorithm

For our simulations, we used:

- for the prior over β: $\mathcal{U}_{[0,1]}$. So $\frac{p(\beta')}{p(\beta)} = 1$
- for the proposal: $q(\cdot|\beta, \mathbf{y}) = \mathcal{N}(\beta, \sigma)$ where σ is to be tuned
- for the likelihood $f(\cdot|\beta)$: $\mathbf{y} \mapsto \exp\left(\beta \sum_{(i,j)\in V} y_i y_j\right)$ from the Ising model.
- the Gibbs sampler to generate $\beta' \sim q(\cdot|\beta, \mathbf{y})$

Results

We run the algorithm with $T = 10^4$, $N_{gs} = 10^3$ and $\sigma = 0.1$

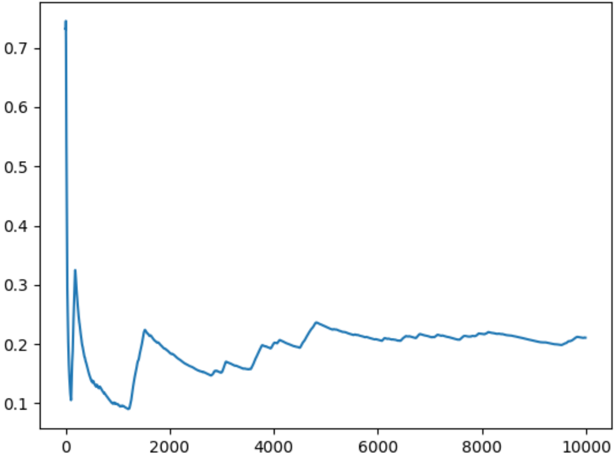

Figure: Evolution of the empirical mean of β over the iterations

Impact of the variance on the limits

We also run the algorithm with other values of the variances.

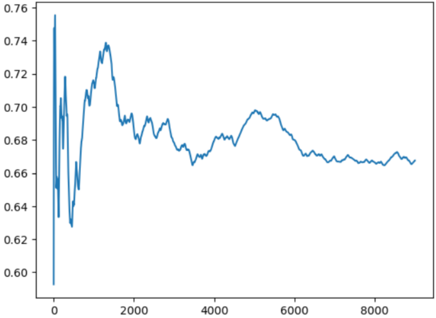

Figure: Evolution when $\sigma = 0.01$

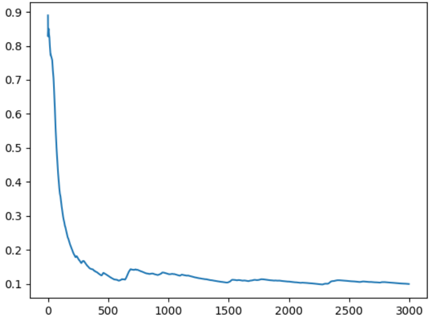

Figure: Evolution when $\sigma = 0.5$

Impact of N_{gs} on the limits

... and with different values of N_{gs}

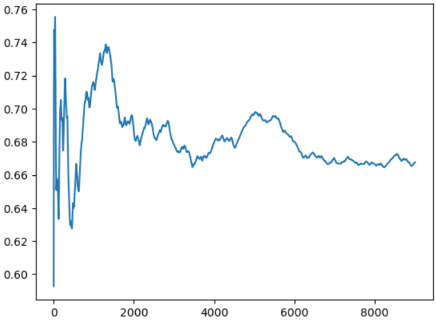

Figure: Evolution when $N_{gs} = 100$

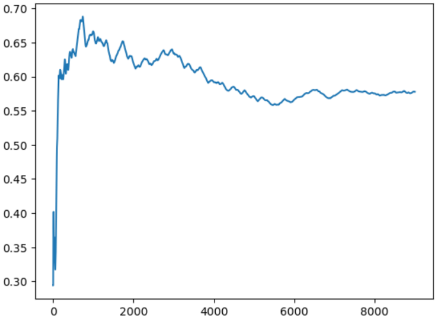

Figure: Evolution when $N_{gs} = 5300$

A way to estimate $\mathcal{Z}(\beta)$

Another idea is to infer $p(\beta|\mathbf{y}) = \frac{f(\mathbf{y}|\beta)}{\mathcal{Z}(\beta)}p(\beta)\frac{1}{p(\mathbf{y})}$ is to construct an estimator of $\frac{1}{\mathcal{Z}(\beta)}$.

Given an instrumental density $q(\mathbf{y})$, it holds

$$\frac{1}{\mathcal{Z}(\beta)} = \frac{1}{\mathcal{Z}(\beta)}\int q(\mathbf{y})d\mathbf{y} = \int \frac{q(\mathbf{y})}{f(\mathbf{y}|\beta)}p(\mathbf{y}|\beta)d\mathbf{y}$$

One can estimate this integral with:

$$\int \frac{q(\mathbf{y})}{f(\mathbf{y}|\beta)}p(\mathbf{y}|\beta)d\mathbf{y} \approx \frac{1}{T}\sum_{t=1}^{T}\frac{q(\mathbf{y}_t)}{f(\mathbf{y}_t|\beta)}$$

Warning: $\mathbf{y}_t \neq y_k$ The first one is a grid of spins, the second one belongs to $\{-1;+1\}$

Generalities on Russian Roulette methods

We consider an infinite sum S defined by : $S = \sum_{k=0}^{+\infty} u_k$

We also consider τ a finite random time, and we define for $n \geq 0$, $p_n := \mathbb{P}(\tau \geq n) > 0$.

We define $S_0 = u_0$, and for $k \geq 1$, $S_k = u_0 + \sum_{j=1}^{k} \frac{u_j}{p_j}$

Property: unbiased estimator of S

The Russian Roulette random truncation approximation of S, defined by $\hat{S} = S_\tau$ is an unbiased estimator of S.