ISUP

Stochastic Modeling
MCMC for estimating the cardinality of a set

Th objective is to study to Monte Carlo methods for estimating the cardinality $N(d)$ of the set $A_d = \{(x_1, \dots, x_d) \in \mathbb{Z}^d \cap [0,5]^d,\ x_1 + \dots + x_d = 5\}$. The first method is a crude Monte Carlo method (in this case equivalent to a reject method), the second one is a more advanced MCMC method. The aim is to show the non efficiency of the first one when d is too big and the relative efficiency of the second one.

Remark : It is a difficult problem and both methods fail when the dimension d is too big.

Preliminaries

Calculate $N(1)$, $N(2)$ and $N(3)$? It starts to be tedious, is not it ? Easier one is smart and good in combinatorics and one finds the exact calculation $N(d) = \binom{d+4}{d-1}$, or one uses MC methods.

Crude Monte Carlo method

1. Give the a.s. limit of the sequence $\frac{1}{n}\sum_{i=1}^{n} 1_{Y_i \in A_d}$, $n \to \infty$, for iid Y_is uniformly distributed over $\mathbb{Z}^d \cap [0,5]^d$?
2. Write a function `MCsimple` with arguments the dimension d and the sample size n providing an estimation $\hat{N}_1(d)$ of $N(d)$ based on the crude Monte Carlo method over n samples. Test the function to estimate $N(2)$ and $N(3)$.
3. Illustrate the randomness of the algorithm thanks to boxplots for $d = 3, 5$ et 10 pour $n = 1000$. What is happening when the dimension is too big ? Print the function $d \mapsto 6^{-d}\binom{d+4}{d-1}$. Is it reasonable to use the crude MC method for estimating $N(10)$?

MCMC method

3. What is the code below producing ?

```
walkd <- function(d){
        walk <- rep(0,d)
        u <- runif(1)
        coord1 <- ceiling(d*u)
        v <- runif(1)
        coord2 <- ceiling(d*v)
        if (coord1 != coord2) {
                walk[coord1] <- 1
                walk[coord2] <- -1
                }
        return(walk)
        }
```

Use it to create a random walk on $\mathbb{Z}^d$ on the set $\{(x_1, \dots, x_d) \in \mathbb{Z}^d,\ x_1 + \dots + x_d = 5\}$ starting from $(5, 0, \dots, 0)$ and represent a trajectory of length 1000 for $d = 2$.

4. Modify the previous code in order that the chain stays on the boundary of $[0,5]^d$ instead of exiting $[0,5]^d$. The chain restricted in $[0,5]^d$ is denoted X.
5. Check that the Markov chain is reversible. Show the existence and unicity of the stationary measure of X. Determine the stationary measure which is also the invariant measure. What are the ergodic properties of X ?
6. We want to estimate $N(4)$ thanks to X. Give the cardinality of $B_4 \subset A_4$ defined as

$$B_4 := \{(x_1, x_2, x_3, 0) \in \mathbb{N}^4,\ x_1 + \cdots + x_3 = 5\}.$$

   Give the a.s. limit ot $\frac{1}{n}\sum_{i=1} \mathbf{1}_{X_i \in B_4}$, $n \to \infty$? Estimate $N(4)$ (choose $n = 1000$).
7. Define recursively

$$B_d := \{(x_1, \ldots, x_{d-1}, 0) \in \mathbb{N}^d\, x_1 + \cdots + x_{d-1} = 5\}, \qquad d \geq 1.$$

   What is the cardinality of B_d ? Deduce a method to estimate recursively $N(k)$ for $k = 4, \ldots, d$. Let $\hat{N}_2(d)$ the corresponding estimation of $N(d)$.
8. Compare the estimators $\hat{N}_1(d)$ and $\hat{N}_2(d)$ for increasing values of d thanks to boxplots on which you represent the true value of $N(d)$.