TD3 : RANDOM SAMPLING

Exercice 1

Let \( \{X_n\}_{n \in \mathbb{N}} \) be a Markov chain on a finite state space \( E \) (for which the stationary distribution is unknown, and \( \pi \) a probability distribution. The goal of the exercise is to modify the transition probabilities of \( \{X_n\} \) so that the stationary distribution is \( \pi \) and the chain reversible.

To do that, we modify the transition from one state to another the following way : if at time \( n \), the chain is in state \( x \), if \( y \) is the next state drawn according to the transition probability of the chain, at time \( n + 1 \), the state is \( y \) with probability \( a(x, y) \) and \( x \) with probability \( 1 - a(x, y) \).

1. Give the transition probabilities of the Markov chain described above. Remember that the chain must be reversible.

Correction : Let \( q(x, y) \) be the transition probability from \( x \) to \( y \) for the new Markov chain. Then \( q(x, y) = p(x, y)a(x, y) \) for \( x \neq y \) and \( q(x, x) = p(x, x) + \sum_{y \neq x} (1 - a(s, y))p(x, y) \).

2. How to choose \( (a(x, y))_{x, y} \) so that the stationary distribution of this chain is \( \pi \)? Can we maximize the values of \( a(x, y) \)? Remember that the chain must be reversible.

Correction : As we want the new chain to be reversible, the equality

\[
\pi(x)p(x, y)a(x, y) = \pi(y)p(y, x)a(y, x)
\]

must hold for all \( x \neq y \) (and there is nothing to check for \( y = x \), this gives the possible choices for the values \( a(x, y) \).

Next, as \( a(x, y) \in [0, 1] \) for all \( x, y \), we have the relation

\[
a(x, y) = \frac{\pi(y)p(y, x)a(y, x)}{\pi(x)p(x, y)}.
\]

To maximize \( a(x, y) \), one must choose \( a(x, y) = \min(1, \frac{\pi(y)p(y, x)}{\pi(x)p(x, y)}) \) : Either \( a(x, y) \) or \( a(y, x) \) will be equal to 1, and the other less than one.

3. Let \( G \) be an undirected connected graph, that is not known entirely, and consider a random walk on this graph. When visiting one state, the information available on this graph is only the set of neighbors and their degree. We want to sample a state uniformly at random.

How to do ?

Correction : The random walk on an undirected graph is done by choosing an edge uniformly at random among the neighbors : \( p(x, y) = \frac{1}{d(x)} \), where \( d(x) \) is the degree of vertex \( x \). So one can choose \( a(x, y) = \min(1, \frac{d(x)}{d(y)}) \).

Note, it is not necessary to know the stationary distribution of the random walk in this exercise, but it can be easily computed : \( \pi(x)p(x, y) = \frac{\pi(x)}{d(x)} \), so if we choose \( \pi(x) = Ad(x) \), we have \( \pi(x)p(x, y) = A = \pi(y)p(y, x) \). The detailed balance equations are satisfied, to the Markov chain is reversible, and the stationary distribution is proportional to the degree. By normalizing, we find \( \pi(x) = \frac{d(x)}{\sum d(x)} \).

Exercice 2

Perfect sampling with a coin

Let \( \mathcal{X} \) be a finite set and \( P \) a probability distribution on \( \mathcal{X} \). The goal of this exercise is to sample an random variable \( X \) according to \( P \) exactly with a non-biased. We also want to minimize average the number of time we throw the coin, that is \( \mathbb{E}[T] \), where \( T \) is the number of throws.
1. Show that all strategies can be represented by a binary tree (potentially infinite), where leaves are labeled by elements of \( \mathcal{X} \). What is the condition so that the number of throws is almost surely finite and the distribution sampled \( P \)?

**Correction:** Let \( F \) be the set of leaves on the tree, and the ensemble des feuilles de l’arbre, \( \lambda(f) \in \mathcal{X} \) be the label of leaf \( f \) and \( h(f) \) its depth. To have an almost surely finite number of throw, one must have for all \( x \in \mathcal{X} \), \( \sum_f |\lambda(f)|_x 2^{-h(f)} = p(x) \)

Let us denote \( H(X) = \sum_{x \in \mathcal{X}} p(x) \log_2(1/p(x)) \) the entropy of \( X \).

2. Show that \( \mathbb{E}[T] \geq H(X) \).

**Correction:** On remarque tout d’abord que pour tout \( x \in \mathcal{X} \), pour tout \( f \in F(x) \), \( 2^{-h(f)} \leq p(x) \), et donc \( h(f) \geq \log \frac{1}{p(x)} \). Ainsi, on peut écrire

\[
\mathbb{E}[T] = \sum_{f \in F} 2^{-h(f)} h(f) = \sum_{x \in \mathcal{X}} \sum_{f \in F(x)} 2^{-h(f)} h(f) \\
\geq \sum_{x \in \mathcal{X}} \sum_{f \in F(x)} 2^{-h(f)} \log \frac{1}{p(x)} = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} = H(X).
\]

3. Give a necessary and sufficient condition on \( P \) so that this bound can be reached.

**Correction:** Il y a égalité si et seulement s’il y a égalité dans pour \( 2^{-h(f)} \leq p(x) \). Il faut donc que pour tout \( x \), \( p(x) \) soit de la forme \( 2^{-n_k} \).

Let \( q \in [0,1] \), and its dyadic decomposition \( q = \sum_{j \in \mathbb{N}} q_j \) avec \( q_j \in \{0,2^{-j}\} \). We note \( T_q = \sum_{j \in \mathbb{N}} j q_j \), and first admit that \( T_q < -q \log q + 2q \).

4. In the general case, propose a strategy the guaranties \( H(X) \leq \mathbb{E}[T] \leq H(X) + 2 \).

**Correction:** On considère la décomposition dyadique de \( p(x) \):

\[
p(x) = \sum_{j \geq 1} p_j(x)
\]

with \( p_j(x) = 2^{-j} \) ou 0 (les atomes). La stratégie proposée est d’allouer un atome de probabilité \( 2^{-j} \) à une feuille se situant à la profondeur \( j \). Comme la somme totale des atomes vaut un \( \sum_{x,j} p_j(x) = 1 \), on peut toujours construire un tel arbre.

Montrons alors que dans ce cas \( E(T) \leq H(X) + 2 \). Tout d’abord, par construction :

\[
E(T) = \sum_x \sum_{j \geq 1, j p_j(x) > 0} j 2^{-j} := \sum_x T_x
\]

Si on montre que \( T_x < -p(x) \log p(x) + 2p(x) \) alors on aura montré l’inégalité souhaitée. On remarque que \( p(x) \) peut être encadré par \( 2^{-(n-1)} > p(x) \geq 2^{-n} \) pour un certain entier \( n \), soit \( n-1 < -\log p(x) \leq n \). On étudie alors :

\[
T_x + p(x) \log p(x) - 2p(x) < T_x - p(x)(n-1) - 2p(x) \]
\[
= T_x - (n+1)p(x) \]
\[
= \sum_{j \geq n, j p_j(x) > 0} (j - n - 1) 2^{-j} \]
\[
= -2^{-n} + \sum_{k \geq 1, k 2^{-k(n+1)} > 0} k 2^{-(k+n+1)} \]
\[
\leq -2^{-n} + \sum_{k \geq 1} k 2^{-(k+n+1)} \]
\[
= -2^{-n} + 2^{-n+1}/2 = 0
\]
5. Show that $T_q < -q \log q + 2q$.

Exercise 3 Independent sets of fixed size

Let $G = (S, A)$ be an undirected graph. The goal of this exercise is to sample an independent set of size $k$ uniformly at random. This is still a problem that to not have a satisfactory solution yet.

Consider the Markov chain describes by the Gibbs sampler:

1. choose $v \in X_n$ and $w \in S$ uniformly;
2. if $X_n \cup \{w\} \setminus \{v\}$ is an independent set of size $k$, then $X_{n+1} = X_n \cup \{w\} \setminus \{v\}$. Else $X_{n+1} = X_n$.

1. Is this chain aperiodic?

Let $\Delta$ be the maximum degree of a vertex. We now assume that $k \leq |S|/(3\Delta + 3)$.

2. Show that the chain is then irreducible.

3. Show that the stationary distribution is the uniform distribution on the independent sets of size $k$.

Let $(X^1_n)$ and $(X^2_n)$ be two Markov chains on the independent sets of size $k$, whose transition probabilities is according to the Gibbs sampler. We define the coupling as follows:

- Choose the same $w$ for the two chains;
- choose $v^1 \in X^1_n$ uniformly in $X^1_n$. If $v^1 \in X^2_n$, choose $v^2 = v^1$, otherwise, choose $v^2$ uniformly in $X^2_n \setminus X^1_n$.

4. Show that the choice of $v^2$ is uniform.

We are now interested in the coupling time. For this, we focus on $d_n = |X^1_n - X^2_n|$. There is coupling when $d_n = 0$.

5. Describe the possible possibilities for $d_{n+1}$ in function of $d_n$ (the possible values and their probabilities).

6. Show that

$$E[d_{n+1} \mid d_n = \ell] \leq \ell \left(1 - \frac{|S| - (\Delta + 1)(3k - 3)}{|S|k}\right)$$

7. Show that

$$E[d_{n+1}] \leq k \left(1 - \frac{|S| - (\Delta + 1)(3k - 3)}{|S|k}\right)^n.$$

8. Deduce an upper bound for the mixing time of the chain.