TP 1: Reminder on Markov Chains and Stochastic Gradient Descent

Samar Rabeh

MVA - computational statistics, ENS-Paris-Saclay 2024 - 2025

Exercise 1: Box-Muller and Marsaglia-Bray Algorithm

1. Prove that $X = R\cos(\Theta)$ and $Y = R\sin(\Theta)$ have N(0,1) distribution and are independent

Let R be a random variable with a Rayleigh distribution with parameter 1, and Θ a random variable uniformly distributed on $[0, 2\pi]$, with R and Θ being independent.

We define:

$$X = R\cos(\Theta)$$
 and $Y = R\sin(\Theta)$.

We aim to show that $X \sim N(0,1), Y \sim N(0,1),$ and that X and Y are independent.

Step 1: Distribution of R

The probability density function (PDF) of a Rayleigh distribution with parameter 1 is:

$$f_R(r) = re^{-\frac{r^2}{2}}, \quad r \ge 0.$$

Step 2: Joint Distribution of X and Y

Since R and Θ are independent, the joint distribution of (R,Θ) is given by:

$$f_{R,\Theta}(r,\theta) = f_R(r)f_{\Theta}(\theta),$$

where $f_{\Theta}(\theta) = \frac{1}{2\pi}$ for $\theta \in [0, 2\pi]$.

Now, we use the transformation from polar to Cartesian coordinates:

$$X = R\cos(\Theta), \quad Y = R\sin(\Theta).$$

The Jacobian determinant for this transformation is r, so the joint PDF of (X,Y) becomes:

$$f_{X,Y}(x,y) = f_{R,\Theta}(r,\theta) \cdot \frac{1}{r} = \left(re^{-\frac{r^2}{2}}\right) \cdot \frac{1}{2\pi} \cdot \frac{1}{r} = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}.$$

This is the joint PDF of two independent normal random variables X and Y, each with mean 0 and variance 1.

Step 3: Marginal Distributions of X and Y

To confirm, we can integrate the joint PDF $f_{X,Y}(x,y)$ over y (or x) to find the marginal distribution of X (or Y). The resulting marginal distribution for both X and Y is:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

indicating that $X \sim N(0,1)$ and $Y \sim N(0,1)$.

Step 4: Independence of X and Y

The joint PDF $f_{X,Y}(x,y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$ factors as:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y),$$

showing that X and Y are independent.

Conclusion

We have shown that X and Y are independent and both follow a standard normal distribution:

$$X, Y \sim N(0, 1)$$
, and $X \perp Y$.

2. Write an algorithm for sampling an independent Gaussian distribution N(0,1)

1. Generate two independent random variables U_1 and U_2 from a uniform distribution U([0,1]).

2. Compute:

$$Z_0 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2),$$

$$Z_1 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2).$$

3. Return (Z_0, Z_1) as two independent samples from N(0, 1).

Python Code Implementation

Below is a Python implementation of the Box-Muller transform:

```
import numpy as np

def box_muller():
    # Step 1: Generate two independent U(0,1) variables
    u1, u2 = np.random.rand(2)

# Step 2: Apply the Box-Muller transform
    z0 = np.sqrt(-2 * np.log(u1)) * np.cos(2 * np.pi * u2)
    z1 = np.sqrt(-2 * np.log(u1)) * np.sin(2 * np.pi * u2)
    return z0, z1 # Two independent N(0,1) samples
```

Conclusion

The Box-Muller transform provides an efficient way to generate pairs of independent standard normal random variables. This method is widely used in simulations and statistical applications where Gaussian-distributed samples are needed.

3(a). What is the distribution of (V_1, V_2) at the end of the "while" loop?

In this question, we analyze the distribution of (V_1, V_2) in the Marsaglia-Bray algorithm after exiting the "while" loop.

Marsaglia-Bray Algorithm Recap Recall the steps involved in the Marsaglia-Bray algorithm: 1. We generate two independent random variables U_1 and U_2 uniformly distributed on [0,1]. 2. We transform them to $V_1 = 2U_1 - 1$ and $V_2 = 2U_2 - 1$, which are then uniformly distributed over [-1,1]. 3. We compute $S = V_1^2 + V_2^2$. 4. The algorithm repeats until S < 1, meaning (V_1, V_2) lies inside the unit circle. When this condition is satisfied, we exit the loop.

Distribution of (V_1, V_2) Since (V_1, V_2) are uniformly distributed over $[-1, 1] \times [-1, 1]$ before entering the loop, and we only exit the loop when $S = V_1^2 + V_2^2 < 1$, the points (V_1, V_2) are uniformly distributed within the unit disk:

$$D = \{(v_1, v_2) \in \mathbb{R}^2 : v_1^2 + v_2^2 < 1\}.$$

Thus, at the end of the "while" loop, (V_1, V_2) follows a **uniform distribution over the unit disk** D.

3(b). What is the expected number of steps in the "while" loop?

To find the expected number of steps in the "while" loop of the Marsaglia-Bray algorithm, we need to determine the probability of a single iteration satisfying S < 1, where $S = V_1^2 + V_2^2$.

Probability of S < 1 1. Since V_1 and V_2 are uniformly distributed over $[-1, 1] \times [-1, 1]$, the initial sample space is the square with area 4 (side length 2).

- 2. The condition $S = V_1^2 + V_2^2 < 1$ requires (V_1, V_2) to lie within the unit disk, which has an area of π .
- 3. Therefore, the probability that a randomly chosen point (V_1, V_2) falls within the unit disk is:

$$P(S < 1) = \frac{\text{Area of unit disk}}{\text{Area of square}} = \frac{\pi}{4}.$$

Expected Number of Steps Since each iteration of the loop is independent, the "while" loop can be modeled as a geometric distribution with success probability $p = \frac{\pi}{4}$.

For a geometric distribution with success probability p, the expected number of trials to achieve the first success is $\frac{1}{p}$. Therefore, the expected number of steps in the "while" loop is:

$$\mathbb{E}[\text{steps}] = \frac{1}{P(S < 1)} = \frac{1}{\frac{\pi}{4}} = \frac{4}{\pi}.$$

3(c). Show that T_1 and V are independent, $V \sim U([0,1])$, and T_1 has the same distribution as $\cos(\Theta)$ with $\Theta \sim U([0,2\pi])$

To address this question, let's recall the definitions: - We have two variables V_1 and V_2 uniformly distributed over the unit disk, so $S = V_1^2 + V_2^2$ is uniformly distributed in the interval [0,1]. - Let $T_1 = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}$ and $V = V_1^2 + V_2^2$.

We need to show that T_1 and V are independent, $V \sim U([0,1])$, and that T_1 has the same distribution as $\cos(\Theta)$ where $\Theta \sim U([0,2\pi])$.

Step 1: Distribution of V Since (V_1, V_2) is uniformly distributed within the unit disk, the probability density of S = V (the squared radius) is uniform over the interval [0, 1]. Therefore, $V \sim U([0, 1])$.

Step 2: Distribution of T_1 We can interpret T_1 as the cosine of the angle Θ formed by the vector (V_1, V_2) with the positive x-axis:

$$T_1 = \cos(\Theta)$$
 where $\Theta = \tan^{-1}\left(\frac{V_2}{V_1}\right)$.

Since Θ is uniformly distributed over $[0, 2\pi]$, T_1 has the same distribution as $\cos(\Theta)$ for $\Theta \sim U([0, 2\pi])$.

Step 3: Independence of T_1 and V The radial component $V = S = V_1^2 + V_2^2$ is independent of the angular component Θ , as (V_1, V_2) is uniformly distributed within the unit disk. Since T_1 is a function of Θ alone and V depends only on the radius S, T_1 and V are independent.

3(d). What is the distribution of the output (X,Y)?

1. We begin by sampling two independent uniform random variables $U_1, U_2 \sim U([0,1])$. 2. We set $V_1 = 2U_1 - 1$ and $V_2 = 2U_2 - 1$, so that V_1 and V_2 are uniformly distributed over [-1,1]. 3. We calculate $S = V_1^2 + V_2^2$ and repeat this process until S < 1, which restricts (V_1, V_2) to the unit disk. 4. Once S < 1, we set:

$$X = V_1 \sqrt{\frac{-2\ln(S)}{S}}$$
 and $Y = V_2 \sqrt{\frac{-2\ln(S)}{S}}$.

Distribution of X and Y The transformation applied to (V_1,V_2) maps points uniformly distributed over the unit disk into points that follow a Gaussian distribution. This is achieved by: - Scaling V_1 and V_2 by $\sqrt{\frac{-2\ln(S)}{S}}$, which introduces the correct radial decay to produce a normal distribution.

Since X and Y are linear transformations of V_1 and V_2 , and the scaling factor $\sqrt{\frac{-2\ln(S)}{S}}$ is independent of the angle Θ , the distribution of (X,Y) is rotationally symmetric.

Exercise 2: Invariant Distribution

1. Prove the Transition Kernel Formula

We are given a Markov chain $(X_n)_{n\geq 0}$ with values in [0,1]. Given the current state X_n :

- If $X_n = \frac{1}{k}$ for some positive integer k:
 - $-X_{n+1} = \frac{1}{k+1}$ with probability $1 X_n^2$.
 - $X_{n+1} \sim U([0,1])$ (uniformly distributed on [0,1]) with probability X_n^2 .
- If $X_n \neq \frac{1}{k}$, then $X_{n+1} \sim U([0,1])$.

We need to show that the transition kernel P(x, A) is given by:

$$P(x,A) = \begin{cases} x^2 \int_{A \cap [0,1]} dt + (1-x^2) \delta_{\frac{1}{k+1}}(A), & \text{if } x = \frac{1}{k} \\ \int_{A \cap [0,1]} dt, & \text{otherwise} \end{cases}$$

where δ_{α} is the Dirac measure at α .

Solution:

To derive P(x,A), we examine the cases for $x=\frac{1}{k}$ and $x\neq\frac{1}{k}$.

Case 1: $x = \frac{1}{k}$

When $X_n = \frac{1}{k}$, the next state X_{n+1} depends on the value of $X_n^2 = \left(\frac{1}{k}\right)^2$.

• With probability $X_n^2 = \frac{1}{k^2}$, X_{n+1} is chosen from a uniform distribution on [0,1]. Therefore, the probability that X_{n+1} falls within $A \cap [0,1]$ is:

$$\frac{1}{k^2} \int_{A \cap [0,1]} dt.$$

• With probability $1 - X_n^2 = 1 - \frac{1}{k^2}$, X_{n+1} takes the specific value $\frac{1}{k+1}$. This means that $X_{n+1} = \frac{1}{k+1}$ with probability $1 - \frac{1}{k^2}$, which we express using the Dirac measure $\delta_{\frac{1}{k+1}}(A)$.

Combining these two probabilities, we get:

$$P\left(\frac{1}{k}, A\right) = \frac{1}{k^2} \int_{A \cap [0,1]} dt + \left(1 - \frac{1}{k^2}\right) \delta_{\frac{1}{k+1}}(A).$$

Case 2: $x \neq \frac{1}{k}$

When $X_n \neq \frac{1}{k}$, the rule for the chain states that X_{n+1} is uniformly distributed over [0,1]. Thus, the probability that X_{n+1} falls within $A \cap [0,1]$ is simply:

$$P(x,A) = \int_{A \cap [0,1]} dt.$$

Conclusion

Summing up both cases, we conclude that the transition kernel P(x,A) is:

$$P(x,A) = \begin{cases} x^2 \int_{A \cap [0,1]} dt + (1-x^2) \delta_{\frac{1}{k+1}}(A), & \text{if } x = \frac{1}{k} \\ \int_{A \cap [0,1]} dt, & \text{otherwise.} \end{cases}$$

This completes the proof.

2. Prove the Invariance of the Uniform Distribution

We need to prove that the uniform distribution on [0,1] is an invariant distribution for the transition kernel P of the Markov chain $(X_n)_{n\geq 0}$. In other words, we want to show that if $X_0 \sim U([0,1])$, then $X_n \sim U([0,1])$ for all $n\geq 1$.

Solution:

Let π denote the uniform distribution on [0,1], so that for any measurable set $A \subseteq [0,1]$, we have:

$$\pi(A) = \int_{A \cap [0,1]} dt.$$

To prove that π is invariant, we need to verify that $\pi P = \pi$, which means:

$$\int_{[0,1]} P(x,A) \, d\pi(x) = \pi(A),$$

for any measurable set $A \subseteq [0, 1]$.

We consider the two cases for x in the transition kernel P(x, A):

Case 1: $x = \frac{1}{k}$

For $x = \frac{1}{k}$, the transition kernel P(x, A) is given by:

$$P\left(\frac{1}{k}, A\right) = \frac{1}{k^2} \int_{A \cap [0, 1]} dt + \left(1 - \frac{1}{k^2}\right) \delta_{\frac{1}{k+1}}(A).$$

Integrating $P\left(\frac{1}{k},A\right)$ with respect to $\pi(x)$ over [0,1], we get:

$$\int_{[0,1]} P\left(\frac{1}{k},A\right) \, d\pi(x) = \int_{[0,1]} \left(\frac{1}{k^2} \int_{A \cap [0,1]} dt + \left(1 - \frac{1}{k^2}\right) \delta_{\frac{1}{k+1}}(A)\right) d\pi(x).$$

Since π is uniform, the integral simplifies to:

$$\pi(A) = \int_{A \cap [0,1]} dt.$$

Case 2: $x \neq \frac{1}{k}$

When $x \neq \frac{1}{k}$, we have:

$$P(x,A) = \int_{A \cap [0,1]} dt.$$

In this case, integrating P(x, A) with respect to $\pi(x)$ over [0, 1], we get:

$$\int_{[0,1]} P(x,A) \, d\pi(x) = \int_{[0,1]} \int_{A \cap [0,1]} dt \, d\pi(x) = \int_{A \cap [0,1]} dt = \pi(A).$$

Conclusion

Since both cases yield $\pi(A)$ as the result, we conclude that the uniform distribution π on [0,1] is indeed invariant for the Markov chain. Thus, we have shown that $\pi P = \pi$, proving that the uniform distribution on [0,1] is an invariant distribution for P.

3. Compute Pf(x) and $\lim_{n\to+\infty} P^n f(x)$

Let f be a bounded measurable function on [0,1]. We need to compute $Pf(x) = \mathbb{E}[f(X_1)|X_0 = x]$ and then find $P^nf(x)$ for all $n \geq 1$. Finally, we will compute $\lim_{n \to +\infty} P^nf(x)$ in terms of $\int f(x)\pi(x)dx$, where π is the invariant distribution.

Solution:

The operator Pf(x) represents the expectation of $f(X_1)$ given that $X_0 = x$. Using the transition kernel P(x, A), we can express Pf(x) as:

$$Pf(x) = \int_{[0,1]} f(y)P(x, dy).$$

We examine two cases for x based on the transition kernel P(x, A):

Case 1: $x = \frac{1}{k}$

When $x = \frac{1}{k}$ (for some integer k > 0), the transition probabilities are:

$$P\left(\frac{1}{k}, A\right) = \frac{1}{k^2} \int_{A \cap [0,1]} dt + \left(1 - \frac{1}{k^2}\right) \delta_{\frac{1}{k+1}}(A).$$

Therefore, $Pf\left(\frac{1}{k}\right)$ can be written as:

$$Pf\left(\frac{1}{k}\right) = \frac{1}{k^2} \int_{[0,1]} f(y) \, dy + \left(1 - \frac{1}{k^2}\right) f\left(\frac{1}{k+1}\right).$$

Case 2: $x \neq \frac{1}{k}$

If $x \neq \frac{1}{k}$, then $X_{n+1} \sim U([0,1])$ and we have:

$$Pf(x) = \int_{[0,1]} f(y) \, dy.$$

This is the expectation of f under the uniform distribution on [0,1], which we denote as $\int f(y) d\pi(y)$, where π is the uniform distribution.

Result for $P^n f(x)$

Since the uniform distribution π is invariant, applying the transition operator P repeatedly will lead the distribution of X_n to converge to π as $n \to +\infty$, regardless of the starting point x. Thus, for any $x \in [0, 1]$,

$$\lim_{n \to +\infty} P^n f(x) = \int_{[0,1]} f(y) \, d\pi(y) = \int_0^1 f(y) \, dy.$$

4(a). Compute $P^n(x, \frac{1}{m+n})$

Base Cases

• If n = 1:

$$P(x, \frac{1}{m+1}) = P\left(\frac{1}{m}, \frac{1}{m+1}\right)$$

$$= P\left(X_{n+1} = \frac{1}{m+1} \mid X_n = \frac{1}{m}\right)$$

$$= 1 - X_n^2 = 1 - \left(\frac{1}{m}\right)^2.$$

• If n = 2:

$$\begin{split} P^2(x,\frac{1}{m+2}) &= P\left(P(x,\frac{1}{m+2})\right) \\ &= \int P(y,\frac{1}{m+2})P(x,dy) \\ &= \int P(y,\frac{1}{m+2})\left[x^2 \int_{dy \cap [0,1]} dt + (1-x^2)\delta_{\frac{1}{m+1}}(dy)\right]. \end{split}$$

The last equality follows from the result of question 1, where $x = \frac{1}{m}$ and δ_{α} is the Dirac measure at α .

Since $\int_{dy\cap[0,1]} dt = 0$, the expression simplifies to:

$$P^{2}(x, \frac{1}{m+2}) = \int P(y, \frac{1}{m+2})(1-x^{2})\delta_{\frac{1}{m+1}}(dy)$$

$$= (1-x^{2}) \int P(y, \frac{1}{m+2})\delta_{\frac{1}{m+1}}(dy)$$

$$= \left(1 - \left(\frac{1}{m}\right)^{2}\right) P\left(\frac{1}{m+1}, \frac{1}{m+2}\right).$$

Using the previous result that $P\left(\frac{1}{m+1}, \frac{1}{m+2}\right) = 1 - \left(\frac{1}{m+1}\right)^2$, we get:

$$P^{2}(x, \frac{1}{m+2}) = \left(1 - \left(\frac{1}{m}\right)^{2}\right) \left(1 - \left(\frac{1}{m+1}\right)^{2}\right)$$
$$= \prod_{i=0}^{1} \left(1 - \left(\frac{1}{m+i}\right)^{2}\right).$$

Inductive Step

We hypothesize that for n, the expression is:

$$P^{n}(x, \frac{1}{m+n}) = \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{m+i}\right)^{2}\right).$$

To prove it for n+1:

$$\begin{split} P^{n+1}(x,\frac{1}{m+n+1}) &= P\left(P^n(x,\frac{1}{m+n+1})\right) \\ &= \int P^n(y,\frac{1}{m+n+1})P(x,dy) \\ &= \int P^n(y,\frac{1}{m+n+1})\left[x^2\int_{dy\cap[0,1]}dt + (1-x^2)\delta_{\frac{1}{m+1}}(dy)\right] \\ &= \left(1-\left(\frac{1}{m}\right)^2\right)\int P^n(y,\frac{1}{m+n+1})\delta_{\frac{1}{m+1}}(dy) \\ &= \left(1-\left(\frac{1}{m}\right)^2\right)P^n\left(\frac{1}{m+1},\frac{1}{m+n+1}\right). \end{split}$$

Let m' = m + 1. Then we have:

$$P^{n+1}(x, \frac{1}{m+n+1}) = \left(1 - \left(\frac{1}{m}\right)^2\right) P^n\left(\frac{1}{m'}, \frac{1}{m'+n}\right).$$

Using the inductive hypothesis:

$$P^{n+1}(x, \frac{1}{m+n+1}) = \left(1 - \left(\frac{1}{m}\right)^2\right) \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{m'+i}\right)^2\right)$$

$$= \left(1 - \left(\frac{1}{m}\right)^2\right) \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{m+i+1}\right)^2\right)$$

$$= \left(1 - \left(\frac{1}{m}\right)^2\right) \prod_{i=1}^n \left(1 - \left(\frac{1}{m+i}\right)^2\right)$$

$$= \prod_{i=0}^n \left(1 - \left(\frac{1}{m+i}\right)^2\right).$$

Thus, by induction, we have shown that:

$$P^{n}(x, \frac{1}{m+n}) = \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{m+i}\right)^{2}\right).$$

4(b). Determine if
$$\lim_{n\to+\infty} P^n(x,A) = \pi(A)$$
 when $A = \bigcup_{q\in\mathbb{N}} \left\{\frac{1}{m+1+q}\right\}$

To determine if $\lim_{n\to+\infty} P^n(x,A) = \pi(A)$, we will calculate $\pi(A)$ and analyze the behavior of $P^n(x,A)$ as $n\to+\infty$.

First

Let π be the uniform distribution on [0, 1]. Then,

$$\pi(A) = \int_{A \cap [0,1]} dt$$

$$= \int_{\bigcup_{q \in \mathbb{N}} \left(\frac{1}{m+1+q}\right) \cap [0,1]} dt$$

$$= \sum_{q \in \mathbb{N}} \int_{\frac{1}{m+1+q} \cap [0,1]} dt$$

$$= 0.$$

Thus, $\pi(A) = 0$.

Second

Next, we compute $P^n(x, A)$:

$$P^{n}(x,A) = P^{n}\left(x, \bigcup_{q \in \mathbb{N}} \left\{\frac{1}{m+1+q}\right\}\right)$$

$$= \sum_{q \in \mathbb{N}} P^{n}\left(x, \frac{1}{m+1+q}\right)$$

$$= \sum_{q \in \mathbb{N}, q \neq n-1} P^{n}\left(x, \frac{1}{m+1+q}\right) + P^{n}\left(x, \frac{1}{m+n}\right) \ge P^{n}\left(x, \frac{1}{m+n}\right).$$

We know that:

$$P^{n}\left(x, \frac{1}{m+n}\right) = \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{m+i}\right)^{2}\right).$$

Also, we observe that:

$$\begin{split} m & \geq 2 \Longrightarrow m+i \geq 2+i, \\ & \Longrightarrow 1 - \left(\frac{1}{m+i}\right)^2 \geq 1 - \left(\frac{1}{i+2}\right)^2, \\ & \Longrightarrow \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{m+i}\right)^2\right) \geq \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{i+2}\right)^2\right). \end{split}$$

Therefore,

$$P^{n}\left(x, \frac{1}{m+n}\right) \ge \prod_{i=0}^{n-1} \left(1 - \left(\frac{1}{i+2}\right)^{2}\right)$$

$$= \prod_{i=0}^{n-1} \frac{(2+i)^{2} - 1}{(2+i)^{2}}$$

$$= \prod_{i=0}^{n-1} \frac{(i+1)(i+3)}{(i+2)^{2}}$$

$$= \prod_{i=0}^{n-1} \frac{i+1}{i+2} \prod_{i=0}^{n-1} \frac{i+3}{i+2}.$$

Breaking these products down, we get:

$$= \prod_{i=1}^{n} \frac{i}{i+1} \prod_{i=2}^{n+1} \frac{i+1}{i}$$
$$= \frac{1}{2} \frac{n+2}{n+1}$$
$$= \frac{1}{2} \left(1 + \frac{1}{n+1} \right).$$

Thus,

$$\implies P^{n}(x,A) \ge \frac{1}{2} \left(1 + \frac{1}{n+1} \right)$$

$$\implies \lim_{n \to +\infty} P^{n}(x,A) \ge \lim_{n \to +\infty} \frac{1}{2} \left(1 + \frac{1}{n+1} \right) = \frac{1}{2}$$

$$\implies \lim_{n \to +\infty} P^{n}(x,A) \ge \frac{1}{2}.$$

Conclusion

Since $\lim_{n\to+\infty} P^n(x,A) \neq \pi(A)$, we conclude that:

$$\lim_{n \to +\infty} P^n(x, A) \neq \pi(A).$$