Université Nice-Sophia Antipolis
SMEFA302 - ECUE Probabilistic computational methods, 2024-2025
Sylvain Rubenthaler (https://math.unice.fr/~rubentha/)

Mid-term exam (duration : 2h)

Documents, calculators and phones forbidden (except your personnal cheat sheet, which cannot exceed one sheet of paper). Grading will take into account how well you justify your answers. Number of points per questions will be fixed during the grading process. You can write your answers in French or English.

Let $(\epsilon_k)_{k\geq 0}$ be independent and identically distributed random variables with common law $\mathbb{P}(\epsilon_k = 1) = 1 - \mathbb{P}(\epsilon_k = -1) = p \in (0,1)$. We consider the simple random walk X_n on $\mathbb{Z}$ defined by $X_n = \sum_{k=0}^n \epsilon_k$. Suppose we want to evaluate (using a Monte Carlo scheme) the probability that X_n enters a subset $A \subset \mathbb{N}^*$ (think of $A = \{n \in \mathbb{Z} : n \geq 10\}$). If we have $p < 1/2$, then the random walk X_n tends to move to the left. One natural way to increase the probability that the random walk visits the set A is to replace p by some $\overline{p} \in (p, 1)$. Let $(\overline{\epsilon}_k)_{k\geq 0}$ be a sequence defined as $(\epsilon_k)_{k\geq 0}$ by replacing p by $\overline{p} \in (0,1)$. We set $(\forall n \in \mathbb{N})$ $Y_n = \sum_{k=0}^n \overline{\epsilon}_k$. The random walk Y_n is more likely to move to the right and as a result the event $\{Y_n \in A\}$ is more likely than $\{X_n \in A\}$. The expected value of $f(X_n) = \mathbb{1}_A(X_n)$ and the approximation mean using the standard Monte Carlo method are given respectively by (for $n \in \mathbb{N}$)

$$\mathbb{E}(f(X_n)) = \mathbb{P}(X_n \in A)\,,\ \overline{f(X_n)}_N = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_A(X_n^i)\,,$$

where $(X_n^i)_{i\geq 1}$ is a collection of independent copies of X_n and $N \in \mathbb{N}^*$.

(1) Let $(u_0, u_1, \dots, u_n) \in \{-1, +1\}^{n+1}$ (for some $n \in \mathbb{N}$). Let $z = \#\{k : u_k = +1\}$.

(a) Show that $\sum_{k=0}^n u_k = 2z - (n+1)$.

$\sum_{k=0}^n u_k = z \times 1 + (n+1-z) \times (-1) = 2z - (n+1)$

(b) Show that

$$\mathbb{P}((\epsilon_0, \dots, \epsilon_n) = (u_0, \dots, u_n)) = (1-p)^{n+1}\left(\frac{p}{1-p}\right)^z .$$

$\mathbb{P}((\epsilon_0, \dots, \epsilon_n) = (u_0, \dots, u_n)) = p^z(1-p)^{n+1-z} = (1-p)^{n+1}\left(\frac{p}{1-p}\right)^z$

(c) Show that

$$\mathbb{P}((\epsilon_0, \dots, \epsilon_n) = (u_0, \dots, u_n)) = \left(\frac{p}{1-p}\right)^{\frac{1}{2}\sum_{k=0}^n u_k} (p(1-p))^{\frac{n+1}{2}} .$$

Using 1a and 1b, we get :

$$\mathbb{P}((\epsilon_0, \dots, \epsilon_n) = (u_0, \dots, u_n)) = (1-p)^{n+1}\left(\frac{p}{1-p}\right)^{\frac{1}{2}\left(\sum_{k=0}^{n+1} u_k + (n+1)\right)}$$

$$= \left(\frac{p}{1-p}\right)^{\frac{1}{2}\sum_{k=0}^{n+1} u_k} (p(1-p))^{\frac{n+1}{2}} .$$

(2) Let $(u_0, u_1, \dots, u_n) \in \{-1, +1\}^{n+1}$ (for some $n \in \mathbb{N}$). Show that

$$\frac{\mathbb{P}((\epsilon_0, \dots, \epsilon_n) = (u_0, \dots, u_n))}{\mathbb{P}((\overline{\epsilon}_0, \dots, \overline{\epsilon}_n) = (u_0, \dots, u_n))} = G_n\left(\sum_{k=0}^n u_k\right) ,$$

$$\text{where } G_n(x) = \left(\frac{p(1-p)}{\overline{p}(1-\overline{p})}\right)^{\frac{n+1}{2}} \left(\frac{p(1-\overline{p})}{\overline{p}(1-p)}\right)^{\frac{x}{2}} .$$

We set $S=\sum_{k=0}^{n+1} u_k$. Using 1c, we get

$$\begin{aligned}\frac{\mathbb{P}((\epsilon_0,\dots,\epsilon_n)=(u_0,\dots,u_n))}{\mathbb{P}((\overline{\epsilon}_0,\dots,\overline{\epsilon}_n)=(u_0,\dots,u_n))} &= \left(\frac{p}{1-p}\right)^{\frac{S}{2}}(p(1-p))^{\frac{n+1}{2}}\left(\frac{1-\overline{p}}{\overline{p}}\right)^{\frac{S}{2}}\frac{1}{(\overline{p}(1-\overline{p}))^{\frac{n+1}{2}}}\\ &= \left(\frac{p(1-p)}{\overline{p}(1-\overline{p})}\right)^{\frac{n+1}{2}}\left(\frac{p(1-\overline{p})}{\overline{p}(1-p)}\right)^{\frac{S}{2}}.\end{aligned}$$

(3) Check that $\mathbb{E}(h(X_n))=\mathbb{E}(h(Y_n)G_n(Y_n))$ (for any measurable h).

We have

$$\mathbb{E}(h(X_n)) = \sum_{(u_0,\dots,u_n)\in\{-1,1\}^{n+1}} h(u_0+\dots+u_n)\mathbb{P}((\epsilon_0,\dots,\epsilon_n)=(u_0,\dots,u_n))$$

$$= \sum_{(u_0,\dots,u_n)\in\{-1,1\}^{n+1}} h(u_0+\dots+u_n)\mathbb{P}((\overline{\epsilon}_0,\dots,\overline{\epsilon}_n)=(u_0,\dots,u_n))\frac{\mathbb{P}((\epsilon_0,\dots,\epsilon_n)=(u_0,\dots,u_n))}{\mathbb{P}((\overline{\epsilon}_0,\dots,\overline{\epsilon}_n)=(u_0,\dots,u_n))}$$

$$(\text{using } 2) = \sum_{(u_0,\dots,u_n)\in\{-1,1\}^{n+1}} h(u_0+\dots+u_n)G_n(u_0+\dots+u_n)\mathbb{P}((\overline{\epsilon}_0,\dots,\overline{\epsilon}_n)=(u_0,\dots,u_n))$$

$$= \mathbb{E}(h(Y_n)G_n(Y_n)).$$

(4) Let $n\in\mathbb{N}$. Let $h:\mathbb{R}^{n+1}\to\mathbb{R}$ be a bounded measurable function. Let $(Y_n^i)_{i\ge1}$ be a collection of independent copies of Y_n. We set

$$\overline{h(Y_n)G_n(Y_n)}_N = \frac{1}{N}\sum_{i=1}^N h(Y_n^i)G_n(Y_n^i),\ \forall N\in\mathbb{N}^*.$$

(a) Prove that

$$W_n^N(h)=\sqrt{N}\left[\overline{h(X_n)}_N-\mathbb{E}(h(X_n))\right]$$

converges in law, as $N\to+\infty$ to a Gaussian variable with mean 0 and variance

$$\sigma_n^2(h)=\mathbb{E}(h(X_n)^2)-\mathbb{E}(h(X_n))^2.$$

The function h is bounded, so the central-limit Theoreom gives the answer.

(b) Prove that

$$\overline{W}_n^N(h)=\sqrt{N}\left[\overline{h(Y_n)G_n(Y_n)}_N-\mathbb{E}(h(X_n))\right]$$

converges in law, as $N\to+\infty$ to a Gaussian variable with mean 0 and variance

$$\begin{aligned}\overline{\sigma}_n^2(h) &= \mathbb{E}(h(Y_n)^2G_n(Y_n)^2)-\mathbb{E}(h(X_n))^2\\ &= \sigma_n^2(h)+\mathbb{E}(h(X_n)^2(G_n(X_n)-1)).\end{aligned}$$

The function $h\times G$ is bounded, so the central-limit Theoreom gives us that $\overline{W}_n^N(h)$ converges in law to a centered Gaussian with variance

$$\begin{aligned}\overline{\sigma}_n^2(h) &= \mathbb{E}(h(Y_n)^2G_n(Y_n)^2)-\mathbb{E}(h(X_n))^2\\ (\text{by } 3) &= \mathbb{E}(h(X_n)^2G_n(X_n))-\mathbb{E}(h(X_n))^2\\ &= \sigma_n^2(h)+\mathbb{E}(h(X_n)^2(G_n(X_n)-1)).\end{aligned}$$

(5) Let $n\in\mathbb{N}$. Prove that for a function of the form $f=\mathbb{1}_A$ with $A\subset\{x\in\mathbb{R}:G_n(x)\le 1/a_n\}$, for some $a_n\ge1$, we have

$$\overline{\sigma}_n^2(f)\le\frac{\mathbb{P}(X_n\in A)}{a_n}-\mathbb{P}(X_n\in A)^2\le\sigma_n^2(f).$$

Algorithm 1 6b

```
import numpy as np
import scipy.stats as sps
n=50
p=0.4
pb=0.6
a=((p*p)/((1-p)*(1-p)))**(-5)
l=10
def X(n,p):
  u=sps.bernoulli.rvs(p,size=n)
  v=[2*x-1 for x in u]
  return(np.sum(v))
def G(n,x):
  return(((p**2)/((1-p)**2))**(x/2))
def MC1(N):
  tab=[X(n,p) for i in range(N)]
  tabf=[int(x>(l-1)) for x in tab]
  return(np.mean(tabf),np.std(tabf))
```

Algorithm 2 6c

```
def MC2(N):
  tab=[X(n,pb) for i in range(N)]
  tabf=[int(x>(l-1))*G(n,x) for x in tab]
  return(np.mean(tabf),np.std(tabf))
```

We have (by 4b)

$$\begin{aligned}\overline{\sigma}_n^2(f) &= \mathbb{E}(\mathbb{1}_A(Y_n)^2 G_n(Y_n)^2) - \mathbb{E}(\mathbb{1}_A(X_n))^2 \\ &\leq \frac{\mathbb{E}(\mathbb{1}_A(Y_n)^2 G_n(Y_n))}{a_n} - \mathbb{E}(\mathbb{1}_A(X_n))^2 \\ &= \frac{\mathbb{P}(X_n \in A)}{a_n} - \mathbb{P}(X_n \in A)^2 \\ &\leq \mathbb{P}(X_n \in A) - \mathbb{P}(X_n \in A)^2 = \sigma_n^2(f)\,.\end{aligned}$$

(6) In the following, we will use the following numerical values : $n = 50$, $p = 0.4$, $\overline{p} = 0.6$, $A = \{x \in \mathbb{Z} : x \geq 10\}$, $a_n = \left(\frac{p^2}{(1-p)^2}\right)^{-5}$.

(a) Show that $A \subset \{x \in \mathbb{Z} : G_n(x) \leq 1/a_n\}$.

We have $\overline{p} = 1 - p$. So, using 2, we get

$$G(x) = \left(\frac{p^2}{(1-p)^2}\right)^{x/2}$$

(which is a decreasing function of x). So, for $x \in \mathbb{Z}$, $x \geq 10$,

$$G(x) \leq \left(\frac{p^2}{1-p^2}\right)^5 = \frac{1}{a_n}\,.$$

(b) Write a python function `MC1(N)` that compute $\overline{f(X_n)}_N$ and an approximation of $\sigma_n(f)$ (for a given N). Remember that for coding questions, the less loops you use, the more points you get.

(c) Write a python function `MC2(N)` that compute $\overline{f(X_n)}_N$ and an approximation of $\overline{\sigma}_n(f)$ (for a given N)