Chapitre XV - Concentration - Loi des grands nombres

Rémi Caneri

Table des matières

I. Inégalité de Bienaymé-Tchebichev

Propriété

Soit X une variable aléatoire d'espérance μ et de variance V.

$$\forall \delta \in \mathbb{R}_+^*, \mathbb{P}(|X-\mu| \geqslant \delta) \leqslant \frac{V}{\delta^2}$$

Démonstration

Soit $\delta \in \mathbb{R}_+^*$.
Soit X une variable aléatoire d'espérance μ et de variance V.
X prend les valeurs $x_1, x_2, ..., x_n$.
Soit F l'ensemble des valeurs x_k telles que $|x_k - \mu| \geqslant \delta$.

$V = \displaystyle\sum_{i=1}^{n} \mathbb{P}(X = x_k)(x_k - E(X))^2$
$V \geqslant \displaystyle\sum_{x_k \in F} \mathbb{P}(X = x_k)(x_k - \mu)^2$
$V \geqslant \displaystyle\sum_{x_k \in F} \mathbb{P}(X = x_k)\delta^2$
$V \geqslant \delta^2 \displaystyle\sum_{x_k \in F} \mathbb{P}(X = x_k)$
Or $\displaystyle\sum_{x_k \in F} \mathbb{P}(X = x_k) = \mathbb{P}(F) = \mathbb{P}(|X-\mu| \geqslant \delta)$
Donc $V \geqslant \delta^2 \mathbb{P}(|X-\mu| \geqslant \delta)$
$\dfrac{V}{\delta^2} \geqslant \mathbb{P}(|X-\mu| \geqslant \delta)$
Ou encore $\mathbb{P}(|X-\mu| \geqslant \delta) \leqslant \dfrac{V}{\delta^2}$

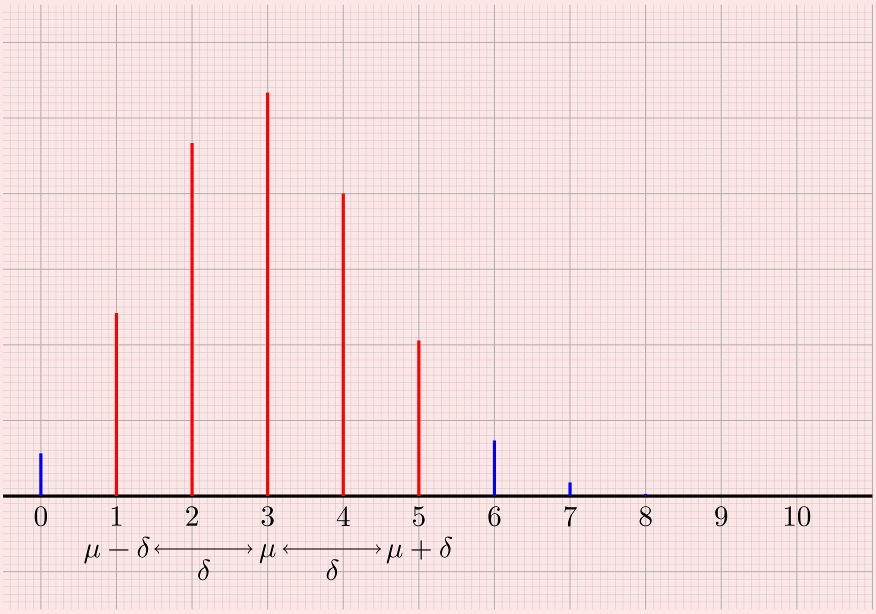
Interprétation

L'inégalité de Bienaymé-Tchebichev indique que la probabilité d'obtenir une valeur distante de μ de plus de δ est inférieure à $\frac{V}{\delta^2}$.

De plus, $\mathbb{P}(|X-\mu| \geqslant \delta) = 1 - \mathbb{P}(|X-\mu| < \delta)$
Donc $\mathbb{P}(|X-\mu| \geqslant \delta) \leqslant \frac{V}{\delta^2} \iff 1 - \mathbb{P}(|X-\mu| < \delta) \leqslant \frac{V}{\delta^2} \iff 1 - \frac{V}{\delta^2} \leqslant \mathbb{P}(|X-\mu| < \delta)$

Donc l'inégalité de Bienaymé-Tchebichev permet d'écrire que la probabilité d'obtenir une valeur distante de μ de moins de δ est supérieure ou égale à $1 - \frac{V}{\delta^2}$.

Interprétation - suite



Loi binomiale de parmètres 10 et 0,3

La probabilité d'être dans le bleu est inférieure à $\frac{V}{\delta^2}$ et celle d'être dans le rouge est supérieure à $1-\frac{V}{\delta^2}$.

Exemple

Au basket, lorsqu'elle tente un lancer franc, Claire marque dans 70% des cas.
A l'entraînement, sur une série de 100 lancers, on note X le nombre de paniers réussis par Claire.

1. Déterminer la loi de probabilité de X.
2. Déterminer l'espérance μ et l'écart type V de X.
3. Donner alors l'inégalité de Bienaymé-Tchebichev. Interpréter cette inégalité.

Réponse

1. Lorsqu'on tente un lancer franc, soit on le marque (dans 70% des cas), soit on ne le marque pas. Il s'agit donc d'une épreuve de Bernoulli de paramètre $0,7$.
   Notre expérience est donc une répétition de 100 épreuves de Bernoulli identiques et indépendantes. Donc $X \rightsquigarrow \mathcal{B}(100;0,7)$.
2. $\mu = 100\times 0,7 = 70$ et $V = 100\times 0,7\times(1-0,7) = 70\times 0,3 = 21$
3. $\forall \delta \in \mathbb{R}_+^*, \mathbb{P}(|X-70| \geqslant \delta) \leqslant \frac{21}{\delta^2}$
   La probabilité de réussir plus de $70+\delta$ ou moins de $70-\delta$ lancers francs est inférieure à $\frac{21}{\delta^2}$.
   Par exemple, si on prend, $\delta = 10$, la probabilité de réussir plus de 80 ou moins de 60 lancers francs est inférieure à $\frac{21}{10^2} = 0,21$

Remarque

Soit $k \in \mathbb{N}^*$.
Choisissons $\delta = k\sigma$. Rappelons que $\sigma = \sqrt{V}$, c'est-à-dire que $V = \sigma^2$.
L'inégalité de Bienaymé-Tchebichev devient $\mathbb{P}(|X-\mu| \geqslant k\sigma) \leqslant \frac{\sigma^2}{(k\sigma)^2}$
$\mathbb{P}(|X-\mu| \geqslant k\sigma) \leqslant \frac{\sigma^2}{k^2\sigma^2}$
$\mathbb{P}(|X-\mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}$

- $k = 1$:
  $\mathbb{P}(|X-\mu| \geqslant 1\sigma) \leqslant \frac{1}{1^2} \iff \mathbb{P}(|X-\mu| \geqslant \sigma) \leqslant 1$
  Aucun intérêt, on s'en doutait.

- $k = 2$:
  $\mathbb{P}(|X-\mu| \geqslant 2\sigma) \leqslant \frac{1}{2^2} \iff \mathbb{P}(|X-\mu| \geqslant 2\sigma) \leqslant 0,25$
  La probabilité d'obtenir une valeur à une distance supérieure à 2σ de μ est inférieure à $0,25$

- $k = 4$:
  $\mathbb{P}(|X-\mu| \geqslant 4\sigma) \leqslant \frac{1}{4^2} \iff \mathbb{P}(|X-\mu| \geqslant 4\sigma) \leqslant 0,0625$
  La probabilité d'obtenir une valeur à une distance supérieure à 4σ de μ est inférieure à $0,0625$

- $k = 5$:
  $\mathbb{P}(|X-\mu| \geqslant 5\sigma) \leqslant \frac{1}{5^2} \iff \mathbb{P}(|X-\mu| \geqslant 5\sigma) \leqslant 0,04$
  La probabilité d'obtenir une valeur à une distance supérieure à 5σ de μ est inférieure à $0,04$

II. Inégalité de concentration

Propriété

Soit $n \in \mathbb{N}$ avec $n \geqslant 2$.
On considère n variables aléatoires X_1, X_2, ..., X_n indépendantes, identiques suivant la même loi de probabilité d'espérance μ et de variance V.
On note $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ la variable aléatoire moyenne.

$$\forall \delta \in \mathbb{R}_+^*, \mathbb{P}(|M_n - \mu| \geqslant \delta) \leqslant \frac{V}{n\delta^2}$$

Démonstration

Soit $n \in \mathbb{N}$ avec $n \geqslant 2$.
On considère n variables aléatoires X_1, X_2, ..., X_n indépendantes, identiques suivant la même loi de probabilité d'espérance μ et de variance V.
On note $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ la variable aléatoire moyenne.

Soit $\delta \in \mathbb{R}_+^*$.
D'après l'inégalité de Bienaymé-Tchebichev, on a :
$\mathbb{P}(|M_n - E(M_n)| \geqslant \delta) \leqslant \frac{V(M_n)}{\delta^2}$
Or $E(M_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$
et $V(M_n) = V\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i\right) = \dfrac{1}{n^2}V\left(\displaystyle\sum_{i=1}^{n} X_i\right) = \dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} V(X_i) = \dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} V = \dfrac{nV}{n^2} = \dfrac{V}{n}$
Donc l'inégalité de Bienaymé-Tchebichev devient :
$\mathbb{P}(|M_n - \mu| \geqslant \delta) \leqslant \frac{V}{n\delta^2}$

Exemple

Dans une société de démarchage par téléphone, on estime que 40% des personnes appelées répondent effectivement.
On appelle n personnes, avec $n \in \mathbb{N}^*$.
La variable aléatoire X_k avec $k \in [\![1;n]\!]$ donne 1 si la k-ième personne appelée répond et 0 sinon.
Les variables X_k sont identiques et indépendantes.

1. Quelle loi suivent les variable X_k ?
2. Déterminer l'espérance et la variance des variables aléatoire X_k.
3. Soit $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ la variable aléatoire moyenne des X_k.
   (a) Rappeler l'inégalité de concentration dans notre cas.
   (b) On appelle précision la valeur δ et risque la valeur $\frac{V}{n\delta^2}$. On appelle 1000 personnes. Déterminer le risque lorsqu'on a une précision de $0,1$.
   (c) Combien de personnes faudrait-il appeler pour avoir un risque inférieure ou égal à 1% lorsqu'on a une précision de $0,1$?

Réponse

1. $\forall k \in [\![1;n]\!]$, la variable X_k donne 1 si la k-ième personne appelée répond et 0 sinon.
   Donc les variables X_k suivent des lois de Bernoulli de paramètre $p = \frac{40}{100} = 0,4$.
2. $E(X_k) = p = 0,4$ et $V(X_k) = p(1-p) = 0,4(1-0,4) = 0,24$
3. $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ est la variable aléatoire moyenne des X_k.
   (a) $\mathbb{P}(|M_n - 0,4| \geqslant \delta) \leqslant \frac{0,24}{n\delta^2}$
   (b) Le risque vaut $\frac{0,24}{1000\times 0,1^2} = \frac{0,24}{10} = 0,024$
   Donc $\mathbb{P}(|M_n - 0,4| \geqslant 0,1) \leqslant 0,024$
   Ainsi, sur 1000 personnes appelées, la probabilité que le nombre de personnes qui répondent soit en dehors de $]300;500[$ est inférieure à 0,024.
   (c) $\frac{0,24}{n0,1^2} \leqslant \frac{1}{100} \iff \frac{24}{n} \leqslant 0,01 \iff \frac{24}{0,01} \leqslant n \iff n \geqslant 2400$
   Il faudrait appeler au moins 2400 personnes pour que le risque soit inférieur à 1%.

III. Loi des grands nombres

Propriété

Soit $n \in \mathbb{N}$ avec $n \geqslant 2$.
On considère n variables aléatoires X_1, X_2, ..., X_n indépendantes, identiques suivant la même loi de probabilité d'espérance μ et de variance V.
On note $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ la variable aléatoire moyenne.

$$\forall t \in \mathbb{R}_+^*, \lim_{n\to+\infty} \mathbb{P}(|M_n - \mu| \geqslant t) = 0$$

Démonstration

Soit $n \in \mathbb{N}$ avec $n \geqslant 2$.
On considère n variables aléatoires X_1, X_2, ..., X_n indépendantes, identiques suivant la même loi de probabilité d'espérance μ et de variance V.
On note $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ la variable aléatoire moyenne.
L'inégalité de concentration nous donne : $\forall \delta \in \mathbb{R}_+^*, \mathbb{P}(|M_n - \mu| \geqslant \delta) \leqslant \frac{V}{n\delta^2}$
Soit $t \in \mathbb{R}_+^*$.
On a $\mathbb{P}(|M_n - \mu| \geqslant t) \leqslant \frac{V}{nt^2}$
Or $\mathbb{P}(|M_n - \mu| \geqslant t) \geqslant 0$
Donc $0 \leqslant \mathbb{P}(|M_n - \mu| \geqslant t) \leqslant \frac{V}{nt^2}$
Or $\lim\limits_{n \to +\infty} \frac{V}{nt^2} = 0$
Donc d'après le théorème des gendarmes, $\lim\limits_{n \to +\infty} \mathbb{P}(|M_n - \mu| \geqslant t) = 0$.

Ainsi $\forall t \in \mathbb{R}_+^*, \lim\limits_{n \to +\infty} \mathbb{P}(|M_n - \mu| \geqslant t) = 0$

Interprétation

La loi des grands nombres indique que peu importe la valeur de $t > 0$, la probabilité de l'événement $|M_n - \mu| \geqslant t$ se rapproche de 0 lorsque n devient très grand.

Exemple

Soient $X_1, X_2, ..., X_n$ n variables aléatoires identiques et indépendantes d'espérance μ.
On pose $M_n = \frac{X_1 + X_2 + ... + X_n}{n}$.
Expliquer pourquoi, pour n assez grand, $\mathbb{P}(|M_n - \mu| \geqslant 0,1) \leqslant 10^{-3}$.

Réponse

Soient $X_1, X_2, ..., X_n$ n variables aléatoires identiques et indépendantes d'espérance μ.
$M_n = \frac{X_1 + X_2 + ... + X_n}{n}$.
Donc d'après la loi des grands nombres, $\forall t \in \mathbb{R}_+^*, \lim\limits_{n \to +\infty} \mathbb{P}(|M_n - \mu| \geqslant t) = 0$.
En particulier, pour $t = 0,1$, on a :
$\lim\limits_{n \to +\infty} \mathbb{P}(|M_n - \mu| \geqslant 0,1) = 0$
Et donc par définition de la limite, on a :
$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \in \mathbb{N}, (n \geqslant N \implies |\mathbb{P}(|M_n - \mu| \geqslant 0,1) - 0| \leqslant \epsilon)$
$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \in \mathbb{N}, (n \geqslant N \implies \mathbb{P}(|M_n - \mu| \geqslant 0,1) \leqslant \epsilon)$
En particulier pour $\epsilon = 10^{-3}$, on a :
$\exists N \in \mathbb{N}, \forall n \in \mathbb{N}, (n \geqslant N \implies \mathbb{P}(|M_n - \mu| \geqslant 0,1) \leqslant 10^{-3})$
Donc pour n assez grand, $\mathbb{P}(|M_n - \mu| \geqslant 0,1) \leqslant 10^{-3}$

IV. Algorithmes

1. Loi binomiale et inégalité de Bienaymé-Tchebichev

Soit S une variable aléatoire qui suit une loi binomiale $\mathcal{B}(100; 0,7)$.

1. Calculer l'espérance et l'écart type de S.
2. Déterminer, à la calculatrice, la probabilité $\mathbb{P}(|S - 70| \geqslant 10)$.
3. Utiliser l'inégalité de Bienaymé-Tchebichev et comparer le résultat obtenu.
4. Cas général
   Soit S_n une variable aléatoire qui suit une loi binomiale $\mathcal{B}(n; p)$.

(a) Compléter la fonction `simulation_binomiale(n, p)` qui prend en paramètres les valeurs de n et p ed la loi binomiale et qui renvoie l'écart $|x - np|$ où x est la valeur prise par la variable aléatoire.

```python
from random import random()
def simulation_binomiale(n, p):
    x = 0
    for k in range(n):
        a = random()
        if a <= p:
            x = ........
    return abs(.............)

```

(b) Écrire une fonction `echantillons(Nombre, n, p)` de paramètres $Nombre$, le nombre d'échantillon, n et p, les paramètres de la loi binomiale et qui renvoie la proportion des valeurs x prises par X telles que $|x - np| \geqslant \sqrt{n}$.

(c) Tester cette fonction avec $n = 1000$ et $p = 0,4$ pour 100 échantillons.

(d) Comparer avec l'inégalité de Bienaymé-Tchebichev.

5. Que donne l'inégalité de Bienaymé-Tchebichev lorsque $\sqrt{n}$, dans le cas d'une loi binomiale de paramètres n et p.

Réponse

1. $E(S) = 100 \times 0,7 = 70$ et $V(S) = 100 \times 0,7 \times (1 - 0,7) = 21$

2. $\mathbb{P}(|S - 70| \geqslant 10) \approx 0,012$.

3. L'inégalité de Bienaymé-Tchebichev donne :
$\mathbb{P}(|S - 70| \geqslant 10) \leqslant \frac{21}{10^2}$
$\mathbb{P}(|S - 70| \geqslant 10) \leqslant 0,21$
L'inégalité de Bienaymé-Tchebichev n'est pas optimale.

4. (a)

```python
from random import random()
def simulation_binomiale(n, p):
    x = 0
    for k in range(n):
        a = random()
        if a <= p:
            x = x+1
    return abs(x-n*p)

```

(b)

```python
from math import sqrt
def echantillons(Nombre, n, p):
    y = 0
    for k in range(Nombre):
        if simulation_binomiale(n, p) >= sqrt(n):
            y = y + 1
    return y/n

```

Réponse - suite

4. (c)
```
>>> echantillons(100, 1000, 0.4)
0.005
>>> echantillon(100, 1000, 0.4)
0.002
>>> echantillons(100, 1000, 0.4)
0.004
```

(d) L'inégalité de Bienaymé-Tchebichev donne dans ce cas :
$\mathbb{P}(|S-40| \geqslant 10) \leqslant \frac{24}{\sqrt{1000}^2}$
$\mathbb{P}(|S-40| \geqslant 10) \leqslant 0,024$
La simulation montre que l'inégalité de Bienaymé-Tchebichev n'est pas optimale.

5. L'inégalité de Bienaymé-Tchebichev donne :
$\mathbb{P}(|S-np| \geqslant \sqrt{n}) \leqslant \frac{np(1-p)}{\sqrt{n}^2}$
$\mathbb{P}(|S-np| \geqslant \sqrt{n}) \leqslant \frac{np(1-p)}{n}$
$\mathbb{P}(|S-np| \geqslant \sqrt{n}) \leqslant p(1-p)$

2. Marche aléatoire

On considère un individu sur un escalier infini dans les deux sens. Il commence sur la marche numéro 0. Les marches sont numérotées par des nombres entiers relatifs : lorsqu'on monte le numéro de la marche augmente de 1, et lorsqu'on descend, il diminue de 1.
Il tire à pile ou face avec une pièce parfaitement équilibrée pour décider si le prochain pas permet de monter ou de descendre une marche.
La probabilité de monter ou de descendre est donc $p = \frac{1}{2}$.

Soit $n \in \mathbb{N}^*$.
Appelons A_n le nombre de fois ou l'individu a monté une marche après n pas.
Appelons X_n le numéro de la marche sur laquelle se trouve l'individu après n pas.

1. **Préambule** Soit Y une variable aléatoire prenant les valeurs $y_1, y_2, ..., y_k$ avec $k \in \mathbb{N}^*$.
Soit $d \in \mathbb{R}$.
Démontrer que $V(Y+d) = V(Y)$.

2. **Etude de la variable X_n**

   (a) Expliquer pourquoi la variable aléatoire A_n suit une loi binomiale $\mathcal{B}\left(n; \frac{1}{2}\right)$.

   (b) Exprimer l'espérance et la variance de A_n en fonction de n.

   (c) Expliquer pourquoi $X_n = 2A_n - n$.

   (d) En déduire que $E(X_n) = 0$ et $\sigma(X_n) = \sqrt{n}$

3. **Simulation de Q échantillons de taille N de X_n**
On considère le programme ci-dessous :

```python
from random import random
from math import sqrt

def X(n):
    x = 0
    for i in range(n):
        a = random()
        if a < 0.5:
            x = x - 1
        else:
            x = x + 1
    return x

def moyenne(n, N):
    som = 0
    for i in range(N):
        som = som + X(n)
    return som / N

def proportion(n, N, Q, delta):
    l = 0
    for i in range(Q):
        if abs(moyenne(n, N)) > = delta:
            l = l + 1
    return l / Q
```

(a) Expliquer pourquoi la fonction `X(n)` permet de simuler la variable aléatoire X_n.

(b) On réalise une simulation d'un échantillon de taille N de X_n.
Quel résultat la fonction `moyenne(n, N)` renvoie-t-elle ?

(c) Lors d'une simulation de Q échantillons de taille N de X_n, on obtient Q moyennes $m_1, m_2, ..., m_Q$.
Dans la fonction `proportion(n, N, Q, delta)`, que représente la variable ℓ ?
Interpréter le résultat renvoyé par cette fonction.

4. **Observation des résultats**

(a) Exécuter `proportion(100, 100, 1000,2)` plusieurs fois de suite.
Interpréter les résultats obtenus.

(b) Dans un échantillon de taille N, on note M_N la moyenne des N valeurs prises par X_n.
Démontrer que $\forall \delta > 0, \mathbb{P}(|M_N| \geqslant \delta) \leqslant \frac{n}{N\delta^2}$

(c) Comparer les résultats obtenus par simulation à celui donné par l'inégalité de la question précédente.

Réponse

1. $E(Y+d) = E(Y)+d$
   $V(Y+d) = \sum_{i=1}^{k} \mathbb{P}(Y=y_i)((y_i+d)-E(Y+d))^2 = \sum_{i=1}^{k} \mathbb{P}(Y=y_i)(y_i+d-(E(Y)+d))^2$
   $V(Y+d) = \sum_{i=1}^{k} \mathbb{P}(Y=y_i)(y_i+d-E(Y)-d)^2 = \sum_{i=1}^{k} \mathbb{P}(Y=y_i)(y_i-E(Y))^2 = V(Y).$

2. (a) Dans notre expérience aléatoire, soit on monte (succès), soit on descend(échec). Il s'agit donc d'une épreuve de Bernoulli.
   On répète cette épreuve n fois. Toutes les épreuves ainsi répétée sont identiques et indépendantes.
   La variable aléatoire A_n compte le nombre de succès.
   Donc A_n suit une loi binomiale $\mathcal{B}\left(n; \frac{1}{2}\right)$

   (b) $E(A_n) = np = n \times \frac{1}{2} = \frac{n}{2}$ et $V(A_n) = np(1-p) = n \times \frac{1}{2}\left(1-\frac{1}{2}\right) = n \times \frac{1}{2} \times \frac{1}{2} = \frac{n}{4}$

   (c) X_n est le numéro de la marche où se trouve l'individu après n pas.
   A_n est le nombre de fois où il a monté d'une marche.
   Donc $n - A_n$ est le nombre de fois où il a descendu une marche.
   Ainsi $X_n = A_n - (n - A_n) = A_n - n + A_n = 2A_n - n$

   (d) $E(X_n) = E(2A_n - n) = 2E(A_n) - n = 2 \times \frac{n}{2} - n = n - n = 0$
   $V(X_n) = V(2A_n - n) = V(2A_n) = 2^2 V(A_n) = 4 \times \frac{n}{4} = n$
   donc $\sigma(X_n) = \sqrt{V(X_n)} = \sqrt{n}$

3. (a) La fonction `X(n)` permet de simuler la variable aléatoire X_n car elle répète n fois l'augmentation ou la diminution de 1 de la variable x. Cette dernière est initialisée à 0 donc elle représente le numéro de la marche après n pas.

   (b) La fonction `moyenne(n, N)` renvoie le numéro moyen de la marche sur lequel se trouve l'individu après n pas.

   (c) La variable ℓ représente le nombre moyennes de l'échantillon de taille Q qui sont supérieures au égales à `delta`. La fonction renvoie donc la proportion de moyennes de l'échantillon des Q moyenne, supérieures ou égales à `delta`.

4. (a)
```
>>> proportion(100, 100, 1000,2)
0.052
>>> proportion(100, 100, 1000,2)
0.041
>>> proportion(100, 100, 1000,2)
0.045
```
   Après simulation de 1000 échantillons de taille 100 de la variable X_n, il y a 5,2% des échantillons dont le numéro moyen de la marche est supérieure ou égale à 2.

   (b) D'après l'inégalité de concentration, on a :
   $\forall \delta > 0, \mathbb{P}(|M_N - E(M_N)| \geqslant \delta) \leqslant \frac{V(X_n)}{N\delta^2}$
   Soit $\forall \delta > 0, \mathbb{P}(|M_N| \geqslant \delta) \leqslant \frac{n}{N\delta^2}$

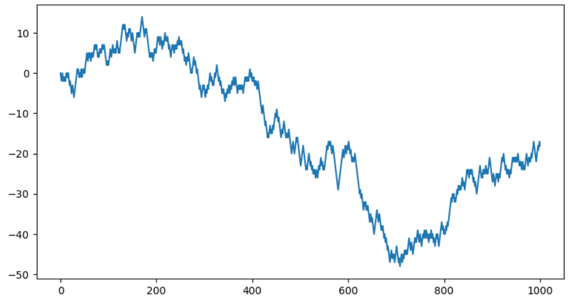
   (c) Dans la simulation, on a $n = 100$, $N = 100$, $Q = 1000$ et $\delta = 2$.
   Donc notre inégalité devient :
   $\mathbb{P}(|M_{100}| \geqslant 2) \leqslant \frac{100}{100 \times 2^2}$
   $\mathbb{P}(|M_{100}| \geqslant 2) \leqslant \frac{1}{4}$
   $\mathbb{P}(|M_{100}| \geqslant 2) \leqslant 0,25$
   L'inégalité de concentration n'est pas optimale.

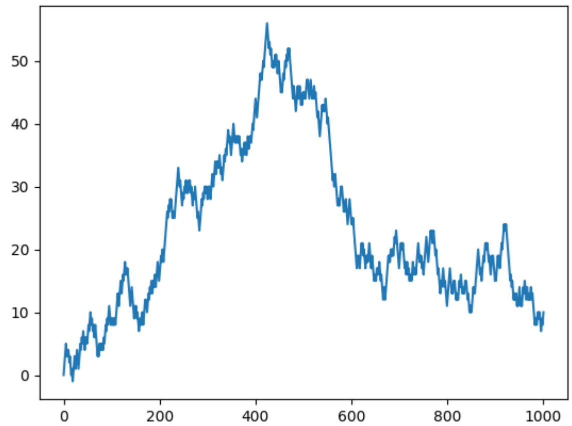
V. Approfondissements

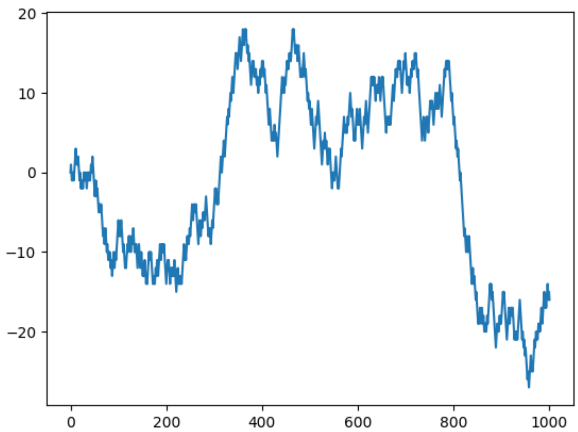
1. Marche aléatoire

On considère un individu sur un escalier infini dans les deux sens. Il commence sur la marche numéro 0. Les marches sont numérotées par des nombres entiers relatifs : lorsqu'on monte le numéro de la marche augmente de 1, et lorsqu'on descend, il diminue de 1.
Il tire à pile ou face avec une pièce parfaitement équilibrée pour décider si le prochain pas permet de monter ou de descendre une marche.
La probabilité de monter ou de descendre est donc $p = \frac{1}{2}$.

Soit $n \in \mathbb{N}^*$.
Appelons A_n le nombre de fois ou l'individu a monté une marche après n pas.
Appelons X_n le numéro de la marche sur laquelle se trouve l'individu après n pas.
Voici trois marches aléatoires de 1000 pas : en abscisse le numéro du pas et en ordonnée le numéro de la marche.







Dans notre expérience aléatoire, soit on monte (succès), soit on descend(échec). Il s'agit donc d'une épreuve de Bernoulli.
On répète cette épreuve n fois. Toutes les épreuves ainsi répétée sont identiques et indépendantes.
La variable aléatoire A_n compte le nombre de succès.
Donc A_n suit une loi binomiale $\mathcal{B}\left(n;\frac{1}{2}\right)$

Ainsi $\forall k \in [\![1;n]\!], \mathbb{P}(A_n = k) = \begin{pmatrix} n \\ k \end{pmatrix} \left(\dfrac{1}{2}\right)^k \left(1-\dfrac{1}{2}\right)^{n-k} = \begin{pmatrix} n \\ k \end{pmatrix} \left(\dfrac{1}{2}\right)^k \left(\dfrac{1}{2}\right)^{n-k} = \begin{pmatrix} n \\ k \end{pmatrix} \left(\dfrac{1}{2}\right)^n = \begin{pmatrix} n \\ k \end{pmatrix} \dfrac{1}{2^n}$

$E(A_n) = np = n \times \frac{1}{2} = \frac{n}{2}$
$V(A_n) = np(1-p) = n \times \frac{1}{2}\left(1-\frac{1}{2}\right) = n \times \frac{1}{2} \times \frac{1}{2} = \frac{n}{4}$

X_n est le numéro de la marche où se trouve l'individu après n pas.
A_n est le nombre de fois où il a monté d'une marche.
Donc $n - A_n$ est le nombre de fois où il a descendu une marche.
Ainsi $X_n = A_n - (n - A_n) = A_n - n + A_n = 2A_n - n$

$\forall k \in [\![1;n]\!], \mathbb{P}(X_n = k) = \mathbb{P}(2A_n - n = k) = \mathbb{P}(2A_n = n + k) = \mathbb{P}\left(A_n = \dfrac{n+k}{2}\right)$
$\forall k \in [\![1;n]\!], \frac{n+k}{2} \leqslant n$ donc :
$\forall k \in [\![1;n]\!], \mathbb{P}(X_n = k) = \begin{cases} \begin{pmatrix} n \\ \frac{n+k}{2} \end{pmatrix} \dfrac{1}{2^n} \text{ si } n \text{ et } k \text{ sont de même parité} \\ \qquad 0 \text{ sinon} \end{cases}$
$E(X_n) = E(2A_n - n) = 2E(A_n) - n = 2 \times \frac{n}{2} - n = n - n = 0$
$V(X_n) = V(2A_n - n) = V(2A_n) = 2^2 V(A_n) = 4 \times \frac{n}{4} = n$
donc $\sigma(X_n) = \sqrt{V(X_n)} = \sqrt{n}$

L'inégalité de Bienaymé-Tchebichev nous donne :
$\forall \delta \in \mathbb{R}_+^*, \mathbb{P}(|X_n - 0| \geqslant \delta) \leqslant \frac{n}{\delta^2}$

Soit $N \in \mathbb{N}^*$.
On place maintenant N individus sur l'escalier, tous sur la marche 0, qui procèdent de la même façon que précédemment pour monter ou descendre l'escalier.
Ils font tous n pas.
On a donc N variables aléatoire $X_{n,1}, X_{n,2}, ..., X_{n,N}$ identiques et indépendantes suivant la même loi binomiale $\mathcal{B}\left(n, \frac{1}{2}\right)$ d'espérance 0 et de variance $\sqrt{n}$.
Posons M_N la variable aléatoire moyenne des N variables aléatoires $X_{n,1}, X_{n,2}, ..., X_{n,N}$.
On a donc $M_N = \frac{X_{n,1}+X_{n,2}+...+X_{n,N}}{N}$
L'inégalité de concentration nous donne :
$\forall \delta \in \mathbb{R}_+^*, \mathbb{P}(|M_N| \geqslant \delta) \leqslant \frac{\sqrt{n}}{N\delta^2}$
Autrement dit, pour tout réel strictement positif δ, la probabilité que le numéro moyen de la marche d'arrivée des N individus après n pas n'appartienne pas à l'intervalle $]-\delta;\delta[$ est inférieure à $\frac{\sqrt{n}}{N\delta^2}$.

Et donc la loi des grands nombres nous donne : $\forall t \in \mathbb{R}_+^*, \lim\limits_{N \to +\infty} \mathbb{P}(|M_N| \geqslant t) = 0$ Autrement dit, pour tout réel strictement positif t, la probabilité que le numéro moyen de la marche d'arrivée d'une infinité d'individus après n pas n'appartienne pas à l'intervalle $]-t;t[$ est nulle.