Variables aléatoires

I Quelques définitions

Définition n°1. Variable aléatoire

Étant donnée une expérience aléatoire, définir une variable aléatoire, c'est associer à chaque issue de l'expérience un nombre réel. On la note à l'aide d'une majuscule : X ou Y par exemple.

Exemple n°1. Désigner un événement

$\{X = a\}$: X prend la valeur a
$\{X < a\}$: X prend des valeurs strictement inférieures à a.

Remarque n°1. Une autre notation possible

$\{X \in a\}$ pour $\{X = a\}$
$\{X \in]-\infty\,;\,a[\,\}$ pour $\{X < a\}$

Exemple n°2. Le jeu auquel tout le monde veut jouer

On lance deux fois de suite une pièce de monnaie équilibrée. Si on obtient $Pile$, on gagne 1€, sinon on gagne 2€.
On définit la variable aléatoire X qui, à chaque issue de fin du jeu, associe la somme gagnée par le joueur.

Notre jeu est une expérience aléatoire à deux épreuves indépendantes.
- Épreuve n°1 : on lance la pièce, on note P quand on obtient $Pile$ et F quand on obtient $Face$.

L'univers est $\Omega_1=\{P,F\}$
- Épreuve n°2 : Exactement la même chose, l'univers est alors $\Omega_2=\{P,F\}$
- L'univers de notre expérience aléatoire est alors $\Omega=\{(P,P),(P,F),(F,P),(F,F)\}$

À chacune des quatre issues, on associe un nombre réel :
(P,P) est associé au nombre 2.
(P,F) et (F,P) sont, chacune, associées au nombre 3.
(F,F) est associé au nombre 4.

Nous avons ainsi défini notre variable aléatoire.
$\{X = 2\}$ correspond à (P,P)
$\{X = 3\}$ correspond à (P,F) ou (F,P)
$\{X = 4\}$ correspond à (F,F)

$\{X = 5\}$ correspond à l'ensemble vide…
« $\{X = \text{quelque chose autre que } 2{,}3 \text{ ou } 4\}$ » correspond à l'ensemble vide…

$\{X \leqslant 3\}$ correspond à (P,P) ou (P,F) ou (F,P)
etc...

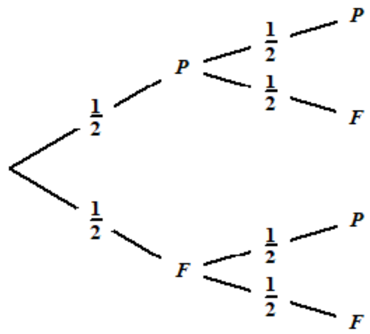
Définition n°2. ***Loi de probabilité d'une variable aléatoire***

Étant donnée une expérience aléatoire sur laquelle on a défini une variable aléatoire X, définir la loi de probabilité de X, c'est associer, à chaque valeur de cette variable aléatoire, la probabilité de l'événement associé.

Exemple n°3. ***Avec le jeu auquel tout le monde veut jouer***

Loi de probabilité de X				
a_i	2	3	4	Total
$P(X = a_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

En effet, grâce à un arbre pondéré :



(P, P) qui correspond à $\{X = 2\}$

(P, F) qui correspond à $\{X = 3\}$

(F, P) qui correspond à $\{X = 3\}$

(F, F) qui correspond à $\{X = 4\}$

On peut déterminer chaque valeur prise par X et la probabilité de l'événement qui lui est associé.

$$P(X = 2) = \frac{1}{2}\times\frac{1}{2} = \frac{1}{4} ,$$

$$P(X = 3) = \frac{1}{2}\times\frac{1}{2} + \frac{1}{2}\times\frac{1}{2} = \frac{2}{4} = \frac{1}{2}$$

$$P(X = 3) = \frac{1}{2}\times\frac{1}{2} = \frac{1}{4}$$

Définition n°3. ***Espérance***

On donne une expérience aléatoire (à n issues possibles) sur laquelle on a défini une variable aléatoire X dont la loi de probabilité est donné ci-dessous.

Loi de probabilité de X					
a_i	a_1	a_2	…	a_n	Total
$P(X=a_i)$	p_1	p_2	…	p_n	1

On appelle espérance de X et on note $E(X)$, le nombre défini par

$$E(X) = p_1\times a_1 + p_2\times a_2 + \ldots + p_n\times a_n$$

Remarque n°2.

Cela représente, la valeur moyenne de X que l'on peut espérer obtenir en répétant l'expérience aléatoire un très grand nombre de fois.

Exemple n°4. ***Toujours avec le jeu auquel tout le monde veut jouer***

$$E(X) = \frac{1}{4}\times 2 + \frac{1}{2}\times 3 + \frac{1}{4}\times 4 = 3$$

En moyenne, on peut espérer gagner trois euros à chaque fois que l'on joue…

II Le cas Bernoulli

Définition n°4. Loi de Bernoulli

On se donne une épreuve de Bernoulli (rappel ici en page3) de paramètre p
On définit une variable aléatoire X qui prend la valeur 1 en cas de succès et 0 en cas d'échec.
On appelle loi de Bernoulli de paramètre p, la loi suivie par X

Loi de Bernoulli de paramètre p notée $B(p)$			
a_i	0	1	Total
$P(X=a_i)$	$1-p$	p	1

Propriété n°1. Espérance d'une variable aléatoire suivant une loi de Bernoulli

Soit X une variable aléatoire suivant une loi de Bernoulli de paramètre p alors : $\boxed{E(X)=p}$

preuve :

$$E(X) = 0\times(1-p) + 1\times p = p$$

III Simulations et échantillons

Définition n°5. échantillon de taille n associé à une épreuves de Bernoulli

Lorsqu'on répète n fois, de façon indépendante, une épreuve de Bernoulli de paramètre p, on obtient une série de n résultats que l'on appelle échantillon de taille n associé à une épreuve de Bernoulli.

Exemple n°5. Simuler un échantillon de taille n

```python
import random

def Bernoulli(p):
    """ epreuve de Bernoulli, p
        est la probabilité de succés
        on renvoie alors 1
        (sinon on renvoie 0)."""
    if random.random() <= p:
        return 1
    else:
        return 0

def echantillon(n,p=0.5):
    """echantillon de taille n
       associé à une épreuve de
       Bernoulli de parametre p
       qui vaut de base 0.5"""
    result = []
    for i in range(n):
        result.append(Bernoulli(p))
    return result

def frequence_des_uns(echantillon):
    """renvoie la frequence des 1
       dans l'echantillon (liste) donné"""
    numerateur = 0
    denominateur = len(echantillon)
    for i in range(denominateur):
        numerateur += echantillon[i]
    return numerateur/denominateur
```

Vous pouvez télécharger le programme ci-dessus en cliquant dessus. (Il comporte quelques annotations supplémentaires pour vous aider à le comprendre)

Plutôt que de truquer une pièce de monnaie afin qu'elle ait 40 % de chance tomber sur Pile puis de la lancer 20 fois en notant le résultat, on préfère effectuer une simulation.

```
>>> mon_echantillon=echantillon(20,0.4)
>>> mon_echantillon
[0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1]
>>> frequence_des_uns(mon_echantillon)
0.35
>>>
```

On simule un échantillon de taille 20 associé à une épreuve de Bernoulli de paramètre 0,4.
Cette simulation peut se faire avec une feuille de calcul ou comme ici avec un programme informatique.
L'intérêt est que l'on peut facilement augmenter la taille de l'échantillon.

IV Étude des échantillons

Définition n°6. Fluctuation d'échantillonnage

Si on réalise plusieurs échantillons de même taille, la fréquence du succès obtenues sur chaque échantillon va varier (fluctuer). On appelle cela la fluctuation d'échantillonnage.

Exemple n°6.

On reprend l'exemple précédent et on simule 10 échantillons de taille 20 d'une épreuve de Bernoulli de paramètre 0,4.

```
>>> for i in range(10):
        print(frequence_des_uns(echantillon(20,0.4)))


0.5
0.3
0.4
0.35
0.3
0.35
0.6
0.65
0.3
0.3
>>> 
```

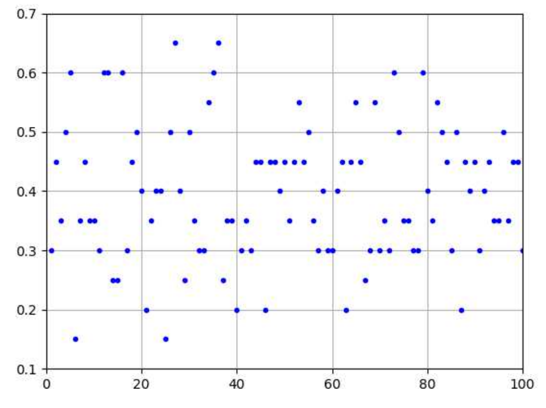
On constate que la fréquence du succès fluctue..

Remarque n°3.

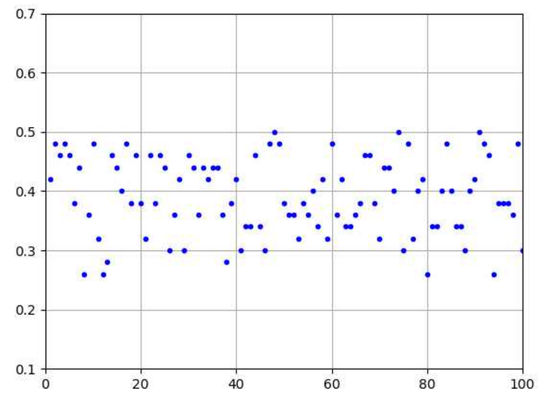
Par contre, plus la taille des échantillons augmente, plus la fluctuation diminue.

Dans les graphiques suivants, chaque point représente la fréquence du succès d'un échantillon.

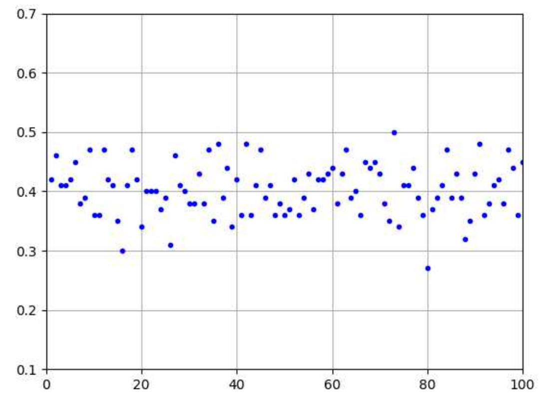
100 échantillons de taille 20
la probabilité de succès de l'épreuve est 0,4



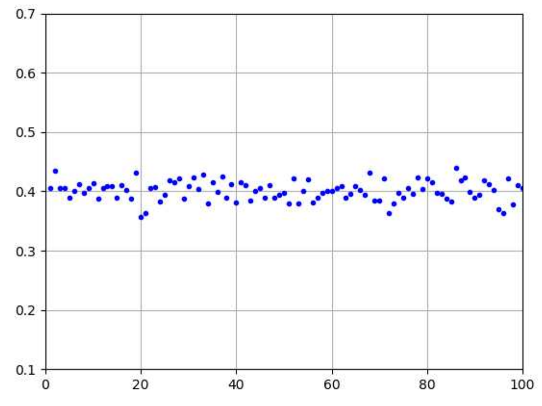
100 échantillons de taille 50
la probabilité de succès de l'épreuve est 0,4



100 échantillons de taille 100
la probabilité de succès de l'épreuve est 0,4



100 échantillons de taille 1000
la probabilité de succès de l'épreuve est 0,4



Remarque n°4.

On peut remarquer que en moyenne :

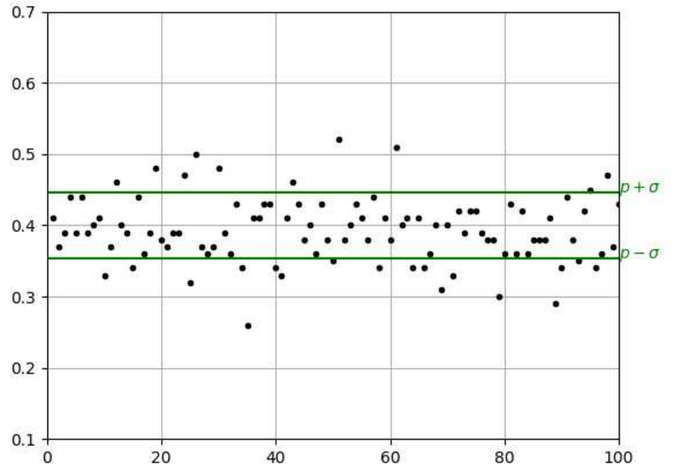
environ 68 % des fréquences sont dans l'intervalle $[p-\sigma\ ;\ p+\sigma]$
environ 95% des fréquences sont dans l'intervalle $[p-2\sigma\ ;\ p+2\sigma]$
environ 99% des fréquences sont dans l'intervalle $[p-3\sigma\ ;\ p+3\sigma]$

où σ est l'écart-type de la série des fréquences (voir la page 5 de ce cours)

Il faut retenir que :

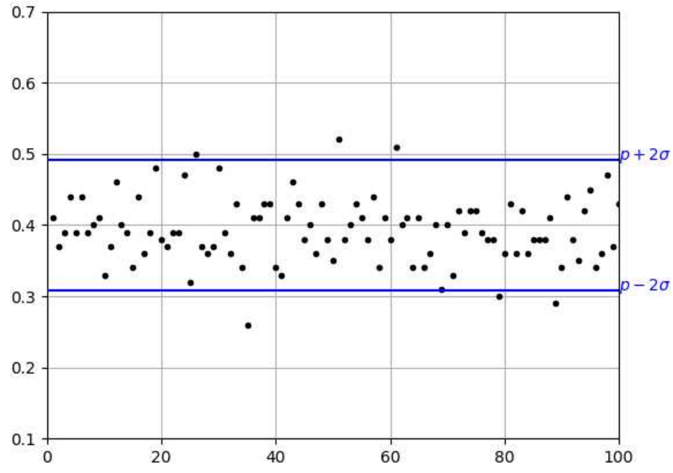
▪ Il n'est pas rare d'obtenir un échantillon dont la fréquence s'écarte de plus d'un écart-type de la fréquence théorique :
cela arrive dans environ 32% des cas
($100\%-68\%$)

Effectivement, il y a pas mal de points à l'extérieur des lignes vertes.



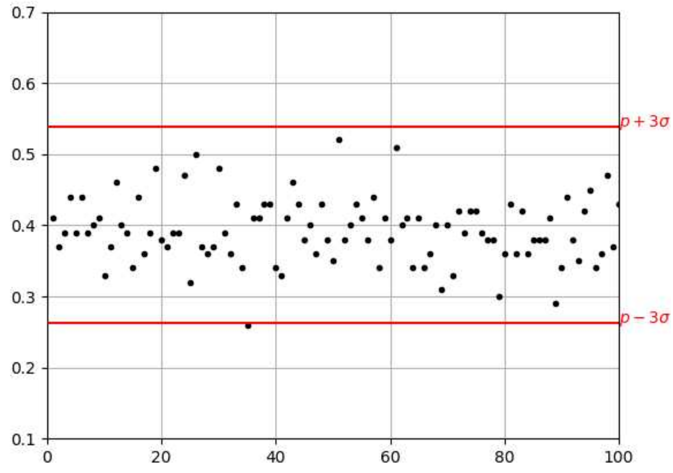
▪ Il est par contre très rare que la fréquence d'un échantillon s'écarte de plus de deux écarts-types de la fréquence théorique :
Cela arrive dans 5 % des cas.

Effectivement, il y a peu de points à l'extérieur des lignes bleues.



▪ Il est encore beaucoup plus rare que la fréquence d'un échantillon s'écarte de plus de trois écarts-types de la fréquence théorique : Cela arrive dans 1 % des cas.

Effectivement, il y a très peu de points à l'extérieur des lignes rouges.



Propriété n°2. Lien entre l'écart-type σ et la taille n de l'échantillon (admis)

La valeur σ n'est pas très éloignée de $\dfrac{1}{2\sqrt{n}}$

Remarque n°5.

Cela signifie qu'en pratique, on prendra $\dfrac{1}{2\sqrt{n}}$ comme valeur de σ .