Échantillonnage

Plan du T.P.

Retour au sommaire

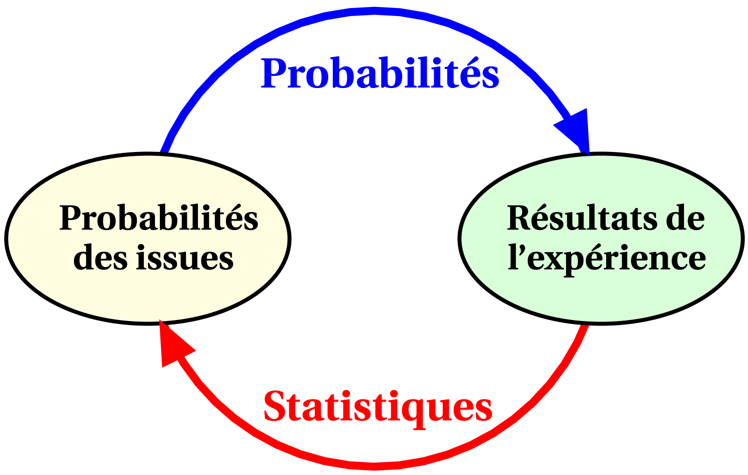
Quelques grandes figures des Probabilités et des Statistiques

Cliquer une image au choix

Blaise Pascal - Jacques Bernoulli - Carl Friedrich Gauss

Pierre-Simon de Laplace - Adolphe Quetelet - Andreï Kolmogorov

1 Probabilités et Statistiques



Probabilités et Statistiques sont deux démarches "inverses"

1.1 Les Probabilités

Exercice 1 :
On dispose dé truqué dont on connaît Les probabilités d'apparition de chaque face. La face 6 apparaît avec probabilité 0,5, et les autres faces apparaissent, chacune, avec probabilité 0,1. Lançons ce dé.

Quelle est la probabilité d'obtenir un nombre pair ? ....

- En probabilités, les **probabilités des issues** de l'expérience sont **connues**.
- Les **résultats de l'expérience** sont ainsi **inconnus**.
- On tente de "prédire" les résultats qu'on l'on peut espérer si l'on réalise l'expérience.

1.2 Les Statistiques

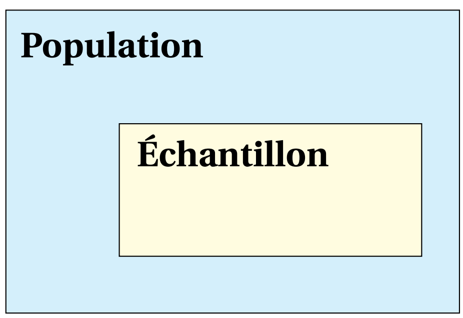
Exercice 2 :
On nous offre un dé truqué.
Comment estimer les probabilités d'apparition de chacune des faces ? ....

À l'aide de quel objet statistique estimerons-nous ces probabilités ? ....

- En Statistiques, les **résultats de l'expérience** sont **connus**.
- Les **probabilités des issues** de l'expérience sont **inconnues**.
- On tente d'estimer ces dernières, à l'aide des résultats de l'expérience.

1.3 L'échantillonnage

- En Statistiques, un échantillon est un ensemble d'individus prélevé dans une population.
- L'échantillonnage désigne les méthodes de sélection et d'étude des échantillons.
- L'objectif est d'estimer, à l'aide d'un échantillon, les caractéristiques de l'ensemble de la population.



2 TP introductif

2.1 Simuler le hasard

Le module `random` offre des fonctions permettant de simuler le hasard.

Exercice 3 : dans la console

a. Commencer par importer dans la console la fonction `randint` depuis la bibliothèque `random` :

```
>>> from random import randint
```

b. simuler des lancers de dé dans la console :

```
>>> randint(1,6)
```

c. Que saisir pour obtenir un nombre entier aléatoire compris entre 1 et 10 ?

....

Exercice 4 : fonction de lancer d'un dé
Écrire dans l'éditeur une fonction `lancer()`, renvoyant le résultat d'un lancer de dé.

2.2 Échantillon

Exercice 5 : lancers successifs

a. Voici une fonction utilisant la fonction `lancer()` de l'exercice précédent :

```
• Fonction frequence_6(n) :
    succes = 0
    Pour ............ :
        de = lancer()
        Si de == 6 :
            ............
    Retourner ............
```

Quel est le rôle de cette fonction ?

....

b. Écrire dans l'éditeur une fonction `frequence_6(n)` :

- prennant un nombre n de lancers en argument,
- renvoyant la fréquence de 6 obtenus parmi ces n lancers.

Définition :
- Un **échantillon** aléatoire de taille n est la liste des résultats obtenus lors de n répétitions indépendantes de la même expérience aléatoire.

2.3 Fluctuation d'échantillonnage

Exercice 6 : échantillons successifs

Á l'aide de la fonction de l'exercice précédent, réaliser plusieurs échantillons de n=100 lancers de dé.

a. La fréquence de 6 obtenus est-elle toujours la même ?

. . . .

b. Autour de quelle valeur semble fluctuer la fréquence f de 6 ?

. . . .

c. Compléter le schéma suivant :

Population

p = . . .

Échantillon

n = . . .

f

Propriété :

- Lorsqu'on prélève des échantillons de taille n dans une population où la proportion d'un caractère est p, on observe que les fréquences f obtenues dans les échantillons, fluctuent autour de la proportion p.
- On parle de **fluctuation d'échantillonnage**.

2.4 La loi des grands nombres

Exercice 7 : lorsque n devient grand

a. Réaliser maintenant des échantillons de lancers de dé de plus grandes tailles.

b. Qu'observe-t-on lorsque n augmente ?

. . . .

Population

p

Échantillon

n

f

Loi des grands nombres :

- On prélève un échantillon de taille n dans une population où la proportion d'un caractère est p.
- Lorsque n est grand, sauf exception, la fréquence f observée dans l'échantillon est proche de la proportion p de la population.

3 Exercices

3.1 Échantillon

Exercice 8 : Pile ou Face
On réalise 5 lancers successifs d'une pièce de monnaie supposée équilibrée. les résultats de chacun des lancers sont notés.

a. Quelle est l'expérience à deux issues qui est ici répétée ?

....

b. Quelles sont les deux issues possibles et leurs probabilités respectives ?

....

c. Quelle est la taille de l'échantillon obtenu ?

....

d. Donner un exemple d'échantillon que nous pourrions ainsi obtenir.

....

e. On souhaite simuler cette expérience à l'aide de Python : pile sera représenté par la valeur 0 et face par 1. Compléter la fonction suivante pour qu'elle renvoie la fréquence de piles obtenue :

```
from random import randint

def lancer() :
        pile = .....
        for compteur in range( ..... ) :
                if randint(0,1) == ..... :
                        ...............
        return ...............
```

Exercice 9 : Contrôle de qualité
Un société produit et commercialise des ampoules électriques. Afin de tester la qualité de sa production, 1000 ampoules sont prélevées au hasard et testées.
Si une ampoule fonctionne moins de 10 000 heures, elle considérée comme non conforme et le résultat de l'expérience est codé par le nombre 0. Sinon, le résultat de l'expérience est codé par le nombre 1.

a. Quelle est l'expérience à deux issues qui est ici répétée ?

....

b. Quelles sont les deux issues possibles ?

....

c. Quelle est la taille de l'échantillon obtenu ?

....

d. Donner un exemple d'échantillon que nous pourrions ainsi obtenir.

....

e. Pourquoi ne pas tester l'ensemble de la production ?

....

Exercice 10 :
On considère l'algorithme suivant :

```
• Fonction mystere(n) :
      nombre ← 0
      Pour compteur allant de 1 à n :
            lancer ← nombre entier aléatoire entre 1 et 6
            Si lancer = 2 ou 4 ou 6 :
                  nombre ← nombre + 1
      Retourner nombre / n
```

a. Quelle expérience aléatoire est ici reproduite n fois ?

....

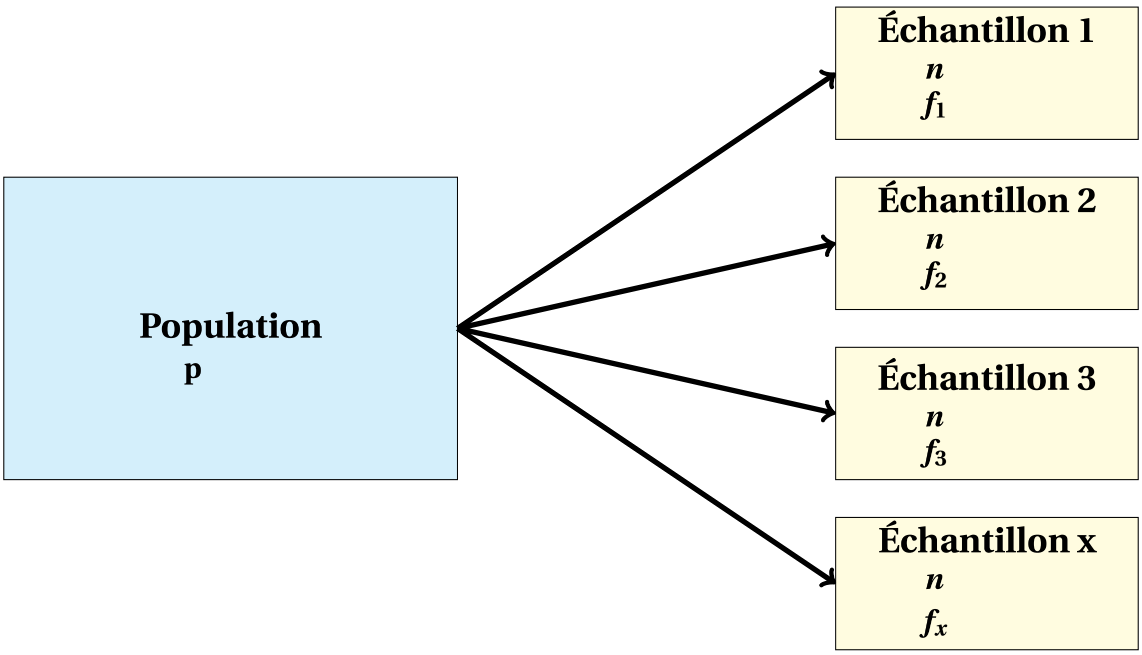
b. Que renvoie cette fonction ?

....

3.2 Fluctuation d'échantillonnage et loi des grands nombres

Nous allons maintenant :

- prélever plusieurs échantillons au sein d'une même population,
- observer les fluctuations d'échantillonnage d'un échantillon à l'autre,
- voir comment limiter ce phénomène de fluctuation d'échantillonnage.

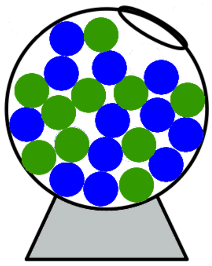


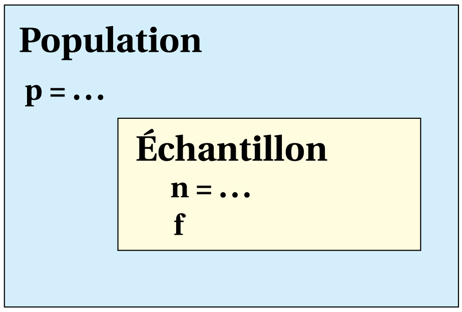
Fluctuation d'échantillonnage

Exercice 11 : Urne

Une urne contient 300 boules bleues et 700 boules vertes.
On a constitué un échantillon de taille 100 de l'expérience consistant à : prélever une boule dans l'urne, noter sa couleur, la remettre dans l'urne.

1. Préciser sur le schéma ci-dessous la population, la proportion p de boules bleues et la taille de l'échantillon.





On constitue maintenant 20 échantillons de taille 100.

Échantillon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Nombres de boules bleues	29	31	40	28	35	26	27	28	33	30	28	25	40	25	36	26	29	25	27	36
Fréquence des boules bleues																				

2. **a.** Calculer la fréquence de boules bleues obtenue dans chaque échantillon.

b. Quelle est l'étendue obtenue entre la plus petite et la plus grande fréquence obtenue ?

....

3. Afin d'obtenir des échantillons de tailles plus significatives, nous allons regrouper, dans le tableau ci-dessous, les 4 premiers échantillons ensembles, les 4 suivants ensembles ...

a. Quelle est la taille des nouveaux échantillons ?

....

b. De combien d'échantillons disposons-nous ainsi ?

....

c. Compléter le tableau ci-dessous.

Échantillon	1	2	3	4	5
Nombres de boules bleues					
Fréquence des boules bleues					

d. Quelle est l'étendue obtenue entre la plus petite et la plus grande fréquence ainsi obtenues ?

....

4. Comparer et expliquer les différences obtenues entre les questions 1 et 2.

....

....

3.3 Estimation d'une proportion dans une population

Exercice 12 : Estimer à l'aide d'un échantillon

Considérons maintenant une situation inverse à celle de l'exercice précédent : la proportion de boules bleues présentes dans l'urne est inconnue et nous allons chercher à l'estimer à l'aide d'un échantillon. On prélève pour cela un échantillon de taille 100 et l'on observe une fréquence 0,43 de boules bleues.

Population
p inconnu

Échantillon
n = . . .
f = . . .

a. Compléter le schéma ci-dessus.

b. Par quelle valeur pourrions-nous estimer la proportion p de boules bleues présentes dans l'urne ?

. . . .

c. Cette estimation fournit-t-elle la proportion exacte de boules bleues ?

. . . .

d. Comment affiner cette estimation ?

. . . .

Exercice 13 : Sondage

En vue de la prochaine élection présidentielle, un candidat fait réaliser un sondage. L'institut de sondage interroge 10 000 électeurs et recueille 1700 intentions de vote pour ce candidat.

Population
p inconnu

Échantillon
n = . . .
f = . . .

a. Compléter le schéma ci-dessus.

b. Par quelle valeur pourrions-nous estimer la proportion p d'électeurs ayant l'intention de voter pour ce candidat ?

. . . .

c. Comment affiner cette estimation ?

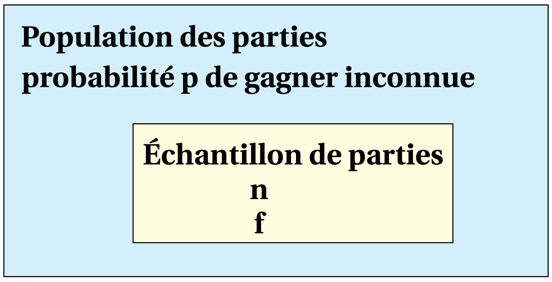
. . . .

Propriété :
Estimation d'une proportion dans une population par une fréquence dans un échantillon

- Considérons une population dans laquelle on cherche à estimer la proportion p des individus qui possèdent un caractère. Prélevons un échantillon de taille n dans la population. Alors la fréquence f observée dans l'échantillon fournit une **estimation** de p.
- Plus la taille n de l'échantillon est grande, plus l'estimation est fiable.

4 T.P. : Estimation d'une probabilité

Exercice 14 : jeu de hasard
Un jeu consiste à lancer 10 fois de suite une pièce équilibrée. Pour gagner, il faut obtenir 5 piles. On souhaite dans cet exercice estimer la probabilité de gagner.



a. Écrire une fonction `partie()` :
- réalisant 10 lancer d'une pièce équilibrée,
- et renvoyant le nombre de pile obtenus.

b. Écrire une seconde fonction `echantillon_partie(n)` :
- réalisant un échantillon de n parties de ce jeu,
- et renvoyant la fréquence des parties gagnées.

a. Proposer une estimation à deux décimales de la probabilité de gagner. On veillera à choisir n suffisamment grand pour stabiliser les fréquences observées dans les échantillons.

Exercice 15 : Simuler le hasard avec la fonction random()
Jusqu'ici, pour générer des nombres aléatoires, nous n'avons utilisé que la fonction randint :

```
from random import randint

nombre_aleatoire = randint(a,b)
```

Remarquons que cette fonction renvoie :
- un nombre aléatoire entier,
- avec des probabilités **égales** pour chacun des entiers compris entre a et b.

a. Quel type de situation, rencontrée dans le cours de probabilités, peut facilement être simulée à l'aide de la fonction `randint` ?

....

En dehors de ce cas, les issues d'une expérience aléatoire n'ont pas toutes les mêmes probabilités de se réaliser (dé truqué, pièce mal équilibrée ...). La fonction `random()` va nous permettre d'obtenir des probabilités différentes pour les différentes issues.

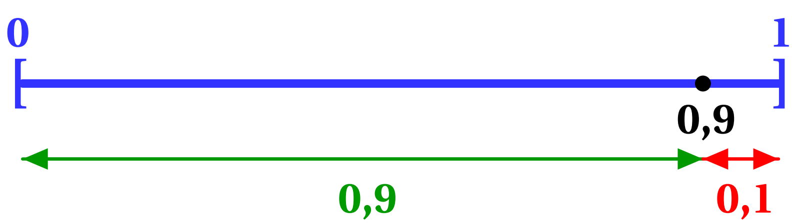
b. Importer dans la console la fonction random() depuis la librairie random :

```
>>> from random import random
```

c. Saisir plusieurs de suite dans la console l'instruction : `>>> random()`, et deviner ce qu'elle renvoie.

....

d. Quelle est la probabilité pour un nombre aléatoire entre 0 et 1 d'être inférieur à 0,9 ?



....

e. Quelles sont les deux valeurs que peut prendre l'expression `random() < 0.9` ?

. . . .

f. Avec quelles probabilités sont prises chacune de ces deux valeurs ?

. . . .

g. Comment obtenir un nombre aléatoire prenant les valeurs :
- `True`, avec la probabilité 0,3
- `False`, avec la probabilité 0,7

. . . .

h. Comment obtenir un nombre aléatoire prenant les valeurs :
- `True`, avec la probabilité 0,49
- `False`, avec la probabilité 0,51

. . . .

Exercice 16 : familles de trois enfants
En France, on estime que la proportion de filles parmi les naissances s'élève à 49%.
On souhaite estimer la probabilité pour un couple projetant d'avoir 3 enfants d'avoir au moins une fille.

Nous représenterons, dans cet exercice, les filles par la valeur `True`, et les garçons par la valeur `False`.

a. Réaliser un schéma explicitant population et échantillon.

b. Écrire une fonction `enfant()` renvoyant :
- `True` (fille), avec probabilité 0,49
- `False` (garçon), avec probabilité 0,51

c. Écrire une fonction `famille()` :
- simulant trois naissances successives dans une famille (à l'aide de la fonction `enfant()`),
- et renvoyant le nombre de filles parmi ces trois enfants.

d. Écrire une fonction `echantillon_famille(n)` :
- constituant un échantillon de n familles (à l'aide de la fonction `famille()`),
- et renvoyant la fréquence dans cet échantillon des familles ayant au moins une fille.

e. Proposer une estimation à deux décimales de la probabilité d'avoir au moins une fille pour une famille de trois enfants. On veillera à choisir n suffisamment grand pour stabiliser les fréquences observées dans les échantillons.

Propriété :
Estimation d'une probabilité par une fréquence observée dans un échantillon

- Considérons une expérience aléatoire et un événement A dont on souhaite estimer la probabilité p. Réalisons un échantillon de taille n de cette expérience. Alors, la fréquence dans l'échantillon des cas où l'événement A est réalisé, fournit une **estimation** de la probabilité p.
- Plus la taille n de l'échantillon est grande, plus l'estimation est fiable.

5 Synthèse

Définition :

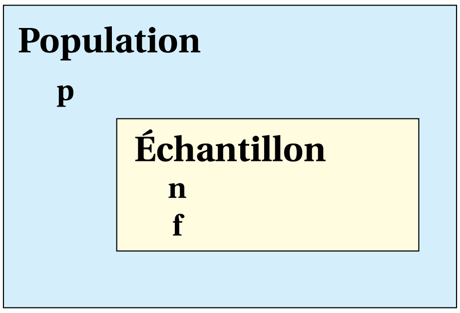
- Un **échantillon** aléatoire de taille n est la liste des résultats obtenus lors de n répétitions indépendantes de la même expérience aléatoire.

Propriété :

- Lorsqu'on prélève des échantillons de taille n dans une population où la proportion d'un caractère est p, on observe que les fréquences f obtenues dans les échantillons, fluctuent autour de la proportion p.
- On parle de **fluctuation d'échantillonnage**.

Loi des grands nombres :

- On prélève un échantillon de taille n dans une population où la proportion d'un caractère est p.
- Lorsque n est grand, sauf exception, la fréquence f observée dans l'échantillon est proche de la proportion p de la population.



Propriété :
Estimation d'une proportion dans une population par une fréquence dans un échantillon

- Considérons une population dans laquelle on cherche à estimer la proportion p des individus qui possèdent un caractère. Prélevons un échantillon de taille n dans la population. Alors la fréquence f observée dans l'échantillon fournit une **estimation** de p.

- Plus la taille n de l'échantillon est grande, plus l'estimation est fiable.

Propriété :
Estimation d'une probabilité par une fréquence observée dans un échantillon

- Considérons une expérience aléatoire et un événement A dont on souhaite estimer la probabilité p. Réalisons un échantillon de taille n de cette expérience. Alors, la fréquence dans l'échantillon des cas où l'événement A est réalisé, fournit une **estimation** de la probabilité p.

- Plus la taille n de l'échantillon est grande, plus l'estimation est fiable.