

◆ TP9 – Simulations de variables aléatoires liées à un schéma de Bernoulli

On rappelle que le module `random` comporte la fonction `random()` qui renvoie un nombre flottant au hasard appartenant à l'intervalle $]0;1[$ et la fonction `randint(a,b)` qui renvoie un entier aléatoire entre `a` et `b` compris.

Question 1.

1. Quelle instruction doit-on écrire pour importer toutes les fonctions du module `random` ?
2. Quelle instruction doit-on écrire si on ne souhaite importer que les fonctions `random` et `randint` du module `random` ?

I. — Simulation de variables suivant des lois usuelles

1) Loi de Bernoulli

Question 2. On considère la fonction `simul` suivante :

```
def simul():  
    return random()
```

Cette fonction simule une variable aléatoire X . Quelle est la loi de X ?

Question 3. On considère la variable aléatoire X de la question précédente. Si $p \in]0;1[$, que vaut $\mathbf{P}(X \leq p)$?

Question 4. Utiliser ce qui précède pour écrire une fonction `bernoulli` prenant en argument un réel `p` appartenant à $]0;1[$ et qui simule une variable aléatoire suivant une loi de Bernoulli de paramètre `p`.

2) Loi binomiale

Question 5. On souhaite écrire une fonction `bernoulli_liste` qui simule, sous la forme d'une liste, n variables aléatoires indépendantes X_1, X_2, \dots, X_n suivant toutes la même loi de Bernoulli de paramètre p .

Parmi les 4 fonctions proposées ci-dessous, certaines conviennent et d'autres non. Identifier lesquelles et expliquer pourquoi il faut éliminer les autres. On vérifiera ensuite ses réponses à l'aide de l'ordinateur.

```
def bernoulli_liste1(n,p):  
    L=[bernoulli(p)]*n  
    return L
```

```
def bernoulli_liste2(n,p):  
    L=[bernoulli(p) for i in range(n)]  
    return L
```

```
def bernoulli_liste3(n,p):  
    L=[]  
    for i in range(n):  
        L+=bernoulli(p)  
    return L
```

```
def bernoulli_liste4(n,p):  
    L=[]  
    for i in range(n):  
        L.append(bernoulli(p))  
    return L
```

Question 6. On rappelle que si X_1, X_2, \dots, X_n sont n variables aléatoires indépendantes suivant toutes la même loi de Bernoulli de paramètre p alors la variable aléatoire $S = X_1 + X_2 + \dots + X_n$ suit une loi binomiale de paramètres n et p .

En utilisant une des fonctions `bernoulli_liste` convenables de la question précédente, écrire une fonction `binomiale` prenant en argument un entier naturel non nul n et un nombre réel p appartenant à $]0;1[$ et qui simule une variable aléatoire suivant une loi binomiale de paramètres n et p .

3) Loi géométrique

Question 7. Rappeler l'interprétation d'une loi géométrique de paramètre p dans le cadre d'un schéma de Bernoulli infini où le succès a une probabilité p .

En déduire une fonction `geometrique` prenant en argument un réel p compris entre 0 et 1 et qui simule une variable aléatoire suivant une loi géométrique de paramètre p . On utilisera la fonction `bernoulli`.

4) Approximation de la loi de Poisson par une loi binomiale

On fixe un réel $\lambda > 0$. Pour tout $n \in \mathbb{N}^*$, on considère une variable aléatoire X_n suivant une loi binomiale de paramètres n et $\frac{\lambda}{n}$. On désigne, de plus, par Y une variable aléatoire suivant une loi de Poisson de paramètre λ .

On peut démontrer (c'est le but de l'exercice 4) que lorsque n devient grand, la variable aléatoire X_n est une bonne approximation de Y . Autrement dit, lorsque n devient grand, on peut approcher la loi de Poisson $\mathcal{P}(\lambda)$ par la loi binomiale $\mathcal{B}\left(n, \frac{\lambda}{n}\right)$.

Question 8. Écrire une fonction `approx_Poisson` qui prend en arguments un réel `lamb` strictement positif et un entier naturel `n` et qui simule une variable aléatoire dont la loi est approximativement une loi de Poisson pour `n` assez grand.

Question 9. Écrire une fonction `liste_approx_Poisson` qui prend en arguments un entier naturel `N` non nul, un réel strictement positif `lamb` et un entier naturel `n` et qui renvoie une liste de `N` résultats renvoyés par `approx_Poisson(lamb,n)`.

Question 10. Écrire une fonction `frequence` prenant en arguments un nombre `x` et une liste de nombres `L` et qui renvoie la fréquence d'apparition de `x` dans `L` i.e. le nombre d'occurrences de `x` dans la liste `L` divisé par le nombre total d'éléments de la liste.

Question 11. Le tableau ci-dessous donne la valeur arrondie à 10^{-3} près de $\mathbf{P}(Y = k)$ pour différentes valeurs de λ et de k .

$\lambda \backslash k$	0,5	1	1,5	2	2,5	3	4	5	6
0	0,607	0,368	0,223	0,135	0,082	0,050	0,018	0,007	0,002
1	0,303	0,368	0,335	0,271	0,205	0,149	0,073	0,037	0,015
2	0,076	0,184	0,251	0,271	0,257	0,224	0,147	0,084	0,045
3	0,013	0,061	0,126	0,180	0,214	0,224	0,195	0,140	0,089
4	0,002	0,015	0,047	0,090	0,134	0,168	0,195	0,175	0,134
5	0,000	0,003	0,014	0,036	0,067	0,101	0,156	0,175	0,161
10	0,000	0,000	0,000	0,000	0,000	0,001	0,005	0,018	0,041

Comparer ces valeurs avec les résultats renvoyés, pour différentes valeurs de `N` et de `n`, par `frequence(k,liste_approx_Poisson(N,lamb,n))` lorsque `k` prend la valeur k et `lamb` la valeur λ .

II. — Estimation de l'espérance

D'après la loi des grands nombres, que nous verrons en fin d'année, lorsqu'on fait la moyenne des valeurs prises par un grand nombre de variables aléatoires indépendantes suivant toutes la même loi alors on obtient un résultat proche de l'espérance commune à ces variables. On va utiliser cela pour estimer les espérances de variables suivant l'une des lois précédentes.

Question 12. Écrire une fonction `moyenne` qui prend en argument une liste `L` de nombres (entiers ou flottants) et qui renvoie la moyenne des éléments de `L`.

Question 13. Écrire une fonction `estimation_bernoulli` qui prend en argument un entier naturel non nul `N` et un réel `p` compris entre 0 et 1 et qui renvoie une estimation de l'espérance d'une variable suivant une loi de Bernoulli de paramètre p à partir de `N` valeurs d'une telle variable. On utilisera les fonctions `bernoulli_liste` et `moyenne`.

Question 14. Afficher le résultat renvoyé par `estimation_bernoulli(10000,p)` pour différentes valeurs de `p`. Ceci est-il en accord avec la valeur de l'espérance d'une variable aléatoire suivant une loi de Bernoulli ?

Question 15. Reprendre, en adaptant, les deux questions précédentes pour estimer l'espérance d'une variable aléatoire suivant une loi binomiale puis pour estimer l'espérance d'une variable aléatoire suivant une loi géométrique.

III. — Estimation d'un paramètre par intervalle de confiance

On considère un schéma de Bernoulli dont on ne connaît pas le paramètre p . On aimerait obtenir une estimation de p .

Si on considère la variable aléatoire X égale à 1 en cas de succès et 0 en cas d'échec alors $X \hookrightarrow \mathcal{B}(p)$ et $\mathbf{E}(X) = p$. La question 11 donne une façon d'estimer l'espérance de X et donc d'estimer p .

Cependant, dans la pratique, il n'est pas toujours possible de procéder de la sorte pour des raisons de coup financier, de temps disponible ou pour d'autres raisons matérielles. Par exemple, si on veut tester si des boîtes de conserve sont conformes ou pas et qu'il est nécessaire de les ouvrir pour cela, on ne va pas tester toute la production, et même faire un test sur un échantillon de grande taille peut s'avérer extrêmement coûteux.

On va considérer la situation suivante. Une entreprise fabrique des objets. Chaque objet, indépendamment des autres, a une probabilité $p \in [0; 0,1[$ d'être défectueux.

On aimerait estimer la probabilité p à partir d'un échantillon, aussi petit que possible, d'objets prélevés au hasard dans la production.

Question 16. Si on entre l'instruction suivante, à quel intervalle appartient le nombre `p` ?

```
p=random()/10
```

Question 17. Écrire une fonction `production` qui prend en argument un entier `N` et qui renvoie une liste `L` de `N` valeurs d'une variable $X \hookrightarrow \mathcal{B}(p)$ où p est un réel inconnu choisi aléatoirement entre 0 et 0,1. On utilisera la fonction `bernoulli_liste`.

Cette liste représente la production de l'usine, la valeur 0 signifiant que l'objet n'a pas de défaut et la valeur 1 signifiant que l'objet possède un défaut.

Question 18. Écrire une fonction `echantillon` qui prend en arguments une liste `L` et un entier `k` inférieur à la longueur de la liste et qui renvoie une liste de `k` éléments choisis aléatoirement à des rangs distincts dans `L`. On pourra utiliser la méthode de listes `.pop()` mais on écrira une fonction sans effet de bord.

Par exemple, si la liste est `L` est `[0,1,1,0,0,0]` et si `k` vaut 4 alors une liste possible est la liste `[1, 0, 0, 0]` obtenue en prenant dans `L` les termes de rangs respectifs 2, 5, 0 et 3.

La fonction `echantillon` permet d'obtenir un échantillon aléatoire de k valeurs de la liste, ce qui représente donc un échantillon de k objets de la production. Notons f la fréquence de 1 dans cet échantillon, ce qui représente la fréquence d'objets défectueux dans l'échantillon.

Un théorème permet d'affirmer que, dans au moins 95% des cas, la probabilité p appartient à l'intervalle $\left[f - \frac{1}{\sqrt{k}}; f + \frac{1}{\sqrt{k}}\right]$. Cet intervalle est appelé intervalle de confiance de p au seuil de confiance 95%.

Question 19. Soit $m \in \mathbb{N}$. Quelle doit-être la taille k d'un échantillon pour obtenir, grâce à l'intervalle de confiance, un encadrement d'amplitude 10^{-m} de p . (On rappelle que l'amplitude d'un encadrement du type $a \leq p \leq b$ est le réel $b - a$).

Question 20. Compléter le code de la fonction `int_conf` qui prend en arguments deux entiers naturels `N` et `m`, qui simule, sous forme d'une liste, une production de `N` objets, sélectionne un échantillon dans cette liste de sorte à calculer et renvoyer un encadrement d'amplitude 10^{-m} de la probabilité p .

```
from math import sqrt

def int_conf(N,m):
    L=...
    k=...
    M=...
    f=...
    return ... , ...
```

Faire différents tests avec la fonction `int_conf` avec $N = 100\,000$ ou $N = 1\,000\,000$ et $m = 1$ ou $m = 2$ en prenant garde au fait que la taille de l'échantillon doit rester inférieur à la taille de la liste.

Question 21. En utilisant certaines des fonctions précédentes, écrire une fonction `confiance` qui prend en argument une probabilité p , une taille de production `N`, une taille d'échantillon `k` et un nombre de simulations `S` et qui renvoie le nombre de points de pourcentage d'intervalles de confiance qui contiennent effectivement la probabilité p lors de `S` simulations.

Vérifier, à l'aide de cette fonction, le seuil de confiance de 95% garanti par la théorie.

IV. — Exercices

Exercice 1. Deux personnes disposent chacune d'une pièce de monnaie qui tombe sur « face » avec une probabilité $p \in]0; 1[$. Ces deux personnes lancent simultanément leur pièce et répètent les lancers jusqu'à obtenir « face » pour la première fois en même temps. On note X la variable aléatoire égale au nombre de lancers nécessaires.

1. Écrire une fonction `nb_lancers` qui prend en argument un réel p compris entre 0 et 1 et qui simule la variable aléatoire X .
2. Écrire une fonction `estimation_esp` qui prend en arguments un réel p strictement compris entre 0 et 1 et un entier N strictement positif et qui permet d'estimer l'espérance de X à partir de N simulations de X . On utilisera la fonction `nb_lancers`.

3. En utilisant la fonction précédente avec N égal à 10 000, estimer l'espérance de X pour p valant 0,3, 0,5 et 0,9.
4. Montrer que X suit une loi géométrique dont on exprimera le paramètre en fonction de p puis vérifier la pertinence des estimations de la question précédente.

Exercice 2. On considère une urne qui contient initialement 1 boule blanche et 1 boule noire. On effectue des tirages successifs dans cette urne de la manière suivante :

- si la boule tirée est noire, on s'arrête ;
- si la boule tirée est blanche, on la remet dans l'urne accompagnée d'une autre boule blanche.

On note X la variable aléatoire égale au nombre de tirages effectués i.e. au nombre de tirages nécessaires pour obtenir la boule noire.

1. Écrire une fonction `simul_X` (sans argument) qui simule la variable aléatoire X .
2. Écrire une fonction `loi_X` qui prend en arguments deux entiers n et N et qui permet d'obtenir une estimation de la probabilité de l'évènement $\{X = n\}$ à partir de N simulations de X .
3. En prenant N égal à 10 000, estimer les probabilités que X soit égale à 1, à 2, à 3 et à 4.
4. a. Démontrer que, pour tout $n \in \mathbb{N}^*$, $\mathbf{P}(X = n) = \frac{1}{n(n+1)}$.
 b. Vérifier la pertinence des estimations précédentes.
 c. La variable aléatoire X admet-elle une espérance ? Si oui, la calculer.

Exercice 3. Soit $n \in \mathbb{N}^*$. On considère une urne contenant n boules numérotées de 1 à n . On tire successivement et avec remise deux boules dans l'urne et on note Y le plus grand des deux numéros obtenus.

1. Écrire une fonction `simul_Y` qui prend en argument un entier non nul n et qui simule Y .
2. Écrire une fonction `loi_Y` qui prend en arguments deux entiers n et N , qui simule N réalisations de Y et qui renvoie une liste de longueur n contenant les fréquences des évènements $\{Y = 1\}$, $\{Y = 2\}$, ..., $\{Y = n\}$.
3. Déterminer la loi de Y .
4. Vérifier la pertinence des résultats renvoyés par la fonction `loi_Y` pour différentes valeurs de n et de N .

Exercice 4. On fixe un réel $\lambda > 0$ et un entier $k \in \mathbb{N}$. Pour tout $n \geq k$, on considère une variable aléatoire X_n suivant une loi binomiale de paramètres n et $\frac{\lambda}{n}$. On désigne, de plus, par Y une variable aléatoire suivant une loi de Poisson de paramètre λ .

1. Soit un entier $n \geq k$. Rappeler les valeurs de $\mathbf{P}(X_n = k)$ et $\mathbf{P}(Y = k)$.
2. Soit un entier $n \geq k$. Rappeler les valeurs de $\mathbf{E}(X_n)$, $\mathbf{V}(X_n)$, $\mathbf{E}(Y)$ et $\mathbf{V}(Y)$ puis vérifier que $\mathbf{E}(X_n) \xrightarrow{n \rightarrow +\infty} \mathbf{E}(Y)$ et $\mathbf{V}(X_n) \xrightarrow{n \rightarrow +\infty} \mathbf{V}(Y)$.
3. Montrer que, lorsque n tend vers $+\infty$, $\binom{n}{k} \sim \frac{n^k}{k!}$.
4. Soit un entier $n \geq k$. Écrire $\left(1 - \frac{\lambda}{n}\right)^{n-k}$ à l'aide de la fonction exponentielle.
5. Dédire de la question précédente que $\lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}$.
6. Conclure que $\mathbf{P}(X_n = k) \xrightarrow{n \rightarrow +\infty} \mathbf{P}(Y = k)$.