

# Challenge Graines de Sondeur

## Sujet 3 : algorithmes de sélection d'un échantillon

L'étude statistique d'une population se base généralement sur l'étude d'un échantillon. Supposons que l'on souhaite estimer l'argent de poche reçu en moyenne chaque mois par les N=20 élèves d'une classe. Pour éviter d'avoir à interroger chacun des 20 élèves, on peut par exemple sélectionner un échantillon de n=4 élèves selon un **tirage avec remise**.

Num de tirage	Individus de la population																			
1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
4	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T

Dans l'exemple ci-dessus, l'individu F est sélectionné 2 fois (tirages 1 et 3), et les individus C et S sont sélectionnés chacun 1 fois. Si l'argent de poche mensuel des individus C, F et S est de 30, 40 et 10 € respectivement, l'argent de poche moyen calculé sur l'échantillon des élèves vaut :

$$(1 \times 30\text{€} + 2 \times 40\text{€} + 10\text{€}) / 4 = 30\text{€}.$$

On peut l'utiliser comme **estimation** (cf. glossaire) de l'argent de poche moyen de l'ensemble de la classe. On peut montrer que cette estimation coïncide en **espérance** (cf. glossaire) avec la véritable valeur moyenne de l'argent de poche dans la classe.

*Le tirage avec remise est une méthode possible permettant de sélectionner un échantillon dans une population, mais en pratique on utilise plutôt des méthodes de tirage **sans remise** (i.e., un individu ne peut être sélectionné qu'une seule fois). Parmi les méthodes de sélection les plus utilisées, nous donnons deux exemples ci-après.*

**Echantillonnage systématique** : on tire d'abord au hasard un individu parmi les N/n=5 premiers. Si l'individu numéroté i est sélectionné, on tire également tous les individus espacés de 5 par rapport à l'individu i, c'est-à-dire les individus i, i+5, i+10 et i+15.

Num de tirage	Individus de la population																				
1	A	B	C	D	E																
2						F	G	H	I	J											
3											K	L	M	N	O						
4																P	Q	R	S	T	

Dans l'exemple, le 1<sup>er</sup> tirage ne concerne que les individus de A à E, et on tire l'unité D. On prend ensuite dans l'échantillon les individus I (espacé de 5 par rapport à D), N (espacé de 10) et S (espacé de 15). On peut là aussi utiliser comme estimation l'argent de poche moyen calculé sur l'échantillon.

**Echantillonnage simple sans remise** : on procède comme pour l'échantillonnage avec remise, mais si un individu est sélectionné lors d'un tirage, il ne peut plus être sélectionné lors des tirages suivants.

Num de tirage	Individus de la population																			
1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2	A	B	C	D	E	F	G	H	I	J	K	L	M	N		P	Q	R	S	T
3	A	B	C	D	E	F	G		I	J	K	L	M	N		P	Q	R	S	T
4	A	B	C	D	E	F	G			J	K	L	M	N		P	Q	R	S	T

# Challenge Graines de Sondeur

Dans l'exemple, on tire l'individu O lors du 1<sup>er</sup> tirage : cet individu ne peut donc pas être concerné par les tirages suivants. On sélectionne ensuite successivement les individus H, I et T dans l'échantillon. On peut là aussi utiliser comme estimation l'argent de poche moyen calculé sur l'échantillon.

## **Travail à réaliser**

L'objectif de ce sujet est d'étudier les propriétés de l'un de ces deux algorithmes sans remise, et de les comparer à la méthode d'échantillonnage avec remise.

### **1) Programmation**

Proposer un algorithme permettant de sélectionner un échantillon selon la méthode de tirage avec remise. De même, proposer un algorithme permettant de sélectionner un échantillon selon une des méthodes de tirage sans remise.

Les implémenter à l'aide du langage Python, ou d'un autre langage de programmation.

### **2) Avantages et inconvénients des méthodes**

Pour voir les avantages et les inconvénients d'une méthode de tirage, vous devez tirer un grand nombre d'échantillons (par exemple, 100) selon cette méthode. Vous devez calculer ensuite l'argent de poche moyen sur chaque échantillon tiré. Donc, vous obtiendrez un tableau avec 100 estimations de l'argent de poche obtenues selon la méthode de tirage choisie.

En réalisant ces tirages répétés, trouvez les avantages et les inconvénients des méthodes proposées : le tirage avec et sans remise, le tirage systématique.

Si vous voulez réaliser une enquête à la sortie du lycée, laquelle de ces méthodes vous semble adaptée?

### **3) Comparaison des méthodes**

Pouvez-vous vérifier qu'avec chacune des méthodes, chaque individu a la même chance d'être sélectionné dans l'échantillon? Vous pouvez procéder en calculant directement ces probabilités, ou en répétant de nombreux tirages par exemple.

Pouvez-vous trouver des cas où un des trois algorithmes fournit, par rapport aux deux autres, une estimation généralement plus proche de la vraie moyenne? Là encore, on pourra procéder en répétant un grand nombre de tirages.