

Capture - Marquage - Recapture

Savoir-faire

- Quantifier l'effectif d'une population ou d'un taxon plus vaste à partir de résultats d'échantillonnage.
- Estimer une abondance par la méthode de capture, marquage, recapture, fondée sur le calcul d'une quatrième proportionnelle.
- Simuler des échantillons de même effectif pour visualiser la fluctuation d'échantillonnage.
- En utilisant une formule donnée pour un intervalle de confiance au niveau de confiance de 95 %, estimer un paramètre inconnu dans une population de grande taille à partir des résultats observés sur un échantillon.

Partie 1 : Fluctuation d'échantillonnage

On s'intéresse à une très grande collection de billes. On décide d'en marquer 30% avec de la peinture avant de les mélanger avec les autres.

On souhaite maintenant effectuer des prélèvements de taille variable dans notre stock de billes et, dans chacun de ces échantillons, compter la part d'entre elles qui sont marquées avec de la peinture.

Cette part sera-t-elle proche des 30% effectivement marquées dans la population initiale ?

On peut réaliser cette expérience avec de vraies billes. On peut aussi réaliser l'expérience à l'échelle de la classe, avec des feuilles, des stylos, ou le matériel à disposition : essayez vous-mêmes !

Cependant, pour pouvoir réaliser nos expériences sur des échantillons en nombre et en taille variable (et surtout pour économiser du temps et des litres de peinture !), nous allons réaliser une simulation en Python.

Simulation d'un échantillon

On considère un échantillon de taille t , c'est-à-dire composés de t individus. Parmi eux, 30% sont marqués. La proportion p d'individus marqués dans cette population est égale à 0.3.

Pour représenter cette expérience avec une simulation Python, on génère t nombres aléatoires compris entre $[0; 1[$ à l'aide de la commande **random**. Tous les nombres inférieurs à 0.3 symbolisent des individus marqués, les autres non. On compte ensuite dans l'échantillon le nombre d'individus marqués, c'est-à-dire le nombre de valeurs inférieures à 0.3.

1. Télécharger le script¹ puis compléter la fonction **echantillon(t)**. On donne ci-dessous la première

1. <https://my.numworks.com/python/elodie-gamot/fluctuation>

partie du script :

```
1 from math import *
2 from random import random
3 from matplotlib.pyplot import scatter, show
4 #L'import de ce module est nécessaire pour la suite du code
5 def echantillon(t):
6     marque= ...
7     for i in range(t):
8         individu=random()
9         if individu< ... :
10             marque+=1
11     return ...
```

2. Exécuter sur la calculatrice `echantillon(100)` trois fois de suite. Quelles sont les valeurs obtenues ? Quelle interprétation peut-on en faire dans le contexte de notre énoncé ?
3. Modifier le programme pour qu'il retourne non plus le nombre de valeurs inférieures à 0.3 mais leur fréquence par rapport à la taille de l'échantillon. Tester plusieurs fois la fonction avec `echantillon(50)` puis `echantillon(500)`. Qu'observe-t-on ?

Représentation graphique avec matplotlib

On souhaite maintenant procéder en même temps à l'étude de n échantillons de taille t . Pour cela, nous allons représenter graphiquement les différents résultats obtenus pour observer leur fluctuation avec la fonction `graph(n,t)` présente dans le même script.

1. Expliquer ce que fait le programme de la fonction `graph(n,t)` sachant que la fonction `scatter(X,Y)` du module `matplotlib.pyplot` permet de représenter un nuage de points utilisant la liste `X` en abscisses et la liste `Y` en ordonnées.

On ne s'intéresse pas pour le moment aux lignes précédées du symbole `#` et qui ne sont pour le moment pas actives.

```
1 def graph(n,t):
2     x=[k for k in range(n)]
3     y=[echantillon(t) for k in x]
4     scatter(x,y)
5     show()
```

2. Lancer `graph(120,100)` plusieurs fois. Autour de quelle valeur se répartissent les fréquences observées dans les échantillons ? Retrouve-t-on les résultats obtenus précédemment ?
3. On supprime maintenant les symboles `#`, ce qui permet l'activation des deux lignes en fin de programme. Elles vont permettre de calculer la fréquence de résultats compris dans l'intervalle $\left[0.3 - \frac{1}{\sqrt{n}}; 0.3 + \frac{1}{\sqrt{n}}\right]$.

```
1 def graph(n,t):  
2     x=[k for k in range(n)]  
3     y=[echantillon(t) for k in x]  
4     compte=len([k for k in y if k>=0.3-1/sqrt(n) and k<=0.3+1/sqrt(n)])  
5     print(compte/n)  
6     scatter(x,y)  
7     show()
```

Lancer à nouveau `graph(120,100)` plusieurs fois.

Quelle valeur est retournée après l'affichage de la représentation graphique ?

4. Quelle conclusion peut-on faire à partir de toutes nos observations ?

Partie 2 : Capture Marquage Recapture

On s'intéresse à une méthode de recensement de la population animale appelée couramment « Capture Marquage Recapture » (CMR). Cette technique consiste à capturer une partie de la population étudiée et à en marquer tous les individus.

Par « marquage », on entend n'importe quelle technique permettant d'identifier un animal. On peut par exemple prendre en photographie ses marques naturelles. Il faut seulement s'assurer que ces marques sont uniques et ne peuvent pas être perdues, et s'il s'agit d'un marquage effectué par les chercheurs, elles ne doivent évidemment pas affecter la survie des individus. Les individus marqués sont ensuite relâchés dans leur environnement.

On procède par la suite à la capture d'un nouvel échantillon, dans lequel on recense les individus marqués. En faisant l'hypothèse que la fréquence d'individus marqués dans cet échantillon est proportionnelle au nombre d'individus marqués au sein de la population entière, on espère pouvoir établir un recensement relativement fiable de cette population.

On procède ainsi à la capture et au marquage de 654 perruches dans une zone donnée. Lors de la recapture, on dénombre 289 individus marqués sur un échantillon de 1189 perruches.

1. En suivant le principe de la méthode CMR, à combien peut-on estimer la population totale de perruches dans cette zone ?
2. Pourquoi s'agit-il d'une estimation et non d'un chiffre exact ?
3. On appelle **intervalle de confiance** l'intervalle dans lequel se trouve une proportion p avec une certitude de 95%.

Cet intervalle se calcule $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}} \right]$ avec f la fréquence du caractère observée dans un échantillon de taille n .

Dans notre exemple, quelle fréquence de perruches marquées observe-t-on lors de la recapture ? En déduire l'intervalle dans lequel se trouve la proportion p (à 10^{-3} près).

4. En déduire un intervalle dans lequel se trouve la population totale de perruches.
5. L'intervalle obtenu par cette méthode doit cependant lui aussi être envisagé avec beaucoup de précautions. Selon vous, quelles sont les limites de la méthode CMR ?

Astuce : Il est possible de calculer un intervalle de confiance avec la calculatrice. Attention toutefois, car la formule utilisée pour ce calcul n'est pas au programme du lycée et aboutit à un résultat beaucoup plus précis ! Cela permet toutefois de vérifier la cohérence des calculs.

Il suffit, dans l'application **Probabilités**, de sélectionner le menu **Intervalles > Une proportion**.

Dans notre exemple, le nombre de succès correspond au nombre de perruches marquées, soit 289, pour un échantillon de taille 1189. Par défaut, on peut laisser le seuil de confiance à 95% mais il est possible d'augmenter ce seuil de confiance pour obtenir un résultat moins précis. Ici, un seuil de confiance de 98% nous donne un résultat proche de celui que nous avons calculé.

Une fois les données entrées, il suffit ensuite de passer aux écrans suivants pour aboutir à l'intervalle de confiance.

rad PROBABILITES	
Intervalle z pour une proportion	
x Nombre de succès	289
n Taille de l'échantillon	1189
Seuil de confiance	0.98
Suivant	

