Compter sans compter

Année 2023– 2024

Anaïs Cambria-Pisciotta, Lilah Chambet-Weil, Camille Faucillion, Ida Wiechers (3e)

Tommaso Lapeyrere (2nde)

Adrien Tapissier et Mehdi Touhami (1ere)

Établissement : Lycée français Vincent van Gogh - La Haye, Pays-Bas

Enseignant·e(s) : Stéphane BERINGUE, Florence DECOOL

Chercheur : Pierre ALBERT, université d'Utrecht aux Pays-Bas.

Sommaire

I. Problématique

Les biologistes des poissons recherchent une méthode pour estimer le nombre de poissons (d'une certaine espèce) présents dans un lac.
Ils ont prévu de le faire en attrapant et en marquant les poissons puis ils les relâchent dans le lac. Lendemain, ils attrapent des poissons et comptent ceux qui sont marqués.
Qu'en pensez-vous ?

II. Présentation du problème

Nous pouvons reformuler le problème avec les notations suivantes :

> Un lac contient un nombre N de poissons. Nous cherchons à déterminer cette population.
>
> Le premier jour, on capture une quantité M de poissons. On les marque puis on les relâche.
>
> Le deuxième jour, on repêche un nombre n de poissons et on compte le nombre m de poissons marqués.
>
> Comment déterminer la quantité N de poissons ?

On définit pour l'ensemble du document :

- N correspond à la population totale
- M correspond à la population capturée
- n correspond à la population recapturée
- m correspond à la population recapturée marquée

Nous définissions également deux situations particulières qui sont à interdire:

- nous ne piochons pas de poissons marqués $m = 0$.
- nous avons aucun poisson dans le lac $N = 0$.

III. Calculs et Lois

Exemple d'une situation :

Le premier jour on a marqué 50 poissons, donc $\boldsymbol{M} = 50$.

Le deuxième jour on repêche 30 poissons, on en a 15 déjà marqué donc $\boldsymbol{n} = 30$ et $\boldsymbol{m} = 15$.

On a alors une fréquence d'observation de $f = \frac{15}{30} = \frac{1}{2}$

En gardant les mêmes proportions, on peut alors supposer qu'il y a 100 poissons au total dans le lac, soit $\boldsymbol{N} = 100$.

Loi

La **loi capture-marqué-recapture** permet de calculer une population de façon indirecte mais approximative :

$$\frac{\boldsymbol{M}}{\boldsymbol{N}} = \frac{\boldsymbol{m}}{\boldsymbol{n}}$$

On peut en déduire :

- Fréquence d'observation $f = \frac{\boldsymbol{m}}{\boldsymbol{n}}$
- $\boldsymbol{N} = \boldsymbol{M} \times \frac{\boldsymbol{n}}{\boldsymbol{m}} = \boldsymbol{m} \times \frac{1}{f}$

IV. Les différentes recherches

A. Premières tentatives

1) Exemple

Pour rappel, on a noté :

- N: Population totale
- M: Population capturée
- $\frac{M}{N} = \frac{m}{n}$ d'où $N = M \times \frac{n}{m}$
- n: Population recapturée
- m: Population recapturée marquée

Lors d'un exemple, on a fixé N = 70 et en faisant une simulation on obtient:

- $M_{\text{simulation}} = 30$
- $n_{\text{simulation}} = 30$
- $m_{\text{simulation}} = 17$

Par proportionnalité, on obtient $N_{simulation} = M \times \frac{n}{m} = 30 \times \frac{30}{17} = 52{,}94$

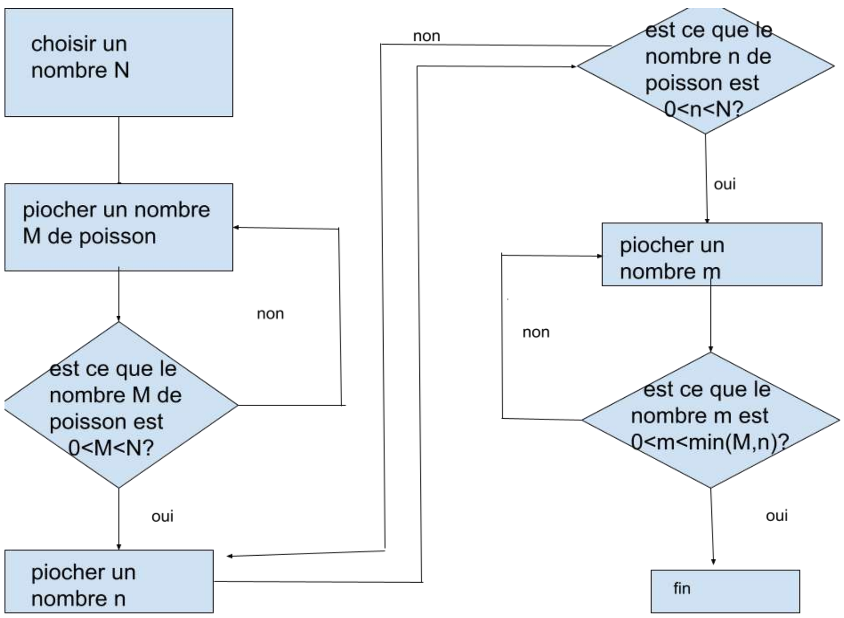
On trouve $N_{simulation} =$ 53 au lieu que notre $N =$ 70 ce qui nous fait un **écart de 32 %**

$$\frac{V_F - V_I}{V_I} \times 100 = \frac{70 - 53}{53} \times 100 = 32\ \%$$

<u>Notre analyse du résultat :</u> on trouve que le résultat obtenu très loin du N initial. On suppose donc que notre échantillon était trop petit

Alors nous avons eu l'idée de faire des moyennes avec 10 captures mais on n'obtenait toujours pas des résultats satisfaisants

2) Script de résolution en langage naturel



B. Autres analyses du problème

Une de nos premières analyses du problème a été d'utiliser la proportionnalité, dans un tableau Excel.

B	C	D	E	F	G	H	I	J
Nb poissons marqués	7	7	7	7	7	7	7	7
Nb poissons piochés marqués	1	5	2	2	1	1	1	1
Nombre n	49	9.8	24.5	24.5	49	49	49	49
Vrai nombre n	50	50	50	50	50	50	50	50
	22.06944444							
Nb poissons marqués	4	5	7	2	6	7	3	5
Nb poissons piochés marqués	2	1	3	1	2	6	2	3
Nb n	8	25	16.33333333	4	18	8.166666667	4.5	8.333333333
Vrai nb n	50	50	50	50	50	50	50	50
	11.54166667							
Nb poissons marqués	5	5	5	5	5	5	5	
Nb poissons piochés marqués	5	2	2	5	1	5	2	
Nb n	5	12.5	12.5	5	25	5	12.5	
Vrai nb n	50	50	50	50	50	50	50	

On fixait le nombre de poissons marqués M à 7, la population recapturée n à 7 poissons et la population totale du lac N à 50 poissons.

Nous utilisions alors un produit en croix : $\frac{M}{N} = \frac{m}{n}$.

En premier lieu, nous avions été surprises d'obtenir 49 comme résultat. En effet, 49 est très proche de 50.

Néanmoins, après mûre réflexion, nous nous sommes rendus compte qu'il s'agissait seulement du fruit du hasard. Effectivement, $\frac{7}{N_{simulation}} = \frac{1}{7}$ donne $N_{simulation} = 7 \times \frac{7}{1} = 7^2 = 49$.

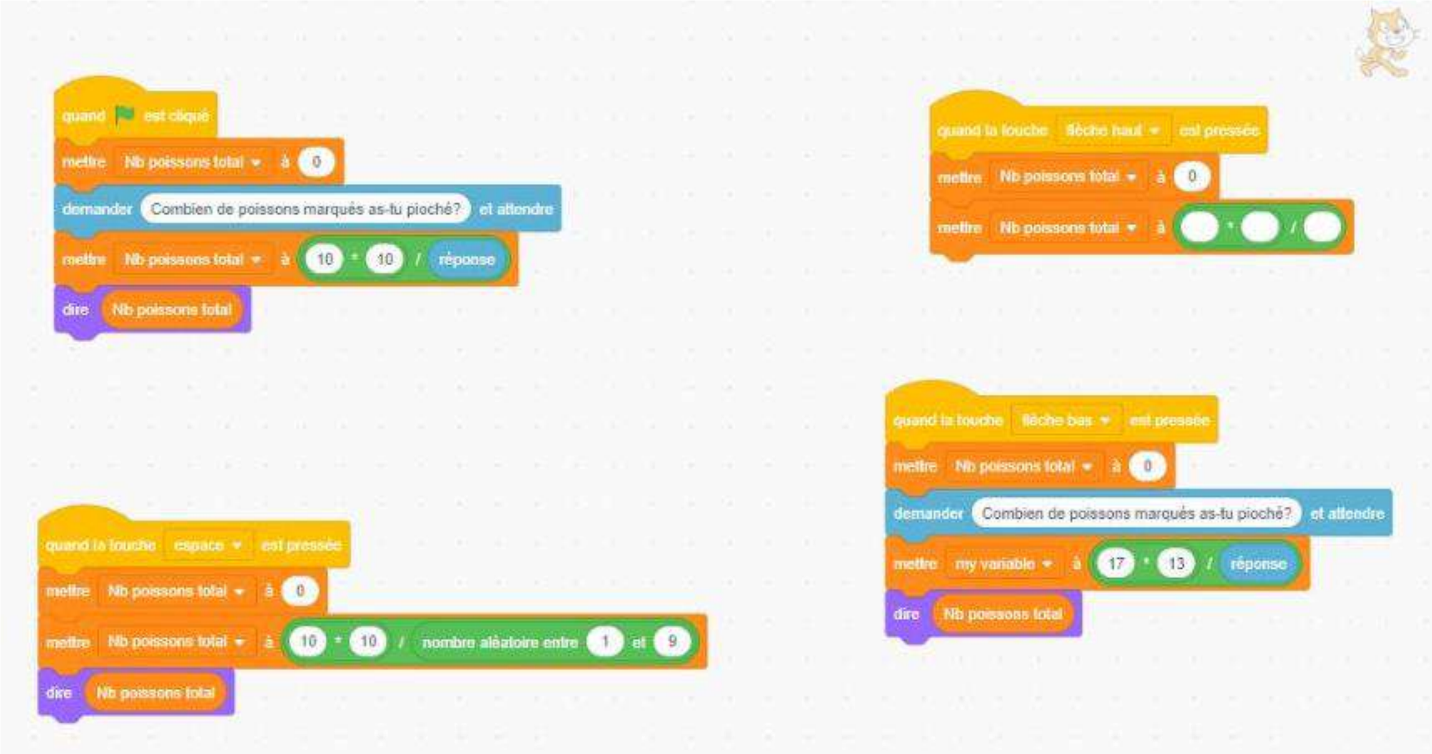
Nous avons fait de nombreux essais où on ne connaissait pas N, le nombre total de poissons.

Mise en situation					
Nb poissons marqués	3	3	10	5	5
Nb poissons piochés marqués	2	2	1	1	2
Nb n				20	40
Vrai nb n	?	?	?	24	24

En conclusion, nous avons fait plusieurs conjectures :

- Plus on en marque, moins on en pioche, plus le résultat sera précis. Proposition envisageable.
- Moins on en marque, plus on en pioche, moins le résultat sera précis. Proposition envisageable.
- Plus on en pioche, plus le résultat sera précis. Proposition fausse.

C. Programmation avec Scratch



Nous avons utilisé la programmation Scratch pour faciliter nos calculs.

Nous avons utilisé la même formule qu'avant : $\frac{M}{N} = \frac{m}{n}$.

Nous avons vite abandonné Scratch pour continuer avec Excel.

D. Analyse du lien entre m et N

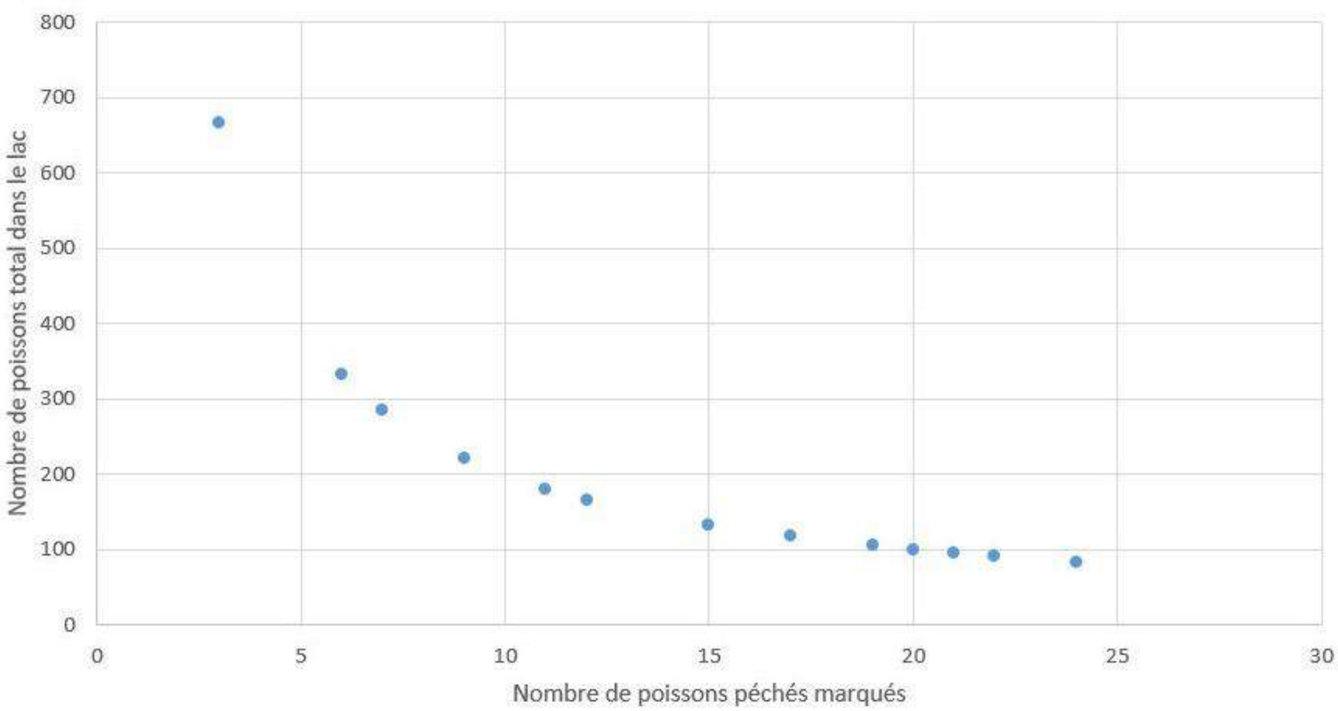
Nous avons ensuite décidé d'analyser le lien entre la population recapturée marquée m et la population totale du lac N.

Pour se faire, nous avons élaboré un tableau Excel où nous fixions la population marquée M à 25 et la population recapturée n à 80.

En utilisant la formule $\frac{M}{N} = \frac{m}{n}$, on a complété un tableur :

n	80	80	80	80	80	80	80	80	80	80	80	80	80		
résulat	100	666,67	333,34	285,7	222,2	181,2	166,67	133,3	117,6	105,2	95,3	90,9	83,3		
m	20	3	6	7	9	11	12	15	17	19	21	22	24		avec n = 80
résultat	100	666,67	333,34	285,7	222,2	181,2	166,67	133,3	117,6	105,2	95,3	90,9	83,3		
résulat	100	666,67	333,34	285,7	222,2	181,2	166,67	133,3	117,6	105,2	95,3	90,9	83,3		

Nous avons ensuite transformé ce tableau en graphique :



On y voit une allure décroissante et on remarque que plus la population recapturée marquée m est élevée, moins il y aura de poissons dans le lac.

En effet, si nous marquons des poissons et que, lors de la repêche, nous retrouvons ces mêmes poissons, alors il n'y a en effet pas beaucoup d'autres poissons dans le lac.

E. Analyse du lien entre n et N

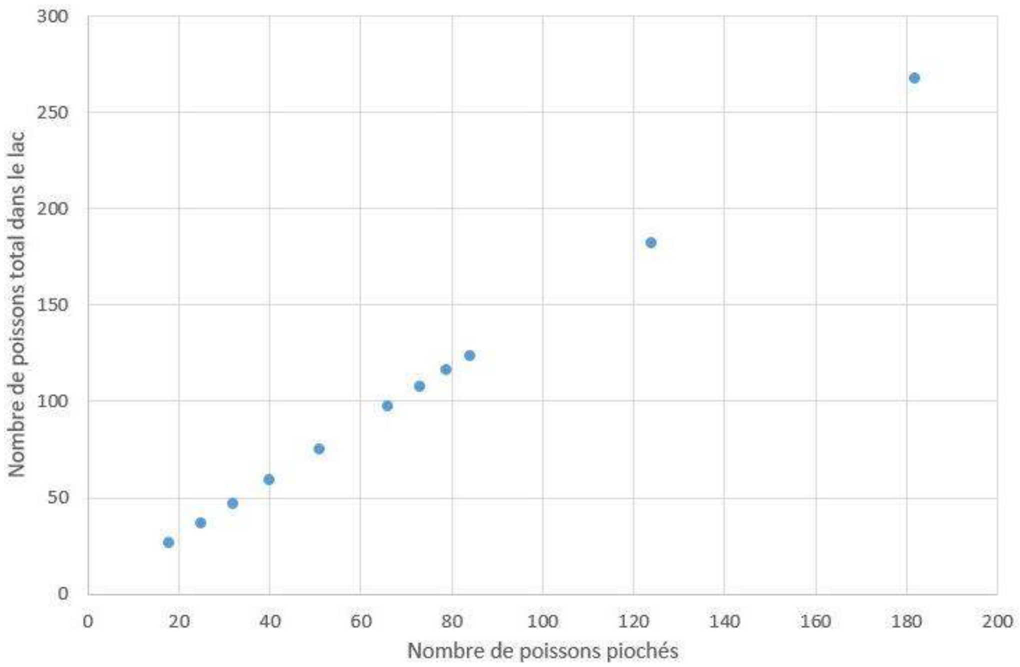
Nous avons ensuite établi le lien entre le nombre de poissons piochés $\boldsymbol{n}$ et le nombre de poissons total dans le lac $\boldsymbol{N}$.

Nous avons fait de nombreuses expériences en fixant la population marquée $\boldsymbol{M}$ à 25 et la population recapturée marquée $\boldsymbol{m}$ à 17.

On cherche donc à savoir combien de poissons $\boldsymbol{n}$ il faut tirer pour avoir 17 poissons marqués ($\boldsymbol{m}$).

n	18	25	32	40	51	66	73	79	84	124	182	
résultat	26,5	36,7	47	58,8	75	97,1	107,4	116,2	123,5	182,4	267,6	
m	17	17	17	17	17	17	17	17	17	17	17	avec m = 17
résultat	26,5	36,7	47	58,8	75	97,1	107,4	116,2	123,5	182,4	267,6	
résultat	26,5	36,7	47	58,8	75	97,1	107,4	116,2	123,5	182,4	267,6	

Nous l'avons ensuite transformé en graphique.



Cela nous a donné l'allure d'une droite passant par l'origine, nous en avons donc déduit que la fonction représentant cette situation est une fonction linéaire.

Le coefficient de proportionnalité de cette droite est d'environ 1,5.

Donc $\boldsymbol{N} = \mathbf{1,5}\ \boldsymbol{n}$

On en conclut que plus la population totale du lac sera élevée $\boldsymbol{N}$, plus il faudra de nombre de poissons recapturés $\boldsymbol{n}$ pour obtenir 17 poissons marqués ($\boldsymbol{m}$).

Le résultat est logique puisque $\frac{25}{\boldsymbol{N}} = \frac{17}{\boldsymbol{n}}$ $\qquad \boldsymbol{N} = \frac{25}{17}\boldsymbol{n} \approx \mathbf{1,47}\ \boldsymbol{n}$

V. Intervalle de confiance et interprétation

Intervalle de confiance au seuil de 95%:

Soit f la fréquence observée d'un caractère dans un échantillon de taille n.

Sous les conditions usuelles d'approximation $n \geq 30, n \times p \geq 5$ et $n \times (1-p) \geq 5,$

l'intervalle $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right]$ est un intervalle de confiance au niveau de confiance 0,95 de la proportion inconnue p dans la population.

Remarque :

Les résultats anormaux sont des résultats en dehors de l'intervalle de confiance

Application de l'intervalle de confiance

On sait que $\frac{M}{N} = \frac{m}{n}$ et la fréquence d'observation de $f = \frac{m}{n}$

Alors $\frac{M}{N} = f$ d'où $N = \frac{M}{f}$

En calculant l'intervalle de confiance on trouve : $[f_{borne\ inf}\ ;\ f_{borne\ sup}\ ]$

Alors $0 < f_{borne\ inf} < f < f_{borne\ sup}$

$$\frac{1}{f_{borne\ inf}} > \frac{1}{f} > \frac{1}{f_{borne\ sup}}$$

$$\frac{M}{f_{borne\ inf}} > \frac{M}{f} > \frac{M}{f_{borne\ sup}}$$

Donc $\frac{M}{f_{borne\ sup}} < N < \frac{M}{f_{borne\ inf}}$

Application de l'intervalle de confiance

- Situation 1 : $M = 50$, $n = 30$ et $m = 15$

On obtient une fréquence d'observation de $f = \frac{m}{n} = \frac{15}{30} = \frac{1}{2}$

L'intervalle de confiance $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right] = \left[\frac{1}{2} - \frac{1}{\sqrt{30}}; \frac{1}{2} + \frac{1}{\sqrt{30}}\right] \approx [0{,}317; 0{,}683]$

Alors $\frac{50}{0{,}683} < N < \frac{50}{0{,}317}$

$73{,}20 < N < 157{,}73$

On peut donc estimer que la population totale de poissons est entre 73 et 158.

- Situation 2 : le même exemple que dans la partie A 1) où l'on avait obtenu $N_{simulation} = 53$ au lieu que notre $N = 70$ ce qui donnait un **écart de 32%**
  On a $= 30$, $n = 30$ et $m = 17$

  On obtient une fréquence d'observation de $f = \frac{m}{n} = \frac{17}{30}$

  L'intervalle de confiance $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right] = \left[\frac{17}{30} - \frac{1}{\sqrt{30}}; \frac{17}{30} + \frac{1}{\sqrt{30}}\right] \approx [0{,}384; 0{,}749]$

  Alors $\frac{30}{0{,}749} < N < \frac{30}{0{,}384}$

  $40{,}1 < N < 78{,}1$

  On peut donc estimer que la population totale de poissons est entre 40 et 78.

  Notre résultat qui nous semblait très éloignés de la réalité est malgré tout dans l'intervalle de confiance au seuil de 95%. Il est donc tout à fait acceptable.

VI. Conclusion

Ce sujet de Math-en-Jeans nous a permis de se rendre compte que cette méthode peut être utilisée pour déterminer le nombre d'individus d'une espèce en voie de disparition, comme le tigre.

Cette méthode est également utilisée pour estimer une population d'insectes spécifiques afin d'éviter un risque de contamination tel que la proportion de moustiques tigres parmi les moustiques. Elle peut également servir pour éviter la surpêche, le thon rouge par exemple.

En conclusion, grâce à cette méthode nous pouvons déterminer un nombre total sans compter chaque individu : d'où le nom de "Compter sans Compter".