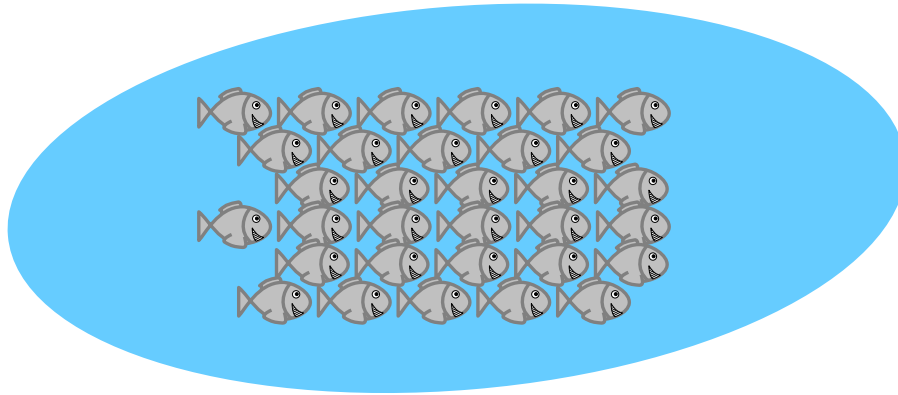


# ◆ Thème 3.1. La biodiversité et son évolution

## I. — Estimation d'une abondance par la méthode CMR

Dans un milieu donné, on considère une population d'animaux d'une certaine espèce. On souhaite déterminer combien d'animaux de cette espèce sont présents dans ce milieu. C'est ce qu'on appelle l'**abondance** de cette espèce dans le milieu considéré.

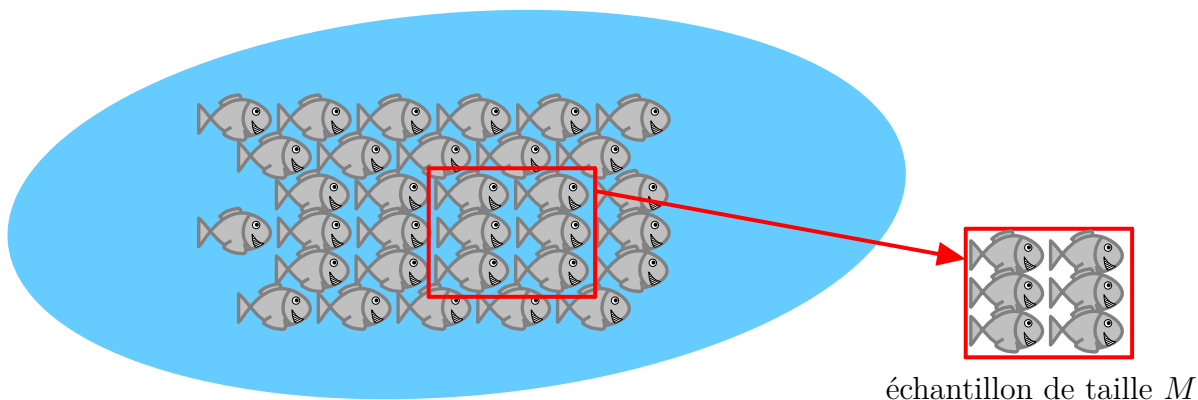
Par exemple, on souhaite estimer l'abondance de truites dans un lac donné c'est-à-dire le nombre  $N$  de truites qui vivent dans ce lac.



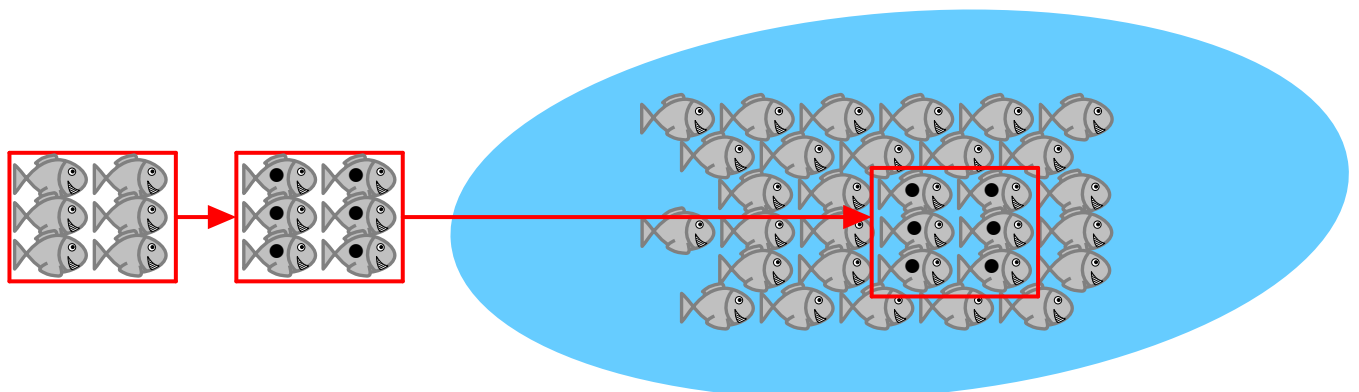
Il est impossible de toutes les compter une à une. On peut utiliser une méthode d'estimation appelée méthode « Capture-Marquage-Recapture » (en abrégé CMR).

Son principe est le suivant.

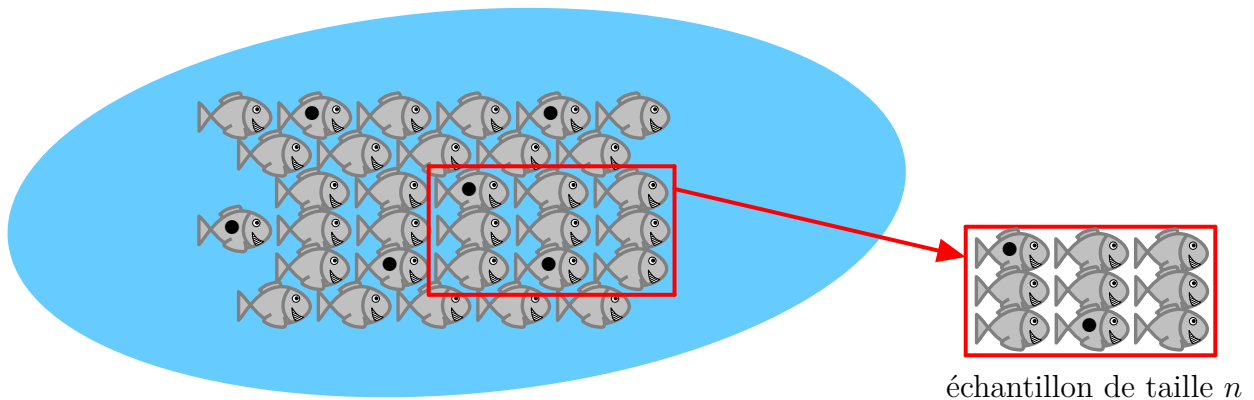
**1ère étape : Capture.** On capture un certain nombre  $M$  de truites. C'est ce qu'on appelle un échantillon et  $M$  est la taille de cet échantillon.



**2ième étape : Marquage.** On marque les truites capturées (à l'aide d'une bague, d'une marque de couleur, d'un transpondeur...) puis on relâche les truites dans le lac.



**3ième étape : Recapture.** Après un certain temps, on effectue une nouvelle capture dans le lac d'un échantillon de taille  $n$  quelconque et on compte le nombre  $m$  de truites marquées dans ce nouvel échantillon.



En faisant l'hypothèse que la proportion de truites marquées est la même dans le nouvel échantillon que dans la population totale, on a l'égalité

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

On en déduit que

$$N = \frac{n \times M}{m}.$$

Imaginons, par exemple, qu'on capture 150 truites pour le premier échantillon et qu'on capture 50 truites pour le second échantillon dont 8 sont marquées. On peut estimer le nombre total de truites dans le lac à

$$N = \frac{50 \times 150}{8} \approx 938.$$

### Conditions d'application et limites de la méthode

Pour que la méthode soit efficace, il est nécessaire que :

1. la population étudiée n'évolue pas (ou peu) entre les deux captures donc il ne faut pas que des individus puissent la quitter ou y entrer, par exemple à l'occasion de flux migratoires. De même, le temps écoulé entre le marquage et la recapture doit être assez court pour éviter les naissances et les décès mais suffisamment important pour assurer un brassage uniforme des individus marqués dans l'ensemble de la population.
2. les animaux marqués ne soient pas affectés par le marquage (que ce soit dans leur comportement ou leur espérance de vie) et les marques ne soient pas perdues. Ainsi, les deux premières étapes (capture et marquage) doivent être aussi courtes que possibles et il faut utiliser soit des marques indélébiles soit un double marquage ;
3. la probabilité de capturer un animal marqué doit être la même que n'importe quel animal de la population.

On peut mettre en évidence certaines limites de la méthode :

1. il n'est pas toujours facile de capturer suffisamment d'individus pour obtenir un résultat satisfaisant et certaines catégories, comme les jeunes, sont souvent plus difficiles à capturer donc sous-représentés dans les échantillons ;
2. le marquage peut affecter le comportement des animaux, les rendre vulnérables face aux prédateurs ou affecter leur place hiérarchique dans leur société ;
3. un animal capturé lors de la constitution du premier échantillon peut se montrer par la suite plus méfiant et avoir une probabilité plus faible d'être recapturé qu'un individu non marqué.

## II. — Fluctuation d'échantillonnage et intervalle de confiance

### 1) Fluctuation d'échantillonnage

Dans la méthode CMR, on fait l'hypothèse que la proportion  $\frac{m}{n}$  d'animaux marqués dans le second échantillon est égale à (ou en tout cas proche de) la proportion  $\frac{M}{N}$  d'animaux marqués dans l'ensemble de la population. La proportion  $\frac{m}{n}$  s'appelle la **fréquence observée** d'animaux marqués dans l'échantillon.

Dans les faits, cette fréquence dépend de l'échantillon. Voici, par exemple, les résultats obtenus lors de 10 recaptures de 50 truites après avoir marquées 150 truites lors de la première capture.

capture	1	2	3	4	5	6	7	8	9	10
Nombre de truites marquées dans l'échantillon	8	10	7	5	6	4	10	4	3	7
Proportion de truites marquées dans l'échantillon	0,16	0,2	0,14	0,1	0,12	0,08	0,2	0,08	0,06	0,14
Estimation de l'abondance	937	750	1071	1500	1250	1875	750	1875	2500	1071

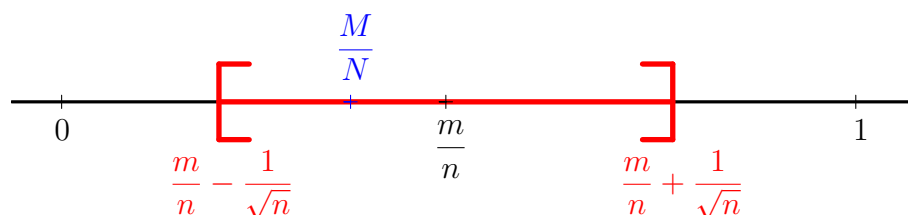
Cette variation est ce qu'on appelle la **fluctuation d'échantillonnage**.

Une question naturelle est alors la suivante : puisque  $\frac{m}{n}$  varie, cette proportion est-elle toujours une bonne approximation de  $\frac{M}{N}$  ?

La réponse est non. On peut très bien, par exemple, avoir un échantillon qui ne contient pas d'animaux marqués, ce qui impliquera que  $\frac{m}{n} = 0$  alors que  $\frac{M}{N} \neq 0$ . Cependant, on peut montrer que, dans la plupart des cas,  $\frac{m}{n}$  est bien une bonne approximation de  $\frac{M}{N}$  et qu'elle est d'autant meilleure que  $n$  est grand.

### 2) Intervalle de confiance

Plus précisément, on peut montrer que, dans 95% des cas, la  **marge d'erreur**  c'est-à-dire l'écart entre la fréquence observée  $\frac{m}{n}$  et la proportion  $\frac{M}{N}$  dans la population totale est inférieure à  $\frac{1}{\sqrt{n}}$ . Autrement dit, pour 95% des échantillons, la proportion  $\frac{M}{N}$  appartient à l'intervalle  $\left[ \frac{m}{n} - \frac{1}{\sqrt{n}} ; \frac{m}{n} + \frac{1}{\sqrt{n}} \right]$ .



On dit que l'intervalle

$$\left[ \frac{m}{n} - \frac{1}{\sqrt{n}} ; \frac{m}{n} + \frac{1}{\sqrt{n}} \right]$$

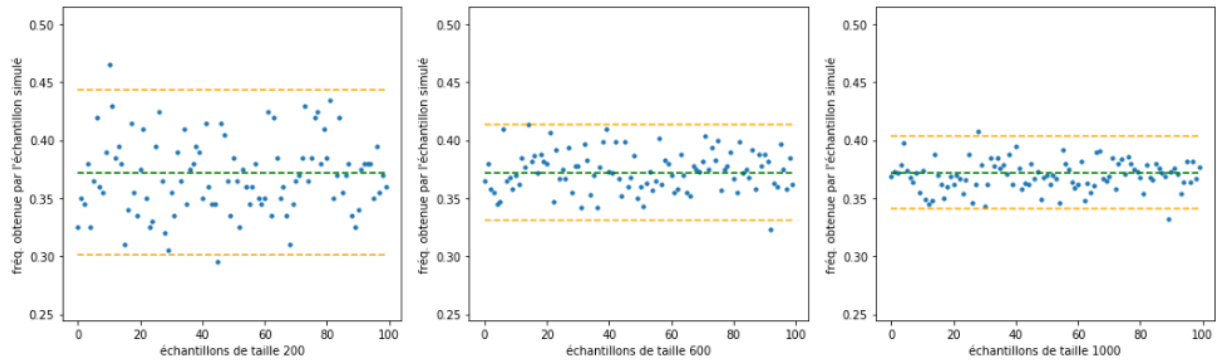
est un **intervalle de confiance** au niveau de confiance 95% de la proportion  $\frac{M}{N}$ .

Si on reprend l'exemple précédent où  $M = 150$ ,  $n = 50$  et  $m = 8$  alors on peut affirmer, avec un niveau de confiance 95% que la proportion  $\frac{M}{N}$  d'animaux marqués dans la population totale est comprise entre  $\frac{8}{50} - \frac{1}{\sqrt{50}} \approx 0,018$  et  $\frac{8}{50} + \frac{1}{\sqrt{50}} \approx 0,302$ . On en déduit que  $N$  est compris entre  $\frac{150}{0,302} \approx 497$  et  $\frac{150}{0,018} \approx 8333$ .

On voit qu'on obtient une estimation très imprécise. Cela vient du fait que la taille de l'échantillon ( $n = 50$ ) est trop petite.

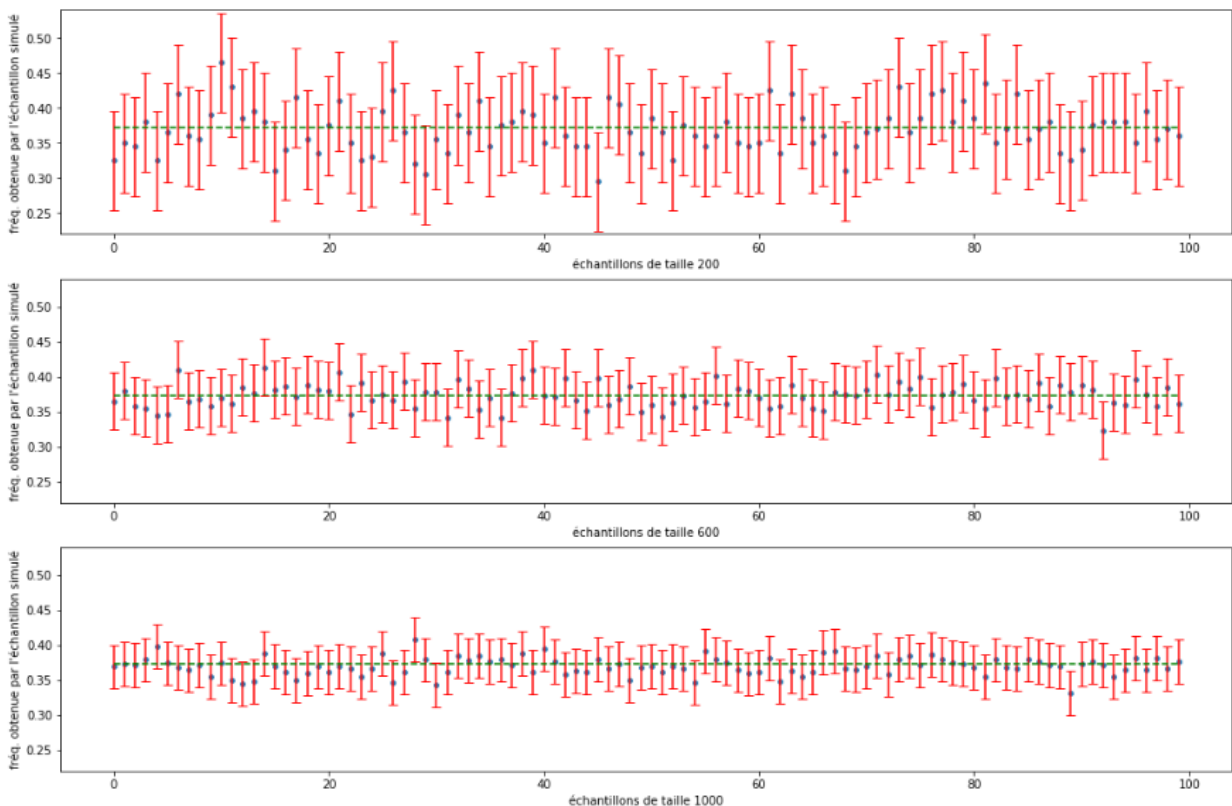
L'amplitude de l'intervalle de confiance, c'est-à-dire l'écart entre les deux bornes de l'intervalle est  $\frac{2}{\sqrt{n}}$ . Cette amplitude diminue lorsque  $n$  augmente. Autrement dit, plus on prend un échantillon de taille  $n$  grande, plus on aura un encadrement précis. Remarquons cependant qu'augmenter la valeur de  $n$  ne change pas le niveau de confiance qui reste égal à 95%.

Sur les graphiques ci-dessous, on a représenté des simulations de 100 recaptures différentes pour des échantillons de tailles  $n = 200$ ,  $n = 600$  et  $n = 1000$ . Les pointilles centraux représentent la proportion réelle de  $\frac{M}{N}$  et les pointillés supérieurs et inférieurs délimitent un écart de plus ou moins  $\frac{1}{\sqrt{n}}$  par rapport à  $\frac{M}{N}$ .



On constate que plus  $n$  augmente, plus les valeurs obtenues se resserrent autour de  $\frac{M}{N}$  mais qu'il y a toujours des valeurs qui s'écartent de plus de  $\frac{1}{\sqrt{n}}$ .

Sur les graphiques ci-dessous, on a représenté les mêmes situations mais en traçant les intervalles de confiance autour de chaque proportion.



On constate que plus  $n$  augmente plus la taille des intervalles diminue mais qu'il y a toujours des intervalles qui ne contiennent pas la proportion  $\frac{M}{N}$  (symbolisée par la ligne en pointillés).

Dans 95%, la proportion  $\frac{M}{N}$  appartient à l'intervalle  $\left[ \frac{m}{n} - \frac{1}{\sqrt{n}} ; \frac{m}{n} + \frac{1}{\sqrt{n}} \right]$  donc la marge d'erreur est inférieur à  $\frac{1}{\sqrt{n}}$ . Ainsi, pour obtenir une approximation avec une marge d'erreur inférieur à une valeur  $E$ , il suffit de choisir  $n$  de telle sorte que  $\frac{1}{\sqrt{n}} \leq E$  c'est-à-dire  $n \geq \frac{1}{E^2}$ .

Par exemple, pour avoir une estimation avec une marge d'erreur inférieure ou égale à 1%, il faut prendre un échantillon de taille au moins  $\frac{1}{0,01^2} = 10000$ .

### III. — Modèle de Hardy-Weinberg

#### 1) Fréquences génotypiques et fréquences alléliques

Dans une population, on considère un caractère déterminé par un gène présent sur deux chromosomes homologues  $c_1$  et  $c_2$ . On suppose que ce gène peut prendre deux formes différentes appelées **allèles** : l'allèle  $A$  et l'allèle  $a$ .

Le **génotype** d'un individu est la composition allélique c'est-à-dire la donnée des deux allèles présents sur ses gènes. Dans notre cas, on a les génotypes ci-contre.

$c_2 \backslash c_1$	$A$	$a$
$A$	$AA$	$Aa$
$a$	$aA$	$aa$

Ainsi, il y a 3 génotypes différents : 2 allèles  $A$  ( $AA$ ), 1 allèle  $A$  et 1 allèle  $a$  ( $Aa$  ou  $aA$ ) et 2 allèles  $a$  ( $aa$ ). Les individus ayant un génotype  $AA$  ou  $aa$  sont dit homozygotes et les autres, ayant un génotype  $Aa$ , sont appelés hétérozygotes.

Dans une population donnée, on peut définir deux types de fréquences :

**les fréquences génotypiques** : la fréquence d'un génotype est la proportion d'individus possédant ce génotype ;

**les fréquences alléliques** : la fréquence d'un allèle est la proportion de gènes portant cet allèle.

Par exemple, supposons que dans une population de 1200 individus, il y en ait 350 qui possèdent le génotype  $AA$ , 560 le génotype  $Aa$  et 290 le génotype  $aa$ .

Alors les fréquences génotypiques sont :

$$f(AA) = \frac{350}{1200} = \frac{7}{24} \quad f(Aa) = \frac{560}{1200} = \frac{7}{15} \quad f(aa) = \frac{290}{1200} = \frac{29}{120}.$$

Pour ce qui est des fréquences alléliques, on remarque que chaque individu possède deux gènes donc le nombre total de gènes est  $2 \times 1200 = 2400$ . Ensuite, chaque individu du génotype  $AA$  possède deux allèles  $A$  et chaque individu du génotype  $Aa$  possèdent un allèle  $A$  donc la fréquence de l'allèle  $A$  est

$$f(A) = \frac{2 \times 350 + 560}{2400} = \frac{21}{40}.$$

De même, chaque individu du génotype  $aa$  possède deux allèles  $a$  et chaque individu du génotype  $Aa$  possèdent un allèle  $a$  donc la fréquence de l'allèle  $a$  est

$$f(a) = \frac{2 \times 290 + 560}{2400} = \frac{19}{40}.$$

Remarquons que, comme il n'y a que 3 génotypes possibles,  $f(AA) + f(Aa) + f(aa) = 1$  et, comme il n'y a que deux allèles possibles,  $f(A) + f(a) = 1$ .

Considérons une population de taille  $N$ . Notons  $p$  la fréquence de  $AA$ ,  $q$  la fréquence de  $Aa$  et  $r$  la fréquence de  $aa$ .

Alors, le nombre d'individus ayant un génotype  $AA$  est  $p \times N$ , le nombre d'individus ayant un génotype  $Aa$  est  $q \times N$  et le nombre d'individus ayant un génotype  $aa$  est  $r \times N$ .

Comme chaque individu possède deux gènes, le nombre total de gènes est  $2N$ . Chaque individu ayant le génotype  $AA$  apporte 2 allèles  $A$  et chaque individu ayant un génotype  $Aa$  apporte 1 allèle  $A$  donc le nombre total d'allèles  $A$  dans la population est  $2 \times p \times N + 1 \times q \times N$ . La fréquence allélique de  $A$  est donc

$$f(A) = \frac{2 \times p \times N + q \times N}{2 \times N} = \frac{(2p + q)N}{2N} = \frac{2p + q}{2} = \frac{2p}{2} + \frac{q}{2} = p + \frac{1}{2}q.$$

De la même manière, la fréquence allélique de  $a$  est

$$f(a) = \frac{2 \times r \times N + q \times N}{2 \times N} = \frac{(2r + q)N}{2N} = \frac{2r + q}{2} = \frac{2r}{2} + \frac{q}{2} = r + \frac{1}{2}q.$$

## 2) Évolution sur plusieurs générations

On s'intéresse maintenant à l'évolution des fréquences génotypiques sur plusieurs générations.

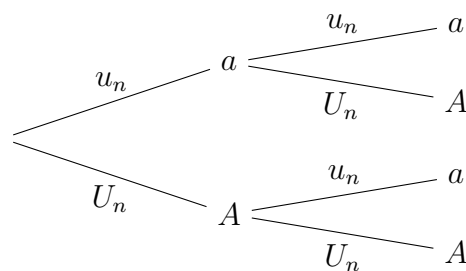
On numérote les générations en considérant que la génération initiale est la génération 0. On note, pour tout entier  $n$  :

•  $p_n$  la fréquence de  $AA$       •  $q_n$  la fréquence de  $Aa$       •  $r_n$  la fréquence de  $aa$

•  $U_n$  la fréquence de  $A$       •  $u_n$  la fréquence de  $a$

à la génération  $n$ . Ainsi, par exemple,  $q_0$  est la fréquence du génotype  $Aa$  à la génération initiale et  $U_3$  est la fréquence de l'allèle  $A$  à la génération 3.

Lors d'une reproduction sexuée, chaque parent va transmettre au descendant un de ses deux allèles. En supposant que la population est suffisamment grande, que les couples se forment au hasard et que la transmission des allèles se fait également au hasard, on peut modéliser la transmission des allèles d'une génération à l'autre comme le choix de deux allèles dans l'ensemble des allèles de la population. Autrement dit, on peut calculer les fréquences génotypiques de la génération  $n + 1$  à l'aide des fréquences alléliques de la génération  $n$  :



On en déduit que  $p_{n+1} = U_n^2$ ,  $q_{n+1} = u_n U_n + U_n u_n = 2u_n U_n$  et  $r_{n+1} = u_n^2$ . Or, d'après ce qui précède,  $U_n = p_n + \frac{1}{2}q_n$  et  $u_n = r_n + \frac{1}{2}q_n$  donc on a les relations suivantes entre les fréquences génotypiques de deux générations successives :

$$p_{n+1} = \left(p_n + \frac{1}{2}q_n\right)^2 \quad q_{n+1} = 2\left(p_n + \frac{1}{2}q_n\right)\left(r_n + \frac{1}{2}q_n\right) \quad r_{n+1} = \left(r_n + \frac{1}{2}q_n\right)^2.$$

Lorsqu'on étudie numériquement l'évolution de ces fréquences, on s'aperçoit que les fréquences génotypiques sont constantes à partir de la génération 1 et que les fréquences alléliques sont constantes dès la génération 0 et ceci quelles que soient les fréquences initiales.

	A	B	C	D	E	F
1	n	p_n	q_n	r_n	U_n	u_n
2	0	0,4	0,1	0,5	0,45	0,55
3	1	0,2025	0,495	0,3025	0,45	0,55
4	2	0,2025	0,495	0,3025	0,45	0,55
5	3	0,2025	0,495	0,3025	0,45	0,55
6	4	0,2025	0,495	0,3025	0,45	0,55
7	5	0,2025	0,495	0,3025	0,45	0,55
8	6	0,2025	0,495	0,3025	0,45	0,55
9	7	0,2025	0,495	0,3025	0,45	0,55

En effet, on a, pour tout entier naturel  $n$ ,

$$U_{n+1} = p_{n+1} + \frac{1}{2}q_{n+1} = U_n^2 + \frac{1}{2} \times 2u_n U_n = U_n^2 + u_n U_n = U_n(U_n + u_n) = U_n$$

car  $U_n + u_n = 1$ . Dès lors,  $u_{n+1} = 1 - U_{n+1} = 1 - U_n = u_n$ .

Ainsi, les fréquences alléliques sont constantes dès la génération initiale.

De plus, pour tout entier naturel  $n$ ,  $p_{n+1} = U_n^2$  donc  $p_n$  est constante à partir de  $n = 1$  et, de même pour  $q_n$  et  $r_n$ .

Ainsi, les fréquences atteignent un état d'équilibre appelé **équilibre de Hardy-Weinberg**. Ce phénomène a été découvert en 1908 de manière indépendante par le mathématicien anglais Godfrey Hardy et le médecin allemand Wilhelm Weinberg.

On a vu précédemment qu'il est facile d'exprimer les fréquences alléliques en fonction de fréquences génotypiques. Dans le cadre du modèle d'Hardy-Weinberg, il est également simple d'exprimer les fréquences génotypiques à l'aide des fréquences alléliques (au moins à partir de la deuxième génération) : on a en effet, dans ce cadre,

$$f(AA) = f(A)^2 \quad f(Aa) = 2f(A)f(a) \quad f(aa) = f(a)^2.$$

### 3) Écart par rapport au modèle

La validité du modèle de Hardy-Weinberg repose sur plusieurs hypothèses.

- La taille de la population doit être très grande.
- Le caractère étudié doit être lié à des cellules diploïdes c'est-à-dire des cellules contenant des paires de chromosomes homologues.
- La population étudiée doit être panmictique (c'est-à-dire dotée d'un système de reproduction sexuée où les fécondations se font au hasard). En particulier, les couples doivent se former au hasard (panmixie) et la transmission des allèles doit également se faire au hasard (pangamie).
- Il ne doit y avoir dans la population ni sélection, ni mutation, ni migration.
- Il ne doit pas y avoir de reproduction intergénérationnelle.
- Les génotypes sont tous viables et féconds.

Lorsqu'on constate des écarts entre les fréquences observées et l'équilibre prévu par le modèle d'Hardy-Weinberg, cela signifie qu'une ou plusieurs des conditions précédentes ne sont pas remplies. Cela est dû principalement à des **forces évolutives** qui vont faire évoluer les fréquences génotypiques. Les principales forces évolutives sont :

- **la dérive génétique** : de petits écarts initiaux dans la transmission peuvent entraîner une grande disparité dans la répartition des allèles après plusieurs générations voire la disparition pure et simple d'une allèle ; cette dérive a d'autant plus de risques de se produire que la taille population initiale est faible ;
- **la sélection naturelle** : les individus qui sont les plus aptes à survivre peuvent influencer les proportions d'allèles transmis à la génération suivante ;
- **la mutation génétique** : des mutations aléatoires d'allèles peuvent s'opérer et se transmettre à la génération suivante, modifiant ainsi les proportions de chaque allèle.

Par exemple, les îles Galapagos abritent treize espèces différentes de pinsons qui se différencient par la taille de leur corps, ainsi que par la forme et la taille de leur bec. Ces trois caractères, et notamment la taille du bec, se transmettent fortement d'une génération à la suivante.

Deux biologistes, Peter et Rosemary Grant, ont suivi l'évolution sur trente ans des populations de pinsons sur l'île de Daphne Major et ont pu détecter sur cette période des événements sélectifs importants.

Par exemple, la fin des années 1970 a été marquée par une sécheresse importante, entraînant une raréfaction des graines de petite taille et qui a coïncidé avec une sélection très forte des individus à gros bec chez l'espèce *G. Fortis*. Chez cette espèce, la consommation de petites graines est préférée quand celles-ci sont abondantes, alors qu'en cas de sécheresse la consommation de graines plus grosses, accessibles uniquement aux individus à gros bec, est favorisée. L'épisode de sécheresse, en causant la raréfaction des graines de petite taille, a entraîné une mortalité plus importante des individus à petit bec alors que les individus à gros bec ont survécu en plus grand nombre. Ces derniers se sont donc en moyenne plus reproduits ce qui a entraîné un déplacement de caractère à la génération suivante. On a donc ici un exemple de l'effet de la sélection naturelle sur la répartition des génotypes.

## La méthode Capture-Marquage-Recapture (CMR)

**Exercice 1.** Les oposums sont des marsupiaux. Les individus de l'espèce *Trichosurus cunninghami* vivent dans les forêts australiennes et se nourrissent principalement de fruits et de feuilles. Entre février et mars 2009, des feux ont ravagé un grand nombre de forêts dans l'état de Victoria, ce qui a eu un fort impact sur les populations de *Trichosurus cunninghami* (voir le tableau ci-dessous).

Année	2004	2005	2006	2007	2008	2009
Nombre d'individus	3214	3215	3237	2984	3314	1209

En 2019, des chercheurs de l'université de Camberra en Australie ont procédé à une campagne CMR dans les forêts brûlées en 2019. Ils ont capturé et marqué 82 individus qu'ils ont ensuite relâché. Deux semaines plus tard, ils ont capturé 67 individus dont 2 étaient marqués.

Peut-on considérer que la population de *Trichosurus cunninghami* a retrouvé un niveau équivalent à celui d'avant 2009 ?

**Exercice 2.** En 1998, une équipe de chercheurs a estimé le nombre de jeunes otaries à fourrure nées dans une population en Australie. Lors de la première capture, 1291 jeunes otaries ont été marquées en coupant une mèche de fourrure. Les jeunes de cette colonie ont ensuite été recapturés visuellement plusieurs fois, ce qui a permis d'estimer leur nombre. Les résultats obtenus lors des 4 recaptures sont rassemblés dans le tableau suivant.

Numéro de la recapture	1	2	3	4
Taille de l'échantillon	1080	1224	1107	1233
Nombre d'individus marqués	391	378	363	357

1. Quel est ici le procédé de marquage ? Donner deux avantages de celui-ci.
2. Pour chacune des 4 recaptures, estimer l'abondance de jeunes otaries nées en 1998.
3. Calculer l'abondance moyenne à partir des 4 résultats de la question 2.
4. Quel est l'intérêt de réaliser plusieurs recaptures ?

**Exercice 3.** La cistude d'Europe (*Emys orbicularis*) est une espèce de tortue vivant notamment dans le sud de la France. Les marais de l'étang de l'Or abritent la plus grande population de cistude d'Europe dans le département de l'Hérault. Un suivi des effectifs de la population a été réalisé par la mise en œuvre sur 4 ans (2015-2018) d'une campagne de CMR. En voici les résultats où N.A.C.M. signifie « nombre d'animaux capturés et marqués », N.A.M.R. signifie « nombre d'animaux marqués capturés lors de la recapture » et N.A.C.R signifie « nombres d'animaux capturés lors de la recapture ».

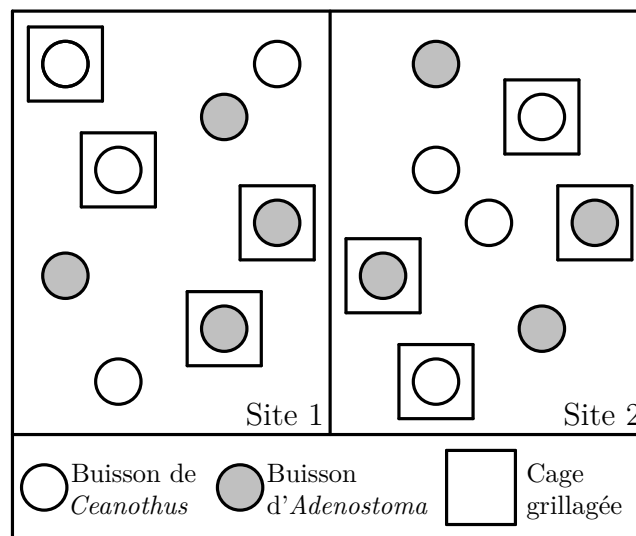
Marais	Candillargues					du Grès St Nazaire					Tartugien Lansargus				
	2014	2015	2016	2017	2018	2014	2015	2016	2017	2018	2014	2015	2016	2017	2018
N.A.C.M.	84	32	28	×	×	84	66	64	48	62	×	42	76	69	100
N.A.M.R.	28	3	3			23	13	10	6	12		8	15	14	25
N.A.C.R.	40	6	9			27	19	23	15	21		17	27	28	39

Estimer, pour chaque marais et chaque année pour laquelle cela est possible, le nombre total de cistude d'Europe.

**Exercice 4.** Patrick Nosil, chercheur, a étudié, par CMR, l'effet de la présence de prédateurs et de la sélection naturelle sur les effectifs de deux populations de phasmes de l'espèce *Timema cristinae* : l'une vit préférentiellement sur les buissons d'*Adenostoma fasciculatum*, l'autre sur les buissons de *Ceanothus spinosus*. Les différences entre ces deux populations reposent sur la couleur générale de leur corps (teinte, saturation et luminosité). Les individus vivant sur les *Adenostoma* ont tendance à présenter des rayures plus grandes et plus brillantes, des corps moins brillants et une taille plus courte que les individus vivant sur les *Ceanothus*.

Les chercheurs ont identifié que la répartition de ces deux populations dans un environnement ne se faisait pas au hasard. L'une se positionne plus particulièrement sur des plantes hôtes de couleur foncée (les buissons d'*Adenostoma*) alors que les individus de l'autre population se positionnent plus particulièrement sur des plantes hôtes claires (buisson de *Ceanothus*).

Le principe de la démarche a été de concevoir, sur le terrain, des zones aléatoirement disposées avec quatre niveaux de conditions expérimentales (individus de chacun des deux types et présence ou non de prédateurs). La figure schématique ci-dessous illustre le montage expérimental : sur le site 1 sont déposés les lots d'individus marqués vivant sur les *Adenostoma* et sur le site 2 sont déposés les individus marqués vivant sur les *Ceanothus*. Pour réaliser les zones sans prédateurs, certains buissons ont été entourés de cages grillagées empêchant l'accès par des oiseaux.



Sur chaque buisson, 24 individus marqués ont été relâchés. Chaque lot comprenait autant d'individus mâles que femelles. Les marquages individuels ont été réalisés par un rond avec un marqueur permanent à pointe fine sur l'abdomen, de sorte que la marque ne soit pas visible par le prédateur lorsque l'insecte est dans sa position naturelle sur la plante hôte. Tous les individus ont été marqués et relâchés dans les 5 jours suivant leur première capture. Des relevés de recapture ont été menés 3, 10, 17 et 24 jours après la libération. Pour cette étape de recapture, les membres de l'équipe ont placé, sous chaque buisson testé, un drap blanc puis ont inspecté visuellement le buisson et ont secoué chaque branche de sorte que les insectes tombent du buisson sur le drap blanc. Ils ont ainsi réussi à enregistrer les insectes présents à la date de recapture, puis chaque spécimen a été relâché sur le même site de capture. Une session de recapture a été considérée comme terminée si aucun phasme n'avait été trouvé après 15 minutes de secousse des branches du buisson testé.

Dans cette étude, la « recapture » est assimilée à un indicateur de survie. Ainsi, si un individu n'est pas recapturé, on le considère comme ayant été prélevé par un prédateur.

1. Combien d'animaux ont été marqués au total ?
2. Quelle est la spécificité de l'utilisation de la méthode CMR dans cette étude par rapport à l'usage traditionnel vu en cours ?
3. Relever dans le texte les précautions prises par les scientifiques pour satisfaire les conditions d'application de la méthode CMR.

# La méthode Capture-Marquage-Recapture (CMR)

## Corrigés

**Exercice 1.** On peut estimer le nombre d'animaux en 2019 à

$$N = \frac{82 \times 67}{2} = 2747.$$

Ainsi, la population de *Trichosurus cunninghami* a retrouvé un niveau sensiblement équivalent à celui d'avant 2009.

**Exercice 2.**

1. Ici, les otaries sont marquées en coupant une mèche de fourrure. Cette méthode a plusieurs avantages :
  - Dans un temps assez court, elle est indélébile et, dans un temps long, elle disparaît (avec la repousse des poils) ce qui ne perturbe pas la vie de l'animal sur le long terme ;
  - elle permet une recapture visuelle c'est-à-dire par simple observation des animaux (et donc sans avoir à la recapter physiquement)
2. Pour la capture 1, on peut estimer l'abondance à  $\frac{1291 \times 1080}{391} \approx 3566$  individus, pour la capture 2, à  $\frac{1291 \times 1224}{378} \approx 4180$  individus, pour la capture 3, à  $\frac{1291 \times 1107}{363} \approx 3937$  individus et pour la capture 4, à  $\frac{1291 \times 1233}{357} \approx 4459$  individus
3. L'abondance moyenne (estimée) est environ

$$\frac{3566 + 4180 + 3937 + 4459}{4} \approx 4036.$$

4. On voit que les estimations diffèrent selon les captures en raison de la fluctuation d'échantillonnage (le résultat obtenu varie en fonction de l'échantillon). Réaliser plusieurs recaptures permet de faire une moyenne sur plusieurs estimations ce qui permet de lisser les écarts dus à la fluctuation d'échantillonnage.

**Exercice 3.** Pour chaque colonne, on effectue le calcul  $\frac{(N.A.C.M) \times (N.A.C.R)}{N.A.M.R}$ . On obtient les résultats suivants.

Marais	Candillargues					du Grès St Nazaire					Tartugien Lansargus				
	2014	2015	2016	2017	2018	2014	2015	2016	2017	2018	2014	2015	2016	2017	2018
N.A.C.M.	84	32	28			84	66	64	48	62		42	76	69	100
N.A.M.R.	28	3	3	×	×	23	13	10	6	12	×	8	15	14	25
N.A.C.R.	40	6	9			27	19	23	15	21		17	27	28	39
Abondance	120	64	84	×	×	99	96	147	120	109	×	89	137	138	156

**Exercice 4.**

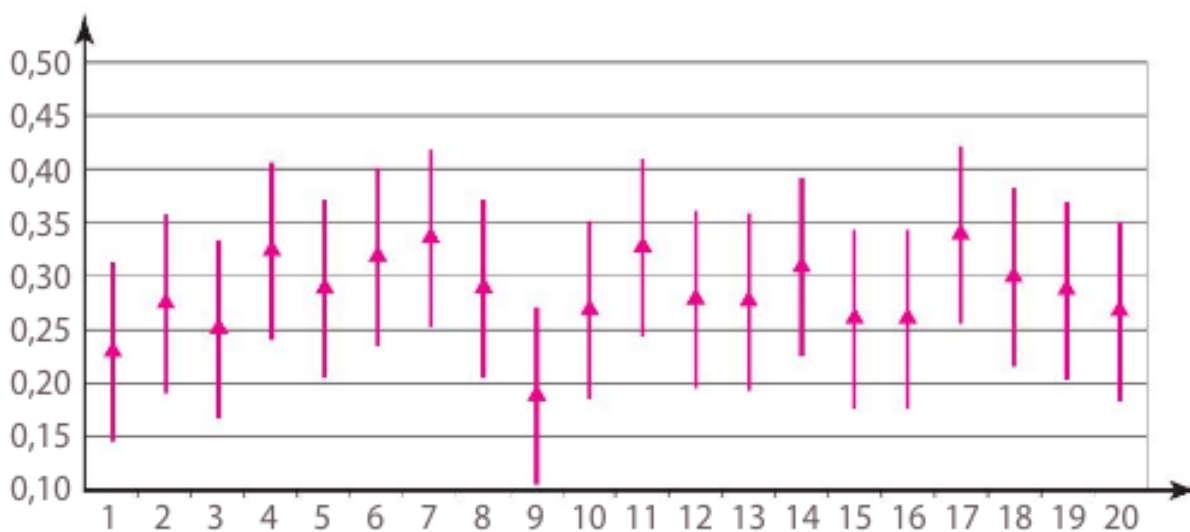
1. Il y a 4 buissons de chaque type dans chacune des deux zones donc  $4 \times 2 \times 2 = 16$  buissons. Sur chaque buisson, on a relâché 24 animaux soit  $16 \times 24 = 384$  animaux au total.
2. Ici, on marque tous les animaux qu'on étudie et on recapture tous les animaux sauf ceux qui sont morts. Le « marquage » alors est le fait de ne peut pas être recapturé.
3. Chaque lot comprend autant d'individus mâles que femelles pour éviter les biais liés au sexe. Le marquage individuel à l'aide d'un marqueur permanent à pointe fine sur l'abdomen permet d'avoir une marque indélébile mais qui n'affecte pas le comportement de l'animal ni son espérance de vie. Le temps entre le marquage et la réintroduction des animaux dans le milieu est court (5 jours) et les recaptures sont faites à des temps suffisamment courts (3, 10, 17 et 24 jours) pour limiter les modifications de la population (naissance, morts naturelles...).

## Estimation par intervalle de confiance

**Exercice 1.** Sur 900 poissons pêchés dans un lac, on a observé que 180 d’entre eux étaient porteurs de parasites.

- Déterminer la proportion de poissons porteurs de parasites parmi les poissons pêchés.
- Déterminer un intervalle de confiance au niveau de confiance 95% de la proportion de poissons porteurs de parasites dans ce lac.

**Exercice 2.** Dans le nord des Alpes, 30% des renards sont infectés par l’échinococcose alvéolaire. À l’aide d’un tableur, on a simulé le prélèvement de 20 échantillons de même taille  $n$  dans la population de renards et on a représenté graphiquement les intervalles de confiance au niveau de confiance 95% associés.



- Que représentent les nombres sur l’axe des abscisses ? sur l’axe des ordonnées ?
- Que représentent les triangles sur les segments verticaux ?
- Quel phénomène explique qu’on obtienne des intervalles de confiance différents ?
- Représenter sur le graphique ci-dessus, à l’aide d’une droite horizontale, la proportion de renards infectés dans l’ensemble de la population.
- Calculer le pourcentage d’intervalles de confiance qui ne contiennent pas cette proportion. Ce résultat était-il prévisible ?
- Estimer la valeur de  $n$ .

**Exercice 3.** On s’intéresse à la proportion d’un caractère dans une population de lapin. Pour ce faire, on prélève une échantillon. Si un individu possède le caractère, on note « 1 » et, sinon, on note « 0 ».

Voici la liste des résultats obtenus :

0 0 1 0 1 0 1 0 0 0 1 1 0 0 1 1 1 0 0 0 1 0 1 1 0 1 0 0 1 0 0 0 1 0 0.

1. **a.** Déterminer la taille  $n$  de l'échantillon et le nombre  $m$  d'animaux possédant le caractère étudié dans l'échantillon. En déduire la fréquence observée du caractère dans cet échantillon.
- b.** Déterminer un intervalle de confiance au niveau de confiance 95% de la proportion de lapin possédant le caractère dans l'ensemble de la population.
2. Dans la même population, on a prélevé un nouvel échantillon et on obtient les résultats suivants :

1 1 0 0 1 0 0 0 0 0 1 1 1 0 1 0 0 1 1 0 0 0 1 1 0 1 0 0 1 1 1 1 0 1 0 1 0 0 0 0 1 1 1 0 0 0 1 0 0 0 0  
 1 1 0 0 1 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 0 1 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0

Reprendre les question **1.a.** et **1.b.** avec ce nouvel échantillon.

3. Les résultats obtenus sur les deux échantillons sont-ils en accord ?

**Exercice 4.** On désire évaluer l'abondance d'une population de manchots empereurs vivant sur une île. Pour cela, on capture 232 individus qui sont marqués puis relâchés. On recapture ultérieurement 400 animaux parmi lesquels on dénombre 116 animaux marqués.

1. Déterminer un intervalle de confiance au niveau de confiance 95% de la proportion  $p$  d'animaux marqués.
2. En déduire un encadrement de l'abondance des manchots empereurs sur l'île.

**Exercice 5.** À l'aide d'un échantillon de taille  $n$ , on estime qu'en France la proportion de personnes du groupe sanguin AB est de 10% avec une marge d'erreur d'au plus 3%.

1. Déterminer un intervalle de confiance au niveau de confiance 95% associé à cet échantillon.
2. Déterminer  $n$ .

**Exercice 6.** La coryza du chat est une maladie virale associée à un syndrome respiratoire qui touche principalement les chatons. Pour estimer la prévalence (c'est-à-dire la proportion de présence) dans un département, on prélève un échantillon de 145 chatons et 25 s'avèrent porteurs de cette maladie.

1. Calculer la fréquence observée de chatons porteurs du coryza dans cet échantillon.
2. Déterminer un intervalle de confiance au niveau de confiance 95% pour la proportion de chatons touchés par la maladie dans le département.
3. L'amplitude de l'intervalle étant trop grande, on souhaite prélever un nouvel échantillon. Quelle doit être sa taille pour que la marge d'erreur soit inférieure ou égale à 4% ?

**Exercice 7.** À la fin des années 1970, une étude anglophone affirmait que les gauchers représentaient entre 8% et 15% de la population mondiale.

Quelle était la taille de l'échantillon sur lequel se basait cette étude ?

# Estimation par intervalle de confiance

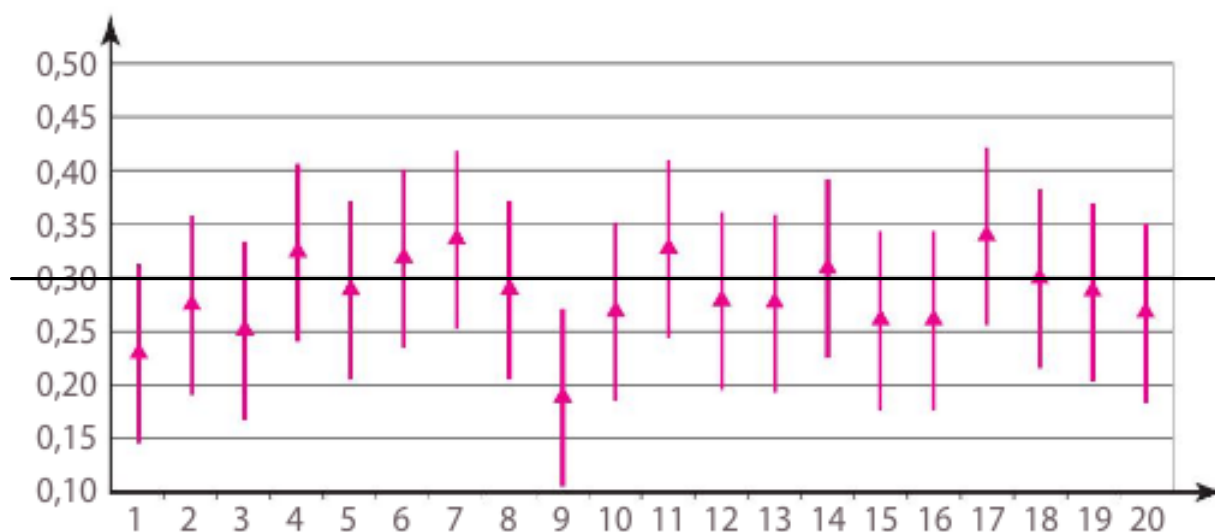
## Corrigés

### Exercice 1.

1. La proportion de poissons porteurs de parasites parmi les poissons pêchés est  $\frac{180}{900} = \frac{1}{5}$ .
2. On a  $\frac{180}{900} - \frac{1}{\sqrt{900}} = \frac{1}{6}$  et  $\frac{180}{900} + \frac{1}{\sqrt{900}} = \frac{7}{30}$  donc un intervalle de confiance au niveau de confiance 95% de la proportion de poissons porteurs de parasites dans ce lac est  $\left[\frac{1}{6}; \frac{7}{30}\right]$  soit environ  $[0,16; 0,24]$ .

### Exercice 2.

1. Les nombres sur l'axe des abscisses représentent les numéros des simulations et les nombres sur l'axe des ordonnées représentent les fréquences (simulées) d'animaux infectés.
2. Les triangles sur les segments verticaux représentent les fréquences observées d'animaux infectés pour chaque simulation. Ces triangles se situent également au milieu de chaque intervalle.
3. Les intervalles de confiance diffèrent en raison de la fluctuation d'échantillonnage.
- 4.



D'après l'énoncé, la proportion de renards infectés est  $30\% = 0,3$  donc on trace une droite horizontale passant en ordonnée par  $0,3$ .

5. Parmi les 20 intervalles, seul 1 ne contient pas la fréquence de  $0,3$ . Ainsi, le pourcentage d'intervalles de confiance qui ne contiennent pas la proportion de renards infectés est  $\frac{1}{20} = 0,05 = 5\%$ . Ceci était prévisible puisque les intervalles représentés sont des intervalles de confiance au niveau de confiance 95%.
6. Grâce au graphique, on peut estimer que l'amplitude des intervalles est environ  $0,17$ . Ainsi, on a  $\frac{2}{\sqrt{n}} \approx 0,17$  donc  $\frac{1}{\sqrt{n}} \approx 0,085$ . Dès lors,  $\sqrt{n} \approx \frac{1}{0,085}$  donc  $n \approx \left(\frac{1}{0,085}\right)^2$  c'est-à-dire  $n \approx 140$ .

### Exercice 3.

- La taille de l'échantillon est  $n = 35$  et le nombre d'animaux possédant le caractère étudié dans l'échantillon est  $m = 14$ . Ainsi, la fréquence observée du caractère dans cet échantillon est  $\frac{14}{35} = \frac{2}{5} = 0,4$ .
  - On a  $\frac{2}{5} - \frac{1}{\sqrt{35}} \approx 0,23$  et  $\frac{2}{5} + \frac{1}{\sqrt{35}} \approx 0,57$  donc un intervalle de confiance au niveau de confiance 95% de la proportion de lapin possédant le caractère dans l'ensemble de la population est  $[0,23; 0,57]$ .
- La taille de l'échantillon est  $n = 100$  et le nombre d'animaux possédant le caractère étudié dans l'échantillon est  $m = 40$ . Ainsi, la fréquence observée du caractère dans cet échantillon est  $\frac{40}{100} = 0,4$ .  
On a  $0,4 - \frac{1}{\sqrt{100}} \approx 0,3$  et  $0,4 + \frac{1}{\sqrt{100}} \approx 0,5$  donc un intervalle de confiance au niveau de confiance 95% de la proportion de lapin possédant le caractère dans l'ensemble de la population est  $[0,3; 0,5]$ .
- Sur les deux échantillons, on trouve les mêmes fréquences observées donc les résultats sont en accord.

### Exercice 4.

- La fréquence observée d'animaux marqués est  $\frac{116}{400} = 0,29$ . De plus,  $0,29 - \frac{1}{\sqrt{400}} = 0,24$  et  $0,29 + \frac{1}{\sqrt{400}} = 0,34$  donc un intervalle de confiance au niveau de confiance 95% de la proportion  $p$  d'animaux marqués est  $[0,24; 0,34]$ .
- Comme  $\frac{232}{0,34} \approx 680$  et  $\frac{232}{0,24} \approx 970$ , on peut estimer l'abondance de manchots empereurs sur l'île entre 680 et 970.

### Exercice 5.

- Comme la fréquence observée est  $10\% = 0,1$  et que la marge d'erreur est d'au plus  $3\% = 0,03$ , un intervalle de confiance au niveau de confiance 95% associé à cet échantillon est  $[0,1 - 0,03; 0,1 + 0,03]$  soit  $[0,07; 0,13]$ .
- On doit avoir  $\frac{1}{\sqrt{n}} = 0,03$  donc  $\sqrt{n} = \frac{1}{0,03}$  c'est-à-dire  $n = \left(\frac{1}{0,03}\right)^2 \approx 1110$ .

### Exercice 6.

- La fréquence observée de chatons porteurs du coryza dans cet échantillon est  $\frac{25}{145} = \frac{5}{29} \approx 0,17$ .
- On a  $\frac{5}{29} - \frac{1}{\sqrt{145}} \approx 0,08$  et  $\frac{5}{29} + \frac{1}{\sqrt{145}} \approx 0,26$  donc un intervalle de confiance au niveau de confiance 95% pour la proportion de chatons touchés par la maladie dans le département est  $[0,08; 0,26]$ .
- Pour avoir une marge d'erreur inférieure ou égale à 4%, il suffit de prendre un échantillon de taille  $n$  telle que  $\frac{1}{\sqrt{n}} \leq 0,04$  c'est-à-dire  $\sqrt{n} \geq \frac{1}{0,04}$  soit  $n \geq \left(\frac{1}{0,04}\right)^2 = 625$ . Ainsi, il suffit de prendre un échantillon de taille au moins égale à 625.

**Exercice 7.** L'intervalle de confiance est ici  $[0,08; 0,15]$ . L'amplitude de cet intervalle est  $0,15 - 0,08 = 0,07$  donc, si on note  $n$  la taille de l'échantillon, on a  $\frac{2}{\sqrt{n}} = 0,07$ . Ainsi,  $\frac{1}{\sqrt{n}} = 0,035$  donc  $\sqrt{n} = \frac{1}{0,035}$  c'est-à-dire  $n = \left(\frac{1}{0,035}\right)^2 \approx 816$ .

Ainsi, l'étude se basait sur un échantillon d'environ 800 personnes.

## Modèle de Hardy-Weinberg

**Exercice 1.** Aux USA, l'étude d'une population de 1000 plants de maïs a donné les fréquences génotypiques suivantes pour un gène donné ayant deux allèles possibles  $B$  ou  $b$  :

Génotype	$BB$	$Bb$	$bb$
fréquence observée	0,35	0,5	0,15

Calculer les fréquences alléliques des allèles  $B$  et  $b$  dans cette population.

**Exercice 2.** La mucoviscidose est une maladie qui frappe 1 enfant sur 2500. L'étude de sa transmission a montré qu'elle est due à l'état homozygote  $aa$  d'un certain gène. Les individus hétérozygotes  $Aa$  sont sains et il est même impossible de détecter chez eux l'allèle pathogène. On suppose que la population est à l'équilibre de Hardy-Weinberg pour ce gène. Calculer :

1. la fréquence de l'allèle pathogène.
2. la proportion de personnes qui portent un allèle  $a$  sans être atteintes par la maladie.

**Exercice 3.** La phénylcétonurie est une maladie génétique rare. Elle est due à un défaut porté par un allèle  $a$  récessif (c'est-à-dire que cet allèle ne s'exprime pas si l'autre chromosome homologue possède le caractère  $A$  dit dominant). La maladie apparaît uniquement à l'état homozygote  $aa$ , mais un sujet non malade hétérozygote (c'est-à-dire ayant pour génotype  $Aa$ ) peut être porteur du gène défectueux. Dans la population, la fréquence de l'allèle  $a$  est  $\frac{1}{100}$ . En supposant que la population est à l'équilibre de Hardy-Weinberg pour ce gène,

1. calculer la proportion de sujets atteints par la phénylcétonurie ;
2. calculer la proportion de porteurs de l'allèle défectueux mais non atteints par la maladie.

**Exercice 4.** Chez les êtres humains, on considère que le gène de la pigmentation de la peau est présent sous forme de deux allèles : l'allèle  $A$  dominant et l'allèle  $a$  responsable de l'albinisme chez les individus homozygotes  $aa$ . Des études ont montré que la fréquence de l'allèle  $a$  dans la population humaine est 0,005%.

1. Calculer la fréquence de l'allèle  $A$ .
2. En supposant que la population est à l'équilibre de Hardy-Weinberg pour ce gène, calculer la fréquence de porteurs sains du caractère d'albinisme.

**Exercice 5.** On cherche à savoir si des populations sont à l'équilibre de Hardy-Weinberg pour un gène codant des molécules à la surface de globules rouges. Il existe deux allèles pour ce gène  $M$  et  $N$ . Une étude menée au Japon en 1958 dans la ville minière de Ashibetsu a donné les résultats suivants :

Génotypes	$MM$	$MN$	$NN$
effectifs	406	744	332

1. Déterminer la taille de la population étudiée.
2. Calculer les fréquences des allèles  $M$  et  $N$  dans la population.
3. En supposant que la population est à l'équilibre d'Hardy-Weinberg, calculer les fréquences génotypiques théoriques à partir des fréquences alléliques obtenus à la question précédente.
4. Comparer les fréquences théoriques obtenues à la question précédente avec les fréquences génotypiques réellement observées. Peut-on valider l'hypothèse selon laquelle la population est en équilibre d'Hardy-Weinberg pour ce gène ?

**Exercice 6.** Le gène de la calpastatine a un effet majeur sur la croissance musculaire des animaux. Cher le mouton, il est situé sur le cinquième chromosome et possède deux allèles  $M$  et  $N$ . L'allèle  $M$  provoque une croissance plus importante des muscles. Certaines populations de moutons ont subi une sélection afin d'obtenir des animaux ayant un poids plus important. En particulier, il a été montré que les individus possédant le génotype  $NN$  avait un poids inférieur en moyenne à ceux des autres génotypes.

Une étude a été menée en Turquie sur deux espèces de moutons : les moutons de Kivircik (KIV) et les moutons Karacabey Merino (KM). Voici les résultats obtenus :

Génotype	$MM$	$MN$	$NN$
KIV	245	79	12
KM	166	65	17

En reprenant la démarche de l'exercice précédent, étudier si ces deux populations sont en équilibres d'Hardy-Weinberg et, si ce n'est pas le cas, expliquer les raisons probables des écarts observés.

**Exercice 7.** Les trois populations états-uniennes de mouche *Drosophila pseudoobscura* sont génétiquement proches pour les différents allèles codant pour la synthèse des enzymes Ptérine-8 et Xanthine déhydrogénase. En revanche, la population colombienne de Bogota est très originale ne possédant même qu'un seul allèle du gène codant pour la xanthine déhydrogénase. On explique ceci par le fait que la population de mouches de Bogota est sans doute issue d'un petit nombre d'individus ayant migré accidentellement d'Amérique du nord, ne véhiculant avec eux qu'une partie des allèles des populations du nord.

Quelle force évolutive est mise en évidence dans le texte précédent ?

**Exercice 8.** La drépanocytose est une maladie génétique résultant de la mutation d'un gène de l'hémoglobine. L'allèle  $A$  code la présence d'une hémoglobine saine HbA tandis que l'allèle  $S$  code la présence d'une hémoglobine mutée HbS. Les individus de génotypes  $AA$  et  $AS$  ne souffrent pas de la maladie. En revanche, les individus de génotypes  $SS$  sont malades.

Il a été prouvé que les individus  $AS$  présentent une résistance accrue au paludisme.

Des relevés statistiques ont donné les résultats suivants en République démocratique du Congo.

Génotype	$AA$	$AS$	$SS$
Nombre d'individus	3182	838	96

tableau 1. Répartition des génotypes dans une population de 4116 personnes adultes

Génotype	$AA$	$AS$	$SS$
Nombre d'individus	270550	71400	8050

tableau 2. Répartition des génotypes dans une population de 350000 nouveaux-nés

1. Calculer les fréquences génotypiques dans la population adulte.
2. Calculer la fréquence de l'allèle  $S$  dans la population adulte.
3. En supposant que la population est à l'équilibre de Hardy-Weinberg, déterminer le nombre théorique d'enfants de chaque génotype dans une population de 350000 nouveaux-nés.
4. Comparer les résultats de la question précédente avec les données du tableau 2. Que peut-on en conclure ?
5. Quelle explication peut-on trouver aux résultats de la question précédente ?

# Modèle de Hardy-Weinberg

## Corrigés

**Exercice 1.** Par propriété, on a

$$f(B) = f(BB) + \frac{1}{2}f(Bb) = 0,35 + \frac{1}{2} \times 0,5 = 0,6$$

et

$$f(b) = f(bb) + \frac{1}{2}f(Bb) = 0,15 + \frac{1}{2}0,5 = 0,4.$$

*Remarque.* On vérifie qu'on a bien  $f(B) + f(b) = 1$ .

**Exercice 2.**

1. D'après l'énoncé,  $f(aa) = \frac{1}{2500}$ . Comme on suppose que la population est à l'équilibre de Hardy-Weinberg,  $f(aa) = f(a)^2$  donc  $f(a) = \sqrt{f(aa)} = \sqrt{\frac{1}{2500}} = \frac{1}{50}$ .
2. On en déduit que  $f(A) = 1 - f(a) = 1 - \frac{1}{50} = \frac{49}{50}$  donc, comme la population est à l'équilibre de Hardy-Weinberg,  $f(Aa) = 2f(a)f(A) = 2 \times \frac{1}{50} \times \frac{49}{50} = \frac{49}{1250}$ . Ainsi, la proportion de personnes qui portent un allèle  $a$  sans être atteintes par la maladie est  $\frac{49}{1250}$ .

**Exercice 3.**

1. Comme la population est à l'équilibre de Hardy-Weinberg,  $f(aa) = f(a)^2 = \left(\frac{1}{100}\right)^2 = \frac{1}{10000}$ . Ainsi, la phénylcétonurie touche 1 personne sur 10000.
2. On a  $f(A) = 1 - f(a) = 1 - \frac{1}{100} = \frac{99}{100}$  donc, Comme la population est à l'équilibre de Hardy-Weinberg,  $f(Aa) = 2f(A)f(a) = 2 \times \frac{99}{100} \times \frac{1}{100} = \frac{99}{5000}$ . Ainsi, la proportion de porteurs de l'allèle défectueux mais non atteints par la maladie est  $\frac{99}{5000}$ .

**Exercice 4.**

1. On a  $f(A) = 1 - f(a) = 1 - \frac{0,005}{100} = 0,99995$ .
2. Comme on suppose que la population est à l'équilibre de Hardy-Weinberg pour ce gène,  $f(Aa) = 2f(A)f(a) = 2 \times 0,99995 \times 0,00005 = 0,000099995$ . Ainsi, la fréquence de porteurs sains du caractère d'albinisme est  $0,000099995 = 0,0099995\%$ .

**Exercice 5.**

1. La taille de la population est  $406 + 744 + 332 = 1482$ .
2. On a donc  $f(MM) = \frac{406}{1482} = \frac{203}{741}$ ,  $f(MN) = \frac{744}{1482} = \frac{124}{247}$  et  $f(NN) = \frac{332}{1482} = \frac{166}{741}$ .  
On en déduit que  $f(M) = f(MM) + \frac{1}{2}f(MN) = \frac{203}{741} + \frac{1}{2} \times \frac{124}{247} = \frac{389}{741}$  et, par suite,  $f(N) = 1 - f(M) = \frac{352}{741}$ .
3. En supposant que le population est à l'équilibre de Hardy-Weinberg, on a

- $f(MM) = f(M)^2 = \left(\frac{389}{741}\right)^2 = \frac{151321}{549081} \approx 0,276,$
- $f(MN) = 2f(M)f(N) = 2 \times \frac{389}{741} \times \frac{352}{741} = \frac{273856}{549081} \approx 0,499,$
- $f(NN) = f(N)^2 = \left(\frac{352}{741}\right)^2 = \frac{123904}{549081} \approx 0,226.$

4. Les fréquences observées sont  $f(MM) = \frac{203}{741} \approx 0,274$ ,  $f(MN) = \frac{124}{247} \approx 0,502$  et  $f(NN) = \frac{166}{741} \approx 0,224.$

Ainsi, les valeurs observées sont très proches des valeurs théoriques calculées à la question précédente donc on peut valider l'hypothèse d'une population à l'équilibre de Hardy-Weinberg.

**Exercice 6.** Commençons par la population de moutons Kivircik.

La taille de la population est  $245 + 79 + 12 = 336$ . On a donc  $f(MM) = \frac{245}{336} = \frac{35}{48}$ ,  $f(MN) = \frac{79}{336}$  et  $f(NN) = \frac{12}{336} = \frac{1}{28}$ .

On en déduit que  $f(M) = f(MM) + \frac{1}{2}f(MN) = \frac{35}{48} + \frac{1}{2} \times \frac{79}{336} = \frac{569}{672}$  et, par suite,  $f(N) = 1 - f(M) = \frac{103}{672}$ .

En supposant que la population est à l'équilibre d'Hardy-Weinberg, on a

- $f(MM) = f(M)^2 = \left(\frac{569}{672}\right)^2 = \frac{323761}{451584} \approx 0,717,$
- $f(MN) = 2f(M)f(N) = 2 \times \frac{569}{672} \times \frac{103}{672} = \frac{58607}{225792} \approx 0,260,$
- $f(NN) = f(N)^2 = \left(\frac{103}{672}\right)^2 = \frac{10609}{451584} \approx 0,023.$

Les fréquences observées sont  $f(MM) = \frac{35}{48} \approx 0,729$ ,  $f(MN) = \frac{79}{336} \approx 0,235$  et  $f(NN) = \frac{1}{28} \approx 0,036.$

Ainsi, les valeurs observées sont relativement proches des valeurs théoriques calculées précédemment donc on peut valider l'hypothèse d'une population à l'équilibre de Hardy-Weinberg.

Continuons avec la population de moutons Karacabey Merino.

La taille de la population est  $166 + 65 + 17 = 248$ . On a donc  $f(MM) = \frac{166}{248} = \frac{83}{124}$ ,  $f(MN) = \frac{65}{248}$  et  $f(NN) = \frac{17}{248}$ .

On en déduit que  $f(M) = f(MM) + \frac{1}{2}f(MN) = \frac{83}{124} + \frac{1}{2} \times \frac{65}{248} = \frac{397}{496}$  et, par suite,  $f(N) = 1 - f(M) = \frac{99}{496}$ .

En supposant que la population est à l'équilibre d'Hardy-Weinberg, on a

- $f(MM) = f(M)^2 = \left(\frac{397}{496}\right)^2 = \frac{157609}{246016} \approx 0,641,$
- $f(MN) = 2f(M)f(N) = 2 \times \frac{397}{496} \times \frac{99}{496} = \frac{39303}{123008} \approx 0,320,$
- $f(NN) = f(N)^2 = \left(\frac{99}{496}\right)^2 = \frac{9801}{246016} \approx 0,040.$

Les fréquences observées sont  $f(MM) = \frac{83}{124} \approx 0,670$ ,  $f(MN) = \frac{65}{248} \approx 0,262$  et  $f(NN) = \frac{17}{248} \approx 0,069$ .

On voit qu'ici les valeurs sont beaucoup plus éloignées (surtout pour les génotypes  $MN$  et  $NN$  donc on ne peut pas valider l'hypothèse d'une population à l'équilibre de Hardy-Weinberg. L'écart constaté s'explique probablement par les sélections opérées afin d'obtenir des animaux ayant un poids plus important.

**Exercice 7.** La force évolutive mise en évidence ici est la dérive génétique : la population de mouche ayant migré n'avait pas les mêmes fréquences génotypiques que la population initiale est cette différence a engendré une dérivé vers un autre équilibre que celui de la population initiale.

**Exercice 8.**

1. Les fréquences génotypiques dans la population adulte sont  $f(AA) = \frac{3182}{4116} = \frac{1591}{2058}$ ,  
 $f(AS) = \frac{838}{4116} = \frac{419}{2058}$  et  $f(SS) = \frac{96}{4116} = \frac{8}{343}$ .

2. La fréquence de l'allèle  $S$  dans la population adulte est  $f(S) = f(SS) + \frac{1}{2}f(AS) = \frac{8}{343} + \frac{1}{2} \times \frac{419}{2058} = \frac{515}{4116}$ .

3. On déduit de la question précédente que  $f(A) = 1 - f(S) = \frac{3601}{4116}$ . En supposant que la population est à l'équilibre de Hardy-Weinberg, la fréquence de chaque génotype à la génération suivante est  $f(AA) = f(A)^2 = \left(\frac{3601}{4116}\right)^2 = \frac{12967201}{16941456}$ ,  $f(AS) = 2f(A)f(S) = 2 \times \frac{3601}{4116} \times \frac{515}{4116} = \frac{1854515}{8470728}$  et  $f(SS) = f(S)^2 = \left(\frac{515}{4116}\right)^2 = \frac{265225}{16941456}$ .

On en déduit que le nombre théorique d'enfants de chaque génotype dans une population de 350000 nouveaux-nés est :

- pour le génotype  $AA$ ,  $\frac{12967201}{16941456} \times 350000 \approx 267900$ ;
- pour le génotype  $AS$ ,  $\frac{1854515}{8470728} \times 350000 \approx 76600$ ;
- pour le génotype  $SS$ ,  $\frac{265225}{16941456} \times 350000 \approx 5500$ ;

4. On constate des différences importantes entre les génotypes théoriques et ceux observés. On en déduit que l'hypothèse d'une population à l'équilibre de Hardy-Weinberg n'est pas correcte.

5. Comme les individus ayant un génotype  $AS$  ont une résistance plus importante au paludisme, la sélection naturelle explique l'écart entre la réalité et le modèle de Hardy-Weinberg. En effet, les individus ayant ce génotype vont davantage survivre au paludisme que les autres et donc l'allèle  $S$  va se retrouver plus représentée dans la population que ce qu'elle le serait à l'équilibre de Hardy-Weinberg. En effet, dans la population adulte, la fréquence de l'allèle  $S$  est  $f_a(S) = \frac{545}{4116} \approx 0,12512$  alors que, dans la population de nouveaux-nés, elle est égale à  $f_{nn}(S) = \frac{8050}{350000} + \frac{1}{2} \times \frac{71400}{350000} = \frac{1}{8} = 0,125$ . On a donc  $f_{nn}(S) < f_a(S)$  ce qui confirme que l'allèle  $S$  est plus fréquente dans la population adulte que dans la population de nouveaux-nés et ainsi la résistance au paludisme modifie bien la répartition allélique pour la génération adulte.