

Extraits de la feuille de TD n° 1
Modélisation probabiliste

3 La proportion normale de polynucléaires basophiles est de 0,5 pour 100 leucocytes. Quelle est la probabilité de n'en trouver aucun parmi 200 leucocytes observés dans ces conditions ?

Comprendre les données :

Quand on prélève un échantillon de N leucocytes, il y en a « normalement » $0,05 \times N$ (environ) qui sont polynucléaires basophiles.

Si on prend 200 leucocytes, il y en a « normalement » 1 qui est polynucléaire basophile. Mais il peut y en avoir 0, il peut y en avoir 1, 2, 3... ou 200 : le nombre de polynucléaires basophiles dans un échantillon de 200 leucocytes est une variable *aléatoire*, on ne peut donc pas connaître sa valeur à l'avance. La seule chose qu'on peut faire c'est estimer cette valeur et éventuellement estimer la probabilité qu'elle soit dans une certaine fourchette. Pour cela, on a besoin de construire une modélisation probabiliste.

Solution :

On considère ici la fonction X , qui à tout échantillon ω de 200 leucocytes associe le nombre $X(\omega)$ de polynucléaires basophiles dans cet échantillon. Pour estimer $P(X = 0)$ on doit déterminer *a priori* la loi de X .

L'expérience qui permet de mesurer $X(\omega)$ consiste, en principe, à effectuer un test sur chacun des leucocytes pour déterminer s'il est polynucléaire basophile ou pas.

- On effectue 200 tests (un sur chaque leucocyte).
- Chaque test peut fournir deux résultats : vrai ou faux. D'après l'énoncé, la probabilité qu'un test soit positif est de 0,005.
- Les tests sont indépendants.

En vertu d'un certain théorème du cours, il s'en suit que X suit une loi binomiale $\mathcal{B}(200; 0,005)$, on a donc :

$$P(X = 0) = C_{200}^0 \times 0,005^0 \times 0,995^{200} \approx 0,367$$

Soit 36,7 % de chances d'avoir 0 polynucléaires basophiles dans un échantillon de 200 leucocytes.

5 Les centres de transfusion sanguine diffusent le tableau suivant, donnant la répartition en France des principaux groupes sanguins.

	O	A	B	AB
Rhésus +	37,0 %	38,1 %	6,2 %	2,8 %
Rhésus –	7,0 %	7,2 %	1,2 %	0,5 %

- 1) Quelle est la probabilité qu'une personne ait un sang de facteur Rhésus – ?
- 2) Dix personnes, prises au hasard en France, donnent leur sang. Soit X la variable aléatoire qui prend pour valeurs le nombre de personnes appartenant au groupe A. Calculer $P(X = 4)$.

Comprendre l'énoncé :

La question 1) ne pose pas trop de problème. Un traitement élémentaire est possible, nous ne donnons ci-après un traitement plus élaboré que par souci de précision.

Pour la question 2) on lit sur le tableau, par un raisonnement analogue à celui de la question 1), que la probabilité qu'une personne en France soit de groupe A est 45,3 %. Cela ne signifie pas que sur 10 personnes prises au hasard il y en a 4,53 de groupe A ! Ni même 4 ou 5.

Le groupe sanguin d'une personne ne peut pas être connu avant d'être mesuré. On peut seulement estimer la probabilité qu'il soit tel ou tel. Sur 10 personnes, il se peut parfaitement qu'on en ait 0 de groupe A, ou 1, ou 2... ou 10.

Solution :

1) Notons X la fonction qui à un individu ω dans la population française associe son groupe sanguin avec Rhésus. Cette fonction peut prendre 8 valeurs, qui sont O+, O-, A+, A-, etc. Un individu ω est de facteur Rhésus négatif si $X(\omega)$ vaut O-, A-, B- ou AB-. Par la propriété d'additivité des mesures de probabilités on a donc :

$$P(X \in \{O-, A-, B-, AB-\}) = P(X = O-) + P(X = A-) + P(X = B-) + P(X = AB-)$$

La probabilité d'avoir un Rhésus négatif est donc $7 + 7,2 + 1,2 + 0,5 = 15,9\%$.

2) Par le même raisonnement qu'en 1), la probabilité d'être de groupe sanguin A est de $38,1 + 7,2 = 45,3\%$. On note maintenant G la fonction qui à un échantillon ω de 10 personnes associe le nombre $G(\omega)$ de personnes de groupe A dans cet échantillon. La mesure de G consiste à tester le groupe sanguins des 10 personnes de l'échantillon.

- On répète 10 fois le même test.
- Chaque test donne vrai (groupe A) ou faux (pas groupe A). La probabilité d'un test positif est 0,453.
- Les tests sont indépendants.

La variable aléatoire G suit donc une loi $\mathcal{B}(10; 0,453)$. On en déduit :

$$P(X = 4) = C_{10}^4 \times 0,453^4 \times (1 - 0,453)^6 \approx 0,237$$

Soit 23,7 % de chances (seulement !) d'avoir quatre personnes de groupe A dans un échantillon de 10 personnes.

9 Dans une population de veaux, la masse d'un animal pris au hasard est une variable aléatoire X qui suit une loi normale d'espérance mathématique 500 kg et d'écart-type 40 kg. On prélève un échantillon de 80 veaux.

- 1) Estimer le nombre de veaux de plus de 560 kg ?

Comprendre l'énoncé :

Si X suit une loi normale, alors $P(X < 0)$ n'est pas nulle. Pourtant le poids d'une veau ne peut assurément pas être négatif ! On comprends donc qu'ici X n'est pas exactement

la variable aléatoire « poids des veaux » mais un modèle probabiliste qui, contrairement aux exercices précédents, nous est donné au départ.

Noter que l'énoncé ne demande pas « combien de veaux pèsent plus de 560 kg ? ». Une telle formulation du problème, bien que courante dans les manuels et dans la pratique, constitue un abus de langage.

On n'a en effet aucun moyen de *connaître* à l'avance le nombre de veaux dont le poids va se situer dans une fourchette donnée. On peut seulement utiliser notre modèle pour *estimer ce nombre*, mais rien ne garantit que dans la réalité on trouvera effectivement le nombre de veaux prédit par cette estimation.

Solution :

On peut s'attendre, sous l'hypothèse que nous avons à faire à un échantillon de veaux représentatif de la population générale, à ce que la proportion de veaux de plus de 560 kg soit environ égale à $P(X \geq 560)$, autrement dit à ce que le nombre N de veaux de plus de 560 kg dans un échantillon de 80 veaux soit :

$$N \approx P(X \geq 560) \times 80$$

Par hypothèse $X \sim \mathcal{N}(500; 40)$ donc (par théorème) $Y = (X - 500)/40 \sim \mathcal{N}(0, 1)$. Or :

$$X \geq 560 \iff \frac{X - 500}{40} \geq \frac{560 - 500}{40} \iff Y \geq 1,5$$

D'où $P(X \geq 560) = P(Y \geq 1,5) = 1 - P(Y \leq 1,5)$. Sur la table 1 on lit que $P(Y \leq 1,5) \approx 0,933$. On en déduit :

$$N \approx P(X \geq 560) \times 80 \approx 0,067 \times 80 = 5,36$$

Soit une estimation de 5 veaux de plus de 560 kg dans un échantillon *représentatif* de 80 veaux.