Le problème des anniversaires

① Le problème des anniversaires est facile à énoncer : **étant donné un groupe de *n* personnes, quelle est la probabilité p_n qu'au moins deux d'entre elles soient nées le même jour[1] ?** La solution est simple : on calcule la probabilité q_n que les n personnes soient toutes nées un jour différent.

Par exemple, si $n=2$, il y a $365\times365=(365)^2$ paires de jours de naissance possibles, et il y a 365×364 paires de jours de naissance différents possibles. Donc $q_2=\dfrac{364}{365}$ et $p_2=1-q_2=\dfrac{1}{365}\cong 0{,}3\%$.

La formule générale est $q_n=\dfrac{365\times364\times363\times...\times(365-n+1)}{(365)^n}$. Ainsi, pour $n=10$, on obtient $q_{10}=\dfrac{365\times364\times363\times...\times356}{(365)^{10}}\cong 88\%$, donc $p_{10}=1-q_{10}\cong 12\%$; ce n'est pas énorme, mais pas négligeable non plus. À combien estimez-vous p_{20} ? (réponse en bas de la page)[2]

On constate que $p_{22}\cong 47{,}6\%$ et $p_{23}\cong 50{,}8\%$. Donc, **dans un groupe d'au moins 23 personnes, il y a plus d'une chance sur deux qu'au moins deux d'entre elles soient nées le même jour**. Et $p_{64}\approx 99{,}7\%$: il aurait donc été très surprenant qu'en PCSI+PC, tous les élèves soient nés un jour différent.

② Attention ! Il existe un autre problème des anniversaires : **dans un groupe de *n* personnes dont fait partie Alfred, quelle est la probabilité r_n qu'une autre personne soit née le même jour qu'Alfred ?**

Ici $r_n=1-\dfrac{(364)^{n-1}}{(365)^{n-1}}$. On a $r_{23}\cong 6\%$ et, pour que r_n dépasse 50%, **le groupe doit comporter au moins 253 personnes** : $r_{252}\cong 49{,}9\%$ et $r_{253}\cong 50{,}005\%$.

③ La différence importante entre les nombres 23 et 253 s'explique mieux si on remarque que, dans ce groupe de n personnes dont fait partie Alfred, il y a au total $\dfrac{n(n-1)}{2}$ paires de personnes, mai seulement $n-1$ paires de personnes contenant Alfred.

④ La distinction entre ces deux problèmes des anniversaires est cruciale pour justifier l'utilisation judiciaire des tests ADN. Ces tests consistent à enregistrer des séquences ADN présentes dans certains emplacements –appelés locus– de certains chromosomes. Pour chaque

---

[1] mais pas forcément la même année.

[2] $p_{20}\approx 41\%$. C'est peut-être plus que ce que vous pensiez ?

locus, la même séquence se retrouve chez 5 à 20% de la population, et les locus ont été choisis en s'assurant que leurs caractéristiques soient mutuellement indépendantes.

En France le fichier national des empreintes génétiques (FNAEG) repose sur l'examen de 13 locus et on estime à 1 sur 3.10^{12}, soit 1 sur 3 000 milliards[3], la probabilité que les profils génétiques de deux personnes différentes soient identiques. Comme le FNAEG fiche environ 3 millions de personnes, la probabilité qu'un profil génétique établi à partir d'un prélèvement sur une scène de crime soit identique à un profil contenu dans le FNAEG est de 3 millions sur 3 000 milliards, soit 1 sur 1 million.

Par contre la probabilité qu'il existe des paires de profils identiques dans le FNAEG n'est pas négligeable. En effet, le FNAEG contient $\dfrac{3.10^6 \times (3.10^6 - 1)}{2} \cong 9.10^{12}$, soit $9\,000$ milliards de paires de profils. On peut donc s'attendre à y trouver environ $\dfrac{9\,000}{3\,000} = 3$ paires de profils identiques.

Bien entendu, tous ces calculs de probabilités ne tiennent pas compte des nombreux problèmes pratiques et erreurs humaines qui peuvent survenir dans la constitution ou dans l'utilisation du FNAEG : confusion d'échantillons, mauvaises lectures de résultats, pollution d'un échantillon pendant le prélèvement, échantillons de mauvaise qualité ne permettant pas d'examiner les 13 locus, etc.

Sources :

Leila SCHNEPS & Coralie COLMEZ. *Les maths au tribunal.* Éditions du Seuil (2015).

https://fr.wikipedia.org/wiki/Fichier_national_automatis%C3%A9_des_empreintes_g%C3%A9n%C3%A9tiques

https://fr.wikipedia.org/wiki/Empreinte_g%C3%A9n%C3%A9tique

[3] Remarque : $1/3.10^{-12} \approx (11\%)^{13}$. Donc tout se passe comme si chaque locus était identique dans 11% de la population.