Probabilités conditionnelles
Problème de recherche : le détecteur de tricheurs

L'éditeur d'un jeu en ligne compétitif utilise un logiciel anti-triche. Des études ont établi que :

- 1 % des joueurs trichent ;
- quand un joueur triche, le logiciel le signale dans 99 % des cas ;
- quand un joueur ne triche pas, le logiciel le signale quand même par erreur dans 2 % des cas.

L'éditeur envisage de bannir automatiquement tous les joueurs signalés.

La mission

1. Un joueur vient d'être signalé. Quelle est la probabilité qu'il soit vraiment un tricheur ?
2. Le bannissement automatique est-il une bonne idée ? Proposer une procédure plus juste et l'évaluer avec des calculs.

Pour aller plus loin

Dans une ligue de très haut niveau, 20 % des joueurs trichent. Le logiciel est-il plus fiable dans ce contexte ?

Coups de pouce

Bloqué ? Des indices progressifs sont disponibles en ligne. N'en ouvrir qu'un à la fois, et seulement après avoir cherché !

Éléments de correction

1. Probabilité qu'un joueur signalé triche

On note T : « le joueur triche » et S : « le joueur est signalé ». D'après l'énoncé : $P(T) = 0{,}01$, $P_T(S) = 0{,}99$, $P_{\overline{T}}(S) = 0{,}02$.
D'après la formule des probabilités totales (arbre pondéré) :

$$P(S) = 0{,}01 \times 0{,}99 + 0{,}99 \times 0{,}02 = 0{,}0099 + 0{,}0198 = 0{,}0297.$$

$$P_S(T) = \frac{P(T \cap S)}{P(S)} = \frac{0{,}0099}{0{,}0297} = \frac{1}{3} \approx 0{,}33.$$

Seulement un joueur signalé sur trois est un tricheur!
Explication avec des effectifs : sur 10 000 joueurs, 100 trichent (99 signalés) et 9 900 sont honnêtes (198 signalés à tort). Il y a deux fois plus d'innocents que de tricheurs parmi les signalés, car les honnêtes sont beaucoup plus nombreux.

2. Une procédure plus juste

Bannir automatiquement punirait deux innocents pour un tricheur : ce n'est pas acceptable. On peut, par exemple, faire passer un **second test indépendant** aux joueurs signalés. Pour un joueur signalé deux fois, la probabilité de tricher devient, avec $P(T) = \frac{1}{3}$ cette fois :

$$\frac{\frac{1}{3} \times 0{,}99}{\frac{1}{3} \times 0{,}99 + \frac{2}{3} \times 0{,}02} = \frac{0{,}33}{0{,}3433} \approx 0{,}961.$$

Après deux signalements, il y a environ 96 % de chances que le joueur triche. On peut aussi faire vérifier les cas par un modérateur humain.

Pour aller plus loin

Avec $P(T) = 0{,}2$: $P_S(T) = \frac{0{,}2 \times 0{,}99}{0{,}2 \times 0{,}99 + 0{,}8 \times 0{,}02} = \frac{0{,}198}{0{,}214} \approx 0{,}925$. Le logiciel n'a pas changé, mais le résultat est beaucoup plus fiable : la valeur d'un signalement dépend de la proportion de tricheurs.

Ce qu'on a utilisé du chapitre

— Arbre pondéré ; formule des probabilités totales.
— Probabilité conditionnelle $P_S(T)$: « inverser » un conditionnement.

Pour aller plus loin (enrichissement)

Le calcul de $P_S(T)$ à partir de $P_T(S)$ s'appelle la *formule de Bayes*. Le même raisonnement s'applique aux tests médicaux de dépistage, aux filtres anti-spam ou aux portiques de sécurité.