Pas de malchance

1 Un peu de géométrie

▷ **Question 1** *Une feuille de TD contient 4 coquilles. À chaque relecture, une coquille non corrigée est corrigée avec probabilité 1/3. Les relectures sont indépendantes les unes des autres. Combien de relectures faut-il faire pour que la probabilité qu'il ne subsiste aucune coquille soit supérieure à 0,9 ?*
Réponse: P[La coquille n° i est corrigée en au plus n relectures]
$= 1 - P$[Elle n'est pas corrigée après n relectures] $= 1 - (\frac{2}{3})^n$

Donc P[Les 4 coquilles sont corrigées en au plus n relectures] $= \left(1 - (\frac{2}{3})^n\right)^4$

Enfin, puisque n est entier, $\left(1 - (\frac{2}{3})^n\right)^4 \geqslant 0{,}9 \iff n \geqslant 10$ (presque 9, mais rigoureusement c'est 10).

2 Moi d'abord

▷ **Question 2** *Alan et Beth lancent une pièce à tour de rôle, le premier qui fait face a gagné. Alan commence, quelle est la probabilité qu'il gagne ?*
Réponse: Alan gagne si la pièce tombe sur face lors d'un lancé de numéro impair (le premier lancé est numéroté 1).

$$\begin{cases} P[\text{Alan gagne}] = \sum_{i=0}^{\infty}(\frac{1}{2})^{2i+1} \\ P[\text{Beth gagne}] = \sum_{i=0}^{\infty}(\frac{1}{2})^{2i+2} = \frac{1}{2}P[\text{Alan gagne}] \end{cases}$$

Donc P[Alan gagne] $= \frac{2}{3}$.

▷ **Question 3** *Peut-on truquer la pièce pour rendre le jeu équitable ?*
Réponse: Posons $p := P[\text{face}]$.

$$\begin{cases} P[\text{Alan gagne}] = \sum_{i=0}^{\infty}(1-p)^{2i} \cdot p \\ P[\text{Beth gagne}] = \sum_{i=0}^{\infty}(1-p)^{2i+1} \cdot p = (1-p) \cdot P[\text{Alan gagne}] \end{cases}$$

Donc P[Alan gagne] $= P$[Beth gagne] $\iff p = 0$. Le seul moyen de truquer la pièce est de faire en sorte que personne ne gagne.

▷ **Question 4** *Cette fois ils lancent deux dés. Alan gagne s'il obtient 7, Beth gagne si elle obtient 6. Elle s'est déjà fait avoir une fois et exige de commencer. Est-ce un jeu équitable ?*
Réponse: Posons $\begin{cases} P_A = P[\text{somme des dés} = 7] = \frac{6}{36} \\ P_B = P[\text{somme des dés} = 6] = \frac{5}{36} \end{cases}$.

$P[\text{Beth gagne}] = \sum_{i\geqslant 0}(1-P_B)^i(1-P_A)^i \cdot P_B$

$P[\text{Alan gagne}] = (1-P_B)\sum_{i\geqslant 0}(1-P_A)^i(1-P_B)^i \cdot P_A = (1-P_B)\frac{P_A}{P_B}P[\text{Beth gagne}] = \frac{31}{30}P[\text{Beth gagne}]$.
Alan est avantagé.

▷ **Question 5** *L'une de ces parties risque t-elle de se prolonger indéfiniment ?*
Réponse: Si la pièce est truquée et ne tombe jamais sur face, le jeu est infini.

Sinon, soit $0 < q \leqslant \min\left(P[\text{un lancer d'Alan est gagant}], P[\text{un lancer de Beth est gagant}]\right)$. La probabilité que le jeu dure au moins n étapes est majorée par $(1-q)^n$, qui tend vers 0 en $n \to \infty$.

3 Estimateurs

Soient $X_1, ..., X_n$ des variables aléatoires réelles indépendantes, de même espérance m et de même variance $\sigma^2 > 0$.

▷ **Question 6** *On pose $\overline{X_n} := \frac{1}{n}\Sigma_{i=1}^n X_i$. Calculer l'espérance et la variance de $\overline{X_n}$. Pourquoi appelle-t-on $\overline{X_n}$ la moyenne empirique ?*
Réponse: $E[\overline{X_n}] = \frac{1}{n}\sum_{i=1}^n E[X_i] = m$.

$\mathrm{Var}[\overline{X_n}] = \mathrm{Var}[\frac{1}{n}\sum X_i] = \frac{1}{n^2}\mathrm{Var}[\sum X_i] = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$.

On réalise une expérience avec n tirages, $\overline{X_n}$ est une estimation de m, mais c'est une variable aléatoire dont la valeur dépend de l'expérience. « Moyenne empirique » signifie que c'est la moyenne obtenue selon le résultat de l'expérience, il peut y avoir une erreur par rapport à l'espérance m.

▷ **Question 7** *On pose $S_n^2 = \frac{1}{n-1}\Sigma_{i=1}^n (X_i - \overline{X_n})^2$. Calculer l'espérance de S_n^2. Quel est l'intérêt de cette variable ?*
Réponse: (Toutes les sommes sont pour $1 \leqslant i \leqslant n$.)

$$\begin{aligned}
E(S_n^2) &= E\left[\tfrac{1}{n-1}\sum (X_i - \overline{X_n})^2\right] \\
&= \tfrac{1}{n-1}E\left[\sum {X_i}^2 - 2\sum X_i\overline{X_n} + \sum \overline{X_n}^2\right] \\
&= \tfrac{1}{n-1}E\left[\sum {X_i}^2 - 2n\overline{X_n}^2 + n\overline{X_n}^2\right] \\
&= \tfrac{1}{n-1}\sum E\left[{X_i}^2\right] - \tfrac{n}{n-1}E\left[\overline{X_n}^2\right] \\
&= \tfrac{n}{n-1}E\left[{X_1}^2\right] - \tfrac{1}{n(n-1)}\left(E\left[\sum {X_i}^2\right] + E\left[\sum_{i\neq j} X_iX_j\right]\right) \\
&= \tfrac{n}{n-1}E\left[{X_1}^2\right] - \tfrac{1}{n(n-1)}\left(nE\left[{X_1}^2\right] + n(n-1)E\left[X_1\right]^2\right) \\
&= E\left[{X_1}^2\right] - E\left[X_1\right]^2 = \sigma^2
\end{aligned}$$

S_n^2 est donc un moyen d'estimer σ^2.

▷ **Question 8** *Quelle est la différence avec la variance empirique $V_n := \frac{1}{n}\Sigma_{i=1}^n (X_i - \overline{X_n})^2$?*
Réponse: V_n est la variance observée sur les n tirages. Si $X_1, ..., X_n$ n'est qu'un échantillon de la population, V_n a un *biais* par rapport à la variance de la variable aléatoire X_1 et n'est donc pas une bonne manière d'estimer σ^2.

▷ **Question 9** *Exprimer S_n^2 et V_n à l'aide de $\sum_{i=1}^n X_i$, $\sum_{i=1}^n {X_i}^2$ et n. Quel est l'intérêt ?*
Réponse: $S_n^2 = \frac{1}{n-1}\left(\sum {X_i}^2 - \frac{1}{n}(\sum X_i)^2\right)$. $V_n = \frac{n-1}{n}S_n^2$.

L'intérêt est de limiter la mémoire nécessaire (par exemple dans une calculatrice). Après avoir effectué k mesures, on ne retient que $\sum_{i=1}^k X_i$, $\sum_{i=1}^k {X_i}^2$ et k. On peut mettre à jour ces variables au fur et à mesure que l'expérience avance.

On suppose de plus que pour tout i, X_i est une variable aléatoire uniformément distribuée dans $[0;\theta]$.

▷ **Question 10** *Estimer θ à l'aide de $\overline{X_n}$. Quelle est la variance de cet estimateur ?*
Réponse: La densité de X est $\frac{1}{\theta}$. $E(\overline{X_n}) = E(X) = \int_0^\theta \frac{x}{\theta}\mathrm{d}x = \frac{\theta}{2}$. Un estimateur de θ est donc $2\overline{X_n}$. Noter que $E(2\overline{X_n})$ n'est pas donné par l'expérience, on ne peut donc pas dire que c'est un estimateur.

$\mathrm{Var}(2\overline{X_n}) = 4\frac{1}{n}\mathrm{Var}(x) = \frac{4}{n}(\int_0^\theta \frac{x^2}{\theta}\mathrm{d}x - E(X)^2) = \frac{4}{n}(\frac{\theta^2}{3} - \frac{\theta^2}{4}) = \frac{\theta^2}{3n}$.

▷ **Question 11** *On pose $M_n := \max(X_1, ..., X_n)$. Estimer θ à l'aide de M_n. Indice : quelle est l'espérance de M_n ?*
Réponse: $P[M_n \leqslant x] = P[\bigcap_{i=1}^n X_i \leqslant n] = P[X \leqslant x]^n = (\frac{x}{\theta})^n$. M_n a donc pour densité

$$\begin{cases} nx^{n-1}/\theta^n & \text{si } 0 \leqslant x \leqslant \theta \\ 0 & \text{sinon} \end{cases}.$$

$E[M_n] = \int_0^\theta x(nx^{n-1})/\theta^n \mathrm{d}x = \frac{n}{n+1}\theta$.

Un estimateur de θ est donc $\frac{n}{n+1}M_n$.

▷ **Question 12** *Calculer la variance de ce nouvel estimateur et comparer au précédent.*
Réponse: $E[{M_n}^2] = \int_0^\theta x^2(nx^{n-1})/\theta^n \mathrm{d}x = \frac{n}{n+2}\theta^2$.
$\mathrm{Var}[M_n] = E[M_n^2] - E[M_n]^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n\theta}{(n+2)(n+1)^2}$
$\mathrm{Var}[M_n] = \Omega(\frac{1}{n^2})$ tandis que $\mathrm{Var}[\overline{X_n}] = \Omega(\frac{1}{n})$. Ce second estimateur est donc plus précis.