TD (Thème Statistiques)

Exercice 1 (Fraude et statistiques) On suppose qu'il y a une probabilité égale à p d'être contrôlé lorsqu'on prend le tram. Monsieur Bernoulli fait n voyages par an sur cette ligne.

1. On suppose que $p = 0.10$, $n = 700$.
   - Quelle est la probabilité que Monsieur Bernoulli soit contrôlé entre 60 et 80 fois dans l'année ?
   - Monsieur Bernoulli voyage en fait toujours sans ticket. Afin de prendre en compte la possibilité de faire plusieurs passages avec le même ticket, on suppose que le prix d'un ticket est de 1.12 euros. Quelle amende minimale la compagnie doit-elle fixer pour que le fraudeur ait, sur une période d'une année, une probabilité supérieure à 0.75 d'être perdant ?
2. On suppose que $p = 0.5$, $n = 300$. Monsieur Bernoulli voyage toujours sans ticket. Sachant que le prix d'un ticket est de 1.12 euros, quelle amende minimale la compagnie doit-elle fixer pour que le fraudeur ait, sur une période d'une année, une probabilité supérieure à 0.75 d'être perdant ?

1. Posons le nombre de contrôles comme une variable aléatoire X. Elle obéit à une loi binomiale $X \sim \mathcal{B}(700, 0.1)$. On peut l'approcher par une loi normale $\mathcal{N}(70, \sqrt{63})$.
   - Pour calculer la probabilité que Monsieur Bern soit contrôlé entre 60 et 80 fois dans l'année, on évalue la probabilité suivante où $U \sim \mathcal{N}(0, 1)$:

     $$\mathbb{P}(60 \leq X \leq 80) \sim \mathbb{P}\left(-\frac{10}{\sqrt{63}} \leq U \leq \frac{10}{\sqrt{63}}\right) = 2F_U\left(\frac{10}{\sqrt{63}}\right) - 1 = 0.79$$

     La probabilité d'être contrôlé entre 60 et 80 fois dans l'année est donc de 79%.
   - Calculons le prix que devrait payer le voyageur : $1.12 \times 700 = 784$ euros. Il est perdant si les amendes cumulées dépassent ce prix, c'est-à-dire si $aX \geq 784$, où a est le prix de l'amende fixée par la compagnie. On cherche donc a pour que :

     $$\mathbb{P}(aX \geq 784) \geq 0.75 \Leftrightarrow \mathbb{P}(aX \leq 784) \leq 0.25 \Leftrightarrow \mathbb{P}\left(U \leq \frac{784/a - 70}{\sqrt{63}}\right) \leq 0.25 \; .$$

     La table de la loi normale nous fournit u_α tel que $\mathbb{P}(|U| \geq u_\alpha) = \alpha$ soit encore tel que $\mathbb{P}(U \leq -u_\alpha) = \frac{\alpha}{2}$. On prend donc $\alpha = 0.5$ (pour que $\alpha/2 = 0.25$) et on lit dans la table $u_{0.5} = 0.6745$, soit $\mathbb{P}(U \leq -u_{0.5}) = 0.25$. L'égalité a ainsi lieu pour

     $$a = \frac{784}{70 - u_{0.5}\sqrt{63}} \approx 12.13 \; .$$

     Il faut donc que l'amende dépasse 13 euros.
2. On fait de même avec $X \sim \mathcal{B}(300, 0.5) \approx \mathcal{N}(150, \sqrt{75})$ et un prix $1.12 \times 300 = 336$ euros que devrait payer le voyageur. On cherche donc de même b tel que

   $$\mathbb{P}(bX \geq 336) \geq 0.75 \Leftrightarrow \mathbb{P}(bX \leq 336) \leq 0.25 \Leftrightarrow \mathbb{P}\left(U \leq \frac{336/b - 150}{\sqrt{75}}\right) \leq 0.25 \; ,$$

et on obtient cette fois

$$b = \frac{336}{150 - u_{0.5}\sqrt{75}} \approx 2.33 \ .$$

Il faut donc que l'amende dépasse 2 euros 30.

Exercice 2 (Notes) Voulant évaluer rapidement les résultats obtenus par ses 200 étudiants lors d'un partiel, un professeur décide de corriger quelques copies tirées au hasard. Il apprend de sa collègue corrigeant l'autre moitié de la promo que les notes sont dispersées avec une variance égale à 4, et admet qu'elles suivent une loi normale dont il veut estimer la moyenne.

1. Le professeur corrige un échantillon de 7 copies et trouve une moyenne de 11. Quel est l'intervalle de confiance à 95% de la moyenne des 200 copies ?
2. Combien de copies le professeur doit-il corriger s'il veut situer la moyenne générale de ses élèves dans un intervalle de confiance d'amplitude 2, avec un risque de 5% ?
3. En trouvant une moyenne égale à 11, combien de copies le professeur devrait- il corriger pour pouvoir dire, avec un risque de 1%, que la moyenne de tous les élèves est supérieure à 10 ?

1. L'intervalle de confiance de la moyenne des 200 copies au risque 5% est :

$$\left[11 - 1.96 \times \frac{2}{\sqrt{7}}, 11 + 1.96 \times \frac{2}{\sqrt{7}}\right] \approx [9.52, 12.48] \ .$$

2. Si l'amplitude de l'intervalle de confiance est égale à 2, on doit avoir :

$$1.96 \times \frac{2}{\sqrt{n}} = 1 \Rightarrow n \approx 15.4 \ ,$$

   donc l'enseignant peut en corrigeant 16 copies situer la moyenne de ses étudiants.
3. Il faut que l'intervalle de confiance à 99% soit égal à $[10, 12]$. On doit donc avoir :

$$2.575 \times \frac{2}{\sqrt{n}} = 1 \Rightarrow n \approx 26.5 \ ,$$

   donc si l'enseignant corrige 27 copies et trouve une moyenne de 11, alors il peut dire que la moyenne de ses étudiants est supérieure à 10, avec un risque d'erreur de 1%.

Exercice 3 (Porte alcootest) Après une journée de vendanges dans cette vigne, Mr Brown rentre parfois chez lui en zigzaguant. Une fois devant sa porte, il sort de sa poche un trousseau de $k \geq 2$ clefs, dont une seule ouvre la porte. Lorsqu'il est à jeun, il essaye une des clefs au hasard, puis, si elle ne fonctionne pas, la met de côté et essaye une des clefs restantes, répétant cette opération après chaque échec. Tandis que lorsqu'il est ivre, il recommence avec cette fois le trousseau complet.
On note A l'évènement « Mr Brown est à jeun » et B l'évènement « Mr Brown est ivre ».

1. Calculer, sachant que Mr Brown est à jeun, la probabilité qu'il parvienne à ouvrir la porte au terme exactement de $n > 0$ tentatives. On désigne par X_A la variable aléatoire représentant le nombre d'essais nécessaires pour parvenir à ouvrir la porte quand il est

à jeun. Quelle loi suit la variable X_A ? Que vaut son espérance ?

2. Mêmes questions sachant que Mr Brown est ivre avec la variable aléatoire X_B.
3. On suppose que Mr Brown a la même probabilité de rentrer à jeun ou ivre $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$. Calculer la probabilité p_n qu'il soit ivre, sachant qu'il est parvenu à ouvrir la porte au terme d'exactement n essais. On notera U_n l'évènement « Mr Brown a ouvert la porte au terme de n essais exactement » et on distinguera les cas où $n \leq k$ et $n > k$. *Indice : utilisez la formule de Bayes.*

1. Puisqu'il y a k clefs, la probabilité que « le premier essai d'ouverture soit un succès » (évènement S_1) est
$$\mathbb{P}(\{X_A = 1\}) = \mathbb{P}(S_1) = \frac{1}{k} \ .$$
On ouvre la porte au bout de 2 essais (évènement $\{X_A = 2\}$) quand « le premier essai se solde par un échec » (évènement E_1) et que « le second essai se solde par un succès » (évènement S_2), soit :
$$\mathbb{P}(\{X_A = 2\}) = \mathbb{P}(E_1 \cap S_2) = \mathbb{P}(E_1) \times \mathbb{P}(S_2 \,|\, E_1) \ ,$$
avec $\mathbb{P}(E_1) = 1 - \mathbb{P}(S_1) = \frac{k-1}{k}$ et $\mathbb{P}(S_2 \,|\, E_1) = \frac{1}{k-1}$ car le trousseau ne comprend plus que $k-1$ clefs au second essai, d'où encore :
$$\mathbb{P}(\{X_A = 2\}) = \frac{1}{k} \ .$$
Montrons alors que la quantité $\mathbb{P}(\{X_A = n\})$ est constante égale à k quel que soit $n \leq k$. En effet, en procédant de même on ouvre la porte au bout de n essais (évènement $\{X_A = n\}$) quand « les $n-1$ premiers essais se soldent par un échec » (évènement $F_{n-1} = E_1 \cap \cdots \cap E_{n-1}$) et que « le dernier essai se solde par un succès » (évènement S_n), soit :
$$\mathbb{P}(\{X_A = n\}) = \mathbb{P}(F_{n-1} \cap S_n) = \mathbb{P}(F_{n-1}) \times \mathbb{P}(S_n \,|\, F_{n-1}) \ ,$$
avec $\mathbb{P}(S_n \,|\, F_{n-1}) = \frac{1}{k-(n-1)}$ car le trousseau ne comprend plus que $k-(n-1)$ clefs au n-ième essai. Enfin pour calculer $\mathbb{P}(F_n)$ on applique toujours la formule conditionnelle :
$$\begin{aligned}
\mathbb{P}(F_n) &= \mathbb{P}((E_1 \cap \cdots \cap E_{n-1}) \cap E_n) \ , \\
&= \mathbb{P}(E_1 \cap \cdots \cap E_{n-1}) \times \mathbb{P}(E_n \,|\, E_1 \cap \cdots \cap E_{n-1}) \ , \\
&= \mathbb{P}(F_{n-1}) \times \mathbb{P}(E_n \,|\, F_{n-1}) \ , \\
&= \mathbb{P}(F_{n-1}) \times (1 - \mathbb{P}(S_n \,|\, F_{n-1})) \ , \\
&= \mathbb{P}(F_{n-1}) \times \frac{k-n}{k-(n-1)} \ .
\end{aligned}$$
On a donc un produit télescopique s'arrêtant à $\mathbb{P}(F_1) = \mathbb{P}(E_1) = \frac{k-1}{k}$ c.-à-d. :
$$\mathbb{P}(F_n) = \frac{k-n}{k-(n-1)} \times \frac{k-(n-1)}{k-(n-2)} \times \cdots \times \frac{k-1}{k} = \frac{k-n}{k} \ .$$

On en conclut bien que

$$\mathbb{P}(\{X_A = n\}) = \mathbb{P}(F_{n-1}) \times \mathbb{P}(S_n \,|\, F_{n-1}) = \frac{k-(n-1)}{k} \times \frac{1}{k-(n-1)} = \frac{1}{k} \; .$$

Pour $n > k$ on a $\mathbb{P}(\{X_A = n\}) = 0$ car on aura à coup sûr trouver la bonne clef avant.
Remarque 1 : au terme de $k-1$ essais infructueux il ne reste plus qu'une clef dans le trousseau qui est forcément celle qui ouvre la porte.
La v.a X_A suit donc la loi discrète uniforme sur $\{1, \ldots, k\}$, et ainsi son espérance est :

$$\mathbb{E}[X_A] = \frac{1}{k}(1 + 2 + \cdots + k) = \frac{k+1}{2} \; .$$

2. Lorsqu'il est ivre il recommence avec le trousseau complet, donc les essais sont indépendants avec pour chaque essai i une probabilité constante de succès $\mathbb{P}(S_i) = \frac{1}{k}$ et d'échec $\mathbb{P}(E_i) = 1 - \frac{1}{k}$. Par conséquent :

$$\mathbb{P}(\{X_B = n\}) = \mathbb{P}(E_1 \cap \cdots \cap E_{n-1} \cap S_n) = \mathbb{P}(E_1) \times \cdots \times \mathbb{P}(E_{n-1}) \times \mathbb{P}(S_n) = \left(1 - \frac{1}{k}\right)^{n-1} \frac{1}{k} \; .$$

Ainsi la variable aléatoire X_B suit une loi géométrique de paramètre $p = \frac{1}{k}$, et donc son espérance est :

$$\mathbb{E}[X_B] = \frac{1}{p} = k \; .$$

3. On sait d'après les deux questions précédentes que si $n \leq k$:

$$\mathbb{P}(U_n \,|\, A) = \frac{1}{k}, \quad \mathbb{P}(U_n \,|\, B) = \left(1 - \frac{1}{k}\right)^{n-1} \frac{1}{k} \; .$$

On cherche à calculer la probabilité d'être ivre sachant que la porte s'est ouverte au terme d'exactement n essais, soit $p_n = \mathbb{P}(B \,|\, U_n)$. On applique alors la formule de Bayes :

$$\mathbb{P}(B \,|\, U_n) = \frac{\mathbb{P}(U_n \,|\, B)\mathbb{P}(B)}{\mathbb{P}(U_n \,|\, A)\mathbb{P}(A) + \mathbb{P}(U_n \,|\, B)\mathbb{P}(B)} \; ,$$

et comme ici $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$ on a donc :

$$p_n = \frac{\left(1 - \frac{1}{k}\right)^{n-1} \frac{1}{k}}{\frac{1}{k} + \left(1 - \frac{1}{k}\right)^{n-1} \frac{1}{k}} = \frac{1}{1 + \left(1 + \frac{1}{k-1}\right)^{n-1}}, \quad \forall n \leq k \; .$$

Quand $n > k$ on a montré que $\mathbb{P}(U_n \,|\, A) = 0$ donc on obtient dans la formule de Bayes :

$$p_n = 1, \quad \forall n > k \; .$$

Remarque 2 : quand $n \leq k$ on constate que $p_n \leq 1/2$ et donc on peut difficilement faire le pari qu'il est ivre quand il parvient à rentrer avec moins d'essais que le nombre de clefs du trousseau, tandis qu'il est évidemment certain qu'il l'est ($p_n = 1$) quand le nombre d'essais surpasse le nombre de clefs du trousseau, ce que l'on avait déjà attesté dans la Remarque 1 avec du bon sens et sans besoin de considérations statistiques. Cet alcootest n'est donc pas très élaboré.

Remarque 3 : la probabilité p_k est la valeur la plus faible, ainsi le fait d'être ivre quand le nombre d'essai est exactement égal à k est le cas le plus improbable, alors même que lorsqu'on est ivre on fait en moyenne k essais pour ouvrir la porte. Cette probabilité p_k ne tend ceci dit par vers 0 quand le nombre de clefs du trousseau k augmente, mais converge vers $1/(1+e) \approx 0.27$.