I3-I4, Statistiques *mai 2018*

Devoir

Documents et calculatrice sont autorisés.

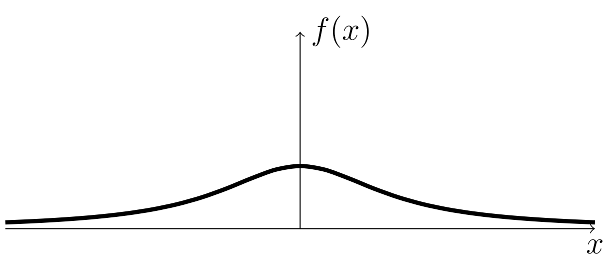
Exercice 1 : roulette faussée ?

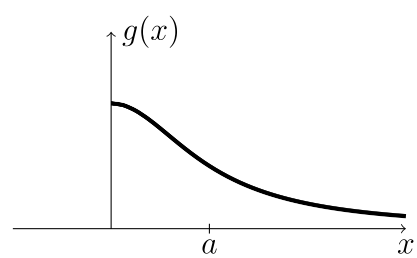
Un casino a un doute sur une de ses roulettes : il semble que le nombre 7 apparaisse moins d'une fois sur 37. On a en effet effectué 1000 tirages et il n'est sorti que 20 fois.

1. Estimer la probabilité p de tomber sur 7 à l'aide d'une moyenne.
2. Réaliser un test de moyenne afin de répondre aux doutes du casino.
3. Combien faudrait-il réaliser de tirages pour estimer p avec une précision de $0,001$?
4. Nous proposons maintenant de répondre au casino, à partir des mêmes données, à l'aide d'un test d'adéquation à la loi uniforme sur $[\![0, 37]\!]$.
   Diviser l'univers en deux sous-ensembles $\{7\}$ et $[\![0, 37]\!] \setminus \{7\}$ et effectuer le test d'adéquation.
5. Laquelle des deux méthodes employées semble la plus précise et pertinente ?

Exercice 2 : loi de Cauchy

La loi de Cauchy de paramètre a, notée $\mathcal{C}(a)$, est la loi de densité $f(x) = \dfrac{1}{\pi} \dfrac{a}{x^2 + a^2}$. L'objectif de cet exercice est d'estimer a à partir d'un échantillon.





1. On dispose de l'échantillon $x_1 = 1$, $x_2 = -2$ pour estimer a.
   Déterminer l'estimateur du maximum de vraisemblance de a correspondant à ces deux valeurs.
2. Généraliser pour un échantillon $x_1, \dots, x_n$ de taille n : déterminer en fonction de ces x_i et de n l'équation que l'on doit résoudre pour déterminer l'estimateur du maximum de vraisemblance de a. (On ne demande pas de la résoudre.)

On souhaite estimer a autrement. La loi de Cauchy n'ayant ni espérance, ni variance, on ne peut pas utiliser les estimateurs standards. On propose d'utiliser la médiane.

La médiane d'une variable aléatoire est la valeur M telle que $\mathbf{P}(X \geqslant M) = \mathbf{P}(X < M) = \frac{1}{2}$. Par exemple, la médiane de la loi de Cauchy est 0, ce qui ne nous aidera pas à déterminer a. On va alors s'intéresser à la variable $|X|$ où X suit la loi de Cauchy $\mathcal{C}(a)$. C'est une variable positive dont la densité est $g(x) = \frac{2}{\pi}\frac{a}{x^2+a^2}$ pour $x \in \mathbf{R}_+$.

3. Montrer que la médiane de $|X|$ est $M = a$.
   Indication : quelle est la dérivée de $x \mapsto \arctan(\frac{x}{a})$ *?*

Il est donc possible d'estimer naturellement a à partir d'un échantillon en prenant la valeur médiane $\overline{M}$ de l'échantillon. On dispose même d'un théorème permettant d'établir des intervalles de confiance : lorsque la taille de l'échantillon n tend vers $+\infty$, la variable aléatoire $2g(M)\sqrt{n}(\overline{M} - M)$ converge en loi vers la loi normale centrée réduite : autrement dit pour n assez grand, on peut considérer que

$$2g(M)\sqrt{n}\left(\overline{M} - M\right) \sim \mathcal{N}(0,1).$$

4. On dispose d'un échantillon de taille 100. La valeur médiane des valeurs absolues de l'échantillon est $\overline{M} = 1,2$.
   Déterminer un intervalle de confiance au risque de 5% pour a.