

Ce sujet s'intéressait à des problèmes de biais statistiques.

Partie I - Biais par la taille, exemples discrets

1. On suppose que le nombre d'enfants dans une famille française est une variable aléatoire X . Pour connaître la loi de X , une idée serait d'interroger les élèves d'une école pour connaître le nombre d'enfants dans leur famille. On suppose que X suit la loi binomiale de paramètres $n = 10$ et $p = 1/5$. On note $p_k = \mathbb{P}(X = k)$ pour $k \in \{0, 1, \dots, 10\}$.

a) i. D'après le cours sur la loi binomiale : $\forall k \in \{0, 1, \dots, 10\}$, $p_k = \binom{10}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{10-k}$.

ii. Toujours d'après le cours sur la loi binomiale : $E(X) = 10 \times \frac{1}{5} = 2$.

iii. Enfin, d'après le cours : $\text{Var}(X) = 10 \times \frac{1}{5} \times \frac{4}{5} = \frac{8}{5}$, et d'après la formule de Koenig-Huygens :

$$\text{Var}(X) = E(X^2) - E(X)^2 \iff E(X^2) = \text{Var}(X) + E(X)^2 = \frac{8}{5} + 4 = \frac{28}{5}$$

b) Soit M_k le nombre de famille à k enfants, $M = \sum_{k=0}^{10} M_k$ le nombre total de familles

(donc $p_k = M_k/M$).

Soit N_k le nombre total d'enfants (c'est-à-dire dans toute la population) qui font partie d'une famille à k enfants, et $N = \sum_{k=0}^{10} N_k$ le nombre total d'enfants dans la population.

i. Le nombre total d'enfants qui font partie d'une famille à k enfants est logiquement égal à $k \times M_k$, soit $N_k = k \times \frac{M_k}{M} \times M = kp_k M$.

ii. On a alors : $N = \sum_{k=0}^{10} N_k = \sum_{k=0}^{10} kp_k M = M \sum_{k=0}^{10} kp_k$, où $\sum_{k=0}^{10} kp_k$ est par définition l'espérance de X , qui vaut 2, donc : $N = 2M \iff N/M = 2$.

iii. La proportion des enfants qui proviennent d'une famille à k enfants est donc :

$$p_k^* = \frac{N_k}{N} = \frac{kp_k M}{2M} = \frac{kp_k}{2}$$

d'après les deux questions précédentes.

c) On choisit une personne au hasard dans la rue, à qui l'on demande combien d'enfants ses parents ont eu (lui ou elle inclus). On note Y ce nombre d'enfants.

i. Pour tout entier k de $\{1, 2, \dots, 10\}$, la probabilité $\mathbb{P}(Y = k)$ est celle d'avoir interrogé une personne qui vient d'une famille à k enfants, qui correspond bien à la proportion $p_k^* = kp_k/2$.

- ii. La variable aléatoire Y a pour univers-image $\{1, 2, \dots, 10\}$, elle est donc finie et admet une espérance donnée par la formule :

$$E(Y) = \sum_{k=1}^{10} k\mathbb{P}(Y = k) = \sum_{k=1}^{10} k^2 p_k / 2 = \frac{1}{2} \sum_{k=0}^{10} k^2 p_k = \frac{E(X^2)}{2} = \frac{E(X^2)}{E(X)}$$

- iii. Avec les valeurs obtenues en 1.a) : $E(Y) = \frac{28/5}{2} = \frac{14}{5} = 2.8 > 2 = E(X)$.

Le biais et l'espérance supérieure pour Y est ici logique : si on interroge une personne, alors elle fait forcément partie d'une famille ayant eu au moins un enfant ! Toutes les familles sans enfant sont donc ignorées dans le calcul de Y .

2. Soit X une variable aléatoire à valeurs dans \mathbb{N} , non identiquement nulle et admettant une espérance.

Pour tout entier $i > 0$, on pose $q_i = \frac{i}{E(X)} \mathbb{P}(X = i)$.

- a) Comme on a supposé que X admet une espérance, alors la série de terme général

$q_i = \frac{1}{E(X)} \cdot i\mathbb{P}(X = i)$ est à un facteur près, celle qui définit cette espérance, donc elle converge et a pour somme totale :

$$\sum_{i=1}^{+\infty} q_i = \frac{1}{E(X)} \sum_{i=1}^{+\infty} i\mathbb{P}(X = i) = \frac{1}{E(X)} \sum_{i=0}^{+\infty} i\mathbb{P}(X = i) = \frac{1}{E(X)} \cdot E(X) = 1$$

(le terme de la somme pour $i = 0$ étant nul, on peut le faire apparaître à loisir dans la somme.)

La suite $(q_i)_{i>0}$ définie ci-dessus définit donc bien une loi de probabilité. On considère la variable aléatoire X^* dont la loi est donnée par les q_i : $\forall i \in \mathbb{N}^*, \mathbb{P}(X^* = i) = \frac{i}{E(X)} \mathbb{P}(X = i)$.

On dit que X^ suit la loi de X biaisée par la taille.*

- b) On suppose que X admet un moment d'ordre 2 : la série $\sum_{i>0} i^2 \mathbb{P}(X = i)$ est donc convergente,

c'est donc aussi le cas de la série de terme général (positif) $i\mathbb{P}(X^* = i) = \frac{1}{E(X)} \cdot i^2 \mathbb{P}(X = i)$, donc X^* admet une espérance qui vaut :

$$E(X^*) = \frac{1}{E(X)} \sum_{i=1}^{+\infty} i^2 \mathbb{P}(X = i) = \frac{1}{E(X)} \sum_{i=0}^{+\infty} i^2 \mathbb{P}(X = i) = \frac{E(X^2)}{E(X)}$$

- c) Si $E(X^2)$ existe, alors X admet aussi une variance donnée par la formule de Koenig-Huygens :

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{E(X^2)}{E(X)} \cdot E(X) - E(X)^2 = E(X) \cdot E(X^*) - E(X)^2 = E(X) \cdot (E(X^*) - E(X))$$

- d) Une variance est toujours positive, et X est une variable aléatoire à valeurs dans \mathbb{N} non identiquement nulle, donc $E(X) > 0$. Les règles de signes du produit assurent alors que :

$$\text{Var}(X) \geq 0 \iff E(X)(E(X^*) - E(X)) \geq 0 \implies E(X^*) - E(X) \geq 0 \implies E(X^*) \geq E(X)$$

3. a) Soit λ un réel strictement positif. On suppose que X est une variable aléatoire qui suit la loi de Poisson de paramètre λ . Soit X^* une variable aléatoire suivant la loi de X biaisée par la taille.

- i. D'après le cours sur la loi de Poisson : $E(X) = \lambda$, donc la loi de X^* est donnée par :

$$\forall i \in \mathbb{N}^*, \mathbb{P}(X^* = i) = \frac{1}{\lambda} \cdot i \cdot e^{-\lambda} \cdot \frac{\lambda^i}{i!} = e^{-\lambda} \cdot \frac{\lambda^{i-1}}{(i-1)!}$$

- ii. Les variables aléatoires X^* et $X+1$ ont alors le même univers-image \mathbb{N}^* , et le calcul précédent fait clairement apparaître que pour tout i de \mathbb{N}^* :

$$\mathbb{P}(X^* = i) = \mathbb{P}(X = i - 1) = \mathbb{P}(X + 1 = i)$$

Donc X^* et $X + 1$ suivent la même loi.

- b) Réciproquement, on suppose que X est une variable aléatoire à valeurs dans \mathbb{N} admettant une espérance non nulle, telle que X^* et $X + 1$ suivent la même loi.

- i. Pour tout entier $k \geq 1$, on a donc :

$$\mathbb{P}(X^* = k) = \mathbb{P}(X+1 = k) \iff \frac{1}{E(X)} \cdot k \mathbb{P}(X = k) = \mathbb{P}(X = k-1) \iff \mathbb{P}(X = k) = \frac{E(X)}{k} \mathbb{P}(X = k-1)$$

- ii. On peut alors enchaîner avec une récurrence simple pour montrer que la propriété

$\mathcal{P}(k)$: " $\mathbb{P}(X = k) = \frac{E(X)^k}{k!} \mathbb{P}(X = 0)$ " est vraie pour tout $k \in \mathbb{N}$:

[I.] pour $k = 0$: $\frac{E(X)^0}{0!} \mathbb{P}(X = 0) = \mathbb{P}(X = 0)$, donc $\mathcal{P}(0)$ est vraie.

[H.] Supposons $\mathcal{P}(k)$ vraie pour un certain $k \in \mathbb{N}$, et montrons qu'alors $\mathcal{P}(k+1)$ est encore vraie, soit : $\mathbb{P}(X = k+1) = \frac{E(X)^{k+1}}{(k+1)!} \mathbb{P}(X = 0)$.

D'après la question i. : $\mathbb{P}(X = k+1) = \frac{E(X)}{k+1} \mathbb{P}(X = k) \stackrel{H.R.}{=} \frac{E(X)}{k+1} \times \frac{E(X)^k}{k!} \mathbb{P}(X = 0) = \frac{E(X)^{k+1}}{(k+1)!} \mathbb{P}(X = 0)$, donc $\mathcal{P}(k+1)$ est vraie si $\mathcal{P}(k)$ l'est.

[C.] La propriété est initialisée et héréditaire : elle est donc vraie pour tout $k \in \mathbb{N}$, d'après le principe de récurrence.

- iii. On a donc : $\forall k \in \mathbb{N}$, $\mathbb{P}(X = k) = \frac{E(X)^k}{k!} \mathbb{P}(X = 0)$, où on doit encore déterminer $\mathbb{P}(X = 0)$; par définition d'une variable aléatoire, ici à valeurs dans \mathbb{N} , on doit avoir :

$$\sum_{k=0}^{+\infty} \mathbb{P}(X = k) = 1 \iff \mathbb{P}(X = 0) \sum_{k=0}^{+\infty} \frac{E(X)^k}{k!} = 1 \iff \mathbb{P}(X = 0) \cdot e^{E(X)} = 1 \iff \mathbb{P}(X = 0) = e^{-E(X)}$$

De sorte que : $\forall k \in \mathbb{N}$, $\mathbb{P}(X = k) = e^{-E(X)} \cdot \frac{E(X)^k}{k!}$, c'est-à-dire que la variable aléatoire X suit une loi de Poisson dont le paramètre est logiquement égal à son espérance.

On vient donc de montrer que les seules variables discrètes X dont la loi biaisée par la taille, est la même que celle de $X + 1$, sont celles qui suivent une loi de Poisson.

4. *Le paradoxe du temps d'attente du bus.* Soit $n \geq 1$ un entier naturel, et soit x une variable aléatoire à valeurs dans $\{1, \dots, n\}$ telle que pour tout $1 \leq k \leq n$, $\mathbb{P}(X = k) > 0$. On suppose qu'à un arrêt de bus donné, les intervalles de temps entre deux bus consécutifs, exprimés en minutes, sont des variables aléatoires indépendantes, de même loi que X . Une personne arrive à cet arrêt à un instant aléatoire, et se demande combien de temps elle va attendre.

- a) Une première idée est que la personne arrive à un instant uniforme entre deux arrivées de bus, séparées par un intervalle de X minutes. On note T la variable aléatoire qui représente le temps d'attente, à valeurs dans $\{1, \dots, n\}$, et on suppose donc que pour tout entier k de $\{1, \dots, n\}$, $\mathbb{P}_{[X=k]}(T = j) = 1/k$ si $1 \leq j \leq k$, et $\mathbb{P}_{[X=k]}(T = j) = 0$ si $j > k$.

- i. Pour tout entier $k \in \{1, \dots, n\}$:

$$\sum_{j=1}^n j \mathbb{P}_{[X=k]}(T = j) = \sum_{j=1}^k \frac{j}{k} + \sum_{j=k+1}^n j \cdot 0 = \frac{1}{k} \sum_{j=1}^k j = \frac{1}{k} \times \frac{k(k+1)}{2} = \frac{k+1}{2}$$

ii. Par conséquent :

$$\begin{aligned} \sum_{k=1}^n \sum_{j=1}^n j \mathbb{P}(X = k) \mathbb{P}_{[X=k]}(T = j) &= \sum_{k=1}^n \mathbb{P}(X = k) \cdot \sum_{j=1}^n j \cdot \mathbb{P}_{[X=k]}(T = j) = \sum_{k=1}^n \mathbb{P}(X = k) \cdot \frac{k+1}{2} \\ &= \frac{1}{2} \sum_{k=1}^n (k+1) \mathbb{P}(X = k) = \frac{E(X+1)}{2} \end{aligned}$$

(d'après le théorème de transfert pour la dernière étape.)

iii. La variable aléatoire T est finie puisqu'à valeurs dans $\{1, \dots, n\}$, donc elle admet une espérance qui vaut : $E(T) = \sum_{j=1}^n j \mathbb{P}(T = j)$.

Or chaque probabilité $\mathbb{P}(T = j)$ peut être exprimée à l'aide de la formule des probabilités totales avec le système complet d'événements $([X = k])_{1 \leq k \leq n}$ associé à X :

$$\forall j \in \{1, \dots, n\}, \quad \mathbb{P}(T = j) = \sum_{k=1}^n \mathbb{P}(X = k) \cdot \mathbb{P}_{[X=k]}(T = j)$$

On a donc en effet :

$$E(T) = \sum_{j=1}^n j \sum_{k=1}^n \mathbb{P}(X = k) \cdot \mathbb{P}_{[X=k]}(T = j) = \sum_{j=1}^n \sum_{k=1}^n j \mathbb{P}(X = k) \cdot \mathbb{P}_{[X=k]}(T = j)$$

iv. Dans cette sommation double, les deux symboles sommes ne sont pas interdépendants donc on peut les changer d'ordre sans que cela ne change la valeur finale de la somme. Les résultats des deux questions précédentes sont donc égaux, et on a bien $E(T) = \frac{E(X+1)}{2}$.

b) En réalité, en arrivant à l'arrêt de bus, on "tombe" dans un intervalle entre deux bus de manière proportionnelle à sa taille (plus l'intervalle est long, plus on a de chances de "tomber" dedans) : l'intervalle de temps est X^* , suivant la loi de X biaisée par la taille. Le temps d'attente T^* vérifie donc en fait, pour tout $k \in \{1, \dots, n\}$, $\mathbb{P}_{[X^*=k]}(T^* = j) = 1/k$ si $j \in \{1, \dots, k\}$ et $\mathbb{P}_{[X^*=k]}(T^* = j) = 0$ si $j > k$.

i. C'est le même calcul qu'en a)i. : pour tout entier $k \in \{1, \dots, n\}$,

$$\sum_{j=1}^n j \mathbb{P}_{[X^*=k]}(T^* = j) = \sum_{j=1}^k \frac{j}{k} + \sum_{j=k+1}^n j \cdot 0 = \frac{1}{k} \sum_{j=1}^k j = \frac{k(k+1)}{2k} = \frac{k+1}{2}$$

ii. La variable aléatoire T^* est toujours finie, et admet une espérance qui vaut :

$$E(T^*) = \sum_{j=1}^n j \mathbb{P}(T^* = j) = \sum_{j=1}^n j \sum_{k=1}^n \mathbb{P}(X^* = k) \cdot \mathbb{P}_{[X^*=k]}(T^* = j) = \sum_{j=1}^n \sum_{k=1}^n j \mathbb{P}(X^* = k) \cdot \mathbb{P}_{[X^*=k]}(T^* = j)$$

iii. Si on échange les deux symboles sommes (opération licite), on obtient :

$$E(T^*) = \sum_{k=1}^n \mathbb{P}(X^* = k) \sum_{j=1}^n j \mathbb{P}_{[X^*=k]}(T^* = j) = \sum_{k=1}^n \mathbb{P}(X^* = k) \times \frac{k+1}{2} = \frac{E(X^* + 1)}{2}$$

iv. On a vu à la question 2.d) que $E(X^*) \geq E(X)$, donc $E(X^*) + 1 \geq E(X) + 1$
 $\iff E(X^* + 1) \geq E(X + 1)$ par linéarité de l'espérance, ce qui démontre bien que $E(T^*) \geq E(T)$.

Deuxième partie : biais par la taille, propriétés

5. *Biais par la taille : le cas de variables à densité.* Soit X une variable aléatoire **positive** de densité f et admettant une espérance $E(X)$ strictement positive (donc $f(x) = 0$ pour tout x strictement négatif).

On définit la fonction g par $g(x) = \frac{x}{E(X)}f(x)$ pour tout x réel.

a) La fonction g est nulle sur $] -\infty; 0[$ car f l'est, et pour tout $x \geq 0$, $g(x) = \frac{x}{E(X)}f(x)$ est positif puisque $x \geq 0$, $f(x) \geq 0$ (f est une densité) et $E(X) > 0$.

La fonction g est continue sur $] -\infty; 0[$ comme fonction constante, continue sur $]0; +\infty[$ comme produit de fonctions continues sur cet intervalle ($E(X)$ étant un terme constant), donc g est continue sur \mathbb{R} sauf peut-être en 0.

Enfin, sous réserve de convergence :

$$\int_{-\infty}^{+\infty} g(x)dx = \int_{-\infty}^0 0 dx + \frac{1}{E(X)} \int_0^{+\infty} xf(x)dx$$

Comme f est nulle sur $] -\infty; 0[$ et X admet une espérance, alors $\int_0^{+\infty} xf(x)dx$ converge et vaut $E(X)$, donc $\int_{-\infty}^{+\infty} g(x)dx = \frac{1}{E(X)} \cdot E(X) = 1$, ce qui achève de démontrer que g est bien une densité de probabilité d'une variable aléatoire positive (puisque $g(x)$ est nul pour tout x strictement négatif).

Soit une variable aléatoire X^* dont la densité est g . On dit que X^* suit la loi de X biaisée par la taille.

b) Soit a un réel strictement positif.

i. D'après le cours : la variable aléatoire aX , transformée affine de X , a pour densité la fonction $x \mapsto \frac{1}{|a|}f\left(\frac{x-0}{a}\right) = \frac{1}{a}f\left(\frac{x}{a}\right)$.

Si on veut vraiment refaire la démonstration, on calcule la fonction de répartition de aX : pour tout réel x ,

$$F_{aX}(x) = \mathbb{P}(aX \leq x) \stackrel{a \geq 0}{=} \mathbb{P}\left(X \leq \frac{x}{a}\right) = F_X\left(\frac{x}{a}\right)$$

Comme X est une variable à densité, F_X est continue sur \mathbb{R} , de classe \mathcal{C}^1 sauf peut-être en un nombre fini de points. La fonction affine $x \mapsto \frac{x}{a}$ étant de classe \mathcal{C}^1 sur \mathbb{R} , par composition F_{aX} est bien continue sur \mathbb{R} , de classe \mathcal{C}^1 sauf en un nombre fini de points et aX est une variable à densité, dont une densité est obtenue par dérivation de F_{aX} là où c'est possible :

$F'_{aX}(x) = \frac{1}{a} \cdot F'_X\left(\frac{x}{a}\right) = \frac{1}{a} f\left(\frac{x}{a}\right)$ là où la dérivée est bien définie, et on peut arbitrairement choisir l'expression finale $\frac{1}{a} f\left(\frac{x}{a}\right)$ pour tout x réel, pour une densité de aX .

ii. Au vu de ce qui précède :

- une densité de la variable aléatoire $(aX)^*$ est la fonction $x \mapsto \frac{x}{E(aX)} \cdot \frac{1}{a} f\left(\frac{x}{a}\right) = \frac{x}{a^2 E(X)} f\left(\frac{x}{a}\right)$ par linéarité de l'espérance.
- une densité de $a \times X^*$ est la fonction $x \mapsto \frac{1}{a} g\left(\frac{x}{a}\right) = \frac{1}{a} \cdot \frac{x/a}{E(X)} f\left(\frac{x}{a}\right) = \frac{x}{a^2 E(X)} f\left(\frac{x}{a}\right)$.

Les variables aléatoires $(aX)^*$ et $a \times X^*$ admettent donc la même densité, et suivent par conséquent la même loi.

c) *Une propriété importante.* Soit $h : [0; +\infty[\rightarrow \mathbb{R}$ une fonction bornée et continue sauf éventuellement en un nombre fini de points ; il existe donc un réel positif M tel que : $\forall x \in \mathbb{R}, |h(x)| \leq M$.

D'après le théorème de transfert, l'espérance $E(Xh(X))$ est alors bien définie si et seulement si l'intégrale $\int_{-\infty}^{+\infty} xh(x)f(x)dx$ est absolument convergente.

La fonction $x \mapsto xh(x)f(x)$ est nulle sur $] - \infty; 0[$ puisque f l'est, et pour tout $x \geq 0$, $|xh(x)f(x)| = xf(x)|h(x)| \leq M.xf(x)$.

Or l'intégrale $M. \int_0^{+\infty} xf(x)dx$ converge puisque X admet une espérance. Par comparaison d'intégrales de fonctions continues (sauf en un nombre fini de points) et positives, l'intégrale $\int_0^{+\infty} xh(x)f(x)dx$ est absolument convergente ; ainsi, $E(Xh(X))$ existe et

$$E(Xh(X)) = \int_0^{+\infty} xh(x)f(x)dx = E(X). \int_0^{+\infty} h(x). \frac{x}{E(X)}f(x)dx = E(X) \int_0^{+\infty} h(x).g(x)dx = E(X).E(h(X^*))$$

toujours d'après le théorème de transfert et le fait que g est une densité de X^* nulle sur $] - \infty; 0[$. La convergence absolue de l'intégrale impliquait au passage le fait que X^* admet une espérance, et donne donc la relation :

$$E(Xh(X)) = E(X).E(h(X^*)) \iff E(h(X^*)) = \frac{1}{E(X)}E(Xh(X))$$

On pose alors la définition suivante (que la variable X soit à densité ou non) : si X est une variable aléatoire réelle positive d'espérance $E(X)$ strictement positive, on dit que la variable aléatoire positive Y suit la loi de X biaisée par la taille si on a

$$E(h(Y)) = \frac{1}{E(X)}E(Xh(X))$$

pour toute fonction $h : [0; +\infty[\rightarrow \mathbb{R}$ bornée et continue sur \mathbb{R} sauf éventuellement en un nombre fini de points.

6. Dans cette question, on se fixe $f : \mathbb{R} \rightarrow \mathbb{R}$ et $g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions croissantes. Soit X une variable aléatoire telle que les espérances $E(f(X))$, $E(g(X))$ et $E(f(X)g(X))$ soient bien définies.

a) Pour tous réels x_1 et x_2 :

si $x_1 \leq x_2$, alors par croissance de f et g sur \mathbb{R} , $f(x_1) \leq f(x_2)$ et $g(x_1) \leq g(x_2)$

donc $(f(x_1) - f(x_2))(g(x_1) - g(x_2))$ est positif comme produit de deux facteurs négatifs.

Sinon bien sûr, $x_1 > x_2$ donc $f(x_1) \geq f(x_2)$ et $g(x_1) \geq g(x_2)$ et $(f(x_1) - f(x_2))(g(x_1) - g(x_2))$ est encore positif comme produit de deux facteurs positifs cette fois.

b) Soient X_1, X_2 deux variables aléatoires indépendantes, de même loi que X : vu les hypothèses faites sur X , on peut utiliser la linéarité de l'espérance pour écrire, sous réserve d'existence :

$$E((f(X_1) - f(X_2))(g(X_1) - g(X_2))) = E(f(X_1)g(X_1)) - E(f(X_1)g(X_2)) - E(f(X_2)g(X_1)) + E(f(X_2)g(X_2))$$

L'indépendance de X_1 et X_2 entraîne celle de $f(X_1)$ et $g(X_2)$, ainsi que celle de $g(X_1)$ et $f(X_2)$ d'après le lemme des coalitions, donc $E(f(X_1)g(X_2)) = E(f(X_1)) \times E(g(X_2))$ et $E(f(X_2)g(X_1)) = E(f(X_2)) \times E(g(X_1))$.

De plus, comme X_1 et X_2 suivent la même loi que X , alors $E(f(X_1)) = E(f(X)) = E(f(X_2))$, $E(g(X_1)) = E(g(X)) = E(g(X_2))$ et $E(f(X_1)g(X_1)) = E(f(X)g(X)) = E(f(X_2)g(X_2))$, donc l'espérance envisagée au départ existe bien et vaut :

$$E((f(X_1) - f(X_2))(g(X_1) - g(X_2))) = 2E(f(X)g(X)) - 2(f(X))E(g(X))$$

- c) La question a) a prouvé que la variable aléatoire $(f(X_1) - f(X_2))(g(X_1) - g(X_2))$ est positive ; la propriété de positivité de l'espérance (qui existe bien pour cette variable aléatoire) assure alors que :

$$E((f(X_1) - f(X_2))(g(X_1) - g(X_2))) \geq 0 \iff 2E(f(X)g(X)) - 2(f(X))E(g(X)) \geq 0$$

$$\iff E(f(X)g(X)) \geq E(f(X))E(g(X))$$

7. Dans cette question, on suppose que X est une variable aléatoire positive d'espérance strictement positive, et telle que $E(X^{m+1})$ existe pour un entier $m \geq 1$ donné.

- a) Soit p un entier naturel tel que $1 \leq p \leq m$.

- i. Soit x un réel positif, on distingue deux cas :

- soit $0 \leq x \leq 1$, alors $0 \leq x^p \leq 1 \leq 1 + x^{m+1}$.
- soit $x \geq 1$, et alors puisque $p < m + 1$, $x^p \leq x^{m+1} \leq 1 + x^{m+1}$.

On a bien démontré que : $\forall x \in \mathbb{R}_+, 0 \leq x^p \leq 1 + x^{m+1}$.

- ii. De ce qui précède on déduit que pour la variable aléatoire X positive introduite plus haut, $0 \leq X^p \leq 1 + X^{m+1}$; comme $E(X^{m+1})$ existe, alors $E(1 + X^{m+1})$ existe également et par majoration (et croissance de l'espérance), on en déduit que $E(X^p)$ existe.

On utilise ici une propriété très générale pas tout à fait dans le programme officiel : il aurait sûrement fallu préciser si X était une variable aléatoire discrète ou à densité, car dans chacun des deux cas on disposait d'un critère pour rédiger précisément.

Si par exemple X est une variable à densité et positive : alors d'après le théorème de transfert, $E(X^p)$ est bien défini si et seulement si l'intégrale $\int_{-\infty}^{+\infty} x^p f(x) dx$ est absolument convergente.

Comme $x \mapsto x^p f(x)$ est nulle sur $] -\infty; 0[$ et positive sur $[0; +\infty[$, cela revient à montrer la convergence simple de $\int_0^{+\infty} x^p f(x) dx$.

Or pour tout $x \geq 0$: $0 \leq x^p \leq 1 + x^{m+1} \xrightarrow{f(x) \geq 0} 0 \leq x^p f(x) \leq f(x) + x^{m+1} f(x)$.

Or $E(X^{m+1})$ existe, donc $\int_0^{+\infty} (f(x) + x^{m+1} f(x)) dx$ converge et vaut $1 + E(X^{m+1})$: par comparaison d'intégrales de fonctions continues (sauf peut-être en un nombre fini de points) et positives, l'intégrale $\int_0^{+\infty} x^p f(x) dx$ converge et $E(X^p)$ existe.

- b) Avec les fonctions $f : x \mapsto x$ et $g : x \mapsto \begin{cases} x^m & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$, bien définies et croissantes sur \mathbb{R} :

d'après ce qui précède, et puisque X est une variable aléatoire à valeurs positives, $E(f(X)) = E(X)$, $E(g(X)) = E(X^m)$ et $E(f(X)g(X)) = E(X^{m+1})$ sont bien définies, donc le résultat de la question 6.c) s'applique, qui s'écrit dans ce cas :

$$E(X^{m+1}) \geq E(X)E(X^m)$$

- c) On a dans cette question un problème de définition : travaille-t-on toujours avec une variable à densité comme dans la question 5., à l'issue de laquelle pourtant l'énoncé définissait plus généralement la loi de X biaisée par la taille ? Le problème est que la fonction $x \mapsto x^m$ n'est sûrement pas bornée sur \mathbb{R} ...

Rédigeons donc cette question en supposant que X est une variable aléatoire positive dont la densité f a les propriétés annoncées au début de la question 5.

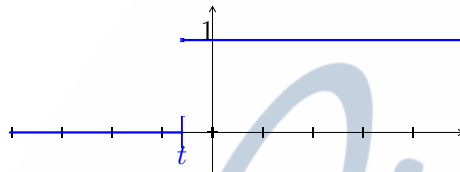
D'après le théorème de transfert, $E((X^*)^m)$ existe si et seulement si l'intégrale

$\int_{-\infty}^{+\infty} x^m g(x) dx = \int_{-\infty}^{+\infty} \frac{1}{E(X)} x^{m+1} f(x) dx$, est absolument convergente. Comme il s'agit, à un facteur $\frac{1}{E(X)}$ près, de l'intégrale qui définit $E(X^{m+1})$ qui existe, cette intégrale est bien absolument convergente, et X^* admet un moment d'ordre m qui vérifie :

$$E((X^*)^m) = \frac{E(X^{m+1})}{E(X)} \geq E(X^m) \quad \text{d'après la question précédente}$$

8. Pour A un événement, on note $\mathbb{1}_A$ la variable aléatoire définie par $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbb{1}_A(\omega) = 0$ sinon. Pour tout réel t , on définit la fonction $g_t : x \mapsto \mathbb{1}_{]t;+\infty[}(x)$.

a) La fonction $g_t : x \mapsto \begin{cases} 0 & \text{si } x \leq t \\ 1 & \text{si } x > t \end{cases}$ est représentée ci-dessous (t est arbitrairement choisi) :



Il s'agit bien d'une fonction croissante sur \mathbb{R} : prouvons-le en prenant deux réels x et y quelconques tels que $x \leq y$. Il y a alors trois possibilités :

- soit $x \leq y < t$, alors $g_t(x) = 0 \leq 0 = g_t(y)$
- soit $x < t \leq y$, alors $g_t(x) = 0 < 1 = g_t(y)$
- soit $t \leq x \leq y$, alors $g_t(x) = 1 \leq 1 = g_t(y)$

On a donc bien : $\forall (x, y) \in \mathbb{R}^2, \quad x \leq y \implies g_t(x) \leq g_t(y)$.

b) Soit X une variable aléatoire positive admettant une espérance. Cette fois la fonction g_t est bien bornée sur \mathbb{R} (par 0 et 1) en plus d'être croissante, et continue sur \mathbb{R} sauf au point t . On sait donc que $E(Xg_t(X))$ existe, de même que $E(X)$, ainsi que $E(g_t(X))$; le résultat de 6.c) s'applique avec les fonctions $x \mapsto x$ et g_t toutes deux croissantes sur \mathbb{R} , qui donne :

$$E(Xg_t(X)) \geq E(X).E(g_t(X))$$

Or $E(g_t(X)) = \mathbb{P}(X > t)$: en effet, $g_t(X(\omega)) = 1$ si $X(\omega) > t$ et 0 sinon, donc $g_t(X)$ est la variable de Bernoulli de succès $[X > t]$, et par conséquent son espérance est la probabilité de succès : $E(g_t(X)) = \mathbb{P}(X > t)$.

c) Comme de même, $E(g_t(X^*)) = \mathbb{P}(X^* > t)$, la définition posée à la fin de la question 5. permet d'écrire :

$$E(g_t(X^*)) = \frac{1}{E(X)} E(Xg_t(X)) \geq \frac{1}{E(X)} . E(X) . \mathbb{P}(X > t) \implies \mathbb{P}(X^* > t) \geq \mathbb{P}(X > t)$$

Mais là encore les problèmes de définition persistent... Ce résultat signifie en tout cas que X^* domine stochastiquement X .

9. Soient X_1, X_2, \dots, X_n des variables aléatoires positives, indépendantes, non nécessairement de même loi. On suppose qu'elles admettent toutes une espérance strictement positive, et on note $\mu_i = E(X_i)$.

De plus, on pose $\mu = \sum_{i=1}^n \mu_i$ et $S_n = \sum_{i=1}^n X_i$.

a) Puisque les X_i admettent chacune une espérance, la linéarité de l'espérance permet d'affirmer que S_n admet une espérance qui vaut :

$$E(S_n) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \mu_i = \mu$$

b) Soit J une variable aléatoire à valeurs dans $\{1, \dots, n\}$, de loi $\mathbb{P}(J = k) = \mu_k/\mu$.

Si les variables aléatoires X_i sont de même loi, alors elles ont aussi la même espérance :

$\forall k \in \llbracket 1, n \rrbracket$, $\mu_k = \mu_1$ (par exemple), mais alors $\mu = n \cdot \mu_1$ et $\mathbb{P}(J = k) = \frac{\mu_1}{n\mu_1} = \frac{1}{n}$: comme on pouvait s'y attendre, J suit dans ce cas la loi uniforme sur $\{1, \dots, n\}$.

On considère X_1^*, \dots, X_n^* des variables aléatoires indépendantes, indépendantes de X_1, \dots, X_n , telles que, pour tout $1 \leq i \leq n$, X_i^* suit la loi de X_i biaisée par la taille.

Soit aussi J une variable aléatoire de loi $\mathbb{P}(J = k) = \mu_k/\mu$, indépendante de $X_1, X_1^*, \dots, X_n, X_n^*$.

On considère la variable aléatoire $X_J = \sum_{j=1}^n X_j \mathbb{1}_{[J=j]}$ et on définit $T_n = S_n - X_J + X_J^*$. Autrement

dit, on choisit un indice aléatoire J et, dans la somme $\sum_{i=1}^n X_i$, on remplace X_J par X_J^* .

c) Soit $h : [0; +\infty[\rightarrow \mathbb{R}$ une fonction bornée et continue sauf éventuellement en un nombre fini de points.

i. Dans la somme $\sum_{i=1}^n \mathbb{1}_{[J=i]}$, un et un seul des termes est égal à 1 puisque par définition d'une variable aléatoire, un seul des événements $[J = i]$ pour $1 \leq i \leq n$, est réalisé. Cette somme est donc constamment égale à 1, et on peut effectivement écrire :

$$h(T_n) = h(T_n) \sum_{i=1}^n \mathbb{1}_{[J=i]} = \sum_{i=1}^n h(T_n) \mathbb{1}_{[J=i]} = \sum_{i=1}^n h(S_n - X_J + X_J^*) \mathbb{1}_{[J=i]} = \sum_{i=1}^n h(S_n - X_i + X_i^*) \mathbb{1}_{[J=i]}$$

puisque en toute rigueur, et selon le principe expliqué dans l'énoncé,

$$h(S_n - X_J + X_J^*) = \sum_{k=1}^n h(S_n - X_k + X_k^*) \mathbb{1}_{[J=k]}$$

$$\text{et donc } h(S_n - X_J + X_J^*) \mathbb{1}_{[J=i]} = \sum_{k=1}^n h(S_n - X_k + X_k^*) \mathbb{1}_{[J=k]} \cdot \mathbb{1}_{[J=i]}$$

$$\text{où : } \mathbb{1}_{[J=k]} \cdot \mathbb{1}_{[J=i]} = \begin{cases} 1 = \mathbb{1}_{[J=i]} & \text{si } i = k \\ 0 & \text{si } i \neq k \end{cases}, \text{ ce qui signifie que seul le terme pour } k = i$$

subsiste de la dernière somme, d'où le résultat voulu.

ii. Par linéarité de l'espérance (toutes les espérances existent car h est bornée), on en déduit :

$$E(h(T_n)) = \sum_{i=1}^n E(h(S_n - X_i + X_i^*) \mathbb{1}_{[J=i]}) \stackrel{(1)}{=} \sum_{i=1}^n E(\mathbb{1}_{[J=i]}) \cdot E(h(S_n - X_i + X_i^*)) \stackrel{(2)}{=} \sum_{i=1}^n \mathbb{P}(J = i) E(h(S_n - X_i + X_i^*))$$

(1) : par indépendance de J vis-à-vis des X_i et X_i^* (et donc, par le lemme des coalitions, $\mathbb{1}_{[J=i]}$ et $h(S_n - X_i + X_i^*)$ sont indépendantes).

(2) : parce que comme on a déjà eu l'occasion de le voir, $E(\mathbb{1}_A) = \mathbb{P}(A)$ pour tout événement A , donc $E(\mathbb{1}_{[J=i]}) = \mathbb{P}(J = i)$.

d) Pour $i \in \{1, \dots, n\}$ et tout réel s : la fonction $k : x \mapsto h(s + x)$ est bornée et continue sur \mathbb{R} sauf peut-être en un nombre fini de points, car c'est déjà le cas de la fonction h .

On peut donc écrire, en application de ce qui a été définie pour la loi biaisée par la taille :

$$E(h(s + X_i^*)) = E(k(X_i^*)) = \frac{1}{E(X_i)} E(X_i k(X_i)) = \frac{1}{\mu_i} E(X_i k(X_i)) = \frac{1}{\mu_i} E(X_i h(s + X_i))$$

L'énoncé admettait alors que ceci permettait d'en déduire l'égalité

$$E(h(S_n - X_i + X_i^*)) = \frac{1}{\mu_i} E(X_i h(S_n)).$$

e) En combinant les deux résultats précédents, on déduit :

$$\begin{aligned}
 E(h(T_n)) &= \sum_{i=1}^n \mathbb{P}(J = i) \cdot \frac{1}{\mu_i} E(X_i h(S_n)) \\
 &\stackrel{(1)}{=} E\left(\sum_{i=1}^n X_i \cdot \mathbb{P}(J = i) \cdot \frac{1}{\mu_i} h(S_n)\right) \stackrel{(2)}{=} E\left(h(S_n) \sum_{i=1}^n X_i \cdot \frac{1}{\mu}\right) \\
 &\stackrel{(3)}{=} \frac{1}{\mu} E(S_n h(S_n)) = \frac{E(S_n h(S_n))}{E(S_n)}
 \end{aligned}$$

(1) : par linéarité de l'espérance

(2) : puisque $\mathbb{P}(J = i) = \frac{\mu_i}{\mu}$, alors $\mathbb{P}(J = i) \cdot \frac{1}{\mu_i} = \frac{1}{\mu}$.

(3) : par linéarité de l'espérance $\frac{1}{\mu} = \frac{1}{E(S_n)}$ étant une constante, et parce que $S_n = \sum_{i=1}^n X_i$.

f) Le résultat précédent étant vrai pour toute fonction h bornée, continue sur \mathbb{R} sauf peut-être en un nombre fini de points : T_n correspond, d'après la définition donnée à la fin de la question 5., à une variable aléatoire qui suit la loi de S_n biaisée par la taille.

Troisième partie : Applications en Statistique

On s'intéresse maintenant au cas où le biais par la taille peut être utilisé en statistique, pour construire des estimateurs non biaisés. Une compagnie d'électricité possède n clients ($n \in \mathbb{N}^*$ donné). Lors de n années écoulées, le i^e client a payé x_i euros ($x_i > 0$), mais a en réalité consommé une quantité d'électricité correspondant à y_i euros ($y_i > 0$). La compagnie sait combien ses clients ont payé, et elle souhaite estimer le rapport

$$\theta = \left(\sum_{i=1}^n y_i\right) / \left(\sum_{i=1}^n x_i\right),$$

pour déterminer à quel point elle a mal facturé ses clients.

10. Soit m un entier fixé tel que $1 \leq m \leq n$. On note \mathcal{P}_m l'ensemble des parties $A \subset \{1, \dots, n\}$ de cardinal m : on sait que leur nombre est égal à $\binom{n}{m}$, et on considère alors une variable aléatoire R , à valeurs dans \mathcal{P}_m et de loi uniforme, c'est-à-dire que chacune de ces parties qui sont les combinaisons d'ordre m parmi n , soit choisie avec la même probabilité : pour toute partie A de \mathcal{P}_m , $\mathbb{P}(R = A) = \frac{1}{\binom{n}{m}}$.

On souhaite écrire un programme pour choisir l'ensemble R au hasard.

Remarque : on travaille donc ici avec une variable aléatoire R qui est à valeurs dans un ensemble de parties, et pas dans \mathbb{R} : toute cette partie est donc totalement hors-programme...

a) On considère la procédure suivante : on prend un premier élément s_1 uniformément dans $\{1, \dots, n\}$, puis un deuxième élément s_2 uniformément dans $\{1, \dots, n\} \setminus \{s_1\}$, etc... puis un m -ième élément s_m uniformément dans $\{1, \dots, n\} \setminus \{s_1, \dots, s_{m-1}\}$. On note $S = (s_1, \dots, s_m)$, qui est un m -uplet aléatoire.

i. Au vu de la procédure de choix des m éléments de S , on a pour tout m -uplet (a_1, \dots, a_m) d'éléments distincts de $\{1, \dots, n\}$:

$$\begin{aligned}
 \mathbb{P}(S = (a_1, \dots, a_m)) &= \mathbb{P}([s_1 = a_1] \cap \dots \cap [s_m = a_m]) \\
 &= \mathbb{P}(s_1 = a_1) \times \mathbb{P}_{[s_1 = a_1]}(s_2 = a_2) \times \dots \times \mathbb{P}_{[s_1 = a_1] \cap \dots \cap [s_{m-1} = a_{m-1}]}(s_m = a_m) \\
 &= \frac{1}{n} \times \frac{1}{n-1} \times \dots \times \frac{1}{n-(m-1)} = \frac{1}{(n-m+1)\dots(n-1)n} = \frac{(n-m)!}{n!}
 \end{aligned}$$

ii. On note $R = \{s_1, \dots, s_m\}$ l'ensemble des entiers tirés lors de la procédure décrite plus haut (l'ordre dans lesquels ils ont été tirés n'importe plus).

Tout sous-ensemble $A = \{a_1, \dots, a_m\} \subset \{1, \dots, n\}$ dans lequel l'ordre ne compte pas, peut voir ses éléments permutés de toutes les façons possibles pour en faire à chaque fois un m -uplet $S = (b_1, \dots, b_m)$ où cette fois l'ordre compte ; par exemple, le sous-ensemble $\{1, 3, 4\}$ donne les triplets $(1, 3, 4), (1, 4, 3), (3, 1, 4), (3, 4, 1), (4, 1, 3), (4, 3, 1)$.

Pour chaque partie A de $\{1, \dots, n\}$ à m éléments, il y a $m!$ façons de permuter les éléments de cette partie, et donc autant de m -uplets qui ne correspondent qu'à cette partie.

La probabilité que le m -uplet tiré donne lieu à la partie A lorsqu'on ne tient plus compte de l'ordre, est donc égale à :

$$m! \times \frac{(n-m)!}{n!} = \frac{m!(n-m)!}{n!} = \frac{1}{\binom{n}{m}}$$

L'ensemble R a donc bien été choisi uniformément dans \mathcal{P}_m .

b) Soit U une variable aléatoire suivant la loi uniforme (à densité) sur $[0; 1[$: alors nU est à valeurs dans $[0; n[$, et $X = 1 + \lfloor nU \rfloor$ est à valeurs entières, comprises entre $1 + 0 = 1$ et $1 + (n-1) = n$.

Pour tout entier $j \in \{1, \dots, n\}$:

$$\mathbb{P}(X = j) = \mathbb{P}(\lfloor nU \rfloor = j - 1) = \mathbb{P}(j - 1 \leq nU < j) = \mathbb{P}\left(\frac{j-1}{n} \leq U < \frac{j}{n}\right)$$

Or $\left[\frac{j-1}{n}; \frac{j}{n}\right]$ est un intervalle de longueur $\frac{1}{n}$ entièrement inclus dans $[0; 1[$, donc :

$$\mathbb{P}\left(\frac{j-1}{n} \leq U < \frac{j}{n}\right) = \frac{1}{n} = \mathbb{P}(X = j)$$

et ce pour tout entier j de $\{1, \dots, n\}$, ce qui prouve bien que $X = 1 + \lfloor nU \rfloor$ suit la loi uniforme discrète sur $\{1, \dots, n\}$.

On en déduit alors la fonction suivante, qui est en fait très classiquement utilisée dès la ECE1, avant même d'avoir pu donner toutes ces justifications théoriques :

```

1  function x = Uniforme(n)
2      x = 1+floor(n*rand())
3  endfunction

```

c) Le script suivant définit une fonction **Selection** qui prend en argument un vecteur V et renvoie un élément x de V pris au hasard avec équiprobabilité parmi tous les éléments de V , ainsi que le vecteur W , qui est le vecteur V auquel on a retiré l'élément x .

```

1  function [x,W] = Selection(V)
2      n = length(V)
3      i = Uniforme(n)
4      x = V(i)
5      W = [V(1:(i-1)), V((i+1):n)]
6  endfunction

```

Dans ce script, $V(1:(i-1))$ est la tranche du vecteur V formée des éléments $V(1), \dots, V(i-1)$, qu'on concatène avec la tranche formée des éléments $V(i+1), \dots, V(n)$: le résultat est bien le vecteur V auquel on a retiré $x=V(i)$ choisi au hasard grâce à son indice.

Pour utiliser cette fonction et bien récupérer les deux variables de sortie, il faudra écrire dans la console, une syntaxe du type :

```
[x,W] = Selection([3 7 9 11 4 31])
```

qui rendra par exemple : $\mathbf{x} = 11$ et $W = [3 \ 7 \ 9 \ 4 \ 31]$.

d) Le programme suivant prend alors en argument deux entiers n et m avec $m \leq n$, et renvoie un vecteur R de m entiers distincts, pris uniformément dans $\{1, \dots, m\}$:

```

1  function R = Choix(n,m)
2      V = 1:n
3      R = []
4      for i = 1:m
5          [x,W] = Selection(V)
6          R = [R,x]
7          V = W
8      end
9  endfunction

```

Il suffit d'utiliser m fois de suite la fonction `Selection` : l'élément extrait du vecteur est à chaque fois rajouté au vecteur R , et V est amputé de x avant la sélection suivante.

11. Pour une partie $A \in \mathcal{P}_m$, on définit $\bar{x}_A = \frac{1}{m} \sum_{i \in A} x_i$, $\bar{y}_A = \frac{1}{m} \sum_{i \in A} y_i$ (moyennes arithmétiques respectivement des montants payés et des montants dûs par les clients appartenant à l'ensemble A), et aussi $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

La compagnie décide d'utiliser $\theta_R = \bar{y}_R / \bar{x}_R$ comme estimateur de θ .

a) On définit deux variables aléatoires $X = \bar{x}_R = \frac{1}{m} \sum_{i \in R} x_i$ et $Y = \bar{y}_R = \frac{1}{m} \sum_{i \in R} y_i$, qui correspondent aux montants moyens payés et consommés par les m clients du groupes tiré au hasard.

i. La variable aléatoire $X = \bar{x}_R$ est ici vue comme une fonction de la variable aléatoire R dont l'univers-image \mathcal{P}_m est fini : le théorème de transfert, en version adaptée à la variable aléatoire non réelle R s'applique, qui donne :

$$E(X) = E(\bar{x}_R) = \sum_{A \in \mathcal{P}_m} \bar{x}_A \mathbb{P}(R = A) = \binom{n}{m}^{-1} \sum_{A \in \mathcal{P}_m} \bar{x}_A$$

puisque $\mathbb{P}(R = A) = \binom{n}{m}^{-1}$ ne dépend pas de la partie A choisie.

ii. Soit $1 \leq i \leq n$ un entier naturel : toute partie A de \mathcal{P}_m qui contient i , contient aussi $m - 1$ autres éléments pris parmi l'ensemble $\{1, \dots, n\} \setminus \{i\}$: il y a donc $\binom{n-1}{m-1}$ parties $A \in \mathcal{P}_m$ telles que $i \in A$.

iii. Dans la double somme $\sum_{A \in \mathcal{P}_m} \sum_{i \in A} x_i$, chaque indice $i \in \{1, \dots, n\}$ est présent autant de fois qu'il existe une partie $A \in \mathcal{P}_m$ telle que $i \in A$. On vient de voir que cela représente $\binom{n-1}{m-1}$ parties A différentes, ce qui ne dépend pas de la valeur de i ; ainsi, chaque x_i est additionné $\binom{n-1}{m-1}$ fois exactement, et par conséquent on a bien :

$$\sum_{A \in \mathcal{P}_m} \sum_{i \in A} x_i = \sum_{i=1}^n \binom{n-1}{m-1} x_i = \binom{n-1}{m-1} \sum_{i=1}^n x_i.$$

iv. La double somme précédente se réécrit, au vu des définitions posées au début de la question 11 :

$$\sum_{A \in \mathcal{P}_m} \sum_{i \in A} x_i = \sum_{A \in \mathcal{P}_m} m \bar{x}_A \stackrel{11.a).i.}{=} m \binom{n}{m} E(X) \iff m \binom{n}{m} E(X) = \binom{n-1}{m-1} \underbrace{\sum_{i=1}^n x_i}_{=n\bar{x}}$$

$$\iff \binom{n}{m} E(X) = \underbrace{\frac{n}{m} \binom{n-1}{m-1}}_{=\binom{n}{m}: \text{formule sans nom}} \cdot \bar{x} \iff E(X) = \bar{x}$$

L'énoncé admettait de même que $E(Y) = \bar{y}$.

v. Par définition : $\theta = \left(\sum_{i=1}^n y_i \right) / \left(\sum_{i=1}^n x_i \right) = \frac{n\bar{y}}{n\bar{x}} = \frac{\bar{y}}{\bar{x}} = \frac{E(Y)}{E(X)}$.

b) Y a-t-il un piège ? L'énoncé appelle $X = \bar{x}_R$ et $Y = \bar{y}_R$, donc $\theta_R = \frac{Y}{X}$ et comme il s'agit d'une variable aléatoire finie, on peut définir $E(\theta_R) = E\left(\frac{Y}{X}\right)$.

c) On donne l'inégalité de Cauchy-Schwarz : si W et Z sont deux variables aléatoires strictement positives, admettant un moment d'ordre deux, $E(WZ) \leq E(W^2)^{1/2} E(Z^2)^{1/2}$, avec égalité si et seulement si il existe un $\alpha > 0$ tel que $W = \alpha Z$.

i. Les variables aléatoires X et $1/X$ sont finies et positives, donc les variables aléatoires \sqrt{X} et $1/\sqrt{X}$ sont bien définies et admettent un moment d'ordre 2 ; l'inégalité de Cauchy-Schwarz s'applique alors avec ces deux variables aléatoires, pour donner :

$$E(\sqrt{X} \cdot 1/\sqrt{X}) \leq E(X)^{1/2} E(1/X)^{1/2} \iff E(1) \leq \sqrt{E(X)E(1/X)} \iff 1/E(X) \leq E(1/X)$$

puisque $E(X) > 0$.

ii. Le cas d'égalité a lieu, comme l'indique l'énoncé, si et seulement si il existe un réel $\alpha > 0$ tel que : $\sqrt{X} = \alpha \cdot 1/\sqrt{X} \iff X = \alpha$, c'est-à-dire si et seulement si X est une variable aléatoire constante.

Comme d'après 11.a)iv., on a toujours $E(X) = \bar{x}$, alors X est constante égale à \bar{x} .

iii. Si pour tout i , $x_i = \bar{x}$, alors quelle que soit la valeur de la partie R dans \mathcal{P}_m ,

$X = \frac{1}{m} \sum_{i \in R} x_i = \frac{1}{m} \cdot m \cdot \bar{x} = \bar{x}$, donc X est constante égale à \bar{x} et $E(1/X) = 1/E(X)$ d'après l'équivalence démontrée à la question précédente.

Réciproquement : si $E(1/X) = 1/E(X)$, alors X est constante égale à \bar{x} , et par conséquent, pour toute partie A de \mathcal{P}_m , $\bar{x}_A = \frac{1}{m} \sum_{i \in A} x_i = \bar{x}$.

Si donc i et j sont deux entiers distincts de $\{1, \dots, n\}$, et B une partie à $m-1$ éléments de $\{1, \dots, n\} \setminus \{i, j\}$: alors $B \cup \{i\}$ et $B \cup \{j\}$ appartiennent à \mathcal{P}_m , donc :

$$\frac{1}{m} \sum_{k \in B \cup \{i\}} x_k = \frac{1}{m} \sum_{k \in B \cup \{j\}} x_k \iff \sum_{k \in B} x_k + x_i = \sum_{k \in B} x_k + x_j \iff x_i = x_j$$

et on prouve ainsi que les $(x_i)_{1 \leq i \leq n}$ sont tous égaux : chacun d'eux est alors aussi égal à la moyenne \bar{x} de ces n nombres égaux.

On a donc bien démontré l'équivalence : $E(1/X) = 1/E(X) \iff x_i = \bar{x}$ pour tout i compris entre 1 et n .

d) Si on suppose que X et Y sont indépendantes : alors par le lemme des coalitions, Y et $1/X$ le sont aussi, et on peut écrire :

$$E(\theta_R) = E\left(\frac{Y}{X}\right) = E(Y)E(1/X) \geq E(Y) \cdot 1/E(X) \implies E(\theta_R) \geq E(Y)/E(X) = \theta$$

d'après le résultat de 11.c)i., et parce que $E(Y) > 0$.

D'après 11.c)ii. et iii., il y a égalité si et seulement si $E(1/X) = 1/E(X)$, donc si et seulement si $x_i = \bar{x}$ pour tout i .

Ainsi, $E(\theta_R)$ n'est pas forcément égal à θ : on dit que θ_R est un estimateur biaisé de θ .

12. a) Si A est une partie de \mathcal{P}_m : pour que $[R = A]$ soit réalisé, c'est-à-dire que les m clients choisis forment exactement le groupe A , avec la nouvelle procédure décrite, il a fallu choisir d'abord l'un des m clients, correspondant à un entier $i \in A$, puis le choix des $m - 1$ clients suivants devait correspondre au groupe $A \setminus \{i\}$, c'est-à-dire qu'on a l'égalité d'événements :

$$A = \bigcup_{i \in A} ([J = i] \cap [V = A \setminus \{i\}])$$

L'union est disjointe à cause du choix particulier et unique du premier client, donc :

$$\mathbb{P}(R = A) = \sum_{i \in A} \mathbb{P}([J = i] \cap [V = A \setminus \{i\}]) = \sum_{i \in A} \mathbb{P}(J = i) \times \mathbb{P}_{[J=i]}(V = A \setminus \{i\})$$

- b) Les hypothèses faites au début de la question 12. permettent de réécrire cette probabilité sous la forme :

$$\mathbb{P}(R = A) = \sum_{i \in A} \frac{x_i}{n\bar{x}} \cdot \frac{1}{\binom{n-1}{m-1}} = \frac{1}{n \binom{n-1}{m-1} \bar{x}} \sum_{i \in A} x_i = \frac{1}{m \binom{n}{m} \bar{x}} \cdot m\bar{x}_A = \frac{1}{\binom{n}{m}} \cdot \frac{\bar{x}_A}{\bar{x}}$$

On a notamment réutilisé la formule sans nom : $m \binom{n}{m} = n \binom{n-1}{m-1}$.

13. Une fois choisi le groupe de clients R (par la procédure de la question 12), on définit $\hat{\theta}_R = \bar{y}_R / \bar{x}_R$.

- a) La variable aléatoire $\hat{\theta}_R$ est finie, donc admet une espérance qui vaut (toujours d'après le théorème de transfert pour la variable aléatoire R) :

$$E(\hat{\theta}_R) = \sum_{A \in \mathcal{P}_m} \frac{\bar{y}_A}{\bar{x}_A} \mathbb{P}(R = A) = \sum_{A \in \mathcal{P}_m} \frac{\bar{y}_A}{\bar{x}_A} \cdot \frac{1}{\binom{n}{m}} \frac{\bar{x}_A}{\bar{x}} = \frac{1}{\binom{n}{m}} \sum_{A \in \mathcal{P}_m} \frac{\bar{y}_A}{\bar{x}}$$

(le coefficient $\binom{n}{m}$ ne varie pas en fonction de $A \in \mathcal{P}_m$).

- b) En reprenant alors le résultat de 11.a)i. dont l'énoncé admettait qu'il s'applique aussi à Y , on peut écrire : $E(Y) = \frac{1}{\binom{n}{m}} \sum_{A \in \mathcal{P}_m} \bar{y}_A = \bar{y}$, et donc la formule précédente se réécrit :

$$E(\hat{\theta}_R) = \frac{1}{\bar{x}} \cdot \frac{1}{\binom{n}{m}} \sum_{A \in \mathcal{P}_m} \bar{y}_A = \frac{\bar{y}}{\bar{x}} = \theta$$

On a donc construit un estimateur non biaisé de θ .

★★★ FIN DU SUJET ★★★