
Concours Blanc 3

Code sujet : 287

Conception : ESSEC BS

OPTION ÉCONOMIQUE

MATHÉMATIQUES II

Jeudi 30 avril 2020, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Aucun document n'est autorisé. L'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

Lorsque l'on effectue des sondages, de nombreux biais statistiques peuvent apparaître : on peut par exemple avoir considéré un échantillon non-représentatif de la population, il peut y avoir un biais dans les réponses des personnes sondées... On va s'intéresser dans ce problème à ce que l'on appelle le biais par la taille : il provient du fait que si l'on choisit une personne au hasard dans la population, celle-ci a plus de chances de faire partie d'une catégorie nombreuse de la population.

Le biais par la taille est la source de nombreux "paradoxes" probabilistes, comme le fait que les gagnants du loto vivent en moyenne plus longtemps (parce que les gagnants sont ceux qui ont pu jouer au loto plus longtemps) ou le fait que vos amis ont en moyenne plus d'amis que vous (car les gens qui ont un très grand nombre d'amis font sûrement partie de vos amis). On verra ici comment formaliser le biais par la taille, et l'utiliser dans différents contextes.

Toutes les variables aléatoires intervenant dans le problème sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

Pour toute variable aléatoire X , on note $E(X)$ son espérance (resp. $\text{Var}(X)$ sa variance) lorsqu'elles existent.

Première partie : 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 va voir que cette approche introduit un biais, en considérant une situation particulière. Supposons que X suive la loi binomiale de paramètres $n = 10$ et $p = 1/5$. On note $p_k = P(X = k)$ pour $k \in \{0, 1, \dots, 10\}$.

- a)
 - i) Rappeler l'expression de p_k pour $k \in \{0, 1, \dots, 10\}$.
 - ii) Que vaut $E(X)$?
 - iii) Donner $\text{Var}(X)$, et en déduire $E(X^2)$.
- b) Soit M_k le nombre de familles 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 de la population.
 - i) Montrer que $N_k = kp_k M$.
 - ii) Montrer que $N/M = 2$.
 - iii) Montrer que la proportion des enfants provenant d'une famille à k enfants est $p_k^* = kp_k/2$.

- 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 élément de $\{1, 2, \dots, 10\}$, montrer que $P(Y = k) = kp_k/2$.
 - ii) Montrer que $E(Y) = E(X^2)/E(X)$.
 - iii) En déduire $E(Y)$ et le comparer à $E(X)$.

- 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

$$q_i = \frac{i}{E(X)} P(X = i).$$

- a) Calculer $\sum_{i=1}^{\infty} q_i$.

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 , c'est-à-dire que pour tout entier $i > 0$ non nul

$$P(X^* = i) = \frac{i}{E(X)} P(X = i).$$

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

- b) On suppose que X admet un moment d'ordre 2. Montrer que $E(X^*) = E(X^2)/E(X)$.
- c) En déduire que si $E(X^2)$ existe, on a

$$\text{Var}(X) = E(X)(E(X^*) - E(X)).$$

- d) Conclure que $E(X^*) \geq E(X)$.

- 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) Donner la loi de X^* .
 - ii) Vérifier que X^* suit la même loi que $X + 1$.
- 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) Montrer que pour tout $k \geq 1$,

$$P(X = k) = \frac{E(X)}{k} P(X = k - 1).$$

- ii) Montrer que pour tout k entier naturel,

$$P(X = k) = \frac{E(X)^k}{k!} P(X = 0).$$

- iii) En déduire la loi de X .

- 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 $1 \leq k \leq n$, $P(X = k) > 0$. 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 pour tout entier k élément de $\{1, \dots, n\}$, $P_{|X=k}(T = j) = 1/k$ si $j \in \{1, \dots, k\}$ et $P_{|X=k}(T = j) = 0$ si $j > k$.
- i) Montrer que pour tout entier $k \in \{1, \dots, n\}$ on a

$$\sum_{j=1}^n j P_{|X=k}(T = j) = \frac{k+1}{2}.$$

- ii) En déduire que

$$\sum_{k=1}^n \sum_{j=1}^n j P(X = k) P_{|X=k}(T = j) = \frac{E(X+1)}{2}.$$

iii) Montrer que

$$E(T) = \sum_{j=1}^n \sum_{k=1}^n j P(X=k) P_{|X=k}(T=j).$$

iv) Montrer que

$$E(T) = \frac{E(X+1)}{2}.$$

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 pour tout $k \in \{1, \dots, n\}$, $P_{|X^*=k}(T^*=j) = 1/k$ si $j \in \{1, \dots, k\}$ et $P_{|X^*=k}(T^*=j) = 0$ si $j > k$.

i) Montrer que pour tout entier $k \in \{1, \dots, n\}$ on a

$$\sum_{j=1}^n j P_{|X^*=k}(T^*=j) = \frac{k+1}{2}.$$

ii) Montrer que

$$E(T^*) = \sum_{j=1}^n \sum_{k=1}^n j P(X^*=k) P_{|X^*=k}(T^*=j).$$

iii) Montrer que

$$E(T^*) = \frac{E(X^*+1)}{2}.$$

iv) En déduire qu'on a $E(T^*) \geq E(T)$.

Deuxième partie : Biais par la taille, propriétés

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$. On dit que f est une *densité* si $f \geq 0$, que f est continue par morceaux, et que $\int_{-\infty}^{\infty} f(t) dt = 1$.

Si f est densité, on dit que f admet une espérance, noté $E(f)$, si l'intégrale ci-dessous converge absolument :

$$E(f) = \int_0^{\infty} t f(t) dt.$$

Enfin, si H est une fonction et que f est une densité, on définit, lorsque l'intégrale ci-dessous converge absolument,

$$E(H(f)) = \int_{-\infty}^{+\infty} f(t) H(t) dt.$$

C'est le théorème de transfert.

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

$$g(x) = \frac{x}{E(f)} f(x)$$

pour tout x réel.

a) Montrer que g définit une densité.

On dit que g est la densité de f biaisée par la taille, et on notera par la suite $g = f^*$.

b) Soit α un réel strictement positif.

i) Montrer que la fonction $f_{\alpha} x \mapsto \frac{1}{\alpha} f\left(\frac{x}{\alpha}\right)$ est une densité.

ii) Démontrer que $x \rightarrow \frac{1}{\alpha} g\left(\frac{x}{\alpha}\right)$ et $(f_{\alpha})^*$ sont identiques.

c) Une propriété importante. Soit $H : [0, +\infty[\rightarrow \mathbb{R}$ une fonction bornée et continue sauf éventuellement en un nombre fini de points. Montrer que $E(fH(f))$ est bien définie et que

$$E(H(f^*)) = \frac{1}{E(f)} E(fH(f)).$$

On pose alors la définition suivante si f est une densité positive d'espérance $E(f)$ strictement positive, on dit que la densité g suit la loi de f biaisée par la taille si on a

$$E(H(g)) = \frac{1}{E(f)} E(fH(f))$$

pour toute fonction $H : [0, +\infty[\rightarrow \mathbb{R}$ bornée et continue sauf éventuellement en un nombre fini de points. On admet que cette définition coïncide dans le cas d'une variable aléatoire discrète : si X est une variable aléatoire discrète, alors X^* est la loi de X biaisée par la taille si, et seulement si, pour tout $h : \mathbb{R} \rightarrow \mathbb{R}$ bornée,

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

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

a) Montrer que quels que soient les réels x_1 et x_2 , on a

$$(\phi(x_1) - \phi(x_2))(\psi(x_1) - \psi(x_2)) \geq 0.$$

b) On admet l'existence de l'intégrale double ci-dessous. *Le candidat, s'il utilise la linéarité de l'intégrale, fera attention à écrire des intégrales convergentes.* Montrer que

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} (\phi(t) - \phi(x))(\psi(t) - \psi(x))f(t)f(x)dt \right) dx = 2E(\phi(f)\psi(f)) - 2E(\phi(f))E(\psi(f)).$$

c) En déduire que

$$E(\phi(f)\psi(f)) \geq E(\phi(f))E(\psi(f)).$$

- 7) Dans cette question, on suppose que f est une densité d'espérance strictement positive, et telle que $E(f^{m+1})$ existe pour un entier $m \geq 1$ donné. *Cela revient à appliquer la formule de transfert avec $H : X \rightarrow X^{m+1}$.*

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

i) Montrer que pour tout réel $x \geq 0$, on a

$$0 \leq x^p \leq 1 + x^{m+1}.$$

ii) Montrer que $E(f^p)$ existe.

b) Montrer que

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

c) En déduire que

$$E((f^*)^m) \geq E(f^m).$$

- 8) Pour A un événement, on note I_A la variable aléatoire définie par $I_A(\omega) = 1$ si $\omega \in A$ et $I_A(\omega) = 0$ sinon. Pour tout t réel, on définit la fonction

$$g_t(x) = I_{[t, +\infty[}(x).$$

a) Montrer que la fonction $x \mapsto g_t(x)$ est croissante sur $\mathbb{R}$.

b) Soit f une densité positive (i.e. $f(x) = 0$ lorsque $x \leq 0$), admettant une espérance. Montrer que pour tout t réel, $E(fg_t(f))$ est bien définie et que

$$E(fg_t(f)) \geq E(f)E(g_t(f)).$$

c) Montrer que pour tout t réel,

$$E(g_t(f^*)) \geq E(g_t(f)).$$

On dit que f^* domine stochastiquement f .¹

- 9) *Retour au cas discret.* On admet que tous les résultats ci-dessus restent valables en remplaçant f par X une variable aléatoire discrète.

Soient $X_1, \dots, X_n$ des variables aléatoires positives, discrètes, 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

$$\mu = \sum_{i=1}^n \mu_i, \quad S_n = \sum_{i=1}^n X_i.$$

1. On vient de démontrer que $P(X^* \geq t) \geq P(X \geq t)$ en langage probabiliste.

- a) Montrer que $E(S_n) = \mu$.
b) Soit J une variable aléatoire à valeurs dans $\{1, \dots, n\}$, de loi $P(J = k) = \mu_k/\mu$. Quelle est la loi de J si les variables X_i sont toutes de même loi ?

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

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

$$X_J = \sum_{j=1}^n X_j I_{[J=j]}$$

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.

- i) Montrer que

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

- ii) En déduire que

$$E(h(T_n)) = \sum_{i=1}^n P(J = i) E(h(S_n - X_i + X_i^*)).$$

- d) Pour $i \in \{1, \dots, n\}$, montrer que pour tout réel s ,

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

On admettra qu'on en déduit l'égalité

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

- e) En déduire que

$$E(h(T_n)) = \frac{E(S_n h(S_n))}{E(S_n)}.$$

- f) Conclure que T_n suit la loi de S_n biaisée par la taille.

Troisième partie : Applications en Statistique

Le paragraphe ci-dessous n'est pas utile : il n'y a rien d'utile pour la question 10. Il servait à introduire l'objectif de la partie.

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 où n est un entier naturel non nul donné. Lors de l'année écoulée, 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 considère une variable aléatoire R , à valeurs dans $\mathcal{P}_m$ et de loi uniforme, c'est-à-dire telle que pour toute partie $A \in \mathcal{P}_m$, $P(R = A) = \frac{1}{\binom{n}{m}}$.

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

- 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 $\{1, \dots, n\} \setminus \{s_1\}$, ..., 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) Montrer que pour tout m -uplet $(a_1, \dots, a_m)$ d'entiers distincts de $\{1, \dots, n\}$, on a

$$P(S = (a_1, \dots, a_m)) = \frac{(n-m)!}{n!}.$$

ii) On note $R = \{s_1, \dots, s_m\}$ l'ensemble des entiers tirés lors de la procédure décrite plus haut (l'ordre dans laquelle ils ont été tirés ne joue plus). Montrer que pour tout ensemble $A = \{a_1, \dots, a_m\} \subset \{1, \dots, n\}$ de cardinal m , on a

$$P(R = A) = \frac{m!(n-m)!}{n!}.$$

En déduire que l'ensemble R a été choisi uniformément dans $\mathcal{P}_m$.

- b) Pour un réel x , on note $[x]$ sa partie entière, c'est-à-dire le plus grand entier naturel inférieur ou égal à x . Montrer que si U est un nombre pris au hasard uniformément entre $[0, 1[$, alors $X = 1 + [nU]$ suit la loi uniforme sur $\{1, \dots, n\}$.
- c) On rappelle que la fonction `random()` renvoie un nombre aléatoire de loi uniforme sur $[0, 1[$ et que `floor(x)` renvoie la partie entière de x . Écrire une fonction `Uniforme` en Python qui prend en argument un entier n et renvoie un nombre (aléatoire), uniforme sur $\{1, \dots, n\}$.

```
def Uniforme(n):  
    ...
```

- d) Écrire une fonction `Selection`, qui prend en argument un vecteur V et renvoie un élément x de V pris de manière aléatoire parmi tous les éléments de V , ainsi que le vecteur W , égal au vecteur V auquel on a enlevé l'élément x . L'instruction `len(V)` renvoie le nombre d'éléments du vecteur V .

```
def Selection(V):  
    ...
```

- e) Compléter le programme suivant, qui prend en argument deux entiers n et m avec $m \leq n$, et renvoie un vecteur R de m entiers distincts, pris uniformément dans $\{1, \dots, n\}$:

```
def Choix(m, n):  
    V = list(range(1, n + 1))  
    R = []  
    for i in range(m):  
        ...  
    return R
```

Le reste n'est pas très accessible cette année... Il restait deux questions, et 10 sous-questions.