



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2 — Option économique

2020

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Sujet

Lorsqu'on effectue des sondages, de nombreux biais statistiques peuvent apparaître : échantillon non représentatif de la population, biais dans les réponses des personnes sondées, etc. On s'intéresse ici au biais par la taille : une personne choisie au hasard dans la population a davantage de chances d'appartenir à une catégorie nombreuse. Ce biais explique divers paradoxes probabilistes, par exemple la durée moyenne de jeu des gagnants du loto ou le nombre moyen d'amis de ses amis. On étudiera sa formalisation et ses applications.

Toutes les variables aléatoires du problème sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. Pour toute variable aléatoire X , on note $E(X)$ son espérance et $V(X)$ 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 sa loi, on pourrait interroger les élèves d'une école sur le nombre d'enfants de leur famille. On va voir que cette approche introduit un biais dans une situation particulière. On suppose que X suit la loi binomiale de paramètres $n = 10$ et $p = 1/5$. On note $p_k = P(X = k)$ pour $k \in \{0, \dots, 10\}$.

1. a) i) Rappeler l'expression de p_k pour $k \in \{0, \dots, 10\}$.

1. a) ii) Que vaut $E(X)$?

1. a) iii) Donner $V(X)$ et en déduire $E(X^2)$.

1. b) Soit M_k le nombre de familles à k enfants et

$$M = \sum_{k=0}^{10} M_k$$

le nombre total de familles, de sorte que $p_k = M_k/M$. Soit N_k le nombre total d'enfants

appartenant à une famille à k enfants et

$$N = \sum_{k=0}^{10} N_k$$

le nombre total d'enfants de la population.

1. b) i) Montrer que $N_k = kp_k M$.

1. b) ii) Montrer que $N/M = 2$.

1. b) iii) Montrer que la proportion d'enfants provenant d'une famille à k enfants est $p_k^* = kp_k/2$.

1. c) On choisit une personne au hasard dans la population des enfants ainsi considérée, et on lui demande combien d'enfants ses parents ont eus, elle comprise. On note Y ce nombre d'enfants.

1. c) i) Pour tout $k \in \{1, \dots, 10\}$, montrer que $P(Y = k) = kp_k/2$.

1. c) ii) Montrer que $E(Y) = E(X^2)/E(X)$.

1. c) iii) En déduire $E(Y)$ et le comparer à $E(X)$.

2. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$, vérifiant $P(X > 0) > 0$ et admettant une espérance. Pour tout entier $i > 0$, on pose $q_i = \frac{i}{E(X)} P(X = i)$.

2. a) Calculer

$$\sum_{i=1}^{+\infty} q_i.$$

Cette suite définit donc une loi de probabilité. On considère une variable aléatoire X^* dont la loi est donnée par

$$P(X^* = i) = \frac{i}{E(X)} P(X = i), \quad i \in \mathbb{N}^*.$$

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

2. b) On suppose que X admet un moment d'ordre 2. Montrer que

$$E(X^*) = E(X^2)/E(X).$$

2. c) En déduire que, si $E(X^2)$ existe, $V(X) = E(X)(E(X^*) - E(X))$.

2. d) Sous l'hypothèse de la question b), conclure que $E(X^*) \geq E(X)$.

3. a) Soit $\lambda > 0$. On suppose que X suit la loi de Poisson de paramètre λ et que X^* suit la loi de X biaisée par la taille.

3. a) i) Donner la loi de X^* .

3. a) ii) Vérifier que X^* suit la même loi que $X + 1$.

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

3. b) i) Montrer que, pour tout $k \geq 1$, $P(X = k) = \frac{E(X)}{k} P(X = k - 1)$.

3. b) ii) Montrer que, pour tout entier naturel k , $P(X = k) = \frac{E(X)^k}{k!} P(X = 0)$.

3. b) iii) En déduire la loi de X .

4. Le paradoxe du temps d'attente du bus. Soit $n \geq 1$ et X à valeurs dans $\{1, \dots, n\}$ avec $P(X = k) > 0$ pour tout k de cet ensemble. Les intervalles entre deux bus

consécutifs, exprimés en minutes, sont supposés indépendants et de même loi que X . Une personne arrive à l'arrêt à un instant aléatoire et se demande combien de temps elle va attendre.

4. a) Une première modélisation suppose une arrivée à un instant uniforme entre deux arrivées de bus séparées par X minutes. On note T le temps d'attente à valeurs dans $\{1, \dots, n\}$ et on suppose que, pour $1 \leq k \leq n$, $P_{[X=k]}(T = j) = 1/k$ si $1 \leq j \leq k$, et 0 si $j > k$.

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

$$\sum_{j=1}^n j P_{[X=k]}(T = j) = (k+1)/2.$$

4. a) ii) En déduire que

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

4. a) iii) Montrer que

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

4. a) iv) Montrer que $E(T) = E(X+1)/2$.

4. b) En arrivant à l'arrêt de bus, on tombe dans un intervalle de manière proportionnelle à sa taille. Cet intervalle a donc une durée X^* suivant la loi de X biaisée par la taille. Le temps d'attente T^* vérifie $P_{[X^*=k]}(T^* = j) = 1/k$ si $1 \leq j \leq k$, et 0 si $j > k$.

4. b) i) Montrer que, pour tout $k \in \{1, \dots, n\}$,

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

4. b) ii) Montrer que

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

4. b) iii) Montrer que $E(T^*) = E(X^* + 1)/2$.

4. b) iv) En déduire que $E(T^*) \geq E(T)$.

Deuxième partie : Biais par la taille, propriétés

On démontre plusieurs propriétés des variables aléatoires biaisées par la taille.

5. Biais par la taille : le cas des variables à densité. Soit X une variable aléatoire positive de densité f et d'espérance $E(X) > 0$. On choisit $f(x) = 0$ pour $x < 0$. On définit g sur $\mathbb{R}$ par $g(x) = \frac{x}{E(X)} f(x)$.

5. a) Montrer que g est une densité d'une variable aléatoire positive.

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

5. b) Soit $a > 0$.

5. b) i) Montrer que aX possède pour densité $x \mapsto \frac{1}{a} f(x/a)$.

5. b) ii) En déduire que $(aX)^*$ et aX^* possèdent la même loi.

5. c) Une propriété importante. Soit $h : [0, +\infty[\rightarrow \mathbb{R}$ bornée et continue sauf éventuellement en un nombre fini de points. Montrer que $E(Xh(X))$ est bien définie et que $E(h(X^*)) = \frac{1}{E(X)}E(Xh(X))$.

On pose la définition suivante, que X soit à densité ou non : si X est positive et d'espérance strictement positive, on dit que la variable aléatoire positive Y suit la loi de X biaisée par la taille si

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

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

6. Dans cette question, $f, g : \mathbb{R} \rightarrow \mathbb{R}$ sont croissantes. Soit X telle que $E(f(X))$, $E(g(X))$ et $E(f(X)g(X))$ soient bien définies.

6. a) Montrer que, pour tous réels x_1, x_2 , $(f(x_1) - f(x_2))(g(x_1) - g(x_2)) \geq 0$.

6. b) Soient X_1, X_2 indépendantes et de même loi que X . Montrer que

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

6. c) En déduire que $E(f(X)g(X)) \geq E(f(X))E(g(X))$.

7. On suppose que X est positive, d'espérance strictement positive et que $E(X^{m+1})$ existe pour un entier $m \geq 1$ donné.

7. a) Soit $p \in \{1, \dots, m\}$.

7. a) i) Montrer que, pour tout réel $x \geq 0$, $0 \leq x^p \leq 1 + x^{m+1}$.

7. a) ii) Montrer que $E(X^p)$ existe.
7. b) Montrer que $E(X^{m+1}) \geq E(X)E(X^m)$.
7. c) On admet que, dans le cadre de cette question, X^* possède un moment d'ordre m et que $E((X^*)^m) = E(X^{m+1})/E(X)$. En déduire que $E((X^*)^m) \geq E(X^m)$.
8. Pour un événement A , on note $\mathbf{1}_A$ la variable valant 1 sur A et 0 ailleurs. Pour tout réel t , on définit $g_t(x) = \mathbf{1}_{]t, +\infty[}(x)$.
8. a) Montrer que g_t est croissante sur $\mathbb{R}$.
8. b) Soit X positive et admettant une espérance. Montrer que, pour tout réel t , $E(Xg_t(X))$ est bien définie et que $E(Xg_t(X)) \geq E(X)P(X > t)$.
8. c) Si $E(X) > 0$, montrer que, pour tout réel t , $P(X^* > t) \geq P(X > t)$. On dit que X^* domine stochastiquement X .
9. Soient $X_1, \dots, X_n$ des variables aléatoires positives, indépendantes, non nécessairement de même loi, d'espérances strictement positives $\mu_i = E(X_i)$. On pose

$$\mu = \sum_{i=1}^n \mu_i$$

et

$$S_n = \sum_{i=1}^n X_i.$$

9. a) Donner $E(S_n)$.

9. b) Soit J à valeurs dans $\{1, \dots, n\}$ de loi $P(J = k) = \mu_k/\mu$. Quelle est sa loi si les X_i sont de même loi ?

On considère $X_1^*, \dots, X_n^*$ indépendantes, indépendantes de $X_1, \dots, X_n$, telles que X_i^* suit la loi de X_i biaisée par la taille. On choisit aussi J de la loi précédente, indépendante de toutes ces variables. On pose

$$X_J = \sum_{j=1}^n X_j \mathbf{1}_{[J=j]},$$

$$X_J^* = \sum_{j=1}^n X_j^* \mathbf{1}_{[J=j]}$$

et $T_n = S_n - X_J + X_J^*$. On remplace donc dans la somme une variable choisie au hasard par sa version biaisée par la taille.

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

9. c) i) Montrer que

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

9. c) ii) En déduire que

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

9. d) Pour $i \in \{1, \dots, n\}$ et tout $s \geq 0$, montrer que $E(h(s + X_i^*)) = \frac{1}{\mu_i} E(X_i h(s + X_i))$.

On admet qu'on en déduit l'égalité particulière $E(h(S_n - X_i + X_i^*)) = \frac{1}{\mu_i} E(X_i h(S_n))$.

9. e) En déduire que $E(h(T_n)) = E(S_n h(S_n)) / E(S_n)$.

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

Troisième partie : Applications en statistique

On souhaite construire des estimateurs non biaisés grâce au biais par la taille. Une compagnie d'électricité possède $n \geq 1$ clients. Durant l'année écoulée, le client i a payé $x_i > 0$ euros, mais a consommé une quantité correspondant à $y_i > 0$ euros. La compagnie connaît ses clients et leurs paiements et veut estimer

$$\theta = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i},$$

afin de déterminer à quel point elle les a mal facturés. Les x_i, y_i sont des données fixes. L'aléa vient du groupe de clients tiré. Un estimateur de θ est une variable calculée à partir de ce groupe ; on le dit non biaisé si son espérance est θ , et biaisé sinon.

10. Soit m fixé avec $1 \leq m \leq n$. On note $\mathcal{P}_m$ l'ensemble des parties de $\{1, \dots, n\}$ de cardinal m . On considère une variable aléatoire R uniforme sur $\mathcal{P}_m$: pour tout $A \in \mathcal{P}_m$, $P(R = A) = 1/\binom{n}{m}$. On veut programmer le choix de R .

10. a) On tire successivement s_1 uniformément dans $\{1, \dots, n\}$, puis s_2 uniformément parmi les indices restants, et ainsi de suite jusqu'à s_m . On pose $S = (s_1, \dots, s_m)$.

10. a) i) Montrer que, pour tout m -uplet $(a_1, \dots, a_m)$ d'entiers distincts de $\{1, \dots, n\}$, $P(S = (a_1, \dots, a_m)) = (n - m)!/n!$.

10. a) ii) On note $R = \{s_1, \dots, s_m\}$ l'ensemble obtenu sans tenir compte de l'ordre. Montrer que, pour toute partie A de cardinal m , $P(R = A) = m!(n - m)!/n!$. En déduire que R est uniforme sur $\mathcal{P}_m$.

10. b) Pour tout réel x , $\lfloor x \rfloor$ est sa partie entière, c'est-à-dire le plus grand entier inférieur ou égal à x . Montrer que, si U suit la loi uniforme sur $[0, 1[$, $1 + \lfloor nU \rfloor$ suit la loi uniforme sur $\{1, \dots, n\}$.

10. c) On utilise les imports suivants :

Python

```
import numpy as np
import numpy.random as rd
```

La commande `rd.random()` renvoie un réel de loi uniforme sur $[0, 1[$; `np.floor(x)` renvoie la partie entière de x sous forme numérique, et `int(...)` la convertit en entier Python. Écrire une fonction `Uniforme` qui prend un entier $n \geq 1$ et renvoie un entier uniforme dans $\{1, \dots, n\}$.

Python

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

10. d) Écrire une fonction `Selection` qui prend une liste non vide V d'éléments distincts, renvoie un élément x choisi uniformément parmi ses éléments et une liste W obtenue en enlevant x de V . `len(V)` est le nombre d'éléments de V . Les listes sont indexées à partir de zéro.

Python

```
def Selection(V):
    n = len(V)
    ...
    return x, W
```

10. e) Compléter la fonction suivante. Elle prend $1 \leq m \leq n$ et renvoie une liste de m entiers distincts, tirés comme à la question a). `list(range(1, n+1))` crée la liste $\{1, \dots, n\}$ et `R.append(x)` ajoute x à la fin de R .

Python

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

11. Pour $A \in \mathcal{P}_m$, on définit

$$\bar{x}_A = \frac{1}{m} \sum_{i \in A} x_i,$$

$$\bar{y}_A = \frac{1}{m} \sum_{i \in A} y_i,$$

et

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

La compagnie choisit $\theta_R = \bar{y}_R / \bar{x}_R$ comme estimateur de θ . Dans cette question, R reste uniforme sur $\mathcal{P}_m$.

11. a) On pose

$$X = \bar{x}_R = \frac{1}{m} \sum_{i \in R} x_i$$

et

$$Y = \bar{y}_R = \frac{1}{m} \sum_{i \in R} y_i.$$

11. a) i) Montrer que

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

11. a) ii) Pour $1 \leq i \leq n$, calculer le nombre de parties $A \in \mathcal{P}_m$ contenant i .

11. a) iii) En déduire que

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

11. a) iv) Conclure que $E(X) = \bar{x}$. On admet que, de même, $E(Y) = \bar{y}$.

11. a) v) Exprimer θ en fonction de $E(X)$ et $E(Y)$.

11. b) Montrer que $E(\theta_R) = E(Y/X)$.

11. c) On donne l'inégalité de Cauchy-Schwarz suivante : si W, Z sont strictement positives et admettent un moment d'ordre 2, alors $E(WZ) \leq \sqrt{E(W^2)E(Z^2)}$, avec égalité si et seulement s'il existe $\alpha > 0$ tel que $W = \alpha Z$ presque sûrement.

11. c) i) Montrer que $E(1/X) \geq 1/E(X)$.

11. c) ii) Montrer qu'il y a égalité si et seulement si X est constante, c'est-à-dire $X = E(X) = \bar{x}$ presque sûrement.

11. c) iii) Si $m < n$, en conclure que $E(1/X) = 1/E(X)$ si et seulement si $x_i = \bar{x}$ pour tout i . Pour le sens nécessaire, on pourra comparer deux groupes de cardinal m ayant $m - 1$ éléments communs. Que se passe-t-il si $m = n$?
11. d) Si X, Y sont indépendantes, montrer que $E(\theta_R) \geq \theta$. Si $m < n$, montrer qu'il y a égalité si et seulement si $x_i = \bar{x}$ pour tout i . Traiter aussi le cas $m = n$.

Ainsi $E(\theta_R)$ n'est pas nécessairement égal à θ : θ_R peut être un estimateur biaisé.

12. On choisit maintenant le groupe de manière biaisée par la taille. On tire d'abord $J \in \{1, \dots, n\}$ avec

$$P(J = i) = x_i / \sum_{r=1}^n x_r.$$

Étant donné J , on choisit uniformément un groupe V de $m - 1$ clients parmi les $n - 1$ clients différents de J . Pour toute partie $A \in \mathcal{P}_m$ et tout $i \in A$,

$$P_{[J=i]}(V = A \setminus \{i\}) = \frac{1}{\binom{n-1}{m-1}}.$$

Le groupe examiné est $R = V \cup \{J\}$.

12. a) i) Pour $A \in \mathcal{P}_m$, montrer que

$$P(R = A) = \sum_{i \in A} P(J = i) P_{[J=i]}(V = A \setminus \{i\}).$$

12. a) ii) En déduire que $P(R = A) = \frac{1}{\binom{n}{m}} \frac{\bar{x}_A}{\bar{x}}$.

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

13. a) Montrer que

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

13. b) Conclure que $E(\hat{\theta}_R) = \theta$. On a donc construit un estimateur non biaisé de θ .