



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2 — Option économique

2022

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Sujet

Toutes les variables aléatoires intervenant dans le problème sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. Pour une variable aléatoire X , on note $E(X)$ son espérance et $V(X)$ sa variance lorsqu'elles existent.

On se donne une suite $(X_i)_{i \geq 1}$ de variables aléatoires réelles indépendantes, de même loi, admettant un moment d'ordre deux. On note $\mu = E(X_i)$ et $\sigma^2 = V(X_i)$, avec $\sigma > 0$. Pour tout entier $n \geq 1$, on définit

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad Z_n = \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu).$$

Première partie : Autour du théorème limite central

1) Soit Z une variable aléatoire de loi normale $\mathcal{N}(0, 1)$, de densité $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

On note Φ sa fonction de répartition, définie par $\Phi(x) = \int_{-\infty}^x f_Z(t) dt$.

1(a) Montrer que Φ est continue sur $\mathbb{R}$.

1(b) Montrer que Φ est strictement croissante.

1(c) Montrer que Φ est une bijection de $\mathbb{R}$ sur $]0, 1[$.

1(d) Montrer que, pour tout $x \in \mathbb{R}$, $\Phi(-x) = 1 - \Phi(x)$.

2(a) Énoncer la loi faible des grands nombres pour la suite $(X_i)_{i \geq 1}$.

2(b) Rappeler les hypothèses du théorème limite central et en déduire que, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} P(Z_n \leq x) = \Phi(x). \quad (*)$$

3) Donnons une première application du théorème limite central, pour un joueur de fléchettes. Au i -ème lancer, le score est une variable aléatoire X_i prenant ses valeurs dans $\{0, 2, 5, 10\}$. Les X_i sont indépendantes, de même loi donnée par

$$P(X_i = 0) = \frac{1}{5}, \quad P(X_i = 2) = \frac{1}{2}, \quad P(X_i = 5) = \frac{1}{5}, \quad P(X_i = 10) = \frac{1}{10}.$$

3(a) Calculer $E(X_i)$.

3(b) Calculer $V(X_i)$.

3(c) Soit $f : [0, 1] \rightarrow \mathbb{R}$ définie par $f(x) = 0$ si $x \in [0, \frac{1}{5}[$, $f(x) = 2$ si $x \in [\frac{1}{5}, \frac{7}{10}[$,
 $f(x) = 5$ si $x \in [\frac{7}{10}, \frac{9}{10}[$ et $f(x) = 10$ si $x \in [\frac{9}{10}, 1]$.

3(c)(i) Soit U de loi uniforme sur $[0, 1]$. Montrer que $f(U)$ a la même loi que X_i .

3(c)(ii) Compléter le programme Python suivant, qui génère un nombre aléatoire de même loi que X_i . La commande `rd.random()` renvoie un réel aléatoire uniforme sur $[0, 1[$.

Python

```
import numpy.random as rd

def X():
    U = rd.random()
    # À compléter
```

Après n lancers, le score du joueur est

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

3(d) Exprimer Z_n en fonction de S_n et de n .

3(e) Le joueur lance $n = 200$ fléchettes. En utilisant l'approximation normale fournie par le théorème limite central pour la somme de ces variables indépendantes, de même loi et de variance non nulle, montrer que $P(S_{200} \leq 500)$ est approximativement $\Phi(-2,5)$. Cette probabilité vaut environ $6 \cdot 10^{-3}$.

- 4) Soit $N \geq 1$ un entier fixé. Pour $k \in \{1, \dots, 2N-1\}$, on pose $x_k = \Phi^{-1}(\frac{k}{2N})$. On pose aussi $x_0 = -\infty$, $x_{2N} = +\infty$, avec les conventions $\Phi(x_0) = 0$, $\Phi(x_{2N}) = 1$,
 $P(Z_n \leq -\infty) = 0$ et $P(Z_n \leq +\infty) = 1$.

4(a) Montrer qu'il existe n_0 , dépendant de N , tel que pour tout $n \geq n_0$,

$$\max_{k \in \{0, \dots, 2N\}} |P(Z_n \leq x_k) - \Phi(x_k)| \leq \frac{1}{2N}.$$

On divise $\mathbb{R}$ en intervalles $I_k =]x_{k-1}, x_k]$ pour $k \in \{1, \dots, 2N-1\}$, avec $I_1 =]-\infty, x_1]$ et $I_{2N} =]x_{2N-1}, +\infty[$.

4(b) Soit $k \in \{1, \dots, 2N\}$ et $x \in I_k$. Soit $n \geq n_0$.

4(b)(i) Montrer que $P(Z_n \leq x) - \Phi(x) \leq P(Z_n \leq x_k) - \Phi(x_{k-1})$.

4(b)(ii) En déduire que $P(Z_n \leq x) - \Phi(x) \leq \Phi(x_k) - \Phi(x_{k-1}) + \frac{1}{2N} = \frac{1}{N}$.

4(b)(iii) Montrer de même que $\Phi(x) - P(Z_n \leq x) \leq \frac{1}{N}$.

4(c) En déduire que, pour tout $x \in \mathbb{R}$ et tout $n \geq n_0$, $|P(Z_n \leq x) - \Phi(x)| \leq \frac{1}{N}$.

4(d) On cherche une suite de réels positifs $(M_n)_{n \geq 1}$ telle que $|P(Z_n \leq x) - \Phi(x)| \leq M_n$ pour tout $x \in \mathbb{R}$ et tout $n \geq 1$. Montrer qu'on peut choisir cette suite avec

$$\lim_{n \rightarrow +\infty} M_n = 0. \quad (E)$$

On pourra définir M_n comme la borne supérieure de l'ensemble borné des valeurs $|P(Z_n \leq x) - \Phi(x)|$, pour $x \in \mathbb{R}$.

Le résultat (E) est plus fort que (*).

5) Soit $x \in \mathbb{R}$ fixé.

5(a) Soit $(x_n)_{n \geq 1}$ une suite réelle convergeant vers x .

5(a)(i) Montrer que $\Phi(x_n) \rightarrow \Phi(x)$.

5(a)(ii) En appliquant (E), montrer que $|P(Z_n \leq x_n) - \Phi(x_n)| \rightarrow 0$.

5(a)(iii) En déduire que $P(Z_n \leq x_n) \rightarrow \Phi(x)$.

5(b)(i) Montrer que, pour tout $n \geq 1$, $P(Z_n \leq x - \frac{1}{n}) \leq P(Z_n < x) \leq P(Z_n \leq x)$.

5(b)(ii) En déduire que $P(Z_n < x) \rightarrow \Phi(x)$.

5(c) Montrer que, pour tous réels $a < b$, $P(Z_n \in [a, b]) \rightarrow \Phi(b) - \Phi(a)$.

Deuxième partie : Applications en statistique

Dans toute cette partie, les X_i sont indépendantes et de loi de Bernoulli de paramètre $p \in]0, 1[$. On modélise une élection entre les candidats A et B : $X_i = 1$ si la i -ème personne vote pour A, $X_i = 0$ si elle vote pour B. Le paramètre inconnu p est la proportion de voix obtenues par A. Le sondage porte sur un échantillon indépendant de n personnes et

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

est la proportion observée en faveur de A.

6(a) Démontrer que $E(X_i) = p$ et $V(X_i) = p(1 - p)$.

6(b) On note désormais $\sigma = \sqrt{p(1 - p)}$. Montrer que $\sigma \leq \frac{1}{2}$.

6(c) Montrer que $E(\bar{X}_n) = p$.

6(d) Montrer que $V(\bar{X}_n) = \frac{\sigma^2}{n}$.

7) On construit des intervalles de confiance pour p à partir du théorème limite central.

Un intervalle aléatoire de niveau de confiance asymptotique γ contient le paramètre fixe p avec une probabilité tendant vers γ lorsque la taille de l'échantillon tend vers l'infini.

7(a) Montrer que, pour tout $a > 0$,

$$\lim_{n \rightarrow +\infty} P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - p) \in [-a, a]\right) = \Phi(a) - \Phi(-a).$$

7(b) En déduire que

$$\lim_{n \rightarrow +\infty} P\left(p \in \left[\bar{X}_n - a \frac{\sigma}{\sqrt{n}}, \bar{X}_n + a \frac{\sigma}{\sqrt{n}}\right]\right) = 2\Phi(a) - 1.$$

7(c) Une table donne $\Phi(1,96) \simeq 0,975$. En déduire que l'intervalle $[\bar{X}_n - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1,96 \frac{\sigma}{\sqrt{n}}]$ a un niveau de confiance asymptotique voisin de 95%.

La quantité σ dépend de p , qui est inconnu.

7(d) Montrer que l'intervalle $[\bar{X}_n - \frac{0,98}{\sqrt{n}}, \bar{X}_n + \frac{0,98}{\sqrt{n}}]$ a une couverture asymptotique au moins égale à $2\Phi(1,96) - 1$, soit environ 95%.

Une autre solution consiste à utiliser les observations pour estimer σ .

8) Pour $n \geq 1$, on pose $V_n = \bar{X}_n(1 - \bar{X}_n) + \frac{1}{n}$. Soit $\varepsilon > 0$.

8(a) Montrer que $V_n - \sigma^2 - \frac{1}{n} = (\bar{X}_n - p)(1 - \bar{X}_n - p)$.

8(b) En déduire que $|V_n - \sigma^2| \leq 2|\bar{X}_n - p| + \frac{1}{n}$.

8(c) Montrer que $P(|V_n - \sigma^2| > \varepsilon) \leq P(|\bar{X}_n - p| > \frac{\varepsilon}{2} - \frac{1}{2n})$.

8(d) Montrer que, pour n assez grand, $P(|\bar{X}_n - p| > \frac{\varepsilon}{2} - \frac{1}{2n}) \leq P(|\bar{X}_n - p| > \frac{\varepsilon}{4})$.

8(e) Conclure que, pour tout $\varepsilon > 0$, $P(|V_n - \sigma^2| > \varepsilon) \rightarrow 0$.

9) On pose $W_n = \frac{\sqrt{n}}{\sqrt{V_n}}(\bar{X}_n - p) = \frac{\sigma}{\sqrt{V_n}}Z_n$. On fixe $x \in \mathbb{R}$ et on note

$$R_n = \frac{\sqrt{V_n}}{\sigma} > 0.$$

9(a) Soit $\varepsilon > 0$.

9(a)(i) En distinguant selon que $|R_n - 1| \leq \varepsilon$ ou non, montrer que

$$P(W_n \leq x) \leq P(Z_n \leq x + \varepsilon|x|) + P(|R_n - 1| > \varepsilon).$$

9(a)(ii) Montrer que $P(|R_n - 1| > \varepsilon) \rightarrow 0$. On pourra utiliser $|R_n^2 - 1| = |R_n - 1|(R_n + 1)$.

9(a)(iii) Montrer que $P(Z_n \leq x + \varepsilon|x|) \rightarrow \Phi(x + \varepsilon|x|)$.

9(a)(iv) En déduire qu'il existe n_ε tel que, pour tout $n \geq n_\varepsilon$,

$$P(W_n \leq x) \leq \Phi(x + \varepsilon|x|) + \varepsilon.$$

9(b) On admet que, symétriquement, pour tout $\varepsilon > 0$ et tout n assez grand,

$$P(W_n \leq x) \geq \Phi(x - \varepsilon|x|) - \varepsilon.$$

En conclure que $P(W_n \leq x) \rightarrow \Phi(x)$.

10(a) Montrer que, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} P\left(p \geq \bar{X}_n - x \frac{\sqrt{V_n}}{\sqrt{n}}\right) = \Phi(x).$$

10(b) Le candidat A remporte l'élection lorsque $p \geq \frac{1}{2}$. Une semaine avant l'élection, un sondage de $n = 1000$ personnes donne $\bar{x}_n = 0,52$ et $v_n = 0,2506$. Pour l'intervalle unilatéral $[\bar{X}_n - x \frac{\sqrt{V_n}}{\sqrt{n}}, +\infty[$, déterminer la valeur observée de x pour laquelle sa borne inférieure vaut $\frac{1}{2}$, puis le niveau de confiance asymptotique associé. On donne

$$\frac{\sqrt{1000}}{\sqrt{0,2506}} \left(0,52 - \frac{1}{2}\right) \simeq 1,26, \quad \Phi(1,26) \simeq 0,896.$$

Préciser ce que cette lecture en termes de confiance permet de dire, et pourquoi elle n'est pas une probabilité attribuée à p après observation du sondage.

- 11) Certains électeurs déclarent A puis votent B. On suppose que, avec une probabilité connue $q \in [0, 1[$, un électeur votant B a déclaré A. La réponse enregistrée est $Y_i = X_i + (1 - X_i)T_i$, où les T_i sont indépendantes de loi de Bernoulli de paramètre q , et où toute la famille des T_i est indépendante de celle des X_i .

11(a) Montrer que Y_i suit une loi de Bernoulli dont on déterminera le paramètre r .

On définit

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$$

et $U_n = \bar{Y}_n(1 - \bar{Y}_n) + \frac{1}{n}$. Comme à la question 9, on admet que, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} P \left(\frac{\sqrt{n}}{\sqrt{U_n}} (\bar{Y}_n - r) \leq x \right) = \Phi(x).$$

11(b) Montrer que, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} P \left(p \geq \frac{1}{1-q} \left(\bar{Y}_n - x \frac{\sqrt{U_n}}{\sqrt{n}} - q \right) \right) = \Phi(x).$$

11(c) Un sondage donne les valeurs observées $\bar{y}_n$ et $u_n = \bar{y}_n(1 - \bar{y}_n) + \frac{1}{n}$. Déterminer la valeur observée de x pour laquelle la borne de confiance précédente vaut $\frac{1}{2}$, et en déduire le niveau de confiance asymptotique associé à cette borne.

11(d) Avec $n = 1000$, $\bar{y}_n = 0,52$ et $q = 0,04$, montrer que cette valeur de x est nulle, et que le niveau de confiance correspondant est $\frac{1}{2}$. Comparer avec la question 10(b), en distinguant confiance et probabilité après observation.

Troisième partie : Démonstration du théorème limite central

Dans cette partie, on suppose en outre que les X_i admettent un moment absolu d'ordre trois, $E(|X_i|^3) < +\infty$. On se propose de démontrer (*) sous cette hypothèse

supplémentaire. On suppose, pour simplifier, $\mu = E(X_i) = 0$ et $\sigma^2 = V(X_i) = 1$, d'où

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i.$$

On fixe un réel x pour toute la suite.

Une fonction est dite de classe C^3 lorsqu'elle est trois fois dérivable et que ses dérivées jusqu'à l'ordre trois sont continues. Pour une fonction bornée f , M_f désigne un majorant positif de $|f|$ sur $\mathbb{R}$.

12(a)(i) Montrer, par intégrations par parties successives, que

$$\int_0^1 u^3(1-u)^3 du = \frac{1}{20} \int_0^1 u^6 du.$$

12(a)(ii) En déduire que $\int_0^1 u^3(1-u)^3 du = \frac{1}{140}$.

On définit $h : \mathbb{R} \rightarrow \mathbb{R}$ par $h(z) = 0$ si $z \leq 0$, $h(z) = 140 \int_0^z u^3(1-u)^3 du$ si $0 < z < 1$, et $h(z) = 1$ si $z \geq 1$.

12(b) Montrer que h est continue sur $\mathbb{R}$.

12(c) Montrer que $0 \leq h(z) \leq 1$ pour tout $z \in \mathbb{R}$.

On admet que h' , h'' et h''' sont continues sur $\mathbb{R}$.

13) On pose $a_n = n^{-1/12}$ et $g_n(z) = 1 - h\left(\frac{z-x}{a_n}\right)$.

13(a)(i) Montrer que g_n est continue et que $0 \leq g_n(z) \leq 1$ pour tout $z \in \mathbb{R}$.

On admet que g'_n , g''_n et g'''_n sont continues et bornées sur $\mathbb{R}$.

13(a)(ii) Montrer qu'on peut choisir $M_{g'''_n} \leq n^{1/4} M_{h'''}$.

13(b)(i) Montrer que $g_n(z) = 1$ si $z \leq x$ et $g_n(z) = 0$ si $z \geq x + a_n$.

Pour un événement A , on définit l'indicatrice $\mathbf{1}_A$ par $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ sinon.

13(b)(ii) Montrer que, pour toute variable aléatoire réelle X ,

$$\mathbf{1}_{\{X \leq x\}} \leq g_n(X) \leq \mathbf{1}_{\{X \leq x + a_n\}}.$$

13(c) Montrer que $E(g_n(Z_n + a_n)) \leq P(Z_n \leq x) \leq E(g_n(Z_n))$.

Il suffit donc de montrer que ces deux espérances tendent vers $\Phi(x)$. On se concentre d'abord sur la seconde.

14) Soit g de classe C^3 sur $\mathbb{R}$.

14(a) Par des intégrations par parties successives, montrer que, pour tous $z, u \in \mathbb{R}$,

$$\frac{1}{2} \int_z^{z+u} (z+u-t)^2 g'''(t) dt = -\frac{1}{2} u^2 g''(z) - u g'(z) + g(z+u) - g(z).$$

14(b) En déduire que, pour tous $u, v, z \in \mathbb{R}$,

$$g(z+u) - g(z+v) = g'(z)(u-v) + \frac{1}{2} g''(z)(u^2 - v^2) + R(z, u, v),$$

où

$$R(z, u, v) = \frac{1}{2} \int_z^{z+u} (z+u-t)^2 g'''(t) dt - \frac{1}{2} \int_z^{z+v} (z+v-t)^2 g'''(t) dt.$$

14(c) Montrer que, si g''' est bornée, $|R(z, u, v)| \leq \frac{1}{6} M_{g'''}(|u|^3 + |v|^3)$.

15) Soit $(Y_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes de loi normale $\mathcal{N}(0, 1)$, indépendante de la famille des X_j .

15(a) Soit un entier $n \geq 2$.

15(a)(i) Justifier que

$$\sum_{i=1}^n Y_i$$

suit une loi normale.

15(a)(ii) On pose

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i.$$

Déterminer sa loi.

15(b) Pour $k \in \{1, \dots, n\}$, on pose

$$W_k = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{k-1} Y_i + \sum_{i=k+1}^n X_i \right),$$

avec la convention qu'une somme vide vaut zéro.

15(b)(i) Montrer que, pour $k \in \{1, \dots, n-1\}$, $W_k + \frac{1}{\sqrt{n}} Y_k = W_{k+1} + \frac{1}{\sqrt{n}} X_{k+1}$.

15(b)(ii) En déduire que

$$g_n(Z_n) - g_n(T_n) = \sum_{k=1}^n \left[g_n \left(W_k + \frac{X_k}{\sqrt{n}} \right) - g_n \left(W_k + \frac{Y_k}{\sqrt{n}} \right) \right].$$

On calcule l'espérance de chacun des termes de la somme.

15(c)(i) Montrer que $E((X_k - Y_k)g'_n(W_k)) = 0$.

15(c)(ii) Montrer que $E((X_k^2 - Y_k^2)g''_n(W_k)) = 0$.

15(c)(iii) En déduire que

$$E \left[g_n \left(W_k + \frac{X_k}{\sqrt{n}} \right) - g_n \left(W_k + \frac{Y_k}{\sqrt{n}} \right) \right] = E \left[R \left(W_k, \frac{X_k}{\sqrt{n}}, \frac{Y_k}{\sqrt{n}} \right) \right],$$

où R est le reste de la question 14 appliqué à g_n .

15(d) Montrer que

$$|E(g_n(Z_n)) - E(g_n(T_n))| \leq \sum_{k=1}^n E \left(\left| R \left(W_k, \frac{X_k}{\sqrt{n}}, \frac{Y_k}{\sqrt{n}} \right) \right| \right).$$

15(e) Montrer que

$$|E(g_n(Z_n)) - E(g_n(T_n))| \leq \frac{M_{g_n}'''}{6\sqrt{n}} (E(|X_1|^3) + E(|Y_1|^3)).$$

15(f) En déduire que $E(g_n(Z_n)) - E(g_n(T_n)) \rightarrow 0$.

16(a) Montrer que $P(T_n \leq x) \leq E(g_n(T_n)) \leq P(T_n \leq x + a_n)$.

16(b) En déduire que $E(g_n(T_n)) \rightarrow \Phi(x)$.

16(c) Conclure que $E(g_n(Z_n)) \rightarrow \Phi(x)$. Pour achever la démonstration annoncée, on pourra appliquer le même raisonnement à la fonction $z \mapsto g_n(z + a_n)$, dont les dérivées ont les mêmes majorants, puis utiliser la question 13(c).