



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2 — Option économique

2022

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Corrigé

Première partie : Autour du théorème limite central

1(a) La densité f_Z est continue sur $\mathbb{R}$. Pour un réel a fixé, la relation de Chasles donne $\Phi(x) = \Phi(a) + \int_a^x f_Z(t) dt$. Le théorème fondamental de l'intégration assure donc que Φ est dérivable, de dérivée $\Phi' = f_Z$. Elle est en particulier continue sur $\mathbb{R}$.

1(b) Méthode 1 — Signe de la dérivée

Pour tout x , $\Phi'(x) = f_Z(x) > 0$. Le critère de stricte monotonie donne la stricte croissance de Φ sur $\mathbb{R}$.

Méthode 2 — Comparaison des probabilités

Si $a < b$, $\Phi(b) - \Phi(a) = \int_a^b f_Z(t) dt$. La fonction intégrée est continue et strictement positive sur le segment non réduit $[a, b]$, donc cette intégrale est strictement positive. Ainsi $\Phi(a) < \Phi(b)$. Cette preuve relie directement la croissance à la masse de probabilité de chaque intervalle.

1(c) Une fonction de répartition vérifie $\lim_{x \rightarrow -\infty} \Phi(x) = 0$ et $\lim_{x \rightarrow +\infty} \Phi(x) = 1$. Ici $\Phi(x) > 0$ et $1 - \Phi(x) > 0$, car les intégrales de la densité strictement positive respectivement sur $[x-1, x]$ et sur $[x, x+1]$ sont positives. La continuité et le théorème des valeurs intermédiaires donnent donc $\Phi(\mathbb{R}) =]0, 1[$. Sa stricte croissance donne l'injectivité. Elle est ainsi bijective de $\mathbb{R}$ sur $]0, 1[$.

1(d) La densité est paire : $f_Z(-t) = f_Z(t)$. Pour $A > |x|$, le changement affine $u = -t$ sur le segment $[-A, -x]$ donne $\int_{-A}^{-x} f_Z(t) dt = \int_x^A f_Z(u) du$. Le passage à la limite, autorisé par la convergence des intégrales définissant la loi normale, donne

$$\Phi(-x) = \int_x^{+\infty} f_Z(u) du = 1 - \Phi(x).$$

2(a) Pour tout $\varepsilon > 0$, la loi faible des grands nombres donne

$$\lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

Les hypothèses sont satisfaites : les variables sont indépendantes, de même espérance finie et de même variance finie. On peut également établir directement le résultat par $V(\bar{X}_n) = \sigma^2/n$ et l'inégalité de Bienaymé-Tchebychev, qui donne $P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \sigma^2/(n\varepsilon^2)$.

2(b) Le théorème limite central s'applique à une suite de variables indépendantes, de même loi, admettant une variance finie strictement positive. Il affirme que les moyennes centrées réduites $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ convergent en loi vers $\mathcal{N}(0, 1)$. Toutes ces hypothèses figurent dans le préambule. La fonction de répartition limite Φ étant continue en tout réel, la définition de la convergence en loi donne, pour tout $x \in \mathbb{R}$, $P(Z_n \leq x) \rightarrow \Phi(x)$.

3(a) Le support est fini, donc tous les moments existent. Par transfert,

$$E(X_i) = 0 \cdot \frac{1}{5} + 2 \cdot \frac{1}{2} + 5 \cdot \frac{1}{5} + 10 \cdot \frac{1}{10} = 3.$$

3(b) Le second moment vaut $E(X_i^2) = 4 \cdot \frac{1}{2} + 25 \cdot \frac{1}{5} + 100 \cdot \frac{1}{10} = 17$. La formule de Koenig-Huygens donne $V(X_i) = 17 - 3^2 = 8$. Ainsi $\sigma = \sqrt{8} = 2\sqrt{2}$.

3(c)(i) La probabilité qu'une variable uniforme sur $[0, 1]$ appartienne à un intervalle est sa longueur; chaque extrémité isolée a une probabilité nulle. Les quatre intervalles de définition de f ont respectivement pour longueurs $1/5$, $7/10 - 1/5 = 1/2$, $9/10 - 7/10 = 1/5$ et $1 - 9/10 = 1/10$. Ils forment une partition de $[0, 1]$. Par conséquent, $f(U)$ prend les valeurs 0, 2, 5, 10 avec les probabilités demandées : sa loi est celle de X_i .

3(c)(ii) Question adaptée au programme actuel. Code adapté en Python.

Python

```
import numpy.random as rd

def X():
    U = rd.random()
```

```

if U < 1 / 5:
    return 0
elif U < 7 / 10:
    return 2
elif U < 9 / 10:
    return 5
else:
    return 10

```

Les tests successifs découpent l'intervalle uniforme dans l'ordre des quatre branches. Chaque appel effectue un seul tirage et renvoie le score associé. L'exclusion de la valeur 1 par `rd.random()` ne modifie aucune probabilité, puisque la loi uniforme ne donne aucune masse à ce point.

3(d) On a $\bar{X}_n = S_n/n$, $\mu = 3$ et $\sigma = 2\sqrt{2}$, d'où

$$Z_n = \frac{S_n - 3n}{2\sqrt{2n}}.$$

3(e) Question adaptée au programme actuel : les hypothèses justifiant l'approximation sont explicitées.

Les scores sont indépendants, de même loi et de variance $8 > 0$; le théorème limite central fournit donc l'approximation normale de leur somme centrée réduite. Pour $n = 200$, $3n = 600$ et $2\sqrt{2n} = 40$. Ainsi

$$P(S_{200} \leq 500) = P\left(Z_{200} \leq \frac{500 - 600}{40}\right) = P(Z_{200} \leq -2,5) \simeq \Phi(-2,5) \simeq 6 \cdot 10^{-3}.$$

Il s'agit d'une approximation pour un effectif fini, et non d'une égalité découlant du théorème limite central.

4(a) Pour chaque $k \in \{1, \dots, 2N - 1\}$, (*) donne un entier n_k à partir duquel $|P(Z_n \leq x_k) - \Phi(x_k)| \leq 1/(2N)$. Il n'y a qu'un nombre fini de tels indices. Choisissons $n_0 = \max(1, n_1, \dots, n_{2N-1})$. Pour $k = 0$ et $k = 2N$, la différence est exactement nulle par

convention. Toutes les inégalités sont donc simultanément satisfaites pour $n \geq n_0$, ce qui établit l'inégalité sur le maximum.

4(b)(i) Notons $F_n(t) = P(Z_n \leq t)$. Les fonctions F_n et Φ sont croissantes. Puisque $x_{k-1} < x \leq x_k$, $F_n(x) \leq F_n(x_k)$ et $\Phi(x) \geq \Phi(x_{k-1})$. Soustraire ces deux inégalités donne $F_n(x) - \Phi(x) \leq F_n(x_k) - \Phi(x_{k-1})$. Le même raisonnement vaut aux intervalles extrêmes avec les conventions de l'énoncé.

4(b)(ii) Décomposons le majorant précédent :

$$F_n(x_k) - \Phi(x_{k-1}) = [F_n(x_k) - \Phi(x_k)] + [\Phi(x_k) - \Phi(x_{k-1})].$$

Le premier terme est au plus $1/(2N)$ par 4(a) ; le second vaut exactement $1/(2N)$, y compris aux extrémités. On obtient $F_n(x) - \Phi(x) \leq 1/N$.

4(b)(iii) Cette fois, $\Phi(x) \leq \Phi(x_k)$ et $F_n(x) \geq F_n(x_{k-1})$. Ainsi

$$\Phi(x) - F_n(x) \leq \Phi(x_k) - F_n(x_{k-1}) = [\Phi(x_k) - \Phi(x_{k-1})] + [\Phi(x_{k-1}) - F_n(x_{k-1})] \leq \frac{1}{N}.$$

4(c) Tout réel x appartient à l'un des intervalles I_k . Les deux inégalités précédentes donnent $-1/N \leq F_n(x) - \Phi(x) \leq 1/N$, donc $|F_n(x) - \Phi(x)| \leq 1/N$, uniformément pour $x \in \mathbb{R}$ et $n \geq n_0$.

4(d) Question adaptée au programme actuel : la borne supérieure utilisée pour construire les majorants est explicitée dans le sujet.

L'ensemble des valeurs $|F_n(x) - \Phi(x)|$, pour $x \in \mathbb{R}$, est non vide et inclus dans $[0, 1]$. Sa borne supérieure existe ; posons $M_n = \sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)|$. Alors $0 \leq M_n \leq 1$ et chaque différence est majorée par M_n . Pour $\varepsilon > 0$, choisissons N tel que $1/N < \varepsilon$. La question 4(c) donne $M_n \leq 1/N < \varepsilon$ pour tout $n \geq n_0(N)$. Cela prouve $M_n \rightarrow 0$.

5(a)(i) La fonction Φ est continue en x et $x_n \rightarrow x$. La caractérisation séquentielle de la continuité donne $\Phi(x_n) \rightarrow \Phi(x)$.

5(a)(ii) Par définition des majorants M_n , $0 \leq |P(Z_n \leq x_n) - \Phi(x_n)| \leq M_n$. Le majorant tend vers zéro, donc la différence absolue tend vers zéro.

5(a)(iii) L'inégalité triangulaire donne

$$|P(Z_n \leq x_n) - \Phi(x)| \leq |P(Z_n \leq x_n) - \Phi(x_n)| + |\Phi(x_n) - \Phi(x)|.$$

Les deux termes tendent vers zéro d'après les deux sous-questions précédentes. Ainsi $P(Z_n \leq x_n) \rightarrow \Phi(x)$.

5(b)(i) Les inclusions $\{Z_n \leq x - 1/n\} \subset \{Z_n < x\} \subset \{Z_n \leq x\}$ résultent de $1/n > 0$. La croissance des probabilités donne l'encadrement demandé.

5(b)(ii) La borne inférieure tend vers $\Phi(x)$ par 5(a)(iii), appliquée à $x_n = x - 1/n$; la borne supérieure tend vers $\Phi(x)$ par (*). Le théorème d'encadrement donne donc $P(Z_n < x) \rightarrow \Phi(x)$.

5(c) Comme $a < b$, $\{Z_n \leq b\}$ est la réunion disjointe de $\{Z_n < a\}$ et $\{a \leq Z_n \leq b\}$. Ainsi $P(a \leq Z_n \leq b) = P(Z_n \leq b) - P(Z_n < a)$. Les deux termes tendent respectivement vers $\Phi(b)$ et $\Phi(a)$ d'après (*) et 5(b)(ii), ce qui donne la limite annoncée.

Deuxième partie : Applications en statistique

6(a) Une Bernoulli prend la valeur 1 avec probabilité p et 0 avec probabilité $1 - p$. Elle est bornée et tous ses moments existent. On a $E(X_i) = p$ et $X_i^2 = X_i$, donc $E(X_i^2) = p$. Par Koenig-Huygens, $V(X_i) = p - p^2 = p(1 - p)$.

6(b) L'identité $p(1 - p) = \frac{1}{4} - (p - \frac{1}{2})^2$ montre que $0 < p(1 - p) \leq 1/4$. En prenant les racines carrées positives, on obtient $0 < \sigma \leq 1/2$.

6(c) Par linéarité de l'espérance d'une somme finie,

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{np}{n} = p.$$

6(d) Les X_i sont indépendantes et admettent une variance. L'additivité des variances donne

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

7(a) Le théorème limite central s'applique puisque $0 < p < 1$ assure $\sigma^2 > 0$. Pour $Z_n = \sqrt{n}(\bar{X}_n - p)/\sigma$, la question 5(c), avec les bornes $-a < a$, donne $P(-a \leq Z_n \leq a) \rightarrow \Phi(a) - \Phi(-a)$.

7(b) Les quantités σ et $\sqrt{n}$ sont strictement positives. Par résolution de la double inégalité,

$$-a \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - p) \leq a \quad \Leftrightarrow \quad \bar{X}_n - a \frac{\sigma}{\sqrt{n}} \leq p \leq \bar{X}_n + a \frac{\sigma}{\sqrt{n}}.$$

Les événements ont donc même probabilité. La limite de 7(a) et la symétrie $\Phi(-a) = 1 - \Phi(a)$ donnent la couverture limite $2\Phi(a) - 1$.

7(c) Question adaptée au programme actuel : l'interprétation est formulée en termes de couverture d'un intervalle aléatoire.

Avec $a = 1,96$, le résultat de 7(b) donne une limite $2\Phi(1,96) - 1 \simeq 2 \times 0,975 - 1 = 0,95$. Avant le sondage, l'intervalle est aléatoire et sa probabilité de contenir le paramètre fixe p est donc voisine de 95% pour n grand. Après observation, ce calcul n'attribue pas une loi de probabilité au paramètre p . De plus, cet intervalle utilise encore la quantité inconnue σ .

7(d) Question adaptée au programme actuel : la borne correcte est « au moins », sans affirmation stricte valable pour tout p .

Comme $1,96\sigma \leq 0,98$, l'intervalle de 7(c) est inclus dans celui de 7(d) pour chaque issue. La probabilité de couverture du second est donc supérieure ou égale à celle du premier. On peut aussi calculer sa limite exactement : l'événement de couverture équivaut à $|Z_n| \leq 0,98/\sigma$, avec une borne positive fixe. Sa probabilité tend vers $2\Phi(0,98/\sigma) - 1$, qui est au moins $2\Phi(1,96) - 1$. Pour $p = 1/2$, $\sigma = 1/2$ et l'égalité est atteinte. Le niveau est strictement supérieur à cette valeur si $p \neq 1/2$.

8(a) Posons $t = \bar{X}_n$. Par définition,

$$V_n - \sigma^2 - \frac{1}{n} = t - t^2 - p + p^2 = (t - p) - (t - p)(t + p) = (t - p)(1 - t - p).$$

C'est l'identité demandée.

8(b) On a $0 \leq \bar{X}_n \leq 1$ et $0 < p < 1$, donc $|1 - \bar{X}_n - p| \leq 1$, en particulier ce module est au plus 2. L'inégalité triangulaire appliquée à 8(a) donne $|V_n - \sigma^2| \leq 2|\bar{X}_n - p| + 1/n$. Une borne plus fine, avec le coefficient 1, serait également possible.

8(c) Si $|V_n - \sigma^2| > \varepsilon$, alors 8(b) impose $2|\bar{X}_n - p| + 1/n > \varepsilon$, donc $|\bar{X}_n - p| > \varepsilon/2 - 1/(2n)$. Cette implication donne une inclusion d'événements, puis l'inégalité de probabilités demandée. Elle reste vraie si le seuil de droite est négatif.

8(d) Pour $n \geq 2/\varepsilon$, $1/(2n) \leq \varepsilon/4$, donc $\varepsilon/2 - 1/(2n) \geq \varepsilon/4$. Dépasser strictement le premier seuil implique de dépasser le second, ce qui établit l'inégalité.

8(e) Méthode 1 — Majoration explicite par Bienaymé-Tchebychev

Pour ces mêmes n , les questions 8(c) et 8(d), puis Bienaymé-Tchebychev et 6(d), donnent

$$0 \leq P(|V_n - \sigma^2| > \varepsilon) \leq P(|\bar{X}_n - p| > \varepsilon/4) \leq \frac{16\sigma^2}{n\varepsilon^2} \rightarrow 0.$$

La conclusion porte sur chaque $\varepsilon > 0$, sans utiliser de règle générale de calcul sur des convergences de variables aléatoires.

Méthode 2 — Loi faible des grands nombres

Pour $\varepsilon > 0$ fixé, on a $P(|V_n - \sigma^2| > \varepsilon) \leq P(|\bar{X}_n - p| > \varepsilon/4)$ dès que $n \geq 2/\varepsilon$. Les X_i étant indépendantes et de même loi de Bernoulli, la loi faible des grands nombres donne directement $P(|\bar{X}_n - p| > \varepsilon/4) \rightarrow 0$. L'encadrement conclut à la même limite. Cette seconde méthode utilise un résultat de convergence déjà établi, tandis que la première fournit en plus une borne numérique explicite en fonction de n .

9(a)(i) Question adaptée au programme actuel. La comparaison du sujet d'origine avec $(1 + \varepsilon)x$ n'est pas valide pour tout $x < 0$. L'encadrement avec $|x|$ rectifie ce défaut.

On a $V_n \geq 1/n > 0$ et $\sigma > 0$, donc $R_n > 0$ et $W_n \leq x$ équivaut à $Z_n \leq xR_n$. Sur l'événement $A_n = \{|R_n - 1| \leq \varepsilon\}$,

$$xR_n = x + x(R_n - 1) \leq x + |x|\varepsilon.$$

Ainsi $\{W_n \leq x\} \cap A_n \subset \{Z_n \leq x + \varepsilon|x|\}$. En réunissant avec A_n^c , puis en majorant la probabilité d'une union par la somme, on obtient

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

9(a)(ii) Question adaptée au programme actuel : on contrôle directement la fluctuation de R_n des deux côtés.

Comme $R_n + 1 \geq 1$, $|R_n - 1| > \varepsilon$ implique $|R_n^2 - 1| > \varepsilon$. Or $R_n^2 = V_n/\sigma^2$, donc

$$P(|R_n - 1| > \varepsilon) \leq P(|V_n - \sigma^2| > \sigma^2\varepsilon) \rightarrow 0$$

d'après 8(e), puisque $\sigma^2\varepsilon > 0$. Aucune propriété générale de composition de convergences n'est nécessaire.

9(a)(iii) Question adaptée au programme actuel : le seuil est corrigé comme en 9(a)(i).

Pour x et ε fixés, $x + \varepsilon|x|$ est un réel fixe. Le théorème limite central de 2(b) s'applique donc à ce seuil et donne $P(Z_n \leq x + \varepsilon|x|) \rightarrow \Phi(x + \varepsilon|x|)$.

9(a)(iv) Question adaptée au programme actuel : on utilise le seuil corrigé.

Pour n assez grand, 9(a)(ii) donne $P(|R_n - 1| > \varepsilon) \leq \varepsilon/2$; 9(a)(iii) donne simultanément $P(Z_n \leq x + \varepsilon|x|) \leq \Phi(x + \varepsilon|x|) + \varepsilon/2$. En prenant le maximum des deux rangs nécessaires, l'inégalité de 9(a)(i) entraîne la majoration demandée avec une erreur totale au plus ε .

9(b) Question adaptée au programme actuel : la borne inférieure est corrigée avec la même convention de signe.

Résultat admis dans cette version adaptée : pour tout $\varepsilon > 0$, à partir d'un certain rang, $P(W_n \leq x) \geq \Phi(x - \varepsilon|x|) - \varepsilon$. Ce résultat permet de compléter la majoration de 9(a)(iv).

Pour vérifier qu'il s'agit bien d'un encadrement particulier, sur A_n on a $xR_n \geq x - \varepsilon|x|$, d'où $\{Z_n \leq x - \varepsilon|x|\} \cap A_n \subset \{W_n \leq x\}$. Cela donne $P(W_n \leq x) \geq P(Z_n \leq x - \varepsilon|x|) - P(A_n^c)$, puis la borne admise à partir des limites déjà établies.

Pour conclure précisément, soit $\eta > 0$. Par continuité de Φ en x , on peut choisir $\varepsilon > 0$ assez petit pour que $\varepsilon < \eta/2$ et $|\Phi(x \pm \varepsilon|x|) - \Phi(x)| < \eta/2$. Les bornes inférieure et supérieure donnent alors, pour tout n assez grand, $|P(W_n \leq x) - \Phi(x)| \leq \eta$. Ainsi $P(W_n \leq x) \rightarrow \Phi(x)$ pour tout réel x , y compris $x = 0$.

10(a) Comme $\sqrt{V_n}/\sqrt{n} > 0$,

$$p \geq \bar{X}_n - x \frac{\sqrt{V_n}}{\sqrt{n}} \iff \frac{\sqrt{n}}{\sqrt{V_n}}(\bar{X}_n - p) \leq x \iff W_n \leq x.$$

La limite de cette probabilité vaut $\Phi(x)$ par 9(b). L'intervalle unilatéral ainsi construit a donc la couverture asymptotique annoncée pour chaque seuil x fixé.

10(b) Question adaptée au programme actuel : l'affirmation d'une probabilité de victoire après observation est remplacée par une lecture correcte de la borne de confiance.

Avec les valeurs observées, l'équation $0,52 - x\sqrt{0,2506}/\sqrt{1000} = 1/2$ a pour solution

$$x_{\text{obs}} = \frac{\sqrt{1000}}{\sqrt{0,2506}}(0,52 - 0,5) \simeq 1,26.$$

Pour le seuil fixé $x = 1,26$, la procédure de 10(a) possède une couverture asymptotique $\Phi(1,26) \simeq 0,896$. Sur cet échantillon, sa borne inférieure est voisine de $1/2$. Le seuil exact calculé à partir des données est environ 1,2634 ; son arrondi à deux décimales est 1,26,

rectifiant la valeur 1,27 imprimée dans le sujet source. Le résultat du sondage situe donc le seuil de victoire au niveau de la borne inférieure d'une procédure de confiance unilatérale d'environ 90%.

Dans le modèle, p est un nombre fixe inconnu. L'événement $p \geq 1/2$ est donc soit vrai, soit faux ; le modèle ne lui attribue pas une probabilité postérieure de 90%. Ce pourcentage décrit la couverture de la procédure sur des échantillons répétés, pour un seuil fixé. Le calcul de x_{obs} sert à lire les bornes de cette famille de procédures ; il ne prouve pas une couverture conditionnelle après sélection d'un seuil à partir des données.

11(a) Question adaptée au programme actuel : $q < 1$ est précisé pour les divisions ultérieures et l'indépendance des deux familles est explicitée.

Si $X_i = 1$, alors $Y_i = 1$. Si $X_i = 0$, alors $Y_i = T_i$. Ainsi Y_i prend seulement les valeurs 0 et 1, et les deux événements disjoints $\{X_i = 1\}$ et $\{X_i = 0, T_i = 1\}$ décrivent $\{Y_i = 1\}$. L'indépendance donne

$$r = P(Y_i = 1) = p + (1 - p)q = q + (1 - q)p.$$

On a donc $Y_i \sim \mathcal{B}(r)$. Puisque $0 < p < 1$ et $0 \leq q < 1$, $0 < r < 1$. En outre les couples (X_i, T_i) sont indépendants entre indices différents, donc le lemme des coalitions assure l'indépendance des Y_i . Leurs lois sont identiques.

Le résultat admis dans le sujet est le suivant : pour tout $x \in \mathbb{R}$, $P(\sqrt{n}(\bar{Y}_n - r)/\sqrt{U_n} \leq x) \rightarrow \Phi(x)$. Il est applicable à ces Bernoulli de paramètre $r \in]0, 1[$ et intervient dans 11(b). Le terme $1/n$ assure $U_n > 0$ même si toutes les réponses observées sont identiques.

11(b) Comme $1 - q > 0$, l'événement étudié est équivalent successivement à

$$(1 - q)p \geq \bar{Y}_n - x \frac{\sqrt{U_n}}{\sqrt{n}} - q,$$

$$r \geq \bar{Y}_n - x \frac{\sqrt{U_n}}{\sqrt{n}},$$

$$\frac{\sqrt{n}}{\sqrt{U_n}}(\bar{Y}_n - r) \leq x.$$

Le résultat particulier admis après 11(a) donne donc exactement la limite $\Phi(x)$.

11(c) Question adaptée au programme actuel : on détermine une borne de confiance, sans attribuer une probabilité au paramètre fixe p .

Réolvons l'équation $\frac{1}{1-q}(\bar{y}_n - x \frac{\sqrt{u_n}}{\sqrt{n}} - q) = 1/2$. Elle équivaut à

$x \frac{\sqrt{u_n}}{\sqrt{n}} = \bar{y}_n - q - (1-q)/2 = \bar{y}_n - (1+q)/2$. Puisque $u_n > 0$, son unique solution est

$$x_{\text{obs}} = \frac{\sqrt{n}}{\sqrt{u_n}} \left(\bar{y}_n - \frac{1+q}{2} \right).$$

La procédure de borne inférieure de 11(b), considérée au seuil fixé égal à cette valeur numérique, a pour niveau de confiance asymptotique $\Phi(x_{\text{obs}})$. Il s'agit, comme en 10(b), d'une lecture de la famille de bornes de confiance. L'observation ne crée pas une loi de probabilité sur p .

11(d) Question adaptée au programme actuel : la conclusion est formulée au niveau de confiance correspondant.

Ici $(1+q)/2 = 1,04/2 = 0,52 = \bar{y}_n$, donc $x_{\text{obs}} = 0$ et $\Phi(x_{\text{obs}}) = \Phi(0) = 1/2$. L'estimation corrigée de p vaut également $(0,52 - 0,04)/(1 - 0,04) = 0,48/0,96 = 1/2$. La correction de la déclaration modifie donc la lecture de la borne de victoire : le niveau associé est 50%, au lieu d'environ 90% sans cette correction. Aucun de ces deux nombres n'est une probabilité postérieure de victoire dans le modèle posé.

Troisième partie : Démonstration du théorème limite central

Question adaptée au programme actuel : le préambule précise le caractère absolu du

moment d'ordre trois et définit la notation C^3 utilisée dans les calculs particuliers suivants.

12(a)(i) Les fonctions intégrées sont polynomiales sur $[0, 1]$, donc les trois intégrations par parties suivantes sont légitimes. À chaque étape, les termes de bord sont nuls : à gauche une puissance strictement positive de u s'annule, à droite une puissance strictement positive de $1 - u$ s'annule.

$$\int_0^1 u^3(1-u)^3 du = \frac{3}{4} \int_0^1 u^4(1-u)^2 du,$$

$$\int_0^1 u^4(1-u)^2 du = \frac{2}{5} \int_0^1 u^5(1-u) du,$$

$$\int_0^1 u^5(1-u) du = \frac{1}{6} \int_0^1 u^6 du.$$

Le produit des coefficients est $\frac{3}{4} \cdot \frac{2}{5} \cdot \frac{1}{6} = \frac{1}{20}$, d'où l'identité demandée.

12(a)(ii) Comme $\int_0^1 u^6 du = [u^7/7]_0^1 = 1/7$, la question précédente donne

$$\int_0^1 u^3(1-u)^3 du = 1/140.$$

Méthode 2 — Développement polynomial

On peut vérifier ce calcul sans intégration par parties : $u^3(1-u)^3 = u^3 - 3u^4 + 3u^5 - u^6$, donc l'intégrale vaut $1/4 - 3/5 + 1/2 - 1/7 = (35 - 84 + 70 - 20)/140 = 1/140$. Cette vérification algébrique est utile pour contrôler la constante de normalisation ; la méthode imposée pour 12(a)(i) reste celle donnée plus haut.

12(b) Sur $] -\infty, 0[$ et $]1, +\infty[$, la fonction est constante, donc continue. Sur $]0, 1[$, elle est l'intégrale à borne variable d'une fonction continue, donc continue. Enfin

$$\lim_{z \rightarrow 0^+} 140 \int_0^z u^3(1-u)^3 du = 0 = h(0) \text{ et}$$

$\lim_{z \rightarrow 1^-} 140 \int_0^z u^3(1-u)^3 du = 140/140 = 1 = h(1)$. Les limites de l'autre côté sont données par les morceaux constants. Ainsi h est continue sur tout $\mathbb{R}$.

12(c) Pour $0 < z < 1$, l'intégrande est positif ou nul sur $[0, 1]$, donc

$$0 \leq \int_0^z u^3(1-u)^3 du \leq \int_0^1 u^3(1-u)^3 du = 1/140. \text{ Après multiplication par } 140, \text{ on}$$

obtient $0 \leq h(z) \leq 1$. À l'extérieur de cet intervalle, h vaut 0 ou 1, donc la conclusion reste vraie.

Résultat admis dans cette version adaptée, comme dans le sujet source : les dérivées h' , h'' , h''' existent et sont continues sur $\mathbb{R}$. Elles sont nulles hors de $[0, 1]$ et continues sur ce segment, donc bornées. En particulier on peut choisir un majorant fini $M_{h''''}$, utilisé en 13(a)(ii).

13(a)(i) Puisque $a_n > 0$, l'application $z \mapsto (z-x)/a_n$ est affine et continue. La continuité de h donne celle de $g_n = 1 - h((\cdot - x)/a_n)$. L'encadrement $0 \leq h \leq 1$ donne $0 \leq g_n \leq 1$.

13(a)(ii) Trois dérivations successives, avec le même facteur constant $1/a_n$ à chaque fois, donnent

$$g'_n(z) = -\frac{1}{a_n} h' \left(\frac{z-x}{a_n} \right), \quad g''_n(z) = -\frac{1}{a_n^2} h'' \left(\frac{z-x}{a_n} \right), \quad g'''_n(z) = -\frac{1}{a_n^3} h''' \left(\frac{z-x}{a_n} \right).$$

Ainsi $|g'''_n(z)| \leq a_n^{-3} M_{h''''} = n^{1/4} M_{h''''}$ pour tout z . On peut choisir $M_{g'''_n} = n^{1/4} M_{h''''}$, ce qui satisfait l'inégalité demandée. Il ne s'agit pas de supposer une formule générale pour les dérivées d'une composée : seuls trois calculs par la règle de dérivation usuelle sont utilisés.

13(b)(i) Si $z \leq x$, alors $(z-x)/a_n \leq 0$, donc $h((z-x)/a_n) = 0$ et $g_n(z) = 1$. Si $z \geq x + a_n$, alors $(z-x)/a_n \geq 1$, donc $h((z-x)/a_n) = 1$ et $g_n(z) = 0$.

13(b)(ii) On raisonne pour chaque valeur réelle t prise par X . Si $t \leq x$, l'indicatrice de gauche et $g_n(t)$ valent 1, tout comme l'indicatrice de droite. Si $x < t \leq x + a_n$, l'indicatrice de gauche est nulle, celle de droite vaut 1, et $0 \leq g_n(t) \leq 1$. Si $t > x + a_n$, les trois quantités sont nulles. L'encadrement est donc valable pour chaque issue.

13(c) Question adaptée au programme actuel. Dans la borne inférieure, le décalage correct est $+a_n$; le décalage $-a_n$ imprimé dans le sujet source inverse le rôle de cette borne.

La question 13(b)(ii), appliquée à $X = Z_n$, donne $\mathbf{1}_{\{Z_n \leq x\}} \leq g_n(Z_n)$. D'autre part, si $Z_n > x$, alors $Z_n + a_n > x + a_n$ et $g_n(Z_n + a_n) = 0$; si $Z_n \leq x$, on a simplement $g_n(Z_n + a_n) \leq 1$. Donc $g_n(Z_n + a_n) \leq \mathbf{1}_{\{Z_n \leq x\}}$. Toutes ces variables sont bornées et admettent une espérance. La croissance de l'espérance, avec $E(\mathbf{1}_A) = P(A)$, donne

$$E(g_n(Z_n + a_n)) \leq P(Z_n \leq x) \leq E(g_n(Z_n)).$$

14(a) Posons $J = \frac{1}{2} \int_z^{z+u} (z+u-t)^2 g'''(t) dt$. L'intégration par parties donne

$$J = \left[\frac{1}{2} (z+u-t)^2 g''(t) \right]_z^{z+u} + \int_z^{z+u} (z+u-t) g''(t) dt.$$

Le terme de bord vaut $-u^2 g''(z)/2$. Une seconde intégration par parties donne

$$\begin{aligned} \int_z^{z+u} (z+u-t) g''(t) dt &= [(z+u-t) g'(t)]_z^{z+u} + \int_z^{z+u} g'(t) dt = -u g'(z) + g(z+u) \\ &\quad - g(z). \end{aligned}$$

En réunissant ces égalités, on obtient la formule demandée. Elles sont valides aussi si $u < 0$, avec la convention des intégrales orientées, et si $u = 0$, les deux membres étant nuls. La fonction g étant de classe C^3 , toutes les intégrations par parties sur les segments concernés sont justifiées. Le résultat est ici démontré directement; aucune formule de reste de développement n'est utilisée comme résultat de cours.

14(b) La question 14(a) s'écrit

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

Écrivons la même égalité avec v à la place de u , puis soustrayons. Les termes $g(z)$ disparaissent et il reste $g'(z)(u-v) + \frac{1}{2}g''(z)(u^2-v^2)$ plus la différence des deux intégrales, qui est précisément $R(z, u, v)$.

14(c) Pour $u \geq 0$, la majoration de $|g'''|$ donne

$$\left| \frac{1}{2} \int_z^{z+u} (z+u-t)^2 g'''(t) dt \right| \leq \frac{1}{2} M_{g'''} \int_z^{z+u} (z+u-t)^2 dt = \frac{1}{6} M_{g'''} u^3.$$

Pour $u < 0$, on renverse les bornes et on obtient de même

$\frac{1}{2} M_{g'''} \int_{z+u}^z (z+u-t)^2 dt = \frac{1}{6} M_{g'''} |u|^3$. La majoration est donc vraie pour tout u . En

l'appliquant aussi à v et en utilisant l'inégalité triangulaire, on obtient

$$|R(z, u, v)| \leq \frac{1}{6} M_{g'''} (|u|^3 + |v|^3).$$

15(a)(i) Les Y_i sont indépendantes et chacune suit une loi normale. La stabilité des sommes de normales indépendantes, admise au programme, assure que

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

suit une loi normale. Son espérance est 0 et sa variance est n , par additivité des variances. Sa loi est donc $\mathcal{N}(0, n)$.

15(a)(ii) Une transformation affine d'une variable normale est normale. Le facteur $1/\sqrt{n}$ multiplie l'espérance par $1/\sqrt{n}$ et la variance par $1/n$. Par conséquent $T_n \sim \mathcal{N}(0, 1)$, pour tout n .

15(b)(i) La définition donne

$$\sqrt{n}W_k + Y_k = \sum_{i=1}^k Y_i + \sum_{i=k+1}^n X_i,$$

$$\sqrt{n}W_{k+1} + X_{k+1} = \sum_{i=1}^k Y_i + \sum_{i=k+2}^n X_i + X_{k+1}.$$

Les seconds membres sont identiques. La division par $\sqrt{n} > 0$ établit l'égalité, y compris pour $k = 1$ et $k = n - 1$ grâce à la convention des sommes vides.

15(b)(ii) Question adaptée au programme actuel. La somme télescopique doit contenir les n remplacements ; la borne supérieure $n - 1$ du sujet source omet le dernier terme.

Posons, pour $j \in \{0, \dots, n\}$,

$$H_j = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^j Y_i + \sum_{i=j+1}^n X_i \right).$$

Alors $H_0 = Z_n$, $H_n = T_n$, $H_{k-1} = W_k + X_k/\sqrt{n}$ et $H_k = W_k + Y_k/\sqrt{n}$. La somme finie

$$\sum_{k=1}^n (g_n(H_{k-1}) - g_n(H_k))$$

est télescopique et vaut $g_n(H_0) - g_n(H_n)$. Cela donne exactement l'identité demandée. Ce calcul explicite les termes extrêmes et contrôle le nombre de remplacements.

15(c)(i) La variable W_k ne dépend ni de X_k ni de Y_k . Les familles X_i et Y_i étant mutuellement indépendantes avec les indépendances internes précisées, le lemme des coalitions montre que W_k est indépendante du couple (X_k, Y_k) . Les variables X_k et Y_k sont intégrables puisqu'elles ont un moment d'ordre deux ; g'_n est bornée. Les produits étudiés sont donc intégrables. Par factorisation des espérances,

$$E((X_k - Y_k)g'_n(W_k)) = (E(X_k) - E(Y_k))E(g'_n(W_k)) = 0,$$

car les deux variables sont centrées.

15(c)(ii) Les carrés X_k^2 et Y_k^2 sont intégrables ; g''_n est bornée. La même indépendance donne

$$E((X_k^2 - Y_k^2)g''_n(W_k)) = (E(X_k^2) - E(Y_k^2))E(g''_n(W_k)) = 0.$$

En effet, les espérances sont nulles et les variances valent 1, donc les deux seconds moments valent 1.

15(c)(iii) Appliquons l'identité démontrée en 14(b) avec $g = g_n$, $z = W_k$, $u = X_k/\sqrt{n}$ et $v = Y_k/\sqrt{n}$. Elle donne une différence égale à

$$\frac{1}{\sqrt{n}}(X_k - Y_k)g'_n(W_k) + \frac{1}{2n}(X_k^2 - Y_k^2)g''_n(W_k) + R\left(W_k, \frac{X_k}{\sqrt{n}}, \frac{Y_k}{\sqrt{n}}\right).$$

Les deux premiers termes sont intégrables et d'espérance nulle par 15(c)(i) et 15(c)(ii). Le dernier est intégrable par 14(c), car $E(|X_k|^3) < +\infty$ et $E(|Y_k|^3) < +\infty$. Pour ce dernier fait, le transfert normal donne

$$E(|Y_k|^3) = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} t^3 e^{-t^2/2} dt = \frac{4}{\sqrt{2\pi}} < +\infty.$$

L'intégrale se calcule d'abord sur $[0, A]$: une primitive de $t^3 e^{-t^2/2}$ est $-(t^2 + 2)e^{-t^2/2}$, puis $A \rightarrow +\infty$ donne la valeur 2 pour l'intégrale. Il n'y a aucun changement de variable non indiqué. On peut donc prendre les espérances dans l'identité, ce qui donne l'égalité demandée.

15(d) Toutes les fonctions g_n sont bornées entre 0 et 1, donc les espérances de gauche existent. Prenons l'espérance dans l'identité de 15(b)(ii) et utilisons 15(c)(iii) pour chacun des n termes. La linéarité d'une somme finie, puis les inégalités

$$\left| \sum a_k \right| \leq \sum |a_k|$$

et $|E(R_k)| \leq E(|R_k|)$, donnent

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

Les restes sont intégrables comme vérifié en 15(c)(iii).

15(e) La majoration de 14(c) donne, pour chaque k ,

$$E \left(\left| R \left(W_k, \frac{X_k}{\sqrt{n}}, \frac{Y_k}{\sqrt{n}} \right) \right| \right) \leq \frac{M_{g_n}''' }{6n^{3/2}} (E(|X_k|^3) + E(|Y_k|^3)).$$

Les X_k ont la même loi et les Y_k aussi. Les espérances de droite sont donc indépendantes de k . La somme comporte n termes, d'où le facteur $n/n^{3/2} = 1/\sqrt{n}$ et

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

15(f) Le nombre $C = E(|X_1|^3) + E(|Y_1|^3)$ est fini et ne dépend pas de n . Par 13(a)(ii), le majorant de 15(e) est au plus $CM_{h'''}/(6n^{1/4})$, qui tend vers zéro. Par encadrement, $E(g_n(Z_n)) - E(g_n(T_n)) \rightarrow 0$.

16(a) La question 13(b)(ii), appliquée à $X = T_n$, donne

$\mathbf{1}_{\{T_n \leq x\}} \leq g_n(T_n) \leq \mathbf{1}_{\{T_n \leq x+a_n\}}$. Les trois variables sont bornées, donc intégrables. La croissance de l'espérance donne l'encadrement demandé.

16(b) Par 15(a)(ii), la fonction de répartition de chaque T_n est exactement Φ . Ainsi l'encadrement précédent s'écrit $\Phi(x) \leq E(g_n(T_n)) \leq \Phi(x + a_n)$. Or $a_n = n^{-1/12} \rightarrow 0$ et Φ est continue en x , donc $\Phi(x + a_n) \rightarrow \Phi(x)$. Le théorème d'encadrement donne $E(g_n(T_n)) \rightarrow \Phi(x)$.

16(c) Question adaptée au programme actuel : une indication permet également de justifier la limite de la borne inférieure corrigée de 13(c) et d'achever la démonstration annoncée.

D'après 15(f), $E(g_n(Z_n)) - E(g_n(T_n)) \rightarrow 0$, tandis que 16(b) donne $E(g_n(T_n)) \rightarrow \Phi(x)$. En additionnant ces deux limites de suites réelles, on obtient d'abord $E(g_n(Z_n)) \rightarrow \Phi(x)$.

Pour l'autre borne, posons $\tilde{g}_n(z) = g_n(z + a_n)$. Cette fonction est bornée, de classe C^3 , et ses dérivées ont les mêmes majorants que celles de g_n . La démonstration par sommes finies et indépendance de 15(b) à 15(f) s'applique donc avec $\tilde{g}_n$ et donne

$$|E(\tilde{g}_n(Z_n)) - E(\tilde{g}_n(T_n))| \leq \frac{CM_{h'''}}{6n^{1/4}} \rightarrow 0.$$

La fonction $\tilde{g}_n$ vaut 1 pour $z \leq x - a_n$, vaut 0 pour $z \geq x$ et reste entre 0 et 1. Par conséquent,

$$\Phi(x - a_n) = P(T_n \leq x - a_n) \leq E(\tilde{g}_n(T_n)) \leq P(T_n \leq x) = \Phi(x).$$

La continuité de Φ donne $E(\tilde{g}_n(T_n)) \rightarrow \Phi(x)$, puis $E(g_n(Z_n + a_n)) = E(\tilde{g}_n(Z_n)) \rightarrow \Phi(x)$. Enfin 13(c) encadre $P(Z_n \leq x)$ entre deux suites tendant vers $\Phi(x)$. On conclut $P(Z_n \leq x) \rightarrow \Phi(x)$ pour tout réel x : le théorème limite central est ainsi démontré sous l'hypothèse supplémentaire de moment absolu d'ordre trois, dans le cas centré de variance un.

Relevé des adaptations

3(c)(ii) — Code adapté en Python. Les quatre branches et les seuils du générateur sont conservés ; les imports et le rôle de la commande aléatoire sont fournis.

3(e) — Les hypothèses justifiant l'approximation normale sont rappelées pour conserver le guidage attendu dans le programme actuel.

4(d) — La construction des majorants par borne supérieure est explicitée pour éviter de supposer une théorie de convergence uniforme connue.

7(c), 7(d), 10(b), 11(c), 11(d) — La lecture statistique est formulée en termes de couverture d'intervalles aléatoires et de niveaux de confiance. Aucun modèle de probabilité sur le paramètre fixe p n'étant donné, les affirmations de probabilité de victoire après observation sont remplacées par les bornes de confiance correspondant aux mêmes données. En 7(d), « au moins » remplace une comparaison strictement supérieure incorrecte lorsque $p = 1/2$. En 10(b), le seuil numérique est arrondi correctement à 1,26 et la valeur normale correspondante est environ 0,896, soit environ 90%.

9(a)(i) à 9(b) — Les seuils sont remplacés par $x \pm \varepsilon|x|$ et la fluctuation de $\sqrt{V_n}/\sigma$ est contrôlée des deux côtés. L'encadrement imprimé avec $(1 + \varepsilon)x$ ne traite pas correctement les réels négatifs. La preuve particulière conserve les inclusions d'événements, le théorème limite central et la continuité de Φ .

11(a) et préambule de 11 — Le domaine $q \in [0, 1[$ et l'indépendance des deux familles sont explicités : ils assurent les divisions par $1 - q$ et l'indépendance des observations Y_i .

Préambule de 12 — Le moment d'ordre trois est précisé comme un moment absolu fini, nécessaire à la majoration de 15(e). La notation C^3 est définie. Le calcul intégral particulier de 14 reste entièrement démontré par intégrations par parties.

13(c) — Le décalage $Z_n - a_n$ du sujet source est rectifié en $Z_n + a_n$ pour que l'espérance constitue bien une borne inférieure de $P(Z_n \leq x)$.

15(b)(ii) — La somme comporte les indices de 1 à n , au lieu de 1 à $n - 1$: les n variables doivent toutes être remplacées. La définition de W_k est écrite en une seule formule avec sommes vides, sans changer les variables définies à l'origine.

16(c) — Une indication complète la conclusion en traitant aussi $g_n(Z_n + a_n)$, afin de justifier les deux limites nécessaires à la conclusion finale.