Examen du cours de Probabilités

Mardi 16 Janvier 2018 (8h30-11h30)

Polycopié et notes de cours autorisés. Téléphones, calculatrices et ordinateurs interdits.

Barême indicatif. L'ensemble de l'examen sera noté sur une trentaine de points, et les points seront approximativement répartis de la façon suivante : 1/5 pour l'exercice 1, 3/10 pour l'exercice 2 et 1/2 pour le problème. Il n'est donc pas nécessaire de tout faire pour avoir la note maximale.

Exercice 1. Soient U_1, U_2 deux variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On pose $X_1 = 1/U_1$ et $X_2 = 1/(U_1U_2)$.

1. Calculer la densité de la loi de X_1.
2. Pour quels réels α a-t-on $\mathbb{E}[|X_1|^\alpha] < +\infty$? $\mathbb{E}[|X_2|^\alpha] < +\infty$?
3. Calculer la loi de (X_1, X_2) et montrer qu'elle admet une densité que l'on précisera. Les variables aléatoires X_1 et X_2 sont-elles indépendantes ?
4. On considère maintenant $U_1, \dots, U_n$, n variables aléatoires indépendantes uniformes sur $[0,1]$ et $X_k = \prod_{i=1}^k U_i^{-1}$ pour $1 \leq k \leq n$. Calculer la loi de $(X_1, \dots, X_n)$.

⋆ ⋆ ⋆

Exercice 2. Soit $a > 0$. Soit $(U_n)_{n\geq 1}$ une suite i.i.d. de loi uniforme sur $[0,1]$.

1. Calculer la fonction de répartition de la loi $\beta(a,1)$. Montrer que $U_1^{1/a} \sim \beta(a,1)$.
2. En déduire une méthode de rejet qui permet de simuler la loi $\beta(a,b)$ pour $b \geq 1$, à partir de la suite $(U_n)_{n\geq 1}$ et d'une autre suite $(V_n)_{n\geq 1}$ i.i.d. de loi uniforme sur $[0,1]$, indépendante de $(U_n)_{n\geq 1}$.

On souhaite calculer $I = \int_0^1 \frac{1}{\sqrt{x}}\varphi(x)dx$ avec $\varphi(x) = \sqrt{1+x}$ par méthode de Monte-Carlo. On propose les deux estimateurs suivants

$$I_n^1 = \frac{1}{n}\sum_{i=1}^n 2\varphi(U_i^2),\ I_n^2 = \frac{1}{n}\sum_{i=1}^n \frac{\varphi(U_i)}{\sqrt{U_i}}.$$

3. Dire si I_n^1 et I_n^2 convergent vers I et en quel sens.
4. La variable $\varphi(U_i^2)$ est elle de carré intégrable ? Et $\frac{\varphi(U_i)}{\sqrt{U_i}}$? Faut il préférer I_n^1 ou I_n^2 pour approcher I ?
5. Montrer que $\mathbf{Var}(2\varphi(U_i^2)) \leq 4\int_0^1 (\varphi(x)-1)^2 \frac{1}{2\sqrt{x}}dx$, puis que $\mathbf{Var}(2\varphi(U_i^2)) \leq \frac{1}{5}$ (Indication : majorer $|\varphi'(x)|$ pour $x \in [0,1]$). En déduire un intervalle de confiance asymptotique à 95% pour I en fonction de I_n^1.
6. Montrer que $\mathbf{Var}(2[\varphi(U_i^2) - U_i^2/2]) \leq 4\int_0^1 (\varphi(x) - 1 - x/2)^2 \frac{1}{2\sqrt{x}}dx$, puis majorer cette variance en utilisant que $|\varphi''(x)| \leq 1/4$ pour $x \in [0,1]$.
7. On pose $J_n = \frac{1}{n}\sum_{i=1}^n U_i^2$. Donner le comportement asymptotique de $I_n^1 - J_n + 1/3$, puis construire un intervalle de confiance à 95% pour I en fonction de I_n^1 et J_n. Comparer la largeur de cet intervalle de confiance avec celle de l'intervalle obtenu uniquement avec I_n^1.

⋆ ⋆ ⋆

Problème. (Résultats identiques à une élection)

Dans ce problème, nous cherchons à estimer la probabilité pour qu'à une même élection, on puisse observer deux fois le même score dans deux villes (ou deux régions) différentes. Les scores publiés sont donnés en pourcentage et arrondis avec deux chiffres après la virgule, c'est à dire en arrondissant à $\epsilon_0 = 10^{-4}$ près.

Première partie : scrutin à 2 choix

On commence par considérer le cas d'un scrutin à deux issues (référendum ou 2ème tour de l'élection présidentielle en France), et on note $(X_i)_{1\leq i\leq n}$ et $(Y_i)_{1\leq i\leq m}$ les votes des électeurs dans chacune des zones. On suppose que ces votes sont indépendants, que X_i suit la loi de Bernoulli de paramètre $p \in]0,1[$, et que Y_i suit la loi de Bernoulli de paramètre $q \in]0,p]$. On suppose également que le nombre de votants dans chaque zone est de même ordre de grandeur, si bien qu'on prendra $m = \lfloor \alpha n \rfloor$ (i.e. le plus grand entier inférieur ou égal à αn) avec $\alpha > 0$. Enfin, on pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $\bar{Y}_m = \frac{1}{m}\sum_{i=1}^m Y_i$. Dans ce cas, on majorera la probabilité d'avoir deux scores identiques par $\mathbb{P}(|\bar{X}_n - \bar{Y}_{\lfloor \alpha n\rfloor}| \leq \epsilon_0)$, avec $\epsilon_0 = 10^{-4}$.

1. Donner le comportement asymptotique lorsque $n \to +\infty$ de $(\bar{X}_n)_{n\geq 1}$ et $(\bar{Y}_{\lfloor \alpha n\rfloor})_{n\geq 1}$.
2. Donner le comportement asymptotique de $\sqrt{n}(\bar{X}_n - p)$ puis de $\sqrt{n}(\bar{Y}_{\lfloor \alpha n\rfloor} - q)$.
3. Montrer que $\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n\rfloor} + q - p)$ converge en loi vers $\mathcal{N}(0,1)$.
4. On suppose $p = q$ et que n est assez grand pour supposer que $\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n\rfloor}) \overset{loi}{\approx} \mathcal{N}(0,1)$. Montrer que
$$\mathbb{P}(|\bar{X}_n - \bar{Y}_{\lfloor \alpha n\rfloor}| \leq \epsilon_0) \approx \sqrt{\frac{2}{\pi}}\int_0^{\sqrt{\frac{n}{p(1-p)(1+1/\alpha)}}\epsilon_0} e^{-x^2/2}dx \leq \sqrt{\frac{n}{p(1-p)(1+1/\alpha)}}\epsilon_0.$$
Donner un majorant de la probabilité pour que deux villes de 500000 habitants obtiennent le même score lorsque $p = 1/2$.
5. Pour $G \sim \mathcal{N}(0,1)$, montrer que pour $x < -1$, $\mathbb{P}(G < x) \leq \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.
On suppose $\Delta = p - q - \epsilon_0 > 0$ et que n est assez grand pour supposer $\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n\rfloor} + q - p) \overset{loi}{\approx} \mathcal{N}(0,1)$ et pour avoir $\Delta\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}} > 1$. Montrer que
$$\mathbb{P}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n\rfloor} \leq \epsilon_0) \leq \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{2\Delta^2 n}{1+\frac{1}{\alpha}}\right).$$
On suppose $\Delta = 0.01$. Donner un majorant de la probabilité pour que deux villes de 500000 habitants obtiennent le même score et commenter.

Deuxième partie : scrutin à d choix, $d \geq 2$

6. On considère $G_1, \ldots, G_d \sim \mathcal{N}(0,1)$ indépendantes si bien que $G \sim \mathcal{N}_d(0, I_d)$. Donner la densité du vecteur G.
7. On considère Γ une matrice carrée symétrique définie positive de dimension d, et C une matrice carrée telle que $CC^\top = \Gamma$ ($C^\top$ désigne la matrice transposée de C). On rappelle que $CG \sim \mathcal{N}_d(0,\Gamma)$. Calculer $f_\Gamma(x), x \in \mathbb{R}^d$ la densité de $\mathcal{N}_d(0,\Gamma)$. Vérifier que $f_\Gamma(0) = \max_{x\in\mathbb{R}^d} f_\Gamma(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Gamma)}}$.

On suppose désormais que les votes $(X_i)_{1\leq i\leq n}$ et $(Y_i)_{1\leq i\leq m}$ sont des variables aléatoires à valeurs dans $\{1,\dots,d\}$. Les votes sont toujours supposés indépendants, et on suppose qu'ils ont même loi : on note pour $k\in\{1,\dots,d\}$, $p_k=\mathbb{P}(X_i=k)=\mathbb{P}(Y_i=k)>0$. On pose également

$$\hat{p}_k^n=\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i=k\}},\ \hat{q}_k^m=\frac{1}{m}\sum_{i=1}^m \mathbf{1}_{\{Y_i=k\}}.$$

Ainsi, p, $\hat{p}^n$ et $\hat{q}^m$ sont des vecteurs de $\mathbb{R}^d$.

8. Calculer Γ^d, la matrice de covariance de $(\mathbf{1}_{X_1=1},\dots,\mathbf{1}_{X_1=d})$. Est-elle inversible ?

On admettra par la suite que $\det(\tilde{\Gamma}^{d-1})=\prod_{k=1}^d p_k>0$, où $\tilde{\Gamma}^{d-1}=(\Gamma_{ij}^d)_{1\leq i,j\leq d-1}$.

9. Donner le comportement asymptotique de $\sqrt{n}(\hat{p}^n-p)$ puis de $\sqrt{n}(\hat{q}^{\lfloor\alpha n\rfloor}-p)$. En déduire que $\sqrt{n}(\hat{p}^n-\hat{q}^{\lfloor\alpha n\rfloor})$ converge en loi vers un vecteur gaussien X centré dont on précisera la matrice de covariance à l'aide de Γ^d.
10. En déduire que pour n grand, $\mathbb{P}(\max_{k\in\{1,\dots,d\}}|\hat{p}_k^n-\hat{q}_k^n|\leq\epsilon_0)\leq\frac{(2\epsilon_0\sqrt{n})^{d-1}}{(2\pi(1+1/\alpha))^{(d-1)/2}\sqrt{\prod_{k=1}^d p_k}}$.
11. Lors d'un dépouillement d'une élection à 6 candidats, sont publiés les mêmes scores après avoir dépouillé un million de bulletins, puis après avoir dépouillé un million et 250000 bulletins [1]. On suppose que $10^5\pi^2\prod_{k=1}^6 p_k\geq 1$ (ce qui est le cas notamment si $p_k=1/6$). Donner un majorant de la probabilité d'un tel événement (on utilisera $\pi^{3/2}\geq 5$). Commenter.

1. Un événement de ce type s'est produit lors d'une primaire à l'élection présidentielle française de 2017, avec ces ordres de grandeur.

Corrigé

Exercice 1.

1. Soit $f : \mathbb{R} \to \mathbb{R}$ bornée mesurable. On a $\mathbb{E}[f(1/U_1)] = \int_0^1 f(1/u)du = \int_1^{+\infty} f(v)\frac{1}{v^2}dv$, et donc $X_1 = 1/U_1$ suit la loi de densité $\mathbf{1}_{v>1}\frac{1}{v^2}$ par le théorème de la fonction muette.
2. $\mathbb{E}[|X_1|^\alpha] = \int_1^{+\infty} v^{\alpha-2}dv < +\infty \iff \alpha < 1$. Comme U_1 et U_2 sont indépendantes et de même loi, $\mathbb{E}[|X_2|^\alpha] = \mathbb{E}[|X_1|^\alpha]^2 < \infty \iff \alpha < 1$.
3. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction bornée mesurable. On a par indépendance de U_1 et U_2

$$\begin{aligned}
\mathbb{E}[f(X_1, X_2)] &= \mathbb{E}[f(1/U_1, 1/(U_1U_2))] \\
&= \int_0^1 \int_0^1 f(1/u_1, 1/(u_1u_2))du_1du_2 \\
&= \int_0^1 \left(\int_0^1 f(1/u_1, 1/(u_1u_2))du_1\right) du_2 \quad \text{(Fubini)} \\
&= \int_0^1 \left(\int_1^{+\infty} f(v_1, v_1/u_2)\frac{1}{v_1^2}dv_1\right) du_2 \quad (v_1 = 1/u_1,\ u_2 \text{ fixé}) \\
&= \int_1^{+\infty} \left(\int_0^1 f(v_1, v_1/u_2)du_2\right) \frac{1}{v_1^2}dv_1 \quad \text{(Fubini)} \\
&= \int_1^{+\infty} \left(\int_1^{+\infty} \mathbf{1}_{v_1<v_2} f(v_1, v_2)\frac{v_1}{v_2^2}dv_2\right) \frac{1}{v_1^2}dv_1 \quad (v_2 = v_1/u_2,\ v_1 \text{ fixé}).
\end{aligned}$$

   Ainsi, (X_1, X_2) suit la loi de densité $\mathbf{1}_{1<v_1<v_2}\frac{1}{v_1v_2^2}$. Cette densité ne se met pas sous forme produit : X_1 et X_2 ne sont pas indépendantes.
4. On pose $\varphi :]0,1[^n \to \{x \in \mathbb{R}^n : 1 < x_1 < \cdots < x_n\}$ définie par $\varphi(u_1, \ldots, u_n) = (1/u_1, \ldots, \prod_{k=1}^k 1/u_k)$. C'est un C^1-difféomorphisme entre deux ouverts, de réciproque

$$\varphi^{-1}(x_1, \ldots, x_n) = (1/x_1, x_1/x_2, \ldots, x_{n-1}/x_n).$$

   La matrice Jacobienne s'écrit,

$$Jac(\varphi^{-1})(x) = \begin{bmatrix} -1/x_1^2 & 0 & 0 & \ldots & 0 \\ 1/x_2 & -x_1/x_2^2 & 0 & \ldots & 0 \\ 0 & 1/x_3 & -x_2/x_3^2 & \ldots & 0 \\ \vdots & \ldots & \ldots & \ldots & \vdots \\ 0 & \ldots & 0 & 1/x_n & -x_{n-1}/x_n^2 \end{bmatrix},$$

   et la valeur absolue du déterminant est donc $1/(x_1 \ldots x_{n-1}x_n^2)$. Pour $f : \mathbb{R}^n \to \mathbb{R}$ mesurable bornée, on a donc

$$\begin{aligned}
\mathbb{E}[f(X_1, \ldots, X_n)] &= \int_{]0,1[^n} f(\varphi(u_1, \ldots, u_n))du_1 \ldots du_n \\
&= \int_{\mathbb{R}^n} f(x_1, \ldots, x_n)\frac{\mathbf{1}_{\{1<x_1<\cdots<x_n\}}}{x_1 \ldots x_{n-1}x_n^2}dx_1 \ldots dx_n.
\end{aligned}$$

   Ainsi, $(X_1, \ldots, X_n)$ suit la loi de densité $\frac{\mathbf{1}_{\{1<x_1<\cdots<x_n\}}}{x_1 \ldots x_{n-1}x_n^2}$.

⋆⋆⋆

Exercice 2.

1. La loi $\beta(a,1)$ est la loi de densité $ax^{a-1}\mathbf{1}_{0<x<1}$. Sa fonction de répartition vaut donc $F(x) = x^a$ pour $x \in [0,1]$. Pour $x \in [0,1]$, $\mathbb{P}(U_1^{1/a} \leq x) = \mathbb{P}(U_1 \leq x^a) = x^a$: $U_1^{1/a}$ a la même fonction de répartition et donc la même loi que $\beta(a,1)$.
2. Soit $p(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}\mathbf{1}_{0<x<1}$ la densité de la loi $\beta(a,b)$. Pour $b \geq 1$, on a a $(1-x)^{b-1} \leq 1$ et donc $p(x) \leq kq(x)$, où $q(x)$ est la densité de la loi $\beta(a,1)$ et $k = \frac{\Gamma(a+b)}{a\Gamma(a)\Gamma(b)} = \frac{\Gamma(a+b)}{\Gamma(a+1)\Gamma(b)}$. On peut alors utiliser la méthode de rejet vue en cours : en posant $N = \inf\{i \geq 1 : kV_i q(U_i^{1/a}) \leq p(U_i^{1/a})\}$, on obtient que $U_N^{1/a}$ suit la loi $\beta(a,b)$.
3. Par la question 1, U_1^2 suit la loi de densité $\frac{1}{2}x^{-\frac{1}{2}}\mathbf{1}_{0<x<1}$. On a donc $\mathbb{E}[\varphi(U_1^2)] = \int_0^1 \varphi(x)\frac{1}{2}x^{-\frac{1}{2}}dx = I/2 < \infty$. Par la loi forte des grand nombres, on en déduit que I_n^1 converge presque sûrement vers I. De la même façon, I_n^2 converge presque sûrement vers I.
4. On a $\mathbb{E}[\varphi(U_1^2)^2] = \int_0^1 \varphi(x)^2\frac{1}{2}x^{-\frac{1}{2}}dx = \int_0^1 \frac{1+x}{2\sqrt{x}}dx < \infty$, et $\varphi(U_1^2)$ est de carré intégrable. En revanche, $\mathbb{E}[\varphi(U_1)^2/U_1] = \int_0^1 \frac{1+x}{x}dx = +\infty$, et $\varphi(U_1)/\sqrt{U_1}$ n'est pas de carré intégrable. On peut donc appliquer le théorème de la limite centrale pour I_n^1, mais pas pour I_n^2. Ainsi, on pourra construire des intervalles de confiance pour I avec une précision en $O(1/\sqrt{n})$ avec I_n^1 et on préfèrera cet estimateur.
5. On a $\mathbf{Var}(2\varphi(U_1^2)) = 4\mathbf{Var}(\varphi(U_1^2)-1) \leq 4\mathbb{E}[(\varphi(U_1^2)-1)^2] = 4\int_0^1(\varphi(x)-1)^2\frac{1}{2\sqrt{x}}dx$. Comme $\varphi'(x) = \frac{1}{2}(1+x)^{-\frac{1}{2}}$, on a $|\varphi'(x)| \leq \frac{1}{2}$ pour $x \in [0,1]$ puis $(\varphi(x)-1)^2 = (\varphi(x)-\varphi(0))^2 \leq \frac{1}{4}x^2$. Cela donne
$$\mathbf{Var}(2\varphi(U_1^2)) \leq \frac{1}{2}\int_0^1 x^{3/2}dx = \frac{1}{5}.$$
L'intervalle $[I_n^1 - \frac{1{,}96}{\sqrt{5n}}, I_n^1 + \frac{1{,}96}{\sqrt{5n}}]$ est donc un intervalle de confiance asymptotique à 95% pour I.
6. On a $\mathbf{Var}(2[\varphi(U_1^2) - U_1^2/2]) = 4\mathbf{Var}(\varphi(U_1^2) - 1 - U_1^2/2) \leq 4\mathbb{E}[(\varphi(U_1^2) - 1 - U_1^2/2)^2] = 4\int_0^1(\varphi(x) - 1 - x/2)^2\frac{1}{2\sqrt{x}}dx$. Comme $\varphi''(x) = -\frac{1}{4}(1+x)^{-\frac{3}{2}}$, on a $|\varphi''(x)| \leq 1/4$ pour $x \in [0,1]$, et donc $(\varphi(x) - 1 + x/2)^2 = (\varphi(x) - \varphi(0) - \varphi'(0)x)^2 \leq \frac{1}{64}x^4$ par l'inégalité de Taylor. On en déduit
$$\mathbf{Var}(2[\varphi(U_1^2) + U_1^2/2]) \leq \frac{1}{32}\int_0^1 x^{7/2}dx = \frac{1}{9 \times 16} = \frac{1}{12^2}.$$
7. Par la loi forte des grands nombres, J_n converge presque sûrement vers $\mathbb{E}[U_1^2] = 1/3$, et donc $I_n^1 - J_n + 1/3$ converge presque sûrement vers I. Ainsi, on obtient que $[I_n^1 - J_n + 1/3 - \frac{1{,}96}{12\sqrt{n}}, I_n^1 - J_n + 1/3 + \frac{1{,}96}{12\sqrt{n}}]$ est un intervalle de confiance asymptotique à 95% pour I. La largeur de l'intervalle de confiance a été diminuée par un facteur multiplicatif $\sqrt{5/144}$. Ainsi, pour une même précision, il faut $\frac{144}{5} = 28{,}8$ fois plus de tirages avec l'estimateur Monte-Carlo naïf qu'avec celui avec réduction de variance.

⋆⋆⋆

Problème.

1. Par la loi forte des grands nombres, $\bar{X}_n$ converge p.s. vers p et $\bar{Y}_{\lfloor \alpha n \rfloor}$ converge p.s. vers q.
2. En utilisant le théorème de la limite centrale, $\sqrt{n}(\bar{X}_n - p)$ converge en loi vers $\mathcal{N}(0, p(1-p))$ lorsque $n \to +\infty$. De même, $\sqrt{\lfloor \alpha n \rfloor}(\bar{Y}_{\lfloor \alpha n \rfloor} - q)$ converge en loi vers $\mathcal{N}(0, q(1-q))$ et donc $\sqrt{n}(\bar{Y}_{\lfloor \alpha n \rfloor} - q)$ converge en loi vers $\mathcal{N}(0, q(1-q)/\alpha)$, puisque $\sqrt{n/\lfloor \alpha n \rfloor} \to \sqrt{1/\alpha}$.
3. Soit $u \in \mathbb{R}$. On note respectivement $\Phi_n^X(u)$, $\Phi_n^Y(u)$ et $\Phi_n(u)$ les fonctions caractéristiques de $\sqrt{n}(\bar{X}_n - p)$, $\sqrt{n}(\bar{Y}_{\lfloor \alpha n \rfloor} - q)$ et $\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor} + q - p)$. D'après la question précédente, on a $\Phi_n^X(u) \underset{n\to+\infty}{\to} \exp(-p(1-p)u^2/2)$ et $\Phi_n^Y(u) \underset{n\to+\infty}{\to} \exp(-\frac{q}{\alpha}(1-q)u^2/2)$. Par indépendance de $\bar{X}_n$ et $\bar{Y}_{\lfloor \alpha n \rfloor}$, on a

$$\begin{aligned}\Phi_n(u) &= \Phi_n^X\left(\frac{u}{\sqrt{p(1-p)+\frac{q}{\alpha}(1-q)}}\right) \times \Phi_n^Y\left(\frac{-u}{\sqrt{p(1-p)+\frac{q}{\alpha}(1-q)}}\right) \\ &\underset{n\to+\infty}{\to} \exp\left(-\frac{p(1-p)u^2}{2(p(1-p)+\frac{q}{\alpha}(1-q))}\right)\exp\left(-\frac{\frac{q}{\alpha}(1-q)u^2}{2(p(1-p)+\frac{q}{\alpha}(1-q))}\right) = \exp(-u^2/2),\end{aligned}$$

ce qui prouve la convergence voulue.
4. Lorsque $p = q$, on a

$$|\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor}| \leq \epsilon_0 \iff \left|\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor} + q - p)\right| \leq \epsilon_0 \sqrt{\frac{n}{p(1-p)(1+1/\alpha)}},$$

ce qui donne le résultat en utilisant l'approximation par la loi normale,$e^{-x^2/2} \leq 1$ et $2 \leq \pi$. Pour $n = 500000$, $\alpha = 1$ et $p = 1/2$, on a $\mathbb{P}(|\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor}| \leq \epsilon_0) \leq \sqrt{2n}\epsilon_0 = 0,1$.
5. On a pour $x < -1$, $\mathbb{P}(G < x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv \leq \int_{-\infty}^x \frac{-v}{\sqrt{2\pi}} e^{-v^2/2} dv = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.
On utilise maintenant que

$$\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor} \leq \epsilon_0 \iff \sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor} + q - p) \leq -\Delta\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}}.$$

Ainsi, en utilisant l'approximation gaussienne, $\Delta\sqrt{\frac{n}{p(1-p)+\frac{q}{\alpha}(1-q)}} > 1$ et la majoration précédente, on obtient

$$\mathbb{P}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor} \leq \epsilon_0) \leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\Delta^2 n}{2[p(1-p)+\frac{q}{\alpha}(1-q)]}\right).$$

Enfin, il est facile de vérifier que le terme de droite est maximal pour $p = q = 1/2$, ce qui donne la majoration annoncée.
Pour $\Delta = 0.01$, on a alors $p = 0,50505$ et $q = 0,49495$. Pour $n = 500000$, $\alpha = 1$, on a $\mathbb{P}(|\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor}| \leq \epsilon_0) \leq \mathbb{P}(\bar{X}_n - \bar{Y}_{\lfloor \alpha n \rfloor} \leq \epsilon_0) \leq \frac{1}{\sqrt{2\pi}} e^{-50}$. Cette probabilité est extrêmement petite ! Même pour une différence de probabilité relativement faible (ici 0.0101) on voit qu'il est impossible statistiquement d'observer le même score dès que les populations ont un vote non homogène.
6. Par indépendance, la densité de G est le produit des densités : $p(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^d x_i^2\right)$.

7. La matrice C est nécessairement inversible puisque $\det(C)^2 = \det(C)\det(C^\top) = \det(\Gamma) > 0$. L'application $x \to Cx$ est un C^1-difféomorphisme de $\mathbb{R}^d$. Comme il s'agit d'une application linéaire, le Jacobien est constant et vaut $\det(C)$. On note pour $x, y \in \mathbb{R}^d$, $x \cdot y = \sum_{i=1}^d x_i y_i$ et $\|x\|_2^2 = x \cdot x$. Pour $f : \mathbb{R}^d \to \mathbb{R}$ bornée mesurable, on a donc

$$\begin{aligned}\mathbb{E}[f(CG)] &= \int_{\mathbb{R}^d} f(Cx)\frac{1}{(2\pi)^{d/2}}\exp\left(-\frac{1}{2}\|x\|_2^2\right)dx_1\dots dx_d \\ &= \int_{\mathbb{R}^d} f(y)\frac{1}{(2\pi)^{d/2}\sqrt{\det(\Gamma)}}\exp\left(-\frac{1}{2}\|C^{-1}y\|_2^2\right)dy_1\dots dy_d \\ &= \int_{\mathbb{R}^d} f(y)\frac{1}{(2\pi)^{d/2}\sqrt{\det(\Gamma)}}\exp\left(-\frac{1}{2}y\cdot(\Gamma^{-1}y)\right)dy_1\dots dy_d.\end{aligned}$$

Ainsi, la densité de $\mathcal{N}_d(0,\Gamma)$ est $\frac{1}{(2\pi)^{d/2}\sqrt{\det(\Gamma)}}\exp\left(-\frac{1}{2}y\cdot(\Gamma^{-1}y)\right)$ et elle est majorée par $\frac{1}{(2\pi)^{d/2}\sqrt{\det(\Gamma)}}$ puisque Γ^{-1} est symétrique définie positive.

8. On a $\mathbf{Var}(\mathbf{1}_{X_1=k}) = p_k(1-p_k)$ et $Cov(\mathbf{1}_{X_1=k}, \mathbf{1}_{X_1=\ell}) = -p_k p_\ell$ pour $k \neq \ell$. Ainsi, on a $\Gamma^d_{ij} = \mathbf{1}_{i=j}p_i - p_i p_j$. Cette matrice n'est pas inversible puisque $0 = \mathbf{Var}(1) = \mathbf{Var}\left(\sum_{k=1}^d \mathbf{1}_{X_1=k}\right) = 1_d \cdot \Gamma^d 1_d$, où 1_d est le vecteur dont les coordonnées sont toutes égales à 1.
Pour le calcul admis de $\det(\tilde{\Gamma}^{d-1})$, on remarque que $\tilde{\Gamma}^{d-1} = \Delta - pp^\top$ où $\Delta = diag(p_1, \dots, p_{d-1})$ est une matrice diagonale et $p \in \mathbb{R}^{d-1}$ est le vecteur des probabilités. Il est facile de voir que pour $x \in \mathbb{R}^{d-1}$, $\det(I_{d-1} - xx^\top) = 1 - \|x\|_2^2$. En effet, $I_{d-1} - xx^\top$ est une matrice symétrique qui a pour vecteurs propres x avec pour valeur propre $1 - \|x\|_2^2$, et une base orthonormée de $Vect(x)^\perp$ avec pour valeur propre 1. Ainsi, $\det(\Delta - pp^\top) = \det(\Delta)\det(I_d - \sqrt{\Delta}^{-1}p(\sqrt{\Delta}^{-1}p)^\top) = (1 - \sum_{i=1}^{d-1} p_i)\prod_{i=1}^{d-1} p_i = \prod_{i=1}^d p_i$.

9. On utilise cette fois ci le TCL multidimensionnel : $\sqrt{n}(\hat{p}^n - p)$ converge en loi vers $\mathcal{N}_d(0,\Gamma^d)$, de même que $\sqrt{\lfloor \alpha n \rfloor}(\hat{q}^{\lfloor \alpha n \rfloor} - p)$, ce qui assure que $\sqrt{n}(\hat{q}^{\lfloor \alpha n \rfloor} - p)$ converge en loi vers $\mathcal{N}_d(0, \Gamma^d/\alpha)$. Ensuite, grâce à l'indépendance entre les suites (X_i) et (Y_i), $\Phi_{\sqrt{n}(\hat{p}^n - \hat{q}^{\lfloor \alpha n \rfloor})}(u) = \Phi_{\sqrt{n}(\hat{p}^n - p)}(u)\Phi_{\sqrt{n}(\hat{q}^{\lfloor \alpha n \rfloor} - p)}(-u)$, ce qui assure la convergence en loi de $\sqrt{n}(\hat{p}^n - \hat{q}^{\lfloor \alpha n \rfloor})$ vers $\mathcal{N}_d(0, \Gamma^d(1 + 1/\alpha))$.

10. Grâce au résultat précédent, les $d-1$ premières coordonnées de $\sqrt{n}(\hat{p}^n - \hat{q}^{\lfloor \alpha n \rfloor})$ convergent en loi vers $\mathcal{N}_{d-1}(0, \tilde{\Gamma}^{d-1}(1 + 1/\alpha))$. Pour n grand, on fait l'approximation gaussienne, et on a

$$\begin{aligned}&\mathbb{P}(\max_{k\in\{1,\dots,d\}} |\hat{p}^n_k - \hat{q}^n_k| \le \epsilon_0) \\ &\le \mathbb{P}(\max_{k\in\{1,\dots,d-1\}} |\hat{p}^n_k - \hat{q}^n_k| \le \epsilon_0) = \mathbb{P}(\max_{k\in\{1,\dots,d-1\}} \sqrt{n}|\hat{p}^n_k - \hat{q}^n_k| \le \epsilon_0\sqrt{n}) \\ &\approx \int_{[-\epsilon_0\sqrt{n},\epsilon_0\sqrt{n}]^{d-1}} \frac{1}{(2\pi)^{(d-1)/2}\sqrt{\det((1+1/\alpha)\tilde{\Gamma}^{d-1})}}\exp\left(-\frac{1}{2}y\cdot(((1+1/\alpha)\tilde{\Gamma}^{d-1})^{-1}y)\right)dy_1\dots dy_{d-1} \\ &\le \frac{(2\epsilon_0\sqrt{n})^{d-1}}{(2\pi(1+1/\alpha))^{(d-1)/2}\sqrt{\prod_{k=1}^d p_k}}.\end{aligned}$$

11. On a $d = 6$, $n = 1000000$, $\alpha = 1/4$. La probabilité d'avoir le même score est donc majorée par $\frac{(2/10)^5}{(10\pi)^{5/2}\sqrt{\prod_{k=1}^{d} p_k}} \leq \frac{(2/10)^5}{(\pi)^{3/2}} \leq 6,4 \times 10^{-5}$. C'est un événement très peu probable ! D'autant plus que nous sommes ici dans l'hypothèse d'un vote homogène.