Examen du cours de Probabilités

Mardi 19 Janvier 2016 (8h30-11h30)

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

Exercice 1. Soient $a, b, c, \theta > 0$. On considère trois variables aléatoires réelles $X \sim \Gamma(a,\theta)$, $Y \sim \Gamma(b,\theta)$ et $Z \sim \Gamma(c,\theta)$ indépendantes. On pose

$$S = X + Y + Z,\ T = \frac{X}{X+Y+Z},\ U = \frac{Y}{X+Y+Z}.$$

1. Donner sans calcul les lois de S, T et U. Calculer $\mathbb{E}[T]$ et $\mathbf{Var}(T)$.
2. Calculer la loi de (S,T,U) (on donnera la densité de ce vecteur aléatoire). Indiquer, de la façon la plus précise possible, quelles variables aléatoires sont indépendantes.

⋆⋆⋆

Exercice 2. On appelle loi de Pareto de paramètre $\alpha > 0$ la loi de densité $\frac{\alpha}{x^{\alpha+1}}\mathbf{1}_{x>1}$ sur $\mathbb{R}$. On suppose que X suit la loi de Pareto de paramètre α et que $U \sim \mathcal{U}([0,1])$.

1. Pour quelles valeurs de α la variable X est-elle intégrable ? Calculer dans ce cas $\mathbb{E}[X]$. De même, préciser sous quelle condition X est de carré intégrable et calculer $\mathbf{Var}(X)$.
2. Calculer pour $x \geq 1$, $\mathbb{P}(X \geq x)$. En déduire que $X \overset{loi}{=} U^{-1/\alpha}$.

On considère une suite $(X_k)_{k\in\mathbb{N}^*}$ i.i.d de variables aléatoires qui suivent la loi de Pareto de paramètre $\boxed{\text{égal à 1}}$. On pose $Z_n = \min_{1\leq k\leq n} X_k$.

3. Calculer pour $x \geq 1$, $\mathbb{P}(Z_n \geq x)$, puis donner la loi de Z_n.
4. Montrer que Z_n converge en probabilité vers 1 lorsque $n \to +\infty$.
5. En utilisant la monotonie de la suite $(Z_n, n \geq 1)$, montrer que Z_n converge presque sûrement vers 1 lorsque $n \to +\infty$.
6. Calculer la densité $q_n(y)$ de la variable aléatoire $Y_n = n(Z_n - 1)$. Calculer la limite de $q_n(y)$ lorsque $n \to +\infty$. En déduire que $n(Z_n - 1)$ converge en loi vers une loi usuelle que l'on précisera. Indication : pour la domination, on pourra utiliser la formule du binôme pour montrer que $q_n(y) \leq \frac{1}{1+y+y^2/2}$.
7. Montrer que pour $f : \mathbb{R} \to \mathbb{R}$ continue bornée, $\mathbb{E}[f(Y_n)] = \mathbb{E}[f(n(U^{-1/n} - 1))]$. Retrouver la convergence en loi obtenue à la question précédente.

⋆⋆⋆

Exercice 3. On considère (X_1, X_2) un vecteur gaussien centré de matrice de covariance $\Gamma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$, avec $\sigma_1, \sigma_2 > 0$ et $\rho \in [-1,1]$. On considère $\varepsilon \sim \mathcal{B}(1/2)$ une variable aléatoire de Bernoulli indépendante de (X_1, X_2) et on définit

$$(Y_1, Y_2) = \varepsilon(X_1, X_2) + (1-\varepsilon)(X_2, X_1).$$

1. Calculer la fonction caractéristique du vecteur (Y_1, Y_2).
2. Calculer l'espérance et la matrice de covariance du vecteur (Y_1, Y_2).
3. Donner une condition nécessaire et suffisante pour que le vecteur (Y_1, Y_2) soit gaussien.

⋆⋆⋆

Exercice 4. Estimation des gains/pertes de la Française des jeux à un tirage du LOTO©.

On considère une suite i.i.d $(X_k)_{k\in\mathbb{N}^*}$ de variables aléatoires de Bernoulli de paramètre $p \in]0,1[$. On note $S_n = \sum_{k=1}^n X_k$. On suppose que l'on connaît p et S_n, et on cherche à construire un intervalle de confiance asymptotique à 95% pour n.

1. Donner le comportement asymptotique de S_n/n et de $\sqrt{n}(S_n/n - p)$ lorsque $n \to +\infty$.
2. En déduire que $\sqrt{S_n}\left(\frac{n}{S_n} - \frac{1}{p}\right) \xrightarrow{loi} \mathcal{N}\left(0, \frac{1-p}{p^2}\right)$
3. Construire un intervalle de confiance asymptotique à 95% pour n en fonction de S_n et p.
4. Montrer que $[\frac{S_n - 2\sqrt{(1-p)S_n}}{p}, +\infty[$ est un intervalle de confiance asymptotique à (au moins) 97,5%, c'est à dire que pour n suffisamment grand, $\mathbb{P}(n \in [\frac{S_n - 2\sqrt{(1-p)S_n}}{p}, +\infty[) \geq 0,975$.

Après chaque tirage, la Française des jeux publie sur son site internet la répartition des gains. Voici les gains au tirage du 9 décembre 2015. Ainsi, la Française des jeux a reversé 3 926 642 €. On rappelle brièvement les règles du jeu : chaque joueur mise 2 € et choisit sur une grille 5 numéros parmi $\{1, \dots, 49\}$ et un numéro chance dans $\{1, \dots, 10\}$. On note $E = \{A \subset \{1, \dots, 49\}, \mathrm{Card}(A) = 5\}$. On suppose que les n joueurs jouent de façon indépendantes et uniforme, et que le tirage des numéros est également indépendant et suit la loi uniforme. Formellement, cela signifie que le tirage (G, N) et les grilles des joueurs $(G_k, N_k)_{1 \leq k \leq n}$ sont indépendants et suivent la même loi

	Nombre de grilles gagnantes	Gains par grille gagnante
5 bons numéros + le N°Chance	0	
5 bons numéros	0	
4 bons numéros	559	1348,30 €
3 bons numéros	26267	8,40 €
2 bons numéros	351347	4,50 €
N° Chance gagnant	685619	mise de 2 € remboursée

$$\forall A \in E, i \in \{1, \dots, 10\},\ \mathbb{P}(G = A, N = i) = \frac{1}{10 \times \binom{49}{5}}.$$

Ici G et N représentent la grille et le numéro chance tirés au sort alors que G_k et N_k représentent la grille et le numéro chance choisis par le k-ème joueur.

5. Montrer que les variables aléatoires $(\mathbf{1}_{N_k = N})_{1 \leq k \leq n}$ sont i.i.d. Indication : on pourra calculer $\mathbb{P}(N_k = N)$ puis $\mathbb{P}(N_{i_1} = N, \dots, N_{i_k} = N)$ pour $1 \leq k \leq n$ et $1 \leq i_1 < \dots < i_k \leq n$.
6. On note $S_n^1 = \sum_{k=1}^n \mathbf{1}_{N_k = N}$. A l'aide des approximations $685619 \approx 685620$, $\sqrt{0,9 \times 685619} \leq 810$ et $3926642 \approx 3930000$, donner un intervalle de confiance à (au moins) 97,5% des bénéfices effectués lors de ce tirage.

On s'interroge désormais sur la validité des hypothèses du modèle. On pose $p_1 = 1/10$, $p_2 = \mathbb{P}(\mathrm{Card}(G_1 \cap G) = 2)$ et on note $S_n^2 = \sum_{k=1}^n \mathbf{1}_{\mathrm{Card}(G_k \cap G) = 2}$.

7. Calculer p_2 et donner une approximation numérique en utilisant que $\frac{42 \times 43 \times 44}{45 \times \cdots \times 49} \approx 3,5 \times 10^{-4}$.
8. Donner le comportement asymptotique du vecteur $\sqrt{n}\left(\frac{1}{n}\begin{pmatrix} S_n^1 \\ S_n^2 \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}\right)$, puis en déduire le comportement asymptotique de $\sqrt{p_1 S_n^1}\left(\frac{S_n^2}{S_n^1} - \frac{p_2}{p_1}\right)$. Indication : écrire
$$\sqrt{p_1 S_n^1}\left(\frac{S_n^2}{S_n^1} - \frac{p_2}{p_1}\right) = \frac{\sqrt{p_1 S_n^1}}{\sqrt{n}}\left[\sqrt{n}\left(\frac{S_n^2/n}{p_1} - \frac{p_2}{p_1}\right) + \frac{S_n^2/n}{p_1 S_n^1/n}\sqrt{n}(p_1 - S_n^1/n)\right].$$
Donner un intervalle de confiance asymptotique à 99% pour $\frac{p_2}{p_1}$ (cet intervalle pourra dépendre de p_1 et p_2).
9. On donne $\frac{2,58}{\sqrt{68561,9}} \approx 10^{-2}$, $\sqrt{\frac{p_2^2}{p_1}(1-p_1) + p_2(1-p_2)} \approx 0,33$ et $\frac{351347}{685619} \approx 0,5125$. Calculer numériquement l'intervalle de confiance à 99% pour $\frac{p_2}{p_1}$, et comparer à la vraie valeur de $\frac{p_2}{p_1}$. Commenter la validité du modèle. Quelle hypothèse faite paraît discutable ?

Corrigé

EXERCICE 1.

1. On applique la Proposition 3.4.4. du polycopié. En remarquant que $X \sim \Gamma(\theta, a)$ et $Y+Z \sim \Gamma(\theta, b+c)$ sont indépendantes (de même pour $Y \sim \Gamma(\theta, b)$ et $X+Z \sim \Gamma(\theta, a+c)$, On obtient que $S \sim \Gamma(a+b+c, \theta)$, $T \sim \beta(a, b+c)$ et $U \sim \beta(b, a+c)$. La densité de T est donc $\frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b+c)} t^{a-1}(1-t)^{b-1}\mathbf{1}_{0<t<1}$. En utilisant que $\Gamma(x+1) = x\Gamma(x)$, on a

$$\mathbb{E}[T] = \int_0^1 \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b+c)} t^a(1-t)^{b-1} dt = \frac{a}{a+b+c} \int_0^1 \frac{\Gamma(a+1+b+c)}{\Gamma(a+1)\Gamma(b+c)} t^a(1-t)^{b-1} dt = \frac{a}{a+b+c}.$$

De même, on obtient $\mathbb{E}[T^2] = \frac{a(a+1)}{(a+b+c)(a+b+c+1)}$ puis

$$\mathbf{Var}(T) = \frac{a(b+c)}{(a+b+c)^2(a+b+c+1)}.$$

2. L'application $\varphi : \begin{array}{l} (\mathbb{R}_+^*)^3 \to \mathbb{R}_+^* \times \{((t,u) \in]0,1[, t+u<1\} \\ (x,y,z) \mapsto (x+y+z, \frac{x}{x+y+z}, \frac{y}{x+y+z}) =: (s,t,u) \end{array}$ est un C^1-diffeomorphisme d'inverse $\varphi^{-1}(s,t,u) = (st, su, s(1-(t+u)))$. On a $|\mathrm{Jac}\varphi^{-1}(s,t,u)| = \left| \det \begin{pmatrix} t & s & 0 \\ u & 0 & s \\ 1-(t+u) & -s & -s \end{pmatrix} \right| = s^2$. Ainsi, pour $f : \mathbb{R}^3 \to \mathbb{R}$ mesurable bornée, on a

$$\mathbb{E}[f(S,T,U)] = \int_{\mathbb{R}_+^*}\int_{\mathbb{R}_+^*}\int_{\mathbb{R}_+^*} f(x+y+z, \frac{x}{x+y+z}, \frac{y}{x+y+z})\theta^{a+b+c}\frac{x^{a-1}y^{b-1}z^{c-1}}{\Gamma(a)\Gamma(b)\Gamma(c)} e^{-\theta(x+y+z)} dxdydz$$
$$= \int_{\mathbb{R}_+^*}\int_{\mathbb{R}_+^*}\int_{\mathbb{R}_+^*} f(s,t,u)\mathbf{1}_{t+u<1}\frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} t^{a-1}u^{b-1}(1-(t+u))^{c-1} \times \frac{\theta^{a+b+c}s^{a+b+c-1}e^{-\theta s}}{\Gamma(a+b+c)} dsdtdu$$

Donc (S,T,U) suit une loi de densité :

$$\mathbf{1}_{t+u<1, t>0, u>0}\frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} t^{a-1}u^{b-1}(1-(t+u))^{c-1} \times \mathbf{1}_{s>0}\frac{\theta^{a+b+c}s^{a+b+c-1}e^{-\theta s}}{\Gamma(a+b+c)}.$$

La variable aléatoire S est indépendante du couple (T,U). En revanche T et U ne sont pas indépendantes.

⋆⋆⋆

EXERCICE 2.

1. La variable X est intégrable si, et seulement si, $\alpha > 1$, et on a alors $\mathbb{E}[X] = \int_1^\infty \frac{\alpha}{x^\alpha} dx = \frac{\alpha}{\alpha-1}$. De même, préciser sous quelle condition X est de carré intégrable si et seulement si $\alpha > 2$. Dans ce cas, $\mathbb{E}[X^2] = \frac{\alpha}{\alpha-2}$ et $\mathbf{Var}(X) = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}$.
2. Pour $x \geq 1$, on a $\mathbb{P}(X \geq x) = \int_x^\infty \frac{\alpha}{z^{\alpha+1}} dz = x^{-\alpha}$. Donc la fonction de répartition de X est $F(x) = 1 - x^{-\alpha}$. On en déduit $F^{-1}(u) = (1-u)^{-1/\alpha}$, et par la méthode d'inversion de la fonction de répartition, $X \overset{loi}{=} F^{-1}(U)$ où U suit une loi uniforme sur $[0,1]$. Comme U a même loi que $1-U$, $X \overset{loi}{=} U^{-1/\alpha}$.
3. Pour $x \geq 1$, $\mathbb{P}(Z_n \geq x) = \mathbb{P}(X_1 \geq x, \ldots, X_n \geq x) = \mathbb{P}(X_1 \geq x)^n = x^{-n}$ par indépendance. On reconnaît la fonction de survie (definie comme un moins la fonction de répartition) calculée à la question 2. Comme la fonction de répartition caractérise la loi, Z_n suit une loi de Pareto de paramètre n.
4. Pour $x > 1$, $\mathbb{P}(Z_n \geq x) = x^{-n} \to 0$ lorsque $n \to +\infty$. Donc Z_n converge en probabilité vers 1.

5. $(Z_n, n \geq 1)$ est une suite décroissante de réels et est minorée par 1, elle converge donc pour tout ω et donc presque sûrement. Comme la convergence presque sûre implique la convergence en probabilité, Z_n converge presque sûrement vers 1.
6. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction bornée mesurable. On a

$$\begin{aligned}\mathbb{E}[f(Y_n)] &= \int_1^\infty f(n(z-1))\frac{n}{z^{n+1}}dz \\ &= \int_0^\infty f(y)\frac{1}{(1+y/n)^{n+1}}dy.\end{aligned}$$

   On a $(1+y/n)^{n+1} \to e^y$, et ainsi $q_n(y) = \frac{1}{(1+y/n)^{n+1}}\mathbf{1}_{y>0}$ converge vers $e^{-y}\mathbf{1}_{y>0}$ lorsque $n \to +\infty$. En utilisant la formule du binôme et $y > 0$, on a $(1+y/n)^{n+1} \geq 1 + \frac{n+1}{n}y + \frac{n(n+1)}{2n^2}y^2 \geq 1 + y + y^2/2$ et donc $q_n(y) \leq \frac{1}{1+y+y^2/2}$ qui est une fonction intégrable. Par le théorème de convergence dominée, $\mathbb{E}[f(Y_n)] \to \int_0^\infty f(y)e^{-y}dy$. ce qui prouve la convergence en loi de Y_n vers $\mathcal{E}(1)$.
7. Soit f une fonction bornée continue. On a $Y_n \overset{loi}{=} n(U^{-1/n}-1)$ et donc $\mathbb{E}[f(Y_n)] = \mathbb{E}[f(n(U^{-1/n}-1))]$. Or $n(U^{-1/n}-1) \to -\log(U)$ p.s et par continuité de f, $f(n(U^{-1/n}-1)) \to f(-\log(U))$ p.s. Comme f est bornée, on en déduit par convergence dominée que $\mathbb{E}[f(n(U^{-1/n}-1))] \to \mathbb{E}[f(-\log(U))]$. Comme $-\log(U) \sim \mathcal{E}(1)$, cela donne la convergence en loi de Y_n vers $\mathcal{E}(1)$.

⋆⋆⋆

Exercice 3.

1. Grâce à l'indépendance entre ε et X, on a pour $u_1, u_2 \in \mathbb{R}$,

$$\begin{aligned}\Phi_Y(u_1,u_2) &= \mathbb{E}[e^{i(u_1Y_1+u_2Y_2)}] = \mathbb{E}[\mathbf{1}_{\varepsilon=1}e^{i(u_1X_1+u_2X_2)}] + \mathbb{E}[\mathbf{1}_{\varepsilon=0}e^{i(u_2X_1+u_1X_2)}] \\ &= \frac{1}{2}\mathbb{E}[e^{i(u_1X_1+u_2X_2)}] + \frac{1}{2}\mathbb{E}[e^{i(u_2X_1+u_1X_2)}] \\ &= \frac{1}{2}e^{-\frac{1}{2}(\sigma_1^2u_1^2+2\rho\sigma_1\sigma_2u_1u_2+\sigma_2^2u_2^2)} + \frac{1}{2}e^{-\frac{1}{2}(\sigma_1^2u_2^2+2\rho\sigma_1\sigma_2u_1u_2+\sigma_2^2u_1^2)}.\end{aligned}$$

2. On a par le même argument en décomposant suivant les valeurs de ε et en utilisant l'indépendance, $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = 0$, $\mathbf{Var}(Y_1) = \mathbb{E}[Y_1^2] = \frac{1}{2}(\sigma_1^2+\sigma_2^2)$, $\mathbf{Var}(Y_2) = \mathbb{E}[Y_2^2] = \frac{1}{2}(\sigma_1^2+\sigma_2^2)$, $\mathbf{Cov}(Y_1,Y_2) = \mathbb{E}[Y_1Y_2] = \rho\sigma_1\sigma_2$.
3. On note $\hat{\Gamma} = \begin{bmatrix} \sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$ et $\bar{\Gamma} = \frac{1}{2}(\Gamma + \hat{\Gamma})$. La première question donne $\Phi_Y(u_1,u_2) = \frac{1}{2}(e^{-\frac{1}{2}u.\Gamma u} + e^{-\frac{1}{2}u.\hat{\Gamma}u})$. D'après la deuxième question, Y est un vecteur gaussien si, et seulement si $\Phi_Y(u_1,u_2) = e^{-\frac{1}{2}u.\bar{\Gamma}u}$. L'inégalité de convexité satisfaite par la fonction exponentielle assure que ceci est vrai si et seulement si $u.\Gamma u = u.\hat{\Gamma}u$ pour tout $u \in \mathbb{R}^2$. Ainsi, Y est Gaussien si, et seulement si $\sigma_1^2 = \sigma_2^2$.

⋆⋆⋆

Exercice 4.

1. La loi forte des grands nombres assure que S_n/n converge presque sûrement vers p. Le théorème de la limite centrale donne la convergence en loi de $\sqrt{n}(S_n/n - p)$ vers une gaussienne centrée de variance $p(1-p)$.
2. On écrit $\sqrt{S_n}\left(\frac{n}{S_n} - \frac{1}{p}\right) = \sqrt{S_n}\frac{n}{pS_n}(p - S_n/n) = \frac{1}{p}\sqrt{\frac{n}{S_n}}\sqrt{n}(p - S_n/n)$. Comme $\frac{1}{p}\sqrt{\frac{n}{S_n}}$ converge p.s. vers $p^{-3/2}$ et $\sqrt{n}(p-S_n/n)$ converge en loi vers $\mathcal{N}(0,p(1-p))$, le théorème de Slutsky assure la convergence en loi de $(\frac{1}{p}\sqrt{\frac{n}{S_n}}, \sqrt{n}(p-S_n/n))$ vers $(p^{-3/2}, \mathcal{N}(0,p(1-p)))$. Le produit étant continu, il en découle que $\sqrt{S_n}\left(\frac{n}{S_n} - \frac{1}{p}\right)$ converge en loi vers $\mathcal{N}(0,(1-p)/p^2)$.

3. Construire un intervalle de confiance asymptotique à 95% pour n en fonction de S_n et p. Lorsque n est grand, $\sqrt{S_n}\left(\frac{n}{S_n}-\frac{1}{p}\right) \overset{loi}{\approx} \frac{\sqrt{1-p}}{p}G$, où $G \sim \mathcal{N}(0,1)$. On obtient donc que

$$\left[\frac{S_n}{p} - 1,96\frac{\sqrt{(1-p)S_n}}{p}, \frac{S_n}{p} + 1,96\frac{\sqrt{(1-p)S_n}}{p}\right]$$

est un intervalle de confiance asymptotique à 95% pour n.

4. On a $\mathbb{P}(G \in [-1,96,+\infty[) \approx 0,975$ donc $\left[\frac{S_n}{p} - 1,96\frac{\sqrt{(1-p)S_n}}{p}, +\infty\right[$ est un intervalle de confiance asymptotique à 97,5% pour n. Comme $1,96 \leq 2$, on en déduit que $[\frac{S_n-2\sqrt{(1-p)S_n}}{p}, +\infty[$ est également un intervalle de confiance asymptotique à 97,5% pour n.

5. Les v.a. $\mathbf{1}_{N_k=N}$ sont des variable de Bernoulli de paramètre $\mathbb{P}(N_k = N) = \frac{1}{100}\sum_{i,j=1}^{10}\mathbf{1}_{\{i=j\}} = \frac{1}{10}$. On a $\mathbb{P}(N_{i_1} = N, \ldots, N_{i_k} = N) = \frac{1}{10^{k+1}}\sum_{j,j_1,\ldots,j_k=1}^{10}\mathbf{1}_{j_1=j,\ldots,j_k=j} = \frac{1}{10^{k+1}}\sum_{i=1}^{10} 1 = \frac{1}{10^k} = \mathbb{P}(N_{i_1} = N) \times \cdots \times \mathbb{P}(N_{i_k} = N)$, ce qui prouve l'indépendance des événements $\{N_k = N\}$, $1 \leq k \leq n$ et donc des variables aléatoires.

6. Grâce à la question 4 et en utilisant les approximations, $[6840000, +\infty[$ est un intervalle de confiance à 97,5% pour n. Ainsi, le bénéfice réalisé est supérieur à $2 \times 6840000 - 3930000 = 9750000$ euros, avec probabilité 97,5%.

7. On a $p_2 = \mathbb{P}(\mathrm{Card}(G_k \cap G) = 2) = \frac{1}{\binom{49}{5}}\binom{5}{2} \times \binom{44}{3} = \frac{120}{45\times\cdots\times 49} \times 10 \times \frac{42\times 43\times 44}{6} \approx 7 \times 10^{-2}$.

8. On note que $\mathbf{1}_{N_k=N}$ et $\mathbf{1}_{\mathrm{Card}(G_k\cap G)=2}$ sont indépendantes. Le théorème de la limite centrale assure que $\sqrt{n}\left(\frac{1}{n}\begin{pmatrix} S_n^1 \\ S_n^2\end{pmatrix} - \begin{pmatrix} p_1 \\ p_2\end{pmatrix}\right)$ converge en loi vers $\mathcal{N}_2\left(0, \begin{pmatrix} p_1(1-p_1) & 0 \\ 0 & p_2(1-p_2)\end{pmatrix}\right)$. On utilise alors l'écriture

$$\sqrt{p_1 S_n^1}\left(\frac{S_n^2}{S_n^1} - \frac{p_2}{p_1}\right) = \frac{\sqrt{p_1 S_n^1}}{\sqrt{n}}\left[\sqrt{n}\left(\frac{S_n^2/n}{p_1} - \frac{p_2}{p_1}\right) + \frac{S_n^2/n}{p_1 S_n^1/n}\sqrt{n}(p_1 - S_n^1/n)\right].$$

La loi forte des grands nombres assure que $\frac{\sqrt{p_1 S_n^1}}{\sqrt{n}} \to p_1$ et $\frac{S_n^2/n}{p_1 S_n^1/n} \to \frac{p_2}{p_1^2}$ presque sûrement. Le théorème de Slutsky et la continuité des opérations assure que $\sqrt{p_1 S_n^1}\left(\frac{S_n^2}{S_n^1} - \frac{p_2}{p_1}\right)$ converge en loi vers $\mathcal{N}\left(0, p_2(1-p_2) + p_2^2\frac{1-p_1}{p_1}\right)$. Donc l'intervalle

$$\left[\frac{S_n^2}{S_n^1} - 2,58\frac{\sqrt{p_2(1-p_2) + p_2^2\frac{1-p_1}{p_1}}}{\sqrt{p_1 S_n^1}}, \frac{S_n^2}{S_n^1} + 2,58\frac{\sqrt{p_2(1-p_2) + p_2^2\frac{1-p_1}{p_1}}}{\sqrt{p_1 S_n^1}}\right]$$

est un intervalle de confiance asymptotique à 99% pour $\frac{p_2}{p_1}$.

9. L'intervalle de confiance à 99% pour $\frac{p_2}{p_1}$ est $[0,5092; 0,5158]$, et la vraie valeur $\frac{p_1}{p_2} \approx 0,7$ est très nettement en dehors de l'intervalle de confiance. Cela tend à indiquer que le modèle utilisé n'est pas tout à fait réaliste. En particulier, l'hypothèse faite sur le caractère uniforme des grilles choisies par les joueurs est très discutable et n'est pas vérifiée en pratique.