Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées

IS Math314 **Année 2004-2005**

Corrigé du partiel du 1^er avril 2005.

Ex 1. *Pétition pour internautes*

1) La v.a. X ayant pour densité f, X^n est intégrable si $\mathbf{E}|X^n| = \int_{-\infty}^{+\infty} |t|^n f(t)\,\mathrm{d}t$ est finie. Or ici

$$\mathbf{E}|X^n| = \int_{-\infty}^{+\infty} |t|^n a\mathrm{e}^{-at}\mathbf{1}_{[0,+\infty[}(t) = \int_0^{+\infty} at^n\mathrm{e}^{-at}\,\mathrm{d}t$$

et cette intégrale généralisée converge clairement dans $\mathbb{R}_+$ pour tout $a > 0$ et tout $n \in \mathbb{N}$, parce que la fonction exponentielle tend plus vite vers 0 en $+\infty$ que t^{-n-2}. Ceci légitime l'existence de $\mathbf{E}X^n$ comme élément de $\mathbb{R}$. On calcule ce moment en écrivant que

$$\mathbf{E}X^n = \int_{-\infty}^{+\infty} t^n a\mathrm{e}^{-at}\mathbf{1}_{[0,+\infty[}(t) = \int_0^{+\infty} at^n\mathrm{e}^{-at}\,\mathrm{d}t.$$

En effectuant dans cette intégrale le changement de variable $u = at$ et en se rappelant la définition de la fonction Γ, il vient

$$\mathbf{E}X^n = a\int_0^{+\infty} \frac{u^n}{a^n}\mathrm{e}^{-u}\frac{\mathrm{d}u}{a} = \frac{1}{a^n}\int_0^{+\infty} u^n\mathrm{e}^{-u}\,\mathrm{d}u = \frac{\Gamma(n+1)}{a^n} = \frac{n!}{a^n}.$$

En appliquant cette formule avec $n = 1$ et $n = 2$, on obtient

$$\mathbf{E}X = \frac{1}{a}$$

et

$$\operatorname{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \frac{2}{a^2} - \frac{1}{a^2} = \frac{1}{a^2}.$$

2) L'instant d'enregistrement de la n^{e} signature est

$$T_n := \sum_{i=1}^n X_i,$$

où les X_i sont indépendantes et de même loi exponentielle de paramètre $a = 1\ \mathrm{mn}^{-1}$. On a alors

$$\mathbf{E}X_i = 1 \text{ mn} \quad \text{d'où} \quad \mathbf{E}T_n = n \ \text{ mn},$$

par linéarité de l'espérance. De plus

$$\operatorname{Var} X_i = 1 \text{ mn}^2 \quad \text{d'où} \quad \operatorname{Var} T_n = \sum_{i=1}^{n} \operatorname{Var} X_i = n \text{ mn}^2,$$

parce que les X_i sont de carré intégrable, indépendantes et de même loi. Ces mêmes propriétés des X_i sont les hypothèses du théorème limite central. La somme centrée réduite T_n^* s'écrit ici

$$T_n^* = \frac{T_n - \mathbf{E}T_n}{\sqrt{\operatorname{Var} T_n}} = \frac{T_n - n}{\sqrt{n}}.$$

Le théorème limite central nous dit que T_n^* converge en loi quand n tend vers $+\infty$ vers Z de loi gaussienne standard $\mathfrak{N}(0,1)$. Nous pourrons donc considérer que la loi de $T_{10\,000}^*$ est approximativement $\mathfrak{N}(0,1)$. Dans une semaine il y a $7 \times 24 \times 60 = 10\,080$ minutes. La probabilité que les initiateurs de la pétition gagnent leur pari est donc

$$\begin{aligned} P(T_{10\,000} \leq 10\,080) &= P\Big(T_{10\,000}^* \leq \frac{10\,080 - 10\,000}{\sqrt{10\,000}}\Big) \\ &= P(T_{10\,000}^* \leq 0{,}8) \\ &\simeq \Phi(0{,}8) \simeq 0{,}788. \end{aligned}$$

Ex 2. *Estimation du rapport signal sur bruit*

1) Voici une première rédaction de cette question supposant déjà une certaine familiarité avec les propriétés des la convergence presque sûre. On la fait suivre de justifications complémentaires à l'usage des débutants. Puisque les X_i sont i.i.d. et de carré intégrable, elles sont aussi intégrables. La suite $(X_i)_{i\geq 1}$ vérifie donc la loi forte des grands nombres. De plus par hérédité de l'indépendance, les X_i^2 sont indépendantes et comme images des variables de même loi X_i par la même fonction borélienne $x \mapsto x^2$, les X_i^2 ont même loi. Comme $\mathbf{E}|X_1^2| = \mathbf{E}X_1^2$ est fini, la suite $(X_i^2)_{i\geq 1}$ vérifie elle aussi la loi forte des grands nombres. On a donc les convergences presque sûres suivantes.

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}X_1. \tag{1}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}(X_1^2). \tag{2}$$

Par continuité sur $\mathbb{R}$ de l'application $x \mapsto x^2$, on déduit de (1) la convergence

$$\Big(\frac{S_n}{n}\Big)^2 \xrightarrow[n\to+\infty]{\text{p.s.}} (\mathbf{E}X_1)^2. \tag{3}$$

En combinant (2) et (3), on obtient

$$\frac{V_n}{n} - \frac{S_n^2}{n^2} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}(X_1^2) - \mathbf{E}(X_1^2) = \operatorname{Var} X_1 = \sigma^2. \tag{4}$$

Comme σ^2 est strictement positif, on déduit de (4) que les variables aléatoires $n^{-1}V_n - n^{-2}S_n^2$ sont presque sûrement strictement positives à partir d'un certain rang (aléatoire). Il en va donc de même pour le dénominateur D_n de R_n puisque

$$D_n := nV_n - S_n^2 = n^2\Big(\frac{V_n}{n} - \frac{S_n^2}{n^2}\Big).$$

Justifications complémentaires. Définissons les évènements

$$\Omega_1 := \{\omega \in \Omega;\ n^{-1}S_n(\omega) \xrightarrow[n\to+\infty]{} \mathbf{E}X_1\}, \quad \Omega_2 := \{\omega \in \Omega;\ n^{-1}V_n(\omega) \xrightarrow[n\to+\infty]{} \mathbf{E}(X_1^2)\}.$$

Par (1), on a $P(\Omega_1) = 1$ et par (2), $P(\Omega_2) = 1$. Par continuité de l'application $x \mapsto x^2$, pour tout $\omega \in \Omega_1$, la suite de nombres réels $\big(n^{-2}S_n^2(\omega)\big)_{n\geq 1}$ converge vers $\mathbf{E}X_1^2$. Comme $P(\Omega_1) = 1$, ceci justifie (3).

Notons $\Omega_3 := \Omega_1 \cap \Omega_2$. Pour tout $\omega \in \Omega_3$, la suite de réels $\big(n^{-1}V_n((\omega) - n^{-2}S_n^2(\omega)\big)_{n\geq 1}$ converge vers $\mathbf{E}(X_1^2) - (\mathbf{E}X_1)^2$ comme différence de deux suites convergentes. L'intersection finie (ou dénombrable) d'évènements de probabilité 1 est encore un évènement de probabilité 1, donc $P(\Omega_3) = 1$, ce qui justifie (4).

En prenant ω quelconque dans Ω_3, et en choisissant $\varepsilon = \sigma^2/2 > 0$, il existe un rang $N(\omega)$ tel que

$$\forall n \geq N(\omega), \quad \frac{V_n(\omega)}{n} - \frac{S_n(\omega)^2}{n^2} > \sigma^2 - \varepsilon = \frac{\sigma^2}{2} > 0.$$

Comme $P(\Omega_3) = 1$, il est donc légitime de dire que $n^{-1}V_n - n^{-2}S_n^2$ est presque-sûrement strictement positive à partir d'un certain rang.

2) Il résulte de la question précédente, que le dénominateur $D_n = nV_n - S_n^2$ est presque-sûrement non nul à partir d'un certain rang. Par conséquent pour étudier la convergence presque-sûre de R_n, il suffit de considérer l'expression

$$R_n = \frac{S_n^2}{nV_n - S_n^2}$$

en se plaçant sur l'évènement Ω_3. Sur cet évènement, on peut écrire

$$R_n = \frac{n^{-2}S_n^2}{n^{-2}(nV_n - S_n^2)} = \frac{\big(\frac{S_n}{n}\big)^2}{\frac{V_n}{n} - \frac{S_n^2}{n^2}}$$

Quand on se restreint à Ω_3, les convergences (3) et (4) deviennent des convergences « pour tout ω » et $R_n(\omega)$ converge donc comme quotient de deux suites convergentes, la deuxième ayant une limite non nulle. La limite est alors le quotient des limites. Comme $P(\Omega_3) = 1$, nous venons d'établir que

$$R_n \xrightarrow[n\to+\infty]{\text{p.s.}} \frac{(\mathbf{E}X_1)^2}{\sigma^2}. \tag{5}$$

3) Un signal est émis sous forme d'impulsions d'intensité inconnue $a > 0$. Pour la i^{e} impulsion, le récepteur capte une valeur $X_i = a + Z_i$, où Z_i représente le « bruit ».

Les Z_i sont i.i.d. centrées, de carré intégrable et d'écart-type inconnu $\sigma > 0$. On cherche à estimer le rapport a/σ, « signal sur bruit », à partir de la seule observation des X_i.

On remarque que les X_i sont indépendantes et de même loi comme images $h(Z_i)$ des Z_i par la même application borélienne $h : x \mapsto a + x$. De plus les X_i sont de carré intégrable car

$$X_i^2 = (a + Z_i)^2 \leq 2a^2 + 2X_i^2, \quad \text{d'où} \quad \mathbf{E}|X_i^2| \leq 2a^2 + 2\mathbf{E}X_i^2 < +\infty.$$

Ceci implique en particulier qu'elles sont intégrables et on a $\mathbf{E}X_i = a + \mathbf{E}Z_i = a$. De plus, comme a est constante,

$$\operatorname{Var} X_i = \operatorname{Var}(a + Z_i) = \operatorname{Var} Z_i = \sigma^2 > 0.$$

Les hypothèses des deux questions précédentes sont donc vérifiées par la suite $(X_i)_{i\geq 1}$ et donc d'après (5), R_n converge presque sûrement vers a^2/σ^2. Par continuité sur $\mathbb{R}_+$ de l'application $x \mapsto \sqrt{x}$, on en déduit en rappelant la positivité de a, que

$$R_n^{1/2} = \frac{|S_n|}{(nV_n - S_n^2)^{1/2}} \xrightarrow[n\to+\infty]{\text{p.s.}} \frac{a}{\sigma}.$$

La variable aléatoire $R_n^{1/2}$ est donc un estimateur p.s. convergent (ou fortement consistant) du rapport signal sur bruit a/σ.

4) Notons Ω' l'ensemble des $\omega \in \Omega$ tels que pour aucun $n \geq 2$, il n'y ait d'*ex-aequo* dans la suite finie $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$. Autrement dit, Ω' est l'ensemble des $\omega \in \Omega$ tels que les termes de la suite infinie de réels $\bigl(X_i(\omega)\bigr)_{i\geq 1}$ soient tous distincts. On suppose dans cette question que les X_i ont une fonction de répartition continue. On sait qu'alors $P(\Omega') = 1$, cf. l'exercice vu en T.D. sur les *ex-ae quo* d'un échantillon. Nous allons utiliser cette propriété pour montrer que presque-sûrement, pour tout $n \geq 2$, $nV_n - S_n^2$ est *strictement* positif.

Rappelons que dans l'espace vectoriel euclidien usuel $\mathbb{R}^n$, le produit scalaire de deux vecteurs quelconques $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ est défini par

$$\langle x, y \rangle := \sum_{i=1}^n x_i y_i.$$

Pour ce produit scalaire, l'inégalité de Cauchy Schwarz s'écrit

$$(\langle x, y \rangle)^2 = \left(\sum_{i=1}^n x_i y_i \right)^2 \leq \|x\|^2 \|y\|^2 = \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right),$$

avec égalité si et seulement si les vecteurs x et y sont *colinéaires*, c'est-à-dire s'il existe un réel c tel que $x_i = cy_i$ pour tout $i = 1, \dots, n$. En particulier en gardant x quelconque et en prenant pour y le vecteur dont *toutes* les composantes valent 1, on obtient :

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad \left(\sum_{i=1}^n x_i \right)^2 \leq n \left(\sum_{i=1}^n x_i^2 \right), \tag{6}$$

avec égalité si et seulement s'il existe c tel que $x_i = c$ pour tout $i = 1, \dots, n$. Autrement dit, l'inégalité large (6) est *stricte*, sauf si *tous* les x_i sont égaux.

Ainsi pour tout $\omega \in \Omega'$ et tout $n \geq 2$, le vecteur $(x_1, \dots, x_n) = \big(X_1(\omega), \dots, X_n(\omega)\big)$ vérifie l'inégalité *stricte* dans (6), d'où $nV_n(\omega) - S_n^2(\omega) > 0$. Or on sait que $P(\Omega') = 1$. Donc presque sûrement, pour tout $n \geq 2$, $nV_n - S_n^2$ est strictement positif.

Ex 3. *La méthode de Monte-Carlo pour le calcul d'intégrales*

Soit $h : [0,1] \to \mathbb{R}$ une fonction continue. On se propose de donner une valeur approchée de

$$I := \int_0^1 h(x)\,\mathrm{d}x,$$

par une méthode probabiliste appelée « méthode de Monte-Carlo ». Pour cela on utilise la simulation informatique d'une suite $(U_i)_{i\geq 1}$ de variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. On pose

$$S_n := \sum_{i=1}^n h(U_i).$$

1) Rappelons un résultat du cours d'I.P.E. relatif à l'existence et au calcul de $\mathbf{E}h(Y)$ lorsque Y est une variable aléatoire à densité.

Proposition 1. *Soit Y une variable aléatoire de densité f et $h : \mathbb{R} \to \mathbb{R}$ une application réglée sur tout intervalle fermé borné de $\mathbb{R}$. Alors*

$$\mathbf{E}|h(Y)| = \int_{-\infty}^{+\infty} |h(x)| f(x)\,\mathrm{d}x. \tag{7}$$

Si $\mathbf{E}|h(Y)| < +\infty$, on a de plus

$$\mathbf{E}h(Y) = \int_{-\infty}^{+\infty} h(x) f(x)\,\mathrm{d}x. \tag{8}$$

En appliquant cette proposition avec $Y = U_1$, de densité $f = \mathbf{1}_{[0,1]}$ et h continue (donc réglée[1]), on obtient

$$\mathbf{E}|X_1| = \mathbf{E}|h(U_1)| = \int_{-\infty}^{+\infty} |h(x)| \mathbf{1}_{[0,1]}(x)\,\mathrm{d}x = \int_0^1 |h(x)|\,\mathrm{d}x.$$

Cette intégrale est une intégrale de Riemann ordinaire de la fonction continue $|h|$ sur l'intervalle fermé borné $[0,1]$, donc est finie. Ceci justifie l'intégrabilité de X_1 et nous permet d'écrire en traduisant (8),

$$\mathbf{E}X_1 = \mathbf{E}h(U_1) = \int_0^1 h(x)\,\mathrm{d}x = I.$$

1. Ici, h n'est définie que sur $[0,1]$, donc en toute rigueur, il faudrait la prolonger à tout $\mathbb{R}$, soit en une fonction continue en posant $h(x) := h(0)$ pour $x < 0$ et $h(x) := h(1)$ pour $x > 1$, soit en une fonction continue par morceaux en posant $h(x) := 0$ pour $x \notin [0,1]$.

2) La suite de variables aléatoires $(X_i)_{i\geq 1}$ est i.i.d. comme image de la suite i.i.d. $(U_i)_{i\geq 1}$ par la fonction continue (donc borélienne) h. On vient de voir que X_1 est intégrable. On peut donc appliquer à la suite $(X_i)_{i\geq 1}$ la loi forte des grands nombres [2], ce qui nous donne

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}X_1 = I.$$

Ce résultat légitime pour n « grand » l'approximation de I par la valeur $S_n(\omega)/n$ calculée à partir de l'échantillon généré par l'ordinateur. On s'intéresse maintenant au contrôle de l'erreur d'approximation.

3) En appliquant à nouveau la proposition 1 avec la fonction continue h^2 à la place de h, on obtient

$$\mathbf{E}|X_1^2| = \mathbf{E}h^2(U_1) = \int_0^1 h(x)^2 \,\mathrm{d}x,$$

ce qui est encore une intégrale de Riemann ordianire de la fonction continue h^2 sur l'intervalle fermé borné $[0,1]$. Cette quantité est donc finie et X_1 est bien de carré intégrable. De plus $\mathbf{E}X_1^2 = \int_0^1 h(x)^2 \,\mathrm{d}x$.

En appliquant la formule de Koenig, on obtient

$$\sigma^2 := \operatorname{Var} X_1 = \mathbf{E}(X_1^2) - (\mathbf{E}X_1)^2 = \int_0^1 h(x)^2 \,\mathrm{d}x - I^2.$$

Dans ce contexte, il n'est pas raisonnable de supposer σ connue, puisqu'on ne sait pas calculer $I = \int_0^1 h(x)\,\mathrm{d}x$. D'ailleurs, sauf cas particulier, le calcul exact de $\int_0^1 h(x)^2\,\mathrm{d}x$ risque lui aussi de ne pas être à notre portée.

Remarque. Pour les étudiants qui ont cru à tort ici qu'on leur demandait de vérifier la positivité de $\int_0^1 h(x)^2\,\mathrm{d}x - I^2$, voici quand même une réponse à cette question non posée. Il suffit d'appliquer l'inégalité de Cauchy-Schwarz pour les intégrales de Riemann au produit de fonctions $h = h \times 1$. Ceci nous donne

$$\left\{\int_0^1 h(x)\,\mathrm{d}x\right\}^2 = \left\{\int_0^1 h(x)\times 1\,\mathrm{d}x\right\}^2 \leq \left\{\int_0^1 h(x)^2\,\mathrm{d}x\right\}\left\{\int_0^1 1^2\,\mathrm{d}x\right\} = \int_0^1 h(x)^2\,\mathrm{d}x,$$

d'où $\int_0^1 h(x)^2\,\mathrm{d}x - I^2 \geq 0$.

4) Nous allons proposer un intervalle de confiance pour I en utilisant le théorème limite central. Ceci nous amène à rajouter l'hypothèse $\sigma^2 > 0$. En fait le cas d'égalité dans l'inégalité de Cauchy-Schwarz pour les intégrales de fonctions *continues* montre que si $\sigma^2 = 0$, alors h est constante sur $[0,1]$. Le calcul de son intégrale est alors trivial et il ne nécessite pas le recours à la méthode de Monte-Carlo ! Une autre façon de légitimer l'hypothèse $\sigma^2 > 0$ est d'utiliser le fait que les variables aléatoires de variance nulle sont constantes presque sûrement. Dans ce cas on aurait $h(U_i) = c$ pour tout i d'où $S_n/n = c$,

2. L'énoncé demandait de détailler soigneusement la vérification des hypothèses de ce théorème. Cela reviendrait à répéter ici l'argumentation déjà donnée dans le corrigé de l'exercice précédent sur l'hérédité de l'indépendance et sur le fait que l'égalité des lois de deux v.a. est préservée par composition de ces v.a. par la même fonction borélienne.

presque-sûrement pour tout n. Alors $S_n/n = I = c = \mathbf{E}h(U_1)$ p.s. pour tout n. Dans ce cas, il est inutile de calculer un grand nombre de X_i pour évaluer I, un seul suffit ! Et il n'y a pas davantage lieu de proposer un intervalle de confiance pour I puisqu'on a immédiatement sa valeur exacte.

Puisque les X_i sont i.i.d., de carré intégrable et $\sigma^2 > 0$, le théorème limite central nous donne la convergence en loi vers une v.a. Z gaussienne $\mathfrak{N}(0,1)$ de la suite des sommes centrées normalisées S_n^* qui s'écrivent ici :

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - nI}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\Bigl(\frac{S_n}{n} - I\Bigr).$$

En notant Φ la f.d.r. de $\mathfrak{N}(0,1)$, la convergence en loi de S_n^* vers Z implique, en tenant compte de la continuité de Φ, que pour tous réels $a < b$,

$$P(S_n^* \in [a,b]) \xrightarrow[n\to+\infty]{} \Phi(b) - \Phi(a).$$

On utilise cette convergence pour légitimer lorsque n est grand, l'approximation

$$P(S_n^* \in [a,b]) \simeq \Phi(b) - \Phi(a).$$

Ensuite, on note les équivalences logiques

$$S_n^* \in [a,b] \;\Leftrightarrow\; \frac{a\sigma}{\sqrt{n}} \le \frac{S_n}{n} - I \le \frac{b\sigma}{\sqrt{n}} \;\Leftrightarrow\; \frac{S_n}{n} - \frac{b\sigma}{\sqrt{n}} \le I \le \frac{S_n}{n} + \frac{a\sigma}{\sqrt{n}}.$$

En prenant $a = -t$ et $b = t$ pour $t > 0$, on obtient un intervalle d'encadrement de I centré sur S_n/n, la probabilité que cet encadrement soit vrai étant :

$$P\Bigl(\frac{S_n}{n} - \frac{t\sigma}{\sqrt{n}} \le I \le \frac{S_n}{n} + \frac{t\sigma}{\sqrt{n}}\Bigr) \simeq \Phi(t) - \Phi(-t) = 2\Phi(t) - 1,$$

la dernière égalité utilisant la parité de la *densité* Φ' de $\mathfrak{N}(0,1)$. En négligeant l'erreur due à l'approximation gaussienne, on peut dire que la probabilité de cet encadrement est au moins 95% pour tout t tel que $2\Phi(t) - 1 \ge 0{,}95$ ou encore $\Phi(t) \ge 0{,}975$. On a évidemment intérêt à choisir l'intervalle de longueur minimale, ce qui revient à prendre pour t la plus petite solution de cette inéquation, soit en raison de la croissance et de la continuité de Φ la solution de l'équation $\Phi(t) = 0{,}975$. La table des valeurs de Φ nous donne comme valeur numérique approchée 1,96. Nous pouvons donc proposer comme intervalle d'encadrement pour I

$$\Bigl[\frac{S_n}{n} - \frac{1{,}96\sigma}{\sqrt{n}}, \frac{S_n}{n} + \frac{1{,}96\sigma}{\sqrt{n}}\Bigr] \quad \text{avec une probabilité d'au moins 95\%,}$$

en négligeant l'erreur due à l'approximation gaussienne.

5) Cet intervalle dépendant de la quantité inconnue σ, n'est pas utilisable en pratique. Une façon de résoudre ce problème est de majorer σ lorsque l'on connaît un majorant M de $\sup_{x\in[0,1]} |h(x)|$. En effet on a alors $h^2 \le M^2$, d'où

$$\sigma^2 = \int_0^1 h(x)^2 \,\mathrm{d}x - I^2 \le \int_0^1 M^2 \,\mathrm{d}x = M^2.$$

On a donc $\sigma \leq M$. Remplacer σ par M élargit l'intervalle proposé ci-dessus et donc augmente (au sens large) la probabilité que cet intervalle aléatoire contienne I. Un intervalle de confiance pour I au niveau 95% est donc

$$\left[\frac{S_n}{n} - \frac{1{,}96M}{\sqrt{n}}, \frac{S_n}{n} + \frac{1{,}96M}{\sqrt{n}}\right]$$

6) Une autre méthode pour construire un intervalle de confiance pour I, consiste à estimer la variance inconnue σ^2 par la variance empirique

$$V_n := \frac{1}{n}\sum_{i=1}^{n} h(U_i)^2 - \Big(\frac{S_n}{n}\Big)^2 = W_n - \Big(\frac{S_n}{n}\Big)^2,$$

qui est calculable à partir de la simulation des U_i. La justification théorique de cette méthode repose sur la version « studentisée » du théorème limite central vue en cours[3]. Ce théorème nous dit que si on remplace σ par $V_n^{1/2}$ dans l'expression de S_n^*, alors la nouvelle suite des variables aléatoires S_n^{**} ainsi obtenues converge encore en loi vers Z de loi $\mathfrak{N}(0,1)$. On a donc

$$S_n^{**} := \frac{S_n - nI}{n^{1/2}V_n^{1/2}} = \frac{n^{1/2}}{V_n^{1/2}}\Big(\frac{S_n}{n} - I\Big) \xrightarrow[n\to+\infty]{\text{loi}} Z.$$

On en déduit comme à la question 4 que

$$P(S_n^{**} \in [a,b]) \xrightarrow[n\to+\infty]{} \Phi(b) - \Phi(a).$$

Une adaptation immédiate de ce qui a été fait à la question 4 conduit à proposer

$$\left[\frac{S_n}{n} - 1{,}96\sqrt{\frac{V_n}{n}}, \frac{S_n}{n} + 1{,}96\sqrt{\frac{V_n}{n}}\right]$$

comme intervalle de confiance pour I au niveau 95%.

3. Ce théorème a été démontré à l'aide du lemme de Slutsky, lors de l'étude des lois des moments empiriques pour un échantillon de grande taille. En cours, nous avons noté S^2 ce qui est ici noté V_n.