INSA

Année Universitaire 2025-2026

3ème Année Mathématiques Appliquées

Probabilités Projet N° 2 : Applications de la simulation probabiliste

1 Convergence de v.a. et simulation probabiliste

1.1 Rappels : loi faible et forte des grands nombres

Les v.a. sont supposées définies sur un espace probabilisé $(\Omega, \mathbb{P}, \mathcal{E})$.

Définition 1.1 (Convergence en probabilité) *Une suite de v.a.* $\{X_n\}_{n\geq 1}$ *converge en probabilité vers une v.a.* X *si*

$$\forall \varepsilon > 0, \quad p_n^c := \mathbb{P}\big\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\big\} \xrightarrow[n\to+\infty]{} 1 \tag{1a}$$

$$\iff \forall \varepsilon > 0, \quad p_n := \mathbb{P}\big\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\big\} \xrightarrow[n\to+\infty]{} 0. \tag{1b}$$

Théorème 1.1 (LfaibleGN) *Pour une suite de v.a.i.i.d.* $\{X_n\}_{n\geq 1}$ *et une fonction borélienne* $h(\cdot)$ *telle que* $\mathbb{E}\big[|h(X_1)|\big] < \infty$*, la suite* $\big\{\overline{h(X)} := \big\{(1/n)\sum_{k=1}^n h(X_k)\big\}_{n\in\mathbb{N}}$ *converge en probabilité vers* $\mathbb{E}[h(X_1)]$.

Si $\mathbb{E}\big[|h(X_1)|^2\big] < \infty$, la LfGN s'obtient à partir de l'inégalité de Tchebychev. Si $\mathbb{E}\big[|h(X_1)|\big] < \infty$, cela résulte de la convergence p.s. garantie par la loi forte des grands nombres (LFGN).

Définition 1.2 (Convergence p.s.) *Une suite de v.a.* $\{X_n\}_{n\geq 1}$ *converge presque sûrement (p.s.) vers une v.a.* X *s'il existe un ensemble* $A \subset \Omega$ *tel que* $\mathbb{P}(A) = 1$ *et*

$$\forall \omega \in A, \quad \lim_{n\to+\infty} X_n(\omega) = X(\omega). \tag{2}$$

ou encore $\mathbb{P}\big\{\omega \in \Omega : \lim_{n\to+\infty} X_n(\omega) = X(\omega)\big\} = 1$

La LFGN est un résultat de convergence central pour (en autre) la simulation probabiliste.

Théorème 1.2 (LForteGN) *Pour une suite de v.a.i.i.d.* $\{X_n\}_{n\geq 1}$*, si* $h(\cdot)$ *est une fonction borélienne telle que* $\mathbb{E}\big[|h(X_1)|\big] < \infty$ *alors il existe* $A \subset \Omega$ *tel que* $\mathbb{P}(A) = 1$ *et*

$$\forall \omega \in A, \quad \lim_{n\to+\infty} \overline{h(X)}_n(\omega) = \mathbb{E}[h(X_1)]. \tag{3}$$

Si on reprend le contexte de l'échantillonnage selon la loi d'une v.a. X alors $\{X_n\}_{n\geq 1}$ est une suite de v.a.i.i.d. qui satisfait (3). Autrement dit, plus on augmente la taille n de l'échantillon $X_1, \ldots, X_n$, plus la valeur de la v.a. $\overline{h(X)}_n$ donne une bonne approximation de $\mathbb{E}[h(X_1)] = \mathbb{E}[h(X)]$.
Le lemme suivant donne une formulation équivalente intéressante pour l'exploration de la convergence p.s. à l'aide de la simulation et pour la comparaison avec le mode de convergence en probabilité.

Lemma 1.1 (Annexe C) *Une suite de v.a.* $\{X_n\}_{n\geq 1}$ *converge p.s. vers une v.a.* X *ssi* $\forall \varepsilon > 0$

$$a_n^c := \mathbb{P}\big\{\omega \in \Omega : \forall k \geq n, |X_k(\omega) - X(\omega)| \leq \varepsilon\big\} = \mathbb{P}\big\{\omega \in \Omega : \sup_{k\geq n} |X_k(\omega) - X(\omega)| \leq \varepsilon\big\} \xrightarrow[n\to+\infty]{} 1 \tag{4a}$$

$$\iff a_n := \mathbb{P}\big\{\omega \in \Omega : \sup_{k\geq n} |X_k(\omega) - X(\omega)| > \varepsilon\big\} \xrightarrow[n\to+\infty]{} 0. \tag{4b}$$

Autrement dit, $\{X_n\}_{n\geq 1}$ *converge p.s. vers* X*, c'est à dire* $\{|X_n - X|\}_{n\geq 1}$ *converge p.s. vers 0, ssi* $\{Z_n := \sup_{k\geq n} |X_k - X|\}_{n\geq 1}$ *converge en probabilité vers 0.*

Il est clair à partir de (4b) et (1b) qu'on a

$$a_n^c \leq p_n^c \leq 1 \quad \text{ou encore} \quad 0 \leq p_n \leq a_n.$$

La convergence p.s. implique donc clairement la convergence en probabilité.

1.2 Illustrations des convergences en probabilité et p.s.

Application 1 (Estimer la valeur d'une probabilité par une fréquence) *Soit X une v.a. réelle et E un borélien quelconque de $\mathbb{R}$. Si $h(\cdot) = 1_E(\cdot)$ alors $h(X)$ suit une loi de Bernoulli de paramètre $p := \mathbb{P}\{h(X) = 1\} = \mathbb{P}\{X \in E\}$ et dans ce cas, la LFGN nous dit que l'on peut estimer cette probabilité par la fréquence d'observation de valeurs dans E pour un n-échantillon $X_1, \ldots, X_n$ de la loi de X :*

$$\overline{1_E(X)}_n = \frac{1}{n}\sum_{k=1}^{n} 1_E(X_k(\omega))$$

ou encore par la proportion de succès sur une réalisation $1_E(x_1), \ldots, 1_E(x_n)$ de $1_E(X_1), \ldots, 1_E(X_n)$.

On suppose que toutes les v.a. sont réelles. La procédure donnée dans Application 1 permet de déterminer des estimations $\widehat{a}_n, \widehat{p}_n$ des probabilités a_n, p_n dans (4b) et (1b). En effet, pour estimer $p := p_n$ **pour** n **fixé**, il suffit :

1. d'obtenir la réalisation $\{(x_n - x)^{(1)}, (x_n - x)^{(2)}, \ldots, (x_n - x)^{(M)}\}$ d'un M-échantillon $\{(X_n - X)^{(1)}, (X_n - X)^{(2)}, \ldots, (X_n - X)^{(M)}\}$ de $X_n - X$. On dira réaliser M répliques (indépendantes) de $X_n - X$.
2. Puis de prendre $\widehat{p}_n := (1/M)\sum_{k=1}^{M} 1_E\big((x_n - x)^{(k)}\big)$ où $E := \{y \in \mathbb{R} : |y| > \varepsilon\}$.

Pour étudier la convergence en probabilité de X_n vers X, il suffit d'explorer la convergence de $\widehat{p}_n$ vers 0. Pour la convergence p.s. on applique cette démarche à la v.a. $Z_n := \sup_{k \geq n} |X_k - X|$. Bien entendu, il faudra fixer le nombre de termes pris en compte dans le calcul du sup, à savoir la longueur de chaque réplique $(X_n - X)^{(\ell)}, (X_{n+1} - X)^{(\ell)}, \ldots$
Ces estimations sont mises en œuvre dans le package `ConvergenceConcepts` pour explorer ces deux modes de convergence vers une limite (connue) (cf Annexe A, p. 6). En particulier, des outils graphiques permettent une « analyse visuelle de la convergence » de $\widehat{p}_n, \widehat{a}_n$ vers 0. On se restreint ici à une limite X constante.

Commentaire 1.1 *Nous avons les paramètres M et n dans les deux étapes proposées. Pour n fixé : la valeur de M pilote le nombre d'échantillon/de répliques de $X_n - X$. L'estimation de $\widehat{p}_n$ repose sur la LFGN appliqué à ce M-échantillon, donc M doit être assez grand pour obtenir une certaine qualité de l'estimation de la probabilité $p_n := \mathbb{P}\{|X_n - X| > \varepsilon\}$. La question initiale est la convergence de p_n quand $n \to +\infty$, donc n devra être pris assez grand pour observer la convergence ou non. Attention, la vitesse de convergence peut être plus ou moins rapide.*

Code R1: Illustration de la convergence en probabilité

On considère un échantillon $\{X_n\}_{n\geq 1}$ d'une loi unif. sur $[1, 9]$. Pour tout $n \geq 1$, on pose $\widehat{b}_n := 2\,\overline{X}_n - 1$.

1. Utiliser `ConvergenceConcepts` pour étudier la convergence en probabilité de $\{\widehat{b}_n\}_{n\geq 1}$ vers 9. L'instruction `cumsum` sera utile dans la construction du générateur de la suite $\widehat{b}_1 - 9, \widehat{b}_2 - 9, \ldots, \widehat{b}_n - 9$ requis par le package.
2. Justifier la convergence en probabilité de la suite $\{\widehat{b}_n\}_{n\geq 1}$. A-t-on convergence p.s. ?

Code R2: Illustration de la convergence p.s.

Soit un échantillon $\{X_k\}_{k\geq 1}$ d'une loi unif. sur $[0, 1]$. Pour tout $n \geq 1$, on pose $\mathcal{M}_n := \max_{k=1,\ldots,n} X_k$. On souhaite explorer la convergence p.s. de $\{\mathcal{M}_n\}_{n\geq 1}$.

1. Donner une fonction `genMn` qui génère, pour n donné, les valeurs de $\mathcal{M}_1, \ldots, \mathcal{M}_n$ (utiliser la fonction `cummax`).
2. Générer $M := 5$ répliques (indépendantes) de la suite $\mathcal{M}_1, \ldots, \mathcal{M}_{100}$, puis construire la représentation graphique superposant les trajectoires des 5 répliques (utiliser la fonction `replicate` pour fabriquer les 5 répliques, la fonction `plot` avec l'option `type="l"` pour la première réplique, puis `lines` pour les 4 autres avec un code couleur différent).
3. Que peut-on conjecturer ?
4. Utiliser `ConvergenceConcepts` pour renforcer votre conviction.
5. Montrer la convergence p.s. de la suite $\{\mathcal{M}_n\}_{n\geq 1}$.

Code R3: Illustration de la différence entre convergence en probabilité et p.s.

Considérons la suite de v.a. indépendantes $\{X_n\}_{n\geq 1}$ où $X_n \sim \text{Ber}(1/\sqrt{n})$.

1. À l'aide de `ConvergenceConcepts`, explorer la convergence en probabilité et p.s. vers 0. Pour la convergence p.s. on pourra également tracer quelques répliques comme dans le CodeR2
2. Justifier théoriquement les conjectures issues de la précédente question.

1.3 Rappels : convergence en loi, Théorème Central Limite (TCL).

Définition 1.3 (Convergence en loi) *La suite de v.a. réelles $\{X_n\}_{n\geq 1}$ converge en loi vers la v.a. X, si pour tout t point de continuité de la fonction de répartition F_X de X, on a*

$$\lim_{n\to+\infty} F_{X_n}(t) = \lim_{n\to+\infty} \mathbb{P}\{X_n \leq t\} = \mathbb{P}\{X \leq t\} = F_X(t).$$

Une conséquence de la LFGN : la fonction de répartition empirique associée à une suite $\{Y_k\}_{k\geq 1}$ de v.a.i.i.d. de même loi de fonction de répartition F, est telle qu'il existe $A \subset \Omega$ tel que $\mathbb{P}(A) = 1$ et

$$\forall t \in \mathbb{R}, \forall \omega \in A, \quad F_M(t,\omega) := \frac{1}{M}\sum_{k=1}^{M} 1_{\{Y_k \leq t\}}(\omega) = \frac{1}{M}\sum_{k=1}^{M} 1_{]-\infty,t]}(Y_k(\omega)) \xrightarrow[M\to+\infty]{} F(t).$$

En fait, cette convergence peut être renforcée en :

Théorème 1.3 (Théorème de Glivenko-Cantelli) *Pour une suite $\{Y_k\}_{k\geq 1}$ de v.a.i.i.d. de loi commune de fonction de répartition F, alors il existe $A \subset \Omega$ tel que $\mathbb{P}(A) = 1$ et*

$$\forall \omega \in A, \quad \lim_{M\to+\infty} \sup_{t\in\mathbb{R}} \left|F_M(t,\omega) - F(t)\right| = 0. \tag{5}$$

Bilan : Dans la Définition 1.3, pour n fixé, on approche la fonction de répartition de X_n (si on ne la connaît pas) par la fonction de répartition empirique d'un M-échantillon de X_n, et, pour des valeurs croissantes de n, on compare la fonction empirique du M-échantillon de X_n à celle de X. L'intérêt de cette méthode est qu'elle s'applique quelque soit le type de v.a. considéré.

Commentaire 1.2 (Statistique de Kolmogorov-Smirnov) *La statistique de Kolmogorov-Smirnov définie par $K_M(\omega) := \sqrt{M}D_M(\omega)$ avec $D_M(\omega) := \sup_{t\in\mathbb{R}} |F_M(t,\omega) - F(t)|$ sert de base d'un test statistique d'adéquation à une loi donnée. La v.a. D_M converge p.s. vers 0 d'après (5). Le test de Kolmogorov-Smirnov (KS) utilise la statistique $K_M := \sqrt{M}D_M$. Si la loi de l'échantillon admet une fonction de répartition F* ***continue****, alors K_M converge vers en loi une v.a. de loi, dite de Kolmogorov-Smirnov, de fonction de répartition*

$$F_{KS}(t) := 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2t^2).$$

Noter que la loi limite ne dépend pas de F. Sous l'hypothèse que l'échantillon est bien de loi F alors il est possible de construire un test de cette hypothèse sur la base de cette convergence . Ce test est donné par la commande `ks.test` *de R. Voir le texte associé sous moodle (et 3MA-SI).*

Code R4: Illustration de la convergence de la fonction de répartition empirique

Illustrer graphiquement l'adéquation d'un générateur d'une loi binomial, unif. sur $[a,b]$, exponentielle de paramètre λ, ... en comparant les fonctions de répartition empirique et théorique. On utilisera `ecdf` pour créer la fonction empirique, `plot` pour obtenir le graphe, puis superposer la fonction de répartition théorique par l'instruction `curve` par exemple. Le cas du générateur d'un échantillon d'une loi unif. sur $[0,1]$ a été traité en Section 2 du Projet N° 1. Puis appliquer le test de KS.

Théorème 1.4 (Théorème Central Limite) *Pour une suite $\{X_n\}_{n\geq 1}$ de v.a.i.i.d. admettant un moment d'ordre deux avec $m := \mathbb{E}[X_1], \sigma^2 = \mathbb{V}(X_1)$, définissons les v.a.*

$$S_n := \sum_{k=1}^{n} X_k, \quad Z_n := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{V}(S_n)}} = \sqrt{n}\left[\frac{\overline{X}_n - m}{\sigma}\right].$$

Alors la suite $\{Z_n\}_{n\geq 1}$ *converge en loi vers une v.a.* Z *de loi* $\mathcal{N}(0,1)$*, c'est à dire*

$$\forall t \in \mathbb{R}, \quad \lim_{n\to+\infty} F_{Z_n}(t) = \lim_{n\to+\infty} \mathbb{P}\{Z_n \leq t\} = F_Z(t) := \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx. \tag{6}$$

Code R5: Illustration du TCL avec des lois du chi-deux

Considérons une suite de v.a. indépendantes $\{X_n\}_{n\geq 1}$ avec $X_n \sim \chi^2_n$. Montrer que le TCL permet d'établir la convergence en loi vers une v.a. $X \sim \mathcal{N}(0,1)$ de la suite de v.a.

$$Z_n := \frac{X_n - n}{\sqrt{2n}}.$$

Illustrer cette convergence à l'aide du package `ConvergenceConcepts` (voir Annexe A).

2 Intégration par des méthodes de Monte Carlo

La LFGN permet d'obtenir des estimations d'une espérance à partir d'une simulation probabiliste donc d'obtenir une valeur approchée d'un certain nombre d'intégrales ou de somme de série en fonction du caractère continu ou discret des v.a. concernées. En effet, il suffit que l'intégrale

$$I := \int_{\mathbb{R}} g(x)\, dx \tag{7a}$$

avec $\int_{\mathbb{R}} |g(x)|\, dx < +\infty$ soit interprétée comme une espérance sous la loi commune de l'échantillon, c'est à dire qu'elle se réécrive sous la forme

$$I = \int_{\mathbb{R}} h(x)\, f(x)\, dx \quad \text{où } h(x) := \frac{g(x)}{f(x)} \tag{7b}$$

où f est une densité de probabilité telle que $f(x) = 0 \Rightarrow g(x) = 0$. Noter que $\int_{\mathbb{R}} |h(x)|\, f(x)\, dx = \int_{\mathbb{R}} |g(x)|\, dx < +\infty$ et $I = \mathbb{E}[h(X)]$ où X est une v.a. admettant la densité f. Il suffit donc de simuler un n-échantillon de $(X_1, \ldots, X_n)$ de la loi de densité f, puis, par la LFGN (légitime car $\mathbb{E}[|h(X)|] < +\infty$), d'utiliser la moyenne empirique

$$\overline{h(X)}_n := \frac{1}{n} \sum_{k=1}^{n} h(X_k)$$

pour estimer la valeur de l'intégrale dans (7b). On adoptera la notation $\mathbb{E}_f$ pour signaler une espérance calculée relativement à la loi de densité f.
Par ailleurs, pour mesurer la qualité de l'approximation, on construit un intervalle de confiance pour I sur la base du TCL (avec estimation de la variance). La construction pour une espérance est donnée en Annexe B. Bien entendu cela requiert que

$$\mathbb{E}_f[h(X)^2] = \int_{\mathbb{R}} \frac{g(x)^2}{f(x)^2} f(x) dx = \int_{\mathbb{R}} \frac{g(x)^2}{f(x)} dx < +\infty.$$

Notons que cette condition n'est pas garantie par $\int_{\mathbb{R}} |g(x)|\, dx < +\infty$ ou $\int_{\mathbb{R}} |g(x)|^2\, dx < +\infty$. Sous la condition $\mathbb{E}_f[h(X)^2] < +\infty$, la variance de $\overline{h(X)}_n$ est

$$\mathbb{V}(\overline{h(X)}_n) = \frac{\mathbb{V}_f(h(X_1))}{n} = \frac{1}{n}\left[\int_{\mathbb{R}} h(x)^2\, f(x)\, dx - I^2\right].$$

Le TCL permet d'écrire

$$\forall t \in \mathbb{R}, \quad \lim_{n\to+\infty} \mathbb{P}\left\{\sqrt{n}\, \frac{\overline{h(X)}_n - I}{\sqrt{\mathbb{V}(h(X_1))}} \leq t\right\} = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\, dx$$

d'où l'intervalle de confiance de niveau 0.95 avec un échantillon de taille n

$$\mathrm{IC}_{n,0.95}(I) = \left[\overline{h(X)}_n \pm 1.96 \sqrt{\frac{\mathbb{V}_f(h(X_1))}{n}}\right].$$

Comme la valeur de I est inconnue, la valeur de $\mathbb{V}(h(X_1))$ l'est également. On la remplace alors par son estimation classique par la variance empirique

$$S_n^2(h) = \overline{h(X)^2}_n - (\overline{h(X)}_n)^2$$

ce qui donne, par le TCL avec estimation consistante de la variance, l'intervalle de confiance de I

$$\mathrm{IC}_{n,0.95}(I) = \left[\overline{h(X)}_n \pm 1.96\sqrt{\frac{S_n^2(h)}{n}}\right].$$

Remarque 1 *On pourra utiliser la fonction* plotCI *du package* plotrix *pour visualiser par exemple : l'évolution de* $\mathrm{IC}_{n,0.95}$ *en fonction de* n*, ou pour identifier pour* M *répliques de* $\mathrm{IC}_{n,0.95}$*, les intervalles ne contiennant pas la valeur théorique de l'intégrale.*

Remarque 2 *Notons que la transformation (7b) s'obtient facilement en posant* $h(x) := g(x)/f(x)$ *pour une quelconque densité de probabilité* f *telle que* $(\forall x : f(x) = 0 \Rightarrow g(x) = 0)$*. Sachant que l'on relie la qualité de l'approximation à l'intervalle de confiance obtenu ou à la variance de l'estimateur* $\overline{h(X)}_n$ *de* I*, et par là même à la valeur de la variance* $\mathbb{V}_f(h(X_1))$*, le choix de* f *peut être guidé par la volonté d'obtenir une variance la plus faible possible. Cette idée nous amène aux techniques dites de « réduction de la variance » en simulation probabiliste (ou aux méthodes d'échantillonnage préférentiel ou « Importance sampling methods »). Ce type de technique sera abordé en DMA09- Sim. et Estim. d'Événements rare.*

Code R6: Exemple jouet

Considérons l'exemple jouet de la fonction $g(x) := [\cos(50x) + \sin(20x)]^2$ et l'intégrale

$$I := \int_0^1 g(x)dx = \int_0^1 [\cos(50x) + \sin(20x)]^2\, dx.$$

Cette dernière se réécrit sous la forme

$$I := \int_{\mathbb{R}} g(x)1_{[0,1]}(x)\, dx = \int_{\mathbb{R}} g(x)f(x)\, dx$$

où f est la densité d'une loi unif. sur l'intervalle $[0, 1]$. Il suffit donc de produire un n-échantillon $X_1, \ldots, X_n$ d'une telle loi et d'estimer l'intégrale par la somme

$$\overline{h(X)}_n = \frac{1}{n}\sum_{k=1}^{n} h(X_k).$$

Réaliser des simulations de sorte à construire un graphe illustrant la convergence en fonction de la taille n sur lequel on superposera les intervalles de confiance associés à chaque estimation obtenue.

Code R7: Fonction à support non borné

On souhaite calculer l'intégrale

$$I := \int_{\mathbb{R}} \frac{1}{1+|x|^3}\, dx.$$

Proposer une estimation de I en introduisant

1. une densité de la loi de Cauchy : $f(x) := \dfrac{1}{\pi(1+x^2)}$;

2. une densité de type Laplace (voir Code R5 du TP N° 1) : $f(x) := \exp\big(-|x|\big)/2$.

Étudier la possibilité de construire un intervalle de confiance pour les deux cas de figure.

Remarque 3 (Monte Carlo vs Intégration numérique) *Les méthodes d'intégration numérique sont très efficaces pour un calcul sur* $\mathbb{R}$ *et la simulation ne peut pas les concurrencer. Par contre, si la dimension du domaine d'intégration augmente, ces méthodes rencontrent des difficultés importantes, en particulier en termes de vitesse de convergence, quand elles convergent. Le choix d'une méthode de Monte Carlo est donc pertinent pour calculer des intégrales multiples. Un des arguments classiques est que la vitesse de convergence (en* $1/\sqrt{n}$*) des méthodes de MC sont indépendantes de la dimension* d *de l'espace et qu'il*

n'y a aucune hypothèse de régularité de la fonction à intégrer (par exemple une fonction indicatrice). En effet les méthodes d'intégration par quadrature

$$\int h(x)\,dx \simeq \sum_{i=1}^{n} p_i h(x_i)$$

admettent une erreur en $O(n^{-s/d})$ où s la régularité de h (ordre de dérivabilité). Le choix des (p_i, x_i) dépend de s.

Remarque 4 (Monte Carlo et les zones de faible probabilité) *La simulation de MC via une mise en œuvre standard présente l'avantage de parcourir les zones d'intégration les plus probables alors qu'une implémentation classique d'une méthode numérique n'incorpore pas ce genre d'information. A contrario, pour estimer une petite probabilité, une mise en œuvre standard d'une méthode de MC ne « visitera » pas suffisamment souvent cette zone de faible probabilité/événement rare pour obtenir de bonnes estimations à partir de simulation d'échantillons de taille raisonnable. Dans un tel cas, des mises en œuvre spécifiques, comme l'échantillonnage préférentiel, sont développées.*

Code R 8: Échantillonnnage préférentiel et estimation de petite probabilité

On souhaite estimer $p = \mathbb{P}\{X > 4\}$ avec $X \sim \mathcal{N}(0,1)$.

1. Mettre en place la méthode classique de Monte Carlo pour estimer la valeur p.
2. Introduire une alternative en échantillonnant suivant la loi d'une v.a. $Y = E + 4$ avec $E \sim \text{Exp}(1)$. On précisera les fonctions g, f et h dans (7a)–(7b).
3. Comparer les deux en termes de taille d'échantillons pour obtenir un même niveau de précision. Conclusion.

A Package `ConvergenceConcepts` [LL09a, LL09b]

L'intérêt de ce package est de proposer des illustrations graphiques prêtes à l'emploi des modes de convergence les plus classiques. Cependant, il ne peut être employé qu'après avoir déterminé un candidat limite. Pensez à installer et charger le package via le menu de l'interface ou utiliser les commandes

```
install.packages("ConvergenceConcepts"); require(ConvergenceConcepts)
```

On pourra consulter l'aide en ligne du package
https://cran.r-project.org/web/packages/ConvergenceConcepts/index.html.

Exemple 1 (Fonction `check.convergence`) *Un appel avec une liste de paramètres :*

```
check.convergence(nmax,M,genXn,argsXn=NULL,mode="p",epsilon=0.05,r=2,nb.sp=10,
density=FALSE,densfunc=dnorm,probfunc=pnorm,tinf=-3,tsup=3,plotfunc=plot,...)
```

avec les arguments

- *nmax* : longueur d'une trajectoire de la suite.
- M : nombre de répliques ou réalisations de trajectoires de longueur *nmax*.
- `genXn` : une fonction qui génère les valeurs de $X_1 - \boldsymbol{X}, \ldots, X_n - \boldsymbol{X}$, ou seulement les valeurs de $X_1, \ldots, X_n$ pour la convergence en loi.
- `argsXn` : une liste d'arguments de `genXn`.
- `mode` : une chaîne de caractère qui spécifie le mode de convergence choisi. Elle doit être "p" (par défault), "as", "r" or "L".
- `epsilon` : une valeur numérique qui fixe la borne de l'intervalle $[-\varepsilon, \varepsilon]$.
- r : si on étudie la convergence en moment d'ordre $r > 0$.
- `nb.sp` : nombre de trajectoires à inclure dans le dessin de gauche.
- `density` : Si `density =TRUE`, alors le graphe de la densité de X et l'histogramme de X_n sont fournis. Si `density =FALSE`, alors le graphe de la fonction de répartition $F_X(t)$ et la fonction de répartition empirique $F_n(t)$ de X_n sont fournis.
- `densfunc` : fonction qui calcule la densité de X.
- `probfunc` : fonction qui calcule la fonction de répartition de X.
- `tinf` : borne inférieure du domaine pour étudier la convergence en loi
  `tsup` : borne supérieure du domaine pour étudier la convergence en loi.

Exemple 2 (Convergence en probabilité) *$X_1, \ldots, X_n$ est un n-échantillon d'une loi $\mathcal{N}(2, 9)$ et on étudie la convergence en probabilité vers 0 des variables $Y_1, \ldots, Y_n$ définie par $Y_k = (1/2)^k X_k$*

```
# définition du générateur de Y_1-0,...,Y_n-0
genYn = function(n) {
                      res = (0.5)^(1:n)*rnorm(n,2,3)  # ATTENTION : écart-type ici
                      return(res)
}
# appel nmax=2000, M=500
check.convergence(2000,500,genYn,mode="p")
```

On peut adapter le code pour avoir une fonction « générateur » qui intègre les paramètres supplémentaires de la moyenne et l'écart-type de la loi normale de X_k. Pour cela on ré-écrit la fonction et l'appel à `check.convergence` par

```
# définition du générateur de Y_1-0,...,Y_n-0
genYn = function(n,m,sigma) {
                      res = (0.5)^(1:n)*rnorm(n,m,sigma)
                      return(res)
}
# appel nmax=2000, M=500 et m=2, sigma=3
check.convergence(2000,500,genYn,argsXn=list(m=2,sigma=3),mode="p")
```

Exemple 3 (Convergence en loi) *$X_1, \ldots, X_n$ est un n-échantillon d'une loi $\mathcal{N}(2, 9)$. On étudie la convergence en loi des variables $S_1, \ldots, S_n$ définie par pour $k \geq 1$ par $S_k = \sum_{i=1}^{k} (1/2)^i X_i$, vers $S \sim \mathcal{N}(2, 3)$ sur la base de $M = 1000$ répliques d'une suite $S_1, \ldots, S_{2000}$:*

```
genSn = function(n) {
                       res = cumsum((0.5)^(1:n)*rnorm(n,2,3))
                       return(res)
}
check.convergence(2000,1000,genSn,mode="L",
 density = F,
 densfunc = function(x){dnorm(x,2,sqrt(3))},
 probfunc=function(x){pnorm(x,2,sqrt(3))},
 tinf = -4, tsup = 8
)
```

Références

[LL09a] P. Lafaye De Micheaux and B. Liquet. ConvergenceConcepts : an R package to investigate various modes of convergence. *R Journal*, 1 :18–26, 2009.

[LL09b] P. Lafaye De Micheaux and B. Liquet. Understanding convergence concepts : A visual-minded and graphical simulation-based approach. *The American Statistician*, 63 :173–178, 2009.

B Construction d'un intervalle de confiance (bilatéral) asymptotique de niveau $1-\alpha$ ($\alpha \in]0,1[$) pour une espérance

Soit $(X_1,\dots,X_n)$ n v.a.i.i.d. d'espérance et variance commune $m := \mathbb{E}[X_1], \sigma^2 := \mathbb{V}(X_1)$. On sait d'après la LFGN que

$$\overline{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{p.s.} m.$$

Un intervalle de confiance (exact et bilatéral) pour m de niveau $1-\alpha$ est un intervalle de la forme $\mathrm{IC}_{n,1-\alpha}(m) := [\overline{X}_n \pm \gamma_n]$ tel que

$$\mathbb{P}\big\{m \in \mathrm{IC}_{n,1-\alpha}(m)\big\} = \mathbb{P}\big\{m \in [\overline{X}_n \pm \gamma_n]\big\} = 1-\alpha. \tag{8}$$

Notons que α représente le risque de ne pas trouver m dans l'intervalle $\mathrm{IC}_{n,1-\alpha}(m)$. Il a donc vocation a être petit. Par ailleurs, (8) revient à trouver γ_n tel que

$$\mathbb{P}\big\{-\gamma_n \le \overline{X}_n - m \le \gamma_n\big\} = 1-\alpha.$$

autrement dit de contrôler les fluctuations de la moyenne empirique autour de m.
Ici on va construire un $\mathrm{IC}_{1-\alpha}(m)$ approché dans le sens où on obtiendra que pour n assez grand

$$\mathbb{P}\big\{m \in \mathrm{IC}_{n,1-\alpha}(m)\big\} \approx 1-\alpha. \tag{9}$$

Pour n assez grand. Le TCL donne les fluctuations de $\overline{X}_n$ autour de l'espérance pour n grand :

$$\sqrt{\frac{n}{\sigma^2}}\,[\,\overline{X}_n - m\,] \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$$

ou encore

$$\forall u>0,\quad \lim_{n\to+\infty}\mathbb{P}\left\{\sqrt{\frac{n}{\sigma^2}}\,|\overline{X}_n - m| \le u\right\} = \int_{-u}^{u}\frac{1}{\sqrt{2\pi}}\exp(x^2/2)dx = 2F_{\mathcal{N}(0,1)}(u)-1$$

avec $F_{\mathcal{N}(0,1)}$ la fonction de répartition d'une loi $\mathcal{N}(0,1)$. De cette dernière limite, comme

$$\sqrt{\frac{n}{\sigma^2}}\,|\overline{X}_n - m| \le u \Longleftrightarrow \overline{X}_n - u\sqrt{\frac{\sigma^2}{n}} \le m \le \overline{X}_n + u\sqrt{\frac{\sigma^2}{n}}$$

on en déduit que pour tout $u>0$

$$\forall u>0,\quad \lim_{n\to+\infty}\mathbb{P}\left\{\overline{X}_n - u\sqrt{\frac{\sigma^2}{n}} \le m \le \overline{X}_n + u\sqrt{\frac{\sigma^2}{n}}\right\} = 2F_{\mathcal{N}(0,1)}(u)-1$$

c'est à dire (9) avec

$$\gamma_n := u\sqrt{\frac{\sigma^2}{n}}.$$

Pour avoir une probabilité d'environ $1-\alpha$ pour n assez grand, il suffit de choisir u tel que

$$2F_{\mathcal{N}(0,1)}(u)-1 = 1-\alpha \Longleftrightarrow F_{\mathcal{N}(0,1)}(u) = 1-\alpha/2 \Longleftrightarrow u = Q_{\mathcal{N}(0,1)}(1-\alpha/2)$$

avec $Q_{\mathcal{N}(0,1)}$ la fonction quantile de la loi normale (cf table ou `qnorm`$(1-\alpha/2)$ sous R). Par exemple si on prend un niveau de confiance à 0.95, i.e. $\alpha = 0.05$ alors $u = 1.96$ et on on obtient finalement un intervalle de confiance (approché et bilatéral) de niveau 0.95 :

$$\mathrm{IC}_{n,1-\alpha}(m) = \left[\,\overline{X}_n \pm 1.96\sqrt{\frac{\sigma^2}{n}}\,\right] \quad \text{avec } \sigma^2 := \sigma^2(X_1).$$

L'intervalle de confiance ainsi obtenu dépend de la connaissance de σ^2. En général, si m est inconnu il en est de même de la variance σ^2. Cependant le TCL à la base de la construction reste valable quand σ^2 est remplacé par un estimateur dit consistent, c'est dire convergent en probabilité vers σ^2 (lemme de Slutsky). L'estimateur usuel, la variance empirique

$$S_n^2 := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 - \left(\overline{X}_n\right)^2,$$

possède cette propriété. En effet il n'est pas difficile de montrer que $\{S_n^2\}_{n\geq 1}$ converge p.s. vers σ^2 avec la LFGN. Au bilan, si σ^2 est inconnu alors l'intervalle de confiance au niveau $1-\alpha$ est

$$\mathrm{IC}_{n,1-\alpha}(m) = \left[\overline{X}_n \pm 1.96\sqrt{\frac{S_n^2}{n}}\right].$$

Par ailleurs, si on réalise M répliques d'un n-échantillon avec n assez grand et $\mathrm{IC}_{n,1-\alpha}^{(k)}(m)$ est l'intervalle de confiance de niveau $1-\alpha$ associé à la réplique n$^{\mathrm{o}}$ k, alors la LFGN

$$\lim_{M\to+\infty} \frac{1}{M}\sum_{k=1}^{M} 1_{\left\{m\in \mathrm{IC}_{n,1-\alpha}^{(k)}(m)\right\}} \overset{p.s.}{=} \mathbb{P}\{m\in \mathrm{IC}_{n,1-\alpha}^{(1)}(m)\} \approx 1-\alpha.$$

Autrement dit, si M est assez grand, on peut considérer qu'$(1-\alpha)\%$ des M intervalles de confiance $\left\{\mathrm{IC}_{n,1-\alpha}^{(k)}(m), k=1,\dots,M\right\}$ devraient contenir la valeur théorique de m.

C Preuve du Lemme 1.1

Une suite de v.a. $\{X_n\}_{n\geq 1}$ converge p.s. vers une v.a. X ssi $\{Z_n := \sup_{k\geq n}|X_k - X|\}_{n\geq 1}$ converge en probabilité vers 0.

1. Une étape préliminaire est de constater que la définition de la convergence d'une suite de réels $\{x_n\}_{n\geq 1}$ vers x

$$\forall \varepsilon>0, \exists k(\varepsilon)\in\mathbb{N}, \forall n\geq k(\varepsilon),\ |x_n - x|\leq \varepsilon$$

s'écrit de manière équivalente en

$$\forall \ell\geq 1, \exists k(\ell)\in\mathbb{N}, \forall n\geq k(\ell),\ |x_n - x|\leq \frac{1}{\ell}.$$

Le sens direct est prendre $\varepsilon = 1/\ell$. La réciproque résulte que pour $\varepsilon>0$ donné, il existe un $\ell(\varepsilon) = \lfloor 1/\varepsilon\rfloor + 1$ tel que $1/\ell\leq\varepsilon$. D'où pour tout $n\geq\ell(\varepsilon), |x_n - x|\leq 1/\ell\leq\varepsilon$.

2. La suite $\{X_n\}_{n\geq 1}$ converge p.s. vers X si l'ensemble $C := \left\{\omega\in\Omega : \lim_n X_n(\omega) = X(\omega)\right\}$ est de probabilité 1. Notons qu'avec la définition de la convergence de la suite de réels $\{X_n(\omega)\}_{n\geq 1}$ vers $X(\omega)$ on peut écrire

$$\begin{aligned}
C &= \left\{\omega\in\Omega : \forall\varepsilon>0, \exists k(\varepsilon,\omega)\in\mathbb{N}, \forall n\geq k(\varepsilon,\omega),\ |X_n(\omega)-X(\omega)|\leq\varepsilon\right\}\\
&= \left\{\omega\in\Omega : \forall\ell\geq 1, \exists k(\ell,\omega)\in\mathbb{N}, \forall n\geq k(\ell,\omega),\ |X_n(\omega)-X(\omega)|\leq\frac{1}{\ell}\right\} \quad \text{avec le point 1.}\\
&= \bigcap_{\ell\geq 1}\bigcup_{k\geq 1}\left\{\omega\in\Omega : \forall n\geq k, |X_n(\omega)-X(\omega)|\leq\frac{1}{\ell}\right\}\\
&\quad \text{cette écriture en termes d'intersection et réunion d'événements montrent que } C \text{ est un événement}\\
&\quad \text{et donc demander une probabilité associée est légitime}\\
&= \bigcap_{\ell\geq 1}\bigcup_{k\geq 1}\left\{\omega\in\Omega : \sup_{n\geq k}|X_n(\omega)-X(\omega)|\leq\frac{1}{\ell}\right\} = \bigcap_{\ell\geq 1}\bigcup_{k\geq 1}\left\{\omega\in\Omega : |Z_k(\omega)|\leq\frac{1}{\ell}\right\}
\end{aligned}$$

On a $\mathbb{P}(C) = 1 \Leftrightarrow \mathbb{P}(\overline{C}) = 0$ avec $\overline{C} = \bigcup_{\ell\geq 1}\bigcap_{k\geq 1}\{\omega\in\Omega : |Z_k|>1/\ell\}$. Et $\mathbb{P}(\overline{C}) = 0 \Leftrightarrow \forall\ell\geq 1, \mathbb{P}(\cap_{k\geq 1}\{\omega\in\Omega : |Z_k(\omega)|>1/\ell\}) = 0$ car proba d'une réunion d'ensembles est nulle ssi toutes les probas des ensembles sont nulles. Maintenant

$$\begin{aligned}
&\forall\ell\geq 1, \quad \mathbb{P}\left(\bigcap_{k\geq 1}\left\{\omega\in\Omega : |Z_k(\omega)|>\frac{1}{\ell}\right\}\right) = 0\\
\Leftrightarrow\quad &\forall\ell\geq 1, \quad \lim_{k\to+\infty}\mathbb{P}\left\{\omega\in\Omega : |Z_k(\omega)|>\frac{1}{\ell}\right\} = 0 \text{ car suite croissante d'événements}\\
\Leftrightarrow\quad &\forall\varepsilon>0, \quad \lim_{k\to+\infty}\underbrace{\mathbb{P}\{\omega\in\Omega : |Z_k(\omega)|>\varepsilon\}}_{a_k \text{ dans (4b)}} = 0 \quad \text{avec un argument de type 1.}\\
\Leftrightarrow\quad &\text{la suite } \{Z_n\}_{n\geq 1} \text{ converge en proba. vers 0.}
\end{aligned}$$