Le Package R ConvergenceConcepts: Un nouvel outil graphique pour l'étude de quelques modes de convergence de variables aléatoires

Pierre Lafaye de Micheaux, Benoit Liquet

LE PACKAGE R CONVERGENCECONCEPTS: UN NOUVEL OUTIL GRAPHIQUE POUR L'ÉTUDE DE QUELQUES MODES DE CONVERGENCE DE VARIABLES ALÉATOIRES

Pierre Lafaye de Micheaux[(a)] & Benoît Liquet[(b)]

[(a)] *Département de Mathématiques et de Statistique*
Université de Montréal CP 6128, succ. Centre-ville
Montréal, Québec H3C 3J7
Canada

[(b)] *INSERM U897, ISPED*
Université Victor Ségalen Bordeaux 2
France

Résumé

Ce travail tente d'éclairer l'intuition sous-jacente à la convergence de suites de variables aléatoires en utilisant un package **R** appelé `ConvergenceConcepts`. Ce package permet de jouer avec de telles suites. Il est ainsi possible de visualiser des trajectoires d'une suite donnée de variables aléatoires et d'étudier, au moyen de représentations dynamiques adaptées, les quatre modes les plus classiques de convergence, à savoir: la convergence presque sûre, la convergence en probabilité, la convergence en moyenne r et la convergence en loi.

Abstract

This paper tries to shed some light on the intuition behind the convergence of sequences of random variables using an **R**-package called `ConvergenceConcepts`. This package enables one to play with such sequences. It is thus possible to visualize sample paths of a given sequence of random variables and to investigate, through adequate graphical dynamical representations, the four classical modes of convergence: almost surely, in probability, in r-th mean and in law.

Mots-clés: Convergence de variables aléatoires, logiciel **R**, simulations.

1 Introduction

L'idée de limite d'une suite est l'un des concepts les plus anciens de l'Analyse Mathématique. Il remonte à Euclide, il y a près de 2000 ans ("*Étant données deux grandeurs inégales, si, de la plus grande on retranche plus que la moitié, et que du reste on retranche plus que la moitié et si l'on continue toujours ainsi, nous aboutirons à une grandeur inférieure à la plus petite des grandeurs donnée*"). Mais, en dépit de sa simplicité apparente, sa définition rigoureuse a été formalisée il y a seulement 300 ans, de façon indépendante par Newton et

Leibniz. De plus, il convient de noter que la définition moderne d'une limite (pour tout ϵ il existe un entier N tel que $\forall n \geq N$, $|x_n - x| < \epsilon$) est encore plus récente, puisqu'elle a été proposée indépendamment par Bernhard Bolzano (Der binomische Lehrsatz, Prague 1816) et par Cauchy dans son "*Cours d'analyse*" (1821).

Ainsi, nous avons maintenant une définition rigoureuse du concept de limite, et une idée intuitive de celle-ci: une suite de points $(x_n)_{n\in\mathbb{N}}$ convergeant vers un point appelé la limite. En effet, supposons que nous ayons une suite de nombre réels. Un point x est la limite d'une telle suite si pour tout voisinage de x, tous les points de la suite, sauf un nombre fini d'entre eux, sont dans ce voisinage. Cela peut être visualisé en prenant un ensemble de sphères de taille décroissante vers zéro, ayant toutes le même centre x, et pour lesquelles seulement un nombre fini de points de la suite se retrouveront en dehors de ces sphères.

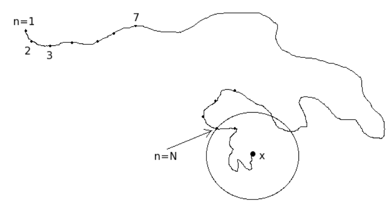


Figure 1: Convergence d'une suite de nombres réels vers une limite fixe.

Maintenant, notez que le concept de convergence d'une suite **aléatoire** de nombres réels n'a émergé qu'il y a un siècle. Par conséquent, il n'est pas surprenant que les étudiants éprouvent de la difficulté à appréhender ce nouveau concept. C'est l'objectif de ce travail que d'en exposer l'intuition sous-jacente, en utilisant un package **R** (Ihaka et Gentleman (1996)) de visualisation appelé `ConvergenceConcepts`, et déjà décrit dans Lafaye de Micheaux et Liquet (2009a,b). Dans la section 2, nous rappelons les définitions pour les quatre modes classiques de convergence de variables aléatoires, c'est-à-dire la convergence presque sûre, en probabilité, en moyenne r et en loi. Nous montrons aussi comment il est possible de les comprendre en visualisant mentalement ces suites aléatoires et en approchant les probabilités associées par des fréquences calculées sur certaines de ces réalisations. Dans la section 3, nous proposons un exemple et montrons comment il est possible de l'étudier en utilisant notre package **R**.

2 Convergence de suites aléatoires

2.1 Généralités

Supposons que $x_n, n = 1, 2, \ldots$ (et x) sont des nombres réels pris au hasard (le type de hasard n'a pas besoin d'être précisé pour le moment). En probabilité, nous pensons à ces valeurs en terme de réalisations d'un certain processus de génération aléatoire, et nous écrivons généralement:

$$x_n = X_n(\omega), \quad n = 1, \ldots \qquad \text{et} \qquad x = X(\omega),$$

où $\omega \in \Omega$ peut être vu comme un “index” ou une étiquette de l'expérience particulière qui a mené à ces valeurs observées.

La question est maintenant: *Quelle pourrait être une définition de la convergence de* $(X_n)_{n\in\mathbb{N}}$ *vers* X*?*.

Cette question est assez difficile puisque, pour chaque $\omega \in \Omega$ nous avons une valeur $X_n(\omega)$ et donc X_n (et X) peut être vu comme une sorte de “nuage” de valeurs réelles! Les probabilistes et les statisticiens ont proposés plusieurs réponses à cette question. Il est possible de les expliquer aux étudiants, **au moyen d'exemples** et en **utilisant notre package R** `ConvergenceConcepts`. De façon plus précise, nous montrerons quelle est la signification d'une écriture telle que celle-ci:

$$X_n \xrightarrow[n\to\infty]{\square} X,$$

où $\square$ sera remplacé par $p.s.$, P, r ou L.

2.2 Convergence presque sûre

Nous disons que X_n converge presque sûrement vers X et nous écrivons $X_n \xrightarrow{p.s.} X$ quand

$$P[\omega \in \Omega; \lim_{n\to\infty} X_n(\omega) = X(\omega)] = 1.$$

Sans perte de généralité, nous prendrons $X = 0$. Notez qu'en utilisant le Lemme 1, il est possible de visualiser ce ce mode de convergence en “regardant” des blocs glissants de longueur infinie, s'appuyant à des valeurs croissantes successives de n. Si, au fur et à mesure que n croit, la fréquence des portions de trajectoires (les portions à l'intérieur du bloc à l'étude) se trouvant à l'intérieur de la bande horizontale $[-\epsilon, +\epsilon]$ va vers 1, alors nous avons un bon indice qu'il y a une telle convergence.

Lemma 1 (Ferguson (1996), p.5)**.** $X_n \xrightarrow{p.s.} X$ *si et seulement si pour tout* $\epsilon > 0$,

$$P[|X_k - X| > \epsilon, \forall k \geq n] \longrightarrow 0, \text{ quand } n \to \infty.$$

La Figure 2 illustre ceci avec 10 réalisations fictives d'une certaine suite aléatoire.

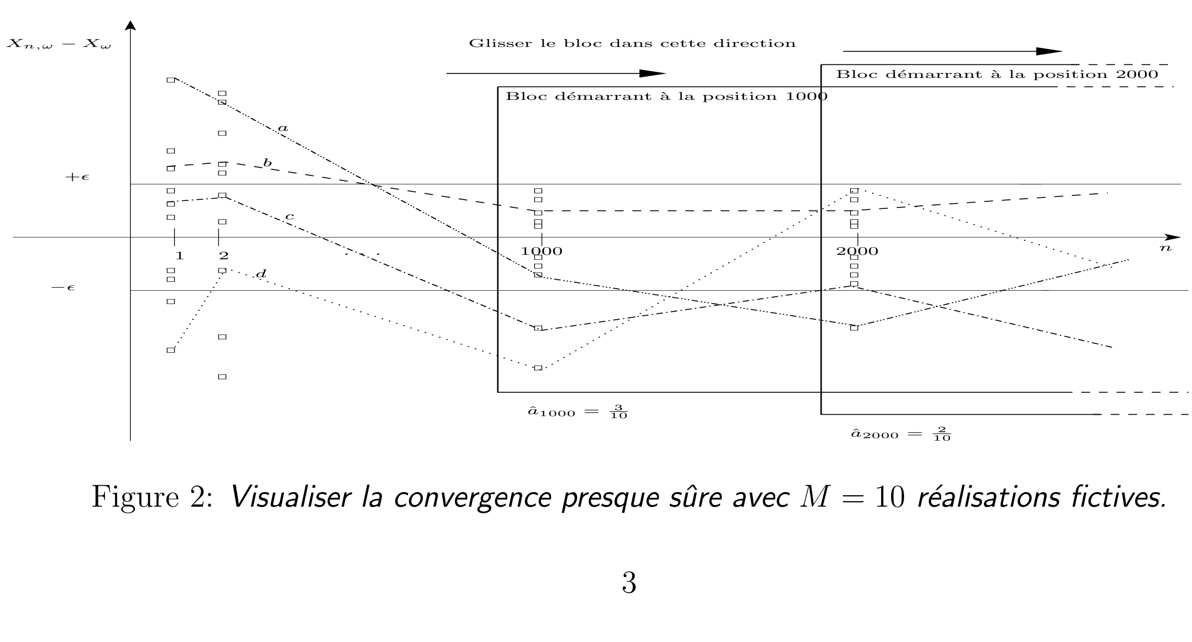


Figure 2: *Visualiser la convergence presque sûre avec* $M = 10$ *réalisations fictives.*

Ici, il est possible d'essayer notre package **R** en sélectionnant la première entrée de la fenêtre s'affichant lorsque l'on tape `investigate()` dans la console de **R**.

2.3 Convergence en probabilité

On dit que X_n converge en probabilité vers X et on écrit $X_n \xrightarrow{P} X$ quand

$$\forall \epsilon > 0,\ p_n = P\left[\omega \in \Omega;\ |X_n(\omega) - X(\omega)| > \epsilon\right] \underset{n\to\infty}{\longrightarrow} 0.$$

Sans perte de généralité, on prendra $X = 0$. Le même type d'intuition que pour la convergence presque sûre peut être utilisé ici. Dans ce cas, il faut visualiser les trajectoires d'une suite aléatoire et imaginer l'évolution en fonction de n de la fréquence de ces trajectoires qui tombent dans la bande horizontale $[-\epsilon, +\epsilon]$ à l'intérieur de la barre verticale glissante. Ceci est illustré sur la Figure 3.

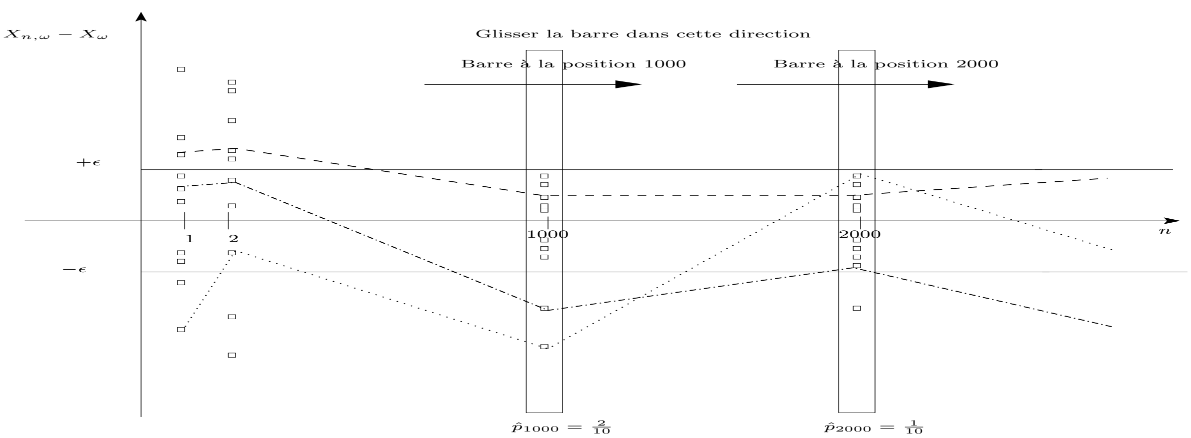


Figure 3: *Visualiser la convergence en probabilité avec $M = 10$ réalisations fictives.*

À ce stade, il est possible d'essayer notre package **R** en sélectionnant la première entrée de la fenêtre qui s'affiche lorsque l'on tape `investigate()` dans la console de **R**, puis de cliquer sur le bouton intitulé *Probability* sur le panneau de droite de la nouvelle fenêtre qui s'affiche.

2.4 Convergence en moyenne r et convergence en loi

On dit que X_n converge en moyenne r vers X et on écrit $X_n \xrightarrow{r} X$ quand $e_{n,r} = E|X_n - X|^r \underset{n\to\infty}{\longrightarrow} 0$. On dit que X_n converge en loi vers X et on écrit $X_n \xrightarrow{L} X$ quand $l_n = |F_n(t) - F(t)| \underset{n\to\infty}{\longrightarrow} 0$ en tout point t pour lequel F est continue.

Dans ce cas, l'intuition est moins attractive. En fait, on tracera seulement l'évolution de $\hat{e}_{n,r} = \frac{1}{M}\sum_{j=1}^{M}|x_n^j - x^j|^r$ en fonction de n dans le premier cas et celle de $\hat{l}_n = |\hat{F}_n(t) - F(t)|$ dans le second, où

$$\hat{F}_n(t) = \frac{\#\{x_n^j; x_n^j \leq t, j = 1, \ldots, M\}}{M}$$

et où M est le nombre de réalisations de la suite aléatoire à l'étude.

À ce stade, il est possible d'essayer notre package **R** en sélectionnant la deuxième entrée dans la fenêtre qui s'affiche après avoir tapé la commande `investigate()` dans la console de **R**.

3 Un exemple

Soient $X_1, X_2, \ldots$ des variables aléatoires *i.i.d.* suivant une loi uniforme sur $[0, 1]$. On définit $M_n = \max\{X_1, \ldots, X_n\}$. On peut maintenant se poser les questions suivantes.

(a) Est-ce que M_n converge en probabilité vers 1?
(b) Est-ce que M_n converge presque sûrement vers 1?
(c) Est-ce que M_n converge en moyenne quadratique vers 1?

La réponse à la question (a) est clairement *oui* puisque, pour chaque valeur croissante de n, on peut imaginer que M_n croitra vers 1, et par conséquent que la proportion de trajectoires qui tomberont dans la bande horizontale $[1-\epsilon, 1+\epsilon]$ à l'intérieur de la barre verticale glissante augmentera vers 1. Ceci peut être visualisé en utilisant la suite d'instructions **R** ci-dessous:

```
Mn.moins.1 <- function(n) cummax(runif(n))-1
null <- check.convergence(nmax=1000,M=500,genXn=Mn.moins.1,mode="p")
```

La réponse à la question (b) est aussi *oui* et peut être visualisée en tapant:

```
null <- check.convergence(nmax=1000,M=500,genXn=Mn.moins.1,mode="as")
```

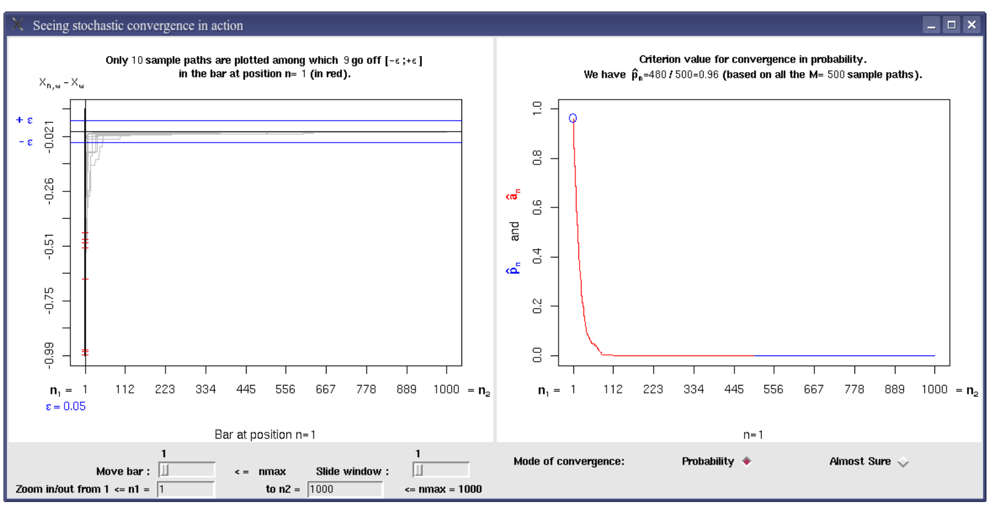
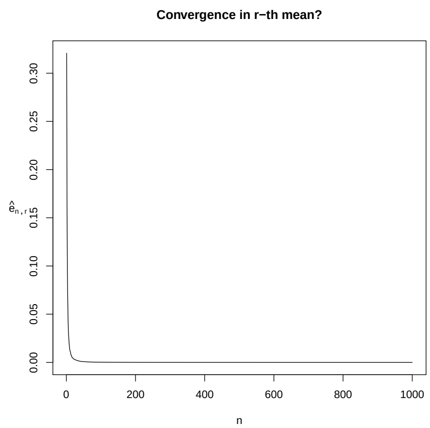


Figure 4: Visualisation de la convergence en probabilité et presque sûre de M_n vers 1.

La réponse à la question (c) est aussi *oui* et peut être visualisée en tapant:

```
check.convergence(nmax=1000,M=500,genXn=Mn.moins.1,mode="r",r=2)
```



4 Conclusion

Nous avons montré les possibilités offertes par le logiciel **R** pour aider les étudiants à forger leur intuition sur des concepts difficiles de la théorie statistique, c'est-à-dire la notion de convergence d'une suite de variables aléatoires.

Bibliographie

[1] T. Ferguson (1996) A course in Large Sample Theory. Chapman and Hall.
[2] R. Ihaka et R. Gentleman, (1996) **R**: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
[3] P. Lafaye de Micheaux et B. Liquet (2009a) Understanding convergence concepts: A visual-minded and graphical simulation-based approach. *The American Statistician*, 63(2), 173–178.
[4] P. Lafaye de Micheaux et B. Liquet (2009b) `ConvergenceConcepts`: An **R** Package to Investigate Various Modes of Convergence, *The* **R** *Journal*, 1(2), 18–25.