

## Cours n° 3 : Suites de variables aléatoires

# 1 Convergence de suites de variables aléatoires

Rappels convergences presque sûre, en loi.

presque sûre : définition

en loi : définition par convergence faible, convergence des fonctions de répartition avec convergence uniforme si limite continue (introduire distance de Kolmogorov, utiliser le second théorème de Dini), Lemme de Scheffé pour fonctions de masse (convergence simple nécessaire et suffisante) et densités (convergence simple suffisante).

### 1.1 Illustration numérique de la convergence

Nous allons voir des méthodes numériques pour illustrer les différents types de convergence de variables aléatoires. Ces méthodes sont principalement basés soit sur des *simulations*, c'est-à-dire de générations de réalisations de la suite de v.a. par des générateurs de nombres pseudo-aléatoires. Elles rentrent alors dans la classe des méthodes dites de *Monte Carlo*. Dans le cadre de la convergence en loi, on peut aussi parfois calculer *numériques les lois des v.a.*, par exemple leur densité ou fonction de masse.

Une convergence est un résultat asymptotique qui porte sur le comportement d'une infinité de variables aléatoires. Dans le cadre du calcul numérique ou de la simulation, l'ordinateur ne peut traiter qu'un nombre fini de variables aléatoires (ou leurs lois). On ne peut donc pas « prouver » par ces méthodes qu'une suite de v.a. converge d'un sens ou d'un autre, ni même quantifier les « chances » (*qu'est-ce que cela veut dire en fait ?*) que la suite converge en fonction des calculs numériques ou des résultats de simulations. Pour cette raison, on ne parle ici que « d'illustration » de la convergence : l'ordinateur apportera alors des « indicateurs » plus ou moins précis qui peuvent confirmer ou infirmer l'hypothèse d'une telle convergence (ou de son absence), mais il sera *a priori* impossible de connaître le degré de « confiance » que l'on peut avoir en les résultats. De surcroît, ceux-ci peuvent parfois être ambigus et difficiles à interpréter.

### 1.2 Illustration de la convergence en loi par calcul de la loi

Dans certains cas, par exemple des sommes de v.a. indépendantes ou des chaînes de Markov sur des espaces d'état petits, on peut calculer explicitement ou approcher numériquement la loi de  $X_n$ , i.e. fonction de répartition ou densité ou fonction de masse. Pour illustrer le fait que  $X_n \rightarrow X$  en loi, on peut alors tracer pour quelques « grandes » valeurs de  $n$  les fonctions de répartition de  $X_n$  et les superposer. Ces fonctions de répartition doivent alors être « proches », ce qu'on vérifiera « à vue d'oeil ». Si on connaît également la loi de  $X$  (ou si on a un candidat explicite), alors on peut également superposer sa fonction de répartition.

Dans le cas où les  $X_n$  sont discrètes et ont le même support, on pourra faire de même avec les fonctions de masse. Dans le cas où  $X$  et  $X_n$  sont à densité, on pourra faire de même avec les densités, en prenant en compte qu'il n'est pas nécessaire que la densité de  $X_n$  converge vers la densité de  $X$  en tout point (celui est cependant suffisant par le *Lemme de Scheffé*).

Qu'est-ce que  $n$  « grand » ? Cela dépend de la vitesse de convergence. Celle-ci se mesure par exemple par la distance de Kolmogorov si la limite est continue, ou par la distance  $\ell_\infty$  ou  $\ell_1$  entre les fonctions de masse dans le cas discret. Si la vitesse est alors exponentielle, alors normalement  $n$  de l'ordre de quelques dizaines suffit

largement. Si celle-ci est polynomiale, il sera éventuellement nécessaire de prendre  $n$  plus grand, par exemple de l'ordre de quelques centaines, voir milliers, en fonction de l'exposant de la vitesse de convergence.

Pour mesurer la vitesse de convergence dans le cas où celle-ci est polynomiale, on peut effectuer en tracé log-log de la distance en fonction de  $n$  : celui-ci doit alors avoir comme asymptote une droite dont la pente est égale à l'exposant du polynôme.

### 1.3 Illustration de la convergence presque sûre par simulation

Contrairement à la convergence en loi, la convergence presque sûre porte sur la loi de la suite entière  $(X_n)_{n \geq 1}$  et non pas seulement sur les lois marginales. Le calcul explicit de la loi de  $(X_1, \dots, X_n)$  pour  $n$  grand est alors impossible en pratique, l'espace d'état étant trop grand. La seule façon utilisable en pratique pour illustrer la convergence presque sûre est alors par simulation, dite « méthode Monte Carlo. »

*A priori* la méthode est évidente : il suffit d'illustrer que pour une réalisation typique  $\omega$ , la suite  $(X_n(\omega))_{n \geq 0}$  s'approche d'une certaine valeur  $X$  (qui peut être elle-même aléatoire). Pour ce faire, on peut échantillonner  $M$  réalisations et tracer pour chacune de ses réalisations les valeurs de  $X_n$  en fonction de  $n$ , pour  $n = 1, \dots, N$  avec  $N$  grand ; on observe alors une convergence presque sûre si ces tracés admettent chacun une asymptote horizontale.

Pourtant, le diable est dans le détail et plusieurs questions se posent :

**Comment choisir  $M$  ?** Supposons pour le moment que nous sachions parfaitement déterminer pour une réalisation  $(X_n(\omega))_{n \geq 0}$  d'une suite de v.a. si elle converge. Que peut-on dire alors si on observe convergence pour une telle réalisation obtenue par simulation ? *A priori*, on peut seulement dire que la probabilité de cette convergence est non-nulle, mais pas qu'elle vaut 1. Pour cela, il faut échantillonner plusieurs, disons  $M$ , réalisations indépendantes de la suite. Si la convergence a lieu presque sûrement, alors on observera convergence dans toutes les  $M$  réalisations. Dans le cas contraire, on observera non-convergence dans une fraction positive des réalisations avec grande probabilité. Si  $M$  est suffisamment petit, alors on peut superposer les différents tracés sur une seule figure.

Comment faut-il choisir  $M$  ? Cela dépend du degré de confiance que l'on souhaite avoir en ces simulations et de la probabilité de l'événement de convergence. Par exemple, si la suite converge avec probabilité  $p = 0.5$ , alors la probabilité qu'on observe convergence dans toutes les  $M$  réalisations vaut  $p^M = 1/2^M$ , donc déjà avec  $M = 5$  il y a moins de 4% de chance que toutes les réalisations montrent convergence. Par contre, si la probabilité de l'événement de convergence est  $p = 95\%$ , alors la probabilité d'observer convergence dans toutes les  $M$  réalisations vaut  $0.95^M \approx e^{-M/20}$ , donc il faudrait choisir environ  $M \approx 100$  pour pouvoir distinguer les cas  $p = 0.95$  et  $p = 1$ .

On peut formaliser le raisonnement ci-dessus par un test statistique, dans lequel on souhaite distinguer entre deux hypothèses, disons  $H_0 : p = 1$  et  $H_1 : p < 1$ , et on rejette  $H_0$  si on observe une réalisation parmi les  $M$  réalisations indépendantes sans convergence. L'erreur de première espèce est alors 0, car on ne rejettéra jamais  $H_0$  à tort, et la puissance du test est  $1 - p^M$ , comme nous l'avons calculée ci-dessus. On choisira alors  $M$  en fonction de la puissance souhaitée.

Evidemment, si on doit choisir  $M$  grand, il devient impraticable d'inspecter chacun des  $M$  tracés et il faudra avoir recours à des méthodes différentes ; on en reparle ci-dessous.

Notons cependant que dans de nombreuses situations, on peu montrer *a priori* que la probabilité de l'événement de convergence presque sûre est soit 0, soit 1, par exemple quand on est dans le cadre de la loi 0-1 de Kolmogorov, ou encore quand il s'agit d'une chaîne de Markov irréductible. Dans ces cas-là, une seule réalisation suffit en théorie pour déterminer laquelle des deux situations prévaut.

**Comment choisir  $N$  ?** Comme pour la convergence en loi ci-dessus, ceci dépend de la vitesse de convergence. Si elle est exponentielle, alors normalement  $n$  de l'ordre de quelques dizaines suffit largement. Si elle est polynomiale, il sera éventuellement nécessaire de prendre  $n$  plus grand, par exemple de l'ordre de quelques centaines, voir milliers, en fonction de l'exposant de la vitesse de convergence. Si la vitesse de convergence est encore plus lente, par exemple logarithmique, alors il est probable que les simulations ne soient pas concluantes.

**Comment s'assurer que l'asymptote que l'on observe en est vraiment une ?** Inversement, comment s'assurer que l'absence d'une asymptote évidente est vraiment un indicateur de non-convergence ou plutôt due à une convergence lente ? Ceci est un problème particulièrement fréquent en probabilités, où on observe souvent des convergences de vitesse  $1/\sqrt{n}$ .

Ceci dépend encore de la vitesse de convergence, ou, à l'inverse, de l'échelle à laquelle la trajectoire « fluctue ». Si la vitesse de convergence est exponentielle, avec une décroissance nette dès le début du tracé, alors on n'aura en général pas de difficulté de l'observer. Inversement, si la suite  $X_n$  fluctue beaucoup, avec des fluctuations qui ne ralentissent pas (c'est-à-dire que les différences  $X_n - X_{n-1}$  ne tendent pas vers 0), alors c'est un indicateur que la suite ne converge pas p.s.

Si, en revanche, la vitesse de convergence est plus lente qu'exponentielle ou si des fluctuations sont présentes à des échelles de plus en plus grandes, alors il convient de faire plusieurs tracés pour de différentes valeurs de  $N$ , afin d'illustrer la décroissance en passant d'une échelle à l'autre, ou, inversement, la présence de fluctuations à chaque échelle. Alternativement, on peut effectuer un seul tracé mais avec un changement d'échelle dans l'abscisse. Par exemple, si la vitesse de convergence est polynomiale en  $n$ , alors on peut faire un tracé à l'échelle logarithmique dans l'abscisse, ce qui rend la convergence exponentielle. De même, un changement d'échelle logarithmique est utile pour illustrer une non-convergence si les écarts entre des fluctuations croissent exponentiellement vite.

## 1.4 Illustration de la convergence en loi par simulation

Méthode de choix si on ne sait pas calculer explicitement la loi. Méthode : générer  $M$  échantillons de la suite et tracer fonctions de répartition ou histogrammes ou fonctions de masse empiriques pour quelques valeurs de  $n$ . L'histogramme est plus compliqué, car il faut choisir le nombre de classes. On choisira  $M$  en fonction de la précision souhaitée. Par exemple, pour les fonctions de répartition empiriques, on peut montrer que l'erreur décroît en  $1/\sqrt{M}$ , il faut donc choisir  $M \approx 1/\varepsilon^2$ , si  $\varepsilon$  est la précision souhaitée. La convergence en loi s'illustre alors par le fait que pour  $n$  grand, les tracés se ressemblent.

Inconvénient de la méthode : on étudie seulement un petit nombre de  $X_n$ . Il faut alors avoir confiance (ou justifier) que les  $X_n$  pour des valeurs intermédiaires de  $n$  se comportent de manière similaire. Sinon, on peut calculer pour tout  $n$  la distance (de Kolmogorov) entre les fonctions de répartition empiriques et celle de la limite et la tracer en fonction de  $n$ . Si  $M$  est choisi suffisamment grand, on voit alors une décroissance vers 0. Problème : très gourmand en calcul : complexité  $M \times N$ , mais ça peut se paralléliser facilement.