Ecole Internationale des Sciences du Traitement de l'Information

CHAPITRE 2 : CONVERGENCES D'UNE SUITE DE VARIABLES ALEATOIRES

Table des matières

1. CONVERGENCE EN LOI

1.1. Définition

Soient l'espace probabilisé $(\Omega,\mathcal{A},P)$ et une suite de variables aléatoires $X_1,\dots,X_n$ sur $(\Omega,\mathcal{A})$. Notons $F_1,\dots,F_n$ les fonctions de répartition correspondantes. On dit que la suite $\{X_1,\dots,X_n\}$converge *en loi* vers une variable aléatoire X de fonction de répartition F si

$$\lim_{n\to+\infty} F_n(t)=F(t)$$

pour tout t appartenant au domaine de continuité de F. Ce type de convergence est noté

$$X_n \xrightarrow{\mathcal{L}} X$$

<u>Remarque</u> : Nous avons vu en probabilités 1 que la loi binomiale converge en loi vers une loi de Poisson.

1.2. Théorème de la limite centrale

<u>Théorème</u> (théorème de la limite centrale)
Soit $X_1,\ldots,X_n$ une suite de variables aléatoires indépendantes de même loi telle que $E(X_i)=\mu$ et $var(X_i)=\sigma^2$. Alors la variable aléatoire

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma},$$

converge en loi vers une variable aléatoire Y de la normale centrée réduite,

$$Y_n \xrightarrow{\mathcal{L}} N(0,1)$$

$$\Leftrightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathcal{L}} N(\mu, \frac{\sigma^2}{n})$$

Ce théorème joue un rôle capital en statistique car l'étude de somme de variables aléatoires indépendantes est primordial.

<u>Exemple 1</u>
Un camion peut transporter une charge maximale de 10 tonnes. Il doit livrer des lots de marchandise dont le poids est une variable aléatoire d'espérance 95 kg et d'écart-type 2kg. Combien peut-il livrer de lot pour être certain à 99% de ne pas dépasser sa charge maximale ?

1.3. Applications du théorème de la limite centrale

<u>Convergence de la loi binomiale vers la loi normale</u>
Soit $X_1,\ldots,X_n$ une suite de variables aléatoires indépendantes de loi binomiale b(n,p), alors

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} N(0,1).$$

En pratique, on utilise cette approximation lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

<u>Convergence de la loi de Poisson vers la loi normale</u>
Soit $X_1,\ldots,X_n$ une suite de variables aléatoires indépendantes de loi de Poisson $P(\lambda)$, alors

$$\frac{X_n - \lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{L}} N(0,1).$$

En pratique, on utilise cette approximation lorsque $\lambda \geq 20$.

Dans les deux cas, on approche une loi discrète par une loi continue. Le calcul de la probabilité P(X=k) ne peut donc se faire que par encadrement et en général on utilise l'approximation,

$$P(X=k) \approx P(k - \tfrac{1}{2} < X < k + \tfrac{1}{2}).$$

C'est ce qu'on appelle la correction de continuité.

<u>Exemple 2</u>

Soit un échantillon de loi binomiale b(50,0.40). La valeur exacte pour

$$P(X=20)=P(X\leq 20)-P(X\leq 19)=0.1145.$$

On a n assez grand et np=20 et n(1-p)=12, on peut donc considérer que X suit une loi normale $N(12.5\ ;3.06^2)$, d'où

$$P(X=20)\approx P(19.5<X<20.5)=P(-0.144<Z<0.144)=0.1114$$

où Z suit une loi N(0,1). L'erreur d'approximation est d'autant plus grand que la taille de l'échantillon est petite.

$$\begin{tikzcd}[column sep=huge, row sep=huge]
& X\sim b(n,p) \arrow[dl, "\begin{cases} n\geq 30 \\ p\leq 0.1\end{cases} \Rightarrow \lambda = np"'] \arrow[dr, "\begin{cases} n\geq 30 \\ np\geq 5 \\ n(1-p)\geq 5\end{cases} \Rightarrow \begin{cases}\mu = np \\ \sigma^2 = np(1-p)\end{cases}"] & \\
X\sim P(\lambda) \arrow[rr, "\lambda\geq 20 \Rightarrow \begin{cases}\mu=\lambda \\ \sigma^2=\lambda\end{cases}"] & & X\sim N(\mu,\sigma^2)
\end{tikzcd}$$

2. CONVERGENCE EN PROBABILITES

2.1. Définition

Soit l'espace probabilisé $(\Omega,\mathcal{A},P)$. Une suite de variables aléatoires $X_1,\ldots,X_n$ sur $(\Omega,\mathcal{A})$ converge *en probabilités* vers une variable aléatoire X si

$$\forall\varepsilon>0,\ \lim_{n\to+\infty} P\left[\left|X_n-X\right|\geq\varepsilon\right]=0.$$

Ce type de convergence est noté

$$X_n \xrightarrow{P} X$$

<u>Exemple 3</u>

Soit (X_n) une suite de variables aléatoires continues de fonction de densité

$$\forall x\in\mathbb{R},\ f_n(x)=\frac{ne^{-nx}}{(1+e^{-nx})^2}.$$

Alors (X_n) converge en probabilité vers 0.

2.2. Loi faible des grands nombres

Soit $X_1,\ldots,X_n$ une suite de variables aléatoires pas nécessairement indépendantes telles que $E(X_i)$ existe. On dit que la suite obéit à la loi faible des grands nombres si la variable aléatoire

$$Y_n=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-E(X_i)\right)=E(\overline{X})-\frac{1}{n}\sum_{i=1}^{n}E(X_i),$$

converge en probabilité 0.

Théorème

Si $X_1,\dots,X_n$ une suite de variables aléatoires indépendantes de même loi telle que $E(X_i)=\mu$, alors la suite obéit à la loi faible des grand nombres, *i.e*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu .$$

Exemple 4

Soit X le nombre de clients se présentant à un guichet entre 10h et 12h. Alors x suit une loi de Poisson $P(\lambda)$. Comment déterminer λ à partir des observations journalières $X_1,\dots,X_n$? Les variables aléatoires X_i sont i.i.d. telles que $E(X_i)=\lambda$, donc d'après la loi des grands nombres

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to+\infty}{\overset{p}{\to}} \lambda .$$

La moyenne des observations donne ainsi une valeur approchée de λ.

3. CONVERGENCE PRESQUE SURE

3.1. Définition

Soit l'espace probabilisé $(\Omega,\mathcal{A},P)$. Une suite de variables aléatoires $X_1,\dots,X_n$ sur $(\Omega,\mathcal{A})$ converge *presque sûrement* vers une variable aléatoire X si

$$P\left[\lim_{n\to+\infty} X_n = X\right] = 1 .$$

Ce type de convergence est noté

$$X_n \xrightarrow{p.s} X$$

3.2. Loi forte des grands nombres

Soit $X_1,\dots,X_n$ une suite de variables aléatoires pas nécessairement indépendantes telles que $E(X_i)$ et $var(X_i)$ existent. On dit que la suite obéit à la loi forte des grands nombres si la variable aléatoire

$$Y_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - E(X_i)\right) = E(\bar{X}) - \frac{1}{n}\sum_{i=1}^{n} E(X_i),$$

converge presque sûrement vers 0.

Théorème

Si $X_1,\dots,X_n$ une suite de variables aléatoires indépendantes de même loi telle que $E(X_i)=\mu$ et $var(X_i)=\sigma^2$, alors la suite obéit à la loi forte des grand nombres, *i.e*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s} \mu .$$

4. CONVERGENCE EN MOYENNE QUADRATIQUE

Soit l'espace probabilisé $(\Omega,\mathcal{A},P)$ et une suite de variables aléatoires $X_1,\ldots,X_n$ sur $(\Omega,\mathcal{A})$ telles que $E(X_i)$ et $E(X_i^2)$ existent. On dit que la suite converge *en moyenne quadratique* vers une variable aléatoire X si

$$\lim_{n\to+\infty} E\left[(X_n - X)^2\right] = 0.$$

Ce type de convergence est noté

$$X_n \xrightarrow{m.q} X$$

<u>Exemple 5</u>

Soit $X_1,\ldots,X_n$ des variables aléatoires i.i.d de loi de Bernoulli B(p). Alors la suite (T_n) définie par

$$T_n = \frac{K_n}{n}, \text{ où } K_n = \sum_{i=1}^{n} X_i$$

converge en moyenne quadratique vers p.

5. DIVERS

5.1. Convergence d'une fonction de variables aléatoires

Soit g une fonction continue de $\mathbb{R}^k$ dans $\mathbb{R}^m$ et (X_n) une suite de variables aléatoires convergente vers X, alors la suite $(g(X_n))$ converge vers g(X) et ce quel que soit le mode de convergence.

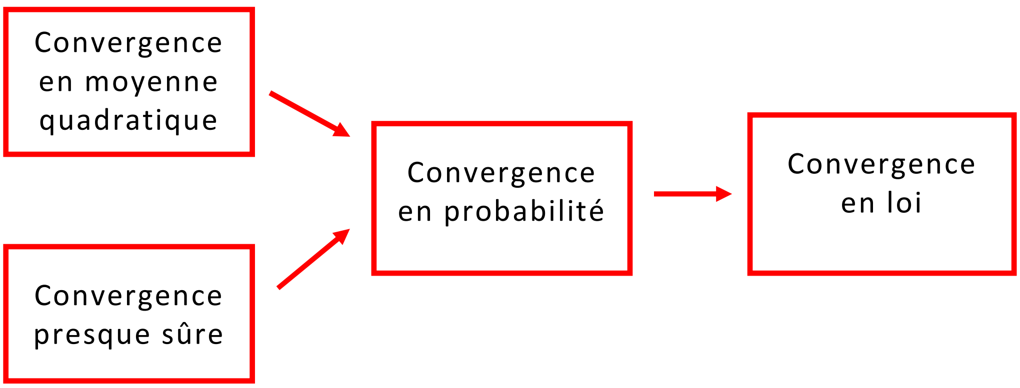
<u>Exemple</u>

Soit $X_1,\ldots,X_n$ des variables aléatoires i.i.d. de loi $N(0,\sigma^2)$ alors d'après la loi des grands nombres

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \underset{n\to+\infty}{\overset{p}{\rightarrow}} E(X_i^2) = \sigma^2$$

On peut en déduire que la suite $(\sqrt{T_n})$ converge en probabilité vers l'écart-type σ.

5.2 Liens entre les différentes convergences



EXERCICES

DEFINITIONS DES CONVERGENCES

Exercice 1

Soit (X_n) une suite de variables aléatoires i.i.d. de loi uniforme sur [0,1]. On note

$$Y_n = \max\{X_1, \ldots, X_n\}.$$

1) Montrer que $n(1-Y_n)$ converge en loi.
2) Montrer que Y_n converge en moyenne quadratique vers 1.

Exercice 2

Soit (X_n) une suite de variables aléatoires telle que

$$\lim_{n\to+\infty} E(X_n) = a \in \mathbb{R} \qquad \text{et} \qquad \lim_{n\to+\infty} V(X_n) = 0.$$

1) Montrer que X_n converge en moyenne quadratique vers *a*.
2) Montrer que X_n converge en probabilité vers *a*.

Exercice 3

Soit (Y_n) une suite de variables aléatoires définies sur {0,n} telles que

$$P(Y_n = 0) = 1 - \frac{1}{n} \qquad \text{et} \qquad P(Y_n = n) = \frac{1}{n}.$$

1) Montrer que Y_n ne converge pas en moyenne quadratique vers 0.
2) Montrer que Y_n converge en probabilité vers 0.

Exercice 4

Soit (X_n) une suite de variables aléatoires i.i.d. On note

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k^2.$$

1) Montrer que si X_n suit une loi $N(0,1)$ alors S_n converge en probabilité vers 1.
2) Si X_n suit une loi $N(\mu,\sigma^2)$, trouver la limite en probabilité de Sn.

APPLICATIONS DES THEOREMES « LIMITE »

Exercice 5

Les statistiques récentes de certains vétérinaires passionnés ont montré que parmi les beaux chats d'Angora le pourcentage d'avoir un oeil bleu et l'autre vert est de 0,01.

Soit X la variable aléatoire représentant le nombre de chats aux yeux bicolores dans un échantillon de 200 chats d'Angora choisis au hasard.

a) En formalisant le problème d'après une variable aléatoire discrète, calculer la probabilité pour qu'il y ait au moins deux chats aux yeux bicolores dans l'échantillon.
b) Une formalisation d'après une variable aléatoire discrète différente (approximation de la précédente) serait-elle possible ? Si oui, calculer de nouveau la probabilité demandée en a).

Exercice 6

Toto a constaté que souvent parmi 100 e-mails qu'il reçoit, il y en a un qui est à rejeter. Soit X_n la variable aléatoire qui représente le nombre de « mauvais e-mails » parmi n reçus. Déterminer une valeur de n à partir de laquelle Toto aura plus de 10 mauvais e-mails dans sa boite avec une probabilité de 2/3.

Exercice 7

Considérons n mesures indépendantes X_1, X_2, ..., X_n de l'énergie moyenne μ d'un faisceau de particules à l'intérieur d'un accélérateur. Supposons que chacune de ces mesures suit une loi d'espérance μ et d'écart-type 1. On utilise la moyenne $\overline{X}$ comme estimateur de μ. Déterminer le nombre minimal de mesures à affecter afin que l'erreur d'estimation soit inférieure à 0,095 avec une probabilité de 0,9.

Exercice 8

Suite à un sondage on constate qu'un habitant sur 200 utilise Internet. Soit X la variable aléatoire représentant le nombre d'habitants utilisant Internet dans une ville de 80 000 habitants comme Pau.

a) Quelle est la loi de probabilité de X ? Quelle est la probabilité pour que plus de 10 000 habitants utilisent Internet?
b) Par quelle autre loi pouvez-vous modéliser X ? Recalculer la probabilité précédente.