Chapitre 5 Notions sur la convergence en loi Théorème limite central

I Fonction caractéristique d'une loi de probabilité (Notions)

Soit X une v.a. réelle de loi de probabilité μ_X

Définition: On appelle *fonction caractéristique* de la v.a. X (ou de la loi μ_X), la fonction $\varphi_X : \mathbb{R} \to \mathbb{C}$ définie par:

$$\forall t \in \mathbb{R}, \quad \varphi_X(t) = \int_{\mathbb{R}} e^{itx} d\mu_X(x)$$

Remarques: 1) si X est discrète, alors

$$\varphi_X(t) = \sum_{x \in X(\Omega)} e^{itx} \mathbb{P}(X=x)$$

2) si X a une densité de probabilité f, on a

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

3) Si pour tout $t \in \mathbb{R}$, on considère la v.a. à valeurs complexes e^{itX}, alors on a:

$$\varphi_X(t) = \mathbb{E}(e^{itX}).$$

4). (formule de transformation linéaire): si X et Y sont des v.a. telles que $Y = aX + b$ (a et b constantes réelles), alors:

$$\varphi_Y(t) = e^{ibt} \varphi_X(at)$$

(vérification immédiate).

Exemple: (fonction caractéristique de la loi normale $N(m, \sigma^2)$):

Si la v.a. X est de loi normale de moyenne m et de variance σ^2, alors:

$$\varphi_X(t) = e^{imt} e^{-\frac{1}{2}\sigma^2 t^2}$$

D'après la remarque 4 ci-dessus, il suffit de montrer que si X est de loi normale $N(0,1)$ i.e. $m=0$ et $\sigma^2 = 1$, on a

$$\varphi_X(t) = e^{-\frac{1}{2}t^2}$$

Démontrons ceci et posons $\varphi(t) = \varphi_X(t)$. On a:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx.$$

Cette fonction est dérivable et on a

$$\varphi'(t) = \int_{-\infty}^{\infty} e^{itx} ix \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \quad (*)$$

(d'après le théorème de dérivation de Lebesgue sous le signe intégrale)

Intégrons par parties $(*)$ avec $u' = \frac{x e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$ et $v = e^{itx}$. Alors:

$$\varphi'(t)=\left[-ie^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right]_{x=-\infty}^{x=+\infty}+\int_{-\infty}^{\infty}i^2te^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx$$

le terme tout intégré est nul, donc :

$$\varphi'(t)=-t\varphi(t)\quad(\text{et } \varphi(0)=1)$$

D'où $\varphi(t)=e^{-\frac{1}{2}t^2}$ cqfd.

Théorème (f^n caractéristique d'une somme de v.a. indépendantes).
Si $X_1, X_2, \dots, X_n$ sont des v.a. indépendantes, la fonction caractéristique φ_Y de la v.a.

$$Y=X_1+\dots+X_n,$$

est donné par la formule :

$$\forall t\in\mathbb{R},\ \varphi_Y(t)=\prod_{i=1}^{n}\varphi_{X_i}(t)$$

dém^n : $\varphi_Y(t)=\mathbb{E}(e^{itY})=$

$$\mathbb{E}(e^{itX_1}e^{itX_2}\dots e^{itX_n})=$$

$$\mathbb{E}(e^{itX_1})\mathbb{E}(e^{itX_2})\dots\mathbb{E}(e^{itX_n})$$

Car les v.a. $e^{itX_1},\dots,e^{itX_n}$ sont indépendantes.

cqfd.

Théorème (d'injectivité de la f^n caractéristique)
Si X et Y sont des v.a. telles que $\varphi_X=\varphi_Y$, alors X et Y ont la même loi (i.e $\mu_X=\mu_Y$ ou de manière équivalentes X et Y ont même fonction de répartition.

dém^n : théorème admis.
Ce théorème justifie le nom de fonction caractéristique car elle caractérise la loi de probabilité.

Exercice : Si X est de loi normale $N(m_1,\sigma_1^2)$ et Y de loi normale $N(m_2,\sigma_2^2)$ et si X et Y sont indépendantes, alors $X+Y$ est de loi normale $N(m_1+m_2,\sigma_1^2+\sigma_2^2)$.

Solution : utiliser les fonctions caractéristiques et les deux théorèmes précédents.

Fonction caractéristique et moments

Théorème : 1) Si les n premiers moments d'une v.a. existent alors $\varphi_X(t)$ est n-fois dérivable et

$$\frac{\varphi_X^{(k)}(0)}{i^k}=\mathbb{E}(X^k)\quad(k=0,1,\dots,n).$$

2) Réciproquement si $\varphi_X(t)$ est n fois dérivable en $t=0$ alors X admet des moments jusqu'à l'ordre n si n est pair et jusqu'à l'ordre $n-1$ si n est impair

dém^n : 1) s'obtient immédiatement en dérivant k fois sous le signe intégrale,

$$\varphi_X(t)=\int_{-\infty}^{\infty}e^{itx}d\mu_X(x)$$

2) admis.

II La convergence en loi :

Soit (X_n) une suite de variables aléatoires et X une variable aléatoire de f.r. respectives F_{X_n} et F_X.

Définition : On dit que la suite X_n converge en loi vers X si

$$\lim_{n\to+\infty} F_{X_n}(t) = F_X(t) \text{ en tout } t\in\mathbb{R} \text{ où } F_X \text{ est continue}$$

(convergence simple)

Si X est une v.a. continue, la convergence $F_{X_n}(t) \to F_X(t)$ $(n\to+\infty)$ doit être en tout $t\in\mathbb{R}$.

Remarques : 1) Quand la fonction F_X est discontinue en un point $a\in\mathbb{R}$, on ne peut pas exiger que $\lim_{n\to\infty} F_{X_n}(a) = F(a)$. Par exemple si $X_n \equiv a_n$ (constante) et $X\equiv a$ (constante) et si $a_n \downarrow a$ (convergence décroissante), les fonctions de répartition F_{X_n} convergent bien vers F_X sauf au point a

2) Si a et $b\in\mathbb{R}$ sont des points de continuité de F_X, alors

$$\lim_{n\to\infty} \mathbb{P}(a < X_n \leq b) = \mathbb{P}(a < X \leq b)$$

(en effet $\mathbb{P}(a<X_n\leq b) = F_{X_n}(b) - F_{X_n}(a) \to F_X(b) - F_X(a)$).

3) L'intérêt pratique de la convergence en loi est qu'on peut assimiler la loi de X_n à la loi de X quand n est assez grand. Ceci est d'un très grand intérêt en Statistiques. Par exemple De-Moivre a montré en 1733 que si X_n est de loi binomiale $\mathcal{B}(n,\frac{1}{2})$, les v.a. centrées réduites

$$\tilde{X}_n = \frac{X_n - \frac{n}{2}}{\frac{1}{2}\sqrt{n}},$$

convergent en loi (quand $n\to+\infty$) vers la loi normale centrée réduite, i.e. :

$$\forall t\in\mathbb{R},\ \lim_{n\to+\infty} \mathbb{P}\left(\frac{X_n-\frac{n}{2}}{\sqrt{n}/2} \leq t\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx.$$

On en déduit par exemple que

$$\mathbb{P}\left(\left|X_n - \frac{n}{2}\right| \leq \frac{\sqrt{n}}{2}t\right) \to \int_{-t}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx \quad (n\to+\infty)$$

Ce qui signifie pratiquement que si n est grand

$$\mathbb{P}\left(\left|X_n - \frac{n}{2}\right| \leq \frac{\sqrt{n}}{2}t\right) \simeq \int_{-t}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx$$

D'où pour $a>0$

$$\mathbb{P}\left(\left|X_n - \frac{n}{2}\right| \leq a\right) \simeq \int_{-\frac{2a}{\sqrt{n}}}^{\frac{2a}{\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx$$

Ce qui montre comment X_n s'écarte de sa valeur moyenne $\frac{n}{2}$.

Remarque : En général la démonstration directe de la convergence en loi par la convergence des fonctions de répartition, n'est pas facile. Heureusement la notion de fonction caractéristique se révèle être un outil très précieux.

Soit (X_n) une suite de variables aléatoires réelles. On note φ_n la fonction caractéristique de X_n.

> **Théorème** : Supposons qu'il existe une v.a. X de fonction caractéristique φ telle que
> $$\forall t \in \mathbb{R}, \ \lim_{n\to+\infty} \varphi_n(t) = \varphi(t)$$
> Alors la suite (X_n) converge en loi vers X

dém^{n} : résultat admis

Remarque : il existe une version plus forte de ce résultat connue sous le nom de théorème de continuité de Paul Lévy et qui sera étudié dans le cours de Probabilités 2.

Exemple : Pour tout entier n supposons que X_n soit une v.a. de loi binomiale $\mathcal{B}(n, p_n)$. On suppose que $\lim_{n\to+\infty} np_n = \lambda > 0$. Alors (X_n) converge en loi vers une v.a. de Poisson de paramètre $\lambda > 0$.

démonstration : La fonction caractéristique φ_n de la v.a. X_n est donnée par :

$$\begin{aligned}\varphi_n(t) &= (p_n e^{it} + 1 - p_n)^n \\ &= (1 + p_n(e^{it}-1))^n = \left(1 + \left(\frac{\lambda}{n} + \frac{\varepsilon_n}{n}\right)(e^{it}-1)\right)^n \\ &= \left(1 + \frac{\lambda}{n}(e^{it}-1) + o\left(\frac{1}{n}\right)\right)^n \\ &\to e^{\lambda(e^{it}-1)} \text{ quand } n \to +\infty\end{aligned}$$

Mais $\varphi(t) = e^{\lambda(e^{it}-1)}$ est la fonction caractéristique de la loi de Poisson de paramètre $\lambda > 0$. D'où le résultat

Exercice : Soient X_n ($n \in \mathbb{N}$) et X des v.a. réelles. On suppose que ces variables sont toutes à valeurs dans $\mathbb{N}$. Montrer que (X_n) converge en loi vers X si et seulement si :
$$\forall k \in \mathbb{N}, \ \lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

Remarque : l'approximation de la loi binomiale par la loi de Poisson vue au chapitre 3 est en fait une convergence en loi.

III) Le théorème limite central

On désigne généralement sous le nom de théorème limite central un résultat de convergence en loi d'une suite de v.a. vers la loi normale.

Soit (X_n) une suite de v.a. indépendantes de même loi ayant un moment d'ordre 2.

On note $m = \mathbb{E}(X_n)$ et $\sigma^2 = Var(X_n)$.

Posons $S_n = X_1 + X_2 + \cdots + X_n$ et

$\widetilde{S}_n = \dfrac{S_n - \mathbb{E}(S_n)}{\sigma_{S_n}} = \dfrac{S_n - nm}{\sigma\sqrt{n}}$ la v.a. centrée réduite.

Théorème: sous les hypothèses précédentes, la suite $(\widetilde{S}_n)$ converge en loi vers une v.a. normale centrée réduite.

démon: $\widetilde{S}_n = \displaystyle\sum_{i=1}^{n} \frac{(X_i - m)}{\sigma\sqrt{n}}$.

Notons φ la fonction caractéristique commune des $X_i - m$. La fonction caractéristique de $\widetilde{S}_n$ est

$$\varphi_n(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \quad (\text{indépendance des } X_i - m)$$

Mais $\varphi(t)$ admet un développement limité

$$\varphi(t) = \varphi(0) + \frac{\varphi'(0)}{1!}t + \frac{\varphi''(0)}{2!}t^2 + o(t^2)$$
$$= 1 - \frac{\sigma^2}{2}t^2 + o(t^2)$$

D'où $\varphi_n(t) = \left(1 - \dfrac{t^2}{2n} + o\left(\dfrac{1}{n^2}\right)\right)^n \longrightarrow e^{-\frac{1}{2}t^2}$ (si $n \to +\infty$)

Or $e^{-\frac{1}{2}t^2}$ est la fonction caractéristique de la loi normale $N(0,1)$; cqfd.

Voici un autre résultat très intéressant où les v.a. X_n ne sont plus supposées de même loi :

Théorème (Lindeberg-Feller): si les (X_n) sont indépendantes et s'il existe $a > 0$ t.q.

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}(X_n \in [-a, a]) = 1$$

(on dit que les X_n sont uniformément bornées)

On suppose que X_i a une variance σ_i^2 et que $\displaystyle\sum_{i=1}^{n} \sigma_i^2 \longrightarrow +\infty$ si $n \to +\infty$. Alors:

la v.a. centrée réduite $\widetilde{S}_n = \dfrac{S_n - \mathbb{E}(S_n)}{\sigma_{S_n}}$

(avec $S_n = X_1 + \cdots + X_n$), converge en loi vers la loi normale $N(0,1)$ quand $n \to +\infty$.

démon: admis

Exemple (d'application): La veille d'un week-end 200 personnes vont s'approvisionner en argent liquide à un distributeur de billets. Le distributeur dispose de 42000 euros. La banque sait par expérience que les sommes prélevées sont des v.a. indépendantes de moyenne 200 euros et d'écart type 50 euros. Quelle est la probabilité que tous les clients du distributeur soient servis ?

Solution: Soit X la somme totale qui sera prélevée. On a $\mathbb{E}(X) = 40000$ et $\sigma^2(X) =$

$200\cdot 25\cdot 10^2 = 50\cdot 10^4 \Rightarrow \sigma(X) = \sqrt{50\cdot 10^4} = 100\sqrt{50} \simeq 707.$

Supposons que les conditions de l'approximation normale soient remplies (TLC si les sommes prélevées par chaque ont la même loi ou ce qui est plus réaliste théorème de Lindeberg-Feller). Alors:

$$\mathbb{P}(X \le 42000) = \mathbb{P}\left(\frac{X-40000}{707} \le 2,82\right)$$

$$\cong \int_{-\infty}^{2,82} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}dx = 0,9976.$$

__Remarque__ : Si les v.a. (X_n) sont toutes de même loi de Bernoulli de paramètre $p \in]0,1[$ et indépendantes, alors $S_n = X_1 + \dots + X_n$ est de loi binomiale $\mathcal{B}(n,p)$. Si $n \to +\infty$, la v.a. centrée réduite

$$\tilde{S}_n = \frac{S_n - np}{\sqrt{np(1-p)}}$$ converge en loi vers la loi normale $N(0,1)$

Thm de De-Moivre-Laplace (1803-04)

> On considère dans la pratique que l'approximation normale est bonne si:
>
> $n \geq 100$ et si $np \geq 50$ et $n(1-p) \geq 50$.

Si la valeur de np ou $n(1-p)$ est trop petite il faut envisager l'approximation par la loi de Poisson.

IV Application du théorème limite central à l'estimation d'une proportion.

En écrivant la v.a. binomiale centrée réduite $\tilde{S}_n$ sous la forme:

$$\tilde{S}_n = \frac{\frac{S_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}},$$

le théorème limite central implique que

$$\forall a<b \in \mathbb{R}, \quad \lim_{n\to\infty} \mathbb{P}\left(a\sqrt{\frac{p(1-p)}{n}} \le \frac{S_n}{n} - p \le b\sqrt{\frac{p(1-p)}{n}}\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx$$

Comme la loi normale est tabulée, ceci permet le calcul des probabilités des écarts $\frac{S_n}{n} - p$ entre la fréquence de succès $\frac{S_n}{n}$ et la probabilité de succès p. De plus en écrivant

$$\mathbb{P}\left(a\sqrt{\frac{p(1-p)}{n}} \le \frac{S_n}{n} - p \le b\sqrt{\frac{p(1-p)}{n}}\right) =$$

$$\mathbb{P}\left(\frac{S_n}{n} - b\sqrt{\frac{p(1-p)}{n}} \le p \le \frac{S_n}{n} - a\sqrt{\frac{p(1-p)}{n}}\right),$$

on voit que l'intervalle aléatoire

$\left[\frac{S_n}{n} - b\sqrt{\frac{p(1-p)}{n}}, \frac{S_n}{n} - a\sqrt{\frac{p(1-p)}{n}}\right]$ est un intervalle de confiance de p au niveau de confiance

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx.$$

Si $a = -b$, l'intervalle de confiance est dit symétrique. L'inconvénient majeur de cet intervalle est qu'il est encore exprimé en fonctions de p (le paramètre à estimer). On a donc recours à des intervalles de confiance approchés. Par exemple, on peut remplacer p par $\frac{S_n}{n}$ ou le produit $p(1-p)$ par $\frac{1}{4}$. Dans le cas symétrique, on peut toutefois calculer un intervalle de confiance exact : écrivons $\frac{S_n}{n} = p_0$, on a

$$\mathbb{P}\left(|p_0 - p| \leq b\sqrt{\frac{p(1-p)}{n}}\right) \simeq \phi(b) = \int_{-b}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

On peut résoudre l'inégalité $|p_0 - p| \leq b\sqrt{\frac{p(1-p)}{n}}$:

$$(p_0 - p)^2 \leq b^2 \frac{p(1-p)}{n}$$

$$\Longleftrightarrow (n + b^2)p^2 - (2p_0 n + b^2)p + np_0^2 \leq 0 \cdot \circledast$$

Le discriminant de ce trinôme en p est

$$\Delta = b^4 + 4nb^2 p_0(1-p_0) \geq 0 .$$

L'inégalité $\circledast$ est donc vérifiée entre les racines du trinôme :

$$\begin{matrix} \alpha = \alpha(p_0, n, b) \\ \beta = \beta(p_0, n, b) \end{matrix} = \frac{2p_0 n + b^2}{2(n+b^2)} \mp b \frac{\sqrt{b^2 + 4np_0(1-p_0)}}{2(n+b^2)}$$

$[\alpha, \beta]$ est donc l'intervalle de confiance exact de p au niveau de confiance $\phi(b)$. On notera que si n est grand, on peut négliger b^2 face à n et à np_0. Alors l'intervalle prend la forme simple :

$$p_0 - b\sqrt{\frac{p_0(1-p_0)}{n}} \leq p \leq p_0 + b\sqrt{\frac{p_0(1-p_0)}{n}}$$