P. Maurer
ENS Rennes

Leçon 261. Loi d'une variable aléatoire : caractérisation, exemples, applications.

Devs :

- Théorème de Radon-Nikodym
- Théorème central limite

Références :

1. Brianes-Pagès, Théorie de l'intégration
2. Foata-Fuchs, Calcul des probabilités
3. Barbe-Ledoux, Probabilités
4. Ouvrard, Probabilités 2
5. Candelpergher, Théorie des probabilités
6. Cadre-Vial, Statistique mathématique
7. Zuily-Quéffélec, Analyse pour l'agrégation

Dans ce qui suit, on se donne $(X, \mathcal{A}, \mathbb{P})$ un espace probabilisé, un ensemble E et une tribu $\mathcal{B}$ sur E (par exemple, $E=\mathbb{R}^d$ et $\mathcal{B}=\mathcal{B}(\mathbb{R}^d)$).

1 Généralités

1.1 Loi d'une variable aléatoire : définition et propriétés

Définition 1. *Soit $X\colon(\Omega,\mathcal{A},\mathbb{P})\to(E,\mathcal{B})$ une variable aléatoire. On appelle loi de X sous $\mathbb{P}$ la mesure de probabilité définie par $\mathbb{P}_X(A)=\mathbb{P}(X^{-1}(A))=\mathbb{P}(X\in A)$.*

Exemple 2. Si $(E,\mathcal{B})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$, on dit que X est une variable aléatoire réelle.

Exemple 3. On lance une pièce non truquée, n fois, de manière indépendante. On a $E=\{\text{Pile},\text{Face}\}$, et si X est la variable aléatoire qui compte le nombre de Faces, la loi de X est déterminée par

$$\mathbb{P}(X=k)=\frac{1}{2^n}\binom{n}{k}.$$

On dit que X suit une loi binomiale de paramètres $(1/2,n)$ et on note $X\sim\mathcal{B}(n,1/2)$.

On considère désormais que $(E,\mathcal{B})=(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ et on se donne $X\colon(\Omega,\mathcal{A},\mathbb{P})\to(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ une variable aléatoire sur $\mathbb{R}^d$.

Définition 4. *Si $X\in L^1(\Omega,\mathcal{A},\mathbb{P})$, on appelle espérance de X l'intégrale*

$$\mathbb{E}[X]=\int_\Omega X(\omega)\,d\mathbb{P}(\omega).$$

On dit que X est centrée si $\mathbb{E}[X]=0$.

Théorème 5. *(De transfert).*
Soit $\phi\colon\mathbb{R}^d\to\mathbb{R}$ une fonction borélienne. Si ϕ est à valeurs positives,

$$\mathbb{E}[\phi(X)]=\int_\Omega(\phi\circ X)(\omega)d\mathbb{P}(\omega)=\int_{\mathbb{R}^d}\phi(x)\,d\mathbb{P}_X(x).$$

Si ϕ est à valeurs quelconques,

$$\phi(X)\in L^1(\Omega,\mathcal{A},\mathbb{P})\Longleftrightarrow\phi\in L^1(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d),\mathbb{P}_X),$$

et dans ce cas, l'égalité précédente a lieu.

Exemple 6. Si $A\in\mathcal{B}(\mathbb{R}^d)$, la fonction $\mathbf{1}_A$ est mesurable, et le théorème de transfert montre que $\mathbb{E}[\mathbf{1}_A(X)]=\mathbb{P}(X\in A)=\mathbb{P}_X(A)$.

Théorème 7. *Notons $\mathcal{C}_K^+(\mathbb{R}^d)$ l'ensemble des fonctions continues, positives, et à support compact sur $\mathbb{R}^d$. Soit μ,ν deux mesures positives sur $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ finies sur tout compact. On suppose que*

$$\forall f\in\mathcal{C}_K^+(\mathbb{R}^d)\quad\int_{\mathbb{R}^d}f d\mu=\int_{\mathbb{R}^d}f d\nu.$$

Alors $\mu=\nu$.

Corollaire 8. *La loi de X est entièrement déterminée par la famille des $\mathbb{E}[f(X)]$ où f parcourt $\mathcal{C}_K^+(\mathbb{R}^d)$. Ceci donne une méthode efficace pour déterminer la loi d'une variable aléatoire.*

Exemple 9. Si $X\sim\mathcal{N}(0,1)$, on montre que $X^2\sim\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$ avec le corollaire 8 et le théorème de transfert.

1.2 Fonction de répartition et fonction quantile

On suppose que X est une variable aléatoire réelle (i.e que $(E,\mathcal{B})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$).

Définition 10. *On appelle fonction de répartition de X, ou de sa loi $\mathbb{P}_X$, et on note F_X, la fonction définie sur $\mathbb{R}$ par*

$$\forall t\in\mathbb{R}\quad F_X(t)=\mathbb{P}_X(]-\infty,t])=\mathbb{P}(X\leq t).$$

Proposition 11. *Une fonction de répartition F vérifie les propriétés suivantes :*

- *On a $0 \leq F \leq 1$,*
- *F est croissante, continue à droite, avec une limite à gauche en tout point,*
- *$\lim_{t \to -\infty} F(t) = 0$ et $\lim_{t \to +\infty} F(t) = 1$.*

Proposition 12. *Réciproquement, toute fonction F vérifiant les trois propriétés de la proposition 11 est la fonction de répartition d'une variable aléatoire réelle.*

Proposition 13. *La fonction de répartition caractérise la loi : si X et Y sont deux variables aléatoires réelles, alors $F_X = F_Y \Longleftrightarrow \mathbb{P}_X = \mathbb{P}_Y$.*

Proposition 14. *Une fonction de réparition admet au plus un nombre dénombrable de discontinuité (cela est liée à son caractère croissant).*

Exemple 15. La fonction de répartition de la loi binomiale $\mathcal{B}(n,p)$ est donnée par

$$F(t) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^k \mathbf{1}_{[k,+\infty[}(t).$$

Exemple 16. La fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$ est donnée par

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

Définition 17. *Soit F une fonction de répartition. On appelle fonction quantile la fonction $F^{\leftarrow}$ définie par $F^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : F(x) > u\}$ pour tout $u \in]0,1[$.*

Théorème 18. *Si $U \sim \mathcal{U}(]0,1[)$, alors $F^{\leftarrow}(U)$ a pour fonction de répartition F.*

1.3 Loi d'un d-uplet de variables aléatoires et indépendance

Définition 19. *Soit $(X_1,\ldots,X_d)$ un d-uplet de variables aléatoires réelles. On appelle loi conjointe la loi $\mathbb{P}_{(X_1,\ldots,X_d)}$, et lois marginales les lois $\mathbb{P}_{X_1},\ldots,\mathbb{P}_{X_d}$ de chaque composante du d-uplet.*

Remarque 20. La donnée de la loi conjointe permet de retrouver les lois marginales, mais la réciproque est fausse en général. On a cependant le théorème suivant.

Théorème 21. *Soit $d \geq 1$ un entier et $(X_1,\ldots,X_d)$ une famille finie de variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, \mathbb{P})$. La loi $\mathbb{P}_{(X_1,\ldots,X_d)}$ du vecteur aléatoire sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est égale au produit des lois marginales $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}$. Réciproquement, si la loi du vecteur est égale au produit des marginales, alors les variables sont indépendantes.*

Théorème 22. *Une famille quelconque de variables aléatoires X_i, $i \in I$ sur $(\Omega, \mathcal{A}, \mathbb{P})$ est indépendante si et seulement si pour toute famille finie $J \subset I$, et toute famille de fonctions boréliennes ϕ_i, $i \in J$ telles que $\phi_i(X_i) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ pour $i \in J$, on a*

$$\mathbb{E}\left[\prod_{i \in J} \phi_i(X_i)\right] = \prod_{j \in J} \mathbb{E}[\phi_i(X_i)].$$

Exemple 23. Soit $(X_1,\ldots,X_n)$ un n-uplet de variables aléatoires discrètes de loi $\mathcal{B}(p)$. Alors $X_1 + \cdots + X_n \sim \mathcal{B}(n,p)$.

Exemple 24. Soit (X_1, X_2) est un couple de variables aléatoires indépendantes de lois marginales respectives $\Gamma(p_1, \lambda_1)$ et $\Gamma(p_2, \lambda_2)$, et $\mu > 0$. Alors

- $X_1 + X_2$ a pour loi $\Gamma(p_1 + p_2, \lambda)$.
- μX_1 a pour loi $\Gamma\left(p_1, \frac{\lambda}{\mu}\right)$.

Exemple 25. Soit $(X_1,\ldots,X_n)$ des variables aléatoires indépendantes, de loi exponentielle de paramètre $\lambda > 0$. Alors $X_1 + \cdots + X_n$ a pour loi $\Gamma(n, \lambda)$.

2 Loi de certaines classes de variables aléatoires

2.1 Loi des variables aléatoires discrètes

Définition 26. *On dit qu'une variable aléatoire est discrète si $X(\Omega)$ est fini ou dénombrable.*

Proposition 27. *La loi d'une variable aléatoire discrète est de la forme $\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$ avec $I \subset \mathbb{N}$, où δ_{x_i} désigne la masse de Dirac en x_i.*

Définition 28. *On définit la fonction (ou série) génératrice de X comme la série entière $G_X(t) := \sum_{k \geq 0} \mathbb{P}(X = k) t^k$.*

Proposition 29. *Le rayon de convergence R de la série entière $G_X(t)$ est supérieur ou égal à 1. La fonction $t \mapsto \mathbb{E}[t^X]$ est bien définie et à valeurs finies pour $t \in]-R, R[\cup \{-1, 1\}$, et vérifie alors $\mathbb{E}[t^X] = G_X(t)$.*

Proposition 30.

1. *La fonction génératrice de X sur $[-1,1]$ détermine entièrement la loi de X.*
2. *Si X et Y sont indépendantes, alors $G_{X+Y} = G_X G_Y$.*

Théorème 31. *La variable aléatoire X est d'espérance finie si et seulement si G_X est dérivable à gauche au point 1, et dans ce cas, on a $\mathbb{E}[X] = G'_X(1)$.*

Exemple 32. Si $X \sim \mathcal{G}(\alpha)$, $G_X(t) = (1-\alpha)\frac{s}{1-\alpha s}$, et on en déduit $\mathbb{E}[X] = \frac{1}{1-\alpha}$.

Exemple 33. On effectue des jeux de pile ou face successifs indépendants, avec probabilité $p \in]0,1[$ de faire face, et $1-p$ de faire pile. On s'arrête lorsque l'on obtient deux faces consécutives. En moyenne, il faut $\frac{1+p}{p^2}$ répétitions.

2.2 Loi des variables aléatoires absolument continues

Définition 34. *Soit μ et ν deux mesures sur $(X, \mathcal{A})$. On dit que ν est absolument continue par rapport à μ, et on note $\nu \ll \mu$, si pour tout $A \in \mathcal{A}$, $\mu(A) = 0 \Longrightarrow \nu(A) = 0$.*

Développement 1 :

Théorème 35. *(de Radon-Nikodym).*

Soit $(X, \mathcal{A})$ un espace mesurable et μ, ν deux mesures positives σ-finies sur $(X, \mathcal{A})$. Il y a équivalence entre :

1. *$\nu \ll \mu$,*
2. *$\exists f : (X, \mathcal{A}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ μ-intégrable telle que $\forall A \in \mathcal{A} \quad \nu(A) = \int_A f d\mu$.*

En outre, la fonction f est unique (à une égalité μ-presque partout près).
On note $f = \frac{d\nu}{d\mu}$ et on dit que f est la dérivée de Radon-Nikodym, ou la densité de ν par rapport à μ.

Application 36. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et X une variable aléatoire réelle sur Ω. On dit que la loi P_X de X admet une densité par rapport à la mesure de Lebesgue λ sur $\mathbb{R}$ si on a $P_X \ll \lambda$.

Dans ce cas, la fonction de répartition F_X de X s'écrit

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\, d\lambda(x),$$

où $f_X = \frac{dP_X}{d\lambda}$ est la dérivée de Radon-Nikodym de P_X par rapport à λ.

Remarque 37. (Admise). Si X admet une densité par rapport à λ, F_X est continue sur $\mathbb{R}$, et λ-presque partout dérivable. On dit que F est une fonction absolument continue.

Dans le cas où F est au moins $\mathcal{C}^1$ par moreceaux, on obtient la densité f_X de X comme la dérivée de F sur les intervalles où elle est dérivable.

Exemple 38. Une variable aléatoire sur Ω à valeurs dans $[0,1]$ est dite de loi $B(r,s)$ (loi bêta de paramètres $r > 0$ et $s > 0$) si elle est absolument continue par rapport à la mesure de Lebesgue et admet pour densité :

$$f(x) = \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1}\mathbf{1}_{[0,1]}(x).$$

Exemple 39. Soit (X, Y) un couple de variables aléatoires réelles, de loi absolument continue de densité $x \mapsto f(x)\, g(x)$ par rapport à la mesure de Lebesgue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.

On suppose de plus que $f \in L^1(\mathbb{R})$ et que $\int_{\mathbb{R}} f(x)\, dx = 1$.

Alors X et Y sont indépendantes, absolument continues par rapport à la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et de densité respective f et g.

2.3 Fonction caractéristique

Définition 40. *Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle fonction caractéristique de X ou de la loi de X, ou transformée de Fourier, et on note φ_X, la fonction à valeurs complexe définie par $\varphi_X(t) = \mathbb{E}[e^{i\langle t, X\rangle}] = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\, d\mathbb{P}_X(x)$.*

Proposition 41. *Si la loi de X a une densité f par rapport à la mesure de Lebesgue, alors*

$$\varphi_X(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} f(x)\, dx.$$

Exemple 42. Soit X une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ suivant une loi normale centrée réduite. Alors la fonction caractéristique φ_X de X vérifie

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = e^{-\frac{t^2}{2}}.$$

Théorème 43. *Soit X et Y deux vecteurs aléatoires de lois $\mathbb{P}_X$ et $\mathbb{P}_Y$ telles que $\varphi_X = \varphi_Y$. Alors $\mathbb{P}_X = \mathbb{P}_Y$: la fonction caractéristique caractérise la loi.*

Proposition 44. *Soit X une variable aléatoire réelle.*

- *Si $\mathbb{E}[|X|^n] < \infty$, alors φ est n-fois dérivable et $\varphi^{(k)}(t) = i^k \int_{\mathbb{R}} x^k e^{itx} d\mathbb{P}_X(x) = i^k \mathbb{E}[X^k e^{itX}]$.*
  En particulier, $\varphi^{(k)}(0) = i^k\, \mathbb{E}[X^k]$.
- *Réciproquement, si n est pair et si φ est n-fois dérivable en zéro, alors X admet tout moment d'ordre plus petit ou égal à n.*

Théorème 45. *(Des moments).*

Soit X, Y des variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un intervalle borné $[a,b]$. On suppose que $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ pour tout $k \in \mathbb{N}$. Alors X et Y ont la même loi.

3 Convergence en loi et applications

3.1 Théorèmes de convergence

Théorème 46. *(Théorème et définition, Paul Levy).*
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires sur $(\Omega,\mathcal{A},\mathbb{P})$.
Les propositions suivantes sont équivalentes :

i. *On a* $\lim\limits_{n\to+\infty} F_{X_n}(t)=F_X(t)$ *en tout point t où F_X est continue.*

ii. *On a* $\lim\limits_{n\to+\infty}\int_\Omega \phi(X_n)\,d\mathbb{P}=\int_\Omega\phi(X)\,d\mathbb{P}$ *pour toute fonction $\phi\colon\mathbb{R}\to\mathbb{R}$ continue bornée.*

iii. *On a* $\lim\limits_{n\to+\infty}\varphi_{X_n}(t)=\varphi_X(t)$ *pour tout $t\in\mathbb{R}$.*

On dit alors que la suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X, et on note $X_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} X$.

Exemple 47. Soit $X\sim\mathcal{N}(0,1)$ et $X_n=(-1)^nX$. Alors $(X_n)_{n\geq0}$ converge en loi vers X, mais ne converge ni presque sûrement ni en probabilités vers X.

Proposition 48. *Si $(X_n)_{n\geq0}$ converge en loi vers une constante $c\in\mathbb{R}$, alors $(X_n)_{n\geq0}$ converge en probabilité vers c.*

Théorème 49. *(Lemme de Slutsky)*
Soit $(X_n)_{n\geq0}$ et $(Y_n)_{n\geq0}$ des suites de variables aléatoires réelles sur $(\Omega,\mathcal{F},\mathbb{P})$. On suppose que $(X_n)_{n\geq0}$ converge en loi vers X et que $(Y_n)_{n\geq0}$ converge en probabilité vers une constante $c\in\mathbb{R}$. Alors le couple (X_n,Y_n) converge en loi vers le couple (X,c).

En particulier, on a $X_n+Y_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} X+c$ *et* $X_nY_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} cX$.

Exemple 50. On suppose que X_n et X sont à valeurs entières. Alors $X_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} X$ si et seulement si on a $\lim\limits_{n\to+\infty}\mathbb{P}(X_n=k)=\mathbb{P}(X=k)$ pour tout $k\in\mathbb{N}$.

Développement 2 :

Théorème 51. *(Théorème central limite)*
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, définies sur $(\Omega,\mathcal{F},\mathbb{P})$. On suppose que $X_0\in L^2(\Omega,\mathcal{A},\mathbb{P})$ et on note $\sigma^2=\mathrm{Var}(X_0)$ et $\mu=\mathbb{E}[X_0]$. On suppose de plus que $\sigma^2>0$.
Par hypothèse, on a donc $X_n\in L^2(\Omega,\mathcal{F},\mathbb{P})$, $\mathrm{Var}(X_n)=\sigma^2$ et $\mathbb{E}[X_n]=\mu$pour tout $n\in\mathbb{N}$.
On note également $S_n=X_1+\cdots+X_n$ pour tout $n\geq1$.
On a alors la convergence en loi

$$\frac{S_n-n\mu}{\sqrt{n}} \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,\sigma^2).$$

3.2 Applications à l'inférence statistique

Proposition 52. *On considère le modèle statistique $(\{0,1\}^n,\{\mathcal{B}(\theta)^{\otimes n}\}_{\theta\in]0,1[})$ du jeu de pile ou face. Alors l'estimateur des moments $\overline{X_n}$ construit avec le n-échantillon $(X_1,\ldots,X_n)$ est asymptotiquement normal, de vitesse $\sqrt{n}$, car le théorème central limite donne :*

$$\sqrt{n}\,(\overline{X_n}-\theta) \underset{n\to+\infty}{\to} \mathcal{N}(0,\theta(1-\theta)).$$

Application 53. (Intervalles de confiance asymptotiques)
Soit $(X_i)_{i\in\mathbb{N}^*}$ des réalisations i.i.d de $\mathcal{B}(p)$ pour $p\in[0,1]$ inconnu. Le théorème central limite donne un intervalle de confiance asymptotique de niveau α pour p en fonction de la moyenne empirique $\widehat{p_n}=\frac{1}{n}\sum_{k=1}^n X_i$: $\mathrm{IC}_\alpha(p)=\left[\widehat{p_n}\pm\frac{q_{1-\alpha/2}}{2}\frac{1}{\sqrt{n}}\right]$, où $q_{1-\alpha/2}=F^{\leftarrow}(1-\alpha/2)$ est le quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$.

Théorème 54. *(δ-méthode).*
Soit $(X_n)_{n\geq1}$ une suite de variables aléatoires d'espérence θ et de variance σ^2.
Soit $\overline{X_n}$ un estimateur de θ.
On suppose que

$$\sqrt{n}\,(\overline{X_n}-\theta) \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,\sigma^2).$$

Alors pour toute fonction $g\colon\mathbb{R}\to\mathbb{R}$ de classe $\mathcal{C}^1$ telle que $g'(\theta)\neq0$ on a

$$\sqrt{n}\,(g(\overline{X_n})-g(\theta)) \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,\sigma^2(g'(\theta))^2).$$

Application 55. On considère $T_1,\ldots,T_n$ des réalisations i.i.d de loi $\mathcal{E}(\lambda)$, où $\lambda\in]0,+\infty[$ est un paramètre inconnu à estimer. L'estimateur du maximum de vraissemblance est donné par

$$\widehat{\lambda_n}=\frac{n}{\sum_{i=1}^n T_i}.$$

Il est fortement consistant, et la δ-méthode montre qu'il est asymptotiquement normal. On en déduit que $\mathrm{IC}_\alpha(\lambda)=\left[\frac{\widehat{\lambda_n}}{1+\frac{q_{1-\alpha/2}}{\sqrt{n}}},\frac{\widehat{\lambda_n}}{1+\frac{q_{\alpha/2}}{\sqrt{n}}}\right]$ est un intervalle de confiance asymptotique pour λ de niveau α.

3.3 Méthode de Monte-Carlo

On souhaite évaluer $\int_{\mathbb{R}^d}f(x)\,g(x)\,dx$, où f est une densité sur $\mathbb{R}^d$ et $g\in L^1(f)$. Si Y est une variable aléatoire de densité f, $I=\mathbb{E}[g(Y)]$. Si $(Y_n)_{n\in\mathbb{N}}$ est un échantillon de Y, alors $\frac{1}{n}\sum_{i=1}^n g(Y_i)\underset{\mathbb{P}-\mathrm{ps}}{\to} I$ d'après la loi des grands nombres.

Cette méthode est moins efficace que les méthodes de quadrature en dimension 1, mais son efficacité ne dépend pas de la dimension de l'espace (elle est en $O(1/\sqrt{n})$), et elle est donc largement préférable en grande dimension.