CHAPITRE 26

Rappels de théorie de l'intégration et des probabilités

26.1 Résultats de théorie de l'intégration

26.1.1 Théorème de dérivation des intégrales à paramètre

On en énonce une version lisible et qui est suffisante pour la plupart des applications. Les hypothèses peuvent cependant être affaiblies, voir par exemple [LG06, Théorème 2.3.2].

Théorème 26.1 *Soit $(\mathcal{X}, \mathcal{F}, \mu)$ un espace mesuré et soit I un intervalle ouvert de $\mathbb{R}$. On considère une application $f : I \times \mathcal{X} \to \mathbb{R}$ vérifiant les conditions suivantes :*

(1) pour tout $t \in \mathbb{R}$, l'application $f(t, \cdot) : \mathcal{X} \to \mathbb{R}$ appartient à $\mathbb{L}^1(\mu)$;

(2) pour μ–presque tout x, l'application $f(\cdot, x) : I \to R$ est dérivable ;

(3) il existe une fonction $g \in \mathbb{L}^1(\mu)$ telle que pour μ–presque tout x,

$$\forall t \in I, \qquad \left|\frac{\partial f}{\partial t}(t, x)\right| \leqslant g(x) .$$

Alors l'application

$$F : t \in I \longmapsto \int_{\mathcal{X}} f(t, x)\, \mathrm{d}\mu(x)$$

est bien définie et est dérivable sur I, de dérivée en $t \in I$ égale à

$$F'(t) = \int_{\mathcal{X}} \frac{\partial f}{\partial t}(t, x)\, \mathrm{d}\mu(x) .$$

26.1.2 Le théorème fondamental du calcul différentiel et intégral

Ce qui suit est essentiellement tiré de [Rud98, chapitre 7] et s'intéresse aux liens réciproques entre intégration et dérivation. On redonne ici brièvement la progression logique et l'enchaînement des résultats. Dans tout ce qui suit, on note λ la mesure de Lebesgue.

La première partie du théorème fondamental indique qu'en dérivant une primitive, on obtient l'intégrande. Sa preuve repose sur la considération des points de Lebesgue de la fonction, dont on a rappelé la définition au paragraphe 24.1.5.

Théorème 26.2 *Soit $f \in \mathbb{L}^1(\lambda)$. On note F la primitive définie pour tout $x \in \mathbb{R}$ par*

$$F(x) = \int_{-\infty}^{x} f \, \mathrm{d}\lambda \, .$$

Alors l'ensemble $\mathcal{D}$ des points où F est dérivable est le complémentaire d'un ensemble de mesure nulle et l'on a $F'(x) = f(x)$ pour $x \in \mathcal{D}$.

La seconde partie du théorème fondamental généralise ce fait et indique une condition nécessaire et suffisante pour qu'une fonction soit l'intégrale de sa dérivée.

Définition 26.3 Soit $I = [a, b]$ un intervalle compact de $\mathbb{R}$. Une fonction $f : I \to \mathbb{R}$ est dite *absolument continue* sur I si pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout entier $n \geqslant 1$ et toute suite finie $\big([a_k, b_k]\big)_{1 \leqslant k \leqslant n}$ de sous-intervalles de I d'intérieurs disjoints,

$$\sum_{k=1}^{n} |b_k - a_k| < \delta \qquad \text{implique} \qquad \sum_{k=1}^{n} \big| f(a_k) - f(b_k) \big| < \varepsilon \, .$$

Une fonction $\mathbb{R} \to \mathbb{R}$ est dite absolument continue si elle est absolument continue sur tout sous-intervalle compact de $\mathbb{R}$.

En particulier, en prenant $n = 1$, on note qu'une fonction absolument continue est continue. Le paragraphe 24.2.1 explique les liens entre absolue continuité au sens des fonctions réelles (énoncée ci-dessus) et absolue continuité $\nu \ll \lambda$ d'une loi ν sur $\mathbb{R}$ par rapport à la mesure de Lebesgue : le théorème suivant indique qu'une condition nécessaire et suffisante pour $\nu \ll \lambda$ est l'absolue continuité (au sens des fonctions réelles) de la fonction de répartition F_ν de ν.

Théorème 26.4 *Soit $I = [a, b]$ un intervalle compact de $\mathbb{R}$. Une fonction $f : I \to \mathbb{R}$ est absolument continue sur I si et seulement si f est dérivable en presque tout point de I, de dérivée (ainsi définie presque partout) notée f' vérifiant en outre $f' \in \mathbb{L}^1(\lambda)$ et*

$$\forall x \in I, \qquad f(x) - f(a) = \int_a^x f' \, \mathrm{d}\lambda \, .$$

Pour la culture, voici, dans la lignée de ce théorème, un autre résultat utile (et dont la preuve est fort différente). On note qu'on requiert ici une dérivabilité en tout point et non pas seulement en presque tout point.

Théorème 26.5 *Soient $I = [a, b]$ un intervalle compact de $\mathbb{R}$ et $f : I \to \mathbb{R}$ une fonction dérivable en* tout *point de I, de dérivée vérifiant $f' \in \mathbb{L}^1(\lambda)$. Alors*

$$\forall x \in I, \qquad f(x) - f(a) = \int_a^x f' \, \mathrm{d}\lambda \, .$$

26.1.3 Le théorème de changement de variables

On fixe ici deux ouverts U et D de $\mathbb{R}^d$, où $d \in \mathbb{N}^*$. On rappelle qu'une application $\varphi : U \to D$ est un difféomorphisme de classe C^1 si φ est bijective et que φ et φ^{-1} sont de classe C^1 respectivement sur U et D. On sait alors que la différentielle $D\varphi(u)$ de φ est inversible en tout point $u \in U$. On appelle jacobien de φ en u le déterminant $\det D\varphi(u)$.

Théorème 26.6 *Soit $\varphi : U \to D$ un difféomorphisme de classe C^1. Alors, pour toute fonction borélienne $f : D \to \mathbb{R}_+$,*

$$\int_D f \, \mathrm{d}\lambda = \int_U f\big(\varphi(u)\big) \big|\det D\varphi(u)\big| \, \mathrm{d}u \, .$$

La version qui précède est proposée dans tous les ouvrages classiques de théorie de l'intégration, voir par exemple [LG06]. On invite le lecteur à consulter également une version légèrement plus générale proposée par [Rud98, Théorème 7.26].

26.2 Résultats de la théorie des probabilités

26.2.1 Convergences

Définitions équivalentes de la convergence en loi $X_n \rightsquigarrow X$

Le lemme suivant donne plusieurs définitions équivalentes de la convergence en loi. On pourra en trouver une preuve dans [vdV98, chapitre 2]. On l'énonce pour un vecteur aléatoire X défini sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{R}^d$, pour $d \in \mathbb{N}^*$, et des vecteurs aléatoires X_n définis chacun sur un espace $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ et à valeurs dans $\mathbb{R}^d$, pour $n \geqslant 1$. On note les espérances correspondantes $\mathbb{E}$ et $\mathbb{E}_n$. Lorsque x et y sont deux vecteurs, la notation $x \leqslant y$ signifie que l'inégalité $x_j \leqslant y_j$ vaut pour chacune des composantes de x et y.

Lemme 26.7 (Portmanteau) Les assertions suivantes sont équivalentes et définissent la convergence en loi $X_n \rightsquigarrow X$ d'une suite de vecteurs aléatoires (X_n) vers X :

(1) en tout point de continuité de $x \mapsto \mathbb{P}\{X \leqslant x\}$, on a la convergence simple

$$\mathbb{P}_n\{X_n \leqslant x\} \longrightarrow \mathbb{P}\{X \leqslant x\} \, ;$$

(2) pour toute fonction continue bornée f,

$$\mathbb{E}_n\big[f(X_n)\big] \longrightarrow \mathbb{E}\big[f(X)\big] \, ;$$

(3) pour toute fonction lipschitzienne bornée f,

$$\mathbb{E}_n\big[f(X_n)\big] \longrightarrow \mathbb{E}\big[f(X)\big] \, ;$$

(4) pour toute fonction f continue et positive,

$$\liminf \ \mathbb{E}_n\big[f(X_n)\big] \geqslant \mathbb{E}\big[f(X)\big]\ ;$$

(5) pour tout ensemble ouvert U,

$$\liminf \ \mathbb{P}_n\{X_n \in U\} \geqslant \mathbb{P}\{X \in U\}\ ;$$

(6) pour tout ensemble fermé F,

$$\limsup \ \mathbb{P}_n\{X_n \in F\} \leqslant \mathbb{P}\{X \in F\}\ ;$$

(7) pour tout ensemble borélien B, de frontière $\delta B = \overline{B} \setminus \overset{\circ}{B}$ telle que $\mathbb{P}\{X \in \delta B\} = 0$,

$$\mathbb{P}_n\{X_n \in B\} \longrightarrow \mathbb{P}\{X \in B\}\ .$$

Ces définitions équivalentes de la convergence en loi montrent en particulier que cette dernière est stable par passage aux fonctions continues.

Lemme de Slutzky

Il n'est en général pas vrai que les convergences marginales $X_n \rightsquigarrow X$ et $Y_n \rightsquigarrow Y$ entraînent la convergence jointe $(X_n, Y_n) \rightsquigarrow (X, Y)$. C'est vrai lorsque pour tout n, les variables aléatoires X_n et Y_n sont indépendantes et que X et Y le sont également. C'est aussi vrai lorsque $X_n \to X$ et $Y_n \to Y$ en probabilité, auquel cas (X_n, Y_n) converge vers (X, Y) en probabilité, et donc en loi.

Un troisième cas plus intéressant que ces deux cas trop particuliers est fourni par le lemme de Slutzky : le cas où Y est presque sûrement constante, égale à c. On rappelle que dans ce cas, la convergence en loi $Y_n \rightsquigarrow c$ est équivalente à la convergence en probabilité $Y_n \to c$. Pour simplifier l'énoncé, on suppose ici que toutes les variables aléatoires sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On pourrait évidemment lever cette hypothèse en indiçant correctement toutes les probabilités et espérances comme nous l'avons fait plus haut.

Lemme 26.8 (Slutzky) Soient (X_n) et (Y_n) deux suites de vecteurs aléatoires prenant leurs valeurs respectivement dans $\mathbb{R}^d$ et $\mathbb{R}^p$, où p et d sont deux éléments de $\mathbb{N}^*$ éventuellement différents. Si $X_n \rightsquigarrow X$ pour un certain vecteur aléatoire X et $Y_n \to c$ en probabilité pour un vecteur déterministe c, alors $(X_n, Y_n) \rightsquigarrow (X, c)$.

Le plus souvent, on applique à cette convergence jointe $(X_n, Y_n) \rightsquigarrow (X, c)$ une fonction continue g et l'on en tire $g(X_n, Y_n) \rightsquigarrow g(X, c)$. Des fonctions souvent considérées sont, dans le cas où $d = p = 1$, la somme $g(x, y) = x + y$, le produit $g(x, y) = xy$ ou la division $g(x, y) = x/y$, que l'on réservera toutefois au cas où $c \neq 0$. Ces formules pour la somme et le produit s'étendent de manière naturelle au cas multidimensionnel lorsque $d = p$.

Preuve On va utiliser la caractérisation (3) du lemme de Portmanteau (Lemme 26.7) et on considère à cet effet une fonction lipschitzienne bornée $f : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}$ arbitraire. On note L sa constante de Lipschitz et B un majorant uniforme sur les valeurs qu'elle prend. Il s'agit de montrer que le terme suivant converge vers 0,

$$\Big|\mathbb{E}\big[f(X_n, Y_n)\big]-\mathbb{E}\big[f(X, c)\big]\Big| \leqslant \Big|\mathbb{E}\big[f(X_n, Y_n)-f(X_n, c)\big]\Big|+\Big|\mathbb{E}\big[f(X_n, c)\big]-\mathbb{E}\big[f(X, c)\big]\Big| .$$

Le second terme dans la majoration tend vers 0 par le point (3) du lemme de Portmanteau appliqué à la fonction $f(\,\cdot\,, c)$ et à la convergence $X_n \rightsquigarrow X$. On montre maintenant que c'est également le cas du premier terme. Pour cela, on fixe dans un premier temps $\varepsilon > 0$ et on décompose ce premier terme en fonction de la valeur de $\|Y_n - c\|$:

$$\Big|\mathbb{E}\big[f(X_n, Y_n) - f(X_n, c)\big]\Big| \leqslant 2B\ \mathbb{P}\big\{\|Y_n - c\| > \varepsilon\big\} + L\varepsilon\ \mathbb{P}\big\{\|Y_n - c\| \leqslant \varepsilon\big\} .$$

Comme $Y_n \to c$ en probabilité, il vient

$$\limsup \Big|\mathbb{E}\big[f(X_n, Y_n) - f(X_n, c)\big]\Big| \leqslant L\varepsilon .$$

Ceci étant vrai pour tout $\varepsilon > 0$, cela montre la convergence vers 0 du membre de gauche et conclut la preuve. □

Extension du théorème de convergence dominée à la convergence en probabilité

On énonce enfin une extension souvent méconnue du théorème de convergence dominée au cas où les variables aléatoires ne convergent pas presque sûrement mais en probabilité. Ici, il est important que toutes les variables aléatoires soient définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On l'énonce pour des variables aléatoires réelles unidimensionnelles mais, quitte à travailler composante par composante, le résultat s'étend à des vecteurs aléatoires.

Théorème 26.9 *Soit (X_n) une suite de variables aléatoires réelles convergeant en $\mathbb{P}$–probabilité vers X et telle que les X_n sont toutes dominées par $Y \in \mathbb{L}^1(\mathbb{P})$. Alors $X \in \mathbb{L}^1(\mathbb{P})$ et $X_n \to X$ dans $\mathbb{L}^1(\mathbb{P})$.*

Preuve On montre que 0 est la seule valeur d'adhérence possible pour la suite

$$(r_n)_{n\geqslant 1} = \Big(\mathbb{E}\big[|X_n - X|\big]\Big)_{n\geqslant 1} .$$

Soit $\ell \in [-\infty, +\infty]$ une de ses valeurs d'adhérence (*a priori*, éventuellement infinie) et soit une sous-suite convergeant vers ℓ, donnée par $\phi : \mathbb{N} \to \mathbb{N}$, i.e., $r_{\phi(n)} \to \ell$.

La suite $\big(X_{\phi(n)}\big)$ converge en particulier en probabilité vers X, on peut donc en extraire une sous-suite convergeant $\mathbb{P}$–presque sûrement vers X, donnée par $\phi \circ \psi$ où $\psi : \mathbb{N} \to \mathbb{N}$:

$$X_{\phi(\psi(n))} \longrightarrow X \qquad \mathbb{P}\text{–p.s.}$$

A la sous-suite $X_{\phi(\psi(n))}$, dominée par Y, on peut alors appliquer le théorème de convergence dominée *ordinaire* et voir qu'elle converge vers X dans $\mathbb{L}^1(\mathbb{P})$, soit $r_{\phi(\psi(n))} \to 0$. Nécessairement, $\ell = 0$ et l'extension est prouvée. □

26.2.2 Inégalités utiles

Retour sur l'inégalité de Hoeffding

On prouve dans un premier temps l'inégalité de Hoeffding (Lemme 3.7). Elle repose sur le lemme suivant, dit lemme de Hoeffding.

Lemme 26.10 (Lemme de Hoeffding) Soit Y une variable aléatoire bornée : il existe deux réels a et b tels que $a \leqslant Y \leqslant b$ presque sûrement. Alors pour tout $s > 0$,

$$\log \mathbb{E}\big[e^{sY}\big] \leqslant s\,\mathbb{E}[Y] + \frac{s^2}{8}(b-a)^2 .$$

On propose une preuve élégante mais plus longue que la preuve plus classique et calculatoire présentée, par exemple, par [GS01].

Preuve On note, pour tout $s \geqslant 0$,

$$\psi(s) = \log \mathbb{E}\big[e^{sY}\big] .$$

Comme Y est bornée, des applications répétées du Théorème 26.1 montrent que les fonctions $s \in \mathbb{R}_+^* \mapsto \mathbb{E}\big[e^{sY}\big]$ et ψ sont bien définies et de classe C^∞ sur $\mathbb{R}_+^*$. Par ailleurs, les dérivées premières et secondes sont données, pour $s > 0$, par

$$\psi'(s) = \frac{\mathbb{E}\big[Ye^{sY}\big]}{\mathbb{E}\big[e^{sY}\big]} \qquad \text{et} \qquad \psi''(s) = \frac{\mathbb{E}\big[Y^2e^{sY}\big]\,\mathbb{E}\big[e^{sY}\big] - \Big(\mathbb{E}\big[Ye^{sY}\big]\Big)^2}{\Big(\mathbb{E}\big[e^{sY}\big]\Big)^2} .$$

On prolonge ces formules par continuité en $s = 0$.

Interprétons $\psi''(s)$ comme une certaine variance :

$$\psi''(s) = \mathbb{E}_{\mathbb{Q}}\big[Y^2\big] - \big(\mathbb{E}_{\mathbb{Q}}[Y]\big)^2 = \mathrm{Var}_{\mathbb{Q}}\,Y ,$$

où la loi $\mathbb{Q}$ est absolument continue par rapport à $\mathbb{P}$, de densité donnée par

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(\omega) = \frac{1}{\mathbb{E}\big[e^{sY}\big]} e^{sY(\omega)} ,$$

pour tout $\omega \in \Omega$, où Ω est l'espace probabilisé sur lequel est définie Y. En particulier, en utilisant la remarque de la page 24, il vient

$$\psi''(s) = \mathrm{Var}_{\mathbb{Q}}\,Y \leqslant \frac{(b-a)^2}{4} .$$

En intégrant deux fois et en utilisant le fait que $\psi'(0) = \mathbb{E}[Y]$ et $\psi(0) = 0$, on obtient que pour tout $s > 0$,

$$\psi(s) \leqslant \frac{(b-a)^2}{4}\,\frac{s^2}{2} + s\,\mathbb{E}[Y] ,$$

qui est le résultat attendu. □

Preuve du Lemme 3.7 La preuve de l'inégalité de Hoeffding se déduit du lemme de Hoeffding par une majoration de Chernoff (i.e., une application de l'inégalité de Markov à l'exponentielle de la somme de variables aléatoires d'intérêt) : pour tout $s > 0$,

$$\mathbb{P}\left\{\sum_{i=1}^{n} Y_i \geqslant t\right\} = \mathbb{P}\left\{e^{s(Y_1+\ldots+Y_n)} \geqslant e^{st}\right\} \leqslant \frac{\mathbb{E}\left[e^{s(Y_1+\ldots+Y_n)}\right]}{e^{st}} .$$

Or, par indépendance des Y_i, leur caractère centré et par le lemme de Hoeffding (Lemme 26.10),

$$\mathbb{E}\left[e^{s(Y_1+\ldots+Y_n)}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{sY_i}\right] \leqslant \prod_{i=1}^{n} e^{s^2(b_i-a_i)^2/8} .$$

Ainsi, pour tout $s > 0$,

$$\mathbb{P}\left\{\sum_{i=1}^{n} Y_i \geqslant t\right\} \leqslant \exp\left(-st + \frac{s^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right) .$$

En considérant

$$s = \frac{4t}{\displaystyle\sum_{i=1}^{n}(b_i - a_i)^2} ,$$

on optimise le membre de droite de la dernière inégalité et on conclut la preuve. □

Le théorème de Berry–Esséen

On énonce enfin le théorème de Berry–Esséen, auquel on avait fait allusion au début du paragraphe 3.4. Ce théorème donne une borne uniforme sur la différence entre deux fonctions de répartition, celle de la statistique considérée dans le théorème de la limite centrale et Φ, celle de la loi limite $\mathcal{N}(0, 1)$.

Théorème 26.11 (Berry–Esséen) *Soit $(Y_1, \ldots, Y_n)$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées, de loi commune admettant un moment d'ordre deux et dont on note respectivement μ et σ^2 la moyenne et la variance. On a alors pour tout $n \geqslant 1$,*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{\frac{\sqrt{n}}{\sigma}\left(\overline{Y}_n - \mu\right) \leqslant t\right\} - \Phi(t) \right| \leqslant 3\,\frac{\mathbb{E}\left[\left|Y_1 - \mu\right|^3\right]}{\sigma^3\sqrt{n}} .$$

On ne propose pas la preuve de ce résultat, mais le lecteur pourra se référer par exemple à [Dur96, paragraphe 2.4.d], qui fournit une preuve complète donnant la constante 3 et indique que cette dernière peut être améliorée (dans l'état actuel de la recherche, sans connaître sa valeur optimale, on sait qu'elle est plus petite que 0.8).