

SÉMINAIRE BRELOT-CHOQUET-DENY. THÉORIE DU POTENTIEL

PAUL-ANDRÉ MEYER

Théorèmes fondamentaux du calcul des probabilités

Séminaire Brelot-Choquet-Deny. Théorie du potentiel, tome 5 (1960-1961), exp. n° 1, p. 1-10

http://www.numdam.org/item?id=SB CD_1960-1961__5__A2_0

© Séminaire Brelot-Choquet-Deny. Théorie du potentiel
(Secrétariat mathématique, Paris), 1960-1961, tous droits réservés.

L'accès aux archives de la collection « Séminaire Brelot-Choquet-Deny. Théorie du potentiel » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

THÉORÈMES FONDAMENTAUX DU CALCUL DES PROBABILITÉS

par Paul-André MEYER

Cet exposé ne prétend pas se substituer à un cours de la théorie de la mesure : il s'adresse à un lecteur déjà familiarisé avec la théorie des mesures de Radon, telle qu'elle figure dans les premiers chapitres de l'intégration de BOURBAKI. Les théorèmes communs à la théorie de la mesure abstraite et à celle de l'intégration dans les espaces localement compacts seront simplement énoncés, dans la terminologie propre au calcul des probabilités. On pourra en trouver la démonstration dans [3], [2] ou [1]. Certains résultats moins familiers seront complètement démontrés. Nous appellerons "espace LCD", dans toute la suite, un espace localement compact à base dénombrable : un tel espace est métrisable, et nous noterons d l'une quelconque des distances compatibles avec sa topologie.

I. Tribus et variables aléatoires.

A. DÉFINITION 1. - On appelle tribu de parties d'un ensemble Ω un ensemble F de parties de Ω , satisfaisant aux propriétés suivantes :

$$1^\circ \emptyset \in F$$

$$2^\circ A \in F \Rightarrow C A \in F.$$

$$3^\circ \text{ Si } \{A_n\}_{n \in \mathbb{N}} \text{ est une suite de parties de } F, \text{ alors } \bigcup_{n \in \mathbb{N}} A_n \text{ appartient à } F.$$

Le couple formé d'un ensemble Ω et d'une tribu F sur Ω sera appelé espace mesurable. Les ensembles de F prendront le nom d'ensembles mesurables, ou d'événements. L'événement $\emptyset$ est dit impossible, l'événement Ω , certain.

B. DÉFINITION 2. - Soient (Ω, F) , (E, T) deux espaces mesurables : on appelle application mesurable (plus précisément : F , T -mesurable) une application $f : \Omega \rightarrow E$, telle que $f^{-1}(T) \subset F$. On note alors : $f : (\Omega, F) \rightarrow (E, T)$. On dit aussi que f est une variable aléatoire (définie sur (Ω, F) , à valeurs dans (E, T)).

THÉORÈME 1. - Soient deux applications mesurables :

$$f : (\Omega, F) \rightarrow (E, T),$$

$$g : (E, T) \rightarrow (H, U).$$

L'application composée $g \circ f : (\Omega, \mathcal{F}) \rightarrow (H, \mathcal{U})$ est mesurable.

C. Tribu engendrée par une famille d'applications.

Comme toute intersection (dans $\mathcal{P}(\Omega)$) d'une famille de tribus sur Ω est une tribu, on peut poser les deux définitions suivantes :

DÉFINITION 3. - Soit A un ensemble de parties de Ω : on appelle tribu engendrée par A , et on note $B(A)$, la plus petite des tribus sur Ω pour lesquelles tous les ensembles de A sont mesurables.

DÉFINITION 4. - Soit Ω un ensemble, $(E_i, \mathcal{T}_i)_{i \in I}$ des espaces mesurables, f_i des applications de Ω dans E_i : on appelle tribu engendrée par les applications f_i , et on note $B(f_i, i \in I)$, la plus petite des tribus sur Ω pour lesquelles toutes les applications f_i sont mesurables.

Exemples de tribus.

1° Soient μ une mesure de Radon sur un espace LCD, $\mathcal{E}$: les ensembles μ -mesurables forment une tribu sur E .

2° Soit E un espace LCD : on appelle tribu borélienne sur E la tribu engendrée par les parties compactes de E : elle est aussi engendrée par les ouverts, par les boules fermées, par les fonctions continues à support compact. Ses éléments seront appelés ensembles boréliens.

Lorsque nous parlerons dans la suite, de variables aléatoires à valeurs dans un espace LCD sans préciser la tribu, il sera sous-entendu qu'il s'agit de la tribu borélienne.

3° Soient $(E_i, \mathcal{T}_i)_{i \in I}$ des espaces mesurables, et $\Omega = \prod_{i \in I} E_i$: on appelle tribu produit sur Ω , et on note $\prod_{i \in I} \mathcal{T}_i$, la tribu engendrée par les applications coordonnées. Elle est aussi engendrée par les "cylindres" de la forme $\prod_{i \in I} A_i$, où $A_i \in \mathcal{T}_i$ pour tout i , $A_i = E_i$ sauf pour un nombre fini d'indices. Si E est un espace LCD, il en est de même de E^n , et la tribu borélienne de E^n est le produit des tribus boréliennes des facteurs.

THÉORÈME 2. - Soient $(E_i, \mathcal{T}_i)_{i \in I}$ des espaces mesurables, f_i des applications de Ω dans E_i , $(H, \mathcal{U})$ un espace mesurable : pour qu'une application $h : (H, \mathcal{U}) \rightarrow (\Omega, B(f_i, i \in I))$ soit mesurable, il faut et il suffit que toutes les applications $f_i \circ h : (H, \mathcal{U}) \rightarrow (E_i, \mathcal{T}_i)$ soient mesurables.

Une remarque qui nous servira souvent sera celle-ci : $B(f_i, i \in I)$ est la réunion des tribus $B(f_i, i \in J)$, où J parcourt la famille des parties dénombrables de I .

D. - Résultats concernant les variables aléatoires numériques.

THÉORÈME 3. - Soit (Ω, F) un espace mesurable :

a. La limite d'une suite (simplement) convergente de variables aléatoires numériques sur Ω est mesurable. L'ensemble de convergence d'une suite quelconque de variables aléatoires numériques est mesurable.

b. Pour qu'une fonction numérique f sur Ω soit mesurable, il faut et il suffit qu'elle soit égale à la limite d'une suite croissante de fonctions numériques mesurables dont l'ensemble des valeurs est dénombrable ("fonctions étagées mesurables").

c. Les variables aléatoires numériques sur Ω forment une algèbre sur R .

d. Pour qu'une fonction f soit mesurable, il faut et il suffit que f^+ et f^- le soient.

REMARQUE. - (a) et (b) se généralisent facilement (au mot "croissante" près) aux variables aléatoires à valeurs dans un espace LCD.

THÉORÈME 4. - Soit $f : (\Omega, F) \rightarrow (E, T)$: pour qu'une variable aléatoire numérique sur Ω soit $B(f)$ -mesurable, il faut et il suffit qu'elle soit de la forme $g \circ f$, où g est une variable aléatoire numérique mesurable sur E .

DÉMONSTRATION. - Les fonctions étagées $B(f)$ -mesurables étant évidemment de cette forme, montrons que, si des $h_n = g_n \circ f$ tendent en croissant vers une fonction h finie, h est de cette forme : or, les g_n croissent sur le sous-ensemble $f(\Omega)$ de E : soit A l'ensemble de convergence de la suite g_n : il est T -mesurable et contient $f(\Omega)$: si l'on pose $g = \lim g_n$ sur A , $g = 0$ hors de A , on a $h = g \circ f$. On conclut par le théorème 3.

THÉORÈME 5. - Les notations sont celles de la définition 4 : soit H un espace vectoriel de fonctions numériques finies (resp. bornées) sur Ω , tel que, pour toute suite croissante (resp. croissante et majorée uniformément) de fonctions positives h_n de H , qui converge vers une fonction finie h , on ait $h \in H$: si H contient tout produit fini de la forme $h_{i_1} \circ f_{i_1} \dots h_{i_n} \circ f_{i_n}$, où chaque

h_i est une fonction numérique bornée sur E_i , T_i -mesurable, H contient toutes les fonctions mesurables (resp. mesurables bornées) par rapport à $B(f_i, i \in I)$.

DÉMONSTRATION. - Construisons par récurrence transfinie les ensembles suivants : K_1 est l'ensemble des combinaisons linéaires finies de produits de la forme ci-dessus. Si b est un ordinal dénombrable de seconde espèce, K_b est la réunion des K_a , $a < b$. Enfin, K_{a+1} est l'ensemble des différences de deux fonctions positives, dont chacune est limite d'une suite croissante (resp. croissante et uniformément bornée) de fonctions positives de K_a . Soit K la réunion des K_a associés aux ordinaux dénombrables : il est clair que chaque K_a , donc K , est contenu dans H ; que chaque K_a , donc K , est une algèbre sur R : l'ensemble des parties de Ω dont la fonction caractéristique appartient à K est donc une tribu, qui contient $B(f_i, i \in I)$: H contient donc les fonctions étagées mesurables par rapport à cette dernière tribu, et on conclut par le théorème 3.

II. Lois de probabilité et espérances mathématiques.

A. Définition des lois de probabilité. - Soit (Ω, F) un espace mesurable : on appelle loi de probabilité sur (Ω, F) une application p de F dans R_+ telle que $p(\Omega) = 1$, et complètement additive, c'est-à-dire telle que pour toute suite $\{A_n\}_{n \in N}$ de parties disjointes de F , on ait $p(\bigcup_{n \in N} A_n) = \sum_{n \in N} p(A_n)$. La valeur $p(A)$ est appelée probabilité de l'évènement A . Un évènement dont la probabilité est 1 (resp. 0) est dit presque certain, ou presque sûr (resp. presque impossible). Deux variables aléatoires égales presque sûrement sont dites équivalentes. Une loi de probabilité p sur (Ω, F) est dite complète si tout ensemble qui est contenu dans une partie de F ; dont la probabilité est 0, appartient à F (sa probabilité est alors 0).

Il est facile de voir que, quelle que soit la loi sur (Ω, F) , la famille des parties de Ω qui ne diffèrent d'un ensemble de F que par un sous-ensemble d'un ensemble de F dont la probabilité est 0, est une tribu F' , et que p se prolonge alors en une application p' de F' dans R_+ de manière unique, de telle sorte que p' soit une loi de probabilité complète sur (Ω, F') . On dit que cette loi est la complétion de p .

EXEMPLE. - Soit E un espace LCD, μ une mesure de Radon positive, de masse 1, sur E . Comme les ensembles boréliens de E sont μ -intégrables, μ détermine une loi de probabilité sur la tribu borélienne de E : une telle loi sera appelée

"loi de Radon" dans la suite. La complétion de cette loi est son extension à la tribu des ensembles μ -mesurables.

B. Espérances mathématiques :

THÉOREME 1. - Soit p une loi de probabilité sur un espace (Ω, F) . Il existe un espace vectoriel, noté $L^1(\Omega, F, p)$, dont les éléments sont des variables aléatoires numériques et une forme linéaire sur cet espace, positive pour l'ordre naturel, notée E , et appelée espérance mathématique, possédant les propriétés suivantes :

1° Si $A \in F$, sa fonction caractéristique χ_A appartient à L^1 , et l'on a : $E(\chi_A) = p(A)$.

2° Pour que $f \in L^1$, il faut et il suffit que f^+ et f^- appartiennent à L^1 .

3° Si une suite croissante de fonctions positives f_n de L^1 converge vers une fonction f , $f \in L^1$ si et seulement si $\sup_n E(f_n) < \infty$, et $E(f) = \sup_n E(f_n)$. L'espace vectoriel L^1 et la forme E sont uniquement déterminés par ces conditions. Une fonction $f \in L^1$ est aussi dite intégrable, et le nombre $E(f)$ est aussi appelé intégrale de f par rapport à p , et noté $\int_{\Omega} f(\omega) dp(\omega)$.

THÉOREME 2 (Théorème de Lebesgue). - Soient f_n une suite de fonctions intégrables, majorées en module par une fonction intégrable fixe, qui convergent presque partout, et f une fonction mesurable égale presque partout à leur limite : f est intégrable, et la suite des $E(f_n)$ converge vers $E(f)$.

Dans le théorème suivant, comme dans toute la suite, nous attribuerons la valeur $+\infty$ au symbole $E(f)$, si f est une fonction positive mesurable, non intégrable. L'énoncé de la condition 3° est encore vrai pour de telles fonctions.

THÉOREME 3 (Lemme de Fatou). - Soit f_n une suite de fonctions positives mesurables : on a l'inégalité :

$$E(\liminf_{n \in \mathbb{N}} f_n) \leq \liminf_{n \in \mathbb{N}} E(f_n) \quad .$$

C. Intégration par rapport à une mesure image.

DÉFINITION. - Soient (Ω, F) , (E, T) deux espaces mesurables, $f : (\Omega, F) \rightarrow (E, T)$ une application mesurable, p une loi de probabilité sur (Ω, F) : on appelle loi image de p par f la loi $f(p) = q$ sur (E, T) définie par :

$$q(A) = p(f^{-1}(A)) \quad \text{si } A \in T.$$

THÉOREME 4. - Soit g une variable aléatoire numérique sur E : pour qu'elle soit q -intégrable, il faut et il suffit que $g \circ f$ soit p -intégrable, et l'on a :

$$\int_E g(e) dq(e) = \int_{\Omega} g \circ f(\omega) dp(\omega) \quad .$$

D. THÉOREME 5. - Soient E un espace LCD, B sa tribu borélienne : toute loi de probabilité p sur (E, B) est une loi de Radon (cf n° 1).

DÉMONSTRATION immédiate. - Toute fonction continue à support compact f est mesurable et bornée, donc p -intégrable : la forme linéaire $f \rightarrow E(f)$ est une mesure de Radon μ , et tout ensemble borélien est μ -mesurable. L'ensemble des fonctions boréliennes bornées f telles que $E(f) = \mu(f)$ est un espace vectoriel, stable pour les passages à la limite croissantes ou décroissantes, uniformément majorés, et contenant les fonctions continues à support compact. On conclut comme dans le théorème 5, paragraphe 1.

Nous ne distinguerons pas par la suite la loi de Radon p et sa mesure associée. Soient (Ω, F, p) une loi de probabilité, (E, T) un espace mesurable, f une application mesurable de Ω dans E : la loi image $f(p)$ sur E sera souvent appelée répartition de la variable aléatoire f . Si E est un espace LCD, T sa tribu borélienne, on appellera aussi répartition de f la mesure de Radon associée à la loi image.

E. Intégration des lois de probabilité.

DÉFINITION. - Soient (X, S) et (Y, T) deux espaces mesurables : on appelle famille S -mesurable de lois de probabilités sur (Y, T) une famille $\{q_x\}_{x \in X}$ de lois de probabilité sur (Y, T) , telle que, pour toute partie A , T -mesurable de Y , la fonction $x \rightarrow q_x(A)$ soit S -mesurable sur X .

Nous ferons constamment dans la suite l'abus de notation suivant : si f est une fonction T -mesurable et positive (par exemple) sur Y , nous écrirons $\int_Y f(y) q(x, dy)$ au lieu de $\int_Y f(y) dq_x(y)$. Les avantages sont d'ordre purement typographique.

THÉOREME 6 (Théorème de Fubini). - Soit Z l'espace $X \times Y$, muni de la tribu $U = S \times T$:

1° Pour toute fonction $f(x, y)$, U -mesurable sur Z , chacune des applications

partielles $x \rightarrow f(x, y)$, $y \rightarrow f(x, y)$ est mesurable sur l'espace facteur correspondant.

2° Soient p une loi de probabilité sur (X, S) et une famille S -mesurable $\{q_x\}$ de lois de probabilité sur (Y, T) : il existe une loi de probabilité et une seule sur (Z, U) , notée $r = \int_X q_x dp(x)$, telle que, pour toute partie de Z de la forme $A \times B$, on ait :

$$r(A \times B) = \int_A dp(x) q_x(B) \quad .$$

3° Soit f une fonction positive, U -mesurable sur Z : la fonction $x \rightarrow \int q(x, dy) f(x, y)$ est S -mesurable, et l'on a :

$$\int f(z) dr(z) = \int_X dp(x) \int_Y q(x, dy) f(x, y) \quad .$$

4° Si toutes les lois q_x sont identiques à une même loi q sur (Y, T) , la loi r est appelée loi produit, et notée $p \otimes q$. Dans ce cas, les lois p et q jouent des rôles symétriques.

REMARQUES. - Si f n'est pas positive, le second membre de la relation (3) peut avoir un sens sans que le premier en ait un.

Si, dans le cas de 4°, les lois (X, S, p) , (Y, T, q) , sont complètes il n'en est pas de même de la loi $(Z, U, p \otimes q)$: si la fonction $f(x, y)$ est mesurable sur $(Z, U, (p \otimes q))$ (loi complétée), l'assertion de 3° reste vraie, à cela près que p -presque toute et non plus toute application partielle $x \rightarrow f(x, y)$ est maintenant S -mesurable.

L'extension des résultats ci-dessus au cas d'un nombre fini quelconque de facteurs ne présente aucune difficulté.

F. Définition de l'indépendance.

DÉFINITION. - Soit (Ω, F, p) une loi de probabilité, et $\{f_i\}_{i \in I}$ une famille finie de variables aléatoires à valeurs dans des espaces (E_i, T_i) : l'application produit $\{f_i\}_{i \in I}$ à valeurs dans $(\prod_{i \in I} E_i, \prod_{i \in I} T_i)$ est mesurable : nous dirons que les variables f_i sont indépendantes si la répartition de la variable aléatoire $\{f_i\}_{i \in I}$ dans l'espace produit, est le produit des répartitions des f_i dans les facteurs : autrement dit, si pour toute famille d'ensembles $A_i \in T_i$, on a :

$$p(f_{i_1} \in A_{i_1}, \dots, f_{i_n} \in A_{i_n}) = p(f_{i_1} \in A_{i_1}) \dots p(f_{i_n} \in A_{i_n}).$$

Les variables aléatoires d'une famille quelconque sont dites indépendantes si celles de toute sous-famille finie le sont.

EXEMPLE. - Soient f et g deux variables numériques indépendantes : μ et ν leurs répartitions : la répartition de leur somme $f + g$ est l'image de la répartition $\mu \otimes \nu$ sur $\mathbb{R}^2$, par l'application $(x, y) \rightarrow x + y$: c'est la convolution $\mu * \nu$.

Indépendance de tribus. - Soient $\mathcal{F}_i$ une famille de sous-tribus de la tribu $\mathcal{F}$ de l'espace Ω , et f_i l'application identique de $(\Omega, \mathcal{F})$ sur $(\Omega, \mathcal{F}_i)$, qui est mesurable : on dit que les tribus $\mathcal{F}_i$ sont indépendantes si les f_i le sont. Inversement, des variables aléatoires g_i sont indépendantes, si et seulement si les tribus $\mathcal{B}(g_i)$ le sont.

Pour cela, il faut et il suffit que, pour toute famille finie $i_1 \dots i_n$ d'indices distincts, on ait, si les $A_{i_k} \in \mathcal{F}_{i_k}$:

$$P\left(\bigcap_k A_{i_k}\right) = \prod_k P(A_{i_k})$$

ou encore, pour des variables aléatoires numériques bornées (par exemple) Y_{i_k} , $\mathcal{F}_{i_k}$ -mesurables :

$$E\left(\prod_k Y_{i_k}\right) = \prod_k E(Y_{i_k})$$

Pour qu'une tribu soit indépendante d'elle-même, il faut et il suffit qu'elle ne contienne que des événements presque certains ou presque impossibles.

EXEMPLE. - Soit $X_1 \dots X_n \dots$ une suite de variables indépendantes, et $\mathcal{F}_\infty$ l'intersection des tribus $\mathcal{F}_n = \mathcal{B}(X_p, p \geq n)$: $\mathcal{F}_\infty$ est indépendante d'elle-même.

G. Appendice : variables aléatoires uniformément intégrables.

DÉFINITION. - Une famille de variables aléatoires numériques $\{f_i\}_{i \in I}$, intégrables sur $(\Omega, \mathcal{F}, p)$ est dite uniformément intégrable (ou les variables aléatoires elles-mêmes sont dites uniformément intégrables) si, lorsque $n \rightarrow \infty$, les intégrales $\int_{\{|f_i| > n\}} |f_i| dp$ tendent vers 0 uniformément en i .

THÉORÈME 7.

1° Pour qu'une famille $\{f_i\}$ soit uniformément intégrable, il faut et il suffit :

a. que les $E(|f_i|)$ soient uniformément bornées.

b. que les intégrales $\int_A |f_i| dp$ tendent vers 0 uniformément en i , lorsque $p(A) \rightarrow 0$ (c'est-à-dire $\forall \varepsilon, \exists \eta$ tel que $p(A) < \eta \Rightarrow \forall i \int_A |f_i| dp < \varepsilon$)

2° Soit f_n une suite de variables aléatoires intégrables qui converge presque partout vers une fonction intégrable f . Pour que les f_n convergent vers f au sens de la norme $\| \cdot \|_1$, il faut et il suffit qu'elles soient uniformément intégrables. Si les f_n sont positives, cela équivaut encore à la convergence de $E(f_n)$ vers $E(f)$.

DÉMONSTRATION.

1° Il suffit évidemment de raisonner dans le cas d'une suite f_n , de variables aléatoires positives : soit c une constante.

Posons $f_n^c(\omega) = f_n(\omega)$ si $f_n(\omega) \leq c$, $f_n^c(\omega) = c$ sinon, et posons $f_{nc} = f_n - f_n^c$. L'hypothèse d'uniforme intégrabilité équivaut à : $E(f_{nc}) \xrightarrow{c \rightarrow \infty} 0$ uniformément

en n . Comme $\int_A f_n dp \leq cp(A) + E(f_{nc})$, l'uniforme intégrabilité implique les conditions (a) et (b). La réciproque est évidente, car $p(\{f_n \geq c\}) \leq \frac{1}{c} E(f_n)$.

2° Si les $f_n \xrightarrow{\| \cdot \|_1} f$, les $E(|f_n|)$ sont uniformément bornées, et :

$$\int_A |f_n| dp \leq \int_A |f| dp + \| |f_n| - |f| \|_1.$$

Les conditions (a) et (b) sont donc vérifiées. Inversement, si les $f_n \geq 0$ sont u.i. et tendent presque partout vers f , f est intégrable (lemme de Fatou) et

$$E(|f_n - f|) \leq E(|f_n^c - f^c|) + E(f_{nc}) + E(f_c).$$

On choisit c de manière à rendre petits les deux derniers termes, le premier $\rightarrow 0$, d'après le théorème de Lebesgue.

Reste à voir que, si les f_n sont positives et tendent presque partout vers f , intégrable, si $E(f_n) \rightarrow E(f)$, alors $E(|f - f_n|) \rightarrow 0$. Comme $f + f_n = \sup(f, f_n) + \inf(f, f_n)$ comme (théorème de Lebesgue) $E[\inf(f, f_n)] \rightarrow E(f)$, il résulte de ce que $E(f_n) \rightarrow E(f)$, que $E[\sup(f, f_n)] \rightarrow E(f)$. On conclut alors en remarquant que

$$E(|f - f_n|) = E[\sup(f, f_n)] - E[\inf(f, f_n)].$$

BIBLIOGRAPHIE

- [1] HALMOS (Paul R.). - Measure theory. - New York, D. Van Nostrand, 1950 (The University Series in higher Mathematics).
 - [2] LOÈVE (Michel). - Probability theory. - New York, D. Van Nostrand, 1955 (The University Series in higher Mathematics).
 - [3] SAKS (Stanislaw). - Theory of the integral 2nd edition (English translation). - New-York, G. E. Stechert and Co, 1937 (Polska Akademia Nauk, Monografie matematyczne, 7).
-