

M. BARBUT

Intégrales et mesures

Mathématiques et sciences humaines, tome 20 (1967), p. 1-28

http://www.numdam.org/item?id=MSH_1967__20__1_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1967, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) 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/>

M. BARBUT

INTEGRALES ET MESURES

Introduction en vue du Calcul des Probabilités

Cette introduction fait suite à mon article: "De Pascal à Savage: un chapitre de l'algèbre linéaire, le Calcul des Probabilités" (M.S.H. n° 15, pp. 15-28, 1966), dans lequel j'essayai de montrer comment, dans le cas fini, l'algèbre linéaire peut être, via les aléas numériques et l'espérance, la clef du Calcul des Probabilités; cet article appelait une suite sur le cas infini, dont la publication est rendue urgente par la mise en place de la seconde année du D.U.E.L. de Psychologie et Sociologie.

Les enseignants devront, en effet, au cours de l'année, non faire une théorie détaillée de l'intégration, de la mesure et des densités, mais essayer de faire

comprendre aux étudiants ce que recouvre un symbole tel que $\int_{-\infty}^x e^{-\frac{t^2}{2}} dt$, et comment ceci se relie à l'algèbre linéaire et au Calcul des Probabilités dans le cas fini, que les étudiants auront étudiés préalablement. Je propose donc aux enseignants deux idées directrices et un parti pédagogique.

Les idées directrices d'abord:

S'appuyer sur une bonne compréhension de ce qu'est une forme linéaire dans le cas fini; procéder par approximations à partir de cas.

Linéarité et approximations sont les deux "idées force"; elles résultent d'ailleurs de la structure de groupe abélien (linéarité) ordonné compact (approximativement fini) de $\mathbb{R}$.

Une seule propriété de $\mathbb{R}$ (qui peut lui servir de définition, quant à son type d'ordre, par complétion d'un ordre dense dénombrable) suffit d'ailleurs à comprendre l'esprit de la plupart des démonstrations: toute suite dénombrable monotone majorée y possède un supremum. On peut soit admettre cette propriété, très intuitive, soit la prendre comme définition, soit la démontrer à partir d'une définition équivalente (par les coupures, par exemple). Ce peut être en tout cas une occasion de faire réfléchir l'étudiant à l'un des sens du mot continu, puis à son lien avec ce qu'on appelle continuité d'une fonction.

Le cas pédagogique ensuite:

Je n'ai traité avec quelque détail que de l'intégration et de la mesure sur un ensemble dénombrable et sur $\mathbb{R}$ (le premier de ces cas permettant de mieux comprendre ce qui se passe dans le second; me semble-t-il), ce sont les seuls cas pratiquement utiles pour l'enseignement que nous visons. En tout cas, je n'introduis des notions telles que celle d'additivité dénombrable, puis de tribu qu'à partir du moment où le problème posé fait comprendre la nécessité de leur intro-

2.

duction. La notion de tribu, par exemple, est parfaitement superflue tant que l'on n'a pas à définir d'intégrales sur $\mathbb{R}$, par rapport à une mesure continue; et rien ne me semble plus redoutable pour les débutants que de leur "balancer" d'emblée le classique triplet $(\Omega, \mathcal{A}, \mathbb{P})$.

Comme j'ai uniquement pensé aux nécessités du Cours de Calcul des Probabilités, je n'ai défini l'intégrale de Lebesgue-Stieltjes que dans le cas positif, et non dans le cas plus général où l'on n'exige que la continuité de l'intégrale, de façon à ne pas allonger encore un exposé déjà volumineux.

D'autre part, dans le paragraphe 1, je n'ai fait que résumer sommairement les points du cas fini sur lesquels on s'appuie; le lecteur trouvera des développements dans mon article M.S.H. n° 15 cité supra.

Lectures complémentaires recommandées:

- Pour un exposé moins archaïque: J. NEVEU, "Les Bases Mathématiques du Calcul des Probabilités", (Masson), Chap. I, II, IV.
- Pour la genèse des idées: H. LEBESGUE, "Leçons sur l'Intégration et la Recherche des Fonctions Primitives", (Gauthier-Villars), ou
E. BOREL, "Leçons sur la Théorie des Fonctions", (Gauthier-Villars).

TABLE

1 - Rappel des idées essentielles du cas fini	p. 4
2 - Cas infini: les principes	p. 7
3 - Mesures et intégrales sur un ensemble dénombrable	p. 11
4 - Mesures et intégrales sur $\mathbb{R}$	p. 15
5 - Densités	p. 24

1.- RAPPEL DES IDEES ESSENTIELLES DU CAS FINI

1.a. Vectoriel réticulé des aléas numériques

Si $A = \{1, 2, \dots, n\}$ est un ensemble fini, muni d'une mesure de probabilité p , dont la distribution de probabilité est $p = (p_1, p_2, \dots, p_n)$, chaque application $f: A \rightarrow \mathbb{R}$, où $\mathbb{R}$ est le corps ordonné des nombres réels, définit un aléa numérique.

L'ensemble des aléas numériques définis sur A est identifiable à l'ensemble $\mathbb{R}^A$ des applications de A dans $\mathbb{R}$. Cet ensemble peut être muni d'une structure de vectoriel réticulé, en posant:

- Addition : $f + g = h \iff \forall x \in A, f(x) + g(x) = h(x)$
- Homothéties : $\lambda f = h \iff \forall x \in A, \lambda f(x) = h(x)$
- Ordre : $f \leq g \iff \forall x \in A, f(x) \leq g(x)$

L'ordre $\leq$, défini par majoration uniforme (sur A) entre applications de A dans $\mathbb{R}$ munit de $\mathbb{R}^A$ d'une structure de treillis (ou ensemble ordonné réticulé) compatible avec la structure de vectoriel.

- Treillis en posant:

- Supremum : $f \vee g = h \iff \forall x \in A, \max(f(x), g(x)) = h(x)$
- Infimum : $f \wedge g = h \iff \forall x \in A, \min(f(x), g(x)) = h(x)$

- Compatible avec la structure de vectoriel:

- $f \leq g \implies \forall h \in \mathbb{R}^A, f + h \leq g + h$
- $f \leq g \implies \forall \lambda > 0, \lambda f \leq \lambda g$

1.b. L'espérance définie par une mesure

Pour un vectoriel réticulé des aléas numériques définis sur A , l'espérance par rapport à la mesure p est donnée par:

$$E_p(f) = \sum_{x \in A} f(x) p_x$$

E_p est une application de $\mathbb{R}^A$ dans $\mathbb{R}$ qui est:

- 1) Linéaire: $\forall \lambda, \forall \mu, E_p(\lambda f + \mu g) = \lambda E_p(f) + \mu E_p(g)$
- 2) Positive: $f \geq 0 \implies E_p(f) \geq 0$ ($f \geq 0$ signifie $\forall x \in A, f(x) \geq 0$)
- 3) Normée : $(\forall x \in A, f(x) = k) \implies E_p(f) = k$

On dit que c'est une forme linéaire, positive, normée du vectoriel $\mathbb{R}^A$.

La positivité équivaut à la monotonie:

$$f \geq g \implies E_p(f) \geq E_p(g)$$

1 et 2 signifient donc que l'espérance est un homomorphisme de $\mathbb{R}^A$ muni de son addition, ses homothéties et de son ordre dans $\mathbb{R}$ muni de son addition, ses homothéties et son ordre.

Etant normée, l'espérance est a fortiori bornée en ce sens que si une fonction f est bornée (uniformément) par un nombre k , son espérance est un nombre fini qui ne dépend que de k et de la fonction f .

1. c. La mesure définie par l'espérance

Réciproquement, si l'on ne suppose pas a priori l'ensemble A probabilisé par une mesure de probabilité p , à toute application I linéaire, positive et normée de $\mathbb{R}^A$ dans $\mathbb{R}$ correspond une mesure de probabilité p_I induite par I sur A .

p_I est défini par la restriction de I aux fonctions caractéristiques des parties de A ; si $B \subset A$, la fonction caractéristique φ_B de B est définie par:

$$\varphi_B(x) = 1 \iff x \in B$$

$$\varphi_B(x) = 0 \iff x \notin B$$

Si l'on pose:

$$\forall B \subset A, \quad p_I(B) = I(\varphi_B)$$

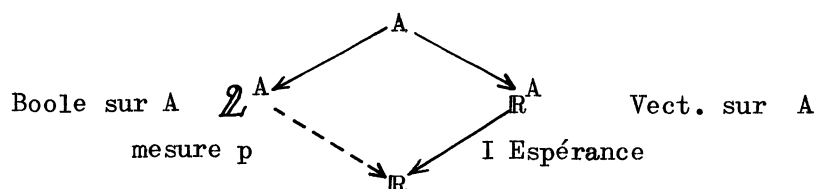
p_I est une mesure de probabilité sur A :

$$- p_I(B \cup C) + (B \cap C) = p_I(B) + p_I(C)$$

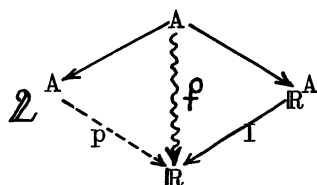
$$- p_I(\emptyset) = 0, \quad p_I(A) = 1$$

$$- \forall B \subset A, \quad p_I(B) \geq 0$$

Désignons par $\mathcal{Z}^A$ l'ensemble des parties de A muni de sa structure d'algèbre booléenne au moyen des opérations $\cup$, $\cap$, et de l'inclusion $\subset$. L'équivalence entre mesures de probabilités $p : \mathcal{Z}^A \rightarrow \mathbb{R}$ et formes linéaires, positives, normées $I : \mathbb{R}^A \rightarrow \mathbb{R}$ est résumée par le diagramme (commutatif):



Si on se donne I (resp. p) on peut compléter le diagramme par p (resp. I); le lien s'établit d'ailleurs par la diagonale du diagramme:



6.

Celle-ci est la distribution de probabilité

$$p = (p_1, p_2, \dots, p_n)$$

de la mesure p , ou encore le vecteur des coefficients (covecteurs) de la forme linéaire I :

$$I(f) = p_1 f_1 = p_2 f_2 + \dots + p_n f_n$$

1.d; Les deux modes de calcul de l'espérance

On a indiqué que l'espérance E d'une application $f : A \longrightarrow \mathbb{R}$ (ou aléa numérique) se calcule par la formule^p:

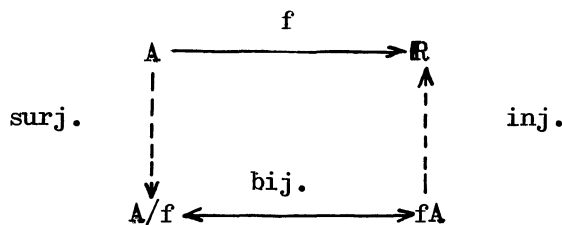
$$(I) \quad E_p(f) = \sum_{x \in A} f(x) p_x$$

La sommation se fait par rapport au départ A de f .

Mais la fonction (application), f détermine sur A une partition en classes A_y , constituées des éléments ayant même image y :

$$A_y = \{ i : i \in A, f_i = y \}$$

L'ensemble des classes A_y , ou quotient A/f , est en correspondance bijective avec l'image fA de f , qui peut être factorisée selon le diagramme



En regroupant dans la somme (I) tous les éléments de A appartenant à une même classe, on obtient une seconde expression de l'espérance, où la sommation se fait par rapport à l'arrivée:

$$(II) \quad E_p(f) = \sum_{y \in fA} y p(A_y)$$

Ce second mode de calcul de l'espérance correspond à une façon très usuelle de calculer les moyennes; sa généralisation au cas où A est infini est l'une des clefs de la définition de l'intégrale selon H. LEBESGUE.

2. - CAS INFINI : LES PRINCIPES

2.a. L'intégrale est définie par approximation

Si A est infini, l'ensemble $\mathbb{R}^A$ des applications de A dans $\mathbb{R}$ est toujours un vectoriel réticulé. L'usage est d'appeler intégrales les formes I linéaires, positives, bornées de ce vectoriel (Cf. 1 - 6). Une intégrale normée (si une fonction f est constante, son intégrale $I(f)$ est égale à cette constante) pourra toujours être interprétée comme l'espérance de f considérée comme un aléa numérique.

Les difficultés et les idées nouvelles vont provenir de ce que les calculs (linéaires) ne vont plus pouvoir se faire que par approximations: approximations au départ $\mathbb{R}^A$ de l'intégrale I , car ce vectoriel ne possède pas de base dénombrable, c'est-à-dire une partie que l'on puisse décrire avec des mots finis et telle que tout vecteur soit combinaison linéaire d'un nombre fini de ces vecteurs; approximations à l'arrivée $\mathbb{R}$, ou plus précisément, continuité de l'intégrale en ce sens que deux fonctions (éléments de $\mathbb{R}^A$) proches l'une de l'autre devront avoir des valeurs de l'intégrale voisine.

La démarche va donc consister à essayer de décrire sinon $\mathbb{R}^A$ tout entier, tout au moins une grande partie de cet ensemble (les fonctions bornées) de façon approximative par des fonctions accessibles au calcul de type fini; puis à définir l'intégrale de sorte que celle des fonctions accessibles au calcul soit elle-même calculable, et qu'elle soit continue: l'intégrale f sera alors approximativement égale à celles (calculables) des approximations (calculables) de f .

2.b. Approximation des fonctions: topologie de la convergence uniforme

Pour définir des approximations (une topologie) dans $\mathbb{R}^A$, il y a bien des façons de procéder; celle que l'on adopte ici découle naturellement de la structure d'ordre de $\mathbb{R}^A$; on a posé en effet:

$$f \leq g \iff \forall x \in A, \quad f x \leq g x$$

f est uniformément (pour tout x) inférieure à g .

Cette définition de l'ordre munit $\mathbb{R}^A$ de sa structure de treillis distributif, compatible avec sa structure de vectoriel ($f \leq g \implies \forall h, \quad f + h \leq g + h$).

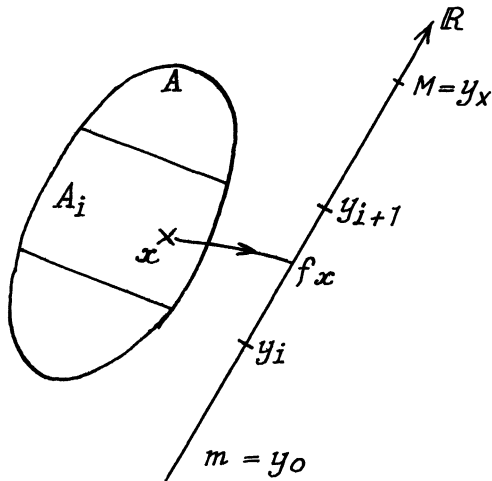
Il est donc naturel de dire que f et g sont deux fonctions proches l'une de l'autre si leur différence (vectorielle) est petite au sens de l'ordre défini supra. On pose ainsi:

$$|f - g| < \underline{\varepsilon} \iff \forall x \in A, \quad |f x - g x| < \varepsilon$$

où $\underline{\varepsilon}$ désigne la fonction constante prenant la valeur ε . $\mathbb{R}^A$ est ainsi muni de la topologie dite de la convergence uniforme.

2.c. Les fonctions étagées approchent uniformément les fonctions bornées

Dans cette topologie, toute fonction bornée peut être arbitrairement approchée par des fonctions étagées, c'est-à-dire ne prenant qu'un nombre fini de valeurs: f est étagée si et seulement si son image fA est finie.



Si en effet f est bornée, c'est-à-dire telle que son image fA soit comprise entre deux nombres m et M , on peut partitionner l'intervalle $[m, M]$ en points de division:

$$m = y_0, < y_1, < \dots < y_i < y_{i+1} < \dots < y_k = M$$

de telle sorte que le plus grand des intervalles (y_i, y_{i+1}) ait une longueur inférieure à ε .

Soit A_i la partie de A dont les éléments ont une image comprise entre y_i (inclus) et y_{i+1} (exclus):

$$A_i = \{x: x \in A, y_i \leq f(x) < y_{i+1}\}$$

Les parties A_i constituent une partition de A . La fonction étagée φ qui pour tout i prend la valeur y_i sur A_i est majorée (uniformément) par f , et est telle que:

$$\varepsilon > f - \varphi \geq 0$$

De même, la fonction étagée ψ qui prend la valeur y_{i+1} sur A_i majore f , et l'on a:

$$\varepsilon > \psi - f \geq 0$$

φ et ψ sont donc deux fonctions étagées qui approchent f à moins de ε , et qui l'encadrent:

$$\varphi \leq f < \psi \quad |\psi - \varphi| < \varepsilon$$

A toute partition finie de l'intervalle $[m, M]$, on peut ainsi faire correspondre un encadrement de f par deux fonctions étagées.

2.d. L'intégration des fonctions étagées: mesures

Or une fonction étagée φ , prenant les valeurs $y_0, y_1, y_2, \dots, y_k$, est combinaison linéaire de fonctions caractéristiques d'ensembles (qui sont des fonctions étagées)

$$\varphi = \sum_{i=0}^k y_i \varphi_{A_i}$$

où φ_{A_i} est la fonction caractéristique de la partie A_i de A sur laquelle φ prend

la valeur y_i (cf. par. 1.c.). Si donc l'intégrale $I(\varphi)$ de φ étagée existe, on a en vertu de la linéarité de l'intégrale:

$$I(\varphi) = \sum_{i=0}^k y_i I(\varphi_{A_i})$$

On sait donc calculer l'intégrale de toutes les fonctions étagées dès que l'on sait calculer celle de toutes les fonctions caractéristiques, c'est-à-dire, comme on l'a rappelé au paragraphe 1.c., dès que l'on sait définir une mesure m sur A .

2.e. L'intégrale des fonctions étagées définit par continuité celle des fonctions bornées

Pour les fonctions étagées, le problème de l'intégrale est ainsi ramené, comme dans le cas fini, à celui des mesures sur A . Avant de traiter cette question, montrons que si l'on a une mesure sur A , alors l'intégrale de toute fonction bornée peut bien être définie par continuité au moyen de celle des fonctions étagées, comme on le souhaitait. Posons comme dans 1.c.:

$$\forall X \subset A \quad I(\varphi_X) = m(X)$$

2.e. 1. - Soient φ et ψ les deux fonctions étagées encadrant f bornée définies comme on l'a indiqué plus haut à partir d'une partition $y_0 < y_1 < \dots < y_k$ de $[m, M]$. On a :

$$I(\psi) - I(\varphi) = I(\psi - \varphi) \quad (\text{linéarité de l'intégrale}).$$

$$I(\psi - \varphi) = \sum_{i=0}^k (y_{i+1} - y_i) m(A_i)$$

Donc :

$$\begin{aligned} I(\psi - \varphi) &\leq (\max_i y_{i+1} - y_i) \sum_i m(A_i) \\ &= (\max_i y_{i+1} - y_i) m(A) < \varepsilon m(A) \end{aligned}$$

$m(A)$ est un nombre fini (1 si l'on veut); $I(\psi - \varphi)$ est donc arbitrairement petite. Puisque d'autre part :

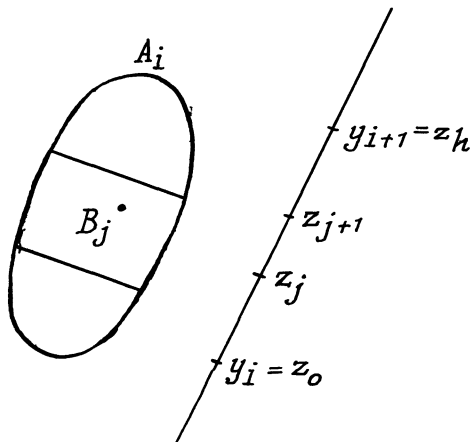
$$\varphi \leq f \leq \psi$$

Alors: $I(\varphi) \leq I(f) \leq I(\psi)$ (monotonie - positivité de l'intégrale)

et $I(f)$ est donc approximable à ε près par $I(\varphi)$ ou $I(\psi)$.

2.e.2. - En outre, lorsque ε tend vers zéro, $I(\varphi)$ et $I(\psi)$ tendent vers un même nombre bien défini, qui est, par définition, l'intégrale de f .

On n'en donnera pas ici de démonstration complète, mais le lecteur saisira ce phénomène de convergence en prenant des partitions de plus en plus fines de $[m, M]$; alors $I(\varphi)$ croît, et $I(\psi)$ décroît. La croissance de $I(\varphi)$, par exemple, se voit en considérant une subdivision $y_i = z_0 < z_1 < \dots < z_j < z_{j+1} < \dots < z_h = y_{i+1}$



de l'intervalle $[y_i, y_{i+1}]$, à laquelle correspond une subdivision en classes $B_0, B_1, \dots, B_h$ de la classe A_i . Pour la fonction étagée φ' associée à cette subdivision, on a, pour le terme correspondant à l'intervalle $[y_i, y_{i+1}]$:

$$\forall j, z_j \geq y_i = z_0; \text{ donc:}$$

$$\begin{aligned} I(\varphi') &= \sum_{j=0}^h z_j m(B_j) > \sum_{j=0}^h y_i m(B_j) \\ &= y_i \sum m(B_j) = y_i m(A_i) = I(\varphi) \end{aligned}$$

Le nombre à déterminer $I(f)$ peut donc être "pris en sandwich" entre deux suites de nombres du type $I(\varphi)$ et $I(\psi)$ calculables (ce sont des sommes finies), la première suite croissant, la seconde décroissant, et l'écart entre les deux suites tendant à zéro; il est bien défini (si les nombres $I(\varphi)$ et $I(\psi)$ le sont).

N.B. - La définition de l'intégrale $I(f)$ d'une fonction bornée f de A dans R au moyen de ses approximations par la minorante $I(\varphi)$ et la majorante $I(\psi)$ repose sur cette propriété du type d'ordre de R que toute suite croissante (resp. décroissante) majorée (resp. minorée) y possède une limite: la suite des $I(\varphi)$ est croissante et majorée par l'une quelconque des $I(\psi)$.

Cette propriété de R , la compacité, qui peut d'ailleurs lui servir de définition à partir du corps $\mathbb{Q}$ des rationnels (qui lui n'est pas compact) est celle qui justifie que l'on prenne, dans les questions traitées ici, R comme ensemble d'arrivée, et non un autre ensemble plus "naturel" de nombres tel que le corps $\mathbb{Q}$ ordonné de rationnels, ou l'anneau ordonné $\mathbb{D}$ des nombres décimaux.

2.f. - L'additivité dénombrable

Mais cette définition par passage à la limite introduit subrepticement une condition supplémentaire sur les mesures. On a en effet fait tendre ε vers zéro; on a ainsi fait tendre vers l'infini le nombre des points de division sur $[m, M]$ et le nombre correspondant des classes A_i de A . Les calculs faits n'ont donc de sens que si l'égalité:

$$m\left(\bigcup_{i=0}^k A_i\right) = \sum_{i=0}^k m(A_i)$$

valable lorsque les A_i sont des parties disjointes reste vrai lorsque k tend vers l'infini, c'est-à-dire si les A_i sont en infinité dénombrable. C'est donc à cause des nécessités de la technique d'approximation que l'on s'intéresse en théorie de l'intégrale non à toutes les mesures, mais seulement à celles qui satisfont à la condition restrictive d'additivité dénombrable (ou σ -additivité).

$$(\sigma) \quad m\left(\bigcup A_i\right) = \sum m(A_i)$$

pour toute famille dénombrable de parties disjointes. On sait que les propriétés (des mesures):

$$m(\emptyset) = 0$$

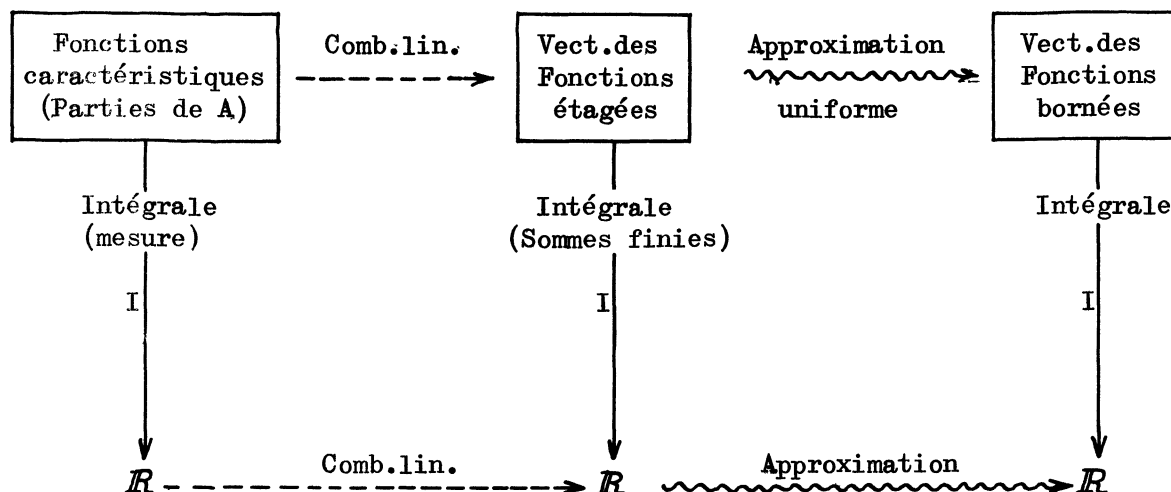
et

$$m(X \cap Y) + m(X \cup Y) = m(X) + m(Y)$$

sont des conséquences de (σ) , et équivalent à l'additivité finie.

2.g. - On est ramené au problème des mesures

La démarche suivie jusqu'ici peut se résumer par le diagramme:



Comme on le verra plus loin, un nouveau passage à la limite portant sur des fonctions bornées, permet d'étendre la définition de l'intégrale à certaines fonctions non bornées.

Tout est donc conditionné par ce qui se passe à la source, c'est-à-dire par la possibilité de définir des mesures qui, on l'a vu, doivent être additivement dénombrables, sur un ensemble A.

C'est ici que la nature de A, qui jusqu'ici n'avait joué aucun rôle, prend une importance capitale.

- ou bien A est tel qu'on puisse définir des mesures, pour lesquelles toute partie de A est mesurable (on verra que c'est le cas si A est dénombrable); alors la question est réglée: à chaque mesure correspond une intégrale, et toute fonction bornée est intégrable.

- ou bien (c'est le cas si $A = \mathbb{R}$) A est tel que l'on ne puisse y définir de mesure que sur une sélection de ses parties, mais non sur toutes. Alors ne seront intégrables que les fonctions bornées telles que les parties A_i du paragraphe 2.c. appartiennent toujours à cette sélection.

Comment il faut faire cette sélection pour que les parties qui la constituent soient support d'une mesure, et pour que les fonctions les plus "utiles" soient intégrables c'est ce que l'on examinera sur le cas $A = \mathbb{R}$ (par. 4).

3. - MESURES ET INTEGRALES SUR UN ENSEMBLE DENOMBRABLE

Si $a = \{1, 2, \dots, i, \dots\}$ est un ensemble dénombrable, une fonction de A dans $\mathbb{R}$ (aléa numérique) est donnée par la suite des valeurs prises:

$$f = (f_1, f_2, \dots, f_i, \dots)$$

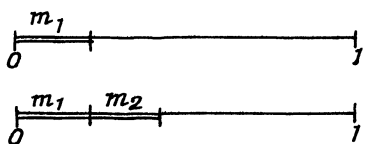
Dire que f est bornée signifie que tous les nombres f_i sont situés dans un intervalle, ou encore qu'il existe un nombre k bornant leurs valeurs absolues:

$\exists k, \forall i, |f_i| < k$ (attention à l'ordre des quantificateurs). On dit souvent que f est une suite bornée. L'ensemble des fonctions bornées de A dans $\mathbb{R}$ forme évidemment un vectoriel.

3.a. - Construction intuitive des mesures

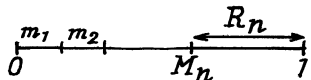
Les mesures positives et bornées sur A peuvent, comme dans le cas fini, être définies de façon atomique c'est-à-dire en affectant à chaque élément $i \in A$ un nombre $m_i \geq 0$, de sorte que $\sum m_i = 1$ (s'il s'agit d'une mesure de probabilité).

On s'en rend facilement compte en considérant l'antique procédure "Achille et la Tortue". Sur le segment $[0,1]$, on choisit un premier intervalle, d'origine 0 et de longueur m_1 ($0 \leq m_1 \leq 1$); si $m_1 < 1$, on choisit ensuite sur l'intervalle restant un segment m_2 ($0 \leq m_2 \leq 1 - m_1$) si $m_1 + m_2 < 1$,



on choisit m_3 , etc Ce procédé peut continuer indéfiniment, en raison de la propriété de $\mathbb{R}$ d'être indéfiniment divisible (dense en soi): entre deux points, on peut toujours en insérer un troisième. Le cas "Achille et la Tortue" est celui où chaque point de subdivision est choisi au milieu de l'intervalle restant.

Si l'on pose:



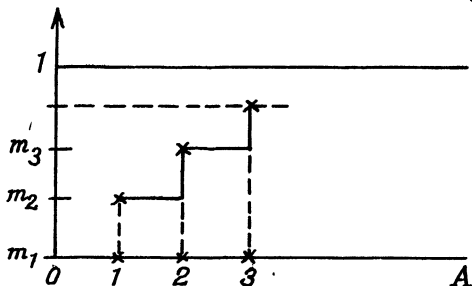
$$M_n = m_1 + m_2 + \dots + m_n$$

M_n est la mesure de la partie finie $\{1, 2, \dots, n\}$ de A , et le reste $R_n (=1 - M_n)$ est celle de la partie complémentaire; lorsque n tend vers l'infini, R_n tend par construction vers zéro. Les distributions "tronquées":

$\mu_n = (m_1, m_2, m_3, \dots, m_n, 0, 0, \dots)$ sont d'ailleurs des fonctions étagées tendant, en croissant, uniformément vers la distribution:

$$\underline{m} = (m_1, m_2, \dots, m_n, m_{n+1}, \dots)$$

Lorsque A est l'ensemble ordonné des entiers naturels $\mathbb{N} = \{1 < 2 < 3 < \dots < n < \dots\}$, le procédé indiqué n'est autre que celui

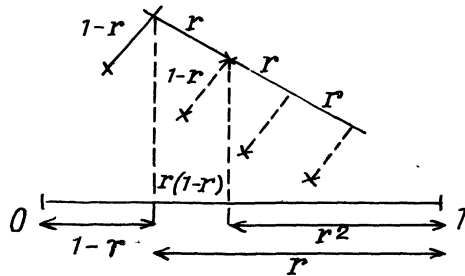


de la construction de la fonction de répartition; mais on se gardera de croire qu'il en est toujours ainsi; A peut n'être pas ordonné, ou bien avoir un ordre qui n'a rien à voir avec l'ordre de numérotation (exemple: A est l'ensemble $\mathbb{Q}$ des rationnels).

3. b. - Exemples

Les deux exemples usuels de telles distributions sont: la distribution de Pascal,

ou géométrique: c'est la généralisation "d'Achille et la Tortue"; à chaque étape, le reste R_n est égal à r^n , où r ($0 < r < 1$) est le paramètre choisi une fois pour toutes de la distribution. On a:



$$m_n = (1 - r) r^{n-1}$$

$$M_n = (1 - r) (1 - r + \dots + r^{n-1}) = 1 - r^n$$

On peut voir cette distribution comme résultant d'un processus visualisé sur la figure ci-contre.

La distribution de Poisson, pour laquelle:

$$m_n = e^{-\lambda} \frac{\lambda^n}{n!}$$

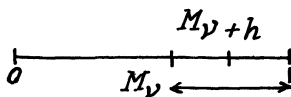
qui intervient en particulier d'une part comme distribution limite de la distribution binomiale $\binom{k}{p} \theta^p (1 - \theta)^{k-p}$ lorsque le paramètre θ de celle-ci est de l'ordre de grandeur $\frac{\lambda}{k}$ et $k \rightarrow \infty$; d'autre part, dans les processus exponentiels.

3. c. - Critère pour reconnaître une mesure bornée

Les mesures bornées sur un ensemble A dénombrable sont définies par des suites de nombres positifs m_i ; mais toute suite de nombres positifs ne définit pas une mesure; il faut non seulement pour cela que le "terme général" m_i tende vers zéro lorsque i tend vers l'infini (contre-exemple classique: $m_i = \frac{1}{i}$; M_i est alors de l'ordre de grandeur de $\log i$) mais que la suite satisfasse au critère plus restrictif dit de Cauchy (dont on remarquera le caractère de finitude):

$$\forall \varepsilon, \exists k(\varepsilon), \forall v > k(\varepsilon), \forall h > 0, \sum_{v}^{v+h} m_i < \varepsilon$$

Ce critère équivaut à dire que le reste R_n de la somme $\sum_{i=1}^n m_i$ tend vers zéro lorsque n tend vers l'infini, c'est une condition nécessaire et suffisante pour que la suite, croissante avec n , des nombres $M_n = \sum_{i=1}^n m_i$ soit bornée, donc



qu'elle converge lorsque $n \rightarrow \infty$. Ce résultat est intuitif (figure ci-contre): choisir $v > k(\varepsilon)$, puis faire tendre h vers l'infini ($v + h = n$); M_{v+h} est

croissante (avec h) et bornée par $M_v + \varepsilon$, donc elle converge: il s'agit toujours de la propriété de compacité de R , qui a déjà été évoquée au Paragraphe 2.e.

On montre par ailleurs (on ne le fera pas ici: technicité de la démonstration, qui est du même type que celle qui vient d'être esquissée) que lorsque la suite $m = (m_1, m_2, \dots, m_i, \dots)$ est à sommes finies M_n bornées ($\forall n$), ou, de façon équivalente, satisfait au critère de Cauchy, c'est bien une mesure en ce sens que pour toute partie X et A , le nombre

$$m(X) = \sum_{i \in X} m_i$$

est bien défini, comme fonction du seul ensemble X :

autrement dit, cette somme ne dépend que de X , et non de l'ordre dans lequel sont pris les éléments de X , lorsqu'on fait les sommations.

3.d.- L'intégrale des fonctions bornées

L'ensemble des mesures positives bornées (ou de probabilité) sur A dénombrable est ainsi bien précisé: c'est l'ensemble des "séries" à termes positifs convergentes. A chaque mesure m correspond une intégrale, I_m , définie sur l'ensemble des fonctions bornées de A dans $\mathbb{R}$. On a:

$$I_m(f) = \sum_{i \in A} m_i f_i$$

Le critère de Cauchy est valable non seulement pour les séries à termes positifs, mais pour toute somme. Il s'écrit alors, pour une somme de termes positifs ou négatifs u_i :

$$\forall \varepsilon, \exists k(\varepsilon), \forall v > k(\varepsilon), \forall h > 0, \left| \sum_v^{v+h} u_i \right| < \varepsilon$$

Ce critère de convergence nous assure de la convergence, lorsque n tend vers l'infini, des sommes partielles:

$$\sum_1^n m_i f_i$$

En effet:

$$\left| \sum_v^{v+p} m_i f_i \right| \leq \sum_v^{v+p} m_i |f_i| \leq (\max |f_i|) \sum_v^{v+p} m_i < K \sum_v^{v+p} m_i$$

Or $\sum_v^{v+p} m_i$ peut être rendu inférieur à $\frac{\varepsilon}{K}$ (ε donné) en choisissant v assez grand.

3.e.- Extension de l'intégrale à des fonctions non bornées

Si, quelque soit la distribution $\underline{m} = (m_1, m_2, \dots, m_i, \dots)$, telle que $\forall i, m_i \geq 0$, et $\sum m_i < \infty$, I_m est définie sur l'ensemble des fonctions bornées, pour chaque distribution $\underline{m}$ donnée, il y a des fonctions non bornées pour lesquelles $I_m(f)$ est encore définie, c'est-à-dire pour lesquelles la série $\sum_{i \in A} m_i f_i$ converge; le cas le plus important est celui où A est l'ensemble des entiers naturels muni de sa structure arithmétique, et où l'on prend pour f soit la fonction identique (on considère l'aléa égal à i avec la probabilité m_i), soit les fonctions puissances (on considère les moments de l'aléa précédent).

Par exemple, pour la distribution de Pascal, on a comme espérance

$$(1-r) \sum_i i r^{i-1} = \frac{1}{1-r}$$

et tous les moments sont définis; de même, pour la distribution de Poisson. L'espérance est:

$$e^{-\lambda} \sum_i i \frac{\lambda^i}{i!} = \lambda$$

et tous les moments sont définis.

Par contre, si l'on prend:

$$m_i = \frac{1}{i^\alpha} \quad (\alpha > 1)$$

ceci définit une mesure bornée, pour laquelle ne sont définis que les moments d'ordre h inférieur à $\alpha - 1$.

N.B. - Pour le vectoriel des fonctions bornées de A (dénombrable) dans $\mathbb{R}$, l'ensemble des covecteurs (formes linéaires sans la condition de positivité) est constitué par l'ensemble des "séries absolument convergentes", c'est-à-dire des suites:

$$\underline{u} = (u_1, u_2, \dots, u_i, \dots)$$

telles que $\sum |u_i|$ converge. L'ensemble de ces $\underline{u}$ constitue le vectoriel dual du vectoriel des fonctions bornées; les mesures positives sont une partie de ce vectoriel (correspondant à l'orthant positif (1) du cas fini).

Sur le vectoriel des fonctions bornées de A dans $\mathbb{R}$, l'intégrale I_u ($I_u(f) = \sum_{i \in A} u_i f_i$), si elle n'est plus monotone (positive) est cependant toujours continue au sens de la topologie de la convergence uniforme.

La condition sur les distributions $\underline{u}$, que $\sum_{i \in A} u_i$ converge, équivaut (critère de Cauchy) à ce que toutes les sommes finies $\sum_{i \in A}^{\nu+p} |u_i|$ soient bornées ($\forall \nu, \forall p$): ces distributions sont souvent dites "à variations bornées".

4. MESURES ET INTEGRATION SUR $\mathbb{R}$

Dans le cas où $A = \mathbb{R}$, l'idée directrice pour la construction des mesures est que l'on veut pouvoir intégrer les "bonnes" fonctions: fonctions monotones, fonctions continues, fonctions en escalier (c'est-à-dire étagées, mais pour lesquelles les ensembles A_y (par. 2.c) sont des unions d'intervalles de $\mathbb{R}$). D'autre part, la structure de groupe abélien ordonné de $\mathbb{R}$ comme ensemble de départ va jouer un rôle essentiel; c'est d'ailleurs par rapport à cette structure que le concept de "bonne" fonction prend son sens.

4.a. Intégrer les bonnes fonctions: mesurer les intervalles

4.a.1.- Il faut des mesures telles que pour toutes les "bonnes" fonctions, et les fonctions étagées leur servant d'approximation, les ensembles

$$A_i = \{x : x \in A, y_i \leq f(x) < y_{i+1}\}$$

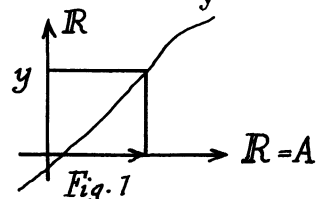
soient mesurables; il est équivalent (additivité) d'exiger que soient mesurables tous les ensembles S_y :

$$S_y = \{x : f(x) < y\}$$

On a en effet: $A_i = S_{y_{i+1}} - S_{y_i}$ (différence ensembliste);

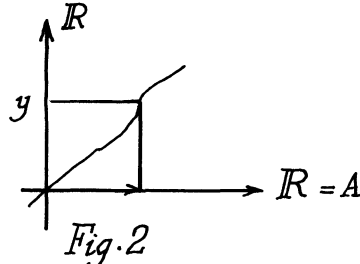
donc: $m(A_i) = m(S_{y_{i+1}}) - m(S_{y_i})$.

4.a.2.- Pour les fonctions f monotones de $\mathbb{R}$ dans $\mathbb{R}$, les ensembles S_y sont les

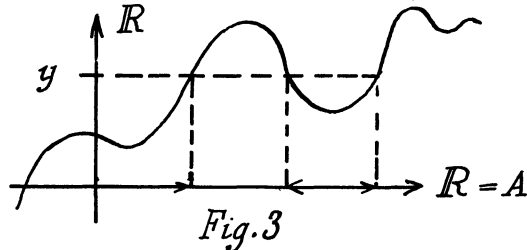


(1) Ensemble des points à coordonnées toutes positives.

demi-droites ouvertes à droite $] - \infty, y [$, dont l'extrémité y est exclue (en tout point de continuité de f , fig. 1), ou fermées à droite $] - \infty, y]$ (pour certaines discontinuités de f , fig. 2). Comme d'autre part, on ne considère pour l'instant que les fonctions bornées, on exclut $-\infty$.



4. a. 3. - Pour toutes les fonctions continues (qui constituent un vectoriel), S_y est une union d'intervalles ouverts, en nombre fini, ou en infinité dénombrable (fig. 3).



4.a.4. - Les fonctions étagées encadrant une fonction continue sont donc des escaliers; pour les escaliers, les S_y sont des unions d'intervalles (ouverts ou fermés: discontinuités).

4.a.5. - Toute mesure pour laquelle sont mesurables les intervalles, c'est-à-dire, on l'a vu, les demi-droites ouvertes $\{x : x < y\} =] - \infty, y [$ de $\mathbb{R}$, permet donc, si elle satisfait à la condition (σ) d'additivité dénombrable d'intégrer:

- les fonctions monotones, donc plus généralement les fonctions "à variations bornées": combinaisons linéaires de fonctions monotones (cf. par. 4).

- les fonctions continues

- les escaliers: donc elle permet le calcul approché des intégrales pour les fonctions à variations bornées et les fonctions continues.

- plus généralement, toutes les fonctions obtenues comme limites uniformes des précédentes; en particulier, les limites uniformes d'escalier, ou fonctions réglées, qui est plus vaste que celui des seules fonctions continues.

4.b. - Fonctions de répartition

4.b.1. - La première chose à définir sur $\mathbb{R}$ est donc la mesure d'une demi-droite $] - \infty, y [$. Posons:

$$M(y) = m(] - \infty, y [)$$

La fonction $M(y)$ doit évidemment être monotone non décroissante:

$\xrightarrow{\quad y \quad y' \quad} \mathbb{R} = A$ 1) $y' \geq y \implies M(y') \geq M(y)$
 puisque $] - \infty, y' [\supset] - \infty, y [$, et qu'une mesure est monotone de l'ordre d'inclusion $\supset$ dans A à l'ordre $\geq$ dans $\mathbb{R}$. Bien entendu, elle est positive:

$$2) \quad \forall y, M(y) \geq 0$$

En outre, on pose par convention ($-\infty$ est exclus des mesures possibles).

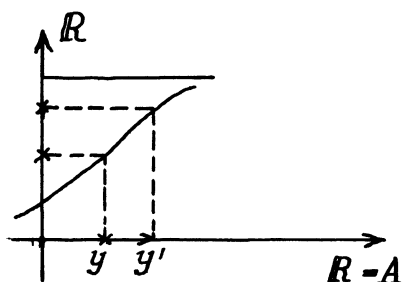
$$3) \quad M(-\infty) = 0$$

Si m doit être une mesure de probabilité, il faut en outre que m soit normée:

$$4) \quad m(A) = M(+\infty) = 1$$

Donc, à chaque mesure m parmi celles qui nous intéressent est associée une fonction $M : \mathbb{R} \longrightarrow \mathbb{R}$ satisfaisant aux conditions 1 à 4 énoncées; M est la fonction de répartition de m .

4.b.2. - Au moyen de la fonction de répartition M , la mesure d'un intervalle



$[y, y' [$ ouvert à droite, fermé à gauche, s'exprime par:

$$m([y, y' [) = M(y') - M(y)$$

Si y' tend en décroissant vers y , l'intervalle $[y, y' [$ tend vers l'ensemble atomique $\{y\}$, qu'il contient quelque soit $y' > y$

On pose par convention:

$$M(y+0) = \lim_{\substack{h > 0 \\ h \rightarrow 0}} M(y+h)$$

Cette limite existe, la suite des nombres $M(y+h)$ étant monotone décroissante, et bornée inférieurement.

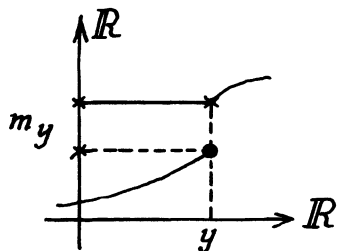
On a donc, pour tout élément $y \in \mathbb{R}$:

$$m(\{y\}) = M(y+0) - M(y)$$

Posons pour abrégé:

$$m(\{y\}) = m_y$$

Deux cas sont alors possibles:



- ou bien, y est un point de continuité de M , ce qui signifie par définition de la continuité que $M(y+0) = M(y)$; alors $m_y = 0$

- ou bien, y est un point de discontinuité de M , et l'on a:

$$M(y+0) - M(y) = m_y > 0$$

Par contre, si y' restant fixe, y tend en croissant vers y' , l'ensemble limite de $[y, y' [$ est vide (y' n'est pas élément de $[y, y' [$); on a donc:

$$\lim_{\substack{h > 0 \\ h \rightarrow 0}} (M(y) - M(y-h)) = m(\emptyset) = 0$$

La fonction de répartition M doit donc être continue à gauche; il faut ajouter une cinquième condition nécessaire pour qu'une fonction M soit fonction de répartition d'une mesure m :

$$5) \text{ Continuité à gauche: } M(y-0) = M(y)$$

Avec la notation $M(y+0)$, les mesure d'intervalles s'expriment par:

$$\text{fermé} \quad : m([y, y']) = M(y'+0) - M(y)$$

$$\text{semi ouvert à gauche: } m(\left] y, y' \right]) = M(y'+0) - M(y+0)$$

$$\text{ouvert} \quad : m(\left] y, y' \right[) = M(y') - M(y+0)$$

4.b.3. - A chaque mesure m est associée une fonction de répartition M , réciproquement, à chaque fonction de répartition M , c'est-à-dire à chaque fonction $M: \mathbb{R} \rightarrow \mathbb{R}$ satisfaisant aux conditions 1 à 5 supra, correspond une mesure m , pour laquelle les intervalles se mesurent comme on vient de l'indiquer; considérons les unions dénombrables d'intervalles:

$$J = J_1 \cup J_2 \cup \dots \cup J_n \cup \dots$$

On peut toujours supposer les J_n disjoints, et l'on a:

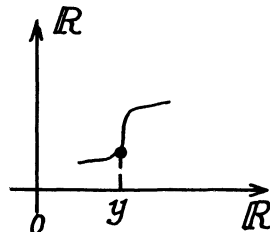
$$m(J) = \sum_n m(J_n)$$

La série du second membre converge, puisqu'elle est à termes positifs et bornée (par $m(\mathbb{R}) = 1$). m ainsi définie à partir d'une fonction de répartition permet donc d'intégrer, on l'a vu, en a), toutes les fonctions utiles; mais en fait, beaucoup d'autres parties de $\mathbb{R}$ sont mesurables pour ces mesures.

4.c. - Fonctions de répartition et mesures totalement discontinues

4.c.1. - Décomposition des fonctions de répartition.

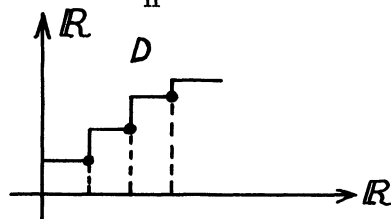
Une fonction de répartition peut présenter, on vient de le voir, des discontinuités, des sauts:



$$M(y+0) - M(y) = m_y > 0$$

D'après la continuité à gauche, la valeur prise par la fonction M au point y est la borne inférieure du saut, et est limite des valeurs prises à gauche de y .

Les sauts d'une fonction de répartition sont en nombre fini, ou en infinité dénombrable; en effet, comme la somme des valeurs m_y de ces sauts est bornée par 1, il y a, pour tout entier n , un nombre fini N de sauts (au plus $n-1$) d'amplitude supérieure $\frac{1}{n}$. Soit $\mathcal{D}$ leur ensemble.



Posons, pour tout $y \in \mathbb{R}$:

$$D(y) = \sum_{\substack{t < y \\ t \in \mathcal{D}}} m_t$$

La fonction D est un escalier (en un sens un peu différent de celui adopté jusqu'ici: l'ensemble des valeurs prises peut être infini dénombrable) monotone croissant et continu à gauche; c'est une fonction de répartition (à l'homothétie

$\frac{1}{D(\infty)}$ près); et l'on a en tout saut:

$$D(y+0) = D(y) + m_y$$

La différence:

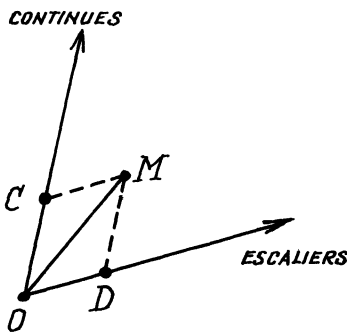
$$C(y) = M(y) - D(y)$$

définit donc de façon évidente une fonction de répartition continue à gauche et à droite.

On a ainsi une décomposition unique de toute fonction de répartition M en une somme de deux fonctions de répartition

$$M = D + C$$

D appartient au vectoriel des escaliers, C à celui des fonctions continues; plus précisément, s'il s'agit des intersections de ces vectoriels avec le $1/2$ cône (orthant positif du cas fini) des fonctions de répartition; ces intersections sont disjointes.



On peut donc étudier séparément les mesures définies par une fonction de répartition en escalier d'une part, continue d'autre part; on verra que c'est des secondes que viennent les difficultés.

4.c.2. - Cas d'une fonction de répartition en escalier.

Soit D une fonction de répartition en escalier de $\mathbb{R}$ dans $\mathbb{R}$, et $\mathcal{D}$ son ensemble dénombrable de sauts: $\mathcal{D}$ est encore appelé le support de D , et de la mesure m_D lui correspondant. Comme la série $\sum_{x \in \mathcal{D}} m_x$ converge, la formule:

$$m_D(X) = \sum_{x \in \mathcal{D} \cap X} m_x$$

définit sans ambiguïté la mesure de toute partie X de $\mathbb{R}$, et l'on vérifie aisément que m_D satisfait à tous les axiomes des mesures; donc à toute fonction de répartition en escalier correspond une mesure définie sur l'ensemble $\mathcal{P}(\mathbb{R})$ des parties de $\mathbb{R}$.

Par contre, toute fonction bornée f est intégrable, et l'on a:

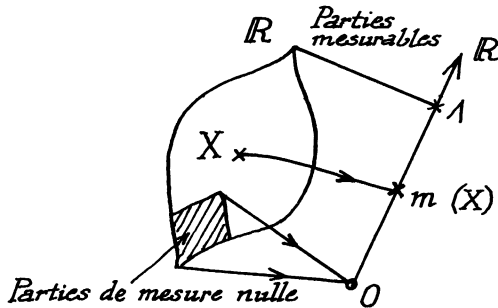
$$I(f) = \sum_{x \in \mathcal{D} \cap X} m_x f(x)$$

On voit que l'on est exactement ramené au cas où A est dénombrable (par. 3) à ceci près que maintenant l'ensemble dénombrable $\mathcal{D}$, support de la mesure, est considérée comme partie de $\mathbb{R}$.

4.c.3. - Parties négligeables.

En particulier, toute partie $X \in \mathcal{R} - \mathcal{D}$ aura une mesure $m_D(X)$ nulle; dans ce cas, la classe un peu mystérieuse des parties de mesure nulle, ou parties négligeables, est extrêmement simple: c'est l'ensemble des parties de $\mathcal{R} - \mathcal{D}$.

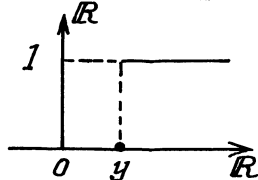
Dans le cas général où M n'est pas un escalier, la mesure m applique



l'ensemble des parties mesurables de $\mathcal{R}$ (on verra que c'est une algèbre de Boole) dans $\mathcal{R}$; l'ensemble des parties négligeables constitue le noyau de cette application: c'est une sous-algèbre, et deux parties X et Y ont même mesure si et seulement si elles diffèrent, au sens ensembliste, d'une partie de mesure nulle.

4.c.4. - Mesures de Dirac.

Un cas particulier important des fonctions de répartition en escalier est



celui où le support $\mathcal{D}$ se réduit à un point: $\mathcal{D} = \{y\}$. En calcul des Probabilités, c'est le cas d'un aléa numérique certain. On a donc:

$D(x) = 0$ si $x \leq y$; $D(x) = 1$ si $x > y$.

La mesure m_D correspondante est appelée mesure de Dirac. L'intégrale correspondante est:

$$I(f) = 1 \times f(y) = f(y)$$

Comme quoi la valeur prise en un point par une fonction numérique est une intégrale, donc une moyenne.

Les autres mesures en escalier peuvent d'ailleurs être considérées comme combinaisons linéaires, de coefficients m_x , des mesures de Dirac correspondant aux éléments x de leur support.

4.d. - Fonctions de répartition et mesures continues - Tribus

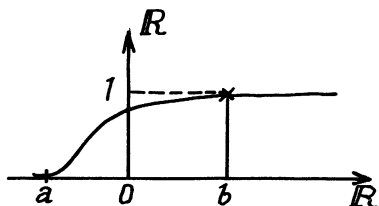
4.d.1. - Le prototype d'une fonction de répartition continue est celui où l'on prend pour M l'application identique:

$$\forall y \in \mathcal{R}, M(y) = y$$

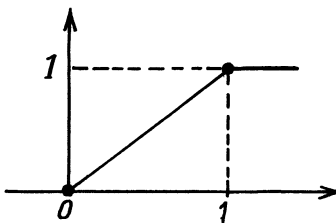
La mesure (non bornée) correspondante est la mesure de Lebesgue; c'est la plus fortement liée à la structure de groupe abélien ordonné de $\mathcal{R}$, puisqu'elle postule l'invariance de la mesure par rapport au groupe des translations.

Il est prudent de commencer par étudier le cas des fonctions de répartition, et des mesures, sur un intervalle de $\mathcal{R}$; la plupart des résultats relatifs au cas général s'obtiennent à partir de celui-ci en faisant tendre vers $+\infty$ et $-\infty$ respectivement les bornes supérieures et inférieures de l'intervalle.

On considère donc des fonctions de répartition continues, nulles si y est inférieur à un nombre a , égales à 1 si y est supérieur à un nombre b .
On peut d'ailleurs supposer $a = 0$ et $b = 1$.



La mesure de Lebesgue correspond alors aux mesures uniformes (équiprobabilité) du Calcul des Probabilités.



4.d.2.- Mesure des ouverts et des fermés

C étant une fonction continue, la mesure d'un intervalle (y, y') , qu'il soit ouvert ou fermé, est toujours:

$$C(y') - C(y)$$

Puisque $C(y + 0) = C(y)$ en tout point y .

On peut donc mesurer à partir de C tout ensemble X union dénombrable d'intervalles (y_i, y'_i) disjoints, par:

$$m_C(X) = \sum_i (C(y'_i) - C(y_i)) \leq 1$$

Le second membre est en effet, une fois de plus, une série à termes positifs et à sommes finies bornées.

En particulier, tout ensemble dit ouvert, c'est-à-dire union dénombrable d'intervalles ouverts disjoints est donc mesurable par rapport à m ; par suite le sont aussi les fermés, c'est-à-dire les complémentaires d'ouverts; on pose, pour un fermé (additivité des mesures):

$$m_C(X) = 1 - m_C(\mathbb{R} - X)$$

4.d.3.- Parties négligeables

D'autre part, sont immédiatement mesurables, et de mesure nulle, parmi les ensembles qui ne sont ni ouverts ni fermés, toutes les parties dénombrables.

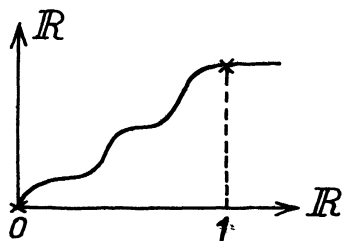
En effet, ε étant donné à l'avance, chaque élément i d'un ensemble $X = \{1, 2, \dots, i, \dots\}$ dénombrable peut être recouvert par un intervalle ouvert $]y_i, y'_i[$ de centre i et tel que (continuité de C):

$$C(y'_i) - C(y_i) < \frac{\varepsilon}{2^i}$$

L'union J de ces intervalles non disjoints constitue un recouvrement ouvert de X , et l'on a (monotonie de m_C):

$$m_C(X) \leq m_C(J) < \sum_i (C(y'_i) - C(y_i)) < \varepsilon \sum_i \frac{1}{2^i} = \varepsilon$$

$m_C(X)$ est ainsi inférieure à tout nombre positif ε ; elle ne peut donc être que nulle.



Bien entendu, la classe des ensembles de mesure nulle comprend en général, pour C donnée, d'autres ensembles que les parties dénombrables: c'est apparent si C reste constante sur certains intervalles. Il semble d'ailleurs que la question de savoir s'il n'y a que les ensembles dénombrables qui soient de mesure nulle pour toute fonction de répartition continue est encore ouverte.

4.d.4.- Extension de la mesure

On a vu que pour les besoins usuels de l'intégration, savoir mesurer les parties qui sont union d'intervalles suffit; si l'on veut néanmoins étendre la mesure à d'autres parties de $\mathbb{R}$, l'idée directrice est celle qui a constamment servi, à savoir celle d'un passage à la limite sur les mesures déjà définies.

Lebesgue, en fait, a procédé comme on vient de le faire pour les ensembles de mesure nulle. Etant donnée une partie X de $\mathbb{R}$, on considère l'ensemble des recouvrements de X par des parties Y ouvertes (donc mesurables) de $\mathbb{R}$. L'ensemble des nombres $m_C(Y)$ est borné inférieurement: il possède donc un infimum, $\mu_C(X)$. Si la mesure de X existe, elle est au plus égale à cet infimum:

$$m_C(X) \leq \mu_C(X).$$

De même, à la partie X' complémentaire de X on peut associer l'infimum $\mu_C(X')$ des mesures des ouverts Y recouvrant X' . Il est clair, d'après la construction, que:

$$\mu_C(X) + \mu_C(X') \geq m(\mathbb{R}) = 1 \quad (\mathbb{R} = X \cup X')$$

Comme l'existence de la mesure de X implique celle de son complémentaire, avec $m_C(X) + m_C(X') = 1$, on a la double inégalité:

$$1 - \mu_C(X') \leq m_C(X) \leq \mu_C(X)$$

X sera par conséquent réputé mesurable (par rapport à C) si et seulement si:

$$\mu_C(X) + \mu_C(X') = 1$$

On a alors: $m_C(X) = \mu_C(X)$

Ce qui est important dans cette définition c'est que de tout recouvrement d'une partie X bornée de $\mathbb{R}$ par des ouverts, on peut extraire un recouvrement fini: c'est là un autre aspect de la compacité de $\mathbb{R}$, déjà évoquée.

Par conséquent, les techniques d'approximation déjà mises en jeu pourront de nouveau jouer à plein; une partie mesurable pourra toujours être approchée par un recouvrement au moyen d'un nombre fini d'intervalles ouverts, c'est-à-dire un escalier (stricto sensu); sa mesure sera donc calculable par approximations).

4.d.5.- Tribus

C'est grâce à la technique des recouvrements finis que l'on peut démontrer (on ne le fera pas ici) que l'ensemble des parties mesurables, par rapport à C donnée, constitue une tribu, (ou σ -algèbre de Boole), c'est-à-dire une sous-algèbre de l'algèbre des parties de $\mathbb{R}$ (ou de l'intervalle considéré dans $\mathbb{R}$) fermée par rapport à l'union et l'intersection dénombrables; en particulier, une union dénombrable de parties mesurables, deux à deux disjointes, est mesurable, et a pour mesure la somme des mesures de ces parties: la propriété (σ) d'additivité dénombrable, qui était requise pour donner un sens à l'intégrale est bien satisfaite.

A chaque mesure m (de fonction de répartition M) correspond ainsi une tribu de parties mesurables par rapport M ; et l'on a vu que c'est la composante continue de M qui est à l'origine de l'introduction de cette notion; on peut en effet montrer que, contrairement à ce qui se passe pour les mesures en escalier, toute partie n'est pas mesurable par rapport à une mesure continue: la tribu est une sous-algèbre propre de $\mathcal{P}(\mathbb{R})$.

A chaque tribu correspond non une, mais une infinité de mesures; elles sont définies par la donnée de la mesure sur une partie génératrice de la tribu.

Les plus triviales des tribus sont les sous-algèbres finies de $\mathcal{P}(\mathbb{R})$; elles sont utiles à manipuler en exercice. Un contre-exemple simple de sous-algèbre qui n'est pas une tribu est fourni par l'ensemble des parties finies ou cofinies (complémentaires de finies) de $\mathbb{R}$.

Mais la plus utile des tribus est bien entendu celle (à laquelle il convient de réserver le nom de tribu borélienne; on dit parfois "corps" de Borel) qui est engendrée par les intervalles de $\mathbb{R}$: on part des intervalles de $\mathbb{R}$, et on considère toutes les parties qu'ils engendrent par complémentarité, union et intersection dénombrables (donc approximables). Toute mesure sur $\mathbb{R}$ admet en effet, d'après la construction qu'on vient de développer cette tribu comme sous-tribu de sa tribu.

4.e. - L'intégration sur $\mathbb{R}$. Intégrale de Lebesgue - Stieltjes

Il n'y a pratiquement plus rien à en dire ici. Pour une mesure m , donnée par sa fonction de répartition M , est intégrable toute fonction f bornée telle que pour tout y l'ensemble $S_y = \{x : f(x) < y\}$ appartient à la tribu associée à M (c'est-à-dire à la composante continue C de M). On peut d'ailleurs décomposer $I_M(f)$ en:

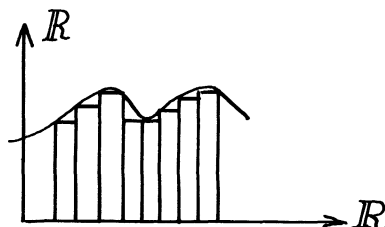
$$I_M(f) = I_D(f) + I_C(f)$$

$I_D(f) = \sum_{x \in \mathcal{D}} m_x f(x)$ est une somme.

Quant à $I_C(f)$, c'est à son propos qu'on introduit traditionnellement la notation $\int_{\mathbb{R}} f dC$.

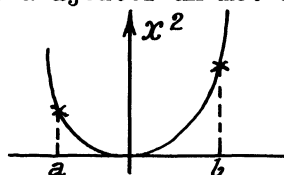
$I_M(f)$ est l'intégrale de Lebesgue - Stieltjes, d'opérateur (on dit encore de noyau) M de la fonction f .

Comme on l'a dit supra, par. 4.a., sont intégrables, parce que pour elles les ensembles S_y appartiennent toujours à la tribu borélienne, toutes les fonctions usuelles.



La composante $I_C(f)$ de l'intégrale est en général calculée par approximations, au moyen d'escaliers (c'est à ce stade qu'on peut introduire la fameuse figure ci-contre).

Il reste à ajouter un mot de l'intégration des fonctions non bornées; consi-



dérons par exemple les fonctions qui sont bornées dans tout intervalle fini (a, b) : cas des fonctions puissances, qui servent pour les calculs de moments, en Calcul des Probabilités.

En ce cas on les intègre (s'il se peut) sur (a, b) , où elles sont bornées; puis on cherche si l'intégrale $I_{(a,b)}(f)$ ainsi définie possède une limite lorsque $a \rightarrow -\infty$ et $b \rightarrow +\infty$; si oui, cette limite est par définition l'intégrale sur $\mathbb{R}$ de la fonction f .

5. - DENSITES

On sait que le problème de l'intégration est lié à celui de la dérivation; les premiers contacts avec la notion d'intégrale se font encore parfois par le biais des fonctions primitives.

Ici, c'est à partir de l'intégrale que vont être introduites densité et dérivée; ceci amènera à voir que la décomposition des fonctions de répartition en une fonction de sauts et une fonction continue doit encore être affinée: la composante continue est elle-même décomposable en une partie non dérivable, et une partie dérivable, ou absolument continue.

5.a. - On peut définir des mesures et des fonctions de répartition au moyen d'une intégrale

- M est une fonction de répartition sur $\mathbb{R}$, m est la mesure correspondante, et T_m la tribu des parties mesurables par rapport à m .

Pour chaque fonction intégrable (par rapport à m), et chaque partie X de $\mathbb{R}$ mesurable ($X \in T_m$), on peut calculer l'intégrale:

$$\int_X \gamma \, dM$$

Si M et γ sont données, on définit ainsi une fonction $g : T_m \rightarrow \mathbb{R}$

$$g(X) = \int_X \gamma \, dM$$

- g est une mesure (si X et Y sont disjointes, $g(X \cup Y) = g(X) + g(Y)$); si en outre γ est positive ($\forall t \in \mathbb{R}, \gamma(t) \geq 0$) on a: $\forall X \in T_m, g(X) \geq 0$. g est une mesure positive; c'est une mesure de probabilité si $g(\mathbb{R})$ est finie.

La f.r. de g est d'ailleurs:

$$G(x) = \int_{-\infty}^x \gamma \, dM$$

On vérifiera ultérieurement que G est continue à gauche.

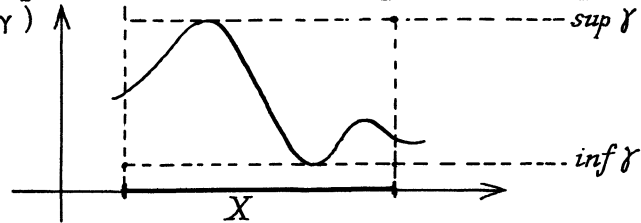
- Ainsi, à chaque fonction positive et intégrable γ, m fait correspondre une mesure g_γ bien définie sur la tribu T_m de m .

Mais, comme on va le voir, deux fonctions distinctes γ et γ' , peuvent avoir même mesure g image; et d'autre part toute mesure ne peut être obtenue par ce procédé: la correspondance n'est ni injective, ni surjective.

L'outil dans l'étude de cette correspondance, c'est l'inégalité de la moyenne; le concept nouveau qu'elle introduit, celui de la densité.

5.b. - Inégalité de la moyenne

Supposons γ bornée; sur chaque partie X de $\mathcal{R}$ elle possède un supremum ($\sup_X \gamma$) et un infimum ($\inf_X \gamma$)



L'intégrale étant monotone, on a:

$$m(X) \inf_X \gamma \leq \int_X \gamma \, dM \leq m(X) \sup_X \gamma$$

Soit encore:

$$m(X) \inf_X \gamma \leq g(X) \leq m(X) \sup_X \gamma$$

C'est l'inégalité (l'une des inégalités) de la moyenne.

Conséquences immédiates:

- toute partie X négligeable par rapport à m ($m(X) = 0$), l'est aussi par rapport à g .
- toute partie X mesurable par rapport à m et sur laquelle γ est nulle, est négligeable par rapport à g .
- Si donc, on ajoute à γ une fonction γ' bornée nulle sauf sur une partie négligeable par rapport à m , on a :

$$g(X) = \int_X (\gamma + \gamma') \, dM = \int_X \gamma \, dM$$

γ et $\gamma + \gamma'$ définissent la même mesure g .

Pour toute partie X de mesure non nulle par rapport à m , l'inégalité de la moyenne s'écrit encore:

$$\inf_X \gamma \leq \frac{g(X)}{m(X)} \leq \sup_X \gamma$$

$\frac{g(X)}{m(X)}$ est un rapport de mesures; on peut l'appeler densité de la mesure g par rapport à la mesure m sur l'ensemble X .

5.c. - Densité locale et dérivées

5.c.1. - Si en particulier $X = \{x\}$, on a :

$$\sup_X \gamma = \inf_X \gamma = \gamma(x)$$

L'inégalité de la moyenne se réduit à :

$$g_x = \gamma(x) m_x$$

où:

$$g_x = g(\{x\}) \quad m_x = m(\{x\})$$

Donc, en tout point x de continuité pour M ($m_x = 0$), on a $g_x = 0$, et G est continue.

- En un point de discontinuité pour M , on a $m_x > 0$, et la densité de p par rapport à m est bien définie au point x , et vaut:

$$\gamma(x) = \frac{g_x}{m_x}$$

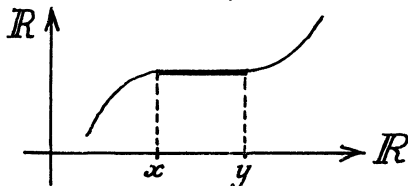
Peut-on définir une densité locale (en un point) là où M est continue ?

5.c.2. - Si $X = \{x \leq t < y\}$, l'inégalité de la moyenne s'écrit:

$$(M(y) - M(x)) \inf_{[x,y]} \gamma \leq G(y) - G(x) < (M(y) - M(x)) \sup_{[x,y]} \gamma$$

D'où:

- en faisant tendre x vers y , continuité à gauche de G .
- Si M est constante sur un intervalle, G l'est aussi



Lorsqu'on décompose M en $D + C$, où D est une fonction de sauts, et C une f.r. continue (cette décomposition est unique), on a:

$$G = G_1 + G_2$$

Avec:

$$G_1(x) = \int_{-\infty}^x \gamma \, dD = \sum_{y < x} \gamma_y m_y$$

$$G_2(x) = \int_{-\infty}^x \gamma \, dC$$

G_1 est la composante totalement discontinue de G , G_2 sa composante continue.

Dans la suite, on peut supposer M continue partout, et se borner au cas des points x où $M(y) - M(x) > 0$ si $y > x$ (leur ensemble est le support de m).

5.c.3. - Pour un tel point, on a, $\forall y$, la double inégalité:

$$\inf_{(x,y)} \gamma \leq \frac{G(y) - G(x)}{M(y) - M(x)} < \sup_{(x,y)} \gamma$$

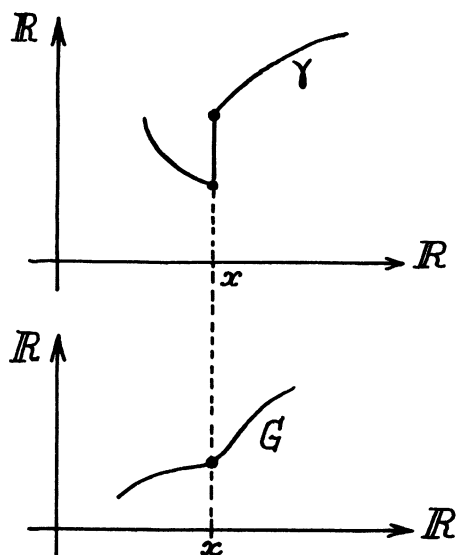
Si ce point x est point de continuité pour γ , il en résulte que le rapport de densité a pour limite $\gamma(x)$ lorsque y tend vers x : G est dérivable par rapport à m , et sa dérivée est $\gamma(x)$; G est une primitive de γ . La densité (en un point) de g par rapport à m est ainsi bien définie.

Dans le cas particulier où pour M on prend l'application identique ($M(x) = x$), on retrouve la définition usuelle de la dérivée.

La densité locale se trouve ainsi bien définie:

1) En tout point de discontinuité de M .

2) En tout point de continuité pour M et pour γ . Dans tous ces cas, elle vaut $\gamma(x)$ au point x .



Pour les points de continuité de M qui sont discontinuité de γ , on peut encore définir non une densité, mais des bornes entre lesquelles on peut la choisir arbitrairement.

Le seul cas que l'on rencontre parfois en Calcul des Probabilités est celui où la fonction γ possède en un point x une discontinuité telle que $\gamma(x-0)$ et $\gamma(x+0)$ soient bien définis.

En ce cas, G a une dérivée à droite, qui est $\gamma(x+0)$; et une dérivée à gauche, qui est $\gamma(x-0)$. Ce type de discontinuité pour γ se traduit par un point anguleux pour le graphique de G .

5.d. - Absolue continuité et fonctions de répartition "singulières"

De l'inégalité de la moyenne, on déduit encore que pour toute partie X mesurable par rapport à m , et pour γ bornée:

$$0 \leq g(X) \leq K m(X)$$

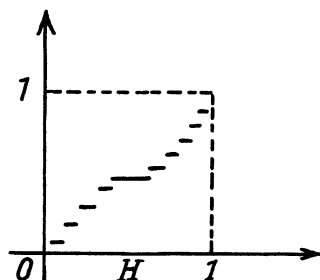
où K est une borne de $|\gamma|$

Si en particulier $m(X) < \varepsilon$, alors $g(X) < K \varepsilon = \eta$

$g(X)$ tend vers zéro avec $m(X)$: c'est l'absolue continuité de g par rapport à m .

Ainsi, toute fonction de répartition G obtenue par intégration par rapport à M d'une fonction bornée γ , et admettant γ pour densité, est absolument continue par rapport à M .

Si l'on montre une fonction de répartition H continue, mais non absolument continue par rapport à une mesure m convenablement choisie, on aura prouvé, comme on l'annonçait en 5.a., que toute f.r. n'est pas définissable comme intégrale d'une densité, comme primitive.



L'exemple classique consiste à prendre pour m la mesure uniforme ($m(x) = x$), et pour H la fonction inverse d'une f.r. totalement discontinue dont le support est partout dense sur l'intervalle $(0,1)$. Par exemple: on applique en sens inverse le procédé "Achille et la Tortue" à l'ensemble des rationnels compris entre 0 et 1. H est continue, car si H admettait une discontinuité, cela impliquerait qu'il existe un intervalle de $(0,1)$ qui ne contienne aucun rationnel; mais H n'est pas absolument continue par rapport à m , puisque son support a pour mesure 0 selon m , et 1 selon H .

28.

Une telle fonction de répartition est dite singulière.

Si d'une f. r. continue on retranche sa composante singulière, la fonction restante est une f. r. absolument continue.

En définitive, toute fonction de répartition M , et par conséquent toute mesure, est décomposable en une somme de f. r.:

$$M = D + C_1 + C_3$$

D est une fonction de sauts,

C_1 est absolument continue (et a une densité) par rapport à la mesure uniforme.

C_3 est continue singulière par rapport à la mesure uniforme.