INSTITUT DE FRANCE
Académie des sciences

Histoire des sciences / Textes scientifiques fondateurs – Avril 2013

À propos de la note de Henri Lebesgue Sur une généralisation de l'intégrale définie

par Jean-Pierre Kahane, membre de l'Académie des sciences

Ce qui suit est une note aux Comptes rendus, de 1901, intitulée « Sur une généralisation de l'intégrale définie »[1]. Vue d'aujourd'hui, c'est la définition de l'intégrale et de la mesure de Lebesgue, deux notions de portée immense en mathématiques et au delà. De plus, c'est un exposé admirablement clair de ces deux notions. Comme la plupart des grandes nouveautés, elle n'a pas été reconnue tout de suite. Le *Jahrbuch über die Forschritte der Mathematik*, le livre allemand qui recense et analyse annuellement toutes les publications mathématiques, avait consacré beaucoup de place, trois pages au total, aux notes précédentes de Lebesgue, qui concernaient les surfaces. Il accorde trois lignes à la note de 1901.

Le centenaire de la note a fait l'objet de nombreux rapports et commentaires. A l'initiative de Gustave Choquet, la note a été rééditée dans les Comptes rendus de l'année 2000, avec des commentaires de Jean-Michel Bony, Gustave Choquet et Gilles Lebeau[2]; j'ai présenté un rapport à l'Académie en mars 2001, reproduit dans la Gazette des mathématiciens et traduit dans plusieurs langues[3]; l'Ecole normale supérieure de Lyon, sous l'impulsion d'Etienne Ghys, a organisé un colloque dont la suite, « Autour du centenaire Lebesgue », a réuni dans un numéro de « Panoramas et synthèses » des études de Gustave Choquet, Thierry de Pauw, Pierre de la Harpe, Jean-Pierre Kahane, Hervé Pajot et Bruno Sevenec, sur l'histoire et les prolongements de la note[4]. Une bonne partie des mathématiques du vingtième siècle s'y rattache : l'analyse de Fourier et l'analyse fonctionnelle, la théorie des probabilités, la théorie géométrique de la mesure, et toutes les variations autour de la mesure et de l'intégration liées aux groupes, aux variables réelles, à la physique, et par là à presque tout le champ des sciences mathématiques.

Je me limiterai ici à quelques points d'histoire et aux réflexions qu'ils inspirent.

---

[1] H.Lebesgue, Sur une généralisation de l'intégrale définie, C.R.Acad.Sci. Pairs 132 (1901), pp. 1025-1027

[2] H.Lebesgue, Sur une généralisation de l'intégrale définie, C.R.Acad.Sci. Pairs 132 (1901), pp. 1025-1027, reproduit dans C.R.Acad.Sci.Paris série 1, mathématique 332 (2001), avec des commentaires de J.-M.Bony, G.Choquet et G.Lebeau.

[3] J.-P.Kahane, Naissance et postérité de l'intégrale de Lebesgue, Gazette des mathématiciens n°89, Juillet 2001, pp.5-20

[4] Autour du centenaire Lebesgue, G. Choquet, T. De Paauw, P. de la Harpe, J.-P. Kahane, H.Pajot, B. Sévennec, Panoramas et synthèses n°18 (2004), Société mathématique de France

Depuis Newton et Leibniz, le calcul différentiel et intégral repose sur deux procédés qui s'appliquent aux fonctions usuelles et sont inverses l'un de l'autre : la dérivation et l'intégration. Par exemple, la dérivée de la fonction sinus est la fonction cosinus, et l'intégration de la fonction cosinus à partir de zéro redonne la fonction sinus. Au cours du 18^{e} siècle ce calcul s'est élargi aux fonctions de plusieurs variables, il a permis l'éclosion de la mécanique rationnelle et de la physique mathématique. Cependant, dès qu'avec l'Alembert en 1749 est apparue l'équation des cordes vibrantes, qui est une équation aux dérivées partielles, s'est posée la question de la validité des solutions : peut-on admettre comme solution une fonction non dérivable ? une fonction discontinue ? La célèbre controverse des cordes vibrantes, qui a opposé Euler et d'Alembert, part de là.

Les mathématiciens ont donc été amenés à élucider et formaliser ce que sont les dérivées, les fonctions dérivables, les limites, la continuité. Le Cours d'analyse de Cauchy à l'Ecole Polytechnique, au début du 19^{e} siècle, contient sur ces notions les définitions qui sont encore de règle aujourd'hui.

Qu'en est-il de l'intégration et des fonctions intégrables ? L'intégration des fonctions continues semblait aller de soi. Celle des fonctions continues par morceaux semblait d'ensuivre. Mais au delà ? Dirichlet, en 1829, osa produire un exemple de fonction définie sur un intervalle et non intégrable : c'est une fonction prenant deux valeurs différentes, l'une sur les rationnels et l'autre sur les irrationnels. La raison qu'il donnait est qu'elle n'est continue sur aucun intervalle.

Le sujet resta stagnant jusqu'à la thèse de Riemann sur les séries trigonométriques, qui fut soutenue en 1854 mais qui ne fut publiée qu'en 1867, après sa mort. En dix lignes, Riemann répond à la question : que doit-on entendre par l'intégrale définie d'une fonction entre deux valeurs a et b de la variable ? Riemann partage l'intervalle (a, b) en petits morceaux, choisit un point dans chaque morceau, multiplie la valeur de la fonction en ce point par la longueur du morceau, fait la somme, et regarde si cette somme a une limite, indépendante du partage et du choix des points, quand la taille des petits morceaux tend vers zéro. Si la somme a une limite, c'est l'intégrale. Après avoir donné la définition, Riemann se préoccupe des fonctions intégrables, et il en donne une caractérisation, que plus tard Lebesgue résumera sous la forme : l'ensemble de leurs points de discontinuité est de mesure nulle. Il y a donc des fonctions intégrables au sens de Riemann qui ne sont continues sur aucun intervalle. Reste que la fonction de Dirichlet est non intégrable au sens de Riemann.

Aussitôt connue, la définition de Riemann devint classique : on tenait la définition d'une fonction intégrable comme on avait déjà celle d'une fonction dérivable. L'intégrale, comme la dérivée, avait le statut d'une notion mathématique clairement définie.

Restait une question : l'intégration reste-t-elle l'opération inverse de la dérivation ? Est-il vrai qu'une fonction dérivée est intégrable au sens de Riemann ? La réponse est négative : l'intégration au sens de Riemann ne permet pas de retrouver toutes les fonctions primitives, celles qui par dérivation donnent la fonction donnée.

Pour justifier sa « généralisation », Lebesgue part de là. L'intégration au sens qu'il va indiquer permet de trouver les primitives de toutes les fonctions dérivées bornées. Lebesgue s'attache donc à l'intégration des fonctions bornées : elles sont définies sur un intervalle (a,b) et prennent leurs valeurs dans un intervalle (m,M). Au lieu de découper (a,b), c'est (m,M) qu'il découpe en morceaux. Il s'impose alors de définir la mesure de l'ensemble sur lequel la fonction prend ses valeurs dans le morceau considéré. C'est la mesure de Lebesgue. Pour respecter l'usage, Lebesgue ne parle pas de fonctions intégrables mais de fonctions sommables, et la terminologie s'est maintenue longtemps, surtout en France où, pour des raisons diverses, l'intégrale de Lebesgue s'est imposée moins vite qu'à l'étranger.

En France, Lebesgue a fait connaître son intégrale par le cours Peccot, qu'il a assuré deux fois[5]. A cette occasion il a étendu la définition à des fonctions non bornées, ce qui s'est avéré essentiel dans la suite. Le rédacteur du second cours Peccot, sur les séries trigonométriques, a été Pierre Fatou, dont la thèse a été la première application de l'intégrale et de la mesure de Lebesgue à la théorie des fonctions analytiques. A l'étranger, l'intégrale de Lebesgue s'est répandue comme un feu de paille : en Angleterre, en Autriche, en Hongrie, en Allemagne, en Belgique, en Russie, en Pologne. Les Anglais Hobson et Hardy l'ont adoptée pour les séries de Fourier, et ont décidé d'appeler intégrables (en renonçant au terme de sommable) les fonctions intégrables au sens de Lebesgue. L'Autrichien Ernst Fischer et le Hongrois Frédéric Riesz ont établi que la transformation de Fourier est, suivant une expression de F. Riesz, « un billet aller et retour permanent » entre les espaces que nous appelons aujourd'hui l^2 et L^2 (constitués respectivement des suites et des fonctions définies sur un intervalle dont les carrés sont sommables). La clé est que l'espace des fonctions de carrés sommables est complet, au sens que la condition nécessaire de convergence de Cauchy y est aussi suffisante. Mais la clé n'a été forgée qu'à la suite des travaux de Riesz et de Fischer de 1907; au départ, c'était un lemme assez difficile à exprimer. C'est Hardy qui a appelé L^p les espaces de fonctions dont la p-ième puissance est intégrable (=sommable), et l'expression « L^p est complet » apparaît à ma connaissance pour la première fois dans l'ouvrage de Stefan Banach « Théorie des opérations linéaires », de 1932, qui a joué un rôle fondateur dans l'analyse fonctionnelle. En Belgique, Charles de la Vallée-Poussin a immédiatement adopté et enseigné l'intégrale de Lebesgue. L'Autrichien Hausdorff et le Polonais Saks ont également exposé l'intégrale et la mesure de Lebesgue comme base de leurs travaux. Ce sont pourtant les Russes Lusin et Souslin qui, corrigeant une vue inexacte de Lebesgue sur les ensembles boréliens, ont complété de la façon la plus originale la théorie des ensembles mesurables. D'un autre côté la mesure de Lebesgue a joué un rôle décisif en probabilités : le Polonais Hugo Steinhaus a considéré l'intervalle (0,1) de la droite réelle, munie de la mesure de Lebesgue, comme modèle pour toute la théorie ; les évènements sont les parties mesurables de l'intervalle, les variables aléatoires sont les fonctions mesurables, les espérances sont les intégrales de Lebesgue. L'Américain Norbert Wiener, au lieu de se restreindre à l'intervalle (0,1), a imité la construction de la mesure de Lebesgue en construisant une mesure sur l'espace des fonctions continues issues d'un point : c'est la mesure de Wiener, et c'est ainsi qu'il introduit « the fundamental random function » que Paul Lévy, un peu plus tard, appellera simplement « le mouvement brownien », le mouvement brownien des mathématiciens. L'axiomatique des probabilités, établie par Kolmogorov en 1933, achève ce mouvement : dans l'héritage de la note de Lebesgue, la théorie des probabilités, sous sa forme actuelle, occupe une place de premier plan.

[5] Les cours Peccot, donnés au Collège de France et accompagnès d'une bourse substantielle, ont été attribués au départ à Borel, Baire et Lebesgue. Ils ont joué un rôle très important à l'époque, et ils continuent jusqu'à présent à valoriser chaque année les travaux d'un mathématicien. Lebesgue a assuré le cours Peccot en 1904 et 1906

Dans les années 1950, l'intégrale de Lebesgue était enseignée dans le monde entier, avec une exception notable : la France. On pouvait être agrégé de mathématiques en ignorant tout de la mesure et de l'intégrale de Lebesgue.
Les causes en sont multiples : il y avait un retard à l'enseignement des sciences
qu'on a peine à imaginer aujourd'hui. Pour une part Lebesgue est responsable du retard concernant son intégrale : dans ses cours au Collège de France, il a toujours choisi d'autres sujets. Outre Fatou, son seul continuateur direct a été Arnaud Denjoy. Lebesgue n'avait résolu le problème de la primitive, au moyen de son intégrale, que pour des fonctions dérivées bornées. Denjoy a mis au point une autre forme d'intégration, qu'il a appelé totalisation, permettant de trouver la primitive de toute fonction dérivée.

Je n'irai pas plus loin dans l'histoire du sujet, mais l'intégration est un sujet mouvant et toujours actuel. On vient de parler de l'intégration selon Cauchy, selon Riemann, selon Denjoy, selon Wiener, et on pourrait continuer longtemps. C'est assez pour voir qu'il n'y a pas de notion d'intégrale formalisable comme celle de dérivée. La fonction de Dirichlet, prenant deux valeurs différentes, l'une sur les rationnels et l'autre sur les irrationnels, est non intégrable au sens de Riemann comme au sens de Cauchy, mais elle est intégrable au sens de Lebesgue. On ne peut pas parler de fonction intégrable sans préciser dans quel sens on l'entend.

On a donc beaucoup de choix pour enseigner l'intégrale, à tous les niveaux. Le choix de l'intégrale de Lebesgue s'impose en physique quantique. Dans d'autres secteurs, l'intégrale de Riemann et ses avatars reste bien adaptée. Dans d'autres au contraire, les distributions de Schwartz sont les bons outils. Le terme même d'intégration présente deux faces très différentes. D'une part, suivant une formule de Youri Manin, intégrer une fonction, c'est trouver la quantité de quelque chose dans un domaine. Et d'autre part, intégrer une équation différentielle, c'est une sorte de généralisation de la recherche de la primitive. Dans l'enseignement de départ, on peut selon le cas et l'époque privilégier le calcul des aires, ou la solution de l'équation $y' = f(x)$ c'est à dire le calcul des primitives. A un niveau plus élevé, l'intégrale d'Itô est bien adaptée à la résolution des équations différentielles stochastiques, et on en fait grand usage dans beaucoup d'applications des probabilités. Ainsi, en brisant le monopole qu'avait acquis pour un temps l'intégrale de Riemann, Lebesgue a révélé que l'intégrale est une notion protéiforme dont il serait illusoire de donner une définition unique. Reste que de grands pans des mathématiques se sont créés ou recréés à la suite de la parution de la note que voici.

ANALYSE MATHÉMATIQUE. — *Sur une généralisation de l'intégrale définie.* Note de M. H. **Lebesgue**, présentée par M. Picard.

« Dans le cas des fonctions continues, il y a identité entre les notions d'intégrale et de fonction primitive. Riemann a défini l'intégrale de certaines fonctions discontinues, mais toutes les fonctions dérivées ne sont pas intégrables, au sens de Riemann. Le problème de la recherche des fonctions primitives n'est donc pas résolu par l'intégration, et l'on peut désirer une définition de l'intégrale comprenant comme cas particulier celle de Riemann et permettant de résoudre le problème des fonctions primitives [1].

» Pour définir l'intégrale d'une fonction continue croissante

$$\gamma(x) \, (a \leqq x \leqq b),$$

on divise l'intervalle (a, b) en intervalles partiels et l'on fait la somme des quantités obtenues en multipliant la longueur de chaque intervalle partiel

[1] Ces deux conditions imposées *a priori* à toute généralisation de l'intégrale sont évidemment compatibles, car toute fonction dérivée intégrable, au sens de Riemann, a pour intégrale une de ses fonctions primitives.

par l'une des valeurs de y quand x est dans cet intervalle. Si x est dans l'intervalle (a_i, a_{i+1}), y varie entre certaines limites m_i, m_{i+1}, et réciproquement si y est entre m_i et m_{i+1}, x est entre a_i et a_{i+1}. De sorte qu'au lieu de se donner la division de la variation de x, c'est-à-dire de se donner les nombres a_i, on aurait pu se donner la division de la variation de y, c'est-à-dire les nombres m_i. De là deux manières de généraliser la notion d'intégrale. On sait que la première (se donner les a_i) conduit à la définition donnée par Riemann et aux définitions des intégrales par excès et par défaut données par M. Darboux. Voyons la seconde.

» Soit la fonction y comprise entre m et M. Donnons-nous

$$m = m_0 < m_1 < m_2 < \ldots < m_{p-1} < \mathrm{M} = m_p;$$

$y = m$, quand x fait partie d'un ensemble E_0; $m_{i-1} < y \leqq m_i$ quand x fait partie d'un ensemble E_i.

» Nous définirons plus loin les mesures λ_0, λ_i de ces ensembles. Considérons l'une ou l'autre des deux sommes

$$m_0\lambda_0 + \Sigma m_i\lambda_i; \quad m_0\lambda_0 + \Sigma m_{i-1}\lambda_i;$$

si, quand l'écart maximum entre deux m_i consécutifs tend vers zéro, ces sommes tendent vers une même limite indépendante des m_i choisis, cette limite sera par définition l'intégrale des y qui sera dite *intégrable*.

» Considérons un ensemble de points de (a, b); on peut d'une infinité de manières enfermer ces points dans une infinité dénombrable d'intervalles; la limite inférieure de la somme des longueurs de ces intervalles est la mesure de l'ensemble. Un ensemble E est dit *mesurable* si sa mesure augmentée de celle de l'ensemble des points ne faisant pas partie de E donne la mesure de (a, b) (1). Voici deux propriétés de ces ensembles : une infinité d'ensembles mesurables E_i étant donnée, l'ensemble des points qui font partie de l'un au moins d'entre eux est mesurable; si les E_i n'ont deux à deux aucun point commun, la mesure de l'ensemble obtenu est la somme des mesures E_i. L'ensemble des points communs à tous les E_i est mesurable.

» Il est naturel de considérer d'abord les fonctions telles que les ensembles qui figurent dans la définition de l'intégrale soient mesurables. On trouve que : *si une fonction limitée supérieurement en valeur absolue est*

(1) Si l'on ajoute à ces ensembles des ensembles de mesures nulles convenablement choisis, on a des ensembles mesurables au sens de M. Borel (*Leçons sur la théorie des fonctions*).

telle que, quels que soient A *et* B, *l'ensemble des valeurs de x pour lesquelles on a* $A < y \leqq B$ *est mesurable, elle est intégrable* par le procédé indiqué. Une telle fonction sera dite *sommable*. L'intégrale d'une fonction sommable est comprise entre l'intégrale par défaut et l'intégrale par excès. De sorte que, *si une fonction intégrable au sens de Riemann est sommable, l'intégrale est la même avec les deux définitions*. Or, *toute fonction intégrable au sens de Riemann est sommable*, car l'ensemble de ses points de discontinuité est de mesure nulle, et l'on peut démontrer que si, en faisant abstraction d'un ensemble de valeurs de x de mesure nulle, il reste un ensemble en chaque point duquel une fonction est continue, cette fonction est sommable. Cette propriété permet de former immédiatement des fonctions non intégrables au sens de Riemann et cependant sommables. Soient $f(x)$ et $\varphi(x)$ deux fonctions continues, $\varphi(x)$ n'étant pas toujours nulle; une fonction qui ne diffère de $f(x)$ qu'aux points d'un ensemble de mesure nulle partout dense et qui en ces points est égale à $f(x) + \varphi(x)$ est sommable sans être intégrable au sens de Riemann. *Exemple :* La fonction égale à 0 si x irrationnel, égale à 1 si x rationnel. Le procédé de formation qui précède montre que l'ensemble des fonctions sommables a une puissance supérieure au continu. Voici deux propriétés des fonctions de cet ensemble.

» 1° *Si f et φ sont sommables, $f + \varphi$ et $f\varphi$ le sont* et l'intégrale de $f + \varphi$ est la somme des intégrales de f et de φ.

» 2° *Si une suite de fonctions sommables a une limite, c'est une fonction sommable.*

» L'ensemble des fonctions sommables contient évidemment $y = k$ et $y = x$; donc, d'après 1°, il contient tous les polynomes et comme, d'après 2°, il contient toutes ses limites, il contient donc toutes les fonctions continues, toutes les limites de fonctions continues, c'est-à-dire les fonctions de première classe (voir Baire, *Annali di Matematica*, 1899), il contient toutes celles de seconde classe, etc.

» En particulier, *toute fonction dérivée, limitee supérieurement en valeur absolue*, étant de première classe, *est sommable* et l'on peut démontrer que *son intégrale, considérée comme fonction de sa limite supérieure, est une de ses fonctions primitives*.

» Voici maintenant une application géométrique : si $|f'|$, $|\varphi'|$, $|\psi'|$ sont limitées supérieurement, la courbe

$$x = f(t), \qquad y = \varphi(t), \qquad z = \psi(t)$$

a pour longueur l'intégrale de $\sqrt{f'^2 + \varphi'^2 + \psi'^2}$. Si $\varphi = \psi = 0$, on a la variation totale de la fonction f à variation limitée. Dans le cas où f', φ', ψ' n'existent pas, on peut obtenir un théorème presque identique en remplaçant les dérivées par les nombres dérivés de Dini. »

ANALYSE MATHÉMATIQUE. — *Sur les intégrales analytiques des équations*