TRIBUNE LIBRE

Naissance et postérité de l'intégrale de Lebesgue

Jean-Pierre Kahane (*Académie des Sciences*)

Il y a quelques mois, Gustave Choquet a proposé à la section de Mathématiques de célébrer le centenaire de l'intégrale de Lebesgue. L'intégrale de Lebesgue, en effet, a profondément marqué le développement des mathématiques au 20ème siècle, et elle est apparue il y a 100 ans, le 29 avril 1901, sous la forme d'une note aux Comptes rendus intitulée : « Sur une généralisation de l'intégrale définie ».

La première suite donnée à la proposition de Gustave Choquet est la note commémorative, signée de Jean-Michel Bony, Gustave Choquet et Gilles Lebeau, qui vient de paraître aux Comptes rendus.

La seconde est cette intervention devant l'Académie[1,2].

Mon programme est de vous parler de la naissance et de la postérité de l'intégrale de Lebesgue. La naissance, je vous l'ai dit, c'est une note aux Comptes rendus, dont nous allons voir reproduite sur l'écran la première page. C'est l'occasion de vous dire un mot des Comptes rendus à cette époque. Puis je vous parlerai un peu de Lebesgue et de ses contemporains, Émile Borel et René Baire. Puis nous ferons des mathématiques. J'essaierai de vous expliquer en quoi l'intégrale de Lebesgue diffère de celle de Riemann, et sa relation avec trois grandes questions qui ont intéressé Lebesgue : le calcul des primitives, la mesure des aires et les séries trigonométriques. Ce sera tout pour la naissance.

Pour la postérité, le choix est immense. L'intégrale et la mesure de Lebesgue ont joué un rôle déterminant dans deux secteurs essentiels des mathématiques du 20ème siècle, l'analyse fonctionnelle et la théorie des probabilités. Je me bornerai à l'évoquer sous deux angles : le théorème de Riesz-Fischer de 1907, et la théorie du mouvement brownien par Norbert Wiener dans les années 1920–1930. Et je conclurai par quelques remarques sur l'évolution des notions de mesure et d'intégrale à travers les âges, et jusqu'à présent.

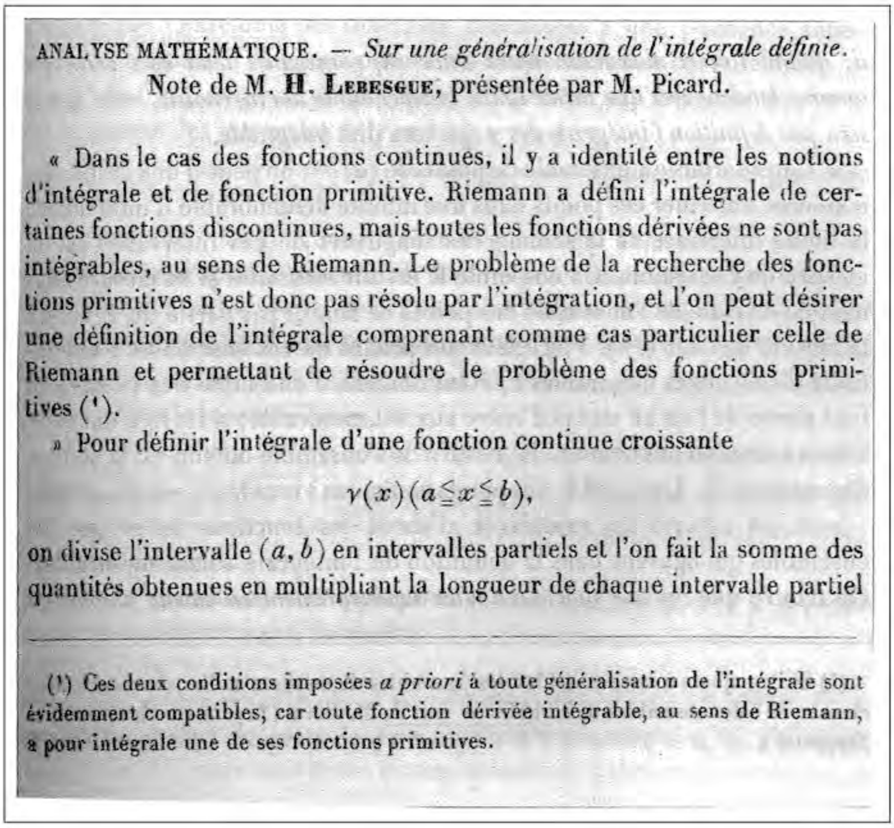
Les archives de l'Académie conservent le manuscrit de la note de Lebesgue du 29 avril 1901. Les quatre feuillets du manuscrit ont été découpés, de façon à permettre à trois typographes de travailler simultanément sur ce texte. Les délais d'impression étaient incroyablement rapides : à cette époque, et c'était

[1] Séance du 5 mars 2001.

[2] La troisième sera un colloque, à l'École Normale Supérieure de Lyon, à l'initiative de notre confrère Étienne Ghys, les 27 et 28 avril, à l'intention des élèves de l'École et des professeurs de la région, sous la double responsabilité de l'ENS de Lyon et de la Société Mathématique de France.

encore le cas au début des années 1950, on pouvait déposer une note à l'Académie le lundi avant 15h, corriger les épreuves le mercredi matin, et voir la note imprimée le lundi suivant.

Les Comptes rendus étaient un grand moyen de communication en mathématiques. Dans les années 1900 et 1901, on trouve les noms de Poincaré, Picard, Painlevé, Hadamard, Borel, Baire, Lebesgue parmi les Français, et parmi les étrangers de Stekloff, Liapounoff, Mittag-Leffler, Levi-Civita, Lindelöf, Fejér, Tzitzeica, Von Koch... C'est un grand journal international, très apprécié par les étrangers comme par les Français. La simple collection des notes en mathématiques de cette époque constituerait un bon document de synthèse sur une bonne partie de l'activité mathématique mondiale.



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.

Note de Lebesgue aux CRAS
reproduite avec l'aimable autorisation de la bibliothèque de l'IHP.

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 o 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. »

Revenons aux mathématiciens français que j'ai mentionnés. En 1901, Poincaré a 47 ans, Picard 45, Painlevé 38, Hadamard 36, Borel a 30 ans, Baire 27 ans, Lebesgue va avoir 26 ans.

Émile Borel est le chef de file de la nouvelle génération. Il a été enfant prodige, son intelligence est extrêmement vive, il a des intuitions fulgurantes. En 1896, il pressent ce que devrait être une théorie moderne des probabilités, faisant intervenir une infinité dénombrable d'évènements. En 1898, il publie un livre sur la théorie des fonctions dont l'essentiel est consacré aux ensembles de points sur la droite que l'on peut obtenir, à partir des intervalles, par les opérations de réunion dénombrable et de passage au complémentaire ; c'est alors qu'il introduit une notion nouvelle de mesure des ensembles, en en dégageant les propriétés principales, mais sans en achever la théorie. Il va diriger une collection de monographies, chez Gauthier-Villars, où il accueillera en particulier les travaux de Baire et de Lebesgue.

René Baire a passé sa thèse sur les fonctions discontinues. Il les divise en classes, selon la manière dont elles se représentent comme sommes de séries de fonctions continues. Il distingue parmi les ensembles de points les ensembles de 1ère et les ensembles de 2ème catégorie. La théorie de Baire est de nature topologique ; c'est un éclairage différent de celui de Borel, avec une mise au point d'emblée parfaite.

Henri Lebesgue connait bien les travaux de Borel et de Baire, il admire beaucoup Baire et il le tutoie. Par contre, il vouvoie Borel et, dans ses premières lettres, il lui donne du « Cher Monsieur ». À l'École Normale Supérieure, où il a connu Baire, il a eu pour camarades de promotion Paul Montel et Paul Langevin, qu'il admire également et considère non seulement comme un physicien, mais comme un mathématicien, « une espèce à part dans le genre mathématicien », écrira-t-il plus tard. Il a le goût de la géométrie, il connait bien l'analyse classique, et aussi les travaux de Cantor et des Italiens, Dini, Peano, Volterra qui s'intéressent à ce que l'école française dominante, Hermite ou Poincaré, considère comme une tératologie. C'est l'époque où Hermite écrivait à Stieltjes avec quelque malice qu'il se détournait « avec effroi et horreur de cette plaie lamentable des fonctions continues qui n'ont pas de dérivées ». Et Poincaré exprimait encore sa méfiance de façon encore plus carrée : « Autrefois, quand on inventait une fonction nouvelle, c'était en vue de quelque but pratique ; aujourd'hui on les invente tout exprès pour mettre en défaut les raisonnements de nos pères, et on n'en tirera jamais que cela. » Or, dès ses années d'École, Lebesgue est attiré par ce qui est simple et qui sort des sentiers battus. Classiquement, on établit qu'une surface développable, c'est à dire applicable sur un plan, est formée de droites ; effectivement, c'est le cas quand on tord une feuille de papier. Mais, observe Lebesgue, ce n'est plus le cas quand on la froisse. Quelle est donc la théorie des surfaces, hors du cadre classique, qui permet de rendre compte de la situation générale ? Il s'était entretenu de cette question avec Paul Montel à l'École, donc avant 1897, il s'y attache pendant deux année, 1898 et 1899, où il est boursier après l'agrégation, et encore, à partir de 1900, quand il est professeur de classe préparatoire à l'École Centrale au lycée de Nancy. En 1899 et 1900, il publie 5 notes aux Comptes rendus qui concernent toutes les fonctions de plusieurs variables, les surfaces, et particulièrement l'aire des surfaces.

La note de 1901 s'appelle modestement « Sur une généralisation de l'intégrale définie ». Vue d'aujourd'hui, c'est la définition de l'intégrale de Lebesgue et de la mesure de Lebesgue, deux notions de portée immense. De plus, c'est un exposé admirablement clair et succinct de ces deux notions. Comme la plupart des grandes nouveautés, elle n'a pas été reconnue tout de suite. Le Jahrbuch über die Fortschritte der Mathematik, le livre allemand qui, chaque année, recense et analyse toutes les publications mathématiques, a consacré beaucoup de place, plusieurs pages au total, aux notes de Lebesgue sur les surfaces. Il consacre 3 lignes à la note de 1901.

Avant Lebesgue, la conception dominante de l'intégrale est celle de Riemann. Pour intégrer une fonction $y = f(x)$, Riemann découpe l'intervalle (a, b) où varie x et il considère les sommes :

$$\sum y_i(x_{i+1} - x_i) \ ,$$

où y_i désigne une valeur prise par y sur l'intervalle (x_i, x_{i+1}). Si ces sommes tendent vers une limite quand on raffine le découpage, cette limite est l'intégrale $\int_a^b f(x)dx$, et on dit que la fonction f est intégrable au sens de Riemann sur (a, b). Riemann indique une condition nécessaire et suffisante pour qu'il en soit ainsi, que Lebesgue traduira de façon très simple grâce à sa notion de mesure : il faut et il suffit que l'ensemble des points x où la fonction est discontinue soit de mesure nulle.

Lebesgue, lui, découpe l'intervalle où varie y, et il associe à chaque intervalle (y_j, y_{j+1}) de ce découpage la mesure de l'ensemble des x tels que

$$y_j \leqslant f(x) < y_{j+1}$$

Si cette mesure est m_j, une valeur approchée de l'intégrale sera

$$\sum m_j y_j$$

Si ces sommes tendent vers une limite quand on raffine le découpage, c'est l'intégrale au sens de Lebesgue, et Lebesgue dit que la fonction est sommable. Dans sa note de 1901, Lebesgue s'attache uniquement au cas où les deux intervalles où varient x et y sont bornés. Plus tard, il levera cette restriction, qui aurait interdit beaucoup de développements ultérieurs.

Voici comment Lebesgue explique son point de vue, dans une conférence à Copenhague en 1926.

« Avec le procédé de Riemann, on sommait les indivisibles dans l'ordre où ils étaient fournis par la variation de x. On opérait donc comme le ferait un commerçant sans méthode qui compterait pièces et billets au hasard de l'ordre où ils lui tomberaient sous la main ; tandis que nous opérons comme le commerçant méthodique qui dit :

j'ai $m(E_1)$ pièces de une couronne, valant $1.m(E_1)$
j'ai $m(E_2)$ pièces de deux couronnes, valant $2.m(E_2)$
j'ai $m(E_5)$ pièces de cinq couronnes, valant $5.m(E_5)$
etc. Donc j'ai en tout

$$S = 1m(E_1) + 2m(E_2) + 5m(E_5) + \ldots \text{»}$$

Vous voyez les notations qu'utilise Lebesgue dans cet exemple. Au lieu d'écrire simplement m_1, m_2, m_5, il met en évidence un ensemble, E_j, et la mesure de cet ensemble, $m(E_j)$. Dans le cas général d'une fonction à valeurs réelles, et non nécessairement entières, il apparait des ensembles E, définis par les inégalités du type $u \leqslant f(x) \leqslant v$, et il s'agit de donner un sens à $m(E)$.

Lebesgue avait en vue un théorème du type

$$\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n$$

qu'il démontre, en effet, sous l'hypothèse que l'intégrale est prise sur un intervalle borné et que les fonctions f_n sont uniformément bornées, et, plus tard, sous l'hypothèse plus générale que les valeurs absolues $|f_n|$ sont toutes majorées par une fonction sommable. Le théorème de Lebesgue est la clé de voûte de la théorie. En même temps, il veut naturellement que l'intégrale soit additive, c'est-à-dire

$$\int (f+g) = \int f + \int g$$

Si l'on impose ces deux conditions, on est amené à une formule du type

$$\int \sum_1^\infty f_n = \sum_1^\infty \int f_n$$

qui doit être satisfaite, en particulier, si les f_n sont les fonctions indicatrices d'ensembles E_n disjoints. Cela veut dire

$$m(\sum_1^\infty E_n) = \sum_1^\infty m(E_n).$$

C'est l'additivité totale de la mesure telle que Borel l'a définie : la mesure d'une réunion dénombrable d'ensembles disjoints est la somme des mesures de ces ensembles. Borel part de la mesure des intervalles, qui est leur longueur, puis il passe aux réunions dénombrables d'intervalles et à leurs complémentaires, puis il répète les opérations de réunion dénombrable et de complémentation, autant de fois qu'il le veut, et il obtient ainsi les ensembles que nous appelons aujourd'hui boréliens. Il est convaincu de pouvoir construire une mesure totalement additive sur la famille en question, que nous appelons aujourd'hui la tribu borélienne, mais il ne le fait pas — ni en 1898 dans ses Leçons sur la théorie des fonctions, ni jamais plus tard.

Lebesgue a donc besoin de la mesure de Borel, mais la mesure de Borel n'existe encore qu'à l'état de projet, et sa construction pas à pas, en suivant les procédés de construction des ensembles boréliens, présente des difficultés. Au lieu de suivre cette voie, Lebesgue prend de la hauteur. Tout ensemble de points sur la droite peut être recouvert par une réunion d'intervalles. À chaque recouvrement on associe la somme des longueurs des intervalles. Puis on considère la borne inférieure de ces sommes pour tous les recouvrements possibles : c'est ce que Lebesgue appelle la mesure extérieure de l'ensemble. Très simplement, il définit une notion complémentaire, la mesure intérieure. Quand la mesure intérieure et la mesure extérieure coïncident, leur valeur commune s'appelle la mesure de l'ensemble, et l'ensemble est dit mesurable. C'est ce qu'aujourd'hui nous appelons la mesure de Lebesgue, définie sur la tribu de Lebesgue. La tribu de Lebesgue est bien plus grosse que la tribu de Borel, mais la mesure de Lebesgue, sur la tribu de Borel, n'est autre que la mesure de Borel.

C'est donc Lebesgue, indiscutablement, qui a achevé la théorie de la mesure de Borel, tout en introduisant une nouvelle classe d'ensembles, parmi lesquels les ensembles de mesure nulle. Comme tout ensemble mesurable au sens de Lebesgue est la réunion d'un borélien et d'un ensemble de mesure nulle, il est arrivé à Borel de dire que l'apport de Lebesgue était d'avoir introduit les ensembles de mesure nulle. Lebesgue a fort mal pris la chose, et s'est brouillé avec Borel pour cette raison. Il s'était déjà brouillé avec Baire. Les lettres de Lebesgue dans le fonds Borel, qui est depuis un an déposé aux Archives de l'Académie, expliquent les circonstances de ces dissentiments. Au début des années 1900, Borel semblait choyé par la vie, époux d'une femme remarquable qui était la fille du mathématicien Paul Appell, Baire était sans cesse malade, et Lebesgue était pauvre, surchargé d'enseignement (21 heures par semaine au lycée de Nancy), accablé de problèmes familiaux et financiers.

C'est juste avant 1900 qu'avait été créée la fondation Peccot au Collège de France, pour permettre à un jeune mathématicien — moins de 30 ans — d'exposer ses travaux en lui accordant quelques subsides. Les cours Peccot existent toujours et ils sont prestigieux dans le monde mathématique. Les premiers titulaires en ont été Borel pendant trois ans, puis Lebesgue à la place de Baire, qui était malade, puis Baire, puis de nouveau Lebesgue. Les querelles autour du cours Peccot, comme autour des nominations universitaires, n'ont qu'une valeur anecdotique, mais ces cours Peccot ont eu un impact considérable.

Après la note de 1901, Lebesgue a developpé la théorie de l'intégrale dans trois œuvres majeures : sa thèse, passée en 1902, qui s'appelle « Intégrale, longueur, aire » — et le titre indique bien le lien entre l'intégrale et les questions géométriques qui avaient fait l'objet de ses premières notes — puis, suite au premier cours Peccot, les « Leçons sur l'intégration et la recherche des fonctions primitives » et, suite au second cours Peccot, les « Leçons sur les séries trigonométriques ». Je vais tenter de donner un aperçu des matières traitées dans ces travaux. Mais auparavant, je veux souligner le rôle qu'ont eu les cours Peccot pour la continuation des travaux de Baire et de Lebesgue. Les auditeurs de ces trois cours, de Lebesgue, Baire et Lebesgue, étaient peu nombreux, mais il y avait parmi eux Arnaud Denjoy, qui a rédigé le cours de Baire et qui a été l'auteur d'une nouvelle théorie de l'intégration, la « totalisation », fondée à la fois sur les travaux de Cantor, de Baire et de Lebesgue, et Pierre Fatou, dont la thèse en 1906, à la suite du second cours Peccot de Lebesgue, sur « Séries trigonométriques et séries de Taylor », a montré tout le parti que l'analyse classique pouvait tirer du nouvel outil qu'était l'intégrale de Lebesgue.

Ainsi Fatou et Denjoy ont-ils été, en France et à cette époque, les principaux continuateurs de Lebesgue. À l'étranger, en Angleterre, en Belgique, en Autriche et en Hongrie, en Russie et en Pologne, l'intégrale de Lebesgue a été immédiatement adoptée, exposée et utilisée. Mais en France, elle n'était pas enseignée. Lebesgue lui-même ne l'a exposée qu'à l'occasion des cours Peccot. Après qu'il ait été nommé professeur au Collège de France, en 1921, ses cours, sauf rares exceptions, ont porté sur de tout autres sujets. En 1950, alors que dans tous les pays du monde l'intégrale de Lebesgue était un sujet classique, on pouvait être en France agrégé de mathématiques sans en avoir entendu parler. Szolem Mandelbrojt a évoqué dans ses souvenirs la déception qu'il avait éprouvée, en arrivant en France, de ne voir enseignée nulle part l'intégrale de Lebesgue et les théories qui en étaient issues.

Revenons donc aux trois écrits principaux de Lebesgue, de 1902, 1904 et 1906. Le premier, la thèse, est d'inspiration géométrique. Quoiqu'il contienne et complète la matière de la note de 1901 sur l'intégrale, il est inspiré du début à la fin par les questions qui l'avaient intéressé en 1899 et en 1900 : la longueur d'une courbe, et surtout l'aire d'une surface. Je vais me borner ici aux aires planes et à leur relation avec la théorie de la mesure. Camille Jordan avait exposé une théorie des aires planes qui a fortement influencé Lebesgue. Étant donné un domaine plan D, Jordan considère tous les polygones qui contiennent D, et tous les polygones qui sont contenus dans D ; la borne inférieure des aires des polygones contenant D est « l'étendue extérieure » de D, la borne supérieure des polygones contenus dans D est « l'étendue intérieure ». Lorsque les deux coïncident, on dit que le domaine est quarrable et que son aire est la valeur commune de l'étendue extérieure et de l'étendue intérieure. On voit que Lebesgue a tiré l'idée des mesures extérieure et intérieure de l'approche de Jordan — mais les définitions devaient être modifiées de façon radicale.

Pour faire saisir la différence entre l'étendue de Jordan et l'aire selon Lebesgue, je vais me contenter d'une figure, analogue à celle qu'expose Lebesgue dans une note de sa thèse. Partons d'un triangle ABC, dont la surface est S. Choisissons sur le côté BC deux points B' et C', et appelons S_1 la surface du triangle $AC'B'$. Lorsque l'on ôte du triangle ABC le $AC'B'$, on obtient deux nouveaux triangles, ABC' et $AB'C$, dont la surface totale est $S - S_1$. Puis on évide le triangle ABC' et le triangle $AB'C$ de la même manière, en leur ôtant des triangles de sommets C' et B' respectivement, dont les surfaces sont S_2 et S_3. On obtient maintenant quatre triangles dont la somme des surfaces est $S - S_1 - S_2 - S_3$. En continuant indéfiniment de cette manière, on a une suite de domaines emboités, formés de triangles, contenus dans le triangle ABC, et dont le complément relativement à ABC est constitué de triangles de surfaces S_1, S_2, S_3, S_4 etc. À la limite, on obtient une courbe Γ, contenue dans le triangle ABC, dont le complément est constitué d'une infinité de triangles dont la somme des surfaces est $\sum_1^\infty S_n$. Si l'on complète cette courbe Γ par un arc de cercle joignant B et C, à l'extérieur du triangle ABC, on obtient une courbe fermée simple qui délimite un domaine D. Lorsque $\sum_1^\infty S_n = S$, le domaine D est quarrable au sens de Jordan, et il ne l'est pas quand $\sum_1^\infty S_n < S$. Par contre, il est toujours mesurable au sens de Lebesgue. Au sens de Lebesgue, la courbe Γ a toujours pour aire $S - \sum_1^\infty S_n$, que cette quantité soit positive ou nulle. Aujourd'hui, nous appellerions une telle courbe une fractale.

La note de 1901 met en évidence une propriété importante de l'intégrale de Lebesgue : elle résout le problème des primitives pour des fonctions bornées. Cela veut dire que si une fonction f est dérivable sur un intervalle et que sa dérivée, f', est bornée, la différence

$$f(x) - \int_0^x f'(t)dt$$

est une constante. Si f' est intégrable au sens de Riemann, il en est ainsi. Mais il existe des cas où f' existe et est bornée, et n'est pas intégrable au sens de Riemann. Volterra en a donné un exemple, qui est reproduit dans la thèse de Lebesgue. Les « Leçons sur l'intégration » de 1904 contiennent la description et l'analyse des notions d'intégrale qui se sont succédées depuis Cauchy, en particulier avec Dirichlet et Riemann, la caractérisation des fonctions intégrables au sens de Cauchy-Dirichlet et au sens de Riemann, et leur relation à la théorie des fonctions dérivées. C'est Denjoy, en 1912, qui a généralisé l'intégrale de Lebesgue de façon à calculer les primitives des fonctions dérivées les plus générales, c'est à dire à s'affranchir de l'hypothèse que f' est bornée.

À la suite des « Leçons sur l'intégration » d'excellents livres ont exposé l'intégrale de Lebesgue et ses applications, en particulier par Charles de la Vallée Poussin en 1916 (« Fonctions d'ensembles, intégrales de Lebesgue, classes de Baire ») et par Felix Hausdorff en 1927 (« Mengenlehre »). Mais la bible sur le sujet est l'ouvrage de Stanislas Saks paru en 1933 dans la collection polonaise des Monographies mathématiques, il est écrit en français et a pour titre « Théorie de l'intégrale ». Voici son appréciation sur le lien réalisé par Lebesgue entre intégration et dérivation :

« Le mérite de M. Lebesgue ne se réduit pas seulement à la création d'une notion nouvelle et plus générale de l'intégrale, ni même à la mise de cette notion en relation étroite avec la théorie de la mesure, la valeur de son œuvre consiste en premier lieu dans sa théorie de la dérivabilité, qu'il a construite parallèlement à celle de l'intégrale. C'est grâce à elle que la découverte de M. Lebesgue a trouvé tant d'applications dans les branches les plus diverses de l'analyse... »

La première de ces branches est la théorie des séries trigonométriques. Historiquement, il y a un rapport étroit entre la notion d'intégrale et les séries trigonométriques. Cela s'explique par le fait que les coefficients d'une série trigonométrique se calculent au moyen d'intégrales, par les formules de Fourier : si

$$f(t) = \sum c_n e^{int}$$

on a

$$c_n = \int_0^{2\pi} f(t) e^{-int} dt/2\pi.$$

La notation de l'intégrale définie, $\int_a^b$, a été introduite par Fourier à cette occasion, et Fourier attribuait à ces formules une valeur uiverselle. En fait, une bonne partie de la théorie des séries, de celle de l'intégrale, et plus tard de l'analyse fonctionnelle allait consister à leur donner un sens. Quelques étapes marquantes sont l'article de Dirichlet de 1830, la thèse de Riemann sur les séries trigonométriques de 1854, la thèse de Cantor en 1870, le théorème de Fejér de 1900 (encore une note aux Comptes rendus !), les leçons de Lebesgue publiées en 1906, le théorème de Riesz-Fischer en 1907, la totalisation de Denjoy sous sa seconde forme en 1921, la théorie des distributions de Schwartz en 1949, et la liste n'est pas close.

Pour Dirichlet, il s'agissait d'intégrer des fonctions continues, en raccordant les intégrales aux points de discontinuité. Comme exemple de fonction non intégrable, il donnait la fonction définie sur $(0,1)$, qui vaut 1 sur les irrationnels et 0 sur les rationnels. Pour Riemann, nous l'avons vu, la notion d'intégrale est plus générale ; la fonction de Dirichlet est toujours non intégrale. Pour Lebesgue au contraire, c'est une fonction intégrable (pour éviter la confusion, Lebesgue dit « sommable ») et son intégrale vaut 1.

Riemann dans sa thèse avait mené l'étude des fonctions qui sont sommes en tout point d'une série trigonométrique. Il les avait caractérisées sans avoir utilisé le calcul des coefficients, puisqu'en général ces fonctions ne sont pas intégrables au sens de Riemann. Cantor avait montré que les coefficients sont bien déterminés par la somme de la série. Un très beau résultat de Lebesgue est que, si une fonction somme d'une série trigonométrique est bornée, elle est intégrable en son sens, et les coefficients s'obtiennent par les formules de Fourier. Pour lever la restriction que la fonction est bornée, il a fallu attendre Denjoy et sa seconde totalisation.

Les leçons de Lebesgue sur les séries trigonométriques contiennent une foule de résultats qui, joints à la thèse de Fatou, ont immédiatement conquis le public mathématique en dehors de la France. Elles ont aussi deux particularités. C'est la première fois, que dès le départ, Lebesgue donne la définition générale de l'intégrale, des fonctions non-bornées comme des fonctions bornées. Et c'est la dernière fois que Lebesgue présente un exposé de son intégrale.

En évoquant la naissance de l'intégrale de Lebesgue j'ai été amené, déjà, à parler de sa descendance. S'agissant de la postérité de l'intégrale de Lebesgue, je me bornerai maintenant à trois aspects : l'analyse fonctionnelle, la théorie moderne des probabilités, et quelques prolongements directs de la mesure et de l'intégrale de Lebesgue.

Pendant l'année 1907, le Hongrois Frédéric Riesz et l'Autrichien Ernst Fischer ont publié une série de 5 notes aux Comptes rendus qui ont donné un nouvel éclairage à l'intégrale de Lebesgue. Par des méthodes légèrement différentes, ils donnent une caractérisation des coefficients de Fourier des fonctions dont le carré est intégrable sur l'intervalle $(0, 2\pi)$: les carrés de leurs valeurs absolues forment une série convergente. Ainsi, non seulement on a l'égalité à laquelle Fatou a donné le nom de Parseval,

$$\sum_{-\infty}^{\infty} |c_n|^2 = \int_0^{2\pi} |f(t)|^2 \frac{dt}{2\pi},$$

lorsque les c_n sont les coefficients de Fourier-Lebesgue (je veux dire, obtenus par les formules de Fourier en utilisant l'intégrale de Lebesgue), mais encore, comme l'explique Frédéric Riesz de manière plaisante dans un article de 1949, les formules de Fourier constituent « un billet d'aller et retour permanent » entre deux espaces à une infinité de dimensions, l'espace ℓ^2 des suites de carré sommable et l'espace L^2 des fonctions de carré intégrable. En termes plus savants, la transformation de Fourier est un isomorphisme entre l'espace L^2 et l'espace ℓ^2, déjà introduit par Hilbert.

La clé de la démonstration, chez Fischer comme chez Riesz, tient en trois mots : « L^2 est complet ». Plus tard, Frédéric Riesz explicite une version plus générale de ce résultat : « L^p est complet », en désignant par L^p l'espace des fonctions dont la p-ième puissance de la valeur absolue est intégrable, p étant un nombre réel quelconque $\geqslant 1$.

L'expression « L^2 est complet » signifie que l'on peut traiter L^2 à la manière d'un espace euclidien, comme Fischer et Riesz en avaient clairement l'intuition en parlant l'un et l'autre d'une sorte de « géométrie des fonctions ». L'idée de traiter les fonctions comme des points dans un espace abstrait n'était pas nouvelle : on la trouve chez Volterra, Hadamard et surtout dans la thèse de Maurice Fréchet de 1906, mais elle prenait là une forme frappante. Pour les fondateurs de la mécanique quantique, l'espace L^2 a été un outil immédiatement disponible. D'autres classes de fonctions ou de distributions peuvent être construits, à partir de L^2, en conservant cette structure géométrique d'espaces de Hilbert : ce sont les espaces de Sobolev, essentiels dans la théorie des équations aux dérivées partielles.

Les espaces L^p avec $p \neq 2$ ont une géométrie différente. Ce sont des espaces de Banach. C'est d'ailleurs Banach, dans sa « Théorie des opérations linéaires », le premier volume des *Monographies mathématiques polonaises*, qui a donné la forme actuelle aux théorèmes « L^2 est complet », « L^p est complet » : dans les écrits de Fischer et de Riesz, il fallait de longues phrases pour exprimer cela. En 1907, les espaces « L^p » n'étaient pas encore nommés, et l'adjectif « complet » avait un autre sens. Banach, en définissant « complet » avec le sens actuel et en nommant les espaces L^p (d'après l'initiale de Lebesgue), a fait passer dans les définitions l'essence de ces propositions, et c'est pourquoi, maintenant, elles s'énoncent de façon si simple.

Ainsi, au départ, il y a un problème sur les coefficients de Fourier, et une solution élégante, c'est le théorème de Riesz-Fischer. Puis une proposition intermédiaire, utilisée comme lemme, émerge comme résultat important, et prend le nom de théorème. Elle est encore longue et compliquée à énoncer. Enfin, deux définitions viennent permettre d'énoncer ce théorème en trois mots. L'élixir de la théorie est passée dans les définitions.

L'analyse fonctionnelle contemporaine utilise bien d'autres espaces que les espaces L^p, mais elle ne peut pas se passer des espaces L^p. Il n'est pas abusif de placer Lebesgue à la base de l'analyse fonctionnelle, tant par l'importance de l'intégrale comme forme linéaire que par les espaces fonctionnels qu'elle a permis de considérer.

L'influence de Lebesgue sur la théorie des probabilités a suivi deux voies principales, auxquelles on peut attacher les noms du Polonais Hugo Steinhaus et de l'Américain Norbert Wiener, au cours de la même décennie, 1920–1930.

L'idée de Steinhaus est de fonder les probabilités sur la considération de l'intervalle $(0, 1)$ de la droite réelle, que je désignerai par I. Il constitue ainsi un dictionnaire : un évènement est une partie de I mesurable au sens de Lebesgue, sa probabilité est la mesure de Lebesgue de cette partie ; une variable aléatoire est une fonction définie sur I et mesurable au sens de Lebesgue, sa distribution est l'image de la mesure de Lebesgue par cette fonction, et son espérance, quand elle existe, est l'intégrale de Lebesgue de cette fonction, et ainsi de suite. Il montre comment construire sur I une suite de variables aléatoires indépendantes ayant toutes pour distribution la mesure de Lebesgue sur I. À partir de là, il a un modèle mathématique pour toutes les suites ou séries de variables aléatoires indépendantes. La première application qu'il donne est celle-ci : si, dans une série de Taylor admettant un cercle C comme cercle de convergence, on change les phases des coefficients de manière aléatoire suivant la probabilité naturelle, on obtient presque sûrement (c'est-à-dire avec probabilité égale à 1) une fonction qui ne se prolonge pas analytiquement hors de C. C'était là une idée de Borel, qui datait de 1896, mais que ni Borel ni personne d'autre n'avait jusqu'alors converti en énoncé mathématique.

L'idée de Wiener était tout autre. Wiener avait lu Einstein et Jean Perrin, qui avaient établi, théoriquement et expérimentalement, une théorie physique du mouvement brownien. Jean Perrin avait observé que les trajectoires du mouvement brownien rappelaient les fonctions continues sans dérivées des mathématiciens, que l'on tenait jusque là pour des curiosités tout à fait étrangères au monde physique. La réalité physique du mouvement brownien suggérait qu'il y en ait un modèle mathématique. Le programme de Wiener était de munir l'ensemble des fonctions continues du temps t, partant d'un point donné au temps $t = 0$, d'une mesure de probabilité telle que les équations d'Einstein soient satisfaites et que, presque sûrement, ces fonctions jouissent de propriétés convenables, dont celle de n'être dérivables nulle part. Construire une mesure sur un espace plus gros que l'intervalle I n'était pas chose nouvelle. Mais la construction de la mesure de Wiener était un tour de force. Il s'agissait, à partir de conditions données a priori sur le processus — c'est un processus gaussien et ses accroissements sont indépendants — de construire la mesure de façon que ces propriétés soient vérifiées. Une fois la mesure de probabilité construite, on peut traiter la fonction comme un objet aléatoire. Wiener l'appelait « the fundamental random function », puis, suivant Paul Lévy, on l'a appelé tout simplement « le mouvement brownien ». C'est un objet fascinant, d'usage constant dans plusieurs domaines des mathématiques, et sur lequel physiciens et mathématiciens fournissent en permanence de merveilleuses conjectures et de superbes théorèmes.

Ainsi Wiener n'utilisait pas la mesure de Lebesgue, mais il en construisait une variante sur un espace de fonctions. Au contraire Steinhaus prenait comme espace de probabilité universel l'intervalle $(0, 1)$ de la droite réelle, muni de la mesure de Lebesgue. On peut parfaitement construire le mouvement brownien à partir du point de vue de Steinhaus, au moyen de la série de Fourier-Wiener, qui est une série trigonométrique dont les coefficients sont des variables aléatoires gaussiennes indépendantes, et c'est ce que Wiener a fait dans le dernier exposé qu'il a présenté de sa « fundamental random function » en 1934.

Or, au moment précis ou Wiener se ralliait au point de vue de Steinhaus, Kolmogorov reprenait l'idée de Wiener, de construire un espace de probabilité adapté à un processus donné par une loi. Les Grundbegriffe der Wahrscheinlichkeitsrechnung de Kolmogorov définissent un espace de probabilité comme un ensemble Ω, muni d'une tribu de parties qu'on appelle les évènements, et d'une mesure totalement additive sur cette tribu qu'on appelle la probabilité. C'est le triplet $(\Omega, \mathcal{A}, P)$ qui, depuis lors, est le signe de ralliement des probabilistes. Le point de vue de Kolmogorov est plus flexible que celui de Steinhaus parce qu'en cours de route, dans le développement d'une théorie, on peut changer d'espace Ω ou de tribu $\mathcal{A}$. Mais les espaces de probabilité les plus importants sont isomorphes à l'espace de Lebesgue, c'est à dire à l'intervalle $(0, 1)$ de la droite réelle, muni de la tribu et de la mesure de Lebesgue.

La postérité de l'intégrale de Lebesgue serait impressionnante au seul titre des espaces fonctionnels et des espaces de probabilité. Mais elle comporte bien d'autres descendants, que je veux évoquer rapidement.

Je commencerai par la théorie géométrique de la mesure, où l'article fondateur est celui de Felix Hausdorff, en 1919. Hausdorff transpose à une situation plus générale la construction de la mesure extérieure de Lebesgue. Au lieu de la droite, il considère un espace sans spécification particulière. Au lieu d'intervalles, il dispose d'un stock de petits corps pesants, susceptibles de recouvrir un objet quelconque situé dans l'espace. Étant donné un tel objet, il attache à chaque recouvrement une masse, qui est la masse totale des corps qui participent au recouvrement — en général, une infinité de corps est nécessaire. Puis, il prend la borne inférieure de ces masses, pour tous les recouvrements astreints à une certaine condition de finesse — par exemple que le diamètre des corps ne dépasse pas un ε donné — puis la limite de cette borne inférieure lorsque l'on affine cette condition — par exemple, lorsque l'on fait tendre ε vers zéro. Il obtient alors pour l'objet en question une mesure, qui a toutes les propriétés de la mesure extérieure de Lebesgue, et qui peut être finie ou infinie. C'est la mesure de Hausdorff. Si l'espace est muni d'une distance, on peut choisir comme stock de petits corps l'ensemble de toutes les boules, et pour masse d'une boule une puissance donnée du diamètre, disons, d^α. On obtient alors la mesure de Hausdorff en dimension α. Elle est en général infinie lorsque α est petit, et nulle lorsque α est grand. La dimension de Hausdorff est la valeur de α qui délimite ces deux situations.

Avant même la grande vogue de la fractalité, mesure et dimension de Hausdorff ont joué un rôle important en analyse harmonique et en théorie du potentiel. En théorie du potentiel la notion naturelle n'est plus la mesure, mais la capacité d'un ensemble. Choquet, dans sa théorie des capacités, est un héritier lointain de Lebesgue.

Une autre généralisation de la mesure de Lebesgue concerne les espaces dans lesquels on peut définir des objets congruents. Selon certaines conditions, on peut construire une mesure, unique à une normalisation près, telle que deux objets congruents aient la même mesure. C'est là une découverte du Hongrois Alfred Haar, simplifiée ensuite par Banach, et exposée dans la « Théorie de l'intégrale » de Saks sous forme définitive. La mesure de Haar est un outil essentiel dans l'étude des groupes topologiques.

Sans que le terme de tribu soit utilisé à l'époque, il semblait à Borel qu'aucune opération de l'analyse ne ferait jamais sortir de la tribu borélienne. C'était aussi l'avis de Lebesgue, et il avait cru le démontrer en 1905. Rarement erreur a été plus fructueuse. Au début de l'année 1917, les Comptes rendus publient deux notes des Russes Nicolas Lusin et M. Ya. Souslin qui établissent le contraire : on peut trouver une suite de polynômes $P_n(x)$, qui convergent vers une valeur $y(x)$ quand x appartient à un certain ensemble borélien, mais tels que l'ensemble des valeurs prises par $y(x)$ ne soit pas borélien. Ainsi l'image borélienne d'un borélien n'est pas nécessairement un borélien. Il y a mieux : l'ensemble des couples $(x, y(x))$ est un borélien, mais sa projection sur l'axe des y, qui est l'ensemble des $y(x)$, ne l'est pas. La projection d'un borélien n'est pas nécessairement un borélien. L'analyse classique force donc à sortir de la tribu borélienne. Entre la tribu de Borel et celle de Lebesgue se trouve la tribu de Lusin, constituée par les ensembles que Lusin appelle analytiques et qui sont des images continues de boréliens.

Au cours des années 1920 s'est développée à Moscou une école mathématique extrêment brillante, dont Lusin a été le fondateur. Ainsi Lebesgue et son œuvre ont été beaucoup mieux connus à Moscou qu'ils ne l'étaient en France. Hongrie, Pologne et Russie ont été les foyers de rayonnement de la pensée de Lebesgue et de son héritage.

Je terminerai par quelques mots sur l'intégrale, autrefois et aujourd'hui. À l'époque de Cauchy se sont dégagées des notions qui sont restées intangibles, telles que la convergence des suites ou des séries, la continuité, la dérivabilité, l'analyticité. La notion générale de fonction s'est aussi établie à cette époque. Il était naturel de chercher à préciser ce qu'est une fonction intégrable. Après les tentatives de Cauchy et de Dirichlet, de rattacher l'intégrabilité à la continuité des fonctions sur certains intervalles, l'intégrale de Riemann a fait l'unanimité pendant plusieurs décennies : une fonction était intégrable si elle admettait une intégrale au sens de Riemann. C'est pourquoi Lebesgue utilise un autre adjectif, sommable, pour qualifier les fonctions susceptibles d'intégration selon sa methode. C'est avec Hardy et Littlewood que l'on a commencé à parler de fonctions intégrables au sens de Lebesgue. L'intégrale qui s'est imposée au 20ème siècle est l'intégrale de Lebesgue, mais elle a subi bien des extensions et modifications : les intégrales de Denjoy, de Perron, de Henstock ; celle de Radon, de Saks, de Haar ; celles de Wiener, d'Itô, de Feynman. Ces différentes intégrales interviennent dans l'étude des fonctions d'une variable réelle, en analyse fonctionnelle, en probabilités et en physique théorique. L'intégrale de Feynman a donné du fil à retordre aux mathématiciens. Une approche rappelle celle de Lebesgue, et elle est due à Pierre Cartier et Cécile de Witt : elle consiste à imposer des conditions à l'intégrale en question qui en assurent l'unicité, et alors seulement à donner un procédé de construction.

On a beaucoup de choix pour enseigner l'intégrale, à tous les niveaux, parce que c'est une notion polysémique. On peut dire, avec Youri Manin, que l'intégrale, c'est la quantité de quelque chose dans un domaine. Mais d'autre part, intégrer une équation différentielle, c'est partir d'une donnée sur la variation instantanée pour décrire l'évolution d'un phénomène. Ainsi au lycée, on peut privilégier la solution de $y' = f(x)$, comme on l'a fait pendant quelques années, ou privilégier le calcul des aires, comme on le fait maintenant : en tout cas, il faut tenir les deux bouts. Au niveau supérieur, on peut rattacher l'intégrale aux fonctions de variables réelles, à l'analyse fonctionnelle ou aux probabilités. Bourbaki avait choisi de présenter l'intégrale comme forme linéaire sur un espace de fonctions continues, donc le substrat était un espace topologique. Denjoy, fidèle à la distinction entre le point de vue de Baire, topologique, et celui de Lebesgue, métrique, avait violemment attaqué Bourbaki, et aujourd'hui le point de vue probabiliste donne raison à Denjoy. Cependant, à lire l'autobiographie de Laurent Schwartz, on voit que cette erreur de Bourbaki, si erreur il y a, a joué un rôle décisif pour découvrir la bonne définition des distributions, comme formes linéaires sur certains espaces de fonctions.

La conclusion est que, quelle que soit la perfection d'un exposé mathématique, et l'exposé de Lebesgue en 1901, d'emblée, atteint une sorte de perfection, les mathématiques sont beaucoup plus riches que tout exposé qui peut en être fait. L'efflorescence mathématique qui a suivi l'intégrale de Lebesgue m'a paru justifier ce long compte rendu.