L'intégrale de Stieltjes - Un précurseur de l'intégrale de Lebesgue

1 Quelques rappels d'intégration de Riemann

Revenons pour un moment dans un monde pré-Lebesgue. On y dispose de l'intégrale de Riemann, définie comme suit. Soit $f : [a,b] \to \mathbb{R}$ une fonction bornée dont on souhaite connaître l'aire sous la courbe. On considère une subdivision $\sigma = (a = t_0 < x_1 < \cdots < t_n = b)$ de l'intervalle $[a,b]$ en sous-intervalles $[t_i, t_{i+1}]$, et, sur chaque sous intervalle, on approche f par une fonction constante. On peut choisir $\sup_{[t_i,t_{i+1}]} f$ (on approche f "par le haut"), ou $\inf_{[t_i,t_{i+1}]} f$ (on approche f "par en-dessous").

Alors l'aire sous la courbe de ces fonctions en escalier est facile à calculer : c'est un paquet de rectangles. On obtient ainsi

$$I^+(f,\sigma) = \sum_{i=0}^{n-1} (t_{i+1} - t_i) \sup_{[t_i,t_{i+1}]} f, \quad I^-(f,\sigma) = \sum_{i=0}^{n-1} (t_{i+1} - t_i) \inf_{[t_i,t_{i+1}]} f$$

où $I^+(f,\sigma)$ est légèrement plus grande que l'aire sous la courbe de f et $I^-(f,\sigma)$ est légèrement plus petite. Si on prend des subdivisions de plus en plus fines (et que f n'est pas trop méchante), on devrait approcher de mieux en mieux l'aire sous la courbe de f. Plus rigoureusement, on définit donc

$$I^+(f) = \inf\{I^+(f,\sigma) \,|\, \sigma \text{ subdivision de } [a,b]\}, \; I^-(f) = \sup\{I^-(f,\sigma) \,|\, \sigma \text{ subdivision de } [a,b]\}$$

et ont dit que f est intégrable si $I^+(f) = I^-(f)$. Dans ce cas, on appelle cette quantité intégrale de f, notée $\int_a^b f(t)dt$. Les fonctions intégrables incluent les fonctions continues, les fonctions monotones, et bien sûr les fonctions en escaliers.

Si f est Riemann-intégrable, alors on peut considérer la subdivision régulière σ_n de $[a,b]$ en n sous-intervalles de même taille, donnée par $t_i = a + i\frac{b-a}{n}$. Alors, pour $t_i \leq c_i \leq t_{i+1}$,

$$R_n(f) = \sum_{i=0}^{n-1} f(c_i)(t_{i+1} - t_i) = \frac{b-a}{n} \sum_{i=0}^{n-1} f(c_i) \to \int_a^b f(t)dt$$

Le membre de gauche est appelé "somme de Riemann".

Rappelons qu'on a le critère d'intégrabilité suivant, obtenu en utilisant les définitions de sup et inf :

Proposition 1. *Une fonction $f : [a,b] \to \mathbb{R}$ bornée est intégrable si et seulement si, pour tout $\varepsilon > 0$, il existe une subdivision σ_ε telle que $I^+(f,\sigma_\varepsilon) - I^-(f,\sigma_\varepsilon) < \varepsilon$.*

De manière équivalente, f est Riemann-intégrable si et seulement si, pour tout $\varepsilon > 0$, il existe deux fonctions en escaliers h_ε^+ et h_ε^- telles que $h_\varepsilon^- \leq f \leq h_\varepsilon^+$ et telles que

$$\int_a^b h_\varepsilon^+(t)dt - \int_a^b h_\varepsilon^-(t)dt < \varepsilon$$

On a déjà discuté des limites de l'intégrale de Riemann, et de l'intérêt de la généraliser. Une direction de généralisation est l'intégrale de Lebesgue, mais il en existe une autre, plus ancienne (1894) et due à Thomas Stieltjes, et qui part dans une direction assez différente.

2 Un nouvel espoir : l'intégrale de Stieltjes

2.1 Motivation

On peut voir l'intégrale de f entre a et b, (plus précisément, la quantité $\frac{1}{b-a}\int_a^b f(t)dt$) comme une "valeur moyenne" de f sur $[a,b]$. Dans ce calcul de moyenne, tous les points de l'intervalle sont traités de la même façon :

1. chaque point a une contribution individuelle nulle (on dit qu'il est *négligeable*). Les points ne contribuent à l'intégrale que groupés dans des sous-intervalles ;
2. de plus, le "poids" de chaque sous-intervalle ne dépend que de sa longueur, et pas du tout de sa position dans $[a,b]$.

On peut toutefois imaginer des situations où l'on voudrait donner plus d'importance aux valeurs de f dans une certaine région de $[a,b]$ que dans une autre, ou garder trace de la valeur de f en des points spécifiques [1]. C'est ce que l'intégrale de Stieltjes (ou l'intégrale de Thomas, pour les intimes), se propose de faire.

2.2 Construction

Pour remplir ce contrat, on introduit une fonction $p : [a,b] \to \mathbb{R}$ *croissante* qui va nous permettre de contrôler le poids donné à différents points et sous-intervalles. Soit $f : [a,b] \to \mathbb{R}$ une fonction bornée et $\sigma = (a = t_0 < \cdots < t_n = b)$ une partition de $[a,b]$. On définit

$$S^+(f,\sigma) = \sum_{i=0}^{n-1}(p(t_{i+1}) - p(t_i)) \sup_{[t_i,t_{i+1}]} f, \quad S^-(f,\sigma) = \sum_{i=0}^{n-1}(p(t_{i+1}) - p(t_i)) \inf_{[t_i,t_{i+1}]} f$$

Ainsi, par rapport à Riemann, on remplace la longueur des sous intervalles par $p(t_{i+1}) - p(t_i)$, une longueur "pondérée" par p (mais toujours positive, puique p est croissante). Et on poursuit la construction de la même façon : on pose

$$S^+(f) = \inf\{S^+(f,\sigma) \,|\, \sigma \text{ subdivision de } [a,b]\}, \; S^-(f) = \sup\{S^-(f,\sigma) \,|\, \sigma \text{ subdivision de } [a,b]\}$$

et on dit que f est *Riemann-Stieltjes-intégrable* (qu'on abrègera en RS-intégrable) si $S^+(f) = S^-(f)$. On note alors cette quantité $\int_a^b f dp$ ou $\int_a^b f(t)dp(t)$ et on l'appelle l'intégrale de Riemann-Stieltjes de f par rapport à p.

Remarque : Si on prend $p(x) = x$, on retombe sur l'intégrale de Riemann.

Comme précédemment, si $t_i = a + i\frac{b-a}{n}$, les sommes de Riemann-Stieltjes convergent vers l'intégrale :

$$RS_n(f) = \sum_{i=0}^{n-1} f(t_i)(p(t_{i+1}) - p(t_i)) \to \int_a^b f(t)dp(t)$$

Si h est une fonction en escaliers, mettons $h = \sum_k \alpha_k \mathbb{1}_{I_k}$, où les $I_k = [x_k, x_{k+1}]$ forment une partition de $[a,b]$ en sous-intervalles, alors on a le résultat attendu :

$$\int_a^b h(t)dp(t) = \sum_k \alpha_k(p(x_{k+1}) - p(x_k))$$

On a donc le critère d'intégrabilité suivant, analogue au cas de Riemann :

Proposition 2. *Une fonction $f : [a,b] \to \mathbb{R}$ bornée est intégrable si et seulement si, pour tout $\varepsilon > 0$, il existe une subdivision σ_ε telle que $S^+(f,\sigma_\varepsilon) - S^-(f,\sigma_\varepsilon) < \varepsilon$.*

De manière équivalente, f est Riemann-intégrable si et seulement si, pour tout $\varepsilon > 0$, il existe deux fonctions en escaliers h_ε^+ et h_ε^- telles que $h_\varepsilon^- \le f \le h_\varepsilon^+$ et telles que

$$\int_a^b h_\varepsilon^+(t)dp(t) - \int_a^b h_\varepsilon^-(t)dp(t) < \varepsilon$$

1. Cela peu sembler un peu flou, voir capillotracté, mais on verra un peu plus bas des applications en probabilités de ce genre d'idées.

On récupère les propriétés attendues de l'intégrale :

Proposition 3. *Soient $f, g : [a, b] \to \mathbb{R}$ RS-intégrables, $\lambda \in \mathbb{R}$, alors*

1. *$f + \lambda g$ est RS-intégrable, et l'intégrale est linéaire : $\int_a^b f + \lambda g dp = \int_a^b f dp + \lambda \int_a^b g dp$;*
2. *Pour $c \in [a, b]$, $\int_a^b f dp = \int_a^c f dp + \int_c^b f dp$;*
3. *Si $f \leq g$ sur $[a, b]$, $\int_a^b f dp \leq \int_a^b g dp$;*
4. *$\left|\int_a^b f dp\right| \leq \int_a^b |f| dp$*

De plus, le critère d'intégrabilité ci-dessus permet de démontrer, exactement comme on le fait pour l'intégrale de Riemann, que

- les fonctions continues sont RS-intégrables
- les fonctions monotones sont RS-intégrables.

Pour mieux se faire une idée de cette intégrale, testons-la sur quelques exemples.
Exemple 1 : Calculons $\int_a^b 1 dp$. Soit $\sigma = (a = x_0 < \cdots < x_n = b)$ une subdivision, alors

$$\sum_{i=0}^{n-1} 1 \cdot (p(x_{i+1}) - p(x_i)) = p(x_n) - p(x_0) = p(b) - p(a)$$

donc $\int_a^b 1 dp = p(b) - p(a)$.
Exemple 2 : Soit $p(x) = e^x$, calculons $\int_0^1 x dp(x)$. Considérons la subdivision régulière $t_i = i\frac{1}{n}$ de $[0, 1]$. Alors la somme associée est

$$RS_n(f) = \sum_{i=0}^{n-1} f(t_i)(p(t_{i+1}) - p(t_i)) = \sum_{i=0}^{n-1} t_i(e^{t_{i+1}} - e^{t_i})$$

D'après le théorème des accroissements finis, pour chaque i, il existe $c_i \in [t_i, t_{i+1}]$ tel que $e^{t_{i+1}} - e^{t_i} = e^{c_i}(t_{i+1} - t_i) = \frac{b-a}{n} e^{c_i}$, donc

$$RS_n(f) = \frac{1}{n} \sum_{i=0}^{n-1} t_i e^{c_i}$$

Or, par continuité de exp, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $|c_i - t_i|$ soit suffisamment petit pour que $|e^{c_i} - e^{t_i}| < \varepsilon$ pour tous les i, et alors

$$\left| RS_n(f) - \frac{1}{n} \sum_{i=0}^{n-1} t_i e^{t_i} \right| \leq \frac{1}{n} \sum |t_i||e^{c_i} - e^{t_i}| < M\varepsilon.$$

Or $\frac{1}{n} \sum_{i=0}^{n-1} t_i e^{t_i}$ est la somme de Riemann associée à la fonction xe^x, et celle-ci converge vers $\int_0^1 xe^x dx$. Mais la majoration de $|RS_n(x) - R_n(xe^x)|$ que l'on vient d'obtenir montre que ces deux sommes ont la même limite quand $n \to \infty$, d'où

$$\int_0^1 x d\exp(x) = \int_0^1 x \exp(x) dx$$

En fait, on montre, exactement de la même façon, le résultat suivant :

Proposition 4. *Si $p : [a, b] \to \mathbb{R}$ est croissante et de classe $\mathcal{C}^1$, alors, pour toute fonction RS-intégrable f,*

$$\int_a^b f(t) dp(t) = \int_a^b f(t) p'(t) dt$$

Il semble donc que le cas où p est une brave fonction $\mathcal{C}^1$ ne soit pas le plus intéressant, puisqu'on retombe alors bêtement sur l'intégrale de Riemann classique. Regardons donc ce qui se passe quand p est plus irrégulière.

Exemple 3 : Sur $I = [0,1]$, considérons la fonction poids

$$p(x) = \begin{cases} 0 \text{ si } 0 \leq x \leq \frac{1}{2} \\ 1 \text{ si } \frac{1}{2} < x \leq 1 \end{cases}$$

Soit f une fonction RS-intégrable, que l'on suppose continue en $\frac{1}{2}$, quelle est son intégrale de Stieltjes avec ce poids ? Si on considère la subdivision régulière $t_i = i\frac{1}{n}$, la somme associée est

$$RS_n(f) = \sum_{i=0}^{n-1} f(t_i)(p(t_{i+1}) - p(t_i))$$

Or, $p(t_{i+1}) - p(t_i) = 0$, puisque p est constante sur le sous-intervalle $[t_i, t_{i+1}]$, sauf si celui-ci contient $\frac{1}{2}$: il existe i_n tel que $\frac{i_n}{n} \leq \frac{1}{2} \leq \frac{i_n+1}{n}$, et alors $p(t_{i_n+1}) - p(t_{i_n}) = 1$. On a donc $RS_n(f) = f\left(\frac{i_0}{n}\right)$, et lorque $n \to \infty$, on a donc

$$\int_a^b f(t)dp(t) = \lim_{n\to\infty} RS_n(f) = f\left(\frac{1}{2}\right).$$

Ce choix de poids nous permet donc d'isoler la valeur que prend f en $\frac{1}{2}$. Si on avait choisi une fonction poids qui "saute" en un autre point x_0 de $[0,1]$, on aurait obtenu que l'intégrale de Stieltjes avec ce poids est égale à $f(x_0)$

Remarque : ça ne marche que si f est continue en x_0 ! Sinon, on n'est pas sûrs que $f\left(\frac{i_0}{n}\right) \to f(x_0)$.

Plus généralement, si p est une fonction en escalier croissante, associée à une subdivision $(x_0 < \cdots < x_p)$, on obtient de la même façon que si f est RS-intégrable et n'est discontinue en aucun des x_k alors

$$\int_a^b f(t)dp(t) = \sum_{k=0}^{n-1} \delta_k f(x_k),$$

où δ_k est le "saut" de p en x_k : $\delta_k = p(x_k^+) - p(x_k^-)$.

Exemple 4 : Considérons une fonction continue f sur $\mathbb{R}^+$. Alors, on obtient que la n-ième somme partielle de la série de terme général $f(k)$ est une intégrale de Stieltjes,avec le poids $p(x) = E(x)$ (partie entière) :

$$S_n = \sum_{k=0}^{n} f(k) = \int_0^n f(t)dE(t).$$

C'est là que la force de l'intégrale de Stieltjes apparaît : en choisissant soigneusement la fonction p et ses discontinuités, on peut traiter dans un même cadre des phénomènes continus et discrets.

2.3 Applications en probabilités

Considérons une mesure de probabilité P sur un espace mesurable $(\Omega, \mathcal{T})$. Une *variable aléatoire* est une fonction mesurable $X : \Omega \to \mathbb{R}$, autrement dit,

$$\forall\, a \in \mathbb{R},\ X^{-1}(]-\infty, a]) = \{\omega \in \Omega, X(\omega) \leq a\} \in \mathcal{T}.$$

On peut donc définir la *fonction de répartition* de X par

$$F_X : \mathbb{R} \to [0,1],\ a \mapsto P(X(\omega) \leq a)$$

C'est une fonction croissante sur $\mathbb{R}$ (tiens donc).

Maintenant, une quantité importante en probabilités est l'*espérance* $\mathbb{E}[X]$ de la variable aléatoire X. Si X est une variable aléatoire discrète (suivant, par exemple, une loi binomiale, ou de Poisson), on s'attend à avoir

$$\mathbb{E}[X] = \sum_k x_k P(X = x_k)$$

tandis que si X est une variable aléatoire continue et admettant une densité f_X, on s'attend plutôt à

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Or, si X est discrète, alors $P(X = x_k) = P(X \leq x_k) - P(X \leq x_{k-1}) = F_X(x_k) - F_X(x_{k-1})$, où F_X est, dans ce cas, une fonction en escaliers avec des sauts à chaque x_k. D'una autre côté, si f_X est une densité pour X, alors F_X est dérivable et $F_X' = f_X$.

Dans les deux cas, on a donc

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x dF_X(x),$$

où on utilise F_X comme poids de l'intégrale de Stieltjes. Et donc, si on utilise cette formule comme définition de l'espérance d'une variable aléatoire, on peut traiter, dans le même cadre, les variables aléatoires discrètes et continues.

En particulier, en étendant l'intégrale de Stieltjes à $\mathbb{R}^2$, de façon à pouvoir traiter deux variables aléatoires en même temps, on peut définir l'espérance de $X + Y$, où X est une v.a. continue et Y une v.a. discrète, ce qui n'est pas aisé dans le cadre habituel. C'est donc un outil qui permet de simplifier les preuves de résultats théoriques (même si dans la pratique, on calcule les espérances avec de bonnes vieilles intégrales classiques).

2.4 Généralisons toujours plus

De là, on peut partir dans différentes directions :

- Comme pour l'intégrale de Riemann, on peut définir des intégrales généralisées $\int_a^\infty f dp$, par passage à la limite.

- On peut généraliser la fonction poids p : si p n'est pas croissante, mais que $p = p_1 - p_2$ (autrement dit, p est la somme d'une fonction croissante et d'une fonction décroissante) on peut raisnnablement définir une intégrale de Stieltjes de poids p par

$$\int_a^b f dp = \int_a^b f dp_1 - \int_a^b f dp_2$$

De telles fonctions p sont dites "à variation bornée". On a alors :

Théorème 5 (Théorème de représentation de Riesz). *Soit $L : (\mathcal{C}^0([a,b]), \|.\|_\infty) \to \mathbb{R}$ une forme linéaire continue, alors il existe une fonction à variation bornée p telle que, pour tout $f \in \mathcal{C}^0([a,b])$,*

$$L(f) = \int_a^b f dp$$

Autrement dit, n'importe quelle façon raisonnable (c'est à dire linéaire continue) d'associer un réel à une fonction continue peut s'écrire sous la forme d'une intégrale de Stieltjes.

- De même qu'on a généralisé l'intégrale de Riemann en intégrale de Lebesgue via la mesure de Borel, on peut de même construire une mesure sur les boréliens qui généralise cette autre définition de longueur d'intervalle ($p(b) - p(a)$ au lieu de $b - a$). C'est la mesure de Borel-Stieltjes, notée μ_p, à partir de laquelle on peut bien sûr construire une intégrale (l'intégrale de Lebesgue-Stieltjes) qui dispose de toutes les propriétés des intégrales définies à partir d'une mesure.