Sur l' analyse générale IV.

(Entropie dans la cybernétique)

Memoria di Tokui Satō (à Kōbe, Japon)

Résumé. - *Donner un exmple d' applications d' analyse générale comme suite des articles* [3] *et* [4] (1).

1. **Introduction.** - Nous avons donné dans [3] les intégrales

$$\int_S f(X)\, d\mu(X)$$

de fonctions d' ensemble, et leur application dans le calcul de probabilité. Cette intégrale nous amène facilement à considérer celui de fonctions d' ensemble contenant un paramètre : $\int_S f(X, t)\, d\mu(X, t)$, où t est un paramètre réel et appelé le temps. On pourrait alors interpréter cette intégrale comme une sorte de problèmes typiques dans la théorie du processus stochastique.

Cela posé, il est bien connu que l' entropie joue un rôle essentiel dans la cybernétique. A un certain point de vue, nous pourrions considérer une entropie variant avec le temps, c' est-à-dire entropie qui se définit sous la forme $\int_S f(X, t)\, d\mu(X, t)$.

Nous somme ainsi ramenés à un problème de l' intégrale contenant le paramètre qui va être étudié ci-après.

Pour cela, nous devons ajouter quelques compléments au mémoire précédent [3].

Soient S un ensemble quelconque (mais fixe), et $|B$ un corps d' ensemble borelien (sur l' ensemble S).

Dans cet article nous appellerons $\mu(X)$ *mesure dans* $|B$, qui est une fonction σ-additive dans $|B$ telle que

$$-\infty < \mu(X) < +\infty \qquad (X \in |B).$$

(1) Les chiffres dans les crochets renvoient aux références placées à la fin de cet article.

Un *espace de mesure* est le triple : un ensemble S, un corps borelien $\mathbb{B}$ sur S et une mesure $\mu(X)$ dans $\mathbb{B}$. Nous désignons cet espace de mesure par $S(\mathbb{B}, \mu)$.

La mesure $\mu(X)$ a trois propriétés suivantes bien connues.

(I) *La variation totale* $\|\mu\|$ *de* $\mu(X)$ *ou* $\|\mu\| = \sup\limits_{\Delta} \Sigma_{k=1}^{n} |\mu(X_k)|$ *est finie, où* $\sup\limits_{\Delta}$ *représente la borne supérieure par rapport à toute division* Δ *de* S.

(II) (*Décomposition de* JORDAN). - $\mu(X)$ *est une différence de deux mesures non négatives* $\mu^{+}(X)$ *et* $\mu^{-}(X)$, *c'est-à-dire*

$$\mu(X) = \mu^{+}(X) - \mu^{-}(X).$$

(III) (*Décomposition de* HAHN). - *Pour la mesure* $\mu(X)$, *il existe toujours un ensemble* $E(\in \mathbb{B})$ *tel que* $\mu(X) \geqq 0$ *pour tout ensemble* $X \subseteqq E$ $(X \in \mathbb{B})$ *et* $\mu(X) \leqq 0$ *pour tout ensemble* $X \in E^c$ $(X \in \mathbb{B})$.

Soit $f(X)$ une fonction définie dans $\mathbb{E}(\subseteqq \mathbb{B})$. Nous définirons l'intégrale de $f(X)$ par

$$\int_A f(X)\, d\mu(X) = \int_A f(X)\, d\mu^{+}(X) - \int_A f(X)\, d\mu^{-}(X) \qquad (A \in \mathbb{E}).$$

Lorsque $\int_A f(X)\, d\mu(X)$ prend une valeur finie, nous appellerons $f(X)$ *fonction μ-sommable sur A.*

Par définition, nous avons facilement deux propositions suivantes.

(IV) *Soient* C_1 *et* C_2 *deux constantes, et* $\mu_1(X)$, $\mu_2(X)$ *deux mesures dans* $\mathbb{B}$. $\mu(X) = C_1\mu_1(X) + C_2\mu_2(X)$ *est aussi une mesure dans* $\mathbb{B}$.

Si $f(X)$ *est* μ_1*- et* μ_2*-sommable sur* A, *elle est aussi* μ*-sommable sur* A, *et on a*

$$\int_A f(X)\, d\mu(X) = C_1 \int_A f(X)\, d\mu_1(X) + C_2 \int_A f(X)\, d\mu_2(X).$$

(V) *Supposons que, pour tout point* x *de* S, $\{x\}$ *appartienne à* $\mathbb{B}$.

Si $f(X)$ *est une fonction R-intégrable sur* A, *on a*

$$(R)\int_A f(X)\, d\mu(X) = (R)\int_A f(x)\, d\mu(X).$$

2. **Continuité.** - Considérons la continuité de l'intégrale $\int_A f(X, t) d\mu(X, t)$ par rapport à t.

Soit I un intervalle fermé de t.

LEMME. - Soit $\{f_\lambda(t)\}$ *une famille de fonctions continues* $f_\lambda(t)\,(\lambda \in \wedge)$ *dans* I.

Si $\{f_\lambda(t)\}$ *est également continue dans* I, *la fonction* $F(t) = \sup\limits_{\lambda \in \wedge} f_\lambda(t)$ *est uniformément continue dans* I.

En effet, on peut prende $\delta (> 0)$ indépendant de λ tel que

$$|f_\lambda(t_1) - f_\lambda(t_2)| < \varepsilon/2, \quad |t_1 - t_2| < \delta,$$

ε étant un nombre positif donné à l'avance. Par définition, on a

$$F(t_1) \geqq f_\lambda(t_1), \qquad F(t_2) \geqq f_\lambda(t_2),$$

λ étant arbitraire, et il existe λ_1 et λ_2 tels qu'on ait

$$F(t_1) - \varepsilon/2 < f_{\lambda_1}(t_1) \qquad \lambda_1 \in \wedge,$$

$$F(t_2) - \varepsilon/2 < f_{\lambda_2}(t_2) \qquad \lambda_2 \in \wedge.$$

On obient donc

$$|F(t_1) - F(t_2)| < \varepsilon. \qquad \text{C. Q. F. D.}$$

TÉHORÈME 1. - *Soit* $\mu(X, t)$ *une fonction définie dans* $|B \times$ I. *Si* $\mu(X, t)$ *est une mesure dans* $|B$ *pour chaque* $t (\in \mathrm{I})$, *et* $\mu(X, t)$ *considérée comme fonction de* t *est également continue dans* I *pour* $X \in |B$, $\mu^+(X, t)$ *et* $\mu^-(X, t)$ *sont aussi également continues dans* I *pour* $X \in |B$.

En effet, $\mu^+(X, t) = \sup \{\mu(Y, t);\ X \supset Y \in |B\}$, $\mu^-(X, t) = -\inf \{\mu(Y, t);\ X \supset Y \in |B\}$ ayant lieu, d'après le lemme nous avons le théorème.

TÉHORÈME 2. - *Soient* $\mu(X, t)$ *une fonction satisfaisant à l'hypothèse du théorème* 1 *et* $f(X, t)$ *une fonction bornée dans* $|B \times$ I, *et* μ*-sommable dans* $A \cap |B$ $(A \in |B)$ *pour chaque* t $(\in \mathrm{I})$.

Si $f(X, t)$ *considérée comme fonction de* t *est également continue dans* I *pour* $X (\in A \cap |B)$, *l'intégrale* $\int_A f(X, t)\, d\mu(X, t)$ *est continue dans* I.

PREUVE. - En vertu de (IV), on a

$$\left| \int_A f(X, t)\, d\mu(X, t) - \int_A f(X, t_0)\, d\mu(X, t_0) \right|$$

$$\leqq \left| \int_A (f(X, t) - f(X, t_0))\, d\mu(X, t_0) \right|$$

$$+ \left| \int_A f(X, t)\, d(\mu(X, t) - \mu(X, t_0)) \right|$$

Par hypothèse, on peut prendre $\delta (> 0)$ indépendant de $X (\in A \cap \mathfrak{B})$ tel que

$$|f(X, t) - f(X, t_0)| < \varepsilon \qquad |t - t_0| < \delta,$$

où ε est un nombre positif donné à l'avance. D'après le théorème 1, on a, sans perdre la généralité,

$$\| \mu(t) - \mu(t_0) \| < \varepsilon\ (^2) \qquad |t - t_0| < \delta .$$

En vertu de (I) on peut prendre une constante M telle que

$$\| \mu(X, t_0) \| \leqq M .$$

On a donc l'inégalité

$$\left| \int_A f(X, t)\, d\mu(X, t) - \int_A f(X, t_0)\, d\mu(X, t_0) \right| \leqq \varepsilon (F + M)$$

pour $|t - t_0| < \delta$, où $|f(X, t)| \leqq F$. C. Q. *F*. D.

3. **Dérivabilité.** - Considérons la dérivabilité par rapport à t de l'intégrale

$$\int_A f(X, t)\, d\mu(X, t) .$$

Téorème 3. - *Soit $\mu(X, t)$ une fonction dans $\mathfrak{B} \times \mathrm{I}$, qui est une mesure dans $\mathfrak{B}$ pour chaque $t \in \mathrm{I}$, et la dérivée partielle $\mu_t'(X, t)$ est continue par rapport à t.*

Si $\mu_t'(X, t)$ considérée comme fonction de t est une famille normale dans I, *$\mu_t'(X, t)$ est une mesure dans $\mathfrak{B}$ pour chaque $t (\in \mathrm{I})$.*

Preuve. - En vertu de (II) et (III) on peut supposer que $\mu(X, t)$ est non négative dans $\mathfrak{B} \times$ I. Il suffit de montrer que $\mu'_t(X, t)$ est σ-additive.

(²) Nous désignons par $\| \mu(t) \|$ la variation totale de la mesure $\mu(X, t)$ dans $\mathfrak{B}$.

Soient $X_n \in \mathfrak{B}(n = 1.2 \ldots)$ disjoints deux à deux et $X = U_{n=1}^{\infty} X_n$.

$$\mu(X, t) = \mu(U_{n=1}^{\infty} X_n, t) = \Sigma_{n=1}^{\infty} \mu(X_n, t)$$

ayant lieu, d'après le théorème de DINI, la série $\Sigma_{n=1}^{\infty} \mu(X_n, t)$ converge uniformément dans I.

D'autre part, on a $\mu'_t(U_{k=1}^{n} X_k, t) = \Sigma_{k=1}^{n} \mu'_t(X_k, t)$, et $\{\Sigma_{k=1}^{n} \mu'_t(X_k, t)\}$ est normal dans I. On a (³) donc

$$\mu'_t(X, t) = \Sigma_{n=1}^{\infty} \mu'_t(X_n, t) \qquad \text{C. Q. F. D.}$$

THÉORÈME 4. – *Soit* $\mu(X, t)$ *une fonction bornée dans* $\mathfrak{B} \times \mathrm{I}$ *qui est continue par rapport à* t *et une mesure dans* $\mathfrak{B}$ *pour chaque* $t(\in \mathrm{I})$.

Si $\varphi(t)$ *est une fonction continue dans* I, *l'intégrale*

$$\int_a^t \varphi(t) \mu(X, t)\, dt \qquad (a, t \in \mathrm{I})$$

est une mesure dans $\mathfrak{B}$ *pour chaque* $t(\in \mathrm{I})$.

En effet, d'après le théorème d'Arzelà, on a

$$\int_a^t \varphi(t) \mu(U_{n=1}^{\infty} X_n, t)\, dt = \Sigma_{n=1}^{\infty} \int_a^t \varphi(t) \mu(X_n, t)\, dt,$$

où $\quad X_n \in \mathfrak{B} \; (n = 1, 2, \ldots), \; X_i \cap X_j = \emptyset \qquad (i \neq j)$ C. Q. F. D.

Soit $\{f_\lambda(T)\}$ une famille de fonctions dérivables $f_\lambda(t)$ $(\lambda \in \Lambda)$. Lorsqu'on peut prendre un nombre δ indépendant de λ et t_0 tel que

$$\left| \frac{f_\lambda(t) - f_\lambda(t_0)}{t - t_0} - f_\lambda'(t_0) \right| < \varepsilon, \; |t - t_0| < \delta$$

pour $\varepsilon(>0)$ donné à l'avance, nous dirons que la famille $\{f_\lambda(t)\}$ est *également dérivable dans* I.

Par définition $f_\lambda'(t)$ est continue dans I, si $\{f_\lambda(t)\}$ est également dérivable dans I.

THÉORÈME 5. – *Soient* $\mu(X, t)$ *une fonction satisfaisant à l'hypothèse du théorème* 3, *et* $f(X, t)$ *une fonction définie et bornée dans* $A \cap \mathfrak{B} \times \mathrm{I}$ $(A \in \mathfrak{B})$

(³) Voir [5].

est sommable par rapport à la mesure $\mu(X, t)$ *pour chaque* $t(\in \mathrm{I})$.

Si $f(X, t)$ *considérée comme fonction de* t *pour* $X(\in A \cap \mid B)$ *est également dérivable dans* I *et si l'on a*

$$\lim_{t \to t_0} \left\| \frac{\mu(t) - \mu(t_0)}{t - t_0} - \mu_t'(t_0) \right\| = 0,$$

l'intégrale $\int_A f(X, t)\, d\mu(X, t)$ *est dérivable par rapport à* t, *et on a*

$$\frac{d}{dt}\int_A f(X, t)\, d\mu(X, t) = \int_A f_t'(X, t)\, d\mu(X, t) +$$

$$+ \int_A f(X, t)\, d\mu_t'(X, t).$$

PREUVE. - En vertu de (IV), on a

$$\frac{1}{t - t_0}\left\{\int_A f(X, t)\, d\mu(X, t) - \int_A f(X, t_0)\, d\mu(X, t_0)\right\} =$$

$$= \int_A \frac{f(X, t) - f(X, t_0)}{t - t_0}\, d\mu(X, t_0) + \int_A f(X, t)\, d\left(\frac{\mu(X, t) - \mu(X, t_0)}{t - t_0}\right).$$

Par hypothèse, on peut prendre $\delta(>0)$ indépendent de $X(\in \cap \mid B)$ tel que

$$\left| \frac{f(X, t) - f(X, t_0)}{t - t_0} - f_t'(X, t_0) \right| < \varepsilon, \qquad |t - t_0| < \delta,$$

ε étant un nombre positif donné à l'avance. On obient donc

$$\left| \frac{1}{t - t_0}\left\{\int_A f(X, t)\, d\mu(X, t) - \int_A f(X, t_0)\, d\mu(X, t_0)\right\} - \right.$$

$$\left. - \left\{\int_A f_t'(X, t_0)\, d\mu(X, t_0) + \int_A f(X, t)\, d\mu_t'(X, t_0)\right\}\right|$$

$$< \varepsilon F + M \left\| \frac{\mu(t) - \mu(t_0)}{t - t_0} - \mu_t'(t_0) \right\|,$$

où $|f(X, t)| < F, \|\mu(X, t_0)\| < M$.

En vertu de $\lim_{t \to t_0} \left\| \frac{\mu(t) - \mu(t_0)}{t - t_0} - \mu_t'(t_0) \right\| = 0$, il en résulte que

$$\frac{d}{dt} \int_A f(X, t)\, d\mu(X, t) = \int_A f_t'(X, t)\, d\mu(X, t)$$

$$+ \int_A f(X, t)\, d\mu_t'(\mathrm{X}, t). \qquad \text{C. Q. F. D.}$$

D' après les théorèmes 4 et 5, nous avons facilement le théorème suivant.

THÉORÈME 6. - *Soient $\mu(X, t)$ et $f(X, t)$ fonctions satisfaisant aux hypothèses des théorèmes 4 et 5 respectivement.*
Si $\varphi(t)$ est une fonction continue dans I, *on a*

$$\frac{d}{dt} \int_A f(X, t)\, d\left(\int_a^t \varphi(t)\, \mu(X, t)\, dt \right)$$

$$= \int_A f_t'(X, t)\, d\left(\int_a^t \varphi(t)\, \mu(X, t)\, dt \right) + \int_A f(X, t)\, \varphi(t)\, d\mu(X, t).$$

4. Entropie - D' abord rappelons nous la définition de l' entropie dans la cybernétique.

Soient $E_1, E_2, \ldots, E_n$ un nombre fini d' événements et $p_1, p_2, \ldots, p_n$ ses probabilités. Il est bien connu que M. C. E. SHANNON a défini [4] l' entropie de cette répartition de probabilité par

(1) $$H = - \Sigma_{k=1}^{n} (\log_2 p_k) p_k.$$

Soit $S(|B, \mu)$ un espace de probabilité. Nous nous proposons de donner la définitions de l' entropie de la répartition de probabilité $\mu(X)$ par l' intégrale

(2) $$H = - \int_A \log_2 \mu(X)\, d\mu(X).$$

---

[4] Voir [6].

Si l' espace $S(|B, \mu)$ est μ-dense, on a [5]

$$H = -\int_S \log_2 \mu(X)\, d\mu(X) = \mu(S) \log_2 e =$$

$$= \log_2 e.$$

Il est remarquable que cette intégrale est indépendante de $\mu(X)$. Ce qui perd, me semble-*t*-il quelque chose d' essentiel contenu dans l' expréssion (1). Un mémoire [1] de A. J. Chintschin montre [6] implicitement que l' entropie est une sorte d' échelle en des répartitions de probabilité. Pour réaliser concrétement cette assertion, il est bien raisonnable que nous prenons une mesure $\sigma(X)$ comme base. Nous appellerons $\sigma(X)$ *base-répartition de probabilité.*

Soit $\mu(X)$ une mesure dans $|B$.

D' aprés le théorème généralisé de Radon-Nikodym [7], $\mu(X)$ s' exprime sous la forme

$$\mu(X) = \mu(X \cap N) + \int_X m(X)\, d\sigma(X),$$

où $N \in |B$, $\sigma(N) = 0$; $m(X)$ est une (S)-(M)-fonction par rapport à $\sigma(X)$. $\mu(X \cap N)$ étant σ-singulière, si l' espace $S(|B, \sigma)$ de mesure est σ-dense et $m(X)$ est σ-continue, on a [8]

$$\frac{d\mu(X)}{d\sigma(X)} = m(X). \tag{3}$$

En général, nous appellerons donc $m(X)$ *dérivée de* Radon-Nikodym et désignons celui-ci par (3). Nous proposons définir l' entropie de la répartition de probabilité $\mu(X)$ au lieu de (2), par l' intégrale

$$H = -\int_S \log_2 |\mu(X \cap N) + m(X \cap N^c)|\, d\mu(X). \tag{4}$$

D' abord, $\mu(X) = \int_X 1 \; d\mu(X)$ ayant lieu, lorsqu' on prend $\mu(X)$ comme base-répartition de probabilité l' entropie de $\mu(X)$ est nulle, ce qui se raccorde bien avec notre sens.

---

(5) Voir [3].
(6) Voir aussi [2].
(7) Voir [3].
(8) Voir [3].

En particulier, si l'espace $S(\mathfrak{B}, \sigma)$ de mesure est σ-dense, et $m(X)$ est σ-continue, on a

$$(5) \qquad H = -\left\{ \int_N \log_2 \mu(X)\, d\mu(X) + \int_{S-N} \left(\log_2 \frac{d\mu(X)}{d\sigma(X)} \right) \frac{d\mu(X)}{d\sigma(X)}\, d\sigma(X) \right\}.$$

Ensuite, soit $f(x)$ une fonction non-négative et continue dans R^1 telle que

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1\,.$$

Beaucoup de mathématiciens [9] définissent l'entropie de la répartition

$$\mu(x) = \int_{-\infty}^{x} f(x)\, dx$$

de probabilité par l'intégrale

$$(6) \qquad H = -\int_{+\infty}^{-\infty} (\log_2 f(x))\, f(x)\, dx\,.$$

Dans ce cas en vertu de (V), (4) (ou (5)) se réduit à (6).

A ce point de vue, nous pouvons considérer une entropie pour ainsi dire, variant avec le temps.

Soient $\mu(X, t)$ une fonction non négative dans $\mathfrak{B} \times \mathrm{I}$ qui est une mesure ($\mu(S, t) = 1$) dans $\mathfrak{B}$ pour chaque $t(\in \mathrm{I})$, et $\sigma(X)$ une base-répartition de probabilité. On obtient alors

$$H(t) = -\int_S \log_2 \{(\mu(X \cap N, t) + m(X \cap N^c, t)\}\, d\mu(X, t)\,,$$

où

$$\mu(X, t) = \mu(X \cap N, t) + \int_X m(X, t)\, d\sigma(X)\,.$$

REMARQUE. - Le théorème 4 a des rapport intimes avec l'intégrale de WIENER dans la théorie du processus stochastique.

---

[9] Voir [7].

Soient I un intervalle $[a, b]$ sur R^1 avec la mesure ordinaire de LEBESGUE λ et $\varphi(t)$ une fonction sommable dans I. On voit aisément que le théorème 4 subsiste aussi au cas où $\mu(X, t)$ est une fonction bornée dans $|B \times \mathrm{I}$ qui est une mesure dans $|B$ pour chaque $t(\in \mathrm{I})$ et sommable par rapport à t.

Soit $\sigma(X)$ une base-répartition de probabilité dans $|B$. D'apèrs le théorème généralisé de RADON-NIKODYM, on a

$$\mu(X, t) = \mu(X \cap N, t) + \int_X m(X, t)\, d\sigma(X),$$

o ù $N \in |B$, $\sigma(N) = 0$, $m(X, t)$ est une (S)-(M)-fonction dans $|B$ par rapport à $\sigma(X)$ pour chaque $t(\in \mathrm{I})$.

Soient $S(|B, \sigma)$ et $S \times \mathrm{I}(|B \times \mathrm{I}, \sigma \times \lambda)$ deux espaces bornés et complets

D'après le théorème généralisé de FUBINI (10), si $m(X, t)$ est une (S)-(M)-fonction non négative dans $|B \times \mathrm{I}$ et $\varphi(t)$ est non négative, on a

$$\int_a^t \varphi(t)\mu(X, t)\,dt = \int_a^t \varphi(t)\mu(X \cap N, t)\,dt + \int_a^t dt \int_X \varphi(t) m(X, t)\, d\sigma(X)$$

$$= \int_a^t \varphi(t)\mu(X \cap N, t)\,dt + \int_X d\sigma(X) \int_a^t \varphi(t) m(X, t)\,dt.$$

RÉFÉRENCES

[1] A. J. CHINTSCHIN, *La notion de l'entropie dans le calcul de probabilité, y C π.* Matem. HAYK, T. **8**, B. 3 (55), (1953), 3-20, (en russse).

[2] D. K. FADDEJEW, *A la notion de l'entropie d'un nombre fini de schemas stochatiques,* ibid., T. **11**, B. 1 (66), (1956), 227-231, (en russe).

[3] T. SATO *Sur l'analyse générale* (*Théorie des fonctions d'ensemble*), Annali di Math., **47** (1959), 251-317.

[4] ——, II (*Intégrales d'applications de treillis*), ibid., **52** (1960), 363-383. III (*Différentiation, d'applications de fonctions*) ibid., **59** (1962), 55-64.

[5] ——, *Pri la limo de funkcisekvajho,* Mem. Fac. Sc., KYUSYU Univ., A, **4** (1949), 23-27,

[6] C. E. SHANNON, *A mathematical theory of communication*, Bell Syst. TECHN J. **27** (1948) 379-423, 623-656.

[7[N. WIENER, *Cybernetics,* (The technology press), (1948).

(10) Voir [3].