SÉMINAIRE DE PROBABILITÉS (STRASBOURG)

ALBERT TORTRAT

Désintégration d'une probabilité. Statistiques exhaustives

Séminaire de probabilités (Strasbourg), tome 11 (1977), p. 539-565
<http://www.numdam.org/item?id=SPS_1977__11__539_0>

Laboratoire de Probabilités
Associé au C.N.R.S. N° 224

DESINTEGRATION D'UNE PROBABILITE,

STATISTIQUES EXHAUSTIVES

par

A. TORTRAT

1. Introduction.

Ce texte est surtout didactique. Tout en essayant de donner une vue relativement complète du problème de la désintégration en liaison avec "mesures parfaites", "mesures compactes au sens abstrait", et la construction de mesures sur les espaces produits, il vise à montrer qu'il y a deux théorèmes (pas très différents) de désintégration, qui contiennent l'essentiel du sujet.

Le plus simple des deux suffit déjà à presque tous les besoins et date de 1954 [1]. Le deuxième, suivant une étude assez fournie de M. Valadier (cf.[9]) est de 1971 (cf.[6]) . Mentionnons l'exposé qu'en fait M. Chatterji dans [1] , et notons, pour exemple, l'extension (légère) de [8] .

Les éléments des n^os 2 à 5 sont dans [5], mais paraissent encore peu connus. Il nous semble utile de les reprendre (avec du recul) sous une forme plus complète [1].

---

(1) L'article [7] est paru en Russe (traduit à l'Institut Henri Poincaré). Nous en avons dans [5] simplifié la présentation qui faisait appel à un double encadrement, inutile dans le cas d'une algèbre. Ce théorème est rarement donné dans sa généralité. Le fait (conséquence de la preuve) que la même classe compacte $\mathcal{C}$, dénombrable, approche une algèbre dénombrable - pour la loi désintégrée et ses désintégrées - (donc $\mathcal{C}_\delta$ de même pour la tribu) est important, et il est légitime d'attribuer à Jirina la part essentielle qui lui revient, sans la tronquer.

Notre point de vue est qu'on doit utiliser beaucoup plus systématiquement les "vraies" probabilités conditionnelles, en statistique en particulier (cf. le § 9) : elles existent en général à condition de ne pas vouloir poser les définitions sur un triplet $(\Omega,\mathcal{A},P)$ informel, beaucoup trop fruste (cf. le § 7).

2. Position des problèmes.

Le problème de la désintégration n'a un sens concret, le plus souvent, que posé dans le cadre d'un espace produit $\mathcal{T} \times \mathcal{Y}$ (muni de la tribu produit $\mathcal{E} \times \mathcal{B}$, pour l'instant). La question est alors exactement d'étendre la formule d'intégrations successives (intégrale double de Fubini), c'est à dire d'écrire la probabilité élémentaire

(1) $\quad P(dt \times dy) = P(dt)\, P_t(dy).$

Dans (1) $P(dt)$ est la loi marginale (de la v.a. T, à valeurs dans l'espace $\mathcal{T}$), et $P_t(.)$ [2] une famille de lois sur $\mathcal{B}$, dépendant mesurablement de t, en ce sens que chaque $P_t(B)$, $B \in \mathcal{B}$ est une fonction $\mathcal{E}$ - mesurable de t.

On peut bien sûr, penser les v.a. T et Y comme des fonctions à valeurs dans un $(\Omega,\mathcal{A},P)$ abstrait (c.à.d. inconnu, n'ayant d'autre propriété que la σ-additivité de P sur $\mathcal{A}$). En fait dans tout problème "concret" la probabilité P est transportée dans un espace produit (défini par toutes les v.a. considérées [3]) puis réduite à un sous-espace de bonnes trajectoires, spécialement lorsque le produit en question est non dénombrable : à ce 2ème stade on peut retrouver des difficultés, mais seulement si cet espace n'est pas polonais.

---

(2) Il n'est légitime (et encore) de les appeler probabilités de transition que si $\mathcal{T}$ et $\mathcal{Y}$ sont les mêmes. Ce sont les lois conditionnelles des $(Y|t)$, notées $(Y|t)$.

(3) ou par une "base linéaire" de l'ensemble de ces v.a., par exemple, si celles-ci sont réelles.

T est la v.a. de conditionnement; ce peut être une fonction de la v.a. échantillon Y , alors T s'appelle une statistique et la considération de la loi $(Y|t)$ de Y conditionnée par $\{T=t\}$ est habituelle en statistique appliquée (dans le cas ici précisé : pour la définition des statistiques exhaustives).

Sous la forme (1) nous désintégrons en fait la loi de Y . Si $E \in \mathcal{C} \times \mathcal{B}$, on a

(2) $$PE = \int P(dt)\, P_t(E_t) \quad , \quad E_t \text{ section de } E \text{ par } t \ ,$$

comme on le prouve immédiatement par le même raisonnement que pour la preuve du théorème de Fubini. Suivant (2), on peut écrire

(2') $$P(E|t) = P_t(E_t) \ .$$

C'est dire que la seule désintégration de la loi de Y , par rapport à t fournit celle de P elle-même (dans $\mathcal{T} \times \mathcal{Y}$) , et que $P(E|t)$ ne dépend en fait que de la section de E par t .

Le problème abstrait, de désintégration de P sur $(\Omega,\mathcal{A})$, par rapport à la v.a. T (à valeurs dans un espace quelconque) se ramène au précédent en transportant P dans $\mathcal{T} \times \Omega$, sur la tribu $\mathcal{C} \times \mathcal{A}$. Comme lorsque T est une fonction de la seule v.a. Y (ici de ω) se pose l'exigence qui en général ne peut être satisfaite : que P_t soit portée par l'ensemble $\{\omega : T(\omega) = t\}$ (cf. la proposition 5).

La désintégration de Y par rapport à T se posera, abstraitement, dans Ω , sous la forme de la définition d'une famille P_t de lois sur la sous-tribu $\mathcal{B}$ liée à Y , soit $\bar{\omega}(A,\omega)$; qui soit $\mathcal{A}_T$ - mesurable en ω , en désignant ici par $\mathcal{A}_T$ et $\mathcal{A}_Y$ les sous-tribus de $\mathcal{A}$ images inverses (pour T et Y respectivement) dans Ω des tribus $\mathcal{C}$ et $\mathcal{B}$ dans les espaces $\mathcal{T}$ et $\mathcal{Y}$. $\bar{\omega}(A,\omega)$ ne dépend donc que de $T(\omega)$, si $\mathcal{A}_T$ est liée à une v.a. T (dès qu'une tribu $\mathcal{A}'$ est à base dénombrable on peut la rattacher à une v.a. réelle).

3. Théorème 1 : Si dans $(\Omega,\mathcal{A},P)$, $\mathcal{A}'\subset\mathcal{A}$ est engendrée par $\{A_i\}$, algèbre dénombrable séparant les points de Ω , il n'existe de probabilité conditionnelle, par rapport à $\mathcal{A}'$, sur aucune sur-tribu vraie de $\mathcal{A}'$, c.à.d. sur aucune tribu $\mathcal{B}\supset\mathcal{A}'$, qui ne soit pas contenue dans $\mathcal{A}'_p$ tribu complétée de $\mathcal{A}'$ par rapport à P .

Preuve : Soit $\bar{\omega}(\omega,.)$ une famille $\mathcal{A}_T$-mesurable de lois sur $\mathcal{B}$. Hors un $N\in\mathcal{A}_T$, P-nul, on a $\bar{\omega}(\omega,A_i) = 1_{A_i}(\omega)$ (puisque $P^{\mathcal{A}'}A_i \equiv 1_{A_i}$ vu $A_i\in\mathcal{A}'$) . Pour ces $\omega\notin N$, la loi $\bar{\omega}(\omega,.)$ est définie sur $\mathcal{A}'$ par ses valeurs sur l'algèbre $\{A_i\}$, et comme $\mathcal{A}'$ sépare les points de $\Omega(\{\omega\}\in\mathcal{A}')$, c'est $\delta(\omega)$. C'est donc aussi $\delta(\omega)$ sur $\mathcal{B}$, soit

(3) $\qquad \bar{\omega}(\omega,B) = 1_B(\omega)$, tout $\omega\notin N$, tout $B\in\mathcal{B}$.

Ce serait dire, puisque $\bar{\omega}(\omega,B)$ est $\mathcal{A}'$-mesurable que B équivaut à un élément de $\mathcal{A}'$, modulo un N , P-nul, de $\mathcal{A}'_p\big((B-N) + \{\omega : \omega\in N,\bar{\omega}(\omega,B) = 1\}\in\mathcal{A}'_p\big)$. ∎

Corollaire 1 : On suppose que pour toute sous-tribu $\mathcal{A}'$, il existe une désintégration $\bar{\omega}(\omega,.)$ sur $\mathcal{A}$ par rapport à $\mathcal{A}'$. Alors, s'il existe dans $\mathcal{A}$ une partie dénombrable $\{A_i\}$ séparant les points ω , cette partie engendre $\mathcal{A}$: engendre $\mathcal{A}'_p$ contenant $\mathcal{A}$. En effet prenant $\mathcal{A}'$ engendrée par ces A_i , et $\mathcal{B} = \mathcal{A}$, le théorème assure $\mathcal{B}\subset\mathcal{A}'_p$.

Remarque 1 : Le théorème ci-dessus formalise l'exemple dû à Dieudonné (cf. [5] p.232), antérieur au théorème qui suit. Des conséquences en seront tirées après ce théorème (de Jirina).

4. <u>Théorème 2</u> (Jirina) : Soit dans $(\Omega,\mathcal{A},P)$, $\mathcal{A}_o$ une sous-tribu engendrée par l'algèbre dénombrable $\mathcal{Q}_o = \{A_n\}$; on suppose P compacte sur $\mathcal{A}_o$ (au sens abstrait, c'est-à-dire qu'il existe $\mathcal{C} \subset \mathcal{A}_o$, classe compacte approchant $\mathcal{A}_o$ p.r. à P : $A \in \mathcal{A}_o \Rightarrow PA = \sup_{C \subset A} PC$) .

a) Alors pour toute sous-tribu $\mathcal{B}$ de $\mathcal{A}$, il existe une désintégration, par rapport à $\mathcal{B}$, de P sur $\mathcal{A}_o$.

b) Les $\overline{\omega}(\omega,A)$ (famille $\mathcal{B}$-mesurable de lois sur $\mathcal{A}_o$) sont toutes compactes, avec la même classe $\mathcal{C}$ (si $\mathcal{C} = \mathcal{C}_\delta$: est stable pour les intersections dénombrables, ce qu'on peut toujours supposer).

a')Si $\mathcal{B}$ est seulement dans $\mathcal{A}$, la conclusion a) subsiste, mais non b).

<u>Preuve.1</u> : Soit $PA_n = \lim \uparrow PC_{ni}$ avec $C_{ni} \uparrow$, $C_{ni} \subset A_n$, $C_{ni} \in \mathcal{C}$.

Soit $\mathcal{Q}$ l'algèbre (dénombrable) engendrée par $\mathcal{Q}_o$ et l'ensemble des C_{ni} , et $\overline{\omega}(\omega,A)$ un choix de $P^{\mathcal{B}}(A)$ fait pour tout $A \in \mathcal{Q}$ (valant 1 pour $A = \Omega$, 0 pour $A = \emptyset$) .

On voit aisément que, hors d'un N élément P-nul de $\mathcal{B}$, les $\overline{\omega}(\omega,A)$ sont additives sur $\mathcal{Q}$ et satisfont à

$$\sup_i \overline{\omega}(\omega,C_{ni}) = \overline{\omega}(\omega,A_n) \quad , \quad \text{tout } n \, . \tag{4}$$

En effet, posant

$$\lim_{i\uparrow\infty} \uparrow \overline{\omega}(\omega,C_{ni}) = \overline{\omega}'(\omega,A_n) \leqq \overline{\omega}(\omega,A_n) \ ,$$

on a

$$\int \overline{\omega}'(\omega,A_n)\, dP = \lim \uparrow PC_{ni} = PA_n = \int \overline{\omega}(\omega,A_n)\, dP \ .$$

(4) assure la σ-additivité sur $\mathcal{Q}_o$ de ces $\overline{\omega}(\omega,.)$ $(\omega \notin N)$, donc leur prolongement à $\mathcal{A}_o$.

Pour $\omega \in N$, on posera (par exemple, c'est indifférent) $\overline{\omega}(\omega,.) = P$. Cela assure a'). En effet la formule (les $\overline{\omega}(\omega,A)$, $A \in \mathcal{A}_o$ sont $\mathcal{C}$ -mesurables si elles le sont pour $A \in \mathcal{Q}_o$)

$$\mu(AB) = \int_B \overline{\omega}(\omega,A)\, dP \ , \ B \in \mathcal{C} \ , \ A \in \mathcal{A}_o \ ,$$

définit pour tout $B \in \mathcal{C}$ fixé, une fonction σ-additive de $A \in \mathcal{C}_o$ qui égale $P(AB)$ sur $\mathcal{Q}_o$ donc aussi sur $\mathcal{A}_o$. $\overline{\omega}(\omega,.)$ est bien la désintégration cherchée.

2 : Supposons $\mathcal{E} \subset \mathcal{A}_o$. Alors (il est classique que) $\mathcal{E}_\delta$ ($= \mathcal{E}$ par hypothèse) approche $\mathcal{A}_o$ relativement à chaque $\overline{\omega}(\omega,.)$ (tout $\omega \notin N$) , puisque $\mathcal{E}$ approche $\mathcal{Q}_o$.

Soit en effet (ω fixé) B de $\mathcal{A}_o$, et $A = \bigcap_j A^j$, $A^j \in \mathcal{Q}_o$ tel que $\overline{\omega}A > \overline{\omega}B - \varepsilon$ et $A \supset B$.

Si $\overline{\omega}C^j > \overline{\omega}A^j - \varepsilon/2^j$ avec $C^j \subset A^j$,

on a $C = \bigcap_j C_j \subset B$ et $\overline{\omega}C > \overline{\omega}B - 2\varepsilon$. ■

<u>Corollaire 2</u> : Pour désintégrer P sur $(\mathcal{T},\mathcal{C}) \times (\mathcal{Y},\mathcal{B})$ par rapport à $\mathcal{C}$, il suffit que $\mathcal{B}$ soit à base dénombrable et que la loi P (marginale) sur $\mathcal{B}$, soit compacte.

<u>Remarque 2</u> : Pour $\mathcal{Y} = R^{\mathbb{N}}$, une preuve directe est plus simple, se ramène à $Y = R$. Cette preuve directe vaut encore pour $\mathcal{Y}$ polonais, car $\mathcal{Y}$ est alors identifiable (topologiquement à un G_δ de $R^{\mathbb{N}}$: une intersection dénombrable d'ouverts de $R^{\mathbb{N}}$) . (Cf. le § 8).

5. Mesures parfaites.

Définition 1 : Une probabilité (ou mesure ≥ 0 , bornée, sur une tribu) est dite parfaite si elle est compacte sur toute sous-tribu à base dénombrable.

Corollaire 3 (du théorème de Jirina) : Dans Ω , espace métrique séparable, une loi P définie sur une tribu $\mathcal{A}$ contenant les ouverts n'est parfaite que si $\mathcal{A} \subset \mathcal{B}_P$ tribu borélienne complétée. Il en est de même si $(\Omega,\mathcal{A})$ étant quelconque, il existe $\mathcal{B}$ à base dénombrable séparant les points de Ω .

Preuve : Soit $A \in \mathcal{A} - \mathcal{B}_P$, et $\mathcal{A}_o = \sigma(A,\mathcal{B})$, tribu engendrée par A et $\mathcal{B}$.

Suivant le théorème, $\overline{\omega}(\omega,.)$ existerait sur $\mathcal{A}_o$, désintégration de P par rapport à $\mathcal{B}$, contredisant le théorème 1 (pour $\mathcal{A}' = \mathcal{B}$, $\mathcal{A}_o$ surtribu vraie de $\mathcal{A}'$) .

Lemme 1 : Si P est parfaite sur $(\Omega,\mathcal{A})$, toute v.a. X à valeurs dans $\mathcal{X}$, induit sur la tribu image $X\mathcal{A}$ une loi μ image de P , qui est parfaite. En particulier, si X est mesurable pour une tribu $\mathcal{B}$ de $\mathcal{X}$ séparant les points et à base dénombrable, alors on a $X\mathcal{A} \subset \mathcal{B}_\mu$, vu le corollaire qui précède.

Preuve : Soit $\mathcal{A}' \subset X\mathcal{A}$, à base dénombrable, et $\mathcal{E} \subset X^{-1}\mathcal{A}'$ une classe compacte approchant $X^{-1}\mathcal{A}' \subset \mathcal{A}$ ($X^{-1}\mathcal{A}'$ est à base dénombrable avec $\mathcal{A}'$) p.r. à P (suivant l'hypothèse P parfaite). La classe $\mathcal{E}'$ image de $\mathcal{E}$, par X , approche $\mathcal{A}'$ p.r. à μ et est compacte :

si $\bigcap_1^\infty C'_i = \emptyset$, $C_i = X^{-1} C'_i \in \mathcal{E}$ et $\bigcap_1^\infty C_i = \emptyset$,

donc une $\bigcap_I C_i$ est vide (I fini) et $\bigcap_I C'_i = \emptyset$. ∎

On notera que les C_i sont saturés (pour $X(\omega)$) et que $\mathcal{E}'$ est l'image directe de $\mathcal{E}$ (contrairement à $X\mathcal{A}$) : $\mathcal{E}'$ est dans $X\Omega$ et $\mathcal{E}' \subset X\mathcal{A}$.

<u>Proposition 1</u> : P sur $(\Omega,\mathcal{A})$ est parfaite si et seulement si $X\Omega \in \mathcal{B}_\mu$ pour <u>toute</u> v.a. réelle X .

P est compacte sur $(\Omega,\mathcal{A}_o)$, $\mathcal{A}_o$ à base dénombrable, si cette condition est vraie pour <u>une</u> v.a. X réelle, ou à valeurs dans un espace polonais, engendrant $\mathcal{A}_o$.

<u>Preuve</u> : Soit $\{A_i\}$ une partie dénombrable de $\mathcal{A}$, engendrant la tribu $\mathcal{A}_o$ et X la v.a.

(5) $$X(\omega) = 2 \sum_1^\infty 1_{A_i} / 3^i .$$

X prend ses valeurs dans l'ensemble de Cantor habituel C , du segment $[0,1]$, et $\mathcal{A}_o$ est en correspondance biunivoque avec la trace $\mathcal{B}_E$ de $\mathcal{B}(R)$ sur l'image $E = X\Omega$ (et égale $X^{-1}\mathcal{B}$) . <u>Suivant le lemme 1</u>, $E \in \mathcal{B}_\mu$ est condition nécessaire de "P parfaite". <u>Si inversement</u> $E \in \mathcal{B}_\mu$, (pour une X réelle quelconque engendrant $\mathcal{A}_o$) on a $E \supset E_o$, $E_o \in \mathcal{B}$, $\mu(E-E_o) = 0$, et $\mathcal{B}_E$ est approchée par la classe $\mathcal{E}'$ des compacts $\subset E_o$, relativement à μ . La classe $\mathcal{E} = X^{-1}\mathcal{E}'$ est dans $\mathcal{A}_o$, approche $\mathcal{A}_o$ p.r. à P et est compacte : si $\bigcap C_i = \emptyset$ et $C_i = X^{-1} K_i$, $\bigcap K_i = \emptyset$, <u>car</u> ces K sont dans E , donc une même intersection finie de C_i et K_i est vide. La preuve est identique dans le cas où X prend ses valeurs dans un espace polonais.

Corollaire 4 (Sazonov) : Si Ω est métrique séparable, $\mathcal{A}$ sa tribu borélienne, P parfaite sur $(\Omega,\mathcal{A})$ est de Radon : $\mathcal{A}$ est approchée (relativement à P) par la classe (compacte) des compacts de Ω .

Preuve : Suivant le lemme 1, appliqué au plongement de Ω dans son complété $\hat{\Omega}$, on a $\Omega \in \hat{\mathcal{A}}_{\hat{P}}$, où $\hat{P}$ est la loi induite par P sur la tribu borélienne $\hat{\mathcal{A}}$ de $\hat{\Omega}$.

$\hat{\Omega}$ étant polonais, $\hat{P}$ est, on le sait de Radon, et (comme ci-dessus pour E), c'est dire que $\mathcal{A}$ étant la trace de $\hat{\mathcal{A}}$ sur Ω , et Ω approché par des compacts K_ε de $\hat{\Omega}$ (donc compacts dans Ω) relativement à $\hat{P}$, P est de Radon.

Remarque 3 : Il n'est pas évident qu'une mesure compacte est parfaite, même si $\mathcal{A}$ est à base dénombrable, car si $\mathcal{C}$ approche $\mathcal{A}$, cela n'assure pas que les éléments de $\mathcal{C}$ approchant ceux de $\mathcal{A}_o$ à base dénombrable $\subset \mathcal{A}$, sont (peuvent être choisis) dans $\mathcal{A}_o$. Ce qui suit le montrera, et fournira aussi une preuve de la proposition de Sazonov et de la proposition 1, indépendante de la précédente (qui utilisait le théorème de Jirina, par le lemme 1).

Proposition 2 : Si P sur $(\Omega,\mathcal{A})$ est compacte, ou parfaite, elle est quasi-compacte, c'est-à-dire que pour toute suite A_n de $\mathcal{A}$, et tout $\eta > 0$, il existe A (de $\mathcal{A}$) tel que :

(6) $\quad PA > 1-\eta$ et la classe $\{AA_n\}$ est compacte.

Preuve : Soit $C_n \subset A_n$ avec $\mu(A_n - C_n) < \eta/2^{n+1}$. Ces C_n sont pris dans la classe $\mathcal{C}$ relative à P et soit à $\mathcal{A}$ (si P est compacte) soit à la tribu engendrée par les A_n si P est parfaite.

De même on approche les A_n^c par C_n' (4) :

$$C_n C_n' = \emptyset \quad \text{et} \quad P\{\Gamma_n = C_n + C_n'\} > 1 - \eta/2^n .$$

On a donc

$$P(\Gamma \cup \Gamma_n^c) < \eta \Rightarrow P(A = \bigcap \Gamma_n) > 1 - \eta .$$

A répond à la proposition car $A_n C_n' = \emptyset \Rightarrow A_n A \subset A_n C_n \subset C_n \subset A_n$ et

(7) $$C_n A \subset A_n A \subset C_n A \Rightarrow AA_n = C_n \bigcap_{i \neq n} (C_i + C_i') .$$

(7) assure la compacité de $\{AA_n\}$ partie de $\mathscr{C}_{s\delta}$ (ensemble des $\bigcap$ dénombrables d'unions finies d'éléments de $\mathscr{C}$) . ∎

<u>Proposition 3</u> : Si P sur $(\Omega, \mathscr{A})$ est quasi-compacte elle satisfait au critère de la proposition 1 donc est parfaite. Ainsi si $\mathscr{A}$ est à base dénombrable, "P compacte" équivaut à "P parfaite".

<u>Preuve</u> : Soit $\{I_n\}$ l'ensemble des intervalles ouverts de R à extrémités dyadiques, et $X(\omega)$ une v.a. réelle. Associons aux η_i d'une suite $\downarrow 0$, les A de la proposition 2 , notés A^i , relatifs aux $A_n = X^{-1} I_n$. <u>Chaque</u> XA^i <u>est fermé</u> car si a est un réel, adhérent à XA^i , $a = \lim \downarrow I_{n_k}$ (pour une sous-suite convenable $\{n_k\}$), avec $I_{n_k} \cap XA^i \neq \emptyset$ pour tout k . Les $A_{n_k} A^i \downarrow$ (avec 1/k) étant non vides, $\bigcap_k A_{n_k}$ coupe A^i (car pour i fixé, la classe $A_n A^i$ est compacte) donc $a \in XA^i$. Puisque $\bigcup_i XA^i$ est un borélien dans R de μ-mesure 1 (μ image de P), on a bien $X\Omega \in \mathscr{B}_\mu$. ∎

---

(4) L'intersection est parfois notée par l'absence de signe, lorsqu'il n'y a pas ambiguité. + désigne une réunion à éléments disjoints, - une différence (propre ou non).

Preuve directe du théorème de Sazonov : Soit $\Omega = \sum_{i=1}^{\infty} A_{ni}$ un partage de Ω en boréliens de diamètres $< 2^{-n}$, et A'_k un ordre de l'ensemble des A_{ni}. Soit $X(\omega)$ l'application (5) définie par les A'_k, de Ω dans l'ensemble C de Cantor (dans $[0,1]$).

Si P est compacte sur la tribu borélienne de Ω, on a $E = X\Omega \in \mathcal{B}_\mu$ (vu les propositions 2 et 3). Mais X^{-1} est continue sur E car si $a_n \to a$, dans E, pour $n \geq N_K$, les coordonnées triadiques correspondantes sont égales jusqu'au rang K. Si $X^{-1} a \in A_{\ell i_\ell}$ (pour tout ℓ), $X^{-1} a_n$ y appartient aussi (tout $n \geq N_K$) pour les $\ell = 1,2,\ldots \ell_o$ tels que les $A_{\ell i_\ell}$ soient des A'_k de rang $\leq K$. Puisque $\ell_o \uparrow \infty$ avec K, cela prouve la continuité de X^{-1}.
La classe compacte de la proposition 1 (image par X^{-1} de compacts $\subset E$) est donc formée de compacts de Ω. ∎

Proposition 4 : Une condition nécessaire et suffisante pour qu'il existe sur la tribu borélienne $\mathcal{A}$ de Ω métrique séparable une loi non compacte est que Ω ne soit pas universellement mesurable dans son complété $\hat{\Omega}$, c'est-à-dire qu'il existe une loi μ sur la tribu borélienne $\mathcal{A}$ de $\hat{\Omega}$ telle que $\Omega \notin \hat{\mathcal{A}}_\mu$.

Preuve : Si la trace P de μ sur $(\Omega, \mathcal{A})$ était compacte, on devrait avoir suivant la proposition 3, $\Omega \in \hat{\mathcal{A}}_\mu$, donc la condition de l'énoncé suffit. Elle est nécessaire, car si $\Omega \in \hat{\mathcal{A}}_\mu$, μ est de Radon (suivant la première preuve du théorème de Sazonov).

Remarque 4 : La séparabilité de Ω dans le théorème de Sazonov peut être remplacée par l'hypothèse P τ-régulière, qui dans un espace métrique (et sur la tribu borélienne $\mathcal{A}$) équivaut à "le support-topologique, toujours séparable de P est de P-mesure 1 ".

En fait toute loi sur un tel $(\Omega,\mathcal{A})$ est τ-régulière (moyennant les axiomes, qu'on sait admissibles, du choix, du continu, et d'accessibilité de tout cardinal). Ainsi sur $\mathcal{A}$ tribu borélienne d'un espace métrique, la compacité "abstraite" de P équivaut à sa compacité au sens "de Radon".

Remarque 5 : Le théorème de Jirina est simple, et essentiel. On ne peut faire beaucoup mieux, seulement lever l'hypothèse $\mathcal{A}_o$ à base dénombrable en (renforçant) remplaçant P compacte, sur $\mathcal{A}_o$, par P de Radon. Nous donnons ce théorème dans le cadre d'un espace produit (auquel on peut toujours se ramener) pour la clarté.

L'hypothèse "P compacte sur $\mathcal{A}_o$" (du théorème de Jirina) ou celle, plus forte "P parfaite sur $\mathcal{A}$" ne peut être levée, comme le montre le théorème 1 (avec Ω métrique séparable, a fortiori si Ω est abstrait), même si $\mathcal{A}_o$ est liée à une v.a. réelle Y (ce qui est réalisable pour $\mathcal{A}_o$ à base dénombrable, suivant (5)). Ainsi la désintégration dans l'espace $\mathcal{T} \times R$, R portant Y, de la loi de Y ne peut se transporter dans Ω, cf. le § 7.

6. Le théorème "savant" de désintégration.

Théorème : Soit P une probabilité dans l'espace $\mathcal{T} \times \mathcal{Y}$, sur la tribu produit $\mathcal{C} \times \mathcal{B}$. Notons μ et ν les lois marginales P_T et P_Y, et supposons $\mathcal{Y}$ espace topologique, $\mathcal{B}$ sa tribu borélienne, ν de Radon et $\mathcal{C}$ complète (pour μ).

Alors il existe une désintégration de ν en la famille ν_t de lois de Radon

(8) $\quad P(dt \times dy) = \mu(dt)\ \nu_t(dy)$.

Preuve : Choisissant des compacts K_i disjoints dans Y , portant ν , on voit qu'il suffit de prouver le théorème pour Y compact (et les $\nu_t K_i$ seront égales à νK_i , dans le cas général).

A. Soit $f \in \mathcal{C}^+ = \mathcal{C}^+(Y)$, c'est-à-dire continue ≥ 0 sur Y .

$P(1_A f) = \int_{A\times\mathcal{Y}} f(y)\, P(dt\times dy)$, $A \in \mathcal{T}$, est une mesure sur $\mathcal{C}$ majorée par νf . Elle admet une densité p.r. à μ , bornée par νf , soit $h_f(t)$ de $\mathcal{L}^\infty(\mathcal{T}, \mathcal{C})$. Soit $h \to \tilde{h}$ un relèvement de $\mathcal{L}^\infty$ (5) . Puisque $h \underset{\text{p.s.}}{\leq} h' \Rightarrow \tilde{h} \leq \tilde{h}'$ partout, $\tilde{h}_f(t)$ est, pour chaque t , une fonction linéaire positive sur $\mathcal{C}^+$, ainsi est défini l'opérateur linéaire ν_t sur $\mathcal{C}(Y)$ (mesure de Radon au sens de Bourbaki) désintégrant les νf :

$$(9) \qquad \nu f = \int_{\mathcal{T}} \nu_t f\ \mu(dt) \ , \qquad (9') \qquad \tilde{h}_f(t) = \nu_t\ f \ ,$$

$$P(1_A\ f) = \int_A \nu_t\ f\ .\ \mu(dt)\ .$$

Le problème qui demeure est d'étendre (9), aux ouverts $\mathcal{O}$ dans Y , puis (c'est élémentaire) aux boréliens.

B. Soit $\mathcal{O}$ un ouvert dans Y et $0 \leq f_\alpha \uparrow 1_{\mathcal{O}}$, une famille filtrante $\uparrow$ de $\mathcal{C}^+$ d'enveloppe $1_{\mathcal{O}}$.

Posons $h_\alpha(t) = \nu_t\ f_\alpha$ $\left(h_\alpha = \tilde{h}_\alpha\right)$ et $h_\alpha(t) \uparrow \nu_t \mathcal{O} = h(t) \leq 1$.

---

(5) Cf. [3] . Un relèvement linéaire suffirait ; rappelons qu'<u>à la classe d'équivalence</u> de h on fait correspondre un élément h de cette classe, à 1 correspond 1 , à $h \geq 0$ p.s. correspond $\tilde{h} \geq 0$, et $ah + bh'$ se relève en $a\tilde{h} + b\tilde{h}'$.

Nous ne savons pas encore h mesurable. Mais considérons les mesures

$$P(1_A f_\alpha) = \mu_\alpha A = \int_A h_\alpha \, d\mu \leq \mu A \quad , \quad A \in \mathscr{C} .$$

On sait (et il est nécessaire de le prouver) que la borne supérieure des μ_α, comme fonction additive sur $\mathscr{C}$, des μ_α additives (borne éventuellement infinie pour une famille μ_α quelconque)

i) est définie par $\mu_o A = \sup\limits_{A=\Sigma_I A_i} \Sigma_I \mu_{\alpha_i} A_i$ ($\{A_i\}$ décomposition finie de A) ;

ii) si la famille μ_α est (comme ici) filtrante $\uparrow$ (pour l'ordre naturel) $\mu_o A = \sup \mu_\alpha A$;

iii) si les μ_α sont σ-additives μ l'est.

Ici il est évident que $\mu_o \leq \mu$, donc que μ_o est σ-additive.

Ainsi il existe une densité $h_o = d\mu_o/d\mu \overset{\text{p.s.}}{\geq}$ chaque h_α (et ici ≤ 1), dite $\sup \operatorname{ess} h_\alpha$.

C. h_o égale p.s. h. En effet la preuve habituelle du théorème de Radon-Nicodym (faite avec toutes les mesures $\leq \mu_o$ et $\ll P$, valable aussi bien avec la famille filtrante μ_α) montre qu'on peut prendre $h_o = \lim \uparrow \sup (h_1,\ldots,h_n)$, définie par une <u>suite</u> de h_α, pour $\sup \operatorname{ess} h_\alpha$: si on choisit h_n de sorte que

$$\int_{\mathcal{T}} h_n \, d = \mu_n \mathcal{T} \uparrow \sup \mu_\alpha \mathcal{T} = \mu_o \mathcal{T} ,$$

et pose $h'_o = \lim \uparrow h_n$, on a

$$\mu'_o \leq \mu_o \quad \text{et} \quad (\mu_o - \mu'_o) \mathcal{T} = 0 \quad \text{donc} \quad \mu'_o = \mu_o .$$

Ainsi $h_o \overset{p.s.}{\leq} h$. Alors $\tilde{h}_o \leq \tilde{h}$ et $\tilde{h}_o \geq$ chaque h_α donc à h assurent $h \sim h_o$. On a bien h $\mathscr{C}$-mesurable, car $\mathscr{C}$ est μ complète. Mais $P(A\times B)$ est pour A fixé de Radon $(K \subset B$ et $\nu(B-K) < \varepsilon \Rightarrow P\{A\times(B-K)\} < \varepsilon)$. On a donc (par τ-régularité)

$$P(A\times\theta) = \sup \int_A h_\alpha(t)\mu(dt) = \sup \mu_\alpha A = \mu_o A = \int_A h(t)\ dt\ ,\ A\in\mathscr{C}\ ,$$

avec $\quad h(t) = \sup \nu_t\ f_\alpha = \nu_t\ \theta$.

Alors la formule $P'(E) = \int_T \nu_t(E_t)\ d\mu$ définit sur $\mathscr{C}\times\mathscr{B}$ une loi qui coïncide avec P sur les $A\times\theta$, donc lui est identique.∎

Corollaire 5 (de forme analogue au théorème de Jirina) : Soit $(\Omega,\mathscr{B},P)$ une probabilité de Radon, $\mathscr{C}$ une tribu dans Ω complète (pour elle-même p.r. à P)(6), P ayant une extension à une tribu $\mathscr{A}$ contenant $\mathscr{C}$ et $\mathscr{B}$.

Alors il existe une famille $P(\omega,.)$ de lois de Radon sur $\mathscr{B}$, désintégrant P par rapport à $\mathscr{C}$.

Preuve : Appliquer (Ω,P) dans $(\Omega,\mathscr{C})\times(\Omega,\mathscr{B})$. La désintégration dans cet espace produit (de l'image de P), du théorème précédent, fournit la solution :

$$P(A\cap B) = \int_A P(d\omega)\overline{\omega}(\omega,B)\quad ,\ \text{avec}\ \ \overline{\omega}(\omega,.) = \nu_\omega(.)\ \ \text{du théorème.}$$

---

(6) Il n'est pas nécessaire que $\mathscr{C}$ contienne tous les nuls de $\mathscr{B}_P$. Il suffit que $\mathscr{C}$ soit la complétée d'une sous-tribu, par exemple celle liée à une v.a. T .

La proposition suivante répond au souci, lorsque $\mathcal{C}$ est liée à la v.a. $T(\omega)$, que ν_ω (qui ne dépend que de t) soit portée par $\{\omega : T(\omega) = t\}$. Cette réponse est extrêmement restrictive.

<u>Proposition 5</u> : Soit $(\Omega,\mathcal{B},P)$ une probabilité de Radon ($\mathcal{B}$ tribu borélienne de Ω espace topologique), sous tendue par des compacts métrisables (à moins que $\mathcal{T}$ ci-après ne soit à base dénombrable). On suppose que $\omega \to t(\omega)$ est une application Lusin-mesurable dans l'espace topologique T, de tribu borélienne $\mathcal{C}$ (: mesurable suffit, si T est métrique séparable).

Alors les lois P_t désintégrant P par rapport à la v.a. $t(\omega)$ sont "bien portées" : portées par $\{\omega : t(\omega) = t\}$.

<u>Preuve</u> : Soit ψ l'application $\omega \to t(\omega) \times \omega$ dans $T \times \Omega$. Par définition de la Lusin-mesurabilité (conséquence de la mesurabilité lorsque T est métrique séparable), il existe des compacts K_ε dans Ω, sur lesquels $t(.)$ est continue ($PK_\varepsilon > 1-\varepsilon$). ψ est donc aussi continue sur ces K_ε et $\psi K = K'_\varepsilon$ est un compact de $\mathcal{T} \times \Omega$.

Soit $\Omega_o = \bigcup K_\varepsilon$ (une suite $\varepsilon_i \downarrow 0$), et $\mathcal{B}_o = \mathcal{B}(\Omega_o)$. Dès que $\mathcal{T}$ ou Ω_o (c.à.d. les K_ε) est à base dénombrable de voisinages, tout ouvert de $T \times \Omega_o$ est réunion dénombrable de pavés ouverts, donc $\mathcal{C} \times \mathcal{B}_o$ est la tribu borélienne de $\mathcal{T} \times \Omega_o$, donc contient les K'_ε.

Ainsi $E = \psi\,\Omega \supset E_o \in \mathcal{C} \times \mathcal{B}$ (E_o équivalent à E), et la désintégration de l'image P' de P dans $T \times \Omega$ (du corollaire ci-dessus - on complètera $\mathcal{C}$) donne

$$P'\,E_o^c = 0 = \int_{\mathcal{T}} P'\,E_{0t}^c\,P'(dt) \Rightarrow P'_t\,E_{0t} = 1 \quad \text{p.s.}$$

Ainsi pour chaque t , P'_t est portée par $\{\omega : T(\omega) = t\}$, et c'est, dans Ω , la famille P_t cherchée.

On notera qu'en général $E = \psi\Omega$ est seulement, pour P', de probabilité extérieure $P'^{*} E = 1$ pour P' , ce qui ne donne pas de renseignements sur les $P'^{*}_t E_t$.

7. Discussion de ces résultats.

Soit T sur $(\Omega,\mathcal{A},P)$ une v.a. à valeurs dans un espace quelconque $\mathcal{T}$, de tribu $\mathcal{E}$, et Y une v.a. à valeurs dans $\mathcal{Y}$ muni d'une tribu $\mathcal{B}$ à base dénombrable et séparant les points (de $\mathcal{Y}$) .

a) Si $\mathcal{Y}$ est polonais (et $\mathcal{B}$ sa tribu borélienne), une famille $(Y|t)$ de lois désintégrant la loi de Y par rapport à t existe toujours, dans $\mathcal{T} \times \mathcal{Y}$, mais pour obtenir, dans Ω , sur $\mathcal{A}_Y = Y^{-1}\mathcal{B}$, une famille $\mathcal{A}_T$-mesurable $P(.|T(\omega)) = \overline{\omega}(\omega,.)$, on doit

i) supposer P compacte sur $\mathcal{A}_Y$ (pour utiliser le théorème 2)

ou

ii) supposer que $Y\Omega \in \mathcal{B}_P$ (tribu $\mathcal{B}$ complétée pour P sur $\mathcal{B}$).

En effet dans ce dernier cas, on a $Y\Omega \supset E_o$, avec $E_o \in \mathcal{B}$, $P\,E_o = 1$. Ainsi

$$(10) \qquad P\,E_o = \int_{\mathcal{T}} P_t\,E_o\,P(dt) = 1 \Rightarrow P_t\,E_o = 1 \quad \text{p.s.}$$

La désintégration dans $\mathcal{T} \times \mathcal{Y}$ peut se transporter dans Ω , et cela donne les conclusions du th.2 , pour $\mathcal{A}_o = \mathcal{A}_Y$ ($\mathcal{E}$ devenant $\mathcal{A}_T$) .

On notera que suivant la proposition 1, les conditions i) et ii) sont équivalentes.

b) Cette équivalence, <u>et</u> ce transport de la désintégration, de $\mathcal{T} \times \mathcal{Y}$ dans Ω , vaut encore si, $\mathcal{Y}$ étant quelconque (mais sa tribu $\mathcal{B}$ séparante et à base dénombrable), on remplace ii) par ii') $Y\Omega \in \mathcal{B}_P$ <u>et</u> P est compacte sur $\mathcal{B}$.

En effet, suivant le lemme 1, i) implique ii') .

Que ii) entraine i) a la même preuve qu'en a) à ceci près que la classe $\mathcal{E}_E$ restriction de $\mathcal{E} \subset \mathcal{B}$ ($\mathcal{E}$ approchant $\mathcal{B}$ par rapport à P) à $E_o \in \mathcal{B}$, n'est plus formée de compacts (son image par Y^{-1} assure toujours i) , et (10) assure le transport des P_t en les $P_{T(\omega)}(.) = \bar{\omega}(\omega,.)$ sur $\mathcal{A}_Y$). En fait le cas le plus important est celui où $\mathcal{B}$ est la tribu borélienne de $\mathcal{Y}$ métrique séparable, alors, suivant le corollaire 3, $\mathcal{E}$ est en fait formée de compacts (peut être choisie telle). Il reste que, contrairement au cas a), il <u>faut</u> supposer P compacte (donc en fait de Radon) sur $\mathcal{B}$, pour désintégrer dans $\mathcal{T} \times \mathcal{Y}$.

Ainsi dans ces deux cas, a) ou b), il est équivalent de démontrer le théorème de Jirina dans $\mathcal{T} \times \mathcal{Y}$, ou dans Ω (avec les sous-tribus $\mathcal{A}_T$ et $\mathcal{A}_Y$), dès que $\mathcal{B}$ (nécessairement à base dénombrable) sépare les points de $\mathcal{Y}$.

Mais <u>l'existence de cette désintégration dans $\mathcal{T} \times \mathcal{Y}$ ne requière aucune condition</u> si $\mathcal{Y}$ est polonais, et seulement que P soit de Radon dans $\mathcal{Y}$, si $\mathcal{Y}$ est métrique séparable, alors que, même si T et Y sont réelles, il "faut" P compacte sur $\mathcal{A}_Y$. Cela tient à ce que pour tirer parti de la seule connaissance $P^*(Y\Omega) = 1$, après la désintégration de P dans $\mathcal{T} \times \mathcal{Y}$, il faudrait prendre en considération <u>tous</u> les $B \in \mathcal{B}$, ne coupant pas $Y\Omega$ et P-nuls, qui sont en quantité non dénombrable.

c) Si Y réelle est intégrable, on a ($A \in \mathcal{C}$, A représente l'événement $T \in A$)

$$\int_A P(dt) \int_{\mathcal{Y}} y\, P_t(dy) = \int_{\mathcal{T} \times \mathcal{Y}} Y\, 1_A\, dP \ , \quad \text{espérance de} \quad (Y|A).$$

Ainsi $\int_{\mathcal{Y}} y\, dP_{T(\omega)}(y) = E^{\mathcal{A}_T} Y$ est, dans Ω , "une densité de Radon Nicodym" de la mesure signée $\nu A = \int_{T^{-1}A} Y(\omega)\, dP$ " , une $\mathcal{A}_T$-régularisée" de la v.a. $Y(\omega)$. Il serait préférable de se limiter à une de ces deux expressions pour désigner $E^{\mathcal{A}_T} Y$, si on ne veut pas définir ses valeurs comme de vraies espérances, de vraies intégrales. Elles le sont pourtant, à condition qu'on ne veuille pas considérer $(\Omega, \mathcal{A}, P)$ comme un support unique de l'épreuve aléatoire considérée, mais seulement comme une désignation formelle commode de tous les supports équivalents : on ne perd aucune information concernant les v.a. étudiées en transportant P dans un espace produit où elles sont représentées (le plan R^2 s'il s'agit des seules v.a. T et Y réelles). Les propriétés d'une suite $E^{\mathcal{A}_T} Y_n$ se ramènent à celles des intégrales des Y_n pour les $P_t(dy)$, avec $y = \prod_1^{\infty} y_n \in R^{\mathbb{N}}$.

d) Dans $\mathcal{T} \times \mathcal{Y}$, le problème des P_t bien portées ne se pose pas. A supposer que les P_t existent dans Ω , la proposition 5 assure que les P_t sont, dans Ω , portées par $\{\omega : T(\omega) = t\}$. Elle n'a d'intérêt que par exemple, dans les conditions qui suivent.

Supposons (cf. le § 2) que Y soit un échantillon, fini ou dénombrable de v.a. à valeurs dans $\mathcal{Y}_0$ polonais. Alors Y prend ses valeurs dans un $\mathcal{Y}$ polonais lui aussi. Soit T une fonction (certaine) de Y . Si T est mesurable comme v.a. à valeurs dans $\mathcal{T}$ métrique séparable, alors la désintégration (dans $\mathcal{Y}$ ou dans $\mathcal{T} \times \mathcal{Y}$ par rapport à $T^{-1}\mathcal{E}$, ou $\mathcal{E}$, c'est-à-dire p.r. à) (existe toujours et) est bien portée.

On notera que le passage par la proposition 5 est inutile si on peut paramètrer les "surfaces" $\{y : T(y) = t\}$ de sorte que $\mathcal{Y}$ s'identifie à $\mathcal{T} \times \mathcal{Y}'$ (ce qui en pratique sera souvent réalisable).

Pourquoi, alors, ne pas définir d'emblée "T exhaustive pour le paramètre θ" indexant la famille P_θ de lois de l'échantillon Y par :

"Il existe un choix, en t , des lois $P_\theta(Y|t)$, indépendant de θ " (soit $P_t(dy)$) , ?

Des exemples élémentaires (comme la loi de Bernoulli, θ décrivant $]0,1[$) illustrent cette définition de façon éclatante, et celle-ci est naturelle puisque, en introduisant une loi a priori $G(d\theta)$, quelconque, pour le paramètre θ (qui devient valeur de la v.a. Θ) , on voit aisément, puisque

$$(11) \qquad G(d\theta)\, P_\theta(dy) = G(d\theta)\, P_\theta(dt)\, P_t(dy) \Longrightarrow P(dy) = P(dt)\, P_t(dy)$$

pour les lois marginales de T et Y , qu'on a

$$(12) \qquad P(d\theta|y) = G(d\theta)\, P_\theta(dt)/P(dt) = P(d\theta|t) \ .$$

Pour assurer la désintégration (12) on supposera Θ à valeurs dans un espace polonais. Puisque la loi a postériori $P(d\theta|y)$ porte toute l'information donnée par y (: l'échantillon), (12) exprime l'idée, évidente dès (11), que la connaissance de t suffit, que celle de y n'apporte plus d'information concernant θ .

e) Lorsqu'on utilise que le théorème 2, la désintégration est unique à une équivalence près ; deux systèmes $\bar{\omega}(\ ,.)$ et $\bar{\omega}'(\ ,.)$ désintégrant P sur $\mathcal{A}_0$ p.r. à la tribu $\mathcal{E}$, sont p.s. identiques sur l'algèbre dénombrable $\mathcal{Q}_0$, donc également sur $\mathcal{A}_0$.

8. <u>Le théorème de Tulcea contient le théorème de Kolmogorov-Marcewsky</u> (cf.[2] et [5]) .

Soit $\mathcal{X} = \prod_T \mathcal{X}_t$ un produit quelconque, indexé par $t \in T$, d'espaces $\mathcal{X}_t$ munis de tribus $\mathcal{B}_t$ et $\mathcal{A}$ l'algèbre $\bigcup_I \mathcal{B}^I$ où $\mathcal{B}^I = \underset{t\in I}{} \mathcal{B}_t$ est la tribu produit, limitée à I partie finie de T .

$\mathcal{A}$ engendre la tribu produit des $\mathcal{B}_t$ (t décrivant T) .

Supposons que la fonction P sur $\mathcal{A}$ soit σ-additive sur chaque $\mathcal{B}^I$, et que les lois marginales P_t soient parfaites. Prouvons, sans utiliser la méthode directe, que μ est σ-additive sur $\mathcal{A}$ (ce qui constitue le théorème de Kolmogorov-Marcewsky).

Soit $D = \{t_1,\dots,t_n,\dots\}$ une partie dénombrable de T . D'après le théorème de Jirina, les lois $P_{t_1\dots t_n}(dx)$ désintégrant $P_{t_{n+1}}$ par rapport à $x_{t_1}\ \dots\ x_{t_n}$ existent sur toute $\mathcal{B}'_{t_{n+1}}$ à base dénombrable. Le théorème de Tulcea (cf.[2] p.614) assure la σ-additivité de P sur $\bigcup_{n=1}^{\infty} \mathcal{B}'_{t_1\dots t_n})$ pour tout choix de $\mathcal{B}'_{t_n}$ dans $\mathcal{B}_{t_n}$, donc aussi lorsque ces choix, et celui de D , varient, c'est-à-dire sur $\mathcal{A}$ elle-même.

Autre application du théorème de Tulcea :

Revenons à la remarque 2. Si la v.a. Y égale $\prod_1^\infty Y_i$ (Y_i réelles), et si on choisit successivement les systèmes de lois $P_t(dy_1), P_{ty_1}(dy_2), \ldots, P_{ty_1\ldots y_I}(dy_{I+1})$, le théorème de Tulcea assure que pour chaque t fixé, l'ensemble des $P_{ty_1\ldots y_I}(dy_{I+1})$ définit une loi P_t dans $\mathcal{Y}$, donc que $P(dt)\, P_t(dy)$ coïncide avec la probabilité donnée sur $\mathcal{E} \times \mathcal{B}$ puisqu'elle coïncide sur $\mathcal{E} \times \mathcal{A}$.

Si Y est portée par $E = \bigcap_1^\infty \mathcal{O}_n$, les $P_t(dy)$ le sont également puisque E mesurable entraine $P_t\, E = 1$ p.s.

9. Les statistiques exhaustives.

Lemme 2 : Soit Ω un espace polonais de tribu borélienne $\mathcal{B}$. Il existe dans $\mathcal{B}$ une algèbre dénombrable $\mathcal{A}$ (engendrant $\mathcal{B}$) et dans $\mathcal{A}$ une infinité de suites $C_{nk} \underset{k\uparrow\infty}{\uparrow} C_n$ (avec $C_n \in \mathcal{A}$, telles que les relations

$$P\, C_{nk} \underset{k\uparrow\infty}{\uparrow} P\, C_n \quad , \quad n = 1,2,\ldots \quad ,$$

soient (nécessaires et) suffisantes pour assurer la σ-additivité sur $\mathcal{A}$ de n'importe quelle fonction P (≥ 0 , bornée, par exemple $P \Omega = 1$) additive sur $\mathcal{A}$.

Remarque 6 : Ce lemme très important est dû à Harris (cf. (4) p.111) et a été prouvé par lui sous la seule hypothèse de (monotonie et) sous-additivité (alors la conclusion est $A_n \uparrow A \Rightarrow P\, A_n \uparrow PA$) , et étendu à n'importe quelle mesure (: fonction additive de $\mathcal{A}$ dans $[0,\infty[$) par Acquaviva. Nous en donnons une preuve simplifiée dont le principe parait n'être pas valable pour une mesure non bornée.

Preuve a) : Nous identifions Ω avec une $\lim \downarrow \mathcal{O}_n$, $\mathcal{O}_n$ ouvert de $\mathcal{Y} = [0,1]^{\mathbb{N}}$, et considérons dans $\mathcal{Y}$ l'algèbre $\tilde{\mathcal{A}}$ cylindrique engendrée par les pavés (k,ℓ entiers, y_i coordonnées de y)

$$(13)\ \prod_1^I \Delta_i \ , \text{ où } \Delta_i = \{y_i : y_i < \frac{k_i}{2^{\ell_i}}\} \quad \text{ou} \quad \{y_i \leq \frac{k_i}{2^{\ell_i}}\} \ .$$

Dans Ω nous prenons pour $\mathcal{A}$ l'algèbre trace $\tilde{\mathcal{A}} \cap \Omega$. Les cyclindres (13) à Δ_i tous ouverts étant numérotés de 1 à ∞ (: on ordonne l'ensemble des $\{I,k_i,\ell_i \ , \ i = 1,\dots I\}$) par n , sont désignés par $\tilde{C}_n$, les $\tilde{C}_n\Omega$ par C_n et C_{nk} ($k = 1,2,\dots$) est pour chaque n une suite tirée de $\{C_n\}$ et croissant vers C_n . $\mathcal{O}_n$ étant ouvert est une réunion (dénombrable) de cyclindres ouverts, chacun d'eux est une limite $\uparrow$ de cyclindres compacts (choisir la base compacte, et multiplier par $\prod_{I+1}^{\infty} [0,1]_i$ si la base est dans $\prod_1^I [0,1]_i$) de $\tilde{\mathcal{A}}$. On en déduit des K_{nk} de $\tilde{\mathcal{A}}$ qui $\uparrow \mathcal{O}_n$ lorsque $k \uparrow \infty$. On ajoute les suites $K_{nk}\Omega$ aux précédentes ; obtenant la classe dénombrable $\mathcal{E}$.

b) Donnons nous m additive (≥ 0, $m\,\Omega = 1$) sur $\mathcal{A}$, et posons

$$\tilde{m}\,\tilde{A} = m(\tilde{A}\,\Omega) \ , \qquad \tilde{A} \in \tilde{\mathcal{A}} \ .$$

La σ-additivité supposée de m sur les suites $C_{nk} \uparrow C_n$, assure la définition par $\tilde{m}$ d'une P probabilité sur $\mathcal{Y}$ (par le système cohérent de ses fonctions de répartition). Si nous prouvons que $P\,\Omega = 1$, cela démontrera que la restriction de P à Ω coïncide avec m sur $\mathcal{A}$: m est σ-additive sur $\mathcal{A}$.

c) Prouvons que $m(K_{nk}\ \Omega) \leq P(K_{nk})$, en nous plaçant dans le produit $[0,1]^I$ où un tel $K = K_{nk}$ a sa base. On a $K = \lim \downarrow \tilde{A}_\ell$, où $\tilde{A}_\ell$ est une somme de pavés "semi-ouverts à droite", donc appartenant à l'anneau $\tilde{\mathcal{A}}_o$ sur lequel P égale $\tilde{m}$. D'où $PK = \lim \downarrow P\,\tilde{A}_\ell = \lim \downarrow m(\tilde{A}_\ell\ \Omega) \geq mK\Omega$. Considérons un $\mathcal{O}_n = \mathcal{O}$ fixé. Il existe un $\tilde{A} = \bigcap_1^\infty \tilde{A}_k$, avec $\tilde{A}_k \downarrow$ et compact, tel que

$$P\,\tilde{A} > P\,\mathcal{O}^c - \varepsilon \ ,$$

et pour chaque $K_i = K_{ni}$, $\tilde{A}\,K_i = \emptyset \Rightarrow A_k\,K_i$ vide pour tout $k \geq K(n,i)$, d'où

$$m(K_i\ \Omega) \uparrow 1 \Rightarrow P\,K_i^c \downarrow 0 \Rightarrow P\,\tilde{A}_k \downarrow 0 \Rightarrow P\,\mathcal{O}_n^c = 0 \Rightarrow P\,\Omega = 1 \ .$$

Remarque 4 : Il faut noter que l'utilisation de cette représentation de Ω polonais est un peu délicate, et, peut être, artificielle car Ω n'est pas complet dans $\mathcal{Y}$ n'y étant pas fermé (s'il n'est pas compact) : une structure uniforme sur Ω (pratiquement celle définie par sa distance) ne peut être la trace de celle de $\mathcal{Y}$, sur les bords de Ω elle est nécessairement plus fine que celle définie par $\mathcal{Y}$.

Théorème 4 : Soit P_θ une famille de probabilités dans l'espace polonais $\mathcal{Y}$ (ce peut être l'espace de représentation d'un échantillon dénombrable $Y = \{Y_1, Y_2, \ldots, Y_n, \ldots\}$ pour une v.a. à valeurs dans un espace polonais $\mathcal{Y}_o$: $\mathcal{Y} = \mathcal{Y}_o^{\mathbb{N}}$) .

Soit T une fonction (certaine) de Y, qu'on peut toujours supposer mesurable en mettant, dans l'espace $\mathcal{T} = T\mathcal{Y}$ la tribu image $\mathcal{E} = T\mathcal{B}$ (En pratique on se placera dans les conditions de la proposition 5 assurant des $p(A,t)$ bien portées).

Pour que T soit exhaustive (pour le paramètre θ, cf. le § 7-d)), il suffit que pour chaque C de la famille dénombrable $\mathcal{C}$ définie au lemme 2 il existe (dans l'espace $\mathcal{T} \times \mathcal{Y}$ ou dans $\mathcal{Y}$ avec la sous-tribu $\mathcal{B}_T = T^{-1}\mathcal{E}$), des "densités" $p(A,t)$, espérances conditionnelles $P_\theta^{\mathcal{E}} A$, indépendantes de θ :

$$(14) \qquad B \in \mathcal{B}_T \Rightarrow P_\theta(A\,B) = \int_B p(A,t)\, P_\theta(dt)\ .$$

<u>Preuve</u> : Elle est la même que celle du théorème de Jirina : les couples (A,A') (A et A' disjoints) de l'algèbre $\mathcal{A}$ du lemme 2 sont dénombrables et

$$P_\theta\{B(A+A') = P_\theta BA + P_\theta BA' \Rightarrow P_\theta\{t : p(A+A',t) \neq p(A,t)+p(A',t)\} = 0$$

pour tout θ.

De même, pour chaque suite $C_{nk} \uparrow C_n$ de $\mathcal{C}$ (incluant les $K_{nk}\Omega$ du lemme 2), la relation

$$P_\theta\, C_{nk} = \int p(C_{nk},t)\, P_\theta(dt) \uparrow \int p(C_n,t)\, P_\theta(dt) = P_\theta\, C_n\ ,$$

donne

$$P_\theta\{t : p(C_{nk},t) \underset{k\to\infty}{\not\to} p(C_n,t)\} = 0 \text{ pour tout } \theta\ .$$

Ainsi, en réunissant tous ces ensembles P_θ-nuls (<u>pour tout</u> θ) on obtient N également nul, tel que pour $t \notin N$, $p(A,t)$ soit σ-additive sur $\mathcal{A}$ (suivant le lemme 2) et définisse (pour $t \in N$ on posera $P(.,t)$ égale à une loi constante, quelconque) une désintégration des P_θ dans $\mathcal{Y}$, par rapport à la v.a. T, qui ne dépend pas de θ :

$$\int_{B_1 \in \mathcal{B}_T} p(B,t)\, P_\theta(dt) = P'(B\, B_1) \ ,$$

avec $\quad P'(B\, B_1) = P(B\, B_1)$ pour tous $B \in \mathcal{A}$, $B_1 \in \mathcal{B}_T$,

donc tous B , B_1 .

Remarque 6 : Deux familles $p(A,t)$ et $p'(A,t)$, désintégrations indépendantes de θ sont égales sauf sur un $N \in \mathcal{C}$, nul pour toute P_θ , puisqu'il en est ainsi sur $\mathcal{A}$ pour chaque A , donc sur $\mathcal{B}$.

Remarque 7 : Il ne semble pas qu'on puisse rattacher le lemme 2 à l'existence d'une classe compacte fixe (ne dépendant pas de la probabilité) et dénombrable. Dans $R^{\mathbb{N}}$ une telle classe existe : celle des cylindres à base compacte (somme finie de pavés fermés dyadiques), dans $[0,1]^{\mathbb{N}}$ ce sont des compacts, mais cette propriété ne passe pas à la restriction à $\Omega = \bigcap \mathcal{O}_n$.

Pour terminer, recommandons la preuve incomparable, par Hansel, du théorème de relèvement, et remercions M. Valadier de nous avoir adressé son travail [9] , et M. P.A. Meyer d'accueillir ce texte dans un environnement beaucoup plus savant.

BIBLIOGRAPHIE

(1) CHATTERJI S.D. : Les martingales et leurs applications analytiques.
Ecole d'été de St Flour (1971) . Lecture notes n°307.

(2) DOOB J.L. : Stochastic processes.
Ed. Wiley (1953).

(3) HANSEL G. : Le théorème de relèvement.
Ann. Inst. H. Poincaré VIII-4 (1972).

(4) HARRIS T.E. : Counting measures, Monotone random Set Functions.
Zeitschrift W. 10 (1968) , 102-119.

(5) HENNEQUIN et TORTRAT : Théorie des probabilités et quelques applications.
Ed. Masson 1965.

(6) HOFFMANN - JORGENSEN J. : Existence of conditional probabilities.
Math. Scand. 28 (1971) 257-264.

(7) JIRINA M. : Probabilités conditionnelles sur des algèbres à base dénombrable.
Czechosloviak Math. J., 4 (1954) pp. 372-380.

(8) MOKOBODSKY : Relèvement borélien compatible avec une classe d'ensembles négligeables.
Application à la désintégration de mesures.
Séminaire de Probabilités IX (1973-4) p. 442
Strasbourg, Lecture notes n° 465.

(9) VALADIER : Comparaison de trois théorèmes de désintégration.
Séminaire d'analyse convexe (n°10) Montpellier 1972.