SÉMINAIRE DE PROBABILITÉS (STRASBOURG)

NASSIF GHOUSSOUB

An integral representation of randomized probabilities and its applications

Séminaire de probabilités (Strasbourg), tome 16 (1982), p. 519-543
<http://www.numdam.org/item?id=SPS_1982__16__519_0>

An integral representation of randomized probabilities and its applications.

By

Nassif Ghoussoub.

0. Introduction: In this paper, we study the topological and extremal properties of the set of random probabilities on a compact space. Our main goal is to give a Choquet-type integral representation for the measure theoretic notions to which one can associate a random probability.

The representation applied, for instance, to randomized stopping times shows that they are averages of true stopping times. The behaviour of some stochastic processes on randomized stopping times can then be easily understood from the behaviour of these processes on the genuine ones. As an immediate application, we give a proof of the Baxter-Chacon compactness argument and of an optimal stopping problem.

Applied to positive operators on L_1 and $C(K)$, the representation implies that such operators are averages of point transformations: a useful fact for extending some properties which are easily verifiable in the case of operators induced by point transformations to more general operators. For an example we give a proof of the Ricsz-Thorin convexity theorem.

The above representation also implies that operators on L_1 of a compact group are actually randomized multipliers and they become convolution operators, that is averages of translations, only if they

---

Department of Mathematics - University of British Columbia - Vancouver, B.C., Canada.

commute with these translations. We also discuss the possibility of associating to any transition probability on a compact space K, a Markov chain induced by a random walk on a group operating "measurably" on K.

Finally, we study the connection between the various types of convergence of a sequence of operators on L_1 and the convergence of the measures supported by the space of point transformations, which represent these operators. We show that while vague convergence of the representing measures already implies mean convergence of the operators, almost sure convergence is implied by a stronger topology naturally imposed on these representing measures.

It is my pleasure to thank M. Emery, J. Fournier and E. Perkins for the very stimulating and helpful discussions during the preparation of this paper.

I. Integral representation of random probabilities:

Let K be a compact separable space and let $(\Omega,\mathcal{F},P)$ be a probability space. Denote by

$\mathcal{C}$ the Banach space $L_1(\Omega; C(K))$ of all $C(K)$-valued Bochner integrable random variables. Recall [13] that the dual of $\mathcal{C}$ is $L_{\infty *}(\Omega,\mathcal{M}(K))$ of all random measures $\mu:(\Omega,\mathcal{F},P)\to\mathcal{M}(K)$ measurable for the weak-star Borel subsets of $\mathcal{M}(K)$ and such that $\text{ess sup}|\mu_w|(K)$ is finite.

Consider the set $D = \{\mu \in \mathcal{C}^* ; \mu \geq 0, \|\mu\| \leq 1 \text{ and } \mu(1) = 1\}$ where 1 is the unit process in $\mathcal{C}$. Clearly D is a weak-star compact convex subset of $\mathcal{C}^*$. In the sequel we want to identify the extreme points of D in order to apply Choquet's integral representation. For that, consider the convex set D_1 of all probability measures on $K\times\Omega$ (equipped with the product σ-field) such that their projection on Ω is P. Let $\mathcal{P}(K)$ be the set of probability measures on K.

Lemma I.1: There exists an affine bijection between D_1 and D.

Proof: Let ν be in D_1. Since K is compact, every probability measure on K is tight, hence by [8], there exists a strict disintegration of ν with respect to the projection $pr: K \times \Omega \to \Omega$ and P. That is an application $\bar{\nu}: w \to \nu_w$ from Ω into $\mathcal{P}(K)$ such that for every Borel subset B of K, $w \to \nu_w(B)$ is measurable. It is then clear that $\bar{\nu} \in D$.

On the other hand, if $\bar{\nu} \in D$, define the measure ν on $K \times \Omega$ by $\nu(A \times B) = \int_B \nu_w(A) dP(w)$ whenever A is Borel in K and $B \in \mathcal{F}$.

It is clear that ν extends to the product σ-field on $K \times \Omega$ and that the projection of ν on Ω is P. The representation is unique, since the processes of the form $f(t) \cdot g(w)$ where $f \in C(K)$ and $g \in L_1(\Omega)$ belong to $\mathcal{C}$ and they generate the whole product σ-field by the monotone class theorem. The uniqueness implies that $\nu \to \bar{\nu}$ is affine.

For any measurable function $\sigma: \Omega \to K$, denote by δ_σ the random measure associated to the measure on the product $K \times \Omega$ defined by $X \to E[X_\sigma]$ for any $X \in \mathcal{C}$.

Proposition I.2: a) D is $\sigma(\mathcal{C}^*, \mathcal{C})$ compact and convex and is separable whenever the σ-field $\mathcal{F}$ is.

b) The extreme points of D are the random measures δ_σ where σ is a measurable function from Ω into K.

Proof: a) If $\mathcal{F}$ is a separable σ-field, then $L_1(\Omega, C(K))$ is a separable Banach space since $C(K)$ is. It follows that the dual ball is a metrizable weak-star compact convex set.

b) Follows from the first lemma and the well known fact that the extreme points of the set of probability measures on $\Omega \times K$ whose projection on Ω is P, are the measures supported by the measurable graphs from Ω into K. (See for instance [15]).

Denote now by G the space of all measurable functions from Ω to K and let d be the metric on K. We say that (σ_n) in G converges in probability to σ in G, if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{w \in \Omega \,;\, d(\sigma_n(w),\sigma(w)) \geq \varepsilon\} = 0$$

Lemma I.3: If σ_n, σ are in G, then $\sigma_n \to \sigma$ in probability if and only if δ_{σ_n} converges to δ_σ in $\sigma(C^*,C)$.

Proof: Suppose $\sigma_n \to \sigma$ in probability. For any $f \in C(K)$, $\big(f(\sigma_n)\big)_n$ converges then in probability to $f(\sigma)$ and since $(f(\sigma_n))_n$ is bounded in $L_\infty(\Omega,F,P)$, it converges in $\sigma(L_\infty, L_1)$, hence for every $g \in L_1(\Omega,F,P)$ we get

$$\int g(w)f(t)\delta_{\sigma_n} \,.\, dP \to \int g(w) \,.\, f(t) \,.\, \delta_\sigma \,.\, dP \,.$$

Since the linear span of $\{g(w) \,.\, f(t) \,;\, g \in L_1(\Omega),\ f \in C(K)\}$ is dense in C, it follows that $\delta_{\sigma_n} \to \delta_\sigma$ in $\sigma(C^*,C)$.

Suppose now $\delta_{\sigma_n} \to \delta_\sigma$ in $\sigma(C^*,C)$. Take $0 < \varepsilon < \frac{1}{2}$. There exists a partition $\{B_1,B_2,\dots,B_m,1\}$ of Ω and elements $t_1,t_2,\dots,t_m$ in K such that for every w in B_i, $d(\sigma(w),t_i) < \varepsilon^2/2$.

Let f_i be the function in $C(K)$ defined by $f_i(t) = d(t,t_i) \cdot \frac{\varepsilon^2}{2}$ and let $g_i = \frac{\chi_{B_i}}{P(B_i)}$.

Since $\delta_{\sigma_n} \to \delta_\sigma$ in $\sigma(C^*,C)$ we get that for every $g \in L_1(\Omega)$ and $f \in C(K)$, there exists N so that if $n \geq N$ we have

$$\left| \int_\Omega\int_K g(w)f(t)(\delta_{\sigma_n} - \delta_\sigma)dP \right| \leq \varepsilon^2/2 \,.$$

Applying this to the finite family $(f_i,g_i)_{i=1}^m$, we get N so that if $n \geq N$

$$\frac{1}{P(B_i)} \left| \int_{B_i} d(\sigma_n,t_i) - d(\sigma,t_i))dP \right| \leq \varepsilon^2/2 .$$

for any $1 \leq i \leq m$.

Since $\frac{1}{P(B_i)} \int_{B_i} d(\sigma,t_i)dP \leq \varepsilon^2/2$,

It follows that for any $1 \leq i \leq m$ and $n \geq N$

$$\int_{B_i} d(\sigma_n,t_i)dP \leq \varepsilon^2 P(B_i)$$

That is

$$P((d(\sigma_n,t_i) > \varepsilon) \cap B_i) \leq \varepsilon P(B_i)$$

Hence

$$P((d(\sigma_n,\sigma) \geq \varepsilon) \cap B_i) \leq \varepsilon P(B_i)$$

and $$P(d(\sigma_n,\sigma) \geq \varepsilon) \leq \varepsilon .$$

Identify now the elements in G which are equal almost everywhere. The metric $\bar{d}$ defined on the equivalence classes of elements in G by

$$\bar{d}(\sigma,\tau) = \int_\Omega \frac{d(\sigma,\tau)}{1+d(\sigma,\tau)} \cdot dP$$

defines clearly a topology on G which coincides with the topology of convergence in probability and that $(G,\bar{d})$ is a separable complete metric space.

Denote by $A(D)$ the space of all affine continuous functions on D . For each element $X = (X_t(w))$ in C and each random measure $(\mu_w)_w$ in C^* , the duality map is then defined by

$$\langle X,\mu\rangle = \int_\Omega \left(\int_K X_t(w)d\mu_w(t)\right)dP = E\int_K X_t(w)d\mu_w(t)$$

where E is the expectation with respect to P. If $\sigma \in G$, we shall denote by $E[X_\sigma]$ the expression $\langle X,\delta_\sigma\rangle$.

Theorem I.4: To any random probability μ in D, we can associate a probability Radon measure $\tilde{\mu}$ on $(G,\bar{d})$ such that

1) For any X in C, $\langle X,\mu\rangle = \int_G E[X_\sigma]d\tilde{\mu}(\sigma)$

2) $A(D)$ is dense in $L^1(G,\tilde{\mu})$.

Proof: By the Choquet representation theorem [4] applied to the convex compact set D and μ, there exists a maximal and simplicial Radon measure $\tilde{\mu}$ on the extreme points of D (that is $(G,\bar{d})$) such that $\overline{A(D)} = L_1(\tilde{\mu})$ and for any $h \in A(D)$ we have

$$h(\mu) = \int_G h(\sigma)d\tilde{\mu}(\sigma)$$

Now, it is just enough to notice that any $X \in C$, defines an element in $A(D)$ by the map $\nu \to \langle X,\nu\rangle$.

We shall see later that this representation is not unique, that is D is not a simplex.

II. Increasing processes and related notions:

a) Increasing processes: Suppose now K to be the interval $[0,\infty]$. An increasing process is a map $A:[0,\infty] \to L_1(\Omega,F,P)$ satisfying the following properties

(i) (A_t) is right continuous from $[0,\infty]$ into L_1.

(ii) $0 \le A_t \le A_s$ for all $t \le s$

Let D_2 be the set of all increasing processes (A_t) such that $A_\infty = 1$. It is then easy to show that to any random probability

(μ_w) in D, one can associate a unique increasing process in D_2 such that

$$A_t(w) = \mu_w([0,t]) \quad \text{a.s.}$$

The extreme points of D_2 are then the increasing processes of the form $A_t(w) = I_{[0,t]}(\sigma(w))$, where σ is a measurable map from Ω to $[0,\infty]$.

Let now B_1 = the space of measurable processes $(X_t(w))$ so that $\sup_t |X_t| \in L^1$. Then every element $X \in B_1$ defines a bounded affine function on D_2 via the map

$$\overline{X} : \ (A_t) \to E \int_0^\infty X_t dA_t$$

where a cadlag version of A has been chosen. The above representation says then that for any (A_t) in D_2, there exists a Radon probability measure $\tilde{\mu}$ on the space G_2 of all measurable maps from Ω into $[0,\infty]$ so that for any $X \in C$,

$$E\int_0^\infty X_t dA_t = \int_{G_2} E[X_\sigma] d\tilde{\mu}(\sigma)$$

The above equation holds also for any $X \in B^1$ since they verify the barycentric formula.

In this case, $\tilde{\mu}$ can be chosen in a natural way, since if we take the increasing process (B_t) which is the inverse of (A_t), that is $B_t = \inf\{s \ ; \ A_s > t\}$ then

$$E\int_0^\infty X_t dA_t = \int_0^1 E[X_{B_t}] dt$$

and $\tilde{\mu}$ can be chosen to be the image of the Lebesgue measure on $[0,1]$ by the map $B : [0,1] \to G_2$. (It is easy to see that B is measurable when G_2 is equipped with the Borel σ-field generated by the topology of convergence in probability).

We noted that C embeds in $A(D)$ in a natural way and in general $C + \mathbb{R}$ is dense in $A(D)$. However, in case $K = [0,\infty]$, one can associate to any affine and continuous function h on D_2, a process Y in $L^1(\Omega:C([0,\infty]))$ such that

$$n(\delta_\sigma) = E[Y_\sigma]$$

To sketch a proof of this fact, define for each t in $[0,\infty]$ and $H \in F$, the function $\sigma_{t,H}$ from Ω into $[0,\infty]$ equal to t on H and 0 elsewhere.

For each t, define

$$Q_t(H) = h(\delta_o) - h(\delta_{t,H})$$

Q_t is additive since, if $H \cap H' = \phi$ then

$$Q_t(H \cup H') = h(\delta_o) - h\left(\delta_{\sigma_{t,H\cup H'}}\right) = h(\delta_o) - h\left(\delta_{\sigma_{t,H}} \vee \delta_{\sigma_{t,H'}}\right)$$

$$= h(\delta_o) - h\left(\delta_{\sigma_{t,H}} - \delta_{\sigma_{t,H'}}\right) + h\left(\delta_{\sigma_{t,H}} \wedge \delta_{\sigma_{t,H'}}\right)$$

$$= h(\delta_o) - h\left(\delta_{\sigma_{t,H}}\right) + h(\delta_o) - h\left(\delta_{\sigma_{t,H'}}\right)$$

$$= Q_t(H) + Q_t(H') .$$

If $H_n \downarrow \phi$, then σ_{t,H_n} converges to 0 in probability and $Q_t(H_n)$ converges to zero by the continuity of h. It is also clear that Q_t is absolutely continuous with respect to P. Let $X_t = \frac{dQ_t}{dP}$.

Since h is affine and continuous, it is Lipschitz with Lipschitz constant equal to K say. For all t and s we have

$$\int |X_t - X_s| dP = \mathrm{Var}(Q_t - Q_s)(\Omega) =$$

$$\sup \left\{ \sum_i |Q_t(H_i) - Q_s(H_i)| \; ; \; (H_i) \text{ partition of } \Omega \right\} \leq$$

$$\sup \left\{ \sum_i K\, d\,(\sigma_{t,H_i},\ \sigma_{s,H_i})\ ;\ \text{partition of}\ \ \Omega \right\} \le$$

$$\sup \left\{ \sum_i K|t-s|P(H_i)\ ;\ (H_i)\ \text{partition of}\ \ \Omega \right\} = K|t-s|$$

It is standard to show that there exists a modification of X which is separable and measurable. The same proof as above shows that for any simple functions σ and τ in G we have

$$\int |X_\tau - X_\sigma|\, dP \le K\, d\,(\tau,\sigma)$$

It follows that (X_t) has a modification which is continuous. Note now that for any sequence of simple σ_n's, we have $\int |X_o - X_{\sigma_n}|\, dP \le K\, d\,(0,\sigma_n) \le K$; That is $\sup_n \int |X_{\sigma_n}|\, dP < \infty$. By Fatou's lemma, we get that $\sup_{\sigma \in G} \int |X_\sigma| < \infty$, let now

$\tau = \min\{t\ ;\ X_t(w) = \max_{t \in [0,\infty]} X_t(w)\}$, we get that

$$\int \sup_t |X_t|\, dP \le \int |X_\tau|\, dP < \infty \quad \text{and} \quad X \in \mathcal{C}.$$

The process $Y_t = h(\delta_o) - X_t$ will do the job.

Note that in general (unless $\Omega = \{w\}$ or $(K) = \mathbb{R}$) D(hence D_1 and D_2) is not a Choquet simplex, that is the maximal representing measure is not unique. For an example it is enough to take $\Omega = \{0,1\}$ and P the probability assigning 1/2 to each of the sets $\{0\}$ and $\{1\}$.

Let $A_t(0) = A_t(1) = \begin{cases} 1/2 & \text{if} \quad 0 \le t < 1 \\ 1 & \text{if} \quad t = 1 \end{cases}$

Let σ_i , $i = 1,2,3,4$ be maps from $\{0,1\}$ to $[0,1]$ defined by

$$\sigma_1 = \begin{cases} 1 & \text{if} \quad w = 0 \\ 0 & \text{if} \quad w = 1 \end{cases} \qquad\qquad \sigma_2 = \begin{cases} 0 & \text{if} \quad w = 0 \\ 1 & \text{if} \quad w = 1 \end{cases}$$

$$\sigma_3 = \begin{cases} 0 & \text{if } w=0 \\ 0 & \text{if } w=1 \end{cases} \qquad \sigma_4 = \begin{cases} 1 & \text{if } w=0 \\ 1 & \text{if } w=1 \end{cases}$$

It is immediate that

$$A_t = \frac{1}{2} I_{[0,t]}(\sigma_1) + \frac{1}{2} I_{[0,t]}(\sigma_2) = \frac{1}{2} I_{[0,t]}(\sigma_3) + \frac{1}{2} I_{[0,t]}(\sigma_4) \ .$$

However, we have the following

Proposition II.1 : The map $A \to B \circ \lambda$ is a simplicial selection from D_2 onto $P(G_2)$ where B is the inverse of A and $B \circ \lambda$ is the probability on G_2, image by B of the Lebesgue measure λ on $[0,1]$.

Sketch of proof: To prove that $A(D_2)$ is dense in $L_1(G_2, B \circ \lambda)$ it is enough to show that the space

$$X = \left\{ f \in L_1[0,1] \,\middle|\, \text{there exists } X_t(w) \text{ lower semi-continuous on } D_2 \text{ and that } f(t) = E[X_I \right.$$

is dense in $L_1[0,1]$. But this follows from the fact that X contains the intervals (a,b) since $X_{(a,b)}(s) = E[X_{]]B_a, B_b[[} \circ B_s]$.

Let now (A_t^n) be a sequence of increasing processes in D_2 and let $(\tilde{\mu}_n)$ be a sequence of representing probabilities on G_2. It is clear that there exists a subsequence $(\tilde{\mu}_{n_k})$ which is vaguely convergent to say $\tilde{\mu}$ on $C(D)$, hence (A_t^n) converges to (A_t^∞) (the barycenter of $\tilde{\mu}$) on every continuous function on D which verifies the barycentric formula. That is essentially the Baxter-Chacon compactness argument in the case we are dealing with a constant filtration.

Let Λ^1 be the space of optional processes of class (D). Every element X in Λ^1 defines then a bounded affine function on D_3 via the map $\overline{X} : A_t \to E\int_0^\infty X_t dA_t$. (See [10]).

Again, by the representation theorem we get a Radon probability measure $\tilde{\mu}$ on the space G_3 of all F_t-stopping times so that for any $X \in \Lambda_1$ we have

$$E\int_0^\infty X_t dA_t = \int_{G_3} E[X_\sigma] d\tilde{\mu}(\sigma) \ .$$

The following proposition reduces the topology of Baxter-Chacon [1] to the vague topology on $M(D_3)$.

<u>Proposition II.2</u> : (Baxter-Chacon-Meyer) Every sequence (A_t^n) of randomized stopping times has a subsequence $(A_t^{n_k})$ such that for any regular process (X_t) of class (D) we have

$$E\int_0^\infty X_t dA_t^{n_k} \to E\int_0^\infty X_t dA_t$$

where (A_t) is also a randomized stopping time.

For the proof it is enough to take $\tilde{\mu} \in M(D_3)$ to be a cluster point in the vague topology of a sequence $(\tilde{\mu}_n)_n$ in $M(G_3)$ representing $(A_t^n)_n$. Then A_t = barycenter of $\tilde{\mu}$ is a limit of $(A_t^n)_n$ on the regular optional processes of class (D) since they induce bounded affine maps on D_3 , verify the barycentric formula and are continuous on G_3 .

Again, one can show that the map $A_t \to B_t \circ \lambda$ is a simplicial selection from D_3 onto $P(G_3)$ where B_t is the time change associated to A_t .

Note that the vague convergence of $(\tilde{\mu}_n)$ is stronger than the convergence of Baxter-Chacon, since the elements of $C(D_3)$ are not necessarily induced by processes and we do not know if the space $A(D_3)$ is strictly larger than the space of optional regular processes of class (D) .

Another immediate application of the representation above is the following optimal stopping rule.

<u>Proposition II.3</u> For any regular process of class (D) (X_t) there exists a stopping time σ_0 such that $E[X_{\sigma_0}] = \sup_{\sigma \in G_3} E[X_\sigma]$

<u>Proof</u>: It is enough to notice that $\overline{X} : A_t \to E\int_0^\infty X_t dA_t$ is affine and continuous on the convex compact D_3 , hence it attains its maximum on an extreme point.

b) Vector measures:

Suppose now $K = [0,1]$. Recall that an $L_1(\Omega,\mathcal{F},P)$-valued vector measure F is a set function F from $\mathcal{B}$ (the Borel functions of $[0,1]$) into $L_1(\Omega,\mathcal{F},P)$ so that

(i) $F(\phi) = 0$ a.e.

(ii) $F(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} F(B_i)$ for any disjoint $B_1, B_2, \ldots$ in $\mathcal{B}$.

Let D_4 be the set of all $L_1(\Omega,\mathcal{F},P)$-valued, positive vector measures such that $F[0,1] = 1$ a.e. One can show (see [2]) the existence of a unique increasing process (A_t) in D_2 so that $A_t = F[0,t]$ a.e.

The extreme points of D_4 are then the vector measures F of the form

$$F(A) = \chi_{\sigma^{-1}(A)} \quad \text{for some} \quad \sigma : \Omega \to [0,1]$$

These are exactly the lattice orthogonally scattered measures introduced in [14].

III. Integral representation of operators:

Let now K be a Hausdorff topological space with a countable basis and let λ be a Radon probability on K . Let T be a bounded linear operator from $L_1(K,\lambda)$ into $L_1(\Omega,P)$. A disintegration theorem of Fakhoury [9], asserts that there exists and application $\mu : \Omega \to \mathcal{M}(K)$ so that

(i) $\mu_w = T1(w)\, \nu_w$ where ν_w are Radon probability measures on K

(ii) $w \to \mu_w$ is measurable if $\mathcal{M}(K)$ is equipped with the σ-field of weak-star Borel subsets.

(iii) Every f in $L_1(K,\lambda)$ is $|\mu_w|$-integrable for P-almost all w and $Tf(w) = \int_K f(t)d\mu_w(t)$.

(iv) $\int |\mu_w| dP(w) \leq \|T\| \cdot \lambda$

By combining this disintegration result and the representation of section I we get

Theorem III.1: If K is compact and T is a positive bounded linear operator from $L_1(K,\lambda)$ into $L_1(\Omega,F,P)$, then there exists a probability Radon measure $\tilde{\nu}$ on $(G,\bar{d})$ such that

1) For any f in $L_1(K,\lambda)$ we have

$$Tf(w) = T1(w)\int_G f(\sigma(w))d\tilde{\nu}(\sigma) \quad \text{for } P\text{-almost all } w$$

2) $A(D)$ is dense in $L^1(G,\tilde{\nu})$

Proof: Associate to T the random probability $\nu = (\nu_w)$ in D. By Theorem I.4, there exists a probability measure $\tilde{\nu}$ on $(G,\bar{d})$ verifying 2) and for any X in $L^1(\Omega,C(K))$

$$\langle X,\nu\rangle = \int E[X_\sigma]d\tilde{\nu}(\sigma) .$$

For any B in F and f in $C(K)$, the process $X(t,w) = T1(w)X_B(w).f(t)$ belongs to C. Hence,

$$\int_B(\int_K f(t)T1(w)d\mu_w(t))dP = \int_G(\int_B T1(w)f(\sigma(w))dP(w))d\tilde{\nu}(\sigma)$$

That is

$$\int_B Tf(w)dP(w) = \int_G(\int_B T1(w)f(\sigma(w))dP)\ d\tilde{\nu}(\sigma)$$

In order to apply Fubini's theorem on $G \times \Omega$, we still have to prove that for any $B \in F$ and $f \in C(K)$ the map

$$\psi : G \times \Omega \to \mathbb{R}$$

defined by $\psi(\sigma,w) = X_B(w)f(\sigma(w))$ is measurable for the $\tilde{\nu} \otimes P$ completion of the product σ-field on $G \times \Omega$. Actually, we prove the existence of a measurable version of the map $(\sigma,w) \to \sigma(w)$. That is a measurable map $\psi : G \overleftrightarrow{\times} \Omega \to \mathbb{R}$ so that for each $\sigma \in G$, we have for P-almost all w.

$$f \circ \psi(\sigma,w) = f(\sigma(w)) \text{ for all } f \text{ in } C(K).$$

It is clear that whenever we integrate with P, we can use $f(\sigma(w))$ instead of $f \circ \psi(\sigma,w)$, and we shall do so for almost everywhere equalities.

Since $(G,\bar{d})$ is separable, let (σ_n) be a dense sequence in G and let $A_{n,r}$ be the closed ball centered at σ_n and of radius $r > 0$.

For any $r > 0$, we have $G = \cup_n A_{n,r}$. Let $B_{n,r} = A_{n,r} \setminus \cup_{m<n} A_{m,r}$ and let $\tau_{n,r} \in B_{n,r}$ if $B_{n,r} \neq \emptyset$.

For any $k > 0$, we have $G = \cup_n B_{n,\frac{1}{k}}$. Define now $\psi_k(\sigma,w) = \sum_{n=1}^{\infty} \tau_{n,k} \chi_{B_{n,\frac{1}{k}}}(\sigma,w)$. That is $\psi_k(\sigma,w) = \tau_{n,\frac{1}{k}}(w)$ whenever $\sigma \in B_{n,\frac{1}{k}}$.

For any $\varepsilon > 0$, let $\ell' \geq \ell \geq 2(\frac{1+\epsilon}{\epsilon^2})$

We have

$$\begin{aligned}
&\tilde{\nu} \otimes P \{(\sigma,w); d(\psi_\ell(\sigma,w),\psi_{\ell'}(\sigma,w)) > \varepsilon\} = \\
&\sum_{n=1}^{\infty} \int_{B_{n,\frac{1}{\ell'}}} P\{w \in \Omega ; d(\tau_{n,\ell}(w), \tau_{n,\ell'}(w)) > \epsilon\} \\
&\leq \sum_{n=1}^{\infty} \int_{B_{n,\frac{1}{\ell'}}} \frac{1+\epsilon}{\epsilon} \cdot \bar{d}(\tau_{n,\ell}, \tau_{n,\ell'}) \cdot d\tilde{\nu}(\sigma) \\
&\leq \sum_{n=1}^{\infty} \int_{B_{n,\frac{1}{\ell'}}} \frac{1+\epsilon}{\epsilon} \cdot \frac{2}{\ell} \cdot d\tilde{\nu}(\sigma) \leq 2\frac{(1+\epsilon)}{\epsilon} \cdot \frac{\epsilon^2}{2(1+\epsilon)} = \varepsilon
\end{aligned}$$

(ψ_k) is then Cauchy in probability, hence it converges to ψ. A similar argument shows that for $\tilde{\nu}$ almost all $\sigma \in G$, $\psi(\sigma,w) = \sigma(w)$

P-almost everywhere. (Note that if $K = [0,\infty]$ one can choose ψ independently of T).

By applying Fubini's theorem for any $B \in F$, we get

$$\int_B Tf\, dP = \int_B T1(w) \int_G f(\psi(\sigma,w))d\tilde{\nu}(\sigma)dP = \int_B T1(w) \int_G f(\sigma(w))d\tilde{\nu}(\sigma)dP$$

That is for any $f \in C(K)$ $\quad Tf(w) = T1(w) \int f(\sigma(w))d\tilde{\nu}(\sigma)$ a.e.

Since $C(K)$ is dense in $L^1(K,\lambda)$, it is easy to show that the above equation extends to all functions in $L^1(K,\lambda)$.

Corollary III.1: If T is any bounded linear operator from $L_1(K,\lambda)$ into $L_1(\Omega,F,P)$, then there exists two probability measures $\tilde{\nu}_1$ and $\tilde{\nu}_2$ on $(G,\bar{d})$ such that for any f in $L_1(K,\lambda)$ we have

$$Tf(w) = T^+1(w) \int_G f(\sigma(w))d\tilde{\nu}_1(\sigma) - T^-1(w) \int_G f(\sigma(w))d\tilde{\nu}_2(\sigma)$$

For a proof it is enough to notice that every bounded linear operator on L_1 is the difference of two positive operators T^+ and T^-.

If now T is a positive operator from C(K) into $L_1(\Omega,F,P)$, then T extends to an operator from $L_1(K,\lambda)$ into $L_1(\Omega,P)$ where λ is the probability measure on K equal to $\frac{P \circ T}{\|T1\|_1}$. If we recall that an operator from C(K) into $L_1(\Omega,F,P)$ is said to be regular if it is the difference of two positive operators from C(K) into $L_1(\Omega,F,P)$, then the above theorem applies and we get

Corollary III.2: If T is a regular operator from C(K) into $L_1(\Omega,F,P)$, then there exists two probability Radon measures $\tilde{\nu}_1$ and $\tilde{\nu}_2$ on G so that for any $f \in C(K)$ we have

$$Tf(w) = T^+1(w) \int_G f(\sigma(w))d\tilde{\nu}_1(\sigma) - T^-1(w) \int_G f(\sigma(w))d\tilde{\nu}_2(\sigma) .$$

Corollary III.3: a) If $(A_t)_t$ is an increasing process on $[0,\infty]$, then there exists a probability Radon measure $\tilde{\mu}$ on G_2 such that

$$A_t = A_\infty \cdot \int I_{[0,t]}(\sigma)d\tilde{\mu}(\sigma) = A_\infty \cdot \tilde{\mu}\{\sigma; \cdot\, \sigma(w) \le t\}$$

b) If (A_t) is a randomized stopping time then G_2 may be replaced by the set G_3 of F_t-stopping times.

Proof: It is enough to notice that each increasing process defines an operator from the cadlag functions on $[0,1]$ into $L_1(\Omega,F,P)$ by

$$Tf = \int_0^\infty f(t)dA_t$$

where the integral is in the sense of Lebesque-Steltjes. Applying now Corollary III.2 to the function $\chi_{[0,t]}$ and note that $A_t = T\chi_{[0,t]}$ a.e.

Corollary III.4: If F is a positive $L_1(\Omega,F,P)$ - valued vector measure on the Borel subsets of $[0,1]$, then there exists a Radon probability measure on G_4 such that

$$F(A) = F[0,1] \int_{G_4} \chi_{\sigma^{-1}(A)} d\tilde{\mu}(\sigma) \quad \text{for any} \quad A \in \mathcal{B} .$$

Proof: Following [6], there exists a measure λ on $[0,1]$ so that $\lambda << F$. Moreover F defines a positive operator from $L_\infty(\lambda)$ into $L_1(\Omega,F,P)$ by

$$Tf = \int fdF$$

where the integral is in the sense of Bartle-Dunford and Schwartz.

Recall that a tree in $L_1(\Omega,F,P)$ is a family of functions $\{\psi_{n,k};\ n=0,1,\ldots;\ k=1,2,\ldots,2^n\}$ in $L_1(\Omega,F,P)$ verifying

$$2\psi_{n,k} = \psi_{n+1,2k-1} + \psi_{n+1,2k} \quad \text{for each } n,k .$$

Let $I_{n,k} = \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right]$ be the diadic intervals of $[0,1]$.

Corollary III.5: If $(\psi_{n,k})$ is a bounded positive tree in $L_1(\Omega,F,P)$, then there exists a Radon probability $\tilde{\mu}$ on G such that for any n and $1 \le k \le 2^n$.

$$\psi_{n,k} = 2^n \psi_{0,1} \int_G \chi_{\sigma^{-1}(I_{n,k})} d\tilde{\mu}(\sigma)$$

Proof: Associate to the tree $(\psi_{n,k})$ the operator T from $L_1[0,1]$ into $L_1(\Omega,F,P)$ defined by $T\frac{\chi_{I_{n,k}}}{2^{-n}}=\psi_{n,k}$ for each n and $1 \le k \le 2^n$ and which can be extended by linearity and continuity. Apply then Theorem III.1 to T.

The above representation can be useful for extending some properties which are easily verifiable in the case of operators induced by point transformations to more general operators. Here is an immediate application of this representation.

The Riesz-Thorin convexity theorem: Every bounded linear operator on L_1, whose restriction on L_∞ is also bounded, induces a bounded operator on each L_p, $1 < p < \infty$.

Proof: Suppose that T is a positive operator on L_1 so that $T1 \in L_\infty$. For any f in L_p, $(|f^p| \in L_1)$ we have:

$$\int_\Omega |Tf|^p dP = \int_\Omega |T1(w)|^p \left| \int_G f(\sigma(w)) d\tilde{\nu}(\sigma) \right|^p dP \le$$

$$\int_\Omega |T1|^{p-1} \int_G |T1| \left| f^p(\sigma) \right| d\tilde{\nu}(\sigma)\ dP \le$$

$$\|T1\|_\infty^{p-1} . \int_\Omega T|f|^p . dP \le \|T1\|_\infty^{p-1} . \|T1\|_1 . \|f^p\|_1$$

Hence $$\|T\|_p \le \|T1\|_\infty^{1-\frac{1}{p}} . \|T\|_1^{\frac{1}{p}} .$$

. Convolutions and multipliers:

Let K be a compact abelian group with a Haar measure λ, and let μ be a positive Radon measure on K. Define the convolution operator on $L_1(G)$ by

$$Tf(x) = \int_K f(y^{-1}x) d\mu(y)$$

It is clear that μ is a representation of T in the sense of theorem III.1. In this case $\tilde{\mu}$ is actually supported by the set of translations on K which can be identified with K and is a subset of G = {measurable transformations on K }.

More generally, let G be a group operating on a topological space K. If μ is a positive Radon measure on G, one can associate an operator on the bounded Borel functions on K by

$$Tf(x) = \int_G f(gx)d\mu(g) .$$

The above theorem shows that any operator T on $L_1(K,\lambda)$ with $T1 = 1$, is a "generalized" convolution where the canonical semi-group operating "measurably" on K is the non-abelian semi-group of the measurable transformations on K, equipped with the composition operation.

We may also write $Tf = f*\tilde{\mu}$, which makes T appear like a "randomized" multiplier. Note also that

$$T^2f(w) = \int_G \int_G f(\sigma\tau(w))d\tilde{\mu}(\tau)d\tilde{\mu}(\sigma)$$

That is the n^{th} iterate of T is given by the formula $T^nf = f*\tilde{\mu}^{*n}$.

It is enlightening at this stage to recall Wendell's theorem [12], which asserts that an operator T on L_1 of a compact abelian group K, which commutes with translations can be written as $Tf = f * \mu$ where μ is a Radon measure on K. The above representation shows that the measure μ always exists and that if T commutes with translations, then μ is supported by the group of translations. To give a proof of this fact, it is simpler to use the first representation, that is if T is a bounded linear operator on $L_1(K,\lambda)$, there exists a random measure $\mu : K \to \mathcal{M}(K)$ such that

$$Tf(x) = \int_K f(t)d\mu_x(t) \quad \text{for } \lambda\text{-almost all } x .$$

For each $y \in K$, denote by τ_y the translation operator associated to y. That is $\tau_y(x) = y^{-1}x$, $(\tau_y f)(x) = f(y^{-1}x)$ and $(\tau_y \mu)(f) = \mu(\tau_y f)$, where f is in $L_1(K)$ and μ is a Radon measure on K. Let e be the unit element in K.

The fact that T commutes with translations means that for any x and y in K we have $\tau_x \mu_y = \mu_{\tau_x y}$.

If $y = e$, we have $\tau_{x^{-1}} \mu_e = \mu_x$ for any $x \in K$. That is for any f in $L_1(K)$,

$$Tf(x) = \mu_x(f) = \int_K f(xy)d\mu_e(y) = \int_K f(y^{-1}x)d\nu_e(y)$$

where ν_e is the image of μ_e by the map $y \to y^{-1}$.

Another way to see it, is to show that the extreme points of the subset D_5 of D defined by

$$D_5 = \{(\mu_x) \in D \; ; \; (\tau_y \mu_x)_x = (\mu_{\tau_y x})_x \text{ for each } y \in K\},$$

are the random probabilities of the form δ_{τ_a} for some a in K.

. <u>Markov chains and random walks</u>:

Let $(K, \mathcal{B}, \lambda)$ be a compact separable space with a Radon probability measure λ on its Borel σ-field $\mathcal{B}$. Let P be a transition probability.

Let G be the semi-group of measurable transformations from K into itself. By Theorem [4] of [9] and Corollary (2), there exists a probability Radon measure $\tilde{\mu}$ on G such that for any $A \in B$ we have

$$P(x,A) = \int_G \chi_A(\sigma x)d\tilde{\mu}(\sigma) .$$

Let $\tilde{P}$ be the transition probability on G defined for any $\sigma \in G$ and bounded Borel function g on G by

$$\tilde{P}(\sigma,g) = \tilde{\mu} * \delta_\sigma(g) = \int_G g(\tau\sigma)d\tilde{\mu}(\tau) .$$

Let $(\Omega, F, F_n, X_n, P_\sigma)$ be the canonical Markov chain associated to it [11]; that is $X_n : (\Omega, F) \to G$ is a homogenuous Markov chain with respect to the σ-field (F_n) with transition probability $\tilde{P}$ and starting measures $(P_\sigma)_{\sigma\in G}$.

Let $\bar{\Omega} = K \times \Omega$, $\bar{F} = B \otimes F$, $\bar{F}_n = B \otimes F_n$ and let e be the identity transformation on K , and P_e the probability associated with the starting measure δ_e . If ν is a probability measure on K , we denote by $\bar{P}_\nu$ the probability measure $\nu \otimes P_e$ on $(\bar{\Omega}, \bar{F})$.

For $\bar{w} = (x,w)$, set $Y_o(\bar{w}) = x$ and

$$Y_n(\bar{w}) = X_n(w)(x) \quad \text{(the transformation } X_n(w) \text{ applied to } x)$$

One can show that (Y_n) is a Markov chain on K with respect to the σ-algebras $\bar{F}_n$ with transition probability equal to P and for any starting probability measure $\bar{P}_\nu$.

The chain is not the canonical chain associated to P . But it might be of interest to know that one can associate to any transition probability, a Markov chain induced by a "pseudo random walk" on a canonical semi-group G . The case of interest might be when $\tilde{\mu}$ is supported on the group G_6 in G of all the invertible transformations, since then $X_n(w)$ can be written as $X_0 Z_1 \ldots Z_n$ where the Z_i's are independent, identically distributed and of law $\tilde{\mu}$.

An interesting problem will be then to characterize the set D_6 of transition probabilities on K , whose extreme points are the transition probabilities induced by the invertible point transformations. It is clear that in case K is a group, D_6 contains D_5, and the Markov chains associated to elements in D_6 are the natural extensions of the random walks.

VI. Stronger topologies on $M(D)$:

Let (μ_n) be a sequence of random probabilities in D and let $(\tilde{\mu}_n)$ be a sequence of probabilities on G representing $(\tilde{\mu}_n)$. We have already seen that the vague convergence of $(\tilde{\mu}_n)$ is stronger than the weak convergence of Baxter-Chacon, since the elements of C(D) unlike those of A(D) are not induced by processes in $\mathcal{C}$. We shall see in the sequel that $(\tilde{\mu}_n)$ may converge on a large space of functions than C(D) and this convergence is strong enough to imply in some cases almost sure convergence.

Since D is not a simplex, the representing measures on (μ_n) are not unique. To keep more control on the representation, we prefer to select for each $\nu \in D$, a maximal measure $\tilde{\nu}$ on G which is simplicial; that is one, which is extreme in the set of all maximal measures representing ν. If $\tilde{\nu}$ is such a measure, then A(D) is dense in $L^1(G,\tilde{\nu})$.

Let now $\nu = \sum_{n=1}^{\infty} 2^{-n}\mu_n$; since the map $\nu \to \tilde{\nu}$ is not in general linear and continuous, we shall need the following lemma.

Lemma VI .1: There exists maximal Radon probabilities $\tilde{\mu}_n$ (resp $\tilde{\nu}$) supported on G, representing μ_n (resp ν) so that

(1) A(D) is dense in $L^1(\tilde{\nu})$

(2) $\sum_{n=1}^{\infty} 2^{-n}\tilde{\mu}_n = \tilde{\nu}$

Proof: Let $\tilde{\nu}$ be a simplicial and maximal measure on G representing ν. The two measures on D, $\tilde{\nu}$ and $\sum_{n=1}^{\infty} 2^{-n}\delta_{\mu_n}$ have the same barycenter ν, hence by [4], there exists a measurable map $\phi: D \to M(D)$ such that

(i) $\phi(\mu_n)$ is maximal for each n.

(ii) $\phi(\mu_n)$ and μ_n have the same barycenter for each n.

(iii) $\sum_{n=1}^{\infty} 2^{-n}\delta_{\mu_n}(\phi) = \tilde{\nu}$.

It is clear that $\tilde{\mu}_n = \phi(\mu_n)$ verify (1) and (2).

Note now that $A(D) \subset C(D) \subseteq L^1(\tilde{\nu})$ and that $\tilde{\mu}_n \in L^\infty(\tilde{\nu})$ for each n. But even though $(\tilde{\mu}_n)$ is relatively compact for the vague topology $\sigma(M(D), C(D))$, it is not necessarily bounded in $L^\infty(\tilde{\nu})$. Therefore, it is natural to put the $(\tilde{\mu}_n)$ in a space, where they are bounded, and such that this space is conjugate to a space between $C(D)$ and $L^1(\tilde{\nu})$.

For that let ν^* be the (possibly infinite) subadditive and positively homogenuous map on $L^1(\tilde{\nu})$ defined by

$$\nu^*(f) = \sup_n |\tilde{\mu}_n(f)|$$

Let E be the completion of the space $\{f \in L^1(\nu) ; \nu^*(|f|) < \infty\}$. It is immediate to see that E is a Banach lattice and that

$$C \subseteq \mathcal{C}(D) \subseteq E \subseteq L_1(\tilde{\nu})$$

where the injection maps are continuous.

The crucial fact is, of course, that $(\tilde{\mu}_n)$ is a $\sigma(E^*, E)$ relatively compact sequence in E^*, since it is in the positive ball of E^*. But, in general, E fails to be separable, hence we can only expect to find a subnet of $(\tilde{\mu}_n)$ which is convergent on the elements of E.

VII. Convergence of normalized positive operators:

Suppose now λ is a Radon probability measure on K and let (S_n) be a sequence of positive operators from $L_1(K,\lambda)$ into $L_1(\Omega,F,P)$ and let (μ_n) be the random probabilities associated to them. Set $\nu = \sum_{n=1}^{\infty} 2^{-n}\mu_n$. By combining the representation of section III, with the results of last section, we obtain that there exists Radon probability measures $(\tilde{\mu}_n)$, $\tilde{\nu}$ supported on G so that

(1) $\tilde{\nu} = \sum_{n=1}^{\infty} 2^{-n}\tilde{\mu}_n$

(2) A(D) is dense in $L_1(\tilde{\nu})$

(3) For each f in $L_1(K)$, we have

$$S_n f(w) = S_n 1(w) \int f o\psi(\sigma,w) d\tilde{\mu}_n(\sigma)$$

where $\psi : G \times \Omega \times \mathbb{R}$ is measurable with respect to the $\tilde{\nu} \otimes P$ completion of the product σ-field on $G \times \Omega$.

Note that each S_n can be extended to the cone of all positive and finite measurable functions on K , since if h is a function in this cone which is not necessarily in $L_1(K)$, we can define $S_n h$ as the limit of the sequence $(S_n h_m)_m$ where (h_m) is a sequence in $L_1(K,\lambda)$ increasing to h . It is standard to show that $S_n h$ does not depend on the particular sequence (h_m) and it is easy to see that equation (3) still holds for such an h .

Suppose now that $S_n 1 = 1$ and notice that if we fix $f \in L_1(K,\lambda)$, then almost all w's define an integrable function w_f in $L_1(G,\tilde{\mu}_n)$ via the map $w_f(\sigma) = f o\psi(\sigma,w)$ and that if $\sup_n |S_n f(w)| < \infty$, then w_f belongs to the space E since $\tilde{\mu}_n(w_f) = S_n f(w)$.

The connection between the convergence of the operators and the convergence of their representing measures is given by the following.

Proposition VII.1: a) If $\tilde{\mu}_n$ converges vaguely to $\tilde{\mu}$ then there exists an operator $S: L_1(K,\lambda) \to L_1(\Omega, \ ,P)$ so that $S_n f$ converges weakly to Sf for each f in $L_1(K,\lambda)$.

b) If $\tilde{\mu}_n$ converges to $\tilde{\mu}$ in $\sigma(E^*,E)$ then for every $f \in L_1(K,\lambda)$, $S_n f$ converges almost everywhere on the set $\{w; \sup_n S_n f(w) < \infty\}$.

If now a sequence of normalized operators is given, one can always find a subsequence of their representing measure which is convergent vaguely. Unfortunately, it is not the case for the almost sure convergence and in general one cannot expect to have convergence in $\sigma(E^*,E)$ of a subsequence of $(\tilde{\mu}_n)$ but of a subnet. The natural question is to find conditions on the sequence of operators (S_n) to insure that the ball of E^* is weak*-sequentially compact.

Another approach might be to find the right conditions to insure the existence of a simplicial $\tilde{\nu}$ so that $\tilde{\mu}_n$ are bounded in $L_\infty(\tilde{\nu})$. In this case the weak convergence of the operators will imply automatically the almost sure convergence (up to a maximal inequality) since $\mathcal{A}(D)$ is then dense in $L_1(G,\tilde{\nu})$.

Addendum: After this paper was written, M. Talagrand showed me a recent paper of Edgar-Millet-Sucheston entitled "On compactness and optimality of stopping times" in which the Choquet representation of randomized stopping times is used to deal with some optimal stopping problems.

References:

[1] S.R. Baxter, R.V. Chacon: "Compactness of stopping times." Z. Wahrs verw 40, p.169-181 (1977).

[2] I. Berkes, H.P. Rosenthal: "Almost exchangeable sequences of random variables." (Revised version) Preprint (1981).

[3] S.D. Chatterji:Ecole d'été de probabilités de Saint-Flour-Springer-Verlag. Berlin - Heidelberg - New York 307, (1973).

[4] G. Choquet: "Lectures in Analysis" Volume II, W.A. Benjamin Inc. Amsterdam (1969).

[5] C. Dellacherie: "Capacités et processes stochastiques" Vol. 67, Springer-Verlag, Berlin, Heidelberg, New York (1972).

[6] J. Diestel, J.J. Uhl: "Vector measures" Mathematical Surveys, No. 15, A.M.S., Providence (1977).

[7] N. Dunford, J. Schwartz: "Linear operators", Volume I, Interscience New York, (1958).

[8] G. Edgar: "Disintegration of measures and the vector-valued Radon-Nikodym theorem": Duke Math J. 42, 3, p.447-450 (1975).

[9] H. Fakhoury: "Représentation d'opérateurs ā valeurs dans $L_1(X,\Sigma,\mu)$: Math Annalen.

[10] P.A. Meyer: "Convergence faible et compacité des temps d'arrêt" d'après Baxter et Chacon. Sem de probabilites XII, 649, Springer-Verlag (1978).

[11] D. Revuz: "Markov Chains" North-Holland, American Elsevier (1975).

[12] W. Rudin: "Fourier analysis on groups." Interscience, New York, (1962).

[13] H.H. Schaeffer: "Banach lattices and positive operators": Berlin, Heidelberg, New York, Springer-Verlag (1974).

[14] K. Sundaresan, W. Woycynski: "L-orthogonally scattered measures." Pacific. J. of Math. 43, 785-797 (1972).

[15] M. Yor: "Sous-espaces denses dans L^1 on H^1 et représentation des martingales." Séminaire de probabilités XII, Lecture notes in Mathematics, 649, Springer-Verlag (1978).