PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 10, Fasc. 1 (1989), p. 75–92

INTEGRAL REPRESENTATION IN THE SET OF TRANSITION KERNELS*

BY

HARALD LUSCHGY (RHEINE)

Abstract. We prove a Choquet-type representation and uniqueness theorem for noncompact convex sets of transition kernels between a measurable space and a separable metrizable Radon space. Applications to sets of equivariant kernels and kernels with prescriped values are given. Furthermore, in the framework of statistical decision theory the representation is applied to sets of decision rules.

1. Introduction. Let $(X, \mathscr{B}(X), \mu)$ be a probability space and let Y be a separable metrizable Radon space equipped with its Borel σ-algebra $\mathscr{B}(Y)$. We will address ourselves to the study of topological and geometrical properties of sets of transition kernels from X to Y. Our main goal is to give a Choquet-type integral representation in closed convex sets $\mathscr{D}$ of kernels (Section 2). To this end the set of all kernels is embedded in a locally convex space of bilinear forms on $L^1(\mu)\times C(Y)$. Then each kernel $\delta\in\mathscr{D}$ is the barycenter of a probability measure ϱ on the extreme boundary ex $\mathscr{D}$ of $\mathscr{D}$ in the following sense: ϱ is defined on the cylinder σ-algebra on ex $\mathscr{D}$ and

$$\int_X \varepsilon_x\otimes\delta(x)\,d\mu(x) = \int_{\operatorname{ex}\mathscr{D}}\int_X \varepsilon_x\otimes\varphi(x)\,d\mu(x)\,d\varrho(\varphi)$$

holds. In particular, ex $\mathscr{D}\neq\emptyset$ if $\mathscr{D}\neq\emptyset$. Conversely, the barycenter of each probability measure on $\mathscr{D}$ is contained in $\mathscr{D}$ (Theorem 2.5). Such results are known for the case where $\mathscr{D}$ is the set of all kernels (see the classical paper of Wald and Wolfowitz [23] and [1], [11], [12], [20]). Related results for special sets of kernels occur in [25-28]. In Section 3 it is shown that the extreme boundary of the sets in question is measurable (Proposition 3.2) and an analogue of Choquet's uniqueness theorem holds provided $L^1(\mu)$ is separable (Theorem 3.3).

In Section 4 the above representation is applied to the set of all transition kernels which are equivariant with respect to the action of a group. The results of this section, except for Theorem 4.1, specify indications of Ferguson ([7], Chap. 4.2).

* This research was supported by the Deutsche Forschungsgemeinschaft.

The representation applied to sets of kernels with prescriped values yields an extension of Strassen's ([22], Theorem 3) generalization of the Blackwell–Stein–Sherman–Cartier theorem (Corollary 5.2). In a statistical framework the representation also implies that the risk function of a decision rule, equivariant rule, Bayes rule is a mixture of the risk functions of nonrandomized rules, nonrandomized equivariant rules, nonrandomized Bayes rules, respectively (Section 5).

Now we fix some notations and recall some definitions. Let $(X, \mathscr{B}(X))$ and $(Y, \mathscr{B}(Y))$ be measurable spaces. If Y is a topological space, then $\mathscr{B}(Y)$ denotes its Borel σ-algebra. $X \times Y$ is always equipped with the product σ-algebra $\mathscr{B}(X) \otimes \mathscr{B}(Y)$. Let $M(Y)$ be the space of all signed finite measures on $\mathscr{B}(Y)$ and $M_+^1(Y)$ the subset of all probability measures on $\mathscr{B}(Y)$. $M_+^1(Y)$ is equipped with the σ-algebra $\Sigma(M_+^1(Y))$ generated by the functions $\{Q \mapsto Q(C)\colon C \in \mathscr{B}(Y)\}$. A transition kernel δ from X to Y is a measurable map $\delta\colon X \to M_+^1(Y)$. We denote by $\mathscr{M}$ the set of all transition kernels from X to Y and by $\mathscr{N}$ the subset of all kernels $x \mapsto \varepsilon_{\varphi(x)}$ arising from measurable maps $\varphi\colon X \to Y$. ε_y is the point measure at $y \in Y$. If $(Z, \mathscr{B}(Z))$ is another measurable space, $\mathscr{M}^Z$ is the set of all kernels from X to Z. For each $\nu \in M(X)$ and $\delta \in \mathscr{M}$ let $\nu \otimes \delta$ denote the mixture $\int_X \varepsilon_x \otimes \delta(x)\, d\nu(x)$ in $M(X \times Y)$, and $\nu(\delta)$ the mixture $\int_X \delta(x)\, d\nu(x)$ in $M(Y)$. If $P \in M_+^1(X)$, denote by P_* the inner measure formed from P. $\mathscr{B}(X)_u$ stands for the universal completion of $\mathscr{B}(X)$.

A separable metrizable space is *universally measurable* (u.m.) if it is universally Borel measurable in its completion with respect to some and hence any metric defining the topology. It is well known that a separable metrizable space is u.m. if and only if it is a Radon space. A measurable space is u.m. if it is isomorphic to a separable metrizable u.m. space with its Borel σ-algebra. A Hausdorff topological space is *Souslin* if it is the continuous image of a Polish space; it is *Lusin* if it is the continuous injective image of a Polish space. A measurable space is *Souslin* if it is isomorphic to a Souslin topological space with its Borel σ-algebra. Every Souslin measurable space is u.m. ([8], III. 2.3, and [21], p. 124, Corollary 1). Note that there are Hausdorff topological spaces Y such that the measurable space $(Y, \mathscr{B}(Y))$ is Souslin without the topology being Souslin ([4], p. 9).

2. Integral representation and measure convexity. Fix a nonempty subset $\mathfrak{P}$ of $M_+^1(X)$. We denote by $L(\mathfrak{P})$ the band generated by $\mathfrak{P}$ in $M(X)$. Subsets $\mathscr{D}$ of $\mathscr{M}$ are always equipped with the σ-algebra $\Sigma(\mathscr{D})$ generated by the functions $\{\delta \mapsto \nu(\delta)(C)\colon \nu \in L(\mathfrak{P}),\ C \in \mathscr{B}(Y)\}$. Note that $\Sigma(\mathscr{D}) = \mathscr{D} \cap \Sigma(\mathscr{M})$ holds. By the monotone class theorem, $\Sigma(\mathscr{D})$ is also generated by the functions

$$\{\delta \mapsto \int_{X \times Y} u\, d\nu \otimes \delta\colon\ \nu \in L(\mathfrak{P}),\ u \in B(X \times Y)\},$$

where $B(X \times Y)$ denotes the space of all bounded measurable real-valued functions on $X \times Y$. A transition kernel $\delta \in \mathcal{M}$ is said to be *represented by* $\varrho \in M_+^1(\mathcal{D})$, $\mathcal{D} \subset \mathcal{M}$, if

$$P \otimes \delta = \int_{\mathcal{D}} P \otimes \varphi d\varrho(\varphi) \quad \text{for every } P \in M_+^1(X) \cap L(\mathfrak{P}).$$

A convex subset $\mathcal{D}$ of $\mathcal{M}$ is said to have the *integral representation property* if for each kernel $\delta \in \mathcal{D}$ there is a measure $\varrho \in M_+^1(\operatorname{ex} \mathcal{D})$ which represents δ. For $\mathcal{M}$ we obviously obtain $\operatorname{ex} \mathcal{M} = \mathcal{N}$ provided $\mathcal{B}(Y)$ is countably generated and contains singletons. A subset $\mathcal{D}$ of $\mathcal{M}$ is called *measure convex* if for each $\varrho \in M_+^1(\mathcal{D})$ there is a kernel $\delta \in \mathcal{D}$ which is represented by ϱ.

The following proposition is an immediate consequence of Kirschner's [12] extension to Souslin topological spaces of a result of Wald and Wolfowitz [23] on randomization in statistics. It will be needed in Section 4.

PROPOSITION 2.1. *Let $(Y, \mathcal{B}(Y))$ be a Souslin measurable space. Then $\mathcal{M}$ has the integral representation property.*

Proof. Let $\delta \in \mathcal{M}$. Since Y can be equipped with a Souslin topology compatible with the given measurable structure, there is a measure $\varrho \in M_+^1(\mathcal{N})$ such that

$$v(\delta) = \int_{\mathcal{N}} v(\varphi)\, d\varrho(\varphi) \quad \text{for every } v \in L(\mathfrak{P})$$

(Kirschner [12]). Then

$$P \otimes \delta (A \times C) = \int_{\mathcal{N}} P \otimes \varphi (A \times C)\, d\varrho(\varphi)$$

for every $P \in M_+^1(X) \cap L(\mathfrak{P})$, $A \in \mathcal{B}(X)$, and $C \in \mathcal{B}(Y)$, because $P(A \cap \cdot) \in L(\mathfrak{P})$. This yields $P \otimes \delta = \int_{\mathcal{N}} P \otimes \varphi d\varrho(\varphi)$. So δ is represented by ϱ.

It would be interesting to know whether $\mathcal{M}$ is measure convex for arbitrary sets $\mathfrak{P}$.

For the rest of this section we assume $\mathfrak{P} = \{\mu\}$ for some probability measure μ on $\mathcal{B}(X)$. Then by the Radon-Nikodym theorem, $L(\mathfrak{P}) = \{f \cdot \mu \colon f \in L^1(\mu)\}$, where $f \cdot \mu(A) = \int_A f d\mu$ for every $A \in \mathcal{B}(X)$. Let us identify transition kernels that differ only on a μ-null set. This corresponds to considering the set $\mathcal{M}(\mu)$ of all equivalence classes of kernels from X to Y. For $\mathcal{D} \subset \mathcal{M}$ let $\mathcal{D}(\mu)$ be the members of $\mathcal{M}(\mu)$ which contains a representant in $\mathcal{D}$. Generally, we will not distinguish in our notation between equivalence classes and their representants. It is not hard to verify that $\operatorname{ex} \mathcal{M}(\mu) = \mathcal{N}(\mu)$ holds provided $\mathcal{B}(Y)$ is countably generated and contains singletons.

Now suppose that Y is a separable metrizable space or a completely regular Souslin space. Since $\mathcal{B}(Y)$ is countably generated and coincides with the Baire σ-algebra on Y ([8], III.2.1), the map from $\mathscr{M}(\mu)$ to the space $B(L^1(\mu), C(Y))$ of all continuous bilinear forms on $L^1(\mu)\times C(Y)$, defined by

$$\delta \mapsto \Big((f, k) \mapsto \int_Y kdf \cdot \mu(\delta)\Big),$$

is injective, where $C(Y)$ is the space of all bounded continuous real-valued functions on Y. We thus can identify $\mathscr{M}(\mu)$ with a subset of $B(L^1(\mu), C(Y))$. The locally convex Hausdorff topology $\sigma\big(B(L^1(\mu), C(Y)), L^1(\mu)\otimes C(Y)\big)$ on $B(L^1(\mu), C(Y))$ and the induced topology on any subset of $\mathscr{M}(\mu)$ will be denoted by $\tau(Y)$ or, simply by τ.

LEMMA 2.2. *$\Sigma(\mathscr{D})$ is generated by $\{u|\mathscr{D}\colon u\in\big(B(L^1(\mu), C(Y)), \tau\big)'\}$ for every $\mathscr{D}\in\mathscr{M}(\mu)$.*

Proof. Let $\Sigma_0(\mathscr{D})$ denote the σ-algebra on $\mathscr{D}$ generated by the τ-continuous linear functionals. Since $\Sigma(\mathscr{D})$ is obviously generated by the functions

$$\{\delta \mapsto \int_Y kdf \cdot \mu(\delta)\colon f\in L^1(\mu),\ k\in B(Y)\},$$

we have $\Sigma_0(\mathscr{D})\subset\Sigma(\mathscr{D})$. To prove the converse inclusion, let $f\in L^1(\mu)$ and

$$V = \{k\in B(Y)\colon \delta\mapsto \int_Y k\,df\cdot\mu(\delta) \text{ is } \Sigma_0(\mathscr{D})\text{-measurable}\}.$$

Then V is a vector space which contains $C(Y)$ and is closed under bounded monotone convergence. By the functional form of the monotone class theorem ([5], p. 15), this implies $V = B(Y)$ and hence $\Sigma(\mathscr{D})\subset\Sigma_0(\mathscr{D})$.

The proof of the following lemma is left to the reader.

LEMMA 2.3. *Let $\mathscr{D}\subset\mathscr{M}(\mu)$, $\delta\in\mathscr{M}(\mu)$, and $\varrho\in M_+^1(\mathscr{D})$. Then the following statements are equivalent:*

(i) *δ is represented by ϱ.*

(ii) $\mu\otimes\delta = \int_{\mathscr{D}} \mu\otimes\varphi d\varrho(\varphi)$.

(iii) *$\delta = r(\varrho)$ in $\big(B(L^1(\mu), C(Y)), \tau\big)$, where $r(\varrho)$ denotes the barycenter of ϱ.*

We also need the following information.

LEMMA 2.4. *Let $(X, \mathscr{B}(X))$ and $(Y, \mathscr{B}(Y))$ be measurable spaces.*

(a) $\mathscr{B}(X)_u\otimes\mathscr{B}(Y)_u\subset\big(\mathscr{B}(X)\otimes\mathscr{B}(Y)\big)_u$.

(b) *If $B\in\big(\mathscr{B}(X)\otimes\mathscr{B}(Y)\big)_u$, then the section B_x belongs to $\mathscr{B}(Y)_u$ for every $x\in X$.*

(c) *Let $B\in\big(\mathscr{B}(X)\otimes\mathscr{B}(Y)\big)_u$ and $\delta\in\mathscr{M}$. Then the function $X\to\mathbf{R}$, $x\mapsto\delta(x, B_x)$ is universally measurable. In particular, $x\mapsto\delta(x, C)$ is universally measurable for every $C\in\mathscr{B}(Y)_u$.*

(d) *Let $\mathscr{D}\subset\mathscr{M}(\mu)$ and $B\in\big(\mathscr{B}(X)\otimes\mathscr{B}(Y)\big)_u$. Then the function $\mathscr{D}\to\mathbf{R}$,*

$\delta \mapsto \mu \otimes \delta(B)$ is universally measurable. In particular, $\delta \mapsto \mu(\delta)(C)$ is universally measurable for every $C \in \mathscr{B}(Y)_u$.

Proof. The proofs of (a)-(c) are certainly well known and omitted.

(d) Let $\varrho \in M_+^1(\mathscr{D})$. There exist $B_1, B_2 \in \mathscr{B}(X) \otimes \mathscr{B}(Y)$ such that $B_1 \subset B \subset B_2$ and

$$\int_{\mathscr{D}} \mu \otimes \delta(B_2 \setminus B_1)\, d\varrho(\delta) = 0.$$

We obtain $\varrho(\{\delta \in \mathscr{D}: \mu \otimes \delta(B_2 \setminus B_1) = 0\}) = 1$. Since the functions $\mathscr{D} \to \boldsymbol{R}$, $\delta \mapsto \mu \otimes \delta(B_i)$ are measurable ($i = 1, 2$), we conclude that $\delta \mapsto \mu \otimes \delta(B)$ is measurable with respect to the ϱ-completion of $\Sigma(\mathscr{D})$ and hence, ϱ being arbitrary, is universally measurable.

The narrow topology $\sigma(M(Y), C(Y))$ on $M(Y)$ and the induced topology on any subset of $M_+^1(Y)$ is denoted by $w(Y)$ or, simply, by w.

We come to the main result of this section.

THEOREM 2.5. *Let Y be a separable metrizable u.m. space. Then every $\tau(Y)$-closed convex subset of $\mathscr{M}(\mu)$ has the integral representation property and is measure convex.*

Proof. Choose a totally bounded metric inducing the topology of Y. Then the completion Z of Y is compact and $\mathscr{B}(Y) \subset \mathscr{B}(Z)_u$ holds. It is known that $\mathscr{M}^Z(\mu)$ is a $\tau(Z)$-compact subset of $B(L^1(\mu), C(Z))$ (Farrell [6] or Luschgy and Mussmann [16]). Since, by Portmanteau's theorem, the map i: $(M_+^1(Y), w(Y)) \to (\{Q \in M_+^1(Z): Q(Y) = 1\}, w(Z))$ defined by $i(Q)(C) = Q(Y \cap C)$ for every $C \in \mathscr{B}(Z)$ is a homeomorphism, it follows that the map

$$(\mathscr{M}(\mu), \tau(Y)) \to (\{\delta \in \mathscr{M}^Z(\mu): \mu(\delta)(Y) = 1\}, \tau(Z)),$$

$\delta \mapsto$equivalence class of $i \circ \delta'$ for some representant δ' of δ is also a homeomorphism. Note that Lemma 2.4(c) assures that this map is surjective. In the following we identify $(\mathscr{M}(\mu), \tau(Y))$ with the subspace $\{\delta \in \mathscr{M}^Z(\mu): \mu(\delta)(Y) = 1\}$ of $(\mathscr{M}^Z(\mu), \tau(Z))$.

Now let $\mathscr{D}$ be a $\tau(Y)$-closed convex subset of $\mathscr{M}(\mu)$ and $\delta \in \mathscr{D}$. Let $\mathscr{D}^-$ denote the $\tau(Z)$-closure of $\mathscr{D}$ in $\mathscr{M}^Z(\mu)$. Then $\mathscr{D}^-$ is $\tau(Z)$-compact and convex. From Lemma 2.2 and the Stone–Weierstrass theorem it follows that $\Sigma(\mathscr{D}^-)$ coincides with the Baire σ-algebra on $\mathscr{D}^-$. Therefore, by the theorem of Bishop-de Leeuw, there exists a $\varrho \in M_+^1(\operatorname{ex} \mathscr{D}^-)$ such that $\delta = r(\varrho)$ in $(B(L^1(\mu), C(Z)), \tau(Z))$. In view of Lemmas 2.3 and 2.4(a) this implies

$$\mu \otimes \delta(B) = \int_{\operatorname{ex} \mathscr{D}^-} \mu \otimes \varphi(B)\, d\varrho(\varphi)$$

for every $B \in \mathscr{B}(X) \otimes \mathscr{B}(Z)_u$. In particular, choosing $B = X \times Y$, we have

$$1 = \mu(\delta)(Y) = \int_{\operatorname{ex} \mathscr{D}^-} \mu(\varphi)(Y)\, d\varrho(\varphi).$$

Since, by Lemma 2.4(d), the function $\operatorname{ex}\mathscr{D}^- \to \boldsymbol{R}$, $\varphi \mapsto \mu(\varphi)(Y)$ is universally measurable, there is a ϱ-null set $N \in \Sigma(\operatorname{ex}\mathscr{D}^-)$ such that $\mu(\varphi)(Y) = 1$ for every $\varphi \in \operatorname{ex}\mathscr{D}^- \setminus N$, that is, $\operatorname{ex}\mathscr{D}^- \setminus N \subset \mathscr{M}(\mu)$. Since $\mathscr{D}$ is $\tau(Y)$-closed in $\mathscr{M}(\mu)$, we have $\mathscr{D}^- \cap \mathscr{M}(\mu) = \mathscr{D}$ and, therefore, $\operatorname{ex}\mathscr{D}^- \setminus N \subset \operatorname{ex}\mathscr{D}$. Furthermore, $\mathscr{D}$ is an extremal subset of $\mathscr{D}^-$ which yields $\operatorname{ex}\mathscr{D} = \mathscr{D} \cap \operatorname{ex}\mathscr{D}^-$. Thus we obtain $\varrho_*(\operatorname{ex}\mathscr{D}) = 1$. So we can define a probability measure ϱ_0 on $\Sigma(\operatorname{ex}\mathscr{D})$ by $\varrho_0(\operatorname{ex}\mathscr{D} \cap F) = \varrho(F)$, $F \in \Sigma(\operatorname{ex}\mathscr{D}^-)$. Then it is clear that

$$\mu\otimes\delta = \int_{\operatorname{ex}\mathscr{D}} \mu\otimes\varphi \, d\varrho_0(\varphi)$$

holds, and, by Lemma 2.3, this implies that ϱ_0 represents δ.

In order to prove that $\mathscr{D}$ is measure convex, let $\varrho \in M_+^1(\mathscr{D})$. Then $r(\varrho) \in \mathscr{D}^-$, where $r(\varrho)$ denotes the barycenter of ϱ in $\big(B(L^1(\mu), C(Z)), \tau(Z)\big)$ ([19], Proposition 1.1). By Lemmas 2.3 and 2.4(a), for the kernel $\delta = r(\varrho)$ we have

$$\mu\otimes\delta(B) = \int_{\mathscr{D}} \mu\otimes\varphi(B) \, d\varrho(\varphi)$$

for every $B \in \mathscr{B}(X)\otimes\mathscr{B}(Z)_u$. In particular,

$$\mu(\delta)(Y) = \int_{\mathscr{D}} \mu(\varphi)(Y) \, d\varrho(\varphi) = 1$$

holds and hence $\delta \in \mathscr{D}$.

If Y is not u.m., then τ-closed convex subsets of $\mathscr{M}(\mu)$ do not always have extreme points, even when $|X| = 1$ ([24], Counterexample 3).

For nonmetrizable spaces Y we have the following version of the theorem:

COROLLARY 2.6.(a) *Let Y be a completely regular Souslin space. Then every $\tau(Y)$-closed convex subset of $\mathscr{M}(\mu)$ is measure convex.*

(b) *Let Y be a completely regular Lusin space. Then every $\tau(Y)$-closed convex subset of $\mathscr{M}(\mu)$ has the integral representation property.*

Proof. (a) Let $\mathscr{D}$ be a $\tau(Y)$-closed convex subset of $\mathscr{M}(\mu)$ and $\varrho \in M_+^1(\mathscr{D})$. Since Y can be equipped with a metrizable Souslin topology compatible with the given Borel structure ([8], III.2.3), it follows from the preceding theorem that there exists a $\delta \in \mathscr{M}(\mu)$ which is represented by ϱ. An application of the Hahn-Banach theorem yields $\delta \in \mathscr{D}$.

(b) Choose a Polish topology on Y finer than the given one and observe that it is compatible with the given Borel structure [21], p. 108, Lemma 17). Hence, the assertion follows from Theorem 2.5.

Theorem 2.5 provides, in case $\mathfrak{P} = \{\mu\}$, an extension of Proposition 2.1.

COROLLARY 2.7. *Let $\big(Y, \mathscr{B}(Y)\big)$ be u.m. Then $\mathscr{M}(\mu)$ has the integral representation property and is measure convex.*

This corollary comprises the representation theorems given in [1], [11], and [20].

By the way, under the above hypotheses the measure convexity of $\mathscr{M}$ may also be proved as follows. Let $\varrho \in M_+^1(\mathscr{M})$ and let Q be the probability measure $\int_{\mathscr{M}} \mu \otimes \varphi d\varrho(\varphi)$ on $\mathscr{B}(X) \otimes \mathscr{B}(Y)$ with X-marginal μ. By a well known disintegration theorem, there is a kernel $\delta \in \mathscr{M}$ such that $Q = \mu \otimes \delta$. This implies that ϱ represents δ.

The final result in this section will be needed in Section 4. Let Γ_0 be a multifunction from X to Y whose graph $\mathrm{Gr}(\Gamma_0) = \{(x, y) \in X \times Y: y \in \Gamma_0(x)\}$ belongs to $(\mathscr{B}(X) \otimes \mathscr{B}(Y))_u$. Define a multifunction Γ from X to $M_+^1(Y)$ by

$$\Gamma(x) = \{Q \in M_+^1(Y): Q(\Gamma_0(x)) = 1\}.$$

Γ is well defined, since, by Lemma 2.4(b), $\Gamma_0(x) = \mathrm{Gr}(\Gamma_0)_x \in \mathscr{B}(Y)_u$. Put

$$\mathscr{M}_\Gamma = \{\delta \in \mathscr{M}: \delta(x) \in \Gamma(x) \text{ for } \mu\text{-almost every } x \in X\}$$

and $\mathscr{N}_\Gamma = \mathscr{N} \cap \mathscr{M}_\Gamma$.

Corollary 2.8. *Let $(Y, \mathscr{B}(Y))$ be u.m. Then $\mathscr{M}_\Gamma(\mu)$ has the integral representation property and is measure convex. Furthermore,* $\operatorname{ex} \mathscr{M}_\Gamma(\mu) = \mathscr{N}_\Gamma(\mu)$ *holds.*

Proof. Note that

$$\mathscr{M}_\Gamma(\mu) = \{\delta \in \mathscr{M}(\mu): \mu \otimes \delta(\mathrm{Gr}(\Gamma_0)) = 1\}.$$

Since $\mathscr{M}_\Gamma(\mu)$ is an extremal subset of $\mathscr{M}(\mu)$, we obtain $\operatorname{ex} \mathscr{M}_\Gamma(\mu) = \mathscr{N}_\Gamma(\mu)$. Let $\delta \in \mathscr{M}_\Gamma(\mu)$. By Corollary 2.7, there is a measure $\varrho \in M_+^1(\mathscr{N}(\mu))$ which represents δ. In particular, we have

$$1 = \mu \otimes \delta(\mathrm{Gr}(\Gamma_0)) = \int_{\mathscr{N}(\mu)} \mu \otimes \varphi(\mathrm{Gr}(\Gamma_0)) d\varrho(\varphi).$$

Since, by Lemma 2.4(d), the function $\mathscr{N}(\mu) \to \boldsymbol{R}$, $\varphi \mapsto \mu \otimes \varphi(\mathrm{Gr}(\Gamma_0))$ is universally measurable, there is a ϱ-null set $N \in \Sigma(\mathscr{N}(\mu))$ such that $\mu \otimes \varphi(\mathrm{Gr}(\Gamma_0)) = 1$ for every $\varphi \in \mathscr{N}(\mu) \setminus N$, that is, $\mathscr{N}(\mu) \setminus N \subset \mathscr{N}_\Gamma(\mu)$. Thus we obtain $\varrho_*(\mathscr{N}_\Gamma(\mu)) = 1$. So the probability measure ϱ_0 on $\Sigma(\mathscr{N}_\Gamma(\mu))$, defined by $\varrho_0(\mathscr{N}_\Gamma(\mu) \cap F) = \varrho(F)$ for every $F \in \Sigma(\mathscr{N}(\mu))$, represents δ. Now let $\varrho \in M_+^1(\mathscr{M}_\Gamma(\mu))$. By Corollary 2.7, there is a kernel $\delta \in \mathscr{M}(\mu)$ which is represented by ϱ. Since

$$\mu \otimes \delta(\mathrm{Gr}(\Gamma_0)) = \int_{\mathscr{M}_\Gamma(\mu)} \mu \otimes \varphi(\mathrm{Gr}(\Gamma_0)) d\varrho(\varphi) = 1,$$

δ belongs to $\mathscr{M}_\Gamma(\mu)$.

A more general version of the corollary will be proved in Section 5.

3. Uniqueness and measurability of the set of extreme points. In this section we assume $\mathfrak{P} = \{\mu\}$ for some $\mu \in M_+^1(X)$ such that $L^1(\mu)$ is separable. $\mathscr{M}(\mu)$ is equipped with the topology τ. We begin with some topological properties of $\mathscr{M}(\mu)$.

PROPOSITION 3.1.(a) *Let Y be a separable metrizable space. Then $\mathscr{M}(\mu)$ is also separable metrizable. Further $\mathscr{M}(\mu)$ is compact, respectively Polish, Lusin, Souslin, u.m. if and only if Y has the same property.*

(b) *Let Y be a completely regular Souslin space. Then $\mathscr{M}(\mu)$ is also Souslin. Further $\mathscr{M}(\mu)$ is Lusin if and only if Y is Lusin.*

Proof. (a) Let Z be the compact completion of Y with respect to a totally bounded metrization. Then by Portmanteau's theorem, $\mathscr{M}(\mu)$ is homeomorphic with a subspace of the compact space $(\mathscr{M}^Z(\mu), \tau(Z))$. Since $L^1(\mu)$ and $C(Z)$ are separable, $(\mathscr{M}^Z(\mu), \tau(Z))$ is metrizable and hence $\mathscr{M}(\mu)$ is separable metrizable. If Y is Polish, so is $\mathscr{M}(\mu)$ ([1], 5.2). If Y is Lusin, Souslin respectively, then, by (b), $\mathscr{M}(\mu)$ has the same property. If Y is u.m., then $\mathscr{M}(\mu)$ is homeomorphic with $(\{\delta \in \mathscr{M}^Z(\mu)\colon \mu(\delta)(Y) = 1\}, \tau(Z))$ and, in view of Lemmas 2.2 and 2.4(d), this implies that $\mathscr{M}(\mu)$ is u.m. In order to prove the converse, let δ_y be the kernel $x \mapsto \varepsilon_y$ for $y \in Y$. Then the map $Y \to \{\delta_y\colon y \in Y\}(\mu)$, $y \mapsto$equivalence class of δ_y is a homeomorphism and $\{\delta_y\colon y \in Y\}(\mu)$ is a closed subset of $\mathscr{M}(\mu)$ ([18], Lemmas II.6.1 and II.6.2). Thus, if $\mathscr{M}(\mu)$ has one of the above properties, then Y has the same property ([21], p. 95, Theorem 2, p. 96, Theorem 3, p. 118, Proposition 8).

(b) Let $p\colon Z \to Y$ be a continuous surjection of a Polish space Z onto Y. Then the image measure map

$$\bar{p}\colon \big(M_+^1(Z), w(Z)\big) \to \big(M_+^1(Y), w(Y)\big)$$

is also a continuous surjection ([5], III.45) and the same is true for the map $(\mathscr{M}^Z(\mu), \tau(Z)) \to \mathscr{M}(\mu)$, $\delta \mapsto$equivalence class of $\bar{p} \circ \delta'$ for some representant δ' of δ. Indeed, this map is clearly continuous. Further let $\varphi \in \mathscr{M}(\mu)$ and φ' be a representant of φ. Since $(M_+^1(Y), w)$ is Souslin and $\Sigma(M_+^1(Y)) = \mathscr{B}(M_+^1(Y), w)$ ([21], p. 385, Theorem 7, and p. 387, Theorem 8), $\bar{p}$ admits a universally measurable right inverse q ([8], III.11.7). Choosing a kernel $\delta' \in \mathscr{M}^Z$ such that $\delta' = q \circ \varphi'$ μ-almost everywhere, we obtain $\bar{p} \circ \delta' = \varphi'$ μ-almost everywhere. Thus the above map is surjective. In view of (a) this implies that $\mathscr{M}(\mu)$ is Souslin. Since Y is homeomorphic with the sequentially closed subset $\{\delta_y\colon y \in Y\}(\mu)$ of $\mathscr{M}(\mu)$, Y is Lusian when $\mathscr{M}(\mu)$ has this property ([21], p. 102, Corollary 1, and p. 95, Theorem 2). Conversely, if Y is Lusin and $\mathscr{O}$ a Polish topology on Y finer than the given one, then $\mathscr{B}(Y, \mathscr{O}) = \mathscr{B}(Y)$, $\tau(Y, \mathscr{O})$ is finer than $\tau(Y)$ and, by (a), $(\mathscr{M}(\mu), \tau(Y, \mathscr{O}))$ is Polish. Hence $\mathscr{M}(\mu)$ is Lusin.

The next proposition deduces the measurability of the set of extreme points of a closed convex subset of $\mathcal{M}(\mu)$ from the topological properties of $\mathcal{M}(\mu)$.

PROPOSITION 3.2. (a) *If Y is a separable metrizable space or a completely regular Souslin space, then $\Sigma(\mathcal{D}) = \mathcal{B}(\mathcal{D})$ for every subset $\mathcal{D}$ of $\mathcal{M}(\mu)$.*

(b) *If Y is a completely regular Souslin space, then $\operatorname{ex}\mathcal{D} \in \Sigma(\mathcal{D})_u$ for every closed convex subset $\mathcal{D}$ of $\mathcal{M}(\mu)$.*

(c) *If Y is a separable metrizable u.m. space or a completely regular Lusin space, then $\operatorname{ex}\mathcal{D} \in \Sigma(\mathcal{D})$ for every closed convex subset $\mathcal{D}$ of $\mathcal{M}(\mu)$.*

Proof. (a) It suffices to prove the assertion for $\mathcal{D} = \mathcal{M}(\mu)$. According to Proposition 3.1, $\mathcal{M}(\mu)$ is strongly Lindelöf. Since the topology τ has a base consisting of $\Sigma(\mathcal{M}(\mu))$-measurable sets, $\Sigma(\mathcal{M}(\mu)) = \mathcal{B}(\mathcal{M}(\mu))$ follows from Lemma 2.2 and the strong Lindelöf property.

(b) Let $\mathcal{D}$ be a closed convex subset of $\mathcal{M}(\mu)$. By Proposition 3.1(b), $\mathcal{M}(\mu)$ is Souslin. Then $\mathcal{D}$ is also Souslin and hence $\operatorname{ex}\mathcal{D} \in \mathcal{B}(\mathcal{D})_u$ (Jayne and Rogers [10]). The assertion now follows from (a).

(c) If Y is separable metrizable u.m., then the assertion follows from (a), Proposition 3.1(a), and Proposition 1.3 in [19] by embedding $\mathcal{M}(\mu)$ in $(\mathcal{M}^Z(\mu), \tau(Z))$ for some compact metrizable space Z. If Y is completely regular Lusin, then the assertion follows from the preceding (see the proof of Corollary 2.6(b)).

Now we show that in the situation of Theorem 2.5 an analogue of Choquet's uniqueness theorem holds.

THEOREM 3.3. *Let Y be a separable metrizable u.m. space and $\mathcal{D}$ a closed convex subset of $\mathcal{M}(\mu)$. Then, for each kernel in $\mathcal{D}$, there is a unique representing measure in $M_+^1(\operatorname{ex}\mathcal{D})$ if and only if $\mathcal{D}$ is a simplex.*

Proof. The "only if" part. By Theorem 2.5, $r(\varrho) \in \mathcal{D}$ holds for each $\varrho \in M_+^1(\operatorname{ex}\mathcal{D})$. The barycentric map $r\colon M_+^1(\operatorname{ex}\mathcal{D}) \to \mathcal{D}$ is an affine bijection. This implies that $\mathcal{D}$ is a simplex.

The "if" part. Choose a totally bounded metric on Y defining the topology and let Z be the (compact) completion of Y. We can identify $\mathcal{M}(\mu)$ with the subspace $\{\delta \in \mathcal{M}^Z(\mu)\colon \mu(\delta)(Y) = 1\}$ of $(\mathcal{M}^Z(\mu), \tau(Z))$. Let $\mathcal{D}^-$ denote the $\tau(Z)$-closure of $\mathcal{D}$. Then by Proposition 3.1(a), $\mathcal{D}^-$ is $\tau(Z)$-compact metrizable and convex. Note that $\mathcal{D}^-$ is contained in the $\tau(Z)$-closed hyperplane $\{T \in B(L^1(\mu), C(Z))\colon 1_X \otimes 1_Z(T) = 1\}$. Let $\delta \in \mathcal{D}$. We denote by K_1 the cone generated by $\mathcal{D}$, and by K_2 the cone generated by $\mathcal{D}^-$. Further let $\leqslant_i$ denote the induced orderings on K_i, i.e. $\psi \leqslant_i \varphi$ if and only if $\varphi - \psi \in K_i$, $i = 1, 2$. Since $\mathcal{D}^- \cap \mathcal{M}(\mu) = \mathcal{D}$, $\varphi \in K_1$ and $\psi \in K_2$ $(\psi \leqslant_2 \varphi)$ imply $\psi \in K_1$ and both orderings coincide on K_1. Thus

$$\{\varphi \in K_1\colon \varphi \leqslant_1 \delta\} = \{\varphi \in K_2\colon \varphi \leqslant_2 \delta\}.$$

By assumption, K_1 is a lattice (in the ordering $\leqslant_1$) and so

$\{\varphi \in K_1: \varphi \leqslant_1 \delta\}$ is a lattice. Hence $\{\varphi \in K_2: \varphi \leqslant_2 \delta\}$ is a lattice (in the ordering $\leqslant_2$). In particular, $\{\varphi \in K_2: \varphi \leqslant_2 \delta\}$ has the Riesz decomposition property. Furthermore, by Proposition 3.2(a), $\Sigma(\mathscr{D}^-) = \mathscr{B}(\mathscr{D}^-)$ holds. Now the stage has been set for an application of the Loomis uniqueness theorem: there is a unique measure $\varrho \in M^1_+(\mathrm{ex}\,\mathscr{D}^-)$ such that $\delta = r(\varrho)$ in $(B(L^1(\mu), C(Z)), \tau(Z))$ ([17], XI.24, and the proof of XI.29). By Theorem 2.5, there is a measure $\varrho_0 \in M^1_+(\mathrm{ex}\,\mathscr{D})$ such that $\delta = r(\varrho_0)$ in $(B(L^1(\mu), C(Y)), \tau(Y))$. Since $\mathrm{ex}\,\mathscr{D} = \mathscr{D} \cap \mathrm{ex}\,\mathscr{D}^-$, we can define a probability measure ϱ_1 on $\Sigma(\mathrm{ex}\,\mathscr{D}^-)$ by $\varrho_1(F) = \varrho_0(\mathrm{ex}\,\mathscr{D} \cap F)$. Then $\delta = r(\varrho_1)$ in $(B(L^1(\mu), C(Z)), \tau(Z))$ and hence $\varrho_1 = \varrho$. This yields the uniqueness of ϱ_0 (cf. Lemma 2.3).

By the way, the separability of $L^1(\mu)$ is not used in the "only if" part of the theorem. This part also holds for completely regular Souslin spaces Y. The "if" part is also valid for completely regular Lusin spaces Y. Simple examples show that $\mathscr{M}(\mu)$ is no simplex.

4. Integral representation in the set of equivariant transition kernels. This section was inspired by Ferguson ([7], Chap. 4.2). Let G be a group which acts (from the left) on X and Y. G is equipped with a σ-algebra $\mathscr{B}(G)$ and we assume that the maps $G \to G$, $g \mapsto g^{-1}$, $G \times X \to X$, $(g, x) \mapsto gx$, $G \times Y \to Y$, $(g, y) \mapsto gy$ are measurable. Then G acts on $M^1_+(Y)$ by the map $(g, Q) \mapsto gQ$, $gQ(C) = Q(g^{-1}C)$ for every $C \in \mathscr{B}(Y)$, and one easily verifies that this map is measurable. A transition kernel δ from X to Y is said to be *equivariant* if $\delta(gx) = g\delta(x)$ for every $g \in G$, $x \in X$. We denote by $\mathscr{M}_G$ the set of all equivariant kernels from X to Y and by $\mathscr{N}_G$ the subset $\mathscr{M}_G \cap \mathscr{N}$. Note that if $\mathscr{B}(Y)$ is separated, then $\mathscr{N}_G$ is the set of all kernels $x \mapsto \varepsilon_{\varphi(x)}$ arising from equivariant measurable maps $\varphi: X \to Y$. A probability measure μ on $\mathscr{B}(X)$ is said to be *quasi-invariant* if $\{g\mu: g \in G\} \ll \mu$. We shall apply the results of Section 2 to the set $\mathscr{M}_G$.

THEOREM 4.1. *Let Y be a separable metrizable u.m. space, G a locally compact σ-compact group, and $\mathfrak{P} = \{\mu\}$ for some quasi-invariant probability measure μ. Further assume that G acts continuously on Y (by which it is meant that the induced maps on Y are continuous). Then $\mathscr{M}_G(\mu)$ has the integral representation property and is measure convex.*

Proof. According to Theorem 2.5, it suffices to show that the convex set $\mathscr{M}_G(\mu)$ is a τ-closed subset of $\mathscr{M}(\mu)$. We may assume $\mathscr{M}_G \neq \emptyset$. Since μ is quasi-invariant, G acts on $\mathscr{M}(\mu)$ by $g\delta$ = equivalence class of $g\delta'$ for some representant δ' of δ, where $(g\delta')(x) = g\delta'(g^{-1}x)$ for every $x \in X$. Further, G acts on $L^1(\mu)$ by $gf(x) = f(g^{-1}x)(dg\mu/d\mu)(x)$, $x \in X$, and on $C(Y)$ by $gk(y) = k(g^{-1}y)$, $y \in Y$. We have

$$f \otimes k(g\delta) = g^{-1} f \otimes g^{-1} k(\delta)$$

for every $\delta \in \mathcal{M}(\mu)$, $f \in L^1(\mu)$, $k \in C(Y)$. Thus the map $\delta \mapsto g\delta$ is τ-continuous for every $g \in G$. This implies that the set $\mathcal{M}(\mu)_G$ of all fixed points in $\mathcal{M}(\mu)$ under the action of G is τ-closed. Furthermore, $\mathcal{M}_G(\mu) = \mathcal{M}(\mu)_G$ holds (see Berk and Bickel [2]). This completes the proof.

By Corollary 2.6, the measure convexity of $\mathcal{M}_G(\mu)$ also holds for completely regular Souslin spaces Y and the integral representation property for completely regular Lusin spaces Y.

For statistical applications it is desirable that kernels in $\mathcal{M}_G$ have a representing measure which is supported by $\mathcal{N}_G$. This is not always possible. Let G_x be the isotropy group of $x \in X$ in G, $Y_K = \{y \in Y\colon\ gy = y$ for every $g \in K\}$ for $K \subset G$, and $Y_x = Y_{G_x}$. Then in order for $\delta \in \mathcal{M}_G$ to be represented by a measure in $M_+^1(\mathcal{N}_G)$ it is necessary that $\delta(x, Y_x) = 1$ P-almost everywhere for every $P \in \mathfrak{P}$, since $\varphi(x, Y_x) = 1$ for every $\varphi \in \mathcal{N}_G$, $x \in X$, provided $\{(x, y) \in X \times Y\colon\ y \in Y_x\} \in \big(\mathcal{B}(X) \otimes \mathcal{B}(Y)\big)_u$ and $\mathcal{B}(Y)$ is separated. To prove that this condition is also sufficient we need the following assumptions:

(A4.1) There is a measurable map $S\colon X \to G$ such that the (measurable) map $T\colon X \to X$, defined by $T(x) = S(x)^{-1}x$, is invariant, i.e. $T(gx) = T(x)$ for every $g \in G$, $x \in X$.

(A4.2) The action of the isotropy group G_x on Y is trivial for every $x \in X$.

(A4.3) $G_x = H$ for some subgroup H of G and every $x \in T(X)$.

Let $Z = T(X)$, $\mathcal{B}(Z) = Z \cap \mathcal{B}(X)$, and, if $Y_x \in \mathcal{B}(Y)_u$ for every $x \in X$,

$$\mathcal{D}_G = \{\delta \in \mathcal{M}_G\colon\ \delta(x, Y_x) = 1 \text{ for every } x \in X\}.$$

THEOREM 4.2. *Let* $(X, \mathcal{B}(X))$ *or* $(G, \mathcal{B}(G))$ *be u.m. and assume* (A4.1).

(a) *Let* $(Y, \mathcal{B}(Y))$ *be a Souslin measurable space and assume* (A4.2). *Then* $\mathcal{M}_G$ *has the integral representation property and* $\operatorname{ex} \mathcal{M}_G = \mathcal{N}_G$ *holds.*

(b) *Let* $(Y, \mathcal{B}(Y))$ *be a Souslin measurable space. Assume* (A4.3) *and* $Y_H \in \mathcal{B}(Y)$. *Then* $\mathcal{D}_G$ *has the integral representation property and* $\operatorname{ex} \mathcal{D}_G = \mathcal{N}_G$ *holds.*

(c) *Let* $(Y, \mathcal{B}(Y))$ *be u.m.,* $\mathcal{N}_G \neq \emptyset$, *and* $\mathfrak{P} = \{\mu\}$ *for some probability measure* μ. *Assume* $\{(t, y) \in Z \times Y\colon\ y \in Y_t\} \in \big(\mathcal{B}(Z) \otimes \mathcal{B}(Y)\big)_u$. *Then* $\mathcal{D}_G(\mu)$ *has the integral representation property and is measure convex. Furthermore,* $\operatorname{ex} \mathcal{D}_G(\mu) = \mathcal{N}_G(\mu)$ *holds.*

Proof. Assume $Y_x \in \mathcal{B}(Y)_u$ for every $x \in X$. Let $\mathcal{M}_1$ denote the set of all transition kernels from Z to Y and $\mathcal{N}_1$ the subset of all kernels $t \mapsto \varepsilon_{\psi(t)}$ arising from measurable maps $\psi\colon Z \to Y$. We put

$$\mathcal{D}_1 = \{\varphi \in \mathcal{M}_1\colon\ \varphi(t, Y_t) = 1 \text{ for every } t \in Z\}.$$

For each $\varphi \in \mathcal{D}_1$ the kernel $x \mapsto S(x)\varphi(Tx)$ belongs to $\mathcal{D}_G$. In fact, since $T/Z = id_Z$ and T is invariant, we have $S(gt)^{-1}g \in G_t$ for every $t \in Z$, $g \in G$.

Furthermore, $\varphi(t)$ is a G_t-invariant probability measure on $\mathscr{B}(Y)$ for every $t \in Z$. This yields $S(gx)\varphi(Tx) = S(gS(x)T(x))\varphi(Tx) = gS(x)\varphi(Tx)$ for every $x \in X$, $g \in G$, and from $Y_x = S(x)Y_{T(x)}$ follows

$$\varphi(T(x), S(x)^{-1}Y_x) = \varphi(T(x), Y_{T(x)}) = 1$$

for every $x \in X$. Thus we can define a map $i\colon \mathscr{D}_1 \to \mathscr{D}_G$ by $i(\varphi)(x) = S(x)\varphi(Tx)$ for every $x \in X$. Then i is an affine bijection with $i(\mathscr{D}_1 \cap \mathscr{N}_1) = \mathscr{N}_G$ and $i^{-1}(\delta) = \delta|Z$ for every $\delta \in \mathscr{D}_G$. Considering T as a map from X onto Z we obtain

$$(*)\qquad \begin{aligned} P\otimes i(\varphi)(B) &= \int_X\int_Y 1_B(x, y)\, i(\varphi)(x, dy)\, dP(x) \\ &= \int_X\int_Y 1_B(S(x)T(x), S(x)y)\,\varphi(T(x), dy)\, dP(x) \\ &= \int_{Z\times Y}\int_G 1_B(st, sy)\, Q_P(t, ds)\, dP^T\otimes\varphi(t, y) \end{aligned}$$

for every $\varphi \in \mathscr{D}_1$, $P \in M_+^1(X)$, and $B \in \mathscr{B}(X)\otimes\mathscr{B}(Y)$, where Q_P denotes the regular T-conditional distribution of S under P.

Subsets $\mathscr{D}$ of $\mathscr{M}_1$ are equipped with the σ-algebra $\Sigma(\mathscr{D})$ defined with respect to $\mathfrak{P}^T = \{\mu^T\colon \mu \in \mathfrak{P}\}$. We claim that i is an isomorphism between the measurable spaces $(\mathscr{D}_1, \Sigma(\mathscr{D}_1))$ and $(\mathscr{D}_G, \Sigma(\mathscr{D}_G))$. If P is a probability measure in $L(\mathfrak{P})$, then $P^T \in L(\mathfrak{P}^T)$. If Q is a probability measure in $L(\mathfrak{P}^T)$, then there is a countable subset $\{\mu_n\colon n \in I\}$ of $\mathfrak{P}$, $I \subset N$, such that $Q \ll \{\mu_n^T\colon n \in I\}$ ([14], Lemme 1). Let $\lambda = \sum 2^{-n}\mu_n$, $n \in I$. Then by the Radon–Nikodym theorem, $Q = f\cdot\lambda^T$ for some $f \in L^1(\lambda^T)$ and for the measure $P = f\circ T\cdot\lambda$ on $\mathscr{B}(X)$ we obtain $P^T = Q$ and $P \in L(\mathfrak{P})$. Thus

$$(**)\qquad M_+^1(Z)\cap L(\mathfrak{P}^T) = \{P^T\colon P \in M_+^1(X)\cap L(\mathfrak{P})\}.$$

Therefore, i is measurable, since

$$P(i(\varphi))(C) = \int_{Z\times Y}\int_G 1_C(sy)\, Q_P(t, ds)\, dP^T\otimes\varphi(t, y),$$

$$(t, y) \mapsto \int_G 1_C(sy)\, Q_P(t, ds) \in B(Z\times Y)$$

for every $\varphi \in \mathscr{D}_1$, $P \in M_+^1(X)\cap L(\mathfrak{P})$, $C \in \mathscr{B}(Y)$, and i^{-1} is measurable, since

$$P^T(i^{-1}(\delta))(C) = \int_{X\times Y} 1_C(S(x)^{-1}y)\, dP\otimes\delta(x, y),$$

$(x, y) \mapsto 1_C(S(x)^{-1}y) \in B(X\times Y)$ for every $\delta \in \mathscr{D}_G$, $P \in M_+^1(X)\cap L(\mathfrak{P})$, $C \in \mathscr{B}(Y)$. Clearly, the restriction $i_0\colon \mathscr{D}_1 \cap \mathscr{N}_1 \to \mathscr{N}_G$ of the map i to $\mathscr{D}_1 \cap \mathscr{N}_1$ is also an isomorphism between the measurable spaces $(\mathscr{D}_1 \cap \mathscr{N}_1, \Sigma(\mathscr{D}_1 \cap \mathscr{N}_1))$ and $(\mathscr{N}_G, \Sigma(\mathscr{N}_G))$.

Assume that $\mathscr{B}(Y)$ is countably generated and contains singletons. Since $\mathscr{D}_1$ is an extremal subset of $\mathscr{M}_1$ and $\operatorname{ex}\mathscr{M}_1 = \mathscr{N}_1$, we obtain $\operatorname{ex}\mathscr{D}_1 = \mathscr{D}_1 \cap \mathscr{N}_1$. From $i(\operatorname{ex}\mathscr{D}_1) = \operatorname{ex} i(\mathscr{D}_1)$ follows $\operatorname{ex}\mathscr{D}_G = \mathscr{N}_G$.

(a) By (A4.2), $Y_x = Y$ for every $x \in X$. Hence $\mathscr{D}_1 = \mathscr{M}_1$ and $\mathscr{D}_G = \mathscr{M}_G$ hold. Let $\delta \in \mathscr{M}_G$. According to Proposition 2.1, there is a measure $\varrho_1 \in M_+^1(\mathscr{N}_1)$ which represents the kernel $i^{-1}(\delta)$ with respect to $\mathfrak{P}^T$. Then, by (*) and (**), the image measure $\varrho \in M_+^1(\mathscr{N}_G)$ of ϱ_1 under i_0 represents δ.

(b) By (A4.3) we have $Y_x = S(x) Y_H$ and hence $Y_x \in \mathscr{B}(Y)$ for every $x \in X$. Since $\mathscr{D}_1 = \{\varphi \in \mathscr{M}_1\colon \varphi(t, Y_H) = 1 \text{ for every } t \in Z\}$, we can identify $\mathscr{D}_1$ with the set of all kernels from Z to Y_H, where Y_H is equipped with the σ-algebra $\mathscr{B}(Y_H) = Y_H \cap \mathscr{B}(Y)$. Since $(Y_H, \mathscr{B}(Y_H))$ is a Souslin measurable space ([21], p. 96, Theorem 3), the integral representation property of $\mathscr{D}_G$ follows from Proposition 2.1 as in (a).

(c) In view of Lemma 2.4(b) we have $Y_t \in \mathscr{B}(Y)_u$ for every $t \in Z$ and hence $Y_x \in \mathscr{B}(Y)_u$ for every $x \in X$. Define $j\colon \mathscr{D}_1(\mu^T) \to \mathscr{D}_G(\mu)$ by $j(\varphi) = \mu$-equivalence class of $i(\varphi')$ for some representant $\varphi' \in \mathscr{D}_1$ of φ. Then j is affine and an isomorphism between the measurable spaces $(\mathscr{D}_1(\mu^T), \Sigma(\mu^T))$ and $(\mathscr{D}_G(\mu), \Sigma(\mathscr{D}_G(\mu))$ with $j(\mathscr{D}_1 \cap \mathscr{N}_1(\mu^T)) = \mathscr{N}_G(\mu)$. Define a multifunction Γ from Z to $M_+^1(Y)$ by $\Gamma(t) = \{Q \in M_+^1(Y)\colon Q(Y_t) = 1\}$. Since $\mathscr{D}_1 \cap \mathscr{N}_1 \neq \emptyset$, we obtain $\mathscr{D}_1(\mu^T) = \mathscr{M}_{1\Gamma}(\mu^T)$ and $\mathscr{D}_1 \cap \mathscr{N}_1(\mu^T) = \mathscr{N}_{1\Gamma}(\mu^T)$. By (*) and (**), the assertion now follows from Corollary 2.8.

The following corollary is an immediate consequence of part (c).

COROLLARY 4.3. *Let $(X, \mathscr{B}(X))$ or $(G, \mathscr{B}(G))$ be u.m., let $(Y, \mathscr{B}(Y))$ be u.m., and $\mathfrak{P} = \{\mu\}$ for some probability measure μ. Assume* (A4.1) *and* (A4.2). *Then $\mathscr{M}_G(\mu)$ has the integral representation property, is measure convex, and $\operatorname{ex}\mathscr{M}_G(\mu) = \mathscr{N}_G(\mu)$ holds.*

Remarks. (1) The measurability condition in Theorem 4.2(c) is satisfied if $\mathscr{B}(X)$ is countably generated and contains singletons and $(G, \mathscr{B}(G))$ is a Souslin measurable space.

To see this put $B_1 = \{(g, t, y) \in G \times Z \times Y\colon gt = t\}$ and $B_2 = \{(g, t, y) \in G \times Z \times Y\colon gy = y\}$. Then we have $B_1, B_2 \in \mathscr{B}(G) \otimes \mathscr{B}(Z) \otimes \mathscr{B}(Y)$ ([5], I.12) and $\{(t, y) \in Z \times Y\colon y \in Y_t\}^c$ is the projection of $B_1 \cap B_2^c$ to $Z \times Y$. From a projection theorem ([3], III.23) it immediately follows that $\{(t, y) \in Z \times Y\colon y \in Y_t\} \in (\mathscr{B}(Z) \otimes \mathscr{B}(Y))_u$.

(2) In Theorem 4.2(a) the assumption (A4.2) is essential for $\operatorname{ex}\mathscr{M}_G = \mathscr{N}_G$ to hold. Let $X = Y = \{-1, 0, 1\}$ and $G = \{e, g\}$ with identity e, $g^{-1} = g$, and the action $gx = -x$, $gy = -y$. Define $S\colon X \to G$ by $S(-1) = g$ and $S(0) = S(1) = e$. Then (A4.1) is satisfied, but (A4.2) does not hold. We obtain $\mathscr{D}_G = \{\delta \in \mathscr{M}_G\colon \delta(0, \{0\}) = 1\}$, $\operatorname{ex}\mathscr{D}_G = \mathscr{N}_G$, and $\operatorname{ex}\mathscr{M}_G = \mathscr{N}_G \cup \{\delta_1, \delta_2, \delta_3\}$, where $\delta_1(x) = \varepsilon_x$, $\delta_2(x) = \varepsilon_{-x}$, $\delta_3(x) = \varepsilon_0$ for $x = \pm 1$ and $\delta_i(0) = (\varepsilon_1 + \varepsilon_{-1})/2$ for $i = 1, 2, 3$.

(3) For Polish groups G and closed subgroups H of G the assumptions (A4.1) and (A4.3) are equivalent to a product representation $X = G/H \times Z$. To see this observe that, if the space G/H of left cosets of H is equipped with the quotient topology, the canonical surjection π: $G \to G/H$ admits a Borel measurable right inverse ψ. This follows from a selection theorem of Kuratowski and Ryll-Nardzewski [13]. Let $X = G/H \times Z$ and $\mathscr{B}(X) = \mathscr{B}(G/H) \otimes \mathscr{B}(Z)$ for an arbitrary measurable space $(Z, \mathscr{B}(Z))$. G acts on X by $(g, (g'H, z)) \mapsto (gg'H, z)$ and this map is measurable. Define S: $X \to G$ by $S((gH, z)) = \psi(gH)$. Then (A4.1) and (A4.3) are satisfied. Conversely, if these conditions hold, put $Z = T(X)$ and $\mathscr{B}(Z) = Z \cap \mathscr{B}(X)$. Then the map $X \to G/H \times Z$, $x \mapsto (\pi \circ S(x), T(x))$ is an equivariant isomorphism between the measurable spaces $(X, \mathscr{B}(X))$ and $(G/H \times Z, \mathscr{B}(G/H) \otimes \mathscr{B}(Z))$.

5. Applications. At first we treat transition kernels with prescriped values. Let Y be a metrizable Souslin space and $\mu \in M_+^1(X)$. Let Γ be a multifunction from X to $M_+^1(Y)$ such that $\Gamma(x)$ is nonempty w-closed and convex for every $x \in X$ and $\mathrm{Gr}(\Gamma) \in \mathscr{B}(X)_u \otimes \Sigma(M_+^1(Y))$. We denote by $\mathrm{ex}\,\Gamma$ the multifunction $x \mapsto \mathrm{ex}\,\Gamma(x)$.

THEOREM 5.1. *$\mathscr{M}_\Gamma(\mu)$ is a nonempty τ-closed measure convex subset of $\mathscr{M}(\mu)$ and has the integral representation property with respect to $\mathfrak{P} = \{\mu\}$. Furthermore,* $\mathrm{ex}\,\mathscr{M}_\Gamma(\mu) = \mathscr{M}_{\mathrm{ex}\,\Gamma}(\mu)$ *holds.*

Proof. Since $(M_+^1(Y), w)$ is a Souslin space and $\Sigma(M_+^1(Y)) = \mathscr{B}(M_+^1(Y), w)$ holds, it follows from a selection theorem ([3], III.22) that $\mathscr{M}_\Gamma \neq \emptyset$. Clearly, $\mathscr{M}_\Gamma(\mu)$ is convex. The assertion $\mathrm{ex}\,\mathscr{M}_\Gamma(\mu) = \mathscr{M}_{\mathrm{ex}\,\Gamma}(\mu)$ follows from Theorem IV.15 of [3]. According to Theorem 2.5, it remains to show that $\mathscr{M}_\Gamma(\mu)$ is a τ-closed subset of $\mathscr{M}(\mu)$. (For constant multifunctions this has been proved in [9].) Let $\delta \in \mathscr{M}(\mu) \setminus \mathscr{M}_\Gamma(\mu)$ and put

$$A = \{x \in X:\ \delta(x) \notin \Gamma(x)\}.$$

Then A^c is the projection of $\mathrm{Gr}(\delta) \cap \mathrm{Gr}(\Gamma) \in \mathscr{B}(X)_u \otimes \Sigma(M_+^1(Y))$ to X, hence, by a projection theorem ([3], III.23), $A \in \mathscr{B}(X)_u$. Choose a totally bounded metric inducing the topology of Y and denote by $U(Y)$ the space of all bounded uniformly continuous real-valued functions on Y. Using Portmanteau's theorem we conclude that $\Gamma(x)$ is a $\sigma(M(Y), U(Y))$-closed subset of $M(Y)$ for every $x \in X$. Therefore, by the Hahn-Banach theorem, for each $x \in A$ there is a function $k_x \in U(Y)$ such that

$$\sup\{\int_Y k_x\,dQ:\ Q \in \Gamma(x)\} < \int_Y k_x\,d\delta(x).$$

We can assume $k_x \in V$ for some countable norm dense subset V of $U(Y)$. Setting

$$A(n, k) = \{x \in X:\ \sup\{\int_Y k\,dQ:\ Q \in \Gamma(x)\} \leqslant \int_Y k\,d\delta(x) - 1/n\}$$

for $n \in N$ and $k \in V$ yields

$$A = \bigcup_{\substack{n \in N \\ k \in V}} A(n, k).$$

The function

$$x \mapsto \sup\{\int_Y k dQ\colon Q \in \Gamma(x)\}$$

is universally measurable ([3], III.39) and hence $A(n, k) \in \mathscr{B}(X)_u$ for every $n \in N$ and $k \in V$. Because $\mu(A) > 0$, there exist $n \in N$ and $k \in V$ such that $\mu(A(n, k)) > 0$. Now for each member φ of the τ-closure $\mathscr{M}_\Gamma(\mu)^-$ of $\mathscr{M}_\Gamma(\mu)$ in $\mathscr{M}(\mu)$ we have

$$\int_{A(n,k)} \int_Y k(y) \varphi(x, dy) d\mu(x) \leqslant \int_{A(n,k)} \int_Y k(y) \delta(x, dy) d\mu(x) - \mu(A(n, k))/n$$

and thus $\delta \notin \mathscr{M}_\Gamma(\mu)^-$.

We can infer the following version for metrizable Souslin spaces of a result due to Strassen ([22], Theorem 3). Strassen proved the equivalence of (i) and (ii) for Polish spaces Y (under a slightly weaker measurability assumption on Γ).

COROLLARY 5.2. *Let $\lambda \in M_+^1(Y)$. Then the following statements are equivalent:*

(i) *There is a kernel $\delta \in \mathscr{M}_\Gamma$ such that $\lambda = \mu(\delta)$.*

(ii) $\int_Y k d\lambda \leqslant \int_X \sup\{\int_Y k dQ\colon Q \in \Gamma(x)\} d\mu(x)$ *for every* $k \in C(Y)$.

(iii) $\int_Y k d\lambda \leqslant \int_X \sup\{\int_Y k dQ\colon Q \in \operatorname{ex} \Gamma(x)\} d\mu(x)$ *for every* $k \in C(Y)$.

Proof. (i) $\Rightarrow$ (iii) follows from the integral representation property of $\mathscr{M}_\Gamma(\mu)$ and (iii) $\Rightarrow$ (ii) is obvious.

(ii) $\Rightarrow$ (i). The set $K = \{\mu(\delta)\colon \delta \in \mathscr{M}_\Gamma(\mu)\}$ is convex. We claim that K is a w-closed subset of $M(Y)$. Suppose (δ_α) is a net in $\mathscr{M}_\Gamma(\mu)$ such that $\mu(\delta_\alpha) \to Q$ for some $Q \in M(Y)$. Then the set $\{\delta_\alpha\}$ is relatively τ-compact in $\mathscr{M}(\mu)$. This fact is known and easily seen by embedding $\mathscr{M}(\mu)$ in $\mathscr{M}^Z(\mu)$ for some compact metrizable space Z. Therefore (δ_α) has a cluster point δ which belongs to $\mathscr{M}_\Gamma(\mu)$ since, by the preceding theorem, $\mathscr{M}_\Gamma(\mu)$ is τ-closed in $\mathscr{M}(\mu)$. Clearly $Q = \mu(\delta)$. Thus $Q \in K$ and our claim is proved. Now by the Hahn-Banach theorem, $\lambda \in K$ if

$$\int_Y k d\lambda \leqslant \sup\{\int_Y k d\mu(\delta)\colon \delta \in \mathscr{M}_\Gamma(\mu)\}$$

for every $k \in C(Y)$. But the right-hand side of this inequality equals the right-hand side of (ii). Indeed, let $k \in C(Y)$ and $\varepsilon > 0$ and consider the multifunction Γ_ε from X to $M_+^1(Y)$ defined by

$$\Gamma_\varepsilon(x) = \{P \in \Gamma(x)\colon \int_Y k dP \geqslant \sup\{\int_Y k dQ\colon Q \in \Gamma(x)\} - \varepsilon\}.$$

Then $\Gamma_\varepsilon(x) \neq \emptyset$ for every $x \in X$ and $\mathrm{Gr}(\Gamma_\varepsilon) \in \mathscr{B}(X)_u \otimes \Sigma(M_+^1(Y))$. Hence, according to a selection theorem ([3], III.22), $M_{\Gamma_\varepsilon} \neq \emptyset$.

The second application is concerned with the risk-equivalence of two methods of randomization in statistical decision theory. Let Θ be an index set. For each $\vartheta \in \Theta$, let P_ϑ be a probability measure on $\mathscr{B}(X)$ and $L(\vartheta, \cdot, \cdot)\colon X \times Y \to [0, \infty]$ a measurable loss function. Decision rules are transition kernels from X to Y. In the decision problem $(\{P_\vartheta\colon \vartheta \in \Theta\}, Y, L)$ the risk function R, defined for $\vartheta \in \Theta$ and $\delta \in \mathscr{M}$, is given by

$$R(\vartheta, \delta) = \int\limits_{X \times Y} L(\vartheta, x, y)\, dP_\vartheta \otimes \delta(x, y).$$

Assume that $\{P_\vartheta\colon \vartheta \in \Theta\} \ll \mu$ for some $\mu \in M_+^1(X)$ and $\mathfrak{P} = \{\mu\}$. Let $\mathscr{D} \subset \mathscr{M}(\mu)$ and $\mathscr{D}_1 \subset \mathscr{D}$. $\mathscr{D}$ and $M_+^1(\mathscr{D}_1)$ are said to be *risk-equivalent* if for each $\delta \in \mathscr{D}$ there is a measure $\varrho \in M_+^1(\mathscr{D}_1)$ such that

$$R(\vartheta, \delta) = \int\limits_{\mathscr{D}_1} R(\vartheta, \varphi)\, d\varrho(\varphi)$$

for every $\vartheta \in \Theta$ and vice versa. Then, under the assumption of Corollary 2.7, $\mathscr{M}(\mu)$ and $M_+^1(\mathscr{N}(\mu))$ are risk-equivalent. If a group G acts on X and Y, then, under the assumptions of Corollary 4.3, $\mathscr{M}_G(\mu)$ and $M_+^1(\mathscr{N}_G(\mu))$ are risk-equivalent.

In order to treat Bayes rules, we assume additionally that Θ is equipped with a σ-algebra $\mathscr{B}(\Theta)$ such that L and $(\vartheta, x) \mapsto (dP_\vartheta/d\mu)(x)$ are measurable. We denote by $\mathscr{D}$ the set of all Bayes rules with respect to a σ-finite measure λ on $\mathscr{B}(\Theta)$. Let Y be a metrizable Souslin space and let $L(\vartheta, x, \cdot)$ be lower semicontinuous for every $\vartheta \in \Theta$ and $x \in X$.

PROPOSITION 5.3. *$\mathscr{D}(\mu)$ and $M_+^1(\mathscr{D} \cap \mathscr{N}(\mu))$ are risk-equivalent.*

Proof. According to Lemma 2 of [15], the Bayes risk function

$$\int\limits_{\Theta} R(\vartheta, \cdot)\, d\lambda(\vartheta)\colon\ \mathscr{M}(\mu) \to [0, \infty]$$

is lower τ-semicontinuous. This implies that $\mathscr{D}(\mu)$ is a τ-closed subset of $\mathscr{M}(\mu)$. Furthermore, $\mathscr{D}(\mu)$ is a convex extremal subset of $\mathscr{M}(\mu)$, hence, $\mathrm{ex}\,\mathscr{D}(\mu) = \mathscr{D} \cap \mathscr{N}(\mu)$. The assertion now follows from Theorem 2.5.

REFERENCES

[1] E. J. Balder, *Mathematical foundations of statistical decision theory: the modern viewpoint*, Preprint Nr. 199, Department of Mathematics, University of Utrecht, 1981.

[2] R. H. Berk and P. J. Bickel, *On invariance and almost invariance*, Ann. Math. Statist. 39 (1968), p. 1573–1576.

[3] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Math. 580, Springer, Berlin 1977.

[4] J. P. R. Christensen, *Topology and Borel Structure*, North-Holland, Amsterdam 1974.

[5] C. Dellacherie and P.-A. Meyer, *Probabilities and Potential*, North-Holland, Amsterdam 1978.

[6] R. H. Farrell, *Weak limits of sequences of Bayes procedures in estimation theory*, Proc. Fifth Berkeley Symp. Math. Statist. Probab. Univ. of Calif. 1 (1967), p. 83–111.

[7] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York 1967.

[8] J. Hoffmann-Jorgensen, *The Theory of Analytic Spaces*, Various Publ. Series Nr. 10, Matematisk Institut, Aarhus Universitet, 1970.

[9] A. Irle, *On the compactness of vector-valued transition measures*, Trans. of the Eighth Prague Conf. on Information Theory, Statist. Dec. Functions, Random Processes, Vol. A (1978), p. 325–334.

[10] J. E. Jayne and C. A. Rogers, *The extremal structure of convex sets*, J. Funct. Anal. 26 (1977), p. 251–288.

[11] H.-P. Kirschner, *On the risk-equivalence of two methods of randomization in statistics*, J. Multivariate Anal. 6 (1976), p. 159–166.

[12] – *On a theorem of Wald and Wolfowitz on randomization in statistics*, Manuscripta Math. 18 (1976), p. 249–256.

[13] K. Kuratowski and C. Ryll-Nardzewski, *A general theorem on selectors*, Bull. Acad. Polon. Sci., Sér. Sci. Math. Astronom. Phys., 13 (1965), p. 397–403.

[14] H. Luschgy, *Sur l'existence d'une plus petite sous-tribu exhaustive par paire*, Ann. Inst. H. Poincaré, Sect. B, 14 (1978), p. 391–398.

[15] – *Elimination of randomization and Hunt-Stein type theorems in invariant statistical decision problems*, Statistics 18 (1987), p. 99–111.

[16] H. Luschgy and D. Mussmann, *A characterization of weakly dominated statistical experiments by compactness of the set of decision rules*, to appear in Sankhyā, Series A, 49 (1987).

[17] P.-A. Meyer, *Probabilités et Potentiel*, Hermann, Paris 1966.

[18] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York 1967.

[19] R. R. Phelps, *Lectures on Choquet's Theorem*, Van Nostrand, New York 1966.

[20] S. Portnoy, *On fundamental theorems in decision theory*, Department of Statistics, Havard University, 1972.

[21] L. Schwartz, *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*, Oxford University Press, London 1973.

[22] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. 36 (1965), p. 423–439.

[23] A. Wald and J. Wolfowitz, *Two methods of randomization in statistics and the theory of games*, Ann. of Math. 53 (1951), p. 581–586.

[24] H. v. Weizsäcker and G. Winkler, *Integral representation in the set of solutions of a generalized moment problem*, Math. Ann. 246 (1979), p. 23–32.

[25] G. A. Edgar, A. Millet and L. Sucheston, *On compactness and optimality of stopping times*, Lecture Notes in Math. 939 (*Martingale Theory in Harmonic Analysis and Banach Spaces*), p. 36–61, Springer, Berlin 1982.

[26] E. A. Fainberg, *Non-randomized Markov and semi-Markov strategies in dynamic programming*, Theory Probab. Appl. 27 (1982), p. 116–126.

[27] N. Ghoussoub, *An integral representation of randomized probabilities and its applications*, Lecture Notes in Math. 920 (Sém. Prob. XVI), p. 519–543, Springer, Berlin 1982.

[28] D. Kadelka, *On randomized policies and mixtures of deterministic policies in dynamic programming*, Methods of Oper. Research 46 (1982), p. 67–75.

Villa Gröning
Feuerstiege 13
D-4440 Rheine 11, FRG

Received on 26. 3. 1986