THE EXISTENCE OF PROBABILITY MEASURES WITH GIVEN MARGINALS[1]

By V. Strassen

Göttingen University

1. Summary. First an integral representation of a continuous linear functional dominated by a support function in integral form is given (Theorem 1). From this the theorem of Blackwell-Stein-Sherman-Cartier [2], [20], [4], is deduced as well as a result on capacities alternating of order 2 in the sense of Choquet [5], which includes Satz 4.3 of [23] and a result of Kellerer [10], [12], under somewhat stronger assumptions. Next (Theorem 7), the existence of probability distributions with given marginals is studied under typically weaker assumptions, than those which are required by the use of Theorem 1. As applications we derive necessary and sufficient conditions for a sequence of probability measures to be the sequence of distributions of a martingale (Theorem 8), an upper semimartingale (Theorem 9) or of partial sums of independent random variables (Theorem 10). Moreover an alternative definition of Lévy-Prokhorov's distance between probability measures in a complete separable metric space is obtained (corollary of Theorem 11). Section 6 can be read independently of the former sections.

2. Acknowledgments. This work started from a conversation with Paul André Meyer about the similarity of the Blackwell-Stein-Sherman-Cartier theorem and Satz 4.3 [23]. We both found independently generalizations of the two results, Professor Meyer earlier than I. His result is included here with his kind permission as the corollary of Theorem 10. My knowledge of the relevant literature has been considerable improved by contact with Professors W. J. Bade, K. Ito, J. L. Kelley, H. Kellerer and L. Le Cam, whom I thank very much. I am indebted to H. Kellerer for a critical remark.

3. Notation. If $(\Omega, \mathcal{B})$ and (R, ϑ) are measurable spaces, then a Markov kernel P from Ω to R is a real function on $\vartheta \times \Omega$ such that for any $\omega \varepsilon \Omega$, $P(\cdot, \omega)$ is a probability measure on ϑ and for any $E \varepsilon \vartheta$, $P(E, \cdot)$ is $\mathcal{B}$-measurable.

If μ is a probability measure on $\mathcal{B}$, $P\mu$ is a probability measure on ϑ defined by

$$(1) \qquad (P\mu)(E) = \int P(E, \omega)\mu(d\omega)$$

for $E \varepsilon \vartheta$. $P \times \mu$ is a probability measure on $\vartheta \times \mathcal{B}$ defined by

$$(2) \qquad (P \times \mu)(E \times F) = \int_F P(E, \omega)\mu(d\omega)$$

for $E \varepsilon \vartheta$, $F \varepsilon \mathcal{B}$.

Received 17 August 1964.

[1] This paper was prepared with the partial support of the Office of Naval Research, Contract Nonr-222-43. This paper in whole or in part may be reproduced for any purpose of the United States Government.

If z is a bounded ϑ-measurable function on R, zP is a bounded $\mathcal{B}$-measurable function on Ω defined by

(3) $$(zP)(\omega) = \int z(r)P(dr, \omega).$$

4. An application of the Hahn-Banach theorem. Let X be a real linear space, elements of which we denote by x, y. A real function h on X is called a support function if and only if it is subadditive, i.e.,

$$h(x + y) \leqq h(x) + h(y)$$

and nonnegative homogeneous, i.e.,

$$h(ax) = ah(x)$$

for any $a \geqq 0$.

We list two well-known properties of support functions h.

(4) Hahn-Banach theorem (see, e.g., [15], p. 21):
If X_0 is a linear submanifold of X and l_0 a linear functional on X_0 satisfying $l_0(x) \leqq h(x)$ for all $x \in X_0$, then there is a linear functional l on X which extends l_0 and satisfies $l(x) \leqq h(x)$ for all $x \in X$.
Let X be a Banach space and X^* be its dual (we denote elements of X^* by x^*, y^*). Then h is norm continuous if and only if it is norm bounded, i.e., if and only if

$$\|h\| =_{\text{def}} \sup\{|h(x)|: \|x\| \leqq 1\} < \infty.$$

(5) Moreover, for any continuous support function h let $\varphi(h)$ be the set of all linear functionals on X dominated by h (these are then automatically continuous), i.e.,

$$\varphi(h) = \{x^*: x^* \in X^* \quad \text{and} \quad x^* \leqq h\}.$$

Then $h = \sup \varphi(h)$. φ maps the set of continuous support functions one-to-one onto the set of nonempty, convex and weak* compact subsets of X^*.

Now let X be a separable Banach space, $(\Omega, \mathcal{B}, \mu)$ a probability space. We denote elements of Ω by ω, η. Let $\omega \rightarrow h_\omega$ be a map from Ω into the set of continuous support functions on X, which is weakly measurable in the sense that for every x the map $\omega \rightarrow h_\omega(x)$ is $\mathcal{B}$-measurable. Then also $\omega \rightarrow \|h_\omega\|$ is $\mathcal{B}$-measurable (because X is separable). Assume

(6) $$\int \|h_\omega\| \, \mu(d\omega) < \infty.$$

The integral

$$h(x) = \int h_\omega(x)\mu(d\omega)$$

defines a continuous support function on X.

THEOREM 1. *If $x^* \in X^*$, then the following are equivalent:*

(i) *x^* is dominated by h*, i.e., $x^* \in \varphi(h)$,

(ii) *there is a map* $\omega \to x_\omega^*$ *from* Ω *into* X^* *which is weakly measurable in the sense that for all* x *the map*

$$\omega \to x_\omega^*(x) \tag{7}$$

is measurable, which satisfies

$$x_\omega^*(x) \leqq h_\omega(x) \tag{8}$$

for all x *and* ω (i.e., $x_\omega^* \varepsilon \varphi(h_\omega)$ *for all* ω) *and*

$$x^*(x) = \int x_\omega^*(x)\mu(d\omega) \tag{9}$$

for all x.

PROOF. Let $\mathfrak{L}$ be the linear space of measurable maps ξ from Ω into X which assume only finitely many values, ξ and ξ' being considered equal if they agree μ-almost surely. Put $H(\xi) = \int h_\omega(\xi(\omega))\mu(d\omega)$. H is a support function on $\mathfrak{L}$.

Now assume (i). On the linear subspace $\tilde{\mathfrak{L}}$ of $\mathfrak{L}$, which consists of all μ-almost everywhere constant maps $\tilde{\xi}$ from Ω into X, the linear functional $\tilde{Q}$ defined by $\tilde{Q}(\tilde{\xi}) = x^*(x)$, where x is the μ-sure value of $\tilde{\xi}$, is dominated by H:

$$\tilde{Q}(\tilde{\xi}) = x^*(x) \leqq h(x) = \int h_\omega(x)\mu(d\omega) = \int h_\omega(\tilde{\xi}(\omega))\mu(d\omega) = H(\tilde{\xi}).$$

By the Hahn-Banach theorem there is a linear functional Q on $\mathfrak{L}$ dominated by H such that for all $\tilde{\xi} \varepsilon \tilde{\mathfrak{L}}$, $Q(\tilde{\xi}) = \tilde{Q}(\tilde{\xi})$. Now let $x \varepsilon X$, $\Lambda \varepsilon \mathcal{B}$ and I_Λ be the indicator of Λ. Then

$$Q(xI_\Lambda) \leqq H(xI_\Lambda) = \int_\Lambda h_\omega(x)\mu(d\omega) \tag{10}$$

and

$$-Q(xI_\Lambda) = Q(-xI_\Lambda) \leqq \int_\Lambda h_\omega(-x)\mu(d\omega), \tag{11}$$

so that $\Lambda \to Q(xI_\Lambda)$ is a finite signed measure μ_x such that $\mu_x \ll \mu$ with a density $\omega \to q_\omega(x)$ bounded from above by $\omega \to h_\omega(x)$ μ-almost everywhere. Because Q is a linear functional, we have for any x, $y \varepsilon X$ and a, b real

$$q_\omega(ax + by) = aq_\omega(x) + bq_\omega(y), \qquad (\mu\text{-almost all } \omega),$$

so that if we exclude a μ-null set Λ_0, q_ω may be assumed rational linear and dominated by h_ω on some countable dense subset X_0 of X closed under rational linear combinations. Without loss of generality $\Lambda_0 = \varnothing$, because there is a map $\omega \to x_\omega^*$ from Λ_0 to X^* such that (7) and (8) hold (this can be seen by adapting K. Jacobs' construction ([23], p. 295) to this case: one replaces the sets A^k there by sets

$$A^{x,r} = \{x^*:x^*(x) \leqq r\}$$

for rational r, and x from a countable dense subset of X. If one is willing to assume $(\Omega, \mathcal{B}, \mu)$ to be complete, the problem becomes of course trivial).

Let x_ω^* be the unique continuous extension to X of the restriction to X_0 of

q_ω. Then $x_\omega^* \varepsilon X^*$ and

$$x_\omega^*(x) \leqq h_\omega(x) \tag{12}$$

for all x. Moreover for any x, $\omega \to x_\omega^*(x)$ is $\mathcal{B}$-measurable (as a pointwise limit of a sequence of measurable functions). Also, because for any sequence $(x_n)_{n \geqq 1}$ of elements in X norm-converging to x

$$\lim_n \sup_{\Lambda \varepsilon \mathcal{B}} |\mu_{x_n}(\Lambda) - \mu_x(\Lambda)| = 0$$

(see (10), (11)) and therefore

$$\lim_n \int |q_\omega(x_n) - q_\omega(x)| \, \mu(d\omega) = 0,$$

we have for any x

$$x_\omega^*(x) = q_\omega(x), \qquad (\mu\text{-almost all } \omega).$$

So $\omega \to x_\omega^*(x)$ is a density of μ_x with respect to μ. We conclude

$$x^*(x) = \tilde{Q}(\tilde{\xi}) = Q(\tilde{\xi}) = Q(xI_\Omega) = \mu_x(\Omega) = \int x_\omega^*(x)\mu(d\omega),$$

which together with (12) proves (ii). We remark that the theorem follows very easily from (5) when Ω is finite. After having finished this paper I learned from Hans Kellerer that he has already proved a result implying Theorem 1 in the case $X = R^n$ in [ ] with a different method.

5. Convex sets of probability measures. Let Ω be a convex compact metrizable subset of a locally convex topological vector space. For Borel probability measures μ and ν write $\mu \prec \nu$ if and only if for all $y \varepsilon S = \{$all continuous concave functions on $\Omega\}$

$$\int y \, d\mu \geqq \int y \, d\nu. \tag{13}$$

A dilatation P is a Markov kernel from Ω to Ω such that for all continuous affine functions z on Ω, $zP = z$ (see (3)).

THEOREM 2 (Hardy-Littlewood-Pólya-Blackwell-Stein-Sherman-Cartier-Fell-Meyer [2], [20], [4]). *$\mu \prec \nu$ if and only if there is a dilatation P such that $\nu = P\mu$ (see (1)).*

PROOF. The "if" part follows from Jensen's inequality (see [7]). We apply Theorem 1 to prove the "only if" part, putting $X = C(\Omega) = \{$all continuous real functions on $\Omega\}$, $x^* = \nu$ (by Riesz's theorem we identify here and in the following certain linear functionals with probability measures) and

$$h_\omega(x) = \inf\{y(\omega) : y \varepsilon S \quad \text{and} \quad y \geqq x\}.$$

One easily verifies the assumptions made before Theorem 1. Moreover, (i) is satisfied: From the separability of $\{y : y \varepsilon S$ and $y \geqq x\}$ it follows that $\omega \to h_\omega(x)$ is pointwise limit of a nonincreasing sequence of functions in S, so that from (13)

$$x^*(x) = \int x \, d\nu \leqq \int h_\omega(x)\nu(d\omega) \leqq \int h_\omega(x)\mu(d\omega).$$

(ii) of Theorem 1 provides us with a function $\omega \to x_\omega^*$. Because

$$h_\omega(x) \leqq \sup \{x(\omega) : \omega \varepsilon \Omega\}$$

and the right side is just the support function defining the set of probability measures, x_ω^* is a probability measure for each ω. Because of (7), $P(\cdot, \omega) = x_\omega^*$ defines a Markov kernel. (9) yields $\nu = P\mu$. From (8) we get for $z \varepsilon S$

$$(zP)(\omega) = x_\omega^*(z) \leqq h_\omega(z) = z(\omega),$$

so for continuous affine z, $zP = z$, which completes the proof.

For the application of the above theorem to the comparison of experiments see Blackwell [1]. It would be interesting to know whether the more general stability result of Le Cam [18] (Theorem 2 of that paper) can be obtained by our methods.

Now let R be a Polish space, i.e., a Hausdorff space metrizable by a complete separable distance (see [3]). Denote by $C(R)$ the Banach space of bounded continuous real functions on R. Denote by $\mathfrak{R}$ the set of nonempty convex sets of (Borel-) probability measures in R which are closed with respect to the weak* topology (generated by the functionals of the form $\nu \to \int z d\nu$ for $z \varepsilon C(R)$).

THEOREM 3. *Let $\omega \to K_\omega$ be a map from Ω into $\mathfrak{R}$ such that for any $z \varepsilon C(R)$*

$$\omega \to h_\omega(z) = \sup \{\textstyle\int z \, d\alpha : \alpha \varepsilon K_\omega\}$$

is $\mathfrak{L}$-measurable. Let ν be a probability measure in R. In order that there exist a Markov kernel P from Ω to R such that

$$\nu = P\mu \tag{14}$$

and

$$P(\cdot, \omega) \varepsilon K_\omega \tag{15}$$

for μ-almost all ω, it is necessary and sufficient that

$$\textstyle\int z \, d\nu \leqq \int h_\omega(z) \mu(d\omega) \tag{16}$$

for all $z \varepsilon C(R)$.

PROOF. If $K \varepsilon \mathfrak{R}$, then

$$h(z) = \sup \{\textstyle\int z \, d\alpha : \alpha \varepsilon K\}$$

defines a continuous support function h on $C(R)$ (see (5)) satisfying

$$h(z) \leqq \sup z(R). \tag{17}$$

If α is a probability measure, write $\alpha \leqq h$ instead of $\int z \, d\alpha \leqq h(z)$ for all $z \varepsilon C(R)$. Then $\alpha \leqq h$ for a probability measure α is equivalent to $\alpha \varepsilon K$ (see (5)). So we can express (15) by

$$P(\cdot, \omega) \leqq h_\omega . \tag{15'}$$

Clearly (14), (15′) imply (16). So assume (16). Let $\mathfrak{A}$ be a class of closed sets in R containing all compact sets and all complements of some countable base closed under finite union and intersection of the topology of R. Let X be a separable subspace of $C(R)$ such that for any $A \in \mathfrak{A}$ and any $\epsilon > 0$ there is an $x_{A,\epsilon} \in X$ such that

$$x_{A,\epsilon}(A) \subseteq \{1\}, x_{A,\epsilon}(R - A_\epsilon) \subseteq \{0\}, x_{A,\epsilon}(R) \subseteq \langle 0, 1 \rangle,$$

where A_ϵ is the ϵ-neighborhood of A and $\langle 0, 1 \rangle$ denotes the closed unit interval.

The assumptions of Theorem 1 are satisfied for X, and putting $\nu = x^*$ we see that (i) follows from (16). So Theorem 1 provides us with a map $\omega \rightarrow x_\omega^*$ such that $x_\omega^*(x) \leqq h_\omega(x)$ and

$$\int x \, d\nu = \int x_\omega^*(x)\mu(d\omega) \tag{19}$$

for $x \in X$. Applying the Hahn-Banach theorem, extend x_ω^* to a linear functional $\tilde{x}_\omega^*$ on $C(R)$ dominated by h_ω. Because $h_\omega(z) \leqq \sup z(R)$ for all $z \in C(R)$ (see (17)), $\tilde{x}_\omega^*$ is a nonnegative linear functional on $C(R)$ (consider negative z).

By a theorem of Alexandroff (see [21], p. 161) there is a finitely additive nonnegative set function $P(\cdot, \omega)$ defined on the set algebra generated by the topology satisfying

$$P(E, \omega) = \sup\{P(A, \omega): A \text{ closed}, A \subseteq E\},$$

which is connected with $\tilde{x}_\omega^*$ by

$$\tilde{x}_\omega^*(z) = \int z(r)P(dr, \omega) \tag{20}$$

for $z \in C(R)$ and by

$$P(A, \omega) = \lim_{k\to\infty} \tilde{x}_\omega^*(x_{A,1/k}) \tag{20a}$$

for $A \in \mathfrak{A}$.

We show that $P(\cdot, \omega)$ is a probability measure for μ-almost all ω.

Let $\delta > 0$, $n \geqq 1$ and $A^{(n)}$ be a compact set such that

$$\nu(A^{(n)}) > 1 - \delta 2^{-2n}$$

(see [21], p. 161). Then by (19) and (20a)

$$\mu\{\omega: P(A^{(n)}, \omega) > 1 - 2^{-n}\} > 1 - \delta 2^{-n},$$

so that

$$\mu\{\omega: P(A^{(n)}, \omega) > 1 - 2^{-n} \text{ for all } n\} > 1 - \delta.$$

For ω in this latter set $P(\cdot, \omega)$ is a measure (see [21], p. 161), so that $P(\cdot, \omega)$ is a probability measure for μ-almost all ω. $P(A, \cdot)$ is $\mathfrak{B}$-measurable for any $A \in \mathfrak{A}$, so P is a Markov kernel (because $\mathfrak{A}$ contains enough sets). For the same reason (20) implies (14). (15′) follows from (20) and the fact that $\tilde{x}_\omega^*$ is dominated by h_ω. P. A. Meyer's result (see Acknowledgement) can be stated as the following

Corollary. *Let R and Ω be compact metric and let $\omega \rightarrow K_\omega$ be a map from Ω into $\mathfrak{R}$ such that $\bigcup_{\omega \in \Omega} \{\omega\} \times K_\omega$ is closed within $\Omega \times$ the set of probability measures in R*

endowed with the weak**-topology. Let* ν *be a probability measure in* R. *In order that there exist a Markov kernel* P *from* Ω *to* R *such that* $\nu = P\mu$ *and* $P(\cdot, \omega) \varepsilon K_\omega$ *for* μ*-almost all* ω, *it is necessary and sufficient that*

$$\int z\, d\nu \leqq \int \sup \{\int z\, d\alpha : \alpha \varepsilon K\omega\} \mu(d\omega)$$

for all $z \varepsilon C(R)$.

It is important to know under what assumptions one can replace (15′) by a similar condition involving only sets instead of continuous functions. Let $\mathfrak{U}$ be the topology of R. Then a (normalized) capacity alternating of order 2 (see Choquet [5]) is a real function f on $\mathfrak{U}$ such that

$$f(\varnothing) = 0, \qquad f(R) = 1, \qquad f(U) \leqq f(V)$$

whenever $U \subseteq V$,

$$f(U) + f(V) \geqq f(U \cup V) + f(U \cap V),$$

and

$$f(U) = \lim_n f(U_n)$$

whenever U is increasing and $U = \bigcup_{n \geqq 1} U_n$. Let $(\Omega, \mathfrak{B}, \mu)$ be a probability space. A kernel alternating of order 2 from Ω to R is a function F on $\mathfrak{U} \times \Omega$ such that for any $\omega \varepsilon \Omega$, $F(\cdot, \omega)$ is a capacity alternating of order 2 and for any $U \varepsilon \mathfrak{U}$, $F(U, \cdot)$ is $\mathfrak{B}$-measurable. Define

$$f(U) = \int F(U, \omega) \mu(d\omega)$$

for $U \varepsilon \mathfrak{U}$. Then f is a capacity alternating of order 2.

THEOREM 4. *Let* ν *be a Borel probability measure in* R. *Then* $\nu \leqq f$ (i.e., $\nu(U) \leqq f(U)$ *for every* $U \varepsilon \mathfrak{U}$) *if and only if there is a Markov kernel* P *from* Ω *to* R *such that* $\nu = P\mu$ *and*

$$P(\cdot, \omega) \leqq F(\cdot, \omega)$$

for μ*-almost all* ω. (We remark that $\nu \leqq h$ and $\nu \leqq f$ have a different meaning: for support functions h it means $\int z d\nu \leqq h(z)$ for all $z \varepsilon C(R)$, for capacities f it means $\nu(U) \leqq f(U)$ for all $U \varepsilon \mathfrak{U}$.)

PROOF. The "if" part is trivial. So assume $\nu \leqq f$. For any ω

$$(21) \quad h_\omega(z) = \inf z(R) + \int_0^\infty F(r : z(r) > \inf z(R) + t, \omega)\, dt \leqq \sup z(R)$$

defines a support function on $C(R)$ such that for any probability measure π, $\pi \leqq F(\cdot, \omega)$ is equivalent to $\pi \leqq h_\omega$ (Choquet [5]). To see, e.g., that h_ω is a support function, remark that $\inf z(R)$ in (21) can be replaced by any smaller constant (the same in both occurrences) without changing the value of $h_\omega(z)$ and use the defining properties of capacities alternating of order 2. The subadditivity of h_ω follows from Theorem 54.1 in [5].

Because an analogous statement is true for f in place of $F(\cdot, \omega)$, Theorem 4 follows immediately from Theorem 3.

Now let R be a complete separable metric space with distance d. Then the set Ω of all nonempty closed subsets ω of R endowed with the distance (introduced by Hausdorff)

$$\delta(\omega, \tilde{\omega}) = \max\{\sup_{r\varepsilon\omega} d'(r, \tilde{\omega}), \sup_{s\varepsilon\tilde{\omega}} d'(\omega, s)\}$$

where

$$d'(r, \tilde{\omega}) = \inf_{s\varepsilon\tilde{\omega}} \min\{d(r, s), 1\}$$

is a complete separable metric space (compare [16], p. 20).

Let μ be a Borel probability measure in Ω. The set function f defined on all open subsets U of R by

$$f(U) = \mu\{\omega : \omega \cap U \neq \emptyset\}$$

is the capacity alternating of order ∞ corresponding to μ in the sense of Choquet [5].

The following result has been proved in [23] for compact metric R in a different way.

THEOREM 5. *If ν is a Borel probability measure in R, then $\nu \leqq f$* (i.e., $\nu(U) \leqq f(U)$ *for every open* U) *if and only if there is a probability measure α in $R \times \Omega$ with marginals ν and μ such that*

$$\alpha\{r \varepsilon \omega\} = 1.$$

PROOF. Put

$$F(U, \omega) = 1 \qquad \text{if } U \cap \omega \neq \emptyset$$
$$= 0 \qquad \text{otherwise.}$$

Then F is a kernel alternating of order 2 and

$$f(U) = \mu\{\omega : \omega \cap U \neq \emptyset\} = \int F(U, \omega)\mu(d\omega).$$

So by Theorem 4 $\nu \leqq f$ implies the existence of a Markov kernel from Ω to R such that $\nu = P\mu$ and $P(R - \omega, \omega) \leqq F(r - \omega, \omega) = 0$, therefore $P(\omega, \omega) = 1$. $\alpha = P \times \mu$ (see (2)) has marginals ν and μ and

$$\alpha\{r \varepsilon \omega\} = \int P(\omega, \omega)\mu(d\omega) = 1.$$

The converse is trivial.

As an application, let ξ be a random variable with values in R. An observation of ξ will usually be corrupted by errors, so that it may be considered as a random variable η different from ξ. About the connection between ξ and η one might assume, e.g., that

$$\Pr\{d(\xi, \eta) > \delta \mid \xi\} \leqq \epsilon \tag{22}$$

for certain positive ϵ and δ. Or more generally

$$\Pr\{d(\xi, \eta) > a \mid \xi\} \leqq \varphi(a, \xi), \tag{23}$$

where φ is a nonnegative measurable function on $R^1 \times R$ (R^1 = real line) such that $\varphi(0, s) = 1$ and $\varphi(\cdot, s)$ is right continuous and nonincreasing for every $s \varepsilon R$. The question naturally arises about the possible distributions of η for a given distribution of ξ (assuming (23)). For simplicity let us suppose that (R, d) has the property that for any $U \varepsilon \mathfrak{U}$, $s \varepsilon U$ and $r \varepsilon R - U$

$$d(r, s) > d(r, U),$$

where $d(r, U) = \inf_{u \varepsilon U} d(r, u)$. This is not a serious restriction because, e.g., Banach spaces and complete Riemann manifolds have this property. Then we claim that a probability distribution ν is a possible distribution of a random variable η satisfying (23) if and only if

$$\nu(U) \leqq E\varphi(d(\xi, U), \xi) \tag{24}$$

for any $U \varepsilon \mathfrak{U}$, where E means expectation. Also, the right side of (24) is a capacity alternating of order ∞.

It follows from the above assumption on (R, d) that if η satisfies (23) then

$$\begin{aligned}\Pr\{\eta \varepsilon U\} &\leqq \Pr\{d(\xi, \eta) > d(\xi, U) \text{ or } \xi \varepsilon U\} \\ &= \int_{R-U} \Pr\{d(r, \eta) > d(r, U) \mid \xi = r\} \Pr\{\xi \varepsilon dr\} + \Pr\{\xi \varepsilon U\} \\ &\leqq E\varphi(d(\xi, U), \xi).\end{aligned}$$

Therefore the distribution ν of ξ satisfies (24).

Conversely if ν satisfies (24), let Γ be a random variable with values in Ω such that almost surely Γ is a sphere in R with center ξ and radius ρ_Γ (we allow $\rho_\Gamma = \infty$, in which case $\Gamma = R$), where $\Pr\{\rho_\Gamma > a \mid \xi\} = \varphi(a, \xi)$ for any $a \geqq 0$. So $\Pr\{\Gamma = R \mid \xi\} = \lim_{a \to \infty} \varphi(a, \xi)$. Let f be the capacity alternating of order ∞ corresponding to the distribution Γ. Then

$$f(U) = \Pr\{\Gamma \cap U \neq \emptyset\} = \Pr\{\rho_\Gamma > d(\xi, U)\} = E\varphi(d(\xi, U), \xi),$$

so that by (24) $\nu \leqq f$. By Theorem 5 there is a random variable η with distribution ν such that $\Pr\{\eta \varepsilon \Gamma\} = 1$. But then

$$\Pr\{d(\xi, \eta) > a \mid \xi\} \leqq \Pr\{\rho_\Gamma > a \mid \xi\} = \varphi(a, \xi),$$

which proves the claim.

In [23] an analogue of the lemma of Neyman and Pearson has been proved for capacities alternating of order ∞ (such as the right side of (24)) in the case where R is finite. It would be worthwhile to generalize this result to Polish spaces R (with some modifications involving, e.g., the role of the relative entropy). It appears that the mentioned result holds true if one replaces capacities of order ∞ by those of order 2 and that in fact capacities of order 2 may in a sense be characterized this way. In this connection it is important to observe that the set of possible distributions of the observed random variable η given the distribution of ξ will not be defined by a capacity of order ∞ if one replaces (22) by $\Pr\{d(\xi, \eta) > \delta\} \leqq \epsilon$, but it will be defined by a capacity of order 2 (see (31)).

Now let (R, ϑ, ν) and $(\Omega, \mathfrak{B}, \mu)$ be probability spaces, m a measure on $\vartheta \times \mathfrak{B}$ with a σ-finite $\mathfrak{B}$-marginal m_0. Fréchet [8] has asked under what conditions there exists a probability measure α on $\vartheta \times \mathfrak{B}$ with marginals ν and μ such that

$$\alpha \leqq m.$$

For finite R and Ω a necessary condition (see (26) below) has been suggested by Fréchet and its sufficiency has been proved by Berge using graph theory (unpublished) and by Dall'Aglio [6]. The general case has been settled at the same time in a different context by Kellerer [10], see also [11], [12], [13], [14]. We will prove below a slight variant of Kellerer's result under the assumption that R is Polish and ϑ is the σ-algebra of Borel sets. Then there is a Markov kernel M from Ω to R such that for any $D \in \vartheta$, $M(D, \cdot)$ is a copy of the conditional expectation of D with respect to $\mathcal{B}$ and the measure m, i.e.,

$$m(D \times B) = \int_B M(D, \omega) m_0(d\omega), \tag{25}$$

where $D \varepsilon \vartheta$, $B \varepsilon \mathcal{B}$.

THEOREM 6. *There is a probability measure $\alpha \leqq m$ on $\vartheta \times \mathcal{B}$ with marginals ν and μ if and only if for all $D \varepsilon \vartheta$, $B \varepsilon \mathcal{B}$*

$$\nu(D) + \mu(B) \leqq 1 + m(D \times B). \tag{26}$$

PROOF. The necessity of (26) is trivial. So assume (26). Without loss of generality we may also assume $m_0 \ll \mu$. Put

$$F(U, \omega) = \min \left\{ \frac{dm_0}{d\mu}(\omega) M(U, \omega), 1 \right\}$$

for open $U \subseteq R$ and $\omega \varepsilon \Omega$, where $dm_0/d\mu$ denotes a fixed $\mathcal{B}$-measurable copy of the density. From (26) follows $\mu \leqq m_0$, so that $dm_0/d\mu \geqq 1$ may be assumed everywhere. It is easily checked that F is a kernel alternating of order 2. Now if

$$B = \left\{ \omega: \frac{dm_0}{d\mu}(\omega) M(U, \omega) \leqq 1 \right\}$$

then by (26) and (25)

$$\begin{aligned} \nu(U) &\leqq \mu(\Omega - B) + m(U \times B) \\ &= \int_{\Omega - B} 1 \, d\mu + \int_B M(U, \omega) \frac{dm_0}{d\mu}(\omega) \mu(d\omega) \\ &= \int F(U, \omega) \mu(d\omega), \end{aligned}$$

so that by Theorem 4 there is a Markov kernel P from Ω to R such that $P\mu = \nu$ and $P(\cdot, \omega) \leqq F(\cdot, \omega)$ for μ-almost all ω. So if $\alpha = P \times \mu$, then the marginals of α are ν and μ and

$$\begin{aligned} \alpha(U \times B) &= \int_B P(U, \omega) \mu(d\omega) \\ &\leqq \int_B F(U, \omega) \mu(d\omega) \\ &\leqq \int_B M(U, \omega) m_0(d\omega) \\ &= m(U \times B) \end{aligned}$$

for open $U \subseteq R$ and $B \varepsilon \mathcal{B}$. The theorem follows by approximation.

6. Probability measures with given marginals. Let S, T be complete separable metric spaces, $\hat{y}$ and $\hat{z}$ positive (not necessarily bounded above) continuous functions on S and T respectively, bounded away from 0. Put

$$\hat{x} = \hat{y} \circ p_S + \hat{z} \circ p_T ,$$

where p_S (resp. p_T) is the projection of $S \times T$ onto S (resp. T), and let X be the Banach space of continuous real functions x on $S \times T$ such that

$$\|x\| = \sup\{|x(s, t)|/\hat{x}(s, t) : s \varepsilon S, t \varepsilon T\} < \infty.$$

Similarly let Y (resp. Z) be the Banach space of continuous functions y on S (resp. z on T) such that $\|y\| = \sup(|y|/\hat{y})(S) < \infty$ (resp. $\|z\| = \sup(|z|/\hat{z})(T) < \infty$). Then, e.g., $y \varepsilon Y$ is equivalent to $y \circ p_S \varepsilon X$.

Let Π be the set of all (Borel-) probability measures π in $S \times T$ such that $\hat{x}$ is π-integrable, endowed with the topology $\mathfrak{J}$ generated by the functionals $\pi \to \int x \, d\pi$ for $x \varepsilon X$ (i.e., the relativized weak* topology, when Π is considered as a subset of X^*). Let Λ be a nonempty $\mathfrak{J}$-closed convex subset of Π and μ and ν (Borel-) probability measures in S and T respectively such that $\hat{y}$ and $\hat{z}$ are μ- and ν-integrable, respectively.

THEOREM 7. *A necessary and sufficient condition for the existence of a probability measure* $\lambda \varepsilon \Lambda$ *such that*

$$\mu = \lambda \circ p_S^{-1} \qquad (27)$$
$$\nu = \lambda \circ p_T^{-1}$$

is that

$$\int y \, d\mu + z \, d\nu \leqq \sup\{\textstyle\int (y \circ p_S + z \circ p_T) \, d\gamma : \gamma \varepsilon \Lambda\} \qquad (28)$$

for all $y \varepsilon Y$, $z \varepsilon Z$.

PROOF. The necessity of (28) is trivial. So assume (28). Let M be the set of all pairs (α, β), where α and β are probability measures in S and T respectively such that $\hat{y}$ is α-integrable and $\hat{z}$ is β-integrable. Then $(\mu, \nu) \varepsilon M$. M is a subset of $(Y \times Z)^*$ by the agreement

$$(\alpha, \beta)(y, z) = \int y \, d\alpha + \int z \, d\beta.$$

The relativized weak* topology in M is denoted by $\mathfrak{J}_1$. $\mathfrak{J}_1$ is metrizable: Replacing $S \times T$ by S for simplicity it is sufficient to show that the set Π_S of probability measures α in S such that $\hat{y}$ is α-integrable is a metrizable subset of Y^* in the weak* topology. But if

$$(\varphi_S \alpha)(E) = \int_E \hat{y} \, d\alpha$$

for any Borel set $E \subseteq S$, then φ_S is a homeomorphism from Π_S onto the closed subset $\{m : \int (1/\hat{y}) \, dm = 1\}$ of the set of finite positive measures in S in the weak* topology, i.e., endowed with the Lévy-Prokhorov distance (see [21], p. 166). Therefore Π_S is metrizable. Consider the set M_Λ consisting of all $(\alpha, \beta) \varepsilon M$ such that there is a $\gamma \varepsilon \Lambda$ with α and β as marginals. M_Λ is convex. (μ, ν) is in the $\mathfrak{J}_1$-closure of M_Λ, for otherwise (μ, ν) would not be in the weak* closure of M_Λ (within $(Y \times Z)^*$), so that there would exist $(y, z) \varepsilon Y \times Z$ with

$$(\mu, \nu)(y, z) > \sup\{(\alpha, \beta)(y, z):(\alpha, \beta) \varepsilon M_\Lambda\}$$

(see [15] 14.4 and 17.6), in contradiction to (28). Let $(\alpha_n, \beta_n)_{n \geqq 1}$ be a sequence of elements of M_Λ $\mathfrak{J}_1$-converging to (μ, ν) and let $\lambda_n \varepsilon \Lambda$ have marginals α_n and β_n. Then

$$\lim_n \int \hat{x}\, d\lambda_n = \lim_n \left(\int \hat{y}\, d\alpha_n + \int \hat{z}\, d\beta_n\right) = \int \hat{y}\, d\mu + \int \hat{z}\, d\nu,$$

so that $\sup_n \int \hat{x}\, d\lambda_n < \infty$. It follows from [21], p. 170 and the above discussion about the map φ_S that for any $\epsilon > 0$ there are compacts $K_S^\epsilon \subseteq S$ and $K_T^\epsilon \subseteq T$ such that

$$\sup_n \int_{S-K_S^\epsilon} \hat{y}\, d\alpha_n < \epsilon$$

and

$$\sup_n \int_{T-K_T^\epsilon} \hat{z}\, d\beta_n < \epsilon.$$

Therefore if $K^\epsilon = K_S^\epsilon \times K_T^\epsilon$ we have

$$\sup \int_{(S\times T)-K^\epsilon} \hat{x}\, d\lambda_n < 2\epsilon,$$

so if φ is defined by $(\varphi\gamma)(E) = \int_E \hat{x}\, d\gamma$ for any Borel set E of $S \times T$, it follows from [21], p. 170 that the sequence $(\varphi\lambda_n)_{n \geqq 1}$ is relatively compact. The same is true for $(\lambda_n)_{n \geqq 1}$ (with respect to the $\mathfrak{J}$-topology). If λ is any cluster point of $(\lambda_n)_{n \geqq 1}$, then $\lambda \varepsilon \Lambda$ (because Λ is closed), and (27), because the projections are continuous. It is clear from this proof that an entirely analogous result holds if one replaces S, T by any finite number $S, T, \cdots, R$ of Polish spaces.

Theorem 7 enables us to obtain a generalization of the Blackwell-Stein-Sherman theorem to the noncompact case. We remark that in the following the k-dimensional vectorspace R^k can easily be replaced by any separable Banach space, if one keeps in mind the two well known facts: Any continuous concave function is the infimum of continuous affine functions, and, any lower- (or upper-) semicontinuous concave function is continuous.

THEOREM 8. *Let $(\mu_n)_{n \geqq 1}$ be a sequence of probability measures in R^k. Then a necessary and sufficient condition for the existence of a k-dimensional martingale $(\xi_n)_{n \geqq 1}$ (see [7], [19]) such that the distribution of ξ_n is μ_n for all n is that all μ_n have means and that for any concave function x on R^k the sequence $(\int x\, d\mu_n)_{n \geqq 1}$ is nonincreasing (the values of the integrals may be $-\infty$).*

PROOF. The necessity of the condition is well known (Jensen's inequality, see [7]). To prove sufficiency, it is enough to prove the following: If two probability measures μ and ν in R^k have means and satisfy $\int x\, d\mu \geqq \int x\, d\nu$ for every concave x, then there is a probability measure λ in $R^k \times R^k$ with marginals μ and ν such that the conditional expectation of the last k coordinates given the first k coordinates is the first k coordinates. For if we know this, we can clearly construct a martingale $(\xi_n)_{n \geqq 1}$ for the theorem as a Markov process. To prove the existence of a λ we use Theorem 7, putting $S = T = R^k$,

$$\hat{y}(t) = \hat{z}(t) = 1 + |t|$$

($|t|$ being the Euclidean length of t) and

$\Lambda = \{\lambda : \lambda \varepsilon \Pi$ and for all bounded continuous functions y on S

$$\int p_T(y \circ p_S)\, d\lambda = \int p_S(y \circ p_S)\, d\lambda\},$$

i.e., Λ is the set of all probability measures in $S \times T$ which are the joint distribution of some k-dimensional martingale (the last equation simply is the martingale equation written in a form which makes it transparent that Λ is $\mathfrak{I}$-closed). Because the assumptions before Theorem 7 are satisfied, our proof will be completed, if we can show (28).

Let z_0 be the smallest concave function $\geqq z$, i.e., z_0 is the infimum of the set of affine functions on R^k which are $\geqq z$ if this set is nonempty and $z_0 = \infty$ otherwise. Then

$$\int y\, d\mu + \int z\, d\nu \leqq \int y\, d\mu + \int z_0\, d\nu \leqq \int (y + z_0)\, d\mu \leqq \sup_{s \varepsilon S} (y(s) + z_0(s)),$$

where $y + z_0 = \infty$ if $z_0 = \infty$. Let r be any real number $< \sup\{y(s) + z_0(s) : s \varepsilon S\}$. Then for some $s \varepsilon S$, $r < y(s) + z_0(s)$. We have to show

$$r < \sup\{\int (y \circ p_S + z \circ p_T)\, d\gamma : \gamma \varepsilon \Lambda\}. \tag{29}$$

For any $t \varepsilon T$ let Λ_t be the set of probability measures in T with expectation t. The function z_1 on T defined by

$$z_1(t) = \sup\{\int z\, d\alpha : \alpha \varepsilon \Lambda_t\}$$

is concave and $\geqq z$, so $z_1 \geqq z_0$ and hence $r < y(s) + z_1(s)$. By definition of $z_1(s)$ there is an $\alpha \varepsilon \Lambda_z$ with

$$r < y(s) + \int z\, d\alpha = \int (y \circ p_S + z \circ p_T)\, d\gamma,$$

where $\gamma = \delta_s \times \alpha \varepsilon \Lambda$. That proves (28) and the theorem.

In an analogous way one gets the following

THEOREM 9. *Let $(\mu_n)_{n \geqq 1}$ be a sequence of probability measures in R^1. Then a necessary and sufficient condition for the existence of an upper semimartingale* ([8], *submartingale in the sense of* [19]) *$(\xi_n)_{n \geqq 1}$ such that the distribution of ξ_n is μ_n for all n is that all μ_n have means and that for any concave nonincreasing function x the sequence $(\int x\, d\mu_n)_{n \geqq 1}$ is nonincreasing.*

Now let S be a topological group whose topology is Polish (so, e.g., any second countable locally compact group). For any two (Borel-) probability measures μ and α in S the convolution $\mu * \alpha$ is defined by

$$(\mu * \alpha)(E) = \int \mu(Es^{-1})\alpha(ds) = \int \alpha(s^{-1}E)\mu(ds)$$

for E Borel.

THEOREM 10. *Let μ and ν be probability measures in S. Then $\mu * \alpha = \nu$ for some probability measure α in S if and only if for all bounded continuous real functions x*

on S

$$\int x\,d\nu \leqq \sup\{\int x(ts)\mu(dt): s \in S\}.$$

PROOF. This can be deduced from Theorem 7. However, a direct proof seems as simple. Let X be the Banach space of bounded continuous real functions on S. The set

$$K = \{\beta:\beta = \mu * \alpha \text{ for some probability measure } \alpha\}$$

is convex. It is also a closed subset of the set of all probability measures in S with respect to the relativized weak* topology (i.e., with respect to the Prokhorov distance [21]), for if $\beta_n = \mu * \alpha_n$ for $n \geqq 1$ and $\lim_n \beta_n = \beta$, then by Theorem 1.12 of [21] or Proposition 1 of [17] given $\epsilon > 0$ there are compact subsets K_ϵ and K of R such that

$$\int \alpha_n(s^{-1}K_\epsilon)\mu(ds) = (\mu * \alpha_n)(K_\epsilon) > 1 - \epsilon$$

for all n and $\mu(K) > 1 - \epsilon$. Hence

$$\alpha_n(K^{-1}K_\epsilon) \geqq \int_K \alpha_n(s^{-1}K_\epsilon)\mu(ds) > 1 - 2\epsilon$$

for all n, so that $\{\alpha_n:n \geqq 1\}$ is relatively compact and therefore $(\alpha_n)_{n\geqq 1}$ has a cluster point α. Clearly $\beta = \mu * \alpha$.

Therefore, applying (5) to the closure $\bar{K}$ of K in X^* we see that for any probability measure ν, $\nu \in K$ if and only if

$$\int x\,d\nu \leqq \sup\{\int x\,d\beta:\beta \in \bar{K}\}$$

for all $x \in X$. But

$$\begin{aligned} \sup\{\int x\,d\beta:\beta \in \bar{K}\} &\geqq \sup\{\int x\,d\beta:\beta \in K\} \\ &= \sup\{\int (\int x(st)\mu(dt))\alpha(ds): \alpha \text{ any probability measure in } S\} \\ &= \sup\{\int x(st)\mu(dt): s \in S\}. \end{aligned}$$

H. Kellerer told me an even simpler proof of Theorem 10 applying the fact that a positive linear functional l can always be extended (where $l(y) = \int x\,d\nu$ if $y(s) = \int x(ts)\mu(dt)$).

As in the case of Theorem 3 it is important to know under which conditions one can replace functions by sets in (28). One would expect that this is possible if the set Λ can be defined in terms of a capacity f alternating of order 2 (compare Theorem 4). That, however, is not true in general even if f is alternating of order ∞.

One needs the additional assumption that the probability measure in the space of closed subsets of $S \times T$ corresponding to f (f alternating of order ∞) has a support which is linearly ordered by inclusion. We will prove only a special case of this, which seems to imply the more interesting applications.

Let ω be a nonempty closed subset of $S \times T$ and $\epsilon \geqq 0$.

THEOREM 11. *There is a probability measure* λ *in* $S \times T$ *with marginals* μ *and* ν

such that $\lambda(\omega) \geqq 1 - \epsilon$, *if and only if for all open* $U \subseteq T$

$$\nu(U) \leqq \mu(p_S(\omega \cap (S \times U))) + \epsilon. \tag{30}$$

($p_S(\omega \cap (S \times U))$ is analytic (Souslinien [3]) and therefore μ-measurable).

REMARKS. In (30) open sets U can be replaced by closed sets.

$$\min \{\mu(p_S(\omega \cap (S \times U))) + \epsilon, 1\} \tag{31}$$

as a function of U is a capacity alternating of order 2.

Condition (30) can be replaced by the following: for all open $U \subseteq T$ and $V \subseteq S$

$$\mu(V) + \nu(U) \leqq \sup \{\gamma(V \times T) + \gamma(S \times U):$$
$$\gamma \text{ is a probability measure in } S \times T \text{ and } \gamma(\omega) \geqq 1 - \varepsilon\}.$$

So Theorem 11 serves the purpose of replacing functions by sets in condition (28). We know that Theorem 7 holds for any finite number of spaces $S, T, \cdots, R$. This is not true for Theorem 11 even if $\epsilon = 0$, as the following example shows (compare Kellerer [12].

$$S = T = \{1, 2\}, \qquad R = \{1, 2, 3\},$$

$$\omega = \{(1, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 3)\},$$

$$\mu\{1\} = \mu\{2\} = \nu\{1\} = \nu\{2\} = \tfrac{1}{2}, \qquad \alpha\{1\} = \tfrac{1}{3}, \qquad \alpha\{2\} = \tfrac{1}{2}, \qquad \alpha\{3\} = \tfrac{1}{6},$$

$$\text{then} \quad \mu(V) + \nu(U) + \alpha(W) \leqq \sup \{\gamma(V \times T \times R) + \gamma(S \times U \times R)$$
$$+ \gamma(S \times T \times W) : \gamma(\omega) = 1\}$$

for all $V \subseteq S$, $U \subseteq T$, $W \subseteq R$, but there is no γ in $S \times T \times R$ with marginals μ, ν and α such that $\gamma(\omega) = 1$.

PROOF OF THEOREM 11. The necessity of (30) is clear. To prove sufficiency we apply Theorem 7, putting $\hat{y} = \hat{z} = 1$, $\Lambda = \{\lambda : \lambda \in \Pi, \lambda(\omega) \geqq 1 - \epsilon\}$. We have to verify (28). We may assume that y and z are positive (by adding constants) and that $p_S(\omega) = S$ (by adjoining a point t_0 to T and enlarging ω by $S \times \{t_0\}$). Then using (30)

$$\int z \, d\nu = \int_0^{\sup z(T)} \nu\{t : z(t) > r\} \, dr$$
$$\leqq \int_0^{\sup z(T)} \min \{\mu\{s : \sup z(\omega_s) > r\} + \epsilon, 1\} \, dr,$$

where $\omega_s = \{t : (s, t) \in \omega\}$. Now if $z_0(s) = \sup z(\omega_s)$, then z_0 is a bounded non-negative μ-measurable function and

$$\int z \, d\nu \leqq \int_0^{\sup z(T)} \min \{\mu\{s : z_0(s) > r\}, 1 - \epsilon\} \, dr + \epsilon \sup z(T)$$
$$= \int_0^{\sup z_0(S)} \min \{\mu\{s : z_0(s) > r\}, 1 - \epsilon\} \, dr + \epsilon \sup z(T)$$
$$= \sup \{\textstyle\int z_0 \, d\tilde{\mu} : \tilde{\mu} \text{ is a finite measure on } S, \tilde{\mu}(S) = 1 - \epsilon$$
$$\text{and} \quad \tilde{\mu} \leqq \mu\} + \epsilon \sup z(T),$$

as is easily seen. Therefore

$$\int y\,d\mu + \int z\,d\nu \leqq \sup\{\int (y + z_0)\,d\tilde{\mu} : \tilde{\mu}(S) = 1 - \epsilon \quad \text{and} \quad \tilde{\mu} \leqq \mu\}$$
$$+ \epsilon(\sup y(S) + \sup z(T))$$
$$\leqq (1 - \epsilon) \sup (y + z_0)(S) + \epsilon \sup (y(S) + z(T))$$
$$= (1 - \epsilon) \sup \{y(s) + z(t) : (s, t) \in \omega\} + \epsilon \sup (y(S) + z(T))$$
$$= \sup\{\int (y \circ p_S + z \circ p_T)\,d\gamma : \gamma \in \Lambda\},$$

proving the theorem.

As an application one may take $S = T$ and ω a closed nonempty partial ordering.

Another application concerns the Lévy-Prokhorov distance:

Let (S, d) be a complete separable metric space. As has been shown by Prokhorov [21], the set of (Borel-) probability measures in S with the distance

$$L(\mu, \nu) = \inf\{\epsilon : \nu(A) \leqq \mu(A_\epsilon) + \epsilon \quad \text{for all closed} \quad A \subseteq S\}$$
$$= \inf\{\epsilon : \mu(A) \leqq \nu(A_\epsilon) + \epsilon \quad \text{for all closed} \quad A \subseteq S\}$$

where $A_\epsilon = \{s : d(s, A) \leqq \epsilon\}$, is a complete separable metric space whose topology is the relativized weak* topology.

Corollary. $L(\mu, \nu) = \min\{\epsilon$: *there is a probability measure* λ *in* $S \times S$ *with marginals* μ *and* ν *such that* $\lambda\{d(s, s') > \epsilon\} \leqq \epsilon\}$.

That the two distances whose equality is asserted in the corollary induce the same topology is known (see Hammersley [9] if S is a finite dimensional vector space and Skorokhod ([22], p. 281) if S is Polish, were an elegant proof of an even stronger statement concerning almost everywhere convergence is given).

REFERENCES

[1] Blackwell, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 93–102.

[2] Blackwell, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

[3] Bourbaki, N. (1948). *Topologie générale*, (Chap. 9). Hermann, Paris.

[4] Cartier-Fell-Meyer. To appear at the end of 1964 in the *Bull. Soc. Math. de France*.

[5] Choquet, G. (1953). Theory of capacities. *Ann. Inst. Fourier* **5** 131–292.

[6] Dall'Aglio, G. (1961). Sulle distribuzioni doppie con margini assegnati soggette a delle limitazioni. *Giorn. Ist. Ital. Attuari* **94**.

[7] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[8] Fréchet, M. (1957). Les tableaux de corrélation dont les marges et des bornes sont données. *Ann. Univ. Lyon Sect. A* **20** 13–31.

[9] Hammersley, J. M. (1952). An extension of the Slutzky-Fréchet theorem. *Acta Math.* **87** 243–257.

[10] Kellerer, H. (1961). Funktionen auf Produkträumen mit vorgegebenen Marginalfunktionen. *Math. Ann.* **144** 323–344.

[11] Kellerer, H. G. (1964). Bemerkung zu einem Satz von H. Richter. *Archiv der Mathematik* **15** 204–207.

[12] Kellerer, H. G. (1964). Maßtheoretische Marginalprobleme. *Math. Annalen* **153** 168–198.

[13] KELLERER, H. G. (1964). Marginalprobleme für Funktionen. *Math. Annalen* **154** 147–156.

[14] KELLERER, H. G. (1964). Schnittmaß-Funktionen in mehrfachen Produkträumen. *Math. Annalen* **155** 369–391.

[15] KELLEY, J. L. and NAMIOKA, I. (1963). *Linear Topological Spaces.* Van Nostrand, Princeton.

[16] KURATOWSKI, C. (1950). *Topologie* **2**. Warsaw.

[17] LE CAM, L. (1957). Convergence in distribution of stochastic processes. *Univ. California Publ. Statist.* **2** 207–236.

[18] LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.

[19] LOÈVE, M. (1963). *Probability Theory,* (3rd ed.). Van Nostrand, Princeton.

[20] MEYER, P. A. (1964). Progrès récents dans la théorie des cônes convexe à base compacte. Séminaire Brelot-Choquet-Deny **8** No. 1.

[21] PROKHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** 157–214.

[22] SKORODHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.

[23] STRASSEN, V. (1964). Messfehler und Information. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **2** 273–305.