*The Annals of Probability*
1993, Vol. 21, No. 1, 136–142

# THE EXISTENCE OF PROBABILITY MEASURES WITH GIVEN MARGINALS

BY HEINZ J. SKALA

*University of Paderborn*

In an important paper, Strassen discussed the existence of measures on product spaces with given marginals in the context of Polish spaces. We generalize his results to arbitrary Hausdorff spaces and indicate some applications such as measures with given support and stochastic inequalities on partially ordered Hausdorff spaces. In the second part of our paper, we state two results on the general moment problem, thus generalizing earlier theorems due to Kemperman.

**1. Introduction and notation.** In a celebrated paper, Strassen (1965) stated a necessary and sufficient condition for the existence of probability measures with given marginals. It turns out that his theorem still holds if we replace Polish spaces by Hausdorff spaces and restrict the possible solutions to a narrowly closed, convex set of Radon measures. Strassen's theorem and some of its consequences have been extended by Edwards (1978) to completely regular spaces. His technique depends on compactification arguments and is different from ours. An equivalent of our Corollary 6 was established by Hansel and Troallic (1986) by first solving the easier content version and then using Henry's extension theorem. Essentially the same result occurs in Kellerer's (1984) important paper which was pointed out by a referee to the author. Kellerer's paper contains a lot of results related to ours. His main arguments strongly depend on the well-known duality theorem of linear programming. Plebanek (1989) obtained a version of Corollary 6 where only one of the marginals is assumed to be Radon.

Let $X$ be a Hausdorff space and denote by $M_+^b(X)$ ($M_+^1(X)$) the nonnegative bounded (probability) Radon measures on $X$. We endow $M_+^b(X)$ with the narrow topology [weak topology in the sense of Topsøe (1970)], that is, the weakest topology for which all maps $\lambda \to \int h\, d\lambda = \lambda(h)$ are lower semicontinuous (l.s.c.) whenever $h$ is a bounded l.s.c. function. In this topology, a net $(\lambda_i)$ converges towards $\lambda$ (we write $\lambda_i \to_n \lambda$) if and only if $\lim \lambda_i(X) = \lambda(X)$ and $\liminf \lambda_i(G) \geq \lambda(G)$ for all open subsets $G \subset X$. (Equivalent conditions are given in Topsøe's Portmanteau theorem.)

$Bd_{B0}(X)$ denotes the bounded Borel-measurable functions on $X$. As usual we shall write $\mathscr{F}(X)$, $\mathscr{K}(X)$, $\mathscr{G}(X)$ and $\mathscr{B}(X)$ for the pavings on $X$ of closed, compact, open and Borel sets, respectively.

For further basic definitions and results we refer the reader to Bourbaki (1987), Kelley (1955), Schwartz (1973) and Topsøe (1970).

Received October 1990; revised June 1991.
*AMS* 1991 *subject classifications*. Primary 60B05; secondary 60B99.
*Key words and phrases*. Probability measures with given marginals, moment problem.

**2. On two theorems of Strassen.**

THEOREM 1. *Let $S$ and $T$ be Hausdorff spaces and let $\Lambda \neq \varnothing$ be a narrowly closed convex subset of $M_+^1(S \times T)$. In order that there exists a $\lambda \in \Lambda$ with given marginals $\mu \in M_+^1(S)$ and $\nu \in M_+^1(T)$, it is necessary and sufficient that*

$$\int f\,d\mu + \int g\,d\nu \leq \sup\left\{\int (f \oplus g)\,d\gamma\colon \gamma \in \Lambda\right\} \tag{1}$$

*for all bounded Borel-measurable functions $f\colon S \to \mathbb{R}$ and $g\colon T \to \mathbb{R}$. [$f \oplus g(s,t) =_{\text{def}} f(s) + g(t)$ for all $(s,t) \in S \times T$.]*

PROOF. The necessity of (1) being clear, we only prove sufficiency. Assume (1) and denote by $M_\Lambda$ the set of all pairs $(\alpha, \beta) \in M_+^1(S) \times M_+^1(T)$ such that there exists a $\gamma \in \Lambda$ with $(\alpha, \beta)$ as marginals.

The pairing

$$\langle(\alpha, \beta), (f, g)\rangle = \int f\,d\alpha + \int g\,d\beta$$

puts the two vector spaces $M(S) \times M(T)$ and $Bd_{B0}(S) \oplus Bd_{B0}(T)$ in duality and we observe that the topology

$$\sigma = \sigma(M(S) \times M(T), Bd_{B0}(S) \oplus Bd_{B0}(T))$$

relative to $\langle(\alpha, \beta), (f, g)\rangle \to \int f\,d\alpha + \int g\,d\beta$ is just the product of the topologies $\sigma(M(S), Bd_{B0}(S))$ and $\sigma(M(T), Bd_{B0}(T))$ [see Bourbaki (1987), page 50].

That $(\mu, \nu)$ is in the closure $\mathrm{cl}_\sigma M_\Lambda$ follows from the Hahn–Banach theorem together with the representation theorem of weakly continuous linear functionals.

Therefore there exists a net $((\alpha, \beta)_i)$ of elements in $M_\Lambda$ converging weakly to $(\mu, \nu)$ which is the case if and only if $(\alpha_i)$ converges weakly to $\mu$ and $(\beta_i)$ converges weakly to $\nu$. As the relative topologies of the weak topologies on $M_+^b(S)$ and $M_+^b(T)$ are finer than the narrow topologies $nM_+(S)$ and $nM_+(T)$, respectively, we infer $\alpha_i \to_n \mu$ and $\beta_i \to_n \nu$. Assign to every $(\alpha, \beta)_i$ a $\gamma_i \in \Lambda$ with marginals $(\alpha, \beta)_i$ and assume for the moment that the net $(\gamma_i)$ is compact, that is, universal subnets of $(\gamma_i)$ converge. Let $(\gamma_{i_j})$ be a universal subnet and assume $\gamma_{i_j} \to_n \gamma$; then $\gamma \in \Lambda$ as $\Lambda$ is narrowly closed. That the marginal measures $\alpha_{i_j}$ and $\beta_{i_j}$ of $\gamma_{i_j}$ converge narrowly to the marginal measures $\alpha$ and $\beta$ of $\gamma$ follows from Topsøe's [(1970), page 40] Portmanteau theorem.

It remains to show that $(\gamma_i)$ is compact. This is the case if and only if for every subclass $\mathscr{G}$ of $\mathscr{G}(S \times T)$ which dominates $\mathscr{K}(S \times T)$ (i.e., $\forall\, K \in \mathscr{K}(S \times T)\ \exists\, G \in \mathscr{G}\colon G \supset K$) we have

$$\inf_{\mathscr{G}'} \limsup_i \min_{G \in \mathscr{G}'} \gamma_i(G^c) = 0, \tag{2}$$

where the infimum is taken over all finite subclasses $\mathscr{G}'$ of $\mathscr{G}$ [see Topsøe (1970), page 43] and $\limsup \gamma_i(S \times T) < \infty$. The latter condition is trivially satisfied and we verify a somewhat stronger condition than (2). For a given $\varepsilon > 0$, there exist compact sets $K_S$ and $K_T$ such that $\mu(K_S^c) \le \varepsilon$ and $\mu(K_T^c) \le \varepsilon$. Let $G \in \mathscr{G}$, $G \supset K_S \times K_T$, then there are open sets $G_S$ and $G_T$ such that $K_S \times K_T \subset G_S \times G_T \subset G$ [cf. Kelley (1955), page 142]. By the Portmanteau theorem [Topsøe (1970), page 40], $\limsup \gamma_i(G_S^c \times T) = \limsup \alpha_i(G_S^c) \le \mu(G_S^c) \le \mu(K_S^c) \le \varepsilon$ and $\limsup \gamma_i(S \times G_T^c) = \limsup \beta_i(G_T^c) \le \nu(G_T^c) \le \nu(K_T^c) \le \varepsilon$. As $G^c \subset (G_S \times G_T)^c = (G_S^c \times T) \cup (S \times G_T^c)$, we obtain

$$\begin{aligned}\limsup \gamma_i(G^c) &\le \limsup \gamma_i(G_S \times G_T)^c \\ &= \limsup \gamma_i((G_S^c \times T) \cup (S \times G_T^c)) \\ &\le \limsup(\gamma_i(G_S^c \times T) + \gamma_i(S \times G_T^c)) \\ &\le \limsup \gamma_i(G_S^c \times T) + \limsup \gamma_i(S \times G_T^c) \le 2\varepsilon,\end{aligned}$$

which obviously implies (2) and our theorem is proved. □

REMARK. It is sufficient to assume (1) for all bounded l.s.c. functions.

COROLLARY 2. *Let $S$ and $T$ be Radon spaces and $\mu, \nu$ Borel probability measures on $S$ and $T$, respectively. Let $\Lambda \ne \varnothing$ be a narrowly closed convex subset of Borel probability measures on $S \times T$. Then the following are equivalent*: (i) *There exists a Radon measure $\lambda \in \Lambda$ with marginals $\mu$ and $\nu$.* (ii) *There exists a $\lambda \in \Lambda$ with marginals $\mu$ and $\nu$.* (iii) $\mu(f) + \nu(g) \le \sup\{\gamma(f \oplus g)$: $\gamma \in \Lambda\}$ *for all* $f \in Bd_{B0}(S)$, $g \in Bd_{B0}(T)$.

PROOF. Trivially (i) ⇒ (ii) ⇒ (iii).

(iii) ⇒ (i): For every $\gamma \in \Lambda$, we define $\tilde{\gamma}(B) = \sup_{K \subset B} \inf_{G \supset K} \gamma(G)$, $B \in \mathscr{B}(S \times T)$. Then $\tilde{\gamma}$ is a Radon probability measure which coincides with $\gamma$ on $\mathscr{B}(S) \otimes \mathscr{B}(T)$ and is dominated on $\mathscr{G}(S \times T)$ by $\gamma$ (i.e., $\tilde{\gamma}(G) \le \gamma(G)$ for all $G$]. See Topsøe [(1970), page 29]. By the Portmanteau theorem, the constant net $\gamma_i = \gamma$ converges to $\tilde{\gamma}$. Thus the narrow closure $\tilde{\Lambda}$ of $\{\tilde{\gamma}: \gamma \in \Lambda\}$ in $M_+^1(S \times T)$ is a subset of $\Lambda$. As $\tilde{\gamma}(f \oplus g) = \gamma(f \oplus g)$, we can apply Theorem 1 to $\tilde{\Lambda}$ and obtain the desired result. □

COROLLARY 3. *In the following cases, Corollary* 2 *holds*: (a) *$S$ and $T$ are Souslin spaces.* (b) *$S$ and $T$ are Lusin spaces.* (c) *$S$ and $T$ are Polish spaces.*

PROOF. Souslin, Lusin and Polish spaces are Radon spaces. □

REMARKS. (i) Corollary 3(c) is essentially Strassen's (1965) Theorem 7. (Put $\hat{y} = 1$ and $\hat{z} = 1$.)

(ii) Corollary 2 cannot be inferred directly from Theorem 1 because products of Radon spaces need not be Radon, in general.

It seems to be an open question under which set-theoretical assumptions it can be proved that the product of Radon spaces is again a Radon space.

(iii) As the narrow topology defined on the space of *all Borel* measures on $S \times T$ is not Hausdorff if $S \times T$ is not a Radon space, Corollary 2 is not applicable for $\Lambda = \{\lambda_0\}$ if $\lambda_0$ is not Radon.

An easy consequence of Theorem 1 is:

COROLLARY 4. *Let $S$, $T$, $\mu$, $\nu$ be as in Theorem 1 and let $F \neq \varnothing$ be a closed subset of $S \times T$. There exists a Radon probability measure $\lambda$ with given marginals $\mu$, $\nu$ and $\lambda(F) = 1$ iff*

$$(3) \qquad \int f\,d\mu + \int g\,d\nu \leq \sup\{f(s) + g(t): (s,t) \in F\}.$$

REMARK. This corollary was stated by Edwards (1978) for completely regular spaces.

As an interesting application of Theorem 1, we prove a counterpart to Strassen's (1965) Theorem 11.

THEOREM 5. *Let $S$ and $T$ be Hausdorff spaces, $F \neq \varnothing$ a closed subset of $S \times T$, $\varepsilon \geq 0$ a given real number and let $\Lambda = \{\gamma \in M_+^1(S \times T): \gamma(F) \geq 1 - \varepsilon\}$. There exists a Radon probability measure $\lambda \in \Lambda$ with given marginals $\mu \in M_+^1(S)$ and $\nu \in M_+^1(T)$ if and only if*

$$(4) \qquad \mu(B_S) + \nu(B_T) \leq \sup\{\gamma(B_S \times T) + \gamma(S \times B_T): \gamma \in \Lambda\}$$

*for all Borel $B_S \subset S$ and $B_T \subset T$. [It suffices to assume* (4) *for all open $G_S \subset S$ and $G_T \subset T$].*

PROOF. $1_F$ is u.s.c., thus by the definition of the narrow topology, $\gamma \to \gamma(1_F)$ is u.s.c. and therefore the set $\Lambda$ is narrowly closed. Trivially $\Lambda$ is convex and nonempty.

The necessity of (4) being clear, we prove sufficiency by establishing (1) for nonnegative Borel-step functions $f$ and $g$ with representations $f = \sum_{i=1}^m \alpha_i 1_{A_i}$, $\alpha_i \geq 0$, and $g = \sum_{j=1}^n \beta_j 1_{B_j}$, $\beta_j \geq 0$, with $S = A_1 \supseteq \cdots \supseteq A_m \supseteq A_{m+1} = \varnothing$, $\mu(A_i \setminus A_{i+1}) > 0$ and $T = \bigcup B_j$ with $B_j$ pairwise disjoint and $\nu(B_j) > 0$. As

$$\sup\left\{\int h\,d\gamma: \gamma \in \Lambda\right\} = (1 - \varepsilon)\sup_F h + \varepsilon \sup_{S \times T} h$$

for every bounded Borel measurable function $h$ on $S \times T$, we infer from (4) that

$$(5) \qquad \mu(B_S) + \nu(B_T) \leq 1 + \varepsilon \quad \text{whenever } B_S \times B_T \cap F = \varnothing.$$

Furthermore, $\sup\{\int f \oplus g\,d\gamma: \gamma \in \Lambda\} = (1 - \varepsilon)c + \varepsilon(a + b)$, where $a = \sum_{i=1}^m \alpha_i$, $b = \max_{j \leq n} \beta_j$ and $c = \sup_F f \oplus g$. To give a formula for $c$, we define the sets $J_i = \{j: A_i \times B_j \cap F \neq \varnothing\}$ and denote $\tilde{\mathscr{B}}_i = \bigcup\{B_j: j \in J_i\}$, $\varepsilon_i = \nu(\tilde{\mathscr{B}}_i^c)$ and $a_j = \sum_{i:\, J_i \ni j} \alpha_i$. We obtain $c = \max\{\beta_j + a_j: j \in J_1\}$. Let $k$ be the biggest $i$

with $\varepsilon_i \leq \varepsilon$. Using (5) for $B_S = A_i$ and $B_T = \tilde{B}_i^c$, we obtain $\mu(A_i) \leq \nu(\tilde{\mathscr{B}}_i) + \varepsilon$ for $i > k$, while $\mu(A_i) \leq 1$ for $i \leq k$. If $\alpha = \sum_{i \leq k} \alpha_i \geq c - b$, then we put $l = k$, otherwise $l = k + 1$, so that $(b + \alpha - c)\varepsilon_l \leq (b + \alpha - c)\varepsilon$ in any case. We then obtain the desired formula:

$$\begin{aligned}\int f\,d\mu + \int g\,d\nu &= \sum_{i=1}^{m} \alpha_i \mu(A_i) + \sum_{j=1}^{n} \beta_j \nu(B_j) \\ &\leq \varepsilon \sum_{i>k} \alpha_i + \alpha + \sum_j \Bigg(\beta_j + \sum_{i:\, J_i \ni j,\, i>k} \alpha_i\Bigg) \nu(B_j) \\ &= \varepsilon(a - \alpha) + \alpha\varepsilon_l + \sum_{j \in J_l} (\beta_j + a_j)\nu(B_j) + \sum_{j \notin J_l} \beta_j \nu(B_j) \\ &\leq \varepsilon(a - \alpha) + \alpha\varepsilon_l + c(1 - \varepsilon_l) + b\varepsilon_l \\ &= c + \varepsilon(a - \alpha) + (b + \alpha - c)\varepsilon_l \\ &\leq (1 - \varepsilon)c + \varepsilon(a + b). \qquad \square\end{aligned}$$

A short look at the proof of Theorem 5 shows:

COROLLARY 6. *Let $S$, $T$, $F$ and $\varepsilon$ be as in Theorem 5. Then there exists a Radon probability measure $\lambda$ with given marginals $\mu \in M_+^1(S)$ and $\nu \in M_+^1(T)$ such that $\lambda(F) \geq 1 - \varepsilon$ iff* (5) *holds, that is, $\mu(B_S) + \nu(B_T) \leq 1 + \varepsilon$ whenever $B_S \times B_T \cap F = \varnothing$.*

REMARKS. (i) If $F = F_S \times F_T$, then (5) reduces to $\mu(F_S) \geq 1 - \varepsilon$ and $\nu(F_T) \geq 1 - \varepsilon$. (ii) If $S = T$ and $F$ is the diagonal $\{(s, t)\colon s = t\}$, then (5) reduces to $|\mu - \nu| \leq \varepsilon$.

A further immediate consequence of Theorem 5 results in a generalization of the key implication of Theorem 1 in Kamae, Krengel and O'Brien (1977). Let $S$ be a Hausdorff space endowed with a closed partial ordering and denote by $\mathscr{I}(S)$ the family of increasing Borel sets, that is, $B \in \mathscr{I}(S)$ iff $s \in B$ and $s \leq t$ together imply $t \in B$. We shall write $\mu \prec \nu$ iff $\mu(B) \leq \nu(B)$ for all $B \in \mathscr{I}(S)$.

COROLLARY 7. *Let $\mu, \nu \in M_+^1(S)$; then $\mu \prec \nu$ implies the existence of a $\lambda \in M_+^1(S \times S)$ with support in $F = \{(s, t) \in S \times S\colon s \leq t\}$ such that $\mu$ and $\nu$ are the marginals of $\lambda$.* [*It suffices to assume $\mu(G) \leq \nu(G)$ for all open increasing sets.*]

Using different arguments, this result has also been obtained by Kellerer (1984). It is therefore not uninteresting to observe that a basic duality result by Kellerer from which other duality theorems may be derived is also an easy consequence of Theorem 5.

COROLLARY 8. *Let S, T, F be as in Theorem 5 and let* $\Lambda$ *be the set of all Radon probability measures on* $S \times T$ *with* $\mu$ *and* $\nu$ *as marginals. Then*

$$\begin{aligned} \sup\{\gamma(F): \gamma \in \Lambda\} &= \inf\left\{\int f\,d\mu + \int g\,d\nu: 1_F \le f \oplus g;\ f, g \text{ Borel}\right\} \\ &= \inf\{\mu(B_S) + \nu(B_T): F \subset (B_S \times T) \\ &\qquad \cup (S \times B_T): B_S, B_T \text{ Borel}\} \end{aligned} \tag{6}$$

*and the supremum is attained.*

PROOF. $\Lambda$ is narrowly compact [this follows from our proof of Theorem 10 or from Kellerer (1984)]. Hence $\gamma \to \gamma(1_F)$ assumes its supremum. Trivially sup $\le$ inf. An easy calculation shows that sup $<$ inf leads to a contradiction with Theorem 5. □

**3. A general moment problem.** The following theorem is due to Kemperman (1983) for completely regular spaces. The proof for arbitrary Hausdorff spaces is similar to Kemperman's and is left to the reader.

THEOREM 9. *Let X be a Hausdorff space and let* $\Lambda \neq \varnothing$ *be a convex and narrowly compact set of Radon probability measures. For an arbitrary set of bounded l.s.c. functions* $h_i$, $i \in I$, *and given real numbers* $c_i$ *the following conditions are equivalent*: (i) *There exists a* $\lambda \in \Lambda$ *such that* $\lambda(h_i) \le c_i$ *for all* $i \in I$. (ii) *For every nonnegative linear combination* $\sum_{k=1}^n a_k h_{i_k}$ *there exists a* $\lambda \in \Lambda$ *such that* $\lambda(\sum_{k=1}^n a_k h_{i_k}) \le \sum_{k=1}^n a_k c_{i_k}$.

REMARK. It is well known that $M_+^1(X)$ is narrowly compact iff $X$ is a compact space.

Frequently, moment conditions come together with given marginal measures. Therefore we state:

COROLLARY 10. *Let S and T be Hausdorff spaces with given marginals* $\mu \in M_+^1(S)$ *and* $\nu \in M_+^1(T)$. *Denote by* $\Lambda$ *the set of all Radon probability measures on* $S \times T$ *with* $\mu$ *and* $\nu$ *as marginals. For an arbitrary set of bounded l.s.c. functions* $h_i$: $S \times T \to \mathbb{R}$, $i \in I$, *and given real numbers* $c_i$, *there exists a* $\lambda \in \Lambda$ *such that* $\lambda(h_i) \le c_i$ *for all* $i \in I$ *iff condition* (ii) *of Theorem 9 holds.*

PROOF. It is well known that there exists a unique Radon probability measure $\lambda$ on $S \times T$ such that $\lambda(B_S \times B_T) = \mu(B_S) \cdot \nu(B_T)$ for all Borel subsets $B_S$ and $B_T$ of $S$ and $T$, respectively. Therefore we have $\Lambda \neq \varnothing$. [See, e.g., Schwartz (1973), page 63, and observe that for probability measures the essential outer measure coincides with the measure of a set.] Trivially, $\Lambda$ is convex and only its narrow compactness remains to be proved. As $\mu$ and $\nu$ are Radon, $\Lambda$ is tight, that is, for every $\varepsilon > 0$, there exists a compact $K \subset S \times T$ such that $\lambda(K^c) \le \varepsilon$ for all $\lambda \in \Lambda$. Trivially $\sup\{\lambda(S \times T): \lambda \in \Lambda\} < \infty$ and therefore $\Lambda$ is net-compact (in fact, relatively compact) for the narrow topology [Topsøe (1970), page 43]. Let $(\lambda_i)$ be a net, $\lambda_i \in \Lambda$, such that $\lambda_i \to_n \gamma$. Then the marginals of the $\lambda_i$ (i.e., $\mu$ and $\nu$) must converge narrowly towards the marginals of $\gamma$. (Take, e.g., an open subset $G$ of $S$ and apply the Portmanteau theorem to $G \times T$.) Thus $\gamma \in \Lambda$ and $\Lambda$ is narrowly closed and hence narrowly compact. □

**Acknowledgments.** The author wants to thank Th. Landes for helpful discussions and an anonymous referee for pointing out additional relevant references and improvements in the presentation of this material.

## REFERENCES

BOURBAKI, N. (1987). *Topological Vector Spaces*. Springer, New York.

EDWARDS, D. A. (1978). On the existence of probability measures with given marginals. *Ann. Inst. Fourier* (*Grenoble*) **28** 53–78.

HANSEL, G. and TROALLIC, J. P. (1986). Sur le probléme des marges. *Probab. Theory. Related Fields* **71** 357–366.

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912.

KELLERER, H. G. (1984). Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete* **67** 399–432.

KELLEY, J. L. (1955). *General Topology*. Van Nostrand, New York.

KEMPERMAN, J. H. B. (1983). On the role of duality theory in the theory of moments. In *Semi-Infinite Programming and Applications* (A. V. Fiacco, K. O. Kortanek, eds.) 63–92. Springer, New York.

PLEBANEK, G. (1989). Measures on two-dimensional products. *Mathematika* **36** 253–258.

SCHWARTZ, L. (1973). *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*. Oxford Univ. Press.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

TOPSØE, F. (1970). *Topology and Measure*. Springer, New York.

UNIVERSITY OF PADERBORN
INSTITUTE OF STATISTICS AND ECONOMETRICS
WARBURGER STRAßE 100
4790 PADERBORN
GERMANY