*The Annals of Probability*
1985, Vol. 13, No. 3, 898–922

# DE FINETTI-TYPE THEOREMS: AN ANALYTICAL APPROACH[1]

BY PAUL RESSEL

*Stanford University and Katholische Universität Eichstätt*

A famous theorem of De Finetti (1931) shows that an exchangeable sequence of {0, 1}-valued random variables is a unique mixture of coin tossing processes. Many generalizations of this result have been found; Hewitt and Savage (1955) for example extended De Finetti's theorem to arbitrary compact state spaces (instead of just {0, 1}).

Another type of question arises naturally in this context. How can mixtures of independent and identically distributed random sequences with certain specified (say normal, Poisson, or exponential) distributions be characterized among all exchangeable sequences?

We present a general theorem from which the "abstract" theorem of Hewitt and Savage as well as many "concrete" results—as just mentioned—can be easily deduced. Our main tools are some rather recent results from harmonic analysis on abelian semigroups.

**1. Introduction.** De Finetti's famous classical theorem (1931) says that an exchangeable sequence of {0, 1}-valued random variables is a unique mixture of i.i.d. Bernoulli sequences. More precisely: let $X_1, X_2, \cdots$ be {0, 1}-valued and assume that

$$P(X_1 = x_1, \cdots, X_n = x_n) = P(X_1 = x_{\sigma(1)}, \cdots, X_n = x_{\sigma(n)})$$

holds for all $n \in \mathbb{N}$, $x_1, \cdots, x_n \in \{0, 1\}$ and all permutations $\sigma$ of $\{1, \cdots, n\}$, then for some unique probability measure $\mu$ on [0, 1]

$$P(X_1 = x_1, \cdots, X_n = x_n) = \int_0^1 p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}\, d\mu(p). \tag{1}$$

Many generalizations and similar results have been found; in fact since around 1970 a renewed strong interest in this area can be observed and many exciting new results were found. Several excellent survey articles are available; see for example Kingman (1978), Diaconis and Freedman (1984), Aldous (1983) and Lauritzen (1984).

One early aim was of course to extend De Finetti's theorem from {0, 1} to more general state spaces and a most satisfactory result was obtained by Hewitt and Savage (1955); they found that {0, 1} can indeed be replaced by any compact Hausdorff space, from which it can immediately be extended to Borel subsets of compact spaces, in particular to all polish or locally compact spaces. Without

---

Received December 1983; revised August 1984.

[1] This research was supported by the Stiftung Volkswagenwerk, Hannover.

*AMS* 1980 *subject classifications.* Primary 60E05, 43A35; secondary 60B99, 62A05, 44A05.

*Key words and phrases.* De Finetti's theorem, exchangeability, Hewitt and Savage's theorem, (completely) positive definite functions on *-semigroups, Radon-presentability, integrated Cauchy functional equation, convolution semigroups, Schoenberg triples, multivariate survival function.

any topological assumptions the result may fail to hold as was shown by Dubins and Freedman (1979).

Given any fixed state space, another type of question seems natural: what are (necessary and sufficient) conditions on a given sequence of exchangeable random variables in order that it turns out to be a mixture of i.i.d. sequences of a particular type, say of normally or exponentially or Poisson distributed random variables. See Freedman (1962) for many interesting examples.

Our aim in the present article is to prove a general theorem from which both the "abstract" representation of Hewitt and Savage and also many "concrete" results (as just mentioned) may be easily deduced. The methods involved are of analytical nature; they use in an essential way some results from harmonic analysis on semigroups, a good introduction to which may be found in Berg, Christensen and Ressel (1984).

**2. Some basic results from harmonic analysis on semigroups.** Let $S$ denote an abelian semigroup, written additively, with a neutral element called 0, and provided also with an *involution* $*: S \to S$ such that $(s+t)^* = s^* + t^*$ and $(s^*)^* = s$ for all $s, t \in S$. In many cases this involution will be the identical one, i.e. $s^* = s$ for all $s \in S$. An example with a nontrivial involution we are going to use later on will be $S = \mathbb{R} \times \mathbb{R}_+$ with the usual addition and the involution $(s_1, s_2)^* = (-s_1, s_2)$.

A function $\varphi: S \to \mathbb{C}$ is called *positive definite* if

$$\textstyle\sum_{j,k=1}^{n} c_j \bar{c}_k \varphi(s_j + s_k^*) \geq 0$$

for all $n \geq 1$, $s_1, \cdots, s_n \in S$ and $c_1, \cdots, c_n \in \mathbb{C}$. The set $\mathscr{P}(S)$ of all positive definite functions on $S$ is a convex cone containing the set $S^*$ of *semicharacters* on $S$ which by definition are those functions $\rho: S \to \mathbb{C}$ for which $\rho(0) = 1$, $\rho(s+t) = \rho(s)\rho(t)$ and $\rho(s^*) = \overline{\rho(s)}$. If $\mu$ is a (nonnegative) Radon measure on $S^*$ such that $\int |\rho(s)| \, d\mu(\rho) < \infty$ for all $s \in S$ (in particular then $\mu(S^*) < \infty$), the function $\varphi(s) := \int \rho(s) \, d\mu(\rho)$ is positive definite since

$$\sum c_j \bar{c}_k \varphi(s_j + s_k^*) = \int |\sum c_j \rho(s_j)|^2 \, d\mu(\rho) \geq 0,$$

and one of the main problems of harmonic analysis on semigroups is to establish the converse which does not hold i.g. without further conditions on $\varphi$ or $S$, see Berg, Christensen and Jensen (1979). One particularly useful condition on $\varphi$ is to assume that $\varphi$ is *exponentially bounded*; this means that there is a function $\alpha: S \to \mathbb{R}_+$ fulfilling $\alpha(0) = 1$, $\alpha(s^*) = \alpha(s)$ and $\alpha(s+t) \leq \alpha(s)\alpha(t)$ ($\alpha$ is then called an *absolute value*) such that $|\varphi(s)| \leq C\alpha(s)$ for some $C \in \mathbb{R}_+$ and all $s \in S$. In Berg and Maserick (1984) the fundamental result is proved that $\mathscr{P}_1^{\alpha}(S)$, the set of $\alpha$-bounded positive definite functions on $S$, normalized by $\varphi(0) = 1$, is a Bauer simplex whose extreme points is precisely the set $S^{\alpha}$ of $\alpha$-bounded semicharacters (it is easily seen that a semicharacter $\rho$ is $\alpha$-bounded iff $|\rho(s)| \leq \alpha(s)$ for all $s \in S$). This implies that each function $\varphi \in \mathscr{P}_1^{\alpha}(S)$ has a unique integral representation

$$\varphi(s) = \int \rho(s)\, d\mu(\rho), \quad s \in S \tag{2}$$

where $\mu$ is a Radon probability measure on $S^*$, concentrated on the compact set $S^\alpha$. In particular if $\varphi$ is bounded (i.e. $\alpha$-bounded w.r.t. the absolute value $\alpha \equiv 1$) then the representing measure $\mu$ is concentrated on $\hat{S} := \{\rho \in S^* \mid |\rho(s)| \le 1$ for all $s \in S\}$, the set of all bounded semicharacters.

In most of the examples later on we actually want the measure $\mu$ to be concentrated on the set $S_+^*$ of nonnegative semicharacters. A necessary condition for this to hold is that $\varphi$ even is *completely positive definite*, by which we mean that not only $\varphi$ but also each translate $\varphi_a$, defined by $\varphi_a(s) = \varphi(a+s)$, is positive definite. In fact

$$\textstyle\sum_{j,k=1}^n c_j \bar{c}_k \varphi_a(s_j + s_k^*) = \displaystyle\int \rho(a) \,|\textstyle\sum_{j=1}^n c_j \rho(s_j)|^2 \, d\mu(\rho) \ge 0$$

if $\mu(S^* \setminus S_+^*) = 0$. The converse statement holds also, as long as we only consider exponentially bounded functions, and is proved now.

**PROPOSITION 1.** *Let $\varphi$ be a completely positive definite exponentially bounded function on S. Then the unique representing measure $\mu$ for $\varphi$ is concentrated on $S_+^*$.*

**PROOF.** Let $\varphi$ be bounded w.r.t. the absolute value $\alpha$, i.e.

$$|\varphi(s)| \le C\alpha(s) \quad \text{for all} \quad s \in S \quad \text{and some} \quad C \in \mathbb{R}_+.$$

Then for $a \in S$ we have $|\varphi_a(s)| = |\varphi(a+s)| \le C\alpha(a+s) \le C\alpha(a)\alpha(s)$ so that all translates of $\varphi$ are likewise $\alpha$-bounded. Therefore we find uniquely determined Radon measures $\mu_a$, $a \in S$, on $S^*$ (which are in fact concentrated on $S^\alpha$) such that

$$\varphi(a+s) = \int \rho(s)\, d\mu_a(\rho), \quad s \in S, \quad a \in S.$$

From $\varphi(a+s) = \int \rho(a+s)\, d\mu(\rho) = \int \rho(s)\rho(a)\, d\mu(\rho)$ we may conclude

$$d\mu_a(\rho) = \rho(a)\, d\mu(\rho) \quad \text{and} \quad d\mu_{a^*}(\rho) = \overline{\rho(a)}\, d\mu(\rho).$$

For the open subset $U_a := \{\rho \in S^* \mid \operatorname{Re} \rho(a) < 0\}$ of $S^*$ we find

$$0 \le (\mu_a + \mu_{a^*})(U_a) = \int_{U_a} [\rho(a) + \overline{\rho(a)}]\, d\mu(\rho) = 2\int_{U_a} \operatorname{Re} \rho(a)\, d\mu(\rho) \le 0$$

whence $\mu(U_a) = 0$, and since $\mu$ is a Radon measure,

$$\mu(\textstyle\bigcup_{a \in S} U_a) = 0$$

showing that $\mu$ is concentrated on those $\rho$ for which $\operatorname{Re}(\rho) \ge 0$. This last condition however implies already $\rho(S) \subseteq [0, \infty[$, because if $0 \ne \rho(s) = |\rho(s)|\, e^{i\theta}$ for some $\theta \in\, ]0, \pi/2]$ (without restriction, since then $\rho(s^*) = |\rho(s^*)|\, e^{-i\theta}$), then for the minimal $n \in \mathbb{N}$ such that $n\theta > \pi/2$ we have $\rho(ns) = |\rho(s)|^n e^{in\theta}$ and therefore Re $\rho(ns) < 0$. Hence $\mu(S^* \backslash S_+^*) = 0$ as asserted. □

REMARK 1. In Ressel (1982, Theorem 2) the above result was shown under the two restrictions of $\varphi$ being bounded and $S$ carrying the identical involution. In this case the representing measure concentrates on $\hat{S}_+ = \{\rho \in S^* \mid 0 \le \rho \le 1\}$ so that by a famous result of Choquet the two classes of bounded completely positive definite and completely monotone functions coincide.

One of the great advantages of the general integral representation for positive definite functions is certainly that the determination of the extreme points is reduced to a rather simple functional equation (although it is not always obvious how the semicharacters of a given semigroup look like). But still the problem remains to establish that certain functions are in fact (completely) positive definite. It turned out that the following result may be very helpful.

APPROXIMATION LEMMA. *Let $(a_{jk})$ be a complex $p \times p$-matrix, let $M \neq \emptyset$ be a set and let $\Phi: M \times M \to \mathbb{C}$ be a bounded positive definite kernel. Suppose that for every $n \in \mathbb{N}$ there exist $\{x_{j,\sigma}^n \mid j = 1, \cdots, p;\ \sigma = 1, \cdots, n\} \subseteq M$ such that*

$$\Phi(x_{j,\sigma}, x_{k,\tau}) = \begin{cases} a_{jk} & \text{if } j \neq k \\ a_{jj} & \text{if } j = k \quad \text{but} \quad \sigma \neq \tau. \end{cases}$$

*Then the matrix $(a_{jk})$ is positive (semi) definite, too.*

This result appears in Christensen and Ressel (1982, Lemma 5), but it was used for special matrices already in Berg and Ressel (1978) and in Ressel (1976).

**3. The main theorem.** Let $\mathscr{F}$ be the family of all complex-valued measurable functions on some measurable space $(\mathscr{X}, \mathscr{B})$ which are bounded by 1. Let $S$ denote an abelian semigroup and let $Z$ denote an abstract set with involution, i.e. there is a mapping $*: Z \to Z$ such that $(z^*)^* = z$ for all $z \in Z$. Furthermore three mappings $\theta: Z \to \mathscr{F}$, $t: Z \to S$, $\beta: Z \to \mathbb{C}\backslash\{0\}$ are given such that $\theta(z^*) = \overline{\theta(z)}$, $\beta(z^*) = \overline{\beta(z)}$, $t(z^*) = (t(z))^*$ and such that $t(Z)$ generates $S$; this means that each $s \in S\backslash\{0\}$ is a finite sum of elements in $t(Z)$.

THEOREM 1. *Let under the assumptions just given $X_1, X_2, \cdots$ be a sequence of $\mathscr{X}$-valued random variables such that*

$$E[\textstyle\prod_{j=1}^n \theta(z_j) \circ X_j] = \prod_{j=1}^n \beta(z_j)\varphi(\sum_{j=1}^n t(z_j)) \tag{3}$$

*for all $n \ge 1$ and all $z_1, \cdots, z_n \in Z$, where $\varphi: S \to \mathbb{C}$ is some function normalized by $\varphi(0) = 1$. Then $\varphi$ is exponentially bounded and positive definite. In case the functions in $\mathscr{F}_0 := \theta(Z)$ are nonnegative and $\beta > 0$ $\varphi$ is even completely positive definite.*

REMARK 2. The seemingly more general case that the function $\varphi$ in (3) depends also explicitly on $n$ can easily be reduced to the case where $\varphi$ is the same function for all $n \in \mathbb{N}$. We define $\tilde{t}: Z \to S \times \mathbb{N}_0$ by $\tilde{t}(z) := (t(z), 1)$ and consider $S \times \mathbb{N}_0$ as a semigroup w.r.t. componentwise addition and the involution $(s, n)^* = (s^*, n)$. Let $\tilde{S}$ be the subsemigroup of $S \times \mathbb{N}_0$ generated by $\tilde{t}(Z)$. Then

$$\varphi_n(\textstyle\sum_{j=1}^n t(z_j)) = \varphi(\sum_{j=1}^n t(z_j), n) = \varphi(\sum_{j=1}^n \tilde{t}(z_j))$$

and Theorem 1 applies. This idea is due to S. Lauritzen.

PROOF. Let $s_1, \cdots, s_p \in S$ be given and let $n \in \mathbb{N}$ be fixed. If none of the $s_j$ is zero we have by assumption

$$s_j = \textstyle\sum_{\ell=1}^{n_j} t(z_{j,\ell}), \quad j = 1, \cdots, p$$

for suitable $z_{j,\ell} \in Z$. We choose $np$ disjoint subsets $N_{j,\sigma} \subseteq \mathbb{N}, j = 1, \cdots, p, \sigma = 1, \cdots, n$ such that $|N_{j,\sigma}| = n_j$, define $\gamma_j := \prod_{\ell=1}^{n_j} \beta(z_{j,\ell})$ and complex-valued random variables $Y_{j,\sigma}$ by

$$Y_{j,\sigma} := \textstyle\prod_{\ell=1}^{n_j} \theta(z_{j,\ell}) \circ X_{k_\ell}, \quad \text{where} \quad N_{j,\sigma} = \{k_1, k_2, \cdots, k_{n_j}\}.$$

Then

$$E(Y_{j,\sigma}\overline{Y_{k,\tau}}) = \begin{cases} \gamma_j\bar{\gamma}_k\varphi(s_j + s_k^*), & \text{if } j \neq k \\ |\gamma_j|^2\varphi(s_j + s_j^*), & \text{if } j = k \text{ but } \sigma \neq \tau, \end{cases}$$

so that by the Approximation Lemma the matrix $(\gamma_j\overline{\gamma_k}\varphi(s_j + s_k^*))_{j,k\le p}$ is positive semidefinite. Since $\gamma_j \neq 0$ for all $j$ this implies $(\varphi(s_j + s_k^*))$ to be positive semidefinite, too.

In case $\theta(Z)$ consists only of nonnegative functions and $\beta > 0$ let $a \in S\backslash\{0\}$ be of the form $a = \sum_{\ell=1}^m t(z_\ell)$ and choose $N_a \subseteq \mathbb{N}$ disjoint of all the $N_{j,\sigma}$ and of cardinality $|N_a| = m$. Put

$$Y_a := \textstyle\prod_{\ell=1}^m \theta(z_\ell) \circ X_{\nu_\ell} \quad \text{where} \quad N_a = \{\nu_1, \cdots, \nu_m\}$$

and $\gamma_a := \prod_{\ell=1}^m \beta(z_\ell)$. Then

$$E(Y_a Y_{j,\sigma} Y_{k,\tau}) = \begin{cases} \gamma_a\gamma_j\gamma_k\varphi(a + s_j + s_k), & \text{if } j \neq k \\ \gamma_a\gamma_j^2\varphi(a + 2s_j), & \text{if } j = k \text{ but } \sigma \neq \tau \end{cases}$$

(note, that the involution on $S$ is now necessarily the identical one) and since $Y_a \geq 0$ we see as before that $(\varphi(a + s_j + s_k))_{j,k\le p}$ is positive semidefinite.

If $0 \in S$ is not a finite sum of elements in $t(Z)$ we must show that the matrices $(\varphi(s_j + s_k^*))$ resp. $(\varphi(a + s_j + s_k^*))$ are also positive semidefinite if one of the $s_j$, say $s_1$ equals 0. In this case we put $Y_{1,\sigma} := 1$ for $\sigma = 1, \cdots, n$ and likewise $\gamma_1 := 1$ and get the result as before.

We still have to show that $\varphi$ is exponentially bounded. From (3) we see that for $s = \sum_{j=1}^n t(z_j)$

$$|\varphi(s)| \leq \prod_{j=1}^n \frac{1}{|\beta(z_j)|}$$

and hence

$$|\varphi(s)| \leq \alpha(s) := \inf\left\{\prod_{j=1}^n \frac{1}{|\beta(z_j)|} \,\middle|\, s = \sum_{j=1}^n t(z_j), n \in \mathbb{N}, z_j \in Z\right\}.$$

Then $\alpha$ is well-defined on $S\backslash\{0\}$ and we complete the definition by $\alpha(0) := 1$. Now a routine argument shows $\alpha(s + s') \le \alpha(s)\alpha(s')$ for all $s, s' \in S$ and $\alpha(s^*) = \alpha(s)$, i.e. $\alpha$ is an absolute value majorizing $\varphi$. This finishes the proof of Theorem 1. □

In all subsequent applications the set $\mathscr{F}_0 = \theta(Z)$ of measurable functions will be rich enough to separate probability measures not only on $(\mathscr{X}, \mathscr{B})$ but also on $(\mathscr{X}^n, \mathscr{B}^n)$ for all $n \ge 1$. This means that given two probability measures $P$, $Q$ on $(\mathscr{X}^n, \mathscr{B}^n)$ such that

$$\int f_1 \otimes f_2 \otimes \cdots \otimes f_n \, dP = \int f_1 \otimes f_2 \otimes \cdots \otimes f_n \, dQ$$

for all $f_j \in F_0$ then $P = Q$ (here $f_1 \otimes \cdots \otimes f_n(x_1, \cdots, x_n) = \prod_{j=1}^n f_j(x_j)$ denotes the usual tensor product of functions). Let us agree to call $\mathscr{F}_0$ *fully separating* in this case.

Corollary 1. *Let in the situation of Theorem 1 the family $\mathscr{F}_0$ be fully separating. Then the sequence $X_1, X_2, \cdots$ is exchangeable.*

By the integral representation theorem mentioned earlier there is a unique Radon probability measure on $S^*$ (supported even by the compact set of $\alpha$-bounded semicharacters, $\alpha$ being derived from $\beta$ as described in the proof of Theorem 1) such that

$$\varphi(s) = \int \rho(s) \, d\mu(\rho), \quad s \in S,$$

implying

$$E\left[\prod_{j=1}^n \theta(z_j) \circ X_j\right] = \int \prod_{j=1}^n \beta(z_j)\rho(t(z_j)) \, d\mu(\rho) \tag{4}$$

for all $n \in \mathbb{N}$ and $z_1, \cdots, z_n \in Z$. In most of the applications below $\theta$ will be one-to-one so that $Z$ might be identified with the subset $\mathscr{F}_0$ of $\mathscr{F}$, closed under complex conjugation. The remaining task will then be to show that $\mu$ is concentrated on the "right" set of semicharacters, i.e. on those $\rho \in S^*$ such that the function $z = f \mapsto \beta(f)\rho(t(f))$ is given as $\int f \, d\kappa$ for some probability measure $\kappa$ on $(\mathscr{X}, \mathscr{B})$. Since in all our applications $\mathscr{F}_0$ will be fully separating, the measure is unique if existent. Note that for general $\rho \in S^*$ the functional $\beta(f)\rho(t(f))$ cannot be expected to be linear.

**4. The theorem of Hewitt and Savage.** Let $\mathscr{X}$ denote a compact Hausdorff space and $\mathscr{B}$ the family of Borel subsets of $\mathscr{X}$. Furthermore let $\mathscr{C}(\mathscr{X})$ be the space of all continuous real-valued functions on $\mathscr{X}$, $\mathscr{C}_+(\mathscr{X})$ the nonnegative members of $\mathscr{C}(\mathscr{X})$ and let $Z = \mathscr{F}_0 = \{f \in \mathscr{C}(\mathscr{X}) \mid 0 \le f \le 1\}$, $\theta = \mathrm{id}$.

Lemma 1. *If $\tau: \mathscr{F}_0 \to [0, 1]$ fulfills $\tau(f) + \tau(g) = 1$ whenever $f, g \in \mathscr{F}_0$, $f + g = 1$, and $\tau(f) + \tau(g) + \tau(h) = 1$ whenever $f, g, h \in \mathscr{F}_0$, $f + g + h = 1$, then $\tau(f) = \int f \, d\kappa$ for some uniquely determined Radon probability measure $\kappa$ on $\mathscr{X}$.*

PROOF. From $0 + 1 = 1$ and $0 + 0 + 1 = 1$ we infer $\tau(0) + \tau(1) = 1 = 2\tau(0) + \tau(1)$, hence $\tau(0) = 0$ and $\tau(1) = 1$. Suppose now $f \leq g$ for $f, g \in \mathscr{F}_0$; then $f + (g - f) + (1 - g) = 1$ implying $\tau(f) + \tau(g - f) + \tau(1 - g) = 1 = \tau(g) + \tau(1 - g)$ and therefore $\tau(f) + \tau(g - f) = \tau(g)$ so that $\tau(f) \leq \tau(g)$, i.e. $\tau$ is monotone. Obviously $\tau(\alpha f) = \alpha\tau(f)$ for all $f \in \mathscr{F}_0$ and all $\alpha \in \mathbb{Q} \cap [0, 1]$, and using the monotonicity of $\tau$ this holds even for all $\alpha \in [0, 1]$. Extending $\tau$ to $\mathscr{C}_+(\mathscr{X})$ by $\tau(f) := \|f\|\,\tau(f/\|f\|)$ for $f \neq 0$ and then in the usual way to $\mathscr{C}(\mathscr{X})$ by $\tau(f) := \tau(f^+) - \tau(f^-)$ shows $\tau$ to be a positive linear functional on $\mathscr{C}(\mathscr{X})$ which by Riesz's representation theorem is induced by some uniquely determined Radon probability measure on $\mathscr{X}$. □

REMARK 3. If $\tau: \mathscr{F}_0 \to [0, 1]$ only fulfills $\tau(f) + \tau(1 - f) = 1$ for all $f \in \mathscr{F}_0$ and $\tau(1) = 1$, it need not be induced by a measure. For example fix some $x_0 \in \mathscr{X}$ and define $\tau$ by $\tau(f) = 0$ if $f(x_0) < ½$, $\tau(f) = ½$ if $f(x_0) = ½$ and $\tau(f) = 1$ if $f(x_0) > ½$.

THEOREM 2. *Let $P$ denote an exchangeable Radon probability measure on the countable infinite product $\mathscr{X}^\infty$ of some compact Hausdorff space $\mathscr{X}$. Then for some uniquely determined Radon probability measure $\mu$ on $M_+^1(\mathscr{X})$, the space of all Radon probabilities on $\mathscr{X}$, we have*

$$P(A) = \int \kappa^\infty(A)\, d\mu(\kappa) \tag{5}$$

*for each Borel set $A \subseteq \mathscr{X}^\infty$. (Here $\kappa^\infty$ denotes the unique extension to a Radon measure on $\mathscr{X}^\infty$ of the usual Kolmogoroff product measure $\kappa \otimes \kappa \otimes \cdots$.)*

PROOF. Let $P$ be the distribution of $X = (X_1, X_2, \cdots)$ and let as before $\mathscr{F}_0$ be the set of all continuous functions $f: \mathscr{X} \to [0, 1]$. Since $P$ is exchangeable, the expectation

$$E(\textstyle\prod_{j=1}^n f_j \circ X_j), \quad f_1, \cdots, f_n \in \mathscr{F}_0$$

depends only on the number of times each $f \in \mathscr{F}_0$ appears among $f_1, \cdots, f_n$; but this means that it can be factorized over the free abelian semigroup over $\mathscr{F}_0$, denoted $S = \mathbb{N}_0^{(\mathscr{F}_0)} = \{s: \mathscr{F}_0 \to \mathbb{N}_0 \mid |\{s \neq 0\}| < \infty\}$. i.e.

$$E(\textstyle\prod_{j=1}^n f_j \circ X_j) = \varphi(\sum_{j=1}^n \delta_{f_j})$$

for some function $\varphi$ on $S$, where $\delta_f \in S$ is one at $f$ and zero at $\mathscr{F}_0\backslash\{f\}$. With $t(f) := \delta_f$ and $\beta \equiv 1$ we may apply Theorem 1 to conclude that $\varphi$ is a (bounded) completely positive definite function. A moment's reflection shows that the dual semigroup $S^*$ can be identified with $\mathbb{R}^{\mathscr{F}_0}$, similarly $\hat{S} \simeq [-1, 1]^{\mathscr{F}_0}$ and $\hat{S}_+ \simeq [0, 1]^{\mathscr{F}_0}$. Equation (4) now takes the form

$$E(\textstyle\prod_{j=1}^n f_j \circ X_j) = \int \prod_{j=1}^n \rho(f_j)\, d\mu(\rho), \quad n \in \mathbb{N}, \quad f_1, \cdots, f_n \in \mathscr{F}_0$$

with $\mu \in M_+^1([0, 1]^{\mathscr{F}_0})$, where—slightly abusing notation—we have identified $\delta_f$ with $f$.

Let $f, g \in \mathscr{F}_0$ be given, $f + g = 1$. Then $1 = E[(f+g) \circ X_1] = E(f \circ X_1) + E(g \circ X_1) = \int [\rho(f) + \rho(g)]\, d\mu(\rho)$, and for $n \in \mathbb{N}$

$$1 = E[\textstyle\prod_{j=1}^n (f+g) \circ X_j] = \sum_{k=0}^n \sum_{\alpha \subseteq \{1,\cdots,n\}; |\alpha| = k} E[\prod_{i\in\alpha} f(X_i) \prod_{j \in \alpha^c} g(X_j)]$$

$$= \sum_{k=0}^n \binom{n}{k} \int (\rho(f))^k (\rho(g))^{n-k}\, d\mu(\rho) = \int [\rho(f) + \rho(g)]^n\, d\mu(\rho)$$

so that $\mu(\{\rho \mid \rho(f) + \rho(g) = 1\}) = 1$ and therefore, $\mu$ being a Radon measure,

$$\mu(\cap_{f\in\mathscr{F}_0}\{\rho \mid \rho(f) + \rho(1-f) = 1\}) = 1.$$

Similarly, if $f + g + h = 1$ for $f, g, h \in \mathscr{F}_0$ we find

$$\begin{aligned} 1 &= E[\textstyle\prod_{j=1}^n (f+g+h) \circ X_j] \\ &= \sum_{k_1,k_2,k_3 \ge 0; k_1+k_2+k_3 = n} \int (\rho(f))^{k_1} (\rho(g))^{k_2} (\rho(h))^{k_3}\, d\mu(\rho) \\ &= \int [\rho(f) + \rho(g) + \rho(h)]^n\, d\mu(\rho) \end{aligned}$$

so that $\rho(f) + \rho(g) + \rho(h) = 1$ $\mu$-a.s.; we see that $\mu$ is in fact concentrated on those $\rho\colon \mathscr{F}_0 \to [0, 1]$ for which the conditions of Lemma 1 are fulfilled, i.e. on the set $M_+^1(\mathscr{X})$ of Radon probabilities on $\mathscr{X}$. Since the pointwise topology on $[0, 1]^{\mathscr{F}_0}$ induces the usual weak (or vague) topology on $M_+^1(\mathscr{X})$, $\mu$ is a Radon measure on $M_+^1(\mathscr{X})$. Now

$$\begin{aligned} E(\textstyle\prod_{j=1}^n f_j \circ X_j) &= \int f_1 \otimes \cdots \otimes f_n\, dP = \varphi(\sum_1^n \delta_{f_j}) = \int \prod_1^n \rho(\delta_{f_j})\, d\mu(\rho) \\ &= \int_{M_+^1(\mathscr{X})} \prod_1^n \int f_j\, d\kappa\, d\mu(\kappa) \\ &= \int_{M_+^1(\mathscr{X})} \int_{\mathscr{X}^\infty} f_1 \otimes \cdots \otimes f_n\, d\kappa^\infty\, d\mu(\kappa) \end{aligned}$$

for all $n \in \mathbb{N}$ and $f_1, \cdots, f_n \in \mathscr{F}_0$, so that routine arguments show

$$P(A) = \int \kappa^\infty(A)\, d\mu(\kappa)$$

at least for all $A \in \mathscr{B} \otimes \mathscr{B} \otimes \cdots$, a $\sigma$-field being contained and not always equal to $\mathscr{B}(\mathscr{X}^\infty)$, the Borel field of $\mathscr{X}^\infty$. Since for any open subset $G \subseteq \mathscr{X}^\infty$ the function $\nu \mapsto \nu(G)$ is lower semicontinuous on $M_+^1(\mathscr{X}^\infty)$ and since furthermore $\kappa \mapsto \kappa^\infty$ is a continuous transformation from $M_+^1(\mathscr{X})$ to $M_+^1(\mathscr{X}^\infty)$ (even a much more general result of this type holds, see for example Ressel (1977, Theorem 2)), we get that $\kappa \mapsto \kappa^\infty(A)$ is Borel measurable on $M_+^1(\mathscr{X})$ for each Borel set $A \subseteq \mathscr{X}^\infty$. If $G \subseteq \mathscr{X}^\infty$ is open then $G = \cup_{\lambda\in\Lambda} G_\lambda$ for an increasingly filtered family of open sets

$G_\lambda \in \mathscr{B} \otimes \mathscr{B} \otimes \cdots$, so that standard results about Radon measures show

$$P(G) = \sup P(G_\lambda) = \sup \int \kappa^\infty(G_\lambda)\, d\mu(\kappa)$$
$$= \int \sup \kappa^\infty(G_\lambda)\, d\mu(\kappa) = \int \kappa^\infty(G)\, d\mu(\kappa)$$

and this equality then extends immediately to all Borel sets in $\mathscr{X}^\infty$, thus finishing the proof. □

REMARK 4. Hewitt and Savage (1955) showed the validity of (5) for all Baire sets in $\mathscr{X}^\infty$, i.e. for all $A \in \mathscr{B}_0 \otimes \mathscr{B}_0 \otimes \cdots$, where $\mathscr{B}_0$ is the $\sigma$-field generated by $\mathscr{C}(\mathscr{X})$. The extension to the Borel subsets of $\mathscr{X}$ was shown in Diaconis and Freedman (1980).

In analogy with a terminology introduced by Hewitt and Savage one might call a Hausdorff space $\mathscr{X}$ *Radon-presentable* if the result of Theorem 2 holds for $\mathscr{X}$. It might be true that all Hausdorff spaces have this property.

THEOREM 3. *Completely regular Hausdorff spaces are Radon-presentable.*

PROOF. Let $\mathscr{Y}$ be completely regular, and let $\mathscr{X}$ denote its Stone-Čzech compactification. If $P \in M_+^1(\mathscr{Y}^\infty)$ then $\tilde{P}: \mathscr{B}(\mathscr{X}^\infty) \to [0, 1]$, defined by $\tilde{P}(B) := P(B \cap \mathscr{Y}^\infty)$ defines a Radon measure $\tilde{P}$ on $\mathscr{X}^\infty$ which is exchangeable if $P$ was. In this case $\tilde{P} = \int \tilde{\kappa}^\infty\, d\tilde{\mu}(\kappa)$ by Theorem 2, where $\tilde{\mu} \in M_+^1(M_+^1(\mathscr{X}))$. Since $P$ is Radon, so is its projection onto the first coordinate, hence for suitable compact sets $K_1 \subseteq K_2 \subseteq K_3 \subseteq \cdots \subseteq \mathscr{Y}$

$$1 = \lim_{j\to\infty} P(K_j \times \mathscr{Y} \times \mathscr{Y} \times \cdots) = \lim_{j\to\infty} \tilde{P}(K_j \times \mathscr{X} \times \mathscr{X} \times \cdots)$$
$$= \lim_{j\to\infty} \int_{M_+^1(\mathscr{X})} \tilde{\kappa}(K_j)\, d\tilde{\mu}(\tilde{\kappa}) = \int_{M_+^1(\mathscr{X})} \tilde{\kappa}\left(\textstyle\bigcup_{j=1}^\infty K_j\right) d\tilde{\mu}(\tilde{\kappa})$$

so that $\tilde{\mu}(\{\tilde{\kappa} \mid \tilde{\kappa}(K) = 1\}) = 1$ for $K = \bigcup_{j=1}^\infty K_j \subseteq \mathscr{Y}$. The set $\tilde{M} := \{\kappa \in M_+^1(\mathscr{X}) \mid \tilde{\kappa}(K) = 1\}$ is a Borel set in $M_+^1(\mathscr{X})$, on which $\tilde{\mu}$ is concentrated, therefore $\tilde{\mu}$ is a Radon measure on $\tilde{M}$ (w.r.t. the trace topology from $M_+^1(\mathscr{X})$).

Since any bounded real-valued continuous function on $\mathscr{Y}$ has a unique continuous extension to $\mathscr{X}$, the natural injection

$$\begin{aligned} \tilde{M} &\to M_+^1(\mathscr{Y}) \\ \tilde{\kappa} &\mapsto \kappa \end{aligned}$$

is continuous, ensuring that $\mu$, defined as the image of $\tilde{\mu}$ under this injection, is again a Radon measure (on the Hausdorff space $M_+^1(\mathscr{Y})$ w.r.t. the usual weak topology).

If now $A \in \mathscr{B}(\mathscr{Y}^\infty)$ then $A = B \cap \mathscr{Y}^\infty$ for some $B \in \mathscr{B}(\mathscr{X}^\infty)$ and

$$P(A) = \tilde{P}(B) = \int_{M^1_+(\mathscr{X})} \tilde{\kappa}^\infty(B)\, d\tilde{\mu}(\tilde{\kappa}) = \int_{\tilde{M}} \tilde{\kappa}^\infty(B \cap K^\infty)\, d\tilde{\mu}(\tilde{\kappa})$$

$$= \int_{M^1_+(\mathscr{Y})} \kappa^\infty(A \cap K^\infty)\, d\mu(\kappa) = \int_{M^1_+(\mathscr{Y})} \kappa^\infty(A)\, d\mu(\kappa)$$

since $\mu(\{\kappa \mid \kappa(K) = 1\}) = 1$. □

The above result implies that all polish and all locally compact spaces are Radon-presentable. It is also possible to show that all analytic spaces (i.e. Hausdorff images of polish spaces under continuous mappings) are Radon-presentable: If $f\colon \mathscr{Y} \to \mathscr{X}$ is a continuous surjection of a polish space $\mathscr{Y}$ onto an analytic space $\mathscr{X}$, then $f$ admits a universally measurable right inverse and using this fact the proof is rather straightforward.

**5. Countable state space.** Let $\mathscr{X}$ be a nonempty finite or countable set (and $\mathscr{B}$ equal the power set of $\mathscr{X}$). In this case it is possible to specify probabilities on $\mathscr{X}$ or $\mathscr{X}^n$ by the complete list of point probabilities. We obtain the framework of Theorem 1 by putting $Z = \mathscr{F}_0 := \{1_{\{x\}} \mid x \in \mathscr{X}\}$, $\theta = \mathrm{id}$, and then identifying $1_{\{x\}}$ with $x$. Hence the mappings $\beta$ and $t$ are now defined on $\mathscr{X}$, i.e. $t\colon \mathscr{X} \to S$ such that $t(\mathscr{X})$ generates $S$; in particular $S$ is countable, too.

THEOREM 4. *Let $P \in M^1_+(\mathscr{X}^\infty)$ have the property*

$$P(x_1, \cdots, x_n) = \textstyle\prod_{j=1}^n \beta(x_j)\varphi(\sum_{j=1}^n t(x_j)) \tag{6}$$

*for all $n \geq 1$ and all $x_1, \cdots, x_n \in \mathscr{X}$, where $\beta\colon \mathscr{X} \to ]0, \infty[$ and $t\colon \mathscr{X} \to S$, a semigroup generated by $t(\mathscr{X})$, and where $\varphi\colon S \to \mathbb{R}$ is normalized by $\varphi(0) = 1$. Then $\varphi$ is an exponentially bounded completely positive definite function on $S$ whose uniquely determined representing measure is concentrated on the relatively compact Borel set*

$$W = W_\beta := \{\rho \in S^*_+ \mid \textstyle\sum_{x \in \mathscr{X}} \beta(x)\rho(t(x)) = 1\}.$$

*$W$ is compact in case $\mathscr{X}$ is finite. Conversely for each $\mu \in M^1_+(W)$ the function $\varphi(s) := \int \rho(s)\, d\mu(\rho)$ defines via* (6) *a probability measure on $\mathscr{X}^\infty$, i.e. the set of all $P$ satisfying* (6) *is affinely isomorphic to the face $\{\mu \in M^1_+(S^*) \mid \mu(W) = 1\}$.*

PROOF. From Theorem 1 we know that $\varphi(s) = \int \rho(s)\, d\mu(\rho)$ where $\mu \in M^1_+(S^*_+)$ has compact support. Equation (6) then becomes

$$P(x_1, \cdots, x_n) = \int \textstyle\prod_{j=1}^n \beta(x_j)\rho(t(x_j))\, d\mu(\rho)$$

implying

$$1 = \sum_{x_1,\cdots,x_n\in\mathscr{X}} P(x_1, \cdots, x_n) = \int [\sum_{x\in\mathscr{X}} \beta(x)\rho(t(x))]^n \, d\mu(\rho)$$

for all $n \in \mathbb{N}$, hence $\mu(W) = 1$. It is easy to see that $W$ in fact is contained in the compact set of all $\alpha$-bounded semicharacters. See the proof of Theorem 1.

The converse statement is nearly immediate. □

REMARK 5. The set $W$ may be empty in which case no probability $P$ on $\mathscr{X}^\infty$ fulfills (6). If for example $S$ is an idempotent semigroup (i.e. $s + s = s$ for all $s \in S$) then all semicharacters are $\{0, 1\}$-valued, so that $W$ would be empty as soon as $\sum \beta(x) < 1$.

REMARK 6. In case $0 = \sum_{i=1}^m t(y_i)$ for $\{y_1, \cdots, y_m\} \subseteq \mathscr{X}$ we have $P(y_1, \cdots, y_m) = \prod_{i=1}^m \beta(y_i)\varphi(0) = \sum_{x\in\mathscr{X}} P(y_1, \cdots, y_m, x) = \sum_x \prod_{i=1}^m \beta(y_i)\beta(x)\varphi(t(x)) = \prod_{i=1}^m \beta(y_i)$, so that necessarily $\varphi(0) = 1$.

REMARK 7. In Lauritzen (1975) a good deal of Theorem 4 was already proved. What was not shown was that $\varphi$ is an exponentially bounded and completely positive definite function. There was also a slight restriction concerning the semigroup $S$ which was assumed to be a subsemigroup of $T \times (\mathbb{N}_0, +)$ for some semigroup $T$, and being generated by some subset of $T \times \{1\}$. This restriction is fulfilled in the following 4 Examples, two of which (Examples 3 and 4) are also contained in Lauritzen's paper, but it is not assumed in Theorem 5 below. More detailed information may be found in Lauritzen (1982, 1984). For given $\beta$, $t$ and $S$ the family

$$\mathscr{E} = \{\kappa \in M_+^1(\mathscr{X}) \mid \kappa(x) = \beta(x)\rho(t(x)),\ \rho \in W\}$$

is a *general exponential model* in the sense of Lauritzen (1975, Definition 3.1).

EXAMPLE 1. Let $\mathscr{X} = \{0, 1, 2, \cdots, k\}$ where $k \in \mathbb{N}$. We wish to determine all $P \in M_+^1(\mathscr{X}^\infty)$ such that

$$P(x_1, \cdots, x_n) = \varphi_n(\textstyle\sum_{i=1}^n x_i) = \varphi(\sum_{i=1}^n (x_i, 1)) \tag{7}$$

for all $x_1, \cdots, x_n \in \mathscr{X}$, and all $n \in \mathbb{N}$. Here $\beta \equiv 1$, $t(x) = (x, 1)$ and $S_k \subset \mathbb{N}_0 \times \mathbb{N}_0$ is the semigroup generated by $t(\mathscr{X}) = \{(0, 1), (1, 1), \cdots, (k, 1)\}$, i.e. $S_k = \{(m, n) \in \mathbb{N}_0 \times \mathbb{N}_0 \mid m \leq kn\}$.

We want to determine the set $W$ from Theorem 4; so we first have to know how the nonnegative semicharacters of $S_k$ look like. See Figure 1.

Let $\rho \in S_{k,+}^*$ and define $u, v \in \mathbb{R}_+$ by $u^k := \rho(k, 1)$ and $v^k := \rho(0, 1)$. We assert that

$$\rho(m, n) = u^m v^{kn-m} \quad \text{for all} \quad (m, n) \in S_k.$$

If $1 \leq \ell < k$ then $(\rho(\ell, 1))^k = \rho(\ell k, k) = \rho(\ell k, \ell)\rho(0, k - \ell) = u^{k\ell} v^{k(k-\ell)}$, hence

$$\rho(\ell, 1) = u^\ell v^{k-\ell}.$$

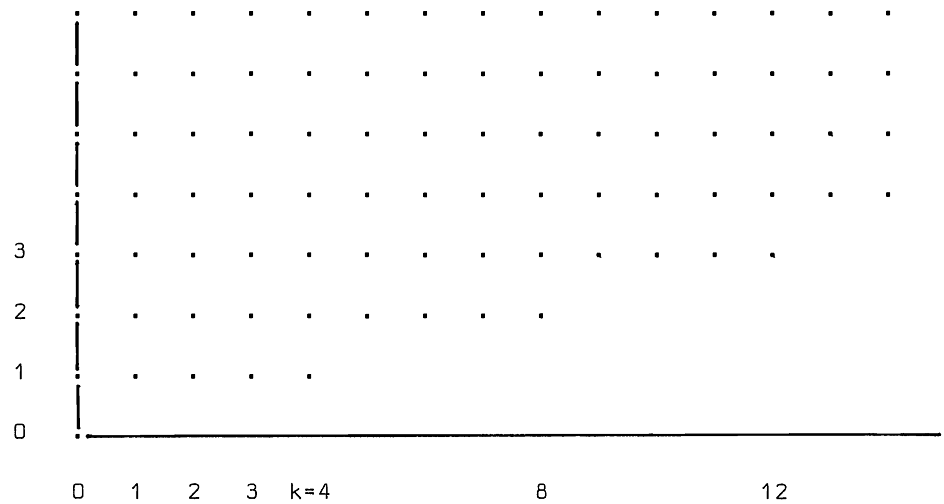


FIG. 1. *The semigroup* $S_4$

Let now $(m, n) \in S_k$ be given. If $m = pk$ for some $p \in \mathbb{N}$ then $(m, n) = (pk, p) + (0, n-p)$ and $\rho(m, n) = u^{kp}v^{k(n-p)} = u^m v^{kn-m}$. If $k$ does not divide $m$ then $kp < m < (p+1)k$, so $n \geq m/k > p$ and $(m, n) = (pk, p) + (m - pk, 1) + (0, n-p-1)$, implying

$$\begin{aligned}\rho(m, n) &= (\rho(k, 1))^p \rho(m - pk, 1)(\rho(0, 1))^{n-p-1}\\ &= u^{kp}u^{m-pk}v^{k-(m-pk)}v^{k(n-p-1)}\\ &= u^m v^{kn-m}.\end{aligned}$$

Therefore $\rho \in W$ iff $\rho(0, 1) + \rho(1, 1) + \cdots + \rho(k, 1) = 1$ iff

$$v^k + uv^{k-1} + u^2v^{k-2} + \cdots + u^{k-1}v + u^k = 1$$

which describes a compact curve-piece $\mathscr{C}_k$ inside the unit square. The corresponding generalized exponential model is given by

$$\mathscr{E} = \{\kappa \in M_+^1(\{0, 1, \cdots, k\}) \mid \kappa(\{j\}) = u^j v^{k-j}, (u, v) \in \mathscr{C}_k\}$$

which for $k = 1$ reduces to all of $M_+^1(\{0, 1\})$; this is de Finetti's original theorem.

EXAMPLE 2. Let now $\mathscr{X} = \{0, 1, 2, \cdots\}$ be all natural numbers, including 0. Again we want to know all solutions to (7). We have to consider the semigroup $S$ generated by $\{(0, 1), (1, 1), (2, 1), \cdots\}$ inside $\mathbb{N}_0 \times \mathbb{N}_0$, i.e. $S = (\mathbb{N}_0 \times \mathbb{N}) \cup \{(0, 0)\}$.

To determine $S_+^*$ let $\rho$ be a nonnegative semicharacter. Put $v := \rho(0, 1)$ and $w := \rho(1, 1)$. If $v = 0$ then $w^2 = \rho(2, 2) = \rho(2, 1)\, v = 0$ and then $\rho = 1_{\{(0,0)\}}$. For $v \neq 0$ put $u = w/v$; then $\rho(m, 1)(\rho(0, 1))^{m-1} = \rho(m, m) = w^m$ so that $\rho(m, 1) = u^m v$ and $\rho(m, n) = u^m v\, v^{n-1} = u^m v^n$ for all $(m, n) \in S$. We see that $\rho \in W$ if and only if $\sum_{m=0}^{\infty} \rho(m, 1) = \sum_{m=0}^{\infty} u^m v = 1$ iff $u \in [0, 1[$ and $v = 1 - u$, i.e. the extreme solutions of (7) are now given by the geometrical distributions on $\mathbb{N}_0$. Note that $W$ is no longer compact in this case.

EXAMPLE 3. Let again $\mathscr{X} = \mathbb{N}_0$, but this time we choose a nontrivial function $\beta$, namely $\beta(x) = (x!)^{-1}$, hence we are going to determine all $P \in M_+^1(\mathscr{X}^\infty)$ such that

$$P(x_1, \cdots, x_n) = (\prod_{j=1}^n x_j!)^{-1}\varphi_n(\sum_{j=1}^n x_j) = (\prod_{j=1}^n x_j!)^{-1}\varphi(\sum_{j=1}^n (x_j, 1)),$$

where $\varphi$: $S \to \mathbb{R}$ with the same semigroup $S$ as in Example 2. Let $\rho \in S_+^*$ be given by $\rho(m, n) = u^m v^n$, $u, v \in \mathbb{R}_+$. Then $\rho \in W_\beta$ iff

$$1 = \sum_{m=0}^\infty (1/m!)u^m v \Leftrightarrow v = e^{-u}$$

and the corresponding exponential model is just the set of Poisson distributions. See Freedman (1962, Theorem 4) where also the two examples $\beta(m) = \binom{N}{m}$ and $\beta(m) = \binom{N+m-1}{N-1}$ are considered, leading to mixtures of Binomial and Inverse Binomial distributions, respectively.

It may be of some interest to see what it means in this case that all semicharacters in $W_\beta$ are $\alpha$-bounded where

$$\alpha(m, n) = \inf\{\prod_{j=1}^n x_j! \mid \sum_1^n x_j = m\}.$$

Since

$$u^m e^{-un} \le \alpha(m, n) \quad \text{for all} \quad u \ge 0$$

we may maximize the left-hand side to obtain the lower bound

$$(m/n)^m e^{-m} \le \alpha(m, n)$$

(begin nontrivial of course only for $m > n$).

The reader might wonder at this point if it really holds in general that $W_\beta$ is relatively compact, since in the present example the subset $\{(u, e^{-u}) \mid u \in \mathbb{R}_+\} \subseteq \mathbb{R}_+^2$ describing $W_\beta$ is obviously unbounded. The explanation lies in the fact that the identification of $\mathbb{R}_+^2$ with $S_+^*$ is not a topological one (in this case; had we considered $\mathbb{N}_0^2$ instead of $S$, the topologies would coincide, too). In fact $\lim_{u\to\infty}(u, e^{-u}) = (0, 0)$ in the topology of $S_+^*$.

EXAMPLE 4. We let $\mathscr{X} = \mathbb{N}$ and want to determine all $P \in M_+^1(\mathbb{N}^\infty)$ such that

$$P(x_1, \cdots, x_n) = \varphi_n(\max_{1\le j\le n} x_j) = \varphi(\textstyle\sum_{j=1}^n (x_j, 1)) \tag{8}$$

where the notation $\sum_{j=1}^n (x_j, y_j) := (\max_{1\le j\le n} x_j, \sum_{j=1}^n y_j)$ is used in order to avoid misunderstandings. Let $S$ be the subsemigroup of $(\mathbb{N}, \vee) \times (\mathbb{N}_0, +)$ generated by $t(\mathbb{N})$, $t(x) := (x, 1)$. As a set $S$ looks similar to the semigroup considered in the two previous examples, we have $S = (\mathbb{N} \times \mathbb{N}) \cup \{(1, 0)\}$.

Let $\rho \in S_+^*$, then $\rho(m, n) = (\rho(m, 1))^n$ and for $m' \le m$

$$\rho(m', 1)\rho(m, 1) = \rho(m, 2) = (\rho(m, 1))^2;$$

hence $\rho(m', 1) = 0$ implies $\rho(m, 1) = 0$ for all $m \ge m'$, and if $\rho(m, 1) \ne 0$ then

$\rho(m', 1) = \rho(m, 1)$ for all $m' \le m$, so that

$$\rho(m, n) = 1_{\{1,2,\cdots,k\}}(m)v^n$$

for some $k \in \overline{\mathbb{N}}$ and $v \in \mathbb{R}_+$. We see that $\rho \in W$ iff

$$\sum_{x\in\mathscr{X}} \rho(x, 1) = \sum_{x=1}^{\infty} 1_{\{1,\cdots,k\}}(x)v = 1$$

iff $k \in \mathbb{N}$ and $v = 1/k$. The extreme solutions of (8) are therefore given as uniform distributions on one of the discrete intervals $\{1, \cdots, k\}$, $k \in \mathbb{N}$.

REMARK 8. In the situation given by Theorem 4 the statistic $t_n: \mathscr{X}^n \to S$ defined by $t_n(x_1, \cdots, x_n) = \sum_{j=1}^{n} t(x_j)$ is obviously sufficient for the set of all $P \in M_+^1(\mathscr{X}^n)$ fulfilling (6). In fact if $s \in t_n(\mathscr{X}^n)$ and $\varphi(s) > 0$ we have

$$0 < \gamma(s) := \sum_{\sum_{j=1}^{n} t(y_j)=s} \prod_{j=1}^{n} \beta(y_j) < \infty$$

and

$$P(X_1 = x_1, \cdots, X_n = x_n \mid \sum_{j=1}^{n} t(X_j) = s) = \frac{1}{\gamma(s)} \prod_{j=1}^{n} \beta(x_j) 1_{\{t_n=s\}}(x_1, \cdots, x_n),$$

whereas $\varphi(s) = 0$ implies $P(\sum_{j=1}^{n} t(X_j) = s) = 0$.

As another interesting application of Theorem 4 we will derive the general solution of the so-called *Integrated Cauchy Functional Equation* (ICFE) on countable abelian semigroups. This equation was first considered by Deny (1960) and has recently found a lot of interest mainly in connection with probabilistic characterization problems, see for example, Davies (1980), Lau and Rao (1982) and Richards (1982). With the exception of the last mentioned paper the ICFE is only considered on (locally compact) abelian groups; Richards (1982) characterizes the bounded solutions of the ICFE on abelian semigroups.

THEOREM 5. *Let $S$ be a countable abelian semigroup and let $\beta: S \to \mathbb{R}_+$ be a given function such that $\{s \in S \mid \beta(s) > 0\}$ generates $S$. Then $\varphi: S \to \mathbb{R}_+$ is a solution of the* ICFE

$$\varphi(s) = \sum_{s'\in S} \beta(s')\varphi(s + s') \quad \textit{for all} \quad s \in S \tag{9}$$

*if and only if $\varphi$ is a completely positive definite and exponentially bounded function whose representing measure is concentrated on $\{\rho \in S_+^* \mid \sum_{s\in S} \beta(s)\rho(s) = 1\}$.*

PROOF. One direction is, of course, obvious. Assume now that $\varphi$ satisfies (9). If $\varphi(0) = 0$, a moment's reflection shows that $\varphi$ is identically zero. Without restriction we may therefore assume $\varphi(0) = 1$. Let $\mathscr{X} := \{\beta > 0\}$ and define $P \in M_+^1(\mathscr{X}^\infty)$ by

$$P(x_1, \cdots, x_n) := \prod_{j=1}^{n} \beta(x_j)\varphi(\sum_{j=1}^{n} x_j).$$

Then $\sum_{x\in\mathscr{X}} P(x) = \sum_{x\in\mathscr{X}} \beta(x)\varphi(x) = 1$ and for $n \ge 2$

$$\begin{aligned}\sum_{x_n\in\mathscr{X}} P(x_1, \cdots, x_n) &= \prod_{j=1}^{n-1} \beta(x_j) \sum_{x_n\in\mathscr{X}} \beta(x_n)\varphi(\sum_{j=1}^{n-1} x_j + x_n) \\ &= \prod_{j=1}^{n-1} \beta(x_j)\varphi(\sum_{j=1}^{n-1} x_j) = P(x_1, \cdots, x_{n-1})\end{aligned}$$

showing that $P$ indeed is a probability measure on $\mathscr{X}^\infty$. The proof is finished by an application of Theorem 4. □

**6. De Finetti's Theorem for convolution semigroups.** Already long ago Schoenberg (1938) proved the theorem that a continuous function $\varphi$ on $\mathbb{R}_+$ such that $\varphi(\| x \|^2)$ is positive definite on $\mathbb{R}^n$ for all $n \in \mathbb{N}$ ($\| \cdot \|$ denoting Euclidean norm) is necessarily completely monotone. His result can easily be translated into a de Finetti-type theorem: let $X_1, X_2, \cdots$ be a sequence of real-valued random variables such that the characteristic function of $X_{(n)} := (X_1, \cdots, X_n)$ is spherically symmetric for all $n$, i.e.

$$E(e^{i\langle y, X_{(n)}\rangle}) = \varphi_n(\| y \|^2), \quad y \in \mathbb{R}^n, \quad n \in \mathbb{N},$$

then we see immediately that $\varphi_1 = \varphi_2 = \cdots$; let us denote this function $\varphi$. By Schoenberg's theorem

$$\varphi(s) = \int_0^\infty e^{-\lambda s} \, d\mu(s)$$

for a uniquely determined $\mu \in M_+^1(\mathbb{R}_+)$ so that

$$E(e^{i\langle y, X_{(n)}\rangle}) = \int_0^\infty \prod_{j=1}^n e^{-\lambda y_j^2} \, d\mu(\lambda),$$

implying by the unicity of characteristic functions

$$P(X_{(n)} \in B) = \int_0^\infty N(0, 2\lambda)^{\otimes n}(B) \, d\mu(\lambda)$$

for $n$-dimensional Borel sets $B$, and this holding for all $n$ we have in fact

$$P(X \in B) = \int_0^\infty N(0, 2\lambda)^\infty(B) \, d\mu(\lambda)$$

for all Borel sets $B \subseteq \mathbb{R}^\infty$, i.e. $X$ is a variance mixture of centered i.i.d. normal sequences.

Now the family of centered normal distributions on $\mathbb{R}$ with varying variance is in fact a *one-parameter convolution semigroup*, i.e. $\lambda \mapsto N(0, \lambda)$ is a continuous semigroup homomorphism from $(\mathbb{R}_+, +)$ into $M_+^1(\mathbb{R})$ (being considered as a semigroup w.r.t. to convolution) such that $0 \mapsto \varepsilon_0$. Replacing in this definition $(\mathbb{R}_+, +)$ by other topological abelian semigroups we obtain more general convolution semigroups. Using our main theorem we will be able to generalize the argument given above and characterize mixtures of many one-parameter convolution semigroups and some more general convolution semigroups, too. Our starting point will be the following "abstract" result.

THEOREM 6. *Let $R$ and $S$ be abelian semigroups with involution, let $t\colon R \to S$ be a mapping fulfilling $t(0) = 0$, $t(r^*) = (t(r))^*$ and such that $t(R)$ generates $S$. Let further $\beta\colon R \in \mathbb{C} \setminus \{0\}$ fulfill $\beta(0) = 1$, $\beta(r^*) = \overline{\beta(r)}$ and suppose $\varphi\colon S \to \mathbb{C}$ has the property that*

$$\Phi_n(r_1, \cdots, r_n) := \prod_{j=1}^n \beta(r_j)\varphi\left(\sum_{j=1}^n t(r_j)\right) \tag{10}$$

*is bounded and positive definite on $R^n$ for all $n \in \mathbb{N}$. Then $\varphi$ is an exponentially bounded positive definite function on $S$ whose representing measure is concentrated* *on the compact set $W = W_{t,\beta}$ consisting of those semicharacters $\sigma \in S^*$ such that $r \mapsto \beta(r)\sigma(t(r))$ is bounded and positive definite on $R$.*

*In case the functions $\Phi_n$ are even completely positive definite for all $n$ and $\beta$ is positive, $\varphi$ is completely positive definite, too.*

PROOF. Without loss of generality let $\varphi(0) = 1$. Then $\Phi_1, \Phi_2, \cdots$ have the representations

$$\Phi_n(r_1, \cdots, r_n) = \int \prod_{j=1}^n \rho_j(r_j)\, d\nu_n(\rho_1, \cdots, \rho_n) \tag{11}$$

for uniquely determined $\nu_n \in M_+^1(\hat{R}^n)$. Since $\Phi_n(r_1, \cdots, r_{n-1}, 0) = \Phi_{n-1}(r_1, \cdots, r_{n-1})$ the $\nu_n$'s are compatible so that for some $\nu \in M_+^1(\hat{R}^\infty)$ the measures $\nu_n$ are the projections of $\nu$ onto the first $n$ coordinates.

In the terminology of Theorem 1 we put $Z = R$ (with the involution given in $R$) and define $\theta(r)$ as the continuous function $\rho \mapsto \rho(r)$ on $\mathscr{X} = \hat{R}$. (Note that here $\theta$ is one-to-one iff the bounded semicharacters on $R$ separate points.) Instead of (10) we can now write, denoting $X_1, X_2, \cdots$ the coordinate projections on $\hat{R}^\infty$,

$$E[\prod_{j=1}^n \theta(r_j) \circ X_j] = \prod_{j=1}^n \beta(r_j)\varphi(\textstyle\sum_{j=1}^n t(r_j))$$

and Theorem 1 tells us that $\varphi$ is positive definite and exponentially bounded. Hence $\varphi$ has the unique integral representation

$$\varphi(s) = \int \sigma(s)\, d\mu(\sigma), \quad s \in S$$

where $\mu \in M_+^1 S^*)$ is concentrated on $S^\alpha$, the $\alpha$-bounded semicharacters of $S$, with $\alpha$ given by

$$\alpha(s) = \inf\{|\prod_{j=1}^n \beta(r_j)|^{-1} | \textstyle\sum_{j=1}^n t(r_j) = s,\, n \in \mathbb{N},\, r_1, \cdots, r_n \in R\}.$$

From (10) and (11) follows that $\nu$ is exchangeable, hence, invoking Theorem 2, there is a Radon probability measure $\eta$ on $M_+^1(\hat{R})$ such that

$$\nu = \int \kappa^\infty\, d\eta(\kappa)$$

in the sense of (5), which of course extends from indicator functions of Borel subsets of $\hat{R}^\infty$ to (at least) bounded measureable functions on $\hat{R}^\infty$. In particular we get

$$\begin{aligned}
&\int \prod_{j=1}^n \beta(r_j)\sigma(t(r_j))\, d\mu(\sigma) = \prod_{j=1}^n \beta(r_j)\varphi(\textstyle\sum_{j=1}^n t(r_j)) \\
&\quad = \Phi_n(r_1, \cdots, r_n) = \displaystyle\int_{\hat{R}^\infty} \prod_{j=1}^n \rho_j(r_j)\, d\nu(\rho_1, \rho_2, \cdots) \\
&\quad = \int_{M_+^1(\hat{R})} \left[\int_{\hat{R}^\infty} \prod_{j=1}^n \rho_j(r_j)\, d\kappa^\infty(\rho_1, \rho_2, \cdots)\right] d\eta(\kappa) \\
&\quad = \int_{M_+^1(\hat{R})} \prod_{j=1}^n \hat{\kappa}(r_j)\, d\eta(\kappa)
\end{aligned} \tag{12}$$

denoting by $\hat{\kappa}$ the (abstract Laplace) transform of $\kappa$, i.e. $\hat{\kappa}(r) := \int_{\hat{R}} \rho(r)\, d\kappa(\rho)$, $r \in R$.

Let now be given $N \in \mathbb{N}$, $\{c_1, \cdots, c_N\} \subseteq \mathbb{C}$ and $\{a_1, \cdots, a_N\} \subseteq R$. We want to show that $\mu$-almost surely the (continuous) function

$$f(\sigma) := \textstyle\sum_{j,k=1}^{N} c_j\bar{c}_k\beta(a_j + a_k^*)\sigma(t(a_j + a_k^*)), \quad \sigma \in S^\alpha$$

is nonnegative. In order to prove this we consider the complex-valued Radon measure $\xi := f \cdot \mu$ on $S^\alpha$ and use the fact that $\xi \geq 0$ is equivalent with positive definiteness of $\hat{\xi}$ on $S$.

Given $M \in \mathbb{N}$, $\{s_1, \cdots, s_M\} \subseteq S$ and $\{d_1, \cdots, d_M\} \subseteq \mathbb{C}$ we have to show that

$$\textstyle\sum_{p,q=1}^{M} d_p\bar{d}_q\hat{\xi}(s_p + s_q^*) \geq 0.$$

By assumption $s_p = \sum_{m=1}^{m_p} t(r_{p,m})$ for suitable $r_{p,m} \in R$, $p = 1, \cdots, M$. Put $\gamma_p := \prod_{m=1}^{m_p} \beta(r_{p,m})$, then $\{\gamma_1, \cdots, \gamma_M\} \subseteq \mathbb{C}\backslash\{0\}$ and

$$\begin{aligned}
&\textstyle\sum d_q\bar{d}_p\hat{\xi}(s_p + s_q^*) \\
&= \sum\sum d_p\bar{d}_q c_j\bar{c}_k \int \beta(a_j + a_k^*)\sigma(t(a_j + a_k^*)) \\
&\qquad\qquad \cdot \textstyle\prod_{m=1}^{m_p} \sigma(t(r_{p,m})) \prod_{\ell=1}^{m_q} \sigma(t(r_{q,\ell}^*))\, d\mu(\sigma) \\
&= \sum\sum d_p\bar{d}_q \frac{1}{\gamma_p\bar{\gamma}_q} c_j\bar{c}_k \int \beta(a_j + a_k^*)\sigma(t(a_j + a_k^*)) \\
&\qquad\qquad \cdot \textstyle\prod_{\ell=1}^{m_p} \beta(r_{p,m})\sigma(t(r_{p,m})) \prod_{\ell=1}^{m_q} \beta(r_{q,\ell}^*)\sigma(t(r_{q,\ell}^*))\, d\mu(\sigma) \\
&= \sum\sum \frac{d_p}{\gamma_p}\frac{\overline{d_q}}{\gamma_q} c_j\bar{c}_k \int_{M_+^1(\hat{R})} \hat{\kappa}(a_j + a_k^*) \textstyle\prod_{m=1}^{m_p} \hat{\kappa}(r_{p,m}) \prod_{\ell=1}^{m_q} \hat{\kappa}(r_{q,\ell}^*)\, d\eta(\kappa) \\
&= \int_{M_+^1(\hat{R})} \textstyle\sum_{j,k=1}^{N} c_j\bar{c}_k\hat{\kappa}(a_j + a_k^*) \left| \sum_{p=1}^{M} \frac{d_p}{\gamma_p} \prod_{m=1}^{m_p} \hat{\kappa}(r_{p,m}) \right|^2 d\eta(\kappa) \geq 0
\end{aligned}$$

where we used (12). Knowing that $\xi \geq 0$ we find $\mu(f^{-1}(\mathbb{C}\backslash\mathbb{R}_+)) = 0$, and since $\mu$ is a Radon measure we may indeed conclude that $\mu$ concentrates on those $\sigma \in S^\alpha$ for which $\beta(\sigma \circ t)$ is positive definite on $R$. Recalling that $\sigma \in S^\alpha$ means $\sigma \in S^*$ and $|\sigma(s)| \leq \alpha(s)$ for all $s$ we get $|\sigma(t(r))| \leq \alpha(t(r)) \leq 1/|\beta(r)|$ and this shows that $\mu$ is concentrated on the set $W_{t,\beta}$ defined in the statement of this theorem.

The last assertion concerning complete positive definiteness follows immediately from Theorem 1 and Proposition 1. □

Notice that the set $W_{t,\beta}$ on which $\mu$ concentrates in the above Theorem is minimal in the sense that the (Laplace) transform $\varphi$ of any measure $\mu \in M_+(W_{t,\beta})$ leads to functions $\Phi_n$, defined by (10), and being bounded and positive definite on $R^n$ for all $n$. In fact to see this it is enough to consider a one-point measure $\mu = \varepsilon_\sigma$ with $\sigma \in W_{t,\beta}$ and then

$$\textstyle\prod_{j=1}^{n} \beta(r_j)\sigma(\sum_{j=1}^{n} t(r_j)) = \prod_{j=1}^{n} \beta(r_j)\sigma(t(r_j))$$

is positive definite as a finite product of such functions. In particular we can state the following

COROLLARY 2. *Under the conditions of Theorem* 6 *the set of functions* $\varphi\colon S \to \mathbb{C}$ *such that* $\Phi_n$ *as defined in* (10) *is bounded and positive definite for each* $n$, *normalized by* $\varphi(0) = 1$, *is a Bauer simplex whose extreme points are given by the subset* $W_{t,\beta}$ *of* $S^*$.

In case $\beta \equiv 1$ the set $W_{t,\beta} = W_t$ is a $*$-subsemigroup of $\hat{S}$, i.e. a subsemigroup of the bounded semicharacters being closed under complex conjugation. Particularly important is the case that $W_t$ equals $\hat{S}$ in which case the triple $(R, S, t)$ has been called a *Schoenberg triple*, see Berg, Christensen and Ressel (1984, Definition 5.1.4). Another way to state Schoenberg's classical result mentioned earlier is to say that $(\mathbb{R}, \mathbb{R}_+, t)$ with $t(x) = x^2$ is a Schoenberg triple. Here $\mathbb{R}$ is considered as a $*$-semigroup w.r.t. the involution $r^* = -r$.

Let us examine a little closer Schoenberg triples $(R, S, t)$ where $S = \mathbb{R}_+$ (with usual addition). Since $\hat{S} \simeq [0, \infty]$ via the identification of $\lambda \in [0, \infty]$ with the semicharacter $s \mapsto e^{-\lambda s}$ ($\lambda = \infty$ corresponding to $1_{\{0\}}$) we see that the condition $W_t = \hat{S}$ means that $\exp(-\lambda t)$ is positive definite on $R$ for all $\lambda \geq 0$, but this is equivalent with $t$ being a so-called *negative definite* function (with the additional properties $t \geq 0$ and $t(0) = 0$) on the semigroup $R$ which uniquely determines a one-parameter convolution semigroup $(\kappa_\lambda)_{\lambda \geq 0}$ of Radon probability measures on $\hat{R}$, the connection being given by

$$\hat{\kappa}_\lambda(r) = e^{-\lambda t(r)}, \quad \lambda \geq 0, \quad r \in R,$$

see Berg, Christensen and Ressel (1984, Theorem 4.3.7). If now a sequence $X = (X_1, X_2, \cdots)$ of $\hat{R}$-valued random variables (with Radon measures as distributions, to be precise) has the property that the (generalized) Laplace transform of $X_{(n)} = (X_1, \cdots, X_n)$ only depends on the sum $\sum_{j=1}^n t(r_j)$, i.e.

$$E[\textstyle\prod_{j=1}^n X_j(r_j)] = \varphi_n(\sum_{j=1}^n t(r_j))$$

then $\varphi_1 = \varphi_2 = \cdots = \varphi\colon \mathbb{R}_+ \to \mathbb{R}$, $\varphi(0) = 1$, and by Theorem 6

$$\varphi(s) = \int_{[0,\infty]} e^{-\lambda s}\, d\mu(\lambda), \quad s \in \mathbb{R}_+$$

where $\mu \in M_+^1([0, \infty])$. In most of our examples $\varphi$ will be continuous, i.e. $\mu(\{\infty\}) = 0$, implying

$$E[\textstyle\prod_{j=1}^n X_j(r_j)] = \int_0^\infty \prod_{j=1}^n e^{-\lambda t(r_j)}\, d\mu(\lambda)$$

and showing in this way that $X$ is a mixture of i.i.d. sequences distributed according to some $\kappa_\lambda$, $\lambda \geq 0$, i.e.

$$P(X \in B) = \int_0^\infty (\kappa_\lambda)^\infty(B)\, d\mu(\lambda) \quad \text{for all} \quad B \in \mathscr{B}(\hat{R}^\infty).$$

EXAMPLE 5. For $0 < p \leq 2$ the triple $(\mathbb{R}, \mathbb{R}_+, |r|^p)$ is a Schoenberg triple (since $t(r) = |r|^p$ is negative definite in the group sense on the real line). The corresponding one-parameter convolution semigroup $(\kappa_\lambda)_{\lambda\geq 0}$ consists of symmetric stable measures of index $p$; for $p = 1$ this is the Cauchy-semigroup. A sequence $X = (X_1, X_2, \cdots)$ of real-valued random variables is a mixture of $\kappa_\lambda$-distributed i.i.d. sequences iff

$$E(e^{i\sum r_j X_j}) = \varphi(\textstyle\sum |r_j|^p), \tag{13}$$

a result originally due to Bretagnolle, Dacunha-Castelle and Krivine (1966). If $p > 2$ then for no $\lambda > 0$ the function $\exp(-\lambda|r|^p)$ is positive definite (in the usual group sense) on $\mathbb{R}$, hence by Theorem 6 equation (13) can only hold for the trivial sequence $X \equiv 0$.

EXAMPLE 6. (Mixtures of normal distributions with varying mean and varying variance). If $X_1, \cdots, X_n$ are i.i.d. normal random variables with mean $a$ and variance $\sigma^2$ then the characteristic function of $(X_1, \cdots, X_n)$

$$E[\exp(i \textstyle\sum_{j=1}^n r_j X_j)] = \exp(ia \sum_1^n r_j - (\sigma^2/2) \sum_1^n r_j^2)$$

is a function of $(\sum r_j, \sum r_j^2)$, and this would still hold for any mixture of such i.i.d. sequences. We shall see that this property is characteristic for mixtures of normal i.i.d. sequences. We need the following

LEMMA 2. *Let $R$ be an abelian semigroup with involution and define $S := (R \times\, ]0, \infty[) \cup \{(0, 0)\}$ with coordinatewise addition and the involution $(r, y)^* := (r^*, y)$. Let $\sigma \in \hat{S}$ be given. Then either $\sigma = 1_{\{(0,0)\}}$ or the limit*

$$\rho(r) := \lim_{n\to\infty} \sigma(r, y/2^n)$$

*exists for all $r \in R$, is independent of $y \in\, ]0, \infty[$, $\rho \in \hat{R}$ and*

$$\sigma(r, y) = \rho(r) e^{-\lambda y}$$

*for some unique $\lambda \in \mathbb{R}_+$.*

PROOF. Put $\eta(y) := \sigma(0, y)$, $y \in \mathbb{R}_+$; then $\eta \in \hat{\mathbb{R}}_+$ and two cases will be considered separately.

CASE 1. $\eta = 1_{\{0\}}$; then $\sigma(r, y) = \sigma(r, y/2)\sigma(0, y/2)$ and this is zero if $y > 0$, and one if $y$ (and hence $r$, too) is zero, i.e. $\sigma = 1_{\{(0,0)\}}$.

CASE 2. $\eta(y) = e^{-\lambda y}$ for some $\lambda \in \mathbb{R}_+$. In this case

$$\begin{aligned}
\sigma(r, y) &= \sigma(r, y/2)\sigma(0, y/2) = \sigma(r, y/2)e^{-\lambda(y/2)} \\
&= \sigma(r, y/4)\sigma(0, y/4)e^{-\lambda(y/2)} = \sigma(r, y/4)e^{-\lambda(y/2+y/4)} \\
&= \cdots \\
&= \sigma(r, y2^{-n})\exp[-\lambda y(1/2 + 1/4 + \cdots + 2^{-n})],
\end{aligned}$$

therefore the limit $\lim_{n\to\infty} \sigma(r, y2^{-n})$ exists and equals $\sigma(r, y)e^{\lambda y}$. This limit does not depend on $y \in ]0, \infty[$; for let $0 < y' < y$, then, putting $h := y - y'$,

$$\sigma(r, y)e^{\lambda y} = \sigma(r, y' + h)e^{\lambda(y'+h)} = \sigma(r, y')\sigma(0, h)e^{\lambda(y'+h)}$$
$$= \sigma(r, y')e^{-\lambda h}e^{\lambda(y'+h)} = \sigma(r, y')e^{\lambda y'}.$$

Hence $\rho(r) := \sigma(r, y)e^{\lambda y}$ is well-defined and

$$\rho(r_1)\rho(r_2) = \sigma(r_1, y)e^{\lambda y}\sigma(r_2, y)e^{\lambda y} = \sigma(r_1 + r_2, 2y)e^{2\lambda y}$$
$$= \rho(r_1 + r_2)$$

as well as $\rho(r^*) = \sigma(r^*, y)e^{\lambda y} = \sigma((r, y)^*)e^{\lambda y} = \overline{\sigma(r, y)}e^{\lambda y} = \overline{\rho(r)}$, i.e. $\rho \in \hat{R}$ and $\sigma(r, y) = \rho(r)e^{-\lambda y}$ as asserted. □

Consider now the semigroup $S = (\mathbb{R} \times ]0, \infty[) \cup \{(0, 0)\}$ and define $t\colon \mathbb{R} \to S$ by $t(r) := (r, r^2)$; then $(\mathbb{R}, S, t)$ is a Schoenberg triple. In fact the semigroup generated by $t(\mathbb{R})$ contains $\{0\} \times \mathbb{R}_+$ and is then easily seen to be just $S$; and if $\sigma \in \hat{S}$, $\sigma \neq 1_{\{0\}}$, then by the Lemma above $\sigma(r, y) = \rho(r)e^{-\lambda y}$ for some $\lambda \geq 0$ implying $\sigma(t(r)) = \rho(r)e^{-\lambda r^2}$ to be positive definite on $\mathbb{R}$ (and so is of course the limit, as $\lambda \to \infty$, given by $1_{\{0\}} \circ t$).

Suppose $X = (X_1, X_2, \cdots)$ is a sequence of real-valued random variables such that

$$E[\exp(i \textstyle\sum_{j=1}^n r_j X_j)] = \varphi(\sum_{j=1}^n r_j, \sum_{j=1}^n r_j^2)$$

where $\varphi\colon S \to \mathbb{C}$ and $\varphi(0) = 1$. Then by Theorem 6 $\varphi \in \mathscr{P}^b(S)$ and using Lemma 2 $\varphi$ has the representation

$$\varphi(r, y) = \int_{\hat{R}\times[0,\infty]} \rho(r)e^{-\lambda y}\, d\mu(\rho, \lambda)$$

defining a natural extension of $\varphi$ from $S$ to $\mathbb{R} \times \mathbb{R}_+$. Let $\mu_1$, $\mu_2$ denote the two marginals of $\mu$ on $\hat{\mathbb{R}}$ and $[0, \infty]$ respectively. The continuity of

$$\varphi(0, y) = \int_{[0,\infty]} e^{-\lambda y}\, d\mu_2(\lambda) = E[\exp i\sqrt{y/2}\,(X_1 - X_2)]$$

on $\mathbb{R}_+$ leads to $0 = \mu_2(\{\infty\}) = \mu(\hat{\mathbb{R}} \times \{\infty\})$. For fixed $y > 0$ the function $r \mapsto \varphi(r, y)$ is continuous at $r = 0$ since for $r^2 \leq y$

$$\varphi(r, y) = \varphi(r + \sqrt{(y - r^2)/2} - \sqrt{(y - r^2)/2}, r^2 + (y - r^2)/2 + (y - r^2)/2)$$
$$= E[\exp i\{rX_1 + \sqrt{(y - r^2)/2}\,(X_2 - X_3)\}],$$

hence—being positive definite—this function is continuous everywhere. Now

$$|\varphi(r, 0) - \varphi(r, 1/n)| = \left| \int \rho(r)(1 - e^{-\lambda/n})\, d\mu \right| \leq \int_0^\infty (1 - e^{-\lambda/n})\, d\mu_2(\lambda)$$

which converges to 0 as $n \to \infty$, uniformly in $r \in \mathbb{R}$, so that $\varphi(r, 0)$ is a continuous positive definite function on $\mathbb{R}$, hence by Bochner's theorem $0 = \mu_1(\hat{\mathbb{R}}\backslash\mathbb{R}) = \mu((\hat{\mathbb{R}}\backslash\mathbb{R}) \times [0, \infty])$ showing that $\mu$ is indeed concentrated on $\mathbb{R} \times \mathbb{R}_+$, i.e.

$$\varphi(r, y) = \int_{\mathbb{R}\times\mathbb{R}_+} e^{iar}e^{-\lambda y^2}\, d\mu(a, \lambda)$$

and

$$E[\exp(i \textstyle\sum_{j=1}^n r_jX_j)] = \displaystyle\int \textstyle\prod_{j=1}^n e^{iar_j}e^{-\lambda r_j^2}\, d\mu(a, \lambda);$$

we see that $X$ is in fact a mixture of normal i.i.d. sequences. Note that we were dealing again with a convolution semigroup, namely with

$$\mathbb{R} \times \mathbb{R}_+ \to M_+^1(\mathbb{R}) \qquad (a, \lambda) \mapsto N(a, \lambda)$$

which is some kind of a two-parameter convolution semigroup.

EXAMPLE 7. (Mixtures of normal distributions with varying mean and fixed variance). The natural guess in this case is that

$$E[\exp(i \textstyle\sum_{j=1}^n r_jX_j)] = \exp(-(\sigma_0^2/2) \sum_1^n r_j^2)\varphi(\sum_1^n r_j)$$

should be the necessary and sufficient condition. Putting in Theorem 6 $R = S = \mathbb{R}$, $t = id$, $\beta(r) = \exp(-(\sigma_0^2/2)r^2)$ we get $\varphi(r) = \int e^{iar}\, d\mu(a)$ where $\mu$ may be any probability distribution on $\mathbb{R}$ (since $\beta(r)e^{iar}$ is positive definite for all $a \in \mathbb{R}$), showing the above guess to be correct. Note that $(N(a, \sigma_0^2))_{a\in\mathbb{R}}$ is not a convolution semigroup.

EXAMPLE 8. (Mixtures of gamma-distributions). The exponential distribution $e^{-x}dx$ on $\mathbb{R}_+$ is infinitely divisible and determines the one-parameter convolution semigroup $(\kappa_\lambda)_{\lambda\ge0}$ where $\kappa_\lambda$ has the density $(\Gamma(\lambda))^{-1}x^{\lambda-1}e^{-x}$ on $\mathbb{R}_+$. A sequence of nonnegative random variables $X_1, X_2, \cdots$ is a mixture of $\kappa_\lambda$-distributed i.i.d. sequences iff

$$E[\exp(- \textstyle\sum_{j=1}^n r_jX_j)] = \varphi(\prod_{j=1}^n (1 + r_j))$$

for all $n$. This follows since $((\mathbb{R}_+, +), ([1, \infty[\,, \cdot), 1 + r)$ is a Schoenberg triple, noting that $\kappa_\lambda$ has Laplace transform $(1 + t)^{-\lambda}$.

EXAMPLE 9. (Mixtures of Poisson distributions, a second characterization). Besides the necessary and sufficient condition of Example 3 above, using the point probabilities, there is another one using generating functions. Since it is easily established that $(([-1, 1], \cdot), (\mathbb{R}_+, +), 1 - r)$ is a Schoenberg triple, a sequence $X_1, X_2, \cdots$ of nonnegative integer-valued random variables is a mixture of Poisson distributed i.i.d. sequences if and only if the generating function of $X_1, \cdots, X_n$ has the form

$$E(\textstyle\prod_{j=1}^n r_j^{X_j}) = \varphi(\sum_{j=1}^n (1 - r_j))$$

for all $n \in \mathbb{N}$, $r_1, \cdots, r_n \in [-1, 1]$, and some function $\varphi$ on $\mathbb{R}_+$.

EXAMPLE 10. (Mixtures of normal distributions on the torus). Let $T := \{z \in \mathbb{C} \mid |z| = 1\}$ be the torus group. A probability measure $\kappa$ on $T$ is *normal* iff its Fourier coefficients are given as $\hat{\kappa}(k) = a^k \exp(-\lambda k^2)$ for $k \in \mathbb{Z}$, where $a \in T$ and $\lambda > 0$; see for example Berg (1975). Hence $\kappa$ is symmetric iff $a = \pm 1$. Since $(\mathbb{Z}, \mathbb{N}_0, t(k) = k^2)$ is a Schoenberg triple (see Berg and Ressel, 1978) a sequence $X_1, X_2, \cdots$ of $T$-valued random variables is a mixture of symmetric normal i.i.d. sequences iff its Fourier coefficients have the form

$$E(\prod_{j=1}^n X_j^{k_j}) = \varphi(\sum_{j=1}^n k_j^2)$$

for all $n \in \mathbb{N}$, $k_1, \cdots, k_n \in \mathbb{Z}$. It seems likely that mixtures of the whole normal family on the torus are characterized by writing instead $\varphi(\sum k_j, \sum k_j^2)$, but we have not carried out the details.

EXAMPLE 11. (Mixtures of exponential distributions). If $X$ is a nonnegative i.i.d. sequence of exponentially distributed random variables (say with parameter $\lambda \geq 0$) then the common survival function of $X_1, \cdots, X_n$

$$P(X_1 \geq a_1, \cdots, X_n \geq a_n) = \exp(-\lambda \sum_1^n a_j)$$

depends only on the sum $\sum a_j$. We shall see that this is a characteristic property. Suppose $X_j \geq 0$ and

$$P(X_1 \geq a_1, \cdots, X_n \geq a_n) = \varphi_n(\sum_{j=1}^n a_j)$$

for all $n$, then again $\varphi_1 = \varphi_2 = \cdots = \varphi$ and in order to see that $\varphi$ then necessarily is the (ordinary) Laplace transform of some measure on the half-line, we have to look at the semigroup $(\mathbb{R}_+, \vee)$ w.r.t. the composition $r \vee r' = \max\{r, r'\}$ (the involution being the identical). It is easy to see that positive definiteness for a function $f: (\mathbb{R}_+, \vee) \to \mathbb{R}$ is equivalent with $f$ being nonnegative and decreasing (see Berg, Christensen and Ressel, 1984, Proposition 4.4.18). From this a moment's thought makes it clear that $((\mathbb{R}_+, \vee), (\mathbb{R}_+, +), id)$ defines a Schoenberg triple, and in order to apply Theorem 6 we have to make sure that $(a_1, \cdots, a_n) \mapsto \varphi(\sum_1^n a_j)$ is positive definite on $(\mathbb{R}_+, \vee)^n$. Let $\nu_n$ be the distribution of $(X_1, \cdots, X_n)$. Then

$$\begin{aligned}\varphi(\sum_{j=1}^n a_j) &= \nu_n([a_1, \infty[ \times \cdots \times [a_n, \infty[) \\ &= \int \prod_{j=1}^n 1_{[a_j,\infty[}(x_j)\, d\nu_n(x_1, \cdots, x_n) \\ &= \int \prod_{j=1}^n 1_{[0,x_j]}(a_j)\, d\nu_n(x_1, \cdots, x_n)\end{aligned}$$

and the assertion follows since all the indicator functions $\prod 1_{[0,x_j]}$ are semicharacters of $(\mathbb{R}_+, \vee)^n$. Therefore $\varphi \in \mathscr{P}_1^b(\mathbb{R}_+, +)$ implying $\varphi(s) = \int e^{-\lambda s}\, d\mu(s)$ for some $\mu \in M_+^1([0, \infty])$. Note that $\mu(\{\infty\}) > 0$ might well occur here, corresponding to the possibility that the $X_j$ are 0 with positive probability. Note also that $\lambda \mapsto$ "exponential distribution with parameter $\lambda$" is a one-parameter convolution semigroup in $M_+^1([0, \infty], \wedge)$, $\lambda = 0$ corresponding to $\varepsilon_\infty$.

EXAMPLE 12. (Mixtures of Marshall-Olkin distributions). Similar to the one-dimensional case the multivariate survival-function of an $n$-dimensional random vector $X$ or its distribution $\nu$ is defined by

$$\varphi(x) = P(X_1 \geq x_1, \cdots, X_n \geq x_n) = \nu(\textstyle\prod_{j=1}^n [x_j, \infty[), \quad x \in \mathbb{R}^n.$$

It is immediately clear that $\nu \in M_+^1(\mathbb{R}^n)$ is uniquely determined by its survival function. The bivariate Marshall-Olkin distribution can be defined therefore by its survival function which has the form

$$\varphi(x, y) = \exp\{-[\alpha x + \beta y + \gamma(x \vee y)]\}, \quad x, y \geq 0, \tag{14}$$

where $\alpha$, $\beta$, $\gamma$ are nonnegative parameters; this distribution was introduced by Marshall and Olkin (1967). We shall need the following

LEMMA 3. *If $\varphi$ as defined in* (14) *is a survival function then necessarily* $\{\alpha, \beta, \gamma\} \subseteq \mathbb{R}_+$.

PROOF. Choose $y, r, x \in \mathbb{R}_+$ such that $y + r \leq x$, then, denoting $X$, $Y$ random variables with $\varphi(x, y) = P(X \geq x, Y \geq y)$, we see from

$$P(X \geq x, Y \geq y + r) = e^{-\alpha x - \beta(y+r) - \gamma x} \leq P(X \geq x, Y \geq y) = e^{-\alpha x - \beta y - \gamma x}$$

that $e^{-\beta r} \leq 1$, hence $\beta \geq 0$; similarly $\alpha \geq 0$. For $x > 0$ we have

$$0 \leq P((X, Y) \in [x, 2x] \times [x, 2x]) = \varphi(x, x) + \varphi(2x, 2x) - \varphi(x, 2x) - \varphi(2x, x)$$

or

$$e^{-(\alpha+\beta+\gamma)x} + e^{-2(\alpha+\beta+\gamma)x} \geq e^{-(\alpha+2\beta+2\gamma)x} + e^{-(2\alpha+\beta+2\gamma)x}$$

or

$$e^{(\alpha+\beta+\gamma)x} + 1 \geq e^{\alpha x} + e^{\beta x}$$

with equality for $x = 0$. Taking on both sides the derivative at 0 gives $\alpha + \beta + \gamma \geq \alpha + \beta$, hence $\gamma \geq 0$. □

Let now $Z = (Z_1, Z_2, \cdots)$ be a sequence of $\mathbb{R}_+^2$-valued random vectors $Z_j = (X_j, Y_j)$ and suppose that the common survival function of $Z_1, \cdots, Z_n$ depends only on $\sum x_j$, $\sum y_j$ and $\sum x_j \vee y_j$, i.e.

$$P(X_i \geq x_i, Y_i \geq y_i, 1 \leq i \leq n) = \varphi(\textstyle\sum_{i=1}^n (x_i, y_i, x_i \vee y_i)). \tag{15}$$

We claim that $Z$ must then be a mixture of i.i.d. sequences with Marshall-Olkin distributions. Of course we shall again apply Theorem 6. Let $R := (\mathbb{R}_+^2, \vee)$, i.e. we use the "addition" $(x_1, y_1) \vee (x_2, y_2) = (x_1 \vee x_2, y_1 \vee y_2)$, let $\beta \equiv 1$, define $t$: $\mathbb{R}_+^2 \to \mathbb{R}_+^3$ by $t(x, y) := (x, y, x \vee y)$ and let $S$ be the semigroup generated by $t(\mathbb{R}_+^2)$; on $\mathbb{R}_+^3$ we use the usual addition and in both cases the identical involution. It is easily seen that

$$S = \{(x, y, z) \in \mathbb{R}_+^3 \mid x \vee y \leq z \leq x + y\}.$$

Similar to the preceding example we derive from (15)

$$\varphi(\sum_1^n (x_i, y_i, x_i \vee y_i)) = \int_{\mathbb{R}_+^{2n}} \prod_1^n 1_{[0,u_i]}(x_i) 1_{[0,v_i]}(y_i)\, d\nu_n(u_1, v_1, \cdots, u_n, v_n),$$

$\nu_n$ being the distribution of $(Z_1, \cdots, Z_n)$, and this shows that $\varphi(\sum_1^n t(r_j))$ is positive definite on $R^n$ for all $n$. Since $\varphi$ is bounded we get $\varphi(s) = \int \rho(s)\, d\mu(s)$ with $\mu$ concentrated on those $\sigma \in \hat{S}_+$ for which $\sigma \circ t \in \mathscr{P}_1^b(R)$. (Since $S$ is 2-divisible, we have in fact $\hat{S}_+ = \hat{S}$; and since $R$ is idempotent, $\mathscr{P}_1^b(R) = \mathscr{P}_1(R)$.) It is not difficult to find that $\hat{S}$ is parametrized by the "cone"

$$\mathscr{C} = \{(u, v, w) \in (\mathbb{R} \cup \{\infty\})^3 \mid u + w \geq 0,\ v + w \geq 0,\ u + v + w \geq 0\}$$

via the usual identification of $(u, v, w)$ with the semicharacter $\sigma(x, y, z) = \exp[-(ux + vy + wz)]$. Note that $\mathscr{C}$ contains vectors with negative components, for example $(-1, -2, 3) \in \mathscr{C}$. Note also that the parametrization is not one-to-one on the "boundary", i.e. on those $(u, v, w) \in \mathscr{C}$ where at least one coordinate is infinite. Since any left continuous function in $\mathscr{P}_1^b(R)$ is a survival function, Lemma 3 implies that only vectors $(u, v, w) \in \mathscr{C}$ with nonnegative components lead to a survival function $\sigma \circ t$. Hence

$$\varphi(x, y, z) = \int_{[0,\infty]^3} \exp[-(ux + vy + wz)]\, d\mu(u, v, w)$$

for some probability measure $\mu$ on $[0, \infty]^3$, and together with (15) we see that $Z$ indeed is a mixture of Marshall-Olkin distributed i.i.d. sequences. Several "degenerate" cases are possible in this mixture; the vector $(\infty, v, w) \in [0, \infty]^3$ for example corresponds to $(0, Y)$ where $Y$ is exponentially distributed (if $0 < v + w < \infty$), the vector $(0, 0, w)$ with $0 < w < \infty$ corresponds to an exponential distribution concentrated on the diagonal in $\mathbb{R}_+^2$, and all the vectors $(u, v, \infty)$ in $\mathscr{C}$ lead to $(X, Y) = (0, 0)$.

**Acknowledgement.** Thanks are due to Persi Diaconis and to Steffen Lauritzen for many stimulating discussions, in particular during our special seminar at the Department of Statistics in Stanford.

## REFERENCES

ALDOUS, D. J. (1983). Exchangeability and related topics. Ecole d'été St. Flour. *Springer Series Lecture Notes in Mathematics.* To appear.

BERG, C. (1975). On the relation between Gaussian measures and convolution semigroups of local type. *Math. Scand.* **37** 183–192.

BERG, C., CHRISTENSEN, J. P. R. and JENSEN, C. U. (1979). A remark on the multidimensional moment problem. *Math. Ann.* **243** 163–169.

BERG, C., CHRISTENSEN, J. P. R. and RESSEL, P. (1984). *Harmonic Analysis on Semigroups.* Graduate Texts in Mathematics, Vol. 100, Springer-Verlag.

BERG, C., and MASERICK, P. H. (1984). Exponentially bounded positive definite functions. *Illinois J. Math.* **28** 162–171.

BERG, C., and RESSEL, P. (1978). Une forme abstraite du théorème de Schoenberg. *Arch. Math.* **30** 55–61.

BRETAGNOLLE, Y., DACUNHA-CASTELLE, D. and KRIVINE, Y. L. (1966). Lois stables et espaces $L^p$. *Ann. Inst. H. Poincaré*, Sect. B II **3** 231–259.

CHRISTENSEN, J. P. R. and RESSEL P. (1982). Positive definite kernels on the complex Hilbert sphere. *Math. Z.* **180** 193–201.

DAVIES, L. (1981). A theorem of Deny with applications to characterization problems. In: *Lecture Notes in Mathematics* **861** (eds. Dugué, Lukacs and Rohatgi), 35–41. Springer Verlag, Berlin.

DE FINETTI (1931). Funzione caratteristica di un fenomeno allatorio. *Atti della R. Accademia Nazionale dei Lincii Ser. 6, Memorie, Classe di Scienze, Fisiche, Matematiche e Naturali, Vol.* 4, 251–299.

DENY, J. (1960). Sur l'équation de convolution $\mu = \mu * \sigma$. Séminaire Brelot-Choquet-Deny (Théorie du Potentiel) 4[e] année, n° 5.

DIACONIS, P. and FREEDMAN, D. (1980). De Finetti's theorem for Markov chains. *Ann. Probab.* **8** 115–130.

DIACONIS, P. and FREEDMAN, D. (1984). Partial exchangeability and sufficiency. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on "Statistics: Applications and New Directions", Calcutta 1981 (to appear in Sankhyā A).

DUBINS, L. and FREEDMAN, D. (1979). Exchangeable processes need not be mixtures of independent identically distributed random variables. *Z. Wahrsch. verw. Gebiete* **48** 115–132.

FREEDMAN, D. (1962). Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Statist.* **33** 916–923.

HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.

KINGMAN, J. (1978). Uses of exchangeability. *Ann. Probab.*, **6**, 183–197.

LAU, K. S. and RAO, C. R. (1982). Integrated Cauchy functional equation and characterizations of the exponential law. *Sankhyā* A **44** 72–90.

LAURITZEN, S. L. (1975). General exponential models for discrete observations. *Scand. J. Statist.* **2** 23–32.

LAURITZEN, S. L. (1982). *Statistical Models as Extremal Families.* Aalborg University Press.

LAURITZEN, S. L. (1984). Extreme point models in statistics. *Scand. J. Statist.* **11** 65–91.

MARSHALL, A. W. and OLKIN, I. (1967). A multivariate exponential distribution. *JASA* **62** 30–44.

RESSEL, P. (1976). A short proof of Schoenberg's theorem. *Proc. Amer. Math. Soc.* **57** 66–68.

RESSEL, P. (1977). Some continuity and measurability results on spaces of measures. *Math. Scand.* **40** 69–78.

RESSEL, P. (1982). A general Hoeffding type inequality. *Z. Wahrsch. verw. Gebiete* **61** 223–235.

RICHARDS, D. St. (1982). *Exponential Distributions on Partially Ordered Abelian Semigroups.* University of N.C. at Chapel Hill, Inst. of Statistics Mimeo Series, No. 1387.

SCHOENBERG, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.* **44** 522–536.

MATH.-GEOGR. FAKULTÄT
RESIDENZPLATZ 12
D-8078 EICHSTÄTT
W. GERMANY