Math. Z. 247, 201–212 (2004) Mathematische Zeitschrift

DOI: 10.1007/s00209-003-0636-6

The Markov moment problem and de Finetti's theorem: Part II

Persi Diaconis[1], **David Freedman**[2]

[1] Department of Mathematics and of Statistics, Stanford University, Stanford, CA 94305, USA

[2] Department of Statistics and of Mathematics, University of California, Berkeley, CA 94720-3860, USA (e-mail: freedman@stat.berkeley.edu)

Received: 1 February 2003; in final form: 18 June 2003 /
Published online: 14 January 2004 –

Abstract This paper gives an abstract version of de Finetti's theorem that characterizes mixing measures with L_p densities. The general setting is reviewed; after the theorem is proved, it is specialized to coin tossing and to exponential random variables. Laplace transforms of bounded densities are characterized, and inversion formulas are discussed.

Introduction

In part I of this paper, we discussed the Hausdorff moment problem on the unit interval, and explained how such problems can be translated into questions about the prior or "mixing" measure in Bayesian statistics. Our object here is to give a version of de Finetti's theorem that characterizes mixing measures with L_p densities, in the general setting described by Diaconis and Freedman (1984), which covers "partial exchangeability." We begin by reviewing the setup and proving general theorems; then we give some examples, showing how the general theory specializes to normal variables, coin tossing, and exponential variables. In connection with the latter, we characterize Laplace transforms of bounded densities and discuss inversion formulas. As will be seen, the abstract theory gives a generalized procedure for inverting probability transforms. Finally, there is a brief literature review. Theorems 2–4 and their corollaries are thought to be new.

The abstract setup can be described as follows. For $i = 1, 2, \ldots$, let Ω_i be a Polish space equipped with the Borel σ-field $\mathcal{F}_i$. Let $\Omega = \prod_{i=1}^{\infty} \Omega_i$ and $\mathcal{F} = \prod_{i=1}^{\infty} \mathcal{F}_i$. Let X_i be the ith coordinate function on Ω. The nth "sufficient statistic" T_n is a Borel mapping from $\prod_{i=1}^{n} \Omega_i$ to a Polish space W_n equipped with its Borel σ-field $\mathcal{B}_n$. In principle, T_n does not act on Ω, although $T_n(X_1, \ldots, X_n)$ does. For each n and $t \in W_n$, let $Q_{n,t}$ be a probability on $\left(\prod_{i=1}^{n} \Omega_i, \prod_{i=1}^{n} \mathcal{F}_i\right)$. It is assumed that $t \to Q_{n,t}$ is Borel.

To illustrate the setup, suppose the X_i are independent normal random variables with common mean 0 and variance $\sigma^2 > 0$. Then Ω_i would be the real line, W_n would be the set of positive real numbers, and $T_n(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i^2$. In this example, $Q_{n,t}$ is uniform on the n-tuples of real numbers $(x_1, \ldots, x_n)$ with $\sum_{i=1}^{n} x_i^2 = t$. Geometrically, this set of n-tuples is the sphere centered at 0 having radius $\sqrt{t}$. Statistically, $Q_{n,t}$ is the conditional distribution of the sample, given the sufficient statistic.

We return to the abstract setting, and define M_Q, the partially exchangeable probabilities, as the set of P on $(\Omega, \mathcal{F})$ such that for each n, given $T_n(X_1, \ldots, X_n) = t$, a regular conditional P-distribution for $X_1, \ldots, X_n$ is $Q_{n,t}$. Informally, $Q_{n,t}$ is the distribution of the data given that the sufficient statistic took the value t. This does not depend on the parameters, i.e., is the same for all $P \in M_Q$. Said another way, M_Q is the set of P for which $Q_{n,t}$ works as advertised. In our normal example, M_Q will turn out to be the set of probability distributions faced by a textbook Bayesian statistician, who is going to observe (by assumption) a sequence of independent normal random variables with mean 0 and variance $\sigma^2 > 0$, and is contemplating all possible prior probabilities for σ^2. That is the content of Theorem 1 below.

We impose the following regularity conditions (which are obvious for the normal, once you decipher the notation).

(1) $Q_{n,t}\{T_n = t\} = 1$.
(2) If $T_n(x_1, \ldots, x_n) = T_n(x_1', \ldots, x_n')$ then $T_{n+1}(x_1, \ldots, x_n, x)$ $= T_{n+1}(x_1', \ldots, x_n', x)$ for all $x \in \Omega_{n+1}$.
(3) For each $s \in W_n$ and $t \in W_{n+1}$, relative to $Q_{n+1,t}$, the kernel $Q_{n,s}$ is a regular conditional distribution for $(X_1, \ldots, X_n)$ given $T_n(X_1, \ldots, X_n) = s$ and $X_{n+1} = x$. Here, the X_i are viewed as the coordinate functions on $\prod_{i=1}^{n+1} \Omega_i$.

We define the partially exchangeable σ-field $\hat{\Sigma}$ as

$$\hat{\Sigma} = \bigcap_{n=1}^{\infty} \hat{\Sigma}(n),$$

where $\hat{\Sigma}(n)$ is spanned by $T_n(X_1, \ldots, X_n), X_{n+1}, X_{n+2}, \ldots$. The main theorem proved in Diaconis and Freedman (1984) is the following.

Theorem 1. *Conditions* (1), (2), *and* (3) *are in force. Then M_Q is convex, and there is a set $G \in \hat{\Sigma}$ with the following properties.*

(i) $P(G) = 1$ *for all* $P \in M_Q$.
(ii) *For each $\omega \in G$, the sequence of probabilities $Q_{n,T_n(\omega)}$ converges weak-star to a limiting probability $Q_\omega \in M_Q$, which is* 0–1 *on* $\hat{\Sigma}$.
(iii) *As ω ranges over G, the kernels Q_ω range over the extreme points of M_Q.*
(iv) *For any $P \in M_Q$, the kernel Q_ω is a regular conditional P-distribution for $X_1, X_2, \ldots$ given $\hat{\Sigma}$, and*

$$P = \int_G Q_\omega \, \hat{P}(d\omega), \tag{4}$$

with $\hat{P}$ *the restriction of* P *to* $\hat{\Sigma}$. *The representation* (4) *is unique, i.e.,* $\hat{P} \leftrightarrow P$. *Moreover,* P *is extreme iff* P *is* 0–1 *on* $\hat{\Sigma}$, *i.e.,*

$$P\{\omega : \omega \in G \,\&\, Q_\omega = P\} = 1. \tag{5}$$

Remark. (i) The σ-field $\hat{\Sigma}$ may be restricted even further, to the σ-field $\check{\Sigma}$ spanned by $\omega \to Q_\omega$. Then $\hat{P}$ is replaced by $\check{P}$, the restriction of P to $\check{\Sigma}$, the advantage being that $\check{\Sigma}$ is a Borel σ-field equivalent to the inseparable $\hat{\Sigma}$ up to sets that have measure 0 for all $P \in M_Q$.

(ii) In this context, P is the mixture and $\check{P}$ is the mixing measure. Equation (4) becomes

$$P = \int_G Q_\omega \, \check{P}(d\omega). \tag{6}$$

(iii) If we identify all points in the same atom of $\check{\Sigma}$, the resulting quotient space $\mathcal{X}$ is analytic. The quotient of $\check{\Sigma}$ is the Borel σ-field in $\mathcal{X}$ and the quotient π of $\check{P}$ is a probability on that σ-field. Then

$$P = \int_{\mathcal{X}} Q_x \, \pi(dx). \tag{7}$$

This may be a more convincing analog to de Finetti's theorem for coin-tossing. To define Q_x, choose any ω in the fiber corresponding to x – it doesn't matter which – and set $Q_x = Q_\omega$. Let X map ω in G to $x \in \mathcal{X}$, so that $\pi = PX^{-1}$, i.e., π is the limiting distribution of the random measures Q_{n,T_n}. Among other things, the mixing measure has been recovered from the mixture. This is a generalized inversion formula; the application to Laplace transforms will be detailed below. Technically, $\mathcal{X}$ can be realized as the image of G under X, and is then an analytic subset of the set of probabilities on Ω; in applications, $\mathcal{X}$ will be a homelier object.

(iv) In our normal example, G can be taken as the set where $(X_1^2 + \cdots + X_n^2)/n$ converges to a finite positive limit L. Then Q_ω makes the coordinates independent normal random variables, with variance $L(\omega)$. The quotient space $\mathcal{X}$ in (iii) is $(0, \infty)$, and the quotient probability π is the prior on σ^2, viz., the distribution of L. Said with less formality, $X_1, X_2, \ldots$ have an orthogonally invariant distribution iff they are scale mixtures of independent normal variables with mean 0 and variance 1. (It is the distributions that are being mixed, not the random variables; the customary informal language is, well, informal.)

In view of Theorem 1(iii) and (5),

$$Q_{\omega'}\{\omega : \omega \in G \,\&\, Q_\omega = Q_{\omega'}\} = 1 \text{ for all } \omega' \in G, \tag{8}$$

an equation that will be used later. Condition (1) implies that $t \to Q_{n,t}$ is 1–1. Hence T_n and Q_{n,T_n} span the same σ-field, and we may view Q_{n,T_n} as the sufficient statistic instead of T_n.

Bounded densities

Our first result characterizes mixtures where the mixing measure has a bounded density. It is the abstract version of Theorem 4 for L_∞ in Part I. Let $P_i \in M_Q$ for $i = 0, 1$, let n_0 be a positive integer, and let c be a positive constant. Conditions (1), (2), and (3) are in force; we use the notation of Theorem 1.

Theorem 2. *Let $P^{(n)}$ be the restriction of P to the σ-field spanned by $X_1, X_2, \ldots, X_n$, and let $U_n = T_n(X_1, \ldots, X_n)$ map Ω to W_n. The following conditions are equivalent.*

(i) $\check{P}_1 \leq c\check{P}_0$.
(ii) $P_1 \leq cP_0$.
(iii) $P_1^{(n)} \leq cP_0^{(n)}$ *for all* $n = n_0, n_0 + 1, \ldots$.
(iv) $P_1 U_n^{-1} \leq cP_0 U_n^{-1}$ *for all* $n = n_0, n_0 + 1, \ldots$.

Proof. Plainly, (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv), the first implication being immediate from (6). Next, (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). The first implication results from sufficiency: if $P \in M_Q$, then

$$P^{(n)} = \int_{W_n} Q_{n,t}\, PU_n^{-1}(dt). \tag{9}$$

The proof of Theorem 2 is complete. From the present perspective, if condition (iii) holds for any n_0, it plainly holds for $n_0 = 1$; this will be helpful in one of the applications below, where the dependence of the conditions on n_0 will be less transparent. Theorem 2 characterizes mixtures with bounded densities, and in the next section, we turn to L_p densities. □

L_p densities

Theorem 3. $P_1 << P_0$ *iff* $\check{P}_1 << \check{P}_0$, *and then* $d\check{P}_1/d\check{P}_0$ *is a version of* dP_1/dP_0.

Proof. If $\check{P}_1 << \check{P}_0$, then $P_1 << P_0$ by (6). The converse is obvious from the fact that $\check{P}$ restricts P to a smaller σ-field. To compute the Radon-Nikodym derivative, suppose $\check{P}_1 << \check{P}_0$. Let $\phi = d\check{P}_1/d\check{P}_0$, and fix $A \in \mathcal{F}$. Then

$$\int_A \phi\, dP_0 = \int_G \left(\int_A \phi(\omega)\, Q_{\omega'}(d\omega) \right) \check{P}_0(d\omega'). \tag{10}$$

But $Q_{\omega'}$ concentrates on the $\check{\Sigma}$-atom containing ω' by (8), and ϕ is $\check{\Sigma}$-measurable, so

$$Q_{\omega'}\{\phi = \phi(\omega')\} = 1,$$

and we may replace $\phi(\omega)$ on the right side of (10) by $\phi(\omega')$. Thus

$$\int_A \phi\, dP_0 = \int_G \left(\int_A \phi(\omega')\, Q_{\omega'}(d\omega) \right) \check{P}_0(d\omega')$$

$$
\begin{aligned}
&= \int_G \left(\int_A Q_{\omega'}(d\omega) \right) \phi(\omega') \check{P}_0(d\omega') \\
&= \int_G Q_{\omega'}(A) \check{P}_1(d\omega') \\
&= P_1(A).
\end{aligned}
$$

□

Corollary 1. *Let $\phi = d\check{P}_1/d\check{P}_0$, with $\phi = \infty$ on the part of the space where $\check{P}_1$ is singular with respect to $\check{P}_0$. Define Φ in the analogous way for P_0 and P_1. Then $\Phi = \phi$ a.e. $[P_0 + P_1]$.*

Proof. This is immediate from Theorem 3, on replacing P_0 by $\frac{1}{2}(P_0 + P_1)$. In principle, Φ need only be $\mathcal{F}$-measurable; in fact, however, Φ is $\check{\Sigma}$-measurable up to null sets. □

Recall that $U_n = T_n(X_1, \ldots, X_n)$. Suppose

$$P_1 U_n^{-1} << P_0 U_n^{-1} \text{ for all } n. \tag{11}$$

Let $h_n = dP_1 U_n^{-1}/dP_0 U_n^{-1}$, a Borel function on W_n. Let

$$c_n = \left(\int_{W_n} h_n^p \, dP_0 U_n^{-1} \right)^{1/p} \tag{12}$$

and

$$H_n = h_n(U_n). \tag{13}$$

Recall that $P^{(n)}$ is the restriction of P to the σ-field spanned by $X_1, \ldots, X_n$, so that $P_1^{(n)} << P_0^{(n)}$ by (9) and (11); and $H_n = dP_1^{(n)}/dP_0^{(n)}$ is a martingale relative to P_0. The proof of the next theorem is omitted as a routine application of differentiation theory (Hewitt and Stromberg, 1969, pp. 369–75).

Theorem 4. *Assume* (11)*, and definitions (*12–13*). Then c_n is non-decreasing as n increases. Moreover, H_n converges a.e. $P_0 + P_1$ to a limit H, which is infinite on the part of the space where P_1 is singular with respect to P_0, and dP_1/dP_0 on the part of the space where $P_1 << P_0$. Finally,* $\lim c_n = (\int H^p \, dP_0)^{1/p}$.

Corollary 2. *$P_1 << P_0$ with an L_p density having norm at most c iff* $\sup c_n \leq c$.

Remark. (i) $P_1 << P_0$ iff H_n is uniformly P_0-integrable; but this amounts to little more than restating the definition of absolute continuity.

(ii) Corollary 2 is the abstract version of Theorem 9 in Part I. Theorems 2 and 3 here capture the reasoning for Theorem 4 in Part I.

Examples

Example 1. Coin tossing. To make contact with de Finetti's original result for coin tossing – Theorem 5 in Part I – let $\Omega_i = \{0, 1\}$ and $W_n = \{0, \ldots, n\}$, with the discrete topology on both. Let $T_n(x_1, \ldots, x_n) = x_1 + \cdots + x_n$. Informally, 1 is heads, 0 is tails; $\omega \in \Omega$ is the record of an infinite number of coin tosses, $X_i(\omega)$ being the outcome on the ith toss; $T_n(X_1, \ldots, X_n)$ is the number of heads in the first n tosses of the coin. For $j = 0, \ldots, n$, let $Q_{n,j}$ be the uniform distribution on the $\binom{n}{j}$ sequences of 0s and 1s of length n whose sum is j. It takes only a few (tedious) minutes to verify the following:

M_Q consists of all the exchangeable probabilities on $(\Omega, \mathcal{F})$.
Conditions (1), (2), and (3) are satisfied.
In Theorem 1, the σ-field $\hat{\Sigma}$ consists of the Borel sets invariant under finite permutations of coordinates.
G can be taken as the set where $(X_1 + \cdots + X_n)/n$ converges as $n \to \infty$; call the limit L.
Q_ω is the probability on Ω making the coordinates X_i independent tosses of a p-coin, where $p = L(\omega)$.
The quotient space $\mathcal{X}$ in Remark (iii) is the unit interval; the quotient σ-field is the Borel σ-field; and the quotient probability is the distribution of L, which is the mixing measure μ on [0, 1] in Theorem 5 of Part I.

Theorem 5 in part I is therefore a special case of Theorem 1 here. Of course, a direct proof is easier. But Theorem 1 does provide a unified framework for de Finetti's theorem and many variations. Corollary 2 here gives Theorem 9 in Part I, and the present Theorem 2 does L_∞. At least for us, the abstract setup makes the structure of the proofs easier to see.

Example 2. Exponential random variables. The random variable $X > 0$ has the exponential distribution with parameter λ if $P(X > x) = \exp(-\lambda x)$ for $x > 0$. Here, $0 < \lambda < \infty$. Mixtures of independent exponentials with a common parameter were characterized by Freedman (1963). Informally, a sequence of positive random variables is a mixture of exponentials iff the sums are sufficient statistics, and given the sum, the summands are uniformly distributed over the simplex. To make the connection with Theorem 1, we take $\Omega_i = W_n = (0, \infty)$ with the Borel σ-field. Let $T_n(x_1, \ldots, x_n) = x_1 + \cdots + x_n$. For $0 < t < \infty$, let $Q_{n,t}$ be the uniform distribution on the positive, finite $x_1, \ldots, x_n$ whose sum is t. Let P_λ be the probability on $(\Omega, \mathcal{F})$ according to which the coordinates X_i are independent exponentials with the common parameter λ. If μ is a probability on $(0, \infty)$, let $P_\mu = \int_0^\infty P_\lambda\, \mu(d\lambda)$, that being the mixture we want to characterize. Abstractly, P on $(\Omega, \mathcal{F})$ admits the representation $P = P_\mu$ iff $P \in M_Q$, and then μ is unique. The G in Theorem 1 is again the set where $(X_1 + \cdots + X_n)/n$ converges to a finite positive limit; denote the latter by L. And $Q_\omega = P_{1/L(\omega)}$ makes the X_i independent exponentials with common parameter $1/L(\omega)$: the inverse results from the fact that an exponential distribution with parameter λ has mean $1/\lambda$. The quotient space $\mathcal{X}$ in Remark (iii) is $(0, \infty)$, the quotient σ-field is the Borel σ-field, and the quotient probability is the distribution of $1/L$, namely, μ. If μ is allowed to have positive mass at 0, the argument is a little more complicated, because the X_i will be infinite with probability $\mu\{0\}$.

Suppose μ and ν are two probabilities on $(0, \infty)$, and c is a positive real number. It is almost obvious from Theorem 2 that $P_\mu \leq cP_\nu$ iff $\mu \leq c\nu$. In the next section, we restate the condition in terms of the Laplace transform, which may be more interesting. We also characterize μ with a bounded density: this is (a little) beyond the scope of our previous theorems, since Lebesgue measure is infinite on $(0, \infty)$.

In these examples, the "sufficient statistic" is the sum, and the conditional distribution is uniform – on $\{0, 1, \ldots, n\}$ for the coin and the simplex for the exponential. In other situations, the sufficient statistic and the conditional will be more complicated: see Diaconis and Freedman (1984) for more examples and discussion.

Laplace transforms

Let μ be a probability on $[0, \infty)$. Its Laplace transform is

$$\phi(x) = \int_0^\infty e^{-\lambda x}\, \mu(d\lambda). \tag{14}$$

We use λ as the variable of integration, in keeping with Example 2, and write ϕ_μ for ϕ if there is any ambiguity. According to a celebrated theorem of Bernstein, Laplace transforms of probabilities on $[0, \infty)$ are characterized as being "completely monotone," and taking the value 1 at $x = 0$; furthermore, μ in (14) is unique. See Widder (1946, pp. 144–163) or Feller (1971, pp. 233, 439), For these purposes, ϕ on $[0, \infty)$ is completely monotone if the nth derivative $\phi^{(n)}$ exists on $[0, \infty)$ for all n, and these functions alternate in sign, so that $(-1)^n\phi^{(n)} \geq 0$ for all $n = 0, 1, \ldots$. Of course, $\phi^{(n)}$ is continuous because $\phi^{(n+1)}$ exists. At 0, continuity and differentiability are from the right: ϕ may not be defined to the left of 0. By convention, $\phi^{(0)} = \phi$.

To avoid technical nuisances, we assume until further notice that $\mu\{0\} = 0$. Recall that $X_1, \ldots, X_n$ are independent exponential random variables relative to P_λ, with common parameter λ. The density of $X_1 + \cdots + X_n$ is

$$x \to \frac{x^{n-1}}{(n-1)!} e^{-\lambda x} \lambda^n \tag{15}$$

for $n = 1, 2, \ldots$. This is a well known formula (Feller 1971, p. 11) and is easy to verify directly. To get the density of the sum relative to P_μ we just the integrate (15) with respect to $\mu(d\lambda)$:

$$\frac{x^{n-1}}{(n-1)!} \int_0^\infty e^{-\lambda x} \lambda^n\, \mu(d\lambda) = (-1)^n \frac{x^{n-1}}{(n-1)!} \phi^{(n)}(x) \tag{16}$$

for $n = 1, 2, \ldots$. The equality in (16) follows by differentiating (14) under the integral sign, n times.

Lemma 1. *Let μ and ν be two probabilities on $(0, \infty)$. Let c be a positive constant. Then $\mu \leq c\nu$ iff $(-1)^n \phi_\mu^{(n)}(x) \leq c(-1)^n \phi_\nu^{(n)}(x)$ for all $n = 0, 1, \ldots$ and $x > 0$. For sufficiency, the upper bound is needed only for large positive n.*

Proof. Combine Theorem 2 and (16), the latter giving the density of the sufficient statistic with respect to P_μ or P_ν. This is where we use n_0 in Theorem 2. □

Corollary 3. *Let ν be exponential with parameter h. Then $\mu \leq c\nu$ iff*

$$(-1)^n \phi_\mu^{(n)}(x) \leq cn!h/(x+h)^{n+1}$$

for all $n = 0, 1, \ldots$ and $x > 0$. For sufficiency, the upper bound is needed only for large positive n.

Proof. The Laplace transform of ν is $\phi_\nu(x) = h/(h+x)$, so $(-1)^n \phi_\nu^{(n)}(x) = hn!/(h+x)^{n+1}$. □

Theorem 5. *Let ϕ be a given function on $[0, \infty)$, and c a given positive real number. Then ϕ is the Laplace transform of a probability μ on $(0, \infty)$ such that μ is absolutely continuous, with a density bounded above by c, iff $\phi(0) = 1$ and*

$$0 \leq (-1)^n \phi^{(n)}(x) \leq cn!/x^{n+1} \tag{17}$$

for all $n = 0, 1, \ldots$ and $x > 0$. Furthermore, μ is unique. For sufficiency, the upper bound is needed only for large positive n.

Proof. For uniqueness, (14) determines μ according to Bernstein's theorem. Suppose that ϕ is the Laplace transform of a probability μ on $(0, \infty)$ with $d\mu/dx \leq c$; the conditions on ϕ and its derivatives follow by routine calculus, proving necessity. For sufficiency, μ exists by Bernstein's theorem. Let

$$\psi(x) = \int_0^\infty e^{-\lambda x} e^{-\lambda} \mu(d\lambda) = \phi(x+1). \tag{18}$$

Plainly, $(-1)^n \psi^{(n)}(x) \leq cn!/(1+x)^{n+1}$. Corollary 3 shows that $e^{-\lambda}\mu(d\lambda) \leq ce^{-\lambda} d\lambda$, which completes the proof. □

Essentially this theorem can be found in Widder (1946, p. 315) or Feller (1971, p. 440); also see Hirschman and Widder, (1955, chap. 7). There are similar – albeit more complicated – results for L_p: see Widder (1946, pp. 288, 312–14). Rather than pursuing this topic, we turn to inversion formulas for the Laplace transform (Widder, 1946, p. 288; Feller, 1971, p. 440). These have always seemed mysterious, at least to us; the theory developed here may help. In Example 2, the P_μ density of $n/(X_1 + \cdots + X_n)$ converges weak-star to μ: indeed, $n/(X_1 + \cdots + X_n)$ converges a.e. $[P_\lambda]$ to λ, by the strong law. The density of the denominator $X_1 + \cdots + X_n$ was computed from the Laplace transform ϕ of μ, in (16). By a change of variables $(y = n/x)$, the density of $n/(X_1 + \cdots + X_n)$ is seen to be

$$f_n(y) = (-1)^n \frac{1}{n!} \left[\phi^{(n)}\left(\frac{n}{y}\right)\right] \left(\frac{n}{y}\right)^{n+1}. \tag{19}$$

As noted above, $f_n(y)dy$ converges to $\mu(dy)$ as n grows, which gives the basic inversion formula (Widder, 1946, p. 288). The convergence is better for smoother μ, but weak-star convergence always holds. We have assumed $\mu\{0\} = 0$. Otherwise, the contribution from 0 needs to be assessed separately: the distribution of $n/(X_1 + \cdots + X_n)$ picks up an atom at 0, whose mass is – naturally – $\mu\{0\}$.

Example 3. Let $0 < \alpha < \infty$. The Γ-density $\lambda \to \lambda^{\alpha-1}e^{-\lambda}/\Gamma(\alpha)$ has Laplace transform $x \to 1/(1+x)^\alpha$ for $0 \le x < \infty$. Now let $f(\lambda) = 1$ for λ near 0, while $f(\lambda) = 1/\sqrt{|1-\lambda|}$ for λ near 1, and $f(\lambda) = e^{-\lambda}$ for λ near ∞. The definition of f on $(0, \infty)$ can be completed so that f is a positive density, and C_∞ except at 1. It is routine to show that the Laplace transform ϕ of f is approximately 1 near 0 and $1/x$ near ∞. Plainly, f is unbounded. In short, the condition $\phi(x) \le c/x$ does not establish the boundedness of f in Theorem 5. In this example, the upper bound in (17) will hold for $n = 0, \ldots, n_0$, although c will depend on n_0. As $n \to \infty$, however, (19) suggests that $(-1)^n\phi^{(n)}(x)x^{n+1}/n!$ will be unbounded for x near n, so the upper bound in (17) fails. We have not verified this directly, but see Widder (1946, p. 288).

Remark. (i) We think that Widder (1946, p. 288, Definition 6) omitted a factor $1/k!$ in the definition of $L_{k.t}$; if so, our (19) matches up; otherwise, we cannot verify the calculations following his definition.

(ii) From the present perspective, Bernstein's theorem can be derived from Hausdorff's solution to the little moment problem – Theorem 2 in Part I. The connection is made by the mapping $\lambda \to -\log\lambda$, which takes the unit interval to the half-line. Bernstein seems to have been unaware of Hausdorff's work; Widder confesses to having rediscovered it for a third time (Widder, 1946, p. 144). With respect to Hausdorff's solution to Markov's problem, we might be in fourth place.

Example 4. Normal random variables with mean 0. In connection with Theorem 1, we considered scale mixtures of normal random variables with common mean 0. There, we used variance as the parameter; here, it will be more convenient to use the "natural parameter" $\lambda = 1/\sigma^2$. See Lehmann (1991, p. 57). Let $\Omega_i = (-\infty, \infty)$, and let P_λ on $\Omega = \prod_i \Omega_i$ make the coordinates X_i independent normal random variables, with mean 0 and common variance $1/\lambda$. For any probability μ on $(0, \infty)$, let $P_\mu = \int P_\lambda\, \mu(d\lambda)$. When does μ have a bounded density with respect to Lebesgue measure? with respect to Haar measure $d\lambda/\lambda$? The nth sufficient statistic will be taken as $T_n = \frac{1}{2}(X_1^2 + \cdots + X_n^2)$. Let $x \to \psi_{\mu,n}(x)$ be the density of T_n with respect to P_μ. By excluding a set of measure 0 with respect to all P_μ, we can assume that our X_i never vanish, so $T_n > 0$. Let $m = n/2$. For $n = 1, 2, \ldots,$ the density of T_n with respect to P_λ is

$$x \to \frac{x^{m-1}}{\Gamma(m)} e^{-\lambda x}\lambda^m.$$

See (Feller, 1971, pp. 47–48). The density with respect to P_μ is therefore

$$x \to \psi_{\mu,2m}(x) = \frac{x^{m-1}}{\Gamma(m)} \int_0^\infty e^{-\lambda x}\lambda^m\, \mu(d\lambda) = (-1)^m \frac{x^{m-1}}{(m-1)!}\phi_\mu^{(m)}(x)$$

for $m = 1, 2, \ldots,$ where $\phi_\mu(x)$ is the Laplace transform of μ; the second equality holds by (16). Now $x^2\psi_{\mu,2m}(x)/m = (-1)^m x^{m+1}\phi_\mu^{(m)}(x)/m!$, and Theorem 5 shows

(20) The mixing measure μ in Example 4 is absolutely continuous with a density bounded above by c iff $\psi_{\mu,2m}(x) \leq cm/x^2$ for all positive x and $m = 1, 2, \ldots$.

Interestingly, this constrains T_n only for positive even n: we are not ready for fractional derivatives, nor is $\psi_{\mu,0}$ defined. (Among other things, $\Gamma(0) = \infty$ and $T_0 = 0$ if it is to be defined at all.) Of course, if $\dot{\mu} \leq c$, then $\phi_\mu(x) \leq c/x$; but going in, the upper bound is unavailable for the Laplace transform itself. That is why we wanted a version of Theorem 5 that requires the upper bound only for derivatives.

Of course, if $\phi(x)$ is the Laplace transform of $\mu(d\lambda)$, then $-\phi'(x)$ is the Laplace transform of $\lambda\mu(d\lambda)$, and Theorem 5 can be applied to the latter. Indeed,

$$(-1)^m \frac{x^{m+1}}{m!}\frac{\partial^m}{\partial x^m}(-\phi_\mu) = (-1)^{m+1}\frac{x^{m+1}}{m!}\frac{\partial^{m+1}}{\partial x^{m+1}}\phi_\mu = x\psi_{\mu,2m+2}(x).$$

Consequently,

(21) μ is absolutely continuous with a density bounded above by $\lambda \to c/\lambda$ iff $\psi_{\mu,2m+2}(x) \leq c/x$ for all positive x and $m = 0, 1, \ldots$.

Example 2 can be handled in a similar way. This is not surprising, since the sum of m exponential variables is distributed as 1/2 times a χ^2 variable with $2m$ degrees of freedom.

(22) The mixing measure μ in Example 2 is absolutely continuous with a density bounded above by c iff the density of the sufficient statistic $X_1 + \cdots + X_m$ is bounded above by cm/x^2 for all positive x and $m = 1, 2, \ldots$.

There is an entertaining geometrical consequence to the connection between the χ^2 and the exponential distributions. Let $X_1, X_2, \ldots, X_{2n-1}, X_{2n}$ be independent normal random variables, with mean 0 and variance 1. Then $(X_1^2 + X_2^2)/2, \ldots, (X_{2n-1}^2 + X_{2n}^2)/2$ are independent standard exponential variables. Given $X_1^2 + X_2^2 + \cdots + X_{2n-1}^2 + X_{2n}^2$, we have on the one hand that

$$X_1, X_2, \ldots, X_{2n-1}, X_{2n}$$

is uniformly distributed over a sphere in R^{2n}; on the other hand,

$$(X_1^2 + X_2^2)/2, \ldots, (X_{2n-1}^2 + X_{2n}^2)/2$$

is uniformly distributed over a simplex in the positive orthant of R^n. Consequently,

(23) Pick a point $(x_1, x_2, \ldots, x_{2n-1}, x_{2n})$ uniformly at random on the surface of a sphere in $2n$-dimensional Euclidean space. Then the point $\big((x_1^2 + x_2^2), \ldots, (x_{2n-1}^2 + x_{2n}^2)\big)$ is uniformly distributed over the corresponding simplex in the positive orthant of n-dimensional space.

In general, as is well known, the partitioned sum of squares has a Dirichlet distribution on the simplex.

Brief literature review

The proof of Theorem 1 is given in Diaconis and Freedman (1984). This follows Oxtoby (1952), who gave a masterful exposition of the Krilov-Bogolioubov theory, presenting stationary processes as mixtures of ergodic processes. Similar techniques were used by Hunt (1960) to develop the Martin boundary for transient Markov chains. The Scandinavian school has worked on such problems from a slightly different perspective: see Martin-Löf (1974), Lauritzen (1988), or Kallenberg (1999). There has been an extensive development of such theories in statistical mechanics; see Ruelle (1984) and Georgii (1988). Aldous (1985) discusses applications to probability theory; and Schervish (1995), to Bayesian statistics. Many other examples, discussed from the perspective of semigroups and Choquet theory, will be found in Berg, Christensen and Ressel (1984); the connection to de Finetti's theorem is explained in Ressel (1985). The characterization of mixtures of normals appears in Freedman (1963). It is often attributed to Schoenberg (1938a): see especially Theorem 2 on p. 817, also see Schoenberg (1938b). But the translation is not without difficulty.

References

Aldous, D.J.: Exchangeability and related topics. École d'été de probabilités de Saint-Flour XIII. edité par P. L. Hennequin. Lecture notes in mathematics, 1117, Springer-Verlag, 1985

Berg, C., Christensen, J.P.R., Ressel, P.: Harmonic analysis on semigroups: theory of positive definite and related functions. Springer Graduate Texts in Mathematics, **100**, 1984

Diaconis, P., Freedman, D.: Partial exchangeability and sufficiency. In: J. K. Ghosh, J. Roy, (eds.), Proc. Indian Statist. Assoc. Golden Jubilee: Applications And New Directions. Indian Statist. Assoc., Calcutta, 1984, pp. 205–236

Feller, W.: An Introduction to Probability Theory and its Applications. Vol. **II**, Second edition, Wiley, New York, 1971

Freedman, D.A.: Invariants under mixing which generalize de Finetti's theorem. Ann. Math. Stat. **34**, 1194–1216 (1963)

Georgii, H.O.: Gibbs Measures and Phase Transitions. de Gruyter, Berlin, 1988

Hewitt, E., Stromberg, K.: Real and Abstract Analysis. Springer, New York, 1969

Hirschman, I.I., Widder, D.V.: The Convolution Transform. Princeton University Press, 1955

Hunt, G.: Markoff chains and Martin boundaries. Ill. J. Math. **4**, 313–40 (1960)

Kallenberg, O.: Multivariate sampling and the estimation problem for exchangeable arrays. J. Theor. Prob. **12**, 859–83 (1999)

Lanford, O., Ruelle, D.: Observables at infinity and states with short range correlations in statistical mechanics. Comm. Math. Phys. **13**, 194–215 (1969)

Lauritzen, S.L.: Extremal Families and Systems of Sufficient Statistics. Springer Lecture Notes in Statistics, **49**, 1988

Lehmann, E.L.: Testing Statistical Hypotheses. 2nd ed. Wadsworth & Brooks/Cole, 1991

Martin-Löf, P.: Repetitive structures and the relation between canonical and microcanonical distributions in statistics and statistical mechanics. In: O. Barndorff-Nielsen, P. Blaesild, G. Schou, (eds.), Proceedings of Conference on Foundational Questions in Statistical Inference. Aarhus, 1974

Oxtoby, J.C.: Ergodic sets. Bull. Amer. Math. Soc. **58**, 116–36 (1952)
Ressel, P.: De Finetti-type theorems: An analytical approach. Ann. Prob. **13**, 898–922 (1985)
Ruelle, D.: Thermodynamic Formalism: The Mathematical Structures of Classical Equilibrium Statistical Mechanics. Cambridge University Press, 1984
Schervish, M.J.: Theory of Statistics. Springer, 1995
Schoenberg, I.: Metric spaces and completely monotone functions. Ann. Math. **39**, 811–41 (1938a)
Schoenberg, I.: Metric spaces and positive definite functions. Trans. Amer. Math. Soc. **44**, 522–36 (1938b)
Widder, D. V.: The Laplace Transform. First printing, 1941, Princeton University Press, 1946