Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

PARAMETRIZING DOUBLY STOCHASTIC MEASURES

BY RICHARD A. VITALE*
University of Connecticut

Doubly stochastic measures can be identified with the trace of a pair of (Lebesgue) measure preserving maps of the unit interval to itself.

It has been of traditional interest in probability theory to produce a random vector (or metric space element), which has a given distribution and is defined on a standard space, such as $[0, 1]$ endowed with Lebesgue measure. In a classic work, Lévy (1937, section 23) used an approach based on conditioning. For the purpose of the Skorokhod representation, Billingsley (1971, Theorem 3.2) considered the case of random elements of a general metric space. Whitt (1976, Lemma 2.7) considered general measures on R^n and employed a Borel isomorphism to treat questions of extremal correlation and minimal variance. Rüschendorf (1983) used a similar approach to consider a general class of optimization problems.

In this note, we revisit the question of representing a random element in the special case of a doubly stochastic measure on the unit square. First we show the existence of a random element (using essentially Whitt's approach) with a refinement to a *canonical* representation. Then we turn to criteria for extremality of a doubly stochastic measure. Our aim is to provide the reader who is interested in extremality with different settings.

This paper was invited for presentation at the AMS-IMS-SIAM Joint Summer Research Conference on Distributions with Fixed Marginals, Doubly-Stochastic Measures, and Markov Operators, July 31–August 6, 1993. At the Conference, the author learned of *A Representation for Doubly Stochastic Measures* by W. F. Darsow and E. T. Olsen. Subsequently, a revision (Darsow and Olsen, 1994) was provided to the author. Interested readers are directed to this work for variant arguments and related topics, including copulas.

---

* Supported in part by ONR Grant N00014-90-J-1641 and NSF Grant DMS-9002655.

AMS 1991 Subject Classification: Primary 62E10; Secondary 60E99.

Key words and phrases: Doubly stochastic measures. .

1. Doubly Stochastic Measures via Pairs of Measure-Preserving Mappings. A doubly stochastic measure is a Borel measure on the square with Lebesgue marginals. We use I and $I \times I$ for the unit interval and unit square respectively. $\mathcal{B}$ and $\mathcal{B} \times \mathcal{B}$ will denote their respective Borel σ-algebras (associated Lebesgue measures m and $m \times m$). U will stand for a random variable uniformly distributed on $[0, 1]$.

DEFINITION. A Borel measurable function $\sigma : I \to I$ that satisfies $m(\sigma^{-1}(B)) = m(B)$ for each $B \in \mathcal{B}$ is *measure-preserving (m.p.)*. The class of such maps will be denoted Σ.

DEFINITION. A doubly stochastic measure μ on $\mathcal{B} \times \mathcal{B}$ is *realized* by a pair (σ, τ) of elements of Σ if for any Borel subset B^2 of the square $\mu(B^2) = m(\sigma, \tau)^{-1}(B^2))$.

We can now assert the following representation.

THEOREM 1. *(i) If $\sigma, \tau \in \Sigma$, then the measure μ on $\mathcal{B} \times \mathcal{B}$ defined by $\mu(B^2) = m((\sigma, \tau)^{-1}(B^2))$ is doubly stochastic. (ii) Conversely, every doubly stochastic measure μ can be realized as a pair $(\sigma, \tau) \in \Sigma \times \Sigma$. (iii) In (ii), one may assume that $(\sigma, \tau)^{-1}(\mathcal{B} \times \mathcal{B}) = \mathcal{B}$. In this case, we say that (σ, τ) is canonical.*

Before proceeding to the proof, we collect some well-known results.

THEOREM 2. (Kuratowski, see Royden (1988)). *Every uncountable, complete, separable metric space S is Borel equivalent to I, that is, there is a $1-1$ map $\varphi : S \to I$ such that both φ and φ^{-1} are Borel measurable.*

PROPOSITION 1. (e.g. Vitale, 1991). *Let X be a random variable with continuous distribution function F. Then $U = F(X)$ is uniformly distributed on $[0, 1]$ and $X \overset{\text{a.s.}}{=} F^{-1}F(X)$, where $F^{-1}(u) = \inf\{x | u \leq F(x)\}$ is left-continuous.*

PROOF OF THEOREM 1. For (i), note that if $B^2 = B^1 \times I$, $B^1 \in \mathcal{B}$, then $\mu(B^2) = m(\sigma^{-1}(B^1)) = m(B^1)$ and likewise for the other marginal.

For (ii), recall the standard fact that there exists a pair of random variables (X, Y) with joint distribution μ. From Theorem 2, there is a Borel equivalence $\varphi : I \times I \to I$. Set $W = \varphi(X, Y)$, and let F be its distribution function. F has no jumps since this would imply an atom in the distribution of (X, Y). By Proposition 1, $U = F(W)$ is uniformly distributed and $W \overset{\text{a.s.}}{=} F^{-1}(U)$. It follows that $(X, Y) \overset{\text{a.s.}}{=} \varphi^{-1}F^{-1}(U)$, and we can take $(\sigma, \tau) = \varphi^{-1}F^{-1} : I \to I \times I$.

For the final assertion note that by Theorem 2 $(\varphi^{-1}F^{-1})^{-1}(\mathcal{B}\times\mathcal{B}) = (F^{-1})^{-1}\varphi(\mathcal{B}\times\mathcal{B}) = (F^{-1})^{-1}(\mathcal{B})$. The last expression is equal to $\mathcal{B}$ since $(F^{-1})^{-1}([0,F^{-1}(u)]) = [0,u]$ for every $u \in I$. ∎

2. Extreme Doubly Stochastic Measures. The class of doubly stochastic measures is a convex sub-class of the Borel measures on $\mathcal{B}\times\mathcal{B}$. The Douglas-Lindenstrauss condition for extreme points (Douglas (1964); Lindenstrauss (1965) takes the following form.

THEOREM 3. *The doubly stochastic measure realized by* $(\sigma,\tau)\in\Sigma\times\Sigma$ *is extreme iff given any Borel-measurable* $f: I\times I\to R$ *such that* $E|f[\sigma(U),\tau(U)]|<\infty$ *and* $\epsilon>0$, *there are functions* $g,h: I\to R$ *such that*

$$E|f[(\sigma,\tau)(U)] - g[\sigma(U)] - h[\tau(U)]| < \epsilon. \tag{1}$$

Approximation by simple functions leads to the following variant, which is more convenient to apply in some cases (see the following example).

THEOREM 4. *The measure realized by* (σ,τ) *is extreme iff for every* $A\in\sigma^{-1}\mathcal{B}$, $B\in\tau^{-1}\mathcal{B}$, *and* $\epsilon>0$ *there are measurable partitions* $A_1,A_2,A_3,\cdots,A_m\in\sigma^{-1}\mathcal{B}$ *and* $B_1,B_2,B_3,\cdots,B_n\in\tau^{-1}\mathcal{B}$ *and constants* $a_i, 1\le i\le m$ *and* $b_j, 1\le j\le n$ *such that*

$$E|1_{A\cap B}(U) - \sum_{i=1}^{m} a_i 1_{A_i}(U) - \sum_{j=1}^{n} b_j 1_{B_j}(U)| < \epsilon. \tag{2}$$

For the sufficiency part, the sets A *and* B *can be specified to be of the form* $(-\infty,t]$.

EXAMPLE. Consider the doubly stochastic measure which places mass 1/3 uniformly on each of the 3 line segments:

$$\begin{aligned} (u,u),&\quad 1/3\le u\le 1\\ (u,1/2(u-1/3)),&\quad 1/3\le u\le 1\\ (u,2u+1/3)),&\quad 0\le u\le 1/3. \end{aligned}$$

This can be parametrized by the pair (σ,τ), where

$$\sigma(u) = \begin{cases} 2u+1/3, & 0\le u<1/3\\ u-1/3, & 1/3\le u<2/3\\ 2u-1, & 2/3\le u\le 1 \end{cases}$$

and

$$\tau(u) = \begin{cases} u, & 0 \le u < 1/3 \\ 2u - 1/3, & 1/3 \le u < 2/3 \\ 2u - 1, & 2/3 \le u \le 1 \,. \end{cases}$$

Note that for any a, b, the set $[\sigma \le a] \cap [\tau \le b]$ is (modulo a null set) a union of sets among $[0, c_1], [1/3, c_2], [2/3, c_3]$, where $0 \le c_1 \le 1/3, 1/3 \le c_2 \le 2/3$ and $2/3 \le c_3 \le 1$. It is enough to observe then that $1_{[0,c_1]} = 1_{[\tau \le c_1]}, 1_{[1/3,c_2]} = 1_{[\sigma \le c_2 - 1/3]}$, and $1_{[2/3,c_3]} = 1_{[1/3 \le \tau \le 2c_3 - 1]} - 1_{[\sigma \le c_3 - 2/3]}$, which implies that the associated doubly stochastic measure is extreme (cf. Shiflett (1972)).

DEFINITION. If $\mathcal{B}_1$ and $\mathcal{B}_2$ are two sub σ-algebras of $\mathcal{B}$, then we write $\mathcal{B}_1 \approx \mathcal{B}_2$ if they differ only by null sets, i.e., for any $B_1 \in \mathcal{B}_1$ ($B_2 \in \mathcal{B}_2$), there is a $B_2 \in \mathcal{B}_2$ ($B_1 \in \mathcal{B}_1$) such that $m(B_1 \Delta B_2) = 0$.

Theorem 4 has the following immediate consequence.

COROLLARY 1. *The property of being extreme for* (σ, τ) *depends only on* $\sigma^{-1}\mathcal{B}$ *and* $\tau^{-1}\mathcal{B}$. *That is, if for any other pair* $(\tilde{\sigma}, \tilde{\tau})$,

$$\tilde{\sigma}^{-1}\mathcal{B} \approx \sigma^{-1}\mathcal{B} \quad \text{and} \quad \tilde{\tau}^{-1}\mathcal{B} \approx \tau^{-1}\mathcal{B},$$

then either both *pairs* $(\sigma, \tau), (\tilde{\sigma}, \tilde{\tau})$ *are extreme or* neither *is.*

One might ask from the Corollary whether *every* pair of candidate sub σ-algebras can be realized (modulo $\approx$) as $\sigma^{-1}\mathcal{B}$ and $\tau^{-1}\mathcal{B}$. The next result implies that this is the case.

PROPOSITION 2. *Given a non-atomic sub* σ*-algebra* $\mathcal{B}_1 \subseteq \mathcal{B}$, *there is* $\sigma \in \Sigma$ *such that* $\sigma^{-1}\mathcal{B} \approx \mathcal{B}_1$.

PROOF. Let $D_1, D_2, \cdots$ be the respective dyadic subintervals of I, *i.e.*, $D_1 = [0, 1/2], D_2 = [0, 1/4] \cup [1/2, 3/4], \cdots$. Their indicator functions are stochastically independent, symmetric Bernoulli random variables. Define a random variable U on I by

$$U = \sum_{i=1}^{\infty} \frac{1}{2^i} 1_{D_i}(\cdot). \tag{3}$$

Suppose in addition that D_0 is any other Borel subset of I. By the Carathéodory isomorphism theorem for measure algebras (Royden (1988) p. 399), there is a system of subsets $B_0, B_1, \cdots$ of $\mathcal{B}_1$ such that

$$m(B_i) = m(D_i) \tag{4}$$

for all i and

$$m(B_i \cap B_j) = m(D_i \cap D_j) \tag{5}$$

for all $i \neq j$. Define the random variable $\sigma : I \to I$ by

$$\sigma = \sum_{i=1}^{\infty} \frac{1}{2^i} 1_{B_i}(\cdot). \tag{6}$$

We claim that

$$\sigma^{-1}(\mathcal{B}) \approx \mathcal{B}_1. \tag{7}$$

Note that $\sigma^{-1}(\mathcal{B})$ is the σ-algebra generated by $B_1, B_2, \cdots$. It is enough to show "inclusion" from the right in (7), so consider $B_0 \in \mathcal{B}_1$ (by the isomorphism theorem, we can assume that such an arbitrary choice is the image of some D_0 under the isomorphism). Now it is standard that $m(D_0 \Delta \liminf_{n\to\infty} D_n^*) = 0$ for a sequence D_n^* in the *algebra* generated by $\{D_1, D_2, \cdots\}$. By the isomorphism, i.e., (4) and (5), it must be that $m(B_0 \Delta \liminf_{n\to\infty} B_n^*) = 0$ for the image sequence B_n^* in the algebra generated by $\{B_1, B_2, \cdots\}$ and hence in $\sigma^{-1}(\mathcal{B})$. ∎

3. A Second Characterization. We turn to another way of characterizing extreme (equivalently, non-extreme) doubly stochastic measures.

DEFINITION. A random variable valued in I is of class BD (Bounded Density) if it has a bounded density with respect to Lebesgue measure. In particular, a uniform random variable is of class BD.

THEOREM 5. *Let (σ, τ) canonically represent a doubly stochastic measure μ. Then μ is* not *extreme iff there is a nonuniform BD random variable X such that both σX and τX are uniform.*

PROOF. (i): Suppose the existence of such an X. Let μ_1 be the bivariate measure induced on the square by $(\sigma, \tau)X$. For any Borel subset B^2 of the square, $\mu_1(B^2) = \Pr(X \in (\sigma, \tau)^{-1}B^2) \leq c \cdot m((\sigma, \tau)^{-1}B^2) = c \cdot \mu(B^2)$. Here $c > 1$ is any number which exceeds the essential supremum of the density of X. Note that $\mu_1 \neq \mu$ since there is some Borel B in the interval such that $\Pr(X \in B) \neq \Pr(U \in B)$. By the canonical nature of (σ, τ), we have $B = (\sigma, \tau)^{-1}B^2$ for some B^2, which implies that $\mu_1(B^2) \neq \mu(B^2)$. Now define $\mu_2 = (1 - 1/c)^{-1}(\mu - (1/c)\mu_1)$. We have $\mu = (1/c)\mu_1 + (1 - 1/c)\mu_2$, which exhibits μ as non-extreme.

(ii): If μ is not extreme, $\mu = \theta\mu_1 + (1-\theta)\mu_2$, $\mu_1 \neq \mu$, then by the Radon-Nikodym theorem, there is a bounded density f, not identically unity, such that for any Borel subset B^2 of the square

$$\mu_1(B^2) = \int_{B^2} f(x,y)\mu(dx,dy),$$

or, using that (σ, τ) realizes μ,

$$\mu_1(B^2) = Ef(\sigma U, \tau U)1_{B^2}(\sigma U, \tau U) = Ef(\sigma U, \tau U)f(\sigma U, \tau U)1_{(\sigma,\tau)^{-1}B^2}(U). \tag{8}$$

Note that

$$Ef(\sigma U, \tau U)1_B(U)$$

defines a nonuniform probability measure on Borel subsets B of the interval. If X has this probability measure, then

$$\Pr(X \in B) = E1(X \in B) = Ef(\sigma U, \tau U)1_B(U),$$

and (8) may be rewritten

$$\mu_1(B^2) = \Pr((\sigma X, \tau X) \in B^2).$$

Since μ_1 is doubly stochastic, it follows that σX and τX are uniform. ∎

REMARK. We should emphasize that this result does not hold if the boundedness of the density of X is relaxed. This can be seen in an example of Losert (1982, p. 391).

References

BILLINGSLEY, P. (1971). *Weak Convergence of Measures: Applications in Probability.* Regional Conference Series in Applied Mathematics No. 5, SIAM, Philadelphia.

DARSOW, W. F. and OLSEN, E. T. (1994). A representation for doubly stochastic measures. *IIT Research Report*, Illinois Institute of Technology.

DOUGLAS, R. G. (1964). On extremal measures and subspace density. *Michigan Math. J.* **11**, 243–246.

LÉVY, P. (1937). *Théorie de l'Addition des Variables Aleatoires.* Gauthier-Villars, Paris.

LINDENSTRAUSS, J. (1965). A remark on extreme doubly stochastic measures. *Amer. Math. Monthly* **72**, 379–382.

LOSERT, V. (1982). Counter-examples to some conjectures about doubly stochastic measures. *Pacific J. Math.* **99**, 387–397.

ROYDEN, H. L. (1988). *Real Analysis* (third edition), Macmillan, New York.

RÜSCHENDORF, L. (1983). Solution of a statistical optimization problem by rearrangement methods. *Metrika* **30**, 55–61.

SHIFLETT, R. C. (1972). Extreme Markov operators and the orbits of Ryff. *Pacific J. Math.* **40**, 201–206.

VITALE, R. A. (1991). Approximation by mutually completely dependent processes. *J. Approx. Theory* **66**, 225–228.

WHITT, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4**, 1280–1289.

DEPARTMENT OF STATISTICS
BOX U-120
UNIVERSITY OF CONNECTICUT
STORRS, CT 06269-3120
rvitale@uconnvm.uconn.edu