

## TWO CHARACTERIZATIONS OF CONDITIONAL PROBABILITY<sup>1</sup>

M. M. RAO

**ABSTRACT.** This note contains two types of functional equations characterizing the conditional probability measures as subclasses of vector measures on general Banach function spaces. One uses the theory of vector integration and the other employs a form of the Šidák identity.

**1. Introduction.** Let  $(\Omega, \Sigma, \mu)$  be a measure space and  $L^1(\Sigma)$  be the usual Lebesgue space of real functions on  $\Omega$ . If  $\mathfrak{B} \subset \Sigma$  is a sub  $\sigma$ -algebra, then the conditional expectation  $E^{\mathfrak{B}}: L^1(\Sigma) \rightarrow L^1(\Sigma)$  exists and the mapping  $P^{\mathfrak{B}}: A \mapsto E^{\mathfrak{B}}(\chi_A) \in L^1(\mathfrak{B})$  is called the conditional probability function ( $\mu(\Omega) = \infty$  being admitted). The fact that  $P^{\mathfrak{B}}$  is (not a scalar but) a vector measure can be recognized in the detailed discussion, perhaps for the first time, by Dubins [2]; but this point of view has not been pursued for several years. In 1965 Olson gave a characterization of  $P^{\mathfrak{B}}$  (when  $\mu(\Omega) = 1$ ) in [5] as a subclass of vector-valued set functions from  $\Sigma \rightarrow L^1(\mathfrak{B})$ , based on the theory and ideas of averaging operators, and indicated that in a future paper a similar result will be presented relating it to the Šidák identity [8]. It did not seem to have appeared. However, much more general results can be given if  $\mu$  is not necessarily finite and the  $L^p$ -spaces are replaced by the Banach function spaces  $L^p$  (cf. [9, Chapter 15] and [4]), and this will illuminate the structure of these measures. The purpose of this note is to present two such characterizations, one based on the Dunford-Schwartz vector integration and the other on a Šidák identity. These will complement and clarify the work in [5]. Since the  $L^p$ -spaces include both the Orlicz and Lebesgue spaces, the generality implied here has wider applications.

Let us first recall the  $L^p(\Sigma)$  and fix some notation. If  $L^0(\Sigma)$  denotes the class of all scalar measurable functions, a mapping  $\rho: L^0(\Sigma) \rightarrow \bar{\mathbf{R}}^+$  is a function norm whenever (i)  $\rho(f) = 0$  iff  $f = 0$ , (ii)  $\rho(f) = \rho(|f|)$ , and  $\rho(af) = a\rho(f)$ ,  $a \in \mathbf{R}^+$ , (iii)  $\rho(f + g) \leq \rho(f) + \rho(g)$ , (iv)  $0 \leq f_1 \leq f_2$  a.e.  $\Rightarrow \rho(f_1) \leq \rho(f_2)$ , and (v)  $0 \leq f_n \uparrow f$  a.e.  $\Rightarrow \rho(f_n) \uparrow \rho(f)$ . Let  $L^\rho(\Sigma) = \{f \in L^0(\Sigma): \rho(f) < \infty\}$ . Then  $L^\rho = (L^\rho(\Sigma), \rho)$  is a Banach space and it reduces to the Orlicz space  $L^\Phi(\Sigma)$  if  $\rho(f) = \inf\{k > 0: \int_\Omega \Phi(|f|/k) d\mu \leq 1\}$  where  $\Phi$  is a convex Young's function. Also  $\rho$  is said to have the *Jensen* or (*J*-) *property* if  $\rho(E^{\mathfrak{B}}(f)) \leq \rho(f)$  for each simple  $f \in L^\rho(\Sigma)$  and  $\sigma$ -algebra  $\mathfrak{B} \subset \Sigma$ . The *J*-property is present for any Orlicz (hence Lebesgue) space, but need not be

---

Received by the editors December 5, 1975.

*AMS (MOS) subject classifications* (1970). Primary 47A50, 46G10, 60B05.

*Key words and phrases.* Conditional probability and expectation, vector integration, Šidák's identity.

<sup>1</sup>Prepared with the partial support of NSF grant MPS 75-13881.

© American Mathematical Society 1976

true in general  $L^p$ -spaces. This property is introduced in [4], and investigated. The  $\rho$  is called an *absolutely continuous norm* (a.c.n.) iff  $\rho(f\chi_{A_n}) \rightarrow 0$  for all  $f \in L^p(\Sigma)$  and  $A_n \searrow, \mu(A_n) \rightarrow 0$ , and *strictly monotone* if in (iv)  $\rho(f_1) < \rho(f_2)$  unless  $f_1 = f_2$  a.e. The properties of  $L^p$ -spaces to be used below are proved for  $\sigma$ -finite  $\mu$  in [4] and [9], but they are now known to be true, as stated here, for more general measures. The example of  $\rho(f) = \int_0^1 |f|(t) dt/t$  shows that even if  $\mu(\Omega) < \infty$ , and  $\rho$  has a.c.n.,  $L^p$  may not have constants in it. So the following concept called a *generalized weak unit* (g.w.u.) will be useful. A family  $F = \{\chi_{A_\alpha}, \alpha \in I\}$  is a g.w.u. if  $\rho(\chi_{A_\alpha}) < \infty, \mu(A_\alpha) < \infty, \cup_{\alpha \in I} A_\alpha \doteq \Omega$  (i.e. both sides differ by a null set), and  $A \in \Sigma \Rightarrow A \doteq \cup_{\alpha \in J} A \cap A_\alpha$  for at most a countable set  $J \subset I$ . If  $\mu$  is  $\sigma$ -finite, a g.w.u. always exists, and if a g.w.u. exists, then  $\mu$  has the direct sum property or is strictly localizable (cf. [9]). With  $\rho$  one can pair an *associate norm*,  $\rho'$  defined by  $\rho'(f) = \sup\{|\int_\Omega fg d\mu|: \rho(g) \leq 1\}$ . This satisfies conditions (i)–(v). Hereafter let  $\Sigma_0 = \{A \in \Sigma: \mu(A) < \infty\}$ . To exclude a trivial case let  $\mu$  have the finite subset property [9, p. 257], i.e.  $A \in \Sigma, \mu(A) > 0 \Rightarrow \exists B \in \Sigma_0, B \subset A$  and  $\mu(B) > 0$ . Then  $\rho'$  is nontrivial. This is always true if there is a g.w.u. in  $L^p$ .

A linear mapping  $T: L^p(\Sigma) \rightarrow L^p(\Sigma)$  is a *Šidák operator* if for any  $f, g$  in  $L^p(\Sigma), T((Tf) \vee (Tg)) = (Tf) \vee (Tg)$  where  $\vee = \max$ . It can be checked that every Šidák operator is a projection. This identity was introduced in [8], when  $L^p = L^2, \mu(\Omega) = 1$  and  $T1 = 1$ . It is now possible to state and prove the results which I had announced before.

**2. Characterizations.** The first result uses a special Šidák identity.

DEFINITION. Let  $\mathfrak{X}$  be a Banach space and  $\nu: \Sigma_0 \rightarrow \mathfrak{X}$  be a vector measure. Then  $\nu$  is of  $\rho'$ -semivariation finite iff (i)  $\nu(A) = 0$  whenever  $\mu(A) = 0$ , (ii)  $\infty > \sup\{\rho'(\nu_B): B \in \Sigma_0\} (= \rho'(\nu), \text{ say})$  where

$$(1) \quad \rho'(\nu_B) = \sup \left\{ \sup \left[ \left| x^* \left( \int_B f d\nu \right) \right| : \rho(f) \leq 1, f \mu\text{-simple} \right], \right. \\ \left. \|x^*\| \leq 1, x^* \in \mathfrak{X}^* \right\}.$$

Here

$$\int_B f d\nu = \sum_{i=1}^n a_i \nu(A_i \cap B) \quad \text{if } f = \sum_{i=1}^n a_i \chi_{A_i} \in L^p(\Sigma).$$

It can be shown that if  $\mathfrak{X} = L^p(\mathfrak{B}), \mu|_{\mathfrak{B}} = \mu_{\mathfrak{B}}$  is localizable and  $\rho$  has the  $J$ -property, then  $\rho'(P^{\mathfrak{B}}) \leq 1$  is true.

THEOREM 1. Let  $(\Omega, \Sigma, \mu)$  be a measure space, and  $L^p(\Sigma)$  be the real function space introduced above. Suppose that there is a g.w.u.  $F_0 = \{\chi_{A_\alpha}, \alpha \in I\} \subset L^p \cap L^{p'}(\Sigma)$ , and  $\rho$  is an a.c.n. with the strictly monotone and  $J$ -properties. Let  $\nu: \Sigma_0 \rightarrow L^p(\Sigma)$  be a vector measure. Then  $\nu$  is a conditional probability  $P^{\mathfrak{B}}$  for a unique  $\sigma$ -field  $\mathfrak{B} \subset \Sigma$  with  $A_\alpha \in \mathfrak{B}, \alpha \in I$ , iff (i)  $0 \leq \nu(A), A \in \Sigma_0$ , (ii)  $\rho'(\nu) \leq 1$ , (iii)  $\nu(A_\alpha) = \chi_{A_\alpha} \in F_0, \alpha \in I$ , and (iv) for any  $A, B, C$  in  $\Sigma_0$ ,

$$(2) \quad \int_\Omega \chi_A (\nu(B) \vee \nu(C)) d\mu = \int_\Omega \nu(A) (\nu(B) \vee \nu(C)) d\mu.$$

When the above conditions hold, then  $\mathfrak{B}$  is given by  $\mathfrak{B} = \sigma\{A \in \Sigma: \nu(A) = \chi_A, \text{ a.e.}\}$ .

To prove Theorem 1 and to present the second result, the integral of a scalar function relative to a vector measure (or the Dunford-Schwartz, or D-S, integral, cf. [3, IV.10]) is needed. As noted after (1), if  $G: \Sigma \rightarrow \mathfrak{X}$ , a Banach space, is a vector measure which is  $\mu$ -continuous, then  $\int_A f dG$  is defined as the obvious sum for simple  $f$ . A function  $f \in L^p(\Sigma)$  is said to be integrable relative to  $G$  if there is a sequence of simple functions  $\{f_n, n \geq 1\} \subset L^p(\Sigma)$  such that  $f_n \rightarrow f$  a.e.  $[\mu]$ , and the sequence  $\{\int_A f_n dG, n \geq 1\} \subset \mathfrak{X}$  is Cauchy for each  $A \in \Sigma$ . Its limit  $\int_A f$  is unique and does not depend on the sequence and the mapping

$$(3) \quad \nu_f: A \rightarrow \int_A f dG = \lim_n \int_A f_n dG \in \mathfrak{X}, \quad A \in \Sigma,$$

is the vector integral, or the D-S integral. The work of [3] now shows that the mapping  $f \mapsto \int_\Omega f dG$  is linear and the dominated convergence theorem holds and  $\lim_{\mu(A) \rightarrow 0} \int_A f dG = 0$  so that  $\nu_f$  is then a vector measure into  $\mathfrak{X}$ . If  $\mathfrak{X}$  is also a lattice (e.g.,  $\mathfrak{X} = L^p(\mathfrak{B})$ ) and  $G \geq 0$ , then one can show that the monotone convergence theorem is again valid. Moreover, if  $\mathfrak{X} = L^p(\mathfrak{B})$ ,  $\rho$  is an a.c.n., and  $G = P^\mathfrak{B}$ , then each  $f \in L^p(\Sigma)$  is  $P^\mathfrak{B}$ -integrable in the above sense and  $E^\mathfrak{B}(f) = \int_\Omega f dP^\mathfrak{B}$ . With these properties, the following result holds.

**THEOREM 2.** Let  $L^p(\Sigma)$  be the function space with a g.w.u.  $F_0$ , the  $J$ -property and the a.c.n. as in the above theorem. Then a vector measure  $\nu: \Sigma_0 \rightarrow L^p(\Sigma)$  is a conditional probability  $P^\mathfrak{B}$  relative to a unique  $\sigma$ -field  $\mathfrak{B} \subset \Sigma$  with  $A_\alpha \in \mathfrak{B}$ ,  $\alpha \in I$ , iff (i)  $\nu(A) \geq 0$ ,  $A \in \Sigma_0$ , (ii)  $\rho'(\nu) \leq 1$ , (iii)  $\nu(A_\alpha) = \chi_{A_\alpha} \in F_0$ ,  $\alpha \in I$ , and (iv) for any  $A \in \Sigma_0$

$$(4) \quad \nu(A) = \int_\Omega \nu(A)(\omega) \nu(d\omega),$$

where the integral is taken in the D-S sense. When the above conditions hold, then  $\mathfrak{B}$  is generated by  $\{A \in \Sigma: \nu(A) = \chi_A \text{ a.e.}\}$ .

If  $L^p = L^p$ ,  $1 \leq p < \infty$ , or more generally,  $L^p = L^\Phi$  an Orlicz space with  $\Phi(2x) \leq C\Phi(x)$ ,  $x \geq 0$ , and  $\mu$  is  $\sigma$ -finite, the hypotheses of the above two theorems are automatic. However, the spaces admitted are more extensive.

### 3. Proofs.

**PROOF OF THEOREM 1.** If  $\nu = P^\mathfrak{B}$  for some  $\sigma$ -field  $\mathfrak{B} \subset \Sigma$  with  $A_\alpha \in \mathfrak{B}$ ,  $\alpha \in I$ , then the discussion in §1 clearly implies that conditions (i)–(iii) are true. ( $P^\mathfrak{B}$  exists since  $\{A_\alpha\}_{\alpha \in I} \subset \mathfrak{B} \Rightarrow \mu_\mathfrak{B}$  is strictly localizable.) Since  $P^\mathfrak{B}(A) = E^\mathfrak{B}(\chi_A)$ , by the standard properties of  $E^\mathfrak{B}$  ( $E^\mathfrak{B}$  is identity on  $L^p(\mathfrak{B})$  and is an averaging) yield

$$\int_\Omega \chi_A(\nu(B) \vee \nu(C)) d\mu = \int_\Omega E^\mathfrak{B}(\chi_A)(\nu(B) \vee \nu(C)) d\mu_\mathfrak{B}$$

which is (2). So consider the converse.

Let  $\nu$  be a vector measure satisfying the conditions of the theorem. Then the linear operator  $T: L^p(\Sigma) \rightarrow L^p(\Sigma)$  given by  $Tf = \int_\Omega f d\nu$ ,  $f \in L^p(\Sigma)$ , is

well defined where the integral is taken in the D-S sense, recalled above. Now (i) implies that  $T$  is positive, and by (ii) together with a property of the integral, and  $\rho'(\nu) \leq 1$ ,

$$(5) \quad \rho(Tf) \leq \rho(f)\rho'(\nu) \leq \rho(f)$$

so that  $T$  is a contraction. (This is easily proved first for simple  $f$  and then the general case follows by the Fatou property of  $\rho$ .) By (iii),  $T\chi_\alpha = \chi_\alpha \in F_0$ ,  $\alpha \in I$  (or  $TF_0 = F_0$ ). Since  $T\chi_A = \nu(A)$ , by (2) (in which both sides are finite, see below)

$$(6) \quad \int_{\Omega} \chi_A (T\chi_B \vee T\chi_C) d\mu = \int_{\Omega} \chi_A T^*(T\chi_B \vee T\chi_C) d\mu,$$

where  $T^*: (L^p(\Sigma))^* = L^{p'}(\Sigma) \rightarrow L^{p'}(\Sigma)$  is the adjoint of  $T$ . The a.c.n. of  $\rho$  and the localizability of  $\mu \Rightarrow (L^p(\Sigma))^* = L^{p'}(\Sigma)$ , and this is used here. It follows from (iii) that  $0 \leq \nu(A) \leq 1$  a.e. and then  $T: L^p \cap L^\infty(\Sigma) \rightarrow L^p \cap L^\infty(\Sigma)$  and  $\nu(A_\alpha) = \chi_{A_\alpha} \in L^p \cap L^{p'}(\Sigma)$ . So  $T\chi_A \in L^p \cap L^{p'}(\Sigma)$  and (6) makes sense. Since  $A \in \Sigma_0$  is arbitrary, (6) yields the important equation

$$(7) \quad T\chi_B \vee T\chi_C = T^*(T\chi_B \vee T\chi_C), \quad \text{a.e.}$$

On the other hand, taking  $C = \emptyset$  in (6) and extending it successively (first for simple functions) using the continuity of  $T$  one gets

$$(8) \quad \int_{\Omega} fTg d\mu = \int_{\Omega} (Tf)(Tg) d\mu, \quad f, g \in L^p \cap L^{p'} \cap L^\infty(\Sigma).$$

Taking  $g = \chi_A$ ,  $A \in \Sigma$ ,  $A \subset A_\alpha$  in (8) one has

$$(9) \quad \int_A Tf d\mu = \int_{\Omega} \chi_A Tf d\mu = \int_{\Omega} (T\chi_A)(Tf) d\mu, \quad f \in L^p \cap L^{p'} \cap L^\infty(\Sigma).$$

Since  $T\chi_A = \nu(A) \leq \nu(A_\alpha) = \chi_{A_\alpha}$  a.e., the following computation is valid:

$$(10) \quad \int_{\Omega} (T\chi_A)(Tf) d\mu = \int_{\Omega} fT\chi_A d\mu = \int_{\Omega} (T^*f)\chi_A d\mu = \int_A (T^*f) d\mu,$$

by (8). Using the fact that  $\cup_{\alpha} A_\alpha \doteq \Omega$ , it follows from (9) and (10) that  $Tf = T^*f$  a.e. This together with (7) yields  $(T\chi_B) \vee (T\chi_C) = T(T\chi_B \vee T\chi_C)$ ,  $B, C \in \Sigma_0$ , which is Šidák's identity for indicator functions on  $\Sigma_0$ . Now taking  $C = \emptyset$  and using the linearity of  $T$ , the preceding identity implies  $T^2f = Tf$  for simple  $f \in L^p(\Sigma) \cap L^{p'}(\Sigma)$ . By the a.c.n. property of  $\rho$  and (iii), such functions are dense in  $L^p(\Sigma)$ , so that one can deduce that  $T: L^p(\Sigma) \rightarrow L^p(\Sigma)$  is a positive contractive projection with  $TF_0 = F_0$ . The conclusion of Theorem 1 is now an immediate consequence of the following known result:

**PROPOSITION 3.** *Let  $L^p(\Sigma)$  be the function space with the properties given in Theorem 1. If  $T: L^p(\Sigma) \rightarrow L^p(\Sigma)$  is a positive contractive projection with  $TF_0 = F_0$ , relative to a g.w.u.  $F_0$ , then  $Tf = E^{\mathfrak{B}}(f)$ ,  $f \in L^p$ , for a unique  $\sigma$ -field  $\mathfrak{B} \subset \Sigma$ .*

If  $L^p = L^p$ ,  $1 \leq p < \infty$ ,  $\mu(\Omega) = 1$ , this was proved in [1] and if  $L^p = L^\Phi$ ,  $\Phi(2x) \leq C\Phi(x)$ ,  $x \geq 0$ ,  $\mu(\Omega) = 1$ , in [6]. The general extension is not difficult. It is found in [7, p. 356].

**PROOF OF THEOREM 2.** If  $\nu = P^{\mathfrak{B}}$  ( $A_\alpha \in \mathfrak{B}$ ,  $\alpha \in I$ ), then (i)–(iii) are again

clearly true. Regarding (4), it is a consequence of the equation

$$(11) \quad E^{\mathfrak{B}}(f) = E^{\mathfrak{B}}(E^{\mathfrak{B}}(f)) = \int_{\Omega} E^{\mathfrak{B}}(f) dP^{\mathfrak{B}}, \quad f \in L^p(\Sigma).$$

Taking  $f = \chi_A$ ,  $A \in \Sigma$ ,  $A \subset A_{\alpha}$ ,  $\alpha \in I$ , (iv) follows.

For the converse, let  $\nu: \Sigma_0 \rightarrow L^p(\Sigma)$  be a vector measure as given. Since  $\rho'(\nu) \leq 1$ , the D-S integral  $Tf = \int_{\Omega} f d\nu$  exists first for simple  $f$  and then generally, and  $\rho(Tf) \leq \rho(f)$ . Thus  $T$  is a positive contractive linear operator on  $L^p(\Sigma) \rightarrow L^p(\Sigma)$ . The result will follow from Proposition 3, as in the preceding proof, if it is shown that  $T^2 = T$ . Thus if  $f = \sum_{i=1}^n a_i \chi_{A_i}$ ,  $A_i \in \Sigma_0$ ,  $f \in L^p(\Sigma)$  (each  $A_i$  may be taken to be contained in some finite union of  $A_{\alpha}$ 's of  $F_0$ ):

$$(12) \quad \begin{aligned} T^2 f &= T(Tf) = \int_{\Omega} Tf d\nu = \int_{\Omega} \left[ \sum_{i=1}^n a_i \nu(A_i) \right] (\omega) \nu(d\omega) \\ &= \sum_{i=1}^n a_i \int_{\Omega} \nu(A_i)(\omega) \nu(d\omega) = \sum_{i=1}^n a_i \nu(A_i) = Tf, \end{aligned}$$

by (4) and the linearity of the D-S integral. Hence  $T^2 = T$  on a dense set, so that the same is true on  $L^p(\Sigma)$ . As noted before, this completes the proof.

**4. Remarks.** (1) If  $L^p = L^1$  then the hypothesis that  $\nu(A) \geq 0$  a.e. can be dropped in Theorem 2.

For, by the above proof  $T^2 = T$ ,  $TF_0 = F_0$  and  $\|T\| \leq 1$ . It is then easy to show that  $T$  is also positive. In particular, if the variation measure  $|\nu|(\cdot)$  of  $\nu$  satisfies  $|\nu| \leq \mu$ , then  $\rho'(\nu) \leq 1$  and the same conclusion obtains. (Otherwise some condition such as positivity is needed!)

(2) The above remark and Theorem 2 imply the following result.

**COROLLARY 4.** Let  $L^p(\Sigma)$  be as in Theorem 2, and  $\nu: \Sigma_0 \rightarrow L^p(\Sigma)$  a vector measure. Then  $\nu = P^{\mathfrak{B}}$  for a  $\mathfrak{B} \subset \Sigma$ , iff (i)  $|\nu|(A) \leq \mu(A)$ ,  $A \in \Sigma_0$ , (ii)  $\nu(A_{\alpha}) = \chi_{A_{\alpha}} \in F_0$ ,  $\alpha \in I$ , and (iii)  $\nu$  satisfies the functional equation (4). The  $\sigma$ -field  $\mathfrak{B}$  is given as in the theorem.

(3) Finally, it should be emphasized that an application of the vector (or D-S) integration theory is a natural part for work with conditional probabilities. Further, to a large extent, one need not assume any regularity hypotheses on these measures if the vector measure point of view is systematically used. The regularity condition comes in when one demands the vector integral, such as (4), to be the Lebesgue integral.

#### REFERENCES

1. T. Andô, *Contractive projections in  $L_p$ -spaces*, Pacific J. Math. **17** (1966), 391–405. MR **33** #566.
2. L. E. Dubins, *Conditional probability distributions in the wide sense*, Proc. Amer. Math. Soc. **8** (1957), 1088–1092. MR **21** #2295.
3. N. Dunford and J. T. Schwartz, *Linear operators. Part I: General theory*, Interscience, New York, 1958. MR **22** #8302.
4. N. E. Gretskey, *Representation theorems on Banach function spaces*, Mem. Amer. Math. Soc. No. 84 (1968).
5. M. P. Olson, *A characterization of conditional probability*, Pacific J. Math. **15** (1965), 971–983. MR **32** #3096.

6. M. M. Rao, *Conditional expectations and closed projections*, Nederl. Akad. Wetensch. Proc. Ser. A **68** = Indag. Math. **27** (1965), 100–112. MR **31** #4056.
7. ———, *Conditional measures and operators*, J. Multivariate Anal. **5** (1975), 330–413.
8. Z. Šidák, *On relations between strict-sense and wide-sense conditional expectations*, Teor. Verojatnost. i Primenen **2** (1957), 283–288 = Theor. Probability Appl. **2** (1957), 267–272. MR **19**, 1086.
9. A. C. Zaanen, *Integration*, 2nd rev. ed., North-Holland, Amsterdam; Interscience, New York, 1967. MR **36** #5286.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, RIVERSIDE, CALIFORNIA 92509