A REPRESENTATION FOR CONDITIONAL EXPECTATIONS GIVEN σ-LATTICES[1]

By Tim Robertson

University of Missouri and University of Iowa

1. Introduction. Conditional expectations are discussed by Brunk in [4] and [5]. (Generally, in the literature "conditional expectation" refers to "conditional expectation given a σ-field." However since we shall only be concerned with conditional expectations given σ-lattices, we shall use this abbreviated terminology for the latter, more general, concept.) As illustrated in these references, conditional expectations have been found to provide solutions for several maximum likelihood estimation problems. The principle result of this paper gives a representation for conditional expectations. Marshall and Proschan [7] and the author [9] have found such a representation for estimates useful in studying their asymptotic properties. Special cases of this representation theorem appear in other papers: Theorem 2.2 in Ayer, Brunk, Reid and Silverman [1] and Theorem 1 in Brunk [2] are instances in which the domain of the functions is finite. Brunk, Ewing and Utz [3] consider another version which we shall discuss in detail before proving the theorem.

Suppose we are given a totally finite measure space $(\Omega, \mathfrak{A}, \mu)$ and a σ-lattice $\mathfrak{L}$ of measurable subsets of Ω. A σ-lattice, by definition, contains both the null set $\varnothing$ and Ω and is closed under countable unions and intersections. The symbol $\mathfrak{L}^c$ will denote the σ-lattice of all subsets of Ω which are complements of members of $\mathfrak{L}$. We say that a random variable X is $\mathfrak{L}$-measurable provided $[X > a] \varepsilon \mathfrak{L}$ for each real number a. Let L_2 denote the class of square integrable random variables and $L_2(\mathfrak{L})$ the collection of all those members of L_2 which are $\mathfrak{L}$-measurable. Let $\mathfrak{B}$ denote the class of Borel subsets of the real line. We adopt the following definition for the conditional expectation, $E(X \mid \mathfrak{L})$, of X given $\mathfrak{L}$ (see Brunk [4]).

Definition. If $X \varepsilon L_2$ then $Y \varepsilon L_2(\mathfrak{L})$ is equal to $E(X \mid \mathfrak{L})$ if and only if Y has both of the following properties:

$$(1) \qquad \int (X - Y)Z \, d\mu \leqq 0 \quad \text{for each} \quad Z \varepsilon L_2(\mathfrak{L})$$

and

$$(2) \qquad \int_B (X - Y) \, d\mu = 0 \quad \text{for each} \quad B \varepsilon Y^{-1}(\mathfrak{B}).$$

(Brunk [4] shows that there is such a random variable Y associated with each $X \varepsilon L_2$ and that it is unique in the sense that if W is any other member of $L_2(\mathfrak{L})$ having these properties, then $\mu[Y \neq W] = 0$.)

Received 13 August 1965; revised 10 February 1966.

[1] This research forms part of the author's doctoral dissertation submitted to the Department of Statistics, University of Missouri. It was sponsored by the U. S. Army Research Office—Durham under contract number DA-31-124-ARO(D)-158.

A modified form of the theorem in this paper can be obtained without the assumption that $\mu(\Omega) < \infty$. In this case the definition of $E(X \mid \mathcal{L})$ is slightly more complicated (see Brunk [4]). Let $\mathcal{B}'$ denote the class of Borel sets of reals which exclude the origin. In case $(\Omega, \mathcal{A}, \mu)$ is not totally finite we obtain a definition for $E(X \mid \mathcal{L})$ by replacing (2) by

$$(3) \qquad B \varepsilon Y^{-1}(\mathcal{B}'), \qquad \mu(B) < \infty \Rightarrow \int_B (X - Y)\, d\mu = 0.$$

2. Results. We are assuming that our measure space is totally finite. Suppose $X \varepsilon L_2$ and $Y = E(X \mid \mathcal{L})$.

LEMMA. *Y has both of the following properties:*

$$(4) \qquad \int_{A \cdot B} (X - Y)\, d\mu \leqq 0 \quad \textit{for} \quad A \varepsilon \mathcal{L} \quad \textit{and} \quad B \varepsilon Y^{-1}(\mathcal{B})$$

and

$$(5) \qquad \int_{A \cdot B} (X - Y)\, d\mu \geqq 0 \quad \textit{for} \quad A \varepsilon \mathcal{L}^c \quad \textit{and} \quad B \varepsilon Y^{-1}(\mathcal{B}).$$

PROOF. Professor Brunk, in a personal communication, observes that for $A \varepsilon \mathcal{L}$ and $B \varepsilon Y^{-1}(\mathcal{B}) \cdot \mathcal{L}^c$ we have

$$(6) \qquad \int_{A \cdot B} (X - Y)\, d\mu = \int_B (X - Y)\, d\mu - \int_{A^c \cdot B} (X - Y)\, d\mu \leqq 0$$

since $\int_B (X - Y)\, d\mu = 0$ by (2) and $\int_{A^c \cdot B} (X - Y)\, d\mu \geqq 0$ by (1). (This result is similar to Corollary 3.2 of Brunk [5].) If B is the inverse image under Y of an interval $(a, b]$ then using (6)

$$\int_{A \cdot B} (X - Y)\, d\mu = \int_{A \cdot [Y > a] \cdot [Y \leqq b]} (X - Y)\, d\mu \leqq 0$$

since $A \cdot [Y > a] \varepsilon \mathcal{L}$ and $[Y \leqq b] \varepsilon Y^{-1}(\mathcal{B}) \cdot \mathcal{L}^c$. Property (4) then follows by first proving it provided B is the inverse image under Y of a finite disjoint union of intervals and then applying the generalized extension theorem which appears on page 90 of Loève [6]. Property (5) follows from (2) and (4).

REMARK. *The random variable $Y \varepsilon L_2(\mathcal{L})$ is equal to $E(X \mid \mathcal{L})$ if and only if*

$$(7) \qquad \int (X - Y) Z \Psi(Y)\, d\mu \leqq 0$$

for each non-negative Borel function Ψ and $\mathcal{L}$-measurable random variable Z such that $Z \cdot \Psi(Y) \varepsilon L_2$.

PROOF. First assume that $Y \varepsilon L_2(\mathcal{L})$ and that Y has property (7). Property (1) follows by taking $\Psi(x) \equiv 1$. Property (2) follows by applying (7) with Ψ the indicator function of a Borel set first with $Z \equiv 1$ and then $Z \equiv -1$.

Conversely suppose that $Y = E(X \mid \mathcal{L})$. Property (7) with Z the indicator function of a set in $\mathcal{L}$ follows from (4) by approximating Ψ by simple non-negative Borel functions. The case when Z is non-negative follows by approximating Z by simple random variables using methods similar to those used by the author [8] together with 3.13 and 3.16 in Brunk [5]. Similarly if Z is a non-negative, $\mathcal{L}^c$-measurable random variable such that $Z \cdot \Psi(Y) \varepsilon L_2$ we can show that $\int (X - Y) Z \Psi(Y)\, d\mu \geqq 0$ by starting with (5) instead of (4). The argument can then be completed by expressing an arbitrary Z as the difference of its positive and negative parts.

For each real number y let $P_y = [Y > y]$ and $T_y = [Y \geqq y]$. Define the set function α by

$$\alpha(A) = [\mu(A)]^{-1} \cdot \int_A X \, d\mu \quad \text{for each} \quad A \in \mathcal{A} \quad \text{such that} \quad \mu(A) > 0.$$

Before proving the theorem, which expresses Y in terms of such an averaging function, we discuss the following special case which appears in Brunk, Ewing and Utz [3]: Suppose Ω is n-dimensional Euclidean space R_n, $\mathcal{A}$ is a σ-field of subsets of Ω and μ is a totally finite measure on $\mathcal{A}$. If $u = (u_1, u_2, \cdots, u_n)$ and $v = (v_1, v_2, \cdots, v_n)$ are elements in Ω we say that $u \leqq v$ if $u_i \leqq v_i$ for each $i = 1, 2, \cdots, n$. This relation partially orders Ω. This partial ordering induces a σ-lattice of subsets of Ω in the following way. A subset of L of Ω is in $\mathcal{L}$ if and only if $L \in \mathcal{A}$ and if $u \in L$ and $u \leqq v$ imply that $v \in L$.

Their theorem states that if for each $\epsilon > 0$, $\mu(L - T_{y+\epsilon}) > 0$ for each $L \in \mathcal{L}$ containing ω and $\mu(P_{y-\epsilon} - L') > 0$ for each $L' \in \mathcal{L}$ not containing ω then

$$Y(\omega) = y = \sup_{L \in \mathcal{L}, L \ni \omega} \inf_{L' \in \mathcal{L}, L' \not\ni \omega} \alpha(L - L').$$

Our result applies to more general spaces Ω and needs less restrictive hypotheses than $\mu(L - T_{y+\epsilon}) > 0$ and $\mu(P_{y-\epsilon} - L') > 0$ for every L in $\mathcal{L}$ which contains ω and every L' in $\mathcal{L}$ which does not contain ω. Moreover, the theorem in Brunk, Ewing and Utz holds in particular when ω is a mass point of μ and our theorem holds more generally when μ assigns positive measure to $[Y = y]$.

THEOREM. *Fix $\omega \in \Omega$ and let $Y(\omega) = y$. Suppose for each $\epsilon > 0$, $\mu(T_y - T_{y+\epsilon}) > 0$ and $\mu(P_{y-\epsilon} - P_y) > 0$. Define the classes $\mathfrak{M}_1$ and $\mathfrak{M}_2$ by:*

$$\mathfrak{M}_1 = \{L; L \in \mathcal{L}, \mu(L - T_{y+\epsilon}) > 0 \text{ for each } \epsilon > 0\}$$

and

$$\mathfrak{M}_2 = \{L'; L' \in \mathcal{L}, \mu(P_{y-\epsilon} - L') > 0 \text{ for each } \epsilon > 0\}.$$

If $L \in \mathfrak{M}_1$ let

$$\mathfrak{N}_1(L) = \{L'; L' \in \mathcal{L}, \mu(L - L') > 0\}.$$

If $L' \in \mathfrak{M}_2$ let

$$\mathfrak{N}_2(L') = \{L; L \in \mathcal{L}, \mu(L - L') > 0\}.$$

By hypothesis $T_y \in \mathfrak{M}_1$ and $P_y \in \mathfrak{M}_2$. If $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are any two collections of subsets of Ω such that $T_y \in \mathfrak{R}_1 \subset \mathfrak{M}_1$ and $P_y \in \mathfrak{R}_2 \subset \mathfrak{M}_2$ then

$$\begin{aligned} y &= \sup_{L \in \mathfrak{R}_1} \inf_{L' \in \mathfrak{N}_1(L)} \alpha(L - L') \\ &= \inf_{L' \in \mathfrak{N}_1(T_y)} \alpha(T_y - L') \\ &= \inf_{L' \in \mathfrak{R}_2} \sup_{L \in \mathfrak{N}_2(L')} \alpha(L - L') \\ &= \sup_{L \in \mathfrak{N}_2(P_y)} \alpha(L - P_y). \end{aligned}$$

PROOF. First suppose $L \varepsilon \mathfrak{R}_1$. Then $L \varepsilon \mathfrak{M}_1$ and $\mu(L - T_{y+\epsilon}) > 0$ for each $\epsilon > 0$. Using (4) we conclude that

$$\int_{L-T_{y+\epsilon}} X \, d\mu \leqq \int_{L-T_{y+\epsilon}} Y \, d\mu < (y + \epsilon)\cdot\mu(L - T_{y+\epsilon}).$$

Hence for every $L \varepsilon \mathfrak{R}_1$ and $\epsilon > 0$, $T_{y+\epsilon} \varepsilon \mathfrak{N}_1(L)$ and $\alpha(L - T_{y+\epsilon}) < (y + \epsilon)$. It follows that

$$\text{(8)} \qquad \inf_{L' \varepsilon \mathfrak{N}_1(L)} \alpha(L - L') \leqq y \quad \text{for each} \quad L \varepsilon \mathfrak{R}_1 .$$

Now suppose $L' \varepsilon \mathfrak{N}_1(T_y)$. Then $\mu(T_y - L') > 0$ and by (5)

$$\int_{T_y - L'} X \, d\mu \geqq \int_{T_y - L'} Y \, d\mu \geqq y\cdot\mu(T_y - L').$$

We conclude that

$$\text{(9)} \qquad y \leqq \inf_{L' \varepsilon \mathfrak{N}_1(T_y)} \alpha(T_y - L').$$

Combining (8) and (9) it follows that

$$y = \sup_{L \varepsilon \mathfrak{R}_1} \inf_{L' \varepsilon \mathfrak{N}_1(L)} \alpha(L - L') = \inf_{L' \varepsilon \mathfrak{N}_1(T_y)} \alpha(T_y - L').$$

The remainder of the theorem follows by a similar argument.

Let $S = [Y = y]$, $\mathfrak{R}_1 = \{L; L \varepsilon \mathfrak{L}, S \subset L\}$ and $\mathfrak{R}_2 = \{L'; L' \varepsilon \mathfrak{L}, L' \subset P_y\}$.

COROLLARY. *If* $\mu(S) > 0$ *then*

$$y = \sup_{L \varepsilon \mathfrak{R}_1} \inf_{L' \varepsilon \mathfrak{N}_1(L)} \alpha(L - L') = \inf_{L' \varepsilon \mathfrak{R}_2} \sup_{L \varepsilon \mathfrak{N}_2(L')} \alpha(L - L').$$

In case μ is not totally finite our representative theorem is still valid if we add the hypothesis that $y - \delta > 0$ and $\mu[T_{y-\delta}] < \infty$ for some $\delta > 0$ and make other obvious changes in the theorem to insure that the sets involved have finite measure. The assumptions $y - \delta > 0$ and $\mu[T_{y-\delta}] < \infty$ do not seem unreasonable in most of the applications which have come to the author's attention. In most of these the random variable Y is non-negative and integrable. Hence the representation is only valid for $y > 0$.

3. Acknowledgment. The author wishes to thank Professor H. D. Brunk for suggesting the investigations in the course of which the subject matter of this paper arose and for his guidance and encouragement. He also wishes to thank the associate editor and the referee for their valuable comments.

REFERENCES

[1] AYER, MIRIAM, BRUNK, H. D., EWING, G. M., REID, W. T., and SILVERMAN, EDWARD (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

[2] BRUNK, H. D. (1955). Maximum likelihood estimation of monotone parameters. *Ann. Math. Statist.* **26** 607–616.

[3] BRUNK, H. D., EWING, G. M., and UTZ, W. R. (1957). Minimizing integrals in certain classes of monotone functions. *Pacific J. Math.* **7** 833–847.

[4] Brunk, H. D. (1963). On an extension of the concept conditional expectation. *Proc. Amer. Math. Soc.* **14** 298–304.

[5] Brunk, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

[6] Loève, Michel (1963). *Probability Theory* (Third Edition). Van Nostrand, New York.

[7] Marshall, Albert W. and Proschan, Frank (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Statist.* **36** 69–77.

[8] Robertson, Tim (1965). A note on the reciprocal of the conditional expectation of a positive random variable. *Ann. Math. Statist.* **36** 1302–1305.

[9] Robertson, Tim (1966). On estimating density which is measurable with respect to a σ-lattice. (Submitted to *Ann. Math. Statist.*)