A NOTE ON THE RECIPROCAL OF THE CONDITIONAL EXPECTATION OF A POSITIVE RANDOM VARIABLE[1]

BY TIM ROBERTSON

University of Missouri

1. Introduction and summary. Brunk [3] discusses conditional expectations given σ-lattices. This note is concerned with the observation that the reciprocal of the conditional expectation given a σ-lattice with respect to some measure μ of a positive random variable X is the conditional expectation of $1/X$ given the complementary σ-lattice with respect to another measure. (For the case of a σ-field this result is equivalent to the familiar result about the reciprocal of a Radon-Nikodym derivative.) Associated with this property of conditional expectations is a mapping of the set of all convex functions on $(0, \infty)$ into itself which leads to alternative ways of formulating certain extremum problems.

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space with $\mu(\Omega) < \infty$. Let I_A denote the indicator function of a set A. We shall let $\mathcal{L}$ denote a σ-lattice of subsets of $\Omega(\mathcal{L} \subset \mathcal{A})$. A σ-lattice is, by definition, closed under countable unions and intersections and contains both Ω and the null set $\varnothing$. If $\mathcal{L}$ is such a σ-lattice then $\mathcal{L}^c$ will denote the σ-lattice of all subsets of Ω which are complements of members of $\mathcal{L}$. We say that a random variable X is $\mathcal{L}$-measurable if $\{X > a\} \varepsilon \mathcal{L}$ for each real number a. Let L_2 denote the class of square integrable random variables and $L_2(\mathcal{L})$ the class of $\mathcal{L}$-measurable, square integrable random variables. Later we shall want to restrict our attention to strictly positive random variables, so for any set S of random variables we let S^+ denote the set of all those members of S which are strictly positive. Let $\mathcal{B}$ denote the class of Borel subsets of the real line. The following is one of several available definitions for the conditional expectation, $E_\mu(X \mid \mathcal{L})$, of X given $\mathcal{L}$ (see Brunk [3]).

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

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

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

(Brunk [3] shows the existence of such a Y and that it is unique in the sense that if W is any other member of $L_2(\mathcal{L})$ having these properties then $W = Y[\mu]$.)

2. Some equivalent definitions for $E_\mu(X \mid \mathcal{L})$. We may replace (2) in the definition of $E_\mu(X \mid \mathcal{L})$ by:

$$(3) \qquad \int (X - Y)\varphi(Y) \, d\mu = 0 \qquad \text{for every Borel function } \varphi \text{ of a real variable such that } \varphi(Y) \varepsilon L_2 .$$

This can be shown by approximating φ by simple random variables.

Received 6 November 1964; revised 17 February 1965.

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

We also remark that:

$$(4) \qquad \int_A (X - Y)\, d\mu \leqq 0 \qquad \text{for each } A \,\varepsilon\, \mathfrak{L}$$

may replace (1) in the definition of $E_\mu(X \mid \mathfrak{L})$. This can be shown by approximating Z by simple random variables and using (2) and (4) together with the identity

$$\sum_{i=1}^{k} a_i I_{A_i} = \sum_{i=1}^{k-1} (a_i - a_{i+1}) I_{B_i} + a_k I_{B_k}$$

where $a_1 > a_2 > \cdots > a_k$, $A_i = [Z = a_i]$, $B_k = \Omega$ and $B_j = \sum_{i=1}^{j} A_i$ $(j = 1, 2, \cdots, k-1)$.

3. The reciprocal of the conditional expectation of a positive random variable.

THEOREM 3.1. *Suppose* $X \,\varepsilon\, L_2^+$, $Y = E_\mu(X \mid \mathfrak{L})$ *and* $1/Y \,\varepsilon\, L_2$. *If the measure* γ *is defined by* $\gamma(A) = \int_A X\, d\mu$ *for each* $A \,\varepsilon\, \mathfrak{A}$ *then*

$$E_\gamma(1/X \mid \mathfrak{L}^c) = [E_\mu(X \mid \mathfrak{L})]^{-1}.$$

PROOF. Since $X \,\varepsilon\, L_2^+$ it follows from (2) that $Y \,\varepsilon\, L_2^+(\mathfrak{L})$. From this and the hypothesis we can conclude that $1/Y \,\varepsilon\, L_2^+(\mathfrak{L}^c)$. Suppose now that $B \,\varepsilon\, (1/Y)^{-1}(\mathfrak{B})$. Then B is also a member of $Y^{-1}(\mathfrak{B})$ and since X is the Radon-Nikodym derivative of γ with respect to μ we have:

$$\int_B (1/X - 1/Y)\, d\gamma = \int_B (1/X - 1/Y) X\, d\mu$$
$$= -\int (X - Y)(1/Y) I_B\, d\mu.$$

However, by (3) this integral must vanish so that:

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

It remains only to verify that $\int (1/X - 1/Y) I_A\, d\gamma \leqq 0$ for each $A \,\varepsilon\, \mathfrak{L}^c$. We have:

$$\int (1/X - 1/Y) I_A\, d\gamma = \int (X - Y)(-I_A/Y)\, d\mu.$$

But, I_A is non-negative, $1/Y$ is strictly positive and both are measurable $\mathfrak{L}^c$ so that $I_A/Y \,\varepsilon\, L_2(\mathfrak{L}^c)$ and $(-I_A/Y) \,\varepsilon\, L_2(\mathfrak{L})$. However, by (1) this integral is non-positive and by (4) the proof is complete.

4. $E_\mu(X \mid \mathfrak{L})$ as a solution to certain extremum problems. Suppose X is any member of L_2^+, $Y = E_\mu(X \mid \mathfrak{L})$ and $1/Y \,\varepsilon\, L_2$. Let Φ be any real valued function of a real variable whose domain includes $(0, \infty)$ and which is convex on this interval. Then for any $Z \,\varepsilon\, L_2^+$ we define the function $J_\Phi(Z; X, \mu)$ by:

$$(5) \qquad J_\Phi(Z; X, \mu) = \int [\Phi(X) - \Phi(Z) - (X - Z)\varphi(Z)]\, d\mu$$

where $\varphi(Z)$ denotes any derivative, say for the sake of definiteness, the left derivative of Φ at Z. We still require that μ be totally finite. $J_\Phi(Z; X, \mu)$ always exists as a positive real number or $+\infty$. Brunk [2] shows the following:

$$\min_{Z \varepsilon L_2^+(\mathcal{L})} J_\Phi(Z; X, \mu) = J_\Phi(Y; X, \mu). \tag{6}$$

However, by the same token, for any such convex Φ we have:

$$\min_{Z \varepsilon L_2^+(\mathcal{L}^c)} J_\Phi(Z; 1/X, \gamma) = J_\Phi[E_\gamma(1/X \mid \mathcal{L}^c); 1/X, \gamma].$$

But by Theorem 3.1, $E_\gamma(1/X \mid \mathcal{L}^c) = [E_\mu(X \mid \mathcal{L})]^{-1}$ so that $Y = E_\mu(X \mid \mathcal{L})$ solves another minimum problem (other than (6)), namely:

$$\min_{Z \varepsilon L_2^+(\mathcal{L})} J_\Phi(1/Z; 1/X, \gamma) \tag{7}$$

or equivalently:

$$\min_{Z \varepsilon L_2^+(\mathcal{L})} \int X[\Phi(1/X) - \Phi(1/Z) - (1/X - 1/Z)\varphi(1/Z)]\, d\mu.$$

The question arises: is this actually a new minimum problem or is the functional $H_\Phi(Z; X, \mu) = J_\Phi(1/Z; 1/X, \gamma)$ of the form $J_{\Phi^*}(Z: X, \mu)$ for some other convex function Φ^*. The answer is the latter. Let $\Phi^*(\lambda) = \lambda\Phi(1/\lambda)$. It is easily verified that Φ^* is also convex on $(0, \infty)$ and that:

$$\Phi^*(X) - \Phi^*(Z) - (X - Z)\varphi^*(Z)$$
$$= X[\Phi(1/X) - \Phi(1/Z) - (1/X - 1/Z)\varphi(1/Z)].$$

Note that since Φ is convex if and only if $-\Phi$ is concave and since $J_{-\Phi}(Z; X, \mu) = -J_\Phi(Z; X, \mu)$, (6) implies that for any concave Θ:

$$\max_{Z \varepsilon L_2^+(\mathcal{L})} J_\Theta(Z; X, \mu) = J_\Theta(Y, X, \mu). \tag{8}$$

Let θ denote the left derivative of Θ. Then $\int(X - Y)\theta(Y)\, d\mu = 0$ and if we let M denote the class of all those members Z of $L_2^+(\mathcal{L})$ for which

$$\int (X - Z)\theta(Z)\, du = 0$$

and $\Theta(Z)$ is integrable then

$$\max_{Z \varepsilon M} [-\int \Theta(Z)\, d\mu] = -\int \Theta(Y)\, d\mu. \tag{9}$$

It is assumed that $\Theta(X)$ and $\Theta(Y)$ are integrable.

To illustrate these comments suppose we are given two k-tuples $(n_1, n_2, \cdots, n_k)$ and $(a_1, a_2, \cdots, a_k)$ of positive real numbers and let $n_1 + n_2 + \cdots + n_k = n$. In certain estimation problems (for example see [4]) we wish to find that k-tuple $(p_1, p_2, \cdots, p_k)$, if it exists, of positive real numbers which satisfies

$$p_1 \geqq p_2 \geqq \cdots \geqq p_k \tag{10}$$

$$a_1p_1 + a_2p_2 + \cdots + a_kp_k = 1 \tag{11}$$

and $p_1^{n_1} \cdot p_2^{n_2} \cdot \cdots \cdot p_k^{n_k} \geqq r_1^{n_1} \cdot r_2^{n_2} \cdot \cdots \cdot r_k^{n_k}$ for every other k-tuple $(r_1, r_2, \cdots, r_k)$ of positive real numbers which satisfies both (10) and (11).

Let $\Omega = \{1, 2, \cdots, k\}$, $\mathcal{A}$ be the collection of all subsets of Ω and $\mathcal{L}$ be the σ-lattice of left intervals in Ω:

$$\mathcal{L} = \{\phi, \{1\}, \{1, 2\}, \cdots, \{1, 2, \cdots, k\}\}.$$

The k-tuples $p = (p_1, p_2, \cdots, p_k)$ of reals can be thought of as functions on Ω and p satisfies (10) if and only if p is measurable $\mathcal{L}$.

Then using $\Theta(\lambda) = \log(\lambda)$ in (9) with X and μ defined on Ω and $\mathcal{A}$ respectively by $X(i) = n \cdot a_i/n_i$ and $\mu(i) = n_i/n$ it follows from the above remarks that the solution for Z is $E_\mu(X \mid \mathcal{L}^c)$, and hence for $1/p$, $[E_\mu(X \mid \mathcal{L}^c)]^{-1}$. By Theorem 3.1, this coincides with $E_\gamma(1/X \mid \mathcal{L})$. In this approach to the problem we made use of the assumption that n_i's were non-zero.

In some applications some of the n_i's might be zero. Considerations mentioned above suggest rephrasing the problem in terms of $\Theta(\lambda) = \lambda \log(1/\lambda)$. With this approach using (8) instead of (9) our solution is given by $E_\gamma(Y \mid \mathcal{L})$ where Y and γ are defined by $Y(i) = n_i/n \cdot a_i$ and $\gamma(i) = a_i$. Here we need to assume that the a_i's are non-zero.

Representations of $E(X \mid \mathcal{L})$ and methods for calculating it are given in [1], in [5] and in [4] where an elegant graphical interpretation is given.

5. Acknowledgments. The author wishes to thank Professor H. D. Brunk for suggesting the investigations in the course of which the subject matter of this note arose and for his critical comments and valuable suggestions. He also wishes to thank Professor Robert Berk for his 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. (1961). Best fit to a random variable by a random variable measurable with respect to a σ-lattice. *Pacific J. Math.* **11** 785–802.

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

[4] GRENANDER, ULF (1956). On the theory of mortality measurement, Part II. *Skand. Aktuarietidskr.* **39** 126–153.

[5] VAN EEDEN, CONSTANCE (1957). Maximum likelihood estimation of partially or completely ordered parameters. *Nederl. Akad. Wetensch. Proc. Ser. A* **60** 128–136; *Indag. Math.* **19** 201–211.