CONDITIONAL EXPECTATION GIVEN A σ-LATTICE AND APPLICATIONS

By H. D. Brunk

University of Missouri

1. Introduction. The applications discussed are applications of an extremizing property of the conditional expectation given a σ-lattice. It has been interesting to the author to observe that a variety of problems of maximum likelihood estimation of parameters, of functions, of probability densities, lead to a class of extremum problems of which the conditional expectation given a σ-lattice furnishes the solution. By a σ-lattice is understood a family of sets closed under countable union and countable intersection, but not necessarily difference or complement. In extending in a natural way to σ-lattices [11] the concept of conditional expectation given a σ-field, one sacrifices a very useful property indeed, linearity. On the other hand, much carries over; in particular, some theorems of martingale type are proved in Sections 2 and 3.

2. Projection on a closed convex set in a complete inner product space. Let H be a complete inner product space: a complete metric space with distance between elements X, Y given by $\|X - Y\|$, where $\|Z\| =_D (Z, Z)^{\frac{1}{2}}$ for $Z \varepsilon H$, and where $(\cdot, \cdot)$ denotes the inner product. If $Y \varepsilon H$, $Z \varepsilon H$, then seg YZ will denote the segment joining Y and Z; i.e., seg $YZ =_D \{W: W \varepsilon H, \exists \alpha \ni 0 \leqq \alpha \leqq 1$ and $W = (1 - \alpha)Y + \alpha Z\}$. A subset A of H will be termed closed if it is closed in the topology of the metric above. The set A is convex if $Y \varepsilon A, Z \varepsilon A \Rightarrow$ seg YZ $\subset A$.

The following existence theorem is well known.

THEOREM 2.1. *Let A be a closed convex set in H, and let $X \varepsilon H$. Then there is a unique closest point Y of A to X.* For let $\{Y_n\}$ be a sequence of points of A whose distances from X approach the infimum of such distances, and set $U_n = Y_n - X$. From the fact that $(Y_n + Y_m)/2 \varepsilon A$ for all n and m and from the equation

$$\|U_n - U_m\|^2 = 2(\|U_n\|^2 + \|U_m\|^2) - \|U_n + U_m\|^2$$

it follows that $\{Y_n\}$ is Cauchy and hence has a limit $Y \varepsilon A$. Uniqueness also follows immediately from the convexity of A and the identity

$$\|(Y_1 + Y_2)/2 - X\|^2 = (\|Y_1 - X\|^2 + \|Y_2 - X\|^2)/2 - \|Y_1 - Y_2\|^2/4.$$

The notation $P(X \mid A)$ will be used for the closest point of a closed convex set A to the point X.

THEOREM 2.2. *If $X \varepsilon H$ and if A is a closed convex subset of H, then $Y = P(X \mid A)$ if and only if*

$$(2.1) \qquad (X - Y, Y - Z) \geqq 0 \quad \textit{for all} \quad Z \varepsilon A.$$

PROOF. This is immediate from Theorem 2.1 and the observation that if

Received 5 April 1965.

$U \varepsilon H$, $V \varepsilon H$, then Y is the closest point of seg UV to X if and only if $(X - Y, Y - Z) \geqq 0$ for all $Z \varepsilon$ seg UV.

COROLLARY 2.1. *If $X \varepsilon H$ and if A is a closed convex subset of H then the element Y of A is $P(X \mid A)$ if and only if*

$$\|X - Z\|^2 \geqq \|X - Y\|^2 + \|Y - Z\|^2. \tag{2.2}$$

THEOREM 2.3. *Projection on a closed convex set is a continuous, indeed, a distance reducing, operator. That is, if $X_i \varepsilon H$, $i = 1, 2$, if A is a closed convex subset of H, and if $Y_i = P(X_i \mid A)$, $i = 1, 2$, then $\|Y_2 - Y_1\| \leqq \|X_2 - X_1\|$.*

PROOF. For by Theorem 2.2, $(X_1 - Y_1, Y_1 - Y_2) \geqq 0$ and $(X_2 - Y_2, Y_2 - Y_1) \geqq 0$. Then

$$\begin{aligned}\|X_2 - X_1\|^2 &= \|X_2 - Y_2 + Y_2 - Y_1 + Y_1 - X_1\|^2 \\ &= \|X_2 - Y_2 + Y_1 - X_1\|^2 + \|Y_2 - Y_1\|^2 + 2(X_2 - Y_2, Y_2 - Y_1) \\ &\quad + 2(X_1 - Y_1, Y_1 - Y_2) \\ &\geqq \|Y_2 - Y_1\|^2.\end{aligned}$$

(This simple proof was suggested by D. L. Hanson.)

COROLLARY 2.2. *If $0 \varepsilon A$ then $\|P(X \mid A)\| \leqq \|X\|$.*

A subset A of H is a *cone* if $X \varepsilon A$ and $k \geqq 0 \Rightarrow kX \varepsilon A$.

COROLLARY 2.3. *If $X \varepsilon H$ and if A is a closed convex cone in H, then $Y = P(X \mid A)$ if and only if*

$$(X - Y, Y) = 0 \tag{2.3}$$

and

$$(X - Y, Z) \leqq 0 \quad \text{for all} \quad Z \varepsilon A. \tag{2.4}$$

This follows from Theorem 2.2 on replacing Z by kY; first, with $0 < k < 1$, then with $k > 1$.

The following is a mean-square convergence theorem of martingale type: one might call a sequence $\{X_n, A_n\}$ a martingale if, for all n, A_n is a closed convex subset of H, $X_n \varepsilon H$, $A_n \subset A_{n+1}$, and for $k \geqq 0$, $X_n = P(X_{n+k} \mid A_n)$.

THEOREM 2.4. *Let $\{A_n\}$, $n = 0, \pm 1, \pm 2, \cdots$, be a two-way sequence of closed convex sets in H such that $A_n \subset A_{n+1}$, $n = 0, \pm 1, \pm 2, \cdots$. Suppose $\{X_n\}$ is a bounded sequence, $\|X_n\| \leqq M$ for some positive constant M, $n = 0, \pm 1, \pm 2, \cdots$, of elements of H such that $X_n = P(X_{n+k} \mid A_n)$, $n = 0, \pm 1, \pm 2, \cdots, k \geqq 0$. Then $\lim_{n \to -\infty} X_n = X_{-\infty}$ and $\lim_{n \to \infty} X_n = X_\infty$ exist, $X_n = P(X_\infty \mid A_n)$, $n = 0, \pm 1, \pm 2, \cdots$, and $X_{-\infty} = P(X_n \mid \bigcap_k A_k)$, for all n.*

PROOF. Suppose $\{X_n\}$, $n = 0, \pm 1, \pm 2, \cdots$, is not a Cauchy sequence. Then there exist $\epsilon > 0$ and pairs of integers $n_1 < n_1' < n_2 < n_2' < \cdots$ such that $\|X_{n_j'} - X_{n_j}\| \geqq \epsilon$, $j = 1, 2, \cdots$. Since $X_{n_j} \varepsilon A_{n_j'}$ and $X_{n_j'} = P(X_{n_{j+1}} \mid A_{n_j'})$, we have from Corollary 2.1 that $\|X_{n_{j+1}} - X_{n_j}\|^2 \geqq \|X_{n_j'} - X_{n_j}\|^2 \geqq \epsilon^2$.

Using induction with Corollary 2.1, we have

$$4M^2 \geqq \|X_{n_{r+1}} - X_{n_1}\|^2 \geqq \sum_{i=1}^{r} \|X_{n_{i+1}} - X_{n_i}\|^2 \geqq r\epsilon^2,$$

a contradiction for sufficiently large r. Thus $\{X_n\}$, $n = 0, 1, 2, \cdots$, is Cauchy, and similarly $\{X_n\}$, $n = 0, -1, -2, \cdots$, is Cauchy, so that $\lim_{n\to-\infty} X_n = X_{-\infty}$ and $\lim_{n\to\infty} X_n = X_\infty$ exist. From the validity for fixed k and $Z \varepsilon A_k$ of $\|X_n - Z\|^2 \geqq \|X_n - X_k\|^2 + \|X_k - Z\|^2$ for all $n \geqq k$ we have

$$\|X_\infty - Z\|^2 \geqq \|X_\infty - X_k\|^2 + \|X_k - Z\|^2$$

so that $X_k = P(X_\infty \mid A_k)$, $k = 0, \pm1, \pm2, \cdots$. Also $\|X_n - Z\|^2 \geqq \|X_n - X_{-\infty}\|^2 + \|X_{-\infty} - Z\|^2$ for all n and for all $Z \varepsilon \bigcap_{k=-\infty}^{\infty} A_k$, so that $X_{-\infty} = P(X_n \mid \bigcap_k A_k)$. This completes the proof of Theorem 2.4.

A conclusion of Theorem 2.4 is that every bounded martingale is of the form $\{P(X \mid A_n)\}$ for some $X \varepsilon H$. On the other hand, if $X_n = P(X \mid A_n)$ for all n, it does not follow that $X_n = P(X_{n+k} \mid A_n)$ for $k \geqq 0$, so that Theorem 2.5 below is more general.

THEOREM 2.5. *Let $\{A_n\}$, $n = 0, \pm1, \pm2, \cdots$, be a two-way sequence of closed convex sets in H such that $A_n \subset A_{n+1}$, $n = 0, \pm1, \pm2, \cdots$. Let $X \varepsilon H$ and set $X_n = P(X \mid A_n)$, $n = 0, \pm1, \pm2, \cdots$. Suppose there are an integer N and a positive number M such that $\|X_n\| \leqq M$ for $n \leqq N$. Then $\lim_{n\to\infty} X_n = X_\infty$ and $\lim_{n\to-\infty} X_n = X_{-\infty}$ exist. Also $X_\infty = P(X \mid A_\infty)$ and $X_{-\infty} = P(X_n \mid A_{-\infty})$ for all n, where A_∞ is the closure of $\bigcup_{k=0}^{\infty} A_k$ and $A_{-\infty} = \bigcap_{k=-\infty}^{0} A_k$.*

The theorem applies, of course, to one-way sequences, using the device of setting all sets with sufficiently large index or sufficiently small index equal to each other. We remark that for one-way expanding sequences of sets the boundedness condition is no restriction.

PROOF. From Corollary 2.1 we have

$$\|X - X_m\|^2 \geqq \|X - X_n\|^2 + \|X_n - X_m\|^2 \quad \text{for} \quad n \geqq m.$$

Thus $\{\|X - X_n\|\}$ is non-increasing as n increases. Further, $\|X_n\| \leqq M$ for $n \leqq N$, so that $\|X - X_n\| \leqq \|X\| + M$ for $n \leqq N$. It follows that $W_\infty = \lim_{n\to\infty} \|X - X_n\|$ and $W_{-\infty} = \lim_{n\to-\infty} \|X - X_n\|$ exist as non-negative numbers. From the inequality $\|X_n - X_m\|^2 \leqq \|X - X_m\|^2 - \|X - X_n\|^2$ for $n \geqq m$ it follows that the sequences $\{X_n\}_{n=0}^{\infty}$ and $\{X_n\}_{n=0}^{-\infty}$ are both Cauchy, and the limits X_∞ and $X_{-\infty}$ exist.

Further, if $Z \varepsilon \bigcup_{k=0}^{\infty} A_k$, then there is an integer m such that $Z \varepsilon A_m$ so that $\|X - Z\|^2 \geqq \|X - X_n\|^2 + \|X_n - Z\|^2$ for $n \geqq m$. Thus $\|X - Z\|^2 \geqq \|X - X_\infty\|^2 + \|X_\infty - Z\|^2$ for $Z \varepsilon \bigcup_{k=0}^{\infty} A_k$ and hence for $Z \varepsilon A_\infty$. Also $X_\infty \varepsilon A_\infty$, so that it is identified by Corollary 2.1 with $P(X \mid A_\infty)$.

Finally, if $Z \varepsilon A_{-\infty}$ then for all n, $\|X - Z\|^2 \geqq \|X - X_n\|^2 + \|X_n - Z\|^2$, hence $\|X - Z\|^2 \geqq \|X - X_{-\infty}\|^2 + \|X_{-\infty} - Z\|^2$. Since $X_{-\infty} \varepsilon A_{-\infty}$ we have again $X_{-\infty} = P(X \mid A_{-\infty})$.

COROLLARY 2.4. *Under the hypotheses of Theorem 2.5, if in addition $U_n \to X$ as $n \to \infty$ then $P(U_n \mid A_n) \to X_\infty$ as $n \to \infty$. If $V_n \to X$ as $n \to -\infty$ then $P(V_n \mid A_n) \to X_{-\infty}$ as $n \to -\infty$.*

PROOF. Set $U_n' = P(U_n \mid A_n)$, $V_n' = P(V_n \mid A_n)$. Then $\|U_n' - X_\infty\| \leqq \|U_n' - X_n\| + \|X_n - X_\infty\|$. But by Theorem 2.3, $\|U_n' - X_n\| \leqq \|U_n - X\|$, and the first conclusion follows. The second follows on replacing U_n and U_n' by V_n and V_n' respectively.

We exhibit now some additional properties of projection on closed convex cones, consequent upon Corollary 2.3. We note that while $P(\cdot \mid A)$ is not necessarily linear, for a closed convex cone A it is positively homogeneous:

THEOREM 2.6. *If A is a closed convex cone, if $k \geqq 0$, and if $X \varepsilon H$ then $P(kX \mid A) = kP(X \mid A)$.*

If A is a closed convex cone, we denote by $-A$ the set of its negatives: $-A \equiv_D \{X: -X \varepsilon A\}$. One verifies immediately also the following.

THEOREM 2.7. *If A is a closed convex cone, and if $X \varepsilon H$, then $P(-X \mid -A) = -P(X \mid A)$.*

The following theorem and corollaries of this section generalize known properties of projection on (linear) subspaces; cf. e.g. [1], Section 33, which contains also the restriction of Theorem 2.5 to projections on subspaces.

THEOREM 2.8. *If A_i is a closed convex cone, $i = 1, 2$, if $A_1 \subset A_2$, and if $X \varepsilon H$ then*

$$\|P(X \mid A_2\|^2 - \|P(X \mid A_1)\|^2 = (P(X \mid A_2) - P(X \mid A_1), X)$$
$$\geqq \|P(X \mid A_2) - P(X \mid A_1)\|^2.$$

PROOF. Set $Y_2 = P(X \mid A_2)$, $Y_1 = P(X \mid A_1)$. The equation $\|Y_2\|^2 - \|Y_1\|^2 = (Y_2 - Y_1, X)$ is immediate from Corollary 2.3. For the inequality, we have

$$\begin{aligned}(Y_2 - Y_1, X) &= (Y_2 - Y_1, X - Y_2) + (Y_2 - Y_1, Y_2 - Y_1) + (Y_2 - Y_1, Y_1)\\ &= -(Y_1, X - Y_2) + \|Y_2 - Y_1\|^2 + (Y_2 - Y_1, Y_1)\\ &= \|Y_2 - Y_1\|^2 - 2(X - Y_2, Y_1),\end{aligned}$$

since $(Y_1, Y_1) = (X, Y_1)$. From Corollary 2.3 we have $(X - Y_2, Y_1) \leqq 0$, so that $(Y_2 - Y_1, X) \geqq \|Y_2 - Y_1\|^2$.

COROLLARY 2.5. *If A_i is a closed convex cone, $i = 1, 2$, and if $A_1 \subset A_2$ then*

$$(P(X \mid A_2), X) \geqq (P(X \mid A_1), X), \quad \text{for all} \quad X \varepsilon H.$$

This property may be expressed by the notation $P(\cdot \mid A_2) \geqq P(\cdot \mid A_1)$.

COROLLARY 2.6. *If A is a closed convex cone and if $X \varepsilon H$ then $(P(X \mid A), X) \leqq \|X\|^2$.*

This property may be expressed as follows: $\|P(\cdot \mid A\| \leqq 1$, where $\|P(\cdot \mid A)\| = \sup_{0 \neq X \varepsilon H} [(P(X \mid A), X)/(X, X)]$. Since $(P(X \mid A), X) = \|P(X \mid A)\|^2$, Corollary 2.6 is a restatement for closed convex cones of Corollary 2.2.

3. Conditional expectation given a σ-lattice. Let $(\Omega, \mathfrak{A}, \mu)$ be a measure space: Ω an abstract set, $\mathfrak{A}$ a σ-field of subsets of Ω, and μ a measure on $\mathfrak{A}$. We shall use the term random variable for an equivalence class of $\mathfrak{A}$-measurable functions, two such representing the same random variable if they differ on a set of measure

0. Let L_1 denote the class of integrable random variables and L_2 the class of random variables whose squares are integrable. Let $\mathfrak{M}$ be a sub-σ-lattice of $\mathfrak{A}$: $\mathfrak{M}$ is closed under countable union and countable intersection, but not necessarily under complementation. Set $\mathfrak{M}^c = \{A: A^c \varepsilon \mathfrak{M}\}$. Let $R(\mathfrak{M})$ denote the class of random variables X such that $[X > a] \varepsilon \mathfrak{M}$ for all real a. (Here $[X > a]$ denotes the equivalence class of sets in $\mathfrak{A}$ differing by sets of measure 0 from the set on which a function representing X is greater than a.) We note that for such X also $[X \geqq a] \varepsilon \mathfrak{M}$ for all real a, and that $X \varepsilon R(\mathfrak{M}) \Leftrightarrow -X \varepsilon R(\mathfrak{M}^c) \Leftrightarrow [X \leqq a] \varepsilon \mathfrak{M}^c$ for all real a. We set $L_i(\mathfrak{M}) = R(\mathfrak{M}) \cap L_i$, $i = 1, 2$; then $L_i = L_i(\mathfrak{A})$, $i = 1, 2$, and the class of all random variables is $R(\mathfrak{A})$. If $M \subset R(\mathfrak{A})$ we write $-M = \{X: -X \varepsilon M\}$, and observe that $R(\mathfrak{M}^c) = -R(\mathfrak{M})$.

It is shown in [11] that a family F of random variables is $R(\mathfrak{M})$ for some sub-σ-lattice $\mathfrak{M}$ of $\mathfrak{A}$ which contains $\varnothing$ and Ω if and only if

(3.1) F is a conditional σ-lattice;

i.e., $X \varepsilon R(\mathfrak{A})$, $X_n \varepsilon F$, $X_n \leqq X$ for $n = 1, 2, \cdots$, imply $\bigvee_n X_n \varepsilon F$; and $X \varepsilon R(\mathfrak{A})$, $X_n \varepsilon F$, $X_n \geqq X$ for $n = 1, 2, \cdots$ imply $\bigwedge_n X_n \varepsilon F$;

(3.2) each constant random variable is in F;

(3.3) F is a convex cone.

LEMMA 3.1. *Let $\mathfrak{M}$ be a sub-σ-lattice of $\mathfrak{A}$. If ϕ is a nondecreasing function on the reals, and if $Z \varepsilon R(\mathfrak{M})$, then $\phi(Z) \varepsilon R(\mathfrak{M})$. If ϕ is nonincreasing then $\phi(Z) \varepsilon R(\mathfrak{M}^c)$. If $W \geqq 0$, $W \varepsilon R(\mathfrak{M})$, $Z \varepsilon R(\mathfrak{M})$, then $WZ \varepsilon R(\mathfrak{M})$. If $W \leqq 0$, $W \varepsilon R(\mathfrak{M}^c)$, $Z \varepsilon R(\mathfrak{M})$, then $WZ \varepsilon R(\mathfrak{M}^c)$.*

The first two statements are immediate. If $W > 0$, the next follows from the identity, for real a:

$$[WZ > a] = \cup [W > r][Z > a/r],$$

the union being extended over positive rationals r. For $W \geqq 0$ and for positive integers n, we set $W_n = W \vee (1/n)$, and observe that $WZ = \bigwedge_n W_n(Z \vee 0) + \bigwedge_n W_n(Z \wedge 0) \varepsilon R(\mathfrak{M})$ by (3.1) and (3.3). The last statement, for $W < 0$, follows from $[WZ > a] = \cup[W > -r][Z < -a/r]$, the union again being extended over positive rationals r; the extension for $W \leqq 0$ is similar to that above for $W \geqq 0$.

It is also shown in [11] that $M = L_2(\mathfrak{M})$ for some sub-σ-lattice $\mathfrak{M}$ of $\mathfrak{A}$ containing $\varnothing$ and Ω if and only if

(3.4) M is a lattice closed in L_2;

(3.5) $X \varepsilon M \Rightarrow L_2 \cap \{I([X > a]): a \text{ is real}\} \subset M$

and $L_2 \cap \{I([X < a]): a \text{ is real}\} \subset -M$;

(3.6) M is a convex cone.

If $\mu(\Omega) < \infty$ then (3.5) may be replaced by:

(3.7) each constant random variable is in M.

The space L_2 becomes a complete inner product space (Hilbert space, if infinite dimensional) with the introduction of the inner product $(X, Y) = \int XY\, d\mu$. Thus the results of Section 2 apply. If $M = L_2(\mathfrak{M})$, we shall denote by $E(X \mid \mathfrak{M})$ the closest point of M to the random variable $X \varepsilon L_2$; i.e., $E(X \mid \mathfrak{M}) = P(X \mid M)$.

As applied in this context, Corollary 2.3 states that $Y = E(X \mid \mathfrak{M})$ if and only if

$$(3.8) \qquad \int (X - Y)Y\, d\mu = 0$$

and

$$(3.9) \qquad \int (X - Y)Z\, d\mu \leqq 0 \quad \text{for all} \quad Z \varepsilon L_2(\mathfrak{M}).$$

REMARK 3.1. It was observed above that if $\mu(\Omega) < \infty$ then $L_2(\mathfrak{M})$ contains the constant random variables; (3.9) then implies $\int (X - Y)\, d\mu = 0$. In case $(\Omega, \mathfrak{A}, \mu)$ is a probability space, (3.8) and (3.9) may be interpreted as follows:

If $V(X) < \infty$, then $Y = E(X \mid \mathfrak{M})$ if and only if Y is $\mathfrak{M}$-measurable, $V(Y) < \infty$, $E(Y) = E(X)$, cov $(X, Y) = V(Y)$ and cov $(X, Z) \leqq$ cov (Y, Z) for all $\mathfrak{M}$-measurable Z such that $V(Z) < \infty$; here $V(\cdot)$ denotes variance. It follows that the variance of Y is not greater than that of X, and that for every $Z \varepsilon L_2(\mathfrak{M})$, the correlation coefficient of X and Z is not greater than that of Y and Z.

Also, a direct interpretation of Corollary 2.1 in the case of a probability space is:

$Y = E(X \mid \mathfrak{M})$ if and only if $Y \varepsilon L_2(\mathfrak{M})$ and $V(X - Z) \geqq V(X - Y) + V(Y - Z)$ for all $Z \varepsilon L_2(\mathfrak{M})$.

REMARK 3.2. $E(-X \mid \mathfrak{M}^c) = -E(X \mid \mathfrak{M})$. (Cf. Theorem 2.7.)

Let $\mathfrak{B}'$ denote the class of Borel sets of reals which exclude the origin. The following theorem is given in [11].

THEOREM 3.1. *If $\mathfrak{M}$ is a σ-lattice, $\mathfrak{M} \subset \mathfrak{A}$, and if $X \varepsilon L_2$, then $Y = E(X \mid \mathfrak{M})$ is the unique random variable in $L_2(\mathfrak{M})$ satisfying*

$$(3.10) \qquad \int XZ\, d\mu \leqq \int YZ\, d\mu \quad \textit{for all } Z \varepsilon L_2(\mathfrak{M})$$

and

$$(3.11) \qquad B \varepsilon Y^{-1}(\mathfrak{B}'), \qquad \mu(B) < \infty \Rightarrow \int_B X\, d\mu = \int_B Y\, d\mu.$$

If also $X \varepsilon L_1$ then $Y \varepsilon L_1$ and

$$(3.12) \qquad B \varepsilon Y^{-1}(\mathfrak{B}') \Rightarrow \int_B X\, d\mu = \int_B Y\, d\mu.$$

If $\mu(\Omega) < \infty$, $\mathfrak{B}'$ may be replaced by $\mathfrak{B}$, the class of all Borel sets of reals.

That the condition $\mu(B) < \infty$ may not in general be deleted from (3.11) is clear, for the integrals need not then exist. That $\mathfrak{B}'$ may not be replaced by $\mathfrak{B}$ in (3.11) and (3.12), even in the σ-finite case, is clear from the following example. Set $\Omega = [0, 1]$, $\mu\{0\} = 1$, $\mu(A) = \int_A dt/t$ for Borel $A \subset (0, 1]$. Let $\mathfrak{M}$ denote the class of intervals $(a, 1]$ or $[a, 1]$ with $0 \leqq a \leqq 1$. Then $R(\mathfrak{M})$ is the class of nondecreasing functions on [0, 1], and each function $f \varepsilon L_2(\mathfrak{M})$ satisfies $f(0) \leqq 0$, since $f \in L_2(\mathfrak{M}) \Rightarrow f(x) \downarrow 0$ as $x \downarrow 0$. Let $X(0) = 1$, $X(t) = t$ on $(0, 1]$, $Y = E(X \mid \mathfrak{M})$. Then $Y(t) = X(t)$ for $t > 0$, and $Y(0) = 0$; thus $\int_{[Y=0]} X \, d\mu \neq \int_{[Y=0]} Y \, d\mu$ even though $\mu[Y = 0] < \infty$.

COROLLARY 3.1. *Under the hypotheses of Theorem* 3.1,

$$\int X\phi(Y) \, d\mu = \int Y\phi(Y) \, d\mu, \tag{3.13}$$

if ϕ *is a real valued function on the reals such that*

$$\phi(Y) \in L_2 \quad \textit{and} \quad \phi(0) = 0; \tag{3.14}$$

or if

$$X \in L_1, \quad \phi \textit{ is bounded,} \quad \textit{and} \quad \phi(0) = 0; \tag{3.15}$$

or if

$$\mu(\Omega) < \infty \quad \textit{and} \quad \phi(Y) \in L_2 . \tag{3.16}$$

PROOF. The validity of (3.13) under hypothesis (3.14), (3.15) or (3.16) follows from the usual approximation by simple functions from (3.11) or (3.12) or the last statement of Theorem 3.1.

COROLLARY 3.2. *If* $A \in \mathfrak{M}^c$, $B \in Y^{-1}(\mathfrak{B}') \cap \mathfrak{M}$, $\mu(B) < \infty$, *then* $\int_{AB} X \, d\mu \geqq \int_{AB} Y \, d\mu$.

PROOF.
$$\int_{AB} X \, d\mu = \int_B X \, d\mu - \int_{A^cB} X \, d\mu$$
$$\geqq \int_B Y \, d\mu - \int_{A^cB} Y \, d\mu,$$

since $A^cB \in \mathfrak{M}$; hence the conclusion.

Corollary 3.2 may be used with a method of Johansen and Karush [16] to obtain almost sure convergence of a sequence $\{X_n\}$, where $X_n = E(X \mid \mathfrak{M}_n)$, and where $\{\mathfrak{M}_n\}$ is an expanding sequence of σ-lattices. For reals a and b with $a < b$, let A denote the event that infinitely many X_n are less than a, and B the event that infinitely many are greater than b. One uses Corollary 3.2 to show that $b\mu(AB) \leqq \int_{AB} X \, d\mu \leqq a\mu(AB)$, whence $\mu(AB) = 0$. To obtain the left inequality, define $A_{qr} = \bigcup_{n=q}^{r}[X_n < a]$, $B_{st} = \bigcup_{n=s}^{t}[X_n < b]$, for positive integers $q \leqq r \leqq s \leqq t$. Then $A_{qr}B_{st} = A_{qr}[X_s > b] + A_{qr}[X_s \leqq b] \cap [X_{s+1} > b] + \cdots + A_{qr}\bigcap_{n=s}^{t-1}[X_n \leqq b] \cap [X_t > b]$. But $A_{qr}\bigcap_{n=s}^{\nu-1} [X_n \leqq b] \in \mathfrak{M}_\nu^{\,c}$ for $s < \nu \leqq t$, while $[X_\nu > b] \in \mathfrak{M}_\nu$ since $X_\nu = E(X \mid \mathfrak{M}_\nu)$. Set $E_\nu = A_{qr}\bigcap_{n=s}^{\nu-1}[X_n \leqq b] \cap [X_\nu > b]$. Then $\int_{E_\nu} X \, d\mu \geqq \int_{E_\nu} X_\nu \, d\mu \geqq b\mu(E_\nu)$ by Corollary 3.2. Addition of these inequalities and passage to the limit yields $b\mu(AB) \leqq \int_{AB} X \, d\mu$. Proof of the inequality on the right is similar.

The application here of Corollary 3.2 actually requires $b > 0$ so that $(b, \infty) \in \mathfrak{B}'$ and $\mu[X_\nu > b] < \infty$. This would suffice if, for example, $X > 0$, hence $X_n > 0$ for all n. Also, the conclusion of Corollary 3.2 holds with $\mathfrak{B}'$ replaced by $\mathfrak{B}$ if $\mu(\Omega) < \infty$, so that the method applies for totally finite measure spaces.

COROLLARY 3.3. *If* $0 \leqq Z \in R(\mathfrak{M})$ *and if* $X \leqq Z$ *then* $Y = E(X \mid \mathfrak{M}) \leqq Z$. *If* $0 \geqq Z \in R(\mathfrak{M})$ *and if* $X \geqq Z$ *then* $Y = E(X \mid \mathfrak{M}) \geqq Z$.

PROOF. Suppose there is a positive number c such that $\mu[Z < c < Y] > 0$.

Then $c\mu[Z < c < Y] > \int_{[Z<c]\cap[Y>c]} X\,d\mu \geqq \int_{[Z<c]\cap[Y>c]} Y\,d\mu$ by Corollary 3.2. But $\int_{[Z<c<Y]} Y\,d\mu > c\mu[Z < c < Y]$, a contradiction. It follows that $Y \leqq Z$. Replacing in the first statement X by $-X$, Z by $-Z$, $\mathfrak{M}$ by $\mathfrak{M}^c$ and using Remark 3.2 yields the second statement.

The observation that every constant is in $R(\mathfrak{M})$ yields the following corollary:

Corollary 3.4. *If k is a nonnegative constant and if $X \leqq k$ then $E(X \mid \mathfrak{M}) \leqq k$. If k is a nonnegative constant and if $X \geqq -k$ then $E(X \mid \mathfrak{M}) \geqq -k$.*

(This corollary is incorrectly stated as part of the conclusion of Theorem 1 in [11].)

The applications discussed in Section 4 are applications of an extremizing property of conditional expectation given a σ-lattice (cf. [8] and [10]), which we now develop. Let Φ be a convex function with domain dom Φ, ϕ any determination of Φ', and define

$$(3.17) \qquad \Delta_\Phi(x, z) = \Phi(x) - \Phi(z) - (x - z)\phi(z) \geqq 0.$$

We remark that Δ_Φ has an obvious geometrical interpretation involving the tangent to the graph of Φ at z, and that Δ_Φ is unchanged by the addition of a linear function to Φ. In particular, if 0ε dom Φ, one can thus arrange that $\Phi(0) = \phi(0) = 0$ without changing Δ_Φ. One verifies that for arbitrary x, y, z in dom Φ,

$$(3.18) \quad \Delta_\Phi(x, z) = \Delta_\Phi(x, y) + \Delta_\Phi(y, z) + (x - y)[\phi(y) - \phi(z)].$$

Theorem 3.2. *If $X \varepsilon L_2$, $Y = E(X \mid \mathfrak{M})$, $Z \varepsilon L_2(\mathfrak{M})$, Φ convex, $\Phi(X)$, $\Phi(Y)$, $\Phi(Z) \varepsilon L_1$, $\phi(Y)$, $\phi(Z) \varepsilon L_2$, and if $X(\Omega)$, $Y(\Omega)$, $Z(\Omega)$ are all in* dom Φ, *then*

$$\int \Delta_\Phi(X, Z)\,d\mu \geqq \int \Delta_\Phi(X, Y)\,d\mu + \int \Delta_\Phi(Y, Z)\,d\mu.$$

Proof. If 0ε dom Φ, we may, if necessary, add a linear function to Φ so as to achieve $\Phi(0) = \phi(0) = 0$, and conclude from Corollary 3.1 that

$$\int (X - Y)\phi(Y)\,d\mu = 0.$$

If $0 \not\varepsilon$ dom Φ, then Y is never zero, and $\int (X - Y)\phi(Y)\,d\mu = 0$ follows again from Corollary 3.1 (in applying Corollary 3.1, we extend ϕ to the reals by setting $\phi = 0$ outside dom Φ). Further, since ϕ is nondecreasing on dom Φ, we have $\phi(Z) \varepsilon R(\mathfrak{M})$, hence $\phi(Z) \varepsilon L_2(\mathfrak{M})$, whence $\int (X - Y)\phi(Z)\,d\mu \leqq 0$. The conclusion of the theorem now follows from (3.18).

Since $\Delta_\Phi \geqq 0$, we conclude that

$\int \Delta_\Phi(X, Y)\,d\mu \leqq \int \Delta_\Phi(X, Z)\,d\mu$ for all $Z \varepsilon L_2(\mathfrak{M})$ such that $\phi(Z) \varepsilon L_2$; i.e., $Y = E(X \mid \mathfrak{M})$ *furnishes the minimum value to* $\int [\Phi(X) - \Phi(Z) - (X - Z)\phi(Z)]\,d\mu$. Since, further, $\int (X - Y)\phi(Y)\,d\mu = 0$, $Y = E(X \mid \mathfrak{M})$ *furnishes the maximum value to* $\int \Phi(Z)\,d\mu$ subject to the conditions $Z \varepsilon L_2(\mathfrak{M})$, $\phi(Z) \varepsilon L_2$, $\int (X - Z)\phi(Z)\,d\mu = 0$.

Remark 3.3. From the inequality $\Phi(X) \geqq \Phi(Y) + (X - Y)\phi(Y)$, it follows on integrating both members that $\int \Phi(X)\,d\mu \geqq \int \Phi(Y)\,d\mu$, if any of the conditions (3.14)–(3.16) is satisfied. On setting $\Phi(x) = x^p/p$, we find that $E(\cdot \mid \mathfrak{M})$

reduces not only the L_2 norm, but also the L_p norm (of elements of $L_2 \cap L_p$, $p > 1$).

4. Applications. The applications to be mentioned are to problems of maximum likelihood estimation in which the conditional expectation given a σ-lattice furnishes the solution of the extremum problem involved. Those problems which have come to the author's attention fall broadly into two similar categories. In the first, an unknown function $\theta(\cdot)$ on a given set Ω is to be estimated. While unknown, θ is known to be measurable with respect to a given σ-lattice of subsets of Ω. Observations are made on random variables whose joint distribution depends on θ, and the maximum likelihood estimate of θ is required.

An example of a problem of this kind is discussed by van Eeden [23]. Let Ω be a finite set of reals, $\{\omega_1, \cdots, \omega_k\}$, with $\omega_1 < \omega_2 < \cdots < \omega_k$, representing, for example, doses of a drug or vitamin. For $i = 1, 2, \cdots, k$, let n_i subjects be treated with dose ω_i, and let $\theta(\omega_i)$ denote the probability of a specified response from a subject treated with dose n_i. Let S_i denote the number of responses among the n_i subjects treated; S_i has the binomial distribution with parameters n_i, $\theta(\omega_i)$. The unknown function θ is assumed nondecreasing; i.e., measurable with respect to the σ-lattice consisting of the sets $\{\omega_k\}$, $\{\omega_{k-1}, \omega_k\}, \cdots$, $\{\omega_1, \cdots, \omega_k\}$. It was a similar problem which led to the work reported in [2]. Van Eeden [21], [22] treated the problem independently and more generally; in particular, more general order restrictions were considered.

Ewing, Utz, and the author, and, independently, W. T. Reid, later recognized [8], [7], [9] that the same solution is obtained for distributions other than the binomial. Let again Ω be a given finite set (not necessarily of real numbers) $\{\omega_1, \cdots, \omega_k\}$, and $\mathfrak{M}$ a given σ-lattice of subsets of Ω. Let there be given also a one parameter exponential family of distributions parametrized by the mean θ. Let $\bar{x}_i$ denote the sample mean of a random sample of size n_i from the distribution of the family having mean $\theta(\omega_i)$. Let θ be known to be $\mathfrak{M}$-measurable. Let Φ be the convex function conjugate to the log moment generating function of a member of the family. Define $X(\cdot)$ on Ω by $X(\omega_i) = \bar{x}_i$ and μ by $\mu\{\omega_i\} = n_i$; and set $\theta_i = \theta(\omega_i)$, $i = 1, 2, \cdots, k$. Then the maximum likelihood estimate $\hat{\theta}(\cdot)$, subject to $\mathfrak{M}$-measurability, minimizes, in $R(\mathfrak{M}) = L_2(\mathfrak{M})$, the sum [9]

$$\sum_{i=1}^{k} n_i[\Phi(\bar{x}_i) - \Phi(\theta_i) - (\bar{x}_i - \theta_i)\phi(\theta_i)] = \int [\Phi(X) - \Phi(\theta) - (X - \theta)\phi(\theta)]\, d\mu.$$

As was shown in Section 3, the solution is $\hat{\theta} = E(X \mid \mathfrak{M})$.

If the populations under observation have Poisson distributions with means $\theta_i = \theta(\omega_i)$, $i = 1, 2, \cdots, k$, then $\Phi(x) = x \log x$ (plus a linear function). The $\mathfrak{M}$-measurable maximum likelihood estimate $\hat{\theta}(\cdot)$ is the solution of the extremum problem which for present purposes we shall call the *Poisson extremum problem*:

to maximize, in the class of $\mathfrak{M}$-measurable functions

(*P*) $\theta(\cdot)$ on Ω,

$$\sum_{i=1}^{k} [\bar{x}_i \log \theta_i + (\bar{x}_i - \theta_i)]n_i = \int [X \log \theta + X - \theta]\, d\mu,$$

or equivalently (cf. end of Section 3)

$$(P') \qquad \text{to maximize, in the class of } \mathfrak{M}\text{-measurable functions } \theta(\cdot) \text{ on } \Omega, \quad \sum_{i=1}^k (\bar{x}_i \log \theta_i) n_i \text{ given } \sum (\bar{x}_i - \theta_i) n_i = 0.$$

If the populations being observed have gamma distributions with fixed parameter α and means $\theta_i = \theta(\omega_i)$, $i = 1, 2, \cdots, k$, then $\Phi(x) = -\alpha \log x$ (plus a linear function). The $\mathfrak{M}$-measurable maximum likelihood estimate $\hat{\theta}(\cdot)$ is the solution of the extremum problem which for present purposes we call the *gamma extremum problem*:

$$(G) \qquad \text{to minimize, in the class of } \mathfrak{M}\text{-measurable functions } \theta \text{ on } \Omega, \quad \sum_{i=1}^k [\log \theta_i + (\bar{x}_i - \theta_i)/\theta_i] n_i = \int [\log \theta + (X - \theta)/\theta]\, d\mu.$$

If the populations are normal, variances and sample sizes combine in a more general estimation problem: let the populations be normal with prescribed variances σ_i^2 and unknown means θ_i, $i = 1, 2, \cdots, k$. One may then set $\Phi(x) = x^2/2$. The $\mathfrak{M}$-measurable maximum likelihood estimate $\hat{\theta}(\cdot)$ is the solution of the extremum problem which we shall here refer to as the *normal extremum problem*:

$$(N) \qquad \text{to minimize, in the class of } \mathfrak{M}\text{-measurable functions } \theta \text{ on } \Omega, \quad \sum_{i=1}^k (\bar{x}_i - \theta_i)^2 n_i/\sigma_i^2 = \int (X - \theta)^2\, d\mu,$$

where in this case $\mu\{\omega_i\} = n_i/\sigma_i^2$.

It is interesting that each of the various problems of maximum likelihood estimation discussed below leads to one of these three extremum problems met in the very special context described above.

We emphasize that the solution to each of these extremum problems is $E(X \mid \mathfrak{M})$. Methods of calculating $E(X \mid \mathfrak{M})$ are discussed in [8], [7], [21], and [22] (though in these papers different notation and terminology are used). For the special case in which $\mathfrak{M}$-measurability imposes a linear (simple) order on $\theta_1, \cdots, \theta_k$, another method of calculating $E(X \mid \mathfrak{M})$ is given in [2] (see also [14]). For the case in which $\mathfrak{M}$-measurability implies a partial order among $\theta_1, \cdots, \theta_k$ in which each has at most one predecessor, [20] gives still another method of calculation.

Bartholomew ([3], [4], [5]) deals with the problem mentioned above of sampling from normal populations with given standard deviations and unknown but ordered means. Included in his work are discussions of the likelihood ratio test against trend and a comparison of its power with that of other tests. Chacko

[12] carries farther certain aspects of Bartholomew's work. He also introduces an interesting test against trend, using the same estimate $\hat{\theta} = E(X \mid \mathfrak{M})$, but applied to ranks in sampling from arbitrary distributions.

Lombard [18] uses $\hat{\theta} = E(X \mid \mathfrak{M})$ as a least-squares estimate ("normal" extremum problem) in an application in which Ω is a planar array: $\omega_i = (x_i, y_i)$ are points in the cartesian plane.

Instances of the "gamma" extremum problem in problems of the first type appear in work of Herbach [15] and Thompson [20]. Herbach finds the maximum likelihood estimators of mean and variance for the balanced one-way classification in the analysis of variance, subject to the condition that estimates of variance be non-negative. Dealing also with multiple classification, Thompson determines "restricted" maximum likelihood estimates of variances, determined from joint densities of translation-invariant sufficient statistics, subject again to the requirement that the estimates be non-negative. He develops also an algorithm for use when the order relations among the θ_i can be described by means of a "rooted tree": each θ_i has at most one immediate predecessor.

The "Poisson" extremum problem arises in the following problem of the first type, studied by Boswell [6]. Let X_t be a process of Poisson type (independent increments), with nondecreasing mean rate of occurrence or intensity $\lambda(t)$; i.e., $EX_t = \int_0^t \lambda(\tau)\, d\tau$ is convex. The problem is to determine the maximum likelihood estimate of λ from observation of X_t.

In each of the problems of the second category, a measure space $(T, \mathcal{B}, \nu)$ is prescribed. Observations are made on T according to an unknown density f (with respect to ν) which is to be estimated. Each of the examples which has come to the author's attention leads to a "Poisson" extremum problem.

Grenander [14] considers maximum likelihood estimation of a nonincreasing density on the interval $[a, \infty)$, where a is given. Here $T = [a, \infty)$, $\mathcal{B}$ is the class of Borel subsets of T, and ν is Lebesgue measure.

Grenander discusses also maximum likelihood estimation of a nondecreasing mortality intensity. In making a further study of this problem, Marshall and Proschan [19] describe it as maximum likelihood estimation of a distribution with increasing failure rate: $r(x) = f(x)/[1 - F(x)]$ is nondecreasing, where $F(x) = \int_{-\infty}^{x} f(t)\, dt$. Here T is the set of real numbers, and $\mathcal{B}$ and ν are again Borel sets and Lebesgue measure respectively. The problem of maximum likelihood estimation here leads again to the "Poisson" extremum problem, in which the unknown θ_i are replaced by the unknown values r_i of $r(x)$ at the observed order statistics. Marshall and Proschan obtain strong uniform consistency theorems for maximum likelihood estimators of failure rate, density, and distribution function.

REFERENCES

[1] AKHIEZER, H. I. and GLAZMAN, I. M. (1961). *Theory of Linear Operators in Hilbert Space, 1*. Ungar, New York. (Translated from the Russian.)

[2] 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.

[3] Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives, I and II. *Biometrika* **46** 36–48, 329–335.

[4] Bartholomew, D. J. (1961). A test of homogeneity for means under restricted alternatives. *J. Roy. Statist. Soc. Ser. B* **23** 239–281.

[5] Bartholomew, D. J. (1961). Ordered tests in the analysis of variance. *Biometrika* **48** 325–332.

[6] Boswell, Marllyn (1965). Estimating and testing trend in a stochastic process of Poisson type. Thesis, University of California, Riverside.

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

[8] 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.

[9] Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann Math. Statist.* **29** 437–454.

[10] 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.

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

[12] Chacko, V. J. (1963). Testing homogeneity against ordered alternatives. *Ann. Math. Statist.* **34** 945–956.

[13] Crawford, G. B. and Saunders, S. C. (1964). Nonparametric maximum likelihood estimation. Boeing Document D1-82-0308.

[14] Grenander, Ulf (1956). On the theory of mortality measurement, Part II. *Skand. Aktuarietidskr.* **39** 125–153.

[15] Herbach, L. H. (1959). Properties of model II-type analysis of variance tests, A: optimum nature of the F-test for model II in the balanced case. *Ann. Math. Statist.* **29** 939–959.

[16] Johansen, S. and Karush, J. (1963). On the semimartingale convergence theorem (Abstract). *Ann. Math. Statist.* **34** 1120.

[17] Loève, Michel (1960). *Probability Theory*, (2nd ed.). Van Nostrand, New York.

[18] Lombard, P. B. and Brunk, H. D. (1963). Evaluating the relation of juice composition of mandarin oranges to percent unacceptance of a taste panel. *Food Technology* **17** 113–115.

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

[20] Thompson, W. A., Jr. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33** 273–289.

[21] Van Eeden, Constance (1956). Maximum likelihood estimation of ordered probabilities. *Indag. Math.* **18** 444–455.

[22] Van Eeden, Constance (1957). Maximum likelihood estimation of partially or completely ordered parameters, I and II. *Indag. Math.* **19** 128–136, 201–211.

[23] Van Eeden, Constance (1960). On distribution-free bio-assay. *Proc. Symp. Quantitative Methods in Pharmacology*, Leiden, 206–210.