Pacific Journal of Mathematics

BEST FIT TO A RANDOM VARIABLE BY A RANDOM VARIABLE MEASURABLE WITH RESPECT TO A σ-LATTICE

HUGH D. BRUNK

Vol. 11, No. 3 BadMonth 1961

BEST FIT TO A RANDOM VARIABLE BY A RANDOM VARIABLE MEASURABLE WITH RESPECT TO A σ-LATTICE

H. D. BRUNK

1. Introduction and summary. Let $(\Omega, \mathscr{S}, \mu)$ be a probability space and f a random variable, an $\mathscr{S}$-measurable function from Ω into the space R of real numbers. Let $\mathscr{S}_0$ be a sub-σ-algebra of $\mathscr{S}$. Let f be integrable; that is, let its expectation $E(f)$ exist. Then the Radon-Nikodym Theorem yields an $\mathscr{S}_0$-measurable function g, the conditional expectation of f given $\mathscr{S}_0$: $g = E(f \mid \mathscr{S}_0)$. The conditional expectation g is, in a strong sense to be made precise below, the best fit to f by an $\mathscr{S}_0$-measurable function. The purpose of the present note is to show that there corresponds to f a function with the same minimizing properties when an arbitrary sub-σ-lattice $\mathscr{L}$ takes the place of $\mathscr{S}_0$.

The conditional expectation $g = E(f \mid \mathscr{S}_0)$ has the property that

$$\int (f - g)h d\mu = 0$$

for $\mathscr{S}_0$-measurable h such that the integral exists. It is then immediate that

$$\int (f - h)^2 d\mu = \int (f - g)^2 d\mu + \int (g - h)^2 d\mu .$$

More generally, the squared difference may be replaced by the W. H. Young form $\Delta_\Phi(\circ, \circ)$ determined by an arbitrary convex function Φ (see §2):

$$\int \Delta_\Phi(f, h) d\mu = \int \Delta_\Phi(f, g) d\mu + \int \Delta_\Phi(g, h) d\mu$$

for $\mathscr{S}_0$-measurable h, provided appropriate integrals exist. (The function $\Delta_\Phi(\circ, \circ)$ is nonnegative and vanishes when the arguments are equal.) Thus, for every Φ, $g = E(f \mid \mathscr{S}_0)$ is the solution of the minimizing problem: given f, to minimize $\int \Delta_\Phi(f, h) d\mu$ in the class of $\mathscr{S}_0$-measurable functions. The conditional expectation therefore enjoys a powerful claim to be the "best" fit to f by an $\mathscr{S}_0$-measurable function. (Blackwell [3] has remarked that for square-integrable functions, the conditional expectation may be regarded as a projection in Hilbert space.)

Received June 15, 1960. Part of the research for this paper was done while the author was a recipient of a U. S. Government Award under Public Law 584 of the 79th Congress—The Fulbright Act, and part was supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract AF 18(600)-1108.

Let now $\mathscr{L}$ be a sub-σ-lattice of $\mathscr{S}$: $\mathscr{L}$ is a class of sets in $\mathscr{S}$ containing the void set ϕ and the whole space Ω, and closed under countable intersections and countable unions. Let h be called $\mathscr{L}$-measurable if for every real t $\{\omega \in \Omega: h(\omega) < t\} \in \mathscr{L}$. It will be shown that given an integrable function f, there exists an $\mathscr{L}$-measurable g such that

$$\int (f-h)^2 d\mu \geqq \int (f-g)^2 d\mu + \int (g-h)^2 d\mu \,, \tag{1.1}$$

and, indeed, such that

$$\int \Delta_\Phi(f,h) d\mu \geqq \int \Delta_\Phi(f,g) d\mu + \int \Delta_\Phi(g,h) d\mu \tag{1.2}$$

for every Φ, provided appropriate integrals exist. Thus g is the "best" fit to f in the class of $\mathscr{L}$-measurable functions. (When f is square-integrable, g may be interpreted in L^2 as the point in the cone of $\mathscr{L}$-measurable functions nearest to the given point f.) To determine g requires the specification not only of f but also of the probability measure μ. Thus it seems appropriate to regard f (and g) as random variables. On the other hand, the "best fit" to a sum need not be sum of the "best fits", so a designation of g as a "conditional expectation given $\mathscr{L}$" does not seem completely appropriate.

Methods used in this paper require that μ be totally finite. It would be of interest to relax this restriction.

The problem of maximum likelihood estimation of parameters subject to order restrictions led to a study of the problem of minimizing $\int \Delta_\Phi(f,h) d\mu$ in a special case ([5], § 4). In that special case, Ω is n-dimensional euclidean space, and $\mathscr{L}$ is the class of sets in $\mathscr{S}$ such that $L \in \mathscr{L}$, $(v_1, v_2, \cdots, v_n) \in L$, $u_1 \leqq v_1$, $u_2 \leqq v_2$, $\cdots$, $u_n \leqq v_n \Rightarrow (u_1, u_2, \cdots, u_n) \in L$. Members of $\mathscr{L}$ were called "lower layers". Methods known from the Radon-Nikodym theory were used, but the connection was not clearly understood. It is the purpose of the present paper not only to replace n-dimensional euclidean space by an arbitrary space Ω, and the class of "lower layers" by an arbitrary σ-lattice, but also to formulate the results so as to include conditional expectation given a sub-σ-field as the special instance occurring when $\mathscr{L}$ is a σ-field.

Special cases occurring in maximum likelihood estimation of ordered parameters are treated in [1], [4], [6], [7] and [8]. In the situation treated in [5], inequality (1.1) was found independently by G. M. Ewing[1] and by W. T. Reid[1]; special cases appear in [4] and [9].

Section 2 of the present paper is devoted to definitions. The problem for square-integrable f is treated as a problem in Hilbert space in § 3.

[1] Private communication.

Results on the minimum problem for arbitrary classes of functions are obtained in § 4, and used in § 5 to yield the principal results, Theorem 5.1 and Theorem 5.2, for integrable f and measurable f. It is shown in § 6 that, given a partial ordering on, Ω, a σ-lattice $\mathscr{L}$ can be introduced such that the $\mathscr{L}$-measurable functions are precisely the order-preserving functions. Application to certain problems of maximum likelihood estimation of a multi-dimensional parameter is mentioned in § 7. It is also remarked that (1.2) may be used in a modification of the proof of the Rao-Blackwell Theorem on sufficient statistics[2].

2. **Definitions.** Let Φ be a convex function of a real variable. Set $G_\Phi \equiv_D \{u: \Phi(u) < \infty\}$. (Symbols $\equiv_D$ and $\Longleftrightarrow_D$ will be used in defining the symbol or relation which appears on the right.) Define (cf. [10])

$$\Psi(z) \equiv_D \sup_u [uz - \Phi(u)] . \tag{2.1}$$

Then (W. H. Young's inequality)

$$0 \leqq \Phi(u) + \Psi(z) - uz \leqq \infty , \qquad u, z \text{ real.} \tag{2.2}$$

The function Ψ is convex, and Φ and Ψ are conjugate in the sense of W. H. Young.

For $u \in G_\Phi$, let $\varphi(u)$ denote the left derivative of Φ at u; φ is continuous from the left.

Consider the graph of $\Phi(u)$ in the cartesian (u, w) plane: $w = \Phi(u)$. For fixed z, the form $zu - \Phi(u)$ represents the vertical directed distance from the graph of Φ to the line $w = zu$. If $z = \varphi(u_0)$ for a number $u_0 \in G_\Phi$ then the directed distance $u\varphi(u_0) - \Phi(u)$ is maximized for $u = u_0$, since the line $w = u\varphi(u_0)$ is parallel to a line of support at u_0. Therefore

$$\Phi(u) + \Psi[\varphi(u)] - u\varphi(u) \equiv 0 , \qquad u \in G_\Phi . \tag{2.3}$$

For $u, v \in G_\Phi$, define

$$\begin{cases} \Delta_\Phi(u, v) \equiv_D \Phi(u) + \Psi[\varphi(v)] - u\varphi(v) \\ \qquad\qquad = \Phi(u) - \Phi(v) - (u - v)\varphi(v) . \end{cases} \tag{2.4}$$

(The subscript Φ will often be omitted.) This form has an obvious geometric interpretation relative to the graph of Φ. It follows from (2.2) and (2.3) that

$$\Delta(u, v) \geqq 0 , \qquad \Delta(u, u) = 0 , \qquad u, v \in G_\Phi . \tag{2.5}$$

Also

[2] That there is a connection between (1.2) and the Rao-Blackwell Theorem was suggested to the writer by Cand. Mag. Brøns of the Statistics Institute, University of Copenhagen.

$$(2.6)\qquad \begin{cases} \Delta(u, v) = \int_{\{t:\, v \leqq t < u\}} (u - t)\,d\varphi(t) & \text{if } v \leqq u\,, \\ \Delta(u, v) = \int_{\{t:\, u \leqq t < v\}} (t - u)\,d\varphi(t) & \text{if } v \geqq u\,. \end{cases}$$

For $u, v, w \in G_\Phi$, (2.4) yields

$$(2.7)\qquad \Delta(u, w) = \Delta(u, v) + \Delta(v, w) + (u - v)[\varphi(v) - \varphi(w)]\,.$$

Let $(\Omega, \mathscr{S}, \mu)$ be a probability measure space. Let ϕ denote the void set. For $A \subset \Omega$, let A^c denote its complement $\Omega - A$. For $\mathscr{S}$-measurable, real functions f, h with ranges in G_Φ, and for $A \in \mathscr{S}$, define

$$(2.8)\qquad J_\Phi(f, h; A) \equiv_D \int_A \Delta_\Phi(f, h)\,d\mu\,.$$

(The subscribt Φ will often be omitted.) Define also

$$(2.9)\qquad J(f, h) \equiv_D J(f, h; \Omega)\,.$$

From (2.5),

$$(2.10)\qquad 0 \leqq J(f, h; A) \leqq J(f, h) \leqq \infty\,.$$

3. **Fitting a square-integrable function.** Let $\mathscr{L}$ be a sub-σ-lattice of $\mathscr{S}$; that is, let $\phi \in \mathscr{L}$, $\Omega \in \mathscr{L}$, $\mathscr{L} \subset \mathscr{S}$, and let $\mathscr{L}$ be closed under countable unions and intersections. Let $\mathscr{C}(\mathscr{L})$ denote the class of real-valued functions h on Ω such that $\{\omega\colon h(\omega) < t\} \in \mathscr{L}$ for real t. "Fitting" a given function f refers to the problem of minimizing $J_\Phi(f, h)$ for $h \in \mathscr{C}(\mathscr{L})$. It will be shown that, broadly speaking, given f there is a function $g \in \mathscr{C}(\mathscr{L})$, independent of Φ, which minimizes $J_\Phi(f, \circ)$ in $\mathscr{C}(\mathscr{L})$ for every Φ. For this function g, indeed,

$$J_\Phi(f, h) \geqq J_\Phi(f, g) + J_\Phi(g, h)$$

for $h \in \mathscr{C}(\mathscr{L})$. In the present approach to the problem, the square-integrable function f is regarded as an element of the Hilbert space of square-integrable functions. (In [11] von Neumann approached the Radon-Nikodym Theorem via Hilbert space.)

Let $\mathscr{H}$ be a real Hilbert space, and $\mathscr{C}$ a closed convex cone in $\mathscr{H}$: $\mathscr{C}$ is closed; $x \in \mathscr{C}$, $a \geqq 0 \Rightarrow ax \in \mathscr{C}$; and $x \in \mathscr{C}$, $y \in \mathscr{C} \Rightarrow x + y \in \mathscr{C}$. The following theorem and argument are familiar ([12], p. 120) when $\mathscr{C}$ is a linear subspace, and perhaps in the present more general situation as well.

The inner product in $\mathscr{H}$ will be denoted by $(\circ, \circ)$ and the norm by $\|\circ\|$.

THEOREM 3.1. *If $f \in \mathscr{H}$ then there exists a $g \in \mathscr{C}$ such that*

$(f-g, h) \leqq 0$ *for all* $h \in \mathscr{C}$. *If there exists* $f_0 \neq 0$ *in* $\mathscr{H}$ *such that* $(f, f_0)f_0/\|f_0\|^2 \in \mathscr{C}$, *then* $(f-g, g) = 0$.

If $\mathscr{C}$ is a linear subspace of $\mathscr{H}$ it follows that $(f-g, h) = 0$ for $h \in \mathscr{C}$. It seems of interest to note, as Blackwell has remarked [3], that in this special case Theorem 3.1 yields at once the conditional expectation of a square-integrable random variable. Let $\mathscr{S}_0$ be a sub-σ-algebra of $\mathscr{S}$, $\mathscr{H}$ the class L^2 of square-integrable functions, and $\mathscr{C}$ the subclass of square-integrable, $\mathscr{S}_0$-measurable functions. The function g furnished by the theorem is then $E(f \mid \mathscr{S}_0)$, for $\int fhd\mu = \int ghd\mu$ for $h \in \mathscr{C}$, and in particular when h is the indicator (characteristic) function of a set in $\mathscr{S}_0$.

Proof of Theorem 3.1. Let N denote the set of all elements of $\mathscr{H}$ of the form $f-h$ for $h \in \mathscr{C}$. Since $\mathscr{C}$ is closed, so is N. Since $\mathscr{C}$ is convex, so is N, for $\lambda(f-h_1) + \mu(f-h_2) = f - (\lambda h_1 + \mu h_2) \in N$ if $0 \leqq \lambda \leqq 1$, $\lambda + \mu = 1$, $h_1, h_2 \in \mathscr{C}$. It follows ([12], Theorem 3, p. 120) that N has an element k of smallest norm. Set $g \equiv_D f - k$; then $g \in \mathscr{C}$. Let $h \in \mathscr{C}$; then if $a \geqq 0$, $g + ah = (a+1)[g/(a+1) + ah/(a+1)] \in \mathscr{C}$. Therefore

$$\begin{aligned} \|k\|^2 \leqq \|f-(g+ah)\|^2 &= \|k-ah\|^2 \\ &= \|k\|^2 - 2a(k,h) + a^2\|h\|^2 . \end{aligned}$$

Suppose there exists $h \in \mathscr{C}$ such that $(k, h) > 0$. Set $a = (k,h)/\|h\|^2$, and find $\|k\|^2 \leqq \|k\|^2 - (k,h)^2/\|h\|^2$, a contradiction. Therefore $(k,h) \leqq 0$ for $h \in \mathscr{C}$, the first conclusion of the theorem.

The second conclusion, $(f-g, g) = 0$, is obvious if $g = 0$. In approaching this conclusion for $g \neq 0$, it is first shown that $g \neq 0$ and $(f, g) \geqq 0$ imply $(f-g, g) = 0$. Set $b \equiv_D (f-g, g)/\|g\|^2 = [(f,g) - \|g\|^2]/\|g\|^2 \geqq -1$. Then $g + bg = (1+b)g \in \mathscr{C}$. Hence $\|k\|^2 \leqq \|f-(g+bg)\|^2 = \|k-bg\|^2 = \|k\|^2 - (k,g)^2/\|g\|^2$, so that $(f-g, g) = (k, g) = 0$. It remains to verify that the hypotheses of the theorem imply $(f, g) \geqq 0$. Set $a = (f, f_0)/\|f_0\|^2$. Since by hypothesis $af_0 \in \mathscr{C}$,

$$\|k\|^2 = \|f-g\|^2 \leqq \|f-af_0\|^2 ,$$

or

$$\|f\|^2 - 2(f,g) + \|g\|^2 \leqq \|f\|^2 - 2a(f,f_0) + a^2\|f_0\|^2 ,$$

so that

$$2(f,g) \geqq \|g\|^2 + (f,f_0)^2/\|f_0\|^2 \geqq 0 .$$

This completes the proof of Theorem 3.1

Let L^2 denote the class of square-integrable functions, and set

$\mathscr{C}_1(\mathscr{L}) = L^2 \cap \mathscr{C}(\mathscr{L})$; $\mathscr{C}_1(\mathscr{L})$ is the class of those $\mathscr{L}$-measurable functions which are square-integrable.

LEMMA 3.1. *If $f \in L^2$, there exists $g \in \mathscr{C}_1(\mathscr{L})$ such that*

$$\int (f-h)^2 d\mu \geqq \int (f-g)^2 d\mu + \int (g-h)^2 d\mu \tag{3.1}$$

for all $h \in \mathscr{C}_1(\mathscr{L})$; g is unique a.e. (μ).

Inequality (3.1) is of the form (1.2) for $\Phi(u) \equiv u^2/2$.

Proof of Lemma 3.1. Lemma 3.1 results from the application of Theorem 3.1 to the Hilbert space L^2, in which the inner product is defined by $(f_1, f_2) \equiv_D \int f_1 f_2 d\mu$ for $f_1, f_2 \in L^2$. In this application the closed convex cone $\mathscr{C}$ of Theorem 3.1 is identified with $\mathscr{C}_1(\mathscr{L})$. It is readily verified that $\mathscr{C}_1(\mathscr{L})$ is a convex cone. Also $\mathscr{C}_1(\mathscr{L})$ is closed in L^2, for if $\| h_n - h \|^2 \to 0$ as $n \to \infty$, then $\{h_n\}$ converges to h in measure, and a subsequence converges to h a.e. (μ); but the limit of a sequence of $\mathscr{L}$-measurable functions is also $\mathscr{L}$-measurable. Let g be the element of $\mathscr{C}_1(\mathscr{L})$ guaranteed by Theorem 3.1. Then

$$\int (f-g) h d\mu \leqq 0 \tag{3.2}$$

for $h \in \mathscr{C}_1(\mathscr{L})$. Further, every constant function is in $\mathscr{C}_1(\mathscr{L})$. Therefore the second hypothesis of Theorem 3.1 is satisfied for $f_0 \equiv_D 1$. It follows that

$$\int (f-g) g \, d\mu = 0 , \tag{3.3}$$

so that

$$\int (f-g)(g-h) d\mu \geqq 0 . \tag{3.4}$$

Inequality (3.1) is now immediate. The uniqueness a.e. (μ) of g is evident from (3.1).

For a real-valued function φ of a real variable, and a function h from Ω into the real line R, let φh denote the composite function: for $\omega \in \Omega$, $\varphi h(\omega) \equiv_D \varphi[h(\omega)]$. Inequality (3.4) is the special instance of

$$\int (f-g)(\varphi g - \varphi h) d\mu \geqq 0 , \tag{3.5}$$

in which $\varphi(u) \equiv u$. From (2.7) it follows that (3.5) is equivalent to

(1.2), given the existence of appropriate integrals. Conditions will now be investigated under which, given f, the same function g satisfies (3.5) for functions φ other than the identity well. Lemma 3.2, below, is phrased more generally than is required for the present application.

Let W be a vector lattice ([2], Chapter XV), so that

$$a, b \in W \Rightarrow a \vee b + a \wedge b = a + b \tag{3.6}$$

(here $a \vee b$ and $a \wedge b$ denote respectively the l.u.b. and g.l.b. of the two elements a and b of W). (For (3.6) it is sufficient that W be a commutative lattice-ordered group; ([2], p. 219).) Let $\mathscr{D}$ be a class of order-preserving maps of W into itself, which is a lattice under the induced partial ordering: $\varphi_1 \leqq \varphi_2 \Longleftrightarrow_D \varphi_1(w) \leqq \varphi_2(w)$ for all $w \in W$ ("$\leqq$" denotes the ordering relation on the partially ordered set W). Let $\mathscr{E}$ be a subclass of $\mathscr{D}$. An intersection of lattices is a lattice, and the intersection of all lattices containing $\mathscr{E}$ is the smallest lattice, $\mathscr{E}^*$, containing $\mathscr{E}$. It may be constructed as follows. For an arbitrary subclass $\mathscr{F}$ of $\mathscr{D}$, define $T\mathscr{F}$ as the class of all elements of $\mathscr{D}$ of the form $\varphi_1 \vee \varphi_2$ or $\varphi_1 \wedge \varphi_2$ for $\varphi_1, \varphi_2 \in \mathscr{F}$. Then

$$\mathscr{E}^* = \lim_n T^n \mathscr{E} = \bigcup_n T^n \mathscr{E}.$$

LEMMA 3.2. *Let L be a nonnegative (or non-positive) linear functional on $\mathscr{D}$. Then $L = 0$ on $\mathscr{E}$ implies $L = 0$ on $\mathscr{E}^*$.*

(This may be regarded as a special instance of the proposition that in a normed lattice the elements of zero norm form a lattice.)

Proof. It suffices to show that $\mathscr{F} \subset \mathscr{D}$ and $L = 0$ on $\mathscr{F}$ imply $L = 0$ on $T\mathscr{F}$. But this is immediate from (3.6) and the assumed linearity and constancy of sign of L.

Lemma 3.2 is applied in proving Theorem 3.2.

THEOREM 3.2. *Let $f \in L^2$ and let g be given by Lemma 3.1. Let Φ be convex, let $\varphi g \in L^2$, and let the range of f be in G_Φ. Then the range of g is in G_Φ (i.e., there is a determination of g in the equivalence class determined by Lemma 3.1 whose range is in G_Φ),*

$$\int (f - g)(\varphi g - \varphi h) d\mu \geqq 0 , \tag{3.7}$$

and

$$J_\Phi(f, h) \geqq J_\Phi(f, g) + J_\Phi(g, h) \tag{3.8}$$

for all $h \in \mathscr{C}(\mathscr{L})$ such that the range of h is in G_Φ and such that $\varphi h \in L^2$.

Proof. Setting h in (3.2) first equal to 1 then equal to -1 yields the result that

$$\int (f-g)d\mu = 0 \; . \tag{3.9}$$

From (3.3) and (3.9) it follows that

$$\int (f-g)(ag+b)d\mu = 0 \; .$$

for real a and b. In applying Lemma 3.2, take for W the real line (a vector lattice) R. For fixed f and hence fixed g, take for $\mathscr{D}$ the class of non-decreasing functions ψ defined on R such that $\psi g \in L^2$. One verifies that $\mathscr{D}$ is a lattice. For $\psi \in \mathscr{D}$, set $L(\psi) \equiv_D \int (f-g)\psi g d\mu$. L is clearly a linear functional on $\mathscr{D}$; from (3.2) it follows that L is non-positive. Let $\mathscr{E}$ denote the subclass of $\mathscr{D}$ consisting of functions ψ of the form $\psi(y) \equiv ay + b$, $a \geqq 0$. For arbitrary real c and d with $c < d$, define ψ_1 by $\psi_1(y) = 0$ for $y \leqq c$, $\psi_1(y) = (y-c)/(d-c)$ for $c < y \leqq d$, $\psi_1(y) = 1$ for $y > d$. Then $\psi_1 \in T^2\mathscr{E}$. By Lemma 3.2, $L(\psi_1) = 0$. Let t be an arbitrary real number. For $n=1, 2, \cdots$, set $c_n = t$, $d_n = t + 1/n$, and define ψ_n as ψ_1 was defined above, with c and d replaced by c_n and d_n respectively. Let ψ_0 denote the step-function: $\psi_0(y) = 0$ for $y \leqq t$, $\psi_0(y) = 1$ for $y > t$. Then $L(\psi_0) = \lim_{n\to\infty} L(\psi_{1_n}) = 0$. That is,

$$\int_{\{\omega: g(\omega) > t\}} [f(\omega) - g(\omega)]d\mu(\omega) = 0 \; .$$

It follows that for every Borel set B of real numbers,

$$\int_{\{\omega: g(\omega) \in B\}} [f(\omega) - g(\omega)]d\mu(\omega) = 0 \; . \tag{3.10}$$

(Equation (3.10) may be interpreted thus: $g = E(f \mid g)$.)

It can be seen as follows that the conclusion that the range of g is in G_φ is a consequence of (3.10). Suppose, for example, that $f(\omega) < a$ for $\omega \in \Omega$. Then

$$a\mu\{g \geqq a\} \leqq \int_{\{g \geqq a\}} g d\mu = \int_{\{g \geqq a\}} f d\mu < a\mu\{g \geqq a\} \; ,$$

unless $\mu\{g \geqq a\} = 0$.

It now follows from (3.10) that $\int (f-g)\varphi g d\mu = 0$. Also, if the range of h is in G_φ and if $\varphi(h) \in L^2$, it follows from (3.2) (with h there replaced by φh) that $\int (f-g)\varphi h d\mu \leqq 0$. Equation (3.7) is then immediate. The proof of Theorem 3.2 is completed by the observation that (3.8) is a consequence of (3.7) and (2.7).

4. Minimizing $J(f, \circ)$. Some theorems on minimzing $J(f, \circ)$ in arbitrary classes of $\mathscr{S}$-measurable functions are given in this section. In § 5 the result of Theorem 3.2 is extended to arbitrary integrable f, using the results of the present section.

LEMMA 4.1. *Let Φ be convex. Let f, h_1, h_2 be $\mathscr{L}$-measurable functions with ranges in G_Φ. Set $E \equiv_D \{\omega\colon h_1(\omega) < h_2(\omega)\}$, and for real t set $E(t) \equiv_D \{\omega\colon h_1(\omega) \leqq t < h_2(\omega)\}$. Then*

$$(4.1)\qquad \begin{aligned} -\infty &\leqq J_\Phi(f, h_2; E) - J_\Phi(f, h_1; E) \\ &= \int d\varphi(t) \int_{E(t)} [t - f(\omega)] d\mu(\omega) \leqq \infty \,, \end{aligned}$$

provided either $J_\Phi(f, h_1; E) < \infty$ or $J_\Phi(f, h_2; E) < \infty$.

Proof. From (2.8) and (2.6),

$$\begin{aligned} J(f, h; A) = &\int_{A \cap \{\omega: h(\omega) < f(\omega)\}} d\mu(\omega) \int_{\{t: h(\omega) \leqq t < f(\omega)\}} [f(\omega) - t] d\varphi(t) \\ &+ \int_{A \cap \{\omega: f(\omega) < h(\omega)\}} d\mu(\omega) \int_{\{t: f(\omega) \leqq t < h(\omega)\}} [t - f(\omega)] d\varphi(t) \,. \end{aligned}$$

Since Δ is nonnegative (inequality (2.5)), Fubini's Theorem ([12], Corollary, p. 95) applies, to yield

$$(4.2)\qquad \begin{aligned} J(f, h; A) = &\int d\varphi(t) \int_{A \cap \{\omega: h(\omega) \leqq t < f(\omega)\}} [f(\omega) - t] d\mu(\omega) \\ &+ \int d\varphi(t) \int_{A \cap \{\omega: f(\omega) \leqq t < h(\omega)\}} [t - f(\omega)] d\mu(\omega) \,. \end{aligned}$$

Set $A = E$ and h first equal to h_2, then equal to h_1. Lemma 4.1 then follows, using the observation that

$$E \cap \{h_1 \leqq t < f\} = E \cap \{h_2 \leqq t < f\} \cup E \cap \{f > t\} \cap \{h_1 \leqq t < h_2\}$$

and

$$E \cap \{f \leqq t < h_2\} = E \cap \{f \leqq t < h_1\} \cup E \cap \{f \leqq t\} \cap \{h_1 \leqq t < h_2\} \,.$$

THEOREM 4.1. *Let $\mathscr{C}$ be a class of $\mathscr{S}$-measurable functions, and f a given, fixed $\mathscr{S}$-measurable function. A sufficient condition that g minimize $J_\Phi(f, \circ)$ in $\mathscr{C}$ for all Φ such that the range of f is in G_Φ is that g be bounded by $\inf_\omega f(\omega)$ and $\sup_\omega f(\omega)$, and that*

$$(4.3)\qquad \int_{\{\omega:\, g(\omega) \leqq t < h(\omega)\}} [f(\omega) - t] d\mu(\omega) \leqq 0 \quad \textit{and} \quad \int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [t - f(\omega)] d\mu(\omega) \leqq 0$$

hold for all real t and every $h \in \mathscr{C}$. If $\mathscr{C}$ is a lattice under the partial ordering $h_1 \leqq h_2 \Longleftrightarrow_D h_1(\omega) \leqq h_2(\omega)$ for $\omega \in \Omega$, then (4.3) *is also necessary.*

Proof of sufficiency. For $h \in \mathscr{C}$, set

$$B_1 \equiv_D \{\omega\colon g(\omega) < h(\omega)\}\ ,$$
$$B_2 \equiv_D \{\omega\colon g(\omega) > h(\omega)\}\ ,$$
$$B_3 \equiv_D \{\omega\colon g(\omega) = h(\omega)\}\ .$$

Then

$$J(f, g) = \sum_{i=1}^{3} J(f, g; B_i)$$

and

$$J(f, h) = \sum_{i=1}^{3} J(f, h; B_i)\ .$$

Clearly $J(f, g; B_3) = J(f, h; B_3)$. In Lemma 4.1 set $h_1 = g$, $h_2 = h$, so that E becomes B_1 and $E(t)$ becomes $\{\omega\colon g(\omega) \leqq t < h(\omega)\}$. From (4.1) and (4.3) follows

$$0 \leqq \int d\varphi(t) \int_{\{\omega:\, g(\omega) \leqq t < h(\omega)\}} [t - f(\omega)]d\mu(\omega) = J(f, h; B_1) - J(f, g; B_1) \leqq \infty\ .$$

Interchanging the roles of g and h in the application of Lemma 4.1 yields

$$0 \geqq \int d\varphi(t) \int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [t - f(\omega)]d\mu(\omega) = J(f, g; B_2) - J(f, h; B_2) \geqq -\infty\ .$$

Subtraction gives $0 \leqq J(f, h) - J(f, g) \leqq \infty$, completing the proof of the sufficiency of condition (4.3).

Proof of necessity. Let t_0 be a real number, and define $\Phi_0(t) \equiv_D |t - t_0|/2$, so that $\varphi_0(t)$ has a unit jump at t_0, with $\varphi_0(t_0) = -1/2$. Applying Lemma 4.1 first with $h_2 = h$, $h_1 = g$, $E = \{g < h\}$ and then with $h_2 = g$, $h_1 = h$, $E = \{h < g\}$, one has

$$\text{(4.4)} \quad \begin{aligned} -\infty &\leqq J_{\Phi_0}(f, h) - J_{\Phi_0}(f, g) \\ &= \int_{\{\omega:\, g(\omega) \leqq t_0 < h(\omega)\}} [t_0 - f(\omega)]d\mu(\omega) + \int_{\{\omega:\, h(\omega) \leqq t_0 < g(\omega)\}} [f(\omega) - t_0]d\mu(\omega)\ . \end{aligned}$$

If g minimizes $J_{\Phi_0}(f, \circ)$ in $\mathscr{C}$, then the left member is nonnegative for every $h \in \mathscr{C}$. Given $h \in \mathscr{C}$, define $h_1 \equiv_D g \wedge h$, and replace h in (4.4) by h_1. One finds

$$0 \leqq J_{\Phi_0}(f, h_1) - J_{\Phi_0}(f, g) = \int_{\{\omega:\, h(\omega) \leqq t_0 < g(\omega)\}} [f(\omega) - t_0]d\mu(\omega)\ ,$$

verifying the second of inequalities (4.3). Similarly, setting $h_1 = g \vee h$ yields the first, completing the proof of Theorem 4.1.

Let f be a given $\mathscr{S}$-measurable function, and $\mathscr{C}$ a class of $\mathscr{S}$-

measurable functions. Consider the following two properties of a function $g \in \mathscr{C}$ which is bounded by $\inf_\omega f(\omega)$ and $\sup_\omega f(\omega)$, and for which $\int |f - g| \, d\mu < \infty$.

For real t and $h \in \mathscr{C}$,

$$\int_{\{\omega:\, g(\omega) \leqq t < h(\omega)\}} [g(\omega) - f(\omega)] d\mu(\omega) \geqq 0 \,, \tag{4.5}$$
$$\int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [f(\omega) - g(\omega)] d\mu(\omega) \geqq 0 \,.$$

For all Φ such that the range of f is in G_Φ and all $h \in \mathscr{C}$ with range in G_Φ,

$$J_\Phi(f, h) \geqq J_\Phi(f, g) + J_\Phi(g, h) \,. \tag{4.6}$$

THEOREM 4.2. *Let f be a given $\mathscr{S}$-measurable function. Suppose that $\inf_\omega f(\omega) \leqq g(\omega) \leqq \sup_\omega f(\omega)$ for $\omega \in \Omega$ and that $\int |f - g| \, d\mu < \infty$. Then* (4.5) $\Longleftrightarrow$ (4.6).

Proof that (4.5) $\Rightarrow$ (4.6). Let $h \in \mathscr{C}$, let Φ be convex, and let f, h have ranges in G_Φ. Set $B_1 \equiv_D \{\omega: g(\omega) < h(\omega)\}$, $B_2 \equiv_D \{\omega: h(\omega) < g(\omega)\}$. Set

$$a \equiv_D \int d\varphi(t) \int_{\{\omega:\, g(\omega) \leqq t < h(\omega)\}} [t - g(\omega)] d\mu(\omega) \geqq 0$$

and

$$b \equiv_D \int d\varphi(t) \int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [g(\omega) - t] d\mu(\omega) \geqq 0 \,.$$

In (4.2), replace f by g and A by Ω, to find

$$J(g, h) = a + b \,.$$

Applying (4.5) and Lemma 4.1, one has

$$a \leqq \int d\varphi(t) \int_{\{\omega:\, g(\omega) \leqq t < h(\omega)\}} [t - f(\omega)] d\mu(\omega) = J(f, h; B_1) - J(f, g; B_1)$$

and

$$b \leqq \int d\varphi(t) \int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [f(\omega) - t] d\mu(\omega) = J(f, h; B_2) - J(f, g; B_2) \,,$$

provided either $J(f, h) < \infty$ or $J(f, g) < \infty$. If both are infinite, (4.6) is granted. If at least one is finite, then

$$J(g, h) = a + b \leqq J(f, h) - J(f, g) \,.$$

Since $J(g, h) \geqq 0$, $J(f, g)$ must then be finite, and (4.6) follows.

Proof that (4.6) $\Rightarrow$ (4.5). From (4.6) and (2.7) it follows that

$$\int(f-g)(\varphi g-\varphi h)d\mu \geqq 0$$

when $h \in \mathscr{C}$, and when the ranges of f and h are contained in G_Φ, provided the integral exists. Let t be a real number, and set $\Phi(u) \equiv_D -(u-t)$ for $u \leqq t$, $\Phi(u) \equiv_D 0$ for; $u > t$. Then

$$\int(f-g)(\varphi g-\varphi h)d\mu = -\int_{\{g \leqq t < h\}}(f-g)d\mu + \int_{\{h \leqq t < g\}}(f-g)d\mu\ ,$$

the integrals existing by hypothesis. Given $h \in \mathscr{C}$, set $h_1 \equiv_D g \wedge h$. Then

$$0 \leqq \int(f-g)(\varphi g-\varphi h_1)d\mu = \int_{\{h \leqq t < g\}}(f-g)d\mu\ .$$

The proof of the first member of (4.5) is similar.

5. Fitting an integrable function in $\mathscr{C}(\mathscr{L})$. Let f be integrable. For positive M, N, define

$$f_{M,N} \equiv_D [-M \vee f] \wedge N\ , \tag{5.1}$$

and

$$f_M \equiv_D \lim_{N\to\infty} f_{M,N}\ , \tag{5.2}$$

so that

$$f = \lim_{M\to\infty} f_M\ . \tag{5.3}$$

For fixed M, N, the function $f_{M,N}$ is square-integrable. Lemma 3.1 makes correspond to $f_{M,N}$ a square-integrable, $\mathscr{L}$-measurable function $g_{M,N}$. It will first be shown that

$$g_M \equiv_D \lim_{N\to\infty} g_{M,N} \tag{5.4}$$

and

$$g \equiv_D \lim_{M\to\infty} g_M \tag{5.5}$$

exist. The principal result of the paper will then be proved:

THEOREM 5.1. *If f is integrable and if the range of f is in G_Φ, then*

$$J_\Phi(f, h) \geqq J_\Phi(f, g) + J_\Phi(g, h)$$

for every $h \in \mathscr{C}(\mathscr{L})$ whose range is in G_Φ.

The proof follows several preliminary lemmas.

LEMMA 5.1. *Let $f \in L^2$ and let g be given by Lemma 3.1. Let t be real, and let $h \in \mathscr{C}(\mathscr{L})$. Then*

$$\tag{5.6} \begin{aligned} &\int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [f(\omega) - t] d\mu(\omega) \\ &\qquad > \int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [f(\omega) - g(\omega)] d\mu(\omega) \geqq 0 \;, \end{aligned}$$

$$\tag{5.7} \begin{aligned} &\int_{\{\omega:\, g(\omega) \leqq t < h(\omega)\}} [f(\omega) - t] d\mu(\omega) \\ &\qquad \leqq \int_{\{\omega:\, g(\omega) \leqq t < h(\omega)\}} [f(\omega) - g(\omega)] d\mu(\omega) \leqq 0 \;, \end{aligned}$$

provided, in (5.6), that the indicated set has positive measure.

Proof. Set $\Phi(u) \equiv_D -(u - t)$ for $u \leqq t$, $\Phi(u) \equiv_D 0$ for $u > t$. Set $h_1 \equiv_D g \wedge h$. Then $\varphi h_1 \in \mathscr{C}(\mathscr{L})$. application of (3.2) with h replaced by φh_1 yields

$$\int_{\{\omega:\, g(\omega) \wedge h(\omega) \leqq t\}} [f(\omega) - g(\omega)] d\mu(\omega) \geqq 0 \;.$$

Also, by (3.10),

$$\int_{\{\omega:\, g(\omega) \leqq t\}} [f(\omega) - g(\omega)] d\mu(\omega) = 0 \;.$$

Since $\{g \wedge h \leqq t\} = \{g \leqq t\} \cup \{h \leqq t < g\}$, it follows that

$$\int_{\{\omega:\, h(\omega) \leqq t < g(\omega)\}} [f(\omega) - g(\omega)] d\mu(\omega) \geqq 0 \;.$$

The first of inequalities (5.6) is clear. The proof of (5.7) is similar.

COROLLARY 5.1. *Let $f_i \in L^2$ and let g_i be determined by f_i through Lemma 3.1, $i = 1, 2$. If $f_1(\omega) \leqq f_2(\omega)$ for $\omega \in \Omega$, then there are determinations of g_1, g_2 such that $g_1(\omega) \leqq g_2(\omega)$ for $\omega \in \Omega$.*

Proof. Suppose that for some real t, $\mu\{\omega:\, g_2(\omega) \leqq t < g_1(\omega)\} > 0$. From (5.6) and (5.7) it follows that

$$\begin{aligned} &\int_{\{\omega:\, g_2(\omega) \leqq t < g_1(\omega)\}} [f_2(\omega) - t] d\mu(\omega) \leqq 0 \\ &\qquad < \int_{\{\omega:\, g_2(\omega) \leqq t < g_1(\omega)\}} [f_1(\omega) - t] d\mu(\omega) \\ &\qquad \leqq \int_{\{\omega:\, g_2(\omega) \leqq t < g_1(\omega)\}} [f_2(\omega) - t] d\mu(\omega) \;, \end{aligned}$$

a contradiction. Thus for every real t, $\mu\{g_2 \leqq t < g_1\} = 0$, so that $g_1 \leqq g_2$ a.e. (μ). One may then suppose g_1, g_2 so chosen that the inequality is satisfied everywhere.

Frow Corollary 5.1 it follows that for fixed M the sequence $g_{M,N}$ is monotone, as is also the sequence g_M. The existence of the limits g_M and g is then guaranteed.

THEOREM 5.2. *If g is $\mathscr{S}$-measurable and if the range of f is in G_Φ, then*

$$J_\Phi(f, h) \geqq J_\Phi(f, g) + J_\Phi(g, h)$$

for all bounded $h \in \mathscr{C}(\mathscr{L})$ with range in G_Φ.

Proof. From the geometric interpretation (cf. (2.4)) of Δ and the boundedness of h it is clear that for fixed M there exists N_0 such that $\Delta[f_{M,N}(\omega), h(\omega)]$ is non-decreasing in N for $N > N_0$, $\omega \in \Omega$. Also there exists M_0 such that $\Delta[f_M(\omega), h(\omega)]$ is non-decreasing in M for $M > M_0$, $\omega \in \Omega$. Therefore

$$\begin{cases} J(f_M, h) = \lim_{N\to\infty} J(f_{M,N}, h) , \\ J(f, h) = \lim_{M\to\infty} J(f_M, h) . \end{cases} \tag{5.8}$$

By Theorem 3.2,

$$J(f_{M,N}, h) \geqq J(f_{M,N}, g_{M,N}) + J(g_{M,N}, h) ;$$

hence

$$\liminf_{N\to\infty} J(f_{M,N}, h) \leqq \liminf_{N\to\infty} J(f_{M,N}, g_{M,N}) + \liminf_{N\to\infty} J(g_{M,N}, h) .$$

By Fatou's lemma,

$$\liminf_{N\to\infty} J(f_{M,N}, g_{M,N}) \geqq J(f_M, g_M)$$

and

$$\liminf_{N\to\infty} J(g_{M,N}, h) \geqq J(g_M, h) .$$

Therefore

$$\liminf_{N\to\infty} J(f_{M,N}, h) \geqq J(f_M, g_M) + J(g_M, h) .$$

From (5.8) it now follows that

$$J(f_M, h) \geqq J(f_M, g_M) + J(g_M, h) .$$

A repetition of the argument yields

$$J(f, h) \geqq J(f, g) + J(g, h) ,$$

completing the proof of Theorem 5.2.

LEMMA 5.3. *If f is integrable, so is g.*

Proof. Let $E_{MN} \equiv_D \{\omega: g_{M,N}(\omega) \geqq 0\}$. The application of (3.10) to $f_{M,N}, g_{m,n}$ gives $\int_{E_{MN}} g_{M,N}d\mu = \int_{E_{MN}} f_{M,N}d\mu$. Therefore

$$\int_{E_{MN}} |g_{M,N}|\, d\mu = \int_{E_{MN}} g_{M,N}d\mu$$
$$= \int_{E_{MN}} f_{M,N}d\mu \leqq \int_{E_{MN}} |f_{M,N}|\, d\mu \leqq \int_{E_{MN}} |f|\, d\mu\ .$$

Similarly

$$\int_{E^c_{NM}} |g_{M,N}|\, d\mu = \int_{E^c_{MN}} - g_{M,N}d\mu$$
$$= \int_{E^{-c}_{MN}} - f_{M,N}d\mu \leqq \int_{E^c_{MN}} |f_{M,N}|\, d\mu \leqq \int_{E^c_{MN}} |f|\, d\mu\ .$$

Addition gives

$$\int |g_{M,N}|\, d\mu \leqq \int |f|\, d\mu\ ,$$

and the integrability of $|g| = \lim_M \lim_N |g_{M,N}|$ follows.

Proof of Theorem 5.1. By hypothesis and Lemma 5.3, both f and g are integrable. Passage to the limit yields (4.5). By Theorem 3.2, $g_{M,N}$ is bounded by $\inf_\omega f_{M,N}(\omega)$ and $\sup_\omega f_{M,N}(\omega)$; therefore also $\inf_\omega f(\omega) \leqq g(\omega) \leqq \sup_\omega f(\omega)$, $\omega \in \Omega$. The conclusion of Theorem 5.1 now follows from Theorem 4.2.

6. σ-lattices determined by partial orderings on Ω. The problem of minimizing $J(f, \circ)$ in $\mathscr{D}(\mathscr{L})$ was discussed in § 4 of [5] for the special case in which Ω is a euclidean space E_n, and in which a partial ordering on E_n is given by

$$\omega = (\omega_1, \cdots, \omega_n) \leqq \xi = (\xi_1, \cdots, \xi_n) \Longleftrightarrow_D \omega_1 \leqq \xi_1,\ \omega_2 \leqq \xi_2, \cdots, \omega_n \leqq \xi_n\ .$$

In [5], classes $\mathscr{L}$ and $\mathscr{U}$ of $\mathscr{S}$-measurable sets were introduced as follows: $L \in \mathscr{L} \Longleftrightarrow_D \xi \in L,\ \omega \leqq \xi \Rightarrow \omega \in L$; $U \in \mathscr{U} \Longleftrightarrow_D U^c \in \mathscr{L}$. The approach in [5] to the minimum problem was through an analoue of the Hahn-Jordan decomposition theorem. The present investigation began with the realization that the methods apply equally well when $\mathscr{L}$ is an arbitrary σ-lattice of sets in $\mathscr{S}$. Indeed, such an approach forms an alternative to that developed in the preceding sections. The present section is devoted to the remark that, given a partial ordering on Ω, the class of $\mathscr{S}$-measurable, order-preserving maps from Ω into R coincides with the class $\mathscr{C}(\mathscr{L})$ for a suitably defined σ-lattice $\mathscr{L}$.

Given a σ-lattice $\mathscr{L} \subset \mathscr{S}$, $\mathscr{C}(\mathscr{L})$ denotes the class of functions h such that for every real $t\{\omega: h(\omega) < t\} \in \mathscr{L}$. For a partial ordering $\mathscr{P}(\leqq)$ of Ω, define $\mathscr{P}^*$ as the class of $\mathscr{S}$-measurable, order-preserving maps of Ω into R. Define also $\mathscr{L}(\mathscr{P})$ as the class of $\mathscr{S}$-measurable sets A such that $\xi \in A$, $\omega \leqq \xi \Rightarrow \omega \in A$. The class $\mathscr{L}(\mathscr{P})$ is a σ-lattice.

The following theorem may be proved by straightforward application of the definitions.

THEOREM 6.1. $\mathscr{C}[\mathscr{L}(\mathscr{P})] = \mathscr{P}^*$.

In should perhaps also be remarked that given a class $\mathscr{C}$ of $\mathscr{S}$-measurable functions, one can determine as follows a σ-lattice $\mathscr{L}$ of $\mathscr{S}$-measurable sets such that $\mathscr{C}$ is embedded in the class $\mathscr{C}(\mathscr{L})$ of $\mathscr{L}$-measurable sets. Define a partial ordering $\mathscr{P}(\mathscr{C})$: $\omega \leqq \xi \Longleftrightarrow_D h(\omega) \leqq h(\xi)$ for all $h \in \mathscr{C}$. Then set $\mathscr{L} = \mathscr{L}[\mathscr{P}(\mathscr{C})]$.

7. Concluding remarks. Let X_0 be a random vector, and $\tau = (\tau_1, \cdots, \tau_n)$ a point of euclidean n-space E_n. Define

$$\Psi(\tau) \equiv_D \log E(e^{X_0 \cdot \tau}) .$$

The function Ψ is convex, defined on a convex subset G_Ψ of E_n. For τ in G_Ψ, $\exp\{x\cdot\tau - \Psi(\tau)\}$ $(x \in E_n)$ is the density function with respect to the distribution of X_0 of a member of the exponential family (Darmois-Koopman class, Koopman-Pitman class, or Laplacian family) of distributions generated by X_0.

For $i = 1, 2, \cdots, k$, let $\tau^i \in G_\Psi$. Let independent random samples of sizes $N_1, \cdots, N_k$ be taken from the distributions corresponding to $\tau^1, \cdots, \tau^k$ respectively. Let $\bar{x}^i$ denote the (vector) sample mean of the sample from the ith population. Then the logarithm of the joint density function is

$$\sum_{i=1}^{k} N_i(\bar{x}^i \cdot \tau^i) - \Psi(\tau^i) . \tag{7.1}$$

For $n = 1$, let Φ denote the convex function conjugate to Ψ in the sense of W. H. Young (§ 2); and define θ^i by $\tau^i = \varphi(\theta^i)$, $i = 1, 2, \cdots, k$. A problem of maximum likelihood estimation of the parameters $\theta^1, \cdots, \theta^k$ is a problem of maximizing (7.1), or equivalently of minimizing, for given $\bar{x}^1, \cdots, \bar{x}^k$,

$$\sum_{i=1}^{k} N_i[\Phi(\bar{x}^i) + \Psi(\tau^i) - \bar{x}^i \tau^i] . \tag{7.2}$$

Let Ω be a space of k distinct points $\omega^1, \cdots, \omega^k$, and μ a measure assigning measure N_i/N to ω^i, $i = 1, 2, \cdots, k$, where $N = \sum_{i=1}^{k} N_i$. Define $f(\omega^i) = \bar{x}^i$, $h(\omega^i) = \theta^i$, $i = 1, 2, \cdots, k$. The sum (7.2) can then be written $NJ_\Phi(f, h)$. The problem of minimizing (7.2) subject to a partial ordering on $\theta^1, \theta^2, \cdots, \theta^k$ is thus a special instance of the problem treated in this paper. (This special problem has been treated in [5], [6], [7], and [1], and a special case in [4].)

Certain problems involving n-dimensional parameters with $n > 1$ reduce to the one-dimensional case.

1^0. Suppose the components $X_{10}, \cdots, X_{n0}$ of X_0 are independent. Then $\Psi(\tau)$ is of the form $\sum_{j=1}^n \Psi_j(\tau_j)$. The form to be minimized can be written

$$\sum_{i=1}^k N_i \Big[\sum_{i=1}^n \Phi_j(\bar{x}_j^i) + \Psi_j(\tau_j^i) - \bar{x}_j^i \tau_j^i \Big] ,$$

or $\sum_{j=1}^n J_{\Phi_j}(f_j, h_j)$. In effect, the components of the n-dimensional parameter can be estimated separately.

The methods of the present paper appear to extend naturally to situations involving convex functions of several real variables only for functions Φ of the form $\sum_{j=1}^n \Phi_j$; and for such functions the one-dimensional treatment suffices. Much of the material in § 3 is meaningful also when Φ is an arbitrary convex function of several real variables; but for such functions generalizations of Theorms 5.1 and 5.2 have escaped the author.

2^0. Suppose that order restrictions are applied only to the first components $\tau_1^1, \cdots, \tau_1^k$ of $\tau^1, \cdots, \tau^k$, and that the other components are required to be independent of i:

$$\tau_2^1 = \cdots = \tau_2^k, \tau_3^1 = \cdots = \tau_3^k, \cdots, \tau_n^1 = \cdots = \tau_n^k \; . \tag{7.3}$$

The minimizing values of $\tau_1^1, \cdots, \tau_1^k$ must minimize also the function of them obtained when the parameters τ_j^i $j = 2, 3, \cdots, n, i = 1, 2, \cdots, k$, are replaced by their minimizing values. But this function is of the form (7.2) (one-dimensional problem) for a certain function Φ depending on the minimizing values of the τ_j^i $(j = 2, 3, \cdots, n, i = 1, 2, \cdots, k)$ subject to (7.3). Since the solution is independent of the particular function Φ, the τ_1^i are determined by the $\bar{x}_1^i$ as in the one-dimensional problem $(i = 1, 2, \cdots, k)$.

This remark is appropriate in particular when $n = 2, X_{01}$ is normal with mean 0 and standard deviation 1, and $X_{02} \equiv X_{01}^2$ (the superscript here indicates the square). The distribution of the exponential fumily generated by X_0, corresponding to the parameter point $\tau = (\tau_1, \tau_2)$ is normal with mean $\tau_1/(1 - 2\tau_2)$ and variance $1/(1 - 2\tau_2)$. Thus if the parameters τ_j^i, $i = 1, 2, \cdots, k, j = 1, 2$ are to be estimated by the maximum likelihood method subject to a partial ordering of the means $\mu_i \equiv_D \tau_1^i/(1 - 2\tau_2^i)$ and subject to the condition that τ_2^i is independent of i, then the μ_i are determined by the sample means as in the one-dimensional problem. This result appears in [7] and in [1].

A final remark is that the inequality (1.2) for the conditional expectation of a random variable can be used in a modification of the proof of the Rao-Blackwell theorem on sufficient sub-σ-fields. Let f be a statistic. Let $\mathscr{T}$ be a sufficient sub-σ-field, i.e., $g = E(f \mid \mathscr{T})$ is independent of the measure μ in the the class of measures considered. Let θ_0 denote the expectation of f. By (1.2),

$$J_\Phi(f, \theta_0) \geqq J_\Phi(f, g) + J_\Phi(g, \theta_0) .$$

Hence

$$J_\Phi(g, \theta_0) \leqq J_\Phi(f, \theta_0) , \tag{7.4}$$

For $\Phi(u) \equiv_D u^2/2$, (7.4) states that g has smaller variance than f. Further, let $L(u, v)$ represent the loss which occurs if the estimate of the parameter $E(f)$ is u when the true value is v. Suppose $L(u, v)$ is convex in u for fixed v. Set $\Phi(u) \equiv_D L(u, \theta_0)$ for constant θ_0—the true parameter value. From (7.4) it is then immediate that the risk is smaller for g than for f, whatever the true value θ_0.

References

1. D. J. Bartholomew, *A test of homogeneity for ordered alternatives*, Biometrika **46** (1959), 36-48.
2. Garrett Birkhoff, *Lattice Theory*, Amer. Math. Soc. Colloq. Publ. Vol. 25, Revised Edition, 1948.
3. David Blackwell, *Conditional expectation and unbiased sequential estimation*, Ann. Math. Stat., **18** (1947), 105-10.
4. Miriam Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and Edward Silverman, *An empirical distribution function for sampling with incomplete information*, Ann. Math. Stat., **26** (1955), 641-47.
5. H. D. Brunk, G. M. Ewing and W. R. Utz, *Minimizing integrals in certain classes of monotone functions*, Pacific J. Math., **7** (1957), 833-46.
6. H. D. Brunk, *Maximum likelihood estimates of monotone parameters*, Ann. Math. Stat., **26** (1955), 607-16.
7. Constance van Eeden, *Maximum likelihood estimation of ordered probabilities*, Proc. Kon. Ned. Akad. v. Wet. A **59** (1956), Indag. Math., **18** (1956), 444-55.
8. ———, *Maximum likelihood estimation of partially or completely ordered parameters*, Proc. Kon. Ned. Akad. v. Wet. A **60** (1957), Indag. Math., **19** (1957), 128-36 and 201-11.
9. ———, *A least squares inequality for maximum likelihood estimates of ordered parameters*, Proc. Kon. Ned. Akad. v. Wet. A **60** (1957), Indag. Math., **19** (1957), 513-21.
10. W. Fenchel, *Convex cones, sets, and functions*, mimeographed notes, Department of Methematics, Princeton University, 1953.
11. J. von Neumann, *On rings of operators* III, Annals of Math., **41** (1940), 94-161.
12. A. C. Zaanen, *An introduction to the theory of integration*, North-Holland Publishing Co., Amsterdam, 1958.

UNIVERSITY OF MISSOURI

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. The price per volume (4 numbers) is $12.00; single issues, $3.50. Back numbers are available. Special price to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $4.00 per volume; single issues, $1.25.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 103 Highland Boulevard, Berkeley 8, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION

Pacific Journal of Mathematics

Vol. 11, No. 3 BadMonth, 1961