A GENERALIZATION OF THE NEYMAN-PEARSON FUNDAMENTAL LEMMA[1]

By Herman Chernoff and Henry Scheffé

University of Illinois and Columbia University

1. Summary. Given $m + n$ real integrable functions $f_1, \cdots, f_m, g_1, \cdots, g_n$ of a point x in a Euclidean space X, a real function $\phi(z_1, \cdots, z_n)$ of n real variables, and m constants $c_1, \cdots, c_m$, the problem considered is the existence of a set S^0 in X maximizing $\phi\left(\int_S g_1\, dx, \cdots, \int_S g_n\, dx\right)$ subject to the m side conditions $\int_S f_i\, dx = c_i$, and the derivation of necessary conditions and of sufficient conditions on S^0. In some applications the point with coordinates $\left(\int_S g_1\, dx, \cdots, \int_S g_n\, dx\right)$ may also be required to lie in a given set. The results obtained are illustrated with an example of statistical interest. There is some discussion of the computational problem of finding the maximizing S^0.

2. The problem. The Neyman-Pearson fundamental lemma concerns the problem, given a number of integrable functions, to form their integrals over a variable set S, and to find a set S^0 (if any) for which one of these integrals is maximum subject to the condition that the others have fixed values. The generalization considered here is to maximize a function of several integrals, subject to similar side conditions.

More precisely, we are given $m + n$ integrable[2] functions $f_1(x), \cdots, f_m(x)$, $g_1(x), \cdots, g_n(x)$ of a point x in a Euclidean space S, a real-valued function $\phi(z_1, \cdots, z_n)$ of n real variables defined on the n-dimensional Euclidean space Z, or at least on a suitable subset of Z to be specified later, m constants $c_1, \cdots, c_m$, and a subset A of Z. Let S denote any Borel set in X and form

$$\phi\left(\int_S g_1 dx, \cdots, \int_S g_n dx\right). \tag{2.1}$$

The problem is the existence and characterization of sets S^0 which maximize (2.1) subject to the m conditions

$$\int_S f_i dx = c_i \qquad (i = 1, \cdots, m) \tag{2.2}$$

and the further condition that the point with the coordinates

$$\left(\int_S g_1 dx, \cdots, \int_S g_n dx\right)$$

[1] Work of Scheffé sponsored by the Office of Naval Research.

[2] With respect to Lebesgue measure on the Borel sets.

lie in the given subset A of Z. If the only side conditions are of the form (2.2) then $A = Z$. The Neyman-Pearson lemma refers to the case where

$$n = 1, \qquad \phi \equiv z_1, \qquad A = Z. \tag{2.3}$$

Briefly the history of the problem is the following. It arises in the Neyman-Pearson theory of optimum statistical tests and its generalizations. It was treated in the important special case (2.3) by Neyman and Pearson [5], [6], who obtained the inequalities (4.1) below, with the symbol X_c in (4.1) replaced by X, as sufficient conditions for a maximizing set S^0. The problems of existence and necessity in the case (2.3) were recently solved in all generality by Dantzig and Wald [1]; the necessity problem in this case had been solved under some restrictions (including $m = 1$) in the original paper [5] of Neyman and Pearson. A statistical example which does not come under the special case was recently investigated by Isaacson [4], who obtained sufficient conditions for his problem; this example falls under our treatment and is discussed in Section 8.

In this paper we obtain an existence theorem, and necessary conditions and sufficient conditions for a maximizing S^0. To obtain these the results of Dantzig and Wald are employed, as well as their device of considering certain vector measures to which the Lyapunov theorem [3] may be applied. Construction and computation of a maximizing S^0 are also considered.

3. Further notation and the condition $\mathfrak{C}$. The symbol S (with or without superscripts) will always be understood to denote a Borel set in the Euclidean space X, and the symbols f_j, g_i will always denote integrable[2] functions of a point x in X. In addition to the n-dimensional Euclidean space Z of points $z = (z_1, \cdots, z_n)$, it is convenient to introduce an m-dimensional Euclidean space Y of points $y = (y_1, \cdots, y_m)$. Furthermore, $y(S)$ will denote the point in Y with the coordinates

$$y_j(S) = \int_S f_j\,dx \qquad (j = 1, \cdots, m),$$

and $z(S)$ the point in Z with the coordinates

$$z_i(S) = \int_S g_i\,dx \qquad (i = 1, \cdots, n).$$

Let c be the point in Y with the coordinates $(c_1, \cdots, c_m)$. Then the quantity (2.1) to be maximized may be written $\phi(z(S))$, and the side conditions as $y(S) = c$ and $z(S)\ \varepsilon\ A$.

Call M the range of the $(m + n)$-dimensional vector measure with components

$$y_1(S), \cdots, y_m(S), z_1(S), \cdots, z_n(S), \tag{3.1}$$

that is, M is the set of all points with coordinates (3.1) generated by the totality of Borel sets S in X. Then M is a set in the $(m + n)$-dimensional space $Y \times Z$. By the Lyapunov theorem [3] M is closed, bounded, and convex. We shall call

M_c the set in Z which is the cross section of M by $y = c$, that is, M_c is the set of all points z for which there exists an S with $z(S) = z$, $y(S) = c$. The projection of M on the space Y will be denoted by N, so N is the set of points y for which there exists an S with $y(S) = y$. Thus the side conditions $y(S) = c$ can be satisfied if and only if $c \varepsilon N$.

From the Lyapunov theorem N is a closed, bounded, and convex set in Y. Let π be the smallest-dimensional linear space containing N. In the following it is crucial whether the given point c is an inner point or boundary point of N with respect to the topology not of Y but of π. A point y of N is called an *inner point* of N if there exists an m-dimensional neighborhood U of y such that $U \cap \pi \subset N$, otherwise it is called a *boundary point*.

If c is a boundary point the inequalities of the Neyman-Pearson lemma and its generalizations have to be considered in a certain subset X_c of X determined by the following definitions. Regard the points in Y as vectors and denote the inner product of two vectors $\xi = (\xi_1, \cdots, \xi_m)$ and $y = (y_1, \cdots, y_m)$ by $\xi \cdot y = \xi_1 y_1 + \cdots + \xi_m y_m$. Then $(\xi^1, \cdots, \xi^r)$ is called a *maximal* set of vectors relative to a boundary point c if

$$\xi^i \cdot \xi^i \neq 0 \qquad (i = 1, \cdots, r), \tag{3.2}$$

$$\xi^1 \cdot y \leqq \xi^1 \cdot c \qquad \text{for all } y \varepsilon N, \tag{3.3}$$

$$\xi^p \cdot y \leqq \xi^p \cdot c \qquad \text{for all } y \varepsilon N \text{ for which } \xi^i \cdot y = \xi^i \cdot c, \quad i = 1, \cdots, p-1;\ p = 2, \cdots, r. \tag{3.4}$$

A maximal set $(\xi^1, \cdots, \xi^r)$ is called a *complete* maximal set relative to c if $(\xi^1, \cdots, \xi^r, \xi^{r+1})$ maximal relative to c implies ξ^{r+1} is a linear combination of $\xi^1, \cdots, \xi^r$. The existence of a complete maximal set relative to every boundary point c is shown by Dantzig and Wald ([1], Lemma 3.1). The set X_c is now defined as X if c is an inner point, while if c is a boundary point it is defined as the subset of X in which

$$\sum_{j=1}^{m} \xi_j^i f_j(x) = 0 \qquad (i = 1, \cdots, r), \tag{3.5}$$

where $(\xi^1, \cdots, \xi^r)$ is a complete maximal set relative to c, and $\xi^i = (\xi_1^i, \cdots, \xi_m^i)$.

If D is the domain of definition of $\phi(z)$ and $\phi(z)$ has a differential at $z^0 = (z_1^0, \cdots, z_n^0) \varepsilon D$, then by definition there exist constants[3] $a_1, \cdots, a_n$ such that for $z \varepsilon D$

$$\phi(z) - \phi(z^0) = \sum_{i=1}^{n} a_i(z_i - z_i^0) + o(\| z - z^0 \|), \tag{3.6}$$

[3] If furthermore z^0 is an interior point of D, then $\partial\phi/\partial z_i$ exist at z^0 and equal a_i $(i = 1, \ldots, n)$. In the converse direction, if $\partial\phi/\partial z_i$ exist in a neighborhood of z^0 and are continuous at z^0, then $\phi(z)$ has a differential at z^0. However, in the applications below z^0 may be a boundary point of D.

where

$$\| z - z^0 \| = \left[\sum_{i=1}^{n} (z_i - z_i^0)^2 \right]^{1/2}.$$

If S^0 is a set such that $\phi(z)$ has a differential at $z^0 = z(S^0)$, and if $a_1, \cdots, a_n$ denote the constants in the differential at z^0 as in (3.6), then S^0 will be said to satisfy the *condition* $\mathcal{C}$ if there exist constants $k_1, \cdots, k_m$ such that

$$\text{(3.7)} \qquad \begin{aligned} \sum_{i=1}^{n} a_i g_i(x) &\geqq \sum_{j=1}^{m} k_j f_j(x) \qquad \text{a.e. in } X_c \cap S^0, \\ \sum_{i=1}^{n} a_i g_i(x) &\leqq \sum_{j=1}^{m} k_j f_j(x) \qquad \text{a.e. in } X_c - S^0. \end{aligned}$$

It should be noted that in general the set S^0 must first be known before the constants a_i can be evaluated. This makes the problem of constructing sets S satisfying the condition $\mathcal{C}$ and the side conditions inherently more difficult in the general case than in the special case (2.3), since in the general case the coefficients a_i in the condition $\mathcal{C}$ are functions of the set S to be determined, while in the special case there is only one a_i which is always unity. Further attention is given to the problem of construction and computation of a maximizing S in Section 9.

About the condition $\mathcal{C}$ we remark also that if c is a boundary point, X_c will frequently be a set of measure zero, in which case the condition $\mathcal{C}$ is vacuous. In this case it may be shown ([1], Lemma 3.2) that the set S satisfying the side conditions $y(S) = c$ is unique up to a set of measure zero.

4. Results of Dantzig and Wald. These concern the special case (2.3). They include an existence theorem which we shall not need (it is covered by our Theorem 5.1) and the following theorem which we shall. Here we write $g_1(x) = g(x)$. The set X_c is defined in connection with (3.5).

THEOREM 4.1 (Dantzig and Wald). *If S^0 is a set satisfying $y(S^0) = c$, then a necessary and sufficient condition that S^0 maximize $\int_S g(s)\, dx$ subject to the condition $y(S) = c$ is that there exist constants $k_1, \cdots, k_m$ such that*

$$\text{(4.1)} \qquad \begin{aligned} g(x) &\geqq \sum_{i=1}^{m} k_i f_i(x) \qquad \text{a.e. } \textit{in } X_c \cap S^0, \\ g(x) &\leqq \sum_{i=1}^{m} k_i f_i(x) \qquad \text{a.e. } \textit{in } X_c - S^0. \end{aligned}$$

5. Existence theorem. For our method of proof of the existence theorem to succeed it is essential that the set A be closed. It may nevertheless be possible to use the theorem in situations where the given A is not closed, by applying it to a closed set A_1 containing A and then arguing that the maximum cannot occur in $A_1 - A$; an example is given in Section 8.

THEOREM 5.1. *If there exists a set S satisfying the side conditions $y(S) = c$ and $z(S) \varepsilon A$, if $\phi(z)$ is continuous in $M_c \cap A$, and if A is closed,*[4] *then there exists a set S^0 maximizing $\phi(z(S))$ subject to the conditions $y(S) = c$ and $z(S) \varepsilon A$.*

PROOF. Since M is closed and bounded, so is M_c, and therefore $M_c \cap A$. Also $M_c \cap A$ is nonempty because there exists an S satisfying $y(S) = c$ and $z(S) \varepsilon A$. Since $\phi(z)$ is continuous in the nonempty closed bounded set $M_c \cap A$, there exists a point $z^0 \varepsilon M_c \cap A$ such that

$$\phi(z^0) = \sup \phi(z) \qquad \text{for } z \varepsilon M_c \cap A.$$

Now $z^0 \varepsilon M_c$ implies the existence of an S^0 with $z(S^0) = z^0$ and $y(S^0) = c$. For any other S satisfying $y(S) = c$ and $z(S) \varepsilon A$ we have $z(S) \varepsilon M_c \cap A$, hence $\phi(z(S)) \leqq \phi(z^0) = \phi(z(S^0))$.

6. Necessary conditions. Suppose $\phi(z)$ takes on its maximum value in $M_c \cap A$ at $z^0 = z(S^0)$. The hypotheses of the following theorem imply that z^0 is an interior (in the topology of the n-dimensional space Z) point of A. This will of course be the case if A is open, and in particular if $A = Z$. On the other hand it is easily seen that z^0 must be a boundary (same topology) point of $M_c \cap A$, unless all the constants a_i (see equation (3.6)) in the differential of $\phi(z)$ at z^0 vanish. An S^0 for which all $a_i = 0$ at $z^0 = z(S^0)$ will always satisfy the condition $\mathcal{C}$ (with all $k_j = 0$).

THEOREM 6.1. *If S^0 is a set for which $z(S^0)$ is an interior point*[5] *of A, if $\phi(z)$ is defined in $M_c \cap A$ and has a differential at $z = z(S^0)$, then a necessary condition that S^0 maximize $\phi(z(S))$ subject to the conditions $y(S) = c$ and $z(S) \varepsilon A$ is that S^0 satisfy the condition $\mathcal{C}$.*

PROOF. Assume S^0 satisfies $y(S) = c$ and $z(S) \epsilon A$, and maximizes $\phi(z(S))$ subject to these conditions. Let $z^0 = z(S^0) = (z_1^0, \cdots, z_n^0)$, let a_i be the constants in the differential of $\phi(z)$ at z^0 as in (3.6), and define

$$\hat{\phi}(z) = a_0 + \sum_{i=1}^{n} a_i z_i,$$

where

$$a_0 = \phi(z^0) - \sum_{i=1}^{n} a_i z_i^0.$$

Then

$$\hat{\phi}(z(S)) = a_0 + \sum_{i=1}^{n} a_i z_i(S) = a_0 + \sum_{i=1}^{n} a_i \int_S g_i \, dx,$$

$$\hat{\phi}(z(S)) = a_0 + \int_S \left(\sum_{i=1}^{n} a_i g_i \right) dx. \tag{6.1}$$

[4] A hypothesis of Theorem 5.1 is that there exists a set S^1 satisfying the side conditions. Let $z^1 = z(S^1)$. The hypothesis that A is closed may be replaced by the sometimes useful weaker hypothesis that the set $\{z \mid z \varepsilon (M_c \cap A) \text{ and } \phi(z) \geqq \phi(z^1)\}$ is closed.

[5] The proof shows that this hypothesis may be replaced by the weaker one that $z^0 = z(S^0)$ is a limit point of $L \cap A$ for every line L in Z through z^0.

It will suffice to prove that S^0 maximizes $\hat{\phi}(z(S))$ subject to $y(S) = c$. If this is true we can apply the necessary condition of Theorem 4.1 to $g(x) = \sum_{i=1}^{n} a_i g_i(x)$, and this necessary condition becomes our condition $\mathcal{C}$.

Suppose then that S^0 does not maximize $\hat{\phi}(z(S))$ subject to $y(S) = c$. Then there exists an S^1 with $y(S^1) = c$ and $\hat{\phi}(z(S^1)) > \hat{\phi}(z(S^0))$. We note that $z^1 = z(S^1)$ is in M_c but not necessarily in A, and that $z^1 \neq z^0$ since $\hat{\phi}(z^1) > \hat{\phi}(z^0)$. Let $\rho = \| z^1 - z^0 \|$, and $h\rho = \hat{\phi}(z^1) - \hat{\phi}(z^0)$, so that $h > 0$. Write

$$z^\lambda = (1 - \lambda)z^0 + \lambda z^1 \qquad (0 \leqq \lambda \leqq 1).$$

Then all $z^\lambda \varepsilon M_c$ since M_c is convex. Because $\hat{\phi}(z)$ is a linear function of z with $\hat{\phi}(z^1) = \hat{\phi}(z^0) + h\rho$, it follows that

$$\hat{\phi}(z^\lambda) = \hat{\phi}(z^0) + \lambda h\rho. \tag{6.2}$$

From (3.6) we have for $z \varepsilon M_c \cap A$

$$\phi(z) = \hat{\phi}(z) + o(\| z - z^0 \|),$$

and hence if $z^\lambda \varepsilon M_c \cap A$,

$$\phi(z^\lambda) = \hat{\phi}(z^\lambda) + o(\lambda\rho).$$

Thus there exists a $\delta > 0$ such that $0 < \lambda\rho < \delta$ and $z^\lambda \varepsilon M_c \cap A$ imply

$$|[\phi(z^\lambda) - \hat{\phi}(z^\lambda)]/(\lambda\rho)| < h,$$

and so

$$\phi(z^\lambda) > \hat{\phi}(z^\lambda) - \lambda h\rho.$$

From this, (6.2), and $\hat{\phi}(z^0) = \phi(z^0)$, we get

$$\phi(z^\lambda) > \phi(z^0) \tag{6.3}$$

if $0 < \lambda < \delta/\rho$ and $z^\lambda \varepsilon M_c \cap A$. Recalling that z^0 is an interior point of A, we see there is a λ', $0 < \lambda' < \delta/\rho$, such that $z^{\lambda'} \varepsilon A$. Also $z^{\lambda'} \varepsilon M_c$, so $z^{\lambda'} \varepsilon M_c \cap A$, and (6.3) is true for $\lambda = \lambda'$. But $z^{\lambda'} \varepsilon M_c \cap A$ also implies that there exists an $S^{\lambda'}$ with $y(S^{\lambda'}) = c$, $z(S^{\lambda'}) = z^{\lambda'} \varepsilon A$. For this $S^{\lambda'}$ we have $\phi(z(S^{\lambda'})) > \phi(z(S^0))$, so S^0 does not maximize $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$. This is a contradiction, and hence S^0 maximizes $\hat{\phi}(z(S))$ subject to $y(S) = c$.

7. Sufficient conditions. It is convenient to introduce a weakened form of the property of concavity of a function; which we shall call quasi-concavity; related concepts have been considered by de Finetti [2]. A function $\phi(z)$ defined in a convex set D is said to be *concave* in D if $z^0 \varepsilon D$, $z^1 \varepsilon D$, $z^\lambda = (1 - \lambda)z^0 + \lambda z^1$, and $0 \leqq \lambda \leqq 1$ imply

$$\phi(z^\lambda) \geqq (1 - \lambda)\phi(z^0) + \lambda\phi(z^1).$$

If D is open and convex, and $\phi(z)$ has continuous second partial derivatives in D, then a necessary and sufficient condition for $\phi(z)$ to be concave in D is that the $n \times n$ matrix

$$(\partial^2\phi/\partial z_i \partial z_j) \tag{7.1}$$

be nonpositive in D, that is, all the characteristic roots be nonpositive in D. We shall say $\phi(z)$ is *quasi-concave* in a convex set D if there exists a real differentiable function $\psi(\phi)$ on an interval I containing the range $\phi(D)$ of $\phi(z)$, with $0 < \psi'(\phi) < +\infty$ for $\phi \varepsilon I$, and such that $\psi(\phi(z))$ is concave in D. We note that concavity implies quasi-concavity (take $\psi(\phi) \equiv \phi$), but not conversely (for example, with $n = 2$ consider $\phi(z) = z_1 z_2$ in the set D where $z_1 > 0$, $z_2 > 0$, and take $\psi(\phi) = \log \phi$).

THEOREM 7.1. *If the set S^0 satisfies the side conditions $y(S) = c$ and $z(S) \varepsilon A$, if $\phi(z)$ is defined and quasi-concave in a convex set containing $M_c \cap A$ and has a differential at $z = z(S^0)$, then a sufficient condition that S^0 maximize $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$ is that S^0 satisfy the condition $\mathcal{C}$.*

PROOF. Suppose first that $\phi(z)$ is concave instead of merely quasi-concave in a convex set $D \supset M_c \cap A$, that the other hypotheses of the theorem are satisfied by $\phi(z)$ and S^0, and that S^0 satisfies the condition $\mathcal{C}$. Write $z^0 = z(S^0)$ and define the linear function $\hat{\phi}(z)$ as in the proof of Theorem 6.1. Then S^0 maximizes $\hat{\phi}(z(S))$ subject to the condition $y(S) = c$, since the condition $\mathcal{C}$ now becomes the sufficient condition of Theorem 4.1 applied to $\hat{\phi}(z(S))$ in the form (6.1).

Next we note that $\phi(z) \leqq \hat{\phi}(z)$ in D. Assume the contrary, that there exists a point $z^1 \varepsilon D$ with $b = \phi(z^1) - \hat{\phi}(z^1) > 0$, so $z^1 \neq z^0$ since $\phi(z^0) = \hat{\phi}(z^0)$. If $z^\lambda = (1 - \lambda)z^0 + \lambda z^1 (0 \leqq \lambda \leqq 1)$, then $z^\lambda \varepsilon D$. Define $\tilde{\phi}(z^\lambda) = (1 - \lambda)\phi(z^0) + \lambda\phi(z^1)$. Then $\phi(z^\lambda) \geqq \tilde{\phi}(z^\lambda)$ since $\phi(z)$ is concave. But $\hat{\phi}(z^\lambda) = \tilde{\phi}(z^\lambda) - \lambda b$, and hence, since $\phi(z)$ has a differential at z^0, $\phi(z^\lambda) < \hat{\phi}(z^\lambda) + \lambda b = \tilde{\phi}(z^\lambda)$ for λ sufficiently small but positive. This contradicts $\phi(z^\lambda) \geqq \tilde{\phi}(z^\lambda)$.

If now S is any set satisfying $y(S) = c$ and $z(S) \varepsilon A$, then $z(S) \varepsilon M_c \cap A \subset D$, hence $\phi(z(S)) \leqq \hat{\phi}(z(S)) \leqq \hat{\phi}(z(S^0)) = \phi(z(S^0))$, the second inequality because S^0 maximizes $\hat{\phi}(z(S))$ subject to $y(S) = c$. The theorem is now proved in the case where $\phi(z)$ is concave.

Suppose next $\phi(z)$ is quasi-concave in D. By definition there exists a differentiable function $\psi(\phi)$ on an interval I containing $\phi(D)$, such that $0 < \psi'(\phi) < +\infty$ for $\phi \varepsilon I$, and $\Phi(z) \equiv \psi(\phi(z))$ is concave in D. Since $\psi(\phi)$ is a strictly increasing function on I, a set S^0 maximizes $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$ if and only if it maximizes $\Phi(z(S))$ subject to the same side conditions. Since $\Phi(z)$ is concave in D we may apply the above result to $\Phi(z)$ after we verify that $\Phi(z)$ has a differential at $z^0 = z(S^0)$. But this is the case since $\phi(z)$ has a differential at z^0 and $\psi'(\phi)$ exists at $\phi = \phi(z^0)$. Let $\gamma = \psi'(\phi(z^0))$. Then the constants a_i in the differential of $\Phi(z)$ at z^0 are equal to γ times those for $\phi(z)$ at z^0. The factor γ can be absorbed into the constants $k_1, \cdots, k_m$ of the condition $\mathcal{C}$ since $0 < \gamma < +\infty$.

The following corollary may be useful in applications where it is easier to prove that $\phi(z)$ is suitably dominated by a quasi-concave function than that $\phi(z)$ is quasi-concave.

COROLLARY 7.1. *If the set S^0 satisfies the side conditions $y(S) = c$ and $z(S) \varepsilon A$, if U is a neighborhood in Z of $z^0 = z(S^0)$, if $\phi(z)$ is defined in a set $D \supset U \cup (M_c \cap$*

A) and has a differential at z^0, if $\phi^*(z)$ is defined and quasi-concave in a convex set $D^* \supset D$, and if $\phi(z) \leqq \phi^*(z)$ in D while $\phi(z^0) = \phi^*(z^0)$, then a sufficient condition that S^0 maximize $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$ is that S^0 satisfy the condition $\mathcal{C}$.

PROOF. The corollary will be an immediate consequence of applying Theorem 7.1 to $\phi^*(z)$ instead of $\phi(z)$, providing we can prove that under the hypotheses of the corollary $\phi^*(z)$ has a differential at z^0, and that this is the same as the differential of $\phi(z)$ at z^0 (else a set of constants $a_1^*, \cdots, a_n^*$ different from $a_1, \cdots, a_n$ would appear in the condition $\mathcal{C}$).

Suppose $\Phi^*(z) = \psi(\phi^*(z))$ is concave in D^*, where $0 < \psi(\phi) < +\infty$ for $\phi \varepsilon I \supset \phi^*(D^*)$. Let $\Phi(z) = \psi(\phi(z))$. Since $\psi(\phi)$ has a single-valued differentiable inverse ψ^{-1} on the interval $J = \psi(I)$, and $\phi^*(z) = \psi^{-1}(\Phi^*(z))$, $\phi(z) = \psi^{-1}(\Phi(z))$, it will suffice to prove that $\Phi^*(z)$ has a differential at z^0 and that this is the same as the differential of $\Phi(z)$ at z^0. Since $\Phi^*(z)$ is concave it has a plane of support $w = \tilde{\Phi}(z)$ at z^0, that is, there exists a linear function $\tilde{\Phi}(z)$ such that $\Phi^*(z) \leqq \tilde{\Phi}(z)$ for $z \varepsilon D^*$ and $\Phi^*(z^0) = \tilde{\Phi}(z^0)$. We observe next that this plane is identical with the tangent plane $w = \hat{\Phi}(z)$ to the surface $w = \Phi(z)$ at z^0. For, suppose the contrary. Then because $\tilde{\Phi}(z)$ and $\hat{\Phi}(z)$ are both linear and $\tilde{\Phi}(z^0) = \hat{\Phi}(z^0)$, there must exist a point z^1 where $\tilde{\Phi}(z^1) < \hat{\Phi}(z^1)$. With $z^\lambda = (1 - \lambda)z^0 + \lambda z^1$, $\tilde{\Phi}(z^\lambda) < \hat{\Phi}(z^\lambda)$ for $\lambda > 0$. Therefore, since $\Phi(z)$ has a differential at z^0, $\Phi(z^\lambda) > \tilde{\Phi}(z^\lambda)$ for λ sufficiently small and positive. But this implies the contradiction $\Phi(z^\lambda) > \Phi^*(z^\lambda)$. From the relation $\Phi(z) \leqq \Phi^*(z) \leqq \hat{\Phi}(z)$ in D, the desired conclusion about the differential of $\Phi^*(z)$ at z^0 easily follows.

8. An example. We will illustrate our results by considering their application in the theory of Type D critical regions for testing simple hypotheses concerning several parameters. Type D regions were recently defined and studied by Isaacson [4]; they are locally optimum unbiased critical regions which are a generalization of the Type A regions of the Neyman-Pearson theory for the one-parameter case.

Suppose X is the sample space and there exists a probability density $p(x, \theta)$ for $\theta = (\theta_1, \cdots, \theta_k)$ in the parameter space Ω. The hypothesis to be tested is is $H_0 : \theta = \theta^0$. We assume that for any set S in X the integral $\int_S p\, dx$ has second partial derivatives with respect to θ_i and $\theta_j (i, j = 1, \cdots, k)$ in a neighborhood of θ^0 which are continuous at θ^0, and that it can be differentiated twice under the integral sign with respect to θ_i and θ_j at θ^0. Denote by $G(S)$ the symmetric matrix $\left(\int_S g_{ij} dx\right)$, where $g_{ij} = [\partial^2 p/\partial\theta_i\partial\theta_j]_{\theta^0}$. Also write

$$(8.1) \qquad f_j = [\partial p/\partial\theta_j]_{\theta^0} \qquad (j = 1, \cdots, k), m = k + 1, f_m = p(x, \theta^0).$$

It is convenient to call a critical region S for testing H_0 *locally unbiased of size* α if

$$(8.2)\qquad \int_S f_j\,dx = 0 \qquad (j = 1, \cdots, m-1),$$
$$\int_S f_m\,dx = \alpha, \qquad G(S) \text{ is positive definite.}$$

If S is locally unbiased of size α the (generalized) Gaussian curvature of the power surface at $\theta = \theta^0$ is the determinant $|G(S)|$. A critical region S^0 is said to be *of Type D* if it maximizes $|G(S)|$ subject to the condition that it be locally unbiased of size α. If S^0 is locally unbiased of size α, Isaacson obtained as a sufficient condition for S^0 to be of Type D the existence of constants $k_1, \cdots, k_m$ such that S^0 satisfies

$$(8.3)\qquad \sum_{i,j=1}^{k} b_{ij}\, g_{ij}(x) \geqq \sum_{j=1}^{m} k_i f_i(x) \qquad \text{a.e. in } S^0,$$
$$\sum_{i,j=1}^{k} b_{ij}\, g_{ij}(x) \leqq \sum_{j=1}^{m} k_i f_i(x) \qquad \text{a.e. in } X - S^0,$$

where the matrix (b_{ij}) of constants is the adjoint matrix of $G(S^0)$.

To make the problem conform better to our previous notation we introduce an n-dimensional space Z of points z, with $n = \frac{1}{2}k(k+1)$, and write the coordinates of z as

$$(z_{11}, z_{12}, \cdots, z_{1k}, z_{22}, \cdots, z_{2k}, z_{33}, \cdots, z_{kk}).$$

Define $\phi(z)$ to be the determinant of the symmetric matrix (z_{ij}), $\phi(z) = |(z_{ij})|$, where $z_{ji} = z_{ij}$. With $z_{ij}(S) = \int_S g_{ij}dx$, we see the problem is to maximize $\phi(z(S))$ subject to the side conditions (8.2). These may be written $y(S) = c$, where $c = (0, \cdots, 0, \alpha)$, and $z(S) \in A$, where A is the part of Z where the matrix (z_{ij}) is positive definite. Since $\phi(z)$ is a polynomial in the coordinates of z it has a differential everywhere. If we write $z^0 = z(S^0)$ and $a_{ij} = [\partial\phi/\partial z_{ij}]_{z^0}$ $(i \leqq j)$, we find $a_{ii} = b_{ii}$, $a_{ij} = 2b_{ij} (i < j)$, and so the condition $\mathfrak{C}$ for the present problem is Isaacson's stated in connection with (8.3), except that the first inequality of (8.3) is asserted a.e. in $X_c \cap S^0$ and the second a.e. in $X_c - S^0$. However, it will be shown later that for all $\alpha \neq 0$ or 1, c is an inner point as defined in Section 3, so that the set X_c is the whole space X, and Isaacson's condition is thus precisely the condition $\mathfrak{C}$ in this case.

To apply our results we need to note that the set A is open and is contained in a closed set A_1 such that $\phi(z) = 0$ in $A_1 - A$. Let $h_i(z)$ with $i = 1, \cdots, 2^k - 1$, denote the determinants of principal minors of the matrix (z_{ij}); these polynomials in the coordinates of z are continuous functions of z. Since A is the set where all $h_i(z) > 0$, A is open. Let A_1 be the set in Z where all $h_i(z) \geqq 0$; then A_1 is closed. In $A_1 - A$ all $h_i(z) \geqq 0$, some $h_j(z) = 0$. Thus (z_{ij}) is positive but not positive definite in $A_1 - A$, and hence its determinant $\phi(z) = 0$ there.

We shall prove first by application of our existence theorem 5.1 that if there exists any locally unbiased critical region of size α there exists one of Type D. Suppose then that there exists a critical region S^1 satisfying the side conditions (8.2). Then $\phi(z(S^1)) = |G(S^1)| > 0$. Theorem 5.1 tells us there exists a solution S^0 to the modified problem of maximizing $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A_1$. For the solution S^0 we must have $z(S^0) \varepsilon A$, else $z(S^0) \varepsilon A_1 - A$, $\phi(z(S^0)) = 0 < \phi(z(S^1))$. Thus S^0 maximizes $\phi(z(S))$ subject to $y(S) = c$ and $z(S) \varepsilon A$.

That any critical region of Type D necessarily satisfies the condition $\mathcal{C}$ follows immediately from Theorem 6.1.

That the condition $\mathcal{C}$ is sufficient for a locally unbiased critical region S^0 of size α to be of Type D may be deduced from Theorem 7.1. All the hypotheses of this theorem will be seen to be satisfied if we show that the set A is convex and the function $\log \phi(z)$ is concave in A. Suppose then that ζ^0 and ζ^1 are any two points of A. It will suffice to prove that $\zeta^\lambda = (1 - \lambda)\zeta^0 + \lambda\zeta^1$ is in A and that

$$\log \phi(\zeta^\lambda) \geqq (1 - \lambda) \log \phi(\zeta^0) + \lambda \log \phi(\zeta^1) \tag{8.4}$$

for all $\lambda(0 \leqq \lambda \leqq 1)$. Let ζ^λ_{ij} be the coordinates of ζ^λ. Then the matrices (ζ^r_{ij}) are positive definite for $r = 0, 1$. There thus exists a real nonsingular matrix H such that both matrices $H'(\zeta^0_{ij})H$ and $H'(\zeta^1_{ij})H$ are diagonal, say $H'(\zeta^r_{ij})H = D^r$, where D^r is a diagonal matrix with positive diagonal elements $d^r_1, \cdots, d^r_k (r = 0, 1)$. Now

$$(\zeta^\lambda_{ij}) = (1 - \lambda)(\zeta^0_{ij}) + \lambda(\zeta^1_{ij}),$$

and so $(\zeta^\lambda_{ij}) = K'D^\lambda K$, where $K = H^{-1}$, and D^λ is a diagonal matrix with ith diagonal element equal to $(1 - \lambda)d^0_i + \lambda d^1_i$. Hence D^λ is positive, and so is (ζ^λ_{ij}); thus ζ^λ is in A. Furthermore,

$$\log \phi(\zeta^\lambda) = 2 \log | K | + \log | D^\lambda |,$$

so to prove (8.4) it is enough to verify that

$$\log | D^\lambda | \geqq (1 - \lambda) \log | D^0 | + \lambda \log | D^1 |,$$

or that

$$\sum_{i=1}^k \log [(1 - \lambda) d^0_i + \lambda d^1_i] \geqq (1 - \lambda) \sum_{i=1}^k \log d^0_i + \lambda \sum_{i=1}^k \log d^1_i.$$

But this follows from the concavity of the function $\log x$.

We shall conclude by proving that c is an inner point of N in this and similar statistical problems with side conditions of the form

$$\int_S p(x, \theta^0)\, dx = \alpha \qquad (0 < \alpha < 1),$$

$$\left[\frac{\partial}{\partial\theta_i} \int_S p(x, \theta)\, dx\right]_{\theta^0} = 0 \qquad (i = 1, \cdots, k),$$

if the integral $\int_S p(x, \theta)\, dx$ can be differentiated once under the integral sign for all (Borel) sets S at $\theta = \theta^0$. With the notation (8.1), and $y_i(S) = \int_S f_i dx$, N is the set of all $y(S)$ in the m-dimensional space Y with $m = k + 1$. We observe that the set N is symmetrical with respect to the point $(0, \cdots, 0, \frac{1}{2})$, that is if $y = (y_1, \cdots, y_{m-1}, y_m)$ is in N so is $(-y_1, \cdots, -y_{m-1}, 1 - y_m)$. For any $y \varepsilon N$ there exists an S such that $y(S) = y$. The point $y(X - S)$ is symmetrically placed with respect to $(0, \cdots, 0, \frac{1}{2})$, since for $i = 1, \cdots, m - 1$,

$$y_i(S) + y_i(X - S) = \left[\frac{\partial}{\partial \theta_i}\left\{\int_S p(x, \theta)\, dx + \int_{X-S} p(x, \theta)\, dx\right\}\right]_{\theta^0} = 0,$$

while $y_m(S) + y_m(X - S) = 1$. On taking S to be the empty set we find the point $y^0 = (0, \cdots, 0, 0)$ in N; by symmetry N contains $y^1 = (0, \cdots, 0, 1)$, and by convexity the line segment L joining y^0 and y^1 and containing $c = (0, \cdots, 0, \alpha)$. Since $0 < \alpha < 1$, c is an inner point of the line segment L. From this and the symmetry of N it may be argued geometrically that c is an inner point of N, but we shall give an analytic proof instead.

We shall suppose now that c is a boundary point of the convex body N and from this derive a contradiction. There exists a linear function $h(y) = h_1 y_1 + \cdots + h_m y_m$ not identically zero in N such that $y = c$ maximizes $h(y)$ for y in N. The maximum value of $h(y)$ is thus $h_m \alpha$, and hence $h(y^1) = h_m \leqq h_m \alpha$ and $h(y^0) = 0 \leqq h_m \alpha$; therefore $h_m = 0$. Since zero is the maximum of $h(y)$ for y in N and $h(y)$ does not vanish identically in N, there exists a set S in the space X such that $h(y(S)) < 0$. But

$$h(y(S)) = \sum_{i=1}^{m-1} h_i y_i(S) = -\sum_{i=1}^{m-1} h_i y_i(X - S) = -h(y(X - S)).$$

Thus $h(y(X - S)) > 0$. But $h(y) \leqq 0$ for y in N. This is the desired contradiction.

9. Remarks on computation of a solution. We have mentioned that the problem of construction of a solution is much more difficult here than in the special case covered by the Neyman-Pearson fundamental lemma. We now sketch a general approach which perhaps might be modified and expanded to a method of numerical computation if desired. The basic idea is that the condition $\mathfrak{C}$ reduces the search for a minimizing set among all Borel sets to that for a minimum of a function of $n + m$ real variables or an equivalent problem.

Denote the $(n + m)$-dimensional vector $(\alpha_1, \cdots, \alpha_n, \kappa_1, \cdots, \kappa_m)$ by $v = (v_1, \cdots, v_{n+m})$, and by $S(v)$ the set $\{x \mid \sum_{i=1}^n \alpha_i g_i(x) \geqq \sum_{j=1}^m \kappa_j f_j(x)\}$. With $y(S)$ and $z(S)$ defined as before, let $Y(v) = y(S(v))$, $Z(v) = z(S(v))$. If $\phi(z)$ has a differential at $z = Z(v)$, denote the differential coefficients there by $\Phi_1(v), \cdots, \Phi_n(v)$. Let $\delta(z, A)$ be a continuous function of z which is nonnegative and vanishes if and only if $z \varepsilon A$ (this implies A is closed): an example is the Euclidean distance from z to A. We now define three functions of v:

$$D(v) = \delta(Z(v), A),$$

$$E(v) = \sum_{j=1}^{m} [Y_j(v) - c_j]^2,$$

where $Y_j(v)$ are the components of $Y(v)$, and

$$F(v) = \sum_{i=1}^{n} [\Phi_i(v) - v_i]^2.$$

The function $F(v)$ is defined only for v such that $\phi(z)$ has a differential at $z = Z(v)$.

We next make the following simplifying assumptions (which would be lightened if the sketch of our method were expanded):

(i). The conditions of our existence theorem, Theorem 5.1, are satisfied.

(ii). $X_c = X$, that is, c is an inner point of N (see Section 3).

(iii). The set $\left\{x \,\middle|\, \sum_{i=1}^{n} \alpha_i g_i(x) - \sum_{j=1}^{m} \kappa_j f_j(x) = 0\right\} \cap (X - X^0)$,

where

$$X^0 = \{x \mid g_1(x) = \cdots g_n(x) = f_1(x) = \cdots = f_m(x) = 0\},$$

has measure zero for all vectors v with the components $\alpha_1, \cdots, \alpha_n$ not all zero.

(iv). For any solution S^0, $z^0 = z(S^0)$ is an interior point of A, and $\phi(z)$ has a nonzero differential at z^0.

(v). $\phi(z)$ is defined and quasi-concave in a convex set containing $A \cap M_c$.

Under assumption (i) a solution of course exists, and under this set of assumptions it is easy to see from Theorems 6.1 and 7.1 that a necessary and sufficient condition for S^0 to be a solution is that, up to a set of measure zero and a subset of X^0, $S^0 = S(v^0)$, where v^0 is a vector v with not all components $v_1, \cdots, v_n$ zero, and

$$D(v^0) = E(v^0) = F(v^0) = 0.$$

The problem has now been reduced to finding a vector v with $v_1, \cdots, v_n$ not all zero satisfying $D(v) = E(v) = F(v) = 0$. This problem can be formulated in various equivalent ways; one is to minimize $D + E + F$.

An inelegant aspect of the above approach is that if v^0 is a solution of the computational problem, then for any positive λ, $S(\lambda v^0) = S(v^0)$ but λv^0 does not satisfy $F(\lambda v^0) = 0$ unless $\lambda = 1$, that is, λv^0 is no longer a solution of the computational problem but $S(\lambda v^0)$ is still the same solution of our actual variational problem. This situation arises from our having required the components $v_i (i = 1, \cdots, n)$ to be equal to $\Phi_i(v)$, when it is sufficient that they be proportional with a positive constant of proportionality. Such a proportionality holds if and only if the function

$$\tilde{F}(v) \equiv \left\{\sum_{i=1}^{n} v_i^2 \sum_{j=1}^{n} [\Phi_j(v)]^2\right\}^{\frac{1}{2}} - \sum_{i=1}^{n} v_i \Phi_i(v)$$

vanishes. The inelegancy could thus be removed by replacing $F(v)$ in the above discussion by $\tilde{F}(v)$. The solution of the computational problem could then be normalized by adding one of the conditions

$$\sum_{i=1}^{n} v_i^2 = 1 \quad \text{or} \quad \sum_{i=1}^{n+m} v_i^2 = 1.$$

REFERENCES

[1] G. B. Dantzig and A. Wald, "On the fundamental lemma of Neyman and Pearson," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 87–93.

[2] B. de Finetti, "Sulle stratificazioni convesse," *Ann. Mat. Pura Appl. Ser.* 4, Vol. 30 (1949), pp. 173–183.

[3] P. R. Halmos, "The range of a vector measure," *Bull. Am. Math. Soc.*, Vol. 54 (1948), pp. 416–421.

[4] S. L. Isaacson, "On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 217–234.

[5] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London Ser. A*, Vol. 231 (1933), pp. 289–337.

[6] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1–37.