ON DISTINCT HYPOTHESES

BY AGNES BERGER AND ABRAHAM WALD

Columbia University

1. Introduction. The following problem was suggested to one of the authors by Professor Neyman:

Let $X = (X_1, X_4, \cdots, X_n)$ be a chance vector and let h denote any simple hypothesis specifying its distribution. Let H_i be the composite hypothesis that some element h of a set of simple hypotheses $\{h\}_i$, $(i = 0, 1)$, is true, and assume that H_0 and H_1 are known to be exhaustive. Let h_i denote an element of $\{h\}_i$ $(i = 0, 1)$.

For any region W of the sample space S, let $P(W \mid h)$ be the probability that the sample point falls in W when h is true.

We shall call H_0 and H_1 *distinct*, if a region W exists for which

$$P(W \mid h_0) \neq P(W \mid h_1), \qquad \begin{array}{c}\text{for all } h_0 \in \{h\}_0\\ \text{and all } h_1 \in \{h\}_1 .\end{array}$$

The problem is to establish necessary and sufficient conditions for two composite hypotheses H_0 and H_1 to be distinct.

For any critical region W for testing H_0 against H_1, let $\gamma(W \mid h)$ be the probability of a wrong decision when h is true, i.e.

$$\gamma(W \mid h) = \begin{cases} P(W \mid h) & \text{for} \quad h \in H_0 \\ 1 - P(W \mid h) & \text{for} \quad h \in H_1 . \end{cases}$$

Suppose now that H_0 and H_1 are not distinct. Then to any W a pair h_0', h_1' exist such that

$$P(W \mid h_0') = P(W \mid h_1'),$$

thus

$$\gamma(W \mid h_0') = 1 - \gamma(W \mid h_1'),$$

and therefore

$$\text{l.u.b. } \gamma(W \mid h) \geq \tfrac{1}{2} \text{ for any } W. \tag{1.1}$$

This property of non-distinct hypotheses leads us to investigate the conditions under which 2 hypotheses allow a test where the maximum probability of a wrong decision is $< \frac{1}{2}$.

The result, in turn, will enable us to state, for an important class of hypotheses a necessary and sufficient condition for 2 composite hypotheses to be distinct.

2. A lemma. We shall now prove the following lemma:

LEMMA 2.1. *Assume that X has a density function $p(x)$ and let $H_i = h_i$ be the simple hypothesis that $p(x) = p_i(x)$, $(i = 0, 1)$. Assume that the set R of x's* *satisfying* $p_0(x) \neq p_1(x)$ *has a positive measure. Then there exists a region* W *such that* $\gamma(W \mid p_i) < \frac{1}{2}$, $i = 0, 1$.

PROOF: Let R_0 be defined by $p_0 = p_1$, R_1 by $p_0 < p_1$, R_2 by $p_0 > p_1$. Since $\int_S p_i(x)\, dx = 1$ and $p_i(x) \geq 0$, $(i = 0, 1)$, R_1 and R_2 are of positive measure. Let

$$\phi(x) = \begin{cases} p_1 \text{ in } R_1 \\ p_0 \text{ in } R_2 \\ p_1 = p_0 \text{ in } R_0 . \end{cases}$$

Then $\int_S \phi(x)\, dx > 1$ and either

$$\text{a)} \qquad \int_{R_1+R_0} p_1\, dx > \tfrac{1}{2} \quad \text{or} \quad \text{b)} \quad \int_{R_2} p_0\, dx > \tfrac{1}{2}$$

or both. Assume first a).

Let $R_3 \subset R_1 + R_0$ and such that $\int_{R_3} p_1\, dx = \frac{1}{2}$, but $\int_{R_3} p_0\, dx < \frac{1}{2}$. This can be done by including into R_3 a part of R_1 of non-zero measure. Let $R_4 \subset R_1 + R_0 - R_3$ and such that $0 < \int_{R_4} p_1\, dx < \frac{1}{2} - \int_{R_3} p_0\, dx$. Then

$\int_{R_4} p_0\, dx \leq \int_{R4} p_1 dx < \frac{1}{2} - \int_{R_3} p_0\, dx$, thus $\int_{R_3+R_4} p_0\, dx < \frac{1}{2}$ but $\int_{R_3+R_4} p_1 dx > \frac{1}{2}$.

Assume now b).

Let $R_5 \subset R_2$ and such that $\int_{R_5} p_0\, dx = \frac{1}{2}$. Then $\int_{R_5} p_1\, dx < \frac{1}{2}$. Let $R_6 \subset R_2 - R_5$ and such that $0 < \int_{R_6} p_0\, dx < \frac{1}{2} - \int_{R_5} p_1\, dx$. Then $\int_{R_5+R_6} p_0\, dx > \frac{1}{2}$ and $\int_{R_5+R_6} p_1\, dx < \frac{1}{2}$.

Thus in case a) $W = R_3 + R_4$, and in case b) $W = S - R_5 - R_6$ is a critical region for which $\gamma(W \mid p_i) < \frac{1}{2}$ $(i = 0, 1)$. This proves the lemma.

3. The main theorem. Assume now X to have a density function $p(x, \mid \theta)$ where $\theta = (\theta_1, \theta_4, \cdots, \theta_k)$ is an unknown parameter point. Let ω_0 and ω_1 be two disjoint, bounded and closed subsets of the k-dimensional θ — space. Let $\Omega = \omega_0 + \omega_1$ and suppose that θ is known to belong to Ω, which therefore will be called the parameter space. Let H_i be the hypothesis that the true parameter point is an element of ω_i, $(i = 0, 1)$.

We shall consider the problem of testing H_0 against H_1. Clearly, $P(W \mid h)$ can now be written as $P(W \mid \theta)$ and $\gamma(W \mid h)$ as $\gamma(W \mid \theta)$.

We shall make the following assumptions concerning $p(x \mid \theta)$:

Assumption 1. *$p(x \mid \theta)$ is continuous in θ.* This is of course always fulfilled if Ω consists only of a finite number of points.

Assumption 2. *For any bounded domain M of the sample space we have*

$$\int_M [\operatorname*{Max}_\theta p(x \mid \theta)]\, dx < \infty.$$

It follows from Assumptions 1. and 2. that

$$\lim_{r \to \infty} \int_{S - S_r} p(x \mid \theta)\, dx = 0 \tag{3.1}$$

uniformly in θ where S_r is the sphere in the sample space with center at the origin and radius r.

In what follows, whenever we shall speak of cumulative distribution function $g(\theta)$ in the k-dimensional parameter space, we shall always mean a cumulative distribution function satisfying the condition

$$\int_\Omega dg\,(\theta) = 1.$$

For any c.d.f. $g(\theta)$ let W_g denote a critical region which contains any sample point x satisfying the inequality

$$\int_{\omega_1} p(x \mid \theta)\, dg(\theta) > \int_{\omega_0} p(x \mid \theta)\, dg(\theta),$$

and does not contain a sample point x for which

$$\int_{\omega_1} p(x \mid \theta)\, dg(\theta) < \int_{\omega_0} p(x \mid \theta)\, dg(\theta).$$

It can easily be verified that W_g minimizes the average risk

$$\int_\Omega \gamma(W \mid \theta)\, dg(\theta), \text{ i.e.,} \qquad \int_\Omega W_g \mid \theta)\, dg(\theta) = \operatorname*{Min}_W \int_\Omega \gamma(W \mid \theta)\, dg(\theta). \tag{3.2}$$

Let Ω_i $(i = 0, 1)$ be the class of all density functions $p(x) = \int_\Omega p(x \mid \theta)\, dg_i(\theta)$ where $g_i(\theta)$ is subject to the condition

$$\int_{\omega_i} dh_i(\theta) = 1.$$

Two density functions $p(x)$ and $q(x)$ are said to be equal if $p(x) \neq q(x)$ holds only in a set of measure zero.

It follows from (3.1) and Assumptions 1. and 2. that $\gamma(W \mid \theta)$ is a continuous function of θ. Let $\gamma(W)$ denote the maximum of $\gamma(W \mid \theta)$ with respect to θ. We shall prove the following theorem:

THEOREM 3.1. *A necessary and sufficient condition for the existence of a region W such that $\gamma(W) < \frac{1}{2}$ is that the classes Ω_c and Ω_1 be disjoint.*

PROOF. Suppose that Ω_0 and Ω_1 are not disjoint. Then there exist two distribution functions $g_0(\theta)$ and $g_1(\theta)$ such that

$$\int_{\omega_0} dg_0(\theta) = \int_{\omega_1} dg_1(\theta) = 1$$

and

$$\int_{\omega_0} p(x \mid \theta)\, dg_0(\theta) = \int_{\omega_1} p(x \mid \theta)\, dg_1(\theta)$$

(except perhaps for points x in a set of measure 0).

Let $g(\theta) = \frac{1}{2} g_0(\theta) + \frac{1}{2} g_1(\theta)$. Clearly, $\gamma(W) \geq \int_\Omega \gamma(W \mid \theta)\, dg(\theta) = \frac{1}{2}$ for any W. This proves the necessity of our condition.

We shall now assume that Ω_0 and Ω_1 are disjoint. First we shall show that the results of [1] can be applied. On pages 297–8 of [1] there are seven conditions listed for the sequential case. For the non-sequential case (the one considered here) the conditions 6 and 7 drop out and the first five conditions can be reduced to the following conditions:

Condition 1: The weight function $W(\theta, d)$ *is bounded.*

Condition 2: For any θ, *the chance vector* X *admits a density function* $p(x \mid \theta)$.

Condition 3: For any sequence $\{\theta_i\}$ ($i = 1, 2, \cdots$, ad inf.) *there exists a subsequence* $\{\theta_{i_j}\}$ ($j = 1, 2, \cdots$) *and a parameter point* θ_0 *such that*

$$\lim_{j=\infty} p(x \mid \theta_{i_j}) = p(x \mid \theta_0)$$

Condition 4: If $\{\theta_i\}$ ($i = 1, 2, \cdots$) *is a sequence of points and* θ_0 *a point such that*

$$\lim_{i=\infty} p(x \mid \theta_i) = p(x \mid \theta_0)$$

then,

$$\lim_{i=\infty} W(\theta_i, d) = W(\theta_0, d)$$

uniformly in d.

Condition 5: The same as our Assumption 2.

In our problem d(the decision of the statistician) can take only two values: acceptance or rejection of H_0. Condition 1 is evidently fulfilled, since $W(\theta, d) = 0$ if a correct decision is made, and $= 1$ if a wrong decision is made. Clearly, Conditions 2–5 are also fulfilled in our problem.

A distribution $g(\theta)$ is said to be least favorable, if it maximizes the minimum average risk, i.e., if it maximizes $\int_\Omega \gamma(W \mid \theta)\, dg(\theta)$ with respect to g.

It follows from Theorems 4.1 and 4.4 of [1] that there exists a least favorable distribution.

Let $g^*(\theta)$ be a least favorable distribution. Then, as has been shown in [1] there exists a W_{g^*} such that

(3.3) $$\operatorname*{Max}_{\theta} \gamma(W_{g^*} \mid \theta) = \int_{\Omega} \gamma(W_{g^*} \mid \theta)\, dg^*(\theta).$$

Thus, our theorem is proved if we can show that

(3.4) $$\int_{\Omega} \gamma(W_{g^*} \mid \theta)\, dg^*(\theta) < \tfrac{1}{2}.$$

Let H_0^* be the hypothesis that the true density is given by

$$p_0(x) = \frac{\int_{\omega_0} p(x \mid \theta)\, dg^*(\theta)}{\int_{\omega_0} dg^*(\theta)},$$

and H_1^* the hypothesis that the true density is given by

$$p_1(x) = \frac{\int_{\omega_1} p(x \mid \theta)\, dg^*(\theta)}{\int_{\omega_1} dg^*(\theta)}.$$

Since Ω_0 and Ω_1 are disjoint, $p_0(x)$ and $p_1(x)$ are different density functions. Hence, according to Lemma 2.1, there exists a critical region W^* for testing H_0^* such that $\alpha^* < \frac{1}{2}$ and $\beta^* < \frac{1}{2}$, where α^* is the probability of type I error, and β^* is the probability of type II error. Clearly,

(3.5) $$\tfrac{1}{2} > \alpha^* \int_{\omega_0} dg^*(\theta) + \beta^* \int_{\omega_1} dg^*(\theta) = \int_{\Omega} \gamma(W^* \mid \theta)\, dg^*(\theta).$$

Hence, our theorem is proved.

It follows from (1.1) that if H_0 and H_1 are not distinct, Ω_0 and Ω_1 are not disjoint.

On the other hand, suppose that Ω_0 and Ω_1 are not disjoint and let

$$\int_{\omega_0} p(x \mid \theta)\, dg_0(\theta) = \int_{\omega_1} p(x \mid \theta)\, dg_1(\theta).$$

Then for every W

(3.6) $$\int_{\omega_0} P(W \mid \theta)\, dg_0(\theta) = \int_{\omega_1} P(W \mid \theta)\, dg_1(\theta).$$

Assume now that ω_i is a connected set ($i = 0, 1$). Then, because of the continuity of $P(W \mid \theta)$ there exist 2 functions $\theta_0(W)$, $\theta_1(W)$, $\theta_i(W)$ belonging to $\omega_i (i = 0, 1)$ such that

$$P(W \mid \theta_0(W)) = \int_{\omega_0} P(W \mid \theta)\, dg_0(\theta)$$

and

$$P(W \mid \theta_1(W)) = \int_{\omega_1} P(W \mid \theta)\, dg_1(\theta)$$

for every W. Hence, because of (3.6),

$$P(W \mid \theta_0(W)) = P(W \mid \theta_1(W))$$

for every W. Thus, we arrive at the following theorem:

THEOREM 3.2. *If ω_i is a connected set ($i = 0, 1$), then, under the assumptions of Theorem 3.1, a necessary and sufficient condition for H_0 and H_1 to be distinct is that the sets Ω_0 and Ω_1 be disjoint.*

REFERENCE

[1] A. WALD, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), pp. 279–313.