APPLIED MATHEMATICS AND STATISTICS LABORATORIES
STANFORD UNIVERSITY
CALIFORNIA

ON THE DISTRIBUTION OF A "CLOSENESS" CRITERION

BY
THEOPHILOS CACOULLOS

TECHNICAL REPORT NO. 75
MAY 9, 1962

PREPARED UNDER CONTRACT Nonr-225(52)
(NR-342-022)
FOR
OFFICE OF NAVAL RESEARCH
ON THE DISTRIBUTION OF A "CLOSENESS" CRITERION

by

Theophilos Cacoullos

TECHNICAL REPORT NO. 75

May 9, 1962

PREPARED FOR ARMY, NAVY AND AIR FORCE UNDER

CONTRACT Nonr-225(52) (NR-342-022)

WITH THE OFFICE OF NAVAL RESEARCH

This work was sponsored by the Army, Navy, and Air Force through the Joint Services Advisory Group for Research Groups in Applied Mathematics and Statistics by Contract Nonr-225(52) (NR-342-022)

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

APPLIED MATHEMATICS AND STATISTICS LABORATORIES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
ON THE DISTRIBUTION OF A "CLOSENESS" CRITERION

by

Theophilos Cacoullos

1. Introduction

The statistic whose distribution we derive in this paper arises in problems such as the following. Let us consider three populations \( \Pi_0 \), \( \Pi_1 \), and \( \Pi_2 \); each of them may be thought of as representative of a certain profession, or a race or a language. For each element in a population a set of measurements of several characteristics can be obtained and thus each individual may be considered as a random vector observation \( x \) from one of these populations. Suppose a sample is available from each population. The question of interest is: Which of \( \Pi_1 \) and \( \Pi_2 \) is closer to \( \Pi_0 \) and in what sense? Note the basic difference between this problem and the more well-known corresponding classification problem in which the population \( \Pi_0 \) is identified with either \( \Pi_1 \) or \( \Pi_2 \) and the sample from \( \Pi_0 \) is to be classified into one of the two populations.

The classification approach is realistic in many taxonomic problems, such as the sexing of skeletal remains, where the possibilities of identification are limited to two. However, when the external evidence is slight or in case the populations are clearly distinct but still overlapping, like the example of a modern language and two relatively older ones where the actual question is which older language is nearer to the modern language, then the question of which is the nearer is quite pertinent and realistic.
A problem that immediately arises in the topothetical approach is the choice of an "appropriate" measure of distance between two populations. For the case of multivariate normal populations with the same covariance matrix, the Mahalanobis generalised distance seems a natural measure of the divergence between them. In [1] a more general distance function is considered (cf. (1) below), which gives rise to the statistic \( d \) in (3); this is essentially the difference of two dependent non-central quadratic forms in normal variables.

2. **Summary**

The criterion \( d \) is introduced below in the form of a non-central bilinear form in normal variates; then, it is reduced to a difference of independent non-central quadratic forms in normal variates. The method of mixtures developed by Robbins and Pitman [2] is employed to yield the distribution function of \( d \) as a mixture of distribution functions of differences of independent chi-square variables.

3. **The Criterion**

In [1] we considered the following problem: Let \( \prod_{i} : N(\mu, \Sigma) \), \( i = 0, 1, \ldots, k \) denote \( p \)-variate normal populations with a common known positive definite covariance matrix \( \Sigma \). Let \( A \) be another positive definite matrix. Define the (squared) distance \( d_{0i}^2 \) between \( \prod_0 \) and \( \prod_i \) by

\[
(1) \quad d_{0i}^2 = (\mu_0 - \mu_i)' A^{-1} (\mu_0 - \mu_i), \quad i = 1, \ldots, k.
\]
Assuming \( \mu_1, \ldots, \mu_k \) known, and on the basis of a sample \( x_1, \ldots, x_{n_0} \) from \( \prod_{0} \), we want to select the nearest \( \prod_1 \) to \( \prod_{0} \), i.e., the one with

\[
d^2_{01} = \min_{1 \leq j \leq k} d^2_{01}, \ldots, d^2_{0k}.
\]

If \( k = 2 \) the problem can be stated in the form of testing the hypothesis

\[
(2) \quad H_1 : d^2_{01} \leq d^2_{02} \quad \text{vs} \quad H_2 : d^2_{01} > d^2_{02}.
\]

It was proven (cf., Sections 1.3 and 1.11 of [1]) that of all level \( \alpha \) tests of (2) which are invariant under linear transformations, the test with critical region

\[
W = \{ \bar{x} : \delta(\bar{x}) > c_\alpha \},
\]

where

\[
\bar{x} = \frac{1}{n_0} \sum_{i=1}^{n_0} x_i,
\]

\[
\delta(\bar{x}) = (2\bar{x} - \mu_1 - \mu_2)' A^{-1} (\mu_2 - \mu_1),
\]

\[
F[\delta(\bar{x}) > c_\alpha \mid \delta(\mu_0) = 0] = \alpha,
\]

is a UMP test.

In practice \( \mu_1 \) and \( \mu_2 \) are unknown and have to be estimated from samples from \( \prod_1 \) and \( \prod_2 \). It seems reasonable that, in the case of unknown mean vectors, we substitute in the statistic \( \delta(\bar{x}) \) the means \( \mu_1 \) and \( \mu_2 \) by their corresponding sample estimates. The best estimates of \( \mu_1 \) and \( \mu_2 \) are the samples means \( \bar{x}_1 \) and \( \bar{x}_2 \) from \( \prod_1 \) and \( \prod_2 \) respectively. Thus, we suggest (cf. Chapter III of [1]) that the statistic

\[
(3) \quad d(\bar{x}) = (2\bar{x} - \bar{x}_1 - \bar{x}_2)' A^{-1} (\bar{x}_2 - \bar{x}_1) \equiv d
\]
be used in the same manner as $8(\hat{x})$ for testing hypothesis (2). We call $d$ (cf. Section 1.7 of [1]) a "topotheical"* criterion since it can be used to locate $\Pi_0$ relative to $\Pi_1$ and $\Pi_2$.

In order to evaluate operating characteristics of tests based on $d$ the probabilities of mislocating $\Pi_0$ with respect to $\Pi_1$ and $\Pi_2$ have to be computed for various cut-off points $d_o$. This requires the knowledge of the distribution function of $d(\hat{x})$.

If $A = \Sigma$ (when $d^2_{o1}$ becomes the Mahalanobis distance $\Delta^2_{o1}$) it was shown [1] that, within a constant multiplier, $d$ is distributed as

$$T_o = U'V = \sum_{i=1}^{p} U_i V_i,$$

where the vectors $U$ and $V$ have a joint 2p-variate normal distribution with covariance matrix

$$\begin{pmatrix} I & \rho I \\ \rho I & I \end{pmatrix}, \quad |\rho| < 1,$$

and

$$E(U) = \mu \neq 0, \quad E(V) = v \neq 0.$$ 

When $A \neq \Sigma$ it will be shown that $d$ is distributed as

$$(4) \quad T = \sum_{i=1}^{p} \alpha_i U_i V_i, \quad \alpha_i > 0, \ i = 1, \ldots, p,$$

where the $U_i$ and $V_i$ are distributed as in $T_0$. To show that the random variable

* = locating, from τοποθετέω = to locate, to place.
\[ d = (2 \bar{x} - \bar{x}_1 - \bar{x}_2)\, A^{-1} (\bar{x}_2 - \bar{x}_1) \]

is distributed like \( T \). Define

\[ X = k_1 (2 \bar{x}_0 - \bar{x}_1 - \bar{x}_2), \]
\[ Y = k_2 (\bar{x}_2 - \bar{x}_1), \]

where
\[
k_1 = \sqrt{\frac{(n_0 n_1 n_2)^{1/2}}{[4 n_1 n_2 + n_0 (n_1 + n_2)]^{1/2}}},
\]
\[
k_2 = \sqrt{\frac{(n_1 n_2)^{1/2}}{(n_1 + n_2)^{1/2}}},
\]

and \( n_i \) is the sample size from \( \prod_i, i = 0, 1, 2 \). Then it is easily seen (cf. Lemma 3.2.1 of [1]) that \( X \) and \( Y \) have a joint 2p-variate normal distribution with covariance matrix

\[
\Sigma^* = \begin{pmatrix}
\Sigma & \rho \Sigma \\
\rho \Sigma & \Sigma
\end{pmatrix},
\]

where

\[
\rho = k_1 k_2 \frac{n_2 - n_1}{n_1 n_2}, \quad (\rho^2 < 1);
\]

the mean vectors are

\[
E(X) = k_1 (2\mu_0 - \mu_1 - \mu_2) = \xi, \text{ say},
\]
\[
E(Y) = k_2 (\mu_2 - \mu_1) = \eta, \text{ say}.
\]

Therefore, \( d \) can be written as

\[
d = \frac{1}{k_1 k_2} X' A^{-1} Y
\]

and, except for a constant multiplier, the distribution of \( d \) is that of \( X' A^{-1} Y \).
Now we show that \( X' A^{-1} Y \) is distributed like \( T \). For this set

\[
U = CX, \quad V = CV,
\]

where \( C \) is a nonsingular matrix such that

\[
\begin{align*}
C A C' &= \Lambda, \\
C E C' &= I,
\end{align*}
\]

and \( \Lambda \) is a diagonal matrix with diagonal elements the roots \( \lambda_1, \ldots, \lambda_p \) of the determinantal equation

\[
|A - \lambda E| = 0;
\]

since \( A \) and \( E \) are assumed positive definite, the \( \lambda_i \) are all positive.

Then the bilinear form \( X' A^{-1} Y \) can be written as

\[
X' A^{-1} Y = U' C^{-1} A^{-1} C^{-1} V = U' (CAC')^{-1} V = V' \Lambda^{-1} V.
\]

By (9) (5) and (7), \( U \) is \( N(C_1, C_1 C') \), \( V \) is \( N(C_2, C_2 C') \), and

\[
\text{Cov}(U, V) = \rho CEC' = \rho I. \quad \text{Hence and by (8) we have}
\]

**Lemmas.** The topographical criterion \( d \) in (3) is distributed as \( T \) in (4).

Noting that if \( n_1 = n_2 \), by (6), \( \rho = 0 \), we have

**Corollary 1.** If the sample sizes from \( \mathbb{T}_1 \) and \( \mathbb{T}_2 \) are equal, that is, if \( n_1 = n_2 \), then \( d \) is distributed as \( T \) in (4) where now the variables \( U_1 \) and \( V_1 \) are completely independent.

4. **Reduction of the bilinear form.**

Let \( \chi^2_{n, \tau} \) denote a chi-square variable with \( n > 0 \) degrees of freedom and noncentrality parameter \( \tau \). If \( \tau \) is zero we shall write, as usual, \( \chi^2_n \) and the corresponding distribution function \( F_n(x) \) is
\( F_n(x) = \frac{1}{2^n \Gamma\left(\frac{1}{2}n\right)} \int_0^x e^{-\frac{1}{2}t} t^{\frac{n}{2}-1} \, dt \quad (x > 0). \)

Theorem 1. The bilinear form \( T \) in (4) is distributed as

\( T = \sum_{i=1}^p \alpha_i \left( \alpha \chi^2_{1, \tau_1} - \beta \chi^2_{1, \delta_i} \right) \)

where the chi-square variables are independent, and

\( \alpha = \frac{1}{2} (1+\rho), \quad \beta = \frac{1}{2} (1-\rho) = 1-\alpha, \)

(13) \( \tau^2_i = \frac{1}{4\alpha} (\mu_i + \nu_i)^2, \quad (i = 1, \ldots, p). \)

\( \delta^2_i = \frac{1}{4\beta} (\mu_i - \nu_i)^2, \)

Proof: Define

\( U^*_i = \frac{U_i + V_i}{\sqrt{2(1+\rho)}}, \quad V^*_i = \frac{U_i - V_i}{\sqrt{2(1-\rho)}}, \quad (i = 1, \ldots, p). \)

The \( U^*_i \) and \( V^*_i \) are completely independent and each of them is normally distributed with unit variance; the means are

\( \mathbb{E}(U^*_i) = \frac{1}{\sqrt{2(1+\rho)}} (\mu_i + \nu_i), \quad (i = 1, \ldots, p). \)

(14) \( \mathbb{E}(V^*_i) = \frac{1}{\sqrt{2(1-\rho)}} (\mu_i - \nu_i), \)

We can write \( T \) as

\( T = \sum \alpha_i \left( \frac{1}{4} [(U^*_i + V^*_i)^2 - (U^*_i - V^*_i)^2] \right) = \)

\( = \sum \alpha_i \left( \alpha U^*_i^2 - \beta V^*_i^2 \right). \)

Hence, by (14) and (13), (12) follows.
From Lemma 1 and the above theorem, we deduce

**Corollary 2.** Denote the distinct positive roots of

$$|A - \lambda \Sigma| = 0$$

by $\lambda_{1}, \lambda_{2}, \ldots, \lambda_{r}$ with multiplicities $m_{1}, m_{2}, \ldots, m_{r}$ respectively ($m_{1} + m_{2} + \cdots + m_{r} = p$). Then the hypothetical criterion $d$ in (3) is distributed as

$$d = \sum_{s=1}^{r} \left( \alpha_{s}^{*} \chi_{m_{s}, \theta_{2}}^{2} - \beta_{s}^{*} \chi_{m_{s}, \xi_{s}}^{2} \right)$$

where the $\alpha_{j}^{*}$ and $\beta_{j}^{*}$ are positive constants and

$$\epsilon_{i}^{2} = \sum_{i} \tau_{i}^{2}, \quad \xi_{s}^{2} = \sum_{i} \delta_{i}^{2}, \quad (s = 1, \ldots, r),$$

where for each $s$ the summation is over all $i$ such that $\lambda_{i} = \lambda_{i_{s}}$.

**Proof:** The $\alpha_{i}$ of Lemma 1 which correspond to equal $\lambda_{i}$ are equal. The rest of the corollary follows from (12) and the fact that if $\chi_{m_{1}, \tau_{1}}^{2}, \chi_{m_{2}, \tau_{2}}^{2}, \ldots, \chi_{m_{n}, \tau_{n}}^{2}$ are independent then

$$X = \sum_{j=1}^{r} \chi_{m_{j}, \tau_{j}}^{2}$$

is $\chi_{M, \tau}^{2}$ where $M = \sum_{i=1}^{r} m_{i}$ and $\tau^{2} = \sum_{j=1}^{n} \tau_{j}^{2}$.

We should remark that $\alpha^{*}_{s} \neq \beta^{*}_{s}$, $s = 1, \ldots, r$, unless by (12) $\alpha = \beta$, which is true if and only if, by (13), $p = 0$.

5. **Application of the method of mixtures.**

Robbins and Pitman [2] used the method of mixtures to obtain the distribution function of the quadratic form
\[(16) \quad x = a_1 \chi^2_m + a_2 \chi^2_{m_1} + \ldots + a_r \chi^2_{m_r}\]

where the chi-square variates are independent and \(a, a_1, \ldots, a_r\) are positive constants such that
\[
a_i \geq 1, \quad (i=1, \ldots, r).
\]

They show that for every \(x\)
\[(17) \quad \Pr[X \leq x] = \sum_{j} c_j F_{M+2j}(x/a),\]
where the constants \(c_j\) are defined by the identity
\[(18) \quad \prod_{i=1}^{r} \left(a_i^{-1/2m_i} \left[1 - (1 - \frac{1}{a_i})Z\right]^{-1/2m_i}\right) = \sum_{j} c_j Z^j,
\]
and satisfy
\[
c_j \geq 0 \quad (j = 0, 1, 2, \ldots) \quad \sum_{j} c_j = 1;
\]
\(F_n(x)\) is given by (11) and
\[
M = m + m_1 + \ldots + m_r.
\]

**Note.** Above and from now on each index of summation will range from 0 to \(\infty\) unless otherwise noted. Below we are going to use (17) to obtain the df of \(d\) in (15) as a mixture of d.f.'s of differences of independent \(\chi^2\) variables. Unfortunately, as expected, this mixture will not involve a single summation as in (17) but a multiple summation. This makes the computation quite impractical unless the multiple series converges very fast.

For our purposes we need the following simple generalisation of (10) in [2].
Lemma 3. If \( U_1, \ldots, U_n \) are independent variables with respective distribution functions \( F_1(x), \ldots, F_n(x) \) we shall denote (as in [2]) the d.f. of any Borel measurable function \( H(U_1, \ldots, U_n) \) by

\[
(H(U_1, \ldots, U_n))(F_1(x), \ldots, F_n(x)).
\]

Suppose that \( F_1(x), \ldots, F_n(x) \) are mixtures of the following form

\[
F_\alpha(x) = \sum_{i_\alpha=0}^{\infty} \sum_{j_\alpha=0}^{\infty} C_{i_\alpha} C_{j_\alpha} F_\alpha(i_\alpha, j_\alpha, x), \quad \alpha = 1, \ldots, n,
\]

where for each \( \alpha \) and \( i_\alpha, j_\alpha \) we have

\[
C_{i_\alpha} \geq 0, \quad (i_\alpha = 0, 1, \ldots), \quad \sum_{i_\alpha=0}^{\infty} C_{i_\alpha} = 1, \quad (\alpha = 1, \ldots, n),
\]

\[
C_{j_\alpha} \geq 0, \quad (j_\alpha = 0, 1, \ldots), \quad \sum_{j_\alpha=0}^{\infty} C_{j_\alpha} = 1, \quad (\alpha = 1, \ldots, n).
\]

Then

\[
(H(U_1, \ldots, U_n))(F_1(x), \ldots, F_n(x)) = \sum_{i_1} \ldots \sum_{i_n} \sum_{j_1} \ldots \sum_{j_n} C_{i_1} \ldots C_{i_n} C_{j_1} \ldots C_{j_n} (H(U_1, \ldots, U_n))(F_1^{i_1}, F_1^{j_1}(x), \ldots, F_n^{i_n}, F_n^{j_n}(x))
\]

Proof. We have

\[
P[H(U_1, \ldots, U_n) \leq x] = \int \ldots \int_{\{H(u_1, \ldots, u_n) \leq x\}} dF_1(u_1) \ldots dF_n(u_n) =
\]

\[
= \int \ldots \int_{\{H(u_1, \ldots, u_n) \leq x\}} \sum_{\alpha=1}^{n} \sum_{i_\alpha} \sum_{j_\alpha} C_{i_\alpha} C_{j_\alpha} F_\alpha(i_\alpha, j_\alpha, x)
\]

\[
= \int \ldots \int_{\{H(u_1, \ldots, u_n) \leq x\}} \sum_{i_1} \sum_{j_1} \ldots \sum_{i_n} \sum_{j_n} C_{i_1} \ldots C_{i_n} C_{j_1} \ldots C_{j_n} \int \ldots \int dF_1^{i_1}(u_1) \ldots dF_n^{i_n}(u_n)
\]

\[
= \sum_{i_1} \ldots \sum_{i_n} \sum_{j_1} \ldots \sum_{j_n} C_{i_1} \ldots C_{i_n} C_{j_1} \ldots C_{j_n} \int \ldots \int_{\{H(u_1, \ldots, u_n) \leq x\}} dF_1^{i_1}(u_1) \ldots dF_n^{i_n}(u_n)
\]

and hence (19) follows.
For some convenience in computing the error term when we replace a repeated infinite series by a finite one, we give the following

**Lemma 4.** Suppose that a d.f. $F(x)$ is a mixture of the form

$$F(x) = \sum_{i_1} \cdots \sum_{i_m} C_{i_1} \cdots C_{i_m} F_{i_1, \ldots, i_m}(x),$$

where

$$C_{i_\alpha} \geq 0, \ (i_\alpha = 0, 1, \ldots), \ \sum_{i_\alpha} C_{i_\alpha} = 1, \ (\alpha = 1, \ldots, m),$$

and for every $(i_1, \ldots, i_m)$ $F_{i_1, \ldots, i_m}(x)$ is a d.f. Then for any integers $s_\alpha, t_\alpha$ such that

$$0 \leq s_\alpha \leq t_\alpha, \ \alpha = 1, \ldots, m,$$

we have

$$0 \leq F(x) - \sum_{i_1=s_1}^{t_1} \cdots \sum_{i_m=s_m}^{t_m} C_{i_1} \cdots C_{i_m} F_{i_1, \ldots, i_m}(x) \leq \prod_{\alpha=1}^{m} \left( 1 - \sum_{\alpha=s_\alpha}^{t_\alpha} C_{i_\alpha} \right)$$

uniformly with respect to $x$.

**Proof:** This follows immediately from (9) of [2], that is, if a d.f. $F(x)$ is a mixture

$$F(x) = \sum_{j=0}^{\infty} C_j F_j(x), \ C_j \geq 0, \ \sum C_j = 1$$

of the sequence

$$F_0(x), F_1(x), \ldots$$

of distribution functions, then for any integers $n, N$ such that

$$0 \leq n \leq N$$

we have
\[ 0 \leq F(x) - \sum_{j=n}^{N} c_j f_j(x) \leq 1 - \sum_{j=n}^{N} c_j. \]

Let us now return to the derivation of the distribution function of \( d \) in (15). First let us write \( d \) in the form

\[ d = a \sum_{s=1}^{r} (a_s x^2_{m_s}, \theta_s - b_s x^2_{m_s}, \zeta_s) \]

where \( a > 0 \) is such that

\[ a_s > 1, \ b_s > 1, \ (s = 1, \ldots r). \]

If \( x^2_m \) and \( x^2_n \) are independent and \( \alpha > 0, \beta > 0 \), we shall write

\[ P[\alpha x^2_m - \beta x^2_n \leq x] = H^m_m (x, \alpha, \beta) \]

As a corollary of Theorem 3.2.1 of [1] which gives the d.f. of \( T_0 = \sum_{i=1}^{p} U_i V_i \) we have

**Lemma 5.** Let \( F_s \) denote the d.f. of

\[ x_s = a_s x^2_{m_s}, \theta_s - b_s x^2_{m_s}, \zeta_s, \quad (s = 1, \ldots, r); \]

\( F_s(x) \) can be expressed as a mixture of the form

\[ F_s(x) = \sum_{l=1}^{L} \sum_{j=1}^{J} p_{lj} (\theta_s, \zeta_s) p_{lj} ^{m_s+2js} (x; a_s, b_s) \]

where \( p_n(\tau) \) denotes the probabilities of a Poisson distribution with parameter equal to \( \frac{1}{2} \tau^2 \), i.e.,

\[ p_n(\tau) = e^{-\frac{1}{2} \tau^2} \frac{(\frac{1}{2} \tau^2)^n}{n!}, \quad n = 0, 1, \ldots, \]

and, for any integers \( m, n \) and any constants \( \alpha > 0, \beta > 0 \) \( H_n^m (x; \alpha, \beta) \) is defined by (21).
Applying Lemma 3 to $H(X_1, \ldots, X_r) = \sum_{s=1}^{r} X_s$ with $F_s$ given by (22), we deduce

**Corollary 3.** Let $G(x)$ denote the d.f. of

\begin{equation}
Y = \frac{1}{a} d;
\end{equation}

then

\begin{equation}
G(x) = \sum_{i_1} \cdots \sum_{i_r} \sum_{j_1} \cdots \sum_{j_r} p_{i_1}(\theta_1) \cdots p_{i_r}(\theta_r)p_{j_1}(\xi_1) \cdots p_{j_r}(\xi_r) \cdot
\end{equation}

\[
\left[ \left( \sum_{s=1}^{r} x_s \right)^{m_s+2i_s} \right]^{m_{i_1}+2i_1} (x; a_1, b_1), \ldots, \left( \sum_{s=1}^{r} x_s \right)^{m_s+2i_s} \right]^{m_{i_r}+2i_r} (x; a_r, b_r))].
\]

In order to obtain the d.f. in brackets in this expression we note that this is the d.f. of

\[
\sum_{s=1}^{r} a_s x_{m_s+s}^2 - \sum_{s=1}^{r} b_s x_{m_s+2s}^2 = Z \text{ (say)}.
\]

For each $(i_1, \ldots, i_r)$ and $(j_1, \ldots, j_r)$ define constants $c_i^{(i_1, \ldots, i_r)}$ and $d_j^{(j_1, \ldots, j_r)}$ by the identities in $Z$, $|Z| < 1$,

\[
\prod_{s=1}^{r} \left\{ \frac{1}{2} (m_s + 2i_s) \right\} \left\{ 1 - \left( 1 - \frac{1}{a_s} \right) Z \right\} \left\{ \frac{1}{2} (m_s + 2i_s) \right\} = \sum_{i} c_i^{(i_1, \ldots, i_r)} Z^i,
\]

\[
\prod_{s=1}^{r} \left\{ \frac{1}{2} (m_s + 2j_s) \right\} \left\{ 1 - \left( 1 - \frac{1}{b_s} \right) Z \right\} \left\{ \frac{1}{2} (m_s + 2j_s) \right\} = \sum_{j} d_j^{(j_1, \ldots, j_r)} Z^j;
\]

it follows, as in (18), that

\[
c_i^{(i_1, \ldots, i_r)} \geq 0, \quad i=0,1, \ldots, \quad \sum_{i} c_i^{(i_1, \ldots, i_r)} = 1,
\]

\[
d_j^{(j_1, \ldots, j_r)} \geq 0, \quad j=0,1, \ldots, \quad \sum_{j} d_j^{(j_1, \ldots, j_r)} = 1.
\]
By (17) the d.f. of \( \sum_{s=1}^{r} a_s \chi^2_{m_s + 2i_s} \) is

\[
\sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_r=1}^{M_{i_r}} \prod_{i=1}^{r} F_{i_i}(x)_{i_i + 2i_i},
\]

where

\[
M_{i_1, \ldots, i_r} = \sum_{s=1}^{r} (m_s + 2i_s).
\]

Similarly, the d.f. of \( \sum_{s=1}^{r} b_s \chi^2_{m_s + 2j_s} \) is

\[
\sum_{j_1=1}^{N_{j_1}} \cdots \sum_{j_r=1}^{N_{j_r}} \prod_{j=1}^{r} F_{j_j}(x)_{j_j + 2j_j},
\]

where

\[
N_{j_1, \ldots, j_r} = \sum_{s=1}^{r} (m_s + 2j_s).
\]

Another application of Lemma 3 gives the d.f. of \( Z \) in the form

\[
P[Z \leq x] = \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_r=1}^{M_{i_r}} \prod_{i=1}^{r} F_{i_i}(x)_{i_i + 2i_i},
\]

where by our notation in (21) \( H_n^m(x; 1, 1) \) is the d.f. of

\[
\chi^2_{m} - \chi^2_{n},
\]

when both chi-square variates are independent. For convenience we shall write

\[
H_n^m(x; 1, 1) = H_n^m(x).
\]

Therefore, by (24) and (25) we get
\begin{equation}
G(x) = \sum_{i_1} \cdots \sum_{i_r} \sum_{j_1} \cdots \sum_{j_r} c_{i_1} \ldots i_r c_j \sum_{d_j} (i_1, \ldots, i_r) (j_1, \ldots, j_r)
\end{equation}

\begin{equation}
p_1(\theta_1) \cdots p_r(\theta_r) p_1(\xi_1) \cdots p_r(\xi_r) c_{i_1} \ldots i_r c_j \sum_{d_j} (i_1, \ldots, i_r) (j_1, \ldots, j_r)
\end{equation}

\begin{equation}
M_{i_1}^{j_1}, \ldots, M_{i_r}^{j_r} + 2i
\end{equation}

\begin{equation}
N_{j_1}, \ldots, N_{j_r} + 2j
\end{equation}

We should remark that the density which corresponds to $H_n^m(x)$, i.e. the density $h_n^m(x)$, say, of

\begin{equation}
x^2 \frac{m}{n} - x^2
\end{equation}

can be very easily obtained (see, e.g., proof of Theorem 3.2.1 of [1]). We give here its expression in terms of the confluent hypergeometric function defined for $t > 0$ by

\begin{equation}
W_{k,s}(t) = \frac{e^{-\frac{1}{2}t}}{\Gamma(s-k+\frac{1}{2})} \int_0^\infty u^{s-k-\frac{1}{2}} (1+\frac{u}{t})^{s-k-\frac{1}{2}} e^{-u} du.
\end{equation}

Thus, we have

\begin{equation}
h_n^m(x) = \frac{1}{1 - \frac{1}{2(m+n)}} x \frac{m+n}{4} - 1 \quad W_{k,s}(x), \quad x > 0,
\end{equation}

(28)

\begin{equation}
h_n^m(x) = h_n^m(-x), \quad x < 0,
\end{equation}

where

\begin{equation}
k = \frac{m-n}{4}, \quad s = \frac{m+n}{4} - \frac{1}{2}.
\end{equation}
Summing up our discussion above we have, by (23) also, the main result of this paper:

**Theorem 2.** The d.f. of the topothetical criterion \( d \) in its reduced form in (20) is

\[
P[d \leq x] = G\left(\frac{x}{a}\right),
\]

where \( G(x) \) is given by (27).

For an application of Lemma 4 it is more practical to use the expression of \( G(x) \) in (24) where the coefficients \( p_{i}^{s}(\theta_{s}) \) and \( p_{j}^{s}(\zeta_{s}) \), \( s = 1, \ldots, r \) are Poisson probabilities (cf. (22)*). Thus by Theorem 2 also, we deduce

**Corollary 4.** For any integers

\[
N_{s} \geq n_{s} \geq 0, \quad s = 1, \ldots, r.
\]

\[
M_{s} \geq m_{s} \geq 0,
\]

we have

\[
P[d \leq x] \leq \prod_{s=1}^{r} \left(1 - \sum_{i_{s}=n_{s}}^{N_{s}} p_{i_{s}}^{s}(\theta_{s})\right) \left(1 - \sum_{j_{s}=m_{s}}^{M_{s}} p_{j_{s}}^{s}(\zeta_{s})\right),
\]

where \( G_{i_{1}, \ldots, i_{r}}^{j_{1}, \ldots, j_{r}} \left(\frac{x}{a}\right) \) we have denoted the d.f. of \( Z \) in (25).
REFERENCES


STANFORD UNIVERSITY
TECHNICAL REPORT DISTRIBUTION LIST

Informe No. CS-69(3)

Commanding Officer
Frankford Arsenal
Library Branch, G700, Bldg. 40
Philadelphia 37, Pennsylvania

Document Library
U.S. Atomic Energy Commission
1942 14th and Constitution Ave. N.W.
Washington 25, D. C.

Commanding General
Rock Island Arsenal
Rock Island, Illinois

Headquarters
Oklahoma City Materiel Area
United States Air Force
Vick Air Force Base, Oklahoma

Commanding General
Redstone Arsenal (OHIO-W-07)
Huntsville, Alabama

Institute of Statistics
North Carolina State College of A & E
Raleigh, North Carolina

Commanding General
White Sands Proving Ground
Las Cruces, New Mexico

Jet Propulsion Laboratory
California Institute of Technology
Attn: A. J. Scoula
4800 Oak Grove Drive
Pasadena 3, California

Commanding General
White Sands Proving Ground
Las Cruces, New Mexico

Librarian
The RAND Corporation
1700 Main Street
Santa Monica, California

Commanding General
Signal Corps Engineering Laboratory
Fort Monmouth, New Jersey

Librarian Division
Naval Missile Center Command
U.S. Naval Missile Center
Attn: J. L. Nickel
Point Magu, California

Commander
Wright Air Development Center
Attn: NRL Tech. Library, WADC
Wright-Patterson Air Force Base, Ohio

Libraries Division
Code 5077
U.S. Naval Ordnance Test Station
China Lake, California

Commander
Western Development Division, WADD
P.O. Box 262
Inglewood, California

Mailroom
U.S. Navy (Office of Aeronautical Sciences)
1254 F Street, N.W.
Washington 25, D.C.

Commander
Western Development Division, WEIT
P.O. Box 262
Inglewood, California

National Science Foundation
National Bureau of Standards
Washington 25, D.C.

Commanding Officer
Waterways Arsenal (OHIO-W-07)
Attn: W. A. Lube
Waterways Arsenal
Waterdown, Massachusetts

Naval Inspector of Ordnance
U.S. Naval Ordnance Factory
Washington 25, D. C.
Attn: Mrs. G. D. Hoke

Commanding Officer
Waterdown, Massachusetts

Office, Ass't Chief of Staff, 5-A
Research Branch, A & D Division
Department of the Army
Washington 25, D. C.

Commanding Officer
Waterdown, Massachusetts

Office of Technical Services
Department of Commerce
Washington 25, D. C.

Commanding Officer
Waterdown, Massachusetts

Technical Information Officer
Naval Research Laboratory
Washington 25, D. C.

Commanding Officer
Waterdown, Massachusetts

Technical Information Service
Attn: Reference Branch
P.O. Box 62
Oak Ridge, Tennessee

Commanding Officer
Waterdown, Massachusetts

Technical Library Branch
Code 29A
U.S. Naval Ordnance Laboratory
Attn: Clayborn Graves
Corona, California

March 31, 1961
<table>
<thead>
<tr>
<th>Name</th>
<th>Position and Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr. James J. Fleming</td>
<td>Head, Operational Research Branch, U.S. Naval Research Laboratory, Washington 25, D.C.</td>
</tr>
<tr>
<td>Mr. Fred Frishman</td>
<td>Chairman, Army Research Office, Arlington Hall Station, Arlington, Virginia</td>
</tr>
<tr>
<td>Mrs. Dorothy M. Gilford</td>
<td>Logistics and Mathematical Statistics Branch, Office of Naval Research, Washington 25, D.C.</td>
</tr>
<tr>
<td>Dr. Robert Lundegard</td>
<td>Logistics and Mathematical Statistics Branch, Office of Naval Research, Washington 25, D.C.</td>
</tr>
<tr>
<td>Dr. Clifford Maloney</td>
<td>Biomathematics Division, Chemical Corps, U.S. Army Biological Laboratories, Fort Detrick, Maryland</td>
</tr>
<tr>
<td>Mr. R. H. Noyes</td>
<td>Inst. for Exploratory Research, USASRD, Fort Monmouth, New Jersey</td>
</tr>
<tr>
<td>Major Oliver A. Shaw, Jr.</td>
<td>Mathematics Division, Air Force Office of Scientific Research, Washington 25, D.C.</td>
</tr>
<tr>
<td>Dr. Horace M. Trent, Head</td>
<td>Head, Applied Mathematics Branch, U.S. Naval Research Laboratory, Washington 25, D.C.</td>
</tr>
<tr>
<td>Mr. J. Weinstein</td>
<td>Institute for Exploratory Research, USASRD, Fort Monmouth, New Jersey</td>
</tr>
</tbody>
</table>