where

$$\lambda_1' = \left(-d_\alpha' \sqrt{\frac{1}{n} + \frac{1}{m}} \pm \Delta\right) \Big/ \sqrt{\frac{F(x_0)[1 - F(x_0)]}{n} + \frac{F'(x_0)[1 - F'(x_0)]}{m}}$$

and

$$\lambda_2' = \left(d_\alpha' \sqrt{\frac{1}{n} + \frac{1}{m}} \pm \Delta\right) \Big/ \sqrt{\frac{F(x_0)[1 - F(x_0)]}{n} + \frac{F'(x_0)[1 - F'(x_0)]}{m}}.$$

Since this lower bound approaches one as n and m approach infinity the power also approaches one and the test is consistent.

REFERENCES

[1] J. Wolfwitz, "Non-parametric statistical inference," *Proceedings of the Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 93–113.
[2] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 279–281.
[3] F. Massey, "A note on the estimation of a distribution function by confidence limits," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 116–120.
[4] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bull. Math. Univ. Moscou, Série Int.*, Vol. 2, fasc. 2 (1939).

ON OPTIMUM SELECTIONS FROM MULTINORMAL POPULATIONS[1]

By Z. W. Birnbaum and D. G. Chapman[2]

University of Washington

1. Introduction. Let $Y_1, Y_2, \cdots, Y_n$ be scores in n admission tests such as those used in educational institutions, personnel selection, or testing of materials, and let these scores be used as a basis for selecting a sub-population Π^* from an initial population Π. This selection is usually performed in such a manner that an achievement or performance score X has a distribution in Π^*, which shows some required improvement over the distribution of X in Π; such an improvement may for example consist in changing the expectation $E(X)$ of X in Π to a pre-assigned value $E^*(X)$ in Π^*. Among all selection procedures based on $Y_1, \cdots, Y_n$ and achieving the required improvement of the distribution of X, it appears desirable to find those which retain as large a portion of Π as possible. It will be shown that under certain assumptions the linear truncations studied in an earlier paper [1] are such optimal selections.

2. Selection, truncation, linear truncation. Let the frequency of individuals with the scores $(X, Y_1, \cdots, Y_n)$ be $F(X, Y_1, \cdots, Y_n)$ in Π and

[1] Presented at the New York meeting of the Institute of Mathematical Statistics on December 27, 1949.
[2] Research done under the sponsorship of the Office of Naval Research.

$$F^*(X, Y_1, \cdots, Y_n)$$

in Π^*. Since Π^* was obtained by selection from Π, we have $F^*/F \leq 1$, and since the selection was made solely on the basis of the values of $Y_1, \cdots, Y_n$, the ratio F^*/F is independent of X. We thus have

$$\frac{F^*(X, Y_1, \cdots, Y_n)}{F(X, Y_1, \cdots, Y_n)} = \varphi(Y_1, \cdots, Y_n)$$

and

$$0 \leqq \varphi(Y_1, \cdots, Y_n) \leq 1. \tag{2.1}$$

Let $N = \iint \cdots \int F(X, Y_1, \cdots, Y_n)\, dX\, dY_1 \cdots dY_n$ and

$$N^* = \iint \cdots \int F^*(X, Y_1, \cdots, Y_n)\, dX\, dY_1 \cdots dY_n$$

be the number of individuals in Π and Π^*, and $f(X, Y_1, \cdots, Y_n)$ and $f^*(X, Y_1, \cdots, Y_n)$ the distribution densities in Π and Π^*, respectively, so that $F = Nf$, $F^* = N^*f^*$ and $\iint \cdots \int f\, dX\, dY_1 \cdots dY_n = \iint \cdots \int f^*\, dX\, dY_1 \cdots dY_n = 1$. We then have

$$N^*f^* = \varphi Nf,$$

and

$$\frac{N^*}{N} = \iint \cdots \int \varphi(Y_1, \cdots, Y_n) f(X, Y_1, \cdots, Y_n)\, dX\, dY_1 \cdots dY_n. \tag{2.2}$$

Thus any selection of a subpopulation Π^* from Π based only on $Y_1, \cdots, Y_n$, defines a $\varphi(Y_1, \cdots, Y_n)$ satisfying (2.1). Conversely, if the frequencies

$$F(X, Y_1, \cdots, Y_n)$$

in Π are given, any measurable $\varphi(Y_1, \cdots, Y_n)$ satisfying (2.1) defines new frequencies $F^* = \varphi F$ and hence a selection from Π based only on $Y_1, \cdots, Y_n$.

These considerations lead to the following definitions:

A measurable function $\varphi(Y_1, \cdots, Y_n)$ which satisfies (2.1) is called a *selection in* $Y_1, \cdots, Y_n$. If, in particular, φ is the characteristic function of a set Ω in $(Y_1, \cdots, Y_n)$, that is $\varphi = 1$ in Ω and $\varphi = 0$ in $\bar{\Omega}$, then the selection φ will be called a *truncation in* $Y_1, \cdots, Y_n$ *to the set* Ω. If Ω is defined by a condition of the form

$$\sum_{j=1}^{n} a_j Y_j \geq t$$

with constant a_j, t, then the truncation to the set Ω will be called a *linear truncation in* $Y_1, \cdots, Y_n$.

In view of (2.2) we will refer to

$$(2.3)\quad r(\varphi) = \int\int \cdots \int \varphi(Y_1, \cdots, Y_n) f(X, Y_1, \cdots, Y_n)\, dX\, dY_1, \cdots dY_n$$

as the *fraction retained in the selection* φ.

3. A lemma. We will need the following slight generalization of the fundamental lemma of Neyman-Pearson (cf. [2]).

LEMMA. *Let* $G(Y_1, \cdots, Y_n)$, $G_1(Y_1, \cdots, Y_n)$, $\cdots$, $G_m(Y_1, \cdots, Y_n)$ *be given integrable functions and* $c_1, \cdots, c_m$ *given constants, and let* (ϕ) *be the family of all measurable functions* $\varphi(Y_1, \cdots, Y_n)$ *which satisfy the conditions*

$$(3.1)\qquad 0 \leq \varphi(Y_1, \cdots, Y_n) \leq 1$$

$$(3.2)\quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi(Y_1, \cdots, Y_n) G_i(Y_1, \cdots, Y_n)\, dY_1 \cdots dY_n = c_i \qquad \text{for} \quad i = 1, \cdots, m.$$

If there exist constants $k_1, \cdots, k_m$ *such that the characteristic function* $\varphi_0(Y_1, \cdots, Y_n)$ *of the set* $\underset{(Y_1,\cdots,Y_n)}{E}\left[G \geq \sum_{i=1}^{m} k_i G_i\right] = E$ *belongs to* (ϕ), *then*

$$(3.3)\quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi_0 G\, dY_1 \cdots dY_n \geq \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi G\, dY_1 \cdots dY_n$$

for any φ *in* (ϕ).

PROOF: We have $\varphi_0' = 1 \geq \varphi$ in E and $\varphi_0 = 0 \leq \varphi$ in $\bar{E}$, hence

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(G - \sum_{i=1}^{m} k_i G_i\right) \varphi_0\, dY_1 \cdots dY_n \geq \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(G - \sum_{i=1}^{m} k_i G_i\right) \varphi\, dY_1 \cdots dY_n,$$

and (3.3) follows since φ_0 and φ fulfill (3.2).

4. Selection from a multivariate normal population, for which the fraction retained is maximum. From now on we assume that the conditional distribution of X for given $Y_1, Y_2, \cdots, Y_n$ is normal with a mean which is a linear function of the Y's and with a variance which is independent of them, i.e.,

$$(4.1)\qquad f(X \mid Y_1, Y_2, \cdots, Y_n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-\left(X - \sum_{i=1}^{n} \rho_i Y_i\right)^2}{2\sigma^2}\right].$$

Let $Q(Y_1, \cdots, Y_n)$ denote the marginal density of $Y_1, \cdots, Y_n$.

THEOREM 1. *A selection such that*

1° *in* Π^* *a proportion at most equal to a given proper fraction* ϵ *has values of* X *below* X_0, *i.e. the* ϵ*-quantile in* Π^* *is greater than or equal to* X_0, *when* X_0 *is a given number greater than the* ϵ*-quantile in* Π,

2° *the fraction retained is maximum,*

is a linear truncation.

PROOF: We have to maximize

$$r(\varphi) = \int \cdots \int \varphi(Y_1, \cdots, Y_n)Q(Y_1, \cdots, Y_n)\, dY_1 \cdots dY_n \tag{4.2}$$

under the condition

$$\frac{\displaystyle\int_{-\infty}^{X_0}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi(Y_1,\cdots,Y_n)Q(Y_1,\cdots,Y_n)f(X \mid Y_1,\cdots,Y_n)\,dY_1\cdots dY_n dX}{\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi(Y_1,\cdots,Y_n)Q(Y_1,\cdots,Y_n)f(X \mid Y_1,\cdots,Y_n)\,dY_1\cdots dY_n dX} \leq \epsilon.$$

Substituting the expression (4.1) for $f(X \mid Y_1, \cdots, Y_n)$ and integrating with respect to X we may rewrite this in the form

$$\begin{aligned} L(\varphi) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} & \varphi(Y_1, \cdots, Y_n)Q(Y_1, \cdots, Y_n) \\ & \cdot \left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma}\right) - \epsilon\right] dY_1 \cdots dY_n \leq 0, \end{aligned} \tag{4.3}$$

where

$$\psi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2}\, dt,$$

and we have to maximize (4.2) under condition (4.3).

Without loss of generality the inequality $L(\varphi) \leq 0$ in (4.3) may be replaced by equality. For if we had a selection φ_1 which maximizes (4.2) and satisfies (4.3) with a strict inequality $L(\varphi_1) < 0$, then φ_1 could not be equal to 1 almost everywhere since then we would have $F^* = F$ almost everywhere and X_0 would be equal to the ϵ-quantile in Π, in contradiction with 1°; hence $\varphi_2 = \varphi_1 + \alpha(1 - \varphi_1)$ for sufficiently small $\alpha > 0$ would also satisfy (4.3) with a strict inequality but would yield $r(\varphi_2) > r(\varphi_1)$.

To solve our problem we now have to maximize (4.2) under the condition

$$L(\varphi) = 0. \tag{4.4}$$

Applying the lemma of Section 3, with $m = 1$, and

$$G(Y_1, \cdots, Y_n) = Q(Y_1, \cdots, Y_n),$$

$$G_1(Y_1, \cdots, Y_n) = Q(Y_1, \cdots, Y_n)\left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma}\right) - \epsilon\right],$$

we conclude that the selection satisfying 1° and 2° will be the characteristic function $\varphi_0(Y_1, \cdots, Y_n)$ of the set defined by

$$k\left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma}\right) - \epsilon\right] \leq 1, \tag{4.5}$$

provided k can be determined so that φ_0 satisfies (4.4).

To find such a k we consider

$$I(t) = \int_{\sum_{i=1}^{n} \rho_i Y_i \geq t} \cdots \int Q(Y_1, \cdots, Y_n) \left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma} \right) - \epsilon \right] dY_1 \cdots dY_n .$$

As t tends to $-\infty$, $I(t)$ tends to $L(1)$, where L was defined by (4.3). Since the ϵ-quantile in Π was less than X_0 it follows that $I(-\infty) = L(1) > 0$. Since $I(t) < 0$ for large t, there exists t_0 such that $I(t_0) = 0$, and clearly,

$$\psi\left(\frac{X_0 - t_0}{\sigma} \right) - \epsilon > 0.$$

Setting in (4.5) $k = [\psi((X_0 - t_0)/\sigma) - \epsilon]^{-1}$, one obtains a φ_0 such that

$$L(\varphi_0) = I(t_0) = 0.$$

The selection φ_0 is the linear truncation to the set $\sum_{i=1}^{n} \rho_i Y_i \geq t_0$.

By a similar and somewhat simpler argument one proves the following theorem.

THEOREM 2. *A selection such that*

1° *in* Π^* *the mean of* X *has a value greater than or equal to a pre-assigned number* $m > 0$,

2° *the fraction retained is maximum,*

is a linear truncation to a set $\sum_{i=1}^{n} \rho_i Y_i \geq t_0$.

An immediate consequence of Theorems 1 and 2 is that a linear truncation, using a properly determined weighted score $\sum_{i=1}^{n} \rho_i Y_i$ and cutting score t_0, is more economical than any truncation to a set $Y_i \geq t_i$, $i = 1, 2, \cdots, n$, that is than any truncation performed on each admission score separately.

REFERENCES

[1] Z. W. BIRNBAUM, "Effect of linear truncation on a multinormal population," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 272–279.

[2] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. I (1936), pp. 1–37, particularly pp. 10–11.

THE DISTRIBUTION OF DISTANCE IN A HYPERSPHERE

BY J. M. HAMMERSLEY

University of Oxford

1. Summary. Deltheil ([1], pp. 114–120) has considered the distribution of distance in an n-dimensional hypersphere. In this paper I put his results (17) in a more compact form (16); and I investigate in greater detail the asymptotic form of the distribution for large n, for which the rather surprising result emerges that this distance is almost always nearly equal to the distance between the