EFFECT OF LINEAR TRUNCATION ON A MULTINORMAL POPULATION[1]

By Z. W. Birnbaum[2]

University of Washington

1. Introduction. In admission to educational institutions, personnel selection, testing of materials, and other practical situations, the following mathematical model is frequently encountered: A $(k + l)$-dimensional random variable $(X_1, X_2, \cdots, X_k, Y_1, Y_2, \cdots, Y_l) = (\mathbf{X}, \mathbf{Y})$ is considered, with a joint probability-distribution assumed to be non-singular multi-normal. The $Y_1, Y_2, \cdots, Y_l$ are scores in admission tests, the $X_1, X_2, \cdots, X_k$ scores in achievement tests. The admission tests are administered to all individuals in the $(\mathbf{X}, \mathbf{Y})$ population to decide on admission or rejection, and (usually at some later time) the achievement tests are administered to those admitted. A set of weights $a_j \geq 0, j = 1, 2, \cdots, l$ is used to define a composite admission test score $U = \sum_{j=1}^{l} a_j Y_j$ and a "cutting score" τ is chosen so that an individual is admitted if $U \geq \tau$, and rejected if $U < \tau$. We will refer to this procedure as *linear truncation of* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathbf{Y}$ *to the set* $U \geq \tau$.

A linear truncation in $\mathbf{Y}$ clearly will change the absolute distribution of $\mathbf{X}$, except in the case of independence. In this paper a study is made of the absolute distribution of $\mathbf{X}$ after linear truncation in $\mathbf{Y}$ in the case $k = 1$; in particular, the possibility is investigated of choosing the a_j and τ in such a way that the distribution of $\mathbf{X}$ after truncation has certain desirable properties. The case $k > 1$ leads to a considerable diversity of problems which are being studied and, it is hoped, will be the subject of a separate paper.

Throughout this paper it will be assumed that all the parameters of $(\mathbf{X}, \mathbf{Y})$, that is the expectations, variances and covariances before truncation, are known. In practical situations it often happens that only the parameters of $Y_1, Y_2, \cdots, Y_l$ before truncation are known, while the first and second moments involving $X_1, X_2, \cdots, X_k$ are only known for the joint distribution after truncation. It can be shown [1] that in such situations the expectations, variances and covariances of $(\mathbf{X}, \mathbf{Y})$ before truncation can always be reconstructed if $(\mathbf{X}, \mathbf{Y})$ has a multinormal distribution.

In the simplest case $k = l = 1$ the probability-density of the original bi-normal random variable (X, Y) may be, without loss of generality, assumed equal to

$$f(X, Y, \rho) = \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-(X^2 - 2\rho XY + Y^2)/2(1-\rho^2)}. \tag{1.1}$$

By truncating this distribution in Y to the set $Y \geq \tau$ one obtains the probability-density

$$g(X, Y; \rho, \tau) = \begin{cases} \psi^{-1}(\tau) f(X, Y; \rho), & \text{for } Y \geq \tau, \\ 0, & \text{for } Y < \tau, \end{cases} \tag{1.2}$$

[1] Presented to the Institute of Mathematical Statistics on June 18, 1949.

[2] Research done under the sponsorship of the Office of Naval Research.

where

$$\psi(\tau) = \frac{1}{\sqrt{2\pi}} \int_\tau^\infty e^{-t^2/2}\, dt. \tag{1.3}$$

For further use we introduce the abbreviations

$$\varphi(\tau) = \frac{1}{\sqrt{2\pi}} e^{-\tau^2/2}, \tag{1.4}$$

$$\lambda(\tau) = \frac{\varphi(\tau)}{\psi(\tau)}. \tag{1.5}$$

We also note the inequalities

$$\tau \leq \lambda(\tau) \tag{1.6}$$

and

$$\lambda(\tau) \leq \frac{\sqrt{4 + \tau^2} - \tau}{2} \tag{1.7}$$

derived in [2] and [3], respectively.[3]

Before proceeding to the more-dimensional case, we will study some properties of the marginal probability-distribution of X after truncation to $Y \geq \tau$

$$\varphi(X; \rho, \tau) = \int_\tau^\infty g(X, Y; \rho, \tau)\, dY. \tag{1.8}$$

2. The moments of $\varphi(X; \rho, \tau)$. In this section all mathematical expectations are computed for the absolute distribution of X after truncation of (X, Y) to $Y \geq \tau$.

We have

$$\varphi(X; \rho, \tau) = \psi^{-1}(\tau)\varphi(X)\psi\left(\frac{\tau - \rho X}{\sqrt{1 - \rho^2}}\right),$$

and hence

$$\begin{aligned} E(X^n) &= \int_{-\infty}^{+\infty} X^n \varphi(X; \rho, \tau)\, dX \\ &= \psi^{-1}(\tau) \int_{-\infty}^{+\infty} \frac{X}{\sqrt{2\pi}} e^{-X^2/2} \frac{X^{n-1}}{\sqrt{2\pi}} \int_{(\tau-\rho X)/\sqrt{1-\rho^2}}^{\infty} e^{-S^2/2}\, dS\, dX \\ &= \psi^{-1}(\tau) \left\{ -\varphi(X) X^{n-1} \psi\left(\frac{\tau - \rho X}{\sqrt{1 - \rho^2}}\right) \Big|_{x=-\infty}^{+\infty} \right. \\ &\quad + \int_{-\infty}^{+\infty} \varphi(X) \left[\frac{dX^{n-1}}{dX} \psi\left(\frac{\tau - \rho X}{\sqrt{1 - \rho^2}}\right)\right. \end{aligned}$$

[3] Implicitly, the inequality (1.6) was known already to Laplace, cf. *Mécanique Céleste*, transl. by Bowditch, Boston 1839, Vol. 4, p. 493.

$$+ X^{n-1} \frac{\rho}{\sqrt{1-\rho^2}} \varphi\left(\frac{\tau - \rho X}{\sqrt{1-\rho^2}}\right)\Bigg] dX \Bigg\}$$

$$= E\left(\frac{dX^{n-1}}{dX}\right) + \frac{\rho}{\psi(\tau)\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} X^{n-1}\varphi(X)\varphi\left(\frac{\tau - \rho X}{\sqrt{1-\rho^2}}\right) dX.$$

From the identity

$$\varphi(X)\varphi\left(\frac{\tau - \rho X}{\sqrt{1-\rho^2}}\right) = \varphi(\tau)\varphi\left(\frac{X - \rho\tau}{\sqrt{1-\rho^2}}\right) \tag{2.0}$$

we obtain

$$\int_{-\infty}^{+\infty} X^{n-1}\varphi(X)\varphi\left(\frac{\tau - \rho X}{\sqrt{1-\rho^2}}\right) dX = \varphi(\tau) \int_{-\infty}^{+\infty} X^{n-1}\varphi\left(\frac{X - \rho\tau}{\sqrt{1-\rho^2}}\right) dX$$

$$= \sqrt{1-\rho^2}\,\varphi(\tau) \int_{-\infty}^{+\infty} (S\sqrt{1-\rho^2} + \rho\tau)^{n-1}\varphi(S)\, dS,$$

and hence

$$E(X^n) = E\left(\frac{dX^{n-1}}{dX}\right) + \rho\lambda(\tau) \int_{-\infty}^{+\infty} (S\sqrt{1-\rho^2} + \rho\tau)^{n-1}\varphi(S)\, dS, \tag{2.1}$$

for $n \geq 1$.

For $n = 1$ this yields for the expectation of X after truncation

$$E(X) = \rho\lambda(\tau). \tag{2.2}$$

For $n = 2$ we have from (2.1)

$$E(X^2) = 1 + \rho^2\tau\lambda(\tau) = 1 + \rho\tau E(X),$$

and hence for the variance of X after truncation the expression

$$\sigma^2(X) = 1 + E(X)[\rho\tau - E(X)], \tag{2.3}$$

or

$$\sigma^2(X) = 1 - \rho^2\lambda(\tau)[\lambda(\tau) - \tau]. \tag{2.31}$$

From (2.2) we see that $E(X)$ always has the sign of ρ, as one would expect. From (2.3) one finds a lower bound for τ

$$\tau > \frac{E^2(X) - 1}{\rho E(X)}. \tag{2.4}$$

From (2.31) and (1.6) one concludes that $\sigma^2(X) < 1$ for $\rho \neq 0$, hence the variance of X after truncation is always less than the variance before truncation, except if $\rho = 0$.

Similarly one computes from (2.1) the third moment about zero

$$E(X^3) = E(X)[3 - \rho^2(1 - \tau^2)]$$

and obtains for the third moment about the expectation

$$E[X - E(X)]^3 = E(X)\rho^2\{[\lambda(\tau) - \tau][2\lambda(\tau) - \tau] - 1\}. \tag{2.5}$$

Numerical computation indicates that the quantity in braces is always >0, which would mean that the skewness of X after truncation has the same sign as $E(X)$ and ρ. No analytic proof of this statement has been obtained.

3. Determination of τ for given expectation or quantile of X after truncation; dependence of this τ on ρ. Let it be required to determine τ so that the expectation of X after truncation assumes a given value m. It follows immediately from (2.2) that this τ is obtained by solving the equation

$$\lambda(\tau) = \frac{m}{\rho} \tag{3.1}$$

for τ, which can be done with the aid of a table[4] of $\lambda(\tau)$.

Another problem which occurs in applications consists in determining τ so that, for given $0 < \alpha < 1$ and X_α, the α-quantile for X after truncation assumes the value X_α, that is so that

$$\int_{-\infty}^{X_\alpha} \varphi(X; \rho, \tau)\, dX = \psi^{-1}(\tau) \int_{-\infty}^{X_\alpha} \int_{\tau}^{\infty} f(X, Y; \rho)\, dY\, dX = \alpha. \tag{3.2}$$

Let

$$P(s, t; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_s^\infty \int_t^\infty e^{-[(X^2-2\rho XY+Y^2)/2(1-\rho^2)]}\, dY\, dX \tag{3.21}$$

denote the volume of the probability solid $Z = f(X, Y; \rho)$ above the quadrant $X \geq s$, $Y \geq t$. Then (3.2) may be written in the form

$$1 - \frac{P(X_\alpha, \tau; \rho)}{\psi(\tau)} = \alpha,$$

or

$$(1 - \alpha)\psi(\tau) = P(X_\alpha, \tau; \rho), \tag{3.3}$$

and this equation can be solved for τ by trial with the aid of tables of $\psi(\tau)$ and Pearson's tables [4] of $P(s, t; \rho)$,

LEMMA 1. *For fixed expectation of X after truncation $E(X) = m$, the solution $\tau(\rho)$ of (3.1) is a strictly decreasing function of the absolute value of ρ for $0 < |\rho| \leq 1$.*

PROOF: Differentiating $m = \rho\lambda(\tau)$ with regard to ρ one obtains

$$0 = \lambda(\tau) + \rho\lambda'(\tau)\frac{d\tau}{d\rho}$$

and, in view of the identity

$$\lambda'(\tau) = \lambda(\tau)[\lambda(\tau) - \tau],$$

the expression

$$\frac{d\tau}{d\rho} = -\frac{1}{\rho[\lambda(\tau) - \tau]}. \tag{3.4}$$

[4] A table of $1/\lambda(\tau)$ is, for example, given in Karl Pearson, *Tables for Statisticians and Biometricians, Part II*, 1931, pp. 11–15.

From (3.4) and (1.6) we see that

$$\text{sign } \frac{d\tau}{d\rho} = -\text{sign } \rho,$$

which proves our lemma.

LEMMA 2. *For fixed* α, X_α, *the solution* $\tau = \tau(\rho)$ *of* (3.3) *is a strictly decreasing function of* $|\rho|$ *for* $0 < |\rho| \le 1$.

PROOF: Differentiating (3.3) with regard to ρ one obtains

$$-(1 - \alpha)\varphi(\tau) \frac{d\tau}{d\rho} = \frac{\partial P}{\partial \tau} \frac{d\tau}{d\rho} + \frac{\partial P}{\partial \rho},$$

and hence

$$\frac{d\tau}{d\rho} = \frac{-\dfrac{\partial P}{\partial \rho}}{\dfrac{\partial P}{\partial \tau} + (1 - \alpha)\varphi(\tau)}. \tag{3.6}$$

From (3.21) one easily verifies that

$$\frac{\partial P(X_\alpha, \tau; \rho)}{\partial \rho} = \varphi(\tau)(1 - \rho^2)^{-\frac{1}{2}} \int_{(X_\alpha - \rho\tau)/\sqrt{1-\rho^2}}^{\infty} te^{-t^2/2}\, dt,$$

and therefore

$$\frac{\partial P(X_\alpha, \tau; \rho)}{\partial \rho} > 0. \tag{3.7}$$

One also computes

$$\frac{\partial P(X_\alpha, \tau; \rho)}{\partial \tau} = -\varphi(\tau)\psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right),$$

so that the denominator of the right hand expression in (3.6) becomes

$$\varphi(\tau)\left[1 - \alpha - \psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right)\right].$$

In view of (3.3) this is equal to

$$\begin{aligned}
\varphi(\tau) &\left[\frac{P(X_\alpha, \tau; \rho)}{\psi(\tau)} - \psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right)\right] \\
&= \lambda(\tau)\left[P(X_\alpha, \tau; \rho) - \psi(\tau)\psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right)\right] \\
&= \lambda(\tau) \frac{1}{2\pi} \int_\tau^\infty e^{-Y^2/2} \int_{(X_\alpha - \rho Y)/\sqrt{1-\rho^2}}^{(X_\alpha - \rho\tau)/\sqrt{1-\rho^2}} e^{-U^2/2}\, dU\, dY \\
&= \lambda(\tau) \frac{1}{2\pi} \int_\tau^\infty h(Y)\, dY.
\end{aligned}$$

If $\rho > 0$, then $\rho Y > \rho\tau$ in the interval of integration $\tau < Y < \infty$, hence $\frac{X_\alpha - \rho Y}{\sqrt{1-\rho^2}} < \frac{X_\alpha - \rho\tau}{\sqrt{1-\rho^2}}$, therefore the integrand $h(Y)$ is positive, and so is the denominator of (3.6). Similarly one sees that if $\rho < 0$ the integrand $h(Y)$ is negative for $\tau < Y < \infty$ and the denominator of (3.6) is negative. In view of (3.7) we conclude

$$\text{sign}\,\frac{d\tau}{d\rho} = -\text{sign}\,\rho \qquad \text{for } \rho \neq 0.$$

4. Linear truncation of $(X, Y_1, Y_2, \cdots, Y_l)$ to the set $\sum_{j=1}^{l} a_j Y_j \geq \tau$ for given expectation or quantile of X, minimizing the rejected part of the population. Let $(X, Y_1, Y_2, \cdots, Y_l)$ be an $(l+1)$-dimensional non-singular normal random variable with all expectations, variances and covariances known. We wish to choose $a_1, a_2, \cdots, a_l$ and τ so that by setting

$$U = \sum_{j=1}^{l} a_j Y_j \tag{4.1}$$

and performing the linear truncation to the set $U \geq \tau$ we obtain for the expectation of X after truncation a pre-assigned value m, and that this is achieved with the least waste of the original population, that is so that for the non-truncated probability-distribution the probability $P(\sum_{j=1}^{l} a_j Y_j < \tau)$ is minimum.

Without loss of generality we may assume that, before truncation, we have

$$E(X) = E(Y_1) = \cdots = E(Y_l) = 0, \tag{4.21}$$

$$\sigma^2(X) = 1, \tag{4.22}$$

and thus

$$E(U) = 0. \tag{4.3}$$

Furthermore, the a_j and τ can always be multiplied by a constant, without changing the set of truncation, so that we have

$$\sigma^2(U) = 1. \tag{4.4}$$

THEOREM 1. *To truncate $(X, Y_1, Y_2, \cdots, Y_l)$ linearly in $Y_1, Y_2, \cdots, Y_l$ so that the expectation of X after truncation has the given value m and that the probability of the rejected part of the original population is minimum, it is necessary and sufficient* (1) *to determine $a_1, a_2, \cdots, a_l$ so that the absolute value of the correlation-coefficient $\rho(X, U)$ becomes maximum under the condition* (4.4), *and* (2) *for U determined by these $a_1, a_2, \cdots, a_l$ and for $\rho = \rho(X, U)$ to solve equation* (3.1) *for τ.*

The proof of this theorem follows immediately from the first paragraph of section 3 and Lemma 1.

Using the second paragraph of section 3 and Lemma 2, one equally easily arrives at the following theorem:

THEOREM 2. *To truncate $(X, Y_1, Y_2, \cdots, Y_l)$ linearly in $Y_1, Y_2, \cdots, Y_l$*

so that the α-quantile of X after truncation has the given value X_α and that the probability of the rejected part of the original population is minimum, it is necessary and sufficient to satisfy (1) *in Theorem* 1 *and then to solve equation* (3.3).

The problem of satisfying requirement (1) of Theorems 1 and 2 can be solved effectively by a method due to Hotelling [5]. It may be worth noting that this method yields two sets of constants, $a_1, a_2, \cdots, a_l$ and $-a_1, -a_2, \cdots, -a_l$ both maximizing $|\rho(X, U)|$ but leading to values of $\rho(X, U)$ with opposite signs. Nevertheless the choice between $a_1, a_2, \cdots, a_l$ and $-a_1, -a_2, \cdots, -a_l$ and the determination of τ are unique for any given m, since (3.1) has a solution for τ only if sign $\rho =$ sign m.

5. Linear truncation of $(X, Y_1, Y_2, \cdots, Y_l)$ to the set $\sum_{j=1}^{l} a_j Y_j \geq \tau$ for given expectation of X after truncation, minimizing the variance of X after truncation. It may be of practical interest to choose $a_1, a_2, \cdots, a_l$ and τ so that, with the notations and under the assumptions of section 4, the expectation of X after truncation becomes equal to a given number m, and the variance after truncation is minimum.

THEOREM 3. *To truncate $(X, Y_1, Y_2, \cdots, Y_l)$ linearly in $Y_1, Y_2, \cdots, Y_l$ so that the expectation after truncation has the given value m and that, under this condition, the variance of X after truncation becomes as small as possible, it is necessary and sufficient to satisfy the conditions* (1) *and* (2) *of Theorem* 1.

The proof of this theorem follows from section 3 and the following lemma:

LEMMA 3. *For fixed $E(X) = m$, the variance $\sigma^2(X)$ after truncation is a strictly decreasing function of the absolute value of ρ for $0 < |\rho| \leq 1$.*

PROOF: According to (2.3) we have

$$\sigma^2(X) = 1 + m(\rho\tau - m).$$

Differentiating with regard to ρ and using (3.4) we have

$$\frac{d\sigma^2(X)}{d\rho} = m\left(\tau + \rho\frac{d\tau}{d\rho}\right) = m\,\frac{\tau[\lambda(\tau) - \tau] - 1}{\lambda(\tau) - \tau}.$$

For $\tau < 0$ this clearly is <0. For $\tau \geq 0$ inequality (1.7) yields

$$\begin{aligned}\tau[\lambda(\tau) - \tau] - 1 &\leq \tfrac{1}{2}(\tau\sqrt{4 + \tau^2} - 3\tau^2 - 2)\\ &\leq \tfrac{1}{2}[\tau(2 + \tau) - 3\tau^2 - 2] = \tau(1 - \tau) - 1,\end{aligned}$$

and this is < 0 for $\tau \geq 0$. Together with (1.6), this proves that

$$\frac{\tau[\lambda(\tau) - \tau] - 1}{\lambda(\tau) - \tau} < 0$$

for all τ, and hence according to (3.1)

$$\text{sign}\,\frac{d\sigma^2(X)}{d\rho} = -\text{sign}\,m = -\text{sign}\,\rho.$$

It may be conjectured that the sign of $d\sigma^2(X)/d\rho$ is opposite to that of ρ also in the case when $\sigma^2(X)$ is the variance after truncation minimized under condition (3.3). This would lead to a theorem stating that the same choice of $a_1, a_2, \cdots, a_l$ and τ which according to Theorem 2 makes the α-quantile after truncation equal to the given number X_α and minimizes the rejected part of the original population, will also minimize the variance of X after truncation.

REFERENCES

[1] Z. W. Birnbaum, E. Paulson and F. C. Andrews, "On the effect of selection performed on some coordinates of a multi-dimensional population", *Psychometrika*, Vol. 15 (1950).

[2] R. D. Gordon, "Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 364–366.

[3] Z. W. Birnbaum, "An inequality for Mill's ratio", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 245–246.

[4] K. Pearson, *Tables for Statisticians and Biometricians, Part II*, 1st ed., Cambridge Univ. Press, 1931, Tables VIII and IX.

[5] H. Hotelling, "Relations between two sets of variates", *Biometrika*, Vol. 28 (1936), pp. 321–377.