*Jour. Ind. Soc. Ag. Statistics*
*Vol. XL, No. 3 (1988), pp. 185-190*

# A SHORT NOTE ON THE VARIANCE OF THE TRUNCATED DISTRIBUTION WITH AN APPLICATION[1]

RAVINDRA KHATTREE[2] and Y. Q. YIN[3]

(Received : January, 1986)

SUMMARY

In this short note, a property of the variance of the truncated distribution, under very mild conditions has been described. Application of this result is shown in the problems of genetic selection.

*Keywords*: Truncated distribution, Monotonicity of variance, Genetic selection.

## Introduction

The properties of the truncated distributions for various families of probability densities have been well discussed in the literature. Johnson and Kotz [2] present an excellent account of these properties almost in every chapter of their four volume reference work on statistical distributions. In this short note a property of variance of the sub-population

[1]This work was done when authors were graduate students at the Department of Mathematics and Statistics, University of Pittsburgh.

[2]Part of the work of this author was sponsored by the Air Force Office of Scientific Research under Contract F46920-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon. Present address : Mathematical Sciences Division, North Dakota State University, Fargo, ND 58105.

[3]Y. Q. Yin was on leave of absence from the China University of Science and Technology. The work of Yin was supported by a Mellon Fellowship at University of Pittsburgh. Present address : Department of Mathematics, University of Arizona, Tucson, AZ 85721.

obtained by truncating the super-population between two points for a certain family of density functions, bearing some mild conditions has been established.

## 2. Notations and a Lemma

We start with the notations. Let $X$ be a random variable with probability density function $f(\cdot) > 0$. Also, let $F(\cdot)$ be the cumulative distribution of $X$. We further assume that $X$ admits the first and second moments.

Let $a$ and $b$ $(a < b)$ be two points on the real line. The probability density of $X$ in the truncated region $a \leq x < b$ is given by

$$g(x) = \frac{f(x)}{F(b) - F(a)}, \qquad a < x < b \tag{2.1a}$$

and mean and variance are then, readily see◇ to be,

$$m = \frac{1}{F(b) - F(a)} \int_a^b xf(x)\,dx \tag{2.1b}$$

$$v = \frac{1}{F(b) - F(a)} \int_a^b x^2 f(x)\,dx - m^2 \tag{2.1c}$$

respectively.

The following Lemma is a consequence of simple calculus and is true for any function, not necessarily a density.

LEMMA 2.1. *Let $f$ be a monotonically decreasing function in an interval $I$. Let $a, b \in I$, $a < b$, be two points such that*

$$\int_a^b f(x)\,dx = \alpha; \qquad \alpha \neq 0 \tag{2.2}$$

*is fixed, then for any odd nonnegative integer $r$,*

$$\frac{1}{\alpha}\int_a^b x^r f(x)\,dx \leqq \frac{a^r + a^{r-1}b + \ldots + ab^{r-1} + b^r}{r+1} \tag{2.3}$$

*Proof.* Since $f$ is monotonically decreasing in $I$, for any $x_1, x_2 \in I$ and for any odd nonnegative integer $r$,

$$(x_1^r - x_2^r)(f(x_1) - f(x_2)) \leqq 0. \tag{2.4}$$

Integrating with respect to $x_1$ and $x_2$ over the rectangle $(a, b)^2 \leqq I^2$, we have the result.

It may be pointed out that if $0 < a < b$, the above result is true for any nonnegative integer. Our interest, however, is in the case $r = 1$, when (2.3) reduces to

$$\frac{1}{\alpha}\int_a^b x f(x)\, dx \leqq \frac{a+b}{2} \tag{2.5}$$

which is, precisely, a bound on the mean of the truncated distribution.

## 3. Variance of the Truncated Distribution

Our result is about the effect of different truncations, but of the same proportion, on the variances of the sub-populations, obtained after truncation. The result shows that if a fixed proportion of the original population is truncated by points $a$ and $b$, belonging to interval $I$, such that $F(b) - F(a) = \alpha$, a constant, then under some mild conditions, the populations become more and more diverse as we move away. We formally state this result in the following theorem :

THEOREM 3.1. *Suppose $X$ has the density function $f(\cdot)$ which is monotonically decreasing in an interval $I$. Let $a, b \in I$, $a < b$; and*

$$\int_a^b f(x)\, dx = \alpha \tag{3.1}$$

*is fixed. Then $V_x(a)$, the variance of $X$ in the truncated sub-population, as a function of $a$ (and hence of $b$ as well) is monotonically increasing in $I$.*

*Proof.* It is enough to show that

$$\frac{\partial V_X(a)}{\partial a} \geqq 0.$$

Note that (3.1) implies

$$\frac{\partial b}{\partial a} = \frac{f(a)}{f(b)}. \tag{3.2}$$

Now from (2.1b) and (2.1c), the variance as a function of $a$ is

$$V_X(a) = \frac{1}{\alpha}\int_a^b x^2 f(x)\,dx - \left(\frac{1}{\alpha}\int_a^b x f(x)\,dx\right)^2 \tag{3.3}$$

Therefore using (3.2) it can be seen that,

$$\frac{\partial V_X(a)}{\partial a} = \frac{2}{a} f(a)(b-a)\left\{\frac{a+b}{2} - \frac{1}{\alpha}\int_a^b x f(x)\,dx\right\}.$$

Note, as $b > a$; quantity outside parenthesis is positive, while that within parenthesis is, using (2.5) nonnegative. Hence the theorem is established.

## 4. Some Applications

When selecting animals or plants for breeding or production, it is often desirable to consider several traits at the same time. Various methods of selection are available. Among these are methods of independent culling (Hazel and Lush, [1]), tandem selection (Hazel and Lush, *loc. cit.*) and the index selection (Smith [3]). Work done in these lines essentially assumes normality of all the variables (or traits) involved. For estimation of various statistical parameters and evaluation of various associated probability integrals, see Tallis [4], [5] and Young and Weller [6].

Usually in the problem of genetic selection, selection is made to maximize the average of the unobserved or unobservable criterion variable, but it is made on the basis of observed values of predictors. If we denote the criterion variable by $y$ and the regression of criterion on all the predictors by $\eta$, then it is well known that the best selection index for $y$ is $\eta$ and thus the best strategy is to select all those for which

$$\eta \geqslant k \tag{4.1}$$

where $k$ is chosen in such a way that proportions of the selected population is $\alpha$, a predicted value between 0 and 1.

There may be, however, situations where a selection region is sought for which the mean of criterion variable is minimum but at the same time, variance of this region is as small as possible. In general, it may not be possible to attain both the goals simultaneously. As an alternative, one could choose to find a region which has variance of $y$ not more than a prespecified limit decided from practical considerations, but for which mean of $y$ is as large as possible.

If we assume that all the predictors and criterion are in the original

population, distributed jointly as multivariate normal with zero mean, then $\eta$ will also be normally distributed with zero mean. Writing $\sigma_y^2$ and $\sigma_\eta^2$ for variances of $y$ and $\eta$ respectively in the original population, and $W$, for a truncated region on $\eta$-axis, we have

$$V(y \mid \eta \in W) = V(E(y \mid \eta) \mid \eta \in W) + E(V(y \mid \eta) \mid \eta \in W)$$

$$\text{or} \quad V(y \mid \eta \in W) = V(\eta \mid \eta \in W) + \sigma_y^2 - \sigma_\eta^2. \tag{4.2}$$

(4.2) shows that $V(y \mid \eta \in W)$ and $V(\eta \mid \eta \in W)$ differ only by a constant for any region $W$ on $\eta$-axis. Hence putting a restriction on variance of $y$ is equivalent to doing that on variance of $\eta$. Now if our policy for selection was as in (4.1) with $\alpha < \frac{1}{2}$, it would lead to a $\alpha$-proportion subpopulation, even though it maximizes the mean of criterion variable, it is also the most diverse for it. If too much variability is to be avoided and if one seeks a region $W$, for which $V(\eta \mid \eta \in W) \leqq e$, a prespecified quantity, then the region $W$, maximizing mean subject to the above constraint, would be :

$$W^* : k_1 \leqq \eta \leqq k_2 \tag{4.3a}$$

so that

$$P(k_1 \leqq \eta \leqq k_2) = \alpha \tag{4.3b}$$

and that the restriction $V(\eta \mid \eta \in W) \leqq e$ holds as equality constraint

$$V_{W^*}(\eta) = e \tag{4.3c}$$

due to monotonicity property of variance as described in Theorem 3.1.

Of course, to control the variability, one has to sacrifice some of the individual units with high values of criterion variable. But the group of individuals obtained in this way will be more homogeneous.

There may be a situation where, for further experiments, the whole population is to be divided into several groups equal in size on the basis of means of the criterion variable. The theorem says that these groups will differ not only in their mean values but also in the amount of variability and one should possibly take this fact into account while planning for further experiments.

ACKNOWLEDGEMENT

The first author wishes to thank Professor Bimal K. Sinha for a valuable session of discussion.

## REFERENCES

[1] Hazel, L. N. and Lush, J. L. (1942) : The efficiency of three methods of selection, *Jour. Hered.* 33 : 393-399,

[2] Johnson, N. L. and Kotz, S. (1969-1972) : *Distributions in Statistics* (4 volumes), Wiley, New York.

[3] Smith, H. F. (1936) : A discriminant function for plant selection, *Ann. Eugenics* 7 : 240-250.

[4] Tallis, G. M. (1961) : The moment generating function of the truncated multivariate normal distribution, *JRSS* B23 : 223-229.

[5] Tallis, G. M. (1965) : Plane truncation in normal populations, *JRSS* B27 : 301-307.

[6] Young, S. S. Y. and Weller, H. (1961) : Selection for two correlated traits by independent culling levels, *Jour. Genetics* 58 : 329-338.