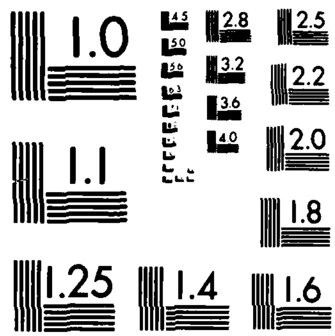
AD-A153 115 AN INEQUALITY AND ITS APPLICATION TO THE TRUNCATED DISTRIBUTIONS(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS R KHATTREE ET AL. FEB 85 1/1

UNCLASSIFIED TR-85-03 AFOSR-TR-85-0347 F49620-85-C-0008 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 85-0347

(2)

AN INEQUALITY AND ITS APPLICATION TO THE
TRUNCATED DISTRIBUTIONS

by

Ravindra Khattree[1] and Y. Q. Yin[2]

Center for Multivariate Analysis

University of Pittsburgh

AN INEQUALITY AND ITS APPLICATION TO THE TRUNCATED DISTRIBUTIONS

by

Ravindra Khattree[1] and Y. Q. Yin[2]

February 1985

Technical Report No. 85-03

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

[1]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.

[2]Y. Q. Yin is on leave of absence from the China University of Science and Technology. The work of Yin is supported by a Mellon Fellowship at the University of Pittsburgh.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) -85-03	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 85-0347
6a. NAME OF PERFORMING ORGANIZATION University of Pittsburgh / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) Center for Multivariate Analysis Fifth Floor, Thackeray Hall Pittsburgh PA 15260	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR / 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0008
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A5; WORK UNIT NO.
11. TITLE (Include Security Classification) AN INEQUALITY AND ITS APPLICATION TO THE TRUNCATED DISTRIBUTIONS	
12. PERSONAL AUTHOR(S) Ravindra Khattree and Y.Q. Yin	
13a. TYPE OF REPORT Technical / 13b. TIME COVERED FROM ____ TO ____	14. DATE OF REPORT (Yr., Mo., Day) FEB 85 / 15. PAGE COUNT 9
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Truncated distribution; bound on mean; monotonicity of variance; genetic selection.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

An inequality is proved and its interpretation is given. Using the inequality, it is shown, under some mild conditions, that for the univariate truncated distributions, the variance of the truncated distribution increases with the value of the truncation point.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL MAJ Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767- 5027 / 22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AN INEQUALITY AND ITS APPLICATION TO THE TRUNCATED DISTRIBUTIONS

by

Ravindra Khattree and Y. Q. Yin

ABSTRACT

An inequality is proved and its interpretation is given. Using the inequality, it is shown, under some mild conditions, that for the univariate truncated distributions, the variance of the truncated distribution increases with the value of the truncation point.

Keywords and Phrases: Truncated distribution, Bound on mean, Monotonicity of Variance, Genetic Selection.

1. INTRODUCTION

The properties of the truncated distributions for the various families of probability densities have been well discussed in the literature. Also, well known are the expressions for mean, variance and higher order moments of truncated distributions, corresponding to certain families. Johnson and Kotz [1] present an excellent account of these properties almost in every chapter of their four-volume reference work on statistical distributions. In this report, ~~we first~~ derive a probability inequality, and then using this inequality, obtain a property of the variance of the subpopulation, obtained by truncating the superpopulation between two points for a certain family of density function bearing some mild conditions.

2. AN INEQUALITY

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

Let $0 \leq a < b$ be any two points. The probability density of X in the truncated region $a \leq x \leq b$ would be given by

$$g(x) = \frac{f(x)}{F(b) - F(a)} \; ; \; 0 \leq a \leq x \leq b \tag{2.1a}$$

and therefore, the mean and variance are readily seen to be

$$m = \int_a^b x \, g(x)dx \tag{2.1b}$$

$$v = \int_a^b x^2 g(x)dx - m^2 \tag{2.1c}$$

Before we prove the main inequality, we will state and prove the following lemma:

<u>Lemma 1</u>. Let $f(x) > 0$ be a continuous integrable density function. Also, let $f(x)$ be monotinically decreasing function of x for $x \geq 0$.

Then,
$$\int_{-c}^{c} y \, f(y+c+a) \leq 0 \text{ for all } c \geq 0, \; a \geq 0. \tag{2.2}$$

<u>Proof</u>. Consider,

$$\int_{-c}^{c} y \, f(y+c+a)dy$$

$$= \int_{-c}^{0} y \, f(y+c+a)dy + \int_0^c y \, f(y+c+a)dy$$

$$= -\int_0^c y \, f(-y+c+a)dy + \int_0^c y \, f(y+c+a)dy$$

$$= \int_0^c y\{f(y+c+a) - f(-y+c+a)\}dy$$

$$\leq 0, \text{ as } f(y+c+a) \leq f(-y+c+a) \; \forall \; a \geq 0, \text{ and } \forall \; 0 \leq y \leq c.$$

We are now in a position where we can prove our inequality which we state in the following lemma:

Lemma 2. Let $0 \leq a < b$, such that $F(b) - F(a) = \alpha$ is fixed, and let $f(.)$ be as defined as in Lemma 1. Then

$$\frac{a+b}{2} \geq \frac{1}{\alpha} \int_a^b x\, f(x)dx \tag{2.3}$$

Proof. We define $y = x - \frac{a+b}{2}$. Then, the right hand side can be written as

$$\frac{1}{\alpha} \int_{-(\frac{b-a}{2})}^{(\frac{b-a}{2})} (y + \frac{a+b}{2})\, f(y + \frac{a+b}{2})dy$$

$$= \frac{1}{\alpha} \int_{-(\frac{b-a}{2})}^{(\frac{b-a}{2})} y\, f(y + \frac{a+b}{2})dy + \frac{a+b}{2\alpha} \int_{-(\frac{b-a}{2})}^{(\frac{b-a}{2})} f(y + \frac{a+b}{2})\, dy.$$

The first integral in the above expression is nonpositive using Lemma 1 with $c = \frac{b-a}{2}$, while the second integral is easily seen to be equal to α (by writing it again in terms of original variable x.) Hence, (2.3) is established.

Remarks. 1. We will first interpret the inequality (2.3). We notice that the right hand side of (2.3) is mean of the truncated random variable X, $0 \leq a \leq x \leq b$. (See (2.1b)). Hence the inequality states that the mean of the truncated distribution is never more than average of the truncation points, under the assumptions already stated.

2. In case X was originally distributed as standard normal, then (2.3) reduces to another interesting inequality

$$\frac{a+b}{2} \geq \frac{\phi(a) - \phi(b)}{\Phi(b) - \Phi(a)}\ ; \quad 0 \leq a < b \tag{2.4}$$

where $\phi(.)$ and $\Phi(.)$ are respectively ordinate and c.d.f. of standard normal distribution.

(2.4) has another interesting interpretation: if we consider ϕ as a function of Φ, then by mean value theorem, there exists a γ; $\Phi(a) \leq \gamma \leq \Phi(b)$ such that

$$\frac{\phi(a) - \phi(b)}{\Phi(b) - \Phi(a)} = \left.\frac{\partial\phi}{\partial\Phi}\right|_{\Phi(.) = \gamma} = \Phi^{-1}(\gamma) = d, \text{ say.}$$

(2.4) states that such a d, corresponding to γ of mean value theorem, will always be less than or equal to midpoint of a and b.

A different proof of (2.4) has been suggested by Dr. Nitish Mukhopadhyaya of Oklahoma State University in a personal communication.

3. In case f(.) was monotinically increasing, the direction of inequalities in (2.2), (2.3) and (2.4) will be reversed. Similar proof will go through with trivial changes.

3. THE VARIANCE OF THE TRUNCATED DISTRIBUTIONS

Our next result is about the effect of different truncations, but of the same proportion, on the variances of the subpopulation obtained after truncation. The result shows that if a fixed proportion of the original population is truncated by points a and b, $0 \leq a < b$, such that $F(b) - F(a) = \alpha$, a constant, then the truncated subpopulation becomes more and more diverse as we move away from the origin, under some mild conditions. We formally state this result in the following theorem:

Theorem. Let X, f(.) F(.), a, b and α be as in Lemma 2, then v, the variances of X in the truncated population, as a function of a (and hence of b as well) is a monotonically increasing function for $a \geq 0$.

Proof. To prove the theorem, it would be enough to show that the derivative of the variance of the truncated population with respect to a is nonnegative.

Note that

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

which implies that

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

now using (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), we have

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

$$- \frac{2}{\alpha^2} \left(\int_a^b x\ f(x)dx\right)\left(bf(b) \frac{f(a)}{f(b)} - af(a)\right)$$

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

Note as $b > a$; quantity outside parentheses is positive, while that within parentheses is, using Lemma 2, nonnegative. Hence,

$$\frac{\partial v_x(a)}{\partial a} \geq 0,$$

which proves our theorem.

Remarks. 1. Theorem can easily be stated for monotonically increasing f(.) with trivial changes.

2. As a corollary, it can be seen that for any probability density symmetric about zero; variance of any α-truncation is an increasing function of $|a|$.

4. SOME APPLICATIONS

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

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

where k is chosen in such a way that proportion of the selected population is α, a predecided value between 0 and 1.

If we assume that all the predictors and criterian are in the original population, distributed jointly as multivariate normal with zero mean, then η will also be normally distributed with zero mean. Writing σ_y^2 and σ_η^2 for variances of y and η respectively in the original population, and W for a truncated region on η axis, we have

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

or

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

(4.2) shows that $V(y|\eta \in W)$ and $V(\eta|\eta \in W)$ differ only by a constant for any region W on η-axis. Now if our policy for selection was as in (4.1), it would lead to a α-proportion subpopulation, even though it maximizes the mean of criterian 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 $V(\eta|\eta \in W) \leq e$, a prespecified quantity, then the region W, maximizing mean subject to the above constraint, would be:

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

so that

$$V_{\omega^*}(\eta) = e \tag{4.3b}$$

and that

$$P(k_1 \leq \eta \leq k_2) = \alpha. \tag{4.3c}$$

Of course, to control the variability, one has to sacrifice some of the individual units with high values of criterian variable.

2. 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 criterian 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] JOHNSON, N. L. and KOTZ, S. (1969-1972). Distributions in Statistics (4 Volumes). Wiley, New York.

END

FILMED

5-85

DTIC