THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ON A CHARACTERIZATION OF THE NORMAL DISTRIBUTION AND ITS APPLICATION TO A GOODNESS-OF-FIT PROBLEM

by

V. Seshadri

Technical Report No. 70
Department of Statistics THEMIS Contract

June 2, 1970

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

ON A CHARACTERIZATION OF THE NORMAL DISTRIBUTION AND ITS APPLICATION TO A GOODNESS-OF-FIT PROBLEM

by

V. Seshadri*
McGill University
Montreal 2, Canada

1. Introduction

In a paper to appear in Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiele, [2], Csörgö and Seshadri have characterized the exponential, gamma and normal laws as also the Poisson process via mappings, onto the unit interval, which are independent of the usually unknown parameters of these families of distribution functions. This paper concerns the characterization of the normal distribution via a more general mapping than mentioned in [2].

Consider (χ, α) any measurable space and denote by $\mathbf{P}$, the set of probability distributions defined on α. Let (y, B) be another measurable space and let $Y = T(X)$, $X \in \chi$ be a measurable mapping of (χ, α) onto (y, B). With this mapping every distribution $P \in \mathbf{P}$ induces on B a corresponding distribution which will be denoted by Q_P^Y. We are concerned with mappings Y which satisfy the following two properties:

P.1. Q_P^Y is the same for all $P \in \mathbf{P}$; in such a case we write $Q_{\mathbf{P}}^Y$

P.2. If for some P' on α, $Q_{P'}^Y = Q_{\mathbf{P}}^Y$ then $P' \in \mathbf{P}$.

For the purposes of the discussion in this paper, (χ, α) will be an

*This research was sponsored by the Office of Naval Research, Contract No. N00014-68-A-0515, Department of Statistics, Southern Methodist University.

n-dimensional Euclidean space of points $X = (X_1, X_2, \cdots, X_n)$ with the σ-algebra of Borel sets and the distributions belonging to $\mathbf{P}$ have a product probability density $f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta)$ where f is a one dimensional density function and θ is a (vector) parameter taking vlaues in an appropriate parameter space. Given some density function of the form described above, with $f(\cdot,\cdot)$ the normal density, we will be interested in mappings Y satisfying P.1. and P.2. in such a way that the induced distribution $Q_{\mathbf{P}}^{Y}$ will be that of $k(k < n)$ ordered random variables from K independent uniformly distributed random variables on [0, 1].

From the point of view of inference such a mapping can be used to replace composite statistical hypotheses by equivalent simple ones in the spirit of Yu. V. Prohorov's paper [6]. The interested reader is referred to [2] and also a forthcoming paper, [3], in the Review of the International Statistical Institute.

2. The Characterization Theorem

Let us first consider a finite sequence of matrices $\{A_i\}$ $(i = 1, 2, \cdots, n-1)$ of size $(n \times n)$ such that the rank of A_i indicated by $r(A_i) = i$ $(i = 1, 2, \cdots, n-1)$. Furthermore let this sequence be characterized by the following property:

P.3. $A_i A_j = A_i \qquad i \leq j$.

It is an easy exercise to show that the matrices $\{A_i\}$ form a Boolean algebra with two operations viz.

a) the product (ordinary) $A_i \times A_j = A_i \quad i \leq j$
$\qquad\qquad = A_j \quad j \leq i$

b) the operation $*$ such that $A_i * A_j = A_i + A_j - A_i \times A_j$ for every (i,j).

The important property from our point of view is that each A_i is idempotent. We shall now consider the following matrices $B_j (j = 1, 2, \cdots, k+1)$ where

$$\left.\begin{aligned} B_1 &= A_2 \\ B_2 &= A_4 - A_2 \\ &\vdots \\ B_j &= A_{2j} - A_{2(j-1)} \\ &\vdots \\ B_{k+1} &= A_{2(k+1)} - A_{2k} \end{aligned}\right\} \text{Let } A_1 = \underset{(1)}{\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \cdots & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & 0 & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdot & \cdots & 0 \end{pmatrix}} . \qquad (2)$$

We have assumed in the above that $n = 2k + 3$, $k \geq 2$. We note at once from (1)

a) $B_r B_s = \phi \quad (r \neq s)$
b) $B_r B_r = B_r$ all r.
c) $r(B_r) = 2$.

Let $X' = (X_1, X_2, \cdots, X_n)$ be a random vector whose components X_i are independent and identically distributed with common mean μ and common variance σ^2, $-\infty < \mu < \infty$, $\sigma > 0$. Consider the quadratic form

$$X'A_1X/\sigma^2 = (X_1 - X_2)^2/2\sigma^2 .$$

Since $(X_1 - X_2)$ has the same distribution as $(X_2 - X_1)$, $X'A_1X/\sigma^2$ has a chi-square distribution with one degree of freedom if and only if $(X_1 - X_2)/\sqrt{2}\sigma$ has a normal distribution with zero mean and unit variance. By Cramer's Theorem [1] this implies that X_1 and X_2 are $N(\mu, \sigma^2)$. Since

X_i are i.i.d., it then follows that $X \sim N(\mu, \sigma^2 I)$ if and only if $X'A_1X/\sigma^2 \sim \chi_1^2$. Since each B_i $(i = 1, 2, \cdots, k+1)$ is idempotent of rank 2 it then follows from the given structure of the matrices A_i that each quadratic form $X'B_iX/\sigma^2$ has χ_2^2 or an exponential distribution. Furthermore, since $B_iB_j = \phi$ $(i \neq j)$ it implies that $X'B_iX/\sigma^2$ is independent of $X'B_jX/\sigma^2$. Consider then the following quadratic forms

$$X'B_1X, \quad X'(B_1+B_2)X, \quad X'(B_1+B_2+B_3)X, \cdots, \quad X'\left(\sum_{i=1}^{k+1} B_i\right)X \quad .$$

From (1) these are simply

$$X'A_2X, \quad X'A_4X, \quad X'A_6X, \cdots, \quad X'A_{2(k+1)}X = X'A_{n-1}X \quad .$$

Therefore we see that

$$\eta_1 = \frac{X'A_2X}{X'A_{n-1}X}, \quad \eta_2 = \frac{X'A_4X}{X'A_{n-1}X}, \cdots, \quad \eta_k = \frac{X'A_{2k}X}{X'A_{n-1}X}$$

are order statistics fo k independent uniformly distributed random variables in [0, 1]. This statement follows from <u>Lemma 1.</u> given below.

<u>Lemma 1.</u> Let $Y_1, Y_2, \cdots, Y_{k+1}$ be (k+1) independent identically distributed positive random variables with a continuous density function and mean $B > 0$. Let $S = \sum_{i=1}^{k+1} Y_i$. Define $Z_r = \sum_{i=1}^{r} Y_i/S$, $r = 1, 2, \cdots, k$. Then the Z_r act like k order statistics of k independent random variables from u(0, 1) [uniform] if and only if the Y_i are exponentially distributed viz. $f(y_i) = \frac{1}{B} e^{-y_i/B}$.

To apply this lemma to our problem we note that $Y_i = X'B_iX/\sigma^2$ and

η_r is the Z_r of the lemma. We have thus proved the following theorem.

<u>Theorem</u>: Let $X' = (X_1 \cdots X_n)$ $n \geq 2k + 3$, $k \geq 2$ be a random vector of independent identically distributed components with mean μ and variance σ^2, $-\infty < \mu < \infty$, $\sigma > 0$. Let $\{A_i\}$ be a sequence of $(n \times n)$ matrices $(i = 1, 2, \cdots, n-1)$ with the following structure

a) $\text{rank } A_i = i$

b) $$A_1 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \cdots & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & 0 & \cdot & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdot & \cdots & 0 \end{pmatrix}$$

c) $A_i A_j = A_i \quad i \leq j$

Then $y_r = \dfrac{X' A_{2r} X}{X' A_{n-1} X}$ $\left(r = 1, 2, \cdots, \frac{n-3}{2}\right)$, act like $\frac{n-3}{2}$ order statistics of $\frac{n-3}{2}$ independent $u(0, 1)$ random variables if and only if X_i has a normal distribution with some mean μ and variance σ^2.

<u>Application</u>. Most goodness-of-fit problems arising in actual practice involve nuisance parameters. The null hypotheses one wishes to test are composite. The unknown parameters are usually estimated by maximum likehood principle and the asymptotic distribution of the thus modified test-statistic is then used to construct probabilities of rejection when the null hypothesis is true. The χ^2 test possesses an element of arbitrariness in the choice of group boundaries and does not yield an exact test.

The Kolmogorov-Smirnov and similar tests are free from these objections and have good asymptotic power properties against alternatives specified in terms of distance between distribution functions Darling [4]. But a serious limitation to the practical usefulness of such tests has been the lack of a simple method of allowing for the presence of nuisance parameters. Several novel methods have been proposed in the recent past by Durbin [5] and Sarkadi [7] to overcome this defect. Csörgö and Seshadri have proposed [2] the method of reducing composite hypotheses to equivalent simple ones in the sense of Prohorov [6] and the theorem of Section 2 does exactly this.

Power considerations are being studied at the moment and a detailed discussion of this aspect will be reported elsewhere.

REFERENCES

[1] Cramér, H. (1963). "Uber eine Eigenschaft der normalen Verteilungsfunktion." Math. Z., 41, 405-414.

[2] Csörgö, M. and Seshadri, V. "Characterizing the Gaussian and the exponential laws via mappings onto the unit interval." To appear in Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiele.

[3] Csörgö, M. and Seshadri, V. "On the problem of replacing composite statistical hypotheses by equivalent simple ones." To appear in the review of the International Statistical Institute.

[4] Darling, D. A. (1957). "The Kolmogorov-Smirnov, Cramer-Von Mises tests," Ann. Math. Stat., 28, 823-38.

[5] Durbin, J. (1961). "Some methods of constructing exact tests," Biometrika, 48, 41-55.

[6] Prohorov, Yu. V. (1966). "Some characterization problems in statistics," Proc. Fifth Berkely Symposium Math. Stat. Prob., 1, 341-49. University of California Press.

[7] Sarkadi, K. (1960). "On testing for normality," Publ. Math. Inst. Hungar. Acad. Sci., 5, 269-75.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) SOUTHERN METHODIST UNIVERSITY	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED
	2b. GROUP UNCLASSIFIED

3. REPORT TITLE

On a Characterization of the Normal Distribution and Its Application to a Goodness-of-Fit Problem

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

V. Seshadri

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 2, 1970	6	6

8a. CONTRACT OR GRANT NO. N00014-68-A-0515	9a. ORIGINATOR'S REPORT NUMBER(S) 70
b. PROJECT NO. NR 042-260	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research

13. ABSTRACT

Goodness-of-fit problems arising in practice often involve nuisance parameters and hence make the hypotheses composite. Exact tests for such problems are hard to come by. A method is proposed in this paper whereby an exact test for normality when both the location and scale parameter are unknown, is possible by eliminating the parameters in a very natural way. A theorem involving a characterization of the normal distribution is stated and proved and its application to a goodness-of-fit study is mentioned.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408