The Annals of Statistics
1979, Vol. 7, No. 5, 1052–1057

CONDITIONAL PROBABILITY INTEGRAL TRANSFORMATIONS AND GOODNESS-OF-FIT TESTS FOR MULTIVARIATE NORMAL DISTRIBUTIONS[1]

BY S. RINCON-GALLARDO, C. P. QUESENBERRY AND FEDERICO J. O'REILLY

North Carolina State University
and IIMAS, Universidad Nacional Autónoma de México

Let $X_1, \cdots, X_n$ be a random sample from a full-rank multivariate normal distribution $N(\mu, \Sigma)$. The two cases (i) μ unknown and $\Sigma = \sigma^2\Sigma_0$, Σ_0 known, and (ii) μ and Σ completely unknown are considered here. Transformations are given that transform the observation vectors to a (smaller) set of i.i.d. uniform rv's. These transformations can be used to construct goodness-of-fit tests for these multivariate normal distributions. Two examples are given to illustrate the application of these tests to numerical problems.

1. Introduction and summary. There is a large literature that considers the multivariate normal distribution. There is, however, no test for multivariate normality that is very widely used, to our knowledge. Most tests for multivariate normality depend upon asymptotic distribution theory. Such tests include χ^2 tests with estimated parameters, and tests based on measures of multivariate skewness and kurtosis posed by Mardia (1970). Wagle (1968) gives transformations on which a test could be based. Recently, Moore (1976) has commented upon the need for tests of multivariate normality. In this work we give transformations which can be used to construct exact size goodness-of-fit tests for multivariate normality, and illustrate their use with two numerical examples.

O'Reilly and Quesenberry (1973), O-Q introduced the conditional probability integral transformations, CPIT's. Transformations were given in that paper for a multivariate normal parent $N(\mu, \Sigma)$ for the case when μ is unknown and $\Sigma = \Sigma_0$ is known. Here we give transformations for the two cases: (i) μ unknown, and $\Sigma = \sigma^2\Sigma_0$ with Σ_0 known, and (ii) μ and Σ unknown. In both of these cases the components of the observation vectors are transformed using certain Student-t distribution functions.

2. Notation and preliminaries. We develop the main transformation result in this and the next section assuming case (ii). The corresponding result for case (i) will be summarized in Theorem 3.2. Let $\mathcal{P}$ denote the class of k-variate full-rank normal distributions with mean vector μ and variance-covariance matrix Σ. For $X_1, \cdots, X_n$ i.i.d. (column) vector rv's from $P \in \mathcal{P}$, with corresponding probability density function f and distribution function F, both defined on R_k, a complete

Received December 1977; revised August 1978.
[1]Supported by IIMAS, UNAM, CONACYT and National Science Foundation Grant MCS76-82652.
AMS 1970 *subject classifications.* Primary 62H15.
Key words and phrases. Goodness-of-fit, multivariate normal, conditional probability integral transformations.

sufficient statistic for $\mathcal{P}$ is $T_n = (\bar{X}_n, S_n)$, where $\bar{X}_n = (1/n)\Sigma_{i=1}^n X_i$ and $S_n = \Sigma_{i=1}^n X_i X_i' - n\bar{X}_n\bar{X}_n'$. It is readily verified that T_n is doubly transitive (cf. O-Q), i.e., $\sigma(T_n, X_n) = \sigma(T_{n-1}, X_n)$, where $\sigma(W)$ denotes the σ-algebra induced by a statistic W.

Consider the conditional distribution function $F(X_{n-\alpha+1}, \cdots, X_n | T_n)$, of the last α observations given T_n. The largest value of α for which this is the distribution function of an absolutely continuous distribution is called the *absolute continuity rank* of $\mathcal{P}$, which is $\alpha = n - k - 1$. In the following we find formulae to transform the $k(n - k - 1)$ rv's of this conditional distribution to i.i.d. $U(0, 1)$ rv's.

Denote by $\tilde{F}$ the conditional distribution function of a single observation given T_n. For $n > k + 1$, $\tilde{F}$ is absolutely continuous and possesses a density function $\tilde{f}$ which is the minimum variance unbiased, MVU, estimator of the parent density function. These functions were obtained by Ghurye and Olkin (1969), page 1265, cases 3.2 and 3.4, for the cases (i) and (ii) above. The next lemma gives the density $\tilde{f}$ in a form that will be convenient in this work. The indicator function of the set satisfying condition $[\cdot]$ is denoted by $I[\cdot]$.

LEMMA 2.1. *If $X_1, \cdots, X_n$ are* i.i.d. rv's *with a common multivariate normal distribution $P \in \mathcal{P}$, the* MVU *estimator $\tilde{f}$ of the corresponding normal probability density function is*

$$\tilde{f}(x) = \frac{[n/(n-1)]^{(n-3)/2}\Gamma[(n-1)/2]}{\pi^{k/2}\Gamma[(n-k-1)/2]}|S_n|^{-\frac{1}{2}} \cdot \left\{[(n-1)/n] - (x - \bar{X}_n)'S_n^{-1}(x - \bar{X}_n)\right\}^{\frac{1}{2}(n-k-3)} \cdot I\left[(x - \bar{X}_n)'S_n^{-1}(x - \bar{X}_n) < (n-1)/n\right], \qquad n > k + 1. \tag{2.1}$$

PROOF. This result is immediate from Ghurye and Olkin (1969) and the two facts:

(a) For $B(k \times k)$ nonsingular and $x(k \times 1)$, $|B - xx'| = |B|(1 - x'B^{-1}x)$.

(b) If B is p.d. then $B - xx'$ is p.d. iff $x'B^{-1}x < 1$.

LEMMA 2.2. *Suppose Y is a* rv *which has for fixed T_n the conditional density function $\tilde{f}$ of* (2.1). *Then for*

$$Z_n = A_n(Y - \bar{X}_n)/\left\{[(n-1)/n] - (Y - \bar{X}_n)'S_n^{-1}(Y - \bar{X}_n)\right\}^{\frac{1}{2}},$$

where $A_n'A_n = S_n^{-1}$, the conditional density function of Z_n given T_n is

$$\tilde{g}(z) = \Gamma\left[\tfrac{1}{2}(n-1)\right]\left\{\pi^{\frac{1}{2}k}\Gamma\left[\tfrac{1}{2}(n-k-1)\right]\right\}^{-1}\left[1 + z'z\right]^{-\frac{1}{2}(n-1)}, \qquad n > k + 1. \tag{2.2}$$

PROOF. If $z = A_n(y - \bar{X}_n)/\{[(n-1)/n] - (y - \bar{X}_n)'S_n^{-1}(y - \bar{X}_n)\}^{\frac{1}{2}}$, then $y = A_n^{-1}z[(n-1)/n]^{\frac{1}{2}}(1 + z'z)^{-\frac{1}{2}} + \bar{X}_n$. The Jacobian is

$$\left\{[(n-1)/n]/(1 + z'z)\right\}^{+\frac{k}{2}}|A_n|^{-1}|I - zz'/(1 + z'z)|,$$

and using the relations

$$|I - zz'/(1 + z'z)| = (1 + z'z)^{-1} \text{ and}$$
$$(y - \bar{X}_n)'S_n^{-1}(y - \bar{X}_n) = [(n-1)/n]z'z/(1 + z'z),$$

the result follows from Lemma 2.1.

The density function $\tilde{g}$ of (2.2) has the form of a generalized multivariate t distribution. Dickey (1967), Theorems 3.2 and 3.3, gives conditional and marginal distributions for generalized multivariate t distributions from which the conditional and marginal distributions of $\tilde{g}$ of (2.2) can be obtained. Let G_ν denote the distribution function of a univariate Student-t distribution with ν degrees of freedom. Then the following can be obtained from results given by Dickey.

LEMMA 2.3. *Let* $Z' = (Z_1, \cdots, Z_k)$ *denote a vector* rv *with (conditional) probability density function* $\tilde{g}(z)$ *of equation* (2.2). *Then*

(2.3) $$\tilde{P}(Z_i \leqslant z_i | Z_1 = z_1, \cdots, Z_{i-1} = z_{i-1}) = G_{n-k+i-2}\left\{z_i\left[(n-k+i-2)/\left(1 + \Sigma_{j=1}^{i-1} z_j^2\right)\right]^{\frac{1}{2}}\right\},$$

for $i = 1, \cdots, k$.

3. The transformations. Consider again the original sample $X_1, \cdots, X_n$, and put

$$Z_j = A_j(X_j - \bar{X}_j)/\left\{[(j-1)/j] - (X_j - \bar{X}_j)'S_j^{-1}(X_j - \bar{X}_j)\right\}^{\frac{1}{2}},$$

and denote $Z_j' = (Z_{1,j}, \cdots, Z_{k,j})$ for $j = k+2, \cdots, n$. Then the next theorem follows from Lemma 2.3 and a slight extension of Theorem 5.1 of O-Q.

THEOREM 3.1. *The* $k(n-k-1)$ *random variables given by*

(3.1) $$U_{i,j} = G_{j-k+i-2}\left\{Z_{i,j}\left[(j-k+i-2)/\left(1 + Z_{1,j}^2 + \cdots + Z_{i-1,j}^2\right)\right]^{\frac{1}{2}}\right\},$$

for $j = k+2, \cdots, n$ *and* $i = 1, \cdots, k$; *are* i.i.d. $U(0, 1)$ rv's.

We now summarize the results for case (i) when μ is unknown and $\Sigma = \sigma^2\Sigma_0$ for Σ_0 known. For $\bar{X}_n$ and S_n defined above put here

$$A'A = \Sigma_0^{-1}, \qquad s_n = \operatorname{tr} \Sigma_0^{-1} S_n,$$

and

$$Z_j = A(X_j - \bar{X}_j)/\left\{[(j-1)s_j/j] - (X_j - \bar{X}_j)'\Sigma_0^{-1}(X_j - \bar{X}_j)\right\}^{\frac{1}{2}},$$

and denote $Z_j' = (Z_{1,j}, \cdots, Z_{k,j})$ for $j = 3, \cdots, n$.

THEOREM 3.2. *For* $X_1, \cdots, X_n$ i.i.d. *from* $N(\mu, \sigma^2\Sigma_0)$, Σ_0 *known, the* $(n-2)k$ *random variables given by*

(3.2)

$$U_{i,j} = G_{[(j-2)k+i-1]}\left\{Z_{i,j}\left[((j-2)k+i-1)/\left(1 + Z_{1,j}^2 + \cdots + Z_{i-1,j}^2\right)\right]^{\frac{1}{2}}\right\},$$

for $j = 3, \cdots, n$ *and* $i = 1, \cdots, k$, *are* i.i.d. $U(0, 1)$ rv's.

4. Applications to goodness-of-fit tests: examples. After the multivariate sample $X_1, \cdots, X_n$ has been transformed by using either (3.1) or (3.2), then a size α goodness-of-fit test for the corresponding composite multivariate normal null hypothesis class (case (i) or case (ii)) can be made by testing the surrogate simple null hypothesis that the transformed values are i.i.d. $U(0, 1)$. Quesenberry and Miller (1977) and Miller and Quesenberry (1977) have studied power properties of omnibus tests for uniformity and recommend either the Watson U^2 test (Watson (1962)) or the Neyman smooth test (Neyman (1937)) for testing simple uniformity. We shall here use a modified form, U^2_{MOD}, of the Watson U^2 statistic proposed by Stephens (1970), that has the advantage of having upper 1, 5 and 10 percentage points that are approximately .267, .187 and .152 for all $n \geqslant 10$, under the null hypothesis. We shall also compute the Neyman smooth statistic p_4^2 which has an approximate $\chi^2(4)$ distribution, under the null hypothesis for large n.

This approach will now be applied to two numerical examples. In each of these examples the goodness-of-fit null hypothesis class is the case (ii) multivariate normal class. All computations were performed using a program written by the authors.

EXAMPLE 1. *Fisher's iris data.* We consider first the iris data of Fisher (1936). The data consists of samples on three species of iris (setosa, versicolor, and virginica), consisting of 50 observations on each of four variables (sepal length, sepal width, petal length and petal width). We have transformed each of these samples using the transformations (3.1). This gives $180[= (n - k - 1)k]$ transformed values for each sample that are observed values on i.i.d. $U(0, 1)$ rv's, under the multivariate normality null hypothesis. The values of the test statistics U^2_{MOD} and p_4^2, and of the observed significance level of p_4^2, $P(\chi^2(4) > p_4^2)$, are summarized in Table 1. The upper 10 per cent significance level for U^2_{MOD} is 0.152 (cf. Stephens (1970)), and neither test statistic is significant at the 10 per cent level for any of the samples. The small values of the test statistics give no reason to question that multivariate normal distributions fit these data well. Further analysis of the transformed u-values can be performed by graphing the ordered u's against the expected values of uniform order statistics as in Quesenberry, et al (1976). We have made such graphs and they also indicate that normal distributions fit these samples well. Finally, it should be observed that if it is assumed under the null hypothesis that all three samples are from parent distributions of the same functional form, but possibly with different parameter values, then we could pool all of the $(3 \times 180 = 540)$ u-values and test that this common functional form is normal.

TABLE 1

Species	U^2_{MOD}	p_4^2	$P(\chi^2(4) > p_4^2)$
Setosa	.064	4.508	.24
Versicolor	.077	3.325	.50
Virginica	.054	3.622	.46

EXAMPLE 2. *Norton's rate of discount and ratio of reserves to deposits data.* Yule and Kendall (1946), page 201, give bivariate data for (1) call discount rates and (2) percentages of reserves on deposit in New York associated banks for 780 banks. They also plot a bivariate histogram which makes it quite clear that these data are skewed, and not well fitted by a bivariate normal distribution. Mardia (1970) has applied asymptotic tests for skewness and kurtosis to this data, and rejects normality with both tests.

We have drawn subsamples of sizes 50, 100 and 200 from these 780 observations, and then performed the transformations of (3.1), and computed p_4^2 and U_{MOD}^2 on each of these four samples. The results obtained are given in Table 2. Neither test statistic is significant for $n = 50$ or 100. For $n = 200$ p_4^2 is highly significant and the observed value of U_{MOD}^2 is just less than the 1 per cent point of 0.267. Both statistics are highly significant for the entire sample of 780 values.

TABLE 2

Sample Size, n	U_{MOD}^2	p_4^2	$P(\chi^2(4) > p_4^2)$
50	0.089	1.467	0.832
100	0.149	5.097	0.277
200	0.260	22.319	1.73 $E(-4)$
780	2.045	78.406	2.59 $E(-12)$

5. Discussion. The transformations of (3.1) and (3.2) are readily programmed using computer languages with matrix algebra packages. From our (limited) experience with the above and some other examples, the testing procedure above of computing the transformed u-values and U_{MOD}^2 and p_4^2 appears to be a practical procedure, and to be sensitive to at least some departures from normality when the sample size n is in the range of 100 to 200.

A warning is also in order. It should be carefully observed that the values of the u's depend upon the order of the entries in the sample, and different values will, in general, be obtained from different orderings. Care must be exercised to assure that $X_1, \cdots, X_n$ are i.i.d. $N(\mu, \Sigma)$ rv's. In particular, the X_i's must *not* be ordered by the values of one or more of their components.

REFERENCES

DICKEY, J. M. (1967). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist.* **38** 511–518.

FISHER, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, VII, 179–188. Also paper 32 in *Contributions to Mathematical Statistics, R. A Fisher*. Wiley, New York.

GHURYE, S. G. and OLKIN, I. (1969). Unbiased estimation of some multivariate probability densities and related functions. *Ann. Math. Statist.* **40** 1261–1271.

MARDIA, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57(3)** 519–530.

MILLER, F. L., JR. and QUESENBERRY, C. P. (1979). Power studies of tests for uniformity, II. *Commun. Statist. B* **8** (3).

MOORE, D. S. (1976). Recent developments in chi-square tests for goodness-of-fit. Mimeograph Series No. 459, Depart. Statist., Purdue Univ.

NEYMAN, JERZY (1937). "Smooth" test for goodness-of-fit. *Skand. Aktuarietidskr.* **20** 149–199.

O'REILLY, F. J. and QUESENBERRY, C. P. (1973). The conditional probability integral transformation and applications to obtain composite chi-square goodness-of-fit tests. *Ann. Statist.* **1** 74–83.

QUESENBERRY, C. P. and MILLER, F. L., JR. (1977). Power studies of some tests for uniformity. *J. Statist. Comput. Simul.* **5** 169–191.

QUESENBERRY, C. P., WHITAKER, T. B. and DICKENS, J. W. (1976). On testing normality using several samples: an analysis of peanut aflatoxin data. *Biometrics* **32(4)** 753–759.

STEPHENS, M. A. (1970). Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. *J. Roy. Statist. Soc. Ser. B* **32** 115–122.

WAGLE, B. (1968). Multivariate beta distribution and a test for multivariate normality. *J. Roy. Statist. Soc. Ser. B* **30** 511–516.

WATSON, G. S. (1962). Goodness-of-fit tests on a circle. II. Biometrika **49** 57–63.

YULE, G. U. and KENDALL, M. G. (1949). *An Introduction to the Theory of Statistics, 13th ed.* London, Griffin.

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
P. O. BOX 5457
RALEIGH, NORTH CAROLINA 27650

IIMAS
UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO
APDC POSTAL 20-726
MÉXICO 20, DF MEXICO