ON A STOCHASTIC INEQUALITY FOR THE WILKS STATISTIC

A. K. GUPTA

(Received May 22, 1972; revised June 18, 1973)

Summary

The general nonnull distribution of Wilks statistic, the likelihood ratio statistic in MANOVA, can be expressed as a product of conditional beta variables [3]. Making use of this result, in the present paper, an upper bound for the nonnull distribution of Wilks statistic is obtained, which provides a conservative evaluation of the power of the likelihood ratio test for the cases when the alternative hypothesis is of rank 1, 2 or 3. For $p=2$, where p is the number of variables, and large f_2, the degrees of freedom, it has been shown that the results of this paper give a much better approximation to the power of Wilks statistic than Mikhail's approximation [10]. A few percentage points have also been computed for $p=3$ and selected values of the degrees of freedom and the noncentrality parameters, which in the linear case have been compared with the exact values obtained by the author [7].

1. Introduction and notations

Let the columns of a $p\times f_2$ matrix $\boldsymbol{X}=(x_{ij})$ and a $p\times f_1$ matrix $\boldsymbol{Y}=(y_{ij})$, $p\leqq f_1$, be distributed independently in p-variate normal distribution with a common positive definite covariance matrix $\boldsymbol{\Sigma}$ and let $\mathrm{E}(\boldsymbol{X})=\boldsymbol{M}$, $\mathrm{E}(\boldsymbol{Y})=\boldsymbol{0}$. The likelihood ratio criterion for testing H_0: $\boldsymbol{M}(p\times f_2)=\boldsymbol{0}$ against $\boldsymbol{M}\neq\boldsymbol{0}$ can be expressed in terms of the following criterion suggested by Wilks [16] and Pearson and Wilks [11],

$$\Lambda=|\boldsymbol{YY}'|/|\boldsymbol{XX}'+\boldsymbol{YY}'| .$$

In the context of multivariate analysis of variance, $\boldsymbol{YY}'$ and $\boldsymbol{XX}'$ are the sums of product matrices for error and hypothesis respectively, and f_1 and f_2 are the corresponding degrees of freedom.

Let $0<\delta_p^2\leqq\cdots\leqq\delta_1^2<\infty$, be the roots of the determinantal equation, $|\boldsymbol{MM}'-\delta^2\boldsymbol{\Sigma}|=0$, and δ_i be the non-negative square root of δ_i^2. The sampling distribution of Λ denoted by $W_p(f_1, f_2; \delta_1^2,\cdots,\delta_r^2)$ where $r=\mathrm{rank}(\boldsymbol{M})$, will be called the Wilks distribution with p dimensions, f_1

and f_2 degrees of freedom and noncentrality parameter vector $(\delta_1^2,\cdots,\delta_r^2)$. In this paper this distribution will be studied for $p=2$ and 3. If $p=1$, the problem is trivial. In the following, we shall assume $p>1$.

2. General nonnull distribution of Λ

To study the general nonnull distribution of Λ, it can be assumed, without loss of generality, that

$$\boldsymbol{\Sigma}=\boldsymbol{I}_p\,, \qquad \boldsymbol{M}=\begin{pmatrix} \boldsymbol{D}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$

where $\boldsymbol{D}_r=\mathrm{diag}\,(\delta_1^2,\cdots,\delta_r^2)$ and δ_i^2 $(i=1,\cdots,r)$ are the positive eigenvalues of $\boldsymbol{MM}'$ (or of $\boldsymbol{M}'\boldsymbol{M}$). Then, using the random orthogonal transformation due to Wijsman [15], Asoh and Okamoto [3] have shown that for $r=1$, i.e. in the linear case (Anderson [1], Anderson and Girshick [2]),

$$W_p(f_1, f_2;\delta_1^2)=B(f_1/2, f_2/2;\delta_1^2)\,W_{p-1}(f_1-1, f_2)\,, \tag{2.1}$$

a product of a noncentral beta variate and an independent Wilks variate. Gupta [7] has obtained the distribution of Λ in this case for $p=2$, 3, 4 and 5 and has also given the general form of the distribution for any p. For the case $r=2$, i.e. in the planar case it has been shown [3] that

$$W_p(f_1, f_2;\delta_1^2,\delta_2^2)=B(f_1/2, f_2/2;\delta_1^2)B((f_1/2-1), f_2/2;\Delta_2^2)\cdot W_{p-2}(f_1-2, f_2) \tag{2.2}$$

where

$$\Delta_2^2=\delta_2^2\left(1-\frac{x_{12}^2}{x_{1\cdot}x_{1\cdot}'+y_{1\cdot}y_{1\cdot}'}\right),$$

and $x_{1\cdot}$ and $y_{1\cdot}$ are the ith rows of the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. Since $\Delta_2<\delta_2$ with probability one, it follows [3] that

$$W_p(f_1, f_2;\delta_1^2,\delta_2^2)\succ B(f_1/2, f_2/2;\delta_1^2)B((f_1/2-1), f_2/2;\delta_2^2)\cdot W_{p-2}(f_1-2, f_2) \tag{2.3}$$

where the symbol "$\succ$" denotes the relation "is stochastically larger than" and the three factors in the right-hand side are distributed independently.

In general we get [3] the following stochastic inequality:

$$W_p(f_1, f_2;\delta_1^2,\cdots,\delta_r^2)\succ\prod_{i=1}^{r}B\left(\frac{1}{2}(f_1+1-i),\,\frac{1}{2}f_2;\delta_i^2\right), \tag{2.4}$$

where the r factors on the right-hand side are distributed independently. Hence if we can evaluate the distribution function $F_B(x)$ of the product of noncentral beta variates in the right-hand side of (2.4), then it will give an upper bound of the distribution function $F_W(x)$ of the Wilks statistic.

3. Method of derivation of density of the product of noncentral beta variates

Let us denote the density of a noncentral beta variate occuring in (2.4) by

$$\begin{aligned}(3.1)\qquad \mathcal{L}(X_i)&=B\left[\frac{1}{2}(f_1+1-i),\ \frac{1}{2}f_2;\ \delta_i^2;\ X_i\right]\\ &=K_iX_i^{(f_1+1-i)/2-1}(1-X_i)^{f_2/2-1}\\ &\quad\cdot {}_1F_1\left(\frac{1}{2}(f_1+f_2+1-i),\ \frac{1}{2}f_2,\ \frac{1}{2}\delta_i^2(1-X_i)\right), \qquad 0<X_i<1\end{aligned}$$

where

$$(3.2)\qquad K_i=\left[e^{-\delta_i^2/2}\Big/B\left(\frac{1}{2}(f_1+1-i),\ \frac{1}{2}f_2\right)\right],$$

and ${}_1F_1$ denotes the confluent hypergeometric function, defined below.

$$(3.3)\qquad {}_1F_1(a,b,z)=\sum_{n=0}^{\infty}\frac{(a)_n}{(b)_n}\frac{z^n}{n!}$$

where $(a)_n=a(a+1)\cdots(a+n-1)$.

Now substitute in (3.1) for ${}_1F_1$ from (3.3) and transform $Y_i=-\log X_i$, then the density of Y_i, after binomial expansion, takes the following form.

$$(3.4)\qquad \mathcal{L}(Y_i)=K_i\sum_{l=0}^{\infty}a_{il}\sum_{k=0}^{b+l}(-1)^k\binom{b+l}{k}e^{-(f_1+2k-i+1)Y_i/2}, \qquad Y_i\geqq 0,$$

where $\upsilon=f_1+f_2$, $b=f_2/2-1$ and

$$(3.5)\qquad a_{il}=\frac{((\upsilon+1-i)/2)_l(\delta_i^2/2)^l}{(b+1)_l\, l!}.$$

To find the distribution of $\prod_{i=1}^{p}X_i$, we first find the distribution of $-\log\left(\prod_{i=1}^{p}X_i\right)=\sum_{i=1}^{p}(-\log X_i)=\sum_{i=1}^{p}Y_i$, which is the sum of independently distributed random variables, and then make inverse transformation to get the required distribution. The latter distribution can be derived by evaluating the successive convolutions (see Gupta [6], Pillai and Gupta [12], Schatzoff [14]). Hence for $p=2$ we obtain the pdf of $U=X_1X_2$ as

$$(3.6)\quad 2K_1K_2U^{(f_1-2)/2}\sum_{j,l=0}^{\infty}a_{1j}a_{2l}\sum_{k=0}^{b+j}\sum_{m=0}^{b+l}\frac{(-1)^{k+m}}{2m-2k-1}\binom{b+j}{k}\binom{b+l}{m}(U^k-U^{m-1/2}),$$

where K_1 and K_2 are given by (3.2) for $i=1,2$ and a_{1j}, a_{2l} are given by (3.5) for $i=1$, $l=j$ and $i=2$ respectively.

Similarly the pdf of $U=X_1X_2X_3$ is given by

$$(3.7)\quad 2\left(\prod_{i=1}^{3}K_i\right)U^{(f_1-2)/2}\sum_{j,l,n=0}^{\infty}a_{1j}a_{2l}a_{3n}(b_{jl}+c_{jln}+d_{jln}),$$

where K_i $(i=1,2,3)$ are given by (3.2); a_{1j}, a_{2l} and a_{3n} are given by (3.5) for $i=1$, $l=j$; $i=2$, $l=l$ and $i=3$, $l=n$, and

$$b_{jl}=\sum_{k,m}\frac{(-1)^m U^k}{2m-2k-1}f(k,m,k+1)\log U,$$

$$c_{jln}=\sum_{\substack{k,m,r\\ r\neq k+1}}\frac{(-1)^{k+m+r}}{(2m-2k-1)(r-k-1)}f(k,m,r)(U^k-U^{r-1}),$$

$$d_{jln}=2\sum_{k,m,r}\frac{(-1)^{k+m+r}}{(2m-2k-1)(2m-2r-1)}f(k,m,r)(U^{m-1/2}-U^{r-1}),$$

and

$$(3.8)\quad f(k,m,r)=\binom{b+j}{k}\binom{b+l}{m}\binom{b+n}{r}.$$

For $p>3$ the density becomes too involved for presentation as well as for programming. However, the method of this section, provides a basis for a recursive algorithm for deriving the density at successive stages of the convolution process.

4. The stochastic inequality

The stochastic inequality (2.4), for $p=2$, can be written as

$$(4.1)\quad W_2(f_1,f_2;\delta_1^2,\delta_2^2)\succ\mathcal{L}(U(2,f_1,f_2))$$

where $\mathcal{L}(U(2,f_1,f_2))$ is given by (3.6) and for $p=3$, we get

$$(4.2)\quad W_3(f_1,f_2;\delta_1^2;\delta_2^2,\delta_3^2)\succ\mathcal{L}(U(3,f_1,f_2))$$

where $\mathcal{L}(U(3,f_1,f_2))$ is given by (3.7). Now the mathematical simplicity of $\mathcal{L}(U)$ makes possible the derivation of the corresponding cdf by straight forward integration. Indeed, the distribution function

$F_{B_p}(u)$ of the product of noncentral beta variates gives an upper bound of the distribution function $F_{W_p}(u)$ of the Wilks statistic. For $p=2$, we get

$$F_{W_2}(u)<F_{B_2}(u) \tag{4.3}$$

where

$$F_{B_2}(u)=4K_1K_2 \sum_{j,l=0}^{\infty} a_{1j}a_{2l} \sum_{k,m} \frac{(-1)^{k+m}u^{f_1/2}}{2m-2k-1} f(k, m, 0) \cdot \left(\frac{u^k}{f_1+2k}-\frac{u^{m-1/2}}{f_1+2m-1}\right), \tag{4.4}$$

K_i $(i=1, 2)$ are given by (3.2) and $f(k, m, r)$ is defined by (3.8) for $r=0$.

Then (4.4) gives the upper bound of the distribution of the likelihood-ratio test in the linear case (i.e. when the rank of the alternate hypothesis is one) for $\delta_2^2=0$, and also in the planar case (i.e. when the rank of the alternate hypothesis is two). The linear case has been studied by Gupta [7] and exact results here are available. Pillai and Jayachandran [13] have studied the planar case for $p=2$. However, little is known for $p=3$, except in the linear case (see Gupta [7]). The above analysis leads to the following inequality for $p=3$.

$$F_{W_3}(u)<F_{B_3}(u) \tag{4.5}$$

where

$$F_{B_3}(u)=4K_1K_2K_3u^{f_1/2} \sum_{j,l,n=0}^{\infty} a_{1j}a_{2l}a_{3n}(b'_{jl}+c'_{jln}+d'_{jln}), \tag{4.6}$$

$$b'_{jl}=\sum_{k,m} \frac{(-1)^m u^k}{(2m-2k-1)(f_1+2k)^2} f(k, m, k+1)((f_1+2k)\log u-2),$$

$$c'_{jln}= \sum_{\substack{k,m,r\\ r\neq k+1}} \frac{(-1)^{k+m+r}}{(2m-2k-1)(r-k-1)} f(k, m, r)\left(\frac{u^k}{f_1+2k}-\frac{u^{r-1}}{f_1+2r-2}\right),$$

$$d'_{jln}=2 \sum_{k,m,r} \frac{(-1)^{k+m+r}}{(2m-2k-1)(2m-2r+1)} f(k, m, r) \cdot \left(\frac{u^{m-1/2}}{f_1+2m-1}-\frac{u^{r-1}}{f_1+2r-2}\right),$$

and K_i $(i=1, 2, 3)$ are given by (3.2) and $f(k, m, r)$ is defined in (3.8).

Let us define the case when the alternative hypothesis is of rank 3, as spacial. Then (4.6) gives the upper bound of the distribution of the likelihood-ratio test in the linear, planar and spacial cases for $p=3$.

5. Computational value of the results

The expressions (4.4) and (4.6) were programmed and it has been verified that the total integral of the series obtained by taking a few terms only, rapidly approaches the theoretical value one as more terms are taken into account. Since $\Delta_2 \to \delta_2$ in probability as $f_1 \to \infty$ and/or $f_2 \to \infty$, a conservative evaluation of the power of the likelihood-ratio test criterion is provided by the results obtained in Section 4, which is expected to be a good approximation when f_1 and/or f_2 are large.

Table I. Exact and approximate powers of Wilks criterion for $p=2$, $f_1=72$ and $f_2=7$

δ_1^2	δ_2^2	Exact	Approx. (Mikhail)	A
0	1	0.07329	0.080	0.073
0	2	.1008	.120	.101
1	1	.1014	.120	.101
0	4	.1670	.234	.167
1	3	.1688	·234	.168
2	2	.1694	.234	.173

The power of Wilks statistic was approximately evaluated using (4.4) for $p=2$, $f_1=72$, $f_2=7$ and for various values of the noncentrality parameters δ_1^2 and δ_2^2. The results are given in Table I under A. Pillai and Jayachandran [13] have computed exact power in this case and have made accuracy checks of the approximation suggested by Mikhail [10]. It is evident from Table I that our approximation is excellent and a considerable improvement over Mikhail's approximation. For small

Table II. Percentage points for $p=3$ and $\alpha=0.95$

f_2	f_1	δ_1^2	δ_2^2	δ_3^2	u	E
2	8	0.5	0	0	.77569	.77570
		1.0	0	0	.76094	.76096
		4.0	0	0	.67010	.67011
	20	0.5	0	0	.91088	.91089
		1.0	0	0	.90402	.90403
		4.0	0	0	.85744	.85745
4	14	0	0	0.5	.67676	
		0	0.5	0.5	.66780	
		0.25	0.25	0.25	.67202	
		0	0	1	.66655	
		0	1	1	.64658	
		1	1	1	.62718	
		3	3	1	.51492	

values of the deviation parameters, Λ provides about three decimal place accuracy in the power. However, considering larger deviations also, our approximation is much better than Mikhail's. It may be added that notation is not standard from one author to another and our f_1 and f_2 are their (Pillai and Jayachandran [13]) $2n+p+1$ and $2m+p+1$ respectively.

The percentage points u for $p=3$, $\alpha=0.95$, $f_2=2$, $f_1=8, 20$ and $f_2=4$, $f_1=14$ have been computed for selected values of the noncentrality parameters δ_i^2, $i=1, 2, 3$, using the approximate distribution (4.6). These results are presented in Table II. In the last column we have given exact percentage points E in the linear case computed by the author [7]. A look at the table reveals the closeness of the approximate and exact percentage points and also the numerical feasibility of the formulae obtained here, even for small values of f_1 and f_2.

6. General remarks

This is well known (Ghosh [5], Kiefer and Schwartz [9]) that Wilks' Λ is unbiased, consistent and admissible and its power function is a monotonic-increasing function of each of the noncentrality parameters (Dasgupta, Anderson and Mudholkor [4]). Very little, however, is known about the actual magnitude of the power and this is due to the fact that the noncentral distribution of the test criterion has not been expressed in a numerically feasible form. From a practical standpoint, the results of this paper make possible for the first time the computation of the upper bound of the distribution for $p=3$, which in fact provides a good approximation for large f_1 and/or f_2.

The results obtained are valid for a class of alternatives called the class of "linear, planar or spacial" alternatives. This class in many important cases is completely general and includes all possible alternatives. Among these important cases are the multivariate test of equality of mean vectors in two, three or four groups, any two or three-variate test, the test of main effects in multivariate factorial experiments at two, three or four levels, and any other multivariate analysis of variance or regression analysis with one, two or three degrees of freedom for hypothesis. For cases where both the number of response variables and the degrees of freedom for hypothesis are greater than three, the class of linear, planar or spacial alternatives is not completely general.

UNIVERSITY OF MICHIGAN

REFERENCES

[1] Anderson, T. W. (1946). The noncentral Wishart distribution and certain problems of multivariate statistics, *Ann. Math. Statist.*, **17**, 409-431.

[2] Anderson, T. W. and Girshick, M. A. (1944). Some extensions of the Wishart distribution, *Ann. Math. Statist.*, **15**, 345-357.

[3] Asoh, Y. H. and Okamoto, M. (1969). A note on the nonnull distribution of the Wilks statistic in MANOVA, *Ann. Inst. Statist. Math.*, **21**, 67-72.

[4] Dasgupta, S., Anderson, T. W. and Mudholkar, G. S. (1964). Monotonocity of power functions of some tests of the multivariate linear hypothesis, *Ann. Math. Statist.*, **35**, 200-205.

[5] Ghosh, M. N. (1964). On the admissibility of some tests of MANOVA, *Ann. Math. Statist.*, **35**, 789-794.

[6] Gupta, A. K. (1968). Some central and noncentral distribution problems in multivariate analysis, *Mimeo. Series*, No. 139, Department of Statistics, Purdue University.

[7] Gupta, A. K. (1971). Noncentral distribution of Wilks statistic in MANOVA, *Ann. Math. Statist.*, **42**, 1254-1261.

[8] Katti, S. K. (1961). Distribution of the likelihood ratio for testing multivariate linear hypothesis, *Ann. Math. Statist.*, **32**, 333-335.

[9] Kiefer, J. and Schwartz, R. (1965). Admissible Bayes character of T^2-, R^2- and other fully invariant tests for classical multivariate normal problems, *Ann. Math. Statist.*, **36**, 747-770.

[10] Mikhail, N. N. (1965). A comparison of tests of the Wilks-Lawley hypothesis in multivariate analysis, *Biometrika*, **52**, 149-156.

[11] Pearson, E. S. and Wilks, S. S. (1933). Methods of analysis appropriate for k samples of two variables, *Biometrika*, **25**, 353-378.

[12] Pillai, K. C. S. and Gupta, A. K. (1969). On the exact distribution of Wilk's criterion, *Biometrika*, **56**, 109-118.

[13] Pillai, K. C. S. and Jayachandran, K. (1967). Power comparisons of tests of two multivariate hypotheses based on four criteria, *Biometrika*, **54**, 195-210.

[14] Schatzoff, M. (1966). Exact distribution of Wilks' likelihood-ratio criterion, *Biometrika*, **53**, 347-358.

[15] Wijsman, R. A. (1957). Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis, *Ann. Math. Statist.*, **28**, 415-423.

[16] Wilks, S. S. (1932). Certain generalisations in the analysis of variance, *Biometrika*, **24**, 471-494.

CORRECTION TO

"ON A STOCHASTIC INEQUALITY FOR THE WILKS STATISTIC"

A. K. GUPTA

In the above titled paper (this Annals 27 (1975), 341–348), the following correction should be made:

On page 346, last line: The value of δ_3^2 should read 3.