# On the Exact Distribution of Wilks' Criterion

By

S. Sylvia Chu and K. C. S. Pillai
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series No. 78-7

June, 1978

Abstract

On the Exact Distribution of Wilks' Criterion

by

S. Sylvia Chu & K. C. S. Pillai
Purdue University

The distribution of Wilks' criterion $W^{(p)}$ has been studied based on variable transformation. Exact distributions of $W^{(p)}$ were given explicitly for p=3 and 4.

On the Exact Distribution of Wilks' Criterion[1]

by

S. Sylvia Chu and K. C. S. Pillai
Purdue University

1. Introduction. The joint distribution of p non-zero characteristic roots of a matrix in multivariate analysis given by Fisher [1], Hsu [2] and Roy [7] can be expressed in the form:

$$(1.1)\qquad C(p,m,n) \prod_{i=1}^{p} b_i^n(1-b_i)^m \prod_{i<j} (b_i-b_j),\ 0 < b_p \leq \ldots \leq b_1 < 1,$$

where $C(p,m,n) = \{\pi^{\frac{1}{2}p} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(2m+2n+p+i+2)]\}$

$$/ \{ \prod_{i=1}^{p} \Gamma[\tfrac{1}{2}(2m+i+1)]\Gamma[\tfrac{1}{2}(2n+i+1)]\Gamma(\tfrac{1}{2}i)\},$$

and the parameters m and n are defined differently for various situations as described by Pillai [4,5]. Now Wilks' criterion $W^{(p)}$ may be defined as

$$(1.2)\qquad W^{(p)} = \prod_{i=1}^{p} b_i.$$

The moments of $W^{(p)}$ are obtained readily from (1.1) and (1.2) as follows:

$$(1.3)\qquad E[(W^{(p)})^h] = C(p,m,n)/C(p,m,n+h).$$

Using (1.3), Wilks [9] showed that $W^{(p)}$ can be expressed as a product of p independent beta variables and obtained the density of $W^{(2)}$, [10]. Schatzoff [8] has used the convolution method to compute exact percentage points for p up to 10 and values of n such that $p(2n+p+1) \leq 70$ through computer algorithm, but has not given the distribution explicitly for any p. Pillai and Gupta [6]

AMS 1970 Subject Classifications: 62H10, 62H15.

[1]This research was supported by David Ross Grant from Purdue University.

Key words and phrases: Exact distribution, Wilks' criterion, transformation of variables.

have found the exact distribution of $-\log W^{(p)}$ for p=3 to 6 by successive convolutions and extended the tables for these values of p and selected degrees of freedom for hypothesis. Lee [3] has also obtained the exact distribution of $(W^{(p)})^{\frac{1}{2}}$ for odd p in the form of an integral which could not be evaluated except numerically. In this paper, an attempt is made to obtain the exact distribution of $W^{(p)}$ by variable transformation. Explicit expressions for the distribution are given here for p=3 and 4.

2. Method of derivation of the density of $W^{(p)}$. Consider the transformation

$$q_1=b_1, q_i=b_i/b_{i-1}, i=2,\dots,p-2, q_{p-1}=b_{p-1}/b_{p-2}, q_p=b_p/b_{p-2}.$$

Then from (1.1) the joint density of $q_1,\dots,q_p$ is given by

$$C(p,m,n)\Big[\prod_{j=1}^{p-2} q_j^{k_0}\Big(1-\prod_{k=1}^{j} q_k\Big)^m\Big](q_{p-1}q_p)^n\Big[\prod_{j=2}^{p-2}\prod_{i=2}^{p-j}\Big(1-\prod_{k=i}^{i+j-2} q_k\Big)\Big]\Big[1-\Big(\prod_{j=1}^{p-2} q_j\Big)q_{p-1}\Big]^m$$
$$\Big[1-\Big(\prod_{j=1}^{p-2} q_j\Big)q_p\Big]^m\Big\{\prod_{j=2}^{p-2}\Big[1-\Big(\prod_{i=j}^{p-2} q_i\Big)(q_{p-1}+q_p)+\Big(\prod_{i=j}^{p-2} q_i^2\Big)q_{p-1}q_p\Big]\Big\}(1-q_{p-1})(1-q_p)(q_{p-1}-q_p), \tag{2.1}$$

where $0 < q_i < 1$, $i=1,\dots,p-2$, $0 < q_p \le q_{p-1} < 1$ and $k_0 = (p-j+1)n + (p-j) + \frac{1}{2}(p-j+1)(p-j)$. Further, transform $t = (1-q_{p-1})(1-q_p)$ and $g = q_{p-1}q_p$. Then the joint density of $q_1,\dots,q_{p-2},g,t$ is obtained as

$$C(p,m,n)\Big[\prod_{j=1}^{p-2} q_j^{k_0}\Big(1-\prod_{k=1}^{j} q_k\Big)^m\Big]g^n\Big[\prod_{j=2}^{p-2}\prod_{i=2}^{p-j}\Big(1-\prod_{k=i}^{i+j-2} q_k\Big)\Big]$$
$$\Big\{\prod_{j=2}^{p-2}\big[(1-a_j)(1-a_jg)+a_jt\big]\Big\}\big[(1-a_1)(1-a_1g)+a_1t\big]^m t, \tag{2.2}$$

where $0 < q_i < 1$, $i=1,\dots,p-2$, $g^{\frac{1}{2}} + t^{\frac{1}{2}} \le 1$ and $a_j = \prod_{i=j}^{p-2} q_i, j=1,\dots,p-2$. Now integrate out t and transform $w = \Big(\prod_{j=1}^{p-2} q_j^{p-j+1}\Big)g$ which is in fact Wilks' criterion $W^{(p)}$. Then the joint density of $q_1,\dots,q_{p-2}$, w is given by

(2.3) $$C(p,m,n)\Big[\prod_{j=1}^{p-2} q_j^{\frac{1}{2}(p-j+1)(p-j)-1}\Big(1-\prod_{k=1}^{j} q_k\Big)^m\Big] w^n \Big[\prod_{j=2}^{p-2}\prod_{i=2}^{p-j}\Big(1-\prod_{k=i}^{i+j-2} q_k\Big)\Big]$$

$$\Big[\sum_{h=0}^{m}\sum_{\ell=0}^{p-3}(h+\ell+2)^{-1} d_h C_\ell \Big(1-w^{\frac{1}{2}}\prod_{j=1}^{p-2} q_j^{-\frac{1}{2}(p-j+1)}\Big)^{2(h+\ell+2)},$$

where $d_h = \binom{m}{h}\Big[(1-a_1)\Big(1-a_1 w\prod_{j=1}^{p-2} q_j^{-p+j-1}\Big)\Big]^{m-h} a_1^h$, $h=0,\ldots,m$,

$$C_\ell = \sum_{j_1=2}^{p-2}\cdots\sum_{j_\ell=2}^{p-2}\Big(\prod_{i=1}^{\ell} a_{j_i}\Big)\Big[\prod_{k=2}^{p-2}(1-a_k)\Big(1-a_k w\prod_{j=1}^{p-2} q_j^{-p+j-1}\Big)\Big],\quad \ell=0,\ldots,p-3,$$

$$j_x > j_y \text{ if } x > y \qquad k \neq j_1,\ldots,j_\ell$$

and $w^{1/3}\prod_{j=1}^{p-3} q_j^{-(p-j+1)/3} < q_{p-2} < 1$, $w^{1/4}\prod_{j=1}^{p-4} q_j^{-(p-j+1)/4} < q_{p-3} < 1$,

$\ldots$, $w^{1/(p-1)} q_1^{-p/(p-1)} < q_2 < 1$, $w^{1/p} < q_1 < 1$, $0 < w < 1$, where integration is in the order $q_{p-2}, q_{p-3},\ldots,q_2,q_1$. From (2.4), the density function of Wilks' criterion $W^{(p)}$ can be obtained by integrating out $q_{p-2}, q_{p-3},\ldots,q_2,q_1$ successively. We may illustrate the method by considering p=3 and 4.

3. Exact distribution of $W^{(3)}$. Putting p=3 in (2.2), we get the joint density of $q_1$, g, t as

$$C(3,m,n)q_1^{3n+5}(1-q_1)^m g^n[(1-q_1)(1-q_1 g) + q_1 t]^m t,$$

where $g^{\frac{1}{2}} + t^{\frac{1}{2}} \leq 1$. Now integrate out t and transform $w=q_1^3 g$, the joint density of $q_1$,w becomes

$$C(3,m,n)\textstyle\sum_1 K_1 w^{n+k+\frac{1}{2}h} q_1^{m+2-i+j-2k-3h/2}, \quad 0 < w < q_1^3 < 1, \text{ where}$$

$$\textstyle\sum_1 = \sum_{i=0}^{m}\sum_{j=0}^{m+i}\sum_{k=0}^{i}\sum_{h=0}^{2m-2i+4},\quad K_1 = (m-i+2)^{-1}\binom{m}{i}\binom{m+i}{j}\binom{i}{k}\binom{2m-2i+4}{h}(-1)^{j+k+h}.$$

Now integrate out $q_1$, we get the density of $W^{(3)}$ as follows:

$$C(3,m,n)\{\sum_{\substack{1\\ a\neq 0}} K_1 a^{-1} w^{n+k+\frac{1}{2}h}(1-w^{a/3}) - \sum_{\substack{1\\ a=0}} \tfrac{1}{3}K_1 w^{n+k+\frac{1}{2}h}\ln w\}, \tag{3.1}$$

where $a=m+3-i+j-2k-3h/2$.

<u>m=0</u>: Put m=0 in (3.1) and we have the density

$$C(3,0,n)w^n[1-8w^{\frac{1}{2}}+8w^{3/2}-w^2-6w\ln w]/6, \tag{3.2}$$

and the exact cdf

$$P(W^{(3)} < w) = C(3,0,n)w^{n+1}\{[1/(n+1)]-[16/(2n+3)]w^{1/2}+[16/(2n+5)]w^{3/2} -[1/(n+3)]w^2-[6w/(n+2)^2][(n+2)\ln w-1]\}/6. \tag{3.3}$$

<u>m=1</u>: Put m=1 in (3.1) and we have the density

$$C(3,1,n)w^n[1-16w^{1/2}-65w+160w^{3/2}-65w^2-16w^{5/2}+w^3-30w(1-w)\ln w]/30, \tag{3.4}$$

and the exact cdf

$$P(W^{(3)} < w) = C(3,1,n)w^{n+1}\{[1/(n+1)]-[32/(2n+3)]w^{1/2}-[65/(n+2)]w +[320/(2n+5)]w^{3/2}-[65/(n+3)]w^2-[32/(2n+7)]w^{5/2}+[1/(n+4)]w^3 -[30w/(n+2)^2][(n+2)\ln w-1]+[30w^2/(n+3)^2][(n+3)\ln w-1]\}/30. \tag{3.5}$$

<u>m=2</u>: Put m=2 in (3.1) and we have the density

$$C(3,2,n)w^n[5-128w^{1/2}-1428w+4480w^{3/2}-4480w^{5/2}+1428w^3+128w^{7/2}-5w^4 -420w(1-5w+w^2)\ln w]/420, \tag{3.6}$$

and the exact cdf

$$P(W^{(3)} < w) = C(3,2,n)w^{n+1}\{[5/(n+1)]-[256/(2n+3)]w^{1/2}-[1428/(n+2)]w +[8960/(2n+5)]w^{3/2}-[8960/(2n+7)]w^{5/2}+[1428/(n+4)]w^3+[256/(2n+9)]w^{7/2} -[5/(n+5)]w^4-[420w/(n+2)^2][(n+2)\ln w-1]+[2100w^2/(n+3)^2][(n+3)\ln w-1] -[420w^3/(n+4)^2][(n+4)\ln w-1]\}/420. \tag{3.7}$$

4. <u>Exact distribution of $W^{(4)}$</u>. Putting p=4 in (2.2), we get the joint density of $q_1, q_2, q, t$ as

$$C(4,m,n)q_1^{4n+9}q_2^{3n+5}(1-q_1)^m(1-q_1q_2)^m g^n[(1-q_1q_2)(1-q_1q_2g)+q_1q_2t]^m$$
$$(1-q_2)[(1-q_2)(1-q_2g)+q_2t]t,$$

where $g^{\frac{1}{2}} + t^{\frac{1}{2}} \leq 1$. Now integrate out t and transform $w = q_1^4q_2^3g$, the joint density of $q_1, q_2$, w becomes

$$C(4,m,n)[\sum_2 K_2 w^c q_1^{a+5} q_2^{b+2}(1-q_2)^2(1-wq_1^{-4}q_2^{-2})+\sum_3 K_3 w^c q_1^{a+5}q_2^{b+3}(1-q_2)],$$

where $w^{\frac{1}{3}} q_1^{-\frac{4}{3}} < q_2 < 1$, $w^{\frac{1}{4}} < q_1 < 1$, $0 < w < 1$,

$$\sum_2 = \sum_{i=0}^{m}\sum_{j=0}^{m}\sum_{k=0}^{m+i}\sum_{\ell=0}^{i}\sum_{h=0}^{2m-2i+4},\quad K_2 = (m-i+2)^{-1}\binom{m}{i}\binom{m}{j}\binom{m+i}{k}\binom{i}{\ell}\binom{2m-2i+4}{h}(-1)^{j+k+\ell+h},$$

$$\sum_3 = \sum_{i=0}^{m}\sum_{j=0}^{m}\sum_{k=0}^{m+i}\sum_{\ell=0}^{i}\sum_{h=0}^{2m-2i+6},\quad K_3 = (m-i+3)^{-1}\binom{m}{i}\binom{m}{j}\binom{m+i}{k}\binom{i}{\ell}\binom{2m-2i+6}{h}(-1)^{j+k+\ell+h},$$

$a = m-i+j+k-3\ell-2h$, $b=m-i+k-2\ell-\frac{3}{2}h$, $c=n+\ell+\frac{1}{2}h$.

Now integrate out $q_2$ to get the joint density of $q_1$,w: (Each summation is taken such that the denominator of the factor does not vanish.)

$$C(4,m,n)[\sum_4\alpha_{05} + \sum_5\alpha_{05}(\ell n\ w-\ell n\ q_1) + \sum_6\alpha_{11} + \sum_7\alpha_{11}(\ell n\ w-\ell n\ q_1)-\sum_2 \frac{K_2}{b+3}\beta_{00}$$

$$+ \sum_2 \frac{2K_2}{b+4}\beta_{14}-\sum_2 \frac{K_2}{b+5}\beta_{28} + \sum_2 \frac{K_2}{b+1}\gamma_{00}-\sum_2 \frac{2K_2}{b+2}\gamma_{14} + \sum_2 \frac{K_2}{b+3}\gamma_{28}-\sum_3 \frac{K_3}{b+4}\beta_{14}+\sum_3 \frac{K_3}{b+5}\beta_{28}],$$

where $\sum_4 = \sum_2 \frac{K_2}{b+3} - \sum_2 \frac{2K_2}{b+4} + \sum_2 \frac{K_2}{b+5} + \sum_3 \frac{K_3}{b+4} - \sum_3 \frac{K_3}{b+5}$,

$$\sum_5 = -\frac{1}{3}\left(\sum_{\substack{2\\b+3=0}} K_2 - \sum_{\substack{2\\b+4=0}} 2K_2 + \sum_{\substack{2\\b+5=0}} K_2 + \sum_{\substack{3\\b+4=0}} K_3 - \sum_{\substack{3\\b+5=0}} K_3\right),$$

$$\sum_6 = -\sum_2 \frac{K_2}{b+1} + \sum_2 \frac{2K_2}{b+2} - \sum_2 \frac{K_2}{b+3},\quad \sum_7 = \frac{1}{3}\left(\sum_{\substack{2\\b+1=0}} K_2 - \sum_{\substack{2\\b+2=0}} 2K_2 + \sum_{\substack{2\\b+3=0}} K_2\right),$$

$r = c + (b+3)/3$, $s = c+1+(b+1)/3$, $d = a+5-4(b+3)/3$, $e = a+1-4(b+1)/3$,

$\alpha_{ij} = w^{c+i} q_1^{a+j}$, $\beta_{ij} = w^{r+i/3} q_1^{d-j/3}$, $\gamma_{ij} = w^{s+i/3} q_1^{e-j/3}$.

Finally integrate out $q_1$, we get the density of $W^{(4)}$ as follows:

$$(4.1) \qquad C(4,m,n)\{[\textstyle\sum_4 \frac{1}{a+6} + \sum_5 \frac{4}{(a+6)^2}]\phi_{06} + \sum_5 \frac{1}{a+6}\varepsilon_{01} - \sum_{4\,(a+6=0)} \frac{1}{4}\varepsilon_{01} - \sum_{5\,(a+6=0)} \frac{1}{8}\varepsilon_{02}$$

$$+[\textstyle\sum_6 \frac{1}{a+2} + \sum_7 \frac{4}{(a+2)^2}]\phi_{12} + \sum_7 \frac{1}{a+2}\varepsilon_{11} - \sum_{6\,(a+2=0)} \frac{1}{4}\varepsilon_{11} - \sum_{7\,(a+2=0)} \frac{1}{8}\varepsilon_{12}$$

$$-\textstyle\sum_2 \frac{K_2}{(b+3)(d+1)}\delta_{0,-4} + [\sum_2 \frac{6K_2}{(b+4)(3d-1)} - \sum_3 \frac{3K_3}{(b+4)(3d-1)}]\delta_{11}$$

$$-[\textstyle\sum_2 \frac{3K_2}{(b+5)(3d-5)} - \sum_3 \frac{3K_3}{(b+5)(3d-5)}]\delta_{25} + \sum_2 \frac{K_2}{(b+1)(e+1)}\theta_{0,-4}$$

$$-\textstyle\sum_2 \frac{6K_2}{(b+2)(3e-1)}\theta_{11} + \sum_2 \frac{3K_2}{(b+3)(3e-5)}\theta_{25} + [\sum_{2\,(d+1=0)} \frac{K_2}{b+3} w^r$$

$$-(\textstyle\sum_{2\,(3d-1=0)} \frac{2K_2}{b+4} - \sum_{3\,(3d-1=0)} \frac{K_3}{b+4})w^{r+1/3} + (\sum_{2\,(3d-5=0)} \frac{K_2}{b+5} - \sum_{3\,(3d-5=0)} \frac{K_3}{b+5})w^{r+2/3}$$

$$-\textstyle\sum_{2\,(e+1=0)} \frac{K_2}{b+1} w^s + \sum_{2\,(3e-1=0)} \frac{2K_2}{b+2} w^{s+1/3} - \sum_{2\,(3e-5=0)} \frac{K_2}{b+3} w^{s+2/3}]\ln w/4\},$$

where $\phi_{ij} = w^{c+i}(1-w^{\frac{1}{4}(a+j)})$, $\varepsilon_{ij} = w^{c+i}(\ln w)^j$, $\delta_{ij} = w^{r+i/3}(1-w^{\frac{1}{4}(d-j)})$, $\theta_{ij} = w^{s+i/3}(1-w^{\frac{1}{4}(e-j)})$.

m=0: Put m=0 in (4.1) and we have the density

(4.2) $\quad C(4,0,n)w^{n}[1-24w^{1/2}-375w+375w^{2}+24w^{5/2}-w^{3}-90w(1+240w^{1/2}+w)\ell n\ w]/180,$

and the exact cdf

(4.3) $$P(W^{(4)} < w) = C(4,0,n)w^{n+1}\{[1/(n+1)]-[48/(2n+3)]w^{1/2}-[375/(n+2)]w$$
$$+[375/(n+3)]w^{2}+[48/(2n+7)]w^{5/2}-[1/(n+4)]w^{3}-[90w/(n+2)^{2}][(n+2)\ell n\ w-1]$$
$$-[480w^{3/2}/(2n+5)^{2}][(2n+5)\ell n\ w-2]-[90w^{2}/(n+3)^{2}][(n+3)\ell n\ w-1]\}/180.$$

The special cases are exactly the same as those from the results of Pillai and Gupta [6]. However, there are some additional terms in the general formulae derived here in comparison to theirs but they cancel out under specialization.

## References

[1] Fisher, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. Ann. Eugenics, 9, 238-249.

[2] Hsu, P. L. (1939). On the distribution of the roots of certain determinantal equations. Ann. Eugenics, 9, 250-258.

[3] Lee, Y. S. (1972). Some results on the distribution of Wilks' likelihood-ratio criterion. Biometrika, 59, 649-664.

[4] Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. Ann. Math. Statist., 26, 117-121.

[5] Pillai, K. C. S. (1956). Some results useful in multivariate analysis. Ann. Math. Statist., 27, 1106-1114.

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

[7] Roy, S. N. (1939). p-statistics or some generalizations in analysis of variance appropriate to multivariate problems. Sankhyā, 4, 381-396.

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

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

[10] Wilks, S. S. (1935). On the independence of k sets of normally distributed variables. Econometrika, 3, 309-326.

S. Sylvia Chu and K. C. S. Pillai

## List of Symbols

Greek letters:

| | | |
|---|---|---|
| Π | u.c. pi | (e.g. p. 1, line 8) |
| π | l.c. pi | (e.g. p. 1, line 9) |
| Γ | u.c. gamma | (e.g. p. 1, line 9) |
| ∑ | u.c. sigma | (e.g. p. 3, line 2) |
| α | l.c. alpha | (e.g. p. 5, line 14) |
| β | l.c. beta | (e.g. p. 5, line 14) |
| γ | l.c. gamma | (e.g. p. 5, line 15) |
| ϕ | l.c. phi | (e.g. p. 6, line 2) |
| ε | l.c. epsilon | (e.g. p. 6, line 2) |
| δ | l.c. delta | (e.g. p. 6, line 3) |
| θ | l.c. theta | (e.g. p. 6, line 5) |

Non-italic symbols:

| | |
|---|---|
| log | (e.g. p. 2, line 1) |
| ℓn | (as in eq. (3.1)) |

Short title:

Distribution of Wilks' criterion