ASYMPTOTIC FORMULAE FOR THE DISTRIBUTION OF HOTELLING'S GENERALIZED T_0^2 STATISTIC

by

Koichi Ito
University of North Carolina

Special report to the Office of Naval Research of work at Chapel Hill under Contract NR 042 031, Project N7-onr-284(02), for research in probability and statistics.

Institute of Statistics
Mimeograph Series No. 133
June 21, 1955

ASYMPTOTIC FORMULAE FOR THE DISTRIBUTION OF HOTELLING'S GENERALIZED T_0^2 STATISTIC

By Koichi Ito
University of North Carolina

1. Summary.

In this paper the asymptotic expansion of a percentage point of Hotelling's generalized T_0^2 distribution is derived in terms of the corresponding percentage point of a χ^2 distribution. Our result generalizes Hotelling's and Frankel's asymptotic expansion for the generalized Student T [2], [3]. The technique used in this paper for obtaining the asymptotic expansion of T_0^2 is an extension of the previous methods of Welch [7] and of James [4], [5], who used them to solve the distribution problem of various statistics in connection wth the Behrens-Fisher problem. An asymptotic formula for the cumulative distribution function (c.d.f.) of T_0^2 is also given together with an upper bound for the error committed when all but the first few terms are omitted in the series. This formula is a sort of multivariate analogue of Hartley's formula of "Studentization" [1].

2. Introduction.

In the multivariate analysis of variance we use the following canonical probability law:

$$(2.1)\qquad P(X_0, X_1) = \text{const.}\exp\left[-\tfrac{1}{2}\mathrm{tr}\,\Lambda(X_1-\xi)(X_1'-\xi')-\tfrac{1}{2}\mathrm{tr}\,\Lambda X_0X_0'\right]dX_0dX_1,$$

where X_1 and X_0 are pxm and pxn matrices respectively, and $\frac{1}{m}X_1X_1' = S_1$ is the sample "between" dispersion matrix and $\frac{1}{n}X_0X_0' = S_0$ is the sample "within" dispersion matrix, the prime denoting the transpose of a matrix. ξ is a $p \times m$ matrix, $\frac{1}{m}\xi\xi'$ being the population "between" dispersion matrix, and Λ is a $p \times p$ symmetric positive definite matrix. It is assumed that m may be $\geq p$ or $< p$, but $n \geq p$. To test the null

[1] Work at Chapel Hill sponsored by the Office of Naval Research under Contract NR 042 031. Reproduction in whole or in part is permitted for any purpose of the United States government.

hypothesis H_0: $\xi = 0$, Hotelling [2] proposed a test based on the statistic:

$$T_0^2 = m \operatorname{tr} S_1 S_0^{-1} \tag{2.2}$$

and derived the exact distribution of this statistic when $p = 2$ and $\xi = 0$. For general values of p the exact distribution of T_0^2 is not available at present, even in the null case $\xi = 0$.

3. Derivation of asymptotic formula of T_0^2.

For general values of p it is known that the statistic

$$X^2 = m \operatorname{tr} S_1 \Lambda \tag{3.1}$$

has a X^2 distribution with m p degrees of freedom. That is to say, we have

$$\Pr \left\{ m \operatorname{tr} S_1 \Lambda \leq 2\theta \right\} = G_\rho(\theta) , \tag{3.2}$$

where 2θ denotes the tabled value of X^2 for a particular level of significance, $\rho = mp/2$, and

$$G_\rho(\theta) = [\Gamma(\rho)]^{-1} \int_0^\theta t^{\rho-1} e^{-t} \, dt. \tag{3.3}$$

Hence, if Λ is known, the statistic X^2 given by (3.1) may be used to test H_0 exactly, and if Λ is unknown but if S_0 is based on a large number of degrees of freedom, i.e., if n is large, we may use as an approximation the result:

$$\Pr \left\{ m \operatorname{tr} S_1 S_0^{-1} \leq 2\theta \right\} = G_\rho(\theta). \tag{3.4}$$

This suggests that in the general case we try to find a function $h(S_0)$ of the elements of S_0 such that

$$\Pr \left\{ m \operatorname{tr} S_1 S_0^{-1} \leq 2h(S_0) \right\} = G_\rho(\theta) . \tag{3.5}$$

When n is large, $2h(S_0)$ will approach $2\theta \equiv \chi^2$, and we now expect to write $2h(S_0)$ as a series with χ^2 as its first term and successive terms of decreasing order of magnitude.

Now

$$(3.6)\quad \Pr\left\{m \operatorname{tr} S_1S_0^{-1} \leq 2h(S_0)\right\} = \int_R \Pr\left\{m \operatorname{tr} S_1S_0^{-1} \leq 2h(S_0)\,|S_0\right\} \Pr\left\{d\, S_0\right\} \quad ,$$

where the first expression on the right denotes the conditional probability of the relation indicated for fixed values of the elements of S_0, and the second denotes the probability element of S_0, which has a Wishart distribution with n degrees of freedom, and the domain of integration R is over all possible values of the elements of S_0. Now we may expand $\Pr\left\{m \operatorname{tr} S_1S_0^{-1} \leq 2h(S_0)\,|S_0\right\}$ about an origin $(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp}, \sigma_{12}, \ldots, \sigma_{p-1,p})$ in a Taylor series, where

$$\Lambda^{-1} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} .$$

Thus,

$$(3.7)\quad \Pr\left\{m \operatorname{tr} S_1S_0^{-1} \leq 2h(S_0)\,|S_0\right\}$$

$$= \left\{\exp\left[\sum_{i\leq j=1}^{p} (s_{0ij}-\sigma_{ij}) \frac{\partial}{\partial\sigma_{ij}}\right]\right\} \Pr\left\{m \operatorname{tr} S_1\Lambda \leq 2h(\Lambda^{-1})\right\}$$

$$= \left\{\exp\left[\operatorname{tr}(S_0 - \Lambda^{-1})\partial\right]\right\} \Pr\left\{m \operatorname{tr} S_1\Lambda \leq 2h(\Lambda^{-1})\right\} \quad ,$$

where s_{0ij} is the i-th row, j-th column element of S_0, and ∂ denotes the matrix of derivative operators:

$$(3.8) \qquad \partial = \begin{pmatrix} \frac{\partial}{\partial\sigma_{11}} & \frac{1}{2}\frac{\partial}{\partial\sigma_{12}} & \cdots & \frac{1}{2}\frac{\partial}{\partial\sigma_{1p}} \\ \frac{1}{2}\frac{\partial}{\partial\sigma_{21}} & \frac{\partial}{\partial\sigma_{22}} & \cdots & \frac{1}{2}\frac{\partial}{\partial\sigma_{2p}} \\ \cdot & \cdot & \cdots & \cdot \\ \frac{1}{2}\frac{\partial}{\partial\sigma_{p1}} & \frac{1}{2}\frac{\partial}{\partial\sigma_{p2}} & \cdots & \frac{\partial}{\partial\sigma_{pp}} \end{pmatrix}$$

its typical element being $\partial_{ij} = \frac{1}{2}(1 + \delta_{ij}) \frac{\partial}{\partial\sigma_{ij}}$, where δ_{ij} is the Kronecker delta.

Whether uniformly convergent or not, the right hand side of (3.7) is an asymptotic representation of $\Pr\left\{m \text{ tr } S_1 S_0^{-1} \le 2h(S_0) \,|\, S_0\right\}$, for sufficiently large values of n. Hence, substitution of (3.7) into (3.6) and term by term integration which may be done legitimately, yields:

$$(3.9) \quad G_\rho(\theta) = \int_R \exp\left[\text{tr}(S_0 - \Lambda^{-1})\partial\right] \Pr\left\{m \text{ tr } S_1\Lambda \le 2h(\Lambda^{-1})\right\} \Pr\left\{d\, S_0\right\}$$

$$= \Theta \Pr\left\{m \text{ tr } S_1\Lambda \le 2h(\Lambda^{-1})\right\} ,$$

where

$$\Theta = \int_R \exp\left[\text{tr}(S_0 - \Lambda^{-1})\partial\right] \Pr\left\{d\, S_0\right\} .$$

Since S_0 has a Wishart distribution with n degrees of freedom, we have

$$\Theta = \exp\left[-\text{tr}\,\Lambda^{-1}\partial\right]\cdot\text{const.}|\Lambda|^{\frac{n}{2}} \int_R |S_0|^{\frac{n-p-1}{2}} \exp\left[\text{tr}\,(S_0\,\partial - \tfrac{n}{2}\Lambda S_0)\right] d\,S_0$$

$$= \exp\left[-\text{tr}\,\Lambda^{-1}\partial\right]\cdot\text{ const. } |\Lambda|^{\frac{n}{2}} \int_R |S_0|^{\frac{n-p-1}{2}} \exp\left[-\tfrac{n}{2}\,\text{tr}(\Lambda - \tfrac{2}{n}\partial)S_0\right] d\,S_0$$

$$= \exp\left[-\text{ tr }\Lambda^{-1}\partial\right]\cdot |\Lambda|^{\frac{n}{2}}\, |\Lambda - \tfrac{2}{n}\partial|^{-\frac{n}{2}}$$

$$= \exp\left[-\text{tr }\Lambda^{-1}\partial\right] \cdot |I - \tfrac{2}{n}\Lambda^{-1}\partial|^{-\frac{n}{2}} ,$$

where I is the p x p identity matrix. Now using [5],

(3.10) $$-\log|I - Y| = \mathrm{tr}\, Y + \tfrac{1}{2}\mathrm{tr}\, Y^2 + \tfrac{1}{3}\,\mathrm{tr}\, Y^3 + \dots,$$

we obtain

(3.11)
$$\begin{aligned}\Theta &= \exp\left[-\mathrm{tr}\,\Lambda^{-1}\partial - \tfrac{n}{2}\log\left|I - \tfrac{2}{n}\Lambda^{-1}\partial\right|\right] \\ &= \exp\left[-\mathrm{tr}\,\Lambda^{-1}\partial + \tfrac{n}{2}\left\{\mathrm{tr}(\tfrac{2}{n}\Lambda^{-1}\partial) + \tfrac{1}{2}\mathrm{tr}(\tfrac{2}{n}\Lambda^{-1}\partial)^2 + \tfrac{1}{3}\mathrm{tr}(\tfrac{2}{n}\Lambda^{-1}\partial)^3 + \dots\right\}\right] \\ &= \exp\left[\tfrac{1}{n}\,\mathrm{tr}(\Lambda^{-1}\partial)^2 + \tfrac{4}{3n^2}\,\mathrm{tr}\,(\Lambda^{-1}\partial)^3 + \dots\right] \\ &= 1 + \tfrac{1}{n}\,\mathrm{tr}(\Lambda^{-1}\partial)^2 + \tfrac{1}{n^2}\left\{\tfrac{4}{3}\mathrm{tr}(\Lambda^{-1}\partial)^3 + \tfrac{1}{2}(\mathrm{tr}(\Lambda^{-1}\partial)^2)^2\right\} + O(n^{-3})\,.\end{aligned}$$

It is to be noted here that in (3.11) the operator ∂ does not act on Λ^{-1} present in Θ itself, and it is more useful for our purpose to write (3.11) in suffix form:

(3.12)
$$\begin{aligned}\Theta &= 1 + \tfrac{1}{n}\Sigma\,\sigma_{rs}\,\sigma_{tu}\,\partial_{st}\,\partial_{ur} \\ &+ \tfrac{1}{n^2}\left\{\tfrac{4}{3}\Sigma\,\sigma_{rs}\sigma_{tu}\sigma_{vw}\partial_{st}\partial_{uv}\partial_{wr} + \tfrac{1}{2}\Sigma\,\sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}\partial_{st}\partial_{ur}\partial_{wx}\partial_{yv}\right\} \\ &+ O(n^{-3}),\end{aligned}$$

where Σ denotes the summation over all suffixes r, s, ... each of which ranges from 1 to p.

Now we represent $h(S_0)$ as

(3.13) $$h(S_0) = \theta + h_1(S_0) + h_2(S_0) + \dots,$$

$h_s(S_0)$ being of order n^{-s}, i.e., we write $h(S_0)$ as an asymptotic series such that

$$\left|n^s\left\{h(S_0) - \theta - h_1(S_0) - \dots - h_s(S_0)\right\}\right|$$

is made arbitrarily small for sufficiently large values of n. Then (3.13) may be substituted into $\Pr\left\{m\,\mathrm{tr}\,S_1\Lambda \le 2h(\Lambda^{-1})\right\}$, and by Taylor's expansion we have:

$$(3.14)\quad \Pr\left\{m \operatorname{tr} \tilde{S}_1 \Lambda \leq 2h(\Lambda^{-1})\right\}$$

$$= \exp\left[\left\{h_1(\Lambda^{-1}) + h_2(\Lambda^{-1}) + \ldots\right\} D\right] \Pr\left\{m \operatorname{tr} S_1 \Lambda \leq 2\theta\right\}$$

$$= \left[1 + \left\{h_1(\Lambda^{-1}) + h_2(\Lambda^{-1}) + \ldots\right\} D + \frac{1}{2}\left\{h_1(\Lambda^{-1}) + h_2(\Lambda^{-1}) + \ldots\right\}^2 D^2 + \ldots\right]$$

$$\times \Pr\left\{m \operatorname{tr} S_1 \Lambda \leq 2\theta\right\},$$

where $D \equiv \frac{\partial}{\partial\theta}$. By substituting (3.12) and (3.14) into (3.9) we obtain

$$(3.15)\quad G_\rho(\theta) = \left[1 + \frac{1}{n}\Sigma\, \sigma_{rs}\sigma_{tu}\partial_{st}\partial_{ur} + \frac{1}{n^2}\left\{\frac{4}{3}\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\partial_{st}\partial_{uv}\partial_{wr}\right.\right.$$

$$\left.\left. + \frac{1}{2}\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}\partial_{st}\partial_{ur}\partial_{wx}\partial_{yv}\right\} + O(n^{-3})\right]$$

$$\times \left[1 + h_1(\Lambda^{-1})D + \left\{h_2(\Lambda^{-1})\, D + \frac{1}{2}h_1^2(\Lambda^{-1})\, D^2\right\} + O(n^{-3})\right]$$

$$\times \Pr\left\{m \operatorname{tr} S_1 \Lambda \leq 2\theta\right\}.$$

By equating terms of successive order in (3.15) we obtain

$$(3.16)\quad \left\{h_1(\Lambda^{-1})D + \frac{1}{n}\Sigma\, \sigma_{rs}\sigma_{tu}\partial_{st}\partial_{ur}\right\} \Pr\left\{m \operatorname{tr} S_1 \Lambda \leq 2\theta\right\} = 0,$$

$$(3.17)\quad \left[h_2(\Lambda^{-1})D + \frac{1}{2}h_1^2(\Lambda^{-1})D^2 + \frac{1}{n}\Sigma\, \sigma_{rs}\sigma_{tu}\left\{h_1^{(st,ur)}(\Lambda^{-1})\, D\right.\right.$$

$$\left. + 2h_1^{(st)}(\Lambda^{-1})\, \partial_{ur}D + h_1(\Lambda^{-1})\, \partial_{st}\partial_{ur}D\right\}$$

$$\left. + \frac{4}{3n^2}\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\partial_{st}\partial_{uv}\partial_{wr} + \frac{1}{2n^2}\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}\partial_{st}\partial_{ur}\partial_{wx}\partial_{yv}\right]$$

$$\times \Pr\left\{m \operatorname{tr} S_1 \Lambda \leq 2\theta\right\} = 0,$$

and so on, where $h_1^{(st)}(\Lambda^{-1}) = \partial_{st}h_1(\Lambda^{-1})$ and $h_1^{(st,\, ur)}(\Lambda^{-1}) = \partial_{ur}\partial_{st}\, h_1(\Lambda^{-1})$.

It now remains to carry out the operations ∂ and D indicated in (3.16) and (3.17) in order to obtain $h_1(\Lambda^{-1})$, $h_2(\Lambda^{-1})$ and hence $h_1(S_0)$, $h_2(S_0)$. These operators will operate on $\Pr\{m \operatorname{tr} S_1\Lambda \leq 2\theta\}$ which is a p x m fold integral, and the operations may be thought of as differentiations, with respect to the boundary only, of the integral of the probability density function of the X_1 throughout a region in the space of X_1. The method used to evaluate $\partial_{st}\partial_{ur} \Pr\{m \operatorname{tr} S_1\Lambda \leq 2\theta\}$, $\partial_{st}\partial_{uv}\partial_{wr} \Pr\{m \operatorname{tr} S_1\Lambda \leq 2\theta\}$, ..., is to change the boundary slightly, expand the integral in powers of the quantities specifying this change, and obtain the derivatives by comparison with Taylor's expansion. We consider

$$J = \Pr\left\{m \operatorname{tr} S_1(\Lambda^{-1} + \epsilon)^{-1} \leq 2\theta\right\}, \tag{3.18}$$

where ϵ is a p x p symmetric matrix. Then by Taylor expansion we have

$$J = \Big\{1 + \Sigma\, \epsilon_{rs}\partial_{rs} + \frac{1}{2!}\Sigma\, \epsilon_{rs}\epsilon_{tu}\partial_{rs}\partial_{tu} + \frac{1}{3!}\Sigma\, \epsilon_{rs}\epsilon_{tu}\epsilon_{vw}\partial_{rs}\partial_{tu}\partial_{vw} + \frac{1}{4!}\Sigma\, \epsilon_{rs}\epsilon_{tu}\epsilon_{vw}\epsilon_{xy}\partial_{rs}\partial_{tu}\partial_{vw}\partial_{xy} + \dots\Big\} \Pr\left\{m \operatorname{tr} S_1\Lambda \leq 2\theta\right\}. \tag{3.19}$$

On the other hand, J is, by definition, written as

$$J = \frac{|\Lambda|^{\frac{m}{2}}}{(2\pi)^{\frac{pm}{2}}} \int_{R'} \exp\left[-\tfrac{1}{2}\operatorname{tr} \Lambda X_1 X_1'\right] dX_1, \tag{3.20}$$

where $X_1X_1' = mS_1$, and domain of integration R' ranges over all possible values of the elements of X_1 such that $m \operatorname{tr} S_1(\Lambda^{-1} + \epsilon)^{-1} \leq 2\theta$. It is now easy to show that integration of (3.20) yields

$$J = \left(\frac{|I-D_\eta E|}{|I-D_\eta|}\right)^{-\frac{m}{2}} G_\rho(\theta), \tag{3.21}$$

where D_η is a diagonal matrix which satisfies

(3.22) $$X_1(p \times m) = \Gamma(p \times p)\, \Xi(p \times m),$$

$$\tfrac{1}{2}\Gamma'(\Lambda^{-1}+\epsilon)^{-1}\,\Gamma = I(p),$$

and $$\tfrac{1}{2}\Gamma'\Lambda\,\Gamma = I(p) - D_\eta,$$

Γ being a non-singular matrix, and E is an operator such that

$$E\, G_\rho(\theta) = G_{\rho+1}(\theta).$$

Now letting $\Delta = E - 1$ and using (3.22) we have

$$\frac{|I-D_\eta E|}{|I-D_\eta|} = \frac{|I-D_\eta-D_\eta\Delta|}{|I-D_\eta|}$$

$$= \frac{|\frac{1}{2}\Gamma'\Lambda\,\Gamma - \{\frac{1}{2}\Gamma'(\Lambda^{-1}+\epsilon)^{-1}\Gamma - \frac{1}{2}\Gamma'\Lambda\,\Gamma\}\Delta|}{|\frac{1}{2}\Gamma'\Lambda\,\Gamma|}$$

$$= \frac{|\Lambda - \{(\Lambda^{-1}+\epsilon)^{-1} - \Lambda\}\Delta|}{|\Lambda|} = |I - \{\Lambda^{-1}(\Lambda^{-1}+\epsilon)^{-1} - I\}\Delta|$$

$$= |I - X\Delta|$$

where $X = \Lambda^{-1}(\Lambda^{-1}+\epsilon)^{-1} - I$. Hence (3.21) becomes

(3.23) $$J = |I - X\Delta|^{-\frac{m}{2}}\, G_\rho(\theta)$$

Now using (3.10) again, we rewrite (3.23) as

(3.24) $$J = \exp\left\{-\frac{m}{2}\log|I - X\Delta|\right\} G_\rho(\theta)$$

$$= \exp\left\{\frac{m}{2}\operatorname{tr} X\Delta + \frac{m}{4}\operatorname{tr} X^2\Delta^2 + \frac{m}{6}\operatorname{tr} X^3\Delta^3 + \frac{m}{8}\operatorname{tr} X^4\Delta^4 + \dots\right\} G_\rho(\theta)$$

$$= [1 + \frac{m}{2}\text{tr } X\Delta + \left\{\frac{m}{4}\text{tr } X^2 + \frac{m^2}{8}(\text{tr } X)^2\right\}\Delta^2 + \left\{\frac{m}{6}\text{ tr } X^3 + \frac{m^2}{8}(\text{tr } X)(\text{tr } X^2) + \frac{m^3}{48}(\text{tr } X)^3\right\}\Delta^3$$

$$+ \left\{\frac{m}{8}\text{ tr } X^4 + \frac{m^2}{12}(\text{tr } X)(\text{tr } X^3) + \frac{m^2}{32}(\text{tr } X^2)^2\right.$$

$$\left. + \frac{m^3}{32}(\text{tr } X)^2(\text{tr } X^2) + \frac{m^4}{384}(\text{tr } X)^4\right\}\Delta^4 + \ldots]G_\rho(\Theta).$$

X can be represented as

$$(3.25)\quad X = \Lambda^{-1}(\Lambda^{-1}+\epsilon)^{-1} - I = \Lambda^{-1}(\Lambda^{-1} + \Sigma\epsilon_{rs}\Lambda_{rs}^{-1})^{-1} - I = (I + \Sigma\epsilon_{rs}\Lambda_{rs}^{-1}\Lambda)^{-1} - I$$

$$= -\Sigma\epsilon_{rs}(\Lambda_{rs}^{-1}\Lambda) + \Sigma\epsilon_{rs}\epsilon_{tu}(\Lambda_{rs}^{-1}\Lambda)(\Lambda_{tu}^{-1}\Lambda) - \Sigma\epsilon_{rs}\epsilon_{tu}\epsilon_{vw}(\Lambda_{rs}^{-1}\Lambda)(\Lambda_{tu}^{-1}\Lambda)(\Lambda_{vw}^{-1}\Lambda)$$

$$+ \Sigma\,\epsilon_{rs}\epsilon_{tu}\epsilon_{vw}\epsilon_{xy}(\Lambda_{rs}^{-1}\Lambda)(\Lambda_{tu}^{-1}\Lambda)(\Lambda_{vw}^{-1}\Lambda)(\Lambda_{xy}^{-1}\Lambda) - \ldots,$$

where Λ_{rs}^{-1} is a p x p matrix obtained by operating ∂_{rs} on Λ, i.e., Λ_{rs}^{-1} has its i-th row, j-th column element, $\frac{1}{2}(\delta_{ri}\delta_{sj} + \delta_{si}\delta_{rj})$. Writing

$$\text{tr}(\Lambda_{rs}^{-1}\Lambda) = (rs),$$

$$\text{tr}(\Lambda_{rs}^{-1}\Lambda)(\Lambda_{tu}^{-1}\Lambda) = (rs|tu),$$

$$\text{tr}(\Lambda_{rs}^{-1}\Lambda)(\Lambda_{tu}^{-1}\Lambda)(\Lambda_{vw}^{-1}\Lambda) = (rs|tu|vw),$$

$$\text{tr}(\Lambda_{rs}^{-1}\Lambda)(\Lambda_{tu}^{-1}\Lambda)(\Lambda_{vw}^{-1}\Lambda)(\Lambda_{xy}^{-1}\Lambda) = (rs|tu|vw|xy),$$

and substituting (3.25) into (3.24), we obtain

$$(3.26)\quad J = [1 + \Sigma\epsilon_{rs}\left\{-\frac{m}{2}(rs)\Delta\right\} + \frac{1}{2!}\Sigma\epsilon_{rs}\epsilon_{tu}\left\{(rs|tu)(m\Delta + \frac{m}{2}\Delta^2) + \frac{m^2}{4}(rs)(tu)\Delta^2\right\}$$

$$+\frac{1}{3!}\Sigma\epsilon_{rs}\epsilon_{tu}\epsilon_{vw}\left\{(rs|tu|vw)(-3m\Delta - 3m\Delta^2 - m\Delta^3) + (rs)(tu|vw)(-\frac{3}{2}m^2\Delta^2 - \frac{3}{4}m^2\Delta^3) - \frac{m^3}{8}(rs)(tu)(vw)\Delta^3\right\}$$

$$+\frac{1}{4!}\Sigma\epsilon_{rs}\epsilon_{tu}\epsilon_{vw}\epsilon_{xy}\left\{(rs|tu|vw|xy)(12m\Delta + 18m\Delta^2 + 12m\Delta^3 + 3m\Delta^4)\right.$$

$$+ (rs)(tu|vw|xy)(6m^2\Delta^2 + 6m^2\Delta^3 + 2m^2\Delta^4) + (rs|tu)(vw|xy)(3m^2\Delta^2 + 3m^2\Delta^3 + \frac{3}{4}m^2\Delta^4)$$

$$+(rs)(tu)(vw|xy)(\tfrac{3}{2}m^3\Delta^3+\tfrac{3}{4}m^3\Delta^4) + (rs)(tu)(vw)(xy)\frac{m^4}{16}\Delta^4\Big\} + \ldots\Big]\, G_\rho(\theta).$$

Then term by term comparison between two expansions for J, (3.19) and (3.26), gives $\partial_{rs}\Pr\{m \text{ tr } S_1\Lambda \leq 2\theta\}$, $\partial_{rs}\partial_{tu}\Pr\{m \text{ tr } S_1\Lambda \leq 2\theta\}$, etc., but in doing so we must take such a care that, for example,

$$\Sigma\, a_{ijk}\,\epsilon_i\epsilon_j\epsilon_k = \Sigma\, b_{ijk}\,\epsilon_i\epsilon_j\epsilon_k$$

implies $a_{ijk} = b_{ijk}$ if both a_{ijk} and b_{ijk} are completely symmetrical in their suffices. With this in mind and using the relation

$$\Delta\, G_\rho(\theta) = -\, E\, g_\rho(\theta),$$

where $g_\rho(\theta) = D\, G_\rho(\theta)$, we obtain

$$(3.27)\qquad \partial_{rs}\Pr\{m \text{ tr } S_1\Lambda \leq 2\theta\} = \frac{m}{2}(rs)\, E g_\rho(\theta),$$

$$(3.28)\qquad \partial_{rs}\partial_{tu}\Pr\{m \text{ tr } S_1\Lambda \leq 2\theta\} = -\Big\{\frac{m}{2}(rs\,|\,tu)(E^2+E)+\frac{m^2}{4}(rs)(tu)(E^2-E)\Big\}\, g_\rho(\theta)$$

$$(3.29)\qquad \partial_{rs}\partial_{tu}\partial_{vw}\Pr\{m \text{ tr } S_1\Lambda \leq 2\theta\} = \Big\{m(rs|tu|vw)(E^3+E^2+E)+\frac{m^2}{4}\Big[(rs)(tu|vw)$$
$$+(tu)(rs|vw)+(vw)(rs|tu)\Big](E^3-E)+\frac{m^3}{8}(rs)(tu)(vw)(E^3-2E^2+E)\Big\}$$
$$\cdot\, g_\rho(\theta),$$

$$(3.30)\qquad \partial_{rs}\partial_{tu}\partial_{vw}\partial_{xy}\Pr\{m \text{ tr } S_1\Lambda \leq 2\theta\} = -\Big\{m\Big[(rs|tu|vw|xy)+(rs|vw|xy|tu)$$
$$+(rs|xy|tu|vw)\Big](E^4+E^3+E^2+E)+\frac{m^2}{2}\Big[(rs)(tu|vw|xy)+(xy)(tu|vw|rs)$$
$$+(vw)(tu|xy|rs)+(tu)(vw|xy|rs)\Big](E^4-E)+\frac{m^2}{4}\Big[(rs|tu)(vw|xy)$$
$$+(rs|vw)(tu|xy)+(rs|xy)(tu|vw)\Big](E^4+E^3-E^2-E)+\frac{m^3}{8}\Big[(rs)(tu)(vw|xy)$$
$$+(rs)(vw)(tu|xy)+(rs)(xy)(tu|vw)+(tu)(vw)(rs|xy)+(tu)(xy)(rs|vw)$$
$$+(vw)(xy)(rs|tu)\Big](E^4-E^3-E^2+E)+\frac{m^4}{16}(rs)(tu)(vw)(xy)(E^4-3E^3+3E^2-E)\Big\}\, g_\rho(\theta).$$

Upon substituting (3.28) into (3.16) we obtain

$$h_1(\Lambda^{-1}) = \frac{1}{4n} \Sigma\, \sigma_{rs}\sigma_{tu} \left[2m(st|ur)\left\{ \frac{\theta^2}{\rho(\rho+1)} + \frac{\theta}{\rho} \right\} + m^2(st)(ur)\left\{ \frac{\theta^2}{\rho(\rho+1)} - \frac{\theta}{\rho} \right\} \right] .$$

Now $(st) = \mathrm{tr}\, \Lambda^{-1}_{st}\Lambda = \frac{1}{2} \sum_{i,j} (\delta_{si}\delta_{tj} + \delta_{ti}\delta_{sj})\sigma^{ji} = \frac{1}{2}(\sigma^{ts} + \sigma^{st}) = \sigma^{st}$

and also,

$$(st|ur) = \mathrm{tr}\, (\Lambda^{-1}_{st}\Lambda)(\Lambda^{-1}_{ur}\Lambda) = \frac{1}{2}(\sigma^{rs}\sigma^{tu} + \sigma^{su}\sigma^{tr}) .$$

Hence we have

$$\Sigma\, \sigma_{rs}\sigma_{tu}(st|ur) = \frac{1}{2}p(p+1)$$

and

$$\Sigma\, \sigma_{rs}\sigma_{tu}(st)(ur) = p.$$

We also note that $2\theta = \chi^2$, $\rho = mp/2$. Therefore we finally obtain, after some simplification,

$$(3.31) \qquad h_1(\Lambda^{-1}) = \frac{1}{4n}\left\{ \frac{p+m+1}{mp+2}\, \chi^4 + (p-m+1)\, \chi^2 \right\} .$$

In a similar way we substitute (3.29), (3.30) and (3.31) into (3.17) to evaluate $h_2(\Lambda^{-1})$. We note here that since $h_1(\Lambda^{-1})$ given by (3.31) is independent of Λ^{-1}, the terms involving $h_1^{(st)}(\Lambda^{-1})$ and $h_1^{(st,ur)}(\Lambda^{-1})$ in (3.17) do not appear. As before, it can be easily shown that

$$\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}(st|uv|wr) = \frac{1}{8}\, p(p^2+3p+4),\ \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}(st)(uv|wr)$$

$$= \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}(uv)(st|wr) = \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}(wr)(st|uv) = \frac{1}{2}p(p+1),$$

$$\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}(st)(uv)(wr) = p,\ \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st|ur|wx|yv)$$

$$= \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st|wx|yv|ur) = \frac{1}{4}\, p(p+1)^2,$$

$$\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st|yv|ur|wx) = \tfrac{1}{4}p(p+3), \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st)(ur|wx|yv)$$

$$= \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(yv)(ur|wx|st) = \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(wx)(ur|yv|st)$$

$$= \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(ur)(wx|yv|st) = \tfrac{1}{2}p(p+1),$$

$$\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st|ur)(wx|yv) = \tfrac{1}{2}p^2(p+1)^2, \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st|wx)(ur|yv)$$

$$= \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st|yv)(ur|wx) = \tfrac{1}{2}p(p+1),$$

$$\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st)(ur)(wx|yv) = \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(wx)(yv)(st|ur) = \tfrac{1}{2}p^2(p+1),$$

$$\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st)(wx)(ur|yv) = \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st)(yv)(ur|wx)$$

$$= \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(ur)(wx)(st|yv) = \Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(ur)(yv)(st|wx) = p.$$

and

$$\Sigma\, \sigma_{rs}\sigma_{tu}\sigma_{vw}\sigma_{xy}(st)(ur)(wx)(yv) = p^2.$$

Using these results we obtain from (3.17), after some simplification,

$$(3.32)\quad h_2(\Lambda^{-1}) = \frac{1}{48n^2}\Big[\frac{-6(p-1)(p+2)(m-1)(m+2)}{(mp+2)^2(mp+4)(mp+6)}\chi^8$$

$$+ \frac{4mp^3+2(3m^2+3m+10)p^2+2(2m^3+3m^2+17m+18)p+4(5m^2+9m+2)}{(mp+2)^2\,(mp+4)}\chi^6$$

$$+ \frac{13p^2+24p-11m^2+7}{mp+2}\chi^4 + \left\{7p^2+(-12m+12)p+(7m^2-12m+1)\right\}\chi^2\Big],$$

which is independent of Λ^{-1} just as $h_1(\Lambda^{-1})$.

Now we substitute (3.31) and (3.32) into (3.13) to obtain

$$(3.33)\quad T_0^2 = 2h(S_0) = 2\theta + 2h_1(S_0)+2h_2(S_0)+O(n^{-3}) = \chi^2+\frac{1}{2n}\left\{\frac{p+m+1}{mp+2}\chi^4+(p-m+1)\chi^2\right\}$$

$$+\frac{1}{24n^2}\Big\{\frac{6(p-1)(p+2)(m-1)(m+2)}{(mp+2)^2(mp+4)(mp+6)}\chi^8+\frac{4mp^3+2(3m^2+3m+10)p^2+2(2m^3+3m^2+17m+18)p+4(5m^2+9m+2)}{(mp+2)^2\,(mp+4)}\chi^6$$

$$+\frac{13p^2+24p-11m^2+7}{mp+2}\chi^4+\left[7p^2+(-12m+12)p+(7m^2-12m+1)\right]\chi^2\Big\} + O(n^{-3}),$$

which is the asymptotic expression of a percentage point of the T_0^2 distribution in terms of the corresponding percentage point of the X^2 distribution with mp degrees of freedom.

If we put m = 1 in (3.33), we have

$$(3.34)\quad T^2 = X^2 + \frac{1}{2n}\left\{X^4 + pX^2\right\} + \frac{1}{24n^2}\left\{4X^6 + (13p-2)X^4 + (7p^2-4)X^2\right\} + O(n^{-3}),$$

which is the asymptotic expression of a percentage point of the generalized Student T distribution. This result (3.34) was previously obtained by Hotelling and Frankel [2], [3].

4. Asymptotic formula for the c.d.f. of T_0^2.

Let $F(2\theta_1)$ be the c.d.f. of T_0^2, i.e.,

$$(4.1)\qquad F(2\theta_1) = \Pr\left\{m \operatorname{tr} S_1 S_0^{-1} \le 2\theta_1\right\} .$$

Then, as (3.6), we can write

$$(4.2)\quad \Pr\left\{m \operatorname{tr} S_1 S_0^{-1} \le 2\theta_1\right\} = \int_R \Pr\left\{m \operatorname{tr} S_1 S_0^{-1} \le 2\theta_1 \mid S_0\right\} \Pr\left\{dS_0\right\}$$

$$= \Theta \Pr\left\{m \operatorname{tr} S_1 \Lambda \le 2\theta_1\right\} ,$$

where Θ is given by (3.12). Upon substituting (3.28), (3.29) and (3.30) into (4.2) we obtain, after some simplification,

$$(4.3)\quad F(2\theta_1) = G_\rho(\theta_1) - \frac{1}{2n}\left\{\frac{2(p+m+1)\theta_1^2}{mp+2} + (p-m+1)\theta_1\right\} g_\rho(\theta_1)$$

$$- \frac{1}{48n^2}\Big[\frac{24\left\{mp^3+2(m^2+m+4)p^2+(m^3+2m^2+21m+20)p+8m^2+20m+20\right\}\theta_1^4}{(mp+2)(mp+4)(mp+6)}$$

$$+ \frac{4\left\{3mp^3-2(3m^2-3m-4)p^2-3(3m^3+2m^2+11m-4)p-40m^2-36m-4\right\}\theta_1^3}{(mp+2)(mp+4)}$$

$$- \frac{2\left\{3mp^3+2(3m^2+3m-4)p^2-3(3m^3-2m^2-5m+4)p-8m^2+12m+4\right\}\ \theta_1^2}{mp+2}$$

$$- \left\{3mp^3-2(3m^2-3m+4)p^2+3(m^3-2m^2+5m-4)p-8m^2+12m+4\right\}\ \theta_1 \ \rbrack g_\rho(\theta_1) + O(n^{-3}) \ ,$$

where $G_\rho(\theta_1) = \lbrack \Gamma(\rho) \rbrack^{-1} \int_0^{\theta_1} t^{\rho-1} e^{-t} dt$, $g_\rho(\theta_1) = \frac{\partial}{\partial \theta_1} G_\rho(\theta_1)$, and $\rho = mp/2$.

(4.3) is a sort of multivariate analogue of Hartley's formula of "Studentization." In fact it can be shown that when $p = 1$, (4.3) coincides with Hartley's formula for the c.d.f. of the univariate analysis of variance F statistic. (See equation (28), p. 178, [1].)

5. Discussion of the error and remarks.

In view of the methods used in sections 3 and 4, it is rather difficult to set a bound for the error committed by omitting all terms after the first few terms in the asymptotic formula for T_0^2 (3.33) or in the asymptotic formula for the c.d.f. of T_0^2 (4.3). There is, however, a method to find lower and upper bounds to the c.d.f. of T_0^2 which is fairly good for large values of n, and they can be used to set a bound for $O(n^{-3})$, say, in the asymptotic expansion of the c.d.f. of T_0^2.

We shall first obtain lower and upper bounds for the c.d.f. of T_0^2. It is well known (e.g. see [6]) that the joint probability law of the characteristic roots $e_1, e_2, \ldots, e_s$ of $m \operatorname{tr} S_1 S_0^{-1}$ under the null hypothesis H_0 is given by

$$(5.1)\quad P(e_1, e_2, \ldots, e_s) = C(s,t,p,n) \prod_{i=1}^{s} e_i^{\frac{t-s-1}{2}} \left(1+\frac{e_i}{n}\right)^{-\frac{m+n}{2}} de_i \prod_{i<j=1}^{s-1} (e_i - e_j),$$

where $0 \le e_s \le e_{s-1} \le \cdots \le e_1 < \infty$, $s = \min(p,m)$, $t = \max(p,m)$ and

$$C(s,t,p,n) = \frac{\pi^{\frac{s}{2}}}{n^{\frac{st}{2}}} \prod_{i=1}^{s} \frac{\Gamma\left\{\frac{1}{2}(n+t-p+i)\right\}}{\Gamma\left\{\frac{1}{2}(t-s+i)\right\}\ \Gamma\left\{\frac{1}{2}(n-p+i)\right\}\ \Gamma(\frac{1}{2})} \ .$$

The statistic T_0^2 is expressed as

(5.2) $$T_0^2 = m \operatorname{tr} S_1 S_0^{-1} = \sum_{i=1}^{s} e_i ,$$

and the c.d.f. of T_0^2 is given by

(5.3) $$F(2\theta_1) = C(s,t,p,n) \int \cdots \int_{R_1} \prod_{i=1}^{s} e_i^{\frac{t-s-1}{2}} (1+ \frac{e_i}{n})^{-\frac{m+n}{2}} de_i \prod_{i<j=1}^{s-1} (e_i - e_j),$$

where R_1 is the domain of integration such that $0 \leq e_s \leq e_{s-1} \leq \cdots \leq e_1 < \infty$ and $0 \leq \sum_{i=1}^{s} e_i \leq 2\theta_1$. Now for any non-negative values of e_i and n, the following inequality holds:

$$\log (1 + \frac{e_i}{n}) \leq \frac{e_i}{n}$$

for i = 1, ..., s, where equality holds when $e_i = 0$ or $n \longrightarrow \infty$. Hence we have

$$\prod_{i=1}^{s} (1 + \frac{e_i}{n})^{-\frac{m+n}{2}} \geq e^{-\frac{m+n}{2n} \sum_{i=1}^{s} e_i} .$$

Therefore the probability law (5.1) is bounded from below as follows:

(5.4) $$P_1(e_1,\ldots,e_s) \leq P(e_1,\ldots,e_s)$$

where

$$P_1(e_1,\ldots,e_s) = C(s,t,p,n) \prod_{i=1}^{s} e_i^{\frac{t-s-1}{2}} de_i \exp\left[- \frac{m+n}{2n} \sum_{i=1}^{s} e_i\right] \prod_{i<j=1}^{s-1} (e_i - e_j).$$

It must be noted here that $P_1(e_1,\ldots,e_s)$ is not a probability law although it is non-negative for all e_i such that $0 \leq e_s \leq \cdots \leq e_1 < \infty$. Now integrating both sides of (5.4) in R_1 we obtain

(5.5) $$F_1(2\theta_1) \leq F(2\theta_1) ,$$

where

$$F_1(2\theta_1) = C(s,t,p,n) \int \cdots \int_{R_1} \prod_{i=1}^{s} e_i^{\frac{t-s-1}{2}} de_i \exp\left[-\frac{m+n}{2n} \sum_{i=1}^{s} e_i\right] \prod_{i<j=1}^{s-1} (e_i - e_j),$$

and also integrating both sides of (5.4) in R_2 where $0 \le e_s \le \cdots \le e_1 < \infty$ and $2\theta_1 \le \sum_{i=1}^{s} e_i < \infty$ and subtracting each from 1, we have

$$(5.6) \qquad F(2\theta_1) \le F_2(2\theta_1),$$

where

$$F_2(2\theta_1) = 1 - C(s,t,p,n) \int \cdots \int_{R_2} \prod_{i=1}^{s} e_i^{\frac{t-s-1}{2}} de_i \exp\left[-\frac{m+n}{2n} \sum_{i=1}^{s} e_i\right] \prod_{i<j=1}^{s-1} (e_i - e_j).$$

In order to evaluate $F_1(2\theta_1)$ and $F_2(2\theta_1)$, we observe that as n tends to ∞, $T_0^2 = \sum_{i=1}^{s} e_i$ has a χ^2 distribution with st degrees of freedom in the limit, i.e., we have

$$(5.7) \quad K(s,t,p) \int \cdots \int_{R_1} \prod_{i=1}^{s} e_i^{\frac{t-s-1}{2}} de_i \exp\left[-\frac{1}{2} \sum_{i=1}^{s} e_i\right] \prod_{i<j=1}^{s-1} (e_i - e_j) = G_{\rho_1}(\theta_1),$$

where

$$K(s,t,p) = \lim_{n \to \infty} C(s,t,p,n) = \frac{\pi^{\frac{s}{2}}}{2^{\frac{st}{2}}} \frac{1}{\prod_{i=1}^{s} \Gamma\left\{\frac{1}{2}(t-s+i)\right\} \Gamma(\frac{i}{2})}$$

and $\rho_1 = st/2$. Hence integration of (5.5) yields

$$(5.8) \qquad F_1(2\theta_1) = L(s,t,p,n)\, G_{\rho_1}\left(\frac{m+n}{n} \theta_1\right),$$

where

$$L(s,t,p,n) = \frac{C(s,t,p,n)}{K(s,t,p)} \left(\frac{n}{m+n}\right)^{\frac{st}{2}} = \left(\frac{2}{m+n}\right)^{\frac{st}{2}} \prod_{i=1}^{s} \frac{\Gamma(\frac{n+t-p+i}{2})}{\Gamma(\frac{n-p+i}{2})} .$$

Similarly we obtain from (5.6)

$$(5.9) \qquad F_2(2\theta_1) = 1 - L(s,t,p,n) \left\{1 - G_{\rho_1}\left(\frac{m+n}{n} \theta_1\right)\right\} .$$

Now if we write (4.3) as

$$(5.10) \qquad F(2\theta_1) = a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + R_3 ,$$

where R_3 is the error committed by omitting all terms except the first three terms in the asymptotic series of $F(2\theta_1)$, the absolute value of R_3 has the following upper bound:

$$(5.11) \qquad |R_3| \leq \max \left\{ \left|F_1(2\theta_1) - a_0 - \frac{a_1}{n} - \frac{a_2}{n^2}\right|, \left|F_2(2\theta_1) - a_0 - \frac{a_1}{n} - \frac{a_2}{n^2}\right| \right\} ,$$

where $F_1(2\theta_1)$ and $F_2(2\theta_1)$ are given by (5.8) and (5.9), respectively.

The actual manner in which (3.33) converges to the true value T_0^2 or (4.3) to the true value $F(2\theta_1)$, is not known, but it is hoped that the use of the first few corrective terms may result in a test which is more accurate than the χ^2 approximation, at any rate for moderately large values of n. In the case of the asymptotic formula for the c.d.f. of T_0^2 (4.3) we may judge the magnitude of the error involved in using the first few terms of the series by (5.11), which turns out to be rather small numerically when n is sufficiently large.

The author wishes to express his indebtedness to Professor Harold Hotelling for suggesting this problem and for his guidance in the preparation of this paper.

REFERENCES

[1] H. O. Hartley, "Studentization, or the elimination of the standard deviation of the parent population from the random sample-distribution of statistics," Biometrika, Vol. 33 (1944), pp. 173-180.

[2] H. Hotelling, "A generalized T test and measure of multivariate dispersion," Proc. Second Berkeley Symposium on Math. Stat. and Probability, 1950, pp. 23-41.

[3] H. Hotelling and L. R. Frankel, "The transformation of statistics to simplify their distribution," Annals of Mathematical Statistics, Vol. 9 (1938), pp. 87-96.

[4] G. S. James, "The comparison of several groups of observations when the ratios of the population variances are unknown," Biometrika, Vol. 38 (1951), pp. 324-329.

[5] G. S. James, "Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown," Biometrika, Vol. 41 (1954), pp. 19-43.

[6] S. N. Roy, "p-statistics and some generalizations in analysis of variance appropriate to multivariate problems," Sankhya, Vol. 4 (1939), pp. 381-396.

[7] B. L. Welch, "The generalization of 'Student's' problem when several different population variances are involved," Biometrika, Vol. 34 (1947), pp. 28-35.