A REMARK ON WALD'S PAPER: "ON A STATISTICAL PROBLEM ARISING IN THE CLASSIFICATION OF AN INDIVIDUAL INTO ONE OF TWO GROUPS

by

Junjiro Ogawa

This investigation was done at the Institute of Statistics, University of North Carolina, Chapel Hill, under the sponsorship of the Office of Naval Research through Contract NR 042031.

Institute of Statistics
Mimeograph Series No. 143
January, 1956

A REMARK ON WALD'S PAPER: "ON A STATISTICAL PROBLEM ARISING IN THE CLASSIFICATION OF AN INDIVIDUAL INTO ONE OF TWO GROUPS"[1]

by Junjiro Ogawa

The Institute of Statistics, University of North Carolina and
The Department of Mathematics, Faculty of Science, Osaka University

Summary.

In 1944, the late Professor A. Wald proposed a statistic for use in classification procedure in the above mentioned paper [1]. That statistic U is as follows:

$$U = \sqrt{\frac{N_1 + N_2}{N_1 N_2}} \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} z_i z_j^* \tag{1}$$

where $\underline{z}' = (z_1, \ldots, z_p)$ and $\underline{z}^{*\prime} = (z_1^*, \ldots, z_p^*)$ are distributed independently of each other according to the non-singular p-dimensional normal distributions with the same covariance matrix $\|\sigma_{ij}\|$ and means

$$\mathcal{E}(\underline{z}) = \underline{\mu} \text{ or } \underline{\nu}, \quad \mathcal{E}(\underline{z}^*) = \sqrt{\frac{N_1+N_2}{N_1 N_2}}\,(\underline{\nu}-\underline{\mu})$$

respectively, and putting as usual

$$\|s_{ij}\|^{-1} = \|s^{ij}\| ,$$

s_{ij}, $i \leq j$, $i, j = 1, \ldots, p$ are distributed according to the Wishart distribution with the same $\|\sigma_{ij}\|$ as the population dispersion matrix, and with degrees of freedom $n = N_1 + N_2 - 2$. Further $\underline{z}$, $\underline{z}^*$ and $\|s_{ij}\|$ are mutually independent in the stochastic sense.

---

1. This investigation was done at the Institute of Statistics, University of North Carolina, Chapel Hill, under the sponsorship of the Office of Naval Research through Contract NR 042031. The author expresses here his hearty thanks to Professor Harold Hotelling by whom this chance was given to him.

Wald considered the exact sampling distribution of U, and taking account of the fact that the distribution of U is essentially the same as that of

$$V = \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} t_{i,n+1}\, t_{j,n+2}\,, \tag{2}$$

where, of course,

$$\|s^{ij}\| = \|s_{ij}\|^{-1}, \quad s_{ij} = \frac{1}{n} \sum_{\alpha=1}^{n} t_{i\alpha}\, t_{j\alpha}, \ i,j = 1,2,\ldots,p$$

and the probability element of the joint distribution of $t_{i,\alpha}$, $i = 1, \ldots, p$; $\alpha = 1, \ldots, n+2$ is given by

$$\frac{1}{(2\pi)^{\frac{p(n+2)}{2}}} \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{p}\sum_{\alpha=1}^{n} t_{i\alpha}^2 + \sum_{i=1}^{p}(t_{i,n+1}-\rho_i)^2 + \sum_{i=1}^{p}(t_{i,n+2}-\zeta_i)^2\right\}\right]$$

$$\cdot \prod_{i=1}^{p} \prod_{\alpha=1}^{n+2} dt_{i\alpha}, \tag{3}$$

where $\underline{\rho}' = (\rho_1,\ldots,\rho_p)$ and $\underline{\zeta}' = (\zeta_1,\ldots,\zeta_p)$ are certain known functions of $\underline{\mu}$ and $\underline{v}$ respectively, he concluded that the distribution of V is the same as that of

$$-n \frac{m_3}{m_3^2 - (1-m_1)(1-m_2)} \tag{4}$$

calculated from the joint probability distribution of m_1, m_2, m_3 which has the following probability element:

$$\text{const.} \exp\left\{-\frac{1}{2}\left(m_1 \sum_{i=1}^{p} \rho_i^2 + 2m_3 \sum_{i=1}^{p} \rho_i \zeta_i + m_2 \sum_{i=1}^{p} \zeta_i^2\right)\right\} K(m_1,m_2,m_3) \frac{1}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}} \tag{5}$$

$$\times F_p(m_1) F_p(m_2) \varnothing_p\left(\frac{m_3}{\sqrt{m_1 m_2}}\right) F_{n+2-p}(1-m_1) F_{n+2-p}(1-m_2) \varnothing_{n+2-p}\left(\frac{-m_3}{\sqrt{(1-m_1)(1-m_2)}}\right) dm_1 dm_2 dm_3$$

in the domain $0 \leq m_1 \leq 1$, $0 \leq m \leq 1$, $-\sqrt{m_1 m_2} < m_3 \leq \sqrt{m_1 m_2}$ and 0 otherwise, where

$$F_k(t) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} t^{\frac{k}{2}-1} e^{-\frac{t}{2}}, \quad \emptyset_k(t) = \frac{\Gamma(\frac{k}{2})}{\sqrt{\pi}\,\Gamma(\frac{k-1}{2})} (1 - t^2)^{\frac{k-3}{2}} \tag{6}$$

and

$$K(m_1, m_2, m_3) = \mathcal{E}\left\{ \left| \begin{vmatrix} r_{11} & \cdots & r_{1p} \\ \cdot & \cdots & \cdot \\ r_{p1} & \cdots & r_{pp} \end{vmatrix}^{\frac{n+2-p}{2}} \right| u, v \right\}. \tag{7}$$

The relation (7) follows from Wald's Lemma 8.

Wald's proof of this result just described was divided into nine lemmas and each lemma was proved by an ingenious method. The author of this note believes that the key point of Wald's proof lies in his Lemma 6, which states that it is sufficient for the calculation of the sampling distribution of $\overline{V}$ which is the same as that of V to start from the conditional distribution of random 2-frame (u, v) under the condition that the random p-plane L_p is fixed. But the proof of Lemma 6 seems to be incomprehensive at least for the author himself, because Wald had not used explicitly the concept of the invariant measure defined on the Grassmann manifold consisting of p-planes in (n+2)-dimensional euclidean space [2]. Consequently the proof of Lemma 7 was also insufficient.

The author shall attempt new proofs of these two lemmas as an example of applications of the theory of invariant measure defined on the analytic manifolds such as orthogonal group, Stiefel, and Grassmann manifold [3].

Explanation of the Problem and Proofs of Lemma 6 and 7.

Our purpose is the new proofs of Wald's Lemma 6 and 7, but for the sake of readers' convenience, we shall describe the problem and follow Wald's proofs and explain the necessary notations.

Mathematically speaking our problem which requires to find the sampling distribution of the classification statistic U is as follows: i.e.

Let the probability element of the joint distribution of $p(n+2)$ variates $t_{i\alpha}$; $i = 1, \ldots, p$; $\alpha = 1, \ldots, n+2$ be

$$\frac{1}{(2\pi)^{\frac{p(n+2)}{2}} \sigma^{n+2}} \exp \left[-\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} \left\{ \sum_{\alpha=1}^{n} t_{i\alpha} t_{j\alpha} + (t_{i,n+1} - \xi_i)(t_{j,n+1} - \xi_j) \right.\right.$$

$$\left.\left. + (t_{i,n+2} - \eta_i)(t_{j,n+2} - \eta_j) \right\} \right] \prod_{i=1}^{p} \prod_{i=1}^{n+2} dt_{i\alpha} , \tag{8}$$

where of course $\|\sigma_{ij}\|$ is non-singular and symmetric variance-covariance matrix of the order $p \times p$ and

$$\|\sigma^{ij}\| = \|\sigma_{ij}\|^{-1}$$

and

$$\sigma^2 = \text{det.} \|\sigma_{ij}\| ,$$

then we are to find the probability distribution of the statistic

$$V = \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} t_{i,n+1} t_{j,n+2} , \tag{9}$$

where

$$\|s^{ij}\| = \|s_{ij}\|^{-1} \text{ and } s_{ij} = \frac{1}{n} \sum_{\alpha=1}^{n} t_{i\alpha} t_{j\alpha}, \quad i, j = 1, \ldots, p .$$

We shall sketch Wald's reasonings and results obtained up to his Lemma 6 briefly.

The statistic V given by (9) is invariant under the non-singular p-dimensional linear transformations of $(t_{1\alpha}, \ldots, t_{p\alpha})$, $\alpha=1, \ldots, n+2$, so we can take, without any loss of generality, that $\|\sigma_{ij}\| = I$ (unit matrix of p dimensions) in (8). In other words, in considering the probability distribution of V, the probability element of $t_{i\alpha}$, $i = 1, \ldots, p$; $\alpha = 1, \ldots, n+2$ may be taken as

$$\frac{1}{(2\pi)^{\frac{p(n+2)}{2}}} \exp\left[-\frac{1}{2}\left\{ \sum_{i=1}^{p}\sum_{\alpha=1}^{n} t_{i\alpha}^2 + \sum_{i=1}^{p}(t_{i,n+1}-\rho_i)^2 + \sum_{i=1}^{p}(t_{i,n+2}-\zeta_i)^2 \right\} \right]$$

$$\times \prod_{i=1}^{p}\prod_{\alpha=1}^{n+2} dt_{i\alpha} \,. \tag{10}$$

This was the result stated in Lemma 1 and 2.

We take two sets of real numbers

$$(u_1, \ldots, u_{n+2}) \quad \text{and} \quad (v_1, \ldots, v_{n+2})$$

satisfying the restrictions

$$\sum_{\alpha=1}^{n+2} u_\alpha^2 = \sum_{\alpha=1}^{m+2} v_\alpha^2 = 1, \quad \sum_{\alpha=1}^{n+2} u_\alpha v_\alpha = 0, \tag{11}$$

and put

$$\bar{t}_{i,n+1} = \sum_{\alpha=1}^{n+2} t_{i\alpha}u_\alpha, \quad \bar{t}_{i,n+2} = \sum_{\alpha=1}^{n+2} t_{i\alpha}v_\alpha \,. \tag{12}$$

Hereafter we shall stand on the geometrical point of view. In (n+2)-dimensional euclidean space with the origin $O = (0, \ldots, 0)$, we put

$R_i = (t_{i1}, \ldots, t_{i,n+2})$, $i = 1, \ldots, p$,

$P = (u_1, \ldots, u_{n+2})$ and $Q = (v_1, \ldots, v_{n+2})$

with reference to a certain fixed orthogonal coordinates system, then the vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ are two unit vectors orthogonal to each other and the system of such vectors $\underline{u} = \overrightarrow{OP}$ and $\underline{v} = \overrightarrow{OQ}$ is called a "2-frame" in algebraic geometry. $\bar{t}_{i,n+1}$ and $\bar{t}_{i,n+2}$ defined in (12) are the orthogonal projections of $\underline{t}_i = \overrightarrow{OR_i}$ on $\underline{u}$ and $\underline{v}$ respectively.

Let the orthogonal projection of the point R_i into the orthogonal complement of the linear subspace spanned by $\underline{u}$ and $\underline{v}$ be $\bar{R}_i$ with the coordinates $(r_{i1}, \ldots, \ldots, r_{i,n+2})$, and put

$$\bar{s}_{ij} = \frac{1}{n} \sum_{\alpha=1}^{n+2} r_{i\alpha} r_{j\alpha} \tag{13}$$

and

$$\bar{V} = \sum_{i=1}^{p} \sum_{j=1}^{p} \bar{s}^{ij}\, \bar{t}_{i,n+1}\, \bar{t}_{j,n+2}\,, \tag{14}$$

where as usual

$$\|\bar{s}^{ij}\| = \|\bar{s}_{ij}\|^{-1}$$

then $\bar{s}_{ij}$, $i,j = 1,\ldots,p$ are invariant against the rotations of coordinate axes. Therefore, if we rotate the coordinate axes such that the $(n+1)$-th and the last $(n+2)$-th axis will coincide with the vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ respectively, then the coordinates of $\bar{R}_i$ with reference to this new coordinates system will be of the form $(\bar{t}_{i,1}, \ldots, \bar{t}_{i,n}, 0, 0)$ and consequently

$$\bar{s}_{ij} = \frac{1}{n} \sum_{\alpha=1}^{n} \bar{t}_{i\alpha}\, \bar{t}_{j\alpha} \tag{15}$$

Lemma 3 states that if we calculate the distribution of $\overline{V}$ given by (14) from the joint probability distribution of $t_{i\alpha}$, $i = 1,\ldots,p$; $\alpha = 1, \ldots, n+2$, which has the probability element

$$\frac{1}{(2\pi)^{\frac{p(n+2)}{2}}} \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{p}\sum_{\alpha=1}^{n+2}(t_{i\alpha}-\rho_i u_\alpha-\zeta_i v_\alpha)^2\right\}\right]\prod_{i=1}^{p}\prod_{\alpha=1}^{n+2} dt_{i\alpha}, \tag{16}$$

where the 2-frame $(\underset{\sim}{u},\underset{\sim}{v})$ is fixed, then the distribution of $\overline{V}$ is independent of $(\underset{\sim}{u},\underset{\sim}{v})$ and is the same as the required distribution of V.

The 2-frame $(\underset{\sim}{u},\underset{\sim}{v})$ may be considered as a point of the algebraic surface defined by (11) in the $(2n+4)$-dimensional euclidean space, and this $(2n+1)$-dimensional algebraic surface is called the Stiefel manifold, which we shall denote by $V_{2,n}$. By Lemma 3 we know that the distribution of $\overline{V}$ is uniform, so to speak, on the Stiefel manifold $V_{2,n}$.

Taking account of the fact that the distribution given by (16) is invariant under the $(n+2)$-dimensional orthogonal transformations, we introduce an invariant probability measure on $V_{2,n}$ (which is essentially the Haar measure on $V_{2,n}$, so it exists and is unique), and denote it by

$$D^{-1}dS = D^{-1}\prod_{j=1}^{n} \underset{\sim}{d}_j' d\underset{\sim}{u}\, \underset{\sim}{d}_j' d\underset{\sim}{v}, \tag{17}$$

where D is constant and $\underset{\sim}{d}_1, \ldots, \underset{\sim}{d}_n$ is a n-frame in the orthogonal complement of the linear sub-space spanned by the 2-frame $(\underset{\sim}{u},\underset{\sim}{v})$ and the product means the exterior product of differentials (see the proof of Lemma 6 later). Then the distribution of $\overline{V}$ is the same as before if it is calculated from

$$\frac{1}{(2\pi)^{p(n+2)/2}\,D} \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{p}\sum_{\alpha=1}^{n+2}(t_{i\alpha}-\rho_i u_\alpha-\zeta_i u_\alpha)^2\right\}\right]\prod_{i=1}^{p}\prod_{\alpha=1}^{n+2} dt_{i\alpha}\, dS, \tag{18}$$

and the distribution (18) is invariant under (n+2)-dimensional orthogonal transformations. This was the result of Lemma 5.

Lemma 4 and 4' give the geometrical interpretation of the statistic $\overline{V}$ in (14): Let the linear sub-space which is spanned by p vectors $\overrightarrow{OR_1}, \ldots, \overrightarrow{OR_p}$ be L_p, which is called "p-plane" in (n+2)-dimensional euclidean space, and let the orthogonal projections of P and Q into L_p be $\overline{P}$ and $\overline{Q}$ respectively, and put

$$< PO\overline{P} = \theta_1, < PO\overline{Q} = \theta_1'; < QO\overline{Q} = \theta_2, < QO\overline{P} = \theta_2', < \overline{P}O\overline{Q} = \theta_3 \tag{19}$$

and further put

$$\begin{aligned} &a_1 = \cos^2\theta_1,\ a_2 = \cos\theta_1' \cos\theta_2,\ b_1 = \cos\theta_1 \cos\theta_2',\ b_2 = \cos^2\theta_2 \\ &a_{11} = \frac{1}{n}(\cos^2\theta_1 - a_1^2 - b_1^2), \quad a_{22} = \frac{1}{n}(\cos^2\theta_2 - a_2^2 - b_2^2) \\ &a_{12} = \frac{1}{n}(\cos\theta_1 \cos\theta_2 \cos\theta_3 - a_1a_2 - b_1b_2), \end{aligned} \tag{20}$$

then $\overline{V}$ can be written as

$$\overline{V} = \begin{vmatrix} 0 & a_1 & a_2 \\ b_1 & a_{11} & a_{12} \\ b_2 & a_{12} & a_{22} \end{vmatrix} \Bigg/ \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} . \quad \text{(Lemma 4)}$$

In other words, the statistic $\overline{V}$ depends only on the relative position of L_p and the 2-frame $(\underline{u}, \underline{v})$. Further $\overline{V}$ may be expressed as a function of θ_1, θ_2, θ_3 only (Lemma 4'), i.e.,

$$\overline{V} = -n \frac{m_3}{m_3^2 - (1-m_1)(1-m_2)} \tag{21}$$

where $m_1 = \cos^2\theta_1$, $m_2 = \cos^2\theta_2$, $m_3 = \cos\theta_1 \cos\theta_2 \cos\theta_3$.

Now we came to Lemma 6. Lemma 6 states that for the calculation of the sampling distribution of $\bar{V}$ we can start from the conditional marginal distribution of $\underline{u}$, $\underline{v}$ which is derived from (18) under the condition that the p-plane L_p was fixed.

Proof of Lemma 6.

We shall arrange the p(n+2) variates $t_{i\alpha}$ in the following matrix form, i.e.,

$$T = \begin{bmatrix} t_{11} & t_{21} & \cdots & t_{p1} \\ t_{12} & t_{22} & \cdots & t_{p2} \\ \vdots & \vdots & & \vdots \\ t_{1n+2} & t_{2n+2} & \cdots & t_{pn+2} \end{bmatrix} = [\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_p] \text{ (say)} \qquad (22)$$

N w since we have assumed that the joint probability distribution of $t_{i\alpha}$ $i = 1, \ldots, p$; $\alpha = 1, \ldots, n+2$ is continuous, the probability of the set of $t_{i\alpha}$ in p(n+2)-dimensional space for which p column vectors $\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_p$ are linearly dependent is zero, so the p x p matrix T'T is symmetric and positive definite with probability one. If we choose an orthogonal matrix G of the type p x p appropriately, it can be put

$$G'T'TG = L^2 , \qquad (23)$$

where L is a diagonal matrix

$$L = \begin{bmatrix} \ell_1 & & & 0 \\ & \ell_2 & & \\ & & \ddots & \\ 0 & & & \ell_p \end{bmatrix} , \qquad (24)$$

and we may consider that $\ell_p > \ell_{p-1} > \dots > \ell_1 > 0$ with probability one. Further if we confine ourselves such that all the elements of the first row of the orthogonal matrix G are positive, i.e.,

$$g_{1j} > 0, \; j = 1, \dots, p,$$

then the orthogonal matrix G satisfying the relation (23) is determined uniquely with probability one.

Put

$$A \equiv TGL^{-1} \tag{25}$$

then it is clear that

$$A' A = I_p \tag{26}$$

i.e., A is a p-frame of (n+2)-dimensions.

From (25) we have the following factorization of T in the unique manner; i.e,

$$T = A \, L \, G' \tag{27}$$

In other words, p(n+2)-dimensional euclidean space may be considered as the direct product space of Stiefel manifold $V_{p,n+2-p}$ and p-dimensional simplex L, and p-dimensional orthogonal group manifold. Elementarily speaking, since p-frame A has $p(n+2)-\frac{p(p+1)}{2}$ independent variables, the orthogonal matrix G has $p^2 - \frac{p(p+1)}{2}$ independent variables and simplex L has p independent variables, the matrix equation (27) may be considered as the transformation of the original independent variables $t_{i\alpha}$, $i = 1, \dots, p$; $\alpha = 1, \dots, n+2$ into new independent variables.

If we do not want to specify the independent variables explicitly in the change of variables, it will be of great convenience to make use of the "exterior products" of differentials. The differential of a function $f(u_1, \dots, u_n)$ of independent variables $u_1, \ldots, u_n$ is defined as

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial u_i} du_i , \tag{28}$$

and the exterior product of differentials is defined as the product which satisfies the associative law and anti-commutative; i.e.,

$$df\, dg = -dg\, df, \text{ and in particular } df\, df = 0. \tag{29}$$

It is easily seen that if the components of a column vector $d\underset{\sim}{x}$ are the linear combinations of the differentials $du_1, \ldots, du_n$; i.e., in matrix notation

$$d\underset{\sim}{x} = J\, d\underset{\sim}{u},$$

then we have

$$\prod_{i=1}^{n} dx_i = \det. J \prod_{i=1}^{n} du_i , \tag{30}$$

where the products of both sides mean the exterior products of differentials [4].

Taking differentials of both sides of (27), we have

$$dT = dA\, L\, G' + A dL\, G' + A\, L\, dG' . \tag{31}$$

Since A was a p-frame, we can choose a matrix B of the type $(n+2) \times (n+2-p)$ such that $[A : B]$ is a $(n+2)$-dimensional orthogonal matrix, in other words, B is a $(n+2-p)$-frame which lies entirely in the orthogonal complement of the linear sub-space which is spanned by the p-frame A. Premultiplying $\begin{bmatrix} A' \\ \cdots \\ B' \end{bmatrix}$ and post-multiplying G into both sides of (31), we have

$$\begin{bmatrix} A' \\ \cdots \\ B' \end{bmatrix} dTG = \begin{bmatrix} A'dAL \\ \cdots \\ B'dAL \end{bmatrix} + \begin{bmatrix} dL \\ \cdots \\ 0 \end{bmatrix} + \begin{bmatrix} LdG'G \\ \cdots \\ 0 \end{bmatrix} = \begin{bmatrix} A'dAL + dL - LG'dG \\ \cdots \\ B'dAL \end{bmatrix} \tag{32}$$

Here we have made use of the fact that from $G'G = L_p$ follows

$$dG'G = -G'dG$$

The exterior product of the left hand side of (32) is seen by use of (30), to be

$$\det \left[\begin{bmatrix} A' \\ \cdots \\ B' \end{bmatrix}^{p} G'^{\,n+2} \right] \prod_{i=1}^{p} \prod_{\alpha=1}^{n+2} dt_{i\alpha} \tag{33}$$

On the other hand the elements of the matrix $A'dA\,L + dL - L\,G'\,dG$ are as follows: i.e., let the p column vectors of A and G be $\underline{a}_i$ and $\underline{g}_i$, $i = 1, \ldots, p$ respectively, then, because of the facts that $\underline{a}_i'\,d\underline{a}_i = 0$, $\underline{g}_i'\,d\underline{g}_i = 0$, it follows that

(i,i) element is $\quad d\ell_i$

and for $i \neq j$ we have as

(i,j) element $\quad \underline{a}_i' d\underline{a}_j \ell_j - \ell_i \underline{g}_i' d\underline{g}_j$

and (j,i) element $\quad \underline{a}_j' d\underline{a}_i \ell_i - \ell_j \underline{g}_j' d\underline{g}_i$.

Therefore the exterior product of (i,j) element and (j,i) element is

$$(\underline{a}_i' d\underline{a}_j \ell_j - \ell_i \underline{g}_i' d\underline{g}_j)(\underline{a}_j' d\underline{a}_i \ell_i - \ell_j \underline{g}_j'\, d\underline{g}_i)$$

$$= \underline{a}_i' d\underline{a}_j \ell_j \underline{a}_j' d\underline{a}_i \ell_i - \ell_i \underline{g}_i' d\underline{g}_j \underline{a}_j' d\underline{a}_i \ell_i - \underline{a}_i' d\underline{a}_j \ell_j^2 \underline{g}_j' d\underline{g}_i + \ell_i \underline{g}_i' d\underline{g}_j \ell_j \underline{g}_j' d\underline{g}_i$$

$$= (\ell_j^2 - \ell_i^2)\, \underline{a}_i' d\underline{a}_j\; \underline{g}_i' d\underline{g}_j \;. \tag{34}$$

Here we have made use of the facts that for $i \neq j$

$$\underline{a}_j'd\underline{a}_i = -\underline{a}_i'd\underline{a}_j$$

$$\underline{g}_j'd\underline{g}_i = -\underline{g}_i'd\underline{g}_j$$

and assume that $j > i$. It follows that the exterior product of the elements of the matrix $A'\,dA\,L + dL - LG'\,dG$ is

$$\prod_{i<j} (\ell_j^2 - \ell_i^2) \prod_{i=1}^{p} d\ell_i \prod_{i<j} \underline{a}_i'\,d\underline{a}_j \prod_{i<j} \underline{g}_i'\,d\underline{g}_j \,. \tag{35}$$

In a similar manner we have the exterior product of the elements of the matrix $B'\,dA\,L$, i.e., put the column vector of B be $\underline{b}_j$, $j = 1, 2, \ldots, n+2-p$, this is

$$(\det.L)^{n+2-p} \prod_{j=1}^{n+2-p} \prod_{i=1}^{p} \underline{b}_j'd\underline{a}_j. \tag{36}$$

Consequently we have the following relation in their absolute values or in considering it as the relation of volume elements

$$\prod_{i=1}^{p} \prod_{\alpha=1}^{n+2} dt_{i\alpha} = (\prod_{1}^{p} \ell_i)^{n+2-p} \prod_{i<j}(\ell_j^2 - \ell_i^2) \prod_{i=1}^{p} d\ell_i \prod_{i<j} \underline{g}_i'd\underline{g}_j \prod_{i<j} \underline{a}_i'd\underline{a}_j$$
$$\cdot \prod_{j=1}^{n+2-p} \prod_{i=1}^{p} \underline{b}_j'\,d\underline{a}_i \,. \tag{37}$$

Now from the general theory of manifolds we know that $\prod_{i<j} \underline{g}_i'd\underline{g}_j$ represents the Haar measure on the p-dimensional orthogonal group manifold and $\prod_{i<j} \underline{a}_i'd\underline{a}_j \cdot \prod_{j=1}^{n+2-p} \prod_{i=1}^{p} \underline{b}_j'\,d\underline{a}_i$ represents the Haar measure on the Stiefel manifold $V_{p,n+2-p}$.

Let the p-frame which specifies the p-plane L_p be H (i.e. if we consider the p-plane L_p as an element of Grassmann manifold and denote it by $\mathfrak{f}$, then H is an analytic function of $\mathcal{J}$ at least locally) and C be a p-dimensional orthogonal matrix, then

$$A = HC \tag{38}$$

is a p-frame lying in L_p. Taking differentials of (38), we have

$$dA = H.dC + dH\,C \ . \tag{39}$$

Premultiplying B' and A' into both sides of (39), we have the relations

$$\begin{aligned} B'dA &= B'\,dH\,C \\ A'dA &= AC'H'dA = C'\,dC + C'H'\,dH\,C \end{aligned} \quad . \tag{40}$$

From (40) it follows that

$$\prod_{j=1}^{n+2-p} \prod_{i=1}^{p} \underline{b}_j' d\underline{a}_i \prod_{i<j} \underline{a}_i' d\underline{a}_j = \prod_{j=1}^{n+2-p} \prod_{i=1}^{p} \underline{b}_j' d\underline{h}_i \prod_{i<j} \underline{c}_i' d\underline{c}_j \ . \tag{41}$$

This means that the Haar measure on the Stiefel manifold $V_{p,n+2-p}$ decomposes into the product of the Haar measure on the Grassmann manifold $\prod_j \prod_i \underline{b}_j'\, d\underline{h}_i$ and that on the orthogonal group manifold $\prod_{i<j} \underline{c}_i' d\underline{c}_j$. Hence we have the following decomposition of the volume element of p(n+2)-dimensional euclidean space

$$\prod_{i=1}^{p} \prod_{\alpha=1}^{n+2} dt_{i\alpha} = \left(\prod_{i=1}^{p} \ell_i\right)^{n+2-p} \prod_{i<j} (\ell_j^2 - \ell_i^2) \prod_{i=1}^{p} d\ell_i \prod_{i<j} \underline{g}_i' d\underline{g}_j$$

$$\cdot \prod_{j=1}^{n+2-p} \prod_{i=1}^{p} \underline{b}_j' d\underline{h}_i \prod_{i<j} \underline{c}_i' d\underline{c}_j \ . \tag{42}$$

Now put the linkage factor between T and $\underset{\sim}{u},\underset{\sim}{v}$

$$\underset{\sim}{\Phi} = \text{const. } \exp \left[- \frac{1}{2} \psi \, (L,G,C,H;\ \underset{\sim}{u}\underset{\sim}{v};\ \underset{\sim}{\rho},\ \underset{\sim}{\zeta}) \right], \tag{43}$$

then we have

$$\psi(L.G.C.H;\underset{\sim}{u},\underset{\sim}{v};\underset{\sim}{\rho},\underset{\sim}{\zeta}) = \sum_{i=1}^{p} \sum_{\alpha=1}^{n+2} (t_{i\alpha} - \rho_i u_\alpha - \zeta_i v_\alpha)^2$$

$$= \sum_{i=1}^{p} (t_i - \rho_i \underset{\sim}{u} - \zeta_i \underset{\sim}{v})'(t_i - \rho_i \underset{\sim}{u} - \zeta_i \underset{\sim}{v}) = \operatorname{tr} \quad (T - \underset{\sim}{u}\underset{\sim}{\rho}' - \underset{\sim}{v}\underset{\sim}{\zeta}')'(T - \underset{\sim}{u}\underset{\sim}{\rho}' - \underset{\sim}{v}\underset{\sim}{\zeta}')$$

$$= \operatorname{tr}(L^2) - 2\operatorname{tr} \left\{ GLC(H'u\underset{\sim}{\rho}' + H'v\underset{\sim}{\zeta}') \right\} + \underset{\sim}{\rho}'\underset{\sim}{\rho} + \underset{\sim}{\zeta}'\underset{\sim}{\zeta} \ . \tag{44}$$

From (44) it is easily seen that if we put

$$\tilde{H} = OH,\ \tilde{\underset{\sim}{u}} = O\underset{\sim}{u},\ \hat{\underset{\sim}{v}} = O\underset{\sim}{v},$$

where O is an orthogonal matrix of order n+2, then

$$\psi(L,G,\tilde{H};\ \tilde{\underset{\sim}{u}},\ \tilde{\underset{\sim}{v}};\ \underset{\sim}{\rho},\underset{\sim}{\zeta}) = \psi(L,G,H;\ \underset{\sim}{u},\underset{\sim}{v};\ \underset{\sim}{\rho},\underset{\sim}{\zeta}). \tag{45}$$

The marginal distribution of the random p-plane derived from (18) is independent of H, so the conditional distribution of $\underset{\sim}{u}$, $\underset{\sim}{v}$ derived from (18) under the condition that L_p is fixed depends only on the relative positions of $\underset{\sim}{u},\underset{\sim}{v}$ and H, i.e.,

$$H'\underset{\sim}{u} \text{ and } H'\underset{\sim}{v} \ .$$

If we complete H to an orthogonal matrix of (n+2)-dimensions by adjoining $F(\overline{n+2} \times \overline{n+2-p})$ and put

$$\begin{bmatrix} H' \\ \cdots \\ F' \end{bmatrix} \underset{\sim}{u} = \begin{bmatrix} \underset{\sim}{u}_1^* \\ \\ \underset{\sim}{u}_2^* \end{bmatrix} \equiv \underset{\sim}{u}^* , \quad \begin{bmatrix} H' \\ \cdots \\ F' \end{bmatrix} \underset{\sim}{v} = \begin{bmatrix} \underset{\sim}{v}_1^* \\ \\ \underset{\sim}{v}_2^* \end{bmatrix} \equiv \underset{\sim}{v}^*,$$

then $\underline{u}^*$, $\underline{v}^*$ is a point in $V_{2,n}$ and because of the invariance of dS we have

$$dS = dS^* ,$$

and consequently we have the conditional distribution of $\underline{u}^*$, $\underline{v}^*$ under the condition that L_p is fixed as follows: i.e.,

$$\phi(\underline{u}_1^*, \underline{v}_1^*)\, dS^*. \tag{46}$$

Because of the facts that $\cos\theta_1 = |\underline{u}_1^*|$, $\cos\theta_2 = |\underline{v}_1^*|$ and $\cos\theta_3 = \underline{u}_1^{*\prime}\, \underline{v}_1^* / |\underline{u}_1^*|\, |\underline{v}_1^*|$, the statistic $\bar{V}$ depends on $\underline{u}_1^*$, $\underline{v}_1^*$ only, hence Lemma 6 follows.

Lemma 7 states that the probability element of the conditional distribution of u, v under the condition that $H = M_p$, where M_p is the p-frame consisting of p unit vectors along the first p axes of the coordinates system, is proportional to

$$\exp \left[-\frac{1}{2} \sum_{\gamma=p+1}^{n+2} \sum_{i=1}^{p} (\rho_i u_\gamma + \zeta_i v_\gamma)^2 \right] f(\underline{u},\underline{v})dS , \tag{47}$$

where

$$f(\underline{u},\underline{v}) = \mathcal{E} \left\{ \left| \begin{matrix} r_{11} & \cdots & r_{1p} \\ \cdot & \cdots & \cdot \\ r_{1p} & \cdots & r_{pp} \end{matrix} \right|^{\frac{n+2-p}{2}} \middle|\; \underline{u},\underline{v} \right\} , \tag{48}$$

and

$$r_{ij} = \sum_{\alpha=1}^{p} t_{i\alpha} t_{j\alpha}, \quad i,j = 1, \ldots, p . \tag{49}$$

Proof of Lemma 7.

We shall calculate the probability element of the probability distribution of $\underline{u},\underline{v}$, when the p-plane L_p is fixed to M_p from (18). In this case

$$H = \begin{bmatrix} 1 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} ,$$

therefore, because of

$$A = HC, \quad T = ALG' = HCLG'$$

$$H'T = H'HCLG' = CLG',$$

we have

$$CLG' = \begin{vmatrix} t_{11} & \cdots & t_{p1} \\ \vdots & & \vdots \\ t_{1p} & \cdots & t_{pp} \end{vmatrix} , \tag{50}$$

and consequently we have

$$GLC'CLG' = GL^2G' = (r_{ij}) , \quad i,j = 1, \ldots, p . \tag{51}$$

Taking the determinants of both sides of (49), it follows that

$$\prod \ell_i = |r_{ij}|^{1/2} . \tag{52}$$

So the conditional distribution is proportional to

$$\exp\left[- \frac{1}{2} \psi (L,G,C, H ,\underline{u},\underline{v};\underline{\rho},\underline{\zeta})\right] |r_{ij}|^{\frac{n+2-p}{2}} \prod_{i<j} (\ell_j^2-\ell_i^2) \prod_i d\ell_i \prod_{i<j} \underline{g}_2' d\underline{g}_j$$

$$\cdot \prod_{i<j} \underline{c}_i' \, d\underline{c}_j \, dS . \tag{53}$$

But it is clear from (50) that

$$\prod_{i<j} (\ell_j^2 - \ell_i^2) \prod d\ell_i \prod_{i<j} \underline{g}_i' \, d\underline{g}_j \prod_{i<j} \underline{c}_i' \, d\underline{c}_j = \prod_{i=1}^{p} \prod_{\alpha=1}^{p} dt_{i\alpha} \tag{54}$$

so the conditional distribution of $t_{i\alpha}$, $i = 1, \ldots, p$, $\alpha = 1, \ldots, p$ and $\underline{u}$, $\underline{v}$ under the condition that $H = M_p$ is proportional to

$$\exp \left[-\frac{1}{2} \sum_{\gamma=p+1}^{n+2} \sum_{i=1}^{p} (\rho_i u_\gamma + \zeta_i v_\gamma)^2 \right] |r_{ij}|^{\frac{n+2-p}{2}} \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{p} \sum_{i=1}^{p} (t_{i\alpha} - \rho_i u_\alpha \right.$$

$$\left. - \zeta_i v_\alpha)^2 \right] \prod_{i=1}^{p} \prod_{\alpha=1}^{p} dt_{i\alpha} \, dS \tag{55}$$

Consequently we have the conditional marginal distribution of $\underline{u}$, $\underline{v}$ as the one given in (47).

This completes the proof of Lemma 7. From this point on, Wald's original proofs may be followed.

REFERENCES

[1] A. Wald, "On a statistical problem arising in the classification of an individual into one of two groups," Ann. Math. Stat., Vol. 15 (1944).

[2] Shung-Chern, "On Grassmann and differential rings and their relations to the theory of multiple integrals," Sankhyā, Vol. 7 (1945-46).

[3] A. T. James, "Normal multivariate analysis and the orthogonal group," Ann. Math. Stat., Vol. 25 (1954).

[4] A. T. James, loc. cit., Lemma 4.1.