Journal of Statistical Planning and Inference 138 (2008) 3419–3428

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

A tale of two countries: The Craig–Sakamoto–Matusita theorem

Junjiro Ogawa[1], Ingram Olkin[a,b,*,2]

[a] Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305-4065, Canada
[b] University of Calgary, Canada

ARTICLE INFO

Article history:
Received 9 April 2005
Accepted 2 September 2005
Available online 15 March 2008

MSC:
15A24
01A27
60E05
62E10

Keywords:
Quadratic form
Cochran's theorems
Matrix equations
Chi-square distribution

ABSTRACT

This paper has a long history. Junjiro Ogawa and I have been friends and collaborators for many years. We started this joint paper in the early 1990s and prepared a draft in September 1995. However, we were trying to improve and simplify some of the proofs, and we revised the paper in January 1998. But circumstances were such that we did not submit the paper, but planned to find further simplifications. Junjiro's untimely death in 2003 brought it forth again. Junjiro always liked this topic, and I am pleased to submit this joint paper in his memory. This paper contains both an historical summary as well as several new results. (Ingram Olkin)

1. Introduction

During World War II there was little or no communication between Japan and English speaking countries. But research in the mathematical sciences continued in both Japan and the English speaking countries. Thus, it is not surprising that related results would be obtained in one country without knowledge of results in the other. However, because English journals flourished and few Japanese journals were translated, the work of the Japanese statisticians during this period was not sufficiently known. In this article we provide a history that puts into perspective the work in Japan with that in English journals (see also Ogawa, 1993).

A bibliography by Dumas and Styan (1998) on the distribution of quadratic forms and the Craig–Sakamoto theorem lists over 300 entries. In the present survey we focus primarily on the subset that relates to the collaboration of the authors. In this we offer several new proofs. Two surveys that relate to the present theme are by Scarowsky (1973) and Dumas (1999); see also the book by Mathai and Provost (1992).

2. The central case

The problem of the independence of quadratic forms originated with Cochran (1934) who showed that if $x_1, \ldots, x_n$ are independent and identically distributed with a common standard normal distribution, then the quadratic forms

$$Q_1 = x'Ax \quad \text{and} \quad Q_2 = x'Bx,$$

* Corresponding author at: Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305-4065, USA. Tel.: +1 650 857 1497; fax: +1 725 8977.
E-mail address: iolkin@stat.stanford.edu (I. Olkin).

[1] Professor Emeritus; deceased 2003.

[2] Supported in part by National Science Foundation Grant DMS 93-01366.

0378-3758/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.jspi.2005.09.010

where A, B are symmetric, are independent if and only if the determinantal equation

$$|I-\xi A-\eta B|=|I-\xi A||I-\eta B|, \quad \forall \xi,\eta \tag{C}$$

holds.

At times two alternative expressions may be useful:

$$|I-\eta B(I-\xi A)^{-1}|=|I-\eta B|, \tag{C.1}$$

$$|I+\xi\eta AB(I-\xi A-\eta B)^{-1}|=1, \quad \forall \xi,\eta. \tag{C.2}$$

Obviously, (C) holds if $AB=0$. The tantalizing problem is the converse.

Craig (1943) states a theorem that condition (C) implies that

$$AB=0. \tag{O}$$

Craig proof is incorrect in that he assumed that A and B can be diagonalized simultaneously, that is, that A and B are commutative. With this assumption the proof is straightforward in that $I-\xi A$, $I-\eta B$ and $I-\xi A-\eta B$ can be diagonalized simultaneously.

Hotelling (1944) noted the error in Craig's proof and provided a proof based on the argument that if Q, Q_1, Q_2 are quadratic forms in the x's, and Q is independent of Q_1+Q_2 and Q is independent of Q_1, then Q is independent of Q_2. This is a more subtle argument, which also is false as was noted by Ogawa (1949).

Let us now turn to events on the Japanese side. About the same time as the appearance of Craig's statement, Sakamoto (1944) conjectured that condition (C) implied condition (O), but he did not provide a proof. Ogawa (1946, in Japanese) gave a proof using a linear algebraic argument, but again the proof was incorrect.

At this point in time we have a joint conjecture by Craig and Sakamoto, together with three flawed proofs. The first correct proof was that of Matusita (1949). This proof involves an expansion of the determinants together with an analysis of the eigenvalue structure of the matrices. In a footnote Matusita states that he had obtained this result in 1944, but only published the proof after noting the interest generated by the Craig–Sakamoto papers. Thus the name Craig–Sakamoto–Matusita.

Ogawa's (1949) paper reviews some history; he credits S. Nabeya with noting an error in his 1946 paper. In this paper he provides a detailed proof that (C) implies (O); and also discusses of Craig's (1947) paper (see Section 5). Two interesting lemmas due to Nabeya underlie the proof.

Lemma 1. *Let A and B be symmetric matrices of rank r and s, respectively, and $C=A+B$ of rank t. If $t=r+s$ and C is idempotent, then $AB=0$, and A and B are idempotent.*

Lemma 2. *Let the non-zero eigenvalues of the symmetric matrices A and B be $\xi_1,\dots,\xi_r$ and $\eta_1,\dots,\eta_s$, respectively, and of $C=A+B$ be $\gamma_1,\dots,\gamma_t$. If $t=r+s$ and $\Pi\gamma_i=(\Pi\xi_i)(\Pi\eta_i)$, then $AB=0$.*

A third correct proof is that of Aitken (1950), which is similar to that of Matusita. This proof is somewhat terse, and is explicated by Lancaster (1954) who also provides a review of the subject but without referencing some of the Japanese papers. About the same time Ogasawara and Takahashi (1951) gave a proof that is direct and straightforward. It is based on the expansion

$$-\log|I-\xi A|=\sum_1^{\infty}\frac{\operatorname{tr}(\xi A)^k}{k}. \tag{2.1}$$

Using (2.1) condition (C) can be rewritten as

$$\sum\frac{1}{k}\operatorname{tr}[(\xi A)^k+(\eta B)^k-(\xi A+\eta B)^k]=0, \quad \forall \xi,\eta. \tag{2.2}$$

By equating the coefficients of $\xi^2\eta^2$ in the expansion of (2.2) they obtain the condition

$$\operatorname{tr}(AB+BA)^2+2\operatorname{tr}(AB)(AB)'=0. \tag{2.3}$$

Although AB is not symmetric, $AB+BA$ is symmetric and its square is non-negative definite. Each term in (2.3) is non-negative, so that each term must be zero. But $\operatorname{tr}XX'=\Sigma x_{ij}^2=0$ implies that $X=0$, that is, $AB=0$.

3. Some further history

We note that neither Aitken nor Ogasawara and Takahashi were aware of the proof of Matusita. Aitken references a paper by Matérn (1949) who showed that if A and B are non-negative definite, rather than the weaker assumption of symmetry, then the independence of the quadratic forms $Q_1=x'Ax$ and $Q_2=x'Bx$ holds if Q_1 and Q_2 are uncorrelated, which is equivalent to $\operatorname{tr}AB=0$. If A and B are symmetric but not non-negative definite, then $\operatorname{tr}AB=0$ does not imply that $AB=0$. But the non-negative definiteness condition of A and B implies that $\operatorname{tr}AB=\operatorname{tr}A^{1/2}BA^{1/2}=0$, so that $A^{1/2}B^{1/2}=0$, and hence $AB=0$.

Few textbooks give a discussion or proof that (C) implies (O). Exceptions are Hogg and Craig (1978), Searle (1971), Guttman (1982), Kendall and Stuart (1963, 1968). See also Johnson and Kotz (1970). These books generally provide only a limited history. The book by Mathai and Provost (1992) is more specialized and provides considerable detail about the independence of linear and quadratic forms, but the Japanese history therein is incomplete.

As noted, extensive surveys are provided by Scarowsky (1973) and Dumas (1999). The paper by Driscoll and Gundberg (1986) gives a good history of Craig's work and much of the results in English language journals. But it does not provide an adequate discussion of the role of the Japanese researchers.

Kawada (1950) shows that for two quadratic forms $Q_1 = x'Ax$ and $Q_2 = x'Bx$, where $x_i \sim \mathcal{N}(0,1), i = 1, \ldots, n$, the covariance conditions

$$F_{ij} = \mathrm{cov}(Q_1^i, Q_2^j) = 0, \quad i,j = 1,2 \tag{K}$$

imply that $AB=0$. The conditions (K) are considerably weaker than the independence of Q_1 and Q_2, but stronger than the condition of zero correlation, as used by Matern (1949).

The proof that (K) implies (O) is based on the computation of moments to yield

$$\begin{aligned} F_{11} &= 2\,\mathrm{tr}\,AB = 0, \\ F_{12} &= 8\,\mathrm{tr}\,AB^2 + 4(\mathrm{tr}\,AB)(\mathrm{tr}\,B) = 0, \\ F_{21} &= 8\,\mathrm{tr}\,BA^2 + 4(\mathrm{tr}\,AB)(\mathrm{tr}\,A) = 0, \\ F_{22} &= 32\,\mathrm{tr}\,A^2B^2 + 16\,\mathrm{tr}\,ABAB + 16(\mathrm{tr}\,AB^2)(\mathrm{tr}\,A) + 16(\mathrm{tr}\,A^2B)(\mathrm{tr}\,B) \\ &\quad + 8(\mathrm{tr}\,AB)(\mathrm{tr}\,A)(\mathrm{tr}\,B) + 8(\mathrm{tr}\,AB)^2 = 0. \end{aligned} \tag{3.1}$$

Using F_{11}, F_{12} and F_{21} in F_{22} yields

$$2\,\mathrm{tr}\,A^2B^2 + \mathrm{tr}\,ABAB = 0. \tag{3.2}$$

With $G=AB$, (3.2) becomes $2\,\mathrm{tr}\,GG' + \mathrm{tr}\,G^2 = 0$. The problem here is that G is not symmetric, so that $\mathrm{tr}\,G^2$ need not be non-negative, as can be seen from

$$A = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

so that $\mathrm{tr}\,G^2 = -2$. However,

$$2\,\mathrm{tr}\,GG' + \mathrm{tr}\,G^2 = \mathrm{tr}\,GG' + \tfrac{1}{2}\,\mathrm{tr}\,(G+G')^2 = 0, \tag{3.3}$$

from which it follows that $G = 0$.

Note that Matern (1949) obtains F_{11}, which is not sufficient to prove that $AB = 0$, whereas Kawada (1950) makes the stronger assumption that (K) holds.

To summarize the status of the central case as of 1950, there are correct proofs by Matusita (1949), Ogawa (1949), Aitken (1950), and Kawada (1950). Schematically, we have the following implications:

$$(\mathrm{K}) \iff (\mathrm{O}) \iff (\mathrm{C}).$$

Most proofs that (C) implies (O) depend on a power series expansion. Three proofs that do not depend on power series expansions are those of Taussky (1958), Olkin (1997) and Li (2000). The proof by Taussky (1958) is based on an examination of the eigenvalues in

$$\Pi(1 - \xi\alpha_i)(1 - \eta\beta_i) = \Pi(1 - \xi\alpha_i - \eta\beta_i), \quad \forall\ \xi, \eta, \tag{3.4}$$

where α_i and β_i are the eigenvalues of A and B. The key point is that a pair of matrices A and B are said to have property L if $\xi A + \eta B$ has eigenvalues $\xi\alpha_i + \eta\beta_i$ for all ξ, η. Such matrices were introduced by Mark Kac and studied by Motzkin and Taussky (1952). An examination of (3.4) then shows that for each i either α_i or $\beta_i = 0$, from which we obtain that $AB = 0$.

The proof of Olkin is based on a determinantial lemma that has some intrinsic interest (see Marcus, 1998 for a generalization).

Lemma 3. *Let $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$ be a symmetric matrix with $B_{11} : m \times m$ and $B_{22} = n - m \times n - m$. Further define the 2×2 determinant $B(i,j) = \begin{vmatrix} b_{ii} & b_{ij} \\ b_{ij} & b_{jj} \end{vmatrix} = b_{ii}b_{jj} - b_{ij}^2$. If $\mathrm{tr}\,B = 0$ and*

$$\sum_{\substack{i,j \varepsilon \mathscr{A} \\ i<j}} B(i,j) + \sum_{j \varepsilon \mathscr{B}} \sum_{i \varepsilon \mathscr{A}} B(i,j) = 0, \tag{3.5}$$

where $\mathscr{A} = \{1, 2, \ldots, m\}, \mathscr{B} = \{m+1, \ldots, n\}$, then $B_{11} = 0, B_{12} = 0$.

The proof of Li (2000) is based on the following lemma.

Lemma 4. *Let C be a symmetric matrix with largest eigenvalue λ_1. If $c_{ii} = \lambda_1$ then $c_{ij} = 0$ for all $j \neq i$.*

Proof. With no loss in generality let $c_{11} = \lambda_1$. Further let $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)$, with the one in the k-th entry, and let $C = GDG'$, where G is orthogonal and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Then

$$c_{11} = e_1 C e_1' = e_1 GDG' e_1' = (g_{11}, \ldots, g_{1n}) D (g_{11}, \ldots, g_{1n})' = \Sigma g_{1j}^2 \lambda_j = \lambda_1. \tag{3.6}$$

Consequently, $g_{ij} = \pm 1, g_{12}^2 = \cdots = g_{1n}^2 = 0$. By orthogonality,

$$\Sigma g_{1j} g_{kj} = g_{1j} g_{kj} = \pm g_{kj} = 0, \quad k = 2, \ldots, n, \tag{3.7}$$

and hence

$$\begin{aligned} c_{1k} &= e_1 C e_k' = (g_{11}, \ldots, g_{1n}) D (g_{k1}, \ldots, g_{kn})' \\ &= g_{11} \lambda_1 g_{k1} = 0, \quad k = 2, \ldots, n. \qquad \square \end{aligned} \tag{3.8}$$

4. The central case: two new results

The condition (C) is a strong condition in that it holds for all ξ and all η. A condition weaker than (C) is

$$\begin{aligned} &\text{(a)} \quad |I - \xi A - \xi B| = |I - \xi A||I - \xi B|, \\ &\text{(b)} \quad |I - \xi A + \xi B| = |I - \xi A||I + \xi B|, \quad \forall \xi. \end{aligned} \tag{C*}$$

Ogawa conjectured that (C*) implies that $AB = 0$. That (C*) holds is indeed surprising. Although there is no natural statistical motivation for (C*), it has considerable interest in linear algebra. The following proof is due to the second author.

It is immediate that $AB = 0$ implies (C*). To prove the converse, expand (C*(a)) using (2.1) and equate the coefficients of ξ^4:

$$2\,\mathrm{tr}\, A^2B^2 + \mathrm{tr}\, ABAB + 2\,\mathrm{tr}(AB^2 + A^2B) = 0. \tag{4.1}$$

Now expand (C*(b)) to yield

$$2\,\mathrm{tr}\, A^2B^2 + \mathrm{tr}\, ABAB - 2\,\mathrm{tr}\,(AB^2 + A^2B) = 0. \tag{4.2}$$

Adding (4.1) and (4.2) yields (3.2), which implies that $AB = 0$.

The weakest assumptions (that is, the strongest result) that we are aware of are

$$|I - \xi A - \xi B| = |I - \xi A||I - \xi B|, \quad \forall \xi. \tag{C**}$$

Ogawa conjectured that (C**) implies that $AB = 0$ and supplied a proof similar to the method used by Taussky (1958), but which yields this stronger result. This proof is lengthy, and is provided in the Appendix.

5. The central case with a non-identity covariance matrix

It has been noticed that if $(x_1, \ldots, x_n)$ have a joint normal distribution with zero means and positive definite covariance matrix Σ, the transformation $x \to \Sigma^{-1/2}x, A \to \Sigma^{1/2}A\Sigma^{1/2}, B \to \Sigma^{1/2}B\Sigma^{1/2}$ reduces the problem to the case that the x's are independent, identically distributed standard normal random variables. The condition $AB = 0$ then becomes $(\Sigma^{1/2}A\Sigma^{1/2})(\Sigma^{1/2}B\Sigma^{1/2}) = 0$, that is $A\Sigma B = 0$.

This point was noted by Matusita (1949), by Aitken (1950), and by Ogasawara and Takahashi (1951). When Σ is singular an additional discussion is required to reduce the model to a subspace, or alternatively, to use generalized inverses (see Nagase and Banerjee, 1976; Rayner, 1974; Styan, 1970).

Good (1963) attempts to extend the determinantal condition (C) to the case that A and B are arbitrary matrices, instead of symmetric matrices, when x has a multivariate normal distribution with mean zero and covariance matrix Σ, not necessarily non-singular. As noted by Shanbhag (1966) the result obtained needs to be modified, and this was done in Good (1966), in which the main condition becomes $\Sigma A \Sigma B \Sigma = 0$.

6. The non-central case

When the mean of x is μ, the condition for the independence of $Q_1 = x'Ax$ and $Q_2 = x'Bx$ becomes

$$|I - \xi A||I - \eta B|/|I - \xi A - \eta B| \\ = \exp \mu'\{(I - \xi A)^{-1} + (I - \eta B)^{-1} - (I - \xi A - \eta B)^{-1} - I\}\mu \tag{NC}$$

for all ξ, η.

Clearly, if $AB = 0$, then each side of (NC) is unity, so that $AB = 0$ implies independence. As in the central case, the converse is the difficult part.

The non-central case was first proved by Ogawa (1950), and later by Laha (1956), who was unaware of Ogawa's solution. In both papers, a form of the following lemma provides a key step.

Lemma 5. *If $\phi_j(\xi, \eta)$ are rational polynomials in ξ and η, and*

$$\frac{\phi_1(\xi,\eta)}{\phi_2(\xi,\eta)} = \exp\left(\frac{\varphi_3(\xi,\eta)}{\varphi_4(\xi,\eta)}\right) \tag{6.1}$$

for all real ξ and η, then φ_1/φ_2 and ϕ_3/ϕ_4 are constants.

Of course, with the use of this lemma as applied to condition (NC), we obtain that the left-hand side of (NC) is a constant, which must be unity (take $\xi = \eta = 0$ say). Thus the left-hand side of (NC) reduces to the central case (C), from which it follows that $AB = 0$.

This lemma itself has some history. Ogawa (1950) gave a function-theoretic proof. Laha (1956) stated the lemma and asserted that it can be easily proved, but did not provide a proof. Driscoll and Gundberg (1986) discuss the non-central case and include two supplements, one due to Driscoll and one to Searle (1984), that provide details for the proof of the lemma. Both authors were unaware of Ogawa's proof. Another discussion of the lemma is provided by Harville and Kempthorne (1994). In a sequel paper Reid and Driscoll (1988) acknowledge Ogawa's priority; they also clarify some misstatements and provide the reduction (NC) $\Rightarrow$ (C), using an approach that does not require the lemma. A second sequel by Driscoll and Krasnicka (1995) provides further insights into the general case.

We discuss the non-central case afresh and provide a proof (due to the second author) that avoids the use of the lemma. However, it has an additional assumption. We comment on this assumption at the end of the proof. Indeed, with this assumption the central and non-central case can be treated simultaneously. In the central case an expansion of (C) using (2.1) yields Eq. (2.2)

$$\text{tr}(AB + BA)^2 + 2\,\text{tr}\,A^2B^2 = 0,$$

which implies that $AB = 0$. In the non-central case an expansion of (NC) using (2.1) yields the equation

$$\text{tr}\{[(AB + BA)^2 + (A^2B^2 + B^2A^2)]H\} = 0,$$

where $H = \frac{1}{4}I + \mu\mu'$ is positive definite

Commentary. It is perhaps surprising that if U, V and H are non-negative definite, that $\text{tr}\,UV \geqslant 0$ but $\text{tr}\,UVH$ need not be non-negative. Throughout, the study of condition (C) leads to the quantity $\text{tr}\,(A^2B^2+B^2A^2)=2\,\text{tr}\,AB^2A$, which is non-negative. However, $Q = A^2B^2 + B^2A^2$ need not be non-negative, as can be seen from the fact that

$$A^2B^2 + B^2A^2 = (A, B)\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} \tag{6.2}$$

and $\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$ has eigenvalues ± 1. Consequently, it remains unclear as to conditions for $\text{tr}(A^2B^2 + B^2A^2)H$ to be non-negative. If A and B commute, then $\text{tr}\,A^2B^2H = \text{tr}\,ABHBA \geqslant 0$, or if $A = I + ae'e$ and $eB = be$, $a, b \geqslant 0$, then $\text{tr}(A^2B^2 + B^2A^2)H \geqslant 0$. For further discussion see Werner and Olkin (2004).

Theorem. *Let A and B be $n \times n$ symmetric matrices such that $A^2B^2 + B^2A^2$ is non-negative definite, and $\mu = (\mu_1, \ldots, \mu_n)'$. If*

$$\log|I - \xi A| + \log|I - \eta B| - \log|I - \xi A - \eta B| \\ = \mu'\{(I - \xi A)^{-1} + (I - \eta B)^{-1} - (I - \xi A - \eta B)^{-1} - I\}\mu, \tag{6.3}$$

then $AB = 0$.

Proof. Expand (6.3) in a series:

$$-\sum_{k=1}^{\infty}\frac{1}{k}\operatorname{tr}[(\xi A)^k+(\eta B)^k-(\xi A+\eta B)^k]=\mu'\left\{\sum_{k=0}^{\infty}[(\xi A)^k+(\eta B)^k-(\xi A+\eta B)^k-I]\right\}\mu$$
$$=\sum_{k=1}^{\infty}\operatorname{tr}\{[(\xi A)^k+(\eta B)^k-(\xi A+\eta B)^{-k}](\mu\mu')\},$$

or equivalently

$$\sum_{1}^{\infty}\operatorname{tr}\left([(\xi A)^k+(\eta B)^k-(\xi A+\eta B)^k]\left\{\frac{I}{k}+\mu\mu'\right\}\right)=0,\quad \forall \xi,\eta. \tag{6.4}$$

Setting the coefficient of $\xi^2\eta^2$ equal to zero gives

$$\operatorname{tr}[(A^2B^2+B^2A^2)H]+\operatorname{tr}[(AB+BA)^2H]=0, \tag{6.5}$$

where $H=\frac{1}{4}I+\mu\mu'$ is positive definite. The second term is equal to $\operatorname{tr}(AB+BA)H(AB+BA)$, which is non-negative because H is positive definite; the first term is non-negative by assumption. □

7. The general central case: second degree polynomial statistics

One extension of the independence of two quadratic forms is to include a linear term:

$$P_1(x)=x'Ax+x'a,\qquad P_2(x)=x'Bx+x'b. \tag{7.1}$$

This was first examined by Laha (1956) when $x_1,\dots,x_n$ are independent identically distributed standard normal random variables. The condition for independence is

$$\frac{|I-\xi A||I-\eta B|}{|I-\xi A-\eta B|}=\exp\left\{\frac{1}{4}[(\xi a+\eta b)'(I-\xi A-\eta B)^{-1}(\xi a+\eta b)\right.$$
$$\left.-\xi^2a'(I-\xi A)^{-1}a-\eta^2b'(I-\eta B)^{-1}b]\right\}. \tag{7.2}$$

Laha uses Lemma 5 to separate each part, from which he shows that

$$AB=0,\quad a'B=0,\quad b'A=0,\quad a'b=0. \tag{7.3}$$

The implication that $AB=0$ follows from the central case by setting the left-hand side equal to unity. Setting the right-hand side of (7.2) equal to unity leads to

$$(\xi a+\eta b)'(I-\xi A-\eta B)^{-1}(\xi a+\eta b)-\xi^2a'(I-\xi A)^{-1}a-\eta^2b'(I-\eta B)^{-1}b=0 \tag{7.4}$$

for all ξ,η.

The coefficients of $\xi\eta$ and $\xi^2\eta^2$ yield

$$a'b=0,\quad a'B^2a+b'A^2b=0,$$

from which the result follows.

The non-central case follows from (7.2) by the transformation $x\to x-\mu$, $a\to a+2A\mu$, $b\to b+2B\mu$. The conditions (7.2) remain unchanged. For example

$$(a'b)\to(a'+2\mu'A)(b+2B\mu)=(a'b)+2\mu'(Ab)+2(a'B)\mu+4\mu'(AB)\mu=0.$$

Provost (1996) provides two proofs of the general case, one of which depends on Lemma 5 and the other is obtained by equating coefficients in (7.4). However, for the non-central case it is necessary to examine more combinations of coefficients than for the central case.

Kac (1945) notes that the linear form $L=x'a$ and quadratic form $Q=x'Ax$ are independent if and only if $Aa=0$. Necessity in this result follows from (C) because of the independence of Q and $L^2=x'aa'x$, so that $Aaa'=0$, which implies $Aa=0$.

8. Bilinear forms

Bilinear forms can be created in different ways. Craig (1947) considers independent bivariate random variables (x_i, y_i), $i = 1, \ldots, n$, each pair having a bivariate normal distribution with means zero, unit variances and correlation ρ. If $x = (x_1, \ldots, x_n)'$, $y = (y_1, \ldots, y_n)'$, then $Q_1 = x'Ay$ and $Q_2 = x'By$ with A and B symmetric are independent if and only if

$$\begin{aligned} &|I - 2\rho\xi A - (1-\rho^2)\xi^2 A^2|\ |I - 2\rho\eta B - (1-\rho^2)\eta^2 B^2| \\ &\quad = |I - 2\rho(\xi A + \eta B) - (1-\rho^2)(\xi A + \eta B)^2|, \quad \forall \xi, \eta. \end{aligned} \tag{8.1}$$

If $AB = 0$ then (8.1) holds. To show the converse note that the coefficient of $\xi\eta$ yields $\operatorname{tr} AB = 0$ and the coefficient of $\xi^2\eta^2$ yields $\operatorname{tr} ABAB + 2 \operatorname{tr} A^2B^2 = 0$, which is the same as (3.2). Consequently (O) holds.

A second way to generate a bilinear form was noted by Aitken (1950). Let $x = (x_1, \ldots, x_r)'$, $y = (y_1, \ldots, y_s)'$, where x and y are jointly normally distributed with covariance matrix Σ of dimension $r + s$. Aitken noted that results for the independence of the bilinear forms $Q_1 = x'Ay$ and $Q_2 = x'By$ can be obtained from the independence of quadratic forms by noting that

$$x'Ay = \frac{1}{2}\,(x'\ y')\begin{pmatrix} 0 & A' \\ A & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \equiv z'\tilde{A}z.$$

Thus the independence of Q_1 and Q_2 is equivalent to the independence of $z'\tilde{A}z$ and $z'\tilde{B}z$. The condition $\tilde{A}\tilde{B} = 0$ becomes

$$\begin{pmatrix} 0 & A' \\ A & 0 \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} 0 & B' \\ B & 0 \end{pmatrix} = \begin{pmatrix} A'\Sigma_{22}B & A'\Sigma_{21}B' \\ A\Sigma_{12}B & A\Sigma_{11}B' \end{pmatrix} = 0.$$

Mathai and Provost (1992) consider the independence of $Q_1 = x'Ay$ and $Q_2 = y'By$, where x is p-dimensional, y is q-dimensional, A is a $p \times q$ matrix and B is a $q \times q$ symmetric matrix. The vector (x, y) has a normal distribution with mean zero and non-singular covariance matrix $\Sigma = \begin{pmatrix} I & \Psi \\ \Psi' & I \end{pmatrix}$. (The transformation $x \to \Sigma_{11}^{-1/2}x$, $y \to \Sigma_{22}^{-1/2}y$ permits this canonical form.) From the joint moment generating function, Q_1 and Q_2 are independent if and only if

$$\begin{aligned} &|I - \xi(\Psi'A + A'\Psi) - \eta B - \xi^2 A'(I - \Psi\Psi')A| \\ &\quad = |I - \xi(\Psi'A + A'\Psi) - \xi^2 A'(I - \Psi\Psi')A|\ |I - \eta B|, \quad \forall \xi, \eta. \end{aligned} \tag{8.2}$$

The pair of conditions $AB = 0$ and $A'\Psi B = 0$ imply (8.2). As before, the converse is more troubling. Expanding (8.2) in a series and setting the coefficient of $\xi^2\eta^2$ equal to 0 yields

$$\begin{aligned} 0 &= 2 \operatorname{tr} BA'(I - \Psi\Psi')AB + 2 \operatorname{tr} (\Psi'A + A'\Psi)^2 B^2 + \operatorname{tr} [(\Psi'A + A'\Psi)B]^2 \\ &= \{2 \operatorname{tr} [BA'(I - \Psi\Psi')^{1/2}]\ [(I - \Psi\Psi')^{1/2}AB]\} \\ &\quad + \{\operatorname{tr} [(\Psi'A + A'\Psi)B]^2 + 2 \operatorname{tr} (\Psi'A + A'\Psi)B^2(\Psi'A + A'\Psi)\}. \end{aligned}$$

The first term is non-negative, the second and third terms together are non-negative from (2.5), so that each term must be equal to zero, from which the result follows.

Although at first glance this result has the appearance of being a generalization, it follows from the result on quadratic forms. Rewrite

$$Q_1 = (x'\ y')\begin{pmatrix} 0 & A \\ A' & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 2x'Ay,$$

$$Q_2 = (x'\ y')\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = y'By.$$

The condition for independence is

$$\begin{pmatrix} 0 & A \\ A' & 0 \end{pmatrix}\begin{pmatrix} I & \Psi \\ \Psi' & I \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} 0 & AB \\ 0 & A'\Psi B \end{pmatrix} = 0,$$

so that $AB = 0$ and $A'\Psi B = 0$, as was to be proved.

9. Multivariate versions

Because the Wishart distribution is a multivariate version of the chi-square distribution, we can expect to find matrix extensions of the basic results. In the multivariate version let

$$R_1 = XAX', \quad R_2 = XBX',$$

where now X is a $p \times n$ matrix whose columns have a p-variate normal distribution with means zero and covariance matrix I.

This model can be reduced to the univariate case by letting

$$Q_1 = s'R_1 s = (s'X)A(X's) \equiv z'Az,$$

$$Q_2 = s'R_2 s = (s'X)B(X's) \equiv z'Bz,$$

where z is an n-dimensional vector that has a normal distribution with mean 0, zero covariances and variances $s's$, which we can take to be unity (or alternatively, start with $z = s/\sqrt{s's}$). Consequently, the multivariate model reduces to a linear model.

When the mean is $EX = \Theta$, Cov $(x_{ij}, x_{\ell m}) = \sigma_{i\ell}\psi_{jm}$, and

$$Q_1 = XAX' + \tfrac{1}{2}(LX' + XL'),$$

$$Q_2 = XBX' + \tfrac{1}{2}(MX' + xM'),$$

the problem becomes more complicated. This general model is studied by Khatri (1963), who apparently was unaware of previous results.

10. Extensions

A variety of extensions can be posed, some of which have received attention. Although quadratic forms arise naturally in a statistical context, higher degree polynomials constitute one direction of study. It is known that the independence of higher degree polynomials in normal variables does not lead to simple conditions.

A second direction is the extension of the chi-square distribution to the Wishart distribution as described in Section 6.

Another direction by Letac and Massam (1995) is an extension to a Wishart distribution on symmetric cones; their paper includes generalized versions of the proofs by Ogasawara and Takahashi (1951) and of Matusita (1949) and Lancaster (1954).

Acknowledgments

The authors are most grateful to Michael F. Driscoll, himself an author of several papers on this subject, for his many most helpful comments and suggestions on several versions of this paper.

Appendix

We here provide a proof due to Ogawa that (C**) implies that $AB = 0$.

The condition (C**): $|I - \xi A - \xi B| = |I - \xi A||I - \xi B|$ can be restated in a homogeneous form as

$$x^n|xI - A - B| = |xI - A||xI - B|. \tag{A.1}$$

This means that the set of non-zero eigenvalues of $A + B$ is the union of sets of non-zero eigenvalues of A and B.

We consider matrices A, B and $C = A + B$ as the linear transformations in the n-dimensional vector space $\mathscr{L}$ over the field $\mathscr{R}$ of the real numbers. Let the non-zero eigenvalues of A be $\alpha_1, \ldots, \alpha_r$, the non-zero eigenvalues of B be $\beta_1, \ldots, \beta_s$ and non-zero eigenvalues of C be $\gamma_1, \ldots, \gamma_t$. Then $t = r + s$, and

$$\{\gamma_1, \ldots, \gamma_t\} = \{\xi_1, \ldots, \xi_r\} + \{\eta_1, \ldots, \eta_s\}. \tag{A.2}$$

Let the decomposition of the whole space $\mathscr{L}$ into the range and its orthocomplement corresponding to A, B and C be

$$\begin{aligned} &\mathscr{L} = \mathscr{L}_A + \mathscr{L}_A^+, \quad \mathscr{L}_A^+ = \mathscr{L}_A^0, \\ &\mathscr{L} = \mathscr{L}_B + \mathscr{L}_B^+, \quad \mathscr{L}_B^+ = \mathscr{L}_B^0, \\ &\mathscr{L} = \mathscr{L}_C + \mathscr{L}_C^+, \quad \mathscr{L}_C^0 = \mathscr{L}_C^0, \end{aligned} \tag{A.3}$$

respectively, where $\mathscr{L}_A$ is the range of A, $\mathscr{L}_A^+$ is the orthocomplement of $\mathscr{L}_A$ in $\mathscr{L}$ and $\mathscr{L}_A^0$ is the eigenspace of A corresponding to the eigenvalue 0. Let the orthonormal systems of eigenvectors spanning $\mathscr{L}_A$, $\mathscr{L}_B$ and $\mathscr{L}_C$ be $\{a\}$, $\{b\}$, and $\{c\}$, respectively.

Because the three matrices A, B and C are symmetric, $\dim \mathscr{L}_A = r$, $\dim \mathscr{L}_B = s$, $\dim \mathscr{L}_C = t$, and $\dim \mathscr{L}_C \leqslant \dim \mathscr{L}_{A\cup B} = \dim \mathscr{L}_A + \dim \mathscr{L}_B - \dim(\mathscr{L}_A \cap \mathscr{L}_B)$, from which it follows that $\dim(\mathscr{L}_A \cap \mathscr{L}_B) = 0$. The condition $AB = 0$ is equivalent to the conditions that

$$AB\phi = 0, \quad \forall \phi \in \mathscr{L}. \tag{A.4}$$

Because $B\phi \in \mathscr{L}_B$, $B\phi$ is a linear combination of eigenvectors $b_1, \ldots, b_s$ corresponding to non-zero eigenvalues $\eta_1, \ldots, \eta_s$ of B, that is, $B\phi = \lambda_1 b_1 + \cdots + \lambda_s b_s$; it then follows that

$$AB\phi = \lambda_1 Ab_1 + \cdots + \lambda_s Ab_s. \tag{A.5}$$

Expressing b_j as the linear combination of eigenvectors $a_1, \ldots, a_r$ corresponding to the non-zero eigenvalues $\xi_1, \ldots, \xi_r$ of A, one obtains

$$b_j = d_{1j}a_1 + d_{2j}a_2 + \cdots + d_{rj}a_r + \sigma_{jj}, \quad d_{ij} = a_i' b_j, \quad t_j \in \mathscr{L}_A^0,$$

$$Ab_j = d_{1j}\xi_1 a_1 + \cdots + d_{rj}\xi_r a_r, \quad j = 1, \ldots, s.$$

Therefore, the condition (A.4) is equivalent to the condition $D = (d_{ij}) = 0$.

$$D = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1s} \\ \vdots & \vdots & & \vdots \\ d_{r1} & d_{r2} & \ldots & d_{rs} \end{bmatrix} = 0.$$

If one puts

$$f_1 = a_i - d_{i1}b_1 - \cdots - d_{is}b_s \in \mathscr{L}_B^0, \quad i = 1, \ldots, r,$$

then the Gramian of the vector set $\{f_1, \ldots, f_s\}$ is $|I - DD'|$ because

$$f_i' = \delta_{ij} - \sum_{k=1}^{s} d_{ik}\, d_{jk}, \quad i, j = 1, \ldots, r.$$

By the Hadamard inequality, $|I - DD'| \leqslant \Pi_{i=1}^{s}(1 - \Sigma_{j=1}^{r}\ d_{ij}^2)$, so that if $|I - DD'| = 1$, then

$$\sum_{i=1}^{s} d_{ij}^2 = 0, \ \ i = 1, \ldots, s \quad \text{and} \quad D = 0.$$

Now we show that $D = 0$ holds.

Let the matrix representation of C with respect to the eigenvectors $\{t\}$ of C corresponding to non-zero eigenvalues $\{\gamma\}$ be

$$\begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \quad \text{where } \Lambda_1 = \text{diag}\,(\xi_1, \ldots, \xi_r),\ \ \Lambda_2 = \text{diag}\,(\eta_1, \ldots, \eta_s),$$

$$C(u_1, \ldots, u_t) = (Cu_1, \ldots, Cu_t) = (u_1, \ldots, u_t)\begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}.$$

Because

$$Aa_i = \xi_i a_i, \quad Bb_i = \sum_{j=1}^{s} \eta_j d_{ij} b_j,$$

$$Ab_j = \sum_{i=1}^{r} \xi_i d_{ij} a_i, \quad Bb_j = \eta_j b_j,$$

we obtain

$$Ca_i = Aa_i + Ba_i = \xi_i a_i + \sum_{j=1}^{s} \eta_j d_{ij} b_j,$$

$$Cb_j = Ab_j + Bb_j = \sum_{i=1}^{r} \xi_i d_{ij} a_i + \eta_j b_j,$$

$$C(a_1, \ldots, a_r, b_1, \ldots, b_s) = (a_1, \ldots, a_r, b_1, \ldots, b_s)\begin{pmatrix} \Lambda_1 & \Lambda_1 D' \\ \Lambda_2 D & \Lambda_2 \end{pmatrix}, \tag{A.6}$$

that is, the matrix representaiton of the linear transformation C with respect to the basis vectors $\{a, b\}$ of $\mathscr{L}_C$ is given by

$$R_C = \begin{pmatrix} \Lambda_1 & \Lambda_1 D' \\ \Lambda_2 D & \Lambda_2 \end{pmatrix} = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}\begin{pmatrix} I & D' \\ D & I \end{pmatrix}. \tag{A.7}$$

If we denote the transformation matrix between $\{a, b\}$ and $\{t\}$ by

$$(u_1, \dots, u_t) = (a_1, \dots, a_r, b_1, \dots, b_s)Q = (a_1, \dots a_r, b_r, \dots, b_s) \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix},$$

then

$$\begin{pmatrix} \Lambda_1 & \Lambda_1 D' \\ \Lambda_2 D & \Lambda_2 \end{pmatrix} Q = Q \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}, \tag{A.8}$$

and taking the determinants of both sides, and noting (A.7), we obtain that $|I - DD'| = |I - D'D| = 1$, as was to be proved.

Zielinski (1985) maintains that the Gramians of $\{a, b\}$ and $\{u\}$ are equal. But this is not true, because

$$(u_i, u_j) = Q' \begin{pmatrix} a'a & a'b \\ b'a & b'b \end{pmatrix} Q,$$

taking the determinant of both sides yields Gramian of $\{a, b\} = |QQ'|^{-1}$.

If the transformation matrix Q is orthogonal, then Zielinski's proposition is correct.

References

Aitken, A.C., 1950. On the statistical independence of quadratic forms in normal variates. Biometrika 37, 93–96.
Cochran, W.G., 1934. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. Proc. Cambridge Philos. Soc. 30, 178–191.
Craig, A.T., 1943. Note on the independence of certain quadratic forms. Ann. Math. Statist. 14, 195–197.
Craig, A.T., 1947. Bilinear forms in normally correlated variables. Ann. Math. Statist. 18, 565–573.
Driscoll, M.F., Gundberg Jr., W.R., 1986. A history of the development of Craig's theorem. Amer. Statist. 42, 139–142.
Driscoll, M.F., Krasnicka, B., 1995. An accessible proof of Craig's theorem in the general case. Amer. Statist. 49, 59–62.
Dumas, M.F., 1999. The Craig–Sakamoto theorem, Dissertation, McGill Univeristy.
Dumas, M.F., Styan, G.P.H., 1998. Private communication.
Good, I.J., 1963. On the independence of quadratic expressions (with an appendix by L.R. Welch.). J. Roy. Statist. B 25, 377–382 (Corrigenda: J. Roy. Statist. Soc. B 28 (1966) 584).
Good, I.J., 1966. On the independence of quadratic expressions: corrigenda. J. Roy. Statist. Soc. B 28, 584.
Guttman, I., 1982. Linear Models. Wiley, New York.
Harville, D.A., Kempthorne, O., 1994. An alternative way to establish the necessity parts of the classical results on the distribution of quadratic forms. Reprint 94-32, Statistical Laboratory, Iowa State University.
Hogg, R.V., Craig, A.T., 1978. Introduction to Mathematical Statistics. fourth ed. Macmillan, New York.
Hotelling, H., 1944. A note on a matrix theorem of A.T. Craig. Ann. Math. Statist. 15, 427–429.
Johnson, N.L., Kotz, S., 1970. Distributions in Statistics, Continuous and Univariate Distributions–2. Wiley, New York.
Kac, M., 1945. A remark on independence of linear and quadratic forms involving independent Gaussian variables. Ann. Math. Statist. 16, 400–401.
Kawada, Y., 1950. Independence of quadratic forms in normally correlated variables. Ann. Math. Statist. 21, 614–615.
Kendall, M.G., Stuart, A., 1963, 1968. The Advanced Theory of Statistics, Vol. 1: Distribution Theory–2. Hafner Publishing, New York.
Khatri, C.G., 1963. Further contributions to Wishartness and independence of second-degree polynomials in normal vectors. J. Indian Statist. Assoc. 1, 61–70.
Laha, R.G., 1956. On the stochastic independence of two second degree polynomial statistics in normally distributed covariates. Ann. Math. Statist. 27, 790–796.
Lancaster, H.O., 1954. Traces and cumulants of quadratic forms in normal variables. J. Roy. Statist. Soc. Ser. B 16, 247–254.
Letac, G., Massam, H., 1995. Craig–Sakamoto's theorem for the Wishart distributions on symmetric cones. Ann. Inst. Statist. Math. Tokyo 47, 785–799.
Li, C.-K., 2000. A simple proof of the Craig–Sakamoto theorem. Linear Algebra Appl. 321, 281–283.
Marcus, M., 1998. On a determinant result of I. Olkin. Linear Algebra Appl. 277, 237–238.
Matérn, B., 1949. Independence of non-negative quadratic forms in normally correlated variables. Ann. Math. Statist. 20, 119–120.
Mathai, M.A., Provost, S.B., 1992. Quadratic Forms in Random Variables, Theory and Applications. Marcell Dekker, New York.
Matusita, K., 1949. Note on the independence of certain statistics. Ann. Inst. Statist. Math. 1, 79–82.
Motzkin, T.S., Taussky, O., 1952. Pairs of matrices with property L. Trans. Amer. Math. Soc. 73, 108–114.
Nagase, G., Banerjee, K.S., 1976. Independence of second degree polynomial statistics in normal variates. Amer. Statist. 30, 201.
Ogasawara, T., Takahashi, M., 1951. Independence of quadratic quantities in a normal system. J. Hiroshima Univ. Ser. A 15, 1–3.
Ogawa, J., 1946. On the independence of quadratic forms. Res. Memoire Inst. Statist. Math. Tokyo 2, 98–111 (in Japanese).
Ogawa, J., 1949. On the independence of bilinear and quadratic forms of a random sample from a normal population. Ann. Inst. Statist. Math. 1, 83–108.
Ogawa, J., 1950. On the independence of quadratic forms in a non-central normal system. Osaka Math. J. 2, 151–159.
Ogawa, J., 1993. A history of the development of Craig–Sakamoto's theorem viewed from the Japanese standpoint. Proc. Ann. Inst. Statist. Math. 41, 47–59 (in Japanese).
Olkin, I., 1997. A determinantal proof of the Craig–Sakamoto theorem. Linear Algebra Appl. 264, 217–223.
Provost, S.B., 1996. On Craig's theorem and its generalizations. J. Statist. Plann. Inference 53, 311–321.
Rayner, A.A., 1974. Quadratic forms and degeneracy: a personal view. Unpublished paper, Annual Conference of the South African Statistical Association, Pietermaritzburg, South Africa.
Reid, J.G., Driscoll, M.F., 1988. An accessible proof of Carig's theorem in the non-central case. Amer. Statist. 42, 139–142.
Sakamoto, H., 1944. Independence of two statistics. Res. Meuoirs Inst. Statist. Math. Tokyo 1 (9), 1–25 (in Japanese).
Scarowsky, I., 1973. Quadratic forms in normal variables, Unpublished Master's thesis, Department of Mathematics, McGill University.
Searle, S.R., 1971. Linear Models. Wiley, New York.
Searle, S.R., 1984. Detailed proofs (class notes) of necessary and sufficient conditions for independence and chi-square distribution properties of quadratic forms of normal variables, Technical Report BU-834-M, Cornell University, Biometrics Unit Series.
Shanbhag, D.N., 1966. On the independence of quadratic forms. J. Roy. Statist. Soc. B 28, 582–583.
Styan, G.P.H., 1970. Notes on the distribution of quadratic forms in singular normal variables. Biometrika 57, 567–572.
Taussky, O., 1958. On the matrix theorem of A.T. Craig and H. Hotelling. Indag. Math. 20, 139–141.
Werner, H.J., Olkin, I., 2004. On permutations of the factors in a marix product, Unpublished Manuscript.
Zielinski, A., 1985. A note on Nabeya's proof of the Sakamoto–Craig lemma. Statist. Decisions Supplement 2, 299–300.