A NOTE ON CALCULATING TRACE($\mathbf{X}^2$)

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, New York 14853

BU-1162-M June, 1992

Abstract

Simplified expressions are given for $\text{tr}(\mathbf{X}^2)$ for certain forms of $\mathbf{X}$ that arise in the variance of a quadratic form, $\text{var}(\mathbf{y}'\mathbf{A}\mathbf{y}) = 2\,\text{tr}[(\mathbf{V}\mathbf{A})^2]$, when $\mathbf{y} \sim N(\mathbf{0}, \mathbf{V})$.

Introduction

For a vector of random variables, $\mathbf{y}$, having a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{V}$, it is well known that $\mathbf{y}'\mathbf{A}\mathbf{y}$ has variance $2\,\text{tr}[(\mathbf{A}\mathbf{V})^2] + 4\boldsymbol{\mu}'\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu}$, where $\text{tr}(\mathbf{X})$ is the trace of $\mathbf{X}$, the sum of its diagonal elements. In representing $\mathbf{A}\mathbf{V}$ by $\mathbf{X}$ it is therefore of interest to consider the calculation of $\text{tr}(\mathbf{X}^2)$. The starting point is the well-known result that

$$\text{tr}(\mathbf{X}\mathbf{X}') = \text{tr}(\mathbf{X}'\mathbf{X}) = \textstyle\sum_i \sum_j x_{ij}^2 \,, \tag{1}$$

the sum of squares of every element x_{ij} of $\mathbf{X}$.

Then, for

$$\text{symmetric } \mathbf{X}: \qquad \text{tr}(\mathbf{X}^2) = \text{tr}(\mathbf{X}\mathbf{X}') = \textstyle\sum_i \sum_j x_{ij}^2 \,, \tag{2}$$

and for

$$\text{skew-symmetric } \mathbf{X}: \qquad \text{tr}(\mathbf{X}^2) = \text{tr}[\mathbf{X}(-\mathbf{X})'] = -\textstyle\sum_i \sum_j x_{ij}^2 = -2 \sum_i \sum_{j>i} x_{ij}^2 \,. \tag{3}$$

And, of course, for

$$\text{idempotent } \mathbf{X}: \qquad \text{tr}(\mathbf{X}^2) = \text{tr}(\mathbf{X}) = r(\mathbf{X}) \,,$$

where $r(\mathbf{X})$ is the rank of $\mathbf{X}$.

For non-symmetric $\mathbf{X}$ only the case of square $\mathbf{X}$ need be considered because only then does $\text{tr}(\mathbf{X}^2)$ exist. The latter can then be expressed in a form analogous to (1) by writing

$$\mathbf{X} = \tfrac{1}{2}(\mathbf{S} + \mathbf{D})$$

for

$$\mathbf{S} = \mathbf{S}' = \mathbf{X} + \mathbf{X}' = \{s_{ij} = x_{ij} + x_{ji}\}$$

and

$$\mathbf{D} = -\mathbf{D}' = \mathbf{X} - \mathbf{X}' = \{d_{ij} = x_{ij} - x_{ji}\} \,.$$

Then

$$\mathrm{tr}(\mathbf{X}^2) = \tfrac{1}{4}[\mathrm{tr}(\mathbf{S}^2) + \mathrm{tr}(\mathbf{D}^2) + 2\,\mathrm{tr}(\mathbf{SD})] \,.$$

And, because

$$\begin{aligned} \mathrm{tr}(\mathbf{SD}) &= \mathrm{tr}[(\mathbf{X} + \mathbf{X}')(\mathbf{X} - \mathbf{X}')] \\ &= \mathrm{tr}(\mathbf{X}^2) - \mathrm{tr}[(\mathbf{X}')^2] = \mathrm{tr}(\mathbf{X}^2) - \mathrm{tr}\{[(\mathbf{X}')^2]'\} \\ &= \mathrm{tr}(\mathbf{X}^2) - \mathrm{tr}(\mathbf{X}^2) = 0 \,, \end{aligned}$$

utilizing (2) and (3) gives

$$\mathrm{tr}(\mathbf{X}^2) = \tfrac{1}{4}[\mathrm{tr}(\mathbf{S}^2) + \mathrm{tr}(\mathbf{D}^2)] = \tfrac{1}{4}\textstyle\sum_i \sum_j (s_{ij}^2 - d_{ij}^2) \,,$$

and this is, of course,

$$\mathrm{tr}(\mathbf{X}^2) = \textstyle\sum_i \sum_j x_{ij}\, x_{ji} \,.$$

Dispersion Matrices

When $\mathbf{X}$ has the form $\mathbf{AV}$ for $\mathbf{V}$ being a dispersion matrix, $\mathbf{V}$ often has a form that utilizes $\mathbf{J}$-matrices, which are square with every element unity. For example, in the 1-way classification, random model, with a classes and n_i observations in class i, $\mathbf{V} = \left\{_d\ \sigma_\alpha^2 \mathbf{J}_{n_i} + \sigma_e^2 \mathbf{I}_{n_i}\right\}_{i=1}^{a}$ a block diagonal matrix with $\mathbf{J}_{n_i}$ being an $n_i \times n_i$ matrix of ones, and $\mathbf{I}_{n_i}$ being an identity matrix of order n_i. Moreover

$$\mathbf{J}_{n_i} = \mathbf{1}_{n_i}\mathbf{1}'_{n_i}$$

where $\mathbf{1}_{n_i}$ is the summing vector of order n_i, a vector of n_i ones. It is therefore of interest to consider $\mathrm{tr}(\mathbf{AT})^2$ for $\mathbf{T}$ being various matrices involving $\mathbf{J}$, $\mathbf{I}$ and $\mathbf{1}$. Examples follow.

$$\mathrm{tr}(\mathbf{1a}') = \mathrm{tr}(\mathbf{a}'\mathbf{1}) = \textstyle\sum_i a_i \; , \tag{4}$$

$$\mathrm{tr}(\mathbf{AJ}) = \mathrm{tr}(\mathbf{1}'\mathbf{A1}) = \textstyle\sum_i \sum_j a_{ij} = a_{\cdot\cdot} \quad \text{for} \quad a_{\cdot\cdot} = \sum_i \sum_j a_{ij} \; , \tag{5}$$

$$\mathrm{tr}(\mathbf{AJ})^2 = \mathrm{tr}(\mathbf{A11}'\mathbf{A11}') = \mathrm{tr}(\mathbf{1}'\mathbf{A11}'\mathbf{A1}) \left(\textstyle\sum_i \sum_j a_{ij}\right)^2 = a^2_{\cdot\cdot} \; , \tag{6}$$

$$\begin{aligned} \mathrm{tr}(\mathbf{AJB})^2 &= \mathrm{tr}(\mathbf{A11}'\mathbf{BA11}'\mathbf{B}) = \mathrm{tr}(\mathbf{1}'\mathbf{BA11}'\mathbf{BA1}) \\ &= \left(\textstyle\sum_j b_{\cdot j} a_{j\cdot}\right)^2 \; \text{for} \; b_{\cdot j} = \sum_i b_{ij} \; \text{and} \; a_{j\cdot} = \sum_k a_{jk} \\ &= [\text{inner product of (the row of column sums of } \mathbf{B}) \\ &\quad\; \text{with (the column of row sums of } \mathbf{A}]^2 \; ; \end{aligned} \tag{7}$$

$$\begin{aligned} \mathrm{tr}(\mathbf{AJA}')^2 &= [\text{inner product of (the row of row sums of } \mathbf{A}) \text{ with itself}]^2 \\ &= \left(\textstyle\sum_i a^2_{i\cdot}\right)^2 . \end{aligned} \tag{8}$$

$$\begin{aligned} \mathrm{tr}[\mathbf{A}(\alpha\mathbf{I} + \beta\mathbf{J})]^2 &= \alpha^2 \mathrm{tr}(\mathbf{A}^2) + \beta^2 \mathrm{tr}(\mathbf{AJ})^2 + 2\alpha\beta\, \mathrm{tr}(\mathbf{A}^2\mathbf{J}) \\ &= \alpha^2 \mathrm{tr}(\mathbf{A}^2) + \beta^2 a^2_{\cdot\cdot} + 2\alpha\beta\,(\mathbf{1}'\mathbf{A}^2\mathbf{1}) \\ &= \alpha^2 \mathrm{tr}\,\mathbf{A}^2 + \beta^2 a^2_{\cdot\cdot} + 2\alpha\beta \textstyle\sum_j a_{\cdot j} a_{j\cdot} . \end{aligned} \tag{9}$$

and for symmetric $\mathbf{A}$ this is

$$\mathrm{tr}[\mathbf{A}(\alpha\mathbf{I} + \beta\mathbf{J})]^2 = \alpha^2 \textstyle\sum_i \sum_j a^2_{ij} + \beta^2 a^2_{\cdot\cdot} + 2\alpha\beta \sum_i a^2_{i\cdot} \; . \tag{10}$$

Examples

$\mathrm{var}(\mathbf{y}'\mathbf{Ay})$ is often wanted when $\mathbf{y}'\mathbf{Ay}$ is a sum of squares, in which case $\mathbf{A}$ frequently has the form $\mathbf{A} = r\mathbf{I}_n + s\mathbf{J}_n$ for scalars r and s. It is therefore of interest to evaluate some of the preceding expressions for this $\mathbf{A}$, wherein $a_{ii} = r + s$, $a_{ij} = s$ for $i \neq j$, and hence $a_{i\cdot} = r + ns$, and $a_{\cdot\cdot} = n(r + ns)$. Hence, on using (6) and (8) respectively,

$$\mathrm{tr}[(r\mathbf{I} + s\mathbf{J})\mathbf{J}]^2 = a^2_{\cdot\cdot} = n^2(r + ns)^2 \text{ , and}$$

$$\mathrm{tr}[(r\mathbf{I} + s\mathbf{J})\mathbf{J}(r\mathbf{I} + s\mathbf{J})]^2 = \left(\textstyle\sum_i a^2_{i\cdot}\right)^2 = n^2(r + ns)^4 \; ;$$

$$\begin{aligned} \mathrm{tr}[(r\mathbf{I} + s\mathbf{J})(\alpha\mathbf{I} + \beta\mathbf{J})]^2 &= \alpha^2 \textstyle\sum_i \sum_j a^2_{ij} + \beta^2 a^2_{\cdot\cdot} + 2\alpha\beta \sum_i a^2_{i\cdot}, \quad \text{from (10)} \\ &= \alpha^2[n(r + s)^2 + n(n - 1)s^2] + \beta^2 n^2 (r + ns)^2 + 2\alpha\beta n(r + ns)^2 \end{aligned}$$

$$= (r + ns)^2(\alpha + n\beta)^2 + (n - 1)r^2\alpha^2 ,$$

after a little reduction.

Results of this nature can also be obtained, as pointed out by C.E. McCulloch, from using

$$\mathrm{tr}(a\mathbf{I}_n + b\mathbf{J}_n) = (a + b)n$$

and

$$(a\mathbf{I}_n + b\mathbf{J}_n)(p\mathbf{I}_n + q\mathbf{J}_n) = ap\mathbf{I}_n + (aq + bp + bqn)\mathbf{J}_n .$$

For example,

$$\mathrm{tr}[(r\mathbf{I} + s\mathbf{J})\mathbf{J}]^2 = \mathrm{tr}[(r + sn)\mathbf{J}]^2 = \mathrm{tr}[(r + sn)^2 n\mathbf{J}] = (r + sn)^2 n^2 .$$

A specific example of this kind of matrix is the variance of MSA = SSA/(a − 1) for

$$\mathrm{SSA} = \sum_{i=1}^{a} n(\bar{y}_{i.} - \bar{y}_{..})^2 = \mathbf{y}'\mathbf{A}\mathbf{y} \quad \text{where} \quad \mathbf{A} = (\mathbf{I}_a - \bar{\mathbf{J}}_a) \otimes \bar{\mathbf{J}}_n$$

with

$$\mathbf{V} = \mathbf{I}_a \otimes \left(\sigma_e^2 \mathbf{I}_n + \sigma_\alpha^2 \mathbf{J}_n\right),$$

where $\otimes$ is the direct product operator and $\bar{\mathbf{J}}_n = \mathbf{J}_n \big/ n$. Then

$$\mathrm{var(SSA)} = 2\,\mathrm{tr}(\mathbf{AV})^2 = 2\,\mathrm{tr}\left[(\mathbf{I}_a - \bar{\mathbf{J}}_a) \otimes \left(\sigma_e^2 \mathbf{I}_n + \sigma_\alpha^2 \mathbf{J}_n\right)\bar{\mathbf{J}}_n\right]^2 .$$

With the general result $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$, this gives

$$\mathrm{var(SSA)} = 2(a - 1)\mathrm{tr}\left[\left(\sigma_e^2 \mathbf{I}_n + \sigma_\alpha^2 \mathbf{J}_n\right)\mathbf{J}_n \big/ n\right]^2 = 2(a - 1)\left(\sigma_e^2 + n\sigma_\alpha^2\right)^2, \tag{11}$$

on using (6). Hence, as is well known,

$$\mathrm{var(MSA)} = \frac{\mathrm{var(SSA)}}{(a - 1)^2} = \frac{2\left(\sigma_e^2 + n\sigma_\alpha^2\right)^2}{a - 1} = \frac{2[\mathrm{E(MSA)}]^2}{a - 1} .$$

In the case of unbalanced data, with

$$\mathbf{A} = \left\{_d \ \bar{\mathbf{J}}_{n_i}\right\}_{i=1}^{a} - \bar{\mathbf{J}}_N \quad \text{for} \quad N = n_. = \sum_{i=1}^{a} n_i$$

and

$$\mathbf{V} = \left\{_d \ \sigma_e^2 \mathbf{I}_{n_i} + \sigma_\alpha^2 \mathbf{J}_{n_i}\right\}_{i=1}^{a}$$

$$\mathrm{tr}(\mathbf{AV})^2 = \mathrm{tr}\left[\left\{_d \left(\sigma_e^2\mathbf{I}_{n_i} + \sigma_\alpha^2\mathbf{J}_{n_i}\right)\mathbf{J}_{n_i} \Big/ n_i\right\}_{i=1}^{a}\right]^2 + \mathrm{tr}\left[\left\{_d \left(\sigma_e^2\mathbf{I}_{n_i} + \sigma_\alpha^2\mathbf{J}_{n_i}\right)\right\}\mathbf{J}_N \Big/ N\right]^2$$

$$- 2\mathrm{tr}\left[\left\{_d\ \sigma_e^2\mathbf{I}_{n_i} + \sigma_\alpha^2\mathbf{J}_{n_i}\right\}\left\{_d\ \bar{\mathbf{J}}_{n_i}\right\}\left\{_d\ \sigma_e^2\mathbf{I}_{n_i} + \sigma_\alpha^2\mathbf{J}_{n_i}\right\}\mathbf{J}_N \Big/ N\right].$$

On using (6) for the first two terms this becomes, where $\sum$ represents $\sum_i$

$$\mathrm{tr}(\mathbf{AV})^2 = \sum\frac{1}{n_i^2}\left(n_i\sigma_e^2 + n_i^2\sigma_\alpha^2\right)^2 + \frac{1}{N^2}\left[\sum\left(n_i\sigma_e^2 + n_i^2\sigma_\alpha^2\right)\right]^2$$

$$- 2\,\mathrm{tr}\left[\left\{_d \left(\sigma_e^4 \big/ n_i + \sigma_\alpha^4 n_i + 2\sigma_e^2\sigma_\alpha^2\right)\mathbf{J}_{n_i}\right\}\mathbf{J}_N \Big/ N\right];$$

and then using (5) gives

$$\mathrm{tr}(\mathbf{AV})^2 = \sum\left(\sigma_e^2 + n_i\sigma_\alpha^2\right)^2 + \left[\sigma_e^2 + \left(\sum n_i^2 \big/ N\right)\sigma_\alpha^2\right]^2$$

$$-(2/N)\sum n_i\left(\sigma_e^4 + n_i^2\sigma_\alpha^4 + 2n_i\sigma_e^2\sigma_\alpha^2\right)$$

$$= \sum\left(\sigma_e^2 + n_i\sigma_\alpha^2\right)^2 + \left[\sigma_e^2 + \left(\sum n_i^2 \big/ N\right)\sigma_\alpha^2\right]^2 - (2/N)\sum n_i\left(\sigma_e^2 + n_i\sigma_\alpha^2\right)^2$$

$$= \sigma_e^4(a + 1 - 2N/N) + 2\sigma_e^2\sigma_\alpha^2\left(N + \sum n_i^2 \big/ N - 2\sum n_i^2 \big/ N\right)$$

$$+ \sigma_\alpha^4\left[\sum n_i^2 + \left(\sum n_i^2\right)^2 \big/ N^2 - 2\sum n_i^3 \big/ N\right]$$

$$= (a-1)\sigma_e^4 + 2\left(N - \sum n_i^2 \big/ N\right)\sigma_e^2\sigma_\alpha^2 + \sigma_\alpha^4\left[\sum n_i^2 + \left(\sum n_i^2\right)^2 \big/ N - 2\sum n_i^3 \big/ N\right].$$

It is easily established that this expression with $n_i = n\ \forall\ i$ does, of course, lead to (11) for balanced data.