Notes for 751-752
Sections 8-9

Martin Lindquist[*]

September 17, 2017

[*]Thanks to Brian Caffo and Ingo Ruczinski for supplying their notes, upon which much of this document is based.

8 Design Matrices of Less Than Full Rank

Recall the least squares criteria:

$$f(\boldsymbol{\beta}) \quad = \quad ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2.$$

The solution is obtained by solving the so-called normal equations:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}.$$

Note that the matrix $\mathbf{X}'\mathbf{X}$ retains the same rank as $\mathbf{X}$. If it is a full rank $p \times p$ matrix and invertible, we can solve the normal equations as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

However, if the $n \times p$ design matrix $\mathbf{X}$ has rank $r < p$, there is not a unique solution $\hat{\boldsymbol{\beta}}$ to the normal equations.

Here we describe three ways to find *a* solution $\hat{\boldsymbol{\beta}}$ and *the* orthogonal projection $\hat{\mathbf{Y}}$:

1. Reducing the model to one of full rank.
2. Finding a generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$.
3. Imposing identifiability constraints.

8.1 Reducing the Model to One of Full Rank

Let $\mathbf{X}_1$ consist of r linearly independent columns from $\mathbf{X}$ and let $\mathbf{X}_2$ consist of the remaining columns. Then $\mathbf{X_2} = \mathbf{X_1F}$ because the columns of $\mathbf{X}_2$ are linearly dependent on the columns of $\mathbf{X}_1$.

$$\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}) = (\mathbf{X_1}, \mathbf{X_1F}) = \mathbf{X_1}(\mathbf{I_{r\times r}}, \mathbf{F}).$$

This is a special case of the rank factorization $\mathbf{X} = \mathbf{KL}$, where $\text{rank}(\mathbf{K}_{n\times r}) = r$ and $\text{rank}(\mathbf{L}_{r\times p}) = r$. Now, we can write: $\mathbf{X}\boldsymbol{\beta} = \mathbf{KL}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\alpha}$.

Since $\mathbf{K}$ has full rank, the least squares estimate of $\boldsymbol{\alpha}$ is $\hat{\boldsymbol{\alpha}} = (\mathbf{K}'\mathbf{K})^{-\mathbf{1}}\mathbf{K}'\mathbf{y}$ and the orthogonal projection is $\hat{\mathbf{y}} = \mathbf{K}\hat{\boldsymbol{\alpha}} = \mathbf{K}(\mathbf{K}'\mathbf{K})^{-\mathbf{1}}\mathbf{K}'\mathbf{y}$. Therefore, $\mathbf{P} = \mathbf{K}(\mathbf{K}'\mathbf{K})^{-\mathbf{1}}\mathbf{K}'$, or equivalently $\mathbf{P} = \mathbf{X_1}(\mathbf{X_1'}\mathbf{X_1})^{-\mathbf{1}}\mathbf{X_1'}$.

Example: 8.1 (One-way ANOVA with 2 groups).

Consider the model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad \text{for } \; j = 1, \ldots n_i; \;\; i = 1, 2$$

or

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

Let $\mathbf{X}_1$ consist of the first 2 columns of $\mathbf{X}$. Then

$$\mathbf{X} = \mathbf{X}_1 \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix},$$

and $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\alpha}$, where

$$\boldsymbol{\alpha} = \begin{pmatrix} \mu + \alpha_2 \\ \alpha_1 - \alpha_2 \end{pmatrix}.$$

Then

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_j y_{1j} + \sum_j y_{2j} \\ \sum_j y_{1j} \end{pmatrix} \\ &= \begin{pmatrix} n_2^{-1} & -n_2^{-1} \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} \end{pmatrix} \begin{pmatrix} \sum_j y_{1j} + \sum_j y_{2j} \\ \sum_j y_{1j} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y}_2 \\ \bar{y}_1 - \bar{y}_2 \end{pmatrix}, \end{aligned}$$

and hence $\hat{\mathbf{y}} = \mathbf{X}_1\hat{\boldsymbol{\alpha}} = (\bar{y}_{1\cdot}, \ldots, \bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \ldots, \bar{y}_{2\cdot})'$.

8.2 Generalized Inverse

Definition: 8.2 For an $m \times n$ matrix $\mathbf{A}$, a generalized inverse of $\mathbf{A}$ is an $n \times m$ matrix $\mathbf{A}^-$ satisfying $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$.

Note that a generalized inverse always exists but is not unique except when $\mathbf{A}$ is nonsingular, in which case $\mathbf{A}^- = \mathbf{A}^{-1}$.

Let $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$, where $\mathbf{X}_1$ consists of r linearly independent columns from $\mathbf{X}$. Then a generalized inverse of $\mathbf{X}'\mathbf{X}$ is

$$(\mathbf{X}'\mathbf{X})^- = \begin{pmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Thus, a solution to the normal equations is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$ with fitted values $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} = \mathbf{P}\mathbf{y}$, where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$. Note that this also gives $\mathbf{P} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$.

This result is a special case of the following theorem.

Theorem: 8.3 Let the matrix $\mathbf{W}_{p\times p}$ have rank r and be partitioned as

$$\mathbf{W} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix},$$

where $\mathbf{A}$ has rank r. Then a generalized inverse of $\mathbf{W}$ is

$$\mathbf{W}^- = \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Example: 8.4 (One-way ANOVA with 2 groups, continued). We have

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{pmatrix}.$$

If $\mathbf{X}_1$ consists of the first 2 columns of $\mathbf{X}$, then

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} \end{pmatrix}.$$

and generalized inverse of $\mathbf{X}'\mathbf{X}$ is

$$(\mathbf{X}'\mathbf{X})^- = \begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Now a solution to the normal equations is

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \sum_j y_{1j} + \sum_j y_{2j} \\ \sum_j y_{1j} \\ \sum_j y_{2j} \end{pmatrix} = \begin{pmatrix} \bar{y}_2 \\ \bar{y}_1 - \bar{y}_2 \\ 0 \end{pmatrix},$$

and $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = (\bar{y}_{1\cdot}, \ldots, \bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \ldots, \bar{y}_{2\cdot})'$, as before.

Definition: 8.5 A matrix $\mathbf{A}^+$ satisfying the following conditions is called the Moore-Penrose inverse:

1. $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
2. $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$
3. $(\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A}$
4. $(\mathbf{A}\mathbf{A}^+)' = \mathbf{A}\mathbf{A}^+$

$\mathbf{A}^+$ is unique. Using the Moore-Penrose inverse provides the minimum norm solution to the least squares problem. Note it is sometimes referred to as the pseudo-inverse.

8.3 Imposing Identifiability Constraints

A final approach is to impose $s = p - r$ constraints on $\boldsymbol{\beta}$ in order to make $\boldsymbol{\beta}$ uniquely determined (identifiable), i.e. such that for any $\boldsymbol{\theta} \in \mathcal{R}(\mathbf{X})$, there is a unique $\boldsymbol{\beta}$ satisfying

$$\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\theta} \quad \text{and} \quad \mathbf{H}\boldsymbol{\beta} = \mathbf{0}.$$

This can be written

$$\begin{pmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix} \boldsymbol{\beta} \equiv \mathbf{G}\boldsymbol{\beta}.$$

Now when is there a unique solution?

Theorem: 8.6 A unique solution exists if and only if $\mathbf{G}$ has rank p and the rows of $\mathbf{H}$ are linearly independent of the rows of $\mathbf{X}$.

Theorem: 8.7 A unique solution exists if and only if $\mathbf{G}$ has rank p and $\mathbf{H}$ has rank $p - r$.

To estimate $\boldsymbol{\beta}$, we solve $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\mathbf{H}\hat{\boldsymbol{\beta}} = \mathbf{0}$, i.e. we solve the augmented normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ and $\mathbf{H}'\mathbf{H}\hat{\boldsymbol{\beta}} = \mathbf{0}$. This gives us $(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})\hat{\boldsymbol{\beta}} = (\mathbf{G}'\mathbf{G})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. Therefore,

$$\hat{\boldsymbol{\beta}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{X}'\mathbf{y}, \text{ and } \hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{y}, \text{ where } \mathbf{P} = \mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{X}'.$$

Example: 8.8 (One-way ANOVA with 2 groups, cont.)

Set the constraint $\alpha_1 + \alpha_2 = 0$, i.e.

$$\mathbf{H}\boldsymbol{\beta} \equiv (0, 1, 1)\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = 0.$$

Suppose $n_1 = n_2 = m$. Then it can be shown that

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \bar{Y}_{\cdot\cdot} \\ \frac{1}{2}(\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}) \\ \frac{1}{2}(\bar{Y}_{2\cdot} - \bar{Y}_{1\cdot}) \end{pmatrix}$$

satisfies the normal equations, and clearly satisfies the constraint $\alpha_1 + \alpha_2 = 0$. Therefore, we have as before $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = (\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \ldots, \bar{Y}_{2\cdot})'$.

9 Random Vectors and Matrices

Up to this point, our exploration of linear models only relied on least squares and projections. We now begin now discussing the statistical properties of our estimators. We start by defining expected values. We assume that the reader has basic univariate mathematical statistics.

9.1 Random Vectors and Matrices

A random vector is defined as a vector of random variables, i.e.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

Similarly, a random matrix is a matrix of random variables $\mathbf{Z} = (Z_{ij})$.

9.2 Expected values

If X is a random variable having density funciton f, the k^{th} moment is defined as

$$E[X] = \int_{-\infty}^{\infty} x^k f(x) dx.$$

In the multivariate case where $\mathbf{X}$ is a random vector then the k^{th} moment of element i of the vector is given by

$$E[X_i^k] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i^k f(x_1, \ldots, x_n) dx_1, \ldots, dx_n.$$

It is worth asking if this definition is consistent with all of the subdistributions defined by the subvectors of $\mathbf{X}$. Let $i_1, \ldots, i_p$ is any subset of indices of $1, \ldots, n$ and $i_{p+1}, \ldots, i_n$ are the remaining, then the joint distribution of $(X_{i_1}, \ldots, X_{i_p})^t$ is

$$g(x_{i_1}, \ldots, x_{i_p}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_{i_{p+1}}, \ldots, dx_{i_n}.$$

The k^{th} moment of X_{i_j} for $j \in \{1, \ldots, p\}$ is equivalently:

$$\begin{aligned} E[X_{i_j}] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_{i_j}^k g(x_{i_1}, \ldots, x_{i_p}) dx_{i_1}, \ldots, d_{x_{ip}} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_{i_j}^k f(x_1, \ldots, x_n) dx_1, \ldots, dx_n. \end{aligned}$$

(HW, prove this.) Thus, if we know only the marginal distribution of X_{i_j} or any level of joint information, the expected value is the same.

The mean or expectation of a random vector $\mathbf{X}$ is defined as

$$E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}.$$

Similarly, the mean or expectation of a random vector $\mathbf{Z}$ is defined as $E[\mathbf{Z}] = (E[Z_{ij}])$. Thus, if $\mathbf{X}$ is any random vector or matrix, then $E[\mathbf{X}]$ is simply the elementwise expected value defined above. Often we will write $E[\mathbf{X}] = \boldsymbol{\mu}$, or some other Greek letter, adopting the notation that population parameters are Greek.

The rules of expected values in the univariate setting translate well to the multivariate settings. For example:

- A constant vector $\mathbf{a}$ (i.e., a vector of constants) and a constant matrix $\mathbf{A}$ (i.e., a matrix of constants) satisfy $E[\mathbf{a}] = \mathbf{a}$ and $E[\mathbf{A}] = \mathbf{A}$.
- $E[\mathbf{X} + \mathbf{Y}] = E[\mathbf{X}] + E[\mathbf{Y}]$
- $E[\mathbf{AX}] = \mathbf{A}E[\mathbf{X}]$ if $\mathbf{A}$ is a constant matrix.
- $E[\mathbf{AZB} + \mathbf{C}] = \mathbf{A}E[\mathbf{Z}]\mathbf{B} + \mathbf{C}$ if $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are constant matrices.
- $E[\mathbf{X}'] = E[\mathbf{X}]'$
- $E[\text{tr}(\mathbf{X})] = \text{tr}(E[\mathbf{X}])$

9.3 Variance

The multivariate variance of random vector $\mathbf{X}$ is defined as

$$\text{var}(\mathbf{X}) = \boldsymbol{\Sigma} = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t].$$

Alternatively, we can write it as follows: defined as

$$\text{var}(\mathbf{X}) \equiv [\text{cov}(X_i, X_j)] \equiv \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{var}(X_n) \end{pmatrix}.$$

Note that this term is alternatively, written as the variance, covariance matrix, or variance-covariance matrix.

Example: 9.1 If $X_1, \ldots, X_n$ are independent, then the covariances are 0 and the covariance matrix is equal to $\operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, or $\sigma^2 \mathbf{I}_n$ if the X_i have common variance σ^2.

Properties of variance-covariance matrices:

- $\operatorname{var}(\mathbf{X}) = [\operatorname{var}(\mathbf{X})]'$.
- $\operatorname{var}(\mathbf{X} + \mathbf{a}) = \operatorname{var}(\mathbf{X})$ if $\mathbf{a}$ is a constant vector.
- $\operatorname{var}(\mathbf{AX}) = \mathbf{A}\operatorname{var}(\mathbf{X})\mathbf{A}'$ if $\mathbf{A}$ is a constant matrix.
- $\operatorname{var}(\mathbf{X})$ is positive semidefinite.
- $\operatorname{var}(\mathbf{X})$ is positive definite provided no linear combination of the X_i is a constant.
- $\operatorname{var}(\mathbf{X}) = E[\mathbf{XX}'] - E[\mathbf{X}](E[\mathbf{X}])'$

9.4 Multivariate covariances

The multivariate covariance is given by

$$\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = E[(\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)'] = E[\mathbf{XY}'] - \boldsymbol{\mu}_x \boldsymbol{\mu}_y'.$$

Alternatively, we can write

$$\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = [\operatorname{cov}(X_i, Y_j)] \equiv \begin{pmatrix} \operatorname{cov}(X_1, Y_1) & \operatorname{cov}(X_1, Y_2) & \cdots & \operatorname{cov}(X_1, Y_n) \\ \operatorname{cov}(X_2, Y_1) & \operatorname{cov}(X_2, Y_2) & \cdots & \operatorname{cov}(X_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_m, Y_1) & \operatorname{cov}(X_m, Y_2) & \cdots & \operatorname{cov}(X_m, Y_n) \end{pmatrix}.$$

Note that this definition applies even if $\mathbf{x}$ and $\mathbf{y}$ are of different length. Further notice the multivariate covariance is not symmetric in its arguments. Moreover,

$$\operatorname{cov}(\mathbf{X}, \mathbf{X}) = \operatorname{var}(\mathbf{X}).$$

If $\mathbf{A}$ and $\mathbf{B}$ are constant matrices, then $\operatorname{cov}(\mathbf{AX}, \mathbf{BY}) = \mathbf{A} \operatorname{cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}'$. In addition, it further holds that $\operatorname{cov}(\mathbf{X} + \mathbf{Y}, \mathbf{Z}) = \operatorname{cov}(\mathbf{X}, \mathbf{Z}) + \operatorname{cov}(\mathbf{Y}, \mathbf{Z})$.

Theorem: 9.2 Let $\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$. Then $\operatorname{cov}(\mathbf{Z}) = \begin{pmatrix} \operatorname{var}(\mathbf{X}) & \operatorname{cov}(\mathbf{X}, \mathbf{Y}) \\ \operatorname{cov}(\mathbf{Y}, \mathbf{X}) & \operatorname{var}(\mathbf{Y}) \end{pmatrix}$.

Example: 9.3 If $X_1, \ldots, X_n$ are exchangeable, they have a constant variance σ^2 and a constant correlation ρ between any pair of variables. Thus

$$\text{cov}(\mathbf{X}) = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

This is sometimes called an exchangeable covariance matrix.

Multivariate covariances are useful for sums of random vectors.

$$\text{var}(\mathbf{X} + \mathbf{Y}) = \text{var}(\mathbf{X}) + \text{var}(\mathbf{Y}) + \text{cov}(\mathbf{X}, \mathbf{Y}) + \text{cov}(\mathbf{Y}, \mathbf{X}).$$

A nifty fact from covariances is that the covariance of $\mathbf{AX}$ and $\mathbf{BX}$ is $\mathbf{A\Sigma B}'$. Thus $\mathbf{AX}$ and $\mathbf{BX}$ are uncorrelated iff $\mathbf{A\Sigma B}' = \mathbf{0}$.

9.5 Correlation matrix

The correlation matrix of $\mathbf{x}$ is defined as

$$\text{corr}(\mathbf{X}) = [\text{corr}(X_i, X_j)] \equiv \begin{pmatrix} 1 & \text{corr}(X_1, X_2) & \cdots & \text{corr}(X_1, X_n) \\ \text{corr}(X_2, X_1) & 1 & \cdots & \text{corr}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{corr}(X_n, X_1) & \text{corr}(X_n, X_2) & \cdots & 1 \end{pmatrix}.$$

If we denote $\text{cov}(\mathbf{X})$ by $\mathbf{\Sigma} = (\sigma_{ij})$, then the correlation matrix and covariance matrix are related by

$$\text{cov}(\mathbf{X}) = \text{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}}) \times \text{corr}(\mathbf{X}) \times \text{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}}).$$

This is easily seen using $\text{corr}(X_i, X_j) = \text{cov}(X_i, X_j)/\sqrt{\sigma_{ii}\sigma_{jj}}$.

9.6 Quadratic form moments

Theorem: 9.4 Let $E[\mathbf{X}] = \boldsymbol{\mu}$ and $\text{cov}(\mathbf{X}) = \mathbf{\Sigma}$ and $\mathbf{A}$ be a constant matrix. Then

$$E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] = \text{tr}(\mathbf{A\Sigma}).$$

Theorem: 9.5 $E[\mathbf{X}'\mathbf{A}\mathbf{X}] = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

Example: 9.6 Let $X_1, \ldots, X_n$ be independent random variables with common mean μ and variance σ^2. Then the sample variance $S^2 = \sum_i (X_i - \bar{X})^2/(n-1)$ is an unbiased estimate of σ^2.