APPENDIX 1

PROPERTIES OF MATRICES

Properties of matrices are reported in this appendix. Familiarity with the elementary algebra of matrices is assumed, and properties are stated but not proved. For proofs of most of the propositions, consult Graybill (1969).

Definitions

A *matrix* is a two-dimensional array of numbers denoted by capital boldface letters such as **A**. If the array has only one column, it is called a *vector* and is indicated (usually) by lowercase boldface letters such as **a**. The *ij*th element of **A** is indicated by A_{ij}, and the symbol $\{A_{ij}\}$ stands for a matrix with elements A_{ij}. An $m \times n$ matrix has m rows and n columns.

(D1) If **A** and **B** are $m \times n$ matrices, the *sum* of **A** and **B** is defined by

$$\mathbf{A}+\mathbf{B}=\{A_{ij}+B_{ij}\}.$$

(D2) If **A** is $m \times n$ and **B** is $n \times k$, then the *product* of **A** and **B** is defined by

$$\mathbf{AB}=\left\{\sum_{p=1}^{n} A_{ip}B_{pj}\right\}$$

(D3) *Multiplication* of a matrix **A** by a *scalar* α is defined by

$$\alpha\mathbf{A}=\{\alpha A_{ij}\}.$$

(D4) The *transpose* of a matrix $\{A_{ij}\}$ is the matrix

$$\mathbf{A}'=\{A_{ji}\}$$

(D5) A matrix is *symmetric* if it is equal to its transpose $\mathbf{A}=\mathbf{A}'$.

(D6) A *diagonal matrix* is a square matrix with zeroes off the main diagonal, $A_{ij}=0$, for $i \neq j$. It is indicated by $\mathbf{A}=\text{diag}\{d_1,d_2,\ldots,d_n\}$ where $A_{ii}=d_i$.

(D7) An *identity matrix* of order p is the $p \times p$ diagonal matrix

$$\mathbf{I}_p=\text{diag}\{1,1,\ldots,1\}.$$

If the order is obvious, the notation **I** suffices.

(D8) The notation $\mathbf{1}_n$ indicates an $n \times 1$ *vector of ones*. Where convenient, the subscript n is suppressed.

(D9) The *trace* of a square $n \times n$ matrix **A** is the sum of its diagonal elements

$$\text{tr}\mathbf{A}=\sum_{i=1}^{n} A_{ii}$$

(D10) Matrices **A** and **B** are said to *commute* if $\mathbf{AB}=\mathbf{BA}$.

(D11) A real symmetric matrix **A** is said to be *positive definite* if for any vector $\mathbf{x}\neq\mathbf{0}$, $\mathbf{x}'\mathbf{A}\mathbf{x}>0$, and is said to be *positive semi-definite* if $\mathbf{x}'\mathbf{A}\mathbf{x}=0$ for at least one $\mathbf{x}\neq\mathbf{0}$ and $\mathbf{x}'\mathbf{A}\mathbf{x}>0$ otherwise. (*Negative definite* and *negative semi-definite* are defined analogously.)

(D12) A square matrix **A** is said to be *orthogonal* if $\mathbf{AA}'=\mathbf{I}$. (As a consequence, $\mathbf{A}'=\mathbf{A}^{-1}$ and $\mathbf{A}'\mathbf{A}=\mathbf{I}$.)

(D13) A square symmetric matrix **A** is said to be *idempotent* if $\mathbf{AA}=\mathbf{A}$. (Sometimes symmetry is not included in the definition.)

(D14) The determinant of an $n \times n$ matrix **A** is

$$|\mathbf{A}|=\sum(\pm)A_{1i_1}A_{2i_2}\cdots A_{ni_n}$$

where $(i_1,i_2,\ldots,i_n)$ is a permutation of the first n integers, the summation extends over all $n!$ permutations, and the sign is + if the permutation is even or − if the permutation is odd.

(D15) The *inverse* of a square matrix **A** is a matrix $\mathbf{A}^{-1}$ such that $\mathbf{AA}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$.

ι 1

of a matrix is the number of linearly independent number of linearly independent rows).

natrix **A** is said to be *nonsingular* if and only if the rank

$\times n$) and **B** is $(p \times q)$ the *Kronecker product* of **A** and **B** is rix

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \dots & A_{1n}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22} & B & A_{2n}\mathbf{B} \\ \vdots & & \vdots & \\ A_{m1}\mathbf{B} & A_{m2}\mathbf{B} & \dots & A_{mn}\mathbf{B} \end{bmatrix}$$

ɔperties are easily verifiable, assuming that the operations

$$\operatorname{tr}\mathbf{AB} = \operatorname{tr}\mathbf{BA}$$

$$\operatorname{tr}(\mathbf{A}+\mathbf{B}) = \operatorname{tr}\mathbf{A} + \operatorname{tr}\mathbf{B}$$

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$$

$$|\alpha\mathbf{A}| = \alpha^n|\mathbf{A}| \qquad (\mathbf{A} \text{ is } (n \times n))$$

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$\mathbf{BCB}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1}$$

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{B}^{-1}$$

ratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ be indicated by $Q(\mathbf{x}, \mathbf{A})$; then assuming **H** ɔle

$$\cdot \mathbf{b}), \mathbf{H}) + Q((\boldsymbol{\beta} - \mathbf{b}^*), \mathbf{H}^*)$$

$$)(\boldsymbol{\beta} - \mathbf{b}^{**}, \mathbf{H}+\mathbf{H}^*) + Q(\mathbf{b}-\mathbf{b}^*, \mathbf{H}^*(\mathbf{H}+\mathbf{H}^*)^{-1}\mathbf{H})$$

where $\quad \mathbf{b}^{**} = (\mathbf{H}+\mathbf{H}^*)^{-1}(\mathbf{H}\mathbf{b} + \mathbf{H}^*\mathbf{b}^*)$.

Theorems concerning the rank of a matrix are

(T11) $\qquad \operatorname{rank}(\mathbf{AB}) \leqslant \min[\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})]$

(T12) $\qquad \operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}') = \operatorname{rank}(\mathbf{AA}')$.

The existence of an inverse is implied by the following theorem.

(T13) If **A** is an $n \times n$ matrix, its inverse exists if and only if $|\mathbf{A}| \neq 0$.

A square matrix may be partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}$$

where **E** and **H** are themselves square. Then

(T14) $$\begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{E}^{-1}(\mathbf{I} + \mathbf{F}\mathbf{C}^{-1}\mathbf{G}\mathbf{E}^{-1}) & -\mathbf{E}^{-1}\mathbf{F}\mathbf{C}^{-1} \\ -\mathbf{C}^{-1}\mathbf{G}\mathbf{E}^{-1} & \mathbf{C}^{-1} \end{bmatrix}$$

where $\mathbf{C} = \mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F}$.

(T15) $$\left|\begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}\right| = |\mathbf{E}||\mathbf{H} - \mathbf{G}\mathbf{E}^{-1}\mathbf{F}| = |\mathbf{H}||\mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{G}|.$$

An implication of (T15) is $|\mathbf{E} + \mathbf{x}\mathbf{x}'| = |\mathbf{E}|(1 + \mathbf{x}'\mathbf{E}^{-1}\mathbf{x})$ for **E** a matrix and **x** a vector.

The following theorems apply to positive definite (p.d.) and positive semi-definite (p.s.d.) matrices

(T16) If **A** is $n \times n$ p.d., **P** is $n \times m$ with rank $m \leqslant n$, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is p.d.

(T17) If **A** is p.d., then $\mathbf{A}^{-1}$ is p.d.

(T18) If **A** is p.d., then $|\mathbf{A}| > 0$.

(T19) If **A** is p.d., and **B** p.s.d., then $\mathbf{A}+\mathbf{B}$ is p.d.

The following results apply to idempotent matrices.

(T20) If **A** is idempotent, the values of λ satisfying $|\mathbf{A} - \lambda\mathbf{I}| = 0$ are either one or zero.

(T21) If **A** is idempotent, $\operatorname{tr}(\mathbf{A}) = \operatorname{rank}(\mathbf{A})$.

On the assumption that the operations are well defined, the Kronecker

‹ 1

llowing properties:

$$\mathbf{A}\otimes(\mathbf{B}+\mathbf{C})=\mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}$$

$$(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C}=\mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C})$$

$$(\mathbf{A}\otimes\mathbf{B})'=\mathbf{A}'\otimes\mathbf{B}'$$

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=\mathbf{AC}\otimes\mathbf{BD}$$

$$(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$$

$$\mathbf{B}|=|\mathbf{A}|^m|\mathbf{B}|^n, \qquad (\mathbf{A} \text{ is } (n\times n), \mathbf{B} \text{ is } (m\times m))$$

$$\mathrm{tr}|\mathbf{A}\otimes\mathbf{B}|=\mathrm{tr}(\mathbf{A})\cdot\mathrm{tr}(\mathbf{B})$$

iation

a scalar function of the vector **x**, and let the vector of with respect to **x** be denoted by

$$\frac{\partial y}{\partial \mathbf{x}}=\left\{\frac{\partial y}{\partial x_i}\right\}.$$

a vector function of the vector **x**, and let the *matrix* of with respect to **x** be

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}=\left\{\frac{\partial y_i}{\partial x_j}\right\}.$$

r function of the matrix **A**, and let the *matrix* of derivatives ct to **A** be

$$\frac{\partial y}{\partial \mathbf{A}}=\left\{\frac{\partial y}{\partial A_{ij}}\right\}.$$

rix function of the scalar t, and let the matrix of derivatives ect to t be

$$\frac{\partial \mathbf{A}}{\partial t}=\left\{\frac{\partial A_{ij}}{\partial t}\right\}.$$

ng formulas taken from Dwyer (1967), are straightforwardly **A** and **B** are matrices, **x** is a vector, and t is a scalar.

(T29) $$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}}=\mathbf{A}$$

(T30) $$\frac{\partial \mathbf{x}'A\mathbf{x}}{\partial \mathbf{x}}=2\mathbf{Ax} \qquad (\text{with } \mathbf{A}=\mathbf{A}')$$

(T31) $$\frac{\partial\, \mathrm{tr}\,\mathbf{A}}{\partial \mathbf{A}}=\mathbf{I}$$

(T32) $$\frac{\partial |\mathbf{A}|^r}{\partial \mathbf{A}}=r|\mathbf{A}|^r(\mathbf{A}^{-1})'$$

(T33) $$\frac{\partial\, 1n|\mathbf{A}|}{\partial \mathbf{A}}=(\mathbf{A}^{-1})'$$

(T34) $$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{A}}=\mathbf{xx}' \qquad (\mathbf{A} \text{ symmetric})$$

(T35) $$\frac{\partial \mathbf{xA}^{-1}\mathbf{x}}{\partial \mathbf{A}}=-\mathbf{A}^{-1}\mathbf{xx}'\mathbf{A}^{-1} \qquad (\mathbf{A} \text{ symmetric})$$

(T36) $$\frac{\partial\, \mathrm{tr}\,\mathbf{AB}}{\partial \mathbf{A}}=\mathbf{B}'$$

(T37) $$\frac{\partial \mathbf{AB}}{\partial t}=\mathbf{A}\frac{\partial \mathbf{B}}{\partial t}+\frac{\partial \mathbf{A}}{\partial t}\mathbf{B}$$

(T38) $$\frac{\partial \mathbf{A}^{-1}}{\partial t}=-\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{-1}$$

Gradients, Normals, and Tangent Hyperplanes

Let $f(\mathbf{x})$ be a scalar valued function of the vector **x**. The differential of f is

$$df=\sum_i \frac{\partial f}{\partial x_i}dx_i$$
$$=\left(\frac{\partial f}{\partial \mathbf{x}}\right)'d\mathbf{x}.$$

Setting this differential to zero and solving for $d\mathbf{x}$ determines a direction in which there is no differential change in the function f. Geometrically, the equation $0=(\mathbf{x}-\mathbf{x}_0)'(\partial f/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_0})$ thus defines a hyperplane tangent to the surface $f(\mathbf{x})=f(\mathbf{x}_0)$ at $\mathbf{x}_0$. The direction $\partial f/\partial \mathbf{x}$ is orthogonal to this hyperplane and is called the *gradient vector* of the function or the (inward or outward) normal of the surface. The quadratic form $f(\mathbf{x})=\mathbf{x}'\mathbf{Ax}$ has the ε 1

e at $\mathbf{x}_0$ given by $\mathbf{x}_0'\mathbf{A}\mathbf{x}_0=\mathbf{x}'\mathbf{A}\mathbf{x}_0$, and $\mathbf{Ax}$ is the normal of

Ellipsoids

f this section is illustrated in Figure A1: If $\mathbf{A}$ is a real ve definite matrix, the equation $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ defines an irections of tangency between a unit sphere and an $_2$ in the figure, are the eigenvectors of the matrix $\mathbf{A}$, and r axes of the ellipsoid. The corresponding eigenvalues are eeded to produce a tangency between the unit sphere and $=r^2$. The relative eigenvalues are just the relative lengths es of the ellipse, the longer axis having the smaller

square $(k\times k)$ matrix, $\mathbf{x}$ is a $(k\times 1)$ vector of variables, calar, then $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ is the equation of a *quadratic surface*. $\mathbf{A}=\mathbf{I}_k$, then $\mathbf{x}'\mathbf{A}\mathbf{x}=\mathbf{x}'\mathbf{x}=r^2$ is the equation of a sphere igin with radius r. A *family of concentric quadratic surfaces* ices $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ for a given $\mathbf{A}$ and for any $r\geq 0$.
of tangencies between a family of concentric quadratic e unit sphere $\mathbf{x}'\mathbf{x}=1$ can be found by maximizing or

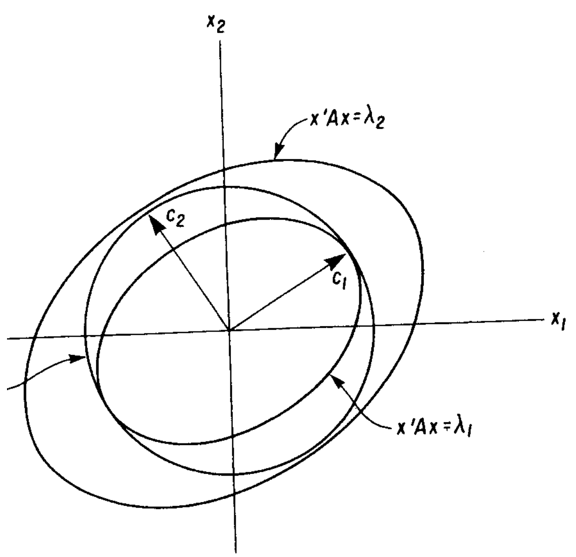

Fig. A.1 Eigenvectors and eigenvalues.

minimizing the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ subject to the constraint $\mathbf{x}'\mathbf{x}=1$. This is a simple Lagrangian problem:

$$\mathbf{0}=\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x}-\lambda[\mathbf{x}'\mathbf{x}-1])}{\partial\mathbf{x}}=2\mathbf{A}\mathbf{x}-2\lambda\mathbf{x},$$

where λ is the Lagrange multiplier. A tangency direction is thus a direction $\mathbf{x}$ satisfying $\mathbf{A}\mathbf{x}=\lambda\mathbf{x}$, or

$$(\mathbf{A}-\lambda\mathbf{I}_k)\mathbf{x}=\mathbf{0}_k.$$

If the matrix $(\mathbf{A}-\lambda\mathbf{I}_k)$ were invertible, the only solution to this set of equations would be $\mathbf{x}=(\mathbf{A}-\lambda\mathbf{I}_k)^{-1}\mathbf{0}_k=\mathbf{0}_k$. Thus a condition for a nontrivial solution is that this matrix is not invertible and hence that the determinant is zero:

$$|\mathbf{A}-\lambda\mathbf{I}_k|=\sum_{i=0}^{k}p_i\lambda^i=0.$$

This polynomial in λ of degree k is known as the *characteristic equation* of the matrix $\mathbf{A}$, and the roots of the polynomial are known as the *characteristic* (or eigen, or latent) *values* of the matrix $\mathbf{A}$. If λ_i is a root of the polynomial $|\mathbf{A}-\lambda\mathbf{I}_k|=0$, the tangency direction $\mathbf{c}_i$ satisfying $(\mathbf{A}-\lambda_i\mathbf{I}_k)\mathbf{c}_i=\mathbf{0}_k$ and normalized such that $\mathbf{c}_i'\mathbf{c}_i=1$, is known as a *characteristic* (or eigen, or latent) *vector* of the matrix $\mathbf{A}$.

Example. Let

$$\mathbf{A}=\begin{bmatrix}4 & 2\\ 2 & 4\end{bmatrix}.$$

The characteristic polynomial is $|\mathbf{A}-\lambda\mathbf{I}_2|=(4-\lambda)^2-4$, with roots $\lambda_1=2$, $\lambda_2=6$. The characteristic vector corresponding to $\lambda_1=2$ is a solution to

$$\begin{bmatrix}4-2 & 2\\ 2 & 4-2\end{bmatrix}\begin{bmatrix}c_{11}\\ c_{12}\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}.$$

Thus $c_1'=(1,-1)/\sqrt{2}$, and similarly $c_2'=(1,1)/\sqrt{2}$.

If a matrix $\mathbf{B}$ is not symmetric, there is a symmetric matrix $\mathbf{A}$ such that $\mathbf{x}'\mathbf{B}\mathbf{x}=\mathbf{x}'\mathbf{A}\mathbf{x}$ for all $\mathbf{x}$. Merely set $A_{ij}=A_{ji}=(B_{ij}+B_{ji})/2$. Thus for the analysis of quadratic surfaces we need consider only symmetric matrices.

THEOREM 30. *Given a symmetric matrix* $\mathbf{A}$, *the characteristic vectors corresponding to unequal characteristic values are orthogonal.*

Proof: $0=\mathbf{c}_i'\mathbf{A}\mathbf{c}_j-\mathbf{c}_i'\mathbf{A}\mathbf{c}_j=\mathbf{c}_i'\mathbf{A}\mathbf{c}_j-\mathbf{c}_j'\mathbf{A}'\mathbf{c}_i=\mathbf{c}_i'\mathbf{A}\mathbf{c}_j-\mathbf{c}_j'\mathbf{A}\mathbf{c}_i=\mathbf{c}_i'\mathbf{c}_j\lambda_j-\mathbf{c}_j\mathbf{c}_i=(\lambda_j-\lambda_i)\mathbf{c}_i'\mathbf{c}_j$. Thus $\lambda_j\neq\lambda_i$ implies $\mathbf{c}_i'\mathbf{c}_j=0$.

ssociated with multiple roots occurs, for example, when on $\mathbf{x}$ is a direction of tangency between the unit sphere mily of concentric spheres $\mathbf{x}'\mathbf{x}=r^2$. Thus any vector $\mathbf{x}$ is a tor. [The characteristic polynomial in this case is $(\lambda-1)^k$ roots all equal to one.] Since any vector is a characteristic ɔle to select a set of k orthogonal characteristic vectors. In ence of multiple roots implies some freedom in choosing ctors, but it is always possible to choose them to be ıere are no multiple roots, then the set of characteristic :.

The characteristic values of a real symmetric matrix are

: that the characteristic polynomial has the complex root with characteristic vector $\mathbf{c}=\mathbf{x}+\mathbf{y}(-1)^{1/2}$. Equating the ex parts of $\mathbf{A}(\mathbf{x}+\mathbf{y}(-1)^{1/2})=(\mathbf{x}+\mathbf{y}(-1)^{1/2})(a+b(-1)^{1/2})$ $-\mathbf{y}b$ and $\mathbf{A}\mathbf{y}=\mathbf{y}a+\mathbf{x}b$. Premultiplying these two expres- $\mathbf{x}'$ and then subtracting yields $0=\mathbf{y}'\mathbf{x}a-\mathbf{y}'\mathbf{y}b-\mathbf{x}'\mathbf{y}a-\mathbf{x}'\mathbf{x}b$, which implies $b=0$; thus λ is real. of the equation $\mathbf{A}\mathbf{c}_i=\mathbf{c}_i\lambda_i$, we have $\lambda_i=\mathbf{c}_i'\mathbf{A}\mathbf{c}_i/\mathbf{c}_i'\mathbf{c}_i=c_i'\mathbf{A}\mathbf{c}_i$.

:. *If* $\mathbf{A}$ *is positive definite, all of its characteristic roots are* $\mathbf{A}$ *is positive semi-definite, all its characteristic roots are*

ɛ can be written as a linear combination of the k characteris- $\Sigma_i \mathbf{c}_i z_i=\mathbf{C}\mathbf{z}$, where $\mathbf{C}$ is a $(k\times k)$ matrix whose columns are rs of $\mathbf{A}$. The quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ can then be written as

$$\mathbf{x}'\mathbf{A}\mathbf{x}=\mathbf{z}'\mathbf{C}'\mathbf{A}\mathbf{C}\mathbf{z}=\mathbf{z}'\Lambda\mathbf{z},$$

liagonal matrix with the eigenvalues λ_i on the diagonal. The $\mathbf{C}=\Lambda$ follows from the fact that $\mathbf{c}_j'\mathbf{A}\mathbf{c}_i=\mathbf{c}_j'\mathbf{c}_i\lambda_i$. It is also easy $\mathbf{C}'\mathbf{C}=\mathbf{I}_k$. Collecting these results together, we have the ›rem:

3. *For any* $k\times k$ *real symmetric matrix* $\mathbf{A}$ *there exists a* $k\times k$ $\mathbf{C}$ *such that* $\mathbf{C}'\mathbf{A}\mathbf{C}=\Lambda$ *where* Λ *is a diagonal matrix and* $\mathbf{C}$ *is* $\mathbf{C}'\mathbf{C}=\mathbf{I}_k$.

nns of $\mathbf{C}$ are used as basis vectors then the quadratic surface

$\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ can be expressed in its *canonical form*

$$\mathbf{z}'\Lambda\mathbf{z}=\sum_i \lambda_i z_i^2=r^2$$

where $\mathbf{x}=\mathbf{C}\mathbf{z}$. If the eigenvalues λ_i are all positive, this is the equation of an ellipsoid centered at the origin with axes of length $r/\lambda_i^{1/2}$. An axis is taken by the transformation $\mathbf{x}=\mathbf{C}\mathbf{z}$ into a column of $\mathbf{C}$, that is, into an eigenvector.

The principal geometric result of this section can now be restated. If $\mathbf{A}$ is a square symmetric positive definite matrix, the quadratic surface $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ is an ellipsoid located at the origin with axes equal to $\mathbf{c}_i r/(\lambda_i)^{1/2}$ where $\mathbf{c}_i$ is a characteristic vector of $\mathbf{A}$ and λ_i is the corresponding characteristic value.

The following result describes the projection of an ellipsoid onto an axis:

THEOREM 34 *The extreme values of the function* $\psi'\mathbf{x}$ *evaluated over the surface* $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ *(where* $\mathbf{A}$ *is invertible) occur at the points* $\mathbf{x}=\pm\mathbf{A}^{-1}\psi\sqrt{r^2/\psi'\mathbf{A}^{-1}\psi}$. *The function at these points takes on the values* $r\sqrt{\psi'\mathbf{A}^{-1}\psi}$. *(A corollary is that the orthogonal projection of the surface* $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ *onto the ith axis is the interval* $|x_i|\leqslant[\mathbf{A}^{-1}]_{ii}^{\frac{1}{2}}r$.)

Proof: Maximization of $\psi'\mathbf{x}$ subject to the constraint $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ leads to the Lagrangian derivatives $\mathbf{0}=\psi+\lambda\mathbf{A}\mathbf{x}$ where λ is the Lagrange multiplier. Solving for $\mathbf{x}$ yields $\mathbf{x}=-\lambda^{-1}\mathbf{A}^{-1}\psi$, which can be used to determine λ: $r^2=\mathbf{x}'\mathbf{A}\mathbf{x}=\psi'\mathbf{A}^{-1}\psi\lambda^{-2}$. Thus $\lambda^2=\psi'\mathbf{A}^{-1}\psi/r^2$, and the two $\mathbf{x}$ vectors are $\mathbf{x}=\pm\mathbf{A}^{-1}\psi r\sqrt{(\psi'\mathbf{A}^{-1}\psi)^{-1}}$.

Conjugate Axes

The coordinate system of the eigenvectors of a symmetric positive definite matrix $\mathbf{A}$ is not necessarily the only coordinate system in which $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$ assumes the canonical form of an ellipsoid. The transformation $\mathbf{z}=\mathbf{P}^{-1}\mathbf{x}$ takes $\mathbf{x}'\mathbf{A}\mathbf{x}$ into $\mathbf{z}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{z}$, which is the canonical form of an ellipsoid if $\mathbf{P}'\mathbf{A}\mathbf{P}$ is a positive-diagonal matrix. Indicating a column of $\mathbf{P}$ by $\mathbf{P}_i$, the matrix $\mathbf{P}'\mathbf{A}\mathbf{P}$ is a positive-diagonal matrix if $\mathbf{P}_i'\mathbf{A}\mathbf{P}_j=0$ for $i\neq j$. Such a $\mathbf{P}$ can be constructed in the following way. Choose any vector $\mathbf{P}_1$ to be the first column. For $\mathbf{P}_2$ choose any vector satisfying $\mathbf{P}_2'\mathbf{A}\mathbf{P}_1=0$, which from the discussion above requires that $\mathbf{P}_2$ be in the hyperplane tangent to the ellipsoid at $\mathbf{P}_1$. Next find a $\mathbf{P}_3$ in the intersection of the tangent hyperplanes at $\mathbf{P}_1$ and $\mathbf{P}_2$, etc. Such a sequence of $\mathbf{P}_i$ vectors is called a *set of conjugate axes* of the ellipsoid $\mathbf{x}'\mathbf{A}\mathbf{x}=r^2$, and is illustrated in Figure A2.

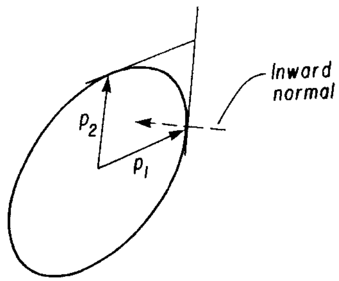

Fig. A.2 Conjugate axes.

te Axes

ds, $\mathbf{x}'\mathbf{A}\mathbf{x}=r_1^2$ and $\mathbf{x}'\mathbf{B}\mathbf{x}=r_2^2$ where **A** and **B** are positive of common conjugate axes, illustrated in Figure A3. This l algebraically in Theorem 35.

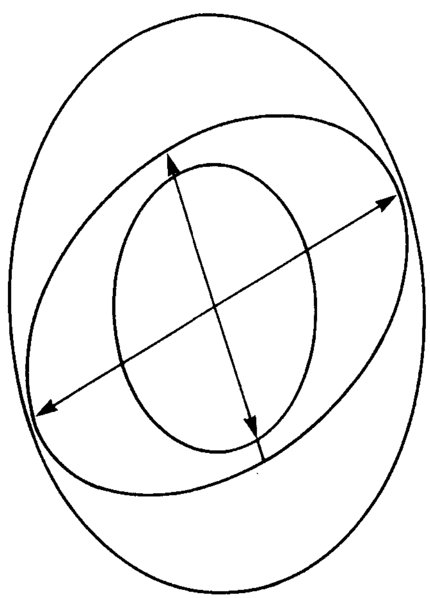

Fig. A.3 Common conjugate axes.

Given a pair of real $(m\times m)$ symmetric matrices **A** *and* **B**, *ı $(m\times m)$ nonsingular matrix* **P** *such that* **P′BP** *is a diagonal* **P′AP** *is a diagonal matrix with ones and zeroes on the*

Proof: Using Theorem 33, find a matrix **C** such that **C′AC** is a diagonal matrix $\boldsymbol{\Lambda}$ with any zero diagonal elements in the lower-right corner,

$$\boldsymbol{\Lambda}=\begin{bmatrix}\boldsymbol{\Lambda}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}.$$

Let **E** be **C′BC**, partition so that $\mathbf{E}_{22}$ corresponds to the full block of zeroes of **C′AC**. Let $(\mathbf{E}_{22}^{-1}\mathbf{E}_{21})$be a matrix such that $\mathbf{0}=\mathbf{E}_{21}-\mathbf{E}_{22}(\mathbf{E}_{22}^{-1}\mathbf{E}_{21})$, and let

$$\mathbf{F}=\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ -\mathbf{E}_{22}^{-1}\mathbf{E}_{21} & \mathbf{I}\end{bmatrix}$$

and note that $\mathbf{F}'\boldsymbol{\Lambda}\mathbf{F}=\boldsymbol{\Lambda}$ and

$$\mathbf{F}'\mathbf{E}\mathbf{F}=\begin{bmatrix}\mathbf{E}_{11}-\mathbf{E}_{12}\mathbf{E}_{22}^{-1}\mathbf{E}_{21} & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_{22}\end{bmatrix}.$$

Then find a matrix $\mathbf{R}_1$ such that $\mathbf{R}_1'\boldsymbol{\Lambda}_1^{-1/2}(\mathbf{E}_{11}-\mathbf{E}_{12}\mathbf{E}_{22}^{-1}\mathbf{E}_{21})\boldsymbol{\Lambda}_1^{-1/2}\mathbf{R}_1=\mathbf{D}_1$, a diagonal matrix, and $\mathbf{R}_1'\mathbf{R}_1=\mathbf{I}$; also find a matrix $\mathbf{R}_2$ such that $\mathbf{R}_2'\mathbf{E}_{22}\mathbf{R}_2=\mathbf{D}_2$, a diagonal matrix, and $\mathbf{R}_2'\mathbf{R}_2=\mathbf{I}$. Then let

$$\mathbf{P}=\mathbf{C}\mathbf{F}\begin{bmatrix}\boldsymbol{\Lambda}_1^{-1/2}\mathbf{R}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_2\end{bmatrix}.$$

If **A** is invertible the diagonal elements of **D** can be shown to be the roots of the polynomial $|\mathbf{B}-\lambda\mathbf{A}|=0$, the columns of **P** satisfy the eigenvector equation $(\mathbf{B}-d_i\mathbf{A})\mathbf{P}_i=\mathbf{0}$, and the matrix **P** is unique if there are no multiple roots $d_i\neq d_j$, for $i\neq j$.

APPENDIX 2

PROBABILITY DISTRIBUTIONS

The probability distributions used in this book are reported in this appendix. Relevant properties are listed. Johnson and Kotz (1972) and Raiffa and Schlaifer (1961) are useful sources of additional details and proofs.

Definitions

A multivariate random variable $\mathbf{x}$ is said to have the probability density function (p.d.f.) $f(\mathbf{x})$ if the probability that $\mathbf{x}$ is in any region R is

$$P(\mathbf{x} \in R) = \int_R f(\mathbf{t})\,dt$$

where the symbols stand for the integral of f over the region R. If the symbol R is suppressed, the integral by convention extends over the whole domain of definition of f. Thus, for example $\int f(\mathbf{x})\,d\mathbf{x} = 1$ indicates that the integral of a p.d.f. over its domain is equal to one.

The *marginal* p.d.f. of the subvector $\mathbf{x}_I$, where $\mathbf{x}' = (\mathbf{x}_I'\mathbf{x}_J')$ is

$$g(\mathbf{x}_I) = \int_{\mathbf{x}_J} f(\mathbf{x})\,d\mathbf{x}_J$$

where the symbols stand for integration over the domain of $\mathbf{x}_J$. The *conditional* p.d.f. of the subvector $\mathbf{x}_I$ given $\mathbf{x}_J$ is

$$h(\mathbf{x}_I|\mathbf{x}_J) = \frac{f(\mathbf{x})}{g(\mathbf{x}_J)}.$$

Whereas here we have selected the symbols f, g, and h to distinguish three different densities, henceforth we, without confusion, write $f(\mathbf{x})$, $f(\mathbf{x}_I)$, and $f(\mathbf{x}_I|\mathbf{x}_J)$.

The mean or expectation of an element x_i is

$E(x_i) = \int x_i f(\mathbf{x})\,d\mathbf{x}$. The mean vector of a multivariate random variable is $E\mathbf{x} = \{E(x_i)\} = \int \mathbf{x} f(\mathbf{x})\,d\mathbf{x}$. The variance-covariance matrix of the random vector $\mathbf{x}$ is

$$V(\mathbf{x}) = E(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))'.$$

Beta Distribution

The beta p.d.f. is

$$f_\beta(p|r,n) = [B(r,n-r)]^{-1} p^{r-1}(1-p)^{n-r-1}, \quad 0 \leqslant p \leqslant 1$$

where $B(r,n-r) = (r-1)!(n-r-1)!/(n-1)!$, and $r,n > 0$. The first two moments of a beta random variable are

$$E(p) = \frac{r}{n}$$

$$V(p) = \frac{r(n-r)}{n^2(n+1)}.$$

Multivariate Normal Distributions

The nondegenerate r-variate normal p.d.f. is

$$f_N^r(\mathbf{x}|\mathbf{u},\Sigma) = (2\pi)^{-r/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

where $\mathbf{x}$ and $\boldsymbol{\mu}$ are $r \times 1$ vectors and $\boldsymbol{\Sigma}$ is an $r \times r$ symmetric positive definite matrix. The mean and variance are

$$E(\mathbf{x}) = \boldsymbol{\mu}$$

$$V(\mathbf{x}) = \boldsymbol{\Sigma}.$$

Partition $\mathbf{x}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ conformably:

$$\mathbf{x}' = (\mathbf{x}_I'\mathbf{x}_J'), \boldsymbol{\mu} = (\boldsymbol{\mu}_I'\boldsymbol{\mu}_J) \qquad \text{and}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{II} & \boldsymbol{\Sigma}_{IJ} \\ \boldsymbol{\Sigma}_{JI} & \boldsymbol{\Sigma}_{JJ} \end{bmatrix}$$

with $\mathbf{x}_I$ having q elements and $\mathbf{x}_J$ having $r-q$ elements. The marginal and conditional distributions are

$$f(\mathbf{x}_I) = f_N^q(\mathbf{x}_I|\boldsymbol{\mu}u_I, \boldsymbol{\Sigma}_{II})$$

$$f(\mathbf{x}_I|\mathbf{x}_J) = f_N^q(\mathbf{x}_I|\boldsymbol{\mu}_I + \boldsymbol{\Sigma}_{IJ}\boldsymbol{\Sigma}_{JJ}^{-1}(\mathbf{x}_J - \boldsymbol{\mu}_J), \boldsymbol{\Sigma}_{II} - \boldsymbol{\Sigma}_{IJ}\boldsymbol{\Sigma}_{JJ}^{-1}\boldsymbol{\Sigma}_{JI}).$$

Notation. $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for "$\mathbf{x}$ is normally distributed with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$."

$\sim N(\mu, \Sigma)$ *with* Σ $r\times r$ *positive definite and if* $\mathbf{A}$ *is* $m\times r$ *then*

$$\mathbf{z}=(\mathbf{a}+\mathbf{A}\mathbf{x})\sim N(\mathbf{a}+\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}').$$

ɔrmal distribution is an r-dimensional distribution that robability on a subspace of dimension $q<r$. The covari- a degenerate normal is singular. The density function is conditionally on the subspace in which the random must lie.

rmal distribution has a density in the form of a normal oes not integrate to one because Σ^{-1} is singular. After a necessary, an improper normal can be written as the r dimensional proper normal times an improper uniform e other variables:

$$f(\mathbf{x}_I,\mathbf{x}_J)\propto f_N(\mathbf{x}_I).$$

ıe marginal and conditional distributions of $\mathbf{x}_I$ are taken distribution $f_N(\mathbf{x}_I)$.

on

. is

$$,\nu)=\left[\left(\frac{\nu}{2}-1\right)!\right]^{-1}\left[\frac{\nu s^2}{2}\right]^{\nu/2} h^{\frac{\nu}{2}-1}\exp\left[-\frac{1}{2}\nu s^2 h\right]$$

$^2>0$. The moments of the gamma distribution are

$$E(h)=s^{-2}$$

$$V(h)=\left(\frac{\nu s^4}{2}\right)^{-1}.$$

(s^2,ν) stands for "h is gamma distributed with parameters

lent Distribution

ɪrate r-variate Student p.d.f. is the marginal of a normal-

$$=\int_0^\infty f_N^r\left(\mathbf{t}|\mu,(h\mathbf{N})^{-1}\right)f_\gamma\left(h|s^2,\nu\right)dh$$

$$\frac{\left(\frac{1}{2}\nu+\frac{1}{2}r-1\right)!|\mathbf{H}|^{\frac{1}{2}}}{\ldots}\left[\nu+(\mathbf{t}-\mu)'\mathbf{H}(\mathbf{t}-\mu)\right]^{-(\nu+r)/2}$$

where $\mathbf{H}=\mathbf{N}s^{-2}$, $\nu>0$, and $\mathbf{H}$ is symmetric positive definite. The first two moments are

$$E(\mathbf{t})=\mu \qquad \nu>1$$

$$V(\mathbf{t})=\mathbf{H}^{-1}\frac{\nu}{\nu-2} \qquad \nu>2$$

Note that the Student distribution and also the normal distribution are written most conveniently in terms of the precision matrices $\mathbf{H}=\mathbf{V}^{-1}$ instead of the variance matrices. More importantly, prior to posterior analysis of location parameters involves formulas that are additive in the precision matrices. Bayesians such as Raiffa and Schlaifer (1961) for these reasons often express densities in terms of the precision parameters. In consideration of students who are more comfortable working with variances, I have adopted the schizophrenic notation above.

Partitioning as in the case of the normal distribution and letting $\mathbf{H}^{-1}=\mathbf{V}$, the marginal Student distribution is

$$f(\mathbf{t}_I)=f_S^q\left(\mathbf{t}_I|\mu_I,(\mathbf{H}^{-1})_{II},\nu\right)$$

where

$$(\mathbf{H}^{-1})_{II}=\left(\mathbf{H}_{II}-\mathbf{H}_{IJ}\mathbf{H}_{JJ}^{-1}\mathbf{H}_{JI}\right)^{-1}=\mathbf{V}_{II}$$

and the conditional p.d.f. is

$$f(\mathbf{t}_I|\mathbf{t}_J)=f_S^q\left(\mathbf{t}_I|\mu_I^*,(\mathbf{H}_{II})^{-1}s^2,\nu+r-q\right)$$

where

$$\mu_I^*=\mu_I+\mathbf{V}_{IJ}\mathbf{V}_{JJ}^{-1}(\mathbf{t}_J-\mu_J)=\mu_I-\mathbf{H}_{II}^{-1}\mathbf{H}_{IJ}(\mathbf{t}_J-\mu_J)$$

$$(\mathbf{H}_{II})^{-1}=\mathbf{V}_{II}-\mathbf{V}_{IJ}\mathbf{V}_{JJ}^{-1}\mathbf{V}_{JI}$$

$$s^2=\frac{\nu+(\mathbf{t}_J-\mu_J)'\mathbf{V}_{JJ}^{-1}(\mathbf{t}_J-\mu_J)}{\nu+r-q}$$

Notation. $\mathbf{t}\sim S(\mu,\mathbf{V},\nu)$ stands for "$\mathbf{t}$ has a Student distribution with parameters μ, $\mathbf{V}$, and ν."

Wishart Distribution

A $k\times k$ symmetric positive definite matrix Ω is said to have a Wishart distribution if its density function is

$$f_W^k(\Omega|\mathbf{S},\nu)=c|\Omega|^{(\nu-k-1)/2}\exp\left[-\tfrac{1}{2}\mathrm{tr}\,\Omega\mathbf{S}\right]$$

where

a $(k \times k)$ symmetric positive-definite matrix. Properties of a
ution are discussed in Zellner (1971, pp. 389–394). The one
in this book is the following relationship between the
udent distributions.

If conditional on $\boldsymbol{\Omega}$, *the random vector* $\mathbf{t}$ *is normally distrib-*
an $\boldsymbol{\mu}$ *and covariance matrix* $(T\boldsymbol{\Omega})^{-1}$ *and if the matrix* $\boldsymbol{\Omega}$ *has*
stribution with parameters $\mathbf{S}$ *and* ν, *then marginally* $\mathbf{t}$ *has a*
Student distribution with parameters $\boldsymbol{\mu}$, $\mathbf{S}/T$ *and* $\nu+1-k$.

int distribution of $\mathbf{t}$ and $\boldsymbol{\Omega}$ can be written as

$$f(\mathbf{t},\boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{1/2}\exp\left[-\frac{T}{2}(\mathbf{t}-\boldsymbol{\mu})'\boldsymbol{\Omega}(\mathbf{t}-\boldsymbol{\mu})\right]$$

$$|\boldsymbol{\Omega}|^{(\nu-k-1)/2}\exp\left[-\frac{1}{2}\operatorname{tr}\boldsymbol{\Omega}\mathbf{S}\right]$$

$$=|\boldsymbol{\Omega}|^{(\nu-k+1-1)/2}\exp\left[-\frac{1}{2}\operatorname{tr}\boldsymbol{\Omega}\mathbf{W}\right]$$

$\cdot\, T(\mathbf{t}-\boldsymbol{\mu})(\mathbf{t}-\boldsymbol{\mu})'$. The last line in this expression has the form
distribution, and the matrix $\boldsymbol{\Omega}$ can be integrated from the
rely by inserting the appropriate normalizing constant:

$$f(\mathbf{t}) \propto |\mathbf{W}|^{-(\nu+1)/2}$$

$$=\left(|\mathbf{S}|\left(1+T(\mathbf{t}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\mathbf{t}-\boldsymbol{\mu})\right)\right)^{-(\nu+1)/2}$$

$$\propto f_s^k(\mathbf{t}|\,\boldsymbol{\mu},\mathbf{S}/T,\nu+1-k).$$

APPENDIX 3

PROOF OF THEOREMS 5.5 AND 5.8[1]

Proof of Theorem 5.5

The variables are first scaled so that the prior precision is $\mathbf{D}^* = d\mathbf{I}$. We let C_n be the set of all combinations of the first k integers taken n at a time. Define $H(I,J)$ as the minor of $\mathbf{H}$ formed by deleting the rows $i \in I$ and columns $j \in J$ with $I,\ J \in C_n$ for some $n \leqslant k$. Furthermore, let $H(Im,Jn)$ be the minor formed by deleting the rows I and m, and the columns J and n.

Expanding the characteristic polynomial as in Gantmacher (1959, p. 70) yields

$$|\mathbf{H}+d\mathbf{I}| = \sum_{j=0}^{k} p_j d^j \quad \text{with} \quad p_j = \sum_{I\in C_j} H(I,I) = \sum_{I\in C_j} |\mathbf{H}_I|,$$

where $|\mathbf{H}_I| = H(I,I)$ and where we let $|\mathbf{H}_I| = 1$ for $I = C_k$.

We can derive a similar expansion for the adjoint, which takes the form

$$|\mathbf{H}+d\mathbf{I}|(\mathbf{H}+d\mathbf{I})^{-1} = \sum_{j=0}^{k-1} \mathbf{B}_j d^j.$$

We derive a formula for $\mathbf{B}_j$ by collecting all terms that involve d^j. In our notation, the cofactor of the (m,n)th element of $\mathbf{H}+d\mathbf{I}$ is

$$(-1)^{m+n}(H+dI)(m,n).$$

In the expansion of this determinant, d^j occurs in any term that contains exactly j diagonal elements:

$$\prod_{i\in I}(h_{ii}+d),$$

[1]This material is taken from Leamer and Chamberlain (1976).

3

-1, and I contains neither m nor n. This product of j s multiplied by $(H+dI)(Im,In)$, which can be written $n,In)$ plus terms that contain d. The latter are thereby ;ideration, since they would create higher powers of d. $'m,In)=0$ for m or $n \in I$, the (m,n)th element of $\mathbf{B}_j$ is

$$(-1)^{m+n}\Sigma_{I\in C_j} H(Im,In).$$

"deleted" inverse $\mathbf{H}_I^{-1}$ to be the $k \times k$ matrix whose

$$(Im,In)/|\mathbf{H}_I|, \qquad \text{we have} \quad \mathbf{B}_j = \sum_{I \in C_j} |\mathbf{H}_I|\mathbf{H}_I^{-1}.$$

$=\mathbf{H}_I^{-1}\mathbf{Hb}$ is the restricted least-squares point with $\beta_i = 0$ can write the conditional posterior mean as

$$(\mathbf{H}+d\mathbf{I})^{-1}\mathbf{Hb} = \Sigma_{j=0}^{k} d^j \sum_{I \in C_j} |\mathbf{H}_I|\mathbf{b}_I / |\mathbf{H}+d\mathbf{I}| \qquad (1)$$

$$|\mathbf{H}+d\mathbf{I}| = \Sigma_{j=0}^{k} d^j \sum_{I \in C_j} |\mathbf{H}_I|. \qquad (2)$$

ınvolves a straightforward transformation of $\mathbf{H}$ in these for the arbitrary diagonal matrix $\mathbf{D}$ in the place of $d\mathbf{I}$.

5.8

rem involves a slight rewriting of equations (1) and (2).

APPENDIX **4**

ASSORTED PROBLEMS

Problems. ***Chapter 2, Sections 2.1–2.2***

1. Determine a probability for each of the following events, and explain what is meant by probability in each case.

(a) A one in the next roll of a die.
(b) R is the first letter of the first name of the 168th entry in the Los Angeles phone book.
(c) R is the first letter of the last name of the eighth president of the United States.
(d) R is the first letter of the last name of the president of the United States in the year 2000.

2. A coin is to be flipped. If it lands heads up, the following statement is made: "Nixon's weight is negative." Otherwise, the statement is made: "Nixon's weight is positive." Which of the following statements are true?

(a) The probability that a true statement will be made is one half.
(b) The probability that a true statement will be made is either one or zero, depending on which statement is made.
(c) If we get a head and say "Nixon's weight is positive," the probability that this is true is one-half.

3. What is the probability of an event A if the odds against A are (a) 2 to 1 (b) 3 to 1 (c) 3 to 2

4. In a three-horse race the odds against the favorite are even (one to one); against the second horse, two to one; and against the "sleeper," three to one. Find stakes that make you a sure winner.

ɔwn that coherence implies the conditional probability $P(B) = P(A \cap B)$ when A is a proper subset of B. Prove ɛnts A and B.

ed. If it lands heads up a ball is drawn from an urn red and three black balls. Otherwise, a ball is drawn ntaining three red and one black balls. Given that a red m this experiment, what is the probability that it came ırn?

s even odds that the Yankees will win the World Series. nces that the Yankees are two-to-one favorites to have ıns. Given that they have more home runs, they are orites to win the series. Construct a set of bets that make ıre loser.

·r 2, *Sections 2.3–2.5*

llowing table where r_1 and n_1 are beta prior parameters ɔle consists of one success in five trials.

95% interval	$E(\tilde{p})$	r_2	n_2	95% interval	$E(\tilde{p})$

proximate prior distributions on the binomial parameter om variable X takes on the value one for the event: a ɛcted United States citizen

than 48 chromosomes.
teeth.
than 82,496 hairs on his body.

ɩhese events we observe 1 success ($X=1$) in a sample of ɔare the probability of a one on the first trial with the ɔf a one on the sixth trial. Give explanations for your

3. Using a beta prior on p, give an example when the posterior variance exceeds the prior variance.

4. (a) If your opinions about p are described by a beta distribution with parameters r and n, what is the probability of getting a success on the next trial?
(b) Starting with a beta prior with parameters r_1 and n_1 and observing r successes in n trials, what is the probability of another success?
(c) Suppose you have observed n successes in n trials, with the "noninformative" prior $n_1 = r_1 = 0$. What is the probability of another success, or is it defined?
(d) Formulate a beta prior for p=proportion of coins that land heads up. What is the probability of another head if no heads are observed, $n = 1, 10, 100, 1000$?

5. (a) Suppose that X and Y are both Bernoulli random variables (r.v.) with the same parameter p. (i) Given that p is a random variable in the sense that it has a prior distribution, can X and Y be independent? (ii) What do we mean when we say that the number of successes in n *independent* Bernoulli trials has a binomial distribution? (Call this *conditional independence*.)
(b) If X is a Bernoulli r.v. with parameter p_x and Y is a Bernoulli r.v. with parameter p_y, what restriction on the joint prior on (p_x, p_y) implies X and Y (marginally) independent?

6. A sample of size 10 with mean 50 was taken from a normal population with unknown mean $\tilde{\mu}$ and variance 100. A normal prior for $\tilde{\mu}$ was formed with mean m_1 and variance v_1. Fill in the following table.

		Prior				Posterior	
m_1	v_1	95% interval	$P(\tilde{\mu}>48)$	m_2	v_2	95% interval	$P(\tilde{\mu}>48)$
50	1000						
50	1						
40	1000						
40	1						

7. Consider a population of size N consisting of pN ones and $(1-p)N$ twos. How does one make inferences based on a sample of size n about the mean $\mu = p + 2(1-p) = 2 - p$ when the sample (with replacement) results in s two's and f one's?

proximate prior distributions for the following means:
IQ of Harvard students.
weight in pounds of Harvard students.
length in tenths of an inch of wooden foot-long rulers.
weight in marspounds of Martians.

ample these populations, and in a random sample of size e a mean of $m=121$ and a standard deviation of $s=5$. posterior probabilities that you assign to the events $\leqslant 122$.

her it is more appropriate when reporting results of a eriment to provide the data, the sufficient statistics, or p.d.f.

t we begin with a normal prior for μ with mean m_1 and Two samples from a normal distribution are taken with nd m_b and with the sample sizes n_a and n_b. Given σ^2,

osterior" distribution for μ given m_a alone.
sterior distribution for μ given m_a and m_b by using the f part (a) as a prior for sample b.
sterior for μ given a sample mean $m=(m_a n_a+m_b n_b)/(n_a$ ith sample size n_a+n_b. Compare your answers to (b) and

ng are given: a normal population with uncertain mean μ variance α^2, a random sample of size n, and a normal prior for μ with mean m_1 and variance v_1. Let $m_2=E(\mu|m)$ the sample mean.

down the posterior mean m_2 given the sample mean, m.
ute the expected value of m_2 given μ but not m; given m

ute the expected value of m_2 given neither m nor μ.
ute the posterior variance v_2 of μ given the sample mean

ute the expected variance given μ but not m; given m but

ute the distribution of m_2 given neither m nor μ.

12. (a) One thousand measurements with a ruler of the width of a finger yielded a mean of 1.9682 and a standard deviation $s=.06$ centimeters. What is a 95% interval for the width of the finger, assuming a diffuse prior where relevant?
(b) A single measurement with a micrometer yielded a value of 1.96105. Reconcile (a) and (b).

13. A random sample of size n is to be taken from a normal population with *known* mean and the *unknown* variance σ^2. Beginning with a prior for $h=\sigma^{-2}$ in the gamma family, compute the posterior distribution for h.

Problems. Chapter 3

1. Find the mean and variance of the constrained estimator (3.15).

2. Show that the F statistic for testing $\mathbf{R}\boldsymbol{\beta}=\mathbf{r}$ is

$$F=\frac{(\mathbf{Rb}-\mathbf{r})'\left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1}(\mathbf{Rb}-\mathbf{r})}{ps^2}$$

where p is the rank of $\mathbf{R}$.

3. Find the posterior distribution of the residual vector $\mathbf{u}$, $f(\mathbf{u}|\mathbf{Y},\mathbf{X})$, given a conjugate prior for $\boldsymbol{\beta},\sigma^2$.

4. Given the regression equation $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$, with $\mathbf{u}$ normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{V}$, with $\mathbf{V}\neq\mathbf{I}$, show that the generalized least-squares estimator

$$\mathbf{b}(\mathbf{V})=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

is the maximum likelihood estimator, given $\mathbf{V}$. Show also that it is the best linear unbiased estimator of $\boldsymbol{\beta}$.

5. Let a regression equation be $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{\gamma}+\mathbf{u}$, where $\mathbf{X}$ and $\mathbf{Z}$ are observable matrices, $\mathbf{u}$ is unobservable and distributed normally with mean $\mathbf{0}_T$, and variance is $\sigma^2\mathbf{I}_T$. Let $\mathbf{M}_z=\mathbf{I}_T-\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

(a) Show that $\mathbf{M}_z\mathbf{M}_z=\mathbf{M}_z$
(b) Show that the least-squares estimate of $\boldsymbol{\beta}$ is $(\mathbf{X}'\mathbf{M}_z\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_z\mathbf{Y}$
(c) If $\boldsymbol{\gamma}$ has a prior distribution with mean $\mathbf{0}$ and variance $v\mathbf{I}$, what is the variance of $\boldsymbol{\varepsilon}=\mathbf{Z}\boldsymbol{\gamma}+\mathbf{u}$? Find the generalized least-squares

of $\boldsymbol{\beta}$ given the regression equation $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$. Show estimator converges to the least-squares estimator of $\boldsymbol{\beta}$ in part (b) as v goes to infinity.

$Y=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$, $\mathbf{u}\sim N(0,\sigma^2\mathbf{I}_T)$, and

$$\mathbf{X}'\mathbf{X}=\frac{1}{3}\begin{pmatrix}1 & -1\\ -1 & 4\end{pmatrix} \qquad \mathbf{Y}'\mathbf{Y}=8$$

$$\mathbf{X}'\mathbf{Y}=\begin{pmatrix}1\\1\end{pmatrix}$$

$$\sigma^2=1$$

e the least-squares estimate of $\boldsymbol{\beta}$.
e a 95% confidence interval for β_1.
hypothesis $\beta_1=0$ versus the alternative $\beta_1\neq 0$.
e a 95% confidence interval for $\beta_1+\beta_2$.
hypothesis $\beta_1+\beta_2=0$ versus the alternative $\beta_1+\beta_2\neq 0$.
e the least-squares estimate of β_1, and β_2 given $\beta_1+\beta_2=$

t assume that σ^2 is unknown and that $\mathbf{Y}'\mathbf{Y}=8$ and $T=12$.

ta in problem 6 and a normal prior for $\boldsymbol{\beta}$ with mean $\mathbf{0}$ $\mathbf{I}$, determine the posterior distribution of $\boldsymbol{\beta}$. What are r intervals for β_1, β_2, and $\beta_1+\beta_2$?

ta in problem 6 and the prior in problem 8 with variance to a scale factor λ, find the equation for the curve

2 is unknown as in problem 7, compute a posterior using a normal-gamma prior for $(\boldsymbol{\beta},\sigma^2)$ with parameters

$$\mathbf{b}^*=(0,0)$$

$$\mathbf{N}^*=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

$$\nu^*=10$$

$$s^{*2}=1.$$

5% posterior intervals for β_1, β_2, and $\beta_1+\beta_2$?

Problems. Chapter 4

1. It is known that x is normally distributed with unknown mean μ and variance $\sigma^2=1$.

(a) Construct a .05 level test of $H_0:\mu=0$ versus $H_1:\mu=1$.
(b) Write down the function $P(H_0|x)/P(H_1|x)$.
(c) If $P(H_0)/P(H_1)=1$, what is the significance level of the following test: accept H_0 if $P(H_0|x)/P(H_1|x)>1$; otherwise reject H_0.

2. Suppose that $\bar{x}\sim N(\mu,\sigma^2/n)$ where $\sigma^2=1$ and n is the sample size.

(a) Construct a .05 level test of $H_0:\mu=0$ versus $H_1:\mu\neq 0$, given $n=1$, 10, and 100. For each value of n, draw approximately the error characteristic curve, $P(\text{error}|\mu)$.
(b) Construct a .05 level test of $H_0':|\mu|\leqslant .1$ versus $H_1':|\mu|>.1$, for $n=1$, 10, 100. For each value of n, draw approximately the error characteristic curve, $P(\text{error}|\mu)$.
(c) Suppose that you meant to test H_0' against H_1' but instead constructed a .05 level test for H_0 versus H_1. What is the error characteristic curve for $n=1, 10, 100$. What is the actual significance level of the test?

3. Make use of the data reported in problem 6, Chapter 3 to do the following problem.

(a) Let the prior allocate positive probability to the hypothesis $\beta_1=0$, and let it otherwise be (appropriately) diffuse. Compute for $T=10$, $T=100$, and $T=1000$, $P(\beta_1=0|\mathbf{Y})/P(\beta_1\neq 0|\mathbf{Y})$.
(b) Compute the classical t value for testing $\beta_1=0$ for $T=10$, $T=100$, and $T=1000$. Is the hypothesis $\beta_1=0$ accepted or rejected?
(c) Compute $P(\beta_1=0|\mathbf{Y})/P(\beta_2=0|\mathbf{Y})$, using as a prior the uniform distribution over the lines $\beta_1=0$ and $\beta_2=0$.
(d) Suppose in testing $\beta_1=0$ versus $\beta_2=0$ the first hypothesis is accepted if it yields an equation with the higher R^2. Let $g(\beta_1,\beta_2)$ be the probability of making error of accepting the wrong model. Write down an expression for $g(\beta_1,\beta_2)$ and explain why it depends on β_1 and β_2. What is the maximum value of $g(\beta_1,\beta_2)$? What is the type I error probability if $\beta_1=0$ is the null hypothesis and $\beta_2=0$ is the alternative.

constrained least-squares points given constraints of
$=\mathbf{Mr}$, where $\mathbf{R}$ is a given $p\times k$ matrix with rank p,
a given $p\times 1$ vector, and where $\mathbf{M}$ is an arbitrary $p\times p$

ıtion of this chapter, let the least-squares estimator of
ıd let a constrained least-squares estimator be $\psi'\hat{\beta}(\mathbf{R})$
a row vector and where $\hat{\beta}(\mathbf{R})=\mathbf{b}-(\mathbf{X}'\mathbf{X})^{-1}$
$^{1}\mathbf{R}')^{-1}\mathbf{R}')^{-1}\mathbf{Rb}$. Show that $\psi'\hat{\beta}(\mathbf{R})$ has a smaller mean-
than $\psi'\mathbf{b}$ if and only if the "true" squared t is less than
$^{2}/\sigma^{2}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$.

mal linear regression model with known variance and a
normal prior distribution for the coefficient vector,
gencies between a posterior ellipsoid and either a prior
l ellipsoid lie on the information contract curve.

ely on the following statements that have been made
ar regression model $\mathbf{Y}=\mathbf{x}\beta+\mathbf{z}\gamma+\mathbf{u}$.

nexperimental sciences the explanatory variables $\mathbf{x}$ and
t subject to control, and they may be highly collinear.
es the variables are collinear because they are drawn
"collinear" population; alternatively, $\mathbf{x}$ and $\mathbf{z}$ may be
lently distributed, but by chance are correlated in this
r sample. It is important to distinguish these two possi-
A useful test of the "collinearity problem" is a test of the
sis that $\mathbf{x}$ and $\mathbf{z}$ are distributed independently.
rity between the explanatory variables is less of a prob-
he variables are negatively correlated than if they are
y correlated.

lowing data, find the (four) principal component regres-
s and comment on the desirability of estimation subject
omponent restrictions.

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix}3 & -1 & 0\\ -1 & 3 & 0\\ 0 & 0 & 1\end{bmatrix}\qquad \mathbf{X}'\mathbf{Y}=\begin{bmatrix}8\\8\\8\end{bmatrix}$$

6. Making use of the data given in problem 6, Chapter 3,

(a) Draw the ellipsoid of constrained estimates.
(b) Assuming the number of observations T is equal to 10, compute the pretest estimator (5.10) of β_1, with a test level of .05 and the hypothesis $\beta_2=0$.
(c) Can a Stein–James estimate of $\boldsymbol{\beta}$ be computed? Is the least-squares estimate admissible?
(d) Making use of a spherical prior, graph the information contract curve, and find the rotation invariant average regressions.
(e) Find the principal component regression estimates of $\boldsymbol{\beta}$.
(f) Compute the conditional 95% confidence interval for β_1 given β_2.
(g) If the prior distribution is spherical, is there a collinearity problem?
(h) For each coefficient, compute the measures of collinearity c_1, and c_2.

7. T observations of a two-variable regression model yielded the following moments

$$\mathbf{Y}'\mathbf{Y}=2 \qquad T=12$$
$$\mathbf{Y}'\mathbf{X}=\begin{pmatrix}1\\1\end{pmatrix}\qquad \mathbf{X}'\mathbf{X}=\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}$$

Assume that the matrix $\mathbf{X}$ is fixed.

(a) Find a 2×1 vector $\boldsymbol{\lambda}$ such that $\mathbf{X}\boldsymbol{\lambda}=0$.
(b) Find a vector $\boldsymbol{\beta}$ that is observationally equivalent to the vector (3,2).
(c) Is the function $\beta_1+\beta_2$ identified?
(d) Draw a graph indicating the likelihood contours.
(e) Find an identifying restriction.
(f) If the prior is spherical, find a linear combination of β_1 and β_2 about which the experiment is personally uninformative.
(g) Find a linear combination about which the experiment is publicly informative.