

Contents

A	Linear algebra basics	1
A.1	Basic operations	1
A.2	Fundamental theorem of linear algebra	4
A.2.1	Part one	4
A.2.2	Part two	5
A.3	Nature of the solution	7
A.3.1	Exactly-identified	7
A.3.2	Under-identified	8
A.3.3	Over-identified	10
A.4	Matrix decomposition and inverse operations	12
A.4.1	LU factorization	12
A.4.2	Cholesky decomposition	16
A.4.3	Eigenvalues and eigenvectors	17
A.4.4	Singular value decomposition	23
A.4.5	Spectral decomposition	28
A.4.6	quadratic forms, eigenvalues, and positive definiteness	29
A.4.7	similar matrices, Jordan form, and generalized eigen- vectors	29
A.5	Gram-Schmidt construction of an orthogonal matrix	32
A.5.1	QR decomposition	34
A.5.2	Gram-Schmidt QR algorithm	34
A.5.3	Accounting example	35
A.5.4	The Householder QR algorithm	36
A.5.5	Accounting example	36

A.6	Computing eigenvalues	39
A.6.1	Schur's lemma	39
A.6.2	Power algorithm	40
A.6.3	QR algorithm	40
A.6.4	Schur decomposition	42
A.7	Some determinant identities	46
A.7.1	Determinant of a square matrix	46
A.7.2	Identities	47
A.8	Matrix exponentials and logarithms	49
B	Iterated expectations	51
B.1	Decomposition of variance	53
B.2	Jensen's inequality	54
C	Multivariate normal theory	55
C.1	Conditional distribution	57
C.2	Special case of precision	61
C.3	Truncated normal distribution	63
D	Projections and conditional expectations	71
D.1	Gauss-Markov theorem	71
D.2	Generalized least squares (GLS)	74
D.3	Recursive least squares	76
E	Two stage least squares IV (2SLS-IV)	79
E.1	General case	79
E.2	Special case	81
F	Seemingly unrelated regression (SUR)	83
F.1	Classical	84
F.2	Bayesian	84
F.3	Bayesian treatment effect application	85
G	Maximum likelihood estimation of discrete choice models	87
H	Optimization	89
H.1	Linear programming	89
H.1.1	basic solutions or extreme points	90
H.1.2	fundamental theorem of linear programming	91
H.1.3	duality theorems	91
H.1.4	example	92
H.1.5	complementary slackness	92
H.2	Nonlinear programming	93
H.2.1	unconstrained	93
H.2.2	convexity and global minima	93
H.2.3	example	94

H.2.4	constrained — the Lagrangian	94
H.2.5	Karash-Kuhn-Tucker conditions	95
H.2.6	example	96
H.3	Theorem of the separating hyperplane	97
I	Quantum information	99
I.1	Quantum information axioms	99
I.1.1	The superposition axiom	99
I.1.2	The transformation axiom	100
I.1.3	The measurement axiom	100
I.1.4	The combination axiom	101
I.2	Summary of quantum "rules"	103
I.3	Observables and expected payoffs	104
I.4	Density operators and quantum entropy	105
J	Common distributions	109

Appendix B

Iterated expectations

Along with Bayes' theorem (the glue holding consistent probability assessment together), iterated expectations is extensively employed for connecting conditional expectation (regression) results with causal effects of interest.

Theorem 11 *Law of iterated expectations*

$$E[Y] = E_X[E[Y | X]]$$

Proof.

$$\begin{aligned} E_X[E[Y | X]] &= \int_{\underline{x}}^{\bar{x}} E[Y | X] f(x) dx \\ &= \int_{\underline{x}}^{\bar{x}} \left[\int_{\underline{y}}^{\bar{y}} y f(y | x) dy \right] f(x) dx \end{aligned}$$

By Fubini's theorem, we can change the order of integration

$$\int_{\underline{x}}^{\bar{x}} \left[\int_{\underline{y}}^{\bar{y}} y f(y | x) dy \right] f(x) dx = \int_{\underline{y}}^{\bar{y}} y \left[\int_{\underline{x}}^{\bar{x}} f(y | x) f(x) dx \right] dy$$

The product rule of Bayes' theorem, $f(y | x) f(x) = f(y, x)$, implies iterated expectations can be rewritten as

$$E_X[E[Y | X]] = \int_{\underline{y}}^{\bar{y}} y \left[\int_{\underline{x}}^{\bar{x}} f(y, x) dx \right] dy$$

Finally, the summation rule integrates out X , $\int_{\underline{x}}^{\bar{x}} f(y, x) dx = f(y)$, and produces the result.

$$E_X [E [Y | X]] = \int_{\underline{y}}^{\bar{y}} y f(y) dy = E [Y]$$

■

B.1 Decomposition of variance

Corollary 1 *Decomposition of variance.*

$$\text{Var} [Y] = E_X [\text{Var} [Y | X]] + \text{Var}_X [E [Y | X]]$$

Proof.

$$\begin{aligned} \text{Var} [Y] &= E [Y^2] - E [Y]^2 \\ &= E_X [E [Y^2 | X]] - E_X [E [Y | X]]^2 \\ \text{Var} [Y] &= E_X [E [Y^2 | X]] - E [Y]^2 \\ &= E_X [\text{Var} [Y | X] + E [Y | X]^2] - E_X [E [Y | X]]^2 \\ &= E_X [\text{Var} [Y | X]] + E_X [E [Y | X]^2] - E_X [E [Y | X]]^2 \\ \text{Var} [Y] &= E_X [\text{Var} [Y | X]] + \text{Var}_X [E [Y | X]] \end{aligned}$$

The second line draws from iterated expectations while the fourth line is the decomposition of the second moment. ■

In analysis of variance language, the first term is the residual variation (or variation unexplained) and the second term is the regression variation (or variation explained).

B.2 Jensen's inequality

For any concave function $g(x)$, $E[g(x)] \leq g(E[x])$. Likewise, for any convex function $h(x)$, $E[h(x)] \geq h(E[x])$. Hence, utility functions exhibiting concavity characterize risk aversion or positive risk premia while utility functions exhibiting convexity characterize risk-seeking preferences or negative risk premia.

Further, Jensen's inequality tells us the geometric mean, $G(x)$, is less than or equal to the arithmetic mean, $A(x)$, with equality only when all outcomes are the same.

$$G(x) \equiv \prod_{i=1}^n x_i^{p_i} \leq A(x) \equiv \sum_{i=1}^n p_i x_i$$

To see this result, let $g(\cdot)$ be the logarithm (a monotone increasing, concave function) for x nonnegative (if any $x_i = 0$ then the inequality is trivially satisfied as the geometric mean is zero if any $x_i = 0$)

$$E[g(x)] = \sum_{i=1}^n p_i \ln x_i \leq g(E[x]) = \ln \sum_{i=1}^n p_i x_i$$

Let $p_i = \frac{w_i}{w}$ where $w = \sum_{i=1}^n w_i$, then

$$\sum_{i=1}^n \frac{w_i}{w} \ln x_i \leq \ln \sum_{i=1}^n \frac{w_i}{w} x_i$$

To recover geometric and arithmetic means exponentiate (a monotone increasing function) both sides

$$\sqrt[w]{\prod_{i=1}^n x_i^{w_i}} = G(x) \leq \sum_{i=1}^n \frac{w_i}{w} x_i = A(x)$$