ENCYCLOPEDIA OF STATISTICAL SCIENCES

VOLUME 5

Samuel Kotz
Norman L. Johnson
Editors-in-Chief

ENCYCLOPEDIA OF STATISTICAL SCIENCES

VOLUME 5

LINDEBERG CONDITION
to MULTITRAIT-MULTIMETHOD MATRICES

A WILEY-INTERSCIENCE PUBLICATION
John Wiley & Sons
NEW YORK · CHICHESTER · BRISBANE · TORONTO · SINGAPORE

Library of Congress Cataloging in Publication Data:

Main entry under title:

Encyclopedia of statistical sciences.

"A Wiley-Interscience publication."

Includes bibliographies.

Contents: v. 1. A to Circular probable error—
v. 3. Faà di Bruno's formula to Hypothesis testing—
[etc.]—v. 5. Lindeberg condition to multitrait-
multimethod matrices.

1. Mathematical statistics—Dictionaries.
2. Statistics—Dictionaries. I. Kotz, Samuel.
II. Johnson, Norman Lloyd. III. Read, Campbell B.
QA276.14.E5 1982 519.5'03'21 81-10353
ISBN 0-471-05552-2 (v. 5)

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

R. J. Adler, *Technion, Haifa, Israel.* Local Time; Markov Random Fields

M. C. Agrawal, *University of Delhi, Delhi, India.* Mean Reciprocal Values

I. A. Ahmad, *University of Maryland, Catonsville, Maryland.* Matusita's Distance

P. Albrecht, *Universität Mannheim, Federal Republic of Germany.* Mixed Poisson Processes

F. B. Alt, *University of Maryland, College Park, Maryland.* Matching Problems

I. Amara, *University of North Carolina, Chapel Hill, North Carolina.* Log-Rank Scores, Statistics and Tests

C. E. Antle, *Pennsylvania State University, University Park, Pennsylvania.* Lognormal Distribution

A. A. Assad, *University of Maryland, College Park, Maryland.* Management Science, Statistics in

J. K. Baksalary, *Academy of Agriculture in Poznan, Poznan, Poland.* Milliken's Estimability Criterion

Y. Bard, *IBM, Cambridge, Massachusetts.* Maximum Entropy Principle, Classical Approach

D. J. Bartholomew, *The London School of Economics, London, England.* Manpower Planning

C. Benedetti, *Università degli Studi di Roma, Rome, Italy. Metron*

J. O. Berger, *Purdue University, West Lafayette, Indiana.* Minimax Estimation

U. N. Bhat, *Southern Methodist University, Dallas, Texas.* Markov Processes

B. B. Bhattacharyya, *North Carolina State University, Raleigh, North Carolina.* Multinomial Probit: Multinomial Logit

H. M. Blalock, Jr., *University of Washington, Seattle, Washington.* Multiple Indicator Approach

D. Boos, *North Carolina State University, Raleigh, North Carolina. L*-Statistics

K. O. Bowman, *Union Carbide Corporation, Oak Ridge, Tennessee.* Method of Moments

B. M. Brown, *The University of Tasmania, Hobart, Tasmania, Australia.* Median Estimates and Sign Tests

M. W. Browne, *University of South Africa, Pretoria, South Africa.* Multitrait-Multimethod Matrices

R. J. Buehler, *University of Minnesota, Minneapolis, Minnesota.* Minimum Variance Unbiased Estimation

O. Bunke, *Humboldt-Universität zu Berlin, Berlin, German Democratic Republic. Mathematische Operationsforschung und Statistik, Series Statistics*

J. A. Cadzow, *Arizona State University, Tempe, Arizona.* Maximum Entropy Spectral Analysis

S. Cambanis, *University of North Carolina, Chapel Hill, North Carolina.* Measure Theory in Probability and Statistics

V. Chew, *University of Florida, Gainesville, Florida.* Multiple Range and Associated Test Procedures

P. L. Cornelius, *University of Kentucky, Lexington, Kentucky.* Linear Models with

Crossed-Error Structure; Linear Models with Nested-Error Structure

J. A. Cornell, *University of Florida, Gainesville, Florida.* Mixture Experiments

R. G. Cornell, *University of Michigan, Ann Arbor, Michigan.* Most Probable Number

P. R. Cox, *Mayfield, Sussex, England.* Marriage; Migration

E. M. Cramer, *University of North Carolina, Chapel Hill, North Carolina.* Multicollinearity

C. Dagum, *University of Ottawa, Ottawa, Canada.* Lorenz Curve

E. B. Dagum, *Statistics Canada, Ottawa, Canada.* Moving Averages

J. N. Darroch, *The Flinders University, Bedford Park, South Australia.* Marginal Symmetry

M. H. DeGroot, *Carnegie-Mellon University, Pittsburgh, Pennsylvania.* Minimax Decision Rules; Multiple Decision Procedures

J. L. Denny, *University of Arizona, Tucson, Arizona.* Markovian Dependence

D. J. De Waal, *University of the Orange Free State, Bloemfontein, South Africa.* Matrix-Valued Distributions

C. A. Drossos, *University of Patras, Patras, Greece.* Metamathematics of Probability

R. M. Dudley, *Massachusetts Institute of Technology, Cambridge, Massachusetts.* Manifolds

S. Edwards, *University of North Carolina, Chapel Hill, North Carolina.* Logistic Regression

B. S. Everitt, *University of London, London, England.* Mixture Distributions

P. C. Fishburn, *Bell Laboratories, Murray Hill, New Jersey.* Mean-Variance Analyses

M. A. Fligner, *Ohio State University, Columbus, Ohio.* Location Tests

M. R. Frankel, *National Opinion Research Center, New York, New York.* Master Samples

G. H. Freeman, *National Vegetable Research Station, Wellesbourne, Warwick, England.* Magic Square Designs

J. Gani, *University of Kentucky, Lexington, Kentucky.* Literature, Statistics in; Meteorology, Statistics in

S. I. Gass, *University of Maryland, College Park, Maryland.* Linear Programming

J. L. Gastwirth, *George Washington University, Washington, DC.* Measures of Uniformity

S. Geman, *Brown University, Providence, Rhode Island.* Method of Sieves

J. E. Gentle, *IMSL, Houston, Texas.* Monte Carlo Methods

B. L. Golden, *University of Maryland, College Park, Maryland.* Management Science, Statistics in

J. C. Gower, *Rothamsted Experimental Station, Harpenden, Herts, England.* Measures of Similarity, Dissimilarity and Distance

P. E. Green, *University of Pennsylvania, Philadelphia, Pennsylvania.* Marketing, Statistics in

S. T. Gross, *Baruch College, New York, New York.* Median Estimation, Inverse

F. E. Grubbs, *U.S. Army, Aberdeen Proving Ground, Maryland.* Mann–Grubbs Method

W. C. Guenther, *University of Wyoming, Laramie, Wyoming.* Modified (or Curtailed) Sampling Plans

R. C. Gupta, *University of Maine, Orono, Maine.* Modified Power Series Distribution

P. Hackl, *Wirtschaftsuniversität Wien, Vienna, Austria.* Moving Sums (MOSUM)

P. Hall, *Australian National University, Canberra, Australia.* Martingales

E. J. Hannan, *Australian National University, Canberra, Australia.* Linear Systems, Statistical Theory

H. L. Harter, *Wright State University, Dayton, Ohio.* Method of Averages; Method of Group Averages; Method of Least Absolute Values; Method of Least p-th Powers; Minimax Method; Most Approximative Method

D. M. Hawkins, *National Research Institute for Mathematical Sciences, Pretoria, South Africa.* Masking and Swamping

T. P. Hettmansperger, *Pennsylvania State*

University, University Park, Pennsylvania. Mann's Test for Trend; Mathisen's Median Test

C. C. Heyde, *University of Melbourne, Parkville, Victoria, Australia.* Multidimensional Central Limit Theorem

R. R. Hocking, *Texas A & M University, College Station, Texas.* Linear Regression

C. Hsiao, *University of Toronto, Toronto, Ontario, Canada.* Minimum Chi Square

R. Hultquist, *Pennsylvania State University, University Park, Pennsylvania.* Multiple Linear Regression

J. S. Hunter, *Princeton University, Princeton, New Jersey.* Measurement Errors

C. M. Jarque, *Secretaria de Programacion y Presupuesto, Mexico City, Mexico.* Multistratified Sampling

N. T. Jazairi, *York University, Downsview, Ontario, Canada.* Log-Change Index Numbers; Marshall–Edgeworth–Bowley Index

E. H. Jebe, *Ann Arbor, Michigan.* Multiphase Sampling

D. E. Johnson, *Kansas State University, Manhattan, Kansas.* Messy Data

B. L. Joiner, *University of Wisconsin, Madison, Wisconsin.* MINITAB

K. G. Jöreskog, *University of Uppsala, Uppsala, Sweden.* LISREL

V. M. Joshi, *University of Western Ontario, London, Ontario, Canada.* Linear Sufficiency

K. Kafadar, *Hewlett-Packard, Palo Alto, California.* Monte Carlo Swindle

G. Kallianpur, *University of North Carolina, Chapel Hill, North Carolina.* Mahalanobis, Prasantha Chandra

N. Keyfitz, *Harvard University, Cambridge, Massachusetts.* Malthus, Thomas Robert; Malthusian Parameter

G. G. Koch, *University of North Carolina, Chapel Hill, North Carolina.* Logistic Regression; Log-Rank Scores, Statistics and Tests; Multiple-Record Systems

K. Kocherlakota, *University of Manitoba, Winnipeg, Canada.* Multinormal Distributions

S. Kocherlakota, *University of Manitoba, Winnipeg, Canada.* Multinormal Distributions

H. C. Kraemer, *Stanford University Medical Center, Stanford, California.* Measures of Agreement

R. A. Kronmal, *University of Washington, Seattle, Washington.* Marsaglia's Table Method; Mixture Method

S. Kullback, *George Washington University, Washington, D.C.* Minimum Discrimination Information Estimation

H. E. Kyburg, *University of Rochester, Rochester, New York.* Logic in Statistical Reasoning

N. M. Laird, *Harvard University, Cambridge, Massachusetts.* Missing Information Principle

D. Lambert, *Carnegie-Mellon University, Pittsburgh, Pennsylvania.* Minimax Tests; Most Stringent Test

L. R. LaMotte, *University of Houston, Houston, Texas.* Linear Estimators, Bayes

C. Leach, *University of Newcastle upon Tyne, Newcastle upon Tyne, England. Mathematical and Statistical Psychology, British Journal of*

E. A. Lew, *Punta Gorda, Florida.* Medico-Actuarial Investigations

K. McConway, *The Open University, Milton Keynes, England.* Marginalization

H. MacGillivray, *University of Queensland, St. Lucia, Queensland, Australia.* Mean, Median, Mode and Skewness

H. J. Malik, *University of Guelph, Ontario, Canada.* Logistic Distribution

B. F. J. Manly, *University of Otago, Dunedin, New Zealand.* Manly–Parr Estimators

N. R. Mann, *University of California, Los Angeles, California.* Mann–Fertig Statistic; Mann–Grubbs Method

K. V. Mardia, *The University of Leeds, Leeds, England.* Mardia's Test of Multivariate Normality

D. G. Marx, *University of the Orange Free State, Bloemfontein, South Africa.* Matric-*t* Distribution

P. W. Mielke, *Colorado State University, Fort Collins, Colorado.* Mood's Dispersion Test; Multi-Response Permutation Procedures

R. G. Miller, *Stanford University, Stanford, California.* Multiple Comparisons

G. A. Milliken, *Kansas State University, Manhattan, Kansas.* Messy Data

L. E. Moses, *Stanford University, Stanford, California.* Matched Pairs; Matched Pairs *t* Test

G. S. Mudholkar, *University of Rochester, Rochester, New York.* Multiple Correlation Coefficient

B. Natvig, *University of Oslo, Oslo, Norway.* Multistate Coherent Systems

D. G. Nel, *University of the Orange Free State, Bloemfontein, South Africa.* Matric-*t* Distribution

K. O'Brien, *East Carolina University, Greenville, North Carolina.* Mantel–Haenszel Statistic

J. K. Ord, *Pennsylvania State University, University Park, Pennsylvania.* Linear-Circular Correlation

T. Papaioannou, *University of Ioannina, Ioannina, Greece.* Measures of Information

W. C. Parr, *University of Florida, Gainesville, Florida.* Minimum Distance Estimation

E. Parzen, *Texas A & M University, College Station, Texas.* Multiple Time Series

G. P. Patil, *Pennsylvania State University, University Park, Pennsylvania.* Linear Exponential Family; Logarithmic Series Distribution

A. V. Peterson, *University of Washington, Seattle, Washington.* Marsaglia's Table Method; Mixture Method

A. N. Pettitt, *University of Technology, Loughborough, Leicester, England.* Mann–Whitney–Wilcoxon Statistic

E. C. Pielou, *University of Lethbridge, Alberta, Canada.* Line Intercept Sampling; Line Intersect Sampling; Line Transect Sampling

K. C. S. Pillai, *Purdue University, West Lafayette, Indiana.* Mahalanobis *D* Square

J. H. Pollard, *Macquarie University, North Ryde, New South Wales, Australia.* Mathematical Theory of Population

D. Pregibon, *Bell Laboratories, Murray Hill, New Jersey.* Link Tests

R. L. Prentice, *The Fred Hutchinson Cancer Research Center, Seattle, Washington.* Linear Rank Tests

C. P. Quesenberry, *North Carolina State University, Raleigh, North Carolina.* Model Construction, Selection of Distributions

D. Raghavarao, *Temple University, Philadelphia, Pennsylvania.* Main Effects

C. R. Rao, *University of Pittsburgh, Pittsburgh, Pennsylvania.* Matrix Derivatives, Applications in Statistics; MINQE

M. V. Ratnaparkhi, *Wright State University, Dayton, Ohio.* Liouville–Dirichlet Distribution; Multinomial Distributions

C. B. Read, *Southern Methodist University, Dallas, Texas.* Linear Hazard Rate Distribution; Markov Inequality; Mathematical Functions, Approximations to; Maxwell, James Clerk; Mean Deviation; Mean Squared Error; Median Unbiased Estimators; Midranges; Mills' Ratio; Morbidity

J. Rissanen, *IBM, San Jose, California.* Minimum-Description-Length Principle

T. Robertson, *University of Iowa, Iowa City, Iowa.* Monotone Relationships

V. K. Rohatgi, *Bowling Green State University, Bowling Green, Ohio.* Moment Problem

D. Ruppert, *University of North Carolina, Chapel Hill, North Carolina. M*-Estimators

B. F. Ryan, *Pennsylvania State University, University Park, Pennsylvania.* MINITAB

T. A. Ryan, *Pennsylvania State University, University Park, Pennsylvania.* MINITAB

D. Sankoff, *University of Montréal, Montréal, Canada.* Linguistics, Statistics in

F. W. Scholz, *Boeing Computer Services Company, Tukwila, Washington.* Maximum Likelihood Estimation

H. T. Schreuder, *U.S. Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.* Mickey's Unbiased Ratio and Regression Estimators

S. J. Schwager, *Cornell University, Ithaca, New York.* Mean Slippage Problems

H. Seal*, *École Polytechnique Federale de*

*Deceased

Lausanne, Lausanne, Switzerland. Multiple Decrement Tables

A. R. Sen, *Calgary, Alberta, Canada.* Location Parameter, Estimation of

P. K. Sen, *University of North Carolina, Chapel Hill, North Carolina.* Locally Optimal Statistical Tests; Log-Rank Scores, Statistics and Tests; Multidimensional Contingency Tables

E. Seneta, *University of Sydney, Sydney, New South Wales, Australia.* Markov, Andrei Andreevich; Montmort, Pierre Rémond de

G. Shafer, *University of Kansas, Lawrence, Kansas.* Miller's Paradox; Moral Certainty

M. Shaked, *University of Arizona, Tucson, Arizona.* Majorization and Schur Convexity

A. Shapiro, *University of South Africa, Pretoria, South Africa.* Minimum Rank Factor Analysis; Minimum Trace Factor Analysis

L. R. Shenton, *University of Georgia, Athens, Georgia.* Method of Moments

D. M. Shoemaker, *U.S. Department of Education, Washington, D.C.* Multiple Matrix Sampling

G. Simons, *University of North Carolina, Chapel Hill, North Carolina.* Measure Theory in Probability and Statistics

B. Singer, *The Rockefeller University, New York, New York.* Longitudinal Data Analysis

N. D. Singpurwalla, *The George Washington University, Washington, D.C.* Military Standards for Fixed-Length Life Tests; Military Standards for Sequential Life Testing

H. Solomon, *Stanford University, Stanford, California.* Military Statistics

G. W. Somes, *East Carolina University, Greenville, North Carolina.* Mantel–Haenszel Statistic; McNemar Statistics; Miettinen's Statistic

A. P. Soms, *University of Wisconsin, Madison, Wisconsin.* Lindstrom–Madden Method

F. W. Steutel, *Technisch Hogeschool Eindhoven, Eindhoven, The Netherlands.* Log-Concave and Log-Convex Distributions

G. W. Stewart, *University of Maryland, College Park, Maryland.* Linear Algebra, Computational; Matrix, Ill-Conditioned

S. Stigler, *University of Chicago, Chicago, Illinois.* Mansfield, Merriman

D. J. Strauss, *University of California, Riverside, California.* Luce's Choice Axioms and Generalizations

R. Syski, *University of Maryland, College Park, Maryland.* Multi-Server Queues

H. Thomas, *Pennsylvania State University, University Park, Pennsylvania.* Measurement Structures and Statistics

D. M. Titterington, *University of Glasgow, Glasgow, Scotland.* Medical Diagnosis, Statistics in

J. W. Tolle, *University of North Carolina, Chapel Hill, North Carolina.* Mathematical Programming

R. Trader, *University of Maryland, College Park, Maryland.* Moments, Partial

R. C. Tripathi, *University of Texas, San Antonio, Texas.* Modified Power Series Distribution

W. Uhlmann, *Institut für Angewandte Mathematik und Statistik, Würzburg, Federal Republic of Germany. Metrika*

T. S. Wallsten, *University of North Carolina, Chapel Hill, North Carolina.* Measurement Theory

E. J. Wegman, *Office of Naval Research, Arlington, Virginia.* Military Statistics

L. Weiss, *Cornell University, Ithaca, New York.* Maximum Probability Estimation

D. J. White, *Manchester University, Manchester, England.* Markov Decision Processes

H. White, *University of California, La Jolla, California.* Misspecification, Tests for

Y. Wind, *University of Pennsylvania, Philadelphia, Pennsylvania.* Marketing, Statistics in

D. B. Wolfson, *McGill University, Montreal, Canada.* Lindeberg–Feller Theorem; Lindeberg–Lévy Theorem

F. W. Young, *University of North Carolina, Chapel Hill, North Carolina.* Multidimensional Scaling

M. Zeleny, *Fordham University at Lincoln Center, New York, New York.* Multiple Criteria Decision Making

L *continued*

LINDEBERG CONDITION *See* LINDEBERG–FELLER THEOREM

LINDEBERG–FELLER THEOREM

HISTORY

The Lindeberg–Feller central limit theorem gives necessary and sufficient conditions for the convergence in distribution of a sequence of suitably standardized sums of independent random variables with finite variance, to a standard normal random variable. An important forerunner to the Lindeberg–Feller theorem was proved by Liapunov* in 1900, who established sufficient conditions for convergence to normality assuming the existence of moments of order $2+\delta$ for some $\delta > 0$. Lindeberg [5] strengthened Liapunov's theorem* by assuming the existence of the second moments while Feller [1] established the necessity of Lindeberg's condition for convergence to normality, provided that the uniformly asymptotically negligible condition is satisfied. The theorem given below is for a single sequence of random variables. A more general version for double arrays of random variables may be found in most textbooks on probability theory (see, e.g., Gnedenko and Kolmogorov [4]).

STATEMENT OF THE LINDEBERG–FELLER THEOREM

Let $X_1, X_2, \ldots,$ be a sequence of independent random variables and let the distribution function of X_k be $F_k(x)$. Suppose that $\operatorname{var}(X_k) = \sigma_k^2 < \infty$ for all k. Let $S_n = \sum_{k=1}^{n}(X_k - E[X_k])/s_n$, where $s_n^2 = \sum_{k=1}^{n}\operatorname{var}(X_k)$. Then in order that as $n \to \infty$,

$$\text{(a)} \quad \Pr[S_n \leqslant x] \to (1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-t^2/2}\,dt \qquad \text{for all} \quad x > 0,$$

$$\text{(b)} \quad \lim_{n\to\infty} \max_{1 \leqslant k \leqslant n} \Pr[|(X_k - E[X_k])/s_n| > \epsilon] = 0 \qquad \text{for all} \quad \epsilon > 0,$$

it is necessary and sufficient that for each $\eta > 0$, we have

$$\text{(L)} \quad (1/s_n^2)\sum_{k=1}^{n}\int_{|x| > \epsilon s_n} x^2\,dF_k(x + E[X_k]) \to 0 \qquad \text{for each} \quad \epsilon > 0.$$

IDEA OF PROOF: SUFFICIENCY

Form a sequence of truncated bounded random variables and using the Lindeberg condition together with a limit result due to Liapunov, establish sufficiency for this truncated sequence. Finally, show that the standardized sums arising from the truncated random variables behave the same way in the limit as do the standardized sums arising from the original random variables.

An alternative approach using convolution operators may be found in Feller [2].

NECESSITY

The proof is technical and involves manipulations with characteristic functions.*

REMARKS

1. Condition (b), sometimes called the uniformly asymptotically negligibility (UAN) condition, ensures that the relative contribution of each of the summands of S_n tends to zero as $n \to \infty$.
2. Condition (L) is referred to the Lindeberg condition. Again, it is a condition on the relative smallness of each summand, and may be shown to imply that $\max_{1 \leqslant j \leqslant n} \sigma_j / s_n \to 0$ as $n \to \infty$.
3. Condition (L) may be easily shown to imply the Liapunov condition.*
4. As in the case of the Lindeberg–Lévy theorem,* there exist various extensions of the Lindeberg–Feller theorem to sums of random vectors [3] in $\mathbb{R}^k$ and to sums of dependent random variables. (*See* LIMIT THEOREM, CENTRAL.)

References

[1] Feller, W. (1935). *Math. Zeit.*, **40**, 521–559.

[2] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. 2. Wiley, New York.

[3] Gnedenko, B. V. (1944). *Uspekhi Mat. Nauk*, **10**, 230 (in Russian).

[4] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, Mass.

[5] Lindeberg, J. W. (1922). *Math. Zeit.*, **15**, 221.

(LIMIT THEOREM, CENTRAL
LINDEBERG–LÉVY THEOREM)

D. WOLFSON

LINDEBERG–LÉVY THEOREM

HISTORY

The Lindeberg–Lévy central limit theorem is widely known more simply as "the" central limit theorem, although it is, strictly speaking, a single representative of central limit theorems in general. It refers to the convergence in distribution of a suitably standardized sum of independent, identically distributed random variables with finite variance, to a normal random variable with zero mean and unit variance. The form given below is due, independently, to Lévy [1] and Lindeberg [2].

STATEMENT OF THE LINDEBERG–LÉVY CENTRAL LIMIT THEOREM

Let $X_1, X_2, X_3, \ldots$ be a sequence of independent and identically distributed random variables with $E[X_i] = \mu$ and $0 < \operatorname{var}(X_i) = \sigma^2 < \infty$. Then

$$\Pr\left[\sum_{i=1}^{n} (X_i - \mu)/(\sqrt{n}\,\sigma) \leqslant x\right] \to (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

for all $x \in (-\infty, \infty)$, as $n \to \infty$.

IDEA OF PROOF

The characteristic function* of

$$\sum_{i=1}^{n} (X_i - \mu)/(\sqrt{n}\,\sigma)$$

is expanded using Taylor's theorem and the

assumed existence of a second moment. This expanded characteristic function (ch.f) is then shown to converge pointwise to $e^{-t^2/2}$, the ch.f. of the standard normal distribution. An application of the continuity theorem then leads to the stated result.

EXTENSIONS AND GENERALIZATIONS

The Lindeberg–Lévy theorem may be extended to include random vectors:

Random Vector Lindeberg–Lévy Theorem. Let $\mathbf{X}_n = (X_{n1}, X_{n2}, \ldots, X_{nk})$ be independent, identically distributed random vectors with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$, where $\mu_j = E[X_{nj}]$, $j = 1, 2, \ldots, k$. Suppose further that $E[X_{nj}^2] < \infty$ and that the covariance matrix * of $\mathbf{X}_n$ is $\boldsymbol{\Sigma}$. Then $\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu})/\sqrt{n}$ converges in distribution to a multivariate normal* random vector having mean zero and covariance matrix $\boldsymbol{\Sigma}$.

A second extension, to random sums of random variables, is exemplified by the following theorem:

Random Sums Lindeberg–Lévy Theorem. Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables with $E[X_i] = \mu$ and $\operatorname{var}(X_i) = \sigma^2$. Let $\{\nu_n, n \geqslant 1\}$ be a sequence of random variables taking on only strictly positive integer values such that $\nu_n/n \to \alpha$ in probability, where α is a constant satisfying $0 < \alpha < \infty$. Then

$$\Pr\left[\sum_{i=1}^{\nu_n}(X_i - \mu)/\left(\sigma\sqrt{\nu_n}\right) \leqslant x\right] \to \left(1/\sqrt{2\pi}\right)\int_{-\infty}^{x} e^{-t^2/2}\,dt$$

for all $x \in (-\infty, \infty)$ as $n \to \infty$.

Related to the central limit theorem but requiring more delicate analyses are the laws of the iterated logarithm,* which describe the transient behavior of the standardized sums as n increases. These results complement those describing the limiting behavior given by the central-limit-type theorems.

Other generalizations of the Lindeberg–Lévy theorem may be found in LIMIT THEOREMS, CENTRAL, which includes examples of central-limit-type theorems for dependent random variables; also discussed are the concepts of infinite divisibility* and stability as well as the famous Berry–Esseen theorem on rate of convergence to normality (*see* ASYMPTOTIC NORMALITY).

APPLICATIONS

Either in its original form or in one of its extended forms, the Lindeberg–Lévy theorem lies at the heart of most of the asymptotic theory of statistics. The construction of confidence intervals, indeed much of the theory and practice of statistical inference* and hypothesis testing,* is facilitated by the availability of the central limit theorem. The frequent occurrence of sums of random variables in statistics is the reason for the wide applicability of this important theorem.

References

[1] Lévy, P. (1925). *Calcul des Probabilités*. Gauthier-Villars, Paris.

[2] Lindeberg, J. W. (1922). *Math. Zeit.*, **15**, 221.

Further Reading

In addition to the references just given, the following texts contain many of the topics referred to in this article.

Billingsley, P. (1979). *Probability and Measure*. Wiley, New York.

Feller, W. (1966), *An Introduction to Probability Theory and Its Applications*, Vol. 2. Wiley, New York.

Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, Mass.

All of these books are rigorous in spirit and require some mathematical sophistication.

(LIMIT THEOREM, CENTRAL)

D. WOLFSON

LINDLEY'S EQUATION

Let W_n be the waiting time of the nth customer in a queue, S_n be the corresponding service time, and X_{n+1} be the length of time between the nth and $(n+1)$th arrivals. Then, for a general $G/G/1$ queueing system (with one customer),

$$W_{n+1} = \max(0, W_n + S_n - X_n).$$

This basic result, due to Lindley [2], implies that the distribution function

$$F_n(x) = \Pr(W_n \leqslant x)$$

of the Ws converges as $n \to \infty$ to some limit function F. For more details see Lindley [2] and Cohen [1].

References

[1] Cohen, J. W. (1969). *The Single Server Queue*. North-Holland, Amsterdam, London.

[2] Lindley, D. V. (1952). *Proc. Cambridge Phil. Soc.*, **48**, 177–184.

(QUEUEING THEORY)

LINDSTROM–MADDEN METHOD

The Lindstrom–Madden method is a method for constructing approximate lower confidence limits* on the reliability of a series system, given Bernoulli subsystem data. This is a fundamental problem in reliability theory* and is discussed in ref. 6, together with the Lindstrom–Madden method. A series system of independent components functions only if all the components function. More precisely, let Y_i, $i = 1, 2, \ldots, k$, be independent binomial random variables with parameters (n_i, p_i), $n_1 \leqslant n_2 \leqslant \cdots \leqslant n_k$, where p_i is the probability that the ith subsystem will function, and let the observed values be $y_1, y_2, \ldots, y_k$, with $x_i = n_i - y_i$, $i = 1, 2, \ldots, k$. The reliability of the system is $\prod_{i=1}^{k} p_i$. A general method of constructing a lower $1 - \alpha$ level confidence limit for $\prod_{i=1}^{k} p_i$ was given in ref. 1. However, this is very difficult to implement in practice. Lipow and Riley [5] constructed the exact lower confidence limit for $\prod_{i=1}^{k} p_i$ for specific values of x_i, n_i, $k = 2, 3$, and Lloyd and Lipow [6] noted that the tabulated values were close to an approximation, the Lindstrom–Madden method, described below. Consider building systems by randomly selecting without replacement a single test result (success or failure) from each of the k subsystem data. Then there are n_1 systems and the expected number of failures is $z_1 = n_1 q_0$, $q_0 = 1 - \prod_{i=1}^{k}((n_i - x_i)/n_i)$. Let

$$I_p(r,s) = \frac{1}{\beta(r,s)} \int_0^p t^{r-1}(1-t)^{s-1}\,dt;$$

i.e., $I_p(r,s)$ is the incomplete beta function. Then if y is an integer, $y < n$, we have

$$\sum_{i=0}^{y} \binom{n}{i} p^{n-i} q^i = I_p(n-y, y+1).$$

A complete discussion of the above can be found in ref. 4. For $0 \leqslant y < n$, real, define $u(n, y, \alpha)$ by $\alpha = I_{u(n,y,\alpha)}(n-y, y+1)$. Thus, for integer values of y, $u(n, y, \alpha)$ is a $100(1-\alpha)\%$ lower confidence limit for p. Then the Lindstrom–Madden method consists of using $u(n_1, z_1, \alpha)$ as an approximation to the exact lower confidence limit b and reduces to the usual method for putting a lower confidence limit on the success probability if z_1 is an integer.

Denote by $[x]$ the integral part of x, x real. Sudakov [7] showed that $u(n_1, z_1, \alpha) \leqslant b \leqslant u(n_1, [z_1], \alpha)$, and hence if z_1 is an integer, then the Lindstrom–Madden method is exact. The special case when only x_1 is nonzero (in this case $z_1 = x_1$) was proved by Winterbottom [8]. Sudakov's results were simplified and generalized by Harris and Soms [2]. Also, Harris and Soms [3] improved the lower bound $u(n_1, z_1, \alpha)$ by the use of a short FORTRAN program, whose listing they provide. We now give two examples.

Example 1. Let $\alpha = 0.05$, $(n_1, n_2, n_3) = (5, 10, 20)$, and $(x_1, x_2, x_3) = (1, 0, 0)$. Then $z_1 = 1$, $u(5, 1, 0.05) = 0.343$, so $b = 0.343$.

Example 2. Let $\alpha = 0.05$, $(n_1, n_2, n_3, n_4) = (10, 15, 20, 25)$, and $(x_1, x_2, x_3, x_4) = (1, 3, 2, 4)$. Then $z_1 = 4.557$, $u(10, 4.557, 0.05) = 0.257$, $u(10, 4, 0.05) = 0.304$, so $0.257 \leqslant b \leqslant 0.304$. Using the program in Harris and Soms [3], this can be improved to $0.291 \leqslant b \leqslant 0.304$.

Addendum—Added in Proof An error has been discovered in the proof of one of the lemmas employed in the proof of Sudakov's inequality. This is documented in "The Theory of Optimal Confidence Limits for Systems Reliability with Counterexamples for Results on Optimal Confidence Limits for Series Systems", B. Harris and A. P. Soms, Tech. Rep. 643, Dept. of Statistics, University of Wisconsin-Madison, Madison, Wis. Numerical evidence indicates that for confidence levels of practical interest the inequality is still valid.

References

[1] Buehler, R. J. (1957). *J. Amer. Statist. Ass.*, **52**, 482–493. (The fundamental paper on one-sided confidence limits on parametric functions.)

[2] Harris, B. and Soms, A. P. (1980). Bounds for Optimal Confidence Limits for Series Systems. *Tech. Rep. No. 2093*, Mathematics Research Center, University of Wisconsin–Madison, Madison, Wis. (Generalizes and simplifies the results of ref. 7.)

[3] Harris, B. and Soms, A. P. (1981). Improved Sudakov-Type Bounds for Optimal Confidence Limits on the Reliability of Series Systems. *Tech. Rep. No. 637*, Dept. of Statistics, University of Wisconsin–Madison, Madison, Wis. (Further sharpening of the lower bound for the optimal lower confidence limit on system reliability.)

[4] Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Discrete Distributions*. Wiley, New York, Chap. 3. (Provides an excellent reference for the binomial distribution.)

[5] Lipow, M. and Riley, J. (1959). *Tables of Upper Confidence Bounds on Failure Probability of 1, 2, and 3 Component Serial Systems*. Vols. 1 and 2. Space Technology Laboratories, Los Angeles, Calif. (Often used as a standard of comparison for various approximate methods; contains only equal sample sizes.)

[6] Lloyd, D. K. and Lipow, M. (1962). *Reliability: Management, Methods, and Mathematics*. Prentice-Hall, Englewood Cliffs, N. J., Chap. 9. (An early and often cited text on all aspects of reliability.)

[7] Sudakov, R. S. (1974). *Eng. Cybern.*, **12**, 55–63. (Contains the first theoretical justification of the Lindstrom–Madden method; difficult to read.)

[8] Winterbottom, A. (1974). *J. Amer. Statist. Ass.*, **69**, 782–788. (Proves a special case of ref. 7 and gives a survey of different methods.)

Bibliography

Harris, B. (1977). In *Theory and Applications of Reliability: With Emphasis on Bayesian and Nonparametric Methods*, C. P. Tsokos and I. N. Shimi, eds., Vol. 2. Academic Press, New York, pp. 275–297. (Comprehensive survey paper of the many approaches used on this problem.)

Harris, B. and Soms, A. P. (1981). Recent Advances in Statistical Methods for System Reliability Using Bernoulli Sampling of Components. *Tech. Rep. No. 643*, Dept. of Statistics, University of Wisconsin–Madison, Madison, Wis. (An update of the previous paper.)

Johns, M. V., Jr. (1975). Reliability Assessment for Highly Reliable Systems. *Tech. Rep. No. 1*, Dept. of Statistics, Stanford University, Stanford, Calif. (Uses the Poisson approximation and has tables for $k = 2$.)

(RELIABILITY
SYSTEM RELIABILITY ANALYSIS)

ANDREW P. SOMS

LINEAR ALGEBRA, COMPUTATIONAL

Computational linear algebra may be divided into three parts. The first part deals with discrete computations, such as manipulating the graph of a matrix or performing exact calculations on matrices with integer or rational elements. The second part is concerned with constrained optimization problems that are defined in terms of matrices—for example, linear and quadratic programming problems. The third part is often called numerical linear algebra. It deals with the solution in floating-point arithmetic of such problems as solving linear systems of equations, linear least-squares problems, the computation of eigenvalues, etc.

Because of limitations of space this article will be restricted to numerical linear algebra.

It should be appreciated, however, that even this restriction is insufficient. Specifically, a matrix may be too large to fit into the high-speed memory of the computer in question, and it is a matter of considerable algorithmic ingenuity to solve such problems. Most of the recent research in this area has been devoted to *sparse matrices*, most of whose elements are zero. No attempt will be made to survey sparse matrix computations here; instead, this article will treat matrix problems that can be held in the high-speed memory of the computer or at least can be solved with one or two passes of the data through memory.

It is hardly necessary to argue the importance of numerical linear algebra in statistics. This is obvious for areas like multivariate analysis,* where matrix notation has become standard. It is less well appreciated that a significant amount of matrix computations is essential for maximum likelihood estimation,* robust estimation,* and other essentially nonlinear statistical problems. This is because nonlinear problems are frequently solved by solving a sequence of linearized approximations.

Modern numerical linear algebra has five principal features:

1. The systematic use of matrix decompositions
2. The use of updating methods to recompute decompositions of slightly altered matrices
3. The use of backward rounding-error analysis to assess the stability of algorithms
4. The use of perturbation theory to assess the accuracy of computed solutions
5. The implementation of matrix algorithms in high-quality mathematical software

In selecting or developing methods for solving a problem, it is necessary to have an understanding of each of these five features. Accordingly, they will be discussed separately in the next five sections. The treatment is primarily didactic; technical comments and references will be found in the section "Bibliographical Notes."

DECOMPOSITIONS

The method of Gaussian elimination for solving the linear equation

$$Ax = b \tag{1}$$

of order p will serve to introduce the use of matrix decompositions in numerical linear algebra. The result of performing Gaussian elimination on the system is to produce a lower triangular matrix L and an upper triangular matrix U such that

$$A = LU. \tag{2}$$

Given this *LU decomposition*, one may solve the system (1) by first solving the upper triangular system

$$Uy = b$$

(this is equivalent to performing the elimination operations on the vector b). The upper triangular system

$$Lx = b$$

is then solved for x (this is usually called the back-substitution phase of the algorithm); *see* GAUSS – JORDAN ELIMINATION.

From the foregoing it is evident that the LU decomposition (2) and its computation are distinct from the problem (1) and its solution. Once an LU decomposition has been computed [an $O(p^3)$ process], it may be used again and again to solve systems of the form (1) at a cost of $O(p^2)$ work per solution. Moreover, the same LU decomposition can be used to solve the system $A^T x = b$. Finally, if the LU decomposition of A is available, it is unnecessary to form A^{-1} to solve (1), even though the formula $x = A^{-1}b$ suggests that it is. In fact, the calculation of A^{-1} is not only unnecessary, but it is expensive [an extra $O(p^3)$ work] and undesirable on grounds of numerical stability.

The foregoing comments about the LU decomposition hold generally for the decom-

positional approach to numerical linear algebra. A decomposition may be regarded as a computational platform from which a variety of problems may be solved. It may be relatively expensive to compute; but once it is available, it can be used repeatedly at little additional cost. Moreover, a decomposition may make it unnecessary to compute such notationally convenient but numerically tricky objects as inverses or generalized inverses*. Finally, although it is not evident from the LU decomposition, a well-chosen decomposition may aid in the mathematical analysis of a statistical process.

There are a large number of decompositions that are used in numerical linear algebra, and it can be a matter of some delicacy to select the right one. Accordingly, the rest of this section is devoted to a survey of the most frequently used matrix decompositions.

Pivoted LU Decomposition

The LU decomposition (2) need not exist, and even when it does it may be impossible to compute it in a stable manner. The cure for this situation is to interchange rows and columns of A during the computations in order to ensure numerical stability, a process that is called *pivoting*. The result is a decomposition of the form

$$P_r A P_c = LU,$$

where P_r and P_c are the permutation matrices corresponding to the row and column interchanges. There are two common strategies for choosing interchanges. At the kth step of the algorithm, *partial pivoting* interchanges rows to bring the largest element in column k into the (k,k)-position. *Complete pivoting* interchanges both rows and columns to place the largest element in the matrix in the (k,k)-position. Although one can prove stronger theorems about the stability of complete pivoting, in practice the simpler partial-pivoting strategy is just as stable and is consequently the method of choice.

The principal application of the LU decomposition is to the solution of linear equations and, where it is required, the computation of inverses. It is computed by Gaussian elimination or one of its variants, such as the algorithms of Crout or Doolittle. It requires $O(p^3)$ work.

Cholesky Decomposition

If A is a $p \times p$ positive-definite matrix (which here implies symmetry), it can be factored uniquely in the form

$$A = R^T R, \tag{3}$$

where R is upper triangular with positive-diagonal elements. This factorization is called the *Cholesky decomposition*.

There is an important variant in which pivoting is used to produce a decomposition of the form

$$P^T A P = R^T R,$$

where P is a permutation matrix. The interchanges may be chosen so that the elements of R satisfy

$$r_{kk}^2 \geqslant \sum_{i=k}^{j} r_{ij}^2 \qquad (j = k+1, k+2, \ldots, p).$$

This implies that the diagonals of R are decreasing. In particular, if A is near a semidefinite matrix, then a trailing principal submatrix of A will be small.

The Cholesky decomposition is used to solve linear systems involving positive-definite matrices. In statistical applications the matrix is often a correlation matrix. The Cholesky decomposition of the augmented cross-product matrix $(X\ y)^T(X\ y)$ is used in the analysis of the linear model $y = XB + e$. The Cholesky decomposition is also used in the solution of symmetric eigenvalue problems of the type $Ax = \lambda Bx$, a problem that occurs frequently in statistical applications. The algorithm for computing the decomposition is called variously the Cholesky algorithm or the square-root method.[1] It is simply an adaptation of Gaussian elimination and requires $O(p^3)$ work.

QR Decomposition

Given any $n \times p$ $(n > p)$ matrix X there is an $n \times n$ orthogonal matrix Q such that

$$Q^T X = \begin{bmatrix} R \\ O \end{bmatrix},$$

where R is upper triangular with nonnegative diagonal elements. If Q is partitioned in the form

$$Q = (\overset{p}{Q_X} \quad \overset{n-p}{Q_\perp}),$$

then

$$X = Q_X R, \tag{4}$$

an expression that is sometimes called the *QR factorization*. The *QR* decomposition is closely related to the Cholesky factorization, as the relation

$$X^T X = R^T R \tag{5}$$

shows.

The *QR* decomposition is extraordinarily versatile. To cite one example, if X is of rank p, $Q_\perp Q_\perp^T$ is the projection onto the orthogonal complement of the column space of X. Thus $Q_\perp Q_\perp^T y$ is the residual vector $y - X\hat{b}$ of linear regression. As another example, the generalized inverse $X^\dagger$ of X is given by $X^\dagger = R^{-1} X^T$. Thus operations involving $X^\dagger$ can be replaced by a multiplication by X^T followed by the solution of upper triangular systems.

The *QR* decomposition may be computed by three distinct algorithms: the Golub–Householder algorithm, the method of plane rotations, and the Gram–Schmidt* method with reorthogonalization. They each have their own advantages and drawbacks. Each requires $O(np^2)$ work.

Spectral Decomposition

It is well known that a symmetric matrix A of order p has a set of orthonormal eigenvectors $v_1, v_2, \ldots, v_p$ satisfying

$$Av_i = \lambda_i v_i \qquad (i = 1, 2, \ldots, p). \tag{6}$$

If $V = (v_1, v_2, \ldots, v_p)$, then V is orthogonal and it follows from (6) that

$$V^T A V = \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p). \tag{7}$$

The decomposition (7) is called the *spectral decomposition* of A.

The spectral decomposition is widely used both inside and outside of statistics. For example, the spectral decomposition of a sample correlation matrix gives estimates of the principal components (*see* COMPONENTS ANALYSIS).

All methods for computing eigenvalues and eigenvectors are necessarily iterative. For the spectral decomposition the procedure is to perform a direct reduction to a tridiagonal matrix* followed by the iterative *QR* algorithm (not to be confused with the *QR* decomposition). The amount of work varies, but it is always $O(p^3)$. This algorithm supersedes the older Jacobi algorithm, which, unfortunately, is still to be found in some statistical programs.

Singular Value Decomposition

Given any $n \times p$ $(n \geqslant p)$ matrix X, there is an $n \times n$ orthogonal matrix U and a $p \times p$ orthogonal matrix V such that

$$U^T X V = \begin{bmatrix} \Psi \\ 0 \end{bmatrix}.$$

where

$$\Psi = \operatorname{diag}(\psi_1, \psi_2, \ldots, \psi_p)$$

with

$$\psi_1 \geqslant \psi_2 \geqslant \cdots \geqslant \psi_p \geqslant 0.$$

The numbers ω_i are called the *singular values* of X, the columns of U the *left singular vectors*, and the columns of V the *right singular vectors*. The singular value decomposition is related to the spectral decomposition of $X^T X$ in much the same way as the *QR* and Cholesky decompositions are related; compare the relation

$$V^T X^T X V = \Psi^2$$

with (7).

Since there are relatively few computational tasks that require a singular value decomposition and since the decomposition

is expensive to compute, it is less used than the other decompositions. On the other hand, it has many interesting properties that make it of theoretical interest to both numerical analysts and statisticians. For example, the matrix $\overline{X}$ of rank k nearest X in the least-squares sense may be obtained by setting $\overline{\Psi} = \text{diag}(\psi_1, \psi_2, \ldots, \psi_k, 0, \ldots, 0)$ and forming

$$\overline{X} = U \begin{bmatrix} \overline{\Psi} \\ 0 \end{bmatrix} V^T.$$

The customary way of calculating the singular value decomposition is by an initial reduction to bidiagonal form followed by a variant of the QR algorithm. When $n > p$, a great deal of work can be saved by first calculating the QR decomposition of X and then the singular value decomposition of R, from which the singular value decomposition of X can be reconstituted.

Schur Decomposition

It is natural to attempt to generalize the spectral decomposition of a symmetric matrix to nonsymmetric matrices by asking for a nonsingular matrix W such that $W^{-1}AW$ is diagonal, in which case the columns of W are eigenvectors of A. Unfortunately, such a decomposition need not exist, and even when it does, the matrix W may be so near a singular matrix as to be computationally useless. If W is restricted to be unitary, then it can be determined so that

$$W^H A W = T,$$

where T is an upper triangular matrix. The diagonal elements of T, which are eigenvalues of A, can be made to appear in any order, although usually they are thought of as appearing in descending order of magnitude. Whatever the order, the decomposition is called a *Schur decomposition* and the columns of W are called *Schur vectors*. The first k Schur vectors span the invariant subspace corresponding to the first k eigenvalues of A as they appear in T.

When A is real, it is desirable on grounds of computational efficiency to remain in the real field, even when A has complex eigenvalues. It can be shown that there is an orthogonal matrix W such that W^TAW is quasi-triangular; i.e., W^TAW is block upper triangular with at most 2×2 blocks. The 1×1 blocks are eigenvalues of A, while the 2×2 blocks contain complex conjugate pairs of eigenvalues. The decomposition is sometimes called a real Schur decomposition.

The Schur decomposition is used in the computation of eigenvectors and other objects related to the spectrum of A. In fact, many problems that would seem to require eigenvectors for their solution actually require no more than the Schur decomposition. The decomposition is computed by an initial reduction to Hessenberg form followed by the QR algorithm at a cost of $O(p^3)$ work.

Generalized Schur Decomposition

Although a theory of canonical forms exists for the generalized eigenvalue problem $Ax = \lambda Bx$, an attempt to compute one of these forms may lead to numerical difficulties. If one restricts oneself to unitary equivalences of the form $(A, B) \rightarrow (Y^HAW, Y^HBW)$, then one can find unitary matrices Y and W such that Y^HAW and Y^HBW are both upper triangular. This decomposition may be called a *generalized Schur decomposition*. The eigenvalues of the problem are ratios of the corresponding diagonal elements of Y^HAW and Y^HBW, and they may be made to appear in any order. If B is nonsingular, this decomposition is related to the Schur decomposition of $B^{-1}A$, since $W^HB^{-1}AW$ is upper triangular. When A and B are real, there is a variant decomposition in which B is quasi-triangular.

The generalized Schur decomposition is used principally to solve the generalized eigenvalue problem. It is computed by a variant of the QR algorithm.

For square matrices, the total work required to compute any of the decompositions described here is $O(p^3)$; however, the order constants vary widely. The following is