ENCYCLOPEDIA OF STATISTICAL SCIENCES

VOLUME 1 A–Circular Probable Error

SAMUEL KOTZ & NORMAN L. JOHNSON
EDITORS-IN-CHIEF

CAMPBELL B. READ
ASSOCIATE EDITOR

ENCYCLOPEDIA OF STATISTICAL SCIENCES

VOLUME 1

A to CIRCULAR PROBABLE ERROR

1807 1982

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."
Contents: v. 1. A–Circular probable error.
1. Mathematical statistics—Dictionaries.
2. Statistics—Dictionaries. I. Kotz, Samuel.
II. Johnson, Norman Lloyd. III. Read, Campbell B.
QA276.14.E5 519.5'03'21 81-10353
ISBN 0-471-05546-8 (v. 1) AACR2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

H. Ahrens, *Zentralinstitut für Mathematik und Mechanik der Adw, Berlin, East Germany*. Biometrical Journal

F. B. Alt, *University of Maryland, College Park, Maryland*. Bonferroni Inequalities and Intervals

O. D. Anderson, *University of Nottingham, Nottingham, England*. Box-Jenkins Model

D. F. Andrews, *University of Toronto, Toronto, Canada*. Andrews Function Plots

S. Asmussen, *University of Copenhagen, Copenhagen, Denmark*. Branching Processes

C. R. Baker, *University of North Carolina, Chapel Hill, North Carolina*. Absolute Continuity

V. P. Barabba, *Bureau of the Census, Washington, D.C.* Bureau of the Census

G. A. Barnard, *University of Waterloo, Ontario, Canada*. Causation

D. Basu, *Florida State University, Tallahassee, Florida*. Basu's Theorems

R. J. Beckman, *Los Alamos Scientific Laboratory, Los Alamos, New Mexico*. Chemistry, Statistical Methods in

B. Benjamin, *City University, London, England*. Census

V. P. Bhapkar, *University of Kentucky, Lexington, Kentucky*. Chi-Square Tests

G. K. Bhattacharyya, *University of Wisconsin, Madison, Wisconsin*. Bennett's Bivariate Sign Test

Y. M. M. Bishop, *Harvard University, Cambridge, Massachusetts*. Biometric Society

K. O. Bowman, *Union Carbide Corporation, Oak Ridge, Tennessee*. Approximations to Distributions

L. D. Broemeling, *Oklahoma State University, Stillwater, Oklahoma*. Box and Cox Transformation

F. B. Cady, *Cornell University, Ithaca, New York*. Biometrics

J. M. Cameron, *Wheaton, Maryland*. Calibration

G. J. Chaitin, *IBM, Yorktown Heights, New York*. Algorithmic Information Theory

J. Chambers, *Bell Laboratories, Murray Hill, New Jersey*. Algorithms, Statistical

S. K. Chatterjee, *Calcutta Statistical Association, Calcutta, India*. Calcutta Statistical Association Bulletin

H. Chernoff, *Massachusetts Institute of Technology, Cambridge, Massachusetts*. Chernoff Faces

E. Çinlar, *Princeton University, Princeton, New Jersey*. Chung Processes

W. G. Cochran, *Harvard University, Cambridge, Massachusetts*. Biometric Society

D. R. Cox, *Imperial College, London, England*. Binary Data; Biometrika

H. T. David, *Iowa State University, Ames, Iowa*. Chance (I)

N. R. Draper, *University of Wisconsin, Madison, Wisconsin*. Backward Elimination Selection Procedure.

B. S. Duran, *Texas Tech University, Lubbock, Texas*. Barton-David Test; Capon Test

M. R. Frankel, *City University of New York, New York*. Balanced Repeated Replications

G. H. Freeman, *National Vegetable Research Station, Warwick, England*. Agriculture, Statistics in

W. A. Fuller, *Iowa State University, Ames, Iowa*. Area Sampling

K. R. Gabriel, *University of Rochester, Rochester, New York*. Biplots

J. Galambos, *Temple University, Philadelphia, Pennsylvania*. Characterizations of Distributions

J. Gani, CSIRO, *University of Kentucky, Lexington, Kentucky*. Applied Probability; Applied Probability Journals

A. E. Gelfand, *University of Connecticut, Storrs*. Archaeology, Statistics in

J. D. Gibbons, *University of Alabama, University, Alabama*. Brown-Mood Median Test

R. E. Glaser, *Lawrence Livermore National Laboratory, Livermore, California*. Bartlett's Test of Homogeneity of Variances

A. Goldman, *University of Nevada, Las Vegas, Nevada*. Blackjack

W. A. Golomski, *Chicago, Illinois*. American Society for Quality Control (ASQC)

I. J. Good, *Virginia Polytechnic Institute and State University, Blacksburg, Virginia*. Axioms of Probability

B. G. Greenberg, *University of North Carolina, Chapel Hill, North Carolina*. Biostatistics

S. J. Haberman, *University of Chicago, Chicago, Illinois*. Association, Measures of

D. M. Hawkins, *National Research Institute for Mathematical Statistics, Pretoria, South Africa*. Branch-and-Bound Methods

T. P. Hettmansperger, *Pennsylvania State University, University Park, Pennsylvania*. Bowker's Test for Symmetry

C. C. Heyde, CSIRO, *Canberra, Australia*. Australian Journal of Statistics

M. Hills, *British Museum, London, England*. Anthropology, Statistics in; Allometry

D. V. Hinkley, *Stanford University, Stanford, California*. Annals of Statistics

W. Hoeffding, *University of North Carolina, Chapel Hill, North Carolina*. Asymptotic Normality

R. V. Hogg, *University of Iowa, Iowa City, Iowa*. Adaptive Methods

G. M. Jenkins, *Lancaster, England*. Autoregressive-Integrated Moving Average (ARIMA) Models; Autoregressive-Moving Average (ARMA) Models

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

J. D. Kalbfleisch, *University of Waterloo, Waterloo, Ontario, Canada*. Ancillary Statistics

G. K. Kanji, *Sheffield City Polytechnic, Sheffield, England*. Bulletin in Applied Statistics (BIAS)

G. V. Kass, *University of the Witwatersrand, Johannesburg, South Africa*. Automatic Interaction Detection (AID) Techniques

J. R. Kettenring, *Bell Laboratories, Murray Hill, New Jersey*. Canonical Analysis

G. G. Koch, *University of North Carolina, Chapel Hill*. Chi-Square Tests; Chi-Square Tests, Numerical Examples

H. C. Kraemer, *Stanford University School of Medicine, Stanford, California*. Biserial Correlation

K. Krickeberg, *U.E.R. de Mathematiques, Logique Formelle et Informatique, Université Rene Descartes, Paris, France*. Bernoulli Society

R. G. Laha, *Bowling Green State University, Bowling Green, Ohio*. Characteristic Functions

H. O. Lancaster, *The University of Sydney, Sydney, New South Wales, Australia*. Chi-Square Distribution

F. C. Leone, *American Statistical Association, Washington, D.C.* American Statistical Association

E. A. Lew, *Punta Gorda, Florida*. Actuarial Statistics—Life

D. V. Lindley, *University College, London, England*. Bayesian Inference

J. I. McCool, *SFK Industries, Inc., King of Prussia, Pennsylvania*. Censored Data

B. H. Margolin, *National Institute for Environmental Health Sciences, Research Triangle Park, North Carolina.* Blocks, Balanced Incomplete; Blocks, Randomized Complete

D. H. Mellor, *Darwin College, Cambridge University, England.* Chance (II)

W. Morris, *Iowa State University, Ames, Iowa.* Chance (I)

J. Neter, *University of Georgia, Athens, Georgia.* American Statistican, The

H. Niederhausen, *Stanford University, Stanford, California.* Butler–Smirnov Test

R. M. Norton, *College of Charleston, Charleston, South Carolina.* Arc-Sine Distribution

J. L. Norwood, *Bureau of Labor Statistics, Washington, D.C.* Bureau of Labor Statistics

J. K. Ord, *Pennsylvania State University, University Park, Pennsylvania.* Aggregation

H. D. Patterson, *University of Edinburgh, Edinburgh, Scotland.* Changeover Designs

S. C. Pearce, *The University, Canterbury, England.* Analysis of Variance; Analysis of Covariance

S. K. Perng, *Kansas State University, Manhattan, Kansas.* Bahadur Efficiency

C. B. Read, *Southern Methodist University, Dallas, Texas.* Catastrophe Theory

G. K. Robinson, *Australian Road Research Board, Nunawading, Victoria, Australia.* Behrens-Fisher Problem

R. N. Rodriguez, *General Motors Laboratories, Warren, Michigan.* Burr Distributions

E. G. Schilling, *General Electric Company, Cleveland, Ohio.* Acceptance Sampling

G. A. F. Seber, *University of Auckland, Auckland, New Zealand.* Capture-Recapture Methods

E. Seneta, *University of Sydney, Sydney, New South Wales, Australia.* Abbe, Ernst; Bernstein, Sergei Natanovich; Bienaymé, Irenée-Jules; Boltzmann, Ludwig Edward; Boscovich, Ruggiero Giuseppe; Cauchy, Augustin-Louis; Chebyshev, Pafnuty Lvovich; Chuprov, Alexander Alexandrovich

R. J. Serfling, *Johns Hopkins University, Baltimore, Maryland.* Asymptotic Expansion

G. Shafer, *University of Kansas, Lawrence, Kansas.* Belief Functions; Bernoulli's, The

L. R. Shenton, *University of Georgia, Athens, Georgia.* Approximations to Distributions

W. L. Smith, *University of North Carolina, Chapel Hill, North Carolina.* Birth and Death Processes

B. Sowan, *Imperial College, London, England.* Biometrika

M. A. Stephens, *Simon Fraser University, British Columbia, Canada.* Anderson-Darling Test for Goodness of Fit

S. Stigler, *Center for Advanced Study in the Behavioral Sciences, University of Chicago, Chicago, Illinois.* Arithmetic Mean

M. E. Stokes, *University of North Carolina, Chapel Hill, North Carolina.* Chi-Square Tests, Numerical Examples

G. P. H. Styan, *McGill University, Montreal, Canada.* Canadian Journal of Statistics, The (La Revue Canadienne de Statistique)

L. Takács, *Case Western Reserve University, Cleveland, Ohio.* Ballot Problems

H. Taylor, *Cornell University, Ithaca, New York.* Brownian Motion

G. L. Tietjen, *Los Alamos Scientific Laboratory, Los Alamos, New Mexico.* Chemistry, Statistical Methods in

D. S. Tracy, *University of Windsor, Ontario, Canada.* Angle Brackets

R. K. Tsutakawa, *University of Missouri, Columbia, Missouri.* Bioassay, Statistical Methods in

G. J. G. Upton, *University of Essex, Colchester, England.* Categorical Data

G. L. Yang, *University of Maryland, College Park, Maryland.* Biometric Functions

S. L. Zabell, *Northwestern University, Evanston, Illinois.* Bortkiewicz, Ladislaus von

PREFACE

The purpose of this encyclopedia is to provide information about an extensive selection of topics concerned with statistical theory and the applications of statistical methods in various more or less scientific fields of activity. This information is intended primarily to be of value to readers who do not have detailed information about the topics but have encountered references (either by field or by use of specific terminology) that they wish to understand. The entries are not intended as condensed treatises containing all available knowledge on each topic. Indeed, we are on guard against placing too much emphasis on currently fashionable, but possibly ephemeral, matters. The selection of topics is also based on these principles. Nevertheless, the encyclopedia was planned on a broad basis—eight volumes, each of approximately 550 pages—so that it is possible to give attention to nearly all the many fields of inquiry in which statistical methods play a valuable (although not usually, or necessarily, a predominant) role.

Beyond the primary purpose of providing information, we endeavored to obtain articles that are pleasant and interesting to read and encourage browsing through the volumes. There are many contributors, for whose cooperation we are grateful, and a correspondingly wide range of styles of presentation, but we hope that each is attractive in its own way. There is also, naturally and inevitably, a good deal of variation among the (mathematical and technical-scientific) levels of the entries. For some topics, considerable mathematical sophistication is needed for adequate treatment; for others, it is possible to avoid heavy reliance on mathematical formulation.

We realize that even an eight-volume compendium cannot incorporate all of the terms, notions, and procedures that have appeared in statistical literature during the last century. There are also contributions by scientists who paved the way, as early as the seventeenth century, toward the statistical sciences as they are known today. We endeavored to include historical background and perspective when these seem important to the development of statistical methods and ideas.

It is to be expected that most readers will disagree with the relative amount of emphasis accorded to certain fields, and will find that some topics of considerable interest have been omitted. While this may reflect a lack of judgment or knowledge (or both) on our part, it is inevitable, because each person has a specific, idiosyncratic viewpoint on statistical matters (as on others). Our intention is to mirror the state of the art in the last quarter of the twentieth century, including terms (in particular mathematical) that found a place in the language of statistical methodology during its formative years.

We have two ways of cross-referencing: First, when a possibly unfamiliar term appears in an entry, reference to another entry is indicated by an asterisk, or by direct reference (e.g., *See* HISTOGRAMS). An asterisk sometimes refers to the preceding word but quite frequently to the preceding phrase. For example, " . . . random variable*" refers to the entry on random variables rather than on variables. We feel that this notation is the simplest possible. Second, most articles conclude with a list of related entries of potential interest for further reading. These two sets of cross-references may overlap but are usually not identical. The starred items are for utility, whereas the list is more for interest. Neither set is exhaustive and we encourage individual initiative in searching out further related entries.

Since our primary purpose is to provide information, we neither avoid controversial topics nor encourage purely polemic writing. We endeavor to give fair representation to different viewpoints but cannot even hope to approximate a just balance (if such a thing exists).

In accordance with this primary purpose, we believe that the imposition of specific rules of style and format, and levels of presentation, must be subordinate to the presentation of adequate and clear information. Also, in regard to notation, references, and similar minutiae, we did not insist on absolute uniformity although we tried to discourage very peculiar deviations that might confuse readers.

The encyclopedia is arranged lexicographically in order of entry titles. There are some inconsistencies; for example, we have "CHEMISTRY, STATISTICS IN" but "STATISTICS IN ASTRONOMY." This simply reflects the fact that the encyclopedia is being published serially, and the second of these entries was not available when the first volume was in production. (This volume does, however, contain the "dummy" entry "ASTRONOMY, STATISTICS IN *See* STATISTICS IN ASTRONOMY.")

We are indeed fortunate that Professor Campbell B. Read joined us as Associate Editor on October 1, 1980. Professor Read's active participation in the editorial process and the numerous improvements he contributed to this project have been invaluable. The Co-Editors-in-Chief express their sincerest appreciation of his expertise.

We also express our thanks to the members of the Advisory Board for their valuable advice and expert suggestions; to the Editorial Assistant, Ms. June Maxwell, for her devotion and for contributions to the project far beyond the call of duty; and last, but certainly not least, to all the contributors, who responded enthusiastically to our call for partnership in this undertaking.

Unsigned entries are contributed by the Editors—Samuel Kotz, Norman L. Johnson, and Campbell B. Read—either jointly or individually.

SAMUEL KOTZ
NORMAN L. JOHNSON

College Park, Maryland
Chapel Hill, North Carolina
January 1982

A

ABAC

A graph from which numerical values may be read off, usually by means of a grid of lines corresponding to argument values.

(NOMOGRAM)

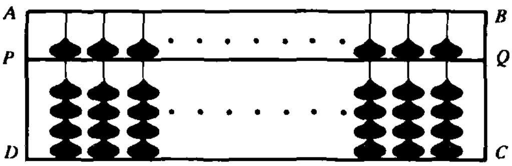

Figure 1. Diagrammatic representation of the form of abacus presently in common use.

ABACUS

A simple instrument to facilitate numerical computation. There are several forms of abacus. The one in most common use at present is represented diagramatically in Fig. 1. It consists of a rectangular framework *ABCD* with a cross-piece *PQ* parallel to the longer sides, *AB* and *CD*, of the rectangle. There are a number (at least eight, often more) of thin rods or wire inserted in the framework and passing through *PQ*, parallel to the shorter sides, *AD* and *BC*. On each rod there are threaded four beads between *CD* and *PQ*, and one bead between *PQ* and *AB*.

Analogously to the meaning of position in our number system, the extreme right-hand rod corresponds to units; the next to the left, tens; the next to the left, hundreds; and so on. Each bead in the lower rectangle (*PQCD*) counts for 1, when moved up, and each bead in the upper rectangle (*ABQP* counts for 5. The number shown in Fig. 2 would be 852 if beads on all rods except the three extreme right-hand ones are as shown for the three extreme left-hand rods (corresponding to "zero").

The Roman abacus consisted of a metal plate with two sets of parallel grooves, the lower containing four pebbles and the upper one pebble (with a value five times that of

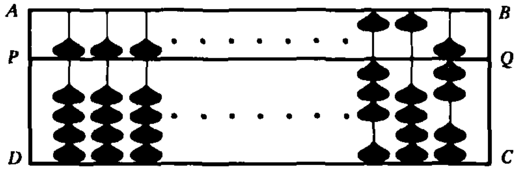

Figure 2. Abacus that would be showing the number 852 if beads on all rods except the three extreme right-hand ones are as shown for the three extreme left-hand rods (corresponding to "zero").

each pebble in the corresponding groove of the lower set). The Japanese and Chinese abacus (still in use) consists of a frame with beads on wires. The Russian abacus, which originated in the sixteenth century (the modern version in the eighteenth century), is also still in use.

Bibliography

Dilson, J. (1968). *The Abacus: A Pocket Computer*. St. Martin's Press, New York.

Gardner, M. (1979). *Mathematical Circus*. Alfred A. Knopf, New York, Chap. 18.

Pullan, J. (1969). *The History of the Abacus*. F. A. Praeger, New York.

ABBE, ERNST

Born: January 23, 1840, in Eisenach, Germany.

Died: January 14, 1905, in Jena, Germany.

Contributed to: theoretical and applied optics, astronomy, mathematical statistics.

The recognition of Abbe's academic talent by those in contact with him overcame a childhood of privation and a financially precarious situation very early in his academic career, when he completed "On the Law of Distribution of Errors in Observation Series," his inaugural dissertation for attaining a lectureship at Jena University at the age of 23 [1]. This dissertation, partly motivated by the work of C. F. Gauss*, seems to contain his only contributions to the probability analysis of observations subject to error.These contributions constitute a remarkable anticipation of later work in distribution theory and time-series* analysis, but they were overlooked until the late 1960s [5, 8], and almost none of the early bibliographies on probability and statistics (a notable exception being ref. 10) mention this work. In 1866, Abbe was approached by Carl Zeiss, who asked him to establish a scientific basis for the construction of microscopes; this was the beginning of a relationship that lasted throughout his life, and from this period on his main field of activity was optics [9] and astronomy.

Abbe shows, first, that the quantity $\Delta = \sum_{i=1}^{n} Z_i^2$, where Z_i, $i = 1, \ldots, n$, are n independently and identically distributed $N(0, 1)$ random variables, is described by a chi-square* density with n degrees of freedom [5, 8], although this discovery should perhaps be attributed to I. J. Bienaymé* [4]. Second, again initially by means of a "discontinuity factor" and then by complex variable methods, Abbe obtains the distribution of $\Theta = \sum_{j=1}^{n} (Z_j - Z_{j+1})^2$, where $Z_{n+1} = Z_1$, and ultimately that of Θ/Δ, a ratio of quadratic forms* in $Z_1, \ldots, Z_n$ very close in nature to the definition of what is now called the first circular serial correlation coefficient,* and whose distribution under the present conditions is essentially that used to test the null hypothesis of Gaussian white noise* against a first-order autoregression alternative, in time-series* analysis [3]. (The distribution under such a null hypothesis was obtained by R. L. Anderson in 1942.) Knopf [6] expresses Abbe's intention in his dissertation as being to seek a numerically expressible criterion to determine when differences between observed and sought values in a series of observations are due to chance alone.

References

[1] Abbe, E. (1863). *Über die Gesetzmässigkeit der Vertheilung der Fehler bei Beobachtungsreihen*. Hab. schrift., Jena (Reprinted as pp. 55–81 of ref. 2.)

[2] Abbe, E. (1906). *Gesammelte Abhandlungen*, Vol. 2. G. Fischer, Jena.

[3] Hannan, E. J. (1960). *Time Series Analysis*. Methuen, London, pp. 84–86.

[4] Heyde, C. C. and Seneta, E. (1977). *I. J. Bienaymé: Statistical Theory Anticipated*. Springer-Verlag, New York, p. 69.

[5] Kendall, M. G. (1971). *Biometrika*, **58**, 369–373. (Sketches Abbe's mathematical reasoning in relation to the contributions to mathematical statistics.)

[6] Knopf, O. (1905). *Jahresber. Dtsch. Math.-Ver.*, **14**, 217–230. [One of several obituaries by his associates; nonmathematical, with a photograph.

Another is by S. Czapski (1905), in *Verh. Dtsch. Phys. Ges.*, 7, 89–121.]

[7] Rohr, L. O. M. von (1940). *Ernst Abbe*. G. Fischer, Jena. (Not seen.)

[8] Sheynin, O. B. (1966). *Nature (Lond.)*, **211**, 1003–1004. (Notes Abbe's derivation of the chi-square density.)

[9] Volkman, H. (1966). *Appl. Opt.*, **5**, 1720–1731. (An English-language account of Abbe's life and contributions to pure and applied optics; contains two photographs of Abbe, and further bibliography.)

[10] Wölffing, E. (1899). *Math. Naturwiss. Ver. Württemberg* [Stuttgart], *Mitt.*, (2) **1**, 76–84. [Supplements the comprehensive bibliography given by E. Czuber (1899), in *Jahresber. Dtsch. Math.-Ver.*, 7 (2nd part), 1–279.]

(CHI-SQUARE DISTRIBUTION
QUADRATIC FORMS
SERIAL CORRELATION
COEFFICIENT
TIME-SERIES ANALYSIS)

E. SENETA

ABEL'S FORMULA

(Also known as the Abel identity.) If each term of a sequence of real numbers $\{a_i\}$ can be represented in the form $a_i = b_i c_i$, $i = 1, \ldots, n$, then $a_1 + a_2 + \cdots + a_n$ can be expressed as

$$s_1(b_1 - b_2) + s_2(b_2 - b_3) + \cdots + s_{n-1}(b_{n-1} - b_n),$$

where $s_i = c_1 + \cdots + c_i$. Equivalently,

$$\sum_{k=n}^{m} b_k c_k = B_m c_{m+1} - B_{n-1} c_n + \sum_{k=n}^{m} B_k (c_k - c_{k+1}),$$

where $B_k = \sum_{l=1}^{k} b_l$.

This representation is usually referred to as Abel's formula, due to Norwegian mathematician Niels Henrik Abel (1802–1829). (The continuous analog of this formula is the formula of integration by parts.) It is useful for manipulations with finite sums.

Bibliography

Knopp, K. (1951). *Theory and Application of Infinite Series*, 2nd ed. Blackie, London/Dover, New York.

ABSOLUTE ASYMPTOTIC EFFICIENCY (AAE) *See* ESTIMATION, POINT

ABSOLUTE CONTINUITY (of Measures on Infinite-Dimensional Linear Spaces)

Absolute continuity of measures, the Radon–Nikodym theorem*, and the Radon–Nikodym derivative* are subjects properly included in any basic text on measure and integration. However, both the mathematical theory and the range of applications can best be appreciated when the measures are defined on an infinite-dimensional linear topological space. For example, this setting is generally necessary if one wishes to discuss hypothesis testing* for stochastic processes with infinite parameter set. In this article we first define basic concepts in the area of absolute continuity, state general conditions for absolute continuity to hold, and then specialize to the case where the two measures are defined on either a separable Hilbert space or on an appropriate space of functions. Particular attention is paid to Gaussian measures.

The following basic material is discussed in many texts on measure theory; see, e.g., ref. 23. Suppose that (Ω, β) is a measurable space, and that μ_1 and μ_2 are two probability measures on (Ω, β). μ_1 is said to be absolutely continuous with respect to μ_2 ($\mu_1 \ll \mu_2$) if A in β and $\mu_2(A) = 0$ imply that $\mu_1(A) = 0$. This is equivalent to the following: $\mu_1 \ll \mu_2$ if and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that $\mu_2(A) < \delta$ implies that $\mu_1(A) \leqslant \epsilon$. Similar definitions of absolute continuity can be given for nonfinite signed measures; this article, however, is restricted to probability measures. When $\mu_1 \ll \mu_2$, the Radon–Nikodym theorem states that there exists a real-valued β-measurable function f such that $\mu_1(A) = \int_A f \, d\mu_2$ for all A in β. The function f, which belongs to $L_1[\Omega, \beta, \mu_2]$ and is unique up to μ_2-equivalence, is called the Radon–Nikodym derivative of μ_1 with respect to μ_2, and is commonly denoted by $d\mu_1/d\mu_2$. In statistical and engineering applications $d\mu_1/d\mu_2$ is usually called the likelihood ratio*, a term that has its genesis in maximum likelihood estimation*.

Absolute continuity and the Radon–Nikodym derivative have important applications in statistics. For example, suppose that $X : \Omega \to \mathbb{R}^N$ is a random vector. Suppose also that under hypothesis H_1 the distribution function of X is given by $F_1 = \mu_1 \circ X^{-1}$ $[F_1(x) = \mu_1\{\omega : X(\omega) \leqslant x\}]$, whereas under H_2, X has the distribution function $F_2 = \mu_2 \circ X^{-1}$. F_i defines a Borel measure on $\mathbb{R}^1$; one says that F_i is induced from μ_i by X. A statistician observes one realization (sample path) of X, and wishes to design a statistical test to optimally decide in favor of H_1 or H_2. Then, under any of several classical decision criteria of mathematical statistics (e.g., Bayes risk, Neyman–Pearson*, minimum probability of error), an optimum decision procedure* when $\mu_1 \ll \mu_2$ is to form the test statistic* $\Lambda(X) = [dF_1/dF_2](X)$ and compare its value with some constant, C_0; the decision is then to accept H_2 if $\Lambda(X) \leqslant C_0$, accept H_1 if $\Lambda(X) > C_0$. The value of C_0 will depend on the properties of F_1 and F_2 and on the optimality criterion. For more details, see HYPOTHESIS TESTING*.[1]

Two probability measures μ_1 and μ_2 on (Ω, β) are said to be equivalent $(\mu_1 \sim \mu_2)$ if $\mu_1 \ll \mu_2$ and $\mu_2 \ll \mu_1$. They are orthogonal, or extreme singular $(\mu_1 \perp \mu_2)$ if there exists a set A in β such that $\mu_2(A) = 0$ and $\mu_1(A) = 1$. For the hypothesis-testing problem discussed above, orthogonal induced measures permit one to discriminate perfectly between H_1 and H_2. In many practical applications, physical considerations rule out perfect discrimination. The study of conditions for absolute continuity then becomes important from the aspect of verifying that the mathematical model is valid.

In the framework described, the random vector has range in $\mathbb{R}^N$. However, absolute continuity, the Radon–Nikodym derivative, and their application to hypothesis-testing problems are not limited to such finite-dimensional cases. In fact, the brief comments above on hypothesis testing apply equally well when X takes its value in an infinite-dimensional linear topological space, as when $X(\omega)$ represents a sample path* from a stochastic process* (X_t), $t \in [a, b]$. (The infinite-dimensional case does introduce interesting mathematical complexities that are not present in the finite-dimensional case.)

GENERAL CONDITIONS FOR ABSOLUTE CONTINUITY

We shall see later that special conditions for absolute continuity can be given when the two measures involved have certain specialized properties, e.g., when they are both Gaussian. However, necessary and sufficient conditions for absolute continuity can be given that apply to any pair of probability measures on any measurable space (Ω, β). Further, if (Ω, β) consists of a linear topological space Ω and the smallest σ-field β containing all the open sets (the Borel σ-field), then additional conditions for absolute continuity can be obtained that apply to any pair of probability measures on (Ω, β). Here we give one well-known set of general necessary and sufficient conditions. First, recall that if (Ω, β, P) is a probability space and F a collection of real random variables on (Ω, β), then F is said to be uniformly integrable with respect to P [23] if the integrals $\int_{\{\omega : |f(\omega)| \geqslant c\}} |f(\omega)|\, dP(\omega)$, $c > 0$, f in F, tend uniformly to zero as $c \to \infty$. An equivalent statement is the following: F is uniformly integrable (P) if and only if

(a) $$\sup_F \int_\Omega |f(\omega)|\, dP(\omega) < \infty$$

and

(b) For every $\epsilon > 0$ there exists $\delta > 0$ such that $P(A) < \delta$ implies that

$$\sup_F \int_A |f(\omega)|\, dP(\omega) \leqslant \epsilon.$$

Theorem 1. Suppose that μ_1 and μ_2 are two probability measures on a measurable space (Ω, β). Suppose that $\{\mathcal{F}_n,\ n \geqslant 1\}$ is an increasing family of sub-σ-fields of β such that β is the smallest σ-field containing $\cup_n \mathcal{F}_n$. Let μ_i^n be the restriction of μ_i to $\mathcal{F}_n$. Then $\mu_1 \ll \mu_2$ if and only if

(a) $\mu_1^n \ll \mu_2^n$ for all $n \geqslant 1$,

and

(b) $\{d\mu_1^n / d\mu_2^n,\ n \geqslant 1\}$ is uniformly integrable (μ_2).

When $\mu_1 \ll \mu_2$, then $d\mu_1/d\mu_2 = \lim_n d\mu_1^n/d\mu_2^n$ almost everywhere (a.e.) $d\mu_2$.

Condition (a) of Theorem 1 is obviously necessary. The necessity of (b) follows from the fact that $\{d\mu_1^n/d\mu_2^n, \mathcal{F}_n : n \geqslant 1\}$ is a martingale* with respect to μ_2. This property, and the martingale convergence theorem, yield the result that $d\mu_1/d\mu_2 = \lim_n d\mu_1^n/d\mu_2^n$ a.e. $d\mu_2$. Sufficiency of (a) and (b) follows from the second definition of uniform integrability given above and the assumption that β is the smallest σ-field containing $\cup_n \mathcal{F}_n$.

Conditions (a) and (b) of Theorem 1 are also necessary and sufficient for $\mu_1 \ll \mu_2$ when the family of increasing σ-fields $(\mathcal{F}_t)$ has any directed index set.

A number of results frequently used to analyze absolute continuity can be obtained from Theorem 1. This includes, for example, Hájek's divergence criterion [20] and Kakutani's theorem on equivalence of infinite product measures [29] (a fundamental result in its own right).

The conditions of Theorem 1 are very general. However, in one respect they are somewhat unsatisfactory. They usually require that one specify an infinite sequence of Radon–Nikodym derivatives $\{d\mu_1^n/d\mu_2^n,\ n \geqslant 1\}$. It would be preferable to have a more direct method of determining if absolute continuity holds. One possible alternative when the measures are defined on a separable metric space involves the use of chacteristic functions*. The characteristic function of a probability measure defined on the Borel σ-field of a separable metric space completely and uniquely specifies the measure [38]. Thus in such a setting, two characteristic functions contain all the information required to determine whether absolute continuity exists between the associated pair of measures. The use of characteristic functions offers a method for attacking the following problem. For a given measure μ on (Ω, β) determine the set $\mathcal{P}_\mu$ of all probability measures on (Ω, β) such that $\nu \ll \mu$ for all ν in $\mathcal{P}_\mu$. Some results on this problem are contained in ref. 3; further progress, especially detailed results for the case of a Gaussian measure μ on Hilbert space, would be useful in several important applications areas (detection of signals in noise, stochastic filtering*, information theory*).

PROBABILITY MEASURES ON HILBERT SPACES

There has been much recent activity in the study of probability measures on Banach spaces [1, 4, 5, 31]. Here we restrict attention to the case of probabilities on Hilbert spaces; this is the most important class of Banach spaces for applications, and the theory is relatively well developed in this setting.

Let $\mathbf{H}$ be a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and Borel σ-field Γ. Let μ be a probability measure on Γ. For any element y in $\mathbf{H}$, define the distribution function F_y by $F_y(a) = \mu\{x : \langle y, x\rangle \leqslant a\}$, a in $(-\infty, \infty)$. μ is said to be Gaussian if F_y is Gaussian for all y in $\mathbf{H}$. It can be shown that for every Gaussian μ there exists a self-adjoint trace-class nonnegative linear operator R_μ in $\mathbf{H}$ and an element m_μ in $\mathbf{H}$ such that

$$\langle y, m_\mu \rangle = \int_{\mathbf{H}} \langle y, x\rangle \, d\mu(x) \tag{1}$$

and

$$\langle R_\mu y, v\rangle = \int_{\mathbf{H}} \langle y - m_\mu, x\rangle \langle v - m_\mu, x\rangle \, d\mu(x) \tag{2}$$

for all y and v in $\mathbf{H}$. R_μ is called the covariance (operator) of μ, and m_μ is the mean (element). Conversely, to every self-adjoint nonnegative trace-class operator R_μ and element m in $\mathbf{H}$ there corresponds a unique Gaussian measure μ such that relations (1) and (2) are satisfied. Non-Gaussian measures μ may also have a covariance operator R_μ and mean element m_μ satisfying (1) and (2); however, the covariance R_μ need not be trace-class. For more details on probability measures on Hilbert space, see refs. 17, 38, and 53.

Elegant solutions to many problems of classical probability theory (and applications) have been obtained in the Hilbert space framework, with methods frequently making use of the rich structure of the theory of linear operators. Examples of such problems include Sazanov's solution to obtaining necessary and sufficient conditions for a complex-valued function on $\mathbf{H}$ to be a characteristic function* [49]; Prohorov's conditions for weak compactness of families of probability measures, with applications to convergence of stochastic processes [43]; the results of Mourier on laws of large numbers* [34]; the results of Fortét and Mourier on the central limit theorem* [15, 34]; and conditions for absolute continuity of Gaussian measures. The latter problem is examined in some detail in the following section. The study of probability theory in a Hilbert space framework received much of its impetus from the pioneering work of Fortét and Mourier (see refs. 15 and 34, and the references cited in those papers). Their work led not only to the solution of many interesting problems set in Hilbert space, but also to extensions to Banach spaces and more general linear topological spaces [1, 4, 5, 15, 31, 34].

The infinite-dimensional Hilbert spaces $\mathbf{H}$ most frequently encountered in applications are $L_2[0, T]$ ($T < \infty$) and l_2. For a discussion of how Hilbert spaces frequently arise in engineering applications, *see* STATISTICAL COMMUNICATION THEORY. In particular, the interest in Gaussian measures on Hilbert space has much of its origin in hypothesis-testing and estimation problems involving stochastic processes: detection and filtering of signals embedded in Gaussian noise. For many engineering applications, the noise can be realistically modeled as a Gaussian stochastic process with sample paths almost surely (a.s.) in $L_2[0, T]$ or a.s. in l_2. When $\mathbf{H}$ is $L_2[0, T]$, a trace-class covariance operator can be represented as an integral operator whose kernel is a covariance function. Thus suppose that (X_t), $t \in [0, T]$, is a measurable zero-mean stochastic process on (Ω, β, P), inducing the measure μ on the Borel σ-field of $L_2[0, T]$; $\mu(A) = P\{\omega : X(\omega) \in A\}$. Then $E\int_0^T X_t^2(\omega)\,dt < \infty$ if and only if μ has a trace-class covariance operator R_μ defined by $[R_\mu f](t) = \int_0^T R(t,s) f(s)\,ds$, f in $L_2[0, T]$, where R is the covariance function of (X_t). If R_μ is trace-class, then $E\int_0^T X_t^2(\omega)\,dt = \text{trace } R_\mu$.

ABSOLUTE CONTINUITY OF PROBABILITY MEASURES ON HILBERT SPACE

If $\mathbf{H}$ is finite-dimensional and μ_1 and μ_2 are two zero-mean Gaussian measures on Γ, it is easy to see that μ_1 and μ_2 are equivalent if and only if their covariance matrices have the same range space. However, if $\mathbf{H}$ is infinite-dimensional, this condition (on the ranges of the covariance operators) is neither necessary nor sufficient for $\mu_1 \sim \mu_2$. The study of conditions for absolute continuity of two Gaussian measures on function space has a long and active history. Major early contributions were made by Cameron and Martin [6, 7] and by Grenander [18]. The work of Cameron and Martin was concerned with the case when one measure is Wiener measure (the measure induced on $C[0, 1]$ by the Wiener process*) and the second measure is obtained from Wiener measure by an affine transformation. Grenander obtained conditions for absolute continuity of a Gaussian measure (induced by a stochastic process with continuous covariance) with respect to a translation. Segal [50] extended the work of Cameron and Martin to a more general class of affine transformations of Wiener measure. Segal also obtained [50] conditions for absolute continuity of Gaussian "weak distributions." These necessary and sufficient conditions can be readily applied to obtain sufficient conditions for equivalence of any pair of Gaussian measures on **H**; they can also be used to show that these same conditions are necessary. Complete and general solutions to the absolute continuity problem for Gaussian measures were obtained by Feldman [12] and Hájek [21]. Their methods are quite different. The main result, in each paper, consists of two parts: a "dichotomy theorem," which states that any two Gaussian measures are either equivalent or orthogonal; and conditions that are necessary and sufficient for equivalence. The following theorem for Gaussian measures on Hilbert space is a modified version of Feldman's result [12]; several proofs have been independently obtained (Kallianpur and Oodaira [30], Rao and Varadarajan [44], Root [45]).

Theorem 2. Suppose that μ_1 and μ_2 are two Gaussian measures on Γ, and that μ_i has covariance operator R_i and mean m_i, $i = 1$, 2. Then:

1 either $\mu_1 \sim \mu_2$ or $\mu_1 \perp \mu_2$;
2 $\mu_1 \sim \mu_2$ if and only if all the following conditions are satisfied:
   (a) range $(R_1^{1/2}) =$ range $(R_2^{1/2})$;
   (b) $R_1 = R_2^{1/2}(I + T)R_2^{1/2}$, where I is the identity on **H** and T is a Hilbert–Schmidt operator in **H**:
   (c) $m_1 - m_2$ is in range $(R_1^{1/2})$.

Various specializations of Theorem 2 have been obtained; see the references in refs. 8 and 47. Two of the more interesting special cases, both extensively analyzed, are the following: (1) both measures induced by stationary Gaussian stochastic processes; (2) one of the measures is Wiener measure. In the former case, especially simple conditions can be given when the two processes have rational spectral densities; see the papers by Feldman [13], Hájek [22], and Pisarenko [40, 41]. In this case, when the two measures have the same mean function, $\mu_1 \sim \mu_2$ if and only if $\lim_{|\lambda|\to\infty} f_1(\lambda)/f_2(\lambda) = 1$, where f_i is the spectral density* of the Gaussian process inducing μ_i. Moreover, this occurs if and only if the operator T appearing in Theorem 2 is also trace-class [22]. For the case where one of the measures is Wiener measure, see the papers by Shepp [51], Varberg [54, 55], and Hitsuda [24].

The problem of determining the Radon–Nikodym derivative for two equivalent Gaussian measures on a Hilbert space has been studied, especially by Rao and Varadarajan [44]. For convenience, we use the notation of Theorem 2 and assume now that all covariance operators are strictly positive. In the case where the Hilbert space is finite-dimensional, the log of the Radon-Nikodym derivative $d\mu_1/d\mu_2$ (log-likelihood ratio*) is easily seen to be a quadratic-linear form; that is, $\log \Lambda(X) = \langle x, Wx\rangle + \langle x, b\rangle +$ constant, where the linear operator $W = \frac{1}{2}(R_2^{-1} - R_1^{-1})$, $b = R_1^{-1}m_1 - R_2^{-1}m_2$, and $\log \equiv \log_e$. However, when **H** is infinite-dimensional, the log-likelihood ratio need not be a quadratic-linear form defined by a bounded linear operator. This holds true even if the operator T of Theorem 2 is not only Hilbert–Schmidt, but is also trace class. However, when T is Hilbert-Schmidt, one can always express the log of the Radon–Nikodym derivative as an almost surely convergent series [44]. The essential difficulty in characterizing the likelihood ratio for infinite-dimensional Hilbert space is that the operators R_1 and R_2 cannot have *bounded* inverses and these two inverses need not have the same domain of definition. Even if range $(R_1) =$ range(R_2), so that $R_2^{-1} - R_1^{-1}$ is defined on range (R_1), it is not necessary that $R_2^{-1} - R_1^{-1}$ be bounded on range (R_1).

In the finite-dimensional case, if $R_1 = R_2$, then $\log \Lambda(X) = \langle x, b\rangle +$ constant, with b defined as above, so that the log-likelihood ratio is a bounded linear form. This need not be the case for infinite-dimensional Hilbert space; in general, $\log \Lambda(X)$ will be a bounded linear form (when $R_1 = R_2$) if and only if $m_1 - m_2$ is in the range of R_1. As can be seen from Theorem 1, this condition is strictly stronger than the necessary and sufficient condition for $\mu_1 \sim \mu_2$, which (with $R_1 = R_2$) is that $m_1 - m_2$ be in range $(R_1^{1/2})$.

If the two measures are induced by stationary Gaussian processes with rational spectral densities, expressions for the likelihood ratio can be given in terms of the spectral densitites; see the papers by Pisarenko [41] and Hájek [22].

In many applications, only one of the two measures can be considered to be Gaussian. For this case, a useful sufficient condition for absolute continuity is given in ref. 2. This condition can be applied when the two measures are induced by stochastic processes (X_t) and (Y_t), where (Y_t) is a function of (X_t) and a process (Z_t) that is independent of (X_t). In particular, if (X_t) is Gaussian and $(Y_t) = (X_t + Z_t)$, then conditions for absolute continuity can be stated in terms of sample path properties of the (Z_t) process (absolute continuity, differentiability, etc.). Such conditions can often be verified in physical models by knowledge of the mechanisms generating the observed data, when the distributional properties of the (Z_t) process are unknown. When (X_t) is the Wiener process on $[0, T]$, conditions for absolute continuity of the induced measures on $L_2[0, T]$ can be obtained from the results of refs. 10, 27, and 28. Some of these results do not require independence of (X_t) and (Z_t).

Other results on absolute continuity of measures on Hilbert space have been obtained for infinitely divisible measures [16], measures induced by stochastic processes with independent increments [16], admissible translations of measures [42, 52], and for a fixed measure and a second measure obtained from the first measure by a nonlinear transformation [16]. With respect to admissible translates, Rao and Varadarajan [44] have shown that if μ is a zero-mean measure having a trace-class covariance operator, R, then the translate of μ by an element y is orthogonal to μ if y is not in range $(R^{1/2})$. A number of these results are collected in the book by Gihman and Skorohod [17], which also contains much material on basic properties of probability measures on Hilbert space, and on weak convergence*. The book by Kuo [33] contains not only basic material on probability measures on Hilbert spaces (including absolute continuity), but also an introduction to some topics in probability on Banach spaces.

ABSOLUTE CONTINUITY OF MEASURES INDUCED BY STOCHASTIC PROCESSES

Many problems involving stochastic processes are adequately modeled in the framework of probability measures on Hilbert space, provided that the sample paths of each process of interest belong almost surely to some separable Hilbert space. However, this condition is not always satisfied; even when it is satisfied, one may prefer conditions for absolute continuity stated in terms of measures on $\mathbb{R}^T$ (the space of real-valued functions on T), where T is the parameter set of the process. For example, a class of stochastic processes frequently considered are those having almost all paths in $D[0, 1]$. $D[0, 1]$ is the set of all real-valued functions having limits from both left and right existing at all points of $(0, 1)$, with either left-continuity or right-continuity at each point of $(0, 1)$, and with a limit from the left (right) existing at 1(0). $D[0, 1]$ is a linear metric space* under the Skorohod metric [38], but this metric space is not a Hilbert space.

The general conditions for absolute continuity stated in Theorem 1 apply in any setting. Moreover, necessary and sufficient conditions for equivalence of measures (most frequently on $\mathbb{R}^T$) induced by two Gaussian stochastic processes can be stated in a number of ways: The reproducing kernel Hilbert space (r.k.H.s.) of the two covariance functions [30, 37, 39]; operators and elements in an L_2 space of real-valued random functions [12]; operators and elements in an L_2-space of random variables [46]; and tensor products [35]. Hájek's conditions for absolute continuity in terms of the divergence [21] apply to the general case. Sato [48] has stated conditions for absolute continuity in terms of a representation for all Gaussian processes whose induced measure on $\mathbb{R}^T$ is equivalent to the measure induced by a given Gaussian process. Several of these results are presented in [8]. Many other papers on absolute continuity for measures induced by two Gaussian processes have appeared; space does not permit an attempt at a complete bibliography.

Use of the r.k.H.s. approach to study linear statistical problems in stochastic processes was first explicitly and systematically employed by Parzen; the r.k.H.s. approach was also implicit in the work of Hájek (see the papers by Hájek [22] and Parzen [39] and their references).

For non-Gaussian processes, results on absolute continuity have been obtained for Markov processes* [16, 32], diffusion processes* [36], locally infinitely divisible processes [16], semimartingales* [25], point processes* [26], and non-Gaussian processes equivalent to the Wiener process [9, 10, 27, 28].

Dudley's result [9] is of particular interest to researchers interested in Gaussian measures. Suppose that (W_t) is the Wiener process on $[0, 1]$ with zero mean and unity variance parameter, and that $\beta(\cdot, \cdot)$ is a continuous real-valued function on $\mathbb{R} \times [0, 1]$. Let $Y_t = \beta(W_t, t)$. Dudley shows in ref. 9 that the measure on function space induced by (Y_t) is absolutely continuous with respect to Wiener measure if and only if $\beta(u, t) = u + \phi(t)$ or $\beta(u, t) = -u + \phi(t)$, where ϕ is in the r.k.H.s. of the Wiener covariance $\min(t, s)$. The methods used to prove this result rely heavily on some of the special properties of the Wiener process, such as the fact that (W_t) has the strong Markov property*, and laws of the iterated logarithm* for the Wiener process (obtained in ref. 9). A characterization of admissible β's for other Gaussian processes with continuous paths would be of much interest; such characterizations would necessarily require a different approach, and this problem is very much open at present.

The absolute continuity problem discussed in refs. 10, 27, and 28 has received much attention, partly because of its connection to signal detection* and nonlinear filtering*. One considers a measurable process (Y_t) defined by $Y_t = \int_0^t h_s\, ds + W_t$, $0 \leqslant t \leqslant T$, where (W_t) is a zero-mean Wiener process and (h_s) is a stochastic process with sample paths a.s. in $L_1[0, T]$. Let μ_Y and μ_W be the measures induced by (Y_t) and (W_t) on the space of continuous functions on $[0, 1]$. Conditions for $\mu_Y \ll \mu_W$, $\mu_Y \sim \mu_W$, and results on the Radon–Nikodym derivative have been obtained in refs. 10, 27, and 28. In the special case where (h_s) is independent of (W_t), a sufficient condition for $\mu_Y \sim \mu_W$ is that $\int_0^T h_s^2\, ds < \infty$ for almost all sample paths of (h_s). This condition is also sufficient for $\mu_Y \ll \mu_W$ if the process (h_s) is only assumed independent of future increments of (W_t).

Finally, we mention a result of Fortét [14], who has obtained a sufficient condition for orthogonality of two measures when one is Gaussian, expressed in terms of the r.k.H.s. of the two covariances. Suppose that μ_i is a probabliity measure on $\mathbb{R}^T$, $T = [0, 1]$, with r.k.H.s. H_i and mean function m_i. Then if μ_1 is Gaussian, μ_1 and μ_2 are orthogonal unless both the following conditions are satisfied: (a) $H_1 \subset H_2$; and (b) $m_1 - m_2 \in H_2$.

NOTE

1. The ubiquitous nature of the Radon–Nikodym derivative in various hypothesis-testing applications can be attributed to its being a necessary and sufficient statistic* [11].

References

[1] Aron, R. M. and Dineen, S., eds. (1978). Vector Space Measures and Applications, I. *Lect. Notes Math.*, **644**. (Contains most of the papers on probability theory presented at the Conference on Vector Space Measures and Applications, Dublin, 1978.)

[2] Baker, C. R. (1973). *Ann. Prob.*, **1**, 690–698.

[3] Baker, C. R. (1979). *Lect. Notes Math.*, **709**, 33–44.

[4] Beck, A., ed. (1976). Probability in Banach spaces. *Lect. Notes Math.*, **526**. (Collection of papers on various topics; Proceedings of First International Conference on Probability in Banach Spaces, Oberwolfach, 1975.)