

**Citations for the IMS Bulletin Article on 215  
Influential Ideas and Discoveries in Statistics,  
1650 – 2010**

Anirban DasGupta, Purdue University, USA

## 1 215 Influential Publications in Statistics

On July 18, 2013, I wrote an email to Steve Stigler, which said: "Dear Steve, has any well known person made a list of what she or he considers to be the most major inventions/discoveries in statistics? .. I ask because I prefer to use another person's list than write one myself." On July 22, Steve thus replied: "Anirban, some time ago there was a list of 10 discoveries in the sciences and the chi-sq was one. I am working on a book on the 5 most consequential ideas in stat, but it is quite different in scope and flavor. So go ahead and make a list - my guess is that different people's lists will show little overlap."

So, blame it on Steve, here, in my last column for the Bulletin, I enter into the absurdly perilous territory of making a personal list of 215 *influential and original developments* in statistics, primarily to provide a sense of our lavish and multifarious heritage to a fresh PhD in statistics. I choose items that have influenced the research or practice or education or thinking of many people across the world; the criterion was not deep theorems per se, rather innovations and publications with a universal and scopic impact. The choices are of course personal, but not prejudicial; no one should take them too literally. I had it looked over by thirteen very senior world statisticians. There is, predictably, an overlap of this expanded list with the Springer *Breakthroughs in Statistics* volumes. I list the items followed by what I know to be the first serious suggestion or origin of the particular item; this is of course very difficult and some errors are likely!

*215 Developments in Statistics*: Likelihood 1657, WLLN 1713, CLT 1738, Bayes' thm 1763, scan statistics 1767, latin sq 1782, sampling/surveys (The Bible)/1786, least squares 1805, normal distn 1809, Poisson process 1837, outlier detection 1852, Chebyshev ineq 1853, optimal design 1876, regression 1877, correlation 1888, Edgeworth expansion 1889, histogram 1891, mixture models 1894, Pearson family 1895, periodogram 1898, chi-sq test 1900, Gauss-Markov thm 1900,  $P$ - values 1900, Wiener process 1900, PCA 1901, Factor anal. 1904, meta analysis 1904, Lorenz curve 1905,  $t$  test 1908, maximum likelihood, 1912, variance stabilization 1915, SEM 1921, ANOVA 1921, sufficiency 1922, Fisher-Yates test 1922, Pólya urns 1923,

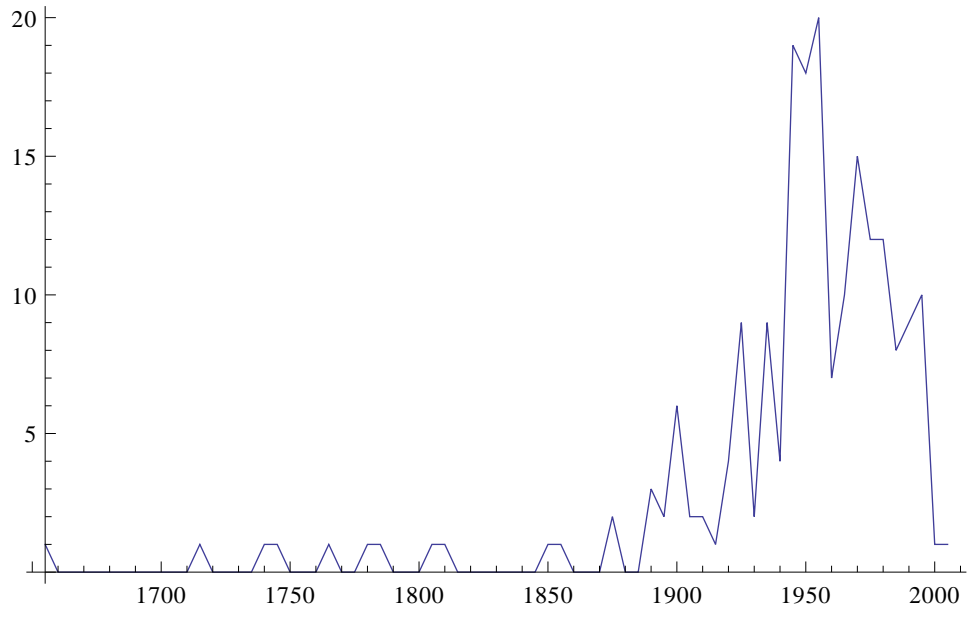
exchangeability 1924, Slutsky 1925, stable laws 1925, Fisher inf 1925, factorial designs 1926, normal extremes 1927, Wishart distn 1928, Control charts 1931, Neyman-Pearson tests 1933, K-S test 1933, martingales 1935, Exp family 1936, Fisher's LDF 1936, Mahalanobis distance 1936, canonical corr 1936, conf intervals 1937, permutation tests 1937,  $F$  test 1937, Wilks' thm 1938, large dev. 1938, Cornish-Fisher expansions 1938, Kendall's  $\tau$  1938, Pitman estimates 1939, admissibility/minimaxity 1939, errors in variables 1940, BIBD 1940, Berry Esseen thm 1941, Wald test 1943, Ito integral 1944, Wilcoxon's test 1945, SPRT 1945, Cramér-Rao ineq 1945, two stage estimators 1945, asymp normality MLE 1946, Jeffreys prior 1946, Plackett-Burman designs 1946, delta method 1946, resampling 1946, Monte Carlo 1946, Kendall and Stuart 1st ed 1946, Hunt-Stein thm 1946, Rao Blackwell 1947, Mann Whitney test 1947, contingency table testing 1947, orthogonal arrays 1947, score test 1948,  $U$  statistics 1948, Anscombe residuals 1948, Pitman efficiency 1948, Bernstein- von Mises 1949, jackknife 1949, test for additivity 1949, Feller Vol I, 1950, Lehmann-Scheffe thm 1950, Durbin-Watson test 1950, NN classification 1951, long range dep. 1951, invariance principle 1951, ARMA models 1951, thresholding 1952, Horvitz-Thompson est 1952, Lehmann's TSH 1953, semiparametric models 1953, Scheffe intervals 1953, Whittle estimator 1953, MCMC 1953, cusum charts 1954, ranking and selection 1954, saddlepoint approx 1954, test for change point 1955, Basu's thm 1955, shrinkage 1956, kernel density est 1956, empirical Bayes 1956, monotone likelihood ratio 1956, DKW 1956, adaptive inference 1956, Grenander est 1956,  $k$ -means clustering 1957, Roy's root test 1957, Kaplan-Meier 1958, FDA 1958, Copulas 1959,  $D$ -optimality 1959, LAN 1960, Kalman filter 1960, HMM 1960, conjugate priors 1961, Bayes factor 1961, James Stein est 1961, Hartley-Rao method 1962, kriging 1963, variogram 1963, ridge regr. 1963, Monte Carlo tests 1963,  $M$  est 1964, nonparametric regression 1964, Box-Cox transformations 1964, growth curve models 1964, Shapiro -Wilk test 1965, random coefficient regression 1965, Jolly-Seber models 1965, regression with  $n < p$  1966, Bahadur expansion 1966, Bahadur slope 1967, robust Bayes 1967, total positivity 1968, q-q plot 1968, longitudinal data 1968, fractional Brownian motion 1968, Grubbs' tests 1969, AIC 1969, repeated sig tests 1969, Box Jenkins method 1970, empirical risk min 1971, VC classes 1971, penalized

density est 1971, proportional hazards 1972, Stein's method 1972, GLM 1972, isotonic regression 1972, Stein's identity 1973,  $C_p$  1973, robust regression 1973, Ferguson priors 1973, influence function 1974, proj pursuit 1974, opt est eqns 1974, covariance regularization 1975, partial likelihood 1975, KMT embedding 1975, imputation 1976, CV 1977, EDA 1977, EM 1977, consistent nonparametric regression 1977, BIC 1978, quantile regression 1978, Donsker classes 1978, bootstrap 1979, reference priors 1979, unit root test 1979, bootstrap consistency 1981, Pickands dependence function 1981, ARCH 1982, graphical models 1983, magic formula 1983, propensity scores 1983, CART 1984, block bootstrap 1985, Chinese restaurant process 1985, Bayesian networks 1985, generalized est eqns 1986, prepivoting 1987, Daubechies wavelets 1988, EL 1988, multiresolution analysis 1989, landmark gaussian distn 1989, boosting 1990, smoothing splines 1990, deconvolution ests 1990, SIR 1991, automatic bandwidths 1991, functional regression 1991, supersaturated designs 1993, local linear smoothers 1993, oracles 1994, bagging 1994, stationary bootstrap 1994, SureShrink 1995, reversible jump MCMC 1995, FDR 1995, soft margin SVM 1995, Lasso 1996, block thresholding 1999, LAR 2004, group lasso 2006, sparse PCA 2006, covariance banding and tapering 2008.

Here is a breakdown of the items into categories: descriptive statistics 1, books 3, sampling and design 11, non and semiparametrics 18, probability 36, parametric inference 66, models and methodology 80.

Let me also report a time series plot of occurrences of these developments in 5 year intervals. We cannot expect that innovations have checked in at a uniform rate over 350 years. There are some seven major peaks in the plot around special periods, corresponding, apparently, to the Pearson age, the Fisher age, the optimality/nonparametrics age, the methodology age, the robustness/heavy math age, the bootstrap/computer age, and the HD age. Someone else's list may produce other peaks.

Time Series Plot of Statistical Discoveries



## References

- Akaike, H. (1969) (*AIC*). Fitting autoregressive models for prediction, Ann. Inst. Stat. Math., 21, 243-247.
- Aldous, D. (1985) (*Chinese restaurant process*). Exchangeability and related topics, In Ecole d'ete de probabilités de Saint-Flour, XIII, 1-198, Springer, Berlin.
- Anscombe, F. (1948) (*Anscombe Residuals*). The transformation of Poisson, Binomial, and negative-binomial data, Biometrika, 35, 246-254.
- Armitage, P., McPherson, C. and Rowe, B. (1969) (*Repeated significance tests*). Repeated significance tests on accumulating data, JRSSA, 132, 235-244.
- Bachelier, L. (1900) (*Wiener process*). The Theory of Speculation, PhD Dissertation.
- Bahadur, R.R. (1966) (*Bahadur expansion*). A note on quantiles in large samples, Ann. Math. Statist., 37, 577- 580.
- Bahadur, R.R. (1967) (*Bahadur slope*). Rates of convergence of estimates and test statistics, Ann. Math. Statist., 38, 303-324.
- Barlow, R. et al. (1972) (*Isotonic regression*). Statistical Inference under Order Restrictions: Theory and Application of Isotonic Regression, John Wiley.
- Barnard, G. (1947) (*Contingency table testing*). Significance tests for  $2 \times 2$  tables, Biometrika, 34, 123-138.
- Barnard, G. (1963) (*Monte Carlo tests*). Discussion on The Spectral Analysis of Point Processes by M.S. Bartlett, JRSSB, 25, 294.
- Barndorff-Nielsen, O. (1983) (*Magic formula*). On a formula for the distribution of a maximum likelihood estimator, Biometrika, 70, 343-385.
- Basu, D. (1955) (*Basu's theorem*). On statistics independent of a complete sufficient statistic, Sankhya, 15, 377-380.
- Bayes, T. (1763) (*Bayes' theorem*). An essay toward solving a problem in the doctrine of chances, Philos Trans Royal Soc London, 53, 370-418.
- Bechhofer, R. (1954) (*Ranking and selection*). A single-sample multiple-decision procedure for ranking means of normal populations with known variances, Ann. Math. Statist., 25, 16-39.
- Benjamini, Y. and Hochberg, Y. (1995) (*FDR*). Controlling the false discov-

ery rate: A practical and powerful approach to multiple testing, JRSSB, 57, 289-300.

Beran, R. (1987) (*Prepivoting*). Prepivoting to reduce level error of confidence sets, Biometrika, 74, 457-468.

Bernardo, J. (1979) (*Reference priors*). Reference posterior distributions for Bayesian inference, JRSSB, 41, 113-147.

Bernoulli, J. (1713) (*WLLN*). Ars Conjectandi.

Berry, A. (1941) (*Berry-Esseen theorem*). The accuracy of the Gaussian approximation to the sum of independent variates, Trans. Amer. Math. Soc., 49, 122-136.

Bickel, P. and Freedman, D. (1981) (*Bootstrap consistency*). Some asymptotic theory for the bootstrap, Ann. Statist., 9, 1196-1217.

Bickel, P. and Levina, E. (2008) (*Covariance banding and tapering*). Regularized estimation of large covariance matrices, Ann. Statist., 36, 199-227.

Bienayme, L. (1853) (*Chebyshev's inequality*). Considerations al'appui de la decouverte de Laplace, Comptes Rendus de l'Academie des Sciences, 37, 309-324.

Blackwell, D. (1947) (*Rao-Blackwell theorem*). Conditional expectation and unbiased sequential estimation, Ann. Math. Statist., 18, 105-110.

Blum, J. and Rosenblatt, J. (1967) (*Robust Bayes*). On partial apriori information in statistical inference, Ann. Math. Statist., 38, 1671-1678.

Bosq, D. (1991) (*Functional regression*). Modelization, nonparametric estimation and prediction for continuous time processes, In Nonparametric Functional Estimation and Related Topics, G. Roussas Ed, 335, 509-529, Kluwer.

Box, G. and Cox, D. (1964) (*Box-Cox transformations*). An analysis of transformations, JRSSB, 26, 211-252.

Box, G. and Jenkins, G. (1970) (*Box-Jenkins method*). Time Series Analysis: Forecasting and Control, Holden Day.

Breiman, L. et al. (1984) (*CART*). Classification and Regression Trees, Wadsworth.

Breiman, L. (1994) (*Bagging*). Bagging Predictors, Tech report, Univ Calif. Berkeley.

Cai, T. (1999) (*Block thresholding*). Adaptive wavelet estimation: A block

thresholding and oracle inequality approach, *Ann. Statist.*, 27, 898-924.

Cleveland, W. (1979) (*LOESS*). Robust locally weighted regression and smoothing scatterplots, *JASA*, 74, 829-836.

Cornish, E. and Fisher, R.A. (1938) (*Cornish-Fisher expansions*). Moments and cumulants in the specification of distributions, *Revue de l'Institut Internat. de Statistique*, 5, 307-322.

Cortes, C. and Vapnik, V. (1995) (*Soft margin SVM*). Support vector networks, *Machine Learning*, 20, 273-297.

Cox, D. (1972) (*Proportional hazards*). Regression models and life-tables, *JRSSB*, 34, 187-220.

Cox, D. (1975) (*Partial likelihood*). Partial likelihood, *Biometrika*, 62, 269-276.

Cramér, H. (1938) (*Large deviations*). Sur un nouveau theoreme-limite de la theorie des probabilites, *ASI*, 736, 2-23.

Cramér, H. (1946) (*Asymptotic normality of MLE*). *Mathematical Methods of Statistics*, Princeton Univ. Press.

Daniels, H. (1954) (*Saddlepoint approximations*). Saddlepoint approximations in statistics, *Ann. Math. Statist.*, 25, 631-650.

Daubechies, I. (1988) (*Daubechies wavelets*). Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, 41, 909-996.

de Moivre, A. (1738) (*CLT*). *The Doctrine of Chances*, 2nd Ed.

Dempster, A., Laird, N. and Rubin, D. (1977) (*EM algorithm*). Maximum likelihood from incomplete data via the EM algorithm, *JRSSB*, 39, 1-38.

Dickey, D. and Fuller, W. (1979) (*Unit root test*). Distribution of the estimators for autoregressive time series with a unit root, *JASA*, 74, 427-431.

Donoho, D., and Johnstone, I. (1994) (*Oracles*). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 81, 425-455.

Donoho, D., and Johnstone, I. (1995) (*SureShrink*). Adapting to unknown smoothness via wavelet shrinkage, *JASA*, 90, 1200- 1224.

Donsker, M. (1951) (*Invariance principle*). The invariance principle for certain probability limit theorems, *Mem Amer Math Soc*, 6, 1-10.

Doob, J. (1949) (*Bernstein-von Mises*). Applications of the theory of martingales, *Colloq Intem du C.N.R.S (Paris)*, 13, 22-28.

Dudley, R. (1978) (*Donsker classes*). Central limit theorems for empirical



measures, *Ann. Prob.*, 6, 899-929.

Durbin, J. and Watson, G. (1950) (*Durbin-Watson test*). Testing for serial correlation in least squares regression, *Biometrika*, 37, 409-428.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956) (*DKW inequality*). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *Ann. Math. Statist.*, 27, 642-669.

Efron, B. (1979) (*Bootstrap*). Bootstrap methods: Another look at the Jack-knife, *Ann. Statist.*, 7, 1-26.

Efron, B. et al. (2004) (*Least angle regression*). Least angle regression, With discussion and a rejoinder by the authors, *Ann. Statist.*, 32, 407-499

Eggenberger, F. and Pólya, G. (1923) (*Pólya urns*). Über die Statistik verketteter Vorgänge, *ZAMM*, 3, 4, 279-289.

Engle, R. (1982) (*ARCH models*). Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation, *Econometrica*, 50, 987-1008.

Euler, L. (1782) (*Latin squares*). Recherches sur une nouvelle espèce de quarrés magiques, *Verh. Zee Genoot Weten Vliss*, 9, 85-239.

Fan, J. (1993) (*Local linear smoothers*). Local linear regression smoothers and their minimax efficiencies, *Ann. Statist.*, 21, 196-216.

Ferguson, T. (1973) (*Ferguson priors*). Bayesian analysis of some nonparametric problems, *Ann. Statist.*, 1, 209-230.

Fisher, R.A. (1912) (*Maximum likelihood*). On an absolute criterion for fitting frequency curves, *Messenger of Math.*, 41, 155-160.

Fisher, R.A. (1915) (*Variance stabilization*). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population, *Biometrika*, 10, 4, 507-521.

Fisher, R.A. (1921) (*ANOVA*). On the probable error of a coefficient of correlation deduced from a small sample, *Metron*, 1, 3-32.

Fisher, R.A. (1922) (*Fisher-Yates test*). On the interpretation of  $\chi^2$  from contingency tables and the calculation of  $P$ , *JRSS*, 85, 1, 87-94.

Fisher, R.A. (1922) (*Sufficiency*). On the Mathematical Foundations of Theoretical Statistics, *Philos Trans Royal Soc London, Ser A*, 222, 309-368.

Fisher, R.A. (1925) (*Fisher information*). Theory of Statistical Estimation, *Proc. Cambridge Philos. Soc.*, 22, 700-725.

- Fisher, R.A. (1926) (*Factorial designs*). The arrangement of field experiments, Jour Ministry Agr Great Britain, 33, 503-513.
- Fisher, R.A. (1936) (*Fisher's LDF*). The use of multiple measurements in taxonomic problems, Ann. Eugenics, 179-188.
- Fisher, R.A. (1940) (*BIBD*). An examination of the different possible solutions of a problem in incomplete blocks, Ann. Eugenics, 10, 52-75.
- Fix, E. and Hodges, J. (1951) (*NN classification*). Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties, Tech Report, USAF, School of Aviation Medicine, Randolph Field, TX.
- Feller, W. (1950). An Introduction to Probability Theory and its Applications, Vol I, John Wiley.
- Frechet, M. (1927) (*Normal extremes asymptotic distributions*). Sur la loi de probabilit e de l'ecart maximum, Ann. Soc. Polon Math., 6, 93.
- Friedman, J. and Tukey, J. (1974) (*Projection pursuit*). Projection pursuit algorithm for exploratory data analysis, IEEE Trans. Computers-C, 23, 881-890.
- Galton, F. (1877) (*Regression*). Typical Laws of Heredity, Nature, 15, 492-495, 512-514, 532-533.
- Galton, F. (1888) (*Correlation*). Correlation (refer to Bulmer, M., 2003, Francis Galton: Pioneer of Heredity and Biometry, Johns Hopkins Univ Press).
- Gauss, C. (1809) (*Normal distribution*). Theoria motus corporum coelestium in sectionibus conicis solem ambientium.
- Godambe, V. and Thompson, M. (1974) (*Optimal estimating equations*). Estimating equations in the presence of a nuisance parameter, Collection of Articles Dedicated to Jerzy Neyman, Ann. Statist., 2, 568-571.
- Good, I.J. and Gaskins, R.A. (1971) (*penalized density estimation*). Nonparametric roughness penalties for probability densities, Biometrika, 58, 255-277.
- Gosset, W. (1908) (*t test*). The probable error of a mean, Biometrika, 6, 1, 1-25.
- Green, P. (1995) (*Reversible jump MCMC*). Reversible jump MCMC computation and Bayesian model determination, Biometrika, 82, 711-732.
- Grenander, U. (1956) (*Grenander estimator*). On the theory of mortality measurement, Skand. Aktuarietidskr, 39, 70-96.

- Grubbs, F. (1969) (*Grubbs' tests for outliers*). Procedures for detecting outlying observations in samples, *Technometrics*, 11, 1-21.
- Hall, P. (1985) (*Block and tile bootstrap*). Resampling a coverage pattern, *Stoch. Processes Appls.*, 20, 231-246.
- Hall, P. et al. (1991) (*Automatic bandwidth selection*). On optimal data-based bandwidth selection in kernel density estimation, *Biometrika*, 78, 263-289.
- Hampel, F. (1974) (*Influence function*). The influence curve and its role in robust estimation, *JASA*, 69, 383-393.
- Hartley, H. and Rao, J.N.K. (1962) (*Hartley-Rao method*). Sampling with unequal probabilities and without replacement, *Ann. Math. Statist.*, 33, 350-374.
- Hodges, J. (1952) (*Thresholding*). Superefficiency, Unpublished.
- Hoeffding, W. (1948) (*U-statistics*). A class of statistics with asymptotically normal distributions, *Ann. Math. Statist.*, 19, 293-325.
- Horvitz, D. and Thompson, D. (1952) (*Horvitz-Thompson estimate*). A generalization of sampling without replacement from a finite universe, *JASA*, 47, 663-685.
- Hotelling, H. (1936) (*Canonical correlation*). Relations between two sets of variates, *Biometrika*, 28, 321-377.
- Huber, P.J. (1964) (*M-estimates*). Robust estimation of a location parameter. *Ann. Math. Statist.*, 35, 73-101.
- Huber, P.J. (1973) (*Robust regression*). Robust regression: Asymptotics, conjectures, and Monte Carlo, *Ann. Statist.*, 5, 799-821.
- Hunt, G. and Stein, C. (1946) (*Hunt-Stein theorem*). Maximal invariant statistics and minimaxity (refer to Lehmann, E., *Testing Statistical Hypotheses*, 1986 Edition).
- Hurst, H. (1951) (*Long range dependence*). Long-term storage capacity of reservoirs, *Trans Amer Soc Civil Engineers*, 116, 770-808.
- Huygens, C. (1657) (*Likelihood*). Van rekeningh in spelen van geluck (On Reasoning in Games of Chance).
- Ito, K. (1944) (*Stochastic integrals*). Stochastic integral, *Proc. Imperial Acad Tokyo*, 20, 519-524.
- James, W. and Stein, C. (1961) (*James-Stein estimator*). Estimation with

- quadratic loss, Proc Fourth Berkeley Symp Math. Statist. Prob., 1, 361-379, Univ California Press.
- Jeffreys, H. (1946) (*Jeffreys prior*). An invariant form for the prior probability in estimation problems, Proc Royal Soc of London, Ser A, 186, 453-461.
- Jeffreys, H. (1961) (*Bayes factor*). The Theory of Probability, Oxford.
- Johnson, W. (1924) (*Exchangeability*). Logic: Part III, The Logical Foundations of Science (refer to Zabell, S. (1992), Predicting the unpredictable, Synthese, 90, 205-).
- Jolly, G. (1965) (*Jolly-Seber models*). Explicit estimates from capture-recapture data with both death and immigration - stochastic model, Biometrika, 52, 225-247.
- Kalman, R. (1960) (*Kalman filter*). A new approach to linear filtering and prediction problems, Jour Basic Engineering, 82, 35-45.
- Kaplan, E. and Meier, P. (1958) (*Kaplan-Meier estimate*). Nonparametric estimation from incomplete observations, JASA, 53, 457-481.
- Karlin, S. (1968) (*Total positivity*). Total Positivity, Stanford Univ Press.
- Karlin, S. and Rubin, H. (1956) (*Monotone likelihood ratio*). The theory of decision procedures for distributions with monotone likelihood ratio, Ann. Math. Statist., 27, 272-299.
- Kendall, M. (1938) (*Kendall's  $\tau$* ). A new measure of rank correlation, Biometrika, 30, 81-89.
- Kendall, M. and Stuart, A. (1946). The Advanced Theory of Statistics, 1st Ed, C. Griffin, London.
- Kiefer, J. and Wolfowitz, J. (1959) (*D-optimality*). Optimum designs in regression problems, Ann. Math. Statist., 30, 271-294.
- Koenker, R. and Bassett, G. (1978) (*Quantile regression*). Regression quantiles, Econometrica, 46, 33-50.
- Kolmogorov, A. (1933) (*Kolmogorov-Smirnov test*). Sulla determinazione empirica di una legge de distribuzione, G. Ist. Ital. Attuari, 4, 83.
- Komlos, J., Major, P. and Tusnady, G. (1975) (*KMT embedding*). An approximation of partial sums of independent random variables and the sample distribution function I, Z. Wahr. Verw. Geb., 32, 111-131.
- Koopman, B. (1936) (*Exponential family*). On distributions admitting a sufficient statistic, Trans. Amer. Math. Soc., 39, 3, 399-409.

- Krige, D. (1951) (*Kriging*). A statistical approach to some mine valuations and allied problems at the Witwatersrand, Masters Thesis at the Univ of Witwatersrand.
- Laplace, P. S. (1786) (*Survey sampling*). Estimates of the French Population.
- Le Cam, L. (1960) (*Local asymptotic normality*). Locally asymptotically normal families of distributions, Univ Calif Publ in Statist., 3, 37-98.
- Legendre, A-M. (1805) (*Least squares*). Nouvelles Methodes pour la Determination des Orbites des Cometes.
- Lehmann, E. (1953). Testing Statistical Hypotheses, 1st Ed, John Wiley.
- Lehmann, E. (1953) (*Semiparametric models*). The power of rank tests, Ann. Math. Statist., 24, 23-43.
- Lehmann, E. and Scheffe, H. (1950) (*Lehmann-Scheffe theorem*). Completeness, similar regions, and unbiased estimation, Sankhya, 10, 305-340.
- Levy, P. (1925) (*Stable laws*). Calcul des Probabilites, Gauthier-Villars, Paris.
- Levy, P. (1935) (*Martingales*). Proprietes asymptotiques des sommes de variables aleatoires enchainees, Bull. Sci. Math., 59, 84-96, 109-128.
- Li, K-C. (1991) (*Sliced inverse regression*). Sliced inverse regression for dimension reduction, JASA, 86, 316-327.
- Liang, K. and Zeger, S. (1986) (*Generalized estimating equations*). Longitudinal data analysis using generalized linear models, Biometrika, 73, 13-22.
- Lorenz, M. (1905) (*Lorenz curve*). Methods of measuring the concentration of wealth, Publ. ASA, 9, 209-219.
- Mahalanobis, P.C. (1936) (*Mahalanobis distance*). On the generalized distance in statistics, Proc. National Inst. of Sci. of India, 2, 49-55.
- Mahalanobis, P.C. (1946) (*Resampling*). Sample surveys of crop yields in India, Sankhya, 7, 29-106, 269-280 (strongly influenced by Hubback, J., 1927, according to Mahalanobis's own statements).
- Mallat, S. (1989) (*Multiresolution analysis*). A theory for multiresolution signal decomposition: The wavelet representation, Pattern Anal Machine Intel, IEEE, 11, 674-693.
- Mallows, C. (1973) (*Mallows'  $C_p$* ). Some comments on  $C_p$ , Technometrics, 15, 661-675.
- Mandelbrot, B. and van Ness, J. (1968) (*Fractional Brownian Motion*). Frac-

tional Brownian motions, fractional noises and applications, *SIAM Rev*, 10, 422-437.

Mann, H. and Whitney, D. (1947) (*Mann-Whitney test*). On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Statist.*, 18, 50-60.

Mardia, K. and Dryden, L. (1989) (*Landmark Gaussian distribution*). Shape distributions for landmark data, *Adv. Appl. Prob.*, 21, 742-755.

Markov, A. (1900) (*Gauss-Markov theorem*). *Wahrscheinlichkeitsrechnung*, Tebner, Leipzig.

Matheron, G. (1963) (*Variogram*). Principles of geostatistics, *Econ. Geology*, 58, 1246-1266.

Metropolis, N. et al. (1953) (*MCMC*). Equations of state calculations by fast computing machines, *Jour Chemical Phys*, 21, 1087-1092.

Michell, J. (1767) (*Scan statistics*). An enquiry into the probable parallax and magnitude of the fixed stars from the quantity of light which they afford us, and the particular circumstances of their situation, *Philos. Trans.*, 57, 234-264.

Nadarya, E. (1964) (*Nadarya-Watson estimator*). On estimating regression, *Theory Prob. Appls*, 9, 141-142.

Nelder, J. and Wedderburn, R. (1972) (*Generalized linear models*). Generalized linear models, *JRSSA*, 135, 370-384.

Neyman, J. (1937) (*Confidence intervals*). Outline of a theory of statistical estimation based on the classical theory of probability, *Philos Trans Royal Soc London, Ser A*, 236, 333-380.

Neyman, J. and Pearson, E. (1933) (*Neyman-Pearson tests*). On the problem of the most efficient tests of statistical hypotheses, *Philos. Trans. Royal Soc.*, Ser A, 231, 289-337.

Owen, A. (1988) (*Empirical likelihood*). Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

Page, E. (1954) (*CUSUM charts*). Continuous inspection scheme, *Biometrika*, 41, 100-115.

Page, E. (1955) (*Test for change point*). A test for a change in a parameter occurring at an unknown point, *Biometrika*, 42, 523-527.

Pearl, J. (1985) (*Bayesian networks*). Bayesian networks: A model of self-

activated memory for evidential reasoning, Tech Report, UCLA.

Pearson, K. (1891) (*Histogram*). Lectures on the geometry of statistics (refer to Stigler, S., 1986, *The History of Statistics: The Measurement of Uncertainty before 1900*, Harvard Univ Press).

Pearson, K. (1894) (*Mixture models*). On the dissection of asymmetrical frequency curves.

Pearson, K. (1895) (*Pearson family*). Contributions to the mathematical theory of evolution II, Philos Trans. Royal Soc London, 186, 343-414.

Pearson, K. (1900) (*P-values*). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philos Mag Ser. 5, 50, 157-175.

Pearson, K. (1901) (*PCA*). On lines and planes of closest fit to systems of points in space, Philos Mag., 2, 559-572.

Pearson, K. (1904) (*Meta-analysis*). Meta-Analysis, Typhoid inoculation studies, British Med J.

Peirce, B. (1852) (*Outlier detection*). Criterion for the rejection of doubtful observations, Astr. J, II, 45.

Peirce, C. (1876) (*Optimal design*). Note on the theory of economy of research, Coast Survey Report, 197-201.

Pickands, J. (1981) (*Pickands dependence function*). Multivariate extreme value distributions (with discussion), Proc 43rd Session ISI, Bull Inst. Internat. Statist., 49, 859-878.

Pitman, E. (1937) (*Permutation tests*). Significance tests which may be applied to samples from any population, RSS Suppl 4, 119-130, 225-232.

Pitman, E. (1939) (*Pitman estimates*). The estimation of the location and the scale parameters of a continuous population of any given form, Biometrika, 30, 391-421.

Pitman, E. (1948) (*Pitman efficiency*). Lecture Notes on Nonparametric Statistical Inference, Columbia Univ., NY.

Plackett, R. and Burman, J. (1946) (*Plackett-Burman designs*). The design of optimum multifactorial experiments, Biometrika, 33, 305-325.

Poisson, S. (1837) (*Poisson process*). Probabilite des juggements en matiere criminelle et en matiere civile, precedes des .. probabilites, Bachelier, Paris.

- Politis, D. and Romano, J. (1994) (*Stationary bootstrap*). The stationary bootstrap, JASA, 89, 1303-1313.
- Potthoff, R. and Roy, S.N. (1964) (*Growth curve models*). A generalized multivariate analysis of variance model useful for growth curve problems, Biometrika, 51, 313-326.
- Quenouille, M. (1949) (*Jackknife*). Approximate tests of correlation in time series, JRSSB, 11, 68-84.
- Raiffa, H. and Schlaiffer, R. (1961) (*Conjugate priors*). Applied Statistical Decision Theory, Harvard Business School Publ.
- Rao, C.R. (1945) (*Cramér- Rao inequality, Rao-Blackwell theorem*). Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc., 37, 81-89.
- Rao, C.R. (1947) (*Orthogonal arrays*). Factorial experiments derivable from combinatorial arrangements of arrays, JRSS Suppl., 9, 128-139.
- Rao, C.R. (1948) (*Score test*). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, Proc Cambridge Philos Soc, 44, 50-57.
- Rao, C.R. (1958) (*Functional data analysis*). Some statistical methods for the comparison of growth curves, Biometrics, 14, 1-17.
- Rao, C.R. (1965) (*Random coefficient regressions*). The theory of least squares when the parameters are stochastic and its applications, Biometrika, 52, 447-458.
- Rao, C.R. (1966) (*Regression with  $n < p$* ). Generalized inverse for matrices and its applications in mathematical statistics, In Research Papers in Statistics, Festschrift for J. Neyman, F.N. David Ed., John Wiley (Rao (1962), JRSSB, is a shorter paper on regression with  $n < p$ ).
- Robbins, H. (1956) (*Empirical Bayes*). An empirical Bayes approach to Statistics, Proc Third Berkeley Symp Math. Statist. Prob., Univ California Press.
- Rosenbaum, P. and Rubin, D. (1983) (*Propensity scores*). The central role of the propensity score in observational studies for causal effects, Biometrika, 70, 41-55.
- Rosenblatt, M. (1956) (*Kernel density estimates*). Remarks on some non-parametric estimates of a density function, Ann. Math. Statist., 27, 832-837.



- Roy, S.N. (1957) (*Largest root test*). Some Aspects of Multivariate Analysis, John Wiley.
- Rubin, D. (1976) (*Imputation*). Inference and missing data, *Biometrika*, 63, 581-592.
- Schapiro, R. (1990) (*Boosting*). The strength of weak learnability, *Machine Learning*, 5, 197-227.
- Scheffe, H. (1953) (*Scheffe confidence intervals*). A method for judging all contrasts in the analysis of variance, *Biometrika*, 40, 87-104.
- Schuster, A. (1898) (*Periodogram*). On the investigation of hidden periodicities with applications to supposed 26 day period of meteorological phenomena, *Terres Magnetism*, 3, 13-41.
- Schwarz, G. (1978) (*BIC*). Estimating the dimension of a model, *Ann. Statist.*, 6, 461-464.
- Shapiro, S. and Wilk, M.B. (1965) (*Shapiro-Wilk test*). An analysis of variance test for normality (complete samples), *Biometrika*, 52, 591-611.
- Singh, K. (1981) (*Bootstrap consistency*). On the asymptotic accuracy of Efron's bootstrap, *Ann. Statist.*, 9, 1187-1195.
- Shewart, W. (1931) (*Control charts*). Economic Control of Quality of Manufactured Product, van Nordstrom.
- Sklar, A. (1959) (*Copulas*). Fontions de repartition á  $n$  dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, 8, 229-231.
- Slutsky, E. (1925) (*Slutsky's theorem*). Uber stochastische Asymptoten und Grenzwerte, *Metron*, 5, 3-89.
- Snedecor, G. (1937) (*Fisher-Snedecor F test*). *Statistical Methods*, 1st Ed, Iowa State Univ Press.
- Spearman, C. (1904) (*Factor analysis*). General Intelligence, Objectively Determined and Measured, *Amer J of Psychology*, 15, 2, 201-292.
- Stefanski, L. and Carroll, R. (1990) (*Deconvolution density estimators*). Deconvoluting kernel density estimators, *Statistics*, 21, 169-184.
- Stein, C. (1945) (*Two stage estimation*). A two sample test for a linear hypothesis whose power is independent of the variance, *Ann. Math. Statist.*, 43, 243-258.
- Stein, C. (1956) (*Shrinkage*). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proc Third Berkeley Symp Math.*

- Statist. Prob., Univ California Press.
- Stein, C. (1956) (*Adaptive estimation*). Efficient nonparametric testing and estimation, Proc Third Berkeley Symp Math. Statist. Prob., Univ California Press.
- Stein, C. (1972) (*Stein's method*). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, Proc Sixth Berkeley Symp Math. Statist. Prob., Univ California Press.
- Stein, C. (1973) (*Stein's identity*). Estimation of the mean of a multivariate normal distribution, Proc Prague Symp Asymp Statist., 345-381.
- Stein, C. (1975) (*Covariance regularization*). The Rietz Lecture, IMS.
- Steinhaus, H. (1957) (*k-means clustering*). Sur la division des corps materiels en parties, Bull. Acad. Polon Sci, 4, 801-801.
- Stone, C. (1977) (*Consistent nonparametric regression*). Consistent nonparametric regression, Ann. Statist., 5, 595-620.
- Stone, M. (1977) (*CV*). Asymptotics for and against cross-validation, Biometrika, 64, 29-35.
- Stratonovich, R. (1960) (*Hidden Markov models*). Conditional Markov processes, Theory Prob. Appls., 5, 156-178.
- Thiele, T. (1889) (*Edgeworth expansions*). Semi-invariants (refer to Cramér, H., Mathematical Methods of Statistics, 1946, Princeton Univ Press).
- Tibshirani, R. (1996) (*Lasso*). Regression shrinkage and selection via the lasso, JRSSB, 58, 267-288.
- Tikhonov, A. (1963) (*Ridge regression*). Solution of incorrectly formulated problems and the regularization method, Doklady Akademie Nauk SSSR, 151, 501-504.
- Tukey, J. (1949) (*Test for additivity*). One degree of freedom for non-additivity, Biometrics, 5, 232-242.
- Tukey, J. (1977) (*EDA*). Exploratory Data Analysis, Addison-Wesley.
- Ulam, S. and von Neumann (1946) (*Monte Carlo Sampling*). Refer to *Monte Carlo Methods*, Wikipedia.
- Vapnik, V. and Chervonenkis, A. (1971) (*VC classes*). On the uniform convergence of relative frequencies of events to their probabilities, Theory Prob. Appls, 16, 264-280.
- Wahba, G. (1990) (*Smoothing splines*). Spline Models for Observational

- Data, SIAM.
- Wald, A. (1939) (*Admissibility and minimaxity*). Contributions to the theory of statistical estimation and testing hypotheses, Ann. Math. Statist., 10, 299-326.
- Wald, A. (1940) (*Errors in variables*). The fitting of variables if both variables are subject to error, Ann. Math. Statist., 11, 3, 284-300.
- Wald, A. (1943) (*Wald test*). Tests of statistical hypotheses concerning several parameters when the number of observations is large, Trans. Amer. Math. Soc., 5, 426-482.
- Wald, A. (1945) (*SPRT*). Sequential tests of statistical hypotheses, Ann. Math. Statist., 16, 117-186.
- Watson, G. (1964) (*Nonparametric regression*). Smooth regression analysis, Sankhya, Ser A, 26, 359-372.
- Wermuth, N. and Lauritzen, S. (1983) (*Graphical models*). Graphical and recursive models for contingency tables, Biometrika, 70, 537-552.
- Whittle, P. (1951) (*ARMA models*). Hypothesis testing in time series analysis, Almqvist and Wicksell.
- Whittle, P. (1953) (*Whittle estimator*). Estimation and information in stationary time series, Ark Mat, 2, 423-434.
- Wilcoxon, F. (1945) (*Wilcoxon signed-rank test*). Individual comparisons by ranking methods, Biometrics Bulletins, 1, 80-83.
- Wilk, M. and Gnanadesikan, R. (1968) (*q-q plots*). Probability plotting methods for the analysis of data, Biometrika, 55, 1-17.
- Wilks, S. (1938) (*Asymptotic distribution of LRT*). The large sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Math. Statist., 9, 60-62.
- Wishart, J. (1928) (*Wishart distribution*). The generalised product moment distribution in samples from a normal multivariate population, Biometrika, 20A, 1-2, 32-52.
- Wright, S. (1921) (*Structural equation models*). Correlation and Causation, Jour Agr Res., 20, 557-585.
- Wu, C.J. (1993) (*Supersaturated designs*). Construction of supersaturated designs through partially aliased interactions, Biometrika, 80, 661-669.
- Yuan, M. and Lin, Y. (2006) (*Group lasso*). Model selection and estimation

in regression with grouped variables, JRSSB, 68, 49-67.

Zeller, M. (1968) (*Longitudinal data analysis*). Stimulus control with fixed-ratio reinforcement, Jour Experimental Anal. Behavior, 11, 107-115.

Zou, H., Hastie, T., and Tibshirani, R. (2006) (*Sparse PCA*). Sparse principal components analysis, JCGS, 15, 262-286.