Decentralized predictive symbiotic corporate compliance mining

Emelyne Umunoza Gasana[1,2], Dietrich von Rosen[3] and Martin Singull[2]

[1] *Department of Mathematics, Linköping University, Sweden*
[2] *Department of Mathematics, University of Rwanda, Rwanda*
[3] *Department of Energy and Technology, Swedish University of Agricultural Sciences, Sweden*

Abstract

The exact distribution of a classification function is often complicated to allow for easy numerical calculations of misclassification errors. The use of expansions is one way of dealing with this difficulty. In this paper, approximate probabilities of misclassification of the maximum likelihood based discriminant function are established via an Edgeworth-type expansion based on the standard normal distribution for discriminating between two multivariate normal populations.

Keywords

Classification rule, Discriminant analysis, Edgeworth-type expansion, Misclassification errors.

References:

Anderson, T. W. (1951). Classification by multivariate analysis. *Psychometrika*, 16:31–50.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. John Wiley & Sons, Inc., Hoboken.

Barnard, M. M. (1935). The secular variations of skull characters in four series of Egyptian skulls. *Annals of Eugenics*, 6:352–371.

Cornish, E. A. and Fisher, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique*, pages 307–320

Davis, A. (1976). Statistical distributions in univariate and multivariate edgeworth populations. *Biometrika*, 63:661–670.

Edgeworth, F. (1905). The law of error. *Transactions of the Cambridge Philosophical Society*, 20:36–66; 113–141.

Edgeworth, F. (1907). On the representation of statistical frequency by a series. *Journal of the Royal Statistical Society*, 70(1):102–106.

Fisher, R. A. (1936). The use of multiple measurements in 10 taxonomic problems. *Annals of Eugenics*, 7:179–188.

Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*, 8:376–386.

Fujikoshi, Y., Ulyanov, V. V., and Shimizu, R. (2011). *Multivariate Statistics: High-dimensional and Large-Sample Approximations*, volume 760. John Wiley & Sons, Inc., Hoboken.

Gram, J. P. (1879). *Om Rækkeudviklinger, bestemte ved Hjælp af de mindste Kvadraters Methode.* PhD thesis, Andr. Fred. Høst & Son, København.

Gupta, S. S. and Panchapakesan, S. (1982). Edgeworth expansions in statistics: A brief review. *Technical report, Purdue university Lafayette in department of statistics.*

Hald, A. (2000). The early history of the cumulants and the gramcharlier series. *International Statistical Review*, 68(2):137–153.

Hald, A. and Steffensen, J. (2002). *On the History of Series Expansions of Frequency Functions and Sampling Distributions*, 1873-1944. Det Kongelige Danske Videnskabernes Selskab, Copenhagen.

Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, volume 579. Springer Science & Business Media, Dordrecht.

Kudo, A. (1959). The classificatory problem viewed as a two-decision problem. *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics*, 13:96–125.

Kudo, A. (1960). The classifactory problem viewed as a a two-decision problem-II. *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics*, 14:63–83.

Laplace, P.-S. (1811). Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations. *Mémoires de l'Académie Royale des Sciences de Paris*, pages 279–347.

Lauritzen, S. L., Thiele, T. N., Thiele, J., and Hald, A. (2002). *Thiele: Pioneer in Statistics.* Clarendon Press.

Mahalanobis, P. C. (1925). Analysis of race-mixture in Bengal. *Journal and Proceedings of Asiatic Society of Bengal New series*, 23:301–333.

Memon, A. Z. and Okamoto, M. (1971). Asymptotic expansion of the distribution of the Z statistic in discriminant analysis. *Journal of Multivariate Analysis*, 1(3):294–307.

Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *The Annals of Mathematical Statistics*, 34:1286–1301.

Pearson, K. (1915). On the problem of sexing osteometric material. *Biometrika*, 10(4):479–487.

Ranga Rao, R. (1960). *Some Problems in Probability Theory*. PhD thesis, Calcutta University.

Siotani, M. (1977). *Asymptotic Expansions for Error Rates and Comparison of the W-procedure and the Z-procedure in Discriminant Analysis*. North-Holland, Amsterdam.

Siotani, M. and Wang, R.-H. (1975). *Further Expansion Formulae for Error Rates and Comparison of the W-and Z-procedures in Discriminant Analysis*. Technical Report No. 33, Department of Statistics, Kansas State University, Manhattan, Kansas.

Sitgreaves, R. (1952). On the distribution of two random matrices used in classification procedures. *The Annals of Mathematical Statistics*, 23:263–270.

Sitgreaves, R. (1961). Some results on the distribution of the W-classification statistic. *Chapter 19 in Solomon, H. (ed.), Studies in Item Analysis and Prediction*, pages 241–261.

Srivastava, M. S. and Khatri, C. (1979). *An Introduction to Multivariate Statistics*. Elsevier North-Holland, New York.

Thiele, T. N. (1873). Om en tilnærmelsesformel. *Tidsskrift for mathematik*, 3:22–31.

Thiele, T. N. (1889). *Forlaesinger over Almindelig Iagttagelseslaere: Sandsynlighedsregning og Mindste Kvadraters Methode*. København: Reitzel.

Wald, A. (1944). On a statistical problem arising in the classification of an individual into one of two groups. *The Annals of Mathematical Statistics*, 15:145–162.