Triangular systems for symmetric binary variables

Nanny Wermuth, *Department of Mathematical Sciences, Chalmers/Gothenburg University, Sweden*
Giovanni M. Marchetti, *Department of Statistics, University of Florence, Italy*
D.R. Cox, *Nuffield College, Oxford, UK*

Abstract

We introduce and study distributions of sets of binary variables that are symmetric, that is each has equally probable levels. The joint distribution of these special types of binary variables, if generated by a recursive process of linear main effects is essentially parametrized in terms of marginal correlations. This contrasts with the log-linear formulation of joint probabilities in which parameters measure conditional associations given all remaining variables. The new formulation permits useful comparisons of different types of graphical Markov models and leads to a close approximation of Gaussian orthant probabilities.

AMS 2000 subject classifications: Primary 62E10; secondary 62H17, 62H20.

Keywords: Graphical Markov models, linear in probability models, log-linear models, recursive generating processes.

Full Text: PDF

Wermuth, Nanny, Marchetti, Giovanni M., Cox, D.R., Triangular systems for symmetric binary variables, Electronic Journal of Statistics, 3, (2009), 932-955 (electronic). DOI: 10.1214/09-EJS439.

References

Bahadur, R.R. (1961). A representation of the distribution of joint responses to n dichotomous items. In: Studies in item analysis and prediction. H. Solomon (ed.) 158–176. Stanford University Press. MR0121893

Bartlett, M.S. (1935). Contingency table interactions. Suppl. J. Roy. Statist. Soc., 2, 248–252.

Bergsma, W., Croon, M. & Hagenaars, J.A. (2009) Marginal Models. Springer, New York.

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. J. Roy. Statist. Soc. B 25, 220–233. MR0168065

Cheng, M.C. (1969). The orthant probabilities of four Gaussian variates. Ann. Math. Statist. 40, 152–161. MR0235596

Cochran, W.G. (1938). The omission or addition of an independent variate in multiple linear regression. Suppl. J. Roy. Statist. Soc. 5, 171–176.

Cox, D.R. (2006). Principles of Statistical Inference. Cambridge University Press.

Cox, D.R. (2007). On a generalization of a result of W.G. Cochran. Biometrika, 94, 400–410. MR2410022

Cox, D.R. & Wermuth, N. (1994). A note on the quadratic exponential binary distribution. Biometrika, 81, 403–408. MR1294901

Cox, D.R. & Wermuth, N. (1996). Multivariate Dependencies – Models, Analysis and Interpretation. London: Chapman & Hall. MR1456990

Cox, D.R. & Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. J. Roy. Statist. Soc. B, 65, 937–941. MR2017879

Cramèr, H. (1946). Mathematical methods of statistics. Princeton University Press. MR1816288

Darroch, J.N. & Ratcliff, D. (1972) Generalized iterative scaling for log-linear models. Ann. Math. Statist. , 43, 1470–1480. MR0345337

Darroch, J.N., Lauritzen, S.L. & Speed, T.P. (1980). Markov fields and log-linear models for contingency tables. Ann. Statist. 8, 522–539. MR0568718

Dempster, A. P. (1972). Covariance selection. Biometrics 28, 157–175.

Drton, M. (2009). Discrete chain graph models. Bernoulli. to appear.

Fienberg, S. (2007). The analysis of cross-classified categorical data. Springer (2007), New York.

Frydenberg, M. (1990). The chain graph Markov property. Scan. J. Statist. 17, 333–353. MRMR1096723

Glonek, G.J.N. & McCullagh, P. (1995). Multivariate logistic models. J. Roy. Statist. Soc. B, 57, 533–546.

Goldberger, A.S. (1964). Econometric Theory. Wiley, New York. MR0158754

Good, I. J. (1958). The interaction algorithm and practical Fourier analysis. J. Roy. Statist. Soc. B 20, 361–372. MR0126937

Goodman, L.A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. Biometrika, 60, 179–192. MR0415839

Green, P.J., Hjort, N. & Richardson, S. (eds.) (2003). Models for highly structured stochastic systems. Oxford University Press.

Holland, P.W. & Rosenbaum, P. (1986). Conditional association and unidimensionality in monotone latent variable models. Ann. Statist., 14, 1523–1543. MR0868316

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1995). Distributions in statistics. Continuous univariate distributions, Vol. 2, 2nd edn., Wiley, New York.

Lancaster, H.O. (1969). The Chi-Squared Distribution. Wiley, New York. MR0253452

Lazarsfeld, P.F. (1961). The algebra of dichotomous systems. In: Studies in item analysis and prediction. H. Solomon (ed.) 111-157. Stanford University Press. MR0121892

Ma, Z.M., Xie, X.C. & Geng, Z. (2006). Collapsibility of distribution dependence. J. Roy. Statist. Soc. B, 68, 127–133. MR2212578

Marchetti G.M. & Lupparelli, M. (2009). Chain graph models of multivariate regression type for categorical data. Submitted.

Marchetti, G.M. & Wermuth, N. (2009). Matrix representations and independencies in directed acyclic graphs. Ann. Statist., 37, 961–978. MR2502657

McFadden, J.A. (1955). Urn models of correlation and a comparison with the multivariate normal integral. Ann. Math. Statist. 26, 478–489. MR0070863

McFadden, J.A. (1956). An approximation for the symmetric quadrivariate normal integral. Biometrika 43, 206–207. MR0077051

Moran, P.A.P. (1956). The numerical evaluation of a class of integrals. Proc. Cambridge Philos. Soc., Statist. 52, 230–233. MR0076450

Pearl, J. & Wermuth, N. (1994). When can association graphs admit a causal interpretation? In: Models and Data, Artificial Intelligence and Statistics IV. 205–214. P. Cheeseman and W. Oldford (eds.). New York: Springer.

Schläfli, L. (1858). On the multiple integral. Quart. J. Math., 2, 269–301, 3 54-68, 97–108.

Sheppard, W.F. (1898). On the geometrical treatment of the 'normal curve' of statistics, with special reference to correlation and to the theory of error. Proc. Roy. Soc. London, 62, 171–173.

Streitberg, B. (1990). Lancaster interactions revisited. Ann. Statist., 18, 1878–1885. MR1074442

Streitberg, B. (1999). Exploring interactions in high-dimensional tables: a bootstrap alternative to log-linear models. Ann. Statist., 27, 405–413. MR1701118

Wermuth, N. (1976). Analogies between multiplicative models for contingency tables and covariance selection. Biometrics, 32, 95–108. MR0403088

Wermuth, N. (1980) Linear recursive equations, covariance selection, and path analysis. J. Amer. Statist. Assoc. 75 963–972. MR0600984

Wermuth, N. (1998). Pairwise independence. Encyclopedia of Biostatistics. P. Armitage and T. Colton (eds). New York: Wiley, 3244-3246.

Wermuth, N. & Cox, D.R. (1998). On association models defined over independence graphs. Bernoulli, 4, 477–495. MR1679794

Wermuth, N. & Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. J. Roy. Statist. Soc. B, 66, 687–717. MR2088296

Wermuth, N., Cox, D.R. & Marchetti, G.M. (2006). Covariance chains. Bernoulli, 12, 841–862. MR2265345

Wermuth, N., Wiedenbeck, M. & Cox, D.R. (2006). Partial inversion for linear systems and partial closure of independence graphs. BIT, Numerical Mathematics, 46, 883–901. MR2285213

Wright, S. (1934). The method of path coefficients. Ann. Math. Statist., 5, 161–215.

Xie, X.C., Ma, Z.M. & Geng, Z. (2008). Some association measures and their collapsibility. Stat. Sinica. 1165–1183. MR2440406

Yates, F. (1937). The Design and Analysis of Factorial Experiments. Harpenden: Imperial Bureau of Soil Science.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J. Amer. Statist. Assoc. 57, 348–368. MR0139235