Fan, Jianqing; **Peng, Heng**

Nonconcave penalized likelihood with a diverging number of parameters. (English)

Zbl 1092.62031
Ann. Stat. 32, No. 3, 928-961 (2004).

Summary: A class of variable selection procedures for parametric models via nonconcave penalized likelihood was proposed by *J. Fan* and *R. Li* [ibid. 30, 74–99 (2002; Zbl 1012.62106); J. Am. Stat. Assoc. 96, No. 456, 1348–1360 (2001; Zbl 1073.62547)] to simultaneously estimate parameters and select important variables. They demonstrated that this class of procedures has an oracle property when the number of parameters is finite. However, in most model selection problems the number of parameters should be large and grow with the sample size. In this paper some asymptotic properties of the nonconcave penalized likelihood are established for situations in which the number of parameters tends to ∞ as the sample size increases.

Under regularity conditions we have established an oracle property and the asymptotic normality of the penalized likelihood estimators. Furthermore, the consistency of the sandwich formula of the covariance matrix is demonstrated. Nonconcave penalized likelihood ratio statistics are discussed, and their asymptotic distributions under the null hypothesis are obtained by imposing some mild conditions on the penalty functions. The asymptotic results are augmented by a simulation study, and the newly developed methodology is illustrated by an analysis of a court case on the sexual discrimination of salary.

MSC:

62F12 Asymptotic properties of parametric estimators
62E20 Asymptotic distribution theory in statistics
62F03 Parametric hypothesis testing

Keywords:

model selection; nonconcave penalized likelihood; diverging parameters; oracle property; asymptotic normality; standard errors; likelihood ratio statistic

Software:

Excel

Full Text: DOI arXiv

References:

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Proc. 2nd International Symposium on Information Theory (V. Petrov and F. Csáki, eds.) 267–281. Akadémiai Kiadó, Budapest. · Zbl 0283.62006

[2] Albright, S. C., Winston, W. L. and Zappe, C. J. (1999). Data Analysis and Decision Making with Microsoft Excel . Duxbury, Pacific Grove, CA.

[3] Antoniadis, A. (1997). Wavelets in statistics: A review (with discussion). J. Italian Statist. Soc. 6 97–144.

[4] Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations (with discussion). J. Amer. Statist. Assoc. 96 939–967. · Zbl 1072.62561 · doi:10.1198/016214501753208942

[5] Bai, Z. D., Rao, C. R. and Wu, Y. (1999). Model selection with data-oriented penalty. J. Statist. Plann. Inference 77 103–117. · Zbl 0926.62045 · doi:10.1016/S0378-3758(98)00168-2

[6] Blake, A. (1989). Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. IEEE Trans. Pattern Anal. Machine Intell. 11 2–12. · Zbl 0682.90087 · doi:10.1109/34.23109

[7] Blake, A. and Zisserman, A. (1987). Visual Reconstruction . MIT Press, Cambridge, MA.

[8] Breiman, L. (1996). Heuristics of instability and stabilization in model selection. Ann. Statist. 24 2350–2383. · Zbl 0867.62055 · doi:10.1214/aos/1032181158 ·

[9] Cox, D. D. and O'Sullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. Ann. Statist. 18 1676–1695. JSTOR: · Zbl 0719.62051 · doi:10.1214/aos/1176347872 · links.jstor.org

[10] De Boor, C. (1978). A Practical Guide to Splines . Springer, New York. · Zbl 0406.41003

[11] Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. Aide-Memoire of a Lecture at AMS Conference on Math Challenges of the 21st Century.

[12] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika 81 425–455. · Zbl 0815.62019 · doi:10.1093/biomet/81.3.425

[13] Fan, J. (1997). Comments on "Wavelets in statistics: A review," by A. Antoniadis. J. Italian Statist. Soc. 6 131–138.

[14] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96 1348–1360. · Zbl 1073.62547 · doi:10.1198/016214501753382273

[15] Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. Ann. Statist. 30 74–99. · Zbl 1012.62106 · doi:10.1214/aos/1015362185 ·

[16] Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometric regression tools (with discussion). Technometrics 35 109–148. · Zbl 0775.62288 · doi:10.2307/1269656

[17] Fu, W. J. (1998). Penalized regression: The bridge versus the LASSO. J. Comput. Graph. Statist. 7 397–416.

[18] Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., eds. (1996). Markov Chain Monte Carlo in Practice . Chapman and Hall/CRC Press, London. · Zbl 0832.00018

[19] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell. 6 721–741. · Zbl 0573.62030 · doi:10.1109/TPAMI.1984.4767596

[20] Green, P. J. and Silverman, B. W. (1994). Nonparametric Regression and Generalized Linear Models : A Roughness Penalty Approach . Chapman and Hall, London. · Zbl 0832.62032

[21] Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. Ann. Statist. 1 799–821. JSTOR: · Zbl 0289.62033 · doi:10.1214/aos/1176342503 · links.jstor.org

[22] Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. J. Amer. Statist. Assoc. 96 1387–1396. · Zbl 1073.62539 · doi:10.1198/016214501753382309

[23] Knight, K. and Fu, W. J. (2000). Asymptotics for Lasso-type estimators. Ann. Statist. 28 1356–1378. · Zbl 1105.62357 · doi:10.1214/aos/1015957397 ·

[24] Lehmann, E. L. (1983). Theory of Point Estimation . Wiley, New York. · Zbl 0522.62020

[25] Mallows, C. L. (1973). Some comments on \(C_p\). Technometrics 12 661–675. · Zbl 0269.62061 · doi:10.2307/1267380

[26] McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models , 2nd ed. Chapman and Hall, London. · Zbl 0588.62104

[27] Murphy, S. A. (1993). Testing for a time dependent coefficient in Cox's regression model. Scand. J. Statist. 20 35–50. · Zbl 0782.62053

[28] Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica 16 1–32. · Zbl 0034.07602 · doi:10.2307/1914288

[29] Nikolova, M., Idier, J. and Mohammad-Djafari, A. (1998). Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF. IEEE Trans. Image Process. 7 571–585. · Zbl 0973.94008 · doi:10.1109/83.663502

[30] Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. Ann. Statist. 16 356–366. JSTOR: · Zbl 0637.62026 · doi:10.1214/aos/1176350710 · links.jstor.org

[31] Schwarz, G. (1978). Estimating the dimension of a model. Ann. Statist. 6 461–464. JSTOR: · Zbl 0379.62005 · doi:10.1214/aos/1176344136 · links.jstor.org

[32] Shen, X. T. and Ye, J. M. (2002). Adaptive model selection. J. Amer. Statist. Assoc. 97 210–221. · Zbl 1073.62509 · doi:10.1198/016214502753479356

[33] Stone, C. J., Hansen, M., Kooperberg, C. and Truong, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion). Ann. Statist. 25 1371–1470. · Zbl 0924.62036 · doi:10.1214/aos/1031594728 ·

[34] Tibshirani, R. J. (1996). Regression shrinkage and selection via the Lasso. J. Roy. Statist. Soc. Ser. B 58 267–288. · Zbl 0850.62538

[35] Tibshirani, R. J. (1997). The Lasso method for variable selection in the Cox model. Statist. Medicine 16 385–395.

[36] van der Vaart, A. W. (1998). Asymptotic Statistics. Cambridge Univ. Press. · Zbl 0910.62001 · doi:10.1017/CBO9780511802256

[37] Wahba, G. (1990). Spline Models for Observational Data. SIAM, Philadelphia. · Zbl 0813.62001