Böhringer, Stefan; Lohmann, Dietmar

**Exact model comparisons in the plausibility framework.** (English) [Zbl 1480.62143]
J. Stat. Plann. Inference 217, 224-240 (2022).

Summary: Plausibility is a formalization of exact tests for parametric models and generalizes procedures such as Fisher’s exact test. The resulting tests are based on cumulative probabilities of the probability density function and evaluate consistency with a parametric family while providing exact control of the \( \alpha \) level for finite sample size. Model comparisons are inefficient in this approach. We generalize plausibility by incorporating weighing which allows to perform model comparisons. We show that one weighing scheme is asymptotically equivalent to the likelihood ratio test (LRT) and has finite sample guarantees for the test size under the null hypothesis unlike the LRT. We confirm theoretical properties in simulations that mimic the data set of our data application. We apply the method to a retinoblastoma data set and demonstrate a parent-of-origin effect. Weighted plausibility also has applications in high-dimensional data analysis and \( P \)-values for penalized regression models can be derived. We demonstrate superior performance as compared to a data-splitting procedure in a simulation study. We apply weighted plausibility to a high-dimensional gene expression, case-control prostate cancer data set. We discuss the flexibility of the approach by relating weighted plausibility to targeted learning, the bootstrap, and sparsity selection.

**MSC:**
- 62J05  Linear regression; mixed models
- 62F03  Parametric hypothesis testing
- 62P10  Applications of statistics to biology and medical sciences; meta analysis

**Keywords:**
- plausibility; exact testing; likelihood; high-dimensional data; model comparison; retinoblastoma

**Full Text:** DOI arXiv

**References:**

[1] Agresti, Alan, A survey of exact inference for contingency tables, Statist. Sci., 7, 1, 131-153 (1992) · Zbl 0955.62587
[2] Agresti, Alan; Coull, Brent A., Approximate is better than “exact” for interval estimation of binomial proportions, Amer. Statist., 52, 2, 119-126 (1998)
[3] Ahdesmaki, Miika; Zuber, Verena; Gibb, Sebastian; Strimmer, Korbinian, Sda: Shrinkage discriminant analysis and CAT score variable selection (2015)
[4] Bremner, Rod; Sage, Julien, Cancer: The origin of human retinoblastoma, Nature, 514, 7522, 312-313 (2014)
[5] Clopper, Charles J.; Pearson, Egon S., The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika, 26, 4, 404-413 (1934) · Zbl 60.1175.02
[6] Fisher, R. A., On the interpretation of \( \chi^2 \) from contingency tables, and the calculation of \( P \), J. R. Stat. Soc., 85, 1, 87-94 (1922)
[7] Friedman, Jerome; Hastie, Trevor; Tibshirani, Robert, The Elements of Statistical Learning, Vol. 1 (2001), Springer series in statistics New York · Zbl 0973.62007
[8] Friedman, Jerome; Hastie, Trevor; Tibshirani, Rob, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw., 33, 1, 1-22 (2010)
[9] Goeman, Jelle J.; Mansmann, Ulrich, Multiple testing on the directed acyclic graph of gene ontology, Bioinformatics, 24, 4, 537-544 (2008)
[10] Hoerl, Arthur E.; Kennard, Robert W., Ridge regression: Biased estimation for nonorthogonal problems, Technometrics, 12, 1, 55-67 (1970) · Zbl 0202.17205
[11] Kundson, Alfred G., Mutation and cancer: Statistical study of retinoblastoma, Proc. Natl. Acad. Sci., 68, 4, 820-823 (1971)
[12] Martin, Ryan, Plausibility functions and exact frequentist inference, J. Amer. Statist. Assoc., 110, 512, 1552-1561 (2015) · Zbl 1373.62027
[13] Meinshausen, Nicolai, Hierarchical testing of variable importance, Biometrika, 95, 2, 265-278 (2008) · Zbl 1437.62557
[14] Meinshausen, Nicolai; Meier, Lukas; Bühlmann, Peter, P-values for high-dimensional regression, J. Amer. Statist. Assoc., 104, 488, 1671-1681 (2009) · Zbl 1205.62089
[15] Singh, Dinesh; Febbo, Phillip G.; Ross, Kenneth; Jackson, Donald G.; Manola, Judith; Ladd, Christine; Tamayo, Pablo; Renshaw, Andrew A.; D’Amico, Anthony V.; Richie, Jerome P., Gene expression correlates of clinical prostate cancer behavior, Cancer Cell, 1, 2, 203-209 (2002)

[16] Tibshirani, R., Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol., 58, 1, 267-288 (1996) · Zbl 0850.62538

[17] Van der Laan, Mark J.; Rose, Sherri, Targeted Learning: Causal Inference for Observational and Experimental Data (2011), Springer Science & Business Media · Zbl 1218.62121

[18] Wasserman, Larry; Roeder, Kathryn, High dimensional variable selection, Ann. Statist., 37, 5A, 2178 (2009) · Zbl 1173.62054

[19] Yuan, Yinyin, Lol: Lots of Lasso (2019)

[20] Zheng, Wenjing, Laan, Mark van der, 2010. Asymptotic theory for cross-validated targeted maximum likelihood estimation, U.C. Berkeley Division of Biostatistics Working Paper Series.

[21] Zou, Hui; Hastie, Trevor, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser. B Stat. Methodol., 301-320 (2017) · Zbl 1069.62054

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.