Better predictions when models are wrong or underspecified

Ommen, M. van

Author: Ommen, Thijs van
Title: Better predictions when models are wrong or underspecified
Issue Date: 2015-06-10

Bibliography

R. K. Ahuja, J. B. Orlin, C. Stein, and R. E. Tarjan. Improved algorithms for bipartite network flow. *SIAM Journal on Computing*, 23(5):906–933, 1994.

H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csáki, editors, *2nd International Symposium on Information Theory*, pages 267–281, 1973.

H. Akaike. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, 66(2):237–242, 1979.

T. Ando. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, 94(2):443–458, 2007.

J. Y. Audibert. *PAC-Bayesian statistical learning theory*. PhD thesis, Université Paris VI, 2004.

J. Y. Audibert. Progressive mixture rules are deviation suboptimal. In *NIPS*, 2007.

R. E. Barlow, D. J. Bartholomew, J. M. Bremmer, and H. D. Brunk. *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley, New York, 1972.

A. R. Barron. Are Bayes rules consistent in information? In *Open Problems in Communication and Computation*, pages 85–91. Springer, 1987.

A. R. Barron. Information-theoretic characterization of Bayes performance and the choice of priors in parametric and nonparametric problems. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 6, pages 27–52. Oxford University Press, Oxford, 1998.

A. R. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Transactions on Information Theory*, 37(4):1034–1054, 1991.

A. R. Barron and N. Hengartner. Information theory and superefficiency. *Annals of Statistics*, 26(5):1800–1825, 1998.

A. R. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44(6):2743–2760, 1998. Special Commemorative Issue: Information Theory: 1948-1998.

A. R. Barron, M. J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, 27(2):536–561, 1999.

S. Ben-David, T. Lu, T. Luu, and D. Pál. Impossibility theorems for domain adaptation. In *Artificial Intelligence and Statistics (AISTATS)*, pages 129–136, 2010.

R. Berk. Limiting behavior of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics*, 37:51–58, 1966.

J. M. Bernardo. Expected information as expected utility. *The Annals of Statistics*, 7(3):686–690, 1979.

J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, Chichester, 1994.

J. Besag. Statistical analysis of non-lattice data. *The statistician*, pages 179–195, 1975.

P. Bissiri, C. Holmes, and S. Walker. A general framework for updating belief distributions. *arXiv preprint arXiv:1306.6430*, 2013.

O. Bochet, R. İlkılıç, H. Moulin, and J. Sethuraman. Balancing supply and demand under bilateral constraints. *Theoretical Economics*, 7:395–423, 2012.

O. Bochet, R. İlkılıç, and H. Moulin. Egalitarianism under earmark constraints. *Journal of Economic Theory*, 148:535–562, 2013.

G. E. P. Box. Robustness in the strategy of scientific model building. In R. L. Launer and G. N. Wilkinson, editors, *Robustness in Statistics*, New York, 1979. Academic Press.

G. E. P. Box. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, pages 383–430, 1980.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

L. Breiman. Statistical modeling: The two cultures (with discussion). *Statistical Science*, 16(3):199 –215, 2001.

K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: A practical information-theoretic approach*. Springer, New York, second edition, 2002.

O. Catoni. A mixture approach to universal model selection. Preprint LMENS-97-30, 1997. Available from `http://www.math.ens.fr/edition/publis/Index.97.html`.

O. Catoni. *PAC-Bayesian Supervised Classification*. Lecture Notes-Monograph Series. IMS, 2007.

O. Catoni. Discussion on the paper 'Catching up Faster by Switching Sooner' by Van Erven, Grünwald and De Rooij. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):399–400, 2012.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, Cambridge, UK, 2006.

G. Claeskens and N. L. Hjort. The focused information criterion. *Journal of the American Statistical Association*, 98:900–916, 2003. With discussion.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.

I. Csiszár and P. Shields. The consistency of the BIC Markov order estimator. *Annals of Statistics*, 28:1601–1619, 2000.

N. V. Cuong, W. S. Lee, N. Ye, K. M. A. Chai, and H. L. Chieu. Active learning for probabilistic hypotheses using the maximum Gibbs error criterion. In *Advances in Neural Information Processing Systems 26*, 2013.

A. S. Dalalyan and A. B. Tsybakov. Mirror averaging with sparsity priors. *Bernoulli*, 18(3):914–944, 2012.

P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *The Annals of Statistics*, 29:1–65, 2001.

A. P. Dawid. The well-calibrated Bayesian. *Journal of the American Statistical Association*, 77:605–611, 1982. Discussion: pages 611–613.

A. P. Dawid. Present position and potential developments: Some personal views, statistical theory, the prequential approach. *Journal of the Royal Statistical Society, Series A*, 147(2):278–292, 1984.

A. P. Dawid. Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis. *Statistical Science*, 19:44–57, 2004.

P. De Blasi and S. G. Walker. Bayesian asymptotics with misspecified models. *Statistica Sinica*, 23:169–187, 2013.

J. L. Doob. Application of the theory of martingales. In *Le Calcul de Probabilités et ses Applications. Colloques Internationaux du Centre National de la Recherche Scientifique*, pages 23–27, Paris, 1949.

A. Doucet and N. Shephard. Robust inference on parameters via particle filters and sandwich covariance matrices. Technical Report 606, University of Oxford, Department of Economics, 2012.

L. Dümbgen, R. Samworth, and D. Schuhmacher. Approximation by log-concave distributions, with applications to regression. *The Annals of Statistics*, 39(2):702–730, 2011.

D. B. Dunson and J. A. Taylor. Approximate Bayesian inference for quantiles. *Nonparametric Statistics*, 17(3):385–400, 2005.

T. van Erven and P. Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

T. van Erven, P. D. Grünwald, and S. de Rooij. Catching up faster in Bayesian model selection and model averaging. In *Advances in Neural Information Processing Systems (NIPS)*, 2007.

T. van Erven, P. D. Grünwald, W. M. Koolen, and S. de Rooij. Adaptive hedge. In *Advances in Neural Information Processing Systems 24 (NIPS-11)*, 2011.

T. van Erven, P. D. Grünwald, and S. de Rooij. Catching up faster by switching sooner: A predictive approach to adaptive estimation with an application to the AIC-BIC dilemma. *Journal of the Royal Statistical Society, Series B*, 74(3): 361–417, 2012. With discussion.

T. E. Feenstra. Conditional prediction without a coarsening at random condition. Master's thesis, Leiden University, 2012. Thesis adviser: P. D. Grünwald.

T. S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York, 1967.

Y. Freund, Y. Mansour, and R. E. Schapire. Generalization bounds for averaged classifiers (how to be a Bayesian without believing). *Annals of Statistics*, 32 (4):1698–1722, 2004.

G. Gallo, M. D. Grigoriadis, and R. E. Tarjan. A fast parametric maximum flow algorithm and applications. *SIAM Journal on Computing*, 18(1):30–55, 1989.

A. E. Gelfand and S. K. Ghosh. Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85(1):1–11, 1998.

A. Gelman. Bayes and Popper. Entry in A. Gelman's blog on Statistical Modeling, Causal Inference, and Social Science, 2004.

A. Gelman and C. Shalizi. Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 2012.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, third edition, 2013.

A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24:997–1016, 2014.

S. Ghosal, J. Ghosh, and A. van der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531, 2000.

R. D. Gill. The Monty Hall problem is not a probability puzzle (It's a challenge in mathematical modelling). *Statistica Neerlandica*, 65:58–71, 2011.

R. D. Gill and P. D. Grünwald. An algorithmic and a geometric characterization of coarsening at random. *The Annals of Statistics*, 36:2409–2422, 2008.

R. D. Gill, M. J. van der Laan, and J. M. Robins. Coarsening at random: Characterizations, conjectures, counter-examples. In D. Y. Lin and T. R. Fleming, editors, *Proc. First Seattle Symposium in Biostatistics: Survival Analysis*, pages 255–294, 1997.

A. V. Gnedin. The Monty Hall problem in the game theory class. *arXiv preprint arXiv:1107.0326*, 2011.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, predictions, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

A. V. Goldberg and R. E. Tarjan. A new appoach to the maximum-flow problem. *Journal of the Association for Computing Machinery*, 35(4):921–940, 1988.

G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

I. J. Good. *Good thinking: The foundations of probability and its applications*. University of Minnesota Press, 1983.

P. D. Grünwald. *The Minimum Description Length Principle and Reasoning under Uncertainty*. PhD thesis, University of Amsterdam, The Netherlands, 1998. Available as ILLC Dissertation Series 1998-03; see `www.grunwald.nl`.

P. D. Grünwald. Viewing all models as "probabilistic". In *Proceedings of the Twelfth ACM Conference on Computational Learning Theory (COLT' 99)*, pages 171–182, 1999.

P. D. Grünwald. *The Minimum Description Length Principle*. MIT Press, Cambridge, MA, 2007.

P. D. Grünwald. That simple device already used by Gauss. In P. D. Grünwald, P. Myllymäki, I. Tabus, M. Weinberger, and B. Yu, editors, *Festschrift in Honor of Jorma Rissanen on the Occasion of his 75th Birthday*, pages 293–304. Tampere University Press, Tampere, Finland, 2008.

P. D. Grünwald. Safe learning: Bridging the gap between Bayes, MDL and statistical learning theory via empirical convexity. In *Proceedings of the Twenty-Fourth Conference on Learning Theory (COLT' 11)*, 2011.

P. D. Grünwald. The safe Bayesian: Learning the learning rate via the mixability gap. In *Proceedings 23rd International Conference on Algorithmic Learning Theory (ALT '12)*. Springer, 2012.

P. D. Grünwald. Safe Bayesian learning theory. Manuscript in Preparation, 2014.

P. D. Grünwald and A. P. Dawid. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *The Annals of Statistics*, 32 (4):1367–1433, 2004.

P. D. Grünwald and J. Y. Halpern. Updating probabilities. *Journal of Artificial Intelligence Research*, 19:243–278, 2003.

P. D. Grünwald and J. Y. Halpern. When ignorance is bliss. In *Proceedings of the Twentieth Annual Conference on Uncertainty in Artificial Intelligence (UAI 2004)*, Banff, Canada, 2004.

P. D. Grünwald and J. Langford. Suboptimality of MDL and Bayes in classification under misspecification. In *Proceedings of the Seventeenth Conference on Learning Theory (COLT' 04)*, New York, 2004. Springer-Verlag.

P. D. Grünwald and J. Langford. Suboptimal behavior of Bayes and MDL in classification under misspecification. *Machine Learning*, 66(2-3):119–149, 2007. DOI 10.1007/s10994-007-0716-7.

J. Gu and S. Ghosal. Bayesian ROC curve estimation under binormality using a rank likelihood. *Journal of Statistical Planning and Inference*, 139(6):2076–2083, 2009.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Verlag, 2001.

D. F. Heitjan and D. B. Rubin. Ignorability and coarse data. *The Annals of Statistics*, 19:2244–2253, 1991.

D. P. Helmbold and M. K. Warmuth. Some weak learning results. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 399–412. ACM, 1992.

U. Hjorth. Model selection and forward validation. *Scandinavian Journal of Statistics*, 9:95–105, 1982.

P. Hoff and J. Wakefield. Bayesian sandwich posteriors for pseudo-true parameters. *arXiv preprint arXiv:1211.0087*, 2012.

C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

R. İlkılıç and Ç. Kayı. Allocation rules on networks. *Social Choice and Welfare*, 43:877–892, 2014.

M. Jaeger. Ignorability for categorical data. *The Annals of Statistics*, 33:1964–1981, 2005a.

M. Jaeger. Ignorability in statistical and probabilistic inference. *Journal of Artificial Intelligence Research*, 24:889–917, 2005b.

L. Jansen. Robust Bayesian inference under model misspecification. Master's thesis, Leiden University, 2013.

H. Jeffreys. *Theory of Probability*. Oxford University Press, London, 3rd edition, 1961.

A. Juditsky, P. Rigollet, and A. B. Tsybakov. Learning by mirror averaging. *The Annals of Statistics*, 36(5):2183–2206, 2008.

S. Jukna. *Extremal Combinatorics: With Applications in Computer Science*. Springer, Berlin, 2001.

B. Kleijn and A. van der Vaart. Misspecification in infinite-dimensional Bayesian statistics. *Annals of Statistics*, 34(2), 2006.

S. Konishi and G. Kitagawa. Generalised information criteria in model selection. *Biometrika*, 83(4):875–890, 1996.

W. Kotłowski, P. D. Grünwald, and S. de Rooij. Following the flattened leader. In *Conference on Learning Theory (COLT)*, pages 106–118, 2010.

S. Krogdahl. The dependence graph for bases in matroids. *Discrete Mathematics*, 19:47–59, 1977.

S. Lacoste-Julien, F. Huszár, and Z. Ghahramani. Approximate inference for the loss-calibrated Bayesian. *AISTATS 2011 - Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 15:416–424, 2011.

H. Leeb. Evaluation and selection of models for out-of-sample prediction when the sample size is small relative to the complexity of the data-generating process. *Bernoulli*, 14(3):661–690, 2008.

F. B. Lempers. *Posterior Probabilities of Alternative Linear Models*. University Press, Rotterdam, 1971.

J. Li, Y. Liu, L. Huang, and P. Tang. Egalitarian pairwise kidney exchange: Fast algorithms via linear programming and parametric flow. In A. Lomuscio, P. Scerri, A. Bazzan, and M. Huhns, editors, *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*, pages 445–452, 2014.

J. Q. Li. *Estimation of Mixture Models*. PhD thesis, Yale University, New Haven, CT, 1999.

F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481), 2008.

E. Mammen and S. van de Geer. Locally adaptive regression splines. *The Annals of Statistics*, 25:387–413, 1997.

D. McAllester. PAC-Bayesian stochastic model selection. *Machine Learning*, 51 (1):5–21, 2003.

N. Megiddo. Optimal flows in networks with multiple sources and sinks. *Mathematical Programming*, 7:97–107, 1974.

C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, San Fransisco, 1973.

R. H. Möhring and F. J. Radermacher. Substitution decomposition for discrete structures and connections with combinatorial optimization. *Annals of Discrete Mathematics*, 19:257–356, 1984.

H. Moulin and J. Sethuraman. The bipartite rationing problem. *Operations Research*, 61:1087–1100, 2013.

U. K. Müller. Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica*, 81(5):1805–1849, 2013.

N. Murata, S. Yoshizawa, and S. Amari. Network Information Criterion — Determining the number of hidden units for an artificial neural network model. *IEEE Transactions on Neural Networks*, 5(6):865–872, 1994.

J. Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.

A. O'Hagan. Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B*, 57(1):99–138, 1995. With discussion.

H. Owhadi and C. Scovel. Brittleness of Bayesian inference and new Selberg formulas. *arXiv preprint arXiv:1304.7046*, 2013.

J. Oxley. *Matroid Theory*. Oxford University Press, New York, second edition, 2011.

S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

T. Park and G. Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

N. Quadrianto and Z. Ghahramani. A very simple safe-Bayesian random forest. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2014. in press.

A. E. Raftery, D. Madigan, and J. A. Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92 (437):179–191, 1997.

R. V. Ramamoorthi, K. Sriram, and R. Martin. On posterior concentration in misspecified models. *arXiv preprint arXiv:1312.4620*, 2013.

J. Rissanen. Universal coding, information, prediction and estimation. *IEEE Transactions on Information Theory*, 30:629–636, 1984.

J. Robins and L. Wasserman. The foundations of statistics: A vignette. *Journal of the American Statistical Association*, 2000.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, New Jersey, 1970.

S. de Rooij, T. van Erven, P. D. Grünwald, and W. M. Koolen. Follow the leader if you can, Hedge if you must. *Journal of Machine Learning Research*, 2014.

A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125:151–188, 2005.

R. Royall and T.-S. Tsou. Interpreting statistical evidence by using imperfect models: Robust adjusted likelihood functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):391–404, 2003.

A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*, volume A. Springer, Berlin, 2003a.

A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*, volume B. Springer, Berlin, 2003b.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978.

M. Seeger. PAC-Bayesian generalization error bounds for Gaussian process classification. *Journal of Machine Learning Research*, 3:233–269, 2002.

S. Selvin. A problem in probability. *The American Statistician*, 29:67, 1975. Letter to the editor.

C. Shalizi. Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics*, 3:1039–1074, 2009.

R. Shibata. Statistical aspects of model selection. In J. C. Willems, editor, *From data to model*, pages 215–240. Springer-Verlag, 1989.

D. D. Sleator and R. E. Tarjan. A data structure for dynamic trees. *Journal of Computer and System Sciences*, 26:362–391, 1983.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64(4):583–639, 2002. With discussion.

J. P. Spinrad. *Efficient Graph Representations*. Number 19 in Fields Institute monographs. American Mathematical Society, Providence, RI, 2003.

V. Spokoiny. Parametric estimation. Finite sample theory. *The Annals of Statistics*, 40(6):2877–2909, 2012.

K. Sriram, R. V. Ramamoorthi, and P. Ghosh. Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Analysis*, 8(2):479–504, 2013.

M. Sugiyama and M. Kawanabe. *Machine learning in non-stationary environments: Introduction to covariate shift adaptation*. MIT Press, Cambridge (MA), 2012.

M. Sugiyama and H. Ogawa. Subspace information criterion for model selection. *Neural Computation*, 13(8):1863–1889, 2001.

M. Tooley. *Electronic Circuits: Fundamentals and Applications*. Routledge, London, third edition, 2006.

V. N. Vapnik. *Statistical learning theory*. Wiley, New York, 1998.

J. von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928.

M. vos Savant. Ask Marilyn. *Parade Magazine*, 15:15, 1990.

V. G. Vovk. Aggregating strategies. In *Proc. COLT' 90*, pages 371–383, 1990.

S. Walker and N. L. Hjort. On Bayesian consistency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):811–821, 2002.

S. G. Walker. Bayesian inference with misspecified models. *Journal of Statistical Planning and Inference*, 143(10):1621–1633, 2013.

S. Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3591, 2010.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

H. Whitney. On the abstract properties of linear dependence. *American Journal of Mathematics*, 57(3):509–533, 1935.

F. Willems, Y. Shtarkov, and T. Tjalkens. The context-tree weighting method: basic properties. *IEEE Transactions on Information Theory*, 41:653–664, 1995.

P. M. Williams. Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, 31(2):131–144, 1980.

H. Wong and B. Clarke. Improvement over Bayes prediction in small samples in the presence of model uncertainty. *Canadian Journal of Statistics*, 32(3): 269–283, 2004.

Y. Yang. Combining different procedures for adaptive regression. *Journal of multivariate analysis*, 74:135–161, 2000.

Y. Yang. Can the strenghts of AIC and BIC be shared? *Biometrika*, 92(4):937–950, 2005.

Y. Yang. Prediction/estimation with simple linear models: is it really that simple? *Econometric Theory*, 23:1–36, 2007a.

Y. Yang and A. R. Barron. Information-theoretic determination of minimax rates of convergence. *Annals of Statistics*, 27:1564–1599, 1999.

Z. Yang. Fair-balance paradox, star-tree paradox, and Bayesian phylogenetics. *Journal of Molecular Biology and Evolution*, 24(8):1639–1655, 2007b.

A. Zellner. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. K. Goel and A. Zellner, editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 223–243. North–Holland, Amsterdam, 1986.

T. Zhang. From ϵ-entropy to KL entropy: Analysis of minimum information complexity density estimation. *Annals of Statistics*, 34(5):2180–2210, 2006a.

T. Zhang. Information theoretical upper and lower bounds for statistical estimation. *IEEE Transactions on Information Theory*, 52(4):1307–1321, 2006b.

Index