

Bibliography

- Axelrod, Brian et al. (July 2019). “A Polynomial Time Algorithm for Log-Concave Maximum Likelihood via Locally Exponential Families”. In: arXiv: [1907.08306 \[cs.DS\]](#).
- Azzalini, A and A W Bowman (1990). “A look at some data on the old faithful geyser”. In: *J. R. Stat. Soc. Ser. C Appl. Stat.* 39.3, p. 357.
- Bagnoli, Mark and Ted Bergstrom (2005). “Log-Concave Probability and Its Applications”. In: *Economic Theory* 26.2, pp. 445–469. URL: <http://www.jstor.org/stable/25055959>.
- Baker, Charles R. (1973). “Joint Measures and Cross-Covariance Operators”. In: *Transactions of the American Mathematical Society* 186, pp. 273–289. (Visited on 06/13/2022).
- Balabdaoui, Fadoua, Kaspar Rufibach, and Jon A Wellner (June 2009). “Limit Distribution Theory for Maximum Likelihood Estimation of a Log-Concave Density”. In: *Ann. Stat.* 37.3, pp. 1299–1331.
- Barndorff-Nielsen, O (May 2014). *Information and Exponential Families: In Statistical Theory*. en. John Wiley & Sons.

- Bauer, Frank, Sergei Pereverzev, and Lorenzo Rosasco (Feb. 2007). “On regularization algorithms in learning theory”. In: *J. Complex.* 23.1, pp. 52–72.
- Bauschke, Heinz H. and Patrick L. Combettes (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 1st. Springer Publishing Company. ISBN: 9781441994660.
- Bertinet, A. and Thomas C. Agnan (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers.
- Birgé, Lucien (1989). “The Grenander Estimator: A Nonasymptotic Approach”. In: *The Annals of Statistics* 17.4, pp. 1532–1549.
- Birgé, Lucien (June 1997). “Estimation of unimodal densities without smoothness assumptions”. In: *Ann. Stat.* 25.3, pp. 970–981.
- Boyd, Stephen and Lieven Vandenberghe (Mar. 2004). *Convex Optimization*. en. Cambridge University Press.
- Brockwell, Peter J and Richard A Davis (Nov. 2013). *Time Series: Theory and Methods*. en. Springer Science & Business Media.
- Brown, Lawrence D (1986). *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*. en. IMS.
- Bühlmann, Peter and Bin Yu (June 2003). “Boosting with the L^2 Loss”. In: *J. Am. Stat. Assoc.* 98.462, pp. 324–339.
- Canu, Stéphane and Alex Smola (Mar. 2006). “Kernel methods and the exponential family”. In: *Neurocomputing* 69.7, pp. 714–720.
- Caponnetto, A and E De Vito (Aug. 2006). “Optimal Rates for the Regularized Least-Squares Algorithm”. en. In: *Found. Comput. Math.* 7.3, pp. 331–368.

- Casella, George and Roger L Berger (2002). *Statistical inference*. Vol. 2. Duxbury Pacific Grove, CA.
- Chen, Wenyu, Rahul Mazumder, and Richard J Samworth (May 2021). “A new computational framework for log-concave density estimation”. In: arXiv: [2105.11387 \[stat.CO\]](#).
- Chen, Yining and Richard J Samworth (2013). “Smoothed Log-concave Maximum Likelihood Estimation with Applications”. In: *Stat. Sin.* 23.3, pp. 1373–1398.
- Cook, R Dennis (Feb. 1977). “Detection of Influential Observation in Linear Regression”. In: *Technometrics* 19.1, pp. 15–18.
- Cook, R Dennis and Sanford Weisberg (1982). *Residuals and influence in regression*. New York: Chapman and Hall.
- Cule, M, R Samworth, and M Stewart (2010). “Maximum likelihood estimation of a multi-dimensional log-concave density”. In: *Journal of the Royal Statistical Society* 72, pp. 545–607.
- Dai, Bo et al. (Nov. 2018). “Kernel Exponential Family Estimation via Doubly Dual Embedding”. In: arXiv: [1811.02228 \[cs.LG\]](#).
- Darmois, Georges (1935). “Sur les lois de probabilité à estimation exhaustive”. In: *CR Acad. Sci. Paris* 260.1265, p. 85.
- Debruyne, Michiel, Mia Hubert, and Johan A K Suykens (2008). “Model selection in kernel based regression using the influence function”. In: *Journal of machine learning research* 9, pp. 2377–2400.

- Denkowski, Zdzislaw, Stanislaw Migórski, and Nikolaos S Papageorgiou (Dec. 2013). *An Introduction to Nonlinear Analysis: Theory*. en. Springer Science & Business Media.
- Diestel, Joseph and John Jerry Uhl (June 1977). *Vector Measures*. en. American Mathematical Soc.
- Doss, Charles R and Jon A Wellner (Apr. 2016). “Global Rates of Convergence of the MLEs of Log-concave and s -concave Densities”. en. In: *Ann. Stat.* 44.3, pp. 954–981.
- Dümbgen, Lutz, Andre Huesler, and Kaspar Rufibach (July 2007). “Active Set and EM Algorithms for Log-Concave Densities Based on Complete and Censored Data”. In: arXiv: [0707.4643 \[stat.ME\]](#).
- Dümbgen, Lutz and Kaspar Rufibach (Feb. 2009). “Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency”. en. In: *Bernoulli* 15.1, pp. 40–68.
- (2010). “logcondens: Computations related to univariate log-concave density estimation”. In: *Journal of Statistical Software*, to appear.
- Dümbgen, Lutz, Richard Samworth, and Dominic Schuhmacher (2011). “Approximation by Log-concave Distributions, with Applications to Regression”. In: *Ann. Stat.* 39.2, pp. 702–730.
- Engl, Heinz Werner, Martin Hanke, and A Neubauer (July 1996). *Regularization of Inverse Problems*. en. Springer Science & Business Media.

- Fisher, Ronald A (1922). “On the mathematical foundations of theoretical statistics”.
In: *Philosophical Transactions of the Royal Society of London. Series A, Contain-
ing Papers of a Mathematical or Physical Character* 222.594-604, pp. 309–368.
- Freedman, David and Persi Diaconis (Dec. 1981). “On the histogram as a density
estimator:L 2 theory”. en. In: *Z. Wahrscheinlichkeitstheorie verw Gebiete* 57.4,
pp. 453–476.
- Fukumizu, Kenji (2005). “Infinite dimensional exponential families by reproducing
kernel Hilbert spaces”. In: *2nd International Symposium on Information Geometry
and its Applications*, pp. 324–333.
- Fukumizu, Kenji, Francis R Bach, and Arthur Gretton (2007). “Statistical Consis-
tency of Kernel Canonical Correlation Analysis”. In: *J. Mach. Learn. Res.* 8.2.
- Fukumizu, Kenji, Francis R Bach, and Michael I Jordan (2004). “Dimensionality
Reduction for Supervised Learning with Reproducing Kernel Hilbert Spaces”. In:
J. Mach. Learn. Res. 5.Jan, pp. 73–99.
- (Aug. 2009). “Kernel dimension reduction in regression”. en. In: *aos* 37.4, pp. 1871–
1905.
- Fukumizu, Kenji et al. (2007). “Kernel measures of conditional dependence”. In: *Adv.
Neural Inf. Process. Syst.* 20.
- Fukunaga, K and L Hostetler (Jan. 1975). “The estimation of the gradient of a density
function, with applications in pattern recognition”. In: *IEEE Trans. Inf. Theory*
21.1, pp. 32–40.
- Good, I J and R A Gaskins (1971). “Nonparametric Roughness Penalties for Proba-
bility Densities”. In: *Biometrika* 58.2, pp. 255–277.

- Grenander, Ulf (1956). “On the theory of mortality measurement: part ii”. In: *Scand. Actuar. J.* 1956.2, pp. 125–153.
- Gretton, Arthur et al. (2005). “Kernel Methods for Measuring Independence”. In: *Journal of Machine Learning Research* 6.70, pp. 2075–2129. URL: <http://jmlr.org/papers/v6/gretton05a.html>.
- Gretton, Arthur et al. (2012). “A Kernel Two-Sample Test”. In: *J. Mach. Learn. Res.* 13.25, pp. 723–773.
- Groeneboom, P (1984). *Estimating a Monotone Density*. en. Centrum voor Wiskunde en Informatica.
- Gu, Chong (June 1993). “Smoothing Spline Density Estimation: A Dimensionless Automatic Algorithm”. In: *J. Am. Stat. Assoc.* 88.422, pp. 495–504.
- Gu, Chong and Chunfu Qiu (1993). “Smoothing Spline Density Estimation: Theory”. In: *Ann. Stat.* 21.1, pp. 217–234.
- Hampel, Frank R (1968). “Contributions to the theory of robust estimation”. PhD thesis. University of California.
- (June 1974). “The Influence Curve and its Role in Robust Estimation”. In: *J. Am. Stat. Assoc.* 69.346, pp. 383–393.
- Hampel, Frank R et al. (1986). *Robust Statistics: The Approach Based on Influence Functions*. en. John Wiley & Sons.
- Huber, Peter J. (1964). “Robust Estimation of a Location Parameter”. In: *The Annals of Mathematical Statistics* 35.1, pp. 73–101. DOI: [10.1214/aoms/1177703732](https://doi.org/10.1214/aoms/1177703732). URL: <https://doi.org/10.1214/aoms/1177703732>.

- Huber, Peter J (1967). “The behavior of maximum likelihood estimates under non-standard conditions”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1, pp. 221–233.
- Hyvärinen, Aapo (2005). “Estimation of Non-Normalized Statistical Models by Score Matching”. In: *J. Mach. Learn. Res.* 6.Apr, pp. 695–709.
- Ibragimov, I A (Jan. 1956). “On the Composition of Unimodal Distributions”. In: *Theory Probab. Appl.* 1.2, pp. 255–260.
- Izenman, Alan J (Mar. 2009). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. en. Springer Science & Business Media.
- Jegelka, Stefanie et al. (2009). “Generalized Clustering via Kernel Embeddings”. In: *KI 2009: Advances in Artificial Intelligence*. Springer Berlin Heidelberg, pp. 144–152.
- Kim, Arlene K H and Richard J Samworth (Dec. 2016). “Global rates of convergence in log-concave density estimation”. en. In: *aos* 44.6, pp. 2756–2779.
- Kimeldorf, George and Grace Wahba (Jan. 1971). “Some results on Tchebycheffian spline functions”. In: *J. Math. Anal. Appl.* 33.1, pp. 82–95.
- Koenker, Roger and Ivan Mizera (Mar. 2007). “Density Estimation by Total Variation Regularization”. In: *Advances in Statistical Modeling and Inference*. Vol. 3. Series in Biostatistics. World Scientific, pp. 613–633.
- Koh, Pang Wei and Percy Liang (2017). “Understanding Black-box Predictions via Influence Functions”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, pp. 1885–1894.

- Koopman, B O (1936). “On Distributions Admitting a Sufficient Statistic”. In: *Trans. Amer. Math. Soc.* 39.3, pp. 399–409.
- Le, Song (2008). *Learning Via Hilbert Space Embedding of Distributions*. en. University of Sydney.
- Leonard, Tom (Jan. 1978). “Density Estimation, Stochastic Processes and Prior Information”. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 40.2, pp. 113–132.
- Lin, Junhong et al. (Oct. 2018). “Optimal rates for spectral algorithms with least-squares regression over Hilbert spaces”. In: *Appl. Comput. Harmon. Anal.*
- Lo Gerfo, L et al. (July 2008). “Spectral algorithms for supervised learning”. en. In: *Neural Comput.* 20.7, pp. 1873–1897.
- McCullagh, P. and J.A. Nelder (1989). *Generalized Linear Models, Second Edition*. Monographs on Statistics and Applied Probability Series. Chapman & Hall. ISBN: 9780412317606.
- Muandet, Krikamol et al. (2016). “Kernel mean shrinkage estimators”. In: *J. Mach. Learn. Res.* 17.1, pp. 1656–1696.
- Muandet, Krikamol et al. (2017). “Kernel Mean Embedding of Distributions: A Review and Beyond”. In: *Foundations and Trends® in Machine Learning* 10.1-2, pp. 1–141.
- Pal, Jayanta Kumar, Michael Woodroffe, and Mary Meyer (2007). “Estimating a Polya Frequency Function₂”. In: *Lect. Notes Monogr. Ser.* 54, pp. 239–249.
- Parry, Matthew, A Philip Dawid, and Steffen Lauritzen (Feb. 2012). “Proper local scoring rules”. en. In: *Ann. Stat.* 40.1, pp. 561–592.

- Pitman, E J G (Dec. 1936). “Sufficient statistics and intrinsic accuracy”. In: *Math. Proc. Cambridge Philos. Soc.* 32.4, pp. 567–579.
- Raj, Anant et al. (Oct. 2020). “Model-specific Data Subsampling with Influence Functions”. In: arXiv: [2010.10218 \[cs.LG\]](#).
- Rao, B L S Prakasa (1969). “Estimation of a Unimodal Density”. In: *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 31.1, pp. 23–36.
- Raskutti, Garvesh, Martin J. Wainwright, and Bin Yu (2014). “Early Stopping and Non-parametric Regression: An Optimal Data-dependent Stopping Rule”. In: *Journal of Machine Learning Research* 15.11, pp. 335–366.
- Rastogi, Abhishake and Sivananthan Sampath (2017). “Optimal Rates for the Regularized Learning Algorithms under General Source Condition”. In: *Frontiers in Applied Mathematics and Statistics* 3, p. 3.
- Rathke, Fabian and Christoph Schnörr (Dec. 2019). “Fast multivariate log-concave density estimation”. In: *Comput. Stat. Data Anal.* 140, pp. 41–58.
- Reed, Michael and Barry Simon (Dec. 2012). *Methods of Modern Mathematical Physics: Functional Analysis*. en. Elsevier. ISBN: 9780125850506.
- Ripley, Brian D. (1996). *Pattern Recognition and Neural Networks*. en. Cambridge University Press. DOI: [10.1017/CB09780511812651](#).
- Royden, H. L. and P.M. Fitzpatrick (2018). *Real Analysis*. eng. Fourth edition [2018 reissue]. Pearson modern classic. New York, NY: Pearson. ISBN: 9780134689494.
- Rufibach, Kaspar (July 2007). “Computing maximum likelihood estimators of a log-concave density function”. In: *J. Stat. Comput. Simul.* 77.7, pp. 561–574.

- Samworth, Richard J (Nov. 2018). “Recent Progress in Log-Concave Density Estimation”. en. In: *Stat. Sci.* 33.4, pp. 493–509.
- Sardy, Sylvain and Paul Tseng (Jan. 2010). “Density Estimation by Total Variation Penalized Likelihood Driven by the Sparsity ℓ_1 Information Criterion: Total variation density estimation”. In: *Scand. Stat. Theory Appl.* 37.2, pp. 321–337.
- Schölkopf, Bernhard, Ralf Herbrich, and Alex J Smola (2001). “A Generalized Representer Theorem”. In: *Computational Learning Theory*. Springer Berlin Heidelberg, pp. 416–426.
- Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller (July 1998). “Non-linear Component Analysis as a Kernel Eigenvalue Problem”. In: *Neural Comput.* 10.5, pp. 1299–1319.
- Silverman, B W (Sept. 1982). “On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method”. en. In: *Ann. Stat.* 10.3, pp. 795–810.
- (1986). *Density Estimation for Statistics and Data Analysis*. Boston, MA: Chapman and Hall. ISBN: 978-0412246203.
- Smale, Steve and Ding-Xuan Zhou (Mar. 2007). “Learning Theory Estimates via Integral Operators and Their Approximations”. en. In: *Constr. Approx.* 26.2, pp. 153–172.
- Smola, Alex et al. (2007). “A Hilbert Space Embedding for Distributions”. In: *Algorithmic Learning Theory*. Springer Berlin Heidelberg, pp. 13–31.

- Song, Le, Kenji Fukumizu, and Arthur Gretton (2013). *Kernel Embeddings of Conditional Distributions: A Unified Kernel Framework for Nonparametric Inference in Graphical Models*.
- Song, Le, Arthur Gretton, and Carlos Guestrin (2010). “Nonparametric Tree Graphical Models”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterton. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, pp. 765–772.
- Song, Le et al. (Jan. 2014). “Nonparametric Latent Tree Graphical Models: Inference, Estimation, and Structure Learning”. In: arXiv: [1401.3940](https://arxiv.org/abs/1401.3940) [stat.ML].
- Sriperumbudur, Fukumizu, and others (2011). “Universality, Characteristic Kernels and RKHS Embedding of Measures”. In: *Journal of Machine Learning Research* 12.70, pp. 2389–2410.
- Sriperumbudur et al. (2010). “Hilbert space embeddings and metrics on probability measures”. In: *Journal of Machine Learning Research* 11, pp. 1517–1561.
- Sriperumbudur, Bharath et al. (2017). “Density Estimation in Infinite Dimensional Exponential Families”. In: *Journal of Machine Learning Research* 18.57, pp. 1–59. URL: <http://jmlr.org/papers/v18/16-011.html>.
- Steinwart, Ingo and Andreas Christmann (Sept. 2008). *Support Vector Machines*. en. Springer Science & Business Media.
- Sun, Siqi, Mladen Kolar, and Jinbo Xu (2015). “Learning structured densities via infinite dimensional exponential families”. In: *Advances in Neural Information*

- Processing Systems*. Ed. by C. Cortes et al. Vol. 28. Curran Associates, Inc., pp. 2287–2295.
- Sutherland, Dougal J et al. (May 2017). “Efficient and principled score estimation with Nyström kernel exponential families”. In: arXiv: [1705.08360 \[stat.ML\]](#).
- Ting, Daniel and Eric Brochu (2018). “Optimal Subsampling with Influence Functions”. In: *Adv. Neural Inf. Process. Syst.* 31.
- Tsybakov, Alexandre B (2009). *Introduction to Nonparametric Estimation*. Springer series in statistics. Dordrecht: Springer.
- Wainwright, Martin J and Michael I Jordan (2008). “Graphical Models, Exponential Families, and Variational Inference”. In: *Foundations and Trends® in Machine Learning* 1.1–2, pp. 1–305.
- Walther, Guenther (June 2002). “Detecting the Presence of Mixing with Multiscale Maximum Likelihood”. In: *J. Am. Stat. Assoc.* 97.458, pp. 508–513.
- Wasserman, Larry (Sept. 2006). *All of Nonparametric Statistics*. en. Springer Science & Business Media.
- Wegman, Edward J (1970a). “Maximum Likelihood Estimation of a Unimodal Density Function”. In: *Ann. Math. Stat.* 41.2, pp. 457–471.
- (1970b). “Maximum Likelihood Estimation of a Unimodal Density, II”. In: *Ann. Math. Stat.* 41.6, pp. 2169–2174.
- White, Halbert (1982). “Maximum Likelihood Estimation of Misspecified Models”. In: *Econometrica* 50.1, pp. 1–25.
- Williams, Christopher K I and Matthias Seeger (2001). “Using the Nyström Method to Speed Up Kernel Machines”. In: *Advances in Neural Information Processing*

- Systems 13*. Ed. by T K Leen, T G Dietterich, and V Tresp. MIT Press, pp. 682–688.
- Yao, Yuan, Lorenzo Rosasco, and Andrea Caponnetto (Aug. 2007). “On Early Stopping in Gradient Descent Learning”. In: *Constr. Approx.* 26.2, pp. 289–315.
- Yu, Shiqing, Mathias Drton, and Ali Shojaie (Sept. 2020). “Generalized Score Matching for General Domains”. In: arXiv: [2009.11428 \[stat.ME\]](#).
- Zhang, Tong and Bin Yu (Aug. 2005). “Boosting with early stopping: Convergence and consistency”. en. In: *aos* 33.4, pp. 1538–1579.
- Zhou, Ding-Xuan (Oct. 2008). “Derivative reproducing properties for kernel methods in learning theory”. In: *J. Comput. Appl. Math.* 220.1, pp. 456–463.