

Haziq Jamil

*Department of Statistics*

*London School of Economics and Political Science*

PhD thesis: ‘Regression modelling using priors depending on Fisher information covariance kernels’

11 October 2018 (v1.10f378773)

# Chapter 1

## Preceding chapters



# Bibliography

- Agresti, Alan and Jonathan Hartzel (2000). “Tutorial in biostatistics: Strategies comparing treatment on binary response with multi-centre data”. In: *Statistics in Medicine* 19, pp. 1115–1139.
- Akaike, Hirotogu (1973). “Information theory and an extension of the maximum likelihood principle”. In: *2nd International Symposium on Information Theory*. Akadémiai Kiadó, pp. 267–281.
- Albert, James H. and Siddhartha Chib (1993). “Bayesian Analysis of Binary and Polychotomous Response Data”. In: *Journal of the American Statistical Association* 88.422, pp. 669–679. DOI: [10.2307/2290350](https://doi.org/10.2307/2290350).
- Alpay, Daniel (1991). “Some Remarks on Reproducing Kernel Krein Spaces”. In: *The Rocky Mountain Journal of Mathematics* 21.4, pp. 1189–1205. DOI: [10.1216/rmj/1181072903](https://doi.org/10.1216/rmj/1181072903).
- Alvarez, Ignacio, Jarad Niemi, and Matt Simpson (2014). “Bayesian inference for a covariance matrix”. In: *26th Annual Conference on Applied Statistics in Agriculture*. DOI: [10.4148/2475-7772.1004](https://doi.org/10.4148/2475-7772.1004).
- Balakrishnan, Alampallam V. (1981). *Applied Functional Analysis*. 2nd ed. Springer-Verlag. ISBN: 978-1-4612-5867-4. DOI: [10.1007/978-1-4612-5865-0](https://doi.org/10.1007/978-1-4612-5865-0).
- Banner, Katharine M. and Megan D. Higgs (2017). “Considerations for assessing model averaging of regression coefficients”. In: *Ecological Applications* 27.1, pp. 78–93. DOI: [10.1002/eap.1419](https://doi.org/10.1002/eap.1419).
- Barbieri, Maria Maddalena and James O. Berger (2004). “Optimal predictive model selection”. In: *Annals of Statistics* 32.3, pp. 870–897. DOI: [10.1214/009053604000000238](https://doi.org/10.1214/009053604000000238).
- Bates, Douglas M., Martin Mächler, Ben Bolker, and Steve Walker (2015). “Fitting Linear Mixed-Effects Models Using **lme4**”. In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Beal, Matthew James (2003). “Variational algorithms for approximate Bayesian inference”. PhD thesis. Gatsby Computational Neuroscience Unit, University College London.
- Beal, Matthew James and Zoubin Ghahramani (2003). “The Variational Bayesian EM Algorithm for Incomplete Data: with Application to Scoring Graphical Model Structures”. In: *Bayesian Statistics 7*. Proceedings of the Seventh Valencia International Meeting. Ed. by José M. Bernardo, A. Philip Dawid, James O. Berger, Mike West,

- David Heckerman, M. J. (Susie) Bayarri, and Adrian F. M. Smith. Oxford University Press, pp. 453–464. ISBN: 978-0-19-852615-5.
- Berger, James O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag. ISBN: 978-0-387-96098-2. DOI: [10.1007/978-1-4757-4286-2](https://doi.org/10.1007/978-1-4757-4286-2).
- (2006). “The Case for Objective Bayesian Analysis”. In: *Bayesian Analysis* 1.3, pp. 385–402. DOI: [10.1214/06-BA115](https://doi.org/10.1214/06-BA115).
- Bergsma, Wicher (2018). *Regression and classification with I-priors*. Manuscript in submission. ARXIV: [1707.00274](https://arxiv.org/abs/1707.00274) [math.ST].
- Berlinet, Alain and Christine Thomas-Agnan (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Boston, MA: Springer. ISBN: 978-1-4613-4792-7. DOI: [10.1007/978-1-4419-9096-9](https://doi.org/10.1007/978-1-4419-9096-9).
- Bickel, Peter, David Choi, Xiangyu Chang, and Hai Zhang (2013). “Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels”. In: *The Annals of Statistics* 41.4, pp. 1922–1943. DOI: [10.1214/13-AOS1124](https://doi.org/10.1214/13-AOS1124).
- Bishop, Christopher (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag. ISBN: 978-0-387-31073-2.
- Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe (2017). “Variational Inference: A Review for Statisticians”. In: *Journal of the American Statistical Association* 112.518, pp. 859–877. DOI: [10.1080/01621459.2017.1285773](https://doi.org/10.1080/01621459.2017.1285773).
- Bouboulis, Pantelis and Sergios Theodoridis (2011). “Extension of Wirtinger’s Calculus to Reproducing Kernel Hilbert Spaces and the Complex Kernel LMS”. In: *IEEE Transactions on Signal Processing* 59.3, pp. 964–978. DOI: [10.1109/TSP.2010.2096420](https://doi.org/10.1109/TSP.2010.2096420).
- Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Breiman, Leo and Jerome H. Friedman (1985). “Estimating Optimal Transformations for Multiple Regression and Correlation”. In: *Journal of the American Statistical Association* 80.391, pp. 590–598. DOI: [10.1080/01621459.1985.10478157](https://doi.org/10.1080/01621459.1985.10478157).
- Breslow, Norman E. and David G. Clayton (1993). “Approximate Inference in Generalized Linear Mixed Models”. In: *Journal of the American Statistical Association* 88.421, pp. 9–25. DOI: [10.2307/2290687](https://doi.org/10.2307/2290687).
- Bunch, David S. (1991). “Estimability in the multinomial probit model”. In: *Transportation Research Part B: Methodological* 25.1, pp. 1–12. DOI: [10.1016/0191-2615\(91\)90009-8](https://doi.org/10.1016/0191-2615(91)90009-8).
- Bürkner, Paul-Christian (2017). “brms: An R Package for Bayesian Multilevel Models Using Stan”. In: *Journal of Statistical Software* 80.1, pp. 1–28. DOI: [10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01).
- Butler, John S. and Robert Moffitt (1982). “A computationally efficient quadrature procedure for the one-factor multinomial probit model”. In: *Econometrica* 50.3, pp. 761–764. DOI: [10.2307/1912613](https://doi.org/10.2307/1912613).
- Cade, Brian S (2015). “Model averaging and muddled multimodel inferences”. In: *Ecology* 96.9, pp. 2370–2382. DOI: [10.1890/14-1639.1](https://doi.org/10.1890/14-1639.1).

- Cannings, Timothy I. and Richard J. Samworth (2017). “Random-projection ensemble classification”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.4, pp. 959–1035. DOI: [10.1111/rssb.12228](https://doi.org/10.1111/rssb.12228).
- Carlin, Bradley P. and Siddhartha Chib (1995). “Bayesian Model Choice via Markov Chain Monte Carlo Methods”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 57.3, pp. 473–484.
- Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell (2017). “Stan: A Probabilistic Programming Language”. In: *Journal of Statistical Software* 76.1, pp. 1–32. DOI: [10.18637/jss.v076.i01](https://doi.org/10.18637/jss.v076.i01).
- Casella, George (1985). “An Introduction to Empirical Bayes Data Analysis”. In: *The American Statistician* 39.2, pp. 83–87. DOI: [10.2307/2682801](https://doi.org/10.2307/2682801).
- Casella, George and Roger L. Berger (2002). *Statistical Inference*. 2nd ed. Pacific Grove, CA: Duxbury. ISBN: 978-0-534-24312-8.
- Casella, George, F. Javier Girón, M. Lina Martínez, and Elías Moreno (2009). “Consistency of Bayesian procedures for variable selection”. In: *The Annals of Statistics* 37.3, pp. 1207–1228. DOI: [10.1214/08-AOS606](https://doi.org/10.1214/08-AOS606).
- Casella, George and Elías Moreno (2006). “Objective Bayesian Variable Selection”. In: *Journal of the American Statistical Association* 101.473, pp. 157–167. DOI: [10.1198/016214505000000646](https://doi.org/10.1198/016214505000000646).
- Chan, Jennifer S. K. and Anthony Y. C. Kuk (1997). “Maximum Likelihood Estimation for Probit-Linear Mixed Models with Correlated Random Effects”. In: *Biometrics* 53.1, pp. 86–97. DOI: [10.2307/2533099](https://doi.org/10.2307/2533099).
- Chen, Dong, Peter Hall, and Hans-Georg Müller (2011). “Single and Multiple Index Functional Regression Models with Nonparametric Link”. In: *The Annals of Statistics* 39.3, pp. 1720–1747. DOI: [10.1214/11-AOS882](https://doi.org/10.1214/11-AOS882).
- Chen, Yen-Chi, Y. Samuel Wang, and Elena A. Eroshova (2018). “On the use of bootstrap with variational inference: Theory, interpretation, and a two-sample test example”. In: *Annals of Applied Statistics* to appear. ARXIV: [1711.11057](https://arxiv.org/abs/1711.11057) [[stat.ME](https://arxiv.org/archive/stat)].
- Cheng, Ching-An and Byron Boots (2017). “Variational Inference for Gaussian Process Models with Linear Complexity”. In: *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Ed. by Isabelle Guyon, Ulrike Von Luxburg, Samy Bengio, Hanna Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, pp. 5184–5194. ARXIV: [1711.10127](https://arxiv.org/abs/1711.10127) [[stat.ML](https://arxiv.org/archive/stat)].
- Chipman, Hugh, Edward I. George, and Robert E. McCulloch (2001). “The Practical Implementation of Bayesian Model Selection”. In: *Model Selection*. Ed. by P. Lahiri. Vol. 38. Institute of Mathematical Statistics, pp. 65–134. DOI: [10.1214/lnms/1215540964](https://doi.org/10.1214/lnms/1215540964).
- Chopin, Nicolas (2011). “Fast simulation of truncated Gaussian distributions”. In: *Statistics and Computing* 21.2, pp. 275–288. DOI: [10.1007/s11222-009-9168-1](https://doi.org/10.1007/s11222-009-9168-1).
- Cohen, Serge (2002). “Champs localement auto-similaires”. In: *Lois d’échelle, fractales et ondelettes*. Ed. by Patrice Abry, Paulo Gonçalves, and Jacques Lévy Véhel. Vol. 1. Hermès Sciences Publications. ISBN: 978-2-7462-0409-6.

- Damien, Paul and Stephen G. Walker (2001). “Sampling Truncated Normal, Beta, and Gamma Densities”. In: *Journal of Computational and Graphical Statistics* 10.2, pp. 206–215. DOI: [10.1198/10618600152627906](https://doi.org/10.1198/10618600152627906).
- Dansie, Brenton R. (1985). “Parameter estimability in the multinomial probit model”. In: *Transportation Research Part B: Methodological* 19.6, pp. 526–528. DOI: [10.1016/0191-2615\(85\)90047-5](https://doi.org/10.1016/0191-2615(85)90047-5).
- Davidian, Marie and David M. Giltinan (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC. ISBN: 978-0-412-98341-2.
- Davison, Anthony Christopher (2003). *Statistical Models*. Cambridge University Press. ISBN: 978-0-511-81585-0. DOI: [10.1017/CB09780511815850](https://doi.org/10.1017/CB09780511815850).
- Dean, Angela and Daniel Voss (1999). *Design and Analysis of Experiments*. Springer. ISBN: 978-0-387-98561-9. DOI: [10.1007/978-3-319-52250-0](https://doi.org/10.1007/978-3-319-52250-0).
- Dellaportas, Petros, Jonathan J. Forster, and Ioannis Ntzoufras (2002). “On Bayesian model and variable selection using MCMC”. In: *Statistics and Computing* 12.1, pp. 27–36. DOI: [10.1023/A:1013164120801](https://doi.org/10.1023/A:1013164120801).
- Dempster, Arthur P, Nan M Laird, and Donald B. Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 1–38.
- Denwood, Matthew (2016). “**runjags**: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS”. In: *Journal of Statistical Software* 71.9, pp. 1–25. DOI: [10.18637/jss.v071.i09](https://doi.org/10.18637/jss.v071.i09).
- Deterding, David Henry (1990). “Speaker Normalization for Automatic Speech Recognition”. PhD thesis. University of Cambridge.
- Diggle, Peter, Paula Moraga, Barry Rowlingson, and Benjamin Taylor (2013). “Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm”. In: *Statistical Science* 28.4, pp. 542–563. DOI: [10.1214/13-STS441](https://doi.org/10.1214/13-STS441).
- Diggle, Peter, Pingping Zheng, and Peter Durr (2005). “Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.3, pp. 645–658. DOI: [10.1111/j.1467-9876.2005.05373.x](https://doi.org/10.1111/j.1467-9876.2005.05373.x).
- Duane, Simon, Anthony D Kennedy, Brian J. Pendleton, and Duncan Roweth (1987). “Hybrid Monte Carlo”. In: *Physics Letters B* 195.2, pp. 216–222. DOI: [10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X).
- Durrande, Nicolas, David Ginsbourger, Olivier Roustant, and Laurent Carraro (2013). “ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis”. In: *Journal of Multivariate Analysis* 115, pp. 57–67. DOI: [10.1016/j.jmva.2012.08.016](https://doi.org/10.1016/j.jmva.2012.08.016).
- Duvenaud, David (2014). “Automatic Model Construction with Gaussian Processes”. PhD thesis. University of Cambridge.
- Eddelbuettel, Dirk and Romain Francois (2011). “**Rcpp**: Seamless R and C++ Integration”. In: *Journal of Statistical Software* 40.8, pp. 1–18. DOI: [10.18637/jss.v040.i08](https://doi.org/10.18637/jss.v040.i08).

- Efron, Bradley and David V. Hinkley (1978). “Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information”. In: *Biometrika* 65.3, pp. 457–483. DOI: [10.1093/biomet/65.3.457](https://doi.org/10.1093/biomet/65.3.457).
- Embrechts, Paul and Makoto Maejima (2002). *Selfsimilar Processes*. Princeton, NJ: Princeton University Press. ISBN: 978-0-691-09627-8.
- Fernández, Carmen, Eduardo Ley, and Mark F. J. Steel (2001). “Benchmark priors for Bayesian model averaging”. In: *Journal of Econometrics* 100.2, pp. 381–427. DOI: [10.1016/S0304-4076\(00\)00076-2](https://doi.org/10.1016/S0304-4076(00)00076-2).
- Ferraty, Frédéric and Philippe Vieu (2006). *Nonparametric Functional Data Analysis*. New York: Springer. ISBN: 978-0-387-30369-7. DOI: [10.1007/0-387-36620-2](https://doi.org/10.1007/0-387-36620-2).
- Fisher, Ronald Aylmer (1922). “On the mathematical foundations of theoretical statistics”. In: *Philosophical Transactions of the Royal Society A* 222.594-604, pp. 309–368. DOI: [10.1098/rsta.1922.0009](https://doi.org/10.1098/rsta.1922.0009).
- Fouskakis, Dimitris and David Draper (2008). “Comparing Stochastic Optimization Methods for Variable Selection in Binary Outcome Prediction, With Application to Health Policy”. In: *Journal of the American Statistical Association* 103.484, pp. 1367–1381. DOI: [10.1198/016214508000001048](https://doi.org/10.1198/016214508000001048).
- Fowlkes, Charless, Serge Belongie, Fan Chung, and Jitendra Malik (2004). “Spectral grouping using the Nyström method”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.2, pp. 214–225. DOI: [10.1109/TPAMI.2004.1262185](https://doi.org/10.1109/TPAMI.2004.1262185).
- Fowlkes, Charless, Serge Belongie, and Jitendra Malik (Dec. 2001). “Efficient Spatiotemporal Grouping Using the Nyström Method”. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, HI*. Vol. 1, pp. 231–238. DOI: [10.1109/CVPR.2001.990481](https://doi.org/10.1109/CVPR.2001.990481).
- Friedman, Jerome H., Trevor Hastie, and Robert Tibshirani (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer-Verlag. ISBN: 978-0-387-84857-0. DOI: [10.1007/978-0-387-84858-7](https://doi.org/10.1007/978-0-387-84858-7).
- George, Edward I. and Dean P. Foster (2000). “Calibration and Empirical Bayes Variable Selection”. In: *Biometrika* 87.4, pp. 731–747. DOI: [10.1093/biomet/87.4.731](https://doi.org/10.1093/biomet/87.4.731).
- George, Edward I. and Robert E. McCulloch (1993). “Variable Selection Via Gibbs Sampling”. In: *Journal of the American Statistical Association* 88.423, pp. 881–889. DOI: [10.2307/2290777](https://doi.org/10.2307/2290777).
- Geweke, John (1989). “Bayesian Inference in Econometric Models Using Monte Carlo Integration”. In: *Econometrica* 57.6, pp. 1317–1339. DOI: [10.2307/1913710](https://doi.org/10.2307/1913710).
- (1991). “Efficient Simulation from the Multivariate Normal and Student-*t* Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities”. In: *Computing Science and Statistics: the Twenty-Third Symposium on the Interface*. Ed. by Elaine M. Keramidas, pp. 571–578.
- (1996). “Variable Selection and Model Comparison in Regression”. In: *Bayesian Statistics 5*. Proceedings of the Fifth Valencia International Meeting. Ed. by José M. Bernardo, James O. Berger, A. Philip Dawid, and Adrian F. M. Smith. Oxford University Press. ISBN: 978-0-19-852356-7.



- Geweke, John, Michael Keane, and David Runkle (1994). “Alternative Computational Approaches to Inference in the Multinomial Probit Model”. In: *The Review of Economics and Statistics* 76.4, pp. 609–632. DOI: [10.2307/2109766](https://doi.org/10.2307/2109766).
- Girolami, Mark and Simon Rogers (2006). “Variational Bayesian Multinomial Probit Regression with Gaussian Process Priors”. In: *Neural Computation* 18.8, pp. 1790–1817. DOI: [10.1162/neco.2006.18.8.1790](https://doi.org/10.1162/neco.2006.18.8.1790).
- Goia, Aldo and Philippe Vieu (2014). “Some Advances in Semiparametric Functional Data Modelling”. In: *Contributions in Infinite-Dimensional Statistics and Related Topics*. Societa Editrice Esculapio, pp. 135–141. DOI: [10.15651/978-88-748-8763-7](https://doi.org/10.15651/978-88-748-8763-7).
- Gretton, Arthur, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf (Oct. 2005). “Measuring Statistical Dependence with Hilbert-Schmidt Norms”. In: *Proceedings of the Sixteenth International Conference of Algorithmic Learning Theory (ALT 2005), Singapore*. Ed. by Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita. Springer-Verlag, pp. 63–77. DOI: [10.1007/11564089\\_7](https://doi.org/10.1007/11564089_7).
- Groves, Theodore and Thomas Rothenberg (1969). “A note on the expected value of an inverse matrix”. In: *Biometrika* 56.3, pp. 690–691. DOI: [10.1093/biomet/56.3.690](https://doi.org/10.1093/biomet/56.3.690).
- Gu, Chong (2013). *Smoothing Spline ANOVA Models*. New York: Springer-Verlag. ISBN: 978-1-4614-5368-0. DOI: [10.1007/978-1-4614-5369-7](https://doi.org/10.1007/978-1-4614-5369-7).
- Güvenir, H. Altay, Burak Acar, Gulsen Demiroz, and Ayhan Cekin (1997). “A supervised machine learning algorithm for arrhythmia analysis”. In: *Computers in Cardiology 1997*. Lund, Sweden, pp. 433–436. DOI: [10.1109/CIC.1997.647926](https://doi.org/10.1109/CIC.1997.647926).
- Hajivassiliou, Vassilis and Daniel McFadden (1998). “The Method of Simulated Scores for the Estimation of LDV Models”. In: *Econometrica* 66.4, pp. 863–896. DOI: [10.2307/2999576](https://doi.org/10.2307/2999576).
- Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud (1996). “Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results”. In: *Journal of Econometrics* 72.1–2, pp. 85–134. DOI: [10.1016/0304-4076\(94\)01716-6](https://doi.org/10.1016/0304-4076(94)01716-6).
- Hall, Peter, Tung Pham, Matt P. Wand, and Shen S. J. Wang (2011). “Asymptotic normality and valid inference for Gaussian variational approximation”. In: *The Annals of Statistics* 39.5, pp. 2502–2532. DOI: [10.1214/11-AOS908](https://doi.org/10.1214/11-AOS908).
- Hastie, Trevor and Robert Tibshirani (1986). “Generalized Additive Models”. In: *Statistical Science* 1.3, pp. 297–310. DOI: [10.1214/ss/1177013604](https://doi.org/10.1214/ss/1177013604).
- Hein, Matthias and Olivier Bousquet (2004). *Kernels, Associated Structures and Generalizations*. Tech. rep. Max Planck Institute for Biological Cybernetics, Tübingen, Germany.
- Hensman, James, Nicolo Fusi, and Neil D. Lawrence (Aug. 2013). “Gaussian Processes for Big Data”. In: *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI 2013), Bellevue, WA*. Ed. by Ann Nicholson and Padhraic Smyth. Arlington, VA: AUAI Press. ISBN: 978-0-9749039-9-6. ARXIV: [1309.6835 \[cs.LG\]](https://arxiv.org/abs/1309.6835).



- Hoerl, Arthur E. and Robert W. Kennard (1970). “Ridge Regression: Biased Estimation for Nonorthogonal Problems”. In: *Technometrics* 12.1, pp. 55–67. DOI: [10.2307/1267351](https://doi.org/10.2307/1267351).
- Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky (1999). “Bayesian Model Averaging: A Tutorial”. In: *Statistical science* 14.4, pp. 382–401. DOI: [10.1214/ss/1009212519](https://doi.org/10.1214/ss/1009212519).
- Hofmann, Thomas, Bernhard Schölkopf, and Alexander J. Smola (2008). “Kernel Methods in Machine Learning”. In: *The Annals of Statistics* 36.3, pp. 1171–1220. DOI: [10.1214/009053607000000677](https://doi.org/10.1214/009053607000000677).
- Itzykson, Claude and Jean-Michel Drouffe (1991). *Statistical Field Theory*. Vol. 2: Strong Coupling, Monte Carlo Methods, Conformal Field Theory and Random Systems. Cambridge University Press. ISBN: 978-0-511-62278-6. DOI: [10.1017/CB09780511622786](https://doi.org/10.1017/CB09780511622786).
- Jamil, Haziq (2017). **iprior**: *Regression Modelling using I-Priors*. R package version 0.7.1. URL: <https://cran.r-project.org/web/packages/iprior/>.
- (2018a). **ipriorBVS**: *Bayesian Variable Selection using I-priors*. R package version 0.1.1. URL: <https://github.com/haziqj/ipriorBVS>.
- (2018b). **iprobit**: *Binary and Multinomial Probit Regression with I-priors*. R package version 0.1.1. URL: <https://github.com/haziqj/iprobit>.
- Jamil, Haziq and Wicher Bergsma (2017). “**iprior**: An R Package for Regression Modelling using I-priors”. R package vignette. URL: [https://cran.r-project.org/web/packages/iprior/vignettes/iprior\\_paper.pdf](https://cran.r-project.org/web/packages/iprior/vignettes/iprior_paper.pdf).
- Jaynes, Edwin Thompson (1957a). “Information Theory and Statistical Mechanics”. In: *Physical Review* 106.4, p. 620. DOI: [10.1103/PhysRev.106.620](https://doi.org/10.1103/PhysRev.106.620).
- (1957b). “Information Theory and Statistical Mechanics II”. In: *Physical Review* 108.2, p. 171. DOI: [10.1103/PhysRev.108.171](https://doi.org/10.1103/PhysRev.108.171).
- (2003). *Probability Theory: The Logic of Science*. Ed. by G. Larry Bretthorst. Cambridge University Press. ISBN: 978-0-521-59271-0.
- Jeffreys, Harold (1946). “An invariant form for the prior probability in estimation problems”. In: *Proceedings of the Royal Society A* 186.1007, pp. 453–461. DOI: [10.1098/rspa.1946.0056](https://doi.org/10.1098/rspa.1946.0056).
- Jordan, Michael I., Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul (1999). “An Introduction to Variational Methods for Graphical Models”. In: *Machine Learning* 37.2, pp. 183–233. DOI: [10.1023/A:1007665907178](https://doi.org/10.1023/A:1007665907178).
- Kadane, Joseph B. (2011). *Principles of Uncertainty*. Chapman & Hall/CRC. ISBN: 978-1-4398-6161-5.
- Kahle, David and Hadley Wickham (2013). “**ggmap**: Spatial Visualization with **ggplot2**”. In: *The R Journal* 5.1, pp. 144–161.
- Kalaitzis, Alfredo, Antti Honkela, Pei Gao, and Neil D. Lawrence (2014). **gptk**: *Gaussian Processes Tool-Kit*. R package version 1.08. URL: <https://CRAN.R-project.org/package=gptk>.

- Kalaitzis, Alfredo and Neil D. Lawrence (May 2011). “A Simple Approach to Ranking Differentially Expressed Gene Expression Time Courses through Gaussian Process Regression”. In: *BMC Bioinformatics* 12.1, p. 180. DOI: [10.1186/1471-2105-12-180](https://doi.org/10.1186/1471-2105-12-180).
- Kammar, Ohad (2016). *A note on Fréchet differentiation under Lebesgue integrals*. URL: <https://www.cs.ox.ac.uk/people/ohad.kammar/notes/kammar-a-note-on-frechet-differentiation-under-lebesgue-integrals.pdf>.
- Karatzoglou, Alexandros, Alexander J. Smola, Kurt Hornik, and Achim Zeileis (2004). “**kernlab** - An S4 Package for Kernel Methods in R”. In: *Journal of Statistical Software* 11.9, pp. 1–20. DOI: [10.18637/jss.v011.i09](https://doi.org/10.18637/jss.v011.i09).
- Kass, Robert E. and Adrian E. Raftery (1995). “Bayes Factors”. In: *Journal of the American Statistical Association* 90.430, pp. 773–795. DOI: [10.2307/2291091](https://doi.org/10.2307/2291091).
- Keane, Michael (1992). “A Note on Identification in the Multinomial Probit Model”. In: *Journal of Business & Economic Statistics* 10.2, pp. 193–200. DOI: [10.2307/1391677](https://doi.org/10.2307/1391677).
- (1994). “A Computationally Practical Simulation Estimator for Panel Data”. In: *Econometrica* 62.1, pp. 95–116. DOI: [10.2307/2951477](https://doi.org/10.2307/2951477).
- Keane, Michael and Kenneth Wolpin (1994). “The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence”. In: *The Review of Economics and Statistics* 76.4, pp. 648–672. DOI: [10.2307/2109768](https://doi.org/10.2307/2109768).
- Kenward, Michael G. (1987). “A Method for Comparing Profiles of Repeated Measurements”. In: *Journal of the Royal Statistical Society C (Applied Statistics)* 36.3, pp. 296–308. DOI: [10.2307/2347788](https://doi.org/10.2307/2347788).
- Kimeldorf, George S and Grace Wahba (1970). “A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines”. In: *The Annals of Mathematical Statistics* 41.2, pp. 495–502. DOI: [10.1214/aoms/1177697089](https://doi.org/10.1214/aoms/1177697089).
- Kokoszka, Piotr and Matthew Reimherr (2017). *Introduction to Functional Data Analysis*. Chapman & Hall/CRC. ISBN: 978-1-4987-4634-2.
- Krée, Paul (1974). “Produits tensoriels complétés d’espaces de Hilbert”. In: *Séminaire Paul Krée* 1.7, pp. 1974–1975.
- Kuhn, Max et al. (2017). **caret**: *Classification and Regression Training*. R package version 6.0–77. URL: <https://CRAN.R-project.org/package=caret>.
- Kuo, Frances Y., Ian H. Sloan, Grzegorz Wasilkowski, and Henryk Woźniakowski (2010). “On decompositions of multivariate functions”. In: *Mathematics of Computation* 79.270, pp. 953–966. DOI: [10.1090/S0025-5718-09-02319-9](https://doi.org/10.1090/S0025-5718-09-02319-9).
- Kuo, Lynn and Bani Mallick (1998). “Variable selection for regression models”. In: *Sankhyā: The Indian Journal of Statistics, Series B* 60.1, pp. 65–81.
- Kuss, Malte and Carl Edward Rasmussen (2005). “Assessing Approximate Inference for Binary Gaussian Process Classification”. In: *Journal of Machine Learning Research* 6, pp. 1679–1704.
- Kyung, Minjung, Jeff Gill, Malay Ghosh, and George Casella (2010). “Penalized regression, standard errors, and Bayesian lassos”. In: *Bayesian Analysis* 5.2, pp. 369–411. DOI: [10.1214/10-BA607](https://doi.org/10.1214/10-BA607).

- Lange, Kenneth (1995). “A quasi-Newton acceleration of the EM algorithm”. In: *Statistica Sinica* 5.1, pp. 1–18.
- Lee, Kyeong Eun, Naijun Sha, Edward R. Dougherty, Marina Vannucci, and Bani Mallick (2003). “Gene selection: a Bayesian variable selection approach”. In: *Bioinformatics* 19.1, pp. 90–97. DOI: [10.1093/bioinformatics/19.1.90](https://doi.org/10.1093/bioinformatics/19.1.90).
- Leisch, Friedrich and Evgenia Dimitriadou (2010). *mlbench: Machine Learning Benchmark Problems*. R package version 2.1-1.
- Lian, Heng and Gaorong Li (2014). “Series Expansion for Functional Sufficient Dimension Reduction”. In: *Journal of Multivariate Analysis* 124, pp. 150–165. DOI: [10.1016/j.jmva.2013.10.019](https://doi.org/10.1016/j.jmva.2013.10.019).
- Liang, Feng, Rui Paulo, German Molina, Merlise A. Clyde, and James O. Berger (2008). “Mixtures of  $g$  Priors for Bayesian Variable Selection”. In: *Journal of the American Statistical Association* 103.481, pp. 410–423. DOI: [10.1198/016214507000001337](https://doi.org/10.1198/016214507000001337).
- Lifshits, Mikhail (2012). “Lectures on Gaussian processes”. In: *Lectures on Gaussian Processes*. Springer-Verlag, pp. 1–117. DOI: [10.1007/978-3-642-24939-6](https://doi.org/10.1007/978-3-642-24939-6).
- Lindley, Dennis V. (1957). “A statistical paradox”. In: *Biometrika* 44.1–2, pp. 187–192. DOI: [10.1093/biomet/44.1-2.187](https://doi.org/10.1093/biomet/44.1-2.187).
- Liu, Chuanhai, Donald B. Rubin, and Ying Nian Wu (1998). “Parameter expansion to accelerate EM: The PX-EM algorithm”. In: *Biometrika* 85.4, pp. 755–770. DOI: [10.1093/biomet/85.4.755](https://doi.org/10.1093/biomet/85.4.755).
- Lloyd, Chris, Tom Gunter, Michael Osborne, and Stephen Roberts (July 2015). “Variational Inference for Gaussian Process Modulated Poisson Processes”. In: *Proceedings of the Thirty-Second International Conference on Machine Learning (ICML 2015), Lille, France*. Ed. by Francis Bach and David Blei. Vol. 37, pp. 1814–1822. ARXIV: [1411.0254](https://arxiv.org/abs/1411.0254) [stat.ML].
- Louppe, Gilles (Oct. 2014). “Understanding Random Forests: From Theory to Practice”. PhD thesis. University of Liege, Belgium. ARXIV: [1407.7502](https://arxiv.org/abs/1407.7502) [stat.ML].
- Lunn, David J., Andrew Thomas, Nicky Best, and David Spiegelhalter (Oct. 2000). “WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility”. In: *Statistics and Computing* 10.4, pp. 325–337. DOI: [10.1023/A:1008929526011](https://doi.org/10.1023/A:1008929526011).
- MacDonald, Blake, Pritam Ranjan, and Hugh Chipman (2015). “**GPfit**: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs”. In: *Journal of Statistical Software* 64.12, pp. 1–23. DOI: [10.18637/jss.v064.i12](https://doi.org/10.18637/jss.v064.i12).
- Madigan, David and Adrian E. Raftery (1994). “Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam’s Window”. In: *Journal of the American Statistical Association* 89.428, pp. 1535–1546. DOI: [10.2307/2291017](https://doi.org/10.2307/2291017).
- Mallows, Colin L. (1973). “Some comments on  $C_p$ ”. In: *Technometrics* 15.4, pp. 661–675. DOI: [10.2307/1267380](https://doi.org/10.2307/1267380).
- Mandelbrot, Benoit B. and John W. Van Ness (1968). “Fractional Brownian Motions, Fractional Noises and Applications”. In: *SIAM Review* 10.4, pp. 422–437.

- Marsaglia, George and Wai Wan Tsang (2000). “The Ziggurat Method for Generating Random Variables”. In: *Journal of Statistical Software* 5.8, pp. 1–7. DOI: [10.18637/jss.v005.i08](https://doi.org/10.18637/jss.v005.i08).
- Mary, Xavier (2003). “Hilbertian subspaces, subdualities and applications”. PhD thesis. INSA de Rouen.
- McCullagh, Peter and John A. Nelder (1989). *Generalized Linear Models*. 2nd ed. Chapman & Hall/CRC. ISBN: 978-0-412-31760-6.
- McCulloch, Robert E., Nicholas G. Polson, and Peter E. Rossi (2000). “A Bayesian analysis of the multinomial probit model with fully identified parameters”. In: *Journal of Econometrics* 99.1, pp. 173–193. DOI: [10.1016/S0304-4076\(00\)00034-8](https://doi.org/10.1016/S0304-4076(00)00034-8).
- McCulloch, Robert E. and Peter E. Rossi (1994). “An exact likelihood analysis of the multinomial probit model”. In: *Journal of Econometrics* 64.1, pp. 207–240. DOI: [10.1016/0304-4076\(94\)90064-7](https://doi.org/10.1016/0304-4076(94)90064-7).
- McDonald, Gary C. and Richard C. Schwing (1973). “Instabilities of Regression Estimates Relating Air Pollution to Mortality”. In: *Technometrics* 15.3, pp. 463–481. DOI: [10.2307/1266852](https://doi.org/10.2307/1266852).
- McLachlan, Geoffrey and Thriyambakam Krishnan (2007). *The EM Algorithm and Extensions*. 2nd ed. John Wiley & Sons. ISBN: 978-0-471-20170-0. DOI: [10.1002/9780470191613](https://doi.org/10.1002/9780470191613).
- Meila, Marina (2003). “Data centering in feature space.” In: *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics (AISTATS)*. Ed. by Christopher Bishop and Brendan Frey. Society for Artificial Intelligence and Statistics.
- Meng, Xiao-Li and Donald B. Rubin (1993). “Maximum likelihood estimation via the ECM algorithm: A general framework”. In: *Biometrika* 80.2, pp. 267–278. DOI: [10.1093/biomet/80.2.267](https://doi.org/10.1093/biomet/80.2.267).
- Meng, Xiao-Li and David Van Dyk (1997). “The EM Algorithm—an Old Folk-song Sung to a Fast New Tune”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59.3, pp. 511–567. DOI: [10.1111/1467-9868.00082](https://doi.org/10.1111/1467-9868.00082).
- Micchelli, Charles A., Yuesheng Xu, and Haizhang Zhang (2006). “Universal Kernels”. In: *Journal of Machine Learning Research* 7, pp. 2651–2667.
- Microsoft Corporation and Stephen Weston (2017). **doSNOW**: *Foreach Parallel Adaptor for the snow Package*. R package version 1.0.15. URL: <https://CRAN.R-project.org/package=doSNOW>.
- Miller, Alan (2002). *Subset Selection in Regression*. Chapman & Hall/CRC. ISBN: 978-1-58488-171-1.
- Minka, Thomas P. (Aug. 2001). “Expectation propagation for approximate Bayesian inference”. In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI 2001)*, Seattle, WA. Ed. by Daphne Koller John Breese. San Francisco, CA: Morgan Kaufmann Publishers Inc., pp. 362–369. ISBN: 1-55860-800-1. ARXIV: [1301.2294](https://arxiv.org/abs/1301.2294) [cs.AI].
- Mitchell, Toby J. and John J. Beauchamp (1988). “Bayesian Variable Selection in Linear Regression”. In: *Journal of the American Statistical Association* 83.404, pp. 1023–1032. DOI: [10.2307/2290129](https://doi.org/10.2307/2290129).

- Neal, Radford M. (1999). “Regression and Classification using Gaussian Process Priors”. In: *Bayesian Statistics 6*. Proceedings of the Sixth Valencia International Meeting. Ed. by José M. Bernardo, James O. Berger, A. Philip Dawid, and Adrian F. M. Smith. Oxford University Press, pp. 475–501. ISBN: 978-0-19-850485-6.
- (2011). “MCMC using Hamiltonian dynamics”. In: *Handbook of Markov Chain Monte Carlo*. Ed. by Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. Chapman & Hall/CRC. ISBN: 978-1-4200-7941-8. ARXIV: [1206.1901 \[stat.CO\]](#).
- Nobile, Agostino (1998). “A hybrid Markov chain for the Bayesian analysis of the multinomial probit model”. In: *Statistics and Computing* 8.3, pp. 229–242. DOI: [10.1023/A:10089053](#).
- Nocedal, Jorge and Stephen Wright (2006). *Numerical Optimization*. New York: Springer-Verlag. ISBN: 978-0-387-30303-1. DOI: [10.1007/978-0-387-40065-5](#).
- Ntzoufras, Ioannis (2011). *Bayesian Modeling Using WinBUGS*. Wiley. ISBN: 978-0-470-14114-4. DOI: [10.1002/9780470434567](#).
- O’Hara, Robert B. and Mikko J. Sillanpää (2009). “A Review of Bayesian Variable Selection Methods: What, How and Which”. In: *Bayesian Analysis* 4.1, pp. 85–117. DOI: [10.1214/09-BA403](#).
- Ong, Cheng Soon, Xavier Mary, Stéphane Canu, and Alexander J. Smola (July 2004). “Learning with non-positive kernels”. In: *Proceedings of the Twenty-First International Conference on Machine Learning (ICML 2004)*, Banff, Alberta, Canada. Ed. by Russ Greiner and Dale Schuurmans.
- Ormerod, John T., Chong You, and Samuel Müller (2017). “A variational Bayes approach to variable selection”. In: *Electronic Journal of Statistics* 11.2, pp. 3549–3594. DOI: [10.1214/17-EJS1332](#).
- Pan, Jianxin and Yi Pan (2017). “**jmcm**: An R Package for Joint Mean-Covariance Modeling of Longitudinal Data”. In: *Journal of Statistical Software* 82.1, pp. 1–29. DOI: [10.18637/jss.v082.i09](#).
- Park, Trevor and George Casella (2008). “The Bayesian Lasso”. In: *Journal of the American Statistical Association* 103.482, pp. 681–686. DOI: [10.1198/016214508000000337](#).
- Pawitan, Yudi (2001). *In All Likelihood*. Statistical Modelling and Inference Using Likelihood. Oxford University Press. ISBN: 978-0-19-850765-9.
- Petersen, Kaare Brandt and Michael Syskind Pedersen (2012). *The Matrix Cookbook*. Technical University of Denmark. URL: <http://www2.imm.dtu.dk/pubdb/p.php?3274>.
- Pinheiro, José C. and Douglas M. Bates (2000). *Mixed-Effects Models in S and S-plus*. Springer-Verlag. ISBN: 978-0-387-98957-0. DOI: [10.1007/b98882](#).
- Pinheiro, José C., Douglas M. Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team (2017). **nlme**: *Linear and Nonlinear Mixed Effects Models*. R package version 3.1-131. URL: <https://CRAN.R-project.org/package=nlme>.
- Plummer, Martyn (Mar. 2003). “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling”. In: *Proceedings of the Third International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria. Ed. by Kurt Hornik, Friedrich Leisch, and Achim Zeileis.



- Quiñonero-Candela, Joaquin and Carl Edward Rasmussen (Dec. 2005). “A Unifying View of Sparse Approximate Gaussian Process Regression”. In: *Journal of Machine Learning Research* 6, pp. 1939–1959.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
- Rabe-Hesketh, Sophia and Anders Skrondal (2012). *Multilevel and Longitudinal Modeling Using Stata*. 3rd ed. Stata Press. ISBN: 978-1-59718-108-2.
- Raftery, Adrian E., David Madigan, and Jennifer A Hoeting (1997). “Bayesian Model Averaging for Linear Regression Models”. In: *Journal of the American Statistical Association* 92.437, pp. 179–191. DOI: [10.1080/01621459.1997.10473615](https://doi.org/10.1080/01621459.1997.10473615).
- Ramsay, J. O., Hadley Wickham, Spencer Graves, and Giles Hooker (2017). *fda: Functional Data Analysis*. R package version 2.4.7. URL: <https://CRAN.R-project.org/package=fda>.
- Ramsay, James and Bernard W. Silverman (2005). *Functional Data Analysis*. New York: Springer-Verlag. ISBN: 978-1-4757-7107-7. DOI: [10.1007/978-1-4757-7107-7](https://doi.org/10.1007/978-1-4757-7107-7).
- Rasmussen, Carl Edward and Christopher K. I. Williams (2006). *Gaussian Processes for Machine Learning*. The MIT Press. ISBN: 0-262-18253-X. URL: <http://www.gaussianprocess.org/gpml/>.
- Raykar, Vikas C and Ramani Duraiswami (Mar. 2007). “Fast large scale Gaussian process regression using approximate matrix-vector products”. In: *Learning Workshop 2007*. San Juan, Puerto Rico.
- Reed, Michael (1972). *Methods of Modern Mathematical Physics*. Functional Analysis. 1st ed. Academic Press. ISBN: 978-0-323-15500-7.
- Revolution Analytics and Steve Weston (2015). *foreach: Provides Foreach Looping Construct for R*. R package version 1.4.3. URL: <https://CRAN.R-project.org/package=foreach>.
- Robbins, Herbert (1956). “An Empirical Bayes Approach to Statistics”. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Ed. by Jerzy Neyman. Vol. 1: Contributions to the Theory of Statistics. Berkeley, CA: University of California Press, pp. 157–163.
- (1983). “Some Thoughts on Empirical Bayes Estimation”. In: *The Annals of Statistics* 11.3, pp. 713–723.
- Robert, Christian (1995). “Simulation of truncated normal variables”. In: *Statistics and Computing* 5.2, pp. 121–125. DOI: [10.1007/BF00143942](https://doi.org/10.1007/BF00143942).
- (2007). *The Bayesian Choice*. From Decision-Theoretic Foundations to Computational Implementation. New York: Springer-Verlag. ISBN: 978-0-387-95231-4. DOI: [10.1007/0-387-71599-1](https://doi.org/10.1007/0-387-71599-1).
- (2014). “On the Jeffreys-Lindley paradox”. In: *Philosophy of Science* 81.2, pp. 216–232. ARXIV: [1303.5973](https://arxiv.org/abs/1303.5973).
- Robinson, Anthony John (1989). “Dynamic error propagation networks”. PhD thesis. University of Cambridge.

- Rudin, Walter (1987). *Real and Complex Analysis*. 3rd ed. McGraw-Hill Education. ISBN: 978-0-07-100276-9.
- Rue, Håvard, Sara Martino, and Nicolas Chopin (2009). “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392. DOI: [10.1111/j.1467-9868.2008.00700.x](https://doi.org/10.1111/j.1467-9868.2008.00700.x).
- SAS Institute Inc. (2008). *SAS/STAT(R) 9.2 User’s Guide*. 2nd ed. Cary, NC: SAS Institute Inc. ISBN: 978-1-60764-566-5.
- Schoenberg, Isaac J. (1937). “On Certain Metric Spaces Arising From Euclidean Spaces by a Change of Metric and Their Imbedding in Hilbert Space”. In: *Annals of Mathematics* 38.4, pp. 787–793. DOI: [10.2307/1968835](https://doi.org/10.2307/1968835).
- Schölkopf, Bernhard and Alexander J. Smola (2002). *Learning with Kernels*. Support Vector Machines, Regularization, Optimization, and Beyond. The MIT Press. ISBN: 978-0-262-19475-4.
- Schwarz, Gideon (1978). “Estimating the Dimension of a Model”. In: *The Annals of Statistics* 6.2, pp. 461–464. DOI: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136).
- Scott, Steven L. and Hal R. Varian (2014). “Predicting the present with Bayesian structural time series”. In: *International Journal of Mathematical Modelling and Numerical Optimisation* 5.1-2, pp. 4–23. DOI: [10.1504/IJMMNO.2014.059942](https://doi.org/10.1504/IJMMNO.2014.059942).
- Sejdinovic, Dino and Arthur Gretton (2012). *What is an RKHS?* COMP6113 Advanced Topics in Machine Learning: Lectures conducted at University College London. URL: [http://www.gatsby.ucl.ac.uk/~7B~%7Dgretton/coursefiles/RKHS%7B%5C\\_%7DNotes1.pdf](http://www.gatsby.ucl.ac.uk/~7B~%7Dgretton/coursefiles/RKHS%7B%5C_%7DNotes1.pdf).
- Shi, Jian Qing and Yafeng Cheng (2014). **GPFDA**: *Apply Gaussian Process in Functional Data Analysis*. R package version 2.2. URL: <https://CRAN.R-project.org/package=GPFDA>.
- Skorohod, Anatolij Vladimirovič (1974). *Integration in Hilbert Space*. Springer-Verlag. ISBN: 978-3-642-65634-7. DOI: [10.1007/978-3-642-65632-3](https://doi.org/10.1007/978-3-642-65632-3).
- Skrondal, Anders and Sophia Rabe-Hesketh (2004). *Generalized Latent Variable Modeling*. Multilevel, Longitudinal, and Structural Equation Models. Chapman & Hall/CRC. ISBN: 978-1-58488-000-4.
- Sobol, Ilya M (2001). “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”. In: *Mathematics and Computers in Simulation* 55.1–3, pp. 271–280. DOI: [10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6).
- Sriperumbudur, Bharath, Kenji Fukumizu, Arthur Gretton, Aapo Hyvärinen, and Ravant Kumar (2017). “Density Estimation in Infinite Dimensional Exponential Families”. In: *Journal of Machine Learning Research* 18.57, pp. 1–59. ARXIV: [1312.3516 \[math.ST\]](https://arxiv.org/abs/1312.3516).
- Stan Development Team (2016a). **RStan**: *The R Interface to Stan*. R package version 2.14.1. URL: <http://mc-stan.org/>.
- (2016b). **rstanarm**: *Bayesian Applied Regression Modeling via Stan*. R package version 2.13.1. URL: <http://mc-stan.org/>.



- Steinwart, Ingo and Andreas Christmann (2008). *Support Vector Machines*. New York: Springer-Verlag. ISBN: 978-0-387-77241-7. DOI: [10.1007/978-0-387-77242-4](#).
- Steinwart, Ingo, Don Hush, and Clint Scovel (2006). “An Explicit Description of the Reproducing Kernel Hilbert Spaces of Gaussian RBF Kernels”. In: *IEEE Transactions on Information Theory* 52.10, pp. 4635–4643. DOI: [10.1109/TIT.2006.881713](#).
- Sturtz, Sibylle, Uwe Ligges, and Andrew Gelman (2005). “**R2WinBUGS**: A Package for Running WinBUGS from R”. In: *Journal of Statistical Software* 12.3, pp. 1–16. DOI: [10.18637/jss.v012.i03](#).
- Tapia, Richard A. (1971). *The Differentiation and Integration of Nonlinear Operators*. Ed. by Louis B. Rall. DOI: [10.1016/C2013-0-11348-7](#).
- Taylor, Benjamin, Tilman Davies, Barry Rowlingson, and Peter Diggle (2013). “**lgcp**: An R Package for Inference with Spatial and Spatio-Temporal Log-Gaussian Cox Processes”. In: *Journal of Statistical Software* 52.4, pp. 1–40. DOI: [10.18637/jss.v052.i04](#).
- Teh, Yee Whye, David Newman, and Max Welling (2008). “A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation”. In: *Advances in Neural Information Processing Systems 20 (NIPS 2007)*. Ed. by John C. Platt, Daphne Koller, Yoram Singer, and Sam T. Roweis, pp. 1353–1360.
- Thodberg, Hans Henrik (1996). “A review of Bayesian neural networks with an application to near infrared spectroscopy”. In: *IEEE Transactions on Neural Networks* 7.1, pp. 56–72. DOI: [10.1109/72.478392](#).
- Tibshirani, Robert (1996). “Regression Shrinkage and Selection via the Lasso”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 58.1, pp. 267–288. DOI: [10.1111/j.1467-9868.2011.00771.x](#).
- Tibshirani, Robert, Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu (May 2002). “Diagnosis of Multiple Cancer Types by Shrunk Centroids of Gene Expression”. In: *Proceedings of the National Academy of Sciences of the United States of America (PNAS 2002)*. Vol. 99. 10, pp. 6567–6572. DOI: [10.1073/pnas.082099299](#).
- Titsias, Michalis (Apr. 2009). “Variational Learning of Inducing Variables in Sparse Gaussian Processes”. In: *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS 2009), Clearwater Beach, FL*. Ed. by David van Dyk and Max Welling, pp. 567–574.
- Train, Kenneth (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press. ISBN: 978-0-511-80527-1. DOI: [10.1017/CB09780511805271](#).
- van der Vaart, Aad W. and J. Harry van Zanten (2008). “Reproducing kernel Hilbert spaces of Gaussian priors”. In: *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*. Ed. by Bertrand Clarke and Subhashis Ghosal. Beachwood, OH: Institute of Mathematical Statistics, pp. 200–222. DOI: [10.1214/074921708000000156](#).
- Vila, J. P., V. Wagner, and P. Neveu (2000). “Bayesian Nonlinear Model Selection and Neural Networks: A Conjugate Prior Approach”. In: *IEEE Transactions on Neural Networks* 11.2, pp. 265–278. DOI: [10.1109/72.838999](#).
- Wahba, Grace (1990). *Spline Models for Observational Data*. SIAM. ISBN: 978-0-89871-244-5. DOI: [10.1137/1.9781611970128](#).

- Wasserman, Larry (2004). *All of Statistics. A Concise Course in Statistical Inference*. New York: Springer-Verlag. ISBN: 978-0-387-40272-7. DOI: [10.1007/978-0-387-21736-9](#).
- Wassermann, Larry (2006). *All of Nonparametric Statistics*. New York: Springer-Verlag. ISBN: 978-0-387-25145-5. DOI: [10.1007/0-387-30623-4](#).
- Wei, Greg C. G. and Martin A. Tanner (1990). “A Monte Carlo Implementation of the EM Algorithm and the Poor Man’s Data Augmentation Algorithms”. In: *Journal of the American statistical Association* 85.411, pp. 699–704. DOI: [10.2307/2290005](#).
- 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 (NIPS 2000)*. Ed. by Todd K. Leen, Thomas G. Dietterich, and Volker Tresp, pp. 682–688.
- Wood, Simon N. (2017). *Generalized Additive Models. An Introduction with R*. 2nd ed. Chapman & Hall/CRC. ISBN: 978-1-4987-2833-1.
- Wu, C. F. Jeff (1983). “On the Convergence Properties of the EM Algorithm”. In: *The Annals of Statistics* 11.1, pp. 95–103. DOI: [10.1214/aos/1176346060](#).
- Yamamoto, Yutaka (2012). *From Vector Spaces to Function Spaces: Introduction to Functional Analysis with Applications*. SIAM. ISBN: 978-1-61197-230-6. DOI: [10.1137/9781611972313](#).
- Yu, Yaming (2012). “Monotonically overrelaxed EM algorithms”. In: *Journal of Computational and Graphical Statistics* 21.2, pp. 518–537.
- Zafeiriou, Stefanos (Oct. 2012). “Subspace Learning in Krein Spaces: Complete Kernel Fisher Discriminant Analysis with Indefinite Kernels”. In: *Proceedings of the Twelfth European Conference on Computer Vision (ECCV 2012), Florence, Italy*. Ed. by Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid. Berlin, Heidelberg: Springer, pp. 488–501. ISBN: 978-3-642-33764-2. DOI: [10.1007/978-3-642-33765-9\\_35](#).
- Zellner, Arnold (1986). “On Assessing Prior Distributions and Bayesian Regression Analysis with  $g$ -Prior Distributions”. In: *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. New York: Elsevier, pp. 233–243.
- Zhang, Haizhang, Yuesheng Xu, and Jun Zhang (2009). “Reproducing Kernel Banach Spaces for Machine Learning”. In: *Journal of Machine Learning Research* 10, pp. 2741–2775.
- Zhang, Haizhang and Jun Zhang (2012). “Regularized learning in Banach spaces as an optimization problem: representer theorems”. In: *Journal of Global Optimization* 54.2, pp. 235–250. DOI: [10.1007/s10898-010-9575-z](#).
- Zhang, Huamin and Feng Ding (2013). “On the Kronecker Products and Their Applications”. In: *Journal of Applied Mathematics* 296185. DOI: [10.1155/2013/296185](#).
- Zhu, Hongxiao, Fang Yao, and Hao Helen Zhang (2014). “Structured functional additive regression in reproducing kernel Hilbert spaces”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.3, pp. 581–603. DOI: [10.1111/rssb.12036](#).

Zou, Hui and Trevor Hastie (2005). “Regularization and variable selection via the elastic net”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320. DOI: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x).