

To-do list

Contents

1	Preceding chapters	3
	Bibliography	20
	Figures	21
	Tables	22
	Theorems	23
	Definitions	24
	Nomenclature	28
	Abbreviations	29

Haziq Jamil

Department of Statistics

London School of Economics and Political Science

PhD thesis: ‘Regression modelling using Fisher information covariance kernels (I-priors)’

Chapter 1

Preceding chapters

Bibliography

- agresti2000
tutorial 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.
- akaike1973 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.
- albert1993b
ayesian 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.
- alpay1991so
me Alpay, Daniel (1991). “Some remarks on reproducing kernel Krein spaces”. In: *The Rocky Mountain Journal of Mathematics*, pp. 1189–1205.
- alvarez2014
bayesian Alvarez, Ignacio, Jarad Niemi, and Matt Simpson (2014). “Bayesian inference for a covariance matrix”. In: *arXiv preprint arXiv:1408.4050*.
- balakrishna
n1981applied Balakrishnan, Alampallam V (1981). *Applied Functional Analysis*. 2nd ed. Vol. 3. Springer Science & Business Media. DOI: [10.1007/978-1-4612-5865-0](https://doi.org/10.1007/978-1-4612-5865-0).
- Barbieri2004 Barbieri, Maria Maddalena and Jim O Berger (2004). “Optimal predictive model selection”. In: *Annals of Statistics* 32.3, pp. 870–897. ISSN: 00905364. DOI: [10.1214/009053604000000238](https://doi.org/10.1214/009053604000000238).
- bates2014lme4 Bates, Douglas, 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).
- beal2003variational Beal, Matthew James (2003). *Variational algorithms for approximate Bayesian inference*. university of London London.
- beal2003 Beal, Matthew James and Z 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. Bayarri, and Adrian F.M. Smith. Oxford: Oxford University Press, pp. 453–464.
berger2006case	Berger, James et al. (2006). “The case for objective Bayesian analysis”. In: <i>Bayesian analysis</i> 1.3, pp. 385–402.
berger2013statistical	Berger, Jim O (2013). <i>Statistical Decision Theory and Bayesian Analysis</i> . New York: Springer-Verlag. ISBN: 978-0-387-96098-2. DOI: 10.1007/978-1-4757-4286-2 .
bergsma2017	Bergsma, Wicher (2017). “Regression with I-priors”. In: <i>Unpublished manuscript</i> .
berlinet2011reproducing	Berlinet, Alain and Christine Thomas-Agnan (2011). <i>Reproducing Kernel Hilbert Spaces in Probability and Statistics</i> . Springer-Verlag. DOI: 10.1007/978-1-4419-9096-9 .
bernardo2003variational	Bernardo, JM, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AFM Smith, M West, et al. (2003). “The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures”. In: <i>Bayesian statistics</i> 7, pp. 453–464.
bickel2013asymptotic	Bickel, Peter, David Choi, Xiangyu Chang, and Hai Zhang (2013). “Asymptotic normality of maximum likelihood and its variational approximation for stochastic block-models”. In: <i>The Annals of Statistics</i> , pp. 1922–1943.
bishop2006pattern	Bishop, Christopher (2006). <i>Pattern Recognition and Machine Learning</i> . Springer-Verlag.
blei2017variational	Blei, David M, Alp Kucukelbir, and Jon D McAuliffe (2017). “Variational inference: A review for statisticians”. In: <i>Journal of the American Statistical Association</i> just-accepted.
bouboulis2011extension	Bouboulis, Pantelis and Sergios Theodoridis (2011). “Extension of Wirtinger’s calculus to reproducing kernel Hilbert spaces and the complex kernel LMS”. In: <i>IEEE Transactions on Signal Processing</i> 59.3, pp. 964–978.
Breiman1985	Breiman, L and J H Friedman (1985). “Estimating optimal transformations for multiple-regression and correlation”. In: <i>Journal of the American Statistical Association</i> 80.391, pp. 614–619.
breiman2001random	Breiman, Leo (2001). “Random forests”. In: <i>Machine learning</i> 45.1, pp. 5–32.
breslow1993approximate	Breslow, Norman E and David G Clayton (1993). “Approximate inference in generalized linear mixed models”. In: <i>Journal of the American statistical Association</i> 88.421, pp. 9–25.
bunch1991estimability	Bunch, David S (1991). “Estimability in the multinomial probit model”. In: <i>Transportation Research Part B: Methodological</i> 25.1, pp. 1–12.

buerkner2016brms	Bürkner, Paul-Christian (2017). “brms: An R Package for Bayesian Multilevel Models Using Stan”. In: <i>Journal of Statistical Software</i> 80.1, pp. 1–28. DOI: 10.18637/jss.v080.i01 .
butler1982computationally	Butler, John S and Robert Moffitt (1982). “A computationally efficient quadrature procedure for the one-factor multinomial probit model”. In: <i>Econometrica: Journal of the Econometric Society</i> , pp. 761–764.
cannings2017random	Cannings, Timothy I and Richard J Samworth (2017). “Random-projection ensemble classification”. In: <i>Journal of the Royal Statistical Society: Series B (Statistical Methodology) with discussion</i> 79.4, pp. 959–1035.
carlin1995bayesian	Carlin, Bradley P and Siddhartha Chib (1995). “Bayesian model choice via Markov chain Monte Carlo methods”. In: <i>Journal of the Royal Statistical Society. Series B (Methodological)</i> , pp. 473–484.
carpenter2016stan	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: <i>Journal of Statistical Software, Articles</i> 76.1, pp. 1–32. DOI: 10.18637/jss.v076.i01 .
casella1985introduction	Casella, George (1985). “An introduction to empirical Bayes data analysis”. In: <i>The American Statistician</i> 39.2, pp. 83–87.
casella2002statistical	Casella, George and Roger L Berger (2002). <i>Statistical inference</i> . Vol. 2. Duxbury Pacific Grove, CA.
casella2009consistency	Casella, George, F Javier Girón, M Lina Martínez, and Elias Moreno (2009). “Consistency of Bayesian procedures for variable selection”. In: <i>The Annals of Statistics</i> , pp. 1207–1228.
Casella2006	Casella, George and Elías Moreno (2006). “Objective Bayesian Variable Selection”. In: <i>Journal of the American Statistical Association</i> 101.473, pp. 157–167. ISSN: 0162-1459. DOI: 10.1198/016214505000000646 .
chan1997maximum	Chan, Jennifer SK and Anthony YC Kuk (1997). “Maximum likelihood estimation for probit-linear mixed models with correlated random effects”. In: <i>Biometrics</i> , pp. 86–97.
chen2011single	Chen, Dong, Peter Hall, and Hans-Georg Müller (2011). “Single and Multiple Index Functional Regression Models with Nonparametric Link”. In: <i>The Annals of Statistics</i> 39.3, pp. 1720–1747. DOI: 10.1214/11-AOS882 .
chen2017use	Chen, Yen-Chi, Y Samuel Wang, and Elena A Eroshova (2017). “On the use of bootstrap with variational inference: Theory, interpretation, and a two-sample test example”. In: <i>arXiv preprint arXiv:1711.11057</i> .

cheng2017variational	Cheng, Ching-An and Byron Boots (2017). “Variational Inference for Gaussian Process Models with Linear Complexity”. In: <i>Advances in Neural Information Processing Systems</i> , pp. 5190–5200.
Chipman2001	Chipman, Hugh, Edward I George, and Robert E McCulloch (2001). “The Practical Implementation of Bayesian Model Selection”. In: <i>IMS Lecture Notes - Monograph Series</i> 38, pp. 65–134.
chopin2011fast	Chopin, Nicolas (2011). “Fast simulation of truncated Gaussian distributions”. In: <i>Statistics and Computing</i> 21.2, pp. 275–288.
cohen2002	Cohen, S (2002). “Champs localement auto-similaires”. In: <i>Lois d’échelle, fractales et ondelettes</i> . Ed. by Patrice Abry, Paulo Gonçalves, and Jacques Lévy Véhel. Vol. 1. Hermès Sciences Publications.
damien2001sampling	Damien, Paul and Stephen G Walker (2001). “Sampling truncated normal, beta, and gamma densities”. In: <i>Journal of Computational and Graphical Statistics</i> 10.2, pp. 206–215.
dansie1985parameter	Dansie, BR (1985). “Parameter estimability in the multinomial probit model”. In: <i>Transportation Research Part B: Methodological</i> 19.6, pp. 526–528.
davidian1995nonlinear	Davidian, Marie and David M Giltinan (1995). <i>Nonlinear Models for Repeated Measurement Data</i> . Chapman and Hall/CRC.
davison2003statistical	Davison, Anthony Christopher (2003). <i>Statistical models</i> . Vol. 11. Cambridge University Press.
dean1999design	Dean, Angela and Daniel Voss (1999). <i>Design and analysis of experiments</i> . Vol. 1. Springer.
dellaportas2002bayesian	Dellaportas, Petros, Jonathan J Forster, and Ioannis Ntzoufras (2002). “On Bayesian model and variable selection using MCMC”. In: <i>Statistics and Computing</i> 12.1, pp. 27–36.
dempster1977maximum	Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm”. In: <i>Journal of the royal statistical society. Series B (methodological)</i> , pp. 1–38.
denwood2016runjags	Denwood, Matthew (2016). “ runjags : An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS”. In: <i>Journal of Statistical Software</i> 71.9, pp. 1–25. DOI: 10.18637/jss.v071.i09 .
deterding1989speaker	Deterding, David Henry (1989). “Speaker normalization for automatic speech recognition”. PhD thesis. University of Cambridge.

diggle2013spatial	Diggle, Peter J, Paula Moraga, Barry Rowlingson, and Benjamin M Taylor (2013). “Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm”. In: <i>Statistical Science</i> , pp. 542–563.
diggle2005nonparametric	Diggle, Peter, Pingping Zheng, and Peter Durr (2005). “Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK”. In: <i>Journal of the Royal Statistical Society: Series C (Applied Statistics)</i> 54.3, pp. 645–658.
duane1987hybrid	Duane, Simon, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth (1987). “Hybrid monte carlo”. In: <i>Physics letters B</i> 195.2, pp. 216–222.
durrande2013anova	Durrande, Nicolas, David Ginsbourger, Olivier Roustant, and Laurent Carraro (2013). “ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis”. In: <i>Journal of Multivariate Analysis</i> 115, pp. 57–67.
duvenaud2014automatic	Duvenaud, David (2014). “Automatic model construction with Gaussian processes”. PhD thesis. University of Cambridge.
eddelbuettel2011rcpp	Eddelbuettel, Dirk and Romain Francois (2011). “ Rcpp : Seamless R and C++ Integration”. In: <i>Journal of Statistical Software</i> 40.8, pp. 1–18. DOI: 10.18637/jss.v040.i08 .
efron1978assessing	Efron, Bradley and David V Hinkley (1978). “Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information”. In: <i>Biometrika</i> 65.3, pp. 457–483.
embrechts2002selfsimilar	Embrechts, Paul and Makoto Maejima (2002). <i>Selfsimilar Processes. Princeton series in applied mathematics</i> . Princeton University Press, Princeton, NJ.
fernandez2001benchmark	Fernandez, Carmen, Eduardo Ley, and Mark FJ Steel (2001). “Benchmark priors for Bayesian model averaging”. In: <i>Journal of Econometrics</i> 100.2, pp. 381–427.
ferraty2006nonparametric	Ferraty, Frédéric and Philippe Vieu (2006). <i>Nonparametric Functional Data Analysis</i> . 1st. Springer-Verlag. DOI: 10.1007/0-387-36620-2 .
ra1922mathematical	Fisher, RA (1922). “On the mathematical foundations of theoretical statistics”. In: <i>Phil. Trans. R. Soc. Lond. A</i> 222.594-604, pp. 309–368.
Fouskakis2008	Fouskakis, Dimitris and David Draper (2008). “Comparing Stochastic Optimization Methods for Variable Selection in Binary Outcome Prediction, With Application to Health Policy”. In: <i>Journal of the American Statistical Association</i> 103.484, pp. 1367–1381. ISSN: 0162-1459. DOI: 10.1198/016214508000001048 .
fowlkes2004spectral	Fowlkes, Charless, Serge Belongie, Fan Chung, and Jitendra Malik (2004). “Spectral grouping using the Nystrom method”. In: <i>IEEE transactions on pattern analysis and machine intelligence</i> 26.2, pp. 214–225.

fowlkes2001 efficient	Fowlkes, Charless, Serge Belongie, and Jitendra Malik (2001). “Efficient Spatiotemporal Grouping Using the Nyström Method”. In: <i>Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)</i> . Vol. 1, pp. 231–238. DOI: 10.1109/CVPR.2001.990481 .
friedman2001 elements	Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2001). <i>The elements of statistical learning</i> . Vol. 1. Springer series in statistics New York.
George1993	George, Edward I and Robert E McCulloch (1993). “Variable Selection Via Gibbs Sampling”. In: <i>Journal of the American Statistical Association</i> 88.423, pp. 881–889. URL: http://www.jstor.org/stable/2290777?seq=1#page_scan_tab_contents .
geweke1989b bayesian	Geweke, John (1989). “Bayesian inference in econometric models using Monte Carlo integration”. In: <i>Econometrica: Journal of the Econometric Society</i> , pp. 1317–1339.
geweke1991e efficient	— (1991). <i>Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities</i> .
geweke1996v variable	— (1996). “Variable selection and model comparison in regression”. In: <i>In Bayesian Statistics 5</i> .
geweke1994a alternative	Geweke, John, Michael Keane, and David Runkle (1994). “Alternative computational approaches to inference in the multinomial probit model”. In: <i>The review of economics and statistics</i> , pp. 609–632.
girolami2006 variational	Girolami, Mark and Simon Rogers (2006). “Variational Bayesian Multinomial Probit Regression with Gaussian Process Priors”. In: <i>Neural Computation</i> 18.8, pp. 1790–1817.
goia2014some	Goia, Aldo and Philippe Vieu (2014). “Some Advances in Semiparametric Functional Data Modelling”. In: <i>Contributions in Infinite-Dimensional Statistics and Related Topics</i> . Societa Editrice Esculapio, pp. 135–141. DOI: 10.15651/978-88-748-8763-7 .
gretton2005 measuring	Gretton, Arthur, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf (2005). “Measuring Statistical Dependence with Hilbert-Schmidt Norms”. In: <i>Proceedings of the 16th International Conference of Algorithmic Learning Theory</i> . Ed. by Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita. Springer-Verlag, pp. 63–77. DOI: 10.1007/11564089_7 .
groves1969 note	Groves, Theodore and Thomas Rothenberg (1969). “A note on the expected value of an inverse matrix”. In: <i>Biometrika</i> 56.3, pp. 690–691.
gu2013smoothing	Gu, Chong (2013). <i>Smoothing spline ANOVA models</i> . Vol. 297. Springer Science & Business Media.

guvenir1997 supervised	Guvenir, H Altay, Burak Acar, Gulsen Demiroz, and Ayhan Cekin (1997). “A supervised machine learning algorithm for arrhythmia analysis”. In: <i>Computers in Cardiology 1997</i> . IEEE, pp. 433–436.
hajivassiliou1998metho d	Hajivassiliou, Vassilis A and Daniel L McFadden (1998). “The method of simulated scores for the estimation of LDV models”. In: <i>Econometrica</i> , pp. 863–896.
hajivassiliou1996simul ation	Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud (1996). “Simulation of multi-variate normal rectangle probabilities and their derivatives theoretical and computational results”. In: <i>Journal of econometrics</i> 72.1-2, pp. 85–134.
hall2011asy mptotic	Hall, Peter, Tung Pham, Matt P Wand, Shen SJ Wang, et al. (2011). “Asymptotic normality and valid inference for Gaussian variational approximation”. In: <i>The Annals of Statistics</i> 39.5, pp. 2502–2532.
hastie1986	Hastie, Trevor and Robert Tibshirani (Aug. 1986). “Generalized Additive Models”. In: <i>Statist. Sci.</i> 1.3, pp. 297–310. DOI: 10.1214/ss/1177013604 . URL: https://doi.org/10.1214/ss/1177013604 .
hein2004ker nels	Hein, Matthias and Olivier Bousquet (2004). “Kernels, associated structures and generalizations”. In: <i>Max-Planck-Institut fuer biologische Kybernetik, Technical Report</i> .
hensman2013 gaussian	Hensman, James, Nicolo Fusi, and Neil D Lawrence (2013). “Gaussian processes for big data”. In: <i>arXiv preprint arXiv:1309.6835</i> .
hoerl1970ri dge	Hoerl, Arthur E and Robert W Kennard (1970). “Ridge regression: Biased estimation for nonorthogonal problems”. In: <i>Technometrics</i> 12.1, pp. 55–67.
hoeting1999 bayesian	Hoeting, Jennifer A, David Madigan, Adrian E Raftery, and Chris T Volinsky (1999). “Bayesian model averaging: a tutorial”. In: <i>Statistical science</i> , pp. 382–401.
hofmann2008 kernel	Hofmann, Thomas, Bernhard Schölkopf, and Alexander J Smola (2008). “Kernel methods in machine learning”. In: <i>The annals of statistics</i> , pp. 1171–1220.
itzykson199 1statistica l	Itzykson, Claude and Jean Michel Drouffe (1991). <i>Statistical Field Theory: Volume 2, Strong Coupling, Monte Carlo Methods, Conformal Field Theory and Random Systems</i> . Cambridge University Press.
jamil2017ip rior	Jamil, Haziq (2017). iprior : <i>Regression Modelling using I-Priors</i> . R package version 0.7.1. URL: https://cran.r-project.org/web/packages/iprior .
jamil2018ri priorBVS	— (2018a). ipriorBVS : <i>Bayesian Variable Selection using I-priors</i> . R package version 0.1.1. URL: https://github.com/haziqj/ipriorBVS .
jamil2018ri probit	— (2018b). iprobit : <i>Binary and Multinomial Probit Regression with I-priors</i> . R package version 0.1.1. URL: https://github.com/haziqj/iprobit .

jamil2017	Jamil, Haziq and Wicher Bergsma (2017). “iprior: An R Package for Regression Modelling using I-priors”. In: <i>Manuscript in submission</i> .
jaynes1957a	Jaynes, Edwin T (1957a). “Information Theory and Statistical Mechanics”. In: <i>Physical Review</i> 106.4, p. 620.
jaynes1957b	— (1957b). “Information Theory and Statistical Mechanics II”. In: <i>Physical Review</i> 108.2, p. 171.
jaynes2003probability	— (2003). <i>Probability Theory: The Logic of Science</i> . Cambridge University Press.
jeffreys1946invariant	Jeffreys, Harold (1946). “An invariant form for the prior probability in estimation problems”. In: <i>Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences</i> , pp. 453–461.
jordan1999introduction	Jordan, Michael I, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul (1999). “An introduction to variational methods for graphical models”. In: <i>Machine learning</i> 37.2, pp. 183–233.
kadane2011principles	Kadane, Joseph B (2011). <i>Principles of uncertainty</i> . CRC Press.
gptk	Kalaitzis, Alfredo, Antti Honkela, Pei Gao, and Neil D Lawrence (2014). gptk : <i>Gaussian Processes Tool-Kit</i> . R package version 1.08. URL: https://CRAN.R-project.org/package=gptk .
kalaitzis2011simple	Kalaitzis, Alfredo and Neil D Lawrence (May 2011). “A Simple Approach to Ranking Differentially Expressed Gene Expression Time Courses through Gaussian Process Regression”. In: <i>BMC Bioinformatics</i> 12.1, p. 180. DOI: 10.1186/1471-2105-12-180 .
kammar2016	Kammar, Ohad (2016). <i>A note on Fréchet differentiation under Lebesgue integrals</i> . URL: https://www.cs.ox.ac.uk/people/ohad.kammar/notes/kammar-a-note-on-frechet-differentiation-under-lebesgue-integrals.pdf .
zeileis2004kernlab	Karatzoglou, Alexandros, Alexandros Smola, Kurt Hornik, and Achim Zeileis (2004). “ kernlab - An S4 Package for Kernel Methods in R”. In: <i>Journal of Statistical Software</i> 11.9, pp. 1–20. DOI: 10.18637/jss.v011.i09 .
kass1995bayes	Kass, Robert E and Adrian E Raftery (1995). “Bayes factors”. In: <i>Journal of the american statistical association</i> 90.430, pp. 773–795.
keane1994computationally	Keane, Michael P (1994). “A computationally practical simulation estimator for panel data”. In: <i>Econometrica: Journal of the Econometric Society</i> , pp. 95–116.
Keane1992	Keane, Michael P. (1992). “A Note on Identification in the Multinomial Probit Model”. In: <i>Journal of Business & Economic Statistics</i> 10.2, pp. 193–200. ISSN: 0735-0015. DOI: 10.2307/1391677 . URL: http://www.jstor.org/stable/1391677 http://www.jstor.org/stable/pdfplus/1391677.pdf?acceptTC=true .

keane1994solution	Keane, Michael P and Kenneth I Wolpin (1994). “The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence”. In: <i>the Review of economics and statistics</i> , pp. 648–672.
kenward1987method	Kenward, Michael G. (1987). “A Method for Comparing Profiles of Repeated Measurements”. In: <i>Journal of the Royal Statistical Society C (Applied Statistics)</i> 36.3, pp. 296–308. DOI: 10.2307/2347788 .
kimeldorf1970correspondence	Kimeldorf, George S and Grace Wahba (1970). “A correspondence between Bayesian estimation on stochastic processes and smoothing by splines”. In: <i>The Annals of Mathematical Statistics</i> 41.2, pp. 495–502.
kokoszka2017introduction	Kokoszka, Piotr and Matthew Reimherr (2017). <i>Introduction to functional data analysis</i> . CRC Press.
kree1974produits	Krée, Paul (1974). “Produits tensoriels complétés d’espaces de Hilbert”. In: <i>Séminaire Paul Krée</i> 1.7, pp. 1974–1975.
caret	Kuhn, Max et al. (2017). caret : <i>Classification and Regression Training</i> . R package version 6.0–77. URL: https://CRAN.R-project.org/package=caret .
kuo2010decompositions	Kuo, F, I Sloan, G Wasilkowski, and Henryk Woźniakowski (2010). “On decompositions of multivariate functions”. In: <i>Mathematics of computation</i> 79.270, pp. 953–966.
Kuo1998	Kuo, L and B Mallick (1998). “Variable selection for regression models”. In: <i>Sankhya: The Indian Journal of Statistics, Series B</i> 60.1, pp. 65–81.
kuss2005assessing	Kuss, Malte and Carl Edward Rasmussen (2005). “Assessing approximate inference for binary Gaussian process classification”. In: <i>Journal of machine learning research</i> 6.Oct, pp. 1679–1704.
kyung2010penalized	Kyung, Minjung, Jeff Gill, Malay Ghosh, George Casella, et al. (2010). “Penalized regression, standard errors, and Bayesian lassos”. In: <i>Bayesian Analysis</i> 5.2, pp. 369–411.
lange1995quasi	Lange, Kenneth (1995). “A quasi-Newton acceleration of the EM algorithm”. In: <i>Statistica sinica</i> , pp. 1–18.
lee2003gene	Lee, Kyeong Eun, Naijun Sha, Edward R Dougherty, Marina Vannucci, and Bani K Mallick (2003). “Gene selection: a Bayesian variable selection approach”. In: <i>Bioinformatics</i> 19.1, pp. 90–97.
lian2014series	Lian, Heng and Gaorong Li (2014). “Series Expansion for Functional Sufficient Dimension Reduction”. In: <i>Journal of Multivariate Analysis</i> 124.C, pp. 150–165. DOI: 10.1016/j.jmva.2013.10.019 .

liang2008mixtures	Liang, Feng, Rui Paulo, German Molina, Merlise A Clyde, and Jim O Berger (2008). “Mixtures of g Priors for Bayesian Variable Selection”. In: <i>Journal of the American Statistical Association</i> 103.481, pp. 410–423. DOI: 10.1198/01621450700001337 .
lifshits2012lectures	Lifshits, Mikhail (2012). “Lectures on Gaussian processes”. In: <i>Lectures on Gaussian Processes</i> . Springer-Verlag, pp. 1–117. DOI: 10.1007/978-3-642-24939-6 .
liu1998parameter	Liu, Chuanhai, Donald B Rubin, and Ying Nian Wu (1998). “Parameter expansion to accelerate EM: The PX-EM algorithm”. In: <i>Biometrika</i> 85.4, pp. 755–770.
lloyd2015variational	Lloyd, Chris, Tom Gunter, Michael Osborne, and Stephen Roberts (2015). “Variational inference for Gaussian process modulated Poisson processes”. In: <i>International Conference on Machine Learning</i> , pp. 1814–1822.
louppe2014understanding	Louppe, Gilles (Oct. 2014). “Understanding Random Forests: From Theory to Practice”. arXiv:1407.7502. PhD thesis. University of Liege, Belgium.
lunn2000winbugs	Lunn, David J., Andrew Thomas, Nicky Best, and David Spiegelhalter (Oct. 2000). “WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility”. In: <i>Statistics and Computing</i> 10.4, pp. 325–337. DOI: 10.1023/A:1008929526011 .
macdonald2015gpfit	MacDonald, Blake, Pritam Ranjan, and Hugh Chipman (2015). “ GPfit : An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs”. In: <i>Journal of Statistical Software</i> 64.12, pp. 1–23. DOI: 10.18637/jss.v064.i12 .
madigan1994model	Madigan, David and Adrian E Raftery (1994). “Model selection and accounting for model uncertainty in graphical models using Occam’s window”. In: <i>Journal of the American Statistical Association</i> 89.428, pp. 1535–1546.
mallows1973some	Mallows, Colin L (1973). “Some comments on C p”. In: <i>Technometrics</i> 15.4, pp. 661–675.
mandelbrot1968fractional	Mandelbrot, Benoit B and John W Van Ness (1968). “Fractional Brownian motions, fractional noises and applications”. In: <i>SIAM review</i> 10.4, pp. 422–437.
marsaglia2000ziggurat	Marsaglia, George and Wai Wan Tsang (2000). “The ziggurat method for generating random variables”. In: <i>Journal of statistical software</i> 5.8, pp. 1–7.
mary2003hilbertian	Mary, Xavier (2003). “Hilbertian subspaces, subdualities and applications”. PhD thesis. INSA de Rouen.
mccullagh1989	McCullagh, P. and John A. Nelder (1989). <i>Generalized Linear Models</i> . 2nd. Chapman & Hall/CRC Press.
mcculloch2000bayesian	McCulloch, Robert E, Nicholas G Polson, and Peter E Rossi (2000). “A Bayesian analysis of the multinomial probit model with fully identified parameters”. In: <i>Journal of econometrics</i> 99.1, pp. 173–193.

- mcculloch1994exact McCulloch, Robert and Peter E Rossi (1994). “An exact likelihood analysis of the multinomial probit model”. In: *Journal of Econometrics* 64.1, pp. 207–240.
- McDonald1973 McDonald, Gary C and Richard C Schwing (1973). “Instabilities of regression estimates relating air pollution to mortality”. In: *Technometrics* 15.3, pp. 463–481.
- mclachlan2007algorithm McLachlan, Geoffrey and Thriyambakam Krishnan (2007). *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons.
- meila2003data 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.
- meng1993maximum 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.
- meng1997algorithm 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.
- micchelli2006universal Micchelli, Charles A, Yuesheng Xu, and Haizhang Zhang (2006). “Universal kernels”. In: *Journal of Machine Learning Research* 7.Dec, pp. 2651–2667.
- dosnow 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>.
- milller2002subset Miller, Alan (2002). *Subset selection in regression*. CRC Press.
- minka2001expectation Minka, Thomas P (2001). “Expectation propagation for approximate Bayesian inference”. In: *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., pp. 362–369.
- mitchell1988bayesian 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.
- neal1999 Neal, Radford M. (1999). “Regression and Classification using Gaussian Process Priors”. In: *Bayesian Statistics*. Ed. by J M Bernardo, J O Berger, A P Dawid, and A F M Smith. Vol. 6. Oxford University Press. (with discussion), pp. 475–501.
- neal2011mcmc Neal, Radford M et al. (2011). “MCMC using Hamiltonian dynamics”. In: *Handbook of Markov Chain Monte Carlo* 2.11.
- nobile1998hybrid 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.

Ntzoufras2008	Ntzoufras, Ioannis (2011). <i>Bayesian Modeling Using WinBUGS</i> . Wiley, pp. 389–433. DOI: 10.1002/9780470434567.ch11 .
O'Hara2009	O'Hara, R B and M J Sillanpää (2009). “A review of Bayesian variable selection methods: what, how and which”. In: <i>Bayesian Analysis</i> 4.1, pp. 85–117. ISSN: 1936-0975. DOI: 10.1214/09-BA403 . URL: http://projecteuclid.org/euclid.ba/1340370391 .
ong2004learning	Ong, Cheng Soon, Xavier Mary, Stéphane Canu, and Alexander J Smola (2004). “Learning with non-positive kernels”. In: <i>Proceedings of the twenty-first international conference on Machine learning</i> . ACM, p. 81.
ormerod2017variational	Ormerod, John T, Chong You, Samuel Müller, et al. (2017). “A variational Bayes approach to variable selection”. In: <i>Electronic Journal of Statistics</i> 11.2, pp. 3549–3594.
jmcm	Pan, Jianxin and Yi Pan (2017). “ jmcm : An R Package for Joint Mean-Covariance Modeling of Longitudinal Data”. In: <i>Journal of Statistical Software</i> 82.1, pp. 1–29.
park2008bayesian	Park, Trevor and George Casella (2008). “The bayesian lasso”. In: <i>Journal of the American Statistical Association</i> 103.482, pp. 681–686.
pawitan2001all	Pawitan, Yudi (2001). <i>In all likelihood: statistical modelling and inference using likelihood</i> . Oxford University Press.
petersen2008matrix	Petersen, Kaare Brandt and Michael Syskind Pedersen (2008). “The matrix cookbook”. In: <i>Technical University of Denmark</i> 7.15, p. 510.
pinheiro2000mixed	Pinheiro, José C and Douglas M Bates (2000). <i>Mixed-Effects Models in S and S-plus</i> . Springer-Verlag. DOI: 10.1007/b98882 .
nlme	Pinheiro, Joséo, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team (2017). nlme : <i>Linear and Nonlinear Mixed Effects Models</i> . R package version 3.1-131. URL: https://CRAN.R-project.org/package=nlme .
plummer2003jags	Plummer, Martyn (2003). “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling”. In: <i>Proceedings of the 3rd International Workshop on Distributed Statistical Computing</i> . Vol. 124. Vienna, Austria, p. 125.
quinonero2005unifying	Quiñonero-Candela, Joaquin and Carl Edward Rasmussen (Dec. 2005). “A Unifying View of Sparse Approximate Gaussian Process Regression”. In: <i>Journal of Machine Learning Research</i> 6, pp. 1939–1959.
R	R Core Team (2017). <i>R: A Language and Environment for Statistical Computing</i> . Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/ .
skrondal2012multilevel	Rabe-Hesketh, Sophia and Anders Skrondal (2012). <i>Multilevel and Longitudinal Modeling Using Stata, 3rd Edition</i> . 3rd. StataCorp LP. URL: https://EconPapers.repec.org/RePEc:tsj:spbook:mimus2 .

raftery1997 bayesian	Raftery, Adrian E, David Madigan, and Jennifer A Hoeting (1997). “Bayesian model averaging for linear regression models”. In: <i>Journal of the American Statistical Association</i> 92.437, pp. 179–191.
fda	Ramsay, J. O., Hadley Wickham, Spencer Graves, and Giles Hooker (2017). fda : <i>Functional Data Analysis</i> . R package version 2.4.7. URL: https://CRAN.R-project.org/package=fda .
ramsay2005f unctional	Ramsay, James and B. W. Silverman (2005). <i>Functional Data Analysis</i> . Springer.
rasmussen20 06gaussian	Rasmussen, Carl Edward and Christopher K I Williams (2006). <i>Gaussian Processes for Machine Learning</i> . The MIT Press.
raykar2007f ast	Raykar, Vikas C and Ramani Duraiswami (2007). “Fast large scale Gaussian process regression using approximate matrix-vector products”. In:
reed1972met hods	Reed, Michael and Barry Simon (1972). <i>Methods of mathematical physics I: Functional analysis</i> .
foreach	Revolution Analytics and Steve Weston (2015). foreach : <i>Provides Foreach Looping Construct for R</i> . R package version 1.4.3. URL: https://CRAN.R-project.org/package=foreach .
robbins1956 empirical	Robbins, Herbert (1956). <i>An empirical Bayes approach to statistics</i> . Tech. rep. COLUMBIA UNIVERSITY New York City United States.
robbins1985 jerzy	— (1985). “Jerzy Neyman Memorial Lecture Some Thoughts on Empirical Bayes Estimation”. In: <i>Herbert Robbins Selected Papers</i> . Springer, pp. 87–97.
robert2007b ayesian	Robert, Christian (2007). <i>The Bayesian choice: from decision-theoretic foundations to computational implementation</i> . Springer Science & Business Media.
robert1995s imulation	Robert, Christian P (1995). “Simulation of truncated normal variables”. In: <i>Statistics and computing</i> 5.2, pp. 121–125.
robinson198 9dynamic	Robinson, Anthony John (1989). “Dynamic error propagation networks”. PhD thesis. University of Cambridge.
rudin1987re al	Rudin, Walter (1987). <i>Real and complex analysis</i> . Tata McGraw-Hill Education.
SAS2008	SAS Institute Inc. (2008). <i>SAS/STAT(R) 9.2 User’s Guide</i> . 2nd. Cary, NC: SAS Institute Inc. ISBN: 978-1-60764-566-5. URL: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_reg_sect055.htm .
schoenberg1 937	Schoenberg, Isaac J (1937). “On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space”. In: <i>Annals of mathematics</i> , pp. 787–793.

scholkopf2002learning	Schölkopf, Bernhard and Alexander J Smola (2002). <i>Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond</i> . MIT Press.
schwarz1978estimating	Schwarz, Gideon et al. (1978). “Estimating the dimension of a model”. In: <i>The annals of statistics</i> 6.2, pp. 461–464.
scott2014predicting	Scott, Steven L and Hal R Varian (2014). “Predicting the present with bayesian structural time series”. In: <i>International Journal of Mathematical Modelling and Numerical Optimisation</i> 5.1-2, pp. 4–23.
sejdinovic2012	Sejdinovic, Dino and Arthur Gretton (2012). “Lecture notes: What is an RKHS?” In: <i>COMP213 Advanced Topics in Machine Learning. Lecture conducted at University College London</i> , pp. 1–24. URL: http://www.gatsby.ucl.ac.uk/~%7B~%7Dgretton/coursefiles/RKHS%7B%5C_%7DNotes1.pdf .
gpfd	Shi, Jian Qing and Yafeng Cheng (2014). <i>GPFD: Apply Gaussian Process in Functional Data Analysis</i> . R package version 2.2. URL: https://CRAN.R-project.org/package=GPFD .
skorohod2012integration	Skorohod, Anatolij Vladimirovič (2012). <i>Integration in Hilbert Space</i> . Vol. 79. Springer-Verlag. DOI: 10.1007/978-3-642-65632-3 .
skrondal2004generalized	Skrondal, Anders and Sophia Rabe-Hesketh (2004). <i>Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models</i> . Crc Press.
sobol2001global	Sobol, Ilya M (2001). “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”. In: <i>Mathematics and computers in simulation</i> 55.1-3, pp. 271–280.
sriperumbudur2013density	Sriperumbudur, Bharath, Kenji Fukumizu, Arthur Gretton, Aapo Hyvärinen, and Ravent Kumar (2013). “Density estimation in infinite dimensional exponential families”. In: <i>arXiv preprint arXiv:1312.3516</i> .
rstan	Stan Development Team (2016a). RStan: The R Interface to Stan . R package version 2.14.1. URL: http://mc-stan.org/ .
rstanarm	— (2016b). rstanarm: Bayesian Applied Regression Modeling via Stan . R package version 2.13.1. URL: http://mc-stan.org/ .
steinwart2008support	Steinwart, Ingo and Andreas Christmann (2008). <i>Support vector machines</i> . Springer Science & Business Media.
steinwart2006explicit	Steinwart, Ingo, Don Hush, and Clint Scovel (2006). “An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels”. In: <i>IEEE Transactions on Information Theory</i> 52.10, pp. 4635–4643.

sturtz2005r 2winbugs	Sturtz, Sibylle, Uwe Ligges, and Andrew Gelman (2005). “ R2WinBUGS : A Package for Running WinBUGS from R”. In: <i>Journal of Statistical Software</i> 12.3, pp. 1–16. DOI: 10.18637/jss.v012.i03 .
tapia1971di ff	Tapia, R A (1971). <i>The differentiation and integration of nonlinear operators</i> . Ed. by Louis B Rall.
taylor2013l gcp	Taylor, Benjamin M, Tilman M Davies, Barry S Rowlingson, Peter J Diggle, et al. (2013). “lgcp: an R package for inference with spatial and spatio-temporal log-Gaussian Cox processes”. In: <i>Journal of Statistical Software</i> 52.4, pp. 1–40.
teh2007coll apsed	Teh, Yee W, David Newman, and Max Welling (2007). “A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation”. In: <i>Advances in neural information processing systems</i> , pp. 1353–1360.
thodberg199 6review	Thodberg, Hans Henrik (1996). “A Review of Bayesian Neural Networks with an Application to near Infrared Spectroscopy”. In: <i>IEEE Transactions on Neural Networks</i> 7.1, pp. 56–72. DOI: 10.1109/72.478392 .
tibshirani1 996regressi on	Tibshirani, Robert (1996). “Regression shrinkage and selection via the lasso”. In: <i>Journal of the Royal Statistical Society. Series B (Methodological)</i> , pp. 267–288.
tibshirani2 002diagnosi s	Tibshirani, Robert, Trevor Hastie, Balasubramanian Narasimhan, and Gilbert Chu (2002). “Diagnosis of multiple cancer types by shrunken centroids of gene expression”. In: <i>Proceedings of the National Academy of Sciences</i> 99.10, pp. 6567–6572.
titsias2009 variational	Titsias, Michalis (2009). “Variational learning of inducing variables in sparse Gaussian processes”. In: <i>Artificial Intelligence and Statistics</i> , pp. 567–574.
train2009di crete	Train, Kenneth E (2009). <i>Discrete choice methods with simulation</i> . Cambridge university press.
van2008repr oducing	van der Vaart, Aad W and van Zanten (2008). “Reproducing kernel Hilbert spaces of Gaussian priors”. In: <i>Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh</i> . Institute of Mathematical Statistics, pp. 200–222.
vila2000bay esian	Vila, J. P., V. Wagner, and P. Neveu (Mar. 2000). “Bayesian Nonlinear Model Selection and Neural Networks: A Conjugate Prior Approach”. In: <i>IEEE Transactions on Neural Networks</i> 11.2, pp. 265–278. DOI: 10.1109/72.838999 .
wahba1990sp line	Wahba, Grace (1990). <i>Spline models for observational data</i> . Vol. 59. Siam.
wasserman20 13all	Wasserman, Larry (2013). <i>All of statistics: a concise course in statistical inference</i> . Springer Science & Business Media.
wassermann2 006all	Wassermann, Larry (2006). “All of nonparametric statistics”. In: <i>New York</i> .

wei1990monte	Wei, Greg CG and Martin A Tanner (1990). “A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms”. In: <i>Journal of the American statistical Association</i> 85.411, pp. 699–704.
williams2001using	Williams, Christopher K I and Matthias Seeger (2001). “Using the Nyström Method to Speed Up Kernel Machines”. In: <i>Advances in Neural Information Processing Systems 13</i> . The MIT Press, pp. 682–688.
mgcv	Wood, Simon N (2017). <i>Generalized Additive Models: An Introduction with R</i> . 2nd. Chapman and Hall/CRC.
wu1983convergence	Wu, CF Jeff (1983). “On the convergence properties of the EM algorithm”. In: <i>The Annals of statistics</i> , pp. 95–103.
yamamoto2012vector	Yamamoto, Yutaka (2012). <i>From vector spaces to function spaces: introduction to functional analysis with applications</i> . Vol. 127. SIAM.
yu2012monotonically	Yu, Yaming (2012). “Monotonically overrelaxed EM algorithms”. In: <i>Journal of Computational and Graphical Statistics</i> 21.2, pp. 518–537.
zafeiriou2012subspace	Zafeiriou, Stefanos (2012). “Subspace learning in krein spaces: Complete kernel fisher discriminant analysis with indefinite kernels”. In: <i>European Conference on Computer Vision</i> . Springer, pp. 488–501.
zellner1986assessing	Zellner, Arnold (1986). “On assessing prior distributions and Bayesian regression analysis with g-prior distributions”. In: <i>Bayesian inference and decision techniques</i> .
zhang2009reproducing	Zhang, Haizhang, Yuesheng Xu, and Jun Zhang (2009). “Reproducing kernel Banach spaces for machine learning”. In: <i>Journal of Machine Learning Research</i> 10.Dec, pp. 2741–2775.
zhang2012regularized	Zhang, Haizhang and Jun Zhang (2012). “Regularized learning in Banach spaces as an optimization problem: representer theorems”. In: <i>Journal of Global Optimization</i> 54.2, pp. 235–250.
zhang2013kronecker	Zhang, Huamin and Feng Ding (2013). “On the Kronecker products and their applications”. In: <i>Journal of Applied Mathematics</i> 2013.
zhu2014structured	Zhu, Hongxiao, Fang Yao, and Hao Helen Zhang (2014). “Structured Functional Additive Regression in Reproducing Kernel Hilbert Spaces”. In: <i>Journal of the Royal Statistical Society B (Statistical Methodology)</i> 76.3, pp. 581–603. DOI: 10.1111/rssb.12036 .
zou2005regularization	Zou, Hui and Trevor Hastie (2005). “Regularization and variable selection via the elastic net”. In: <i>Journal of the Royal Statistical Society: Series B (Statistical Methodology)</i> 67.2, pp. 301–320.

Figures

Tables

Theorems

Definitions

Nomenclature

As much as possible, and unless otherwise stated, the following conventions are used throughout this thesis.

Conventions

$\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$	Boldface lower case letters denote real vectors
$\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$	Boldface upper case letters denote real matrices
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$	Calligraphic upper case letters denote sets
x'	Primes are used to distinguish elements (not indicate derivatives)
$\hat{\theta}$	Hats are used to denote estimators of parameters

Indexing

$\mathbf{A}_{ij}, A_{ij}, a_{ij}$	The (i, j) 'th element of the matrix \mathbf{A}
$\mathbf{A}_i.$	The i 'th row of the matrix \mathbf{A} as a tall vector (transposed row vector)
$\mathbf{A}.j$	The j 'th column vector of the matrix \mathbf{A}

Symbols

\mathbb{N}	The set of natural numbers (excluding zero)
\mathbb{Z}	The set of integers
\mathbb{R}	The set of real numbers
$\mathbb{R}_{>0}$	The set of positive real numbers, $\{x \in \mathbb{R} x > 0\}$
$\mathbb{R}_{\geq 0}$	The set of non-negative real numbers, $\{x \in \mathbb{R} x \geq 0\}$
\mathbb{R}^d	The d -dimensional Euclidean space
\mathcal{A}^c	The complement of a set \mathcal{A}
$\mathcal{P}(\mathcal{A})$	The power set of the set \mathcal{A}
$\{\}, \emptyset$	The empty set
$\mathbf{0}$	A vector of zeroes
$\mathbf{1}_n$	A length n vector of ones
\mathbf{I}_n	The $n \times n$ identity matrix
\exists	(short hand) There exists
\forall	(short hand) For all
$\lim_{n \rightarrow \infty}$	The limit as n tends to infinity
$\xrightarrow{\text{dist.}}$	Convergence in distribution
$\mathcal{O}(n)$	Computational complexity (time or storage)
Δx	A quantity representing a change in x

Relations

$a \approx b$	a is approximately or almost equal to b
$a \propto b$	a is equivalent to b up to a constant of proportionality
$a \equiv b$	a is identical to b
$A \Rightarrow B$	The statement B being true is predicated on A being true
$A \Leftrightarrow B$	The statement A is true if and only if B is true
$a \in \mathcal{A}$	a is an element of the set \mathcal{A}
$\mathcal{A} \subseteq \mathcal{B}$	\mathcal{A} is a subset of \mathcal{B} which may include itself
$\mathcal{A} \subset \mathcal{B}$	\mathcal{A} is a subset of \mathcal{B} which does not include itself
$a := b, a \leftarrow b$	a is assigned the value b
$X \sim p(X)$	The random variable X is distributed according to the pdf $p(X)$
$X \sim D$	The random variable X is distributed according to the pdf specified by the distribution D , e.g. $D \equiv \mathcal{N}(0, 1)$
$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} D$	Each random variable $X_i, i = 1, \dots, n$ is independently and identically distributed according to the pdf specified by the distribution D
$X Y$	The (random) variable X given/conditional on Y

Functions

$\inf \mathcal{A}$	The infimum of a set \mathcal{A}
$\sup \mathcal{A}$	The supremum of a set \mathcal{A}
$\min \mathcal{A}$	The minimum value of a set \mathcal{A}
$\max \mathcal{A}$	The maximum value of a set \mathcal{A}
$\arg \min_x f(x)$	The value of x which minimises the function $f(x)$
$\arg \max_x f(x)$	The value of x which maximises the function $f(x)$
$ a $ with $a \in \mathbb{R}$	The absolute value of a ; $ a = a$ if a is positive, and $-a$ if a is negative, and $ 0 = 0$
$\delta_{xx'}$	The Kronecker delta; $\delta_{xx'} = 1$ if $x = x'$, and 0 otherwise
$[A]$	The Iverson bracket; $[A] = 1$ if the logical proposition A is true, and 0 otherwise
$\mathbb{1}_{\mathcal{A}}(x)$	The indicator function; $\mathbb{1}_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$, and 0 otherwise
$e^x, \exp(x)$	The natural exponential function
$\log(x)$	The natural logarithmic function
$\frac{d}{dx} f(x), \dot{f}(x)$	The derivative of f with respect to x
$f \circ g$	Composition of functions, i.e. g following f

Abstract vector space operations and notations

\mathcal{V}^\perp	The orthogonal complement of the space \mathcal{V}
\mathcal{V}^\vee	The algebraic dual space of \mathcal{V}
\mathcal{V}^*	The continuous dual space of \mathcal{V}
$\overline{\mathcal{V}}$	The closure of the space \mathcal{V}
$\mathcal{B}(\mathcal{V})$	The Borel σ -algebra of \mathcal{V}
$L^p(\mathcal{X}, \nu)$	The set of p -integrable functions over the space \mathcal{X} with measure ν
$L(\mathcal{V}; \mathcal{W})$	The set of bounded, linear operators from \mathcal{V} to \mathcal{W}
$\dim(\mathcal{V})$	The dimensions of the vector space \mathcal{V}
$\langle x, y \rangle_{\mathcal{V}}$	The inner product between x and y in the vector space \mathcal{V}

$\ x\ _{\mathcal{V}}$	The norm of x in the vector space \mathcal{V}
$D(x, y)$	The distance between x and y
$x \otimes y$	The tensor product of x and y which are elements of a vector space
$\mathcal{F} \otimes \mathcal{G}$	The tensor product space of two vector spaces
$\mathcal{F} \oplus \mathcal{G}$	The direct sum (or tensor sum) of two vector spaces
$df(x), d^2f(x)$	The first and second Fréchet differentials of f at x
$\partial_v f(x), \partial_v^2 f(x)$	The first and second Gâteaux differentials of f at x in the direction v
$\nabla f(x), \nabla^2 f(x)$	The gradient and Hessian of f at x in the direction v (f is a mapping of a Hilbert space)

Matrix and vector operations

$\mathbf{a}^\top, \mathbf{A}^\top$	The transpose of a vector \mathbf{a} or matrix \mathbf{A}
\mathbf{A}^{-1}	The inverse of a square matrix \mathbf{A}
$\ \mathbf{a}\ ^2$	The squared 2-norm the vector \mathbf{a} , equivalent to $\mathbf{a}^\top \mathbf{a}$
$ \mathbf{A} $	The determinant of a matrix \mathbf{A}
$\text{tr}(\mathbf{A})$	The trace of a square matrix \mathbf{A}
$\text{diag}(\mathbf{A})$	The diagonal elements of a square matrix \mathbf{A}
$\text{rank}(\mathbf{A})$	The rank of a matrix \mathbf{A}
$\text{vec}(\mathbf{A})$	The column-wise vectorisation of a matrix \mathbf{A}
$\mathbf{a} \otimes \mathbf{b}$	The outer product of two vectors \mathbf{a} and \mathbf{b}
$\mathbf{A} \otimes \mathbf{B}$	The Kronecker product of matrix \mathbf{A} with matrix \mathbf{B}
$\mathbf{A} \circ \mathbf{B}$	The Hadamard product two matrices \mathbf{A} and \mathbf{B}

Statistical functions

$P(A)$	The probability of event A occurring
$p(X \theta)$	The probability density function of X given parameters θ
$L(\theta X)$	The log-likelihood of θ given data X , sometimes simply $L(\theta)$
$\text{BF}(M, M')$	Bayes factor for comparing two models M and M'
$\mathcal{I}(\theta)$	The Fisher information for θ
$\mathbb{E}[X], \mathbb{E} X$	The expectation ¹ of the random element X
$\text{Var}[X], \text{Var} X$	The variance ¹ of the random element X
$\text{Cov}[X, Y]$	The covariance ¹ between two random elements X and Y
$H(p)$	The entropy of the distribution $p(X)$
$\text{D}_{\text{KL}}(q(x) p(x))$	The Kullback-Leibler divergence from $p(x)$ to $q(x)$, denoted also by $\text{D}_{\text{KL}}(q p)$ for short

Statistical distributions

$N(\mu, \sigma^2)$	Univariate normal distribution with mean μ and variance σ^2
$N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	d -dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\phi(z)$	The standard normal pdf
$\Phi(z)$	The standard normal cdf

¹When there is ambiguity as to which random element the expectation or variance is taken under or what its distribution is, this is explicated by means of subscripting, e.g. $\mathbb{E}_{X \sim N(0,1)} X$ to denote the expectation of a standard normal random variable.

foot:exp

$\phi(x \mu, \sigma^2)$	The pdf of $N(\mu, \sigma^2)$
$\phi(\mathbf{x} \boldsymbol{\mu}, \boldsymbol{\Sigma})$	The pdf of $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$MN_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$	Matrix normal distribution with mean $\boldsymbol{\mu}$ and row variances $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ and column variances $\boldsymbol{\Psi} \in \mathbb{R}^{m \times m}$
${}^tN(\mu, \sigma^2, a, b)$	Truncated univariate normal distribution with mean μ and variance σ^2 restricted to the interval (a, b)
$N_+(0, 1)$	The half-normal distribution with variance σ^2
$N_+(0, \sigma^2)$	The folded-normal distribution with variance σ^2
${}^tN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathcal{A})$	Truncated d -dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ restricted to the set \mathcal{A}
$\Gamma(s, r)$	Gamma distribution with shape s and rate r parameters
$\Gamma^{-1}(s, \sigma)$	Inverse gamma distribution with shape s and scale σ parameters
χ_d^2	Chi-squared distribution with d degrees of freedom
$\text{Bern}(p)$	Bernoulli distribution with probability of success p
$\text{Cat}(p_1, \dots, p_m)$	Categorical distribution with m categories, and each category has probability of success p_j

Abbreviations

ANOVA	Analysis of variance
cdf	cumulative distribution function
CRAN	Comprehensive R Archive Network
DAG	directed acyclic graph
EM	expectation-maximisation
fBm	Fractional Brownian motion
GPR	Gaussian process regression
HPM	highest probability model
IIA	independent of irrelevant alternatives
iid	Identical and independently distributed
Lasso	Least absolute shrinkage and selection operator
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
MMSE	minimum mean squared error
OLS	ordinary least squares
pd/p.d.	positive definite
pdf	probability density function
PIP	posterior inclusion probability
pmf	probability mass function
PMP	posterior model probability
RKHS	Reproducing kernel Hilbert space
RKKS	Reproducing kernel Kreĭn space
RSS	residual sum of squares
SE	Squared exponential (kernel)