New and Evolving Roles of Shrinkage in Large-Scale Prediction and Inference

Edward George (University of Pennsylvania),
Eric Marchand (University of Sherbrooke),
Gourab Mukherjee (University of Southern California),
Debashis Paul (University of California, Davis)

7 April 2019 – 12 April 2019

1 Overview of the Field

The BIRS workshop "New and Evolving Roles of Shrinkage in Large-Scale Prediction and Inference" brought in thirty-six experts in statistical theory, methods and related applied fields to assess the latest developments and explore new directions in the field of shrinkage methodology and algorithm development for large-scale prediction and statistical inference. There were twenty-six talks and two discussion sessions during the workshop focusing on both the current and future state of this rapidly evolving field. The talks were divided into the following themes:

- Empirical Bayes and optimal shrinkage under heterogeneity and asymmetry,
- Shrinkage priors for Bayesian prediction and inference,
- Spectral shrinkage and inferences under unknown covariances,
- Non-linear shrinkage and penalization methods for structured estimation,
- Efficient and scalable shrinkage algorithms for Big Data Analytics.

Overall there was general consensus that shrinkage methods are now omnipresent in modern data science. The traditional roles of shrinkage have been massively revolutionized as complex shrinkage penalties pertinent to modern data structures are used for large-scale predictive modeling. Historically, the elegant mathematical statistics theory developed following the celebrated works of [87, 86, 9] guided and popularized the usage shrinkage perspective in varied applications as a tool to calibrate bias-variance trade-off. Over the past decades, there has been enormous growth of new applied shrinkage methods. The foundational decision theoretic perspectives now needs to be extended to complex high-dimensional framework to be useful for evaluating efficacy of modern methods. The recent developments in shrinkage theory and methods that were presented in this workshop illustrate the conjoint evolution of field in theory and methods. There has been increasing emphasis on non-linear shrinkage methods that can adapt to latent structure in the data generation process. Lasso-type methods in the Frequentist domain and local-global shrinkage prior based methods in the Bayesian domain have been very successful in variable selection applications. Though contemporary Frequentist and Bayesian approaches to structured shrinkage modeling involves different computational tools, the corresponding estimators are often proved to enjoy similar decision theoretic optimality properties. Shape constraint estimation has been a vibrant area in the recent years yielding pragmatic robust methods based on non-parametric maximum likelihood methodology. The workshop also displayed shape-constrained estimation in the Bayesian domain where novel approaches to set-up constrained MCMC samplers are being used. The workshop had talks involving inference in domains where the covariance structure is unknown and needs to be estimated from the data. These recent works involve new mathematical statistics ideas particularly those involving the spectrum of large matrices and yielded powerful methodologies that can be applied to inventory management, weather forecasting, health-care and environmental sciences. The workshop also illustrated recent developments in predictive density estimation. High-dimensional predictive density estimates adapting to various constrains has been developed and they can be used for probability forecasting in a host of modern applications. The power of predictive density estimation methods were displayed by showing their superior risk properties compared to estimative plug-in approaches. The participants agreed that there is a growing need for a confluence of modern theory and large-scale shrinkage applications. This confluence is extremely important for disciplined growth of the subject by understanding the operating characteristics and working principles of innumerable different penalties proposed by practitioners and data enthusiasts.

2 Recent Developments and Open Problems

The workshop brought to fore the latest usage of shrinkage methodology along with their associated theoretical support and guarantees. Perhaps, the Gaussian sequence model [48] is the simplest framework to explicitly study the role of different kinds of linear and non-linear shrinkage [95, 10, 9, 19]. The optimal direction and magnitude of shrinkage has been well-studied in this set-up [14, 32, 33] and these traditional results have lead to an elegant mathematical theory for risk analysis and decision theory in models where the number of parameters increases along with sample size and equals to the number of observations. These form the foundations for shrinkage theory in modern complex models though closed form estimators and exact tractable analysis is not available here. An intrinsic challenge in shrinkage methods is accurate estimation of the shrinkage hyper-parameters so that the resultant algorithm is tuned for optimal estimative or predictive performance. Most of these traditional results dealing with optimal shrinkage consider a parametric family of estimators indexed by shrinkage hyper-parameters that are then optimally tuned [22, 23, 11, 21, 68]. In real-world applications, these parametric families of estimators often have limited usage and there is need to consider flexible non-parametric counter-parts. Conducting optimal non-parametric shrinkage estimators is difficult [13] and very recently powerful theory and methods in that direction have been proposed [46, 47, 53, 54, 85, 18] by using *Non-parametric Maximum likelihood* (NPMLE) based techniques. This workshop had four talks demonstrating the applicability of NPMLE in constructing potent non-parametric shrinkage estimators. These recent works shows that incorporating shape constraints such as monotonicity improves these estimation frameworks and the resultant estimators are robust and adaptive. A convenient aspect of these estimators is that they can be computed using convex optimization solvers. In this context, the recent R package REBayes of [52] has seen popular usage for implementing these methods.

In complex models that are typically used in contemporary data analysis, the empirical Bayes (EB) perspective [79, 80, 96, 1] is used as an efficient interface for conducting shrinkage by pooling information through a hierarchical set-up. Hierarchical modeling has become an increasingly important statistical method in modeling large and complex datasets as it provides an effective tool for combining information and achieving partial pooling of inference [89, 57, 6]. The applications of hierarchical models involve simultaneous inference on different parameters of interest that are related through a higher level similarity and are often well modeled by a second-level prior. The workshop showcased recent developments in optimal selection of shrinkage magnitude and directions in hierarchical models and applications of hierarchical random effects modeling in mortality rate estimation and individualized prediction [37].

In many contemporary applications [42] auxiliary information regarding the second level structures (such as correlation patterns) is available and can only be suitably incorporated in the hierarchical framework through nonexchangeable priors. Recent work of [5] presented in this workshop develops a generic predictive program for constructing efficient shrinkage rules in a non-exchangeable Gaussian model with with an unknown spiked covariance structure. Shrinkage is facilitated through a family of commutative priors for the mean parameter that are governed by a power hyper-parameter which varies over perfect independence to highly dependent scenarios. Bayes predictive rules in such as set-up involve quadratic forms in functionals of the unknown population covariance. Random matrix theory results [75, 74, 49] are used to correctly estimate the shrinkage estimator in such a Bayesian hierarchical model. Recent Bayesian and Frequentist methods for inducing shrinkage in models with unknown covariance is also covered. Consistent boootstrap based programs for estimation of several spectral statistics have been developed [62, 50]. Recently works have lead to several new methods for estimating the spectral distribution such as by empirical tilting. Novel asymptotic estimation of the distributions of eigen vectors of high-dimensional structured random matrices are developed in [25] for applications in network inference. In spatial statistics, statistical computations for large datasets are a challenge, as it is extremely difficult to store a large covariance or an inverse covariance matrix, and compute its inverse, determinant or Cholesky decomposition. In this domain, scalable matrix-free conditional samplings algorithms are being developed for estimation in spatial mixed models [20].

The workshop documented recent progress in predictive density estimation. The goal in *Predictive Density Estimation* (prde) is to use past data to choose a probability distribution that will be good in predicting the behavior of future samples. The problem of predictive density estimation is one of the most fundamental problems in statistical prediction analysis (See [3, 30, 29]). Traditionally, prediction analysis has dealt with extracting as much information as possible from a small data set. However, over the last decade high-dimensional prde has recieved much attention. Recently, decision theoretic parallels have been established between the predictive density estimation and the multivariate normal mean estimation problem [34, 38, 35, 12, 55, 51, 93, 39, 64]. Connections between the two estimation regimes under parametric constraints have been explored by [92], [26], [60], [70] and [69]. Here, recent development in structured predictive density estimation in varied parametric models were discussed. The resultant density estimates possess optimal predictive log-likelihood properties and can be used in a host of modern applications where data sets with large numbers of predictors are increasingly being collected.

While most of these developments in prde are in the sequence models, the workshop covered regression techniques in high-dimensional model with many more covariates than observations. Here, variable selection is important for constructing good estimators of the unknown parameters [2]. Recent developments in this domain has lead to scalable Bayesian shrinkage priors with desirable posterior concentration [7, 8, 61, 91]. Some of the most popular Bayesian variable selection techniques [67, 36, 44] are built on the "spike and slab" prior distribution. Spike and slab approaches and their computationally tractable extensions have recently been very successfully applied in selecting variables in high-dimensional sparse regression models (See [73, 7, 83, 81, 82, 43] and the references therein).Of particular note is the wide usage of continuous spike-and-slab methods exemplified by horse-shoe priors based methods [76, 77, 91, 61]. Bayesian inference methods for adaptation to shape constraints and approximate Bayesian computations reinvigorating traditional MCMC methods for large scale Bayesian shrinkage are vibrant areas of modern research in this stream of works.

The workshop also covered new directions where empirical Bayes and related shrinkage ideas are beginning to be employed for improving existing inference methodologies. The role of shrinkage in testing hypothesis related to functional data [17] as well as empirical Bayes methods for matrix completion [66] and modern causal inference problems were discussed.

2.1 Open Problems

One of the generic themes of the presentations in the workshop was the importance of extending inferential paradigms from point or interval estimation to predictive inference. Here is a brief list of topics that fit into this research agenda.

1. Predictive density estimation has traditionally focused on usage of the Kullback-Leibler loss. However, new phenomena have begun to emerge in the context of statistical models involving high-dimensional parameters with the utilization of a wider variety of losses that include chi-square, total variation and Wasserstein metrics. One broad class of research problems is to conduct systematic studies to develop understanding and effective strategies for predictive inference for various classical models such as Gaussian sequence model, Poisson regression model, linear regression models and various models for longitudinal data under such loss functions.

2. High-dimensional inference poses a multitude of challenges and opportunities. While usage of shrinkage strategies in high-dimensional problems has a long history, the literature on predictive inference for these problems is still relatively limited. Some relevant open problems include (a) choice of appropriate sparse and dense priors for linear models with unknown covariance; and (b) choice of priors for optimal predictive inference for low-dimensional functionals of high-dimensional parameters.

3. Graphical models have emerged as a powerful modeling paradigm for complex multivariate data. The question of predictive inference for such models has been largely unexplored. One of the challenges here is to come up with appropriate models for describing probabilistic structures on graph spaces that are both analytically and computationally tractable.

3 Presentation Highlights

The workshop demonstrated fascinating progress in *predictive density estimation* (prde). Prde has traditionally been one of the most fundamental problems in statistical inference [31, 3]. Over the past decade, tractable decision theoretic progress regarding efficacies of Bayesian and Frequentist approaches to prde has been made. Eric Marchand summarized these recent results across different loss functions and varied useful parametric set-ups [26, 60, 59, 58, 63]. Bill Strawderman showed that under Kullback-Leibler divergence in spherically symmetric distributions, there exists alluring parallels between point prediction and predictive density estimation and minimum risk equivariant densities can be dominated by Harmonic priors [27]. Iain Johnstone showed that in high dimensional predictive density estimation under sparsity constraints there exists decision theoretic contrasts with minimax sparse location estimation results [69]. These contrasts can be explained by analyzing minimax optimal sparse discrete priors [72]. The minimax risk of popular spike-and-slab prdes were also obtained. Fumiyasu Komaki demonstrated the recent progress made in prde in discrete Poisson models. Prediction in discrete models differs in fundamental aspects from prediction in continuous models [56]. In particular, Fumiyasu Komaki proposed shrinkage priors so that its associated non-parametric Bayesian prde under Kullback-Leibler loss. Keisuke Yano presented optimal tuning and usage of prde in poisson models under sparsity constraints [94]. Sparsity in count data implies situations where there exists an overabundance of zeros or near-zero counts and these sparsity constrained framework relates to zero-inflated models. All of the above developments are built in models with known covariances. Takeru Matsuda [65] showed that in models with unknown covariance, predictive density estimates based on singular value shrinkage prior have efficient prediction properties.

Bill Strawderman and Eric Marchand connected frequentist risk in predictive density estimation with the shrinkage location estimators in Gaussian sequence models. Selection of optimal shrinkage parameters in sequence models is a vibrant area of research. Yuzo Maruyama presented ensemble minimaxity [16] of *James-Stein* type [45] estimators in heteroskedastic sequence set-ups. Roger Koenker presented a non-parametric estimation framework for modeling binary response by single-index models with random coefficients [40]. A new approach for computing the Nonparametric maximum likelihood estimation (NPMLE) that significantly increases computational tractability was developed. Aditya Guntuboyina discussed empirical Bayes estimation of multivariate normal means [85] using NPMLE and extended NPMLE based shrinkage estimation to mixture of regression set-ups. Jiaying Gu demonstrated the applicability of Nonparametric empirical Bayes methods in econometric applications dealing with heterogeneity in both the location and scale parameters. She investigated the performance of NPMLE based ranking methods for studying several compound decision problems in teacher quality evaluation using administrative data. The talk illustrated the importance of empirical Bayes ideas in providing elegant interfaces and interpretative robust estimators for dealing with delicate policy questions. It led to much excitement and considerable discussions among the audience. It was highlighted that unlike most state-of-the -art black-box modeling approaches, the rigorous nature of empirical Bayes rules provide the much needed assurances needed for framing data-driven policy in sensitive applications such as evaluations based incentives, remunerative penalties and terminations.

The workshop had several talks on non-linear shrinkage in Bayesian regression models. Edward

George demonstrated the powerful role of continuous shrinkage priors in sparsity constrained high-dimensional models [83]. Malay Ghosh showed that in high-dimensional multivariate regression model, horse-shoe [77, 76] priors can be used for efficient Bayesian estimation [91, 61]. Anirban Bhattacharya presented recent efforts to scale up Bayesian computation in high- dimensional and shape-constrained regression problems. The transition kernel of an exact MCMC algorithm was perturbed to ease the computational cost per step while maintaining accuracy. Debdeep Pati showed that in Bayesian shape constrained estimation, commonly used priors are not suitable and proposed a novel alternative strategy based on shrinking the coordinates using a multiplicative scale parameter. The proposed shrinkage prior [78] guards against the mass shifting phenomenon while retaining computational efficiency.

Another interesting attribute in this workshop was research talks on statistical methods to deal with unknown covariances. Feng Liang presented the problem of estimating a high-dimensional sparse precision matrix in a Bayesian framework. She showed that adaptation in shrinkage and sparsity levels can be induced by a mixture of Laplace priors [28]. Miles Lopes showed that several spectral statistics can be non-parametrically well estimated in high-dimensional set-ups by using bootstrap based methods [62]. Sanjay Chaudhuri presented the problem of estimating the spectral distribution of large covariance matrices by exponential titling in regimes where sample size as well as the dimension increases proportionately. Debashis Mondal presented modern Bayesian and Frequentists approaches to deal with large covariance matrices in saptial mixed models. Tractable shrinkage estimation using matrix-free conditional samplings algorithms was demonstrated [20]. Yingying Fan showed that asymptotic distributions of eigen vectors in large random matrices can be adequately estimated and used for large-scale network inference [25]. Gourab Mukherjee extended the emprical Bayes framework in Gaussian sequence model [89, 90] and considered prediction under unknown covariance in a hierarchical framework guided by a flexible family of non-cummutative priors. Under spiked structure, the resultant Bayes predictors can be well evaluated in such non-exchangeable flexible models [5].

There were several other fascinating attributes of shrinkage methods that were also displayed in the workshop. Xinyi Xu demonstrated the potential of Bayes factors in Bayesian hypothesis testing and model comparisons. Jinchi Lv presented an interesting methodology in high-dimensional non-parametric inference where distance correlation is used for inference pertaining to attributes related to a pair of large random matrices. Holger Dette showed that hypothesis of practical relevance in functional data analysis often can not be tested by existing methodologies and developed a new bootstrap based test for conducting accurate two-sample tests for those purposes [17]. Jialiang Li presented a model averaging method for constructing a prediction function [88]. Several semiparametric models are weighted for prediction of the mean response. The nonparametric regression models are marginally approximated by spline basis functions and fitted by Bayesian MCMC algorithm. Syed Ejaz Ahmed demonstrated the role of shrinkage in removing implicit bias in Big data applications. He showed how shrinkage rules can be re-calibrated to remove post selection bias in complex regression models. Qingyuan Zhao discussed the potential of using an empirical Bayes perspective [95, 96] in modern usage of causal inference in medical and heath care applications.

4 Scientific Progress Made

There were two discussion sessions during the workshop in which the participants summarized the progress made during the meetings by connecting the dots among the rigorous scientific talks. Several new perspectives in regards to modern applications of shrinkage in big-data applications came up in the discussions. These discussions led to a bigger picture on the current themes of research on shrinkage methods as well as on potential application areas. In large-scale predictive modeling in biology, economics, finance, healthcare, management and marketing sciences, it is now commonplace to use some notions of shrinkage to construct robust algorithms [24, 84, 71, 15, 4].

An interesting attribute of this workshop was the assimilation of mathematical theory with statistical methods used in modern applications. Another feature was the increasing involvement of sophisticated Bayesian computational procedures to implement shrinkage strategies involving complex hierarchical models [78, 7, 83, 81, 82, 91]. A further significant development has been in terms successful demonstration of Bayesian shrinkage strategies in “non-standard” problems motivated by modern applications [65, 66]. Finally, there were several talks on novel shrinkage strategies for shape-restricted inference and high-dimensional inference [85, 62, 41], topics that are becoming increasingly important in modern statistics.

5 Outcome of the Meeting

The workshop showcased the latest advances in the domain of predictive inference and shrinkage procedures in statistics. Some of the key areas of development in contemporary statistics have been (i) inference for high-dimensional data; (ii) sparse parameterization of complex models; (iii) inference under geometric constraints; (iv) computational developments for large volumes of data, and (v) enhanced Bayesian methodologies for nonparametric and semiparametric problems. This meeting has been successful in bringing together experts from each of these fields, and thereby providing a platform for an informative exchange of ideas across these related disciplines. Several collaborative research efforts, including new research proposals and exchange visits by various scholars, have been taking shape as a direct consequence of the academic exchanges during this workshop.

One of the notable features of the workshop was the participation of a large number of young researchers in the field. The key member of the organizing committee, Dr. Gourab Mukherjee, is Assistant Professor of Data Sciences and Operations in the University of Southern California. Among the participants, quite a few were either Assistant Professor or Postdoctoral Scholars in various reputed universities. This workshop therefore allowed these young researchers to showcase their exciting research to the some of the most highly respected senior researchers in the field. At the same time the workshop provided a nice networking opportunity to these young researchers, many of them are expected to be leaders in their respective fields of research.

Participants in the workshop also laid out a number of open problems in the areas of predictive inference and shrinkage-based statistical procedure. It is expected that this meeting will work as a catalyst in bringing together researchers across disciplines to solve the mathematical and computational challenges associated with these exciting questions.

References

[1] Syed Ejaz Ahmed and Nancy Reid. *Empirical bayes and likelihood inference*, volume 148. Springer Science & Business Media, 2012.

[2] Syed Ejaz Ahmed and Bahadır Yüzbaşı. High dimensional data analysis: Integrating submodels. In *Big and Complex Data Analysis*, pages 285–304. Springer, 2017.

[3] James Aitchison and I. R. Dunsmore. *Statistical prediction analysis*. Cambridge University Press, Cambridge, 1975.

[4] T Banerjee, G Mukherjee, P Ghosh, and S Dutta. A large-scale constrained joint modeling approach for predicting user activity, engagement and churn with application to freemium mobile games. *Journal of the American Statistical Association*, 2019.

[5] T Banerjee, G Mukherjee, and D Paul. Improved shrinkage prediction under a spiked covariance structure. *available at* `https://gmukherjee.github.io`, 2018.

[6] T Banerjee, G Mukherjee, and W Sun. Adaptive sparse estimation with side information. *available at* `https://arxiv.org/pdf/1811.11930.pdf`, 2018.

[7] Anirban Bhattacharya, Debdeep Pati, Natesh S Pillai, and David B Dunson. Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490, 2015.

[8] Anirban Bhattacharya, Debdeep Pati, Yun Yang, et al. Bayesian fractional posteriors. *The Annals of Statistics*, 47(1):39–66, 2019.

[9] L. D. Brown. Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.*, 42:855–903, 1971.

[10] Lawrence D. Brown. *Fundamentals of statistical exponential families with applications in statistical decision theory*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9. Institute of Mathematical Statistics, Hayward, CA, 1986.

[11] Lawrence D. Brown. In-season prediction of batting averages: a field test of empirical Bayes and Bayes methodologies. *Ann. Appl. Stat.*, 2(1):113–152, 2008.

[12] Lawrence D. Brown, Edward I. George, and Xinyi Xu. Admissible predictive density estimation. *Ann. Statist.*, 36(3):1156–1170, 2008.

[13] Lawrence D Brown and Eitan Greenshtein. Nonparametric empirical bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *The Annals of Statistics*, pages 1685–1704, 2009.

[14] Lawrence D. Brown and Jiunn Tzon Hwang. A unified admissibility proof. In *Statistical decision theory and related topics, III, Vol. 1 (West Lafayette, Ind., 1981)*, pages 205–230. Academic Press, New York, 1982.

[15] Lawrence D Brown, Gourab Mukherjee, Asaf Weinstein, et al. Empirical bayes estimates for a two-way cross-classified model. *The Annals of Statistics*, 46(4):1693–1720, 2018.

[16] Lawrence D Brown, Hui Nie, and Xianchao Xie. Ensemble minimax estimation for multivariate normal means. *The Annals of Statistics*, page 1, 2011.

[17] Holger Dette, Kevin Kokot, and Alexander Aue. Functional data analysis in the banach space of continuous functions. *arXiv preprint arXiv:1710.07781*, 2017.

[18] Lee H Dicker and Sihai D Zhao. High-dimensional classification via nonparametric empirical bayes and maximum likelihood inference. *Biometrika*, 103:21–34, 2016.

[19] David L. Donoho, Iain M. Johnstone, Jeffrey C. Hoch, and Alan S. Stern. Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B*, 54(1):41–81, 1992. With discussion and a reply by the authors.

[20] Somak Dutta and Debashis Mondal. An h-likelihood method for spatial mixed linear models based on intrinsic auto-regressions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):699–726, 2015.

[21] Bradley Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press, 2012.

[22] Bradley Efron and Trevor Hastie. *Computer Age Statistical Inference*, volume 5. Cambridge University Press, 2016.

[23] Bradley Efron and Carl Morris. Stein's estimation rule and its competitors—an empirical bayes approach. *Journal of the American Statistical Association*, 68(341):117–130, 1973.

[24] J. Fan, J. Lv, and L. Qi. Sparse high-dimensional models in economics. *Annual Review of Economics*, 2011.

[25] Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. Asymptotic theory of eigenvectors for large random matrices. *arXiv preprint arXiv:1902.06846*, 2019.

[26] Dominique Fourdrinier, Éric Marchand, Ali Righi, and William E. Strawderman. On improved predictive density estimation with parametric constraints. *Electron. J. Stat.*, 5:172–191, 2011.

[27] Dominique Fourdrinier, Éric Marchand, and William E Strawderman. On efficient prediction and predictive density estimation for normal and spherically symmetric models. *Journal of Multivariate Analysis*, 173:18–25, 2019.

[28] Lingrui Gan, Naveen N. Narisetty, and Feng Liang. Bayesian regularization for graphical models with unequal shrinkage. *Journal of the American Statistical Association*, 0(0):1–14, 2018.

[29] S. Geisser. The future of statistics in retrospect. In *Bayesian statistics, 3 (Valencia, 1987)*, Oxford Sci. Publ., pages 147–158. Oxford Univ. Press, New York, 1988.

[30] Seymour Geisser. The inferential use of predictive distributions. In *Foundations of statistical inference (Proc. Sympos., Univ. Waterloo, Ont., 1970)*, pages 456–469. Holt, Rinehart and Winston of Canada, Toronto, Ont., 1971. With comments by V. P. Godambe, I. J. Good, W. J. Hall and D. A. Sprott and a reply by the author.

[31] Seymour Geisser. *Predictive inference*, volume 55 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993. An introduction.

[32] Edward George. Minimax multiple shrinkage estimation. *The Annals of Statistics*, 14(1):188–205, 1986.

[33] Edward I George. Combining minimax shrinkage estimators. *Journal of the American Statistical Association*, 81(394):437–445, 1986.

[34] Edward I. George, Feng Liang, and Xinyi Xu. Improved minimax predictive densities under Kullback-Leibler loss. *Ann. Statist.*, 34(1):78–91, 2006.

[35] Edward I. George, Feng Liang, and Xinyi Xu. From minimax shrinkage estimation to minimax shrinkage prediction. *Statist. Sci.*, 27(1):82–94, 2012.

[36] Edward I George and Robert E McCulloch. Approaches for bayesian variable selection. *Statistica sinica*, pages 339–373, 1997.

[37] Edward I George, Veronika Ročková, Paul R Rosenbaum, Ville A Satopää, and Jeffrey H Silber. Mortality rate estimation and standardization for public reporting: Medicare's hospital compare. *Journal of the American Statistical Association*, 112(519):933–947, 2017.

[38] Edward I. George and Xinyi Xu. Predictive density estimation for multiple regression. *Econometric Theory*, 24(2):528–544, 2008.

[39] Malay Ghosh, Victor Mergel, and Gauri Sankar Datta. Estimation, prediction and the Stein phenomenon under divergence loss. *J. Multivariate Anal.*, 99(9):1941–1961, 2008.

[40] Jiaying Gu and Roger Koenker. Empirical bayesball remixed: Empirical bayes methods for longitudinal data. *Journal of Applied Econometrics*, 32(3):575–599, 2017.

[41] Jiaying Gu and Roger Koenker. Unobserved heterogeneity in income dynamics: An empirical bayes perspective. *Journal of Business & Economic Statistics*, 35(1):1–16, 2017.

[42] Campbell R Harvey, Yan Liu, and Heqing Zhu. ...and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68, 2016.

[43] Hemant Ishwaran and J Sunil Rao. Spike and slab gene selection for multigroup microarray data. *Journal of the American Statistical Association*, 100(471):764–780, 2005.

[44] Hemant Ishwaran and J Sunil Rao. Spike and slab variable selection: frequentist and bayesian strategies. *Annals of Statistics*, pages 730–773, 2005.

[45] William James and Charles Stein. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley symposium on Mathematical Statistics and Probability*, volume 1, pages 361–379, 1961.

[46] Wenhua Jiang, Cun-Hui Zhang, et al. General maximum likelihood empirical bayes estimation of normal means. *The Annals of Statistics*, 37(4):1647–1684, 2009.

[47] Wenhua Jiang, Cun-Hui Zhang, et al. Empirical bayes in-season prediction of baseball batting averages. In *Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown*, pages 263–273. Institute of Mathematical Statistics, 2010.

[48] I. M. Johnstone. Gaussian estimation: Sequence and wavelet models. Available at: "`http://www-stat.stanford.edu/~imj`", 2012.

[49] Iain M Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 2012.

[50] Noureddine El Karoui and Elizabeth Purdom. The bootstrap, covariance matrices and pca in moderate and high-dimensions. *arXiv preprint arXiv:1608.00948*, 2016.

[51] Kei Kobayashi and Fumiyasu Komaki. Bayesian shrinkage prediction for the regression problem. *J. Multivariate Anal.*, 99(9):1888–1905, 2008.

[52] Roger Koenker and Jiaying Gu. Rebayes: Empirical bayes mixture methods in r. *Journal of Statistical Software*, 82(8):1–26, 2017.

[53] Roger Koenker and Ivan Mizera. Convex optimization, shape constraints, compound decisions, and empirical bayes rules. *Journal of the American Statistical Association*, 109(506):674–685, 2014.

[54] Roger Koenker, Ivan Mizera, et al. Quasi-concave density estimation. *The Annals of Statistics*, 38(5):2998–3027, 2010.

[55] Fumiyasu Komaki. A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, 88(3):859–864, 2001.

[56] Fumiyasu Komaki. Simultaneous prediction of independent Poisson observables. *Ann. Statist.*, 32(4):1744–1769, 2004.

[57] Samuel Kou and Justin J Yang. Optimal shrinkage estimation in heteroscedastic hierarchical linear models. *arXiv preprint arXiv:1503.06262*, 2015.

[58] Tatsuya Kubokawa, Éric Marchand, and William E Strawderman. On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, 142:57–74, 2015.

[59] Tatsuya Kubokawa, Éric Marchand, William E Strawderman, et al. On predictive density estimation for location families under integrated absolute error loss. *Bernoulli*, 23(4B):3197–3212, 2017.

[60] Tatsuya Kubokawa, Éric Marchand, William E Strawderman, and Jean-Philippe Turcotte. Minimaxity in predictive density estimation with parametric constraints. *Journal of Multivariate Analysis*, 116:382–397, 2013.

[61] Minjung Kyung, Jeff Gill, Malay Ghosh, George Casella, et al. Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, 5(2):369–411, 2010.

[62] Miles Lopes, Andrew Blandino, and Alexander Aue. Bootstrapping spectral statistics in high dimensions. *arXiv preprint arXiv:1709.08251*, 2017.

[63] Éric Marchand, François Perron, and Iraj Yadegari. On estimating a bounded normal mean with applications to predictive density estimation. *Electronic Journal of Statistics*, 11(1):2002–2025, 2017.

[64] Yuzo Maruyama and Toshio Ohnishi. Harmonic bayesian prediction under alpha-divergence. *arXiv preprint arXiv:1605.05899*, 2016.

[65] Takeru Matsuda and Fumiyasu Komaki. Singular value shrinkage priors for bayesian prediction. *Biometrika*, 102(4):843–854, 2015.

[66] Takeru Matsuda and Fumiyasu Komaki. Empirical bayes matrix completion. *Computational Statistics & Data Analysis*, 2019.

[67] Toby J Mitchell and John J Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

[68] Carl N Morris. Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55, 1983.

[69] G. Mukherjee and I. M. Johnstone. Exact minimax estimation of the predictive density in sparse gaussian models. *Annals of Statistics*, 43(3):937, 2015.

[70] Gourab Mukherjee. *Sparsity and shrinkage in predictive density estimation*. PhD thesis, Stanford University, 2013.

[71] Gourab Mukherjee, Lawrence D Brown, and Paat Rusmevichientong. Efficient empirical bayes prediction under check loss using asymptotic risk estimates. *arXiv preprint arXiv:1511.00028*, 2015.

[72] Gourab Mukherjee and Iain M Johnstone. On minimax optimality of sparse bayes predictive density estimates. *arXiv preprint arXiv:1707.04380*, 2017.

[73] Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

[74] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617, 2007.

[75] Debashis Paul and Alexander Aue. Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150:1–29, 2014.

[76] Nicholas G Polson and James G Scott. Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian statistics*, 9:501–538, 2010.

[77] Nicholas G Polson, James G Scott, et al. On the half-cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4):887–902, 2012.

[78] Pallavi Ray, Debdeep Pati, and Anirban Bhattacharya. Efficient bayesian shape-restricted function estimation with constrained gaussian process priors. *arXiv preprint arXiv:1902.04701*, 2019.

[79] Herbert Robbins. An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pages 157–163, Berkeley and Los Angeles, 1956. University of California Press.

[80] Herbert Robbins. The empirical bayes approach to statistical decision problems. In *Herbert Robbins Selected Papers*, pages 49–68. Springer, 1985.

[81] Veronika Rocková. Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *Annals of Statistics*, 2017.

[82] Veronika Ročková and Edward I George. Negotiating multicollinearity with spike-and-slab priors. *Metron*, 72(2):217–229, 2014.

[83] Veronika Ročková and Edward I George. The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444, 2018.

[84] Peter E Rossi, Greg M Allenby, and Rob McCulloch. *Bayesian statistics and marketing*. John Wiley & Sons, 2012.

[85] Sujayam Saha and Adityanand Guntuboyina. On the nonparametric maximum likelihood estimator for gaussian location mixture densities with application to gaussian denoising. *arXiv preprint arXiv:1712.02009*, 2017.

[86] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955*, volume 1, pages 197–206. University of California Press, Berkeley and Los Angeles, 1956.

[87] Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151, 1981.

[88] Jingli Wang and Jialiang Li. Semiparametric model averaging prediction: a bayesian approach. *Australian & New Zealand Journal of Statistics*, 60(4):407–422, 2018.

[89] Xianchao Xie, SC Kou, and Lawrence D Brown. Sure estimates for a heteroscedastic hierarchical model. *Journal of the American Statistical Association*, 107(500):1465–1479, 2012.

[90] Xianchao Xie, SC Kou, and Lawrence D Brown. Optimal shrinkage estimation of mean parameters in family of distributions with quadratic variance. *Annals of Statistics*, 44(2):564–597, 2016.

[91] Xiaofan Xu, Malay Ghosh, et al. Bayesian variable selection and estimation for group lasso. *Bayesian Analysis*, 10(4):909–936, 2015.

[92] Xinyi Xu and Feng Liang. Asymptotic minimax risk of predictive density estimation for non-parametric regression. *Bernoulli*, 16(2):543–560, 2010.

[93] Xinyi Xu and Dunke Zhou. Empirical bayes predictive densities for high-dimensional normal models. *J. Multivariate Analysis*, 102(10):1417–1428, 2011.

[94] Keisuke Yano, Ryoya Kaneko, and Fumiyasu Komaki. Asymptotically minimax predictive density for sparse count data. *arXiv preprint arXiv:1812.06037*, 2018.

[95] Cun-Hui Zhang. Empirical bayes and compound estimation of normal means. *Statistica Sinica*, 7(1):181–193, 1997.

[96] Cun-Hui Zhang. Compound decision theory and empirical bayes methods. *The Annals of Statistics*, 31(2):379–390, 2003.