Research statement

David R. Bickel

March 9, 2020

1 Research program 1: Methods of analyzing large-scale data

In empirical Bayes multiple testing, a prior distribution is estimated by considering each gene or other feature corresponding to a null hypothesis to be randomly selected from some reference class. That leads to the problem of selecting a reference class for the application of empirical Bayes methods since each feature may belong to more than one eligible reference class. Two solutions to the reference class problem were developed by Aghababazadeh et al. (2018) and Karimnezhad and Bickel (2018). The latter solution was further studied by Mei et al. (2017).

Bickel and Rahal (2019a) provided methods of transforming estimates of nonlocal false discovery rates to estimates of local false discovery rates by correcting a bias in the former. That work represents an advance over Bickel (2013). Bickel and Rahal (2019b) introduced a new multiple testing procedure and applied it to the analysis of large-scale genomics data.

2 Research program 2: Uncertainty about statistical models

Building on Bickel (2012a), Bickel (2012b) developed a probability-interval framework for blending frequentist and Bayesian methods on the basis of game-theoretic first principles. It enables data analysis on the basis of the posterior distribution that is a blend between a set of plausible Bayesian posterior distributions and a parameter distribution that represents a frequentist method of data analysis by combining strengths of the frequentist and Bayesian approaches. Such blended inference relied on maximum entropy but was extended to a more general information-theoretic framework in Bickel (2019c).

Bickel (2015) introduced a complementary two-stage approach to the problem of uncertainty about the prior. In the first stage, all priors in conflict with the data are eliminated. With this approach, prior conflict may be detected by a number of means, including Bayes factors, proper scoring rules, or the integrated likelihood ratios that Bickel (2012c) used to quantify statistical evidence. In the second stage, a robust Bayes decision rule or distribution-combination method is applied to the posterior distributions corresponding to the remaining priors in order to generate unique inferences and actions such as point estimates.

Bickel (2018) studied fiducial model averaging as a method of combining statistical models in the same way as Bayesian model averaging except that a fiducial distribution over models replaces a Bayesian posterior distribution over models. Unlike conventional Bayesian model averaging, fiducial model averaging can combine frequentist models as well as Bayesian models (Bickel, 2018).

More recently, Bickel (2019e) contributed to the current debate on null hypothesis significance testing by proving that when a null hypothesis has a sufficiently low p value, any Bayesian model that would assign the null hypothesis a large posterior probability would fail a model check. It provides a method of calibrating p values that complements the approaches of Bickel (2019f, 2020).

3 Research program 3: Statistical methods for data science

The analysis of big data can benefit from statistical methods developed in the above research programs. The relevance of the program on analyzing large-scale data (§1) is noted in Section 3.1, and that of the program on model uncertainty (§2) is noted in Section 3.2.

3.1 Assessing the performance of deep neural networks

Bayesian models use posterior predictive distributions to quantify the uncertainty of their predictions. Similarly, the point predictions of neural networks and other supervised learning algorithms may be converted to predictive distributions by various bootstrap methods (Bickel, 2019g, §2). The predictive performance of each algorithm can then be assessed by quantifying the performance of its predictive distribution.

Previous methods for assessing such performance are relative, indicating whether certain algorithms perform better than others (e.g., Quiñonero-Candela et al., 2006). I proposed performance measures that are absolute in the sense that they indicate whether or not an algorithm performs adequately without requiring comparisons to other algorithms (Bickel, 2019g, §1). The first proposed performance measure is a *model predictive p value*, a generalization of a prior predictive p value with the prior distribution equal to the posterior distribution of previous data (Bickel, 2019g, §3). The other proposed performance measures use the model predictive p value for each prediction together with the methods in Bickel and Rahal (2019a), cited in Section 1, to estimate the proportion of target values that are compatible with the predictive distribution (Bickel, 2019g, §4). I illustrated the new performance measures by using them to evaluate the predictive performance of deep neural networks when applied to the analysis of a housing-price data set that is used as a standard in machine learning (Bickel, 2019g, §5).

3.2 Uncertainty about machine learning algorithms

Choosing the supervised learning algorithm that performs the best on test data can overfit the test data in the same way that an individual algorithm can overfit the training data. Such model selection is outperformed by Bayesian model averaging in the presence of a suitable a prior distribution over the algorithms. Bickel (2019d) used a development of Occam's razor (Bickel, 2019a,b) to adjust that prior distribution over models for the complexity of their predictive distributions.

The model predictive p values introduced in Bickel (2019g, §3) can empower the fiducial averaging of supervised learning algorithms in analogy with Bickel (2018)'s use of prior predictive p values for the fiducial averaging of Bayesian models.

4 More research contributions

Additional contributions are organized at `https://davidbickel.com/`.

References

Aghababazadeh, F. A., Alvo, M., Bickel, D. R., 2018. Estimating the local false discovery rate via a bootstrap solution to the reference class problem. PLoS ONE 13, e0206902.

Bickel, D. R., 2012a. Coherent frequentism: A decision theory based on confidence sets. Communications in Statistics - Theory and Methods 41, 1478–1496.

Bickel, D. R., 2012b. Controlling the degree of caution in statistical inference with the Bayesian and frequentist approaches as opposite extremes. Electron. J. Statist. 6, 686–709.

Bickel, D. R., 2012c. The strength of statistical evidence for composite hypotheses: Inference to the best explanation. Statistica Sinica 22, 1147–1198.

Bickel, D. R., 2013. Simple estimators of false discovery rates given as few as one or two p-values without strong parametric assumptions. Statistical Applications in Genetics and Molecular Biology 12, 529–543.

Bickel, D. R., 2015. Inference after checking multiple Bayesian models for data conflict and applications to mitigating the influence of rejected priors. International Journal of Approximate Reasoning 66, 53–72.

Bickel, D. R., 2018. A note on fiducial model averaging as an alternative to checking Bayesian and frequentist models. Communications in Statistics - Theory and Methods 47, 3125–3137.

Bickel, D. R., 2019a. An explanatory rationale for priors sharpened into Occam's razors. Bayesian Analysis, DOI: 10.1214/19-BA1189.
URL `https://doi.org/10.1214/19-BA1189`

Bickel, D. R., 2019b. Confidence intervals, significance values, maximum likelihood estimates, etc. sharpened into Occam's razors. Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2019.1580739.
URL `https://doi.org/10.1080/03610926.2019.1580739`

Bickel, D. R., 2019c. Maximum entropy derived and generalized under idempotent probability to address Bayes-frequentist uncertainty and model revision uncertainty, working paper, DOI: 10.5281/zenodo.2645555.
URL `https://doi.org/10.5281/zenodo.2645555`

Bickel, D. R., 2019d. Model averages sharpened into Occam's razors: Deep learning enhanced by Rényi entropy, working paper, DOI: 10.5281/zenodo.3565931.
URL `https://doi.org/10.5281/zenodo.3565931`

Bickel, D. R., 2019e. Null hypothesis significance testing defended and calibrated by Bayesian model checking. The American Statistician, DOI: 10.1080/00031305.2019.1699443.
URL `https://doi.org/10.1080/00031305.2019.1699443`

Bickel, D. R., 2019f. Null hypothesis significance testing interpreted and calibrated by estimating probabilities of sign errors: A Bayes-frequentist continuum, working paper, DOI: 10.5281/zenodo.3569888.
URL `https://doi.org/10.5281/zenodo.3569888`

Bickel, D. R., 2019g. Testing prediction algorithms as null hypotheses: Application to assessing the performance of deep neural networks, working paper, DOI: 10.1002/sta4.270.
URL `https://doi.org/10.1002/sta4.270`

Bickel, D. R., 2020. Interval estimation, point estimation, and null hypothesis significance testing calibrated by an estimated posterior probability of the null hypothesis, working paper, DOI: 10.5281/zenodo.3694136.
URL `https://doi.org/10.5281/zenodo.3694136`

Bickel, D. R., Rahal, A., 2019a. Correcting false discovery rates for their bias toward false positives. Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2019.1630432.

Bickel, D. R., Rahal, A., 2019b. Model fusion and multiple testing in the likelihood paradigm: Shrinkage and evidence supporting a point null hypothesis. Statistics 53, 1187–1209.

Karimnezhad, A., Bickel, D. R., 2018. Incorporating prior knowledge about genetic variants into the analysis of genetic association data: An empirical Bayes approach. IEEE/ACM Transactions on Computational Biology and Bioinformatics, DOI: 10.1109/TCBB.2018.2865420.
URL `https://ieeexplore.ieee.org/document/8436435/`

Mei, S., Karimnezhad, A., Forest, M., Bickel, D. R., Greenwood, C., 2017. The performance of a new local false discovery rate method on tests of association between coronary artery disease (CAD) and genome-wide genetic variants. PLoS ONE 12, e0185174.

Quiñonero-Candela, J., Rasmussen, C. E., Sinz, F., Bousquet, O., Schölkopf, B., 2006. Evaluating predictive uncertainty challenge. In: Quiñonero-Candela, J., Dagan, I., Magnini, B., d'Alché Buc, F. (Eds.), Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–27.