

DISCUSSION OF THE PAPER “CONNECTING MODEL-BASED AND MODEL-FREE APPROACHES TO LINEAR LEAST SQUARES REGRESSION” BY LUTZ DÜMBGEN AND LAURIE DAVIES (2024)

Larry Wasserman ¹

Department of Statistics & Data Science, Carnegie Mellon, Pittsburgh, PA, USA

1. INTRODUCTION

The authors show that many familiar inferential tools for linear models can be interpreted in a model free framework (Dümbgen and Davies, 2024). This builds on the authors’ previous work where they have developed a substantial theory of model free inference. The connections between the model heavy and model free approaches are interesting. I’d like to comment more broadly about the role of models.

2. ARE MODELS USEFUL?

I think that most statistical models are wrong. In fact, the assumption that data are random draws from some distribution is usually a fiction. But there is a saying (I think due to Tukey) which is: use models but don’t believe them. In other words, statistical models give us a way to think about data analysis even when the models are not correct. However, assuming the model is exactly correct may be too extreme. Model free methods provide an alternative. But there is an approach that lies between these two extremes, namely, the projection approach. Buja *et al.* (2019) call this the *model lean* approach.

3. PROJECTION INFERENCE

One way to formalize the model lean approach is to focus on projections (Park *et al.*, 2023). Let $\mathcal{P} = (P_\theta : \theta \in \Theta)$ be a family of distributions indexed by the parameter θ . We posit that there is indeed a true distribution P but we do not assume that P is in $\mathcal{P}$.

¹ Corresponding Author. E-mail: larry@cmu.edu

Define the projection parameter θ_0 to be the value of θ that minimizes $D(P, P_\theta)$ where D is some discrepancy.

If $\widehat{D}(P, P_\theta)$ is an estimate of D then the minimum distance estimator $\widehat{\theta}$ of θ_0 is the minimizer of $\widehat{D}(P, P_\theta)$. Examples include the minimum Hellinger estimator (Beran, 1977), the minimum power divergence estimator (Basu *et al.*, 1998) and the maximum mean discrepancy estimator (Chérif-Abdellatif and Alquier, 2022). The Hellinger projection is notable for the fact that it is efficient if the model happens to be correct but it requires density estimation. The power divergence is simpler as it does not require density estimation.

Confidence sets for the projection parameter θ_0 can be constructed using standard theory since the minimum distance estimator is an M-estimator. But this requires a number of regularity conditions on the model and on P . In the spirit of reducing assumptions, we can use an relative fit (Park *et al.*, 2023). For each θ , we test the hypothesis that $D(P, P_\theta) \leq D(P, P_{\widehat{\theta}})$. Inverting that test yields a confidence set for θ_0 that is valid under weak assumptions. This is because the estimated discrepancy $\widehat{D}(P, P_\theta)$ at a fixed θ can have good large sample behavior even when the minimum distance estimator $\widehat{\theta}$ does not.

For example, suppose we use the L_2 discrepancy $D(P, P_\theta) = \int (p(x) - p_\theta(x))^2 dx$ which is a special case of the power divergence. We proceed as follows:

1. Split the data $X_1, \dots, X_{2n}$ into two groups $\mathcal{D}_0$ and $\mathcal{D}_1$.
2. From $\mathcal{D}_0$ find $\widehat{\theta}$ to minimize $\int p_\theta^2(x) dx - (2/n) \sum_i p_\theta(X_i)$.
3. From $\mathcal{D}_1$ construct the set

$$C = \left\{ \theta : U(\theta, \widehat{\theta}) \leq \frac{c_\alpha}{\sqrt{n}} \right\}$$

where $U = \widehat{D}(P, P_\theta) - \widehat{D}(P, P_{\widehat{\theta}})$ and c_α is an appropriate critical value.

Under weak conditions, $\liminf_{n \rightarrow \infty} P(\theta_0 \in C) \geq 1 - \alpha$. This does not require any conditions on the behavior of $\widehat{\theta}$ or any regularity of the model. (It is possible to de-randomize the procedure to eliminate the randomness due to sample splitting.)

I am curious about how the authors' opinions of such compromise methods that lie between the extremes.

4. CONCLUSION

I congratulate the authors on an interesting paper and on their development of model free methods. I look forward to further developments in this direction.

REFERENCES

- A. BASU, I. R. HARRIS, N. L. HJORT, M. JONES (1998). *Robust and efficient estimation by minimising a density power divergence*. Biometrika, 85, no. 3, pp. 549–559.
- R. BERAN (1977). *Minimum hellinger distance estimates for parametric models*. The Annals of Statistics, pp. 445–463.
- A. BUJA, L. BROWN, R. BERK, E. GEORGE, E. PITKIN, M. TRASKIN, K. ZHANG, L. ZHAO (2019). *Models as approximations I*. Statistical Science, 34, no. 4, pp. 523–544.
- B.-E. CHÉRIEF-ABDELLATIF, P. ALQUIER (2022). *Finite sample properties of parametric MMD estimation: robustness to misspecification and dependence*. Bernoulli, 28, no. 1, pp. 181–213.
- L. DÜMBGEN, L. DAVIES (2024). *Connecting model-based and model-free approaches to linear least squares regression*. Statistica, 84, no. 2, pp. 65–81.
- B. PARK, S. BALAKRISHNAN, L. WASSERMAN (2023). *Robust universal inference*. arXiv preprint arXiv:2307.04034.