Stimulus Identification from fMRI scans

Charles Zheng and Yuval Benjamini

Stanford University, Department of Statistics

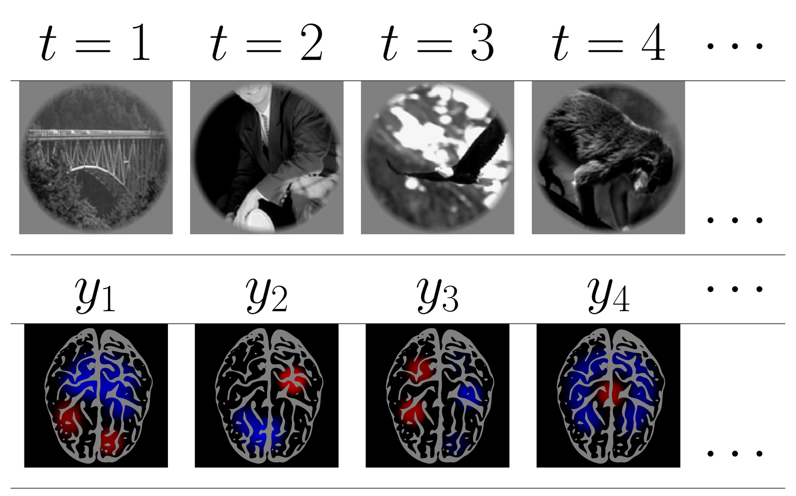
Overview

Seeking to explain the processes behind human perception, scientists employ *forward models* to model the causal relationship between perception of stimuli and neural activity. But how can we measure the quality of these models? Kay et al (2008) introduced the task of *identification* as a way to demonstrate the fidelity and generalizability of the model.

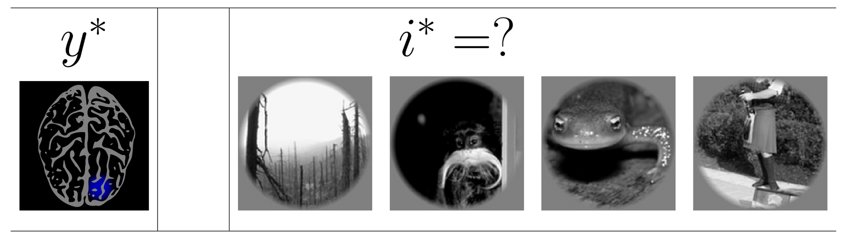
Using the data of Kay *et al.* as a motivating example, we consider the statistical problem of optimal identification. While identification superficially resembles a classification task (with many classes), it combines the challenge of multivariate regression with high-dimensional discrimination.

Data

- Sequence of stimuli (pictures) shown at time $t = 1, \ldots, T = 3500$
- Record subject's multivariate response $y_t \in \mathbb{R}^p$, here $p \approx 20000$ voxels



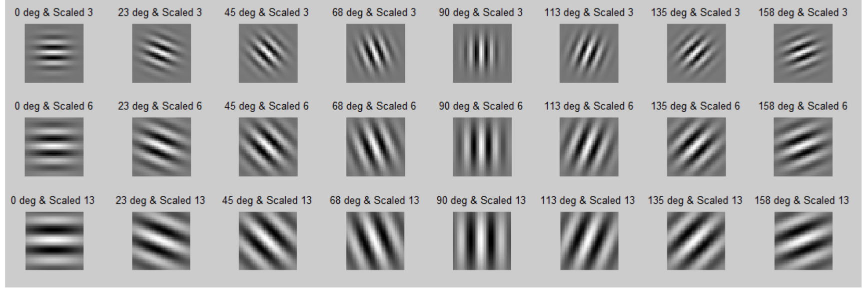
Identification



- Let S be a set of stimuli, possibly outside the training set! $|S|$ can range from 120 to 10000
- Scientist picks a stimulus i^* from S and measures the subject's reponse y^*
- Can the statistician *identify* $i^* \in S$ from y^*?
- *Objective*: minimize misclassification rate

Previous approaches

- In order to generalize to new stimuli, we need to find some quantitative representation
- Kay (2008) uses *Gabor filters* to describe each picture in terms of $q = 10000$ real-valued features



- Notation: write $Y_{T\times p}$ is a matrix containing the T of recorded responses, and where $X_{T\times q}$ is the matrix of the *image features* of the corresponding stimuli

Now consider a parametric model

$$Y \sim F_\theta(X)$$

Such a *forward model* gives the distribution of the response conditional on the stimuli features; while identification requires the converse.
However, the *maximum likelihood* (ML) principle can be invoked to identify the stimuli $i \in S$ "most likely" to have produced y^*.
Let $x_i : i \in S$ denote features of the test stimuli, and identify y^* based on the maximum likelihood (ML) principle

$$i^* = \text{argmax}_i \ell_\theta(y^*|x_i)$$

Example. We take the following as a representative approach, combining features of [1] and [2]:

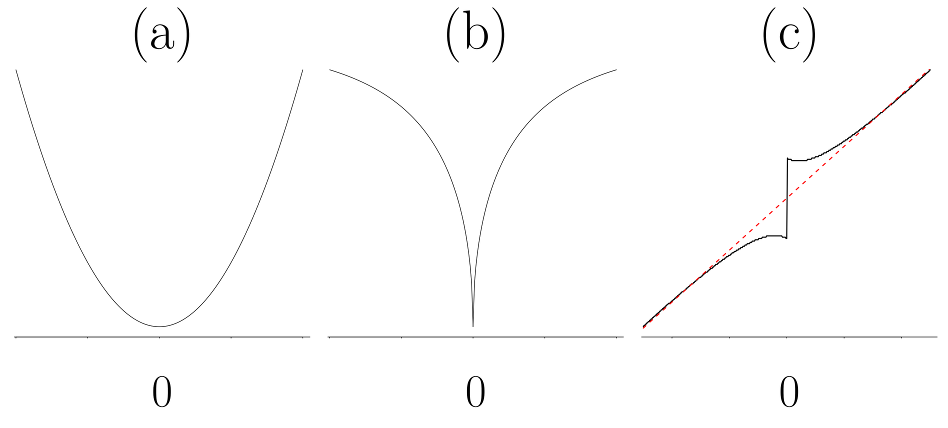
- Assume the normal mutivariate linear model
$$Y \sim N(XB, \Sigma_E), \text{ where } B \in \mathbb{R}^{q\times p}$$
- Obtain point estimates of B using elastic net [4], and Σ_E using the sample covariance of residuals with off-diagonal shrinkage
- The ML rule takes the form
$$i^* = \text{argmin}_i (x_i^T \hat{B} - y^*)^T \hat{\Sigma}_E^{-1}(x_i^T\hat{B} - y^*)$$

Initial Questions

- Can ML (e.g. the method above) be considered an optimal method for identification in any sense?
- If not, how can we do better? Can we find a good method which is computationally tractable?

Limitations of ML

- ML is consistent given the correct model, but can be rather poor in finite samples
- The point estimates $\hat{B}$ is generally obtained by minimizing *prediction error*, but the loss function for identification is different
- Perhaps we can find point estimates tailored for the identification loss function. But the nonconvexity makes it difficult!

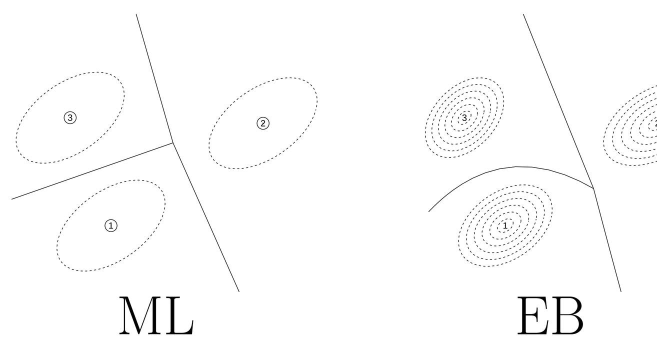


(a) Squared error loss and *(b)* loss function for identification, as a function of the difference between the true mean signal and the predicted signal.

(c) A $p = q = 1$-dimensional example where ML (actually MAP) fails. The Bayes estimate for identification (black) and the MAP estimate (red) diverge sharply when $B \sim N(0, 1)$. The same phenomenon can be found in higher dimensions.

Empirical Bayes

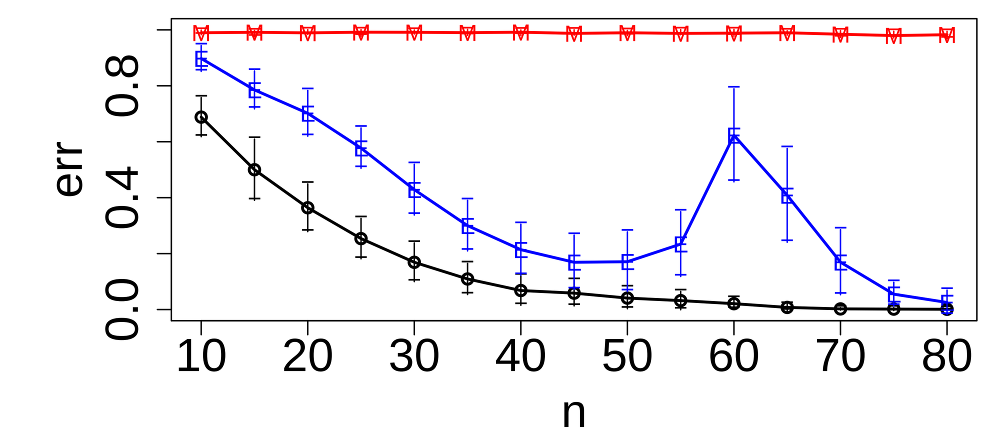
- *Idea*: Unlike ML, the Bayes rule surely optimizes the "correct" objective function. Can we approximate the Bayes rule?
- *Empirical Bayes*: use the data to estimate the covariances Σ_B and Σ_E, then compute posterior distribution of B
- Assume coefficients of B independent; diagonals of Σ_B can be estimated using any estimate of signal strength, e.g. *Eigenprism* [3].
- Similar decision rule to ML
$$\min (x_i^T B - y^*)^T (\text{Cov}(x_i^T B) + \hat{\Sigma}_E)^{-1}(x_i^T B - y^*)$$
but with "added noise" due to uncertainty of B



- *Computation:* requires inverting $pq \times pq$ matrix

Simulation Results

- Parameters $p = q = 60$, random B and Σ_E
- Empirical bayes outperforms ML when $n < q$... however, still unstable!



(E) Empirical Bayes, (M) Maximum likelihood, (o) Bayes risk *(knowing true Σ_B, Σ_E)*

Ongoing Work

- Why does error *increase* with sample size!?
- Refine the crucial step of estimating Σ_B
- Required cost of $O((pq)^3)$ hinders application to real data... develop tractable approximations

Conclusions

- ML-based approaches rely on point estimates, and hence optimize the wrong objective function
- Empirical bayes achieves better performance by approximating the Bayes rule, but the "empirical" part remains tricky
- Better theoretical understanding is needed to explain why EB succeeds (and sometimes fails)

References

[1] Kay et al. *Nature* (2008)
[2] Vu et al. *Annals of Applied Statistics* (2011)
[3] Janson et al. (2015) http://arxiv.org/abs/1505.02097
[4] Zou et al. *J. R. Statist. Soc. B* (2005)

Acknowledgements

This work was supported by an NSF graduate research fellowship. We are also grateful to the travel support provided by the SAND 7 conference.