Stimulus Identification from fMRI scans

Charles Zheng and Yuval Benjamini

Stanford University, Department of Statistics

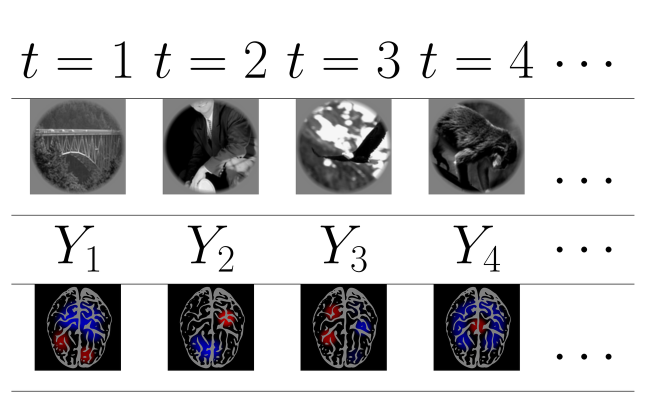
Overview

Seeking to explain the processes behind human perception, scientists employ *forward models* to model the causal relationship between stimulus 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.

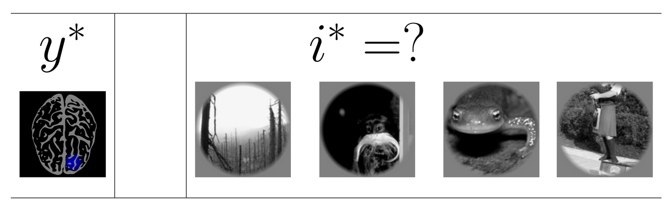
Using the data of Kay *et al.* as a motivating example, we consider the statistical problem of optimal identification. While identification superficially resembles a classification task (with many classes), it combines the challenge of multivariate regression with high-dimensional discrimination.

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 $p \approx 20000$



Identification



- Let S be a set of *new* stimuli–no repeats from the training set! Here $|S| = 120$.
- Scientist picks a stimulus i^* from S and measures the subject's reponse y^*
- Can the statistician *identify* $i^* \in S$ from y^*?

Previous Work

Kay (2008) and Vu (2011) reduce the dimensionality of Y (by applying a filter based on SNR) and consider a linear model

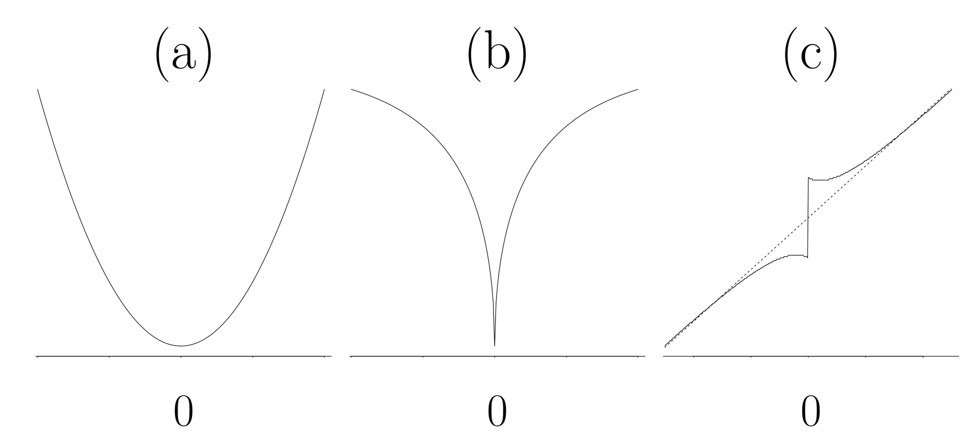
$$Y_{T\times p} = X_{T\times q} B_{q\times p} + E_{T\times p}$$

where X are a set of *image features* (Gabor filters), $q \approx 10000$.
They identify the stimuli based on the *maximum likelihood* (ML) principle:

- Obtain point estimates of coefficients B and noise covariance Σ_E
- E.g. B estimated using elastic net with CV (Zou 2005), shrinkage estimate for covariance
$$\hat{\Sigma}_E = \frac{1}{2}\hat{\text{Cov}}(Y - \hat{Y}) + \frac{1}{2}\text{diag}(\hat{\text{Cov}}(Y - \hat{Y}))$$
where $\hat{\text{Cov}}$ denotes sample covariance
- Obtain predicted means for test stimuli
$$\hat{\mu}_i^{te} = (x_i^{te})^T B$$
- Identify the stimulus i^* by
$$i^* = \text{argmin}_i (\hat{\mu}_i^{te} - y^*)^T \hat{\Sigma}_E^{-1} (\hat{\mu}_i^{te} - y^*)$$

Limitations of ML

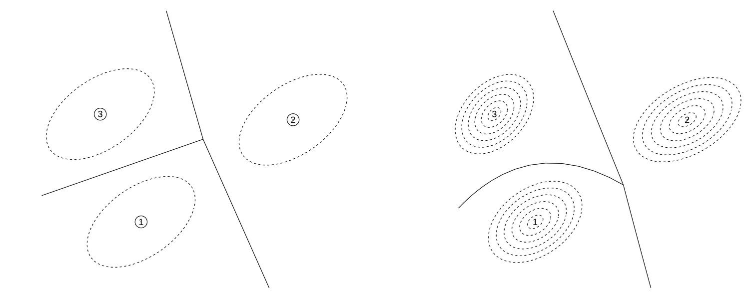
- Estimation of B are motivated by *prediction error*: the loss function for identification is different!
- *Fact:* Any estimate of B which is degenerate (low-rank) is suboptimal



(a) Squared error loss (vertical axis) and *(b)* loss function for identification, as a function of the difference between the true mean signal and the predicted signal.
(c) The optimal point estimate for identification (solid) vs the optimal point estimate for regression (dashed) diverge sharply at 0 in the one-dimensional case. ($B = 0$ a special case of degenerate B)

Empirical Bayes

- *Proposal*: improve on ML by accounting for the uncertainty of the estimate B.
- We consider a hierarchical model with a prior distribution on B, but we use the data to set the prior covariance Σ_B–hence we are being *empirical Bayesians*
- In contrast to ML, which results in *linear decision boundaries* (below: left), Empirical Bayes (EB) results in *quadratic boundaries* (below: right)



ML EB

Technical details

Model

- Noise $E_t \sim N(0, \Sigma_E)$ iid
- Coefficients $B_i \sim N(0, \sigma_i^2 I)$ for $i = 1, \ldots, p$
- X non-random

Estimate hyperparameters

- Use *eigenprism* (Janson 2015) to estimate $\theta_i^2 = ||B_i||^2$ for $i = 1, \ldots, p$
- Set $\sigma_i^2 = \hat{\theta}_i^2 / q$
- Estimate $\hat{B}$ as posterior mean
- Estimate Σ_E (same as in ML)

Compute posterior

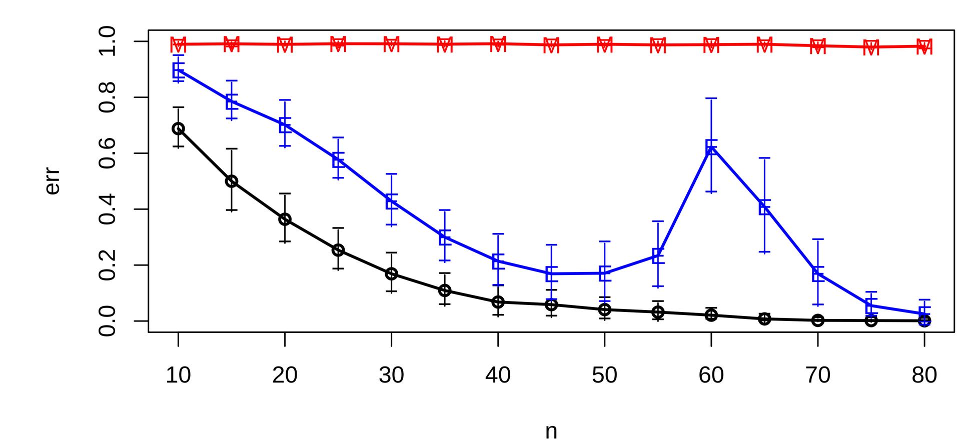
- Closed-form expressions for posteriors of B, μ_i^{te}
- Computational bottleneck: inverting a $pq \times pq$ covariance matrix

Apply Bayes rule

- *Uncertainty* in B is reflected as *added noise*
- Result: posterior $\text{Cov}(y^*|i^*)$ varies, hence *quadratic boundaries*

Simulation Results

- Parameters $p = q = 60$, random coefficients
- 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 covariance estimation methods..
- Required cost of $O((pq)^3)$ unacceptable for real data... develop tractable approximations
- Apply methods to data of Kay (2008)

Conclusions

- We determine theoretical limitations of Maximum Likelihood, and illustrate the promise of Empirical Bayes... but our own method still has obvious flaws
- Many challenges remain for optimal identification!

References

- Kay et al. *Nature* (2008)
- Vu et al. *Annals of Applied Statistics* (2011)
- Janson et al. (2015) http://arxiv.org/abs/1505.02097

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.