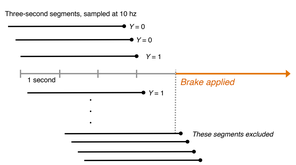
Dimension reduction and binary discrimination for naturalistic driving studies with heterogeneity

Brook Luers
Department of Statistics, University of Michigan
luers@umich.edu

IVBSS Naturalistic Driving Study

In a naturalistic driving study, high frequency kinematic measurements (e.g. speed, lane position) produce repeated measurements of drivers' behavior and driving context during their normal travel routines.

In the Integrated Vehicle-Based Safety Systems (IVBSS) field trial (Sayer. et al., 2011), 108 drivers in southeast Michigan drove instrumented study vehicles during a 40-day study period. This analysis is restricted to 35 of the IVBSS drivers and focuses on a binary measure of **braking behavior**: for each three seconds of driving in which the brake is not yet applied, does the driver apply the brake pedal within the next one second?



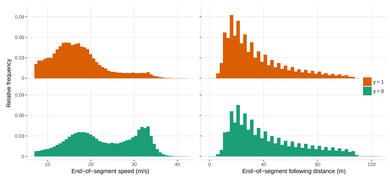
Kinematic measurements were recorded at 10 hertz, and segments in which the vehicle speed is below 7 m/s (about 15 mph) are excluded.

For this analysis, **driving context** within each segment is described by

- speed, in meters per second (10-hertz sampling produces 31 lagged measurements)
- radar-detected following distance, in meters
- steering angle, in degrees (only the 11 lagged measurements from the final one second in each segment)
- interaction between speed and following distance

There are **22,061,603 three-second driving segments** across the 35 drivers in this analysis.

Distributions of end-of-segment speed and following distance are displayed below.



Selection bias

- The relationship between driving context and behavior may be confounded by selection effects.
- Drivers' lifestyles may restrict the contexts in which they are observed: individuals who mostly drive on low-speed, urban roads will brake more frequently than drivers who commute every day on highways.
- Statistical models should highlight variation in driver behavior within the same context, eliminating these selection effects.

Heterogeneity and dimension reduction

These are **multilevel data**:

$$\begin{aligned} \text{covariates: } & x_{ij} \in \mathbb{R}^p \text{ (driving context)} \\ \text{response: } & y_{ij} \in \{0, 1\} \text{ (binary braking indicator)} \\ \text{individuals/groups: } & i = 1, \ldots, N \\ \text{repeated measurements: } & j = 1, \ldots, n_i \end{aligned}$$

These multilevel data can be used to discover heterogeneity in the relationship between driving context (x_{ij}) and braking behavior (y_{ij}).

Objective

Reduce the dimension of x_{ij} with **dimension reduction directions**, $v_1, v_2, v_3, \ldots \in \mathbb{R}^p$, which

- separate the conditional distributions $X \mid Y = 1$ and $X \mid Y = 0$
- reveal heterogeneity in the $X \mid Y$ distributions across individuals
- control for selection-induced heterogeneity in $X \mid Y$ across individuals

Differences of Covariances (DOC)

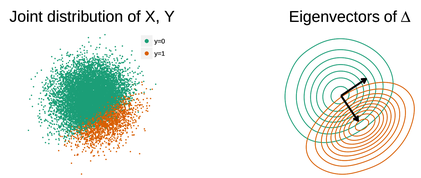
Suppose that, without loss of generality, $\mathrm{E}(X) = 0$ and $\mathrm{Cov}(X) = I$. When Y is binary, the Differences of Covariances (DOC) method uses

$$\delta := \mathrm{E}(X \mid Y = 1) - \mathrm{E}(X \mid Y = 0) \tag{1}$$

and the eigenvectors of

$$\Delta := \mathrm{Cov}(X \mid Y = 1) - \mathrm{Cov}(X \mid Y = 0) \tag{2}$$

as sufficient dimension reduction directions (Cook & Lee, 1999). The following schematic displays the eigenvectors of Δ when X follows a bivariate Gaussian distribution with $\mathrm{Cov}(X) = I$.



DOC and group heterogeneity

With multilevel data, the DOC method can be adapted to within-group and between-group means and covariances. Let D be a categorical random variable representing group membership (e.g. driver identifier in the IVBSS trial), and define the following quantities:

$$\mu_{yi} := \mathrm{E}(X \mid Y = y, D = i) \tag{3}$$
$$\delta_i := \mu_{1i} - \mu_{0i} \tag{4}$$
$$\Delta_i := \mathrm{Cov}(X \mid Y = 1, D = i) - \mathrm{Cov}(X \mid Y = 0, D = i) \tag{5}$$

By conditioning on Y and D, we can extend the DOC algorithm to obtain directions which discriminate $X \mid Y = 1$ and $X \mid Y = 0$ while describing heterogeneity across groups.

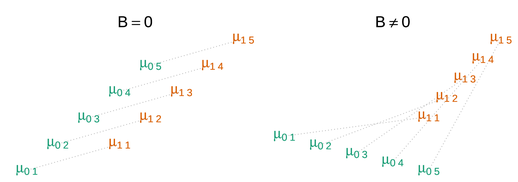
Sources of dimension reduction directions

- Difference in conditional means
$$\delta := \mathrm{E}(\delta_i)$$
- Covariance of conditional mean differences
$$\mathbf{B} := \mathrm{Cov}(\mu_{1i} - \mu_{0i})$$
- Within-group DOC
$$\mathbf{Q} := \mathrm{E}(\Delta_i)$$
- Between-group DOC:
$$\mathbf{R} := \mathrm{Cov}(\mu_{1i}) - \mathrm{Cov}(\mu_{0i})$$

Pooled, within-group, and between-group DOC relationship:

$$\Delta = \mathbf{Q} + \mathbf{R}$$

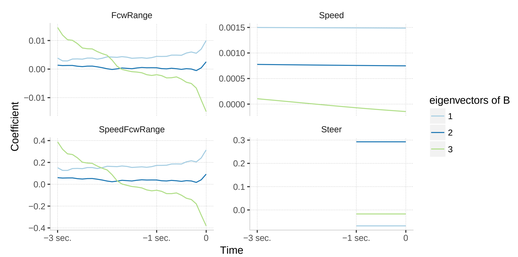
We conjecture that the eigenvectors of $\mathbf{B}$, $\mathbf{Q}$ and $\mathbf{R}$, along with δ, are dimension reduction directions which will describe the conditional distribution $X|Y$ as well as variation in this distribution across individuals. The following schematic illustrates populations in which $\mathbf{B}$ is zero or nonzero.



IVBSS Directions

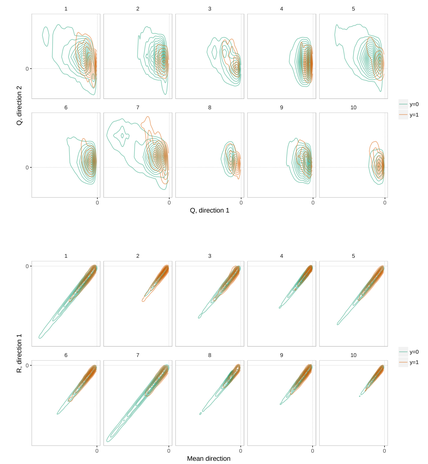
To stabilize estimates of the eigenvectors of $\mathbf{Q}, \mathbf{R}$ and $\mathbf{B}$, these matrices were restricted to the space spanned by the first 20 principal components of the marginal distribution of x_{ij}.

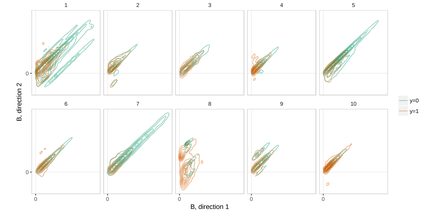
The first three eigenvectors of the sample estimate of $\mathbf{B}$ are displayed below for the IVBSS data.



Low-dimensional summaries of $X \mid Y = y$

The following figures display contours of the conditional distributions $(v_1^\intercal X, v_2^\intercal X) \mid Y = y$, where v_1 and v_2 are two dimension reduction directions. The panels in each figure display contours for a single IVBSS driver.





Findings:

- The eigenvectors of $\mathbf{Q}$ provide substantial separation between $X \mid Y = 1$ and $X \mid Y = 0$.
- The mean difference (δ) and the dominant eigenvector of $\mathbf{R}$ may be redundant.
- Variation among drivers is more apparent in the eigenvectors of $\mathbf{B}$.

References

Cook, R. D. & Lee, H. (1999). Dimension reduction in binary response regression. *Journal of the American Statistical Association, 94*(448), 1187–1200.

Sayer., J., LeBlanc, D., Bogard, S., Funkhouser, D., Bao, S., Buonarosa, M., & Blankespoor, A. (2011, June). ***Integrated vehicle-based safety systems field operational test final program report.*** (Tech. rep. No. UMTRI-2010-36). Transportation Research Institute, University of Michigan. Ann Arbor, Michigan.

Acknowledgements

IVBSS data provided by Carol Flannagan

Building a transportation data ecosystem (MIDAS Challenge Initiative, PI: C. Flannagan)

Advisor: Kerby Shedden (kshedden@umich.edu)