STAT 740: Constrained Estimation and Inference

Class meets: MW 9:00-10:15 (tentatively).

Prerequisites: Stat 530 and Stat 640.

Grading:

- Biweekly homework: 50%.
- Midterm presentation: 20%. The students will choose an appropriate paper to present to the class.
- Final presentation: 30%. Here the students will present a paper and also present (their own) simulations results related to the paper.

Topics: (I have no idea how many we'll be able to cover!)

1. *One-sided testing:* One of the first things we learn in Intro Stats is the two-sample t-test where $H_0 : \mu_1 = \mu_2$. If the alternative is $\mu_1 \neq \mu_2$, we do a two-sided test, but if we know μ_1 can't be larger than μ_2, a one-sided alternative $\mu_1 > \mu_2$ has higher power.

   This is generalized to the ANOVA case where $H_0 : \mu_1 = \cdots = \mu_k$. The general alternative uses an F-test, but what if it is known that $\mu_j \geq \mu_1$, $j = 2, \ldots, k$? For example, μ_1 might represent a placebo effect. Using a one-sided test gives higher power. To derive the estimator and the test statistic, we need:

2. *Cone projection theory:* Many types of estimation and testing involve linear inequality constraints so that the points in the null or alternative space form a polyhedral convex cone. To obtain the estimator one must compute a projection onto a cone, which is less straight-forward than a projection onto a linear space, and does not in general have a closed-form solution.

3. *Constrained parametric regression:* We start with the ordinary least-squares regression $\boldsymbol{y} = \boldsymbol{X\beta} + \sigma\boldsymbol{\varepsilon}$ and consider the constraints $\boldsymbol{A\beta} \geq \boldsymbol{0}$. For example, we could have a surface with interaction and we think that the expected response must be increasing in some or all of the predictors. Or we might want to fit a polynomial that is monotone over some range.

4. *Testing against a set of linear inequality constraints:* Often the constraints are *a priori*, but sometimes the research question involves the constraints. The constrained versus unconstrained test is more challenging than the constant versus constrained test.

5. *Isotonic regression:* There are many methods for estimating a regression function without specifying a parametric model. Most involve a tuning parameter such as a bandwidth or penalty that must be specified. However, we can assume only that the regression function is increasing, and no user-defined parameters are necessary. Further, the isotonic regression has a closed-form solution. For the test of constant versus increasing function, a statistic can be defined that has a null distribution of a mixture of chi-squared random variables.

6. *Convex regression and extensions:* If we specify only that a regression function must be convex, we no longer have a closed form, but the fit can be obtained via cone projection. The test of linear versus convex regression function is similar to the constant versus increasing test.

7. *Categorical data with order restrictions:* Comparing k binomial proportions with ordinal data, treatments with stratified data, odds ratios and monotone dependence.

8. *Decreasing and unimodal density or hazard function estimation:* Maximum likelihood estimation of a decreasing or unimodal density is considered, along with a recently developed least-squares criterion. Hazard functions can be assumed to be increasing, increasing convex, and bathtub-shaped.

9. *Smoothed shape-restricted regression:* Many types of scatterplot smoothers are in the literature: kernel regression, smoothing splines, penalized splines.... often we know something about the shape of the regression function as well. We look at several examples of smoothed monotone regression and extensions.