The geometry of the bootstrap

Patrick Breheny

September 13

Introduction

- In this lecture, we will describe estimators in a slightly different way that leads to an elegant geometrical representation that highlights the connections between the three nonparametric methods we have talked about so far
- Let $\mathbf{w}^*$ denote a vector of weights satisfying $0 \leq w_i^* \leq 1$ and $\sum_i^n w_i^* = 1$, and let $\hat{F}(\mathbf{w}^*)$ denote the CDF which places point mass w_i^* at point x_i
- Now, define our estimator as

$$\hat{\theta}^* = T(\hat{F}(\mathbf{w}^*))$$

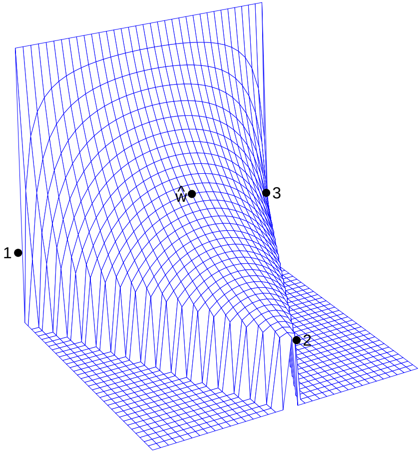
(note that we are now defining our estimator as a function of $\mathbf{w}^*$)

The simplex

- With this representation, our statistic is now a function whose domain is the set of vectors satisfying $0 \leq w_i^* \leq 1$ and $\sum_i^n w_i^* = 1$ (in geometry, such a set of vectors is known as a *simplex*)
- The center of the simplex is $\frac{1}{n}\mathbf{1}$, which we will denote $\hat{\mathbf{w}}$
- Note that $\hat{\theta} = T(\hat{F}(\hat{\mathbf{w}}))$
- Geometrically, the influence function studies the behavior of $\hat{\theta}^*$ in the infinitesimal region around $\hat{\mathbf{w}}$
- Geometrically, the jackknife studies the behavior of $\hat{\theta}^*$ as $\mathbf{w}$ is moved away from $\hat{\mathbf{w}}$ by an amount $1/n$ in the direction opposite the i^{th} vertex

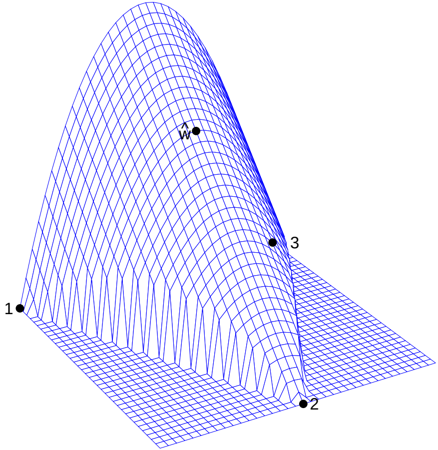
The surface

The estimator (in this case the variance) maps this simplex onto a surface:



The surface (plugin estimate)

If we use the plugin/ML estimator for the variance instead of the unbiased estimator, this surface looks like this:



Bootstrap resampling

- What is the bootstrap doing?
- The bootstrap draws $\mathbf{w}^*$ randomly from the multinomial distribution:

$$\mathbf{w}^* \sim \frac{1}{n}\mathrm{Mult}(n, \hat{\mathbf{w}})$$

- Note that

$$\mathbb{E}(\mathbf{w}^*) = \hat{\mathbf{w}}$$
$$\mathbb{V}(\mathbf{w}^*) = n^{-2}\mathbf{I} - n^{-1}\hat{\mathbf{w}}\hat{\mathbf{w}}'$$

Geometry of delta method, jackknife, and bootstrap

Geometrically, then, one can think of the three nonparametric approaches we have discussed so far as different ways of measuring this surface

- The delta method takes a tangent plane approximation
- The jackknife forms a hyperplane from the leave-one-out support points
- The bootstrap forms a weighted representation of the surface using the multinomial distribution

Equivalence of the three methods for linear statistics

- One might expect, then, that the three methods would agree if the surface was planar
- Suppose $\hat{\theta}^*$ is a linear statistic; then

$$\hat{\theta}^* = \sum_i^n w_i^* a(x_i) = \mathbf{w}'\mathbf{a},$$

  the equation for an n-dimensional hyperplane
- **Theorem:** For any linear statistic,

$$SE_\Delta = SE_{boot} = SE_{jack}\sqrt{\frac{n-1}{n}},$$

  where the $(n-1)/n$ term simply arises from the arbitrary historical decision of the jackknife to use $n-1$ in the denominator of $\tilde{s}^2$, and SE_Δ is the standard error of the functional delta method

Homework

Homework: A rather interesting 1987 paper by DiCiccio and Tibshirani proposes a confidence interval procedure called ABC, for "approximate bootstrap confidence" intervals. Their idea is to use influence functions to approximate what the bootstrap confidence intervals would be, without actually performing any bootstrap resampling. The `boot` package has a function `abc.ci` to compute these intervals. To use it, however, the function that calculates your statistic must take two arguments: the data and a vector of weights. Why would the function require weights, as opposed to the vector of indices that the usual `boot` function requires?

The bootstrap and Bayesian statistics

- Monte Carlo integration also plays a large role in Bayesian statistics; is there a connection?
- Indeed there is
- Suppose we specify a discrete distribution for our data, with $\mathbb{P}(x_i) = w_i$
- Now let us specify a Dirichlet prior for $\{\mathbf{w}\}$:

$$\mathbf{w} \sim \mathrm{Dir}_n(\alpha),$$

where α is a hyperparameter

The bootstrap and Bayesian statistics (cont'd)

- With this model specification, the posterior distribution of $\mathbf{w}$ is

$$\mathbf{w} \sim \mathrm{Dir}_n(\alpha + 1)$$

- As $\alpha \to 0$, the above is very similar to what we have in the bootstrap:

$$\mathbf{w} \sim \frac{1}{n}\mathrm{Mult}(n, \hat{\mathbf{w}}),$$

- Thus, we can think of the bootstrap as carrying out a kind of nonparametric Bayesian analysis in which our bootstrap replications are (nearly) draws from the posterior distribution of θ
- Of course, it cannot be considered a truly Bayesian approach, in that no one could legitimately specify a prior with point masses on the observed data points before collecting the data