Resampling Libraries in S-PLUS and R

Tim Hesterberg

Old and New S-PLUS / R interaction

- **Contrast:**
  - Old: minimal coordination between S-PLUS and R
  - New: cooperation
- **Old example: resampling**
  - Efron/Tibshirani `bootstrap.funs` (S-PLUS library, R package)
  - Davison/Hinkley/Canty `boot` (S-PLUS library, R package)
  - `S+Resample` (S-PLUS library)
- **New example: S+glars (later talk)**
  - Built on `lars` (Hastie & Efron) and `glmpath` (Young and Hastie)
  - Open source
  - Others to come…

S+Resample: Types of Resampling

- **Nonparametric bootstrap (1 and 2 samples)**
- **Parametric bootstrap**
- **Smoothed bootstrap**
- **Permutation test (1 and 2 samples)**
- **Cross validation**
- **Parametric bootstrap tests**
- **Bootstrap prediction errors**
- **Jackknife**
- **Influence**

Sampling Applications

- **Bootstrap**
  - 1 sample
  - 2 samples (difference or ratio)
  - Stratified
  - Sample by subject
- **Permutation Tests**
  - 1 sample
  - 2 samples (difference or ratio)
  - Stratified
  - Sample by subject
  - Permute subset of columns
- **Regression**
  - resample observations or residuals

S+Resample: sampling methods

- **Simple bootstrap**
- **Balanced bootstrap**
- **Reduced size (correct for bias)**
- **Bootknife (correct for bias)**
- **Smoothed bootstrap (correct for bias)**
- **Finite population**
- **Block bootstrap (time series)**
- **Importance sampling**
  - Other variance reduction – control variates, concomitants
- **Permutations, without replacement**

Confidence Intervals

- **Percentiles**
- **BC, BCa**
- **Tilting**
  - 1/37 as many replications as BCa – 60 instead of 2000
- **Bootstrap *t***
- ***T* intervals with bootstrap standard error**

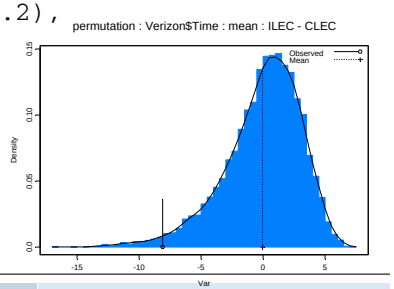
Ease of Use

- **Command line**
  - Difference in two trimmed means, typical syntax:

```
bootstrap(Verizon,
    function(data)
        mean(data$Time[data$Group == "ILEC"], trim=.2) -
        mean(data$Time[data$Group == "CLEC"], trim=.2),
        group = Verizon$Group) # stratified sampling
```

  - Simpler syntax:

```
bootstrap2(Verizon, mean(Time, trim=.2),
           treatment = Group)
```



- **GUI**
  - Introductory Statistics
    - Moore & McCabe
    - Free student version of S-PLUS