STAT 641: BOOTSTRAPPING METHODS

Jiyoun Myung

Department of Statistics and Biostatistics
California State University, East Bay

Spring 2021, Day 3

Bootstrap Estimate For Standard Error

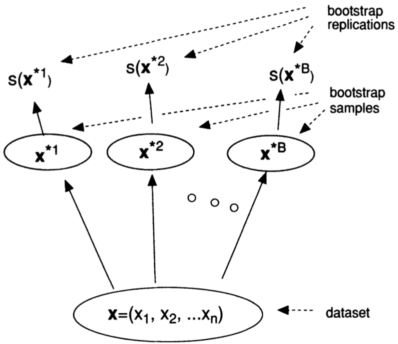


Figure 2.1. *Schematic of the bootstrap process for estimating the standard error of a statistic* $s(\mathbf{x})$. *B bootstrap samples are generated from the original data set. Each bootstrap sample has* n *elements, generated by sampling with replacement* n *times from the original data set. Bootstrap replicates* $s(\mathbf{x}^{*1}), s(\mathbf{x}^{*2}), \ldots s(\mathbf{x}^{*B})$ *are obtained by calculating the value of the statistic* $s(\mathbf{x})$ *on each bootstrap sample. Finally, the standard deviation of the values* $s(\mathbf{x}^{*1}), s(\mathbf{x}^{*2}), \ldots s(\mathbf{x}^{*B})$ *is our estimate of the standard error of* $s(\mathbf{x})$.

Problem

Airline accidents: According to the U.S. National Transportation Safety Board, the number of airline accidents by year from 1983 to 2006 were

```
df <- c(23, 16, 21, 24, 34, 30, 28, 24, 26, 18, 23, 23, 36,
        37, 49, 50, 51, 56, 46, 41, 54, 30, 40, 31)
```

1. For the sample data, compute the mean and its standard error (use sd()), and the median.
2. Using the **sample()** function, bootstrap estimates of the mean and median with estimates of their standard errors, using $B = 1000$. (Use **set.seed(1234)** to allow your results to be reproduced.)

Chapter 4: The empirical distribution and the plug-in principle

Empirical distribution function (EDF)

Having observed a random sample of size n from a probability distribution F,

$$F \rightarrow (x_1, x_2, \cdots, x_n), \tag{4.1}$$

the *empirical distribution function* $\hat{F}$ is defined to be the discrete distribution that puts probability $1/n$ on each value x_i, $i = 1, 2, \cdots, n$. In other words, $\hat{F}$ assigns to a set A in the sample space of x its empirical probability

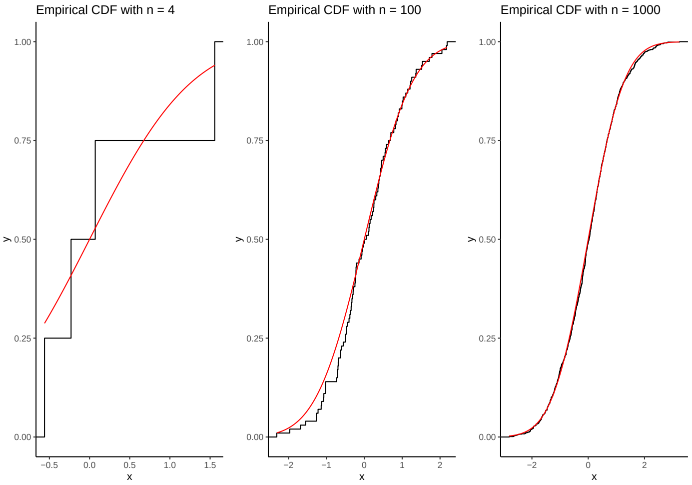
$$\widehat{\text{Prob}}\{A\} = \#\{x_i \in A\}/n, \tag{4.2}$$

the proportion of the observed sample $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ occurring in A. We will also write $\text{Prob}_{\hat{F}}\{A\}$ to indicate (4.2). The hat symbol "$\wedge$" always indicates quantities calculated from the observed data.

Empirical distribution function (EDF)

- $\widehat{F}$ is called the empirical distribution function (EDF) is an estimate of the entire distribution F.
- $\widehat{F}$ can be thought as the distribution which puts mass $\frac{1}{n}$ on each observation x_i (for values that occurs more than once in the sample the mass will be a multiple of $1/n$).
- It can be shown that $\widehat{F}$ is a nonparametric maximum likelihood estimator of F which justifies to estimate F by $\widehat{F}$ if no other information about F is available (such as e.g. F belongs to a parametric family).
- Aside: (A theoretical result) Let $A \subseteq \mathbb{R}$. Then we have $\widehat{\text{Prob}}\{A\} \to P(A)$ as $n \to \infty$.

Example of the EDF



Properties of EDF

For any fixed value of x, we have

- $E\left[\widehat{F}_n(x)\right] = F(x)$,
- $\mathrm{Var}\left[\widehat{F}_n(x)\right] = \frac{1}{n}F(x)[1 - F(x)]$.

The Plug-In Principle

The bootstrap method is a direct application of the plug-in principle.

The plug-in principle is a simple method of estimating parameters from samples. The *plug-in estimate* of a parameter $\theta = t(F)$ is defined to be

$$\hat{\theta} = t(\hat{F}).$$

In other words, we estimate the function $\theta = t(F)$ of the probability distribution F by the same function of the empirical distribution $\hat{F}$, $\hat{\theta} = t(\hat{F})$. (Statistics like (4.13) that are used to estimate parameters are sometimes called *summary statistics*, as well as *estimates* and *estimators*.)

The Plug-In Principle

In general, the plug-in estimate of an expectation $\theta = E_F(x)$ is

$$\hat{\theta} = E_{\hat{F}}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

How good is the plug-in principle? It Is usually quite good, if the only available information about F comes from the sample $\mathbf{x}$.

How far can the plug-in principle be trusted? "Pretty far" is a reasonable summary of current bootstrap research. Simple bootstrap ideas, like resampling from the empirical distribution, work surprisingly well in a surprisingly large catalog of cases.

Chapter 5: Standard errors and estimated standard errors

Standard Error of Mean

Suppose that x is a real-valued random variable with probability distribution F. Let us denote the expectation and variance of F by the symbol μ_F and σ_F^2 respectively,

$$\mu_F = E_F(x), \quad \sigma_F^2 = \mathsf{var}_F(x) = E\left[(x - \mu_F)^2\right].$$

also we write $x \sim (\mu_F, \sigma_F^2)$.

The standard error of the mean $\bar{x}$ is the square root of the variance of $\bar{x}$,

$$\mathsf{se}(\bar{x}) = [\mathsf{var}_F(\bar{x})]^{1/2} = \sigma_F/\sqrt{n}.$$

Plug-In Estimate of Standard Error of Mean

Suppose that we have in hand a random sample of numbers $F \to x_1, x_2, \ldots, x_n$. We want to estimate the standard error of the mean

$$\mathsf{se}(\bar{x}) = \sigma_F/\sqrt{n} = \sqrt{E\left[(x-\mu_F)^2\right]/n}.$$

THe plug-in estimate of σ_F is

$$\hat{\sigma} = \sigma_{\hat{F}} = \left\{ \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{1/2},$$

since $\mu_{\hat{F}} = \bar{x}$ and $E_{\hat{F}} g(x) = \sum_{i=1}^{n} g(x_i)$ for any function g. Thus, the estimated standard error is

$$\widehat{\mathsf{se}}(\bar{x}) = \mathsf{se}_{\hat{F}}(\bar{x}) = \sigma_{\hat{F}}/\sqrt{n} = \left\{ \frac{1}{n^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{1/2}.$$

Chapter 6: The bootstrap estimate of standard error

Background

Resampling methods go back to the 1930's. In 1979, Efron coined Bootstrap and simple approximation to the jackknife (developed by Tukey) for estimating the standard error of $\hat{\theta}$. Efron was the first to show the board applications of bootstrapping to confidence intervals, hypothesis testing and more complex problems.

The bootstrap estimate of standard error requires no theoretical calculations, and is available no matter how mathematically complicated the estimator $\hat{\theta}s(\mathbf{x})$ may be.

In many situations, bootstrap is as good or better than the jackknife as a resampling procedure.

Jackknife and its connection to bootstrap in a few weeks.

The Bootstrap principle

The basic idea of the bootstrapping method is that, in absence of any other information about the distribution, the observed sample contains all the available information about the underlying distribution, and hence resampling the sample is the best guide to what can be expected from resampling from the distribution.

Suppose that a sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is used to estimate a parameter of the distribution and let $\hat{\theta} = s(\mathbf{x})$ be a statistic that estimates. For the purpose of statistical inference on θ, we are interested in the sampling distribution of $\hat{\theta}$ (or certain aspects of it) so as to assess the accuracy of our estimator or to set confidence intervals for our estimate. In many applications, however, the sampling distribution of $\hat{\theta}$ is intractable.

The Bootstrap principle

- If the true distribution F were known, we could draw samples $\mathbf{x}^b, b = 1, \ldots, B$ from F and use Monte Carlo methods to estimate the sampling distribution of our estimate $\hat{\theta}$.
- Usually, since F is unknown and we cannot sample from it, the bootstrapping idea suggests to resample the original sample instead. This distribution from which the bootstrap sample is drawn is the empirical distribution