Chapter 7 - Bootstrap

Yuan Huang

Feb 3, 2013

Standard error of an estimator

Every time we report an estimate, we should give some indication about its precision. Here we introduce a usual measure : the standard error.

Sampling Distribution Definition and Features

- A **sampling distribution** is the probability distribution of a sample statistic. It describes how values of a sample statistic vary across all possible random samples of a specific size that can be taken from a population.
- The **mean value of a sampling distribution** is the mean value of a sample statistic over all possible random samples. For the big five scenarios, this mean equals the value of the population parameter.
- The **standard deviation of a sampling distribution** measures the variation among possible values of the sample statistic over all possible random samples. When referring to such a standard deviation, we include the name of the statistic being studied—for example, *the standard deviation of the mean.*
- The term **standard error** is used to describe the estimated value of the standard deviation of a statistic. Because the formula is different for each statistic, we include the name of the statistic—for example, *the standard error of the mean.*

Figure: Mind on Statistics, 3rd, P348

	Individual	Mean
Population	σ	$\sigma/\sqrt{n}$
	(standard deviation of population)	(standard deviation of $\bar{X}$)
Sample	s	$s/\sqrt{n}$
	(standard deviation of sample)	(standard error of $\bar{X}$)

Bootstrap

- Sometimes we cannot find a specific formula to estimate the standard error of the estimator (e.g. $\hat{\sigma}^2$).
- In this case we can use bootstrap to find an (estimated) standard error.
- Let's assume that we have a random sample from $N(\mu, \sigma^2)$ where μ and σ^2 are both unknown. We estimate them by $\hat{\mu} = \bar{x} = 0.1126$ and $\hat{\sigma}^2 = s^2 = 8.059$.
- In this case, we assume the distribution to be normal. Hence we can apply the parametric Bootstrap. We then use computer to calculate B bootstrap samples from $N(0.1126, 8.059)$. Another way is you can do the usual Bootstrap. In this case, you treat the original sample you collect as the population and doing resample from the original sample.

Bootstrap

- First bootstrap sample: $x_1^*, x_2^*, \ldots, x_n^*$, estimate $\hat{\mu}_1^*, \hat{\sigma}_1^{2*}$.
- Second bootstrap sample: $x_1^*, x_2^*, \ldots, x_n^*$, estimate $\hat{\mu}_2^*, \hat{\sigma}_2^{2*}$.

$\vdots$

- Bth bootstrap sample: $x_1^*, x_2^*, \ldots, x_n^*$, estimate $\hat{\mu}_B^*, \hat{\sigma}_B^{2*}$.

Bootstrap

- In each sample we calculate a new estimate for the mean $\hat{\mu}_B^*$ and a new estimate for the variance $\hat{\sigma}_B^{2*}$.
- We average all the means and all the variances to obtain

$$\bar{\hat{\mu}}^* = \frac{\sum_{i=1}^{B} \hat{\mu}_i^*}{B}, \bar{\hat{\sigma}}^{2*} = \frac{\sum_{i=1}^{B} \hat{\sigma}_i^{2*}}{B}$$

- The bootstrap estimated of the standard error of the two estimators are:

$$S_{\hat{\mu}} = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}\left(\hat{\mu}_i^* - \bar{\hat{\mu}}^*\right)^2}, S_{\hat{\sigma}^2} = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}\left(\hat{\sigma}_i^{2*} - \bar{\hat{\sigma}}^{2*}\right)^2}$$