# Introduction to the bootstrap -

In this course we apply a computer simulation technique called the "bootstrap" or "resampling" to find standard errors and confidence intervals -

Confidence intervals are usually derived using probability theory (Gaussian model for example) or using approximate distributions of the estimators (CLT or Gaussian approximation of the MLE distributions). When the probability/approximation calculations are intractable, we can replace theoretical calculations by Monte Carlo simulations -

The power of the bootstrap lies in the fact that it can be easily applied to assess the accuracy of predictions for a wide range of learning methods -

Suppose that $\hat{\theta}$ is an estimator of $\theta$, based on $n$ iid variables $X_1, \ldots, X_n$. The form of the confidence interval based on $\hat{\theta}$ requires that we know the (approximate) distribution of $\hat{\theta}$.

For example the "two standard deviation rule" $\hat{\theta} \pm 2\,\widehat{s.e.}(\hat{\theta})$ is valid if one is confident that $\hat{\theta}$ has an approximately normal distribution or an exact Student distribution, and if the estimated standard error of $\hat{\theta}$ ($\widehat{s.e.}(\hat{\theta})$) can be calculated -

If we don't know the distribution of $\hat{\theta}$, the idea of the bootstrap is to simulate it.

Idea: How can we simulate samples that are similar to the original sample $X_1, \ldots, X_n$?

↪ we can simulate sampling from the population by sampling from the sample = resampling.

How? we sample from the $n$ observations $n$ times with replacement

= a bootstrap sample

**Definition** A bootstrap sample $(X_1^*, X_2^*, \dots, X_n^*)$ is an i.i.d. sample from the empirical probability that puts mass $\frac{1}{n}$ at each data point $X_1, \dots, X_n$.

↪ each $X_i^*$ is drawn with replacement from $(X_1, \dots, X_n)$

↪ the common CDF of the $X_i^*$s is the ECDF of the $X_i$. As the ECDF is an estimator of the actual distribution of $(X_1, \dots, X_n)$, the bootstrap resampling tries to simulate the original sampling.

From the bootstrap sample $(X_1^*, \dots, X_n^*)$ we calculate the bootstrap estimator $\hat{\theta}^*$ of $\theta$, replacing $X_1, \dots, X_n$ in the formula of $\hat{\theta}$ by the n bootstrap sample values $X_1^*, \dots, X_n^*$.

This procedure is repeated a large number B of times

$$(X_1^*, \dots, X_n^*)^1 \longrightarrow \hat{\theta}^{*(1)}$$
$$(X_1^*, \dots, X_n^*)^2 \longrightarrow \hat{\theta}^{*(2)}$$
$$\vdots \qquad\qquad \vdots$$
$$(X_1^*, \dots, X_n^*)^B \longrightarrow \hat{\theta}^{*(B)}$$

The B estimates $(\hat{\theta}^{*(1)}, \hat{\theta}^{*(2)}, \dots, \hat{\theta}^{*(B)})$ can be regarded as describing an empirical sampling distribution of $\hat{\theta}$.

**Definition** $(\hat{\theta}^{*(1)}, \dots \hat{\theta}^{*(B)})$ is the bootstrap sampling distribution of $\hat{\theta}$

(We can represent it with an histogram)

From the bootstrap sampling distribution, we can estimate the variance of $\hat{\theta}$ and derive confidence intervals for $\theta$.

Bootstrap estimate of $s.e.(\hat{\theta})$

$$\widehat{Var}_{boot}(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{*(b)} - \left(\frac{1}{B}\sum_{j=1}^{B}\hat{\theta}^{*(j)}\right)\right)^2$$

$$\widehat{s.e}_{boot}(\hat{\theta}) = \sqrt{\widehat{Var}_{boot}(\hat{\theta})}$$

## Bootstrap Confidence Intervals

When the bootstrap is valid, the mathematical theorem says that the distribution of $\hat{\theta} - \theta$ can be approximated by the bootstrap distribution of $\hat{\theta}^* - \hat{\theta}$.

Let $b^*_{\alpha/2}$ and $b^*_{1-\alpha/2}$ the $\alpha/2$ and $1-\alpha/2$ sample quantiles of $(\hat{\theta}^{*(1)} - \hat{\theta}, \dots, \hat{\theta}^{*(B)} - \hat{\theta})$. Note that there is no reason that $b^*_{1-\alpha/2} = -b^*_{\alpha/2}$ since we do not expect the bootstrap distribution to be symmetric.

Then $1-\alpha = \mathbb{P}\left(b^*_{\alpha/2} \leq \hat{\theta}^* - \hat{\theta} \leq b^*_{1-\alpha/2}\right)$

$$\approx \mathbb{P}\left(b^*_{\alpha/2} \leq \hat{\theta} - \theta \leq b^*_{1-\alpha/2}\right)$$

$$\Rightarrow IC = \left[\hat{\theta} - b^*_{1-\alpha/2} \; ; \; \hat{\theta} - b^*_{\alpha/2}\right] \text{ is an approximate bootstrap confidence interval with confidence level (approximate) } 1-\alpha.$$

Choice of $B$: $B \geq 200$ to compute $\widehat{Var}_{boot}$

$B \geq 1000$ to compute $b^*_{\alpha/2}$ and $b^*_{1-\alpha/2}$.