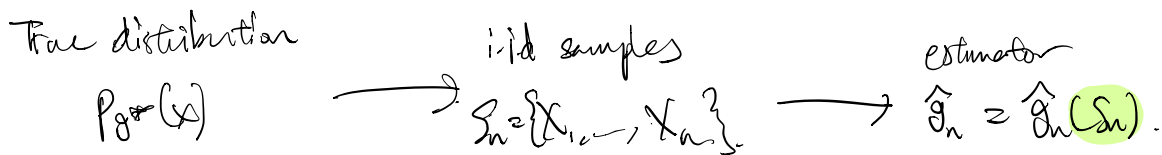


Bootstrap



Goal: Estimate θ^* .

Q. How close is $\hat{\theta}_n$ to θ^* ?

Suppose we had an oracle that told us $P_{\theta^*}(x)$, then we also know the distribution of $\hat{\theta}_n$.

In particular, we know $\text{Var}_{P_{\theta^*}}(\hat{\theta}_n)$.

Idea 1: Replace P_{θ^*} with $\hat{P}_n =$ uniform distribution on $\{X_1, \dots, X_n\}$, i.e. the empirical distribution.

If \hat{P}_n is "close" to P_{θ^*} , then

$$\text{Var}_{P_{\theta^*}}(\hat{\theta}_n) \approx \text{Var}_{\hat{P}_n}(\hat{\theta}_n) \quad \text{---} \quad \hat{\theta}_n(X_1^*, \dots, X_n^*)$$

$X_i^* \sim \text{i.i.d. } \hat{P}_n$

Idea 2: Sample from \hat{P}_n in order to get an approximate of the functional (statistic) that we want to compute.

$$\text{In our case, } \widehat{\text{Var}}_{\hat{P}_n}(\hat{\theta}_n) = \frac{1}{B} \sum_{j=1}^B (\hat{\theta}_n^* - \bar{\theta}_n^*)^2$$

If B is large enough, then

$$\text{Var}_{\hat{P}_n}(\hat{\theta}_n) \approx \widehat{\text{Var}}_{\hat{P}_n}(\hat{\theta}_n)$$

$$\text{Hence, } \text{Var}_{\hat{p}_n}(\hat{\theta}_n) \approx \underset{\uparrow}{\text{Var}_{\hat{p}_n}(\hat{\theta}_n)} \approx \underset{\uparrow}{\widehat{\text{Var}}_{\hat{p}_n}(\hat{\theta}_n)}$$

bootstrap estimate.

Rem! ① Idea applies to other functionals of distributions of $\hat{\theta}_n$, like quantile.

② If we can bootstrap to get CIs, then these are only asymptotically valid.

Bootstrap algorithm:

Input: Sample X_1, \dots, X_n , estimator $\hat{\theta}_n$.

Repeat B times:

- Draw dataset $S_n^{(b)*} = \{X_1^{(b)*}, \dots, X_n^{(b)*}\}$ iid from \hat{p}_n .
- Compute $\hat{\theta}_n^{(b)*} = \hat{\theta}_n(S_n^{(b)*})$.

Use the distribution of $\{\hat{\theta}_n^{(1)*}, \dots, \hat{\theta}_n^{(B)*}\}$ to estimate some statistic that we want.