Recall Have iid samples $x_1, \dots, x_n$ from F which depends on parameter θ. Given an estimator of θ, $\hat{\theta} = s(\underline{x})$ we can estimate the sampling distribution of $\hat{\theta}$ with the <u>nonparametric bootstrap</u> via:

(1) Sample "new" data $\underline{x}^{*b}$ $[n \times 1]$ from $\{x_1, \dots, x_n\}$ with replacement for $b = 1, \dots, B$
$\rightarrow$ bootstrap dataset

(2) Form

$\rightarrow$ bootstrap "replicate" of $\hat{\theta}$

$$\hat{\theta}^*(b) = s(\underline{x}^{*b}) \quad \text{for } b = 1, \dots, B$$

(3) $\hat{\theta}^*(1), \dots, \hat{\theta}^*(B)$ can be thought of as samples from

$$g_{\hat{\theta}} = \text{sampling dist of } \hat{\theta}$$

For example, to estimate standard error of $\hat{\theta}$ could use:

$$\hat{SE}(\hat{\theta}) = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}^*(b) - \frac{1}{B} \sum_{j=1}^{B} \hat{\theta}^*(j) \right)^2}$$

$\Rightarrow$ approx 95% CI $\quad \hat{\theta} \pm 2 \hat{SE}(\hat{\theta})$

Remark.

How big should B be? Bigger = better, usually few hundred/thousand samples.

Parametric bootstrap

In nonp. boot we generate new datasets by sampling $\underline{X}^* \sim F_n$, in parametric bootstrap we generate new bootstrap datasets from a parametric dist $\underline{X}^* \sim F_{\hat{\theta}}$

where $X_i \sim F_\Theta$, using ~~a plug-in estimate~~ of Θ.

$\boxed{\text{Ex}}$ Have data $\underline{x}$ from exponential dist w/ rate λ

$$f_\lambda(x) = \lambda e^{-\lambda x}, \quad x > 0$$

An estimator for rate might be $\hat{\lambda} = \frac{1}{\bar{x}}$

To assess uncertainty in $\hat{\lambda}$ could generate

n iid sample from $\text{Exp}(\hat{\lambda})$, recalculate

$\hat{\lambda}^*$, do this $b = 1, \ldots, B$

$$\hat{\lambda}^*(1), \ldots, \hat{\lambda}^*(B)$$

$\boxed{8.4 \text{ Bootstrap for regression}}$

Suppose we have regression setup with samples

$\underline{z}_1 = (\underline{x}_1, y_1), \ldots, \underline{z}_n = (\underline{x}_n, y_n)$ are obs from

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$$

2 ways to estimate dist of $\hat{\underline{\beta}}_{OLS}$ using bootstrap:

$\boxed{\text{Bootstrapping pairs}}$

Algorithm: For $b = 1, \ldots, B$

(1) Select indep bootstrap samples from $\underline{z}_1, \ldots, \underline{z}_n$,

$\underline{z}^{*1}, \ldots, \underline{z}^{*n}$, with repl.

② Do OLS based on $\{z^*_i\}_{i=1}^n$, call $\hat{\beta}^*(b)$