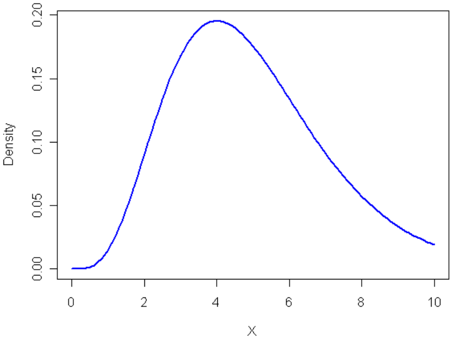
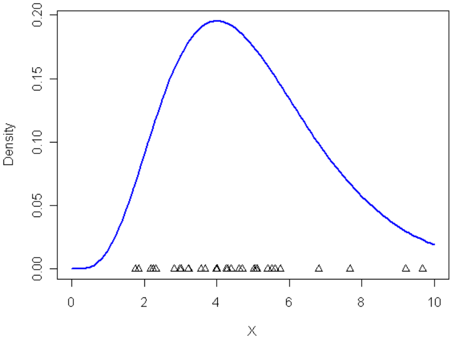
Plug-in Principle



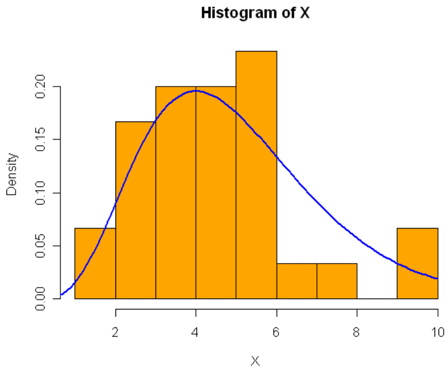
- Goal: Learn something about the underlying probability distribution $F(x)$.

Plug-in Principle



- Goal: Learn something about the underlying probability distribution $F(x)$.
- Observe data $X_1, \ldots, X_n$ drawn independently from $F(x)$.

Plug-in Principle



- Goal: Learn something about the underlying probability distribution $F(x)$.
- Observe data $X_1, \ldots, X_n$ drawn independently from $F(x)$.
- Use the empirical distribution $\hat{F}(x)$ as a surrogate for F. $\hat{F}$ puts probability mass $1/n$ to each of $X_1, \ldots, X_n$.

Plug-in Principle

True CDF:

$$F(x) = P(X \leq x)$$

Empirical CDF:

$$\hat{F}_n(x) = \sum_{i=1}^{n} I(X_i \leq x).$$

We are interested in the quantity $\theta(F)$. Simply estimate using

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

Plug-in Principle Examples

Example 1

density

$$\theta(F) = E_F(X) = \int x\, dF(x) = \int x f(x)\, dx$$

$$\hat{\theta} = \theta(\hat{F}_n) = \frac{1}{n}\sum_{i=1}^n X_i = \text{sample mean} \triangleq \bar{X}$$

Example 2

$$\theta(F) = Var_F(X) = \int (x - E_F X)^2 dF(x)$$

$$= \int (X - E_F X)^2 f(x)\, dx$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$

Example 3.

$$\theta(F) = s.e._F(\bar{X}) = \sqrt{Var_F(\bar{X})}.$$

Any statistic $S(\underset{\sim}{X}) = S(X_1, \ldots, X_n)$, $\theta(F) = Var_F(S(\underset{\sim}{X}))$.

Bootstrap as Plug-in Principle

Let $S(\mathbf{X}) = S(X_1, \dots, X_n)$ be any statistic, $\theta = \sqrt{Var_F(S)}$ be the standard error of S.

What does “plugging in $\hat{F}$ for F” mean in this case?

Real World

$$F_\mu \longrightarrow X_1, X_2, \dots X_n \longrightarrow \hat{\mu} = S(\tilde{X})$$

Bootstrap World

$\hat{F}_n$ mass $\frac{1}{n}$ on each X_i

$$\hat{F}_n \xrightarrow{\text{i.i.d.}} X_1^*, \dots X_n^* \longrightarrow \hat{\mu}^* = S^* = S(\tilde{X}^*)$$

$$\theta(F) = s.e._F[S(\tilde{X})]$$

$$\hat{\theta} = \theta(\hat{F}_n) = \sqrt{Var_{\hat{F}_n}[S(\tilde{X}^*)]}$$

Bootstrap estimate of std. err.

Bootstrap Estimate of Standard Error

$$\hat{F}_n \to \underset{\sim}{X}^{*,1} = (X_1^{*,1}, \ldots, X_n^{*,1}) \to S^{*,1}$$

$$\hat{F}_n \to \underset{\sim}{X}^{*,2} = (X_1^{*,2}, \ldots, X_n^{*,2}) \to S^{*,2}$$

$$\vdots$$

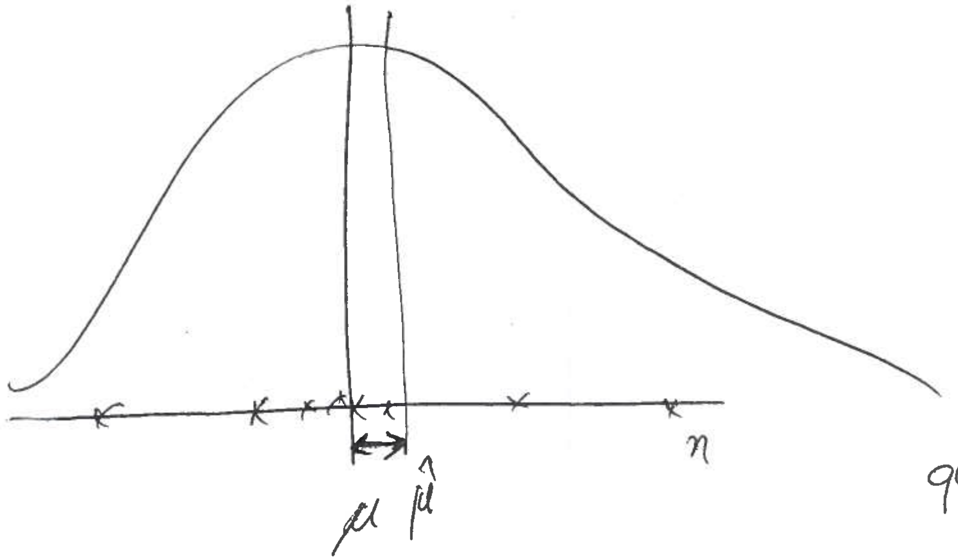
$$\hat{F}_n \to \underset{\sim}{X}^{*,B} = (X_1^{*,B}, \ldots, X_n^{*,B}) \to S^{*,B}$$

$$\hat{\theta}^B = \sum_{b=1}^{B} \left(S^{*,b} - S^{*,\cdot}\right)^2 \Big/ B-1$$

$$S^{*,\cdot} = \frac{1}{B} \sum_{b=1}^{B} S^{*,b}.$$

As $B \to \infty$, $\hat{\theta}^B \longrightarrow \hat{\theta}$ "Ideal Bootstrap estimate."

Bootstrap Percentile Confidence Intervals



Standard t-interval.

$$\hat{\mu} \pm t_{1-\alpha, n-1} \, s.e.(\hat{\mu})$$

Confidence intervals are based on quantiles of the difference $\hat{\mu} - \mu$.

Basic Bootstrap percentile intervals: Use $\underbrace{S^* - \hat{S}}_{\text{the quanties of}}$ for the quantiles of $\hat{S} - S$.

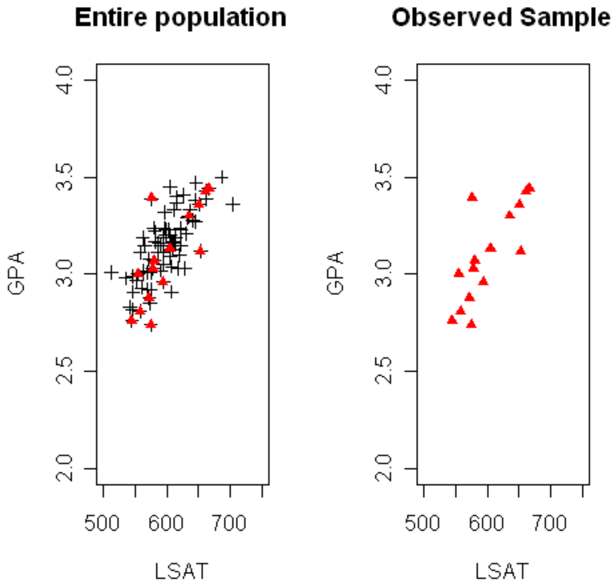
True interval.

$$P_F\left(c_{n,\alpha}(F) \le \hat{\theta} - \theta(F) \le c_{n,1-\alpha}(F) \right) = 1 - 2\alpha.$$

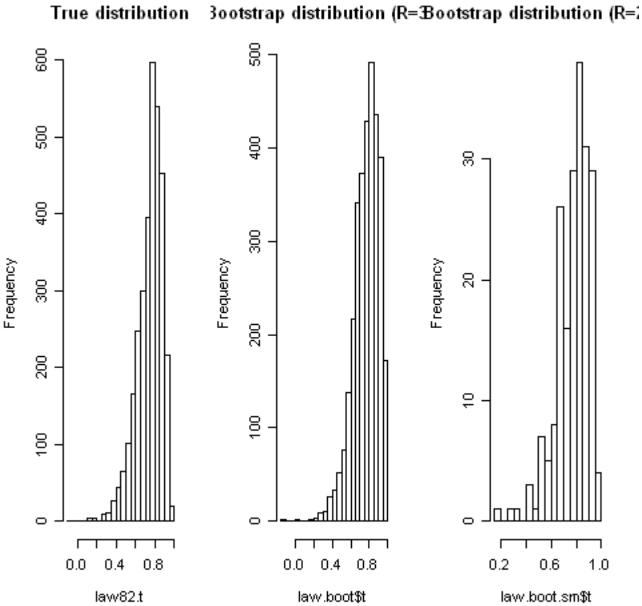
Bootstrap interval:

$$P_{\hat{F}_n}\left(c_{n,\alpha}(\hat{F}_n) \le \hat{\theta}^* - \theta(\hat{F}_n) \le c_{n,1-\alpha}(\hat{F}_n) \right) = 1 - 2\alpha.$$

Example: Law School Data

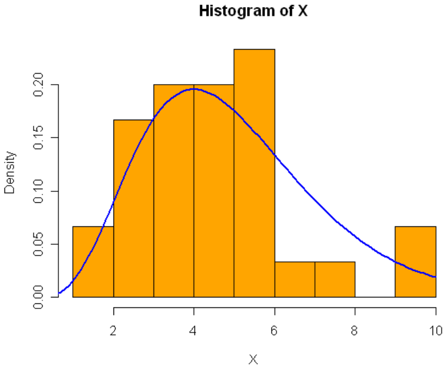


Example: Law School Data



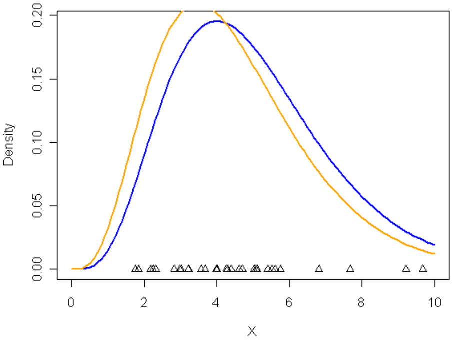
Parametric Bootstrap

Instead of plugging in $\hat{F}$, use $F_{\hat{\theta}}$.



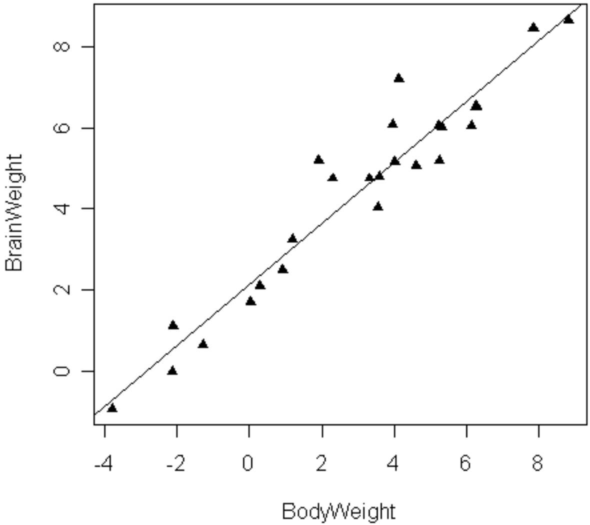
Parametric Bootstrap

Instead of plugging in $\hat{F}$, use $F_{\hat{\theta}}$.



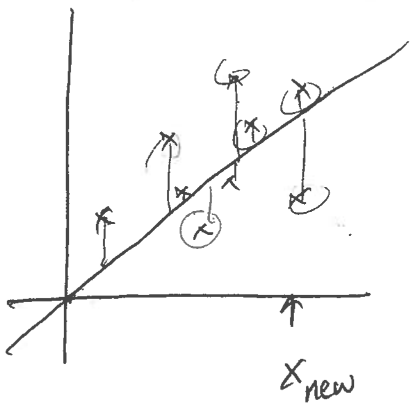
- Nonparametric Bootstrap avoids limiting assumptions.
- Parametric Bootstrap smoothes out discrete data, and is less sensitive to outliers.

Example: Bootstrapping Regressions



Example: Bootstrapping Regressions

$X = (X_1, \dots, X_n)^T$
$Y = (Y_1, \dots, Y_n)^T$



<u>Model</u>: $Y_i = \beta X_i + \varepsilon_i$,

Errors $\varepsilon_i \overset{iid}{\sim} F$ $\quad Var(\varepsilon_i) = \sigma^2$.

$$\hat{\beta} = \frac{X^T Y}{X^T X}$$ least squares estimate of β.

conf. int for β, s.t. err.

$$\hat{Y}_{new} = X_{new} \hat{\beta}$$

<u>Real World</u>: $X \xrightarrow{\beta} X\beta \xrightarrow{F} Y = X\beta + \varepsilon$.

<u>Bootstrap World</u>: sub $\hat{\beta}$ for β.

$\hat{F}_n$ empirical cdf of residuals $Y_i - \hat{Y}_i$ for F

$X \xrightarrow{\hat{\beta}} X\hat{\beta} \xrightarrow{\hat{F}_n} Y^* = X\hat{\beta} + \varepsilon^*$

$$\begin{tikzcd}
F \arrow[d] & \hat{F}_n \arrow[d] & \\
(X_i, Y_i)_{i=1,\dots,n} \arrow[ur] \arrow[d] & (X_i^*, Y_i^*)_{i=1,\dots,n} \arrow[d] & \\
\hat{\beta} = S(\underset{\sim}{X}, \underset{\sim}{Y}) & \hat{\beta}^* = S(X^*, Y^*) . & B \text{ times.}
\end{tikzcd}$$

Example: Bootstrapping Regressions