Bootstrap

We know that asymptotically $\sqrt{n}(\hat{\theta}-\theta) \sim N(0, I(\theta)^{-1})$

but for small n, the sampling distribution can be quite far from this — we may have bias, variance $\neq \frac{I(\theta)^{-1}}{n}$ and a distribution that is far from normal.

$\Rightarrow$ CIs we construct don't have correct coverage.

For some $\hat{\theta}$ we can compute $V(\hat{\theta})$ for finite n analytically

e.g. $\hat{\theta} = \bar{x} \Rightarrow V(\hat{\theta}) = \frac{\sigma^2}{n}$

but for $\hat{\theta}$ that have a complex dependence on x^n it may not be so easy.

* We can use the Delta-method to get an approximate $V(\hat{\theta})$ if $\hat{\theta} = g(\hat{\eta})$ and $\hat{\eta}$ has known form for $V(\hat{\eta})$ and $E(\hat{\eta})$

$g(\hat{\eta})$ approximated by Taylor expansion near $E(\hat{\eta})$

$$\Rightarrow g(\hat{\eta}) \approx g(E(\hat{\eta})) + g'(E(\hat{\eta}))(\hat{\eta} - E(\hat{\eta}))$$

$$\Rightarrow \begin{cases} E(g(\hat{\eta})) \approx g(E(\hat{\eta})) \\ V(g(\hat{\eta})) \approx (g'(E(\hat{\eta})))^2 V(\hat{\eta}) \end{cases}$$

May not be a good approximation if g is far from linear near $E(\hat{\eta})$.

Special case $\hat{\theta} = g(\bar{x})$!

Alternative to this is __bootstrap__

__Idea__ $X_i \sim^{iid} F$ with parameter of interest $\theta(F)$

observe $X^n = (X_1, \ldots, X_n)$

$\Rightarrow$ estimate $\hat\theta_n = \hat\theta(X^n)$

To get sampling distribution of $\hat\theta_n$ we would need repeated sampling from the unknown F

$$\left.\begin{array}{l} X^{n1} \text{ from } F \to \hat\theta_n^1 \\ X^{n2} \text{ from } F \to \hat\theta_n^2 \\ \vdots \\ X^{nM} \text{ from } F \to \hat\theta_n^M \end{array}\right\}$$

Look at distribution of $\hat\theta_n^1, \ldots, \hat\theta_n^M$

$\Rightarrow \frac{1}{M}\sum_{m=1}^{M} \hat\theta_n^m - \theta(F)$ = estimate of bias

$E(\hat\theta_n) - \theta(F)$ etc

but we can't do repeated sampling from F

we only have one data set.

However, F_n = empirical distribution with point mass $\frac{1}{n}$ at each X_i

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \le t\}$$

we can sample from

$X_i^* \sim iid\ F_n$, $i = 1, \ldots, n$ $\iff$ draw X_i^* from $(X_1, \ldots, X_n)$ with replacement.

$\Rightarrow X^* \Rightarrow \hat\theta(X^*)$ is an estimate of $\hat\theta_n$ since $\hat\theta_n$ is the true parameter value under F_n

For $b = 1, \dots, B$

- draw X_i^{*b}, $i = 1, \dots, n$ from $(X_1, \dots, X_n)$ with replacement
- form estimate $\hat{\theta}(X^{*b})$

Now if F_n close to real F, X^{*b} would be like samples from true F and so the distribution of $\hat{\theta}(X^{*b}) = \hat{\theta}^*$ across $b = 1, \dots, B$ would be similar to distribution of $\hat{\theta}_n$ over repeated samples from F

$$\Rightarrow \text{Bias}_F(\hat{\theta}_n) \hat{=} \text{Bias}_{F_n}(\hat{\theta}^*) \hat{=} \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*b} - \hat{\theta}_n$$

$$\Rightarrow V_F(\hat{\theta}_n) \hat{=} V_{F_n}(\hat{\theta}^*) \hat{=} \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^{*b} - \hat{\theta}_n)^2$$

$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta(F))$ has distribution $J_n(F)$: ie.

$$P(\sqrt{n}(\hat{\theta}_n - \theta(F)) \leq t) = J_n(t, F)$$

can be approximated by distribution $J_n(F_n)$ of $\sqrt{n}(\hat{\theta}^* - \hat{\theta}_n)$. $J_n(t, F_n) \hat{=} \frac{1}{B} \sum_{b=1}^{B} 1\{\sqrt{n}(\hat{\theta}^* - \hat{\theta}_n) \leq t\}$

$\Rightarrow$ Quantiles of $\sqrt{n}(\hat{\theta}_n - \theta(F))$ under $F \approx$ quantiles of $\sqrt{n}(\hat{\theta}^* - \hat{\theta}_n)$ under F_n

Goals of bootstrap & related methods

- Estimate & reduce <u>bias</u> of $\hat{\theta}(X^n)$
- Estimate <u>variance</u> of $\hat{\theta}(X^n)$
- Set up CI for θ based on <u>sampling distribution</u> of $\hat{\theta}(X^n)$ - where we estimate the - " - " - "

Bias

$\text{Bias}_F(\hat{\theta}) = E_F(\hat{\theta}) - \theta = E_F(\hat{\theta}(X^n)) - \theta$

often, MLE's have the following property (true for other regular estimators as well)

$$E(\hat{\theta}) = \theta + \frac{b_1}{n} + \frac{b_2}{n^2} + \quad ,$$ where b's may depend on θ but not n.

How to estimate bias?

(1) Push Delta-method to 2^{nd} order

Suppose $\hat{\theta} = g(\bar{x})$ is an estimate of $\theta(\mu)$

$$\Rightarrow E(\hat{\theta}) \approx E\left(g(\mu) + g'(\mu)(\bar{x}-\mu) + \frac{g''(\mu)}{2}(\bar{x}-\mu)^2\right)$$

$$= g(\mu) + \underbrace{\frac{g''(\mu)}{2} V(\bar{x})}_{\text{bias - estimate by } \frac{g''(\mu)\,\hat{\sigma}^2/n}{2} = \hat{V}(\bar{x})}$$

and bias-correction $\hat{\theta} - \frac{g''(\mu)}{2}\hat{V}(\bar{x})$

(2) Jackknife method

$X^n = (X_1, \ldots, X_n) \Rightarrow \hat{\theta}_n = \hat{\theta}(X^n)$

$X^{-i} = (X_1 \ldots X_{i-1}, X_{i+1}, \ldots X_n) \rightarrow \hat{\theta}^{-i} = \hat{\theta}(X^{-i})$ leave-one-out estimate

Bias estimate $\hat{b} = (n-1)(\bar{\hat{\theta}} - \hat{\theta}_n)$ where $\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}^{-i}$

If $E(\hat{\theta}_n)$ of form $\theta + \frac{b_1}{n} + \frac{b_2}{n^2} + \ldots$ Then jackknife

reduces variance

Why?

Well $E(\hat{\theta}_n) = \theta + \frac{b_1}{n} + \frac{b_2}{n^2} \ldots$

means $E(\hat{\theta}^{-i}) = \theta + \frac{b_1}{n-1} + \frac{b_2}{(n-1)^2} \ldots$ for all i

and $E(\bar{\hat{\theta}}) = E\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\theta}^{-i}\right) = \frac{1}{n}\sum_{i=1}^{n} \theta + \frac{b_1}{n-1} + \frac{b_2}{(n-1)^2} \ldots$

$\Rightarrow \hat{\theta}_n - \hat{b} = \hat{\theta}_n - (n-1)(\bar{\hat{\theta}} - \hat{\theta}_n) = n\hat{\theta}_n - (n-1)\bar{\hat{\theta}}$

$\Rightarrow E(\hat{\theta}_n - \hat{b}) = n\left(\theta + \frac{b_1}{n} + \frac{b_2}{n^2} \ldots\right) - \frac{(n-1)}{\cancel{n}}\left(\cancel{n}\theta + \frac{nb_1}{n-1} + \frac{\cancel{n}b_2}{(n-1)^2} \ldots\right)$

$= \theta + b_1 + \frac{b_2}{n} + \ldots - (n-1)\theta - b_1 - \frac{b_2}{(n-1)}$

$= \theta + \frac{b_2}{n} - \frac{b_2}{n-1} + \frac{b_3}{n^2} - \frac{b_3}{(n-1)^2} \ldots$

$= \theta - \frac{b_2}{n(n-1)} + O(n^{-3}) = \theta + \underline{O(n^{-2})}$

Compared with $E(\hat{\theta}_n) = \theta + \underline{O(n^{-1})}$ so bias decreases faster to 0 for $\hat{\theta}_n - \hat{b}$ than $\hat{\theta}$.

(3) Bootstrap method

$X_i \sim^{iid} F \Rightarrow x^n \Rightarrow \hat{\theta}_n = \hat{\theta}(x^n)$

$\Downarrow$

F_n estimate of F (e.g. empirical with $\frac{1}{n}$ point mass @ each x_i

$\downarrow$

repeated sampling from F_n gives

bootstrap data sets $x^{*b} = (x_1^{*b}, \dots x_n^{*b})$

$\downarrow$

bootstrap estimates $\hat{\theta}^{*b}$

Now, under F_n true θ is $\hat{\theta}_n$

$\Rightarrow$ bias estimate under F_n is $\hat{b} = \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}^{*b} - \hat{\theta}_n)$

As long as F_n is a good estimate of F

$\hat{b}$ is a good estimate of bias $E_F(\hat{\theta}_n) - \theta(F)$

Possible F_n :

1) empirical F_n w $\frac{1}{n}$ pointmass at x_i
→ drawing with replacement from x^n
is called non-parametric bootstrap

2) $F_n = \hat{F}$ from model family. Example
$\hat{F} = N(\hat{\mu}, \hat{\sigma}^2)$ is called
parametric bootstrap