Bootstrap

can be multivariate also.

Set up: Data $X_1, \ldots, X_n$ — i.i.d. as X with population cdf F (which is not completely known).

Population: $\mathcal{P}$, **Sample:** $\mathcal{S}$

Parameter of interest: $\theta = g(\mathcal{P})$ or $g(F)$, where g is a feature of interest.

Natural estimator: $\hat{\theta}$ = compute the same feature g from sample data.

Examples: Mean, variance, median, quantiles, etc.

Issue: Need to get the *sampling distribution* of $\hat{\theta}$ so that we can compute, e.g., standard error of $\hat{\theta}$, or confidence interval for θ?

Q: Why not use the methods that we have learnt?

If the estimator is like a sample mean, then we know how to find its dist. But for estimators which are more complicated (e.g., sample median) ~~or ... the ...~~ exact methods are either not available or are complicated or don't work very well in practice.

Basics

Bootstrap: A simulation based technique that allows us to approximate the sampling distribution of $\hat{\theta}$. Assumes large n, but its value needed for validity of bootstrap is typically less than that for the usual large-sample procedure.

Original sample: $X_1, \ldots, X_n \sim$ i.i.d. with cdf F

estimated using $X_1, \ldots, X_n$.

Bootstrap (re)sample: $X_1^*, \ldots, X_n^* \sim$ i.i.d. with cdf $\hat{F}$, where $\hat{F}$ = estimated cdf (which is completely known)

	Population	Estimator
Original sample:	F	$\hat{\theta}$
Bootstrap resample:	$\hat{F}$	$\hat{\theta}^*$

Bootstrap philosophy:

- Take $\hat{F}$ as the "population."
- "Replace the unknown F by known $\hat{F}$" — this typically works because the two are close when n is large (additional conditions are needed, but we won't get into those technical details)
- Works by simulating a large number of resamples from $\hat{F}$.

Parametric bootstrap:

- Functional form of F is known (e.g., normal), but F may depend on unknown parameter θ.
- $\hat{F}$ is same as F but with θ replaced by its MLE $\hat{\theta}$. In other words, $\hat{F}$ is the cdf of the fitted model.
- Ex: $F = N(\mu, \sigma^2)$, $\hat{F} =$ $N[\hat{\mu}_{ML}, \hat{\sigma}^2_{ML}]$.
- Often easy to simulate i.i.d. draws $X_1^*, \ldots, X_n^*$ from $\hat{F}$.

Nonparametric bootstrap:

- Functional form of F is unknown.
- $\hat{F}$ = empirical cdf, where

$F(x) = P[X \le x]$

"Population": $(X_1, \ldots, X_n)$

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le x) =$$ prop of $X_1, \ldots, X_n$ that are $\le x$.

- Think of $\hat{F}$ as a discrete distribution that assigns $1/n$ probability to each of the sample observations, $X_1, \ldots, X_n$.
- Get $X_1^*, \ldots, X_n^*$ by sampling n times **with replacement** from $X_1, \ldots, X_n$. In R: "sample"

Read ~~above~~ for the case when n is very small, in which case we can list all the possible resamples.

Bootstrap distribution of estimator $\hat{\theta} = g(\hat{F})$ of $\theta = g(F)$:

Original sample: $X_1, \ldots, X_n$ — gives $\hat{\theta}$

- Simulate a large number b of *bootstrap resamples*, and compute $\hat{\theta}^* = g(\hat{F}^*)$ from each resample.

Resample #	resample	Estimator
1	$X_{11}^*, X_{12}^*, \ldots, X_{1n}^*$	$\hat{\theta}_1^*$
⋮		⋮
b	$X_{b1}^*, X_{b2}^*, \ldots, X_{bn}^*$	$\hat{\theta}_b^*$

$\left.\begin{matrix}\hat{\theta}_1^* \\ \vdots \\ \hat{\theta}_b^*\end{matrix}\right\}$ bootstrapped estimates.

- This process gives a large number of draws, $\hat{\theta}_1^*, \ldots, \hat{\theta}_b^*$.
- These draws are coming from the *bootstrap distribution* of $\hat{\theta}$.
- How to see this distribution?
- It often provides a good approximation to the *sampling distribution* of $\hat{\theta}$.
- Use the draws $\hat{\theta}_1^*, \ldots, \hat{\theta}_b^*$ to estimate features of sampling distribution of $\hat{\theta}$ that may be of interest.

Estimating a feature η of distribution of $\hat{\theta}$:

- Get a large number b of draws, $\hat{\theta}_1^*, \ldots, \hat{\theta}_b^*$.
- $\hat{\eta}^*$ = same feature computed from these draws.

Ex 1: $\eta = E(\hat{\theta})$. $\hat{\eta}^* = \frac{1}{b} \sum_{k=1}^{b} \hat{\theta}_b^*$

Ex 2: $\eta = var(\hat{\theta})$. $\hat{\eta}^* = \frac{1}{b-1} \sum_{k=1}^{b} \left(\hat{\theta}_b^* - \bar{\hat{\theta}}_b^*\right)^2$

Essentially:

{ replace θ by $\hat{\theta}$ and $\hat{\theta}$ by $\hat{\theta}^*$ to do bootstrap }

Ex 3: $\eta =$ bias of $\hat{\theta} = E(\hat{\theta}) - \theta$ $= E[\hat{\theta} - \theta]$. $\hat{\eta}^* = \frac{1}{b} \sum_{k=1}^{b} \hat{\theta}_k^* - \hat{\theta}$

Ex 4: $\eta = \alpha$-th quantile of $\hat{\theta}$. $\hat{\eta}^* = \alpha$-th quantile of draws of $\hat{\theta}^*$

$= \hat{\theta}^*_{((b+1)\alpha)} = (b+1)\alpha$-th ordered value of the ~~sampled~~ bootstrap draws.

Ex 5: $\eta = \alpha$-th quantile of $\hat{\theta} - \theta$. $\hat{\eta}^* = \hat{\theta}^*_{((b+1)\alpha)} - \hat{\theta}$

Bootstrap Confidence Intervals

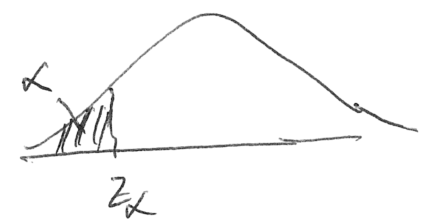
Set up: $\hat{\theta} \approx N(\theta, \hat{V})$ when n is large.

- For example, when $\hat{\theta}$ is MLE and $\hat{V} = \hat{I}^{-1}$.
- Don't need population to be normal.
- works for nonparametric estimators of θ also.

Recall: The standard (approximate) $100(1-\alpha)\%$ CI for θ is:

$$[\hat{\theta} - \underline{z_{1-\alpha/2}}\,\widehat{SE},\ \hat{\theta} - \underline{z_{\alpha/2}}\,\widehat{SE}],$$

pivot.



where $\underline{z_\alpha}$ = $\underline{\alpha\text{-th percentile}}$ of $T = (\hat{\theta} - \theta)/\widehat{SE} \approx N(0,1)$.

Why?

$$
\begin{aligned}
1-\alpha &= P\left[z_{\alpha/2} \le T \le z_{1-\alpha/2}\right] \\
&= P\left[z_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\widehat{SE}} \le z_{1-\alpha/2}\right] \\
&= P\left[-\hat{\theta} + z_{\alpha/2}\widehat{SE} \le -\theta \le -\hat{\theta} + z_{1-\alpha/2}\widehat{SE}\right] \\
&= P\left[\hat{\theta} - z_{1-\alpha/2}\widehat{SE} \le \theta \le \hat{\theta} - z_{\alpha/2}\widehat{SE}\right]
\end{aligned}
$$

Issues: This CI may not be accurate because n may not be large enough for

- normal approximation for T to be good, implying that

  z_α should come from the dist. of T, not normal dist.

  (The distribution of T may not even be symmetric.)
- bias in $\hat{\theta}$ to be negligible, implying that ignoring bias $\hat{\theta}$ is not a good idea.
- $\hat{V}$ to be a good estimate of true V, implying that

  or $\hat{V}$ may not even be available.

  (Often ML-theory based $\hat{V}$ underestimates V.)

Bootstrap CI overcomes these issues to a large extent.

Four Bootstrap CIs for θ

1. Normal approximation CI: Use z critical point but correct $\hat{\theta}$ for bias and use $\hat{V}^*$ to estimate V.

- Estimated bias of $\hat{\theta}$ $(\hat{B}^*)$ $= \frac{1}{b}\sum_{k=1}^{b} \hat{\theta}_k^* - \hat{\theta}$
- CI: $\left[\hat{\theta} - \hat{B}^* - z_{1-\alpha/2}\widehat{SE}^*,\ \hat{\theta} - \hat{B}^* - z_{\alpha/2}\widehat{SE}^*\right]$.

2. Studentized bootstrap CI: Use bootstrap critical point of T instead of z critical point.

$T = \frac{\hat{\theta} - \theta}{\widehat{SE}(\hat{\theta})}$,

- Get $T_1^* = (\hat{\theta}_1^* - \hat{\theta})/\widehat{SE}_1^*, \ldots, T_b^* = (\hat{\theta}_b^* - \hat{\theta})/\widehat{SE}_b^*$
- Estimated α-th percentile of $T =$ $T^*_{((b+1)\alpha)}$
- CI: $\left[\hat{\theta} - t^*_{\left((b+1)(1-\alpha/2)\right)}\widehat{SE},\ \hat{\theta} - t^*_{\left((b+1)(\alpha/2)\right)}\widehat{SE}\right]$.