Bootstrap (ReCap)

- $X_1, X_2, \dots, X_n \sim$ iid F (CDF)
- θ = parameter of interest — feature of F
- $\hat{F}$ = estimated CDF (parametric or nonparametric)
- $\hat{\theta}$ = estimator of θ — same feature of $\hat{F}$
- Bootstrap resample: $X_1^*, X_2^*, \dots, X_n^* \sim$ iid $\hat{F}$
-

Sample	population	Estimator
original	F	$\hat{\theta}$
Bootstrap	$\hat{F}$	$\hat{\theta}^*$

- Generate a large # (b) of resamples:

Resample #	Resample	Estimator (or statistic)
1	$X_{11}^*, X_{12}^*, \dots, X_{1n}^*$	$\hat{\theta}_1^*$
2	$X_{21}^*, X_{22}^*, \dots, X_{2n}^*$	$\hat{\theta}_2^*$
$\vdots$	$\vdots$	$\vdots$
b	$X_{b1}^*, X_{b2}^*, \dots, X_{bn}^*$	$\hat{\theta}_b^*$

$\left.\right\}$ bootstrap dist. of $\hat{\theta}$

- If n is small; take b = # possible resamples.
  In practice: $b = 500$ is good for bias and SE

Need at least $b = 2000$ if estimating quantiles

- To estimate a feature η of sampling dist. of $\hat{\theta}$, use η^* = same feature estimated from the bootstrap draws of $\hat{\theta}^*$
- Essentially: "Replace θ by $\hat{\theta}$, and $\hat{\theta}$ by $\hat{\theta}^*$"
- If $\eta = \alpha$-th quantile of $\hat{\theta} - \theta$; then $\hat{\eta}^* = \alpha$-th sample percentile of $\hat{\theta}_1^* - \hat{\theta}, \dots, \hat{\theta}_b^* - \hat{\theta}$
  $= \hat{\theta}^*_{((b+1)\alpha)} - \hat{\theta}$.

Ex: Nonparametric bootstrap

Resample: Randomly Draw n obs. with replacement from $X_1, \ldots, X_n$.

\# possible resamples: n^n

If $n = 2 \Rightarrow n^n = 2^2 = 4$.

(x_1, x_2)

(x_1^*, x_1^*) $\quad x_2^* = x_2$

(x_1^*, x_2^*) $\quad x_1^* = x_1$

(x_2^*, x_1^*)

(x_2^*, x_2^*)

Bootstrap Confidence Intervals

An approx $100(1-\alpha)\%$ CI for θ: $\hat\theta \pm z_{\alpha/2}\sqrt{\hat V}$

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.

So far: z_α = upper α-th quantile of $N(0,1)$

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

NOW: z_α = α-th quantile of $N(0,1)$

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

Note. Note.

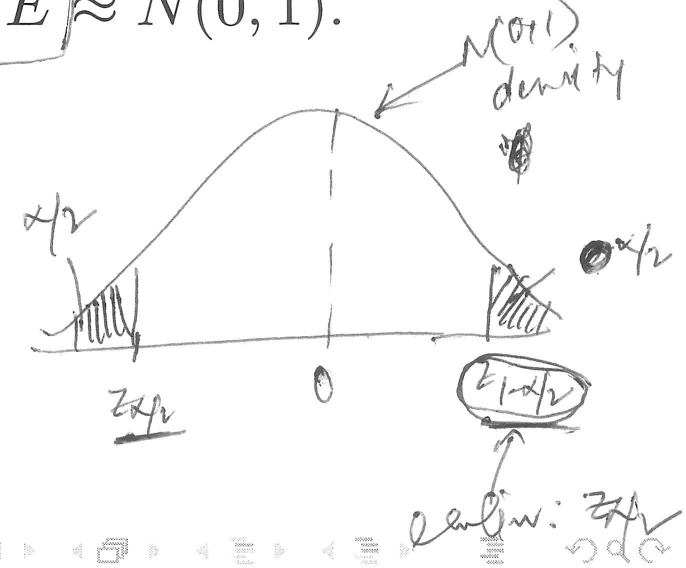
where $z_\alpha = \alpha$-th percentile of $T = (\hat{\theta} - \theta)/\widehat{SE} \approx N(0,1)$.

Why?

$P\left[\hat\theta - z_{1-\alpha/2}\widehat{SE} \le \theta \le \hat\theta - z_{\alpha/2}\widehat{SE}\right]$

$= P\left[-z_{1-\alpha/2} \le \frac{\theta - \hat\theta}{\widehat{SE}} \le -z_{\alpha/2}\right]$

$= P\left[z_{\alpha/2} \le T \le z_{1-\alpha/2}\right] \approx 1-\alpha.$



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

- normal approximation for T to be good, implying that

this CI is not correct

(The distribution of T may not even be symmetric.) NOT

- bias in $\hat{\theta}$ to be negligible, implying that CI may be correct

- $\hat{V}$ to be a good estimate of true V, implying that

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

Bootstrap CI overcomes these issues to a large extent.

Four Bootstrap CIs for θ

bias: $E[\hat{\theta}] - \theta$

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

$V^* =$ bootstrap variance of $\hat{\theta}$

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

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

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

$\frac{\hat{\theta}^* - \hat{\theta}}{\hat{SE}^*}$

- 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 $T =$ $t^*_{((b+1)\alpha)} \rightarrow$ α-th sample percentile of T^*

studentized quantity

- CI: $\left[\hat{\theta} - t^*_{\left((b+1)(1-\alpha/2)\right)}\widehat{SE},\ \hat{\theta} - t^*_{\left((b+1)(\alpha/2)\right)}\widehat{SE}\right].$

3. Basic bootstrap CI: Based on percentiles of $\hat\theta - \theta$ rather than $(\hat\theta-\theta)/\widehat{SE}$. Use bootstrap to estimate them. Notice

$\boxed{a_\alpha = \alpha\text{-th percentile of } \hat\theta - \theta.}$

$$1-\alpha = P(a_{\alpha/2} \le \hat\theta - \theta \le a_{1-\alpha/2})$$

$$= P[\underbrace{\hat\theta - \hat a_{1-\alpha/2}}_{= L} \le \theta \le \underbrace{\hat\theta - a_{\alpha/2}}_{= U}]$$

- Estimated $a_\alpha =$ $\hat\theta^*_{((b+1)\alpha)} - \hat\theta$

lower limit:

$L = \hat\theta - \hat a_{1-\alpha/2}$

$= \hat\theta - \{\hat\theta^*_{((b+1)(1-\alpha/2))} - \hat\theta\}$

$= 2\hat\theta - \hat\theta^*_{((b+1)(1-\alpha/2))}$

- CI: $\left[2\hat\theta - \hat\theta^*_{\left((b+1)(1-\alpha/2)\right)},\ 2\hat\theta - \hat\theta^*_{\left((b+1)(\alpha/2)\right)}\right]$.
- Doesn't require $\widehat{SE}$.

an ↑ fn.

4. **Percentile bootstrap CI:** Works as in basic bootstrap but uses "magic." Suppose there exists a transformation h so that the distribution of $h(\hat{\theta}) - h(\theta)$ is symmetric about zero. Let $U = h(\hat{\theta})$. As before, we can write

$a_\alpha = \alpha$-th percentile of $U - h(\theta)$

density of $U - h(\theta)$

$$1 - \alpha = P\left(a_{\alpha/2} \le U - h(\theta) \le a_{1-\alpha/2}\right)$$

symmetry

$$= P\left(-a_{1-\alpha/2} \le U - h(\theta) \le -a_{\alpha/2}\right)$$

$$= P\left(U + a_{\alpha/2} \le h(\theta) \le U + a_{1-\alpha/2}\right)$$

- Estimated $a_\alpha =$ $U^*_{((b+1)\alpha)} - U$
- $U + a_{\alpha/2} \approx U + \left\{U^*_{((b+1)(\alpha/2))} - U\right\} = U*_{((b+1)(\alpha/2))}$
- Similarly, $U + a_{1-\alpha/2} = U^*_{((b+1)(1-\alpha/2))}$

Therefore,

$$1 - \alpha \approx P\left(U^*_{((b+1)(\alpha/2))} \le h(\theta) \le U^*_{((b+1)(1-\alpha/2))}\right)$$

$$= P\left[h^{-1}\{\ \} \le h^{-1}[h(\theta)] \le h^{-1}\{\ \}\right]$$

$$= P\left[\hat{\theta}^*_{((b+1)\alpha/2)} \le \theta \le \hat{\theta}^*_{((b+1)(1-\alpha/2))}\right]$$

- CI: $\left[\hat{\theta}^*_{\left((b+1)(\alpha/2)\right)},\ \hat{\theta}^*_{\left((b+1)(1-\alpha/2)\right)}\right]$.
- Magic: No need to know h

Q. Which method to use?

Research shows that studentized bootstrap is the best choice, but it requires $\widehat{SE}$. However, if $\widehat{SE}$ is not available, then percentile bootstrap is often the next best choice. More accurate versions of this method are available.

Example: Recall the CPU time data from Example 8.12 on page 217. We had seen that a gamma distribution fit well to these data. Suppose we would like to perform inference on median cpu time.

nonparametric bootstrap.

R code is available in R Handout folder

R code:

```
# use install.packages("boot") to first install
# the package and then load it

library(boot)

# read the cpu data (we have seen these before)

> (cpu <- scan(file="cputime.txt"))
Read 30 items
 [1]  70  36  43  69  82  48  34  62  35  15  59 139
 46  37  42  30  55  56
```

θ = median of the dist.

$\hat{\theta}$ = sample median.

No easy way to compute bias and SE or more generally to have handle on the sampling dist. of $\hat{\theta}$

```
[19]  36  82  38  89  54  25  35  24  22   9  56  19
>

# Parameter of interest: Median

###########################
# Nonparametric Bootstrap #
###########################
```

original data — subject ID's (simply, 1:n)

```
median.npar <- function(x, indices) {
  result <- median(x[indices])
  return(result)
  }
```

resample

original data — b

```
> (median.npar.boot <- boot(cpu, median.npar, R=999,
sim="ordinary", stype="i"))

ORDINARY NONPARAMETRIC BOOTSTRAP
```

```
Call:
boot(data = cpu, statistic = median.npar, R = 999,
sim = "ordinary",  stype = "i")


Bootstrap Statistics :
    original     bias    std. error
t1*     42.5 0.6721722    5.876943
>
```

$\hat{\theta}$ $\hat{B}^*$ $\widehat{SE}^*$

```
# Let's verify the calculations

# See what's else is stored in median.npar.boot

> names(median.npar.boot)
 [1] "t0"         "t"          "R"          "data"
```

$\hat{\theta}$ $\hat{\theta}^*$

```
 "seed"        "statistic"
 [7] "sim"          "call"         "stype"        "strata"
  "weights"
>

> median(cpu)
[1] 42.5
```

$\hat{\theta}$

```
>
> median.npar.boot$t0
```

matches.

```
[1] 42.5
>
> mean(median.npar.boot$t)-median.npar.boot$t0
[1] 0.6721722
>
> sd(median.npar.boot$t)
[1] 5.876943
```

match.

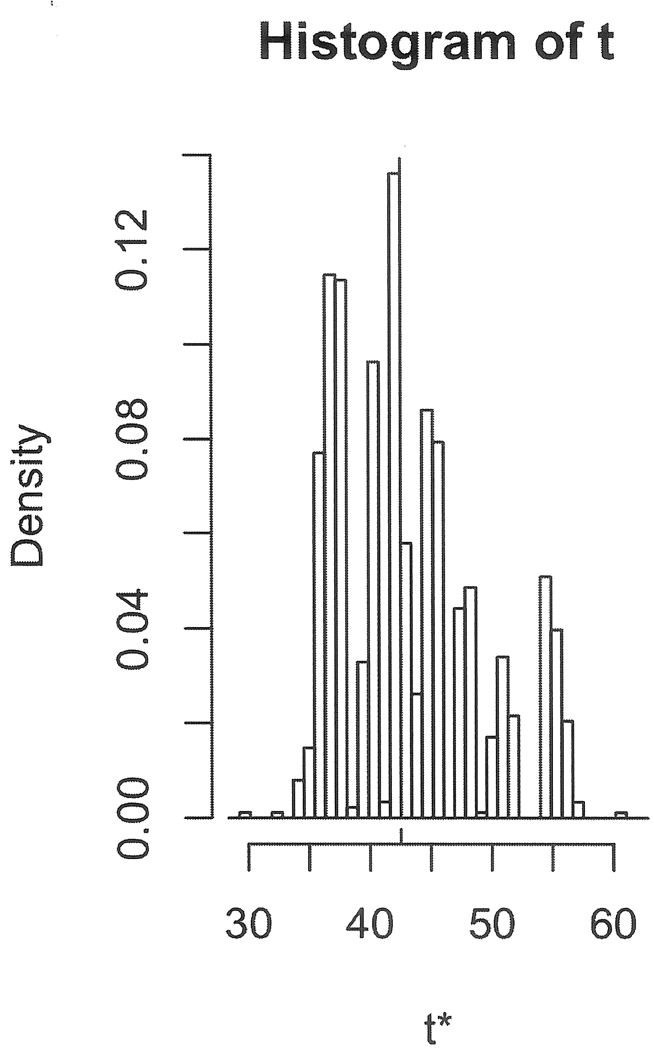
```
Warning message:
sd(<matrix>) is deprecated.
```

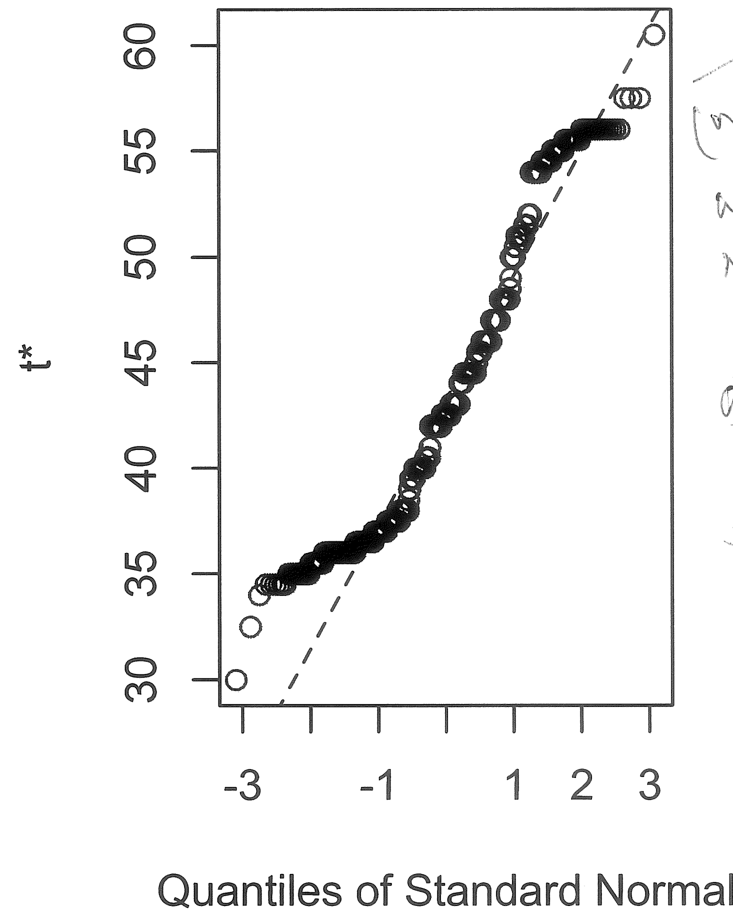
ignore.

```
 Use apply(*, 2, sd) instead.
>

# See the bootstrap distribution of median estimate

plot(median.npar.boot)
```





— cannot (or should not) assume normality for distribution of sample median

```
# Get the 95% confidence interval for median

> boot.ci(median.npar.boot)
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 999 bootstrap replicates

CALL :
boot.ci(boot.out = median.npar.boot)

Intervals :
Level      Normal              Basic
95%   (30.31, 53.35 )   (29.50, 49.50 )

Level     Percentile            BCa
95%   (35.5, 55.5 )   (35.0, 55.5 )
Calculations and Intervals on Original Scale
Warning message:
```

X ignore.

```
In boot.ci(median.npar.boot) :
  bootstrap variances needed for studentized intervals
>
```

ignore the warning.

```
# Let's verify

# Normal approximation method

> c(42.5 - 0.6721722 - qnorm(0.975) * 5.876943,
    42.5 - 0.6721722 - qnorm(0.025) * 5.876943)
[1] 30.30923 53.34642
>
```

matches

```
# Percentile bootstrap method

> sort(median.npar.boot$t)[c(25, 975)]
[1] 35.5 55.5
>
```

matches

```
# Basic bootstrap method

> c(2*42.5-55.5, 2*42.5-35.5)
[1] 29.5 49.5
>
```

matches