Week 10: Bootstrap

MATH-517 Statistical Computation and Visualization

Linda Mhalla

2024-11-22

Bootstrap: Summary

The (non-parametric) Bootstrap

- let $\mathcal{X} = \{X_1, \dots, X_N\}$ be a random sample from F
- quantity of interest: $\theta = \theta(F)$
- (plug-in) estimator: $\hat{\theta} = \theta(\widehat{F}_N)$
  - write $\hat{\theta} = \theta[\mathcal{X}]$, since $\widehat{F}_N$ and thus the estimator depend on the sample
- the distribution $F_{T,N}$ of a scaled estimator $T = g(\hat{\theta}, \theta) = g(\theta[\mathcal{X}], \theta)$ is of interest, e.g., $T = \sqrt{N}(\hat{\theta} - \theta)$

The workflow of the bootstrap is as follows for some $B \in \mathbb{N}$:

Data | Resamples

$$\mathcal{X} = \{X_1, \dots, X_N\} \quad \Longrightarrow \quad \begin{cases} \mathcal{X}_1^\star = \{X_{1,1}^\star, \dots, X_{1,N}^\star\} & \Longrightarrow \quad T_1^\star = g(\theta[\mathcal{X}_1^\star], \theta[\mathcal{X}]) \\ \quad \vdots & \qquad \vdots \\ \mathcal{X}_B^\star = \{X_{B,1}^\star, \dots, X_{B,N}^\star\} & \Longrightarrow \quad T_B^\star = g(\theta[\mathcal{X}_B^\star], \theta[\mathcal{X}]) \end{cases}$$

$F_{T,N}$ now estimated by $\widehat{F}_{T,B}^\star(x) = B^{-1} \sum_{b=1}^{B} \mathbb{I}_{[T_b^\star \leq x]}$

- any characteristic of $F_{T,N}$ can be estimated by the char. of $\widehat{F}_{T,B}^\star(x)$

Confidence Intervals

We want θ_α^U and θ_α^L such that $P\{\theta_\alpha^L \leq \theta \leq \theta_\alpha^U\} = 1 - \alpha$

- $T = \sqrt{N}(\hat{\theta} - \theta) \sim F_{T,N}$ for $\theta \in \mathbb{R}$
- $T_b^\star = \sqrt{N}(\hat{\theta}^{\star b} - \hat{\theta})$ for $b = 1, \dots, B$

$\Rightarrow$ (estimate of) $F_{T,N}$ can be used to construct CI for θ

Asymptotic CI: $q(\alpha)$ is the α-quantile of the asymptotic distribution of T

$$\left(\hat{\theta} - \frac{q(1-\alpha/2)}{\sqrt{N}}, \hat{\theta} - \frac{q(\alpha/2)}{\sqrt{N}}\right)$$

Note: $q(\alpha)$ depends on the asymptotic bias and variance that needs to be estimated (sample/empirical estimates or bootstrap estimates)

E.g., If $\theta = \mathbb{E}(X_1)$, then $q(\alpha)$ is the α-quantile of $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 = \mathrm{Var}(X_1)$ (similar derivation holds for MLEs)

Confidence Intervals

(Basic) Bootstrap CI:

Assuming consistency of the bootstrap, the quantiles of $F_{T,N}$ are estimated by those of the distribution of $T_b^\star$

Let $q_B^\star(\alpha)$ be the empirical α-quantile of $\widehat{F}_{T,B}^\star$, the MC estimate of $F_{T,N}$

$$\left(\hat{\theta} - \frac{q_B^\star(1-\alpha/2)}{\sqrt{N}}, \hat{\theta} - \frac{q_B^\star(\alpha/2)}{\sqrt{N}}\right)$$

We hope that properties of $T_1^\star, \dots, T_B^\star$ mimic effect of sampling from original model $\rightarrow$ false in general, but often more nearly true for a **pivot**

Canonical example: $X_1, \dots, X_N \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. Then,

$$T = \frac{\bar{X} - \mu}{(S^2/N)^{1/2}} \sim t_{N-1}$$

is a pivot as it is independant of the underlying (normal) distribution

Studentized CIs

Exact pivots generally unavailable in non-parametric settings

$\rightarrow$ use studentized statistic

$$T = \frac{\widehat{\theta} - \theta}{V^{1/2}}$$

where $V = \mathrm{Var}(\widehat{\theta})$ is replaced by a consistent estimate

If quantiles $q(\alpha)$ of T are known, then

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

$\Rightarrow$ use bootstrap to estimate the distribution of T

Studentized CIs

- bootstrap sample gives $(\widehat{\theta}^{\star b}, V_b^\star)$ and hence

$$T_b^\star = \frac{\widehat{\theta}^{\star b} - \widehat{\theta}}{V_b^\star}$$

- B bootstrap samples give $T_1^\star, \dots, T_B^\star$

$\Rightarrow$ use $T_1^\star, \dots, T_B^\star$ to estimate distribution of T and denote $q_B^\star(\alpha)$ the estimated α-quantile

- get $1 - \alpha$ confidence interval

$$\widehat{\theta} - V^{1/2} q_B^\star(1 - \alpha/2), \quad \widehat{\theta} - V^{1/2} q_B^\star(\alpha/2)$$

Note Use of studentized statistic reduces error from $\mathcal{O}(N^{-1/2})$ to $\mathcal{O}(N^{-1})$: this is what we showed last week for one-sided CI

$\Rightarrow$ studentization **recommended**, but requires consistent estimation of V

Another Confidence Interval

Percentile CI:

Use empirical quantiles (order statistics) of $\hat{\theta}^{\star 1}, \dots, \hat{\theta}^{\star b}$ to construct CI

$$\hat{\theta}^{\star}_{((B+1)\alpha/2)}, \quad \hat{\theta}^{\star}_{((B+1)(1-\alpha/2))}$$

$\Rightarrow$ tends to be too narrow for small N and coverage is exact up to $\mathcal{O}(N^{-1})$ (same for asymptotic and basic CIs)

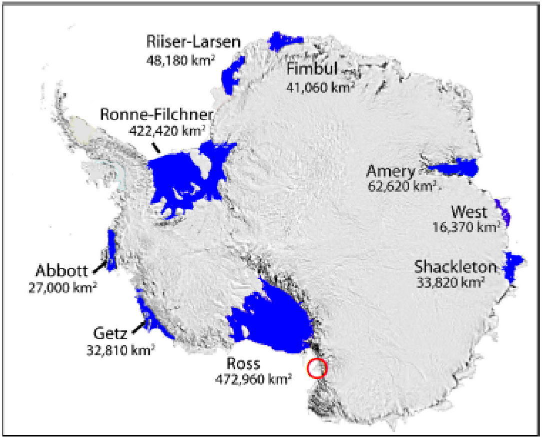
Back to Basic (bootstrap):

$$\hat{\theta} - \{\hat{\theta}^{\star}_{((B+1)(1-\alpha/2))} - \hat{\theta}\}, \quad \hat{\theta} - \{\hat{\theta}^{\star}_{((B+1)\alpha/2)} - \hat{\theta}\}$$

General Comparison

- Asymptotic, basic, and studentized intervals depend on scale
- Percentile interval is transformation-invariant and thus does a better job under skewness (than basic or asymptotic CI). Often too short though
- Studentized interval gives best coverage overall but can be sensitive to V. They are best on transformed scale, where V is approximately constant

Example: Antarctic ice shelves data



Aim: Determine a 90% CI of the median from the log-transformed areas of the 17 ice shelves

Example: Median

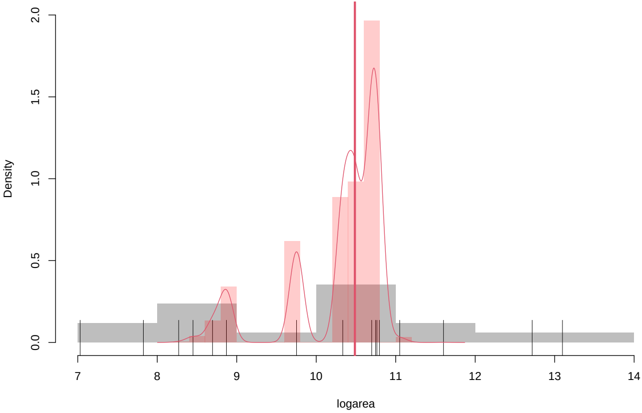
- Antarctic ice shelves data
- interested in the median of the log-area of the ice shelves

```
aa <- read.csv('../data/AAshelves.csv')
  # source: Reinhard Furrer's "Statistical Modeling" lecture at UZH
logarea <- log(aa[[3]]) # log of ice shelf areas
set.seed(517)
N <- length(logarea) #17
B <- 5000
boot_data <- array(sample(logarea, N*B, replace=TRUE), c(B, N))
meds <- apply(boot_data, 1, median)
hist(logarea, col='gray', main='', border=NA)
rug(logarea, ticksize = .04)
abline(v=median(logarea),lwd=2)
```

Example: Median

- Antarctic ice shelves data
- interested in the median of the log-area of the ice shelves

$\rightarrow$ distribution is multimodal due to the small sample size



Example: Median

Is the sample median asymptotically normal?

$$\theta = F^{-1}(1/2) \qquad \& \qquad \hat{\theta} = \widehat{F}_N^{-1}(1/2)$$

$$T = \sqrt{N}(\hat{\theta} - \theta) \overset{?}{\to} \mathcal{N}(0, v)$$

Example: Median

Is the sample median asymptotically normal?

$$\theta = F^{-1}(1/2) \qquad \& \qquad \hat{\theta} = \widehat{F}_N^{-1}(1/2)$$

$T = \sqrt{N}(\hat{\theta} - \theta) \overset{?}{\to} \mathcal{N}(0, v)$

- yes, under some conditions
  - verifying conditions of a general theorem for M-estimator yields assumption:
  - $f(\theta) \neq 0$ and f continuous on some neighborhood of θ

Say we wish to construct a 90% confidence interval for θ

Option I:

- approximate only v using bootstrap and use asymptotic CI

Option II:

- approximate the quantiles of T or $\hat{\theta}$ using bootstrap: basic or percentile CIs

Example: Median

$T^\star = \sqrt{N}(\hat\theta^\star - \hat\theta)$ or just $T^\star = \hat\theta^\star$

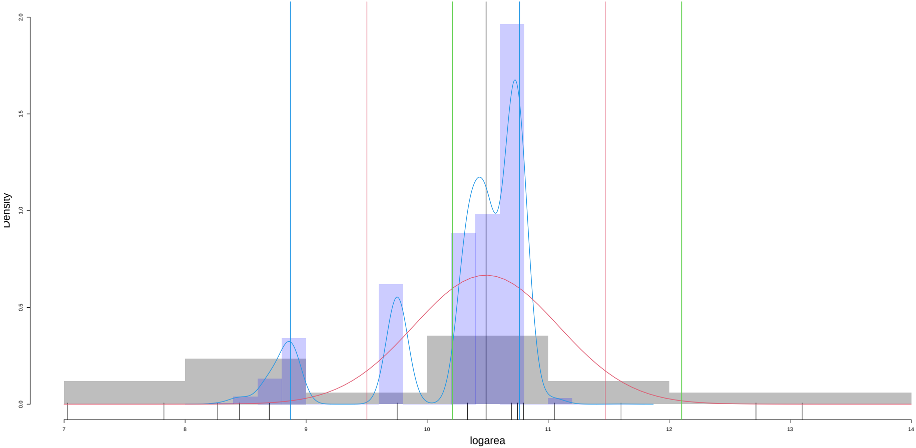
Option I: approximate $\mathrm{aVar}(T^\star)$ using bootstrap

Option II: approximate the quantiles of $T^\star$ using bootstrap

- KDE on the MC draws of $\hat\theta^\star$ can be used to visualize the distribution

```
hist(logarea, prob=TRUE, col='gray', ylim=c(0,2.), main='', border=NA)
rug(logarea, ticksize = .04)
abline(v=median(logarea),lwd=2)
hist(meds, add=T, prob=T, col=rgb(0,0,1,.2), border=NA)
lines(density(meds, adjust=2), col=4, lwd=2)
curve(dnorm(x, median(logarea), sd(meds)), add=T, col=2,lwd=2)
abline(v=c(median(logarea)-qnorm(c(.95,.05), sd = sd(meds))), col=2, lwd=2)
# asymptotic: 9.502293 11.470844
# sd(meds) == sd(sqrt(N)*(meds-median(logarea)))/sqrt(N)
abline(v=c(quantile(meds, c(.05,.95))), col=4, lwd=2)
# percentile: 8.870101 10.763525
abline(v=2*median(logarea)-quantile(meds, c(.95,.05)), col=3, lwd=2)
# basic: 10.20961 12.10304
```

Example: Median



What if we wanted a studentized interval?

$$\widehat{\theta} - V^{1/2} q_B^{\star}(1-\alpha/2), \quad \widehat{\theta} - V^{1/2} q_B^{\star}(\alpha/2)$$

Variance Estimation

- often $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}_p(0, \Sigma)$, but $V = N^{-1}\Sigma$ needs to be estimated

The bootstrap estimator of $N^{-1}\Sigma$ is easy to obtain:

$$\widehat{V}^{\star} = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}^{\star b} - \bar{\theta}^{\star}\right) \left(\hat{\theta}^{\star b} - \bar{\theta}^{\star}\right)^{\top}, \qquad \text{where} \qquad \bar{\theta}^{\star} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{\star b}$$

N^{-1} because one should take $T^{\star} = \sqrt{N}(\hat{\theta}_b^{\star} - \hat{\theta})$, and estimate Σ by

$$\frac{1}{B-1} \sum_{b=1}^{B} \left(T_b^{\star} - \bar{T}^{\star}\right) \left(T_b^{\star} - \bar{T}^{\star}\right)^{\top} \approx N^{-1} \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}^{\star b} - \bar{\theta}^{\star}\right) \left(\hat{\theta}^{\star b} - \bar{\theta}^{\star}\right)^{\top}$$

Often, this is an inner step when computing CIs...

Variance Estimation

Estimation of variance $V = \mathrm{Var}(\widehat{\theta})$ is required for certain types of CIs

E.g., studentized CI are based on the quantiles of $T_b^\star = \frac{\widehat{\theta}^{\star b} - \widehat{\theta}}{V_b^\star}$

$\Rightarrow V_b^\star$ needed

There are several ways to compute this

- iterated (double) bootstrap
- delta method
- jacknife

Iterated Bootstrap

Simple bootstrap: B resamples

Data | Resamples

$$\mathcal{X} = \{X_1, \ldots, X_N\} \implies \begin{cases} \mathcal{X}_1^\star = \{X_{1,1}^\star, \ldots, X_{1,N}^\star\} \implies T_1^\star = g(\theta[\mathcal{X}_1^\star], \theta[\mathcal{X}]) \\ \vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ \mathcal{X}_B^\star = \{X_{B,1}^\star, \ldots, X_{B,N}^\star\} \implies T_B^\star = g(\theta[\mathcal{X}_B^\star], \theta[\mathcal{X}] \end{cases}$$

Iterated Bootstrap

Double bootstrap: $B(C+1)$ resamples

$$
\begin{array}{ccccc}
\text{Data} & & \text{Resamples} & & \text{Re-resamples} \\
\mathcal{X} & \Longrightarrow & \left\{ \begin{array}{l}
\mathcal{X}_1^\star \Longrightarrow \left\{ \begin{array}{c} \mathcal{X}_{1,1}^{\star\star} \\ \vdots \\ \mathcal{X}_{1,C}^{\star\star} \end{array} \right. \\
\vdots \qquad\qquad\qquad \vdots \\
\mathcal{X}_B^\star \Longrightarrow \left\{ \begin{array}{c} \mathcal{X}_{B,1}^{\star\star} \\ \vdots \\ \mathcal{X}_{B,C}^{\star\star} \end{array} \right.
\end{array} \right. & &
\end{array}
$$

E.g., $V_b^\star$ is the sample variance of $\theta_{b,1}^{\star\star}, \ldots, \theta_{b,C}^{\star\star}$

Iterated bootstrap

Why?

- for bias reduction: a j-th iterated bootstrap reduces order of bias from $\mathcal{O}(N^{-1})$ to $\mathcal{O}(N^{-(j+1)})$
- for CIs: each iteration reduces the coverage error by factor $N^{-1/2}$ (one-sided: recall errors of asymptotic and studentized CIs seen last week) or N^{-1} (two-sided)

Choice of C:

- The total cost of implementation is proportional to BC
- Rule of thumb: C should be of the same order as $\sqrt{B}$: a high degree of accuracy in the second stage is less important than for the first stage
- Often reasonable to take $C = 50$ for variance estimation

Example: Median (continued)

Goal: construct CI for the median

Option I: approximate only the asymptotic variance v using bootstrap

- asymptotic

Option II: approximate directly the quantiles of $T^\star$ using bootstrap

- non-studentized CI

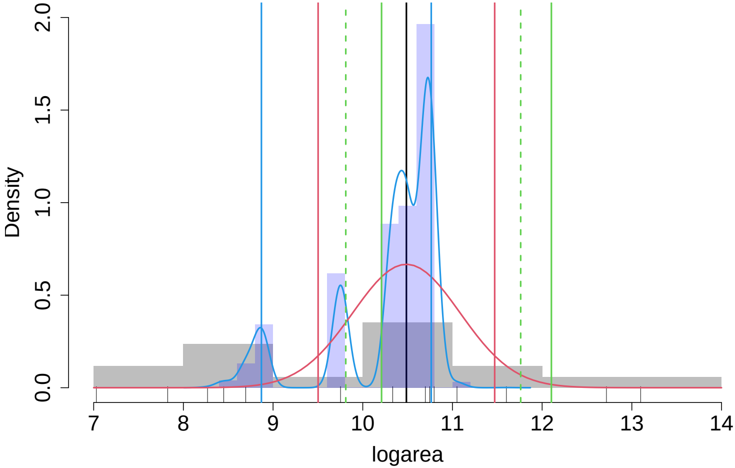
Option III: approximate the quantiles of a studentized statistic using one bootstrap (requires the knowledge of variance, so get that by using another bootstrap)

- studentized CI

Example: Median (continued)

```
set.seed(517)
N <- 17; B <- 5000; C <- 500;
boot_data <- array(sample(logarea, N*B, replace=TRUE), c(B, N))
# Dboot_data <- array(0,c(B,C,N))
# for(b in 1:B){
#   Dboot_data[b,,] <- array(sample(boot_data[b,], N*C, replace=TRUE), c(C, N))
# }
# meds <- apply(boot_data, 1, median)
# Dmeds <- apply(Dboot_data, c(1,2), median)
# sds <- apply(Dmeds, 1, sd)
# T_stars <- sqrt(N)*(meds - median(logarea))/sds #studentized statistic
op <- par(ps=20)
hist(logarea, prob=TRUE, col='gray', ylim=c(0,2.), main='', border=NA)
rug(logarea, ticksize = .04); abline(v=median(logarea),lwd=2)
hist(meds, add=T, prob=T, col=rgb(0,0,1,.2), border=NA)
lines(density(meds, adjust=2), col=4, lwd=2)
curve(dnorm(x, median(logarea), sd(meds)), add=T, col=2,lwd=2)
abline(v=median(logarea)-qnorm(c(.95,.05))*sd(meds), col=2, lwd=2)
### sd(meds) == sd(sqrt(N)*(meds-median(logarea)))/sqrt(N)
abline(v=quantile(meds, c(.05,.95)), col=4, lwd=2)
abline(v=2*median(logarea)-quantile(meds,c(.95,.05)),col=3,lwd=2)
# abline(v=median(logarea)-quantile(T_stars, c(.95,.05))/sqrt(N)*sd(meds), col=3, l
abline(v=c(9.810801, 11.760838), col=3, lwd=2, lty=2) # studentized CI
```

Example: Median (continued)



Is the studentized CI actually better? Simulations!

Delta Method

Computation of variance formulae for functions of averages and other estimators

Suppose $\widehat{\psi} = g(\hat{\theta})$ estimates $\psi = g(\theta)$, and $\hat{\theta} \dot{\sim} \mathcal{N}\left(\theta, \sigma^2/N\right)$

Then under mild conditions and provided $g'(\theta) \neq 0$, Taylor expansion gives

$$\mathrm{E}(\widehat{\psi}) = g(\theta) + O\left(N^{-1}\right)$$
$$\mathrm{Var}(\widehat{\psi}) = \sigma^2 g'(\theta)^2/N + O\left(N^{-3/2}\right)$$

$\Rightarrow \mathrm{Var}(\widehat{\psi}) \doteq \hat{\sigma}^2 g'(\hat{\theta})^2/N = V$

Example: $\hat{\theta} = \bar{X}_N \dot{\sim} \mathcal{N}(\mu, \sigma^2/N)$ and, $\widehat{\psi} = \log\hat{\theta} \dot{\sim} \mathcal{N}\big(\log(\mu), \frac{\sigma^2}{N}\frac{1}{\mu^2}\big)$

Delta Method for Variance Stabilisation

If $\mathrm{Var}(\hat{\theta}) \doteq S(\theta)^2/N$ (depends on θ), find transformation g such that $\mathrm{Var}\{g(\hat{\theta})\} \doteq$ constant

E.g., **Poisson distribution:**

- Let $X_1, \ldots, X_N \overset{i.i.d.}{\sim} \mathrm{Poisson}(\lambda)$
- $\mathbb{E}\left(X_i\right) = \mathrm{Var}\left(X_i\right) = \lambda$
- CLT says $\sqrt{N}\left(\bar{X}_N - \lambda\right) \xrightarrow{d} T \sim \mathcal{N}(0, \lambda)$
- Delta method says

$$\sqrt{N}\{g\left(\bar{X}_N\right) - g(\lambda)\} \xrightarrow{d} g'(\lambda)T = Y \sim \mathcal{N}\left(0, g'(\lambda)^2\lambda\right)$$

- If $g'(\lambda) = \frac{1}{\sqrt{\lambda}}$, then $Y \sim N(0, 1)$

Jackknife

- a predecessor to the bootstrap
  - sometimes can achieve a better trade-off between accuracy and computational costs, but hard to quantify
- used first for bias correction (Quenouille, 1949), later for variance estimation (Tukey, 1958)

Consider $X_1, \dots, X_N$ a random sample from F depending on $\theta \in \mathbb{R}^p$

- $\hat{\theta} = \theta[X_1, \dots, X_N]$
  - interested in some characteristic of the estimator such as the bias

The jacknife method creates resamples of the original sample by leaving out one observation each time and computing

$$\hat{\theta}_{-n} = \theta[X_1, \dots, X_{n-1}, X_{n+1}, \dots, X_N]$$

- consider $\bar{\theta} = N^{-1} \sum_n \hat{\theta}_{-n}$

Jackknife estimator of the bias: $\hat{b} = (N-1)(\bar{\theta} - \hat{\theta})$

Jackknife Bias - a Heuristic

- assume $b = \text{bias}(\widehat{\theta}) = a_1 N^{-1} + a_2 N^{-2} + \mathcal{O}(N^{-3})$ for some constants a_1 and a_2

$$\text{bias}(\widehat{\theta}_{-n}) = a_1 (N-1)^{-1} + a_2 (N-1)^{-2} + \mathcal{O}(N^{-3}) = \text{bias}(\bar{\theta})$$

$$\begin{aligned}
\mathbb{E}\widehat{b} &= (N-1)\big\{\text{bias}(\bar{\theta}) - \text{bias}(\widehat{\theta})\big\} \\
&= (N-1)\left\{a_1\left(\frac{1}{N-1} - \frac{1}{N}\right) + a_2\left(\frac{1}{(N-1)^2} - \frac{1}{N^2}\right) + \mathcal{O}\left(\frac{1}{N^3}\right)\right\} \\
&= a_1 N^{-1} + a_2 N^{-2}\frac{2N-1}{N-1} + \mathcal{O}(N^{-2}) + \mathcal{O}(N^{-3}) \\
&= b + a_2 N^{-2}\frac{N}{N-1} + \mathcal{O}(N^{-2}) = b + \mathcal{O}(N^{-2})
\end{aligned}$$

$\Rightarrow \widehat{b}$ approximates b correctly up to the order N^{-2}, which corresponds to the bootstrap

$\Rightarrow \widehat{\theta}_b^{\star} = \widehat{\theta} - \widehat{b} = N\widehat{\theta} - (N-1)\bar{\theta}$ has bias of order N^{-2}

Jackknife Variance

John W. Tukey defined the "pseudo-values"

$$\theta_n^\star = N\widehat{\theta} - (N-1)\widehat{\theta}_{-n}$$

and conjectured that in some situations these can be treated as i.i.d. with mean θ and variance $N\,\mathrm{Var}(\widehat{\theta})$, and hence we can take

$$\widehat{\mathrm{Var}}(\widehat{\theta}) = \frac{1}{N}\frac{1}{N-1}\sum_{n=1}^{N}\left(\theta_n^\star - \bar{\theta}^\star\right)\left(\theta_n^\star - \bar{\theta}^\star\right)^\top$$

- later shown to actually work ($\approx$ bootstrap via delta method)
- could be used instead of the second bootstrap in our double bootstrap example above
- requires $N+1$ calculations of $\widehat{\theta}$: cheaper than bootstrap
- "works" for smooth statistics (mean, variance, moments) but not for rough statistics (median, maxima, etc)

Hypothesis Testing

- data $X_1, \ldots, X_N$
- hypothesis H_0 to be tested using a test statistic T
- depending on the form of the alternative H_1, evidence against H_0 is
  - large values of T,
  - small values of T, or
  - large values of $|T|$

Assume that large values of T give evidence against H_0

- $t_{obs} = t(X_1, \ldots, X_N)$ the observed value of T
- the p-value

$$p_{\text{obs}} = \Pr_{H_0}(T \geq t_{obs}) = \Pr\left(T \geq t_{\text{obs}} \mid M_0\right)$$

  measures evidence against H_0, i.e., small p_{obs} indicates evidence against the null

$\Rightarrow$ often hard to calculate as it depends on distribution of T under H_0

Hypothesis Testing

- Estimate p_{obs} by simulation from fitted null hypothesis model $\widehat{M}_0$
- **Algorithm:** for $b = 1, \ldots, B$:
  - simulate data set $X_1^\star, \ldots, X_N^\star$ from $\widehat{M}_0$
  - calculate test statistic $t_b^* = t(X_1^\star, \ldots, X_N^\star)$.
- Calculate bootstrap estimate

$$\hat{p} = \frac{\#\left\{t_b^\star \geq t_{\text{obs}}\right\}}{B}$$

of

$$\hat{p}_{\text{obs}} = \Pr\left(T \geq t_{\text{obs}} \mid \widehat{M}_0\right)$$

- Simulation and statistical errors:

$$\hat{p} \approx \hat{p}_{\text{obs}} \approx p_{\text{obs}} = \Pr\left(T \geq t_{\text{obs}} \mid M_0\right)$$

Example

$X_1, \ldots, X_N \overset{i.i.d.}{\sim} Exp(1/2)$ and $H_0 : \mu = 1.78$ vs $H_1 : \mu > 1.78$

```r
set.seed(517)
N      <- 100; B <- 10000
X      <- rexp(N,1/2)
mu_0   <- 1.78   # hypothesized value
T_stat <- (mean(X)-mu_0)/sd(X)*sqrt(N) #asympt. normal under H0
boot_stat <- rep(0,B)
for(b in 1:B){
  Xb           <- sample(X,N,replace=T)
  boot_stat[b] <- (mean(Xb)-mean(X))/sd(Xb)*sqrt(N)
  # Xb           <- rexp(N, rate=1/mu_0)
  # boot_stat[b] <- (mean(Xb)-mu_0)/sd(Xb)*sqrt(N)
}
p_boot    <- mean(boot_stat >= T_stat)
p_obs_hat <- 1-pnorm(T_stat)
c(p_obs_hat, p_boot)
```

```
[1] 0.05919482 0.03760000
```

Example

$H_0 : \mu = 1.78$ vs $H_1 : \mu \neq 1.78$

```
set.seed(517)
N      <- 100;B <- 10000
X      <- rexp(N,1/2)
mu_0   <- 1.78   # hypothesized value # reduce to increase power
T_stat <- (mean(X)-mu_0)/sd(X)*sqrt(N) #asympt. normal under H0
boot_stat <- rep(0,B)
for(b in 1:B){
  # Xb           <- sample(X,N,replace=T)
  # boot_stat[b] <- (mean(Xb)-mean(X))/sd(Xb)*sqrt(N)
  Xb           <- rexp(N, rate=1/mu_0)
  boot_stat[b] <- (mean(Xb)-mu_0)/sd(Xb)*sqrt(N)
}
p_boot    <- mean(abs(boot_stat) >= T_stat)
p_obs_hat <- 2*(1-pnorm(T_stat))

c(p_obs_hat, p_boot)
```

```
[1] 0.1183896 0.1323000
```

Example with Iterated Bootstrap

- $X_1, \dots, X_N \in \mathbb{R}^p$ i.i.d. from a distribution depending on $\theta \in \mathbb{R}^p$
- $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$
- assume $\hat{\theta}$ satisfies $\sqrt{N}(\hat{\theta} - \theta) \stackrel{d}{\to} \mathcal{N}(0, \Sigma)$
- studentized statistic:

$$T = \sqrt{N}\widehat{\Sigma}^{-1/2}(\hat{\theta} - \theta_0) \stackrel{d}{\to} \mathcal{N}(0, I_{p\times p}) \qquad \text{(under } H_0\text{)}$$

  - $\widehat{\Sigma}$ is consistent for Σ
- asymptotic test based on: $\|T\|^2 \stackrel{d}{\to} \chi_p^2$ under H_0

Bootstrap can be used

- instead of using the asymptotic distribution to produce a p-value, or
- when an estimator of Σ is not available

Both of the above combined $\Rightarrow$ double bootstrap

Example with Iterated Bootstrap

$$
\mathcal{X} = \{X_1, \dots, X_N\} \left\{ \begin{array}{lll}
\mathcal{X}_1^\star = \{X_{1,1}^\star, \dots, X_{1,N}^\star\} & \left\{ \begin{array}{c} \mathcal{X}_{1,1}^{\star\star} = \{X_{1,1,1}^{\star\star}, \dots, X_{1,1,N}^{\star\star}\} \\ \vdots \\ \mathcal{X}_{1,M}^{\star\star} = \{X_{1,M,1}^{\star\star}, \dots, X_{1,M,N}^{\star\star}\} \end{array} \right\} & \widehat{\Sigma}_1^{\star\star} \implies T_1^\star \\
 & \vdots & \\
\vdots & \vdots & \\
 & \vdots & \\
\mathcal{X}_B^\star = \{X_{B,1}^\star, \dots, X_{B,N}^\star\} & \left\{ \begin{array}{c} \mathcal{X}_{B,1}^{\star\star} = \{X_{B,1,1}^{\star\star}, \dots, X_{B,1,N}^{\star\star}\} \\ \vdots \\ \mathcal{X}_{B,M}^{\star\star} = \{X_{B,M,1}^{\star\star}, \dots, X_{B,M,N}^{\star\star}\} \end{array} \right\} & \widehat{\Sigma}_B^{\star\star} \implies T_B^\star
\end{array} \right\} \widehat{p}
$$

where

$$
\widehat{\Sigma}_b^{\star\star} = \frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{\theta}_{b,m}^{\star\star} - \bar{\theta}_b^{\star\star}\right)\left(\hat{\theta}_{b,m}^{\star\star} - \bar{\theta}_b^{\star\star}\right)^\top, \quad \text{where} \quad \hat{\theta}_m^{\star\star} = \theta\big[\mathcal{X}_{b,m}^{\star\star}\big] \quad \& \quad \bar{\theta}_b^{\star\star} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_{b,m}^{\star\star},
$$

$$
T_b^\star = \sqrt{N}\left(\widehat{\Sigma}_b^{\star\star}\right)^{-1/2}\left(\hat{\theta}_b^\star - \hat{\theta}\right),
$$

$$
\widehat{p} = \frac{1}{B}\left(\sum_{b=1}^{B} I\big(\|T_b^\star\|^2 \geq \|T\|^2\big)\right),
$$

Parametric Bootstrap and GoF Testing

- $X_1, \dots, X_N \overset{\perp\!\!\!\perp}{\sim} F$
- **goal**: test $H_0 : F \in \mathcal{F} = \{F_\lambda \mid \lambda \in \Lambda\}$ against $H_1 : F \notin \mathcal{F}$
  - if $\mathcal{F} = \{F_0\}$, we could use the KS statistic: $\sup_x \left|\widehat{F}_N(x) - F_0(x)\right|$
- plug in principle: use $t_{obs} = \sup_x \left|\widehat{F}_N(x) - F_{\widehat{\lambda}}(x)\right|$
  - where $\widehat{\lambda}$ is consistent under H_0 (e.g. the MLE)

Bootstrap procedure: **for** $b = 1, \dots, B$

- generate $\mathcal{X}_b^\star = \{X_{b,1}^\star, \dots, X_{b,N}^\star\}$
  - this time not by resampling, but by sampling from $F_{\widehat{\lambda}}$
- estimate $\widehat{\lambda}^{\star b}$ from $\mathcal{X}_b^\star$
- calculate the EDF $\widehat{F}_{N,b}^\star$ from $\mathcal{X}_b^\star$
- set $t_b^\star = \sup_x \left|\widehat{F}_{N,b}^\star(x) - F_{\widehat{\lambda}^{\star b}}(x)\right|$
- estimate the p-value by $\hat{p} = \#\left\{t_b^\star \geq t_{\text{obs}}\right\}/B$

Assignment 7 [5 %]

Go to Assignment 7 for details