Lecture 26: Bootstrap III

Pratheepa Jeganathan

06/03/2019

Recall

- One sample sign test, Wilcoxon signed rank test, large-sample approximation, median, Hodges-Lehman estimator, distribution-free confidence interval.
- Jackknife for bias and standard error of an estimator.
- Bootstrap samples, bootstrap replicates.
- Bootstrap standard error of an estimator.
- Bootstrap percentile confidence interval.
- Hypothesis testing with the bootstrap (one-sample problem.)
- Assessing the error in bootstrap estimates.
- Example: inference on ratio of heart attack rates in the aspirin-intake group to the placebo group.
- The exhaustive bootstrap distribution.

- Discrete data problems (one-sample, two-sample proportion tests, test of homogeneity, test of independence).
- Two-sample problems (location problem - equal variance, unequal variance, exact test or Monte Carlo, large-sample approximation, H-L estimator, dispersion problem, general distribution).
- Permutation tests (permutation test for continuous data, different test statistic, accuracy of permutation tests).
- Permutation tests (discrete data problems, exchangeability.)
- Rank-based correlation analysis (Kendall and Spearman correlation coefficients.)
- Rank-based regression (straight line, multiple linear regression, statistical inference about the unknown parameters, nonparametric procedures - does not depend on the distribution of error term.)
- Smoothing (density estimation, bias-variance trade-off, curse of dimensionality)
- Nonparametric regression (Local averaging, local regression, kernel smoothing, local polynomial, penalized regression)

- Cross-validation, Variance Estimation, Confidence Bands, Bootstrap Confidence Bands.
- Wavelets (wavelet representation of a function, coefficient estimation using Discrete wavelet transformation, thresholding - VishuShrink and SureShrink).
- One-way layout (general alternative (KW test), ordered alternatives), multiple comparison procedure.
- Two-way layout (complete block design (Friedman test)), multiple comparison procedure, median polish, Tukey additivity plot, profile plots.

Better bootstrap confidence intervals

Overview

- We learned
  - Plug-in principal
  - Computing standard error of an estimate
  - Confidence intervals based on bootstrap percentiles
    - Coverage performance (need to do)
  - Hypothesis testing using bootstrap
    - Using bootstrap percentile confidence interval.
    - Using p-value.
  - Exhaustive bootstrap.
- What to cover
  - bootstrap-t interval
  - BCa interval and ABC interval.

Problem

- Inference on one parameter.
- Let $\boldsymbol{x} = \{x_1, \cdots, x_n\} \sim F_\theta(\cdot)$, where θ is an unknown parameter.
- Construct $1 - 2\alpha$ confidence interval for θ.

Confidence interval

- Suppose
$$\hat{\theta} \sim \mathrm{N}\left(\theta, \mathrm{se}^2\right).$$
- Then,
$$Z = \frac{\hat{\theta} - \theta}{\mathrm{se}} \sim \mathrm{N}\left(0, 1\right).$$
- Thus,
$$\mathrm{Prob}_{\theta}\left\{\theta \in \left[\hat{\theta} - z^{(1-\alpha)}\mathrm{se},\quad \hat{\theta} - z^{(\alpha)}\mathrm{se}\right]\right\} = 1 - 2\alpha,$$
where $\mathrm{Prob}_{\theta}\{\}$ is the probability calculated with the true mean equaling θ, so $\hat{\theta} \sim \mathrm{N}\left(\theta, \mathrm{se}^2\right)$.

Coverage of confidence interval

- In the above case,
$$\hat{\theta} \sim \mathrm{N}\left(\theta, \mathrm{se}^2\right),$$
the interval
$$\left[\hat{\theta} - z^{(1-\alpha)}\mathrm{se}, \quad \hat{\theta} - z^{(\alpha)}\mathrm{se}\right]$$
has probability exactly $1 - 2\alpha$ of containing the true value of θ.
    - θ is a constant.
    - Let $\hat{\theta}_{\mathrm{lo}} = \hat{\theta} - z^{(1-\alpha)}\mathrm{se}$ and $\hat{\theta}_{\mathrm{up}} = \hat{\theta} - z^{(\alpha)}\mathrm{se}$. Then, $\hat{\theta}_{\mathrm{lo}}$ and $\hat{\theta}_{\mathrm{up}}$ are random variables.
- equal-tailed CI: If $\mathrm{Prob}_{\theta}\{\theta < \hat{\theta}_{\mathrm{lo}}\} = \alpha$ and $\mathrm{Prob}_{\theta}\{\theta > \hat{\theta}_{\mathrm{up}}\} = \alpha$, then $\left(\hat{\theta}_{\mathrm{lo}}, \hat{\theta}_{\mathrm{up}}\right)$ is an equal-tailed.

Relationship between confidence intervals and hypothesis tests

- $1 - 2\alpha$ confidence interval $\left(\hat{\theta}_{\mathsf{lo}}, \hat{\theta}_{\mathsf{up}}\right)$ is the set of plausible values of θ having observed $\hat{\theta}$.
- Check whether the null value is in the interval.
  - If the null value of θ is not in the interval, reject the null hypothesis.

Standard confidence interval (procedure)

- In most cases,

$$\frac{\hat{\theta}-\theta}{\hat{\mathsf{se}}} \dot{\sim} \mathrm{N}(0,1)$$

.

  - $1-2\alpha$ standard confidence interval for θ is

$$\hat{\theta} \pm z^{(1-\alpha)}\hat{\mathsf{se}},$$

    where z^{α} is the $100 \cdot \alpha$ the percentile point of $\mathrm{N}(0,1)$.

```r
## z^{.05}
qnorm(.05)
```

```
## [1] -1.644854
```

```r
## z^{.95}
qnorm(.95)
```

```
## [1] 1.644854
```

Standard confidence interval (Example)

- **ET** Table 2.1.
- 16 mice were randomly assigned to a treatment or a control group.
- Their survival time in days, following a surgery was recorded.
- Construct 90% confidence interval for the expectation θ of the control group distribution.

```
Table2.1.ET = list(treatment = c(94, 197,
  16, 38, 99, 141, 23),
  control = c(52, 104, 146, 10,
    51, 30, 40, 27, 46))
```

Standard confidence interval (Example)

$1 - 2\alpha = .9$. Thus, $\alpha = .05$.

```
x = Table2.1.ET$control
n = length(x)
theta.hat = round(mean(x), digits = 2)
theta.hat
```

```
## [1] 56.22
```

```
se.theta.hat = round(sd(x)/sqrt(n), digits = 2)
se.theta.hat
```

```
## [1] 14.16
```

```
## z^{1-.05}
z = round(qnorm(.95), digits = 3)
z
```

```
## [1] 1.645
```

Standard confidence interval (Example)

- 90% confidence interval for the expectation θ of the control group distribution is

```
ci.standard = round(theta.hat + c(-1, 1)*z*se.theta.hat,
  digits = 2); ci.standard
```

```
## [1] 32.93 79.51
```

- 90% of time, a random interval constructed in this way will contain the true value θ.

Standard confidence interval (Note)

- $\dfrac{\hat{\theta}-\theta}{\hat{\mathsf{se}}} \dot{\sim} \mathsf{N}(0,1)$ is valid as $n \to \infty$, but is approximation for finite samples.
- Thus, for the example with $n = 9$, actually the standard CI is an approximate CI.
    - The coverage probability is not exactly $1 - 2\alpha$.

Student's t-interval (procedure)

- Improve upon the standard confidence interval.

$$Z = \frac{\hat{\theta} - \theta}{\hat{\mathsf{se}}} \dot{\sim} \mathsf{t}_{n-1},$$

  where t_{n-1} is the Student's t distribution on $n-1$ degrees of freedom.
- Student's t-interval is

$$\left[\hat{\theta} - t_{n-1}^{(1-\alpha)}\hat{\mathsf{se}}, \;\; \hat{\theta} - t_{n-1}^{(\alpha)}\hat{\mathsf{se}}\right].$$

Student's t-interval (Example)

```r
##t^{.05}
qt(.05, df = 8)
```

```
## [1] -1.859548
```

```r
##t^{.95}
qt(.95, df = 8)
```

```
## [1] 1.859548
```

Student's t-interval (Example)

- 90% confidence interval for the expectation θ of the control group distribution is

```
ci.student.t = round(theta.hat + c(qt(.05, df = 9),
  qt(.95, df = (length(x)-1)))*se.theta.hat,
  digits = 2)
ci.student.t
```

```
## [1] 30.26 82.55
```

- Student's t-interval is wider than the standard interval

```
ci.standard
```

```
## [1] 32.93 79.51
```

Student's t-interval (Note)

- Student's t-interval widening the interval to adjust for the fact that the standard error is unknown.

Student's t-interval (Note)

- Increase $n(\geq 20)$, percentiles of t_n distribution don't differ much from the standard normal $\mathrm{N}\,(0, 1)$.

```
##(t^{.05}, t^{.95})
c(qt(.05, df = 50), qt(.95, df = 50))
```

```
## [1] -1.675905  1.675905
```

```
## (z^{.05},  z^{.95})
c(qnorm(.05), qnorm(.95))
```

```
## [1] -1.644854  1.644854
```

Student's t-interval (Note)

- The use of the t distribution doesn't adjust the CI to account for skewness in the underlying population or other errors when $\hat{\theta}$ is not the sample mean (for example, bias of an estimate).

The bootstrap-t interval (overview)

- Adjust for the above errors.
- Construct CI without having

$$Z = \frac{\hat{\theta} - \theta}{\hat{\mathsf{se}}} \,\dot{\sim}\, \mathsf{N}\left(0, 1\right) \quad \text{or} \quad Z = \frac{\hat{\theta} - \theta}{\hat{\mathsf{se}}} \,\dot{\sim}\, \mathsf{t}_{n-1}$$

.

- Estimate the distribution of Z directly from the data.

The bootstrap-t interval (procedure)

- The bootstrap-t method
  - Generate B bootstrap samples $\boldsymbol{x}^{*1}, \boldsymbol{x}^{*2}, \cdots, \boldsymbol{x}^{*B}$.
  - For each compute

$$Z^*(b) = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\text{se}}^*(b)},$$

    where
    - $\hat{\theta}^*$ is the value of $\hat{\theta}$ for the bootstrap sample $\boldsymbol{x}^{*b}$
    - $\hat{\text{se}}^*(b)$ is the estimated standard error of $\hat{\theta}^*$.
  - Let k be the largest integer less than or equal to $(B+1)\alpha$.
    - $\hat{\text{t}}^{(1-\alpha)}$ - the empirical α quantile is the k-th largest value of $Z^*(b)$.
    - $\hat{\text{t}}^{(\alpha)}$ - the empirical $1-\alpha$ quantile is the $(B+1-k)$-th largest value of $Z^*(b)$.
  - The bootstrap-t confidence interval is

$$\left[\hat{\theta} - \hat{\text{t}}^{(1-\alpha)}\hat{\text{se}}, \quad \hat{\theta} - \hat{\text{t}}^{(\alpha)}\hat{\text{se}}\right].$$

The bootstrap-t interval (Example)

- 90% Bootstrap-t interval for the expectation θ of the control group.
- $\alpha = .05$.

```
B = 1000
n = length(x)
theta.hat = mean(x); theta.hat
```

```
## [1] 56.22222
```

```
se.theta.hat = sd(x)/sqrt(n); se.theta.hat
```

```
## [1] 14.1586
```

The bootstrap-t interval (Example)

```
z.star = function(x){
  n = length(x)
  x.star = sample(x, size = n,
    replace = TRUE)
  theta.hat.star = mean(x.star)
  se.theta.hat.star = sd(x.star)/sqrt(n)
  z.star.b = (theta.hat.star -
      theta.hat)/(se.theta.hat.star)
  return(z.star.b)
}
z.star.B = replicate(B, z.star(x))
```

The bootstrap-t interval (Example)

```
#k is the largest integer less than
#or equal to (B+1) * alpha.
alpha = .05
k = ceiling((B+1)*alpha)
t.hat.one.minus.alpha = sort(z.star.B,
  decreasing = TRUE)[k]
t1 = t.hat.one.minus.alpha; t1
```

```
## [1] 1.468585
```

```
k.u = B+1-k
t.hat.alpha = sort(z.star.B,
  decreasing = TRUE)[k.u]
t2 = t.hat.alpha; t2
```

```
## [1] -4.267706
```

The bootstrap-t interval (Example)

- 90% confidence interval for the expectation θ of the control group distribution is

```
ci.bootstrap.t = round(theta.hat -
    c(t1, t2)*se.theta.hat, digits = 2)
ci.bootstrap.t
```

```
## [1]  35.43 116.65
```

- The lower end point is close to the standard interval,

```
ci.standard
```

```
## [1] 32.93 79.51
```

```
- but upper end point is much greater (reflect the two very
```

The bootstrap-t interval (Note)

- For large samples, the coverage of the bootstrap-t interval tends to be closer to the desired interval than the coverage of the standard and Student-t intervals.
- The bootstrap-t table applies only to the given sample.
- Standard and Student-t distributions are symmetric about zero, thus, the CIs are symmetric about $\hat{\theta}$.
- The bootstrap-t percentiles can be asymmetric about 0, so CI can be longer on the left or right.
  - This property improves the coverage of the bootstrap-t CI.

Pivotal statistic

- If
$$Z = \frac{\hat{\theta} - \theta}{\hat{\mathsf{se}}}$$
is called an approximate pivot.
  - The distribution of Z is approximately the same for each value of θ.
- If Z is a pivotal statistic, then the distribution of Z does not depend on any unknown parameters.

The bootstrap-t interval (Note)

- Bootstrap-t particularly applicable to location statistics (sample mean, median, trimmed mean, sample percentile)
  - location statistic: increasing data value x_i by a constant c increases the statistic by c.
- Bootstrap-t may not have the correct coverage with its simple form.
  - For example, CI for correlation coefficient.
- We require computing $\hat{\text{se}}^*(b)$ using bootstrap or jackknife for which there is no simple standard error formula.
  - For the example, where $\hat{\theta}$ is the sample mean, we use the plug-in estimate of $\hat{\text{se}}^*(b)$ for each bootstrap sample $\boldsymbol{x}^{*b}$.

Nested levels of bootstrap sampling (Example)

- Use bootstrap estimate of standard error for each bootstrap sample (two nested levels of bootstrap sampling).
  - Let's choose $B = 25$ to estimate standard error.

Nested levels of bootstrap sampling (Example)

```r
library(magrittr)
B = 1000; B2 = 25
n = length(x); theta.hat = mean(x)
#se.theta.hat = sd(x)/sqrt(n)
bootstrap.results = lapply(as.list(1:B), function(b){
  x.star  = sample(x, size = n, replace = TRUE)
  theta.hat.star = mean(x.star)
  theta.hat.star.star.B2 = lapply(as.list(1:B2),
    function(bb){
    x.star.star = sample(x.star,
      size = n, replace = TRUE)
    theta.hat.star.star = mean(x.star.star)
    return(theta.hat.star.star)
  }) %>% unlist
  return(list(theta.hat.star, theta.hat.star.star.B2))
})
```

Nested levels of bootstrap sampling (Example)

$\hat{\theta}^{*b}$, $b = 1, 2, \cdots, 1000$ and compute $\widehat{\mathrm{se}}\left(\hat{\theta}\right)$.

```
theta.hat.star = lapply(bootstrap.results,
  '[[', 1) %>% unlist
se.theta.hat = sd(theta.hat.star)
se.theta.hat
```

```
## [1] 13.24959
```

Nested levels of bootstrap sampling (Example)

- For each $b = 1, 2, \cdots, 1000$, compute $\hat{\text{se}}^*(b)$ using $\hat{\theta}^{**b2}$, $b2 = 1, 2, \cdots, 25$.

```
theta.hat.star.star.B2 = lapply(bootstrap.results,
  '[[', 2)
se.theta.hat.star = lapply(theta.hat.star.star.B2,
  sd) %>% unlist
z.star.B = (theta.hat.star
  - theta.hat)/se.theta.hat.star
```

Nested levels of bootstrap sampling (Example)

```
alpha = .05
k = ceiling((B+1)*alpha)
t.hat.one.minus.alpha = sort(z.star.B,
  decreasing = TRUE)[k]
t1 = t.hat.one.minus.alpha; t1
```

```
## [1] 1.555962
```

```
k.u = B+1-k
t.hat.alpha = sort(z.star.B,
  decreasing = TRUE)[k.u]
t2 = t.hat.alpha; t2
```

```
## [1] -4.386964
```

Nested levels of bootstrap sampling (Example)

- 90% confidence interval for θ using nested bootstrap

```
ci.bootstrap.t.nested = round(theta.hat -
    c(t1, t2)*se.theta.hat,
  digits = 2)
ci.bootstrap.t.nested
```

```
## [1]  35.61 114.35
```

- Similar to bootstrap-t

```
ci.bootstrap.t
```

```
## [1]  35.43 116.65
```

Transformation and bootstrap-t (overview)

- Use transformation to overcome issues in bootstrap-t interval in small-sample, nonparametric setting.
- Example of Law school data.
  - Parameter of interest is on correlation coefficient θ of LSAT and GPA.

```
library(bootstrap)
data(law)
t(law)
```

```
##           1      2      3      4      5      6   7      8
## LSAT 576.00 635.0 558.00 578.00 666.00 580.00 555 661.00
## GPA    3.39   3.3   2.81   3.03   3.44   3.07   3   3.43
##          11     12     13     14     15
## LSAT 653.00 575.00 545.00 572.00 594.00
## GPA    3.12   2.74   2.76   2.88   2.96
```

Transformation and bootstrap-t (Example)

- Construct CI for θ without any transformation.
- Use two nested levels bootstrap.

```
B = 1000
B2 = 25
n = dim(law)[1]
theta.hat = cor(law$LSAT, law$GPA)
theta.hat
```

```
## [1] 0.7763745
```

Transformation and bootstrap-t (Example)

```
bootstrap.results = lapply(as.list(1:B), function(b){
  x = law
  x.star  = x[sample(1:n, size = n,
    replace = TRUE),]
  theta.hat.star = cor(x.star$LSAT,
    x.star$GPA)
  theta.hat.star.star.B2 = lapply(as.list(1:B2),
    function(bb){
    x.star.star = x.star[sample(1:n, size = n,
      replace = TRUE),]
    theta.hat.star.star = cor(x.star.star$LSAT,
      x.star.star$GPA)
    return(theta.hat.star.star)
  }) %>% unlist
  return(list(theta.hat.star,
    theta.hat.star.star.B2))
})
```

Transformation and bootstrap-t (Example)

- without any transformation

```
theta.hat.star = lapply(bootstrap.results,
  '[[', 1) %>% unlist
se.theta.hat = sd(theta.hat.star)
se.theta.hat
```

```
## [1] 0.1362079
```

Transformation and bootstrap-t (Example)

- without any transformation

```
theta.hat.star.star.B2 = lapply(bootstrap.results,
  '[[', 2)
se.theta.hat.star = lapply(theta.hat.star.star.B2,
  sd) %>% unlist
z.star.B = (theta.hat.star -
    theta.hat)/se.theta.hat.star
```

Transformation and bootstrap-t (Example)

- without any transformation
- 90% confidence interval for θ (correlation coefficient)

```
alpha = .05
k = ceiling((B+1)*alpha)
t.hat.one.minus.alpha = sort(z.star.B,
  decreasing = TRUE)[k]
t1 = t.hat.one.minus.alpha; t1
```

```
## [1] 5.122292
```

```
k.u = B+1-k
t.hat.alpha = sort(z.star.B,
  decreasing = TRUE)[k.u]
t2 = t.hat.alpha; t2
```

```
## [1] -1.336643
```

Transformation and bootstrap-t (Example)

- without any transformation

```
corr.ci.boot.t.no.tran.90 = round(theta.hat -
    c(t1, t2)*se.theta.hat,
  digits = 2)
corr.ci.boot.t.no.tran.90
```

```
## [1] 0.08 0.96
```

Transformation and bootstrap-t (Example)

- without any transformation
- 98% confidence interval for θ (correlation coefficient)

```
alpha = .01
k = ceiling((B+1)*alpha)
t.hat.one.minus.alpha = sort(z.star.B,
  decreasing = TRUE)[k]
t1 = t.hat.one.minus.alpha; t1
```

```
## [1] 12.31386
```

```
k.u = B+1-k
t.hat.alpha = sort(z.star.B,
  decreasing = TRUE)[k.u]
t2 = t.hat.alpha; t2
```

```
## [1] -2.189701
```

Transformation and bootstrap-t (Example)

- without any transformation

```
corr.ci.boot.t.no.tran.98 = round(theta.hat -
    c(t1, t2)*se.theta.hat,
  digits = 2)
corr.ci.boot.t.no.tran.98
```

```
## [1] -0.90  1.07
```

Transformation and bootstrap-t

- Untransformed bootstrap-t procedure may lead intervals which are often too wide and fall outside of allowable range for a parameter.

Transformation and bootstrap-t (Example)

- Let
$$\phi = .5\log\left(\frac{1+\theta}{1-\theta}\right).$$
- Construct CI for ϕ.
- Transform the endpoints back with the inverse transformation
$$\frac{e^{2\phi}-1}{e^{2\phi}+1}$$
to obtain better interval for θ.

```
phi = function(cor.coeff){
  .5*(log(1 + cor.coeff) - log(1 - cor.coeff))
}
```

Transformation and bootstrap-t (Example)

- Confidence interval for ϕ.

```
B = 1000
B2 = 25
n = dim(law)[1]
theta.hat = phi(cor(law$LSAT, law$GPA))
theta.hat
```

```
## [1] 1.036178
```

Transformation and bootstrap-t (Example)

- Confidence interval for ϕ.

```
bootstrap.results = lapply(as.list(1:B), function(b){
  x = law
  x.star  = x[sample(1:n, size = n,
    replace = TRUE),]
  theta.hat.star = phi(cor(x.star$LSAT,
    x.star$GPA))
  theta.hat.star.star.B2 = lapply(as.list(1:B2),
    function(bb){
    x.star.star = x.star[sample(1:n, size = n,
      replace = TRUE),]
    theta.hat.star.star = phi(cor(x.star.star$LSAT, x.star
    return(theta.hat.star.star)
  }) %>% unlist
  return(list(theta.hat.star,
    theta.hat.star.star.B2))
})
```

Transformation and bootstrap-t (Example)

- Confidence interval for ϕ.

```
theta.hat.star = lapply(bootstrap.results,
  '[[', 1) %>% unlist
se.theta.hat = sd(theta.hat.star)
se.theta.hat
```

```
## [1] 0.3723111
```

```
theta.hat.star.star.B2 = lapply(bootstrap.results,
  '[[', 2)
se.theta.hat.star = lapply(theta.hat.star.star.B2,
  sd) %>% unlist
z.star.B = (theta.hat.star -
    theta.hat)/se.theta.hat.star
```

Transformation and bootstrap-t (Example)

- 90% confidence interval for ϕ

```
alpha = .05
k = ceiling((B+1)*alpha)
t.hat.one.minus.alpha = sort(z.star.B,
  decreasing = TRUE)[k]
t1 = t.hat.one.minus.alpha; t1
```

```
## [1] 2.136466
```

```
k.u = B+1-k
t.hat.alpha = sort(z.star.B,
  decreasing = TRUE)[k.u]
t2 = t.hat.alpha; t2
```

```
## [1] -1.642458
```

Transformation and bootstrap-t (Example)

- 90% confidence interval for ϕ

```
ci.phi.bootstrap.t = round(theta.hat -
    c(t1, t2)*se.theta.hat,
  digits = 2)
ci.phi.bootstrap.t
```

```
## [1] 0.24 1.65
```

Transformation and bootstrap-t (Example)

- 90% confidence interval for θ (correlation coefficient)

```
l.phi = round(theta.hat -
    c(t1, t2)*se.theta.hat,
  digits = 2)[1]
u.phi = round(theta.hat -
    c(t1, t2)*se.theta.hat,
  digits = 2)[2]
corr.ci.boot.t.tran.90 = round(c((exp(2*l.phi)-1)/(exp(2*l
  (exp(2*u.phi)-1)/(exp(2*u.phi)+1)), digits = 2)
corr.ci.boot.t.tran.90
```

```
## [1] 0.24 0.93
```

Transformation and bootstrap-t (Note)

- bootstrap-t depends on the scale - some scales better than others.
- $$\phi = .5\text{log}\left(\frac{1+\theta}{1-\theta}\right)$$
  is appropriate when (X, Y) are bivariate normal.
- What transformation to use?
  - Use bootstrap to estimate the appropriate transformation.
  - We need to variance stablize the estimate $\hat{\theta}$.
    - Make variance of $\hat{\theta}$ is approximately contstant.

Transformation and bootstrap-t (Choosing the transformation)

- X is a random variable with mean θ and standard deviataion $s(\theta)$.
- Find a transformation g such that

$$g(x) = \int^{x} \frac{1}{s(u)} du.$$

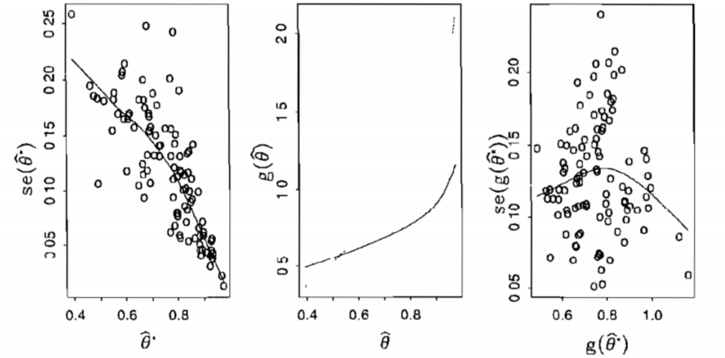
- Then, variance of $g(X)$ is constant.
- $s(u)$ is unknown, but we can write $s(u) = \text{se}\left(\hat{\theta} | \theta = u\right)$.
  - Use bootstrap to estimate $\text{se}\left(\hat{\theta} | \theta = u\right)$.

Transformation and bootstrap-t (Choosing the transformation)

- Generate $B = 100$ of x^{*b}, compute $\hat{\theta}^{*}(b)$.
  - Sample from x^{*b}: $R = 25$ bootstrap samples of x^{**r}.
  - Compute $\hat{\theta}^{**}(r)$ and $\hat{\text{se}}\left(\hat{\theta}^{*}(b)\right)$.

Transformation and bootstrap-t (Choosing the transformation)

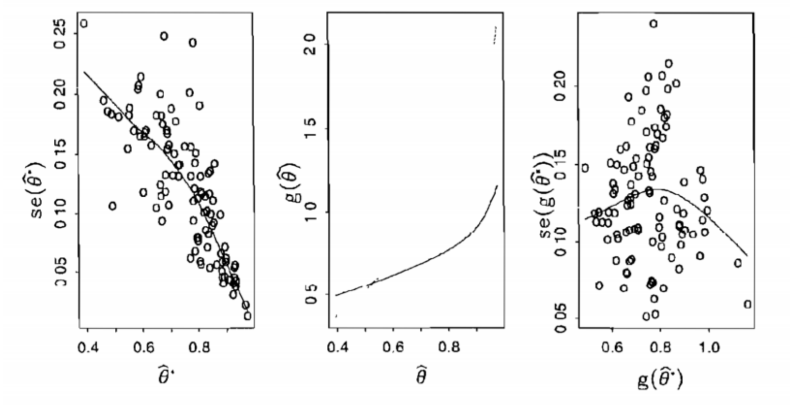
- Fit a curve to the points $\left[\hat{\theta}^*(b),\ \hat{\text{se}}\left(\hat{\theta}^*(b)\right)\right]$.



Source: Efron and Tibshirani (1994)

Transformation and bootstrap-t (Choosing the transformation)

- Estimate the variance stabilizing transformation $g\left(\hat{\theta}\right)$ - use numerical integration.



Source: Efron and Tibshirani (1994)

Transformation and bootstrap-t (procedure)

- Use $B = 1000$ bootstrap samples to construct CI for $\phi = g(\theta)$
  - set the denominator in $\dfrac{g\left(\hat{\theta}^*\right) - g\left(\hat{\theta}\right)}{\hat{\text{se}}^*}$ to 1.

Transformation and bootstrap-t (Example)

- Law school data.
- CI for correlation coefficient θ between LSAT and GPA.

```
xdata = law %>% as.matrix
n = dim(xdata)[1]
theta = function(x, xdata){
  cor(xdata[x,1], xdata[x,2])
  }
results = boott(1:n,theta, xdata,
  VS = TRUE, perc = c(.01,.05, .95, .99))
```

Transformation and bootstrap-t (Example)

- 90% CI

```
round(c(results$confpoints[2],
  results$confpoints[3]), digits = 2)
```

```
## [1] 0.3 0.9
```

- 98% CI

```
round(c(results$confpoints[1],
  results$confpoints[4]), digits = 2)
```

```
## [1] 0.10 0.96
```

- bootstrap-t intervals with transformation are shorter than those without transformation.
- CIs are within the permissible values.
- No need to do nested bootstrap sampling.
- Next, we will work directly with the bootstrap distribution of $\hat{\theta}$ and derive a transformation-respecting confidence interval

Bias-corrected and accelerated bootstrap - BCa (Overview)

- BCa interval endpoints are also given by percentile distribution after correction for bias and skewness.
- Recall: percentile method
  - Bootstrap replicates $\hat{\theta}^*(1), \hat{\theta}^*(2), \cdots, \hat{\theta}^*(B)$.
  - The percentile interval

$$\left(\hat{\theta}_{lo}, \hat{\theta}_{up}\right) = \left(\hat{\theta}^{(\alpha)}, \hat{\theta}^{(1-\alpha)}\right),$$

where $\hat{\theta}^{(\alpha)}$ is the $100 \cdot \alpha$th percentile of B bootstrap replicates.

BCa bootstrap procedure

- Assume that there is a monotone increasing transformation g such that
$$\phi = g\left(\theta\right) \quad \text{and} \quad \hat{\phi} = g\left(\hat{\theta}\right).$$
- The BCa bootstrap bootstrap is based on the following model
$$\frac{\hat{\phi} - \phi}{\sigma_\phi} \sim \mathrm{N}\left(-z_0, 1\right) \quad \text{with} \quad \sigma_\phi = 1 + a\phi.$$
- This is a generalization of th usual normal approxiamtion
$$\frac{\hat{\theta} - \theta}{\mathrm{se}} \sim \mathrm{N}\left(0, 1\right).$$
  - generalization: transformation $g\left(\cdot\right)$, the bias correction z_0, and the acceleration a.

BCa bootstrap procedure

- The BCa interval of intended coverage $1 - 2\alpha$, is given by

$$\left(\hat{\theta}_{lo}, \hat{\theta}_{up}\right) = \left(\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}\right),$$

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{\alpha}}{1 - \hat{a}\left(\hat{z}_0 + z^{\alpha}\right)}\right)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{1-\alpha}}{1 - \hat{a}\left(\hat{z}_0 + z^{1-\alpha}\right)}\right)$$

BCa bootstrap procedure

- and $\Phi(\cdot)$ is the standard normal cumulative distirbution function
- z^{α} - 100αth percentile point of $\mathrm{N}(0,1)$.
- $\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{\theta}^*(b) < \hat{\theta}\}}{B}\right)$,
  - where Φ^{-1} is the inverse function of the $\mathrm{N}(0,1)$.
- $\hat{a} = \frac{\sigma_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left\{\sigma_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right\}^{3/2}}$.
  - Compute the estimate by deleting i-th observation, $\hat{\theta}_{(i)}$.
  - $\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^{n}\hat{\theta}_{(i)}}{n}$.

BCa bootstrap (Example)

- 90% CI for correlation between LSAT and GPA (law school data)

```
library(bootstrap)
xdata = law %>% as.matrix()
n = dim(xdata)[1]
theta = function(x, xdata){
  cor(xdata[x,1], xdata[x,2])
}
results = bcanon(1:n, 100, theta, xdata)
corr.ci.bca.90 = c(results$confpoints[2,2],
  results$confpoints[7,2]);
corr.ci.bca.90
```

```
## bca point bca point
## 0.4096908 0.9176347
```

Comparison of bootstrap intervals (Example)

```
corr.ci.boot.t.no.tran.90
```

```
## [1] 0.08 0.96
```

```
corr.ci.boot.t.tran.90
```

```
## [1] 0.24 0.93
```

```
round(corr.ci.bca.90, digits = 2)
```

```
## bca point bca point
##      0.41      0.92
```

BCa bootstrap(Note)

- The bootstrap-t method is second-order accurate, but not transformation respecting.
- The percentile method is transformation respecting but not second-order accurate.
- BCa is second-order accurate and transformation respecting.
- We can reduce the computation cost for BCa for **smooth estimates**.

The ABC method (Overview)

- The approximate bootstrap confidence intervals
- Approxiamting the BCa interval endpoints analytically - no need Monte Carlo replications.

The ABC method (Example)

- 90% CI for correlation between LSAT and GPA (law school data)

```
x = law %>% as.matrix()
theta = function(p, x){
  x1m = sum(p*x[, 1])/sum(p)
  x2m = sum(p*x[, 2])/sum(p)
  numerator = sum(p*(x[, 1]- x1m)*(x[, 2]- x2m))
  denominator = sqrt(sum(p*(x[, 1] - x1m)^2)*sum(p*(x[, 2]
  return(numerator/denominator)
}

results = abcnon(x, theta)
round(c(results$limits[2,2], results$limits[7,2]), 2)
```

```
##  abc  abc
## 0.44 0.92
```

References for this lecture

ET Chapter 12.