Kapitel 5

Einführung in Bootstrap

Literatur zum Thema:

- Efron B., Tibshirani R.J.: An Introduction to the Bootstrap (1993)
- Hall P.: The Bootstrap and Edgeworth Expansion (1992)
- Davison A.C.: Recent Developments in Bootstrap Methodology, Statistical Science (2003), Vol. 18, No. 2, pp. 141-157

5.1 Einführung

- *Bootstrap* (engl.): Stiefelriemen, Stiefelschlaufe
- „Sich selbst am Schopf aus dem Sumpf ziehen” $\rightarrow$ Lügenbaron Münchhausen (mit Pferd)
- Computergestützte Methode
- Beruht auf wiederholtem Ziehen (*Resampling*) aus den beobachteten Daten.
- *Ziel:* Schätzung von Varianz, Bias oder Verteilung einer Statistik $T = T(X_1, \ldots, X_n)$, Konfidenzintervalle, Tests.
- Wann? In Situationen, in denen
  (a) asymptotische Aussagen fragwürdig sind (kleine Stichprobenumfänge),
  (b) analytische Berechnungen sehr kompliziert oder unmöglich sind, zum Beispiel wenn keine parametrischen Verteilungsannahmen gemacht werden sollen. $\rightarrow$ Bootstrap für nichtparametrische Schätzungen.
- Funktioniert „Bootstrap” immer? Nein, nicht immer (Bootstrap kann *inkonsistent* sein), aber oft.

5.1.1 Grundidee

Einstichproben-Problem: $X = (X_1, \ldots, X_n)$, $X_i \overset{i.i.d.}{\sim} F$, F unbekannt

Interessierende Statistik: $T(X)$

Beobachtete Daten: $x = (x_1, x_2, \ldots, x_n) \to T(x)$

Bootstrap-Stichprobe: Ziehe n mal *mit* Zurücklegen *zufällig* aus $(x_1, \ldots, x_n)$. Wir erhalten

$$x^* = (x_1^*, x_2^*, \ldots, x_n^*) \to T(x^*).$$

Beispiel: $x = (1, 2, 5)$, $n = 3$. $x^* = (1, 1, 5)$ ist eine mögliche Bootstrap-Stichprobe.

Also:

(1) Werte aus der ursprünglichen Stichprobe x können in der Bootstrap-Stichprobe

   (i) einmal vorkommen,

   (ii) mehrfach vorkommen,

   (iii) gar nicht vorkommen.

(2) Die Bootstrap-Stichprobe hat ebenfalls Stichprobenumfang n.

Skizze:

$$
\begin{array}{ccccc}
 & & x = (x_1, \ldots, x_n) \ \text{Daten} & & \\
 \swarrow & \swarrow & \cdots & & \searrow \\
x^{*1} & x^{*2} & & & x^{*B} \\
| & | & & & | \\
T(x^{*1}) & T(x^{*2}) & \ldots & & T(x^{*B})
\end{array}
$$

B: Anzahl von Bootstrap-Stichproben

Mit den berechneten Statistiken $T(x^{*1}), \ldots, T(x^{*B})$ lassen sich Aussagen über die Verteilung von T gewinnen, zum Beispiel

$$\text{Var}_F(T) \ \approx \ \widehat{\text{Var}}_{\text{Boot}}(T) = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} \left[T(x^{*b}) - \bar{T}_{\text{Boot}} \right]^2 \right\}$$

mit

$$\bar{T}_{\text{Boot}} = \frac{1}{B} \sum_{b=1}^{B} T(x^{*b}).$$

5.1.2 Empirische Verteilungsfunktion und das Plug-In-Prinzip

$X = (X_1, \ldots, X_n)$, $X_i \overset{i.i.d.}{\sim} F$, F unbekannt

$x = (x_1, x_2, \ldots, x_n)$ Daten

Empirische Verteilungsfunktion:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x),$$

wobei I die Indikatorfunktion ist.

Plug-In-Prinzip: F durch $\hat{F}_n$ ersetzen.

Beispiel 5.1.

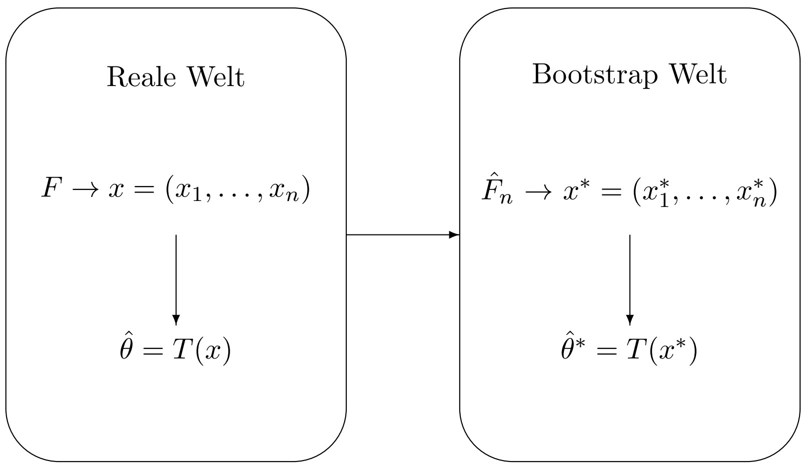
$$
\begin{aligned}
T(F) &= \mu = \int x \mathrm{d}F(x) \\
T(\hat{F}_n) &= \int x \mathrm{d}\hat{F}_n(x) \\
&= \sum_{i=1}^{n} x_i \hat{P}_n(X = x_i) \qquad \textit{(o.w.E. seien alle } x_i \textit{ verschieden)} \\
&= \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}
$$

Plug-In-Prinzip hat Sinn, wenn *keine weiteren Informationen über F* vorhanden sind außer der Stichprobe.

→ „nichtparametrisches Setup”

5.1.3 Reale Welt und Bootstrap-Welt

Wiederum Einstichproben-Fall:



- Die unbekannte Verteilung F liefert x als Zufallsstichprobe.
- Die empirische Verteilung $\hat{F}_n$ liefert x^* als zufällige Bootstrap-Stichprobe.
- Die interessierende Statistik $\hat{\theta} = T(x)$ ist Funktion der Zufallsstichprobe.
- Die Bootstrap-Replikation $\hat{\theta}^* = T(x^*)$ ist Funktion der Bootstrap-Stichprobe.

⇒ Im Allgemeinen kann F bzw. $\hat{F}_n$ in obiger Abbildung durch ein *geschätztes* Wahrscheinlichkeitsmodell P bzw. $\hat{P}_n$ ersetzt werden.

5.1.4 Die ideale Bootstrap-Verteilung

Daten $x = (x_1, x_2, \ldots, x_n)$.

Frage: Wie viele verschiedene Bootstrap-Stichproben gibt es?

Beispiel 5.2. *Sei $x = (1, 2, 5)$. Die Anordnung spielt hier keine Rolle. Wegen $n = 3$ gibt es 10 verschiedene Bootstap-Stichproben (wenn alle x_i verschieden sind):*

$$(1, 1, 1), (2, 2, 2), (5, 5, 5), (1, 1, 2), (1, 1, 5), (2, 2, 5), (1, 2, 2), (1, 5, 5), (2, 5, 5), (1, 2, 5).$$

Die ideale Bootstrap-Schätzung ist die, welche sich durch Berücksichtigung aller möglichen Bootstrap-Stichproben ergibt. Die ideale Bootstrap-Schätzung zum Beispiel für die Varianz von $\hat{\theta} = \text{median}(X)$ in Beispiel 5.2 wäre dabei die Varianz über die 10 Bootstrap-Stichproben. Dabei ist allerdings zu berücksichtigen, dass die Stichproben mit unterschiedlicher Wahrscheinlichkeit gezogen werden.

Beispiel 5.3 (Fortsetzung von Beispiel 5.2)**.** *Mit Hilfe der Multinomialverteilung erhält man*

$$\begin{aligned}
P\Big(x^* = (1,1,1)\Big) &= \frac{3!}{3!0!0!} \cdot \left(\frac{1}{3}\right)^3 \left(\frac{1}{3}\right)^0 \left(\frac{1}{3}\right)^0 = \left(\frac{1}{3}\right)^3 = \left(\frac{1}{27}\right), \\
P\Big(x^* = (2,5,5)\Big) &= \frac{3!}{0!1!2!} \cdot \left(\frac{1}{3}\right)^0 \left(\frac{1}{3}\right)^1 \left(\frac{1}{3}\right)^2 = 3 \cdot \left(\frac{1}{3}\right)^3 = \left(\frac{1}{9}\right), \\
P\Big(x^* = (1,2,5)\Big) &= \frac{3!}{1!1!1!} \cdot \left(\frac{1}{3}\right) \left(\frac{1}{3}\right) \left(\frac{1}{3}\right) = 6 \cdot \left(\frac{1}{3}\right)^3 = \left(\frac{2}{9}\right),
\end{aligned}$$

denn zum Beispiel $(2, 5, 5) \mathrel{\hat{=}} (5, 2, 5) \mathrel{\hat{=}} (5, 5, 2) \mathrel{\hat{=}} (1, 2, 5) \mathrel{\hat{=}} \ldots \mathrel{\hat{=}} (5, 2, 1)$.

Betrachte $\hat{\theta} = \text{median}(X)$. Dann ist $\hat{\theta}(x) = 2$ die Schätzung aus der Stichprobe und

$$\begin{aligned}
\text{Var}_{\hat{F}_n}(\hat{\theta}^*) &= \left(\frac{1}{3}\right)^3 \{(1-c)^2 + (2-c)^2 + (5-c)^2 \\
&\quad + 3 \cdot [(1-c)^2 + (1-c)^2 + (2-c)^2 + (2-c)^2 + (5-c)^2 + (5-c)^2] \\
&\quad + 6 \cdot (2-c)^2\} \\
&= 2.32,
\end{aligned}$$

wobei

$$\begin{aligned}
c = \bar{\hat{\theta}}^* &= \left(\frac{1}{3}\right)^3 [1 + 2 + 5 + 3 \cdot (1 + 1 + 2 + 2 + 5 + 5) + 6 \cdot 2] \\
&= \left(\frac{1}{3}\right)^3 [8 + 3 \cdot 16 + 12] = \left(\frac{1}{3}\right)^3 \cdot 68 = \frac{68}{27} \approx 2.5
\end{aligned}$$

der Mittelwert aller geschätzten Mediane ist.

Allgemein gibt es, sofern alle n Datenpunkte $x_1, \ldots, x_n$ *verschieden* sind, $\binom{2n-1}{n}$ mögliche Bootstrap-Stichproben.

$$\begin{aligned} n = 3: &\quad \binom{5}{3} = 10 \\ n = 15: &\quad \binom{29}{15} = 77\,558\,760 \\ n = 20: &\quad \binom{39}{20} = 68\,923\,264\,410 \end{aligned}$$

Das heißt, wenn n nicht sehr klein ist, dann ist es praktisch nicht möglich, die ideale Bootstrap-Verteilung zu verwenden. Stattdessen begnügt man sich mit einer Anzahl $B \ll \binom{2n-1}{n}$ von Bootstrap-Stichproben.

5.2 Bootstrap–Schätzung eines Standardfehlers

Einstichproben-Fall: $X = (X_1, \ldots, X_n)$, $X_i \overset{\text{i.i.d.}}{\sim} F$, F unbekannt

Daten: $x = (x_1, \ldots, x_n)$

Ziel dieses Abschnitts ist die Schätzung des Standardfehlers eines Schätzers $\hat{\theta} = \hat{\theta}(X)$ für $\theta = T(F)$. Hierbei *kann* $\hat{\theta}(X)$ die Plug-In-Schätzung $T(\hat{F}_n)$ sein, muss aber nicht.

Frage: Wie gut ist die Schätzung $\hat{\theta}$?

5.2.1 Bootstrap-Algorithmus zur Schätzung des Standardfehlers

Algorithmus 12: Bootstrap-Algorithmus zur Schätzung des Standardfehlers

1. Erzeuge B Bootstrap-Stichproben $x^{*1}, \ldots, x^{*B}$.
2. Berechne $\hat{\theta}^*(b)$, $b = 1, \ldots, B$.
3. Schätze den Standardfehler $\mathrm{se}_F(\hat{\theta}) = \sqrt{\mathrm{Var}_F(\hat{\theta})}$ durch

$$\widehat{\mathrm{se}}_B = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} \left[\hat{\theta}^*(b) - \hat{\theta}^*(\cdot) \right]^2 \right\}^{\frac{1}{2}} \quad \text{mit} \quad \hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*(b).$$

Die Bootstrap-Schätzung für den Standardfehler $\mathrm{se}_F(\hat{\theta})$ einer Schätzung $\hat{\theta}$ (Daten aus F) ist also der Standardfehler für zufällige Stichproben vom *Umfang* n gezogen aus $\hat{F}_n$ *mit* Zurücklegen.

Es gilt:

$$\lim_{B \to \infty} \widehat{\mathrm{se}}_B = \mathrm{se}_{\hat{F}_n}(\hat{\theta}^*).$$

Die ideale Bootstrap-Schätzung $\mathrm{se}_{\hat{F}_n}(\hat{\theta}^*)$ und die Approximation $\widehat{\mathrm{se}}_B$ werden oft als *nichtparametrische Bootstrap-Schätzung* bezeichnet, da sie nur auf $\hat{F}_n$ beruhen und $\hat{F}_n$ die nichtparametrische Schätzung für F ist.

$\rightarrow$ Abschnitt 5.2.3: Parametrischer Bootstrap (F wird nicht mehr durch $\hat{F}_n$ geschätzt).

Beispiel 5.4. *Zwei (quasi-) stetige Merkmale Y und Z werden an n Individuen erhoben, d.h.*

$$X = ((Y_1, Z_1), (Y_1, Z_1), \dots, (Y_n, Z_n)), \quad (Y_i, Z_i) \overset{i.i.d.}{\sim} F_{Y,Z}.$$

Gesucht: Schätzung für den Standardfehler des Korrelationskoeffizienten von Y und Z.

5.2.2 Anzahl der Replikationen

Die Anzahl der Replikationen B wird durch folgende Überlegungen bestimmt:

(i) Praktische Überlegungen: Wenn $\hat{\theta}(x^*)$ eine komplizierte Funktion von x^* ist, dann wird B kleiner sein müssen als wenn $\hat{\theta}(x^*)$ eine einfache Funktion von x^* ist.

(ii) Genauigkeitsüberlegungen: Es gilt

$$\mathrm{Var}(\widehat{\mathrm{se}}_B) > \mathrm{Var}(\underbrace{\mathrm{se}_{\hat{F}_n}(\hat{\theta}^*)}_{\text{ideale Bootstrap-Schätzung}}).$$

Die Frage ist, um wieviel die Varianz von $\widehat{\mathrm{se}}_B$ größer ist.

Aus theoretischen Überlegungen ergibt sich, dass $B = 200$ im Einstichproben-Problem in der Regel ausreichend ist zur Schätzung eines Standardfehlers. Für Konfidenzintervalle werden deutlich mehr Replikationen benötigt ($B \approx 2000$).

5.2.3 Parametrischer Bootstrap

Definition 5.1. *Die parametrische Bootstrap-Schätzung des Standardfehlers ist definiert durch*

$$se_{\hat{F}_{n,par}}(\hat{\theta}^*),$$

wobei $\hat{F}_{n,par}$ eine Schätzung von F, abgeleitet aus einem parametrischen Modell, ist.

Beispiel 5.5. *Sei $X = ((Y_1, Z_1)', ..., (Y_n, Z_n)')$ mit*

$$\begin{pmatrix} Y_i \\ Z_i \end{pmatrix} \overset{i.i.d.}{\sim} F_{Y,Z}.$$

Annahme: $F_{Y,Z}$ sei eine bivariate Normalverteilung und

$$\begin{aligned} \hat{\mu} &= \begin{pmatrix} \bar{y} \\ \bar{z} \end{pmatrix}, \\ \hat{\Sigma} &= \frac{1}{n} \left(\begin{array}{c|c} \sum_{i=1}^n (y_i - \bar{y})^2 & \sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z}) \\ \hline \sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z}) & \sum_{i=1}^n (z_i - \bar{z})^2 \end{array} \right). \end{aligned}$$

Das heißt, wir verwenden jetzt $\hat{F}_{n,par} = N_2(\hat{\mu}, \hat{\Sigma})$ als Schätzung für F, und statt Bootstrap-Stichproben aus den Daten zu ziehen, ziehen wir Bootstrap-Stichproben aus dieser bivariaten Normalverteilung:

$$\left.\begin{array}{rcl} x^{*1} & = & ((Y_1^{*1}, Z_1^{*1})', \ldots, (Y_n^{*1}, Z_n^{*1})') \\ & \vdots & \\ x^{*B} & = & ((Y_1^{*B}, Z_1^{*B})', \ldots, (Y_n^{*B}, Z_n^{*B})') \end{array}\right\} \sim N_2(\hat{\mu}, \hat{\Sigma}).$$

Danach geht es weiter wie gewohnt!

Beispiel 5.6 (Standardfehler für die Schätzung des Korrelationskoeffizienten θ)**.**

(i) Vergleich mit der Formel für die bivariate Normalverteilung:

$$\widehat{\text{se}}_{N_2(\mu,\Sigma)}(\hat{\theta}) = \frac{1-\hat{\theta}^2}{\sqrt{n-3}} \; .$$

(ii) Vergleich nach Fisher-Transformation:

$$\hat{\xi} = \frac{1}{2}\log\left(\frac{1+\hat{\theta}}{1-\hat{\theta}}\right) \overset{\text{approx.}}{\sim} N\left[\frac{1}{2}\log\left(\frac{1+\theta}{1-\theta}\right), \left(\frac{1}{\sqrt{n-3}}\right)^2\right] \; .$$

Um dieses Resultat auszunutzen, könnte Inferenz für $\hat{\xi}$ betrieben und anschließend durch Rücktransformation auf den wahren Korrelationskoeffizienten θ übertragen werden.

5.2.4 Ein Beispiel, bei dem der nichtparametrische Bootstrap nicht klappt

Betrachte $X = (X_1, \ldots, X_n)$ mit $X_i \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta)$. Bekannt sei das Maximum $\hat{\theta}_{\text{ML}} = X_{(n)}$. Die Wahrscheinlichkeit, dass $X_{(n)}$ *nicht* in der Bootstrap-Stichprobe auftritt, ist $\left(1 - \frac{1}{n}\right)^n$. Die Wahrscheinlichkeit, dass $X_{(n)}$ in der Bootstrap-Stichprobe vorkommt, ist also

$$1 - \left(1 - \frac{1}{n}\right)^n \to 1 - e^{-1} \approx 0.632 \quad \text{für } n \to \infty \; .$$

Das heißt $P(\hat{\theta}^* = \hat{\theta}_{\text{ML}}) \approx 0.632$ für $n \to \infty$, die Verteilung von $\hat{\theta}^*$ legt also eine Wahrscheinlichkeitsmasse von 0.632 auf den ML-Schätzer. Dieser wird also reproduziert und es gibt damit keinen Informationsgewinn aus diesen Stichproben!

Problem: $\hat{F}_n$ ist keine gute Schätzung für F in den extremen Bereichen von F.

Beim parametrischen Bootstrap gilt dagegen

$$X^* = (X_1^*, \ldots, X_n^*) \text{ mit } X_i^* \sim \text{Unif}(0, \hat{\theta}_{ML})$$

und deshalb

$$\mathbb{P}(\hat{\theta}^* = \hat{\theta}_{ML}) = 0 \; .$$

Also: Nichtparametrischer Bootstrap kann schiefgehen!

5.2.5 Zweistichproben-Problem für unabhängige Stichproben

Seien

$$\left.\begin{array}{l} Y_1,\dots,Y_n \overset{\text{i.i.d.}}{\sim} F \\ Z_1,\dots,Z_m \overset{\text{i.i.d.}}{\sim} G \end{array}\right\} \text{unabhängig, zum Beispiel} \left\{\begin{array}{l} F: \text{ Behandlung} \\ G: \text{ Kontrolle} \end{array}\right.$$

und $X = (Y_1,\dots,Y_n,Z_1,\dots,Z_m)$ bzw. $x = (y_1,\dots,y_n,z_1,\dots,z_m)$.

Ziel: Schätzung des Standardfehlers der Schätzung für die Differenz $\theta = \underbrace{\mu_Y}_{\mathbb{E}(Y_i)} - \underbrace{\mu_Z}_{\mathbb{E}(Z_i)}$.

Betrachte

$$\hat{\theta} = \bar{y} - \bar{z}\ .$$

Vorgehen bei der b-ten Bootstrap-Stichprobe:

$$\begin{array}{rcl} y^{*b} &=& (y_1^{*b},\dots,y_n^{*b}) \text{ zufällig mit Zurücklegen aus } \hat{F}_n \\ z^{*b} &=& (z_1^{*b},\dots,z_m^{*b}) \text{ zufällig mit Zurücklegen aus } \hat{G}_m \end{array}$$

Schätzung:

$$\underbrace{\widehat{\mathrm{se}}_{F,G}(\hat{\theta})}_{\text{Real World}} = \underbrace{\mathrm{se}_{\hat{F}_n,\hat{G}_m}(\hat{\theta}^*)}_{\substack{\text{ideale}\\ \text{Schätzung in}\\ \text{der Bootstrap-}\\ \text{World}}} \approx \underbrace{\widehat{\mathrm{se}}_B}_{\substack{\text{Approx.}\\ \text{der idealen}\\ \text{Bootstrap-}\\ \text{Schätzung}}} = \left\{\frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)\right]^2\right\}^{\frac{1}{2}}$$

mit

$$\hat{\theta}^*(b) = \bar{y}^{*b} - \bar{z}^{*b} = \frac{1}{n}\sum_{i=1}^{n} y_i^{*b} - \frac{1}{m}\sum_{i=1}^{m} z_i^{*b}$$

und

$$\hat{\theta}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B}(\bar{y}^{*b} - \bar{z}^{*b}) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*(b)\ .$$

5.2.6 Bootstrap für eine Zeitreihe

Betrachte die Zeitreihe $y_1, y_2, \dots, y_T$ und die zentrierte Zeitreihe $z_1, z_2, \dots, z_T$ mit $z_t = y_t - \bar{y}$ für $t = 1,\dots,T$.

Annahmen: Es handelt sich um einen AR(1)-Prozess

$$z_t = \beta z_{t-1} + \varepsilon_t \quad (t = 2,\dots,T)$$

mit Anfangsbedingung z_1, $|\beta| < 1$ und $\varepsilon_t \overset{\text{i.i.d.}}{\sim} F$ für $t = 2,\dots,T$, F unbekannt und $\mathbb{E}(\varepsilon_t) = 0$. Die KQ-Schätzung für β lautet:

$$\sum_{t=2}^{T}(z_t - \beta z_{t-1})^2 \to \min_{\beta} \to \hat{\beta}.$$

(Da hier keine Verteilungsannahme getroffen wurde, ist ML-Schätzung nicht möglich.)

Gesucht: Schätzung für $\mathrm{se}_{F,\beta}(\hat{\beta})$.

Idee: Berechne Residuen

$$\left.\begin{array}{rcl} \hat{\varepsilon}_2 & = & z_2 - \hat{\beta} z_1, \\ & \vdots & \\ \hat{\varepsilon}_T & = & z_T - \hat{\beta} z_{T-1}. \end{array}\right\} \; T-1 \text{ Residuen}$$

Bezeichne mit $\hat{F}_{T-1}$ die empirische Verteilungsfunktion der $\hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_T$. Dann erhält man die *b-te Bootstrap-Stichprobe* wie folgt:

(i) Ziehe $\varepsilon_2^{*b}, \ldots, \varepsilon_T^{*b}$ zufällig mit Zurücklegen aus $\hat{F}_{T-1}$.

(ii) Berechne rekursiv

$$\begin{array}{rcl} z_1 & = & y_1 - \bar{y} \\ z_2^{*b} & = & \hat{\beta} z_1 + \varepsilon_2^{*b} \\ z_3^{*b} & = & \hat{\beta} z_2^{*b} + \varepsilon_3^{*b} \\ & \vdots & \\ z_T^{*b} & = & \hat{\beta} z_{T-1}^{*b} + \varepsilon_T^{*b}. \end{array}$$

(iii) Ermittle $\hat{\beta}^{*b}$ mittels KQ aus $z_2^{*b}, \ldots, z_T^{*b}$.

Damit:

$$\widehat{\mathrm{se}}_{F,\beta}(\hat{\beta}) = \mathrm{se}_{\hat{F}_{T-1},\hat{\beta}}(\hat{\beta}^*) \approx \widehat{\mathrm{se}}_B(\hat{\beta}^*) = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} \left[\hat{\beta}^{*b} - \hat{\beta}^*(\cdot) \right]^2 \right\}^{\frac{1}{2}}$$

mit

$$\hat{\beta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}^{*b} \; .$$

Andere Idee: „Moving Block Bootstrap" (vgl. Efron und Tibshirani, 1993).

5.3 Bootstrap-Konfidenzintervalle

5.3.1 Einleitung

Übliches 90%-Konfidenzintervall:

$$\hat{\theta} \pm 1.645 \cdot \widehat{\mathrm{se}}.$$

Übliches 95%-Konfidenzintervall:

$$\hat{\theta} \pm 1.96 \cdot \widehat{\mathrm{se}}.$$

Dabei kann $\widehat{\text{se}}$ auch Bootstrap-Schätzung sein. Die Begründung dafür ist meist:

$$Z = \frac{\hat{\theta} - \theta}{\widehat{\text{se}}} \overset{\text{approx.}}{\sim} N(0,1) \quad \text{(asymptotische Aussage)} \; .$$

Die asymptotische Verteilung ist (approximativ) unabhängig von θ; Z wird approximatives Pivot genannt.

Wenn n klein ist, können die Quantile der Normalverteilung durch die Quantile der t-Verteilung ersetzt werden:

$$\hat{\theta} \pm t_{n-1}^{(1-\alpha/2)} \cdot \widehat{\text{se}} \; .$$

Idee: Annahme der Normalverteilung vermeiden, Verteilung von Z aus den Daten schätzen. Dies wird in den folgenden Abschnitten beschrieben.

5.3.2 Bootstrap-t-Intervall

Betrachte

$$Z = \frac{\hat{\theta} - \theta}{\widehat{\text{se}}} \; , \tag{5.1}$$

wobei $\widehat{\text{se}}$ zunächst irgendeine „vernünftige" Schätzung des Standardfehlers von $\hat{\theta}$ darstellt.

Idee: Schätze Verteilung von Z wie folgt:

1. Generiere B Bootstrap-Stichproben $x^{*1}, \ldots, x^{*B}$.

2. Berechne

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

   wobei $\widehat{\text{se}}^*(b)$ eine Schätzung des Standardfehlers von $\hat{\theta}^*(b)$ ist. Ordne die $Z^*(b)$ aufsteigend der Größe nach.

3. Schätze die Quantile $\hat{t}^{(\alpha)}$ und $\hat{t}^{(1-\alpha)}$ (für ein $(1-2\alpha)$-Konfidenzintervall) als

   $$\frac{\#\left\{Z^*(b) \le \hat{t}^{(\alpha)}\right\}}{B} = \alpha \; .$$

   Dabei bezeichnet $\#A$ die Kardinalität einer Menge A.

   Beispiel: Für $B = 1000$ ist $\hat{t}^{(0.05)}$ der 50. Wert der geordneten $Z^*(b)$-Werte, $\hat{t}^{(0.95)}$ ist der 950. Wert der geordneten $Z^*(b)$-Werte.

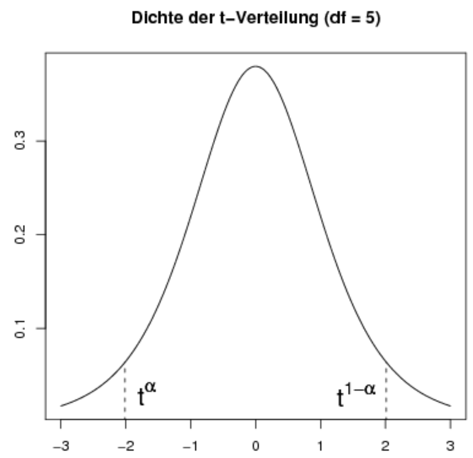
4. Das Bootstrap-t-Intervall zum Vertrauensgrad $1 - 2\alpha$ lautet dann

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

   mit $\widehat{\text{se}}$ aus Formel (5.1).

   Analogie zur t-Verteilung:

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



Beachte: Wenn $B\alpha$ nicht ganzzahlig ist und $\alpha \leq \frac{1}{2}$, dann wähle $k = \lfloor (B+1)\alpha \rfloor$, das ist die größte ganze Zahl $\leq (B+1)\alpha$. Die empirischen Quantile sind dann der k-te Wert der geordneten $Z^*(b)$-Werte und der $(B+1-k)$-te Wert.

Probleme:

1. Das Bootstrap-t-Intervall kann stark durch Ausreißer beeinflusst werden.
2. Betrachte nochmals
$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{\widehat{\text{se}}^*(b)} \ .$$
Wie kann man $\widehat{\text{se}}^*(b)$ schätzen?
   (i) Wenn $\hat{\theta}$ der Mittelwert ist:
$$\widehat{\text{se}}^*(b) = \frac{1}{n} \left\{ \sum_{i=1}^{n} (x_i^{*b} - \bar{x}^{*b})^2 \right\}^{\frac{1}{2}} \quad \text{(Plug-In-Schätzung).}$$
   (ii) Wenn $\hat{\theta}$ komplizierter bzw. keine Standardformel verfügbar ist:
   $\rightarrow$ *Nested Bootstrap:* Es ist eine Bootstrap-Schätzung des Standardfehlers für *jede* Bootstrap-Stichprobe notwendig, zum Beispiel sind für $B = 1000$ und $B^* = 50$
$$BB^* = 1000 \cdot 50 = 50\,000$$
   Stichproben notwendig. Wir samplen also auf zwei verschachtelten Ebenen:

   Real World $\rightarrow$ Bootstrap-World $\rightarrow$ Nested Bootstrap-World.

   Vorteil: Dieser Vorgang ist parallelisierbar (im Gegensatz zu MCMC, wo die Kette nicht parallelisierbar ist).

3. Das Bootstrap-t-Intervall wird von der Skala des Parameters beeinflusst, es ist nicht invariant gegenüber Transformationen. Bei kleinen Stichproben in nichtparametrischem Setup kann irreguläres Verhalten auftreten; hier kann jedoch eine Transformation der Parameter zuverlässigere Ergebnisse liefern.

Beispiel 5.7 (Transformation des Korrelationskoeffizienten)**.** *Sei θ der Korrelationskoeffizient. Ein Konfidenzintervall für θ können wir auf die folgenden zwei Weisen erhalten:*

(i) Bootstrap-t-Intervall für θ direkt.

(ii) Bootstrap-t-Intervall für

$$\phi = \frac{1}{2}\log\left(\frac{1+\theta}{1-\theta}\right) \qquad \text{(Fishersche } Z\text{-Transformation)}$$

und dann Rücktransformation der Endpunkte mittels der Umkehrung

$$\theta = \frac{e^{2\phi}-1}{e^{2\phi}+1}$$

liefert ein kürzeres (= besseres) Konfidenzintervall als das Intervall in (i).

Ergebnis:
1. Bootstrap-t nur für einfache Probleme verwenden, wenn θ ein Lokalisationsparameter, zum Beispiel Median, trimmed mean oder Quantil ist.
2. In komplexen Fällen ist eine Varianzstabilisierung notwendig.

5.3.3 Bootstrap-Perzentil-Intervall

Idee: Verwende direkt die empirische Verteilung der Schätzer $\hat{\theta}^*$ aus den B Bootstrap-Stichproben.

Also:

1. Ziehe

$$\begin{array}{cccl} x^{*1}, & \ldots, & x^{*B} & B \text{ Bootstrap-Replikationen} \\ \downarrow & & \downarrow & \\ \hat{\theta}^*(1), & \ldots, & \hat{\theta}^*(B) & \text{mit } \hat{\theta}^*(b) = T(x^{*b}). \end{array}$$

2. Ordne die $\hat{\theta}^*(b)$ der Größe nach: $\hat{\theta}^*_{(1)}, \ldots, \hat{\theta}^*_{(B)}$.

3. Berechne $B\alpha$ und $B(1-\alpha)$ (bzw. bei nicht-ganzzahliger Anzahl eine Modifikation wie in Abschnitt 5.3.2) und bezeichne mit $\hat{\theta}_B^{*(\alpha)}$ bzw. $\hat{\theta}_B^{*(1-\alpha)}$ die Werte an den jeweiligen Positionen in der sortierten Sequenz der Bootstrap-Schätzungen. Dann ist

$$\left[\hat{\theta}_{\text{lower}}, \hat{\theta}_{\text{upper}}\right] = \left[\hat{\theta}_B^{*(\alpha)}, \hat{\theta}_B^{*(1-\alpha)}\right]$$

ein approximatives $(1-2\alpha)$-Konfidenzintervall.

Beispiel: Für $B = 2000$ und $\alpha = 0.05$ wähle den 100. und 1900. Wert aus der geordneten Liste.

Alternative Schreibweise: Bezeichne mit $\hat{G}_B$ die empirische Verteilung der $\hat{\theta}^*$. Dann ist

$$\left[\hat{\theta}_{\text{lower}}, \hat{\theta}_{\text{upper}}\right] = \left[\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)\right] \ .$$

Vorteile der Perzentil-Methode:

(i) Sie ist invariant gegenüber (streng monotonen) Transformationen.

(ii) Sie ist *range-preserving*, d.h. das Perzentil-Intervall liegt im zulässigen Bereich des Parameters.

Beispiel: Für den Korrelationskoeffizienten liegt das Intervall der Perzentil-Methode im Bereich $[-1, 1]$.

Problem: In der Regel Unterdeckung, d.h. die Intervalle sind häufig zu optimistisch.

Lemma 5.2 (Perzentil-Intervall-Lemma)**.** *Seien $\phi = m(\theta)$ und $\hat{\phi} = m(\hat{\theta})$ eineindeutige Transformationen. Angenommen, $\hat{\phi} = m(\hat{\theta})$ normalisiere die Verteilung von $\hat{\theta}$ perfekt, d.h.*

$$\hat{\phi} \overset{\substack{\textit{exakt,} \\ \textit{nicht nur} \\ \textit{approx.}}}{\sim} N(\phi, c^2)$$

für eine Standardabweichung c.

Dann ist das Perzentil-Intervall basierend auf $\hat{\theta}$ gleich

$$\left[m^{-1}(\hat{\phi} - z^{(1-\alpha)} \cdot c), m^{-1}(\hat{\phi} - z^{(\alpha)} \cdot c)\right]$$

mit den Quantilen $z^{(\alpha)}$, $z^{(1-\alpha)}$ der Standardnormalverteilung.

Das Lemma besagt, dass die Perzentil-Methode immer die korrekte Transformation wählt.

Diskussion:

- Die Perzentil-Methode ist sehr einfach.
- Die Perzentil-Methode ist nicht der Weisheit letzter Schluss. Wenn $\hat{\theta}$ ein Schätzer mit Bias ist, gibt es Alternativen.