Supplement Bootstrap-Konfidenzintervalle

M. Kohl, F. Münch

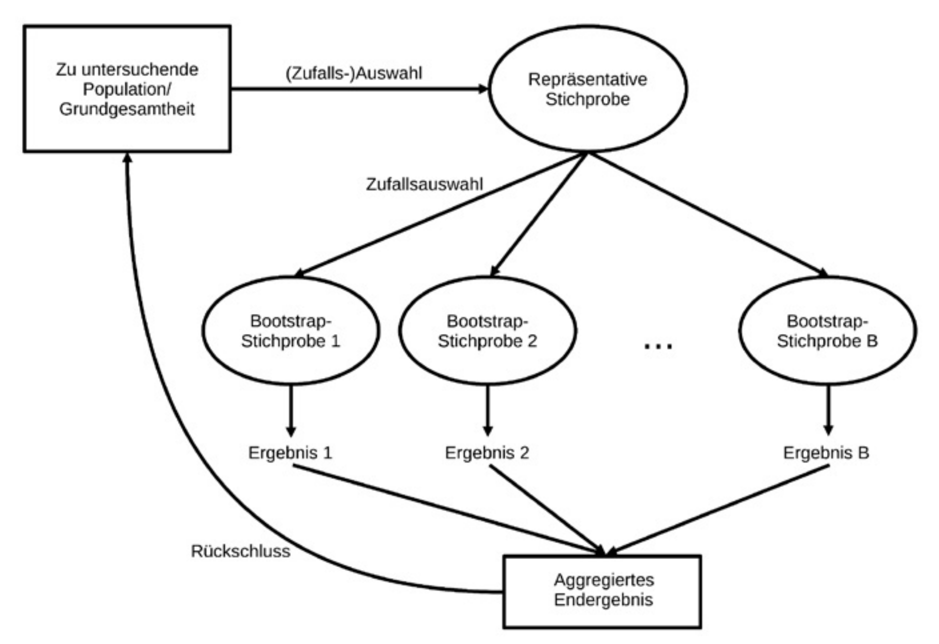

Abb. 1: Schematische Darstellung der Vorgehensweise beim Bootstrapping

Einführung

Der Begriff Bootstrap wurde 1979 von Bradley Efron eingeführt [1]. Abbildung 1 zeigt schematisch die Vorgehensweise beim Bootstrapping.

Wie bei jeder statistischen Analyse sollte der Ausgangspunkt beim Bootstrapping eine repräsentative Stichprobe der Größe n aus einer zu untersuchenden (gut charakterisierten) Population sein. Aus dieser repräsentativen Stichprobe wird dann durch Zufallsauswahl eine große Zahl B von Bootstrap-Stichproben gezogen. Das Ziehen besteht hier in der Regel aus einem Ziehen mit Zurücklegen. In diesem Fall kann insbesondere die Größe der Bootstrap-Stichprobe m identisch mit der Größe der Originalstichprobe n gewählt werden (m = n). Da dieser Ansatz nicht immer funktioniert, wurden in den letzten Jahrzehnten verschiedene Varianten des Bootstraps vorgeschlagen [2].

Typische Werte für die Anzahl B von Bootstrap-Stichproben sind 1000 oder 10.000 in Abhängigkeit von der Rechenzeit, wobei aus technischen Gründen oft ungerade Zahlen für B verwendet werden, also B = 999 oder B = 9999 [3]. Im Fall von Konfidenzintervallen empfehlen sich Werte von B = 999 oder größer. Falls die Rechenzeit gering ist oder keine Rolle spielt, sollte B = 9999 in den meisten Fällen sehr gute und stabile Werte liefern. Ob dies tatsächlich der Fall ist, kann man untersuchen, indem man das Bootstrap-Verfahren mehrfach wiederholt und prüft, ob es relevante Unterschiede zwischen den Ergebnissen gibt. Ist dies der Fall, sollte man B erhöhen oder prüfen, ob es eine alternative Variante des Bootstraps gibt, die besser zum vorliegenden statistischen Problem passt.

Für jede der Bootstrap-Stichproben wird die zu untersuchende Statistik berechnet. Im Fall von Bootstrap-Konfidenzintervallen ist dies der Schätzer für den gesuchten Parameter. Wie wir im nächsten Abschnitt sehen werden, wird bei manchen Varianten der Bootstrap-Konfidenzintervalle auch die Standardabweichung des Schätzers benötigt. Die so entstehenden B Ergebnisse müssen anschließend noch geeignet aggregiert werden, um das Endergebnis zu erhalten. Dieses Endergebnis wiederum ist die Basis für Rückschlüsse auf die zu untersuchende Population.

Vorteile

Es handelt sich um ein sehr einfaches Verfahren, welches auch für komplexe Schätzverfahren und Modelle angewendet werden kann. Außerdem werden hierfür weniger theoretische Annahmen als bei vielen anderen statistischen Verfahren benötigt.

Nachteile

Die Bootstrap-Ergebnisse hängen von der Stichprobe ab, auf deren Basis sie berechnet werden, und es ist nicht bekannt, wie gut die Ergebnisse in einem konkreten Beispiel wirklich sind. Auch kann die Rechenzeit bei komplexen Schätzverfahren recht hoch sein.

DiCiccio und Efron schreiben am Ende des Rejoinders zu ihrer Arbeit über Bootstrap-Konfidenzintervalle [4]:

„Are bootstrap confidence intervals ready for the prime time? If the question is one of always giving highly accurate coverage probabilities in small samples, the answer is no. But this would mean letting the perfect be the enemy of the possible. A more relevant question is whether we can reliably improve upon the standard intervals, and there the answer is yes.“

Im folgenden Abschnitt stellen wir verschiedene Methoden zur Berechnung von Bootstrap-Konfidenzintervallen vor.

Methoden

In den letzten Jahrzehnten wurden verschiedene Methoden entwickelt, mit denen Bootstrap-Konfidenzintervalle berechnet werden können. Wir beschreiben im Folgenden verschiedene Möglichkeiten, die unter anderem auch im Paket boot [5] der Statistiksoftware R [6] enthalten sind. Wir verzichten bei unserer Beschreibung auf Formeln und verweisen für die technischen Details auf die Arbeit von DiCiccio und Efron [4], Kapitel 3 des Buches von Chernick und LaBuddle [2] sowie Kapitel 5 des Buches von Davison und Hinkley [7].

Quantil/Perzentil-Methode

Die einfachste Berechnungsmethode ist das Quantil-/Perzentil-Intervall, welches auf den empirischen Quantilen/Perzentilen der B Bootstrap-Ergebnisse basiert. In diesem Fall entsprechen die untere und obere Grenze des (1-α)-Bootstrap-Konfidenzintervalls gerade dem α/2- und (1-α/2)-Quantil der B Bootstrap-Ergebnisse. Diese intuitive und einfachste Methode zur Berechnung von Bootstrap-Konfidenzintervallen wurde im Artikel verwendet. Man muss davon ausgehen, dass die tatsächliche Überdeckungswahrscheinlichkeit der Perzentil-Intervalle für kleine

Stichprobengrößen n kleiner als 1-α ist, weshalb diese Methode eher bei größeren Stichprobenumfängen n zum Einsatz kommen sollte.

BC_A-Methode

Die BC_a-Methode ist eine Variante der Perzentil-Methode, bei der zusätzlich eine Biaskorrektur (bias-correction = BC) und eine Konvergenzbeschleunigung (acceleration = a) verwendet wird. Hierbei bedeutet Konvergenzbeschleunigung, dass das Bc_a-Konfidenzintervall für wachsenden Stichprobenumfang n schneller gegen das korrekte Konfidenzintervall konvergiert. So nimmt der Fehler, wie bei vielen statistischen Verfahren üblich, nicht mit der Rate $1/\sqrt{n}$, sondern mit der Rate 1/n ab. Entsprechend ist zu erwarten, dass das BC_a-Intervall auch für kleinere Stichproben näher am exakten Intervall liegt und insbesondere die geforderte Überdeckungswahrscheinlichkeit 1-α besser einhält. Simulationsstudien zeigen, dass diese Methode in vielen Situationen sehr gute Ergebnisse liefert.

ABC-Methode

Die ABC-Methode (ABC = Approximate Bootstrap Confidence intervals) ist sehr ähnlich zur BC_a-Methode und liefert aufgrund der gleichen Fehlerrate von 1/n auch annähernd gleich gute Ergebnisse. Die Berechnung der ABC-Intervalle ist etwas einfacher und schneller als für die BC_a-Intervalle. Daher stellt diese Methode bei komplexen Schätzverfahren eine interessante Alternative zur BC_a-Methode dar.

Basic-Methode

Die Basic-Methode, welche auch als reverse Perzentil-Methode bezeichnet wird, ist ähnlich der Perzentil-Methode. Hier wird neben dem α/2- und (1-α/2)-Quantil der B Bootstrap-Schätzungen zusätzlich die Schätzung für die Originalstichprobe verwendet. Die Basic-Methode liefert in vielen Fällen unzureichende Überdeckungswahrscheinlichkeiten und sollte daher in der Praxis eher nicht verwendet werden.

Normalapproximation

Diese Methode basiert auf der Normalapproximation für die gesuchte Statistik, welche auf eine Berechnungsformel führt, die analog zu Formel (4) in Abbildung 3 des Artikels ist. Die Standardabweichung des Schätzers wird in diesem Fall jedoch nicht aus der ursprünglichen Stichprobe, sondern aus den B Bootstrap-Stichproben berechnet. Ist keine explizite Formel für die Berechnung der Standardabweichung des Schätzers bekannt, so besteht ein Ausweg in einem doppelten Bootstrap. Hier werden aus jeder Bootstrap-Stichprobe neue Bootstrap-Stichproben gezogen und diese dazu verwendet, die Standardabweichung des Schätzers zu bestimmen. Alternativ zu den Quantilen der Normalverteilung können natürlich auch die entsprechenden Quantile der t-Verteilung mit n–1 Freiheitsgraden zur Berechnung herangezogen werden (analog zu Formel (3) in Abbildung 3 des Artikels), insbesondere für kleinere Stichprobenumfänge. Simulationsstudien zeigen jedoch, dass andere hier vorgestellte Methoden meistens bessere Ergebnisse erzielen.

t-Methode

Im Fall der t-Methode werden die Bootstrap-Schätzungen an der Schätzung für die Originalstichprobe zentriert und das Ergebnis durch die jeweilige Bootstrap-Schätzung der Standardabweichung des Schätzers dividiert. Man nennt dies auch standardisieren oder studentisieren. Die Quantile der studentisierten Werte werden als Ersatz für die Quantile einer entsprechenden t-Verteilung verwendet (analog zu Formel (3) in Abbildung 3 des Artikels). Wie im Fall der Normalapproximation benötigt man demnach zusätzlich eine Schätzung der Standardabweichung des Schätzers. Folglich ist auch in diesem Fall ein doppeltes Bootstrap zur Berechnung der Standardabweichung des Schätzers eine mögliche Option. Bei der t-Methode wird wie bei der BC_a- und der ABC-Methode eine Fehlerrate von 1/n erreicht. Simulationsstudien zeigen, dass sich mit der t-Methode in der Regel gute Ergebnisse erzielen lassen, insbesondere auch für kleinere Stichprobenumfänge und schiefe Verteilungen.

Zusammenfassung

Für praktische Anwendungen sind vor allem die Perzentil-, die BC_a- und die t-Methode von Interesse, wobei die ABC-Methode bei komplexen Schätzverfahren eine gute Alternative zur BC_a-Methode darstellt. Die Perzentil-Methode ist generell für größere Stichprobenumfänge eine gute Wahl. Die BC_a- und die t-Methode erzielen hingegen auch bei kleineren Fallzahlen gute Ergebnisse. Bei kleinen Stichprobenumfängen und schiefen Verteilungen sollte in erster Linie die t-Methode verwendet werden.

LITERATUR

1. Efron B. Bootstrap Methods: Another Look at the Jackknife. Ann.Statist. 1979;7(1): 1–26.

2. Chernick MR, LaBudde RA. An Introduction to Bootstrap Methods with Applications to R. Wiley & Sons Ltd. 2011.

3. Hall P. On the Number of Bootstrap Simulations Required to Construct a Confidence Interval. Ann.Statist. 1986;14(4): 1453–1462.

4. DiCiccio TJ, Efron B. Bootstrap confidence intervals. Statist.Sci. 1996;11(3): 189–228.

5. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version 1. 2021; 3–28.

6. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. URL https://www.R-project.org/.

7. Davison AC, Hinkley DV. Bootstrap Methods and Their Applications. Cambridge University Press; 1997.