

Bootstrapping

Fabian Koßmann

14. August 2026

1 Was ist Bootstrapping?

Bootstrapping ist eine Methode zur Approximation der Verteilung eines Parameters auf Basis von Stichprobendaten. Das Verfahren wurde 1977 von Bradley Efron in dem Paper [Efr79] vorgestellt und stellt eine Weiterentwicklung der Jackknife-Methode dar, bietet jedoch ein breiteres Anwendungsspektrum.

Dieses Handout sowie die Präsentation basieren auf den Inhalten des genannten Papers.

Ein YouTube Video mit einer sehr einfachen Erklärung ist unter [exp] zu finden.

2 Anwendungsgebiete

Bootstrapping wird in verschiedenen Bereichen der Statistik und des Machine-Learnings eingesetzt. Es dient zur Schätzung von Konfidenzintervallen und zur Durchführung von Hypothesentests, insbesondere dann, wenn keine starken Annahmen über die zugrunde liegende Verteilung der Daten getroffen werden können. Darüber hinaus wird Bootstrapping zur Schätzung von Standardfehlern und zur Bestimmung des Bias in statistischen Schätzern verwendet. Im Bereich des Machine-Learnings findet es Anwendung bei Methoden wie Cross-Validation und der Bewertung der Modellperformance.

3 Problemstellung

Gegeben sei eine Stichprobe $X = (X_1, \dots, X_n)$, gezogen aus einer unbekannten Verteilung F , mit der entsprechenden Realisierung $x = (x_1, \dots, x_n)$. Zudem sei eine Zufallsvariable $R(X, F)$ definiert, die sowohl von der Stichprobe X als auch von der Verteilung F abhängt. Das Ziel ist, die Stichprobenverteilung von R basierend auf den vorliegenden Daten x zu approximieren. Die Stichprobenverteilung ist die Verteilung der Zufallsvariablen R , die durch wiederholtes Ziehen von Stichproben der Größe n aus der zugrunde liegenden Verteilung F entsteht.

4 Bootstrap-Methode

1. Konstruktion einer Verteilung $\hat{F}$ basierend auf der Stichprobe, wobei jedem Datenpunkt $x_1, \dots, x_n$ die gleiche Wahrscheinlichkeitsmasse $\frac{1}{n}$ zugewiesen wird.
2. Ziehen einer Stichprobe $X^* = (X_1^*, \dots, X_n^*)$ aus der Verteilung $\hat{F}$ mit der Realisierung $x^* = (x_1^*, \dots, x_n^*)$. Diese Stichprobe wird als Bootstrap-Stichprobe bezeichnet. Wichtig ist, dass es sich hierbei nicht um eine Permutation der Originalstichprobe handelt, da Werte mehrfach auftreten können. Dies liegt daran, dass das Ziehen mit Zurücklegen erfolgt.
3. Approximation der Stichprobenverteilung der Zufallsvariable $R(X, F)$ durch die Stichprobenverteilung der Zufallsvariable $R(X^*, \hat{F})$. Diese Verteilung wird als Bootstrap-Verteilung bezeichnet.

5 Beispiel

Gegeben sei die Stichprobe $X = (X_1, \dots, X_3)$ mit der Realisierung $(2, 4, 7)$, die aus einer unbekannten Verteilung F stammt. Die Zufallsvariable $R(X, F)$ repräsentiert den Mittelwert der Stichprobe.

1. Konstruktion der Verteilung $\hat{F}$ mit der Zähldichte $\hat{p}$, wobei gilt:

$$\hat{p}(2) = \frac{1}{3}, \quad \hat{p}(4) = \frac{1}{3}, \quad \hat{p}(7) = \frac{1}{3}.$$

2. Bootstrap-Stichproben und zugehörige Mittelwerte: Durch Ziehen mit Zurücklegen aus der Verteilung $\hat{F}$ entstehen folgende Bootstrap-Stichproben und deren Mittelwerte:

Bootstrap-Stichprobe	Zug 1	Zug 2	Zug 3	Mittelwert $R(X^*, \hat{F})$
1	2	2	2	2.00
2	2	2	4	2.67
3	2	2	7	3.67
4	2	4	2	2.67
5	2	4	4	3.33
6	2	4	7	4.33
7	2	7	2	3.67
8	2	7	4	4.33
9	2	7	7	5.33
10	4	2	2	2.67
11	4	2	4	3.33
12	4	2	7	4.33
13	4	4	2	3.33
14	4	4	4	4.00
15	4	4	7	5.00
16	4	7	2	4.33
17	4	7	4	5.00
18	4	7	7	6.00
19	7	2	2	3.67
20	7	2	4	4.33
21	7	2	7	5.33
22	7	4	2	4.33
23	7	4	4	5.00
24	7	4	7	6.00
25	7	7	2	5.33
26	7	7	4	6.00
27	7	7	7	7.00

Tabelle 1: Bootstrap-Stichproben und zugehörige Mittelwerte

3. Bootstrap-Verteilung der Mittelwerte: Die resultierenden Mittelwerte $R(X^*, \hat{F})$ bilden die Bootstrap-Verteilung des Mittelwerts. Diese Verteilung wird in einem Histogramm dargestellt:

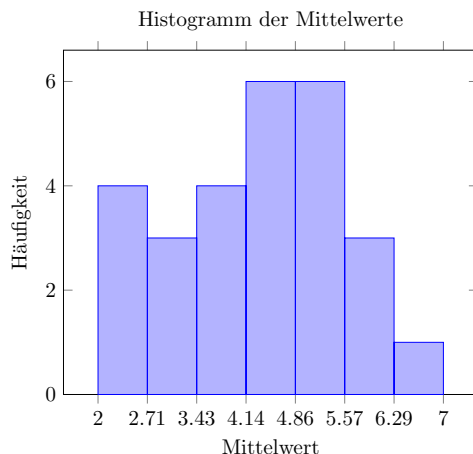


Abbildung 1: Verteilung der Bootstrap-Mittelwerte

6 Monte-Carlo-Methode

Die exakte Berechnung der Bootstrap-Verteilung ist äußerst aufwändig, da es n^n mögliche Stichproben gibt, die aus $\hat{F}$ gezogen werden können. Stattdessen wird häufig eine Monte-Carlo-Simulation verwendet. Dabei werden N Bootstrap-Stichproben $X^{1*}, \dots, X^{N*}$ mit den entsprechenden Realisierungen $x^{1*}, \dots, x^{N*}$ aus der Verteilung $\hat{F}$ gezogen. Diese Methode reduziert den Rechenaufwand erheblich, liefert jedoch eine gute Annäherung an die tatsächliche Bootstrap-Verteilung.

7 Beweisidee

Sei P die wahre Stichprobenverteilung von $R(X, F)$ und P^* die Bootstrap-Stichprobenverteilung von $R(X^*, \hat{F})$.

Der Wert von R kann unter Umständen über eine Funktion f und die zugehörige kumulierte Wahrscheinlichkeitsfunktion bestimmt werden. Dies wird ausgedrückt durch $f(P_{\text{cdf}}(X, F)_n)$ beziehungsweise $f(P_{\text{cdf}}(X^*, \hat{F})_n)$, wobei n die Stichprobengröße ist.

Diese Methode ist für gewisse statistische Größen wie den Mittelwert, die Varianz und die Kovarianz anwendbar.

Die Idee besteht darin zu zeigen, dass die Verteilung $P_{\text{cdf}}(X^*, \hat{F})_n$ die Verteilung $P_{\text{cdf}}(X, F)_n$ annähert bzw. schätzt.

Da die Verteilungen $P_{\text{cdf}}(X^*, \hat{F})_n$ und $P_{\text{cdf}}(X, F)_n$ von der Stichprobengröße n abhängen, werden die Grenzverteilungen (limiting distributions) $P_{L\text{cdf}}(X^*, \hat{F})$ und $P_{L\text{cdf}}(X, F)$ mit den zugehörigen Zähldichten $p_{L\text{cdf}}(X^*, \hat{F})$ und $p_{L\text{cdf}}(X, F)$ betrachtet.

Ziel der Beweisidee ist es zu zeigen, dass $p_{L\text{cdf}}(X^*, \hat{F})$ die Funktion $p_{L\text{cdf}}(X, F)$ annähert.

Dies geschieht durch die uniforme Konvergenz, indem gezeigt wird, dass

$$\sup_x \left| p_{L\text{cdf}}(X^*, \hat{F})(x) - p_{L\text{cdf}}(X, F)(x) \right| \rightarrow 0.$$

Ein gutes YouTube Video zur Beweis Idee ist unter [\[prob\]](#) zu finden und ein vollständig formaler Beweis ist unter [\[proa\]](#) zu finden.

Literatur

[Efr79] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.

[exp] Einfache Erläuterung: <https://www.youtube.com/watch?v=Xz0x-8-cgaQt=404>.

[proa] Formaler Beweis: <https://www.stat.purdue.edu/~dasgupta/bootstrap.pdf>.

[prob] Beweisidee: <https://www.youtube.com/watch?v=BiNcdYbyiWwt=369s>.