

The background features a light-colored world map with a grid, overlaid with a dense field of colorful, semi-transparent bokeh bubbles in shades of blue, orange, red, and grey. The bubbles vary in size and are scattered across the entire frame, creating a vibrant, abstract effect.

# Angewandte Statistik

Julian Hinz — Universität Bielefeld

# Session 9

*Bootstrapping*

# Lernziele

- Wiederholung Stichprobenziehung und
- Einführung in Bootstrapping



**Wiederholung:  
Stichprobenziehung**

# Stichprobenziehung

- Grundgesamtheit häufig schwierig komplett zu beobachten
- daher: Stichproben
- Beispiel: Zensus, Wahlbefragung, ...
- hier: Zufallsstichproben

# Beispiel mit R

- Setting: Europawahl 2024
- Wähler mit Attributen und Wahlpräferenzen
- große Grundgesamtheit, daher sampeln
- Parameter der Grundgesamtheit: z.B. der wahre Anteil von Studierenden unter den Wählern
- Schätzer für Parameter der Grundgesamtheit: Anteil der Studierenden, berechnet aus der zufälligen Stichprobe
  - da Stichprobe zufällig ist, repräsentativ für Grundgesamtheit und Schätzer unverzerrt

```

1 > wähler
2      person_id alter einkommen      beruf verheiratet kinder
3      <int> <int>      <num>      <char>      <char> <num>
4      1:          1    37    48193    Selbständig    Nein    1
5      2:          2    78    13735    Arbeiter/in    Nein    0
6      3:          3    41    20673    Studierende/r    Ja    1
7      4:          4    22    19289    Selbständig    Nein    0
8      5:          5    79    19717    Arbeiter/in    Ja    1
9      ---
10     99996:      99996    27    32280    Arbeiter/in    Ja    0
11     99997:      99997    70    22035    Arbeiter/in    Nein    0
12     99998:      99998    29    21863    Arbeiter/in    Nein    0
13     99999:      99999    24    24526    Angestellte/r    Nein    0
14    100000:     100000    44    19918    Arbeiter/in    Nein    1
15      verkehrsmittel      partei
16      <char>      <char>
17      1:          Fahrrad Konservative Partei
18      2:          Auto      Liberale Partei
19      3:          Auto      Soziale Partei
20      4:          Auto      Konservative Partei
21      5:          Auto      Soziale Partei
22      ---
23     99996:          Auto      Soziale Partei
24     99997:          Auto      Soziale Partei
25     00008:          Fahrrad      Soziale Partei

```



# zufällige Stichprobe

- $n$  Beobachtungen aus der Grundgesamtheit zufällig auswählen
- Beobachtungen sollten nicht mehrfach vorkommen
- Stichprobengröße beeinflusst *Standardfehler*
  - je größer desto repräsentativer!



# Bootstrapping

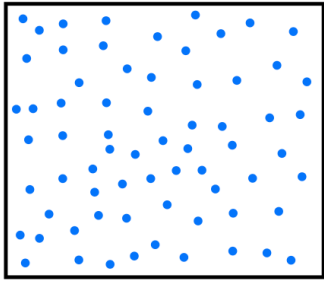
# Einführung in Bootstrapping

- Statistiken aus Stichproben sind *Zufallsvariablen*
  - Stichprobenverteilung mit Mittelwert und Standardfehler
- Beispiel: Anteil der Studierenden in Stichprobe
  - Schätzer für den wahren Anteil der Studierenden bei allen Wählern

# Einführung in Bootstrapping

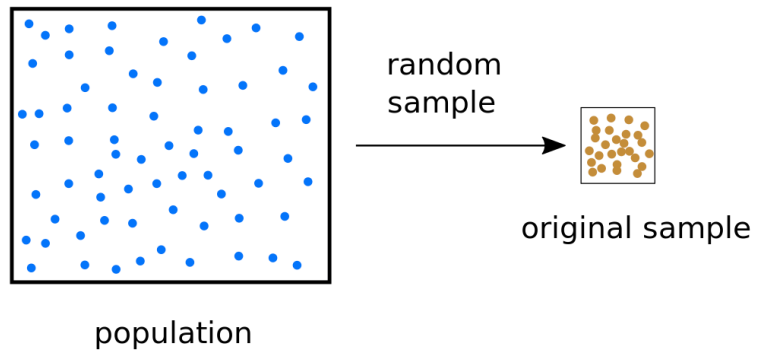
- Problem: In Realität schwierig sehr viele Stichproben zu ziehen
- Lösung: Mit einer einzigen Stichprobe arbeiten!

# Einführung in Bootstrapping

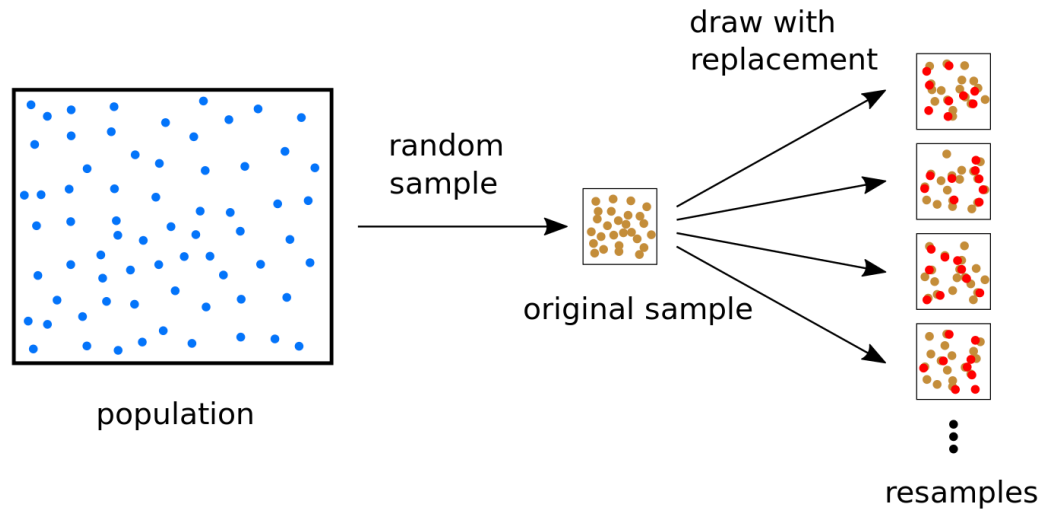


population

# Einführung in Bootstrapping

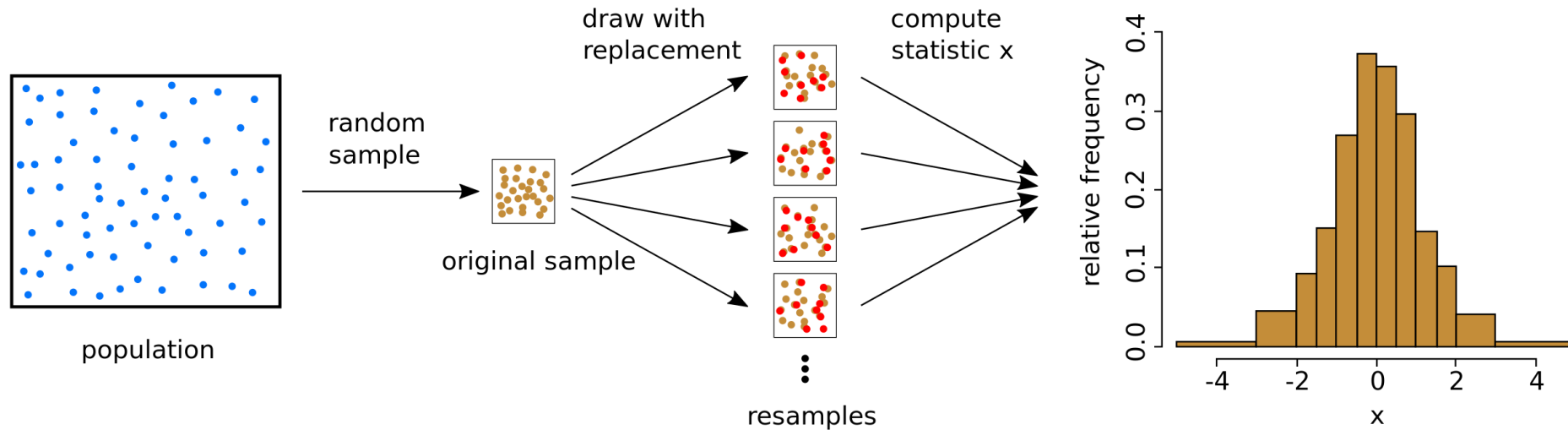


# Einführung in Bootstrapping





# Einführung in Bootstrapping



# Einführung in Bootstrapping

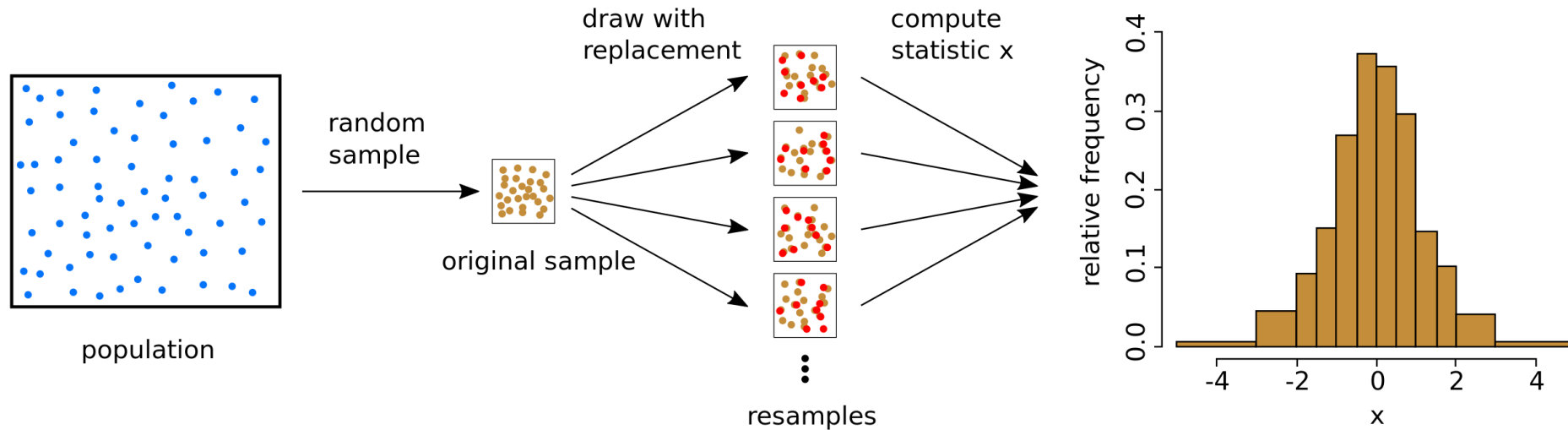


“To pull oneself up by one’s bootstraps” — sich am eigenen Schopf aus dem Sumpf ziehen

# Einführung in Bootstrapping



# Einführung in Bootstrapping



# Ursprung des Bootstrap

- Bradley Efron: “Bootstrap methods: another look at the jackknife” (1979)
- baut auf dem “jackknife” auf
- Bayesianische Erweiterung: Samplen mit Gewichten zwischen 0 und 1
- andere vorgeschlagene Namen für den “bootstrap”: Swiss Army Knife, Meat Axe, Shotgun, ...

# Vorteil gegenüber traditionellen Methoden

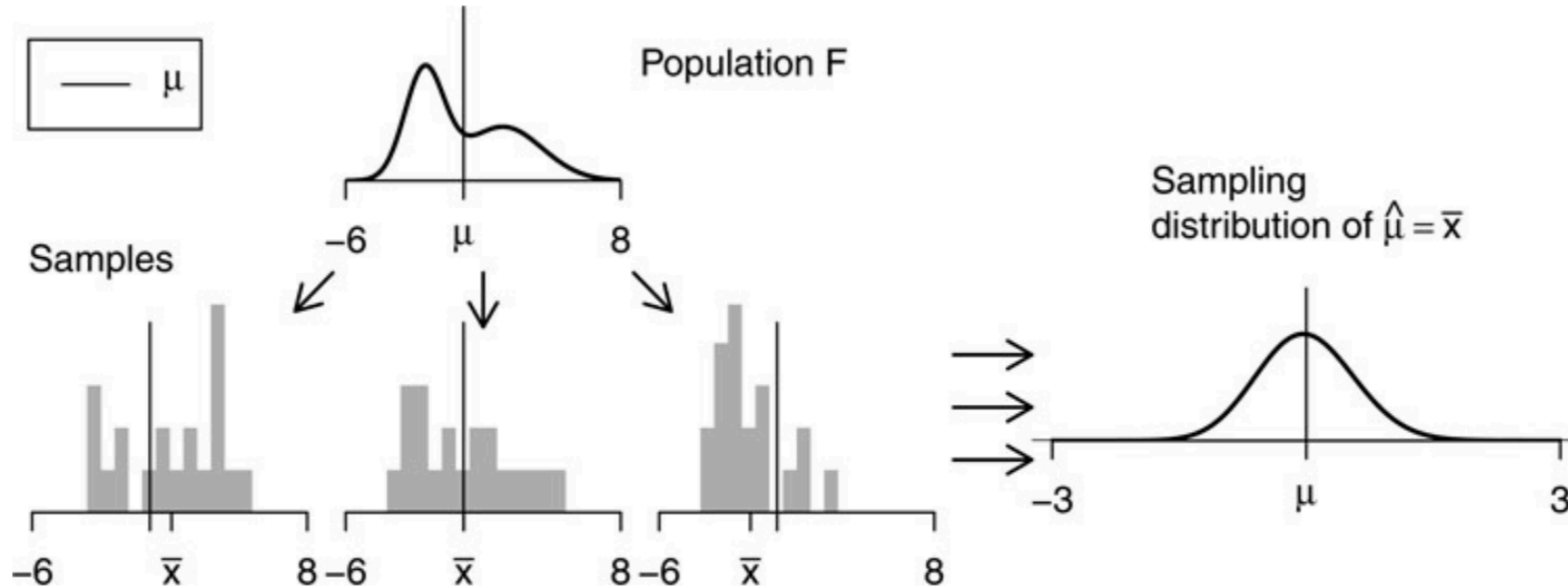
- Häufig Annahmen über Verteilung oder Momente (Mittelwert, Varianz, ...)
- Beispiel: Eine Regression mit normalverteilter Zielvariable
- Annahme bei Bootstrap: empirische Verteilungsfunktion kann tatsächliche Verteilungsfunktion hinreichend gut approximieren
  - Stichprobe nicht zu klein
  - Stichprobe repräsentativ

# Unsicherheit des Schätzers

- Parameter der Grundgesamtheit aus nur einer Stichprobe schätzen
  - z.B. Anteil der Studierenden unter Wählern
- jede Stichprobe zufällig und Schätzer somit eine Zufallsvariable
  - wie gut ist der Schätzer?
- wenn man theoretische Verteilung kennt, mittels Standardfehler Konfidenzintervall berechnen

$$\hat{\mu} \pm 1.96 \cdot SE$$

# Normalerweise

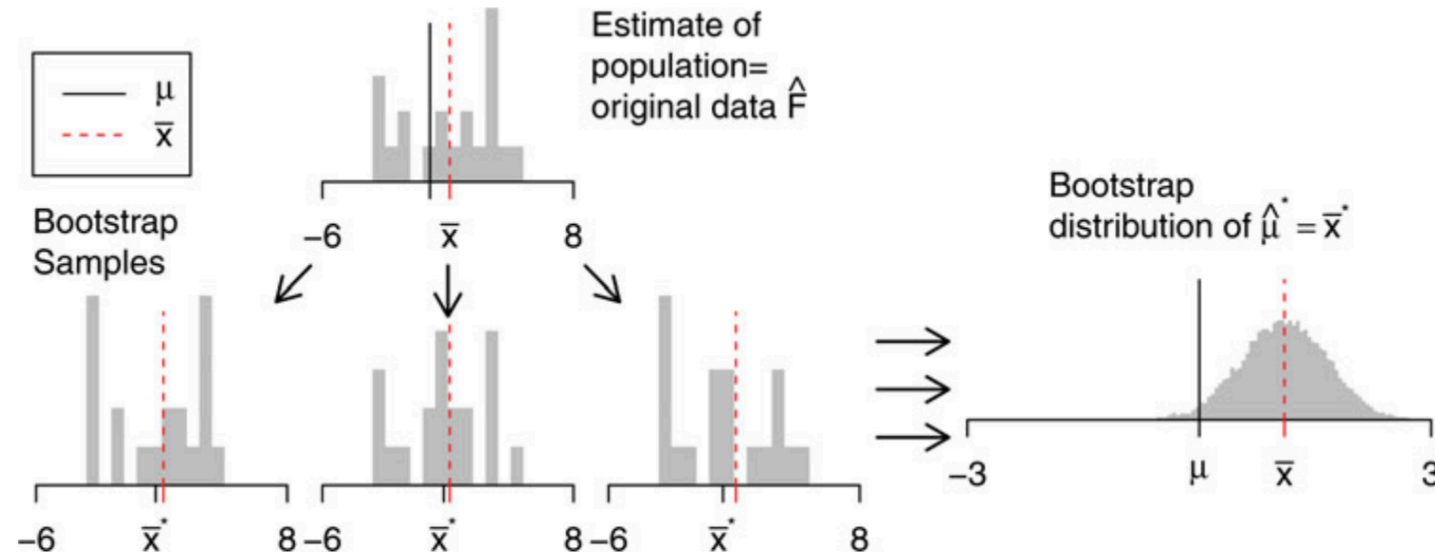


Quelle: Hesterberg (2015)

- mehrere echte Stichproben aus einer Grundgesamtheit
- Stichprobenverteilung und Schätzer für Parameter der Grundgesamtheit
- 95% Konfidenzintervalle aus  $\hat{\mu} \pm 1.96 \cdot SE$



# Bootstrap



Quelle: Hesterberg (2015)

- Idee: Stichprobe eine (repräsentative) Miniatur der Grundgesamtheit
- Bootstrap-Stichproben durch ziehen von neuen Stichprobe *mit Zurücklegen*
- 95% Konfidenzintervall aus 2.5% und das 97.5% Quantil der Bootstrap-Stichproben







USE  
YOUR VOTE.

EUROPEAN ELECTIONS  
**6-9 JUNE 2024**

