Grundlagen der Resampling Methoden

- Angelehnt an eine Vorlesung von Rozenn Dahyot, Trinity College, Dublin.
- Literatur:
  - An Introduction to the Bootstrap, B. Efron und R.J. Tibshirani
  - Computer Intensive Statistical Methods, J.S. Urban Hjorth

Überblick

- Die Resampling Methoden zerfallen grob in drei oder vier Klassen.
- Der Bootstrap, auch Münchhausenmethode, sowohl in der parametrischen, als auch der nicht-parametrischen Variante.
- Der Jackknife.
- (Permutationstests, werden hier nicht behandelt).
- Die Kreuzvalidierung (*cross-validation) CV.*

Einige Notationen

- Gegeben sei eine Stichprobe $\mathbf{x} = (x_1, \ldots, x_n)$ als Realisierung von n i.i.d. Zufallsvariablen $X_i \sim F$, einer unbekannten Verteilungsfunktion.

- Die empirische Verteilungsfunktion wird mit $\hat{F}$ bezeichnet, $\hat{f}$ ist die empirische Dichte.

- Evtl. neu ist die empirische Dichte:

$$\hat{f}(x) := \frac{1}{n} \sum_{1}^{n} \delta(x - x_i),$$

wobei die δ-Funktion (Kroneckers-δ) definiert ist als:

$$\delta(0) = 1,\ \delta(x) = 0 \text{ sonst.}$$

- Eine Statistik oder auch Schätzfunktion $\hat{\theta}(\mathbf{x})$ ist eine Abbildung von einer Stichprobe in den Parameterraum (z.B. $\mathcal{R}$), z.B. für den Mittelwert

$$\hat{\theta}(\mathbf{x}) = \frac{1}{n} \sum x_i \ (= \hat{\mu}).$$

  Um die Abhängigkeit von der Stichprobe zu betonen schreibt man auch $s(\mathbf{x})$ für $\hat{\theta}$.

- Eine andere Sicht auf die Schätzfunktion fasst einen Parameter einer Dichte auf als eine Abbildung t aus der Menge der Dichten in den Parameterraum: $\theta = t(f)$.

- Z.B. ist für den Mittelwert einer bekannten Dichte f

$$\theta = E_f(X) = t(f) = \int_{-\infty}^{\infty} x f(x) dx.$$

- Daraus ergibt sich eine naheliegende Definition für einen Schätzer eines Parameters θ über das sogenannte *plug-in* Prinzip. Es wird statt der unbekannten Dichte f einfach das empirische Gegenstück $\hat{f}$ eingesetzt:

$$\hat{\theta} = t(\hat{f})!$$

- $\hat{\theta}$ heißt der Plug-in Schätzer für θ.

- Im Beispiel für den Mittelwert

$$\hat{\theta} = E_{\hat{f}}(X) = t(\hat{f}) = \int_{-\infty}^{\infty} x\hat{f}(x)dx = \bar{x}$$

- oder für die Varianz

$$\hat{\theta} = t^*(\hat{f}) = \int_{-\infty}^{\infty} (x-\bar{x})^2\hat{f}(x)dx = \frac{1}{n}\sum_{1}^{n}(x_i-\bar{x})^2 = \hat{\sigma}^2.$$

Genauigkeit von Schätzern

- Mittels der Funktion t (bzw. s) kann man nun bei gegebener Stichprobe $\mathbf{x}$ einen Schätzer $\hat{\theta}$ berechnen.

- Bleibt die Frage, wie man die Eignung der Schätzfunktion als Schätzer für den wahren Parameter beurteilt.

- Aus Statistik II sind Erwartungstreue und Konsistenz als Kriterien bekannt, ebenso wie die Standardfehler, die bei der Konstruktion von Konfidenzintervallen genutzt werden.

- Zur Erinnerung: Der Standardfehler ist die Standardabweichung eines Schätzers $\hat{\theta}$ aufgefasst als Zufallsvariable.

$$\mathsf{se}(\hat{\theta}) := \sqrt{\mathsf{var}_f(\hat{\theta})}.$$

- Ist f unbekannt, so kann man auch hier mittels *plug-in* Prinzip zu einer Schätzung des Standardfehlers gelangen. Allgemein also

$$\hat{\mathsf{se}}(\hat{\theta}) = \mathsf{se}_{\hat{f}}(\hat{\theta}) = \sqrt{\mathsf{var}_{\hat{f}}(\hat{\theta})}$$

und für das Beispiel des Standardfehlers bei $\bar{x}$ gilt

$$\hat{\mathsf{se}}(\bar{x}) = \frac{\hat{\sigma}}{\sqrt{n}}.$$

- Entsprechend können über dieses Prinzip auch die bekannten Konfidenzintervalle und ihre Schätzer hergeleitet werden.

Noch Probleme?

- Ja! Denn oft kann man die Funktionen t nicht (leicht) explizit aufschreiben. Wie sieht z.B. die entsprechende Abbildung für den Median aus?
- In der Praxis gibt es nur eine Stichprobe $\mathbf{x}$, also auch nur eine Realisierung der Schätzfunktion $\hat{\theta}$. Wie kann man also z.B. $\mathrm{se}(\hat{\theta})$ angeben, wenn nur eine Stichprobe existiert?
- An dieser Stelle setzen die Resampling Methoden an.

Der Bootstrap

- Gegeben sei eine beliebige Schätzfunktion $\hat{\theta}$, sowie die einzige Stichprobe $\mathbf{x}$ mittels derer ein Schätzwert als Realisierung von $\hat{\theta}$ berechnet werde.
- Wie kann man trotzdem die Güte dieses Schätzers beurteilen? Wie kommt man an Schätzungen für Bias, Varianz etc dieser Schätzfunktion?
- Eine Methode ist die computerintensive Resampling-Technik des Bootstrapping, die im Folgenden vorgestellt wird.
- Der Name kommt übrigens aus den Abenteuern des Baron Münchhausen!
- Der Bootstrap kommt nicht nur mit einer einzigen Stichprobe aus, er verzichtet auch auf jede Modellannahme!

- Die Struktur des Schätzers $\hat{\theta}$ darf beliebig kompliziert sein, da die theoretischen Eigenschaften von θ uns nicht interessieren. Es sind nur die Realisierungen, die aus der Stichprobe gewonnen werden interessant!

- Das hier vorgestellte Prinzip wurde 1979 von Efron entwickelt, um die Standardfehler beliebiger Schätzer zu bestimmen. In den Anwendungen wurde das Vorgehen verfeinert, das Prinzip ist aber unverändert bestehen geblieben.

Die Bootstrap-Stichprobe

- Definition: Eine Bootstrap-Stichprobe $x^* = (x_1^*, \ldots, x_n^*)$ erhält man, indem aus $\mathbf{x} = (x_1, \ldots, x_n)$ n-mal mit Zurücklegen gezogen wird.
- Beispiel: Sei $\mathbf{x} = (1, 2, 3)$, dann sind z.B. (1,1,3) oder (2, 1, 3) mögliche Bootstrap-Stichproben.
- Nun wird endlich klar, warum die Methoden Resampling genannt werden!
- Zu jeder Bootstrap-Stichprobe x^* kann mittels des *plug-in*-Prinzips eine Bootstrap-Schätzung $\hat{\theta}^* = s(\mathbf{x}^*)$ durchgeführt werden. Über viele sogenannte Replikationen kann dann deren Standardfehler empirisch bestimmt werden!
- In `R` wird mit dem Befehl `sample(x, length(x), replace=TRUE)` eine Bootstrap-Stichprobe aus dem Vektor `x` erzeugt.

Welchen Anteil der Beobachtungen verliert man?

- Zieht man mit Zurücklegen aus einem endlichen Vektor, werden manche Einträge doppelt, manche gar nicht im Bootstrap-Sample auftreten.
- Die Wahrscheinlichkeit für eine spezielle Beobachtung x_i in einer Bootstrap-Stichprobe zu fehlen, lässt sich berechen.
- Angenommen alle x_i sind verschieden, dann ist die Wkeit, dass ein Wert x' nicht gezogen wird

$$P(x_i \neq x',\ 1 \leq i \leq n) = (1 - \frac{1}{n})^n.$$

Für $n \to \infty$ konvergiert diese Wkeit gegen $\frac{1}{e} = 0.37$.

Efrons Bootstrap

- Nach diesen Vorarbeiten kann nur für eine beliebige Schätzfunktion $s(\mathbf{x})$ mittels des Bootstrap ein Schätzer für ihren Standardfehler $\widehat{\text{se}}(s(\hat{\mathbf{x}}))$angegeben werden.

- Gegeben seien N Bootstrap-Stichproben $x_1^*, \ldots, x_N^*$ aus $\mathbf{x}$.

- Berechne nun zu jeder Stichprobe $1 \leq k \leq N$

$$\hat{\theta}_k^* = s(x_k^*)$$

- Damit ergibt sich der Bootstrap-Schätzer $\widehat{\text{se}}_B$ für den Standardfehler se($\hat{\theta}$) durch die N Wiederholungen zu

$$\hat{\text{se}}_B = \left[\sum_1^N \left[\hat{\theta}^* - \bar{\theta^*} \right]^2 \right]^{\frac{1}{2}}$$

mit $\bar{\theta^*} = \frac{1}{N} \sum_1^N \hat{\theta}^*$.

- Bei genügend Replikationen lassen sich die vollständigen Verteilungen beliebiger Teststatistiken approximieren!

Liveaufgabe zum Bootstrap

- Erzeugen Sie sich eine Stichprobe vom Umfang 100 aus der N(1,2) Verteilung.
- Bestimmen Sie den Standardfehler für den Mittelwertschätzer mittels Bootstrapping. Nutzen Sie 100, 200, ..., 1000 Replikationen!
- Vergleichen Sie das Ergebnis mit dem theoretischen Wert. Welche Zahl von Replikationen scheint angemessen?
- Schätzen Sie die Verteilung des Schätzers mittels eines Histogramms!

Lösung

```
x <- rnorm(100, mean=1, sd=sqrt(2))

for (runs in seq(100, 1000, 100)){
  bootschaetzer <- replicate(runs,
                    mean(sample(x,length(x), replace=TRUE)))
  cat ("Mittel der Bootstrapschätzer ",mean(bootschaetzer),"\n")
  cat ("Standardfehler für ", runs, " Replikationen: ",
            sd(bootschaetzer), "\n")
}


sqrt(2)/10 ### theoretisch
hist(bootschaetzer)
```

Aufgabe zum Bootstrap

- Die vorhergehende Aufgabe ließe sich auch noch theoretisch lösen. Diese nur sehr schwer ...

- Kombinieren Sie eine Stichprobe vom Umfang 35 aus der Gleichverteilung auf $[0, 6]$ und eine Stichprobe vom Umfang 65 aus $N(-1, 1)$ zu einer Stichprobe vom Umfang 100.

- Schätzen Sie den Median der zugrundeliegenden Verteilung und leiten Sie mittels Bootstrap eine Schätzung des Standardfehlers für diesen Medianschätzer her. (N=200 Replikationen)

- Ensprechend für den Interquartilsabstand. (N=200 Replikationen)

Lösung:

```
replications <- 200
x<- c(6*runif(35), rnorm(65)-1)
median(x)
[1] -0.2982217
sd(replicate(replications,
   median(sample(x,length(x), replace=TRUE))))
[1] 0.2690578
interquartile100 <- function(x){ x<- sort(x); x[75]-x[25]}
interquartile100(x)
[1] 2.26707
sd(replicate(replications,
   interquartile100(sample(x,length(x), replace=TRUE))))
[1] 0.360023
```

Bootstrap-Konfidenzintervalle

- Auch ein Bootstrap-Konfidenzintervall lässt sich mittels der gewonnenen Schätzer definieren.
- Das $(1-\alpha)$% Bootstrap-KI für einen Parameter θ ist gegeben durch

$$\mathsf{KI}_{1-\alpha,B}(\theta) := \hat{\theta} \pm z_{1-\frac{\alpha}{2}} \cdot \hat{\mathsf{se}}_B(\hat{\theta}).$$

Bootstrap-Bias

- Genau wie bei jedem klassischen Schätzer kann man auch für einen Bootstrapschätzer $\hat{\theta}$ den Bias definieren:

$$\text{Bias}_{\hat{f}}(\hat{\theta}) = E_{\hat{f}}s(\mathbf{x}^*) - t(\hat{f}) = E(\hat{\theta}^*) - \hat{\theta}.$$

- Achtung: Der Bias wird hier mit Bezug auf den geschätzten Parameter $\hat{\theta}$ der gesamten Stichprobe berechnet.

- Aufgabe: Schreiben Sie ein Programm, das die Abhängigkeit des Bootstrap-Bias von der Zahl der Replikationen grafisch sichtbar macht!

- Erzeugen Sie die erste Stichprobe $\mathbf{x}$ aus einer Verteilung, bei der 10% der Daten aus einer $N(0,1)$ und 90% der Daten aus einer $N(6,9)$ gezogen werden.

- Als einfaches Beispiel soll der Bias der Mittelwertschätzung betrachtet werden.

- Betrachten Sie die in Zehnerschritten die Replikationszahlen von 10 bis 1000.

- Überlegen Sie zunächst, in welche Einzelschritte das Problem für die Programmierung zerlegt werden kann.

- Berechnen Sie zunächst die Werte für den Bias, überlegen Sie erst dann, wie ein geeigneter Plot aussehen könnte.

- Lösung:

```
samplesize<-100
x<- c(rnorm(samplesize/10),
      rnorm(100 - samplesize/10, mean=6, sd=3))
samplemean <- mean(x)
bootbias <- rep(NA,100)
for (repl in seq(10,1000,10)) {
  bootbias[repl/10] <- mean(
        replicate(repl,
          mean(sample(x,length(x),replace=TRUE)-samplemean)))
}
plot(bootbias, t="l", axes=FALSE)
box(); axis(2)
axis(1,at=seq(10,100,10) , labels=seq(10,1000,100))
abline(h=0)
```

Konvergenzeigenschaften des Bootstrap

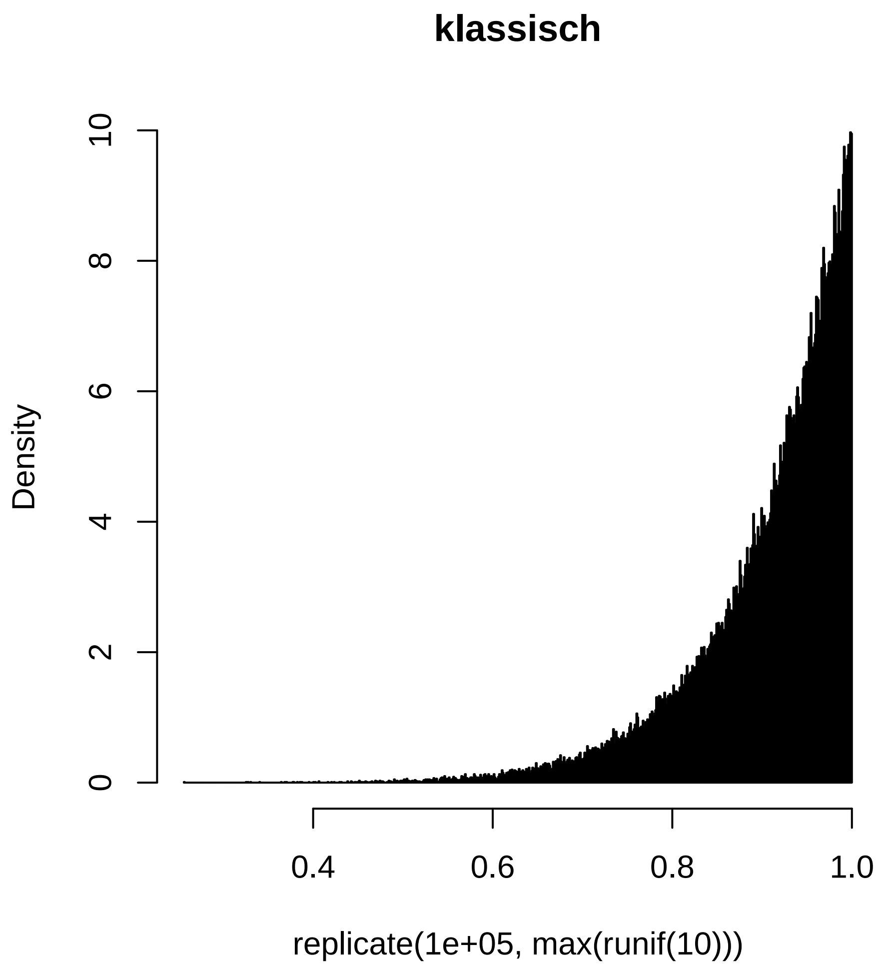
- Es gilt also, dass mit $N \to \infty$: $\text{Bias}_{\hat{f}}(\hat{\theta}) \to 0$!
- Weiterhin gilt, dass mit $n, N \to \infty$: $\hat{\text{se}}_B(\hat{\theta}) \to \text{sd}(\hat{\theta})$ aufgefasst als Zufallsvariable.

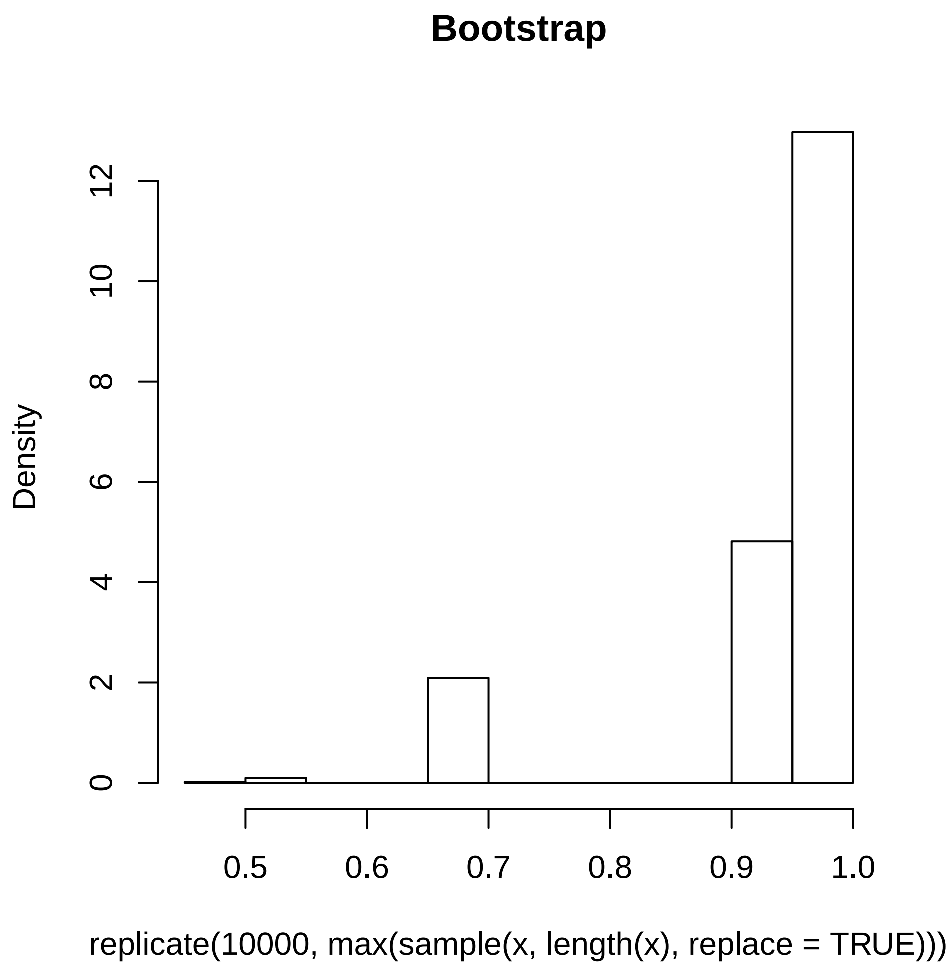
Grenzen des nicht-parametrischen Bootstrap

- Wenn nur eine kleine Stichprobe vorliegt, kann die Verteilung einer Teststatistik nur wenige diskrete Werte annehmen, auch wenn ihre wahre Verteilung stetig ist.

- Beispiel: Gegeben eine Stichprobe $\mathbf{x}$ vom Umfang 10 aus $U_{[0,1]}$. Die gesuchte Statistik ist das Maximum x_{max} für Stichproben vom Umfang 10 aus $U_{[0,1]}$.

- Im folgenden Bild werden einmal die "Dichteschätzer" für das x_{max} von 10000 Stichproben aus $U_{[0,1]}$ und zum anderen die geschätzte Dichte des entsprechende Bootstrapschätzers gegenübergestellt.


klassisch
Density
0
2
4
6
8
10
0.4
0.6
0.8
1.0
replicate(1e+05, max(runif(10)))


Bootstrap
Density
0
2
4
6
8
10
12
0.5
0.6
0.7
0.8
0.9
1.0
replicate(10000, max(sample(x, length(x), replace = TRUE)))

- Das Programm für diese Bilder:

```
par(mfrow=c(1,2))
hist(replicate(10000, max(runif(10))), prob=TRUE, nclass=100)
x <- runif(10)
hist(replicate(10000, max(sample(x,length(x),replace=TRUE))),
     prob=TRUE)
```

- Es ist klar zu erkennen, dass in dieser Situation der Bootstrap völlig versagt. Die kleine Stichprobe enthält einfach zu wenig Information über die zugrundeliegende Verteilung!

- Als Ausweg bleibt nur entweder eine “glatter” Schätzung von $\hat{f}$ durchzuführen, bzw. über f weiter Annahmen zu treffen. Eine typische Annahme ist die Zugehörigkeit von f zu einer bestimmten Verteilungsklasse. Beispielsweise $f = \varphi(\mu, \sigma^2)$, eine Normalverteilungsdichte.

Der parametrische Bootstrap

- Der parametrische Bootstrap liegt relativ nah an den klassischen Methoden und soll die Schwächen des klassischen Bootstrap beheben.
- Er kommt zum Einsatz, wenn die klassischen Methoden dabei versagen, die Genauigkeit der Schätzer zu bestimmen.
- Zum parametrischen Bootstrap gehört immer eine Annahme über die Verteilung f_Θ, aus der die ursprüngliche Stichprobe $\mathbf{x}$ gezogen wurde. Dabei ist $\Theta \in \mathcal{R}^m$, m die Dimension des Parameterraums von f.
- Beispiel: Gleichverteilung auf $[\vartheta_1, \vartheta_2]$ mit $\Theta = (\theta_1, \theta_2)$.

- Aus dieser Stichprobe wird zum einen die gesuchte Teststatistik $\hat{\theta}$ (Maximum, 0.75 Quantil ...) berechnet, zum anderen Schätzer $\hat{\Theta} = (\hat{\vartheta_1}, \ldots, \hat{\vartheta_m})$ für die Parameter der angenommenen Verteilung hergeleitet.

- Der eigentliche Bootstrap-Schritt arbeitet dann nicht mit einem Resampling aus $\mathbf{x}$, sondern mit einem Sampling neuer Stichproben $\mathbf{x}_i^*,\ 1 \leq i \leq N$ des Umfangs n aus $\hat{F}_{\hat{\Theta}}$.

- Die Berechnungen der Bootstrap-Schätzer geschehen dann mit denselben Formeln, wie beim nichtparametrischen Bootstrap.

Beispiel: Parametrischer Bootstrap

- Angenommen unsere Stichprobe vom Umfang $n = 10$ stammt aus einer Gleichverteilung auf $U_{[0,\vartheta]}$, ϑ unbekannt.

```
x<- runif(10)
```

- Gesucht ist die Verteilung des Maximums x_{max}.

- Der Momentenschätzer für den Parameterschätzer $\hat{\vartheta}$ ergibt sich aus

$$E_{f_\vartheta}(X) = \frac{\vartheta}{2}.$$

- Also ist
$$\hat{\vartheta} = 2\bar{X}$$
der Momentenschätzer für die obere Grenze θ der angenommenen Verteilung.

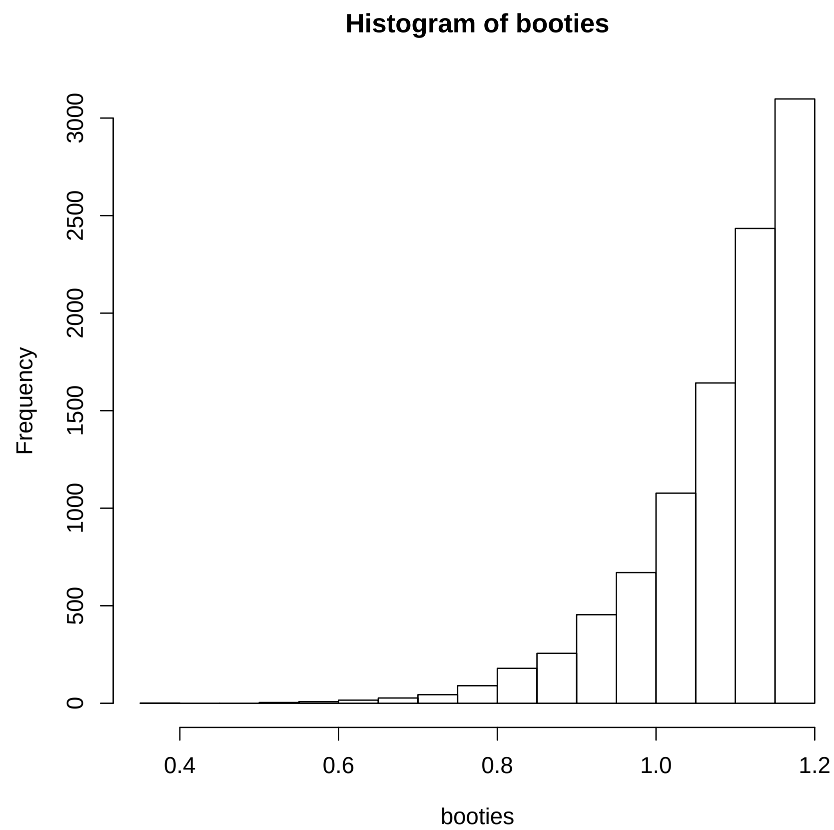
```
hattheta <- 2* mean(x)
```

- Aus dieser geschätzten Verteilung $\hat{f} = f_{\hat{\vartheta}}$ werden nun $N = 10000$ neue Stichproben gezogen und jeweils der ein Bootstrapschätzer für ϑ berechnet.

```
booties <- replicate(10000, max(hattheta*runif(10)))
```

- Das Histogramm kann nun eine erste Schätzung der Verteilung der gesuchten Statistik bieten.

`hist(booties)`



- Aus den Bootstrap Samples kann nun natürlich auch ein Standardfehler für die gesuchte Statistik abgeleitet werden.

```
sd(booties)
```

- Mittels Bootstrapping lassen sich auch empirisch p-Wert berechnen, Konfidenzintervalle etc.

Aufgabe zum parametrischen Bootstrap

- Ihnen liegt eine Stichprobe $\mathbf{x}$ vom Umfang 15 von Wartezeiten an einem Schalter vor (gerundet auf Sekunden).
  x=(260, 522, 619, 1433, 417, 121, 105, 438, 227, 402, 41, 6, 102, 225, 259)
  Sie vermuten, da es sich um Wartezeiten handelt, dass es sich um eine Stichprobe aus einer Exponentialverteilung handelt.

- Den Auftraggeber interessiert ein Konfidenzintervall für die mediane Wartezeit aus der unbekannten, zugrundeliegenden Exponentialverteilung.

- Schätzen Sie mit der Momentenmethode den Parameter λ der Exponentialverteilung.

- Schätzen Sie mittels parametrischem Bootstrap den Standardfehler des Medianschätzers für $N = 250$ Replikationen.

- Geben Sie ein 99%-Bootstrap-Konfidenzintervall für den Median an.