Advanced Econometrics #2: Simulations & Bootstrap

A. Charpentier (Université de Rennes 1)

Université de Rennes 1,
Graduate Course, 2018.
Motivation

Before computers, statistical analysis used probability theory to derive statistical expression for standard errors (or confidence intervals) and testing procedures, for some linear model

\[ y_i = \mathbf{x}_i^T \beta + \epsilon_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{j,i} + \epsilon_i. \]

But most formulas are approximations, based on large samples \((n \to \infty)\).

With computers, simulations and resampling methods can be used to produce (numerical) standard errors and testing procedure (without the use of formulas, but with a simple algorithm).
Overview

Linear Regression Model: \( y_i = \beta_0 + x_i^T \beta + \varepsilon_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i \)

- Nonlinear Transformations: smoothing techniques
- Asymptotics vs. Finite Distance: bootstrap techniques
- Penalization: Parcimony, Complexity and Overfit
- From least squares to other regressions: quantiles, expectiles, etc.
**Historical References**

Permutation methods go back to Fisher (1935) *The Design of Experiments* and Pitman (1937) *Significance tests which may be applied to samples from any population* (there are $n!$ distinct permutations)

Jackknife was introduced in Quenouille (1949) *Approximate tests of correlation in time series*, popularized by Tukey (1958) *Bias and confidence in not quite large samples*

Bootstrapping started with Monte Carlo algorithms in the 40’s, see e.g. Simon & Burstein (1969) *Basic Research Methods in Social Science*

Efron (1979) *Bootstrap methods: Another look at the jackknife* defined a resampling procedure that was coined as “bootstrap”. (there are $n^n$ possible distinct ordered bootstrap samples)
References

Motivation


References

Davison, A.C. & Hinkley, D.V. 1997 Bootstrap Methods and Their Application. CUP.

Efron B. & Tibshirani, R.J. An Introduction to the Bootstrap. CRC Press.


# Preliminaries: Generating Randomness

The following is a table of random digits from the book "A Million Random Digits with 100,000 Normal Deviates" by RAND, 1955.

<table>
<thead>
<tr>
<th>Random Digits</th>
<th>00000</th>
<th>00001</th>
<th>00002</th>
<th>00003</th>
<th>00004</th>
<th>00005</th>
<th>00006</th>
<th>00007</th>
<th>00008</th>
<th>00009</th>
<th>00010</th>
<th>00011</th>
<th>00012</th>
<th>00013</th>
<th>00014</th>
<th>00015</th>
<th>00016</th>
<th>00017</th>
<th>00018</th>
<th>00019</th>
<th>00020</th>
<th>00021</th>
<th>00022</th>
<th>00023</th>
<th>00024</th>
<th>00025</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10097</td>
<td>32533</td>
<td>76520</td>
<td>13586</td>
<td>34673</td>
<td>54876</td>
<td>80959</td>
<td>09117</td>
<td>39292</td>
<td>74945</td>
<td>66065</td>
<td>74717</td>
<td>34072</td>
<td>76850</td>
<td>36697</td>
<td>36170</td>
<td>65813</td>
<td>39885</td>
<td>11999</td>
<td>29170</td>
<td>31060</td>
<td>10805</td>
<td>45571</td>
<td>82406</td>
<td>35303</td>
<td>42614</td>
</tr>
<tr>
<td></td>
<td>08422</td>
<td>68953</td>
<td>19645</td>
<td>09303</td>
<td>23209</td>
<td>02560</td>
<td>15953</td>
<td>34764</td>
<td>35080</td>
<td>33606</td>
<td>00000</td>
<td>01000</td>
<td>00002</td>
<td>00003</td>
<td>00004</td>
<td>00005</td>
<td>00006</td>
<td>00007</td>
<td>00008</td>
<td>00009</td>
<td>00010</td>
<td>00011</td>
<td>00012</td>
<td>00013</td>
<td>00014</td>
<td>00015</td>
</tr>
<tr>
<td></td>
<td>37542</td>
<td>04805</td>
<td>64894</td>
<td>74296</td>
<td>24805</td>
<td>24037</td>
<td>20636</td>
<td>10402</td>
<td>00822</td>
<td>91665</td>
<td>08422</td>
<td>68953</td>
<td>19645</td>
<td>09303</td>
<td>23209</td>
<td>02560</td>
<td>15953</td>
<td>34764</td>
<td>35080</td>
<td>33606</td>
<td>00000</td>
<td>01000</td>
<td>00002</td>
<td>00003</td>
<td>00004</td>
<td>00005</td>
</tr>
<tr>
<td></td>
<td>99019</td>
<td>02529</td>
<td>09376</td>
<td>70715</td>
<td>38311</td>
<td>31165</td>
<td>88676</td>
<td>74397</td>
<td>04436</td>
<td>27659</td>
<td>12807</td>
<td>99970</td>
<td>80157</td>
<td>36147</td>
<td>64032</td>
<td>36653</td>
<td>98951</td>
<td>16877</td>
<td>12171</td>
<td>76833</td>
<td>00005</td>
<td>01000</td>
<td>00002</td>
<td>00003</td>
<td>00004</td>
<td>00005</td>
</tr>
<tr>
<td></td>
<td>96057</td>
<td>43765</td>
<td>62857</td>
<td>53219</td>
<td>09053</td>
<td>57548</td>
<td>29468</td>
<td>28709</td>
<td>83491</td>
<td>25624</td>
<td>95764</td>
<td>59263</td>
<td>31594</td>
<td>29404</td>
<td>02812</td>
<td>31397</td>
<td>53258</td>
<td>09874</td>
<td>25568</td>
<td>28709</td>
<td>00005</td>
<td>01000</td>
<td>00002</td>
<td>00003</td>
<td>00004</td>
<td>00005</td>
</tr>
<tr>
<td></td>
<td>00970</td>
<td>37769</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
<td>00000</td>
</tr>
</tbody>
</table>

Source: A Million Random Digits with 100,000 Normal Deviates, RAND, 1955.
Preliminaries: Generating Randomness

Here random means a sequence of numbers do not exhibit any discernible pattern, i.e. successively generated numbers can not be predicted.

*A random sequence is a vague notion... in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests traditional with statisticians...* Derrick Lehmer, quoted in *Knuth (1997)*

The goal of Pseudo-Random Numbers Generators is to produce a sequence of numbers in $[0, 1]$ that imitates ideal properties of random number.

```r
> runif(30)

[1] 0.3087420 0.4481307 0.0308382 0.4235758 0.7713879 0.8329476
[7] 0.4644714 0.0763505 0.8601878 0.2334159 0.0861886 0.4764753
[13] 0.9504273 0.8466378 0.2179143 0.6619298 0.8372218 0.4521744
[19] 0.7981926 0.3925203 0.7220769 0.3899142 0.5675318 0.4224018
[25] 0.3309934 0.6504410 0.4680358 0.7361024 0.1768224 0.8252457
```
Linear Congruential Method

Produce a sequence of integers $U_1, U_2, \cdots$ between 0 and $m - 1$ following a recursive relationship $X_{i+1} = (aX_i + b) \mod m$, and set $U_i = X_i / m$.

E.g. Start with $X_0 = 17$, $a = 13$, $b = 43$ and $m = 100$. Then the sequence is

$$\{77, 52, 27, 2, 77, 52, 27, 2, 77, 52, 27, 2, 77, \cdots \}$$

Problem: not all values in $\{0, \cdots, m - 1\}$ are obtained, and there is a cycle here.

Solution: use (very) large values for $m$ and choose properly $a$ and $b$.

E.g. $m = 2^{32} - 1$, $a = 16807$ ($= 7^5$) and $b = 0$ (used in Matlab).
Linear Congruential Method

If we start with $X_0 = 77$, we get for $U_{100}, U_{101}, \cdots$

$$\{\cdots, 0.9814, 0.9944, 0.2205, 0.6155, 0.0881, 0.3152, 0.5028, 0.1531, 0.8171, 0.7405, \cdots \}$$

See L’Ecuyer (2017) for an historical perspective.
Randomness?

**Randomness?**

Heuristically,

1. calls should provide a **uniform sample**, \( \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_{u_i \in (a,b)} = b - a \) with \( b > a \),

2. calls should be **independent**, \( \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_{u_i \in (a,b), u_{i+k} \in (c,d)} = (b - a)(d - c) \)

\( \forall k \in \mathbb{N}, \) and \( b > a, \ d > c \).
Monte Carlo: from $\mathcal{U}[0,1]$ to any distribution

Recall that the cumulative distribution function of $Y$ is $F : \mathbb{R} \to [0, 1]$, $F(y) = \mathbb{P}[Y \leq y]$.

Since $F$ is an increasing function, define its (pseudo-)inverse $Q : (0, 1) \to \mathbb{R}$ as

$$Q(u) = \inf \{ y \in \mathbb{R} : F(y) > u \}$$

**Proposition** If $U \sim \mathcal{U}[0,1]$, then $Q(U) \sim F$. 
Monte Carlo

From the law of large numbers, if $U_1, U_2, \cdots$ is a sequence of i.i.d random variables, uniformly distributed on $[0, 1]$, and some mapping $h : [0, 1] \to \mathbb{R}$,

$$\frac{1}{n} \sum_{i=1}^{n} h(U_i) \overset{\text{a.s.}}{\longrightarrow} \mu = \int_{[0,1]} h(u) \, du = \mathbb{E}[h(U)], \text{ as } n \to \infty$$

and from the central limit theorem

$$\sqrt{n} \left( \left( \frac{1}{n} \sum_{i=1}^{n} h(U_i) \right) - \mu \right) \overset{\text{d}}{\rightarrow} \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \text{Var}[h(U)]$, and $U \sim \mathcal{U}_{[0,1]}$. 

@freakonometrics
Monte Carlo

Consider $h(u) = \cos(\pi u/2)$,

```r
> h=function(u) cos(u*pi/2)
> integrate(h,0,1)
0.6366198 with absolute error <7.1e-15
> mean(h(runif(1e6)))
[1] 0.6363378
```

We can actually repeat that a thousand time

```r
> M=rep(NA,1000)
> for(i in 1:1000) M[i]=mean(h(runif(1e6)))
> mean(M)
[1] 0.6366087
> sd(M)
[1] 0.000317656
```
Monte Carlo Techniques to Compute Integrals

Monte Carlo is a very general technique, that can be used to compute any integral.

Let $X \sim \text{Cauchy}$ what is $\mathbb{P}[X > 2]$. Observe that

$$\mathbb{P}[X > 2] = \int_{2}^{\infty} \frac{dx}{\pi(1 + x^2)} \quad (\sim 0.15)$$

since $f(x) = \frac{1}{\pi(1 + x^2)}$ and $Q(u) = F^{-1}(u) = \tan \left( \pi \left[ u - \frac{1}{2} \right] \right)$.

Crude Monte Carlo: use the law of large numbers

$$\hat{p}_1 = \frac{1}{n} \sum_{i=1}^{n} 1(Q(u_i) > 2)$$

where $u_i$ are obtained from i.i.d. $\mathcal{U}([0, 1])$ variables.

Observe that $\text{Var}[\hat{p}_1] \sim \frac{0.127}{n}$. 
Crude Monte Carlo (with symmetry): \( \mathbb{P}[X > 2] = \mathbb{P}[|X| > 2]/2 \) and use the law of large numbers

\[
\hat{p}_2 = \frac{1}{2n} \sum_{i=1}^{n} 1(|Q(u_i)| > 2)
\]

where \( u_i \) are obtained from i.i.d. \( \mathcal{U}([0, 1]) \) variables.

Observe that \( \text{Var}[\hat{p}_2] \sim \frac{0.052}{n} \).

Using integral symmetries:

\[
\int_{2}^{\infty} \frac{dx}{\pi(1 + x^2)} = \frac{1}{2} - \int_{0}^{2} \frac{dx}{\pi(1 + x^2)}
\]

where the later integral is \( \mathbb{E}[h(2U)] \) where \( h(x) = \frac{2}{\pi(1 + x^2)} \).

From the law of large numbers

\[
\hat{p}_3 = \frac{1}{2} - \frac{1}{n} \sum_{i=1}^{n} h(2u_i)
\]
where \( u_i \) are obtained from i.i.d. \( \mathcal{U}([0, 1]) \) variables.

Observe that \( \text{Var}[\hat{p}_3] \sim \frac{0.0285}{n} \).

Using integral transformations:

\[
\int_2^\infty \frac{dx}{\pi(1 + x^2)} = \int_0^{1/2} \frac{y^{-2} dy}{\pi(1 - y^{-2})}
\]

which is \( \mathbb{E}[h(U/2)] \) where \( h(x) = \frac{1}{2\pi(1 + x^2)} \).

From the law of large numbers

\[
\hat{p}_4 = \frac{1}{4n} \sum_{i=1}^{n} h(u_i/2)
\]

where \( u_i \) are obtained from i.i.d. \( \mathcal{U}([0, 1]) \) variables.

Observe that \( \text{Var}[\hat{p}_4] \sim \frac{0.0009}{n} \).
The Empirical Measure

Consider a sample \( \{y_1, y_2, \ldots, y_n\} \). Its empirical cumulative distribution function is

\[
\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} 1_{[-\infty, y]}(y_i).
\]

From Kolmogorov-Smirnov theorem \( \lim_{n \to \infty} \hat{F}_n(y) = F(y) \), while Glivenko–Cantelli theorem, states that the convergence in fact happens uniformly

\[
\|\hat{F}_n - F\|_\infty = \sup_{y \in \mathbb{R}} |\hat{F}_n(y) - F(y)| \xrightarrow{\text{a.s.}} 0.
\]
The Empirical Measure

Furthermore, pointwise, \( \hat{F}_n(y) \) has asymptotically normal distribution with the standard \( \sqrt{n} \) rate of convergence:

\[
\sqrt{n}(\hat{F}_n(y) - F(y)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, F(y)(1 - F(y))\right).
\]

Let \( \hat{Q}_n \) denote the pseudo-inverse of \( \hat{F}_n \). Note that \( \forall u \in (0, 1), \exists i \) such that \( \hat{Q}_n(u) = y_i \). More specifically, if \( y_{1:n} \leq y_{2:n} \leq \cdots \leq y_{n:n} \),

\[
\hat{Q}_n(u) = y_{i:n} \text{ where } i - 1 \leq \frac{u}{n} < i.
\]

**Proposition** Generating numbers from distribution \( \hat{F}_n \) means draw randomly, with replacement, uniformly, in \( \{y_1, \cdots, y_n\} \).
Kolmogorov-Smirnov Test and Monte Carlo

Kolmogorov-Smirnov test, $H_0 : F = F_0$ (against $H_1 : F \neq F_0$). The test statistic for a given cdf $F_0$ is

$$D_n = \sup_x \{|\hat{F}_n(x) - F(x)|\}$$

One can prove that under $H_0$, $\sqrt{n}D_n \xrightarrow{L} \sup_t |B_{F(t)}|$, as $n \to \infty$, where $(B_t)$ is the Brownian bridge on $[0, 1]$.

Consider the height of 200 students.

```r
> Davis = read.table("http://socserv.socsci.mcmaster.ca/jfox/Books/
  Applied-Regression-2E/datasets/Davis.txt")
> Davis[12,c(2,3)]=Davis[12,c(3,2)]
> Y = Davis$height
> mean(Y)
[1] 170.565
> sd(Y)
[1] 8.932228
```
Kolmogorov-Smirnov Test and Monte Carlo

Let us test $F = \mathcal{N}(170, 9^2)$.

```r
> y0 = pnorm(140:205, 170, 9)
> for (s in 1:200) {
+   X = rnorm(length(Y), 170, 9)
+   y = Vectorize(ecdf(X))(140:205)
+   lines(140:205, y)
+   D[s] = max(y-y0)
+ }

while for $\hat{F}_n$,

```
Kolmogorov-Smirnov Test and Monte Carlo

The empirical distribution of $D$ is obtained using

```r
> hist(D, probability = TRUE )
> lines(density(D), col = "blue")
```

Here

```r
> (demp = max(abs(Vectorize(ecdf(Y))(140:205) - y0)))
```

```r
[1] 0.05163936
```

```r
> mean(D > demp)
```

```r
[1] 0.2459
```

```r
> ks.test(Y, "pnorm", 170, 9)
```

```r
D = 0.062969, p-value = 0.406
```
Bootstrap Techniques (in one slide)

Bootstrapping is an asymptotic refinement based on computer based simulations. Underlying properties: we know when it might work, or not. Idea: \( \{(y_i, x_i)\} \) is obtained from a stochastic model under \( \mathbb{P} \). We want to generate other samples (not more observations) to reduce uncertainty.
Heuristic Intuition for a Simple (Financial) Model

Consider a return stochastic model, \( r_t = \mu + \sigma \varepsilon_t \), for \( t = 1, 2, \cdots, T \), with \( (\varepsilon_t) \) is i.i.d. \( \mathcal{N}(0, 1) \) [Constant Expected Return Model, CER]

\[
\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_t \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} [r_t - \hat{\mu}]^2
\]

then (standard errors)

\[
\hat{se}[\hat{\mu}] = \frac{\hat{\sigma}}{\sqrt{T}} \quad \text{and} \quad \hat{se}[\hat{\sigma}] = \frac{\hat{\sigma}}{\sqrt{2T}}
\]

then (confidence intervals)

\[
\mu \in \left[ \hat{\mu} \pm 2\hat{se}[\hat{\mu}] \right] \quad \text{and} \quad \sigma \in \left[ \hat{\sigma} \pm 2\hat{se}[\hat{\sigma}] \right]
\]

What if the quantity of interest, \( \theta \), is another quantity, e.g. a Value-at-Risk ?
Heuristic Intuition for a Simple (Financial) Model

One can use nonparametric bootstrap

1. resampling: generate $B$ “bootstrap samples” by resampling with replacement in the original data,

$$r^{(b)} = \{r_1^{(b)}, \ldots, r_T^{(b)}\}, \text{ with } r_t^{(b)} \in \{r_1, \ldots, r_T\}.$$  

2. For each sample $r^{(b)}$, compute $\hat{\theta}^{(b)}$

3. Derive the empirical distribution of $\hat{\theta}$ from $\{\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}\}$.

4. Compute any quantity of interest, standard error, quantiles, etc.

E.g. estimate the bias

$$\text{bias}[\hat{\theta}] = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{(b)} - \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}$$

bootstrap mean    estimate
Heuristic Intuition for a Simple (Financial) Model

E.g. estimate the standard error

\[
\text{se}[\hat{\theta}] = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left( \hat{\theta}(b) - \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}(b) \right)^2}
\]

E.g. estimate the confidence interval, if the bootstrap distribution looks Gaussian

\[
\theta \in \left[ \hat{\theta} \pm 2\text{se}[\hat{\theta}] \right]
\]

and if the distribution does not look Gaussian

\[
\theta \in \left[ q_{\alpha/2}^{(B)}; q_{1-\alpha/2}^{(B)} \right]
\]

where \( q_{\alpha}^{(B)} \) denote a quantile from \( \{\hat{\theta}(1), \ldots, \hat{\theta}(B)\} \).
Monte Carlo Techniques in Statistics

Law of large numbers (---), if $E[X] = 0$ and $\text{Var}[X] = 1$ : $\sqrt{n} \bar{X}_n \xrightarrow{L} \mathcal{N}(0, 1)$

What if $n$ is small? What is the distribution of $\bar{X}_n$?

Example : $X$ such that $2^{-\frac{1}{2}}(X - 1) \sim \chi^2(1)$

Use Monte Carlo Simulation to derive confidence interval for $\bar{X}_n$ (---).

Generate samples $\{x_1^{(m)}, \cdots, x_n^{(m)}\}$ from $\chi^2(1)$, and compute $\bar{x}_n^{(m)}$

Then estimate the density of $\{\bar{x}_n^{(1)}, \cdots, \bar{x}_n^{(m)}\}$, quantiles, etc.

Problem : need to know the true distribution of $X$.

What if we have only $\{x_1, \cdots, x_n\}$ ?

Generate samples $\{x_1^{(m)}, \cdots, x_n^{(m)}\}$ from $\hat{F}_n$, and compute $\bar{x}_n^{(m)}$ (---)
> n = 20
> ns = 1e6
> xbar = rep(NA,ns)
> for(i in 1:ns){
+ x = (rchisq(n,df=1)-1)/sqrt(2)
+ xbar[i] = mean(x)
+ }
> u = seq(-.7,.8,by=.001)
> v = dnorm(u,sd=1/sqrt(20))
> plot(u,v,col="black")
> lines(density(xbar),col="red")
> set.seed(1)
> x = (rchisq(n,df=1)-1)/sqrt(2)
> for(i in 1:ns){
+ xs = sample(x,size=n,replace=TRUE)
+ xbar[i] = mean(xs)
+ }
> lines(density(xbar),col="blue")
Monte Carlo Techniques in Statistics

Consider empirical residuals from a linear regression, \( \hat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \hat{\beta} \).

Let \( \hat{F}(z) = \frac{1}{n} \sum_{i=1}^{n} 1 \left( \frac{\hat{\varepsilon}_i}{\hat{\sigma}} \leq z \right) \) denote the empirical distribution of Studentized residuals.

Could we test \( H_0 : F = \mathcal{N}(0,1) \)?

```r
1 > X = rnorm(50)
2 > cdf = function(z) mean(X <= z)
```

Simulate samples from a \( \mathcal{N}(0,1) \) (true distribution under \( H_0 \))
Quantifying Bias

Consider $X$ with mean $\mu = \mathbb{E}(X)$. Let $\theta = \exp[\mu]$, then $\hat{\theta} = \exp[\bar{x}]$ is a biased estimator of $\theta$, see Horowitz (1998)

The Bootstrap

**Idea 1**: Delta Method, i.e. if $\sqrt{n}[\hat{\tau}_n - \tau] \xrightarrow{L} \mathcal{N}(0, \sigma^2)$, then, if $g'(\tau)$ exists and is non-null,

$$\sqrt{n}[g(\hat{\tau}_n) - g(\tau)] \xrightarrow{L} \mathcal{N}(0, \sigma^2[g'(\tau)]^2)$$

so $\hat{\theta}_1 = \exp[\bar{x}]$ is asymptotically unbiased.

**Idea 2**: Delta Method based correction,

based on $\hat{\theta}_2 = \exp \left[ \bar{x} - \frac{s^2}{2n} \right]$ where $s^2 = \frac{1}{n} \sum_{i=1}^{n} [x_i - \bar{x}]^2$.

**Idea 3**: Use Bootstrap, $\hat{\theta}_3 = \frac{1}{B} \sum_{b=1}^{n} \exp[\bar{x}^{(b)}]$
Quantifying Bias

$X$ with mean $\mu = \mathbb{E}(X)$. Let $\theta = \exp[\mu]$. Consider three distributions

- **Log-normal**
- **Student $t_{10}$**
- **Student $t_{5}$**

![Graphs showing bias quantification for different distributions](image-url)
```r
> VS = matrix(NA, 15, 3)
> for (s in 1:15) {
+ simu = function(n = 10) {
+   get_i = function(i) {
+     x = rnorm(n, sd = sqrt(6));
+     S = matrix(sample(x, size = n * 10000, replace = TRUE), ncol = 10000)
+     ThetaBoot = exp(colMeans(S))
+     Bias = mean(ThetaBoot) - exp(mean(x))
+     theta = exp(mean(x))/exp(.5 * var(x)/n)
+     c(exp(mean(x)), exp(mean(x)) - Bias, theta)
+   }
+   res = mclapply(1:2000, get_i, mc.cores = 20)
+   res = lapply(1:10000, get_i)
+   res = do.call(rbind, res)
+   bias = colMeans(res - 1)
+   return(bias)
+ }
+ VS[s, ] = simu(10 * s)
+ }
```

```r
# res = mclapply(1:2000, get_i, mc.cores=20)
+ res = lapply(1:10000, get_i)
+ res = do.call(rbind, res)
+ bias = colMeans(res - 1)
+ return(bias)
+ }
+ VS[s, ] = simu(10 * s)
+ ```
Linear Regression & Bootstrap : Parametric

1. sample $\tilde{\varepsilon}^{(s)}_1, \cdots, \tilde{\varepsilon}^{(s)}_n$ randomly from $\mathcal{N}(0, \hat{\sigma})$

2. set $y^{(s)}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \tilde{\varepsilon}^{(s)}_i$

3. consider dataset $(x_i, y^{(b)}_i) = (x_i, y^{(b)}_i)$’s
   and fit a linear regression

4. let $\hat{\beta}^{(s)}_0, \hat{\beta}^{(s)}_1$ and $\hat{\sigma}^2(s)$ denote the estimated values
**Linear Regression & Bootstrap : Residuals**

**Algorithm 6.1.** Davison & Hinkley (1997) *Bootstrap Methods and Applications.*

1. sample \( \hat{\varepsilon}_1^{(b)}, \ldots, \hat{\varepsilon}_n^{(b)} \) randomly with replacement in \( \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\} \)

2. set \( y_i^{(b)} = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\varepsilon}_i^{(b)} \)

3. consider dataset \((x_i, y_i^{(b)}) = (x_i, y_i^{(b)})'s \) and fit a linear regression

4. let \( \hat{\beta}_0^{(b)}, \hat{\beta}_1^{(b)} \) and \( \hat{\sigma}^{2(b)} \) denote estimated values

\[
\hat{\beta}_1^{(b)} = \frac{\sum [x_i - \bar{x}] \cdot y_i^{(b)}}{\sum [x_i - \bar{x}]^2} = \hat{\beta}_1 + \frac{\sum [x_i - \bar{x}] \cdot \hat{\varepsilon}_i^{(b)}}{\sum [x_i - \bar{x}]^2}
\]

hence \( \mathbb{E}[\hat{\beta}_1^{(b)}] = \hat{\beta}_1 \), while

\[
\mathrm{Var}[\hat{\beta}_1^{(b)}] = \frac{\sum [x_i - \bar{x}]^2 \cdot \mathrm{Var}[\hat{\varepsilon}_i^{(b)}]}{(\sum [x_i - \bar{x}]^2)^2} \sim \frac{\sigma^2}{\sum [x_i - \bar{x}]^2}
\]
Linear Regression & Bootstrap : Pairs


1. sample \( \{i_1^{(b)}, \ldots, i_n^{(b)}\} \) randomly with replacement in \( \{1, 2, \ldots, n\} \)
2. consider dataset \((x_i^{(b)}, y_i^{(b)}) = (x_{i_i^{(b)}}, y_{i_i^{(b)}})\)’s and fit a linear regression
3. let \( \hat{\beta}_0^{(b)}, \hat{\beta}_1^{(b)} \) and \( \hat{\sigma}^2(b) \) denote the estimated values

Remark \( \mathbb{P}(i \notin \{i_1^{(b)}, \ldots, i_n^{(b)}\}) = \left(1 - \frac{1}{n}\right)^n \sim e^{-1} \)

Key issue : residuals have to be independent and identically distributed
**Linear Regression & Bootstrap**

Difference between the two algorithms:

1) with the second method, we make no assumption about variance homogeneity potentially more robust to heteroscedasticity

2) the simulated samples have different designs, because the \( x \) values are randomly sampled

Key issue: residuals have to be independent and identically distributed

See discussion below on

- dynamic regression, \( y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \varepsilon_t \)
- heteroskedasticity, \( y_i = \beta_0 + \beta_1 x_i + |x_i| \varepsilon_t \)
- instrumental variables and two-stage least squares
Simulation in Econometric Models

(almost) all quantities of interest can be written \( T(\varepsilon) \) with \( \varepsilon \sim F \).

E.g. \( \hat{\beta} = \beta + (X^TX)^{-1}X^T\varepsilon \)

We need \( \mathbb{E}[T(\varepsilon)] = \int t(\varepsilon)dF(\varepsilon) \)

Use simulations, i.e. draw \( n \) values \( \{\varepsilon_1, \cdots, \varepsilon_n\} \) since

\[
\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} T(\varepsilon_i) \right] = \mathbb{E}[T(\varepsilon)] \quad \text{(unbiased)}
\]

\[
\frac{1}{n} \sum_{i=1}^{n} T(\varepsilon_i) \xrightarrow{L} \mathbb{E}[T(\varepsilon)] \quad \text{as } n \to \infty \quad \text{(consistent)}
\]
Generating (Parametric) Distributions

Inverse cdf Technique:

Let $U \sim \mathcal{U}([0, 1])$, then $X = F^{-1}(U) \sim F$.

Proof 1:

$$
\mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[F \circ F^{-1}(U) \leq F(x)] = \mathbb{P}[U \leq F(x)] = F(x)
$$

Proof 2: set $u = F(x)$ or $x = F^{-1}(u)$ (change of variable)

$$
\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) dF^*(x) = \int_{0}^{1} h(F^{-1}(u)) du = \mathbb{E}[h(F^{-1}(U))]
$$

with $U \sim \mathcal{U}([0, 1])$, i.e. $X \overset{d}{=} F^{-1}(U)$. 
Rejection Techniques

Problem : If $X \sim F$, how to draw from $X^*$, i.e. $X$ conditional on $X \in [a, b]$?

Solution : draw $X$ and use accept-reject method

1. if $x \in [a, b]$, keep it (accept)
2. if $x \notin [a, b]$, draw another value (reject)

If we generate $n$ values, we accept - on average - $[F(b) - F(a)] \cdot n$ draws.
**Importance Sampling**

Problem : If $X \sim F$, how to draw from $X$ conditional on $X \in [a, b]$ ?

Solution : rewrite the integral and use importance sampling method

The conditional censored distribution $X^*$ is

$$dF^*(x) = \frac{dF(x)}{F(b) - F(a)} 1(x \in [a, b])$$

Alternative for truncated distributions : let $U \sim \mathcal{U}([0, 1])$ and set $\tilde{U} = [1 - U]F(a) + UF(b)$ and $Y = F^{-1}(\tilde{U})$
**Going Further : MCMC**

Intuition : we want to use the Central Limit Theorem, but i.id. sample is a (too) strong assumption: if \((X_i)\) is i.id. with distribution \(F\),

\[
\frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} h(X_i) - \int h(x) dF(x) \right) \xrightarrow{L} \mathcal{N}(0, \sigma^2), \text{ as } n \to \infty.
\]

Use the *ergodic theorem*: if \((X_i)\) is a Markov Chain with invariant measure \(\mu\),

\[
\frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} h(X_i) - \int h(x) d\mu(x) \right) \xrightarrow{L} \mathcal{N}(0, \sigma^2), \text{ as } n \to \infty.
\]

See *Gibbs sampler*

**Example** : complicated joint distribution, but simple conditional ones
Going Further : MCMC

To generate $X|X^T 1 \leq m$ with $X \sim \mathcal{N}(0, I)$ (in dimension 2)

1. draw $X_1$ from $\mathcal{N}(0, 1)$
2. draw $U$ from $\mathcal{U}([0, 1])$ and set $\tilde{U} = U \Phi(m - \epsilon_1)$
3. set $X_2 = \Phi^{-1}(\tilde{U})$

See Geweke (1991) Efficient Simulation from the Multivariate Normal and Distributions Subject to Linear Constraints
Monte Carlo Techniques in Statistics

Let \( \{y_1, \cdots, y_n\} \) denote a sample from a collection of \( n \) i.i.d. random variables with true (unknown) distribution \( F_0 \). This distribution can be approximated by \( \hat{F}_n \).

Parametric model: \( F_0 \in \mathcal{F} = \{F_\theta; \theta \in \Theta\} \).

Nonparametric model: \( F_0 \in \mathcal{F} = \{F \text{ is a c.d.f.}\} \)

The statistic of interest is \( T_n = T_n(y_1, \cdots, y_n) \) (see e.g. \( T_n = \hat{\beta}_j \)).

Let \( G_n \) denote the statistics of \( T_n \):

Exact distribution: \( G_n(t, F_0) = \mathbb{P}_{F_0}(T_n \leq t) \) under \( F_0 \)

We want to estimate \( G_n(\cdot, F_0) \) to get confidence intervals, i.e. \( \alpha \)-quantiles

\[
G_n^{-1}(\alpha, F_0) = \inf \{t; G_n(t, F_0) \geq \alpha\}
\]

or \( p \)-values,

\[
p = 1 - G_n(t_n, F_0)
\]
Approximation of $G_n(t_n, F_0)$

Two strategies to approximate $G_n(t_n, F_0)$:

1. Use $G_\infty(\cdot, F_0)$, the asymptotic distribution as $n \to \infty$.

2. Use $G_\infty(\cdot, \hat{F}_n)$

Here $\hat{F}_n$ can be the empirical cdf (nonparametric bootstrap) or $F_\theta$ (parametric bootstrap).
Approximation of $G_n(t_n, F_0)$: Linear Model

Consider the test of $H_0 : \beta_j = 0$, $p$-value being $p = 1 - G_n(t_n, F_0)$

- **Linear Model with Normal Errors** $y_i = \mathbf{x}_i^T \beta + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.
  
  Then $\frac{(\hat{\beta}_j - \beta_j)^2}{\hat{\sigma}_j^2} \sim \mathcal{F}(1, n - k) = G_n(\cdot, F_0)$ where $F_0$ is $\mathcal{N}(0, \sigma^2)$

- **Linear Model with Non-Normal Errors** $y_i = \mathbf{x}_i^T \beta + \varepsilon_i$, with $\mathbb{E}[\varepsilon_i] = 0$.
  
  Then $\frac{(\hat{\beta}_j - \beta_j)^2}{\hat{\sigma}_j^2} \xrightarrow{L} \chi^2(1) = G_\infty(\cdot, F_0)$ as $n \to \infty$. 

@freakonometrics
Approximation of $G_n(t_n, F_0)$: Linear Model

Application $y_i = \mathbf{x}_i^T \beta + \varepsilon_i$, $\varepsilon \sim \mathcal{N}(0, 1)$, $\varepsilon \sim \mathcal{U}([-1, +1])$ or $\varepsilon \sim \mathcal{Std}(\nu = 2)$.

Here $F_0$ is $\mathcal{N}(0, \sigma^2)$
```r
# pvf = function(t) mean((1-pf(t,1,length(t)-2))<.05)
> pvf = function(t) mean((1-pf(t,1,length(t)-2))<.05)
# pvq = function(t) mean((1-pchisq(t,1)<.05)
> pvq = function(t) mean((1-pchisq(t,1)<.05)
# TABLE = function(n=30) {
> TABLE = function(n=30) {
+ ns = 5000
+ x = c(1.0001,rep(1,n-1))
+ e = matrix(rnorm(n*ns),n)
+ e2 = matrix(runif(n*ns,-3,3),n)
+ e3 = matrix(rt(n*ns,2),n)
+ get_i = function(i){
+ r1 = lm(e[,i]~x)
+ r2 = lm(e2[,i]~x)
+ r3 = lm(e3[,i]~x)
+ t1 = r1$coef[2]^2/vcov(r1)[2,2]
+ t2 = r2$coef[2]^2/vcov(r2)[2,2]
+ t3 = r3$coef[2]^2/vcov(r3)[2,2]
+ c(t1,t2,t3)}
+ library(parallel)
+ t = mclapply(1:ns, get_i, mc. cores=50)
+ t = lapply(1:ns, get_i)
+ t = simplify2array(t)
+ rj1 = pvf(t[,1])
+ rj2 = pvf(t[,2])
+ rj3 = pvf(t[,3])
+ rj12 = pvq(t[,1])
+ rj22 = pvq(t[,2])
+ rj32 = pvq(t[,3])
+ ans = rbind(c(rj1,rj2,rj3),c(rj12,rj22,rj32))
+ return(ans) }
> TABLE(30)
```

Approximation of $G_n(t_n, F_0)$: Linear Model

```r
> ns=1e5  
> PROP=matrix(NA,ns,6)  
> n=30  
> VN=seq(10,140,by=10)  
> for (s in 1:ns){  
+ X=rnorm(n)  
+ E=rnorm(n)  
+ Y=1+X+E  
+ reg=lm(Y~X)  
+ T=(coefficients(reg)[2]-1)^2/vcov(reg)[2,2]  
+ PROP[s,1]=T>qf(.95,1,n-2)  
+ PROP[s,2]=T>qchisq(.95,1)  
+ E=rt(n,df=3)  
+ Y=1+X+E
```

```r
1 + reg=lm(Y~X)  
2 + T=(coefficients(reg)[2]-1)^2/ 
3 + vcov(reg)[2,2]  
4 + PROP[s,3]=T>qf(.95,1,n-2)  
5 + PROP[s,4]=T>qchisq(.95,1)  
6 + E=runif(n)*4-2  
7 + Y=1+X+E  
8 + reg=lm(Y~X)  
9 + T=(coefficients(reg)[2]-1)^2/ 
10 + vcov(reg)[2,2]  
11 + PROP[s,5]=T>qf(.95,1,n-2)  
12 + PROP[s,6]=T>qchisq(.95,1)  
13 + }
14 > apply(PROP,mean,2)
```
**Computation of** $G_\infty(t, \hat{F}_n)$

For $b \in \{1, \cdots, B\}$, generate bootstrap samples of size $n$, $\{\hat{\varepsilon}_1^{(b)}, \cdots, \hat{\varepsilon}_n^{(b)}\}$ by drawing from $\hat{F}_n$.

Compute $T^{(b)} = T_n(\hat{\varepsilon}_1^{(b)}, \cdots, \hat{\varepsilon}_n^{(b)})$, and use sample $\{T^{(1)}, \cdots, T^{(B)}\}$ to compute $\hat{G}$,

$$\hat{G}(t) = \frac{1}{B} \sum_{b=1}^{B} 1(T^{(b)} \leq t)$$
Linear Model: computation of $G_\infty(t, \hat{F}_n)$

Consider the test of $H_0 : \beta_j = 0$, $p$-value being $p = 1 - G_n(t_n, F_0)$

1. compute $t_n = \frac{(\hat{\beta}_j - \beta_j)^2}{\hat{\sigma}_j^2}$

2. generate $B$ bootstrap samples, under the null assumption

3. for each bootstrap sample, compute $t_n^{(b)} = \frac{(\hat{\beta}_j^{(b)} - \hat{\beta}_j)^2}{\hat{\sigma}_j^{(b)}^2}$

4. reject $H_0$ if $\frac{1}{B} \sum_{i=1}^{B} 1(t_n > t_n^{(b)}) < \alpha$. 
Linear Model: computation of $G_\infty(t, \hat{F}_n)$

Application $y_i = x_i^T \beta + \varepsilon_i$, $\varepsilon \sim \mathcal{N}(0, 1)$, $\varepsilon \sim \mathcal{U}([-1, +1])$ or $\varepsilon \sim \mathcal{Std}(\nu = 2)$. 

![Sample Size vs Rejection Rate](image.png)
```r
> TABLE2 = function (n = 30) {
+   B = 299
+   sn = sqrt(n / (n - 1))
+   ns = 5000
+   x = rep(1, n)
+   x[1] = 1.0001
+   e = matrix(rnorm(n * ns), n)
+   e2 = matrix(runif(n * ns, -3, 3), n)
+   e3 = matrix(rt(n * ns, 2), n)
+   b0tilde1 = colMeans(e)
+   b0tilde2 = colMeans(e2)
+   b0tilde3 = colMeans(e3)
+   getB_i = function (i) {
+     u1 = (e[,i]- b0tilde1[i]) * sn
+     u2 = (e2[,i]- b0tilde2[i]) * sn
+     u3 = (e3[,i]- b0tilde3[i]) * sn
+     Indic = matrix(sample(n, size = B * n, replace = TRUE), n, B)
+     getB_j = function (j) {
+       y1 = u1[Indic[,j]] + b0tilde1[i]
+       y2 = u2[Indic[,j]] + b0tilde2[i]
+       y3 = u3[Indic[,j]] + b0tilde3[i]
+       r1 = lm(y1 ~ x)
+       r2 = lm(y2 ~ x)
+       r3 = lm(y3 ~ x)
+       t = r1$coef[2]^2 / vcov(r1)[2,2]
+       t2 = r2$coef[2]^2 / vcov(r2)[2,2]
+       t3 = r3$coef[2]^2 / vcov(r3)[2,2]
+       c(t, t2, t3)
+     }
+     res = sapply(1:B, getB_j)
+     rj1 = mean(res[1,] < t[1, i])
+     rj2 = mean(res[2,] < t[2, i])
+     rj3 = mean(res[3,] < t[3, i])
+     c(rj1, rj2, rj3) < 0.05
+     tstar = lapply(1:ns, getB_i)
+     tstar = simplify2array(tstar)
+     tstar <- rowMeans(tstar)
+     return(tstar)
+   }
+   rj1 = mean(res[1,] < t[1, i])
+   rj2 = mean(res[2,] < t[2, i])
+   rj3 = mean(res[3,] < t[3, i])
+   c(rj1, rj2, rj3) < 0.05
+   tstar = lapply(1:ns, getB_i)
+   tstar = simplify2array(tstar)
+   tstar <- rowMeans(tstar)
+   return(tstar)
+ }
```
Linear Regression

What does generate $B$ bootstrap samples, under the null assumption means?

Use residual bootstrap technique:

**Example**: (standard) linear model, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $H_0 : \beta_1 = 0$.

2.1. Estimate the model under $H_0$, i.e. $y_i = \beta_0 + \eta_i$, and save $\{\hat{\eta}_1, \cdots, \hat{\eta}_n\}$

2.2. Define $\tilde{\eta} = \{\tilde{\eta}_1, \cdots, \tilde{\eta}_n\}$ with $\tilde{\eta} = \sqrt{\frac{n}{n-1}} \hat{\eta}$

2.3. Draw (with replacement) residuals $\tilde{\eta}^{(b)} = \{\tilde{\eta}_1^{(b)}, \cdots, \tilde{\eta}_n^{(b)}\}$

2.4. Set $y_i^{(b)} = \hat{\beta}_0 + \tilde{\eta}_i^{(b)}$

2.5. Estimate the regression model $y_i^{(b)} = \beta_0^{(b)} + \beta_1^{(b)} x_i + \varepsilon_i^{(b)}$
Going Further on Linear Regression

Recall that the OLS estimator satisfies

$$\sqrt{n} (\hat{\beta} - \beta_0) = \left( \frac{1}{n} X^T X \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i$$

while for the bootstrap

$$\sqrt{n} (\hat{\beta}^{(b)} - \hat{\beta}) = \left( \frac{1}{n} X^T X \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i^{(b)}$$

Thus, for i.i.d. data, the variance is

$$\mathbb{E} \left[ \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i \right)^T \right] = \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T \varepsilon_i^2 \right]$$
Going Further on Linear Regression

and similarly (for i.id. data)

\[
\mathbb{E} \left[ \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i^{(b)} \right) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i^{(b)} \right)^\top \left| X, Y \right. \right] = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top \widehat{\varepsilon}_i^2
\]
Bootstrap with dynamic regression models

Example: linear model, $y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \varepsilon_t$ with $H_0: \beta_1 = 0$.

2.1. Estimate the model under $H_0$, i.e. $y_t = \beta_0 + \beta_2 y_{t-1} + \eta_i$, and save $\{\hat{\eta}_1, \cdots, \hat{\eta}_n\}$ (estimated residuals from an AR(1))

2.2. Define $\tilde{\eta} = \{\tilde{\eta}_1, \cdots, \tilde{\eta}_n\}$ with $\tilde{\eta} = \sqrt{\frac{n}{n-2}} \hat{\eta}$

2.3. Draw (with replacement) residuals $\tilde{\eta}^{(b)} = \{\tilde{\eta}_1^{(b)}, \cdots, \tilde{\eta}_n^{(b)}\}$

2.4. Set (recursively) $y_t^{(b)} = \hat{\beta}_0 + \hat{\beta}_2 y_{t-1}^{(b)} + \tilde{\eta}_t^{(b)}$

2.5. Estimate the regression model $y_t^{(b)} = \beta_0^{(b)} + \beta_1^{(b)} x_t + \beta_2^{(b)} y_{t-1}^{(b)} + \varepsilon_t^{(b)}$

Remark: start (usually) with $y_0^{(b)} = y_1$
Bootstrap with heteroskedasticity

Example: linear model, \( y_i = \beta_0 + \beta_1 x_i + |x_i| \cdot \varepsilon_t \) with \( H_0 : \beta_1 = 0 \).

2.1. Estimate the model under \( H_0 \), i.e. \( y_i = \beta_0 + \eta_i \), and save \( \{ \hat{\eta}_1, \cdots, \hat{\eta}_n \} \)

2.2. Compute \( H_{i,i} \) with \( H_i = [H_{i,i}] \) from \( H = X[X^T X]^{-1} X^T \).

2.3.a. Define \( \tilde{\eta} = \{ \tilde{\eta}_1, \cdots, \tilde{\eta}_n \} \) with \( \tilde{\eta}_i = \pm \frac{\hat{\eta}_i}{\sqrt{1 - H_{i,i}}} \)

(here \( \pm \) mean \( \{ -1, +1 \} \) with probabilities \( \{ 1/2, 1/2 \} \))

2.4.a. Draw (with replacement) residuals \( \tilde{\eta}_i^{(b)} = \{ \tilde{\eta}_1^{(b)}, \cdots, \tilde{\eta}_n^{(b)} \} \)

2.5.a. Set \( y_i^{(b)} = \hat{\beta}_0 + \hat{\beta}_2 y_{i-1}^{(b)} + \tilde{\eta}_i^{(b)} \)

2.6.a. Estimate the regression model \( y_i^{(b)} = \beta_0^{(b)} + \beta_1^{(b)} x_i + \varepsilon_i^{(b)} \)

This was suggested in Liu (1988) Bootstrap procedures under some non-i.i.d. models
Bootstrap with heteroskedasticity

Example: linear model, $y_i = \beta_0 + \beta_1 x_i + |x_i| \cdot \varepsilon_t$ with $H_0 : \beta_1 = 0$.

2.1. Estimate the model under $H_0$, i.e. $y_i = \beta_0 + \eta_i$, and save $\{\hat{\eta}_1, \ldots, \hat{\eta}_n\}$

2.2. Compute $H_{i,i}$ with $H = [H_{i,i}]$ from $H = X[X^TX]^{-1}X^T$.

2.3.b. Define $\tilde{\eta} = \{\tilde{\eta}_1, \ldots, \tilde{\eta}_n\}$ with $\tilde{\eta}_i = \xi_i \frac{\hat{\eta}_i}{\sqrt{1 - H_{i,i}}}$

(here $\xi_i$ takes values $\left\{\frac{1 - \sqrt{5}}{2}, \frac{1 + \sqrt{5}}{2}\right\}$ with probabilities $\left\{\frac{\sqrt{5} + 1}{2\sqrt{5}}, \frac{\sqrt{5} - 1}{2\sqrt{5}}\right\}$)

2.4.b. Draw (with replacement) residuals $\tilde{\eta}^{(b)} = \{\tilde{\eta}_1^{(b)}, \ldots, \tilde{\eta}_n^{(b)}\}$

2.5.b. Set $y_i^{(b)} = \hat{\beta}_0 + \hat{\beta}_2 y_{i-1}^{(b)} + \tilde{\eta}_i^{(b)}$

2.6.b. Estimate the regression model $y_i^{(b)} = \beta_0^{(b)} + \beta_1^{(b)} x_i + \varepsilon_i^{(b)}$

This was suggested in Mammen (1993) Bootstrap and wild bootstrap for high dimensional linear models, $\xi_i$'s satisfy here $\mathbb{E}[\xi_i^3] = 1$
Bootstrap with heteroskedasticity

Application $y_i = \beta_0 + \beta_1 x_i + |x_i| \cdot \varepsilon_i$, $\varepsilon \sim \mathcal{N}(0, 1)$, $\varepsilon \sim \mathcal{U}([-1, +1])$ or $\varepsilon \sim \mathcal{Std}(\nu = 2)$. 
Bootstrap with 2SLS: Wild Bootstrap

Consider a linear model, \( y_i = x_i^T \beta + \varepsilon_i \) where \( x_i = z_i^T \gamma + u_i \).

Two-stage least squares:

1. regress each column of \( x \) on \( z \), \( \hat{\gamma} = [Z^T Z] Z^T X \) and consider the predicted value

\[
\hat{X} = Z \hat{\gamma} = Z [Z^T Z] Z^T X
\]

2. regress \( y \) on predicted covariates \( \hat{X} \), \( y_i = \hat{x}_i^T \beta + \varepsilon_i \)
Bootstrap with 2SLS: Wild Bootstrap

Example: linear model, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_t$ where $x_i = z_i^T \gamma + u_i$ and $
abla \varepsilon, u \nabla = \rho$, with $H_0 : \beta_1 = 0$.

So called Wild Boostrap, see Davidson & Mackinnon (2009) *Wild bootstrap tests for IV regression*

2.1. Estimate the model under $H_0$, i.e. $y_i = \beta_0 + \eta_i$, by 2SLS and save
$
\hat{u} = \{\hat{\eta}_1, \ldots, \hat{\eta}_n\}
$

2.2. Estimate $\gamma$ from $x_i = z_i^T \gamma + \delta \hat{\eta}_i + u_i$

2.3. Define $\tilde{u} = \{\tilde{u}_1, \ldots, \tilde{u}_n\}$ with $\tilde{u}_i = X_i - z_i^T \hat{\gamma}$

2.4. Draw (with replacement) pairs of residuals $(\hat{\eta}_i^{(b)}, \tilde{u}_i^{(b)})$ of $(\eta_i^{(b)}, \tilde{u}_i^{(b)})$’s

2.5. Set $x_i^{(b)} = z_i^T \hat{\gamma} + \tilde{u}_i^{(b)}$ and $y_i^{(b)} = \hat{\beta}_0 + \hat{\eta}_i^{(b)}$

2.6. Estimate (using 2SLS) the regression model $y_i^{(b)} = \beta_0^{(b)} + \beta_1^{(b)} x_i^{(b)} + \varepsilon_i^{(b)}$, where $x_i^{(b)} = z_i^T \gamma + u_i$
Bootstrap with 2SLS: Wild Bootstrap

See example Section 5.2 in Horowitz (1998) The Bootstrap.
Estimation of Various Quantities of Interest

Consider a quadratic model,

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \]

The minimum is obtained in \( \theta = -\beta_1 / 2\beta_2 \).

What could be the standard error for \( \theta \)?

1. Use of the Delta-Method

\[ \theta = g(\beta_1, \beta_2) = \frac{-\beta_1}{2\beta_2} \]

Since \( \frac{\partial \theta}{\partial \beta_1} = -\frac{1}{2\beta_2} \) and \( \frac{\partial \theta}{\partial \beta_2} = \frac{\beta_1}{2\beta_2^2} \), the variance is

\[
\frac{1}{4} \left[ \frac{-1}{2\beta_2} \frac{\beta_1}{2\beta_2^2} \right] \left[ \begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right] \left[ \begin{array}{c} -1 \\ \frac{\beta_1}{2\beta_2} \end{array} \right]^T = \frac{\sigma_1^2 \beta_2^2 - 2\beta_1 \beta_2 \sigma_{12} + \beta_1^2 \sigma_2^2}{4\beta_2^2}
\]
Estimation of Various Quantifies of Interest

2. Use of Bootstrap

standard deviation of \( \hat{\theta} \),
- delta method vs.
- bootstrap.
**Box-Cox Transform**

\[ y_\lambda = \beta_0 + \beta_1 x + \varepsilon, \text{ with } y_\lambda = \frac{y^{\lambda-1}}{\lambda} \]

with the limiting case \( y_0 = \log[y] \).

We assume that for some (unknown) \( \lambda_0 \), \( \varepsilon \sim \mathcal{N}(0, \sigma^2) \).

As in Horowitz (1998) *The Bootstrap*, use residual bootstrap:

\[ y_i^{(b)} = \left( \lambda[\hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\varepsilon}^{(b)}] \right)^{1/\lambda} \]
Kernel based Regression

Consider some kernel based regression of estimate \( m(x) = \mathbb{E}[Y|X = x] \),

\[
\hat{m}_h(x) = \frac{1}{n h \hat{f}_n(x)} \sum_{i=1}^{n} y_i k \left( \frac{x - x_i}{h} \right) \quad \text{where} \quad \hat{f}_n(x) = \frac{1}{n h} \sum_{i=1}^{n} k \left( \frac{x - x_i}{h} \right)
\]

We have seen that the bias was

\[
b_h(x) = \mathbb{E} [\hat{m}(x)] - m(x) \propto h^2 \left( \frac{1}{2} m''(x) + m'(x) \frac{f'(x)}{f(x)} \right)
\]

and the variance

\[
v_h(x) \propto \frac{\text{Var}[Y|X = x]}{nh f(x)}
\]

Further

\[
Z_n(x) = \frac{\hat{m}_n(x) - m(x) - b_n(x)}{\sqrt{v_h h_n(x)}} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \ n \to \infty.
\]
Kernel based Regression

Idea: convert $Z_n(x)$ into an asymptotically pivotal statistic

Observe that

$$
\hat{m}_h(x) - m(x) \sim \frac{1}{nhf(x)} \sum_{i=1}^{n} [y_i - m(x)]k\left(\frac{x - x_i}{h}\right)
$$

so that $v_n(x)$ can be estimated by

$$
\hat{v}_n(x) = \frac{1}{(nh\hat{f}(x))^2} \sum_{i=1}^{n} [y_i - \hat{m}_h(x)]^2 k\left(\frac{x - x_i}{h}\right)^2
$$

then set

$$
\hat{\theta} = \frac{\hat{m}_h(x) - m(x)}{\sqrt{\hat{v}_n(x)}}
$$

$\hat{\theta}$ is asymptotically $\mathcal{N}(0, 1)$ and it is an asymptotically pivotal statistic
## Poisson Regression

<table>
<thead>
<tr>
<th>Diagnosis period</th>
<th>Reporting-delay interval (quarters):</th>
<th>Total reports to end of 1992</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Quarter</td>
<td>0†</td>
</tr>
<tr>
<td>1988</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>1989</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>1990</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>44</td>
</tr>
<tr>
<td>1991</td>
<td>1</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>63</td>
</tr>
<tr>
<td>1992</td>
<td>1</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>67</td>
</tr>
</tbody>
</table>

Reporting delay can be important

Let $j$ denote year and $k$ denote delay. Assumption

$$N_{j,k} \sim \mathcal{P}(\lambda_{j,k}) \text{ with } \lambda_{j,k} = \exp[\alpha_j + \beta_k]$$

Unreported diagnoses for period $j$:

$$\sum_{k \text{ unobserved}} \lambda_{j,k}$$

Prediction:

$$\sum_{k \text{ unobserved}} \hat{\lambda}_{j,k} = \exp[\hat{\alpha}_j] \sum_{k \text{ unobserved}} \exp[\hat{\beta}_k]$$

Poisson regression is a GLM: confidence intervals on coefficients are asymptotic.

Let $V$ denote the variance function, then Pearson residuals are

$$\hat{\epsilon}_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V[\hat{\mu}_i]}}$$

so here

$$\hat{\epsilon}_{j,k} = \frac{n_{j,k} - \hat{\lambda}_{j,k}}{\sqrt{\hat{\lambda}_{j,k}}}$$
Poisson Regression

So bootstrapped responses are $n_{j,k}^* = \hat{\lambda}_{j,k} + \sqrt{\hat{\lambda}_{j,k} \cdot \hat{\epsilon}_{j,k}}$.
Pivotal Case (or not)

In some cases, $G(\cdot, F)$ does not depend on $F$, $\forall F \in \mathcal{F}$.

Then $T_n$ is said to be pivotal, relative to $\mathcal{F}$.

Example: consider the case of Gaussian residuals, $\mathcal{F} = \mathcal{F}_{\text{gaussian}}$. Then

$$T = \frac{\bar{y} - \mathbb{E}[Y]}{\hat{\sigma}} \sim \text{Std}(n-1)$$

which does not depend on $F$ (but it does depend on $\mathcal{F}$).

If $T_n$ is not pivotal, it is still possible to look for bounds on $G_n(t, F)$,

$$B_n(t) = \left[ \inf_{F \in \mathcal{F}_*} \{G_n(t, F)\}; \sup_{F \in \mathcal{F}_*} \{G_n(t, F)\} \right]$$

for instance, when a set of reasonable values for $\mathcal{F}_*$ is provided, by an expert.
Pivotal Case (or not)

\[ B_n(t) = \left[ \inf_{F \in \mathcal{F}_*} \{ G_n(t, F) \}; \sup_{F \in \mathcal{F}_*} \{ G_n(t, F) \} \right] \]

In the parametric case, set

\[ \mathcal{F}_* = \{ F_\theta, \theta \in IC \} \]

where \( IC \) is some confidence interval.

In the nonparametric case, use Kolomogorov-Smirnov statistics to get bounds, using quantiles of

\[ \sqrt{n} \sup\{ |\hat{F}_n(t) - F_0(t)| \} \]
Pivotal Function and Studentized Statistics

It is interesting to studentize any statistics.

Let $v$ denote the variance of $\hat{\theta}$ (computed using $\{y_1, \cdots, y_n\}$. Then set

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{v}}$$

If quantiles of $Z$ are known (and denoted $z_\alpha$), then

$$P\left(\hat{\theta} + \sqrt{v}z_{\alpha/2} \leq \theta \leq \hat{\theta} + \sqrt{v}z_{1-\alpha/2}\right) = 1 - \alpha$$

Idea: use a (double) bootstrap procedure
Pivotal Function and Double Bootstrap Procedure

1. Generate a bootstrap sample $y^{(b)} = \{y_1^{(b)}, \cdots, y_n^{(b)}\}$
2. Compute $\hat{\theta}^{(b)}$
3. From $y^{(b)}$ generate $\beta$ bootstrap sample, and compute $\{\hat{\theta}_1^{(b)}, \cdots, \hat{\theta}_\beta^{(b)}\}$
4. Compute $\hat{v}^{(b)} = \frac{1}{\beta} \sum_{j=1}^{\beta} (\hat{\theta}_j^{(b)} - \hat{\theta}^{(b)})^2$
5. Set $z^{(b)} = \frac{\hat{\theta}^{(b)} - \hat{\theta}}{\sqrt{\hat{v}^{(b)}}}$

Then use $\{z^{(1)}, \cdots, z^{(B)}\}$ to estimate the distribution of $z$’s (and some quantiles).

$$\mathbb{P}\left(\hat{\theta} + \sqrt{v} z_{\alpha/2}^{(B)} \leq \theta \leq \hat{\theta} + \sqrt{v} z_{1-\alpha/2}^{(B)}\right) = 1 - \alpha$$
Why should we studentize?

Here $Z \xrightarrow{L} \mathcal{N}(0,1)$ as $n \to \infty$ (CLT). Using Edgeworth series,

$$\mathbb{P}[Z \leq z|F] = \Phi(z) + n^{-1/2}p(z)\varphi(z) + O(n^{-1})$$

for some quadratic polynomial $p(\cdot)$. For $Z^{(b)}$

$$\mathbb{P}[Z^{(b)} \leq z|\hat{F}] = \Phi(z) + n^{-1/2}\hat{p}(z)\varphi(z) + O(n^{-1})$$

where $\hat{p}(z) = p(z) + O(n^{-1/2})$, so

$$\mathbb{P}[Z \leq z|F] - \mathbb{P}[Z^{(b)} \leq z|\hat{F}] = O(n^{-1})$$

But if we do not studentize, $Z = (\hat{\theta} - \theta) \xrightarrow{L} \mathcal{N}(0,\nu)$ as $n \to \infty$ (CLT).
Why should we studentize?

Using Edgeworth series,

\[ P[Z \leq z | F] = \Phi \left( \frac{z}{\sqrt{\nu}} \right) + n^{-1/2} p' \left( \frac{z}{\sqrt{\nu}} \right) \varphi \left( \frac{z}{\sqrt{\nu}} \right) + O(n^{-1}) \]

for some quadratic polynomial \( p(\cdot) \). For \( Z^{(b)} \)

\[ P[Z^{(b)} \leq z | \hat{F}] = \Phi \left( \frac{z}{\sqrt{\hat{\nu}}} \right) + n^{-1/2} \hat{p}' \left( \frac{z}{\sqrt{\hat{\nu}}} \right) \varphi \left( \frac{z}{\sqrt{\hat{\nu}}} \right) + O(n^{-1}) \]

recall that \( \hat{\nu} = \nu + O(n^{-1/2}) \), and thus

\[ P[Z \leq z | F] - P[Z^{(b)} \leq z | \hat{F}] = O(n^{-1/2}) \]

Hence, studentization reduces error, from \( O(n^{-1/2}) \) to \( O(n^{-1}) \)
Variance estimation

The estimation of Var[\hat{\theta}] is necessary for studentized bootstrap.

- double bootstrap (used here)
- delta method
- jackknife (leave-one-out)

Double Bootstrap

Requieres $B \times \beta$ resamples, e.g. $B \sim 1,000$ while $\beta \sim 100$

Delta Method

Let $\hat{r} = g(\hat{\theta})$, with $g'(\theta) \neq 0$.

$$\mathbb{E}[\hat{r}] = g(\theta) + O(n^{-1})$$

$$Var[\hat{r}] = Var[\hat{\theta}]g'(\theta)^2 + O(n^{-3/2})$$
**Variance estimation**

Idea: find a transformation such that $\text{Var}[\hat{\tau}]$ is constant. Then

$$\text{Var}[\hat{\theta}] \sim \frac{\text{Var}[\hat{\tau}]}{g'(\hat{\theta})^2}$$

There is also a nonparametric delta method, based on the influence function.
Influence Function and Taylor Expansion

Taylor expansion

\[ t(y) = t(x) + \int_x^y f'(z) dz + t(x) + (y - x) f'(x) \]

\[ t(G) = t(F) + \int_{\mathbb{R}} L_t(z, F)dG(z) \]

where \( L_t \) is the Fréchet derivative,

\[ L_t(z, F) = \left. \frac{\partial[(1 - \epsilon)F + \epsilon \Delta_z]}{\partial \epsilon} \right|_{\epsilon=0} \]

where \( \Delta_z(t) = 1(t > z) \) denote the cdf of the Dirac measure in \( z \).

For instance, observe that

\[ t(\hat{F}_n) = t(F) + \frac{1}{n} \sum_{i=1}^{n} L_t(y_i, F) \]
Influence Function and Taylor Expansion

This can be used to estimate the variance. Set

\[ V_L = \frac{1}{n^2} \sum_{i=1}^{n} L(y_i, F)^2 \]

where \( L(y, F) \) is the influence function for \( \theta = t(F) \) for observation at \( y \) when distribution is \( F \).

The empirical version is \( \ell_i = L(y_i, \hat{F}) \) and set

\[ \hat{V}_L = \frac{1}{n^2} \sum_{i=1}^{n} \ell_i^2 \]

**Example**: let \( \theta = \mathbb{E}[X] \) with \( X \sim F \), then

\[ \hat{\theta} = \bar{y}_n = \sum_{i=1}^{n} \frac{1}{n} y_i = \sum_{i=1}^{n} \omega_i y_i \text{ where } \omega_i = \frac{1}{n} \]
Influence Function and Taylor Expansion

Change \( \omega \)'s in direction \( j \):

\[
\omega_j = \epsilon + \frac{1 - \epsilon}{n}, \quad \text{while } \forall i \neq j, \omega_i = \frac{1 - \epsilon}{n},
\]

then \( \hat{\theta} \) changes in

\[
\ell_j = \left[ y_h - \hat{\theta} \right] \epsilon + \hat{\theta}
\]

Hence, \( \ell_j \) is the standardized chance in \( \hat{\theta} \) with an increase in direction \( j \), and

\[
\hat{V}_L = \frac{n - 1}{n} \frac{\text{Var}[X]}{n}.
\]

Example: consider a ratio, \( \theta = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]} \), then

\[
\hat{\theta} = \frac{\bar{x}_n}{\bar{y}_n} \quad \text{and} \quad \ell_j = \frac{x_j - \hat{\theta} y_j}{\bar{y}_n}
\]
Influence Function and Taylor Expansion

so that

$$
\hat{V}_L = \frac{1}{n^2} \sum_{i=1}^{n} \left( x_j - \frac{\hat{\theta}_j y_j}{\bar{y}_n} \right)^2
$$

Example: consider a correlation coefficient,

$$
\theta = \frac{\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]}{\sqrt{(\mathbb{E}[X^2] - \mathbb{E}[X]^2) \cdot (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2)}}
$$

Let $\bar{xy} = n^{-1} \sum x_i y_i$, so that

$$
\hat{\theta} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\bar{x}^2 - \bar{x}^2) \cdot (\bar{y}^2 - \bar{y}^2)}}
$$
Jackknife

An approximation of $\ell_i$ is $\ell_i^* = (n - 1)(\hat{\theta} - \hat{\theta}_{(-j)})$ where $\hat{\theta}_{(-j)}$ is the statistics computed from sample $\{y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n\}$

One can define Jackknife bias and Jackknife variance

$$b^* = \frac{-1}{n} \sum_{i=1}^{n} \ell_i^* \text{ and } v^* = \frac{1}{n(n-1)} \left( \sum_{i=1}^{n} \ell_i^{*2} - nb^2 \right)$$

cf numerical differentiation when $\epsilon = -\frac{1}{(n - 1)}$. 
Convergence

Given a sample \( \{y_1, \cdots, y_n\} \), i.i.d. with distribution \( F \), set

\[
\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1(y_i \leq t)
\]

Then

\[
\sup \left\{ |\hat{F}_n(t) - F_0(t)| \right\} \xrightarrow{\mathbb{P}} 0, \text{ as } n \to \infty.
\]
How many Bootstrap Samples?

Easy to take \( B \geq 5000 \)

\( R > 100 \) to estimate bias or variance

\( R > 1000 \) to estimate quantiles
How many Bootstrap Samples?

```r
> reg = lm(dist ~ speed, data = cars)
> E = residuals(reg)
> Y = predict(reg)
> beta1 = rep(NA, 2000)
> MB = MV = MQ = matrix(NA, 10, 2000)
> for (i in 1:10) {
+ BIAS = VAR = QUANT = beta1
+ for (b in 1:2000) {
+ carsS = data.frame(speed = cars$speed,
+ dist = Y + sample(E, size = 50, replace = TRUE))
+ beta1[b] = lm(dist ~ speed, data = carsS)$coefficients[2]
+ BIAS[b] = mean(beta1[1:b]) - reg$coefficients[2]
+ VAR[b] = var(beta1[1:b])
+ QUANT[b] = quantile(beta1[1:b], .95)
+ }
+ MB[i,] = BIAS
+ MV[i,] = VAR
+ MQ[i,] = QUANT
```
Consistency

We expect something like

\[ G_n(t, \hat{F}_n) \sim G_\infty(t, \hat{F}_n) \sim G_\infty(t, F_0) \sim G_n(t, F_0) \]

\( G_n(t, \hat{F}_n) \) is said to be consistent if under each \( F_0 \in \mathcal{F} \),

\[
\sup_{t \in \mathbb{R}} \left\{ \left| G_n(t, \hat{F}_n) - G_\infty(t, F_0) \right| \right\} \overset{P}{\rightarrow} 0
\]

Example: let \( \theta = \mathbb{E}_{F_0}(X) \) and consider \( T_n = \sqrt{n}(\bar{X} - \theta) \). Here

\[ G_n(t, F_0) = \mathbb{P}_{F_0}(T_n \leq t) \]

Based on bootstrap samples, a bootstrap version of \( T_n \) is

\[ T_n^{(b)} = \sqrt{n}(\bar{X}^{(b)} - \bar{X}) \text{ since } \bar{X} = \mathbb{E}_{\hat{F}_n}(X) \]

and \( G_n(t, \hat{F}_n) = \mathbb{P}_{\hat{F}_n}(T_n \leq t) \)
Consistency

Consider a regression model \( y_i = \mathbf{x}_i^T \beta + \varepsilon_i \)

The natural assumption is \( \mathbb{E}[\varepsilon_i | \mathbf{X}] = 0 \) with \( \varepsilon_i \)'s i.i.d. \( \sim F \).

The parameter of interest is \( \theta = \beta_j \), and let \( \hat{\beta}_j = \theta(F_n) \).

1. The statistics of interest is \( T_n = \sqrt{n}[\hat{\beta}_j - \beta_j] \).

We want to know \( G_n(t, F_0) = \mathbb{P}_{F_0}(T_n \leq t) \).

Let \( \mathbf{x}^{(b)} \) denote a bootstrap sample.

Compute \( T_{n}^{(b)} = \sqrt{n} (\hat{\beta}_{j}^{(b)} - \hat{\beta}_j) \), and then

\[
G_n(t, F_n) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}(T_{n}^{(b)} \leq t)
\]
Consistency

2. The statistics of interest is \( T_n = \sqrt{n} \frac{[\hat{\beta}_j - \beta_j]}{\sqrt{\text{Var}[\hat{\beta}_j]}} \).

We want to know \( G_n(t, F_0) = \mathbb{P}_{F_0}(T_n \leq t) \).

Let \( x^{(b)} \) denote a bootstrap sample.

Compute \( T_n^{(b)} = \sqrt{n} \frac{[\hat{\beta}_j^{(b)} - \hat{\beta}_j]}{\sqrt{\text{Var}^{(b)}[\hat{\beta}_j]}} \), and then

\[
G_n(t, F_n) = \frac{1}{B} \sum_{b=1}^{B} 1(T_n^{(b)} \leq t)
\]

This second option is more accurate than the first one:
**Consistency**

The approximation error of bootstrap applied to asymptotically pivotal statistic is smaller than the approximation error of bootstrap applied on asymptotically non-pivotal statistic, see Horowitz (1998) *The Bootstrap*.

Here, asymptotically pivotal means that

\[ G_\infty(t, F) = G_\infty(t), \forall F \in \mathcal{F}. \]

Assume now that the quantity of interest is \( \theta = \text{Var}[\hat{\beta}] \).

Consider a bootstrap procedure, then one can prove that

\[
\text{plim}_{B,n \to \infty} \left\{ \frac{1}{B} \sum_{b=1}^{B} \sqrt{n} (\beta^{(b)} - \hat{\beta}) \sqrt{n} (\beta^{(b)} - \hat{\beta})^T \right\} = \text{plim}_{n \to \infty} \left\{ n(\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)^T \right\}
\]
More on Testing Procedures

Consider a sample \( \{y_1, \cdots, y_n\} \). We want to test some hypothesis \( H_0 \). Consider some test statistic \( t(y) \)

Idea: \( t \) takes large values when \( H_0 \) is not satisfied.

The \( p \)-value is \( p = \mathbb{P}[T > t_{\text{obs}} | H_0] \).

Bootstrap/simulations can be used to estimate \( p \), by simulation from \( H_0 \).

1. Generate \( y^{(s)} = \{y_1^{(s)}, \cdots, y_n^{(s)}\} \) generated from \( H_0 \).
2. Compute \( t^{(s)} = t(y^{(s)}) \)
3. Set \( \hat{p} = \frac{1}{1 + S} \left( 1 + \sum_{s=1}^{S} 1(t^{(s)} \geq t_{\text{obs}}) \right) \)

**Example**: testing independence, let \( t \) denote the square of the correlation coefficient.

Under \( H_0 \) variables are independent, so we can bootstrap independently \( x \)'s and \( y \)'s.
With this bootstrap procedure, we estimate

$$\hat{p} = \mathbb{P}(T \geq t_{\text{obs}} | \hat{H}_0)$$

which is not the same as

$$p = \mathbb{P}(T \geq t_{\text{obs}} | H_0)$$
More on Testing Procedures

In a parametric model, it can be interesting to use a sufficient statistic $W$. One can prove that

$$p = \mathbb{P}(T \geq t_{obs} | \hat{H}_0, W)$$

The problem is to generate from this conditional distribution...

**Example** : for the independence test, we should sample from $\hat{F}_x$ and $\hat{F}_y$ with fixed margins.

Bootstrap should be here without replacement.
More on Testing Procedures
More on Testing Procedures

But this nonparametric bootstrap fails when the Gaussian Central Limit
Theorem does not apply (Mammen’s theorem)

Example $X \sim \text{Cauchy}$ : limit distribution $G_\infty(t, F)$ is not continuous, in $F$

Example : distribution of the maximum of the support (see Bickel & Freedman (1981)): $X \sim U([0, \theta_0])$

$T_n = n(\theta_n - \theta_0)$ with $\theta_n = \max\{X_1, \cdots, X_n\}$

Set $T_n^{(b)} = n(\theta_n^{(b)} - \theta_n)$, and $\theta_n^{(b)} = \max\{X_1^{(b)}, \cdots, X_n^{(b)}\}$

Here $T_n \xrightarrow{\mathcal{L}} \mathcal{E}(1)$, exponential distribution, but not $T_n^{(b)}$, since $T_n^{(b)} \geq 0$ (we just resample), and

\[ P[T_n^{(b)} = 0] = 1 - P[T_n^{(b)} > 0] = 1 - \left(1 - \frac{1}{n}\right)^n \sim 1 - e^{-1}. \]
Resampling or Subsampling?

Why not draw subsamples of size $m < n$?

- with replacement, see $m$ out of $n$ bootstrap
- without replacement, see subsampling bootstrap

Less accurate than bootstrap when bootstrap works... but might work when bootstrap does not work

Exemple: maximum of the support, $Y_i \sim \mathcal{U}([0, \theta])$,

$$
\mathbb{P}_{\hat{F}_n} [T_m^{(b)} = 0] = 1 - \left(1 - \frac{1}{n}\right)^m \sim 1 - e^{-m/n} \sim 0
$$

if $m = o(n)$. 

**From Bootstrap to Bagging**

Bagging was introduced in Breiman (1996) Bagging predictors

1. sample a bootstrap sample \((y_i^{(b)}, x_i^{(b)})\) by resampling pairs
2. estimate a model \(\hat{m}^{(b)}(\cdot)\)

The bagged estimate for \(m\) is then

\[
m_{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{m}^{(b)}(x)
\]

**From Bagging to Random Forests**