

Resampling Methods

Outline

- Introduction
- Permutation Test
- Bootstrapping Test
- Jackknife Test
- Other resampling methods
- Conclusion



Definition

Resampling - Any of a variety of methods for
E.E.V.

- **E**stimating the precision of sample statistics (medians, variances, percentiles)
- **E**xchanging labels on data points when performing significance tests
- **V**alidating models by using random subsets

Background

- The theory was introduced in the 1930s by R.A. Fisher & E.J.G. Pitman
- In 1966, Resampling method was first tried with graduate students
- In 1969, the method presented in the edition of *Basic Research Methods in Social Science* (3rd edition, Julian L. Simon and Paul Burstein, 1985).
- In the late 1970s, Efron began to publish formal analyses of the bootstrap—an important resampling application.
- Since 1970, Bootstrap method has been “hailed by an official *American Statistical Association* volume as the only “great breakthrough” in statistics” (Kotz and Johnson, 1992).
- In 1973 Dan Weidenfeld and Julian L. Simon developed the computer language called RESAMPLING STATS (earlier called SIMPLE STATS).

1930

In 1958, Tukey coined
the term Jackknife.

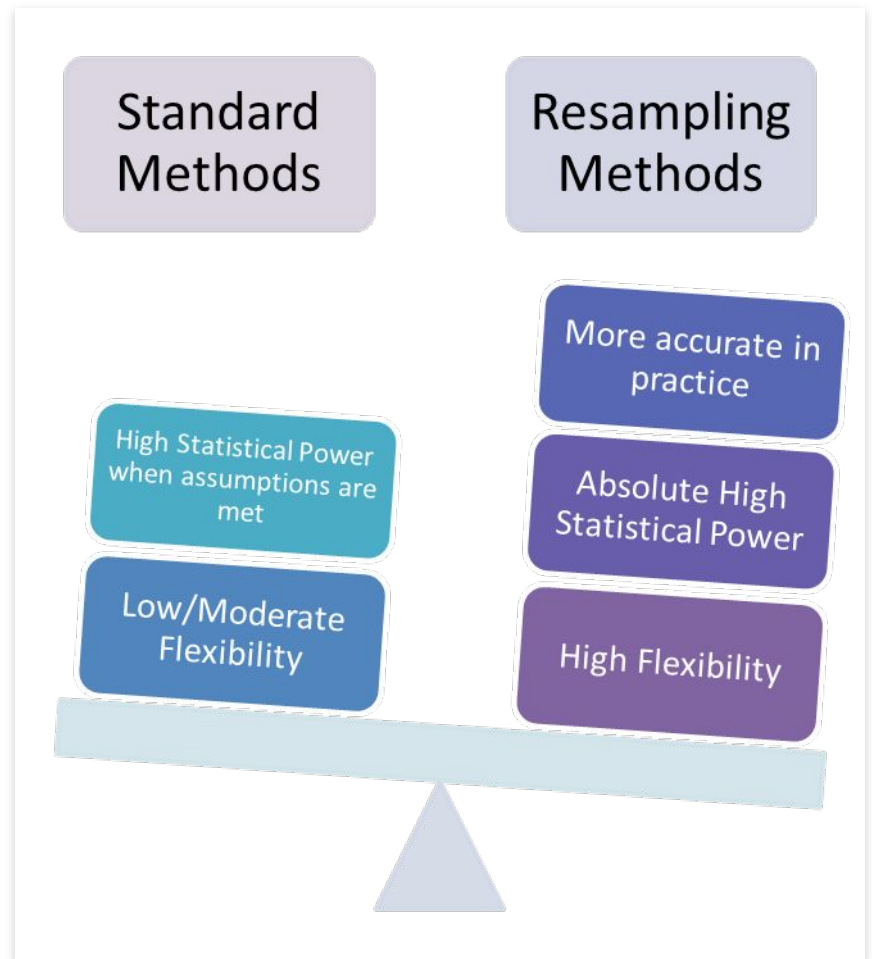
1970s

In 1956, Quenouille
suggested a resampling
technique.

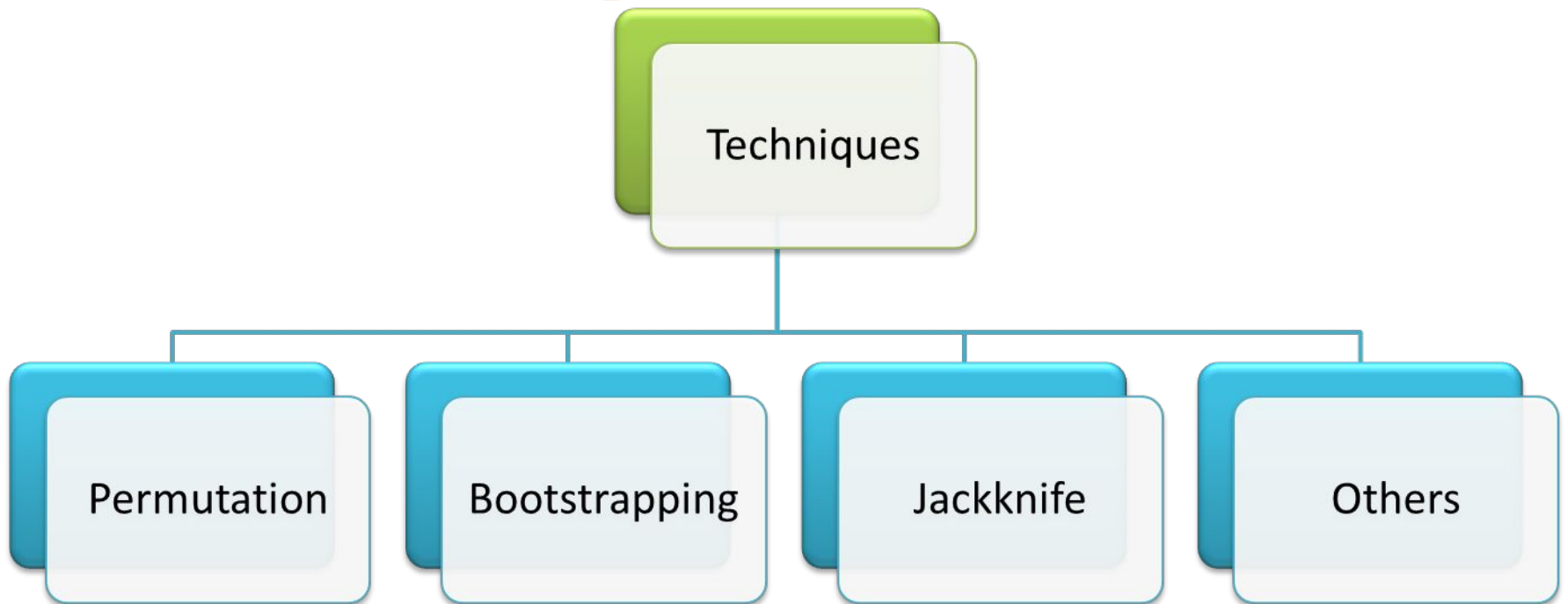
1960s

Why Resampling?

- In most cases, people accept assumptions for standard statistics “as if” they are satisfied.
- Some “awkward” but “interesting” statistics, that standard statistics fail to be applied to.
- Saves us from onerous formulas for different problems.
- More accurate in practice than standard methods.



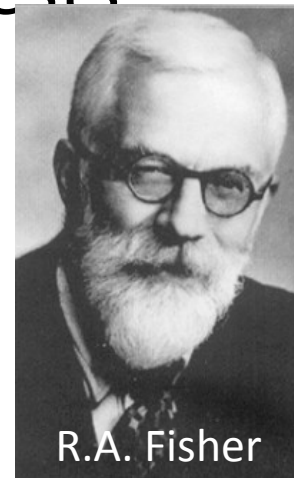
Resampling Techniques



Permutation Tests

1.1 What is Permutation Tests?

- **Permutation tests** are significance tests based on **permutation resamples** drawn at random from the original data. Permutation resamples are drawn **without replacement**.
- Also called randomization tests, re-randomization tests, exact tests
- Introduced by R.A. Fisher and E.J.G. Pitman in the 1930s.



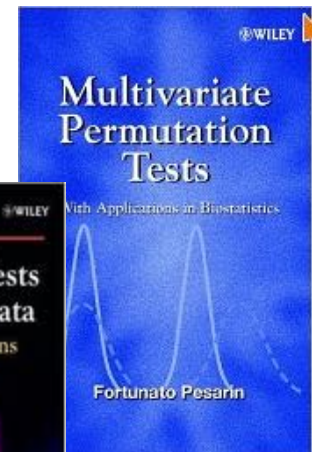
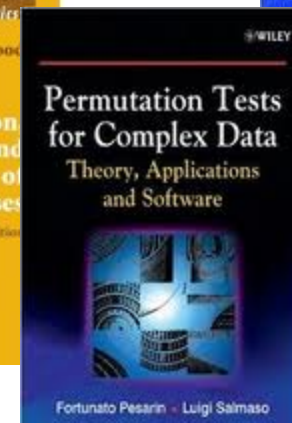
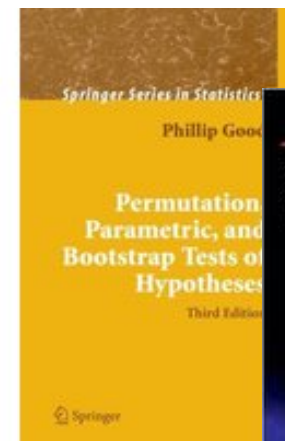
R.A. Fisher



E.J.G. Pitman

When Can We Use Permutation Tests?

- Only when we can see how to **resample** in a way that is **consistent** with the **study design** and with the **null hypothesis**.
- If we cannot do a permutation test, we can often calculate a **bootstrap confidence interval** instead.



Advantages



Exist for **any test statistic, regardless** of whether or not its **distribution is known**



Free to choose the statistic which best discriminates between hypothesis and alternative and which minimizes losses

Can be used for:

- Analyzing **unbalanced designs**;
- Combining **dependent tests** on mixtures of **categorical, ordinal, and metric data**.

Limitations



An Important Assumption

The **observations** are **exchangeable** under the null hypothesis



Consequence

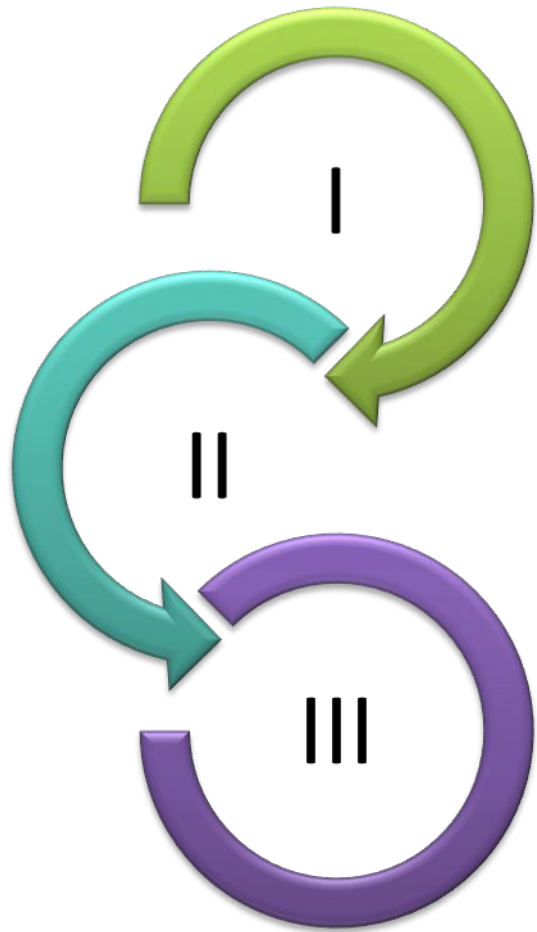
Tests of difference in location (like a permutation t-test) **require equal variance**



In this respect

The **permutation t-test** shares the **same weakness** as the **classical Student's t-test**.

Procedure of Permutation Tests



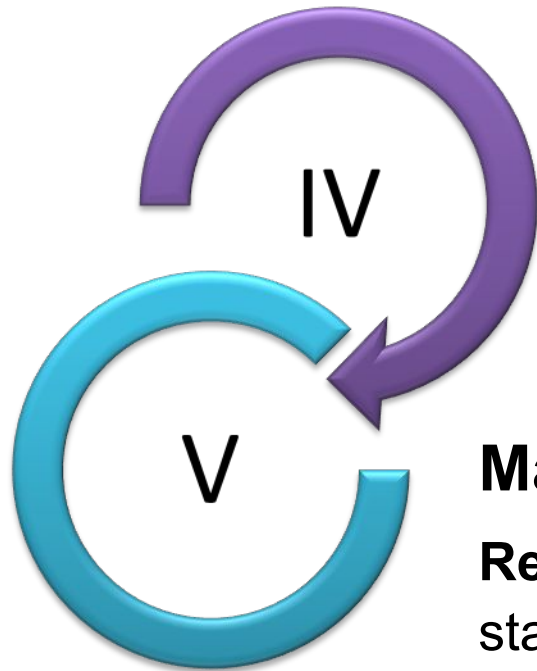
Analyze the problem.

- What is the hypothesis and alternative?
- What distribution is the data drawn from?
- What losses are associated with bad decisions?

Choose a test statistic which will distinguish the hypothesis from the alternative.

Compute the test statistic for the original data of the observations.

Procedure of Permutation Tests



Rearrange the Observations

Compute the test statistic for **all possible permutations** (rearrangements) of the data of the observations

Make a decision

Reject the Hypothesis: if the value of the test statistic for the original data is an **extreme value** in the permutation distribution of the statistic.

Otherwise, accept the hypothesis and reject the alternative.

Permutation Resampling Process

Collect Data from Control & Treatment Groups

5 7 8 4 1

6 8 9 7 5 9

Merge Samples to form a pseudo population

5 7 8 4 1 6 8 9 7 5 9

Sample without replacement from pseudo population to simulate Control and Treatment Groups

5 7 1
5 9

8 4 6
8 9 7

Compute target statistic for each resample

Median (5 7 1 5 9)

Median (8 4 6 8 9 7)

Compute “difference statistic”, save result in table and repeat resampling process 1000+ iterations

Example: “I Lost the Labels”



An physiology experiment to find the relationship between Vitamin E and human “life-extending”

Example: “I Lost the Labels”

- 6 petri dishes:
 - 3 containing standard medium
 - 3 containing standard medium + Vitamin E
- Without the labels, we have no way of knowing which cell cultures have been treated with Vitamin E and which have not.
- There are six numbers “121, 118, 110, 34, 12, 22”, each one belongs to the petri dishes’ results.
- The number belongs to which dishes?

Here is a simple sample:

- data

X			Y		
A	B	C	D	E	F
121	118	110	34	12	22
$\bar{x}_n = 116.33$			$\bar{y}_n = 22.67$		

- random samples $X = \{x_1, \dots, x_{n_x}\}$ and $Y = \{y_1, \dots, y_{n_y}\}$
- test statistic

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{(n_x - 1) + (n_y - 1)}}} \cdot \sqrt{\frac{n_x \cdot n_y}{n_x + n_y}}$$

(\bar{X} , \bar{Y} and S_x^2 , S_y^2 are sample means and sample variances, n_x , n_y sample sizes)

Using the original T-test to Find P-value

- under H_0 holds $T \sim t_{n_x+n_y-2}$ (t-distribution with $n_x + n_y - 2$ d.f.)
- compute the p -value for the observed value t of test statistic T
$$p = 1 - P(|T| \leq |t| | H_0) = 2[1 - P(T \leq |t| | H_0)] = 2[1 - F_{t, n_x+n_y-2}(|t|)]$$
- Decision rule: reject H_0 if $p\text{-value} \leq \alpha$

**T-test statistic: T=13.0875,
two side P-vale=0.0002**

Permutation Resampling Process

Collect Data from Control & Treatment Groups

121 118 110

34 12 22

Merge Samples to form a pseudo population

121 118 110 34 12 22

Sample without replacement from pseudo population to simulate Control and Treatment Groups

121 118
34

110 12
22

Compute target statistic for each resample

Median=91

Median=48

Compute “difference statistic”, save result in table and repeat resampling process 1000+ iterations

After one permutation

- after one permutation:

X			Y		
A	B	D	C	E	F
121	118	34	110	12	22
$\bar{x}_n = 91$			$\bar{y}_n = 48$		

- test statistic

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{(n_x-1) + (n_y-1)}}} \cdot \sqrt{\frac{n_x \cdot n_y}{n_x + n_y}}$$

T-test statistic: T=1.019

Formula in Permutation need

- Note: The number of permutations for comparing two groups of size m and $n - m$ is

$$C_m^n = \binom{n}{m} = \frac{n!}{m! \cdot (n - m)!}$$

- how many permutations exist?

$$C_3^6 = \binom{6}{3} = \frac{6!}{3! \cdot 3!} = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} = 20$$

All 20 Permutation data

permutation	X	Y	\bar{x}_n	\bar{y}_n	$\bar{x}_n - \bar{y}_n$	t
1	ABC	DEF	116.33	22.67	93.67	13.087
2	ABD	CEF	91.00	48.00	43.00	1.019
3	ABE	CDF	87.00	52.00	35.00	0.795
4	ABF	CDE	83.67	55.33	28.33	0.627
5	ACD	BEF	88.33	50.67	37.67	0.866
6	ACE	BDF	84.33	54.67	29.67	0.659
7	ACF	BDE	81.00	58.00	23.00	0.500
8	ADE	BCF	59.00	80.00	-21.00	-0.455
9	ADF	BCE	55.67	83.33	-27.67	-0.611
10	AEF	BCD	51.67	87.33	-35.67	-0.813
11	BCD	AEF	87.33	51.67	35.67	0.813
12	BCE	ADF	83.33	55.67	27.67	0.611
13	BCF	ADE	80.00	59.00	21.00	0.455
14	BDE	ACF	58.00	81.00	-23.00	-0.500
15	BDF	ACE	54.67	84.33	-29.67	-0.659
16	BEF	ACD	50.67	88.33	-37.67	-0.866
17	CDE	ABF	55.33	83.67	-28.33	-0.627
18	CDF	ABE	52.00	87.00	-35.00	-0.795
19	CEF	ABD	48.00	91.00	-43.00	-1.019
20	DEF	ABC	22.67	116.33	-93.67	-13.087

How is the conclusion

- Test decision The absolute value of the test statistic $t \geq |t| = 13.0875$ we obtained for the original labeling.
- We obtain the exact p value $p = 2/20 = 0.1$.
- Note: If both groups have equal size, Only half of permutations is really needed (symmetry)

- data capacitor;
Input group \$ failtime @@;
Datalines;
Control 121 control 118 control 110
Stressed 34 stressed 12 stressed 22
;
Proc multtest data=capacitor permutation nsample=25000
out=results outsamp=samp;
test mean(failtime /lower);
class group;
contrast 'a vs b' -1 1;
Run;
proc print data=samp(obs=18);
run;
proc print data=results;
run;

The Multtest Procedure

Model Information	
Test for continuous variables	Mean t-test
Degrees of Freedom Method	Pooled
Tails for continuous tests	Lower-tailed
Strata weights	None
P-value adjustment	Permutation
Center continuous variables	No
Number of resamples	25000
Seed	863732001

Contrast Coefficients					
Contrast		group			
		Control	Stressed	control	stressed
a vs b	Centered	-1	1	0	0

Continuous Variable Tabulations				
Variable	group	NumObs	Mean	Standard Deviation
failtime	Control	1	121.0000	0.0000
failtime	Stressed	1	34.0000	0.0000
failtime	control	2	114.0000	5.6569
failtime	stressed	2	17.0000	7.0711

p-Values			
Variable	Contrast	Raw	Permutation
failtime	a vs b	0.0053	0.0464

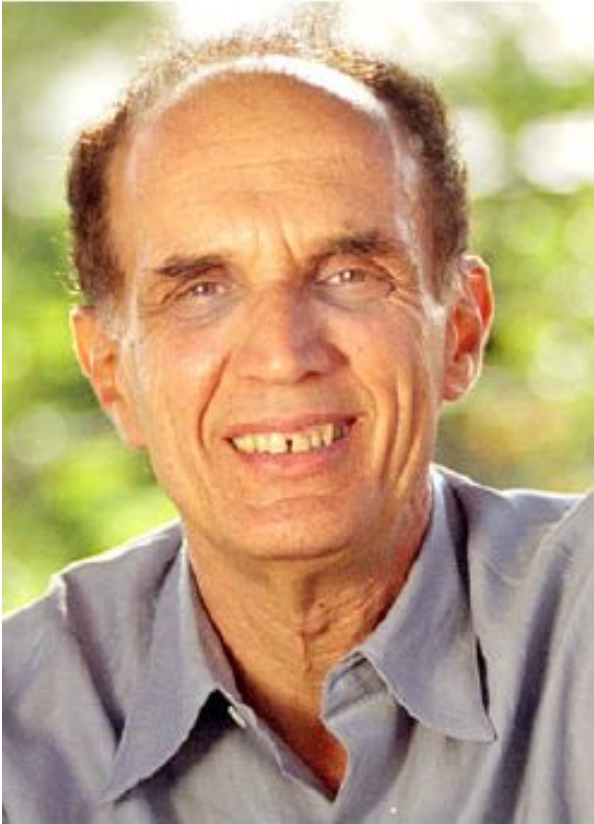
Obs	_sample_	_class_	_obs_	failtime
1	1	Control	2	118
2	1	Stressed	3	110
3	1	control	4	34
4	1	control	1	121
5	1	stressed	5	12
6	1	stressed	6	22
7	2	Control	3	110
8	2	Stressed	6	22

9	2	control	2	118
10	2	control	1	121
11	2	stressed	4	34
12	2	stressed	5	12
13	3	Control	6	22
14	3	Stressed	5	12
15	3	control	4	34
16	3	control	3	110
17	3	stressed	1	121
18	3	stressed	2	118

Obs	_test_	_var_	_contrast_	_value_	_se_	_nval_	raw_p	perm_p	sim_se
1	MEAN	failtime	a vs b	-522	54.3323	2	.005330365	0.04644	.001330914

BOOTSTRAP RESAMPLING

About the author



Bradley Efron

Professor of Statistics and of Health Research and Policy at Stanford University. He received the 2005 National Medal of Science, the highest scientific honor in USA.

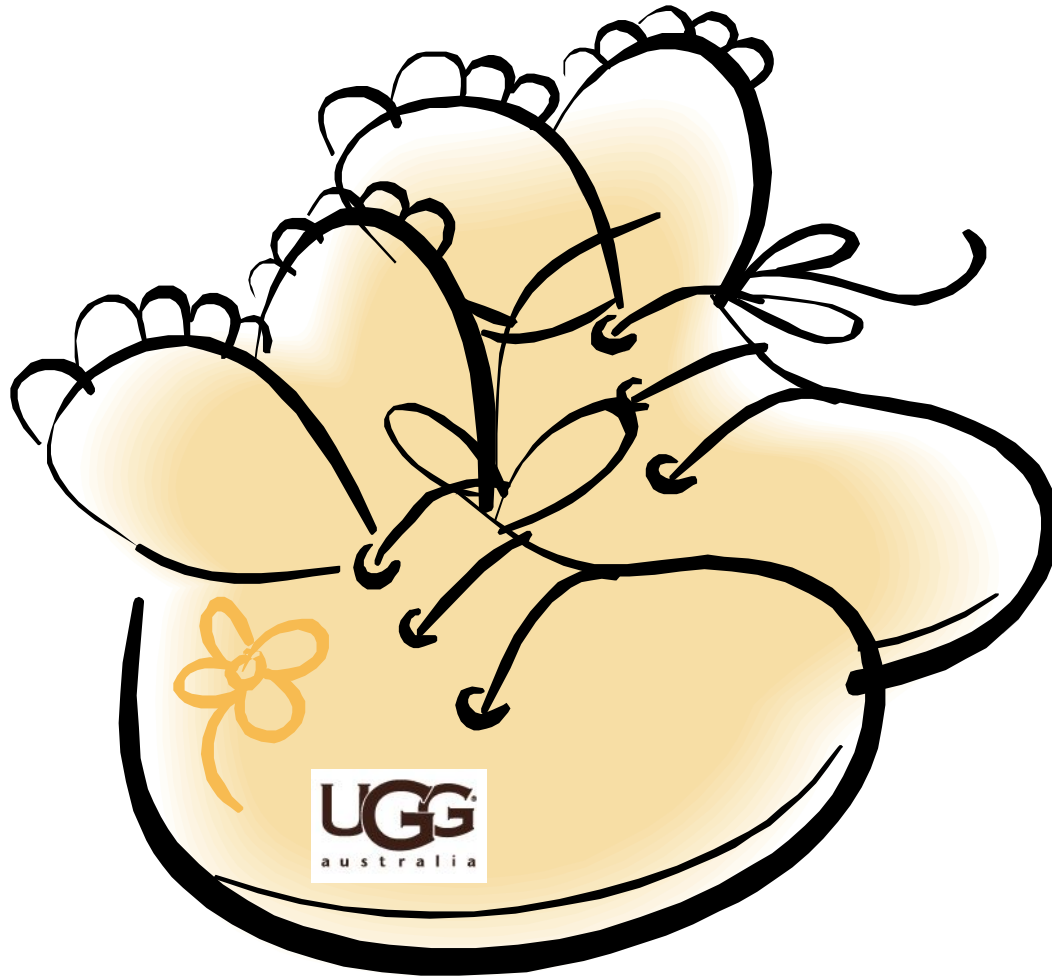
About the author



Rob Tibshirani

Associate Chairman and
Professor of Health
Research and Policy, and
Statistics at Stanford
University.

What is the bootstrap?



What is the bootstrap? In Statistics...

- ▶ Randomly sampling, with replacement, from an original dataset for use in obtaining statistical estimates.
- ▶ A data-based simulation method for statistical inference.
- ▶ A computer-based method for assigning measures of accuracy to statistical estimates.
- ▶ The method requires modern computer power to simplify intricate calculations of traditional statistical theory.

Why use the bootstrap?

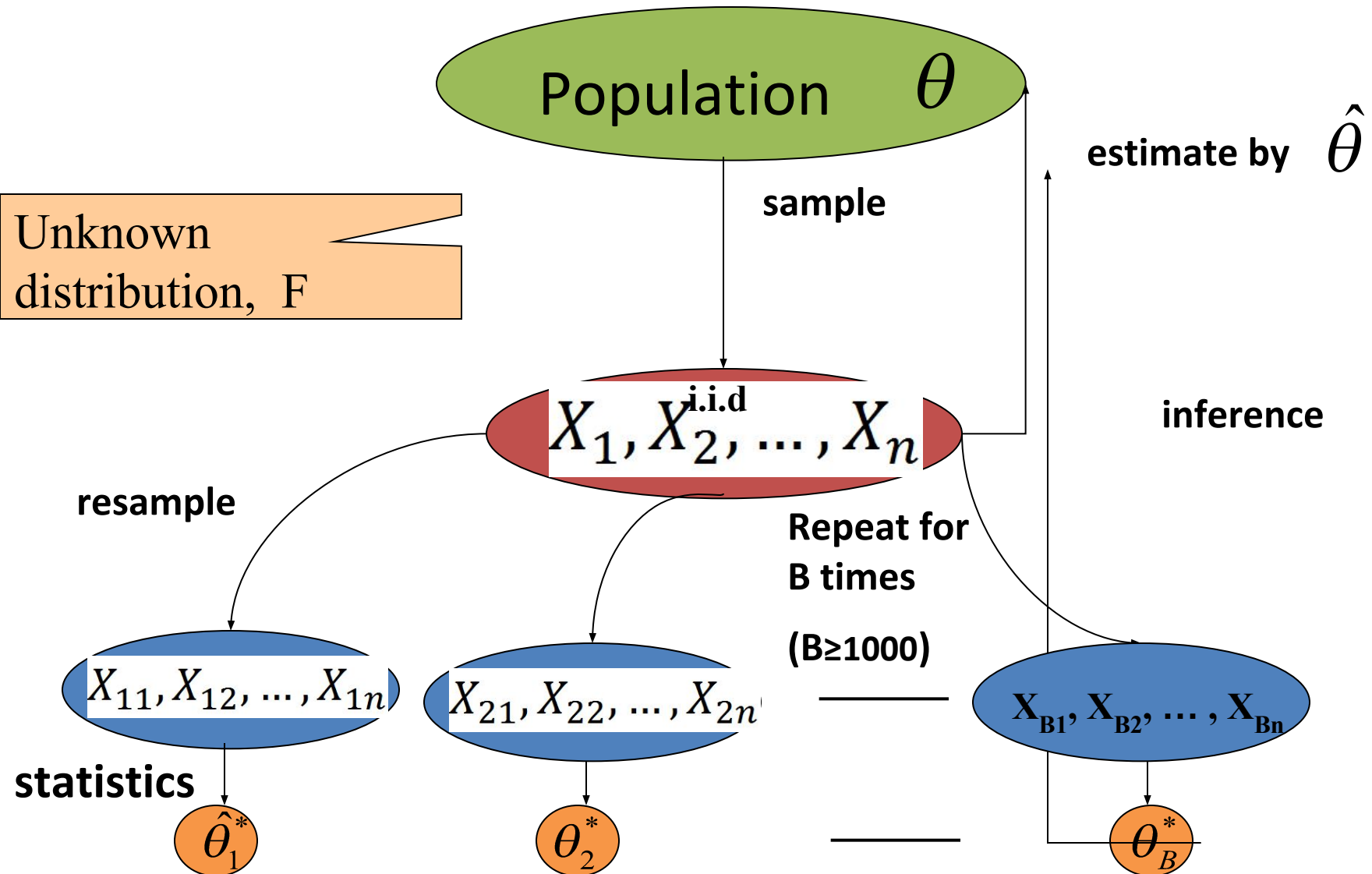
- ✓ Good question.
- ✓ Small sample size.
- ✓ Non-normal distribution of the sample.
- ✓ A test of means for two samples.
- ✓ Not as sensitive to N .

Bootstrap Idea

We **avoid** the task of taking many samples from the population by instead taking **many resamples from a single sample**. The values of *x* from these resamples form the *bootstrap distribution*. We use the bootstrap distribution rather than theory to learn about the sampling distribution.

- Bootstrap draws samples from the *Empirical Distribution* of data $\{X_1, X_2, \dots, X_n\}$ to replicate statistic to obtain its sampling distribution.
- The *Empirical Distribution* is just a *uniform* distribution over $\{X_1, X_2, \dots, X_n\}$. Therefore Bootstrap is just drawing i.i.d samples from $\{X_1, X_2, \dots, X_n\}$. The procedure is illustrated by the following graph.

The nonparametric of Bootstrap



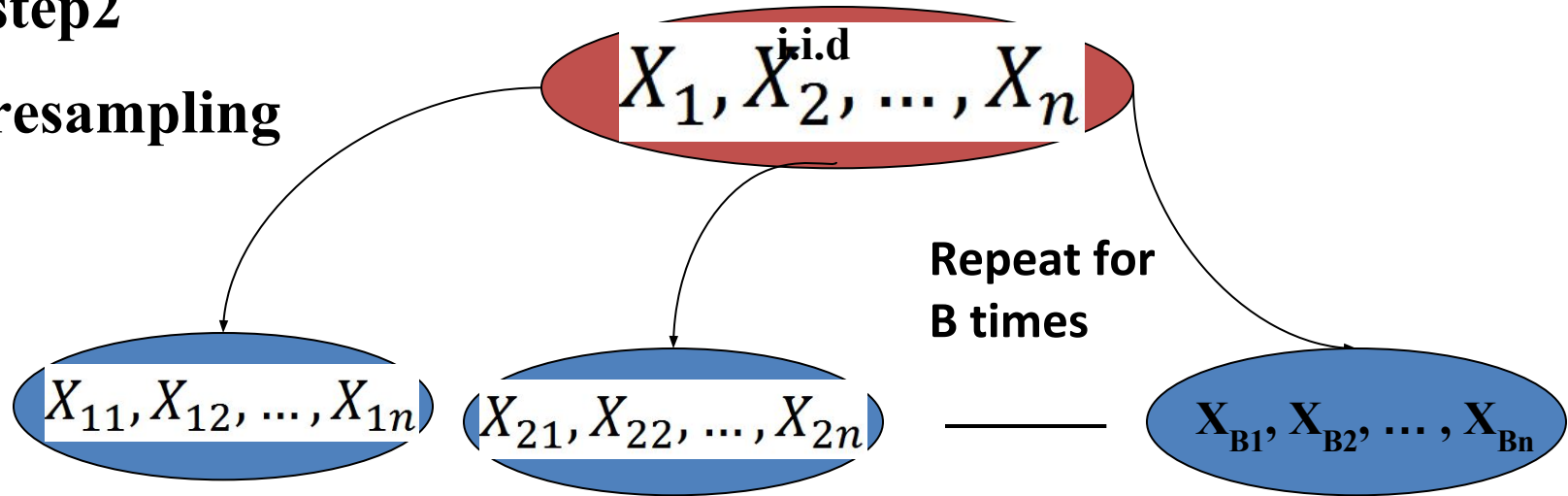
Population θ , with unknown
distribution F

Step1 **sampling**

X_1, X_2, \dots, X_n ^{i.i.d}

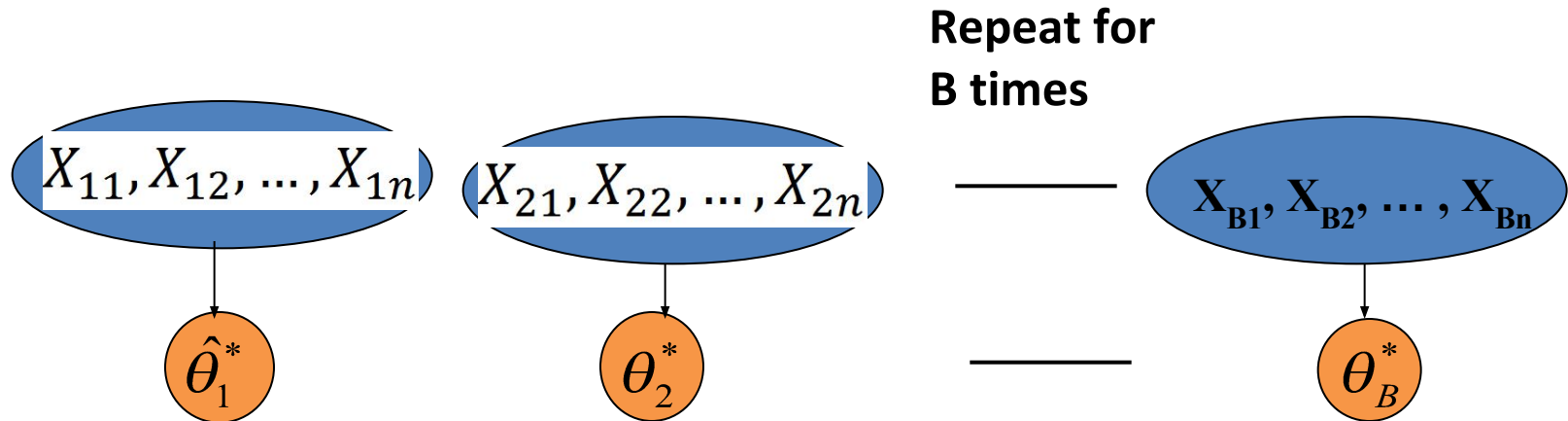
step2

resampling



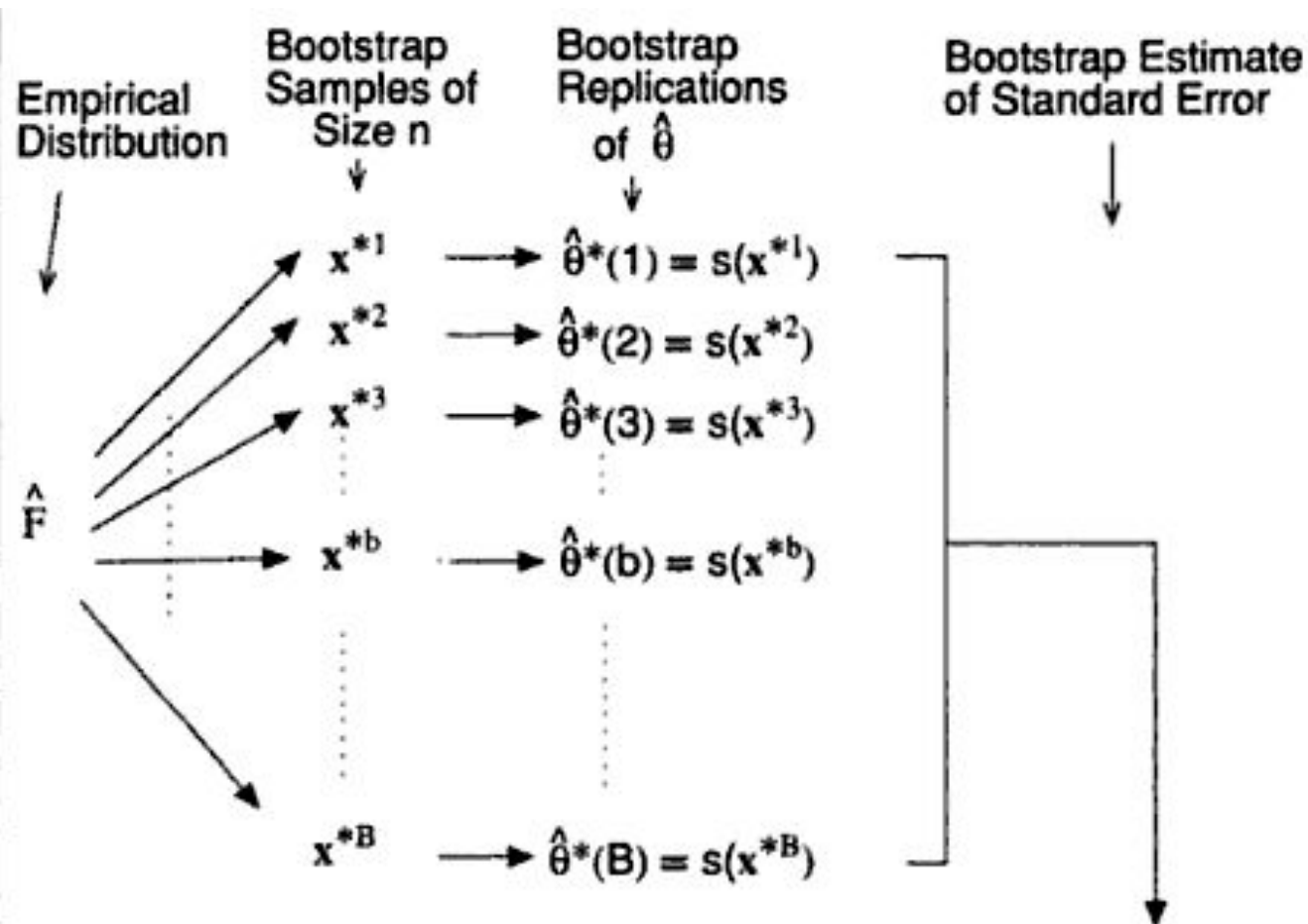
STEP 2: Resampling the data B times with replacement, then you can get many resampling data sets, and use this resampling data instead of real samples data from the population

Step 3: statistics



STEP3: Regard X_1, X_2, \dots, X_n as the new population and resample it B times with replacement, $X_{b1}, X_{b2}, \dots, X_{bn}$ where $i=1, 2, \dots, B$

Bootstrap for Estimating Standard Error of a statistic $s(x)$



$b = 1, 2, \dots, B$

$$\widehat{se}_B = \left[\frac{\sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}{B-1} \right]^{1/2}$$

$$\text{where } \hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}$$

BSE calculation (Continued)

- Bootstrap replicates $s(\mathbf{x}^{*1}), s(\mathbf{x}^{*2}), \dots, s(\mathbf{x}^{*B})$ are obtained by calculating the value of the statistic $s(\mathbf{x})$ on each bootstrap sample.
- The standard deviation of the values $s(\mathbf{x}^{*1}), s(\mathbf{x}^{*2}), \dots, s(\mathbf{x}^{*B})$ is the estimate of the standard error of $s(\mathbf{x})$.

Nonparametric confidence intervals for using Bootstrapping

- Many methods

The simplest : The percentile method

The percentile method

- 1) Construct \hat{F} , the empirical distribution function of the observed data. \hat{F} places probability $1/n$ on each observed data point X_1, X_2, \dots, X_n .
- 2) Draw a bootstrap sample $X_1^*, X_2^*, \dots, X_n^*$ of size n with replacement from \hat{F} .

Then calculate

$$\hat{\theta}^* = \hat{\theta}(X_1^*, X_2^*, \dots, X_n^*).$$

- 3) Repeat Step (2) a large number of times, say 1000, and then rank the values $\hat{\theta}^*$.

The percentile method (Continued)

For a 95% confidence interval, after ranking the bootstrapped theta coefficients, simply take the 2.5 % as the lower confidence limit and the 97.5% as the upper confidence limit.






























The percentile $(1-\alpha)$ 100% confidence interval for a population mean is:

$$(\hat{\theta}^*_{(\alpha/2)}, \hat{\theta}^*_{(1-\alpha/2)})$$

EXAMPLE

Example

Suppose we are interested in the wireless network download speed in the Stony Brook University. It is difficult for us to examine the entire population in the SBU, then the ideology of bootstrap resampling comes in. We take a population sample with 10 data sets, then we resample from the sample we have.

SPEEDTEST.NET 11/24/2010 4:51 AM GMT DOWNLOAD  9.15 Mb/s UPLOAD  8.03 Mb/s  PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 4:54 AM GMT DOWNLOAD  9.46 Mb/s UPLOAD  9.20 Mb/s  PING 4 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 4:20 AM GMT DOWNLOAD  9.25 Mb/s UPLOAD  10.93 Mb/s  PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:00 AM GMT DOWNLOAD  9.14 Mb/s UPLOAD  7.86 Mb/s  PING 5 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 5:08 AM GMT DOWNLOAD  31.94 Mb/s UPLOAD  45.90 Mb/s  PING 13 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:38 AM GMT DOWNLOAD  9.55 Mb/s UPLOAD  5.04 Mb/s  PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 5:02 AM GMT DOWNLOAD  20.69 Mb/s UPLOAD  5.77 Mb/s  PING 4 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:14 AM GMT DOWNLOAD  10.05 Mb/s UPLOAD  14.67 Mb/s  PING 5 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 5:04 AM GMT DOWNLOAD  9.19 Mb/s UPLOAD  7.33 Mb/s  PING 12 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:31 AM GMT DOWNLOAD  5.55 Mb/s UPLOAD  7.94 Mb/s  PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi

Population Sample (Mbps)

5.55, 9.14, 9.15, 9.19, 9.25, 9.46,
9.55, 10.05, 20.69, 31.94

5.55, 9.14, 9.15, 9.19,
9.19, 9.25, 9.25,
10.05, 10.05, 10.05

**Resample
#1**

9.14, 9.15, 9.19, 9.46,
9.46, 9.55, 10.05,
20.69, 20.69, 31.94

**Resample
#2**

5.55, 9.15, 9.15, 9.15,
9.25, 9.25, 9.25, 9.25,
9.46, 9.46

**Resample
#3**

⋮

Repeat for N times

5.55, 9.14, 9.15, 9.19,
9.19, 9.25, 9.25,
10.05, 10.05, 10.05

Resample #1

$$\overline{X_1^*} = 9.087$$

9.14, 9.15, 9.19, 9.46,
9.46, 9.55, 10.05,
20.69, 20.69, 31.94

Resample #2

$$\overline{X_2^*} = 13.932$$

5.55, 9.15, 9.15, 9.15,
9.25, 9.25, 9.25, 9.25,
9.46, 9.46

Resample #3

$$\overline{X_3^*} = 8.892$$

⋮

⋮

$X_{N1}, X_{N2}, X_{N3}, \dots, X_{Nn}$

Resample #N

$$\overline{X_B^*} = \frac{\sum_{i=1}^n X_{Bi}}{n}$$

SAS CODE

```
DATA ONE; /* This is the original data */  
INPUT DOWNLOAD @@;  
DATALINES;  
5.55 9.46 9.25 9.14 9.15 9.19 31.94 9.55 10.05 20.69  
;  
RUN;
```

```

DATA bootsamp;
DO sampnum = 1 to 1000;
/* Create 1000 bootstrap replications */
  DO i = 1 to nobs;
    /* For each replication sample nobs observations with
replacement from the data set ONE. Note the value of      nobs
(=10) is defined in the SET statement below. */
    x = CEIL(RANUNI(0) * nobs);
    /*The command RANUNI is used to generate random numbers
from a uniform distribution ranging from= 0 to 1*/
    SET ONE NOBS = nobs POINT = x;
    OUTPUT;
  END;
END;
STOP;
RUN;

```

```
PROC MEANS DATA=bootsamp noprint nway;  
CLASS sampnum;  
VAR DOWNLOAD;  
OUTPUT out=boot mean=mean var=var n=n;  
/* Save the variables mean, var and n  
into a new data set entitled boot. */  
RUN;  
PROC UNIVARIATE DATA=boot;  
HISTOGRAM mean;  
RUN;
```

The UNIVARIATE Procedure

Variable: mean

Moments			
N	1000	Sum Weights	1000
Mean	12.447439	Sum Observations	12447.439
Std Deviation	2.41351405	Variance	5.82505009
Skewness	0.65788125	Kurtosis	0.80826796
Uncorrected SS	160757.963	Corrected SS	5819.22504
Coeff Variation	19.3896436	Std Error Mean	0.07632202

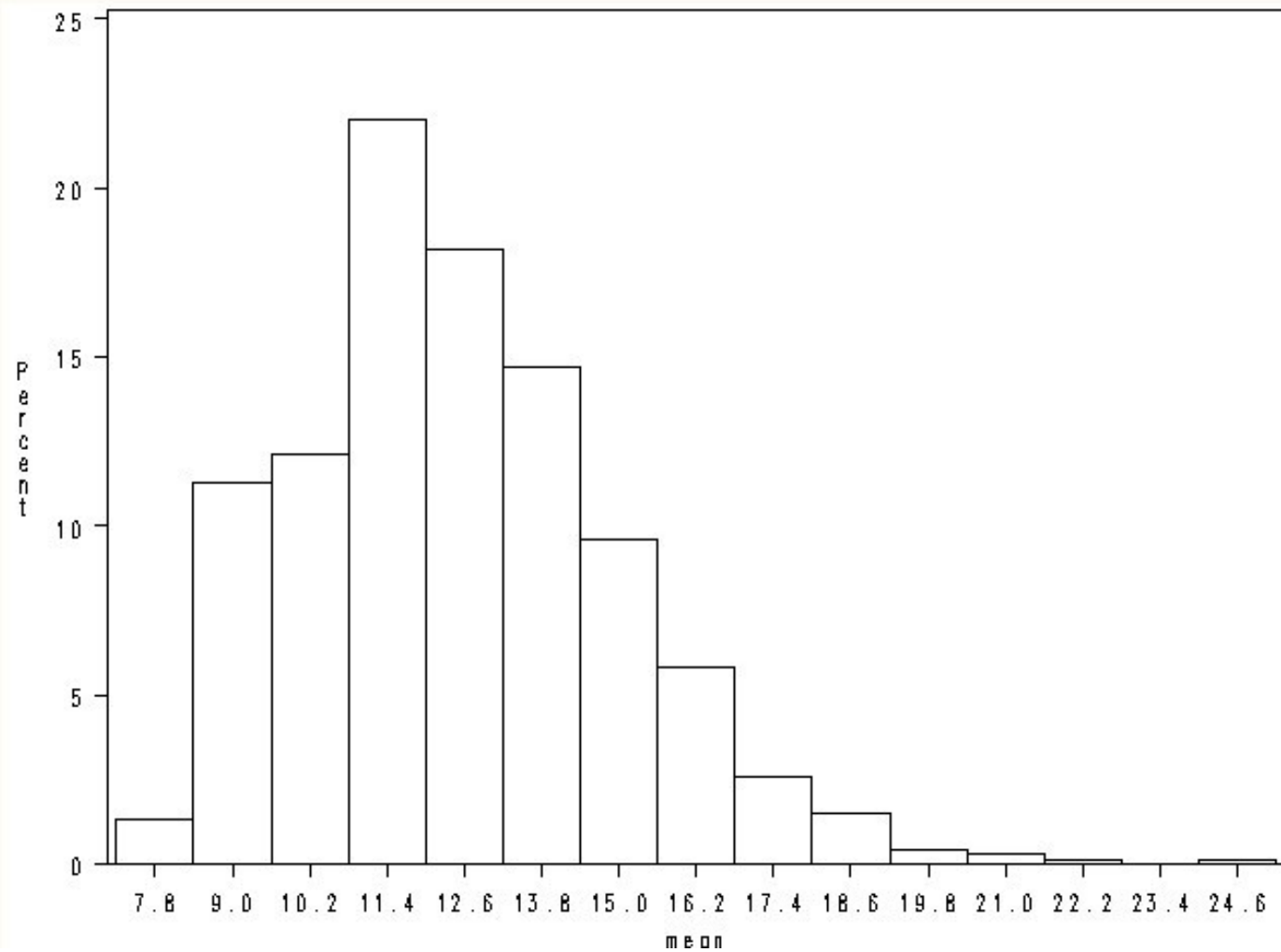
Basic Statistical Measures			
Location		Variability	
Mean	12.44744	Std Deviation	2.41351
Median	12.34050	Variance	5.82505
Mode	11.61000	Range	17.93300
		Interquartile Range	3.11200

Note: The mode displayed is the smallest of 4 modes with a count of 3.

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	25.1930
99%	18.9365
95%	16.6940
90%	15.7320
75% Q3	13.9150
50% Median	12.3405
25% Q1	10.8030
10%	9.3640
5%	8.9555
1%	8.2890
0% Min	7.2600

98% C.I is 8.289 to 18.9365

Variable: mean



We run the code the second time, and we get the result as

Variable: mean

Moments			
N	1000	Sum Weights	1000
Mean	12.457113	Sum Observations	12457.113
Std Deviation	2.42138147	Variance	5.86308822
Skewness	0.6142003	Kurtosis	0.043649
Uncorrected SS	161036.889	Corrected SS	5857.22513
Coeff Variation	19.4377419	Std Error Mean	0.07657081

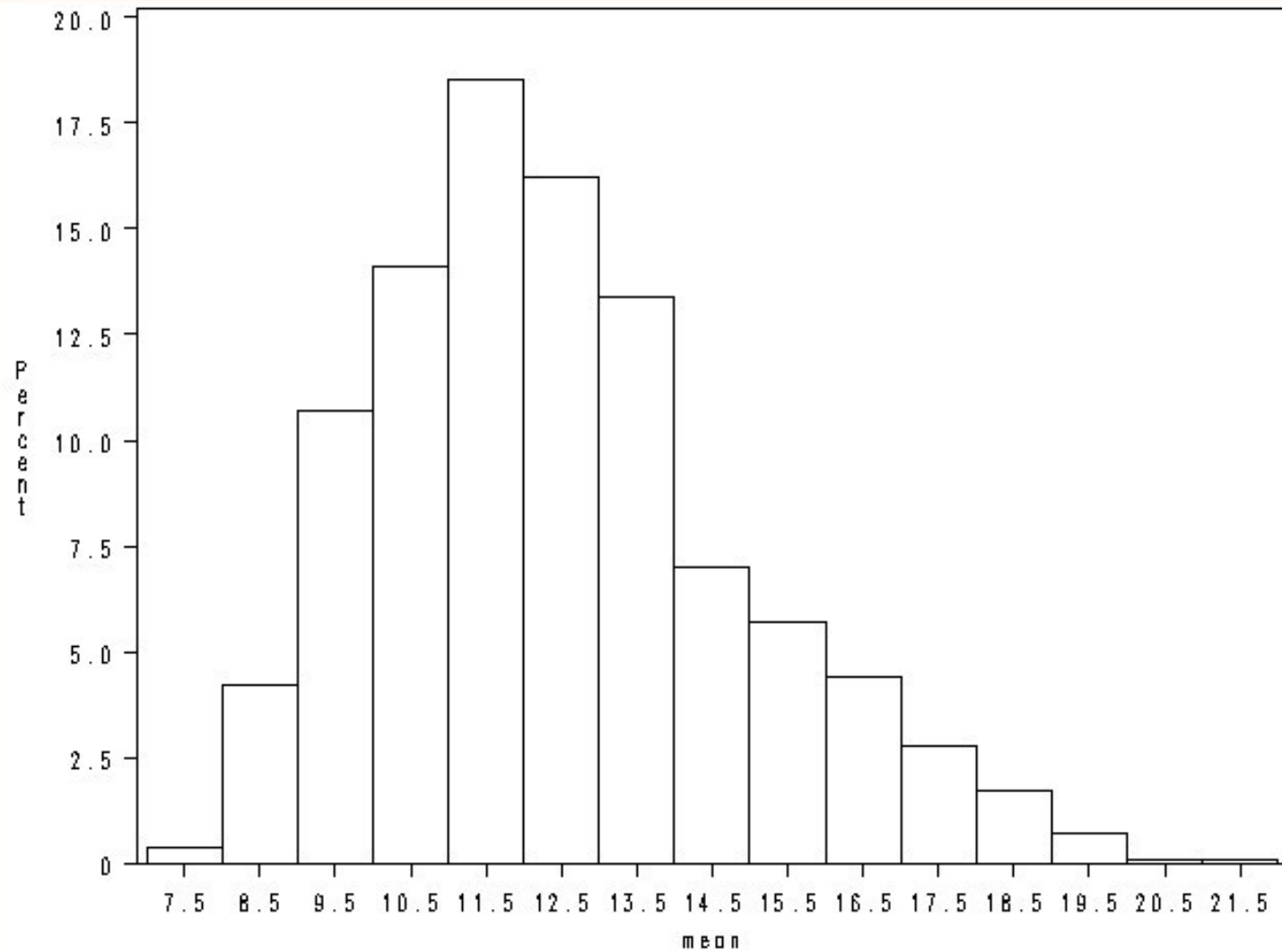
Basic Statistical Measures			
Location		Variability	
Mean	12.45711	Std Deviation	2.42138
Median	12.27100	Variance	5.86309
Mode	11.25100	Range	14.23900
		Interquartile Range	3.18950

Note: The mode displayed is the smallest of 3 modes with a count of 3.

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	21.8030
99%	18.4995
95%	17.2445
90%	15.8610
75% Q3	13.8960
50% Median	12.2710
25% Q1	10.7065
10%	9.4255
5%	9.0175
1%	8.4825
0% Min	7.5640

98% C.I is 8.4825 to 18.4995

Variable: mean

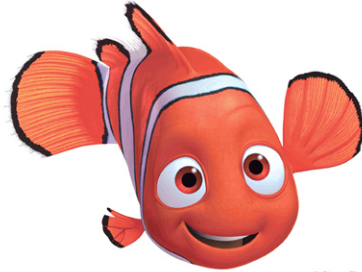


MEAN

12.447

\approx

12.457



© Disney/Pixar

C. I.

(8.289 , 18.9365)

\approx

(8.4825, 18.4995)

JACKKNIFE

The Jackknife



Jackknife methods make use of systematic partitions of a data set to estimate properties of an estimator computed from the full sample.

Quenouille (1949, 1956) suggested the technique to estimate (and, hence, reduce) the bias of an estimator $\hat{\theta}_n$.

Tukey(1958) coined the term jackknife to refer to the method, and also showed that the method is useful in estimating the variance of an estimator.

Why do we need the Jackknife?

For a data set $X = (x_1, x_2, x_3, x_4, x_5)$ the standard deviation of the average is:

$$\sigma = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

For measurements other than the mean, there is no easy way to assess the accuracy.

Jackknife Method

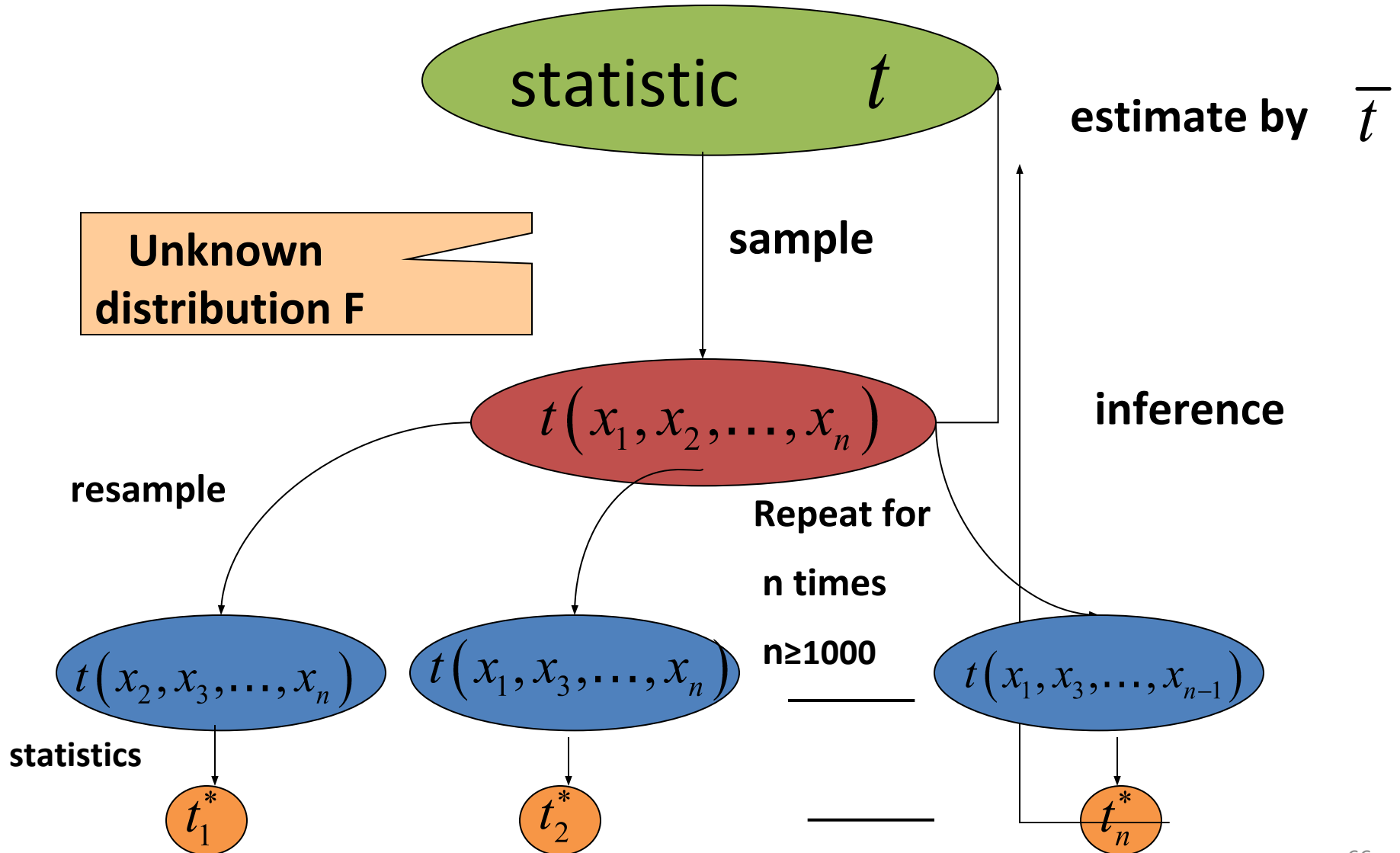
Consider the problem of estimating the standard error of a Statistic $t = t(x_1, \dots, x_n)$ calculated based on a random sample from distribution F . In the jackknife method resampling is done by deleting one observation at a time. Thus we calculate n values of the statistic denoted by

$t_i^* = t(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Let $\bar{t}^* = \sum_{i=1}^n t_i^* / n$. Then the jackknife estimate of $SE(t)$ is given by

$$\hat{SE}(t) = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (t_i^* - \bar{t}^*)^2} = \frac{(n-1)s_{t^*}}{\sqrt{n}} \quad (1)$$

where s_{t^*} is the sample standard deviation of $t_1^*, t_2^*, \dots, t_n^*$.

The nonparametric of Jackknife



The formula is not immediately evident, so let us look at the special case: $t = \bar{x}$. Then

$$t_i^* = \bar{x}_i^* = \frac{1}{n-1} \sum_{j \neq i} x_j = \frac{n\bar{x} - x_i}{n-1} \quad \text{and} \quad \bar{t}^* = \bar{\bar{x}}^* = \frac{1}{n} \sum_{i=1}^n \bar{x}_i^* = \bar{x}.$$

Using simple algebra it can be shown that

$$\sqrt{JSE(t)} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\bar{x}_i^* - \bar{\bar{x}}^*)^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}} = SE(\bar{x}) \quad (2)$$

Thus the jackknife estimate of the standard error (1) gives an exact result for \bar{x} .



Limitations of the Jackknife

- The jackknife method of estimation can fail if the statistic t_i^* is not smooth. Smoothness implies that relatively small changes to data values will cause only a small change in the statistic.
- The jackknife is not a good estimation method for estimating percentiles (such as the median), or when using any other non-smooth estimator.
- An alternate the jackknife method of deleting one observation at a time is to delete d observations at a time ($d \geq 2$). This is known as the delete- d jackknife.
- In practice, if n is large and d is chosen such that $\sqrt{n} < d < n$, then the problems of non-smoothness are removed.



Example for jackknife

SPEEDTEST.NET 11/24/2010 4:51 AM GMT DOWNLOAD 9.15 Mb/s UPLOAD 8.03 Mb/s PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 4:54 AM GMT DOWNLOAD 9.46 Mb/s UPLOAD 9.20 Mb/s PING 4 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 4:20 AM GMT DOWNLOAD 9.25 Mb/s UPLOAD 10.93 Mb/s PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:00 AM GMT DOWNLOAD 9.14 Mb/s UPLOAD 7.86 Mb/s PING 5 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 5:08 AM GMT DOWNLOAD 31.94 Mb/s UPLOAD 45.90 Mb/s PING 13 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:38 AM GMT DOWNLOAD 9.55 Mb/s UPLOAD 5.04 Mb/s PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 5:02 AM GMT DOWNLOAD 20.69 Mb/s UPLOAD 5.77 Mb/s PING 4 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:14 AM GMT DOWNLOAD 10.05 Mb/s UPLOAD 14.67 Mb/s PING 5 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi
SPEEDTEST.NET 11/24/2010 5:04 AM GMT DOWNLOAD 9.19 Mb/s UPLOAD 7.33 Mb/s PING 12 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi	SPEEDTEST.NET 11/24/2010 5:31 AM GMT DOWNLOAD 5.55 Mb/s UPLOAD 7.94 Mb/s PING 9 ms ISP: State University of New ★★★★★ 3.4/5 SERVER: New York, NY DISTANCE: ~ 50 mi

Population Sample (Mbps)

5.55, 9.14, 9.15, 9.19, 9.25, 9.46,
9.55, 10.05, 20.69, 31.94

0, 9.14, 9.15, 9.19,
9.25, 9.46, 9.55,
10.05, 20.69, 31.94

Resample
#1

5.55, 0, 9.15, 9.19,
9.25, 9.46, 9.55,
10.05, 20.69, 31.94

Resample
#2

5.55, 9.14, 0, 9.19,
9.25, 9.46, 9.55,
10.05, 20.69, 31.94

Resample
#3

⋮

Repeat for 10 times

SAS code for jackknife example

```
data in; input n y1-y10 @@;
smean = mean(of y1-y10);
sstd = std(of y1-y10);
svar = var(of y1-y10);
cards;
10
5.55 9.46 9.25 9.14 9.15 9.19 31.94 9.55 10.05 20.69
;
```

```
proc print data=in;
var smean sstd svar;
title 'Values for estimates of original sample';
data jack ; set in;
```

```
array y(10) y1-y10;
do i = 1 to n;
```

use value of n as array length

```
  ytmp = y(i);
```

```
  y(i) = .;
```

```
  jmean = mean(of y1-y10);
```

```
  jstd = std(of y1-y10);
```

```
  jvar = var(of y1-y10);
```

```
output;
```

```
  y(i) = ytmp;
```

```
end;
```

```
drop i n ytmp;
```

Calculate jackknife replicates of the sample mean,
sample standard deviations, and sample variances

SAS code for jackknife example

```
proc means data=jack noprint;  
var jmean jstd jvar;  
output out=jackset  
mean=dotmean dotstd dotvar;
```

Calculate the sample mean, the sample standard deviation, and the sample variance for the full data set

```
data jackset; merge jackset in;  
biasmean= (n-1)*(dotmean - smean);  
biasstd = (n-1)*(dotstd - sstd);  
biasvar = (n-1)*(dotvar - svar);
```

Calculate the jackknife estimates of bias

```
corrmean = smean - biasmean;  
corrstd = sstd - biasstd;  
corrvar = svar - biasvar;
```

Calculate the bias-corrected jackknife estimates

```
sejmean = sqrt((n-1)/n * cssmean);  
sejstd = sqrt((n-1)/n * cssstd);  
sejvar = sqrt((n-1)/n * cssvar);
```

Calculate the standard errors for jackknife estimates

```
keep n dotmean smean sejmean biasmean corrmean  
dotstd sstd sejstd biasstd corrstd  
dotvar svar sejvar biasvar corrvar;
```

SAS code for jackknife example

```
proc print data=jack;  
var y1-y10 jmean jstd jvar;  
format jmean 8.5 jstd 8.5 jvar 8.5 jstdn 8.5 jvarn 8.5;  
title 'Jackknife Samples and Jackknife Replications';
```

```
proc print data=jackset;  
var dotmean smean biasmean corrmean sejmean;  
format dotmean 8.5 smean 8.5 sejmean 8.5 biasmean 8.5;  
title 'Jackknife Estimates';  
title2 'for estimating the mean';
```

```
proc print data=jackset;  
var dotstd sstd biasstd corrstd sejstd;  
format dotstd 8.5 sstd 8.5 sejstd 8.5 biasstd 8.5;  
title2 'for estimating the standard deviation';
```

```
proc print data=jackset;  
var dotvar svar biasvar corrvar sejvar;  
format dotvar 8.5 svar 8.5 sejvar 8.5 biasvar 8.5;  
title2 'for estimating the variance';
```

Output of Jackknife example

Jackknife Samples and Jackknife Replications

Obs	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10
1	.	9.46	9.25	9.14	9.15	9.19	31.94	9.55	10.05	20.69
2	5.55	.	9.25	9.14	9.15	9.19	31.94	9.55	10.05	20.69
3	5.55	9.46	.	9.14	9.15	9.19	31.94	9.55	10.05	20.69
4	5.55	9.46	9.25	.	9.15	9.19	31.94	9.55	10.05	20.69
5	5.55	9.46	9.25	9.14	.	9.19	31.94	9.55	10.05	20.69
6	5.55	9.46	9.25	9.14	9.15	.	31.94	9.55	10.05	20.69
7	5.55	9.46	9.25	9.14	9.15	9.19	.	9.55	10.05	20.69
8	5.55	9.46	9.25	9.14	9.15	9.19	31.94	.	10.05	20.69
9	5.55	9.46	9.25	9.14	9.15	9.19	31.94	9.55	.	20.69
10	5.55	9.46	9.25	9.14	9.15	9.19	31.94	9.55	10.05	.

jmean

jstd

jvar

13.15778	7.97710	63.63412
12.72333	8.30346	68.94738
12.74667	8.29276	68.76993
12.75889	8.28686	68.67209
12.75778	8.28741	68.68112
12.75333	8.28957	68.71698
10.22556	4.13518	17.09975
12.71333	8.30781	69.01968
12.65778	8.32949	69.38037
11.47556	7.78418	60.59350

Output of Jackknife example

Values for estimates of original sample

Obs	smean	sstd	svar
1	12.397	7.89630	62.3515

Jackknife Estimates
for estimating the mean

Obs	dotmean	smean	biasmean	corrmean	sejmean
1	12.39700	12.39700	0.00000	12.397	2.49703

Jackknife Estimates
for estimating the standard deviation

Obs	dotstd	sstd	biasstd	corrstd	sejstd
1	7.79938	7.89630	-0.87222	8.76852	3.70002

Jackknife Estimates
for estimating the variance

Obs	dotvar	svar	biasvar	corrvar	sejvar
1	62.35149	62.35149	-0.00000	62.3515	46.00440

Jackknife
outcomes
for total

Outcome
of original
sample

Bias

Correcte Standard error
d
estimate

s

Other Resampling Methods

- » Cross-validation
 - » Resubstitution
 - » Monte Carlo

Cross-Validation

- Cross-validation is a statistical method for validating a predictive model.
- **Main Idea**: tested with data that are not used to fit the model.

Get out-of-sample tests but **still** use all the data.
The sleight of hand is to do a number of fits, each time **leaving out a different portion** of the data.

Purpose of Cross Validation

- Cross-validation is a way to **predict the fit** of a model to a hypothetical validation set when an explicit validation set is **not** available.



Types of Cross-validation

1

K-fold
cross-validation
n

2

$K \times 2$
cross-validation
n

3

Leave-one-out
cross-validation
n
(LOOCV)

Types of Cross-validation

	K-fold cross-validation	K× 2 cross-validation	Leave-one-out cross-validation (LOOCV)
Method	Randomly divides sample sets into K parts, using one to test the model that was trained on the remaining of K-1 parts.	For each fold, randomly assign data points to two sets, both sets are equal size.	Use a single observation as the validation data, remaining observations as the training data.
Advantage	All observations are used for both training and validation, and each observation is used for validation exactly once.	Good for the large training and test sets. Each data point is used for both training and validation on each fold.	Usually very expensive, because the training process should be repeated many times.

- Cross-validation can **avoid** "self-influence".
- Often used for deciding how many predictor variables to use in regression.

Resubstitution

- Test the model's predictive ability by using the sample cases that were used to develop the model.

Monte Carlo Resampling Method

- Use repeated sampling from populations with **known** characteristics to determine how **sensitive** statistical procedures are to those characteristics.



Conclusion

Comparison and Contrast

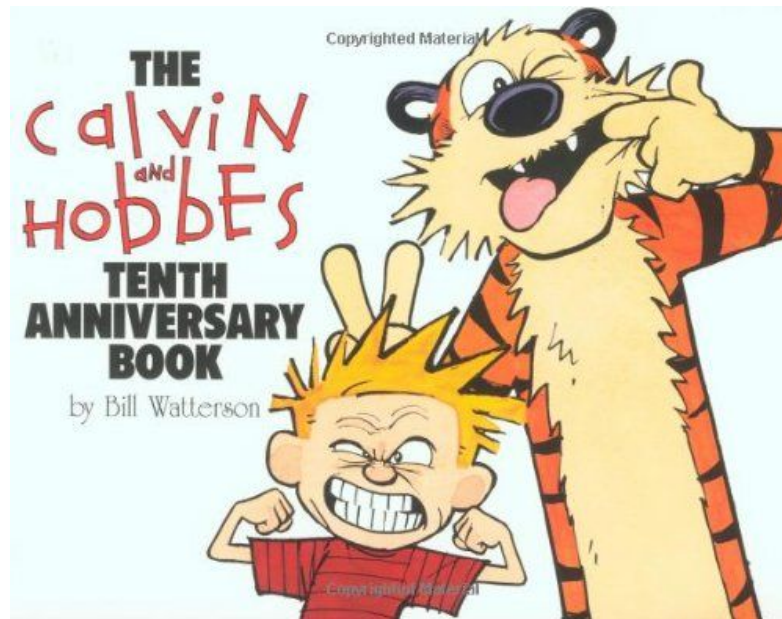
- Comparison:
robust, simpler and more accurate, easily accessible
- Contrast :
Application
Sampling Procedure

Comparison and Contrast

	Application	Sampling Procedure
Permutation	Hypothesis testing	Samples drawn at random, without replacement.
Bootstrap	Standard deviation, confidence interval, hypothesis testing, bias	Samples drawn at random, with replacement
Jackknife	Standard deviation, confidence interval, bias	Samples consist of full data set with one observation left out

Limitation and Implication

Pros: free from highly demanding the
assumption: easily understood



Limitation

Cons:

- Stephen E. Fienberg doubts on the method itself. He argued that resampling methods explored the same data many times and get the the data that can be not justified by any other ways.
- Other critics question the accuracy of resampling estimates especially when there are not enough experimental trials conducted.

Implication

For further studies:

- . further examine and compare their sensitivity to non-normality, non-equivalence of distribution and sample size

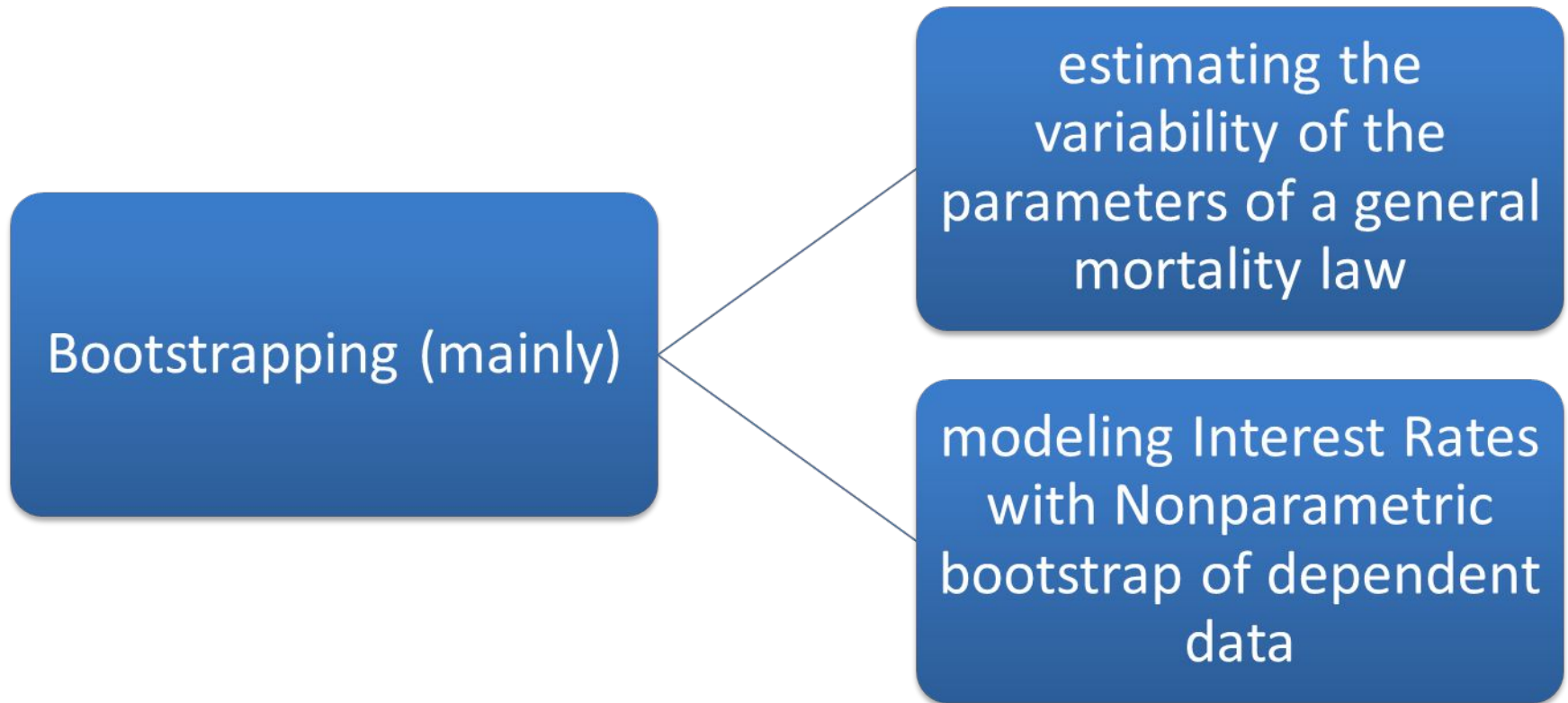


Application

- widely applied in financial field
- Take an example from actuary (e.g. modeling of mortality in life insurance and loss distributions)



Application





Thank You :)