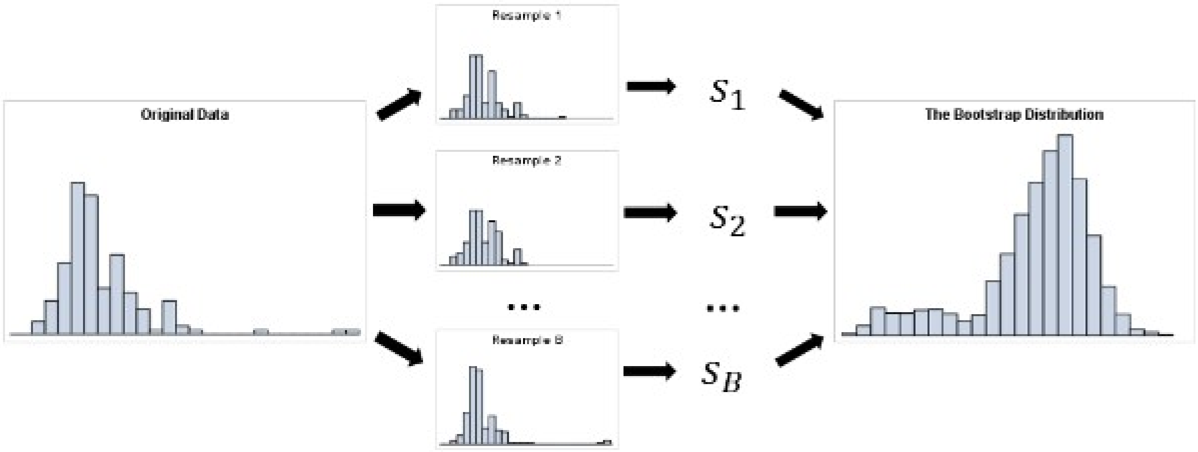


Jackknife Method

Guilherme Souza Rodrigues
guilhermerodrigues@unb.br

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.

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 = (x1, x2, x3, x4, x5)* the standard deviation of the average is:

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

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

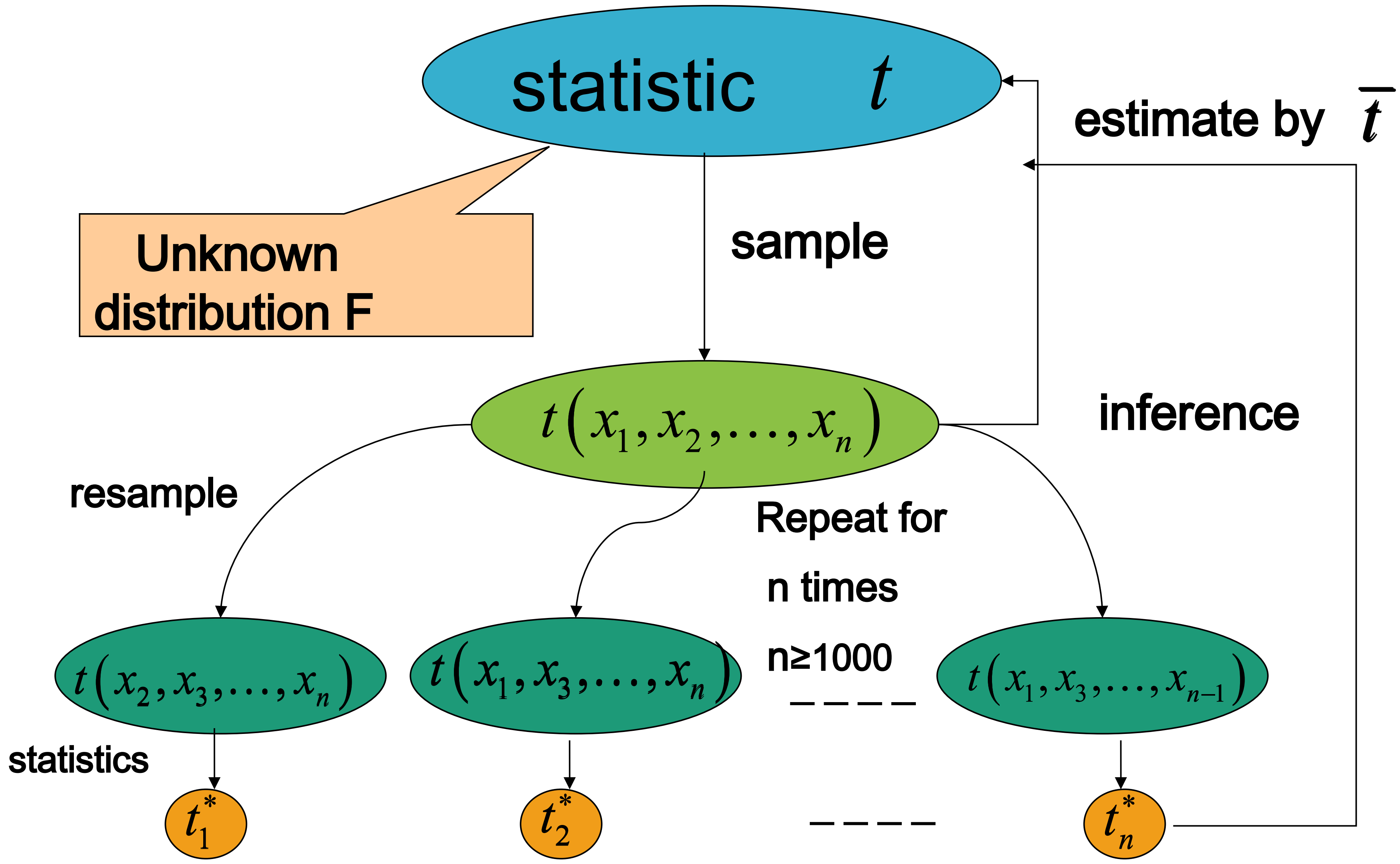
Jackknife Method

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

$$JSE(t) = \sqrt{\frac{n-1}{n} \sum_{i=1}^{n} \left(t_i^* - \bar{t}^*\right)^2} = \frac{(n-1) s_{t^*}}{\sqrt{n}} \tag{1}$$

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

The nonparametric of Jackknife



Jackknife Method

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

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

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.

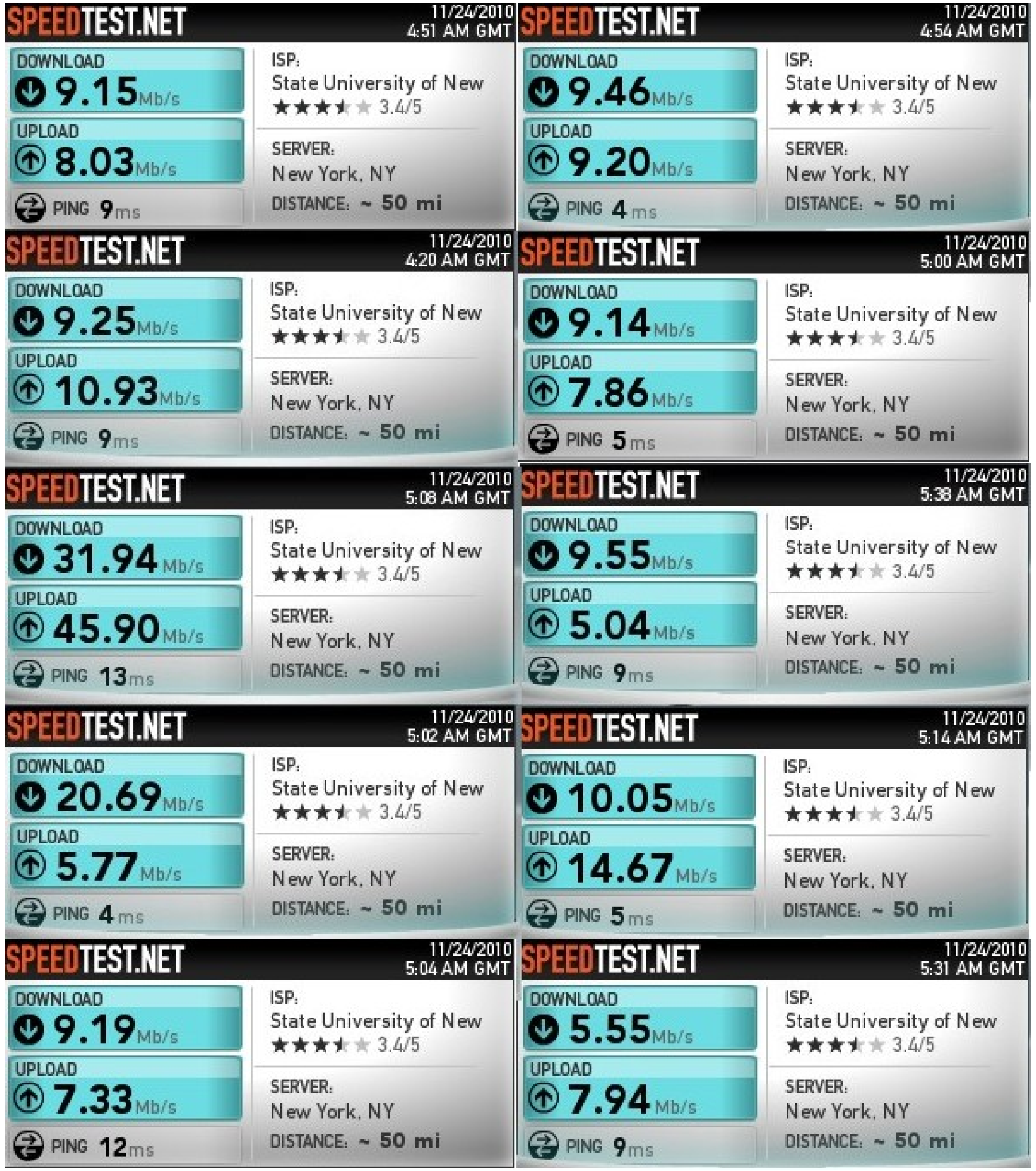
- ✓Smoothness implies that relatively small changes to data values will cause only a small change in the statistic.

- ✓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



Example for jackknife

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

Example for jackknife

The `patch` (`bootstrap`) data from Efron and Tibshirani [84, 10.3] contains measurements of a certain hormone in the bloodstream of eight subjects after wearing a medical patch. The parameter of interest is

$$\theta = \frac{E(new) - E(old)}{E(old) - E(placebo)}.$$

If $|\theta| \leq 0.20$, this indicates bioequivalence of the old and new patches. The statistic is $\overline{Y}/\overline{Z}$.

```
data(patch, package = "bootstrap")
> patch
  subject placebo oldpatch newpatch     z     y
1       1    9243    17649    16449  8406 -1200
2       2    9671    12013    14614  2342  2601
3       3   11792    19979    17274  8187 -2705
4       4   13357    21816    23798  8459  1982
5       5    9055    13850    12560  4795 -1290
6       6    6290     9806    10157  3516   351
7       7   12412    17208    16570  4796  -638
8       8   18806    29044    26325 10238 -2719
```

Compute the jackknife estimate of bias for the patch

Exemplo no R


```
data(patch,package="bootstrap")
n=nrow(patch)
y=patch$y
z=patch$z
theta.hat=mean(y)/mean(z)

#compute jacknife

theta.jack=numeric(n)
for(i in 1:n){
  theta.jack[i]=mean(y[-i])/mean(z[-i])
  }
bias=(n-1)*(mean(theta.jack)-theta.hat)
se=sqrt((n-1)*mean((theta.jack-mean(theta.jack))^2))
print(c(bias,se)), hist(theta.jack)
```

OBRIGADO!