

WE ARE GOING TO WORK WITH  
7 PARAMETRIC MODELS

# Parameter Estimation

---

N. Torelli, G. Di Credico, V. Gioia

Fall 2023

University of Trieste

**Point estimation<sup>1</sup>**

**Interval estimation<sup>2</sup>**

---

<sup>1</sup>Agresti, Kateri: sec 3-4 - 4.1

<sup>2</sup>Agresti, Kateri: sec 4.3 - 4.4

## Point estimation

---

IN THIS PART WE ALWAYS REFER TO PARAMETRIC MODEL THAT MEANS IT HAVE A PARAMETER THAT DEFINE THE MODEL

## The aim of point estimation

Given a model for the data  $\mathbf{y}$ , with parameter  $\theta$ , **point estimation** is concerned with finding a reasonable parameter estimate from the data.

There are several methods for doing this, and the problem can be simply stated as *finding the parameter value most consistent with the data*, a definition that leads to the method of **maximum likelihood estimation**.

We will delve into the details of maximum likelihood estimation in due time, but here we focus on some general aspects of point estimation.

## Example: sample mean and sample variance

A very simple model assumes that the data are a random sample from a normal distribution namely they are the observations of i.i.d. r.v. from  $\mathcal{N}(\mu, \sigma^2)$ .

$\theta$  PARAMETER

Straightforward estimates of  $\mu$  and  $\sigma^2$  are given by **the sample mean**

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

ESTIMATOR

and by the **sample variance**

WE USE A HAT TO  
INDEX A  
ESTIMATOR

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2.$$

IS A STATISTIC THAT IS  
USEFULL TO KNOW  
MORE ON THE  
PARAMETER

WE OBTAIN A VALUE THAT IS CALLED ESTIMATE (OF THE PARAMETER)  
Such estimates are actually sensible anytime we are interested in  
estimating the mean and variance of an i.i.d. sample.

$\theta$  PARAMETER

$\hat{\theta}$  ESTIMATOR

$\hat{\theta} = 3$  ESTIMATE

## Estimation properties

IT'S NOT THE EXACTLY VALUE OF THE  
PARAMETER

To figure out what could be a good estimate, we need to consider *repeated estimation under repeated replication of the data-generating process.* → FOR OBTAINING MORE KNOWLEDGE ABOUT THE PARAMETER

This makes fully sense whenever the available data are a random sample obtained from a large population, like in industrial or social surveys, so that it would perfectly possible to iterate the sampling and obtain further data with the same structure of  $y$ .

However, we apply the same logic even when repetition is just the result of an idealization, like in the case of the temperatures recorded in New Haven of the previous lecture.

The point is: what do we expect to find if we repeat the same analysis to many data sets generated from the same model?

# Unbiasedness

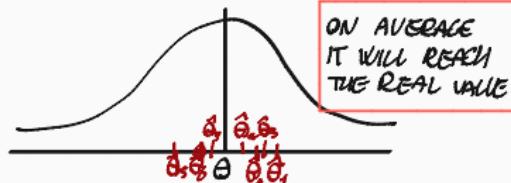
If we replicate the random data and we repeat the estimation process, the result will be a different value of  $\hat{\theta}$  for each replicate.

The values are observations of a random vector, the **estimator** of  $\theta$ , which is usually also denoted by  $\hat{\theta}$  (the context will make clear whether we are referring to the estimator or to the estimate for a given sample).

Since, the estimator is a r.v., it makes fully sense to compute its mean.

For an **unbiased** estimator

LET'S TRY TO  
ESTIMATE IT  
 $\hat{\theta} \Rightarrow$   
 $E(\hat{\theta}) = \theta$ .



Unbiasedness is a desirable property, and we would also like the estimator to have **low variance**.

$\Delta$  IF  $n \rightarrow \infty$  THEN THE BIAS  $\rightarrow 0$   
BE AWARE OF THE LOW SIZE POPULATION

$$E(\bar{Y}) = E\left(\sum_i \frac{Y_i}{n}\right)$$

$$= \sum_i \frac{E(Y_i)}{n}$$

$$= \sum_i \frac{\mu}{n} = \mu$$

$$\text{BIAS} = E(\bar{Y}) - \mu$$

$$= \mu - \mu = 0$$

THE MEAN

IT'S UNBIASED

NOTATION :

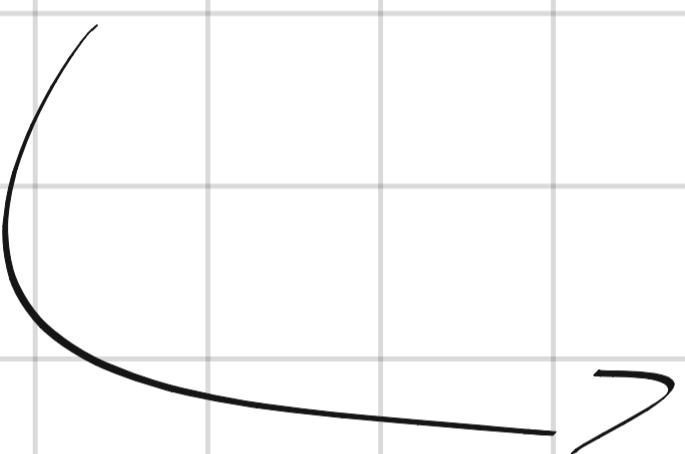
$$\hat{\mu} = \bar{Y}$$

$$\hat{\sigma}^2 = s^2$$

$$\begin{aligned}
 E(\hat{\sigma}^2) &= E\left[\frac{\sum_i (y_i - \bar{y})^2}{n}\right] = E\left[\frac{\sum_i ((y_i - \mu) - (\bar{y} - \mu))^2}{n}\right] \\
 &= E\left[\frac{\sum_i (y_i - \mu)^2}{n} + \frac{\sum (\bar{y} - \mu)^2}{n} - \frac{2\sum (y_i - \mu)(\bar{y} - \mu)}{n}\right] \\
 &\quad \left| \begin{array}{c} n(\bar{y} - \mu) \\ \hline n \end{array} \right. \quad \left. \begin{array}{c} -2(\bar{y} - \mu) (\sum y_i - \sum \mu) \\ n \end{array} \right. \\
 &= -2(\bar{y} - \mu)(\bar{y} - \mu) = -2(\bar{y} - \mu)^2 \\
 &= E\left[\frac{\sum (y_i - \mu)^2}{n} - (\bar{y} - \mu)^2\right] \\
 &\stackrel{1}{=} \sum \frac{E(y_i - \mu)^2}{n} - E(\bar{y} - \mu)^2 \\
 &\stackrel{1}{=} \sum \frac{\text{Var}(y_i)}{n} - \text{Var}(\bar{y}) = \frac{n\sigma^2}{n} - \frac{\sigma^2}{n} = \sigma^2 \left(\frac{n-1}{n}\right)
 \end{aligned}$$

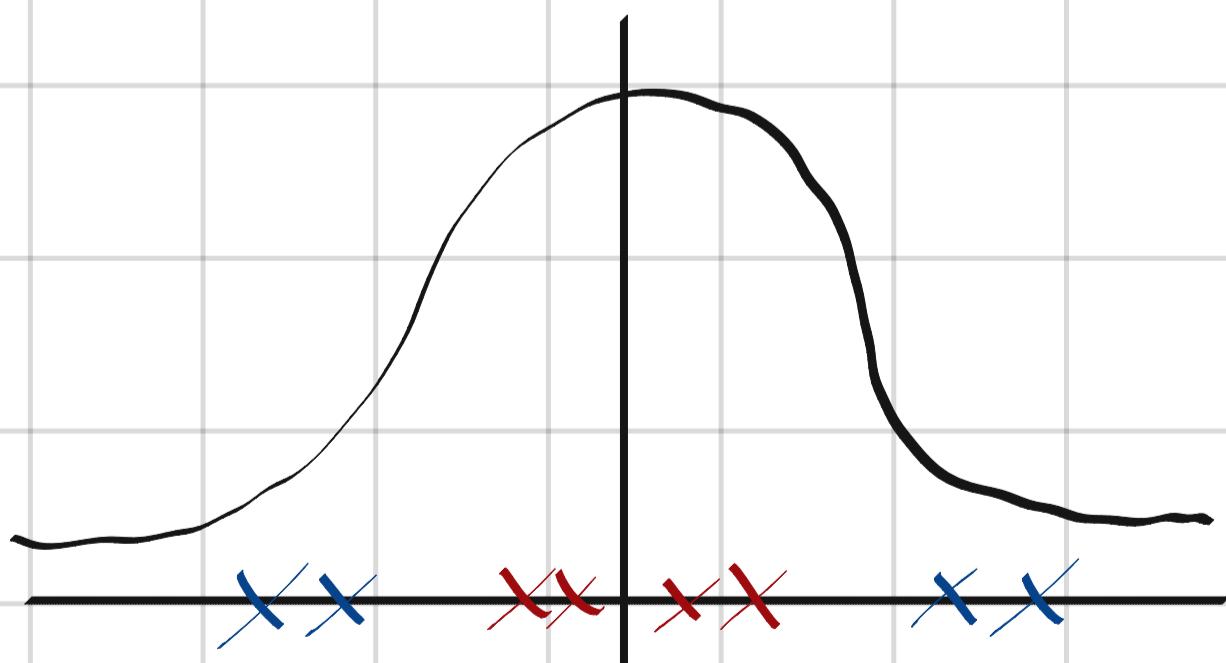
$$\text{Bias}(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = \sigma^2 \left( \frac{n-1}{n} \right) - \sigma^2 = \sigma^2 \left( \frac{n-1-n}{n} \right) = \sigma^2 \left( \frac{-1}{n} \right)$$

ON AVERAGE WE OBTAIN SMALLER VALUE THAN THE TRUTH ONE BECAUSE THE DIFFERENCE IS NEGATIVE AND IT UNDERESTIMATE THE TRUTH OF THE PARAMETER



WE COMPUTE  $s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$   
THE  $n-1$  ALLOW AN UNBIAS VERSION

WE SEEK LOW VARIANCE



WE PREFER THE RED VALUES, AS THE VALUE OF THE ESTIMATE WILL BE MORE CLOSE.

WE USE THE MSE FOR "COMPUTE" THE BIASE AND VARIANCE

## Mean Squared Error

There is *tradeoff* between unbiasedness and low variance, so we usually seek to get both (to some extent): ideally we would target a small **Mean Squared Error (MSE)**

$$\text{MSE}(\hat{\theta}) = E\{(\hat{\theta} - \theta)^2\}.$$

EXPECTED ERROR  $(\text{ESTIMATOR} - \text{PARAMETER})^2$

With some algebra, we obtain

IF BIAS = 0, MSE = VARIANCE

$$\text{MSE}(\hat{\theta}) = \{E(\hat{\theta}) - \theta\}^2 + \text{var}(\hat{\theta}) = \text{Squared bias} + \text{Variance}.$$

IF AN ESTIMATOR IS UNBIASED IT MEANS THAT THE MSE IS ITS VARIANCE.

## Example: normal random sample

For a normal random sample, it is straightforward to verify that

$$E(\bar{Y}) = \mu, \quad \text{var}(\bar{Y}) = \frac{\sigma^2}{n} = \text{MSE}(\bar{Y}).$$

For the sample variance, we use the property that

$$\frac{(n-1)}{\sigma^2} S^2 \sim \chi_{n-1}^2,$$

to obtain

$$E(S^2) = \sigma^2, \quad \text{var}(S^2) = \frac{2\sigma^4}{(n-1)} = \text{MSE}(S^2).$$

*+ BIAS = 0*

The unbiasedness of the sample mean and variance is a general property, holding also for non-normal samples.

# Consistency

THE ABILITY OF A/R ESTIMATOR TO CONVERG TO THE PARAMETER

A (scalar) estimator is said to be **(weakly) consistent** if, for any  $\epsilon > 0$

$$\Pr(|\hat{\theta} - \theta| > \epsilon) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

A sufficient condition for this is that  $MSE(\hat{\theta}) \rightarrow 0$  for large samples, which requires that both bias and variance become negligible.

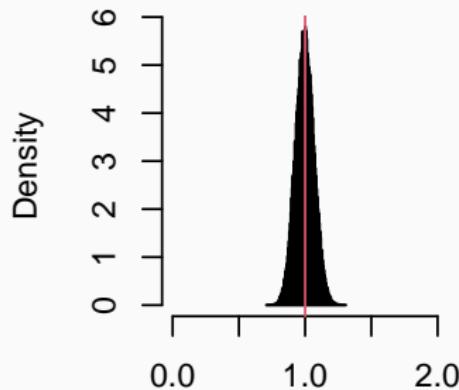
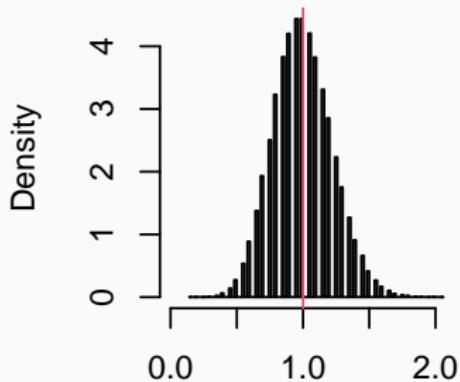
The law of large samples implies that the sample mean is a consistent estimator for the true mean in random samples.

## R lab: consistency of the sample mean

```
M <- 100000; n1 <- 20; n2 <- 200; y1 <- y2 <- rep(NA, M)
for(i in 1:M) {y1[i] <- mean(rpois(n1, 1))
                y2[i] <- mean(rpois(n2, 1))}

par(mfrow=c(1,2))

hist.scott(y1, xlim=c(0,2), main="", xlab="")
hist.scott(y2, xlim=c(0,2), main="", xlab="")
```



# Efficiency

An **efficient estimator** is an estimator that estimates the parameter of interest in some *optimal* manner.

Among estimators with negligible bias, efficiency is associated to small variance. Since this is the case of consistent estimators, they are usually compared in terms of their variance.

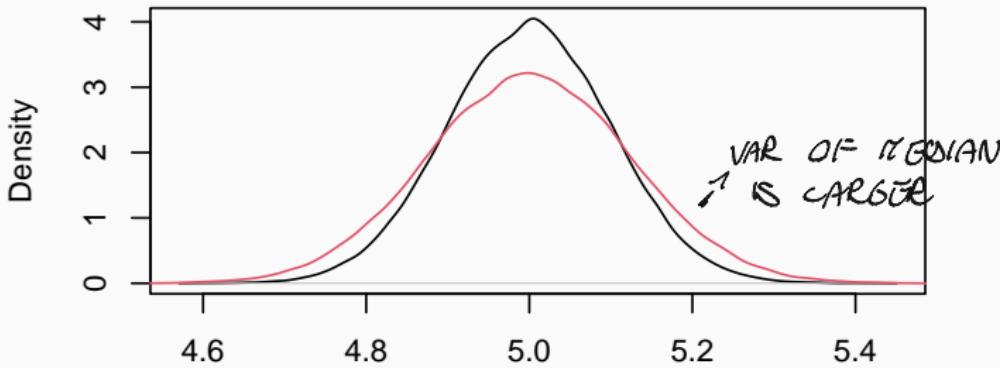
ALSO:

- UNBIASED
- LOW VARIANCE
- CONSISTENT

## R lab: efficiency of the sample mean

For a normal random sample, both the sample mean and sample median are consistent estimators of  $\mu$ . The mean is more efficient.

```
M <- 100000; n <- 100; mat.y <- matrix(NA, nrow = M, ncol = 2)
for(i in 1:M) {y <- rnorm(n, 5)
  mat.y[i,] <- c(mean(y), median(y))}
plot(density(mat.y[,1]), type="l", main="")
lines(density(mat.y[,2]), col=2)
```

$$V(\bar{Y}) = \frac{\sigma^2}{n}$$
$$V(\bar{M}_e) = \frac{25\sigma^2}{n}$$


N = 100000 Bandwidth = 0.008994

RECAP:

WE HAVE SPOKEN ABOUT THE VARIABILITY OF AN ESTIMATOR.  
FOR EXAMPLE, IF WE COLLECT DATA FROM A NORMAL SAMPLE AND  
WE ASSUME ALL VARIABLE i.i.d AND WE COMPUTE THE SAMPLE  
MEAN (WHICH IS AN ESTIMATOR OF PARAMETER  $\mu$  IN NORMAL CASE)  
YOU KNOW WHAT EXACTLY IS THE DISTRIBUTION:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

AGAIN, BECAUSE THE ESTIMATOR IS OBTAINED BY A COMBINATION OF  
RANDOM VARIABLES IS STILL A RANDOM VARIABLE.

THE SINGLE VALUE WE OBTAIN FROM THE ESTIMATOR IS THE ESTIMATE.  
 $\frac{\sigma^2}{n}$  IS THE VARIABILITY AND, IN THIS CASE, THE VARIANCE OF THE  
ESTIMATOR.

## Standard Error

An important quantity defined for a (scalar) estimator is given by its **standard error**, defined as

$\sqrt{\text{STD DEVIATION OF ESTIMATORS}}$        $\text{SE}(\hat{\theta}) = \sqrt{\text{var}(\hat{\theta})}.$

Once a sample is observed, and a numerical estimate of  $\theta$  obtained, then the estimated standard error is obtained by replacing  $\theta$  by  $\hat{\theta}$ .

An example is the **standard error of the mean**  $\text{SE}(\bar{Y}) = \sigma/\sqrt{n}$ , which is estimated by  $s/\sqrt{n}$ .

In applications, the estimated standard error is routinely reported along with the estimate, since it quantifies the **estimation precision**.

PROBLEM: IF WE ARE CONSIDERING THE TRANSFORMATION OF A PARAMETER, OBTAINING THE SAMPLING DISTRIBUTION OF THE ESTIMATOR IT'S NOT SO STRAIGHTFORWARD BECAUSE IN GENERAL WE CAN RELY ON THE CENTRAL LIMIT THEOREM AND KNOW ASYMPTOTICALLY WHAT IS THE DISTRIBUTION OF AN ESTIMATOR

## The delta method

$T = \hat{\theta}$  NOTATION

THE ARE CASE IN WHICH WE APPLY A TRANSFORMATION AND TO ELIMINATE IT'S VARIABILITY WE HAVE TO OBTAIN THE SE AND TO OBTAIN THE SE WE CAN USE THE DELTA METHOD.

Suppose that we are interested in a parameter which is a function of a scalar parameter  $\theta$ , namely  $\psi = g(\theta)$ .  $\hookrightarrow$  A TRANSFORMATION

for a continuous and differentiable function  $g$ .

WE WANT TO FIND  $\hat{\psi} = g(\hat{\theta}) = g(T)$

If  $\hat{\theta}$  is a consistent estimator of  $\theta$ , then the **continuous mapping theorem** ensures that  $g(\hat{\theta})$  is consistent for  $\psi$ .

Its standard error is provided by the **delta method**, stating that

$$\text{SE}(\hat{\psi}) \doteq \text{SE}(\hat{\theta}) |g'(\theta)|,$$

with the approximation becoming more accurate for larger samples.

The result can be extended to settings with multiple parameters.

FOR THE CENTRAL LIMIT THEOREM WE KNOW THAT ASYMPTOTICALLY T FOLLOW A NORMAL DISTRIBUTION

$$T \sim N(\theta, \frac{\sigma^2}{n})$$

$\theta$  := MEAN

$\sigma^2$  := ASSUMED VARIANCE

$\frac{1}{n}$  := BY THE CENTRAL LIMIT THEOREM

WE CAN REWRITE IT:

$$(T - \theta) \sim N(0, \frac{\sigma^2}{n}) = \sqrt{n}(T - \theta) \sim N(0, \sigma^2) \quad (1)$$

BECAUSE  $g(T) \sim ?(?, ?)$  WE USE TAYLOR

$$g(T) \approx g(\theta) + g'(\theta)(T - \theta) + \left( \lim_{n \rightarrow \infty} h(\theta) \right) \quad (2)$$

WE CAN REWRITE (2) IN A WAY TO OBTAIN (1):

$$g(T) - g(\theta) \approx g'(\theta)(T - \theta)$$

MULTIPLY  $\sqrt{n}$  TO BOTH TERMS AND WE OBTAIN (1):

$$\sqrt{n}(g(T) - g(\theta)) \approx g'(\theta)(T - \theta)\sqrt{n} \sim N(0, \sigma^2 g'(\theta)^2) \quad (3)$$

In a logical way, if  $g'(\theta)(T-\theta)\sqrt{n} \approx \sqrt{n}(g(\theta) - g(T))$ , then

$$\Rightarrow \sqrt{n}(g(T) - g(\theta)) \sim N(0, \sigma^2 g'(\theta)^2)$$

$$\Rightarrow g(T) \sim N(g(\theta), g'(\theta)^2 \frac{\sigma^2}{n})$$

$$Var(T) = Var(\hat{\theta})$$

■

THE VARIANCE, AND SO THE SE, DEPENDS ON  $g'$  AND IT'S RATE OF CHANGE.

SO, THERE'S A CORRELATION BETWEEN SE AND  $g'$ , IN FACT IF  $\hat{\gamma} = g(\theta)$ :

$$SE(\hat{\gamma}) = SE(\hat{\theta}) |g'(\theta)|$$

∴ IT'S SO IMPORTANT BECAUSE THE SE IS AN IMPORTANT WAY OF HOW TO EVALUATE AND WE NEED IT.

# Robust estimation

↑ THE ESTIMATOR DOESN'T CHANGE TOO MUCH

A **robust** estimator has good performances across a wide range of statistical models for the data.

The **sample median** is a robust estimation of location, not affected by possible outlying data, quite the opposite of the sample mean.

Robust estimation trades some efficiency with resistance to outliers, and they are often a sensible choice for semi-automatic data analyses.

SAMPLE MEDIAN > SAMPLE MEAN



ORDER AND USE  
THE 2<sup>o</sup> QUANTILE



SENSIBLE TO  
OUTLAYER

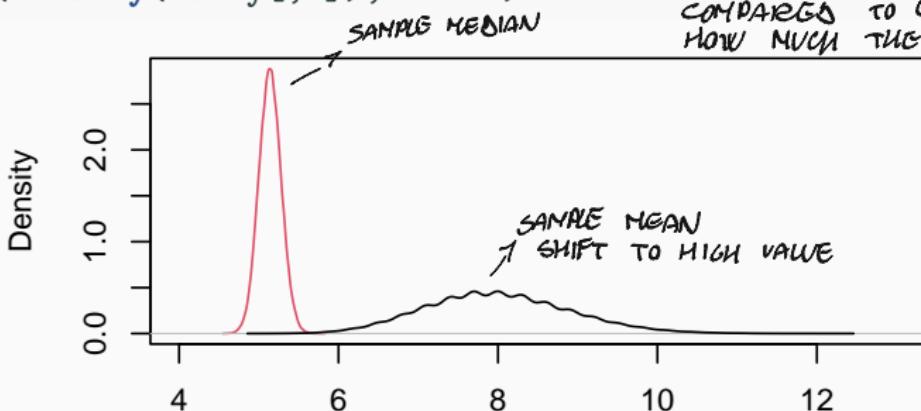
## R lab: robustness of the sample median

```
M <- 100000; n <- 100; mat.y <- matrix(NA, nrow = M, ncol = 2)
for(i in 1:M) { x <- rbinom(n, size = 1, prob = 0.9)
  y <- x * rnorm(n, 5) + (1 - x) * rnorm(n, 35)
  mat.y[i,] <- c(mean(y), median(y))}
```

*KERNEL DENSITY*  
↑  
*plot(density(mat.y[,2]), type="l", main="", xlim=c(4, 13), col = 2)*

```
lines(density(mat.y[,1]), col=1)
```

THE MEAN AND THE  
MEDIAN HAVE TO BE  
COMPARED TO EVALUATE  
HOW MUCH THEY DIFFER



N = 100000 Bandwidth = 0.01255

## Interval estimation

---

## The aim of interval estimation

Confidence intervals provide more satisfactory estimation results than point estimates alone, giving an entire set of values to estimate the model parameter.

They are built by considering a single parameter at a time.

Extensions to multidimensional *confidence regions* exist, but they are seldom used in practice.

## Pivots

Confidence intervals make suitable usage of **pivots**, which are **functions of the data and the parameter whose distribution is known**.

↳ THAT DO NOT DEPEND ON PARAMETERS

A notable example is the following one for a random sample from a  $\mathcal{N}(\mu, \sigma^2)$  distribution, when the parameter of interest is the mean  $\mu$ , and  $\sigma^2$  is not known (so that  $\theta = (\mu, \sigma^2)$ ):

$$T(\mu) = \frac{\bar{Y} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}, \quad \forall \mu \in \mathbb{R}, \sigma^2 > 0$$

WE USE  $\hat{\mu}$  = SAMPLE MEAN

EXAMPLE: WE ASSUME  $\sigma^2$  KNOWN AND WE WANT TO CONSTRUCT A CONFIDENCE INTERVAL FOR  $\mu$ :

$$\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$$

AGAIN, WHAT WE NEED TO BUILD AN INTERVAL? A PIVOT THAT IS A FUNCTION THAT DEPENDS ON THE DATA AND THE PARAMETER. WE APPLY A TRANSFORMATION:

$$\bar{Y} - \mu \sim N(0, \frac{\sigma^2}{n})$$

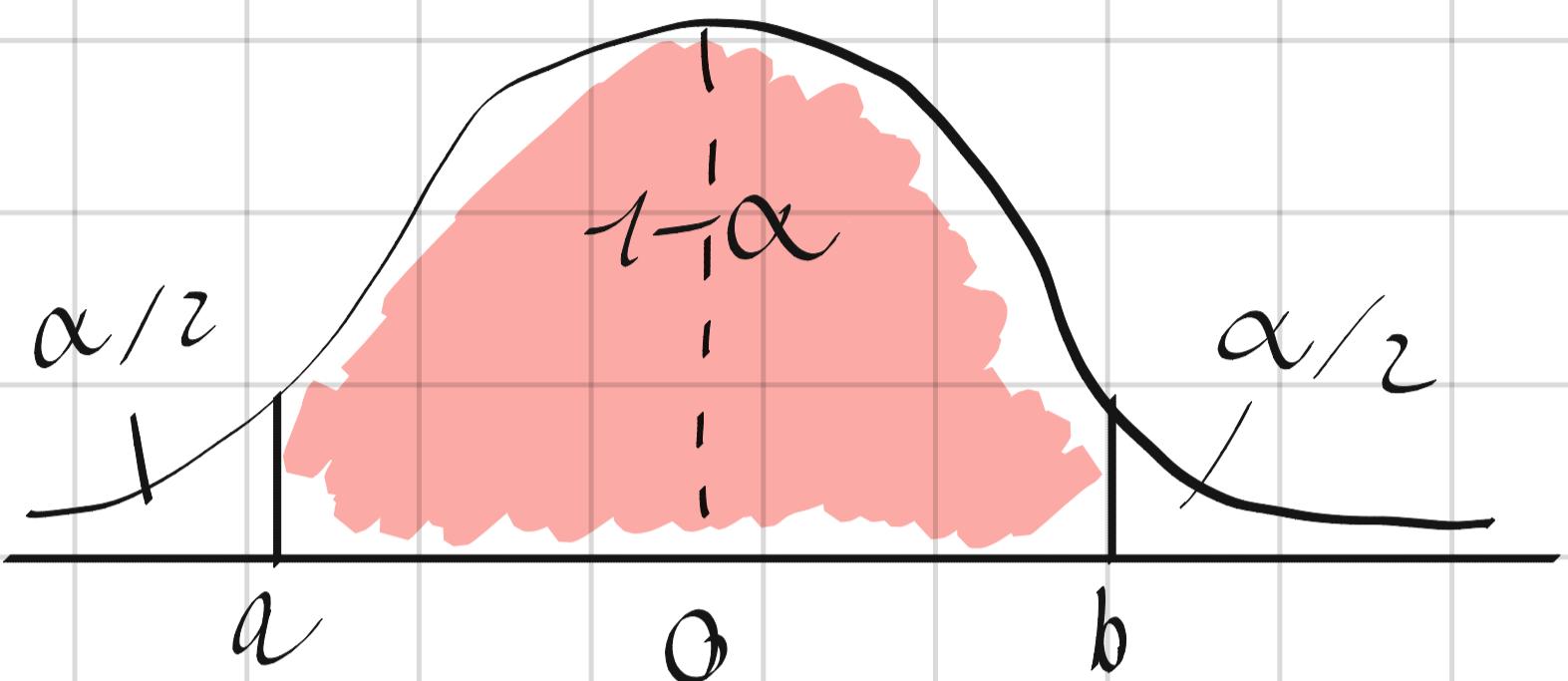
THIS MIGHT BE A PIVOTAL QUANTITY, BUT TO REFER TO A WELL KNOWN DISTRIBUTION WE PREFER THE STANDARD VERSION:

$$T(\mu) = \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

THIS IS OUR PIVOTAL QUANTITIES.

(1)

THE NEXT STEP IS SETTING THE CONFIDENTIAL LEVEL THAT CORRESPOND TO THE AREA UNDER THE CURVE OF THE PILOTAL QUANTITIES DISTRIBUTION



$$1-\alpha = \begin{cases} 0.99 \\ 0.95 \\ 0.90 \end{cases}$$

TYPICALLY ONE  
OF THIS

IF WE SAY  $P(a < T(\mu) < b) = 1-\alpha$  WE HA CONSTRUCT THE FIRST INTERVAL.

NOW, WE HAVE TO ARRANGE SOME TERMS TO COME TO THE FINAL INTERPRETATION

$$P(a < T(\mu) \stackrel{(1)}{=} b) = 1-\alpha$$

$$P\left(a < \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} < b\right) = 1-\alpha$$

=> NEXT PAGE

a AND b ARE TWO QUANTILES OF STD. NORMAL DISTRIBUTION:

$a = z_{\alpha/2}$  AND  $b = z_{(1-\alpha)/2}$   
BUT BECAUSE OF THE SYMMETRY  
WE CAN SAY THAT:

$$\begin{cases} a = -z_{(1-\alpha)/2} \\ b = z_{\alpha/2} \end{cases} \quad (2)$$

$$P\left(z_{\alpha/2} \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < z_{(1-\alpha)/2}\right) = 1 - \alpha$$

We use (2)

$$P\left(z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{Y} - \mu < z_{(1-\alpha)/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(-\bar{Y} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{Y} + z_{(1-\alpha)/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\bar{Y} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{Y} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$= -z_{(1-\alpha)/2}$

$$P\left(\bar{Y} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{Y} + z_{(1-\alpha)/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

THIS INTERVAL IS RANDOM GIVE TO  $\bar{Y}$  AND WE EXPECT THIS INTERVAL COMPUTED ON VERY LARGE NUMBER OF SAMPLES CONTAIN  $\mu$  WITH PROBABILITY  $1 - \alpha$ .

AND IF WE WANT TO BE SURE THAT THE PARAMETER IS INSIDE AND HAVE MORE CONFIDENCE, I WILL LOWER  $\alpha$  AND INCREASE THE QUANTILE AND INCLUDING MORE PROBABILTY VALUES FOR  $\theta$ . OTHERWISE I DECREASE.

## Obtaining a confidence interval

In the normal random sample example, from the previous pivot property it follows that (for  $0 < \alpha < 1$ )

$$\Pr(t_{n-1;\alpha/2} \leq T(\mu) \leq t_{n-1;1-\alpha/2}) = 1 - \alpha,$$

where  $t_{n-1;\alpha}$  is the  $\alpha$  quantile of a  $t_{n-1}$  distribution; due to symmetry of the latter,  $t_{n-1;\alpha/2} = -t_{n-1;1-\alpha/2}$ .

With some simple algebra, the previous property is equivalent to

$$\Pr\left(\bar{Y} - t_{n-1;1-\alpha/2} \sqrt{\frac{S^2}{n}} \leq \mu \leq \bar{Y} + t_{n-1;1-\alpha/2} \sqrt{\frac{S^2}{n}}\right) = 1 - \alpha.$$

## Definition of confidence interval

Hence the *random interval* with endpoints *FOR A NORMAL DISTRIBUTION AND  $\sigma$  UNKNOWN*

$$\bar{Y} - t_{n-1;1-\alpha/2} \sqrt{\frac{S^2}{n}}, \quad \bar{Y} + t_{n-1;1-\alpha/2} \sqrt{\frac{S^2}{n}}$$

contains  $\mu$  with probability  $(1 - \alpha)$ .

This interval is called a  $(1 - \alpha) \times 100\%$  **confidence interval**.

Common choices are  $(1 - \alpha) = 0.95$  or  $(1 - \alpha) = 0.99$ .

## Interpretation

Given a particular set of data  $y_1, \dots, y_n$  we calculate the confidence interval by replacing  $\bar{Y}$  and  $S^2$  with their observed values  $\bar{y}$  and  $s^2$

*observed  
1 NOT r.v.*

$$\bar{y} - t_{n-1;1-\alpha/2} \sqrt{\frac{s^2}{n}}, \quad \bar{y} + t_{n-1;1-\alpha/2} \sqrt{\frac{s^2}{n}}$$

This interval *either does or does not contain the true value of  $\mu$ .*

The probability interpretation previously introduced refers to an *hypothetical sequence of sets of data* generated from the statistical model.

## R lab: confidence interval

Follow with R file

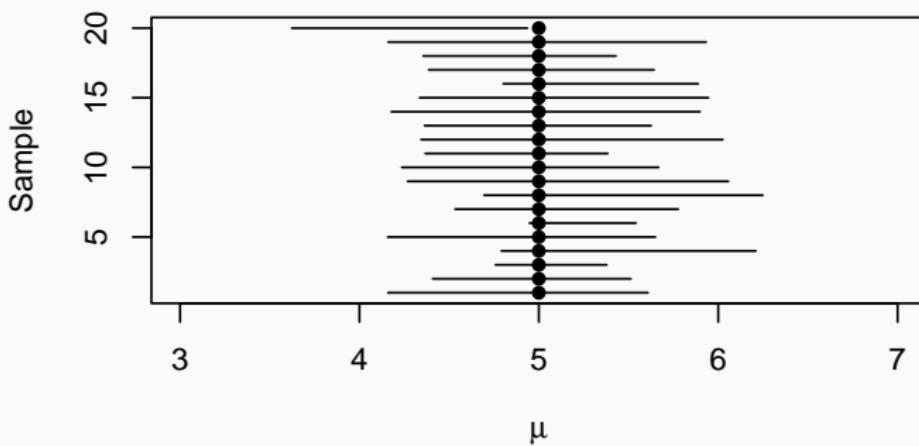
```
M <- 100000; n <- 10; mat.ci <- matrix(NA, nrow = M, ncol = 2)
for(i in 1:M) { y <- rnorm(n, 5)
  se_t <- sqrt(var(y) / n) * qt(0.975, n-1)
  mat.ci[i,] <- mean(y) + se_t * c(-1, 1)}
mean(mat.ci[,1] < 5 & mat.ci[,2] > 5)

## [1] 0.94909
```

## R lab: visualizing confidence intervals

We can visualize the first 20 simulated confidence intervals, expecting that (on average) 19 out of 20 will include the true  $\mu$

```
plot(rep(5, 20), 1:20, pch = 16, ylab="Sample",
      xlab=expression(mu))
for(i in 1:20) segments(mat.ci[i,1], i, mat.ci[i,2], i)
```



## One-sided confidence intervals

If we lift the equi-tailed condition, we can define infinitely many intervals such that

$$\Pr \left( \bar{Y} - t_{n-1;1-\alpha_1} \sqrt{\frac{S^2}{n}} \leq \mu \leq \bar{Y} + t_{n-1;1-\alpha_2} \sqrt{\frac{S^2}{n}} \right) = 1 - \alpha,$$

where  $\alpha_1 + \alpha_2 = \alpha$ .

Other than the standard choice  $\alpha_1 = \alpha_2 = \alpha/2$ , other notable choices are  $\alpha_1 = 0$  (which makes the lower limit equal to  $-\infty$ ) or  $\alpha_2 = 0$  (which makes the upper limit equal to  $\infty$ ).

They are called **one-sided confidence intervals**, and are sometimes employed in applications.

# Approximate confidence intervals & coverage probability

They rest on the fact that we are able to find an approximate distribution for pivotal quantities due to central limit theorem.

Exact pivots are scarce, but approximate ones are easy to find.

A common one is the **Wald pivot** for a generic parameter of interest  $\psi$ , based on a consistent estimator which is approximately normally distributed for large samples

$$Z(\psi) = \frac{\hat{\psi} - \psi}{\text{SE}(\hat{\psi})} \stackrel{\text{STILL STAY NORMAL}}{\sim} \mathcal{N}(0, 1), \quad \forall \psi \in \Psi$$

$\hookrightarrow$  Delta Method

Check if  
the assumption  
are satisfied

The corresponding confidence interval is

$$\hat{\psi} - z_{1-\alpha/2} \text{SE}(\hat{\psi}), \quad \hat{\psi} + z_{1-\alpha/2} \text{SE}(\hat{\psi})$$

The Central Limit Theorem provides such a solution for random samples, when  $\psi$  corresponds to the mean of each variable.

## R lab: approximate confidence intervals

```
M <- 100000; n <- 10; mat.ci <- matrix(NA, nrow = M, ncol = 2)
for(i in 1:M) { y <- rnorm(n, 5)
  se_z <- sqrt(var(y) / n) * qnorm(0.975)
  mat.ci[i,] <- mean(y) + se_z * c(-1, 1)}
mean(mat.ci[,1] < 5 & mat.ci[,2] > 5)
```

```
## [1] 0.91904
```

```
M <- 100000; n <- 100; mat.ci <- matrix(NA, nrow = M, ncol = 2)
for(i in 1:M) { y <- rnorm(n, 5)
  se_z <- sqrt(var(y) / n) * qnorm(0.975)
  mat.ci[i,] <- mean(y) + se_z * c(-1, 1)}
mean(mat.ci[,1] < 5 & mat.ci[,2] > 5)
```

```
## [1] 0.94676
```

## Confidence interval for a proportion

The method for approximate intervals can be readily used for confidence intervals on a proportion  $\pi$ , the success probability of a random sample of  $n$  binary variables,

$$Y_i \sim \mathcal{B}(1, \pi), \quad i = 1, \dots, n.$$

Here the pivot is

$$Z(\pi) = \frac{\bar{Y} - \pi}{\sqrt{\frac{\bar{Y}(1 - \bar{Y})}{n}}} \sim \mathcal{N}(0, 1), \quad \forall \pi \in (0, 1),$$

*SE OF BERNoulli DISTRIBUTION*

since  $\hat{\pi} = \bar{Y}$  and  $SE(\hat{\pi}) = \sqrt{\frac{\pi(1 - \pi)}{n}}$ , which is estimated by plugging-in  $\hat{\pi}$  in place of  $\pi$ .

## R lab: confidence interval for a proportion

```
M <- 100000; n <- 50; mat.ci <- matrix(NA, nrow = M, ncol = 2)
for(i in 1:M) { y <- rbinom(n, size = 1, prob = 0.25)
  p.hat <- mean(y)
  se_z <- sqrt(p.hat * (1 - p.hat) / n)
  se_qz <- se_z * qnorm(0.975)
  mat.ci[i,] <- mean(y) + se_qz * c(-1, 1)}
mean(mat.ci[,1] < 0.25 & mat.ci[,2] > 0.25)

## [1] 0.94063
```

## Confidence interval for a difference of means

An important application concerns the computation of the confidence interval for the difference between two means  $\delta = \mu_X - \mu_Y$ .

For two independent (and large) random samples, the approximate normal pivot is

$$Z(\delta) = \frac{\widehat{\delta} - \delta}{\text{SE}(\widehat{\delta})},$$

with  $\widehat{\delta} = \bar{X} - \bar{Y}$  and  $\text{SE}(\widehat{\delta}) = \sqrt{\text{SE}(\bar{X})^2 + \text{SE}(\bar{Y})^2}$ .

Again, for normal samples, exact solutions exist, both for the case of equal variances and for the case of unequal variances.