

# Lecture on Null Hypothesis Testing & Temporal Correlation

## CS – 590.21 Analysis and Modeling of Brain Networks

[Department of Computer Science](#)

University of Crete



# Acknowledgement – Resources used in the slides

## On statistical hypothesis test

- Yiannis Tsamardinos (University of Crete)
- William Morgan (Stanford University)

# Proving Your Hypothesis

## Mathematics

1. We already know a set of axioms & theorems, say  $K$
2. We want to show the theorem (hypothesis)  $H$
3. We show:  $K, \neg H \Rightarrow \text{False}$  (**contradiction**)
4. Thus, if we trust that  $K$  holds indeed,  $\neg H$  **cannot hold**, and  **$H$  must hold**

## Real World

1. We **already “know”**  $K$
2. We want to show a hypothesis  $H$ , e.g., “ $H$ : medicine  $A$  reduces the mortality of disease  $B$ ”
3. We gather data from the real world. We show that  $K, \neg H$  **makes it very unlikely to observe our data**
4. We conclude that  $\neg H$  **is very unlikely**  
We **reject  $\neg H$** , and **accept  $H$**

# Notation for the following slides

- **Random variables** are denoted with a *capital* letter, e.g.,  $X$
- **Observed quantities** of random variables are denoted with their corresponding *small* letter  $x$

## Example:

- $G$  is the expression level of a specific gene in a patient
- $g$  is the measured expression level of the gene in a ***specific patient***

# The Null Hypothesis

- The hypothesis we hope to accept is called the *Alternative Hypothesis*  
Sometimes denoted as  $H_1$
- The hypothesis we hope to reject, the **negation of the Alternative Hypothesis**, is called the *Null Hypothesis*  
Usually denoted by  $H_0$

Think of the  $H_0$  as the “status quo”

# Standard Single Hypothesis Testing

1. Form the Null & Alternative Hypothesis
2. Obtain related data
3. Find a **suitable test statistic  $T$**
4. Find the **distribution of  $T$  given the null**
5. Depending on the **distribution of  $T$**  & the **observed  $t_o = T(x)$**   
**decide to reject or not  $H_0$**

# Test Statistics

- Test statistic is a function of our data  $X$ :  $T(X)$  (  $X$ : random variable )  
*e.g., if  $X$  contains a single quantity (variable)  $T(X)$  the **mean** value of  $X$*
- $T$  is a random variable (since it depends on  $X$ , our data which is random variable)
- Denote with  $t_o = T(x)$  the observed value of  $T$  in our data
- *Instead of calculating  $P$  ( obtaining data similar to  $X$  |  $H_0$  )*  
Calculate  $P ( T \text{ similar to } t_o / H_0 )$
- If  $P ( T \text{ similar to } t_o / H_0 )$  is **very low, reject  $H_0$**

# Statistical significance tests

- Let's just think about a **two-tailed test**: “difference” or “no difference”
- **Null hypothesis**: there is no difference between A vs. B
- Assume that  $o_A$  &  $o_B$  are “sampled” independently from a “**population**”
- **Test statistic**: a function of the **sample data** on which the decision is to be based
$$t(o_1, o_2) = |e(o_1) - e(o_2)|$$

**e: evaluation metric**
- **Find the distribution of t under the null hypothesis**

**Assume that the null hypothesis is true**
- Where does the  $t(o_A, o_B)$  lie in this distribution?

If **it's somewhere unlikely**, that's evidence that the **null hypothesis is false**





“Welcome to Lake Wobegon, where all the women are strong, all the men are good-looking, and all the children are above average.”

- Garrison Keillor, *A Prairie Home Companion*

The Lake Wobegon Example: “Where all the children are above average!”

- Let  $X$  represent Weschler Adult Intelligence scores (WAIS)
- Typically,  $X \sim \mathbf{N(100, 15)}$  ( $\mu_0 = 100, \sigma = 15$ )
- Obtain data: **9 children** from Lake Wobegon population  
Their scores: {116, 128, 125, 119, 89, 99, 105, 116, 118}  
Average of the observations  $\bar{x} = 112.8$

Does **sample mean** provide strong evidence that **population mean  $\mu > 100$** ?

# One-Sample $z$ Test

1. Hypothesis statements

$$H_0: \mu = \mu_0$$

$H_a: \mu \neq \mu_0$  (two-sided) or

$H_a: \mu < \mu_0$  (left-sided) or

$H_a: \mu > \mu_0$  (right-sided)

3. Test statistic

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} \text{ where } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

4. P-value: convert  $z_{\text{stat}}$  to P value

A. Significance statement (usually not necessary)

# Example: Two-Sided Hypothesis Test “Lake Wobegon”

## 1. Formulation of the Hypotheses:

$$H_0: \mu = 100$$

$$H_a: \mu > 100 \text{ (one-sided)}$$

$$H_a: \mu \neq 100 \text{ (two-sided)}$$

## 2. Obtain data ...

Obtain data: **9 children** from Lake Wobegon population

Their scores: {116, 128, 125, 119, 89, 99, 105, 116, 118}

Average of the observations = 112.8

# Example: Two-Sided Hypothesis Test “Lake Wobegon”

## 3. Test statistic

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{9}} = 5$$

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{112.8 - 100}{5} = 2.56$$

## Classical CLT [\[ edit \]](#)

Let  $\{X_1, \dots, X_n\}$  be a random sample of size  $n$  — that is, a sequence of independent and identically distributed random variables drawn from distributions of expected values given by  $\mu$  and finite variances given by  $\sigma^2$ . Suppose we are interested in the sample average

$$S_n := \frac{X_1 + \dots + X_n}{n}$$

of these random variables. By the law of large numbers, the sample averages converge in probability and almost surely to the expected value  $\mu$  as  $n \rightarrow \infty$ . The classical central limit theorem describes the size and the distributional form of the stochastic fluctuations around the deterministic number  $\mu$  during this convergence. More precisely, it states that as  $n$  gets larger, the distribution of the difference between the sample average  $S_n$  and its limit  $\mu$ , when multiplied by the factor  $\sqrt{n}$  (that is  $\sqrt{n}(S_n - \mu)$ ), approximates the normal distribution with mean 0 and variance  $\sigma^2$ . For large enough  $n$ , the distribution of  $S_n$  is close to the normal distribution with mean  $\mu$  and variance  $\frac{\sigma^2}{n}$ . The usefulness of the theorem is that the distribution of  $\sqrt{n}(S_n - \mu)$  approaches normality regardless of the shape of the distribution of the individual  $X_i$ . Formally, the theorem can be stated as follows:

# Central Limit Theory

Establishes that, in most situations, when **independent random variables are added**, their **properly normalized sum tends toward a normal distribution** even if the **original variables themselves are not normally distributed.**

1. A **sample is obtained containing a large number of observations**, each observation being randomly generated in a way that does not depend on the values of the other observations.
2. If step 1 is performed many times, **the computed values of the average will be distributed according to a normal distribution.**

Example: Flip a coin many times. The probability of getting a given number of heads in a series of **K** flips will approach the normal distr. with mean  $=K/2$



## Classical CLT [\[ edit \]](#)

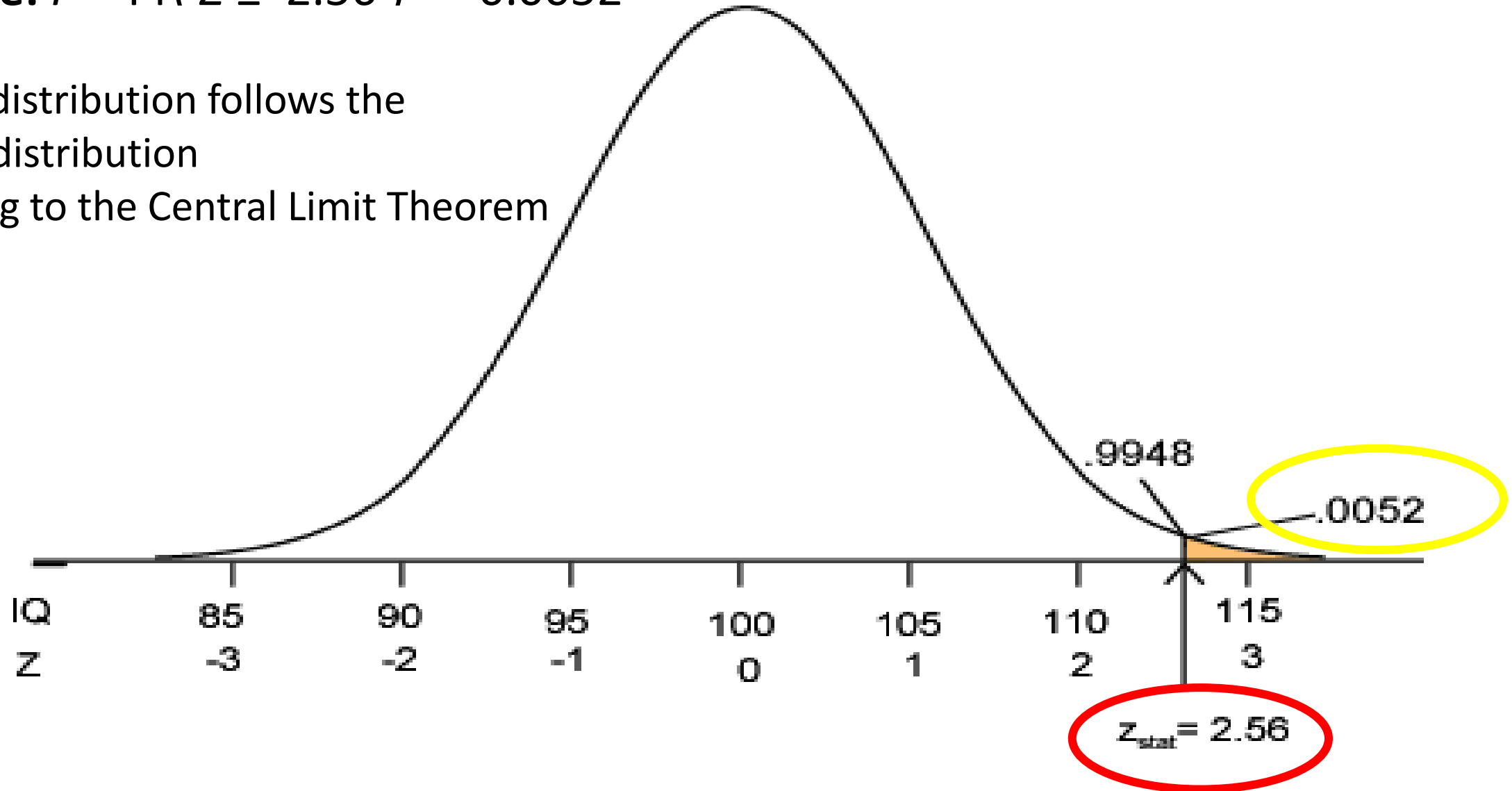
Let  $\{X_1, \dots, X_n\}$  be a random sample of size  $n$  — that is, a sequence of independent and identically distributed random variables drawn from distributions of expected values given by  $\mu$  and finite variances given by  $\sigma^2$ . Suppose we are interested in the sample average

$$S_n := \frac{X_1 + \dots + X_n}{n}$$

of these random variables. By the law of large numbers, the sample averages converge in probability and almost surely to the expected value  $\mu$  as  $n \rightarrow \infty$ . The classical central limit theorem describes the size and the distributional form of the stochastic fluctuations around the deterministic number  $\mu$  during this convergence. More precisely, it states that as  $n$  gets larger, the distribution of the difference between the sample average  $S_n$  and its limit  $\mu$ , when multiplied by the factor  $\sqrt{n}$  (that is  $\sqrt{n}(S_n - \mu)$ ), approximates the normal distribution with mean 0 and variance  $\sigma^2$ . For large enough  $n$ , the distribution of  $S_n$  is close to the normal distribution with mean  $\mu$  and variance  $\frac{\sigma^2}{n}$ . The usefulness of the theorem is that the distribution of  $\sqrt{n}(S_n - \mu)$  approaches normality regardless of the shape of the distribution of the individual  $X_i$ . Formally, the theorem can be stated as follows:

**P-value:**  $P = \Pr( Z \geq 2.56 ) = 0.0052$

Sample distribution follows the Normal distribution according to the Central Limit Theorem



**$P = .0052$**   $\Rightarrow$  it is unlikely the sample came from this null distribution  $\Rightarrow$  strong evidence **against  $H_0$**

# Example - Two-Sided $P$ -value: Lake Wobegon

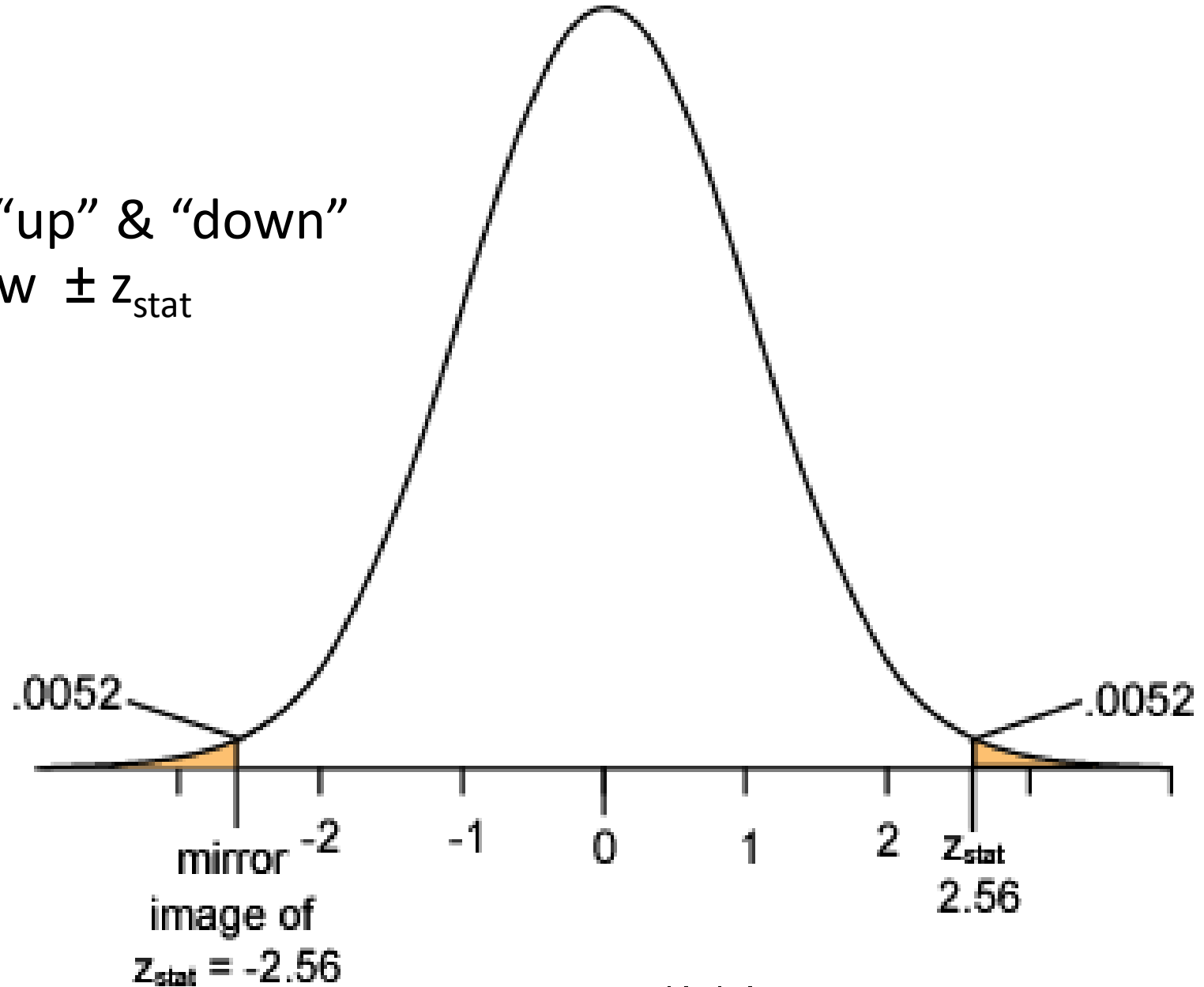
- $H_a: \mu \neq 100$

Considers random deviations “up” & “down” from  $\mu_0 \Rightarrow$  tails above & below  $\pm z_{\text{stat}}$

Thus, **two-sided P**

$$= 2 \times 0.0052$$

$$= 0.0104$$



# Conditions for z Test

1. **Population approximately Normal** or **large sample** (central limit theorem)
2. The population variance is known!

If the **population variance is unknown** (and therefore has to be estimated from the sample itself) & the **sample size is not large** ( $n < 30$ ), the **Student's *t*-test** may be more appropriate.

Another Example

- Background knowledge: Breast Cancer is related to mutations in genes BRCA1 & BRCA2
- Hypothesis: Gene G is expressed differently in breast cancer patients with mutation in BRCA1 than BRCA2
- Data: Obtained 7 patients with BRCA1 mutation & 8 with BRCA2 mutation

Patient number	Expression Level of Gene G $x_i^1$	Have mutation in BRCA1 or not $x_i^2$
1 ( $x_1$ )	98.2244	1
2	69.6810	1
3	118.4339	1
4	115.2322	1
5	150.7729	1
6	117.7385	1
7	80.6921	1
8	142.8455	2
9	156.8692	2
10	151.9287	2
11	147.3357	2
12	131.2094	2
13	150.3127	2
14	147.0670	2
15 ( $x_{15}$ )	122.3306	2

Hedenfalk et al. N Engl J Med. 2001  
Feb 22;344(8):539-48.

# 1. Form the Null Hypothesis

- Gene G is expressed differently in breast cancer patients with mutation in BRCA1 than BRCA2

## Mathematically

- $\mu_1$ : be the mean expression level of gene G in patients with BRCA1 mutation
- $\mu_2$ : be the mean expression level of gene G in patients with BRCA2 mutation

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

## 2. Obtain data....

### DATA: BY THE NUMBERS

NUMBER OF YEARS TO GET DATA: 3

YES!  
FINALLY!



NUMBER OF YEARS TO INTERPRET DATA: 2

what does it  
all mean??



NUMBER OF YEARS TO WRITE ABOUT DATA: 1.5

blah blah  
blah blah...



NUMBER OF SLIDES TO PRESENT DATA: 1

RESULTS

that's  
it?



JORGE CHAM © 2004



### 3. Find a suitable test statistic $T$ (Example)

$$T(x) = \frac{m_1 - m_2}{s \sqrt{\frac{1}{k} + \frac{1}{l}}}$$

Unpaired Two Sample t-test

- The larger the difference of the two means, the larger the statistic
- The larger our sample, the larger the statistic
- The smaller the sample variance, the larger the statistic

**So  $T$  will be quite large (in absolute value), when we can confidently say  $H_0$  does not hold**

3. Find a suitable test statistic  $T$  (cont'd)

$$T(x) = \frac{m_1 - m_2}{s \sqrt{\frac{1}{k} + \frac{1}{l}}}$$

**Unpaired Two Sample t-test**

$$k = \#\{x_i : x_i^2 = 1\}, \quad l = \#\{x_i : x_i^2 = 2\}$$

$$m_1 = \frac{1}{k} \sum_{\{x_i : x_i^2 = 1\}} x_i^1, \quad m_2 = \frac{1}{l} \sum_{\{x_i : x_i^2 = 2\}} x_i^1, \quad m = \frac{1}{k+l} \sum_{\{x_i\}} x_i^1$$

$$s = \sqrt{\frac{1}{k+l-1} \sum (x_i^1 - m)^2}$$

### 3. Find the distribution of $T$ (cont'd)

For the test of this specific example, we will make the following assumptions:

- The **data in both groups are distributed normally** around a mean value  $\mu_1, \mu_2$  respectively
- Their **variance is the same in both groups**
- Each patient was **sampled independently**

and most importantly that **THE NULL HYPOTHESIS HOLDS**

**This is an assumption for ALL tests!**

Then

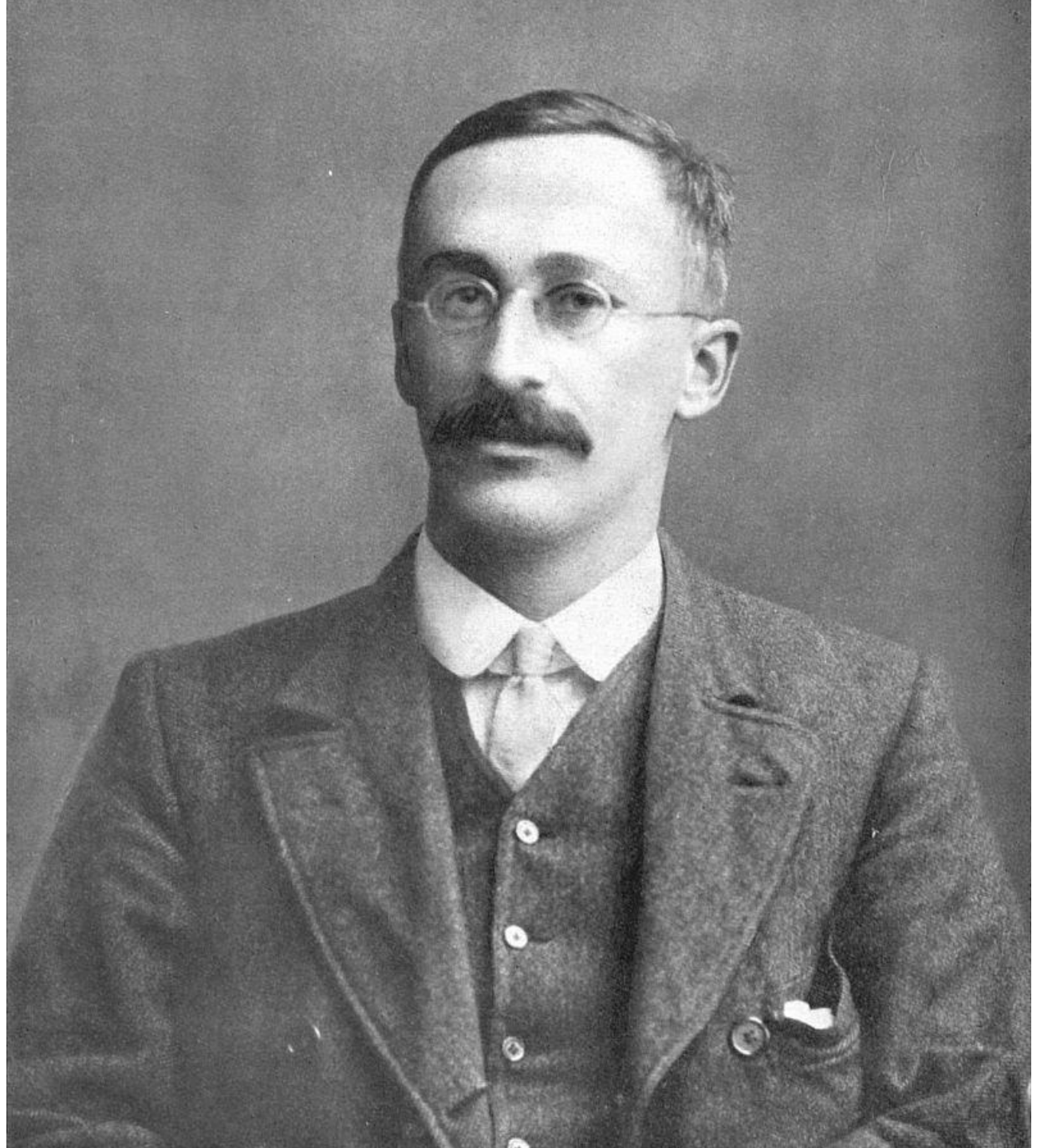
$T(X)$  has a probability density function of:

$$p(t | H_0) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

where the **degrees of freedom of the test  $\nu$**  is

$15 - 2 = 13$  (number of patients - 2)

The  $t$ -statistic was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Dublin. "Student" was his pen name.



## Sampling distribution [\[ edit \]](#)

## One sample T-distribution

Let  $x_1, \dots, x_n$  be the numbers observed in a sample from a continuously distributed population with expected value  $\mu$

$$\bar{x} = \frac{x_1 + \dots + x_n}{n},$$

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

The resulting *t*-value is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

The *t*-distribution with  $n - 1$  degrees of freedom is the sampling distribution of the *t*-value when the samples consist of  $n$  observations from a normally distributed population. Thus for inference purposes *t* is a useful "pivotal quantity" in the case when the mean and variance of the population are unknown.

# Student t-distribution (basics)

## Probability density function [\[ edit \]](#)

Student's **t-distribution** has the [probability density function](#) given by

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

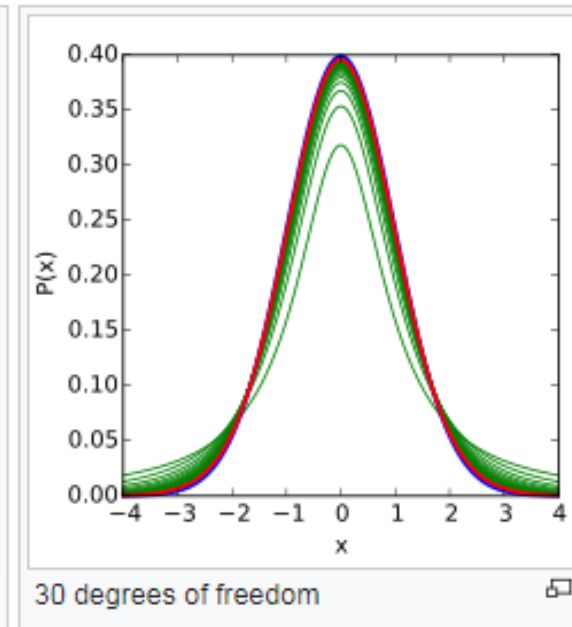
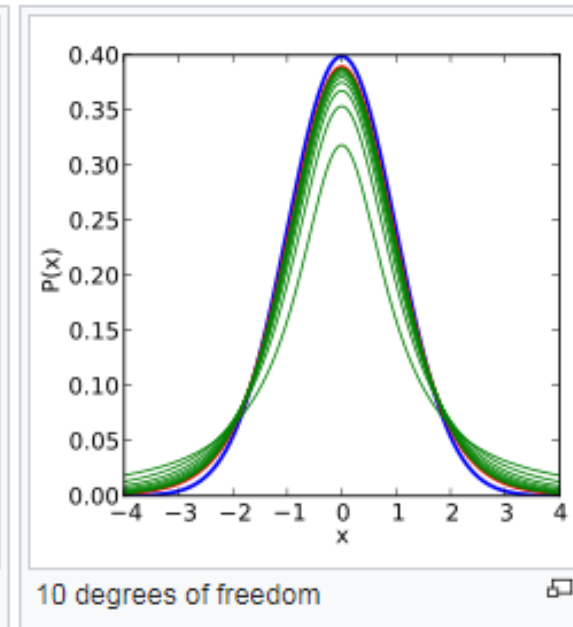
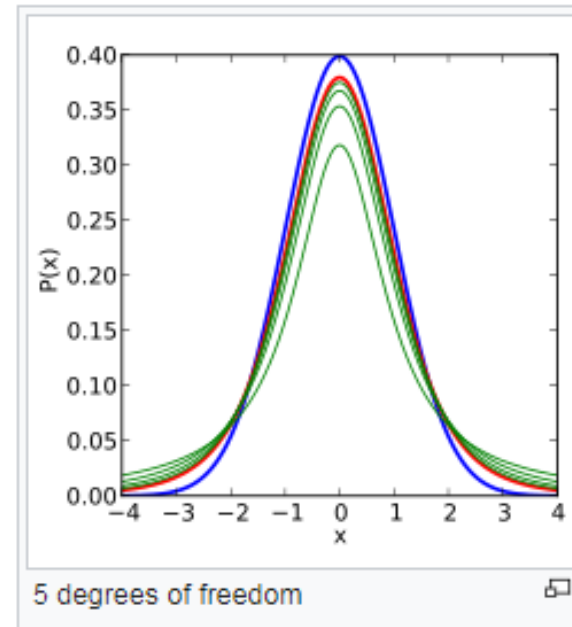
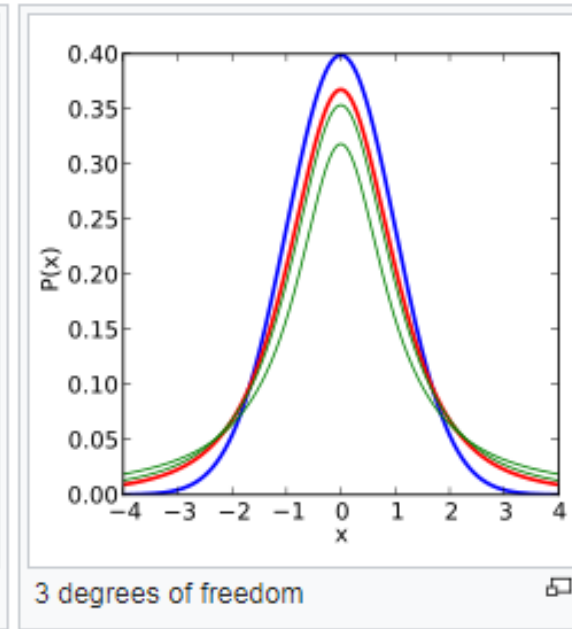
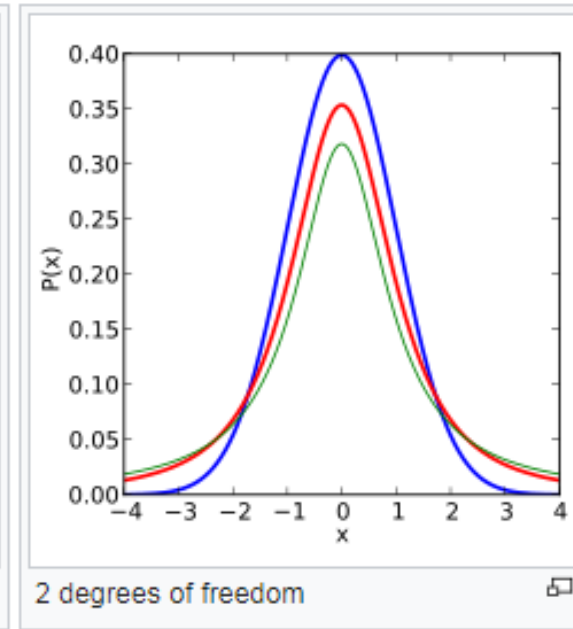
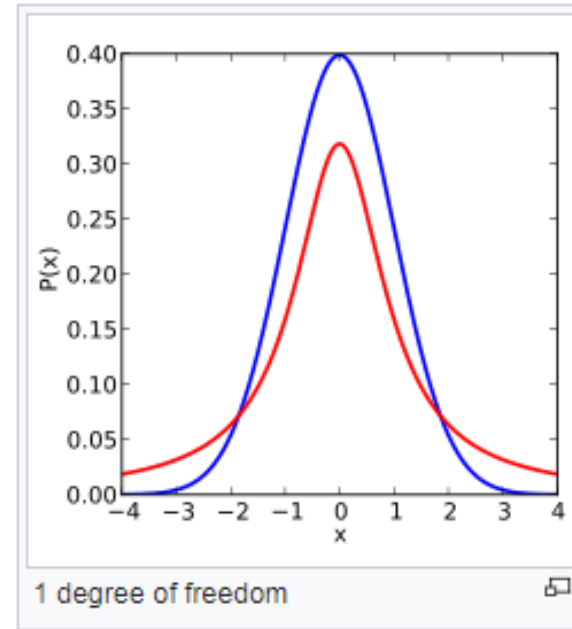
where  $\nu$  is the number of *degrees of freedom* and  $\Gamma$  is the [gamma function](#). This may also be written as

$$f(t) = \frac{1}{\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

# t-distribution (basics)

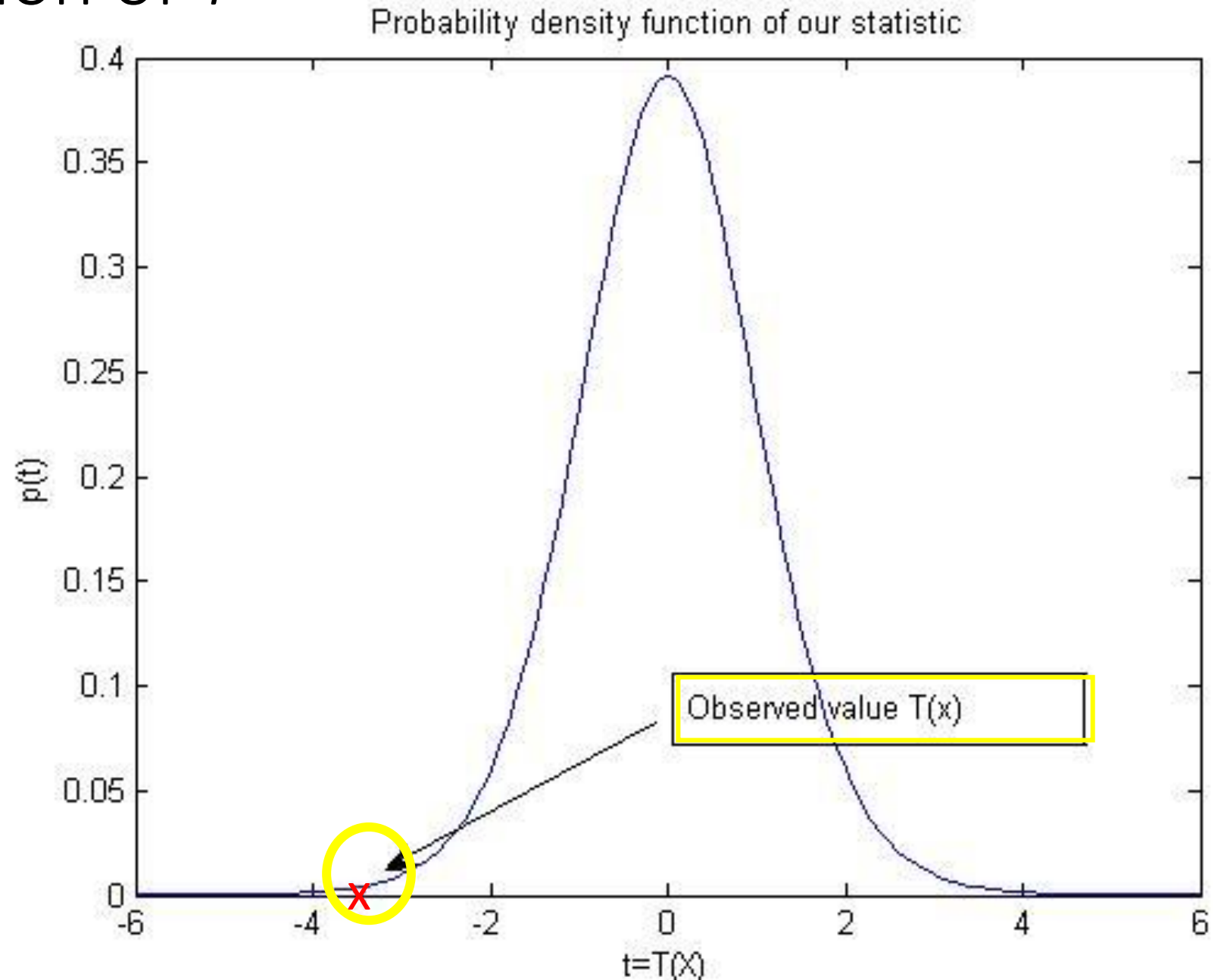
Density of the  $t$ -distribution (red) for 1, 2, 3, 5, 10, and 30 degrees of freedom compared to the standard normal distribution (blue).

Previous plots shown in green.



### 3. Find the distribution of $T$

Example





## 4. Decide on a Rejection Region

- Decide on a **rejection region**  $\Gamma$  in the range of our statistic
- **If  $t_o \in \Gamma$ , then reject  $H_0$**
- If  $t_o \notin \Gamma$ , then **do not reject  $H_0$**   
accept  $H_1$ ?

Since the pdf of  $T$  ***when the null hypothesis holds is known***,

$P(T \in \Gamma | H_0)$  can be calculated

## 4. Decide on a Rejection Region

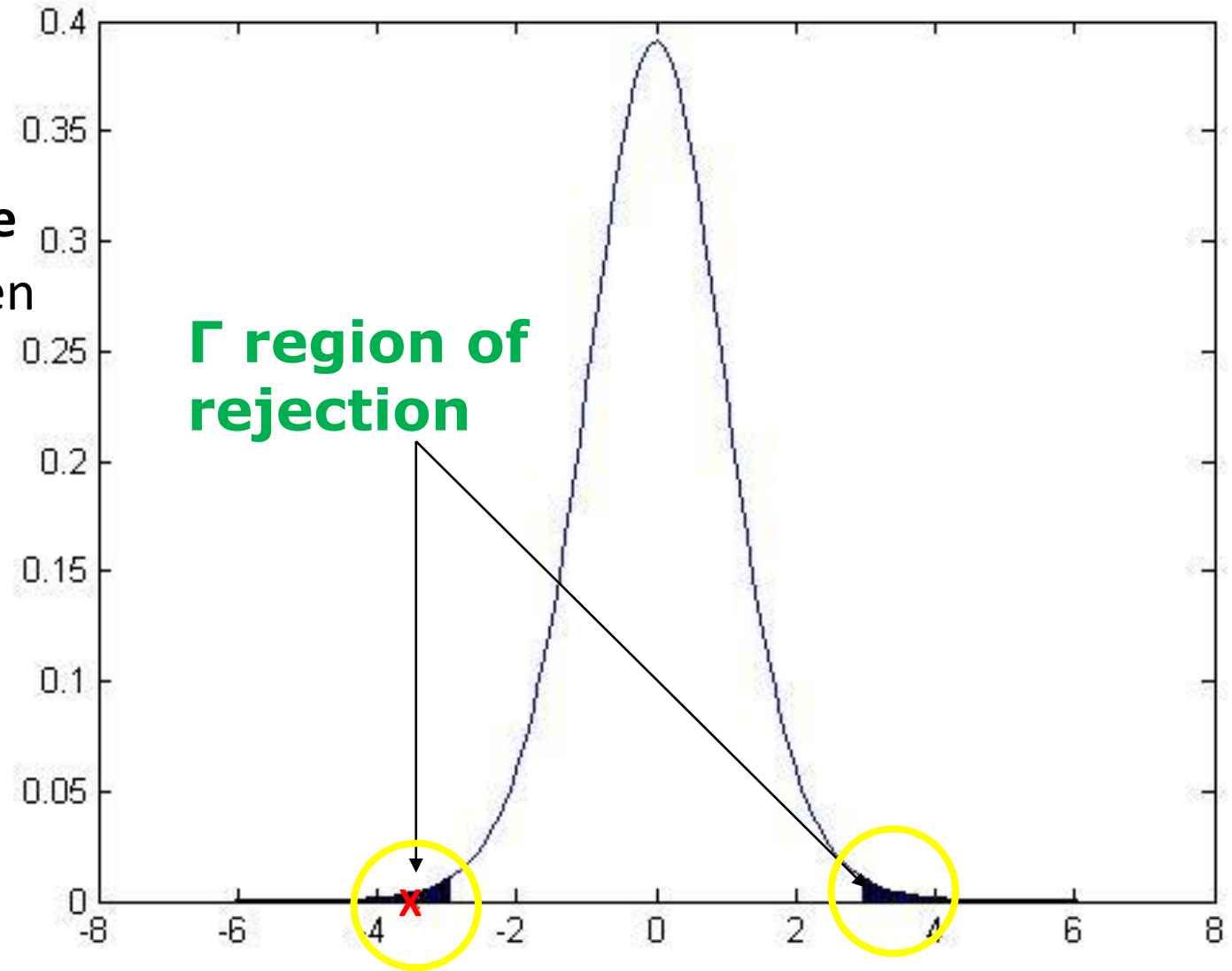
- If  $P(T \in \Gamma \mid H_0)$  **is too low**, we know we are **safely rejecting  $H_0$**
- What should be our rejection region in our example?

# 4. Decide on a Rejection Region

Where **extreme values of  $t_o$**  are:

- **unlikely to come from when  $H_0$  is true**
- could come with high probability, when  $H_0$  is false

$P(T \in \Gamma / H_0)$  is the area of the shaded region (can be calculated)



# Rejection Procedure

- Pre-select a probability threshold  $\alpha$
- Find a **rejection region**  $\Gamma = \{ t: |t| > c \}$ , such that  $\mathbf{P}(T \in \Gamma \mid H_0) = \alpha$
- Decide
  - **Reject  $H_0$** , if  $t_0 \in \Gamma$  (recall:  $t_0$  is the observed  $T$  in our data)
  - **Accept  $H_0$** , otherwise

What values do we usually use for  $\alpha$  in science?

0.05 is typical

Smaller ones are also used: 0.01 , 0.001

**When  $t_0 \in \Gamma$  we say the finding is statistically significant at significance level  $\alpha$**

# Issues to be Considered

- When there exist two or more tests that are appropriate in a given situation, how can the tests be compared to decide which should be used?
- If a **test is derived under specific assumptions** about the **distribution of the population being sampled**, how well will the test procedure work **when the assumptions are violated?**

# Parametric versus non-Parametric Tests

- **Parametric test**

Makes the assumption that the **data are sampled** from a **particular class of distributions**

It then becomes **easier to derive the distribution of the test statistic**

- **Non-Parametric test**

**No assumption** about a particular class of distributions

# Permutation Testing

- Often in biological data, we do ***not know much about the data distribution***
- How do we obtain the distribution of our test statistic?
- Great idea in statistics: permutation testing
- Recently practical because it requires computing power (or a lot of patience)

# Permutation Testing

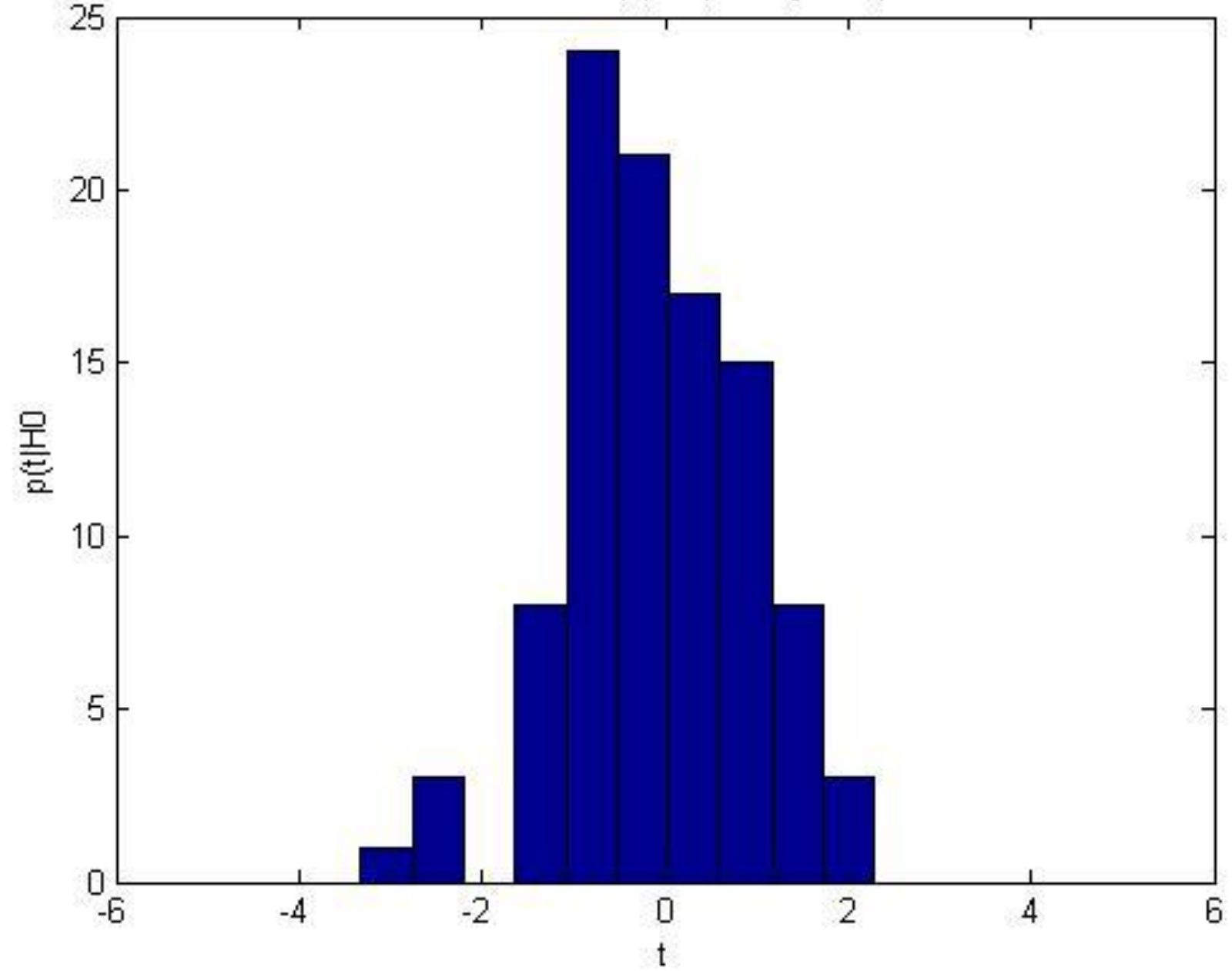
1. In our first example, we want to calculate  $p(t | H_0)$
2. **If  $H_0$** , then it does **not** matter which group each value  $x_i^1$  comes from
3. Then, if we **permute the group labels**, we would get a value for our test statistic **given the null hypothesis holds**
4. If we get a lot of such values, we can estimate (approximate)  $p(t | H_0)$



# Permutation Testing Revisited

- Decide what can be permuted, if the null hypothesis is true
- For all (as many as possible) permutations of the data, calculate the **test statistic on the permuted data:  $t_B$**
- Estimated p-value =  $\#\{ |t_B| \geq |t_o| \} / \#B$

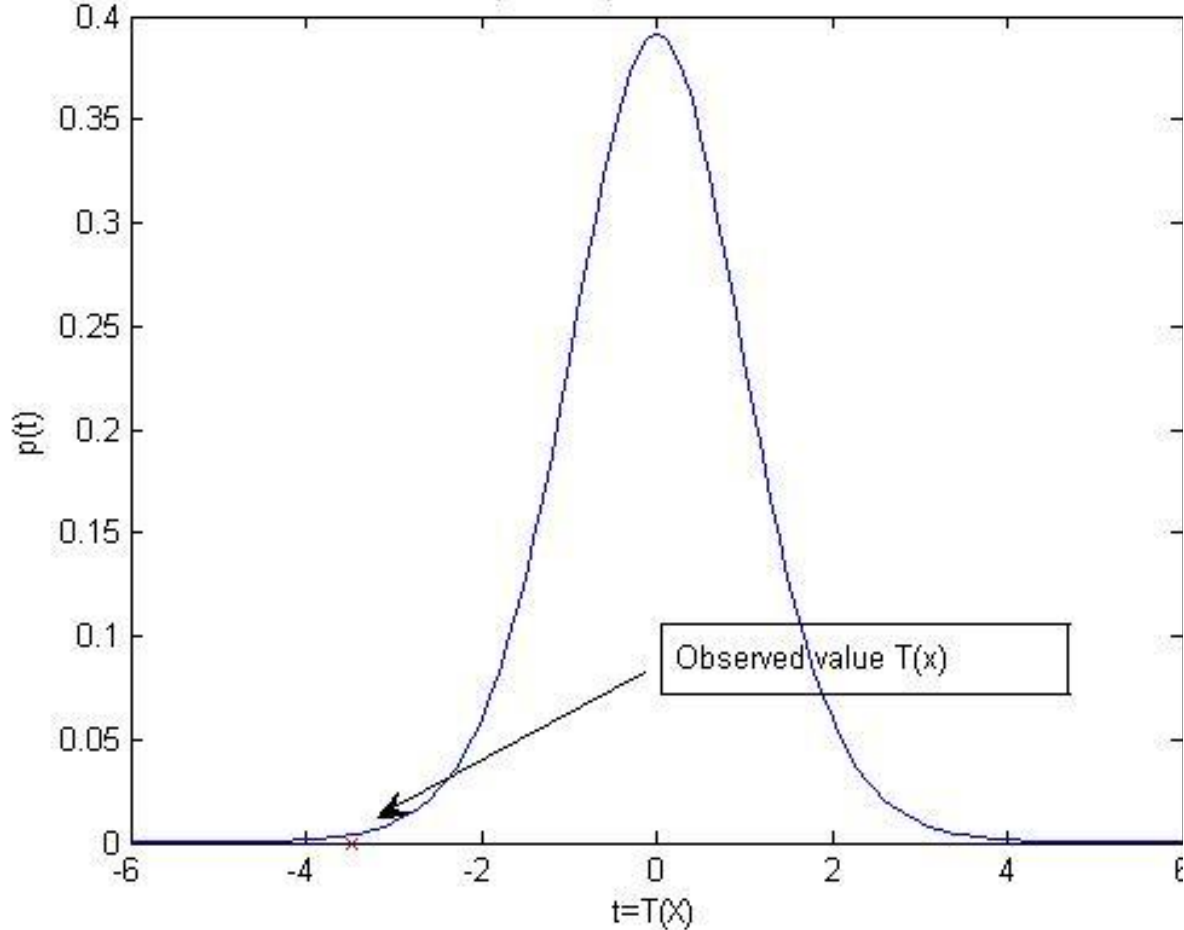
Permutation Estimate of  $p(t|H_0)$  using 100 permutations



Estimated distribution from our data: **100 permutations**

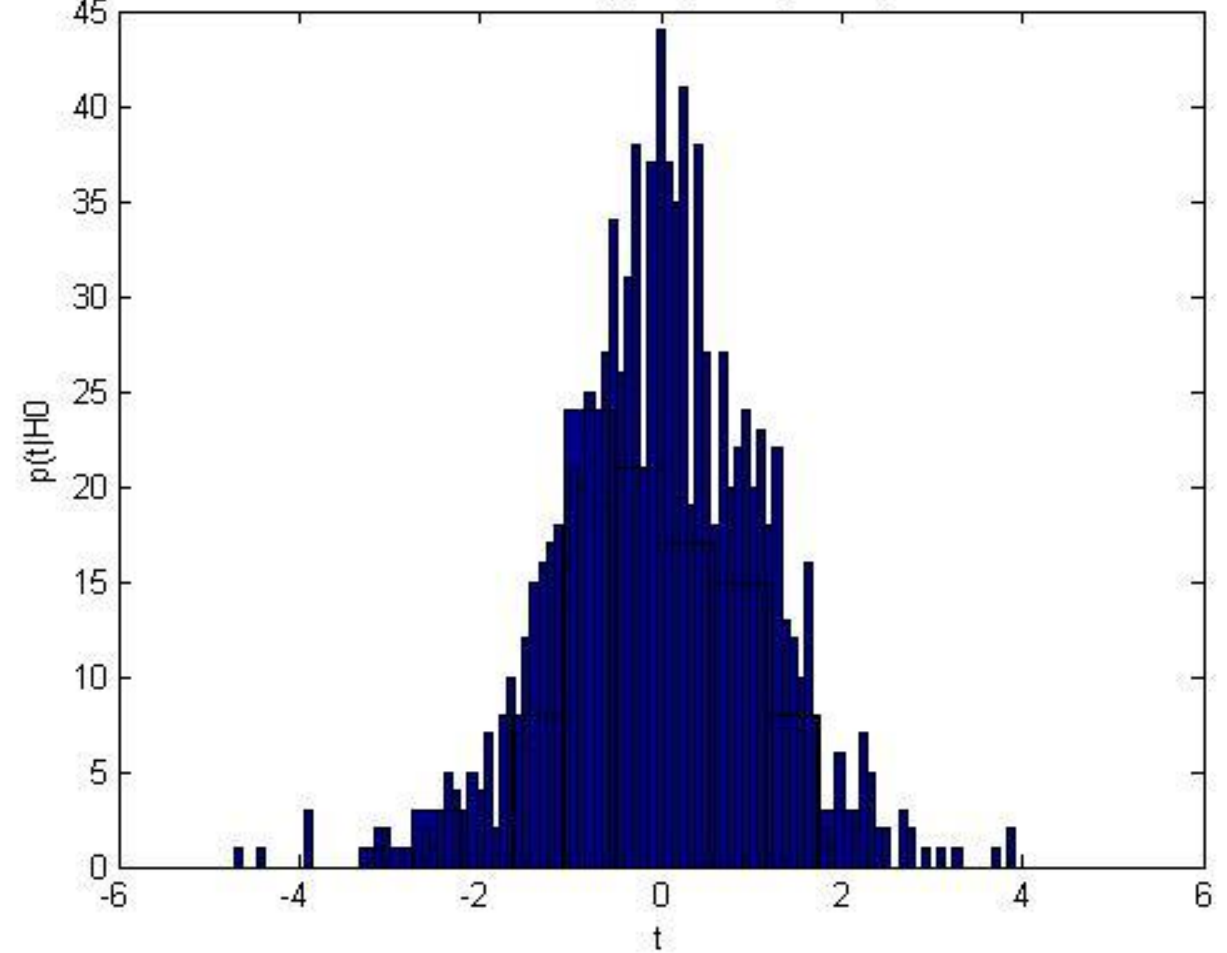
# Does It Really Work?

Probability density function of our statistic



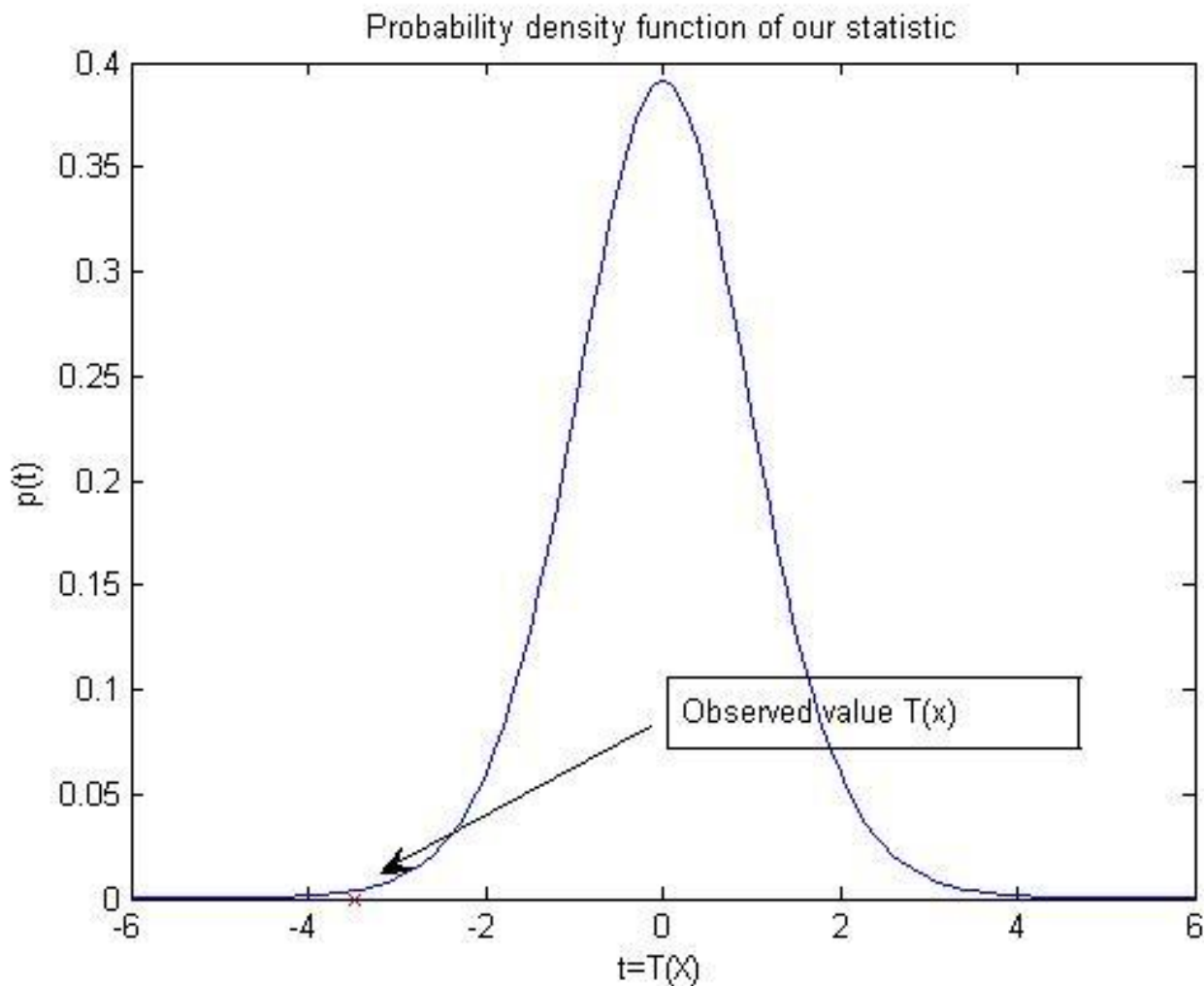
True distribution **calculated**  
**theoretically**

Permutation Estimate of  $p(t|H_0)$  using 1000 permutations

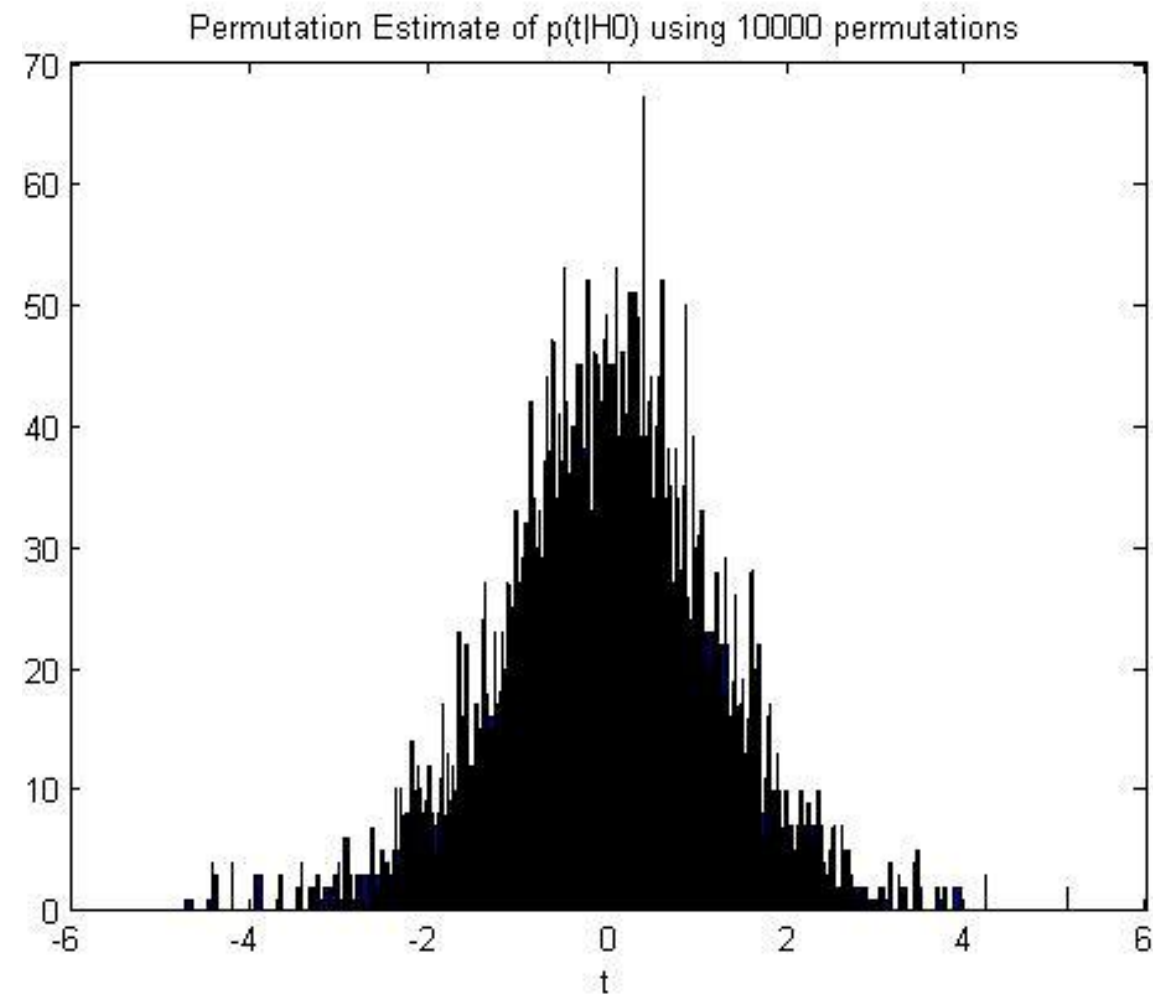


Estimated distribution from our data: **1,000**  
**permutations**

# Does It Really Work?

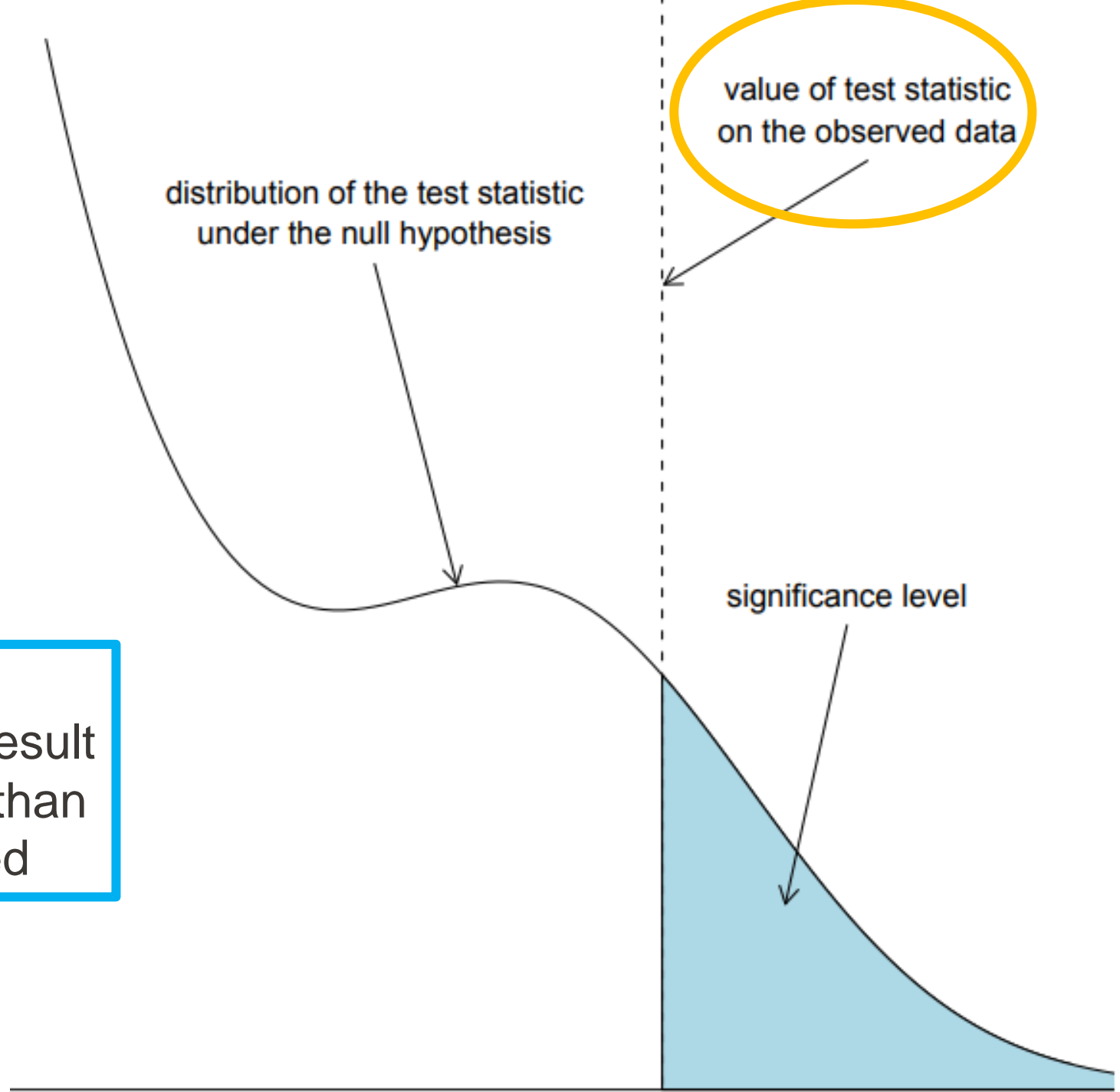


True distribution calculated theoretically



Estimated distribution from our data: **10,000 permutations**

**p-value** is defined as the probability of obtaining a result equal to or more extreme than what was actually observed



## The Significance Level

- The area to the right of  $t(o_A, o_B)$  is the “significance level”—the probability that some  $t^* \geq t(o_A, o_B)$  would be generated *if the null hypothesis were true*.
  - Also called the **p-value**.

Small values suggest the null hypothesis is false, given the observation of  $t(o_A, o_B)$ .

- Corollary: all else being equal, a large difference between  $e(o_A)$  and  $e(o_B)$  yields a smaller significance level (as one would hope!).
- Values below 0.05 are typically considered “good enough.”

So all we have to do is calculate the distribution of  $t$ .

# Calculating the Distribution

The classical approach:

- Keep adding assumptions until we arrive at a known distribution which we can calculate analytically.
- E.g.: Student's t-test.
  - Assume that  $e(o_A)$  and  $e(o_B)$  are sample means from a bivariate Normal distribution with zero covariance. Then we know  $t$  is distributed according to Student's t-distribution if the null hypothesis is true.
- Back in the stone age, computing with rocks and twigs, making those assumptions made the problem tractable.
- But the problem with this approach is that you may falsely reject the null hypothesis if one of the additional assumptions is violated. (Type I error.)



What you  
SHOULD  
do

- Simulate the distribution using a *randomization* test.
- It's just as good as analytical approaches, even when the analytical assumptions are met! (Hoeffding 1952)
- And it's better when they're not. (Noreen 1989)
- Best of all: dirt simple.

Intuition:

- Erase the labels "output of  $A$ " or "output of  $B$ " from all of the observations.
- Now consider the population of every possible labeling. (Order relevant.)
- If the systems are really different, the observed labeling should be unlikely under this distribution.



# Statistical Errors

- **Type 1 Errors**

- Rejecting  $H_0$  when it is actually true

- Concluding a difference when one does ***not actually exist***

- **Type 2 Errors**

- Accepting  $H_0$  when it is actually false (e.g. previous slide)

- Concluding no difference when **one does exist**

Errors can occur due to **biased/inadequate sampling, poor experimental design** or the use of **inappropriate/non-parametric** tests.

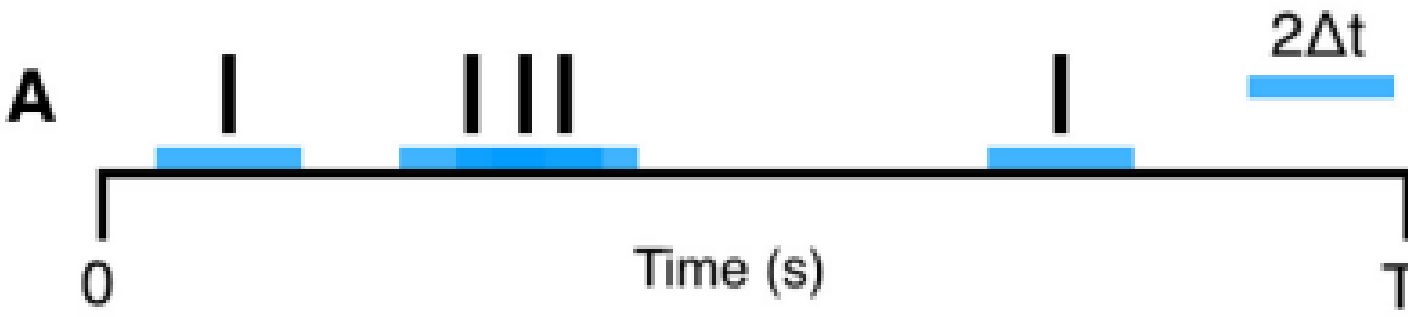
# Regarding the Choice of a Test

**When we cannot reject  $H_0$ , it does not mean  $H_1$  holds!**

- It could be that we do **not have enough power**, i.e.,  
 $H_1$  is not that “**different enough**” from  $H_0$  to distinguish it with the **given sample size**  
of all possible tests for a hypothesis choose the one with the maximum power

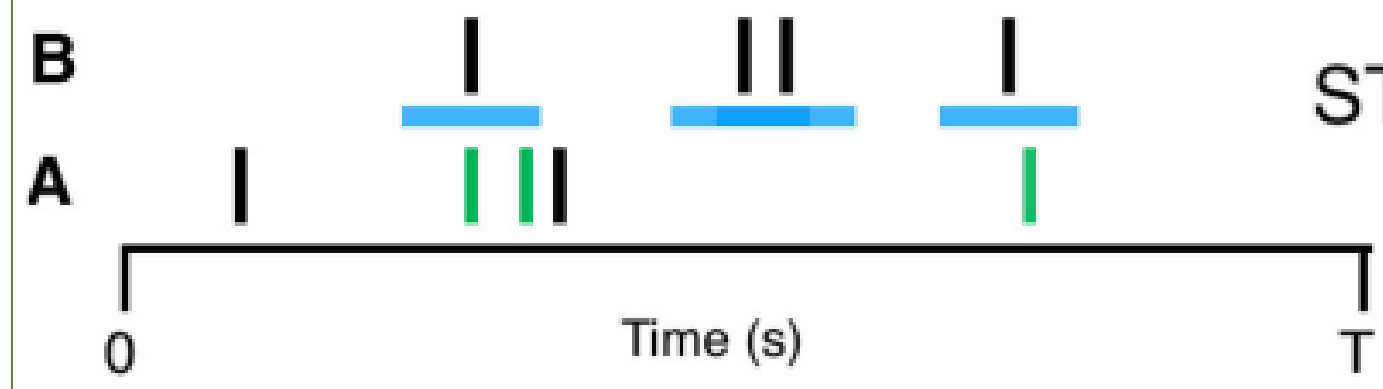
**Power analysis** methods need to be employed.

$T_A$ : the proportion of total recording time which lies within  $\pm\Delta t$  of any spike from A.  $T_B$  calculated similarly.



$T_A$  is given by the fraction of the total recording time (black) which is covered (tiled) by blue bars. Here  $T_A$  is  $1/3$ .

$P_A$ : the proportion of spikes from A which lie within  $\pm\Delta t$  of any spike from B.  $P_B$  calculated similarly.



$P_A$  is the number of green spikes in A (3) divided by the total number of spikes in A (5). Here  $P_A$  is  $3/5$ .

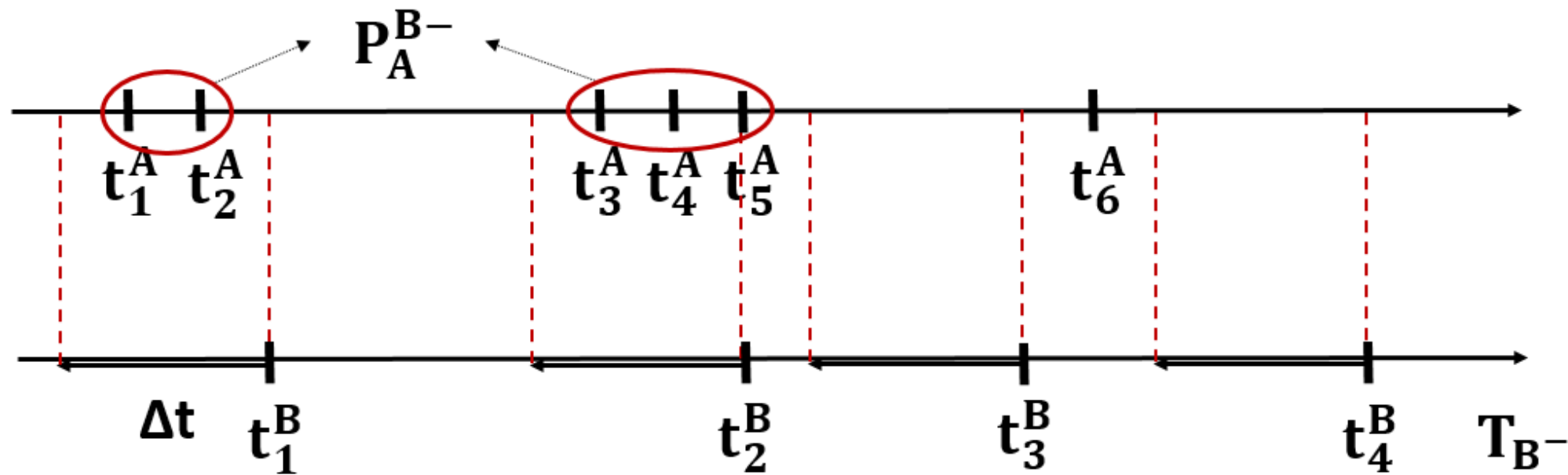
$$STTC = \frac{1}{2} \left( \frac{P_A - T_B}{1 - P_A T_B} + \frac{P_B - T_A}{1 - P_B T_A} \right)$$

# Directional STTC Temporal Correlation Metric

Extended STTC metric to take into consideration the order of the correlation of the spike trains of two neurons

**Directional STTC<sub>AB</sub>** represents a measure of the chance that firing events of A will precede firing events of B

$$STTC_{AB} = \frac{1}{2} \left( \frac{P_A^{B-} - T_{B-}}{1 - P_A^{B-} T_{B-}} + \frac{P_B^{A+} - T_{A+}}{1 - P_B^{A+} T_{A+}} \right)$$



$P_A^{B-}$ : fraction of firing events of A that occur within an interval  $\Delta t$  prior to firing events of B  
 $T_{B-}$ : fraction of total recording time covered by the intervals  $\Delta t$  **prior to each spike of B**  
 $\Delta t$ : specific lag (input in directional STTC)

# Advantages of Directional STTC vs. other correlation metrics

$$STTC_{AB} = \frac{1}{2} \left( \frac{P_A^{B-} - T_{B-}}{1 - P_A^{B-} T_{B-}} + \frac{P_B^{A+} - T_{A+}}{1 - P_B^{A+} T_{A+}} \right)$$

**Relative spike-time shifts (lag parameter)**

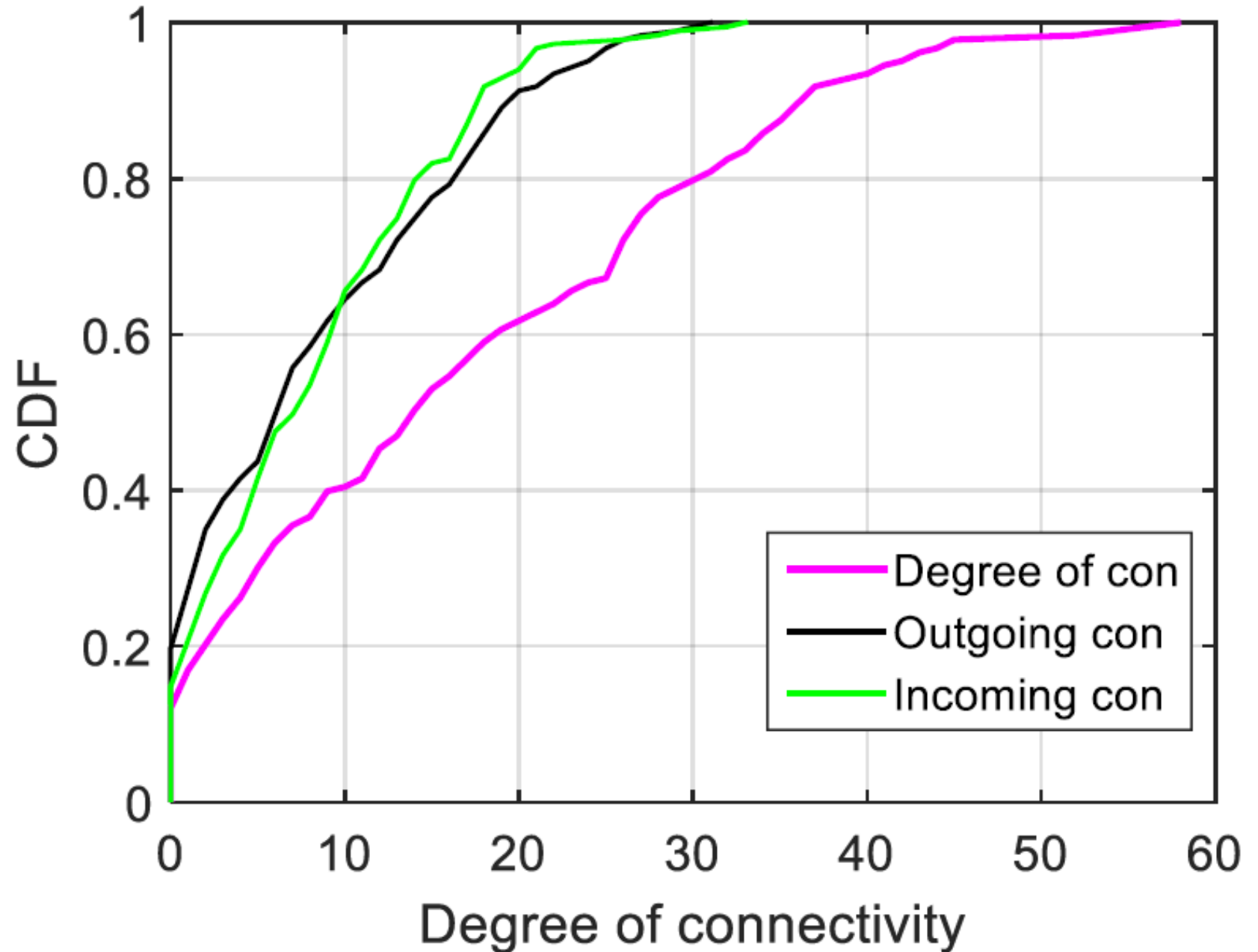
**Local fluctuations of neural activity or noise**

- accounting the amount of correlation expected by chance ( $T_{A+}$  &  $T_{B-}$ )

**The presence of periods without firing events**

- only the firing events contributes on the correlation metric of directional STTC ( $P_A^{B-}$  &  $P_B^{A+}$ )

# Example – Degree of connectivity considering the significant directional STTC edges



Conditional STTC (A->B | C)  $STTC_{AB}^C$

$$STTC_{AB}^C = \frac{1}{2} \left( \frac{\frac{N_{B^-A}^{CA}}{N_A} - T_{B^-}}{1 - \frac{N_{B^-A}^{CA}}{N_A} T_{B^-}} + \frac{\frac{N_{A+B}^{CA}}{N_B} - T_{A^+}}{1 - \frac{N_{A+B}^{CA}}{N_B} T_{A^+}} \right)$$

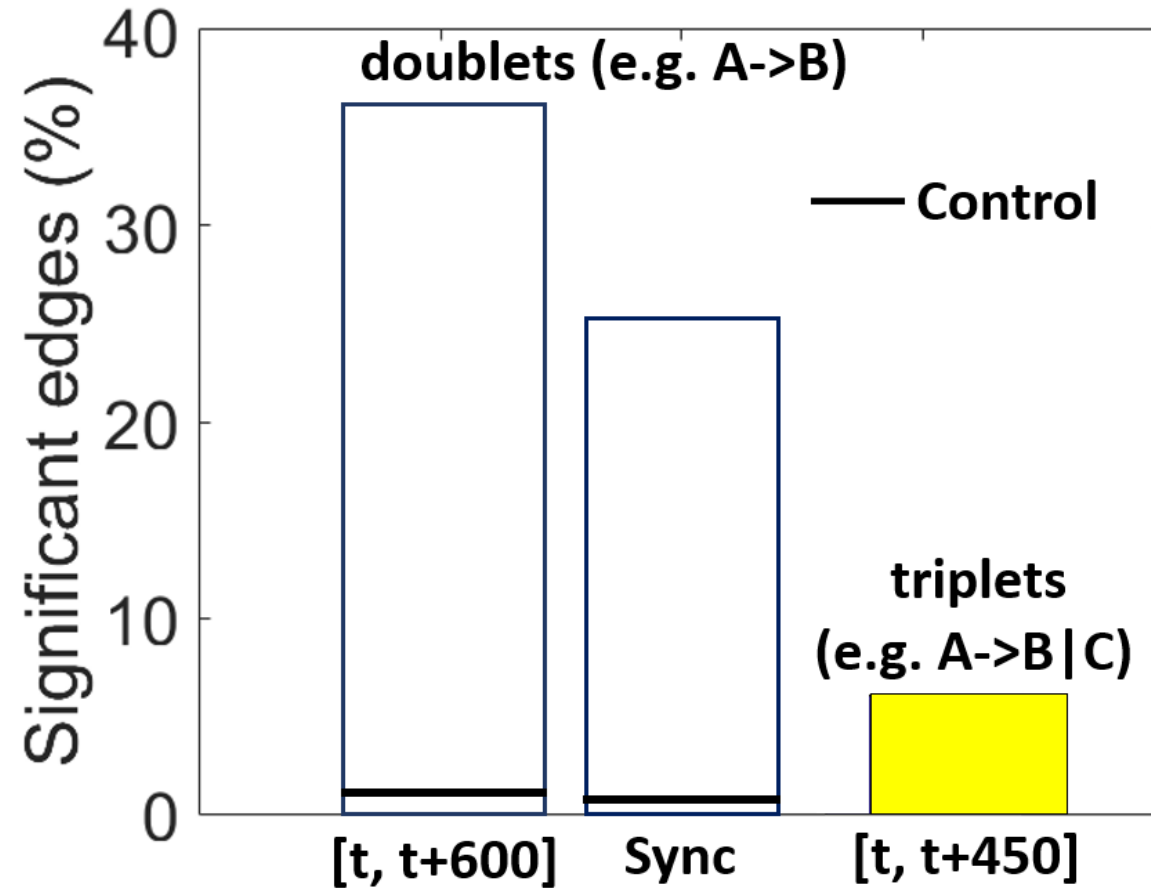
$N_A$  is the number of firing event in A &  $N_B$  is the number of firing event in B.

$T_{A^+}$  is the fraction of the total recording time which is covered by the tiles  $+\Delta t$  after each spike of A, that fall within the tiles  $\Delta t$  after each spike of C.

$T_{B^-}$  is the fraction of the total recording time which is covered by the tiles  $\Delta t$  before each spike of B.



# Significant Motifs



**Directional edge** “A→B” indicate that firing events of A proceed firing events of B by a specific lag  
Circular shuffling by random delays of the neural traces is used to assess **significance** of directional edges

**Null distribution:** STTC values for the circular shifted neurons

**Significant edge:** real STTC value higher than 3 standard deviation of null distribution

# Example: Null distribution test for directional STTC

## For a given pair (A,B)

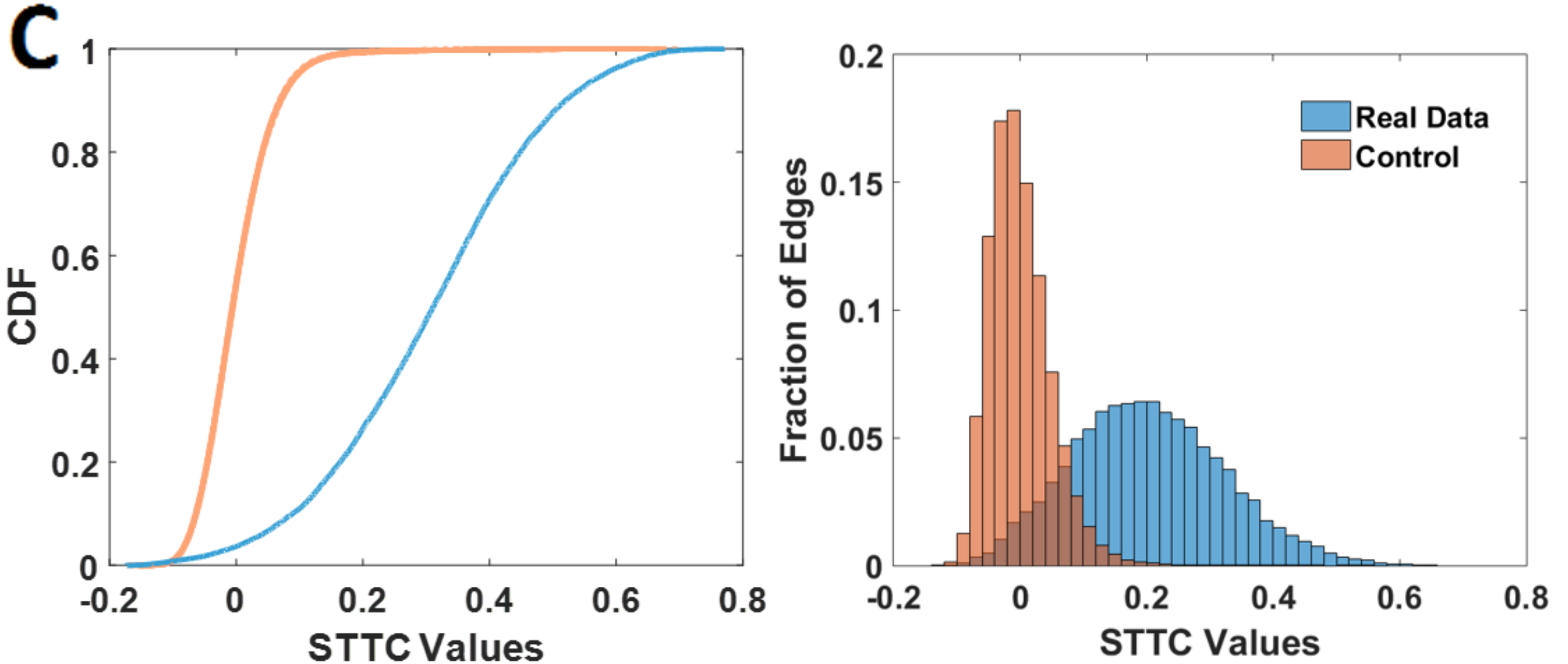
1. Circular shift the spike train of the neuron A,  $A^1$
2. Estimate the directional STTC( $A^1, B$ )
3. Repeat the above steps 100 times
4. Estimate the mean & standard deviation of the obtained STTC values
5. The statistical significant threshold ( $thr$ ) = mean + 3 std dev

If the directional STTC (A, B) >  $thr$  , the edge (A,B) is statistically significant.

The criterion can be strengthened with more repetitions (e.g., **1000**), a larger number of std dev (e.g., **5**).

# Strengthen the Criterion of Significant Edge

- Reject the null hypothesis test
- The total number of spikes of A within a STTC lag of spikes of B is above 3.
- The total number of spikes of B within a STTC lag of spikes of A is above 3.



1. Control group: each neuron trace is circular shifted by random delay
2. For each pair of 'shifted' neurons estimate the directional STTC & null distribution
3. Identify the significant edges of the control group

The **real neuron traces** appear **higher** values of directional STTC & percentage of significant edges

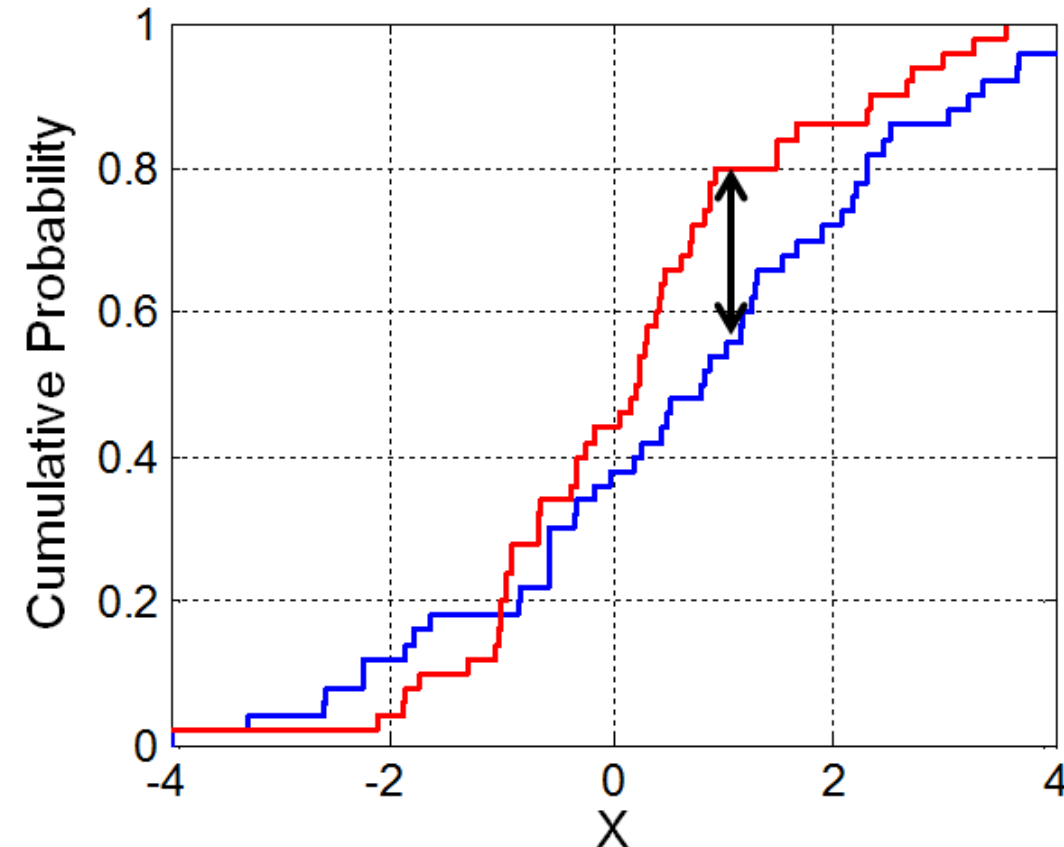
# Kolmogorov-Smirnov (K-S) Test

- **Non-parametric** test of the equality of **continuous 1D** probability distributions
- Quantifies a **distance between two distribution** functions
- Can serve as a **goodness of fit test**

- **Null hypothesis**

$H_0$ : Two samples drawn from **populations with same distribution**

**The maximum absolute difference between the two CDFs**



# Kolmogorov-Smirnov (K-S) Test

- **Non-parametric** test of the equality of **continuous 1D** probability distributions
- Quantifies a **distance between two distribution** functions
- Can serve as a **goodness of fit test**

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

where  $F_{1,n}$  and  $F_{2,m}$  are the empirical distribution functions

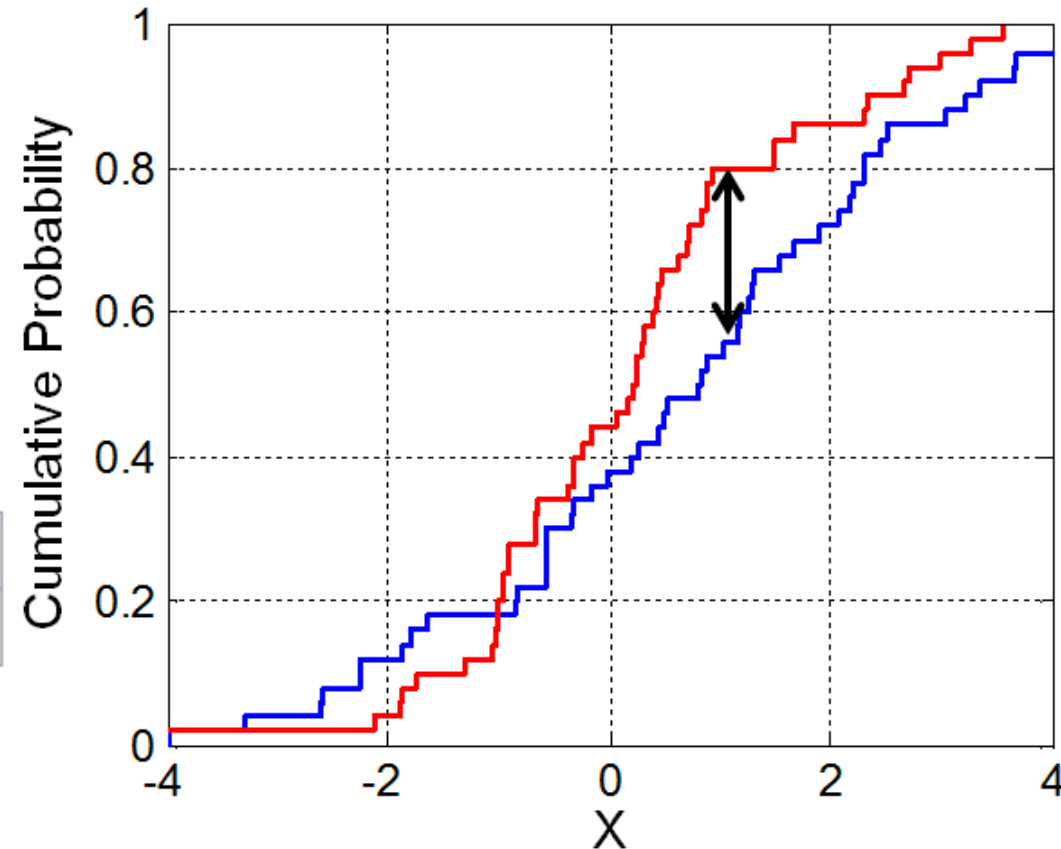
The null hypothesis is rejected at level  $\alpha$  if

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{n \cdot m}} \quad | \quad \mathbf{n \& m: size of the sample datasets}$$

$\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

and in general by

$$c(\alpha) = \sqrt{-\frac{1}{2} \ln\left(\frac{\alpha}{2}\right)}.$$



# Kolmogorov-Smirnov (K-S) Test

- Kolmogorov computed the expected distribution of the distance of the two CDFs when the null hypothesis is true.

# Example: Kolmogorov-Smirnov Test

Lag	Decision		p-value		Distance	
	True Null	Null Null	True Null	Null Null	True Null	Null Null
1	1	0	0	0.5427	0.79	0.0076
2	1	0	0	0.2126	0.78	0.0100
3	1	0	0	0.98485	0.75	0.0043
4	1	0	0	0.9937	0.72	0.0040
5	1	0	0	0.9769	0.68	0.00453

Distance of the two  
distributions  
In Sup norm

For **all neuron pairs** (A, B), populate the following distributions with

Population 1: real STTC of the pair (A,B)

Population 2: random circular shift in one of the two spike trains of (A,B)

Population 3: random circular shift in one of the two spike trains of (A,B)

**True Null: Population 1 vs. Population 2**

**Null Null: Population 2 vs. Population 3**



# Example: Null distribution test for directional STTC

## For a given pair (A,B)

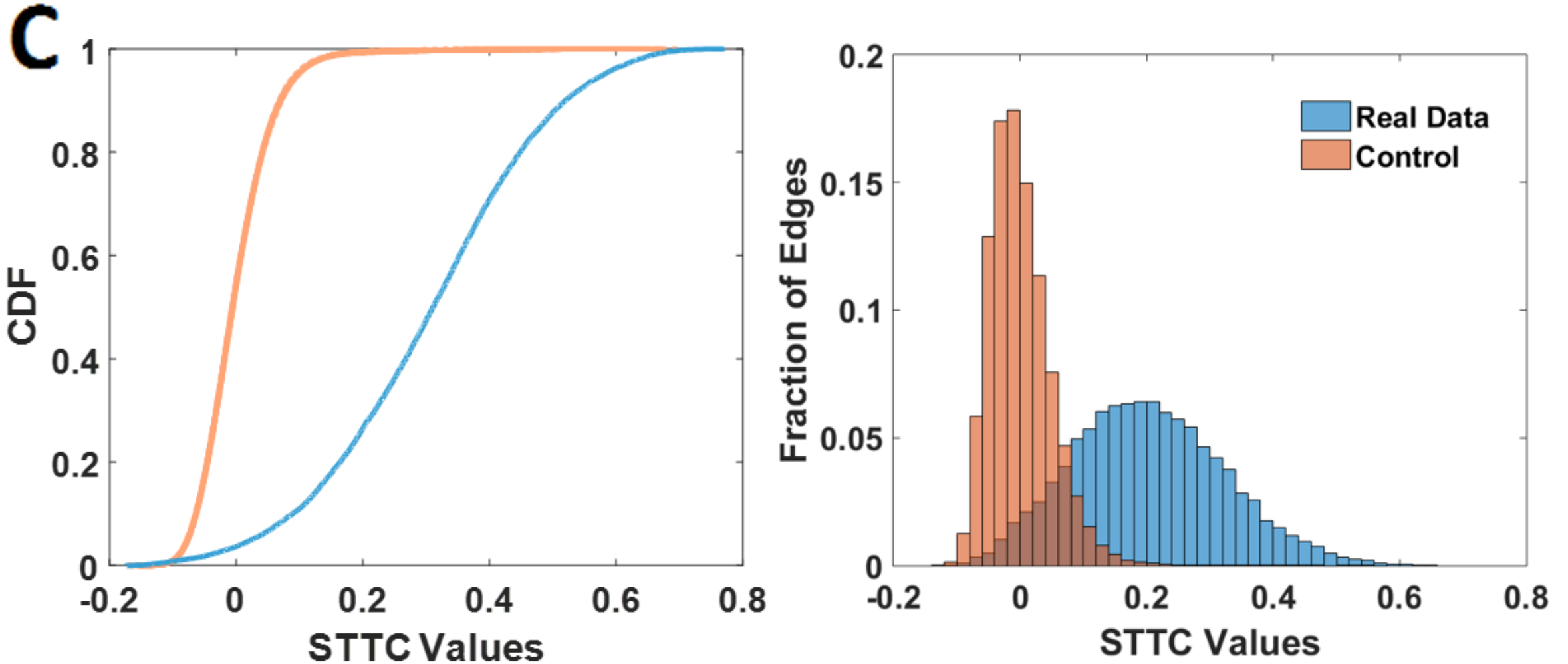
1. Circular shift the spike train of the neuron A,  $A^1$
2. Estimate the directional STTC( $A^1, B$ )
3. Repeat the above steps 100 times
4. Estimate the mean & standard deviation of the obtained STTC values
5. The statistical significant threshold ( $thr$ ) = mean + 3 std dev

If the directional STTC (A, B) >  $thr$  , the edge (A,B) is statistically significant.

The criterion can be strengthened with more repetitions (e.g., **1000**), a larger number of std dev (e.g., **5**).

# Strengthen the Criterion of Significant Edge

- Reject the null hypothesis test
- The total number of spikes of A within a STTC lag of spikes of B is above 3.
- The total number of spikes of B within a STTC lag of spikes of A is above 3.



1. Control group: each neuron trace is circular shifted by random delay
2. For each pair of 'shifted' neurons estimate the directional STTC & null distribution
3. Identify the significant edges of the control group

The **real neuron traces** appear **higher** values of directional STTC & percentage of significant edges

# Kolmogorov-Smirnov (K-S) Test

- **Non-parametric** test of the equality of **continuous 1D** probability distributions
- Quantifies a **distance between two distribution** functions
- Can serve as a **goodness of fit test**

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

where  $F_{1,n}$  and  $F_{2,m}$  are the empirical distribution functions

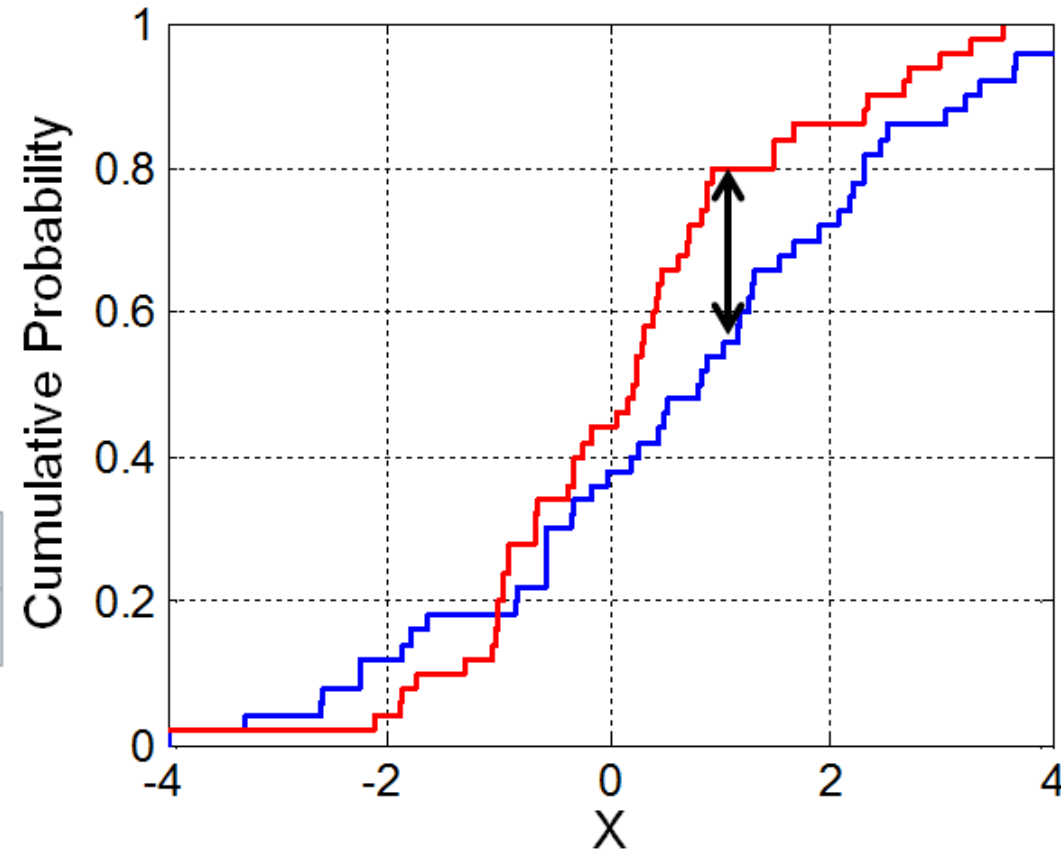
The null hypothesis is rejected at level  $\alpha$  if

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{n \cdot m}} \quad | \quad \mathbf{n \& m: size of the sample datasets}$$

$\alpha$	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

and in general by

$$c(\alpha) = \sqrt{-\frac{1}{2} \ln\left(\frac{\alpha}{2}\right)}.$$



# Kolmogorov-Smirnov (K-S) Test

- Kolmogorov computed the expected distribution of the distance of the two CDFs when the null hypothesis is true.

# Example: Kolmogorov-Smirnov Test

Lag	Decision		p-value		Distance	
	True Null	Null Null	True Null	Null Null	True Null	Null Null
1	1	0	0	0.5427	0.79	0.0076
2	1	0	0	0.2126	0.78	0.0100
3	1	0	0	0.98485	0.75	0.0043
4	1	0	0	0.9937	0.72	0.0040
5	1	0	0	0.9769	0.68	0.00453

Distance of the two  
distributions  
In Sup norm

For **all neuron pairs** (A, B), populate the following distributions with

Population 1: real STTC of the pair (A,B)

Population 2: random circular shift in one of the two spike trains of (A,B)

Population 3: random circular shift in one of the two spike trains of (A,B)

**True Null: Population 1 vs. Population 2**

**Null Null: Population 2 vs. Population 3**