



## Applied Statistics 2019-2020 Permutation Tests

Simone Vantini  
MOX, Dept. of Mathematics, Politecnico di Milano  
Leonardo Campus, Building 14, Floor VI,  
02 2399 4584  
simone.vantini@polimi.it

### Gaussian Assumption and Parametric Tests for the Mean(s)

- $p = \# \text{ random variables (features)}$
- $1 = p < n = \infty$  Thanks to the Central Limit Theorem, Gaussianity is not a key point.
  - $1 = p < n \leq \infty$  The t-distribution is meant to model situations in which the sample size is not very large. So the Gaussianity of data is required for the t-test. Univariate Gaussianity is, anyhow, not difficult to assess (normality tests).
  - $1 \leq p < n \leq \infty$  Hotelling's  $T^2$  test rely on multivariate Gaussianity of data. If  $p$  increases, multivariate Gaussianity can be difficult to assess (curse of dimensionality).
  - $1 \leq n < p \leq \infty$  High-dimensional tests rely on the multivariate Gaussianity of data, and they are not robust with respect to the violation of Gaussianity.  
Powerful Gaussianity tests are not available in the high-dimensional setting.
  - $1 \leq n < p = \infty$  In the functional case, normality is basically an unverifiable assumption.

simone.vantini@polimi.it

POLITECNICO DI MILANO

3

### Gaussian Assumption and Parametric Tests for the Mean(s)

All parametric tests (for the means) are exact  
either **asymptotically** or  
under the **Gaussianity assumption**

and

**not exact otherwise**

simone.vantini@polimi.it

POLITECNICO DI MILANO

4

### At the beginning of permutation tests

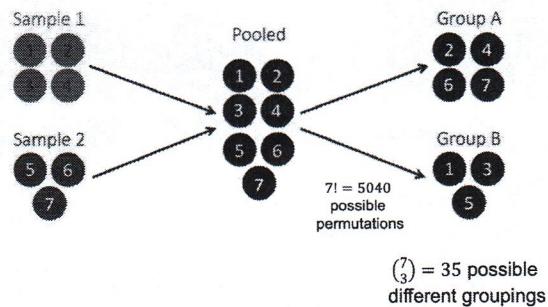
Let us suppose, for example, that we have measurements of the stature of a hundred Englishmen and a hundred Frenchmen. It may be that the first group are, on the average, an inch taller than the second, although the two sets of heights will overlap widely. [...] The simplest way of understanding quite rigorously, yet without mathematics, what the calculations of the test of significance amount to, is to consider what would happen if our two hundred actual measurements were written on cards, shuffled without regard to nationality, and divided at random into two new groups of a hundred each. This division could be done in an enormous number of ways, but though the number is enormous it is a finite and a calculable number. We may suppose that for each of these ways the difference between the two average statures is calculated. Sometimes it will be less than an inch, sometimes greater. If it is very seldom greater than an inch, in only one hundredth, for example, of the ways in which the sub-division can possibly be made, the statistician will have been right in saying that the samples differed significantly. For if, in fact, the two populations were homogeneous, there would be nothing to distinguish the particular subdivision in which the Frenchmen are separated from the Englishmen from among the aggregate of the other possible separations which might have been made. Actually, the statistician does not carry out this very simple and very tedious process, but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method.

Fisher, R. A. (1936). The coefficient of racial likeness and the future of craniometry, *Journal of the Anthropological Institute of Great Britain and Ireland*, pp. 57-63.

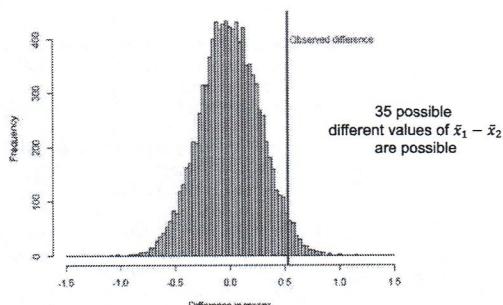
simone.vantini@polimi.it

POLITECNICO DI MILANO

H0:  $m_1 = m_2$  vs H1:  $m_1 > m_2$



### Permutational distribution of $T = \bar{x}_1 - \bar{x}_2$ under the H0



#### Conditional distribution of $T = \bar{x}_1 - \bar{x}_2$ given the pooled sample:

- Under the H0 all the 35 values of  $\bar{x}_1 - \bar{x}_2$  are equally probable
- Under the H1 larger values of  $\bar{x}_1 - \bar{x}_2$  are more probable

$$p = \frac{\sum_{k=1}^{35} I(T^* \geq T_0)}{35}$$

#### Some corner-stones of permutational inference:

- Their aim is making fewer assumptions as possible on data distribution
- How they work:
  - Likelihood-invariant transformations under the H0 (Conditional inference within induced equivalence classes)
  - Selection of the test statistic:
    - no a-priori optimal test statistic its distribution has to be stochastically larger under the «targeted» H1 than under the H0
    - possibility of working in purely metric spaces (i.e., complex data)
- Inferential properties:
  - Finite-sample exactness (differently from bootstrap)
  - Consistency (if the test statistic is properly chosen)
  - Asymptotic equivalence to parametric tests (when the same test statistic is used and the parametric assumptions hold)
- Large computational costs (Conditional Montecarlo)

#### Two-population test and 1-way ANOVA:

→ Value permutations (equivalent to group labels permutation)

#### One-population test and paired two-population test:

→ Recentering in H0 and sign swaps (assuming symmetry)

#### Independence test:

→ Pair Recoupling

#### “F-test” for linear models (linear regression and multi-way ANOVA)

→ Response permutations

#### “T-test” for linear models (linear regression and multi-way ANOVA)

→ Permutations of residuals of restricted model [asymptotic]

→ Permutations of residuals of complete model [asymptotic]

## INTUITION ON PERMUTATION TESTS

Suppose we want to compare the heights of english men vs. french men. We work on two samples (both of 100 units) of english and french men (100 english, 100 french).

We do the mean of each group and we make a comparison. We see that on average english men are 1 inch taller than french men. Is this sufficient to say that the distribution of heights is different?

Saying that there is a difference of 1 inch between two samples is not enough to say that there is difference between two populations.

- Parametric settings: focus on the distributions of the estimate (t-test, ...)
- Nonparametric settings: Fisher's idea:

we take the 200 heights and we write the 200 heights on 200 pieces of paper. Then we randomly assign a piece of paper (i.e. a height) to a guy. Under  $H_0$ , if the two distributions are identical, every sequence of heights ( $200!$  possible sequences) would have the same chance to be observed (even the same of the original sequence).

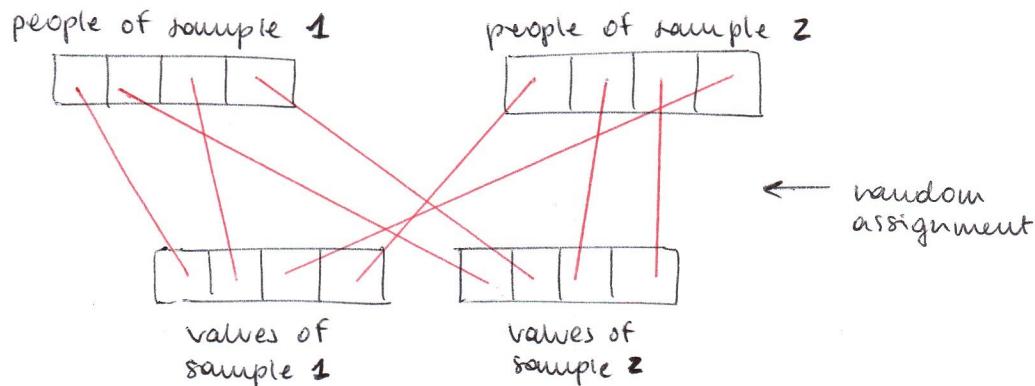
Why? If the two distributions are the same we're drawing 200 realizations from the same distribution (in this case, in fact, the distribution of heights of english and french men).

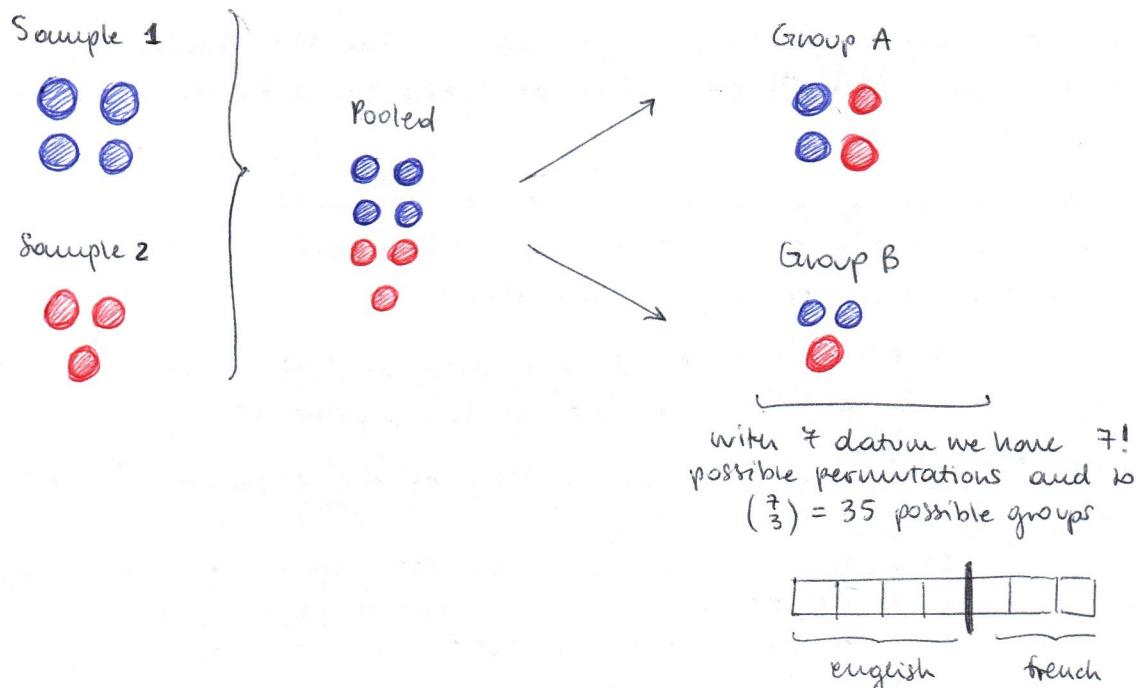
Idea: our sequence has the same probability of being realized than the other sequences under  $H_0$ . Under  $H_1$  (e.g. english men are taller than french ones) some sequences would be more likely to be observed. (sequences with larger values for english men and smaller values for french men would be more likely to be observed than the opposite)

For the p-value we can simply count how many sequences have provided us a difference between the two population larger than the observed one. ( $\rightarrow$  larger than 1 inch)

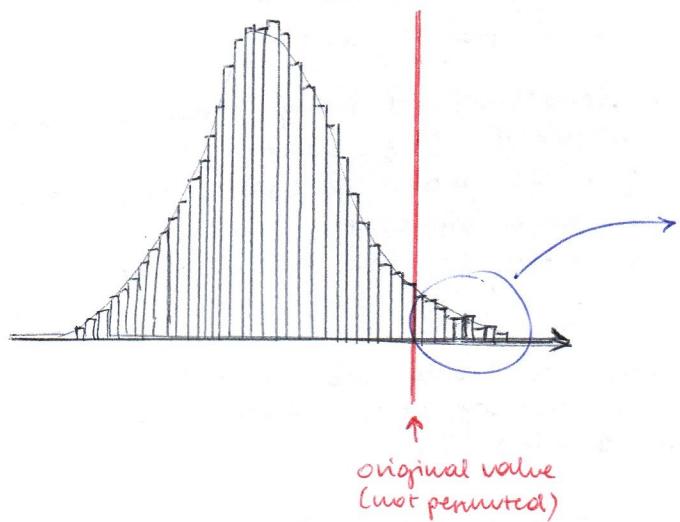
Why is it a valid p-value?

Under  $H_0$  the p-value is uniformly distributed over  $[0,1]$  (in particular we have the exactly same chance to obtain the most extreme dataset and so on... Over all  $200!$  possible datasets obtained by permuting the data) while under  $H_1$  it is more probable that our dataset is one of the most extreme ones.





We do a lot of these permutations and we construct the permutational distribution of  $T = \bar{X}_1 - \bar{X}_2$  under  $H_0$ :



The idea is to compute the distribution (reported as histogram)

The p-value is simply computed as the number of permutations which led us to a test statistics' value larger than the original test stat. value (with no permutations)

**Note:** differently from Sign-Rank tests, the distributions of test statistics are sample-dependent (the distributions do not depend only on the sample size).  
Most famous example: test for comparing two II populations for which 2 II samples has been drawn.

We're testing the equality in distribution of the two random variables. Notice that we have  $|\bar{X}_1 - \bar{X}_2|$  but we can also have  $(\bar{X}_1 - \bar{X}_2)^2$ ,  $\bar{X}_1 - \bar{X}_2$ , variance of the difference or others.  
All these tests are inferentially equivalent.  
(property of permut. tests : if we apply a monotonic transformation to the test statistics, we obtain a new test which is inferentially equivalent to the original test.)

Because of this we use the simplest:  $|\bar{X}_1 - \bar{X}_2|$ .

pooled sample

Under  $H_0$  the permuted vector is still a random sample from the same distribution (if permutation), but only under  $H_0$ !

Under  $H_1$  we may have a new vector (the permuted one) that have a different distribution from the non-permuted one.

Under  $H_0$  permutations preserve distributions, under  $H_1$  permutations may not preserve the distribution of the vector.

**Note:**  
some permutations preserve the distribution also under  $H_1$ :  
pooled:  
 $\Pi(\text{pooled})$ :  
group 1 group 2  
(if we permute intra-grouped then nothing change)

## PERMUTATION TESTS (FISHER 1936)

- FISHER EXAMPLE (SLIDES)

### TWO-SAMPLE (INDEPENDENT) T-TEST

$X_{11}, X_{12}, \dots, X_{1m_1} \sim \text{iid } X_1$   
 $X_{21}, X_{22}, \dots, X_{2m_2} \sim \text{iid } X_2$   $\rightarrow$  INDEPENDENT

$$H_0: X_1 \stackrel{d}{=} X_2 \quad H_1: X_1 \neq X_2$$

$$T_0 = |\bar{X}_1 - \bar{X}_2|$$

IDEA

IF  $H_0$  IS TRUE ( $X_1 \stackrel{d}{=} X_2$ ) WE HAVE THAT:

$$X_{11}, \dots, X_{1m_1}, X_{21}, \dots, X_{2m_2} \stackrel{H_0}{\sim} \text{iid } X_1 \stackrel{d}{=} X_2$$

PERMUTATION OF COMPONENTS

so

$$\text{IF } X_{\text{pooled}}^* = (X_{11}^*, \dots, X_{1m_1}^*, X_{21}^*, \dots, X_{2m_2}^*) = \Pi(X_{\text{pooled}})$$

WE HAVE THAT

$$\Pi(X_{\text{pooled}}) \stackrel{H_0}{\sim} \text{iid } X_1 \stackrel{d}{=} X_2 \quad \forall \Pi \in \Pi$$

WHILE

$$\Pi(X_{\text{pooled}}) \stackrel{H_1}{\not\sim} \text{iid}$$

GROUP OF PERMUTATIONS

$\Rightarrow$  UNDER  $H_0$  ANY  $\Pi(X_{\text{pooled}})$  HAS EXACTLY THE SAME "PROBABILITY" OF BEING OBSERVED (distribution invariant to permutations)

$\Rightarrow$  UNDER  $H_1$  (eg.  $X_2 \stackrel{d}{=} X_1 + c$  WITH  $c \neq 0$ ) SOME  $\Pi(X_{\text{pooled}})$  ARE MORE "PROBABLE" THAN OTHERS (eg.  $\Pi(X_{\text{pooled}})$  ASSOCIATED TO LARGER VALUES OF  $|\bar{X}_1 - \bar{X}_2|$  ARE MORE "PROBABLE.")

two II populations, two II samples, we want to compare the distributions of the populations

under  $H_1$  this is not a proper sample rule  $n_1$  elements come from a distribution and  $n_2$  from another.

Under  $H_0$ , since  $X_{11}, \dots, X_{1m_1}, X_{21}, \dots, X_{2m_2}$  are iid, we have that if we apply a permutation of the components we obtain a new vector (of length  $(n_1+n_2)$ ) having exactly the same distribution than the original one

$X_{\text{pooled}}^* = \text{permuted sample of } X_{\text{pooled}}$   
 $= \Pi(X_{\text{pooled}})$

where  $\Pi$  is the perm. of the components

! DEFINE THE t-VALUE AS THE PROPORTION OF PERMUTED POOLED SAMPLES  $\Pi(X_p)$  THAT ARE MORE EXTREME (IN TERMS OF  $|\bar{X}_1 - \bar{X}_2|$ ) THAN THE ORIGINAL POOLED SAMPLE  $X_{\text{pooled}}$ .

(\*) If we look at the distribution of the test statistics over all the possible  $M$  permutations we obtain the "permutational distribution of the test statistic" (# distribution of the test statistic)

( discrete distribution over all the possible values of the test statistics (values associated to all possible permutations) and this distribution is uniform over the values under  $H_0$  and not uniform under  $H_1$  (under  $H_1$  the probability is still concentrated around the same values but the probability distribution is different (the prob. is concentrated over the values which were associated to larger values of the test statistics)).

ALGORITHM to compute the p-value

$M = (m_1 + m_2)! = \#(\Pi)$ . ~ NUMBER OF POSSIBLE PERMUTATIONS.  
(CARDINALITY OF THE PERMUTATION SPACE)

- DEFINE  $T_0 = T(x_{\text{pooled}})$  → original test statistics
- $\forall \pi \in \Pi$  COMPUTE  $T^* = T(\pi(x_{\text{pooled}}))$  → value of the test statistics for  $\pi$  possible permutation of the pooled sample (\*)

$$\begin{aligned} \text{P-VALUE} &= \frac{1}{M} \sum_{\pi \in \Pi} \mathbb{I}_{\{T(\pi(x_p)) \geq T_0\}} = \\ &= \frac{\#\{T^* \geq T_0\}}{M} \end{aligned}$$

$T_\alpha$  is the value for which  $\alpha$ % of the permutations lead to a value of the test statistics larger than  $T_\alpha$  and  $(1-\alpha)$ % of permutations give us the test statistics smaller than  $T_\alpha$

→ CRITICAL REGION OF SIGNIFICANCE LEVEL  $\alpha$

$$T_0 : T_0 > T_\alpha \text{ WITH } T_\alpha : \frac{1}{M} \sum_{\pi \in \Pi} \mathbb{I}_{\{T(\pi(x_p)) \geq T_\alpha\}} = \alpha$$

superior quantile  
of the permutational  
distribution

$$: \frac{\#\{T^* \geq T_\alpha\}}{M} = \alpha$$

(NB) DIFFERENTLY FROM PARAMETRIC AND SIGN/RANK TESTS  
 $T_\alpha$  DEPENDS ON  $x_{\text{pooled}}$  AND COULD POSSIBLY CHANGE FROM SAMPLE TO SAMPLE.

different samples generate different permutational distr.  
(always discrete uniform but changes the support (= the possible values that this random variable can assume))

(NB)  $M$  COULD BE TERRIBLY LARGE!

# permutations (A MILLION PERMUTATIONS ARE REACHED FOR  $m_1 + m_2 = 10$ )

=> P-VALUE (OR  $T_\alpha$ ) ARE TYPICALLY ESTIMATED VIA MC

Sample specific  
Monte Carlo simulations

CONDITIONAL MC (CMC)

$\Pi$  IS REPLACED BY A RANDOM SUB-SAMPLE OF PERMUTATIONS  $\tilde{\Pi}$

$\tilde{\Pi} \subset \Pi$  WITH  $\#\{\tilde{\Pi}\} = B \ll \#\{\Pi\} = M$

$$\tilde{P} = \frac{1}{B} \sum_{\pi \in \tilde{\Pi}} \mathbb{I}_{\{T(\pi(x_p)) \geq T_0\}}$$

CMC p-value : We randomly pick  $B$  permutations and see for each one if the value of the test statistic is larger/smaller than the value of the test statistic of the original sample.

IDEA: we do not explore all the  $M$  permutations of the data, we use a random selection of permutations:

$\tilde{\Pi}$  := random selection of permutations  
 $\text{card}(\tilde{\Pi}) = B$

$\Rightarrow \tilde{P} \sim B(B, P)$  IN PARTICULAR

$$\tilde{P} \xrightarrow{a.s.} P \quad B \rightarrow \infty$$

R-EXAMPLE 1 - UNIVARIATE - R

What is the prob. distribution of  $\mathbb{I}_{\{T(\pi(x_p)) \geq T_0\}}$ ?

If we take a random permutation, what is the probability that that perm. will give us a test statistic larger than  $T_0$ ? It's equal to the true p-value.

$$\mathbb{I}_{\{... \}} \sim \text{Be}(p\text{-value})$$

p-value := IP over all the permutations that the test stat. is larger than  $T_0$

with this distribution we can build  $CI(p)$  to be sure that the approx. error is smaller than a given threshold (or we can even use it to select  $B$ )

Under  $H_0$  the probability of rejecting  $H_0$  is  $\leq \alpha$ ,  
under  $H_1$  the probability of rejecting  $H_0$  is  $> \alpha$ .

An exact test is a test which controls the probability of type I error and it's a test that at the same time, when  $H_0$  is violated, increases the probability of rejecting  $H_0$  w.r.t. the probability of accepting  $H_0$ .

### IS THE PERMUTATIONAL T-TEST AN EXACT TEST?

$$P(P \leq \alpha) \stackrel{H_0}{\leq} \alpha \text{ AND } P(P \leq \alpha) \stackrel{H_1}{>} \alpha ?$$

THE ANSWER IS "YES"! IN PARTICULAR FOR SOME SPECIFIC VALUES OF  $\alpha$  (THE REACHABLE LEVELS) WE ALSO HAVE:

$$P(P \leq \alpha) \stackrel{H_0}{=} \alpha \text{ AND } P(P \leq \alpha) \stackrel{H_1}{>} \alpha$$

**REACHABLE LEVELS** are multiples of the minimal value p-value (see the last example). The fact that only some p-values are reachable is based on the fact that the distribution of the p-values and the distr. of the test statistic is always discrete (no matter the number of data).

histogram of the T-stat is always discrete, such as the ecdf. Larger is  $n$  and more reachable levels we get.

THE EXACTNESS HOLDS FOR ANY TEST STATISTIC  $T_0 = T(X_p)$  THAT IS STOCHASTICALLY GREATER UNDER  $H_1$  THAN  $H_0$ :

$$P_{H_1}(T(X_p) > t) > P_{H_0}(T(X_p) > t) \quad \forall t \in \mathbb{R}$$

(N.B.) IF THIS ASS. DOES NOT HOLD WE MIGHT LOSE THE PROPERTY  $P(P \leq \alpha) \stackrel{H_1}{\leq} \alpha$  (LOSS OF POWER)  
PROPERTY  $P(P \leq \alpha) \stackrel{H_0}{\leq} \alpha$  IS INSTEAD ALWAYS PRESERVED.

We always have the control of the probability of type I error

(N.B.) THE VALIDITY OF THIS ASS. DEPENDS ON:

- THE CHOICE OF  $T$
- THE DISTRIBUTION OF  $X_1$  AND  $X_2$  UNDER  $H_1$ .

$$\begin{aligned} &\text{ex } X_2 = X_1 + c \quad c \neq 0 \\ &T = |\bar{X}_1 - \bar{X}_2| \end{aligned} \quad \rightarrow \text{ASS. HOLDS.}$$

$$\begin{aligned} &\text{ex } X_1 > 0 \text{ AND } X_2 > 0 \\ &X_2 = cX_1 \quad c > 0 \quad c \neq 1 \\ &T = |\bar{X}_1 - \bar{X}_2| \end{aligned} \quad \rightarrow \text{ASS. HOLDS}$$

### SHRINKAGE

Suppose that the worst that can happen, given two distributions, is that one distribution is a shrinkage of the other (i.e. compressed / expanded).  
Also in this case with:

$T = |\bar{X}_1 - \bar{X}_2|$  we have that the assumptions of the theorem are satisfied ( $\Rightarrow$  exact test).

### SHIFT

Suppose that we're in a context where the worst that can happen is the shift of the two populations (similar to the standard t-test: two gaussians that differ by the mean (and so one is the translation of the other)). In this case with:

$T = |\bar{X}_1 - \bar{X}_2|$  we have that the assumptions of the theorem are satisfied ( $\Rightarrow$  exact test).

→ What we want to focus on, to find out if the assumptions of the theorem are satisfied, is what can happen to the two distributions in the "worst" scenario, i.e. under  $H_1$ .

Because of how the assumptions of the theorem are, the test statistic are expected to move towards larger values when  $H_0$  is violated (to be likely to respect the thm. assumptions).

Settings:  $\mathcal{X}$ ,  $x_p$ ,  $p(x_p)$ ,  $\Pi$ ,  $M$ ,  $[\bar{x}_p]$



### FORMALIZATION OF THE PERMUTATIONAL T-TEST

- $\mathcal{X}$  = image space of the pooled samples  
SET OF ALL POSSIBLE REALIZATIONS OF THE POOLED SAMPLE  
(e.g. IF  $X_1 \sim N(\mu_1, \sigma^2)$  AND  $X_2 \sim N(\mu_2, \sigma^2)$   $\mathcal{X} = \mathbb{R}^{m_1+m_2}$ )
- $\bar{x}_p$  IS THE OBSERVED POOLED SAMPLE.
- $P(\bar{x}_p) = p_1(x_{11}) \dots p_1(x_{1m_1}) p_2(x_{21}) \dots p_2(x_{2m_2})$  IS  
THE DENSITY PROBABILITY OF THE POOLED SAMPLE  
(e.g.  $\frac{1}{(2\pi\sigma^2)^{\frac{m_1+m_2}{2}}} \exp\left(-\frac{1}{2\sigma^2}(x_{11}-\mu_1)^2 + \dots + (x_{2m_2}-\mu_2)^2\right)$ )
- $\Pi$  = GROUP OF PERMUTATIONS = set of all possible permutations  
 $M = \#\Pi$  = CARDINALITY OF  $\Pi$   
(e.g.  $(m_1+m_2)!$ )

\* if the data are independent

probability of the pooled sample observed

### EQUIVALENCE CLASS

all the possible elements of  $\mathcal{X}$  (all the possible samples) that can be obtained by permuting the original sample

- $[\bar{x}_p] = \{\bar{x}_p \in \mathcal{X} : \exists \pi \in \Pi \text{ s.t. } \pi(\bar{x}_p) = \bar{x}_p\}$   
IS THE EQUIVALENCE CLASS (INDUCED BY THE ACTION OF  $\Pi$  ON  $\mathcal{X}$ ) WHICH  $\bar{x}_p$  BELONGS TO.

$$\#[\bar{x}_p] = M$$

poofed sample

We have a space  $\mathcal{X}$  and a group acting on  $\mathcal{X}$ :  $\Pi$ .  
The action of the group on the space generates a set of equivalence relations.  
We can divide  $\mathcal{X}$  in equivalence classes induced by  $\Pi$ .

THE IDEA IS NOT TO MAKE INFERENCE ABOUT  $\bar{x}_p$  UNDER  $H_0$

BUT RATHER ON  $\bar{x}_p | \bar{x}_p \in [\bar{x}_p]$  (THE SO CALLED

PERMUTATIONAL DISTRIBUTION OF THE POOLED SAMPLE)

$$P(\bar{x}_p) \text{ and } P(\bar{x}_p | [\bar{x}_p]) = \frac{P(\bar{x}_p)}{P([\bar{x}_p])} =$$

$$= \frac{P(\bar{x}_p)}{P([\bar{x}_p])} \stackrel{H_0}{=} \begin{cases} 1/M & \bar{x}_p \in [\bar{x}_p] \\ 0 & \bar{x}_p \notin [\bar{x}_p] \end{cases}$$

$$P(\bar{x}_p) = P(\pi(\bar{x}_p)) \forall \pi \in \Pi$$

$$\stackrel{H_1}{=} \begin{cases} > 0 & \bar{x}_p \in [\bar{x}_p] \\ 0 & \bar{x}_p \notin [\bar{x}_p] \end{cases}$$

$$P(\bar{x}_p) \neq P(\pi(\bar{x}_p))$$

### [NOTATIONS]

To avoid confusion:

$\bar{x}_p$  = original sample

$x_p$  = generic sample

$[\bar{x}_p]$  is a weird random variable: its realizations are equivalence classes

We look at the distr. of the pooled sample within a specific equivalence class.  
We condition our distr. w.r.t. the fact that our pooled sample has to belong to the same equivalence class which the original datum belongs to.

Instead of looking at the distribution of the pooled sample (and also the pooled test statistic) in general, over any possible sample we can obtain on the data, we look (compare) only at the samples belonging to the same equivalence class of our original dataset.

Given the fact that we selected a group of permutations and these permutations provide us with samples that have all the same probability under  $H_0 \Rightarrow$  under  $H_0$ :  $f(x_p | [\bar{x}_p]) = \frac{1}{M} \mathbb{1}_{\{x_p \in [\bar{x}_p]\}}$

$H_0 \Rightarrow$  the conditional distribution of the pooled sample given any equivalence class will be always a discrete uniform distribution over  $M$  possible samples belonging to  $[\bar{x}_p]$

$H_1 \Rightarrow$  the distribution remain discrete, the support of the distribution (i.e. the  $M$  possible samples belonging to  $[\bar{x}_p]$ ) remains the same but the probability of observing the specific elements changes ( $\neq \mathbb{1}_{\{x_p \in [\bar{x}_p]\}}$ ). Under  $H_1$  not all the pooled samples will have the same probability of being observed.

→ INFERENCE WITHIN THE EQUIVALENCE CLASS WHICH THE ORIGINAL SAMPLE BELONGS TO.

## Last step: FROM THE POOLED SAMPLE TO THE TEST STATISTIC

When we run a statistical test we look at the sample through the values of the test statistic (that is able to capture relevant features). The critical region is based on a threshold given to the test statistic (we reject/not reject given the value of the test statistic). In parametric settings we look at the distribution of the test statistic under  $H_0$  and we compare our test statistic with all the possible values that could have been observed. When we run a permutation test we reject  $H_0$  comparing our value test statistic w.r.t. the distribution of the test statistic within the equivalence class which our original sample belongs to. We make inference within the equivalence class.

GIVEN  $T_0 = T(m_p)$

WE DO NOT REJECT  $H_0$  RELYING ON THE DISTRIBUTION OF  $T_0$   
BUT ON  $T_0 | X_p \in [m_p]$  UNDER  $H_0$  (PERMUTATIONAL DISTRIBUTION OF  $T_0$ )

instead we use this

$$\text{e.g. } T_0 = |\bar{x}_1 - \bar{x}_2| \stackrel{H_0}{\sim} N(0, \sigma^2(\frac{1}{m_1} + \frac{1}{m_2}))$$

$$T_0 | X_p \in [m_p] \stackrel{H_0}{\sim} U\{t_1, t_2, \dots, t_m\}$$

test statistic  
given the equivalence  
class

THE SET MADE OF  
THE  $m = \frac{(m_1+m_2)}{m_1}$  POSSIBLE

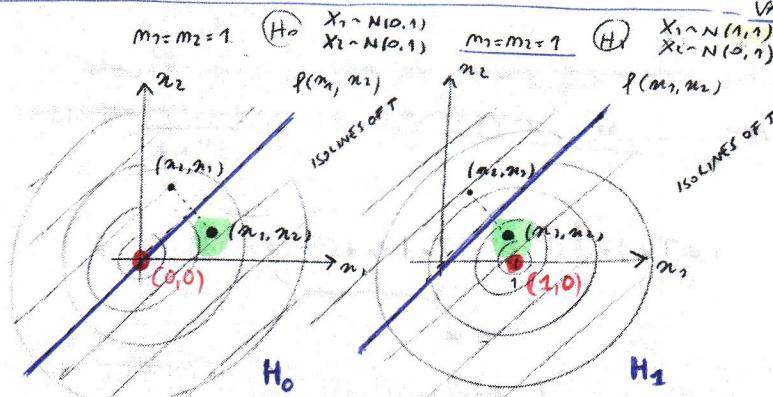
VALUE OF THE TEST STATISTIC  
OBTAINABLE BY PERMUTATIONS

NB THEY DEPEND ON  $m_p$ !

$$T_0 \stackrel{H_1}{\sim} N((\mu_1 - \mu_2), \sigma^2(\frac{1}{m_1} + \frac{1}{m_2}))$$

$T_0 | X_p \in [m_p] \stackrel{H_1}{\sim}$  DISCRETE  $\{t_1, t_2, \dots, t_m\}$  WITH PROBABILITY  
P-values and critical regions  
must be done on the right-tail MASS CONCENTRATED  
TOWARD LARGER VALUES.

## EXAMPLE



obviously we now know these, but since it's non-parametric settings we don't know them officially! That's why we explore distributions numerically, using MC approach

$$f(m_p) = \frac{1}{2\pi} \exp\left(-\frac{m_1^2 + m_2^2}{2}\right)$$

$$f(\pi(m_p)) = \frac{1}{2\pi} \exp\left(-\frac{m_1^2 + m_2^2}{2}\right)$$

$$f(m_p) = \frac{1}{2\pi} \exp\left(-\frac{(x_1-1)^2 + m_2^2}{2}\right)$$

$$f(\pi(m_p)) = \frac{1}{2\pi} \exp\left(-\frac{(x_2-1)^2 + m_2^2}{2}\right)$$

We only have one sample from  $X_1$  and one sample from  $X_2$ . We obtain  $(x_1, x_2)$ . Under  $H_0$  the isolines are centered in  $(0,0)$ , under  $H_1$  they're centered in  $(1,0)$ . Suppose that we want to make permutational inference, and we suppose to observe the point  $(x_1, x_2)$ . What is the equivalence class to which  $(x_1, x_2)$  belongs to? Here the only possible permutation is  $(x_1, x_2) \rightarrow (x_2, x_1)$ . So the equivalence class is made of these two points:  $\{(x_1, x_2), (x_2, x_1)\}$ . For every realization, the equivalence class is made of the original datum and the symmetric data point w.r.t. the bisector of the first quadrant. We look at the probability density within the equivalence class. Under  $H_0$ :  $P((x_1, x_2)) = P((x_2, x_1)) = \frac{1}{2}$ . Under  $H_1$  we expect  $P((x_1, x_2)) > P((x_2, x_1))$  (from graphics). The permutational distribution under  $H_1$  will still be a discrete over  $\{(x_1, x_2), (x_2, x_1)\}$  but with a probability that is longer on  $(x_1, x_2)$  than on  $(x_2, x_1)$ .

$m$  and not  $M$  because it:



(we permute within the two groups) then the test stat.s will be =.

$m$  is the number of different test statistics.

Notice:  $m$  depends on the sample, so  $t_1, \dots, t_m$  depends on the sample as well. (the whole set  $\{t_1, \dots, t_m\}$  depends on the sample).



$$\begin{array}{lll} X_{11}, X_{12} & \stackrel{iid}{\sim} & \text{Be}(p_1) \\ X_{21} & \sim & \text{Be}(p_2) \end{array} \quad \left. \begin{array}{l} n_1 = 2 \\ n_2 = 1 \end{array} \right\} \text{II}$$

Permutational settings. Test statistic :  $T = |\bar{x}_1 - \bar{x}_2|$ .

Suppose we observe:  $(0, 1), (\bar{1})$ .

How many permutations can we do?  $n=3 \Rightarrow n! = 6$

Permutations :

$$\left. \begin{array}{ll} \begin{array}{c} (0, 1) (\bar{1}) \\ (1, 0) (\bar{1}) \\ (0, \bar{1}) (1) \\ (\bar{1}, 0) (1) \end{array} & T(\pi(x_p)) = 0.5 \\ \begin{array}{c} (\bar{1}, 1) (0) \\ (1, \bar{1}) (0) \end{array} & T(\pi(x_p)) = 1 \end{array} \right\} \begin{array}{l} \text{under } H_0: \\ P(\text{test stat} = 0.5) = \frac{2}{3} \\ P(\text{test stat} = 1) = \frac{1}{3} \end{array}$$

Under  $H_0$  we have that the probability of selecting one pooled sample is always the same but also the probability of selecting a given value of the test statistic is the same. With continuous data this is always correct. In the discrete case it may not be correct : it may happen that some tests collapse on the same value. However, practically this is not a problem since we evaluate the p-value by uniformly random permuting the sample, not the test statistic.

## HOW IS THE P-VALUE COMPUTED?

We maintain all the same definitions that we have in parametric settings but we add everywhere (to every probability) the conditionnement on the equivalence class.

$$\text{P-VALUE} = \mathbb{P}_{H_0}(T(X_p) \geq T_0 \mid X_p \in [n_p]) = \frac{\#\{T(\pi(n_p)) \geq T_0\}}{m}$$

P-VALUE  $\stackrel{H_0}{\sim} U\{1/m, 2/m, \dots, 1\}$

P-VALUE  $\stackrel{H_1}{\sim} \text{DISCRETE}\{1/m, 2/m, \dots, 1\}$  WITH MASS PROBABILITY CONCENTRATED TOWARDS SMALLER VALUES

## CRITICAL REGION

$$T_0: T_0 \geq T_\alpha \quad \text{WITH } T_\alpha: \frac{\#\{T(\pi(n_p)) \geq T_\alpha\}}{m} = \alpha$$

Computed in the standard way, but we condition to the permutational distribution

$$\left\{ \begin{array}{l} \mathbb{P}_{H_0}(T_0 \geq T_\alpha \mid X_p \in [n_p]) = \alpha \quad \text{WITH } d = 1/m, 2/m, \dots \\ \mathbb{P}_{H_1}(T_0 \geq T_\alpha \mid X_p \in [n_p]) > \alpha \end{array} \right.$$

\*  $T_\alpha$  is the beginning of the critical region of level  $1-\alpha$ .  $T_0$  is the value that divides all the possible test stat. values in two portions: larger one being on  $1-\alpha\%$ , smaller one being on  $\alpha\%$ .

Validity of permutational tests

PERMUTATIONAL INFERENCE SEEMS TO BE VALID ONLY CONDITIONALLY  
ACTUALLY PERMUTATIONAL INFERENCE IS UNCONDITIONALLY VALID

$$\mathbb{P}_{H_0}(\text{rejection criterion}) = \mathbb{P}_{H_0}(T_0 \geq T_\alpha) = \int_X \mathbb{P}_{H_0}(T_0 \geq T_\alpha \mid X_p \in [n_p]) p(n_p) d n_p = \alpha \quad (\text{by construction})$$

$$= \alpha \int_X p(n_p) d n_p = \alpha \quad \text{FOR ANY DISTRIBUTION OF } X_1, X_2$$

Both  $T_0$  and  $T_\alpha$  are random quantities that depends on the data ( $\neq$  parametric settings where  $T_\alpha$  is not a random quantity)

permutational inference can control the probability of type I error also in a scenario where the pooled sample is sampled not from the equivalence class but from the general population

$$\mathbb{P}_{H_0}(T_0 \geq T_\alpha \mid X_p \in [n_p]) = \alpha \Rightarrow \mathbb{P}_{H_0}(T_0 \geq T_\alpha) = \alpha$$

$$\mathbb{P}_{H_1}(T_0 \geq T_\alpha \mid X_p \in [n_p]) > \alpha \Rightarrow \mathbb{P}_{H_1}(T_0 \geq T_\alpha) > \alpha$$

probability over all the possible samples of reject  $H_0$  using a permutational test.  $\mathbb{P}_{H_0}(T_0 \geq T_\alpha) = \text{level of the test in practice.}$

This probability can be computed as an average probability of rejection over all the equivalence classes (over all the possible samples).

Whatever the distr. of the data, also the unconditional prob. of rejecting  $H_0$  under  $H_0$  is  $\alpha$  (and not only the permutational probability).

This prove that  $\mathbb{P}_{H_0}(\text{reject } H_0)$  will be  $\alpha$  even in the case in which the sample will not be drawn from the equivalence class but from the original population.

And so we conclude

If we use the absolute value or the absolute value<sup>2</sup> or  $\log(\cdot)$ ... we obtain the same p-value. Why? Because its evaluation is based on the ranking of the data, which is preserved by monotonic transformations

(NB) MONOTONIC TRANSFORMATIONS OF THE TEST STATISTIC LEAD TO INFERENTIALLY EQUIVALENT TESTS.

(THIS IS NOT TRUE FOR MONOTONIC TRANSF. OF DATA!)

(NB) PERMUTATION TESTS ARE GENERALLY MORE POWERFUL THAN SIGN AND RANK TESTS.

(NB) LESS ROBUST

THEY CAN BE MADE MORE ROBUST BY USING ROBUST TEST STATISTICS, e.g.: median is more robust than mean

$$|\bar{X}_1 - \bar{X}_2| \text{ and } |\text{MED}(X_1) - \text{MED}(X_2)|$$

(NB) PERMUTATION TESTS AND PARAMETRIC TESTS ARE ASYMPTOTICALLY INFERENTIALLY EQUIVALENT WHEN

- \* • THE SAME TEST STATISTIC IS USED
- \* • ASSUMPTIONS OF PARAMETRIC TESTS HOLD.

(PARAMETRIC TESTS CAN BE SEEN AS COMPUTATIONALLY EFFICIENT APPROXIMATIONS OF PERMUTATION TESTS WHEN SAMPLE SIZE IS LARGE AND SOME ASSUMPTIONS ARE KNOWN TO BE VALID)

PROOF IS BASED ON THE LAW OF LARGE NUMBERS i.e.

PERMUTATIONAL DISTRIBUTION OF T CONVERGES TO THE UNCONDITIONAL DISTRIBUTION OF T

PERMUTING  $\approx$  RESAMPLING  $\approx$  SAMPLING

if  $n$  is large, permuting is "equivalent" to re-sampling

(NB) JOINT USE OF MANY TEST STATISTICS!

REMEMBER OF TAKING CARE OF MULTICLASS CORRECTION!

$$T_A, T_B, T_C \rightarrow P_A, P_B, P_C$$

WE HAVE SMARTER SOLUTIONS THAN BONFERRONI-LIKE PROCEDURES.

FOR EXAMPLE: := EX\*

- $T_{ABC} = T_A + T_B + T_C \rightarrow P_{ABC}$
- $T_{ABC} = \max(T_A, T_B, T_C) \rightarrow P_{ABC}$
- $T_{ABC} = \min(R(T_A), R(T_B), R(T_C)) \rightarrow P_{ABC}$

VIA SYNCHRONIZED PERMUTATIONS

to sample these combined distributions via MC we cannot run 3 || MC simulations (this can destroy the statistical dependence between  $T_A, T_B, T_C$ )

## CHOICE OF THE TEST-STATISTIC

Given an inferential problem many test statistics can be chosen. The nature of the problem may help: if  $H_0$  and  $H_1$  differ on, for example, mean (one population is suspected (and so tested) to be shifted w.r.t. the other) it's more convenient to considerate a test statistic sensitive w.r.t. shifts more than, for instance, test statistics sensitive w.r.t. the variance of data.

Example:  $H_0: X_1 \stackrel{d}{=} X_2$

$$H_0: X_1 \stackrel{d}{=} X_2$$

$$H_1: \mu_1 \neq \mu_2 \rightarrow T := |\bar{X}_1 - \bar{X}_2|$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \rightarrow S := |S_1^2 - S_2^2|$$

specific violation of  $H_0$

convenient test statistics

S and T will react differently to the violation of  $H_0$ :

S will be sensitive w.r.t. changes in the variance,

T will be sensitive w.r.t. changes in the mean value.

That's why sometimes is convenient to run different test: we want to prove  $H_0$ ?

It should be valid both against  $\mu_1 \neq \mu_2$  and  $\sigma_1^2 \neq \sigma_2^2$ .

This approach (using a set of test statistics to test the same  $H_0$ ) is called

MULTIASPECT TESTS (= testing the same  $H_0$  looking at different aspects).

Using multi-aspects test leads us to think about corrections (e.g. Bonferroni's) since the tests must be valid simultaneously.

Simplest solution: run different tests and adjust the p-value with Bonferroni.

However, in permutational framework we have a smarter way:

DIRECT COMBINATION: we make a combination of the different test statistics and we look at the permutational distribution of the global test statistic

Consider: EX\*

We have  $T_A, T_B, T_C$  and we create  $T_{ABC} := T_A + T_B + T_C$ . For valid conclusions we look at the global test (the one with  $T_{ABC}$ ). Notice: usually  $T_A, T_B$  and  $T_C$  are standardized before being summed (just to be comparable). (or they've weighted, always for the sake of being comparable).

Another approach is:  $T_{ABC} = \max\{T_A, T_B, T_C\} \Rightarrow$  if at least one  $T_i$ 's suggest to move under  $H_1$  then we reject  $H_0$ .

Another approach is:  $T_{ABC} = \max\{R(T_A), R(T_B), R(T_C)\}$ , where  $R(T_A) = \text{rank of the test stat } T_A$  (over the permutational distribution).

Notice that this last procedure is an automatic rescaling: technically speaking it's like putting together the p-values of the 3 tests and taking the minimum one.

We proceed now with an overview of the most famous inferential problems:

1. One-way ANOVA
2. Simple regression
3. Center of symmetry
4. Independence test
5. Multivariate:
  1. Two II samples
  2. One sample (or paired)
  3. Multiple regression
    - 3.1. Global test
    - 3.2. Partial test
6. Confidence intervals

Idea: we have to build equivalent classes s.t. the conditional distribution of the transformed pooled sample within the equivalent class is a discrete uniform (under  $H_0$ ).  $H_1$  will be used as a suggestion of a good test statistic. Under  $H_0$  a pooled sample comes from the (same) generic distribution (the one all the  $G$  distrs are equivalent to), and so all the permuted pooled samples come from it. Under  $H_0$  we have the same prob. to get the original pooled sample or the transformed one. Under  $H_1$  this is not true.

$G$  populations ( $G \parallel$  samples)

$H_0$ :  $G$  distributions are =  
 $H_1$ :  $\exists$  a pair of distributions that are different

### ONE-WAY ANOVA

$$X_{11}, \dots, X_{Gm_j} \sim \text{iid } X_i \quad j=1, \dots, G \text{ INDEPENDENT}$$

average distance of each element from the global mean of the pooled sample

$$H_0: X_1 = X_2 = \dots = X_G \quad H_1: \exists j, i=1, \dots, G \text{ s.t. } X_j \neq X_i$$

$$T_0 = \sum_{j=1}^G m_j (\bar{X}_j - \bar{X})^2$$

$$T_0 = \sum_{j=1}^G m_j (\bar{X}_j - \bar{X})^2 / \sum_{j=1}^G m_j = \frac{B}{W}$$

$$T_0 = \sum_{j \neq k} \sum_{i=1}^{m_j} \sum_{i'=1}^{m_k} (X_{ji} - X_{j'i'})^2$$

cumulate on average the distance between any 2 objects belonging to 2 different groups (we look at all the possible pairs)  
⇒ DISTANCE BASED test statistic: they work on the raw data

sensitive to both shifts and changes in variance

$\Pi$  = GROUP OF PERMUTATIONS

$$\#\Pi = (m_1 + m_2 + \dots + m_g)!$$

statistic of the F test (parametric)  
up to some constant:  
= between variance / within variance

This test statistic is targeting the mean of the populations (sensitive w.r.t. shifts)

(H1B) WRT WHICH VIOLATION ARE EXPECTED TEST STATISTICS EXPECTED TO BE SENSITIVE TO?

(the idea is to compare the distance between the group's data clouds (Hierarchical: average linkage))

### SIMPLE REGRESSION

$$Y_i | X_i = x_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

(H1B) ASSUMPTION OF ADDITIVE ERRORS.

$\epsilon_1, \dots, \epsilon_m \sim \text{iid } \mathbb{E}(\epsilon)$  error term (unpredictable)

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$H_0$ : cond. on  $x_1, \dots, x_m$

$$Y_1, \dots, Y_m \sim \text{iid } Y = \beta_0 + \epsilon$$

$H_1$ : cond. on  $x_1, \dots, x_m$

$$Y_1, \dots, Y_m \not\sim \text{iid } Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

### FIRST CASE OF:

The test is non-parametric but here we have a parametric model. We assume:

1. the conditional mean of  $Y$  increases/decreases linearly in the value of  $X$
2. the error is an additive error and the errors are iid  $\epsilon$  (we don't assume that they're gaussian but we assume they're additive and with the same distr.)

⇒ this permutational test is not as generic as the others since we have these assumptions

$\Pi$  = GROUP OF PERMUTATIONS (OF RESPONSES!)

$$\#\Pi = m!$$

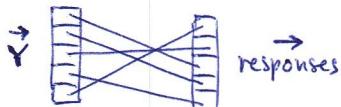
$$T_0 = \hat{\beta}_{\text{OLS}}^2 = \left( \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 \sim \chi^2 \sim R^2$$

$$T_0 = \left( \frac{\hat{\beta}_{\text{OLS}}}{S_{\hat{\beta}_{\text{OLS}}}} \right)^2 = \frac{\left( \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right)^2}{\frac{1}{m-2} \sum_{i=1}^n \hat{\epsilon}_i^2} \sim F$$

test for the significance of the regressor  $X$ .

It's equivalent to test if  $Y$  depends on  $X$ .

Y



under  $H_0$  the new combination  $\tilde{Y}$ -responses will have the same probability of being observed as the original. This is not true under  $H_1$ .

Notice: the group of transformations that we'll use to build equivalence classes are not permutations

### Assumptions:

the distribution is symmetric w.r.t. an unknown point. (If the mean  $\exists$ , the center of symmetry is also equal to the mean value (in general is always equal to the median value))

This is a test for:

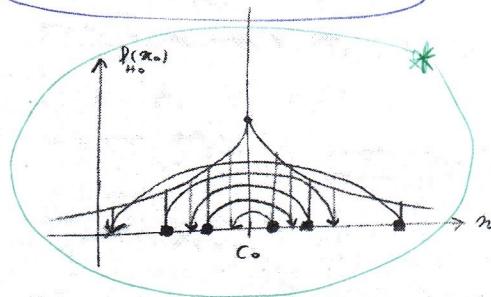
- median value
- mean value (if  $\exists$ )

one population test

### CENTER OF SYMMETRY OF A POPULATION

$$x_1, \dots, x_m \sim \text{ind } X \text{ s.t. } \exists c \quad f(c+h) = f(c-h) \quad \forall h \in \mathbb{R}$$

$$H_0: c = c_0 \quad H_1: c \neq c_0$$



$$X = (x_1, x_2, \dots, x_m)$$

$$\Pi(X) = (x_1 + H_1^*, x_2 + H_2^*, \dots, x_m + H_m^*)$$

$$\text{WITH } H_i^* = \begin{cases} x_i - c_0 \\ -(x_i - c_0) \end{cases} \text{ OR }$$

$\Pi(X) =$  every element can be taken as it is or flipped on the other side of the symmetry

UNDER  $H_0$  BY SWAPPING ONE OR MORE SIGNS OF THE DEVIATIONS  $H_i = x_i - c_0$ . WE OBTAIN A NEW DATASET WITH THE SAME PROB. DENSITY

Idea: under  $H_0$  we transform the original data and we obtain a new data which probability of being obtained is the same as the probability of obtaining the original data.

$\Pi =$  CARTESIAN GROUP OF REFLECTIONS WRT  $c_0$ .

$\# \Pi = 2^m$  ← we can decide if we want to change side for any datum

$$T_0 = |\bar{x} - c_0|$$

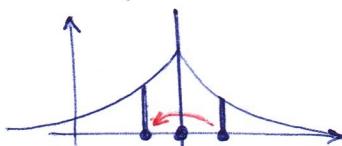
(NB) IF  $c_0 = 0$   $x_i + H_i^* = \begin{cases} x_i \\ -x_i \end{cases} \text{ OR }$  CHANGING SIGNS.

(NB) IF  $x_i$  IS REPLACED BY  $\text{sign}(x_i - c_0)$  WE OBTAIN THE SIGN TEST.

IF  $x_i$  IS REPLACED BY  $\text{sign}(x_i - c_0) R(|x_i - c_0|)$  WE OBTAIN THE SIGNED-RANK TEST.

This transformation is used to make symmetric a distribution that could be not symmetric. (since under  $H_0$  both  $\text{sign}(x_i - c_0)$  and  $R(|x_i - c_0|)$  are symmetric)

\* The likelihood of these data is just the product of the length of the lines:



If we move one point to the other side (of symmetry) the length is still the same ⇒ so the likelihood of data. We can move from one to all the points and under  $H_0$  the likelihood does not change

here transformations are not permutations but "reflections" of the points w.r.t.  $c_0$  (transformations used to build equivalence classes)

## INDEPENDENCE TEST

$$(x_1, y_1), \dots, (x_m, y_m) \sim \text{ind}(X, Y)$$

$$H_0: X \perp\!\!\!\perp Y \quad H_1: X \not\perp\!\!\!\perp Y$$

$\Pi$  = CARTESIAN PRODUCT OF PERMUTATIONS OF  $X$  AND  $Y$  SEPARATELY

$$\#\Pi = m! n!$$

IF  $X$  AND  $Y$  ARE BOTH CONTINUOUS

(n.e. SIMPLE REGRESSION)

$$T_0 = p^2$$

this has low power if the data are quadratically dependent and not linearly dependent

IF  $X$  AND  $Y$  ARE BOTH CATEGORICAL

(n.e. FISHER EXACT TEST)

$$X \in C_1, C_2, \dots, C_a$$

$$Y \in D_1, D_2, \dots, D_b$$

$$T_0 = \sum_{i=1}^a \sum_{j=1}^b \frac{(\hat{P}_{ij} - \hat{P}_i \cdot \hat{P}_j)^2}{\hat{P}_i \cdot \hat{P}_j}$$

it compares the estimated probability of each pair of labels w.r.t. another estimate that is valid only under  $H_0$ .

this can be omitted since it's invariant w.r.t.

IF  $X$  CATEGORICAL AND  $Y$  CONTINUOUS

(n.e. ONE-WAY ANOVA)

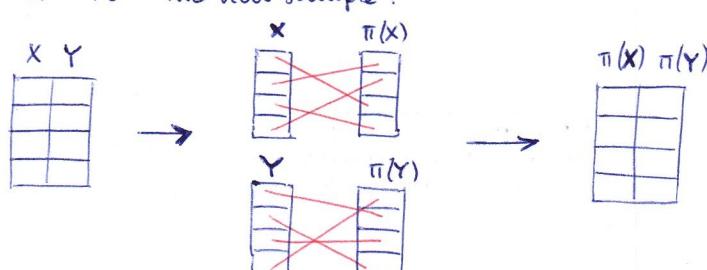
permutations

$$T_0 = \sum_{j=1}^b (\bar{Y}_j - \bar{Y})^2$$

it's comparing the values of the  $Y$ 's given the categories (= given the  $X$ )

this is a re-frame of the one-wayanova that we saw before

Idea: under  $H_0$  the new sample:



$(\pi(X), \pi(Y))$  will have the same probability of being obtained than the original one. This is possible because under  $H_0$  the joint density is the product of the marginals.

Idea: equal to the 1-D case:

Under  $H_0$  the permuted pooled sample has the same probability of being observed as the original pooled sample.  
Under  $H_1$  this is not true.

Note: we permute vectors, not single elements of the vectors (unless we know that the components are II)

## MULTIVARIATE (HOTELLING-LIKE) TESTS

### TWO INDEPENDENT SAMPLES

$$\begin{aligned} X_1, X_{12}, \dots, X_m &\sim \text{iid } \mathbb{X}_1 \in \mathbb{R}^p \\ X_2, X_{22}, \dots, X_{2m_2} &\sim \text{iid } \mathbb{X}_2 \in \mathbb{R}^p \end{aligned} \quad \left. \right\} \text{INDEPENDENT}$$

$$H_0: \mathbb{X}_1 \stackrel{d}{=} \mathbb{X}_2 \quad H_1: \mathbb{X}_1 \neq \mathbb{X}_2$$

(both) possible test statistics

$$\begin{cases} T = \|\bar{\mathbb{X}}_1 - \bar{\mathbb{X}}_2\|_{RP}^2 \\ T = (\bar{\mathbb{X}}_1 - \bar{\mathbb{X}}_2)^T \left( \frac{1}{m_1} S_1 + \frac{1}{m_2} S_2 \right) (\bar{\mathbb{X}}_1 - \bar{\mathbb{X}}_2) \end{cases}$$

SAMPLE COVARIANCE MATRIX OF THE FIRST SAMPLE.

$\Pi$  = GROUP OF PERMUTATIONS (i.e. PERMUTATIONS OF THE  $m_1+m_2$  P-DIMENSIONAL VECTORS)

$$\#\Pi = (m_1+m_2)!$$

ONE SAMPLE (OR PAIRED SAMPLE): center of symmetry of a multivariate population

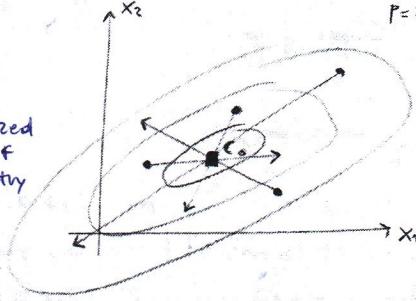
$$\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_m \sim \text{iid } \mathbb{X} \in \mathbb{R}^p \text{ s.t. } \exists c \in \mathbb{R}^p: f(c+\mathbb{X}) = f(c-\mathbb{X}) \forall \mathbb{X} \in \mathbb{R}^p$$

$$H_0: c = c_0 \quad H_1: c \neq c_0$$

$$T = \|\bar{\mathbb{X}} - c_0\|_{RP}^2$$

estimate of the center of symmetry

hypothized center of symmetry ( $\in \mathbb{R}^p$ )



$\Pi$  = CARTESIAN GROUP OF REFLECTIONS WRT  $c_0$  (i.e. REFLECTIONS OF THE ENTIRE VECTOR)

$$\#\Pi = 2^m$$

transformations used instead of the permutations (as in the 1D case)

$$\Pi(\mathbb{X}_1, \dots, \mathbb{X}_m) = (\mathbb{X}_1 + H_1, \dots, \mathbb{X}_m + H_m)$$

$$H_i = \begin{cases} \mathbb{X}_i - c_0 \\ -(\mathbb{X}_i - c_0) \end{cases} \quad \text{OR}$$

Assumptions: linearity of the conditional mean w.r.t. the regressors and an additive action on the response of the error term. (the errors are iid  $\varepsilon$ , not necessarily gaussian but identically distributed)

## MULTIPLE REGRESSION

we consider 2 regressors and 1 response, but everything can be extended

$$Y_i | X_{1i} = x_{1i} \wedge X_{2i} = x_{2i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

$$\varepsilon_1, \dots, \varepsilon_m \sim \text{iid } \varepsilon$$

- **GLOBAL F-TEST** : jointly tests the significance of all the regressors

similar to the test on  $\beta_1$  of the simple regression

$$H_0: (\beta_1, \beta_2) = 0 \quad H_1: (\beta_1, \beta_2) \neq 0$$

CONDITIONALLY ON THE DESIGN MATRIX $Y_1, \dots, Y_m \sim \text{iid } Y$ $Y = \beta_0 + \varepsilon$	CONDITIONALLY ON THE DESIGN MATRIX $Y_1, \dots, Y_m \sim \text{iid } Y$ $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$
---	---

$\Pi$  = GROUP OF PERMUTATIONS (OF RESPONSES)

$$\# \Pi = m!$$

(NB) THE DESIGN MATRIX ALWAYS REMAINS UNCHANGED.

$$T = \left( \frac{\hat{\beta}_{1\text{OLS}}}{S_{\hat{\beta}_{1\text{OLS}}}} \right)^2 + \left( \frac{\hat{\beta}_{2\text{OLS}}}{S_{\hat{\beta}_{2\text{OLS}}}} \right)^2$$

combination of the two marginal test statistics

$$T = \frac{SS_{\text{REG}}}{SS_{\text{RES}}} \sim F$$

this considers also the covariances

- **PARTIAL T-TESTS** : one-at-the-time test the specific regressor

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

COND. ON THE DESIGN MATRIX $Y_i = \beta_0 + \beta_2 x_{2i} + \varepsilon_i$ $Y_i - (\beta_0 + \beta_2 x_{2i}) \sim \text{iid}$	COND. ON THE DESIGN MATRIX $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ $Y_i - (\beta_0 + \beta_2 x_{2i}) \not\sim \text{iid}$
--	---

AN IDEAL BUT UNPRACTICAL SOLUTION WOULD BE PERMUTING THE ERRORS  $\varepsilon_i$  OF THE NULL MODEL:

$\Pi$  = GROUP OF PERMUTATIONS (OF THE ERRORS  $\varepsilon_i = Y_i - (\beta_0 + \beta_2 x_{2i})$  OF THE NULL MODEL)

$$\Pi(Y_i) = \beta_0 + \beta_2 x_{2i} + \Pi(\varepsilon_i)$$

ideal (we don't know  $\varepsilon_i$ )

$$\# \Pi = n!$$

$$T = \left( \frac{\hat{\beta}_{1\text{OLS}}}{S_{\hat{\beta}_{1\text{OLS}}}} \right)$$

under  $H_0$ :

$$Y_i - (\beta_0 + \beta_2 x_{2i}) = \varepsilon_i \sim \text{iid}$$

under  $H_1$ :

$$Y_i - (\beta_0 + \beta_2 x_{2i}) = \beta_1 x_{1i} + \varepsilon_i \not\sim \text{iid}$$

a good idea is to permute the error and see what happens, however the error (true) is not observable

Procedure: we take the dataset and we permute the errors. For every permutation of the errors we obtain a new set of  $Y$ 's (of responses) and for each set of  $Y$ 's we fit a full model (with  $\beta_1$ !) and we estimate  $\beta_1$ .

Then we measure the distance between the estimated  $\beta_1$  and 0 (since  $H_0: \beta_1 = 0$ )

UNFORTUNATELY THE ERRORS  $\epsilon_i = Y_i - (\beta_0 + \beta_1 x_{1i})$  ARE  
NOT KNOWN  $\Rightarrow$  WE NEED APPROXIMATE SOLUTIONS BASED  
 ON GOOD APPROXIMATIONS OF THE ERRORS.

### FREEMAN AND LANE SCHEME (1983)

(RESIDUALS OF THE NULL MODEL)  $\rightarrow$  error under  $H_0 : Y_i - (\text{model}_i)_{H_0}$

$$\epsilon_i \text{ and } \hat{\epsilon}_i^{H_0} = Y_i - (\hat{\beta}_0^{H_0} + \hat{\beta}_1^{H_0} x_{1i}) \quad \rightarrow$$

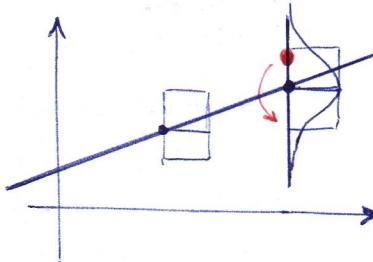
ASYMPTOTIC EXACTNESS

(N.B.) WHEN ERRORS ARE FAR FROM BEING NORMAL  
 ASYMPTOTIC PERMUTATION TESTS REACH THE  
 NOMINAL LEVEL OF SIGNIFICANCE FASTER THAN  
 ASYMPTOTIC PARAMETRIC TESTS.

(N.B.) WHEN THE NULL HYPOTHESIS DOES NOT FULLY  
 DEFINE THE NULL MODEL APPROXIMATE TESTS CAN  
 BE BUILT

(N.B.) IF THE ERRORS ARE KNOWN TO BE SYMMETRICALLY  
 DISTRIBUTED, IT CAN BE ENLARGED ALSO TO SIGN-FLIPPINGS  
 OF THE RESIDUALS.

(the larger the cardinality the more powerful the test)



#### Procedure:

- we fit the null model with the data (here:  $\beta_0 + \beta_1 x_2$ )
- we compute the residuals (estimates of the error)
- we permute the residuals and obtain a permuted set
- for each permutation of the residuals we fit the full model to the data and compute, for instance,  $\hat{\beta}_1$

(or the value of the test statistic in general)

- The null model enters for estimating the residuals
- The complete model enters for the computation of the test statistic

We can use confidence intervals to build tests just introducing as rejection criterium the fact that the CI captures or not the hypothesized value of the parameter. It's a valid procedure because under  $H_0$  the probability that the CI captures the value is  $1-\alpha \rightarrow$  we do not reject when we capture the hypothesized value and we reject when we do not capture it, obtaining a test with significance level  $\alpha$ . On the other hand, the opposite is possible too : from tests to CI :

### PERMUTATIONAL CONFIDENCE INTERVALS

WHENEVER  $H_0$  IS DEFINED AS A FUNCTION OF A PARAMETER A "CI" CAN BE BUILT FROM AN HYPOTHESIS TEST.

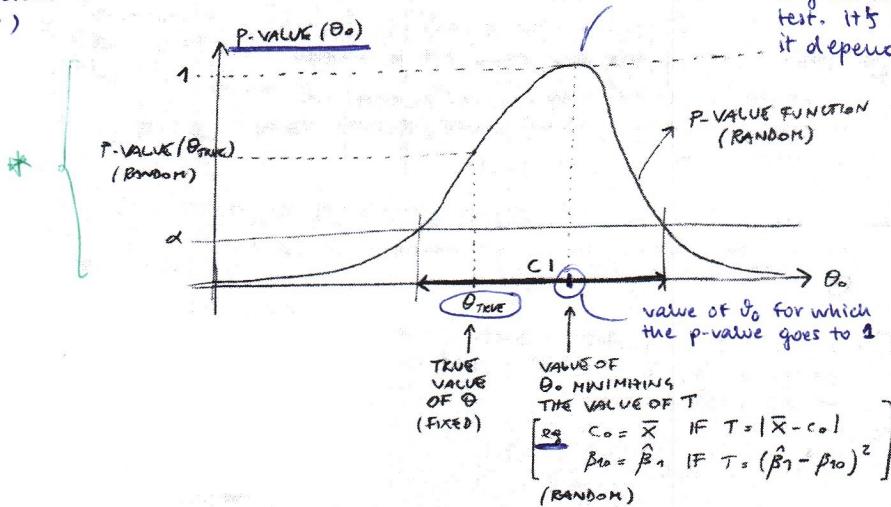
(e.g. CENTER OF SYMMETRY  
REGRESSION COEFFICIENTS)

$$H_0: \theta = \theta_0 \quad H_1: \theta \neq \theta_0$$

$\theta$  generic parameter  
(e.g. center of symmetry,  
value of the coefficient  
of a regressor, ...)

] we have a test that  
provide us a p-value:  
we want to use the  
p-value to build a  
confidence interval

p-value function := given a possible value of  $T_0$   
it gives the p-value of the corresponding  
test. It's a random function because  
it depends on the random data  
that we use to run the test



\* means "go to the first reachable  $\alpha$  (from below)"

$$C_1(1-\alpha) = \{ \theta_0 : P\text{-VALUE}(\theta_0) \geq [1-\alpha] \}$$

random interval (since it depends on random data) which has the probability  $1-\alpha$  of capturing the actual value of the parameter

$$\begin{aligned} &P\text{-VALUE}(\theta_{\text{true}}) \sim \text{UNIF} \\ &P(\theta_{\text{true}} \in C_1) = P(P\text{-VALUE}(\theta_{\text{true}}) \geq [1-\alpha]) \geq 1-\alpha \end{aligned}$$

(N.B.) IT IS FOR INSTANCE POSSIBLE TO BUILD FINITE-SAMPLE EXACT CI FOR THE CENTER OF SYMMETRY OF NON GAUSSIAN DISTRIBUTIONS

(N.B.) IF  $\alpha$  IS A REACHABLE LEVEL OF THE TEST

$$P(\theta_{\text{true}} \in C_1) = 1-\alpha$$

OTHERWISE  $\alpha$  IS FLOORED TO THE FIRST REACHABLE LEVEL  
 $P(\theta_{\text{true}} \in C_1) > 1-\alpha$

(N.B.) COMPUTATION IS NOT TRIVIAL BECAUSE EACH EVALUATION OF THE P-VALUE FUNCTION IS BASED ON A DIFFERENT CMR RUN.

(REMEMBER) in permutation tests, \* is not continuous,  
is discretized ( $\rightarrow$  piecewise constant function)

(Generally we know: the p-value is discretized)

Note: it might be a confidence region instead of the interval, in the sense that it may be:

instead of