Hypothesis Testing summary

Anestis Christidis, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.4.1

Artificial Intelligence &
Information Analysis Lab

Hypothesis Testing

- Elementary Principles
- NSHT & BHT
- Tests
  - Tests comparing mean values (T-test, Z-test)
  - Tests detecting normal distribution
    - Chi-Squared test
    - Mardia’s test
  - Tests determining distribution type
    - Anderson-Darling Test
    - Kolmogorov-Smirnov Test

Hypothesis Testing

- $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ a random variable, where $\mathbf{X}_i$ the vector of measurements for the i^{th} object.

- Statistical Hypothesis:
  - Statement about the distribution of $\mathbf{X}$.
  - Specified a set of *possible* distributions of $\mathbf{X}$.
  - Simple Hypothesis: specifies a single distribution of $\mathbf{X}$.
  - Composite Hypothesis: specifies multiple distributions of $\mathbf{X}$.

Artificial Intelligence & Information Analysis Lab

Hypothesis Testing

- Two hypotheses types, *null* and *alternate*.

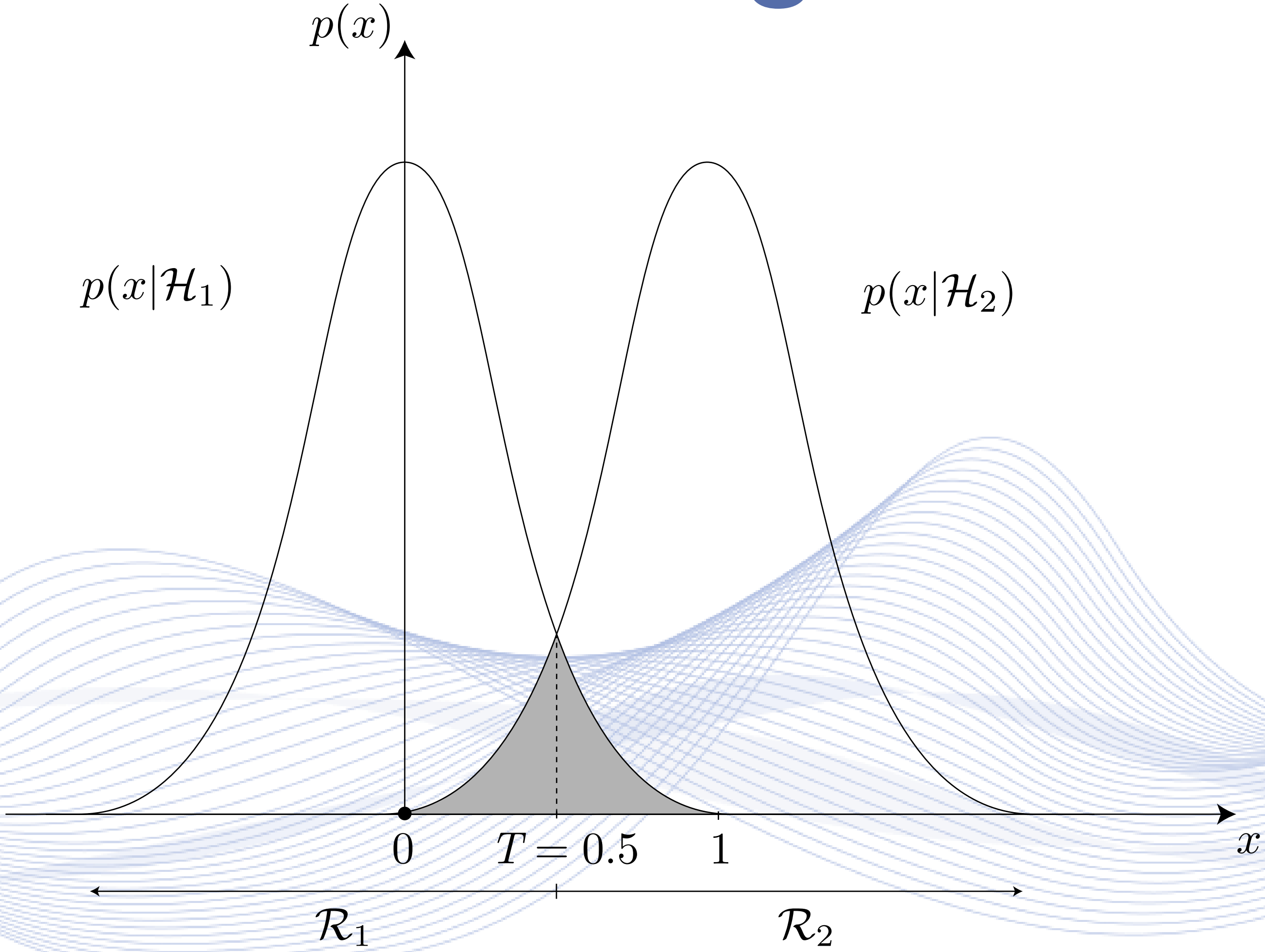
- Null Hypothesis
  - Or Hypothesis 0 ($\mathcal{H}_0$).
  - The default assumption, nothing has changed in the test.
- Alternate Hypothesis
  - Or Hypothesis 1 ($\mathcal{H}_1$).
  - Another hypothesis holds for the test, and $\mathcal{H}_0$ is not valid.

Hypothesis Testing

- One of the two hypotheses ($\mathcal{H}_0$ / $\mathcal{H}_1$) will be true.
- Depending on the ultimate decision there can be error.
- There are two types of error:

- Type 1 Error: reject $\mathcal{H}_0$, when $\mathcal{H}_0$ is true.
- Type 2 Error: fail to reject $\mathcal{H}_0$, when $\mathcal{H}_1$ true.

Hypothesis Testing



Type 1 Error and Type 2 Errors.

Artificial Intelligence & Information Analysis Lab

Hypothesis Testing

- Introduction
- **Elementary Principles**
- NSHT & BHT
- Tests
  - Tests comparing mean values
  - Tests detecting normal distribution
  - Tests detecting distribution type

Significance level

- Significance level / size of critical region α.
- The maximum probability of a type 1 error, over the set of distributions described by $\mathcal{H}_0$.
- Small values (usually 0.1, 0.05, 0.01)

Power

- Suppose $\mathcal{H}_0$ is true
- Thus the distribution of $\mathbf{X}$ is specified by $\mathcal{H}_1$.
- The probability of rejecting $\mathcal{H}_0$, $\mathrm{P}(\mathbf{X} \in \mathcal{R})$, is the power of the test for the specific distribution.

P-values

- $\mathcal{R}_\alpha$: the rejection region for a given significance level α.
- Both are linearly related.

- The P-value of the observed $\mathbf{x}$ of $\mathbf{X}$ ($P(\mathbf{x})$), is the smallest α such that $\mathbf{x} \in \mathcal{R}_\alpha$.

- That is, the smallest significance level for which $\mathcal{H}_0$ is rejected, given $\mathbf{x} = \mathbf{X}$.

Critical Values

- Some tests, instead of P-values, return critical values c and associated significance levels.
- Similar approach to the P-values one. Instead of comparing $P(\mathbf{x})$ to the significance value, we compare it to c.

- If $P(\mathbf{x}) \leq c$, we reject $\mathcal{H}_0$.
- If $P(\mathbf{x}) > c$, we fail to reject $\mathcal{H}_0$.

Hypothesis Testing

- Introduction
- Elementary Principles
- **NSHT & BHT**
- Tests
  - Tests comparing mean values
  - Tests detecting normal distribution
  - Tests detecting distribution type

Null Hypothesis Significance Test

- A test using P-values is called a *Null Hypothesis Significance Test (NHST).*

- While useful in many cases, NHST has some major setbacks regarding data interpretation, and has been subject to criticism.

Bayesian Hypothesis Testing

- *Bayesian Hypothesis Testing (BHT)* is a candidate alternative to the NHST approach.

- In contrast to the latter, BHT calculates new probabilities based on a-priori data, making use of the Bayes Decision Rule.

Maximum A Posteriori Test

- Having the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, we choose the former if and only if:

$$P(\mathcal{H}_0|\mathbf{X}=\mathbf{x}) \geq P(\mathcal{H}_1|\mathbf{X}=\mathbf{x}) \Leftrightarrow$$

$$f_{\mathbf{X}}(\mathbf{x}|\mathcal{H}_0)P(\mathcal{H}_0) \geq f_{\mathbf{X}}(\mathbf{x}|\mathcal{H}_1)P(\mathcal{H}_1)$$

- This can be generalized for any number of hypothesis by opting for the hypothesis $\mathcal{H}_i$ with the highest $f_{\mathbf{X}}(\mathbf{x}|\mathcal{H}_i)P(\mathcal{H}_i)$.

Minimum Cost Hypothesis Test

- Suppose:
  - C_{10}: the cost of choosing $\mathcal{H}_0$ when $\mathcal{H}_1$ is true.
  - C_{01}: the cost of choosing $\mathcal{H}_1$ when $\mathcal{H}_0$ is true.

- Then we choose H_0 if and only if:

$$\frac{f_{\mathbf{X}}(\mathbf{x}|\mathcal{H}_0)}{f_{\mathbf{X}}(\mathbf{x}|\mathcal{H}_1)} \geq \frac{P(\mathcal{H}_1)C_{01}}{P(\mathcal{H}_0)C_{10}} \Leftrightarrow P(\mathcal{H}_0|\mathbf{x})C_{10} \geq P(\mathcal{H}_1|\mathbf{x})C_{01}$$

Artificial Intelligence & Information Analysis Lab

Hypothesis Testing

- Introduction
- Elementary Principles
- NSHT & BHT
- **Tests**
  - Tests comparing mean values
  - Tests detecting normal distribution
  - Tests detecting distribution type

Hypothesis Testing

- We can now move on to more commonly used types of tests.
- The following tests will be analysed:
  - T-test
  - Z-test
  - Chi-Squared test
  - Anderson-Darling test
  - Kolmogorov-Smirnoff test

Hypothesis Testing

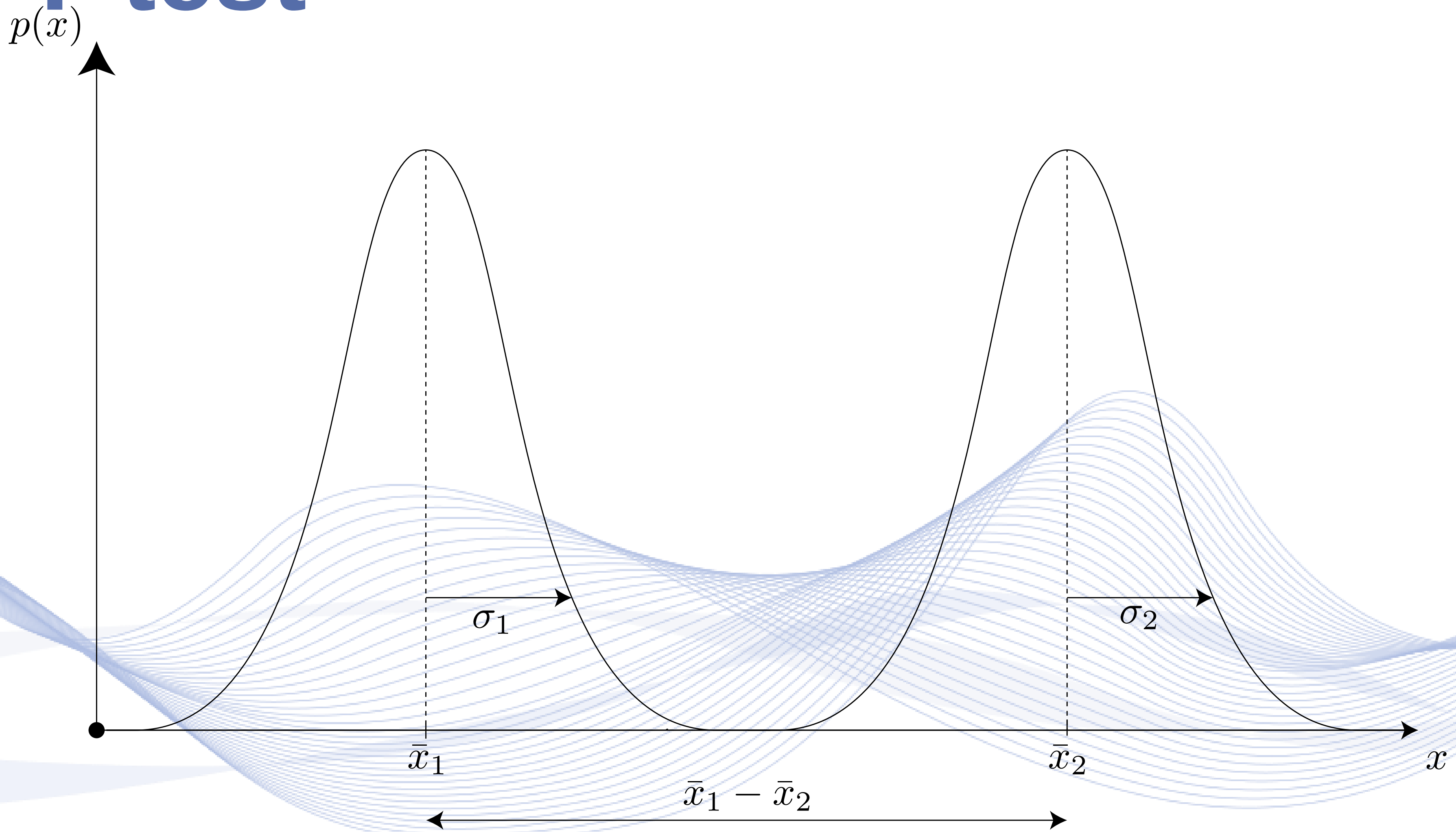
- Introduction
- Elementary Principles
- NSHT & BHT
- Tests
  - **Tests comparing mean values**
  - Tests detecting normal distribution
  - Tests detecting distribution type

T-test

- The T-test measures the significance of the difference between two groups.
- The comparison factor is the mean value of each group.
- The result is expressed with the *t-score*.

- The t-score is a ratio of the difference between two groups, and the one within those groups.
- The larger the t-score, the more different the groups are.

T-test



Independent Samples T-test.

Z-test

- The Z-test is similar in many ways to the T-test.

- However some additional constraints should be satisfied:
  - The sample size n is greater than 30.
  - The data is approximately normally distributed.
  - The data points are independent from each other.
  - The samples are equiprobable in terms of selection.
  - The sample sizes are preferably equal.

Z-test

- First, after a significance level is chosen, we lookup the appropriate Z-score in a Z-table.
- Then the Z statistic is calculated:

$$Z = \frac{\bar{\mathbf{x}} - m}{\sqrt{\frac{\sigma^2}{n}}}$$

- If the calculated Z-score is greater than the Z-score of the chosen significance level, $\mathcal{H}_0$ is rejected.

Hypothesis Testing

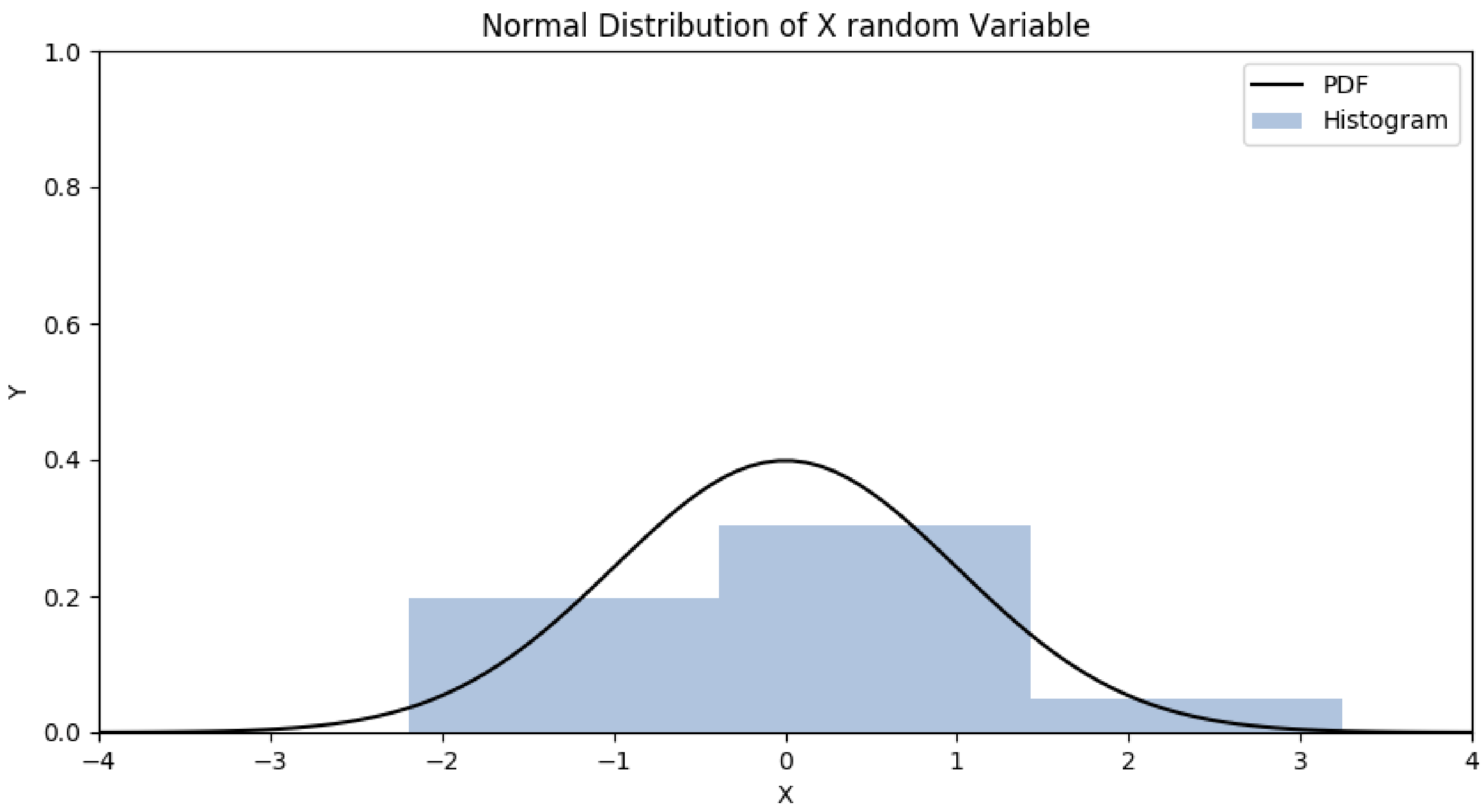
- Introduction
- Elementary Principles
- NSHT & BHT
- Tests
  - Tests comparing mean values
  - **Tests detecting normal distribution**
  - Tests detecting distribution type

Chi-Squared test

- The Chi-Squared test allows us to determine whether a model follows a normal distribution or not.
- The null hypothesis ($\mathcal{H}_1$) is that the data is sampled from a normal distribution.
- The test uses the Chi-Squared statistic:

$$X_c^2 = \sum \frac{(\mathbf{x}_i - \hat{\mathbf{x}}_i)^2}{\hat{\mathbf{x}}_i}$$

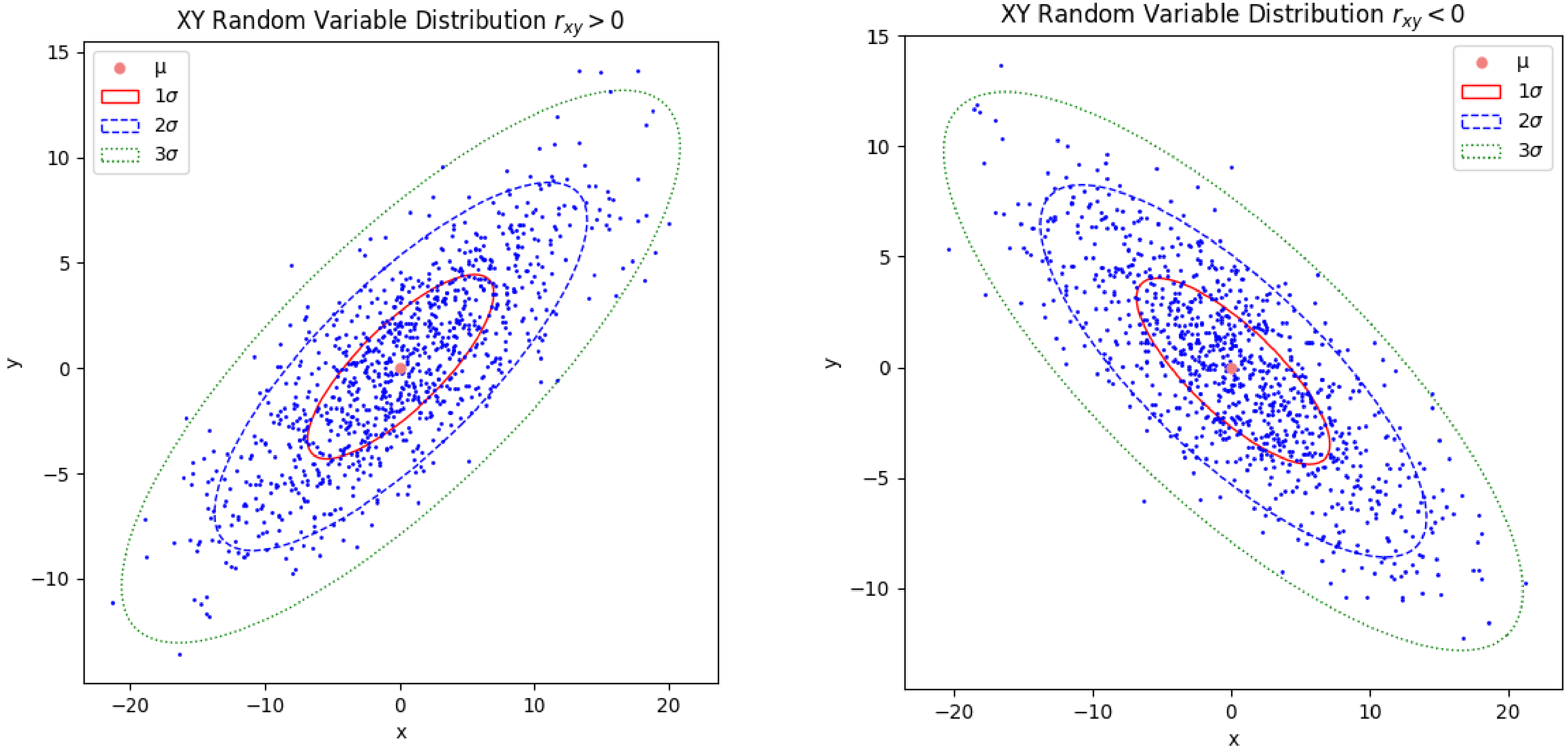
Chi-Squared test



Samples drawn from a 1D Gaussian distribution.

Artificial Intelligence &
Information Analysis Lab

Multivariate Gaussian tests



Data samples drawn from 2D Gaussian pdfs.

Mardia’s test

- However, it is more common to work with multi-variable data.
- Mardia’s test (named after Kantilal Mardia) can help us manage such situations.
- It is based on the comparison of the data’s skewness and kurtosis to the respective ones of a normal distribution.

Mardia's test

- Suppose sample vector $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_\mathrm{n}\}$.
- We calculate the skewness of the p-variate data:

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{C}^{-1} \left(\mathbf{x}_j - \bar{\mathbf{x}} \right) \right]^3$$

and then the kurtosis:

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^{n} [(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{C}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})]^2$$

Hypothesis Testing

- Introduction
- Elementary Principles
- NSHT & BHT
- Tests
  - Tests comparing mean values
  - Tests detecting normal distribution
  - **Tests detecting distribution type**

Anderson-Darling Test

- The Anderson-Darling Test helps us determine how well the sampled data fits a specific distribution.

- It is named after Theodore W. Anderson and Donald A. Darling, who invented it.

- Most commonly used for normal distribution testing but can be applied to other distributions as well.

Anderson-Darling Test

- The statistic used for this test is:

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}\Big((2i-1)\big[ln\Phi(\mathbf{X}_i) + \ln\big(1-\Phi(\mathbf{X}_{n-i+1})\big)\big]\Big)$$

where:

- n is the sample size
- $\Phi(\mathbf{X})$ is the Cumulative Distribution Function for the specified distribution.
- i is the i^{th} sample when the data is ascendingly sorted

Kolmogorov-Smirnov Test

- Named after Andrey Kolmogorov and Nikolai Smirnov.

- This test compares the sampled data with a known distribution and determines whether it comes from the same distribution.

- It is a non-parametric test, meaning nothing is assumed of the underlying distribution.

Kolmogorov-Smirnov Test

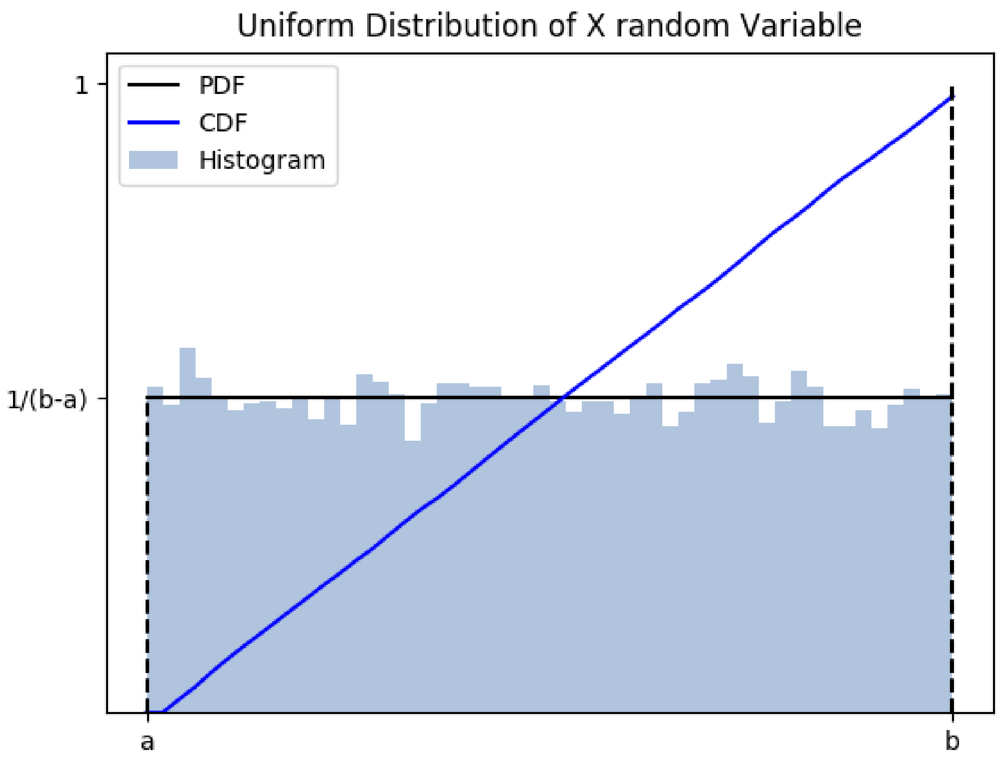
- The test statistic is given by the following formula:

$$D_i = \sup_{\mathbf{x}} |\Phi_i(\mathbf{x}) - \Phi_{data}(\mathbf{x})|$$

where

- $\Phi_i(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} I_{[-\infty,\mathbf{x}]}(\mathbf{X}_i)$, is the cumulative density function (c.d.f.) of the hypothesized distribution, $I_{[-\infty,\mathbf{x}]}$ being the indicator function.
- $\Phi_{data}(\mathbf{x})$ is the empirical distribution function of the observed data.

Kolmogorov-Smirnov Test



Samples drown from a 1D Uniform distribution.

Bibliography

- M.G. Strintzis, Pattern Recognition, 1999.
- S. Theodoridis, K. Koutroumbas, Pattern Recognition, 2011
- The Double Exponential Distribution: Using Calculus to Find a Maximum Likelihood Estimator, Robert M. Norton, The American Statistician, Vol. 38, No. 2 (May 1984), pp. 135-136, Retrieved from: https://www.jstor.org/stable/2683252 .
- Mean and Median, StatTrek.com, viewed: 24/4/2020, Website URL: https://stattrek.com/descriptive-statistics/mean-median.aspx.
- Statistical Hypothesis Tests, MachineLearningMastery.com, viewed: 15/4/2020, Website URL: https://machinelearningmastery.com/statistical-hypothesis-tests/.
- Hypothesis Testing, RandomServices.org, viewed: 15/4/2020, Website URL: http://www.randomservices.org/random/hypothesis/index.html.
- Statistical Tests: When to Use Which, viewed: 15/4/2020, TowardsDataScience.com, Website URL: https://towardsdatascience.com/statistical-tests-when-to-use-which-704557554740.
- Bayesian Hypothesis Testing: An Alternative to Null Hypothesis Significance Testing, InTechOpen.com, viewed: 15/4/2020, Website URL: https://www.intechopen.com/books/bayesian-inference/bayesian-hypothesis-testing-an-alternative-to-null-hypothesis-significance-testing-nhst-in-psychology.
- Chi-Square Test, StatisticsHowTo.com, viewed: 15/4/2020, Website URL: https://www.statisticshowto.com/chi-square-test-normality/.
- Kres H. (1983) The Mardia-Test for Multivariate Normality, Skewness, and Kurtosis: Tables by K. V. Mardia. In: Statistical Tables for Multivariate Analysis. Springer Series in Statistics. Springer, New York, NY.
- T-Test, StatisticsHowTo.com, viewed: 15/4/2020, Website URL: https://www.statisticshowto.com/probability-and-statistics/t-test/.
- Z-Test, StatisticsHowTo.com, viewed: 15/4/2020, Website URL: https://www.statisticshowto.com/z-test/.
- Kolmogorov-Smirnov Test, StatisticsHowTo.com, viewed: 15/4/2020, Website URL: https://www.statisticshowto.com/kolmogorov-smirnov-test/.
- Anderson-Darling Test, StatisticsHowTo.com, viewed: 15/4/2020, Website URL: https://www.statisticshowto.com/anderson-darling-test/.
- Kolmogorov-Smirnov Test, Wikipedia.com, viewed: 15/4/2020, Website URL: https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test.
- Anderson-Darling Test, Wikipedia.com, viewed: 15/4/2020, Website URL: https://en.wikipedia.org/wiki/Anderson%E2%80%93Darling_test.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr