Parameter estimation summary

A. Christidis, Prof. I. Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.3.1

Artificial Intelligence & Information Analysis Lab

Parameter estimation

- Data analysis needs:
  - Probabilistic data modeling
  - Estimation of pdf parameters
- Pdf parameters to be estimated
  - Location and dispersion parameters
- Maximum Likelihood Parameter Estimation
  - ML Estimation for Gaussian Distributions
  - ML Estimation for Laplacian Distributions
  - Robustness of arithmetic mean and median
- Maximum a Posteriori Probability Estimation

Maximum Likelihood Parameter Estimation

- $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ random samples following joint probability distribution $p(\mathbf{x}; \boldsymbol{\theta})$.

- $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$: set of the N data samples.

- Assuming statistical independence between the different samples:

$$p(\mathcal{X}; \boldsymbol{\theta}) \equiv p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N; \boldsymbol{\theta}) = \prod_{k=1}^{N} p(\mathbf{x}_k; \boldsymbol{\theta}).$$

Artificial Intelligence & Information Analysis Lab

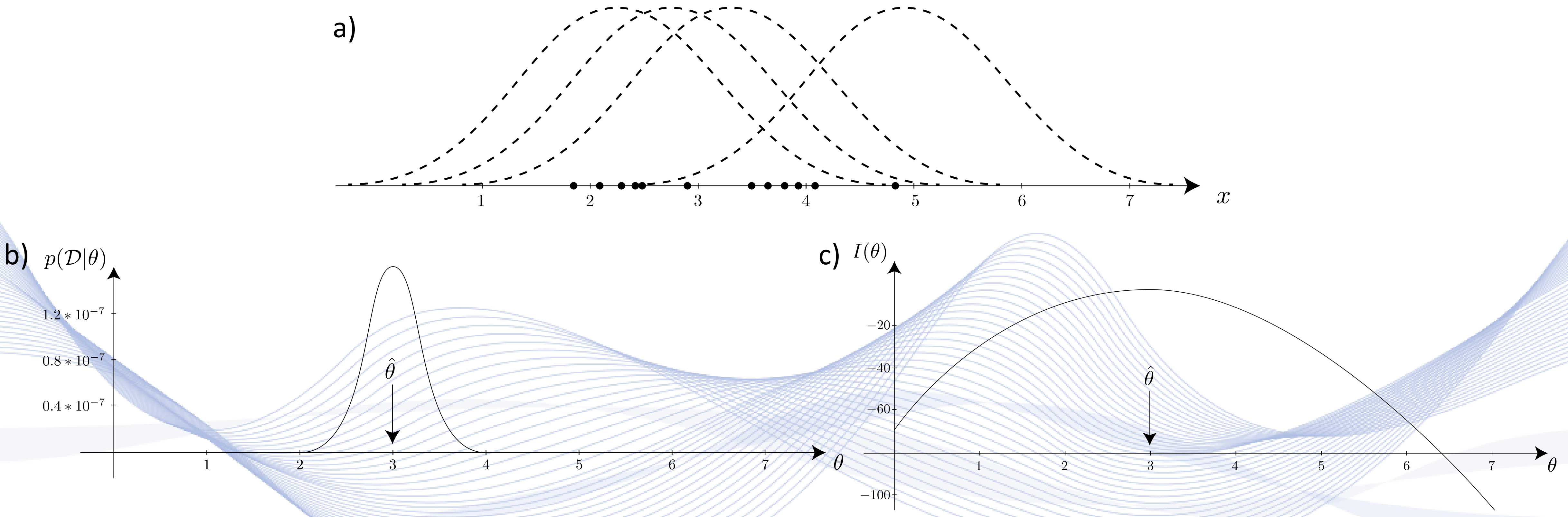
Maximum Likelihood Parameter Estimation

Maximum likelihood (ML) parameter estimation:

- Pdf parameter vector $\boldsymbol{\theta}$ is estimated so that the likelihood function is maximized:

$$\hat{\boldsymbol{\theta}} = arg\max_{\boldsymbol{\theta}} \prod_{k=1}^{N} p(\mathbf{x}_k; \boldsymbol{\theta}).$$

ML Estimation for Gaussian Distributions



a) Pdf of unknown location and various data samples. b) Likelihood function; c) Log-likelihood function.

1D Propability Distributions

- Normal (Gaussian) distribution $N(0,1)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}.$$

- Uniform distribution $\mathrm{U}(0,1)$:

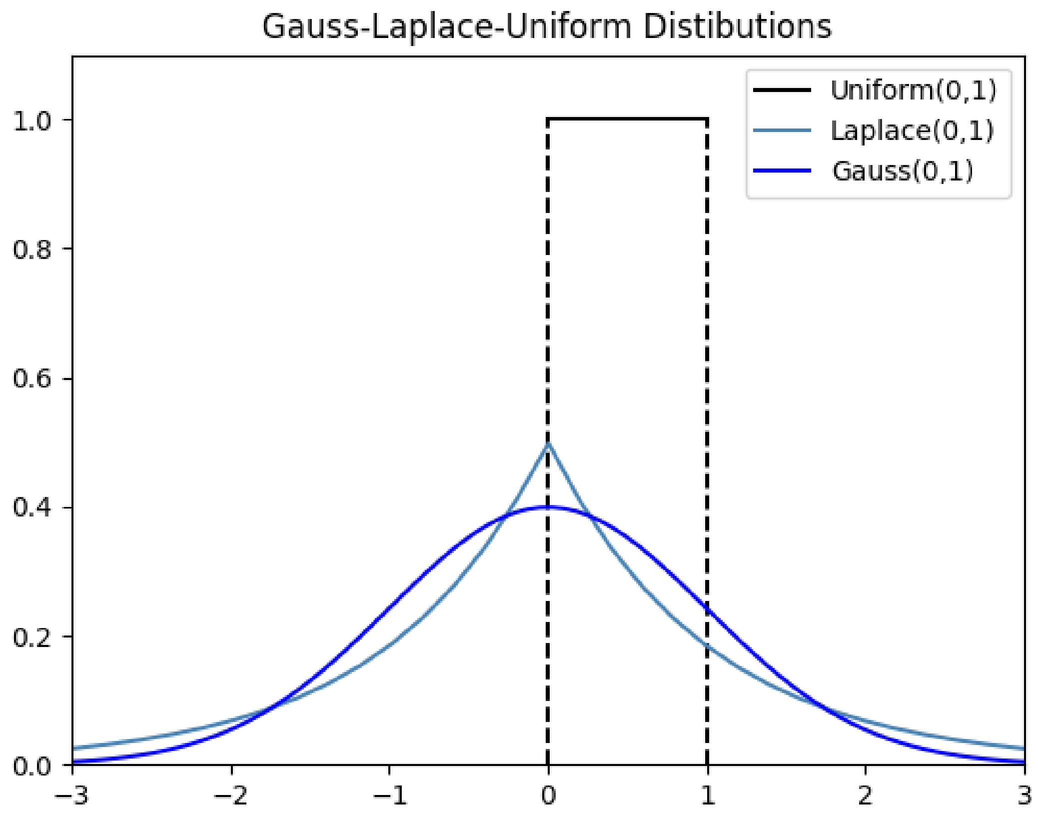
$$f_X(x) = 1, \qquad x \in [0,1].$$

- Laplacian distribution:

$$f_X(x) = \frac{1}{2} e^{-|x|}.$$

1D Propability Distributions

- 1D probability distributions:
  - Uniform (short tailed)
  - Gaussian
  - Laplacian (short tailed)



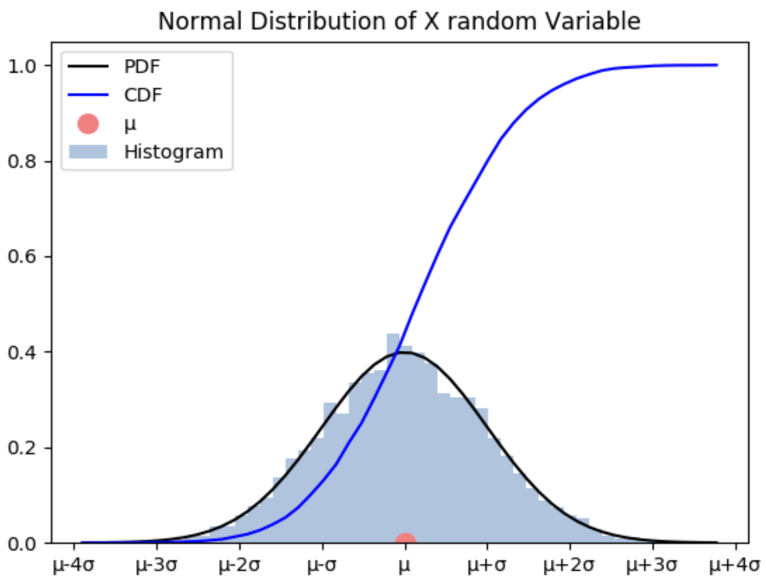
1D Propability Distributions

- Normal (Gaussian) distribution $N(m, \sigma)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}.$$

- 1D Laplacian probability distribution:

$$f_X(x) = \frac{1}{2} e^{-|x-a|}.$$



Multidimension Gaussian distribution

- Multidimension Gaussian distribution:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\mathbf{C})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right\}.$$

- $\mathbf{C}$: covariance matrix.
- $\det(\mathbf{C})$: determinant of $\mathbf{C}$.

Location and dispersion estimation

Typically, probability distributions are characterized by their:

- Location and
- Dispersion

Data samples $\mathbf{x}_1$, $\mathbf{x}_2$,…, $\mathbf{x}_N$ are used to estimate the respective parameters.

- In 1D Gaussian distribution, the parameters to be estimated are: m, σ.
- In multidimensional Gaussian distribution, the parameters to be estimated are: $\mathbf{m}$ (n parameters), $\mathbf{C}$ (n^2 parameters).

ML Estimation for Gaussian Distributions

- In making estimation for Gaussian Distributions, we can discern two cases.

- Unknown mean value $\mathbf{m}$.
- Unknown mean value $\mathbf{m}$, and covariance matrix $\mathbf{C}$.

- There is also the case of estimator biases.

ML Estimation for Gaussian Distributions

- The ML estimation for the mean value should satisfy

$$\sum_{k=1}^{N} \mathbf{C}^{-1}(\mathbf{x}_k - \mathbf{m}) = \mathbf{0}\,.$$

- By multiplying with the covariance matrix, we get:

$$\hat{\mathbf{m}} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_k.$$

ML Estimation for Gaussian Distributions

- *Unknown mean value and covariance matrix*

- We do not know $\mathbf{m}$, $\mathbf{C}$ and assume that their $n+n^2$ unknown parameters form parameter vector $\boldsymbol{\theta}$.

- The case for a single variable will be examined having two unknowns:
  - Location parameter m.
  - Dispersion parameter σ^2.

ML Estimation for Gaussian Distributions

- In the above conditions, $\hat{\theta}_1$ and $\hat{\theta}_2$ are the ML estimations for θ_1 and θ_2 respectively.
- By substituting $\hat{\theta}_1$ with $\widehat{m}$ and $\hat{\theta}_2$ with $\hat{\sigma}^2$, we get the following ML estimations:

$$\widehat{m} = \frac{1}{N}\sum_{k=1}^{N} x_k \,, \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{k=1}^{N} (x_k - \widehat{m})^2.$$

ML Estimation for Gaussian Distributions

- For multiple variable estimations, the formulae adapt to the new conditions:

$$\hat{\mathbf{m}} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_k , \qquad \hat{\mathbf{C}} = \frac{1}{N}\sum_{k=1}^{N} (\mathbf{x}_k - \hat{\mathbf{m}})(\mathbf{x}_k - \hat{\mathbf{m}})^T$$

- Once again, the ML estimation for the mean value, is just the average sample value.

ML Location Estimation for 1D Laplacian Distribution

- Now suppose the samples $x_i, i = 1, \ldots, N$ come from an 1D Laplacian probability distribution:

$$p(x;\theta) = \frac{1}{2} e^{-|x-\theta|}.$$

- The log-likelihood function $L(\theta)$ has the form:

$$L(\theta) = -2ln2 - \sum_{i=1}^{N} |x_i - \theta|.$$

Robustness of arithmetic mean and median

For N data samples $x_i, i = 1, \ldots, N, N = 2\nu + 1$:

- The arithmetic mean is given by:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$$

and the median $x_{(\nu+1)}$.

- Any outlier $c \gg x_i, i = 1, \ldots, N$ or $c \ll x_i,, i = 1, \ldots, N$:
  - influences significantly the arithmetic mean.
  - the median remains robust if less then ν outliers occur.

Maximum a Posteriori Probability Estimation

- We will consider $\boldsymbol{\theta}$ as a random vector.
- Estimate its value on the condition that N data samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ are known.
- Starting point is: $p(\boldsymbol{\theta}|\mathbf{X})$.
- From Bayes theorem:

$$p(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{X}) = p(\boldsymbol{\theta}|\mathbf{X})p(\mathbf{X}) \text{ or } \quad p(\boldsymbol{\theta}|\mathbf{X}) = \frac{p(\boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta})}{p(\mathbf{X})}$$

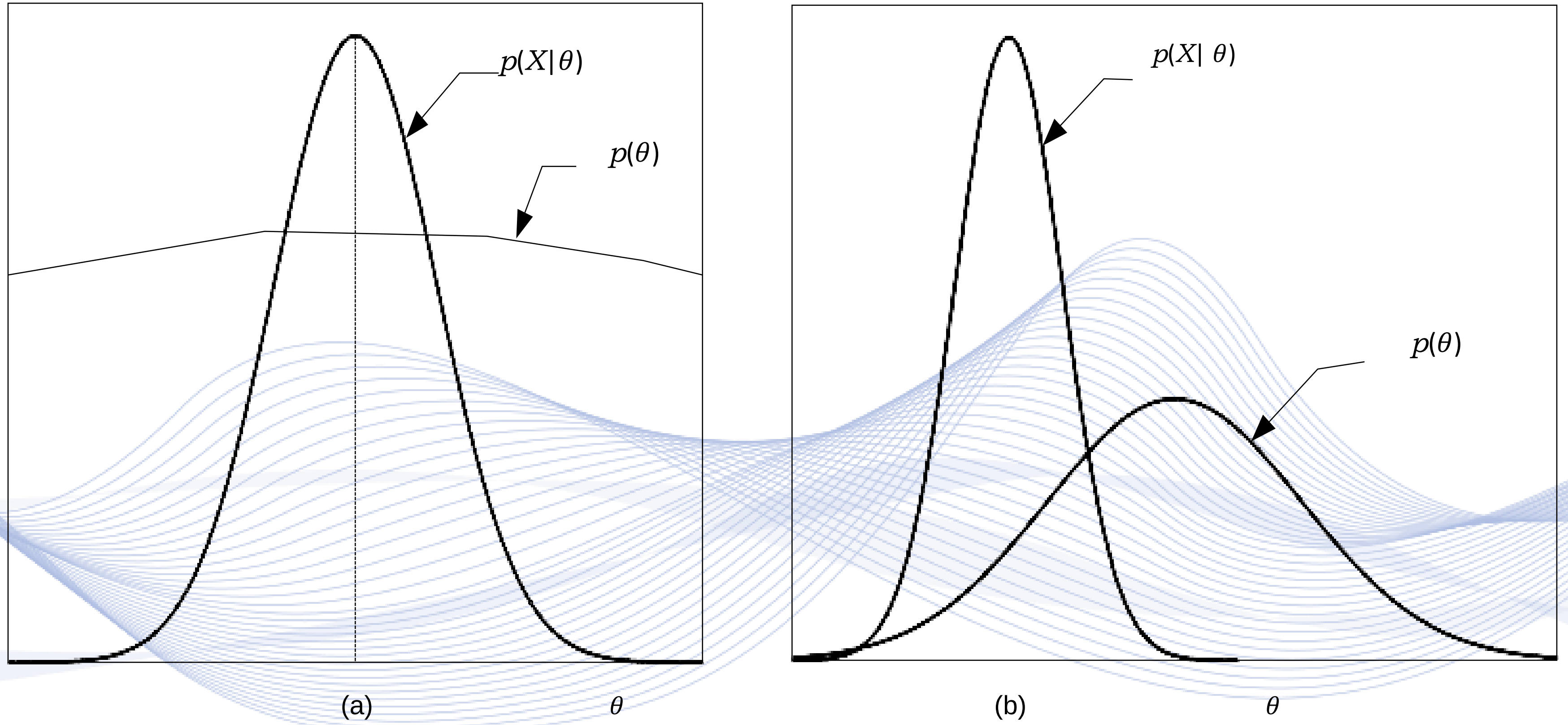
Maximum a Posteriori Probability Estimation

- The maximum a posteriori probability (MAP) estimate $\hat{\boldsymbol{\theta}}_{MAP}$ is defined at the point where $p(\boldsymbol{\theta}|\mathbf{X})$ becomes maximum.

$$\nabla_{\boldsymbol{\theta}}\, p(\boldsymbol{\theta}|\mathbf{X}) = \nabla_{\boldsymbol{\theta}}\big(p(\boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta})\big) = \mathbf{0}.$$

- The difference between the ML and the MAP estimates lies in the involvement of $p(\boldsymbol{\theta})$.
- If it is assumed that $p(\boldsymbol{\theta})$ is uniform distribution, ML and MAP of $\boldsymbol{\theta}$ are identical.

Maximum a Posteriori Probability Estimation



Parameter estimation uses

- Use in Bayesian Learning
- Use in Hypothesis Testing:
  - Whether a data point belongs to a certain pdf
  - Whether a data point belongs to one of two pdfs
  - Whether two pdfs are identical.

References

- 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