# Elements of Statistics (Part II)

Prof. Americo Cunha Jr

Rio de Janeiro State University – UERJ

americo.cunha@uerj.br

www.americocunha.org

@AmericoCunhaJr @AmericoCunhaJr

@AmericoCunhaJr @AmericoCunhaJr

# Nonparametric Estimators

# Empirical Distribution Function

$X_1 < X_2 < \cdots < X_n$ are independent observations of $X$

The empirical distribution function (empirical CDF) is an estimator for distribution function $F_X$ defined by

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}\left(X_i \leq x\right),$$

where

$$\mathcal{I}\left(X_i \leq x\right) = \begin{cases} 1 & \text{if} \quad X_i \leq x \\ 0 & \text{if} \quad X_i > x. \end{cases}$$

*Empirical CDF is consistent and unbiased estimator*

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# Histogram Estimator

$X_1 < X_2 < \cdots < X_n$ are independent observations of $X$

Split the support of $X$ into a denumerable number of bins $\mathcal{B}_m$ with width $h_m$, i.e.,

$$\operatorname{Supp} X = \bigcup_{m=-\infty}^{+\infty} \mathcal{B}_m = \left[(m-1)\, h_m, m\, h_m\right]$$

The histogram is an estimator for probability density function $p_X$ defined by

$$\widehat{p}_n(x) = \sum_{m=-\infty}^{+\infty} \frac{\nu_m}{n\, h_m} \mathbb{1}_{\mathcal{B}_m}(x),$$

where $\nu_m$ is the number of samples of $X$ in $\mathcal{B}_m$ and

$$\mathbb{1}_{\mathcal{B}_m}(x) = \begin{cases} 1 & \text{if} \quad x \in \mathcal{B}_m \\ 0 & \text{if} \quad x \notin \mathcal{B}_m. \end{cases}$$

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# Kernel Density Estimator

$X_1, X_2, \cdots, X_n$ are observations of $X$

The kernel density estimator for the probability density function $p_X$ is defined by

$$\widehat{p}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right),$$

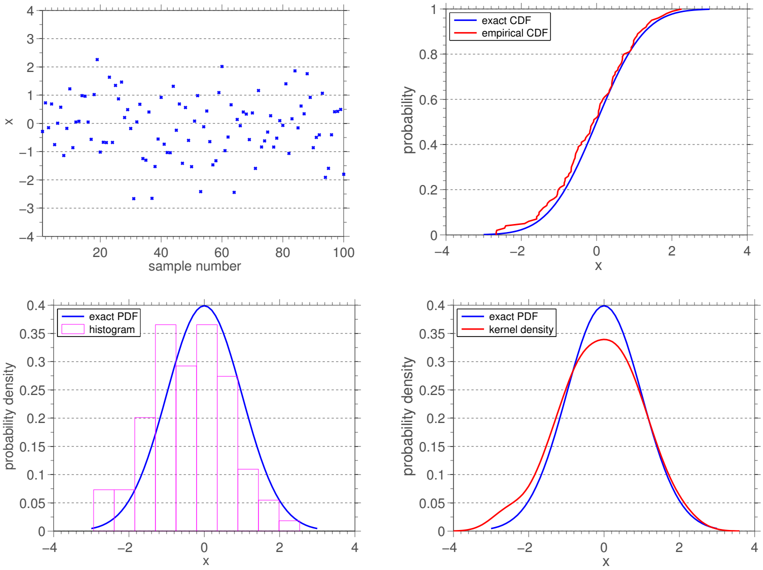
where $h > 0$ is the estimator bandwidth and the kernel $K$ is a smooth function such that

- $K(x) \geq 0$
- $\int_{\mathbb{R}} K(x)\, dx = 1$
- $\int_{\mathbb{R}} x\, K(x)\, dx = 0$
- $\int_{\mathbb{R}} x^2\, K(x)\, dx > 0$

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# An example in nonparametric estimation

100 samples of $X \sim \mathcal{N}(0, 1)$



# Parametric Estimators

# Statistical Moments

Let random variable $X$ be parametrized by vector parameter

$$\theta = (\theta_1, \theta_2, \cdots, \theta_k).$$

For $1 \leq j \leq k$, the j-th moment of $X$ is

$$\alpha_j(\theta_1, \theta_2, \cdots, \theta_k) = \mathbb{E}\left\{X^j\right\} = \int_{\mathbb{R}} x^j \, dF_X(x),$$

while the $j$-th sample moment is defined by

$$\widehat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j,$$

where $X_1, X_2, \cdots, X_n$ are observations of $X$.

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# Moments Estimator

The method of moments estimator $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k)$ is defined to be the value of $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ such that

$$\begin{array}{rcl} \widehat{\alpha}_1 & = & \alpha_1(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k) \\ \widehat{\alpha}_2 & = & \alpha_2(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k) \\ \vdots & \vdots & \vdots \\ \widehat{\alpha}_k & = & \alpha_k(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k). \end{array}$$

*These estimators are very simple and consistent (under very weak assumptions), but they are often biased.*

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# An example of moments estimator

$X_1, X_2, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

- random variables first and second moment

$$\alpha_1(\mu, \sigma^2) = \mu \quad \text{and} \quad \alpha_2(\mu, \sigma^2) = \mu^2 + \sigma^2$$

- method of moments estimator

$$\widehat{\alpha}_1 = \alpha_1(\widehat{\mu}, \widehat{\sigma}^2) \quad \text{and} \quad \widehat{\alpha}_2 = \alpha_2(\widehat{\mu}, \widehat{\sigma}^2)$$

$$\Longleftrightarrow$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \widehat{\mu} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \widehat{\mu}^2 + \widehat{\sigma}^2$$

- parameters estimators

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \left( X_i^2 - \widehat{\mu} \right)$$

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# Likelihood Function

$X_1, X_2, \cdots, X_n$ are independent observations of $X$

The likelihood function is defined by

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} p_X(X_i, \theta).$$

The log-likelihood function is defined by

$$\ell_n(\theta) = \log \mathcal{L}_n(\theta) = \sum_{i=1}^{n} \log p_X(X_i, \theta).$$

The likelihood function is the joint density of the data, except it is treated as a function of the parameter $\theta$.

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# Maximum Likelihood Estimator

The maximum likelihood estimator, denoted by $\widehat{\theta}$, is the parameter vector $\theta$ that maximizes likelihood function $\mathcal{L}_n(\theta)$.

The estimatior $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k)$ is obtained from the solution of

$$
\begin{aligned}
\frac{\partial \mathcal{L}_n}{\partial \theta_1}(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k) &= 0 \\
\frac{\partial \mathcal{L}_n}{\partial \theta_2}(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k) &= 0 \\
\vdots \qquad & \vdots \quad \vdots \\
\frac{\partial \mathcal{L}_n}{\partial \theta_k}(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k) &= 0.
\end{aligned}
$$

MLE is consistent and has the smallest (asymptotically) variance.

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# An example on maximum likelihood estimatation

$$X_1, X_2, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

Likelihood function:

$$\begin{aligned}
\mathcal{L}_n(\mu, \sigma) &= K \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left(-\frac{1}{2\,\sigma^2}(X_i - \mu)^2\right), \\
&= K\,\sigma^{-n} \exp\left(-\frac{1}{2\,\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2\right) \\
&= K\,\sigma^{-n} \exp\left(-\frac{n\,S^2}{2\,\sigma^2}\right) \exp\left(-\frac{n(\overline{X} - \mu)^2}{2\,\sigma^2}\right)
\end{aligned}$$

$$K = \left(\sqrt{2\,\pi}\right)^{-n}, \quad \overline{X} = n^{-1} \sum_{i=1}^{n} X_i \;\; \text{and} \;\; S^2 = n^{-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# An example on maximum likelihood estimatation

Log-likelihood function:

$$\begin{aligned} \ell_n(\mu,\sigma) &= \log\left\{ K\,\sigma^{-n} \exp\left(-\frac{n\,S^2}{2\,\sigma^2}\right) \exp\left(-\frac{n(\overline{X}-\mu)^2}{2\,\sigma^2}\right) \right\} \\ &= \log K - n\,log\sigma - \frac{n\,S^2}{2\,\sigma^2} - \frac{n(\overline{X}-\mu)^2}{2\,\sigma^2} \end{aligned}$$

(log-likelihood or likelihood leads to the same estimator)

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# An example on maximum likelihood estimatation

Maximum log-likelihood estimator:

$$\frac{\partial \ell_n}{\partial \mu}\left(\widehat{\mu}, \widehat{\sigma}\right) = 0 \quad \text{and} \quad \frac{\partial \ell_n}{\partial \sigma}\left(\widehat{\mu}, \widehat{\sigma}\right) = 0$$

$$\Longleftrightarrow$$

$$\frac{n(\overline{X} - \widehat{\mu})}{\widehat{\sigma}^2} = 0 \quad \text{and} \quad -\frac{n}{\widehat{\sigma}} + \frac{n\, S^2}{\widehat{\sigma}^3} + \frac{n(\overline{X} - \widehat{\mu})^2}{\widehat{\sigma}^3} = 0$$

Parameters estimators:

$$\widehat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \widehat{\sigma} = S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - \widehat{\mu}\right)}$$

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

# Computing Maximum Likelihood Estimates

In general MLE estimator is not known analytically.

Log-likelihood expansion around $\theta^j$ gives

$$0 = \ell_n'(\theta) \approx \ell_n'(\theta^j) + (\theta - \theta^j)\,\ell_n''(\theta^j)$$

which provides

$$\theta \approx \theta^j - \frac{\ell_n'(\theta^j)}{\ell_n''(\theta^j)}, \;\; \ell_n''(\theta^j) \neq 0$$

**Newton method for MLE estimation:**

$$\widehat{\theta}^{j+1} = \widehat{\theta}^j - \frac{\ell_n'(\widehat{\theta}^j)}{\ell_n''(\widehat{\theta}^j)}$$

$\widehat{\theta}^0$ defined by moments estimator

L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.

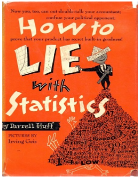
# Final Remarks on Statistics

# Statistical Software

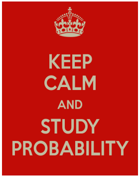
- R (programming language)
  https://www.r-project.org
- Ox (programming language)
  www.oxmetrics.net
- SciPy (Phython library)
  https://www.scipy.org
- GNU Octave
  https://www.gnu.org/software/octave
- Scilab
  http://www.scilab.org
- MATLAB
  https://www.mathworks.com

# Be careful with statistics!


©

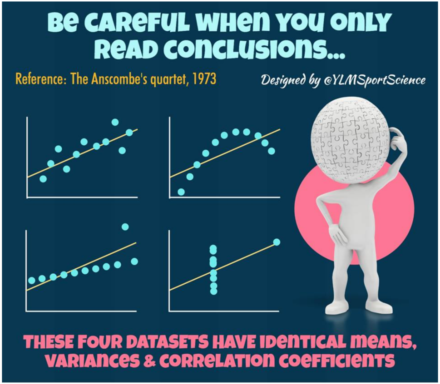



©




©

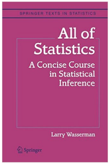

©

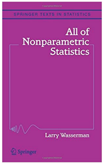


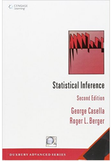

©

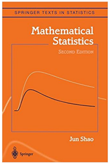
# References

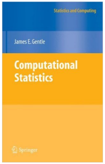

- S. E. Fienberg, *What Is Statistics?*, **Annual Review of Statistics and Its Application**, 1:1–9, 2014.
- V. Stodden, *Reproducing Statistical Results*, **Annual Review of Statistics and Its Application**, 2:1–19, 2015.
- G. Claeskens, *Statistical Model Choice*, **Annual Review of Statistics and Its Application**, 3:233–256, 2016.
- L. Wasserman, **All of Statistics: A Concise Course in Statistical Inference**, Springer, 2004.
- L. Wasserman, **All of Nonparametric Statistics**, Springer, 2007.
- G. Casella, **Statistical Inference**, Thomson Press (India) Ltd, 2008.
- J. Shao, **Mathematical Statistics**, Springer, 2nd Edition, 2007
- J. E. Gentle, **Computational Statistics**, Springer; 2009
- J. E. Gentle, W. K. Härdle, and Y. Mori, **Handbook of Computational Statistics: Concepts and Methods**, Springer, 2nd revised and updated Edition, 2012.

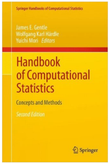



## How to cite this material?

A. Cunha Jr, *Elements of Statistics (Part II)*, Rio de Janeiro State University – UERJ, 2021.

@AmericoCunhaJr @AmericoCunhaJr

@AmericoCunhaJr @AmericoCunhaJr

# © copyright material



- 76% of the statistics are made up:
  https://thecooleststuffever.com/76-statistics-are-made-t-shirt
- How to lie with statistics:
  Wikimedia Commons, File:How to Lie with Statistics.jpg — Wikimedia Commons, the free media repository https://en.wikipedia.org/wiki/File:How_to_Lie_with_Statistics.jpg
- Keep calm and study probability:
  https://keepcalms.com/p/keep-calm-and-study-probability/
- Keep calm and study statistics:
  https://www.keepcalmandposters.com/poster/791012_keep_calm_and_study_statistics
- Be careful when you only read conclusions:
  https://ylmsportscience.com/2016/07/30/stats-be-careful-when-you-only-read-conclusions-by-ylmsportscience/