భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Biostatistics
BT2023

Lecture 12
Statistical significance : test of hypothesis

Himanshu Joshi
15 September 2023

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Advances in Complex Systems
Vol. 21, Nos. 3 & 4 (2018) 1850014 (31 pages)

DOI: 10.1142/S0219525918500145

World Scientific
www.worldscientific.com

TALENT VERSUS LUCK: THE ROLE OF RANDOMNESS IN SUCCESS AND FAILURE

ALESSANDRO PLUCHINO

Department of Physics and Astronomy
University of Catania
Via S.Sofia 64, Catania 95123, Italy
INFN-CT, Via S. Sofia 64, Catania 95123, Italy
alessandro.pluchino@ct.infn.it

Variance: Biased vs Unbiased

$$S^2 = \frac{1}{N}\sum(X_i - \bar{X})^2$$ Biased estimator

$$\sigma^2 = \frac{1}{N}\sum(X_i - \mu)^2$$ Unbiased estimator

$$S^2 = \frac{1}{N-1}\sum(X_i - \bar{X})^2$$ Unbiased estimator

$$\bar{X} \ is \ estimator \ of \ \mu$$

$$E[S^2] \neq \sigma^2$$

$$S^2 \ is \ an \ estimator \ of \ \sigma^2$$

$$But \quad E[\frac{N}{N-1}S^2] = \sigma^2$$

$$\mathbf{E}\left[(\overline{X} - \mu)^2\right] = \frac{1}{n}\sigma^2.$$

Bias of an estimator

$$\begin{aligned}
\mathrm{E}[S^2] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left((X_i-\mu) - (\overline{X}-\mu)\right)^2\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left((X_i-\mu)^2 - 2(\overline{X}-\mu)(X_i-\mu) + (\overline{X}-\mu)^2\right)\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n}(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + \frac{1}{n}(\overline{X}-\mu)^2\sum_{i=1}^{n}1\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n}(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + \frac{1}{n}(\overline{X}-\mu)^2\cdot n\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n}(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + (\overline{X}-\mu)^2\right]
\end{aligned}$$

$$\mathrm{E}\left[(\overline{X}-\mu)^2\right] = \frac{1}{n}\sigma^2.$$

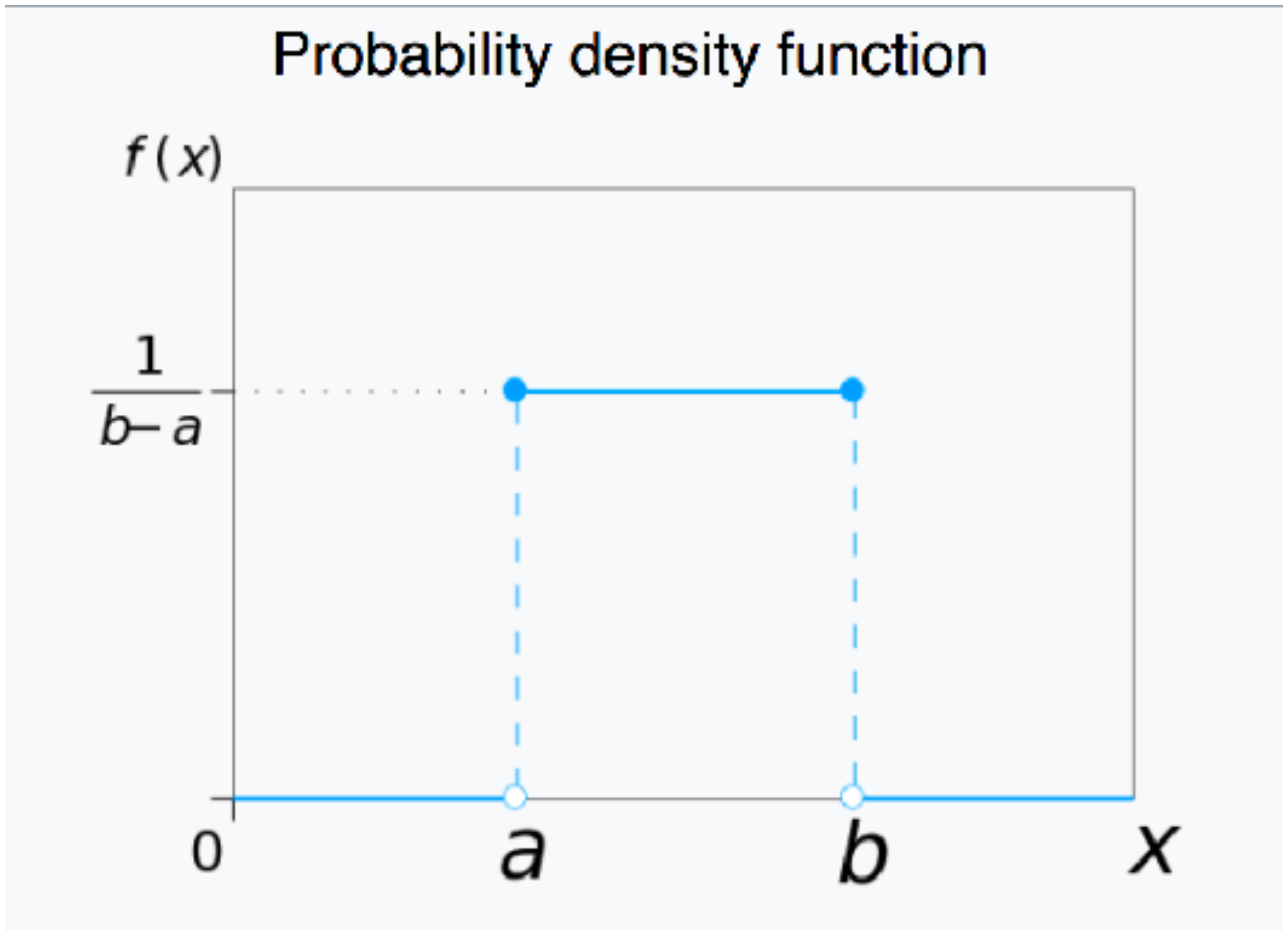
$$\overline{X}-\mu = \frac{1}{n}\sum_{i=1}^{n}X_i - \mu = \frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu).$$

$$\begin{aligned}
\mathrm{E}[S^2] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n}(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + (\overline{X}-\mu)^2\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n}(\overline{X}-\mu)\cdot n\cdot(\overline{X}-\mu) + (\overline{X}-\mu)^2\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - 2(\overline{X}-\mu)^2 + (\overline{X}-\mu)^2\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\overline{X}-\mu)^2\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right] - \mathrm{E}\left[(\overline{X}-\mu)^2\right] \\
&= \sigma^2 - \mathrm{E}\left[(\overline{X}-\mu)^2\right] = \left(1-\frac{1}{n}\right)\sigma^2 < \sigma^2.
\end{aligned}$$

Source: https://en.wikipedia.org/wiki/Bias_of_an_estimator

Probability distributions important to modelling in the life and social sciences

Uniform probability distribution



Probability Distribution Function

$$p(x,k) = \frac{x^k e^{-x}}{k!}$$

$$ke^{-kx}$$

$$f(k;n,p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, ..., n$, where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The null hypothesis

A test of significance

$$H_0 : \mu_1 = \mu_2$$

There no statistically difference between two samples

The alternative hypothesis

$$H_0 : \mu_1 \neq \mu_2$$

Example : Newton's laws of motion

Einstein's theory of gravity

Student's t-Test

Introduced in William Sealy Gosset devised this test to find that quality of barley field depending on the number of sample points

When the sample is small n<30, how do find that the statistics is significant or not

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Next Class

2:30 PM Friday, 22 September 2023