# 18.650 – Fundamentals of Statistics

## 1. Introduction and probability

# Goals

Goals:

- To give you a solid introduction to the mathematical theory behind statistical methods;
- To provide theoretical guarantees for the statistical methods that you may use for certain applications.

At the end of this class, you will be able to

1. From a real-life situation, formulate a statistical problem in mathematical terms
2. Select appropriate statistical methods for your problem
3. Understand the implications and limitations of various methods

# Why statistics?

# In the press

The New York Times

THE UPSHOT

## Nike Says Its $250 Running Shoes Will Make You Run Much Faster. What if That's Actually True?

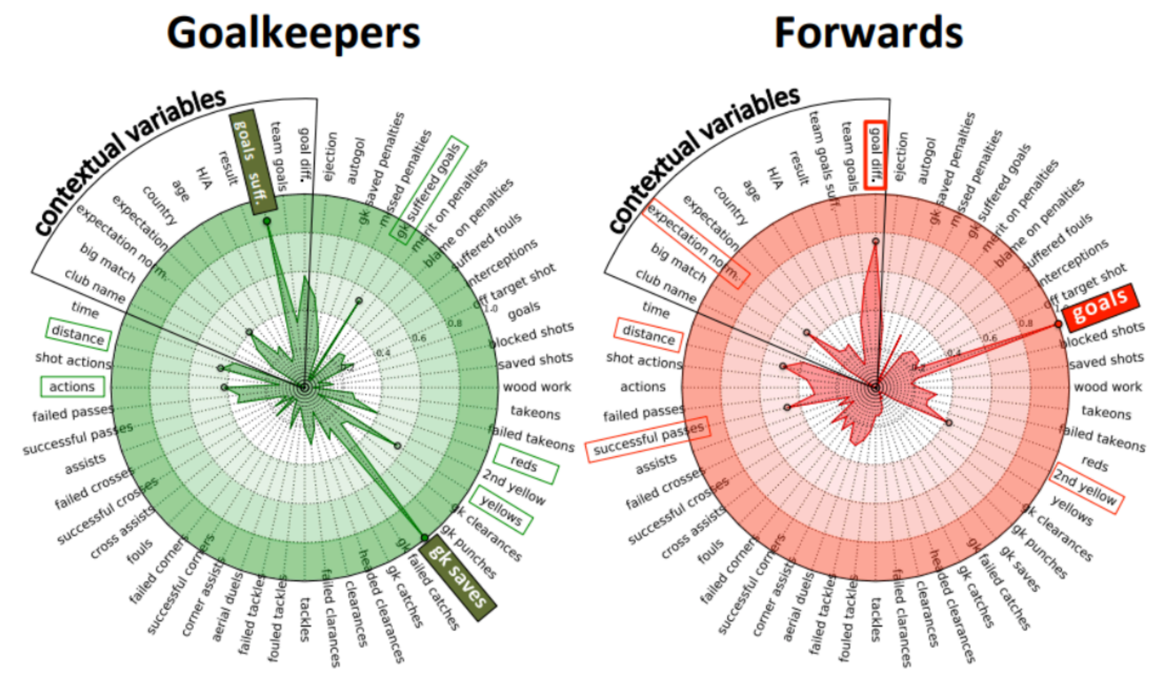
An analysis of nearly 500,000 running times estimates the effect of shoes on race performance.

https:/www.nytimes.com:interactive:2018:07:18:upshot:nike-vaporfly-shoe-strava.html
Citation/Attribution: Article (c) New York Times

MIT Technology Review

## Data Mining Reveals the Way Humans Evaluate Each Other

Vast databases of soccer statistics expose the limited way human observers rate performance and suggest how they can do significantly better.




Object Source/ URL: https://www.technologyreview.com/s/609760/data-mining-reveals-the-way-humans-evaluate-each-other/
Citation/Attribution -- Article from the MIT Technology Review. (c) MIT

# In businesses

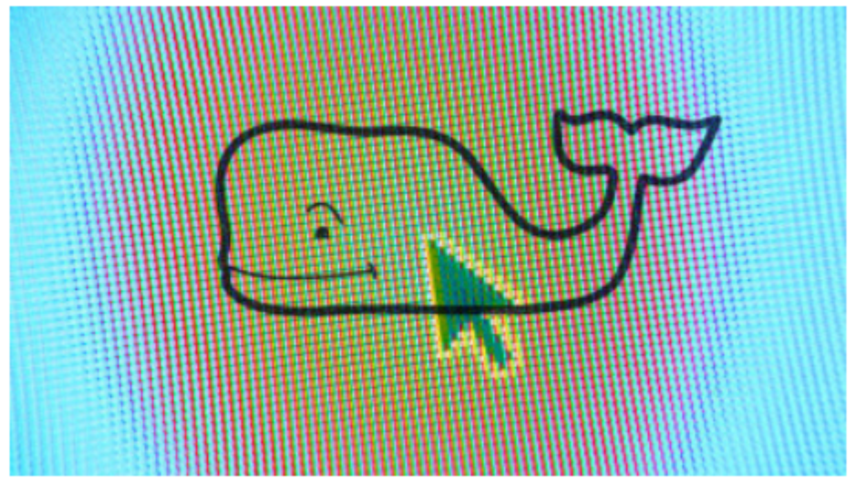
Harvard Business Review

## How Vineyard Vines Uses Analytics to Win Over Customers

TECHNOLOGY DIGITAL ARTICLE by Dave Sutton

A case study on how personalization is changing retail.

SAVE SHARE JUNE 08, 2018



tps://hbr.org/2018/06/how-vineyard-vines-uses-analytics-to-win-over-customers


FAST COMPANY

## AppNexus is key to AT&T's plans to use HBO for more consumer data

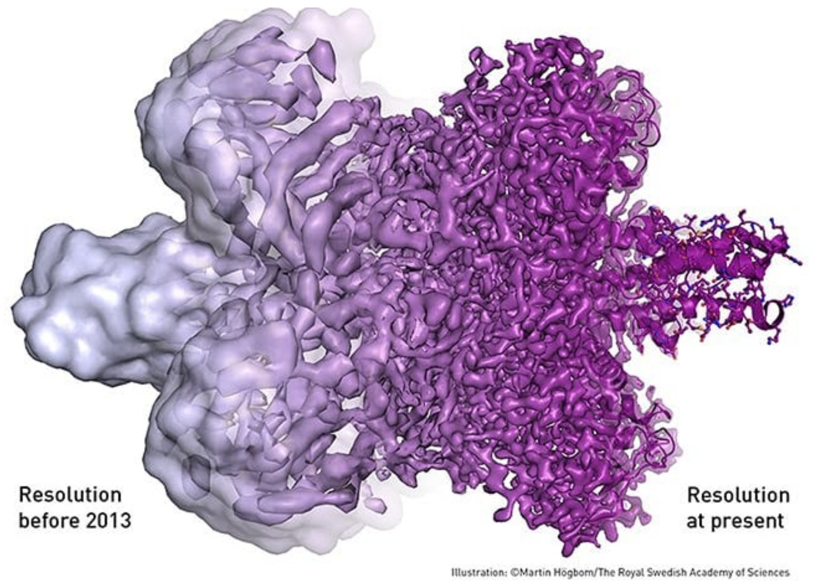
New WarnerMedia CEO John Stankey says HBO is going to "change direction a little bit," and it's all about the advertising.

ttps://www.fastcompany.com/90188017/appnexus-is-key-to-atts-plans-to-use-hbo-for-more-consumer-data


# In science and engineering

The Guardian

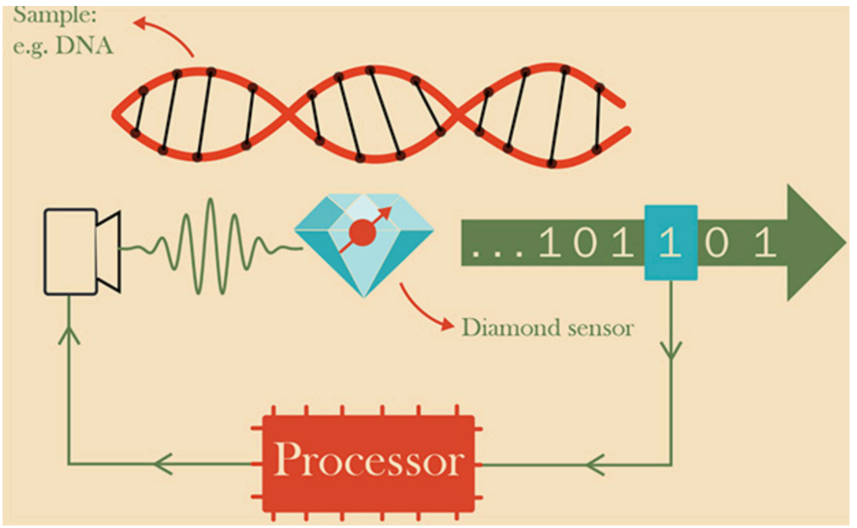
## What is cryo-electron microscopy, the Nobel prize-winning technique?




Object Source / URL*
https://www.theguardian.com/science/2017/oct/04/what-is-cryo-electron-microscopy-the-chemistry-nobel-prize-winning-technique
Citation/Attribution -- Image (c) The Guardian.

IEEE SPECTRUM

## Measuring Tiny Magnetic Fields With an Intelligent Quantum Sensor




https://spectrum.ieee.org/tech-talk/biomedical/devices/measuring-tiny-magnetic-fields-with-an-intelligent-quantum-sensor
Citation/Attribution -- Article Image (c) International Journal of Electrical, Electronics and Data Communication.

# On TV

LAST WEEK TONIGHT WITH JOHN OLIVER

"Last Week Tonight with John Oliver": Scientific Studies




Data Science and the Art of Producing Entertainment at Netflix



# Statistics, Data Science . . . and all that

Statistics, Data Science, Machine Learning, Artificial Intelligence

**What's the difference?**

# Statistics, Data Science . . . and all that

Statistics, Data Science, Machine Learning, Artificial Intelligence

**What's the difference?**

- All use data to gather insight and ultimately make decisions
- Statistics is at the core of the data processing part
- Nowadays, computational aspects play an important role as data becomes larger

# Computational and statistical aspects of data science

- Computational view: data is a (large) sequence of numbers that needs to be processed by a relatively fast algorithm: approximate nearest neighbors, low dimensional embeddings, spectral methods, distributed optimization, etc.
- Statistical view: data comes from a **random process**. The goal is to learn how this process works in order to make predictions or to understand what plays a role in it.

To understand randomness, we need PROBABILITY.

# Probability

- Probability studies randomness (hence the prerequisite)
- Sometimes, the physical process is completely known: dice, cards, roulette, fair coins, ...

Rolling 1 die:

- Alice gets \$1 if # of dots $\leq 3$
- Bob gets \$2 if # of dots $\leq 2$

Who do you want to be: Alice or Bob?

$E[A] = \frac{1}{2} \cdot \$1 = \$.5$

$E[B] = \frac{1}{3} \cdot \$2 = \$.66$

Rolling 2 dice:

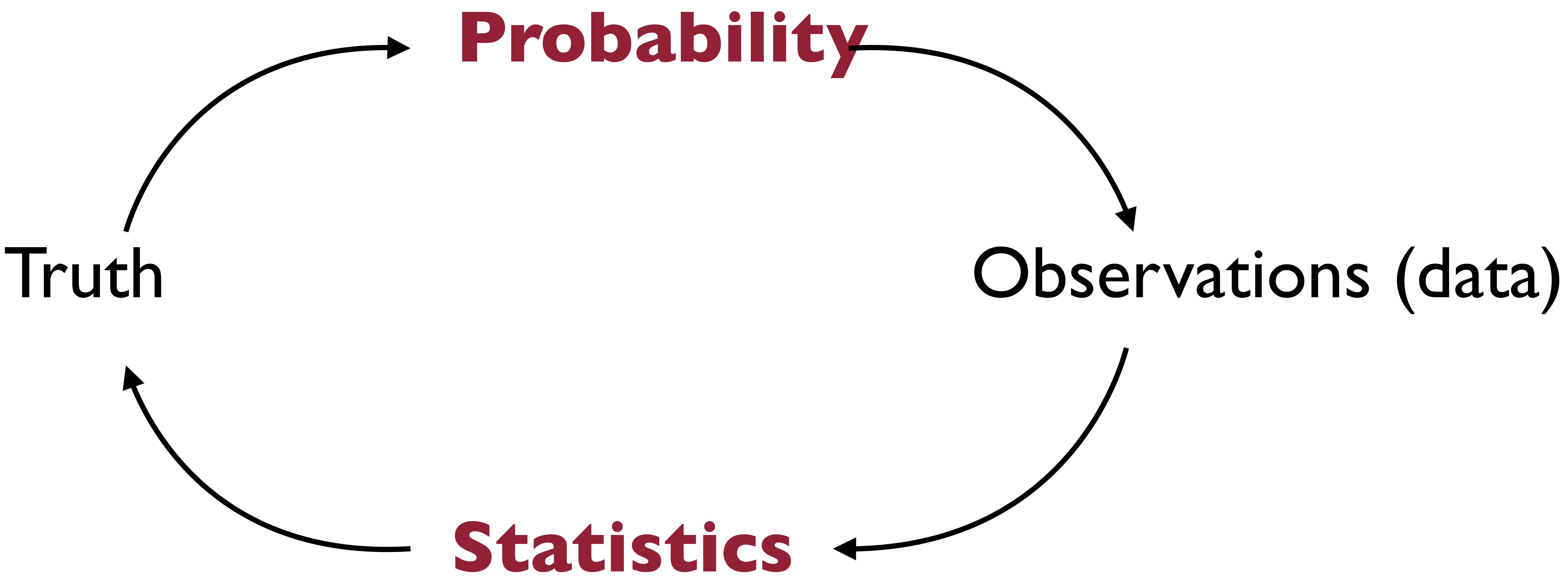
- Choose a number between 2 and 12
- Win \$100 if you chose the sum of the 2 dice

Which number do you choose?

# Statistics and modeling

- Dice are well known random process from physics: $1/6$ chance of each side (no need for data!), dice are independent. We can deduce the probability of outcomes, and expected \$ amounts. This is **probability**.
- How about more complicated processes? Need to estimate parameters from data. This is **statistics**
- Sometimes real randomness (random student, biased coin, measurement error, ...)
- Sometimes deterministic but too complex phenomenon: **statistical modeling**

  Complicated process "=" Simple process + random noise
- (good) Modeling consists in choosing (plausible) simple process **and** noise distribution.


Probability
Truth
Observations (data)
Statistics

# Statistics vs. probability

**Probability** Previous studies showed that the drug was 80% effective. Then we can anticipate that for a study on 100 patients, in average 80 will be cured and at least 65 will be cured with 99.99% chances.

**Statistics** Observe that 78/100 patients were cured. We (will be able to) conclude that we are 95% confident that for other studies the drug will be effective on between 69.88% and 86.11% of patients

# What this course is about

- Understand **mathematics** behind statistical methods
- Justify quantitive statements given modeling assumptions
- Describe interesting mathematics arising in statistics
- Provide a math toolbox to extend to other models.

What this course is **not** about

- Statistical thinking/modeling (e.g., 15.075)
- Implementation (e.g. IDS.012)
- Laundry list of methods (e.g. AP stats)

## What this course is about

- Understand **mathematics** behind statistical methods
- Justify quantitive statements given modeling assumptions
- Describe interesting mathematics arising in statistics
- Provide a math toolbox to extend to other models.

What this course is **not** about

- Statistical thinking/modeling (e.g., 15.075)
- Implementation (e.g. IDS.012)
- Laundry list of methods (e.g. AP stats)

# Let's do some statistics

# The kiss

Le baiser. Auguste Rodin. 1882.

Object Source / URL*
http://www.musee-rodin.fr/en/collections/sculptures/kiss Citation/Attribution Photo (c) Musée Rodin

# The kiss

Le baiser. Auguste Rodin. 1882.

Object Source / URL*
http://www.musee-rodin.fr/en/collections/sculptures/kiss Citation/Attribution Photo (c) Musée Rodin

# The kiss


nature International weekly journal of science

Full text access provided to **Massachusetts Institute of Technology** by **the MIT Libraries**

Cart

Search | go | Advanced search

Journal home > Archive > Brief Communications > Full Text

**Journal content**
- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive
- Supplements

## Brief Communications

*Nature* **421**, 711 (13 February 2003) | doi:10.1038/421711a

### Human behaviour: Adult persistence of head-turning asymmetry

Onur Güntürkün

**A neonatal right-side preference makes a surprising romantic reappearance later in life.**

Top

subscribe to nature

**FULL TEXT**
- Previous | Next
- Table of contents
- Download PDF

Object Source / URL*
https://www.nature.com/articles/421711a Citation/Attribution – SpringerNature.com

## Statistical experiment

*"A neonatal right-side preference makes a surprising romantic reappearance later in life."*

- Let $p$ denote the proportion of couples that turn their head to the right when kissing.
- Let us design a statistical experiment and analyze its outcome.
- Observe $n$ kissing couples times and collect the value of each outcome (say 1 for RIGHT and 0 for LEFT);
- *Estimate* $p$ with the proportion $\hat{p}$ of RIGHT.
- Study: "Human behaviour: Adult persistence of head-turning asymmetry" (Nature, 2003): $n = 124$ and 80 to the right so

$$\hat{p} = \frac{80}{124} = 64.5\,\%$$

## Random intuition

Back to the data:

- 64.5% is much larger than 50% so there seems to be a preference for turning right.
- What if our data was RIGHT, RIGHT, LEFT $(n = 3)$. That's 66.7% to the right. Even better?
- Intuitively, we need a large enough sample size $n$ to make a call. How large?
- Another way to put the problem: for $n = 124$, what is the minimum number of couple "to the right" would you need to see to be convinced that $p > 50\%$? 63? 72? 75? 80?

We need **mathematical modeling** to understand the accuracy of this procedure?

# A first estimator

Formally, this procedure consists of doing the following:

- For $i = 1, \dots, n$, define $R_i = 1$ if the $i$th couple turns to the right RIGHT, $R_i = 0$ otherwise.
- The estimator of $\bar{p}$ is the Sample average

$$\sum_{i=1} R_i = \#\{i : R_i = 1\}$$

$$\hat{p} = \bar{R}_n = \frac{1}{n}\sum_{i=1}^{n} R_i.$$

**What is the accuracy of this estimator ?**

In order to answer this question, we propose a statistical model that describes/approximates well the experiment.

We think of the $R_i$'s as random variables so that $\hat{p}$ is also a random variable. We need to understand its fluctuation.

# Modelling assumptions

Coming up with a model consists of making assumptions on the observations $R_i, i = 1, \ldots, n$ in order to draw statistical conclusions. Here are the assumptions we make:

1. Each $R_i$ is a random variable.
2. Each of the r.v. $R_i$ is Bernoulli with parameter $p$.
3. $R_1, \ldots, R_n$ are mutually independent.

$R_i \sim \mathrm{Ber}(p)$

$\mathbb{P}(R_i = 1) = p$

$\mathbb{P}(R_i = 0) = 1 - p$

# Discussion

Let us discuss these assumptions.

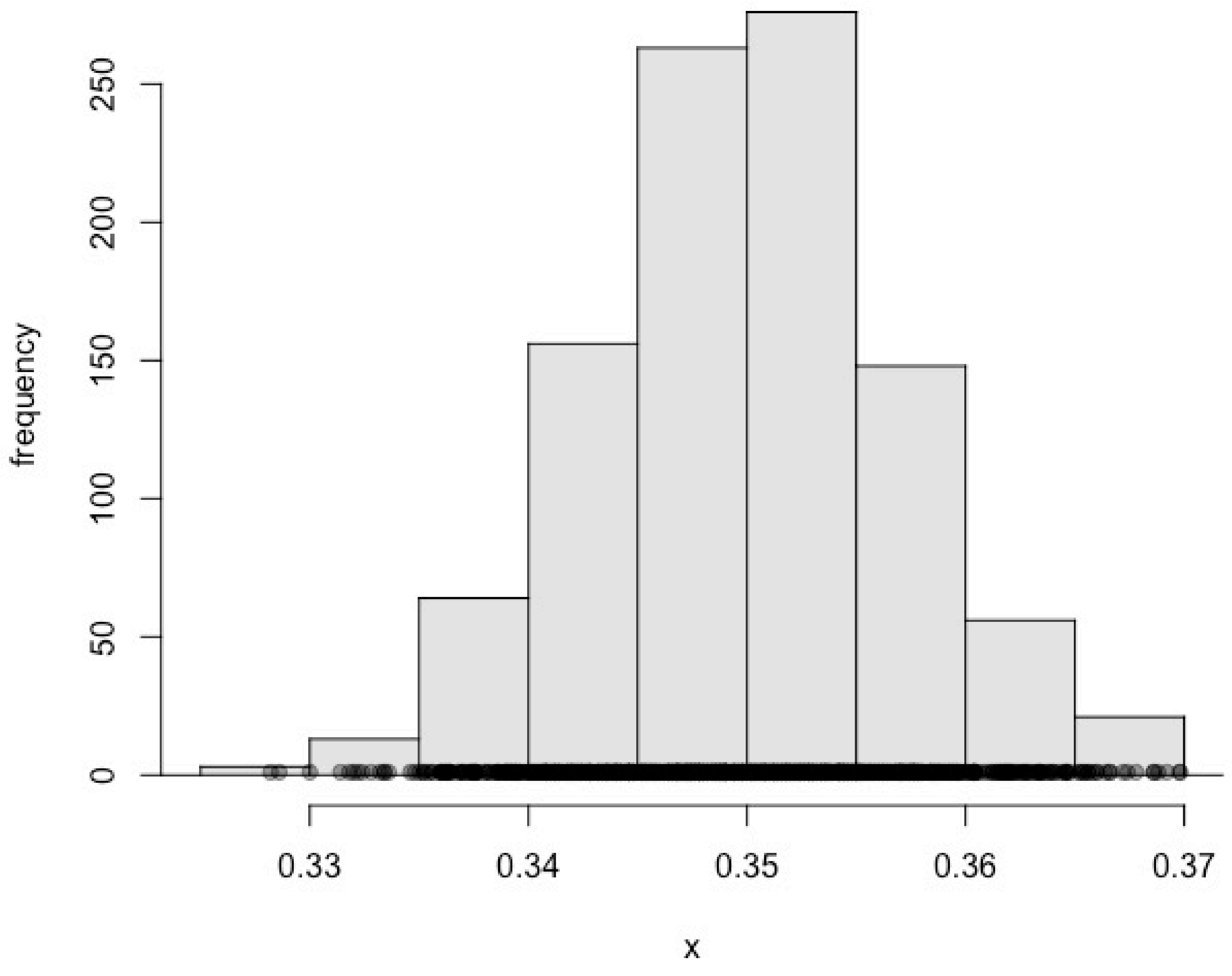
1. Randomness is a way of modeling lack of information; with perfect information about the conditions of kissing (including what goes on in the kissers' mind), physics or sociology would allow us to predict the outcome.

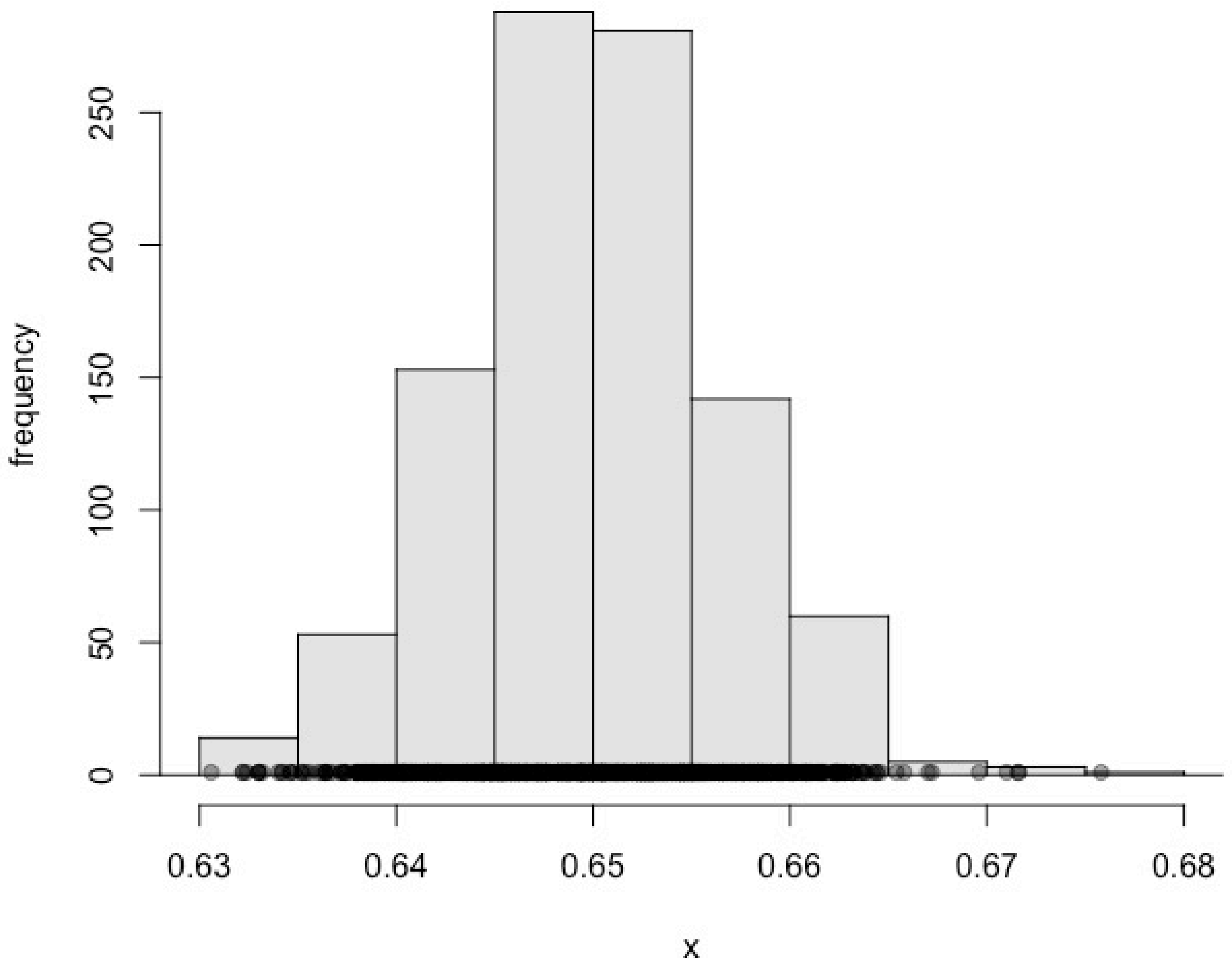
2. Hence, the $R_i$'s are necessarily Bernoulli r.v. since $R_i \in \{0, 1\}$. They could still have a different parameter $R_i \sim \text{Ber}(p_i)$ for each couple but we don't have enough information with the data to estimate the $p_i$'s accurately. So we simply assume that our observations come from the same process: $p_i = p$ for all $i$

3. Independence is reasonable (people were observed at different locations and different times).

# Population vs. Samples

- Assume that there is a total **population** of 5,000 "airport-kissing" couples
- Assume for the sake of argument that $p = 35\%$ or that $p = 65\%$.
- What do **samples** of size 124 look like in each case?



# Why probability?

We need to understand probabilistic aspects of the distribution of the random variable:

$p \to \hat{p}$

$$\hat{p} = \bar{R}_n = \frac{1}{n}\sum_{i=1}^{n} R_i.$$

Specifically, we need to be able to answer questions such as:

- Is the expected value of $\hat{p}$ close to the unknown $p$?
- Does $\hat{p}$ take values close to $p$ with high probability?
- Is the variance of $\hat{p}$ large? I.e. does $\hat{p}$ fluctuate a lot?

We need probabilistic tools! Most of them are about **average of independent random variables**.

$\mathrm{Var}(\bar{R}_n) = ?$

$\mathbb{P}(|\bar{R}_n - p| > 0.1) = ?$

# Probability redux

# Averages of random variables: LLN & CLT

Let $X, X_1, X_2, \ldots, X_n$ be i.i.d. r.v., $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathbb{V}[X]$.

- Laws (weak and strong) of large numbers (LLN):

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P},\ \text{a.s.}} \mu.$$

- Central limit theorem (CLT):

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1).$$

Rule of thumb:

$n \geq 30$

(Equivalently, $\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,\sigma^2)$.)

# Another useful tool: Hoeffding's inequality

What if $n$ is not large enough to apply CLT?

**Theorem** (Hoeffding, 1963)
Let $n$ be a positive integer and $X, X_1, \ldots, X_n$ be i.i.d. r.v. such that $\mu = \mathbb{E}[X]$ and

$X \in [a, b]$ almost surely $\qquad$ ($a < b$ are given numbers)

Then,

$$\mathbb{P}[|\bar{X}_n - \mu| \geq \varepsilon] \leq 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}. \quad \forall \varepsilon > 0$$

This holds even for small sample sizes $n$.

$X_i \overset{iid}{\sim} \mathrm{Ber}(p)$ $\qquad$ $\mathbb{P}\left(|\bar{X}_n - \mu| \geq \frac{c}{\sqrt{n}}\right) \leq 2e^{-2c^2}$

## Consequences

- The LLN's tell us that

$$\bar{R}_n \xrightarrow[n\to\infty]{\mathbb{P},\ \text{a.s.}} p.$$

$X_i \sim Ber(p_i)$ $p_i = p$ $\forall i$

i.i.d

indep.

(what modeling assumptions did we use?)

- Hence, when the size $n$ of the experiment becomes large, $\bar{R}_n$ is a *good* (say *"consistent"*) estimator of $p$.

- The CLT refines this by quantifying *how good* this estimate is: for $n$ large enough the distribution of $\hat{p}$ is almost:

$$\mathbb{P}(|\bar{R}_n - p| \geq \varepsilon) \simeq \mathbb{P}(|\mathcal{N}(0, \frac{p(1-p)}{n})| > \varepsilon)$$

$n \geq 30$

In the Kiss example, $\mathbb{P}(|\bar{R}_n - p| \geq 0.084) \simeq 5\%$

- Hoeffding's inequality tells us that

$$\mathbb{P}(|\bar{R}_n - p| \geq 0.084) \leq 2\exp\left(-\frac{2 \cdot 124 \cdot (0.084)^2}{(1-0)^2}\right) \leq 0.35$$

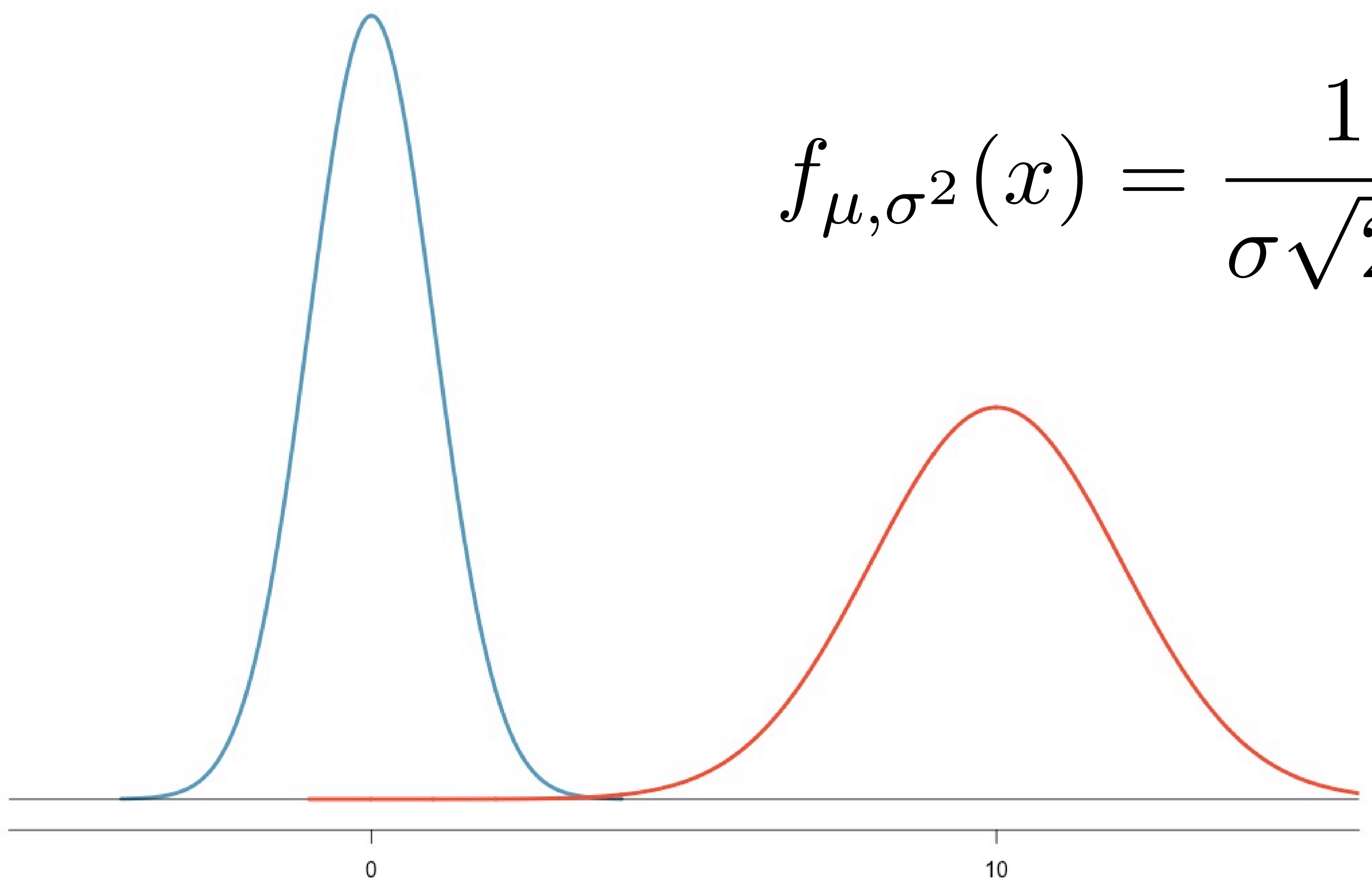
# The Gaussian distribution

Because of the CLT, the Gaussian (a.k.a normal) distribution is ubiquitous in statistics. It is named after German Mathematician Carl Friedrich Gauss (1777–1855) in the context of the method of *least squares* (regression).

- $X \sim \mathcal{N}(\mu, \sigma^2)$
- $\mathbb{E}[X] = \mu$
- $\text{var}(X) = \sigma^2 > 0$


Object Source / URL*
http://mathshistory.st-andrews.ac.uk/PictDisplay/Gauss.html Citation/Attribution Image from the MacTutor History of Mathematics archive (success)

# Gaussian density (pdf)




$$f_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Figure 1: Two pdfs: $\mathcal{N}(0,1)$ and $\mathcal{N}(10,4)$

- Tails decay very fast (like $e^{-\frac{x^2}{2\sigma^2}}$): almost in finite interval.
- There is no closed form for their cumulative distribution function (CDF). We use tables (or computers):

$$\mathbb{P}(X \leq x) = F_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

# Some useful properties of Gaussians

Perhaps the most useful property of the Gaussian family is that it's *invariant under affine transformation*:

- $X \sim \mathcal{N}(\mu, \sigma^2)$, then for any $a, b \in \mathbb{R}$,

$$a \cdot X + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

- **Standardization** (a.k.a Normalization/Z-score): If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Useful to compute probabilities from CDF of $Z \sim \mathcal{N}(0, 1)$:

$$\mathbb{P}(u \leq X \leq v) = \mathbb{P}\left(\frac{u - \mu}{\sigma} \leq Z \leq \frac{v - \mu}{\sigma}\right)$$

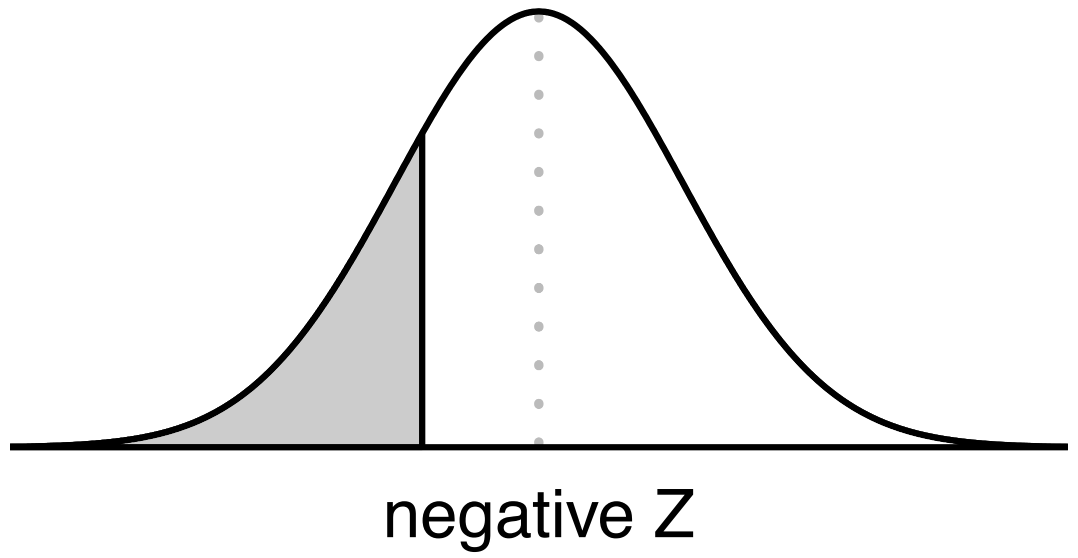
- **symmetry**: If $X \sim \mathcal{N}(\mathbf{0}, \sigma^2)$ then $-X \sim \mathcal{N}(\mathbf{0}, \sigma^2)$: If $x > 0$

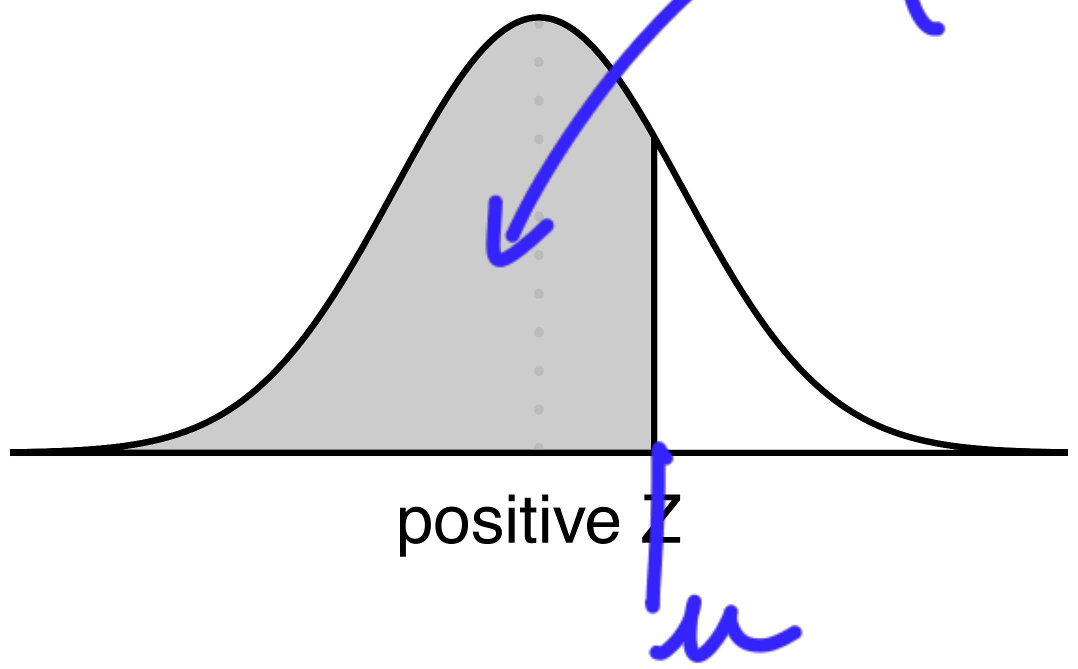
$$\mathbb{P}(|X| > x) = \mathbb{P}(X > x) + \mathbb{P}(-X > +x) = 2\mathbb{P}(X > x)$$

# Gaussian probability tables

$Z \sim N(0,1)$

$F(u) = P(Z \leq u)$

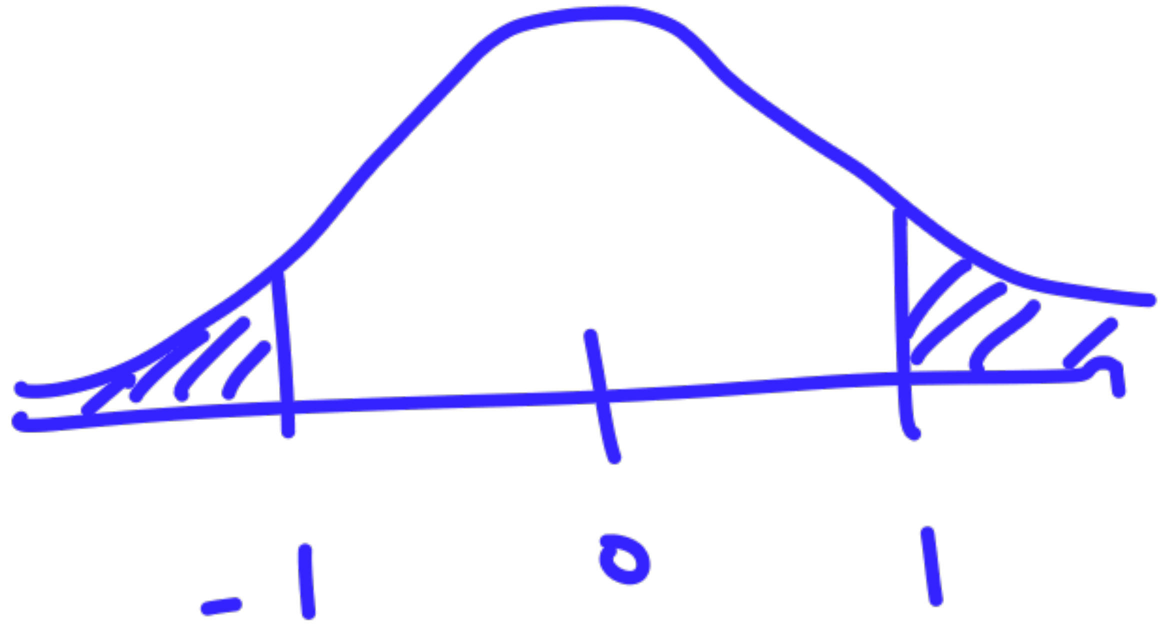



$P(Z \leq 0.76)$
$= 0.7764$

| $Z$ | Second decimal place of $Z$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

# Examples




Assume that $Z \sim \mathcal{N}(0,1)$ and compute

- $\mathbb{P}(Z \leq 1) = 0.8413$
- $\mathbb{P}(Z \geq -1) = \mathbb{P}(-Z \leq 1) = \mathbb{P}(Z \leq 1) = 0.8413$
- $\mathbb{P}(|Z| > 1) = 2\,\mathbb{P}(Z > 1) = 2 \cdot (1 - 0.8413) = 0.31$

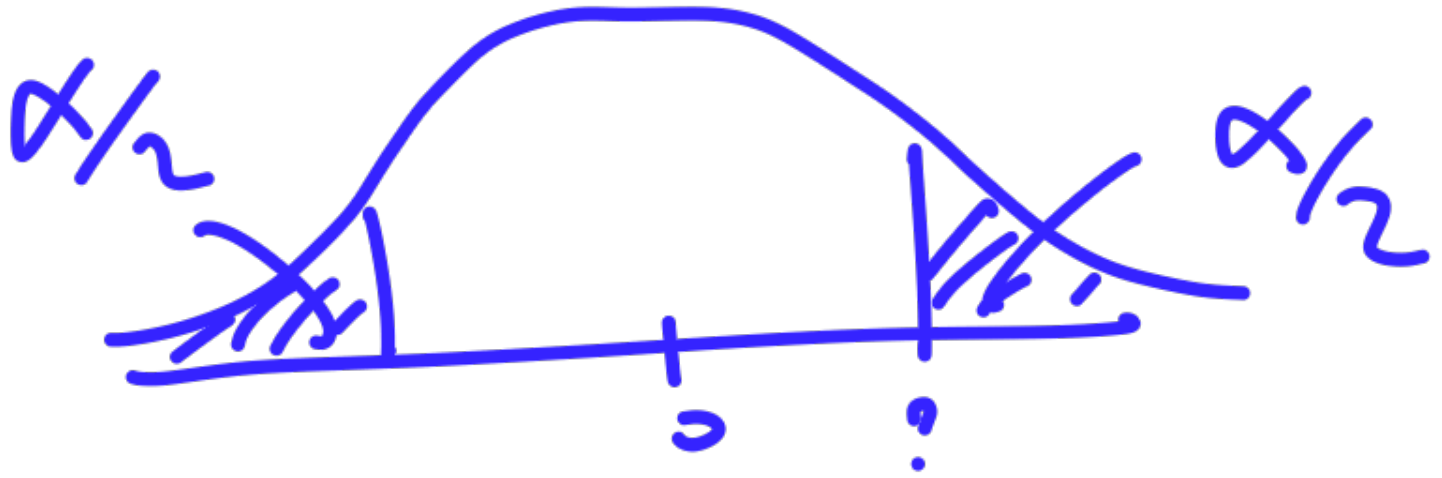
Assume that the score distribution for a final exam is approximately $X \sim \mathcal{N}(85, 4)$, compute

- $\mathbb{P}(X > 90) = \mathbb{P}\left(\frac{X-85}{2} > \frac{90-85}{2}\right) = \mathbb{P}(Z > 2.5) = 0.62\%$
- $\mathbb{P}(80 < X < 90)$

$= \mathbb{P}(-2.5 < Z < 2.5) = 0.9876$

More complicated: what is $x$ such that $\mathbb{P}(X < x) = 90\%$ ($90^{th}$ percentile?). For that we need to read the table backwards.

# Quantiles




## Definition

Let $\alpha$ in $(0,1)$. The quantile of order $1-\alpha$ of a random variable $X$ is the number $q_\alpha$ such that

$$\mathbb{P}(X \le q_\alpha) = 1 - \alpha$$

$\alpha = .1 \Rightarrow q_\alpha$ is the 90th percentile

Let $F$ denote the CDF of $X$:

- $F(q_\alpha) = 1-\alpha$
- If $F$ is invertible, then $q_\alpha = F^{-1}(1-\alpha)$
- $\mathbb{P}(X > q_\alpha) = \alpha$
- If $X = Z \sim \mathcal{N}(0,1)$: $\mathbb{P}(|X| > q_{\alpha/2}) = \alpha$

Some important quantiles of the $Z \sim \mathcal{N}(0,1)$ are:

| $\alpha$ | 2.5% | 5% | 10% |
|---|---|---|---|
| $q_\alpha$ | 1.96 | 1.65 | 1.28 |

We get that $\mathbb{P}(|Z| > 1.96) = 5\%$

# Three types of convergence

- $(T_n)_{n\geq 1}$ is a sequence of random variables
- $T$ is a random variable ($T$ may be deterministic).

- Almost surely (a.s.) convergence:

$$T_n \xrightarrow[n\to\infty]{\text{a.s.}} T \quad \text{iff} \quad \mathbb{P}\left[\left\{\omega : T_n(\omega) \xrightarrow[n\to\infty]{} T(\omega)\right\}\right] = 1.$$

- Convergence in probability:

$$T_n \xrightarrow[n\to\infty]{\mathbb{P}} T \quad \text{iff} \quad \mathbb{P}\left[|T_n - T| \geq \varepsilon\right] \xrightarrow[n\to\infty]{} 0, \quad \forall \varepsilon > 0.$$

- Convergence in distribution:

$$T_n \xrightarrow[n\to\infty]{(d)} T \quad \text{iff} \quad \mathbb{E}[f(T_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[f(T)]$$

for all continuous and bounded function $f$.

# Properties

- If $(T_n)_{n\geq 1}$ converges a.s., then it also converges in probability, and the two limits are equal a.s.
- If $(T_n)_{n\geq 1}$ converges in probability, then it also converges in distribution
- **Convergence in distribution** implies convergence of probabilities if the limit has a density (e.g. Gaussian):

$$T_n \xrightarrow[n\to\infty]{(d)} T \qquad \Rightarrow \qquad \mathbb{P}(a \leq T_n \leq b) \xrightarrow[n\to\infty]{} \mathbb{P}(a \leq T \leq b)$$

# Exercises

**a)** Is the following statement correct? "If $(T_n)_{n\geq 1}$ converges in probability, then it also converges a.s"

1. Yes
2. No

Let $\{X_1, X_2, \ldots, X_n\}$ be a sequence of r.v. such that $\cancel{X}_n \sim \text{Ber}(\frac{1}{n})$. Exercises b), c) and d) are about this sequence.

**b)**Let $\underline{0 < \epsilon < 1,}\ n \geq 1$. What is the value of $P(\{|X_n| > \epsilon\})$?
(answer: $\frac{1}{n}$)

**c)** Does $\{X_n\}$ converges in probability?

1. Yes
2. No

$0 < \varepsilon < 1$

$= \mathbb{P}(X_n > \varepsilon)$

$= \mathbb{P}(X_n = 1) = \frac{1}{n} \xrightarrow[n \to \infty]{} 0$

# Exercises

**d)** Denote by $X$ the limit of $\{X_n\}$ (if it exists) (that is, $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$). What is the value of $X$?

1. $X$ does not exist
2. 0
3. 1
4. None of the above

**e)** Dose $\{X_n\}$ converge in distribution?

1. Yes
2. No

**f)** What is the limit of the sequence $\mathbb{E}[\cos(X_n)]$ as $n$ tends to infinity?

$\mathbb{E}[\cos(0)] = 1$

# Addition, multiplication, division

...only for a.s. and $\mathbb{P}$...

Assume

$$T_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}} T \qquad \text{and} \qquad U_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}} U$$

Then,

- $T_n + U_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}} T + U$,
- $T_n U_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}} TU$,
- If in addition, $U \neq 0$ a.s., then $\dfrac{T_n}{U_n} \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}} \dfrac{T}{U}$.

In general, these rules **do not** apply to convergence $(d)$.

# Slutsky's theorem

Some partial results exist for convergence in distribution on the form of *Slutsky's theorem*.

Let $(X_n), (Y_n)$ be two sequences of r.v., such that:

$$\text{(i) } T_n \xrightarrow[n\to\infty]{(d)} T \qquad \text{and} \qquad \text{(ii) } U_n \xrightarrow[n\to\infty]{\mathbb{P}} u$$

where $T$ is a r.v. and $u$ is a given real number (deterministic limit: $\mathbb{P}(U = u) = 1$). Then,

- $T_n + U_n \xrightarrow[n\to\infty]{(d)} T + u$,
- $T_n U_n \xrightarrow[n\to\infty]{(d)} Tu$,
- If in addition, $u \neq 0$, then $\dfrac{T_n}{U_n} \xrightarrow[n\to\infty]{(d)} \dfrac{T}{u}$.

...

# Taking functions

Continuous functions (for all three types) . If $f$ is a continuous function:

$$T_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbb{P}/(d)} T \quad \Rightarrow \quad f(T_n) \xrightarrow[n\to\infty]{\text{a.s}/\mathbb{P}/(d)} f(T).$$

Continuous Mapping Theorem

**Example:** Recall that by LLN, $\bar{R}_n \xrightarrow[n\to\infty]{\mathbb{P},\text{ a.s.}} p$. Therefore

$$f(\bar{R}_n) \xrightarrow[n\to\infty]{\mathbb{P},\text{ a.s.}} f(p) \quad \text{for any continuous } f$$

(Only need $f$ to be continuous around $p$: f(x)=1/x works if $p > 0$)

We also have by CLT: $\sqrt{n}\frac{\bar{R}_n - p}{\sqrt{p(1-p)}} \xrightarrow[n\to\infty]{(d)} Z$, $Z \sim \mathcal{N}(0,1)$. So

$$f(\sqrt{n}(\bar{R}_n - p)) \xrightarrow[n\to\infty]{(d)} f(Y) \quad Y \sim \mathcal{N}(0, p(1-p))$$

⚠ not the limit of $\sqrt{n}[f(\bar{R}_n) - f(p)]$ !!

Delta ($\Delta$) method

# Recap

- Averages of random variables occur naturally in statistics
- We make modeling assumptions to apply probability results
- For large sample size they are *consistent* (LLN) and we know their distribution (CLT)
- CLT gives the (weakest) convergence in distribution but is enough to compute probabilities
- We use standardization and Gaussian tables to compute probabilities and quantiles
- We can make operations (addition, multiplication, continuous functions) on sequences of random variables

Unit 1 --18.6501x Attribution Credits

1.
Unit 1 Section 2: What is Statistics
Slide # -- page 4
Object Source / URL: file://localhost/--https/::www.nytimes.com:interactive:2018:07:18:upshot:nike-vaporfly-shoe-strava.html
Citation/Attribution: Article (c) New York Times

2.
Unit 1 Section 2: What is Statistics
Slide # page 4
Object Source/ URL: https://www.technologyreview.com/s/609760/data-mining-reveals-the-way-humans-evaluate-each-other/
Citation/Attribution -- Article from the MIT Technology Review. (c) MIT

3.
Unit 1 Section 2: What is Statistics
Slide # page 5
Object Source/ URL:
https://hbr.org/2018/06/how-vineyard-vines-uses-analytics-to-win-over-customers
Citation/Attribution -- Article and Image Copyright © 2019 Harvard Business School Publishing. All rights reserved.

4.
Unit 1 Section 2: What is Statistics
Slide # 5
Object Source / URL*
https://www.fastcompany.com/90188017/appnexus-is-key-to-atts-plans-to-use-hbo-for-more-consumer-data
Citation/Attribution -- Article by Jeff Beer on Fast Company & Inc © 2019 Mansueto Ventures, LLC

5.
Unit 1 Section 2: What is Statistics
Slide # 6
Object Source / URL*
https://www.theguardian.com/science/2017/oct/04/what-is-cryo-electron-microscopy-the-chemistry-nobel-prize-winning-technique
Citation/Attribution -- Image (c) The Guardian.

6.
Unit 1 Section 2: What is Statistics
Slide # 6
Object Source / URL*
https://spectrum.ieee.org/tech-talk/biomedical/devices/measuring-tiny-magnetic-fields-with-an-intelligent-quantum-sensor
Citation/Attribution -- Article Image (c) International Journal of Electrical, Electronics and Data Communication.

7.
Unit 1 Section 2: What is Statistics
Slide # 7
Object Source / URL*
https://www.youtube.com/watch?v=0Rnq1NpHdmw&has_verified=1
Citation/Attribution
Photo of John Oliver © 2019 Home Box Office, Inc. All Rights Reserved.

8.
Unit 1 Section 2: What is Statistics
Slide # 7
Object Source / URL*
https://medium.com/netflix-techblog/studio-production-data-science-646ee2cc21a1
Citation/Attribution
Image on the Medium website (c) Netflix corporation.

9.
Unit 1
Slide # 18, 19
Object Source / URL*
http://www.musee-rodin.fr/en/collections/sculptures/kiss
Citation/Attribution
Photo (c) Musée Rodin.

10.
Unit 1
Slide # 20
Object Source / URL*
https://www.nature.com/articles/421711a
Citation/Attribution – SpringerNature.com

11.
Unit 1
Slide # 23
Object Source / URL*
http://mathshistory.st-andrews.ac.uk/PictDisplay/Gauss.html
Citation/Attribution
Image from the MacTutor History of Mathematics archive
(success)