# Data Representations and Some Concepts of Probability (Review) 

Volker Tresp
Winter 2023-2024

## Preamble

- "Thermodynamik ist ein komisches Fach. Das erste Mal, wenn man sich damit befasst, versteht man nichts davon. Beim zweiten Durcharbeiten denkt man, man hÃđtte nun alles verstanden, mit Ausnahme von ein oder zwei kleinen Details. Das dritte Mal, wenn man den Stoff durcharbeitet, bemerkt man, dass man fast gar nichts davon versteht, aber man hat sich inzwischen so daran gewoehnt, dass es einen nicht mehr stoert." Arnold Sommerfeld


## Measurements and Statements

- Modern science started with the idea of an objective measurement: "This ball has a weight of 1 kg ".
- It proceeded by describing regularities: $F=m a$
- Darwin observed and measured beak shapes among finch species on different islands of the Galapagos. He postulated that the beak of an ancestral species had adapted over time to equip the finches to acquire different food sources
- Researchers tried to generalize this idea to psychology (behavorism), statistics (frequentist statistics), and other fields like econometrics 'If you can't measure it, you can't manage it" (Peter Drucker)"
- Steven Pinker: "It May Not Seem Like It, But The World Is Getting Better" (according to many measurable quantities like percentage of people dying in war, child mortality, ...)


## Beyond Objective Measurements

- Bayesian statistics ("personal belief"); Cognition ("Here is my theory of how the brain works"); Psychoanalytics; case studies
- Criticism of Positivism is philosophy
- We do not speak like Spock: Captain: "Spock: Today is such a wonderful day"; Spock: "I assume what you want to say is ..."
- "Iboprophin lowers fever" (imprecise summary)
- Language is imprecise and often expresses emotions or relates statements to the body (embodiment)
- Write an essay on: "Love is stronger than hate" (we need many words to explain the meaning of this simple sentence!)


## Beyond Objective Measurements (cont'd)

- "For the love of money is a root of all kinds of evils." (from 1 Tim 6:10)
- "Greed is good" (Gordon Gekko)
- The scientific fact is not that "The love of money is a root of all kinds of evils" or that "Greed is good"
- The fact is that that the first statement is written in the Bible and the second from Gordon Gekko in the 1987 film "Wall Street"
- It is absolutely fascinating that we (and GPT) can deal with these ambiguities quite well (most of the time)
- Based on facts (what kind of facts are those) and human understanding of how the world works (what is that) the butler is committed of murder and will spend 20 years in prison


## Outline

- We start with the scientific view, introduce the concept of a random variable and study the laws of probability
- Then we take a look at the Google Knowledge Graph which is supposed to be a fact graph: what kind of facts are there actually?


## Measurements in the Real World

- Our data matrix $\mathbf{D}$ (with entries $x_{i, j}, i=1, \ldots N, j=1, \ldots M$ ) has typically been generated from measurements in the real world
- It is full of numbers but these numbers mean something
- $x_{234,5}=36$ might mean that Jack (with $I D=234$ ) has age (with $I D=5$ ) 36
- As a ground fact: age(Jack, 36)
- "age" is the name of the column, in logics the predicate, in relational databases the attribute, in physics the measurement process
- "Jack" is the name of the row, in logics a constant (entity), in relational databases the key, in physics the object or time instance of measurement
- " 36 " is the table entry, in logics a constant, in physics the outcome of the measurement



## Logic:

-Ground atom (fact): age(Jack, 36) - age is the predicate and Jack and 36 are constants

## Probability:

- Jack is an outcome of the random experiment
- Age is the random variable
- 36 is a value (element of a set)
- Age(Jack) = 36

$$
\cdot P(\text { Age }=36)=0.07
$$

## Measurements in the Real World (cont'd)

- In relational databases, a row is also called a tuple or a record:

$$
\left(\operatorname{Jack}(I D=234), x_{234,1}, \ldots, x_{234, M}\right)
$$

- The data matrix, plus the column 0 for the IDs for the keys and plus names of the attributes (as row 0 )), in relational databases, is called a table or a relation
- For the data matrix we typically do not care about the names of the rows or columns (we do not include row 0 and column 0 )
- In statistics: "Age" is the random variable, and we are interested to find the probability that a randomly selected person has age 36 , written as $P($ Age $=36)$


## Probabilities and Data Matrix

- To do a valid statistical analysis, we need to understand how the data in the data matrix was generated


## Example: Students in Munich

- Let's assume that there are $\tilde{N}=50000$ students in Munich. This set is called the population
- $\tilde{N}$ is the size of the population, often assumed to be infinite
- Formally, I put all 50000 students in an urn (bag)
- I randomly select a student: this is called an (atomic) event or an experiment and defines a random process
- $\omega$ : The selected student is an outcome of the experiment and defines a row in the data matrix; if Jack was selected, then $\omega=$ Jack


## Sample

- A particular student will be picked with elementary probability $1 / \tilde{N}$
- Performing the experiment $N$ times produces a sample (training data set) $\mathbf{D}$ of size $N$
- An analysis of the sample can give us insight about the population (statistical inference)
- Sampling with replacement: I return the student to the urn after the experiment; then $P(\omega=J a c k)=1 / \tilde{N}$
- Sampling without replacement: I do not return the student to the urn after the experiment; this is how a normal data matrix is formed


## Random Variable

- On each selected student, we perform a measurement, i.e., height $H$, and the result of the measurement is a value, e.g., (tiny, small, medium, tall, huge); $H$ is called a random variable
- A random variable (e.g., Height) is a function (measurement) of the outcome of the random experiment); its value depends on the outcome (e.g., "Jack") of the random process
- $\operatorname{Height}(\omega=$ Jack $)=$ tall means the height measurement on the outcome $\omega=$ Jack resulted in the measurement "tall"
- $P(X=x)=P(\{\omega \in \Omega \mid X(\omega)=x\})$
- $\Omega$ : the set of all students in Munich; $\omega \in \Omega$ : one particular student (Jack) who was picked as a result of the random experiment; $x=$ tall: result of the measurement Height
- Since I know the probability of an elementary event, by design, I can estimate this number by simply looking at the training data


## Probability

- If I know the population and know each individual's height, I know, $P(H=$ tall $)$; but often the population concerns entities on which I cannot perform measurements, like the student population in two years
- In statistics, one estimates the probability from the sample (the training data)
- Then the probability that a randomly picked student has height $H=h$ is defined as

$$
P(H=h)=\lim _{N \rightarrow \infty} \frac{N_{h}}{N}
$$

with $0 \leq P(H=h) \leq 1$

- $N_{h}$ is the number of times that a selected student is observed to have height $H=h$


## Sample / Training Data

- I can estimate

$$
\widehat{P}(H=h)=\frac{N_{h}}{N} \approx P(H=h)
$$

- This is the number of times that we observe the value of $h$ in column $H$ in the data matrix, divided by the number of observations $N$
- In statistics one is interested in how well $\widehat{P}(H=h)$ (the probability estimate derived from the sample) approximates $P(H=h)$ (the probability in the population)
- Note the importance of the definition of a population: $P(H=h)$ might be different, when I consider individuals in Munich or Germany
- Thus the population plays an important role in a statistical analysis
- Note that the randomness enters through the sampling process: Jack's height is not random


## Law of Large Numbers

- Law of Large Numbers (Bernoulli)

$$
P\left\{\left|N_{h} / N-P(H=h)\right|<\epsilon\right\} \rightarrow 1 \quad \text { as } \quad N \rightarrow \infty
$$

- Even if there is one 110 m high giant in the population: if the giant shows up in the sample or not: with $n \rightarrow \infty$ the giant becomes irrelevant
- If I predict the height of the next sampled person by the mean (e.g., 175 cm ), the error will be small, but there is a small probability that I pick the giant and I am wrong by more than 100 m


## Statistics and Probability

- Probability is a mathematical discipline developed as an abstract model and its conclusions are deductions based on axioms (Kolmogorov axioms)
- Statistics deals with the application of the theory to real problems and its conclusions are inferences or inductions, based on observations (Papoulis: Probability, Random variables, and Stochastic Processes)
- Frequentist or classical statistics and Bayesian statistics apply probability in slightly different ways


## Joint Probabilities

- Now assume that we also measure weight (size) $S$ with weight attributes very light, light, normal, heavy, very heavy. Thus $S$ is a second random variable
- Similarly

$$
P(S=s)=\lim _{N \rightarrow \infty} \frac{N_{s}}{N}
$$

- We can also count co-occurrences

$$
P(H=h, S=s)=\lim _{N \rightarrow \infty} \frac{N_{h, s}}{N}
$$

This is called the joint probability distribution of $H$ and $S$

## Marginal Probabilities

- It is obvious that we can calculate the marginal probability $P(H=h)$ from the joint probabilities

$$
\begin{gathered}
P(H=h)=\lim _{N \rightarrow \infty} \frac{\sum_{s} N_{h, s}}{N} \\
=\sum_{s} P(H=h, S=s)
\end{gathered}
$$

- This is called marginalization
- I can calculate the marginal probability from the joint probability (without going back to the counts)


## Conditional Probabilities

- One is often interested in the conditional probability. Let's assume that I am interested in the probability distribution of $S$ for a given height $H=h$. Since I need a different normalization I get

$$
P(S=s \mid H=h)=\lim _{N \rightarrow \infty} \frac{N_{h, s}}{N_{h}}
$$

So I count the co-occurrences, but I normalize by $N_{h}$

## Conditional Probabilities (cont'd)

- Then,

$$
P(S=s \mid H=h)=\frac{P(H=h, S=s)}{P(H=h)}
$$

- Relationship to machine learning: $H=h$ is the input and $S=s$ is the output
- Conditioning is closely related to the definition of a population: $P(S=s \mid H=h)$ is the same as $P(S=s)$ in a population which is restricted to students with $H=h$


## Product Rule and Chain Rule

- It follows: product rule

$$
\begin{gathered}
P(S=s, H=h)=P(S=s \mid H=h) P(H=h) \\
=P(H=h \mid S=s) P(S=s)
\end{gathered}
$$

- and chain rule

$$
P\left(x_{1}, \ldots, x_{M}\right)=P\left(x_{1}\right) P\left(x_{2} \mid x_{1}\right) P\left(x_{3} \mid x_{1}, x_{2}\right) \ldots P\left(x_{M} \mid x_{1}, \ldots, x_{M-1}\right)
$$

## Bayes Formula

- If 1 know $P(S=s \mid H=h)$, does it tell me anything about $P(H=h \mid S=s)$ ? Is it the same thing?
- No, but the relationship is given by Bayes formula


## Bayes Formula (con't)

- We use the definition of a conditional probability,

$$
\begin{aligned}
& P(H=h \mid S=s)=\frac{P(H=h, S=s)}{P(S=s)} \\
& P(S=s \mid H=h)=\frac{P(H=h, S=s)}{P(H=h)}
\end{aligned}
$$

- Thus we get Bayes' formula

$$
P(H=h \mid S=s)=\frac{P(S=s \mid H=h) P(H=h)}{P(S=s)}
$$

or

$$
P(H=h \mid S=s)=P(S=s \mid H=h) \frac{P(H=h)}{P(S=s)}
$$

Matlab/octave code to illustrate Bayes formula
$P(X=x \mid Y=y)=P(Y=y \mid X=x) P(X=x) /$ $P(Y=y)$
\% num states of $X$ and $Y$
$m=4 \% x$
$\mathrm{n}=6 \% \mathrm{y}$
\% Some P(X)
$\mathrm{px}=\operatorname{rand}(\mathrm{m}), 1$
$\mathrm{px}=\mathrm{px} / \operatorname{sum}(\mathrm{px})$
\% Some P(YIX)
pygx $=\operatorname{rand}(\mathrm{n}, \mathrm{m})$
pygx $=$ pygx.$/(\text { ones( } \mathrm{n}, 1)^{*} \operatorname{sum}($ pygx $\left.)\right)$
\% Some P(Y)
$\mathrm{py}=\mathrm{pygx}{ }^{*} \mathrm{px}$
\% Bayes Formula: $\mathrm{P}(\mathrm{X}=31 \mathrm{Y}=2)$
Skalar: $\mathrm{P}(\mathrm{X}=3 \mid \mathrm{Y}=2)$

Vector: $P(X=x \mid Y=2)$

Matrix: $P(X=x \mid Y=y)$
pxis3gyis2 $=\operatorname{pygx}(2,3)^{*} \operatorname{px}(3) / p y(2)$
\% Bayes Formula for all X: $\mathrm{P}(\mathrm{XIY}=2)$
pxgyis2 $=\left(p y g x(2,:)^{\prime} . .^{*} p x\right) / p y(2)$
\% Bayess Formula for all x and y
pxgy = pygx' . (px * ones(1, n)) ./ (ones(m, 1) * py')

## Independent Random Variables

- Independence: two random variables are independent, if,

$$
\begin{aligned}
P(S=s, H & =h)=P(S=s) P(H=h \mid S=s) \\
& =P(S=s) P(H=h)
\end{aligned}
$$

## Simplified Notation

- The expression $P(X=x)$ is often simplified as $P(x)$
- Thus instead of writing $P(H=185)$, we write $P(185)$
- Joint: $P(X=x, Y=y) \equiv P(x, y)$
- Marginalization: $P(Y=y)=\sum_{x} P(X=x, Y=y)$ becomes

$$
P(x)=\sum_{x} P(x, y)
$$

- Sometimes $X$ stands for the event $X=x$ with some unspecified $x$; thus one sees also $P(X), P(X, Y)$, and

$$
P(X)=\sum_{X} P(X, Y)
$$

## Summary

- Conditional probability

$$
P(y \mid x)=\frac{P(x, y)}{P(x)} \text { with } P(x)>0
$$

- Product rule

$$
P(x, y)=P(x \mid y) P(y)=P(y \mid x) P(x)
$$

- Chain rule

$$
P\left(x_{1}, \ldots, x_{M}\right)=P\left(x_{1}\right) P\left(x_{2} \mid x_{1}\right) P\left(x_{3} \mid x_{1}, x_{2}\right) \ldots P\left(x_{M} \mid x_{1}, \ldots, x_{M-1}\right)
$$

- Bayes' theorem

$$
P(y \mid x)=\frac{P(x, y)}{P(x)}=\frac{P(x \mid y) P(y)}{P(x)} \quad P(x)>0
$$

- Marginal distribution

$$
P(x)=\sum_{y} P(x, y)
$$

- Independent random variables

$$
P(x, y)=P(x) P(y \mid x)=P(x) P(y)
$$

## Simplifications for Supervised Learning

- Consider that we are interested in $P\left(Y \mid X_{1}, \ldots X_{M}\right)$, e.g., to predict the last component based on all previous ones (supervised learning)
- Then, from a statistical point of view, we do not need to randomly sample the $X_{1}, \ldots X_{M}$, we can design the inputs; thus the term "design matrix"
- The requirement is that the targets $Y$ are randomly sampled from $P\left(Y \mid X_{1}, \ldots X_{M}\right)$
- Each input defines an urn and from that urn, we need to sample randomly
- E.g., if the input is height and the output is weight, I can select systematically people based on height, but I cannot select them based on weight
- E.g., if the input is the cause and the output is effect, I can set the cause (give medication or not) and record the outcome; but and I cannot not only select patients, where the medication has worked worked
- Of course, for many other reasons, the selected inputs should correspond to the population I am interested in




## Marginalization and Conditioning: Basis for Probabilistic Inference

- $P(I, F, S)$ where $I=1$ stands for influenza, $F=1$ stands for fever, $S=1$ stands for sneezing
- What is the probability for influenza, when the patient is sneezing, but temperature is unknown, $P(I \mid S)$ ?
- Thus I need (conditioning) $P(I=1 \mid S=1)=P(I=1, S=1) / P(S=1)$
- I calculate via marginalization

$$
\begin{gathered}
P(I=1, S=1)=\sum_{f} P(I=1, F=f, S=1) \\
P(S=1)=\sum_{i} P(I=i, S=1)
\end{gathered}
$$

## Expected Values

- Expected value

$$
\begin{aligned}
E(X)= & E_{P(x)}(X)=\sum_{i} x_{i} P\left(X=x_{i}\right) \\
& \approx \frac{1}{N} \sum_{k=1}^{N} x_{k}=\operatorname{mean}_{x}
\end{aligned}
$$

(with random observations)

## Variance

- The Variance of a random variable is:

$$
\operatorname{var}(X)=\sum_{i}\left(x_{i}-E(X)\right)^{2} P\left(X=x_{i}\right) \approx \frac{1}{N-1} \sum_{i}\left(x_{i}-\text { mean }_{x}\right)^{2}
$$

- The Standard Deviation is its square root:

$$
\operatorname{stdev}(X)=\sqrt{\operatorname{var}(X)}
$$

## Covariance

- Covariance:

$$
\begin{aligned}
\operatorname{cov}(X, Y)= & \sum_{i} \sum_{j}\left(x_{i}-E(X)\right)\left(y_{j}-E(Y)\right) P\left(X=x_{i}, Y=y_{j}\right) \\
& \approx \frac{1}{N-1} \sum_{i}\left(x_{i}-\operatorname{mean}_{x}\right)\left(y_{i}-\text { mean }_{y}\right)
\end{aligned}
$$

- Covariance matrix:

$$
\Sigma_{[X Y],[X Y]}=\left(\begin{array}{cc}
\operatorname{var}(X) & \operatorname{cov}(X, Y) \\
\operatorname{cov}(Y, X) & \operatorname{var}(Y)
\end{array}\right)
$$

## Covariance, Correlation, and Correlation Coefficient

- Useful identity:

$$
\operatorname{cov}(X, Y)=E(X Y)-E(X) E(Y)
$$

where $E(X Y)$ is the correlation.

- The (Pearson) correlation coefficient (confusing naming!) is

$$
r=\frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X)} \sqrt{\operatorname{var}(Y)}}
$$

- It follows that $\operatorname{var}(X)=E\left(X^{2}\right)-(E(X))^{2}$ and

$$
\operatorname{var}(f(X))=E\left(f(X)^{2}\right)-(E(f(X)))^{2}
$$

## More Useful Rules

- We have, independent of the correlation between $X$ and $Y$,

$$
E(X+Y)=E(X)+E(Y)
$$

and thus also

$$
E\left(X^{2}+Y^{2}\right)=E\left(X^{2}\right)+E\left(Y^{2}\right)
$$

- For the variance of the sum of random variables,

$$
\begin{gathered}
\operatorname{var}(X+Y)=E\left[(X+Y-(E(X)+E(Y)))^{2}\right] \\
=E\left[((X-E(X))+(Y-E(Y)))^{2}\right] \\
=E\left[(X-E(X))^{2}\right]+E\left[(Y-E(Y))^{2}\right]+2 E[(X+E(X))(Y-E(Y)] \\
=\operatorname{var}(X)+\operatorname{var}(Y)+2 \operatorname{cov}(X, Y)
\end{gathered}
$$

- Similarly,

$$
\operatorname{var}(X-Y)=\operatorname{var}(X)+\operatorname{var}(Y)-2 \operatorname{cov}(X, Y)
$$

## Covariance Matrix of Linear Transformation

- Let $\mathbf{w}$ be a random vector with mean $\vec{\mu}_{\mathbf{w}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}}$
- Let

$$
\mathrm{y}=\mathrm{Aw}+\vec{\epsilon}
$$

where $\mathbf{A}$ is a fixed matrix.

- Then $\mathbf{y}$ is a random vector with mean $\vec{\mu}_{y}=\mathbf{A} \vec{\mu}_{w}$ and covariance

$$
\boldsymbol{\Sigma}_{\mathbf{y}}=\mathbf{A} \boldsymbol{\Sigma}_{\mathbf{w}} \mathbf{A}^{T}+\sigma^{2} \mathbf{I}
$$

- Special case (Gaussain distributions): $P(\mathbf{w})=\mathcal{N}\left(\mathbf{w} ; \vec{\mu}_{w}, \mathbf{\Sigma}_{\mathbf{w}}\right), P(\mathbf{y} \mid \mathbf{w})=$ $\mathcal{N}\left(\mathbf{y} \mid \mathbf{A w}, \sigma^{2} \mathbf{I}\right)$ then $P(\mathbf{y})=\mathcal{N}\left(\mathbf{y} ; \mathbf{A} \vec{\mu}_{w}, \mathbf{A} \boldsymbol{\Sigma}_{\mathbf{w}} \mathbf{A}^{T}+\sigma^{2} \mathbf{I}\right)$
- Special case $\left(\sigma^{2}=0\right): \boldsymbol{\Sigma}_{\mathbf{y}}=\mathbf{A} \boldsymbol{\Sigma}_{\mathbf{w}} \mathbf{A}^{T}$


## Continuous Random Variables

- Probability density

$$
f(x)=\lim _{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x+\Delta x)}{\Delta x}
$$

- Thus

$$
P(a<x<b)=\int_{a}^{b} f(x) d x
$$

- The distribution function is

$$
F(x)=\int_{-\infty}^{x} f(x) d x=P(X \leq x)
$$

## Expectations for Continuous Variables

- Expected value

$$
E(X)=E_{P(x)}(X)=\int x P(x) d x
$$

- Variance

$$
\operatorname{var}(X)=\int(x-E(x))^{2} P(x) d x
$$

- Covariance:

$$
\operatorname{cov}(X, Y)=\iint(x-E(X))(y-E(Y)) P(x, y) d x d y
$$



## Joint Gaussian Distributions

- Let $\mathbf{z}=(\mathbf{x} ; \mathbf{y}), \vec{\mu}=\left(\vec{\mu}_{x} ; \vec{\mu}_{y}\right)$
- With

$$
\boldsymbol{\Sigma}=\left(\begin{array}{ll}
\boldsymbol{\Sigma}_{x, x} & \boldsymbol{\Sigma}_{x, y} \\
\boldsymbol{\Sigma}_{y, x} & \boldsymbol{\Sigma}_{y, y}
\end{array}\right)
$$

we get

$$
P(\mathbf{z})=\mathcal{N}(\mathbf{z} ; \vec{\mu}, \mathbf{\Sigma})=\frac{1}{(2 \pi)^{M / 2} \sqrt{|\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2}(\mathbf{z}-\vec{\mu})^{T} \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\vec{\mu})\right)
$$

Here $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$.

## Marginals

- For $\mathbf{x}$,

$$
P(\mathrm{x})=\mathcal{N}\left(\mathbf{x} ; \vec{\mu}_{x}, \mathbf{\Sigma}_{x, x}\right)
$$

- For $\mathbf{y}$,

$$
P(\mathbf{y})=\mathcal{N}\left(\mathbf{y} ; \vec{\mu}_{y}, \Sigma_{y, y}\right)
$$



## 



## Conditional Densities

- For the conditionals, we get

$$
P(\mathbf{x} \mid \mathbf{y})=\mathcal{N}\left(\mathbf{x} ; \vec{\mu}_{x}+\boldsymbol{\Sigma}_{x, y} \boldsymbol{\Sigma}_{y, y}^{-1}\left(\mathbf{y}-\vec{\mu}_{y}\right), \boldsymbol{\Sigma}_{x, x}-\boldsymbol{\Sigma}_{x, y} \boldsymbol{\Sigma}_{y, y}^{-1} \boldsymbol{\Sigma}_{y, x}\right)
$$

- and

$$
P(\mathbf{y} \mid \mathbf{x})=\mathcal{N}\left(\mathbf{y} ; \vec{\mu}_{y}+\boldsymbol{\Sigma}_{y, x} \boldsymbol{\Sigma}_{x, x}^{-1}\left(\mathbf{x}-\vec{\mu}_{x}\right), \boldsymbol{\Sigma}_{y, y}-\Sigma_{y, x} \boldsymbol{\Sigma}_{x, x}^{-1} \boldsymbol{\Sigma}_{x, y}\right)
$$

# Knowledge Graphs 

## Unary and Binary Relations

- A knowledge graph does not just record the attributes of entities Height(Jack, Tall) but also relationships to other entities
- In the statement marriedTo(Jack, Mary), married is a binary predicate (relation) since the arguments are two entities: Jack, Mary; as a triple: (Jack, marriedTo, Mary)
- In the Knowledge Graph ground facts of binary relations are represented as directed links from subject to object
- It contains statements which we, as humans, would interpret as facts but which often are not measurements in the scientific sense
- Can I enter facts in the KG, such that I can be sure with certainty that the butler is the murder
- Can I formulate my background knowledge within the KG or is it outside?


## Jonas Salk



- Google Knwoledge Graph
- "Facts" can be statements (assertions) I find, e.g., in Wikipedia knownFor(JonasSalk, PolioVaccine, Source=Wikipedia)


## Semiotic Triangle

- By introducing a measurement device and a result of the measurement, we imply a form of objectivity; in addition there is the thought in our brain and what is spoken or written (see: semiotic triangle)
- The three views are not identical: I can think and say things which cannot be measured: "Love is stronger than Hate"; reality can produce things, which people cannot understand
- Working definition of the truth value of a statement: For a statement like "Love is stronger than Hate'", find $N$ (random) individuals who claim to be competent to evaluate that statement. The truth value (probability) is the percentage of those individuals who would confirm that the statement is true
- What is in the real world:
- What can be measured
- Independent of humans:



## Concluding Comments

- Modern science started with the concept of an objective measurement
- Statistics permits valid scientific statements of the context of the experiment is clearly defined; you cannot escape your own statistics
- Science and statistics permit predictions about the future (prediction, forecasting)
- What is written and said is related to the scientific facts in the world but also goes beyond
- Written and verbal statements can assume the role of data and permit predictions: Jack likes GWTW, then, I predict, he also likes ... We as humans and science finds patterns everywhere
- Amazingly, we as humans and also GPT can deal with all this complexity


It is difficult to escape statistics:

- So far 77 papers have cited me in 2024
- Maybe no other paper will cite me in 2024
- But it is more likely that approximately 2600 papers will cite me
(1) You
what does it mean that "love is stronger than hate"?


## ChatGPT

The phrase "love is stronger than hate" conveys the idea that the positive and constructive emotions associated with love have more enduring power and influence than the negative and destructive emotions associated with hate. This perspective suggests that acts of kindness, compassion, understanding, and empathy have a greater impact on individuals and society as a whole than acts driven by anger, hostility, and animosity.

In essence, the message is that love has the ability to transcend and overcome the destructive force of hate. Advocates of this viewpoint argue that love promotes unity, healing, and positive change, while hate tends to divide, harm, and create negative consequences. It's a belief in the resilience of goodness and the transformative power of positive emotions over negative ones. This idea is often expressed in various forms in literature, philosophy, and religious teachings as a way to encourage harmony, tolerance, and the pursuit of common humanity.

## You

Mary was found dead on Sunday morning in her London apartment. Jack, her boyfriend, was in Manchester on Sunday morning. Can he be the murderer? The train from London to Manchester takes 3 hours.
(3) ChatGPT

Based on the information provided, it seems unlikely that Jack could be the murderer if he was in Manchester on Sunday morning. If the train from London to Manchester takes 3 hours, and Mary was found dead on Sunday morning in her London apartment, it suggests that Jack would not have had enough time to travel from London to Manchester and commit the crime.

However, it's essential to consider other factors such as the exact time of Mary's death, Jack's alibi, and any additional details that may emerge during the investigation.
Without more information, it is not possible to definitively rule out or implicate Jack as the murderer.
(1) You

Mary was found dead on Sunday morning in her London apartment. Jack, her boyfriend, was in Manchester on Sunday morning. Can he be the murderer? The train from London to Manchester takes 3 hours. Can you generate a story

## ChatGPT

On a chilly Sunday morning in London, the city was shrouded in mist as Detective Inspector Reynolds received a call about a disturbing incident. Mary Thompson, a 28-year-old woman, was found dead in her apartment. The circumstances surrounding her demise were mysterious, and Reynolds knew he had to unravel the truth.

Upon reaching the crime scene, Reynolds began his investigation. Mary's lifeless body lay on the living room floor, and the apartment showed no signs of forced entry. The initial examination hinted at foul play, but the lack of obvious clues left Reynolds puzzled.

Mary's boyfriend, Jack Harrison, was quickly identified as a person of interest. However, an intriguing twist surfaced when Reynolds discovered that Jack had an alibi. Jack claimed he was in Manchester on Sunday morning, miles away from the crime scene.

To validate Jack's alibi, Reynolds checked train schedules and travel times. The train from London to Manchester took approximately three hours. Reynolds obtained surveillance footage from the train station, confirming Jack's presence in Manchester during the crucial time frame.

Despite the seemingly airtight alibi, Reynolds couldn't shake off his suspicions. He delved deeper into Mary and Jack's relationship, uncovering a history of arguments and strained moments. As the pieces of the puzzle started to fit together, Reynolds realized that the key to solving the case might lie within the intricate web of their personal lives.

During the course of the investigation, Reynolds unearthed a series of text messages exchanged between Mary and Jack, revealing a heated argument just days before her death. The messages hinted at unresolved issues and a potential motive for murder.

Determined to get to the bottom of the case, Reynolds continued to explore every lead. As he gathered more evidence, the line between truth and deception blurred. The investigation took unexpected turns, leading Reynolds to question everyone involved in Mary's life.

In the end, the answers to Mary's untimely demise were found in the shadows of secrets and lies. The journey from London to Manchester became a metaphorical passage through the twists and turns of a complex relationship, ultimately exposing the truth behind Mary's tragic end.

