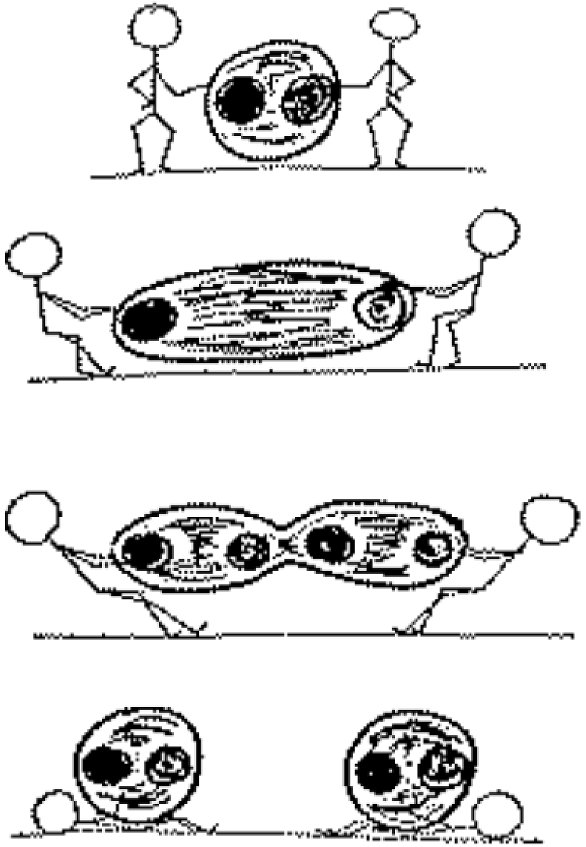
To be(t) or not to be(t): a Bayesian approach to statistical data analysis

Silvia Pisano

Centro Ricerche «Enrico Fermi»
Laboratori Nazionali di Frascati - INFN

Istituto Nazionale di Fisica Nucleare
Laboratori Nazionali di Frascati

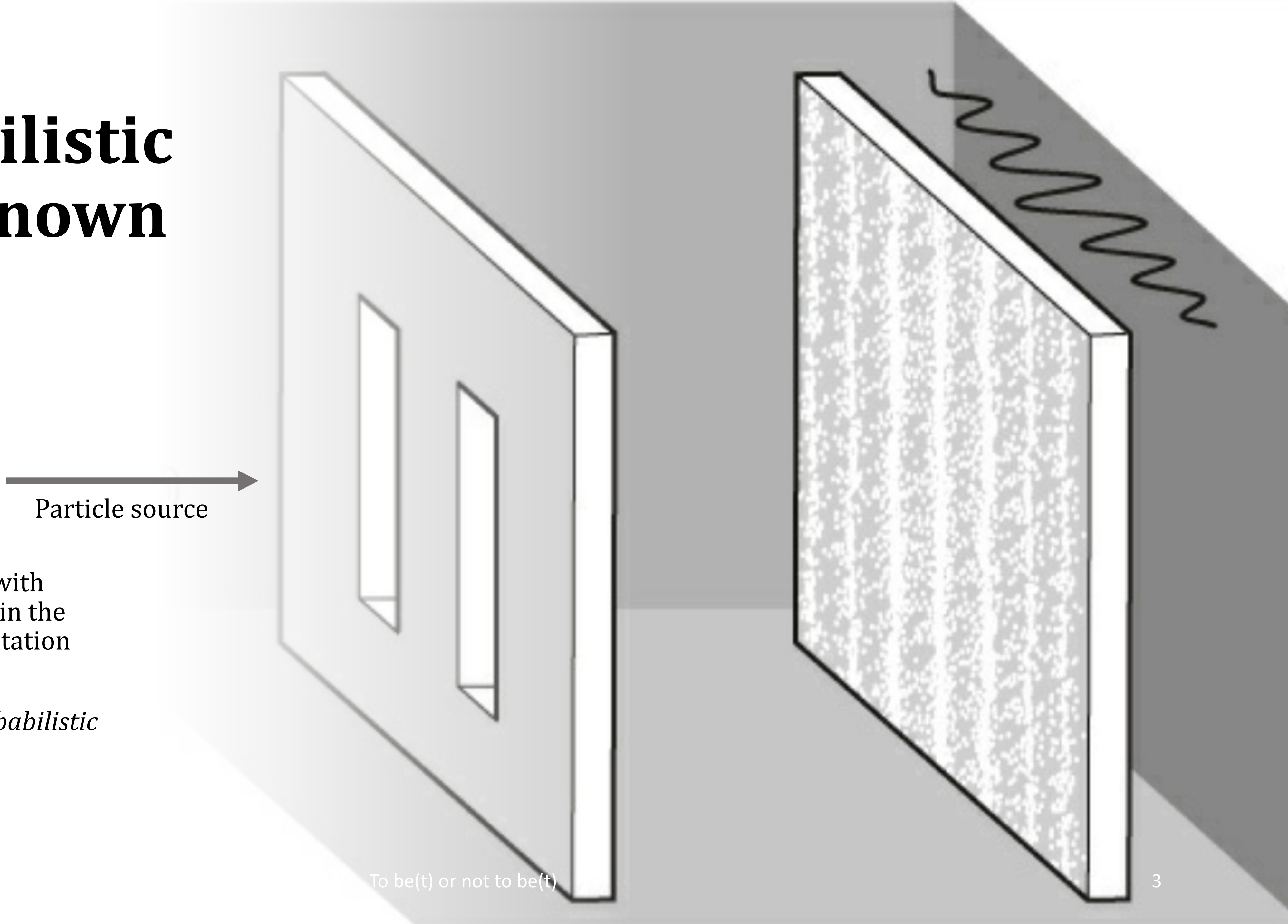


Probabilistic ***vs.*** **unknown**

When tossing a coin in the Trevi fountain, we do not know if it will fall on the head or on the tail side

→ we do not have control on all the conditions affecting the outcome of the toss

Probabilistic *vs.* unknown

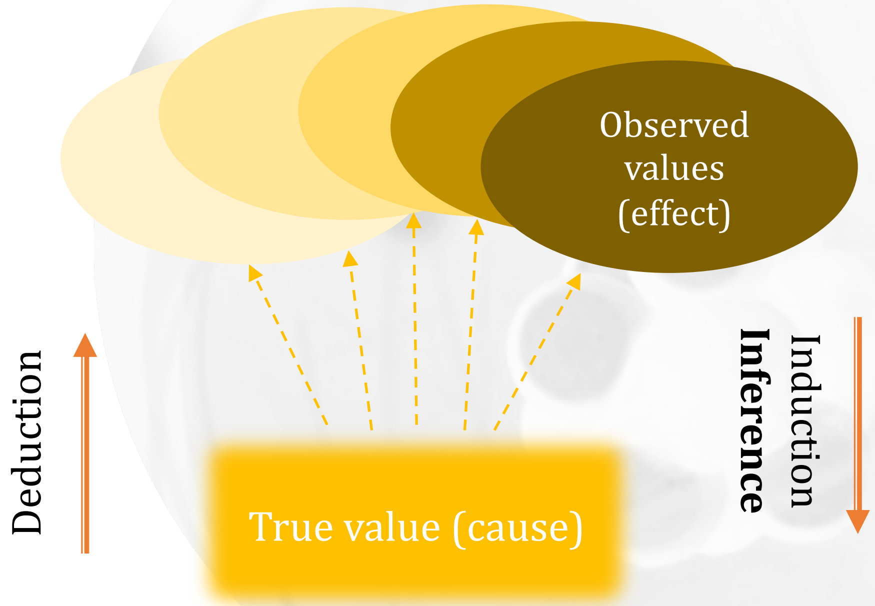


Particle source

Different is the case with Quantum Mechanics in the Copenaghen interpretation

→ *the outcome is probabilistic*

From causes to effects



Induction: process of learning from empirical observations

(Probabilistic) deduction: going from causes to effects

From causes to effects

Example of random variables

1. What is the real weight of the grapes, having observed certain values on the scale?

   → *There are measurement errors*

2. We bought 1kg of grapes: what number will appear on the scale?

   → *The link between observations and theory is alwasy of probabilistic nature*

The Likelihood

It represents the joint probability of the observed data as a function of the parameters of a given model.

$\mathcal{L}(\vartheta|x) \rightarrow$ probability of observing the data x assuming the value ϑ for a given parameter.

ϑ is chosen so to maximize the likelihood.

The Likelihood

It represents the joint probability of the observed data as a function of the parameters of a given model.

$\mathcal{L}(\vartheta|x)$ → probability of observing the data x assuming the value ϑ for a given parameter.

ϑ is chosen so to maximize the likelihood.

Example: I toss a fair coin and obtain 4 consecutive times *head*

The Likelihood

It represents the joint probability of the observed data as a function of the parameters of a given model.

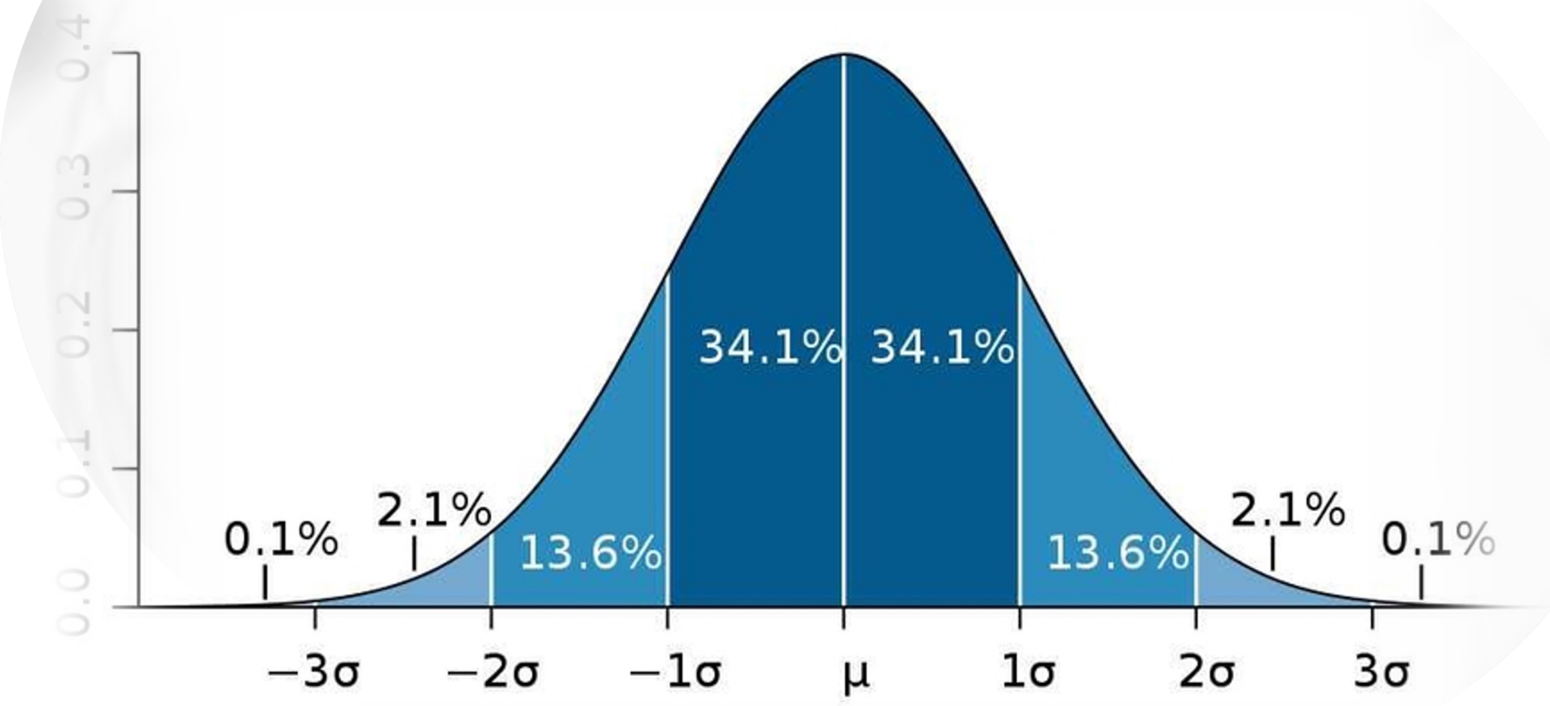
$\mathcal{L}(\vartheta|x) \rightarrow$ probability of observing the data x assuming the value ϑ for a given parameter.

ϑ is chosen so to maximize the likelihood.

Example: I toss a fair coin and obtain 4 consecutive times *head*

...but what it is the probability that the coin is fair?

The Gaussian



Learning from data

We do n independent measurements x_i of an observable X (with n high *enough*) → what we learn about X?

What is X **true value** μ?

$$\mu = \bar{x} \pm {}^{\sigma}/_{\sqrt{n}}$$

Learning from data

We do n independent measurements x_i of an observable X (with n high *enough*) → what we learn about X?

What is X **true value** μ?

$$\mu = \bar{x} \pm {}^{\sigma}/_{\sqrt{n}}$$

We would highly tempted to interpret it as

$$P\left(\bar{x} - {}^{\sigma}/_{\sqrt{n}} < \mu < \bar{x} + {}^{\sigma}/_{\sqrt{n}}\right) = 68\%$$

Learning from data

We do n independent measurements x_i of an observable X (with n high *enough*) → what we learn about X?

What is X **true value** μ?

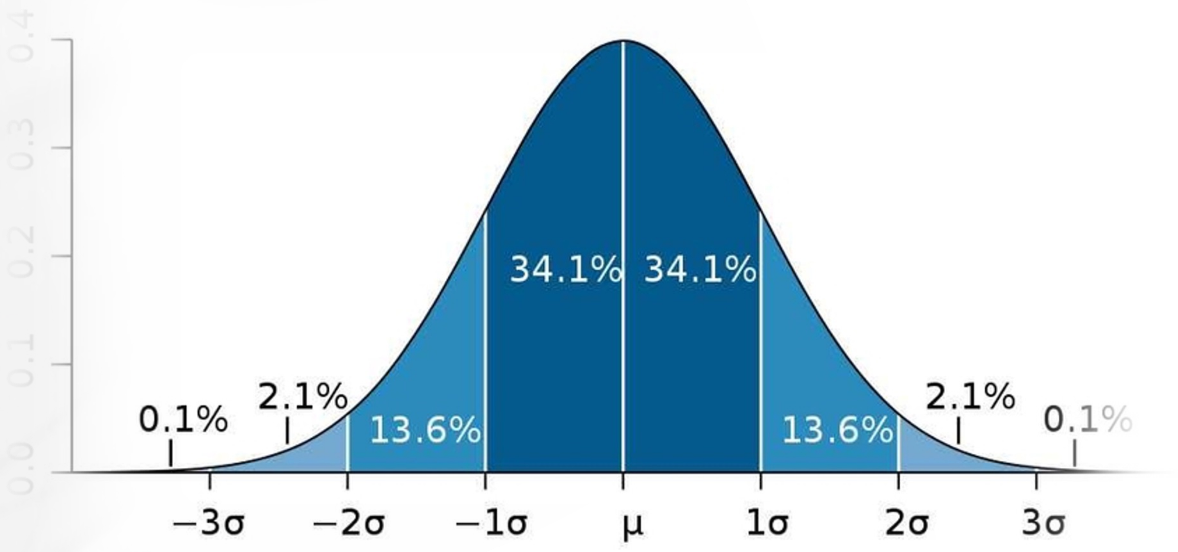
$$\mu = \bar{x} \pm {}^{\sigma}/_{\sqrt{n}}$$

We would highly tempted to interpret it as

$$P\left(\bar{x} - {}^{\sigma}/_{\sqrt{n}} < \mu < \bar{x} + {}^{\sigma}/_{\sqrt{n}}\right) = 68\%$$

while it actually means

$$P\left(\mu - {}^{\sigma}/_{\sqrt{n}} < \bar{x} < \mu + {}^{\sigma}/_{\sqrt{n}}\right) = 68\%$$



Learning from data

We do n independent measurements x_i of an observable X (with n high *enough*) → what we learn about X?

What is X **true value** μ?

$$\mu = \bar{x} \pm {}^{\sigma}/_{\sqrt{n}}$$

We would highly tempted to interpret it as

$$P\left(\bar{x} - {}^{\sigma}/_{\sqrt{n}} < \mu < \bar{x} + {}^{\sigma}/_{\sqrt{n}}\right) = 68\%$$

while it actually means

$$P\left(\mu - {}^{\sigma}/_{\sqrt{n}} < \bar{x} < \mu + {}^{\sigma}/_{\sqrt{n}}\right) = 68\%$$

μ is a constant of unknown value → it cannot be interpreted probabilistically

However, what is in the upper formula is exactly what we want: **having observed the effect E (x) what can we say about the cause H (μ)?**

Probability of the causes

What we *really* want to know is the **probability of the causes**, but we only have **the probability of the data.**

Probability of the causes

What we *really* want to know is the **probability of the causes**, but we only have **the probability of the data.**

...at least through standard statistical tools

Probability of the causes

What we *really* want to know is the **probability of the causes**, but we only have **the probability of the data.**

$\mathrm{P}(H|\,\boldsymbol{data})$ → probability of the hypothesis H having observed the data («*Is this track a* π^- *or a* K^-? »)

$\mathrm{P}(\boldsymbol{data}|H)$ → probability of the data under the hypothesis that H is true («*What is the probability of two consecutive heads when tossing a* ***regular*** *coin*? »)

→ The likelihood $\mathcal{L}(\vartheta|x)$

Conditional probability

The probability that E will occur under the hypothesis that H has occurred

Assessment of the probability of E, under the condition that H is true → it must be independent on the probability of H!

$$\mathrm{P}(E|H) = \frac{P(E \cap H)}{P(H)}$$

P(E) is *practically* always P(E|*something*)! *E.g.*, when throwing a dice we assume it is regular

What would P(E|Ω = *whatever happens*) mean?

Conditional probability

The probability that E will occur under the hypothesis that H has occurred

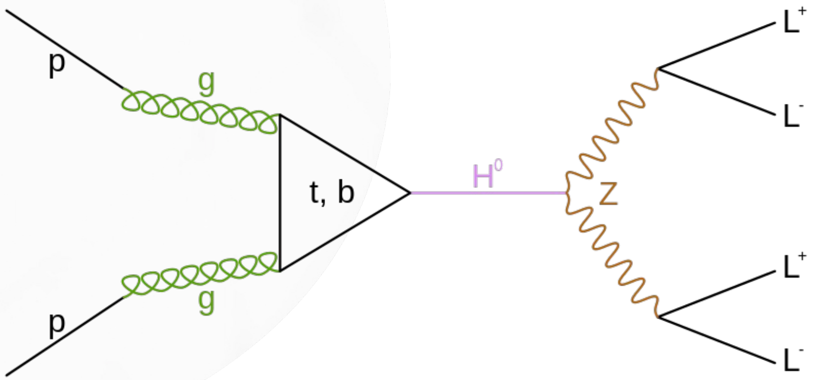
Assessment of the probability of E, under the condition that H is true → it must be independent on the probability of H!

$$P(E|H) = \frac{P(E \cap H)}{P(H)}$$

P(E) is *practically* always P(E|*something*)! *E.g.*, when throwing a dice we assume it is regular

What would P(E|Ω = *whatever happens*) mean?

Example: *«perfectly» efficient detector looking for a rare event*



Conditional probability

The probability that E will occur under the hypothesis that H has occurred

Assessment of the probability of E, under the condition that H is true → it must be independent on the probability of H!

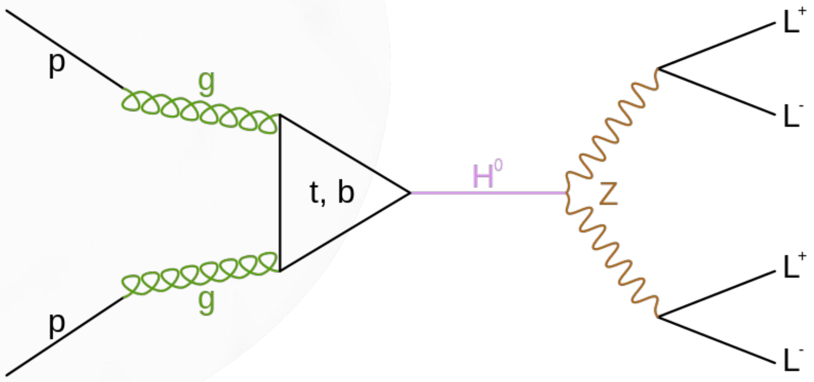
$$\mathrm{P}(E|H) = \frac{P(E \cap H)}{P(H)}$$

P(E) is *practically* always P(E|*something*)! *E.g.*, when throwing a dice we assume it is regular

What would P(E|Ω = *whatever happens*) mean?

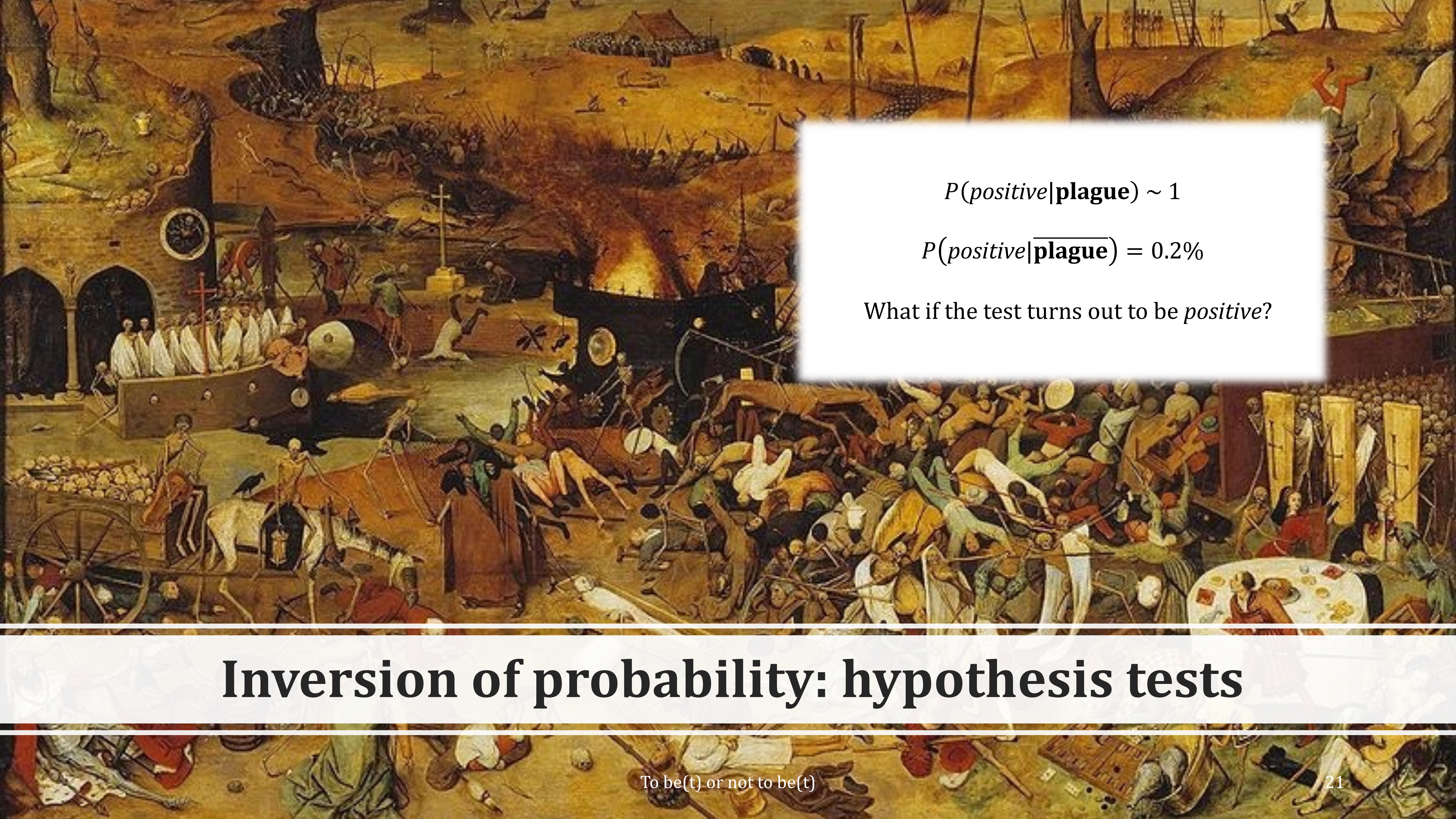
$P(E|U) \neq P(E \cap U)$, *the probability that both events occur.*

$P(E \cap U)$ can be very small, while $P(E|U)$ is very high



Think of the limiting case $P(H) \equiv P(H \cap H) \leq P(H|H) = 1$

Inversion of probability: hypothesis tests



$$P(positive|\mathbf{plague}) \sim 1$$

$$P\left(positive|\overline{\mathbf{plague}}\right) = 0.2\%$$

What if the test turns out to be *positive*?

Inversion of probability: hypothesis tests

The missing ingredient

$$P(positive|\mathbf{plague}) \sim 1$$

$$P(positive|\overline{\mathbf{plague}}) = 0.2\%$$

What if the test turns out to be *positive*?

The missing ingredient

$$P(positive|\mathbf{plague}) \sim 1$$

$$P(positive|\overline{\mathbf{plague}}) = 0.2\%$$

What if the test turns out to be *positive*?

$P(\boldsymbol{data}|H) \equiv P(positive|\overline{\mathbf{plague}})$ → probability of the data under the hypothesis that H is true («*What is the probability that I have a positive test result* ***not having the plague****?* »)

$P(H|\ \boldsymbol{data}) \equiv P(\overline{\mathbf{plague}}|positive)$ → probability of the hypothesis H having observed the data («*What is the probability that I do not have the plague if* ***I have a positive test result***»)

The missing ingredient

$P(positive|\textbf{plague}) \sim 1$

$P(positive|\overline{\textbf{plague}}) = 0.2\%$

The prior $\mathrm{P}(H)$

A priori probability of the hypothesis H

What if the test turns out to be *positive*?

$\mathrm{P}(\boldsymbol{data}|H) \equiv P(positive|\overline{\textbf{plague}})$ → probability of the data under the hypothesis that H is true («*What is the probability that I have a positive test result* ***not having the plague****?* »)

$\mathrm{P}(H|\ \boldsymbol{data}) \equiv P(\overline{\textbf{plague}}|positive)$ → probability of the hypothesis H having observed the data («*What is the probability that I do not have the plague if* ***I have a positive test result***»)

The HERA excess

ELSEVIER

Nuclear Physics B (Proc. Suppl.) 66 (1998) 120–123

Beyond the Standard Model Physics at HERA

C. Diaconu [a]

On behalf of the H1 and ZEUS collaborations

[a]Centre de Physique des Particules de Marseille

163 Av. de Luminy, case 907, 13288 Marseille cedex 09, France

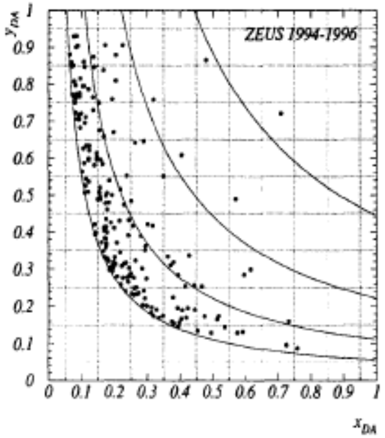


Figure 4. ZEUS selected NC DIS candidate events in the $x_{DA} - y_{DA}$ plane.

The mimimum probablity of 0.6% is obtained for a cut $x_{DA} > 0.57$. The probability to obtain such a statistical fluctuation from the Standard Model for any x_{DA} cut is 7.6%.

The H1 and ZEUS data and expectations as a function of a Q^2 lower limit are shown in table 1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More statistics is needed in order to clarify the excess observed by H1 and ZEUS in the NC DIS high Q^2 regime. If the excess persists, then it could be interpreted in terms of leptoquark production, R-parity violating SUSY process or contact terms.

4. Acknowledgements

The help in preparing this talk and the provision of information and plots by several H1 and ZEUS members are greatly appreciated.

REFERENCES

1. Future Physics at HERA, eds. G. Ingelman, A. de Roeck and R. Klanner, (DESY, Hamburg 1996) and reference therein; R.Cashmore et al., Phys. Reports 122 (1985) 275.
2. K.Hagiwara, S. Komamiya, D. Zeppenfeld, Z. Phys. C29 (1985) 115.
3. H1 Collaboration, S.Aid. et al., Nucl. Phys. B483 (1997) 44.
4. ZEUS Collaboration,
5. H1 Collaboration, S.Aid et al., Phys. Lett. B380 (1996) 461.
6. H1 Collaboration, C.Adloff et al., Z. Phys. C74 (1997) 191.
7. ZEUS Collaboration, J.Breitweg et al., Z. Phys. C74 (1997) 207.

The HERA excess

ELSEVIER

1. The probability to observe such an excess at $Q^2 > 15000\ \mathrm{GeV}^2$ due to statistical fluctuations from the Standard Model is less than 1%. More

Beyond t

C. Diaconu [a]
On behalf of the H1 and ZEUS

[a]Centre de Physique des Particules de Marseille
163 Av. de Luminy, case 907, 13288 Marseille cedex 09, France

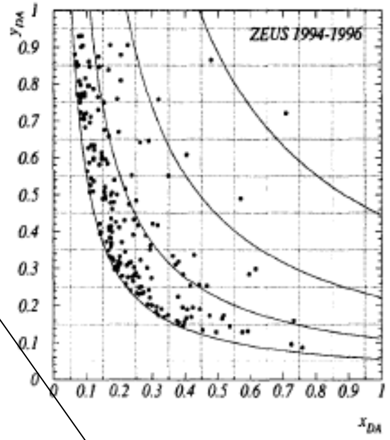


Figure 4. ZEUS selected NC DIS candidate events in the $x_{DA} - y_{DA}$ plane.

The mimimum probablity of 0.6% is obtained for a cut $x_{DA} > 0.57$. The probability to obtain such a statistical fluctuation from the Standard Model for any x_{DA} cut is 7.6%.

The H1 and ZEUS data and expectations as a function of a Q^2 lower limit are shown in table 1. The probability to observe such an excess at $Q^2 > 15000\ \mathrm{GeV}^2$ due to statistical fluctuations from the Standard Model is less than 1%. More statistics is needed in order to clarify the excess observed by H1 and ZEUS in the NC DIS high Q^2 regime. If the excess persists, then it could be interpreted in terms of leptoquark production, R-parity violating SUSY process or contact terms.

4. Acknowledgements

The help in preparing this talk and the provision of information and plots by several H1 and ZEUS members are greatly appreciated.

REFERENCES

1. Future Physics at HERA, eds. G. Ingelman, A. de Roeck and R. Klanner, (DESY, Hamburg 1996) and reference therein; R.Cashmore et al., Phys. Reports 122 (1985) 275.
2. K.Hagiwara, S. Komamiya, D. Zeppenfeld, Z. Phys. C29 (1985) 115.
3. H1 Collaboration, S.Aid. et al., Nucl. Phys. B483 (1997) 44.
4. ZEUS Collaboration,
5. H1 Collaboration, S.Aid et al., Phys. Lett. B380 (1996) 461.
6. H1 Collaboration, C.Adloff et al., Z. Phys. C74 (1997) 191.
7. ZEUS Collaboration, J.Breitweg et al., Z. Phys. C74 (1997) 207.

The HERA excess

ELSEVIER

Beyond t

C. Diaconu [a]
On behalf of the H1 and ZEUS

[a]Centre de Physique des Particules de Marseille
163 Av. de Luminy, case 907, 13288 Marseille cedex 09, France

1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More

$$\mathrm{P}(\boldsymbol{data}|SM) \lesssim 1\%$$

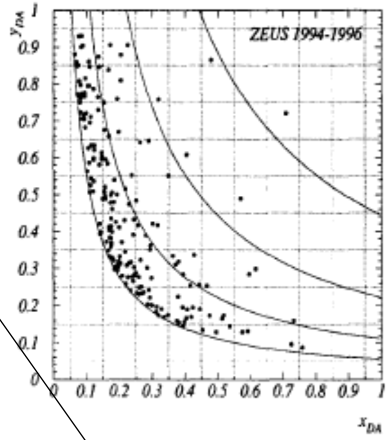


Figure 4. ZEUS selected NC DIS candidate events in the $x_{DA} - y_{DA}$ plane.

The mimimum probablity of 0.6% is obtained for a cut $x_{DA} > 0.57$. The probability to obtain such a statistical fluctuation from the Standard Model for any x_{DA} cut is 7.6%.

The H1 and ZEUS data and expectations as a function of a Q^2 lower limit are shown in table 1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More statistics is needed in order to clarify the excess observed by H1 and ZEUS in the NC DIS high Q^2 regime. If the excess persists, then it could be interpreted in terms of leptoquark production, R-parity violating SUSY process or contact terms.

4. Acknowledgements

The help in preparing this talk and the provision of information and plots by several H1 and ZEUS members are greatly appreciated.

REFERENCES

1. Future Physics at HERA, eds. G. Ingelman, A. de Roeck and R. Klanner, (DESY, Hamburg 1996) and reference therein; R.Cashmore et al., Phys. Reports 122 (1985) 275.
2. K.Hagiwara, S. Komamiya, D. Zeppenfeld, Z. Phys. C29 (1985) 115.
3. H1 Collaboration, S.Aid. et al., Nucl. Phys. B483 (1997) 44.
4. ZEUS Collaboration,
5. H1 Collaboration, S.Aid et al., Phys. Lett. B380 (1996) 461.
6. H1 Collaboration, C.Adloff et al., Z. Phys. C74 (1997) 191.
7. ZEUS Collaboration, J.Breitweg et al., Z. Phys. C74 (1997) 207.

The HERA excess

ELSEVIER

1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More

$P(\boldsymbol{data}|SM) \lesssim 1\%$

Beyond t…

C. Diaconu [a]

On behalf of the H1 and ZEU…

[a]Centre de Physique des Particules de Marseille
163 Av. de Luminy, case 907, 13288 Marseille cedex 09, France

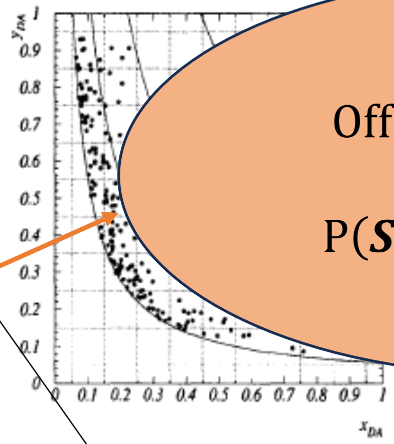


Figure 4. ZEUS selected NC DIS candidate events in the $x_{DA} - y_{DA}$ plane.

The mimimum probablity of 0.6% is obtained for a cut $x_{DA} > 0.57$. The probability to obtain such a statistical fluctuation from the Standard Model for any x_{DA} cut is 7.6%.

The H1 and ZEUS data and expectations as a function of a Q^2 lower limit are shown in table 1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More

…to clarify the excess …NC DIS high …could be …, R- …s.

Official statement:

$P(\boldsymbol{SM}|\boldsymbol{data}) \lesssim 1\%$

…, eds. G. Ingelman, …R. Klanner, (DESY, Hamburg 1996) and reference therein; R.Cashmore et al., Phys. Reports 122 (1985) 275.
2. K.Hagiwara, S. Komamiya, D. Zeppenfeld, Z. Phys. C29 (1985) 115.
3. H1 Collaboration, S.Aid. et al., Nucl. Phys. B483 (1997) 44.
4. ZEUS Collaboration,
5. H1 Collaboration, S.Aid et al., Phys. Lett. B380 (1996) 461.
6. H1 Collaboration, C.Adloff et al., Z. Phys. C74 (1997) 191.
7. ZEUS Collaboration, J.Breitweg et al., Z. Phys. C74 (1997) 207.

The HERA excess

CSN3 Fisica Nucleare

Istituto Nazionale di Fisica Nucleare
Laboratori Nazionali di Frascati

ELSEVIER

Beyond t...

C. Diaconu [a]
On behalf of the H1 and ZEUS...

[a]Centre de Physique des Particules de Marseille
163 Av. de Luminy, case 907, 13288 Marseille cedex 09, France

1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More

$$\mathrm{P}(\boldsymbol{data}|SM) \lesssim 1\%$$

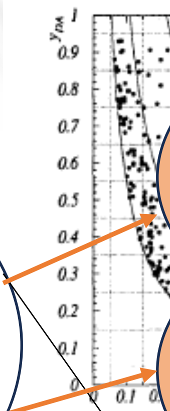


Figure 4. ZEUS selected NC DIS events in the $x_{DA} - y_{DA}$ plane.

The mimimum probablity of 0.6% is obtained for a cut $x_{DA} > 0.57$. The probability to obtain such a statistical fluctuation from the Standard Model for any x_{DA} cut is 7.6%.

The H1 and ZEUS data and expectations as a function of a Q^2 lower limit are shown in table 1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More

... to clarify the excess ... NC DIS high ... could be ... R- ...

... Nucl. Phys.
5. H1 Collaboration, S.Aid et al., Phys. Lett. B380 (1996) 461.
6. H1 Collaboration, C.Adloff et al., Z. Phys. C74 (1997) 191.
7. ZEUS Collaboration, J.Breitweg et al., Z. Phys. C74 (1997) 207.

Official statement:

Media statement:

$$\mathrm{P}(\overline{SM}|\boldsymbol{data}) \lesssim 99\%$$

The HERA excess

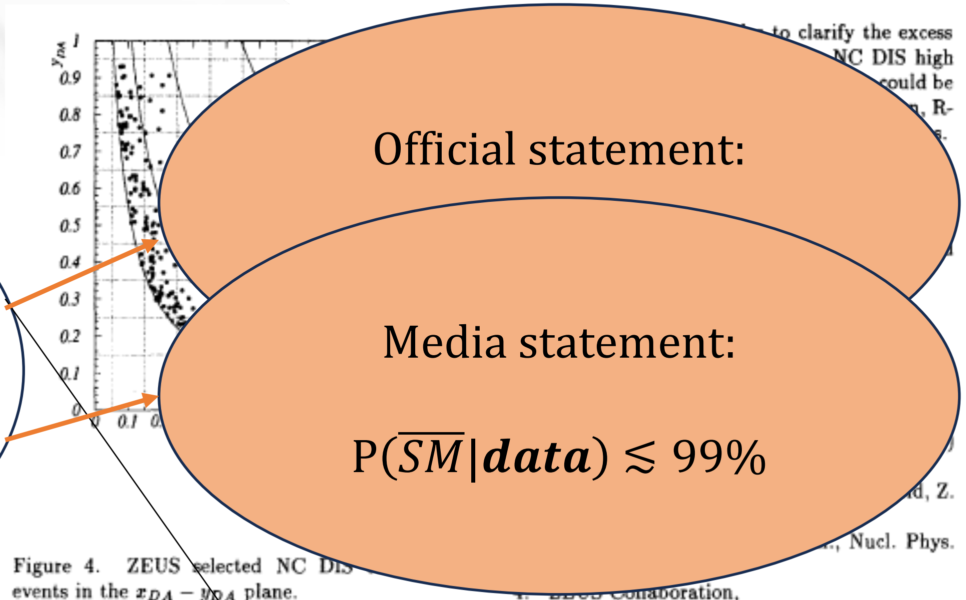


Figure 4. ZEUS selected NC DIS events in the $x_{DA} - y_{DA}$ plane.

The mimimum probablity of 0.6% is obtained for a cut $x_{DA} > 0.57$. The probability to obtain such a statistical fluctuation from the Standard Model for any x_{DA} cut is 7.6%.

The H1 and ZEUS data and expectations as a function of a Q^2 lower limit are shown in table 1. The probability to observe such an excess at $Q^2 > 15000$ GeV^2 due to statistical fluctuations from the Standard Model is less than 1%. More

5. H1 Collaboration, S.Aid et al., Phys. Lett. B380 (1996) 461.
6. H1 Collaboration, C.Adloff et al., Z. Phys. C74 (1997) 191.
7. ZEUS Collaboration, J.Breitweg et al., Z. Phys. C74 (1997) 207.

The Bayes theorem

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Probability. A deeper look at the coin

1. The ratio between the number of favorable cases to the number of all cases (*classical or combinatorial*)

2. The ratio of the number of times an event occurred in a test series to the total number of trials (*frequentistic*).

Probability. A deeper look at the coin

1. The ratio between the number of favorable cases to the number of all cases (*classical or combinatorial*)

   …if all the cases are equally probable!

2. The ratio of the number of times an event occurred in a test series to the total number of trials (*frequentistic*).

   …for a number of trials very large!

Probability. A deeper look at the coin

1. The ratio between the number of favorable cases to the number of all cases (*classical or combinatorial*)

   ...if all the cases are equally probable!

2. The ratio of the number of times an event occurred in a test series to the total number of trials (*frequentistic*).

   ...for a number of trials very large!

Subjective probability

A measure of the degree of belief that an event will occur

Subjective probability

A measure of the degree of belief that an event will occur

1. No issues with circularity of the definition

2. No issues with equiprobability of the possible outcomes

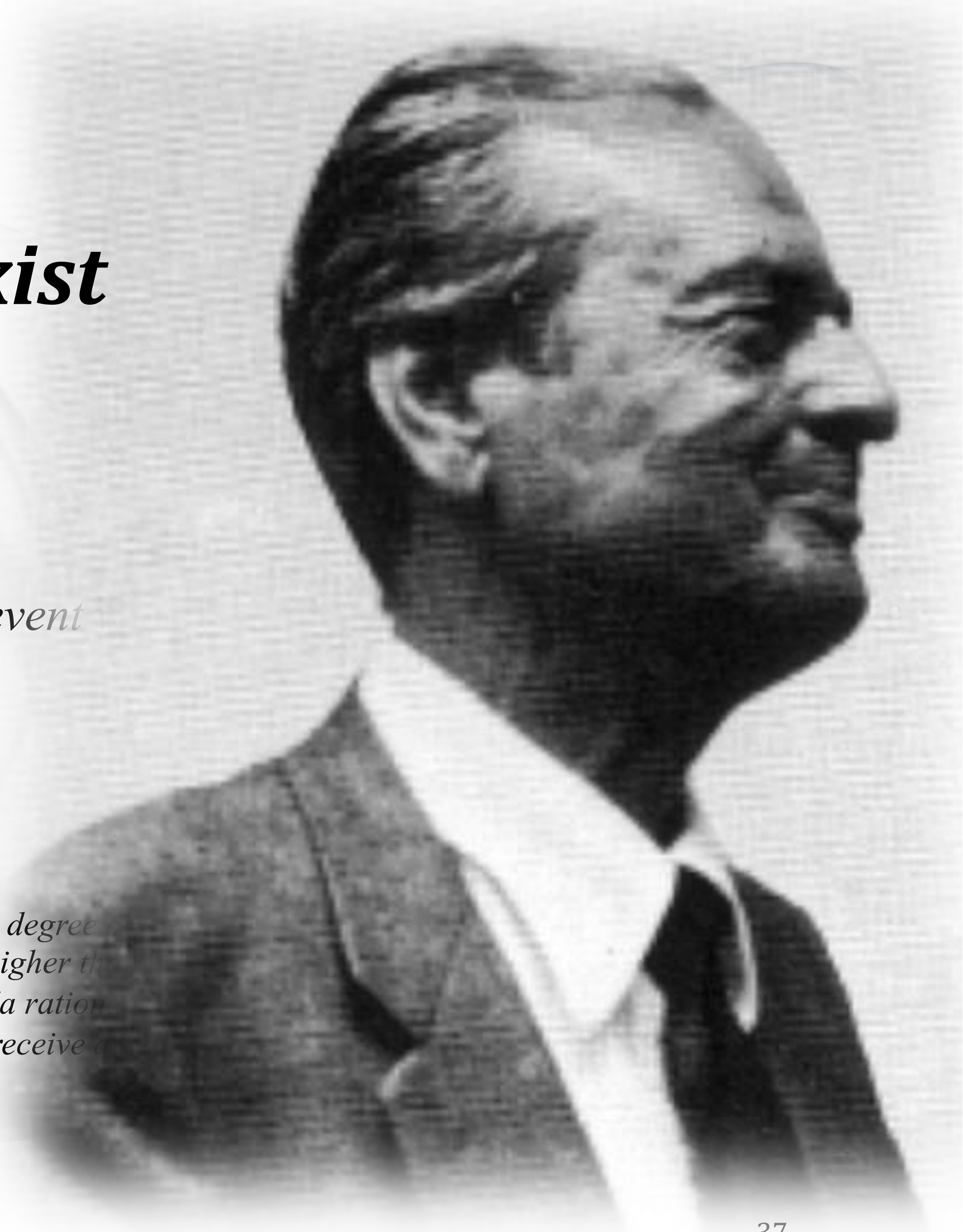
3. No issues with the number of repetitions to be perform

4. **Odds in betting.** *The higher the degree of belief that an event will occur, the higher the amount of money A that someone ("a rational better") is ready to pay in order to receive a sum of money B if the event occurs.*

5. **P(E) is not an intrinsic characteristic of the event E**, *but depends on the state of information available to whoever evaluates P(E).*

probability
does not exist

Bruno de Finetti



A measure of the degree of belief that an event

5. **P(E) is not an intrinsic characteristic of the event E**, *but depends on the state of information available to whoever evaluates P(E).*

4. **Odds in betting.** *The higher the degree belief that an event will occur, the higher th amount of money A that someone ("a ration better") is ready to pay in order to receive sum of money B if the event occurs.*

Inspecting the Bayes theorem

$$P(H|E) = P(E|H)P(H)$$

It explicitly shows how the probability of a certain hypothesis is updated when the state of information changes

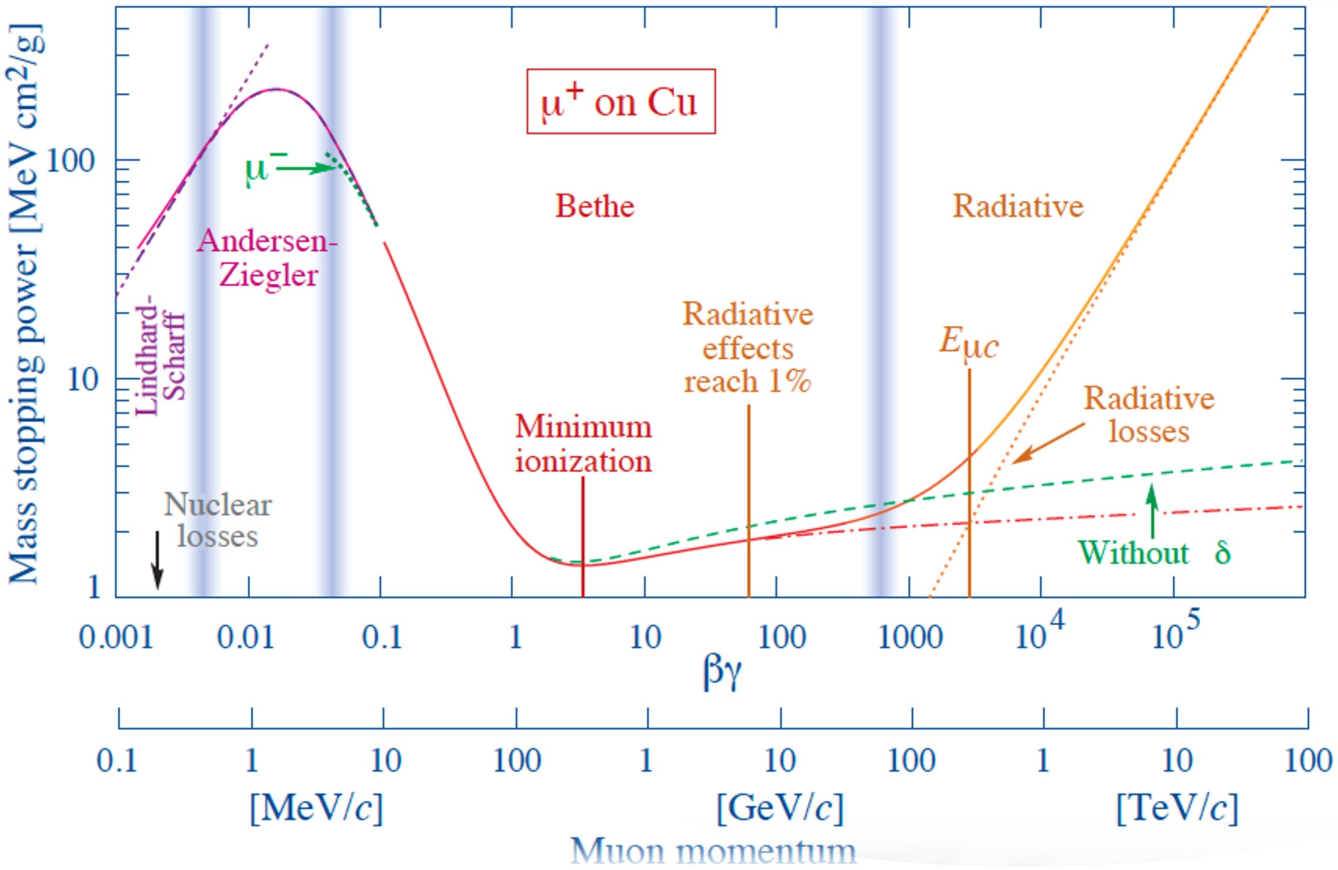
$P(H)$ → *The prior*, *i.e.* the probability of this hypothesis with the state of information available 'before' the knowledge that E has occurred

$P(E|H)$ → *The likelihood*, *i.e.* the likelihood that a cause will produce a given effect

$P(H|E)$ → *The posterior*, *i.e.* the final, or 'a posteriori', probability of Hi 'after' the new information

Formulating the experimental question

…in terms of the Bayes theorem



E: observed event
H_i: mutually exclusive hypothesis that could condition the event E

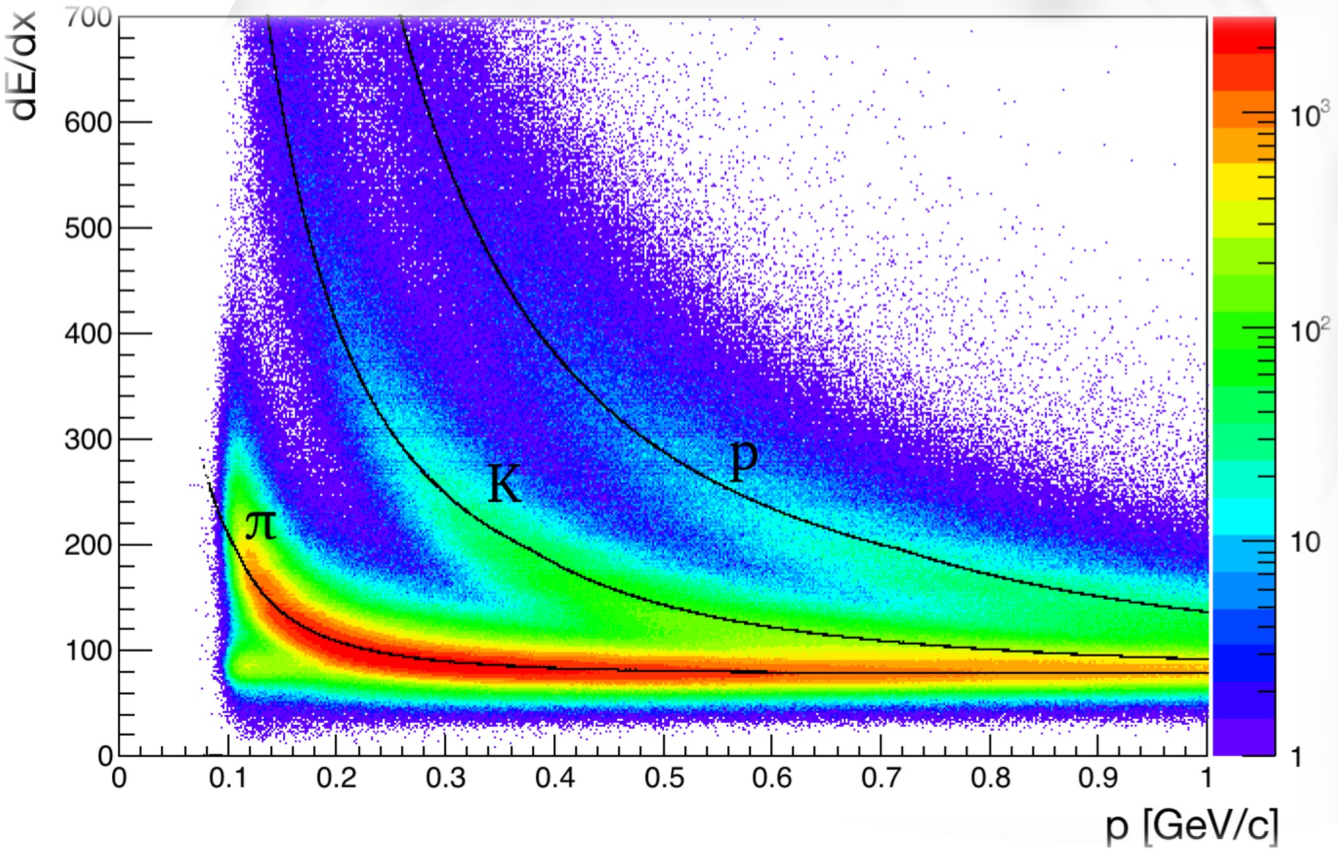
What is the probability of H_i under the hypothesis that E has occurred?

Example: E → energy loss observed in a detector

https://pdg.lbl.gov/2022/reviews/rpp2022-rev-passage-particles-matter.pdf

Formulating the experimental question

…in terms of the Bayes theorem



E: observed event
H_i: mutually exclusive hypothesis that could condition the event E

What is the probability of H_i under the hypothesis that E has occurred?

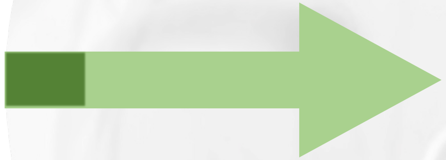
Example: $E \rightarrow$ energy loss observed in a detector

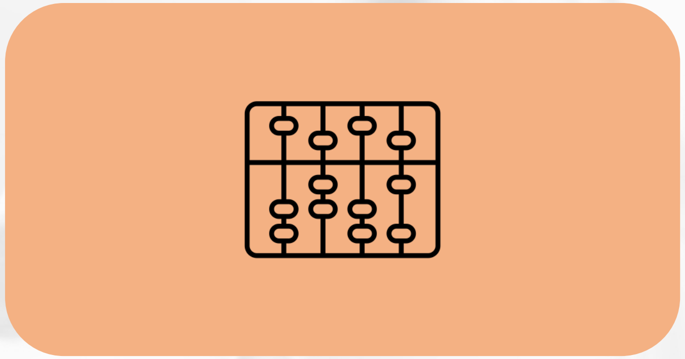
H_i: particle hypothesis, i.e. π, K, p

Application to real life measurements

Particle identification

Beam = 90% π + 10% μ





Identification efficiencies

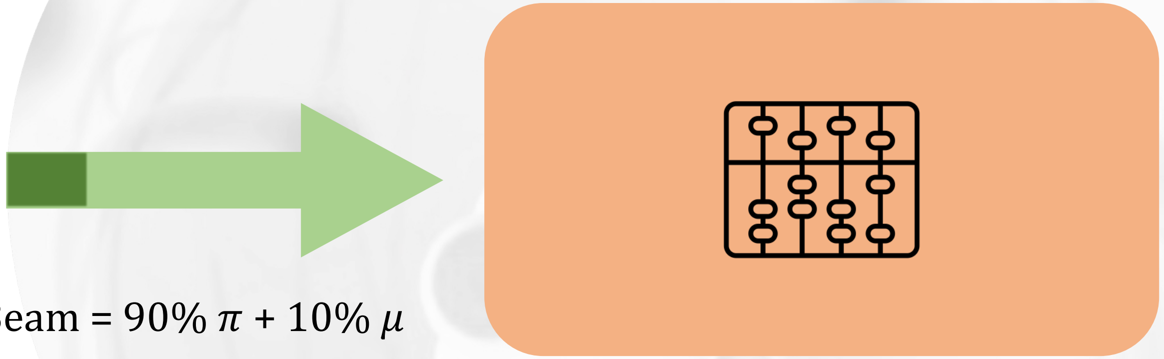
$$\epsilon_{\mu} = 95\%$$
$$\epsilon_{\pi \to \mu} = 2\%$$

The trigger accepts the event if the particle is identified as a μ

Application to real life measurements

Particle identification

Beam = 90% π + 10% μ



Identification efficiencies

$$\epsilon_{\mu} = 95\%$$
$$\epsilon_{\pi\to\mu} = 2\%$$

The trigger accepts the event if the particle is identified as a μ

What is the probability that the trigger is fired by a μ?

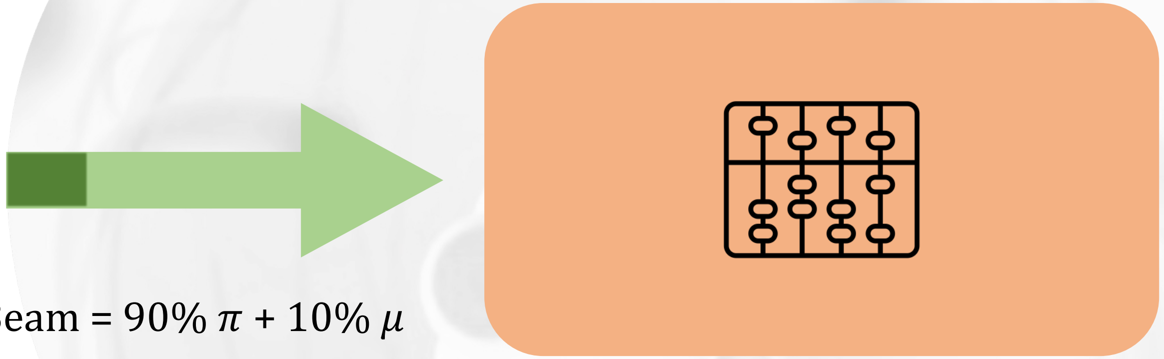
The event E is «The trigger is fired», the hypotheses are «μ» or «π»

$$P(\mu|T) = \frac{P(T|\mu)P_0(\mu)}{P(T|\mu)P_0(\mu) + P(T|\pi)P_0(\pi)} = \frac{0.95\times 0.1}{0.95\times 0.1 + 0.02\times 0.9} = 0.84$$

Application to real life measurements

Particle identification

Beam = 90% π + 10% μ



Identification efficiencies

$$\epsilon_{\mu} = 95\%$$
$$\epsilon_{\pi\to\mu} = 2\%$$

The trigger accepts the event if the particle is identified as a μ

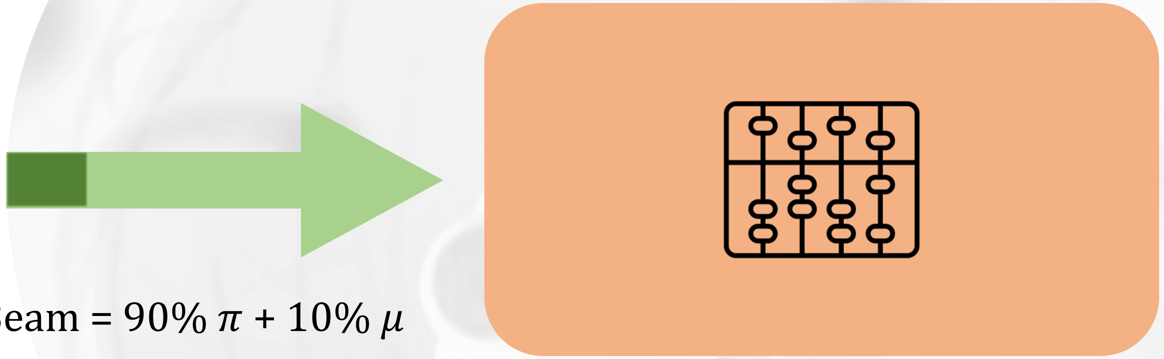
What is the ratio S/N?

The event E is «The trigger is fired», the hypotheses are «μ» or «π»

$$\frac{S}{N} = \frac{P(\mu|T)}{P(\pi|T)} = \frac{P(S|E)}{P(N|E)} = \frac{P(E|S)}{P(E|N)}\frac{P_0(S)}{P_0(N)}$$

Application to real life measurements

Particle identification



Beam = 90% π + 10% μ

Identification efficiencies

$$\epsilon_{\mu} = 95\%$$
$$\epsilon_{\pi\to\mu} = 2\%$$

The trigger accepts the event if the particle is identified as a μ

What is the ratio S/N?

The event E is «The trigger is fired», the hypotheses are «μ» or «π»

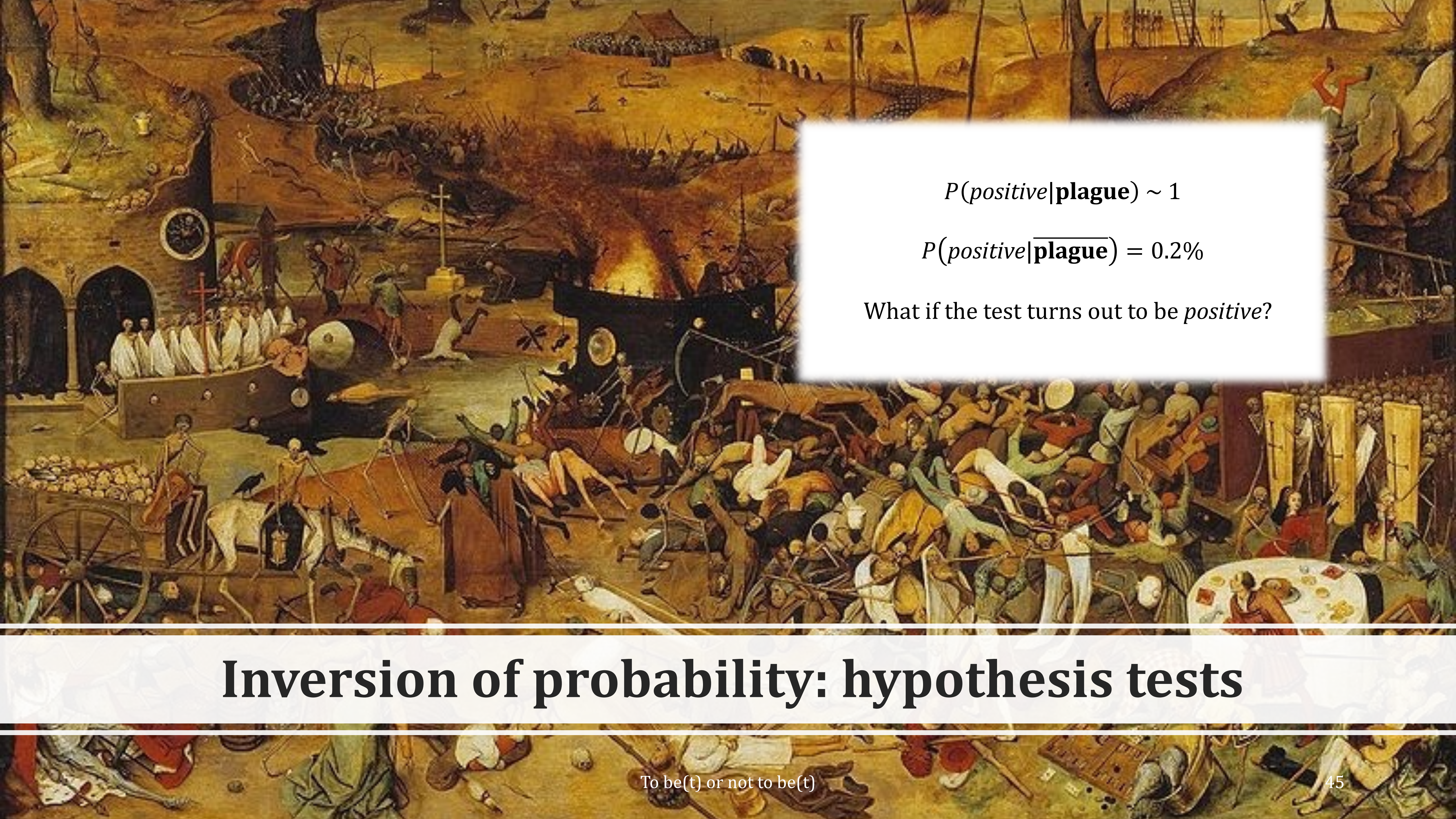
$$\frac{S}{N} = \frac{P(\mu|T)}{P(\pi|T)} = \frac{P(S|E)}{P(N|E)} = \frac{P(E|S)}{P(E|N)}\frac{P_0(S)}{P_0(N)}$$

If conditions are noisy, i.e.

$$P_0(S) << P_0(N)$$

experiment must be very selective:

$$P(E|S) >> P(E|S)$$



$$P(positive|\mathbf{plague}) \sim 1$$

$$P\left(positive|\overline{\mathbf{plague}}\right) = 0.2\%$$

What if the test turns out to be *positive*?

Inversion of probability: hypothesis tests

The Bayesian solution

$P(positive|\mathbf{plague}) \sim 1$

$P\left(positive|\overline{\mathbf{plague}}\right) = 0.2\%$

What if the test turns out to be *positive*?

The Bayesian solution

$P(positive|\mathbf{plague}) \sim 1$

$P(positive|\overline{\mathbf{plague}}) = 0.2\%$

What if the test turns out to be *positive*?

What is the probability of having the plague if the test turns out to be positive?

→ *A prior to describe the «degree of belief» for the hypothesis must be introduced: let's assume only 0.002% of the population is affected by the desease (by previous observations, general knowledge etc)*

The test can turns out to be positive because:

1. $P(positive|\mathbf{plague}) \sim 1$
2. $P(positive|\overline{\mathbf{plague}}) = 0.2\%$

But, a this point, the priors $P(\mathbf{plague})$ and $P(\overline{\mathbf{plague}})$ must be considered

The Bayesian solution

$P(positive|\mathbf{plague}) \sim 1$

$P\left(positive|\overline{\mathbf{plague}}\right) = 0.2\%$

What if the test turns out to be *positive*?

What is the probability of having the plague if the test turns out to be positive?

→ *A prior to describe the «degree of belief» for the hypothesis must be introduced: let's assume only 0.002% of the population is affected by the desease (by previous observations, general knowledge etc)*

Total positive people: 100000 (because infected) + 120000 (not infected, but resulting affected for unefficiency of the test). So, the probability is

$$P(\mathbf{plague}|positive) = \frac{P(positive|\mathbf{plague})P(\mathbf{plague})}{P\left(positive|\overline{\mathbf{plague}}\right)P\left(\overline{\mathbf{plague}}\right) + P(positive|\mathbf{plague})P(\mathbf{plague})} = 0.45$$

The Bayesian solution

$$P(positive|\mathbf{plague}) \sim 1$$

$$P(positive|\overline{\mathbf{plague}}) = 0.2\%$$

What if the test turns out to be *positive*?

What is the probability of having the plague if the test turns out to be positive?

→ *A prior to describe the «degree of belief» for the hypothesis must be introduced: let's assume only 0.002% of the population is affected by the desease (by previous observations, general knowledge etc)*

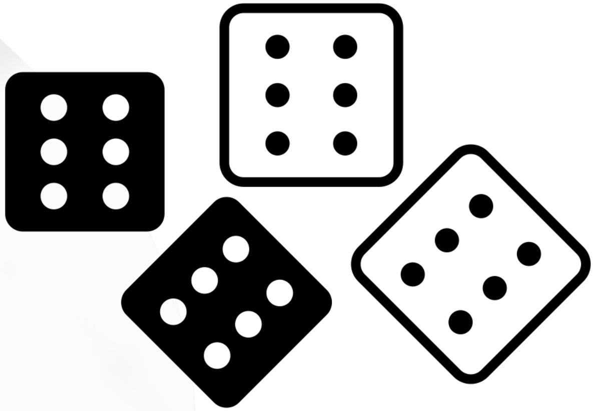
In terms of the ratio S/N:

$$\frac{P(\mathbf{plague}|positive)}{P(\overline{\mathbf{plague}}|positive)} = \frac{P(positive|\mathbf{plague})}{P(positive|\overline{\mathbf{plague}})}\frac{P(\mathbf{plague})}{P(\overline{\mathbf{plague}})} = \frac{\approx 1}{0.002} \times \frac{0.1/60}{\approx 1} \sim 0.83$$

Since the situation is that of *noisy conditions*, the selectivity of the test is not enough

Why it works?

A deeper look at the dice



$\mathcal{L}(x|\mu) \rightarrow$ likelihood that the observation x are due to μ

$\mathcal{L}(x|\mu) \rightarrow f(x|\mu)$

General formula of Bayesian inference:

$$f(\mu|x) = \frac{f(x|\mu) f_0(\mu)}{\int f(x|\mu) f_0(\mu) d\mu}$$

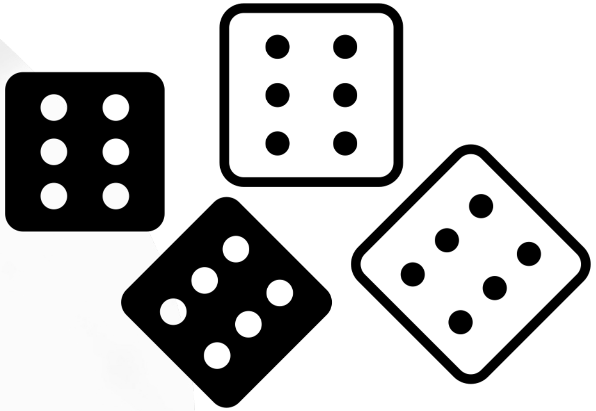
Often, estimators of relevant indicators are enough:

$$\widehat{\mu_i} = E[\mu_i] = \int \mu_i f(x|\mu) d\mu$$

$$\sigma_{\mu_i}^2 \equiv Var(\mu_i) = E\left[\mu_i^2\right] - E^2[\mu_i]$$

Why it works?

A deeper look at the dice



$\mathcal{L}(x|\mu) \rightarrow$ likelihood that the observation x are due to μ

$\mathcal{L}(x|\mu) \rightarrow f(x|\mu)$

General formula of Bayesian inference:

$$f(\mu|x) = \frac{f(x|\mu) f_0(\mu)}{\int f(x|\mu) f_0(\mu) d\mu}$$

Since the denominator is a normalization factor only, *assuming a uniform prior (at least in the region where $f(x|\mu)$ significantly differs from zero)* we have

$$f(\mu|x) \propto f(x|\mu) \equiv \mathcal{L}(x|\mu)$$

If the mean value of $f(\mu|x)$ corresponds to the value that maximizes $\mathcal{L}(x|\mu)$ the conventional results is recovered!

Combining results from different experiments

In conventional statistics, we combine results from different experiments through

$$\mu = \frac{\sum_i d_i/s_i^2}{\sum_i 1/s_i^2}$$

$$\sigma(\mu) = \left(\sum_i 1/s_i^2\right)^{-\frac{1}{2}}$$

Combining results from different experiments

In conventional statistics, we combine results from different experiments through

$$\mu = \frac{\sum_i d_i / s_i^2}{\sum_i 1/s_i^2}$$

$$\sigma(\mu) = \left(\sum_i 1/s_i^2\right)^{-\frac{1}{2}}$$

Let's assume we have to combine three measurements from counting experiments:

1. $4 \pm \sqrt{4}$
2. $7 \pm \sqrt{7}$
3. $10 \pm \sqrt{10}$

From previous equations the combinations gives $6.0 \pm \sqrt{1.4}$. However, if we think of the three experiments as a single one running for the triple of the time we get

$$21 \pm \sqrt{21} = 7.0 \pm 1.5$$

Combining results from different experiments

In conventional statistics, we combine results from different experiments through

$$\mu = \frac{\sum_i d_i/s_i^2}{\sum_i 1/s_i^2}$$

$$\sigma(\mu) = \left(\sum_i 1/s_i^2\right)^{-\frac{1}{2}}$$

By using Bayesian inference through the formula

$$f(\lambda, x) \propto e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i} f_0(\lambda)$$

that leads to the result 7.3 ± 1.6

Combining results from different experiments

In conventional statistics, we combine results from different experiments through

$$\mu = \frac{\sum_i d_i/s_i^2}{\sum_i 1/s_i^2}$$

$$\sigma(\mu) = \left(\sum_i 1/s_i^2\right)^{-\frac{1}{2}}$$

By using Bayesian inference through the formula

$$f(\lambda, x) \propto e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i} f_0(\lambda)$$

that leads to the result 7.3 ± 1.6

Standard approach fails because one of the three assumptions

1. *all the measurements refer to the same quantity*
2. *they are independent*
3. *d_i is distributed around μ as a Gaussian with $\sigma_i = s_i$*

does not hold. *Which one? And why?*

CSN3 Fisica Nucleare

How experiments end?

Istituto Nazionale di Fisica Nucleare
Laboratori Nazionali di Frascati

How the first neutral-current experiments ended

Peter Galison

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

At the beginning of the 1970s there seemed little reason to believe that strangeness-conserving neutral currents existed: theoreticians had no pressing need for them and several experiments suggested that they were suppressed if they were present at all. Indeed the two remarkable neutrino experiments that eventually led to their discovery were designed and built for very different purposes, including the search for the vector boson and the investigation of the parton model. In retrospect we know that certain gauge theories (notably the Weinberg-Salam model) predicted that neutral currents exist. But until 't Hooft and Veltman proved that such theories were renormalizable, little effort was made to test the new theories. After the proof the two experimental groups began to reorient their goals to settle an increasingly central issue of physics. Do neutral currents exist? We ask here: What kind of evidence and arguments persuaded the participants that they had before them a real effect and not an artifact of the apparatus? What eventually convinced them that their experiment was over? An answer to these questions requires an examination of the organization of the experiments, the nature of the apparatus, and the previous work of the experimentalists. Finally, some general observations are made about the recent evolution of experimental physics.

CONTENTS

Over the course of a year and a half—from the fall of 1972 to the spring of 1974—photographs such as Fig. 1 and Fig. 2 that at first appeared to be mere curiosities came to be seen as powerful evidence for the existence of weak neutral currents. Slowly, the experimentalists embedded these photographs in a persuasive demonstration based on a variety of technical, theoretical, and experimental advances. In so doing they presented the physics community with one of the most significant discoveries of recent physics. The subsequent developments in gauge theories and tests of the standard model are well known. But how did the experimentalists themselves come to believe in this result? What persuaded them that they were looking at a real effect and not at an artifact of the machine or the environment?

To understand how the evidence became convincing to the experimentalists we shall need to situate the experiment in the context in which it was planned and built. We need to know something of the experimental and theoretical assumptions held by the physicists involved. Finally, we must trace not only the positive results obtained, but also the myriad of false leads and technical difficulties that arose in the course of the work. In this sense the study will be historical, unlike the excellent and comprehensive review articles that have appeared such as Baltay (1979), Cline and Fry (1977), Cundy (1974), Faissner (1979), Kim *et al.* (1981), Mann (1977), Myatt (1974), and Rousset (1974).

I. INTRODUCTION

The blessing and curse of Fermi's (1934) theory of beta decay was that it skirted the fundamental dynamics of the weak interaction. Certainly the provisional advance afforded by such a move had many historical precedents. A hundred years earlier Ampère unravelled many of the laws of electrodynamics by studying the interactions of electrical currents. Even in the absence of Maxwellian theory much could be learned. Facing the largely unexplored weak interaction, Fermi drew explicitly on the ideas of quantum-electrodynamic currents for his theory. Just as an electron can produce a photon, Fermi reasoned, so could a nucleon emit the light electron and neutrino. The salient difference between electrodynamic and weak currents was this: While an electron retained its charge during the emissive process the nucleon did not—it changed from a neutron to a proton. [See Figs. 3(a) and 3(b).]

Subsequently currents without a change of heavy particle charge were dubbed "neutral" and those with such a change, "charged." For over 30 years after Fermi's paper

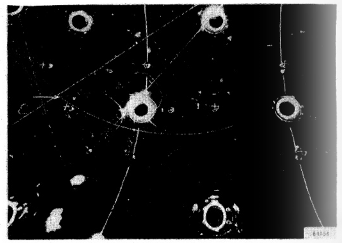


FIG. 1. Neutral-current event. Bubble-chamber photographs from Gargamelle resembling and including this one were at first mistakenly classified as neutron stars. (These are events in which a neutron—putatively at the arrow's end—collided with a nucleus to create a right-moving shower of particles.) Later many of these events were understood to be neutral-current events in which an unseen right-moving neutrino scattered elastically from a quark, creating a right-moving hadronic shower.

assumptions, the experimental technique, and the individual styles of research. When looked at in this way, contemporary experiments suggest that the processes of discovery and justification lose much of their distinct identities.

We need a richer descriptive vocabulary to describe experimentation in a way that will account for the many intermediate steps between the often very subjective working hypotheses of various participants and the logically or empirically based argument that eventually finds its way to publication. Such a vocabulary would be able to depict the degrees of persuasive force that evidence has as it begins to accumulate from diverse considerations. In the process of developing an account like this, we shall come to understand how data is gradually transformed (as in the case of the first muonless-event photograph) from a collection of curiosities to the foundation of a compelling demonstration.

ACKNOWLEDGMENTS

P. Galison, *Reviews of Modern Physics, Vol. 55, No. 2, April 1983*

To bet or not to bet

4. **Odds in betting.** *The higher the degree of belief that an event will occur, the higher the amount of money A that someone ("a rational better") is ready to pay in order to receive a sum of money B if the event occurs.*

Suggested readings

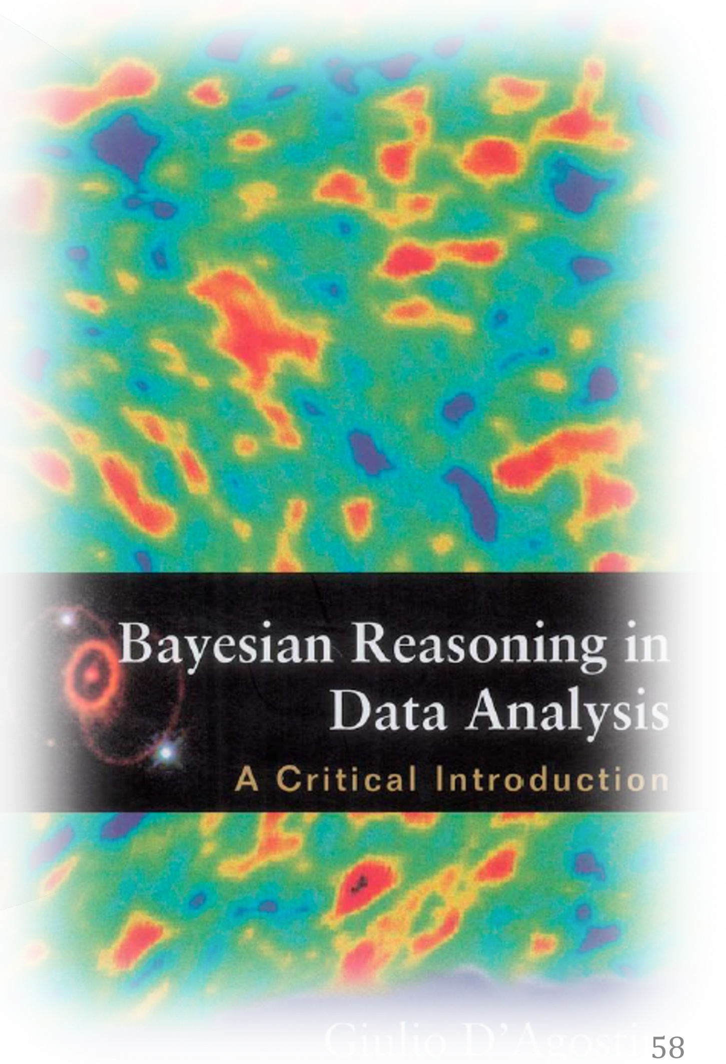
1. «***Bayesian Reasoning in Data Analysis***», **G. D'Agostini** (main reference for the preparation of this lecture)

2. «*How the first neutral-current experiments ended?*», P. Galison, *Reviews of Modern Physics, Vol. 55, No. 2, April 1983*

3. «*Frequentist and Bayesian confidence intervals*», G. Zech, *10.1007/s1010502c0012*

4. «*Why Isn't Everyone a Baysian?*», B. Efron, *The American Statistician, February 1986, Vol. 40, No. 1*

5. «*Beyond the Standard Model Physics at HERA*», C. Diaconu, *Nuclear Physics B (Proc. Suppl.) 66 (1998) 120-123*



backup

CSN3 Fisica Nucleare

Conditional probability

Istituto Nazionale di Fisica Nucleare
Laboratori Nazionali di Frascati

3.5.2 *Conditional probability*

Although everybody knows the formula of conditional probability, it is useful to derive it here in a kind of "standard way". A derivation closer to subjectivist spirit will be given in Sec. 10.3, where the meaning of the resulting formula will be described in more detail.

The notation is $P(E\,|\,H)$, to be read "probability of E given H", where H stands for *hypothesis*. This means: the probability that E will occur under the hypothesis that H has occurred[4].

The event $E\,|\,H$ can have three values:

TRUE: if E is TRUE and H is TRUE;
FALSE: if E is FALSE and H is TRUE;
UNDETERMINED: if H is FALSE; in this case we are merely uninterested in what happens to E. In terms of betting, the bet is invalidated and none loses or gains.

Then $P(E)$ can be written $P(E\,|\,\Omega)$, to state explicitly that it is the probability of E whatever happens to the rest of the world (Ω means all possible events). We realize immediately that this condition is really too vague and nobody would bet a cent on such a statement. The reason for usually writing $P(E)$ is that many conditions are implicitly, and reasonably, assumed in most circumstances. In the classical problems of coins and dice, for example, one assumes that they are regular. In the example of the energy loss of the previous section it was implicit ("obvious") that the high voltage was on (at which voltage?) and that the accelerator was operational (under which condition?). But one has to take care: many riddles are based on the fact that one tries to find a solution which is valid under stricter conditions than those explicitly stated in the question [53], and many people make bad business deals by signing contracts in which what "was obvious" was not explicitly stated (or precisely the contrary was stated explicitly, but in 'small print', as in insurance policies...).

- Equations (3.13)–(3.14) show explicitly how the probability of a certain hypothesis is updated when the *state of information* changes:

  $\boxed{P(H_i\,|\,H_\circ)}$ [also indicated as $P_\circ(H_i)$] is the *initial*, or *a priori*, probability (or simply *'prior'*) of H_i, i.e. the probability of this hypothesis with the state of information available 'before' the knowledge that E has occurred;

  $\boxed{P(H_i\,|\,E, H_\circ)}$ [or simply $P(H_i\,|\,E)$] is the *final*, or *'a posteriori'*, probability of H_i 'after' the new information.

  $\boxed{P(E\,|\,H_i, H_\circ)}$ [or simply $P(E\,|\,H_i)$] is called *likelihood*.

  Note that 'before' and 'after' do not really necessarily imply time ordering, but only the consideration or not of the new piece of information.

To better understand the terms 'initial', 'final' and 'likelihood', let us formulate the problem in a way closer to the physicist's mentality, referring to *causes* and *effects*: 'causes' are all the physical sources capable ofproding a given *observable* (the effect). The 'likelihood' indicates — as the word suggests — *"the likelihood that a cause will produce a given effect"* (not to be confused with "the likelihood that an effect is due to a given cause" which has a different meaning: A 'likelihood' may be arbitrarily small, but in spite of this, it is certain that an effect is due to a given cause, if there are no other causes capable of producing that effect!).

Using our example of the energy loss measurement again, the causes are all the possible charged particles which can pass through the detector; the effect is the amount of observed ionization; the likelihoods are the probabilities that each of the particles give that amount of ionization. Note that

Some standard statistics

Discrete variables: text

Continuous variables: f(x)dx, altrimenti sarebbe 0 (Achille e la tartaruga)