Statistics, Distributions and Probability

Ritesh K. Singh

Indian Institute of Science Education & Research Kolkata
Mohanpur Campus, Nadia

SERC Prep School, IISER Bhopal

in collaboration with

Partha Konar
Physical Research Laboratory
Ahmedabad

Satyaki Bhattacharya
Saha Institute of Nuclear Physics
Kolkata

Why Statistics?

What is Statistics?

- As a discipline
- As a quantifier
- Data Modelling

Probability

- Histograms as Probability Distribution
- Probability Theory

Need for Statistics

It is often said that the language of science is mathematics. It could well be said that the language of experimental science is statistics. It is through statistical concepts that we quantify the correspondence between theoretical predictions and experimental observations.

Kyle Cranmer, `1503.07622`

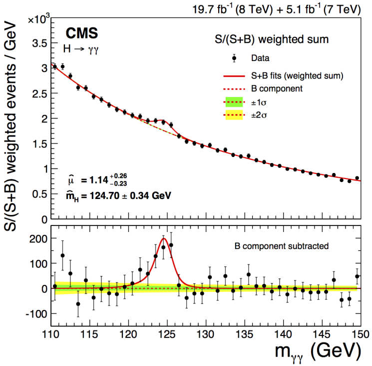
Need for Statistics

... we quantify the correspondence between theoretical predictions and experimental observations.

Positive outcome

Peak in invariant mass distribution of two photons $\Rightarrow$

Higgs mass (in GeV) = $125.09 \pm 0.21(stat.) \pm 0.11(syst.)$ (ATLAS + CMS)



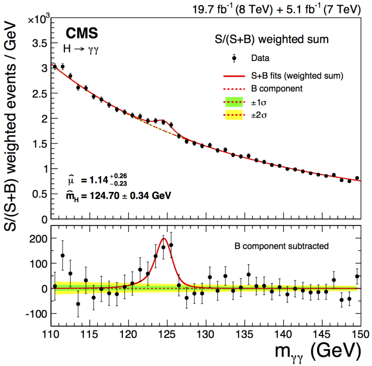
Need for Statistics

... we quantify the correspondence between theoretical predictions and experimental observations.

Positive outcome

Peak in invariant mass distribution of two photons $\Rightarrow$

Higgs mass (in GeV) = $125.09 \pm 0.21(stat.) \pm 0.11(syst.)$ (ATLAS + CMS)



- Parameter estimation
- Hypothesis testing

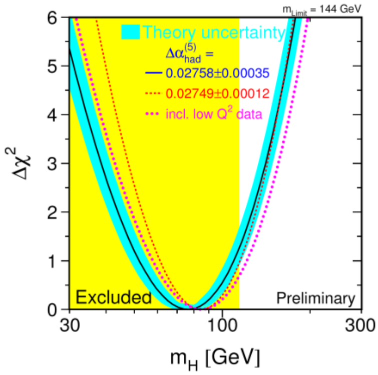
Need for Statistics

... we quantify the correspondence between theoretical predictions and experimental observations.

Null outcome

No conclusive evidence of signals of Higgs boson at LEP ⇒

$m_H > 114.3$ GeV at 95% C.L.
Best fit for m_H
Upper limit on m_H (SM like)



Need for Statistics

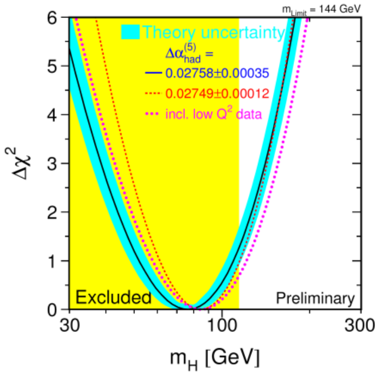
... we quantify the correspondence between theoretical predictions and experimental observations.

Null outcome

No conclusive evidence of signals of Higgs boson at LEP $\Rightarrow$

$m_H > 114.3$ GeV at 95% C.L.
Best fit for m_H
Upper limit on m_H (SM like)

- Parameter limitation
- Hypothesis testing



What is Statistics?

Statistics is a discipline which is concerned with:

What is Statistics?

Statistics is a discipline which is concerned with:

- designing experiments and other data collection (like simulations),

What is Statistics?

Statistics is a discipline which is concerned with:

- designing experiments and other data collection (like simulations),
- summarizing information to aid understanding (in terms of quantifiers, estimators),

What is Statistics?

Statistics is a discipline which is concerned with:

- designing experiments and other data collection (like simulations),
- summarizing information to aid understanding (in terms of quantifiers, estimators),
- drawing conclusions from data (translating quantifiers into model parameters),

What is Statistics?

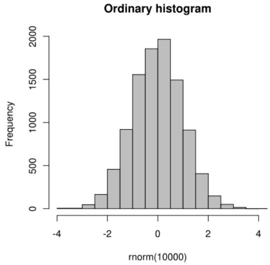
Statistics is a discipline which is concerned with:

- designing experiments and other data collection (like simulations),
- summarizing information to aid understanding (in terms of quantifiers, estimators),
- drawing conclusions from data (translating quantifiers into model parameters),
- estimating the present or predicting the future (post-diction) or (pre-diction)

Statistics ≡ quantifiers

Let $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$ be a large set of data.

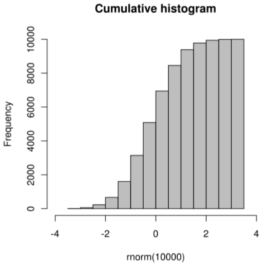
- Histogram, frequency distribution



Statistics ≡ quantifiers

Let $\mathcal{D} = \{x_1, x_2, \dots, x_N\}$ be a large set of data.

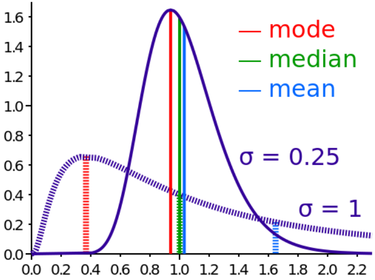
- Histogram, frequency distribution
- Cumulative distribution,



Statistics ≡ quantifiers

Let $\mathcal{D} = \{x_1, x_2, \ \dots \ , x_N\}$ be a large set of data.

- Histogram, frequency distribution
- Cumulative distribution,
- Median center of distribution
- Mode, most frequent
- Mean, center of mass



Moments of distribution

Mean

Center of mass for a given data set

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Mean for a given histogram

$$\langle x \rangle = \left(\sum_{i=1}^{H} w_i \right)^{-1} \sum_{i=1}^{H} w_i \ x_i$$

Here H is number of histograms and w_i is the frequency in the i^{th} histogram.

Higher Moments

For a given data set n^{th} moment is:

$$\langle x^n \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i^n$$

and for a given histogram is:

$$\langle x^n \rangle = \left(\sum_{i=1}^{H} w_i \right)^{-1} \sum_{i=1}^{H} w_i \ x_i^n$$

The mean is the first ($n = 1$) moment. Other important quantifier is related to $\langle x^2 \rangle$.

Moments of distribution

Variance

Mean or average, $\langle x \rangle$ is a measure of center of the data. Variance measures the spread of data around the mean. Since we have:

$$\langle (x - \langle x \rangle) \rangle = 0$$

The variance is defines as

$$V = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$$

Higher Moments

Show:

- $\langle (x - \langle x \rangle)^3 \rangle = \langle x^3 \rangle + 2\langle x \rangle^3 - 3\langle x \rangle \langle x^2 \rangle$
  Skewness
- $\langle (x - \langle x \rangle)^4 \rangle = \langle x^4 \rangle - 3\langle x \rangle^4 + 6\langle x^2 \rangle \langle x \rangle^2 - 4\langle x \rangle \langle x^3 \rangle$
  Kurtosis

Show these relations numerically during Tutorials.

Data Modelling $\equiv$ Data Compression

Data Modelling ≡ Data Compression

- We have $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$, a large set of data points.

Data Modelling ≡ Data Compression

- We have $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$, a large set of data points.
- Chosing H number of bins, the data set can be converted into a histogram (frequency distribution). Now a set of N data points are represented or modelled by set of $2H$ number, H for the bin centers and H for frequencies. This is the first step of modelling, where we replace a large set of data by its frequency distribution.

Data Modelling ≡ Data Compression

- We have $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$, a large set of data points.
- Chosing H number of bins, the data set can be converted into a histogram (frequency distribution). Now a set of N data points are represented or modelled by set of $2H$ number, H for the bin centers and H for frequencies. This is the first step of modelling, where we replace a large set of data by its frequency distribution.
- Next level modelling involves finding a functional form of the frequency distribution. Then we can replace the $2H$ numbers describing the histogram with few moments $\langle x^n \rangle$, which are parameters of its functional approximation.

Data Modelling ≡ Data Compression

- We have $\mathcal{D} = \{x_1, x_2, \ \ldots \ , x_N\}$, a large set of data points.
- Chosing H number of bins, the data set can be converted into a histogram (frequency distribution). Now a set of N data points are represented or modelled by set of $2H$ number, H for the bin centers and H for frequencies. This is the first step of modelling, where we replace a large set of data by its frequency distribution.
- Next level modelling involves finding a functional form of the frequency distribution. Then we can replace the $2H$ numbers describing the histogram with few moments $\langle x^n \rangle$, which are parameters of its functional approximation.
- Example: If the shape of the histogram is Gaussian, it is parametrized with the mean and the variance.

$$\mathcal{D} \Rightarrow \langle x \rangle \pm \sqrt{V} = \langle x \rangle \pm \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

Histogram ≡ Probabilities

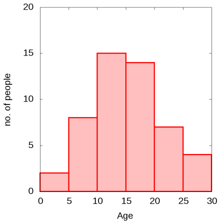
A group of 50 people has following age distribution

Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

Histogram ≡ Probabilities

A group of 50 people has following age distribution

Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

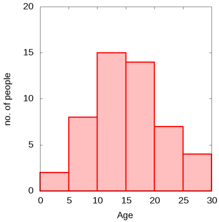


Histogram ≡ Probabilities

A group of 50 people has following age distribution

Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

Randomly choose a person



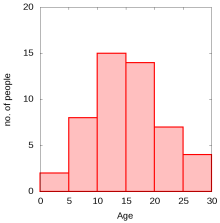
Histogram ≡ Probabilities

A group of 50 people has following age distribution

Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

Randomly choose a person

- What is the probability of age below 15 years?



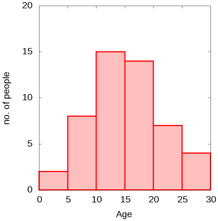
Histogram ≡ Probabilities

A group of 50 people has following age distribution

Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

Randomly choose a person

- What is the probability of age below 15 years?
- Probability of age above 25 year?



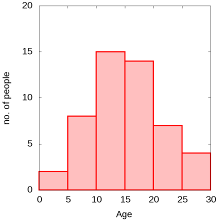
Histogram ≡ Probabilities

A group of 50 people has following age distribution

Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

Randomly choose a person

- What is the probability of age below 15 years?
- Probability of age above 25 year?
- Probability of age between 10 and 20 years?



Histogram ≡ Probabilities

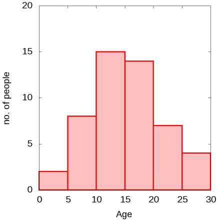
A group of 50 people has following age distribution

Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

Randomly choose a person

- What is the probability of age below 15 years?
- Probability of age above 25 year?
- Probability of age between 10 and 20 years?

Ans: 50%, 8%, 58%



Histogram ≡ Probabilities

A group of 50 people has following age distribution

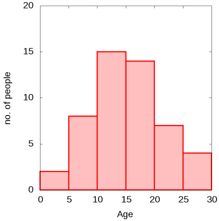
Age	#
0 – 5	2
5 – 10	8
10 – 15	15
15 – 20	14
20 – 25	7
25 – 30	4

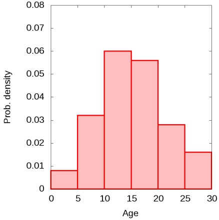
Randomly choose a person

- What is the probability of age below 15 years?
- Probability of age above 25 year?
- Probability of age between 10 and 20 years?

Ans: 50%, 8%, 58%

Probability density = probability/(bin width)





Probability Density Functions

- If $p(x)$ is the PDF of x then $p(x)dx$ is the probability of finding the value between x and $x + dx$.
- $\int_R dx\ p(x) = 1$, R is the range of variable x.
- $p(x) \geq 0$ for $x \in R$

Uniform (x, a, b), $x \in [a, b]$

$$\frac{1}{b-a}$$

Exponential (x, a), $x \in [0, \infty]$

$$\frac{1}{a} \exp(-x/a)$$

Gaussian (x, μ, σ), $x \in [-\infty, \infty]$

$$\frac{1}{\sqrt{2\pi\sigma}} \exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$
  $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

- Probability of A or B is $P(A \cup B)$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

- Probability of A or B is $P(A \cup B)$
- Probability of A and B is $P(A \cap B)$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

- Probability of A or B is $P(A \cup B)$
- Probability of A and B is $P(A \cap B)$
- $P(\bar{A}) = 1 - P(A)$, $\bar{A}$ is complement of A w.r.t. S

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

- Probability of A or B is $P(A \cup B)$
- Probability of A and B is $P(A \cap B)$
- $P(\bar{A}) = 1 - P(A)$, $\bar{A}$ is complement of A w.r.t. S
- $P(A \cup \bar{A}) = 1$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

- Probability of A or B is $P(A \cup B)$
- Probability of A and B is $P(A \cap B)$
- $P(\bar{A}) = 1 - P(A)$, $\bar{A}$ is complement of A w.r.t. S
- $P(A \cup \bar{A}) = 1$
- $P(\phi) = 0$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

- Probability of A or B is $P(A \cup B)$
- Probability of A and B is $P(A \cap B)$
- $P(\bar{A}) = 1 - P(A)$, $\bar{A}$ is complement of A w.r.t. S
- $P(A \cup \bar{A}) = 1$
- $P(\phi) = 0$
- If $A \subset B$, then $P(A) \leq P(B)$

Kolmogorov axiom

Let S is the set of events. Let A and B be the subsets of S.

- For each $A \subset S$, there is a real-valued function P such that $P(A) \geq 0$.
- $P(S) = 1$
- For $A \cap B = \phi$ $P(A \cup B) = P(A) + P(B)$

These are Kolmogorov axioms for the function P to be probability.

Many other properties of probability can be deduced from these.

- Probability of A or B is $P(A \cup B)$
- Probability of A and B is $P(A \cap B)$
- $P(\bar{A}) = 1 - P(A)$, $\bar{A}$ is complement of A w.r.t. S
- $P(A \cup \bar{A}) = 1$
- $P(\phi) = 0$
- If $A \subset B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional Probability

If A and B are independent then probability of A and B

$$P(A \cap B) = P(A)\,P(B)$$

Conditional Probability

If A and B are independent then probability of A and B

$$P(A \cap B) = P(A)\, P(B)$$

- Month = April (say A) and year being even (say B) is independent of each other, since there is April in every year.

Conditional Probability

If A and B are independent then probability of A and B

$$P(A \cap B) = P(A)\, P(B)$$

- Month = April (say A) and year being even (say B) is independent of each other, since there is April in every year.
- 29 February (A) and even/odd-ness of year (B) is not independent.

Conditional Probability

If A and B are independent then probability of A and B

$$P(A \cap B) = P(A)\,P(B)$$

- Month = April (say A) and year being even (say B) is independent of each other, since there is April in every year.
- 29 February (A) and even/odd-ness of year (B) is not independent.

Conditional probability of A given B (with $P(B) \neq 0$) is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional Probability

If A and B are independent then probability of A and B

$$P(A \cap B) = P(A)\ P(B)$$

- Month = April (say A) and year being even (say B) is independent of each other, since there is April in every year.
- 29 February (A) and even/odd-ness of year (B) is not independent.

Conditional probability of A given B (with $P(B) \neq 0$) is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

If A and B are independent

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

i.e. When A and B are independent, then the conditional probability of A given B does not depend upon B at all.

Bayes' Theorem

Probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

and probability of B given A is

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Since $P(A \cap B) = P(B \cap A)$

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

Let's assume S is composed of disjoint subsets A_i, thus $S = \cup_i A_i$
Further

$$B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$$

$$\begin{aligned} P(B) &= \sum_i P(B \cap A_i) \\ &= \sum_i P(B|A_i) P(A_i) \end{aligned}$$

$$P(A|B) = \frac{P(B|A)\ P(A)}{\sum_i P(B|A_i) P(A_i)}$$

An example of Bayes' Theorem

The probability of having AIDS(prior to any tests) for a given population

$P(\text{AIDS}) = 0.001$
$P(\text{no AIDS}) = 0.999$

Consider a test with results + or − and conditional probabilities as

$P(+|\text{AIDS}) = 0.98$
$P(-|\text{AIDS}) = 0.02$

$P(+|\text{no AIDS}) = 0.03$
$P(-|\text{no AIDS}) = 0.98$

Follwing questions can be asked and answered:

- What is the probability for find $+(-)$ result for the population?
  $0.03095\ (0.96905)$
- If the result is $+$, what is the likelihood of person suffring from AIDS?
  0.03166
- If the result is $-$, what is the likelihood of person still suffring from AIDS?
  2.06×10^{-5}