Statistical methods and error analysis

3rd Alice-India School 2020

Satyaki Bhattacharya
Saha Institute of Nuclear Physics

Outlook of the lectures

Statistics is a tool for doing physics. A good physicist understands their tools. Don't just follow without understanding, but read books and conference proceedings, go to seminars, talk to people, experiment with the data, and understand what you are doing. Then you will succeed. And you will have a great time!

Concluding remarks of Roger Barlow,
Asia Pacific School,Vietnam, 2017,
arXiv 1905.12362v1

Introduction

Consider production of Z bosons in high energy proton-proton collisions and Z decaying to $\mu^+\mu^-$. The partonic level reaction is

$$q\bar{q} \rightarrow Z \rightarrow \mu^+\mu-$$

If one measures the mass from the dimuon 4-momenta for every "event" a different mass will be obtained, even if the experimental error is negligible. This follows from quantum mechanics that the mass of each individual Z boson is a random number (and equals the $\hat{s}$ of $q\bar{q}$) distributed according to Breit-Wigner distribution. The PDG value of Z **mass** and width Γ are the mean and full width at half maximum (FWHM) of the Breit-Wigner distribution.

Introduction 2

In a real detector, of course, there will be errors due to stochastic processes involved in the detection process (e.g. the amount of charge produced in a silicon sensor by the muons). This will "smear" the Breit-Wigner distribution. The observed distribution will be a "convolution" of the Breit-Wigner with another distribution describing the random fluctuations in the measurement of Z mass. In addition to the random fluctuations there will be numerous sources of "systematic uncertainties" in the measurements **(e.g.)**.

The estimation of Z mass "fitting" the data to extract the mean and width of the Breit-Wigner and careful estimation of the 'systematic' and 'statistical uncertainties'. The measurement of Z mass is a typical example of point estimation or parameter estimation. We will start from a precise definition of "event" and probability in statistics and then discuss the basics of parameter estimation.

- From the dimuon mass can we conclude we have seen a new particle?
- How do we estimate the uncertainties in Z mass measurement?

Topics to be covered

- Probability, distribution functions, properties
- Point estimation
- Interval estimation
- Hypothesis testing
- Introduction to Monte Carlo methods

References

[1] R J Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley, 1989

[2] G Cowan, *Statistical Data Analysis*, Clarendon Press, 1998

[3] I Narsky and F C Porter, *Statistical Analysis Techniques in Particle Physics*, Wiley, 2014

[4] O Behnke *et al* (Eds.) *Data Analysis in High Energy Phyusics*, Wiley, 2013

[5] L Lyons, *Statistics for Nuclear and Particle Physicists*, Cambridge, 1992

[6] G Bohm and G Zech, *Introduction to Statistics and Data Analysis for Physicists*. DESY, 2010

[7] M R Whalley and L Lhyons (Eds) *Advanced Statistical Techniques in Particle Physics*, Durham report IPPP/02/39, 2002

[8] L Lyons, R Mount and R Reitmayer (Eds) *Proceedings of PHYSTAT03*, SLAC-R-703, 2003

[9] L Lyons and M K Unel (Eds.) *Proceedinss of PHYSTAT05*, Imperial College Press, 2006

[10] R von Mises, *Probability, Statistics and Truth*, Dover, 1957

[11] H Jeffreys, *Theory of Probability*, 3rd Edition, Oxford, 1961

[12] R J Barlow. *A note on* $\Delta \ln L = -\frac{1}{2}$ *Errors*, arXiv:physics/0403046 , 2004

[13] R J Barlow, *Asymmetric Statistical Errors* , arXiv:physics/0406120, 2004

[14] Taken from `https://root.cern.ch/cms-2d-cf-cv-likelihood-profile`, 26/5/19

[15] P R Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw Hill, 1969

Sample space

Definition 1: The set, S, of all possible outcomes of a particular experiment is called sample space of the experiment.

Consider a coin toss experiment. The sample space (S) consists of two outcomes, head (H) and tail (T). For a pair of coin tosses, we have,

$$S = (HH,\, TT,\, TH,\, HT)$$

In the above cases S is "countable". For our Z boson production

$$S = (\,0\,GeV,\, \infty\,GeV\,)$$

event

Definition 2: Event, an event is any collection of possible outcomes of an experiment, i..e. any subset of S (including S itself).

Definition 3: Complementation of A, written as A^c, is the set of all elements that are not in A.

$$A^c = \{x : x \notin A\}$$

Example: event,complement

Event operations: Consider an electron (or a positron) hitting the electron magnetic calorimeter. We are interested in the charge (+ or -) and region of hit (barrel (b) or endcap (ϵ)). The sample space is

$$S = \{b^+, b^-, \epsilon^+, \epsilon^-\}$$

Some events are $A = \{b^+, b^-\}$, $B = \{b^-, \epsilon^+, \epsilon^-\}$

Then the union ($\cup$) and intersection ($\cap$) of A and B are

$$\begin{aligned} & A \cup B = \{b^+, b^-, \epsilon, \epsilon^-\} \\ & A \cap B = \{b^-\} \\ \text{Complement of } A, \quad & A^c = \{\epsilon^+, \epsilon^-\} \end{aligned}$$

Properties of events

Theorem 1: For any three events A, B and C defined on a sample space S,

a. Commutativity: $A \cup B = B \cup A\,, \quad A \cap B = B \cap A$

b. Associativity: $A \cup (B \cup C) = (A \cup B) \cup C\,, \quad A \cap (B \cap C) = (A \cap B) \cap C$

c. Distributive laws: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)\,, \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

d. DeMorgan's laws: $(A \cup B)^c = A^c \cap B^c\,, \quad (A \cap B)^c = A^c \cup B^c$

- Prove the above statements

Definition 4: Events A_1, A_2, ... are disjoint or mutually exclusive if $A_i \cap A_j = \phi \quad \forall \quad i, j$

Definition 5: If A_1, A_2, ... are pairwise disjoint on $\cup_{i=0}^{\infty} A_i = S$ then the collection A_1, A_2, ... form a **partition** of S.

Axiomatic foundation (Kolmogorov)

Definition 7: Given a sample space S and an associated sigma algebra $\mathbb{B}$, a probability function is a function P with domain $\mathbb{B}$ that satisfies:

1. $P(A) \geq 0 \quad \forall\, A \in \mathbb{B}$
2. $P(S) = 1$
3. if $A_1, A_2, ... \in \mathbb{B}$ are pairwise disjoint, then $P(\Sigma_{i=0}^{\infty} \cup A_i) = \Sigma_{i=0}^{\infty} P(A_i)$
4. Axiom of finite additivity: If $A \in B$ and $B \in \mathbb{B}$ are disjoint, then,
   $P(A \cup B) = P(A) + P(B)$
   (A more general axiom is the axiom of countable additivity.)

For the sake of completeness we introduced the concept of sigma algebra, but it is a technicality that we will not have to worry about. For all finite or countable S, $\mathbb{B}$ is just $\{$all subsets of S, including $S\}$.

Properties of probability

Theorem 2: If P is a probability function and A is any set in $\mathbb{B}$, then,

a) $P(\phi) = 0$, where ϕ is the empty set.

b) $P(A) \leq 1$

c) $P(A^c) = 1 - P(A)$

Theorem 3: If P is a probability function, then

a) $P(A) = \Sigma_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2 ...$

b) $P(\cup_{i=1}^{\infty} A_i) \leq \Sigma_{i=1}^{\infty} P(A_i)$ for sets $A_1, A_2, ...$ (Boole's inequality)

Real probability

Also known as Classical probability, this was developed during the 18th-19th century by Pascal, Laplace and others to serve the gambling industry.

If there are several possible outcomes and there is a symmetry between them so they are all, in a sense, identical, then their individual probabilities must be equal. For example, there are two sides to a coin, so if you toss is there must be a probability $\frac{1}{2}$ for each face to land uppermost. Likewise there are 52 cards in a pack, so the probability of a particular card being chosen is $\frac{1}{52}$. In the same way there are 6 sides to a dice, and 33 slots in a roulette wheel,

This enables you to answer questions like 'What is the probability of rolling more than 10 with 2 dice?'. There are 3 such combinations (5-6, 6-5 and 6-6) out of the $6 \times 6 = 36$ total possibilities, so the probability is $\frac{1}{12}$. Compound instances of A are broken down into smaller instances to which the symmetry argument can be applied. This is satisfactory and clearly applicable - you know that if someone offers you a 10 to 1 bet on this dice throw, you should refuse; in the long run knowledge of the correct probabilities will pay off.

The problem arises that this approach cannot be applied to continuous variables. This is brought out in Bertan's paradoxes, one of which runs:

In a circle of radius R an equilateral triangle is drawn. A chord is drawn at random. What is the probability that the length of the chord is greater than the side of the triangle?

Barlow, Asia Pacific School,Vietnam, 2017,
arXiv 1905.12362v1

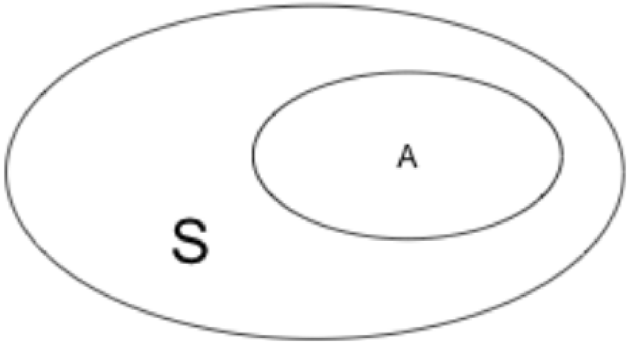
Frequentist Probability

2.4 Frequentist Probability

Because of such difficulties, Real Probability was replaced by Frequentist Probability in the early 20th century, This is the usual definition taught in schools and undergraduate classes. A very readable account is given by von Mises [10].

$$P_A = \lim_{N\to\infty} \frac{N_A}{N}$$

N is the total number of events in the ensemble (or collective). It can be visualised as a Venn diagram, as in Figure 2.



Conditional probability

A familiar example of conditional probability is dice throws. Suppose you have thrown a dice 3 times and got a 6 each time. What is the probability of getting a 6 in the next throw given that you have got three 6's in a three previous throws.

The answer is it's still $\frac{1}{6}$ since the events are independent or the answer is the probability of getting four 6's in a row is $(\frac{1}{6})^4$, so the probability is $(\frac{1}{6})^4$ for getting a 6 in a next throw.

Definition 8: If A and B are events in S, and $P(B)$ 0 then the conditional probability of A given B, written $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

Conditional probability example

Example application of conditional probability:

Draw four cards from top of a deck of cards. What is the probability that all four are aces?

The answer is $\frac{1}{{}^{52}C_4} = \frac{1}{270725}$

But in a alternative way,

Probability of 1^{st} card to be ace is $P(ace) = \frac{4}{52}$

Probability of 2^{nd} card to be ace given first is ace is $P(ace|1^{st}\, is\, ace) = \frac{3}{51}$ and so on.

$\implies P(all\, four\, ace) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49} = \frac{1}{270725}$

Bayes Theorem

Re-expressing (1) we have,

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

Using symmetry,

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

This is often called **Bayes Theorem**.

Theorem 4: Bayes Rule

Let $A_1, A_2, ...$ be a partition of the sample space and let B be any set. Then, for each $i = 1, 2, ...$

$$P(A_i|B) = \frac{\phi(B|A_i)P(A_i)}{\Sigma_{j=0}^{\infty} P(B|A_j)P(A_j)}$$

Example: Suppose we have 30% electron contamination in a pion beam in a calorimeter test beam run, i.e. $P(e) = \frac{3}{10}$ $P(\pi) = \frac{7}{10}$ in beam. Due to reconstruction error, there is a 5% chance that a prion gets reconstructed as electron and vice-versa. If in an event a pron has been detected, what is the probability that truly a pion hit the detector?

Solution: Problem 1.44 (Casella-Berger).

Statistical Independence

Definition 9: Statistical Independece
A collection of events $A_1, A_2, ..., A_n$ are mutually independent if for any sub-collection $A_{i1}, ...A_{ik}$ we have,

$$P(\cap_{j=1}^{k} A_{ij}) = \Pi_{j=1}^{k} P(A_{ij})$$

Example: Throw of two die, toss of three coins etc.

Exercise: Consider events H_1, H_2, H_3 in a three coin toss experiment,

$$H_1 = all\, cases\, where\, the\, first\, coin\, lands\; H = \{HTT, HTH, HHT, HHH\}$$

$$H_2 = all\, cases\, where\, 2^{nd}\, coin\, lands\quad H\; and\, so\, on.$$

Calculate $P(H_1 \cap H_2 \cap H_3)$.

Random variables

Definition 10: A random variable is a function from a sample space S into real numbers.

Examples of random variables:

Experiment	**Random variable**
$e^+e^- \longrightarrow e^+e^-$ collision	p, θ, ϕ
A set of 1 million such collisions	average $p = \hat{p}, \hat{\theta}, \hat{\phi}$ over the sample of 1 million

In experiment rather than dealing with each event we must often use quantities like average or average spread over the sample. To summarize the probability structure of the data all such data summary variables are random variables — they change "unpredictably" as we replace one sample with another statistically independent sample coming from a repeat of the experiment.

Notation: We will denote random variables with upper case letter like X and its realized value in an experiment with the corresponding lower case letters like x.

Distribution of random variables

Let's consider three radioactive nuclei that are being observed for 5 seconds. Let the random variable X be the number of decays occuring in this interval.

All possible cases are

$S:$	111	110	101	100	011	010	001	000
$X(S):$	3	2	2	1	2	1	1	0

(1 denotes nucleus has decayed; 0 denotes nucleus has not decayed)

Range of X is $\chi = 0, 1, 2, 3$. These are the possible outcomes for X, so we have a new sample space χ.Let the probability of decay within 5 seconds be p. Then we can define a induced probability function P_χ as:

x	0	1	2	3
$P_\chi(X = x)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

Cumulative distribution function

In the example above, we can also calculate the probability that $X \leq 0$ or 1 or 2 etc.

$P_X(X \leq 0)$	$P_X(X \leq 0)$	$P_X(X \leq 0)$	$P_X(X \leq 0)$
$(1-p)^3$	$3p(1-p)^2$ $(1-p)^3$	$3p^2(1-p)$ $3p(1-p)^2$ $(1-p)^3$	p^3 $3p^2(1-p)$ $3p(1-p)^2$ $(1-p)^3$
$^1/_8$	$^1/_2$	$^7/_8$	1

P = 1/2

Note that we can ask what is the probability that $X \leq 2.5$ and the answer is $P_X(X \leq 2.5) = 3p^2(1-p) + 3p(1-p)^2 + 91 - p)^3$ and that is in column 3 of the table above. We can calculate their comulative probability for every point on the real line.

Definition 11: Cumulative Distribution Function

The cumulative distribution function or cdf of a random variable X denoted by $F_X(x)$ is defined by,

$$F_X(x) = P_X(X \leq x), \quad \forall x$$

Mass and density functions

Definition 12: Probability Mass Function
The probability mass function of a discrete random variable X is given by,

$$f_X(x) = P(X = x) \quad \forall x$$

Continuing with the example of radioactive decays, if we have N nuclei in a sample, the probability that x of them will decay in 5 seconds is,

$$P(X = x) = {}^N C_x p^r (1-p)^{N-r}$$

This is the binomial probability mass function. We will be using the binomial pmf in many situations in particle experiments, because all particle physics experiments are essentially number counting experiments. N and p are called the parameters of the binomial distribution.

Definition 13: Probability Density Function The probability density function or pdf, $f_X(x)$ of a continuous random variable X is the function that satisfies,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \quad \forall x$$

Is this a distribution function?

Familiar examples are the exponential, normal (or Gaussian), Cauchy distributions.

As a simple demonstration of the properties of a cdf let us consider the function,

$$F_X(x) = \frac{1}{1+e^{-x}}$$

We see that,

$$\lim_{x\to-\infty} F_X(x) = 0 \qquad since, \quad \lim_{x\to-\infty} e^{-x} = \infty$$
$$\lim_{x\to\infty} F_X(x) = 1 \qquad since, \quad \lim_{x\to\infty} e^{-x} = 0$$

and,

$$\frac{dF_X(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} > 0$$

i.e. $F_X(x)$ is a monotonically increasing function.

This one is known as logistic distribution function

Independent but identically distributed (iid)

Let us consider a perfectly homogeneous radioactive sample consisting of $2N$ atoms. Let us divide it into 2 exact halves, A and B such that each piece contains exactly N atoms. Let X denotes the number of decays in A in 5 seconds and Y denotes the decays in B in 5 seconds. Quite clearly both X and Y are binomially distributed and for every $x = 0, 1, 2, 3...N$ we have,

$$P_X(X = x) = P_Y(Y = x)$$

but, X and Y are drawn from two different pieces and are obviously independent statistically. We will say, X and Y are *independent but identically distributed random variable* or *iid*.