Chapter 1

Lesson 1 - April 15th, 2024

Lesson delivered by Zanetti.

1.1 Introduction

The wavelength of the probe must be less than the size of what we're gonna probe. So:

$$\lambda = \frac{2\pi}{\hbar}$$

but also $\hbar k = p\frac{h}{\lambda}$ so we need high momentum to probe it better. The momentum is near E if the mass is much smaller.

Because of the Einstein relation between energy and mass, in order to excite resonances and produce particles, we need sufficient energy (threshold energy).

So we need higher and higher energies to study the particles and resonances.

1.2 Accelerators

We need particles and we need to accelerate them. The energy ΔE is increased by having the particle with charge q to go through a potential ΔV so:

$$\Delta E = q \cdot \Delta V$$

So we use very high voltages to accelerate it. The problem is that we cannot make a potential so big to do it in a single pass. We need a series of accelerating stages in order to achieve the requested energy. When passing from one stage to the next, the voltage is reversed in order to push the particle away and not back in. So the polarity changes at the right moment. This are called radiofrequency cavities and they pump in energy to accelerate the particles. Because the size of the resonators is in the order of the meter, the radiofrequency is in the order of hundreds of megahertz: $\nu \approx 100MHz$.

Now, we can either put many of these accelerating stages in a linear fashion or we can dot it in a circular fashion where the particles are recirculated through the various stages. So we have linear and circular accelerators.

Typical dimensions: in a cavity we have fields of $30MV/m$. With this $\Delta E \approx 30MeV$. So in order to arrive to $7TeV$ we would need an enormous $200km$ cavity which is of course unfeasible. So this is why we need a circular accelerator.

In order to do this we need magnets to bend the path of the particles and guide them through the accelerating stages again.

$$p[GeV] = 0.3R[m]B[T]$$

at LhC we have a radius $R = 3000m$ and the momentum $p = 7000GeV$ and the magnetic fields must be in the range of $B = 8T$ which is a lot. The LhC accelerating stage is actually just 20 meters long. All the other parts of the circle are just used to bend the path of the particles.

1.3 Main features of the colliders

It is important to know the type of particles and the so called luminosity. Typically one collides e^-, e^+, p^+, p^-, $p - p$, etc. Also, the collision doesn't really involve protons but the partons they are made of (so quarks).

An important feature is that of **luminosity**. The collisions are described in terms of probability, so we have cross sections. The theory gives us the probability for a given outcome given an input. Like we have input particles at a certain energy and we get the probability of observing the final state from the theory.

We are interested in rare processes. In order to get the chance to observe rarer and rarer events, one would want a machine capable of producing a sufficient number of events.

1.3.1 Cross section

The cross section is a surface, expressed in barns. 1 barn is $10^{-24}cm^2$. This probability must be translated into a rate of events R. This is related to the prob:

$$\boxed{R = \sigma\alpha}$$

where α is called **istantaneous luminosity** and it relates the rate R and the prob σ. The luminosity is a property of the machine itself (of the beam actually). Now, luminosity follows this formula:

$$\alpha \propto \nu \frac{nN^2}{A} \cdot f(\theta)$$

where ν is the radiofrequency, n is the particle bunches (these are group of particles that are all in phase, so that the accelerating cavities can kick them forward). N is the number of particles per bunch. So the beams must be highly focused at the interaction point where we have a transverse area A and f is a form factor. The transverse area is in the order of μm^2.

At LhC we have $n = 2800$, $N = 1.4 \cdot 10^{11}$ and $A \approx 15\mu m^2$. The luminosity is on the order of $\alpha \approx 2 \cdot 10^{34} cm^{-2} Hz$.

So this is why we need high luminosity: in this way we can compensate for the small values of the σ of the rarest processes and thus observe a decent rate for them.

1.3.2 Project about e^+, e^- - FCC

Now the e^+, e^- collider is way cleaner because these particles are much simpler since they are believed to be pointlike. So all the energy being given to the particles is given to the particles. In the case of the proton collider instead, a fraction of the energy is actually captured since the protons are bags of partons and thus the energy is split between them.

The center of mass system and the LAB system in the case of e^+, e^- coincide, which is a very nice feature. All the kinematic constrants can be used. All of this would be impossible in an hadronic machine like LhC.

So if we create, for instance, a Z-boson creation via e^+, e^- collision. And then the Z boson decays in μ and $\bar{\mu}$. On the other side there's something we don't know:

$$e^- + e^+ \to Z(\mu^+ \mu^-) + X$$

and X is fully unknown. But just by knowing the initial conditions and measuring the upper side of the events, we can infer the properties of the X unknown. We have in fact:

$$E_{tot} = 2E_b$$

and $p_{tot} = 0$. Total energy and momentum are conserved, where E_b is the energy of a single beam. At the end the total energy is $E = E_Z + E_X$ so we have $E_X = 2E_b - E_Z$, measuring the energy of the Z boson. The momentum of X is just the opposite of the measured momentum of the Z. We also know the mass since:

$$m_X^2 = E_x^2 - |\bar{p}_X|^2$$

and:

$$s = 4E_b^2$$

All of this is possible only with collisions with elementary particles.

Looking at the cross section, in this case we have that the dimension of σ is the inverse of the s Lorentz invariant (the s Mandelstam variable).

So we want both σ and s high and they contrast each other. It happens that the other dependencies of the cross section make it so there are other contributions to the σ so it's not a problem. At lower energies instead, the cross section goes down as $1/s$. So, the higher the energy, the smaller σ is. This is why the events are cleaner.

The reason why we also use hadronic machines, is that the charged particles radiates energy via synchrotron radiation. Thus they lose energy, when in circular motion as:

$$\Delta E_{lost} = \frac{4\pi}{3}\alpha^2\frac{\gamma^4}{R}$$

and γ is the Lorentz boost. At the LEP (before LhC) this ΔE was ($R = 3km$, $E_b = 104.5GeV$). In order to compensate for this energy loss, there were radiofrequency accelerators everywhere. In fact $\Delta E/E \approx 3.3\%$. So at every turn the RF needed to provide $3-4GeV$ because of how much energy they lost. So this fixes an upper limit on the feasibility of such accelerators.

When it comes to luminosity of an $e^+.e^-$ machine, there's a problem. Antimatter is scarce. So the number of bunches will be lower (at LEP it was just 10) and $\alpha_{LEP} = 10^{32}cm^{-2}Hz$.

In the case of FCC, it will be up to 10^{34}. These are the reasons behind the hadronic machine.

In the case of hadronic machine, given the mass of the proton is much larger than that of the electron, the synchrotron radiation is not a probem.

1.4 Project with p^+, p^- - LhC

The only limitation to the energy of the protons is the magnetic fields that is bending the particles. In this case we are not colliding point-like particles but bags of particles. We have a fraction f of the probability p_a and p_b, out of all the particles inside the protons, and they collid. They occurr at energy $\hat{s}$ which is smaller than s. The momentum they carry is: $x_a p$ and $x_b p$ so they do not carry all of the momentum.

So the cross section for a given process $\sigma_{p1,p2\to X}$ is:

$$\int_{x_a}^{1} dx_a \int_{x_b}^{1} dx_b P_a(x_a)P_b(x_b)\sigma(a, b \to X)$$

these are the parton probability functions. We put everything together in order to obtain the final cross section of the process. So this $\hat{s}$ is:

$$\hat{s} = x_a x_b s << s$$

So even if we have $7TeV$, the things that collide use a much lower energy. So:

$$E = \frac{(x_a + x_b)}{2} \cdot \sqrt{s}$$

and the longitudinal momentum:

$$p_L = \frac{x_a - x_b}{2}\sqrt{s}$$

in this case the LAB and CoM systems do not coincide. The fractions x_a and x_b are not accessible, are completely stochastic so that's why it's difficult. The

LAB and CoM are thus not easily linkable with a boost because of these things that complicate everything. **We do not know the center of mass system!**

We have to exploit what we know the most: the transverse plane! In this plane we know that the momentum is exactly 0. So since the momentum is conserved, zero momentum in the transverse means transverse in the final state.

1.4.1 Transverse momentum p_t

We also talk about transverse energy E_t which is basically p_t. Then we have the angle that defines the boost that is the rapidity η:

$$\eta = \frac{1}{2}\ln\frac{E+p_z}{E-p_z} = \frac{1}{2}\ln\frac{p+p\cos\theta}{p+p\cos\theta} \approx -\ln\tan\frac{\theta}{2}$$

This is also called pseudo-rapidity. Is the angle that defines boost along the longitudinal direction. The η is invariant. The other important quantity is ϕ. So we have: p_t, η, ϕ. Every particle is gonna be labelled by these three quantities.

So we want the detector to give us:

- Transverse momentum p_t
- Pseudorapidity η gives us the angle with respect to the longitudinal direction
- Angle ϕ

We expect physics to appear in the forward direction because of kinematical constraints. The higher the energy of the process, the more central is the interaction.

In the case of the protons, we have a higher range of processes that can be probed. We have also gluon interactions, gluon-quark, quark-quark and so on. The luminosity is also very high (order to 10^{34} and above). The events though are not clean at all.

Considering the LhC again, the luminosity for a single bunch is:

$$\alpha_b = \frac{N^2\nu}{A_T} = 10^{31}cm^{-2}Hz = 10^4 mb^{-1}Hz$$

The total cross section is in the range of $80mb$ (millibarn). So:

$$R = \alpha_b\sigma_{Inc} = 8\cdot 10^5 Hz$$

The frequency of the LhC is $10^4 Hz$. How is R higher than the machine frequency? This is because everytime the bunch collides, we have $8\cdot 10^5/8\cdot 10^4$ which is roughly 80. So we have 80 collisions of protons each time bunches interact. This is why it's so messy. We don't have just one collision, we have 80 simultaneous collisions and we only care about one of them. The others are just noise.

We have initial and final state particles. We know the colliding particles, we measure the final state particles. In between there's the interesting dynamics of the phenomenon. Knowing initial and final state, we want to know what the interesting dynamics is. That's all about it.

Since it is all dominated by quantum mechanics, a final state does not come univocally from a single initial state. So the analysis cannot be on an event by event fashion, this is because we don't know what happens istantaneously. We need to estimate distributions and here is where analysis comes in.

The events are candidate events since we do not know exactly what it is. The detectors are operated assuming that the final state particles are the standard model ones. We can interpret the data from the detector also assuming that something wierd is going on.

The standard procedure is: we interpret the data assuming that in the final state we just have standard model particles. For our work we're gonna have files, so collection of events for various particles also jets and missing energies. This is after massive data reduction on the raw data.

Typically we reduce the events to a subset that retain the number of features we're interested in. Like if we're interested in a process with 4 muons in the final state, we consider only these ones. So selecting events is done as a function to what we're interested in.

Also, we will be given montecarlo simulations of the physics in order to do comparisons with the data, to see if they somewhat correspond. We will have to try several procedures in order to enhance the signal the most over the noise. At that point we can run the analysis. Also, we need to asses statistical uncertainties (systematic will actually take years to tackle).

Chapter 2

Lesson 2 - April 24th, 2024

Lesson delivered by Zucchetta with incursions by Zanetti.

2.1 Particle processes

The processes we're interested in are the following:

- ttW: top, top, W boson
- ttZ: top, top, Z boson
- tWz: top, W and Z bosons
- tZq: top, Z boson and light quark q. A light quark can be up, down or strange.

Since we're colliding protons with protons, we're not dealing with point-like particles but with "bag" of particles that are smashed together (partons). Each of these processes is characterized by a cross section that determines how frequent it is (i.e. how probable) to observe such process. These are computed by theoretical calculations: These processes furnish the **signal**(s) but of course

Process	Cross Section
ttW	$0.60pb$
ttZ	$0.88pb$
tZq	$88fb$
tWz	$354fb$

these signals are tiny with respect to the background noise being generated by other processes that are of no interest for this analysis. This noise into which the signal is embedded is due to:

- top - antitop pair production: the cross section for this process is actually pretty big at $830pb$.

- W and Z boson processes: W-W, W-Z and Z-Z accounting for $10pb$.

The top quark, as all quarks, cannot exist alone so it fragments before it can hadronize and it goes off like this:

$$t \to Wb$$

where W is a W boson and b is a **bottom quark** (which is quite fat compared to the lighter ones). This W boson is of course unstable and it can decay in a pair of $q\bar{q}$ where q is a light quark or it can decay as a **leptonic pair**: electron and electronic neutrino, muon and muonic neutrino, tau and tauonic neutrino. These are all equiprobable.

While the $q\bar{q}$ can decay in 9 different channels given the color charge of the quarks, the leptons don't have that. We're mainly interested in the leptonic channels with electron and electronic neutrino and muon and muonic neutrino. These two account for 1/9 each of the probability of the decay of the W.

Generally calling l a lepton and ν_l the corresponding neutrino flavour, we thus consider:

$$t \to l\nu_l b$$

Now, the top quark is rarely produced alone: most of the times it is produced in pairs with the anti-top. The top-anti top pair decays in: These **final state**

Decay Channel	Probability
2l-2b	5%
1l-2q-2b	30%
4q-2b	44%

particles are those that are measured by the detector. We cannot measure quarks directly since they're too shortly lived to be measurable.

We now distinguish between two kinds of quark hadronizations:

- q hadronization: a lot of particles are produced by the hadronization process like pions, kaons, mesons. This kind of hadronization involves up, down, strange and charm quarks.

- b hadronization: this is the special case of the b (bottom) quark. The bottom quark hadronizes into a **B meson** which is a bounded state with a bottom quark and a lighter quark (so there's like the B_s meson for bottom+strange, the B_c for bottom+charm, etc). We consider $B^{\pm}$ and B^0 mesons.

The B mesons are much longer lived with respect to the quarks themselves (they can travel for up to $490\mu m$ from the interaction point, also called the **primary vertex**). Actually they travel even farther given the huge Lorentz-boost due to the high energy of the collisions. With a γ factor of almost 10 and $\beta \approx 1$, these particles can travel for half a centimeter when generated at $50GeV$.

The $B^{\pm}$ and B^0 are the ones that travel up to the **secondary vertex** where they decay. So, if we have a secondary vertex, we have a signature that something interesting happened. So **the presence of a secondary vertex in an event is the signature to look for**.

2.2 Detector geometry

The first layer of the CMS detector is the **silicon tracker** which is itself comprised of several layers: an inner tracker (a pixel detector that surrounds the interaction point) and a strip tracker (a detector that tracks the motion of the particle). The silicon tracker can only react to charged particles.

The other layers are calorimeters that measure the energy of the particles by destroying them.

The outermost layers are the muon detectors.

Finally, the detector generates a powerful magnetic field by means of superconducting magnets. The field allows to determine the sign of the charge of the charged particles since their tracks will be bent in different directions depending on their individual charges.

At the LHC, protons are collided in bunches (i.e. groups of protons). Out of these, only a fraction of them actually interact and this number is called, in HEP jargon: **pile up**. So, the higher the pile up, the more difficult the reconstruction becomes.

2.3 Jets and B-Jets

A jet is a "grouping" of multiple particles that originate from the decay of a secondary particle. A jet can be treated as a single entity so it has its own transverse momentum, energy and mass. A special type of jet is the **B-jet**.

The identification of a **B-jet** with respect to a normal jet is absolutely non trivial and it can be done only in $\approx 50\%$ of cases. A specifically purposed algorithm (a discriminator) determines the probability of a jet being a B-jet and that probability is stored in a dedicated branch inside each ROOT file.

2.4 Leptons

Leptons (electrons and muons) are easy to reconstruct. Most of them originate from decay of quarks. How to separate leptons from W decays and from b decays?

The W boson decays in $l\nu$ and typically they are **isolated**. Conversely, the leptons originating from b decays are **not isolated**. For this reason a metric known as **isolation** is defined:

$$I = \frac{\sum p_t}{p_t^l}$$

The idea is to consider, in a cone, all the particle tracks around the lepton of interest. Then, all the momentum of these particles that surround the lepton are summed and the sum is divided by the momentum of the lepton itself, p_t^l. So, if the lepton was isolated, ideally the cone would contain no other particles but the lepton, leading to an isolation of exactly 0. Of course, since background events are present, the cone will contain the tracks of other particles. So, if the lepton was an isolated one, its energy would be independent from the energy of the other particles in the cone and its isolation value would be small. On the contrary, if a lepton wasn't isolated but, for instance, it was generated by the decay of a b quark, its cone would contain a lot of secondary particles and hence its isolation value would become quite large.

Isolation variable I is thus quite useful to discriminate between events.

The particle tracks are reconstructed by an algorithm that is called **particle flow** (PF in the data).

2.5 CMS Trigger

The trigger is made of two layers: L1 and HLT. The level 1 trigger L1 operates at LHC's clock frequency of $40MHz$ and selects events in real time according to its programming (can be set to activate on several conditions depending on the physics process of interest). The data comes out of L1 at a rete of $\approx 100kHz$. This if further reduced by the HLT to $1KHz$, corresponding to a data rate of $\approx 1GB/s$.

The trigger is programmed to select events according to a specific set of rules on the various signals coming from the detector. The list of conditions is called a **menu**. Each of these conditions is sufficient, meaning that the trigger will fire if at least one of those passes and the selected event is thus written.

Inside the data files, the trigger's variables are named like "HLT" and for instance $HLT\_SingleIsoMu27$ means: the rule requires a single isolated muon of transverse momentum $> 27GeV$ for the trigger to fire. This is useful to separate the various events.

2.6 Data structure - ROOT files

Events are stored in a tree structure inside a ROOT file. Each event consists of a series of features (named branches). When an event is read, all the values in the branches at that specific event index are considered.

In order to select the various events, rules must be applied to the features in order to extract them: for instance how many leptons we want, etc.

Chapter 3

Lesson 3 - May 8th, 2024

Lesson delivered by Zucchetta with intro by Zanetti.

Important: lesson material is CMS note 2011/005 available here.

3.1 Introduction

For historical reasons, the High Energy Physics community uses frequentist approach to statistics (actually a mixed approach where some contaminations with Bayesian's). There are also some exceptions: neutrino physics uses bayesian approach as well as some new interpretations of quantum mechanics. Now, the two approaches must of course converge towards the same conclusion albeit with entirely different processes.

3.1.1 Bayesian interpretation

One starts from hist/her personal degree of belief of something being true or false. For example: the Higgs mechanism is a good model. This is our prior. When we do measurements, each one contributes to our belief/disbelief in the prior and hence we update our knowledge as new data comes in. The key point is that the prior can be different for everybody. What is however fundamental is that no matter what the prior is: if two observers have the same information, they should arrive at the same conclusion. Another key point of the Bayesian approach is that it allows to **answer directly** to the question: "does this model describe the data well?".

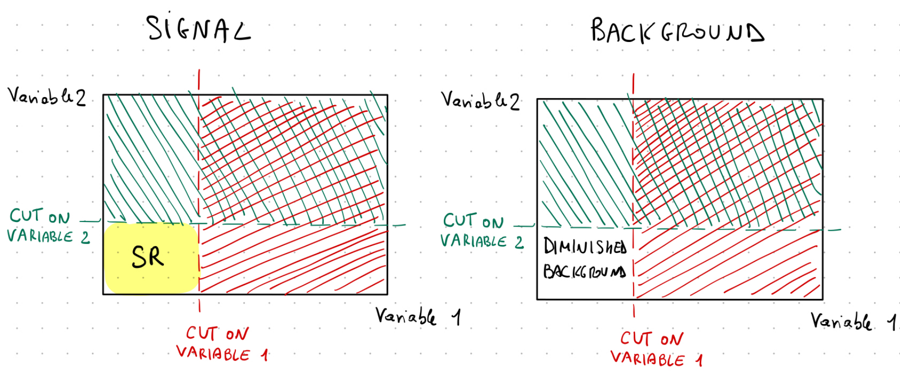
3.1.2 Frequentist interpretation

With the frequentist approach **it is impossible** to answer to that question. It is a fundamental/theoretical impossibility. At the core of the frequentist interpretation is a strong rejection of Bayesian's subjectivity. The only thing we can answer with frequentism is **whether the data is more or less compatible with the model**. We cannot directly say if the model can explain the data but only how probable it is for the data to come from that model under a given hypothesis. This is formalized by means of the **likelihood** function and the **p-values** that are extracted from it.

3.2 Signal and Background

3.2.1 Introduction

When we consider a particular physics process, we say we're interested in a specific **signal**. All the other processes that occurr concurrently or in the "vicinity" of it constitute what is called **background**. As an example we may consider two variables of the dataset and represent them as x and y axes in a square. Actually, for higher dimensional datasets we say that we have an hypercube where each side of the cube is occupied by a variable. Let's go back to the two dimensional case since it's easier to visualize. We draw two such squares: one for the signal and one for the backgrounds as shown below:



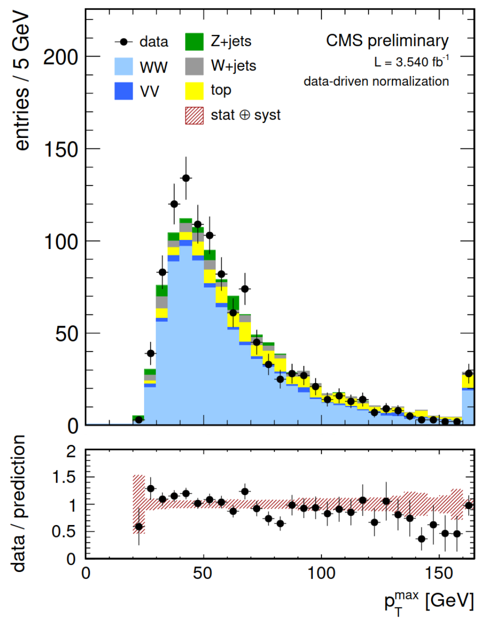
The signal will populate a specific region of the hypercube (or in this case of the square). It is named **SR** which stands for Signal Region. Inside this region we want to enrich the signal and lower the background contributions. We do that by applying **cuts**. We have a cut on variable 1 and another cut on variable 2. Inside the SR region delimited by the cuts we will have the signal + the background contribution in that region. In order to further clean the signal, one could inspect the background in a region far from the SR. Such regions are called **control region**. The idea is to study and model the background in a region where, for sure, we know that no signal is present. In this way we have a faithful representation of the background that will be very useful to further diminish it in the signal region. This could seem simple as a concept but in reality is quite sophisticated and advanced.

Coming back to the SR: we still have a background accompaning the signal. At this point it is important to identify the **most appropriate kinematic variables that will yield the highest discriminating power** so that the greatest reduction in the background could be accomplished without harming the signal yield too much by applying selection rules (i.e. other kinds of cuts and combinations of such) with those.

To make an example: let's consider the top quark reconstruction. The invariant mass of the top is a good kinematic variable. So, for instance, selecting events where the invariant mass lies in a certain interval could help to better isolate the signal.

3.2.2 Plots

The typical plot to consider is a histogram like the following:



The quantities are always binned. In a histogram we always show:

- The actual data: as data points with **poissonian error**.
- The simulated signal and background(s).

It is important to notice that the various bins are **statistically independent** between each other. Each bin has a **content**, named n, that is the **number of events** with the physical quantity being represented lying within the boundaries of that bin.

The first important thing to notice in a plot like this is that **data must follow the background distributions**. If this doesn't happen we need to review the selection rules. The various backgrounds are shown together with the histograms stacked on top of each other.

Now, the data points (which are in on itself frequencies/counts) are inputted in the histogram with weight equal to 1. For the simulation is quite different since the weight is generally different from 1. Note that the weights are per bin. For the Montecarlo, the i-th weight is:

$$w_i = \frac{\mathcal{L} \cdot \sigma}{N_{gen}}$$

where $\mathcal{L}$ is the luminosity, σ is the cross section of the MonteCarlo and N_{gen} is the number of generated events.

Coming back to the data, we have that, for each point (i.e. each bin content) we have uncertainty given by Poisson's statistics. So, if the bin content is n, the uncertainty is just $\sqrt{n}$.

In the case of the Montecarlo, the uncertainty is instead:

$$\sqrt{\sum w_i^2}$$

3.2.3 On the background

The Poisson probability of an observed count n given the expectation ν from the simulations is given by:

$$Poisson(n|\nu) = e^{-\nu} \frac{\nu^n}{n!}$$

The background is a function of a number of parameters. These parameters are called **nuisance parameters**. These same parameters also apply to the signal. These parameters act as uncertainties. They are collectively called $\vec{\theta}$ and they are a **vector**.

We then introduce the parameter μ that is called the **signal strength** which is a scalar and it is a multiplicative factor that scales the signal vertically. So our expectation ν is given by:

$$\nu = \mu \cdot S(\vec{\theta}) + B(\vec{\theta})$$

where S and B are the signal and background functions.

An important thing to notice is that $\vec{\theta}$ is **not a free parameter**. It actually has a pdf that is written as:

$$\rho(\theta|\hat{\theta})$$

where $\hat{\theta}$ is the **true** value of the uncertainty. This quantity is a **posterior pdf** of θ, given $\hat{\theta}$. The frequentist approach actually gives us the likelihood function which is $p(\hat{\theta}|\theta)$. In order to go from ρ to p and viceversa, one needs Bayes' theorem:

$$\rho(\theta|\hat{\theta}) = \frac{p(\hat{\theta}|\theta) \cdot \pi(\theta)}{p(\hat{\theta})}$$

where $p(\hat{\theta})$ is actually a constant and can be dropped and $\pi\theta$ is the **prior** on θ. This prior is extremely import since it encodes the knowledge about the **systematic uncertainties**. These are of **paramount importance** when doing analysis at a professional/publication quality level since, in the case of HEP, $\pi(\theta)$ encodes the information we have about the detector's induced uncertainty in the data without which one cannot hope to extract meaningful signals from the data.

So we have a pdf for θ that will be centered around the true value $\hat{\theta}$. The pdf will be "kind of" gaussian and its width will be directly proportional to the accuracy of the estimation of the true value.

3.3 Inference

The process of statistical inference follows theese steps:

- Define the $\mathcal{H}_0$ and $\mathcal{H}_1$ hypotheses.
  - $\mathcal{H}_0$ is called the **null** hypothesis. It represents the experimental situation where **we have NO signal and ONLY background**, so the strength of the signal $\mu = 0$.
  - $\mathcal{H}_1$ is called the **alternate** hypothesis. It represents the experimental situation where the **strength of the signal** $\mu > 0$.

  The **level of confidence**, indicated by α, is set at 0.05 which is a convention in HEP.

- Define a test statistic. An example of test statistics used in simpler situations is that of χ^2. A test statistic is a variable that is particularly sensitive to difference between signal and background.

- Find the PDF of the test statistic. At the LHC there is no way to know, analytically, the PDF.

- Compute the p-value. One computes the p-value of the $\mathcal{H}_0$ hypothesis being true.

- **If the p-value is less than α, we reject $\mathcal{H}_0$. This is the only thing that frequentist approach allows us to do**.

So this means that we reject the null hypothesis (there's only background) only when the p-value gets smaller than alpha. So we say that we reject it if the probability of the observed data to come from a model where there's only background falls below the fixed threshold of α. We then reject this hypothesis but this does not imply that $\mathcal{H}_1$ is true!

There are two kinds of test statistics we'll see: significance and setting an upper limit.

3.3.1 Test Statistics - Significance

This situation arises when we do observe a signal and we want to calculate how significant it is. We need to consider the likelihood function, starting from the histogram:

$$\mathcal{L}(data|\mu,\theta) = \prod PoissonPdf(data|\mu S(\theta) + B(\theta)) \cdot p(\hat{\theta}|\theta)$$

where "data" is the data, and $\mu S(\theta) + B(\theta)$ is what the histogram gives us and p serves to constrain the θ. The $\prod$ because we make products of all the poissonians, bin per bin. So, this likelihood is constructed numerically starting from the histogram and can become computationally expensive.

Once we've computed $\mathcal{L}$ we can compute the so called **likelihood-ratio**. This is defined as the quantity q_0 so defined:

$$q_0 = -2\log\frac{\mathcal{L}(data|\mu = 0, \hat{\theta_0})}{\mathcal{L}(data|\mu = \hat{\mu}, \hat{\theta \mu})}$$

where $\hat{\mu}$ comes from the best fit to the data and $\hat{\theta \mu}$ refers to the best fit of θ using the best μ. **We use a ratio because of the Neyman-Pearson lemma that guarantees us that this is the most powerful discriminator between the two situations**: the one where the signal is 0 ($\mu = 0$) and the one where the signal is at its best ($\mu = \hat{\mu}$).

We need a pdf for q_0 so that we can compute the p-value. Of course there's no analytical expression. So one must proceed as follows:

- Determine a q_0 on the data that will be called q_0^{obs}.
- For each bin, generate synthetic data so that one can obtain multiple values of q_0.

Plotting these multiple q_0, one will obtain a histogram with a distribution that will be a density estimate of the required pdf that we'll call:

$$f(q_0|\mu = 0, \theta_0)$$

since we're generating under the $\mathcal{H}_0$ hypothesis. **The p-value under $\mathcal{H}_0$ is defined as the probability of observing** $q_0 \geq q_0^{obs}$, so:

$$p_0 = P(q_0 \geq q_0^{obs}) = \int_{q_0^{obs}}^{+\infty} f(q_0|\theta, \hat{\theta_0})dq_0$$

Our case if discrete since we have binned values so the $\int \leftrightarrow \sum$ and we'll sum the bins in the area under the curve for $q_0 \geq q_0^{obs}$ and divide the sum by the total number of generated q_0s.

The **significance** z is related to the μ and it is defined as: **the number of standard deviations in a gaussian with mean** 0 **and** $\sigma = 1$**,** $\mathcal{G}(0,1)$**, such that the integral:**

$$\int_z^{+\infty} \mathcal{G}(\prime, \infty) = p_0 \Rightarrow z = \Phi^{-1}(1 - p_0)$$

where Φ is the CDF of the gaussian. The reason why one uses z instead of p_0 is that p-values become really small and it's more informative to use the number of standard deviations.

For instance, in the Higgs boson discovery, a threshold of 5σ was requested in order to claim the discovery. This meant a p-value of $2.87 \cdot 10^{-7}$. Conversely, a $p_0 = 0.05$ means $z = 1.64$.

3.3.2 Test Statistics - Set an upper limit

This situation arises when we do not observe a signal and we want to set an upper limit on its presence.

This test statistic is also called **CLs methods** (confidence levels methods). In this case we use a likelihood as it was defined before but the q_0 changes. This time q_0 is a function of μ, meaning that:

$$q_0 = -2 \log \frac{\mathcal{L}(data|\mu, \hat{\theta_\mu})}{\mathcal{L}(data|\mu = \hat{\mu}, \hat{\theta \mu})}$$

Since we do not see the signal here we're left with two cases: $\mu = 0$ or $\mu = \mu$ where μ is the parameter chosen. We thus have not one but two PDFs. We have a $q_0^{obs}(\mu)$ coming from the real data and thus we have the two PDFs:

$$f(q(\mu)|0, \hat{\theta_0})$$
$$f(q(\mu)|\mu, \hat{\theta_\mu})$$

And so:

$$P_\mu = \int_{q(\mu)^{obs}}^{+\infty} f(q(\mu)|\mu, \hat{\theta_\mu}) dq(\mu)$$

and:

$$1 - P_b = \int_{q(\mu)^{obs}}^{+\infty} f(q(\mu)|0, \hat{\theta_0}) dq(\mu)$$

The **CLs** is defined as:

$$\frac{P_\mu}{1 - P_b} = \alpha$$

The idea is thus to vary the μ so that the ratio is $\alpha = 0.05$. This will give an upper limit to the cross section. The procedure is thus:

- Scan with different μ values so that the ratio $= \alpha$.
- This tells us we can exclude a signal where the cross section is less than μ with a confidence level of 95%.

This means that: if we repeat the experiment 100 times, we'll find this result 95% of the times.

If we repeat the CLs procedure multiple times, we'll obtain a distribution of μ. We can then calculate the cumulative distribution of the μ. By setting the vertical axis to 0.5 we can project that value towards the cumulative function and obtain the corresponding x value which will be the μ value at 95% of the confidence level. We can also do this by placing not a single value of 0.5 but an interval and projecting it back to the horizontal axis.