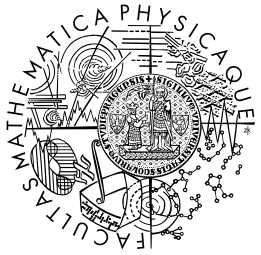

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Jaroslav Říha

The Use of Bayesian Statistical Analysis in Particle Physics

Institute of Particle and Nuclear Physics

Supervisor of the bachelor thesis:	Mgr. Marián Kolesár, Ph.D.
Study programme:	Obecná fyzika
Study branch:	Fyzika

Prague 2019

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb.., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ..... signature

I'd like to thank Mgr. Marián Kolesár, Ph.D. for his patience and dedication.

Název práce: Využití Bayesovské statistické analýzy v částicové fyzice

Autor: Jaroslav Říha

Ústav: Ústav částicové a jaderné fyziky

Vedoucí bakalářské práce: Mgr. Marián Kolesár, Ph.D., Ústav částicové a jaderné fyziky

Abstrakt: Tato bakalářská práce se bude zabývat Bayesovským přístupem ve statistice a tento přístup aplikuje na rozpadové konstanty v "resumované" chirální poruchové teorii. Také bude diskutovat, jak důležité je uvádět předpoklady. Výsledkem práce budou restrikce na parametry použité teorie.

Klíčová slova: Bayesovská statistika, rozpadové konstanty, chirální poruchová teorie, Bayesův teorém, resumovaný přístup

Title: The Use of Bayesian Statistical Analysis in Particle Physics

Author: Jaroslav Říha

Institute: Institute of Particle and Nuclear Physics

Supervisor: Mgr. Marián Kolesár, Ph.D., Institute of Particle and Nuclear Physics

Abstract: This bachelor thesis will consider Bayesian approach to statistics and then apply it on decay constants in 'resummed' chiral perturbation theory. It will also discuss how important it is to state one's assumptions. The results will be in the form of reasonable restrictions one can place on the parameters of the theory.

Keywords: Bayesian statistics, decay constants, chiral perturbation theory, Bayes' theorem, resummed approach

Contents

Introduction

Almost all theories in physics contain constants. Most of these constants cannot be determined from the theory itself, meaning that the theory is somewhat more general than the laws of nature it describes. The theory may place some restrictions on these constants, but their exact value cannot usually be found from it. Thanks to experiments, we can try to extract these values. The situation gets more difficult, when the relation to experimental observables is non-trivial. In such case, one may view these constants as parameters of the theory.

Thankfully, it is still possible to extract limitations of the parameters by using statistical methods. We may not know the exact values, but we may know which part of that multidimensional parameter space can be excluded.

The frequentistic approach is based on the idea of using infinite number of measurements and finding the limit of its relative frequency to calculate the 'real' value of an unknown quantity. The most common method used is finding the minimum of χ^2 distribution (fitting procedure). But this method does have a downside, because it is harder to incorporate assumptions for parameters into those calculations. Often the frequentistic approach uses assumptions such as uniform distributions of unknown variables, but that may not always be the best approximation.

On the other hand, Bayesian approach allows us to use *a priori* assumptions and derive our results from them. One of the advantages of Bayesian approach over frequentistic one is that it is easier to incorporate assumptions for parameters. Bayesian approach can also be used to assign probabilities to limited occurance cases, something the frequentistic approach struggles with.

Effective field theories have the form of an expansion at low energies. Chiral perturbation theory is an effective field theory of quantum chromodynamics, which has relatively slow convergence in the case when three light quark flavors are considered. One way of dealing with the higher order remainders, which will be used in this work, is to resum them and use the Bayesian approach to implement a statistical estimate of their influence.

In the first chapter I will discuss the basics of probability theory and show how naturally we can obtain the Bayes' theorem. Therein I will also discuss the *prior*, its importance and different ways to think about it.

In the second chapter, as a simple demonstration, I will use a few examples of use of the Bayes' theorem.

In the third chapter I will introduce the theory behind the calculations in the fourth chapter. I will discuss some of the properties of strong interactions mostly from a group theory point of view, explain the basics of effective chiral perturbation theory and introduce the decay constants of the light pseudoscalar meson octet.

Then in the fourth chapter I will use the Bayesian approach and apply it on a sector of decay constants F_π,F_K,F_η. I will use two different values of F_η extracted from experiments and obtain restrictions on the parameters of the theory. These restrictions will then be compared.

1. Probability theory

1.1 Introduction to Bayesian approach

Let's formulate the basics of the Bayesian approach (for more information see D'Agostini [1999]). First, we can start by postulating the three axioms of probability.

- The probability of an event is a real number between 0 and 1.

$$P(E) \in <0,1>\ ,\ \forall E \in F, \tag{1.1}$$

  where E is an event in F event space of all possible events.

- The probability of the event space is 1.

$$P(F) = 1 \tag{1.2}$$

- For E_1, E_2,... a countable sequence of disjoint sets

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) \tag{1.3}$$

We are searching for a formula that would let us calculate the probability of A under the assumption of B (written as $P(A|B)$). The probability of intersection of two sets can be written as $P(A \cup B) = P(A|B)P(B)$, where $P(A|B)$ is the probability of "if B, then A". Let A,B be subsets of Ω event space. Then

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A). \tag{1.4}$$

By dividing the whole equation by P(B) we get *the Bayes' theorem*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{1.5}$$

where *P(A)* is called *prior*. $P(A|B)$ is sometimes referred to as *posterior probability*. Here we can use a partition of one - separating a set into finite number of subsets that are all disjoint with each other. This can be written in the form of

$$P(A) = \sum_{n=1}^{N} P(A|B_n)P(B_n). \tag{1.6}$$

Using this on Bayes' theorem, one can calculate *P(B)*, which is often very hard to do. With the partition of one used on (1.5) we get

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_{n=1}^{N} P(B|A_n)P(A_n)}. \tag{1.7}$$

Now we are capable of calculating a probability for a single assumption. Should we continue adding assumptions, we'd get the formula for a probability of A under the assumptions of B_n

$$P(A|B_1 \wedge B_2 ... \wedge B_n) = P(A) \prod_{i=1}^{n} \frac{P(B_i|A)}{P(B_i)}. \tag{1.8}$$

The Bayes' formula (1.7) can also be generalized into an integral form for continuous sets

$$P(x|y) = \frac{P(y|x)P(x)}{\int_\Omega P(y|x)P(x)dx}. \tag{1.9}$$

1.2 More on the prior

As mentioned earlier, the Bayesian approach uses *a priori* assumptions, also called priors. It is essentially our initial prediction.

For example, consider a measurement of an observable. One can then wonder if the data prove their theory or if they are incompatible. Therefore one wants to know the probability of the theory being true given measured data. They use an initial prediction, often total uncertainty and look at posterior probability, mainly to see how it changed from their prior. If it gets bigger, it means that their data somewhat fit the theory, hence the rise in probability. If on the other hand the posterior probability is lower than their prior probability, it means the data do not fit the theory well.

This prior is very commonly thought of as a function depending only on the theory, but an interesting aspect of the Bayesian approach is that one can also change their prior (meaning they change their hypothesis) and they might see different results. This way one can explore the consequences of various assumptions.

It might be tempting to think of the prior as something "subjective" and therefore "not scientific enough", but this can also be its forte. If all people had the same knowledge of the system and the same assumptions, they'd inevitably get the same results. The subjective part is the fact that different people have different levels of knowledge of the problem and can have different assumptions.

1.3 About conditional probability

From (1.5) one can easily see that $P(A|B) = P(B|A)$ only if $P(A) = P(B)$. However, what is obvious from an equation might be easier to overlook in real life. For example, consider an experiment of n measurements with the result

$$\mu = \overline{x} \pm \frac{\sigma}{\sqrt{n}}, \tag{1.10}$$

μ being the real value, $\overline{x}$ being the arithmetic average and σ the experimental uncertainty. Some people might interpret this as

$$P(\overline{x} - \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{\sigma}{\sqrt{n}}) = 68\%. \tag{1.11}$$

But conventional statistics only says

$$P(\mu - \frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + \frac{\sigma}{\sqrt{n}}) = 68\%. \tag{1.12}$$

The capital letters serve to distinguish between those two cases - capital letter variables are random variables of which the small letters are a realization. In the

first case, μ is treated as a variable with $\overline{x}$ and σ being "constant". Which isn't true, only the measurements are supposed to carry an error, not the object of our measurement itself.

This fact is commonly overlooked, because we often encounter cases, where such an inversion is in fact possible. However, it might be due to our implicit assumptions, such as all unknown variables being uniformly distributed.

1.4 Bayesian approach in physics

As seen in previous section, we may often need to "invert" our probability. As results of experiments in physics are often formulated in terms of normal distributions, in the case of k independent experiments 1.12 can be written as

$$P(data|true) = \prod_k \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left[-\frac{(O_k^{exp} - O_k^{true})^2}{2\sigma_k^2}\right], \tag{1.13}$$

where observables O_k^{true} are the true values of observables with experimental uncertainties σ_k. If we have a theory, which we assume can reconstruct our true values using some true values of parameters, we may rewrite this equation into

$$P(data|X_i) = \prod_k \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left[-\frac{(O_k^{exp} - O_k^{theory}(X_i))^2}{2\sigma_k^2}\right]. \tag{1.14}$$

$P(data|X_i)$ is the probability density of obtaining observed values of the observables O_k^{exp} with X_i as the true values of the parameters.

And finally, now we can use the Bayes' theorem (1.8) to invert the probability to obtain $P(X_i|data)$ - the probability of particular values of parameters X_i being true given the experimental data.

2. Simple examples of use

2.1 Drug test

Suppose there is a drug test that yields positive results if a drug user is tested with 99% efficiency (($P(+|user) = 0.99$)) and that there is a 99% probability that it yields negative test results for non-drug users ($P(-|non-user) = 0.99$). Now assuming that there is a drug user in every one hundred people ($P(user) = 0.01$), the possibility of being a drug user if the test is positive is

$$\begin{aligned} P(user|+) &= \frac{P(+|user)P(user)}{P(+)} \\ &= \frac{P(+|user)P(user)}{P(+|user)P(user) + P(+|non-user)P(non-user)} \\ &= \frac{0.99 * 0.01}{0.99 * 0.01 + 0.01 * 0.99} \\ &= 0.5 \end{aligned} \tag{2.1}$$

How interesting! This example actually shows us that the test isn't that much efficient. The reason for this is that the probability of being a drug user is too low when compared to the number of non-users and the sensitivity of the test is nowhere near enough. But still, we didn't know anything about that person, before he took the test. We assumed there was a 1% chance that he was a drug user and now we know that with 50% certainty. If we changed those numbers a bit, i.e. changed $P(user) = P(non-user) = 0.5$, the probability of being a drug user if the test comes positive would actually change to 99%. This shows how big of an influence the *a priori* information has. It also shows us what the probability would be if we tested him a second time.

2.2 Weather wizard

Say a man comes to a village claiming that he can control the weather and summon rain for a price. The village lies in an arid area and it rains only ten days a year (with each day having the same chance of being a rainy day). The village is split in half with half being skeptic, who give the weather wizard 1% that he is the real deal ($P_{skeptic}(wizard) = 0.01$), and the other half being unsure what to believe ($P_{unsure} = 0.5$). Obviously, both groups consider him capable of summoning up the storm if he truly was a wizard ($P(rain|wizard) = 1$). The next day rains. The probabilities these two groups give the weather wizard that

he can do, what he advertised, are

$$\begin{aligned} P_{skeptics}(wizard|rain) &= \frac{P(rain|wizard)P(wizard)}{P(rain|wizard)P(wizard)+P(rain|con)P(con)} \\ &= \frac{1*0.01}{1*0.01+\frac{10}{365}*0.99} \\ &\doteq 0.2694 \\ P_{unsure}(wizard|rain) &= \frac{1*0.5}{1*0.5+\frac{10}{365}*0.5} \\ &\doteq 0.9733 \end{aligned} \tag{2.2}$$

If he promises to summon a storm again and he once again delivers, the probabilities raise to

$$\begin{aligned} P_{skeptics}(wizard|1 \wedge 2) &= \frac{P(2|wizard)P(wizard|1)}{P(2|wizard)P(wizard|1)+P(2|con)P(con|1)} \\ &= \frac{1*0.2694}{1*0.2694+\frac{10}{365}*0.7306} \\ &\doteq 0.9308 \\ P_{unsure}(wizard|1 \wedge 2) &= \frac{1*0.9733}{1*0.9733+\frac{10}{365}*0.0267} \\ &\doteq 0.9992 \end{aligned} \tag{2.3}$$

As you can see all it takes to become a shaman in an arid area is to predict the weather correctly 2 times in a row. But this shows us an interesting phenomenon - if we measure something many times, our a priori prediction doesn't matter. Both probabilities converge to the same number.

3. Physics theory

3.1 Quantum chromodynamics and its symmetries

Quantum chromodynamics (QCD) is a theory of strong interactions - interactions between quarks and gluons (for more information see Donoghue et al. [1992], structure inspired by Kolesár [2008]). An important fact about this theory is that it is based on *local gauge invariance* (meaning it remains invariant after transformations of a local Lie group). The existence of such invariant transformation means that there exists a time invariant. In this case it's called *color*. This symmetry was the basis upon which the theory was built. The symmetry of QCD is SU(3), which stands for *special unitary* 3×3 *matrices*, meaning the theory has a group of matrices with determinant equal to 1 "hidden" in it. The full Langrangian density, which honors the aforementioned restriction, can be written as

$$\mathcal{L}_{QCD} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} + \overline{q}\left(i\gamma_\mu D^\mu - M\right) q, \tag{3.1}$$

where γ^μ are Dirac matrices, $G^{\mu\nu}$ is the antisymmetric gluon field tensor

$$G^{\mu\nu} = \partial^\mu G^\nu - \partial^\nu G^\mu + ig[G^\mu, G^\nu] G_\mu = \frac{1}{2}\lambda^a G^a_\mu. \tag{3.2}$$

G^a_μ are the gluon fields, a their color index, λ^a are the Gell-Mann matrices and q is the quark vector

$$q = \begin{pmatrix} u \\ d \\ s \\ c \\ b \\ t \end{pmatrix}. \tag{3.3}$$

M is the quark mass matrix and D^μ is the covariant derivative

$$D^\mu = \partial^\mu + igG^\mu, \tag{3.4}$$

with g being the strong interaction coupling constant.

The coupling constant g has a very interesting property - it decreases with scale. This unique effect known as *asymptotic freedom* has grave effect on perturbation theory, since on low scales the expansion fails. On the other hand, in the high energy region the quarks can essentially be thought of as free particles inside the hadrons. To have a look at the low energy region, one has to examine more closely the symmetries of this theory. In (3.1) there is one additional symmetry - U(1) for the quark vector. This means that if we transform $q \to e^{i\phi} q$, the form of the Langrangian remains the same. Emma Noether's theorem published in 1918 is one of the most important mathematical theorems applicable for physics. It can be applied every time the system exhibits some sort of continuous symmetry. It states(Thompson [1994]):

If a system has a continuous symmetry property, then there are corresponding quantities whose values are conserved in time.
From the U(1) symmetry we get *the baryon number conservation.*

It is interesting to notice that different quarks differ only in their masses. Thus one could, if the differences between their masses were small, construct an additional approximate symmetry $SU(N_f)$, with N_f denoting the different number of quarks we "consider of the same mass". Suppose we consider the first three quarks in mass (u, d and s) to be such. From this one gets an additional SU(3) approximate symmetry and (3.1) can be rewritten as

$$\begin{aligned}
\mathcal{L}_{QCD} &= \mathcal{L}_0 + \mathcal{L}_{sym.br.} && (3.5)\\
\mathcal{L}_0 &= -\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} + i\overline{q}\gamma_{mu}D^\mu q - \frac{1}{3}\overline{q}M_{u+d+s}q + L_{heavyquarks} && (3.6)\\
\mathcal{L}_{sym.br.} &= -\frac{1}{3}(m_u - m_d)(\overline{u}u - \overline{d}d) - \frac{1}{3}(m_u - m_s)(\overline{u}u - \overline{s}s) - \\
&\quad -\frac{1}{3}(m_d - m_s)(\overline{d}d - \overline{s}s), && (3.7)
\end{aligned}$$

with q now denoting only the light mass quarks and M_{u+d+s} being an identity matrix multiplied by the sum of light quark masses. Now it is obvious that the Langrangian $\mathcal{L}_\prime$ is invariant under unitary transformation of the light quark vector. The symmetry breaking part would vanish if the masses of these 3 quarks were identical.

The mass of d quarks is roughly 4.7 MeV, u's is roughly 2.2 MeV and the mass of s being roughly 95 MeV, taken from Tanabashi et al. [2018]. As one can see, the mass difference between u and d quarks is very small and s's mass is only ten times bigger, which is still relatively small compared to typical scale of strong interactions (usually some GeVs, the energies of hadrons). Therefore it is natural to extend the symmetry to all three of these quarks. Using this approach, we get the isospin symmetry due to u and d being of similar mass and the eightfold way once we include s.

Now, were not only the differences but the masses themselves small too when compared to the scale of the theory, the "left" and "right" components of the quark fields may emerge - two independent SU(3) symmetries with their own generators. The resulting symmetry would be $SU(N_f)_L \times SU(N_f)_R \times U(1)_L \times U(1)_R$ and Lagrangian density could be written as

$$\begin{aligned}
\mathcal{L}_{QCD} &= \mathcal{L}_0 + \mathcal{L}_{sym.br.} && (3.8)\\
\mathcal{L}_0 &= -\frac{1}{4}G^a_{\mu\nu}G^{a\mu} + i\overline{q}_L\gamma_{mu}D^\mu q_L + i\overline{q}_R\gamma_{mu}D^\mu q_R + \mathcal{L}_{heavyquarks} && (3.9)\\
\mathcal{L}_{sym.br.} &= -\overline{q_R}Mq_L + h.c., && (3.10)
\end{aligned}$$

where $q_{L,R} = \frac{1}{2}(1 \mp \gamma_5 q)$.

Since our groups are Lie groups, they can be parametrized by using unitary transformations

$$U = e^{-i\frac{v^a}{2}\lambda^a}, \tag{3.11}$$

where v^a are real parameters and λ^a are Pauli matrices for SU(2) or Gell-Mann matrices for SU(3) (see Georgi [1999] for more information). These generators obey the commutation relations

$$[\lambda^a,\lambda^b] = \sum_c if^{abc}\lambda^c. \tag{3.12}$$

With the $SU(N_f)_L \times SU(N_f)_R$ symmetry, we can parametrize them as

$$U_{L\times R} = e^{-iv_L^a L^a - iv_R^a R^a}. \tag{3.13}$$

Now we may choose a different parametrization, with one group of generators being symmetrical under L-R exchange and one being antisymmetrical

$$S^a = L^a + R^a \tag{3.14}$$
$$A^a = R^a - L^a, \tag{3.15}$$

the commutation relations change accordingly

$$\left[S^a,S^b\right] = \sum_c i f^{abc} S^c \tag{3.16}$$
$$\left[A^a,A^b\right] = \sum_c i f^{abc} A^c \tag{3.17}$$
$$\left[S^a,A^b\right] = \sum_c i f^{abc} A^c. \tag{3.18}$$

If we take a closer look at the first relation, we may notice that the symmetrical part forms a subalgebra. We once again get the *eightfold way* symmetry for SU(3) or *isospin* symmetry for SU(2).

3.2 Spontaneous symmetry breaking

Let Q^a be a generator of symmetry that commutes with Hamiltonian. Now let's see what that can tell us about the spectrum of Hamiltonian:

$$H|n> \;=\; E_n|n>\,,\;\; Q^a|n>=|n'> \tag{3.19}$$
$$E_{n'}|n'> \;=\; H|n'> \;=\; HQ^a|n> \;=\; Q^aH|n> \;=\; E_n|n>\,. \tag{3.20}$$

This means that we would see a degenerate multiplet of energy eigenstates. Usually in quantum field theory the lowest state, called vacuum, is unique. Which means that our generators should annihilate it. But if there were such generators that $Q^a|0>\neq 0$, these states should have energy equal to the physical ground state. Therefore the ground state would not be symmetrical under the whole symmetry group and the corresponding states in the excited multiplets obtained by the perturbative expansion around vacuum would be missing. The Goldstone theorem (Goldstone et al. [1962]) states that for every symmetry breaking generator there exists one independent state of zero energy. These states are observable as massless particles with spin 0 called *Goldstone bosons*.

This symmetry breaking shows us the pattern of hadron multiplets. For SU(3) symmetry, we may observe 8 pseudoscalar Goldstone bosons - octet of pions, kaons and eta with small masses.

So, we started from

$$SU(N_f)_L \times SU(N_f)_R \times U(1)_L \times U(1)_R, \tag{3.21}$$

transformed it into

$$SU(N_f)_S \times SU(N_f)_A \times U(1)_S \times U(1)_A \tag{3.22}$$

and found out that there is spontaneous symmetry breaking due to subgroup $SU(N_f)_A$. We also lose the $U(1)_A$, because it fails to be a symmetry on quantum level due to the Abelian anomaly. Therefore our symmetry group is reduced to

$$G = SU(3)_S \times U(1)_S \tag{3.23}$$

As the Goldstone theorem states, for each generator Q^a for which exists an operator O^a such that

$$< 0|[Q^a,O^a]|0 >\neq 0, \tag{3.24}$$

there must be one independent massless state $|GB >$ which

$$< 0|J_0^{a,A}(0)|GB >< GB|O^a|0 >\neq 0, \tag{3.25}$$

where $J_\mu^{a,A}(0) = \frac{1}{2}\overline{q}\gamma_\mu\gamma_5\lambda^a q$ is the time invariant vector current of Lagrangian. For chiral symmetry, we get

$$< 0|J_\mu^{a,A}(x)|GB^b(p) >= ip_\mu N_p F_{GB} e^{-ipx}\delta^{ab} \tag{3.26}$$

as a consequence of Lorentz invariance and linear realization of SU(3). Now if we identify those Goldstone bosons with our octet of pseudoscalar mesons, our F_{GB} become the *decay constants*. As is well known, decay constants are the form factors, which occur in weak decays of these particles.

3.3 Effective Lagrangian

Very interesting (and also very important) is the idea of an effective Lagrangian formulated by Weinberg [1979]. It is based on the principle of finding the dynamical degrees of freedom that respect certain symmetries and then constructing the Lagrangian from them. The word dynamical was used to show that some degrees of freedom can change with scale. This means that such Lagrangian only works on some scales, below a certain boundary. While it is ineffective above that boundary, below it we can use certain benefits, which can make the theory much easier to work with.

To ensure that the expansion coefficients, called the *low energy coupling constants*, aren't energy dependent, it is necessary to construct the effective Lagrangian as a low energy expansion. In the case of QCD, one needs to make a parallel expansion in quark masses as well. Such a low energy representation of QCD is called *chiral perturbation theory*.

An important thing is that Lorentz invariance permits only even numbered derivatives of the expansion. Written in the form of equation

$$\mathcal{L}_{eff} = \mathcal{L}^{(2)} + \mathcal{L}^{(4)} + \mathcal{L}^{(6)} + ... \tag{3.27}$$

For my calculations, assuming SU(3) symmetry, I will use the standard three flavor effective Lagrangian from Gasser and Leutwyler [1985]

$$\mathcal{L}^{(2)} = \frac{F_0^2}{4}Tr[D_\mu U D^\mu U^+ + \left(U^+\chi + \chi^+ U\right)] \tag{3.28}$$

$$\mathcal{L}^{(4)} = \mathcal{L}^{(4)}(L_1,..,L_{10}) + \mathcal{L}_{WZ}^{(4)} \tag{3.29}$$

and for curiosity

$$\mathcal{L}^{(6)} = \mathcal{L}^{(6)}(C_1,..,C_{90}) + \mathcal{L}^{(6)}_{WZ}(C^W_1,..,C^W_{23}), \tag{3.30}$$

where $U(x)$ and χ are for the purpose of this work

$$U(x) = e^{\frac{i}{F_0}\phi^a(x)\lambda^a} \tag{3.31}$$

$$\chi = 2B_0M. \tag{3.32}$$

D^μ are covariant derivatives, which could possibly allow us to incorporate EM or weak interactions. The non-anomalous part of $\mathcal{L}^{(4)}$ is

$$\begin{aligned}\mathcal{L}^{(4)}(L_1,..,L_{10}) = {} & L_1Tr[D_\mu U^+ D^\mu U]^2 + L_2Tr[D_\mu U^+ D_\nu U]Tr[D^\mu U^+ D^\nu U]+ \\ & +L_3Tr[D_\mu U^+ D^\mu U D_\nu U^+ D^\nu U]+ \\ & +L_4Tr[D_\mu U^+ D^\mu U]Tr[\chi^+ U + \chi U^+]+ \\ & +L_5Tr[D_\mu U^+ D^\mu U(\chi^+ U + U^+\chi)] + L_6Tr[\chi^+ U + \chi U^+]^2+ \\ & +L_7Tr[\chi^+ U - \chi U^+]^2 + L_8Tr[\chi^+ U\chi^+ U + \chi U^+\chi U^+]- \\ & -iL_9Tr[F^{\mu\nu}_R D_\mu U D_\nu U^+ + F^{\mu\nu}_L D_\mu U^+ D_\nu U]+ \\ & +L_{10}Tr[U^+ F^{\mu\nu}_R U F^L_{\mu\nu}],\end{aligned} \tag{3.33}$$

with

$$F^{\mu\nu}_{R,L} = \partial^\mu(v^\nu \pm a^\nu) - \partial^\nu(v^\mu \pm a^\mu) - i\,[v^\mu \pm a^\mu, v^\nu \pm a^\nu] \tag{3.34}$$

$$v_\mu = \frac{1}{2}v^a_\mu\lambda^a,\ a_\mu = \frac{1}{2}a^a_\mu\lambda^a, \tag{3.35}$$

v_μ and a_μ taken from (3.13),(3.14) and (3.15).

As one can see, if one tries to perform calculations using this effective Lagrangian, the next-to-leading order gives us 10 coupling constants and the next-to-next-to leading order 90! Therefore to be able to use this effective Lagrangian, a method to estimate of that many coefficients is needed, because the series doesn't converge fast enough and even the NNLO considerably contributes.

3.4 Resummed approach

The idea of resummed approach comes from Descotes-Genon et al. [2004].

When dealing with a perturbative expansion, very often one takes the first few series members and says that the rest are too small to contribute in a meaningful way. But what if they are not?

For example, let's assume we have a function of 3 variables whose partial derivations aren't in general interchangeable. That means the zeroth term is a constant, the first one has 3 members, the second has 9, the third 27. Not only does the number of coefficients in the expansion grow rapidly, it may also prove more and more difficult to calculate the coefficients of higher orders.

The resummed approach's idea is to resum the remaining members of the series and not neglect it, but calculate with them from then on as one variable called the *higher order remainder*. And then one has to calculate with that remainder algebraically, without reordering the expansion.

The resummed approach thus yields on calculating the higher orders directly, but assumes the sum of those higher orders is equal to the observable itself times an unknown variable. This remainder should of course amount to the difference between the physical value and those few first expansion members. One can then place restrictions on this remainder, when calculating the original observable.

3.5 Decay constant expansions

The decay constants were introduced in (3.26). Now I will show their expansions using the resummed approach (taken from Kolesar and Novotny [2008]).

$$F_\pi^2 = F_0^2(1-4\mu_\pi-2\mu_K)+16B_0\hat{m}(L_4(r+2)+L_5)+\Delta_{F_\pi}^{(4)} \quad (3.36)$$

$$F_K^2 = F_0^2(1-\frac{3}{2}\mu_\pi-3\mu_K-\frac{3}{2}\mu_\eta)+ \\ +16B_0\hat{m}(L_4(r+2)+\frac{1}{2}L_5(r+2))+\Delta_{F_K}^{(4)} \quad (3.37)$$

$$F_\eta^2 = F_0^2(1-6\mu_K)+16B_0\hat{m}(L_4(r+2)+\frac{1}{3}L_5(2r+1))+\Delta_{F_\eta}^{(4)} \quad (3.38)$$

In these equations, $\Delta_{F_{GB}}^{(4)}$ is the higher order remainder of the observable F_{GB}^2,

$$\hat{m} = \frac{(m_u+m_d)}{2} \quad (3.39)$$

$$\mu_i = \frac{m_i^2}{32\pi^2F_0^2}\log\left(\frac{m_i^2}{\mu^2}\right) \quad (3.40)$$

with regularization scale μ and leading order masses

$$m_\pi^2 = 2B_0\hat{m} \quad (3.41)$$

$$m_K^2 = B_0\hat{m}(1+r) \quad (3.42)$$

$$m_\eta^2 = \frac{2}{3}B_0\hat{m}(1+2r). \quad (3.43)$$

Then

$$r = \frac{m_s}{\hat{m}} \quad (3.44)$$

is the mass ratio of the strange to light quarks, and

$$L_4 = -\frac{1}{128\pi^2}\log\left(\frac{\mu}{m_\rho}\right)+L_4^r \quad (3.45)$$

$$L_5 = -\frac{3}{128\pi^2}\log\left(\frac{\mu}{m_\rho}\right)+L_5^r. \quad (3.46)$$

m_ρ sets the scale for L_i^r.

F_0^2, B_0, L_4^r, L_5^r are the low energy coupling constants of the theory. It is also reasonable to define

$$Y = \frac{2\hat{m}B_0}{M_\pi^2}, \quad (3.47)$$

which is the ratio of the squared pion mass at the leading order to its squared physical value. The latest values from Bijnens and Ecker [2014] obtained by standard CHPT at NNLO (with renormalization scale $m_\rho = 0.77GeV$) are in the following table

variable	BE14	free fit
F_0	71 MeV	64 MeV
$L_4^r \times 10^3$	$\equiv 0.3$	0.76(18)
$L_5^r \times 10^3$	1.01(06)	0.50(07)
Y	1.055	0.937

Table 3.1: Values of low energy constants (Bijnens and Ecker [2014])

The fit BE14 preferred by the authors fixes L_4^r by hand, because low values of this constant are in accordance with large N_c. By fixing it, the authors explicitly implemented their assumptions, which in Bayesian approach could be more naturally achieved by using a prior. As we can see the difference in L_5^r that stems from these two different assumptions is much larger than the error from the fit would suggest.

I will use two values of F_η extracted from experiments. The first being

$$F_\eta = (1.18 \pm 0.02) F_\pi = (108.8 \pm 1.8) \text{ MeV} \tag{3.48}$$

(taken from Escribano et al. [2016]) and the other one

$$F_\eta = (1.38 \pm 0.05) F_\pi = (127.3 \pm 4.6) \text{ MeV} \tag{3.49}$$

(taken from Escribano and Frere [2005]). I identify the value of F_η with F_η^8 similarly to Kolesar and Novotny [2008].

These two values of F_η's are obviously incompatible and thus it could be interesting to see how they compare with my theoretical prediction and whether they will give different results for the extraction of the theory's parameters. It should also be noted that both of these results come from Escribano et al. with a ten year gap, it may therefore be reasonable to assume that the lower value is more accurate. The value from year 2016 received new data from BESIII and the authors themselves say that "Moreover, they allowed us to report the most up-to-date results for slope, curvature, and third derivative of the $\eta\prime$TFF, and to update the η-$\eta\prime$ mixing parameters in a mixing scheme compatible with the most general large-N_c ChPT scenario at NLO, thus superseding the values obtained in our previous works..." (Escribano et al. [2016]).

Kolesar and Novotny [2008] also dealt with F_η decay constant in resummed chiral perturbation theory. Their result for F_η obtained by neglecting the dependence on Y and only adding the uncertainties in $\Delta_{F_i}^{(4)}$ in squares to calculate the error gave them the result

$$F_\eta = (1.3 \pm 0.1) F_\pi = (120 \pm 10) \text{MeV}. \tag{3.50}$$

The authors suggest that "$L_5^r < 1 \times 10^{-3}$ and $r \sim 25$ implies $Y > 1.2$, $\Delta_{F_\eta}^{(4)} - \Delta_{F_\pi}^{(4)} > 0.2 F_\eta^2$" using (3.49), so it may be interesting to see how it will compare with my results improved by the Bayesian statistical approach.

3.6 Parameter restrictions (priors)

I placed a restriction on Y - it can only range from 0 to 2.5. This restriction comes from Descotes-Genon et al. [2004] and all of Y's components being positive. This restriction's top boundary depends on the remainders. I will consider Y to be uniformly distributed on this interval.

In our calculations, I will consider the remainders $\Delta_{F_i}^{(4)}$ to be equal to normal distributions with FWHM equal to $0.1 F_i^2$ (estimation inspired by Descotes-Genon et al. [2004]).

I will consider L_5^r to be uniformly distributed in the interval of $< 0{,}2 \times 10^{-3} >$. As the largest value in Bijnens and Ecker [2014] was one half of the top boundary, it doesn't seem that unreasonably large. This interval also includes older values of L_5^r.

4. Results

4.1 One equation

First, I combined (3.36),(3.37) and (3.38) by algebraically substituting for L_4 and L_5 from the first two equations, giving us the equation from Kolesar and Novotny [2008]

$$\begin{aligned} F_\eta^2 &= \frac{1}{3}[4F_K^2 - F_\pi^2 + \frac{M_\pi^2 Y}{16\pi^2}(\log\frac{m_\pi^2}{m_K^2} + (2r+1)\log\frac{m_\eta^2}{m_K^2}) + \\ &\quad +3\Delta_{F_\eta}^{(4)} - 4\Delta_{F_K}^{(4)} + \Delta_{F_\pi}^{(4)}] \end{aligned} \tag{4.1}$$

I used Monte Carlo to calculate F_η by randomly generating 100000 theoretical predictions of F_η's. I used the following values for the parameters - the value of r is from Aoki et al. [2017] and M_π, F_K, F_π are from Tanabashi et al. [2018].

variable	value	σ
r	27.43	0.31
M_π	139.57061 MeV	0.00024 MeV
F_K	110.03 MeV	0.28 MeV
F_π	92.21 MeV	0.15 MeV
$\Delta_{F_K}^{(4)}$	0	$0.1F_K^2$
$\Delta_{F_\pi}^{(4)}$	0	$0.1F_\pi^2$

Table 4.1: Parameter values

Using these values, I got the probability distribution for F_η shown in Fig.4.1. The following mean and standard deviation for F_η were obtained

$$\mu = 118.3 \text{ MeV}, \tag{4.2}$$

$$\sigma = 9.4 \text{ MeV}. \tag{4.3}$$

Comparing that with our experimental data, (108.81 ± 1.84) MeV (3.48) and (127.26 ± 4.61) MeV (3.49), one may notice that it is in between - both values are safely within two sigma confidence level. Also, generating F_η statistically allowed me to improve the precision of the result of Kolesar and Novotny [2008] 120 ± 10MeV.

Then I calculated F_η as a function of the parameter Y, comparing it with experimental values of F_η's in Fig.4.2. It is obvious from Fig.4.2 that there isn't much we can tell about which range of Y's does not fit well with our predictions. This means that our prediction is compatible with experimental values in the whole range of Y. This can be better seen by calculating the probability dependence on Y. Using (1.9) and (1.14), I got the Bayesian probability function of Y for how F_η from Monte Carlo fits the experimental data shown in Fig.4.3. As one can see, larger values of Y's are slightly more compatible for $F_\eta = 1.38F_\pi$ and smaller ones are slightly more compatible for $F_\eta = 1.18F_\pi$. But we cannot put any limitations on Y, because the variation in the probability distribution for different values of Y is rather small.

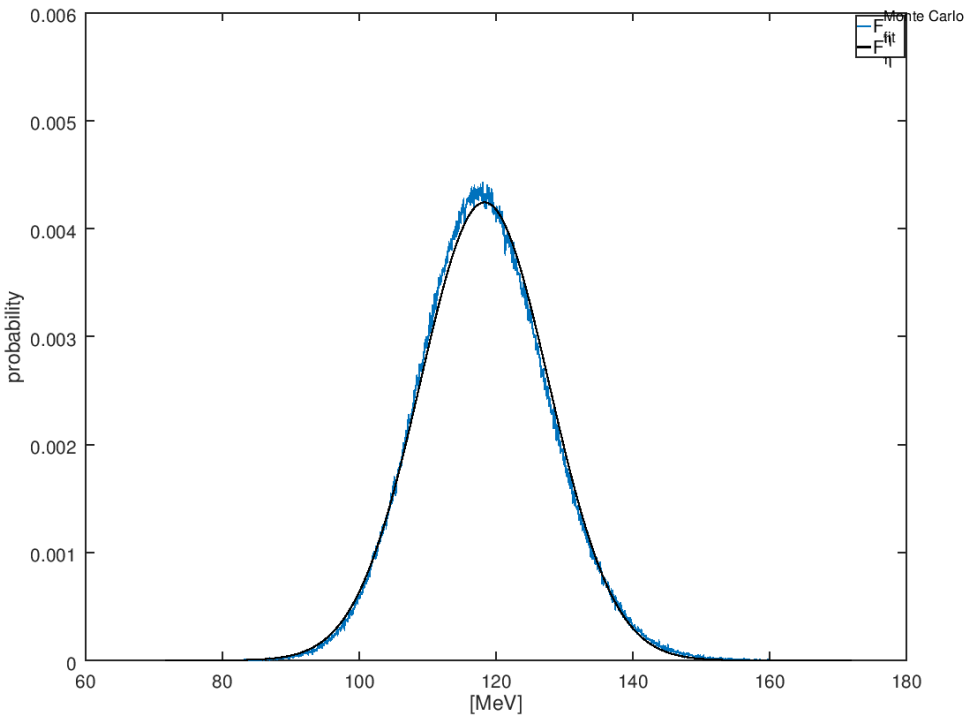

Figure 4.1: Theoretical predictions for F_η generated by Monte Carlo with Gaussian fit, normalized on 0.1 bins

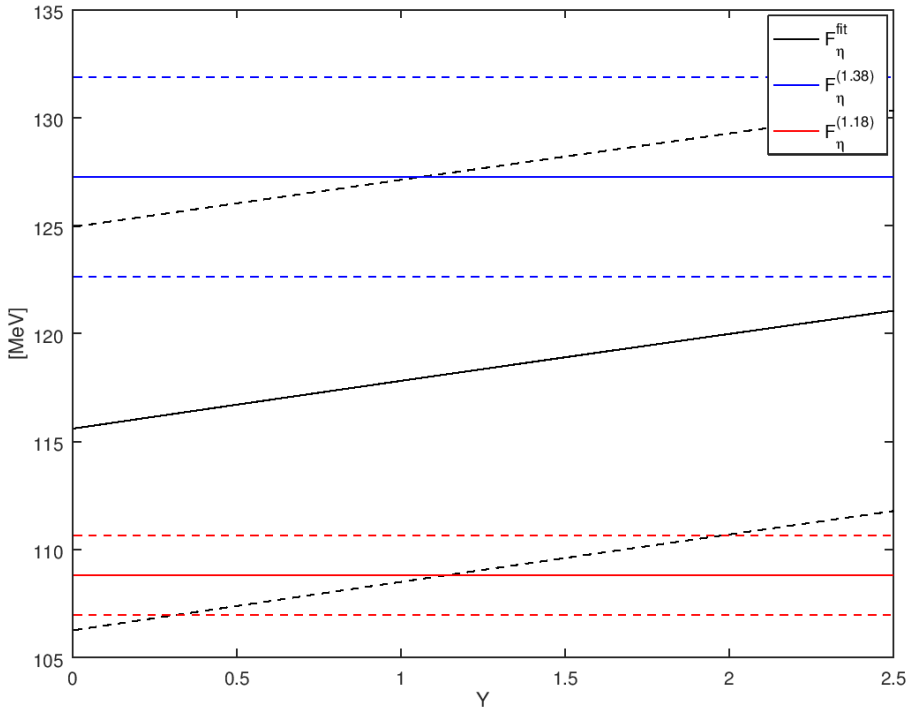

Figure 4.2: Comparison of experimental F_η^{exp}'s (blue and red lines) with the theoretical prediction of F_η (black lines) with dashed lines being one σ confidence levels

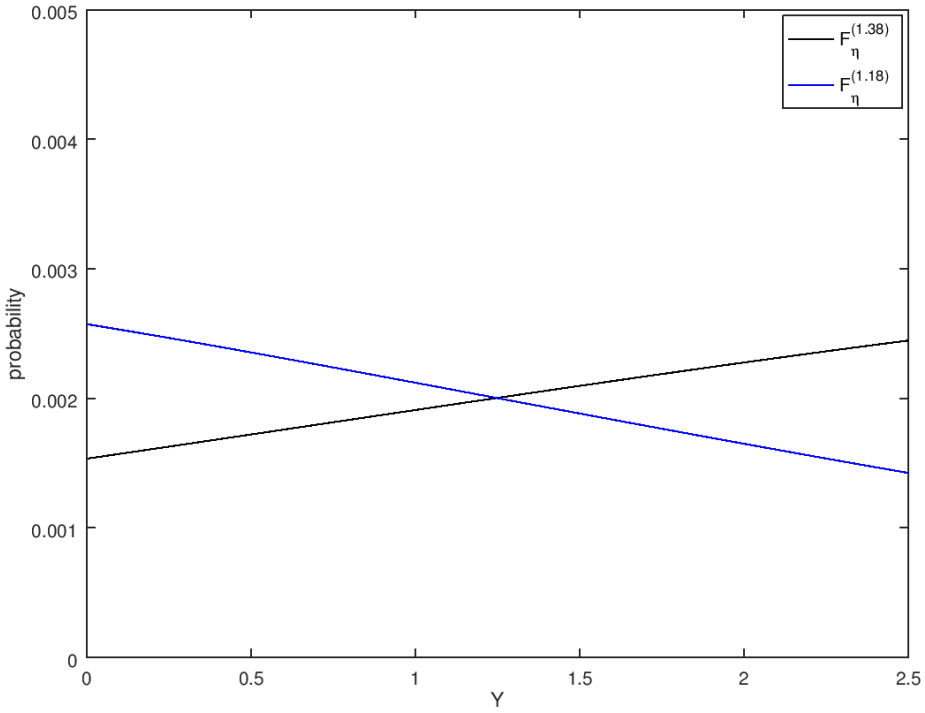

Figure 4.3: Probability of our predictions fitting the experimental data, normalized on bins of 0.005

As a test to see how well we can reconstruct these two experimental values, I calculated the minimum of χ^2 as a function of Y:

$$\chi^2(Y) = \frac{(F_\eta^{th}(Y) - F_\eta^{exp}(Y))^2}{\sigma^2}. \tag{4.4}$$

The graph Fig.4.4 should essentially show us how close we can get with fitting by Monte Carlo in the case of $F_\eta^{(1.38)}$. As can be seen in Fig.4.4, we have no problem to fit our function on the whole scale of Y, so all of our calculations have solid foundations and those performed calculations are valid.

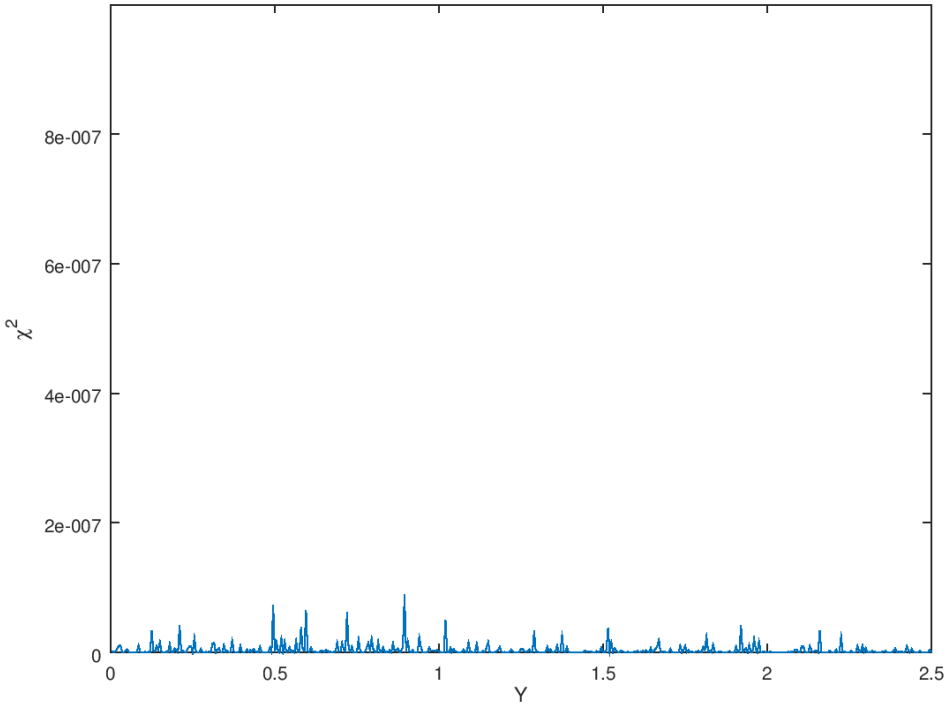

Figure 4.4: Minimum χ^2 for $F_\eta = (1.38 \pm 0.05)F_\pi$

4.2 Two equations

Once again, I started from (3.36),(3.37) and (3.38), but this time I algebraically expressed F_0 from (3.36) and substituted it into (3.37) and (3.38). I also used relation (3.46). This way I got these two following equations for F_K and F_η

$$F_K^2 = \frac{F_\pi^2(1-\delta_{F_\pi}) + \frac{5}{2}\mu_\pi - \mu_K - \frac{3}{2}\mu_\eta + 4YM_\pi^2(\frac{3}{128\pi^2}\log\frac{\mu}{M_\rho} + L_5^r)(r-1)}{1-\delta_{F_K}} \quad (4.5)$$

$$F_\eta^2 = \frac{F_\pi^2(1-\delta_{F_\pi}) + 4\mu_\pi - 4\mu_K + \frac{8}{3}YM_\pi^2(\frac{3}{128\pi^2}\log\frac{\mu}{M_\rho} + L_5^r)(2r-2)}{1-\delta_{F_\eta}} \quad (4.6)$$

with the δ's being normal distributions with FWHM equal to 0.1 and $\Delta_{F_i}^{(4)} = F_i^2\delta_{F_i}$. I set the scale to $M_\rho = 770MeV$. Then I generated F_K and F_η using Monte Carlo with $50Y \times 40L_5^r$ bins with 10000 generated values for each. I plotted 3D graphs for their dependency on Y and L_5^r As we can see, the range of our predictions does include the experimental values of F_K and F_η, but we may receive some limitations for parameters.

Now I used formulas (1.9) and (1.14) to get the probabilities of F_K and F_η in the Bayesian approach. Thus I got Fig.4.7, Fig.4.8 and Fig.4.9. And from Fig.4.7, Fig.4.8 and Fig.4.9 I got the combined probability graphs Fig.4.10 and Fig.4.11. For easier reference, I also plotted σ confidence level graphs. As one can see, these graphs give us certain limitations for parameters, but they differ. I included the F_K-only graph, because there is much smaller uncertainty in value

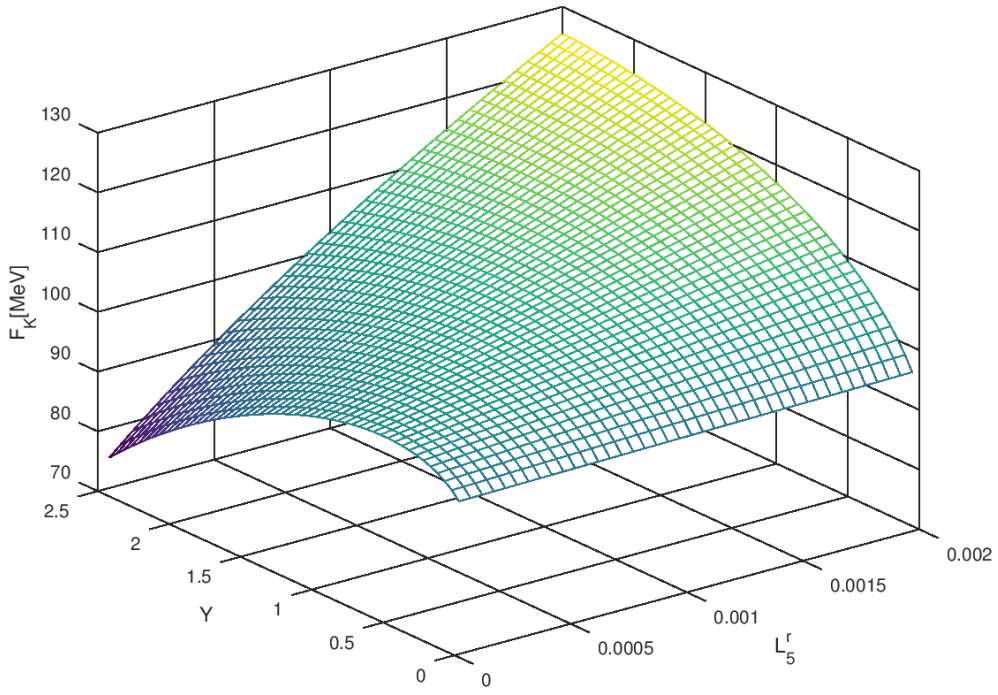

Figure 4.5: F_K^{th} as a function of Y,L_5^r

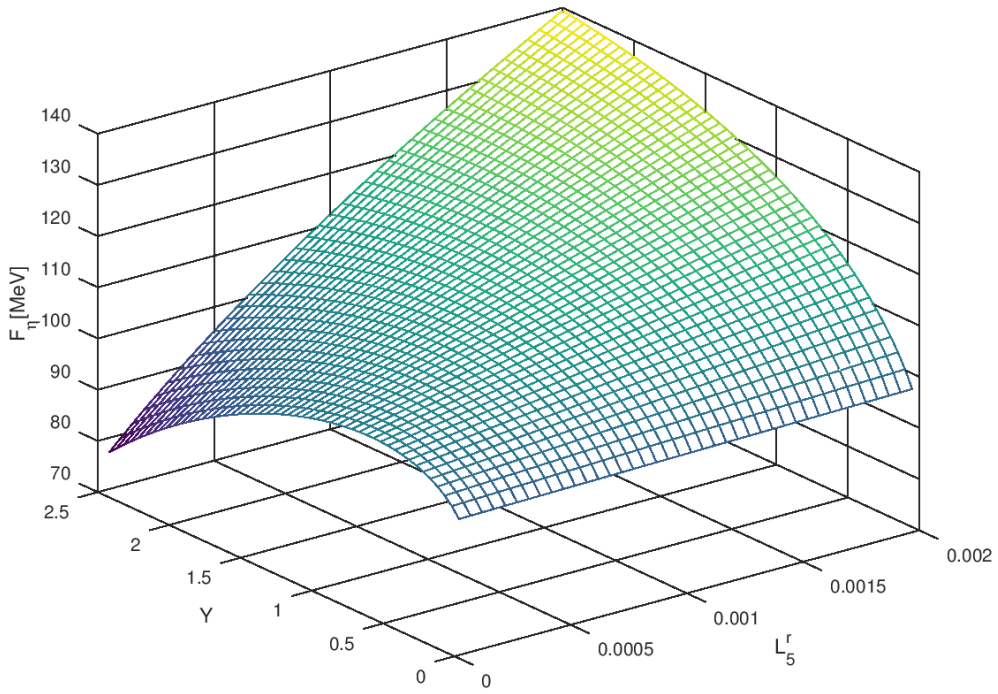

Figure 4.6: F_η^{th} as a function of Y,L_5^r

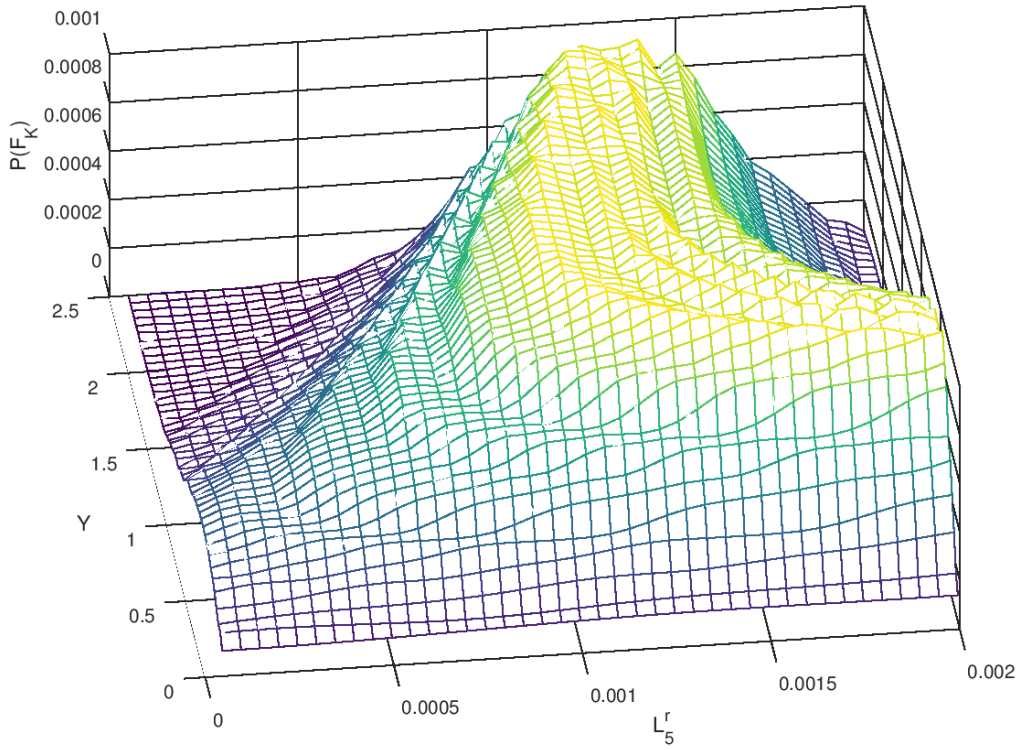

Figure 4.7: $P(Y,L_5^r|F_\eta^{exp})$ as a function of Y,L_5^r for F_K

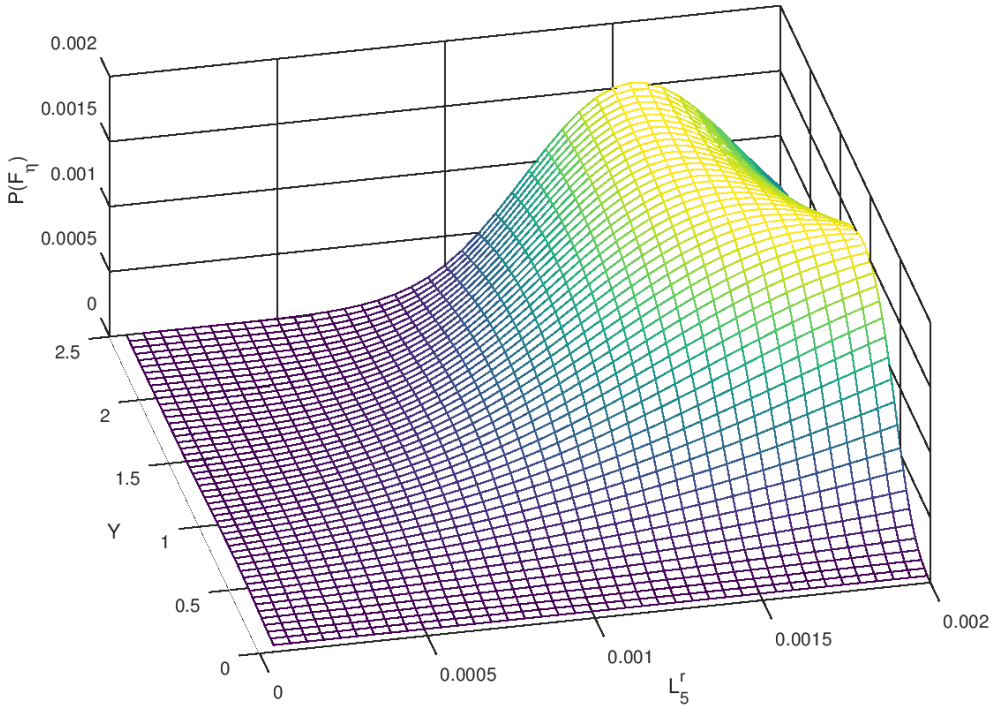

Figure 4.8: $P(Y,L_5^r|F_\eta^{exp})$ as a function of Y,L_5^r for $F_\eta = (1.38 \pm 0.05)F_\pi$

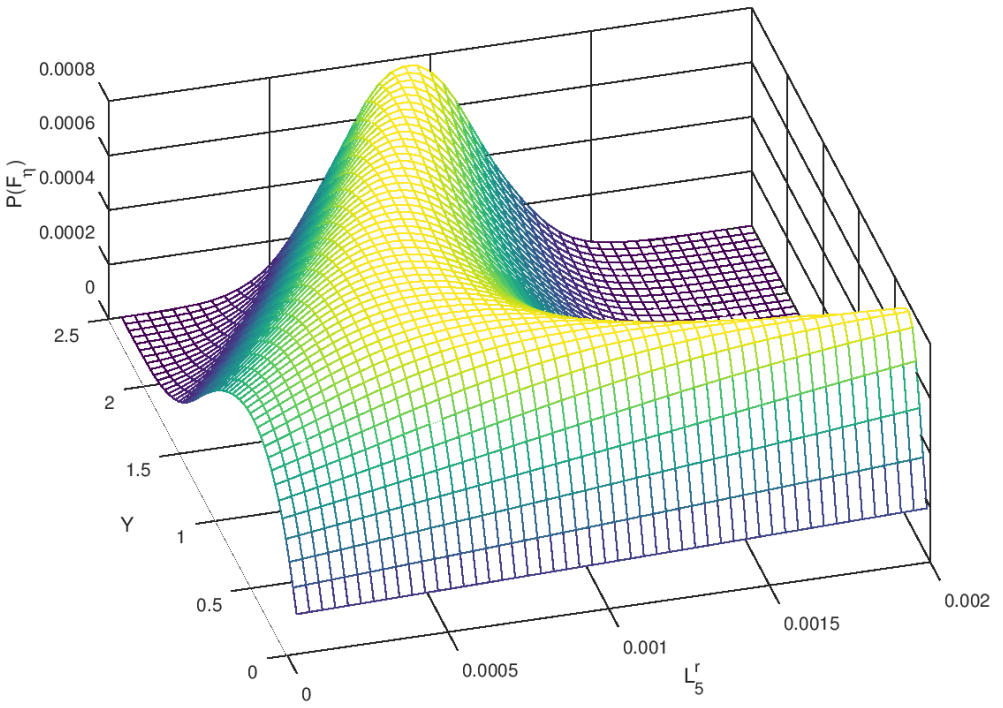

Figure 4.9: $P(Y,L_5^r|F_\eta^{exp})$ as a function of Y,L_5^r for $F_\eta = (1.18 \pm 0.02)F_\pi$

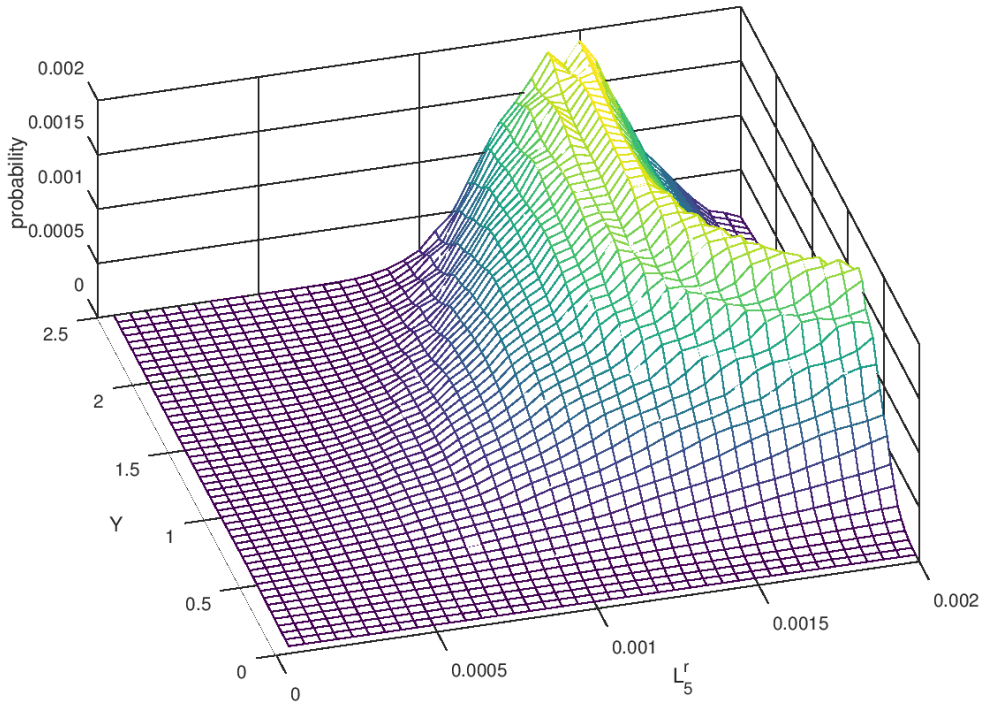

Figure 4.10: $P(Y,L_5^r|F_\eta^{exp})$ as a function of Y,L_5^r for $F_\eta = (1.38 \pm 0.05)F_\pi$

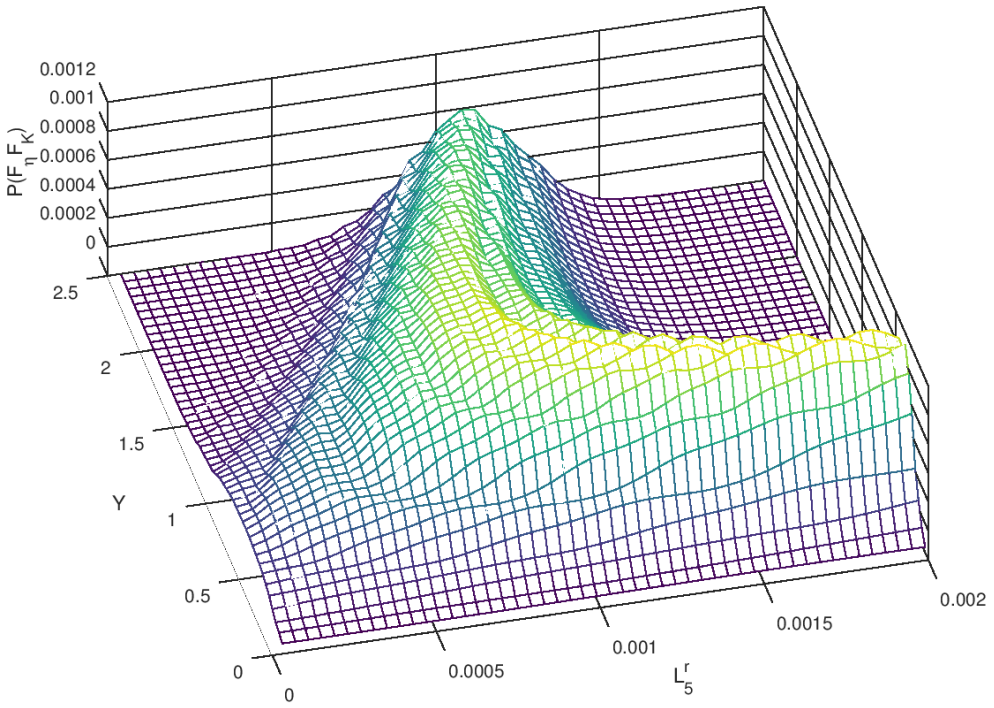

Figure 4.11: $P(Y,L_5^r|F_\eta^{exp})$ as a function of Y,L_5^r for $F_\eta = (1.18 \pm 0.02)F_\pi$

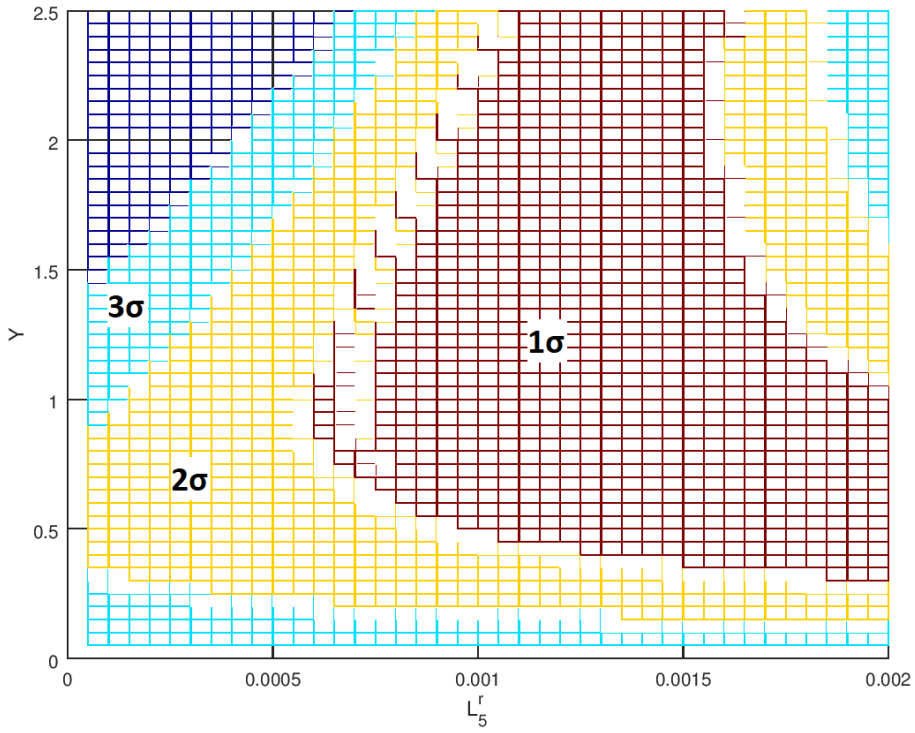

Figure 4.12: Confidence level graph of variables Y,L_5^r for F_K only

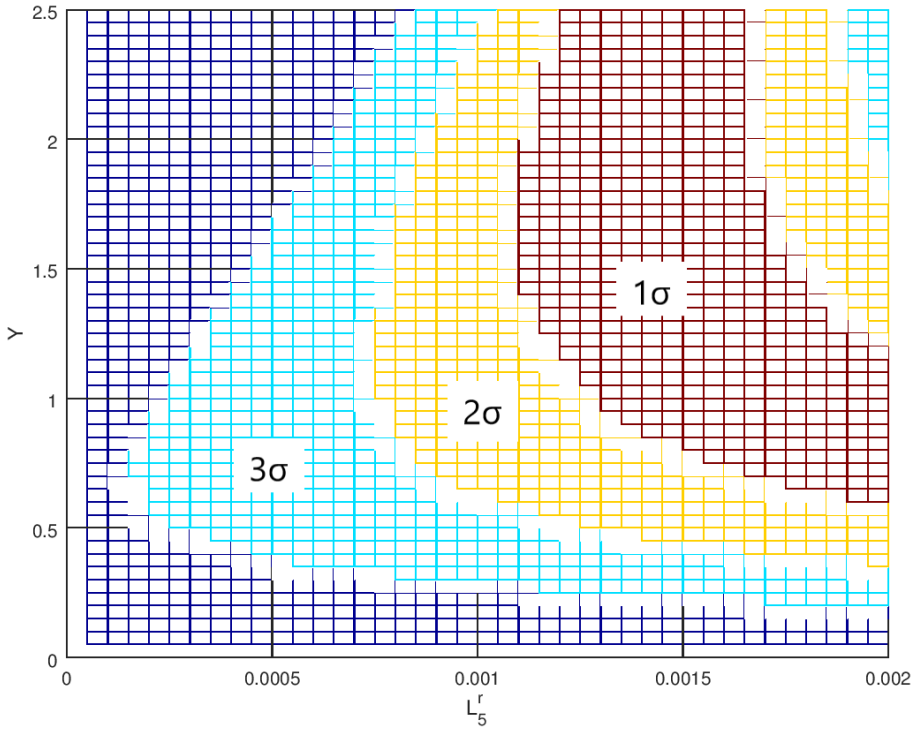

Figure 4.13: Confidence level graph of variables Y,L_5^r for $F_\eta^{exp} = (1.38 \pm 0.05)F_\pi$

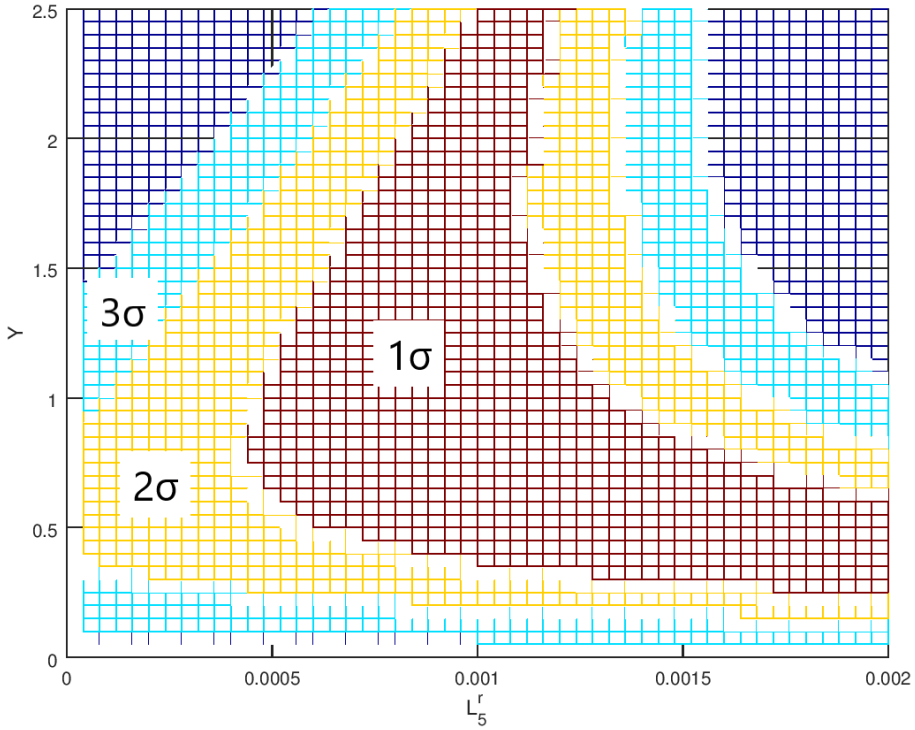

Figure 4.14: Confidence level graph of variables Y,L_5^r for $F_\eta^{exp} = (1.18 \pm 0.02)F_\pi$

of F_K. Notice the significant difference in graphs Fig.4.13 and Fig.4.14. This is natural, because they are based on different input data.

Looking at Fig.4.12, we see that region of $L_5^r < 0.5 \times 10^{-3}$ and $Y > 1.5$ is out of two sigma confidence level (most of it out of three sigma CL).

As for Fig.4.13, an area of $L_5^r < 0.75 \times 10^{-3}$ for all Y's is out of 2 sigma CL, also an area of $Y < 0.25$ for all L_5^r is out of two sigma CL. Summing the graph for all Y actually shows us that we can place the restriction of $L_5^r < 0.9 \times 10^{-3}$ with 2 sigma CL.

Fig.4.14 also shows that region of $L_5^r < 0.5 \times 10^{-3}$ and $Y > 1.5$ is out of two sigma CL and a new region springs up at that region of $L_5^r > 1.5 \times 10^{-3}$ and $Y > 1$.

Looking at all these results together, one may notice an overlay
$L_5^r 10^3 < 0.5$ and $Y > 1.5$. All the results put this region out of 2 sigma CL, therefore this may be a reasonable restriction on our parameters of the theory.

Comparing the confidence level graphs with data from Bijnens and Ecker [2014] in Table 3.1, we can see that their values sit well with my confidence level graphs, except for the free fit compared to Fig.4.13. Their value of $Y = 0.937$ and $L_5^r \times 10^3 = 0.50 \pm 0.07$ is in the 3 sigma CL area, therefore not very compatible with $F_\eta^{exp} = (1.38 \pm 0.05) F_\pi$.

Using Y's from Bijnens and Ecker [2014] on Fig.4.14, we get the restriction $0.1 \times 10^{-3} < L_5^r < 1.6 \times 10^{-3}$ with 2 sigma confidence level.

Comparing the confidence level graphs with Kolesar and Novotny [2008], where they say that "$L_5^r < 1 \times 10^{-3}$ and $r \sim 25$ implies $Y > 1.2$,$\Delta_{F_\eta}^{(4)} - \Delta_{F_\pi}^{(4)} > 0.2 F_\eta^2$", we can see that it is too restrictive and such limitations do not follow from my confidence level graph for $F_\eta^{exp} = (1.38 \pm 0.05) F_\pi$. In Fig.4.13 (which has the same input of F_η^{exp}), for $L_5^r = 0.001$, Y has to be greater than 0.6 to be in the 2 sigma significance region.

Summary

I introduced the reader to the Bayes' theorem and I have shown how different assumptions can lead to different results. I broached the subject of strong interactions at low energies, which I tried to explain using group theory. I introduced the idea of effective Lagrangians and used the not so common resummed approach to chiral perturbation theory. Then I applied all of this on F_π, F_K and F_η decay constant expansions using two different experimental values for F_η. Then I tried to extract limitations on the parameters of the theory using the Bayesian statistical approach.

I tried two methods, the first one was combining all equations into one for F_η. I used Monte Carlo to generate theoretical predictions of F_η, giving me results for the F_η of

$$F_\eta = (118.3 \pm 9.4) \text{ MeV}.$$

The distribution turned out to be an almost perfectly Gaussian. This result wasn't in conflict with either of the two experimental values. Then I used the Bayesian approach to calculate the probability of my theoretical predictions describing the experimental data as a function of parameter Y. Unfortunately, I didn't get any clear restrictions for Y as the dependence turned out to be weak. To confirm that it was indeed correct to use these methods, I calculated the minimum χ^2 to show how well I am able to fit experimental values with my theoretical predictions.

The second approach was to use two equations for F_K and F_η, now with two free parameters Y and L_5^r, the low energy coupling constants. I generated a 2D graph for F_K and F_η and created graphs that would show us the parameter probabilities. Then I converted these 2D graphs into significance graphs and found out that a certain area of that two-dimensional space was out of 2 sigma confidence level for all three constructed sigma CL graphs. The results were compatible with latest results in literature, except for the older and larger value of $F_\eta = (1.38 \pm 0.05)F_\pi$ which seems to be incompatible with $L_5^r < 0.9 \times 10^{-3}$.

Using $Y = 1$ in conjunction with the latest results in literature and the newer experimental value $F_\eta = (1.18 \pm 0.02)F_\pi$, I received a limitation

$$0.1 < 10^3 L_5^r(M_\rho) < 1.6 \text{ at } 2\ \sigma \text{ confidence level.}$$

Literature References

S. Aoki et al. Review of lattice results concerning low-energy particle physics. *Eur. Phys. J.*, C77(2):112, 2017. doi: 10.1140/epjc/s10052-016-4509-7.

Johan Bijnens and Gerhard Ecker. Mesonic low-energy constants. *Ann. Rev. Nucl. Part. Sci.*, 64:149–174, 2014. doi: 10.1146/annurev-nucl-102313-025528.

G. D'Agostini. Bayesian reasoning in high-energy physics: Principles and applications. 1999.

S. Descotes-Genon, N. H. Fuchs, L. Girlanda, and J. Stern. Resumming QCD vacuum fluctuations in three flavor chiral perturbation theory. *Eur. Phys. J.*, C34:201–227, 2004. doi: 10.1140/epjc/s2004-01601-4.

John F Donoghue, Eugene Golowich, and Barry R Holstein. *Dynamics of the Standard Model*. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge Univ. Press, Cambridge, 1992. URL `https://cds.cern.ch/record/238727`. Includes exercises.

Rafel Escribano and Jean-Marie Frere. Study of the eta - eta-prime system in the two mixing angle scheme. *JHEP*, 06:029, 2005. doi: 10.1088/1126-6708/2005/06/029.

Rafel Escribano, Sergi Gonzàlez-Solís, Pere Masjuan, and Pablo Sanchez-Puertas. $\eta\prime$ transition form factor from space- and timelike experimental data. *Phys. Rev.*, D94(5):054033, 2016. doi: 10.1103/PhysRevD.94.054033.

J. Gasser and H. Leutwyler. Chiral Perturbation Theory: Expansions in the Mass of the Strange Quark. *Nucl. Phys.*, B250:465–516, 1985. doi: 10.1016/0550-3213(85)90492-4.

Howard Georgi. Lie algebras in particle physics. *Front. Phys.*, 54:1–320, 1999.

J. Goldstone, A. Salam, and S. Weinberg. Broken Symmetries. *Physical Review*, 127:965–970, August 1962. doi: 10.1103/PhysRev.127.965.

Marian Kolesar and Jiri Novotny. The eta decay constant in 'resummed' chiral perturbation theory. *Fizika*, B17:57–66, 2008.

M. Kolesár. Chiral perturbation theory and the low energy phenomenology of pseudoscalar mesons. *Doctoral Thesis*, 2008.

M. Tanabashi et al. Particle data group - review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018. doi: 10.1103/PhysRevD.98.030001. URL `https://link.aps.org/doi/10.1103/PhysRevD.98.030001`.

W.J. Thompson. *Angular Momentum: An Illustrated Guide to Rotational Symmetries for Physical Systems*. A Wiley-Interscience publication. Wiley, 1994. ISBN 9780471552642. URL `https://books.google.cz/books?id=O25fXV4z0B0C`.

Steven Weinberg. Phenomenological Lagrangians. *Physica*, A96(1-2):327–340, 1979. doi: 10.1016/0378-4371(79)90223-1.

List of Figures

List of Tables