UNIVERSITY OF GRONINGEN

BACHELOR THESIS

Track Finding in Physics

A review of existing methods and an exploration of two new methods

Author:
M. Schut

Supervisor:
Dr. Ir. C.J.G. Onderwater

Second Examiner:
Prof. R.G.E. Timmermans

July 11, 2017

Abstract

The goal of this research paper is to discuss and review several statistical methods used for track finding in (high energy) physics. The method discussed is the Likelihood Method, along with two variations: using a Corridor and so-called Tukey weights. The Likelihood Method has a problem with outliers, they have a large influence on the result. Using one of the mentioned variations reduces this influence, but they use arbitrary parameters that influence the result or are prone to small changes in the hypothesis. These problems will be discussed. Furthermore, two new methods will be offered (the Unlikelihood Method and the Ratio of Distances Method) that may tackle the problems of the earlier methods. The Unlikelihood Method will show to be a prospective method for track finding, the Ratio of Distances Method will not.

Acknowledgements

I would like to thank Gerco Onderwater for his supervision these last two months, for always being available when needed and for his feedback. I would also like to thank Jesse Knol for taking the time to read the earlier versions of this thesis and for his efforts in making the figures in this paper look perfect.

Contents

Introduction and Problem Sketch

Detectors in physics cannot detect particle trajectories, only locations (points in space) at a certain time. When a particle passes through a detector plate, a "hit" is recorded, consisting for instance of the location and time at which the particle interacted with the detector. Many particles (meaning hundreds) passing through a detector plate result in many detected hits in the detector. Many particles passing through many detector plates will result in even more detected hits in the detector. In the last situation it is difficult to reconstruct the trajectory of each individual particle through the detector (see figure 1). Track finding in physics concerns itself with reconstructing trajectories from these hits such that they match the original trajectories of the particles. [A. Strandlie, 2009]

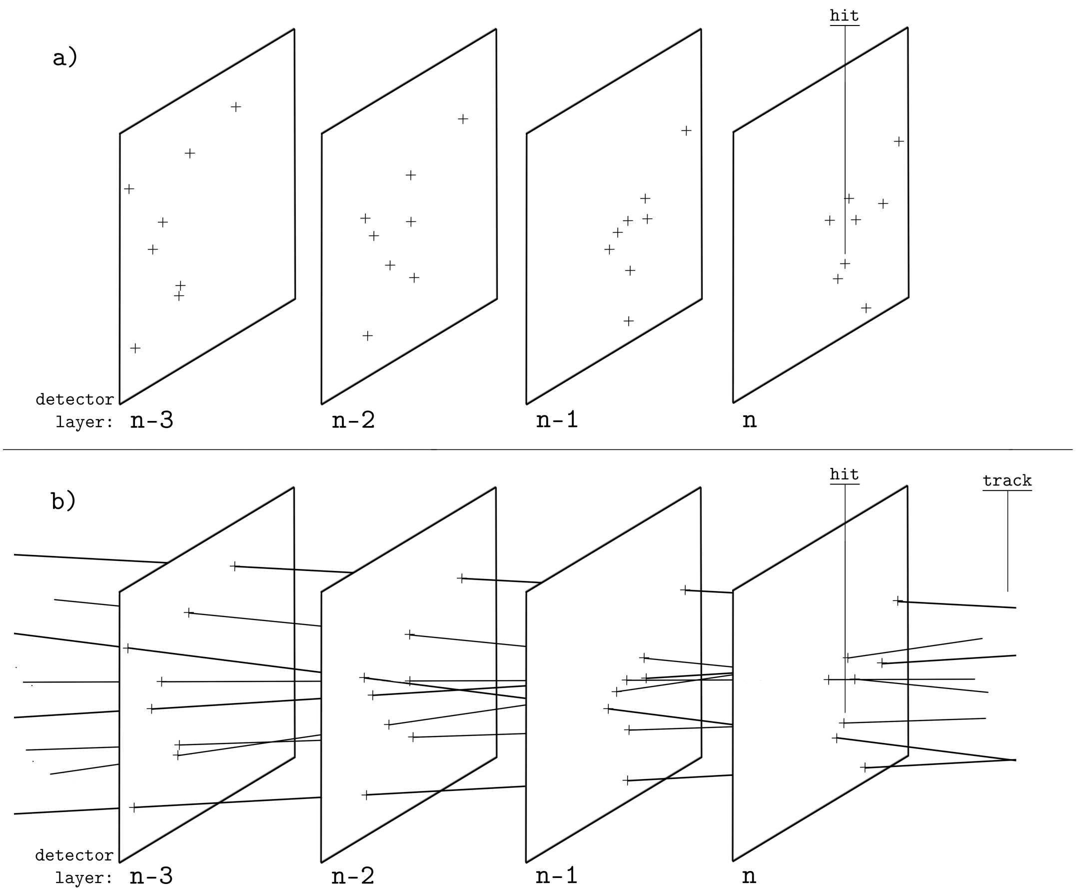

FIGURE 1: The figures show that it is difficult to reconstruct the original particle's trajectory from the detected hits if there are multiple particles passing through the detector layers. In this illustration only eight particles pass through the detector plates, in high energy physics this number will be much larger. a) An illustration of multiple hits detected in multiple detector layers. b) An illustration of the original trajectories of the eight particles passing through the detector.

Track finding methods used nowadays are often based on competition between several hypotheses, the result of the competition depends on the detected hits [A. Strandlie, 2009, p.6]. The methods discussed in the next chapters also use hypothesis competition to find the most likely trajectory of the particle. To find the best hypothesis, a so-called quality parameter is assigned to each hypothesis [R. Frühwirth, 2000, p.159-162]. Bayesian statistics is used in this paper to calculated a quality parameter.

Track finding is important in for example particle physics. In detectors such as the LHC Detector in CERN, track finding is used to reconstruct the original trajectories of high-energy particles, which helps to find all the characteristics of the particles passing through the detector (see Section 1.1). In the LHC Detector this has led to the discovery of new particles such as the Higgs Boson [Cho, 2017].

This paper investigates the track finding method called the Likelihood Method, which uses the statistical concept "likelihood" to check whether a proposed track is close to the true trajectory of the particle through the detector. There are often many trajectories in the detector, this method aims to reconstruct them all which is difficult because there is no a priori knowledge about the number of trajectories (due to particles reacting with each other in the detector for example). Because there is no a priori knowledge about the number of trajectories, each hypothesis needs to be measured separately against the whole data set, even though only a small part of the data belongs to that hypothesis. This causes a problem for the Likelihood Method, namely that the data that does not belong to the trajectory can influence the result such that the true trajectory of a particle can appear as a bad hypothesis. The datapoints that are not part of the trajectory can cause a hypothesis to appear as a bad estimation of the truth even though it is a good fit of the datapoints that are part of the trajectory. This will be explained in Section 2.1.

There are two variations on the Likelihood method that will be discussed, the so-called Tukey weight and the Corridor. These variations solve the problem that the Likelihood Method suffers from, but use arbitrary parameters which may bias the result, which is a new problem. The difficulty is to design a method that can find the one true hit belonging to the particle out of many other hits in each detector layer. The research question for this paper is thus:

"Can we find a quality parameter that finds the true set of tracks and is robust against outliers?"

A proposal for a new method was put forward in an unpublished draft by G. Onderwater ["Retina Thoughts"]. This method was supposed to tackle the problems described above. A method based on this proposal will be discussed in Section 3.1.

This paper is composed of four chapters, the Chapter 1 is an introduction to the definitions used in this paper and into track finding. The tracking methods discussed here are possible in many detectors, but the LHCb VELO detector is taken as a general example throughout this paper and will be shortly discussed in Section 1.1. Although the methods are discussed in light of the LHCb VELO detector, they can be used more generally in physics. Chapter 2 provides an evaluation of current methods (such as the Likelihood Method). In the Chapter 3 and Chapter 4 two new methods are introduced and discussed. Chapter 3 is based on the unpublished draft by Onderwater (the Unlikelihood Method). Chapter 4 is a completely new method, based on the problems that the other methods run into (Ratio of Distances Method). The new methods will show to have (dis)advantages once under investigation.

Chapter 1

Introductory Remarks

This Chapter serves as an introduction into track finding methods and statistics. More background on statistics can be found in Appendix A

1.1 Example Application: The LHCb Detector

An example of an experiment where track finding is important but difficult is the Large Hadron Collider Beauty (LHCb) experiment, which is part of the Large Hadron Collider (LHC) at CERN [*The LHCb Detector*]. Only the LHCb VELO detector is discussed as a general example because this is what a typical detector relevant to this paper looks like. This section provides some background information on how such a detector works.

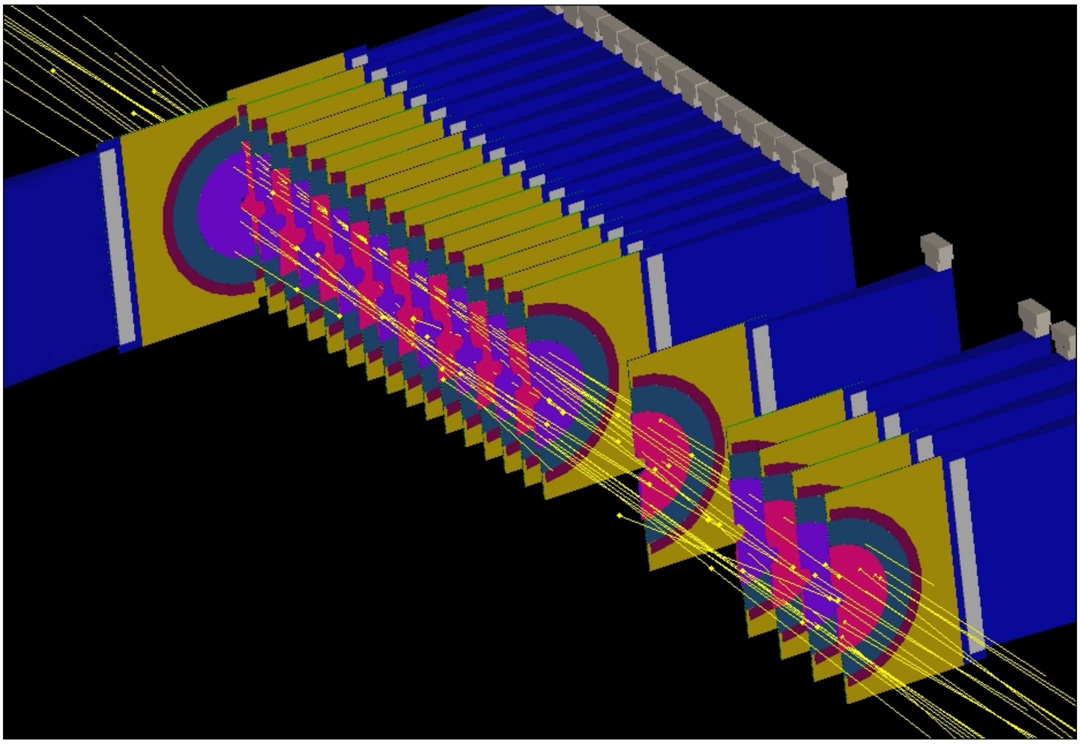

FIGURE 1.1: This illustration shows the LHCb VErtex LOcator (VELO). Many particles pass through the detector, it can be difficult to reconstruct the trajectories shown in the figure from the detected hits. [*The latest from the LHC*]

In the LHC beams of protons collide, producing many different particles. The LHCb experiment records beauty and anti-beauty quarks that can be found in the decay of particles. The VELO part of the LHCb detector (shown in figure 1.1) has 84 silicon sensors which have the shape of half moons. Theses silicon sensors are paired such that the VELO part consists of 42 silicon detector plates in a row in the direction of the beam [*LHCb installs its precision silicon detector, the VELO*]. Just after the collision of the two proton beams the produced particles have straight trajectories that do not deviate much from the original $\hat{z}$-direction of the incoming beam. Due to the absence of magnets most trajectories will go in a straight line in the $\hat{z}$-direction (see figure 1.1). The trajectories can be reconstructed using the detector plates set up in a row. Connecting the hit of a particle in detector plate n with the hit of that particle in detector plate $n+1$ will eventually tell what particle it was. This way the production and decay of particles produced in the collision can be studied. Due to the high energy of the particles in the LHCb detector, new particles and antiparticles can be created randomly, so the VELO detector can also provide information about new particles that come into existence. [*The LHCb Detector*], [*LHCb installs its precision silicon detector, the VELO*]

Many particles (hundreds) pass through the small VELO detector, this makes it extremely difficult to reconstruct the trajectories of the particles.

1.2 Track finding and Statistics Definitions

This section will discuss some definitions necessary to understand the later chapters. Figure 1.2 shows the definitions listed below as well.

- Hit: an intersection between a particle trajectory and a predefined plane fixed at some location in space, z. It is assumed that a single track will generate a single hit in a plane.
- Track: a straight path connecting at least five consecutive hits.
- Detector inefficiencies: hits that are not detected by the detector when a particle passes through it.
- Data: is denoted as $\{x\}$ and represents all hits detected by the detector. A single hit is denoted x_i.
- Hypothesis/proposed track: is denoted as μ and represents the proposed trajectory which needs to be tested against the data.
- True track: the true trajectory of the particle through the detector, it is unknown to the observer.
- Qualifier/quality parameter: is denoted as $\mathcal{Q}$. It is a parameter assigned to the hypothesis to decide how likely the hypothesis is.
- "good" track: A very likely hypothesis. The track is assigned a high likelihood value for example.
- "bad" track: A very unlikely hypothesis. The track is assigned a low likelihood value for example.
- Many Track Problem: the problem of there being a lot of tracks in the detector. This makes it difficult to find a qualifier that reconstructs the individual tracks based on the many hits in the detector layers.

- Outlier: a hit that is not close to the hypothesis, for example a hit from another track that lies a large distance from the proposed track.
- Outlier Problem: a problem in track finding where outliers have a large influence on finding the right track which causes inaccurate qualifiers.
- Arbitrary Parameter Problem: a problem in track finding where a method uses parameters for which there is no preferred value. Thus the parameter gets assigned an arbitrary value which may bias the result.

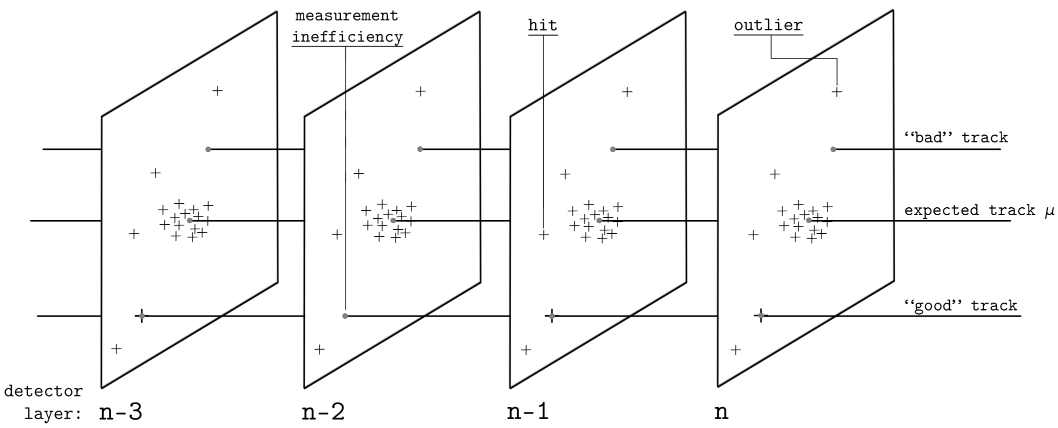

FIGURE 1.2: A visual representation of some important definitions listed in Section 1.2. The symbols shown in this figure for the definitions will be used in more figures in this paper.

For our discussion we will use the LHCb VELO detector. This is done for simplicity, there is no loss of generality. In the VELO part of the LHCb detector the trajectory through the detector is assumed to be straight due to the absence of magnets, which can bend the trajectory of a charged particle [*VELO*]. Due to this the trajectories of the particles in the VELO part have a small angle with respect to the incoming proton beam along the $\hat{z}$-axis. The most likely trajectory of the particle is found by proposing many possible tracks "μ" and doing the same calculation for each possible track. Comparing the results gives the most likely track.

A short list summarizing the notation concerning probabilities, as used in this paper [D. S. Sivia, 2006, p.3-13]:

- $P(A \mid B)$ denotes the probability of getting A given that B is true.
- Prior probability: the uncertainty of the proposed track before the data and given some parameters, $P(\mu \mid \sigma, I)$.
- I: the background information, describing any known or unknown parameters.
- $\neg$: the negation of something. For example: $\neg x$ is the negation of x, it means *not* x.

These notations and definition will be used throughout this paper.

1.3 Track finding: a general understanding

To conclude this introduction into the track finding, this section will explain the way in which a hypothetical track is created. There are many tracks going through the detector (especially in high-energy physics detectors). To find the right set of tracks, several sets are investigated. These sets consist of individual fits (lines fitted to match the data). An example of two competing sets of proposed tracks is given in figure 1.3. In the figure one can see that there are two hypotheses: the black set of tracks and the red set of tracks. To decide which one is better, every track is assigned a quality parameter. A good track finding method would assign a better quality parameter to the black set of tracks than to the red set of tracks. The difficulty is finding such a quality parameter. [A. Strandlie, 2009, p.16-17]

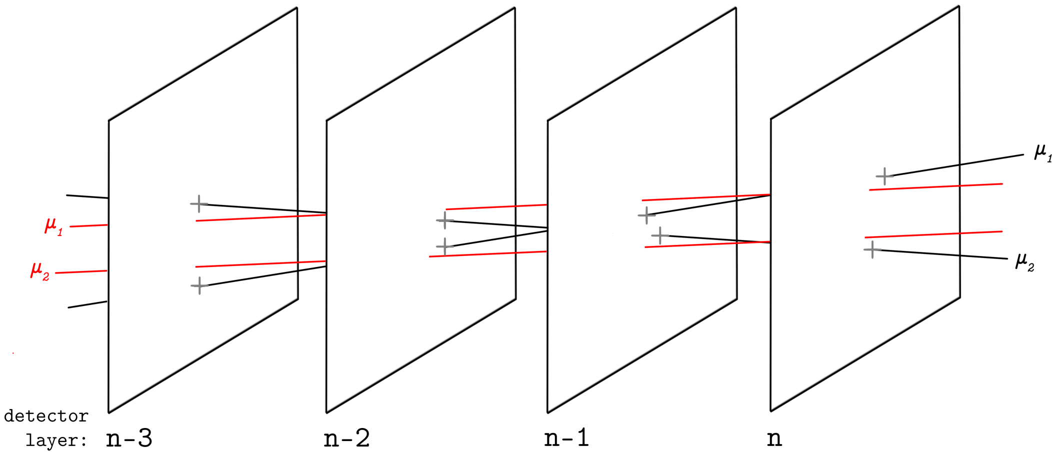

FIGURE 1.3: Two competing sets of tracks: the red set of tracks (μ_1 and μ_2) and the black set of tracks (μ_1 and μ_2). These sets both take all detected hits into account and tracks in the same set do not use the same hits. The black set of tracks is better then the red set of tracks because it is closer to the detected hits.

The situation in figure 1.3 is simplified, in reality there are many more tracks due to the many particles going through detectors in high-energy physics. Therefore there are many different hypotheses. These hypotheses can be created by taking different starting points (hits) and varying the slope of the straight lines. The track has a horizontal slope and a vertical slope so in reality it is not as simple as described here. Using vector decomposition (see figure 1.4) this can also be done for three dimensions. [F.M. Dekking, 2005, p.329-336]

The position of a particle at a time t is given by the initial position (x_0, y_0, z_0) and the distance it traveled calculated from the time (t) and its velocity (v_x, v_y, v_z). The particle's trajectory has a small angle with respect to the proton beam along the $\hat{z}$-axis. Therefore, instead of a time t, the distance $z - z_0$ can be used to find the particle's location and the velocities can be substituted by slopes.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} t = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + \begin{pmatrix} u_x \\ u_y \\ 1 \end{pmatrix} (z - z_0)$$

u_x and u_y are the slopes of the trajectory in the $\hat{x}$- and $\hat{y}$-direction and $u_z \equiv 1$ such that it is just the propagation in the $\hat{z}$-direction.

$$u_x = \frac{v_x}{v_z}, \qquad u_y = \frac{v_y}{v_z}, \qquad u_z = \frac{v_z}{v_z} = 1.$$

Multiplying the slope u_x with the distance traveled in the $\hat{z}$-direction gives the distance traveled in the $\hat{x}$-direction, as one can see in figure 1.4. A detector plate at z is typically placed at a fixed distance $(z - z_0)$ with respect to the starting point (x_0, y_0, z_0). The track is thus specified by five parameters: x_0, y_0, z_0, u_x and u_y. Varying u_x and u_y in some way for each point (x_0, y_0, z_0), gives many possible tracks. These are continuous variables so there are infinitely many options, but this gives all the possible tracks to some degree of preciseness.

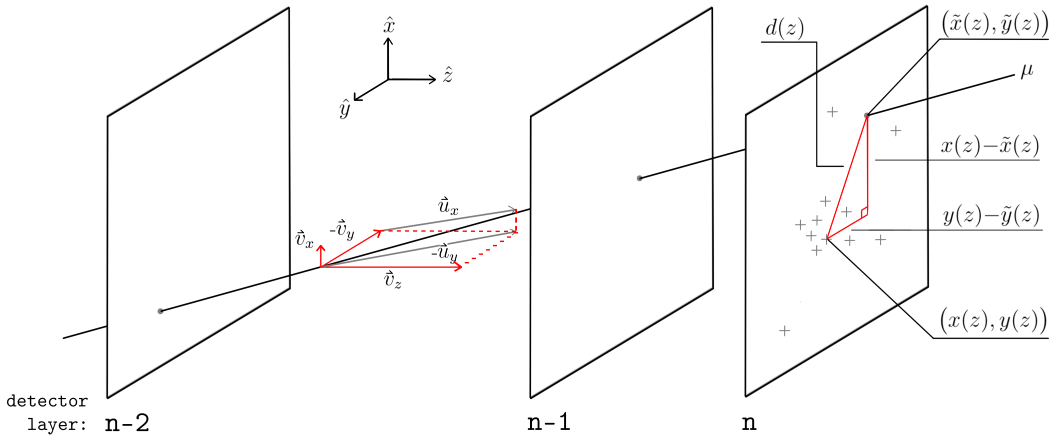

FIGURE 1.4: Vector decomposition between detector layer $n-2$ and $n-1$. The velocity is decomposed into a velocity in the $\hat{x}$- $\hat{y}$ -and $\hat{z}$-direction, $\vec{v} = (v_x, v_y, v_z)$. The slopes u_x and u_y are derived from this vector decomposition. In layer n the definition of the distance, $d(z)$ defined in equation 1.1, is visualized.

The true track has some slope and some initial point and thus corresponds to a point in this five-dimensional space with some value for each of the five parameters. A set of tracks will correspond to a set of points. What values the parameters should have and thus which points a set of tracks should correspond to can be extracted from the data. How well the match between a proposed point in this five-parameter space and the data is will be decided by the qualifier. The distance between a data point and the proposed track in the same plane as defined in equation 1.1 is often used as a qualifier. [A. Strandlie, 2009, p.16-17]

$$d(z) \equiv \sqrt{(x(z) - \bar{x}(z))^2 + (y(z) - \bar{y}(z))^2} \tag{1.1}$$

Here $x(z)$ and $y(z)$ give the measured position of the particle and $\bar{x}(z)$ and $\bar{y}(z)$ give the proposed position of the particle in layer z. This definition is illustrated in figure 1.4. Equation 1.1 above can be linked to the expression of χ^2 as discussed in Appendix A.5.

The distance shows how close the hypothetical and measured positions are. If they are close, then the hypothesis is close to the truth, but if they are not close then the hypothesis is not close to the truth. The distance $d(z)$ is often used to find a qualifier. The methods discussed in later chapters have a probability density function that also uses this distance as a measure of how likely the proposed track is.

When working with these probability distributions, only the two dimensional case is discussed for simplicity.

Chapter 2

Overview of Track-finding Methods

To reconstruct the trajectory of a particle trough the detector, we will use hypothesis testing. A track (some point in the five parameter space) is proposed and investigated, this is done for many tracks resulting in the most likely track. The "investigation" of the proposed tracks can be done with several methods, among which is the Likelihood Method. This method uses the concept of likelihood to find the most likely track. It has a disadvantage, namely that it is heavily influenced by outliers. Two variations have therefore been proposed, the Corridor Method and Tukey weights, that try and tackle this Outlier Problem. Disadvantages of these methods are that they use arbitrary parameters that influence the outcome.

2.1 Likelihood Method

In this section I assume a basic knowledge of Bayesian statistics, likelihood and probability density functions. Background information about this can be found in Appendix A.

Once the hypotheses have been constructed, they need to be tested. A decision criterion is needed to decide whether a hypothesis is correct or not, therefore a qualifier is assigned to each hypothesis, which is based on the hypothesis and the data. For the Likelihood Method this qualifier is the likelihood ($\mathcal{L}$). The likelihood is calculated using Bayes theorem (see Appendix A)

$$\begin{aligned}\mathcal{L} &\equiv P(\mu \mid \{x\}, \sigma, I) \\ &= P(\{x\} \mid \mu, \sigma, I) P(\mu \mid \sigma, I)\end{aligned}$$

The probability, $P(\mu \mid \{x\}, \sigma, I)$, for the proposed track μ to be a "good" track given the data $\{x\}$ can be calculated with the formula above. The likelihood is calculated with the probability that each detected hit x_i is a part of the proposed track. Meaning that *and* for the first, *and* for the second, ..., *and* for the n^{th} hit the data supports or does not support the proposed track:

$$\mathcal{L} = P(x_1 \mid \mu, \sigma, I) P(x_2 \mid \mu, \sigma, I) \cdots P(x_n \mid \mu, \sigma, I) P(\mu \mid \sigma, I) \tag{2.1}$$

The qualifier is the value of the likelihood. For the calculation of the likelihood we need to know whether the data supports the idea that the hits are caused by the proposed track, parametrized by μ. Formulated differently, for each hit we ask the question: "is this hit part of the proposed track?" The answer to this question is a probability: the probability of getting the detected hit given the proposed track, the width of the Gaussian distribution and some background information, of which the limits are shown here:

$$P(x_i \mid \mu, \sigma, I) = \begin{cases} 1 & \text{yes, } x_i \text{ is certainly part of the proposed track} \\ 0 & \text{no, } x_i \text{ is certainly not part of the proposed track} \end{cases}$$

the probability $P(x_i \mid \mu, \sigma, I)$ is calculated by assuming a Gaussian distribution for x_i (the location of the proposed track provides the value of μ in this distribution) as can be seen in figure 2.1.

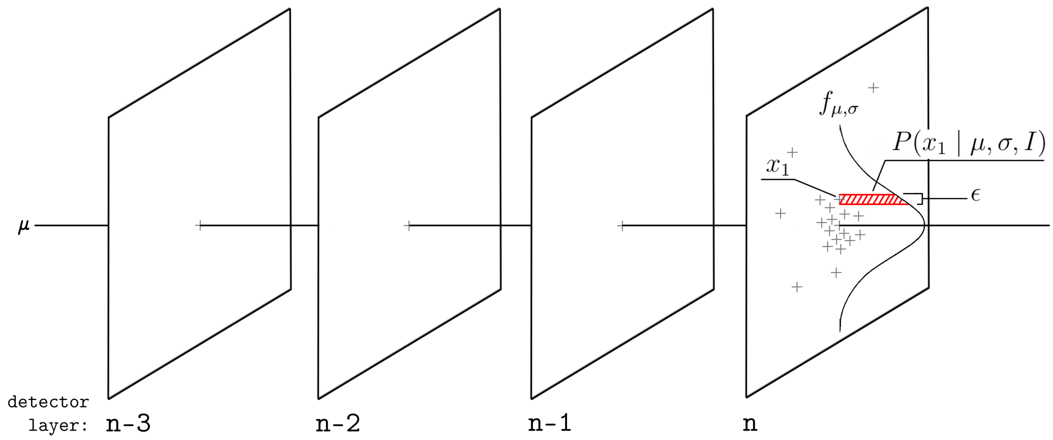

FIGURE 2.1: Each track is assigned a qualifier, the likelihood in this case. The likelihood is calculated by finding the probabilities $P(x_i \mid \mu, \sigma, I) = \epsilon f_{\mu,\sigma}(x_i)$. Which are calculated with a Gaussian distribution ($f_{\mu,\sigma}$) as shown in the figure.

Since a Gaussian distribution is continuous, the probability for exactly some value is always zero. Therefore the probability over an infinitesimal area is calculated, for which the Gaussian probability density function ($f_{\mu,\sigma}(x_i)$, see equation 2.2) for each hit is needed.

$$f_{\mu,\sigma}(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2} \tag{2.2}$$

Since the probability of exactly one value cannot be calculated, the probability is instead given by: (see equation A.2)

$$P(x_1 \in [\tilde{x_1} \pm \epsilon], x_2 \in [\tilde{x_2} \pm \epsilon], \ldots, x_n \in [\tilde{x_n} \pm \epsilon] \mid \mu, \sigma, I) \approx f_{\mu,\sigma}(\tilde{x_1}) f_{\mu,\sigma}(\tilde{x_2}) \cdots f_{\mu,\sigma}(\tilde{x_n})(2\epsilon)^n \tag{2.3}$$

where $\epsilon > 0$ and ϵ is constant. By finding a value for ϵ and σ (with the constraint $\sigma > 0$), the proposed tracks can be tested. Filling these parameters in in formula 2.3 gives the probability that the data matches the hypothesis given that the hypothesis is true. Multiplying this with the probability of the hypothesis (the prior) gives the likelihood. The "best" set of hypotheses is the set with the highest likelihood values since "best" is defined as the maximum likelihood [D. S. Sivia, 2006, p.61-67]. [R. Frühwirth, 2000, p.159].

$$\mathcal{L}_{max} = P(\mu_{\text{most likely track}} \mid \sigma, I) \prod_{i=1}^{n} P(x_i \mid \mu_{\text{most likely track}}, \sigma, I)$$

The background to all these formulae can be found in appendix A

There are two disadvantages to this method: outliers have a large influence on the final result and arbitrary parameters bias the outcome. First we discuss the outlier problem. The assigned probability is between zero and one. Detected hits that lie far away from the proposed track get a very low probability (≈ 0) while hits that are close to the proposed track get a very high probability (≈ 1). As seen in equation 2.1 the product over the probabilities of every detected hit and the prior calculates the total likelihood. Outliers therefore have a big influence on the value of this product.

An example: Imagine a hypothesis that is close to the true track with 11 data points assigned a probability of approximately one, but one hit (an outlier) assigned a probability of approximately zero (see figure 2.2). Then the total product is still approximately zero, even though only one hit does not support the proposed track. The outlier causes the very "good" track to have a very low likelihood (a "bad" qualifier). This is unwanted because this track would have been close to the true track, but this did not show in the calculation because the outlier was included. The calculation of the situation in figure 2.2 would be as follows:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{12} P(x_i \mid \mu, \sigma, I) P(\mu \mid \sigma, I) \\
&\approx (1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 0) P(\mu \mid \sigma, I) \\
&\approx 0
\end{aligned}
$$

So the proposed track gets a low qualifier which is not an adequate representation of all data.

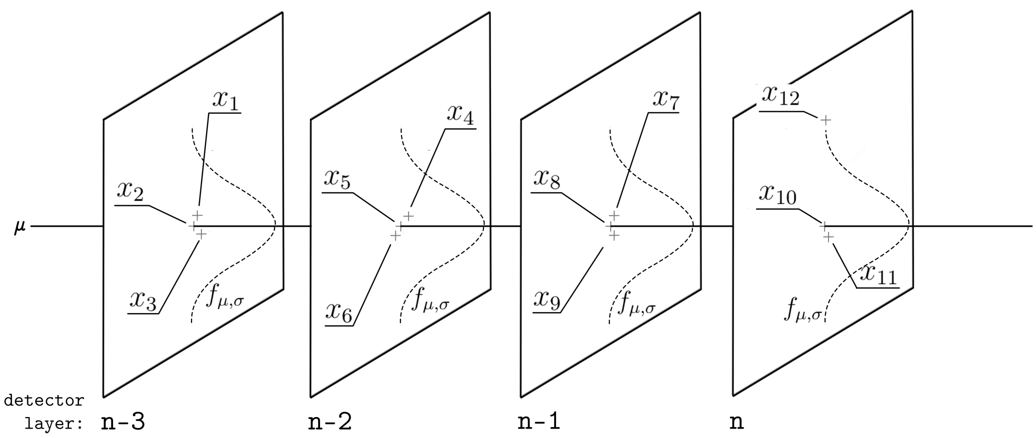

FIGURE 2.2: Situation in which outliers have a big influence when using the Likelihood Method. Even though the proposed track seems like it matches the data very well, it gets assigned a bad qualifier. According to the Likelihood Method this would be a "bad" trajectory.

Another disadvantage is that the assignment of a value to σ and ϵ is arbitrary. For ϵ this does not matter because it attributes the same factor to each likelihood, and this factor can be divided away (see Appendix B.1). But the assignment of σ influences the result. The standard deviation is a measure for the width of the Gaussian distribution. A larger width would mean that values are not as fast assigned a value of zero while a smaller width means more zero-valued probabilities (see Appendix A.3). This matters because a different value of σ influences each hit differently. The probability is not proportional to the distance but to the distance squared [G. Casella, 2002, p.316], so a change is the probability assignment is equivalent to a change in the distance *squared*. It can be seen in figure 2.3 that a distribution with twice the width has a very different ratio between probabilities ($\frac{0.309}{0.0.067} \approx 4600$) then the other distribution ($\frac{0.159}{0.001} = 159$). So the width of the distribution influences how probabilities relate to each other. Thus it influences the outcome of the total likelihood (see figure 2.3).

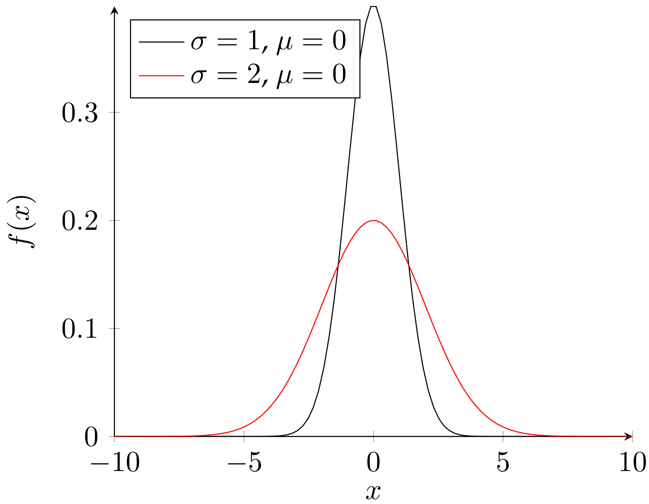

FIGURE 2.3: Illustration of how the width of the distribution influences probability. The ratio in which probabilities relate to each other changes when choosing a different value of σ.
For the black graph: $P(x > 1) = 0.159$ and $P(x > 3) = 0.001$.
For the red graph: $P(x > 1) = 0.309$ and $P(x > 3) = 0.067$.

The way in which the Likelihood Method works is that it constructs a hypothesis: the proposed track is the true track, and then it asks for every detected hit whether the proposed statement is true. Even if this question can be answered *yes* with almost certainty ($P \approx 1$), one cannot draw a solid conclusion from this. A theory is true until proven otherwise as is the rule in science. In the Likelihood Method the hypothesis is assumed to be true and for every hit this is investigated (the statement was assumed to be true, and it is proven that it is), this makes the statement only more likely, not necessarily true. The proposed statement can only be falsified (which can be undesirable when outliers are involved in the calculation of a "good" track). The only result that comes from the Likelihood Method is that the proposed track is not the true track, by falsification of the hypothesis. In Chapter 3 a method will be discussed that does not have this problem.

Some of the problems mentioned above can be made to have a smaller impact with so-called Tukey weights or a Corridor, which will be discussed now.

2.2 Using a Corridor

The source of the problem with outliers that the Likelihood Method has is that it takes into account all data, instead of just the hits that concern the proposed track. To reduce the influence of outliers one can install a "corridor" around the proposed track. A corridor is a region, for example a circle with radius r, around the proposed track. Points outside of this region are not a part of the calculation for the likelihood, while points inside this region are (see figure 2.4) [Steinle, 2012, p.39]. This way, there are few points attributed an excessively small probability and the total probability cannot drop to zero by adding a single data point. So outliers do not have a huge influence, because they are mostly left out of the calculation. If the distances between multiple tracks are sufficiently large such that outliers will be excluded, this approach can be used to solve the Outlier Problem.

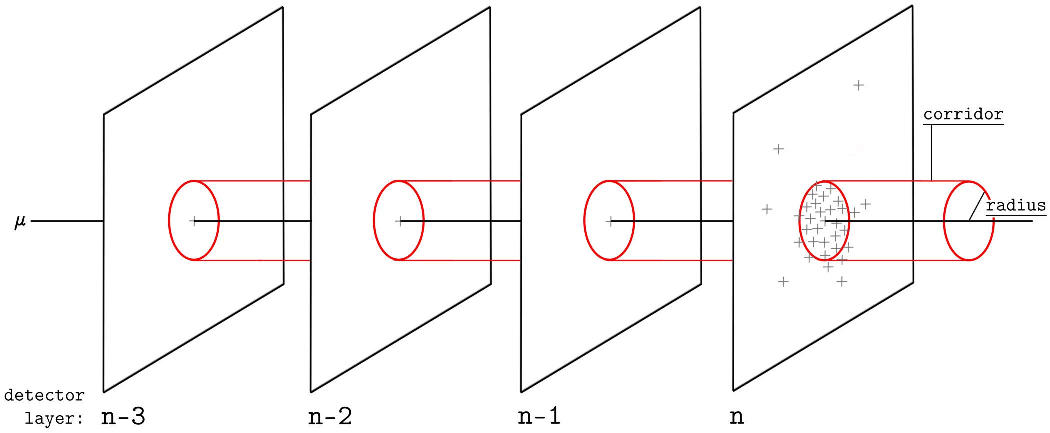

FIGURE 2.4: Schematic representation of the Corridor Method. A corridor of radius r is drawn in the figure, this corridor excludes outliers from the calculation in layer n.

This Corridor Method only enhances the Arbitrary Parameters Problem since there is no argumentation for choosing a specific corridor. It can be argued that a circle is preferable over any other figure because the distance from any point on the circle to the center is the same. For the corridor this would mean that no direction is preferred over another. Still, other figures such as a square or triangle could in principle be used as well. Although the shape is thus not completely arbitrary, the size of the corridor is. That would not matter if the corridor would not influence the result, which it does (see figure 2.5). Hits at the edge of the corridor are added or deleted from the calculation of the likelihood for different radii. This biases the resulting value of the likelihood since these points will or will not be included in the total product calculation dependent on where the boundary of the corridor is installed. The whole reason why the corridor was imposed, was to alter the result such that outliers became less influential. But when is an outlier an outlier and when is it a point that could be on the proposed track? This distinction is not always clear, the width of the corridor cannot be argued to have a specific value.

Not only the width of the corridor causes problems. When slightly adjusting the hypothesis, the new hypothesis may be measured against a different dataset (see figure 2.6). As can be seen in the figure, the hypotheses fit the data equally well, but the red track will be attributed a worse qualifier. Data points at the edge of the corridor slip in and out when slightly adjusting the hypothesis. In for example the LCHb VELO detector this problem will be amplified because there will be many hits in a small area. A slightly different hypothesis could include many more hits.

There is another more philosophical argument against using the Corridor Method. When installing a corridor, you leave out a part of the data. More specifically, you leave out "bad" data and focus only on "good" data. If every experiment would just

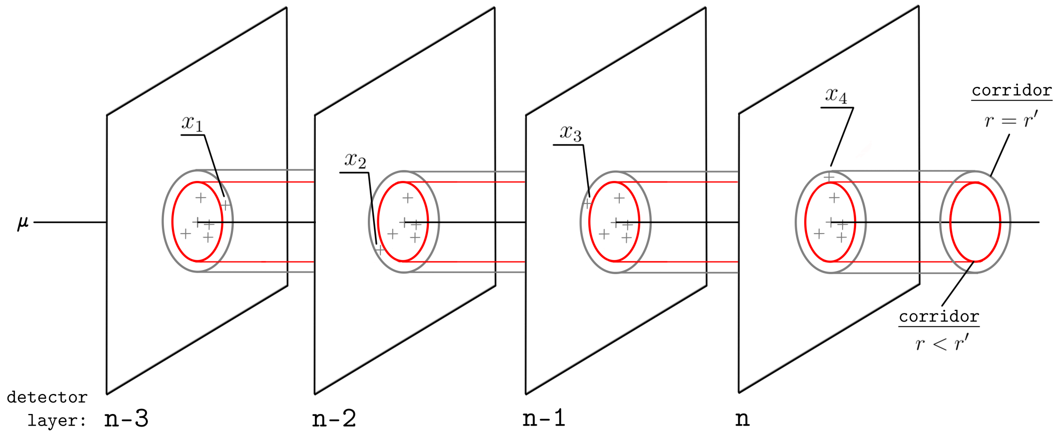

FIGURE 2.5: Illustration of how the width of the Corridor influences the result. A Corridor with a slightly larger radius contains four additional hits (labeled x_1, x_2, x_3 and x_4) that change the value of the qualifier.

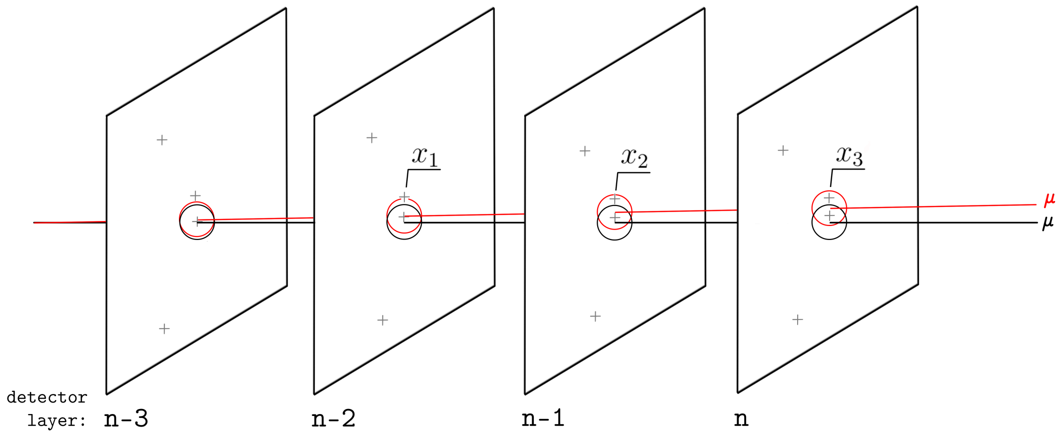

FIGURE 2.6: Illustration of how the result changes drastically with a slightly different hypothesis. The red hypothesis will be assigned a worse qualifier than the black hypothesis because its Corridor contains more "bad" hits even though the black and red hypothesis differ very little. The red track is slightly different but contains the additional hits labeled x_1, x_2 and x_3 with respect to the black track.

leave out the data that deviates from the proposed result, all data would be biased.

Although the Corridor solves the Outlier Problem, it biases the data and hits slip in and out of the Corridor when the hypothesis is changed slightly. There has been research concerning the width and shape of the corridor (e.g. [Steinle, 2012, p.39]), but the other problem remains, especially in high energy physics where there are many particles.

2.3 Using Tukey Weights

A second variation on the Likelihood Method that reduces its disadvantages as well as those of the Corridor Method is using so-called Tukey weights. Using Tukey weights means that every point is assigned a weight, which increases or decreases its influence on the total product, so it can decrease the influence of outliers [A. Strandlie, 2009]. The function

$$w = max(1 - x^2, 0), \tag{2.4}$$

gives a hit the weight $1 - x^2$ or 0, where x represents the position of the particle on the detector plane in relation to the proposed track [E. Etzion, 2006].

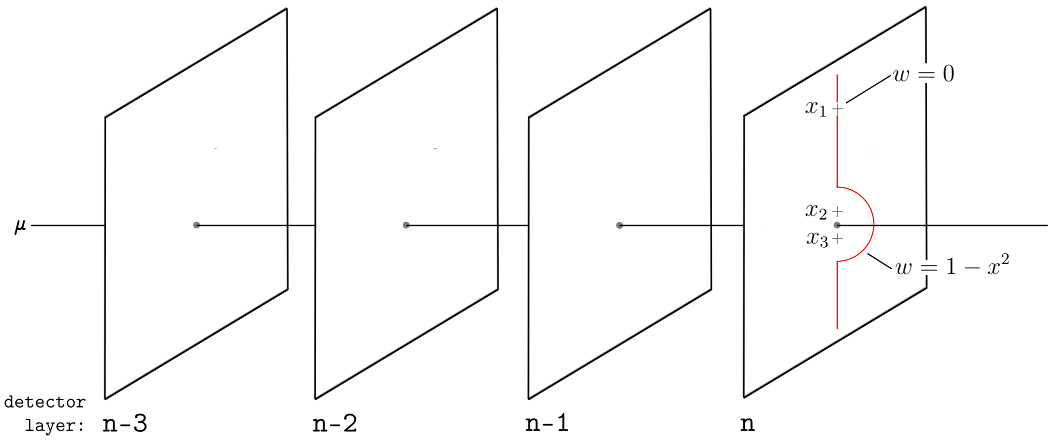

FIGURE 2.7: Schematic representation of Tukey weights. The hits labeled x_2 and x_3 are attributed a weight $1 - x^2$ dependent on their position with respect to the proposed track. The hit labeled x_1 gets attributed zero weight.

Points far away from the proposed track do not contribute to the total likelihood product (since they are assigned zero weight) and points closer to the proposed track have a larger influence on the product since they are assigned a larger weight (see figure 2.7). Tukey weights reduce the influence of outliers by assigning them zero weight [A. Strandlie, 2009]. They are also more efficient than methods such as the Corridor, which have a "hard border" [Hampel, 2001]. With Tukey weight, a slightly different hypothesis does not suddenly contain more hits with a big influence, because it does not have a sudden cut-off like the corridor. Hits that do not have zero weight in one hypothesis can have a weight in a slightly different hypothesis but this weight will be very small.

A disadvantage with respect to the Corridor Method is that using Tukey weights requires more calculations, because in addition to the likelihood, the weights have to be calculated as well and these have to be combined to find the qualifier. So there are more steps to the calculation of the qualifier.

Tukey Weights suffer the same Arbitrary Parameters Problem as the method described before. The "width" of the function seems arbitrary. The function $w = max(1 - (2x)^2, 0)$ makes a weighted corridor of half the width while the function $w = max(1 - (0.5x)^2, 0)$ doubles the width of the weighted corridor (see figure 2.8. There are no a priori arguments why the first would be better than the second, so the width of the function is arbitrary while it influences the qualifier. This means that the influence of outliers can become arbitrarily large or small. One could furthermore consider different weights such as a Gaussian distribution, but all will suffer from the discussed problems.

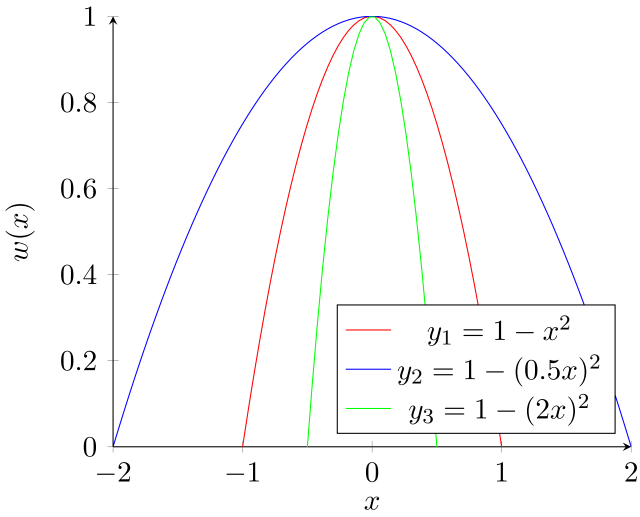

FIGURE 2.8: The plotted functions show the different weighing options. The width differs per function, assigning different weights to the same hits.

2.4 Overview

The Likelihood Method is a way of finding a qualifier that suffers from the Outlier Problem (defined in Section 1.2). There are multiple ways to reduce the influence of outliers: using Tukey weights or a corridor for example. The corridor imposes a "hard cut-off" which excludes part of the data. This makes the product of the likelihood very sensitive to small changes in the hypothesis. Also, the width of the corridor is arbitrary. Tukey weights use a weighing function to assign weights to all hits. This requires extensive calculation for the total likelihood, also the function that assigns the weight is arbitrary.

Both variations solve the Outlier Problem but suffer from the Arbitary Parameter Problem (along with having other disadvantages). The trick seems to be to find a method that solves the outlier problem without needing an arbitrary cut-off or weighing function.

Chapter 3

Unlikelihood Method

In this chapter and in Chapter 4, I propose two new methods for track finding. These methods were designed to solve the Outlier -and Arbitrary Parameter Problem that the methods in the previous chapter suffer from. However, they come with their own problems. I will describe the initial idea, the (dis)advantages each method has and the adjustments made to solve disadvantages.

An unattractive aspect of the Likelihood Method was its philosophical approach. It assumes a hypothesis and then investigates whether the data supports the hypothesis. A better way would be to try and falsify the hypothesis, because only then can something really be known, namely that the hypothesis was false. This "better" way of doing research can be implemented in track finding methods as well, and it has the potential to fix the Outlier Problem. To show the contrast to the Likelihood method, this method will be called the Unlikelihood Method. In the Unlikelihood Method some hypothesis for the possible track is assumed. How do we test whether this hypothesis is correct? By assuming that the hypothesis is not correct, so assuming there is no track there, and falsifying this assumption. If it is false that the there is no track there, the track is there, so we have proven the original hypothesis. In this section I will define two ways in which to define the unlikelihood: $\tilde{\mathcal{L}}$ and $\bar{\mathcal{L}}$. These are defined differently but both are called the unlikelihood. If necessary it will be specified which of the two is meant.

3.1 Option 1: The Unlikelihood $\tilde{\mathcal{L}}$

The draft on which this research is based [“Retina Thoughts”], proposes a new way of assigning a qualifier. This will be called the Unlikelihood Method, which uses the unlikelihood as a qualifier. Based on this draft we defined the unlikelihood ($\tilde{\mathcal{L}}$) as:

$$\tilde{\mathcal{L}} \equiv P(\mu \mid \neg\{x\}, \sigma, I) \tag{3.1}$$

As opposed to the Likelihood Method we want to know whether anything but the data supports the idea that the not the data is caused by the proposed track. Formulated differently, for each hit we ask the question: "is this hit *not* a part of the expected track?" The unlikelihood gives the value of the qualifier, which decides whether the hypothesis is close to the truth.

$$\tilde{\mathcal{L}} = \begin{cases} \text{low} & \text{if the proposed track is a good estimation given the data} \\ \text{high} & \text{if the proposed track is a bad estimation given the data} \end{cases} \tag{3.2}$$

The unlikelihood would be high if there is a high probability that the proposed track is true given that the data is not true. This means that the proposed track is supported by *not* the data. Formulated differently, this meas that the proposed track is disproved by the data. So then the hypothesis is not close to the truth. The Unlikelihood would be low if there is a low probability of getting the proposed track given *not* the data. If *not* the proposed track is probable for *not* the data then the proposed track is probable for the data, so the hypothesis is probable because the data would give a high probability for the proposed track.

3.1.1 Problem: Bayes Theorem

Bayes Theorem will show that the Unlikelihood as defined in equation 3.1 is not possible. Bayes theorem gives (see equation A.1):

$$P(\mu \mid \neg\{x\}, \sigma, I) = \frac{P(\neg\{x\} \mid \mu, \sigma, I) P(\mu \mid \sigma, I)}{P(\neg\{x\})}$$

In track finding the probability that the data is true is often taken to be unity:

$$P(\{x\}) = 1 - P(\neg\{x\}) = 1 \Rightarrow P(\neg\{x\}) = 0$$

Meaning that $P(\neg\{x\}) = 0$ which means that in calculating the unlikelihood one divides by zero, which is not allowed. This might seem the end of the Unlikelihood Method, but it could still provide useful, the mathematical definition just has to be altered. Before going into an altered version of the unlikelihood, the unlikelihood as defined above will be investigated further.

3.1.2 Solution

To solve the problem of division by zero, the probability that the data is not true is taken to be very small. We are free in assigning this probability as long as it is realistic ($P(\{x\}) = 0$ would not be realistic for example, but $P(\{x\}) = 0.99$ could be). We are allowed to do this because no experiment is perfect, there will always be a measurement error in the data. This error is the uncertainty in the data and can be used to define this probability. This gives that:

$$\begin{aligned}
\tilde{\mathcal{L}} &= \frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})} P(\neg\{x\} \mid \mu, \sigma, I) \\
&= \frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})} \prod_{i=1}^{n} P(\neg x_i \mid \mu, \sigma, I) \\
&= \frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})} \prod_{i=1}^{n} \Big[1 - P(x_i \mid \mu, \sigma, I)\Big]
\end{aligned}$$

The last equation above shows that this method solves the Outlier Problem. As an example we will again discuss the situation illustrated in figure 2.2. In Section 2.1 a calculation showed the influence of outliers on the qualifier when the Likelihood

Method was used. The calculation would now be as follows:

$$\begin{aligned}\tilde{\mathcal{L}} &= \frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})} \prod_{i=1}^{n} (1 - P(x_i \mid \mu, \sigma, I)) \\ &\approx \frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})} \cdot (1-1) \cdots (1-1) \cdot (1-0) \\ &\approx \frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})} \cdot 0 \cdots 0 \cdot 1 \\ &= 0\end{aligned}$$

If the proposed track is not supported by the data, this will give $P(x_i \mid \mu, \sigma, I) \approx 0$. Because this is subtracted from one, the contribution to the product will be multiplying by one, meaning that this outlier has no influence on the total product. However, this expression is not yet very nice and can be rewritten in two ways using two different approximation.

3.1.3 First approximation: Product Estimation

Each particle has one true hit in each detector layer so out of the 100 hits in each layer, only 1 will have a high probability. Since the values for $P(x_i \mid \mu, \sigma, I)$ will be small (≈ 0) about 99% of the time (since most hits are not part of a particular track), the following approximation of the product can be made (this calculation is shown more thoroughly in Appendix B.2):

$$\begin{aligned}\tilde{\mathcal{L}} &= \alpha \prod_{i=1}^{n} \Big(1 - P(x_i \mid \mu, \sigma, I)\Big) \\ &\approx \alpha \Big(1 - \sum_{i=1}^{n} P(x_i \mid \mu, \sigma, I) + \ldots\Big) \\ &\approx \alpha \Big(1 - \sum_{i=1}^{n} 2\epsilon(f(\tilde{x}_i) + \ldots)\Big) \\ &\approx \alpha \Big(1 - 2\epsilon(f(\tilde{x}_1) + f(\tilde{x}_2) + ... + f(\tilde{x}_n))\Big)\end{aligned}$$

where in the last line the higher order terms of the approximation are neglected, and with

$$\alpha = \frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})} >> 1.$$

$\alpha >> 1$ because the probability that the data is not true will likely be much smaller than the prior probability. This formula with which the unlikelihood can be calculated solves the Outlier Problem without using any arbitrary parameters (only the width of the distribution is still arbitrary). The value of ϵ is arbitrary, but the assigning of this value can be anything larger than zero as long as that value of ϵ is the same for each proposed track, because it attributes the same factor to each calculation of the unlikelihood (see Appendix B.2). Another constraint that can be placed on the assignment of a value to ϵ is that it should be such that product of 2ϵ and the sum of the values of the pdfs ($f(\tilde{x}_i)$) should be between zero and one. This way, the value of the unlikelihood is kept positive, which simplifies things. An example of such a value is $2\epsilon = \frac{1}{n}$, since the maximum value of $f(x_i)$ is one (in case of a delta function) and thus the maximum value of the sum is n. Multiplying this with $2\epsilon = \frac{1}{n}$ would make sure that the value is within its limits (between zero and one). Because of these constraints, the resulting unlikelihood is never below zero and the smallest value represents the most likely proposed track (see equation 3.2).

A disadvantage of this approximation is that although the resulting formula is simple, the approximation is not always valid. For about 1% of the data (the "good" points), the value of $P(x_i \mid \mu, \sigma, I)$ is *not* very small, and thus this approximation does not hold for those points. There is another way to achieve similar results using a different approximation.

3.1.4 Second approximation: Taylor Expansion

There will be only a small difference between different values of the unlikelihood attributed to different hypotheses because the calculation of the unlikelihood consists of a multiplication of many values between 0 and 1, which will result in a number close to zero. Taking the logarithm of the unlikelihood defined in equation 3.1 makes it numerically easier to compute and will make it easier to distinguish between the qualifiers of different hypotheses. Taking the logunlikelihood ($\log(\tilde{\mathcal{L}})$) and doing a Taylor expansion leads to the following:

$$\begin{aligned}
\log(\tilde{\mathcal{L}}) &= \log\Big(\alpha \prod_{i=1}^{n} P(\neg x_i \mid \mu, \sigma, I)\Big) \\
&= \log(\alpha) + \log\Big(\prod_{i=1}^{n} P(\neg x_i \mid \mu, \sigma, I)\Big) \\
&= \alpha' + \sum_{i=1}^{n} \log\Big(P(\neg x_i \mid \mu, \sigma, I)\Big) \\
&= \alpha' + \sum_{i=1}^{n} \log\Big(1 - P(x_i \mid \mu, \sigma, I)\Big) \\
&\approx \alpha' + \sum_{i=1}^{n} \log(1 - 2\epsilon f(\tilde{x}_i)) \\
&\approx \alpha' - 2\epsilon \sum_{i=1}^{n} \Big(f(\tilde{x}_i) + \frac{1}{2} 2\epsilon f(\tilde{x}_i)^2 + \dots\Big)
\end{aligned}$$

with

$$\alpha' = \log\Big(\frac{P(\mu \mid \sigma, I)}{P(\neg\{x\})}\Big) >> 0$$

and $\log(\alpha) = \alpha' >> 0$ since $\alpha >> 1$ and $\log(1) = 0$.

The Taylor expansion of a logarithm is used to reach the last line of the calculation. But $\log(1-y) \approx y$ only holds if $|y| << 1$, so if $|2\epsilon f(\tilde{x}_i)| << 1$. This places a constraint on the value of ϵ. Again, the contribution of ϵ to the final value, is the same for every expected track as long as the same value of ϵ is used, so it can be chosen arbitrarily (see Appendix B.2). An example of a valid value of ϵ would be $2\epsilon = \frac{1}{n^2}$. The maximum value of $f(\tilde{x}_i)$ is one so the maximum value of the product of the pdf and 2ϵ would be $\frac{1}{n}$. Since n is very large in high energy detectors (many particles are passing through the detector), the product will be a value much smaller than one and the condition is thus satisfied.

The value of the sum will likely be small compared to the value of the constant α'. Using the logarithm helps to make the different values for the logunlikelihood more distinguishable. The value of $\log(\tilde{\mathcal{L}})$ can be any real number, the lowest number again represents the most likely expected track (see equation 3.2).

The two methods above are not perfect, they solve the Outlier Problem without using arbitrary parameters or running into other problems, but they use approximations to calculate the unlikelihood and for the first case this approximation is not always valid. Also, it will be difficult to make the distinction between the values of the unlikelihood because the constant will likely be a large number (due to the small probability to not get the data) while the sum will be a small number. The probability of the data is generally taken to be zero, but this caused a division by zero. So instead of changing the method, I changed the probability to be small. We can also try to change the method by introducing an altered definition of the unlikelihood: $\bar{\mathcal{L}}$.

3.2 Option 2: The Unlikelihood $\bar{\mathcal{L}}$

The idea of the Unlikelihood method sprung from the philosophical approach to scientific research: assuming a hypothesis and trying to falsify it. Therefore we should find the probability that the hypothesis is false. Which leads to another definition of the unlikelihood ($\bar{\mathcal{L}}$), which is somewhat altered with respect to the unlikelihood $\tilde{\mathcal{L}}$.

$$\bar{\mathcal{L}} = P(\neg\mu \mid \{x\}, \sigma, I) \tag{3.3}$$

Here the unlikelihood would be high if there is a high probability of getting a track anywhere but at the proposed track given the data. So if the data supports *not* the proposed line ($\neg\mu$), then the proposed line was not a good estimation of the true track. The unlikelihood would be low if the data disproves the idea that there is no track at the proposed track, meaning that *not* the proposed track was a bad estimate and thus the proposed track is a good estimate of the truth.

$$\bar{\mathcal{L}} = \begin{cases} \text{low} & \text{if the proposed track is a "good" estimation given the data} \\ \text{high} & \text{if the proposed track is a "bad" estimation given the data} \end{cases} \tag{3.4}$$

Using Bayes theorem to find these probabilities can be done without running into trouble now. Bayes theorem gives:

$$P(\neg\mu \mid \{x\}, \sigma, I) = \frac{P(\{x\} \mid \neg\mu, \sigma, I)P(\neg\mu \mid \sigma, I)}{P(\{x\})}$$

and since $P(\{x\}) = 1$,

$$\begin{aligned} \bar{\mathcal{L}} &= P(\{x\} \mid \neg\mu, \sigma, I)P(\neg\mu \mid \sigma, I) \\ &= P(\neg\mu \mid \sigma, I)\prod_{i=1}^{n} P(x_i \mid \neg\mu, \sigma, I). \end{aligned}$$

The resulting formula does not look nice yet, the "$\neg\mu$" in the probability makes calculations difficult. We will simplify the expression for the unlikelihood to make it easier to use. Equation 3.3 is taken as a starting point of the simplification.

$$\begin{aligned}\bar{\mathcal{L}} &= P(\neg\mu \mid \{x\}, \sigma, I)\\ &= 1 - P(\mu \mid \{x\}, \sigma, I)\\ &= 1 - P(\{x\} \mid \mu, \sigma, I)P(\mu \mid \sigma, I)\\ &= 1 - P(\mu \mid \sigma, I)\prod_{i=1}^{n} P(x_i \mid \mu, \sigma, I)\end{aligned} \tag{3.5}$$

This expression does not contain any negations or difficult expressions. It can be checked even that the unlikelihood will always be between zero and one. The product of probabilities and the prior will always be between zero and one and thus subtracting this product from one will also be between zero and one. Equation 3.4 can thus be rewritten as:

$$\bar{\mathcal{L}} = \begin{cases} 1 & \text{if the expected track is a bad estimation given the data} \\ 0 & \text{if the expected track is a good estimation given the data} \end{cases}$$

The unlikelihood $\bar{\mathcal{L}}$ can be related to the likelihood $\mathcal{L}$. Filling in equation 2.1 (the definition of the Likelihood) into equation 3.5 above gives:

$$\bar{\mathcal{L}} = 1 - \mathcal{L}$$

Since the unlikelihood and likelihood are related as in the equation above, we can already expect that it will suffer from the Outlier Problem as well. To show that it does, we repeat the calculation of the situation in Section 2.1 as an example (see figure 2.2).

$$\begin{aligned}\bar{\mathcal{L}} &= 1 - P(\mu \mid \sigma, I)\prod_{i=1}^{n} P(x_i \mid \mu, \sigma, I)\\ &\approx 1 - P(\mu \mid \sigma, I)\epsilon(1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 0)\\ &\approx 1 - P(\mu \mid \sigma, I)\epsilon \cdot 0\\ &\approx 0\end{aligned}$$

The calculation shows that the proposed track is assigned a bad qualifier even though the expected track fits the data very well, the outlier influences the result.

The unlikelihood ($\bar{\mathcal{L}}$) does not provide useful because the outlier problem occurs as much here as it does in the Likelihood Method. This Unlikelihood Method has a philosophically better approach, but there are no other arguments for using this above the Likelihood Method.

This ends the discussion of the Unlikelihood Methods $\bar{\mathcal{L}}$ and $\tilde{\mathcal{L}}$. We can already see that the Unlikelihood Method $\bar{\mathcal{L}}$ has no prospects because it is influenced by outliers. The Unlikelihood Method $\tilde{\mathcal{L}}$ is attractive because it is not influenced by outliers and does not suffer the problems of the Corridor Method or Tukey weights either. Making the approximations can be invalid in some situations. In the Unlikelihood Methods the parameter σ is still arbitrary and influences the assigning of probabilities (see section 2.1).

Chapter 4

Ratio of Differences Method

Every method so far has had the same problem: how to assign a probability to a hit? You can use a Gaussian distribution with a cut-off or a certain weight. But the width of your distribution, the point of the cut-off and the assigning of the weight are all arbitrary choices, influencing the result. Assigning probabilities does not seem problematic at first sight since only the comparison of different qualifiers ($\mathcal{L}$ or $\tilde{\mathcal{L}}$ or $\bar{\mathcal{L}}$) leads to a conclusion which means only the ratio of the products should matter. However the probability and distance do not scale one-to-one. The choice for the width of the distribution influences the product of probabilities, see Section 2.1. The problem is that there are no clearly preferable values of σ.

4.1 Initial Idea

Using distances instead of probability distributions derived from distances would should solve the Arbitrary Parameter Problem described above. To solve the Outlier Problem, these distances would have to be between zero and one, such that outliers have no influence. If outliers are assigned the value one, then they have no influence on the product. The Ratio of Distances (ROD) Method is based on this idea of using distances instead of probabilities (see figure 4.1). The worst possible hits (such as outliers) get value one and the best possible hits get value zero. Taking the product again, and finding the track which gives the lowest value for the product should give the most likely track.

The qualifier in this method is the "ROD"-value (defined in equation 4.2).

$$ROD_i^2 = \frac{\alpha_i^2 + \gamma_i^2}{\alpha^2 + \gamma^2} \leq 1 \tag{4.1}$$

$$\begin{aligned} ROD^2 &= \prod_{i=1}^{n} ROD_i^2 \\ &= \prod_{i=1}^{n} \frac{\alpha_i^2 + \gamma_i^2}{\alpha^2 + \gamma^2} \end{aligned} \tag{4.2}$$

In equation 4.1 and 4.2 α and γ are the lengths of the detector plates in the $\hat{x}$-direction and $\hat{y}$-direction respectively. The sum of these values squared gives the length of the diagonal of the detector plate squared, which is constant. Taking the root gives the farthest distance possible between a hit and a proposed track. The value of α_i and γ_i are the distances from the proposed track to the detected hit x_i in the $\hat{x}$-direction and $\hat{y}$-direction respectively. The sum of these values squared gives the distance from the proposed track to the detected hit squared. Thus, $\alpha_i^2 + \gamma_i^2 \leq \alpha^2 + \gamma^2$ and the ROD-value of x_i will always be between zero and one. These variables are illustrated in figure 4.1.

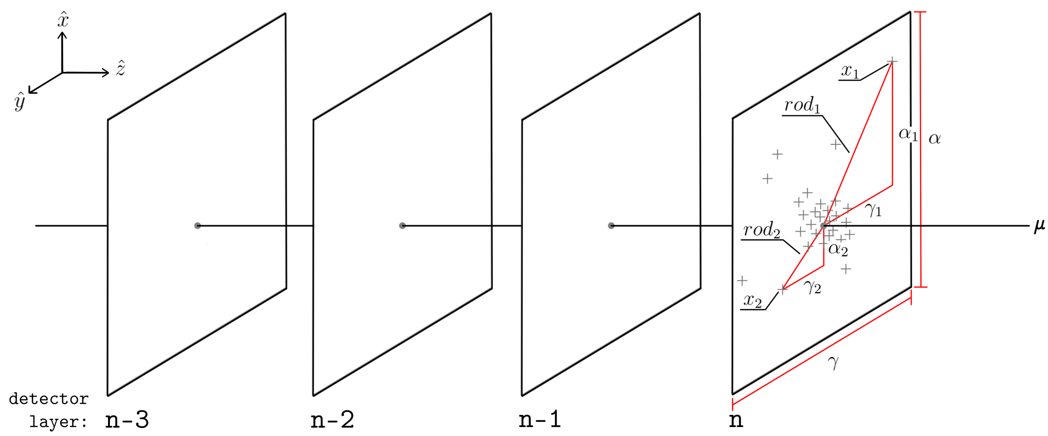

FIGURE 4.1: Schematic representation of Ratio of Distances Method. The variables and constants used in equation 4.2 and equation 4.3 to calculated the Ratio of Distances are illustrated in layer n.

4.2 Taking the sum or the product?

The product in equation 4.2 goes to zero very fast because only a hit which is at the complete other end of the detector plate gets an ROD_i-value of one, and this is a very unlikely event. It will be difficult to distinguish between the ROD-values which are all approximately zero, assuming many particle go through the detector, as is the case in high energy physics (so n is very large). Taking the sum of the ROD_i-values as a qualifier instead of the product would solve this numerical problem. This would result in the following definition, where we took the sum over all hits instead of the product of equation 4.1.

$$ROD^2 = \sum_{i=1}^{n} \left(1 - \frac{\alpha_i^2 + \gamma_i^2}{\alpha^2 + \gamma^2}\right), \tag{4.3}$$

where the variables are as defined in figure 4.1 and where the "$1-$" is introduced to make sure that the outliers contribute a value of zero, which has no influence on the sum. So this method should solve the Outlier Problem.

Taking the sum solves the numerical problem described above. But has taking the sum meaning? In the other methods which use probabilities, the rules for probability decide what to do (namely to take the product). With distance ratios, there is no real meaning in taking a sum, except that then the sum of the ratios is known. This sum cannot be transformed into probabilities, it can only be deduce that the lowest sum is the most likely one because it has the most hits close to the proposed track. With probability calculations the product is taken because it gives the probability for the first hit *and* the second hit *and* ... *and* the n^{th} hit. The ROD method is not bound by an AND/OR condition. As long as every hit is taken into account, either the sum or the product can be taken. The sum being the more practical option here.

Taking the sum has another advantage, namely that outliers cannot influence the outcome as much as when taking the product. Multiplying with 0 has a big effect on the product, but adding 1 is not quite as influential, because it still respects the other values.

4.3 Can anything be deduced?

Is taking the sum of the distance ratios something on which a valid conclusion can be based? A large sum of the distance ratios means that nearly all detected points are far away from the proposed track. A small sum means that nearly all detected points are close to the proposed track. The extremes of equation 4.3 are:

$$ROD^2 = \begin{cases} 0 & \text{if all hits are at the proposed track} \\ n & \text{if all hits are at the maximum distance from the proposed track} \end{cases}$$

where the value of n comes from adding the maximum value 1 for n hits. If there are a lot of "good" points close to the proposed track, then it is more likely that it is correct, but this is not necessarily true. Only one of the points in each layer is the hit belonging to the true track. Having many "good" points makes it more likely that the true point is there as well, but the true point could just as well be surrounded with very few "good" points, resulting in a quite high value for the sum. Such a situation is illustrated in figure 4.2.

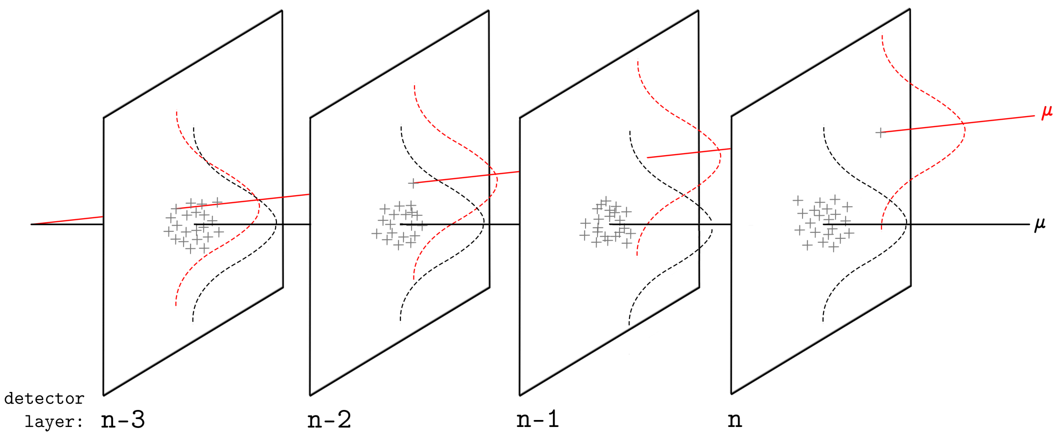

FIGURE 4.2: This is an illustration of a situation in which the ROD Method would result in the wrong track. In this situation where the red line represents the true track, the Likelihood Method with a corridor or Tukey weights or the Unlikelihood Method ($\tilde{\mathcal{L}}$) would choose the red line to be the "best" track. The Ratio of Distances Method would choose the black line to be the "best" track, because most points are close to the black track.

The result of the ROD Method in the situation described in figure 4.2 is not as expected. The method seems to give the wrong result in some situations, can it therefore be used? One could argue that the truth is not known beforehand and thus the only thing that can be done to get closest to the truth is to find the most likely option. Which in this case means the option with the highest number of good options. Many hits supporting the proposed track would indicate that it is more likely that the proposed track is true. More *likely*, it is not said to be true, such a claim cannot be made, we are always dealing with likelihoods and probabilities. The previous methods used a Gaussian distribution with an arbitrary width to determine how likely a track was. The ROD Method uses the number of "good" hits to determine how likely a track is. So there will definitely be a few situations in which the truth is not the most likely option, this is a problem that also occurs in the Likelihood Method due to the influence of outliers.

However, the ROD Method takes this problem one step further. Due to the linear distribution of values, the track with the majority of points will always show the smallest ROD-value, which suggests that it is the most likely track. A majority is most likely in some situations, but not in track finding. For example in the LHCb VELO Detector, where finding the true track using the ROD Method will always turn out to be the track that goes trough the middle of the detector, where the majority of hits will be. This is not true in reality. This deviation from the truth is caused by the linear distribution of values (as opposed to for example an exponential distribution of values which is the case in previous methods). Due to this, a lot of "bad" hits trump a few perfect hits, which is not the intended outcome. Therefore, this method gives an inaccurate representation of the most likely track. This could be solved by imposing an arbitrary cut-off, but that is not desirable for reasons discussed previously.

The Ratio of Distances Method seemed too perfect, and it is. It does not use any arbitrary parameters because there is no probability distribution. It also reduces the influence of outliers because the sum is taken and therefore one outlier cannot turn a good result into a bad result. But a whole new problem appears when examining this method further. Conclusions based on this method are not reliable.

Chapter 5

Conclusion & Discussion

In the past three chapters I have offered five methods of track finding that can be applied to for example the LHCb VELO detector. The purpose of this research was to identify the shortcomings of the old methods and to explore alternatives that solve these shortcoming. I will now discuss the conclusions of each previously discussed method, after which I will conclude on which method(s) could best be used for track finding when there are many tracks. This will answer the question whether we can find a quality parameter that finds the true track and is robust against outliers, which is the research question proposed in the first section: "Introduction and Problem Sketch"

5.1 Conclusion

The Likelihood Method is a method that suffers from the Outlier Problem. It does not have any arbitrary parameters except for the width of the distribution, σ.

Adding a Corridor to the Likelihood Method reduces the influence of outliers, but impose an arbitrary cut-off which can biases the result. Also, the value of the qualifier may change much for slightly different hypotheses.

Using Tukey weights also reduces the influence of outliers. It does not impose a hard cut-off as with the Corridor Method so using a slightly different hypothesis does not give a very different qualifier, and it is efficient. But the weight function w has an arbitrary "width" and the calculation of the qualifier can be extensive.

As an alternative we explored the unlikelihood. There are two definitions of the unlikelihood. The Unlikelihood Method $\bar{\mathcal{L}}$ does not solve the outlier problem. The Unlikelihood Method $\tilde{\mathcal{L}}$ can result in a good track finding method when using the logunlikelihood $\log(\tilde{\mathcal{L}})$, which can solve the outlier problem without using arbitrary parameters that influence the result. A Taylor approximation is used to get a manageable expression. This places a constraint on the value of ϵ. One could also use a product approximation, but this is not always allowed and therefore a lesser option. The chance that the data is true is not taken to be unity to avoid division by zero.

The ROD Method solves the Outlier Problem without using arbitrary parameters but gives rise to a whole new problem: can a conclusion be based on ROD-values? I would say that this is not always the case and thus it is never the case because you do not know when the conclusion is close to the truth and when not.

To conclude, the logunlikelihood method ($\log(\tilde{\mathcal{L}})$) discussed in Section 3.1.4 seems to be the best track finding method. It solves the Outlier Problem without using an arbitrary cut-off or weights that may biases the result. This method assigns a qualifier to hypotheses that is unbiased and not influenced by outliers. The only problem that remains is the assignment of a width to the Gaussian distribution. In section 2.1

it was said that this width (σ) is completely arbitrary, but that might not be entirely true. The standard deviation could represent the measurement inaccuracy, since it represents how sure we are that the data is measured to be exactly where it is [Cornelissen, 2006, p.69] or the standard deviation could be chosen such that the qualifier is optimized [D. S. Sivia, 2006, p.61-67]. So there are ways to find a value for σ. The value of ϵ can still be chosen arbitrarily as long as $|2\epsilon f(x_i)| << 1$. The probability of the data can also be chosen arbitrarily as long as it is not unity and as long as it is realistic.

To answer the research question, using the logunlikelihood method it is possible to find a qualifier that does not suffer from the Outlier Problem and that gives a good representation of the truth in every situation.

5.2 Discussion

From a review of the properties of the various methods it was concluded that the logunlikelihood method is the most promising. This method should be tested using controlled data sets, so that its validity can be established and its shortcomings can be identified.

There are a few things that this paper did not pay attention to which I feel need to be mentioned to get a complete picture.

First of all, measurement inefficiencies have not been discussed although they occur in the detectors. This could turn out to pose an additional problem to the discussed methods.

Furthermore, the assignment and role of the prior distribution has been left out of the discussion. There is much debate concerning the role of the prior (e.g. [Lyons, 2008, p.889-891]).

It should also be mentioned that already much more elegant and efficient adjustments to the likelihood method have been offered (e.g. [A. Abba, 2014]). These were not discussed here because understanding these methods would be a complete research project on its own.

Bibliography

A. Abba F. Bedeschi, M. Citterio F. Caponio A. Cusimano A. Geraci P. Marino M.J. Morello N. Neri G. Punzi A. Piucci L. Ristori F. Spinella S. Strackam D. Tonelli (2014). “The artificial retina processor for track reconstruction at the LHC crossing rate”. URL: https://arxiv.org/pdf/1409.1565.pdf.

A. Strandlie, R. Frühwirth (2009). “Track and Vertex Reconstruction: From Classical to Adaptive Methods”. In: URL: http://www.hephy.at/fileadmin/user_upload/Fachbereiche/ASE/Strandlie.pdf.

Cho, A. (2017). “Physicists detect whiff of new particle at the Large Hadron Collider”. In: *Science* 345. DOI: 10.1126/science.aal1066. URL: http://www.sciencemag.org/news/2017/04/physicists-detect-whiff-new-particle-large-hadron-collider.

Cornelissen, Thijs Gerrit (2006). “Track Fitting in the ATLAS Experiment”. In: URL: https://www.nikhef.nl/pub/services/biblio/theses_pdf/thesis_TG_Cornelissen.pdf#section.3.4.

D. S. Sivia, J. Skilling (2006). *Data Analysis, A Bayesian Tutorial*. Great Clarendon Street, Oxford: Oxford University Press.

E. Etzion D. Primor, G. Mikenberg N. Amram H. Messer (2006). “The use of cluster quality for track fitting in the CSC detector”. URL: https://arxiv.org/ftp/physics/papers/0612/0612129.pdf.

F.M. Dekking C. Kraaikamp, H.P. Lopuhaä L.E. Meester (2005). *A Modern Introduction to Probability and Statistics*. Springer-Verlag London Limited.

G. Casella, R. L. Berger (2002). *Statistical Inference*. Second Edition, International Student Edition. Brooks/Cole, Cengage Learning.

Hampel, F. (2001). “Robust statistics: A brief introduction and overview”. URL: ftp://ftp.sam.math.ethz.ch/sfs/Research-Reports/94.pdf.

LHCb installs its precision silicon detector, the VELO. Acessed: 23-06-2017. URL: https://press.cern/press-releases/2007/11/lhcb-installs-its-precision-silicon-detector-velo.

Lyons, L. (2008). “OPEN STATISTICAL ISSUES IN PARTICLE PHYSICS”. In: *The Annals of Applied Statistics* 2.3, pp. 887–915. URL: http://www.jstor.org/stable/30245113.

Onderwater, C.J.G. “Retina Thoughts”. Unpublished.

R. Frühwirth M. Regler, R. K. Bock H. Grote D. Notz (2000). *Data Analysis Techniques for High-Energy Physics*. Shaftesbury Road Cambridge: Cambridge University Press.

Steinle, C. A. (2012). *A First Level Trigger Approach for the CBM Experiment*. disserta Verlag.

The latest from the LHC. Acessed: 23-06-2017. URL: http://cds.cern.ch/journal/CERNBulletin/2009/26/News%20Articles/1184769?ln=en.

The LHCb Detector. Acessed: 26-05-2017. URL: https://lhcb-public.web.cern.ch/lhcb-public/en/detector/Detector-en.html.

VELO. http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/VELO2-en.html. Acessed: 23-06-2017.

Appendix A

Bayesian Statistics, Likelihood and PDFs

Track finding in this research paper is based on conditional probabilities: what is the chance of getting A given that we have B? The following formula calculates that probability: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. Likewise there is a formula that calculates the chance of getting B given that we have A: $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$ [F.M. Dekking, 2005, p.26]. These two formulae combined give rise to:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)},$$

which is called Bayes Theorem.

A.1 Bayes Theorem in track finding

In track finding, the probability of finding a certain track μ given the data $\{x\}$ is needed. The calculation of this probability could show the most likely track. Using Bayes theorem this can be calculated as follows:

$$P(\mu \mid \{x\}, I) = \frac{P(\{x\} \mid \mu, I)P(\mu \mid I)}{P(\{x\})} \tag{A.1}$$

or in words: [D. S. Sivia, 2006, p.6]

$$P(hypothesis \mid data, I) \propto P(data \mid hypothesis, I)P(hypothesis \mid I)$$

In these formulae the probability $P(\mu \mid I)$ is called the prior. The distribution of this probability gives the chance that the expected line is where it is thought to be. I is used to show any background information (such as the value of σ when using a Gaussian probability density function). In track finding, it is custom to take the chance that the data is there to be unity (so $P(\{x\}) = 1$). [D. S. Sivia, 2006, p.15]

The distribution assigned to the prior differs. It could be an uniform prior, which means that the proposed track can be anywhere. Or it could be a Gaussian distribution, which means the proposed track is most likely to be at some specified position, and less likely to be at places farther away from this specified position. The prior probability density function should describe what is known about the hypothesis before the data was acquired. [D. S. Sivia, 2006, p.61]

A.2 Probability Density Functions

The locations of the hit and the proposed track are a continuous variables, therefore to calculate a probability in track finding we need a continuous distribution. There is an important note to be made on the calculation of continuous distributions such as the Gaussian distribution. For discrete probability distributions there is a probability assigned to each possible value and the total of probability adds up to one. For continuous distributions there are infinitely many possible values, and the sum of the probabilities assigned to all mutually exclusive and exhaustive set of probabilities should be one. Therefore the value assigned to exactly x_i is zero. The probability in continuous distributions cannot be calculated by just filling in a number in the probability density function (pdf) as can be done for discrete distributions. Instead the probability is by definition of the probability density function, the area beneath the curve. The total area beneath a pdf is one. The probability can thus be calculated as shown in equation A.2, which is also illustrated in figure A.1. [F.M. Dekking, 2005, p.316-317]

$$P(x_i - \epsilon \leq X_i \leq x_i + \epsilon) = \int_{x_i-\epsilon}^{x_i+\epsilon} f_{\mu,\sigma}(X_i)dx \approx 2\epsilon f_{\mu,\sigma}(x_i) \tag{A.2}$$

Integrating over a region of the pdf gives the probability that the value of x is within the region that was integrated over. Which turns out to be some constant $\epsilon > 0$ times the value of the pdf at that point, which is basically the width times the height (which is the area).

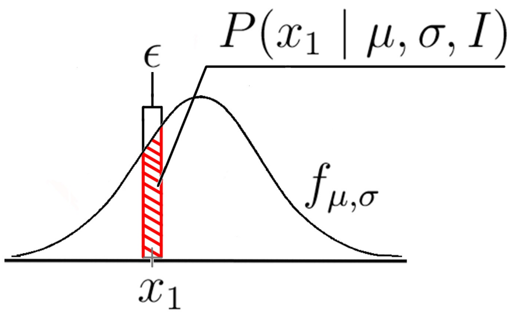

FIGURE A.1: Calculating the probability using the pdf of a normal distribution. The figure shows that $\epsilon f(x_i)$ calculates the area beneath the curve, which is defined as the probability.

A.3 The Gaussian Distribution

The Gaussian or Normal Distribution is widely used in statistics. It plays an important role in track finding as well. In this research the Gaussian distribution will be used to characterize the probabilities.

A Gaussian distribution looks as shown in figure A.1, the width and the height of the two-dimensional distribution is decided by choosing a value for σ, the standard deviation. The position of the distribution is decided by the value of μ, these two parameters decide on the shape of a two-dimensional Gaussian distribution. A

Gaussian distribution has the following probability density formula (which is plotted in figure A.2): [F.M. Dekking, 2005, p.64] [F.M. Dekking, 2005, p.320]

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

In track finding the value of μ is the value of the location of the proposed track and x is location of the hit that has been detected.

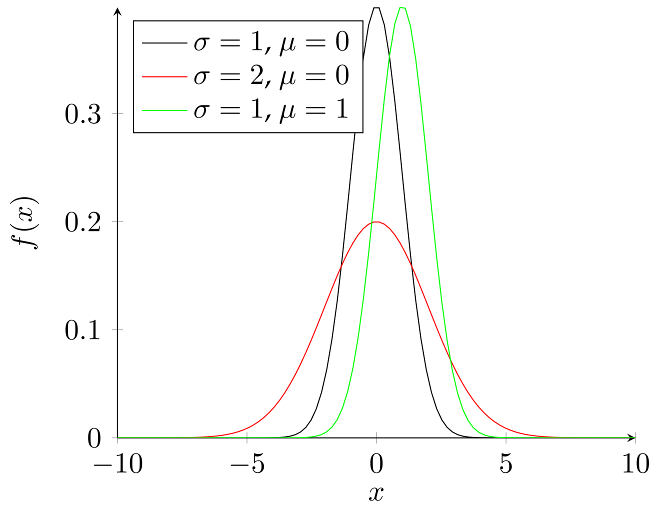

FIGURE A.2: Examples of Gaussian distributions with a varying center and width.

A.4 Likelihood

The likelihood ($\mathcal{L}$) in track finding originates from Bayes theorem. This section shows that derivation. The likelihood is defined as in equation 2.1. The likelihood can be calculated using Bayes theorem (equation A.1):

$$\begin{aligned}
\mathcal{L}(\mu \mid \{x\}, I) &= \prod_{i=1}^{n} P(\mu \mid x_i, I) \\
&= \prod_{i=1}^{n} \frac{P(x_i \mid \mu, I)P(\mu \mid I)}{P(x_i)} \\
&= P(\mu \mid I) \prod_{i=1}^{n} P(x_i \mid \mu, I)
\end{aligned}$$

here the prior probability can be taken out of the product because it is independent of the data and the probability of the data is assumed to be unity. [F.M. Dekking, 2005, p.316-321]

A.5 χ^2

This section explores the meaning of χ^2 and its relation to the distance (equation 1.1) and the likelihood (equation 2.1).

The definition of χ^2 is given in equation A.3. The expression contains the distance between the proposed track and the measurement in the $\hat{x}$-and $\hat{y}$-direction at a fixed location in the $\hat{z}$-direction, denoted $x(z) - \bar{x}(z)$ and $y(z) - \bar{y}(z)$ respectively. The measurement uncertainty σ in the $\hat{x}$-and $\hat{y}$-direction ($\sigma_x(z)$ and $\sigma_y(z)$ respectively) are taken to be independent of each other for simplicity. [D. S. Sivia, 2006, p.61-67]

$$\chi^2(z) \equiv \frac{(x(z) - \tilde{x}(z))^2}{\sigma_x^2(z)} + \frac{(y(z) - \tilde{y}(z))^2}{\sigma_y^2(z)} \tag{A.3}$$

The σ's give the uncertainty in the measurement of the hit. A larger measurement uncertainty results in a smaller value of χ^2 while a larger difference between the expected value and the true value results in a larger value of χ^2. The value of χ^2 is a measure for how likely it is that the found trajectory is due to chance. A large value meaning that it is likely that the trajectory is not close to the truth, but due to chance. A small value meaning that the track is probably close to the truth. [D. S. Sivia, 2006, p.61-67]

The distance is defined in equation 1.1 as:

$$d(z) = \sqrt{(x(z) - \bar{x}(z))^2 + (y(z) - \bar{y}(z))^2}$$

from which it can already be seen that it relates to the value of χ^2. As discussed, the value of χ^2 is calculated using the distance between the proposed track and the hit. We can express χ^2 in terms of $d(z)$ easily if we take the errors in the measurement in the $\hat{x}$- and $\hat{y}$-direction to be the same.

$$\begin{aligned}
\chi^2(z) &= \frac{(x(z) - \bar{x}(z))^2}{\sigma_x^2(z)} + \frac{(y(z) - \bar{y}(z))^2}{\sigma_y^2(z)} \\
&= \frac{(x(z) - \bar{x}(z))^2}{\sigma^2(z)} + \frac{(y(z) - \bar{y}(z))^2}{\sigma^2(z)} \\
&= \frac{(x(z) - \bar{x}(z))^2 + (y(z) - \bar{y}(z))^2}{\sigma^2(z)} = \frac{d(z)^2}{\sigma^2(z)}
\end{aligned}$$

The likelihood can be linked to χ^2 as well when taking the natural logarithm of the likelihood, which is called the loglikelihood function (l) [F.M. Dekking, 2005, p.316-321,p.335-336]. In this calculation we use equation 2.1 to define the likelihood, equation A.2 to go from a probability to a probability density function and equation 2.2 to fill in the pdf. It is also assumed that the measurement error σ is constant for each hit.

$$
\begin{aligned}
l(\mu \mid \{x\}, \sigma, I) &= \ln(\mathcal{L}(\mu \mid \{x\}, \sigma, I)) \\
&= \ln\Big(P(\mu \mid I) \cdot P(\{x\} \mid \mu, I)\Big) \\
&= \ln\Big(P(\mu \mid I) \prod_{i=1}^{n} P(x_i \mid \mu, I)\Big) \\
&= \ln\Big(P(\mu \mid I)\Big) + \ln\Big(\prod_{i=1}^{n} P(x_i \mid \mu, I)\Big) \\
&= \ln\Big(P(\mu \mid I)\Big) + \sum_{i=1}^{n} \ln\Big(P(x_i \mid \mu, I)\Big) \\
&\approx \ln\Big(P(\mu \mid I)\Big) + \sum_{i=1}^{n} \ln\Big(P(x_i \in [\tilde{x_i} \pm \epsilon] \mid \mu, I)\Big) \\
&\approx \ln\Big(P(\mu \mid I)\Big) + \sum_{i=1}^{n} \ln\Big(2\epsilon f_{\mu,\sigma}(\tilde{x_i})\Big) \\
&\approx \ln\Big(P(\mu \mid I)\Big) + \sum_{i=1}^{n} \ln\Big(2\epsilon \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{\tilde{x}_i-\mu}{\sigma})^2}\Big) \\
&\approx \ln\Big(P(\mu \mid I)\Big) + \sum_{i=1}^{n} \Big(\ln\Big(\frac{2\epsilon}{\sqrt{2\pi}\sigma}\Big) + \ln\Big(e^{-\frac{1}{2}(\frac{\tilde{x}_i-\mu}{\sigma})^2}\Big)\Big) \\
&\approx \ln\Big(P(\mu \mid I)\Big) + \ln\Big(\frac{2\epsilon}{\sqrt{2\pi}\sigma}\Big) + \sum_{i=1}^{n} \ln\Big(e^{-\frac{1}{2}(\frac{\tilde{x}_i-\mu}{\sigma})^2}\Big) \\
&\approx \ln\Big(P(\mu \mid I)\Big) + \ln\Big(\frac{2\epsilon}{\sqrt{2\pi}\sigma}\Big) - \sum_{i=1}^{n} \frac{1}{2}\Big(\frac{\tilde{x}_i - \mu}{\sigma}\Big)^2 \ln(e)\Big) \\
&= \mathbb{C} - \frac{1}{2}\sum_{i=1}^{k} \chi_i^2
\end{aligned}
$$

Here the prior is taken into the constant $\mathbb{C}$ together with the value of σ and ϵ, which are all assumed to be constant. The definition of $\mathbb{C}$ is:

$$\mathbb{C} = \ln\Big(P(\mu \mid I)\Big) + \ln\Big(\frac{2\epsilon}{\sqrt{2\pi}\sigma}\Big).$$

This calculation was done in one dimension, in which $\chi^2 = \left(\frac{\tilde{x}-\mu}{\sigma}\right)^2$ and where μ denotes the proposed location (which was denoted $\bar{x}$ in the definition of the distance) and where $\tilde{x}$ denotes the measured location (which was denoted x in the definition of the distance). Thus the likelihood and the value of χ^2 are easily related in the one dimensional case. The negative of the logarithm of the likelihood is the sum over all values of χ^2 for the layers:

$$-l(\mu \mid \{x\}, \sigma, I)) = \frac{\chi^2}{2} + \mathbb{C}'$$

Knowing how these two concepts relate to each other can give more insight into how the likelihood is defined.

Appendix B

Extra Calculations

This Appendix serves as an extra explanation of some statements made in previous chapters. I hope this will clarify some statements made in this paper.

B.1 Calculation supporting Section 2.1

In Section 2.1 the Likelihood Method was discussed and it was claimed that the value of ϵ could be chosen arbitrarily because it has not influence on the result. This section will show the calculation that supports this claim.

In equation 2.1 the likelihood was defined, which is used as a starting point in the following calculation. Take for example the situation in which there are two competing hypotheses: μ_1 and μ_2. These need to be compared to find which one has the highest likelihood (and which one is thus the "best" track). The calculation of the likelihood value for μ_1 would be as follows:

$$
\begin{aligned}
\mathcal{L}_1 &= P(\mu_1 \mid \{x\}, \sigma, I) \\
&= P(\{x\} \mid \mu_1, \sigma, I) P(\mu_1 \mid \sigma, I) \\
&= P(\mu_1 \mid \sigma, I) \prod_{i=1}^{n} P(x_i \mid \mu_1, \sigma, I) \\
&\approx P(\mu_1 \mid \sigma, I) \prod_{i=1}^{n} P(x_i \in [\tilde{x}_i \pm \epsilon] \mid \mu_1, \sigma, I) \\
&\approx P(\mu_1 \mid \sigma, I) \prod_{i=1}^{n} 2\epsilon f_{\mu_1,\sigma,I}(\tilde{x}_i) \\
&\approx (2\epsilon)^n P(\mu_1 \mid \sigma, I) \prod_{i=1}^{n} f_{\mu_1,\sigma,I}(\tilde{x}_i)
\end{aligned}
$$

following a same calculation would result in a similar likelihood value for μ_2.

$$
\mathcal{L}_2 \approx (2\epsilon)^n P(\mu_2 \mid \sigma, I) \prod_{i=1}^{n} f_{\mu_2,\sigma,I}(\tilde{x}_i)
$$

To compare the two likelihood values one could for example take the ratio of the two and see whether the result is larger or smaller than 1. Taking the ratio of the two values also shows that the value of ϵ has no influence on the result since it can be divided away.

$$\begin{aligned}\frac{\mathcal{L}_1}{\mathcal{L}_2} &\approx \frac{(2\epsilon)^n P(\mu_1 \mid \sigma, I) \prod_{i=1}^n f_{\mu_1,\sigma,I}(\tilde{x}_i)}{(2\epsilon)^n P(\mu_2 \mid \sigma, I) \prod_{i=1}^n f_{\mu_2,\sigma,I}(\tilde{x}_i)} \\ &= \frac{P(\mu_1 \mid \sigma, I) \prod_{i=1}^n f_{\mu_1,\sigma,I}(\tilde{x}_i)}{P(\mu_2 \mid \sigma, I) \prod_{i=1}^n f_{\mu_2,\sigma,I}(\tilde{x}_i)}\end{aligned}$$

As can be seen, the expression of the ratio of the two likelihood values does not contain ϵ. The value of ϵ can therefore be anything larger than zero.

B.2 Calculation supporting Section 3.1

The calculations in Section 3.1 show how the two approximations can be used to obtain an expression for the unlikelihood that should be easier to use. The calculation of the product approximation (done in Section 3.1.3) is repeated here in more detail and in addition we will show that the assigning of a value to ϵ can be arbitrary (as long as the constraints discussed in the concerning section are respected). This calculation will only be done for the product estimation, the calculation of the logunlikelihood (discussed in Section 3.1.4) is very similar and will therefore not be repeated as well.

Here we will do the calculation of the unlikelihood $\tilde{\mathcal{L}}$ with respect to the two competing hypotheses μ_1 and μ_2. In line with Section 3.1.3 we have that:

$$\alpha_1 = \frac{P(\mu_1 \mid \sigma, I)}{P(\neg\{x\})} \quad \text{and} \quad \alpha_2 = \frac{P(\mu_2 \mid \sigma, I)}{P(\neg\{x\})}.$$

The calculation of the unlikelihood of the proposed track μ_1 is given as $\tilde{\mathcal{L}}_1$.

$$\begin{aligned}\tilde{\mathcal{L}}_1 &= \alpha_1 \prod_{i=1}^{n} \Big(1 - P(x_i \mid \mu_1, \sigma, I)\Big) \\ &\approx \alpha_1 \Big(1 - \sum_{i=1}^{n} P(x_i \mid \mu_1, \sigma, I) + \ldots\Big) \\ &\approx \alpha_1 \Big(1 - \sum_{i=1}^{n} P(x_i \in [\tilde{x}_i \pm \epsilon] \mid \mu_1, \sigma, I) + \ldots\Big) \\ &\approx \alpha_1 \Big(1 - \sum_{i=1}^{n} 2\epsilon(f_{\mu_1,\sigma,I}(\tilde{x}_i) + \ldots)\Big) \\ &\approx \alpha_1 \Big(1 - 2\epsilon(f_{\mu_1,\sigma,I}(\tilde{x}_1) + f_{\mu_1,\sigma,I}(\tilde{x}_2) + ... + f_{\mu_1,\sigma,I}(\tilde{x}_n) + \ldots)\Big) \\ &\approx \alpha_1 \Big(1 - 2\epsilon(f_{\mu_1,\sigma,I}(\tilde{x}_1) + f_{\mu_1,\sigma,I}(\tilde{x}_2) + ... + f_{\mu_1,\sigma,I}(\tilde{x}_n))\Big)\end{aligned}$$

Here the approximation of the product is used to get from the first to the second line and to get from the third to the forth line the approximation of the probability using the probability density function is used (equation A.2). To reach the last line the higher order terms of the product approximation are neglected. The same calculation for the proposed track μ_2 will show a similar result:

$$\tilde{\mathcal{L}}_2 \approx \alpha_2 \Big(1 - 2\epsilon(f_{\mu_2,\sigma,I}(\tilde{x}_1) + f_{\mu_2,\sigma,I}(\tilde{x}_2) + ... + f_{\mu_2,\sigma,I}(\tilde{x}_n))\Big)$$

To compare the two likelihoods one could for example take the formula $\tilde{\mathcal{L}}_1 = \gamma\tilde{\mathcal{L}}_2$, and look at the value of γ. If $\gamma > 1$ then $\tilde{\mathcal{L}}_1 > \tilde{\mathcal{L}}_2$ and if $\gamma < 1$ then $\tilde{\mathcal{L}}_1 < \tilde{\mathcal{L}}_2$. Filling in the values of $\tilde{\mathcal{L}}_1$ and $\tilde{\mathcal{L}}_2$ as calculated above gives:

$$\alpha_1\Big(1 - \sum_{i=1}^{n} 2\epsilon f_{\mu_1,\sigma,I}(\tilde{x}_i)\Big) = \gamma\alpha_2\Big(1 - \sum_{i=1}^{n} 2\epsilon f_{\mu_2,\sigma,I}(\tilde{x}_i)\Big)$$

and filling in the values of α_1 and α_2,

$$\frac{P(\mu_1 \mid \sigma, I)}{P(\neg\{x\})}\Big(1 - \sum_{i=1}^{n} 2\epsilon f_{\mu_1,\sigma,I}(\tilde{x}_i)\Big) = \gamma\frac{P(\mu_2 \mid \sigma, I)}{P(\neg\{x\})}\Big(1 - \sum_{i=1}^{n} 2\epsilon f_{\mu_2,\sigma,I}(\tilde{x}_i)\Big).$$

In this expression there are two values that are the same for both proposed tracks: ϵ and $P(\neg\{x\})$. The assigning of these values has no influence on the final product. A twice as large value for ϵ will give the same contribution on both sides and should therefore have no influence on the value of γ. It can also be seen that the probability of *not* the data ($P(\neg\{x\})$) can be divided away. This value can therefore also be assigned arbitrarily. However, the value of $P(\neg\{x\})$ should be realistic and cannot be zero for reasons mentioned in section 3.1.3. The value of ϵ should also respect the constrained mentioned in that section. The calculation of the logunlikelihood discussed in Section 3.1.4 would lead to the same conclusion.