Measurement of the tZq Differential Cross-section with the ATLAS Detector at the LHC

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Nilima Akolkar
aus
Vadodara, India

Bonn 2024

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. John Smith
2. Gutachterin: Prof. Dr. Anne Jones

Tag der Promotion:
Erscheinungsjahr:

Contents

Todo list

CHAPTER 1

Statistical Methods and Analysis Setup

In particle physics, the abundance of available data for exploration has dramatically increased with the development of powerful particle colliders. Be it estimating a parameter of the SM or finding evidence of new physics, particle physicists heavily rely on statistical techniques to churn out reliable information from observed data. This chapter provides details on the statistical methods used in this thesis. Basic statistical techniques including parameter estimation and method of maximum likelihood are outlined in Section 1.1. The concept of unfolding is introduced and the techniques used in this thesis are explained in detail. This chapter also discusses the method of profile likelihood unfolding, which is the core technique used in this analysis.

1.1 Statistical Inference

The main goal of a statistical analysis is to infer key information from the observed data. Commonly used approaches of inference are frequentist and Bayesian which are based on the interpretation of probability. In statistics, probability is simply the "chance" of an event ocurring. In the frequentist approach, probability is interpreted as the frequency of an event occurring, whereas in the Bayesian approach, probability is the extent of belief in the occurrence of an event [1].

Statistical problems can be categorised into two types: parameter estimation and hypothesis testing. In parameter estimation, the goal is to determine the "best" possible value of a certain physical quantity or a parameter of a certain mathematical model. It is important to note that the parameter value is always accompanied by an error estimate which quantifies the accuracy of a measurement. A measurement may not be perfect, however the extent of imperfection is concealed in the error values. Therefore, error values play an important role in the interpretation of experimental results.

In hypothesis testing, the main goal is to check if a theory is consistent with the observed data. For this case, the answer is not a numerical value, instead, it is a statement implying how confident we are with the consistency of a theory based on the observed data. In reality, hypothesis testing and parameter estimation are not totally independent of each other. Some problems require estimating a parameter in order to test a hypothesis while in some cases, a parameter is estimated assuming a hypothesis is correct [2]. In this analysis, we focus on parameter estimation. In frequentist statistics, method of least squares and method of maximum likelihood are mainly used for parameter estimation. In this thesis, the maximum likelihood method is used which is discussed in the next section.

Method of maximum likelihood

Statistical models are mathematical expressions that relate observations to underlying theories. These models are characterised by a set of parameters. The possibilities of different outcomes, given a mathematical model, is described by probabilities which is a key concept in statistics. Another important concept which connects observations to parameters, is the likelihood. Likelihood is a measure of how likely a set of parameters can describe the actual observed data.

Consider a random variable x distributed according to a probability distribution function $f(x;\theta)$ where $f(x;\theta)$ represents our assumed hypothesis. Now suppose the functional form of $f(x;\theta)$ is known, however, some parameters are unknown. Suppose a measurement is performed yielding n independent values, denoted by $x_1, x_2, ..., x_n$. The probability of observing this particular set of values is given by the product of the individual probabilities for each value, as given in Eq. (1.1).

$$P(x_1, x_2, ..., x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) dx_i \tag{1.1}$$

If the assumed hypothesis is correct, the probability of observing the data should to be high. This concept leads to the *likelihood function* $L(\theta)$, which is defined as:

$$L(\theta) = L(\theta | x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i, \theta) dx_i \tag{1.2}$$

Here, $L(\theta)$ measures how likely it is to observe the set $x_1, x_2, ..., x_n$, given the parameter θ. Unlike a standard probability function, the likelihood function treats the data as fixed and considers θ as the variable.

The objective of maximum likelihood estimation is to estimate the value of θ that makes the observed data most probable. In other words, we search for the parameter values that maximize the likelihood function, a process referred to as the method of maximum likelihood or maximum likelihood fitting. In practice, it is often more convenient to maximize the logarithm of the likelihood function, known as the log-likelihood. This transformation simplifies the product of probabilities in the likelihood function into a sum, which is mathematically easier to handle while preserving the location of the maximum.

The process of maximising the likelihood depends on the observed data. The efficiency of parameter determination will depend on the extent of reliability on observed data. In particle physics, this fact imposes a certain level of expectations from particle detectors. Even though our detectors are well-developed and efficient, there are some limitations that lead to distortion of observations. Physicists often use mathematical techniques to remove the distortions caused by detectors. These techniques fall under the category of *unfolding*.

1.2 Basic concept of unfolding

In any experiment, the accuracy of measurement depends heavily on the performance of the apparatus used. In particle physics, an ideal detector would capture the original and complete information of collisions without any loss or distortion, accurately preserving the true shape of distributions for any physical observable. However, in reality, the data received from a detector is distorted due to effects

such as limited acceptance and finite resolution of the detector. These distortions may lead to incorrect inferences and therefore, need to be removed.

Unfolding is a mathematical technique used to remove distortions and estimate the original fine structure of detector data. This technique is also called desmearing or deconvolution. In unfolding, the goal is to estimate the true distribution from the observed distribution, using a response matrix. The response matrix is a mathematical construct that characterises the smearing effects introduced by the detector. An illustration visualising unfolding is shown in Fig. 1.1.

Unfolding is essential in various contexts. For instance, it enables results to be combined or compared with those from other experiments that may have different levels of smearing. Additionally, it is not always practical to present results alongside their response matrices. Unfolding becomes crucial when comparing experimental results directly to theoretical predictions without accounting for experimental distortions. Furthermore, non-distorted data is needed when fitting specific parameters to data for tuning Monte Carlo simulations [3].

1.3 Unfolding methodology

There are certain conventions regarding the entities used in unfolding problems within the particle physics community. The distribution of a physical observable obtained from data recorded by the detector is called detector-level distribution. On the other hand, the truth-level distribution represents data that we should have obtained with an ideal detector. This is generated from MC simulations without applying any detector simulations. In unfolding, we also make use of simulated distribution including detector simulation. Since this so-called reconstructed-level distribution is supposed to mimic actual observed data, it is used to validate an unfolding method. Detector-level and reconstructed-level distributions are conceptually the same when discussing unfolding related quantities. Technically, unfolding is a method of estimating the truth-level by correcting detector effects present at detector- or reconstructed-level.

The reconstructed distribution $\vec{x}$ is related to the truth-level distribution $\vec{y}$ by a response matrix R as shown in Eq. (1.3). Here $\vec{b}$ represents backgrounds. In particle physics problems, data is generally organised into histograms with finite bins. In this case the above given relation is given by Eq. (1.4).

$$\vec{x} = R \cdot \vec{y} + \vec{b} \tag{1.3}$$

$$x_i = \sum_{\mathrm{j}\,=\,1}^{\mathrm{bins}} \mathcal{R}_{ij} y_j + \vec{b} \tag{1.4}$$

A response matrix which quantifies the detector effects is computed from simulated data. The quantities used to construct the response matrix are migration matrix and two correction factors namely acceptance and efficiency. A schematic showing reconstructed-level and fiducial truth-level volumes is given in Fig. 1.2. Per-bin acceptance adjusts the number of reconstructed events by the fraction of events that are also present at the fiducial truth-level. It is defined as the ratio of events present at the reconstruction- and truth-levels to the total number of events at the reconstruction-level. Acceptance gives an idea of how well the reconstructed data corresponds the true data. The corrected number of events at the reconstructed-level is given as,

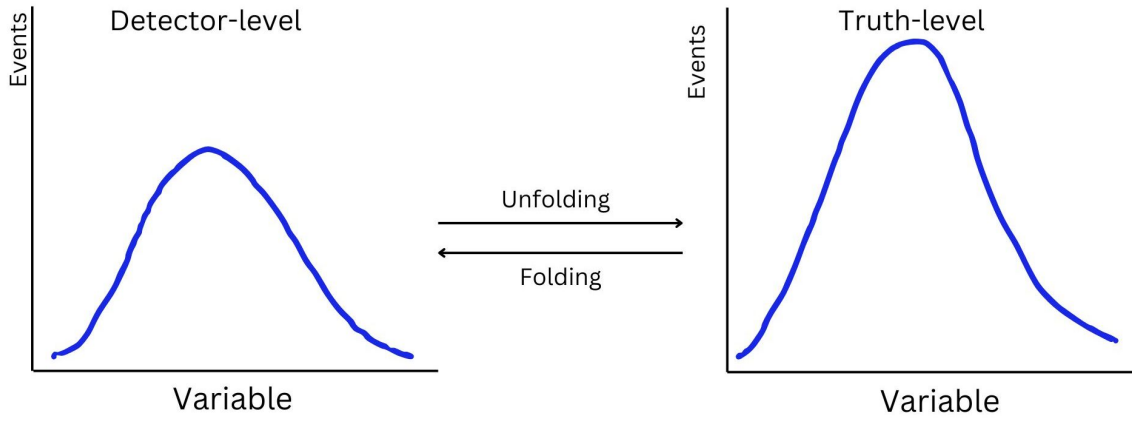

Figure 1.1: An illustration visualising the method of unfolding. The left diagram is an illustration of a histogram obtained from the detector information, called detector-level and the right illustration shows a histogram that is expected form an ideal detector, called truth-level. The difference in the two diagrams depicts the smearing caused by the detector. The procedure to obtain the truth-level information based on the detector-level information is called as unfolding. The reverse procedure is called folding.

$$N_i^{\text{reco}} = x_i * a_i \tag{1.5}$$

Per-bin efficiency adjusts the number of truth-level events by the fraction of truth-level events that are also found at the reconstructed-level. It is defined as the ratio of a number of events present at both the reconstruction-level and the truth-level to the total number of events at the fiducial truth-level. This correction factor gives an idea of the detector's efficiency to reconstruct true events. The corrected number of events at the fiducial truth-level is given as,

$$N_j^{\text{fid}} = y_j * \epsilon_j \tag{1.6}$$

The migration matrix describes the bin-to-bin migrations between truth-level and reconstruction-level histograms. For instance, M_{ij} represents the fraction of events found in bin i at the reconstruction-level while being created in bin j at the truth-level. A migration matrix with maximum diagonal components indicates that most events are reconstructed in the same bin in which they were generated. Equation (1.4) can be re-written as,

$$N_i^{\text{reco}} = \frac{a_i}{\epsilon_j} * \sum_{\text{j}=1}^{\text{bins}} M_{ij} \cdot N_j^{\text{fid}} \tag{1.7}$$

Mathematically, the idea of unfolding is to solve Eq. (1.3) for given $\mathcal{R}$, x and b. The resultant values of y can be interpreted as determined true number of events at the truth-level. One would notice a simple way to find the estimators by inverting the response matrix as shown in Eq. (1.8).

$$\vec{y} = \mathcal{R}^{-1}(\vec{x} - \vec{b}) \tag{1.8}$$

Although matrix inversion method is easy to implement, it is a strategy that one should avoid

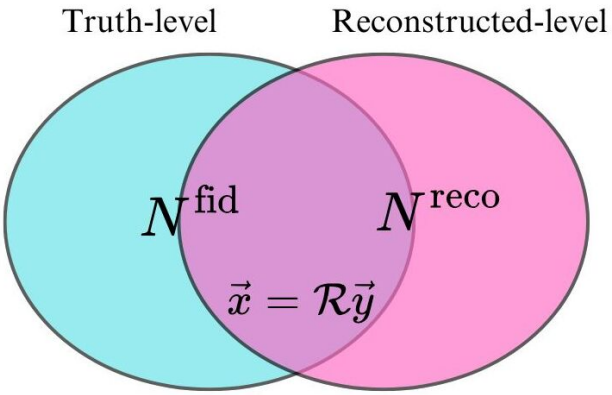

Figure 1.2: A schematic picture showing reconstructed-level volume in blue and truth-level volume in pink.

because of its limitations: in some situations, the response matrix is non-invertible then Eq. (1.8) becomes ill-posed. Even though inversion is possible, there are possible statistical fluctuations in the observed data that may cause negative entries in the inverse matrix. This leads to negative number of events in the unfolded distribution which is unrealistic. When a response matrix acts on a true spectrum, it distorts any fine structure present at the truth-level. Despite that, some residue of this fine structure still remains in the reconstructed spectrum [4]. The inverted matrix, acting on the measured data, assumes its statistical fluctuations are the residual fine structure and restores it. In this way, statistical fluctuations are amplified in the unfolded distribution [5] which is undesirable.

In order to overcome the limitations of matrix inversion, alternate unfolding methods are used in high energy physics. Some of the methods are summarised in [6]. In this thesis, the iterative Bayesian unfolding (IBU) and profile-likelihood unfolding (PLU) are discussed.

1.4 Iterative Bayesian unfolding (IBU)

D'Agostini [7] proposed a method called iterative Bayesian unfolding (IBU) which makes use of Bayes' theorem. To describe this method, consider true events as *causes* ($C_i, i = 1, 2, ..., n_C$) and reconstructed events as *effects* ($E_j, j = 1, 2, ..., n_j$). The conditional probability that a cause C_i gave rise to effect E_j, denoted by $P(C_i|E_j)$, is given by Bayes' theorem:

$$P(C_i|E_j) = \frac{P(E_j|C_i)P(C_i)}{P(E_j)} \tag{1.9}$$

where, $P(E_j|C_i)$ can be interpreted as probability of reconstructed event given true event which is the element M_{ij} of the migration matrix. Consequently, $P(C_i|E_j)$ can be identified as the unfolding matrix. One can determine the number of events ($\hat{n}$) due to cause C_i as

$$\hat{n}(C_i) = \frac{1}{\hat{\epsilon}_i} \sum_{j=1}^{n_E} \hat{n}(E_j) P(C_i|E_j). \tag{1.10}$$

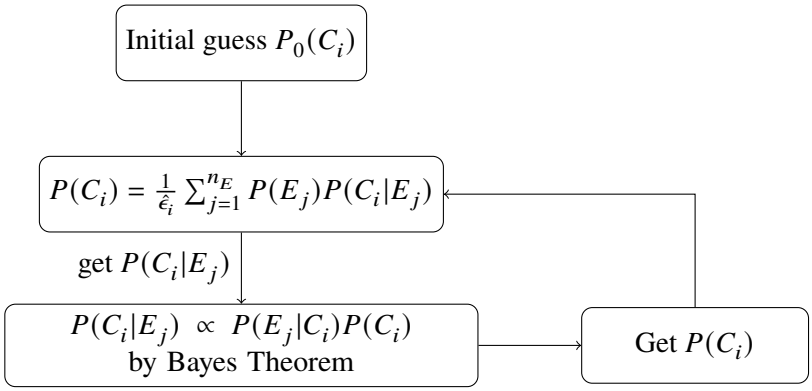

Figure 1.3: Flow chart showing steps followed in iterative Bayesian unfolding

It is important to note that the total number of events due to all causes and all effects are equal because only migration effects are considered so far. By dividing both the sides of Equation 1.10 by total number of events, we obtain

$$P(C_i) = \frac{1}{\hat{\epsilon}_i} \sum_{j=1}^{n_E} P(E_j)P(C_i|E_j). \tag{1.11}$$

Here, $P(C_i)$ is the unfolded distribution. This technique is implemented in the `RooUnfold` [8] software package. The steps performed in iterative Bayesian unfolding to find $P(C_i)$, are shown in Figure 1.3 and explained below:

- An initial guess $P_0(C_i)$ is made and inserted into Equation 1.11. $P(E_j)$ is obtained from MC reconstructed distribution. The solution provides $P(C_i|E_j)$ which is the unfolding matrix.
- The obtained $P(C_i|E_j)$ is used in Bayes' theorem (Equation 1.9) to get a value of $P(C_i)$ which is different from the initial guess.
- This process is repeated for number of iterations specified by user.

1.5 Profile likelihood fitting

Profile likelihood approach is built upon the likelihood function described in Section 1.1. In complex problems such as in high energy physics, the statistical models present many parameters that directly or indirectly affect the measurement. Out of those, at least one parameter is central to the underlying hypothesis, such as the signal strength, which should be estimated. This parameter is called parameter of interest (POI). Furthermore, there are additional parameters that influence the measurement but don't necessarily need to be estimated, for instance, systematic uncertainties. These are called nuisance parameters (NP). Including all these parameters, the likelihood function in Eq. (1.2) can be written as a function of POI (μ) and NP vector (σ) as,

$$L(\theta) = L(x; \mu, \sigma) \tag{1.12}$$

In reality, there are hundreds of NPs affecting the measurement and in such cases it becomes difficult to solve a multidimensional likelihood function. A simpler way is to use a profile likelihood function which is based on the likelihood function but focuses on the POI while considering the effects of the NPs. The profile likelihood function maximises the likelihood over all NPs for each fixed value of the POI. By doing this, the NPs are replaced with their corresponding maximum likelihood estimators $\hat{\sigma}$ and the likelihood becomes a function of μ. The process of using fixed values of σ for a given μ is called "profiling" out the NPs. This reduces the dimensionality of the likelihood. This procedure of estimating the POIs is called profile likelihood fitting.

Formulation

The first step to profile likelihood fitting is to construct a mathematical model that embodies the observed data along with predicted values of model parameters. In high-energy physics analyses, the observed data is typically organised into histograms. As a result, the observable is represented not as individual values but as the content of histogram bins. A likelihood model can be constructed from primary and auxiliary measurements. Primary measurements are based on observed data ($\vec{n}$) and are associated with parameters of interest ($\vec{\mu}$), while auxiliary measurements are existing measurements [1] focusing on systematic variations ($\vec{a}$) of data and are linked to nuisance parameters ($\vec{\sigma}$).

A tool called `HistFactory` [9] is used to construct likelihood functions from histograms based on `ROOT`. A typical likelihood function for binned distributions can be written as follows:

$$\mathcal{L}(\mu, \vec{\sigma}; \vec{n}, \vec{a}) = \prod_{i=1}^{\text{total bins}} P(n_i|\mu S_i(\sigma) + B_i(\sigma)) \cdot \prod_{i}^{\text{total bins}} G(a_i|\sigma_i) \tag{1.13}$$

For bin i, the conditional probability of observing n_i events under the prediction of S_i signal events and B_i background events can be given using a Poisson function. Since the bins are independent, the product of their Poisson probabilities forms the first term of the likelihood function. It is a common practice to introduce a signal strength parameter μ such that $\mu = 1$ corresponds to signal+background hypothesis and $\mu = 0$ corresponds to background-only hypothesis. The goal of fitting is to estimate the signal strength which is our parameter of interest. It is also regarded as the signal normalisation factor.

The second term is a product of per-bin Gaussian probability functions associated with auxiliary measurements. It is called a constraint term because it restricts the likelihood function based on known information, for instance, systematic uncertainties. Given the fact that there can be many systematic uncertainties, it is difficult to handle nuisance parameters all with different values. To simplify the handling, a standard convention is used where the nuisance parameters are redefined by scaling each Gaussian in such a way that its mean is 0 and standard deviation is 1.

In principle, one can optimise the likelihood based on each value of the nuisance parameter in order to capture the effect of that particular systematic variation. However, in practice, this is computationally intensive. Therefore, for each NP we use template histograms: nominal, up variation and down variation. A nominal histogram is filled with predicted events for NP value equal to its mean, i.e. 0. The up(down) variation histograms are filled with predicted events for NP value equal to $+1(-1)$. Now, to get the predicted events for all possible values of the NP, interpolation and extrapolation is used.

[1] In ATLAS, calibrations of various nuisance parameters is performed by dedicated combined-performance (CP) groups. Their results are utilised and validated in physics analysis.

Implementation

In this analysis, a software framework called `TRExFitter` [10] is used to perform profile likelihood fitting. `TRExFitter` builds binned template likelihood functions and performs statistical analysis using tools such as `HistFactory`, `RooStats` [11] and `RooFit` [12].

The input to `TRExFitter` consists of template histograms, which are organised into two main categories:

- **Regions**: A region refers to a subset of the data defined by a set of specific event selection. For instance, a region can be defined by selecting events passing a certain neural network cut. The selection criteria for regions should be decided in such a way that the regions are disjoint from one other. This avoids double counting of events. The fitting is performed in each region.
- **Samples**: A sample corresponds to a specific process such as a signal process or a background process. Each sample is represented by a template histogram describing the expected distribution of events for that process in a given region.

The output of the fitting process includes the following:

- Best-fit values of the POI and the NPs which maximise the likelihood function.
- Best-fit values of the normalisation factors. These factors are used to adjust or scale signal and background contributions such that it aligns with the observed data.
- Uncertainties on the best-fit values quantifying the precision of the fit. `TRExFitter` uses a tool called `MINOS`(CITE) to compute the uncertainties.
- A covariance matrix quantifying the correlation between all parameters.

1.6 Profile likelihood unfolding (PLU)

PLU is a method in which the unfolding problem is translated into a profile likelihood fitting problem as described in Section 1.5. The modified likelihood function used for unfolding is defined in Eq. (1.14).

$$\mathcal{L}(\vec{\mu},\vec{\sigma};\vec{n},\vec{a}) = \prod_{i=1}^{\text{total bins}} P(n_i|\vec{\mu}\vec{S}_i(\sigma)+B_i(\sigma)) \cdot \prod_{i}^{\text{total bins}} G(a_i|\sigma_i) \tag{1.14}$$

The main difference is that instead of having one parameter of interest representing overall normalisation of signal events, there is a normalisation vector with dimensions equal to number of bins in the truth histogram (truth-level bins). In PLU, each truth-level bin is multiplied by the response matrix resulting into sub-histograms per truth-level bin. In unfolding terminology, this is known as *folding* the truth-level bins. Each sub-histogram or folded truth bin is assigned a normalisation factor or scale factor which is a free parameter. The sum of these sub-histograms represent the total signal contribution. Finally, a profile likelihood fit of the reconstructed-level distribution is performed on the sum of the sub-histograms yielding best fit values of the normalisation factors. It is important to note that fitting the normalisation of the sub-histograms is equivalent to fitting the normalisation of the

truth-level bins, which is the desired unfolded result. In this analysis, the normalisation factors are interpreted as differential cross-section.

For a kinematic variable X in bin i, a differential cross-section can be calculated by dividing the unfolded yields by bin width ΔX and luminosity $\mathcal{L}$ as presented in Equation 1.15.

$$\frac{d\sigma_{\text{PLU}}}{dX^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i} N^i_{\text{unf}} \tag{1.15}$$

where unfolded bin contents N_{unf} are calculated by scaling the truth-level bins with the best fit values of the normalisation factors.

normalised differential cross-section

Systematic uncertainties are handled in the same way as in a standard profile likelihood fit. Additionally, for unfolding purpose, PLU requires a response matrix for each systematic variation. Instead of a response matrix, one can also provide migration matrix, acceptance and efficiency.

There are some advantages of using profile likelihood unfolding compared to other unfolding methods. Mainly, PLU does not involve any matrix inversion, hence, the chance of amplifying statistical fluctuations is minimum. Since the `TREXfitter` framework is optimised for handling systematic uncertainties, it is quite easy to utilise it for unfolding as well. The formalism also allows adding free-floating normalisation factors for specific backgrounds that can be constrained by adding control regions in the fit. The *tZq* analysis also includes measurement of inclusive cross-section which is also done using the `TREXfitter` framework. Therefore, using PLU for differential measurement allows to have a common statistical framework for the complete analysis.

Bibliography

[1] H. B. Prosper, *Practical Statistics for Particle Physicists*, arXiv e-prints, arXiv:1504.00945 (2015) arXiv:1504.00945, arXiv: 1504.00945 [stat.ME] (cit. on p. 1).

[2] L. Lyons, *Statistics for Nuclear and Particle Physicists*, Cambridge University Press, 1986 (cit. on p. 1).

[3] L. Lyons, "Unfolding: Introduction", *PHYSTAT 2011*, Geneva: CERN, 2011 225 (cit. on p. 3).

[4] G. Cowan, *A survey of unfolding methods for particle physics*, Oxford University Press, 1998 (cit. on p. 5).

[5] Spanò, Francesco, *Unfolding in particle physics: a window on solving inverse problems*, EPJ Web of Conferences **55** (2013) 03002, URL: https://doi.org/10.1051/epjconf/20135503002 (cit. on p. 5).

[6] Schmitt, Stefan, *Data Unfolding Methods in High Energy Physics*, EPJ Web Conf. **137** (2017) 11008, URL: https://doi.org/10.1051/epjconf/201713711008 (cit. on p. 5).

[7] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **362** (1995) 487, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/016890029500274X (cit. on p. 5).

[8] *RooUnfold*, Accessed on: 12.10.2024, URL: https://gitlab.cern.ch/RooUnfold/RooUnfold (cit. on p. 6).

[9] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., 2012, URL: https://cds.cern.ch/record/1456844 (cit. on p. 7).

[10] *TRExFitter documentation*, Accessed on: 20.11.2024, URL: https://trexfitter-docs.web.cern.ch/trexfitter-docs/ (cit. on p. 8).

[11] L. Moneta et al., *The RooStats Project*, PoS (), arXiv: 1009.1003 [physics.data-an] (cit. on p. 8).

[12] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003, arXiv: physics/0306116 [physics.data-an] (cit. on p. 8).

List of Figures

List of Tables