CHAPTER 10

STATISTICAL METHODS

Once the event selection has been performed and the selected events are binned according to the differential kinematical distributions (as described in Chapter 6), the next step is to perform the statistical analysis in order to extract the confidence intervals (CIs) for the WCs. In Section 10.1 we will step through the relevant statistical concepts and tools. In Section 10.2, the details of the statistical framework will be be described. Finally, in Section 10.3, one of the challenges of multi-dimensional EFT fits will be discussed, and the workaround developed to mitigate this challenge will be explained.

10.1 Likelihood fitting

Let us start by defining the likelihood L as the probability to have measured the observed number of data events, given some theory, i.e. $L = P(\text{data}|\text{theory})$. The number of observed events should follow a Poisson distribution, with a mean corresponding to the number of predicted events. Since each bin is statistically independent, the likelihood will be given by the product of the Poisson probabilities for all of the bins in the analysis.

For many analyses, the predicted number of events in a given bin i can be written as $\mu s_i + b_i$, where b_i is the expected number of background events, s_i is the expected number signal events (according to the SM prediction), and μ is a free parameter. The μ parameter is usually referred to as the signal strength; it must be greater than or equal to 0, and it is constant across all bins. However, in this analysis, we cannot

write the predicted number of events as $\mu s_i + b_i$, because the prediction does not scale linearly with a universal signal strength μ. Rather, the predicted number of events in each bin depends quadratically on the 26 WCs (and the quadratic parameterization is different in each bin); we will write this prediction as $\mu(\theta)_i$, where θ are the values of the WCs. This function $\mu(\theta)_i$ represents the prediction in each bin (which in our case is a 26-dimensional quadratic in terms of the WCs), and should not be confused with the signal strength μ. With this notation, we can write the likelihood as follows:

$$L = \prod_{i=1}^{N} \frac{\mu(\theta)_i^{n_i}}{n_i!} e^{-\mu(\theta)_i}, \tag{10.1}$$

where N is the number of bins, n_i is the number of observed events in bin i, and $\mu(\theta)_i$ is the number of predicted events in bin i (as a function of the WCs θ).

In Eq. 10.1, the θ represents the set of 26 dimension-six WCs studied in this analysis. To understand the likelihood's dependence on the WCs, we could in principle perform a 26-dimensional grid scan. To perform such a scan, we would choose a reasonable range for each WC (based on the estimated sensitivity to the WC), chose a granularity with which to scan, and then proceed to record the likelihood at each point on the 26-dimensional grid. However, this "brute-force" approach scales exponentially with the number of dimensions, and becomes prohibitively expensive when more than a few dimensions are considered. Let us step through an example for 26 dimensions. Even if we chose a very sparse grid with only 5 scan points in each dimension, we would still need to scan 5^{26} points. Assuming one hour per scan point (a typical length of time for the fits in this analysis) and 10k CPU cores (a reasonable amount of resources we could utilize with an opportunistic pool such as Notre Dame's CRC), it would take about 17 billion years to perform the scan. This brute force approach is thus not feasible for our analysis.

Instead of a 26-dimensional scan, we perform a one-dimensional scan for each WC, profiling the other 25 WCs. Continuing to refer to the scanned parameter as

θ, let us refer to the profiled parameters as ν. In the one-dimensional scan, we step along one axis θ in the 26-dimensional space (i.e. we step through a set of values for one WC). At each of the steps along the θ axis, the profiled parameters ν are set to the values that cause the likelihood to be maximized at this given value of the scanned parameter θ. The profile likelihood L_p is thus written as follows:

$$L_p(\theta) = L(\theta, \hat{\hat{\nu}}(\theta)), \tag{10.2}$$

where the double-hat notation denotes the values of the profiled parameters ν that maximize the likelihood for the given θ. The profiled likelihood is thus a function of θ only; it is not a function of ν, since the profiled parameters ν do not freely vary (they are a function of θ). As these scans are only one-dimensional, they are much less computationally expensive than the 26-dimensional scan described above. We can thus perform a one-dimensional scan for each of the WCs in order to determine the one-dimensional profile likelihood for each dimension. In principle we can extend this concept to scan over any number of the WCs (profiling the remaining WCs); however, in practice, the largest number of parameters we can scan is two, since even for a three-dimensional scan, the space is too large to fully explore with our current computational abilities.

Next, we would like to understand how the profile ratio compares to the maximum likelihood as a function of θ. We will refer to the values of θ and ν that globally maximize the likelihood as $\hat{\theta}$ and $\hat{\nu}$, respectively. We can then write the profile likelihood ratio $\lambda_p(\theta)$ as follows:

$$\lambda_p(\theta) = \frac{L(\theta, \hat{\hat{\nu}}(\theta)}{L(\hat{\theta}, \hat{\nu})}, \tag{10.3}$$

where the numerator is the profile likelihood from Eq. 10.2, and the denominator is the value of the likelihood at its global maximum. From the profile likelihood ratio,

we can form the test statistic $-2\ln\lambda_p(\theta)$. Wilks' theorem [70] states that $-2\ln\lambda_p(\theta)$ should approach a χ^2 distribution in the limit where the data sample is large, where the degrees of freedom correspond to the number of free parameters in the $\lambda_p(\theta)$. For example, for a one-dimensional scan, there is one degree of freedom. To find the one-dimensional confidence intervals for a given WC, we would thus need to perform a scan for the WC, finding the $-2\ln\lambda_p(\theta)$ at each scan point; since the $-2\ln\lambda_p(\theta)$ is assumed to follow a χ^2 distribution with one degree of freedom, we can read off the one and two σ confidence intervals by observing where the $-2\ln\lambda_p(\theta)$ crosses one and four, respectively [71].

This section has discussed how the predicted yield in each bin depends on the parameters of interest (the WCs), but the prediction also depends on the systematic uncertainties (enumerated in Chapter 9). The systematic uncertainties are taken into account via additional free parameters in the fit; these degrees of freedom are referred to as nuisance parameters. When finding the profile likelihood $L(\theta, \hat{\hat{\nu}}(\theta))$, the nuisance parameters are profiled (i.e. they may be included in the ν in Eq. 10.3).

10.2 Statistical framework

The CMS Higgs `Combine` software tool [72] is used to perform the likelihood fits. The `Combine` tool uses the `ROOT` framework's `RooFit` tools [73] and the `MINUIT2` software library [74] to numerically minimize the negative log of the profile likelihood function described in Section 10.1. As discussed in Section 3.2.2, the expected yield in each bin is parameterized as a 26-dimensional quadratic in terms of the WCs, as given in Eq. 3.5. In order for `Combine` to calculate the profile likelihood described in Section 10.1, this quadratic dependence must be made known to `Combine`. In principle, a template histogram could be defined for each of the 378 structure constants of the 26-dimensional quadratic, with the appropriate normalizations of the templates specified by the `Combine PhysicsModel`. However, in practice this is not possible

with `Combine`, since the interference terms of the quadratic may be negative, and `Combine` does not handle histograms with negative yields.

To work around this challenge, we use the approach developed in [75]. With this approach, the 378 terms of the quadratic parametrization are rearranged into 378 linear combinations of the original terms, defined such that each term is positive by construction. We can then create a template histogram for each of the rearranged terms, encoding the appropriate normalization of each histogram in the `Combine PhysicsModel`. The template histograms and normalizations encoded in the `PhysicsModel` contain the full description of the 26-dimensional quadratic function, so `Combine` is able to appropriately handle this dependence while performing the likelihood fits.

The expected yields also depend on the experimental and theoretical systematic uncertainties (enumerated in Chapter 9), the effects which are taken into account by a set of nuisance parameters. As mentioned in Section 10.1, the nuisance parameters are profiled in the likelihood fit. The systematic uncertainties may affect either the normalization of the template histograms, or both the normalization and the shape of the template histograms. The former are accounted for via rate systematics in `Combine`, and the latter are accounted for via shape systematics in `Combine`. The systematic uncertainties that affect the shape of the template histograms also carry a 26-dimensional quadratic dependence on the WCs (discussed in Section 9.1), which is accounted for in the same way as the nominal templates.

10.3 Navigating false minima

Multi-dimensional EFT fits can involve features in the likelihood surface that are difficult for profile fits to navigate. Specifically, if these features include local minima, the profile fits may be susceptible to incorrectly identifying a local minimum as the true minimum (i.e. the fit may become "stuck" in the local minimum). Cases

where the fit becomes "stuck" in local minima may lead to false best fit points, discontinuities in the negative log likelihood (NLL) scans, and inaccurate confidence intervals. It is thus important to identify and to mitigate this issue.

Symptoms of this issue had been observed in the predecessor to this analysis (Ref. [10]) in the form of discontinuities in the NLL values obtained in the one-dimensional scans. To work around this issue, Ref. [10] performed two-dimensional scans for pairs of WCs that had been identified as problematic, avoiding the discontinuities by making use of the NLL values obtained in the two-dimensional scans. However, this approach is not only computationally expensive, but it also does not guarantee that the correct minima will be found (as the fit may encounter similar challenges with false minima while profiling the remaining $n-2$ WCs); a more general approach would be beneficial. Since local minima can arise as a result of the interference terms in the n-dimensional quadratic parameterizations, the challenge of navigating local minima seems to be an inherent feature of multi-dimensional EFT likelihood fitting. As the EFT community continues to explore simultaneous fits to larger sets of WCs, this pernicious issue may become increasingly problematic. For these reasons, one of the intermediate goals of this analysis was to develop a more general approach to the navigation of local minima within the `Combine` framework.

Before developing an approach to address these issues, we wanted to first gain a better understanding of the underlying cause. To this end, we worked to reproduce the issue in a much simpler case. In this simplified model, we only considered two WCs ($c_{\varphi Q}^{-}$ and c_{tG}). We then performed a profiled likelihood fit, scanning over $c_{\varphi Q}^{-}$ and profiling c_{tG}; in other words, we asked the fit to step through a set of given points for $c_{\varphi Q}^{-}$, and at each of those points to find the value of c_{tG} that would minimize the NLL. The result of this likelihood fit is shown in the lefthand side of Figure 10.1.

At a value of approximately $c_{\varphi Q}^{-} = 17$, a discontinuity is observed. To understand this discontinuity, we performed a two-dimensional scan over both WCs. This allows

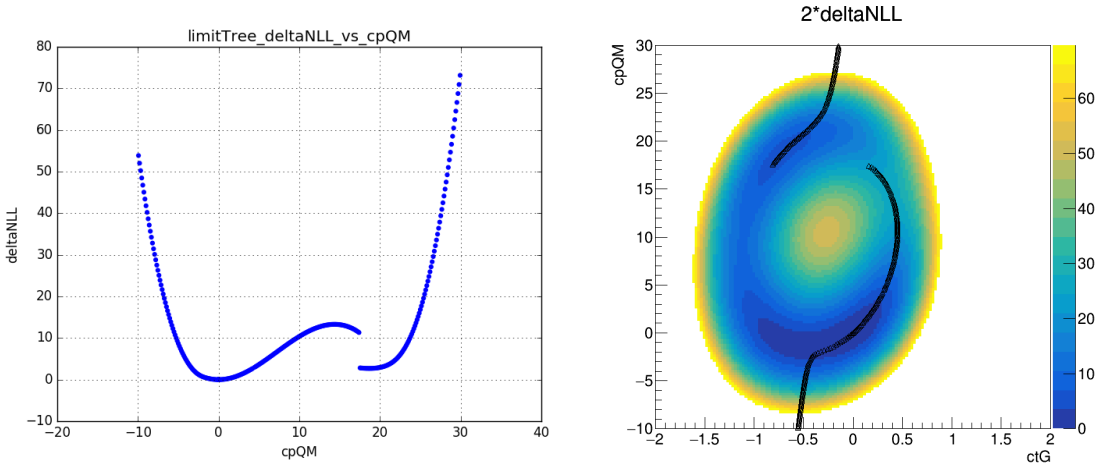

Figure 10.1. Left: NLL for the one-dimensional profiled fit. The $c_{\varphi Q}^{-}$ parameter is scanned while c_{tG} is profiled. Right: NLL for the two-dimensional scan. Here both $c_{\varphi Q}^{-}$ and c_{tG} are scanned. The color scale shows the NLL at each of the two-dimensional scan points. The black overlaid points show the path of the one-dimensional profiled fit.

us to see the complete picture of the space, helping us to visualize why the profiled fit fails. The two-dimensional scan is shown on the righthand side of the Figure 10.1, with the path of the one-dimensional profiled fit overlaid in black points. In the one-dimensional profiled fit, $c_{\varphi Q}^{-}$ was scanned while c_{tG} was profiled, meaning that for every point along the y direction (i.e. the $c_{\varphi Q}^{-}$ direction) of the one-dimensional scan, the fit profiles along the x direction (i.e. the c_{tG} direction) in order to find the c_{tG} value at which the NLL (represented by the color scale) is minimized. Following the path of the profiled fit (the overlaid black points), we see that the fit was correctly identifying the c_{tG} point that minimized the NLL from $c_{\varphi Q}^{-} = 0$ until approximately $c_{\varphi Q}^{-} = 5$. At this point, the true minimum lies on the left of the"hill" in the NLL, but the fit continues around the right of this "hill", subsequently incorrectly identifying these local minima as the best fit points. Once the scan reaches approximately $c_{\varphi Q}^{-} = 17$, the fit suddenly jumps to the deeper minimum on the left side of the "hill", resulting in the discontinuity observed in the one-dimensional NLL plot.

To avoid discontinuities, the ideal solution would be to perform a simultaneous scan over all parameters. However, as discussed in Section 10.1, this approach scales exponentially with the number of parameters, and is infeasible for the case of 26 parameters of interst. As an alternative approach, we introduced an element of randomness into the fit in order to sample from the 26-dimensional space. To motivate this approach, let us revisit the profiled fit discussed above. During a fit, `Combine` always uses the same starting value for the profiled parameters; this is useful for reproducibility, but it means that if the starting point happens to be near a local minimum (where the global minimum lies on the other side of a "hill" in NLL), the fit will always find the local minimum, and never find the correct global minim. If we instead allow the starting point for the profiled fit to be chosen randomly, the starting point will sometimes lie on the other side of the "hill" in NLL, allowing the fit to find the correct minimum.

To test the random starting point method, we modified the `Combine` tool to incorporate random starting points for the profiled parameters. After first generating a list of random starting points, the modified version of the `Combine` script loops through the random starting points, finding the NLL at each, and keeping track of the point that gives rise to the lowest NLL. After trying each of the random starting points, the point that produced the lowest NLL is taken to be the profiled value of the parameter at that point. Applying this approach to the two-dimensional case described above, the NLL is found to be continuous, as shown in Figure 10.2, where the profiled fit finds the correct global minimum at each scan point. For example, at approximately $c^{-}_{\varphi Q} = 5$, the fit is able to identify the deeper minimum on the left side of the "hill" in NLL, jumping to that minimum immediately instead of erroneously continuing around the right side of the "hill" (as had been observed in Figure 10.1).

After the random starting point approach was shown to be successful in the simple two-dimensional case, we generalized the method and tested it on the 16-dimensional

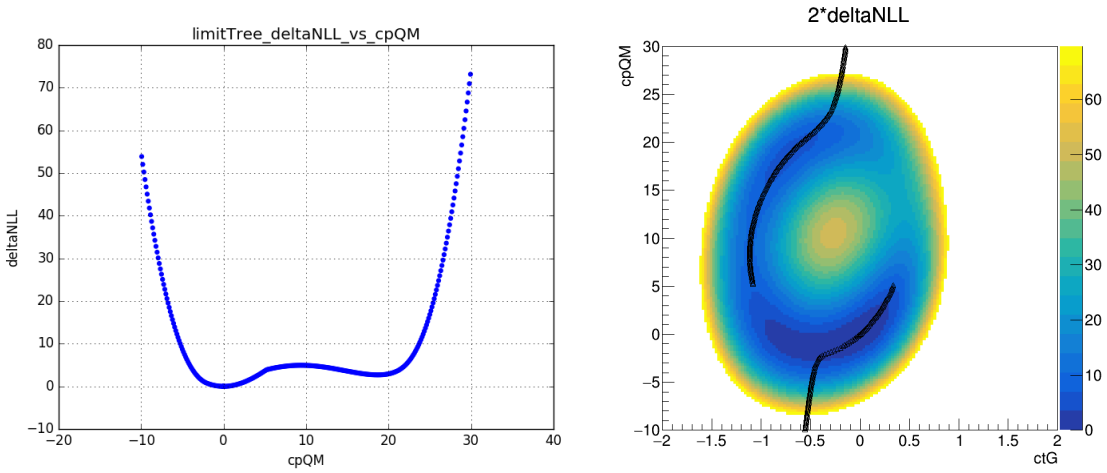

Figure 10.2. Left: NLL for the one-dimensional profiled fit after the implementation of the random starting point approach, where $c_{\varphi Q}^{-}$ is scanned and c_{tG} is profiled. Right: NLL for the two-dimensional scan over $c_{\varphi Q}^{-}$ and c_{tG}. The color scale shows the NLL at each of the two-dimensional scan points. The black overlaid points show the path of the one-dimensional profiled fit after the implementation of the random starting point approach.

fits from [10]. Although the number of random starting points required in order to obtain a smooth NLL curve was larger in the 16-dimensional case than in the two-dimensional case (greater than 50 as opposed to less than 10), the method successfully avoided discontinuities in NLL. The results of this test are shown in Appendix E.

Although this approach is computationally feasible, it is still relatively computationally expensive (as the likelihood fit must be run m times for each scan point, where m is the number of random starting points). In the future, it would be interesting to optimize the approach by considering methods of sampling the space more efficiently. For example, one idea would be to first identify a set of distinct local minima in the space, and using a set of points from these local minima as the starting values instead of choosing the starting points randomly from the full space.