Evaluating and Improving the Truncated Normal Approximation of Likelihoods

Leo Maran

Institute for High Energy Physics

Vienna University of Technology

Project Work

carried out under the guidance of

DI Dr. Wolfgang Waltenberger
MMag Dr. Manfred Jeitler

2021

Abstract

Since the discovery of the Higgs boson, particle physics has been in a difficult situation: while the current Standard Model (SM) is most likely incomplete, no conclusive evidence of new physics has been found experimentally. To simplify the search, the software *SModelS* has been developed in order to more easily compare data from particle detectors to theoretical predictions.

This in turn makes possible the reverse approach to the problem. Instead of comparing theoretical predictions to measured data, the data can be used to infer so called proto-models, whose predictions maximally violate the SM-hypothesis. These proto-models could allow for the detection of signals that might be dispersed in the data and hence went unnoticed up to now, and serve as precursors to beyond the Standard Model (BSM) theories.

In the process of proto-modelling, it is of vital importance to ascribe likelihoods to the signal strength of BSM processes in certain signal regions. While these likelihoods get calculated by the experimental collaborations and enter the calculations of the published results (e.g. upper limits), more often than not they themselves don't get published. In these cases, the likelihoods have to be approximated by means of the data available.

The scope of this project work is to evaluate the quality of the simplest approximation, used in cases, where only the expected and observed 95% CL upper limits of cross sections are available. In order to do this, the likelihoods are compared to ones constructed in a more sophisticated manner, used in cases where more information is available.

More specifically, the data used to construct the more sophisticated likelihoods gets also used to compute 95% CL upper limits, from which in turn the simpler likelihoods get constructed, thus making it possible to compare the two for the same data.

Additionally, a modification to the simplest approximation procedure gets proposed, improving by making it more conservative.

Kurzfassung

Seit der Entdeckung des Higgs Bosons ist die Teilchenphysik in einer schwierigen Situation: Das Standardmodell (SM) scheint unvollständig zu sein, aber bis heute wurden experimentell keine signifikanten Abweichungen von dessen Vorhersagen gefunden. Zur Vereinfachung dieser Suche wurde das Softwarepaket *SModelS* entwickelt, welches den Abgleich von theoretischen Vorhersagen mit experimentellen Daten erleichtert.

Mithilfe von *SModelS* wird es auch möglich einen neuen Ansatz der Theoriebildung zu verfolgen. Statt Vorhersagen von etablierten Modellen mit experimentellen Daten zu vergleichen, können Daten als Ausgangspunkt benutzt werden, um sogenannte Proto-Modelle zu konstruieren, welche die SM-Hypothese maximal verletzen. Das erlaubt es, potenziell signifikante Signale zu finden, die in den Daten verstreut sind, und deshalb bis jetzt unentdeckt blieben. Solch ein Proto-Modell könnte dann als Vorläufer für eine neue, verbesserte Theorie dienen.

Für die Konstruktion von Proto-Modellen wir ein Maß für die Plausibilität von neuer Physik in bestimmten Signalregionen benötigt. Dieses kann ausgedrückt werden durch eine Likelihood-Funktion für die Signalstärke neuer Physik. Obwohl diese in die Berechnung von publizierten Ergebnissen (z.B. 95% obere Konfidenzgrenzen) der Experimentellen Kollaborationen eingehen, werden sie selbst meist nicht veröffentlicht. In solchen Fällen müssen die Likelihoods mithilfe der zur Verfügung stehenden Daten angenähert werden.

Im Zuge dieser Projektarbeit soll die Qualität der einfachsten Näherung überprüft werden. Sie kommt zum Einsatz, wenn nur die erwartete und beobachtete obere Grenzen der Signalstärke zur Verfügung stehen. Als Vergleich wird eine komplexere Likelihood herangezogen, welche berechnet werden kann, wenn mehr Daten verfügbar sind. Um die Likelihoods vergleichen zu können, werden aus den umfangreicheren Daten die erwartete und beobachtete obere Grenze der Signalstärke berechnet, aus welchen dann wiederum die einfachste Näherung berechnet wird. Somit erhält man zwei vergleichbare Likelihoods aus denselben Daten.

Zusätzlich wird im letzten Kapitel eine Modifikation der Näherung vorgeschlagen. Diese Verbessert die Übereinstimmung der beiden Likelihoods und sorgt dafür, dass die einfachste Näherung konservativer wird.

Contents

Chapter 1

Introduction

Since its finalization in the 70s, the Standard Model of particle physics (SM) has been extremely successful at describing the natural world. It is in fact so successful, that since the discovery of the Higgs boson in 2012, which was the last missing particle the SM predicted, particle physics has been in a difficult situation: Despite the overwhelming amount of experimental evidence that suggests the validity of the SM, it fails to account for a series of phenomena. These include the nature of dark matter, the hierarchy problem and the prevalence of matter over antimatter. Moreover, the SM does not include one of the four fundamental forces at all, namely gravity.

It seems clear that the SM is most likely not our final theory. However, to this day no significant deviations from its predictions have been found, even tough there have been some candidates over the years. The most recent one concerns the anomalous magnetic moment of the muon. Measurements suggest with a significance of 4.2σ that its value deviates from the one predicted by the SM [1]. While the significance has been increasing with each published result, it still falls short of the 5σ threshold. This leaves theorists and experimentalist alike with the very difficult task of finding new physics without knowing where to look. Still, over the years many proposals for extensions or even replacements of the SM have been made, of which one of the most promising candidates is called Supersymmetry (SUSY). [2]

SUSY is an extension to the SM that introduces a symmetry transformation between bosons and fermions, which implies that every fermion has a bosonic super-partner and every boson a fermionic one. These super-partners are called sparticles and could solve some problems of the SM. For example the hierarchy problem: If for every loop contributing to the Higgs mass there is a corresponding loop of sparticles that provides a correction with the opposite sign, fine-tuning can be avoided in a fashion that appears much more natural.

Moreover, most forms of SUSY also offer a dark matter candidate. To account for observations like the stability of the proton, a new conserved quantity is introduced, R-parity. All SM particles have odd and all sparticles have even R-parity. This implies two things: sparticles can only be produced pairwise, and every sparticle decays to the lightest supersymmetric particle (LSP), which cannot decay further and is thus stable. The LSP has to be electrically neutral if it is to explain dark matter, which means it can only be observed in particle detectors as missing transversal energy (MET).

Up until today no sparticles have been detected, which can be integrated into SUSY by demanding that supersymmetry is broken (analogous to the broken $SU(2)$ symmetry of the SM).

1.1 SModelS

Experiments at the Large Hadron Collider (LHC) at CERN in Geneva are looking for signs of new physics. However, there are several competing beyond the Standard Model (BSM) theories that each have free parameters (SUSY has more than 100, but their number can be reduced by making some assumptions). This makes the interpretation of the results quite difficult. A fully specified BSM theory usually predicts several observable quantities, e.g. masses and decays of new particles. While it is possible to make exclusions in the parameter space for each model, this would be very expensive in terms of CPU time. But since a wide range of predicted BSM theories lead to similar phenomenologies, simplified model spectra (SMS) are introduced as a tool to interpret LHC data. [3]

Within the SMS framework, the number of detected BSM processes only depends on the event kinematics and not on specific details of the model, and the simplest decay has a branching ratio of 100%. This means, that if one assumes a certain SMS topology (a decay cascade with fixed vertices), the signal efficiency of the corresponding process can be expressed as a function of the mass of the involved BSM particles.

This allows for the calculation of upper limits on cross sections of certain BSM modes in the mass plane of the involved BSM particles. The interpretation of these results in terms of a full BSM theory is however nontrivial. For this purpose, the tool *SModelS* has been developed [4]. It takes a full BSM theory, decomposes it into single topologies and tests those against experimental results, thus allowing to quickly falsify some theories.

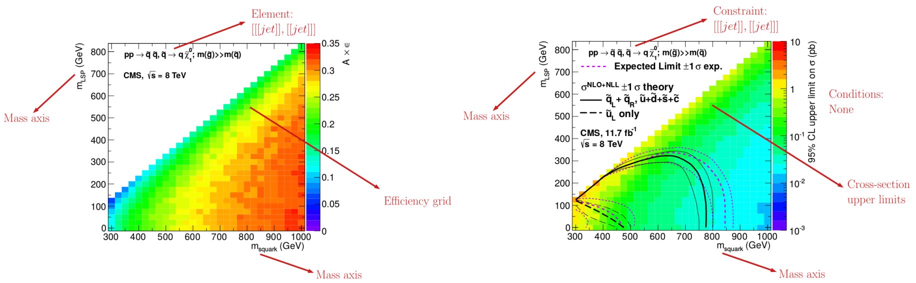
At the heart of *SModelS* lies its extensive database of experimental results. They fall into two main categories [5]:

- Upper Limit (UL) results: These contain 95% confidence level (CL) upper Limits for the production cross section times branching ratio for a simplified model topology. Typically, they are given as Upper Limit maps in BSM mass planes. Some results also include the expected upper limits.

- Efficiency Map (EM) results: These contain signal efficiencies times Acceptance for simplified model topologies. They are given as a function of the BSM masses and also include information for expected and observed events and respective uncertainties for each signal region.

1.2 Proto-models

SModelS and its database of experimental results make another approach of searching for new physics feasible: instead of developing new theories, deducing what predictions they produce, then checking whether experimental evidence is found in the data, the inverse route is taken. This means that rather than



(a) Efficiency Map

(b) UL Map

Figure 1.1: Examples for EM and UL type results in the *SModelS* database [5].

decomposing full models into SMS topologies, so called proto-models are produced. [6] These are defined by the particles and respective masses they contain together with branching ratios and production cross sections. While they are in no way complete BSM theories, ultimately their purpose should be to act as a data driven starting point for the development of new theories. The prime motivation for this approach is the possibility to look for BSM signals that might be dispersed in the data and hence went unnoticed until now. To achieve this, one tries to create proto-models that maximally violate the SM hypothesis.

Hence, a Markov Chain Monte Carlo (MCMC)-inspired algorithm is employed to walk the parameter space, randomly changing the number of new particles, their masses, production rates and branching ratios. This process relies on the construction of an approximate likelihood for the signal, given by

$$L_{BSM}(\mu|D) = P(D|\mu + b + \theta)p(\theta) \tag{1.1}$$

This can be interpreted as the plausibility of the signal strength μ, given the Data D, where θ are the nuisance parameters with distribution $p(\theta)$, describing the uncertainty in the signal μ and background b.

The algorithm roughly consists of the following steps:

1. The current model gets randomly modified, adding or removing particles and changing parameters.

2. The model is checked against the database of results to determine μ_{max}, the upper bound of the signal strength.

3. The possible combinations c of results are checked, for each one a likelihood $L^c_{BSM}(\mu)$ is created.

4. The test statistic K is computed. If the current K is higher than the previous one, the new model is kept, otherwise the step is reverted with a probability of $[\frac{1}{2}\exp(K_i - K_{i-1})]$, where i is the index of the current step.

The test statistic K^c for a combination of results is defined as:

$$K^c := 2\ln\left(\frac{L^c_{BSM}(\hat{\mu}) \times \pi(BSM)}{L^c_{SM} \times \pi(SM)}\right) \tag{1.2}$$

where $\hat{\mu}$ maximises the likelihood and $L_{SM} = L_{BSM}(\mu = 0)$. Furthermore, π_{BSM} and π_{SM} denote the priors of the BSM and SM respectively. These are heuristically chosen and have the function to penalise overly complex models, see reference [6] for additional info. Currently, the best performing proto-model built by said procedure with the *SModelS* database consists of a top partner, a light-flavor quark partner, and a lightest neutral new particle with masses of the order of 1.2 TeV, 700 GeV and 160 GeV, respectively. The corresponding global p-value for the SM hypothesis is $p_{global} \approx 0.19$.

Chapter 2

Constructing Likelihoods

The global likelihood in Equation 1.1 is computed by combining the individual likelihoods of the single analyses. However, likelihoods can only be trivially combined if the analyses are uncorrelated. Since their degree of correlation is mostly unknown, it is binarily approximated: either analyses are considered to be uncorrelated and can hence be combined trivially (i.e. results from different runs, different collaborations or with vastly different final states), or they are not to be combined at all.

In this case, trivial combination means simple multiplication:

$$L_{BSM}(\mu) = \prod_{i=1}^{n} L_i(\mu) \tag{2.1}$$

where the L_i are the individual likelihoods to be combined. How these Likelihoods come about depends heavily on the information provided by the experimental collaborations.

Recently, the ATLAS collaboration started publishing full statistical likelihoods [7]. These describe uncertainties and their correlations across signal regions and are essentially the ones that are used to analyze the experimental data. This is the most desirable form of likelihood for the proto-modelling process, since it contains the most information. However, most often much less information is made public and the likelihoods have to be approximated.

2.1 From Upper Limits

Sometimes only the observed upper limits get published. In this case, no likelihood that would contribute to the global likelihood of equation 2.1 can be constructed. The UL still enters the calculation of μ_{max} however.

Oftentimes both the observed and the expected ULs for μ are available. In this case, the likelihood can be approximated by a truncated normal distribution of the form

$$L(\mu|D) \approx \frac{c}{\sigma_{exp}} e^{-(\mu-\hat{\mu})^2/2\sigma_{exp}^2} \qquad \mu \geq 0 \tag{2.2}$$

which equals a normal distribution centered around $\hat{\mu}$ with standard deviation σ_{exp} that is only defined for positive values of μ, hence the normalisation factor

c. The two parameters of can be evaluated from the two ULs as follows [8]:
Firstly, one demands that the likelihood has to satisfy the observed 95% CL, which can be translated into the constraint

$$0.95 = \frac{\int_0^{\mu_{obs}^{95\%}} d\mu \quad e^{-(\mu-\hat{\mu})^2/2\sigma_{exp}^2}}{\int_0^{\infty} d\mu \quad e^{-(\mu-\hat{\mu})^2/2\sigma_{exp}^2}} = \frac{\mathrm{Erf}\left(\frac{\mu_{obs}^{95\%}-\hat{\mu}}{\sqrt{2}\sigma_{exp}}\right) + \mathrm{Erf}\left(\frac{\hat{\mu}}{\sqrt{2}\sigma_{exp}}\right)}{1 + \mathrm{Erf}\left(\frac{\hat{\mu}}{\sqrt{2}\sigma_{exp}}\right)} \tag{2.3}$$

Secondly, σ_{exp} can be calculated if one assumes the truncated normal likelihood of equation 2.2 with $\hat{\mu} = 0$ (which correspond to the pure background hypothesis) has to satisfy the expected 95% CL

$$0.95 = \frac{\int_0^{\mu_{exp}^{95\%}} d\mu \quad e^{-\mu^2/2\sigma_{exp}^2}}{\int_0^{\infty} d\mu \quad e^{-\mu^2/2\sigma_{exp}^2}} = \mathrm{Erf}\left(\frac{\mu_{exp}^{95\%}}{\sqrt{2}\sigma_{exp}}\right) \tag{2.4}$$

which admits the simple solution

$$\sigma_{exp} = \frac{\mu_{exp}^{95\%}}{1.96} \tag{2.5}$$

With these equations the two parameters of the truncated Gaussian distribution can easily be calculated.

2.2 From numbers of expected and observed events

EM-type results contain more information, namely the number of expected and observed events (including an error on the number of expected events). In these cases it is possible to compute a likelihood as follows:
The probability of observing n events is given by the Poisson distribution

$$p(n, \lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{2.6}$$

where λ corresponds to the number one expects on average. In our case, we take $\lambda = \mu s + b + \theta$, where μs is the number of signal (=new-physics) events, b is the number of background (=SM) events, and θ is a nuisance parameter that gets introduced to account for systematic uncertainties. With some assumptions, one can approximate the distribution of θ to be Gaussian. Therewith a so-called *simplified likelihood* [9] of the form

$$L(\mu|D) = \frac{(\mu s + b + \theta)^{n_{obs}} e^{-(\mu+s+\theta)}}{n_{obs}!} \exp\left(-\frac{\theta^2}{2\delta^2}\right) \tag{2.7}$$

can be constructed, where δ is the signal+background uncertainty. The nuisances θ can be integrated or profiled over, the default when building protomodels is profiling.

Chapter 3

Comparing Likelihoods

Naturally the question arises how good the approximations of chapter 2 are. Answering this was the scope of this project work. More specifically it was investigated, how well the likelihoods constructed from just upper limits by means of a truncated normal distribution correspond to the more sophisticated likelihoods that can be created from EM-type results, and to find potential modifications for the former. For simplicity, from now on the likelihoods constructed by means of a Poisson distribution will be referred to as 'full likelihoods' (even though they are still only approximations).

In order to achieve this, all available EM-type results from the official *SModelS*-database were utilized. For each result the expected and observed upper limits were calculated using the routine implemented in *SModelS*, which is based on the q_μ test statistic of [10]. This allows for the construction of both a full likelihood and a truncated normal likelihood, which were calculated using the functions *profileLikelihood* and *likelihoodFromLimits* (part of the *SModelS* library) respectively. Thus it is possible to compare the results obtained by both methods.

At the time of writing, the official *SModelS* database contains 270 signal refions that contain the necessary information for this type of analysis.

3.1 Visual comparison

Firstly, the likelihoods were compared visually.

There are two distinct cases of results that can be distinguished: overfluctuations and underfluctuations. The former means that more events are observed than were expected, while the latter means that less or just as many events are observed than expected.

Hence, we define overfluctuations to satisfy $\mu_{obs}^{95\%} > \mu_{exp}^{95\%}$ and underfluctuations to satisfy $\mu_{obs}^{95\%} \leq \mu_{exp}^{95\%}$. Of the 270 results, 117 are overfluctuations and 153 are underfluctuations. Amongst the underfluctuations are 47 results where $\mu_{obs}^{95\%} = \mu_{exp}^{95\%}$. Note that the high number of results where the upper limits coincide exactly is likely due to rounding in the computing procedure. In 12 of the 49 cases it holds that $N_{obs} = N_{bg}$, in the others the numbers of observed and expected events are very similar but not exactly equal.

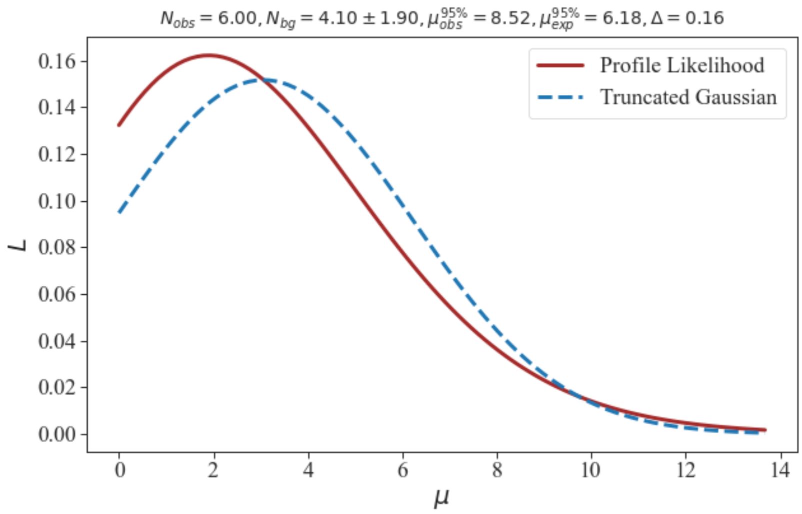


Figure 3.1: Examplary likelihoods. The data that was used to construct them is stated on the top of the plot, $\mu^{95\%}_{obs}$ and $\mu^{95\%}_{exp}$ denote observed and expected upper limit on μ respectively. Furthermore, $\Delta = (\mu^{95\%}_{obs} - \mu^{95\%}_{exp})/(\mu^{95\%}_{obs} + \mu^{95\%}_{exp})$, a quantity that is going to be important later on.

A typical case of overfluctuation is shown in figure 3.1. On it, several characteristic features can be identified. The maximum of the truncated Gaussian likelihood is shifted to higher values of μ when compared to the full likelihood, furthermore it decreases more quickly for high μ. Otherwise, the fit does not seem very bad, from a purely visual standpoint.

In the case of underfluctuations, the two likelihoods are very noticeably different, as shown in figure 3.2. This is because the routine implemented in *SModelS* does not allow the Gaussian distribution to be located at $\mu < 0$. Instead, in the cases where the expected upper limits are smaller than the observed ones, the location of the truncated Gaussian is set to $\hat{\mu} = 0$. This means that the equation 2.3 can not be fulfilled, hence the information how big the underfluctuation was does not enter the likelihood. Consequently, the likelihoods are very different, although they still both have their maximum at $\mu = 0$.

3.2 Comparing moments

The visual comparison of the last section does provide an intuition for the matter, however of course a more thorough investigation is needed. Because of this, the expected value, the variance and the mode of both likelihoods have been calculated for every element of the dataset. Furthermore, also the skewness was calculated, however, since it did not provide valuable information its discussion is omitted.

In the case of the truncated normal distribution the calculations were done analytically using the formulas given in reference [11]. For the full likelihoods the

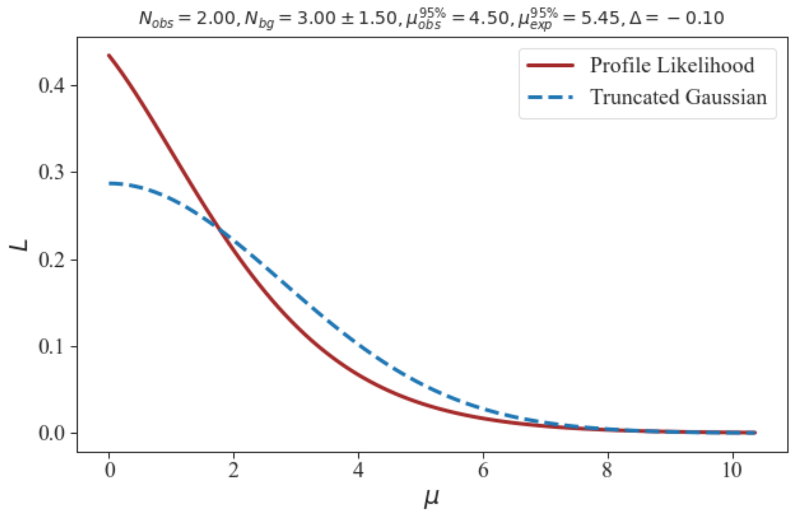


Figure 3.2: Example for an underfluctuation. The truncated Gaussian is located around $\mu = 0$, which clearly does not provide a great fit.

calculations were done numerically. To compare the values, a relative difference

$$\Delta_{rel} = 2\left(\frac{v_{full} - v_{app}}{v_{full} + v_{app}}\right) \tag{3.1}$$

was computed, where v_{full} and v_{app} correspond to the respective values (either expected value, variance or mode) of the profiled and the truncated Gaussian likelihood. The resulting histogram can be found in figure 3.3. On it, several things become apparent that were already visible in the two exemplary likelihoods in figures 3.1 and 3.2.

3.2.1 Expected Values

From the first histogram in figure 3.3 it can be seen that the truncated Gaussian likelihoods tend to have a higher expected value than the full likelihoods. This is not desirable, since it is not conservative when searching for signals. It is apparent, that this mainly stems from the contribution of underfluctuations, since by default for them the Gaussian approximate is located at $\hat{\mu} = 0$, as seen in figure 3.2. It should be noted however, that the relative differences are shifted towards lower values also for the overfluctuations.

3.2.2 Variances

The Variances give a similar picture. The differences corresponding to underfluctuations are shifted towards lower values for the same reason as before. For overfluctuations on the other hand, variances of the approximate likelihoods tend to be smaller than of the full ones. By plotting out the likelihoods, it can be seen that this is caused mainly by the fact that although the likelihoods look similar around their maxima, the full likelihoods tend to have fatter tails. This

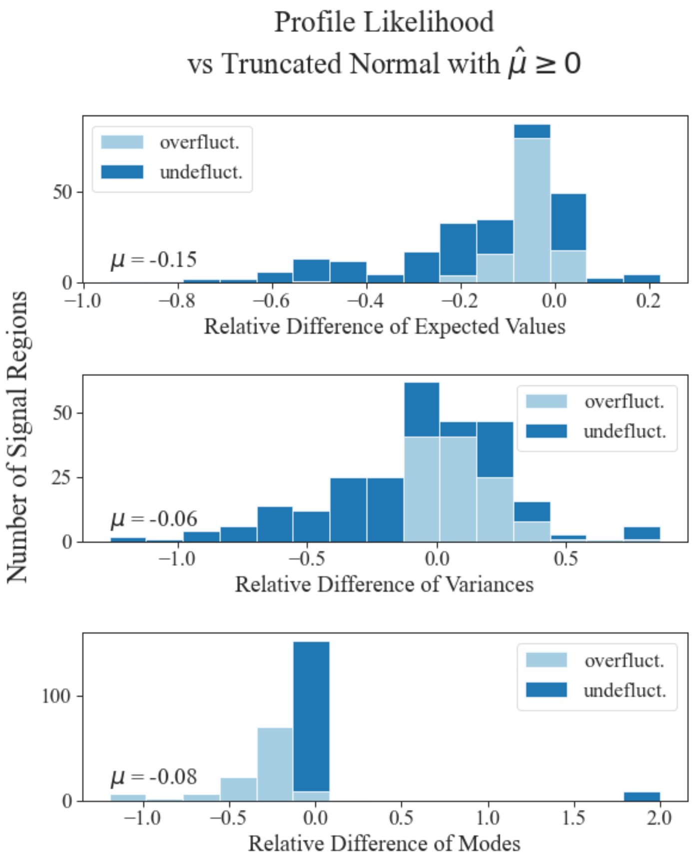


Figure 3.3: Histograms of relative differences of different values for profiled and truncated Gaussian likelihoods. Light and dark blue denote over- and underfluctuations respectively. Negative values mean that the value is higher for the truncated Normal likelihood. The μ in the bottom left corner denotes the arithmetic mean.

can also be seen in the plot above. Generally, one would like the approximate likelihoods to have a bigger variance to be on the conservative side.

3.2.3 Modes

The modes reflect what was said above. Once again, the differences are shifted towards negative values, indicating that the maxima of the approximate likelihoods are bigger than the ones of the full likelihoods.

Note that the few cases where $\Delta_{rel} \approx 2$ are artifacts. These arise when the number of observed events only very slightly exceeds the number of expected events. In this case, both modes are approximately zero, but they nonetheless might be an order of magnitude apart. From equation 3.1 it can be seen easily that in such cases the relative difference is approximately 2.

Ignoring these cases, one can see that the modes indicate even clearer than the expected values that the approximate likelihoods are shifted towards higher values of μ. This is because, as said before, the profiled likelihoods tend to have a fatter tail, which somewhat raises the expected value, cancelling to some extent the shift of the maximum.

The underfluctuations on the other hand show exactly the expected behavior: the likelihoods both have their maximum exactly at 0.

3.3 Dependence on difference of upper limits

In order to come up with modifications for the approximated likelihoods, it was investigated whether the goodness of the Gaussian approximation depends on the difference of expected and observed upper limit. In order to do this, several scatterplots were produced.

3.3.1 Expected value

From figure 3.4 it is clear, that the differences of expected value depend on the differences of upper limits. In the realm of underfluctuations two distinct lines are visible: one for results where the number of observed events was 0, and one where there were observed events. Note the one point with $N_{obs} > 0$ that lies on the line with the results with 0 observed events: it is actually a faulty entry in the *SModelS*-database, since the corresponding entry reads $N_{obs} = 0.00001$.

In the realm of overfluctuations, it can be noted that the approximated likelihoods tend to not be conservative, and that this increases with the difference of upper limits.

From this it might be concluded that a modification that scales with the relative difference of upper limits is a viable approach.

3.3.2 Variance

The variances in figure 3.5 give a very similar picture to the expected values for underfluctuations. Once again, results for $N_{obs} = 0$ (and the faulty database entry) lie on a different line than the other results. For overfluctuations it can be noted that instead of falling with relative difference in upper limits, the difference in variance rises.

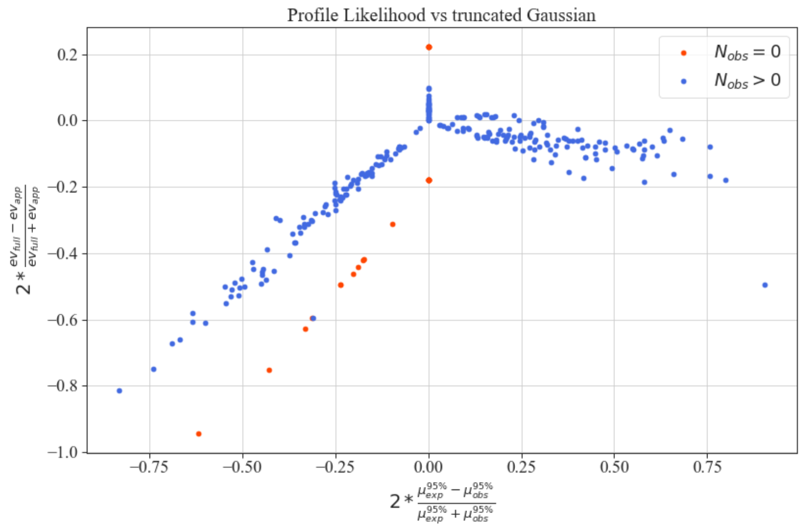


Figure 3.4: Scatterplot of relative differences of expected value over relative differences of upper limits. Overfluctuations are on the right-hand side of the plot, underfluctuations on the left-hand side.

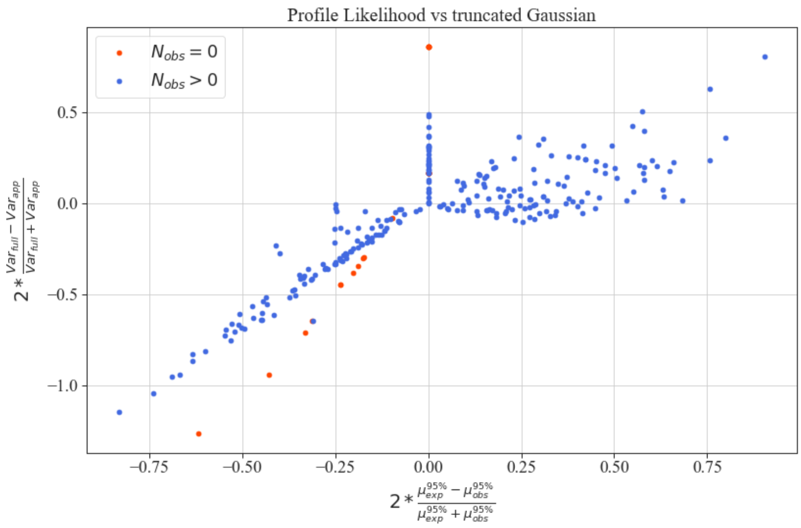


Figure 3.5: Relative differences of variances vs. relative differences of upper limits.

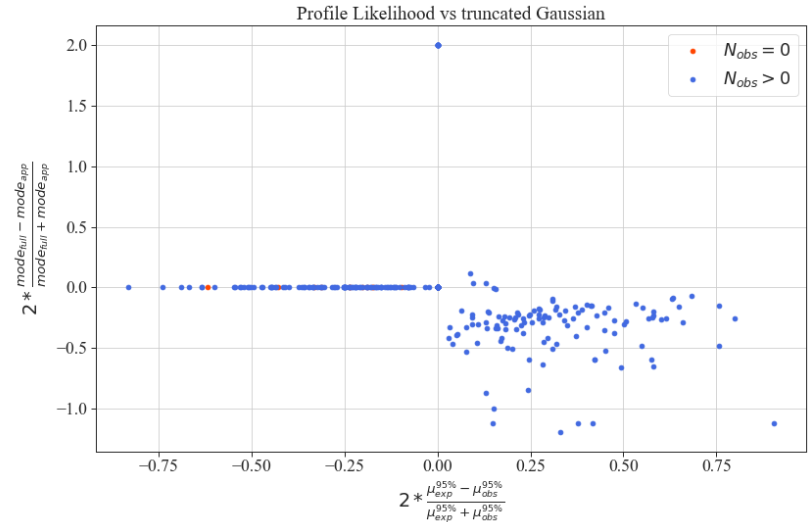


Figure 3.6: Relative differences of modes vs relative differences of upper limits.

3.3.3 Mode

The same scatterplot has been produced for the relative differences of the mode in figure 3.6. It can be seen, that for underfluctuations the relative differences are 0, regardless of the number of observed events. This is expected and has been seen also in the histogram in figure 3.3. The artifacts with $\Delta^{mode}_{rel} = 2$ that were visible in that histogram are also visible in the scatterplot. For overfluctuations, it can be noted that the differences are approximately constant with Δ^{UL}_{rel}.

Chapter 4

Modifying Likelihoods

The aim of the analysis in the previous chapter has been to come up with methods to modify the truncated Gaussian approximation, in order to make it more similar to the full (profiled) likelihood which it was compared to. Ideally, the approximated likelihoods should be modified in a way that satisfies the following criteria:

- the expected value should be unbiased
- the variance should be equal or overestimated
- the mode should approximately coincide

4.1 Underfluctuations

In the case of underfluctuations the deviations between the likelihoods were quite large. As mentioned before, *SModelS* set the truncated Normal distribution in these cases to be located at 0, which is a poor fit in most cases.

Two separate approaches were taken to tackle this. The first and obvious was to allow the Gaussian to be located at negative values, maintaining the conditions in equations 2.3 and 2.5.

The second approach was to approximate the likelihood with an exponential distribution. The condition imposed on it was that it had to satisfy the 95%-CL, i.e.

$$\int_0^{\mu_{obs}^{95\%}} d\mu \quad \lambda e^{-\lambda\mu} = 0.95 \tag{4.1}$$

had to be fulfilled.

In figure 4.1 both modifications have been plotted. It is clear that for this case the normal distribution located at a negative value is the better fit. There are however also cases, as shown in figure 4.2, where the exponential distribution is a better fit. It turns out, that these cases are the cases where $N_{obs} = 0$. This makes sense when looking at equation 2.7 for the simplified likelihood. The Poisson-part of it reduces to just an exponential part if the number of observed events is 0.

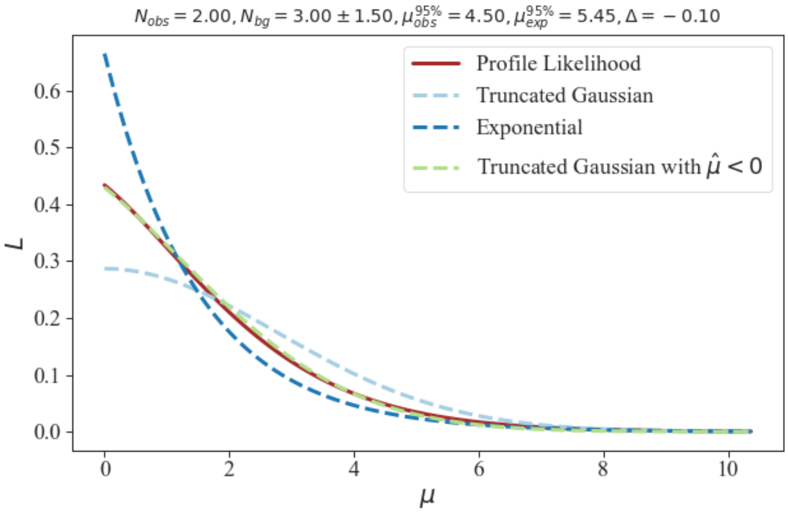


Figure 4.1: The Likelihood of figure 3.2 with added modified likelihoods.

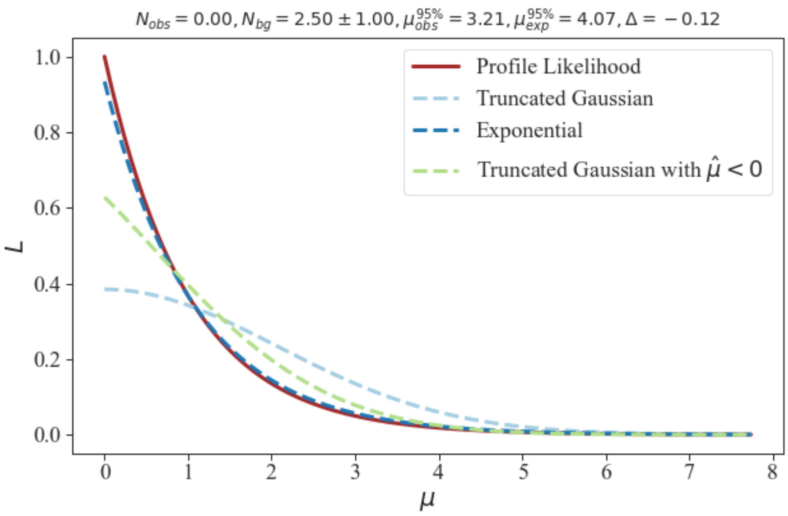


Figure 4.2: Likelihoods for a case with $N_{obs} = 0$.

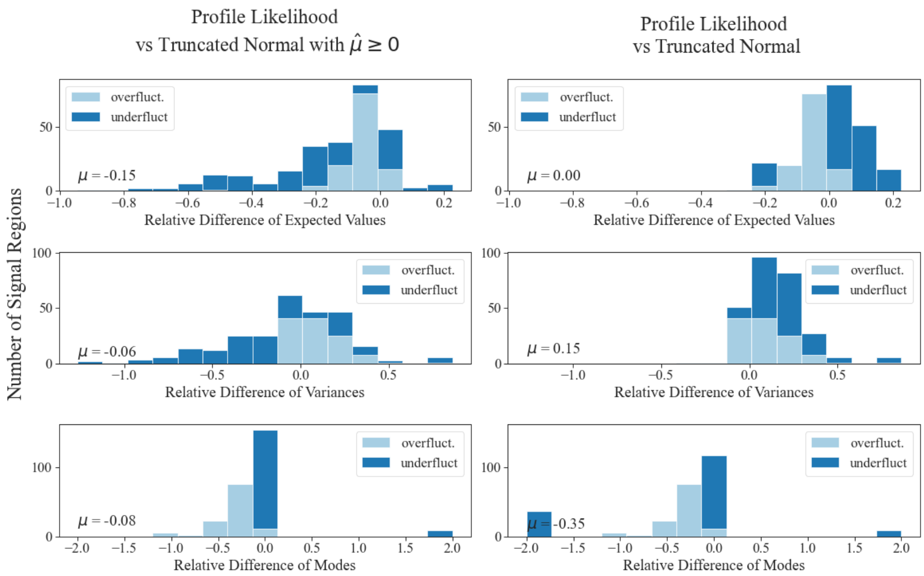


Figure 4.3: Histogram showing the effects of allowing $\hat{\mu} < 0$. Shown are again the relative differences, the left histograms correspond to the unmodified ones in figure 3.3.

It might seem tempting at first to introduce a case differentiation to approximate such likelihoods with an exponential distribution, however this is not possible, because $N_{obs} = 0$ cannot be extrapolated from the upper limits (as is evident for example from figure 3.4). But the upper limits are the only information that is available to compute the likelihoods. And since these cases are a small minority the truncated Gaussian remains the better approximation overall.

In figure 4.3 the effects of allowing negative values of $\hat{\mu}$ can be seen. The expected values for underfluctuations, largely overestimated by the original routine, get mostly underestimated. The ones where the expected value still gets overestimated correspond to the $N_{obs} = 0$ results. This can be seen in figure 4.4 and made plausible by looking at figure 4.2, where it can be seen that the full likelihood, showing exponential behavior, falls a lot quicker than the Gaussian one.

The variances show similar behavior. However, as opposed to the expected values, underestimating is not conservative in this case.

The modes should in theory not be affected by this modification, since overfluctuations don't get modified and underfluctuations should retain 0 as their mode. Nonetheless, a lot of results appear at $\Delta_{rel}^{mode} = -2$. These are all results that satisfy $\mu_{exp}^{95\%} = \mu_{obs}^{95\%}$ and hence fall under our definition of underfluctuation. This means that in the modified procedure, instead of setting $\hat{\mu} = 0$, the location gets computed numerically by solving equation 2.3. And although the resulting values all satisfy $|\hat{\mu}| \leq 10^{-3}$, the relative difference becomes exactly 2, since the mode of the full likelihood gets evaluated to 0. Hence, like the values at $\Delta_{rel}^{mode} = +2$, these can be regarded as artifacts.

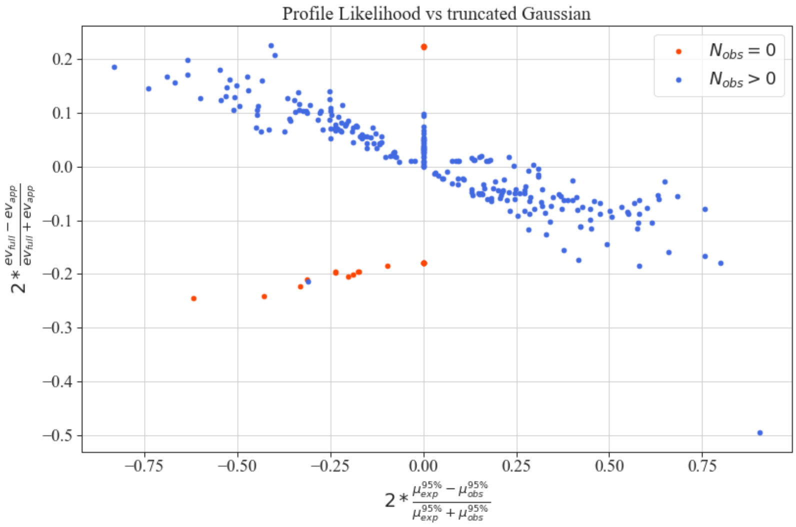


Figure 4.4: Scatterplot of the relative differences of expected values showing the effects of allowing Gaussians located at negative values as approximations.

Overall, just extending the approximation procedure to negative values of $\hat{\mu}$ seems to be a good modification for underfluctuations, since, except for the cases mentioned above, it reliably produces conservative estimations of the expected values. However, it has to be accepted in turn that the variances get underestimated.

4.2 Overfluctuations

For overfluctuations the case is less clear. In section 3.3 it has been seen that on average, the expected values and modes of the approximated likelihoods are too big, while the variances are small. A modification should, ideally, account for all three of those things. In said section it was also made evident that the difference in expected values and variances increases with increasing difference in upper limits. Hence, the strength of a suitable modification to the likelihoods will most likely depend on the relative difference of upper limits. Furthermore, it is important for a correction to be confined in its extent in order for it to never produce nonsensical corrections.

In principle the truncated normal distribution has two parameters that can be adjusted, $\hat{\mu}$ and σ_{exp}, which denote the location and standard deviation before truncation respectively. One might assume that shifting $\hat{\mu}$ changes only the location of the approximated likelihood and adjusting σ_{exp} changes only its width. However, since the conditions 2.3 and 2.5 have to hold, this is not the case.

The modification proposed here increases σ_{exp}. This produces all three effects we are looking for: the truncated Gaussian gets broadened, thereby increasing its variance. In order to still satisfy the 95%-upper limit, $\hat{\mu}$ has to

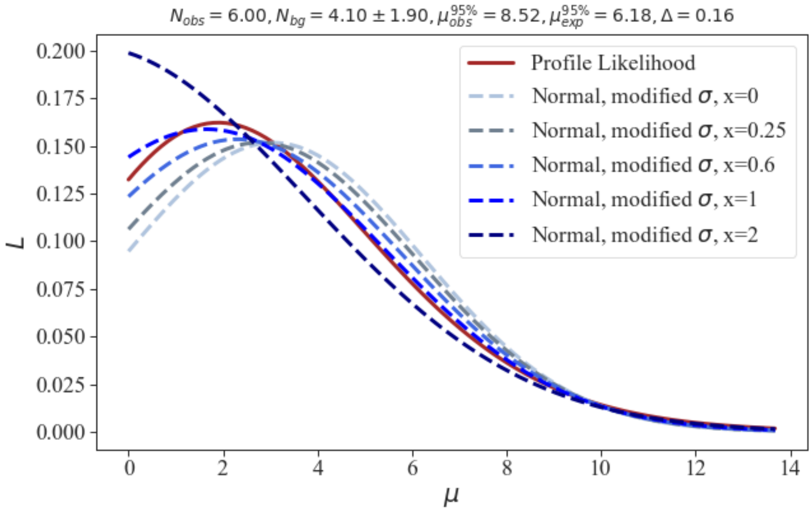


Figure 4.5: Showing the effects of the modification parameter x on the same likelihoods as in figure 3.1. With greater x, the approximated likelihoods get both broader and shifted to lower values of μ.

decrease, with the effect that the expected value as well as the mode decrease as well.

The proposed modification is

$$\sigma_{exp}^{mod} = \frac{\sigma_{exp}}{1 - x\frac{\mu_{obs}^{95\%} - \mu_{exp}^{95\%}}{\mu_{obs}^{95\%} + \mu_{exp}^{95\%}}} \tag{4.2}$$

where x is heuristically chosen and parametrises the strength of the modification. Since this is only applied to overfluctuations where $\mu_{obs}^{95\%} > \mu_{exp}^{95\%}$, the maximal modification is given by

$$\sigma_{exp}^{mod} \leq \sigma_{exp}\frac{1}{1-x} \tag{4.3}$$

If $0 < x < 1$, σ_{exp}^{mod} is always finite and positive.

For $x = 2$, the denominator of equation 4.2 corresponds to $1 - \Delta_{rel}^{UL}$, where Δ_{rel}^{UL} is the difference of the upper limits divided by their arithmetic mean. This is also the quantity on the x-axis of the plot in figure 4.3. As can be seen there, $\Delta_{rel}^{UL} < 1$ for all results in the database used here. Therefore, for the sake of demonstrating the influence of x, several plots have been produced with x ranging from 0 to 2.

In figure 4.5 the effects of the modification can be seen. As anticipated, higher values of the parameter x shift the approximated likelihood to the left while broadening it at the same time. While this plot gives an intuition how the modification manifests itself, it does not suffice to choose an adept value for x. In order to do that, all observations have to be taken into account.

Figures 4.6-4.8 show the effect of the modification. It behaves as expected: for rising x the expected value and mode both fall, while the variance rises. If

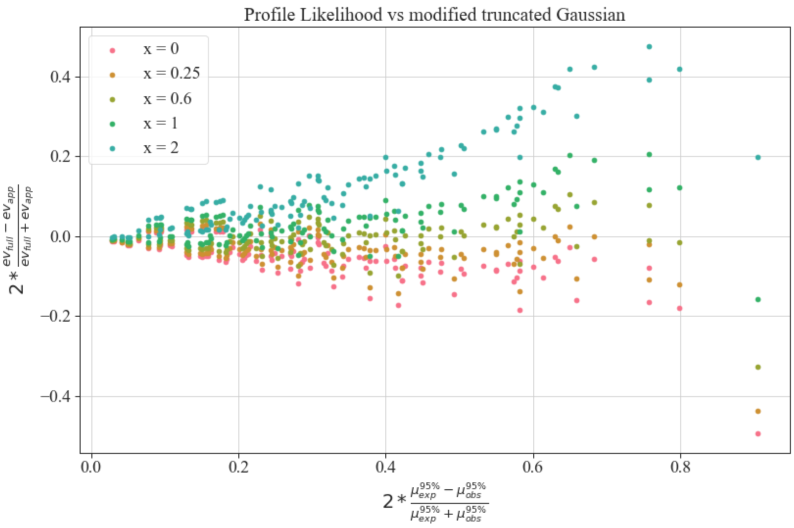


Figure 4.6: Plot showing the overall effects of different modification parameters on the relative difference of expectation values between full and approximated likelihoods. Note that only overfluctuations are pictured, since only they get modified.

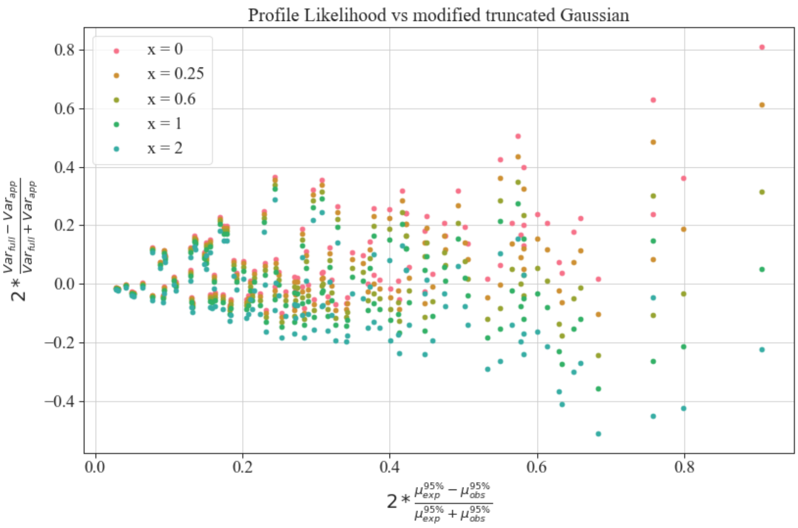


Figure 4.7: Plot showing the effects of different modification parameters x on the variance.

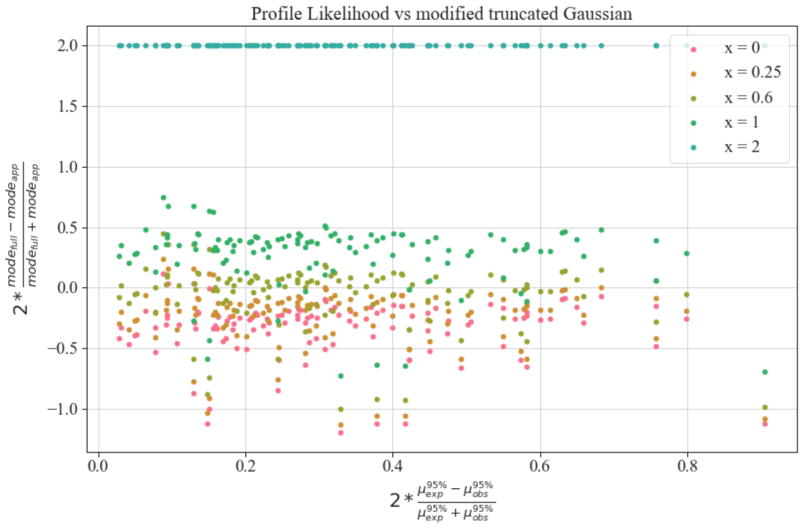


Figure 4.8: Plot showing the effects of different modification parameters x on the mode.

the modification gets to big, the mode of the approximated likelihood tends to zero. While this is not desirable, at least it is conservative.

Considering mainly the mode and expectation value, the best value for x has been estimated to be $x = 0.6$. This means, that the maximal correction for σ is with equation 4.3

$$\sigma_{exp}^{mod} \leq 2.5 * \sigma_{exp} \tag{4.4}$$

4.3 Comparing modified to unmodified likelihoods

Combining the modifications of the last two sections and comparing the modified likelihoods to the unmodified ones, the improvements are clearly visible.

In figure 4.9 it can be seen that the differences between full and modified approximate likelihood are almost 0 on average. The variances of overfluctuations get reduced, however, since the modified underfluctuations have increased variance compared to the unmodified ones, the mean variance of the approximated likelihoods rose with the modifications. The modes on the other hand coincide very well (save the artifacts at ± 2 explained above).

Overall, the proposed modification greatly improves the agreement between the two likelihoods, making the approximated ones more conservative in the case of searching for new physics.

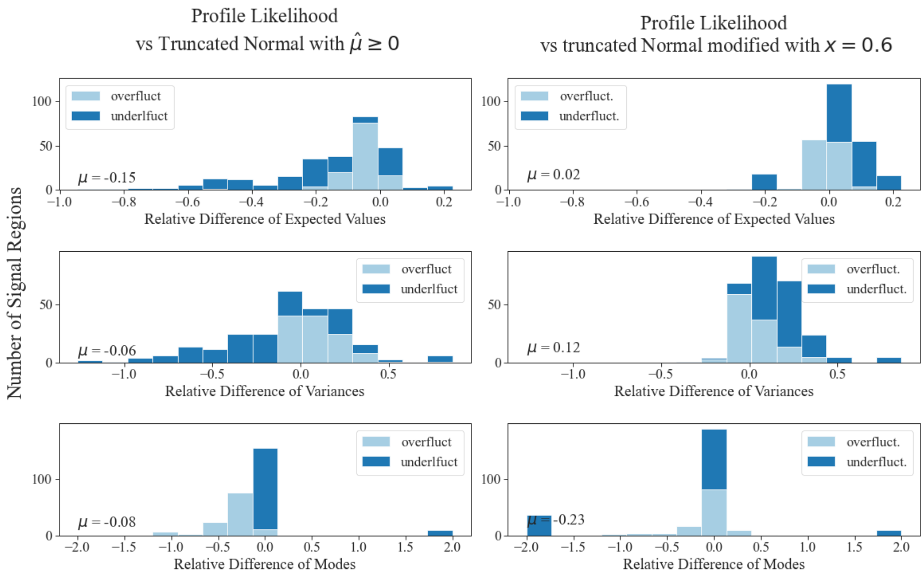


Figure 4.9: Histograms of relative differences as defined in equation 3.1 (negative values mean the respective value of the truncated Gaussian likelihood is higher). On the left are differences for the unmodified approximated likelihoods, on the right differences for the modified ones. Here μ denotes the mean of the values in the respective histograms.

References

[1] B. Abi et al. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021.

[2] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 18:1–98, 1998.

[3] CMS Collaboration. Interpretation of searches for supersymmetry with simplified models. *Phys. Rev. D*, 88:052017, Sep 2013.

[4] Sabine Kraml, Suchita Kulkarni, Ursula Laa, Andre Lessa, Wolfgang Magerl, Doris Proschofsky-Spindler, and Wolfgang Waltenberger. SModelS: a tool for interpreting simplified-model results from the LHC and its application to supersymmetry. *The European Physical Journal C*, 74(5):1–23, 2014.

[5] SModelS user manual. `https://smodels.readthedocs.io/en/latest/`. Accessed: 22-May-2021.

[6] Wolfgang Waltenberger, André Lessa, and Sabine Kraml. Artificial protomodelling: building precursors of a next standard model from simplified model results. *Journal of High Energy Physics*, 2021(3):1–41, 2021.

[7] Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods. Technical report, CERN, Geneva, Aug 2019.

[8] Aleksandr Azatov, Roberto Contino, and Jamison Galloway. Model-independent bounds on a light Higgs. *Journal of High Energy Physics*, 2012(4):127, 2012.

[9] CMS Collaboration. Simplified likelihood for the re-interpretation of public cms results. Technical report, CMS-NOTE-2017-001, 2017.

[10] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011.

[11] Jamol Pender. The truncated normal distribution: Applications to queues with impatient customers. *Operations Research Letters*, 43(1):40–45, 2015.