PROCEEDINGS OF SCIENCE

Confidence intervals for linear combinations of Poisson observations

Francisco Matorras [1]
Instituto de Física de Cantabria (IFCA), Universidad de Cantabria- CSIC
Santander, Spain
E-mail: `francisco.matorras@cern.ch`

Abstract:

Different situations in HEP data analysis involve the calculation of confidence intervals for quantities derived as linear combinations of observations that follow a Poisson law. Although apparently a simple problem, no precise methods exist when asymptotic approximations are not accurate. Existing procedures are reviewed, and new approaches are proposed. Their performance and range of validity is checked in different benchmarks. In general, the simple methods based on error propagation or application of Wilks theorem to MLE show important undercoverage or overcoverage for low number of counts. On the contrary, methods based in profiling the likelihood or projecting the multidimensional confidence regions obtained with the Neyman construction show a much better performance.

XIII Quark Confinement and the Hadron Spectrum - Confinement2018
31 July - 6 August 2018
Maynooth University, Ireland

[1]Speaker

1. Introduction

This work attempts to solve an apparently simple question arising in different fields, in particular in several situations in experimental particle physics, on how to set a confidence interval with good coverage properties for a linear combination of Poisson means.

Some practical examples commonly found in particle physics include:

- Negative weights. Calculation of errors on "counts" when using simulations based on generators which include events with negative weights [1]. In this case, some events are assigned with a negative weight of -1, being the "counts" for a particular range of the kinematics calculated as the difference between the positive and negative counts in that range. Rising question like what to do with the case a negative number for a quantity one would expect to be positive or how to set an error on the case when no counts are observed. This case corresponds to the difference of two Poisson observations.
- The so called "subdominant background". In its simplest version, appears when calculating the expected rates for given analysis, where there are several contributions, one for a channel with no observed events and a large coefficient (i.e. you generated few events because you do not fear that background). For example, if your expectation is given by Nexp=μ1+100 μ2, observing $\vec{n}$=(100,0), how large is the error? This is obviously also a linear combination of Poisson, where one of the coefficients is much larger than the other.
- Background subtraction. Often one gets an estimate of a given signal as the subtraction of two observations (both following a Poisson distribution) with maybe some scaling factor, which turns to be the same problem.

Despite the apparent simplicity, it will be shown that none of the commonly used methods provide good statistical properties when the number of involved counts is small. Several alternatives will be proposed and reviewed.

1.1 Posing the problem

In this work we will address the problem to define a confidence interval of a quantity $\mu' = \sum \beta_i \, \mu_i$, a general linear combination of the means, $\vec{\mu}$, of a set of independent Poisson distributions given a set of observations $\vec{n}$. For each of these observations we expect the probabilities $\mathcal{P}(n_i|m_i)$, to follow a Poisson law. Let me stress that we have one and only one observation for each of the different means.

We will be restricting to the case of only two Poisson distributions. All the described methods can be extended to more dimensions, although for some of them the computation time might become an issue for higher dimensions. All the discussion and examples will be based on giving a central confidence interval at 68.3%, which is the usual definition for an experimental error, but its application to other CL, upper or lower limits is straightforward.

The goal is finding methods providing good frequentist coverage, with the shortest possible interval. We favor methods that provide approximate coverage, permitting moderate undercoverage, with respect to methods guaranteeing coverage above 68.3% but showing in practice large overcoverage. This in general can be tuned in the different models.

Without loss of generality, we can write the linear combination as $\mu' = \beta_1 \, \mu_1 + \beta_2 \, \mu_2 = K(\cos(\theta) \; \mu_1 + \sin(\theta) \; \mu_2)$. The scale K can be ignored in the following calculations, since any

CI obtained with K=1 can just be scaled back multiplying by K. We can then treat the general case with just one parameter that represents a rotation in the plane of the possible values for $\vec{\mu}$. Furthermore, we can build a full rotation matrix R and a full rotated set of means as $\overrightarrow{\mu'} = R\,\vec{\mu}$, the first coordinate in $\overrightarrow{\mu'}$ being the linear combination we are interested in, our parameter of interest. The second represents our nuisance.

1.2 Benchmarks

The performance of the methods is studied on different benchmarks.

In first place the case of negative weights, which correspond to θ=–π/4, is explored for a range of means between 0 and 10, where non-asymptotics effects are important. A special focus is set on the case of subtraction of two null observations $\vec{n} = (0{,}0)$. Checks were also performed on the regime of large means as a comparison with asymptotic predictions.

The second benchmark, subdominant background, corresponds to θ→0.
The case when one of the means is large ~100 and scaled with a small angle and the other is close to 0, will be studied.

Consistency checks were also performed with θ=0, and θ=π/4 (corresponding to the sum of two Poisson observations). Both cases correspond to a pure Poisson law and are confronted with the exact Poisson CI. The case of θ=π/4, is also used for checks, since it corresponds to the sum of two Poisson which follows a Poisson law too.

2. Methods to estimate the confidence intervals

2.1 Error propagation

Many physicists would start by solving this problem with the naïve approach of the so-called "error propagation" of gaussian errors. This method implies two approximations, replace the dependence on the variables by the first order Taylor expansion (which is exact in our linear case) and assume each of the variables follows a Gaussian distribution with $\sigma = \sqrt{n}$, which is not correct unless the number of counts, n, is large. As it will be shown later, this approach significantly undercovers if it is not the case. In the limit of no observations, the predicted interval will have zero-length.

Often this is overcome with the error propagation of the "Garwood intervals" [2], central CI for a Poisson distribution, but it will we shown that on the contrary, this method tends to significantly overcover even for moderate number of counts. As an example, it will predict an interval of ±2.6 for the negative weights case and no observations, while other methods described below provide good coverage for much shorter intervals of ±1.5 or ±2

2.2 Successive approximations

An alternative method, proposed in [3] is based on successive approximations starting from a central limit theorem for the studentized statistic, $Y = \frac{\mu\prime - K_1}{\sqrt{K_2}}$, where K_1 and K_2 are the first two moments in $K_l = \sum_i \beta_i^l n_i$, corresponding to the mean and variance of the observed counts.

Approximate CI are expressed on relatively simple algebraic expressions calculated from these moments.

This method is appealing because it has a simple implementation, even for a high dimensionality and provides a faster convergence to asymptotics with good approximate coverage even for counts ~ 1. However, it still predicts 0-length intervals for $\vec{n}$ = (0,0) and in general undercovers when any of the observations is zero.

In this work, a modification is proposed to overcome this problem. Moments are calculated replacing the cases of 0-counts with one, n→max(n,1), in the calculation. The resulting CI must be corrected for the induced off-set on the mean whenever a 0 is replaced by 1. In this way, a non-zero-length interval centered in the observation is obtained, with rather good empirical coverage properties.

2.3 "MINOS errors"

Another common approach, usually known as MINOS error [4] due to its implementation in the widely used MINUIT program [5], proposes a calculation as follows. One first performs a Maximum Likelihood Estimation, MLE, minimizing the likelihood as a function of the parameter of interest (POI) and nuisances. Then, contours are drawn corresponding to the parameters (both of interest and nuisances) whose log-likelihood varies by ½ from the minimum. The CI on the POI is the projection of this contour onto the axis corresponding to the POI, i.e., the set of all possible values of the POI along that contour.

This is equivalent to the likelihood ratio construction and the asymptotic properties (Wilks theorem [6]) guarantee that for a sufficiently large number of observations it must converge to a χ^2 with ndof equal to the number of parameters. Under these conditions and with two variables, the drawn regions should correspond to a 39.4% interval and its projection into one axis to a 68.3%.

The interpretation as a rotation is convenient in this case because the (orthogonal) change of variables $\overrightarrow{\mu'} = R\,\vec{\mu}$, preserves the MLE properties. Hence one can build the likelihood function in the simplest case, as the product of that of two independent Poisson, calculate the contours on those variables and then project according to a rotated axis with any given angle θ to obtain our desired CI for the variable of interest. This is illustrated in Fig 1, where different contours at shown for the case of observing $\vec{n} = (1{,}1)$. The projection of the 39.4% contours for different rotations is represented by the red arrows. Fig 2 shows other examples of the contours.
This is a relatively simple method, that by construction reproduces the asymptotic expectations, but it will be shown that undercovers for small number of counts.

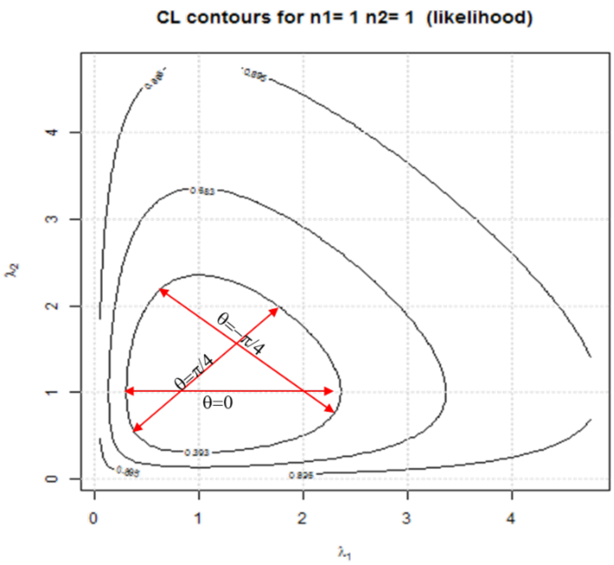

Figure 1: Example of contours for fixed values of the likelihood function for $\vec{n} = (1,1)$, transformed to the corresponding $\chi 2$ tail probabilities for ndof=2. The red arrows show the projected 68.3% CI for different examples of rotations.

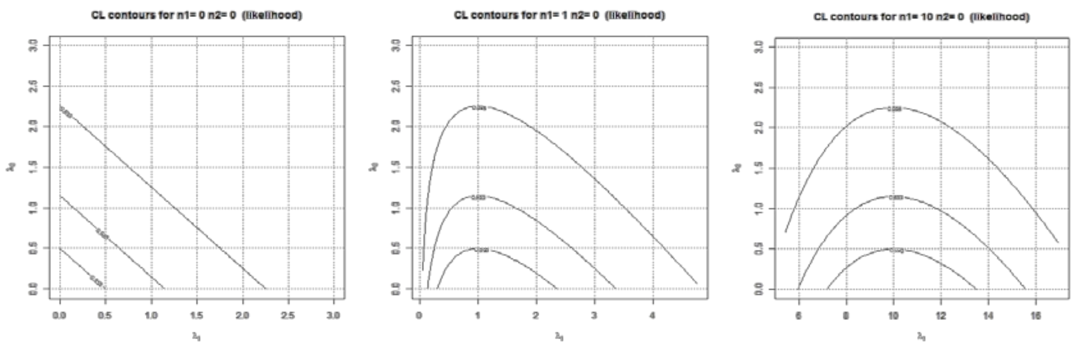

Figure 2: Example of contours for fixed values of the likelihood function for $\vec{n} = (0,0)$, $(1,0)$ and $(1,1)$, transformed to the corresponding $\chi 2$ tail probabilities for ndof=2

2.4 Projection of 2D confidence regions

A new approach is proposed, based on the calculation of the 2D confidence regions using the Neyman construction [7] based on test inversion for the product of two independent Poisson distributions and following a similar approach as described above to project these into a rotated axis corresponding to the concrete linear combination. The calculation would be as follows:

- For a given observation, $\vec{n}$, scan the plane of the possible values for the (non-rotated) Poisson means $\vec{\mu}$
- For each point in this plane, sort/rank the possible $\vec{m}$ according to your preferred ordering rule (in this work probability ordered and Feldman-Cousins were tested)
- For each $\vec{\mu}$ calculate a "confidence level" for $\vec{n}$ from the discrete sum of the Poisson probabilities $\mathcal{P}(m1|\mu1)\ \mathcal{P}(m2|\mu2)$, starting from the highest ranked $\vec{m}$ until $\vec{m} = \vec{n}$, being $\mathcal{P}$ the Poisson probabilities.

- Draw the desired contours separating the regions below and above a given CL, as shown in Figs. 3 and 4.

To obtain the desired CI on the linear combination we need to project these contours using a rotated axis as was done in the previous case. Project the 68.3% CL contours would guarantee coverage of at least 68.3% in the projected interval but will largely overcover in most cases. In this work it is chosen to project the 39.4% instead. This will provide a 68.3% coverage asymptotically. For smaller number of counts, it will be shown the projection implies a moderate undercoverage, that is partially compensated by the overcoverage produced by the discrete nature of the Poisson problem, leading to an acceptable frequentist coverage in most cases. As can be seen in the comparison of Figures 3 and 4, both ordering methods produce rather different contours at low number of counts. This is mostly due to the fact of Feldman-Cousins being more aggressive (produces less overcoverage) especially for the treatment of the case of $\vec{n} = (0,0)$. This point in the plane may have a large probability, but a low likelihood ratio. Therefore, it may be ranked on the top or not depending on the method, thus changing drastically the confidence regions. It is worth noting that the undercoverage appears when projecting, by definition the 2D confidence regions (with one or other ordering method) guarantee the coverage.

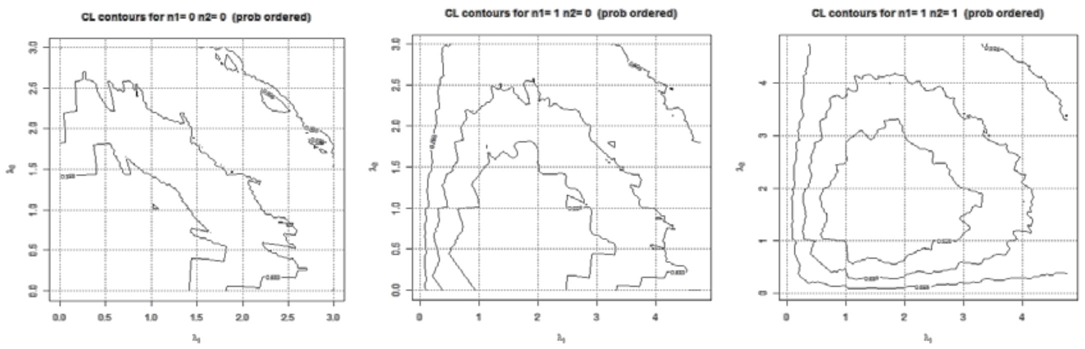

Figure 3: 2D confidence regions at 39.4%, 68.3% and 89.6% obtained with the full Neyman construction at different CL for for $\vec{n} = (0,0),\ (1,0)\ and\ (1,1)$ *, using probability ordering*

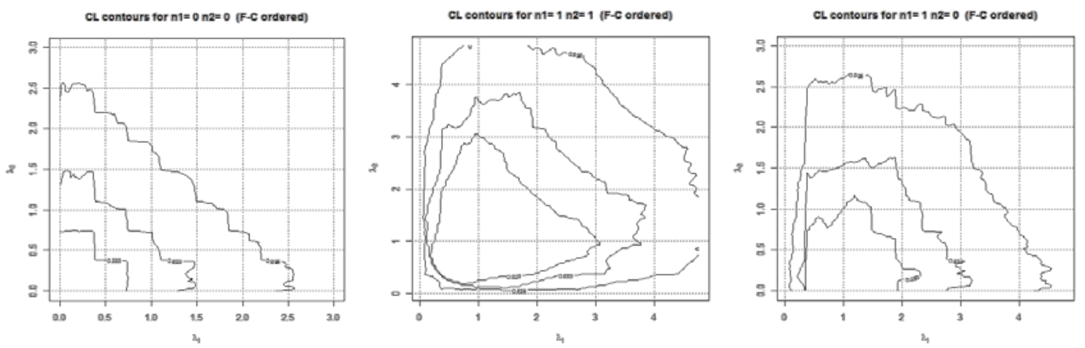

Figure 4: 2D confidence regions at 39.4%, 68.3% and 89.6% obtained with the full Neyman construction at different CL for for $\vec{n} = (0,0),\ (1,0)\ and\ (1,1)$*, using Feldman-Cousins ordering*

2.5 Profile-based methods

Increasing the complexity, one could use a profile-based method, in which the first component of $\overrightarrow{\mu'}$ is treated as POI and the rest as nuisances. There are different alternatives to implement these methods and we have chosen to follow the ideas proposed in [8]: "*a full Neyman construction over both the parameters of interest and the nuisance parameters, using the profile likelihood ratio as an ordering rule*". In practice the same procedure as above is followed, replacing the ordering rule by the profiled likelihood ratio, PLR, where the profiling is performed on the nuisance, in our case the linear combination orthogonal to that of our interest. It is worth noting that the PLR, by definition, does not depend on the nuisance, but there is still a mild dependence on it of the 2D contours as shown in Fig. 5, because the probabilities still have a dependence. Thus, we need to again project onto the rotated axis, although in this case we have only one dof left and hence will project the 68.3% contour. Note also that since the profiling depends on the actual coefficients of the linear combination the 2D contours are not independent of the rotation angle any longer.

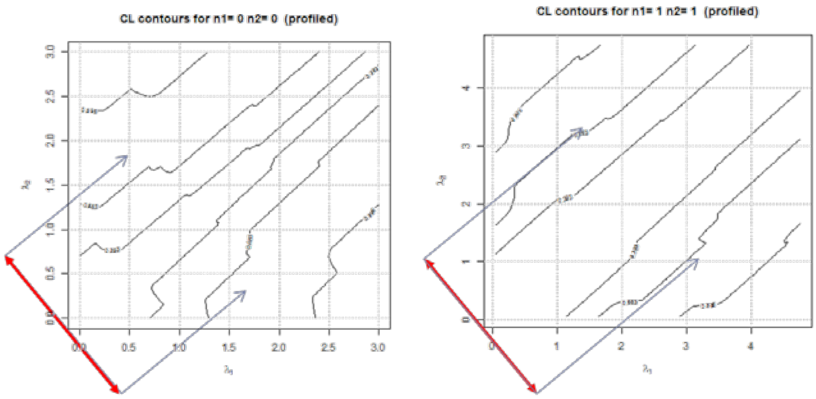

Figure 5:2D confidence regions at 39.4%, 68.3% and 89.6% obtained with the full Neyman construction at different CL for for $\vec{n} = (0,0),\ (1,0)\ and\ (1,1)$, *with an ordering based on the profiled likelihood ratio for* $\theta = -\pi/4$,. *The red arrows show the approximate projected CI.*

2.6 Additional methods

Some other methods that were explored did not provide good results but are outlined here for completeness.

In principle one could try to integrate the nuisance out of the pdf, but no well-motivated priors where found except flat distribution in a bounded box, whose results turn to be dependent on the assumptions made to bound this box.

It was also attempted to take advantage of the "near independence" of the POI and nuisance, trying to marginalize the pdf of the POI for any possible value of the nuisance. Fig. 6 illustrates the situation for two examples when one follows a similar procedure as in the previous cases, with the pdf obtained in that way. One can observe the mild by divergent dependence remaining, which requires to use some alternative method (i.e. Garwood error propagation) to define the bounds, with similar problems as in the previous case.

3.Results

3.1 Confidence Intervals

Figure 6 and Table 1 show some examples of the CI obtained by the different approaches described for different observations in the benchmark I (subtraction of two Poisson). We can appreciate that for moderate number of counts all methods give similar intervals except for “Garwood intervals propagation” and “marginalization”, these giving significantly wider intervals. On the contrary, for low number of counts especially for (0,0), the results are very different, with the mentioned case of the zero-length intervals predicted in some cases. Similarly, Fig 7 shows the large spread of predictions of the different method for the benchmark II.

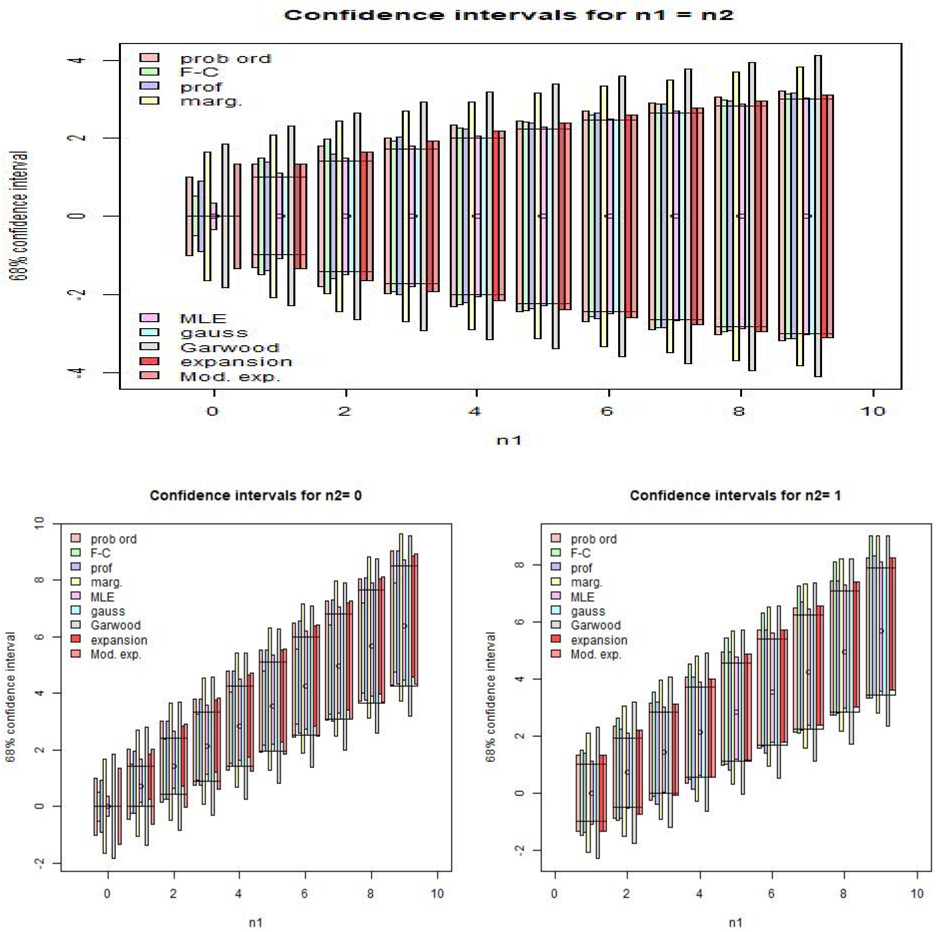

Figure 6: Confidence intervals obtained with the different methods described in the text for benchmark I (negative weights), when the number of positive and negative counts are the same (top), the number of negative counts is fixed (rest of the plots). The black box shows the plain error propagation for comparison.

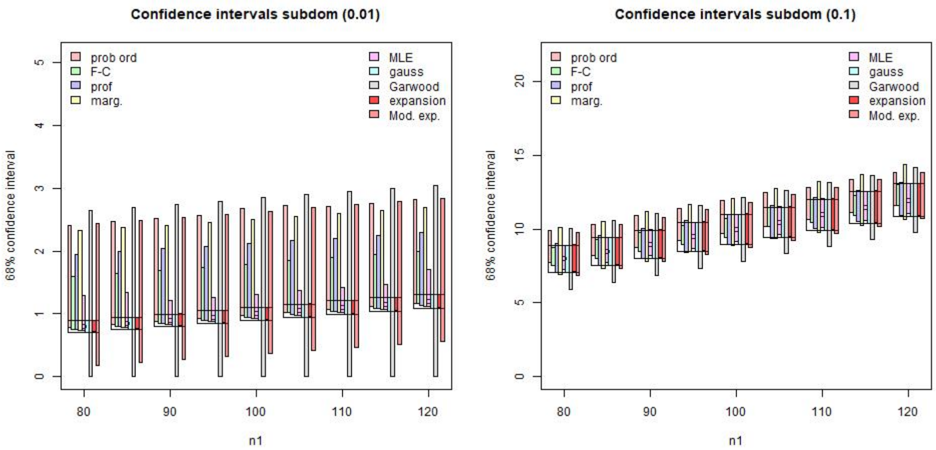

Figure 7: Confidence intervals obtained with the different methods described in the text for benchmark II (subdominant background), for an angle of θ=0.01 (left) and θ=0.1 (right). In all cases n_2=0. The black box shows the plain error propagation for comparison.

Table 1 Confidence intervals according to the different methods for the benchmark I and $\vec{n} = (0{,}0)$, $(1{,}0)$ *and* $(1{,}1)$.

METHOD | $\vec{n}$=	**(0,0)**	**(1,0)**	**(1,1)**
Error prop.	[0.00 0.00]	[0.00 2.00]	[-1.41 1.41]
Garwood prop.	[-2.60 2.60]	[-1.95 3.95]	[-3.25 3.25]
Succ. Approx.	[0.00 0.00]	[0.38 2.63]	[-1.89 1.89]
Mod Approx.	[-1.89 1.89]	[-0.89 2.89]	[-1.89 1.89]
MINOS	[-0.50 0.50]	[0.19 2.36]	[-1.56 1.56]
2D prob-ord	[-1.43 1.43]	[-0.64 2.86]	[-1.89 1.89]
2D FC	[-0.72 0.72]	[-0.35 2.13]	[-2.12 2.12]
2D profile	[-1.29 1.29]	[-0.35 2.75]	[-1.98 1.98]
Marginalized	[-2.34 2.34]	[-1.51 3.82]	[-2.96 2.96]

3.2 Coverage tests

To evaluate the performance of the different approaches, the frequentist coverage is evaluated as a function of the means of the two Poisson observation for the different scenarios. The frequentist coverage for a given $\vec{\mu}$ and θ is calculated as follows. A probability P($\vec{m}$|$\vec{\mu}$) is defined as a product of the two independent Poisson laws, for all $\vec{m}$ with a probability above a given threshold (a cut off such that a total probability close to 100% is achieved with a finite number of observations) a CI is calculated for the given rotation. The coverage is the sum of the probabilities of the $\vec{m}$, such that the CI includes the rotation of the mean μ´. Note, that given the discrete probabilities no toy MC is involved. Figure 8 summarizes the results for benchmark I, the different colors representing the coverage. Our goal is to have orange in its different shades, showing good coverage or moderate over/undercoverage, while regions both with severe over and undercoverage should be avoided (red, yellow, white).

We can draw some conclusions. Gaussian error propagation is naïve, but the coverage is acceptable if both means are above 2 or 3. Garwood error propagation largely overcovers and it is not recommended. MINOS approach improves the coverage, w.r.t. the error propagation, but still shows some severe undercoverage regions. The successive approximation method improves the coverage at intermediate means but also does not perform well for small means. This is cured with the proposed modification of this method. The different 2D approaches provide better coverage properties (especially for profile based) if more accurate CI needed.

Equivalent studies performed on benchmark II lead to similar conclusions.

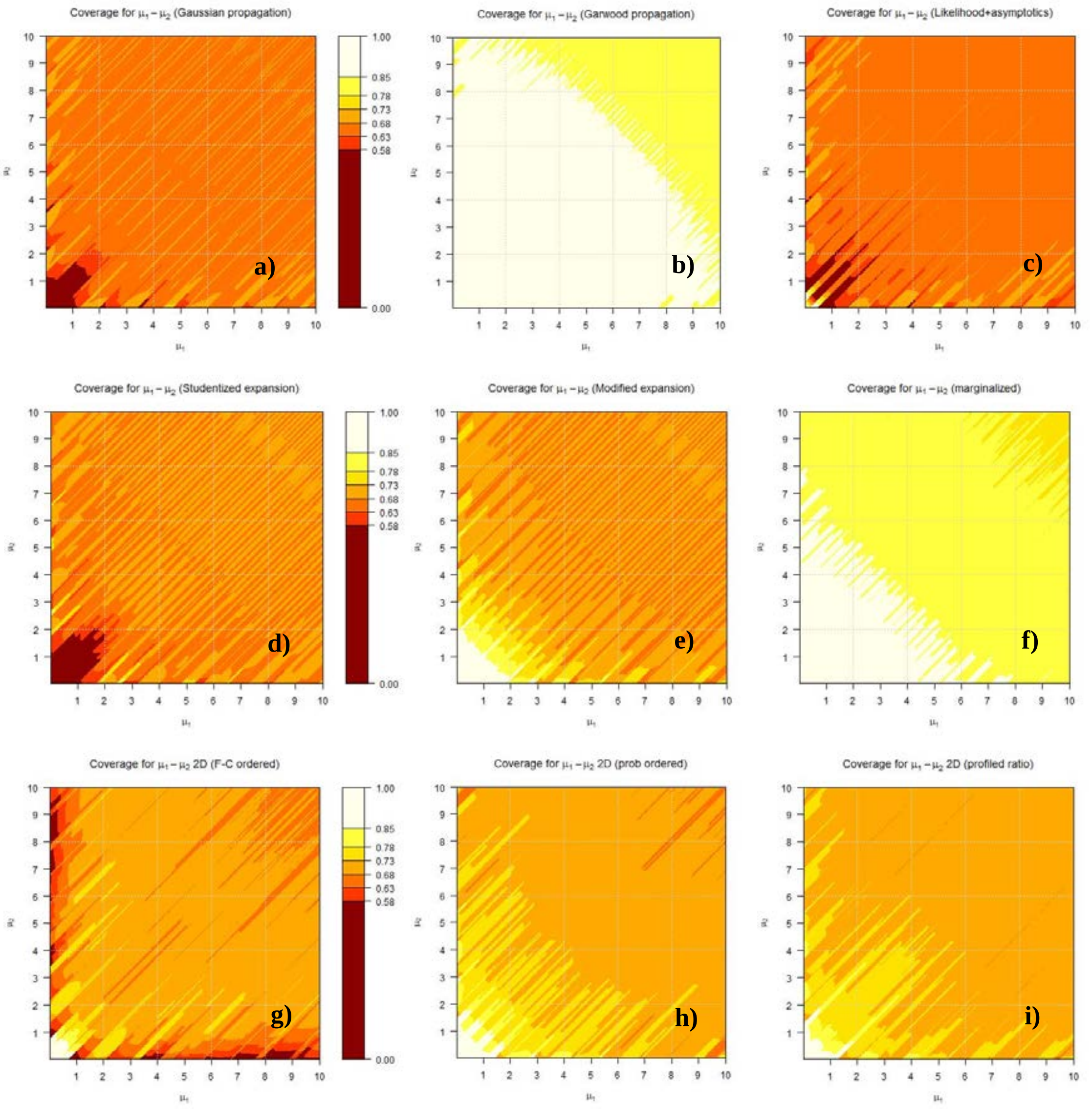

Figure 8: Coverage tests for benchmark I for a) Gaussian error propagation, b) Garwood propagation c) MINOS, d) successive approximations, e) modified successive approximation, f) pseudomarginalization, g-i) 2D projection for Feldman-Cousins, probability and PLR ordering

4. Summary and conclusions

Several situations in Particle Physics data analysis involve the estimation of a confidence interval for a quantity that is derived from a linear combination of others that follow a Poisson distribution. In this work some of the usual approaches to solve the problem are revised,

demonstrating that their performance is non-optimal. Different methods based on the profiling or projection of 2-dimensional Neyman-constructed confidence regions are proposed.
Their performance was evaluated calculating the frequentist coverage for a wide range of parameters. Although none of the methods is perfect in all situations some general conclusions can be drawn. The simple error propagation of “gaussianized” errors, provides reasonable coverage unless the means of the Poisson distributions are at or below 2-3. The common extension of propagate the Garwood intervals is not recommended, since it leads to sizeable overcoverage even for moderate means. The proposed method based on successive approximations has a simple implementation, even for multiple Poisson distributions, and improves the coverage, but still fails to provide good coverage for small means. A promising modification is proposed that cures this problem, providing a good coverage throughout the whole parameters space. However, it does not rely on well-motivated statistics justification, being just empyrical. The MINOS approach also fails to provide good coverage for means at or below 1.

If more accurate CI are needed, three different method are proposed based on the Neyman-constructed 2D confidence regions, which are projected onto a rotated axis whose angle represents the mixing of the two Poisson counts. Three ordering rules where explored, Feldman-cousins, probability ordered and profile likelihood ratio, based. These provide much better coverage properties, the first two additionally permit an interesting visual interpretation. As a counterpart, these are more complex to implement, for example for simple situations like setting an error bar on a histogram. For these cases look-up tables can be provided, since only a finite number of estimations are needed, since the problem is discrete and simple methods can be used once the number of counts is far from zero.

References

[1] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower simulations", JHEP 0206 (2002) 029

[2] F. Garwood “Fiducial Limits for the Poisson Distribution”, Biometrika, Volume 28, Issue 3-4, 1 December *1936*, Pages 437–442, doi: 10.1093/biomet/28.3-4.437

[3] C. S. Withers and S. Nadarajah, “Confidence intervals for linear combinations of Poisson means” *Aust. N. Z. J. Stat. 56(1), 2014, 47*–58 doi:10.1111/anzs.12055

[4] W. T. Eadie, D. Drijard, F. James, M. Roos, and B. Sadoulet. “Statistical Methods in Experimental Physics”. North-Holland, 1971

[5] F. James, M. Roos (CERN), “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations” Comput. Phys. Commun. 10 (1975) 343-367 doi: 10.1016/0010-4655(75)90039-9

[6] Wilks, S. S. (1938). "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". The Annals of Mathematical Statistics. 9: 60–62. doi:10.1214/aoms/1177732360.

[7] J. Neyman, Phil. Trans. Royal Soc. London, Series A, 236 333-80 (1937).

[8] K. Cranmer “Statistical Challenges for Searches for New Physics at the LHC”, (Proc. of Phystat2005) arXiv:physics/0511028 [physics.data-an] 4 Jan 2006 doi:10.1142/9781860948985_0026