Comments to the paper draft
"Measurement of $\sigma(e^+e^- \rightarrow HZ) \times Br(H \rightarrow ZZ^*)$ at the 250 GeV ILC" by E. Antonov and A. Drutskoy

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

ILD Analysis/Software Meeting

July 14, 2021

Reviewer comments

A lot of issues have been already clarified, manuscript has been significantly improved. However, one issue remains, which should still be corrected in my opinion. It refers to the treatment of statistical fluctuations in the event distributions.

Key points:

- rounding of the expected event numbers in the analysis procedure
- not including modeling of the Poisson fluctuations in the "main" fit and its presentation in the paper

Rounding problem

At the begining of section IV authors write (from line 197):

"Then, the signal statistical uncertainties corresponding to the integrated luminosity of 2 ab^{-1} are estimated. For that, the weighted signal and background distributions are summed, the content of each bin is rounded to the integer number and the Poisson uncertainties for the bin contents are assumed to imitate the real data."

I consider the "rounding procedure" described here as "illegal", being against the "good practice" rules. What one get from the generated MC samples are the distributions of the EXPECTED numbers of events and there is no reason to round them to the integer numbers.

Rounding problem

There are two possible ways to address the problem:

- evaluate the expected statistical precision of the normalisation fit from the expected numbers of events
  can be done in a semi-analytical approach
- generate Poisson fluctuations in each bin and repeate the fit procedure many times (so called Toy MC); the spread of the fit results can be considered as a reliable estimate of the statistical fit precision

Semi-analytical approach

Assume signal and background expectations are given by $s_i, b_i, i = 1 \ldots N$, and the number of observed events in each bin is n_i.

We want to find normalisation factors for signal (α) and background (β) defining the number of expected events in each bin:

$$\mu_i = \alpha \cdot s_i \; + \; \beta \cdot b_i$$

We do it by minimising the likelihood which is given by a product of Poissonian probabilities

$$\mathcal{L}(\alpha, \beta) \;\; = \;\; \prod_i \frac{\mu_i^{n_i} \; e^{-\mu_i}}{n!}$$

or minimising the log-likelihood

$$L(\alpha, \beta) \; \equiv \; \log \mathcal{L}(\alpha, \beta) \;\; = \;\; \sum_i n_i \log \mu_i - \mu_i - \log n!$$

Semi-analytical approach

Finding the normalisation is equivalent to solving a set of equations:

$$\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial \beta} = 0$$

We can also calculate second derivatives directly. Taking into account that $\langle \alpha \rangle = \langle \beta \rangle = 1$ and $\langle n_i \rangle = s_i + b_i$ the formula simplifies to:

$$-\frac{\partial^2 L}{\partial \alpha^2} = \sum_i \frac{s_i^2}{s_i + b_i} \equiv S$$

$$-\frac{\partial^2 L}{\partial \beta^2} = \sum_i \frac{b_i^2}{s_i + b_i} \equiv B$$

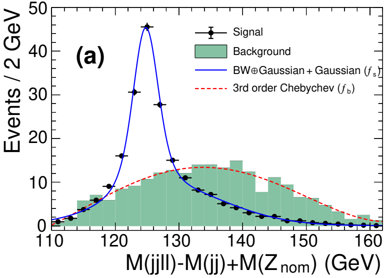
$$-\frac{\partial^2 L}{\partial \alpha \partial \beta} = \sum_i \frac{s_i b_i}{s_i + b_i} \equiv C$$

Semi-analytical approach

And the expected uncertainty on signal normalisation is

$$\sigma_\alpha = \sqrt{\frac{B}{S \cdot B - C^2}}$$

The approach was tested on the data from Fig. 1 of the paper.



Semi-analytical approach

And the expected uncertainty on signal normalisation is

$$\sigma_\alpha = \sqrt{\frac{B}{S \cdot B - C^2}}$$

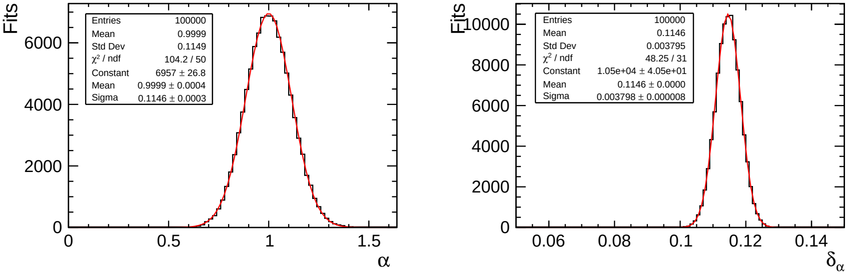
The approach was tested on the data from Fig. 1 of the paper.

Comparison of results:

Formula:	11.467%	No rounding involved !!!
My ToyMC:	11.477%	from RMS of fit results (100'000)
	11.459%	average uncertainty from the fit
Paper:	11.69%	fit with "rounding"
	11.96%	10'000 Toy MC

ToyMC results

Distribution of the fitted signal normalisation (α) and normalisation uncertainty from the fit (σ_α) as obtained from 100'000 MC experiments.

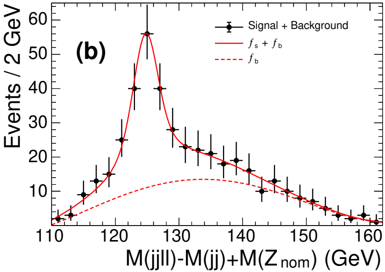


All results are very consistent...

No rounding of the expected event numbers involved !

Poisson fluctuations

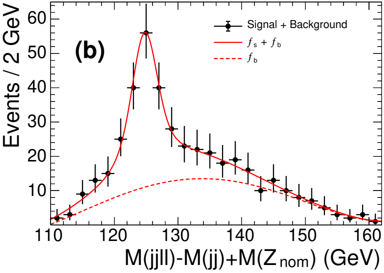
When the expected experimental results are shown on the plot the Poisson uncertainties indicated do not agree with the fluctuations of the data (which are much smaller, resulting from MC statistics).

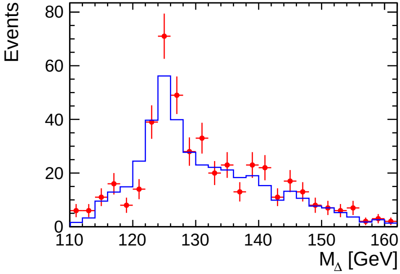


Reviewer comments

Poisson fluctuations

When the expected experimental results are shown on the plot the Poisson uncertainties indicated do not agree with the fluctuations of the data (which are much smaller, resulting from MC statistics).





It makes much more sense, in my opinion, to present one of the ToyMC data sets on the plot, demonstrating the possible impact of fluctuations...