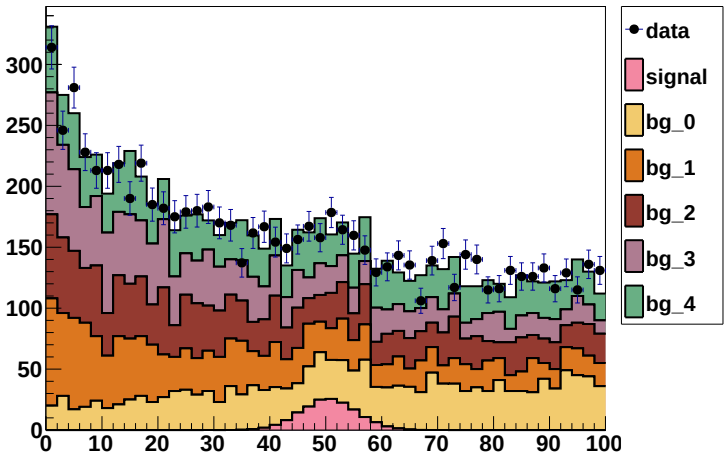
The evaluation of the systematic uncertainties for the finite MC samples in the presence of negative weights

Mandrik P.

NRC «Kurchatov Institute» – IHEP

QFTHEP'2017

Frequently the result of experiment can be represented as "data" distribution and a collection of distributions (templates) from Monte-Carlo simulation corresponding to the signals and backgrounds processes.



Parameters of some "theory" propagate through MC simulation of event reconstruction in detector.

Common way to connect theory with data and incorporate uncertainties is define a likelihood function:

$$\mathcal{L}(data|params, templates) = \prod_{b}^{bins} P(data_b|params, templates_b)$$

and then apply some Bayesian/frequentist methods to estimate parameters, set limits etc.

Negative weights

Events in MC templates with negative weights appeared:

- as a result of systematic correction
- from several Monte-Carlo event generators

While the first case did not change analyses in principle, for the second the following recommendations usually proposed:

- exclude events with negative weights (while they number is small)
- increase bin width to suppress negative events fraction

The motivation of this work is to investigate the correct definition of likelihood function in the presence of negative weights.

MadGraph5_aMC@NLO

Simplistically the cross section in MadGraph5_aMC@NLO[1]:

$$\sigma_{NLO} = \int F_H(x)dx + \int F_S(x)dx$$

where $F_H(x) + F_S(x) > 0$, but for some x: $F_S(x) < 0$.

Functions $|F_H(x)|$ and $|F_S(x)|$ are used to generate two set of events with the weights w equal to $+1$ and -1 (for $F_S(x) < 0$):

$$\sigma_{NLO} \approx \frac{\int |F_H(x)|dx}{N_H} \cdot \sum_i^{N_H} w_i^H + \frac{\int |F_S(x)|dx}{N_S} \cdot \sum_i^{N_S} w_i^S$$

where N_H,N_S - number of events.

[1] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," doi:10.1088/1126-6708/2002/06/029

$\Rightarrow$

- *By construction* and *in the infinite statistics limit*, the prediction of any observable can only be positive in any intervals $[x, x + \Delta x]$.
- Number of negative weights $\propto \int_x^{x+\Delta x} |F_S(x)| \mathcal{H}(-F_S(x)) dx$
- Events with negative and positive weights should be treated in the same way during analyses (same cuts, re-weighting etc).

From the statistical point of view the distribution of events with only positive weights usually described by multiplication of independent Poisson distributions (Poisson regime of multinomial distribution [1]). Distribution of events with only negative weights can be described this way too.

[1] for example C. Walck, "Hand-book on statistical distributions for experimentalists," SUF-PFY/96–01.

Notations

Data histogram:

$$\boldsymbol{X} = (X_1, X_2, ...)$$

MC histograms for process a:

$$t^a = (t_1^a, t_2^a, ...)$$

Parameters of model:

$$\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$$

Initial likelihood function:

$$\mathcal{L}(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{t}) = \prod_i P(X_i|\boldsymbol{\pi}, \boldsymbol{t}_i) = \prod_i P(X_i|\boldsymbol{\pi}, t_i^a, t_i^b, ...)$$

$\mathcal{P}$ - Poisson distribution, $\mathcal{G}$ - Gauss distributions.

Systematic uncertainties due to finite MC statistics

To take into account statistical fluctuations of MC samples Barlow-Beeston transformation [1] could be applied:

$$\prod_i P(X_i|\boldsymbol{\pi}, \boldsymbol{t}_i) \to \prod_i \left[P(X_i|\boldsymbol{\pi}, \boldsymbol{T}_i) \cdot \prod_k \mathcal{P}(t_i^k|T_i^k) \right]$$

where T_i^k - unknown parameter corresponding to the infinite statistics limit.

Because of large number of extra parameters often a "light" version of Barlow-Beeston method is used [2]:

$$\prod_i P(X_i|\boldsymbol{\pi}, \boldsymbol{t}_i) \to \prod_i \left[P(X_i|\boldsymbol{\pi}, \boldsymbol{T}_i) \cdot P(m_i|M_i) \right]$$

where $m_i = f(\boldsymbol{\pi}, \boldsymbol{t}_i)$, M_i - unknown effective parameter.

[1] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples," doi:10.1016/0010-4655(93)90005-W

[2] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra," doi:10.5170/CERN-2011-006.115

Systematic uncertainties due to finite MC statistics in the presence of negative weights

While events with negative and positive weights treated in the same way during analyses they distributions in single bin described as Poisson distributions $\Rightarrow$ the difference described as Skelleman distributions:

$$\mathcal{S}(t|T^+, T^-) = \sum_{s=max(0,t)}^{\infty} \mathcal{P}(s|T^+) \cdot \mathcal{P}(s-t|T^-)$$

And the following transformation could be applied:

$$\prod_i P(X_i|\boldsymbol{\pi}, \boldsymbol{t}_i) \to \prod_i \left[P(X_i|\boldsymbol{\pi}, \boldsymbol{T}_i) \cdot \prod_k \mathcal{S}(t_i^k|T_i^{k+}, T_i^{k-}) \right]$$

But Skelleman distributions is a weak constrain on $T_i^k = T_i^{k+} - T_i^{k-}$ due to loss of information of number of negative and positive events.

If the number of negative and positive events are available for templates (i.e $t_i^k = t_i^{k+} - t_i^{k-}$) the correct transformation will be:

$$\prod_i P(X_i|\boldsymbol{\pi}, \boldsymbol{t}_i) \to \prod_i \left[P(X_i|\boldsymbol{\pi}, \boldsymbol{T}_i) \cdot \prod_k \mathcal{P}(t_i^{k+}|T_i^{k+}) \cdot \mathcal{P}(t_i^{k-}|T_i^{k-}) \right] \tag{1}$$

and "light" version:

$$\prod_i P(X_i|m_i^+ - m_i^-) \to \prod_i P(X_i|M_i^+ - M_i^-) \cdot P(m_i^+|M_i^+) \cdot P(m_i^-|M_i^-) \tag{2}$$

where $m_i^{+(-)} = f(\boldsymbol{\pi}, \boldsymbol{t}_i^{+(-)})$, $M_i^{+(-)}$ - unknown effective parameter. $P(m_i^{+(-)}|M_i^{+(-)})$ depends on the form of f, may be Gauss, Poisson etc.

Analytic minimization

In statistical software often maximize likelihood (or minimize NLL) before the main calculations to exclude parameters from Barlow-Beeston method.

Based on minimum of likelihood function (1):

$$\begin{cases} -\dfrac{\partial \ln \mathcal{L}}{\partial T_i^{k+}} = 1 - \dfrac{\partial \ln P(X_i|\boldsymbol{\pi}, \boldsymbol{T}_i)}{\partial T_i^{k+}} - \dfrac{t_i^{k+}}{T_i^{k+}} = 0 \\ -\dfrac{\partial \ln \mathcal{L}}{\partial T_i^{k-}} = 1 - \dfrac{\partial \ln P(X_i|\boldsymbol{\pi}, \boldsymbol{T}_i)}{\partial T_i^{k-}} - \dfrac{t_i^{k-}}{T_i^{k-}} = 0 \end{cases}$$

Similar for (2).

Comparison for Bayesian analysis

Events sampled from:

$$\mathcal{L}_0 = \mathcal{P}(X|\pi \cdot (C + T^+ - T^-)) \cdot \mathcal{P}(t^+|T^+) \cdot \mathcal{P}(t^-|T^+)$$

Parameters estimated from:

$$\mathcal{L}_n = \mathcal{P}(X|\pi \cdot (C + t^+ - t^-))$$

$$\mathcal{L}_g = \mathcal{P}(X|\pi \cdot (C + T)) \cdot \mathcal{G}(t^+ - t^-|T, \sqrt{t^+ + t^-}) \cdot \mathcal{H}(T)$$

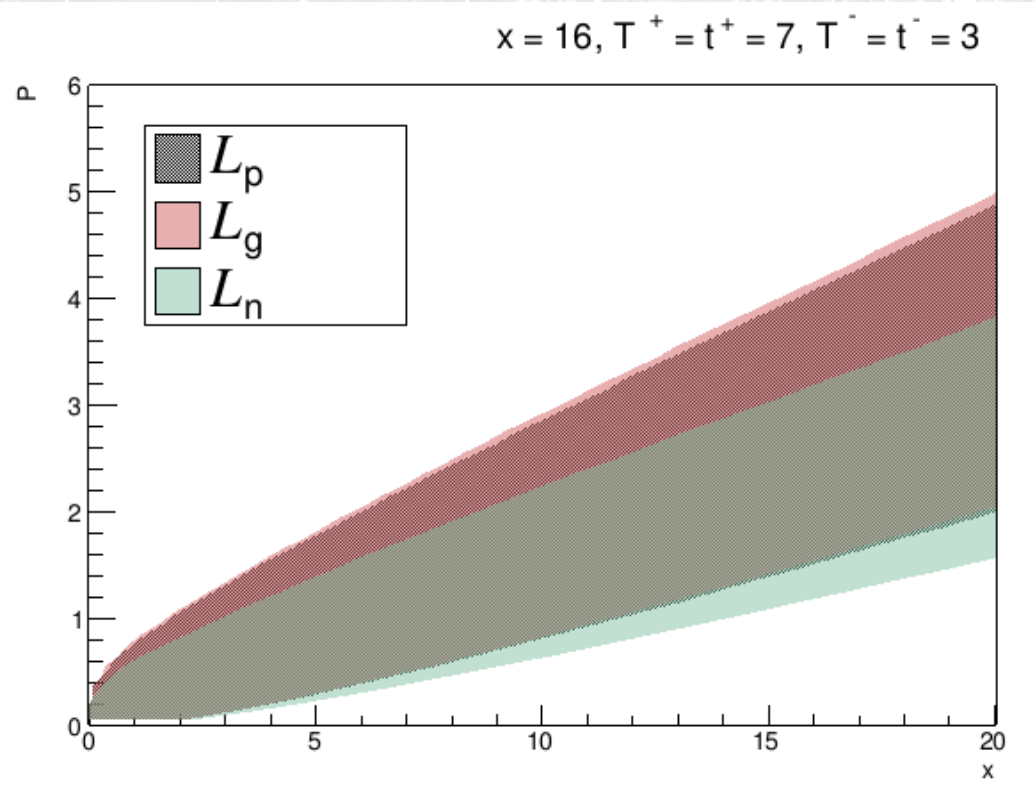
$$\mathcal{L}_p = \mathcal{P}(X|\pi \cdot (C + T^+ - T^-)) \cdot \mathcal{P}(t^+|T^+) \cdot \mathcal{P}(t^-|T^-) \cdot \mathcal{H}(T^+ - T^-)$$

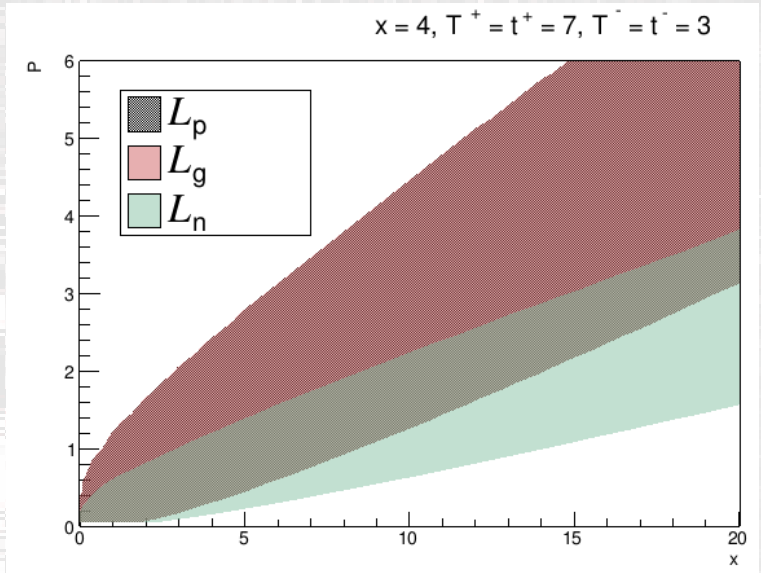
Percent of right estimated parameter π:

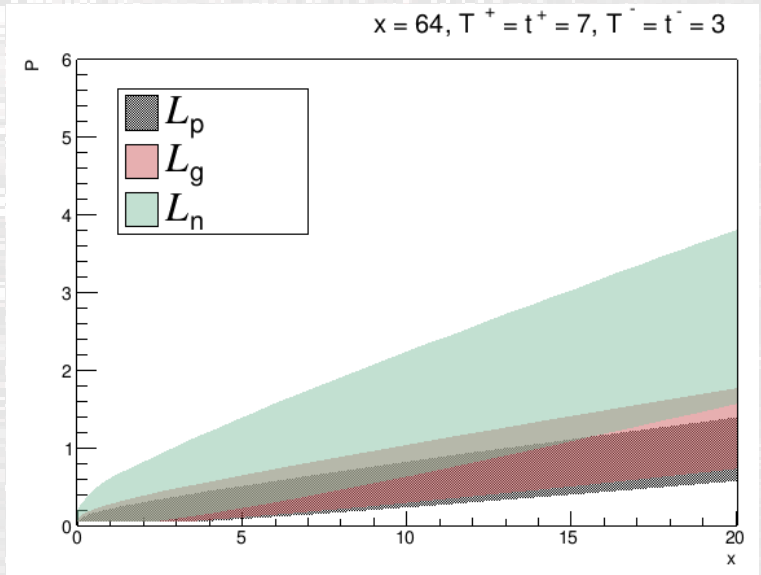
$\mathcal{L}_0$ parameters	CL	$\mathcal{L}_n$	$\mathcal{L}_g$	$\mathcal{L}_p$
$\pi = 3, T^+ = 12, T^- = 4$	1σ	34.71 ± 0.79	67.79 ± 0.75	67.98 ± 0.72
	2σ	62.82 ± 0.70	95.25 ± 0.27	96.14 ± 0.22
$\pi = 3, T^+ = 9, T^- = 7$	1σ	26.41 ± 0.57	77.52 ± 0.63	78.51 ± 0.63
	2σ	49.73 ± 0.71	98.18 ± 0.15	98.21 ± 0.15

Neyman constructions

Symmetric intervals for $CL = 0.95$, maximize likelihood over T^+, T^-:







Conclusion

- The transformation (1), (2) are proposed for construction of likelihood taking into account systematic uncertainties due to finite MC statistics in the presence of negative weights
- Gauss approximation of two Poisson multiplication work quite similar in single bin example $\Rightarrow$ if the your statistical software use Gauss for Barlow-Beeston (or for "light") you can keep use it without changes
- (1), (2) may be useful in more strict analysis with a more bins / more MC templates