Summary of studies on limit setting with CLsGenerator and BayesianMCMC

E. Busato

October 25, 2013

Validation of CLsGenerator without systematics and other checks

Validation of CLsGenerator without systematics

- When there's no systematics, an analytical solution for the limit can be found
  - This solution is used to validate CLsGenerator without systematics
- The analytical solution will also be derived and compared to CLsGenerator in the gaussian case.

Analytical solution (1)

- We have:
  - $N \sim \mathscr{P}(\mu s + b)$ under the $s+b$ hypothesis
  - $N \sim \mathscr{P}(b)$ under the b hypothesis
- Consider first a $CL_{\mathrm{s+b}}$ limit. Upper limit μ_{up} at $1-\alpha$ confidence level given by

$$CL_{\mathrm{s+b}} = \sum_{n=0}^{N\mathrm{obs}} P(n; \mu_{\mathrm{up}} s + b) = \alpha$$

- Poisson c.d.f. given by

$$\sum_{n=0}^{N_{\mathrm{obs}}} P(n; \nu) = 1 - F_{\chi^2}\left(2\nu; 2\left(N_{\mathrm{obs}} + 1\right)\right)$$

- Thus

$$\mu_{\mathrm{up}} = \frac{0.5 \times F^{-1}_{\chi^2}(1-\alpha; 2(N_{\mathrm{obs}}+1)) - b}{s}$$

Analytical solution (2)

- CLsGenerator computes CL_s limits, not $CL_\mathrm{s+b}$. Previous formula can be generalized to CL_s by replacing α by $\alpha \times CL_b$. Indeed:

$$\frac{CL_\mathrm{s+b}}{CL_\mathrm{b}} = \frac{\sum_{n=0}^{N\mathrm{obs}} P(n;\mu_\mathrm{up} s+b)}{\sum_{n=0}^{N\mathrm{obs}} P(n;b)} = \alpha$$

$$\Rightarrow 1 - F_{\chi^2}\left(2(\mu_\mathrm{up} s + b); 2\left(N_\mathrm{obs}+1\right)\right) = \alpha \times \left[1 - F_{\chi^2}\left(2b; 2\left(N_\mathrm{obs}+1\right)\right)\right]$$

- CL_b is independent of the signal, so we just need to replace α by $\alpha \times CL_\mathrm{b}$, yielding

$$\boxed{\mu_\mathrm{up} = \frac{0.5 \times F_{\chi^2}^{-1}\left(1-\alpha\left(1-F_{\chi^2}\left(2b; 2\left(N_\mathrm{obs}+1\right)\right)\right); 2\left(N_\mathrm{obs}+1\right)\right) - b}{s}}$$

Analytical solution in gaussian case (1)

- Gaussian approximation:
  - $N \sim \mathcal{N}(\mu s + b, \sqrt{\mu s + b})$ under the $s+b$ hypothesis
  - $N \sim \mathcal{N}(b, \sqrt{b})$ under the b hypothesis
- Upper $CL_{\mathrm{s+b}}$ limit μ_{up} at $1-\alpha$ confidence level given by

$$
\begin{aligned}
CL_{\mathrm{s+b}} = \int_0^{N_{\mathrm{obs}}} G\left(N; \mu_{\mathrm{up}} s + b, \sqrt{\mu_{\mathrm{up}} s + b}\right) \mathrm{d}N = \alpha \\
\Rightarrow \int_0^{\frac{N_{\mathrm{obs}} - (\mu_{\mathrm{up}} s + b)}{\sqrt{\mu_{\mathrm{up}} s + b}}} G(N; 0, 1) \mathrm{d}N = \alpha \\
\Rightarrow \Phi\left(\frac{N_{\mathrm{obs}} - (\mu_{\mathrm{up}} s + b)}{\sqrt{\mu_{\mathrm{up}} s + b}}\right) = \alpha
\end{aligned}
$$

where Φ is the c.d.f. of the standard normal distribution

$$
\Rightarrow N_{\mathrm{obs}} - (\mu_{\mathrm{up}} s + b) = -\sqrt{\mu_{\mathrm{up}} s + b} \times \Phi^{-1}(1-\alpha)
$$

Analytical solution in gaussian case (2)

- Previous expression can be solved for μ_{up}:

$$\boxed{\mu_{\text{up}} = \frac{N_{\text{obs}} - b}{s} + \frac{\left[\Phi^{-1}(1-\alpha)\right]^2}{2s}\left[1 + \sqrt{1 + 4\frac{Nobs}{\left[\Phi^{-1}(1-\alpha)\right]^2}}\right]}\,(\star)$$

- Previous formula can be generalized to CL_s by replacing α by $\alpha \times CL_b$ (as in the Poisson case). Indeed:

$$\frac{CL_{\text{s+b}}}{CL_{\text{b}}} = \frac{\displaystyle\int_0^{N_{\text{obs}}} G\left(N; \mu_{\text{up}}s + b, \sqrt{\mu_{\text{up}}s + b}\right)\text{d}N}{\displaystyle\int_0^{N_{\text{obs}}} G\left(N; b, \sqrt{b}\right)\text{d}N} = \alpha$$

$$\Rightarrow \Phi\left(\frac{N_{\text{obs}} - (\mu_{\text{up}}s + b)}{\sqrt{\mu_{\text{up}}s + b}}\right) = \alpha \times \Phi\left(\frac{N_{\text{obs}} - b}{\sqrt{b}}\right)$$

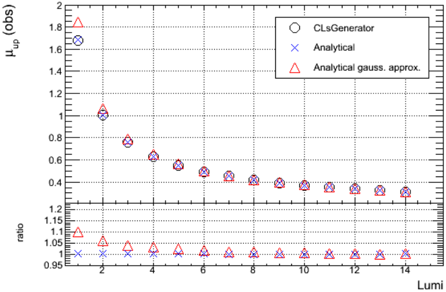
CL_{b} is independent of the signal, so we just need to replace α by $\alpha \times \Phi\left(\frac{N_{\text{obs}} - b}{\sqrt{b}}\right)$ in $(\star)$.

CLsGenerator vs analytical solutions

We take

- $N_{obs} = 1$
- $b = 0.82$
- $s = 2.49$
- No stat/syst uncertainty

and calculate limits for several luminosities (previous numbers correspond to $\mathscr{L} = 1$)

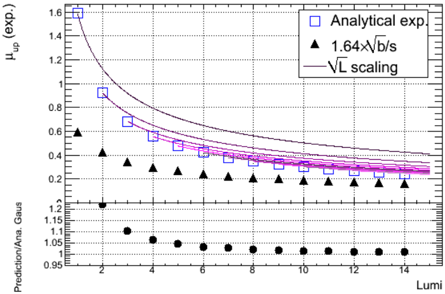


Conclusions:

- CLsGenerator and analytical agree at the 2 per mil level
- CLsGenerator and analytical gaussian approx. agree well when number of events not too small

Comparison to $1.64\frac{\sqrt{b}}{s}$ and scaling with luminosity

- $\mu_{\text{up}} = 1.64\frac{\sqrt{b}}{s}$ is sometimes used to compute limits quickly.
  - From this formula, one often predicts that limit should scale as $\sqrt{L}$
- As before, we take $N_{obs} = 1,\ b = 0.82,\ s = 2.49$ without systematics and compute limit for several luminosities



- Conclusions:
  - $1.64\frac{\sqrt{s}}{b}$ quite different from analytical result.
  - however, $\sqrt{L}$ scaling is a good approximation when number of events not too small

Where does $1.64\frac{\sqrt{b}}{s}$ come from ?

- Let's assume that
  1. $N_{\rm obs} = b$ (i.e. we consider an expected limit)
  2. $b \gg \left[\Phi^{-1}(1-\alpha)\right]^2$
     - note that $\Phi^{-1}(1-\alpha) = 1.64$ for $\alpha = 0.05$, corresponding to $b \gg 2.7$
     - this assumption corresponds more or less to the gaussian approx

  Then, one sees directly from $(\star)$ that

$$\boxed{\mu_{\rm up} \simeq \Phi^{-1}(1-\alpha)\frac{\sqrt{b}}{s} = 1.64\frac{\sqrt{b}}{s}}$$

- Remark: analytical solutions are more correct than the simple $1.64\frac{\sqrt{b}}{s}$ and no more complicated to use in sensitivity/optimization studies $\Rightarrow$ better to use them

Comparison of CLsGenerator and pure bayesian

Purpose of the study

- Purpose:
  - Always interesting to compare various methods
  - Comparing the two approaches tells us how sensitive the analysis is on the treatment of systematics
    - ⋆ CL_s and bayesian with uniform prior are equivalent without systematics $\Rightarrow$ differences arise only from the different treatment of systematics
- Bayesian implementation validated by comparing it to CLsGenerator/analytical without systematics

Bayesian implementation

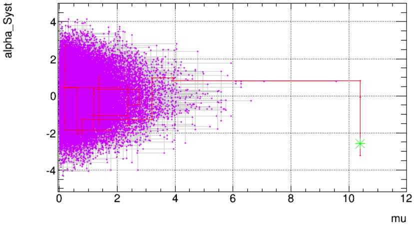
- Statistical model built using RooFit/RooStat
  - It is exactly the same as the CLsGenerator (mclimit) one
  - It is build from the same input files as CLsGenerator
- Marginal likelihood is

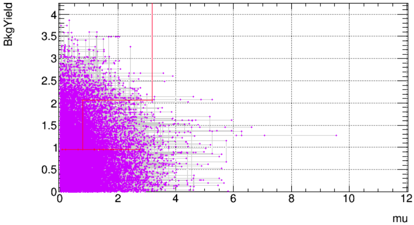
$$\mathscr{L}_{\mathrm{m}}(\mu) = \int \frac{(\mu s + \sum_i b_i)^{N_{obs}}}{N_{obs}!} e^{-(\mu s + \sum_i b_i)} \prod_j e^{-\frac{\eta_j^2}{2}} \prod_i e^{-\frac{(\gamma_i - 1)^2}{2\sigma_i^2}} e^{-\frac{(\gamma_s - 1)^2}{2\sigma_s^2}} \prod_j \mathrm{d}\eta_j \prod_i \mathrm{d}\gamma_i \mathrm{d}\gamma_s$$

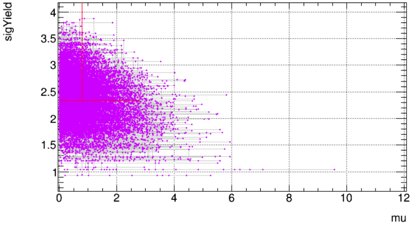
- Marginalization done by Markov Chain Monte Carlo

Visualization of Markov Chain

- Plots below shows markov chain for following configuration:
  +bg Bkg 0.82 1
  .syst Syst 0.02 -0.1
  +sig Sig 2.49 0.4
  +data 1

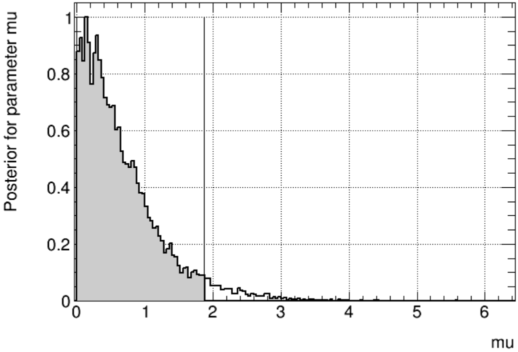






Posterior

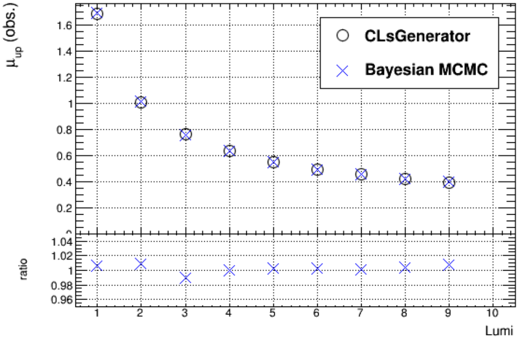
- Plot below shows posterior and 95% CL interval for same configuration as in previous slide



Validation of bayesian implementation

- CLsGenerator and bayesian (uniform prior) should be the same without systematics

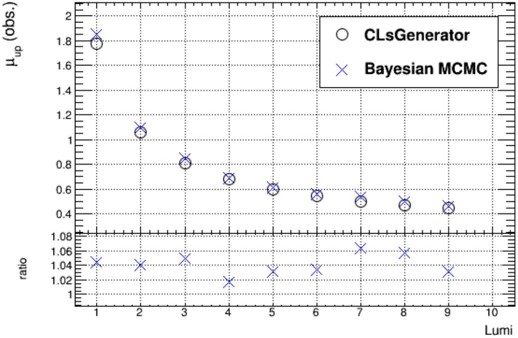
- Configuration
  - $N_{obs} = 1$
  - $b = 0.82$
  - $s = 2.49$
  - No stat/syst uncertainty



- Conclusion: bayesian implementation validated

CLsGenerator/bayesian comparison with systematics

- Configuration
  - $N_{obs} = 1$
  - $b = 0.82$
  - $s = 2.49$
  - Stat uncert. on b: 0.5
  - Stat uncert. on s: 0.3
  - Syst uncert. on s:
    - up=0.05
    - down=−0.05
  - Syst uncert. on b (100% correlated with s):
    - up=0.3
    - down=−0.3



- Effect of systematics treatment difference starts to be visible
- Need to study the effect of systematics further in both cases

Study of increase in lumi vs increase in # of channels

Lumi vs # channels (1)

- We expect that increasing luminosity by factor N is equivalent to increasing number of channels (having the same sensitivity) by the same factor (demo. on next slides)
- Here we check that CLsGenerator and BayesianMCMC are in agreement with this expectation
  - Validates in both cases implementation of channel combination

Demonstration equivalence lumi/#channels (CLsGenerator)

- Single channel: $q_\mu = 2\left[\mu s - N_{\text{obs}} \ln \frac{\mu s + b}{b}\right]$
- N channels: $q'_\mu = 2\left[\mu \sum_{c:channels} s_c - \sum_{c:channels} N^c_{\text{obs}} \ln \frac{\mu s_c + b_c}{b_c}\right]$
  - Assume yields are the same in all channels ($s_c = s,\ b_c = b$, $N^c_{\text{obs}} = N_{\text{obs}}$):
    $q'_\mu = 2\left[\mu N s - N_{\text{obs}} \ln\left(\prod_{c:channels} \frac{\mu s_c + b_c}{b_c}\right)\right] = N \times 2\left[\mu s - N_{\text{obs}} \ln \frac{\mu s + b}{b}\right]$
    $\Rightarrow q'_\mu = N q_\mu$
- Multiplying lumi. by N: $q''_\mu = 2\left[\mu s'' - N''_{\text{obs}} \ln \frac{\mu s'' + b''}{b''}\right]$
  - $s'' = Ns,\ b'' = Nb$ and $N''_{\text{obs}} = N N_{\text{obs}}$

$\Rightarrow q''_\mu = N \times 2\left[\mu s - N_{\text{obs}} \ln \frac{\mu s + b}{b}\right] = N q_\mu$

- Conclusion: $q''_\mu = q'_\mu$

Demonstration equivalence lumi/#channels (BayesianMCMC)

- Single channel: $\mathscr{L}(\mu) = \frac{(\mu s+b)^{N_{\text{obs}}}}{N_{\text{obs}}\,!} e^{-(\mu s+b)}$
- N channels:
  $\mathscr{L}'(\mu) = \prod_{c:channels} \mathscr{L}_c(\mu) = \prod_{c:channels} \frac{(\mu s_c+b_c)^{N^c_{\text{obs}}}}{N^c_{\text{obs}}\,!} e^{-(\mu s_c+b_c)}$
  - Assume yields are the same in all channels ($s_c = s$, $b_c = b$, $N^c_{\text{obs}} = N_{\text{obs}}$):
    $\Rightarrow \mathscr{L}'(\mu) = \left[\frac{(\mu s+b)^{N_{\text{obs}}}}{N_{\text{obs}}\,!} e^{-(\mu s+b)}\right]^N = \frac{1}{(N_{\text{obs}}!)^N} (\mu s + b)^{NN_{\text{obs}}} e^{-N(\mu s+b)}$
- Multiplying lumi. by N: $\mathscr{L}''(\mu) = \frac{(\mu s''+b'')^{N''_{\text{obs}}}}{N''_{\text{obs}}\,!} e^{-(\mu s''+b'')}$
  - $s'' = Ns$, $b'' = Nb$ and $N''_{\text{obs}} = NN_{\text{obs}}$

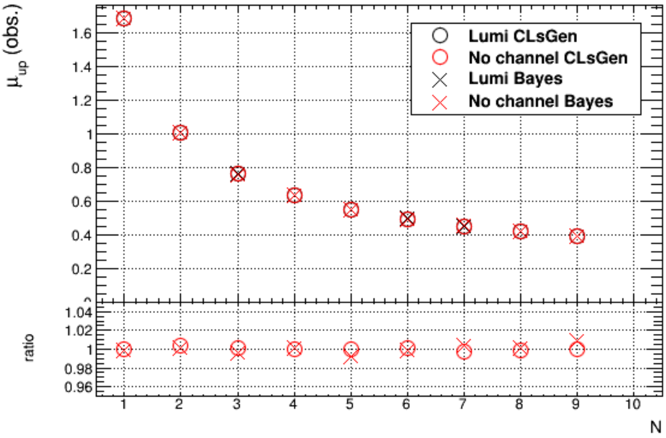
  $\Rightarrow \mathscr{L}''(\mu) = \frac{(\mu Ns+Nb)^{NN_{\text{obs}}}}{(NN_{\text{obs}})!} e^{-(\mu Ns+Nb)} =$
  $\frac{N^{NN_{\text{obs}}}}{(NN_{\text{obs}})!} (\mu s + b)^{NN_{\text{obs}}} e^{-N(\mu s+b)}$
- Conclusion: $\mathscr{L}''(\mu) \propto \mathscr{L}'(\mu)$ so inference on μ is the same

Lumi vs # channels (2)

- We consider this initial situation (no stat./syst. uncertainties):
  - $N_{obs} = 1$
  - $b = 0.82$
  - $s = 2.49$
- Then we compare two limits:

first one computed from one channel with ★ $N_{obs} = 1 \times N$ ★ $b = 0.82 \times N$ ★ $s = 2.49 \times N$	second one computed from N identical channels with ★ $N_{obs} = 1$ ★ $b = 0.82$ ★ $s = 2.49$

Lumi vs # channels (3)



- Conclusion: expectation verified in both CLsGenerator and BayesianMCMC

Studies on interpolation/extrapolation in CLsGenerator

Interpolation/extrapolation of systematics (1)

- For one sample and one systematic uncertainty we know
  - N_{nom}: nominal yield
  - $N_{\uparrow}$: yield with systematic varied 1σ up
  - $N_{\downarrow}$: yield with systematic varied 1σ down
- However, for the purpose of setting limits we need a continuous parametrization of the yield: $N(\eta)$
  - η defined such that $N(\eta = 0) = N_{nom}$, $N(\eta = +1) = N_{\uparrow}$ and $N(\eta = -1) = N_{\downarrow}$
- How do we interpolate for $\eta = [-1, 1]$ and extrapolate for $\eta < -1$ and $\eta > 1$?

Interpolation/extrapolation of systematics (1)

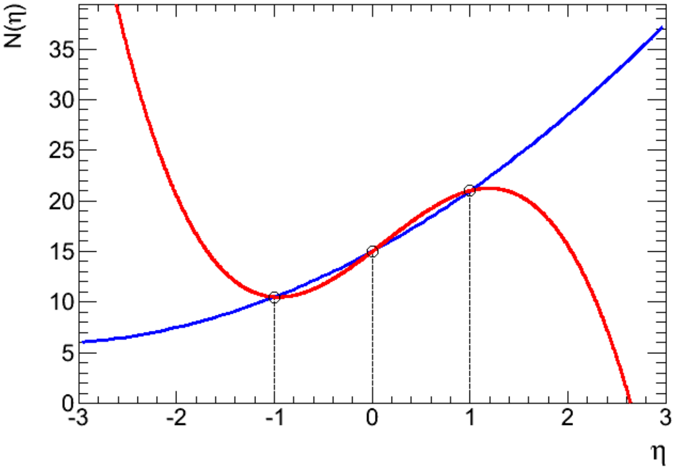
- For one sample and one systematic uncertainty we know
  - N_{nom}: nominal yield
  - $N_{\uparrow}$: yield with systematic varied 1σ up
  - $N_{\downarrow}$: yield with systematic varied 1σ down
- However, for the purpose of setting limits we need a continuous parametrization of the yield: $N(\eta)$
  - η defined such that $N(\eta = 0) = N_{nom}$, $N(\eta = +1) = N_{\uparrow}$ and $N(\eta = -1) = N_{\downarrow}$
- How do we interpolate for $\eta = [-1, 1]$ and extrapolate for $\eta < -1$ and $\eta > 1$?

Interpolation/extrapolation of systematics (1)

- For one sample and one systematic uncertainty we know
  - N_{nom}: nominal yield
  - $N_{\uparrow}$: yield with systematic varied 1σ up
  - $N_{\downarrow}$: yield with systematic varied 1σ down
- However, for the purpose of setting limits we need a continuous parametrization of the yield: $N(\eta)$
  - η defined such that $N(\eta = 0) = N_{nom}$, $N(\eta = +1) = N_{\uparrow}$ and $N(\eta = -1) = N_{\downarrow}$
- How do we interpolate for $\eta = [-1, 1]$ and extrapolate for $\eta < -1$ and $\eta >1$?

Interpolation/extrapolation of systematics (2)

Example : $N_{nom} = 15,\ N_{\uparrow} = 21,\ N_{\downarrow} = 10.5$



Interpolation/extrapolation of systematics (3)

- Rather than using $N_{nom,\uparrow,\downarrow}$, let's use
  - $f^{\uparrow} = \frac{N_{\uparrow} - N_{nom}}{N_{nom}}$
  - $f^{\downarrow} = \frac{N_{\downarrow} - N_{nom}}{N_{nom}}$
  - $f^{syst}(\eta) = \frac{N(\eta)}{N_{nom}}$
- One has
  - $f^{syst}(\eta = 0) = 1$
  - $f^{syst}(\eta = -1) = 1 + f^{\downarrow}$
  - $f^{syst}(\eta = 1) = 1 + f^{\uparrow}$
- Goal: find an inter/extrapolation algorithm such that these relations are satisfied (at least approximately)

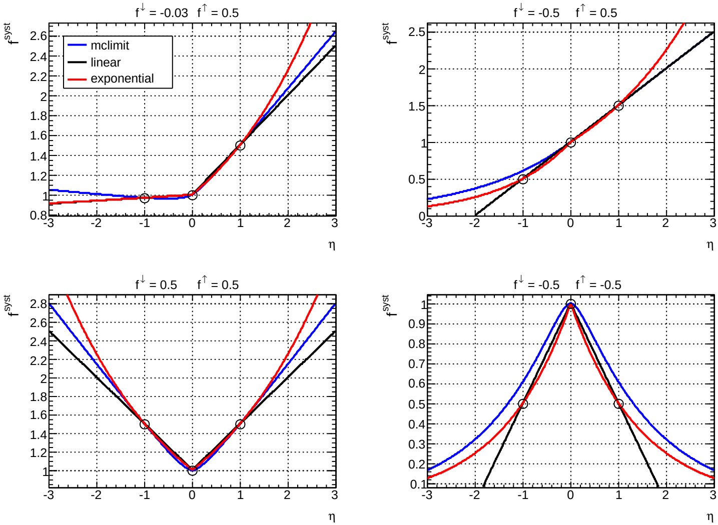
Interpolation/extrapolation of systematics (3)

- Rather than using $N_{nom,\uparrow,\downarrow}$, let's use
  - $f^{\uparrow} = \frac{N_{\uparrow} - N_{nom}}{N_{nom}}$
  - $f^{\downarrow} = \frac{N_{\downarrow} - N_{nom}}{N_{nom}}$
  - $f^{syst}(\eta) = \frac{N(\eta)}{N_{nom}}$
- One has
  - $f^{syst}(\eta = 0) = 1$
  - $f^{syst}(\eta = -1) = 1 + f^{\downarrow}$
  - $f^{syst}(\eta = 1) = 1 + f^{\uparrow}$
- Goal: find an inter/extrapolation algorithm such that these relations are satisfied (at least approximately)

Interpolation/extrapolation of systematics (3)

- Rather than using $N_{nom,\uparrow,\downarrow}$, let's use
  - $f^{\uparrow} = \frac{N_{\uparrow} - N_{nom}}{N_{nom}}$
  - $f^{\downarrow} = \frac{N_{\downarrow} - N_{nom}}{N_{nom}}$
  - $f^{syst}(\eta) = \frac{N(\eta)}{N_{nom}}$
- One has
  - $f^{syst}(\eta = 0) = 1$
  - $f^{syst}(\eta = -1) = 1 + f^{\downarrow}$
  - $f^{syst}(\eta = 1) = 1 + f^{\uparrow}$
- Goal: find an inter/extrapolation algorithm such that these relations are satisfied (at least approximately)

First look at mclimit, linear and exponential algorithms



Closer look at mclimit algorithm (1)

- What does mclimit give for $\eta = -1, 0, +1$?
  - $f^{syst}(\eta = 0) = 1$
  - if $f^{\uparrow} \geq 0$: $f^{syst}(\eta = +1) = 1 + f^{\uparrow}$
  - if $f^{\downarrow} \geq 0$: $f^{syst}(\eta = -1) = 1 + f^{\downarrow}$
  - if $f^{\uparrow} < 0$: $f^{syst}(\eta = +1) = e^{f^{\uparrow}}$
  - if $f^{\downarrow} < 0$: $f^{syst}(\eta = -1) = e^{f^{\downarrow}}$

  In the last two cases, $f^{syst} \simeq 1 + f^{\uparrow(\downarrow)}$ only if $f^{\uparrow(\downarrow)} \simeq 0$ (if $f^{\uparrow(\downarrow)} \simeq -1$, difference can be large).

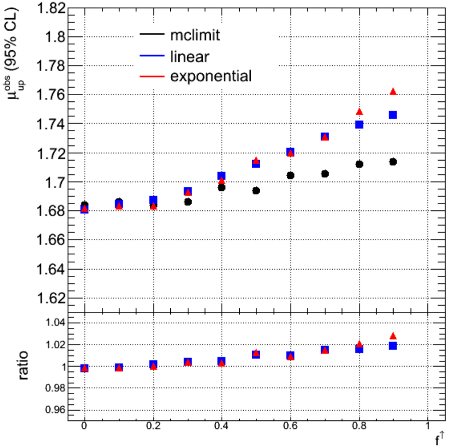
Closer look at mclimit algorithm (2)

- What does mclimit give when uncertainties are symetric: $f^{\uparrow} = -f^{\downarrow}$?
  - $\eta > 0$:
    - ★ $f^{\uparrow} \geq 0$: $f^{syst} = 1 + \eta f^{\uparrow}$
    - ★ $f^{\uparrow} < 0$: $f^{syst} = e^{\eta f^{\uparrow}}$
  - $\eta < 0$:
    - ★ $f^{\downarrow} \geq 0$: $f^{syst} = 1 - \eta f^{\downarrow}$
    - ★ $f^{\downarrow} < 0$: $f^{syst} = e^{-\eta f^{\downarrow}}$

mclimit is equivalent to linear for $\eta > 0$ if $f^{\uparrow} \geq 0$ or $\eta < 0$ if $f^{\downarrow} \geq 0$

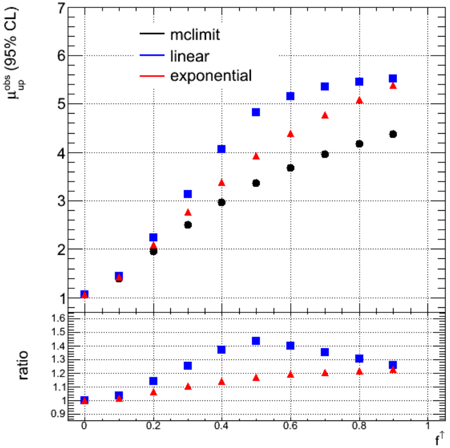
Comparison of algorithms: test 1

- $N_{obs} = 1$
- $b = 0.82$
- $s = 2.49$
- No stat uncertainty
- No syst uncertainty on s
- Syst uncertainty on b: $f^{\uparrow} = -f^{\downarrow} = 0, 0.1, 0.2, \ldots, 0.9$



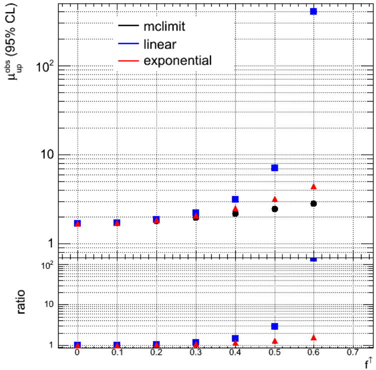
Comparison of algorithms: test 2

- $N_{obs} = 100$
- $b = 100$
- $s = 20$
- No stat uncertainty
- No syst uncertainty on s
- Syst uncertainty on b: $f^{\uparrow} = -f^{\downarrow} = 0, 0.1, 0.2, \ldots, 0.9$



Comparison of algorithms: test 3

- $N_{obs} = 1$
- $b = 0.82$
- $s = 2.49$
- No stat uncertainty
- Syst uncertainty on s and b (100% correlated): $f^{\uparrow} = -f^{\downarrow} = 0, 0.1, 0.2, \ldots, 0.6$

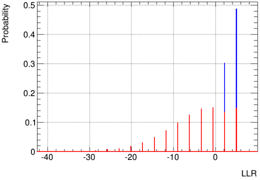


Comments:

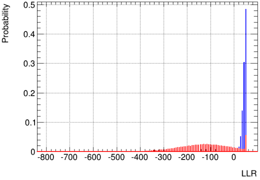
- $q_\mu = 2\left(\mu s - N_{obs} \ln \frac{\mu s + b}{b}\right)$
- When $f^{\downarrow} \ll 0$, $f^{syst} = 0$ quite frequently $\Rightarrow s = b = 0 \Rightarrow N_{obs} = 0 \Rightarrow q_\mu = 2\mu s$ (for both hypothesis)$\Rightarrow$CL$_{s+b}$ and CL$_s$ can't go to very low values as μ increases

Comparison of algorithms: test 3 (cont'd)

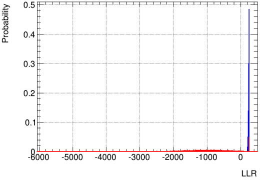
Consider the point $f^{\uparrow} = -f^{\downarrow} = 0.6$



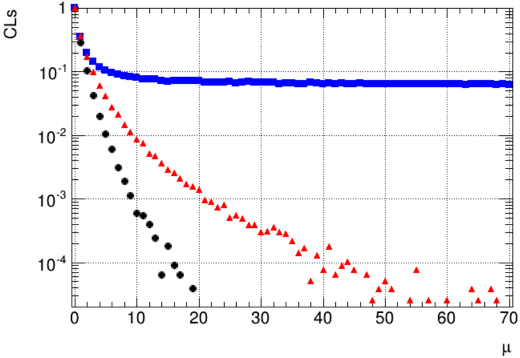
$\mu = 1$



$\mu = 10$



$\mu = 50$



Comparison of algorithms: test 3 (cont'd)

- After having studied a few situations by varying systematics on signal, backgrounds and correlations, it appears that one should be very careful with linear inter/extrapolation
  - when signal systematic uncertainty is large
    - ⋆ when in addition background systematics are large effect is more pronounced
    - ⋆ when in addition background systematics are large and correlated to signal systematics effect can be dramatic (see two previous slides)
- Remark: effect exists only if $f^{\uparrow}$ and/or $f^{\downarrow}$ (for the signal at least) is negative

Comparison of algorithms: LHCP mumu channel

Here we use 4tmm.dat from David (with +data 1)

- Limits

	mclimit	linear	expo
Expected median	1.70	1.70	1.69
Expected $\pm 1\sigma$	1.44-2.08	1.43-2.07	1.43-2.07
Expected $\pm 2\sigma$	1.26-3.52	1.27-3.53	1.27-3.54
Observed	1.67	1.68	1.67

Conclusion: the three algorithms give compatible results

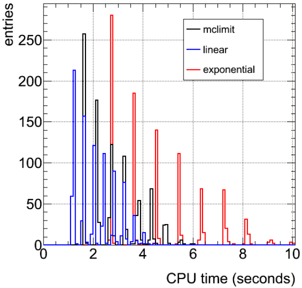
- CPU time

ran

clgen.observedSigStrengthFor95excl(0,1e5,cls)

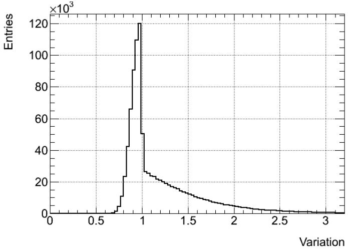
in identical conditions several

times to make this distrib



Understanding f^{syst} distribution: exponential case

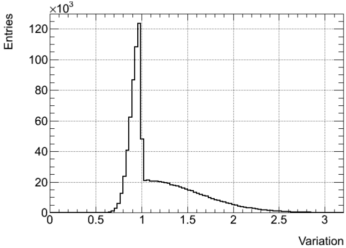
- f^{syst} sometimes looks very weird. Let's consider exponential algo with $f^{\downarrow} = -0.1$ and $f^{\uparrow} = 0.6$ (on the plot, Variation=f^{syst})



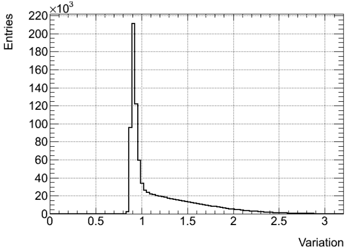
- Where does this funny looking shape comes from ?
  - We have $f^{syst} = (1+f^{\uparrow})^{\eta}$ for $\eta > 0$ ($(1+f^{\uparrow})^{-\eta}$ for $\eta < 0$) with $\eta \sim \mathcal{N}(0,1)$
  - It's straightforward to show that $f^{syst} \sim \frac{1}{\sqrt{2\pi}|\ln\left(1+f^{\uparrow(\downarrow)}\right)|f^{syst}} e^{-\frac{1}{2}\left(\frac{\ln f^{syst}}{\ln\left(1+f^{\uparrow(\downarrow)}\right)}\right)^2}$
  - i.e. $f^{syst} \sim$ piecewise $log\mathcal{N}$
- Note: understanding the shape of a distribution doesn't mean that such a distribution is desirable $\rightarrow$ is it ?

Understanding f^{syst} distribution: linear and mclimit cases

- linear case is obvious: $f^{syst} \sim$ piecewise gaussian truncated at 0. Using values of previous slide we have

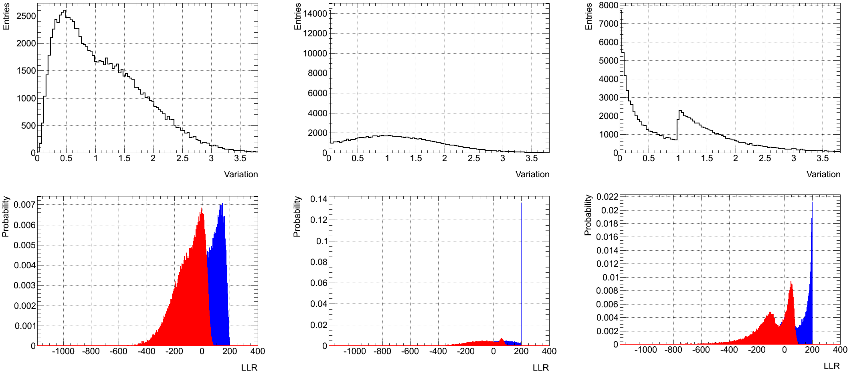


- mclimit case: doesn't seem to be possible to determine analytic solution. Using values of previous slide we have



Coming back to test 2

- In some extreme cases (very large systematics), linear and expo give really really worrying distributions
- Let's consider the point $f^{\uparrow} = -f^{\downarrow} = 0.9$ and $\mu = 5.5$ of test 2.



mclimit linear expo

- Using truncated piecewise gaussian (linear) and piecewise $log\mathscr{N}$ (expo) seems to be absurd $\rightarrow$mclimit seems to be much more reasonable in such large systematics case

Preliminary conclusions (1)

- mclimit, linear and exponential algorithms give similar results for small systematics
  - Any choice seems to be rather safe in this case
- Differences can be significant for large systematics
  - linear shouldn't be used when signal systematics (and background ?) are large
  - expo shouldn't be used when systematics on background and/or signal are large and $f^{\downarrow} \mathrm{or} f^{\uparrow} \ll 0$
  - mclimit seems to be the safest choice when systematics are large
- Argument of continuous derivative which at first sight seemed important only for MINUIT is actually also important when no fitting is done as it provides distributions of f^{syst} that are smooth (particularly important for large systematics).

Preliminary conclusions (2)

- mclimit:
  - pros: f^{syst} smooth, $f^{syst}(\eta) \neq 0 \ \forall \eta$
  - cons: $f^{syst}(\eta = \pm 1) \neq 1 + f^{\uparrow(\downarrow)}$ when $f^{\uparrow(\downarrow)} < 0$,
- linear:
  - pros: simple, fast
  - cons: $f^{syst}(\eta) = 0$ in some cases ($\Rightarrow$ problem when signal syst large)
- expo:
  - pros: simple, $f^{syst}(\eta) \neq 0 \ \forall \eta$
  - cons: f^{syst} can be very discontinuous for large $f^{\uparrow(\downarrow)}$, slow

Preliminary conclusions (3)

- Systematics are such a pain ! It's good that we're not affected too much by them in the same-sign analysis

Ideas for further studies (1)

- Try linear interpolation and expo extrapolation
- Alternative to TMath::Power() for exponential case ?
- Try $\eta \sim log\mathcal{N}$
  - Note that $\eta \sim log\mathcal{N}$ + linear is equivalent to $\eta \sim \mathcal{N}$ + expo (in both cases $f^{syst} \sim log\mathcal{N}$)
    - ⋆ As $\eta \sim \mathcal{N}$ + expo is quite slow, could be interesting to see if it's faster in terms of CPU to do $\eta \sim log\mathcal{N}$ + linear
- Try $\eta \sim Gamma$

Ideas for further studies (2)

Statistical model

$$\mathscr{L}(\mu,\{\eta\},\{\gamma\}) = \frac{(\mu s+\sum_i b_i)^{N_{obs}}}{N_{obs}!} e^{-(\mu s+\sum_i b_i)}$$

where

- $b_i = b_i(\{\eta\},\{\gamma\}) = b_i^{nom} \times \prod_j f_{ij}(\eta_j) \times \gamma_i$
- $s = s(\{\eta\},\{\gamma\}) = s^{nom} \times \prod_j f_j^s(\eta_j) \times \gamma_s$

and

- $\eta_j \sim \mathscr{N}(0,1)$
- $\gamma_i \sim \mathscr{N}(1,\sigma_i)$ and $\gamma_s \sim \mathscr{N}(1,\sigma_s)$

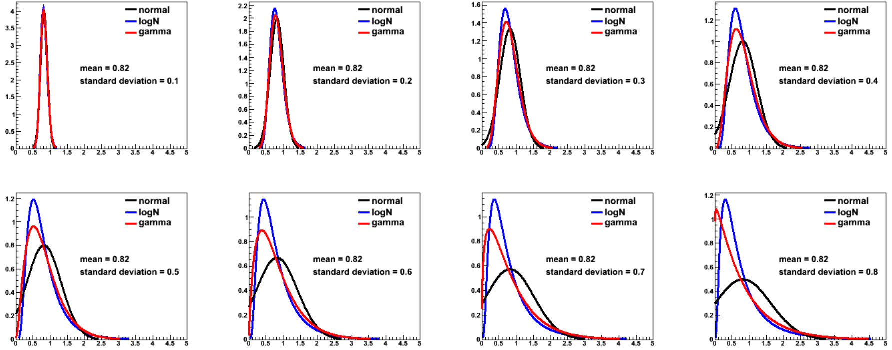
- When generating pseudo exp, N_{obs} generated according to marginal likelihood:
  $$\mathscr{L}_\mathrm{m}(\mu) = \int \frac{(\mu s+\sum_i b_i)^{N_{obs}}}{N_{obs}!} e^{-(\mu s+\sum_i b_i)} \prod_j e^{-\frac{\eta_j^2}{2}} \prod_i e^{-\frac{(\gamma_i-1)^2}{2\sigma_i^2}} e^{-\frac{(\gamma_s-1)^2}{2\sigma_s^2}} \prod_j \mathrm{d}\eta_j \prod_i \mathrm{d}\gamma_i \mathrm{d}\gamma_s$$
  However, $\mathscr{L}(\mu,\{\eta\}=0,\{\gamma\}=1)$ is used to calculate the test
  $\Rightarrow$ try calculating the test with marginal likelihood

Studies on stat uncertainty sampling in CLsGenerator

Purpose of the study

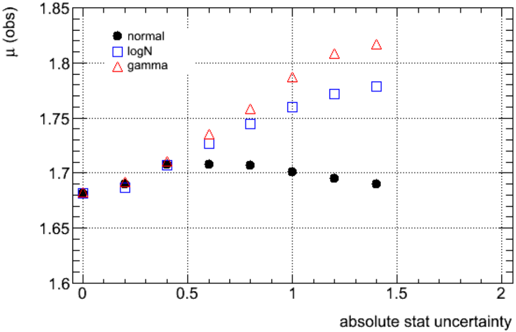
- Nuisance parameters for stat uncertainty sampled from normal distribution by default
- This choice is arbitrary and other distributions could be used
  - In this study normal is compared to logN and gamma
- Following results obtained with the mclimit inter/extrapolation

Comparison of distributions



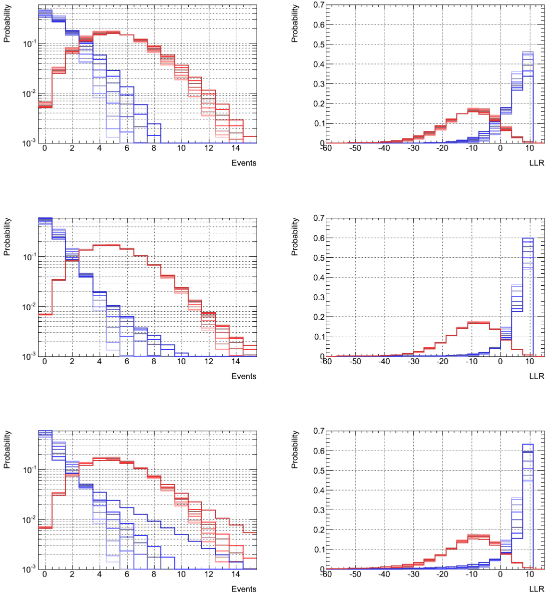
Comparison of samplings for particular case (1)

- $N_{obs} = 1$
- $b = 0.82$
- $s = 2.49$
- No syst uncertainty
- Stat uncertainty on $b = 0.1, ..., 1.4$



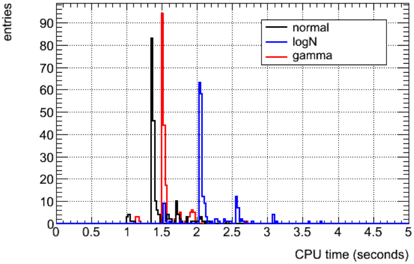
Comparison of samplings for particular case (2)

Normal (top), logN (middle) and Gamma (bottom)



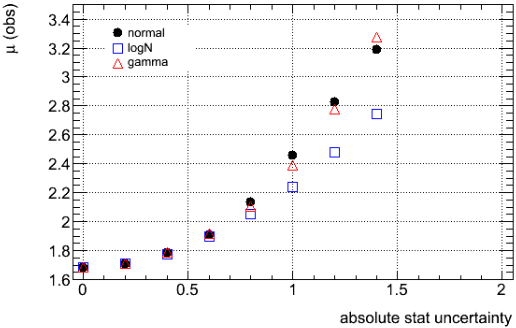
Comparison of samplings for particular case (3)

- Truncation at 0 in normal case doesn't smear distributions but shifts them
  - Happens for both b and s+b distribs $\Rightarrow \mu$ doesn't change much
- logN smears less than gamma $\Rightarrow$ better μ for logN than for gamma
- What is the best choice ? logN ?



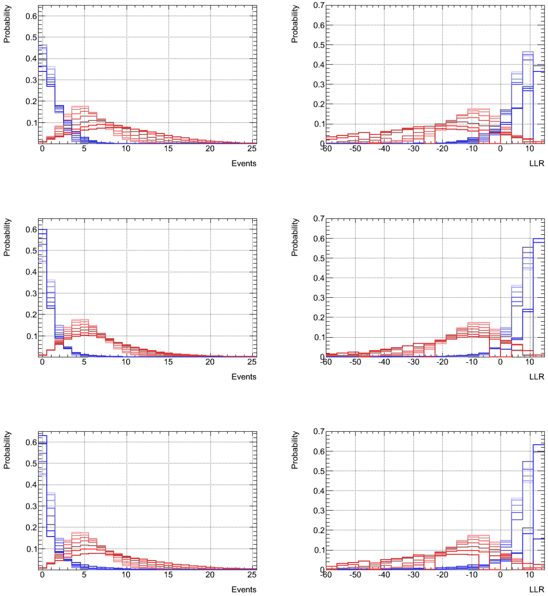
Comparison of samplings for another particular case (1)

- $N_{obs} = 1$
- $b = 0.82$
- $s = 2.49$
- No syst uncertainty
- Stat uncertainty on s and $b = 0.1, ..., 1.4$



Comparison of samplings for another particular case (2)

Normal (top), logN (middle) and Gamma (bottom)



Comparison of samplings for another particular case (3)

- logN smears less than gamma $\Rightarrow$ better μ for logN than for gamma (as in previous case)
- Normal case now very different from previous case
    - s+b distribution is now "more smeared" than b alone distribution $\Rightarrow$ have to go to larger μ to reach given confidence level