Frequentist Limit Recommendation ATLAS

Eilam Gross

for ATLAS Statsitics Forum and Higgs Group

The Test Statistic :

- $$q_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}; & \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

- $$\tilde{q}_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})} & \hat{\mu} < 0 \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})} & 0 < \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

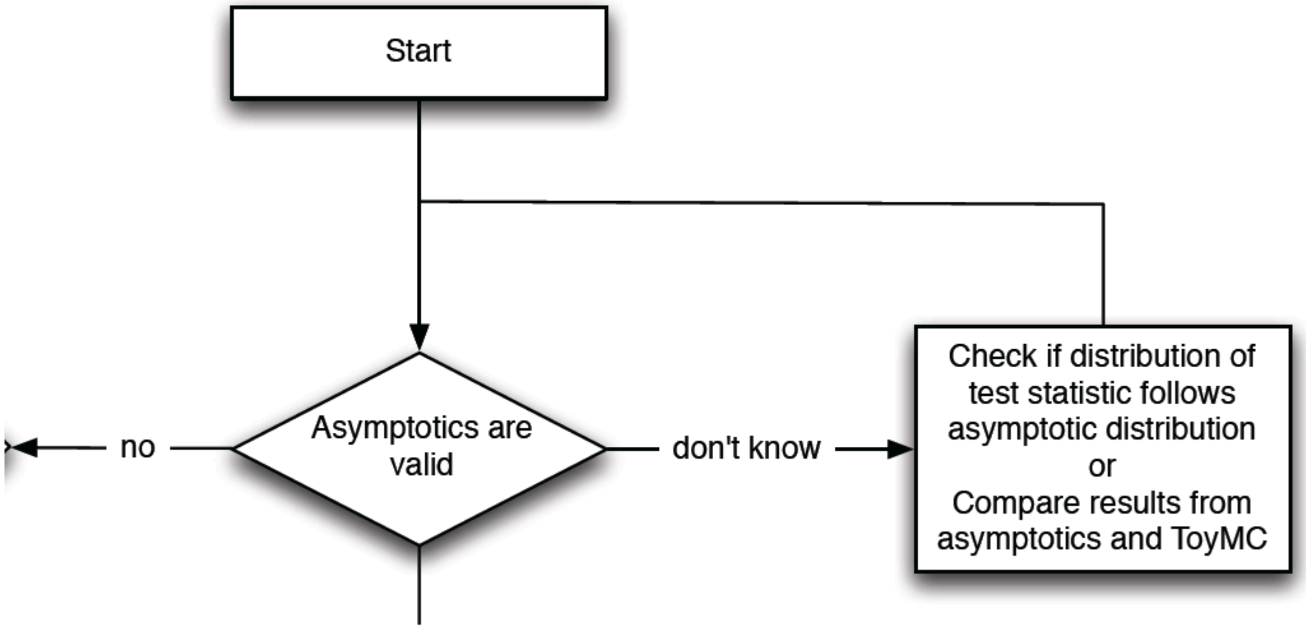
$$q_{\mu,obs} \qquad\qquad \tilde{q}_{\mu,obs}$$

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid \mu)\, dq_\mu \quad \textit{asymotOtic} \approx \quad p_\mu = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu \mid \mu)\, d\tilde{q}_\mu$$

Ref: Asymptotic formulae for likelihood-based tests of new physics ,Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells, arXiv:1007.1727, accepted for publication in EPJC.

Flow Chart

kyle



The Test Statistic :

- $$q_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}; & \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

ATLAS Recommended

- $$\tilde{q}_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})} & \hat{\mu} < 0 \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})} & 0 < \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

$$q_{\mu,obs}$$

$$\tilde{q}_{\mu,obs}$$

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid \mu)\,dq_\mu \quad asymototic \approx \quad p_\mu = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu \mid \mu)\,d\tilde{q}_\mu$$

$$non-asymototic \quad p_\mu = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu \mid \mu, \hat{\hat{\vec{\theta}}}(\mu,\mathrm{obs}))\,d\tilde{q}_\mu$$

A naive model

- You measure n events in the signal region and the unknown background is b (nuisance parameter). Suppose your expectation is 100+-20% (using MC or side band)
- The likelihood can be written as

$$L = Poiss(n \mid \mu s + b) G(m_\delta \mid \delta, 1)$$

$$n = \mu s + b \ ; \quad b = b_0(1 + \epsilon\delta)$$

$$with \ \ b_0 = 100, \epsilon = 0.2$$

- The way to generate toy MCs, is to generate m_δ following a Gaussian around δ =0 with σ =1 and n as a Poisson with expectation μ s+b, where b=b_0 yet keep the nuisance parameter fixed and find its MLEs subject to the generated data
- When b_0 is unknown and measured from data, for s+b experiments use $b_0 = \hat{\hat{b}}(\mu, obs)$ for bg only use $b_0 = \hat{\hat{b}}(\mu = 0, obs)$
- In the asymptotic regime it does not matter

Generate Toy MCs

- Generate toy Monte Carlo experiments to construct the pdf of $\tilde{q}_\mu$ under signal (with strength μ) and background only experiments $f(\tilde{q}_\mu \mid \mu, \hat{\hat{\vec{\theta}}}(\mu,\text{obs}))$ *and* $f(\tilde{q}_\mu \mid 0, \hat{\hat{\vec{\theta}}}(0,\text{obs}))$
.
- Here the $\hat{\hat{\vec{\theta}}}(\mu,\text{obs})$
and $\hat{\hat{\vec{\theta}}}(0,\text{obs})$

  are the conditional MLEs based on the observed data. Also note, that the nuisance parameters are fixed to their conditional MLEs for generating the toy Monte Carlo, but are allowed to float in fits needed to evaluate the test statistic.
- In the asymptotic limit the distribution $f(\tilde{q}_\mu \mid \mu, \vec{\theta})$ is independent of $\vec{\theta}$

1-1. Construct the model (an Example)

$$\mu_T = \sum_l \mu L \sigma_l (1+\epsilon_l^s \delta_{\epsilon^s}) \prod_i (1+\epsilon_{li}^s \delta_i) + \sum_j L \beta_j^0 (1+\epsilon_j^b \delta_{\beta_j}) \prod_i (1+\epsilon_{ji}^b \delta_i)$$

- L is the nominal integrated luminosity,
- μ is the one parameter of interest, the signal strength,
- σ_l is the effective cross section (in pb) for signal events in channel l,
- ϵ_l^s is relative uncertainty on the efficiency of the channel l ,
- β_j^0 is the nominal effective cross section (in pb) for background j,
- ϵ_j^b is the relative uncertainty on the effective cross section for background j,
- ϵ_{li}^s is the relative change in the effective cross-section due to the i^{th} systematic effect on signal channel l, and
- ϵ_{ji}^b is the relative change in the effective cross-section due to the i^{th} systematic effect on channel j.

The nuisance parameters are $\boldsymbol{\theta} = (\beta^0_{j(j\in SB)}; \delta_{\epsilon^s}, \delta_{\beta_j}, \delta_i)$ and the δ are constrained by the normal distribution $N(m_\delta|\delta) = G(m_\delta|\delta, 1)$.

1-2. Construct the model (an Example)

$$L(\mu,\beta^0_{j(j\in SB)};\delta_{\epsilon^s},\delta_{\beta_j},\delta_i) = \mathrm{Pois}(n\,|\,\mu_T)\,N(m_{\delta_s}\,|\,\delta_{\epsilon^s})$$
$$\times\prod_{j\in SB}\mathrm{Pois}(n_j\,|\,\beta^0_j)\prod_j N(m_{\delta_{\beta_j}}\,|\,\delta_{\beta_j})\prod_i N(m_{\delta_i}\,|\,\delta_i)$$

The nuisance parameters are $\boldsymbol{\theta} = (\beta^0_{j(j\in SB)};\delta_{\epsilon^s},\delta_{\beta_j},\delta_i)$ and the δ are constrained by the normal distribution $N(m_\delta|\delta) = G(m_\delta|\delta,1)$.

- n is the number of events in the signal region,
- $m_{\delta_s}, m_{\delta_{\beta_j}}, m_{\delta_i}$ represent auxiliary measurements of the corresponding δ systematic uncertainties. When generating toy Monte Carlo experiments, the m_δ should fluctuate around the value of δ in $\hat{\hat{\boldsymbol{\theta}}}(\mu, obs)$ or $\hat{\hat{\boldsymbol{\theta}}}(0, obs)$. If they are not randomized, this corresponds to a conditional ensemble in which the distribution of the test statistic departs significantly from the asymptotic distributions. $\boldsymbol{\theta}$ is the vector of nuisance parameters given below.
- $n_{j(j\in SB)}$ is the number of events measured in the control sample which is scaled by an extrapolation coefficient τ to estimate the number of events in the signal region. Since τ itself has uncertainty, we standardize it by writing $\tau_{j,j\in SB} = 1 + \epsilon_{\beta_j}\delta_{\beta_j}$,

Find p-value and upper limit

- From the constructed distribution

$$f(\tilde{q}_\mu \mid \mu, \hat{\hat{\vec{\theta}}}(\mu, \text{obs}))$$

find the p-value of the observation

$$p_\mu = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu \mid \mu, \hat{\hat{\vec{\theta}}}(\mu, \text{obs}))\, d\tilde{q}_\mu$$

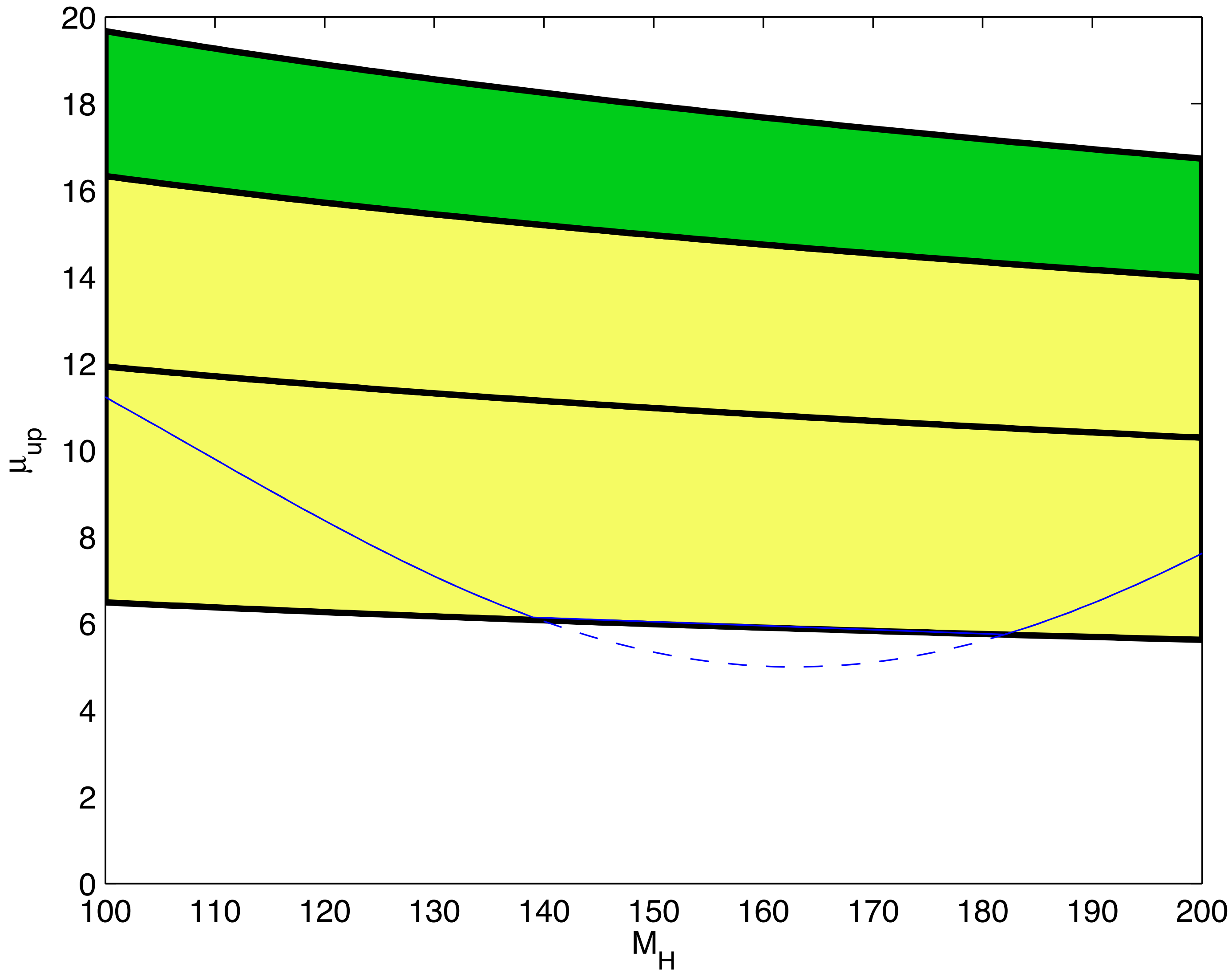
- Find (by iteration or any other way)

$$\mu_{up} \mid \quad p_{\mu_{up}} = 5\%$$

Median Sensitivity, Bands and PCL

- To find the median sensitivity, generate bg-only toy MCs.
- For each one find μ_{up}.
- Draw the μ_{up} distribution, find its median and the 1 and 2 sigma bands.
- To protect against downward fluctuations of the bg which allows to exclude signals to which the experiment is not sensitive to, we recommend to use a power of 16% and not allow the observed limit to go below the -1 σ band.
- Note, without the PCL the -2 σ band extends to $\mu_{up}=0$
- In terms of the yellow/green bands, the observed data should not be hidden - if we have a downward fluctuation, show it.
- But when it passes the power constraint, mark the power constraint solid and the downward fluctuation dotted so that the result we are using is clear.

Illustration



To compare with TEVATRON we recommend to calculate the CLs upper limit

- Define the following ratio of p-values $p'_\mu = \frac{p_\mu}{1-p_b}$
- Generate bg-only experiments and find

$$p_b = 1 - \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu \mid 0, \hat{\hat{\vec{\theta}}}(0,obs)) d\tilde{q}_\mu$$

- By scanning or iteration find $\mu_{up} \mid p'_{\mu_{up}} = 5\%$

This is a very time consuming procedure.

Validations

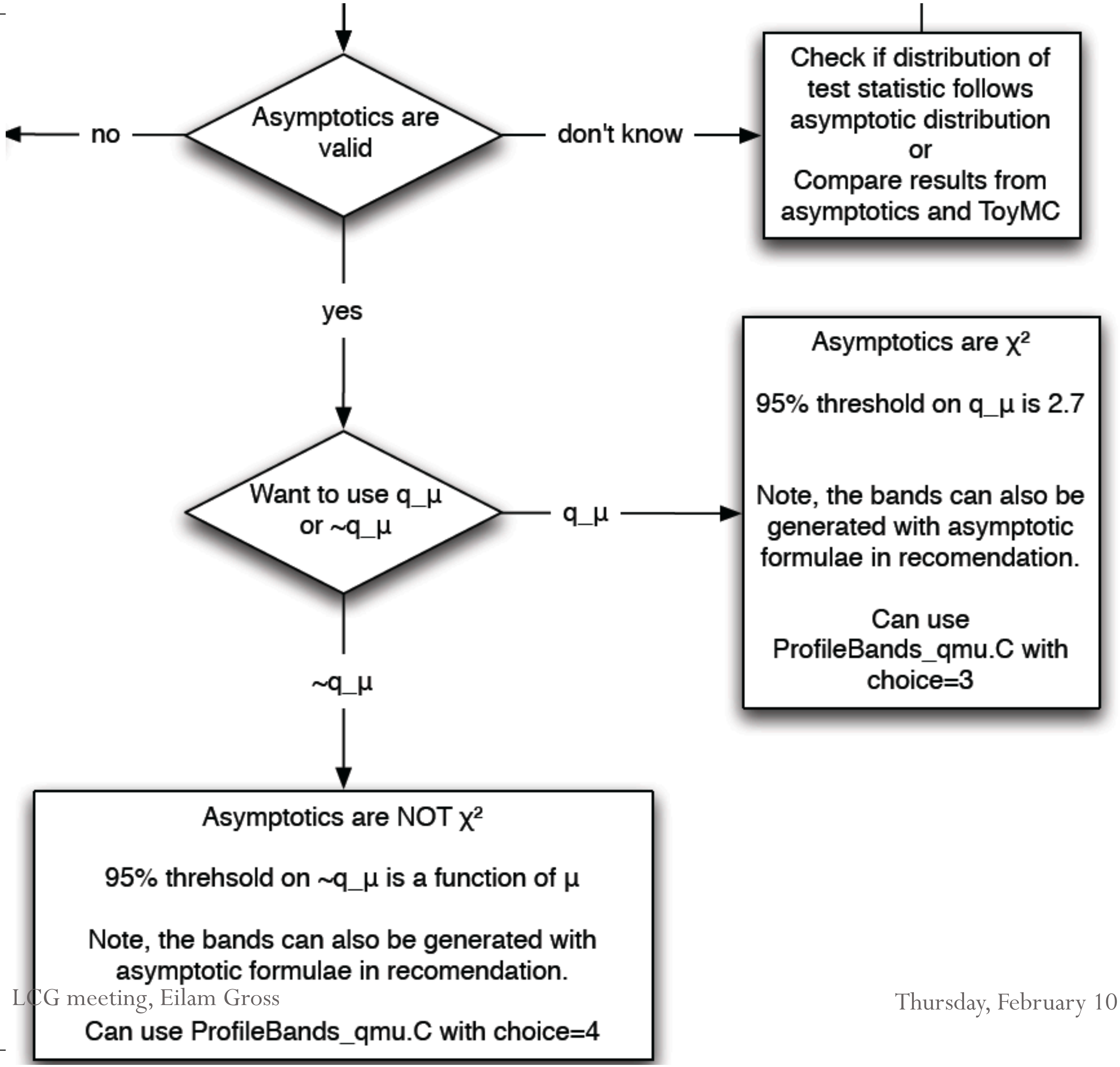
- Two codes, one independent and one with Roostats.
- Apply the asymptotic formulas to see if you get huge deviations, and track them down.
- Asymptotic formulae exist for both

$$q_\mu = \begin{cases} -2\ln\dfrac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}; & \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

$$\tilde{q}_\mu = \begin{cases} -2\ln\dfrac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})} & \hat{\mu} < 0 \\ -2\ln\dfrac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})} & 0 < \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

but as explained, in the asymptotic regime it does not matter.

kyle



Asymptotics are valid
no
don't know
Check if distribution of test statistic follows asymptotic distribution or Compare results from asymptotics and ToyMC
yes
Want to use q_μ or ~q_μ
q_μ
Asymptotics are χ²
95% threshold on q_μ is 2.7
Note, the bands can also be generated with asymptotic formulae in recomendation.
Can use ProfileBands_qmu.C with choice=3
~q_μ
Asymptotics are NOT χ²
95% threhsold on ~q_μ is a function of μ
Note, the bands can also be generated with asymptotic formulae in recomendation.
Can use ProfileBands_qmu.C with choice=4

Asymptotic Formulae for qμ

CL_{s+b}

- $$p_{\mu_{95}} = 1-\Phi(\sqrt{q_\mu}) = 1-\Phi(1.64) = 0.05$$
- Median
$$\mu_{up}^{med} = \sigma\Phi^{-1}(1-\alpha) = \sigma\Phi^{-1}(0.95) = 1.64\sigma$$
- Observed upper limit, solve
$$\mu_{up,obs} \mid q_{\mu_{up},obs} = 1.64^2$$
- Bands:
$$\mu_{up+N} = \sigma(1.64+N)$$
$$\mu_{up+N} = \sigma(\Phi^{-1}(1-\alpha)+N)$$
$$\sigma^2 = \frac{\mu^2}{q_{\mu,A}}$$

CL_s

$$p'_{\mu_{95}} = \frac{1-\Phi(\sqrt{q_{\mu_{95}}})}{\Phi(\sqrt{q_{\mu_{95},A}} - \sqrt{q_{\mu_{95}}})} = 0.05$$

$$\mu_{up}^{med} = \sigma\Phi^{-1}(1-0.5\alpha) = \sigma\Phi^{-1}(0.975)$$

$$\mu_{up,obs} \mid p'_{\mu_{up,obs}} = 0.05$$

$$\mu_{up+N} = \sigma(\Phi^{-1}(1-\alpha\Phi(N))+N)$$
$$\mu_{up+N} = \sigma(\Phi^{-1}(1-0.05\Phi(N))+N)$$
$$\sigma^2 = \frac{\mu^2}{q_{\mu,A}}$$

Ref: Asymptotic formulae for likelihood-based tests of new physics ,Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells, arXiv:1007.1727, accepted for publication in EPJC.

minute made: CL_{s+b} vs CL_s

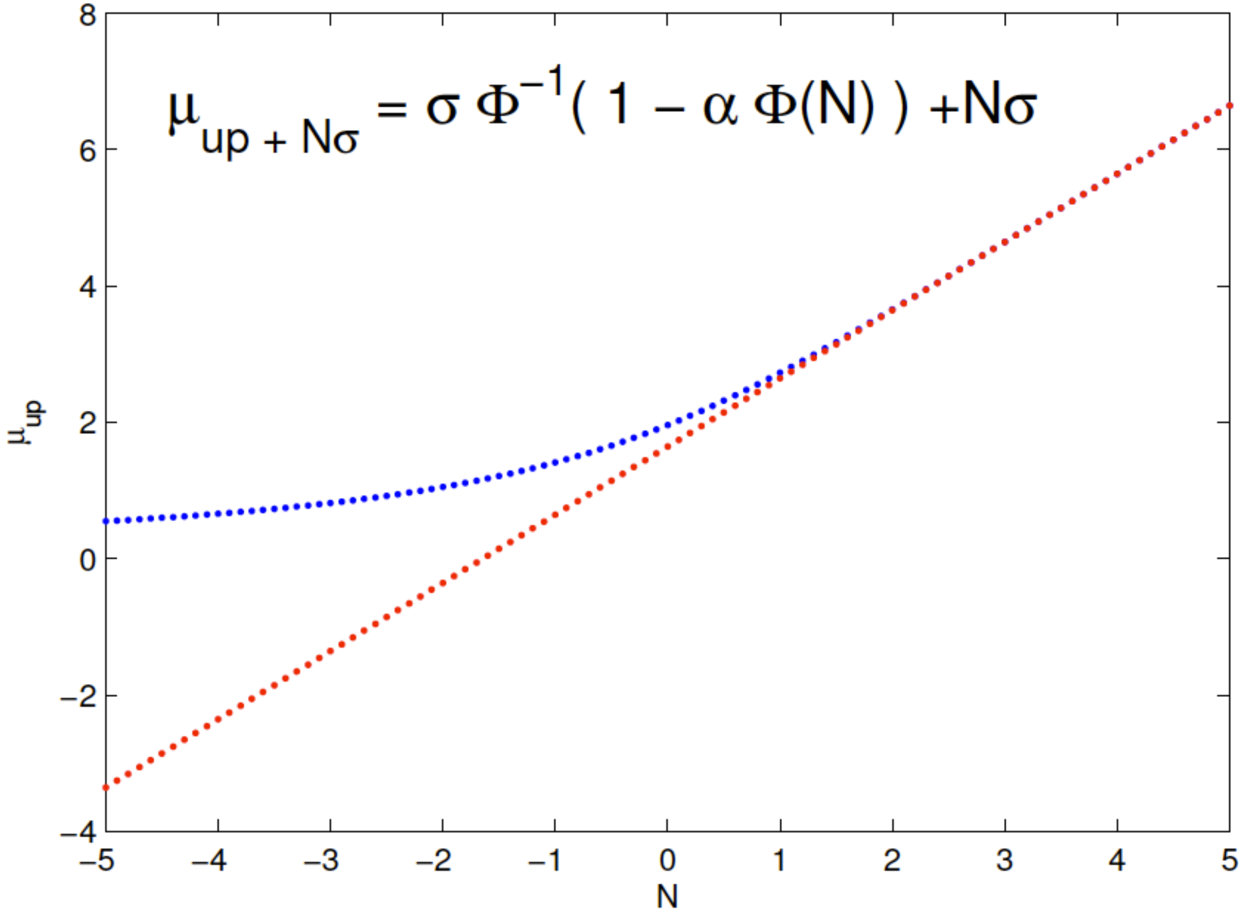


Figure 2: $\mu_{up+N\sigma}$ as a function of N (in units of σ). Red is based on p_s blue is based on p'_s (CL_s).

Summary

- ATLAS recommends to use the profile likelihood as a test statistic
- Toy MCs should be generated (unless the asymptotic is completely verified)
- Use toy MCs to derive the upper limit, the median and the error bands.
- Constrain the upper limit by a power of 16%, the limit cannot go below the -1 sigma band generated by bg-only experiments
- Do not hide results, show the full observed upper limit.
- Use the asymptotic to verify the sense of your results
- Calculate also CLs (for the time being) to compare with TEVATRON (can be easily done with asymptotic)
- Validate with an independent code (for the time being)