Limit Computation Using Asymptotic in CMS

UMN HEP Meeting

Tambe E. Norbert

norbert@physics.umn.edu

September 29, 2014

Outline

1. Motivation
2. Statistical Formalism
3. Test Statistics Distribution
4. Systematics
5. Application Example

Motivation

- Quick computation of limits.
- Study systematic effects.
- Extract experiment sensitivity without generating any Toy Monte Carlo.
- Quantify statistical significance of a signal.

Data Compatibility?

$$p - \text{value} = \begin{cases} 5.7 \times 10^{-7}, & \Rightarrow \text{Reject Bkg-Only Hypo} \\ < 0.05, & \Rightarrow \text{Exclude Sig Hypo} \end{cases}$$

Statistical Test Formalism

Suppose a Poisson, $\mathcal{P}(n)$ of n observed events with expected mean value $E[n] = \mu s + b$ with systematics θ then;

ProfileLikelihood and ratio

$$\mathcal{L}(\mu, b, \theta) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \cdot \mathcal{L}_\theta(\theta)$$

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$

Test Statistics

$$t_\mu = \begin{cases} -2 \ln \lambda(\mu), & \hat{\mu} \leq \mu \\ 0, & \hat{\mu} \geq \mu \end{cases}$$

Test Statistics Distribution

Using the test statistics t_μ, We "find" the pdf of the test statistics $f(t_\mu|\mu)$ assuming a given hypothesis μ.

Asymptotic Method

In Asymptotics, an analytic function of $f(t_\mu|\mu)$ through approximation(Walds).

$$f(t_\mu|\mu') = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}(\sqrt{t_\mu}+\frac{\mu-\mu'}{\sigma})^2\right)+\exp\left(-\frac{1}{2}(\sqrt{t_\mu}-\frac{\mu-\mu'}{\sigma})^2\right)\right]$$

HybridNew Method

Through Monte Carlo or MCMC computations or toy experiments (frequentest) one computes $f(t_\mu|\mu)$ after extracting $f(t_\mu|\mu)$ using Bayesian probability methods.

Computing Probabilities

Using $f(t_\mu|\mu)$, the probabilities (p-values) are computed as:

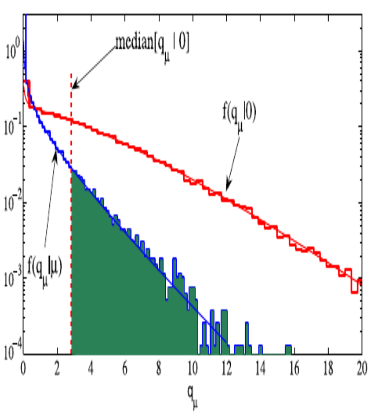
p-values

$$p_u = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu) dt_\mu$$

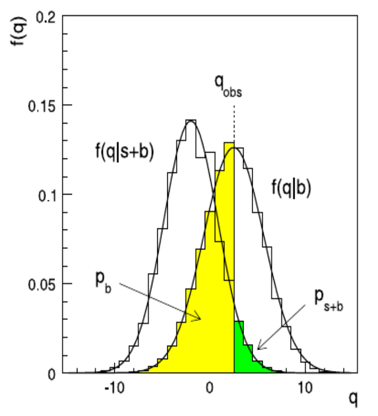
In the typical case where $\mu = 1$ or $\mu = 0$ for $s + b$ or b only hypothesis $t = -2\ln(\frac{\mathcal{L}_{s+b}}{\mathcal{L}_b})$ (Tevatron style)

CL_s Technique

$$CL_s = \frac{p_{s+b}}{1 - p_b} = \frac{CL_{s+b}}{CL_b}$$



median[q_μ | 0]
$f(q_\mu|0)$
$f(q_\mu|\mu)$
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0 2 4 6 8 10 12 14 16 18 20
q_μ



f(q)
q_{obs}
f(q|s+b)
f(q|b)
p_b
p_{s+b}
0.2
0.15
0.1
0.05
0
-10 0 10
q

Systematics

Systematics in Asymptotic

From the definition of $\mathcal{L}(\mu, b, \theta)$, systematics are introduced right on. The approximate formula of the $f(t_\mu|\mu)$ analytically computed embeds in it systematics.

However, in HybridNew approach,

Systematics in HybridNew

Integrate out systematics to get $f(t_\mu|\mu)$.

$$f(t) = \int f(t|\theta)\pi(\theta)d\theta$$

with $\pi(\theta) \propto \mathcal{L}_\theta(\theta)\pi_0(\theta)$ being the prior pdf. This is how systematics are handled by the HybridNew approach.

Application Example

Analysis Result

SM Background/GMSB Signal	Count
Total SM background (b)	1.005 ± 0.001
Data (n_{obs})	1.00
Signal (s)	
GMSB(SPS8) ($\Lambda = 180$ TeV, $c\tau = 500$ mm)	2.341
GMSB(SPS8) ($\Lambda = 180$ TeV, $c\tau = 1000$ mm)	4.585
GMSB(SPS8) ($\Lambda = 180$ TeV, $c\tau = 2000$ mm)	5.704
GMSB(SPS8) ($\Lambda = 180$ TeV, $c\tau = 3000$ mm)	5.386
GMSB(SPS8) ($\Lambda = 180$ TeV, $c\tau = 6000$ mm)	4.096
GMSB(SPS8) ($\Lambda = 180$ TeV, $c\tau = 12000$ mm)	2.772

Application Example

Sources of Systematics

Source	Uncertainty(%)
Photon energy scale	$< 3.0\%$
Jet energy scale	$< 0.05\%$
Jet energy resolution	$< 1.90\%$
PDF uncertainty	$< 1.70\%$
MET resolution	$< 2.8\%$
Signal Eff $\times$ Acceptance	$< 10\%$
ECAL time uncertainty	$< 5.0\%$
Background estimation uncertainty	$\leq 20.0\%$
Luminosity	$< 2.2\%$

Delayed Photon Upper Limit

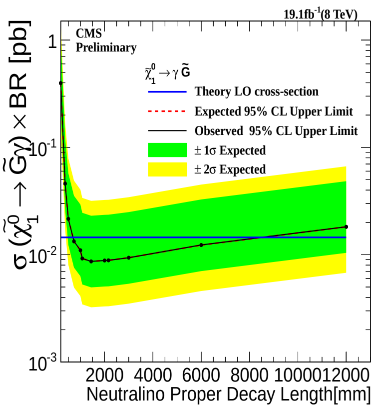


Figure : Systematics Inc.

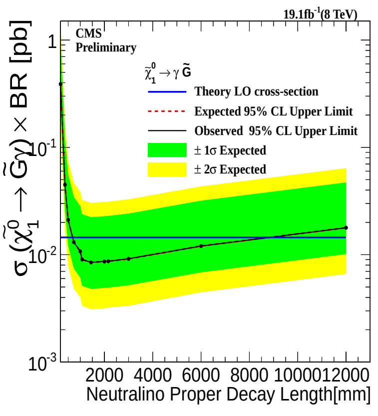
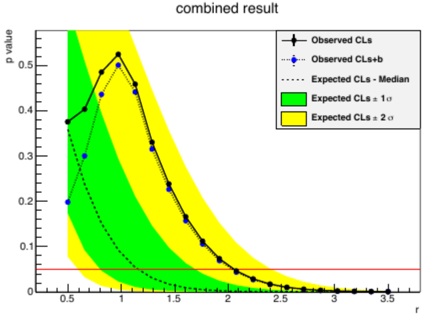
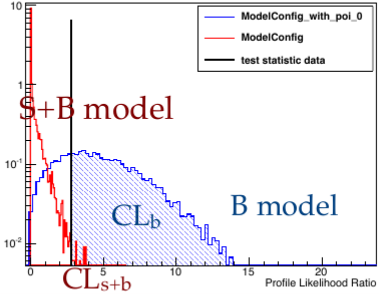
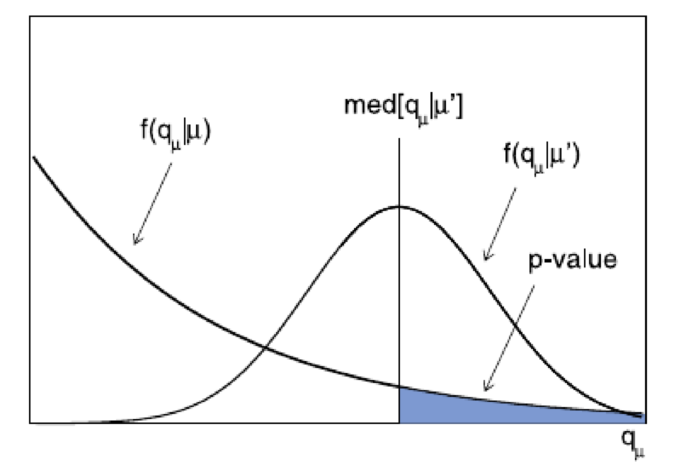


Figure : No Systematics

Limits from P-values






med[q_μ|μ']
f(q_μ|μ)
f(q_μ|μ')
p-value
q_μ

Asymptotic "Asimov" Dataset

"Asimov" Data"

- Its an alternative way to obtain deviations from MLE ($\hat{\mu}$).
- Is a Toy Monte Carlo data (artificial data set),
- Used to estimate the deviations; σ from the median.
- Also used to evaluate the value of $\hat{\mu}$ and $\hat{\theta}$
- Obtaining the true parameters of the MLE.

References

- Glen Cowan et al, *Asymptotic formula for likelihood-based tests of new physics*, Eur.Phys.J.C (2011)71:1554
- A.L. Read, J.Phys. G 28, 2693 (2002)
- A. Wald. *Tests of statistical hypothesis concerning several parameters when the number of observations is large.* Trans. Am. Maths. Soc 54(3), 426-482 (1943)
- Jose Ocariz *Probability and Statistics for Particle Physicists*, arxiv:1405.3402v1
- CERN Summer School: https://indico.cern.ch/event/117033/material/slides/0?contribId=25