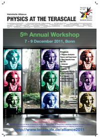
About Interpretation of CMS Search Results

Felix Frensch

KIT - Institut für Experimentelle Kernphysik
5th Annual Workshop of the Helmholtz Alliance "Physics at the Terascale",
7-9 Dec 2011, Bonn (Germany)

8.12.2011



"Statistics is HARD"

- Bob Cousins

Outlook

- Introduction to the problem - counting experiment
- Different approaches
  - Treatment of systematic uncertainties
  - Upper limit calculation
- The CL_s technique - motivation
- Expected upper limit - calculation
- Upper limits observed and expected - interpretations on an example

Scope of work - Counting experiment

The probability to find n events follows a **Possion distribution**.
$\longrightarrow$ Likelihood function for background model:

$$P(n|b(\vec{\nu})) = Poisson(n|b(\vec{\nu})) \cdot \pi(\vec{\nu})$$

$\longrightarrow$ Likelihood function for signal + background model:

$$P(n|s(\vec{\nu}) + b(\vec{\nu})) = Poisson(n|s(\vec{\nu}) + b(\vec{\nu})) \cdot \pi(\vec{\nu})$$

- n is the observed number of events
- $\vec{\nu}$ are systematic uncertainties, so called nuisance parameters
- π are the distributions for the nuisance parameters $\vec{\nu}$
- s/b is the "true"/predicted number of signal/background events, both dependent on nuisance parameters

TASK: Calculating the upper limit on the signal s_{UL}

Bayesian way

1. Use Bayesian-Theorem to calculate the posterior $P(s(\vec{\nu}) + b(\vec{\nu})|n)$:

$$P(s(\vec{\nu}) + b(\vec{\nu})|n) = \frac{P(n|s(\vec{\nu}) + b(\vec{\nu}))\delta(s)\delta(b)\delta(\nu)}{\iiint P(n|s(\vec{\nu}) + b(\vec{\nu}))\delta(s)\delta(b)\delta(\nu)\ db\ ds\ d\nu}$$

- Choice of the priors δ is up to the "degree of believe" or to other informations about the parameters

  $\Rightarrow$ **Main concern of using Bayesian**

2. Marginalizing the background b and other nuisance parameters $\vec{\nu}$:

$$P(s|n) = \iint P(s(\vec{\nu}) + b(\vec{\nu})|n)\ db\ d\nu$$

3. Extract the upper limit s_{UL} (for 95% Confidence Level):

$$1 - \alpha = \int_{-\infty}^{s_{UL}} P(s|n)\ ds = 0.95$$

Hybrid and frequentist ways - Choice of test statistic

LEP and TEV style

- Starting from likelihood ratios:

$$q = -2ln\left(\frac{P(n|s+b)}{P(n|b)}\right)$$

- In case of no nuisance: The **Neyman-Pearson Lemma** proofs this one to be to be the best test statistic
- Details in the Backup

LHC style

- Using the Profile Likelihood

$$q = -2ln\left(\frac{P(n|s+\hat{\hat{b}})}{P(n|\hat{s}+\hat{b})}\right)$$

- $\hat{\hat{b}}$ in the numerator maximizes P for specific s
- $\hat{s}$ and $\hat{b}$ maximize the denominator for all s
- Allows for approximation using **Wilks and Wald theorems** and therefore for much faster computing

- Possibility for pure Frequentist or Hybrid treatment of parameters

- Definition of Hybrid:
  - **Bayesian** treatment of the nuisance parameters
  - **Frequentist** treatment of parameter of interest s

1. Decide what test statistic q to use (LEP, TEV or LHC style)
2. Construct it, here using **Profile Likelihood** and pure frequentist treatment of parameters

$$q = -2ln\left(\frac{P(n|s(\hat{\hat{\nu}}) + \hat{\hat{b}}(\hat{\hat{\nu}}))}{P(n|\hat{s}(\hat{\nu}) + \hat{b}(\hat{\nu}))}\right)$$

   - $\hat{\hat{b}}$ and $\hat{\hat{\nu}}$ in the numerator are maximizing P for specific s
   - $\hat{s}$, $\hat{b}$ and $\hat{\nu}$ are maximizing the denominator for all s
   - This is called "profiling"; nuisance parameters effectively eliminated.

3. Build probability density function for background $f(q|b)$ and signal $+$ background $f(q|s+b)$ by generating Monte Carlo pseudo-data

4 Get the observed value for the test statistic q_{obs} for given s

5 Decide which technique to use for calculating upper limit, here using CL_{s+b}:

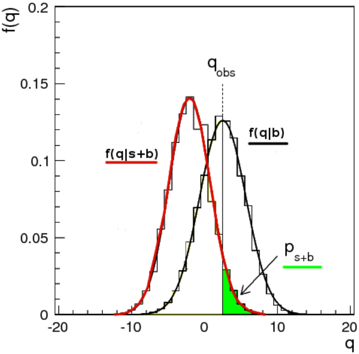
$$p_{s+b} = P(q \geq q_{obs}|s+b) = \int_{q_{obs}}^{\infty} f(q|s+b)\; dq$$

6 Signal + background model is excluded with 95% Confidence Level, if:

$$p_{s+b} < \alpha = 0.05$$

7 Upper limit s_{UL}:

$$max(p_{s+b}) \leq 0.05$$



CL_s

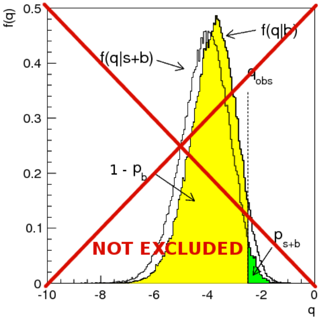
- **Example:** $b = 4; \quad n = 0;$
- **Statistical meaning:** Observation of a downward fluctuation of the background $n < b$
- **Problem:** Find a reasonable upper limit
- **Risk:** Exclusion of the $s + b$ model, although no sensitivity for it!
  - CL_{s+b} upper limit: $s_{UL} = -0.25 \Rightarrow$ even $s = 0$ is excluded!
- **Solution:** Putting the sensitivity into the exclusion procedure
- The CL_s method defines a signal model as excluded, if one finds:

$$CL_s = \frac{p_{s+b}}{p_b} < \alpha = 5\%$$

  - CL_s upper limit: $s_{UL} = 3.23$

CL_s - conclusions

- Saved from **spurious exclusion**
- Therefore "worse" limits, because of overcoverage
  (confidence of excluding > 95%)
  - ⇒ "Worse" limits in the sense of physically better defendable!!!



- CL_s is a conservative, well-working method!

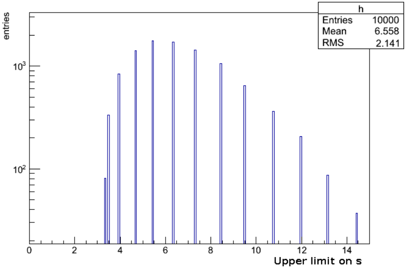
Expected limit

1. Generate a large number of MC toys for the background only hypothesis; $n = Poisson(b(\vec{\nu}))$

- Treating of nuisance parameters, when running MC toys for *b*-only model is nontrivial:
  - B **Bayesian** approaches use predefined distribution
  - F **Frequentist** approaches extract nuisances central values from fitting *b*-only model to data
  - F effectively measuring systematics in data
  - F making the expected limit biased towards the measured one

2 Follow the selected exclusion procedure using the toys as real data

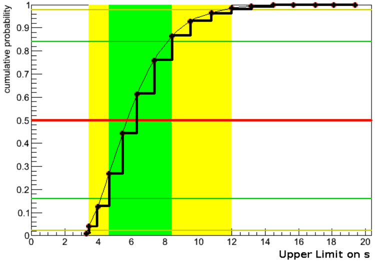
3 For each generated pseudo-data samples find a certain s_{UL}



(Left side) example:

- $s = 1$
- $b = 5$
- no systematic errors

4. Plot the cumulative probability
5. Find the values of s_{UL} where the probability crosses
   - 50% (median expected) quantile
   - 16% / 84% ($\pm 1\sigma$-band; 68%) quantile
   - 2.5% / 97.5% ($\pm 2\sigma$-band; 95%) quantile



$\Rightarrow$ median expected limit: $s_{UL} = 6.3$

$\Rightarrow$ $\pm 1\sigma$-band: $s_{UL} \in [4.7, 8.4]$

$\Rightarrow$ $\pm 2\sigma$-band: $s_{UL} \in [3.5, 12.0]$

$\Rightarrow$ reflects detector sensitivity to the given signature

$\Rightarrow$ $\pm 1\sigma$ / $\pm 2\sigma$-bands shows consistensy between data and expectation

Example - different approaches

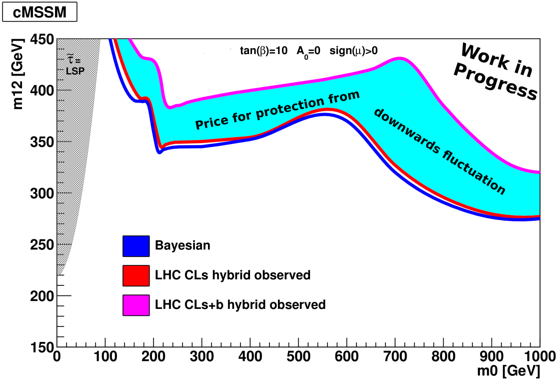
- This is a pure academically study and **NOT** meant as official result!
- Crabbed b and n from CMS PAS SUS-11-010
- Made up s and $\vec{\nu}$ for illustration propose
- 1.4 ± 0.7 background events predicted; 0 observed
- signal s different in each point using the CMSSM SUSY model

On the right side the nuisance parameter $\vec{\nu}$ are specified.

uncertainty	distribution	value
luminosity	logNormal	6%
signal	logNormal	12%
background	logNormal	50%
jet energy scale	logNormal	7.5%
HT Trigger	logNormal	5%
lepton trigger	logNormal	5%
lepton efficiency	logNormal	3%

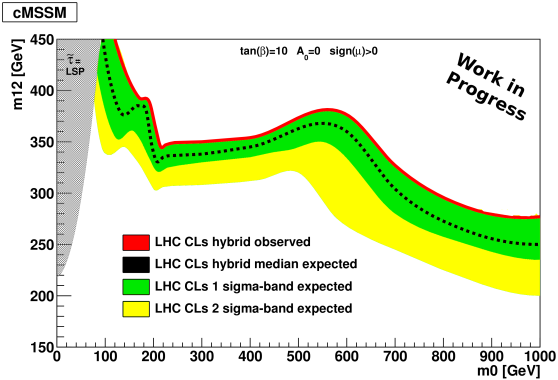
Calculation of the upper limits and extrapolation to the **CMSSM SUSY model**:

Method	s_{UL}
Bayesian	3.16
LHC CL_{s+b} freq	1.46
LHC CL_{s+b} hybrid	1.71
TEV CL_{s+b}	1.66
LEP CL_{s+b}	1.66
LHC CL_s freq	3.24
LHC CL_s hybrid	2.91
TEV CL_s	3.05
LEP CL_s	3.05



- CL_{s+b} excludes more than CL_s due to downward fluctuation
- CL_s agrees good with Bayesian

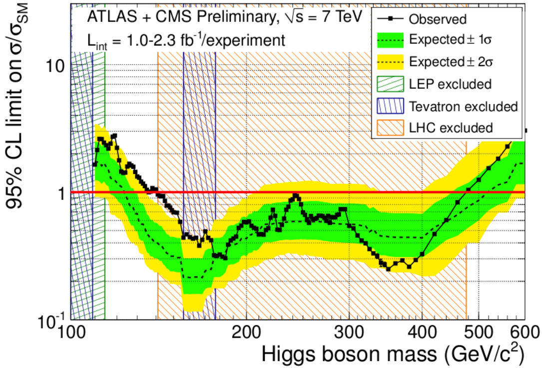
2D plot - **CMSSM SUSY points** m_0:$m_{1/2}$:A_0:$tan(\beta)$:$sign(\mu)$ under the observed line are excluded on 95% Confidence Level



- Observed limit better than expected limit $\Rightarrow$ Maximum downward fluctuation has been observed $0 = n < b$
- The observation is within 1σ (68%) of the background prediction
- The upper limits of the $\pm 1\sigma$ / $\pm 2\sigma$-bands don't have to be equal - here because of the relative small b. For larger statistics this would change.

Higgs Exclusion Plot

1D plot - **SM-Higgs mass** under the red horizontal line are excluded on 95% Confidence Level



- At roughly 119 GeV: $> 2\sigma$ difference from the expected limit and above the red horizontal line - possible excess
- At roughly 140-150 GeV Higgs mass: $> 2\sigma$ deviation from the expected limit, still SM-Higgs is excluded there, but there could be something different
- Consider statistical fluctuations as possible explanation for deviations

Conclusion

- Statistic approaches are nontrivial
- There are different ways refering to the same problem
- Treating systematic uncertanties is ambiguous
- CL_s is used as a reference in HEP
- Expected limit shows the consistensy between data and background only hypothesis
- Exclusion plots deliver many informations, but needs right interpretation

Backup - choice of nuisance paramter distributions

- General way is to use the known/measured distribution of the nuisance paramter, but often one only knows a pair of numbers: mean and width
- Lognormal or Gamma are recommanded - feels more naturally for positive nuisance parameters
- Gaussian truncated to 0 or higher (for positive nuisance parameters) could lead to issues (improper posterior) `http://www.physics.ucla.edu/~cousins/stats/cousins_lognormal_prior.pdf`

Backup - choice of test statistics - LEP

- Nuisance Parameters are treated in a Bayesian way, but do not enter the test statistic

$$q = -2ln\left(\frac{P(n|s,b)}{P(n|b)}\right)$$

- If systematics occur: using to the Bayesian posterior

$$\rho(\nu|\tilde{\nu}) \sim p(\tilde{\nu}|\nu) \cdot \delta(\nu)$$

to modify the $s(\nu)$ and $b(\nu)$ distributions before tossing pseudo MC by drawing random numbers from $\rho(\nu|\tilde{\nu})$

Backup - choice of test statistics - TEV

- Nuisance parameters are treated in a Bayesian way, tossing of pseudo data remains the same, but nuisance parameters additionally enter the test stastic

$$q = -2ln\left(\frac{P(n|s,\hat{\hat{b}},\hat{\hat{\nu}})}{P(n|\hat{b},\hat{\nu})}\right)$$

- If systematics occur: using the Bayesian posterior

$$\rho(\nu|\tilde{\nu}) \sim p(\tilde{\nu}|\nu)\cdot\delta(\nu)$$

  to modify the $s(\nu)$ and $b(\nu)$ distributions before tossing pseudo MC data by drawing random numbers from $\rho(\nu|\tilde{\nu})$

Backup - choice of test statistics - LHC

- Nuisance Parameters are treated in a frequentist way and enter the test stastic

$$q = -2ln\left(\frac{P(n|s,\hat{\hat{b}},\hat{\hat{\nu}})}{P(n|\hat{s},\hat{b},\hat{\nu})}\right)$$

- If systematics occur: using the frequentist "measurement" $p(\tilde{\nu}|\nu)$ to modify the $s(\nu)$ and $b(\nu)$ distributions
- Finding $\hat{\nu}_b$ and $\hat{\nu}_{s+b}$ which describe the observed data in the best way. Then generating pseudo MC data for $f(q|b,\hat{\nu}_b)$ and $f(q|s+b,\hat{\nu}_{s+b})$
- Wilks theorem could be used now: In the asymptotic regime q is expected to have half a χ^2 distribution with one degree of freedom (for $s+b$ experiments), this is technical much faster, because no need to generate MC

Backup - why elimating systematics before building pdf's

- With nuisance parameters: p_{s+b}-value (p_b-value analogous):

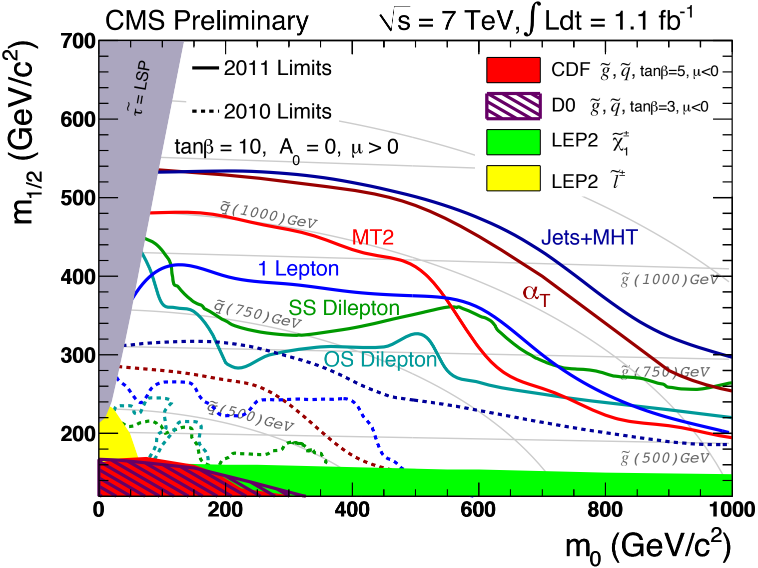
$$p_{s+b} = \int_{q_{obs}}^{\infty} f(q|s+b,\nu)\ dq$$

- In general (excluding the case of profile likelihood, where ν are effectively eliminated), this could lead to value of $s+b$ which would be excluded depending on the value ν - **So when to exclude and when not?**
- Compromises are frequentist or Bayesian treatments of the nuisance parameters:
  - Frequentist approach is to reject $s+b$ model if $p_{s+b} < \alpha$ for the ν that best fits the data
  - Bayesian way of marginalizing the nuisance parameters effectively builds the uncertainty due to ν into the model

Remember: 1.4 ± 0.7 background events predicted $\rightarrow$ 0 observed

Method	s_{UL}	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
Bayesian	3.16	3.16	3.16	4.26	6.95	9.82
LHC CL_s freq	3.24	3.24	3.24	3.44	4.90	7.79
LHC CL_s hybrid	2.91	2.84	2.84	**3.89**	6.61	9.51
TEV CL_s	3.05	3.14	3.14	4.11	6.75	9.51
LEP CL_s	3.05	3.14	3.14	4.11	6.75	9.51

Backup - Official CMS results



Backup - Power Constrained Limits (PCL)

- Statistically no solid foundation for interpreting CL_s, which is a ratio of p-values, as a p-value
- Coverage probability is greater than 95% (remember it's pretty conservative) by an amount which is in general not reported

$\Rightarrow$ In 2010 by Cowan, Cranmer, Gross and Vitells propose an alternative method: Power-Constraint Limits (PCL)

  - Adresses the same problem of **spurious exclusion**

A value of s is excluded, if both of the following conditions are fullfied:

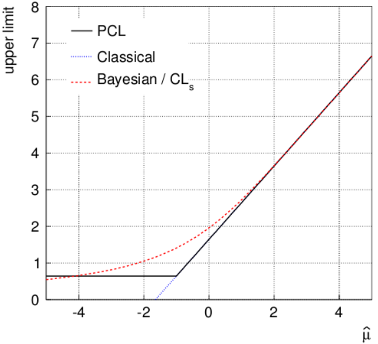
- $p_{s+b} < \alpha$
- sufficient sensitivity: $M_0(s) \geq M_{min}$, with $M_0(s) = P(p_b < \alpha | b)$ and $M_{min} = 0.5(0.16)$

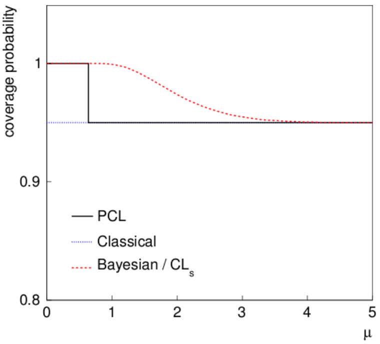
- From $M_0(s) = P(p_b < \alpha | b) = 0.5 \rightarrow$ extract $s_{UL,min}$
- With CL_{s+b} technique $\rightarrow$ get $s_{UL,CL_{s+b}}$

$\Rightarrow$ Power-Constrained Limit calculated via:

$$s_{UL} = max(s_{UL,min}, s_{UL,CL_{s+b}})$$

The converage probability and upper limits for a Gaussian measurement are showed in following plots:





Sources

- http://mschen.web.cern.ch/mschen/Lands/ - LandS Code - tool for statistic calculations
- http://www.physics.ucla.edu/~cousins/stats/cousins_lognormal_prior.pdf - distribution study of nuisance paramters
- http://www.physics.ucla.edu/~cousins/stats/cousins_bounded_gaussian_virtual_talk_12sep2011.pdf - Bob Cousins virtual talk
- http://www.pp.rhul.ac.uk/~cowan/stat_desy.html - Glen Cowan statistic lectures