Limit setting in High Energy Physics

Stefan Schmitt, DESY

- Probability basics
- Limits in HEP, the basic idea
- Systematic uncertainties and background subtraction
- Combining channels
- The CL_s method
- Combining channels and systematic uncertainties
- “Real” limit examples

Probability basics

Probability basics

- Elements of Ω : events, outcomes of an experiment
- Probability of $A \subset \Omega$:

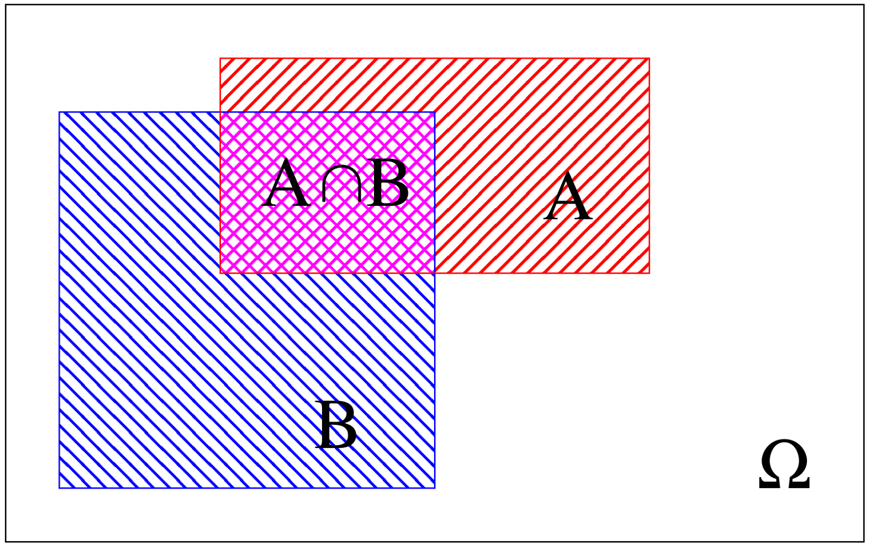
$$0 \leqslant P(A) \leqslant 1, P(\emptyset)=0, P(\Omega)=1$$
$$P(A \cup B)=P(A)+P(B)-P(A \cap B)$$

Example: Poisson distribution

$$P(\{N\})=\frac{e^{-\mu}\mu^{N}}{N!}$$
$$\Omega=\{0,1,2,\ldots\}$$

- Conditional probability of A given B: $$P(A|B)=\frac{P(A \cap B)}{P(B)}$$
- Bayes' law:

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)}$$



Probability densities

- Probabilites on discrete sets: each element has a finite probability
  Example: Poisson distribution $P(\{N\}) = \frac{e^{-\mu}\mu^N}{N!}$
  → Used for event counts $\Omega = \{0, 1, 2, \ldots\}$

- Probability densities: probabilities are defined by integrals
  Example: normal distribution $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
  → Used for systematic errors $\Omega = \mathbb{R}$

$$P(a \leqslant x \leqslant b) = \int_a^b f(x)\,dx$$

Nuisance parameters

- Nuisance: a parameter of a probability density/distribution, not the measurement itself

  Examples:
  - Poisson distribution:

    $$P(\{N\}) = \frac{e^{-\mu}\mu^{N}}{N!}$$

    μ is a nuisance parameter
  - Normal distribution:

    $$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

    μ and σ are nuisance parameters

- Symbol for nuisance parameters: θ

Frequentist/Bayesian probability

- Frequentist view: probabilities describe the outcomes of experiments

  Models have unknown parameters (nuisances). Probabilities (to make an observation) are given as a function of the model parameters

- Bayesian extension: probabilities are also used to describe the "degree of belief" in model parameters.

  → The model parameters (nuisances) themselves can have probabilities assigned.

Bayesian definitions

- $P(A|B)$: Likelihood
  probability of the data A, given the model B
- $P(B)$: Prior
  probability of the model B before looking at the data
- $P(B|A)$: Posterior
  probability of the model B, given the data and the prior knowledge

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

- The normalisation $P(A)$ is often calculated using the relation

$$P(B|A) + P(\neg B|A) = 1$$

Example on Bayes' law

- Consider a disease and test for disease
- 0.1% of population have the disease
- If one has disease, the test is positive at 99% probability
- If one has no disease, the test is positive at 1% probability

- Given a positive test, what is the probability to have the disease?

Example on Bayes' law

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)}$$

- Consider a disease and test for disease
- 0.1% of population have the disease Prior $P(B)=0.001$
- If one has disease, the test is positive at 99% probability: likelihood $P(A|B)=0.99$
- If one has no disease, the test is positive at 1% probability $P(A|\neg B)=0.01$

- Given a positive test, what is the probability to have the disease?
  Normalisation: $P(A)=P(A|B)P(B)+P(A|\neg B)P(\neg B)=11\%$
  Posterior: $P(B|A)=P(A|B)P(B)/P(A)=9\%$

Limits in high energy physics:
the basic idea

Probabilities in high energy physics

- Probability: predict number of events given the theory (parameters) and the experimental setup

- But we want to know what a specific observation tells about the theory

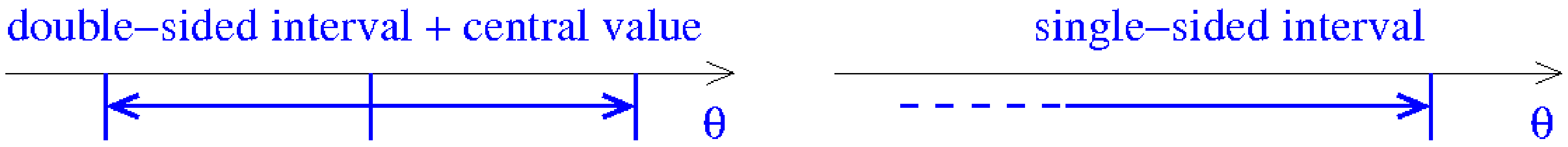
- Frequentist: give for each theory the probability of the observation (no probability for theory, only for observ.)

- Bayes: assign probability (degree of belief) to theories
- High energy physics: make use of both views, but:
  Preference for frequentist, in particular for discoveries

Confidence intervals, Limits

- Confidence intervals tell about parameters of the theory (nuisances)
- Confidence level (CL): associated probability
  - Note: different meaning of CL Frequentist/Bayesian

    Frequentist: $CL \sim P(obs|\theta)$ Bayesian: $CL \sim P(\theta|obs)$
- Double-sided: measurement (usually CL=68%)
- Single-sided: limit (often CL=95%)



$\theta = \mu \pm \sigma$[at 68% CL]

$\theta \leq \theta_{max}$ at 95% CL

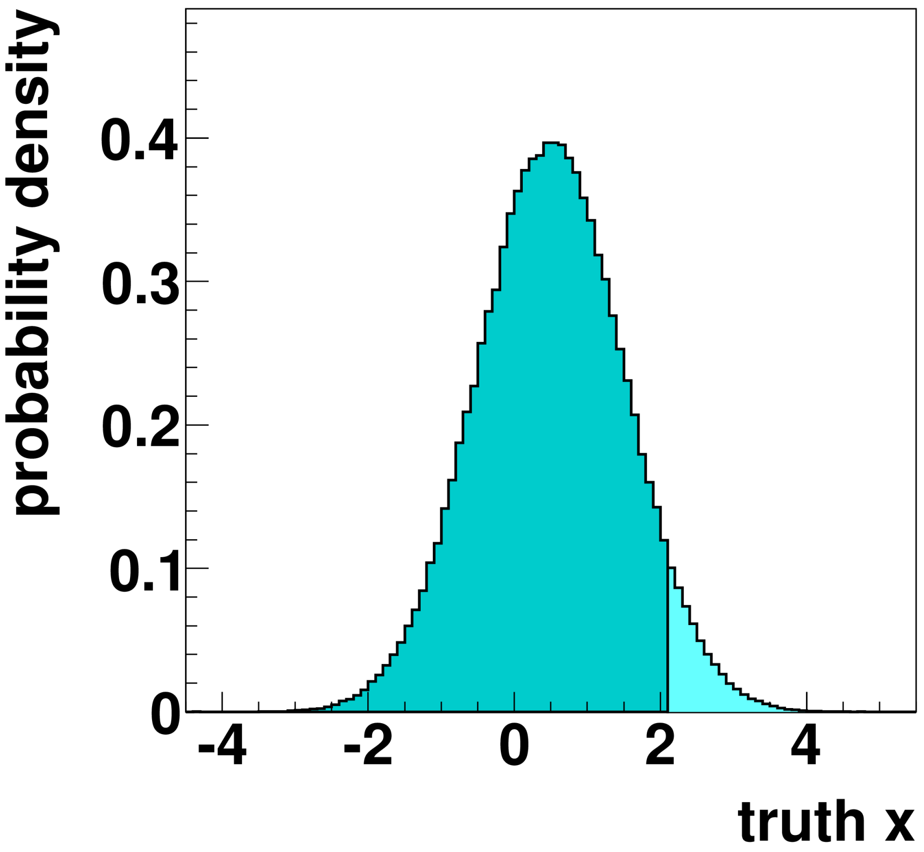
Limits in the Gaussian approximation

- Measurement of a quantity

  $x = \mu \pm \sigma$

- Bayesian interpretation: the truth has a Gaussian probability density around μ with width σ
- Find limit from error function: excluded area has probability 1-CL=5%

- Not correct for low statistics (Poisson)
- Frequentist interpretation?



Limit or measurement?

- Question: when to quote a measurement, when to quote a limit?
- High energy physicist: measurement at ~3σ
- Set limit otherwise
- But: deciding on the statistical procedure after having seen the data is problematic

→ Feldman-Cousins (talk by F. James this morning)

- For the purpose of this lecture, these problems are not discussed

Assumption: we are going to set a limit, no matter what the data say

Limits for low statistic samples

- Imagine a counting experiment, small number of events

$$N_{\text{obs}} = \{0,1,2,\ldots\}$$

- Have to use Poissons law

$$P(N = N_{\text{obs}}) = \frac{e^{-\mu}\mu^{N_{\text{obs}}}}{N_{\text{obs}}!}$$

$$P(N \leqslant N_{\text{obs}}) = \sum_{n=0}^{N_{\text{obs}}} \frac{e^{-\mu}\mu^{n}}{n!}$$

$$= e^{-\mu}\left(1 + \mu + \frac{\mu}{2} + \ldots\right)$$

- Example:

$$N_{\text{obs}} = 0$$

- What is the 95% CL limit on μ?

$$P(N \leqslant 0) = e^{-\mu}$$

- Find μ such that

$$P(N \leqslant 0) < 1 - CL$$

- Result for $N_{obs}=0$

$$\mu > -\ln(1 - CL)$$

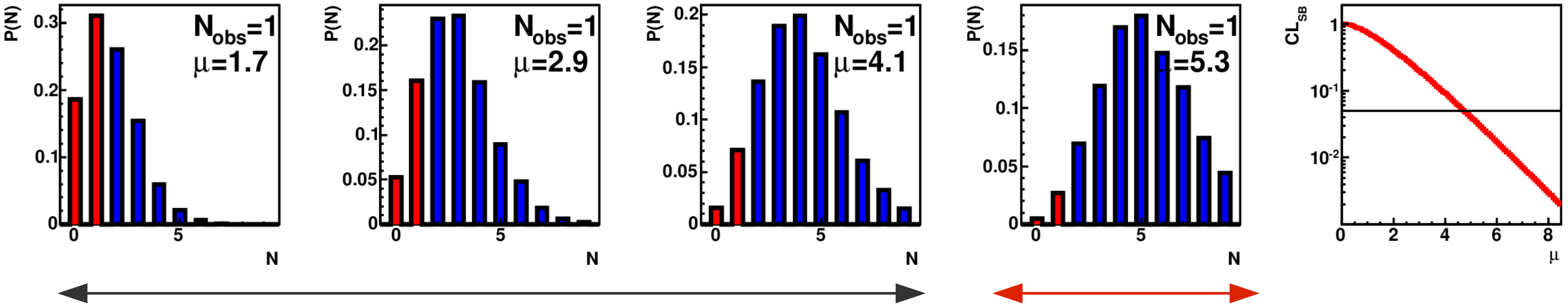
- For $N_{obs}=0$ and $CL=0.95$, exclude $\mu>3$

Frequentist limits (one channel)

- Frequentist limit: “exclude” all models which produce the data at small probability CL_{SB} less than *1-CL* (typically: $CL=0.95$)

$$CL_{SB} = P_{\mu}(N \leq N_{obs}) < 1 - \mathrm{CL}$$

Frequentist limit:
sum (integrate) over observations up to N_{obs}
Repeat for each model



CL_{SB}>0.05, not excluded

CL_{SB}<0.05, excluded

Bayesian limits (one channel)

- Bayesian limit: exclude a set of theories, such that the posterior probability of the excluded theories is 1-CL

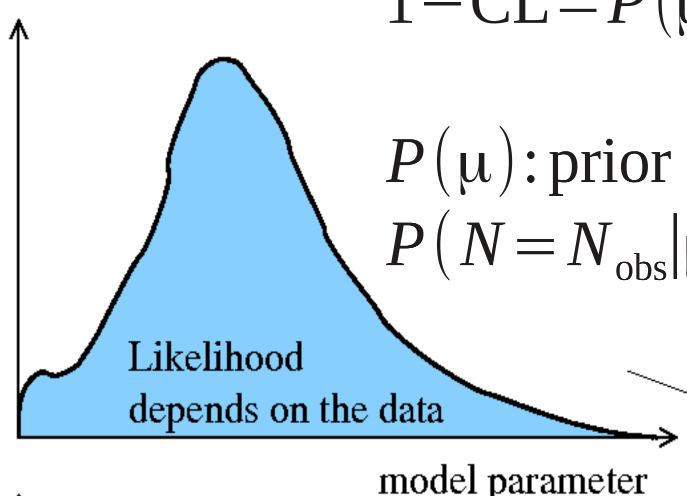
Enumerator: integrate over excluded theories

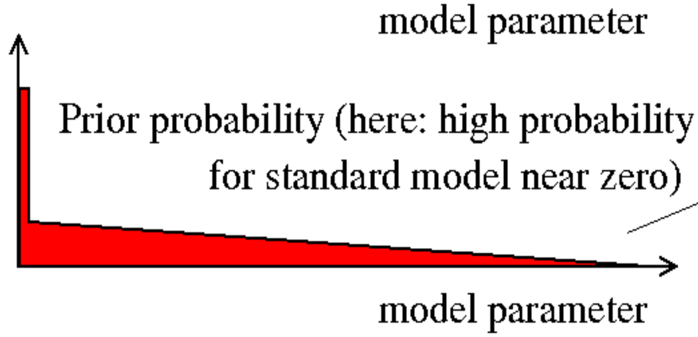
$$1-\mathrm{CL}=P(\mu\geq\mu_0|N=N_{\mathrm{obs}})=\frac{\int_{\mu_0}^{\infty}P(N=N_{\mathrm{obs}}|\mu)\,P(\mu)\,d\mu}{\int_{0}^{\infty}P(N=N_{\mathrm{obs}}|\mu)\,P(\mu)\,d\mu}$$

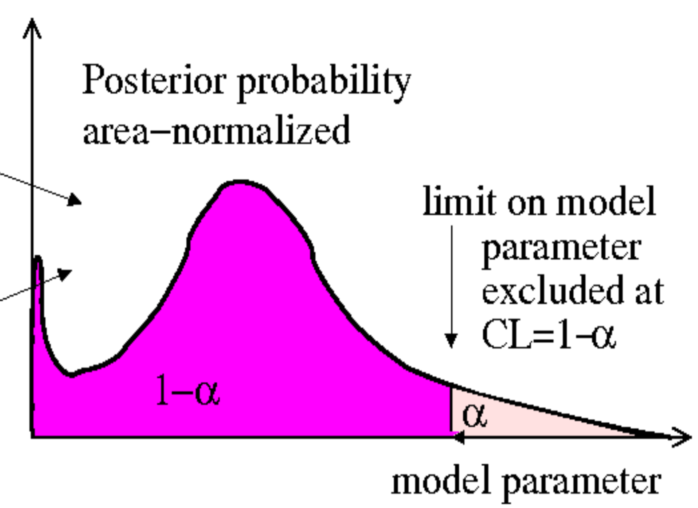
Denominator: integrate all theories (normalisation)

$P(\mu)$: prior probability of the model μ

$P(N=N_{\mathrm{obs}}|\mu)$: Likelihood



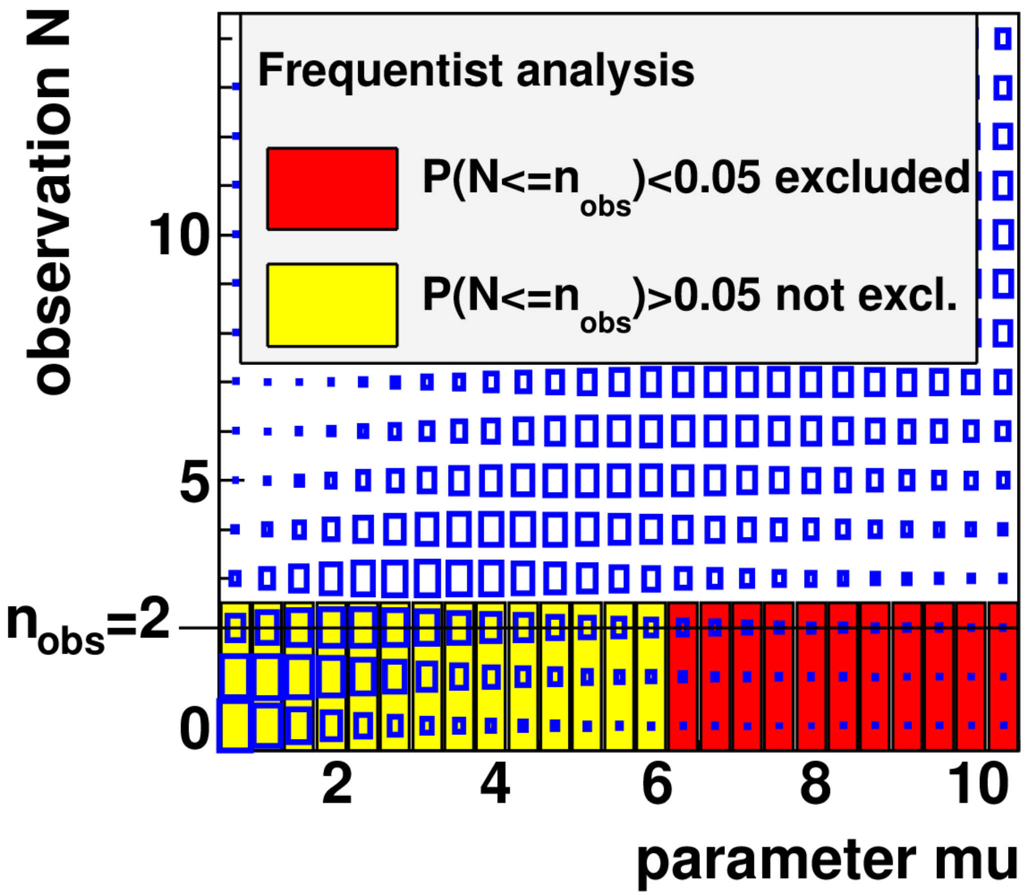




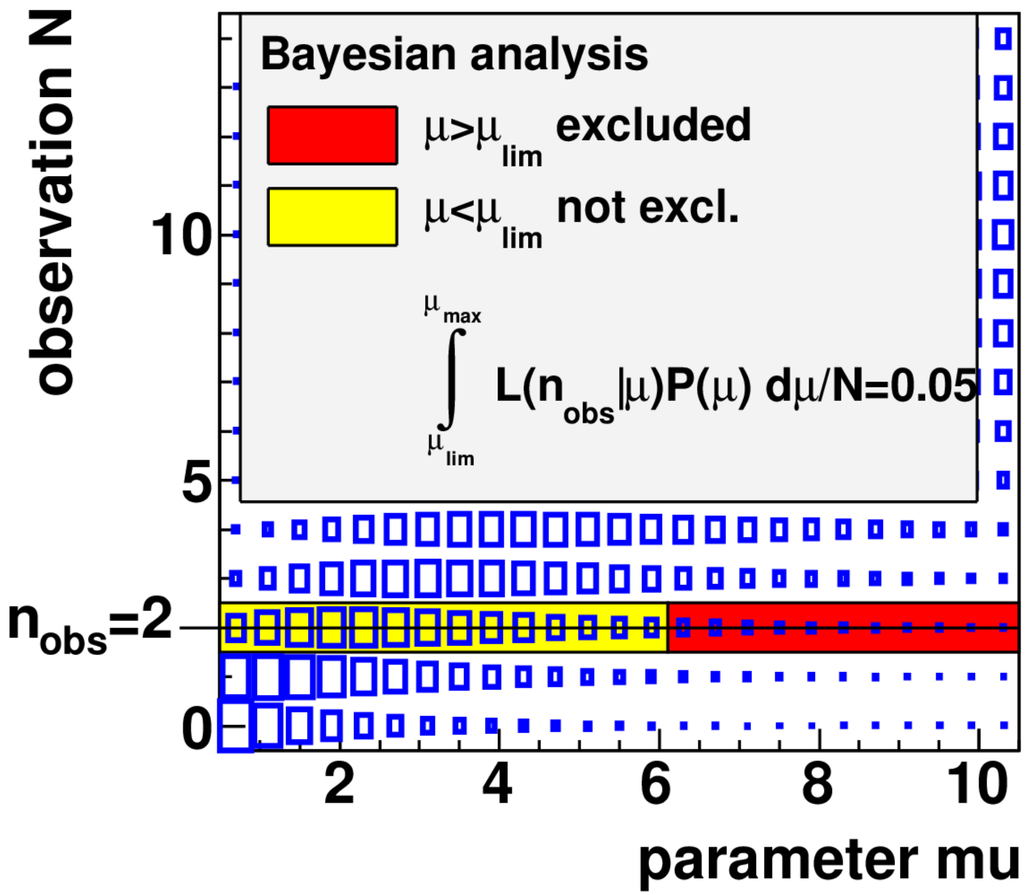
Bayesian limit: integrate over models, fixed N_{obs}

Comparison Limit calculation

- Frequentist
- Test models "μ" one by one



- Bayesian
- Integrate over "μ" to find exclusion interval



Comparison Frequentist/Bayesian

- Frequentist limit tells about the probability of repeated (Gedanken-) experiments
- Calculation is done by summing over possible observations
- Models are tested one by one

Note: dependence on prior is the main weakness of Bayesian methods.

In HEP, Bayesian methods are not used so much.

This talk: Bayesian methods are not discussed in great detail.

- Bayesian limit tells about the model probability
- Calculation is done by integrating over models
- Result depends on prior
- Often used: “flat” prior

$$P(B) = \text{const}$$

- But: this depends on model formulation. For example: “flat” prior in cross section is non-flat in coupling

cross section: $\sigma = k\lambda^4$

probability density: $f(\sigma)\,\mathrm{d}\sigma = 4\mathrm{k}\,\lambda^3 f(k\lambda^4)\,\mathrm{d}\lambda$

Systematic uncertainties
Background subtraction problems

Systematic uncertainties

- Systematic errors: detector effects, hadronisation, etc
- Described by nuisances, usually with given prior distributions
- Bayes: conceptually simple, just integrate over all nuisances (model+systematic effects)

$$1-\mathrm{CL}=P(\mu\geq\mu_0|N=N_{\mathrm{obs}})\propto\int_{\mu_0}^{\infty}d\mu\,P(\mu)\int d\vec{\theta}\,P(\vec{\theta})\,P(N_{\mathrm{obs}}|\mu,\vec{\theta})$$

- Frequentist limits are calculated by “marginalising” (integrating over) systematic parameters, then using Frequentist methods

$$\alpha=P_{\mu}(N\leq N_{\mathrm{obs}})=\int d\vec{\theta}\,P(\vec{\theta})\,P_{\mu}(N\leq N_{\mathrm{obs}}|\vec{\theta})$$

Example with systematic errors

- Consider signal $\mu = L(s+b)$
  L: integrated luminosity
  s, b: signal, background cross sections

  with systematic errors: $L = L_0 \pm \sigma_L$
  $b = b_0 \pm \sigma_b$

- Full probability density has three contributions

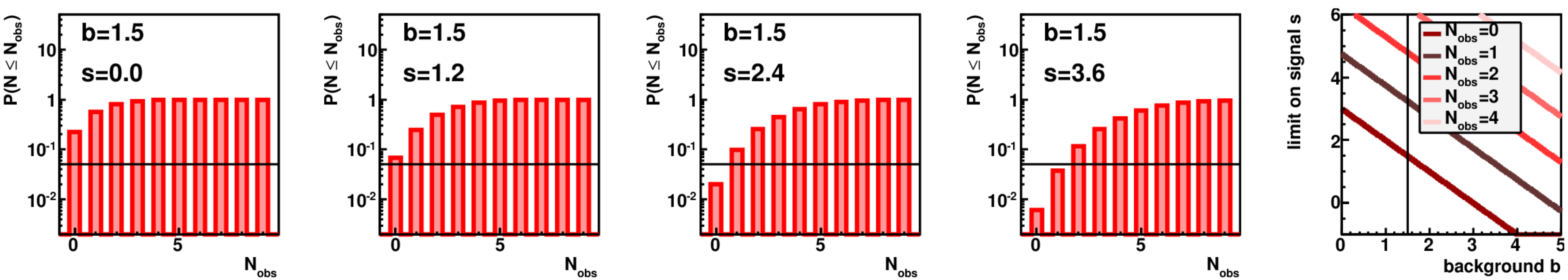
$$P(N|s,L,b) = \underbrace{\frac{e^{-L(s+b)}(L(s+b))^N}{N!}}_{\text{observation}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_L} e^{\frac{-(L-L_0)^2}{2\sigma_L^2}}}_{\text{Prior for syst. error on L}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_b} e^{\frac{-(b-b_0)^2}{2\sigma_b^2}}}_{\text{Prior for syst. error on b}}$$

- Marginalisation: L and b are integrated out

$$P(N|s) = \int dL \int db\, P(N|s,L,b)$$

Example with background

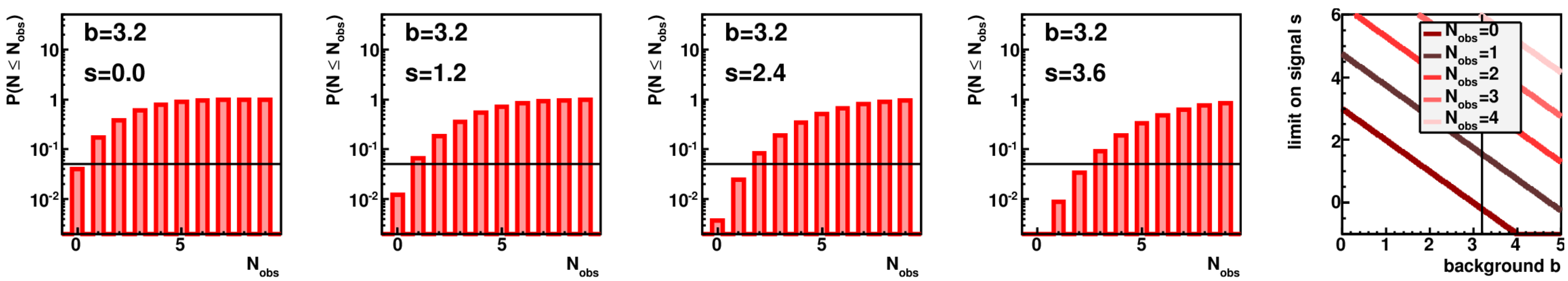
- Example with background (no systematic errors)
- Poisson mean $\mu = s + b$
- Calculate limit on s for a given observation n_{obs} and background b
- Example: $b=1.5$
- Scan signal s
- Set limit $s_{limit} = \mu_{limit} - b$ such that $P(N \le N_{obs}) < 1 - CL = 0.05$
- For $b=1.5$, $N_{obs}=0$: $s_{limit}=1.5$



Strong downward fluctuation

- Example with background (no systematic errors)
- Poisson mean $\mu=s+b$
- Calculate limit on s for a given observation n_{obs} and background b
- Example: $b=3.2$

- Scan signal s
- Set limit $s_{limit}=\mu_{limit}-b$ such that
  $P(N \leq N_{obs})<1-CL=0.05$
- For $b=3.2$, $N_{obs}=0$:
  $s_{limit}=-0.2$???



Strong downward fluctuation

- n_{obs}*=0, b=3.2*
- Limit on s:

$$P(N \leqslant n_{\text{obs}} | s, b) = \exp[-(s+b)]$$
$$\exp[-(s_{\text{limit}}+b)] = 1 - CL = 0.05$$
$$s_{\text{limit}} = 2.996 - b = -0.204$$

- All models with *s>-0.204* are excluded
- Allowed physical region *s≥0*
- The Standard model s=0 is also excluded
- “Empty interval”

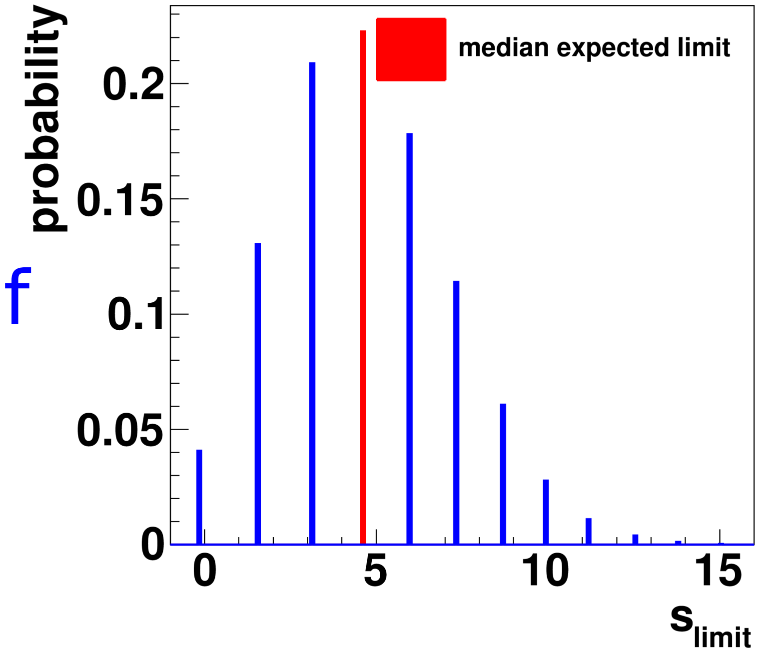
- For *b=3.2* a negative limit is observed with 4% probability → not that unexpected
- Possible solution: Feldman-Cousins
  → talk by F. James this morning

- This talk: more simplified approaches are discussed

The expected limit

- Downward fluctuation: need to quantify “sensititivity”
- For each possible outcome of the experiment (in absence of the signal), calculate the limit
- Define median expected limit and 68% (95%) regions

- Example with b=3.2, no systematics



More realistic examples include systematic errors → distribution is continuous (~Gaussian)

- Expected limit is 4.6
- And 68% of the expected limits are in [1.6,7.3]

How to interpret "unphysical" limits

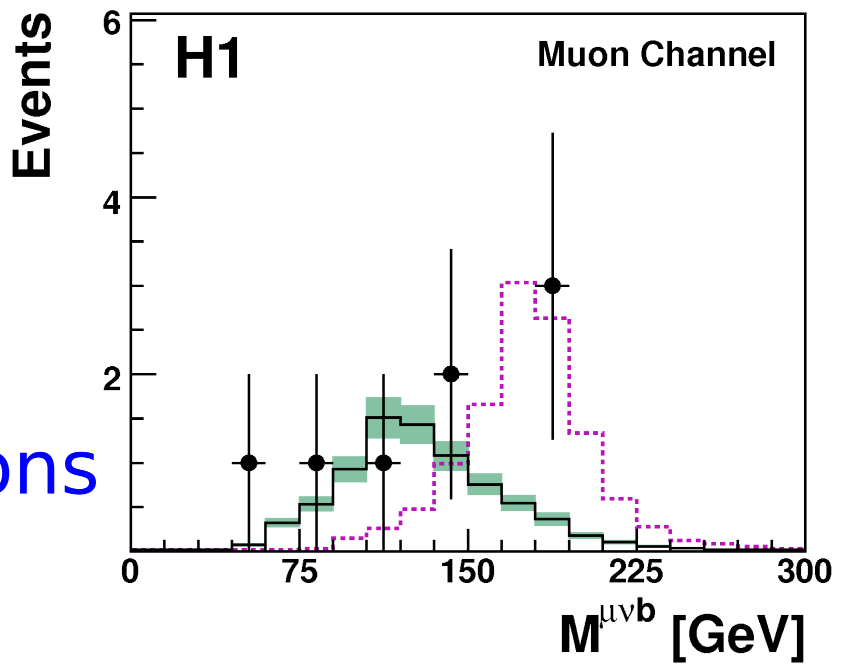
- We observed: s_{obs}=-0.2 (s>-0.2 excluded @95%CL)
- Expected limit is $s_{expected}$=4.6
- 68% range of expected limit is [1.6,7.3]
- Observed limit is much better than expected
  → Downwards fluctuation

- What to write in the publication?
- Power constraint:
  do not quote observed limit, but quote max(s_{obs},$s_{expected}$)
- The "power" of the limit calculation is constrained to $s_{expected}$
- "Standard" method: CL_s
  → explained later in this talk

Combining channels

Combining bins or channels

- Discussed so far: events are counted in a single channel
- More general case: several channels or several bins in one channel
  - Example: mass distribution with N bins (signal/bgr shape) → N channels to be combined
  - 2nd example: "auxillary" selections to measure background



- What is the limit on the total number of signal events, given the shape information in addition to the total number of events?

Combining channels: basic idea

- Bayesian methods: product likelihood of all channels

$$\text{Likelihood} = \prod_{\text{chn}} \frac{e^{-\mu_{\text{chn}}} {\mu_{\text{chn}}}^{N_{\text{obs,chn}}}}{N_{\text{obs,chn}}!} \text{ where } \mu_{\text{chn}} = s_i + b_i$$

  → conceptually simple extension of the 1-dim case

- Frequentist: define “test statistic” X which combines information of several channels, then analyze the probability distribution f(X).

  Properties of X: high X means observation is signal-like, low X means observation is background-like

Choice of the test statistics

- Example: log (likelihood ratio)

$$X = \log \frac{P(N_{\text{obs}} | \text{signal} + \text{bgr})}{P(N_{\text{obs}} | \text{bgr})} = \log L(\text{signal+bgr}) - \log L(\text{bgr})$$

- Other choices are possible, for example: weighted sum

$$X = \sum w_i N_i^{\text{obs}}$$

- weight taken from signal/bgr ratio or something similar

example choice: $w_i = \frac{s_i}{s_i + 2b_i}$ where s_i (b_i): signal (background) in bin i

Note: log of likelihood ratio is a weighted sum:

$$\log(L(\text{signal+bgr}) - \log L(\text{bgr})) \sim \sum_i \underbrace{\log\left(1 + \frac{s_i}{b_i}\right)}_{w_i} N_i^{\text{obs}}$$

Combining bins: analysis procedure

- Channels with background b_i
- Signal described by one unknown parameter s and efficiencies ε_i in each bin

$$\mu_i = s\,\epsilon_i + b_i$$

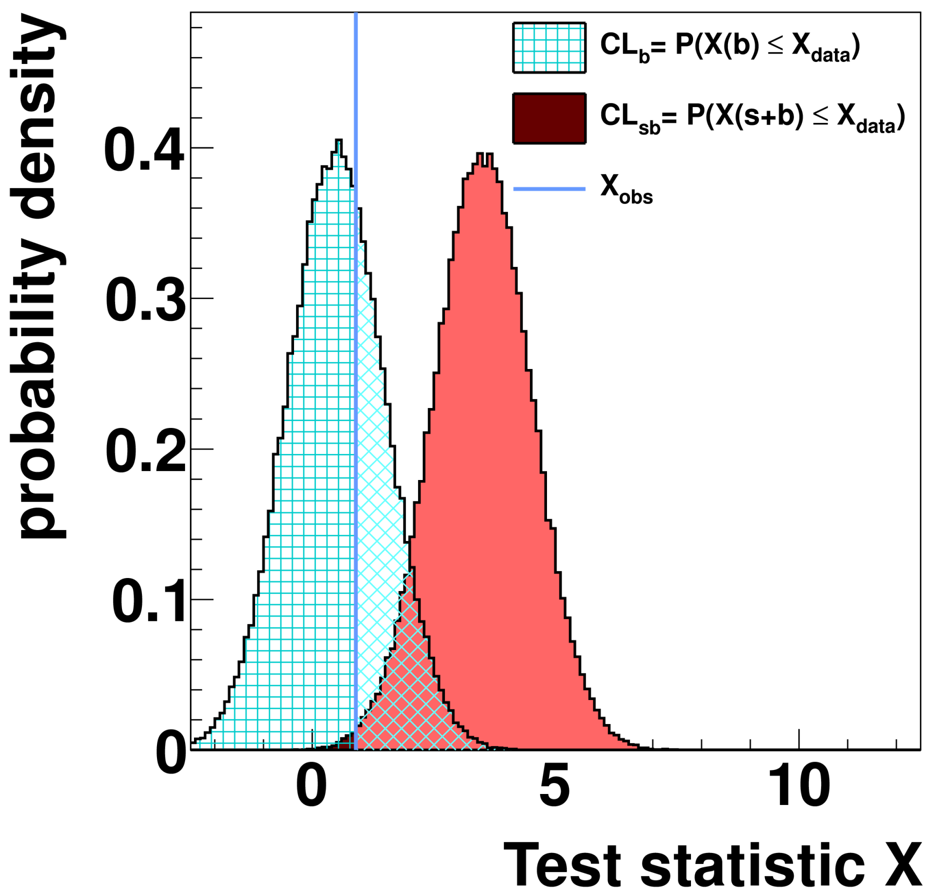
- Choose s and calculate data test statistic

$$X_{\text{data}} = \sum_i \log\left(1 + \frac{s_i}{b_i}\right) N_i^{\text{obs}}$$

- Toy experiments:

$$CL_{sb} = P(X \leqslant X_{\text{data}})$$

- If $CL_{SB} < 1\text{-}CL$ → exclude
- Possible downwards fluctuations → CL_s method



The CL_s method
Also known as “modified Frequentist”

The CL_s method

- Frequentist exclusion limits: exclude if

$$CL_{SB}(s) < 1 - CL$$

- Problem of downward fluctuations: zero-signal model is also excluded

$$CL_{SB}(0) =: CL_B \underbrace{< 1 - CL}_{\text{downwards fluct}}$$

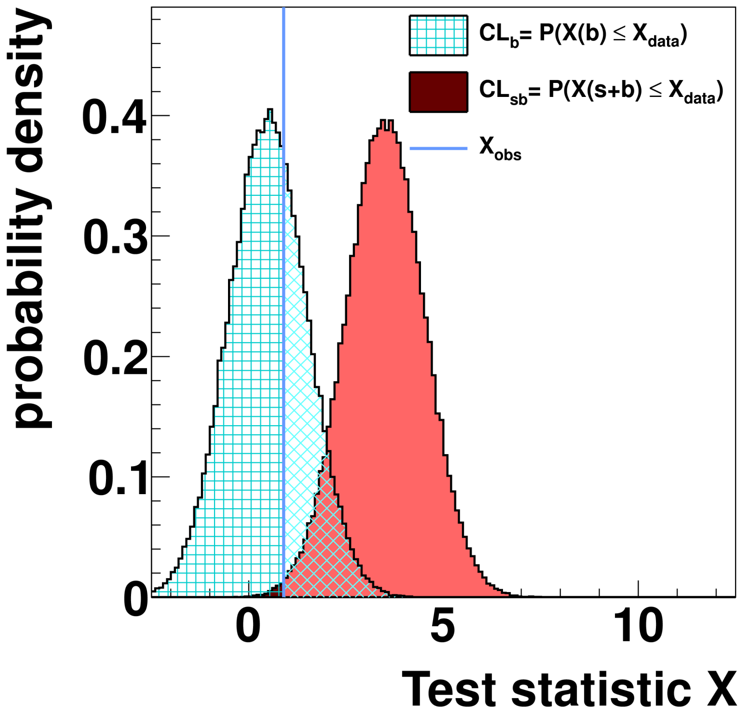
- Idea: "normalise" probability to zero-signal

$$CL_S(s) = \frac{CL_{SB}(s)}{CL_B}$$

- Then:

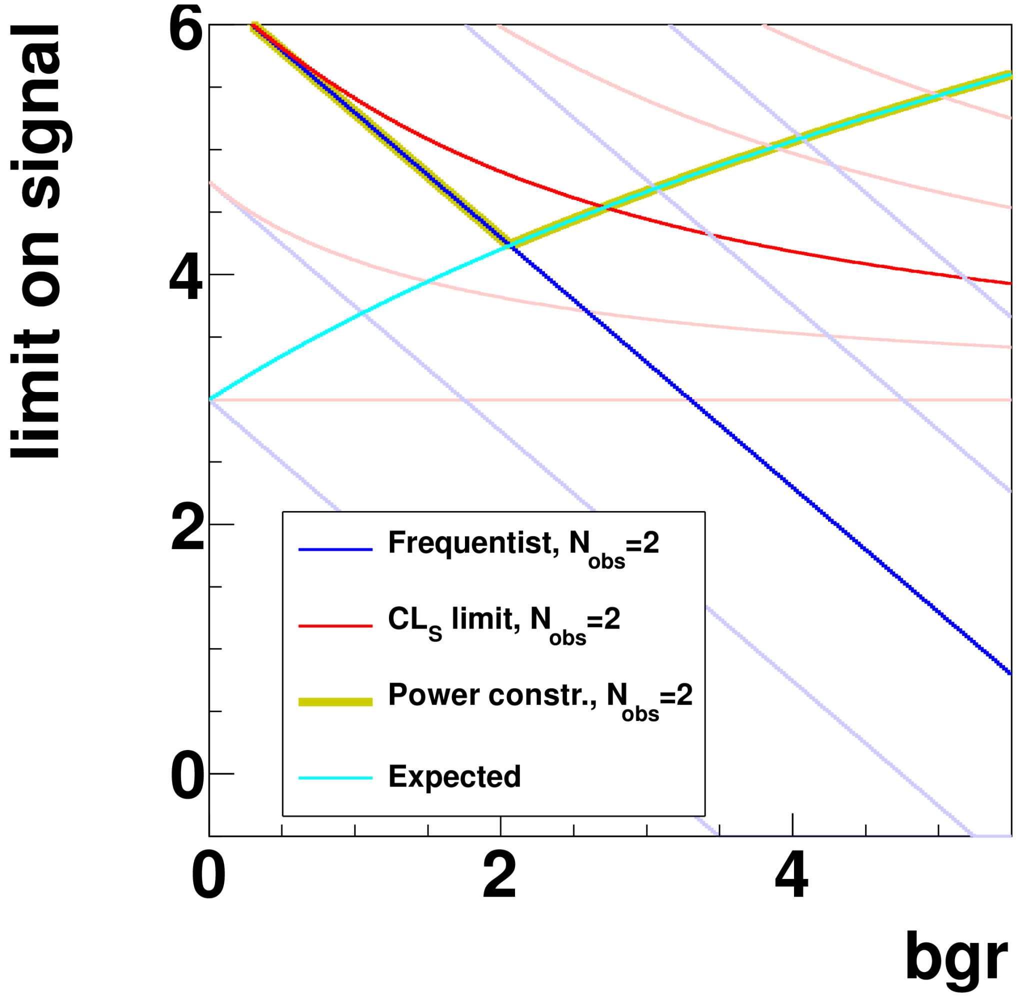
$$CL_S(0) = \frac{CL_{SB}(0)}{CL_B} = 1$$

zero-signal model is never excluded



Comparison CL_s and other methods

- One channel with background $\mu=s+b$
- Compare limits from CL_s, standard frequentist, power constraint (fixed N_{obs}=2)
- Power constraint is most conservative, CL_s in between



Combining channels with systematic errors

Many channels + systematic errors

- HEP problems are of this type
- Bayesian: use N-dim Likelihood + priors for systematic errors and model parameters → limits
- Frequentist: define a “good” test statistic X, then
  - Marginalize systematic errors, calculate confidence levels as a function of model parameters → limits
  - Question: what is a “good” test statistic?

Many channels + systematic errors

- Consider two channels
- Same signal efficiency, same amount of background
- Different amount of background systematics

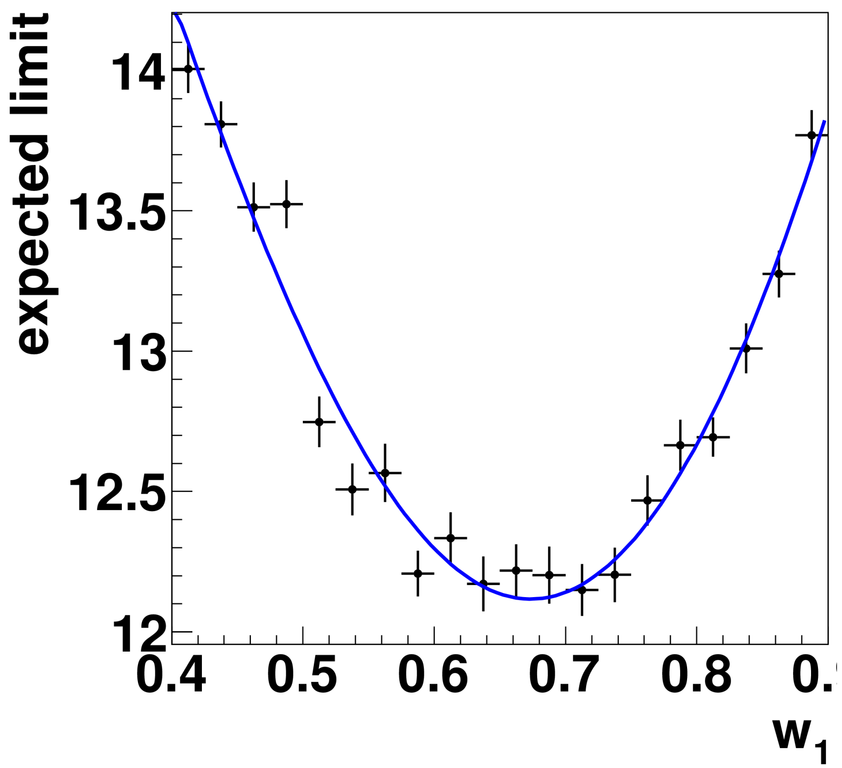
	Channel 1	Channel 2
Efficiency	0.5	0.5
Backgnd	4±0.5	4±3

- Standard log(L) does not know about systematic error → same weight for both channels?

- Test statistic

$X = w_1 N_1 + w_2 N_2$, where $w_2 = 1 - w_1$

- Scan expected limit
- Optimum near w_1=0.7



Choice of test statistic with syst.errors

- No unique method to choose test statistic X
- Requirement: **robustness against systematic errors**
- “Standard” method: **profile likelihood**
  - Use likelihood maximized wrt systematic parameters as test statistic
  - Problem: computational heavy

- Alternative methods exist, for example:

  $X = \sum w_i N_i^{\mathrm{obs}}$

  P. Bock, JHEP 0701 (2007) 080 [arXiv:hep-ex/0405072]

  Weighted sum, bin weights w_i are fixed, and optimized for systematic errors

  Much faster, in practice very similar results to profile likelihood.

Frequentist/Bayesian calculation summary

- Frequentist: calculate

$$CL_{SB}(s)=\int_{X<X_{\text{obs}}} dX \int d\vec{\theta} P(\vec{\theta}) P(X|s,\vec{\theta})$$

- X: profile likelihood or similar
- Exclude if: $CL_S(s)<1-CL$
- Calculation is repeated for many models (scan signal strength s and other par.)

- Bayesian: calculate

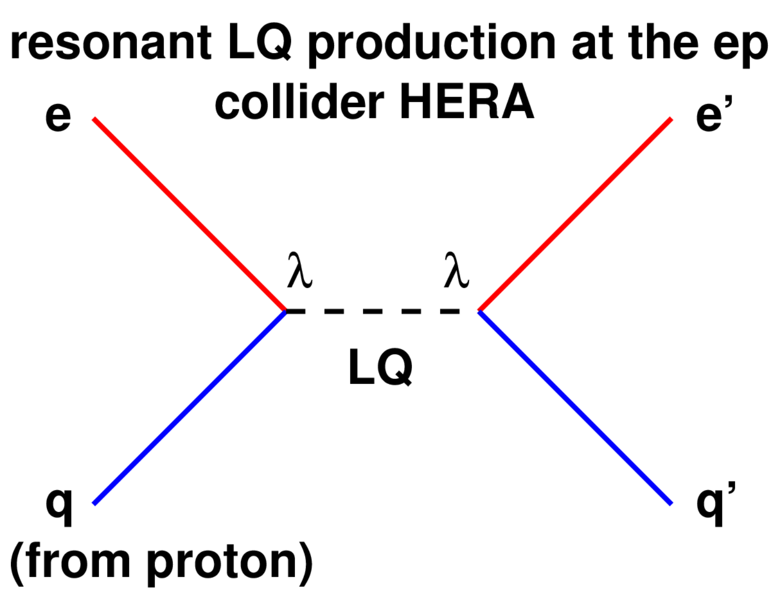
$$I(s)=\int_{s'}^{\infty} ds' P(s') \int d\vec{\theta} P(\vec{\theta}) P(N^i_{\text{obs}}|s',\vec{\theta})$$

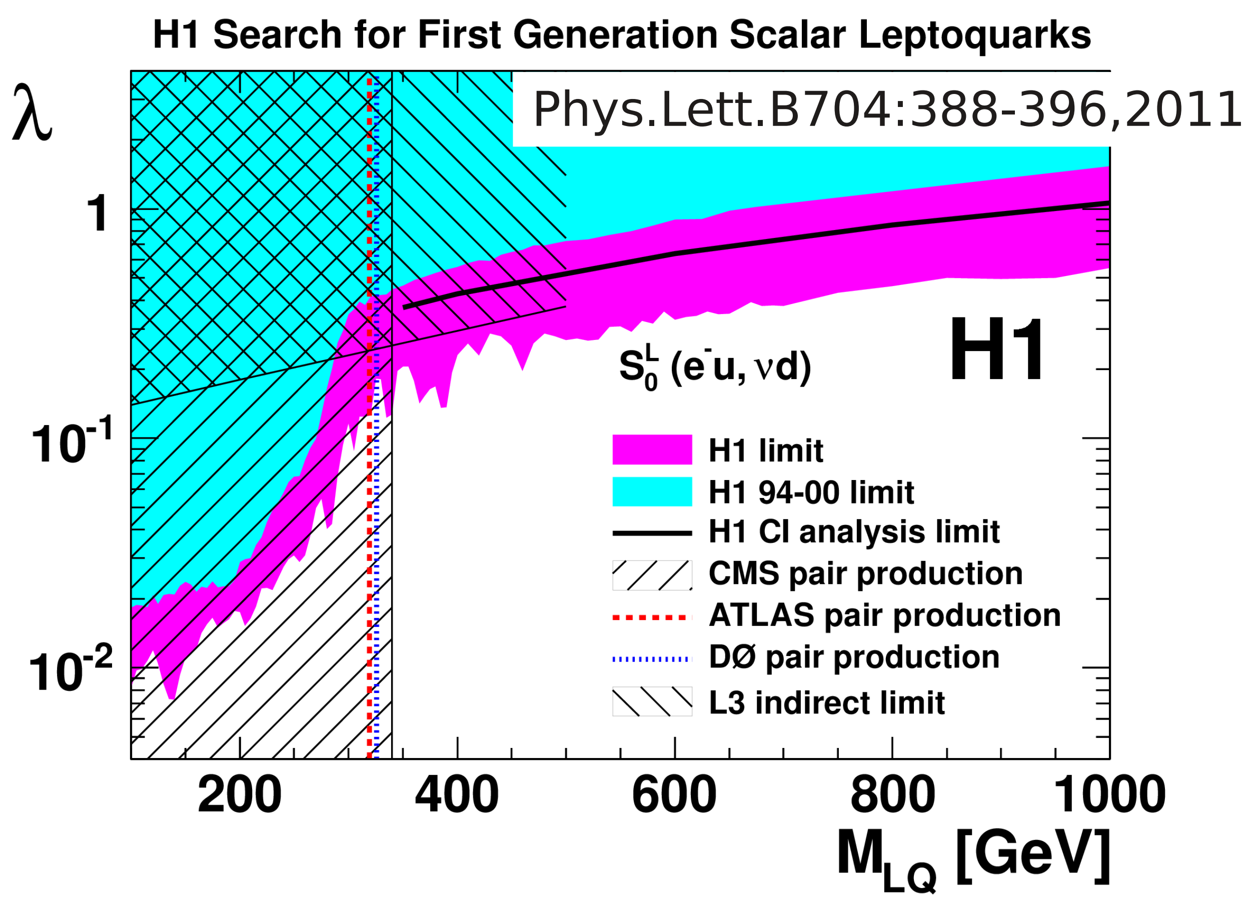
- Exclude $s>s_0$: $\frac{I(s_0)}{I(0)}=1-CL$
- Integrating over s requires sophisticated methods

“Real” examples of limits

Limit example 1: leptoquark search

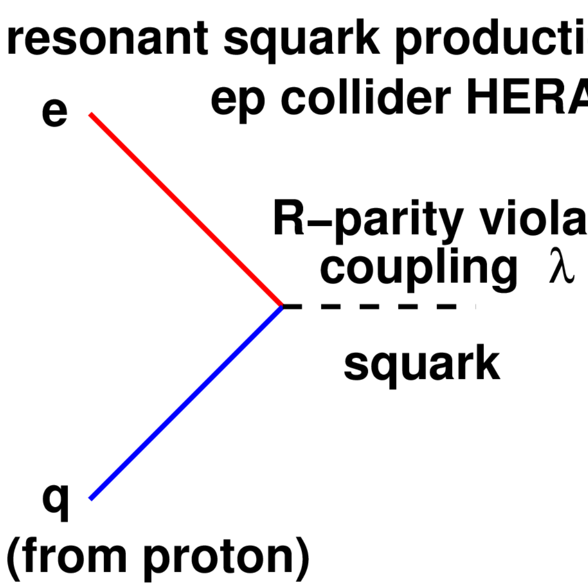
- Limits on one parameter are often shown as a function of another model parameter
- Here: leptoquark coupling (y-axis) and mass (x-axis)
- Colored regions are excluded



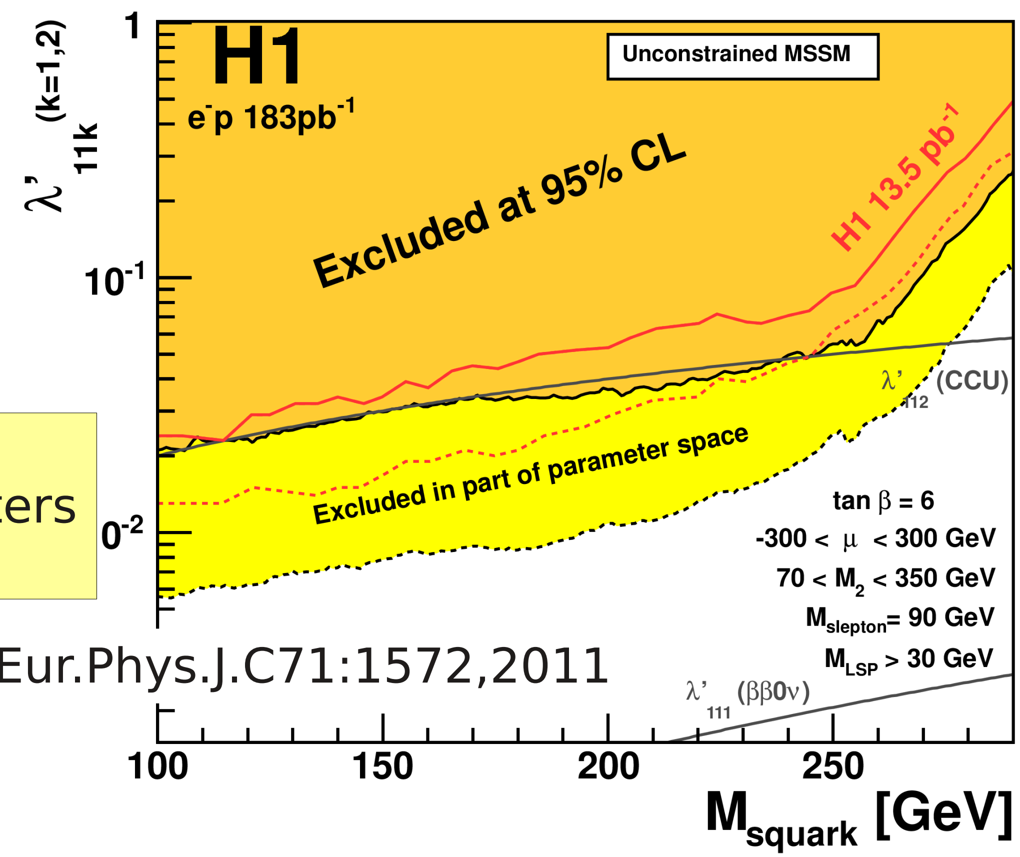


Limit example 2: R-parity violating SUSY

- Show coupling wrt Squark mass
- Scan of other (hidden) model parameters for each point
- Exclude a set of models if the (coupling,mass) is excluded for any setting of the other parameters



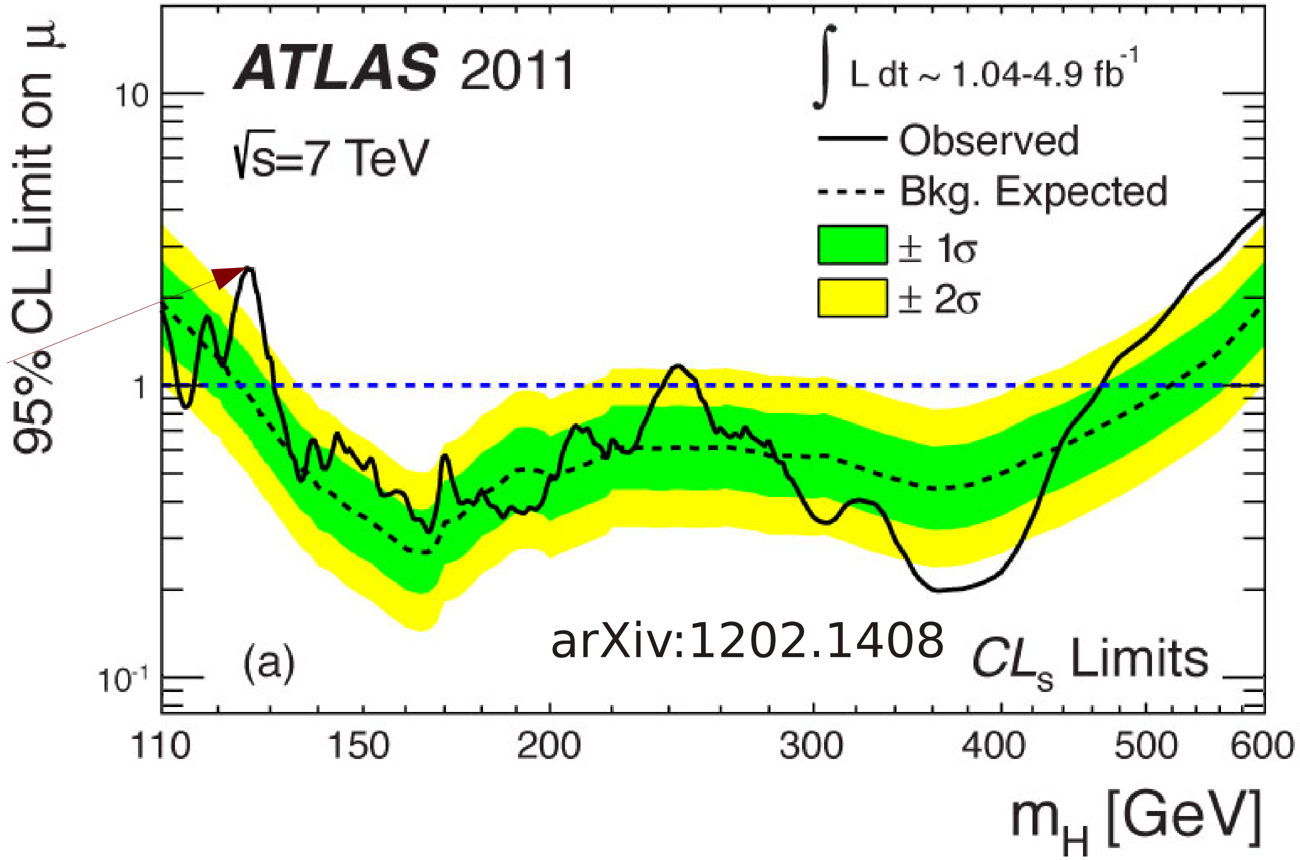
Squark decay depends on other SUSY parameters → scan



Eur.Phys.J.C71:1572,2011

Limit example 3: ATLAS Higgs search

- The model only has one unknown parameter, the Higgs mass
- A scaling factor μ on the Higgs cross section is used as a second parameter
- Solid: observed limit
- Dashed: expected limit
- Green (yellow): 1 (2) σ around expected limit
- Solid line above yellow line → Higgs candidates



Summary

- Overview of limit setting methods
  - Bayesian
  - Frequentist
- Treatment of systematic uncertainties
- Combining channels
- Modifications of the Frequentist method
  - CL_s, power constraint
- Examples of real HEP limits