Review: setting Limits in HEP experiments

Stefan Schmitt, DESY

ATLAS-D meeting

Göttingen September 21, 2011

Outline

- Reminder: some probability theory
- The Frequentist and Bayesian view on probabililty
- Confidence intervals, limits
- Comparison: Frequentist and Bayesian limits
- How to treat systematic uncertianties
- How to combine several channels
- Frequentist specific: CL_S and Power constraints

Probability densities

- Probabilites on discrete sets: each element has a finite probability

  Example: Poisson distribution

  → For event counts

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

$$\Omega = \{0, 1, 2, \ldots\}$$

- Probability densities: probabilities are defined by integrals

  Example: normal distribution

  → For systematic errors

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

$$\Omega = \mathbb{R}$$

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

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

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

- Prior: P(B) where B is the theory
- Likelihood: P(A|B) where A is the measurement
- Posterior: P(B|A) is the result of the analysis

- P(A) has no special name. Normalisation is often calculated by integrating the posterior over all theories: P(B|A)+P(~B|A)=1

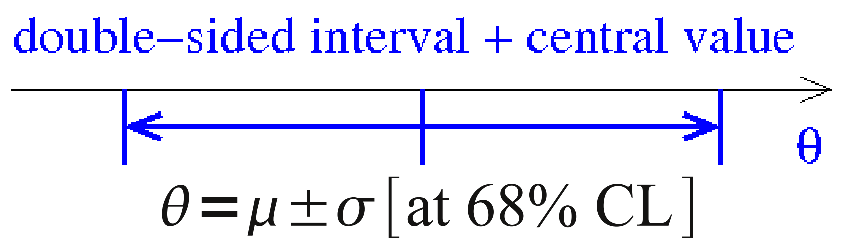
Probabilities in high energy physics

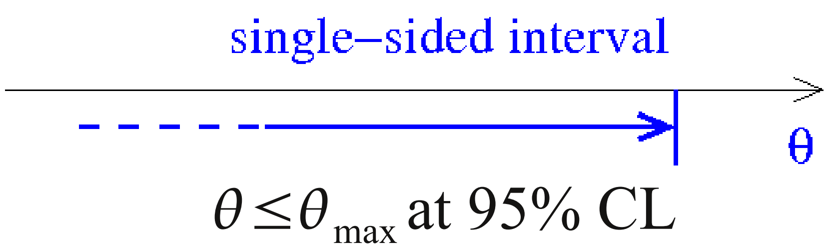
- Probability: predict number of events given the theory (parameters) and the experimental setup
- But we want to know what a specific observation tells about the theory
- Frequentist: give for each theory the probabillity of the observation (there is no probability for a theory)
- Bayes: assign probability (degree of belief) to theories
- High energy physics: make use of both views
  (some 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~P(obs|θ)   Bayesian: CL~P(θ|obs)
- Double-sided: measurement (usually CL=68%)
- Single-sided: limit (often CL=95%)





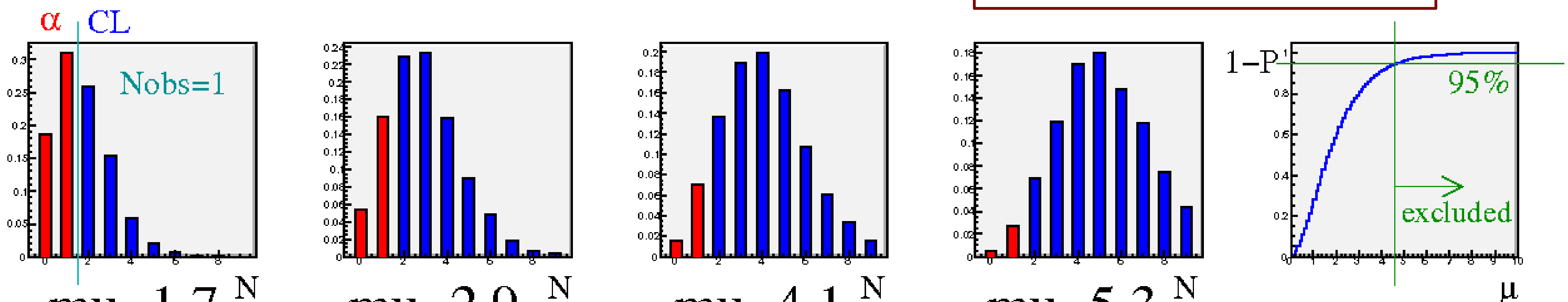
Frequentist limits

- Frequentist limit: exclude all theories which produce the data at small probability α less than 1-CL (typically: CL=0.95)

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

α: also called p-value

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



mu=1.7 P>0.05 | mu=2.9 P>0.05 | mu=4.1 P<0.05 | mu=5.3 P<0.05

Bayesian limits

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

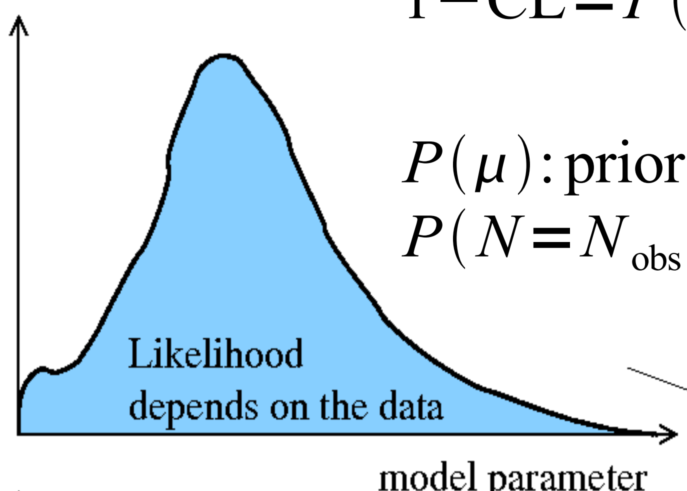
Enumerator: integrate over excluded theories

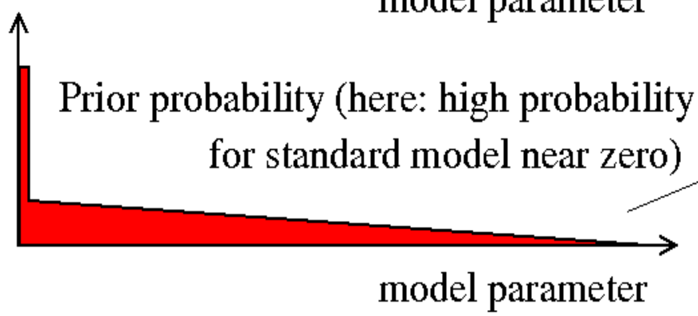
$$1-\mathrm{CL}=P(\mu\geq\mu_0|N=N_{\mathrm{obs}})=\frac{\int_{\mu_0}^{\infty}P(N=N_{\mathrm{obs}}|\mu)P(\mu)d\mu}{\int_{0}^{\infty}P(N=N_{\mathrm{obs}}|\mu)P(\mu)d\mu}$$

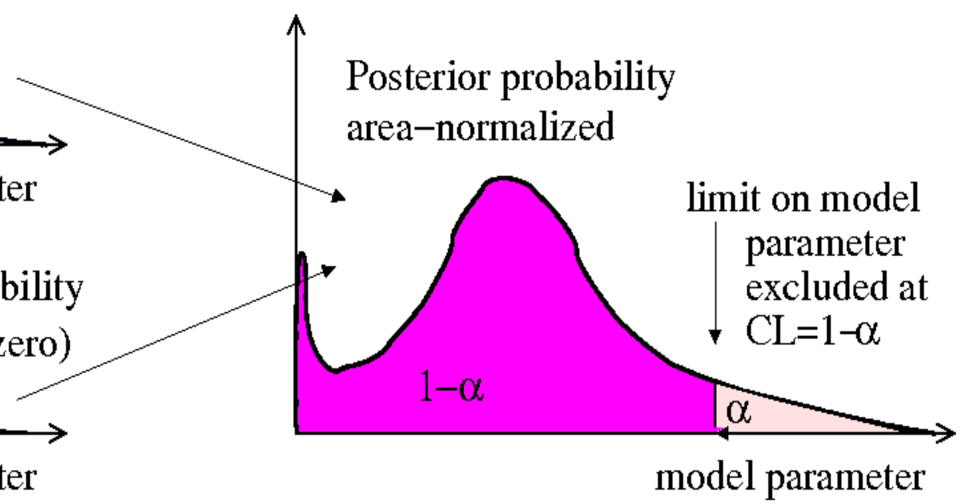
$P(\mu)$: prior probability of the model μ

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

Denominator: integrate all theories (normalisation)







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

Comparison Frequentist/Bayesian

- Frequentist limit tells about the probability of repeated (Gedanken-) experiments
- Calculation is done by integrating over possible observations
- Problem of "Unphysical" limits: CL_s and power constraints
- Systematic uncertainties?
- Combining channels?

- 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: result depends on model formulation. For example: "flat" prior in cross section is non-flat in coupling

Systematic uncertainties

- Systematic errors: detector effects, hadronisation, etc
- Described by nuisances, with given prior distributions
- Bayes: conceptually simple, just integrate over all nuisances

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

- 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

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

- *N* is observed, *L* and *b* are integrated out

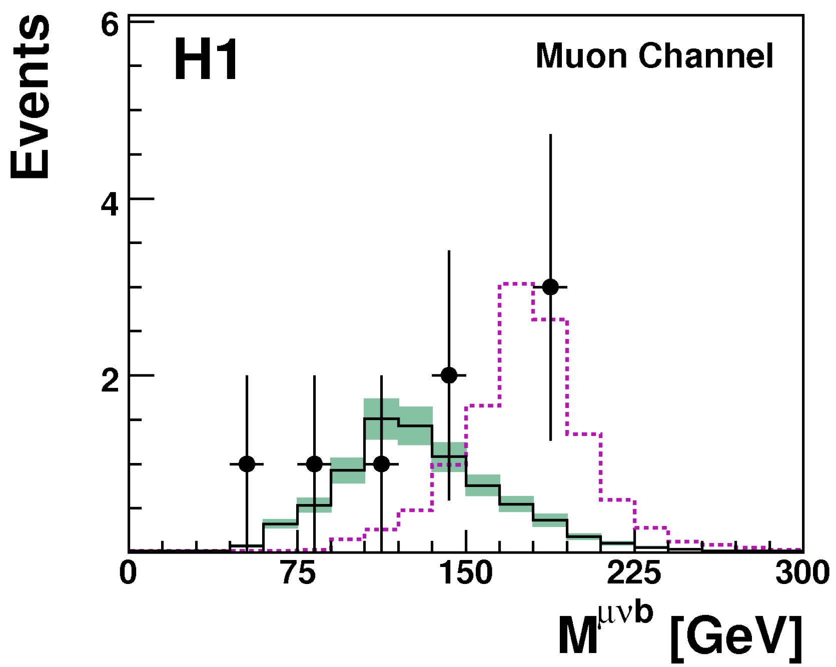
$$P(N|s) = \int dL \int db\, P(N|s,L,b)$$

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



- 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 (2)

- Bayesian methods: use n-dimensional likelihood

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

→ simple extension of the 1-dim case

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

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

Example choices of the test statistics

- Example: likelihood ratio $X = \frac{P(N_{\text{obs}}|\text{signal}+\text{bgr})}{P(N_{\text{obs}}|\text{bgr})} = \frac{L(\text{signal}+\text{bgr})}{L(\text{bgr})}$
- Other choices are possible, for example: weighted sum of all channels, weight taken from signal/bgr ratio or something similar

$$X = \sum w_i N_i^{\text{obs}} \qquad \text{simple choice: } w_i = \frac{s_i}{b_i}$$

- Note: log of likelihood ratio also is a weighted sum:

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

Many channels + systematic errors

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

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. Computational heavy! Example: integrate over O(10000) syst. configurations, call MINUIT 10000 times!
- Alternative methods, e.g. based on weighted sums, $X = \sum w_i N_i^{\mathrm{obs}}$ where bin weights w_i are fixed, optimised for systematic errors

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

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

Frequentist/Bayesian calculation summary

- Frequentist: calculate

$$\alpha(\mu) = \int_{X < X_{\mathrm{obs}}} dX \int d\vec{\theta}\, P(\vec{\theta})\, P(X|\mu, \vec{\theta})$$

- Exclude if: $\alpha(\mu) < 1 - CL$
- Integrating over θ and X is simple, well known probability densities
- Calculation has to be repeated for many models (many choices of μ)

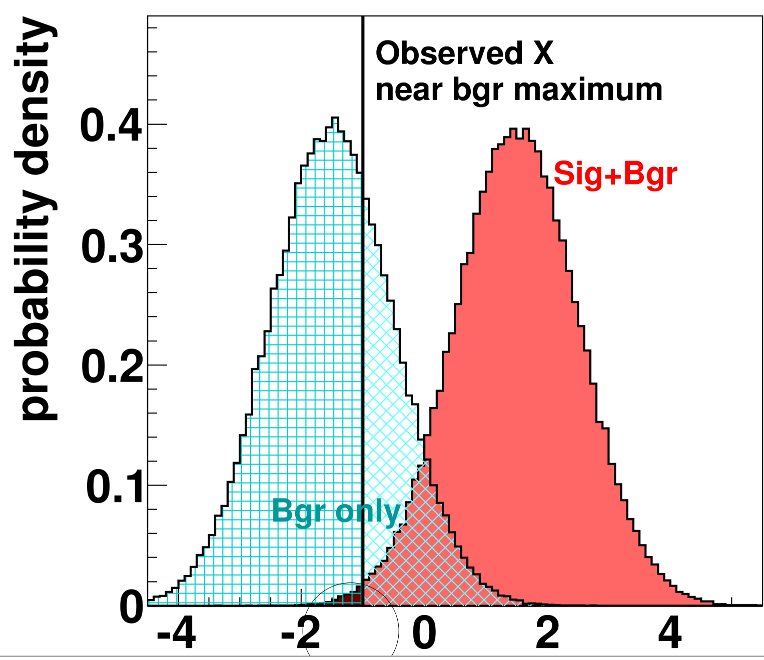
- Bayesian: calculate

$$I(\mu_0) = \int_{\mu_0}^{\infty} d\mu\, P(\mu) \int d\vec{\theta}\, P(\vec{\theta})\, P(N^i_{\mathrm{obs}}|\mu, \vec{\theta})$$

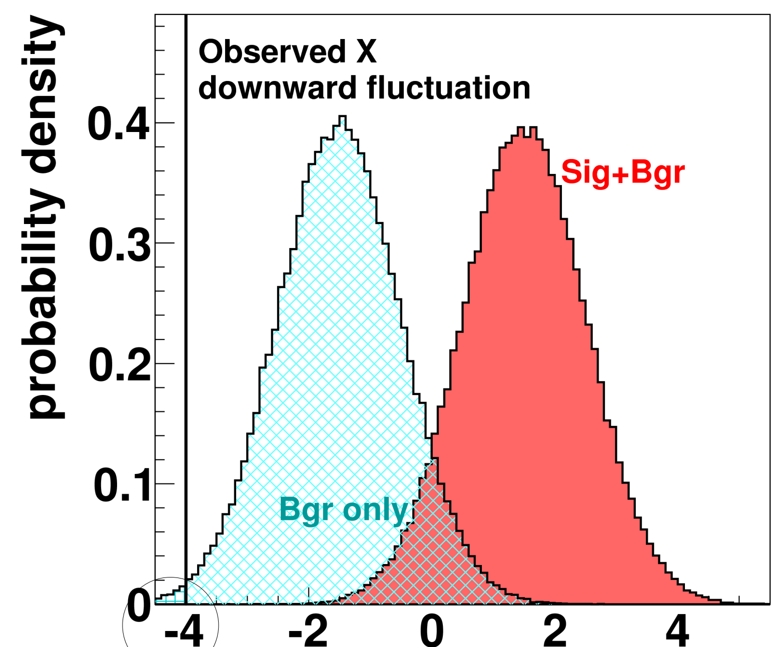
- Exclude $\mu > \mu_0$ $\frac{I(\mu_0)}{I(0)} = 1 - CL$
- Integrating over θ is simple, well known probability densities
- Integrating over μ requires sophisticated methods

Frequentist and downward fluctuations

- General problem with Frequentist methods if α is small for very small or vanishing signals
- Example: theory parameter is the signal cross section, Standard model has signal cross-section zero, observe downward fluctuation in data.



Low α for signal hypothesis → exclude
large α for background hypothesis → keep



Low α for background only hypothesis
Exclude model with zero signal?

Note: this problem is not there for Bayesian limits. The Bayesian integral over excluded models by construction never includes the Standard Model

The CL_S method

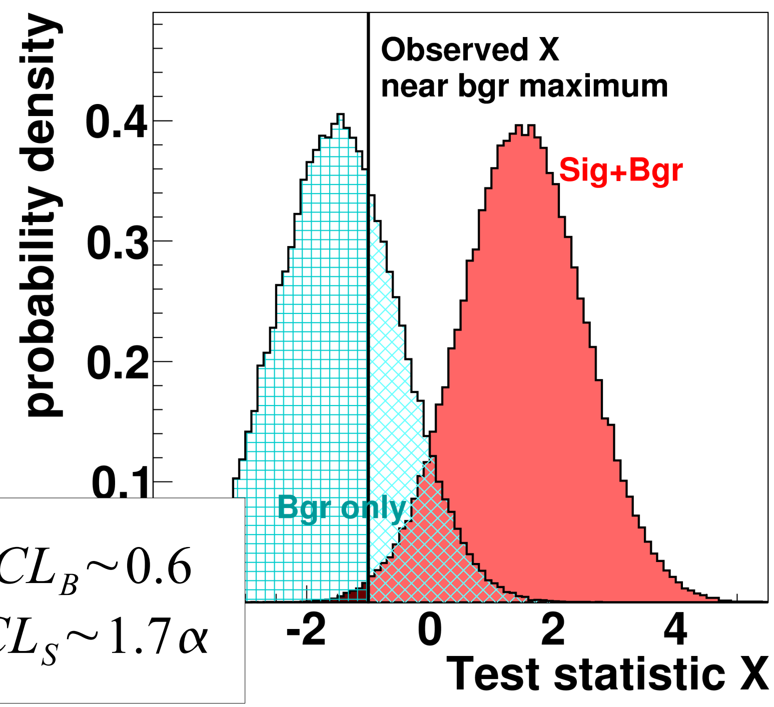
- Use ratio of two probabilities CL_S instead of α to test against CL

$$CL_{SB} = \alpha = \int_{X<X_{\mathrm{obs}}} P(X|\mathrm{signal}+\mathrm{bgr})\,dX$$

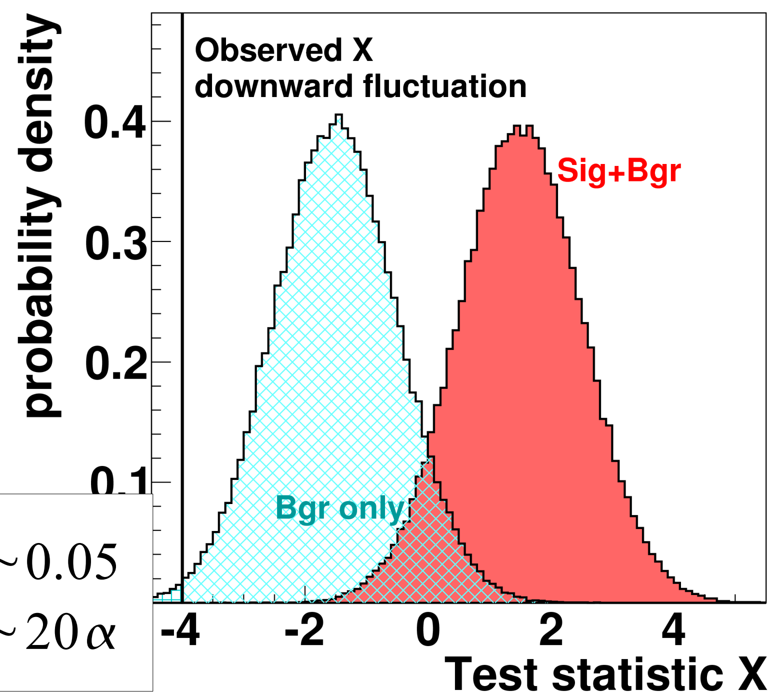
$$CL_B = \int_{X<X_{\mathrm{obs}}} P(X|\mathrm{bgr})\,dX$$

$$CL_S = \frac{CL_{SB}}{CL_B} > \alpha$$

- Standard model has CL_S=1 and is never excluded



$$CL_B \sim 0.6$$
$$CL_S \sim 1.7\alpha$$



$$CL_B \sim 0.05$$
$$CL_S \sim 20\alpha$$

Power constraint limits

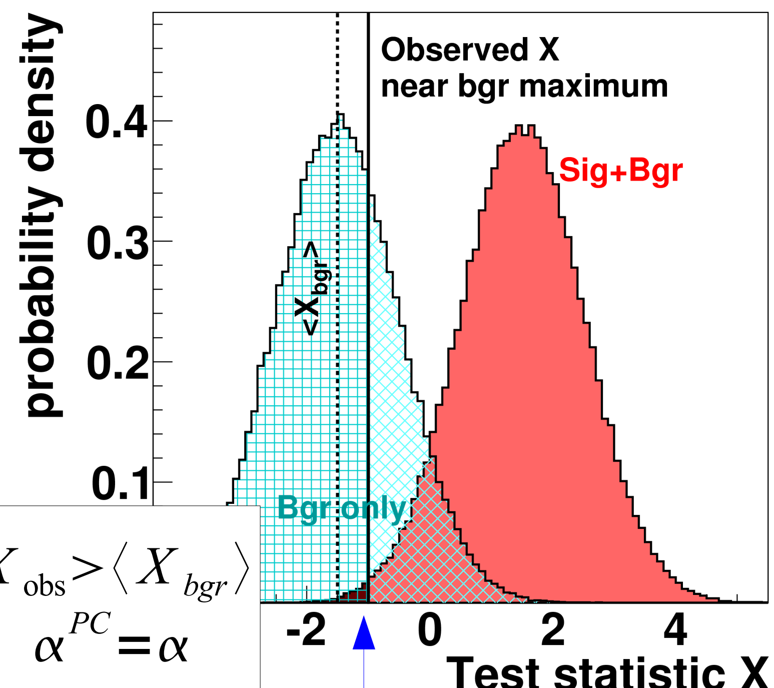
- Conservative version α^{PC} of probability α

$$\alpha = \int_{X < X_{\text{obs}}} P(X|\text{signal}+\text{bgr})\, dX \qquad \alpha^{PC} = \max\left(\alpha, \int_{X < \langle X_{bgr} \rangle} P(X|\text{signal}+\text{bgr})\, dX\right)$$

(~ quote expected limit if it is more conservative)

- Standard model has α^{PC}=0.5 and is never excluded

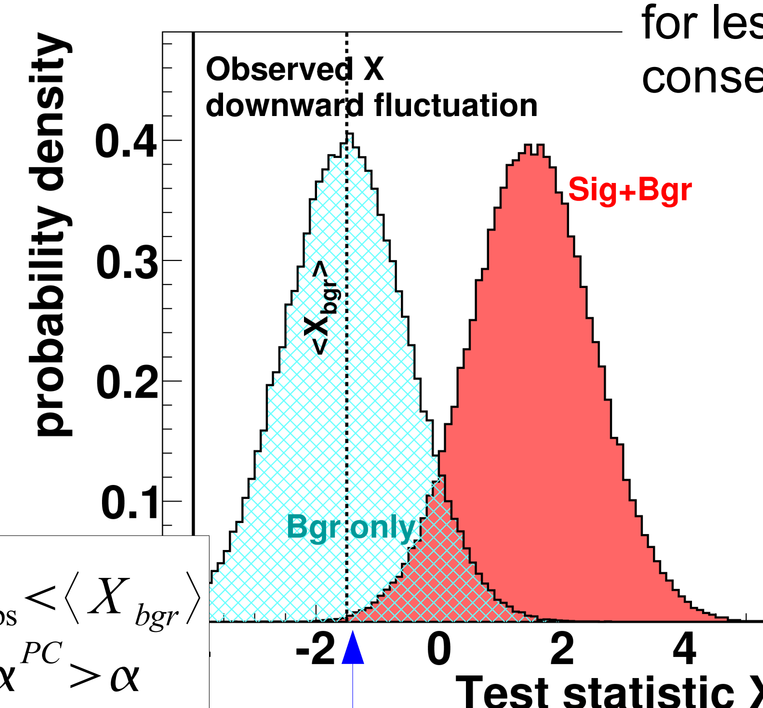
Note: possibility to use $\langle X_{bgr} \rangle$-σ for less conservative PC



$X_{\text{obs}} > \langle X_{bgr} \rangle$

$\alpha^{PC} = \alpha$

Limit taken from data



$X_{\text{obs}} < \langle X_{bgr} \rangle$

$\alpha^{PC} > \alpha$

Limit taken from MC

Summary

- Overview of limit setting methods
  - Bayesian method
  - Frequentist method
- Treatment of systematic uncertainties
- Combining channels
- Modifications of the Frequentist method
  - CL_S
  - Power constraints