Tutorial/Lecture on Limits

Stefan Schmitt, DESY

Tutorial/lecture for the
Terascale Statistics School

Mainz, April 5, 2011

Outline

- Reminder: some probability theory
- The Frequentist and Bayesian view on probabililty
- Confidence intervals, limits
- Frequentist and Bayesian limit examples
- Background, systematic uncertainties
- Combining several bins or channels
- Not covered:
  - Discoveries, p-values, Bayes factors, ...
  - Bayesian objective priors, …
  - Limit tools in Root

Exercises

- Handout with 8 exercises, but time this afternoon is limited: only a selection of the exercises to be worked through in detail
- Procedure: lecture is interrupted a few times for work on exercises, followed by a discussion of the solutions
- Root macros

  Initial version of the macros are on the virtual machine:

  /statistics-school/limits/

  Improved macros are on the web:

  http://www.desy.de/~sschmitt/LimitLecture/

Probability theory: selected items

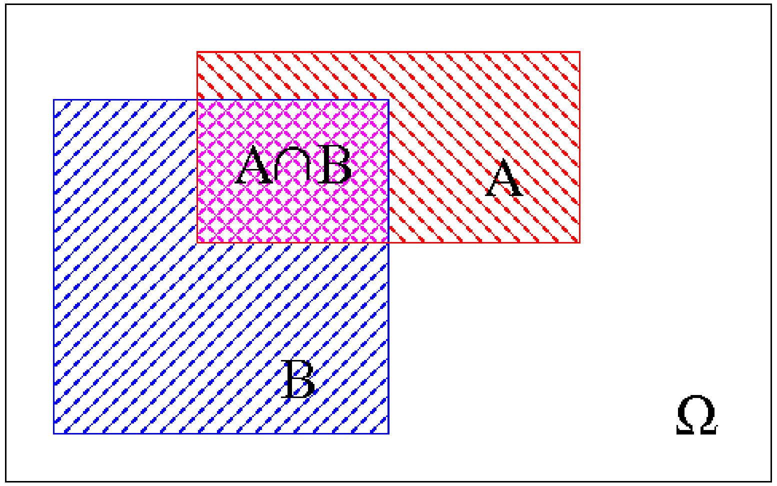
- Elements of Ω : events, outcomes of an experiment
- Probability of A: $0 \leq P(A) \leq 1, P(\Omega)=1$

$$P(\Omega)=1\,,\, P(\emptyset)=0 \qquad P(\Omega \setminus A)=1-P(A)$$

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

- 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

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

  Examples:
  - Poisson distribution:

    μ is a nuisance parameter

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

  - Normal distribution:

    μ and σ are nuisance parameters

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

- 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 using P(B|A)+P(~B|A)=1

Exercise on Bayes' law

- Consider a disease and a test for the disease
- 0.1% of the population have the disease (prior)
- If one has the disease, the test is positive with 99% probability (likelihood)
- If one does not have the disease, the test is positive with 1% probability
- What is the (posterior) probability to have the disease, given a positive test?

Discussion exercise 1

- Prior probability: $P(B)=0.1\%$
- Likelihood: $P(A|B)=99\%$
- Normalisation:

$$P(A) = P(A \cap B)+P(A \cap \sim B)=P(A|B)*P(B)+P(A|\sim B)*P(\sim B)$$
$$= 0.001*0.99+0.01*0.99 = 0.01098$$

- Posterior probability: $P(B|A)=0.99*0.001/0.01098=9\%$
- The posterior is a “Bayesian probability”: there is a true parameter (has disease or not). The “degree of belief” to have the disease is 9% given the positive test.

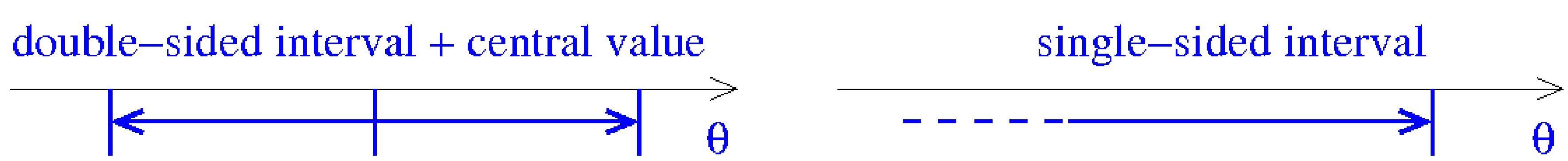
Probabilities in high energy physics

- 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 (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%)



$\theta = \mu \pm \sigma$ [at 68% CL] $\theta \leq \theta_{max}$ at 95% CL

Setting limits: step by step

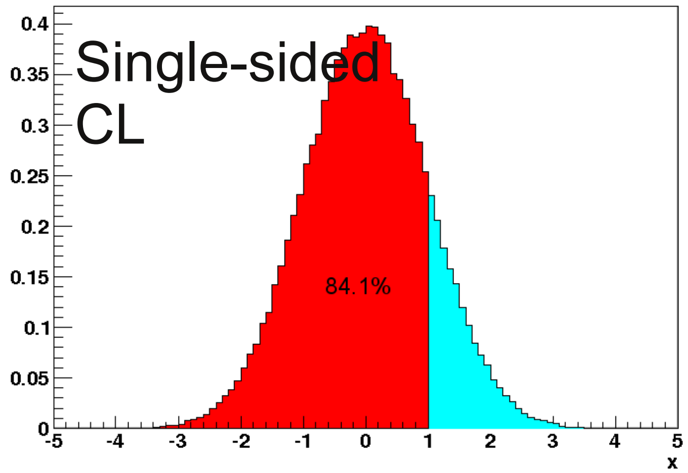
- One channel, no background, no systematics
- One channel, with background, no systematics
- One Channel, background and systematics
- Combining channels, no systematics
- Combining channels with systematics

Limits: Gaussian approximation

- Idea: determine the central value plus error (lecture by Olaf), assume Gauss distribution
- $\Delta\chi^2$= 1,2,3,... corresponds to a certain probability

$\Delta\chi^2$	1	2	3
Single-sided CL	84.1%	97.7%	99.9%

Single-sided CL	95.0%	99.0%
$\Delta\chi^2$	1.64	2.33



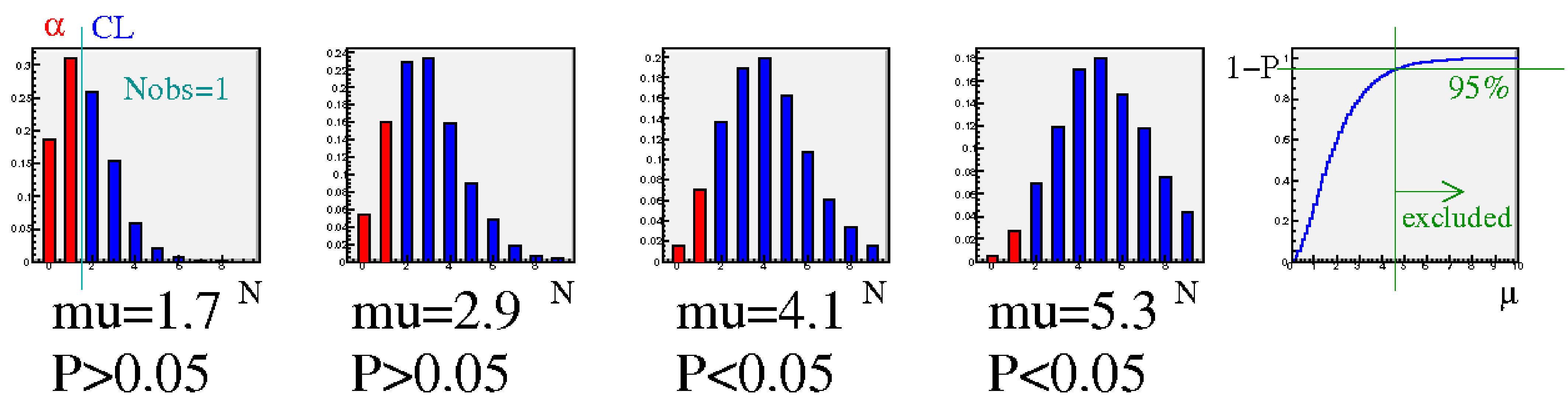
- Problem: several approximations involved: distribution approximated by Gaussian, σ independent of the model and σ,μ are approximated by the measured value and measured error

Frequentist limits

- Frequentist limit: exclude all theories which produce the data at probability less than α=1-CL

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

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



Frequentist limit exercise

- Exercise 2a: counting experiment (Poisson), N_{obs}=0 what is the 95% CL limit on the parameter μ?
  Calculate analytically, using Poisson's law.
  How does the calculation look like for N_{obs}=1,2,3,...?

- Exercise 2b: N_{obs}=2,10,100 and compare to Gaussian approximation (use root macros)

Calculation of Poisson sums

- Sum over Poisson terms is related to χ^2 distribution with number-of-degrees of freedom “k”:

$$\chi^2(x;k)=\frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)} \qquad P(N;\mu)=\frac{e^{-\mu}\mu^N}{N!}$$

- The Poisson sum can be expressed by an integral over the χ^2 distribution (proof by partial integration)

$$\alpha(\mu,N)=\int_{2\mu}^{\infty}\chi^2(x;2(N+1))\,\mathrm{d}x=\sum_{i=0}^{N}P(i;\mu)$$

- Standard functions for χ^2 integrals can be used:

`α( ,N)=TMath::Prob(2* ,2*(N+1))` and

` =0.5*TMath::ChisquareQuantile(1-α,2*(N+1))`

Bayesian limits

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

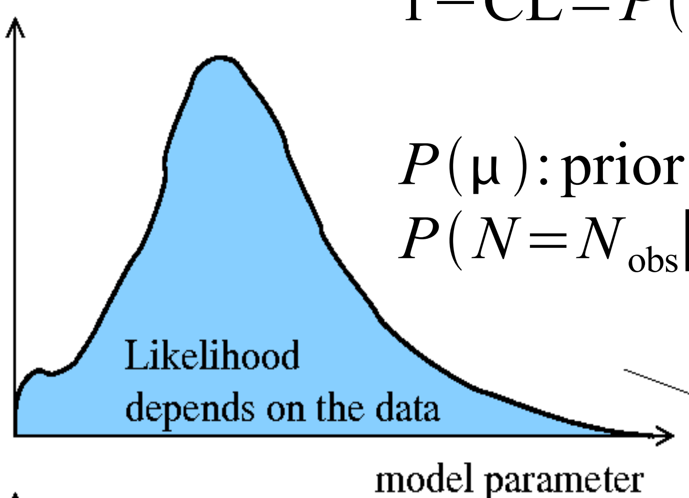
Enumerator: integrate over excluded theories

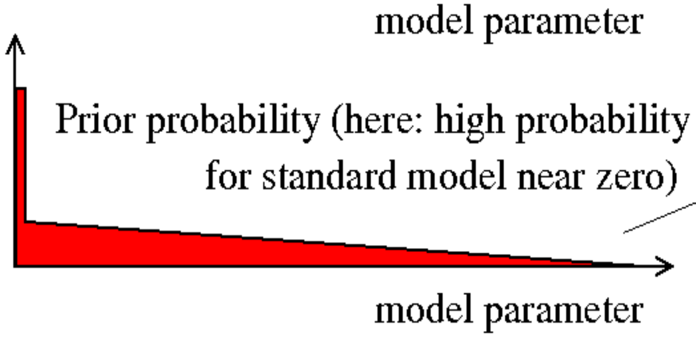
$$1-\mathrm{CL}=P(\mu\geq\mu_0|N=N_{\mathrm{obs}})=\frac{\int_{\mu_0}^{\infty}P(N=N_{\mathrm{obs}}|\mu)\,P(\mu)\,d\mu}{\int_{0}^{\infty}P(N=N_{\mathrm{obs}}|\mu)\,P(\mu)\,d\mu}$$

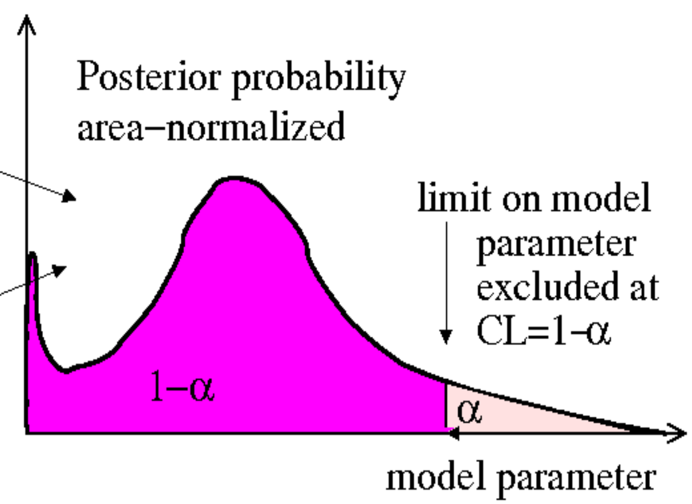
$P(\mu)$: prior probability of the model μ

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

Denominator: integrate all theories (normalisation)







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

Bayesian limit exercise

- Exercise 3a: calculate the Bayesian limit for N_{obs}=0 assuming a “flat prior in N”, $P(\mu)=1$.
  Calculate analytically, using Poisson's law. How does the calculation look like for N_{obs}=1,2,3,...?
- Exercise 3b: calculate the Bayesian limit (root macro) for N_{obs}=2,10,100 (flat prior) & compare to exercise 2b
- Exercise 3c: use a prior $P(\mu)=\mu$, $N_{obs}=\{0,2,10,100\}$
- Exercise 3d: use a flat prior up to μ_{max}=90, set to zero above μ_{max}

Discussion Exercise 2/3

- Gaussian approximation fails for small N_{obs}
- Bayes with "flat" prior and Frequentist accidentially agree for the simple Poisson case (also see page 16)

$$\int_{\mu_0}^{\infty} \exp(-\mu) \frac{\mu^{N_0}}{N_0!} d\mu = \sum_{N=0}^{N_0} \exp(-\mu_0) \frac{\mu_0^N}{N!}$$

	N_{obs}=0	N_{obs}=2	N_{obs}=10	N_{obs}=100
Frequentist	3.0	6.3	17.0	118.1
Gauss.approx	0.0	4.3	15.2	116.4
Bayes flat prior	3.0	6.3	17.0	118.1
Bayes $P(\mu)=\mu$	4.7	7.7	18.2	119.2
Bayes flat up to $\mu=90$	3.0	6.3	17.0	89.7

Discussion exercise 2/3 continued

- Non-flat prior: differences between Bayes and Frequentist limits
- Ill-chosen prior with μ_{max}=90 for N_{obs}=100: limit is defined by prior!
- Dependence on prior: main reason why Bayesian methods are not used that much in HEP

	N_{obs}=0	N_{obs}=2	N_{obs}=10	N_{obs}=100
Frequentist	3.0	6.3	17.0	118.1
Gauss.approx	0.0	4.3	15.2	116.4
Bayes flat prior	3.0	6.3	17.0	118.1
Bayes $P(\mu)=\mu$	4.7	7.7	18.2	119.2
Bayes flat up to μ=90	3.0	6.3	17.0	89.7

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
- Systematic uncertainties?
- Combining channels?
- p-values

Next slides

- Bayesian limit tells about the model probability
- Calculation is done by integrating over models
- Result depends on model formulation, "flat" prior in cross section is non-flat in coupling
- Possibility to have "objective" priors
- Bayes factors

Red: not discussed in this lecture

Setting limits: step by step

- One channel, no background, no systematics
- One channel, with background, no systematics
- One Channel, background and systematics
- Combining channels, no systematics
- Combining channels with systematics

Limits with background

- Expected number of events is given by the sum of a signal and background contribution, both growing with the integrated luminosity

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

- Luminosity and background are known, find limit on the signal contribution
- Frequentist: set limit on μ, then divide by L and subtract b
- Bayesian: use prior which is zero for s<0

Exercise with background

- Exercise 4: calculate Frequentist and Bayesian limits for L=1, N_{obs}={0,2} and b={0.5,2.0,3.5}

	bgr=0.5		bgr=2.0		bgr=3.5	
	$N_{obs}=0$	$N_{obs}=2$	$N_{obs}=0$	$N_{obs}=2$	$N_{obs}=0$	$N_{obs}=2$
Bayesian						
Frequentist						

- Frequentist: set limit on μ, then subtract b
- Bayesian: use prior which is zero for s<0

Exercise with background

- Exercise 4: calculate Frequentist and Bayesian limits for L=1, N_{obs}={0,2} and b={0.5,2.0,3.5}

	bgr=0.5		bgr=2.0		bgr=3.5	
	N_{obs}=0	N_{obs}=2	N_{obs}=0	N_{obs}=2	N_{obs}=0	N_{obs}=2
Bayesian	3.0	5.8	3.0	4.8	3.0	4.3
Frequentist	2.5	5.8	1.0	4.3	-0.5	2.8

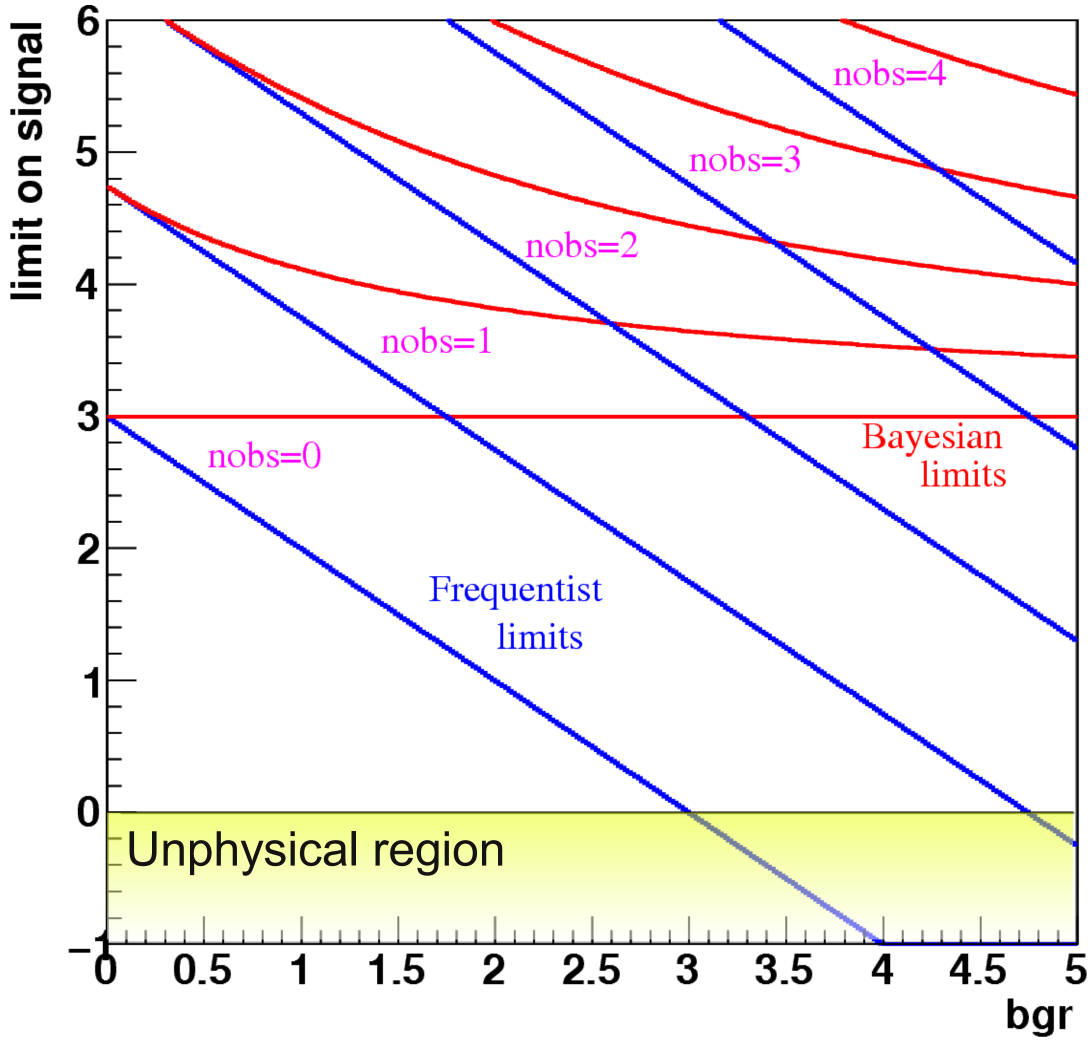
- Problem for Nobs=0 and bgr=3.5: limit excludes all signal above -0.5. Even the “standard model” s=0 is excluded

Discussion Exercise 4

- Frequentist analysis can give limits where all models are "excluded" at a given CL (even the model with s=0)

  N_{obs}=0, μ=s+b, b=3.5

  → limit s<-0.5 @ 95% CL but s>=0 physical bound
- Can not happen for Bayesian limit, because prior knowledge s>=0 is used



Limits near a boundary

- What to do if frequentist analysis excludes parameters beyond the sensitivity of the experiment or beyond boundaries?
- Quote "expected" limit to show the sensitivity of the experiment (limit averaged over many experiments)
- "Modified Frequentist" (CL_S method)

$$\alpha = \mathrm{CL}_S = \frac{\mathrm{CL}_{SB}}{\mathrm{CL}_B} = \frac{P(N \leq N_{\mathrm{obs}}; \mu = S+B)}{P(N \leq N_{\mathrm{obs}}; \mu = B)}$$

- Use Bayesian methods (prior knows about boundaries)
- . . .
- See PDG review on statistics for detailed discussion

Expected limit exercise

- Expected limit: average limit of repeated background experiments (sensitivity), in our case:

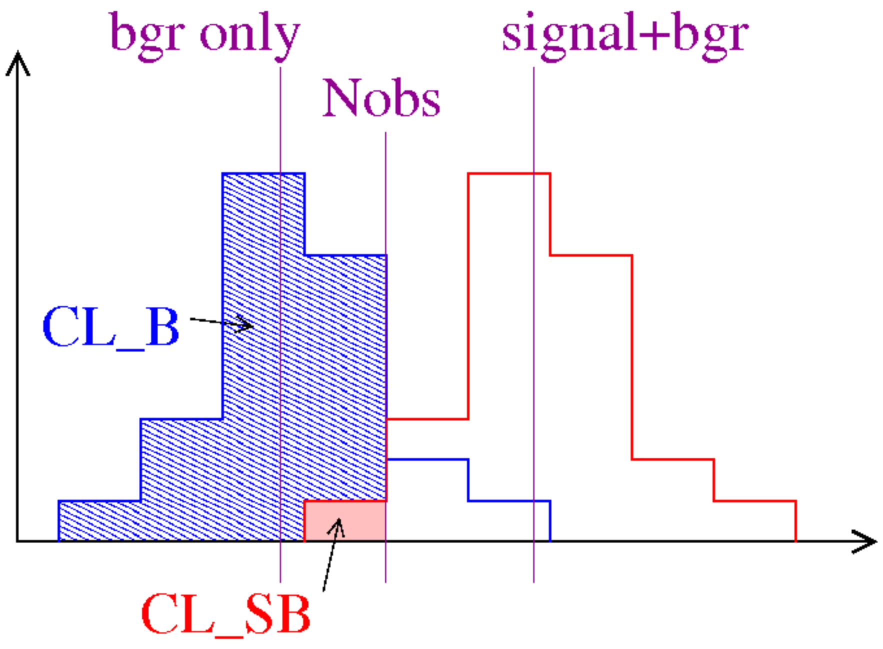
$$\langle \mu_{\mathrm{lim}} \rangle = \sum_{n=0}^{\infty} \frac{e^{-b} b^n}{n!} \mathrm{LimitOnSignal}(b, n)$$

- Exercise 5: calculate expected limits for b={0.5,2.0,3.5} and compare to exercise 4

	bgr=0.5		bgr=2.0		bgr=3.5	
	N_{obs}=0	N_{obs}=2	N_{obs}=0	N_{obs}=2	N_{obs}=0	N_{obs}=2
Bayesian	3.0	5.8	3.0	4.8	3.0	4.3
Frequentist	2.5	5.8	1.0	4.3	-0.5	2.8
Expected	3.3		4.2		4.9	

CL_S: exercise

- Modified Frequentist limit: $$CL_S = \frac{CL_{SB}}{CL_B} = \frac{P(N \le N_{obs}; \mu = S + B)}{P(N \le N_{obs}; \mu = B)}$$
  signal probability is normalized to bgr probability
- At given N_{obs}: for zero signal, CL_S=1. For large signal, CL_S=0
- Use CL_S like $\alpha \rightarrow$ Standard model never excluded
  "conservative", over-coverage
- Exercise 6: calculate limits using the CL_S method



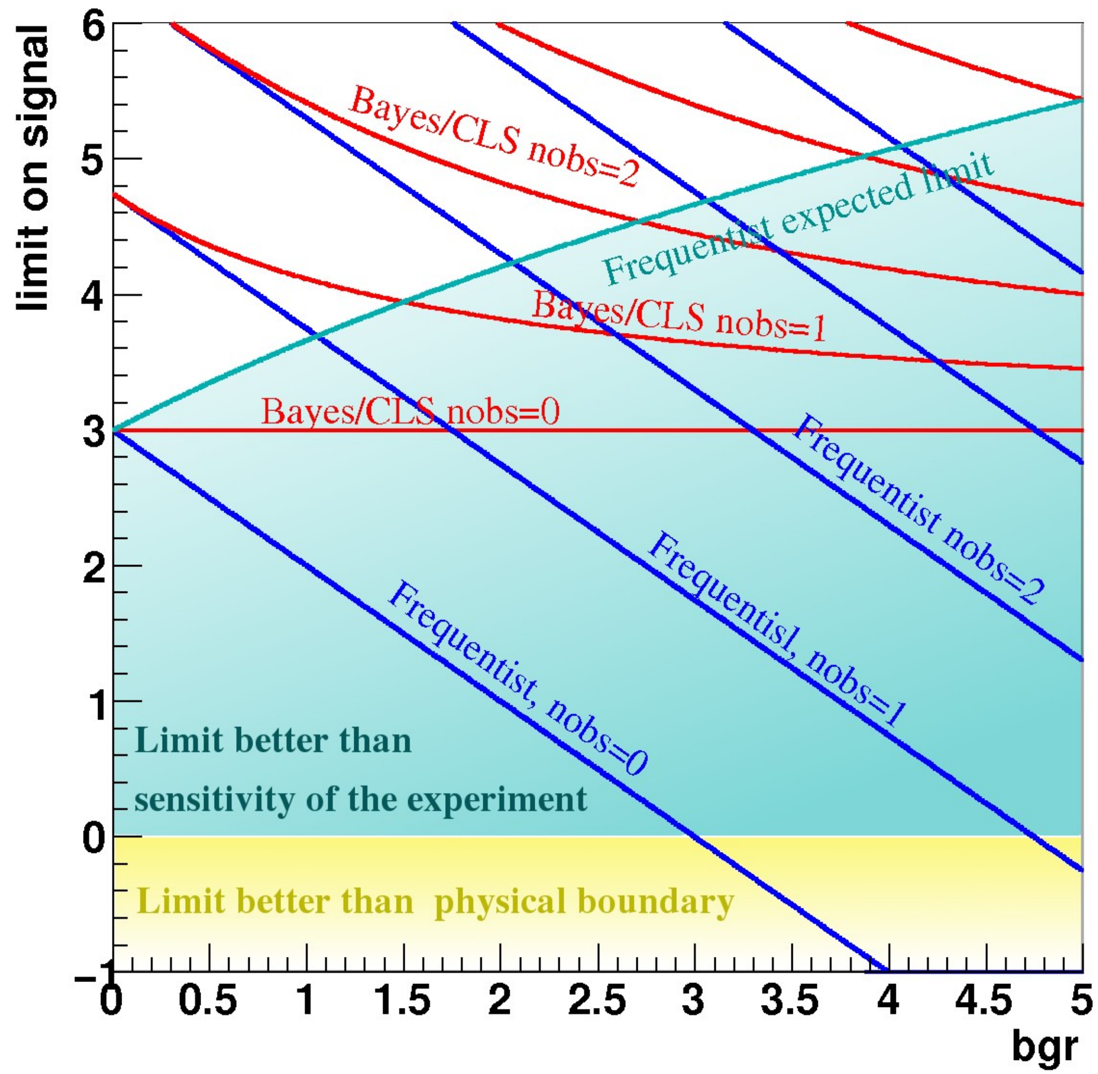
Exercise 6 discussion

- CL_s limit agrees with Bayesian limit for flat prior!
- Reason: identity of Poisson sums and integrals (slide 16)
- Note: agreement is valid only for the simplest case. Picture changes if there are many channels and systematic errors

	bgr=0.5		bgr=2.0		bgr=3.5	
	N_{obs}=0	N_{obs}=2	N_{obs}=0	N_{obs}=2	N_{obs}=0	N_{obs}=2
Bayesian $=CL_s$	3.0	5.8	3.0	4.8	3.0	4.3
Frequentist	2.5	5.8	1.0	4.3	-0.5	2.8
Expected	3.3		4.2		4.9	

Summary limits with background

- Frequentist limit may become "unphysicsal" or "too good"
- Expected limit: sensitivity of the experiment
- CL_s method: agrees with Bayesian (with flat prior) for the case of 1 bin and no syst.
- By construction: CL_s limit never excludes model with zero signal



Setting limits: step by step

- One channel, no background, no systematics
- One channel, with background, no systematics
- One Channel, background and systematics
- Combining channels, no systematics
- Combining channels with systematics

Systematic uncertainties

- Systematic errors: detector effects, hadronisation, etc
- Describe by nuisances, with given prior distributions

  Example: energy scale, measured energies are multiplied by a factor f, with error df

  → prior of f is a Gaussian with $\mu=1$ and $\sigma=df$
- Limits are often calculated by "marginalising" (integrating over) systematic parameters, then using Frequentist methods
- Note: marginalisation is Bayesian → "hybrid method"

Example with systematic errors

- Consider signal

  $$\mu = L(s+b),\ \ L: \text{integrated luminosity},\ s, b: \text{signal, background cross sections}$$

  with systematic errors:

  $$L = L_0 \pm \sigma_L,\ b = b_0 \pm \sigma_b$$

- Full probability density has three contributions

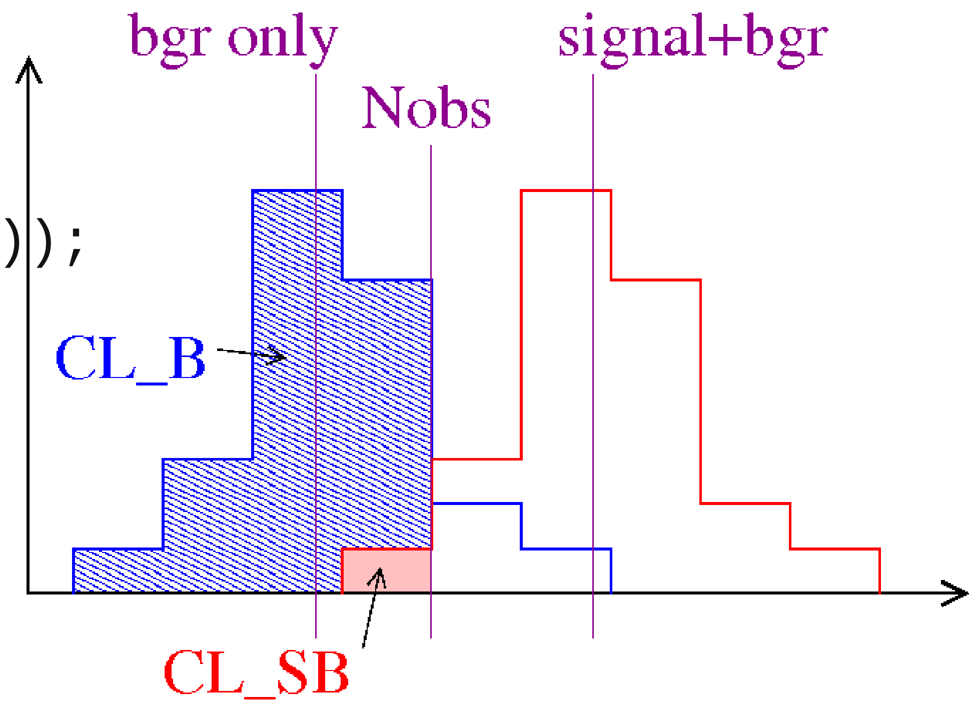
$$P(N, L, b) = \underbrace{\frac{e^{-L(s+b)}(L(s+b))^N}{N!}}_{\text{observation}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_L} e^{\frac{-(L-L_0)^2}{2\sigma_L^2}}}_{\text{syst. error on L}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_b} e^{\frac{-(b-b_0)^2}{2\sigma_b^2}}}_{\text{syst. error on b}}$$

- *N* is observed, *L* and *b* are integrated over
- Exercise 7: limits for N_{obs}={0,2} with/without syst. errors on b, L

Exercise 7 macro

- Typical example for the use of Monte Carlo methods to calculate probabilities
- Probabilities are calculated by counting the outcomes of toy experiments

```
l=rnd->Gaus(1.0,dLumi);
b=rnd->Gaus(bgr,dBgr);
Int_t n_b=rnd->Poisson(l*b);
Int_t n_sb=rnd->Poisson(l*(signal+b));
. . .
if(n_b<=nobs) nexp_b += 1.0;
if(n_sb<=nobs) nexp_sb += 1.0;
. . .
Double_t cl_s=nexp_sb/nexp_b;
```



Discussion exercise 7

- Systematic uncertainties have some impact on the result
- Our example:
  - If background is small, bgr errors have small influence
  - Luminosity affects both signal and background → all limits

CL_S limits	bgr=0.5		bgr=3.5	
	$N_{obs}=0$	$N_{obs}=2$	$N_{obs}=0$	$N_{obs}=2$
No syst	3.0	5.8	3.0	4.3
$\sigma_b/b=50\%$	3.0	5.8	3.1	4.9
$\sigma_L/L=10\%$	3.1	6.0	3.1	4.6
Both syst.	3.1	6.0	3.1	5.0

Setting limits: step by step

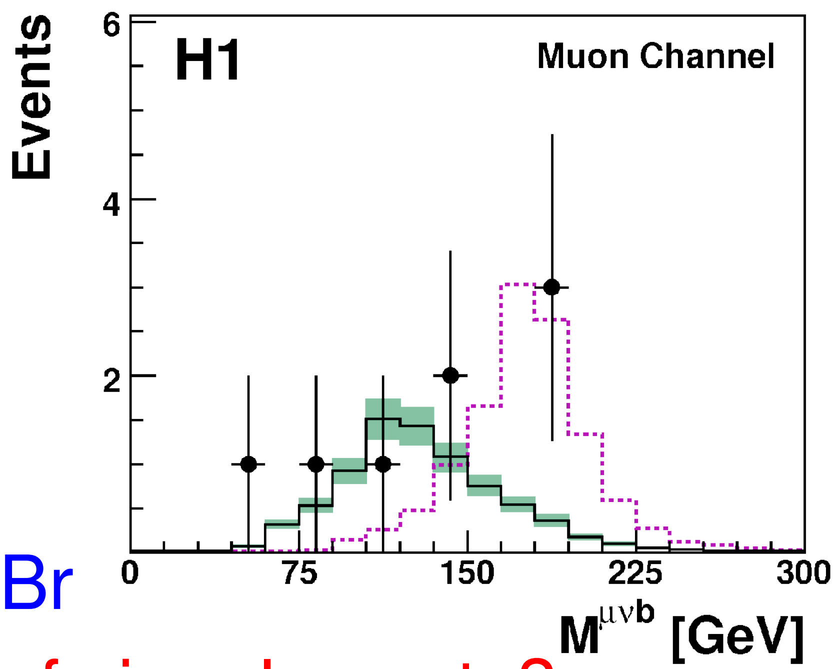
- One channel, no background, no systematics
- One channel, with background, no systematics
- One Channel, background and systematics
- Combining channels, no systematics
- Combining channels with systematics

Combining bins or channels

- Up to now: 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



- For each channel, specify efficiency*Br
- What is the limit on the total number of signal 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 statistics" 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

Choice of the test statistics

- Example: likelihood ratio
- Or likelihood normalised to its maximum

$$X = \frac{L(\text{signal+bgr})}{L(\text{bgr})}$$

$$X = \frac{L(\text{signal+bgr})}{L_{\text{max}}}$$

- 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:

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

Exercise with two channels

- Consider two channels, ε_i=efficiency*BR=0.5 $\mu_i = \epsilon_i * s + b_i$
- One channel dominated by signal, the other dominated by background
- Exercise 8a: calculate the CL_S limit on the number of signal events using only channel 1 or only channel 2
- Exercise 8b: calculate the limit by adding the two channels
- Exercise 8c: calculate the limit using both channels and $X=w_1N_1+w_2N_2$ where $w_i=s_i/b_i$ ($s_i=s*\varepsilon_i$)

	N_{obs}	bgr
Channel 1	7	6.5
Channel 2	2	1.8

Discussion Exercise 8

- The two channels give different limit
- Combined limit is better than each channel alone
- Combined limit is better than the plain sum of the two channels

	N_{obs}=0	bgr	CL_S limit
Channel 1	7	6.5	14.8
Channel 2	2	1.8	9.9
Added	9	8.3	8.2
Weighted sum	(7,2)	(6.5, 1.8)	7.3

Setting limits: step by step

- One channel, no background, no systematics
- One channel, with background, no systematics
- One Channel, background and systematics
- Combining channels, no systematics
- Combining channels with systematics

Many channels + systematic errors

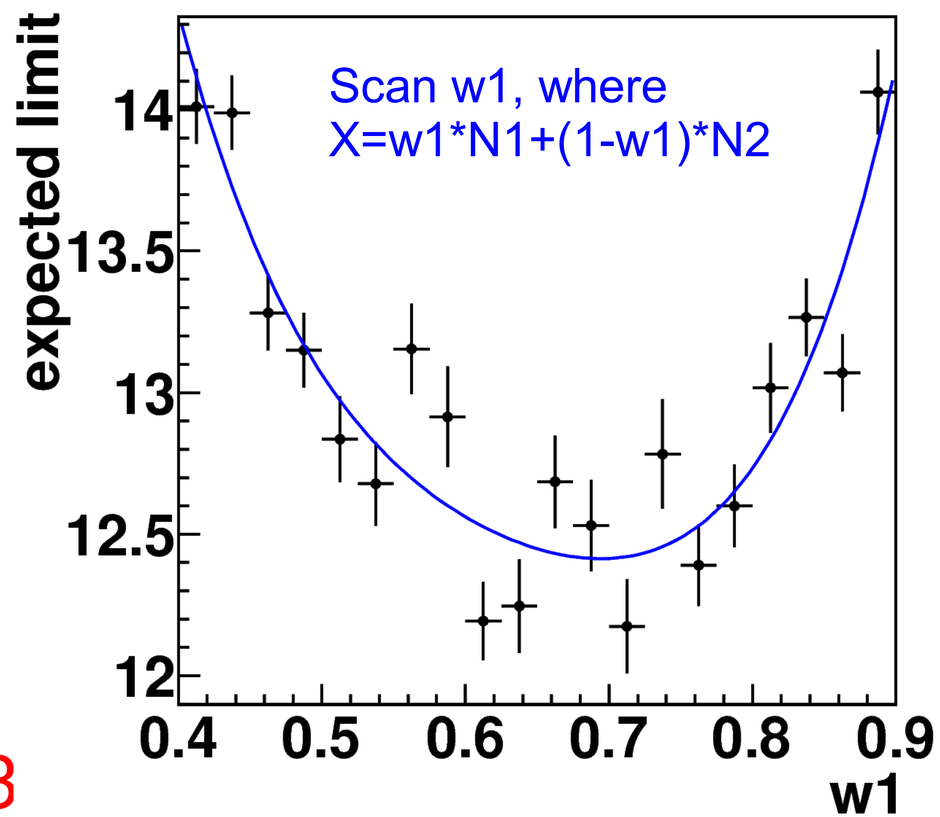
- Most common case in HEP (example: Higgs search)
- Bayesian: use Likelihood and integrate using given priors for systematic errors and models → limits

- Frequentist: define "good" test statistics X, then
  - Calculate confidence levels similar to the case of one channel+systematic errors → limits
  - Question: what is a "good" test statistics?

Many channels + systematics (2)

- Why not to use channel weight $w_i \sim s_i/b_i$ like exercise 8?
- Example 1: two channels with same s/b, but different systematic errors on b

	eff	bgr
Channel 1	0.5	4.0±0.5
Channel 2	0.5	4.0±3.0



→ channel with larger (systematic) error is less sensitive to the signal, it should have a smaller weight.

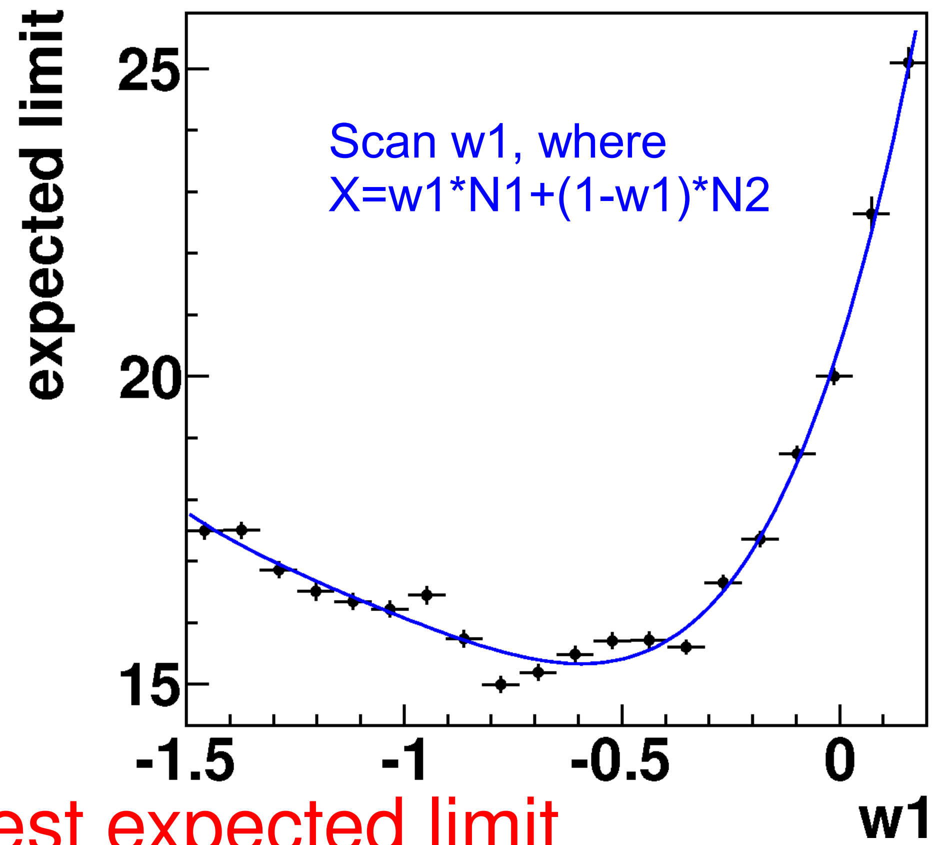
→ $w_i = \varepsilon_i / b_i$ is not the best choice,

best expected limit for $w_1 \sim 0.7$, $w_2 \sim 0.3$

Many channels + systematics (3)

- Example 2: two channels with correlated bgr systematics, one channel with low s/b, one channel with high s/b

	eff	bgr	Bgr norm. error
Channel 1	0.1	20.0	50%
Channel 2	0.9	10.0	



correlation: if bgr is high in channel 1 it is also high in channel 2

→ measure bgr from channel 1 and subtract from channel 2?

→ negative w_1~-0.6 (w_2~1.6) gives best expected limit

Many channels + systematics (4)

- No unique method to set limits for the multi-channel +sys case
- “Standard” method: profile likelihood (RooStat)
  - Use likelihood maximized wrt systematic parameters as test statistics
- Bayesian method: use marginalised likelihood + prior (RooStat)

- Alternative methods, e.g. based on weighted sums, $X = \sum w_i N_i^{\text{obs}}$ where bin weights w_i are optimised for syst. errors

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

Summary

- Basic concepts of setting limits:
  - Frequentist/Bayesian methods
- Examples for specific problems:
  - Signal plus background, expected limit, CL_s method
  - + systematic uncertainties
  - Combining several channels
  - + systematic uncertainties

Not covered:
Bayesian “objective” priors, etc
Discoveries: p-values, Bayes factors, etc
Standard tools in Root
… and many more things

➔ Limit calculation is a wide field. Impossible to do justice to all methods in a few hours