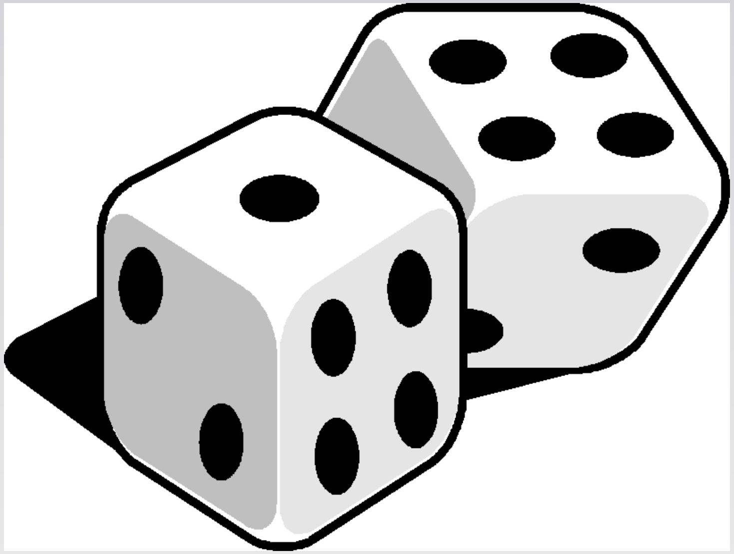
# Statistics In HEP 2

## How do we understand/interpret our measurements



# Outline

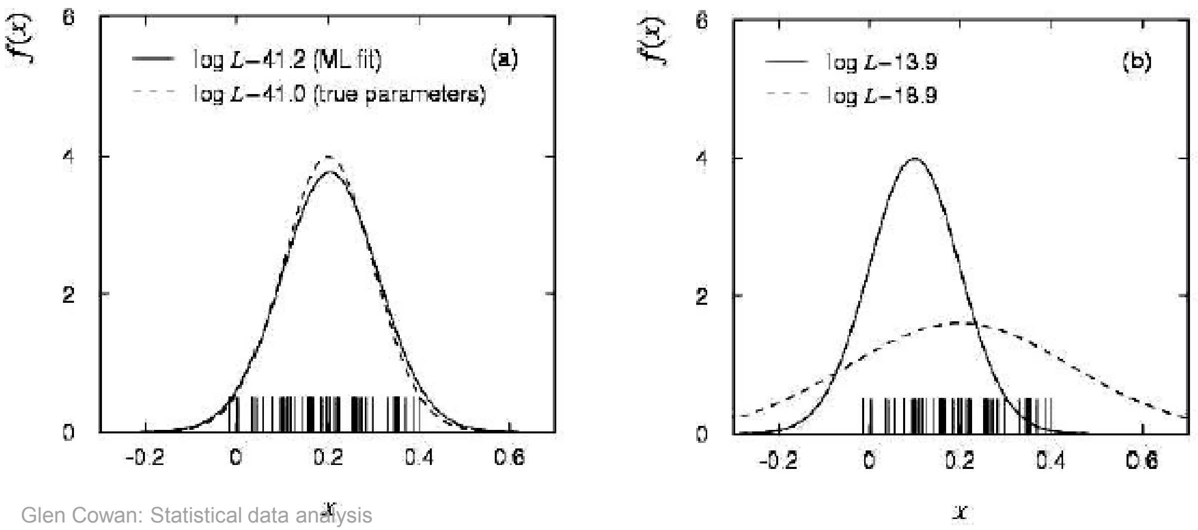
- Maximum Likelihood fit
- strict frequentist Neyman – confidence intervals
  - what “bothers” people with them
- Feldmans/Cousins  confidence belts/intervals
- Bayesian treatement of ‘unphysical’ results
- How the LEP-Higgs limit was derived
- what about systematic uncertainties?
  - Profile Likleihood

# Parameter Estimation

- want to measure/estimate some parameter $\theta$
  - e.g. mass, polarisation, etc..
- observe: $\vec{x}^i = (x_1, \ldots . x_n)_i \quad i = 1, K$
  - e.g $n$ observables for $K$ events
- "hypothesis" i.e. PDF $P(\vec{x};\theta)$ - distribution $\vec{x}$ for given $\theta$
  - e.g. diff. cross section

  → $K$ independent events: $P(\vec{x}^1, \ldots \vec{x}^K; \theta) = \prod_i^K P(\vec{x}^i; \theta)$

- for fixed $\vec{x}$ regard $P(\vec{x};\theta)$ as function of $\theta$ (i.e. Likelihood! $L(\theta)$ )
  - $\theta$ close to $\theta_{true}$ → Likelihood $L(\theta)$ will be large




Glen Cowan: Statistical data analysis

→ try to maximuse $L(\theta)$

→ typically:

→ minimize -2Log($L(\theta)$) → $\hat{\theta}$

→ Maximum Likelihood estimator

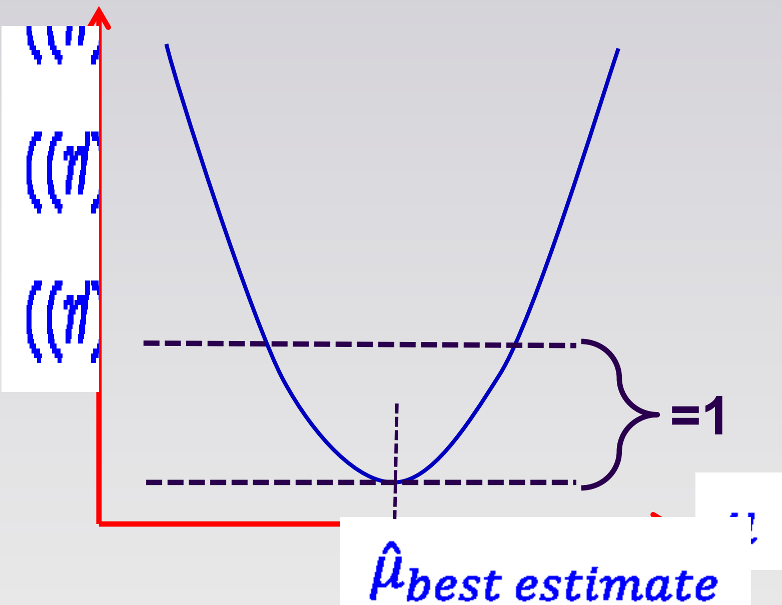
# Maximum Likelihood Estimator

**example:** PDF(x) = Gauss(x,μ,σ) → $L(x|Gauss(\mu)) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

→ estimator for $\mu_{true}$ from the data measured in an experiment $x_1, \ldots x_K$

→ full Likelihood $L(x|\mu) = \prod_i^K \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$

→ typically: $-2\ln(L(x|\mu)) = \sum_i^K \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)$ Note: It's a function of $\mu$ !




For Gaussian PDFs:

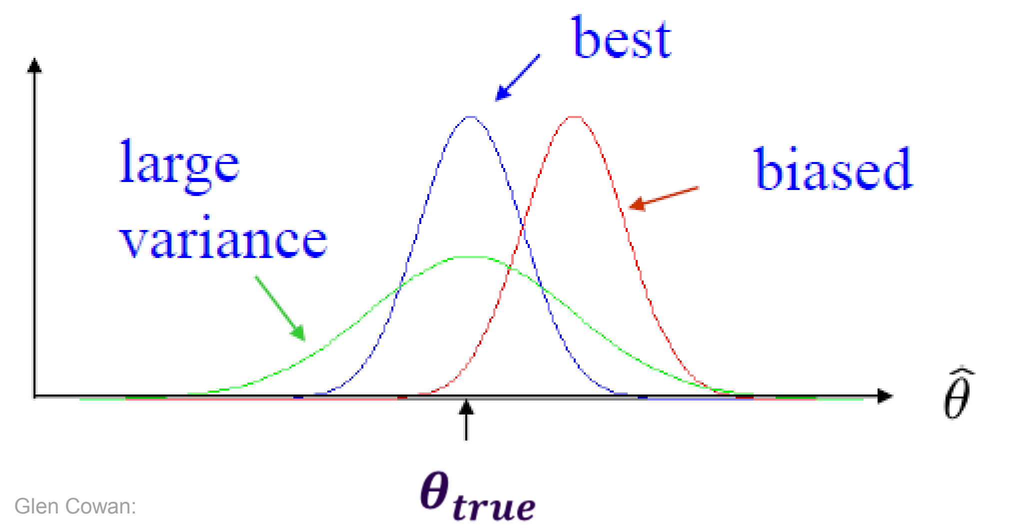
$-2\ln(L(\mu)) == \chi^2$

- $\Delta 2ln(L) = 1$ from $\hat{\mu}$ → interval $[\mu_1; \mu_2]$
→ variance on the estimate $\hat{\mu}$

## properties of estimators

- biased or unbiased
- large or small variance

→ distribution of $\hat{\theta}$ on many measurements ?




- Small bias and small variance are typically "in conflict"
- Maximum Likelihood is typically unbiased only in the limit $K \to \infty$
  - If Likelihood function is "Gaussian" (often the case for large K → central limit theorem)

→ get "error" estimate from or $-2\Delta log(L) = 1$

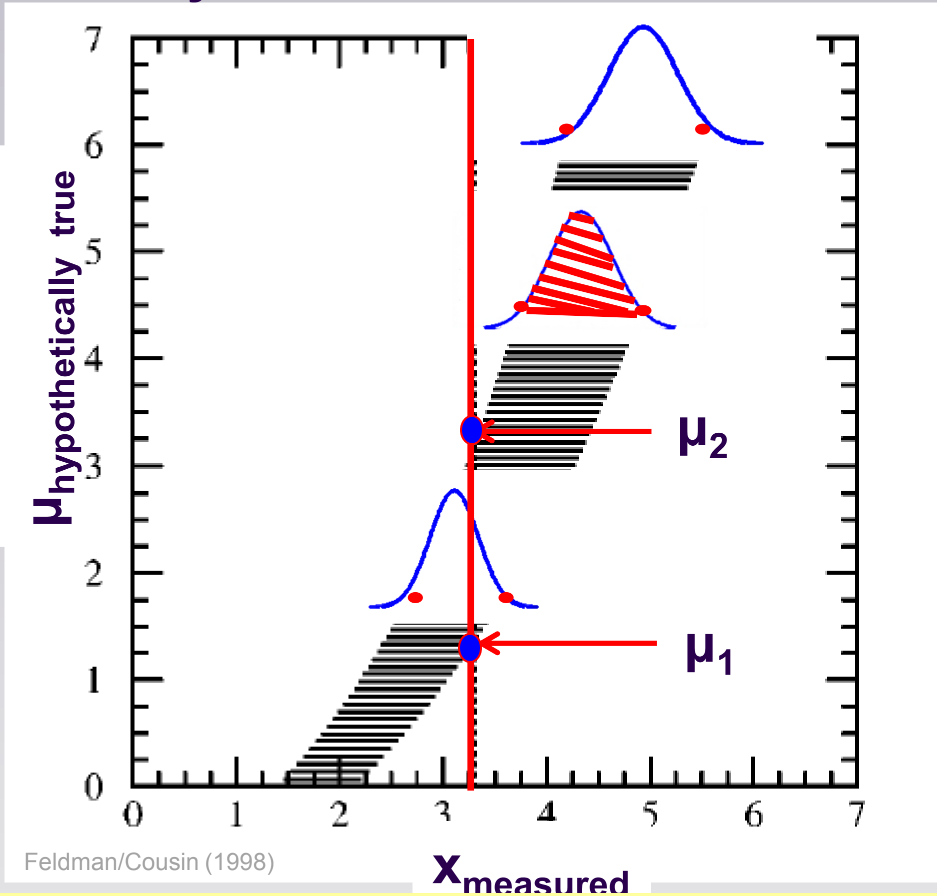
→ If (very) none Gausslan

→ revert typically to (classical) Neyman confidence intervals

# Classical Confidence Intervals

another way to look at a measurement rigorously "frequentist"

- Neymans Confidence belt for CL α (e.g. 90%)




- each $\mu_{hypothetically\ true}$ has a PDF of how the measured values will be distributed
- determine the (central) intervals ("acceptance region") in these PDFs such that they contain α
- do this for ALL $\mu_{hyp.true}$
- connect all the "red dots" → confidence belt

- measure $x_{obs}$ :

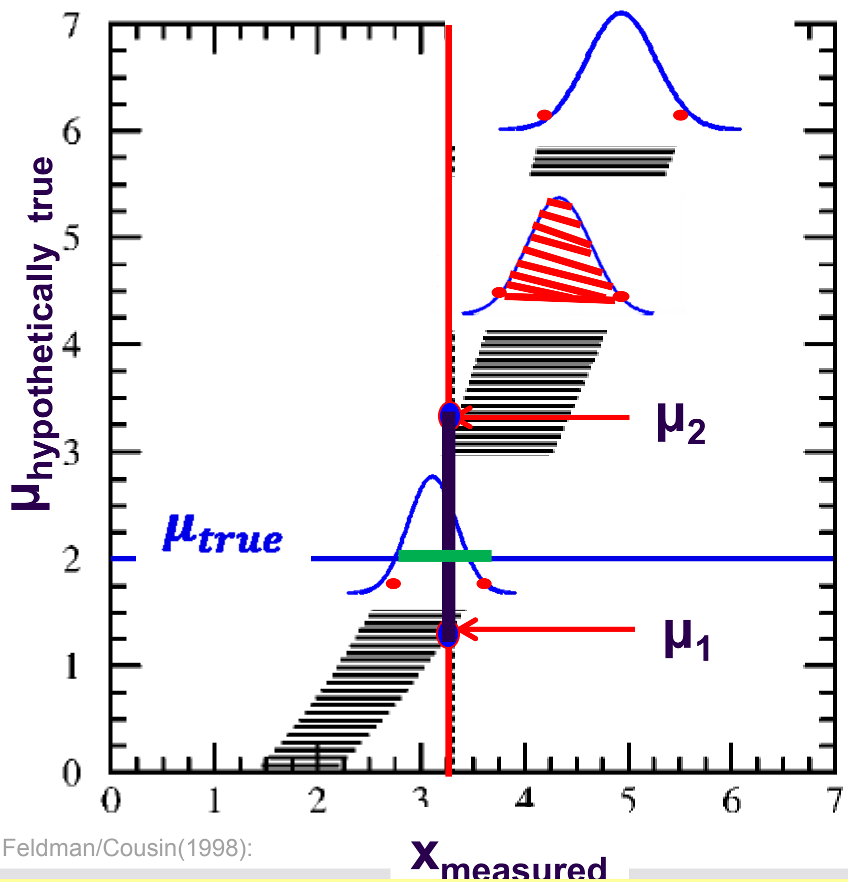
→ conf. interval =[$\mu_1$, $\mu_2$] given by vertical line intersecting the belt.

- by construction: for each $x_{meas.}$ (taken according PDF($\mu_{true}$) the confidence interval [$\mu_1$, $\mu_2$] contains $\mu_{true}$ in α = 90% cases

# Classical Confidence Intervals

another way to look at a measurement rigorously "frequentist"

- Neymans Confidence belt for CL α (e.g. 90%)



Feldman/Cousin(1998):

→conf.interval =[$\mu_1$, $\mu_2$] given by **vertical** line intersecting the belt.

- by construction:
  - $P(x < x_{obs}; \mu_2) = \frac{1-\alpha}{2}$
  - $P(x > x_{obs}; \mu_1) = \frac{1-\alpha}{2}$
- if the true value were $\mu_{true}$

→ lies in $[\mu_1, \mu_2]$ if it intersects ▌

→ $x_{meas}$ intersects ▬ as in 90% (that's how it was constructed)

→ only those $x_{meas}$ give $[\mu_1, \mu_2]$'s that intersect with the ▬

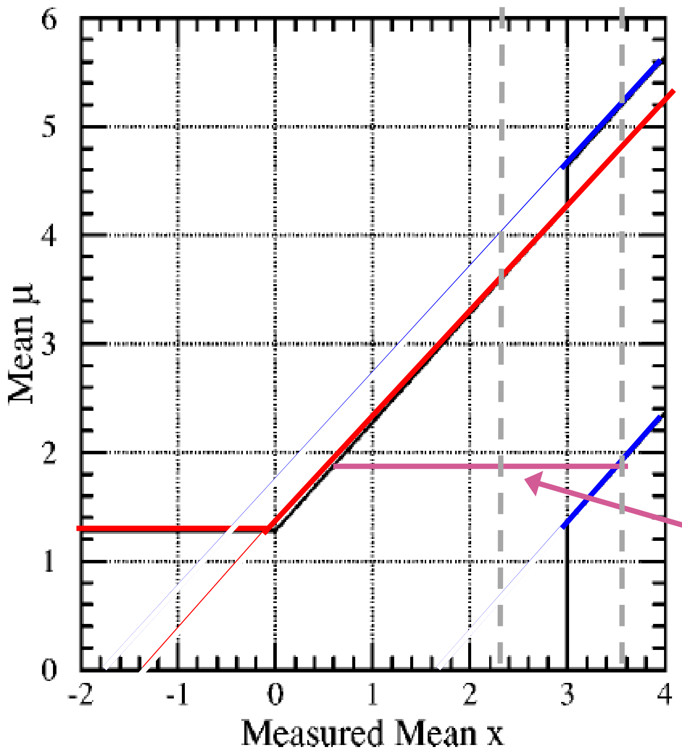
→ 90% of intervals cover $\mu_{true}$

- $P(x; \mu)$ is Gaussian ($\sigma = const$) → central 68% Neyman Conf. Intervals ⇔ Max. Likelihood + its "error" estimate $[\hat{x} - \sigma_{\hat{x}}; \hat{x} + \sigma_{\hat{x}}]$

# Flip-Flop

**When to quote measuremt or a limit!**

- estimate Gaussian distributed quantity $\mu$ that cannot be < 0 (e.g. mass)
- same Neuman confidence belt construction as before:
  - once for measurement (two sided, each tail contains 5% )
  - once for limit (one sided tails contains 10%)




Feldman/Cousins(1998)

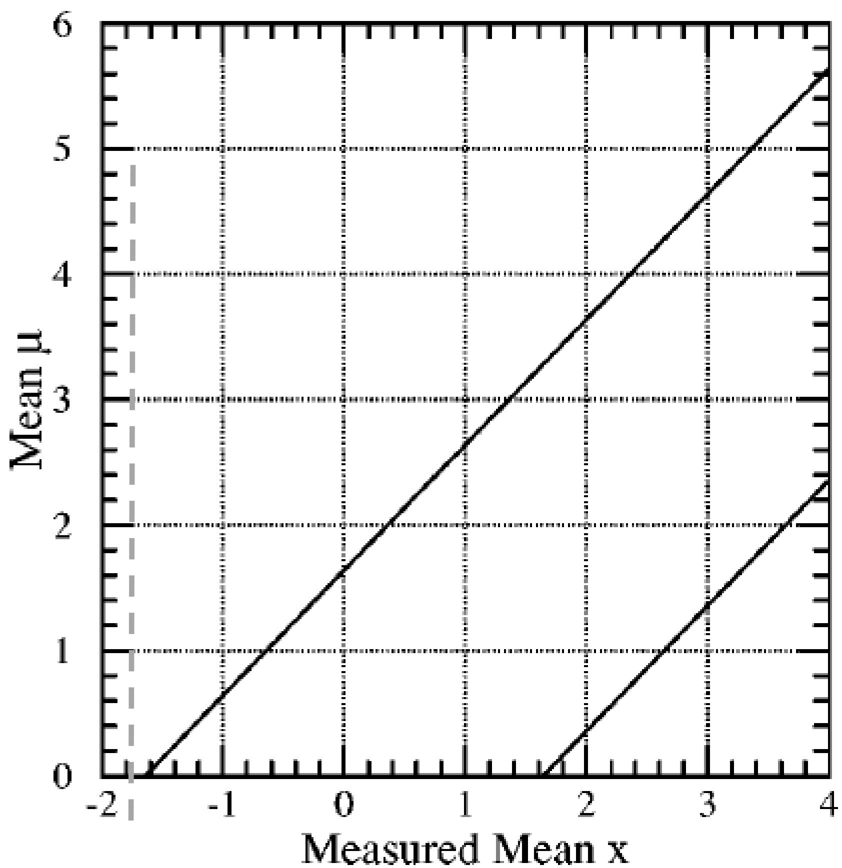
- decide: if $x_{obs}$<0 assume you =0
  → conservative
- if you observe $x_{obs}$ <3
  → quote upper limit only
- if you observe $x_{obs}$>3
  → quote a measurement

→ induces "undercovering" as this acceptance region contains only 85% !!

# Some things people don't like..

**same example:**

- **estimate Gaussian distributed quantity $\mu$ that cannot be < 0 (e.g. mass)**




Feldman/Cousins(1998).

- **using proper confidence belt**
- **assume: $x_{obs} = -1.8$**

  **→ confidence interval is EMPTY!**

- **Note: that's OK from the frequentist interpretation $\mu_{true} \in [conf.interv.]$ in 90% of (hypothetical) measurements.**

**Obviously we were 'unlucky' to pick one out of the remaining 10%**

- **nontheless: tempted to "flip-flop" ??? tsz .. tsz.. tsz..**

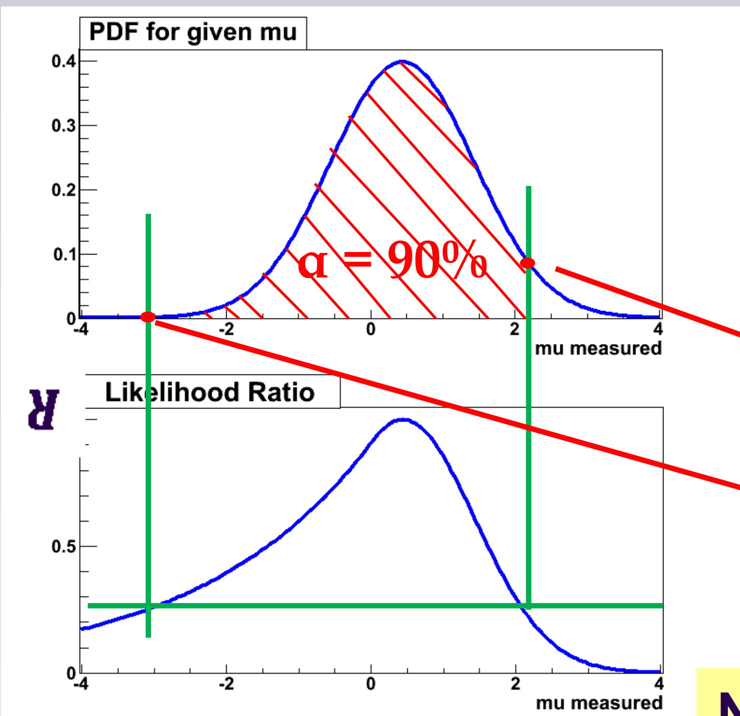
# Feldman Cousins: a Unified Approach

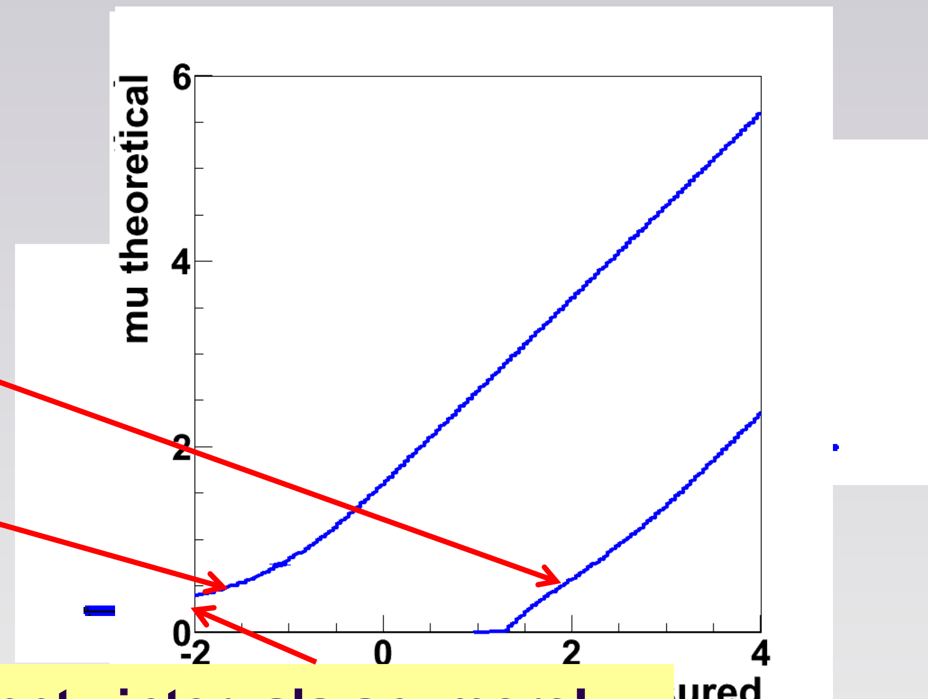
- How we determine the "acceptance" region for each $\mu_{hyp.true}$ is up to us as long as it covers the desired integral of size α (e.g. 90%)

→ include those "$x_{meas.}$" for which the large likelihood ratio first:

$$R = \frac{L(x_{meas}|\mu_{measured})}{L(x_{meas}|\mu_{best\ estimate})}$$

- $\mu_{best\ estimate}$ here: either the observation $x_{meas}$ or the closes ALLOWED $\mu$

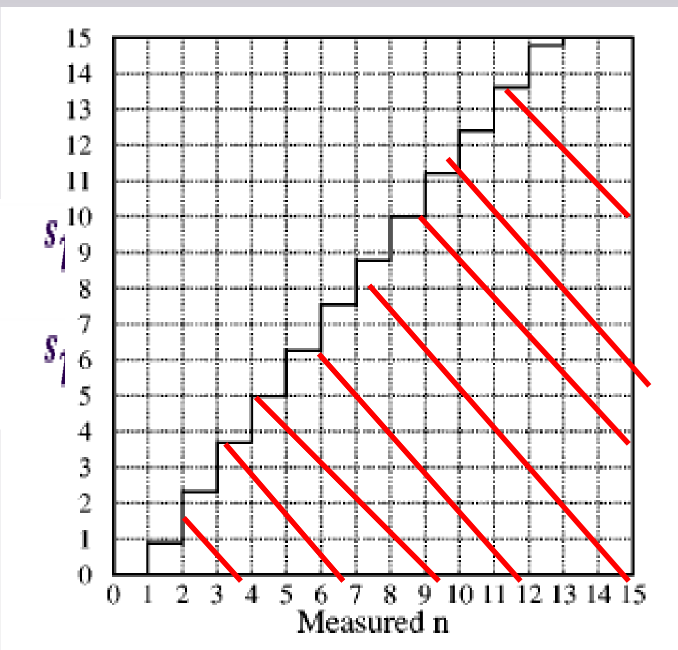



No "empty intervals anymore!

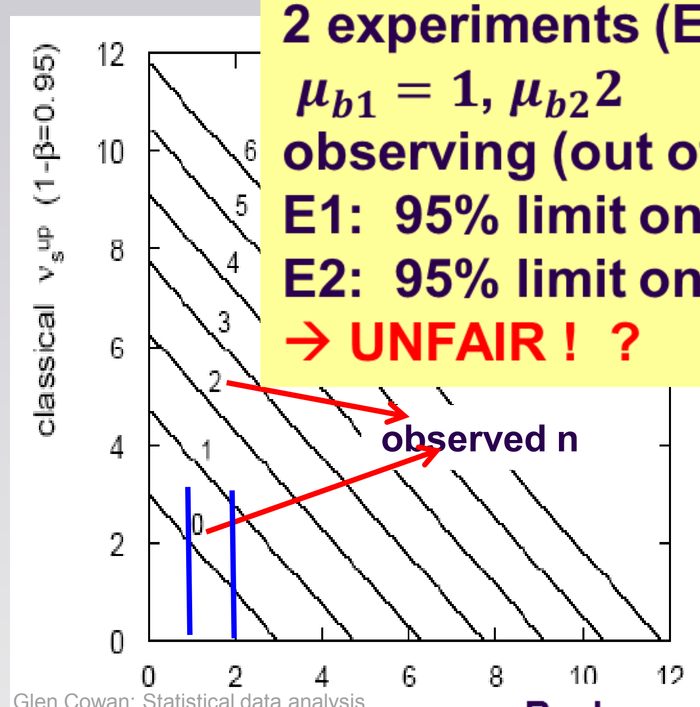
# Being Lucky…

- give upper limit on signal $\mu_s$ on top of know (mean) background $\mu_b$
    - → n=s+b from a possion distribution
    - $P(n) = Poisson(n, \mu_s + \mu_b)$
- Neyman: draw confidence belt with
    - "$\mu_s$" in the "y-axis" (the possible true values of $\mu_s$)




for fixed Background $\mu_b$

sorry… the plots don't match: one is of 90%CL the other for 95%CL




Background $\mu_b$

2 experiments (E1 and E2)
$\mu_{b1} = 1, \mu_{b2} 2$
observing (out of luck) 0
E1: 95% limit on $\mu_s \sim 2$
E2: 95% limit on $\mu_s \sim 1$
→ UNFAIR ! ?

# Being Lucky …

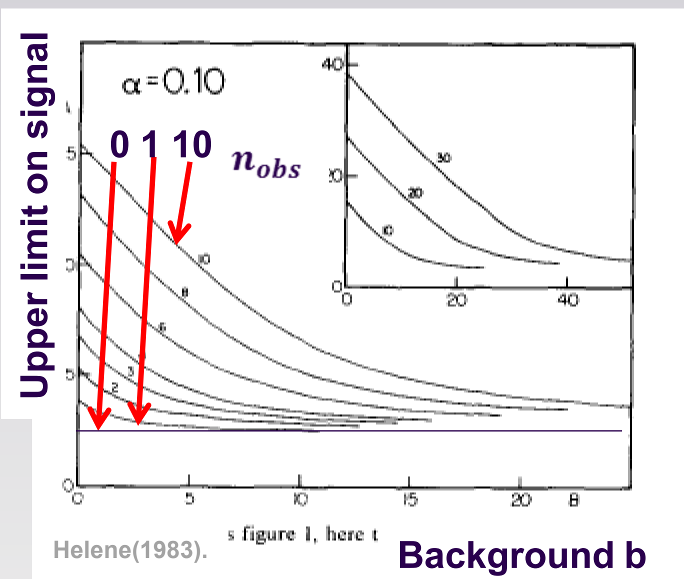
- Feldman/Cousins confidence belts
  →motivated by "popular" 'Bayesian' approaches  to handle such problems.

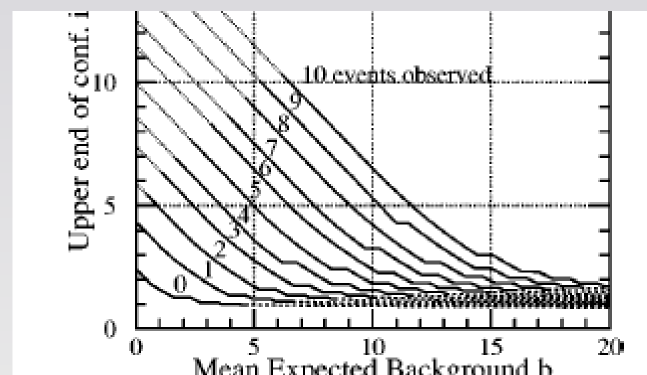
**Bayesian:  rather than constructing Confidence belts:**

- turn Likelihood for $\mu_s$ (on given $n_{obs}$)  into Posterior probability $on\,\mu_s$
  $i.e\ Poisson(n_{obs}; \mu_s + \mu_b)$
- $p(\mu_s|n_{obs}) = L(n_{obs}; \mu_s) * \pi(\mu_s)$add prior probability on "s":
- $\pi(\mu_s) = \begin{cases} 0 & \mu_s < 0 \\ uniform & \mu_s > 0 \end{cases}$

**Feldman/Cousins**

- there is still SOME "unfairness"
- perfectly "fine" in frequentist interpretation:
- should quote "limit+sensitivity"




Feldman/Cousins(1998).

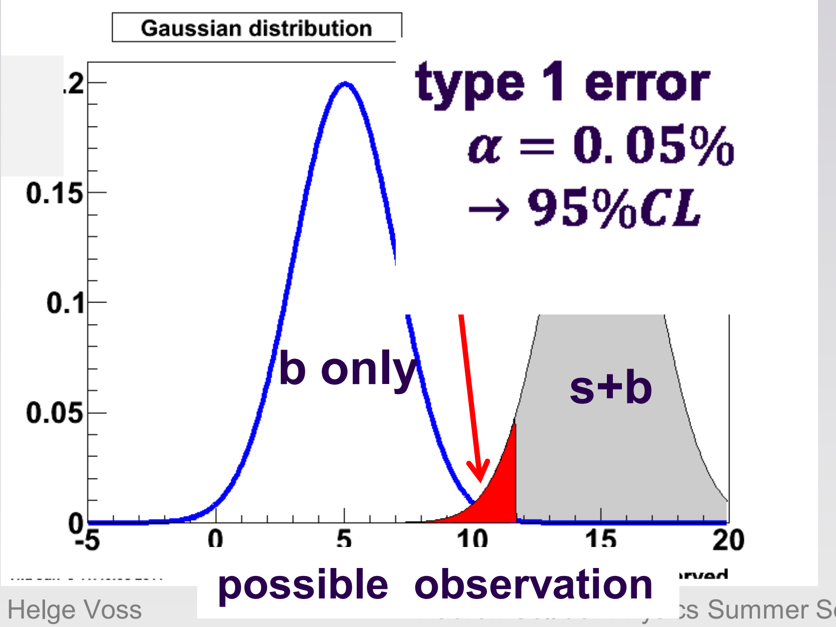
**exclusion limits**

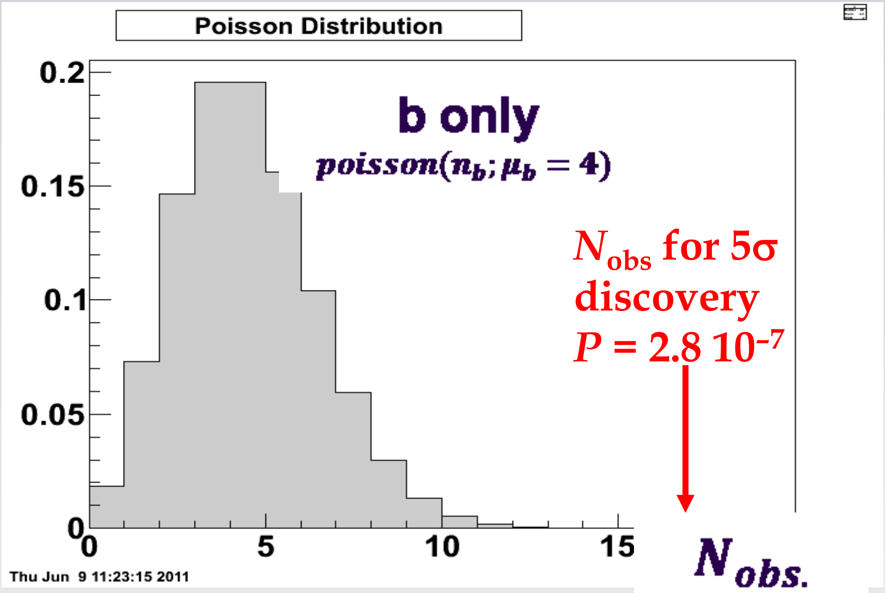
- **upper limit on cross section** (↔lower limit on mass scale)
- (σ < limit as otherwise we would have seen it)

- need to estimate probability of **downward** fluctuation of **s+b**
- try to **"disprove"** $H_0$ = s+b
- better: find **minimal s**, for which you can sill exclude $H_0$ = s+b a pre-specified **Confidence Level**

**discoveries**

- need to estimate probability of **upward** fluctuation of **b**
- try to **disprove** $H_0$ **="background only"**



# Which Test Statistic to used?



**exclusion limit:**

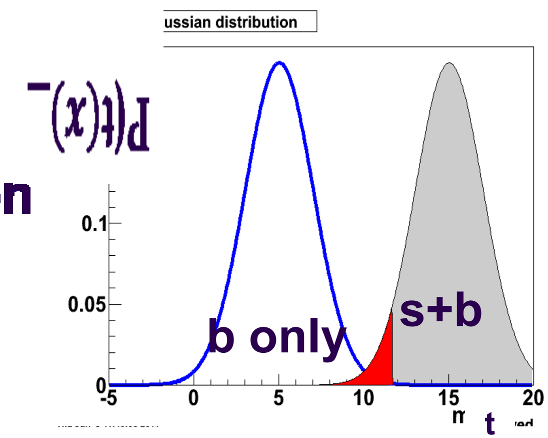
- test statistic does not necessarily have to be simply the counted number of events:
  - → remember Neyman Pearson → Likelihood ratio

→ $t(n_{obs}) = \frac{Poisson(n_{obs};s,b)}{Poisson(n_{obs};b)}$

→ pre-specify $\alpha = 0.05\%$ → 95% CL on exclusion

→ make your measurements

→ If accepted (i.e. in "chritical (red) region") where you decided to "recjet" $H_0$ = s+b



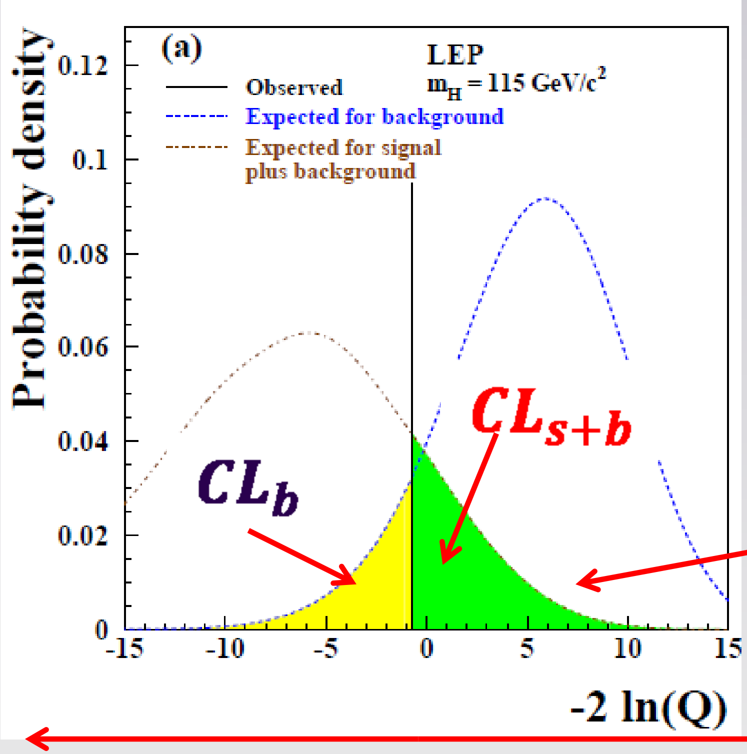

→ $CL_{s+b} = P(t < t_{obs})$

→ (i.e. what would have been the chance for THIS particular measurement to still have been "fooled" and there would have actually BEEN a signal)

# Example :LEP-Higgs search

**Remember: there were 4 experiments, many different search channels**
**→ treat different exerpiments just like "more channels"**

$$Q = \frac{L(x|H_1)}{L(x|H_0)} = \frac{\prod_i^{N_{chan}} Pois(n_i|s_i+b_i) \prod_j^{n_i} \frac{s_i f_s(x_{ij}) + b_i f_b(x_{ij})}{s_i+b_i}}{\prod_i^{N_{chan}} Pois(n_i|b_i) \prod_j^{n_i} f_b(x_{ij})}$$

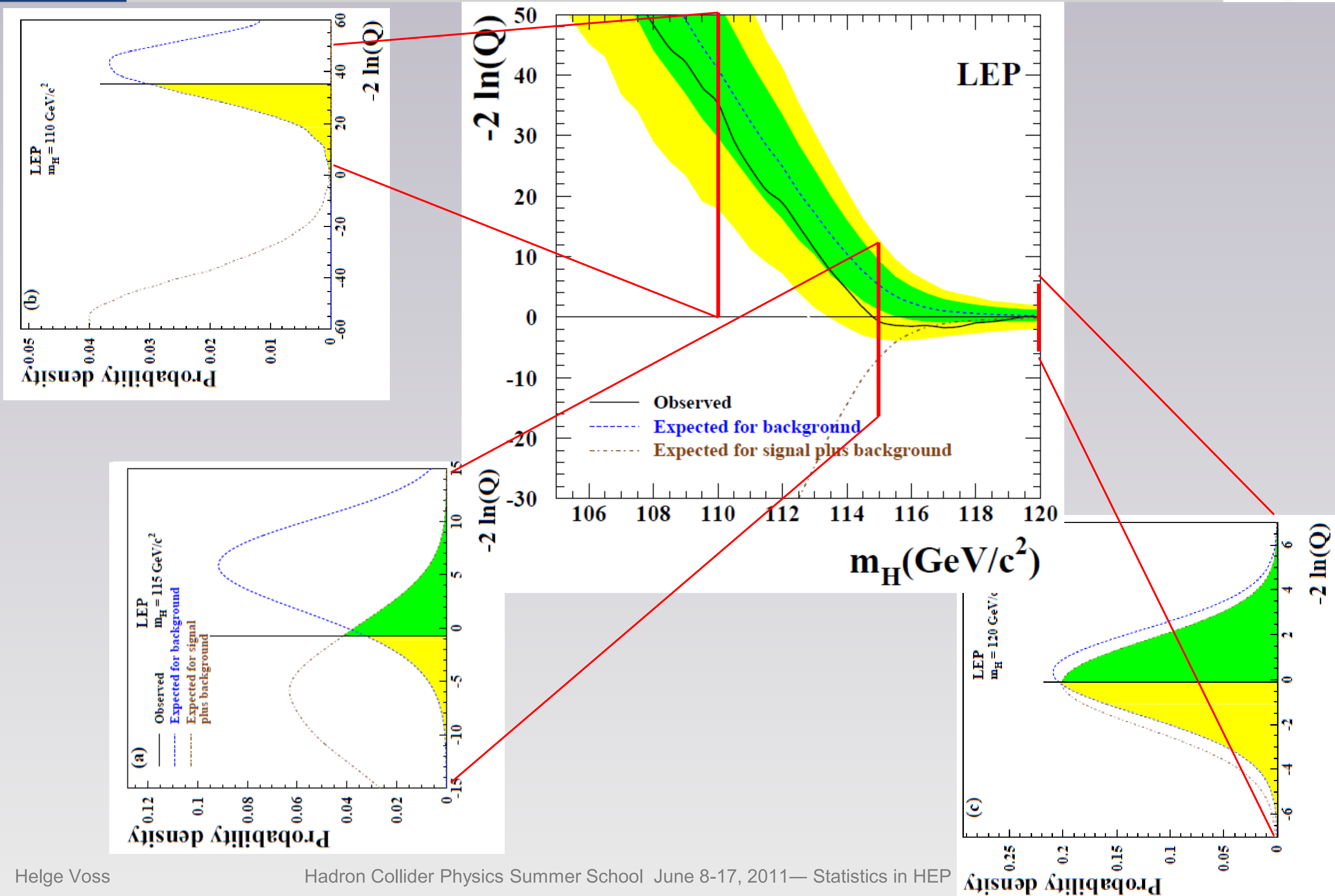



$$q = \ln Q = -s_{tot} + \sum_i^{N_{chan}} \sum_j^{n_i} \ln\left(1 + \frac{s_i f_s(x_{ij})}{b_i f_b(x_{ij})}\right)$$

- Evaluate how the -2lnQ is distributed for
  - background only
  - signal+background
  - (note: needs to be done for all Higgs masses)
- example: $m_H$=115GeV/c²

more signal like ⟷ more background like

# Example: LEP SM Higgs Limit



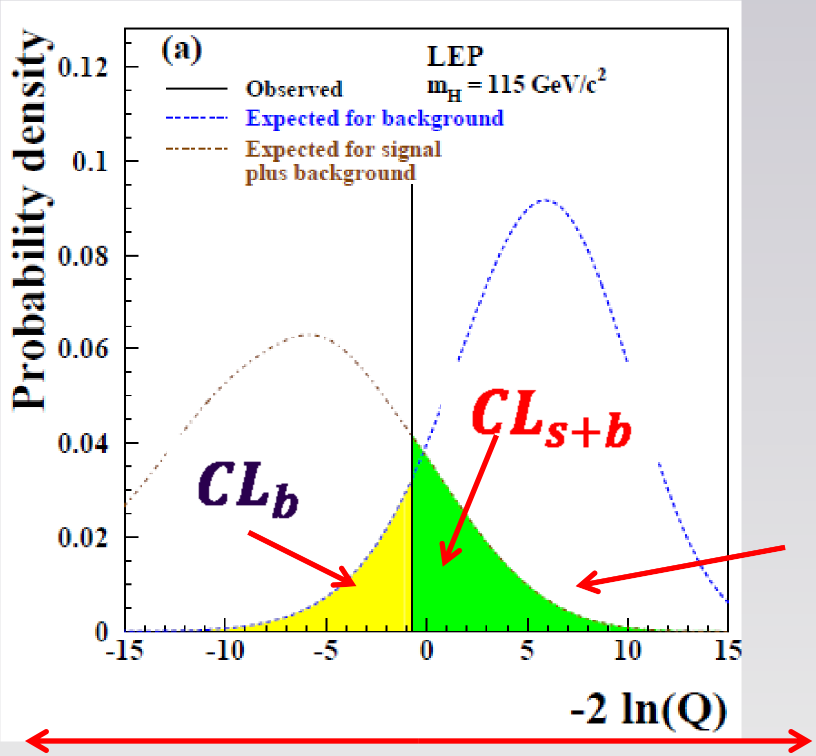
# Example LEP Higgs Search

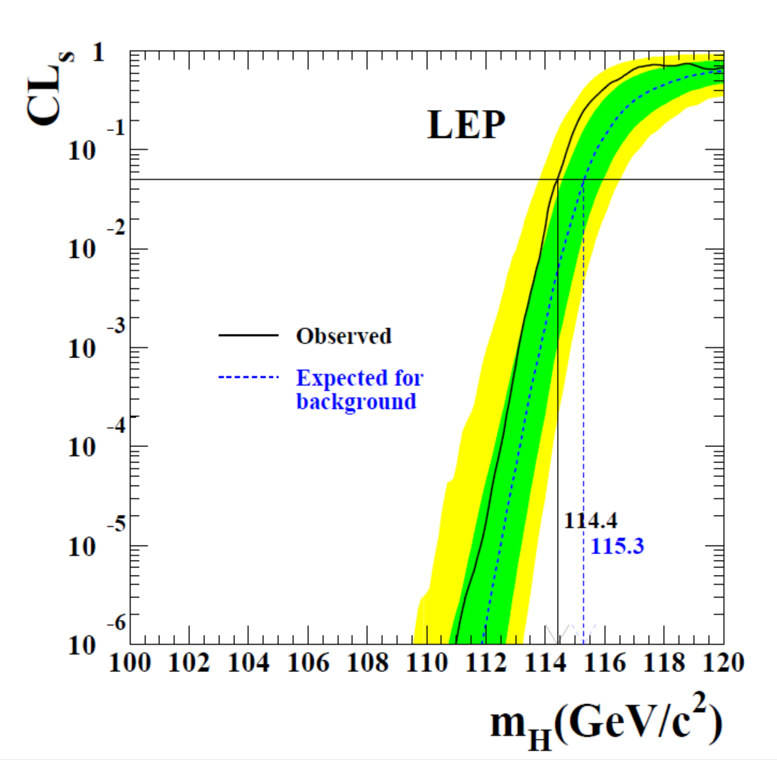
- In order to “avoid” the possible “problem” of Being Lucky when setting the limit

→ rather than “quoting” in addition the expected sensitivity

→ weight your CLs+b by it:

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{P(LLR \geq LLR_{obs}|H_1)}{P(LLR \leq LLR_{obs}|H_0)}$$




more signal like ⟷ more background like

# Systmatic Uncertainties

- standard popular way: (Cousin/Highland)
  - integrate over all systematic errors and their "Probability distribution)
  - → marginalisation of the "joint probability density of measurement paremters and systematic error)

! Bayesian ! (probability of the systematic parameter)

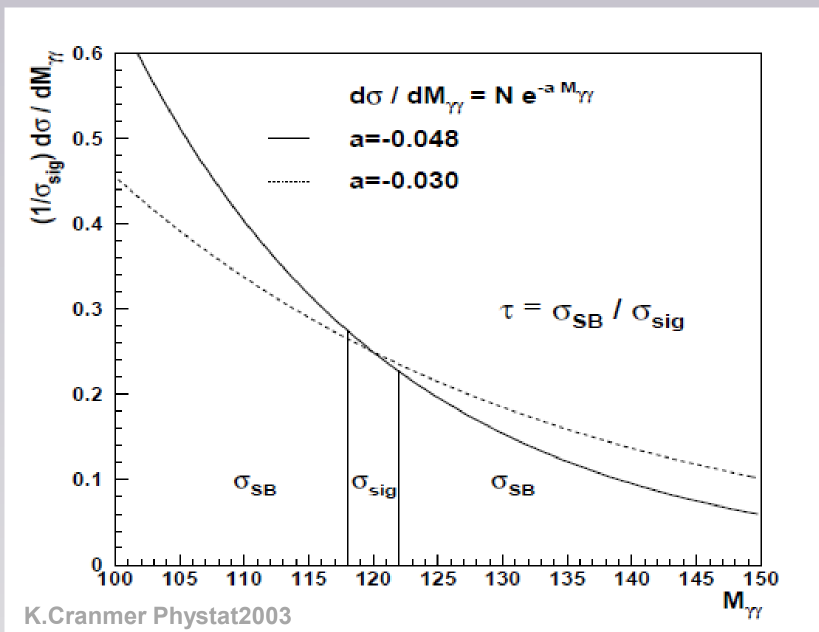
  - "hybrid" frequentist intervals and Bayesian systematic
  - has been shown to have possible large "undercoverage" for very small p-values /large significances (i.e. underestimate the chance of "false discovery" !!

- LEP-Higgs: generaged MC to get the PDFs with "varying" param. with systematic uncertainty

→ essentiall the same as "integrating over" → need probability density for "how these parameters vary"

# Systematic Uncertainties

- Why don't we:
  - include any systematic uncertainly as "free parameter" in the fit




- eg. measure background contribution under signal peak in sidebands
- measurement + extrapolation into side bands have uncertainty
- but you can parametrise your expected background such that:

→ if sideband measurement gives this data → then b=…

**Note: no need to specify prior probability**

$$\underbrace{P(n_{\rm on}, n_{\rm off}|s,b)}_{\text{joint model}} = \underbrace{\text{Pois}(n_{\rm on}|s+b)}_{\text{main measurement}}\ \underbrace{\text{Pois}(n_{\rm off}|\tau b)}_{\text{sideband}}$$

- Build your Likelyhood function such that it includes:
  - your parameters of interest
  - those describing the influcene of the sys. uncertainty
  - → nuisance parameters

# Nuisance Parameters and Profile Liklihood

- Build your Likelyhood function such that it includes:
  - your parameters of interest
  - those describing the influcene of the sys. uncertainty
  - → nuisance parameters

$$\underbrace{P(n_{\text{on}}, n_{\text{off}}|s,b)}_{\text{joint model}} = \underbrace{\text{Pois}(n_{\text{on}}|s+b)}_{\text{main measurement}}\ \underbrace{\text{Pois}(n_{\text{off}}|\tau b)}_{\text{sideband}}$$

$$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}, & 0 \le \hat{\mu} \le \mu \\[2ex] \dfrac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})}, & \hat{\mu} < 0 \end{cases} = \frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}, \quad 0 \le \hat{\mu} \le \mu$$

"ratio of likelihoods", why ?

# Profile Likelihood

$$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}, & 0 \le \hat{\mu} \le \mu \\ \dfrac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\theta})}, & \hat{\mu} < 0 \end{cases} = \dfrac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}, \quad 0 \le \hat{\mu} \le \mu$$

"ratio of likelihoods", why ?

Why not simply using $L(\mu,\theta)$ as test statistics ?

- The number of degrees of freedom of the fit would be $N_\theta + 1_\mu$
- However, we are **not** interested in the values of $\theta$ (→ they are *nuisance* !)
- Additional degrees of freedom dilute interesting information on $\mu$
- The "profile likelihood" (= ratio of maximum likelihoods) concentrates the information on what we are interested in
- It is just as we usually do for chi-squared: $\Delta\chi^2(m) = \chi^2(m,\theta_{best'}) - \chi^2(m_{best},\theta_{best})$
- $N_{d.o.f.}$ of $\Delta\chi^2(m)$ is 1, and value of $\chi^2(m_{best},\theta_{best})$ measures "Goodness-of-fit"

# Summary

- **Maximum Likelihood fit    to estimate paremters**

- **what to do if estimator is non-gaussion:**
  - **Neyman – confidence intervals**
  - **what "bothers" people with them**
- **Feldmans/Cousins  confidene belts/intervals**
  - **unifies "limit" or "measurement" confidence belts**
- **CLs … the HEP limit;**
  - **CLs  … ratio of "p-values" … statisticians don't like that**
  - **new idea: Power Constrained limits**
    - rather than specifying "sensitivity" and "neymand conf. interval"
    - decide beforehand that you'll "accept" limits only if the where your exerpiment has sufficient "power"  i.e. "sensitivity !
- **.. a bit about Profile Likelihood, systematic error.**