Karlsruhe Institute of Technology

Institut für Experimentelle Teilchenphysik

Teilchenphysik II - W, Z, Higgs am Collider

Lecture 08: Statistical Analysis

PD Dr. K. Rabbertz, **Dr. Nils Faltermann** | 16. Juni 2023

Statistics and Particle Physics

Experiment

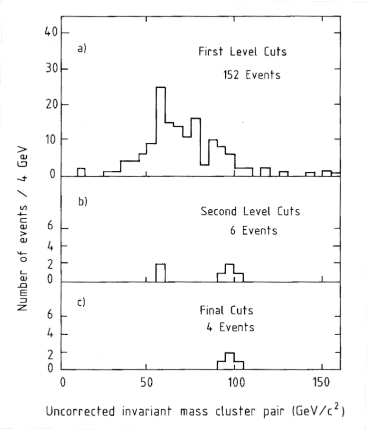
- All measurements are derived from **counting experiments**
- **Millions of billions** of particle collision events each year
- Each event **independent** from and (to many extents) identical to all others

Theory

- Nature intrinsically stochastic
- QM wave functions interpreted as **probability density functions**
- Event-by-event simulation using **Monte Carlo** methods

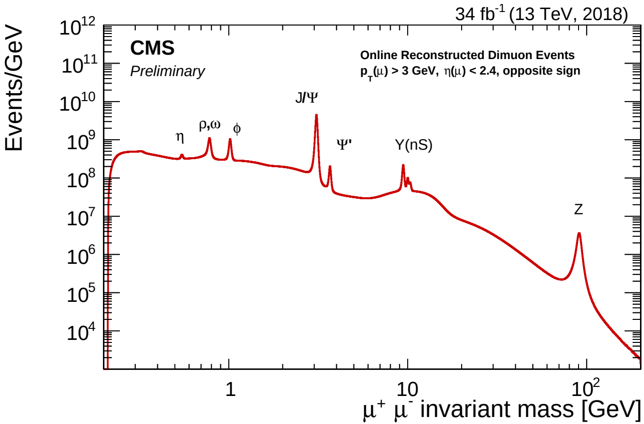
Particle physics experiments perfect application of statistics

Statistics and Particle Physics



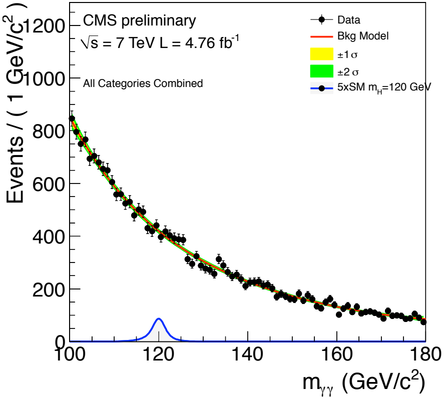
Historically: trust your instincts

Statistics and Particle Physics



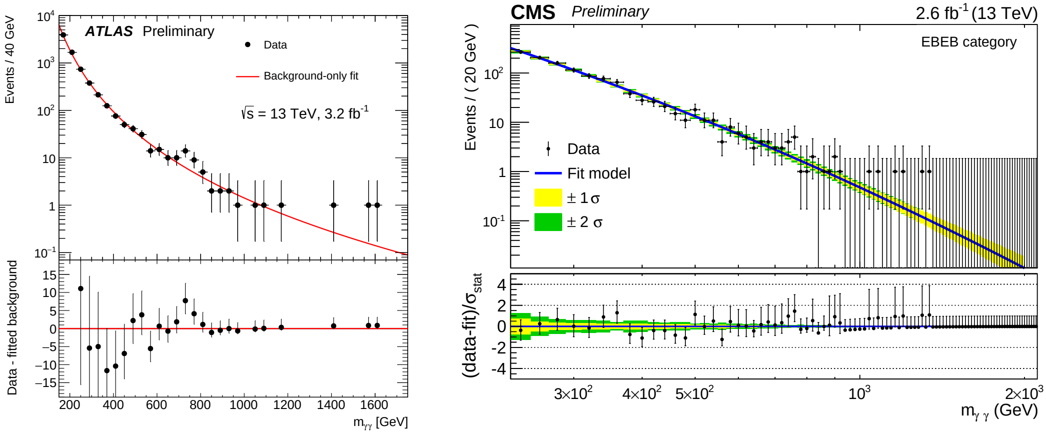
Sometimes unambiguous, clearly there is something!

Statistics and Particle Physics



But most of the time not: is there a signal?

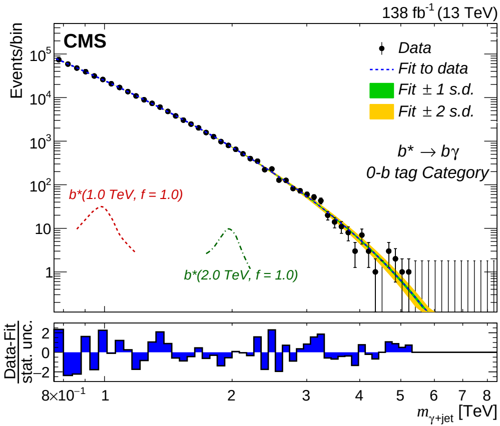
Statistics and Particle Physics



And here at around 750 GeV?

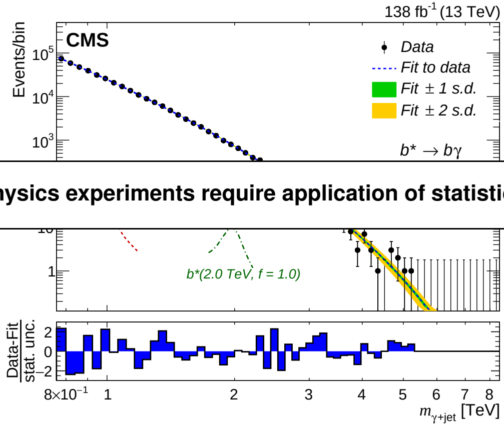
(with more subsequent data: turned out to be nothing)

Statistics and Particle Physics



And what do you do if you don't see anything: just a waste of time?

Statistics and Particle Physics



Particle physics experiments require application of statistical methods

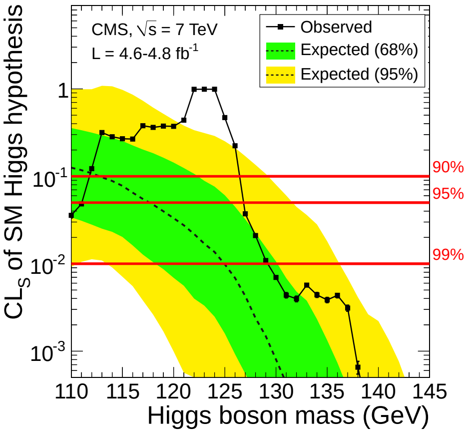
And what do you do if you don’t see anything: just a waste of time?

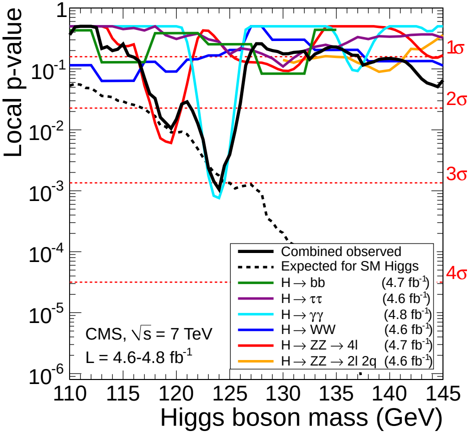
Outline

- Statistics is a vast field!
- Here: focus on concepts and techniques needed to **understand experimental particle physics results**
- Questions we will discuss
    - How to interpret the data: are they compatible with my expectation (model)?
    - How do uncertainties enter the game?
    - Can we distinguish between different models?
    - If we don't see any signal, can we still use the data to make statements about the searched-for process?
- Much more details in lecture **Moderne Methoden der Datenanalyse** (summer term)

Outline

- Goal: we want to understand these plots





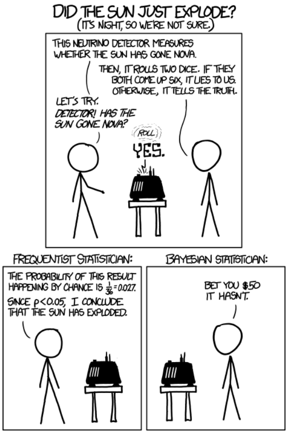
Phys.Lett. B710 (2012) 26–48

Bayesian and Frequentist Approaches

- Two major approaches in statistical analysis:
  - **Frequentist**: probability purely based on observation, objective
  - **Bayesian**: usage of earlier knowledge (prior), subjective
- Will focus on frequentist approach (but for most concepts Bayesian equivalent, see the literature!)

Bayesian and Frequentist Approaches

- Two major approaches in statistical analysis:
  - **Frequentist**: probability purely based on observation, objective
  - **Bayesian**: usage of earlier knowledge (prior), subjective
- Will focus on frequentist approach (but for most concepts Bayesian equivalent, see the literature!)



https://xkcd.com/1132/

Important Probability Density Functions

- Statistical processes underlying the data are assumed to follow a probability density function (pdf) $\mathcal{P}$

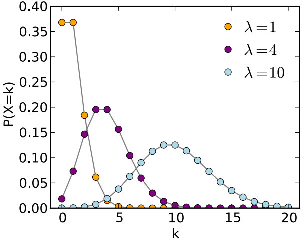
Important Probability Density Functions

- Statistical processes underlying the data are assumed to follow a probability density function (pdf) $\mathcal{P}$

Poisson:

$$\boxed{\mathcal{P}(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}}$$

e. g. counting experiments
(like cross-section measurements)



wikipedia

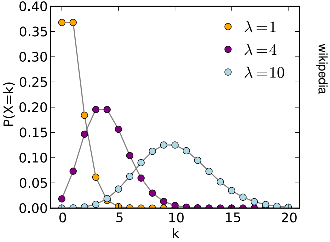
Important Probability Density Functions

- Statistical processes underlying the data are assumed to follow a probability density function (pdf) $\mathcal{P}$

Poisson:

$$\mathcal{P}(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$$

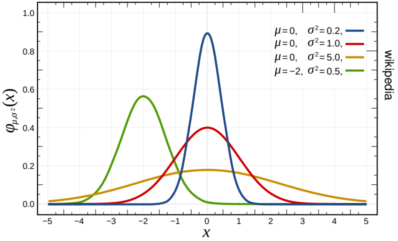
e. g. counting experiments
(like cross-section measurements)



Gaussian:

$$\mathcal{P}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

e. g. parameter estimates
(like mass measurements)



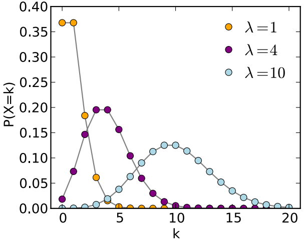
Important Probability Density Functions

- Statistical processes underlying the data are assumed to follow a probability density function (pdf) $\mathcal{P}$

Poisson:

$$\mathcal{P}(x;\lambda) = \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}$$

e. g. counting experiments
(like cross-section measurements)

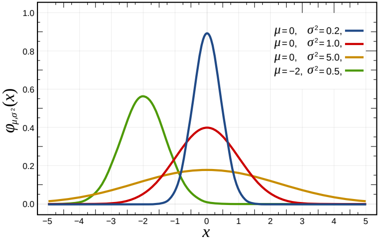


wikipedia

Gaussian:

$$\mathcal{P}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

e. g. parameter estimates
(like mass measurements)



wikipedia

NB: both are in fact limits of the Binomial distribution for a large number of tries

Likelihood

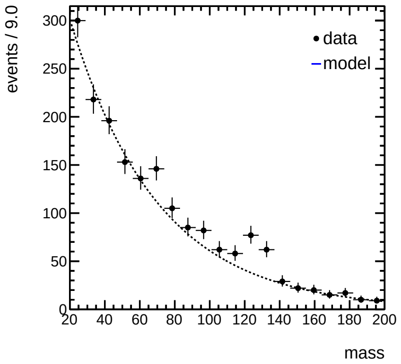
- **The likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},\,i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  $\rightarrow$ **function of model parameters and observed data**

Likelihood

- **The likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  $\rightarrow$ **function of model parameters and observed data**
- Example
  - Data: observed events in 20 bins
  - Model: known SM process (background)

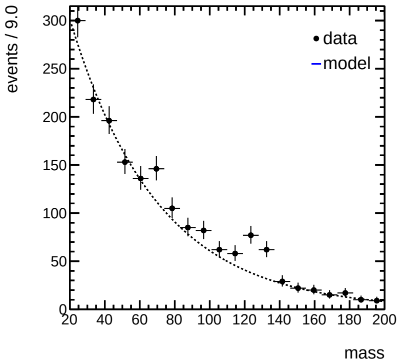


Likelihood

- **The likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  $\rightarrow$ **function of model parameters and observed data**
- Example
  - Data: observed events in 20 bins
  - Model: known SM process (background)

$$\mathcal{L}(n_{\text{obs}}; b) \quad = \prod_i \text{Poisson}(n_{\text{obs},i}, b)$$



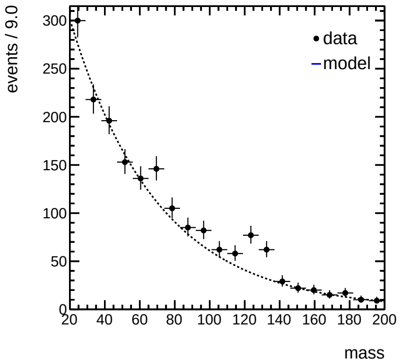
Likelihood

- **The likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  $\rightarrow$ **function of model parameters and observed data**
- Example
  - Data: observed events in 20 bins
  - Model: known SM process (background)

$$\mathcal{L}(n_{\text{obs}}; b) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} e^{-n_{\text{pred},i}}$$

$$n_{\text{pred},i}(b) = \underbrace{b \cdot e^{-\alpha x_i}}_{\text{background}}$$



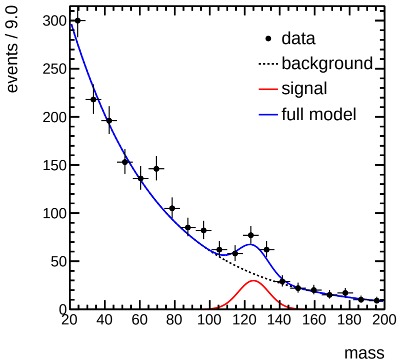
Likelihood

- **The likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  $\rightarrow$ **function of model parameters and observed data**
- Example
  - Data: observed events in 20 bins
  - Model: known SM process (background) + signal

$$\mathcal{L}(n_{\text{obs}}; b, s) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} e^{-n_{\text{pred},i}}$$

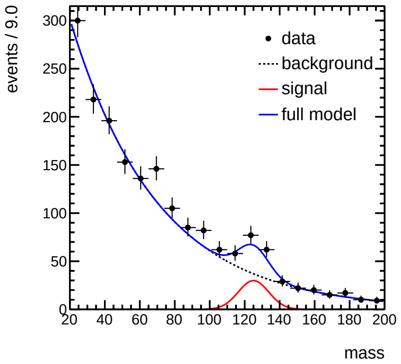
$$n_{\text{pred},i}(b, s) = \underbrace{b \cdot e^{-\alpha x_i}}_{\text{background}} + \underbrace{s \cdot e^{-(m-x_i)^2}}_{\text{signal}}$$



Likelihood

- **The likelihood** $\boxed{\mathcal{L}(n_{\text{obs}}; k) = \prod_i \mathcal{P}(n_{\text{obs},i}; k)}$ **quantifies compatibility of the observed data with a given model**
  - i: independent counting experiments, e. g. bins of a histogram
  - $n_{\text{obs}} = \{n_{\text{obs},i}\}$: number of observed events in i, distributed as $\mathcal{P}$
  - $k = \{k_j\}$: model parameters in $\mathcal{P}$

  $\rightarrow$ **function of model parameters and observed data**
- $\mathcal{L}$ is **not the probability** of a model!
  - Also not the probability of observing the data given the model
- $\mathcal{L}$ is the product of pdfs of the model, used to quantify the compatibility of the data with the model



Parameter Estimation

- Which parameter values of a given model describe best the data?

Parameter Estimation

- Which parameter values of a given model describe best the data?
- **Maximum likelihood (ML) method** (“maximum likelihood fit”):
  Which parameter values lead to the maximum of $\mathcal{L}$?

$\rightarrow$ **Maximum likelihood estimators $\hat{\mathbf{k}}$** of the true values

$$\boxed{\max\left[\mathcal{L}(n_{\text{obs}}, k)\right] = \mathcal{L}(n_{\text{obs}}, \hat{k})}$$

Parameter Estimation

- Which parameter values of a given model describe best the data?
- **Maximum likelihood (ML) method** ("maximum likelihood fit"):
  Which parameter values lead to the maximum of $\mathcal{L}$?

→ **Maximum likelihood estimators $\hat{\mathbf{k}}$** of the true values

$$\boxed{\max\left[\mathcal{L}(n_{\text{obs}}, k)\right] = \mathcal{L}(n_{\text{obs}}, \hat{k})}$$

- Technically: minimisation of negative log-likelihood

$$\text{NLL}(n_{\text{obs}}, k) = -2\ln\mathcal{L}$$

  - Logarithm is monotonic: retains minimum
  - Turns product of pdfs into sum: easier to minimise

Parameter Estimation

- Which parameter values of a given model describe best the data?
- **Maximum likelihood (ML) method** ("maximum likelihood fit"):
  Which parameter values lead to the maximum of $\mathcal{L}$?

$\rightarrow$ **Maximum likelihood estimators $\hat{\mathbf{k}}$** of the true values

$$\boxed{\max\left[\mathcal{L}(n_{\mathrm{obs}}, k)\right] = \mathcal{L}(n_{\mathrm{obs}}, \hat{k})}$$

- Technically: minimisation of negative log-likelihood

$$\mathrm{NLL}(n_{\mathrm{obs}}, k) = -2 \ln \mathcal{L}$$

  - Logarithm is monotonic: retains minimum
  - Turns product of pdfs into sum: easier to minimise
- Arbitrarily complex problem, vast amount of techniques and literature
  - e. g. ATLAS+CMS Higgs couplings combination: $\approx$ 4250 parameters
  - e. g. CMS tracker-alignment problem: $\mathcal{O}(10^6)$ parameters

Maximum-Likelihood Method

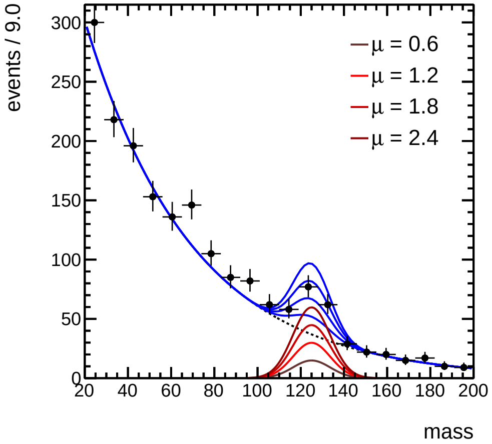
Karlsruhe Institute of Technology

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in analyses at the LHC: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$

$$n_{\text{pred},i}(\mu) = \underbrace{b \cdot e^{-\alpha x_i}}_{\text{background}} + \underbrace{\mu \cdot s_{\text{SM}} \cdot e^{-(m-x_i)^2}}_{\text{signal}}$$

$$\mathcal{L}(n_{\text{obs}};\mu) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} e^{-n_{\text{pred},i}}$$

$\rightarrow$ $\boxed{\hat{\mu}\text{: maximises } \mathcal{L}(n_{\text{obs}};\mu)}$



Maximum-Likelihood Method

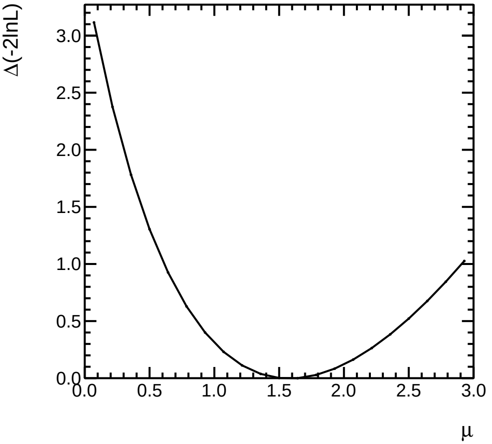
- Example: which *signal-strength modifier* μ describes best the data?
  - Common in analyses at the LHC:
    $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$

$$n_{\text{pred},i}(\mu) = \underbrace{b \cdot e^{-\alpha x_i}}_{\text{background}} + \underbrace{\mu \cdot s_{\text{SM}} \cdot e^{-(m-x_i)^2}}_{\text{signal}}$$

$$\mathcal{L}(n_{\text{obs}};\mu) = \prod_i \frac{(n_{\text{pred},i})^{n_{\text{obs},i}}}{n_{\text{obs},i}!} e^{-n_{\text{pred},i}}$$

$\rightarrow$ $\boxed{\hat{\mu}\text{: minimises } -2\ln \mathcal{L}(n_{\text{obs}};\mu)}$

Note: we are usually not interested in the value of $\mathcal{L}$, only in the difference (Δ) w.r.t. the minimum

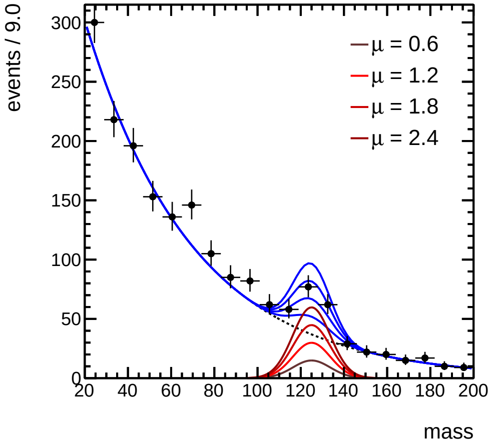


Maximum-Likelihood Method

Karlsruhe Institute of Technology

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in analyses at the LHC: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$
- Important case: **Gaussian distributed measurements**
  - e. g. approximation of Poisson for large number of events (in practice > 10)
- NLL becomes

$$-2 \ln \mathcal{L} = \sum_i \frac{(n_{\text{obs},i} - n_{\text{pred},i})^2}{n_{\text{pred},i}} + \text{const}$$

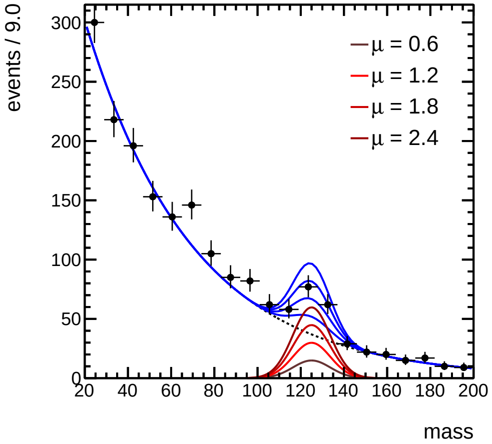


Maximum-Likelihood Method

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in analyses at the LHC: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$
- Important case: **Gaussian distributed measurements**
  - e. g. approximation of Poisson for large number of events (in practice > 10)
- NLL becomes

$$-2 \ln \mathcal{L} = \sum_i \frac{(n_{\text{obs},i} - n_{\text{pred},i})^2}{n_{\text{pred},i}} + \text{const}$$

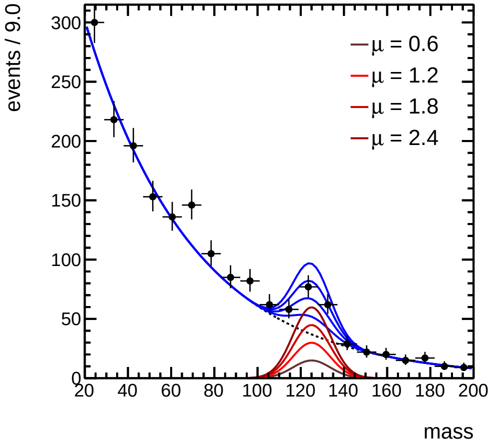
$$\boxed{\chi^2 \equiv \sum_i \frac{(n_{\text{obs},i} - n_{\text{pred},i})^2}{n_{\text{pred},i}}}$$



Maximum-Likelihood Method

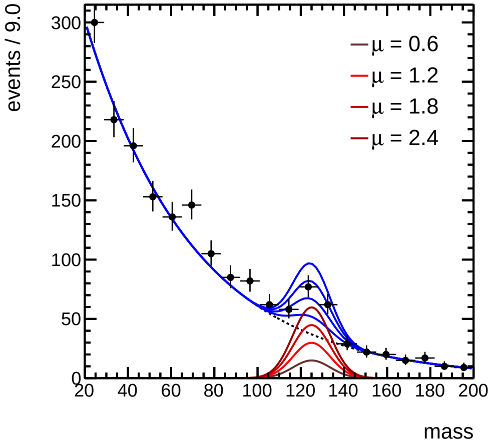
- Example: which *signal-strength modifier* μ describes best the data?
  - Common in analyses at the LHC: $\mu = (\sigma \cdot \mathcal{B})/(\sigma_{\text{SM}} \cdot \mathcal{B}_{\text{SM}})$

ML method obtains parameter values most compatible with the data *for a given model* — **a ML fit does not find the correct model**!



Maximum-Likelihood Method

- Example: which *signal-strength modifier* μ describes best the data?
  - Common in analyses at the LHC:
    $\mu = (\sigma \cdot \mathcal{B}) / (\sigma_{\mathrm{SM}} \cdot \mathcal{B}_{\mathrm{SM}})$



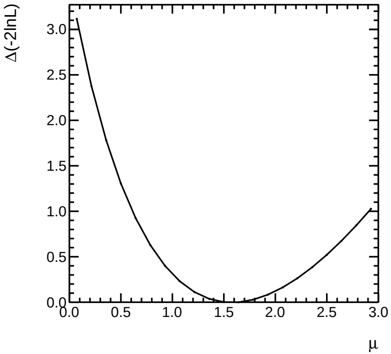
ML method obtains parameter values most compatible with the data *for a given model* — **a ML fit does not find the correct model**!

ML estimator is function of the observed data

Maximum-Likelihood Method

- Uncertainty on $\hat{\mu}$ from scan of NLL = $-2\ln\mathcal{L}$ around minimum
- **Uncertainty due to fluctuations of data**: “statistical uncertainty”
- For Gaussian pdfs: parabola, standard deviation σ follows from

$$\Delta(\mathrm{NLL}) = \mathrm{NLL}(\hat{\mu} \pm r \cdot \sigma) - \mathrm{NLL}(\hat{\mu}) = r$$

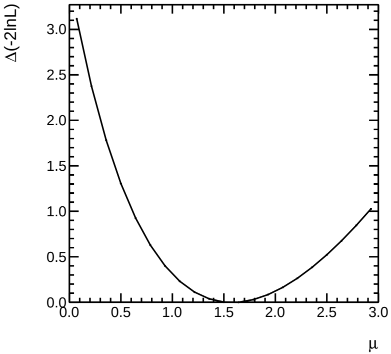


Maximum-Likelihood Method

Karlsruhe Institute of Technology

- Uncertainty on $\hat{\mu}$ from scan of NLL = $-2\ln\mathcal{L}$ around minimum
- **Uncertainty due to fluctuations of data**: “statistical uncertainty”
- For Gaussian pdfs: parabola, standard deviation σ follows from

$$\Delta(\mathrm{NLL}) = \mathrm{NLL}(\hat{\mu} \pm r \cdot \sigma) - \mathrm{NLL}(\hat{\mu}) = r$$



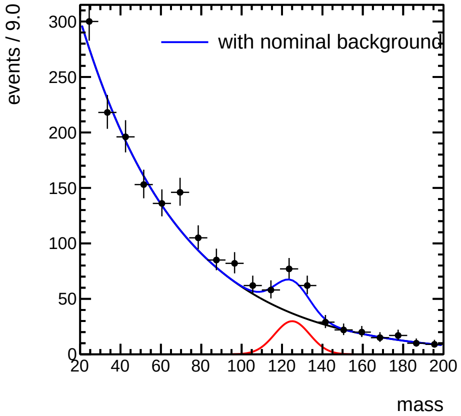
- Other cases can often be approximated by Gaussian case (and true for $n \to \infty$)
- For strongly asymmetric NLL
  - Often asymmetric intervals quoted
  - Better: variable transformation such that NLL symmetric

Uncertainty vs. Error

- **Uncertainty** reflects degree of precision with which one can deduce a parameter value from the data if one does everything correctly
  - Statistical uncertainties stem from the inherent stochastic nature of the data and finite number of observed events (can not be eliminated, only minimised)
  - Systematic uncertainties stem e. g. from the limited knowledge of the precision of the detector or approximations in theory calculations
  - Uncertainties can (in principle) be quantified
- In contrast, **errors are mistakes**
  - e. g. a lose cable or a wrong method
  - Errors can (and should) be eliminated
- In context of statistical data analysis, we mean uncertainties
  - NB: “Error bars” denote uncertainties!

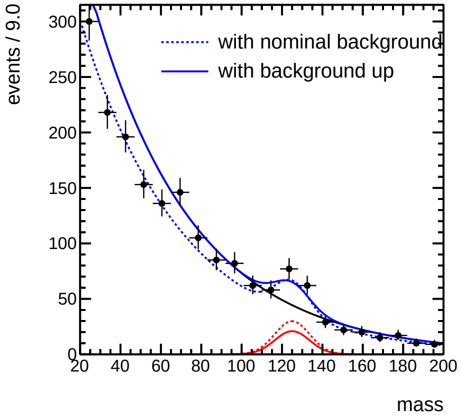
Incorporation of Systematic Uncertainties

- Often, background (and signal) models subject to systematic uncertainties
  - Some (theory or experimental) parameters not exactly known, e. g. cross section or trigger efficiency
- For example, background normalisation *b* not precisely known but with some uncertainty



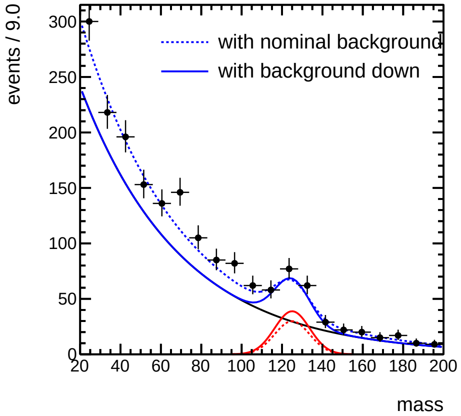
Incorporation of Systematic Uncertainties

- Often, background (and signal) models subject to systematic uncertainties
  - Some (theory or experimental) parameters not exactly known, e. g. cross section or trigger efficiency
- For example, background normalisation *b* not precisely known but with some uncertainty $\rightarrow$ affects determination of signal
  - More background events $\rightarrow$ less signal events
  - Generally leads to **larger uncertainty on $\hat{\mu}$**



Incorporation of Systematic Uncertainties

- Often, background (and signal) models subject to systematic uncertainties
  - Some (theory or experimental) parameters not exactly known, e. g. cross section or trigger efficiency
- For example, background normalisation b not precisely known but with some uncertainty $\rightarrow$ affects determination of signal
  - Less background events $\rightarrow$ more signal events
  - Generally leads to **larger uncertainty on $\hat{\mu}$**

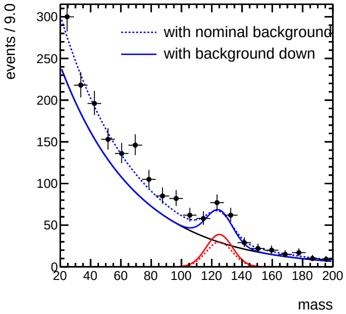


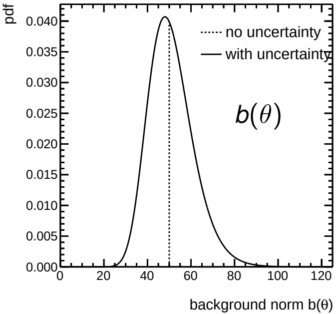
Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\boxed{\mathcal{L}(\mathbf{n}_{\text{obs}};\mu,\theta)=\prod_i \mathcal{P}(n_{\text{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)}$$

- Background normalisation becomes function of θ: $b \to b(\theta)$



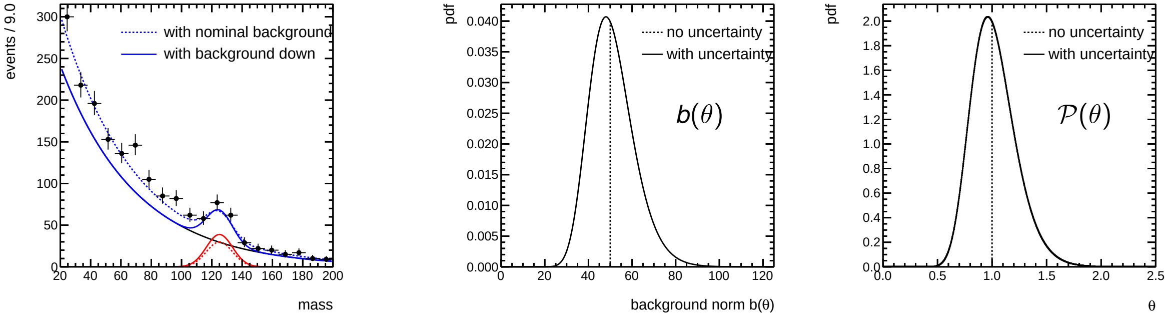


Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\boxed{\mathcal{L}(\mathbf{n}_{\text{obs}};\mu,\theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)}$$

- Background normalisation becomes function of θ: $b \to b(\theta)$
  - θ: assumed true value, parameter of the fit
  - $\tilde{\theta}$: best knowledge, e. g. estimate from independent measurement, distributed as $\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$

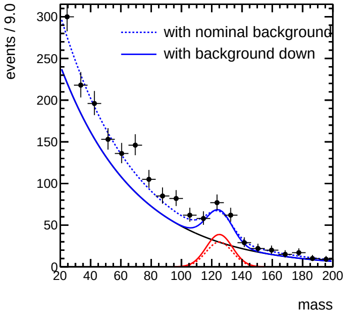


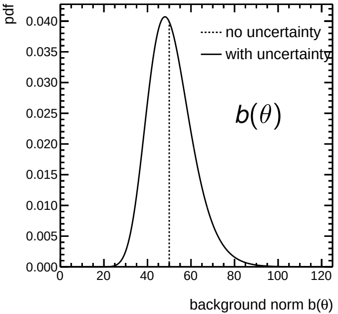
Incorporation of Systematic Uncertainties

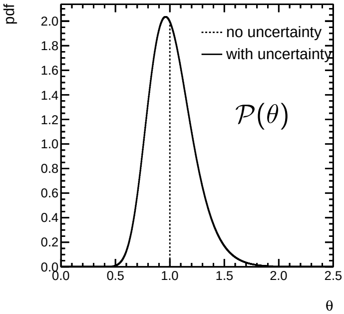
- Incorporated into likelihood via **nuisance parameters** θ

$$\boxed{\mathcal{L}(\mathbf{n}_{\text{obs}}; \mu, \theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i}; \mu, \theta) \cdot \mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)}$$

- Background normalisation becomes function of θ: $b \to b(\theta)$
  - ML fit can adjust θ to a value different than $\tilde{\theta}$ to achieve better description of data but at the cost of reducing value of $\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)$





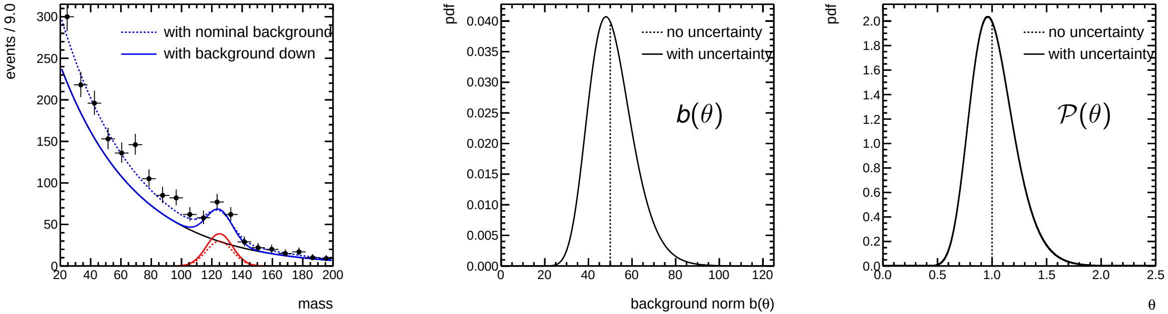


Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\boxed{\mathcal{L}(\mathbf{n}_{\text{obs}};\mu,\theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)}$$

  - μ: *parameter of interest* (POI)
  - θ: *nuisance parameter* (NP)



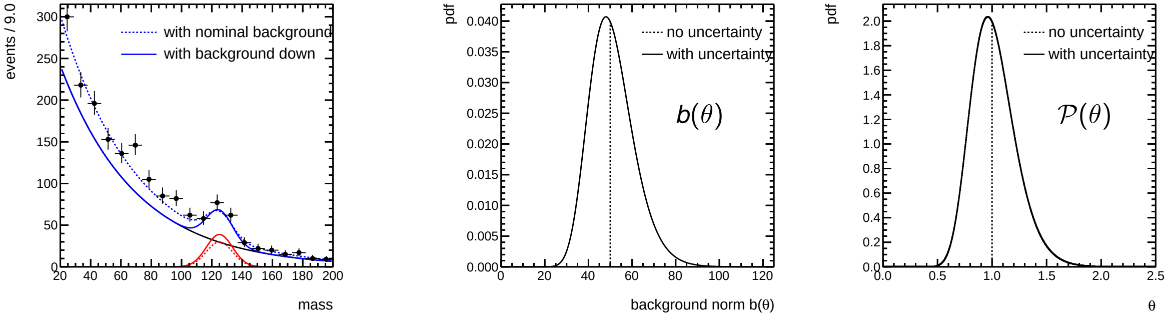
Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\boxed{\mathcal{L}(\mathbf{n}_{\text{obs}};\mu,\theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)}$$

  - μ: *parameter of interest* (POI)
  - θ: *nuisance parameter* (NP)

  generally, (anti-)correlated with μ: increased uncertainty on $\hat{\mu}$ $\rightarrow$ smaller sensitivity



Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\boxed{\mathcal{L}(\mathbf{n}_{\mathrm{obs}};\mu,\theta) = \prod_i \mathcal{P}(n_{\mathrm{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)}$$

  - μ: *parameter of interest* (POI)
  - θ: *nuisance parameter* (NP)
- $\mathcal{L}$ function of μ and θ, but only interested in μ
  $\rightarrow$ can rewrite θ as function of μ

Profile likelihood $\mathcal{L}_{\mathrm{p}}(\mu) \equiv \mathcal{L}(\mu,\hat{\theta}(\mu))$

$\hat{\theta}(\mu)$ maximises $\mathcal{L}$ for given μ (“profiled values” of θ)

Incorporation of Systematic Uncertainties

- Incorporated into likelihood via **nuisance parameters** θ

$$\boxed{\mathcal{L}(\mathbf{n}_{\text{obs}};\mu,\theta) = \prod_i \mathcal{P}(n_{\text{obs}\,i};\mu,\theta)\cdot\mathcal{P}_{\tilde{\theta}}(\tilde{\theta}|\theta)}$$

  - μ: *parameter of interest* (POI)
  - θ: *nuisance parameter* (NP)
- $\mathcal{L}$ function of μ and θ, but only interested in μ
  $\rightarrow$ can rewrite θ as function of μ

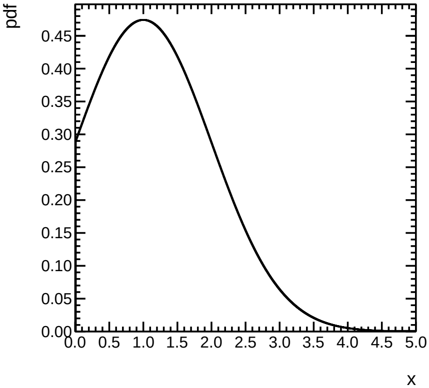
Profile likelihood $\mathcal{L}_{\text{p}}(\mu) \equiv \mathcal{L}(\mu,\hat{\theta}(\mu))$

$\hat{\theta}(\mu)$ maximises $\mathcal{L}$ for given μ (“profiled values” of θ)

  - $\mathcal{L}_{\text{p}}$ in practice computationally advantageous
    - e. g. need only to consider μ instead of full parameter space (μ,θ) when computing uncertainties

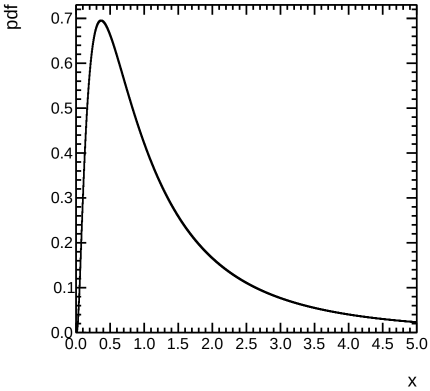
Typical Choices for $\mathcal{P}_\theta(\theta|\tilde{\theta})$

(truncated) Gaussian



$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

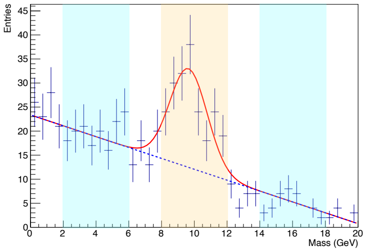
Log-normal



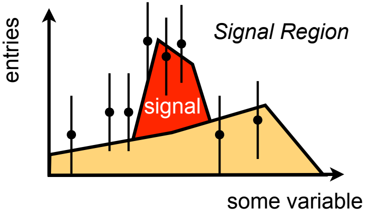
$$\frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

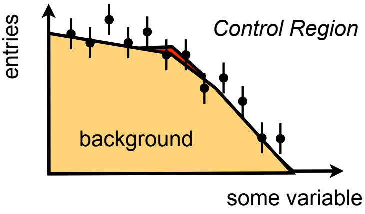
Sidebands and Control Regions

sidebands



signal/control region





- Signal-depleted sideband/control region
  - Determine rate of background process
  - Determine shape of background processes
  - Constrain uncertainties on background prediction
- Can be incorporated into likelihood fit via **nuisance parameters**: **simultaneous determination of POI and background parameters**

Confidence Intervals

- Experiment to determine value of some parameter x measures x_{obs}
  - e. g. from maximum likelihood estimate, tag-and-probe measurement
- Want to quote "uncertainty": interval that reflects statistical precision of measurement
  - Can be estimate of standard deviation from likelihood fit
  - More generally, in particular if non-Gaussian $\mathcal{P}$: **confidence interval**

Confidence Intervals

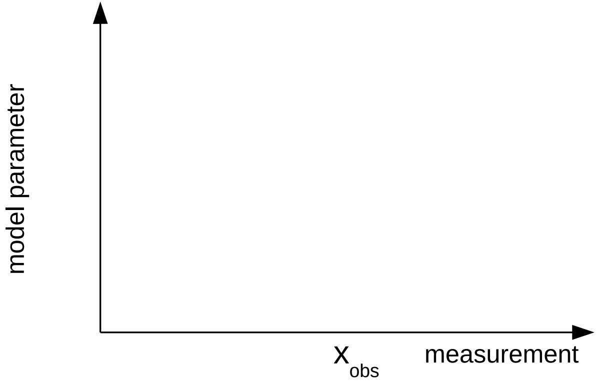
- Experiment to determine value of some parameter x measures x_{obs}
  - e. g. from maximum likelihood estimate, tag-and-probe measurement
- Want to quote “uncertainty”: interval that reflects statistical precision of measurement
  - Can be estimate of standard deviation from likelihood fit
  - More generally, in particular if non-Gaussian $\mathcal{P}$: **confidence interval**
- **Confidence interval** covers on average the true value with a given probability
  - e. g. *90 % confidence level*: if experiment repeated many times, the interval covers the true value in 90 % of the cases

Confidence Intervals

- Experiment to determine value of some parameter x measures x_{obs}
  - e. g. from maximum likelihood estimate, tag-and-probe measurement
- Want to quote “uncertainty”: interval that reflects statistical precision of measurement
  - Can be estimate of standard deviation from likelihood fit
  - More generally, in particular if non-Gaussian $\mathcal{P}$: **confidence interval**
- **Confidence interval** covers on average the true value with a given probability
  - e. g. *90 % confidence level*: if experiment repeated many times, the interval covers the true value in 90 % of the cases
  - Careful: this is **not** the probability of the true value to lie within the interval
    $\rightarrow$ there is only one true value either within or not!

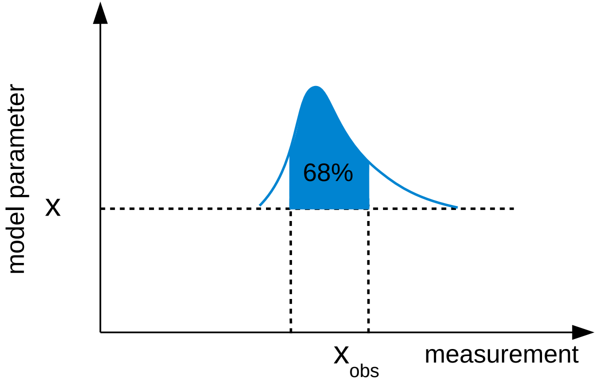
Confidence Intervals

Neyman construction



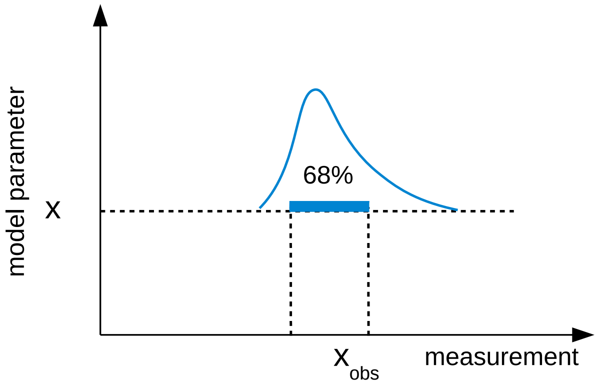
Confidence Intervals

Neyman construction



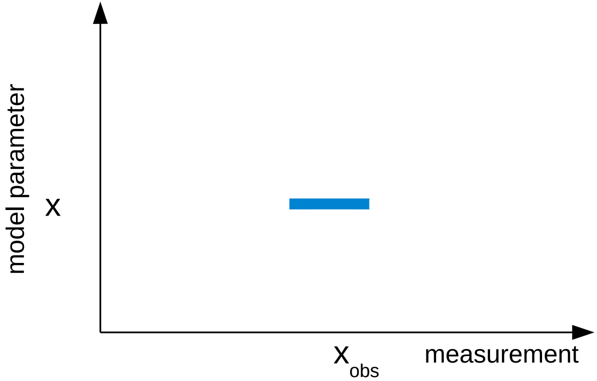
Confidence Intervals

Neyman construction



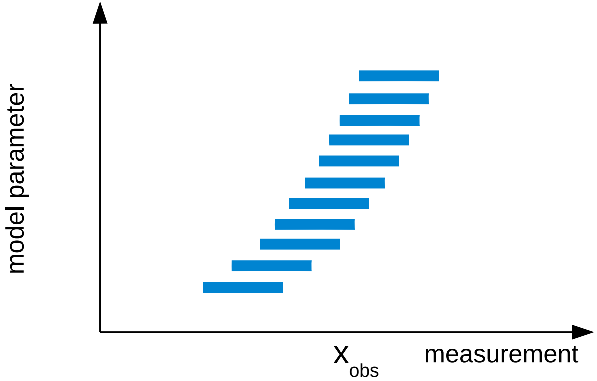
Confidence Intervals

Neyman construction



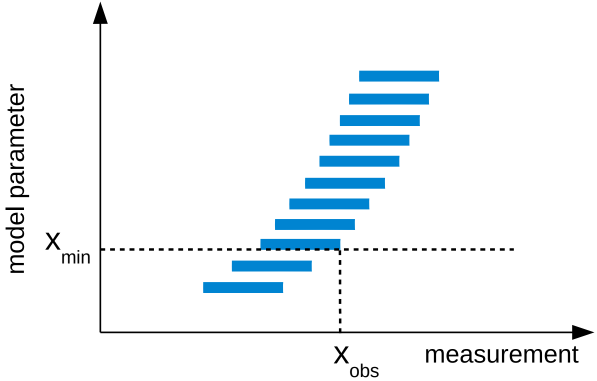
Confidence Intervals

Neyman construction



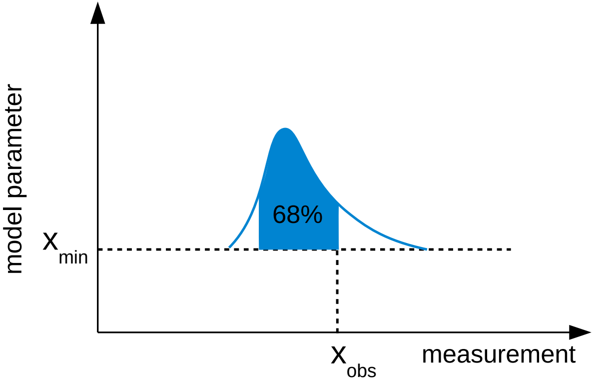
Confidence Intervals

Neyman construction



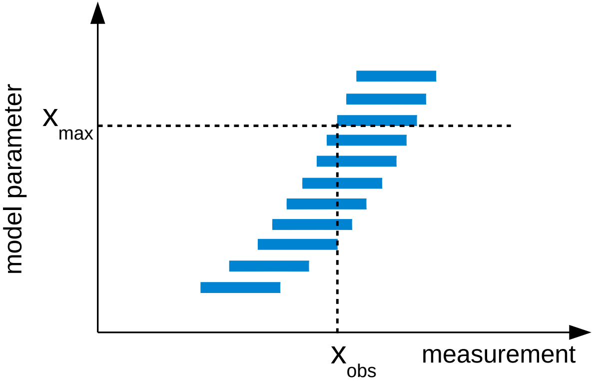
Confidence Intervals

Neyman construction



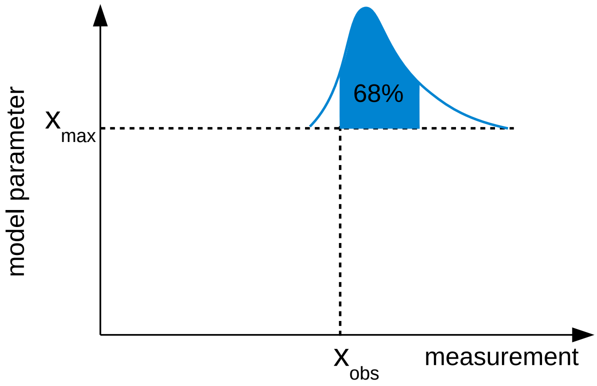
Confidence Intervals

Neyman construction



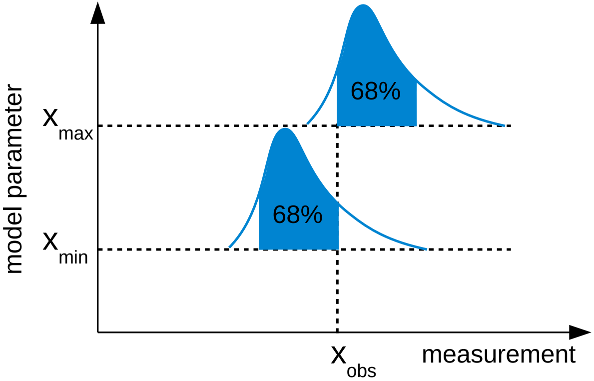
Confidence Intervals

Neyman construction



Confidence Intervals

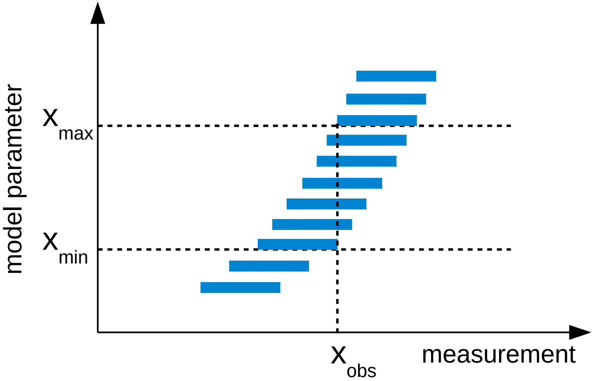
Neyman construction



Confidence interval is $[x_{\mathrm{min}}, x_{\mathrm{max}}]$

Confidence Intervals

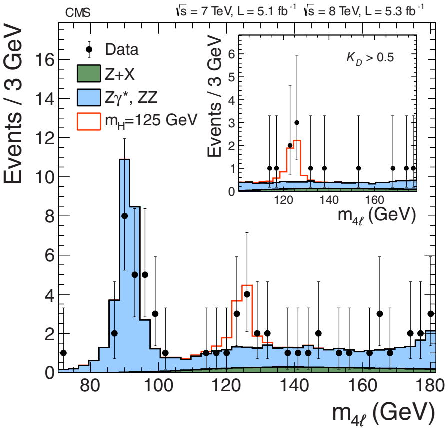
Neyman construction



Confidence interval is $[x_{\text{min}}, x_{\text{max}}]$

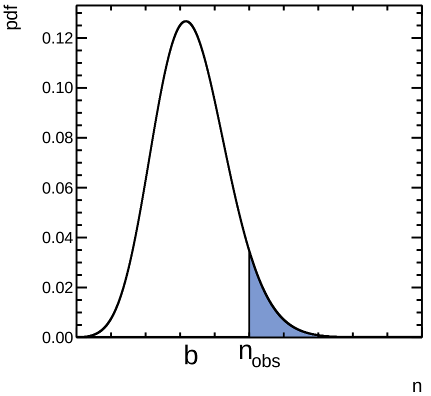
p Value

- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?



p Value

- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?
- Simple example: assume counting experiment
  - b: expected number of background events (=model)
  - n_{obs}: number of observed events

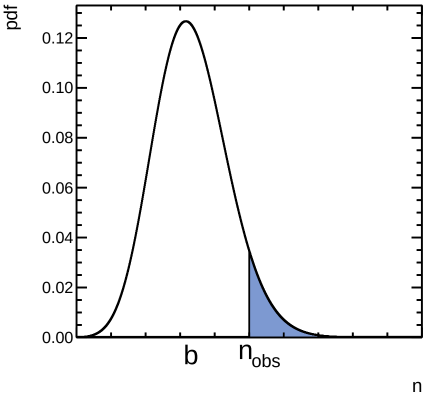


p Value

- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?
- Simple example: assume counting experiment
  - b: expected number of background events (=model)
  - n_{obs}: number of observed events
- "p value": **probability of upward fluctuation as large as or larger than observed in data**

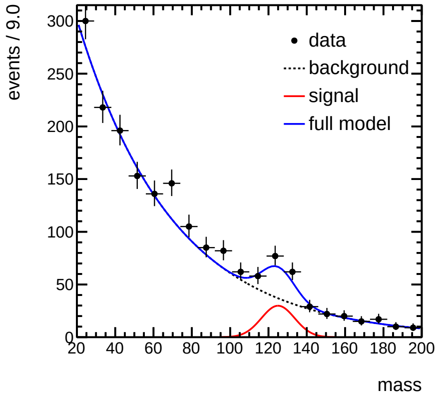
$$\boxed{p \equiv \mathrm{P}(n \geq n_{\text{obs}} | b) = \int_{n_{\text{obs}}}^{\infty} \mathrm{d}n \, \mathcal{P}(n|b)}$$

  - p value is **not the probability of a hypothesis**
    p quantifies level of (dis-)agreement between model and data:
    $\rightarrow$ judgement call whether to keep model or reject it



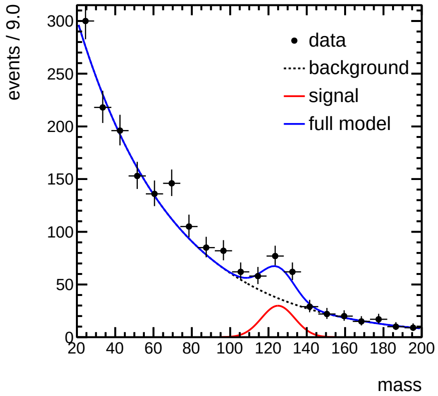
Test Statistic

- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance parameters



Test Statistic

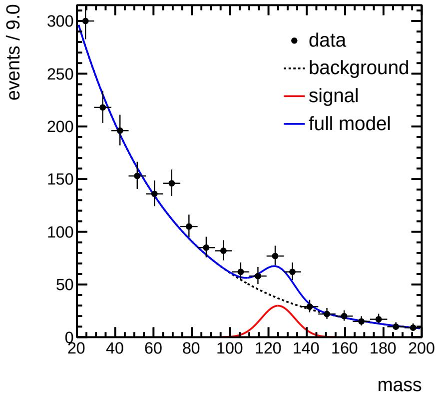
- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance parameters
- Formally: use appropriate **test statistic** $t : \mathbb{R}^D \rightarrow \mathbb{R}$
  - In general any function that **combines relevant information** from an experiment, e. g. number of observed events per bin, **into one single number** reflecting the agreement between data and hypothesis $\rightarrow$ Likelihood is example for test statistic



Test Statistic

- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance parameters
- Formally: use appropriate **test statistic** $t : \mathbb{R}^D \to \mathbb{R}$
  - In general any function that **combines relevant information** from an experiment, e. g. number of observed events per bin, **into one single number** reflecting the agreement between data and hypothesis $\to$ Likelihood is example for test statistic
- t can be used to compute p value for complex models

$$\boxed{p = \int_{t_{\text{obs}}}^{\infty} \mathcal{P}(t)\,\mathrm{d}t}$$

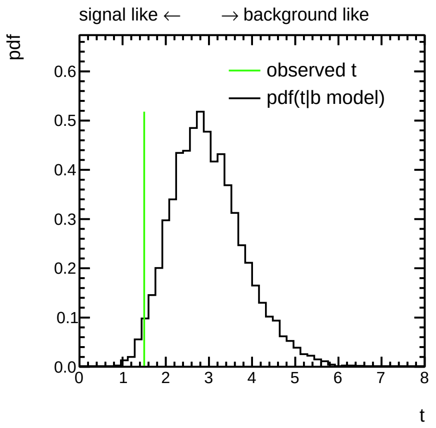


Test Statistic

- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance parameters
- Formally: use appropriate **test statistic** $t : \mathbb{R}^D \to \mathbb{R}$
  - In general any function that **combines relevant information** from an experiment, e. g. number of observed events per bin, **into one single number** reflecting the agreement between data and hypothesis $\to$ Likelihood is example for test statistic
- t can be used to compute p value for complex models

$$\boxed{p = \int_{t_{\mathrm{obs}}}^{\infty} \mathcal{P}(t)\,\mathrm{d}t}$$

  - Requires pdf $\mathcal{P}(t)$ of the test statistic t: typically from simulation (*toy data*)

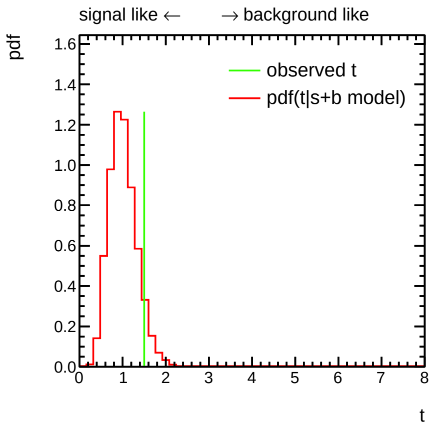


Test Statistic

- Typically problems not as simple as a single measured quantity
  - Multiple channels/measurements, e. g. binned distribution
  - Prediction depends on several parameters, e. g. POI, nuisance parameters
- Formally: use appropriate **test statistic** $t : \mathbb{R}^D \to \mathbb{R}$
  - In general any function that **combines relevant information** from an experiment, e. g. number of observed events per bin, **into one single number** reflecting the agreement between data and hypothesis $\rightarrow$ Likelihood is example for test statistic
- t can be used to compute p value for complex models

$$\boxed{p = \int_{t_{\text{obs}}}^{\infty} \mathcal{P}(t)\,\mathrm{d}t}$$

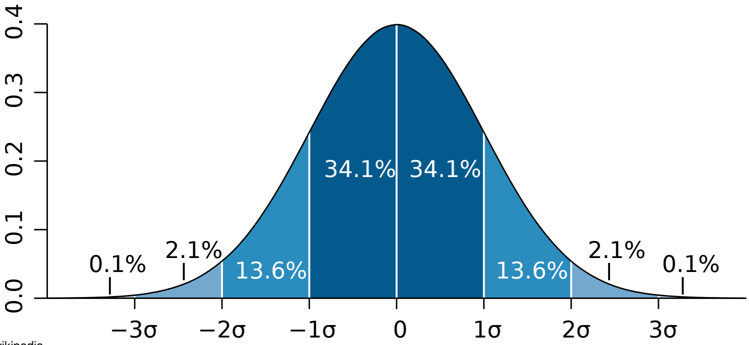
  - Requires pdf $\mathcal{P}(t)$ of the test statistic t: typically from simulation (*toy data*)



Significance

- Often, p value converted into equivalent **significance *Z***:
  upward fluctuation from 0 by Z of normal-distributed variable corresponding to same p value
  - Corresponds to Z standard deviations σ of the Gaussian distribution

$$\boxed{Z = \Phi^{-1}(1-p)}$$ Φ: cumulative (=quantile) function of normal distribution



wikipedia

- **Convention** to classify effects by significance:
  - 3σ: *evidence* for a signal (0.3% chance of background fluctuation)
  - 5σ: *observation* of a signal (0.00006% chance of background fluctuation)

Distinguishing Hypotheses

- **Cannot determine whether a hypothesis is true** (frequentist)
- But can define rules how to reject hypothesis in favour of an alternative hypothesis
- Can determine probability of wrong choice: how often wrong choice is made would the experiment be repeated very often

Distinguishing Hypotheses

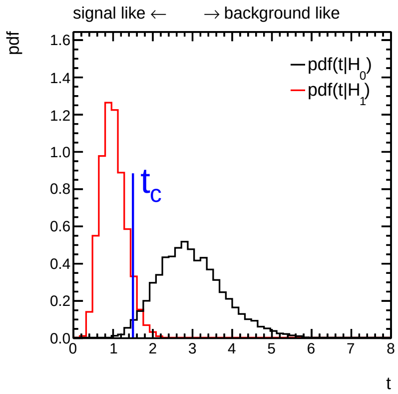
Karlsruhe Institute of Technology

- Classification problem: how to interpret outcome of an experiment?
- Want to distinguish between two alternative hypotheses, e. g.
  - H_0: there are only background events, e. g. no Higgs boson
  - H_1: there is a signal, e. g. a Higgs boson, contribution to the data

Distinguishing Hypotheses

- Classification problem: how to interpret outcome of an experiment?
- Want to distinguish between two alternative hypotheses, e. g.
  - H_0: there are only background events, e. g. no Higgs boson
  - H_1: there is a signal, e. g. a Higgs boson, contribution to the data
- Define **test statistic *t***, e. g. $\mathcal{L}$
- Construct pdf $\mathcal{P}(t|H_i)$ of t under H_0 and H_1
  - In reality often from simulation
- How compatible is t_{obs} with H_i?

$\rightarrow$ Set a critical value t_c and reject H_0 in favour of H_1 if $t_{obs} < t_c$



Distinguishing Hypotheses

- Classification problem: how to interpret outcome of an experiment?
- Want to distinguish between two alternative hypotheses, e. g.
  - H_0: there are only background events, e. g. no Higgs boson
  - H_1: there is a signal, e. g. a Higgs boson, contribution to the data
- Types of errors

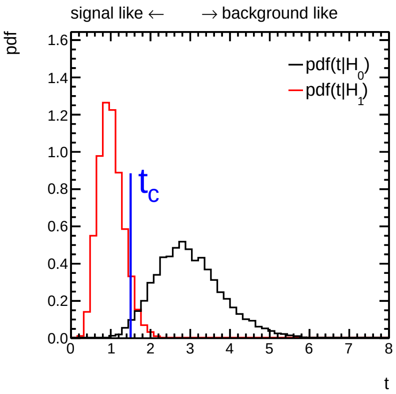
  I: reject H_0 although it is true

  II: accept H_0 although H_1 is true

$$\boxed{\begin{aligned} \text{I: } \mathrm{P}(t < t_c|H_0) &= \int_{-\infty}^{t_c} \mathrm{d}t\, \mathcal{P}(t|H_0) \equiv \alpha \\ \text{II: } \mathrm{P}(t > t_c|H_1) &= \int_{t_c}^{\infty} \mathrm{d}t\, \mathcal{P}(t|H_1) \equiv \beta \end{aligned}}$$

α : significance of test

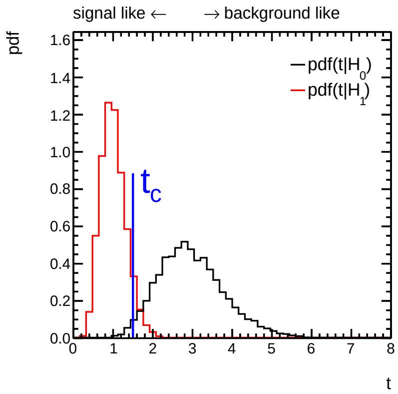
$1-\beta$: power of test



Distinguishing Hypotheses

- Classification problem: how to interpret outcome of an experiment?
- Want to distinguish between two alternative hypotheses, e. g.
  - H_0: there are only background events, e. g. no Higgs boson
  - H_1: there is a signal, e. g. a Higgs boson, contribution to the data
- Types of errors
  - I: reject H_0 although it is true
  - II: accept H_0 although H_1 is true

$$\boxed{\begin{aligned} \text{I: } \mathrm{P}(t < t_c|H_0) &= \int_{-\infty}^{t_c} \mathrm{d}t\, \mathcal{P}(t|H_0) \equiv \alpha \\ \text{II: } \mathrm{P}(t > t_c|H_1) &= \int_{t_c}^{\infty} \mathrm{d}t\, \mathcal{P}(t|H_1) \equiv \beta \end{aligned}}$$



By choice of t_c (choice of α): association to one or the other hypothesis performed "**at the confidence level α**"

Difference Between α and *p* Value?

- In one sense, no difference
  - Both are $\int_x^\infty \mathrm{d}x' \, \mathcal{P}(x')$
- But **conceptually very different**
  - α is computed before one sees the data: predefined property of the test
  - p depends on the data (property of the data) and is a random variable

Neyman-Pearson Lemma

Karlsruhe Institute of Technology

- When performing a test between two hypotheses (models) H_0 and H_1, the **likelihood ratio test**, which rejects H_0 in favour of H_1 if

$$\boxed{Q = \frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} > Q_c} \quad \text{with } \mathrm{P}(Q > Q_c | H_0) = \alpha$$

  is the **most powerful test** at a significance level α (Neyman-Pearson Lemma)
- The test statistic Q is called **likelihood ratio**

Neyman-Pearson Lemma

- When performing a test between two hypotheses (models) H_0 and H_1, the **likelihood ratio test**, which rejects H_0 in favour of H_1 if

$$\boxed{Q = \frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} > Q_c} \quad \text{with } \mathrm{P}(Q > Q_c | H_0) = \alpha$$

  is the **most powerful test** at a significance level α (Neyman-Pearson Lemma)
- The test statistic Q is called **likelihood ratio**
- For a number of reasons, usually

$$\boxed{q = -2 \ln Q}$$

Test Statistic at the LHC

- Used test statistic in particle-physics experiments evolved with time (LEP $\rightarrow$ Tevatron $\rightarrow$ LHC)
  - Different conditions: importance of systematic uncertainties, computational power
- Hypothesis test example
  - H_0: background-only hypothesis, e. g. SM without Higgs boson
  - H_1: Higgs boson with fixed mass and signal strength μ (= signal w.r.t. prediction) present in data

[1] G. Cowan "Asymptotic formulae for likelihood-based tests of new physics", Eur. Phys. J. C71:1554 (2011)

Test Statistic at the LHC

Karlsruhe Institute of Technology

- Used test statistic in particle-physics experiments evolved with time (LEP $\rightarrow$ Tevatron $\rightarrow$ LHC)
  - Different conditions: importance of systematic uncertainties, computational power
- Hypothesis test example
  - H_0: background-only hypothesis, e. g. SM without Higgs boson
  - H_1: Higgs boson with fixed mass and signal strength μ (= signal w.r.t. prediction) present in data
- At the LHC commonly: **profile likelihood ratio test statistic**

$$q_\mu = -2 \ln \frac{\mathcal{L}(n_{\text{obs}}; \mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(n_{\text{obs}}; \hat{\mu}, \hat{\theta})}$$

  - Nuisance parameters are profiled in nominator
  - Global maximum of $\mathcal{L}$ under (μ, θ) as denominator with $0 \leq \hat{\mu} \leq \mu$
  - Allows usage of certain approximations when computing $\mathcal{P}(q_\mu | H_i)$ ("asymptotic formulae" based on Wilks and Wald theorem[1])

[1] G. Cowan "Asymptotic formulae for likelihood-based tests of new physics", Eur. Phys. J. C71:1554 (2011)

Excluding Parameters

- Assume measurement with a given sensitivity: no signal observed
- How much signal can “hide” in the background fluctuations (+uncertainty)?
- **How large could a signal be at most?**

Excluding Parameters

- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2 \ln \frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

Excluding Parameters

- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2 \ln \frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q \, \mathcal{P}(q(\mu_{1-\alpha}) | H_1) \equiv \mathrm{CL}_{\mathrm{s+b}}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$

Excluding Parameters

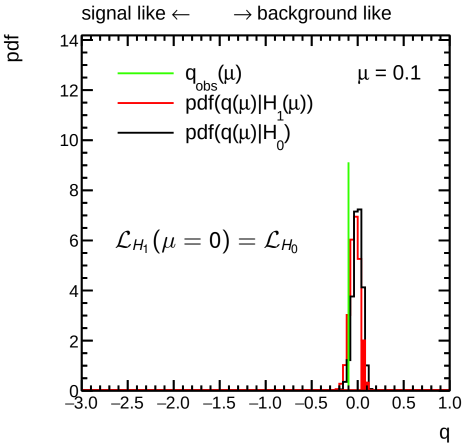
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q\,\mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$



Excluding Parameters

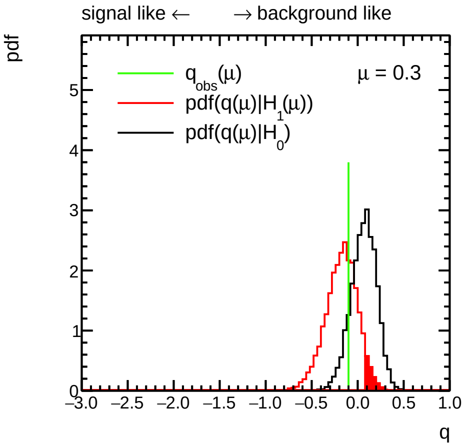
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_\mathrm{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{\mathrm{s+b}}}$$

i. e. for $\mu_{1-\alpha}$: $q_\mathrm{obs} = q_c$



Excluding Parameters

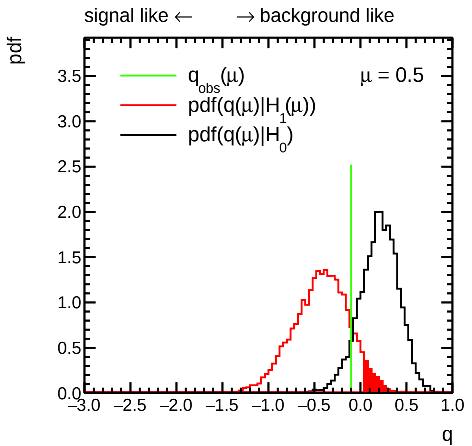
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_{\mathrm{obs}}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{\mathrm{s+b}}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\mathrm{obs}} = q_c$



Excluding Parameters

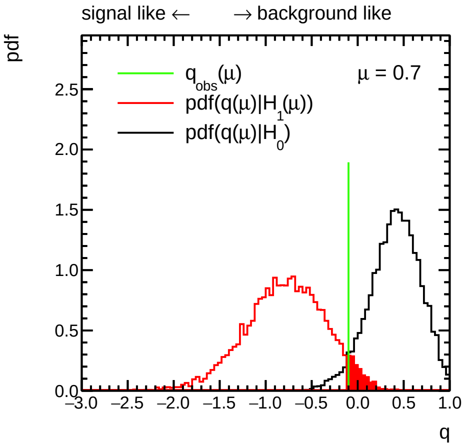
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_\text{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{s+b}}$$

i. e. for $\mu_{1-\alpha}$: $q_\text{obs} = q_c$



Excluding Parameters

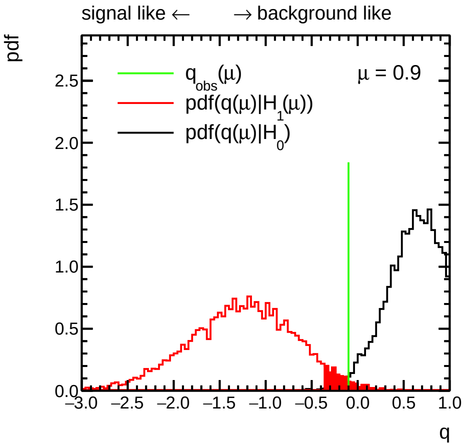
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q\,\mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{\mathrm{s+b}}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$



Excluding Parameters

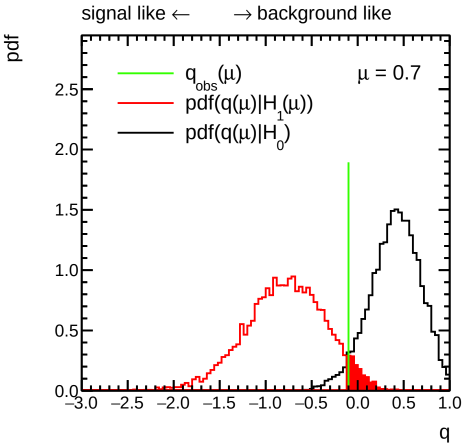
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2 \ln \frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q \, \mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{\text{s+b}}}$$

i. e. for $\mu_{1-\alpha}$: $q_{\text{obs}} = q_c$



Excluding Parameters

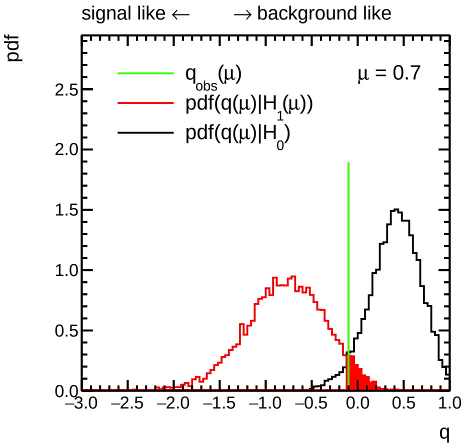
- No signal observed: **how large could a signal be at most?**
- Formally: test statistic q, depending on signal-strength modifier μ

$$q(\mu) = -2\ln\frac{\mathcal{L}_{H_1}(\mu)}{\mathcal{L}_{H_0}}$$

What is largest signal $\mu \equiv \mu_{1-\alpha}$
for which H_1 would be rejected at significance level α?

$$\boxed{\alpha = \int_{q_\text{obs}}^{\infty} \mathrm{d}q\,\mathcal{P}(q(\mu_{1-\alpha})|H_1) \equiv \mathrm{CL}_{s+b}}$$

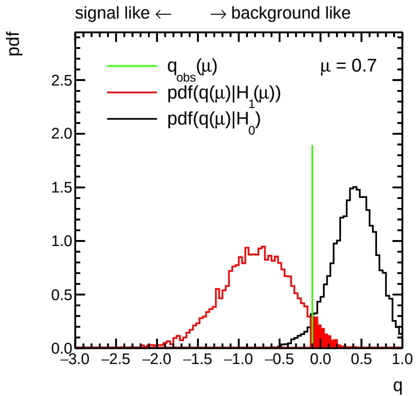
i. e. for $\mu_{1-\alpha}$: $q_\text{obs} = q_c$



$\mu_{1-\alpha}$ is called **upper limit** on μ at confidence level
C.L. $= 1 - \alpha$
In particle physics usually 95 % C.L. limit

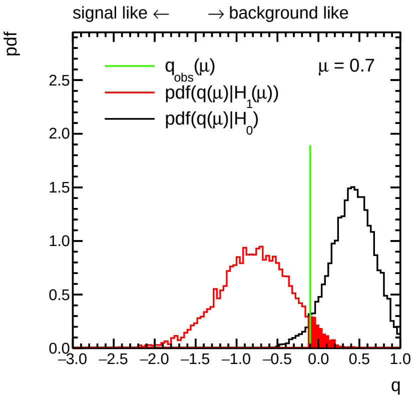
(Observed) Upper Limit: Interpretation

- "Maximal signal that we would still reject"



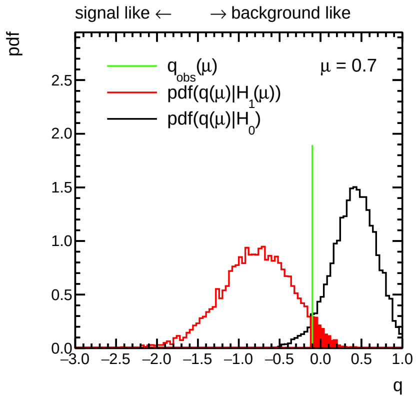
(Observed) Upper Limit: Interpretation

- "Maximal signal that we would still reject"
- 95 % C.L. upper limit on μ: largest value of μ that would still be rejected in a test with significance 5% given the data
  - NB: limit is a **function of the data** (depends on q_{obs})!
- μ for which $\text{CL}_{s+b} = 0.05$: $\boxed{0.05 = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{95})|H_1)}$



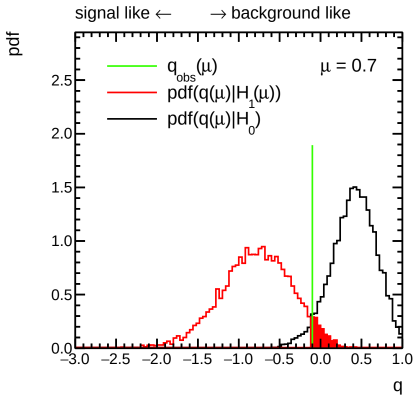
(Observed) Upper Limit: Interpretation

- “Maximal signal that we would still reject”
- 95 % C.L. upper limit on μ: largest value of μ that would still be rejected in a test with significance 5% given the data
  - NB: limit is a **function of the data** (depends on q_{obs})!
- μ for which $\text{CL}_{s+b} = 0.05$: $\boxed{0.05 = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{95})|H_1)}$
- Upper **limit covers true value ($\mu_{\text{true}} < \mu_{95}$) with probability C.L. $=$ 95 %**
  - If the experiment is repeated many times, μ_{95} would be larger than μ_{true} in 95 % of the cases



(Observed) Upper Limit: Interpretation

- "Maximal signal that we would still reject"
- 95 % C.L. upper limit on μ: largest value of μ that would still be rejected in a test with significance 5% given the data
  - NB: limit is a **function of the data** (depends on q_{obs})!
- μ for which $\text{CL}_{s+b} = 0.05$: $\boxed{0.05 = \int_{q_{\text{obs}}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu_{95})|H_1)}$
- Upper **limit covers true value ($\mu_{\text{true}} < \mu_{95}$) with probability C.L. $=$ 95 %**
  - If the experiment is repeated many times, μ_{95} would be larger than μ_{true} in 95 % of the cases
- Still **5 % chance of wrong exclusion**, i. e. that $\mu_{\text{true}} > \mu_{95}$

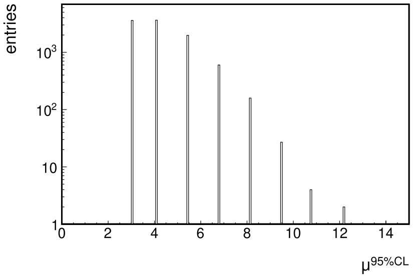


Expected Limit

- **Estimate what observed limit would look like in case of no signal**

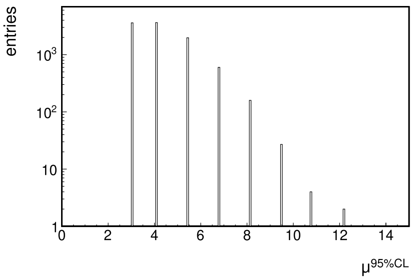
Expected Limit

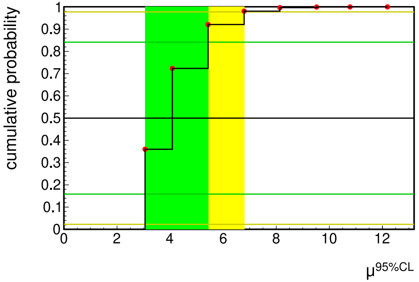
- **Estimate what observed limit would look like in case of no signal**
- Obtained e. g. from *toy dataset*
  - Sample toy data for q under background-only hypothesis from $\mathcal{P}(q|H_0)$
  - Treat each as observation and compute μ_{95} limit
  - Obtain quantiles from distribution of all μ_{95}



Expected Limit

- **Estimate what observed limit would look like in case of no signal**
- Obtained e. g. from *toy dataset*
  - Sample toy data for q under background-only hypothesis from $\mathcal{P}(q|H_0)$
  - Treat each as observation and compute μ_{95} limit
  - Obtain quantiles from distribution of all μ_{95}
- **Expected limit = median of μ_{95} distribution**

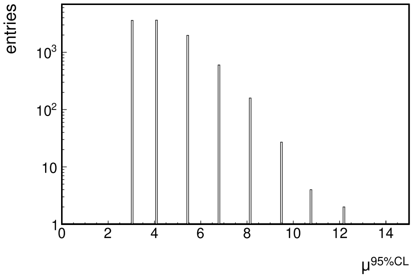


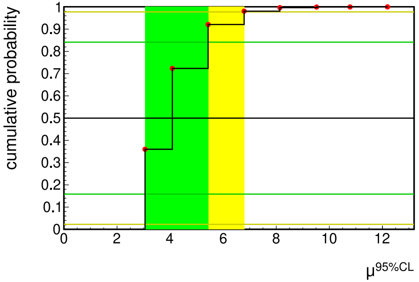


Expected Limit

- **Estimate what observed limit would look like in case of no signal**
- Obtained e. g. from *toy dataset*
  - Sample toy data for q under background-only hypothesis from $\mathcal{P}(q|H_0)$
  - Treat each as observation and compute μ_{95} limit
  - Obtain quantiles from distribution of all μ_{95}
- **Expected limit = median of μ_{95} distribution**
  - 16 and 84% quantiles: 68% confidence interval
  - 2.5 and 97.5% quantiles: 95% confidence interval

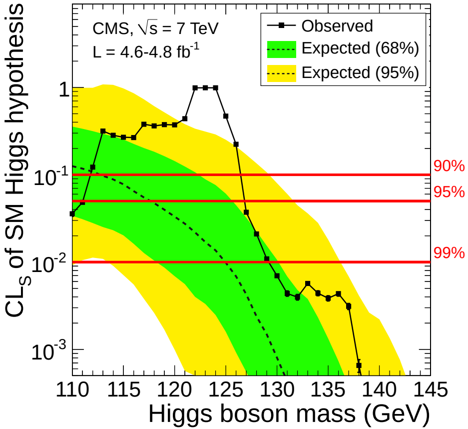
$\rightarrow$ "Brazilian band" plots





Before the Higgs-Boson Discovery

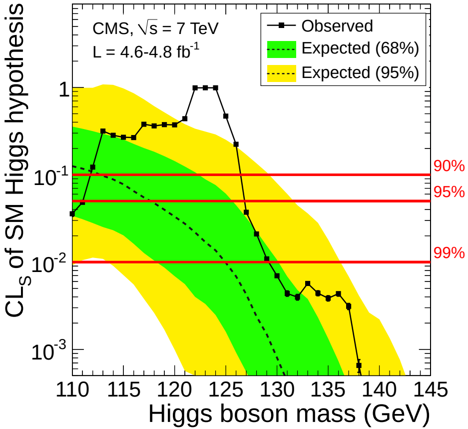
- Combination of Higgs-boson search results by CMS [Phys.Lett. B710 (2012) 26]



- Tested hypotheses
  - H_0: no Higgs boson
  - H_1: SM Higgs boson
- Test statistic q evaluated for SM Higgs boson of different mass ($\mu = 1$ in each case)
- Excluding a SM Higgs boson at 95% CL with masses
  - $m_H > 118\,\text{GeV}$ expected (from toy data under H_0)
  - $m_H > 127\,\text{GeV}$ observed (from real data)

Before the Higgs-Boson Discovery

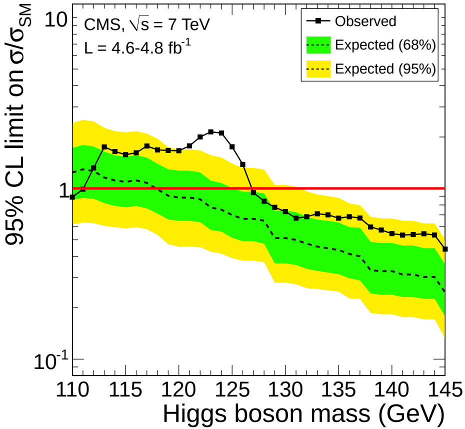
- Combination of Higgs-boson search results by CMS [Phys.Lett. B710 (2012) 26]



- Tested hypotheses
  - H_0: no Higgs boson
  - H_1: SM Higgs boson
- Test statistic q evaluated for SM Higgs boson of different mass ($\mu = 1$ in each case)
- Observed exclusion weaker than expected (smaller mass range)
- Around $m_H = 125\,\text{GeV}$, q_{obs} differs significantly from expectation: **indication that H_0 is wrong!**

Before the Higgs-Boson Discovery

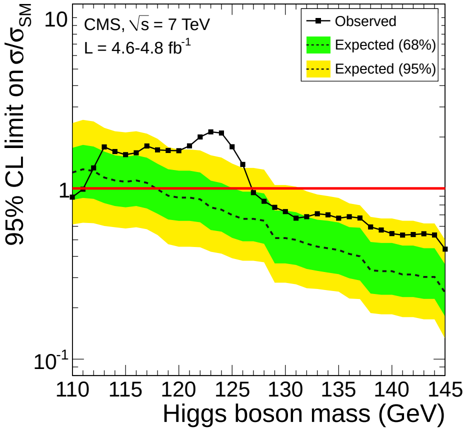
- Combination of Higgs-boson search results by CMS [Phys.Lett. B710 (2012) 26]



- Tested hypotheses
  - H_0: no Higgs boson
  - H_1: a Higgs boson
- Test statistic evaluated for a Higgs boson of different mass and variable signal strength
  - Signal strength μ can vary in each case (not SM any more!)
- Exclusion limits on μ for different masses at 95 % C.L.

Before the Higgs-Boson Discovery

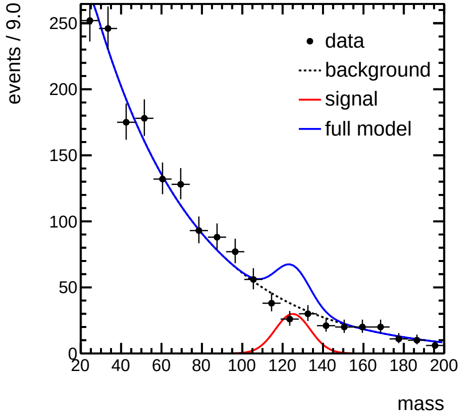
- Combination of Higgs-boson search results by CMS [Phys.Lett. B710 (2012) 26]



- Tested hypotheses
  - H_0: no Higgs boson
  - H_1: a Higgs boson
- Test statistic evaluated for a Higgs boson of different mass and variable signal strength
  - Signal strength μ can vary in each case (not SM any more!)
- Striking: observed limit weaker than expected around $m_H = 125\,\text{GeV}$
- Difference (locally) beyond 2σ: **indication that H_0 is wrong!**

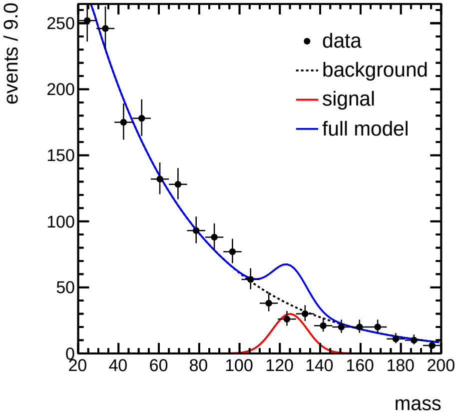
What If the Data Trick Us?

- Suppose **data fluctuates low**, sizably below background expectation
- **CL$_{s+b}$: artificially strong limit** on signal, i. e. $\mu_{1-\alpha}$ is small



What If the Data Trick Us?

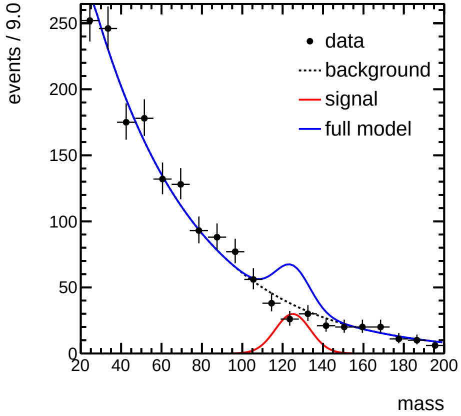
- Suppose **data fluctuates low**, sizably below background expectation
- **$\mathbf{CL_{s+b}}$: artificially strong limit** on signal, i. e. $\mu_{1-\alpha}$ is small



- In extreme case, $\mu_{1-\alpha} \rightarrow 0$
  i. e. exclude signal entirely
- **Not desirable**: just downward fluctuation of the data!
- Often **problem: searches in extreme phase-space regions** with few background events

What If the Data Trick Us?

- Suppose **data fluctuates low**, sizably below background expectation
- **CL_{s+b}: artificially strong limit** on signal, i. e. $\mu_{1-\alpha}$ is small



CL_s method

- Compute both

$$\mathrm{CL}_{s+b} = \int_{q_{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu)|H_1)$$

$$\mathrm{CL}_b \equiv \int_{q_{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu)|H_0)$$

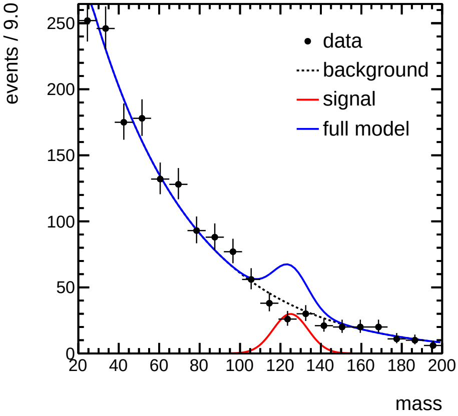
- Define limit as that μ for which

$$\boxed{\mathrm{CL}_s \equiv \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \alpha}$$

normalise CL_{s+b} to “background-only p-value”

What If the Data Trick Us?

- Suppose **data fluctuates low**, sizably below background expectation
- **$\mathrm{CL_{s+b}}$: artificially strong limit** on signal, i. e. $\mu_{1-\alpha}$ is small



$\mathrm{CL_s}$ method

- Compute both

$$\mathrm{CL_{s+b}} = \int_{q_\mathrm{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu)|H_1)$$

$$\mathrm{CL_b} \equiv \int_{q_\mathrm{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu)|H_0)$$

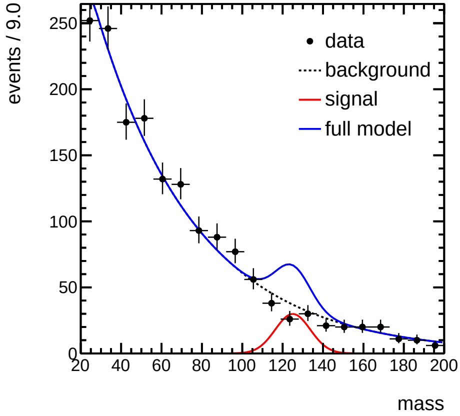
- Define limit as that μ for which

$$\boxed{\mathrm{CL_s} \equiv \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}} = \alpha}$$

- In case of extreme under fluctuation: $\mathrm{CL_s} \to 1$
- H_1 not excluded by mistake – but also weaker limit in case of no signal

What If the Data Trick Us?

- Suppose **data fluctuates low**, sizably below background expectation
- **$\mathrm{CL_{s+b}}$: artificially strong limit** on signal, i. e. $\mu_{1-\alpha}$ is small



$\mathrm{CL_s}$ method

- Compute both

$$\mathrm{CL_{s+b}} = \int_{q_\mathrm{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu)|H_1)$$

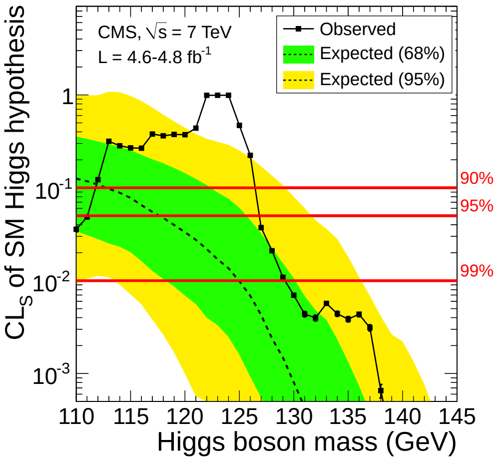
$$\mathrm{CL_{b}} \equiv \int_{q_\mathrm{obs}}^{\infty} \mathrm{d}q\, \mathcal{P}(q(\mu)|H_0)$$

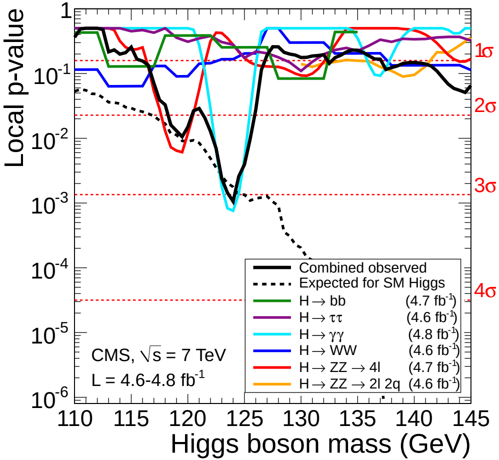
- Define limit as that μ for which

$$\boxed{\mathrm{CL_s} \equiv \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}} = \alpha}$$

$\mathrm{CL_s}$ protects from fluctuations in the data at cost of lower sensitivity
(procedure used in LHC (Higgs boson) searches)

Et Voilà





Phys.Lett. B710 (2012) 26–48

Summary

- **Statistical analysis crucial tool** in particle physics
- Does not tell probability of a certain model (at least not without further assumptions) but allows
  - quantifying the compatibility of the data with a tested model, e. g. via p **value or significance**
  - determining the parameter values of a given model that describe best the data (estimators), e. g. **via maximum-likelihood fit**
- In practice often **comparison of two alternative hypotheses** H_0 and H_1, e. g. background-only and signal+background
  - Rules when to reject H_0 in favour of H_1: allow quantifying type-I/II errors
  - Test statistic combines information of multi-channel data into one single number for application in hypothesis testing
  - **Likelihood ratio is most powerful test statistic**
- Exclusion **limits provide information on model parameter in case no signal found**

Literature

- R. J. Barlow, “Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences”, The Manchester Physics Series **highly recommended**
- G. Cowan, “Statistical Data Analysis”, Oxford Science Publications
- Review “Statistics”, R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022) and 2023 update **very good review**
- The ATLAS and CMS Collaborations, “Procedure for the LHC Higgs boson search combination in Summer 2011”, CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11 **expert document**