Statistical Methods used for Higgs Boson Searches

Roger Wolf
03. June 2014
Recap from Last Time (Simulation of Processes)

- From “paper & pen” statements to high precision predictions on observable quantities (at the LHC):

\[ \sigma_{QCD} = \sum_{jk} \int dx_j dx_k f_j(x_j, \mu_F^2) f_k(x_k, \mu_F^2) \hat{\sigma}(x_j x_k s, \mu_F^2, \mu_R^2) \]

- Discussed in lectures 1-3.
Recap from Last Time (Data Analysis)

- Observable → real measurement:
Recap from Last Time (Data Analysis)

- Observable → real measurement:

Data preparation techniques:

- Reconstruction of traces in the detector units.
- Alignment of track detectors.
- Calibration of energy response.
- Reconstruction & selection efficiency ("Tag & probe", "MC Embedding")
- How well are background processes understood?
Institute of Experimental Particle Physics (IEKP)

of Today

- Observable → real measurement:

Data preparation techniques:
- Reconstruction & selection efficiency ("Tag & probe", "MC Embedding")
- How well are background processes understood?

How to establish a new (small) signal on top of a "reasonably" well known background?

Transverse slice through CMS

Electromagnetic Calorimeter

Hadron Calorimeter

Superconducting Solenoid
Quiz of the Day

- What is the relation between the Binomial, Gaussian & Poisson distribution?
- What is the relation between a minimal $\chi^2$ fit and a Maximum Likelihood fit?
- How exactly do I calculate a 95% CL limit and how does it relate to classical hypothesis tests?
Quiz of the Day

• What is the relation between the Binomial, Gaussian & Poisson distribution?

• What is the relation between a minimal $\chi^2$ fit and a Maximum Likelihood fit?

• How exactly do I calculate a 95% CL limit and how does it relate to classical hypothesis tests? Can you interpret this plot?
Quiz of the Day

- What is the relation between the Binomial, Gaussian & Poisson distribution?

- What is the relation between a minimal $\chi^2$ fit and a Maximum Likelihood fit?

- How exactly do I calculate a 95% CL limit and how does it relate to classical hypothesis tests? Can you interpret this plot?

- What does a “3σ evidence” or a “5σ discovery” mean?
Schedule for Today

1. Probability distributions & Likelihood functions.

2. Parameter estimates (=fits).

3. Limits, p-values, significances.
Schedule for Today

1. Probability distributions & Likelihood functions.
2. Parameter estimates (=fits).
3. Limits, p-values, significances.

Walk through statistical methods that will appear in the next lectures:

- You will see all these methods acting in real life during the next lectures.
- To learn about the interiors of these methods check KIT lectures of Modern Data Analysis Techniques.
Theory:

- QM wave functions are interpreted as probability density functions.

- The Matrix Element, $S_{fi}$, gives the probability to find final state $f$ for given initial state $i$.

- Each of the statistical processes
  $pdf \rightarrow ME \rightarrow hadronization \rightarrow energy \ loss \ in \ material \rightarrow digitization$

  are statistically independent.

- Event by event simulation using Monte Carlo integration methods.
Statistics ↔ Particle Physics

Theory:

• QM wave functions are interpreted as probability density functions.

• The Matrix Element, \( S_{fi} \), gives the probability to find final state \( f \) for given initial state \( i \).

• Each of the statistical processes
  \( pdf \rightarrow ME \rightarrow hadronization \rightarrow energy \ loss \ in \ material \rightarrow digitization \)
  are statistically independent.

• Event by event simulation using Monte Carlo integration methods.

Experiment:

• All measurements we do are derived from rate measurements.

• We record millions of trillions of particle collisions.

• Each of these collisions is independent from all the others.
Statistics ↔ Particle Physics

Theory:
• QM wave functions are interpreted as probability density functions.
• The Matrix Element, $S_{fi}$, gives the probability to find final state $f$ for given initial state $i$.
• Each of the statistical processes $pdf \rightarrow ME \rightarrow$ hadronization $\rightarrow$ energy loss in material $\rightarrow$ digitization are statistically independent.
• Event by event simulation using Monte Carlo integration methods.

Experiment:
• All measurements we do are derived from rate measurements.
• We record millions of trillions of particle collisions.
• Each of these collisions is independent from all the others.

• Particle physics experiments are a perfect application for statistical methods.
Characterization of Probability Distributions

- Expectation Value:

\[ E[x] = \int_{-\infty}^{\infty} x \cdot pdf(x) dx = \mu \]

- Variance:

\[ V[x] = \int_{-\infty}^{\infty} (x - \mu) \cdot pdf(x) dx = \sigma^2 \]

\[ = E[(x - E[x])^2] = E[x^2 - 2xE[x] + E^2[x]] = E[x^2] - E^2[x] \]

- Covariance:

\[ cov[x, y] = E[(x - \mu(x))(y - \mu(y))] = \int_{-\infty}^{\infty} x \cdot y \cdot pdf(x, y) dx = E[xy] - \mu(x)\mu(y) \]

- Correlation coefficient:

\[ \rho(x, y) = \frac{cov[x, y]}{\mu(x)\mu(y)} \]
Probability Distributions

Expectation: $\mu = np$

Variance: $\sigma^2 = np(1 - p)$

$$P(k, n, p) = \binom{n}{k} p^k \cdot (1 - p)^{n-k}$$
(Binomial distribution)
**Probability Distributions**

<table>
<thead>
<tr>
<th>Probability Distribution</th>
<th>Expectation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Gaussian distribution)</td>
<td>$\mu = np$</td>
<td>$\sigma^2 = np(1 - p)$</td>
</tr>
<tr>
<td>(Binomial distribution)</td>
<td>$\mu = np$</td>
<td>$\sigma^2 = np(1 - p)$</td>
</tr>
</tbody>
</table>

$P(k, n, p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}(\frac{k-np}{np(1-p)})^2}$

$n \to \infty$, $p$ fixed

Central limit theorem of de Moivre & Laplace.
Probability Distributions

\[ \mathcal{P}(k, n, p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2} \left( \frac{k-np}{np(1-p)} \right)^2} \]

(Gaussian distribution)

\[ n \to \infty \text{, } p \text{ fixed} \]

Central limit theorem of de Moivre & Laplace.

\[ \mathcal{P}(k, n, p) = \binom{n}{k} p^k \cdot (1-p)^{n-k} \]

(Binomial distribution)

\[ n \to \infty \text{, } np \text{ fixed} \]

Will be shown on next slide.

\[ \mathcal{P}(k, n, p) = \frac{(np)^k}{k!} e^{-np} \]

(Poisson distribution)

<table>
<thead>
<tr>
<th>Expectation:</th>
<th>Variance:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu = np )</td>
<td>( \sigma^2 = np(1 - p) )</td>
</tr>
<tr>
<td>( \mu = np )</td>
<td>( \sigma^2 = np(1 - p) )</td>
</tr>
<tr>
<td>( \mu = np )</td>
<td>( \sigma^2 = \mu = np )</td>
</tr>
</tbody>
</table>
Probability Distributions

\[ \mathcal{P}(k, n, p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2} \left( \frac{k-np}{np(1-p)} \right)^2} \]

(Gaussian distribution)

\[ n \to \infty, \ p \ \text{fixed} \]

Central limit theorem of de Moivre & Laplace.

\[ \mathcal{P}(k, n, p) = \binom{n}{k} p^k \cdot (1-p)^{n-k} \]

(Binomial distribution)

\[ n \to \infty, \ np \ \text{fixed} \]

Will be shown on next slide.

\[ \mathcal{P}(k, n, p) = \frac{(np)^k}{k!} e^{-np} \]

(Poisson distribution)

<table>
<thead>
<tr>
<th>Expectation:</th>
<th>Variance:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu = np )</td>
<td>( \sigma^2 = np(1 - p) )</td>
</tr>
</tbody>
</table>

motivation for uncertainty.
Binomial ↔ Poisson Distribution

\[ P(k, n, p) = \binom{n}{k} p^k \cdot (1 - p)^{n-k} \]

\[ = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \cdot \frac{\mu^k}{n^k} \cdot \left(1 - \frac{\mu}{n}\right)^n \]

\[ = \frac{1 \cdot (1 - \frac{1}{n}) (1 - \frac{2}{n}) \cdots (1 - \frac{k-1}{n})}{(1 - \frac{\mu}{n})^k} \cdot \frac{\mu^k}{k!} \cdot \left(1 - \frac{\mu}{n}\right)^n \]

\[ = \frac{1}{(1 - \frac{\mu}{n})} \cdot \frac{(1 - \frac{2}{n})}{(1 - \frac{\mu}{n})} \cdot \frac{(1 - \frac{2}{n})}{(1 - \frac{\mu}{n})} \cdots \cdot \frac{(1 - \frac{k-1}{n})}{(1 - \frac{\mu}{n})} \cdot \frac{\mu^k}{k!} \cdot \left(1 - \frac{\mu}{n}\right)^n \]

\[ = \frac{\mu^k}{k!} e^{-\mu} \]

\[ \mu = \text{const}, \ n \to \infty \]
Uncertainties on Counting Experiments

\[ P(k, \mu_i) = \frac{\mu_i^k}{k!} e^{-\mu_i} \]

\[ \sqrt{k} \text{ uncertainty} \]

\[ k \]

counting experiment
Uncertainties on Counting Experiments

\[ \mathcal{P}(k, \mu_i) = \frac{\mu_i^k}{k!} e^{-\mu_i} \]

Number of events in \( bin_i \) depends on \( n \) and on probability \( p_i = \int_{i}^{i+\delta} pdf \).

Binned Histogram

\( \sqrt{k} \) uncertainty

Counting experiment

Underlying pdf
Relations between Probability Distributions

Central Limit Theorem:
Random variable variable made up of a sum of many single measurements.

\[ n \to \infty, p = \text{cont} \]

\[ n \to \infty, np = \text{cont} \]

Gaussian

Binomial

Poisson

Look for something that is very rare very often.
Relations between Probability Distributions

Log-normal

Random variable variable made up of a sum of many single measurements.

Central Limit Theorem:
Random variable variable made up of a product of many single measurements.

Gaussian

$n \to \infty, p = \text{cont}$

Binomial

$n \to \infty, np = \text{cont}$

Poisson

Look for something that is very rare very often.

Lognormal Density

$\mu = 0$

$\sigma^2 = 0.1$

$\sigma^2 = 0.5$

$\sigma^2 = 1.0$
Relations between Probability Distributions

Log-normal

Central Limit Theorem:
Random variable variable made up of a sum of many single measurements.

Gaussian

$\chi^2$ Distribution

Random variable variable made up of a product of many single measurements.

Binomial

$n \to \infty, p = \text{cont}$

Poisson

$n \to \infty, np = \text{cont}$

What does the parameter $k$ correspond to in the $\chi^2$ distributions?

Look for something that is very rare very often.
Relations between Probability Distributions

Log-normal

Central Limit Theorem:
Random variable variable made up of a sum of many single measurements.

Random variable variable made up of a product of many single measurements.

\[ \exp \]

Gaussian

\[ n \to \infty, p = \text{cont} \]

\[ \chi^2 \text{ Distribution} \]

Binomial

\[ n \to \infty, np = \text{cont} \]

Poisson

What does the parameter \( k \) correspond to in the \( \chi^2 \) distributions?

Look for something that is very rare very often.
Likelihood Functions

- **Problem**: truth is not known!
- Deduce “truth” from measurements (usually in terms of models).
- **Likeliness of a model to be true** quantified by *likelihood function*
  \[ \mathcal{L}(\{k_i\}, \{\kappa_j\}) \].

*model parameters.*

*measured number of events (e.g. in bins \(i\)).*
Likelihood Functions

- **Problem**: truth is not known!
- Deduce “truth” from measurements (usually in terms of models).
- **Likelihood of a model** to be true quantified by *likelihood function* $\mathcal{L}(\{k_i\}, \{\kappa_j\})$.
- Example: signal on top of known background in a binned histogram:

\[
\mathcal{L}(\{k_i\}, \{\kappa_j\}) = \prod_i \mathcal{P}(k_i, \mu_i(\kappa_j))
\]

Product of *pdfs* for each bin (Poisson).

- Model parameters:
- Measured number of events (e.g. in bins $i$).

\[
\mu_i(\kappa_j) = k_0 \cdot e^{-k_1 x_i} + k_2 \cdot e^{-\left(\kappa_3 - x_i\right)^2}
\]

- Background
- Signal
Parameter Estimates
Parameter Estimates

- **Problem**: find most probable parameter(s) \( \kappa_j \) of a given model.
- Usually minimization of negative \( \ln \) likelihood function (\( NLL \)):
  - \( \ln \) is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians \( NLL \) turns into \( \chi^2 \) minimization:
Parameter Estimates

- **Problem**: find most probable parameter(s) $\kappa_j$ of a given model.

- Usually minimization of negative $\ln$ likelihood function ($NLL$):
  - $\ln$ is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians $NLL$ turns into $\chi^2$ minimization:

  Clear to everybody?
Parameter Estimates

- **Problem**: find most probable parameter(s) $\kappa_j$ of a given model.
- Usually minimization of negative $\ln$ likelihood function ($NLL$):
  - $\ln$ is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians $NLL$ turns into $\chi^2$ minimization:

\[
NLL = -\ln \left( \prod_i \frac{1}{\sigma_i} \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \right) \propto \sum_i \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2
\]

Clear to everybody?

Number of $\mu_i$ determines dimension of the Gaussian distribution.
Parameter Estimates

- **Problem**: find most probable parameter(s) \( \kappa_j \) of a given model.
- Usually minimization of **negative** \( \ln \) likelihood function (**NLL**):
  - \( \ln \) is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians **NLL turns into** \( \chi^2 \) minimization:

\[
NLL = -\ln \left( \prod_i e^{-\frac{1}{2} \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2} \right) \propto \sum_i \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2
\]

- The minimization usually performed:
  - **analytically** (like in an optimization exercise in school).
  - **numerically** (usually the more general solution).
  - by **scan of the NLL** (for sure the most robust method).
Parameter(s) of Interest (POI)

- Each case/problem defines its own parameter(s) of interest (POI's):
  - POI could be the mass $\kappa_3$.

- Example:
  signal on top of known background in a binned histogram:

\[
\mathcal{L}(\{\kappa_i\}, \{\kappa_j\}) = \prod_i \mathcal{P}(\kappa_i, \mu_i(\kappa_j))
\]

Product of pdfs for each bin (Poisson).

\[
\mu_i(\kappa_j) = \kappa_0 \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}
\]

- background
- signal
Parameter(s) of Interest (POI)

- Each case/problem defines its own *parameter(s) of interest (POI's)*:
  - POI could be the mass $\kappa_3$.
  - In our case POI usually is the signal strength $\kappa_2$ for a fixed value for $\kappa_3$.

- Example:
  signal on top of known background in a binned histogram:

\[
\mathcal{L}(\{k_i\}, \{\kappa_j\}) = \prod_i \mathcal{P}(k_i, \mu_i(\kappa_j))
\]

Product of pdfs for each bin (Poisson).

\[
\mu_i(\kappa_j) = \kappa_0 \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}
\]

background  signal
Systematic Uncertainties

- Systematic uncertainties are usually incorporated as *nuisance parameters*:

- Example: assume background normalization $\kappa_0$ is not absolutely known, but with an uncertainty $\sigma(\kappa_0)$:

$$
\mu_i(\kappa_j) = \mathcal{P}'(\tilde{\kappa}_0, \kappa_0, \sigma(\kappa_0)) \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}
$$

- Example: signal on top of known background in a binned histogram:

$$
\mathcal{L}(\{k_i\}, \{\kappa_j\}) = \prod_i \mathcal{P}(k_i, \mu_i(\kappa_j))
$$

Product of pdfs for each bin (Poisson).

$$
\mu_i(\kappa_j) = \kappa_0 \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}
$$

<table>
<thead>
<tr>
<th>mass [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>150</td>
</tr>
</tbody>
</table>

uncertainty

expected value

possible values in single measurements
Hypothesis Tests
Hypothesis Separation

- Start with two alternative hypotheses $H_0$ & $H_1$.
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio ($LR$):

$$q = -2 \ln \left( \frac{L(\text{obs}|H_1)}{L(\text{obs}|H_0)} \right)$$

- $q$ can be calculated for the observation ($\text{obs}$), for the expectation for $H_0$ and for the expectation for $H_1$:
  - Observed is a single value (outcome of measurement).
  - Expectation is a mean value with uncertainties based on toy measurements.
Hypothesis Separation

- Start with two alternative hypotheses $H_0$ & $H_1$.
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio ($LR$).

$$q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$$

- $q$ can be calculated for the observation ($\text{obs}$), for the expectation for $H_0$ and for the expectation for $H_1$:
  - Observed is a single value (outcome of measurement).
  - Expectation is a mean value with uncertainties based on toy measurements.
Test Statistics (LEP)

- Start with two alternative hypotheses $H_0$ & $H_1$.
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio ($LR$):

$$q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$$

\[
\mathcal{L}(n| b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)
\]

\[
\mathcal{L}(n| \mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)
\]

$$q_\mu = -2 \ln \left( \frac{\mathcal{L}(n|\mu s + b)}{\mathcal{L}(n|b)} \right), \quad 0 \leq \mu$$

nuisance parameters $\tilde{\kappa}_j$ integrated out (by throwing toys → MC method) before evaluation of $q_\mu$ (→marginalization).
Test Statistics (Tevatron)

- Start with two alternative hypotheses $H_0$ & $H_1$.
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio ($LR$):

$$q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$$

\[ \mathcal{L}(n|b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j) \]

\[ \mathcal{L}(n|\mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j) \]

$$q_\mu = -2 \ln \left( \frac{\mathcal{L}(n|\mu s(\kappa_\mu) + b(\kappa_\mu))}{\mathcal{L}(n|b(\kappa_\mu=0))} \right), \quad 0 \leq \mu$$

nominator maximized for given $\mu$ before marginalization. Denominator for $\mu = 0$. Better estimates on nuisance parameters. Reduces uncertainties on nuisance parameters.
Test Statistics (LHC)

- Start with two alternative hypotheses $H_0$ & $H_1$.
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio (LR):

$$q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$$

\[
\mathcal{L}(n|b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\bar{\kappa}_j) \\
\mathcal{L}(n|\mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\bar{\kappa}_j) \\
q_\mu = -2 \ln \left( \frac{\mathcal{L}(n|\mu s(\bar{\kappa}_\mu) + b(\bar{\kappa}_\mu))}{\mathcal{L}(n|\bar{\mu} s(\bar{\kappa}_\bar{\mu}) + b(\bar{\kappa}_\bar{\mu}))} \right), \quad 0 \leq \hat{\mu} \leq \mu
\]

nominator maximized for given $\mu$ before marginalization. For the denominator a global maximum is searched for at $\hat{\mu}$. In addition allows use of asymptotic formulas ($\rightarrow$ no need for toys).
Classical Hypothesis Testing

- Classical hypothesis test interested in probability to observe $q_{\text{obs}}$ given that $H_0$ or $H_1$ is true:

$$
\begin{array}{c|c}
q \leq q_{\text{obs}}|_{H_1} & q \geq q_{\text{obs}}|_{H_1} \\
q \leq q_{\text{obs}}|_{H_0} & q \geq q_{\text{obs}}|_{H_0}
\end{array}
$$

$q_{\text{obs}}$ defines upper bound
$q_{\text{obs}}$ defines lower bound

- We are usually interested in “upper limits”, which correspond to “lower bounds” ($\rightarrow$ how often signal $\leq$ observed deviation?).
95% CL Upper Limits

- Our pdf's usually depend on another parameter, which is the actual POI (μ in SM, tan β in MSSM case).
- Traditionally we set 95% CL upper limits on this POI.

- pdf's move apart from each other.
- The more separate the pdf's are the more $H_0$ & $H_1$ are distinguishable.
- Find POI$_i$ for which:

$$\mathcal{I}_{\text{POI}} = \int_{-\infty}^{q_{\text{obs}}} \text{pdf} = 0.05$$

for this POI$_i$ in 95% of all toys $q \geq q_{\text{obs}}$. 
95% CL Upper Limits

- Our pdf's usually depend on another parameter, which is the actual POI (μ in SM, \( \tan \beta \) in MSSM case).

- Traditionally we set 95% CL upper limits on this POI.

95% CL Upper Limit:

- \( POI_i \) is the value at which in case that \( H_1 \) is the true hypothesis, the chance that \( q \geq q_{\text{obs}} \) is 95%.

- Still there is a chance of 5% that \( q < q_{\text{obs}} \).

For this \( POI_i \), in 95% of all toys \( q \geq q_{\text{obs}} \).
95% CL Upper Limits

- Our pdf's usually depend on another parameter, which is the actual POI (μ in SM, \(\tan \beta\) in MSSM case).
- Traditionally we set 95% CL upper limits on this POI.

**95% CL Upper Limit:**

- POI\(_i\) is the value at which in case that \(H_1\) is the true hypothesis the chance that \(q \geq q_{\text{obs}}\) is 95%.
- Still there is a chance of 5% that \(q < q_{\text{obs}}\).
- Assume our POI is \(\mu\): does the 90% CL upper limit on \(\mu\) correspond to a higher or a lower value \(\mu_{90\%}\)?
95% CL Upper Limits

- Our pdf's usually depend on another parameter, which is the actual POI (μ in SM, tan β in MSSM case).

- Traditionally we set 95% CL upper limits on this POI.

95% CL Upper Limit:

- **POI** is the value at which in case that $H_1$ is the true hypothesis the chance that $q \geq q_{obs}$ is 95%.

- Still there is a chance of 5% that $q < q_{obs}$.

- Assume our POI is μ: does the 90% CL upper limit on μ correspond to a higher or a lower value μ$_{90\%}$? → It's lower!

\[ \int_{-\infty}^{q_{obs}} \text{pdf} = 0.05 \]

1% probability of $q$ to be “more background like” than $q_{obs}$.
• In particle physics we set more conservative limits than this, following the CLs method:

• Assume $H_1$ to be signal+background and $H_0$ to be background only hypothesis.

\[
\text{CL}(S + B) = \int_{-\infty}^{q_{\text{obs}}} p df_{H_1}
\]

\[
\text{CL}(B) = \int_{-\infty}^{q_{\text{obs}}} p df_{H_0}
\]

• Find $\text{POI}_i$ for which:

\[
\text{CL}_S = \frac{\text{CL}(S+B)}{\text{CL}(B)} = 0.05
\]
CLs Limits

- In particle physics we set more conservative limits than this, following the CLs method:
- Assume $H_1$ to be signal+background and $H_0$ to be background only hypothesis.

\[
\text{CL}(S + B) = \int_{-\infty}^{q_{\text{obs}}} p df_{H_1} \\
\text{CL}(B) = \int_{-\infty}^{q_{\text{obs}}} p df_{H_0}
\]

- Find POI$_i$ for which:
  \[
  \text{CLS} = \frac{\text{CL}(S+B)}{\text{CL}(B)} = 0.05
  \]
- If $H_0 \& H_1$ are clearly distinguishable $\text{CLS} \rightarrow \text{CL}(S + B)$.
In particle physics we set more conservative limits than this, following the CLs method:

- Assume $H_1$ to be signal+background and $H_0$ to be background only hypothesis.

POI$_i$, POI$_{i+1}$, POI$_{i+2}$

\[
CL(S + B) = \int_{-\infty}^{q_{\text{obs}}} pdf_{H_1}
\]

\[
CL(B) = \int_{-\infty}^{q_{\text{obs}}} pdf_{H_0}
\]

- Find POI$_i$ for which:

\[
CL_S = \frac{CL(S+B)}{CL(B)} = 0.05
\]

- If $H_0$ & $H_1$ are clearly distinguishable $CL_S \rightarrow CL(S + B)$.

- If they cannot be distinguished $CL_S > CL(S + B)$. 

interested in integration of magenta pdf & blue pdf from below.
In particle physics we set more conservative limits than this, following the CLs method:

- Assume $H_1$ to be signal+background and $H_0$ to be background only hypothesis.
**Expected Limit (canonical approach)**

- To obtain the expected limit *mimic calculation of observed*, but base it on toy experiments.

- Make use of the fact that the *pdf*’s do not depend on toys (i.e. schematic plot on the left does not change).

- Throw number of toys under the BG only hypothesis ($H_0$) determine distribution of 95% CL limits on POI.

- Obtain quantiles for expected limit from this distribution.
And if the signal shows up...
p-Value

• How do we know whether what we see is not just a background fluctuation?

• The p-value is the probability \( P(q \geq q_{\text{obs}} | H_0) \) to observe values of \( q \) larger than \( q_{\text{obs}} \) under the assumption that the background only hypothesis \( H_0 \) is the true hypothesis.

• Think of…
  … the limit as a way to falsify the signal plus background hypothesis \( (H_1) \).
  … the p-value as a way to falsify the background only hypothesis \( (H_0) \).
Significance

• If the measurement is normal distributed $q$ is distributed according to a $\chi^2$ distribution.

• The $\chi^2$ probability can then be interpreted as a Gaussian confidence interval.

p-values:
\[ P(q \geq 3\sigma|H_0) = 1 \cdot 10^{-3} \]
\[ P(q \geq 5\sigma|H_0) = 2 \cdot 10^{-5} \]
Significance (in practice)

- If the measurement is normal distributed $q$ is distributed according to a $\chi^2$ distribution.
- The $\chi^2$ probability can then be interpreted as a Gaussian confidence interval.
- Usual approximation in practice is to estimate significances by:

$$S = \frac{n_{\text{obs}} - n_b}{\sqrt{n_b}}$$
Significance (in practice)

- If the measurement is normal distributed, \( q \) is distributed according to a \( \chi^2 \) distribution.

- The \( \chi^2 \) probability can then be interpreted as a Gaussian confidence interval.

- Usual approximation in practice is to estimate significances by:

\[
S = \frac{n_{\text{obs}} - n_b}{\sqrt{n_b}}
\]

![Graph showing expected signal events vs. mass]
Significance (in practice)

- If the measurement is normal distributed, \( q \) is distributed according to a \( \chi^2 \) distribution.
- The \( \chi^2 \) probability can then be interpreted as a Gaussian confidence interval.
- Usual approximation in practice is to estimate significances by:

\[
S = \frac{n_{\text{obs}} - n_b}{\sqrt{n_b}}
\]

expected signal events

Poisson uncertainty on expected background events.
Significance (in practice)

- If the measurement is normal distributed, $q$ is distributed according to a $\chi^2$ distribution.
- The $\chi^2$ probability can then be interpreted as a Gaussian confidence interval.
- Usual approximation in practice is to estimate significances by:

$$S = \frac{n_{\text{obs}} - n_b}{\sqrt{n_b}}$$

expected signal events

Poisson uncertainty on expected background events.
Concluding Remarks

- Reviewed all **statistical tools necessary to search for the Higgs signal** ($\rightarrow$ as a small signal above a known background):
  - Probability distributions, likelihood functions, limits, p-values, ...

- Limits are a usual way to 'exclude' the signal hypothesis ($H_1$).

- p-values are a usual way to 'exclude' the background hypothesis ($H_0$).

- Under the assumption that the test statistic $q$ is $\chi^2$ distributed p-values can be translated into **Gaussian confidence intervals** $\sigma$.

- In particle physics we call an observation with $\geq 3\sigma$ an **evidence**.

- We call an observation with $\geq 5\sigma$ a **discovery**.
Concluding Remarks

- Reviewed all statistical tools necessary to search for the Higgs signal (→ as a small signal above a known background):
  - Probability distributions, likelihood functions, limits, p-values, ...
  - Limits are a usual way to 'exclude' the signal hypothesis \( (H_1) \).
  - p-values are a usual way to 'exclude' the background hypothesis \( (H_0) \).
  - Under the assumption that the test statistic \( q \) is \( \chi^2 \) distributed p-values can be translated into Gaussian confidence intervals \( \sigma \).
  - In particle physics we call an observation with \( \geq 3\sigma \) an evidence.
  - We call an observation with \( \geq 5\sigma \) a discovery.
  - Once a measurement is established the search is over! Measurements of properties are new and different world!
Sneak Preview for Next Week

- Review indirect estimates of the Higgs mass and searches for the Higgs boson that have been made before 2012:
  - Estimates of $m_t$ and $m_H$ from high precision measurements at the Z-pole mass at LEP.
  - Direct searches for the Higgs boson at LEP.
  - Direct searches for the Higgs boson at the Tevatron.

- For the remaining lectures we then will turn towards the discovery of the Higgs boson at the LHC.

During the next lectures we will see 1:1 life examples of all methods that have been presented here.
Backup & Homework Solutions