Statistical Methods used for Higgs Boson Searches

Roger Wolf
Schedule for today

• What is the meaning of the degrees of freedom of the $\chi^2$ function?

• What is the relation between the likelihood function and the $\chi^2$ estimate?

1 Likelihood analyses
2 Parameter estimates
3 $p$-value, significance and limit setting
Statistics vs. particle physics

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.

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.

• Particle physics experiments are a perfect application for statistical methods.
Statistics vs. probability theory (stochastic)

Test statistic:
\[ \Omega^n \to \mathbb{R} : \quad x \to f(x) \]

- NLL (\( q = -2 \ln(\mathcal{L}_1/\mathcal{L}_0) \)).
- Boosted Decision Tree (BDT) output.

Probability (density) function:
\[ \Omega^n \to [0, 1] \subset \mathbb{R} : \quad x \to \mathcal{P}(x) \]

- \( \mathcal{P}("6") = 3.572 \cdot 10^{-6} \).
- Laplacian paradox.

- Problem of statistics is usually *ill-defined*.
- Deduce *truth* from shadows in Platon's cave...
The case of “truth”

- Deduce *truth* from shadows:
  - Usually phrased in form of (nested) models (=*ideas* for Platon):
- Mathematically model = hypothesis.

**Statistics model:**

- Usually not questioned

**Uncertainty model:**

- Usually determined to best knowledge (not questioned)

**Physics model:**

- Usually competing models/ hypotheses will be discussed here!
## Probability distributions

### Expectation: $\mu = np$

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

**Binomial distribution**

$$\mathcal{P}(k, n, p) = \binom{n}{k} p^k \cdot (1 - p)^{n-k}$$
### Probability distributions

#### Gaussian distribution

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

- **Expectation:** \( \mu = np \)
- **Variance:** \( \sigma^2 = np(1 - p) \)

#### Binomial distribution

- **Probability mass function:**
  \[ P(k, n, p) = \binom{n}{k} p^k \cdot (1 - p)^{n-k} \]

- **Expectation:** \( \mu = np \)
- **Variance:** \( \sigma^2 = np(1 - p) \)

---

**Central limit theorem of de Moivre & Laplace.**

- As \( n \to \infty \), \( p \) fixed, the distribution tends to a normal distribution.

---

Institute of Experimental Particle Physics (IEKP)
### Probability distributions

#### (Binomial distribution)
- $n \rightarrow \infty$, $np$ fixed
- Central limit theorem of de Moivre & Laplace.

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

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

#### (Poisson distribution)
- $n \rightarrow \infty$, $np$ fixed
- Will be shown on next slide.

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

- **Expectation:** $\mu = np$
- **Variance:** $\sigma^2 = \mu = np$
\[ P(k, n, p) = \binom{n}{k} p^k \cdot (1 - p)^{n-k} \]

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

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

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

\[ \rightarrow 1 \]

\[ \rightarrow e^{-\mu} \]

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

\[ \mu = \text{const}, \ n \rightarrow \infty \]
Models for counting experiments

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

single experiment

Siméon Denis Poisson
(21.07.1781 – 25.04.1840)
Models for counting experiments

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

many experiments

Siméon Denis Poisson
(21.07.1781 – 25.04.1840)
Model building (likelihood functions)

- Likeliness of a model to be true quantified by **likelihood function** $\mathcal{L}({k_i}, \{\kappa_j\})$.

  $\prod_i \mathcal{P}(k_i, \mu_i) = \prod_i \frac{\mu_i^{k_i}}{k_i!} e^{-\mu_i}$

  - model parameters.
  - measured number of events (e.g. in bins $i$).

- Simple 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
Model building (likelihood functions)

- Likeliness of a model to be true quantified by likelihood function $\mathcal{L}(\{k_i\}, \{\kappa_j\})$.

- Simple 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).

$$
\mu_i(\kappa_j) = k_0 \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}
$$

EX: histogram with 25 bins; for each bin $\mathcal{P}(k_i, \mu_i(\kappa_j)) \geq 0.66$:

\begin{align*}
|\{k_i\}| &= 25 \\
|\{\kappa_j\}| &= 4 \\
\prod \mathcal{P}(k_i, \mu_i(\kappa_j)) &\sim 3.1 \cdot 10^{-5}
\end{align*}

NB: a value of a likelihood function as such is most of the time very close to zero, and w/o a reference in general w/o further meaning.
Distinguishing models \((\text{likelihood ratio})\)

- Task of likelihood analyses:
  do not determine likelihood of an experimental outcome per se, but distinguish models (=hypotheses) and determine the one that explains the experimental outcome best.

**Fundamental lemma of Neyman-Pearson:**

when performing a test between two simple hypotheses \(H_1\) and \(H_0\) the \textit{likelihood ratio test}, which rejects \(H_0\) in favor of \(H_1\) when

\[
Q = \frac{\mathcal{L}_{H_1}(\{k_i\}, \{\kappa_i\})}{\mathcal{L}_{H_0}(\{k_i\}, \{\kappa_i\})} \leq \eta
\]

\[
\mathcal{P}(Q(\{k_i\}, \{\kappa_i\}) \leq \eta | H_i) = \alpha
\]

is the most powerful test at significance level \(\alpha\) for a threshold \(\eta\).

- For \(q = -2 \ln Q\) this ratio turns into a difference \((\Delta \text{NLL})\).
Parameter estimates

Distinguish best parameter (set) in discrete or continuous transformations.
Maximum likelihood fit

- Each likelihood (ratio of) function(s) (with one or more parametric model part(s)) can be subject to a **maximum likelihood fit** (**NB**: negative log-likelihood finds its minimum where the log-likelihood is maximal...).

  - Minimization problem as known from school.
  - In our example e.g. four parameters $\kappa_i$.
  - Parameters can be constrained or unconstrained.

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

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

  Product 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

  The ATLAS+CMS Higgs couplings combined fit has $\mathcal{O}(4250)$ parameters and up to seven POI's.

  The CMS Tracker Alignment problem has $\mathcal{O}(50'000)$ parameters and several thousand POI's.
Parameter(s) of interest (POI)

- In a maximum likelihood fit each case/problem defines its own *parameter(s) of interest (POI's)*:
  - POI could be the mass ($\kappa_3$).

- Simple 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 for each bin (Poisson).

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

background \hspace{1cm} signal

\[ NB: \text{this is a likelihood ratio on its own.} \]
\[ NB: \text{I've also made the scan based on a likelihood ratio.} \]

Likelihood scan
Parameter(s) of interest (POI)

- In a maximum likelihood fit 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 $\kappa_3$).

- Simple 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 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**

**Likelihood scan**

NB: this is a likelihood ratio on its own

NB: I've also made the scan based on a likelihood ratio.
Incorporation of systematic uncertainties

- Systematic uncertainties are usually incorporated in form of *nuisance parameters*:
  - E.g. background normalization $\kappa_0$ not precisely known, but with uncertainty $\sigma(\kappa_0)$:
    \[ \mu_i(\kappa_j) = \kappa_0 \mathcal{P}'(x, 1, \sigma(\kappa_0)) \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2} \]
    - uncertainty
    - possible values of single “measurements” (integrated out)
    - expected value/best knowledge

- Simple example:
  signal on top of known background in a binned histogram:
  \[ \mathcal{L}\left(\{k_i\}, \{\kappa_j\}\right) = \prod_i \mathcal{P}(k_i, \mu_i(\kappa_j)) \]
  Product 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
Incorporation of systematic uncertainties

- Systematic uncertainties are usually incorporated in form of *nuisance parameters*:
  - E.g. background normalization $\kappa_0$ not precisely known, but with uncertainty $\sigma(\kappa_0)$:
    $$\mu_i(\kappa_j) = \kappa_0 \mathcal{P}'(x, 1, \sigma(\kappa_0)) \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}$$

Probability density function ($\mathcal{P}$)

Effect on BG normalization

$$-\ln Q = -\ln \left( \frac{\mathcal{L}_{H_1}(\{\kappa\}, \{\kappa\})}{\mathcal{L}_{H_0}(\{\kappa\}, \{\kappa\})} \right)$$
Relations between probability distributions

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

Log-normal

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

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

Gaussian

Central Limit Theorem:

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

\[ \left( -\ln \left( \frac{\chi^2}{2\pi} \right) - \frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right) dx \]

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

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

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

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

Poisson

Look for something that is very rare very often.
Relations between probability distributions

\[ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{\ln(x) - \mu}{\sigma} \right)^2} \, dx \]

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

Log-normal

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

Gaussian

Binomial

Poisson

\[ \left( -\ln \left( \sqrt{2\pi\sigma^2} \right) - \frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right) \, dx \]

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

Look for something that is very rare very often.
Example: saturated model

- Example of a likelihood ratio:

\[ q_\lambda = -2 \ln \left( \frac{\mathcal{L}(\text{data}_{\text{test}})}{\mathcal{L}(\text{data}_{\text{saturated}})} \right) \]

Model to be tested.
Model w/ as many parameters, \( \lambda_j \), as measurements.

e.g. one shape for each bin.

- Special case: (i) histogram; (ii) no further nuisance parameters; (iii) uncertainties normal distributed:

\[
\mathcal{L}(\text{data}_{\text{test}}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{(d_i - \lambda_i)^2}{2\sigma_i}} \\
\mathcal{L}(\text{data}_{\text{saturated}}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}} \\
q_\lambda = -2 \ln \left( \frac{\mathcal{L}(\text{data}_{\text{test}})}{\mathcal{L}(\text{data}_{\text{saturated}})} \right) = \sum_i \frac{(d_i - \lambda_i)^2}{\sigma_i}
\]

Generalization of the \( \chi^2 \) test.

- General case: (i) many histograms; (ii) many nuisance parameters:

\[
\text{CL of interest: } \int_{q_{\text{obs}}}^{+\infty} P_{\text{test}}
\]

Corresponds to 1.6\( \sigma \) compatibility
Hypothesis testing

Distinguish one preferred hypothesis \((H_0)\) against alternative hypotheses, in general in discrete but in special cases also in continuous transformations.

Full exclusion (here in \(m_{h^{mod}}\) scenario).
All further examples are taken from this very publication:

\[ PRL\ 106\ (2011)\ 231801 \]
Example: test statistics \( (\text{LEP }\sim 2000) \)

- Test signal \((H_1, \text{ for fixed mass, } m, \text{ and fixed signal strength, } \mu)\) vs. background-only \((H_0)\).

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

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

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

nuisance parameters \(\tilde{\kappa}_j\) integrated out before evaluation of \(q_\mu\) (\(\rightarrow\)marginalization).
Example: test statistics (Tevatron ~2005)

- Test signal ($H_1$, for fixed mass, $m$, and fixed signal strength, $\mu$) vs. background-only ($H_0$).

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

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

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

nominator maximized for given $\mu$ before marginalization. Denominator for $\mu = 0$. Better estimates of nuisance parameters w/ reduced uncertainties.
Example: test statistics \((LHC \sim 2010)\)

- Test signal \((H_1,\) for fixed mass, \(m,\) and fixed signal strength, \(\mu)\) vs. background-only \((H_0)\).

\[ \mathcal{L}(n|b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \]
\[ \mathcal{L}(n|\mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \]

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

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 more toys needed!\(^{(\cdot)}\)).

(*) will not be discussed further here.
Test statistic in life

• From the evaluation of the test statistic on data always obtain a plain value $q_{\text{obs}}$
  (in our discussion: $q_{\text{obs}} < 0$ – signal-like; $q_{\text{obs}} > 0$ – background-like).

• → True outcome of the experiment
  (nuisance parameters estimated to best knowledge, no uncertainties involved here)!

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

Product 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
Meaning and interpretation of the test statistic

- How compatible is $q_{\text{obs}}$ with $H_0$ or $H_1$? For this evaluate the test statistic on large number of toy experiments based on $H_0$ or $H_1$.

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

Product for each bin (Poisson).

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

background \hspace{1cm} signal

- Determine toy dataset.
- Determine toy values for all uncertainties.
- Determine value of $-q$ for each toy.
- Proceed as often as possible; do this for $H_0$ & $H_1$. 

$\mathcal{O}(50'000 \, \text{toys})$
Confidence levels ($CL$)

- The association to one or the other hypothesis can be performed up to a given confidence level $\alpha$.

\[
(1 - CL_b) = \int_{-\infty}^{q_{obs}} \mathcal{P}_b \quad (p\text{-value})
\]

\[
CL_{s+b} = \int_{q_{obs}}^{+\infty} \mathcal{P}_{s+b} \quad (CL_{s+b} \text{ confidence})
\]

\[
CL_b = \int_{q_{obs}}^{+\infty} \mathcal{P}_b \quad (CL_b \text{ confidence})
\]

\[
CL_s = \frac{CL_{s+b}}{CL_b} \quad (CL_s \text{ confidence})
\]

Attention: in all plots $-q$ is shown.

Institute of Experimental Particle Physics (IEKP)
The $p$-value

The association to one or the other hypothesis can be performed up to a given confidence level $\alpha$.

Probability to obtain values of $q$, which are at least as signal-like as $q_{\text{obs}}$. If $p$-value is small $H_0$ can be excluded. (*)

\[
(1 - CL_b) = \int_{-\infty}^{q_{\text{obs}}} P_b \quad \text{(p-value)}
\]

\[
CL_{s+b} = \int_{q_{\text{obs}}}^{+\infty} P_{s+b} \quad \text{(CL}_{s+b} \text{ confidence)}
\]

\[
CL_b = \int_{q_{\text{obs}}}^{+\infty} P_b \quad \text{(CL}_b \text{ confidence)}
\]

\[
CL_s = \frac{CL_{s+b}}{CL_b} \quad \text{(CL}_s \text{ confidence)}
\]

(*) Imagine data show a peak. What is the prob. that this is due to an upward fluctuation of the expectation from $H_0$.

Attention: in all plots $-q$ is shown.
Significance

- If the measurement is normal distributed $q$ is distributed according to a $\chi^2$ distribution (cf. slide 21f).
- The resulting $\chi^2$ probability is then equivalent to a Gaussian confidence interval in terms of standard deviations $\sigma$.

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

- If the measurement is normal distributed \( q \) is distributed according to a \( \chi^2 \) distribution (cf. slide 21f).

- The resulting \( \chi^2 \) probability is then equivalent to a Gaussian confidence interval in terms of standard deviations.

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

Deviation from expectation for \( H_0 \).

Poisson uncert. for \( H_0 \).
Excluding parameters

- Sorry, don't see any signal. Up to what size should I definitely have seen it?

- Usually $q$ depends on POI: $q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$
Excluding parameters

- Usually $q$ depends on POI: $q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$
Excluding parameters

- Usually $q$ depends on POI: $q = -2 \ln \left( \frac{\mathcal{L}_{\text{obs}}}{\mathcal{L}_{\text{obs}}|H_0} \right)$
Excluding parameters

- Usually $q$ depends on POI: $q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$
Excluding parameters

Challenging the $H_1$ hypothesis

- Usually $q$ depends on POI: $q = -2 \ln \left( \frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0} \right)$
Observed exclusion

- Traditionally we determine 95% CL exclusions on the POI ($\alpha = 0.05$).
- To be conservative choose probability $\alpha$ that $q$ is more BG-like than $q_{obs}$ low ($\rightarrow$ safer exclusion).
- $P(-q|H_0)$ and $P(-q|H_1)$ move apart from each other with increasing POI.
- The more separated $P(-q|H_0)$ and $P(-q|H_1)$ are the clearer $H_0$ and $H_1$ can be distinguished.
- For 95% CL identify value of POI for which: $CL_{s+b} = \int_{q_{obs}}^{+\infty} P_{s+b} = 0.05$
  for this value $q|H_1$ would have been more signal-like than $q_{obs}$ with 95% probability.
- There is still a 5% chance that we exclude by mistake.
Expected exclusion

- To obtain expected limit mimic calculation of observed; base it on toy datasets.

- Use fact that $P(-q|H_0)$ and $P(-q|H_1)$ do not depend on toys (i.e. schematic plot on the left does not change).

Throw toys under $H_0$ hypothesis; determine distribution of 95% CL limits on $POI$:

Obtain quantiles for expected exclusion from this distribution (expected limit = median).

Challenging the $H_1$ hypothesis
Interpretation issues (increasing pathology)

- Signal and BG hypothesis cannot be distinguished.
- Should this outcome lead to an exclusion of the signal hypothesis?

- $q_{\text{obs}}$ incompatible both with signal and BG hypothesis.
- Should this outcome lead to an exclusion of the signal hypothesis?

- $q_{\text{obs}}$ compatible with BG hypothesis.
- $q_{\text{obs}}$ incompatible with signal hypothesis.
Modified frequentist exclusion method ($CL_s$)

- In particle physics we set more conservative limits, following the $CL_s$ method:

\begin{itemize}
  \item $CL_{s+b} = \int_{q_{obs}}^{+\infty} P_{s+b}$
  \item $CL_b = \int_{q_{obs}}^{\infty} P_b$
  \item Identify value of POI for which:
  \[
  CL_s = \frac{CL_{s+b}}{CL_b} = 0.05
  \]
  \item If $H_0$ and $H_1$ become indistinguishable:
  \[
  CL_{s+b} < CL_s \rightarrow 1
  \]
\end{itemize}
Assume our POI is the signal strength $\mu$ of a new signal: does the 90% CL upper limit on $\mu$ correspond to a higher or a lower value than the 95% CL limit?
Judgment call

- Assume our POI is the signal strength $\mu$ of a new signal: does the 90% CL upper limit on $\mu$ correspond to a higher or a lower value than the 95% CL limit?

It's lower:

\[ \mu_{99\%} \quad \mu_{95\%} \quad \mu_{90\%} \]

\[ 1\% \quad 5\% \quad 10\% \]

\{ probability of $q$ to be “more background like” than $q_{\text{obs}}$. \}
Concluding remarks

- Reviewed all statistical tools necessary to search for the Higgs boson signal (as a small signal above a known background):

- Limits: usual way to 'challenge' signal hypothesis \( (H_1) \).

- \( p \)-values: usual way to 'challenge' 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.

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