Introduction to Statistical Analysis

Nicolas Berger (LAPP Annecy)
Physics measurement data are produced through random processes. Need to be described using a statistical model:

<table>
<thead>
<tr>
<th>Description</th>
<th>Observable</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Counting</strong></td>
<td>( n )</td>
<td>Poisson</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( P(n; S, B) = e^{-(S + B)} \frac{(S + B)^n}{n!} )</td>
</tr>
<tr>
<td><strong>Binned shape analysis</strong></td>
<td>( n_i, i=1..N_{\text{bins}} )</td>
<td>Poisson product</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( P(n_i; S, B) = \prod_{i=1}^{n_{\text{bins}}} e^{-(S f_i^{\text{sig}} + B f_i^{\text{bkg}})} \frac{(S f_i^{\text{sig}} + B f_i^{\text{bkg}})^{n_i}}{n_i!} )</td>
</tr>
<tr>
<td><strong>Unbinned shape analysis</strong></td>
<td>( m_i, i=1..n_{\text{evts}} )</td>
<td>Extended Unbinned Likelihood</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( P(m_i; S, B) = \frac{e^{-(S + B)}}{n_{\text{evts}}!} \prod_{i=1}^{n_{\text{evts}}} S P_{\text{sig}}(m_i) + B P_{\text{bkg}}(m_i) )</td>
</tr>
</tbody>
</table>

Model can include multiple categories, each with a separate description. Includes parameters of interest (POIs) but also nuisance parameters (NPs)
Reminders from Lecture 2: Discovery Significance

Given a statistical model $P(\text{data}; \mu)$, define likelihood $L(\mu) = P(\text{data}; \mu)$

**To estimate a parameter**, use the value $\hat{\mu}$ that maximizes $L(\mu)$.

**To decide between hypotheses** $H_0$ and $H_1$, use the **likelihood ratio** $\frac{L(H_0)}{L(H_1)}$

**To test for discovery**, use $q_0 = -2 \log \frac{L(S=0)}{L(\hat{S})}$ for $\hat{S} \geq 0$

For large enough datasets ($n > 5$), $Z = \sqrt{q_0}$

For a **Gaussian** measurement, $Z = \frac{\hat{S}}{\sqrt{B}}$

For a **Poisson** measurement, $Z = \sqrt{2 \left[ (\hat{S} + B) \log \left( 1 + \frac{\hat{S}}{B} \right) - \hat{S} \right]}$
Reminders from Lecture 2: Limits & Intervals

**Limits**: use LR-based test statistic:

\[ q_{S_0} = -2 \log \frac{L(S=S_0)}{L(\hat{S})} \quad \hat{S} \leq S_0 \]

→ Use **CL$_s$ procedure** to avoid negative limits

**Poisson regime**, \( n=0 \): \( S_{\text{up}} = 3 \) events

**Confidence intervals**: use \( t_{\mu_0} = -2 \log \frac{L(\mu=\mu_0)}{L(\hat{\mu})} \)

→ 1D: crossings with \( t_{\mu_0} = Z^2 \) for \( \pm Z\sigma \) intervals

**Gaussian regime**: \( \mu = \hat{\mu} \pm \sigma_{\text{Gauss}} \) for a 1\( \sigma \) interval
Outline

Expected results and toys
  Pseudo-experiments and Asimov datasets
    Dealing with non-asymptotic situations

Profiling

Look-Elsewhere Effect

Bayesian methods

Presentation of results
Generating Pseudo-data

Model describes the distribution of the observable: \( P(\text{data}; \text{parameters}) \)

\( \Rightarrow \) Possible outcomes of the experiment, for given parameter values

Can draw random events according to PDF: \textit{generate pseudo-data}

\[ P(\lambda = 5) \]

\( 2, 5, 3, 7, 4, 9, \ldots \)

Each entry = separate “experiment”
Expected Results

ATLAS
\( \sqrt{s} = 13 \text{ TeV}, 36.7 \text{ fb}^{-1} \)
Spin-0 Selection
NWA (\( \Gamma_X = 4 \text{ MeV} \))

95% CL Upper Limit on \( \sigma_{\text{fid}} \times B \) [fb]

\[
\text{Observed } \text{CL}_S \text{ limit} \\
\text{Expected } \text{CL}_S \text{ limit} \\
\text{Expected } \pm 1\sigma \\
\text{Expected } \pm 2\sigma
\]
Expected Limits: Toys

**Expected results**: median outcome under a given hypothesis
→ usually B-only for searches, but other choices possible.

Two main ways to compute:
→ **Pseudo-experiments (toys):**
  - Generate a pseudo-dataset in B-only hypothesis
  - Compute limit
  - Repeat and histogram the results
  - Central value = median, bands based on quantiles

68% of toys

95% of toys

*Computation example:

**Diagram:**
- Observed CL\textsubscript{s} limit
- Expected CL\textsubscript{s} limit
- Expected ± 1σ
- Expected ± 2σ

*ATLAS*
\( \sqrt{s} = 13 \text{ TeV}, 36.7 \text{ fb}^{-1} \)
Spin-0 Selection
NWA (\( \Gamma_X = 4 \text{ MeV} \))

**Graph:**
- Number Limit on \( \sigma_{\text{fid}} \times B \) [fb]
- Number of Toys

**Legend:**
- Black line: Observed CL\textsubscript{s} limit
- Dotted blue line: Expected CL\textsubscript{s} limit
- Green area: Expected ± 1σ
- Yellow area: Expected ± 2σ

Repeat for each mass

Expected Limits: Asimov Datasets

**Expected results**: median outcome under a given hypothesis
→ usually B-only by convention, but other choices possible.

Two main ways to compute:

→ **Asimov Datasets**
  - Generate a “perfect dataset” – *e.g.* for binned data, set bin contents carefully, no fluctuations.
  - Gives the median result immediately:
    \[
    \text{median(toy results)} \leftrightarrow \text{result(median dataset)}
    \]
  - Get bands from asymptotic formulas:
    Band width
    \[
    \sigma_{S_0, A}^2 = \frac{S_0^2}{q_{S_0}(\text{Asimov})}
    \]

- Much faster (1 “toy”)
- Relies on Gaussian approximation

Strictly speaking, Asimov dataset if
\[
\hat{X} = X_0 \text{ for all parameters } X,
\]
where \(X_0\) is the generation value.
Outline

Expected results and toys

Pseudo-experiments and Asimov datasets
Dealing with non-asymptotic situations

Profiling

Look-Elsewhere Effect

Bayesian methods

Presentation of results
Beyond Asymptotics: Toys

Asymptotics usually work well, but break down in some cases – e.g. small event counts.

**Solution**: generate *pseudo data* (toys) using the PDF, under the tested hypothesis

→ Also randomize the observable \( \theta^{\text{obs}} \) of each auxiliary experiment:

\[ G(\theta^{\text{obs}}; \theta, \sigma_{\text{syst}}) \]

→ Samples the true distribution of the PLR

→ Integrate above observed PLR to get the p-value

→ Precision limited by number of generated toys,

**Small p-values** (5σ : \( p \approx 10^{-7} \)) ⇔ **large toy samples**
Toys: Example

ATLAS $X \rightarrow Z\gamma$ Search: covers $200 \text{ GeV} < m_X < 2.5 \text{ TeV}$
→ for $m_X > 1.6 \text{ TeV}$, low event counts ⇒ derive results from toys

Asymptotic results (in gray) give optimistic result compared to toys (in blue)
Historical Aside
Classic Discoveries (1)

ψ Discovery

In this graph, the blue data points show a sharp peak in the number of hadrons produced at a narrow range of energies—evidence of the J/ψ particle. The horizontal axis shows the energy of one of the pair of SPEAR beams, measured in GeV. The height of the peak is so great that, to fit the plot on one sheet of graph paper, the vertical axis is compressed into a logarithmic scale.

Huge signal
S/B~50
Several 1000 events

Z⁰ Discovery

(almost) no background

Logbook of J. Rohlf, 1983-05-30
Classic Discoveries (2)

ψ' : discovered in the control room by the (lucky) shifters

First hints of top at D0:
O(10) signal events,
a few bkg events, 2.4σ
And now?

**Short answer:** The high-signal, low-background experiments have been done already (although a surprise would be welcome...)

e.g. at LHC:

- **High background levels**, need precise modeling
- **Large systematics**, need to be described accurately
- **Small signals**: need optimal use of available information:
  - **Shape analyses** instead of counting
  - **Categories** to isolated signal-enriched regions
Outline

Expected results and toys
Pseudo-experiments and Asimov datasets
Dealing with non-asymptotic situations

Profiling

Look-Elsewhere Effect

Bayesian methods

Presentation of results
Reminder: Wilks’ Theorem

Consider \( t_{S_0} = -2 \log \frac{L(S=S_0)}{L(\hat{S})} \)

→ Assume Gaussian regime (e.g. large \( n_{\text{evts}} \), Central-limit theorem) : then:

**Wilk’s Theorem:** \( t_{S_0} \) is distributed as a \( \chi^2 \)
under \( H_{S_0}(S=S_0) \):

\[
f(t_{S_0} \mid S=S_0) = f_{\chi^2(n_{\text{dof}}=1)}(t_{S_0})
\]

⇒ The significance is:

\[
Z = \sqrt{q_0}
\]
**Profiling**

How to deal with nuisance parameters in likelihood ratios?

→ Let the data choose ⇒ use the best-fit values *(Profiling)*

⇒ Profile Likelihood Ratio (PLR)

\[
 t_{s_0} = -2 \log \frac{L(S=S_0, \hat{\theta}(S_0))}{L(\hat{S}, \hat{\theta})}
\]

\(\hat{\theta}(s_0)\) best-fit value for \(S=S_0\) (conditional MLE)

\(\hat{\theta}\) overall best-fit value (unconditional MLE)

**Wilks’ Theorem**: same properties as plain likelihood ratio

\[
f(t_{s_0} \mid S=S_0) = f_{\chi^2(n_{dof}=1)}(t_{s_0})\]

⇒ Profiling “builds in” the effect of the NPs

⇒ Can use \(t_{s_0}\) to compute limits, significance, etc. in the same way as before
Homework 7: Gaussian Profiling

Counting experiment with background uncertainty: \( n = S + B \):

→ **Signal region (SR):** \( n_{\text{obs}} \sim G(S + B, \sigma_{\text{stat}}) \)

→ **Control region (CR):** \( B_{\text{obs}} \sim G(B, \sigma_{\text{bkg}}) \)

Recall: Signal region only (fixed B):

\[
\begin{align*}
S &= (n_{\text{obs}} - B) \pm \sigma_{\text{stat}} \\
\tau_S &= \left( \frac{S - n_{\text{obs}}}{\sigma_{\text{stat}}} \right)^2
\end{align*}
\]

→ Compute the best-fit (MLEs) for S and B

→ Show that the conditional MLE for B is

\[
\hat{B}(S) = B_{\text{obs}} + \frac{\sigma_{\text{bkg}}^2}{\sigma_{\text{stat}}^2 + \sigma_{\text{bkg}}^2} (\hat{S} - S)
\]

→ Compute the profile likelihood \( \tau_S \)

→ Compute the 1σ confidence interval on S

\[
S = (n_{\text{obs}} - B_{\text{obs}}) \pm \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{bkg}}^2} \\
\sigma_S = \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{bkg}}^2}
\]

Stat uncertainty (on n) and systematic (on B) add in quadrature
Systematics Implementation

Prototype: NP measured in a separate auxiliary experiment
- e.g. luminosity measurement

→ Build the combined likelihood of the main+auxiliary measurements

\[ L(\mu, \theta; \text{data}) = L_{\text{main}}(\mu, \theta; \text{main data}) \cdot L_{\text{aux}}(\theta; \text{aux. data}) \]

Gaussian form often used by default: \[ L_{\text{aux}}(\theta; \text{aux. data}) = G(\theta^{\text{obs}}; \theta, \sigma_{\text{syst}}) \]

→ Often no clear setup for auxiliary measurements
- e.g. theory uncertainties on missing HO terms from scale variations
  → Implemented in the same way nevertheless (\text{“pseudo-measurement”})
Uncertainty decomposition

All systematics NPs excluded: statistical uncertainty only

All systematics NPs included: stat+syst uncertainties

\[ \sigma_{\text{syst, tot}} = \sqrt{\sigma_{\text{total}}^2 - \sigma_{\text{stat}}^2} \]

\[
\mu = 0.99 \pm 0.12 \text{ (stat)} \pm 0.06 \text{ (syst)} \pm 0.06 \text{ (theo)}
\]
Profiling Example: $ttH \rightarrow bb$

Analysis uses low-S/B categories to constrain backgrounds.

→ **Reduction in large uncertainties on $tt$ bkg**

→ **Propagates to the high-S/B categories** through the statistical modeling

⇒ **Care needed in the propagation** (e.g. different kinematic regimes)
Profiling Takeaways

When testing a hypothesis, use the best-fit values of the nuisance parameters: Profile Likelihood Ratio.

\[ \frac{L(\mu = \mu_0, \theta_{\mu_0})}{L(\hat{\mu}, \hat{\theta})} \]

Allows to include systematics as uncertainties on nuisance parameters.

Profiling systematics includes their effect into the total uncertainty. Gaussian:

\[ \sigma_{\text{total}} = \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2} \]

Guaranteed to work well as long as everything is Gaussian, but typically also robust against non-Gaussian behavior.

Profiling can have unintended effects – need to carefully check behavior
Outline

Expected results and toys
  Pseudo-experiments and Asimov datasets
  Dealing with non-asymptotic situations

Profiling

Look-Elsewhere Effect

Bayesian methods

Presentation of results
Look-Elsewhere effect

Sometimes, unknown parameters in signal model e.g. p-values as a function of $m_X$

$\Rightarrow$ Effectively: **multiple, simultaneous searches**

$\Rightarrow$ If e.g. small resolution and large scan range, many independent experiments

$\Rightarrow$ More likely to find an excess anywhere in the range, rather than in a predefined location

$\Rightarrow$ **Look-elsewhere effect** (LEE)
Global Significance

Probability for a fluctuation \textbf{anywhere} in the range $\rightarrow$ \textbf{Global p-value}.

at a given location $\rightarrow$ \textbf{Local p-value}

For searches over a parameter range, \textbf{the global p-value is the relevant one} $\rightarrow$ Accounts for the actual search procedure: look for an excess anywhere in the scanned range

$\rightarrow$ Depends on the scanned parameter ranges

\textbf{e.g. } $X \rightarrow \gamma\gamma$

- $200 < m_X < 2000 \text{ GeV}$
- $0 < \Gamma_X < 10\% \, m_X$.

$\rightarrow \, p_{\text{local}}$ is what comes out of the usual formulas

\textbf{How to compute } $p_{\text{global}}$ (or $N_{\text{trials}}$) ?
Global Significance from Toys

**Principle**: repeat the analysis in toy data:
→ generate pseudo-dataset
→ perform the search, scanning over parameters as in the data
→ report the largest significance found
→ repeat many times

⇒ The frequency at which a given $Z_0$ is found is the global p-value

e.g. $X \rightarrow \gamma\gamma$ Search: $Z_{\text{local}} = 3.9\sigma \Rightarrow p_{\text{local}} \sim 5 \times 10^{-5}$,
→ However we are scanning $200 < m_X < 2000$ GeV and $0 < \Gamma_X < 10\% m_X$!
→ Toys: find such an excess 2% of the time somewhere in the range
⇒ $p_{\text{global}} \sim 2 \times 10^{-2}$, $Z_{\text{global}} = 2.1\sigma$ Less exciting, and better indication of true Z!

⊕ **Exact treatment**
⊕ **CPU-intensive** especially for large Z (need $\sim O(100)/p_{\text{global}}$ toys)
**Trials Factor**

**Trials factor** $N = \# \text{ of independent searches:}$

\[
p_{\text{global}} = 1 - (1 - p_{\text{local}})^N \approx N \ p_{\text{local}}
\]

Naively, one could expect

\[
N_{\text{trials}} = N_{\text{indep}} = \frac{\text{scan range}}{\text{peak width}}
\]

However this is usually **wrong**!
Trials Factor from Asymptotics

Asymptotic limit: trials factor \((1 \text{ POI})\) is

\[
N_{\text{trials}} = 1 + \sqrt{\frac{\pi}{2}} \frac{N_{\text{indep}}}{Z_{\text{local}}}
\]

\[N_{\text{indep}} = \frac{\text{scan range}}{\text{peak width}}\]

→ Trials factor is not just \(N_{\text{indep}}\), also depends on \(Z_{\text{local}}\)!

**Why?** Slicing range into \(N_{\text{indep}}\) regions misses peaks sitting on edges between regions → true \(N_{\text{trials}}\) is > \(N_{\text{indep}}\)!

**Search everywhere:**

\[N_{\text{trials}} = 1 + \sqrt{\frac{\pi}{2}} N_{\text{indep}} Z_{\text{local}}\]

**Toy data**

Search in 10 fixed bins: \(N_{\text{trials}} = 10\)
Asymptotic limit: trials factor (1 POI) is

\[ N_{\text{trials}} = 1 + \sqrt{\frac{\pi}{2}} N_{\text{indep}} Z_{\text{local}} \]

→ Trials factor is not just \( N_{\text{indep}} \), also depends on \( Z_{\text{local}} \)!

Why? Slicing range into \( N_{\text{indep}} \) regions misses peaks sitting on edges between regions

\( \Rightarrow \) true \( N_{\text{trials}} \) is \( > N_{\text{indep}} \)!

Search in 10 fixed bins: \( N_{\text{trials}} = 10 \)

Toy data
Outline

Expected results and toys
  Pseudo-experiments and Asimov datasets
  Dealing with non-asymptotic situations

Profiling

Look-Elsewhere Effect

**Bayesian methods**

Presentation of results
Frequentist vs. Bayesian

All methods described so far are frequentist
• Measurement outcomes are random
• Parameters value are fixed but unknown

Must be careful about meaning:

→ “5σ Higgs discovery”
• → if there is really no Higgs, such fluctuations are observed in only one in 3 million experiments: \[ P(\text{data} \mid \text{no Higgs}) \text{ is small} \]

This is not the crucial question! What we would really like to know is
\[ \text{What is the probability that the excess we see is a fluctuation} \]

→ we want \[ P(\text{no Higgs} \mid \text{data}) \] – but all we have is \[ P(\text{data} \mid \text{no Higgs}) \]
However \[ P(\text{no Higgs} \mid \text{data}) \] is not well-defined in the frequentist framework
Can use **Bayes’ theorem** to address this:

\[
P(\text{no Higgs}|\text{data}) = \frac{P(\text{data}|\text{no Higgs})}{P(\text{data})} P(\text{no Higgs})
\]

→ An hypothesis ("no Higgs") is now considered something random
  - Is the presence of the Higgs in an experiment randomly chosen?
  - In fact, different definition of \( p \): **degree of belief**, not from frequencies.
  - \( P(\text{no Higgs}) \) **Prior degree of belief** – critical ingredient in the computation

Compared to frequentist PLR:

- **answers the “right” question**
- **answer depends on the prior**
- **In practice, frequentist and Bayesian methods usually give similar results**

“Bayesians address the questions everyone is interested in by using assumptions that no one believes. Frequentist use impeccable logic to deal with an issue that is of no interest to anyone.” - Louis Lyons
Bayesian methods

**Probability distribution** (= likelihood) :
→ Same as frequentist case, but treat systematics by **integrating over priors**, instead of profiling:

→ Integrate out θ to get P(μ) :

\[ P(\mu) = \int P(\mu, \theta) C(\theta) \, d\theta \]

→ Use probability distribution P(μ) directly for limits & intervals

e.g. define 68% CL (“Credibility Level”) interval (A, B) by:

\[ \int_{A}^{B} P(\mu) \, d\mu = 68 \% \]

⊖ No simple way to test for discovery
⊖ Integration over NPs can be CPU-intensive (but can use MCMC methods)

**Priors** : most analyses use flat priors in the analysis variable(s)
→ **Parameterization-dependent**: if flat in σ×B, then not flat in couplings....
→ Can use the Jeffreys’ or reference priors, but difficult in practice
Homework 8: Bayesian methods and CLs

Gaussian counting problem with systematic on background: \( n = S + B + \sigma_{\text{syst}} \theta \)

\[
P(n; S, \theta) = G(n; S + B + \sigma_{\text{syst}} \theta, \sigma_{\text{stat}}) \cdot G(\theta_{\text{obs}} = 0; \theta, 1)
\]

→ What is the 95% CL upper limit on \( S \), given a measurement \( n_{\text{obs}} \)?

1. CLs computation:
   - Use the result of Homework 7 to compute the PLR for \( S \)
   - Use the result of Homework 6 to compute the CLs upper limit

2. Bayesian computation:
   - Integrate \( P(n; S, \theta) \) over \( \theta \) to get the marginalized \( P(n | S) \)
   - Use Bayes’ theorem to compute \( P(S | n) \propto P(n | S) \cdot P(S) \), with \( P(S) \) a constant prior over \( S > 0 \).
   - Find the 95% CL limit by solving \( \int_{S_{\text{up}}}^{\infty} P(S | n) \, dS = 5\% \)

Solution:

\[
S_{\text{up}}^{\text{CLs}} = n - B + \Phi^{-1}\left(1 - 0.05 \cdot \Phi\left(\frac{n - B}{\sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}}\right)\right) \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}
\]

In both cases
Example: $W' \rightarrow l\nu$ Search

- **POI**: $W' \sigma \times B \rightarrow$ use flat prior over $[0, +\infty]$.  
- **NPs**: syst on signal $\varepsilon$ (6 NPs), bkg (6), lumi (1) $\rightarrow$ integrate over Gaussian priors

---

**Table 1**

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger</td>
<td></td>
</tr>
<tr>
<td>Lepton reconstruction</td>
<td></td>
</tr>
<tr>
<td>and identification</td>
<td></td>
</tr>
<tr>
<td>Lepton momentum</td>
<td></td>
</tr>
<tr>
<td>scale and resolution</td>
<td></td>
</tr>
<tr>
<td>$E_T^{miss}$ resolution and scale</td>
<td></td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td></td>
</tr>
<tr>
<td>Pile-up</td>
<td></td>
</tr>
<tr>
<td>Multijet background</td>
<td></td>
</tr>
<tr>
<td>Top extrapolation</td>
<td></td>
</tr>
<tr>
<td>Diboson extrapolation</td>
<td></td>
</tr>
<tr>
<td>PDF choice for DY</td>
<td></td>
</tr>
<tr>
<td>PDF variation for DY</td>
<td></td>
</tr>
<tr>
<td>EW corrections for DY</td>
<td></td>
</tr>
<tr>
<td>Luminosity</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 1**

- **ATLAS** 
  $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  $W' \rightarrow e\nu$ selection

- **Graph 1**
  $\sigma(pp \rightarrow W') \times \text{BR}(W' \rightarrow l\nu) [\text{pb}]$
  Expected limit
  $W'_{\text{SSM}}$
  $m_{W'}$ [TeV]
  95% CL

---

**Graph 2**

- **ATLAS** 
  $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  $W' \rightarrow e\nu$
  $95\%$ CL

---

arXiv:1706.04786
Why $5\sigma$?

One-sided discovery: $5\sigma \Leftrightarrow p_0 = 3 \times 10^{-7} \Leftrightarrow 1\text{ chance in } 3.5\text{M}

→ Overly conservative?
→ Do we even control such small probabilities?

Reasons for sticking with $5\sigma$ (from Louis Lyons):

- **LEE**: searches typically cover multiple independent regions
  ⇒ Global p-value is the relevant one
  $N_{\text{trials}} \sim 1000$ : local $5\sigma \Leftrightarrow O(10^{-4})$ more reasonable

- **Mismodeled systematics**: factor 2 error in syst-dominated analysis ⇒ factor 2 error on $Z$...

- **History**: $3\sigma$ and $4\sigma$ excesses do occur regularly, for the reasons above

- **“Subconscious Bayes Factor”**: p-value should be at least as small as the subjective $p(S)$:

  \[ \text{Extraordinary claims require extraordinary evidence} \]
  ⇒ Stay with $5\sigma$...
Expected results and toys
   Pseudo-experiments and Asimov datasets
   Dealing with non-asymptotic situations

Profiling

Look-Elsewhere Effect

Bayesian methods

Presentation of results
Reparameterization

Start with basic measurement in terms of e.g. $\sigma \times B$

→ How to measure derived quantities (couplings, parameters in some theory model, etc.)? → just reparameterize the likelihood:

e.g. Higgs couplings: $\sigma_{ggF}, \sigma_{VBF}$ sensitive to Higgs coupling modifiers $\kappa_V, \kappa_F$.

$$L(\sigma_{ggF}, \sigma_{VBF}) \rightarrow L(\sigma_{ggF}(\kappa_V, \kappa_F), \sigma_{VBF}(\kappa_V, \kappa_F)) \equiv L'(\kappa_V, \kappa_F)$$
Reparameterization: Limits

CMS Run 2 Monophoton Search: measured $N_s$ in a counting experiment reparameterized according to various DM models.
Presentation of Results

→ Cannot test every model: need to make enough information public so that others (theorists) are able to do it independently

⇒ **Gaussian case**: sufficient to provide measurements + covariance matrix

→ For example using the HEPData repository.

**Non-Gaussian case**: no simple method
Conclusion

• Significant evolution in the statistical methods used in HEP

• Variety of methods, adapted to various situations and target results

• Allow to
  – model the statistical process with high precision in difficult situations (large systematics, small signals)
  – make optimal use of available information

• Implemented in standard RooFit/RooStat toolkits within the ROOT framework, as well as other tools (BAT)

• Still many open questions and areas that could use improvement → e.g. how to present results with all available information
Homework solutions for Lecture 2
Homework 1: Gaussian Counting

Count number of events $n$ in data
→ assume $n$ large enough so process is Gaussian
→ assume $B$ is known, measure $S$

$$L(S; n) = e^{-\frac{1}{2}(\frac{n-(S+B)}{\sqrt{S+B}})^2}$$

Likelihood :
$$\lambda(S; n) = \left(\frac{n-(S+B)}{\sqrt{S+B}}\right)^2$$

MLE for $S$ : $\hat{S} = n - B$

Test statistic: assume $\hat{S} > 0$, 

$$q_0 = -2 \log \frac{L(S=0)}{L(\hat{S})} = \lambda(S=0) - \lambda(\hat{S}) = \left(\frac{n-B}{\sqrt{B}}\right)^2 = \left(\frac{\hat{S}}{\sqrt{B}}\right)^2$$

Finally:

$$Z = \sqrt{q_0} = \frac{\hat{S}}{\sqrt{B}}$$

Known formula!
→ Strictly speaking only valid in Gaussian regime
Homework 2: Poisson Counting

Same problem but now not assuming Gaussian behavior:

\[ L(S; n) = e^{-(S+B)}(S+B)^n \quad \lambda(S; n) = 2(S+B) - 2n \log(S+B) \]

MLE: \( \hat{S} = n - B \), same as Gaussian

**Test statistic** (for \( \hat{S} > 0 \)):

\[ q_0 = \lambda(S=0) - \lambda(\hat{S}) = -2\hat{S} - 2(\hat{S}+B) \log \frac{B}{\hat{S}+B} \]

Assuming asymptotic distribution for \( q_0 \),

\[ Z = \sqrt{2 \left[ (\hat{S}+B) \log \left( 1 + \frac{\hat{S}}{B} \right) - \hat{S} \right]} \]

See G. Cowan’s slides for case with B uncertainty
Homework 3: Gaussian CL$_{s+b}$

Usual Gaussian counting example with known $B$:

$$\lambda (S) = \left( \frac{n-(S+B)}{\sigma_s} \right)^2$$

Reminder:
Best fit signal : $\hat{S} = n - B$
Significance: $Z = \frac{\hat{S}}{\sqrt{B}}$

Compute the 95% CL upper limit on $S$:

$$q_{S_0} = -2 \log \frac{L(S=S_0)}{L(\hat{S})} = \lambda (S_0) - \lambda (\hat{S}) = \left( \frac{n-(S_0+B)}{\sigma_s} \right)^2 = \left( \frac{S_0-\hat{S}}{\sigma_s} \right)^2$$

for $S_0 > \hat{S}$

so $q_{S_0} = 2.70$ for $S_0 = \hat{S} + \sqrt{2.70} \sigma_s$

And finally $S_{up} = \hat{S} + 1.64 \sigma_s$ at 95 % CL
Homework 4: Gaussian $\text{CL}_s$

Usual Gaussian counting example with known $B$:

$$\lambda(S) = \left( \frac{n - (S+B)}{\sigma_S} \right)^2$$

**Reminder**

Best fit signal: $\hat{S} = n - B$

$\text{CL}_{s+b}$ limit:

$$S_{\text{up}} = \hat{S} + 1.64 \sigma_S \text{ at } 95 \% \text{ CL}$$

$\text{CL}_s$ upper limit: still have $q_{S_0} = \left( \frac{S_0 - \hat{S}}{\sigma_S} \right)^2$ (for $S_0 > \hat{S}$)

so need to solve

$$p_{CL_s} = \frac{p_{S_0}}{p_B} = \frac{1 - \Phi(\sqrt{q_{S_0}})}{1 - \Phi(\sqrt{q_{S_0}} - S_0/\sigma_S)} = 5 \%$$

for $\hat{S} = 0$,

$$S_{\text{up}} = \hat{S} + \left[ \Phi^{-1} \left( 1 - 0.05 \phi \left( \frac{\hat{S}}{\sigma_S} \right) \right) \right] \sigma_S \text{ at } 95 \% \text{ CL}$$

$\hat{S} \sim G(S, \sigma_S)$ so

**Under $H_0(S = S_0)$:**

$$\sqrt{q_{S_0}} \sim G(0, 1)$$

$$p_{S_0} = 1 - \Phi(\sqrt{q_{S_0}})$$

**Under $H_0(S = 0)$:**

$$\sqrt{q_{S_0}} \sim G(S_0/\sigma_S, 1)$$

$$p_B = 1 - \Phi(\sqrt{q_{S_0}} - S_0/\sigma_S)$$
Homework 5: Poisson $\text{CL}_s$

Same exercise, for the Poisson case

**Exact computation**: sum probabilities of cases “at least as extreme as data” ($n$)

$$p_{S_0}(n) = \sum_{k=0}^{n} e^{-(S_0+B)} \frac{(S_0+B)^k}{k!}$$

and one should solve $p_{CL_s} = \frac{p_{S_{up}}(n)}{p_0(n)} = 5\%$ for $S_{up}$

For $n = 0$:

$$p_{CL_s} = \frac{p_{S_{up}}(0)}{p_0(0)} = e^{-S_{up}} = 5\% \Rightarrow S_{up} = \log(20) = 2.996 \approx 3$$

⇒ **Rule of thumb**: when $n_{obs} = 0$, the 95% $\text{CL}_s$ limit is 3 events (for any $B$)

**Asymptotics**: as before,

$$q_{S_0} = \lambda(S_0) - \lambda(\hat{S}) = 2(S_0 + B - n) - 2n \log \frac{S_0 + B}{n}$$

For $n = 0$,

$$q_{S_0}(n=0) = 2(S_0 + B)$$

$$p_{CL_s} = \frac{p_{S_0}}{p_0} = \frac{1 - \Phi(\sqrt{q_{S_0}(n=0)})}{1 - \Phi(\sqrt{q_{S_0}(n=0)} - \sqrt{q_{S_0}(n=B)})} = 5\%$$

⇒ $S_{up} \sim 2$, exact value depends on $B$

⇒ **Asymptotics not valid in this case** ($n=0$) – need to use exact results, or toys
Consider a parameter $m$ (e.g. Higgs boson mass) whose measurement is Gaussian with known width $\sigma_m$, and we measure $m_{\text{obs}}$:

$$\lambda(m; m_{\text{obs}}) = \left( \frac{m - m_{\text{obs}}}{\sigma_m} \right)^2$$

→ Best-fit value (MLE): $\hat{m} = m_{\text{obs}}$.

→ Test statistic: $t_m = \left( \frac{m - m_{\text{obs}}}{\sigma_m} \right)^2$

→ $1\sigma$ Interval $m = m_{\text{obs}} \pm \sigma_m$
Homework solutions for Lecture 3
Homework 7: Gaussian Profiling

Counting experiment with background uncertainty: \( n = S + \theta \):
→ Signal region: \( n \sim \mathcal{G}(S + \theta, \sigma_{\text{stat}}) \)
→ Control region: \( \theta^{\text{obs}} \sim \mathcal{G}(\theta, \sigma_{\text{syst}}) \)

Then:

\[
\lambda(S, \theta) = \left( \frac{n - (S + \theta)}{\sigma_{\text{stat}}} \right)^2 + \left( \frac{\theta^{\text{obs}} - \theta}{\sigma_{\text{syst}}} \right)^2
\]

For \( S = \hat{S} \), matches MLE as it should

MLEs:
\[
\hat{S} = n - \theta^{\text{obs}} \quad \quad \hat{\theta} = \theta^{\text{obs}}
\]

Conditional MLE:
\[
\hat{\theta}(S) = \theta^{\text{obs}} + \frac{\sigma_{\text{syst}}^2}{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2} (\hat{S} - S)
\]

PLR:
\[
s = -2 \log \frac{L(S, \hat{\theta}(S))}{L(\hat{S}, \hat{\theta})} = \lambda(S, \hat{\theta}(S)) - \lambda(\hat{S}, \hat{\theta}) = \frac{(S - \hat{S})^2}{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}
\]

1σ interval
\[
S = \hat{S} \pm \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2} \quad \quad \sigma_S = \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}
\]

Stat uncertainty (on \( n \)) and systematic (on \( \theta \)) add in quadrature
Homework 8: CL$_s$ computation

Gaussian counting with systematic on background: $n = S + B + \sigma_{syst} \theta$

$L(n; S, \theta) = G(n; S + B + \sigma_{syst} \theta, \sigma_{stat}) \ G(\theta_{obs} = 0; \theta, 1)$

**MLE:** $\hat{S} = n - B$

**Conditional MLE:** $\hat{\theta}(\mu) = \frac{\sigma_{syst}}{\sigma_{stat}^2 + \sigma_{syst}^2}(n - S - B)$

**PLR:** $\lambda(\mu) = \left(\frac{S + B - n}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)^2$

This boils down to the Gaussian case of HW 6, so the CL$_s$ limit is

$$\text{CL}_s: \quad S^{\text{CL}_s}_{\text{up}} = n - B + \left[ \Phi^{-1}(1 - 0.05) \Phi\left(\frac{n - B}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right) \right] \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$
Homework 8: Bayesian computation

Gaussian counting with systematic on background: \( n = S + B + \sigma_{\text{syst}} \theta \)

\[ P(n \mid S, \theta) = G(n; S+B+\sigma_{\text{syst}} \theta, \sigma_{\text{stat}}) \ G(\theta \mid 0, 1) \]

Bayesian: \( G(\theta) \) is actually a prior on \( \theta \) ⇒ perform integral (marginalization)

\[ P(n \mid S) = G(S; n-B, \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}) \]

same effect as profiling!

Need \( P(S \mid n) \) ⇒ a prior for \( S \) – take flat PDF over \( S > 0 \)
⇒ Truncate Gaussian at \( S=0 \): \( P(S \mid n) = P(n \mid S) \ P(S) \)

\[ P(S \mid n) = G(S; n-B, \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}) \left[ \Phi \left( \frac{n-B}{\sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}} \right) \right]^{-1} \]

Bayesian Limit:

\[
\int_{S_{\text{up}}}^{\infty} P(S \mid n) \, dS = 5\% = \left[ 1 - \Phi \left( \frac{S_{\text{up}} - (n-B)}{\sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}} \right) \right] \left[ \Phi \left( \frac{n-B}{\sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}} \right) \right]^{-1}
\]

\[ S_{\text{up}}^{\text{Bayes}} = n-B + \left[ \Phi^{-1} \left( 1 - 0.05 \right) \Phi \left( \frac{n-B}{\sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}} \right) \right] \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2} \]

same result as \( \text{CL}_s \)!
Extra Slides
Illustrative Example

Test on a simple example: generate toys with
→ flat background (100 events/bin)
→ count events in a fixed-size sliding window, look for excesses

Two configurations:
1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum
Illustrative Example (2)

Very different results if the excess is near a boundary:

1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum
Illustrative Example (3)

Search in predefined bins: \( N_{\text{trials}} = 10 \)

Search everywhere gives the extra \( Z_{\text{local}} \) dependence
**Z_{Global} Asymptotics Extrapolation**

**Asymptotic trials factor** (1 POI): \( N_{\text{trials}} = 1 + \sqrt{\frac{\pi}{2}} N_{\text{indep}} Z_{\text{local}} \)

How to get \( N_{\text{indep}} \)? Usually work with a slightly different formula:

\[
N_{\text{trials}} = 1 + \frac{1}{p_{\text{local}}} \left\langle N_{\text{up}}(Z_{\text{test}}) \right\rangle e^{\frac{Z_{\text{test}}^2 - Z_{\text{local}}^2}{2}}
\]

⇒ calibrate for small \( Z_{\text{test}} \), apply result to higher \( Z_{\text{local}} \).

Can choose arbitrarily small \( Z_{\text{test}} \)
⇒ many excesses
⇒ can measure \( N_{\text{up}} \) in data (1 “toy”)

Can also measure \( \langle N_{\text{up}} \rangle \) in multiple toys
if large stat uncertainty from too few excesses
In 2D

Generalization to 2D scans: consider sections at a fixed $Z_{\text{test}}$, compute its

**Euler characteristic $\phi$, and use**

$$p_{\text{global}} \approx E[\phi(A_u)] = p_{\text{local}} + e^{-u/2}(N_1 + \sqrt{u}N_2)$$

→ Generalizes 1D bump counting

Now need to determine 2 constants $N_1$ and $N_2$, from Euler $\phi$ measurements at 2 different $Z_{\text{test}}$ values.