Statistics In HEP 2

How do we understand/interpret our measurements
Outline

- Maximum Likelihood fit
  - strict frequentist Neyman – confidence intervals
    - what “bothers” people with them

- Feldmans/Cousins confidence belts/intervals

- Bayesian treatment of ‘unphysical’ results

- How the LEP-Higgs limit was derived

- what about systematic uncertainties?
  - Profile Likelihood
Parameter Estimation

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

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

$\Rightarrow$ try to maximize $L(\theta)$
$\Rightarrow$ typically:
  $\Rightarrow$ minimize - $2\log(L(\theta)) \Rightarrow \hat{\theta}$

$\Rightarrow$ Maximum Likelihood estimator
**Maximum Likelihood Estimator**

**Example:** \( \text{PDF}(x) = \text{Gauss}(x, \mu, \sigma) \) \( \rightarrow \) \( L(x | \text{Gauss}(\mu)) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x-\mu)^2}{2\sigma^2} \right) \)

\( \rightarrow \) estimator for \( \mu_{\text{true}} \) from the data measured in an experiment \( x_1, \ldots, x_K \)

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

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

![Graph of likelihood function](image)

For Gaussian PDFs:

\[ -2\ln(L(\mu)) = \chi^2 \]

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

properties of estimators

- biased or unbiased
- large or small variance
  → distribution of $\hat{\theta}$ on many measurements?

Small bias and small variance are typically “in conflict”
- Maximum Likelihood is typically unbiased only in the limit $K \rightarrow \infty$

- If Likelihood function is “Gaussian” (often the case for large $K$
  → central limit theorem)
  → get “error” estimate from or $-2\Delta \log(L) = 1$
  → If (very) none Gaussian
  → revert typically to (classical) Neyman confidence intervals

Glen Cowan:
another way to look at a measurement rigorously “frequentist”
- Neymans Confidence belt for CL $\alpha$ (e.g. 90%)

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

- measure $x_{\text{obs}}$ $\rightarrow$ conf. interval $= [\mu_1, \mu_2]$ given by vertical line intersecting the belt.

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

Feldman/Cousin (1998)
Classical Confidence Intervals

another way to look at a measurement rigorously “frequentist”

- Neyman’s Confidence belt for CL $\alpha$ (e.g. 90%)

$\rightarrow$ conf. interval $= [\mu_1, \mu_2]$ given by vertical line intersecting the belt.
- by construction:
  - $P(x < x_{\text{obs}}; \mu_2) = \frac{1-\alpha}{2}$
  - $P(x > x_{\text{obs}}; \mu_1) = \frac{1-\alpha}{2}$
- if the true value were $\mu_{\text{true}}$
  $\rightarrow$ lies in $[\mu_1, \mu_2]$ if it intersects $\mu_{\text{true}}$
  $\rightarrow x_{\text{meas}}$ intersects as in 90% (that’s how it was constructed)
  $\rightarrow$ only those $x_{\text{meas}}$ give $[\mu_1, \mu_2]$’s that intersect with the $\mu_{\text{true}}$
  $\rightarrow$ 90% of intervals cover $\mu_{\text{true}}$

- $P(x; \mu)$ is Gaussian ($\sigma = \text{const}$) $\rightarrow$ central 68% Neyman Conf. Intervals
  $\Leftrightarrow$ Max. Likelihood + its “error” estimate $[\hat{x} - \sigma_{\hat{x}}; \hat{x} + \sigma_{\hat{x}}]$
When to quote measurement or a limit!

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

\[ \text{decide: if } x_{\text{obs}} < 0 \text{ assume you } = 0 \]
  \[ \rightarrow \text{ conservative} \]

\[ \text{if you observe } x_{\text{obs}} < 3 \]
  \[ \rightarrow \text{ quote upper limit only} \]

\[ \text{if you observe } x_{\text{obs}} > 3 \]
  \[ \rightarrow \text{ quote a measurement} \]

\[ \rightarrow \text{ induces “undercovering” as this acceptance region contains only 85%!!} \]
Some things people don’t like..

same example:

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

- using proper confidence belt
- assume: $x_{obs} = -1.8$
  $\implies$ confidence interval is EMPTY!

Note: that’s OK from the frequentist interpretation

$\mu_{true} \in [\text{conf. interv.}]$ in 90% of (hypothetical) measurements.

Obviously we were ‘unlucky’ to pick one out of the remaining 10%

- nonetheless: tempted to “flip-flop” ??? tsz .. tsz.. tsz..
• How we determine the “acceptance” region for each $\mu_{\text{hyp.true}}$ is up to us as long as it covers the desired integral of size $\alpha$ (e.g. 90%)
→ include those “$x_{\text{meas.}}$” for which the large likelihood ratio first:

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

• $\mu_{\text{best estimate}}$ here: either the observation $x_{\text{meas}}$ or the closest ALLOWED $\mu$

No “empty intervals anymore!”
Being Lucky…

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

2 experiments (E1 and E2)

$\mu_{b1} = 1, \mu_{b2} = 2$

observing (out of luck) 0

E1: 95% limit on $\mu_s \sim 2$

E2: 95% limit on $\mu_s \sim 1$

$\Rightarrow$ UNFAIR! ?
Feldman/Cousins confidence belts

- motivated by “popular” ‘Bayesian’ approaches to handle such problems.

Bayesian: rather than constructing Confidence belts:

- turn Likelihood for $\mu_s$ (on given $n_{obs}$) into Posterior probability on $\mu_s$
- i.e. $\text{Poisson}(n_{obs}; \mu_s + \mu_b)$
- $p(\mu_s|n_{obs}) = L(n_{obs}; \mu_s) \times \pi(\mu_s)$

- $\pi(\mu_s) = \begin{cases} 0 & \mu_s < 0 \\ \text{uniform} & \mu_s > 0 \end{cases}$

- Feldman/Cousins
  - there is still SOME “unfairness”
  - perfectly “fine” in frequentist interpretation:
  - should quote “limit+sensitivity”
Statistical Tests in Particle Searches

**exclusion limits**

- upper limit on cross section
  \[ \text{lower limit on mass scale} \]
- \( \sigma < \text{limit as otherwise we would have seen it} \)

- need to estimate probability of downward fluctuation of \( s+b \)
- try to "disprove" \( H_0 = \text{background only} \)
- try to disprove \( H_0 = \text{background only} \)

**discoveries**

- need to estimate probability of upward fluctuation of \( b \)

better: find minimal \( s \), for which you can still exclude \( H_0 = s+b \) a pre-specified Confidence Level

**type 1 error**

\[ \alpha = 0.05\% \]
\[ \rightarrow 95\% CL \]

possible observation

\( N_{\text{obs}} \) for \( 5\sigma \) discovery
\[ P = 2.8 \times 10^{-7} \]
exclusion limit:
• test statistic does not necessarily have to be simply the counted number of events:
  → remember Neyman Pearson → Likelihood ratio

→ \( t(n_{\text{obs}}) = \frac{\text{Poisson}(n_{\text{obs}}; s,b)}{\text{Poisson}(n_{\text{obs}}; b)} \)
→ pre-specify \( \alpha = 0.05\% \) → 95% CL on exclusion
→ make your measurements
  → If accepted (i.e. in “critical (red) region”) where you decided to “reject” \( H_0 = s+b \)

→ \( CL_{s+b} = P(t < t_{\text{obs}}) \)
→ (i.e. what would have been the chance for THIS particular measurement to still have been “fooled” and there would have actually BEEN a signal)
Example: LEP-Higgs search

Remember: there were 4 experiments, many different search channels → treat different experiments just like “more channels”

\[ Q = \frac{L(x|H_1)}{L(x|H_0)} = \frac{\prod_{i}^{N_{\text{chan}}} \text{Pois}(n_i|s_i+b_i) \prod_{j}^{n_i} \frac{s_i f_s(x_{ij}) + b_i f_b(x_{ij})}{s_i+b_i}}{\prod_{i}^{N_{\text{chan}}} \text{Pois}(n_i|b_i) \prod_{j}^{n_i} f_b(x_{ij})} \]

- Evaluate how the \(-2\ln Q\) is distributed for
  - background only
  - signal + background
  - (note: needs to be done for all Higgs masses)

example: \( m_H = 115 \text{GeV}/c^2 \)
Example: LEP SM Higgs Limit
Example LEP Higgs Search

- In order to “avoid” the possible “problem” of Being Lucky when setting the limit
  → rather than “quoting” in addition the expected sensitivity
  → weight your CLs+b by it:

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

![Graph showing CL_s and CL_b distributions](image)
Systmatic Uncertainties

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

  ! Bayesian! (probability of the systematic parameter)

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

- LEP-Higgs: generated MC to get the PDFs with “varying” param.
  with systematic uncertainty
  → essentially the same as “integrating over” → need probability density for “how these parameters vary”
Systematic Uncertainties

- Why don’t we:
  - include any systematic uncertainty as “free parameter” in the fit

- eg. measure background contribution under signal peak in sidebands
- measurement + extrapolation into side bands have uncertainty
- but you can parametrise your expected background such that:
  → if sideband measurement gives this data → then b=...

Note: no need to specify prior probability

- Build your likelihood function such that it includes:
  - your parameters of interest
  - those describing the influence of the sys. uncertainty
  → nuisance parameters
Nuisance Parameters and Profile Likelihood

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

\[ l(m) = \frac{L(m, \hat{\theta})}{L(\hat{m}, \hat{\theta})}, \quad 0 \leq \hat{\mu} \leq \mu \]

\[ l(0) = \frac{L(0, \hat{\theta})}{L(\hat{m}, \hat{\theta})}, \quad \hat{\mu} < 0 \]

\[ \tilde{\lambda}(\mu) = \begin{cases} 
  \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})}, & 0 \leq \hat{\mu} \leq \mu \\
  \frac{L(\mu, \hat{\theta})}{L(0, \hat{\theta})}, & \hat{\mu} < 0 
\end{cases} \]

"ratio of likelihoods", why?
Why not simply using $L(\mu, \theta)$ as test statistics?

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

- Maximum Likelihood fit to estimate parameters
  
- what to do if estimator is non-gaussian:
  - Neyman – confidence intervals
  - what “bothers” people with them

- Feldmans/Cousins confidence belts/intervals
  - unifies “limit” or “measurement” confidence belts

- CLs … the HEP limit;
  - CLs … ratio of “p-values” … statisticians don’t like that
  - new idea: Power Constrained limits
    - rather than specifying “sensitivity” and “neyman conf. interval”
    - decide beforehand that you’ll “accept” limits only if the where your experiment has sufficient “power” i.e. “sensitivity” !

- .. a bit about Profile Likelihood, systematic error.