Likelihoods
1) Brief Introduction
2) Do’s & Don’t’s

Louis Lyons
Imperial College & Oxford
CMS

Valencia
June 2018
Topics

What it is
How it works: Resonance
**Uncertainty estimates**
Detailed example: Lifetime
Several Parameters
Extended maximum $\mathcal{L}$

Do’s and Don’t’s with $\mathcal{L}$
Simple example: Angular distribution

\[ y = N (1 + \beta \cos^2\theta) \]
\[ y_i = N (1 + \beta \cos^2\theta_i) \]

= probability density of observing \( \theta_i \), given \( \beta \)

\[ L(\beta) = \prod y_i \]

= probability density of observing the data set \( y_i \), given \( \beta \)

Best estimate of \( \beta \) is that which maximises \( L \)

Values of \( \beta \) for which \( L \) is very small are ruled out

Precision of estimate for \( \beta \) comes from width of \( L \) distribution

CRUCIAL to normalise \( y \)

\[ N = \frac{1}{2(1 + \beta/3)} \]

(Information about parameter \( \beta \) comes from shape of exptl distribution of \( \cos\theta \))
How it works: Resonance

\[ y \sim \frac{\Gamma/2}{(m-M_0)^2 + (\Gamma/2)^2} \]

Vary \( M_0 \)

Vary \( \Gamma \)
Conventional to consider

\[ \ell = \ln(L) = \sum \ln(y_i) \]

For large \( N \), \( L \to \text{Gaussian} \)

"Proof"

Taylor expand \( \ell \) about its maximum

\[ \ell = \ell_{\text{max}} + \frac{1}{2!} \ell'' \left[ \delta \left( \frac{\delta}{\sigma} \right) \right]^2 + \ldots \]

\[ = \ell_{\text{max}} - \frac{1}{2\sigma} \delta^2 + \ldots \]

\[ \therefore L \sim \exp \left( -\frac{\delta^2}{2\sigma} \right) \]
Maximum likelihood uncertainty

Range of likely values of param $\mu$ from width of $L$ or l dists. If $L(\mu)$ is Gaussian, following definitions of $\sigma$ are equivalent:

1) RMS of $L(\mu)$

2) $1/\sqrt{-d^2\ln L / d\mu^2}$ (Mnemonic)

3) $\ln(L(\mu_0 \pm \sigma)) = \ln(L(\mu_0)) - 1/2$

If $L(\mu)$ is non-Gaussian, these are no longer the same

"Procedure 3) above still gives interval that contains the true value of parameter $\mu$ with 68% probability”

Uncertainties from 3) usually asymmetric, and asym uncertainties are messy. So choose param sensibly

e.g $1/p$ rather than $p$; $\tau$ or $\lambda$
Realistic analyses are more complicated than this.

Lifetime Determination

\[ \frac{dN}{dt} = \frac{1}{\tau} e^{-t/\tau} \]

NORMALISATION

Observe \( t_1, t_2, \ldots, t_N \)

Use but to construct

\[ \lambda = \prod \left( \frac{dN}{dt} \right)_i = \prod \frac{1}{\tau} e^{-t_i/\tau} \]

\[ \lambda = \sum \left( -\frac{t_i}{\tau} - \ln \tau \right) \]

\[ \frac{d \lambda}{dt} = \sum \left( -\frac{t_i}{\tau} - \frac{1}{\tau} \right) = 0 \Rightarrow \lambda = \frac{\sum t_i}{\tau} = \bar{t}_i \]

"Obvious"

\[ \frac{d^2 \lambda}{dt^2} = -\sum \frac{2t_i}{\tau} + \sum \frac{1}{\tau^2} = -2\frac{N}{\tau^2} + \frac{N}{\tau^2} = -\frac{N}{\tau^2} \]

\[ \Rightarrow \sigma_\tau = \sqrt{-\frac{d^2 \lambda}{dt^2}} = \bar{t}/\sqrt{N} \]

N.B. 1) Usual \( 1/\sqrt{N} \) behaviour

2) \( \sigma_\tau \propto \tau^{-1/2} \)

Beware for averaging results
Several Parameters

\[ \frac{\partial \ell}{\partial \beta} = 0 \]
\[ \sigma_\beta^2 = \frac{1}{(-\frac{\partial^2 \ell}{\partial \beta^2})} \]

Many dimensions: \[ \ell(\beta_1, \beta_2, \beta_3, \ldots) \]
\[ \beta_1, \beta_2, \beta_3, \ldots \text{ from } \frac{\partial \ell}{\partial \beta_i} = 0 \]

For errors, define \( H_{ij} = -\frac{\partial^2 \ell}{\partial \beta_i \partial \beta_j} = \text{Inverse Error Matrix} \)

\[ E_{ij} = (H^{-1})_{ij} \]

\[ \ell_{\text{prof}} = \ell(\beta, \nu_{\text{best}}(\beta)), \] where
\( \beta = \text{param of interest} \)
\( \nu = \text{nuisance param(s)} \)

Uncertainty on \( \beta \) from decrease in \( \ln(\ell_{\text{prof}}) \) by 0.5

N.B. Profile \( \ell \) contains less information than \( \ell \).
Can be important
Extended Maximum Likelihood

Maximum Likelihood uses shape $\rightarrow$ parameters
Extended Maximum Likelihood uses shape and normalisation
i.e. EML uses prob of observing:
  a) sample of $N$ events; and
  b) given data distribution in $x$,……
  $\rightarrow$ shape parameters and normalisation.

Example: Angular distribution

Observe $N$ events total e.g. 100
  F forward 96
  B backward 4

Rate estimates ML EML
  Total --- 100±10
  Forward 96±2 96±10
  Backward 4±2 4±2
ML and EML

ML uses fixed (data) normalisation
EML has normalisation as parameter

Example 1: Cosmic ray experiment
See 96 protons and 4 heavy nuclei
ML estimate  96 ± 2% protons  4 ±2% heavy nuclei
EML estimate  96 ± 10 protons   4 ± 2 heavy nuclei

Example 2: Decay of resonance
Use ML for Branching Ratios
Use EML for Partial Decay Rates
a) Max like

Prob for fixed \( N = \text{Binomial} \)

\[
\text{Prob of } k \text{ successes} = f(k) = \frac{N!}{k!(N-k)!} \left(1-F\right)^{N-k} F^k
\]

Maximise \( \ln P \) wrt \( f \) \( \Rightarrow f = F/N \)

Error in \( f \) : \( \sigma^2 = \frac{\partial^2 \ln P}{\partial f^2} \)

\[
\sigma^2 = \frac{N}{\hat{f}(1-\hat{f})} \quad \Rightarrow \hat{f} = \frac{N}{\hat{f}(1-\hat{f})}
\]

\( \Rightarrow \) Estimate of \( \hat{F} = NF = F \pm \sqrt{FB/N} \)

\( \hat{B} = N(1-\hat{f}) = B \pm \sqrt{FB/N} \)

b) EML

\( \hat{P} = P \times \frac{e^{-\hat{F}}}{N!} \)

Poisson for overall rate

Maximise \( \ln \hat{P} \) \( \Rightarrow \hat{F} = N\hat{f} = F \pm \sqrt{FB/N} \)

\( \hat{f} = F/N \pm \sqrt{F(1-\hat{f})} \)

For \( \hat{F} + \hat{B} \), either propagate errors for \( \hat{F} = \hat{f} \hat{F} \)

or rewrite eqn \# as product of 2 indep Poissons

\[
\begin{align*}
\hat{F} &= F \pm \sqrt{F} \\
\hat{B} &= B \pm \sqrt{B}
\end{align*}
\]
DO’S AND DONT’S WITH $\mathcal{L}$

- COMBINING PROFILE $\mathcal{L}_s$
- NORMALISATION FOR LIKELIHOOD
- JUST QUOTE UPPER LIMIT
- $\Delta (\ln \mathcal{L}) = 0.5$ RULE
- $\mathcal{L}_{\text{max}}$ AND GOODNESS OF FIT
- $\mathcal{L}$ AND BAYESIAN SMEARING OF $\mathcal{L}$
- USE CORRECT $\mathcal{L}$ (PUNZI EFFECT)
Problems with combining PROFILE $\mathcal{L}$
NORMALISATION FOR LIKELIHOOD

\[ \int P(x \mid \mu) \, dx \quad \text{MUST be independent of } \mu \]

data \quad \text{param}

e.g. Lifetime fit to \( t_1, t_2, \ldots, t_n \) \quad \left[ \tau = \frac{\sum t_i}{N} \right]

INCORRECT \quad P(t \mid \tau) = e^{-t/\tau}

Missing \( 1/\tau \)

\( \tau = \infty \quad \tau \text{ too big} \quad \text{Reasonable } \tau \)
QUOTING UPPER LIMIT

“We observed no significant signal, and our 90% conf upper limit is ….”

Need to specify method e.g.

\[ \mathcal{L} \]

Chi-squared (data or theory error)

Frequentist (Central or upper limit)

Feldman–Cousins

Bayes with prior = const, \( \frac{1}{\mu}, \frac{1}{\sqrt{\mu}}, \mu \) etc

“Show your \( \mathcal{L} \)”

1) Not always practical

2) Not sufficient for frequentist methods
90% C.L. Upper Limits

For Upper Limits

For 2-sided intervals
\[ \Delta \ln \mathcal{L} = -1/2 \text{ rule} \]

If \( \mathcal{L}(\mu) \) is Gaussian, following definitions of \( \sigma \) are equivalent:

1) RMS of \( \mathcal{L}(\mu) \)

2) \( 1/\sqrt{(-d^2 \mathcal{L}/d\mu^2)} \)

3) \( \ln(\mathcal{L}(\mu_0 \pm \sigma) = \ln(\mathcal{L}(\mu_0)) - 1/2 \)

If \( \mathcal{L}(\mu) \) is non-Gaussian, these are no longer the same

"Procedure 3) above still gives interval that contains the true value of parameter \( \mu \) with 68% probability"

Heinrich: CDF note 6438 (see CDF Statistics Committee Web-page)

Barlow: Phystat05
**COVERAGE**

How often does quoted range for parameter include param’s true value?

N.B. Coverage is a property of METHOD, not of a particular exptl result

Coverage can vary with $\mu$

Study coverage of different methods of Poisson parameter $\mu$, from observation of number of events $n$

Hope for:

![Diagram showing coverage $C(\mu)$ vs. $\mu$]
If true for all $\mu$ : “correct coverage”

$P < \alpha$ for some $\mu$ “undercoverage”
(this is serious!)

$P > \alpha$ for some $\mu$ “overcoverage”

Conservative

Loss of rejection
power
Coverage: $\mathcal{L}$ approach (Not Neyman construction)

$$P(n, \mu) = e^{-\mu}\mu^n/n!$$  (Joel Heinrich CDF note 6438)

$$-2 \ln \lambda < 1 \quad \lambda = \frac{P(n, \mu)}{P(n, \mu_{\text{best}})} \quad \text{UNDERCOVERS}$$
Neyman central intervals, NEVER undercover

(Conservative at both ends)
Feldman-Cousins Unified intervals

Neyman construction so NEVER undercovers
Probability ordering

Coverage (C) vs μ: Probability Ordering Intervals  \( (C \to 0.6827 \text{ as } \mu \to \infty) \)
\[ \chi^2 = \frac{(n-\mu)^2}{\mu} \quad \Delta \chi^2 = 0.1 \quad \rightarrow \quad 24.8\% \text{ coverage?} \]

NOT Neyman : Coverage = 0% → 100%
Unbinned $\mathcal{L}_{\text{max}}$ and Goodness of Fit?

Find params by maximising $\mathcal{L}$
So larger $\mathcal{L}$ better than smaller $\mathcal{L}$
So $\mathcal{L}_{\text{max}}$ gives Goodness of Fit??

Monte Carlo distribution of unbinned $\mathcal{L}_{\text{max}}$
Not necessarily: \( \mathcal{L}(\text{data, params}) \)

Contrast \( \text{pdf(data, params)} \)

e.g. \( p(\lambda) = \lambda \exp(-\lambda t) \)

Max at \( t = 0 \)

Max at \( \lambda = 1/t \)
Example 1

Fit exponential to times $t_1, t_2, t_3 \ldots \ldots \ldots$ [Joel Heinrich, CDF 5639]

$L = \prod \lambda \exp(-\lambda t_i)$

$\ln L_{\text{max}} = -N(1 + \ln t_{av})$

i.e. Depends only on AVERAGE $t$, but is

INDEPENDENT OF DISTRIBUTION OF $t$ (except for………)

(Average $t$ is a sufficient statistic)

Variation of $L_{\text{max}}$ in Monte Carlo is due to variations in samples’ average $t$, but

NOT TO BETTER OR WORSE FIT

Same average $t$ \hspace{1cm} same $L_{\text{max}}$
Example 2

\[ \frac{dN}{d \cos \theta} = \frac{1 + \alpha \cos^2 \theta}{1 + \alpha / 3} \]

\[ \mathcal{L} = \prod_i \frac{1 + \alpha \cos^2 \theta_i}{1 + \alpha / 3} \]

pdf (and likelihood) depends only on \( \cos^2 \theta_i \)

Insensitive to sign of \( \cos \theta_i \)

So data can be in very bad agreement with expected distribution

e.g. all data with \( \cos \theta < 0 \)

and \( \mathcal{L}_{\text{max}} \) does not know about it.

Example of general principle
Example 3

Fit to Gaussian with variable $\mu$, fixed $\sigma$

\[
\text{pdf} = \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}
\]

\[
\ln L_{\text{max}} = N(-0.5 \ln 2\pi - \ln \sigma) - 0.5 \sum (x_i - x_{av})^2 / \sigma^2
\]

\[
\text{constant} \quad \sim \text{variance}(x)
\]

i.e. $L_{\text{max}}$ depends only on variance$(x)$,

which is not relevant for fitting $\mu$ \quad ($\mu_{\text{est}} = x_{av}$)

Smaller than expected variance$(x)$ results in larger $L_{\text{max}}$

Worse fit, larger $L_{\text{max}}$ \quad Better fit, lower $L_{\text{max}}$
$\mathcal{L}_{\text{max}}$ and Goodness of Fit?

Conclusion:

$\mathcal{L}$ has sensible properties with respect to parameters

NOT with respect to data

$\mathcal{L}_{\text{max}}$ within Monte Carlo peak is NECESSARY

not SUFFICIENT

(‘Necessary’ doesn’t mean that you have to do it!)
Binned data and Goodness of Fit using $\mathcal{L}$-ratio

$\mathcal{L} = \prod_i p_{ni}(\mu_i)$

$\mathcal{L}_{\text{best}} = \prod_i p_{ni}(\mu_{i,\text{best}})$

$= \prod_i p_{ni}(n_i)$

$\ln[\mathcal{L}\text{-ratio}] = \ln[\mathcal{L}/\mathcal{L}_{\text{best}}]$  

$\xrightarrow{\text{large } \mu_i} -0.5 \chi^2$  
i.e. Goodness of Fit

$\mathcal{L}_{\text{best}}$ is independent of parameters of fit,
and so same parameter values from $\mathcal{L}$ or $\mathcal{L}$-ratio

Baker and Cousins, NIM A221 (1984) 437
Example 1: Poisson

pdf = Probability density function for observing n, given μ

\[ P(n; \mu) = e^{-\mu} \frac{\mu^n}{n!} \]

From this, construct \( \mathcal{L} \) as

\[ \mathcal{L}(\mu; n) = e^{-\mu} \frac{\mu^n}{n!} \]

i.e. use same function of \( \mu \) and \( n \), but for pdf, \( \mu \) is fixed, but for \( \mathcal{L} \), \( n \) is fixed

N.B. \( P(n; \mu) \) exists only at integer non-negative \( n \)
\( \mathcal{L}(\mu; n) \) exists only as continuous function of non-negative \( \mu \)
Example 2  

Lifetime distribution

pdf  \[ p(t; \lambda) = \lambda e^{-\lambda t} \]

So  \[ L(\lambda; t) = \lambda e^{-\lambda t} \]  (single observed \( t \))

Here both \( t \) and \( \lambda \) are continuous

pdf maximises at \( t = 0 \)

\( L \) maximises at \( \lambda = t \)

N.B. Functional form of \( p(t) \) and \( L(\lambda) \) are different
Example 3: Gaussian

$$\text{pdf}(x; \mu) = \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} / (\sigma \sqrt{2\pi})$$

$$\mathcal{L}(\mu; x) = \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} / (\sigma \sqrt{2\pi})$$

N.B. In this case, same functional form for pdf and $\mathcal{L}$

So if you consider just Gaussians, can be confused between pdf and $\mathcal{L}$

So examples 1 and 2 are useful
Transformation properties of pdf and $\mathcal{L}$

Lifetime example: \( \frac{dn}{dt} = \lambda \, e^{-\lambda t} \)

Change observable from \( t \) to \( y = \sqrt{t} \)

\[
\frac{dn}{dy} = \frac{dn}{dt} \frac{dt}{dy} = 2y\lambda \, e^{-\lambda y^2}
\]

So (a) pdf changes, BUT

(b) \[
\int_{t_0}^{\infty} \frac{dn}{dt} \, dt = \int_{\sqrt{t_0}}^{\infty} \frac{dn}{dy} \, dy
\]

i.e. corresponding integrals of pdf are \textbf{IN Variant}
Now for Likelihood

When parameter changes from $\lambda$ to $\tau = 1/\lambda$

(a’) $L$ does not change

$dn/dt = (1/\tau) \exp\{-t/\tau\}$

and so $L(\tau; t) = L(\lambda = 1/\tau; t)$

because identical numbers occur in evaluations of the two $L$’s

BUT

(b’) $\int_{0}^{\lambda_0} L(\lambda; t) \, d\lambda \neq \int_{\tau_0}^{\infty} L(\tau; t) \, d\tau$

So it is NOT meaningful to integrate $L$

(However,.........)
<table>
<thead>
<tr>
<th></th>
<th>pdf(t;λ)</th>
<th>L(λ;t)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Value of function</strong></td>
<td>Changes when observable is transformed</td>
<td>INVARIANT wrt transformation of parameter</td>
</tr>
<tr>
<td><strong>Integral of function</strong></td>
<td>INVARIANT wrt transformation of observable</td>
<td>Changes when param is transformed</td>
</tr>
<tr>
<td><strong>Conclusion</strong></td>
<td>Max prob density not very sensible</td>
<td>Integrating L not very sensible</td>
</tr>
</tbody>
</table>
CONCLUSION:

\[ \int_{p_l}^{p_u} L \, dp = \alpha \quad \text{NOT recognised statistical procedure} \]

[Metric dependent:

\[ \tau \text{ range agrees with } \tau_{\text{pred}} \]
\[ \lambda \text{ range inconsistent with } 1/\tau_{\text{pred}} \]

BUT

1) Could regard as “black box”

2) Make respectable by \( L \) \( \rightarrow \) Bayes’ posterior

\[ \text{Posterior}(\lambda) \sim L(\lambda) \ast \text{Prior}(\lambda) \quad \text{[and Prior}(\lambda)\text{ can be constant]} \]
6) BAYESIAN SHEARING OF $\mathbf{X}$

"USE $\mathbf{X}^*_{\text{DEG}}$ FOR $\mathbf{S} + \sigma_p$.

SHEAR IT TO INCORPORATE SYSTEMATIC UNCERTAINTIES.

SCENARIO:

$\mathbf{N} = \text{POISSON}(\mu = S \mathbf{E} + b)$

PARAM OF INTEREST $\mathbf{S}$ $\mathbf{E}$ $\mathbf{B}$ $\text{BACKGROUND}$

$\text{SIGNAL/ACCEPTANCE/}d\mathbf{X}$

$\text{UNCERTAINTIES}$

$\text{MEASURED IN \text{\textquoteleft\textquoteleft SUBSIDIARY\textquoteright\textquoteright} EXPERIMENT}$

$P(s, e|\mathbf{N}) = \frac{P(n|s, e) \pi(s, e)}{\int_{s} \cdots \cdots \cdots dsde}$

$P(s|\mathbf{N}) = \int_{e} P(s, e|\mathbf{N}) de$

$$= \int_{e} \frac{\int_{s} \cdots \cdots \cdots dsde}{\int_{s} \cdots \cdots \cdots dsde} \pi(s) \pi(e) de$$

e.g. $\pi(s) = \text{truncated EXP.}$. $\pi(e) \sim e^{-\frac{e}{\sigma}}$.

i.e. SHEAR $\mathbf{X}$ (not $\mathbf{X}^*$) by \textit{prior} for $e$. E.g., $\mathbf{X}$. 
Getting $\mathcal{L}$ wrong: Punzi effect

Giovanni Punzi @ PHYSTAT2003
“Comments on $\mathcal{L}$ fits with variable resolution”

Separate two close signals, when resolution $\sigma$ varies event by event, and is different for 2 signals

e.g. 1) Signal 1 $1 + \cos^2 \theta$
    Signal 2 Isotropic
    and different parts of detector give different $\sigma$

2) $M$ (or $\tau$)
    Different numbers of tracks $\rightarrow$ different $\sigma_M$ (or $\sigma_\tau$)
Events characterised by $x_i$ and $\sigma_i$

A events centred on $x = 0$

B events centred on $x = 1$

$L(f)_{\text{wrong}} = \prod [f * G(x_i,0,\sigma_i) + (1-f) * G(x_i,1,\sigma_i)]$

$L(f)_{\text{right}} = \prod [f * p(x_i,\sigma_i;A) + (1-f) * p(x_i,\sigma_i;B)]$

$p(S,T) = p(S|T) * p(T)$

$p(x_i,\sigma_i|A) = p(x_i|\sigma_i,A) * p(\sigma_i|A)$

$= G(x_i,0,\sigma_i) * p(\sigma_i|A)$

So

$L(f)_{\text{right}} = \prod [f * G(x_i,0,\sigma_i) * p(\sigma_i|A) + (1-f) * G(x_i,1,\sigma_i) * p(\sigma_i|B)]$

If $p(\sigma|A) = p(\sigma|B)$, $L_{\text{right}} = L_{\text{wrong}}$

but NOT otherwise
Punzi’s Monte Carlo for

\[ A : G(x,0,\sigma_A) \]
\[ B : G(x,1,\sigma_B) \]

\[ f_A = 1/3 \]

<table>
<thead>
<tr>
<th>( \sigma_A )</th>
<th>( \sigma_B )</th>
<th>( f_A )</th>
<th>( \sigma_f )</th>
<th>( f_A )</th>
<th>( \sigma_f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.336(3)</td>
<td>0.08</td>
<td></td>
<td>Same</td>
</tr>
<tr>
<td>1.0</td>
<td>1.1</td>
<td>0.374(4)</td>
<td>0.08</td>
<td>0.333(0)</td>
<td>0</td>
</tr>
<tr>
<td>1.0</td>
<td>2.0</td>
<td>0.645(6)</td>
<td>0.12</td>
<td>0.333(0)</td>
<td>0</td>
</tr>
<tr>
<td>1 \rightarrow 2</td>
<td>1.5 \rightarrow 3</td>
<td>0.514(7)</td>
<td>0.14</td>
<td>0.335(2)</td>
<td>0.03</td>
</tr>
<tr>
<td>1.0</td>
<td>1 \rightarrow 2</td>
<td>0.482(9)</td>
<td>0.09</td>
<td>0.333(0)</td>
<td>0</td>
</tr>
</tbody>
</table>

1) \( L_{\text{wrong}} \) OK for \( p(\sigma_A) = p(\sigma_B) \), but otherwise BIASSED

2) \( L_{\text{right}} \) unbiased, but \( L_{\text{wrong}} \) biased (enormously)!

3) \( L_{\text{right}} \) gives smaller \( \sigma_f \) than \( L_{\text{wrong}} \)
Explanation of Punzi bias

\[ \sigma_A = 1 \quad \sigma_B = 2 \]

A events with \( \sigma = 1 \)

B events with \( \sigma = 2 \)

Actual Distribution

Fitting function

[N_A/N_B variable, but same for A and B events]

Fit gives upward bias for N_A/N_B because (i) that is much better for A events; and (ii) it does not hurt too much for B events
Another scenario for Punzi problem: PID

 Originally:

Positions of peaks = constant

\(\sigma_i\) variable, \((\sigma_i)_A \neq (\sigma_i)_B\)

\(\sigma_i \sim\) constant, \(p_K \neq p_\pi\)

COMMON FEATURE: Separation/Error \(\neq\) Constant

Where else??

MORAL: Beware of event-by-event variables whose pdf’s do not appear in \(\mathcal{L}\)
Avoiding Punzi Bias

BASIC RULE:
Write pdf for ALL observables, in terms of parameters

- Include $p(\sigma|A)$ and $p(\sigma|B)$ in fit
  (But then, for example, particle identification may be determined more
  by momentum distribution than by PID)

  OR

- Fit each range of $\sigma$ separately, and add $(N_A)_i \rightarrow (N_A)_{total}$, and similarly for B

Incorrect method using $L_{wrong}$ uses weighted average
of $(f_A)_j$, assumed to be independent of $j$

Talk by Catastini at PHYSTAT05
Conclusions

How it works, and how to estimate uncertainties

$\Delta(\ln L) = 0.5$ rule and coverage

Several Parameters

Likelihood does not guarantee coverage

Unbinned $L_{\text{max}}$ and Goodness of Fit

Use correct $L$ (Punzi effect)