evaluating quality of fit in Unbinned Maximum Likelihood Fitting

- Statistical distribution of λ - zero free parameters
- impact of free parameters
- some speculations

K. Kinoshita
University of Cincinnati
Belle Collaboration

Motivation

Unbinned Maximum Likelihood (UMxL) fitting:

- preferred for determining parameter(s) α via parameter-dependent shape f($\mathbf{x}$; α) of distribution in measured $\mathbf{x}$
- maximizes use of $\mathbf{x}$ information, esp. w. limited statistics
- used in many current analyses - CP, lifetime, Dalitz plot, ...

Goodness-of-fit:

- to answer
  - are data statistically consistent with fitted shape (not easily visualized within binless context, esp. in multiple dim.)?
  - is f($\mathbf{x}$; α) a valid parametrization?
- Straightforward for least-squares
- To date, no good test in UMxL - why not?

Unbinned Maximum Likelihood (UMxL) fit

Brief outline: have experiment w N measurements of x -
Maximize under variations in **α** (f(**x**; **α**)=normalized PDF)

$$\mathcal{L}(\alpha) = \prod_i f(x_i; \alpha)$$

Equivalent to maximizing:

$$\ln \mathcal{L}(\alpha) \equiv \lambda(\alpha) = \sum_i^N \ln f(x_i; \alpha)$$

Max. at $\alpha = \bar{\alpha}$

Wish to examine fit quality - questions:

- How are $\lambda(\bar{\alpha})$ distributed in ensemble, if root is $f(x, \bar{\alpha})$?
  - 0 free parameters
  - effect of free parameters
- at what level can other distributions be ruled out?

Distribution with zero free parameters

Mean: limit for large N= expected mean for finite N

$$\lim_{N\to\infty} \lambda(\bar{\alpha}) = N \int dx\, f(x;\bar{\alpha}) \ln f(x;\bar{\alpha}) \equiv N \hat{\lambda}(\bar{\alpha})$$

Variance: of $\ln f(x;\bar{\alpha})$ over PDF = $N f(x;\bar{\alpha})$

$$\lim_{N\to\infty} V[\lambda(\bar{\alpha})] = N\{\int dx\, f(x;\bar{\alpha})[\ln f(x;\bar{\alpha})]^2 - \hat{\lambda}^2(\bar{\alpha})\}$$

$$\equiv N\hat{\sigma}^2(\bar{\alpha})$$

(“Statistical Methods in Experimental Physics”, Eadie, Drijard, James, Roos, & Sadoulet)

Summary:

Ensemble expts w

•N measurements of x

• 0 free parameters

$$E[\frac{\lambda(\bar{\alpha})}{N}] = \hat{\lambda}(\bar{\alpha})$$

$$V[\frac{\lambda(\bar{\alpha})}{N}] = \frac{\hat{\sigma}^2(\bar{\alpha})}{N}$$

UMxL with free parameter(s)

(1) in each experiment, $\lambda(\alpha)$ is maximized - $\lambda(\alpha_{max}) \geq \lambda(\bar{\alpha})$

(2) J. Heinrich note (CDF/MEMO/BOTTOM/CDFR/5639) :

toy MC's for 2 different PDF's by UMxL found

$$\lambda(\alpha_{max}) = N\hat{\lambda}(\alpha_{max})$$

confirmed in analytic calculation

-> conjecture: "CL/goodness" is always 100%

First, examine (2)…

Does fitted α_{max} always give expected $\lambda(\alpha_{max}) \sim N\hat{\lambda}(\alpha_{max})$?

Rewrite parametrized PDF $f(x;\alpha)$ -> $n(\alpha)e^{-h(x;\alpha)}$

measured distribution $g(x) = \sum_i^N \delta(x - x_i)$

$$\lambda(\alpha) = \int dx\ g(x) \ln f(x;\alpha) = \int dx\ g(\ln n - h) = N \ln n - \int dx\ gh$$

To maximize, $\frac{\partial \lambda}{\partial \alpha} = 0 = N \frac{\partial \ln n}{\partial \alpha} - \int dx\ g \frac{\partial h}{\partial \alpha}$

=> $\frac{\partial \ln n}{\partial \alpha} = \frac{1}{N} \int dx\ g \frac{\partial h}{\partial \alpha} \equiv < \frac{\partial h}{\partial \alpha} >$ <- **"expectation value" over PDF** g/N

$$\boxed{\lambda(\alpha_{max})} = N[\ln n(\alpha_{max}) - < h(\alpha_{max}) >]$$

The bottom line:

Just 2 measured numbers characterize data vis-a-vis f:

$< \frac{\partial h}{\partial \alpha}(\alpha_{max}) >$ $< h(\alpha_{max}) >$ **(+maximization of λ constrains 1)**

(averages, not highly correlated w shape of data distribution -> no GoF)

Look at PDF's examined by Heinrich:

(a) $f(x;\alpha) = \frac{1}{\alpha} e^{-x/\alpha}$: **(1 param - 2 measured #'s, 1 constraint)**

Note: $\frac{\partial h}{\partial \alpha} = -\alpha h$ **=> 0 DoF in λ_{max}**

(b) $f(x;\alpha_1,\alpha_2) = \frac{1}{\sqrt{2\pi}\alpha_2} exp(\frac{-(x-\alpha_1)^2}{2\alpha_2^2})$: **(2 params - 3 measured #'s, 2 constraints)**

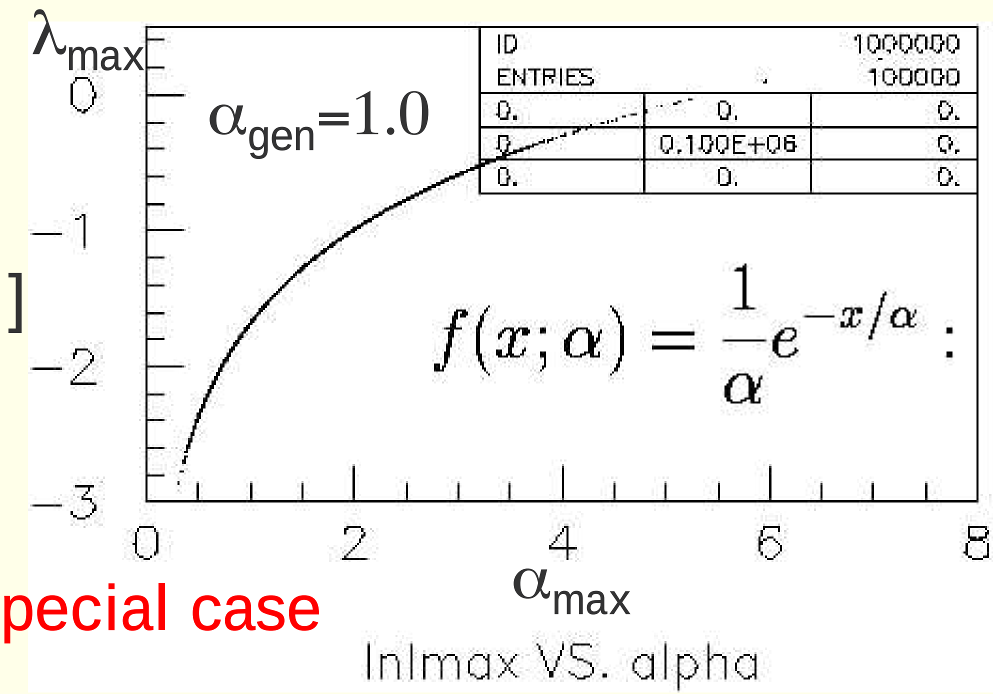
Note: $\frac{\partial h}{\partial \alpha_2} = -\alpha_2 h$ **=> 0 DoF**

i.e. these are special cases where $< \frac{\partial h}{\partial \alpha}(\alpha_{max}) >$ **fixes** $< h(\alpha_{max}) >$

Illustrate:

Seen in λ_{max} vs α_{max}

- Always get $\lambda_{max} = E[\lambda(\alpha_{max})]$



100% correlation is special case

However… there is often a partial correlation …

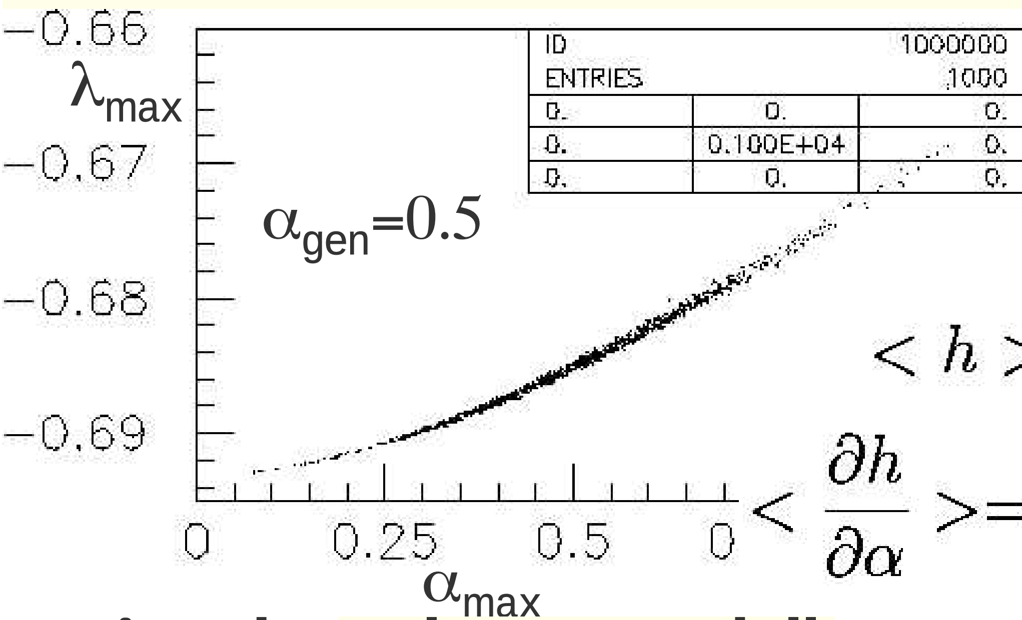
… example

$$f(x;\alpha) = \frac{1+\alpha x^2}{2(1+\alpha/3)}$$

(-1<x<1)

$$h(x;\alpha) = \ln(1+\alpha x^2) = \sum_{i=1}^{\infty} \frac{(-1)^{i+1}(\alpha x^2)^i}{i}$$

$$\frac{\partial h(x;\alpha)}{\partial \alpha} = \frac{1}{\alpha} \sum_{i=1}^{\infty} (-1)^{i+1}(\alpha x^2)^i$$



=> 2 largest terms are highly correlated

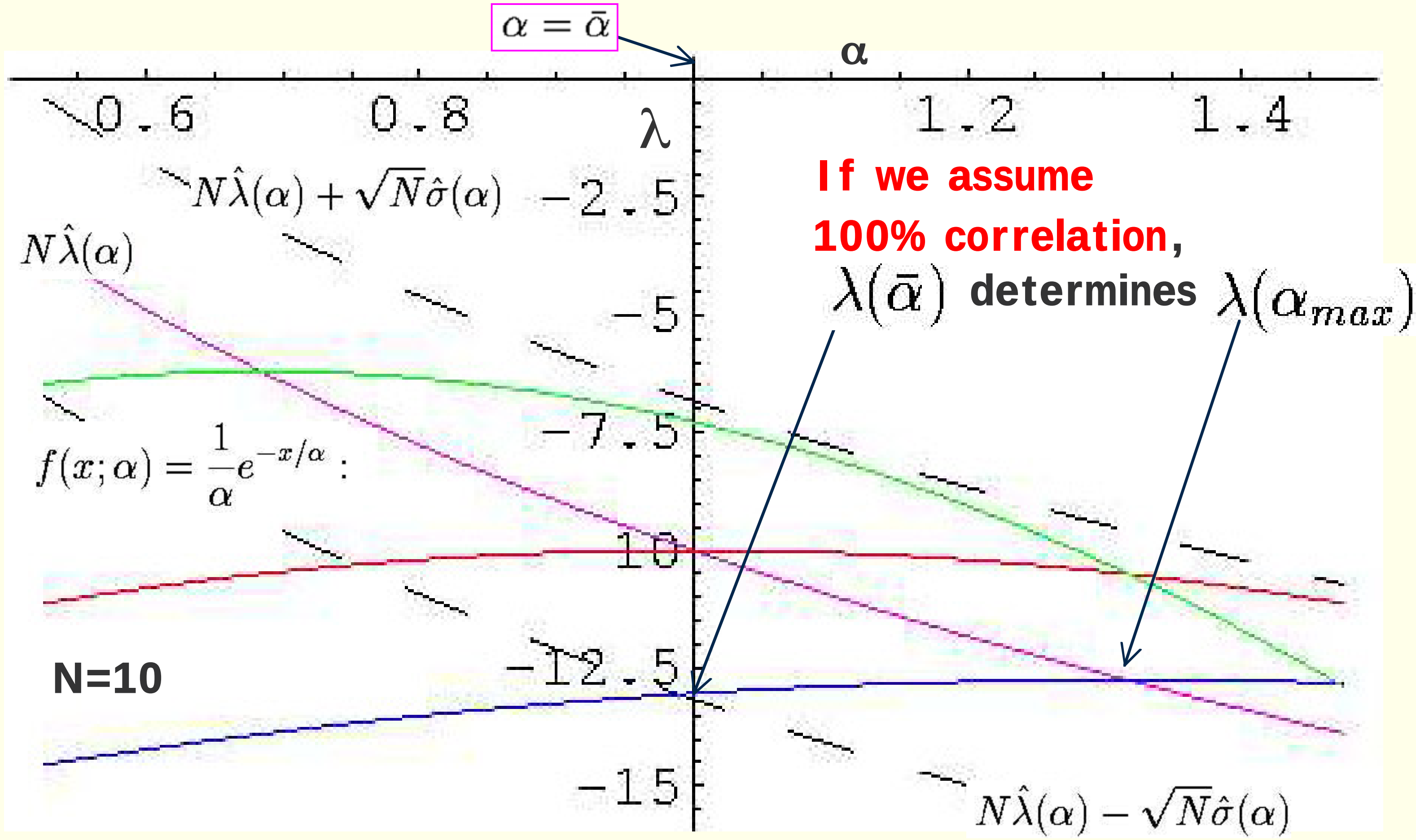
$$< h >= \alpha < x^2 > - \frac{\alpha^2 < x^4 >}{2} + \ldots$$

$$< \frac{\partial h}{\partial \alpha} >= \frac{1}{\alpha}[\alpha < x^2 > -\alpha^2 < x^4 > + \ldots]$$

$\Rightarrow \lambda_{max}$ **is at least partially correlated w measured α_{max},**
INDEPENDENTLY from actual distribution in data->no GoF

Can λ_{max} be used within parametrization to set confidence interval?

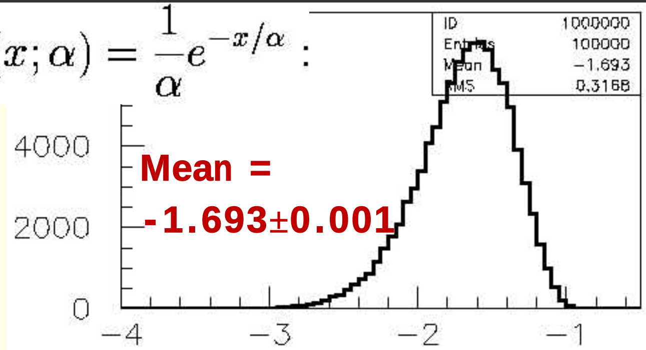
Need distribution in λ_{max} - how much is mean λ shifted by fit?



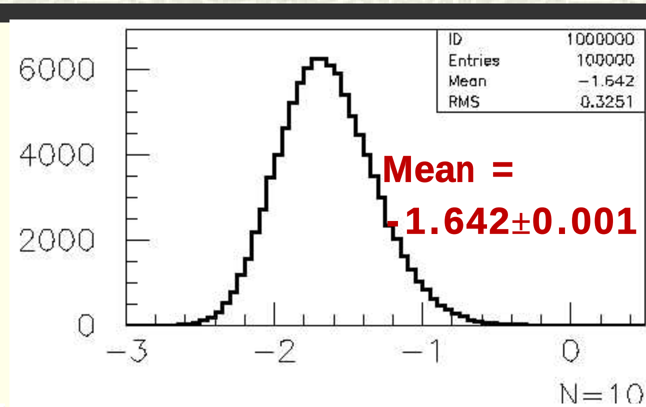
max $\Delta\alpha$=O(σ_α)->conjecture: $\Delta\lambda$=O(0.5) per fitted parameter

Test on the same suspects

$$f(x;\alpha) = \frac{1}{\alpha} e^{-x/\alpha} :$$



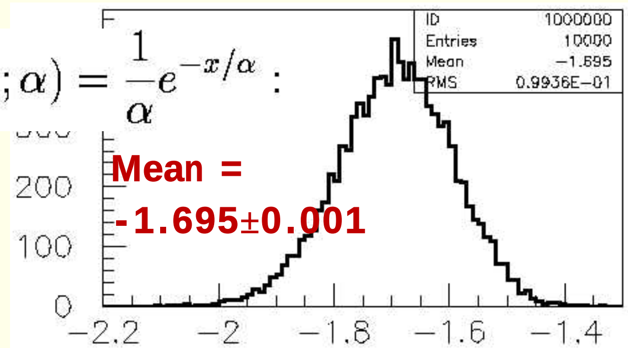
Mean = -1.693±0.001



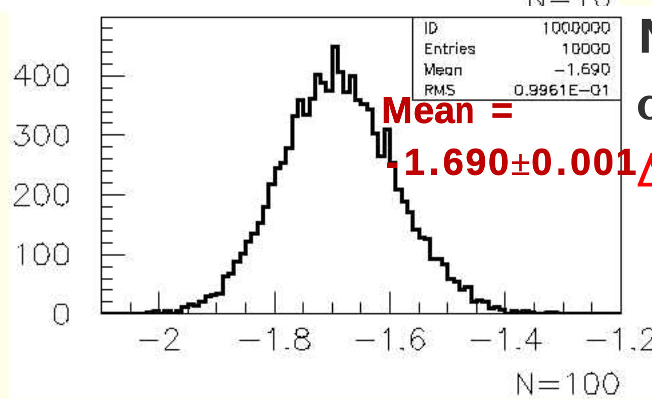
Mean = -1.642±0.001

N=10
α=1.0
Δλ=0.51±0.01

$$f(x;\alpha) = \frac{1}{\alpha} e^{-x/\alpha} :$$



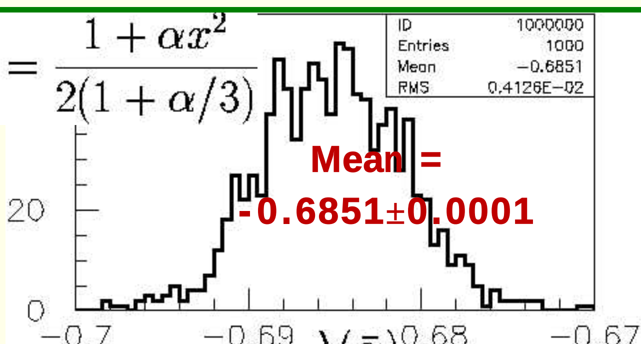
Mean = -1.695±0.001



Mean = -1.690±0.001

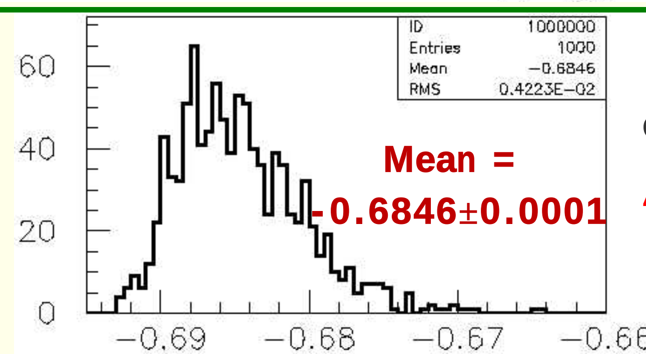
N=100
α=1.0
Δλ=0.5±0.1

$$f(x;\alpha) = \frac{1+\alpha x^2}{2(1+\alpha/3)}$$



Mean = -0.6851±0.0001

$\frac{\lambda(\bar{\alpha})}{N}$



Mean = -0.6846±0.0001

$\frac{\lambda(\alpha_{max})}{N}$

N=1000
α=0.5
Δλ=0.5±0.1

… can λ_{max} be used within parametrization to set confidence interval?

Maybe -

- with stronger demo of distribution shift, width shift, extension to multiple parameters
- nice for multi-parameter fits - reduce to 1-d

Other speculations

tests of fit quality using information generated in UMxL

… without resorting to binning

- Test on subsets of fitted sample, e.g. $\sin 2\phi_1$ result from simultaneous fit over many decay modes - compare λ_{max} in different sets w. expectation - "χ^2"
- event-by-event distribution $\{\lambda_i(\alpha_{max})\}$ - moments, or K-S test

Summary

Goodness-of-fit for UMxL

- sorry, not possible with λ_{max} alone

Other measures of fit quality

Desirable, especially for multiparameter fitting

- steps toward definition of λ_{max} distribution for general PDF
- speculation - exploit info in $\{\lambda_i(\alpha_{max})\}$