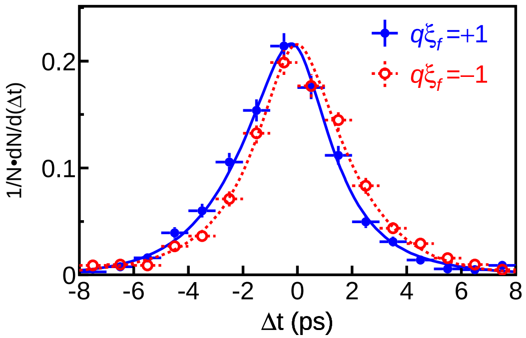
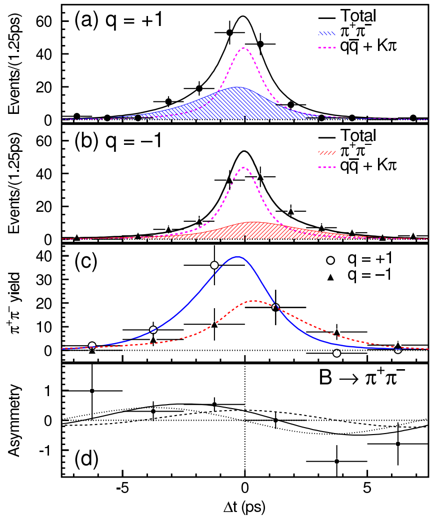
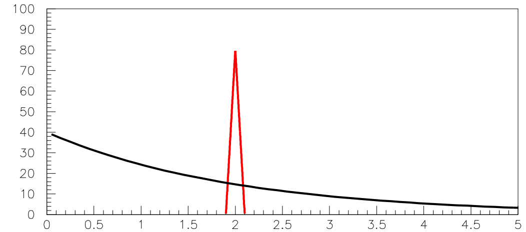
Goodness-of-fit tests

for unbinned maximum likelihood







B.D. Yabsley (Virginia Tech) & K. Kinoshita (U. Cincinnati)

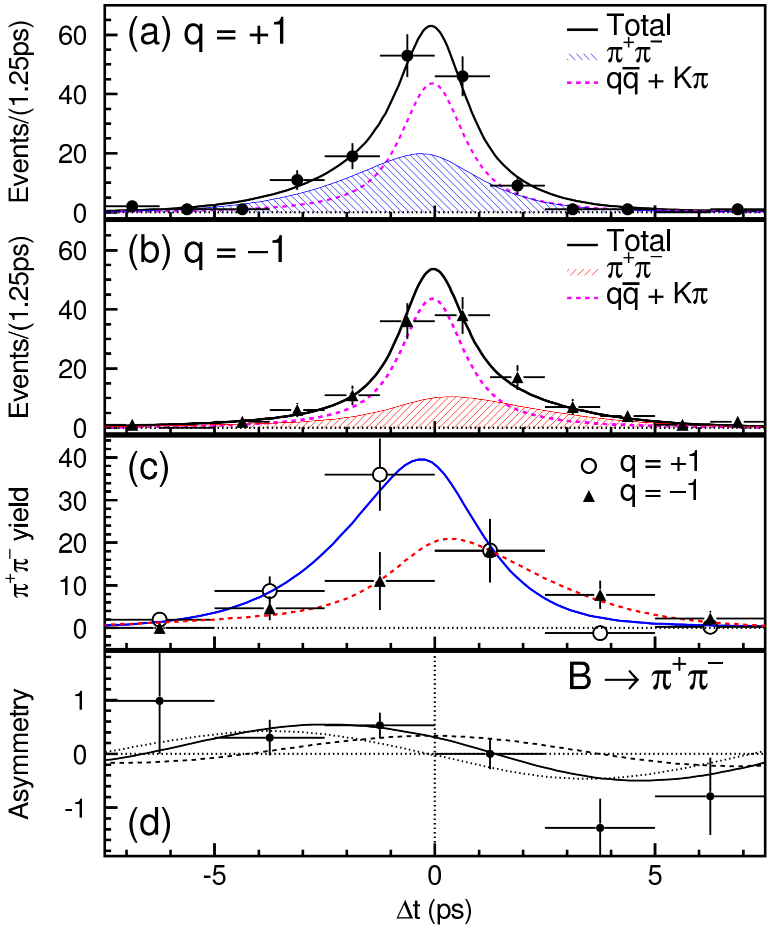
PHYSTAT, SLAC, 8th September 2003

Why we are interested

Belle & other exp'ts make increasing use of unbinned fits:

e.g. CPV params $(A_{\pi\pi}, S_{\pi\pi})$:

- $e^+e^- \to \Upsilon(4S) \to B\overline{B}$
- use $\Delta t \approx \Delta z/\beta\gamma c$; $\sigma_{\Delta z}$ varies
- $S/B \sim 1 \Longrightarrow$ bin by
  flavour-tagging quality
  continuum suppression variable
- $\mathcal{P}_i$ in $\mathcal{L} = \prod_i \mathcal{P}_i$ depends on
  fractions $f_{\pi\pi}$, $f_{K\pi}$, $f_{q\bar{q}}$
  $(f_{\pi\pi} + f_{K\pi} + f_{q\bar{q}} = 1)$
- only 760 events



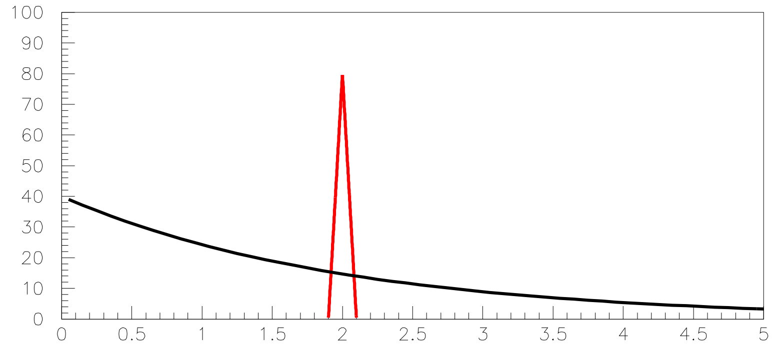
binning is impractical $\Longrightarrow$ UML

Why this presents a problem

. . . the $\mathcal{L}$ value itself seems to have no power
to discriminate against a bad fit

Heinrich's analytic example:

- data: $\delta(x - x_0)$
- fit to $\frac{1}{\alpha}\exp(-x/\alpha)$
- α floats ("compound hypothesis")
- fit finds $\alpha = x_0$
- $-2\ln\mathcal{L}_{\max} = 2N(1 + \ln x_0)$
- expectation for a true exponential, constant α, is $2N(1 + \ln\alpha)$
- $\Longrightarrow \delta$ "looks like" exp



the discrimination power of the $\mathcal{L}$ disappears when its parameter(s) float

(see also Kinoshita (2002))

$\mathcal{L}$ depends on the aggregate $\mathcal{P}_i$ of the points, *not where they are*

To test the fitted *vs* data distributions . . .

- apply a binned χ^2 test to projections of the fit?
  - sensitive to shape differences
  - but fictitious (*e.g.* what is the n.d.f?) $\Longrightarrow$ hard to quantify
  - what about correlations between quantities?
  - what about the multi-dimensional case (e.g. amplitude analysis) $\Longrightarrow$ no intuitive backup
- apply a binning-free goodness-of-fit test to compare (a) the data, to (b) the fitted model [this is "obvious"]

Kay: develop a test for the purpose: $\longrightarrow$ 2^{nd} half of talk

Bruce: take an existing test off-the-shelf and apply it
- I use the *energy test* os Aslan & Zech [hep-ex/0203010]
- this works for *an arbitrary number of dimensions* $\Longrightarrow$ applicable to HEP problems

The energy test

Inspired by the electrostatic energy between distributions of positive charge (say f_0) and negative charge (say f):

$$\phi = \frac{1}{2}\int\int [f(\mathbf{x}) - f_0(\mathbf{x})]\left[f(\mathbf{x}') - f_0(\mathbf{x}')\right] R(\mathbf{x}, \mathbf{x}')\mathbf{dxdx}'$$

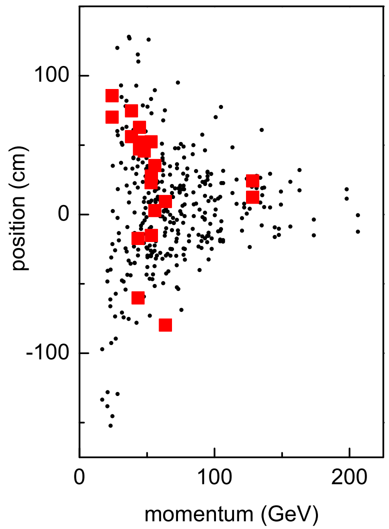
Model by N data pts $(\mathbf{x}_i)$ & M theory pts $(\mathbf{y}_j)$ with $M = 10 \cdot N$ (say):

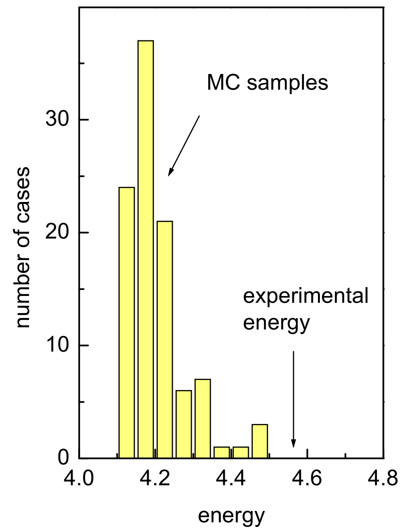
$$\phi_{NM} = \frac{1}{N^2}\sum_{j>i} R(|\mathbf{x}_i - \mathbf{x}_j|) - \frac{1}{NM}\sum_{i,j} R(|\mathbf{x}_i - \mathbf{y}_j|)$$

Gives χ^2-like discrimination in 1D & 2D (presum^{ly} in multi-D)

Choice of weight $R(\mathbf{x}, \mathbf{x}')$):

- logarithmic
- power law
- Gaussian





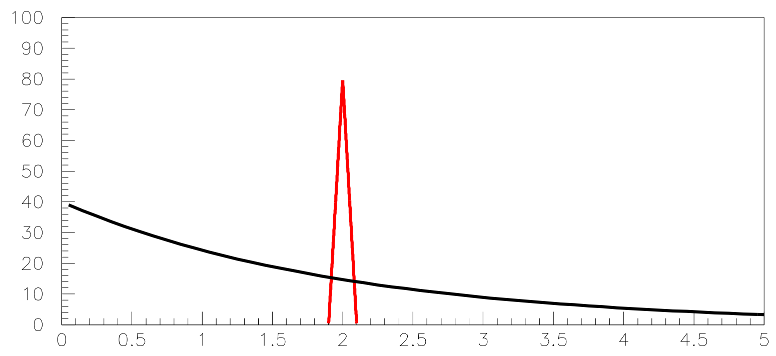
An analytic study of the energy test

Heinrich's example is simple enough to implement on paper

but not for Aslan & Zech's weight f^{ns}; I use $R(\mathbf{x},\mathbf{x}') = \exp\left(-\beta\,|\mathbf{x}-\mathbf{x}')|\right)$

Leave in the constant omitted by A&Z $\longrightarrow E(\exp) = 0$

For exp fitted to $\delta(x - \frac{1}{\alpha})$, after some tedious arithmetic ...



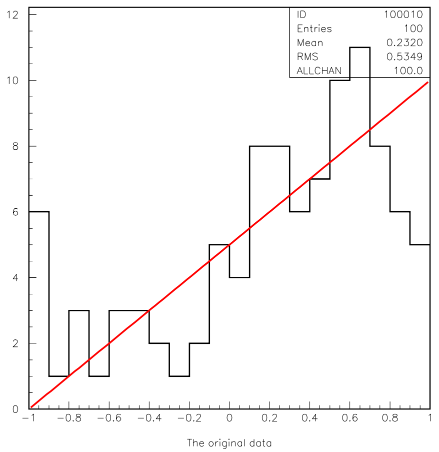
$$\phi = 1 + \frac{1}{\alpha^3(\alpha+\beta)} - \frac{2}{\alpha}\left(\frac{\exp(-\beta/\alpha)}{(\beta-\alpha)}(1-\exp(-(\beta-\alpha)/\alpha)) + \frac{e^{-1}}{(\beta+\alpha)}\right)$$

= self-repuls^n of data + self-repuls^n of th^y − attract^n of data and th^y

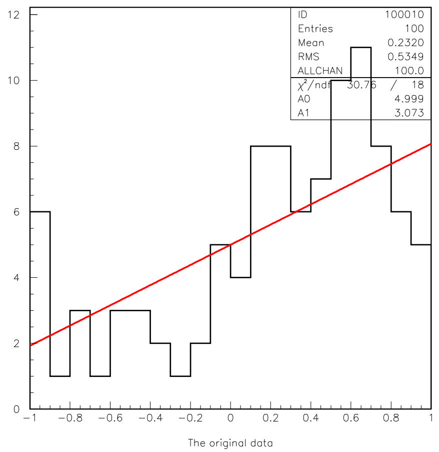
$$\text{For } \beta \gg \alpha,\ \phi \to 1 + \frac{1}{\alpha^3\beta} - \frac{2e^{-1}}{\alpha\beta} > 0$$

The working example

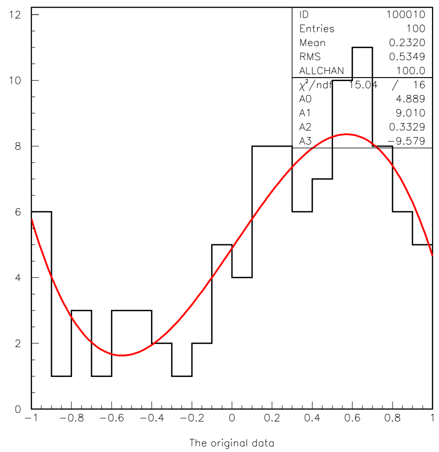
A linear distribution with a contamination thrown in . . .



easy to discriminate against fixed slope



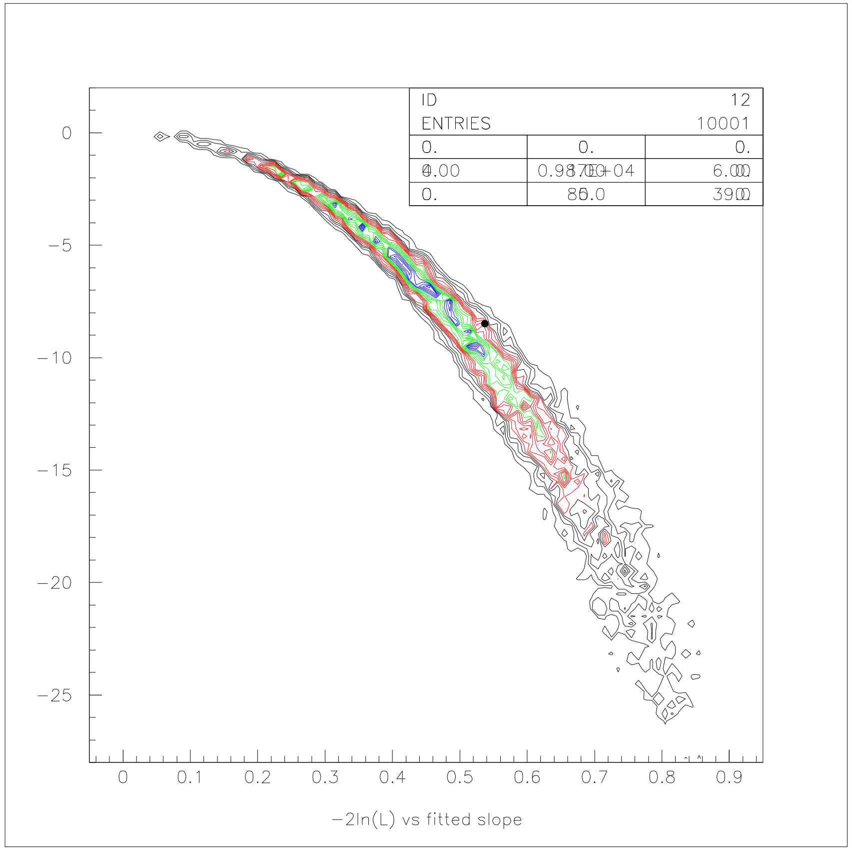
harder to discriminate with floating slope

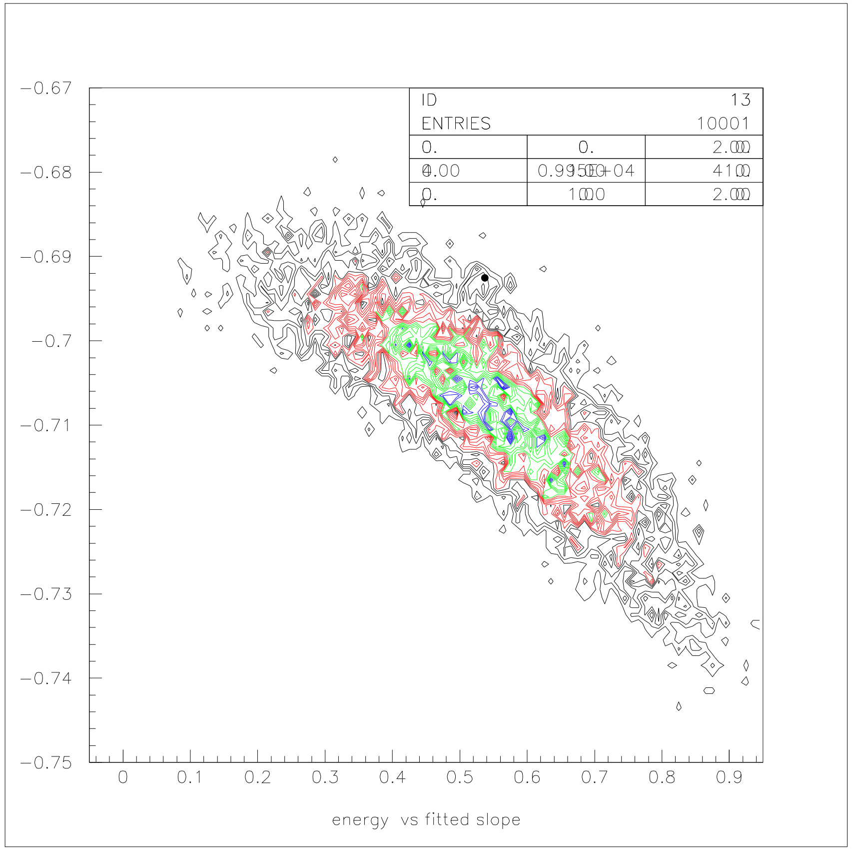


we pretend not to see the true distribution

This is the first of a graded set of examples leading up to pseudo-$B^0 \to \pi^0\pi^0$ analysis . . .

Correlation between fitted param. and measure





The empirical study is a work in progress . . .

An Unbinned Goodness-of-Fit Test Based on the Random Walk

- Test statistic
  - flat distribution
  - null rejection power
  - flattening
  - compound hyopthesis
- multidimensional extension: speculations

Kay Kinoshita
University of Cincinnati
Belle Collaboration

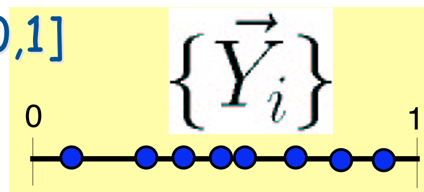
Test statistic

PDF can be flattened:

$$dY = f(X)\, dX$$

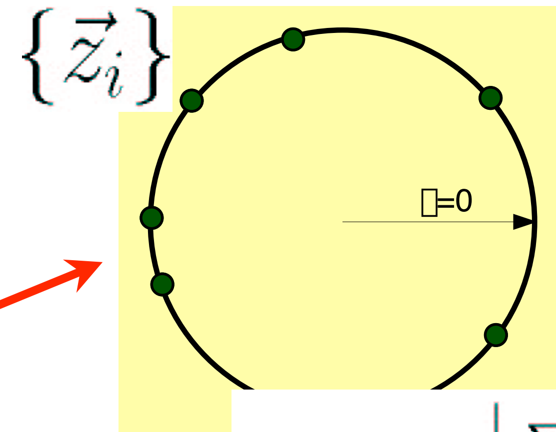
-> $$Y_i = \int_{X_1}^{X_i} f(X) dX$$

on interval [0,1]



map to circle: $\phi_i = 2\pi Y_i$

if flat PDF, 2-d random walk!



statistic: $$K_1 \equiv \frac{|\sum_{i=1}^N \vec{z}_i|^2}{N}$$

(appears in von Mises test)

Equivalent: $$K_1 = \frac{|\mathcal{F}(1)|^2}{N}$$

(1st term, fourier series)

$$\mathcal{F}(k=1) = \int_0^{2\pi} d\phi \sum_{i=1}^N e^{ik\phi} \delta(\phi - \phi_i)$$

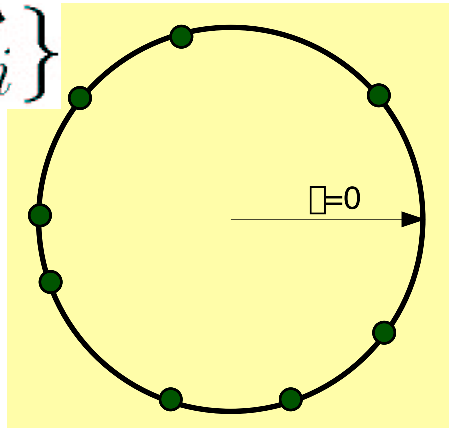
$$= \sum_{i=1}^N e^{i\phi_i}$$

Ensemble distribution:

$$e^{-K_1}$$

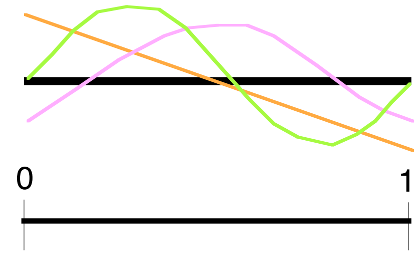
Expectations for K_1

$\{\vec{z}_i\}$

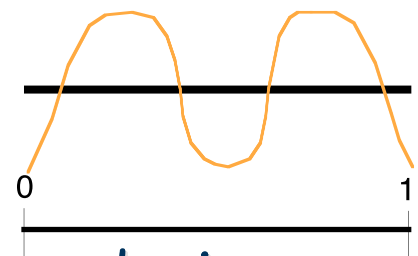


Vector sum:

- sensitive to overall distortions at □□~π



- insensitive to local fluctuations (equal weight for all □ regions)
- insensitive to higher order distortions:



- no binning, scale, data ordering

for sensitivity to higher order, could use

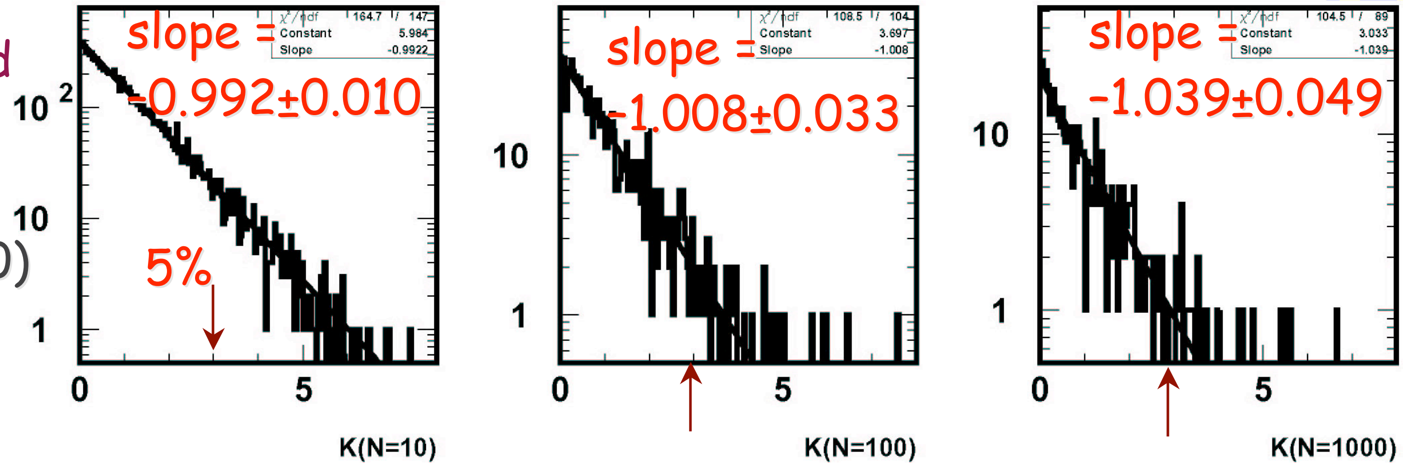
$$K_k = \frac{|\mathcal{F}(k)|^2}{N} \qquad \mathcal{F}(k) = \sum_{i=1}^{N} e^{ik\phi_i}$$

here, examine only k=1

Flat null hypothesis

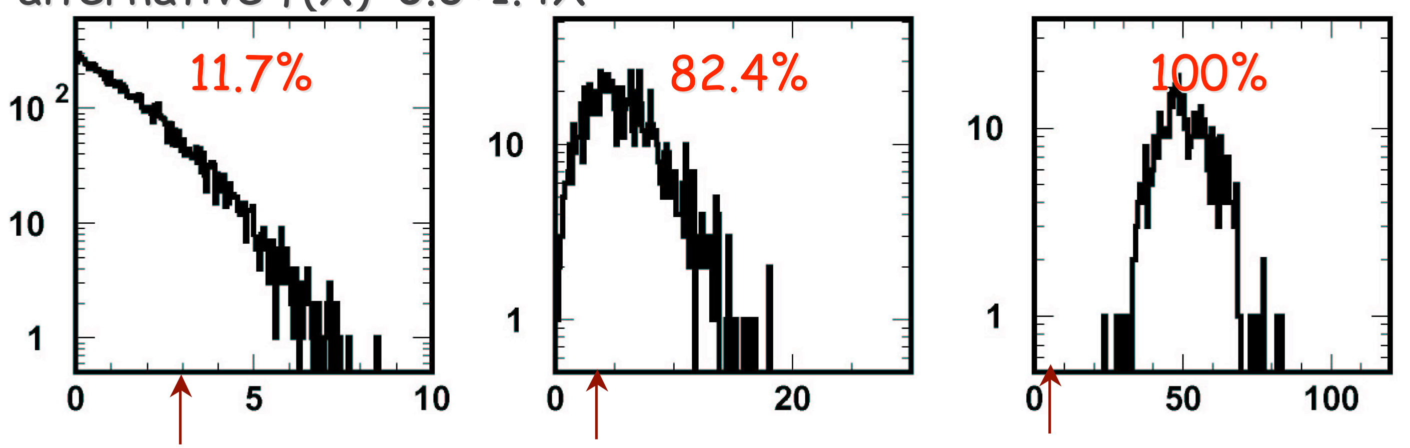
slope = −1/(decay constant)

MC - generated flat ensemble
K_1 distribution
(N=10,100,1000)



rejection: e.g. from Aslan & Zech hep-ex/0203010

alternative f(X)=0.3+1.4X



Null hypotheses

Flat null: alternative hypotheses from Aslan & Zech

Function	Rejection Power		
	$N = 10$	$N = 100$	$N = 1000$
$\mathcal{A}_1(X) = 0.3 + 1.4X$	0.117	0.824	1.00
$\mathcal{A}_2(X) = 0.7 + 0.3[n_2 e^{-64(X-0.5)^2}]$	0.152	0.910	1.00
$\mathcal{A}_3(X) = 0.8 + 0.2[n_3 e^{-256(X-0.5)^2}]$	0.102	0.672	1.00

χ^2, 12 bins
(N=100)
~0.81
~0.85
~0.81

Non-flat null

- Generate according to fixed PDF,
- Flatten " same " "
- Check: ensemble distributionof K_1 is consistent with e^{-K_1}

Compound hypotheses: rejection power

Hypothesis: PDF = $f(\mathbf{X}; \square)$

- Toy MC: generate expts for PDF= $f(\mathbf{X}; \square_0)$

  fit i^{th} expt to $f(\mathbf{X}; \square)$-> $\square_{max,i}$

  flatten, null PDF = $f(\mathbf{X}; \square_{max,i})$

  find K_1 -> ensemble K_1 distribution

Alternate hypothesis: PDF=$g(X)$

- Toy MC: generate expts for PDF= $g(\mathbf{X})$

  fit j^{th} expt to $f(\mathbf{X}; \square)$-> $\square_{max,j}$

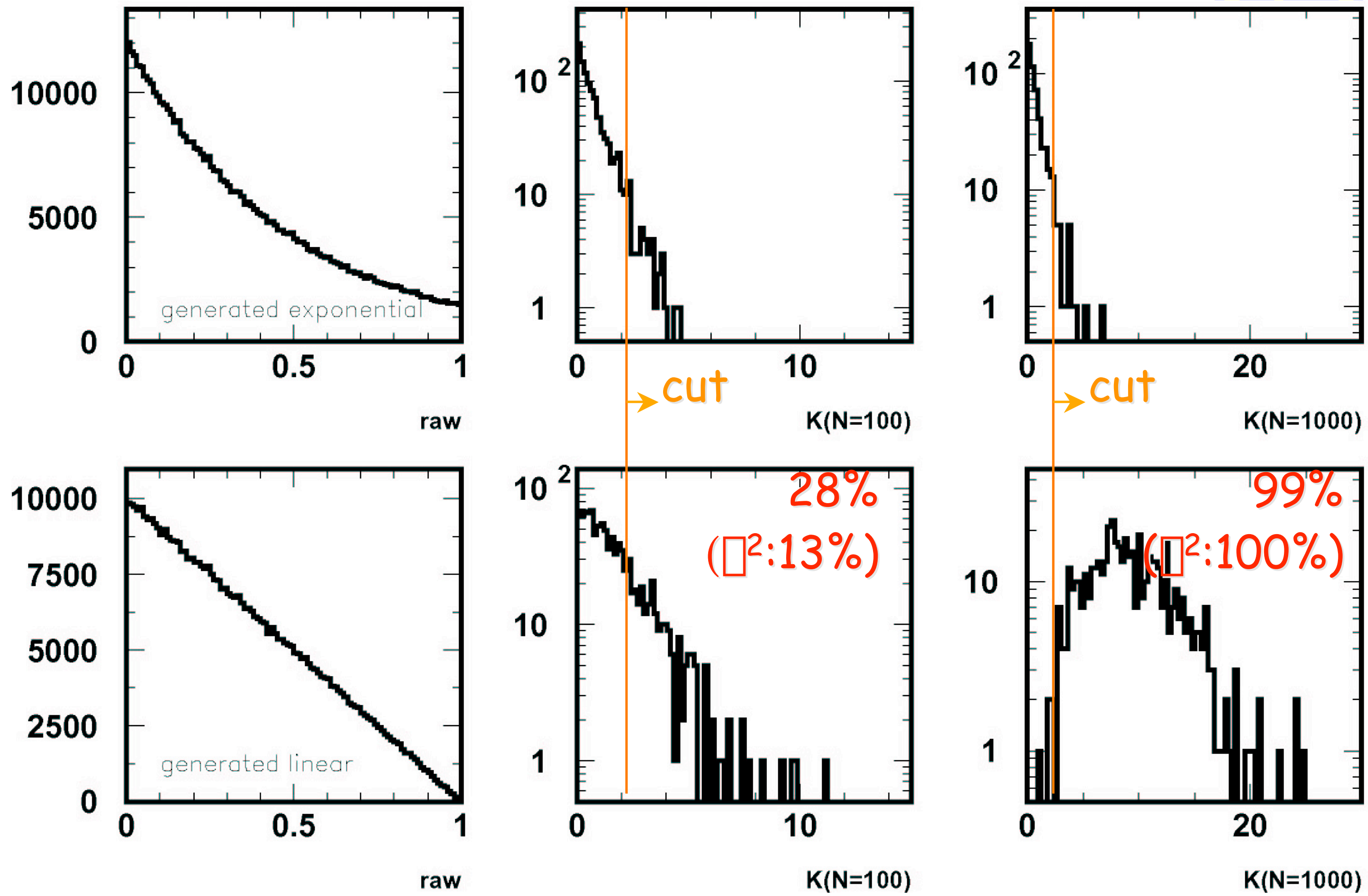
  flatten, null PDF = $f(\mathbf{X}; \square_{max,j})$

  find K_1 -> ensemble K_1 distribution

Compound hypothesis: example

Hypothesis:
Exponential,
Float decay const
(N=100,1000)

Alternate:
$g(X)=2(1-X)$



χ^2: fit w. 20 bins

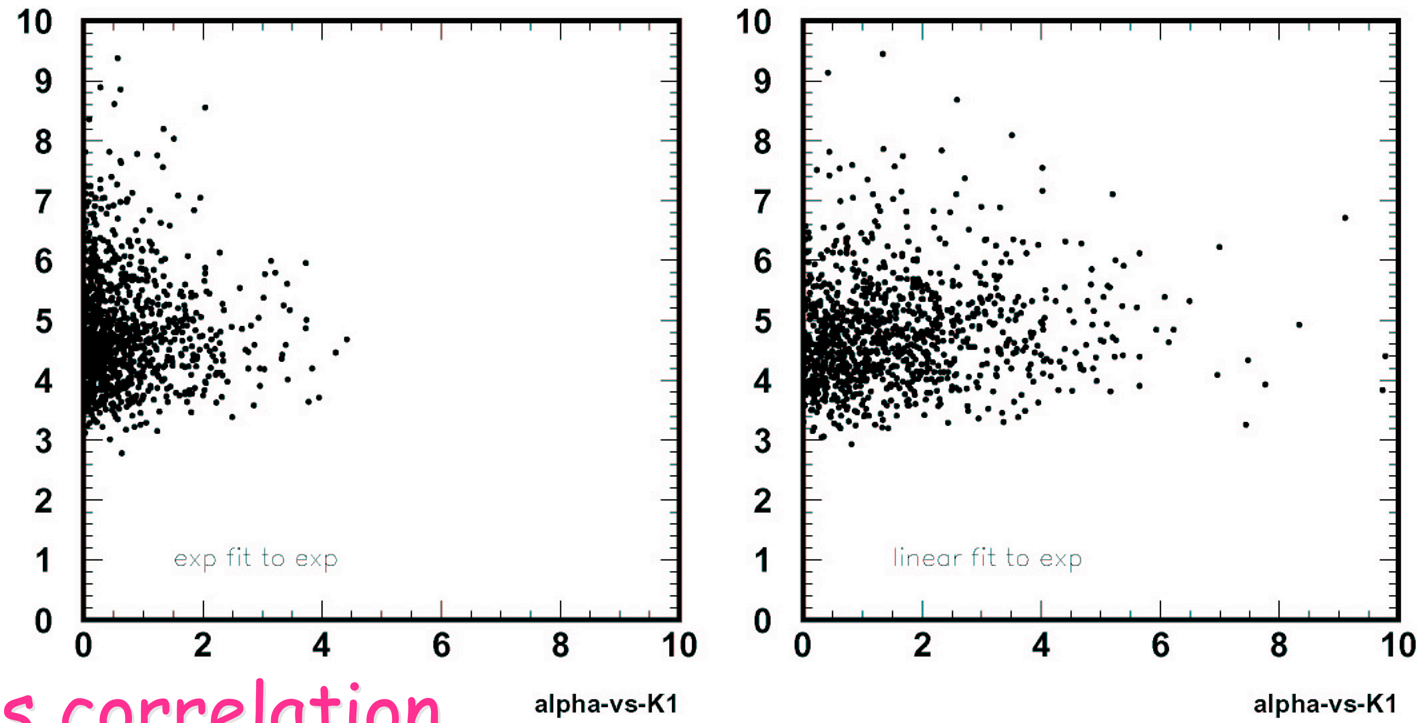
PRELIMINARY: somewhat more powerful than χ^2 for low N

Compound hypothesis: data-parameter correlation

The issue:

Some "goodness-of-fit" params found to correlate w. Data (fitted param) -> not GOF.

□ vs K_1:



no obvious correlation

Compound hypothesis: K_1 distribution

some examples

Form	Generated	Fitted	K_1 (Decay Constant)$^{-1}$ (χ^2/ndf) $N=10$	$N=100$	$N=1000$
$(1-\alpha)+\alpha(2X)$	$\alpha=0.7$	α	–	–	1.28 ± 0.07 (70/67)
$(1-\alpha)+\alpha[n_2 e^{-64(X-0.5)^2}]$	$\alpha=0.3$	α	–	1.90 ± 0.06 (230/80)	1.94 ± 0.09 (223/65)
$(1-\alpha)+\alpha[n_3 e^{-256(X-0.5)^2}]$	$\alpha=0.2$	α	–	1.56 ± 0.05 (203/82)	1.56 ± 0.07 (82/68)
$n_4 e^{-10X/\alpha}$	$\alpha=1.0$	α	1.23 ± 0.01 (147/133)	1.28 ± 0.04 (68/85)	1.28 ± 0.06 (75/76)
$n_5 e^{-[X-(0.5+\alpha_2)]^2/2(\alpha_1/8)^2}$	$\alpha_1=1.0$,	α_1	1.36 ± 0.01 (176/131)	1.38 ± 0.05 (93/85)	1.50 ± 0.07 (56/65)
	$\alpha_2=0$	α_2	1.22 ± 0.01 (154/135)	1.25 ± 0.04 (122/96)	1.28 ± 0.06 (73/72)
		α_1,α_2	1.84 ± 0.019 (148/90)	2.00 ± 0.065 (53/59)	2.13 ± 0.095 (47/47)

Observations

- still resemble exponential decay
- "decay constant" varies w fitted function, parameter (not simple as □^{2}/ndf) BUT approx. consistent for N=10->1000!
  => consistency over changes, in parameter, N.
- very amenable to determination via toy MC

Treatment of multidimensional data

To be developed

- K_1 in 1-d, no ordering of data -> amenable to multi-d?

Look at 2-d

- Data on unit circle -> 2-d toroid in 4-d: what is corresponding vector algebra?
- More obvious: projections - extend Fourier formalism

$$\mathcal{F}(k_1, k_2) = \sum_{i=1}^{N} e^{i(k_1\phi_{i1}+k_2\phi_{i2})}$$

k_1=0 or k_2=0 corresponds to projection onto □$_2$ or □$_1$

(k_1,k_2)= (1,1), (1,–1), etc. " other axes

Develop statistic from orthogonal projections?

Summary

- binning-free goodness-of-fit test needed for unbinned maximum likelihood fits of Belle data
  - must accommodate compound hypotheses, multi-d data.
- 2 studies:
  - Aslan/Zech energy test, extends naturally to multi-d
    - examined behavior under compound hypothesis
    - analytic, Monte Carlo
  - K_1 statistic, may extend to multivariate
    - consistency of distribution over different
      - parameter values, N
    - generality for HEP is good, power comparable to χ^2