

Probability and Statistics

for experimental physicists : part II

José Ocariz
Université Paris Diderot and IN2P3

OUTLINE

Lecture 1 (Monday)

Basic concepts in Probability and Statistics

Lecture 2 (today)

**Maximum Likelihood theorem
Multivariate techniques**

Lecture 3 (Thursday)

**An analysis example from *BaBar*
Hypothesis testing, limit settings**

Disclaimer

**Most, if not all of you, are already familiar with many of these topics...
for consistency, the scope spans from the very general concepts towards
more advanced developments...**

MONDAY'S LECTURE : RECAP

We aim at characterizing some generic PDF : $\mathcal{P}(\vec{x}; \theta_1, \dots, \theta_N)$
that is, estimate its parameters θ

- in some specific cases, empirical estimators are OK (sample mean)

$$\text{Parameter : } \mu = E[x] ; \text{ Estimator : } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\text{Bias : } b = E[\hat{\mu}] - \mu = 0 ; \text{ Variance : } V[\hat{\mu}] = \frac{\sigma^2}{n}$$

- or can be easily adapted (sample variance)

$$\text{Parameter : } \sigma^2 = V[x] ; \text{ Estimator : } (RMS)^2 = \bar{x^2} - (\bar{x})^2$$

$$E[(RMS)^2] = \frac{n-1}{n} \sigma^2 \Rightarrow \text{Bias : } b = E[(RMS)^2] - \sigma^2 \neq 0$$

$$\text{Unbiased estimator : } \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 ; \text{ Variance : } V[\hat{\sigma}^2] = \frac{2\sigma^4}{n}$$

- in the general case, need to check for consistency, bias, efficiency, robustness...
- ... use the Maximum Likelihood Theorem !

MAXIMUM LIKELIHOOD THEOREM (I)

For a sample with n independent outcomes of random variables \vec{x}_i , its Likelihood function is defined in terms of the (multi-parametric) PDF of x :

$$\mathcal{L}(\theta_1, \dots, \theta_N) = \prod_{i=1}^n \mathcal{P}(\vec{x}_i; \theta_1, \dots, \theta_N)$$

Fisher, 1921 : parameter estimation via the Maximum Likelihood Estimator :

$$\mathcal{L}([\vec{x}] | \hat{\theta}_1, \dots, \hat{\theta}_N) = \max_{\theta} \mathcal{L}([\vec{x}] | \theta)$$

in other words : for a fixed (observed) sample, choose the θ values that maximize the probability of observing... the observed sample! ...

There is more than a pleonasm to it : the likelihood has to satisfy a few conditions :

- at least twice derivable with respect to θ
- be (asymptotically) unbiased and efficient (“reaches the Cramer-Rao” bound)
- be (asymptotically) normally distributed :

$$f(\hat{\vec{\theta}}; \vec{\theta}, \Sigma) = \frac{1}{\sqrt{2\pi} \det(\Sigma)} e^{-\frac{1}{2} (\hat{\vec{\theta}} - \vec{\theta})^T \Sigma (\hat{\vec{\theta}} - \vec{\theta})}, \quad \Sigma_{ij}^{-1} = -E \left[\frac{\partial \log \mathcal{L}}{\partial \theta_i} \frac{\partial \log \mathcal{L}}{\partial \theta_j} \right]$$

MAXIMUM LIKELIHOOD THEOREM (II)

$$f(\hat{\vec{\theta}}; \vec{\theta}, \Sigma) = \frac{1}{\sqrt{2\pi} \det(\Sigma)} e^{-\frac{1}{2} (\hat{\vec{\theta}} - \vec{\theta})^T \Sigma (\hat{\vec{\theta}} - \vec{\theta})}, \quad \Sigma_{ij}^{-1} = -E \left[\frac{\partial \log \mathcal{L}}{\partial \theta_i} \frac{\partial \log \mathcal{L}}{\partial \theta_j} \right]$$

Very important property : covariance matrix → errors on parameter estimates

In general, $-2\Delta \log \mathcal{L} = -2[\log \mathcal{L}(\theta) - \log \mathcal{L}(\hat{\theta})] = \sum_{i,j} (\theta_i - \hat{\theta}_i) \Sigma_{ij}^{-1} (\theta_j - \hat{\theta}_j)$

can be used to set uni- or multi-dimensional confidence contours

i.e. for 1 parameter, a change in 0.5 units defines a 68% *confidence interval*

Goodness-of-fit : $-2 \log \mathcal{L}(\hat{\theta})$ follows a χ^2 distr. with $n_{dof} = N(\text{events}) - N(\text{parameters})$

$$\text{p-value} = \int_{-2 \log \mathcal{L}(\hat{\theta})}^{+\infty} dx \chi^2(x; n_{dof})$$

is the probability of obtaining a worse agreement, among a large set of samples, all perfectly described by the underlying PDF

MAXIMUM LIKELIHOOD FITS

$$\mathcal{L}(\theta_1, \dots, \theta_N) = \prod_{i=1}^n \mathcal{P}(\vec{x}_i; \theta_1, \dots, \theta_N)$$

The formalism here developed concerns *unbinned* samples : the PDF of each individual outcome is evaluated and used as argument for the likelihood

There are a few useful variants :

- Various species in the data sample (say, “sig” and “bkg”)

$$\mathcal{L}(\theta_1, \dots, \theta_N) = \prod_{i=1}^n [f_{sig} \mathcal{P}_{sig}(\vec{x}_i; \theta_1, \dots, \theta_N) + (1 - f_{sig}) \mathcal{P}_{bkg}(\vec{x}_i; \theta_1, \dots, \theta_N)]$$



(the common method to estimate observables in presence of backgrounds)

- Extended Maximum Likelihood : treat the sample size as an observable, thus add a Poisson term to the PDF,

$$\mathcal{L}(\lambda_{sig}, \lambda_{bkg}, \theta_1, \dots, \theta_N) = e^{-(\lambda_{sig} + \lambda_{bkg})} \prod_{i=1}^n \mathcal{P}(\vec{x}_i; \theta_1, \dots, \theta_N)$$

- weighted datasets : conceptually straightforward, beware of normalization issues (errors are likely to be incorrectly reported!)

MORE ON UNBINNED ML FITS : EFFICIENCIES

The “intuitive” estimate of efficiency is a simple ratio :

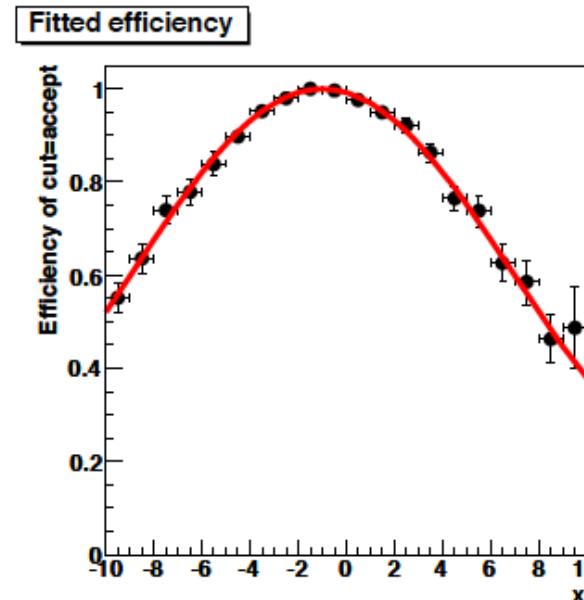
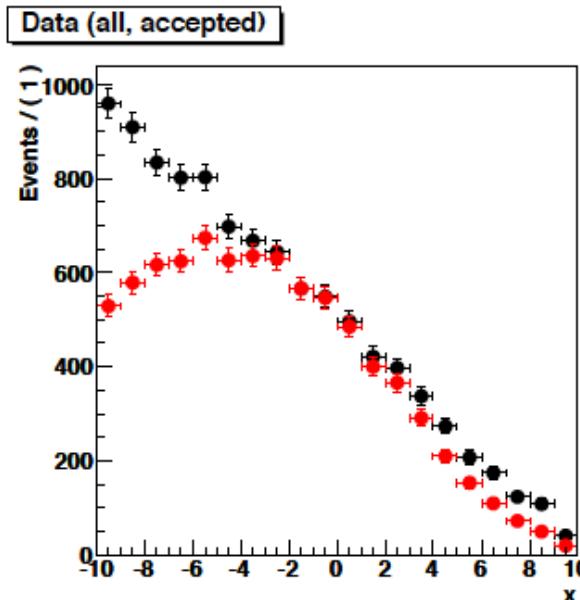
$$\varepsilon = \frac{\text{yes}}{\text{yes} + \text{no}}$$

its uncertainty is often estimated as
(assuming a symmetric binomial behavior)

$$V[\varepsilon] = \frac{\varepsilon(1-\varepsilon)}{n} \quad (n = \text{yes} + \text{no})$$

...this clearly breaks down both in the low- n and very high (in-)efficiency levels...
... which is often the case if one wants to evaluate
ML approach : together with the relevant observables x , (and possible other
parameters θ) including a discrete random variable $c = \{\text{yes}, \text{no}\}$ the PDF becomes

$$P(c|x, \theta) = \delta(c - \text{yes})\varepsilon(x, \theta) + \delta(c - \text{no})[1 - \varepsilon(x, \theta)]$$



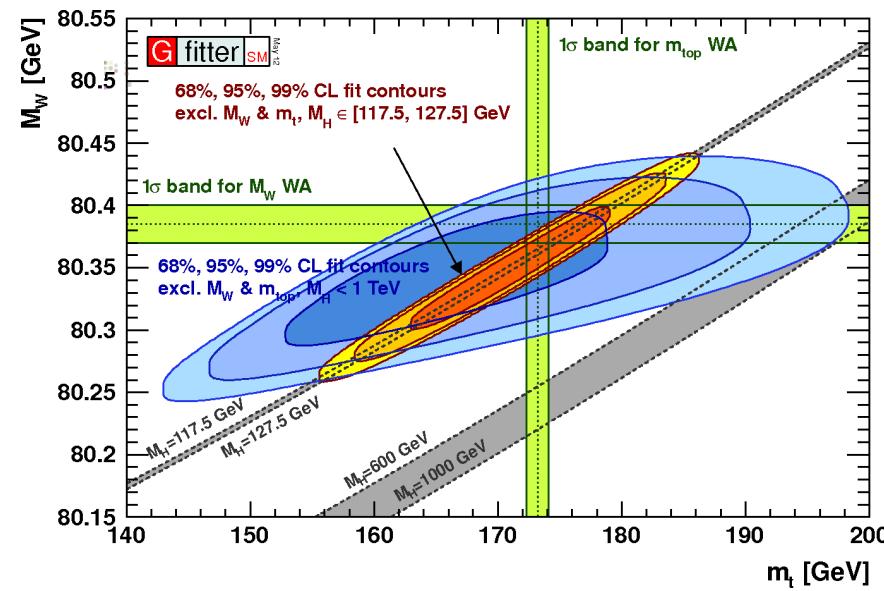
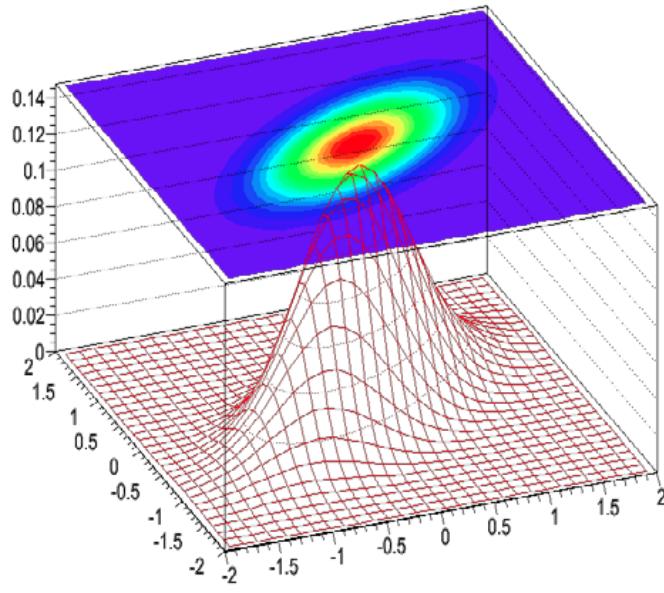
A FEW EXAMPLES (I)

In some cases, the parameter values that maximize the likelihood can be extracted analytically.

- easy example : estimate the distribution of a uni-variate sample with a Gaussian

In general, perform a numerical minimization on $-2\log\mathcal{L}$ (i.e. with Tminuit)

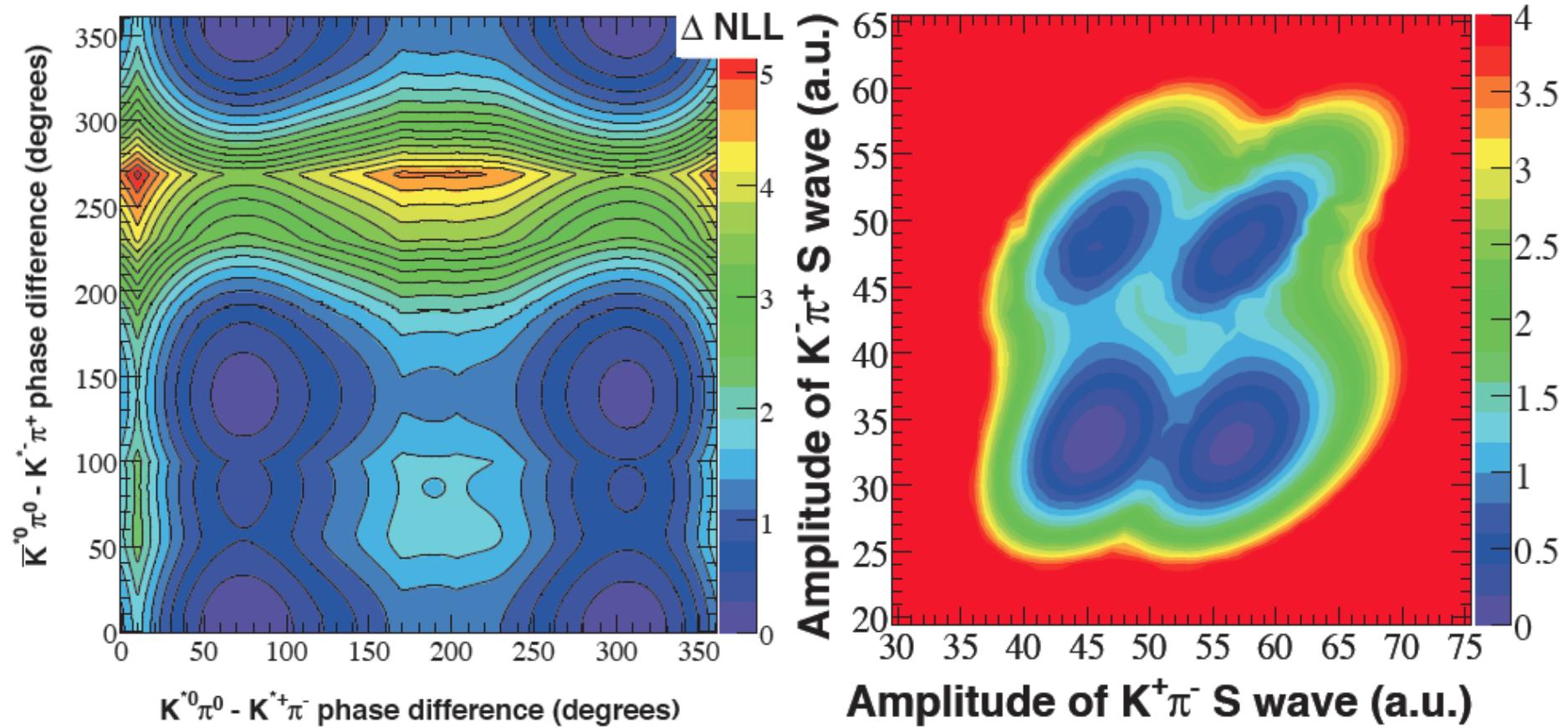
- If the Likelihood has a single maximum, convergence is efficient and safe
- even if the maximum is not exactly parabolic, provide $-2\Delta\log\mathcal{L}$ contours



- beware of ambiguous situations , i.e. local maxima or non-trivial boundaries

A FEW EXAMPLES (II)

- beware of ambiguous situations , i.e. local maxima or non-trivial boundaries



Amplitude analysis : trigonometric ambiguities + degenerate interference patterns
Four local maxima, roughly degenerate in $\min(\log \mathcal{L})$...

A FEW EXAMPLES (III)

- beware of ambiguous situations , i.e. local maxima or non-trivial boundaries

DALITZ PLOT ANALYSIS OF THE DECAY ...

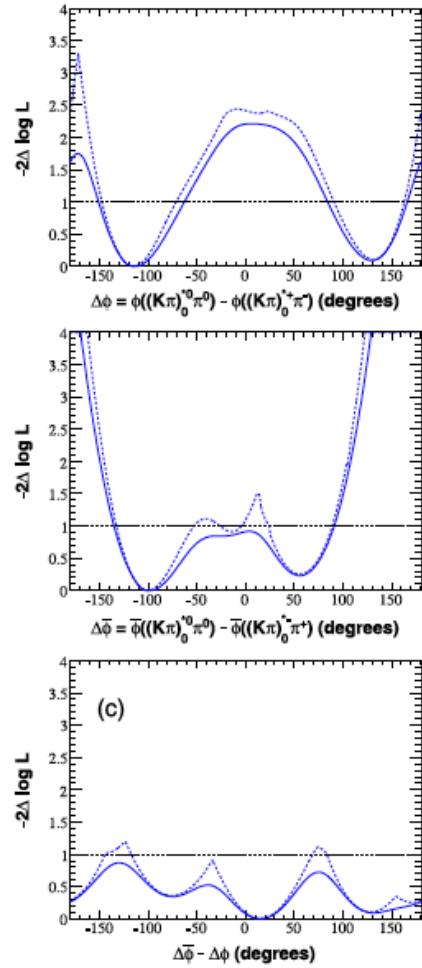


FIG. 11 (color online). The phase difference between the $(K\pi)_0^0$ and the $(K\pi)_0^{*\pm}$ S-waves. The three diagrams are the NLL scans for the B^0 (a) and B^0 (b) decays as well as their differences (c). The lines are drawn as in Fig. 10. The data do not indicate preferred angles. The four fit solutions find their NLL minimum at approximately the same phase differences.

PHYSICAL REVIEW D 78, 052005 (2008)

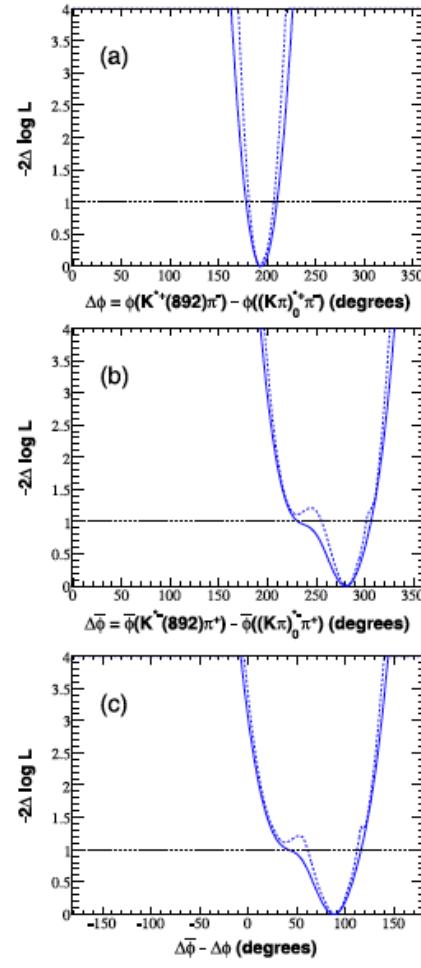


FIG. 12 (color online). The phase difference between the charged $K\pi$ P- and S-waves. The three diagrams are the NLL scans for the B^0 (a) and B^0 (b) decays as well as their differences (c). The lines are drawn as in Fig. 10. The data provide significant constraints on these angles. The four fit solutions find their NLL minimum at approximately the same phase differences.

B. AUBERT *et al.*

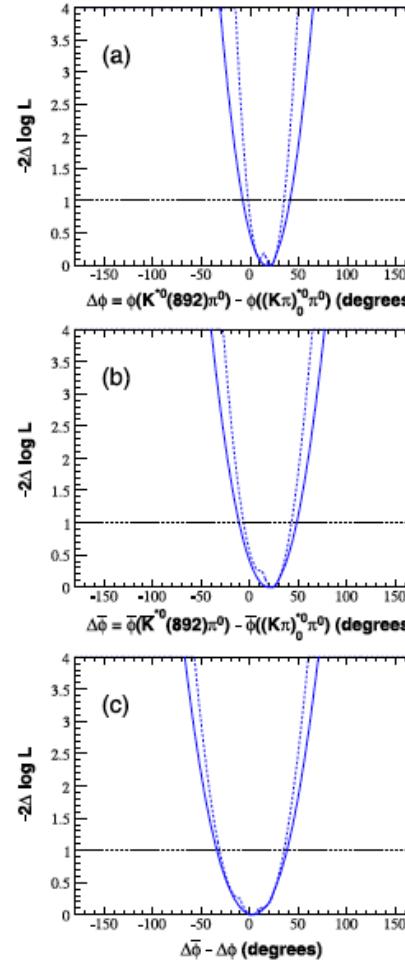


FIG. 13 (color online). The phase difference between the neutral $K\pi$ P- and S-waves. The three diagrams are the NLL scans for the B^0 (a) and B^0 (b) decays as well as their differences (c). The lines are drawn as in Fig. 10. The data provide significant constraints on these angles. The four fit solutions find their NLL minimum at approximately the same phase differences.

PHYSICAL REVIEW D 78, 052005 (2008)

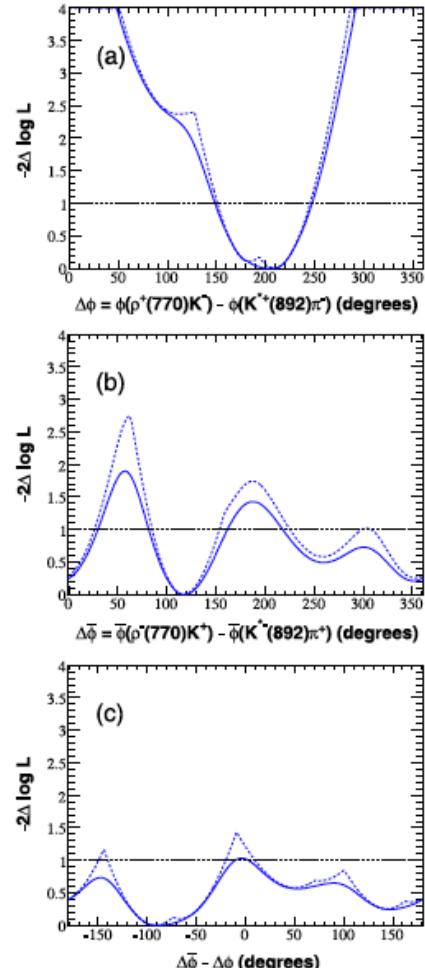


FIG. 14 (color online). The phase difference between the ρK and the charged $K\pi$ P-wave. The three diagrams are the NLL scans for the B^0 (a) and B^0 (b) decays as well as their differences (c). The lines are drawn as in Fig. 10. The vertical scale cuts off $\Delta\chi^2 = 4$; however, it has been checked that all phase differences are consistent with the data at the 3 standard deviation level.

A FEW EXAMPLES (IV)

- a Similar analysis : two degenerate solutions

TIME-DEPENDENT AMPLITUDE ANALYSIS OF ...

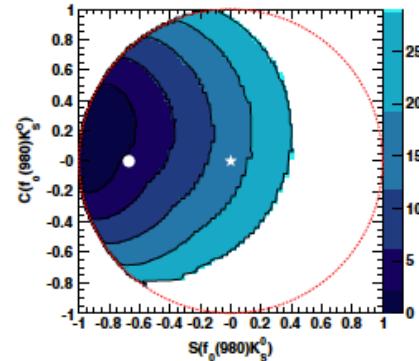


FIG. 10 (color online). Two-dimensional scans of $-2\Delta \log \mathcal{L}$ as a function of (S, C) , for the $f_0(980)K_s^0$ (left) and $\rho^0(770)K_s^0$ (right) isobar components. The value $-\Delta \log \mathcal{L}$ is computed including systematic uncertainties. Shaded areas, from the darkest to the lightest, represent the one to five standard deviations contours. The \bullet (\star) marks the expectation based on the current world average from $b \rightarrow c\bar{c}s$ modes [3] (zero point). The dashed circle represents the physical border $S^2 + C^2 = 1$.

sentation has nonetheless the advantage of allowing direct comparison with the measurement of $\sin 2\beta$ and C in $b \rightarrow c\bar{c}s$ modes. For $f_0(980)K_s^0$, the results agree with the expectation based on $b \rightarrow c\bar{c}s$ to 1.1σ ; for $\rho^0(770)K_s^0$ the agreement is better than 1σ . For the measured values of (β_{eff}, C) for $f_0(980)K_s^0$, CP conservation is excluded at 3.5σ . For $\rho^0(770)K_s^0$, the measurement of (β_{eff}, C) is consistent with CP conservation within 1σ .

The measurement of the phase $\Delta\Phi(K^{*+}(892)\pi^-)$ is presented as a one-dimensional likelihood scan in Fig. 11. For this flavor-specific mode, there is virtually no region in phase space that is accessible both to B^0 and

\bar{B}^0 , thus, sensitivity to this phase difference is limited. Simulation shows that interference of the $K^{*+}(892)\pi^-$ with the $f_0(980)K_s^0$ and $\rho^0(770)K_s^0$ modes (for which B^0 and \bar{B}^0 amplitudes interfere via mixing) provides most of the sensitivity to $\Delta\Phi(K^{*+}(892)\pi^-)$; unfortunately, the overlap in phase space of these resonances is small. As a consequence, only the $[-137, -5]^\circ$ interval is excluded at 95% confidence level. Figure 11 also shows the measurement of the similar phase difference for the $(K\pi)_s^0$ component. As for $K^{*}(892)$, the measurement sets no strong constraint on this phase. Only the interval $[-132, +25]^\circ$ is excluded at 95% confidence level.

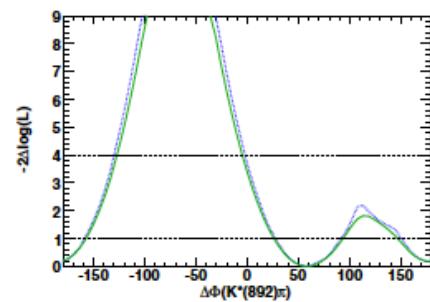


FIG. 11 (color online). Statistical (dashed line) and total (solid line) scans of $-2\Delta \log \mathcal{L}$ as a function of the relative phases $\Delta\Phi(K^{*}(892)\pi)$ (left) and $\Delta\Phi((K\pi)_s^0)$ (right). Horizontal dotted lines mark the one and two standard deviation levels.

B. AUBERT *et al.*

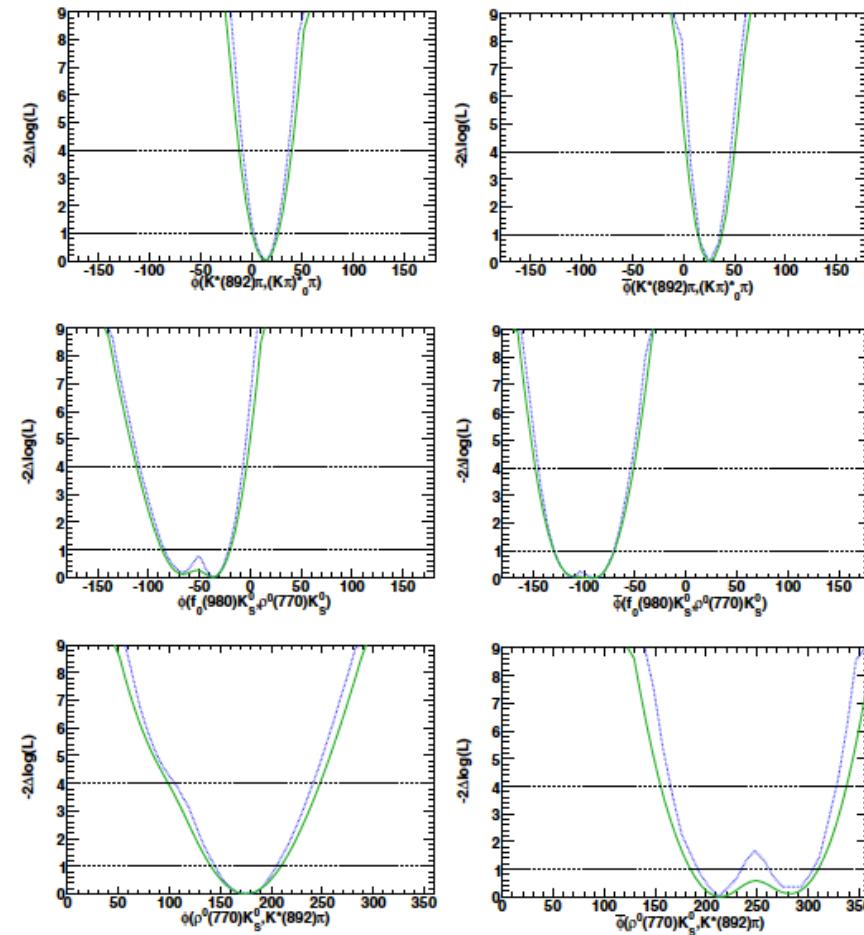


FIG. 12 (color online). Statistical (dashed line) and total (solid line) scans of $-2\Delta \log \mathcal{L}$ as a function of the phase differences $\phi(K^{*}(892)\pi, (K\pi)_s^0)$ (top), $\phi(f_0(980)K_s^0, \rho^0(770)K_s^0)$ (middle), and $\phi(\rho^0(770)K_s^0, K^{*}(892)\pi)$ (bottom). The left (right) column shows $B^0(\bar{B}^0)$ candidates. Horizontal dotted lines mark the one and two standard deviation levels.

A FEW EXAMPLES (V)

- a Similar analysis : two degenerate solutions : provide all info!

B. AUBERT *et al.*

PHYSICAL REVIEW D **80**, 112001 (2009)

TABLE VIII. Full correlation matrix for the isobar parameters of solution I. The entries are given in percent. Since the matrix is symmetric, all elements above the diagonal are omitted.

		c				\bar{c}										
		p^0	K^*	S	f_2	f_X	NR	χ	f_0	p^0	K^*	S	f_2	f_X	NR	χ
c	p^0	100.0														
	K^*	51.9	100.0													
	S	54.0	65.0	100.0												
	f_2	8.4	2.8	21.0	100.0											
	f_X	14.9	23.2	32.2	22.7	100.0										
	NR	5.2	35.0	24.4	12.6	39.3	100.0									
	χ	6.4	9.9	7.8	2.0	7.4	6.1	100.0								
\bar{c}	f_0	31.3	30.3	39.9	25.2	36.7	31.3	8.0	100.0							
	p^0	20.6	48.6	51.2	8.0	27.7	27.5	5.6	17.3	100.0						
	K^*	44.7	73.5	56.3	-4.8	24.9	22.0	9.5	22.6	43.4	100.0					
	S	59.6	71.9	79.7	21.8	39.3	26.9	11.3	35.2	49.4	57.7	100.0				
	f_2	2.4	-10.1	6.3	-56.1	-1.5	3.9	-0.3	10.7	-6.2	-21.5	5.0	100.0			
	f_X	14.5	34.1	12.5	16.1	-23.0	12.4	2.5	34.5	7.3	8.3	12.9	-6.2	100.0		
	NR	17.8	57.6	41.7	12.7	10.1	49.7	2.4	40.0	32.1	25.0	31.7	7.5	46.2	100.0	
	χ	18.9	27.0	30.6	5.8	11.8	9.5	-84.2	21.5	17.8	24.1	27.8	0.8	8.1	202	100.0
arg(c)	p^0	-11.2	13.3	4.0	-16.1	-2.9	-2.1	-0.5	-0.2	24.1	16.3	3.2	-3.3	8.9	2.1	4.2
	K^*	25.0	8.6	-3.2	-0.2	-15.7	-9.7	6.3	-10.4	-3.9	5.5	16.0	3.8	6.3	-6.5	-3.2
	S	33.0	19.6	3.4	-4.7	-17.3	-16.5	6.2	-9.6	1.0	18.7	21.3	-4.2	9.6	-4.2	1.1
	f_2	12.1	-0.6	-9.8	-2.6	-23.1	-27.4	0.9	-16.7	-7.2	2.2	1.1	-10.6	7.2	-14.1	-2.6
	f_X	25.0	10.2	5.4	-0.5	-11.4	-11.8	1.0	-0.8	2.6	8.5	11.8	-3.8	15.6	2.4	0.4
	NR	31.6	17.0	39.3	1.0	-27.1	-31.7	-6.7	11.3	12.8	14.5	19.0	3.3	21.5	19.6	14.2
	χ	8.6	1.8	9.8	0.6	-9.9	-8.9	-7.9	2.8	3.8	1.3	4.2	3.5	7.3	8.9	12.4
arg(\bar{c})	f_0	32.2	11.7	18.9	3.5	-20.3	-26.2	-1.6	-3.6	-6.9	7.3	18.2	1.8	20.3	-7.1	4.3
	p^0	14.5	18.0	14.6	-17.3	-13.4	-21.0	-0.7	-8.7	143	19.8	13.4	1.7	7.2	-4.4	5.4
	K^*	17.1	7.1	22.0	5.2	-13.5	-17.3	-2.1	5.0	7.2	6.5	13.8	8.1	12.8	29.5	9.6
	S	22.5	15.9	25.2	-3.2	-16.9	-21.6	-0.5	4.2	106	17.7	16.1	1.7	14.1	28.8	10.8
	f_2	15.1	4.9	15.5	-5.0	-15.5	-17.9	-2.1	10.0	-2.5	3.9	2.9	11.1	15.7	18.6	7.5
	f_X	8.1	2.7	12.3	-0.6	16.5	-20.4	-0.9	12.2	6.1	3.4	4.8	1.4	-14.6	4.7	6.5
	NR	15.3	4.1	14.5	-3.0	-22.6	-20.8	0.8	1.7	8.2	1.8	5.2	2.6	20.0	15.1	3.2
	χ	10.9	1.1	12.8	0.7	-13.9	-18.0	-4.7	2.1	3.3	0.6	3.9	5.9	9.8	13.4	8.2

		$\arg(c)$				$\arg(\bar{c})$										
		p^0	K^*	S	f_2	f_X	NR	χ	f_0	p^0	K^*	S	f_2	f_X	NR	χ
arg(c)	p^0	100.0														
	K^*	10.4	100.0													
	S	18.2	90.9	100.0												
	f_2	19.6	54.1	61.8	100.0											
	f_X	25.5	49.3	56.9	58.1	100.0										
	NR	24.3	17.2	29.9	31.6	47.8	100.0									
	χ	5.0	6.7	7.9	10.2	17.6	30.8	100.0								
arg(\bar{c})	f_0	18.0	34.3	42.0	39.8	52.9	55.6	23.8	100.0							
	p^0	55.3	22.2	32.4	25.7	36.6	42.2	17.4	58.8	100.0						
	K^*	4.0	21.5	28.0	23.2	36.1	53.9	31.3	46.8	33.5	100.0					
	S	9.6	23.7	35.1	27.8	41.2	60.7	33.3	53.4	42.7	90.9	100.0				
	f_2	5.5	6.4	12.4	1.5	29.3	46.4	23.5	44.1	36.7	56.7	60.8	100.0			
	f_X	1.7	0.0	5.4	13.8	15.5	36.4	19.5	22.2	22.5	42.1	44.8	39.4	100.0		
	NR	7.2	19.2	27.5	28.9	42.3	55.5	32.9	47.3	37.9	63.2	72.5	48.1	48.4	100.0	
	χ	4.1	8.9	13.3	15.5	27.1	43.3	35.9	38.0	26.9	55.9	58.9	40.0	33.6	52.1	100.0

TIME-DEPENDENT AMPLITUDE ANALYSIS OF ...

PHYSICAL REVIEW D **80**, 112001 (2009)

TABLE IX. Full correlation matrix for the isobar parameters of solution II. The entries are given in percent. Since the matrix is symmetric, all elements above the diagonal are omitted.

		c				\bar{c}										
		p^0	K^*	S	f_2	f_X	NR	χ	f_0	p^0	K^*	S	f_2	f_X	NR	χ
c	p^0	100.0														
	K^*	46.9	100.0													
	S	49.1	68.2	100.0												
	f_2	8.7	7.7	25.4	100.0											
	f_X	16.8	40.3	38.5	26.6	100.0										
	NR	-8.4	30.2	21.2	9.4	49.9	100.0									
	χ	5.5	11.7	9.3	3.4	12.1	9.1	100.0								
\bar{c}	f_0	29.2	42.1	50.2	31.5	57.9	34.1	10.0	100.0							
	p^0	61.5	68.1	40.4	6.9	20.6	6.4	6.0	31.6	100.0						
	K^*	39.8	75.7	59.8	0.3	33.1	25.3	10.9	33.2	36.3	100.0					
	S	50.6	75.2	83.2	25.4	49.9	33.4	13.1	51.6	46.0	61.4	100.0				
	f_2	0.8	-6.1	9.6	-53.9	60	13.3	0.2	14.7	5.3	-18.5	10.4	100.0			
	f_X	10.0	-3.3	-0.9	-10.6	-68.7	-17.8	-5.2	-18.4	6.3	-4.0	-4.9	2.2	100.0		
	NR	23.1	68.8	44.7	13.5	39.3	34.8	5.8	45.6	58.3	32.8	45.4	14.7	-13.8	100.0	
arg(c)	χ	-22.3	33.5	37.8	9.8	19.3	9.9	-79.2	31.3	20.7	30.2	36.1	3.3	-2.6	23.3	100.0
	p^0	30.6	2.0	-2.2	-6.3	-16.1	-28.1	-0.0	-15.2	14.3	-1.4	6.1	2.0	19.4	-10.3	0.5
	K^*	38.1	8.9	1.8	-10.1	-17.9	-39.5	-0.1	-15.8	17.4	9.4	7.7	-8.2	19.7	-12.1	3.7
	S	22.2	-7.8	-12.2	-5.9	-7.7	-35.9	-1.7	-14.5	7.8	-5.7	-8.8	-9.9	12.2	-15.2	-3.6
	f_2	18.1	-10.0	-13.7	-7.4	-15.4	-41.3	-2.4	-18.6	1.0	-6.2	-10.2	-12.7	10.7	-21.6	-2.7
	f_X	32.4	-0.4	21.4	0.5	-29.5	-65.2	-10.4	-4.2	12.0	0.2	0.4	-6.4	21.2	-7.0	4.0
	NR	15.4	-2.2	0.2	-1.6	-9.9	-18.3	-4.9	-5.6	5.6	-3.0	-0.2	-0.8	9.2	-5.6	4.0
arg(\bar{c})	f_0	30.1	-8.0	-2.3	-0.9	-13.2	-43.0	-2.8	-16.7	12.1	-7.2	-5.5	-4.9	10.4	-18.7	-1.6
	p^0	7.6	11.4	5.8	-7.5	-1.8	-24.7	0.6	-7.5	4.1	15.1	5.5	-12.6	1.3	-7.0	4.0
	K^*	27.0	0.8	7.6	5.6	28	-27.8	0.6	-2.0	9.1	1.5	7.1	3.2	-6.9	13.9	4.1
	S	32.6	8.0	8.4	-1.1	0.6	-31.3	2.1	-4.1	12.6	12.1	7.6	-5.6	-4.4	12.2	4.7
	f_2	18.7	1.7	6.6	10.1	9.8	-22.9	0.7	7.6	8.6	3.5	0.6	-5.6	-21.6	9.3	4.6
	f_X	21.9	1.8	4.4	9.6	-0.7	-30.2	0.1	-5.0	8.1	28	4.0	-17.3	1.0	-6.6	-0.2
	NR	27.7	-1.9	-3.0	3.9	-0.5	-30.7	2.8	-13.3	7.8	-1.5	-				

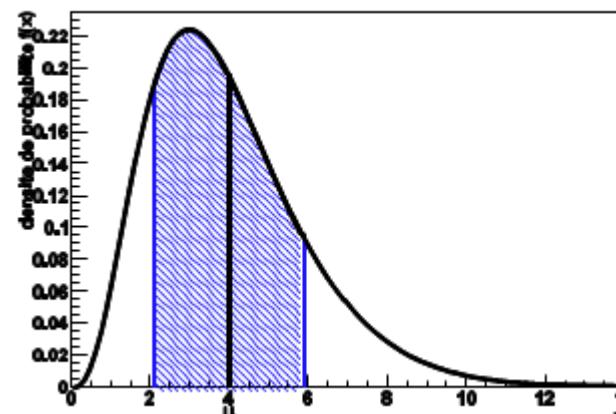
CONFIDENCE INTERVALS (I)

For an parameter θ , define a Confidence Interval with a given Confidence Level as

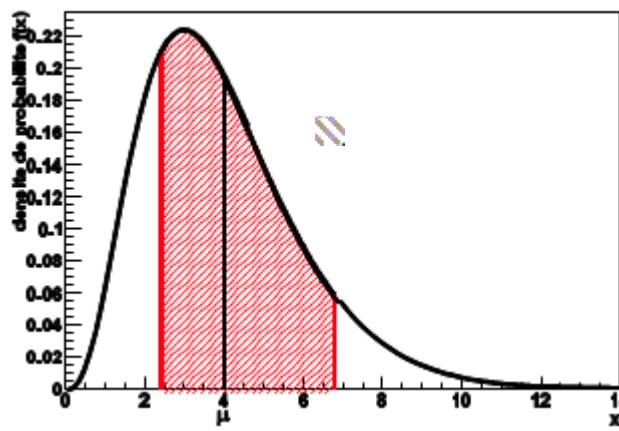
$$P(\theta_a \leq \hat{\theta} \leq \theta_b) = \alpha$$

A few remarks :

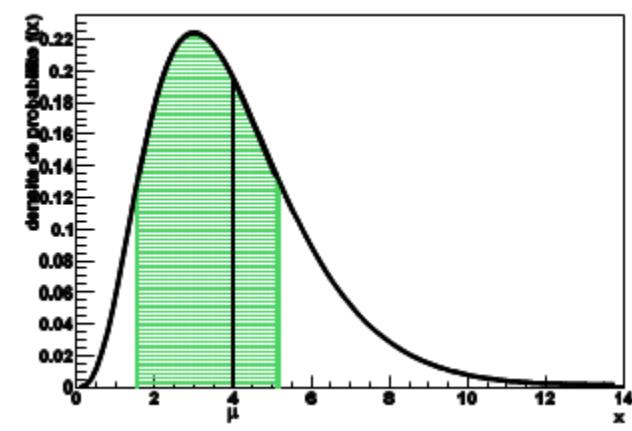
- the Interval boundaries θ_a and θ_b are random variables themselves : different sample realizations will yield different values
- an infinite number of intervals can correspond to a given α value. Examples :



mean-centered, range-symmetric



mean-centered, probability-symmetric

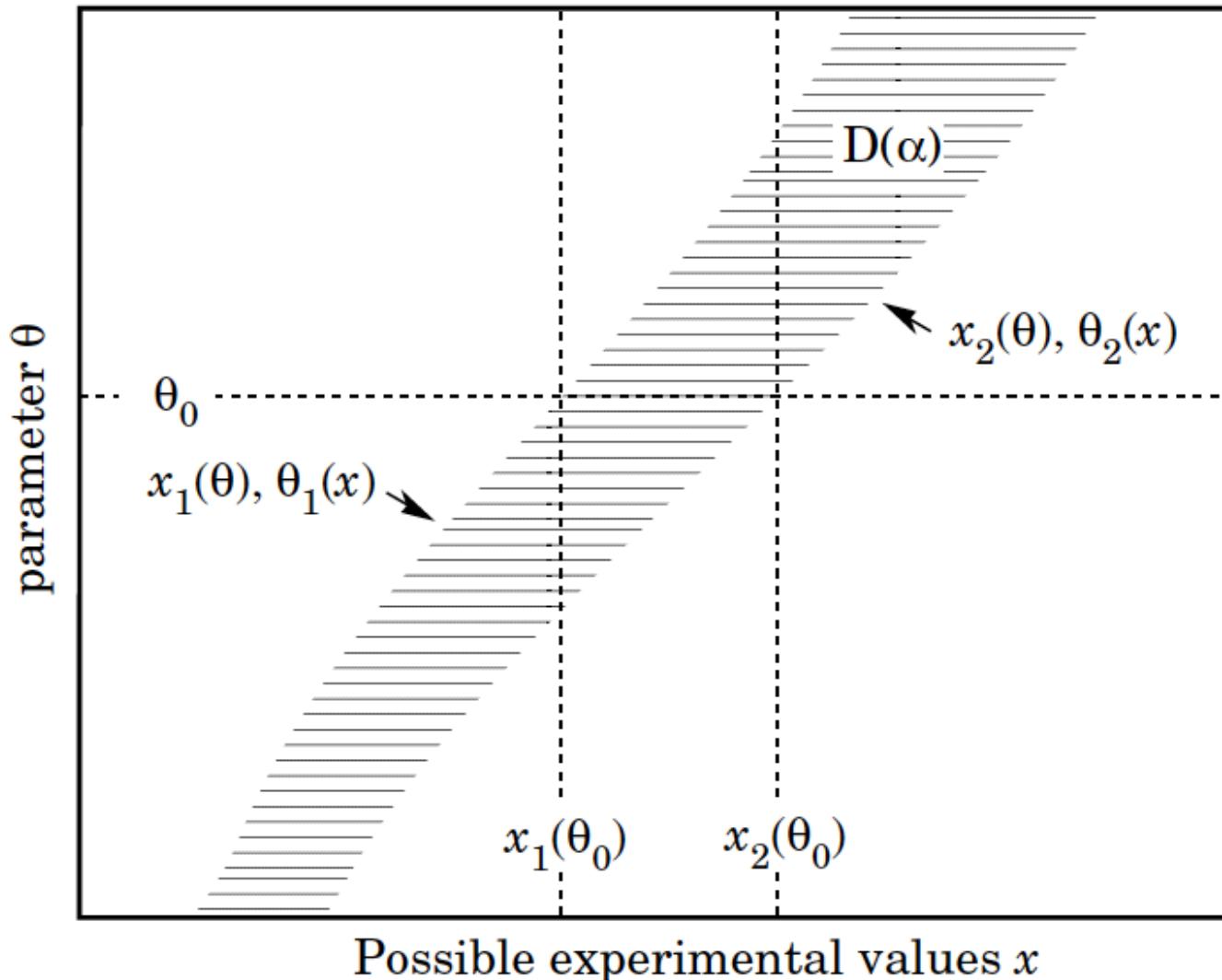


highest-probability

A classical example : the Neyman construction

CONFIDENCE INTERVALS (II)

- The Neyman construction : a confidence belt ensuring coverage



SYSTEMATIC UNCERTAINTIES

The Covariance from ML is an estimator of statistical uncertainties only

- how to deal with systematics?

Distinguish among two categories of parameters in the Likelihood :

$$\mathcal{L}(\mu_1, \dots, \mu_n, \theta_1, \dots, \theta_k)$$

- the μ 's are parameters of interest (PIO) : what you actually want to estimate
- the θ 's are nuisance parameters (NP) : potential sources of systematic biases

Various situations :

- “Type-I” : the sample (or a control) can constrain (some of) the nuisances
 - impact decreases with sample size
- “Type-II” : assumptions in the model, uncontrolled features in data...
 - often independent of sample size

Profile Likelihood : treatment of nuisance parameters as random variables,

$$\mathcal{L}(\mu, \theta, \nu) = \mathcal{L}_\theta(\mu, \theta) \mathcal{L}_\nu(\nu)$$

then minimize the Likelihood with respect to the nuisance parameters.

Clearly robust for Type-I nuisances : i.e. simultaneous fit to the control samples

PROFILE LIKELIHOOD

- Back to the previous example : list of POI's and NP's

TIME-DEPENDENT AMPLITUDE ANALYSIS OF ...

PHYSICAL REVIEW D 80, 112001 (2009)

TABLE V. Summary of measurements of the Q2B parameters for solutions I and II. The first uncertainty is statistical, the second is systematic, and the third represents the DP signal model dependence. We also show the total (statistical and systematic) linear correlations between the parameters β_{eff} (S) and C . Phases are given in degrees and FFs in percent.

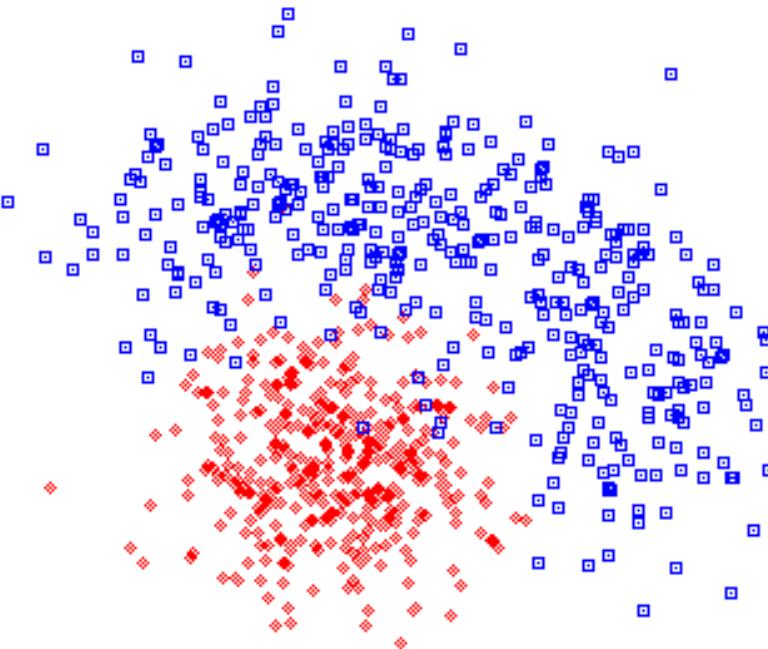
Parameter	Solution I	Solution II
$C(f_0(980)K_s^0)$	$0.08 \pm 0.19 \pm 0.03 \pm 0.04$	$0.23 \pm 0.19 \pm 0.03 \pm 0.04$
$\beta_{\text{eff}}(f_0(980)K_s^0)$	$36.0 \pm 9.8 \pm 2.1 \pm 2.1$	$56.2 \pm 10.4 \pm 2.1 \pm 2.1$
$S(f_0(980)K_s^0)$	$-0.96^{+0.21}_{-0.21} \pm 0.03 \pm 0.02$	$-0.90^{+0.26}_{-0.26} \pm 0.03 \pm 0.02$
$\text{Corr}[\beta_{\text{eff}}(f_0(980)K_s^0), C(f_0(980)K_s^0)]$	-3.1%	-17.0%
$\text{Corr}[S(f_0(980)K_s^0), C(f_0(980)K_s^0)]$	19.7%	12.5%
$FF(f_0(980)K_s^0)$	$13.8^{+1.4}_{-1.4} \pm 0.8 \pm 0.6$	$13.5^{+1.4}_{-1.3} \pm 0.8 \pm 0.6$
$C(\rho^0(770)K_s^0)$	$-0.05 \pm 0.26 \pm 0.10 \pm 0.03$	$-0.14 \pm 0.26 \pm 0.10 \pm 0.03$
$\beta_{\text{eff}}(\rho^0(770)K_s^0)$	$10.2 \pm 8.9 \pm 3.0 \pm 1.9$	$33.4 \pm 10.4 \pm 3.0 \pm 1.9$
$S(\rho^0(770)K_s^0)$	$0.35^{+0.24}_{-0.21} \pm 0.06 \pm 0.03$	$0.91^{+0.19}_{-0.19} \pm 0.06 \pm 0.03$
$\text{Corr}[\beta_{\text{eff}}(\rho^0(770)K_s^0), C(\rho^0(770)K_s^0)]$	-23.0%	-34.0%
$\text{Corr}[S(\rho^0(770)K_s^0), C(\rho^0(770)K_s^0)]$	-21.3%	-10.4%
$FF(\rho^0(770)K_s^0)$	$8.6^{+1.4}_{-1.4} \pm 0.5 \pm 0.2$	$8.5^{+1.3}_{-1.3} \pm 0.5 \pm 0.2$
$A_{CP}(K^*(892)\pi)$	$-0.21 \pm 0.10 \pm 0.01 \pm 0.02$	$-0.19^{+0.10}_{-0.11} \pm 0.01 \pm 0.02$
$\Delta\Phi(K^*(892)\pi)$	$58.3 \pm 32.7 \pm 4.6 \pm 8.1$	$176.6 \pm 28.8 \pm 4.6 \pm 8.1$
$FF(K^*(892)\pi)$	$11.0^{+1.0}_{-1.0} \pm 0.6 \pm 0.8$	$10.9^{+1.0}_{-1.0} \pm 0.6 \pm 0.8$
$A_{CP}(K\pi^*_0\pi)$	$0.09 \pm 0.07 \pm 0.02 \pm 0.02$	$0.12^{+0.07}_{-0.06} \pm 0.02 \pm 0.02$
$\Delta\Phi(K\pi^*_0\pi)$	$72.2 \pm 24.6 \pm 4.1 \pm 4.4$	$-175.1 \pm 22.6 \pm 4.1 \pm 4.4$
$FF(K\pi^*_0\pi)$	$45.2 \pm 2.3 \pm 1.9 \pm 0.9$	$46.1 \pm 2.4 \pm 1.9 \pm 0.9$
$C(f_2(1270)K_s^0)$	$0.28^{+0.21}_{-0.20} \pm 0.08 \pm 0.07$	$0.09 \pm 0.46 \pm 0.08 \pm 0.07$
$\beta_{\text{eff}}(f_2(1270)K_s^0)$	$14.9 \pm 17.9 \pm 3.1 \pm 5.2$	$53.6 \pm 16.7 \pm 3.1 \pm 5.2$
$S(f_2(1270)K_s^0)$	$-0.48 \pm 0.52 \pm 0.06 \pm 0.10$	$-0.95 \pm 0.17 \pm 0.06 \pm 0.10$
$\text{Corr}[\beta_{\text{eff}}(f_2(1270)K_s^0), C(f_2(1270)K_s^0)]$	11.5%	-2.8%
$\text{Corr}[S(f_2(1270)K_s^0), C(f_2(1270)K_s^0)]$	9.9%	21.2%
$FF(f_2(1270)K_s^0)$	$2.3^{+0.5}_{-0.7} \pm 0.2 \pm 0.7$	$2.3^{+0.9}_{-0.7} \pm 0.2 \pm 0.7$
$C(f_3(1300)K_s^0)$	$0.13^{+0.33}_{-0.33} \pm 0.04 \pm 0.09$	$0.30^{+0.34}_{-0.34} \pm 0.04 \pm 0.09$
$\beta_{\text{eff}}(f_3(1300)K_s^0)$	$5.8 \pm 15.2 \pm 2.2 \pm 2.3$	$76.9 \pm 13.8 \pm 2.2 \pm 2.3$
$S(f_3(1300)K_s^0)$	$-0.20 \pm 0.52 \pm 0.07 \pm 0.07$	$-0.42 \pm 0.41 \pm 0.07 \pm 0.07$
$\text{Corr}[\beta_{\text{eff}}(f_3(1300)K_s^0), C(f_3(1300)K_s^0)]$	-27.0%	-9.3%
$\text{Corr}[S(f_3(1300)K_s^0), C(f_3(1300)K_s^0)]$	28.5%	6.1%
$FF(f_3(1300)K_s^0)$	$3.6^{+1.0}_{-0.9} \pm 0.3 \pm 0.9$	$3.5^{+1.0}_{-0.9} \pm 0.3 \pm 0.9$
$C(NR)$	$0.01 \pm 0.25 \pm 0.06 \pm 0.05$	$-0.45^{+0.30}_{-0.30} \pm 0.06 \pm 0.05$
$\beta_{\text{eff}}(NR)$	$0.4 \pm 8.8 \pm 1.9 \pm 3.8$	$51.0 \pm 13.3 \pm 1.9 \pm 3.8$
$S(NR)$	$-0.01 \pm 0.31 \pm 0.05 \pm 0.09$	$-0.87 \pm 0.18 \pm 0.05 \pm 0.09$
$\text{Corr}[\beta_{\text{eff}}(NR), C(NR)]$	10.6%	-37.9%
$\text{Corr}[S(NR), C(NR)]$	10.6%	-91.5%
$FF(NR)$	$11.5 \pm 2.0 \pm 1.0 \pm 0.6$	$12.6 \pm 2.0 \pm 1.0 \pm 0.6$
$C(\chi_{c0}K_s^0)$	$-0.29^{+0.53}_{-0.64} \pm 0.03 \pm 0.05$	$-0.41^{+0.54}_{-0.42} \pm 0.03 \pm 0.05$
$\beta_{\text{eff}}(\chi_{c0}K_s^0)$	$23.2 \pm 22.4 \pm 2.3 \pm 4.2$	$55.2 \pm 23.3 \pm 2.3 \pm 4.2$
$S(\chi_{c0}K_s^0)$	$-0.69 \pm 0.52 \pm 0.04 \pm 0.07$	$-0.85 \pm 0.34 \pm 0.04 \pm 0.07$
$\text{Corr}[\beta_{\text{eff}}(\chi_{c0}K_s^0), C(\chi_{c0}K_s^0)]$	-5.8%	-5.8%
$\text{Corr}[S(\chi_{c0}K_s^0), C(\chi_{c0}K_s^0)]$	-19.1%	-74.2%
$FF(\chi_{c0}K_s^0)$	$1.04^{+0.41}_{-0.33} \pm 0.04 \pm 0.11$	$0.99^{+0.37}_{-0.30} \pm 0.04 \pm 0.11$
total FF	$97.2^{+1.7}_{-1.7} \pm 2.1 \pm 1.15$	$98.3^{+1.7}_{-1.7} \pm 2.1 \pm 1.15$
A_{eff}^{1D}	$-0.01 \pm 0.08 \pm 0.01 \pm 0.01$	$0.01 \pm 0.05 \pm 0.01 \pm 0.01$
$\phi(f_0(980)K_s^0, \rho(770)K_s^0)$	$-35.6 \pm 14.9 \pm 6.1 \pm 4.4$	$-66.7 \pm 18.3 \pm 6.1 \pm 4.4$
$\phi(K^*(892)\pi, (K\pi)_0^*\pi)$	$13.0 \pm 10.9 \pm 4.6 \pm 4.7$	$15.5 \pm 10.2 \pm 4.6 \pm 4.7$
$\phi(\rho(770)K_s^0, K^*(892)\pi)$	$174.3 \pm 28.0 \pm 8.7 \pm 12.7$	$-173.7 \pm 29.8 \pm 8.7 \pm 12.7$
$\phi(\rho(770)K_s^0, (K\pi)_0^*\pi)$	$-172.8 \pm 22.6 \pm 10.1 \pm 8.7$	$-170.8 \pm 26.8 \pm 10.1 \pm 8.7$
$\phi(f_0(980)K_s^0, \rho(770)K_s^0)$	$-89.2 \pm 17.1 \pm 8.5 \pm 7.2$	$-112.2 \pm 17.8 \pm 8.5 \pm 7.2$
$\phi(K^*(892)\pi, (K\pi)_0^*\pi)$	$26.9 \pm 9.2 \pm 4.9 \pm 6.1$	$23.8 \pm 9.1 \pm 4.9 \pm 6.1$
$\phi(\rho(770)K_s^0, (K\pi)_0^*\pi)$	$-147.8 \pm 24.7 \pm 11.3 \pm 11.9$	$-76.5 \pm 24.0 \pm 11.3 \pm 11.9$
$\phi(\rho(770)K_s^0, (K\pi)_0^*\pi)$	$-120.9 \pm 21.6 \pm 8.7 \pm 7.3$	$-52.7 \pm 21.4 \pm 8.7 \pm 7.3$

Parameter Name	Fit Result Sol-I	Fit Result Sol-II
ΔNLL	0.0	0.16
$N(B^0 \rightarrow D^+ \pi^-)$	3361 ± 60	3362 ± 60
$N(B^0 \rightarrow J/\Psi K_s^0)$	1804 ± 44	1803 ± 43
$N(B^0 \rightarrow \eta' K_s^0)$	46 ± 16	44 ± 16
$N(B^0 \rightarrow \Psi(2S)K_s^0)$	142 ± 13	142 ± 13
$N(\text{cont-Lepton})$	46 ± 8.9	47 ± 9
$N(\text{cont-KaonI})$	800 ± 31	800 ± 31
$N(\text{cont-KaonII})$	2127 ± 49	2127 ± 49
$N(\text{cont-KaonPion})$	1775 ± 45	1775 ± 45
$N(\text{cont-Pion})$	2048 ± 48	2048 ± 48
$N(\text{cont-Other})$	1614 ± 42	1614 ± 42
$N(\text{cont-NoTag})$	5829 ± 80	5829 ± 80
$f_{\text{core}}(\Delta E)$ Signal	0.63 ± 0.14	0.63 ± 0.14
$\mu_{\text{core}}(\Delta E)$ Signal	$-1.3 \pm 0.7 \text{ MeV}$	$-1.3 \pm 0.6 \text{ MeV}$
$\sigma_{\text{core}}(\Delta E)$ Signal	$17.1 \pm 1.4 \text{ MeV}$	$17.1 \pm 1.3 \text{ MeV}$
$\mu_{\text{tail}}(\Delta E)$ Signal	$-7.3 \pm 2.9 \text{ MeV}$	$-7.4 \pm 3.0 \text{ MeV}$
$\sigma_{\text{tail}}(\Delta E)$ Signal	$31.2 \pm 4.6 \text{ MeV}$	$31.4 \pm 4.6 \text{ MeV}$
Slope(ΔE) Continuum	-8.51 ± 5.77	-8.49 ± 5.77
$\mu(m_{ES})$ Signal	$5.2788 \pm 0.0001 \text{ GeV}/c^2$	$5.2788 \pm 0.0001 \text{ GeV}/c^2$
$\sigma_L(m_{ES})$ Signal	$2.24 \pm 0.06 \text{ MeV}/c^2$	$2.24 \pm 0.06 \text{ MeV}/c^2$
$\sigma_R(m_{ES})$ Signal	$2.73 \pm 0.07 \text{ MeV}/c^2$	$2.73 \pm 0.07 \text{ MeV}/c^2$
Argus Slope(m_{ES}) Continuum	-0.3 ± 0.2	-0.4 ± 0.2
$a_1(NN)$ Continuum	1.9 ± 0.1	1.9 ± 0.1
$a_2(NN)$ Continuum	3.2 ± 0.4	3.2 ± 0.4
$a_3(NN)$ Continuum	-1.1 ± 0.1	-1.1 ± 0.1
$a_5(NN)$ Continuum	-0.47 ± 0.05	-0.48 ± 0.05
$\mu_{\text{common}}(\Delta t)$ Continuum	$0.018 \pm 0.007 \text{ ps}$	$0.018 \pm 0.007 \text{ ps}$
$\sigma_{\text{core}}(\Delta t)$ Continuum	$1.14 \pm 0.02 \text{ ps}$	$1.14 \pm 0.02 \text{ ps}$
$\sigma_{\text{tail}}(\Delta t)$ Continuum	0.16 ± 0.02	0.16 ± 0.02
$f_{\text{outlier}}(\Delta t)$ Continuum	$2.8 \pm 0.2 \text{ ps}$	$2.8 \pm 0.2 \text{ ps}$
$\sigma_{\text{outlier}}(\Delta t)$ Continuum	0.030 ± 0.004	0.030 ± 0.004

- used to look impressive...
- ... but is now significantly smaller than the workspaces for the Higgs !

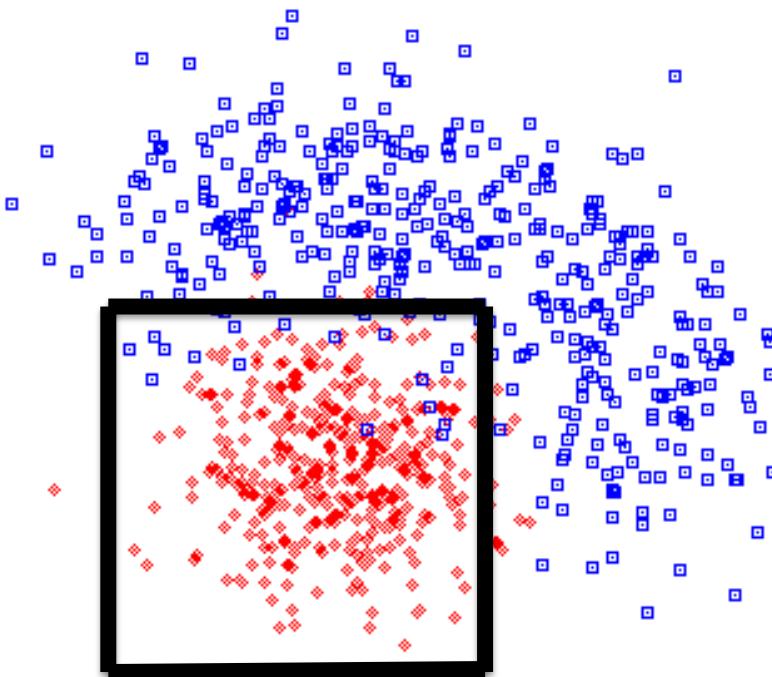
EVENT SELECTION :

- the ML estimation requires control of PDFs over the complete r.v. space
 - moreover, a large sample size is CPU-expensive for the minimization
- it is often useful (even unavoidable) to reduce the sample size
 - restrict ML estimation to “signal-enriched” subsets of r.v. space :



EVENT SELECTION :

- the ML estimation requires control of PDFs over the complete r.v. space
 - moreover, a large sample size is CPU-expensive for the minimization
- it is often useful (even unavoidable) to reduce the sample size
 - restrict ML estimation to “signal-enriched” subsets of r.v. space :

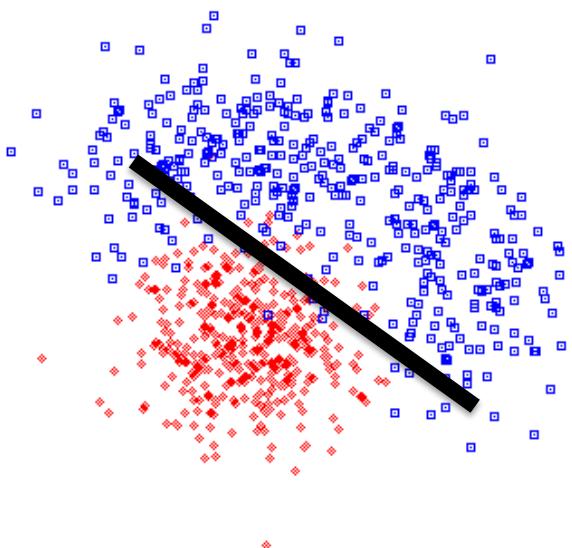


Cut-based selection :

- simplest method
- certainly applied at early stages
 - trigger, offline filters...

EVENT SELECTION :

- the ML estimation requires control of PDFs over the complete r.v. space
 - moreover, a large sample size is CPU-expensive for the minimization
- it is often useful (even unavoidable) to reduce the sample size
 - restrict ML estimation to “signal-enriched” subsets of r.v. space :



Linear decision boundary :

boundary definition can be “optimized” :

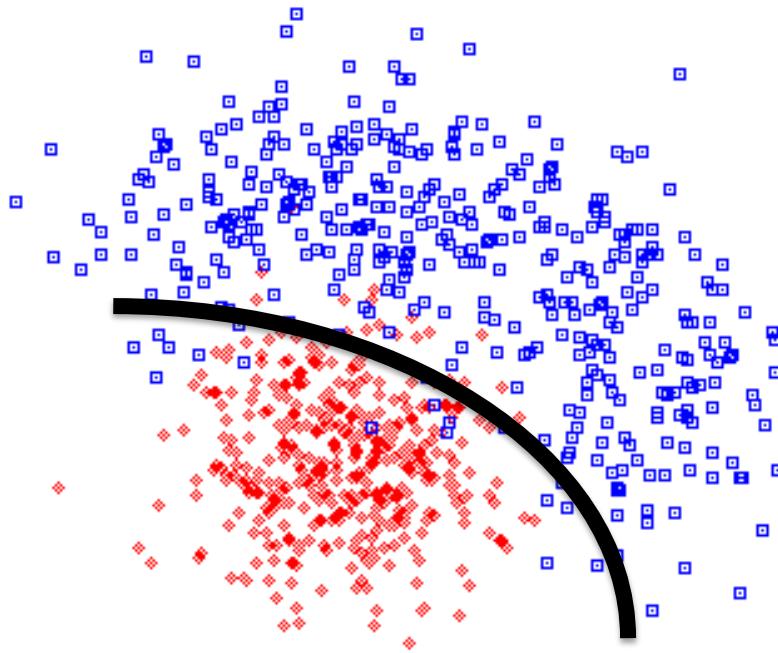
- for example, maximize the *separation*

$$\langle s^2 \rangle = \frac{1}{2} \iint dx dy \frac{[S(x,y) - B(x,y)]^2}{S(x,y) + B(x,y)}$$

- Fisher linear discriminant: optimal usage of linear correlations

EVENT SELECTION :

- the ML estimation requires control of PDFs over the complete r.v. space
 - moreover, a large sample size is CPU-expensive for the minimization
- it is often useful (even unavoidable) to reduce the sample size
 - restrict ML estimation to “signal-enriched” subsets of r.v. space :



Non-linear boundary decisions :
a large variety of tools !

- Neural Networks,
- Likelihood-based,
- Vector Machines,
- Genetic Algorithms,
- (Boosted) Decision Trees ...

NON-LINEAR MVA'S : INCOMPLETE SURVEY

- Likelihood-based MVA : First rotate to uncorrelated variables

$$\vec{x}' = R\vec{x} \quad , \quad C[x'_i, x'_j] = \delta_{ij}\sigma'_i$$

then build the product of 1-dimensional PDFs

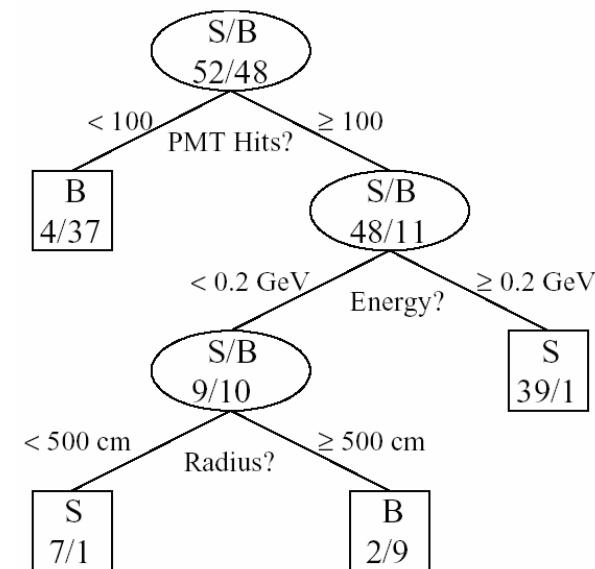
$$f(x) = \prod_i \mathcal{P}_i(x'_i)$$

Does not exploit non-linear, yet performances should be comparable/superior than a linear discriminant...

- Decision tree : training samples are split by successive cuts, until benchmark stop criteria are met.

Often sensitive to fluctuations

- Boosted decision tree: improves stability with respect to fluctuations in training samples.



MULTIVARIATE TECHNIQUES

Very user-friendly, general-purpose, MVA builder tools available : easy!

Beware of potential uncontrollable effects :

- relevance of control samples ,
- under- or over- training ,
- sensitivity to irrelevant information ,
- ...

... make sure you know
which info is being used
by the MVA!

... get a feeling how it is
using it!

... build your list of checks !

