Confidence Regions in 2D with MINUIT MINOS, Numerical Recipes, Feldman-Cousins, and G.O.F.; and use of CL_s to exclude regions in 2D

Bob Cousins, UCLA

LL Statistics Workshop

14 February 2013

N.B.

Much of this talk is well-known to this group, but I hope is can become better known to others.

I will try to go through the more familiar material quickly.

MINUIT MINOS intervals and regions

Computer Physics Communications 10 (1975) 343–367

MINUIT – A SYSTEM FOR FUNCTION MINIMIZATION AND ANALYSIS OF THE PARAMETER ERRORS AND CORRELATIONS

F. JAMES

Data Handling Division, European Organization for Nuclear Research, CH-1211 Geneva 23, Switzerland

and

M. ROOS

Department of Nuclear Physics, University of Helsinki, Sf-00170 Helsinki, Finland

" MINUIT has been under continuous development at CERN since 1967. During this time the program has had enormous exposure to users at CERN and elsewhere. "

MINOS errors from $\Delta(-\ln\mathcal{L})$

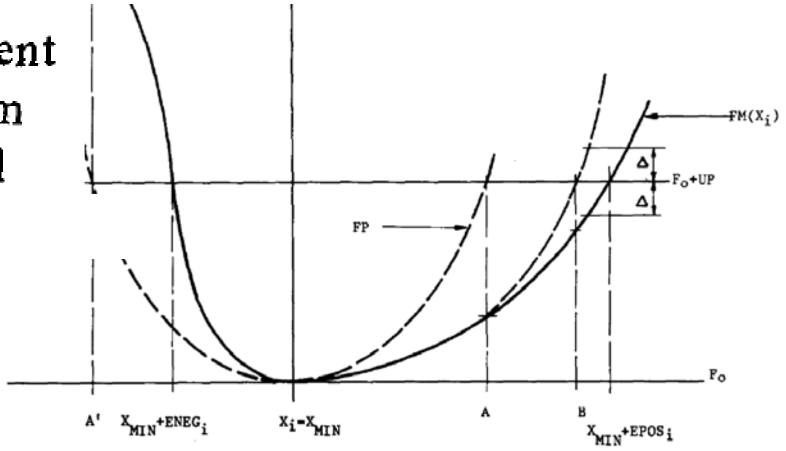


Fig. 2. Calculation of MINOS errors of parameter *i*. The (symmetric) dotted parabola FP is predicted from the covariance matrix, but the nonlinearity of the problem results in the solid curve FM which gives the asymmetric errors EPOS and ENEG (see text).

(Approximate) Confidence Regions Using Δ(-ln$\mathcal{L}$)

(included in appendix to MINUIT users guide)

Computer Physics Communications 20 (1980) 29–35

INTERPRETATION OF THE SHAPE OF THE LIKELIHOOD FUNCTION AROUND ITS MINIMUM

F. JAMES
Data Handling Division, CERN, Geneva, Switzerland

It often happens that the solution of a minimum problem is itself straightforward, but the calculation or interpretation of the resulting parameter uncertainties, as determined by the shape of the function at the minimum, is considerably more complicated. The purpose of this note is to clarify the most commonly encountered difficulties in parameter error determination. These difficulties may arise in connection with any fitting program, but will be discussed here with the terminology of the program MINUIT for the convenience of MINUIT users.

The most common causes of misinterpretation may be grouped into three categories:

1. Proper normalization of the user-supplied chi-square or likelihood function, and appropriate ERROR DEF.
2. Non-linearities in the problem formulation, leading to different errors being calculated by different techniques, such as MIGRAD, HESSE and MINOS.
3. Multiparameter error definition and interpretation.

All these topics are discussed in some detail by Eadie et al., which may be consulted for further details.

Table 1
Table of UP for multiparameter confidence regions

Number of parameters	Confidence level (probability contents desired inside hypercontour of "chi^2 = chi^2 min + UP")				
	50%	70%	90%	95%	99%
1	0.46	1.07	2.70	3.84	6.63
2	1.39	2.41	4.61	5.99	9.21
3	2.37	3.67	6.25	7.82	11.36
4	3.36	4.88	7.78	9.49	13.28
5	4.35	6.06	9.24	11.07	15.09
6	5.35	7.23	10.65	12.59	16.81
7	6.35	8.38	12.02	14.07	18.49
8	7.34	9.52	13.36	15.51	20 09
9	8.34	10.66	14.68	16.92	21.67
10	9.34	11.78	15.99	18.31	23.21
11	10.34	12.88	17.29	19.68	24.71

If FCN is –log(likelihood) instead of chi-square, all values of UP should be divided by 2.

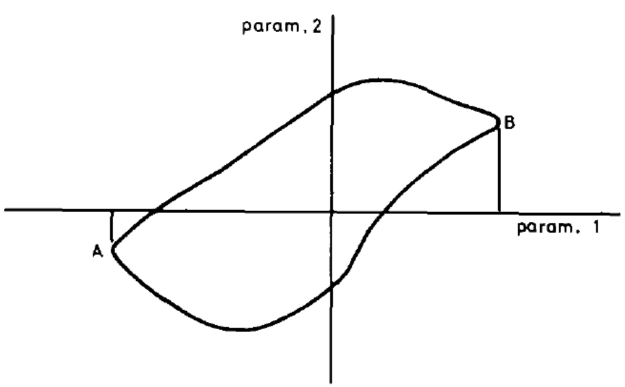


Fig. 1. MINOS errors for parameter 1.

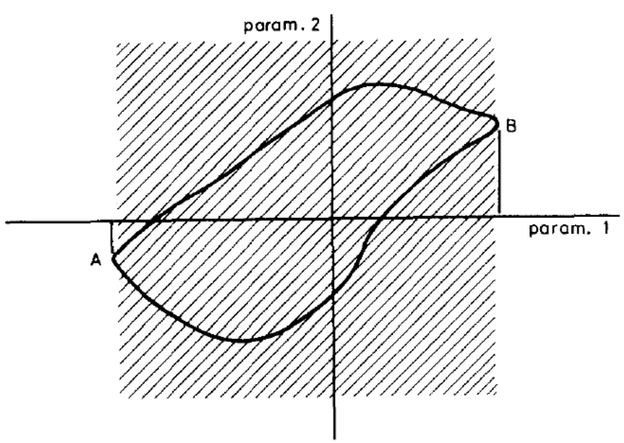


Fig. 2. MINOS error confidence region for parameter 1.

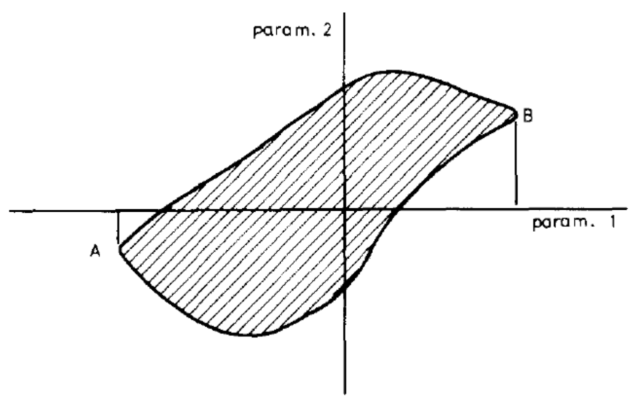


Fig. 4. Optimal confidence region for parameters 1 and 2.

Press et al., Numerical Recipes in C (1992) , p. 694. Gaussian case, all fine, agrees with F. James...

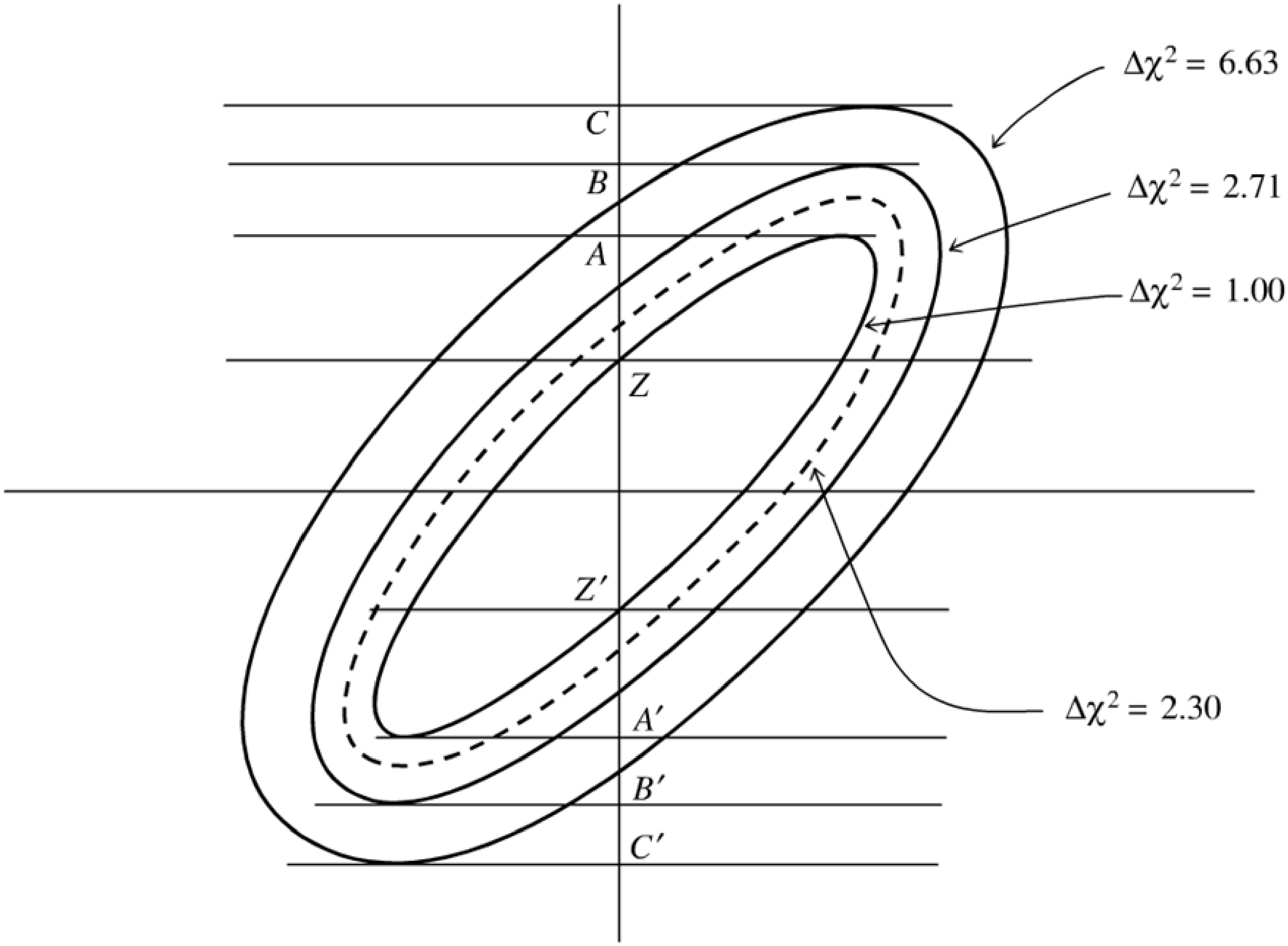


Figure 15.6.4. Confidence region ellipses corresponding to values of chi-square larger than the fitted minimum. The solid curves, with $\Delta\chi^2 = 1.00, 2.71, 6.63$ project onto one-dimensional intervals AA', BB', CC'. These intervals — not the ellipses themselves — contain 68.3%, 90%, and 99% of normally distributed data. The ellipse that contains 68.3% of normally distributed data is shown dashed, and has $\Delta\chi^2 = 2.30$. For additional numerical values, see accompanying table.

Press et al., Numerical Recipes in C (1992) , p. 695… sounds reasonable…

> However, the formal covariance matrix that comes out of a χ^2 minimization has a clear quantitative interpretation only if (or to the extent that) the measurement errors actually are normally distributed. In the case of *non*normal errors, you are "allowed"
>
> - to fit for parameters by minimizing χ^2
> - to use a contour of constant $\Delta\chi^2$ as the boundary of your confidence region
> - to use Monte Carlo simulation or detailed analytic calculation in determining *which* contour $\Delta\chi^2$ is the correct one for your desired confidence level
> - to give the covariance matrix C_{ij} as the "formal covariance matrix of the fit."
>
> You are *not* allowed
>
> - to use formulas that we now give for the case of normal errors, which establish quantitative relationships among $\Delta\chi^2$, C_{ij}, and the confidence level.

Do you see an issue?

Press et al., Numerical Recipes in C (1992) , p. 695… sounds reasonable…

However, the formal covariance matrix that comes out of a χ^2 minimization has a clear quantitative interpretation only if (or to the extent that) the measurement errors actually are normally distributed. In the case of *non*normal errors, you are "allowed"

- to fit for parameters by minimizing χ^2
- to use a contour of constant $\Delta\chi^2$ as the boundary of your confidence region
- to use Monte Carlo simulation or detailed analytic calculation in determining *which* contour $\Delta\chi^2$ is the correct one for your desired confidence level
- to give the covariance matrix C_{ij} as the "formal covariance matrix of the fit."

You are *not* allowed

- to use formulas that we now give for the case of normal errors, which establish quantitative relationships among $\Delta\chi^2$, C_{ij}, and the confidence level.

If the critical value of $\Delta\chi^2$ so determined depends on the location in parameter space, this recipe is impossible!

It is not in general possible to organize table of critical $\Delta\chi^2$ by observed best-fit params, and find such contours.

F-C: Organize by possible true parameter values!

F-C paper: http://arxiv.org/abs/physics/9711021

F-C in 2D (use notation $\Delta\chi^2 \equiv \Delta(-2\ln\mathcal{L})$ even if non-Gaussian)

$$R' \equiv \Delta\chi^2 = \sum_i \left[\frac{(n_i - b_i - \mu_i)^2}{\sigma_i^2} - \frac{(n_i - b_i - \mu_{\mathrm{best}_i})^2}{\sigma_i^2} \right], \tag{1}$$

$$R'' \equiv \Delta\chi^2 = 2\sum_i \left[\mu_i - \mu_{\mathrm{best}_i} + n_i \ln\left(\frac{\mu_{\mathrm{best}_i} + b_i}{\mu_i + b_i} \right) \right], \tag{2}$$

Rewrite key sentences in F-C Sec. V.B. more clearly: Two parameters:

$\kappa_1 \leftrightarrow \sin^2(2\theta)$ in F-C

$\kappa_2 \leftrightarrow \Delta m^2$ F-C

The acceptance region for each point in the $\kappa_1 - \kappa_2$ plane is calculated by performing a Monte Carlo simulation of the results of a large number of experiments for the given set of unknown physical parameters and the known luminosity of the actual experiment.

For each experiment, $\Delta\chi^2$ is calculated using Eq. 1 or 2.

The single number that is needed for each point in the $\kappa_1{-}\kappa_2$ plane is the critical value $\Delta\chi^2_c(\kappa_1,\kappa_2)$, such that CL of the simulated experiments have $\Delta\chi^2 < \Delta\chi^2_c$.

After the data are analyzed, $\Delta\chi^2$ for the data and each point in the $\kappa_1{-}\kappa_2$ plane, i.e. $\Delta\chi^2(N|\kappa_1,\kappa_2)$, is compared to $\Delta\chi^2_c$ and the confidence region is all points such that

$$\Delta\chi^2(N|\kappa_1,\kappa_2) < \Delta\chi^2_c(\kappa_1,\kappa_2). \tag{3}$$

The resulting regions will in general *not* be contours of $\Delta\chi^2$ for the observed χ^2 well (or likelihood function coutours) as impossibly instructed by Numerical Recipes.

N.B. FC intervals/regions are never the empty set, since they always include μ_{best}.

Coverage vs Likelihood Principle

The switch from having critical $\Delta\chi^2 = \Delta(-2\ln\mathcal{L})$ chosen only from *observed* likelihood, to having critical $\Delta\chi^2$ a function of the unknown true values, is where the L.P. is replaced by the Confidence Principle (coverage).

You can't have both L.P. and C.P. (except approximately).

(*Numerical Recipes* implies falsely that you can have both.)

Thus MINUIT MINOS regions, defined by contours in $\Delta\chi^2$, have only approximate coverage, even if one tries to adjust critical $\Delta\chi^2$ depending on observed $\mathcal{L}$.

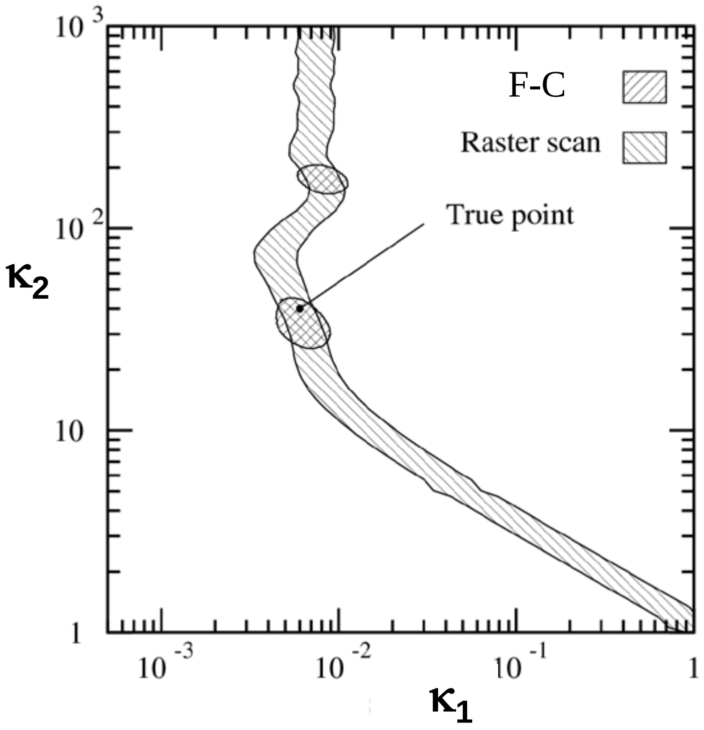
Fred James, smiling, during a break at a PhyStat meeting: "The problem with MINOS is that it obeys the Likelihood Principle."

Lack of full 2D dimensionality

Suppose you only measure one rate, and that the rate depends on some known function f(κ_1,κ_2).

Critical $\Delta\chi^2$ is 1D-like, referenced to the best-fit (κ_1,κ_2).

The confidence region in (κ_1,κ_2) will correspond to a 1D interval for the rate, which will not necessarily be the same interval as that obtained without considering the (κ_1,κ_2) plane.

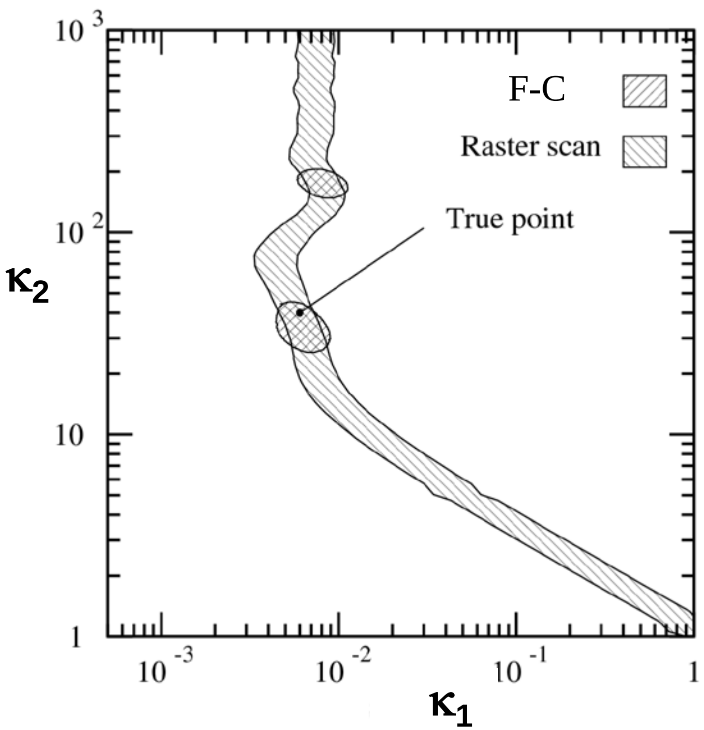


Lack of full 2D dimensionality (cont.)

Suppose that you measure two observables that vary independently with κ_1 and κ_2 .
Then critical $\Delta\chi^2$ is 2D-like.

Can be in between: In nu-osc plane with no event-by-event L/E info, it is the former 1D-likecase. *With* event-by-event L/E info, effective dimensionality varies over the plane (!).

Kyle once made a nice plot of this.



“Raster scan” set of 2D confidence regions:

For each true value of κ_2 , construct 1D acceptance interval for κ_1 .

Upon getting data, the 2D confidence region is the union of all (κ_1 , κ_2) for which the data is in the acceptance interval for κ_1 at that κ_2 .

This set of regions exactly covers.
N.B. κ_2 is *not* being treated as a nuisance parameter that is eliminated.

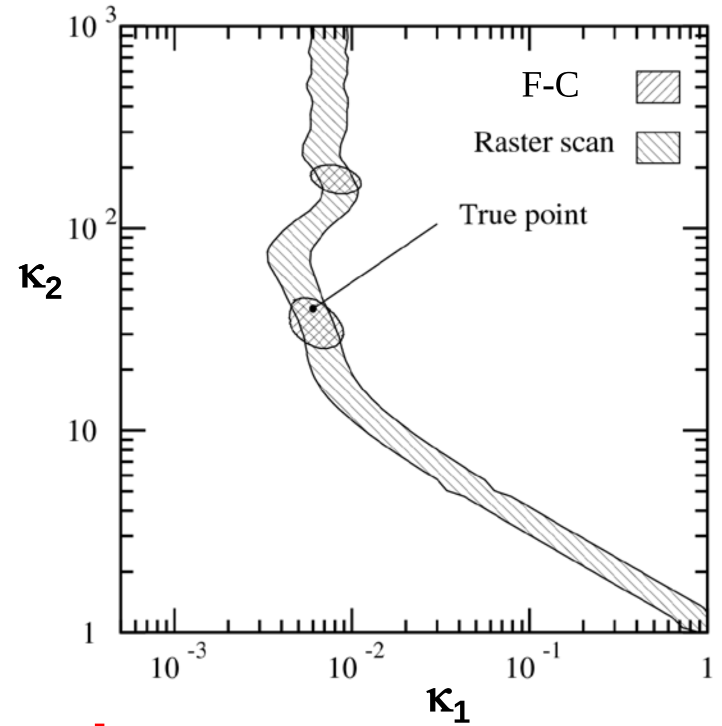
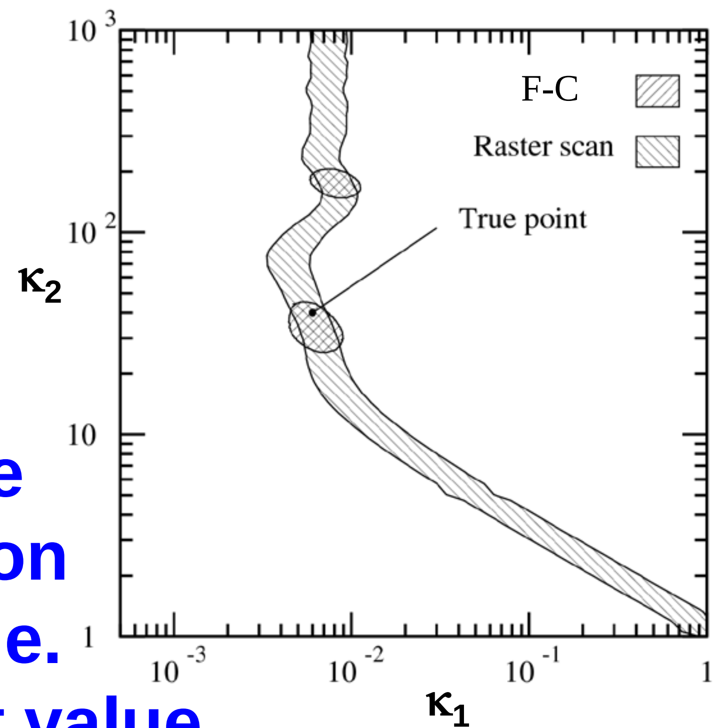


Figure (from F-C) is example of regions for an example of obtained data set in nu osc.

“Raster scan” set of 2D confidence regions (cont.):

Suppose you only care about one value of κ_2 .
The 1D confidence interval for κ_1 for that κ_2 can be read off the 2D raster-scan region.



Note that for multi-D regions such as F-C, the interval for κ_1 from a horizontal slice at one value of κ_2 does not have an interpretation independent of the rest of the plane. (If it is a null interval, it means that value of κ_2 was disfavored, so it cannot be an interval for κ_1 given that that is the true value of κ_2.)

If you want 1D F-C interval, for fixed κ_2, calculate it.

Goodness-of-Fit Intervals and Regions

A known "bad" way to determine confidence intervals and regions for parameters is:

The 95% C.L. interval/region consists of all parameter values for which there is an acceptable fit at that C.L., i.e., *absolute* χ^2 has a p-value of 0.05 or better.

***Perfect coverage*, but volatile, strong dependence on irrelevant information. Rarely, if ever, justifiable.**

A signature feature of g.o.f. intervals/regions: they can be the empty set (!).

Also: In a 2D plane of g.o.f. intervals, the 1D g.o.f. intervals can be read off simply as the naïve slices.

Measurement of the ZZ production cross section and search for anomalous couplings in $2\ell 2\ell'$ final states in pp collisions at $\sqrt{s} = 7$ TeV

JHEP01(2013)063

The CMS collaboration

E-mail: cms-publication-committee-chair@cern.ch

ABSTRACT: A measurement is presented of the ZZ production cross section in the ZZ $\to 2\ell 2\ell'$ decay mode with $\ell = \mathrm{e}, \mu$ and $\ell' = \mathrm{e}, \mu, \tau$ in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS experiment at the LHC. Results are based on data corresponding to an integrated luminosity of $5.0\,\mathrm{fb}^{-1}$. The measured cross section $\sigma(\mathrm{pp} \to \mathrm{ZZ}) = 6.24\,^{+0.86}_{-0.80}\,(\mathrm{stat.})\,^{+0.41}_{-0.32}\,(\mathrm{syst.}) \pm 0.14\,(\mathrm{lumi.})$ pb is consistent with the standard model predictions. The following limits on ZZZ and ZZγ anomalous trilinear gauge couplings are set at 95% confidence level: $-0.011 < f_4^{\mathrm{Z}} < 0.012$, $-0.012 < f_5^{\mathrm{Z}} < 0.012$, $-0.013 < f_4^{\gamma} < 0.015$, and $-0.014 < f_5^{\gamma} < 0.014$.

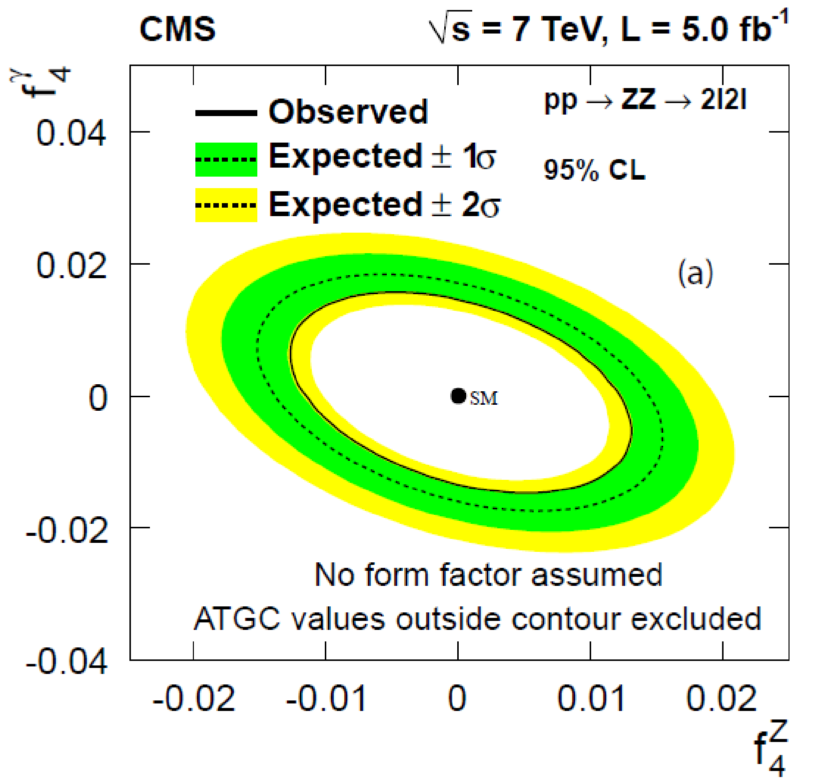
Excerpts:

The limits on ATGCs are calculated with the modified frequentist construction CL_s [37–39] based on the shape of the four-lepton invariant mass distributions, including the 4e, 4μ, and $2e2\mu$ channels in the likelihood combination.

For each distribution only one or two couplings are varied, while all others are set to zero. The fit is performed to find the maximum likelihood value and limits are calculated.

Figure 3 presents the expected and observed two-dimensional exclusion limits at 95% confidence level (CL) on the anomalous neutral trilinear ZZZ and $ZZ\gamma$ couplings. The green and yellow bands represent the one and two standard-deviation variations from the expected limit.

Each point in the plane is tested independently using CL_s and either excluded at 95% C.L., or not.
The 95% C.L. region is the set of points not excluded.



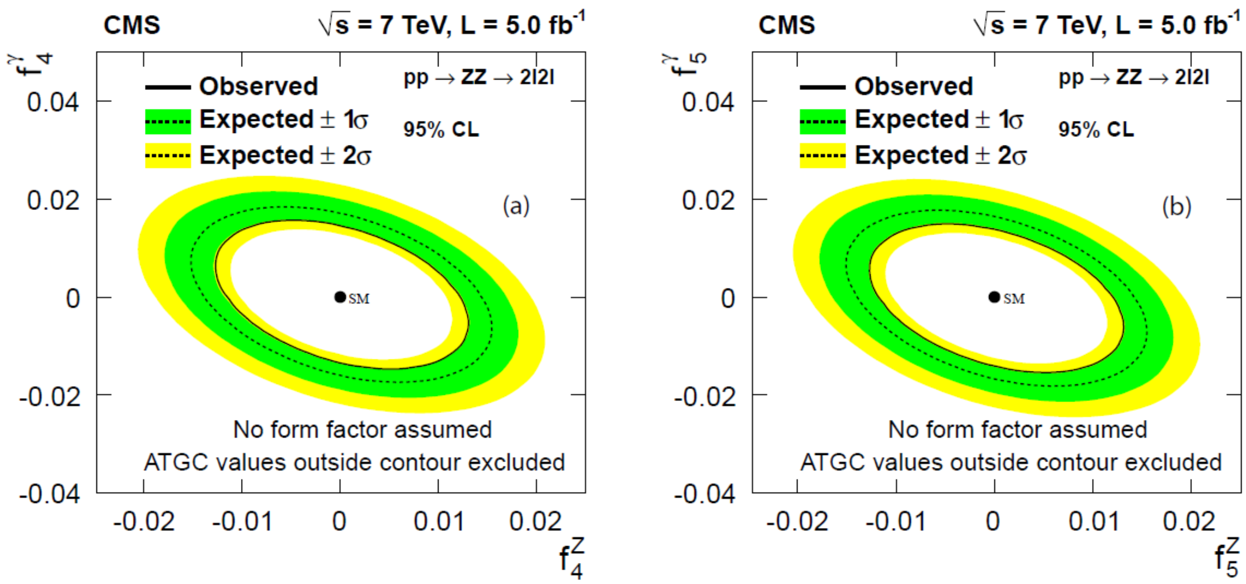


Figure 3. Expected and observed two-dimensional exclusion limits at 95% CL on the anomalous neutral trilinear ZZZ ($f^Z_{4,5}$) and ZZγ ($f^\gamma_{4,5}$) couplings. The green and yellow bands represent the one and two standard-deviation variations from the expected limit. In calculating the limits, the anomalous couplings that are not shown in the figure are set to zero.

One-dimensional 95% CL limits for the $f_4^{Z,\gamma}$ and $f_5^{Z,\gamma}$ anomalous coupling parameters are measured to be

$$-0.011 < f_4^Z < 0.012,\ -0.012 < f_5^Z < 0.012,\ -0.013 < f_4^\gamma < 0.015,\ -0.014 < f_5^\gamma < 0.014.$$

In the one-dimensional fits, all of the ATGC parameters except the one under study are kept fixed to zero.

Comments

Louis raised the point that a 1D interval is read off as a slice of the 2D boundary: no wider or narrower.
This behavior recalls G.O.F. intervals.

Is it possible to eliminate all points in the plane?

Higgs search was actually similar: each mass tested independently for SM Higgs.

One may want to distinguish between the cases:
1) It is known that *there exists* a “true” point somewhere in the plane, as in measurement of couplings, and
2) There might not be any true point (as in Higgs might not exist)

***In either case*, I think that FC is a useful option. Using CL_s to “measure” couplings seems rather odd.**

Backup

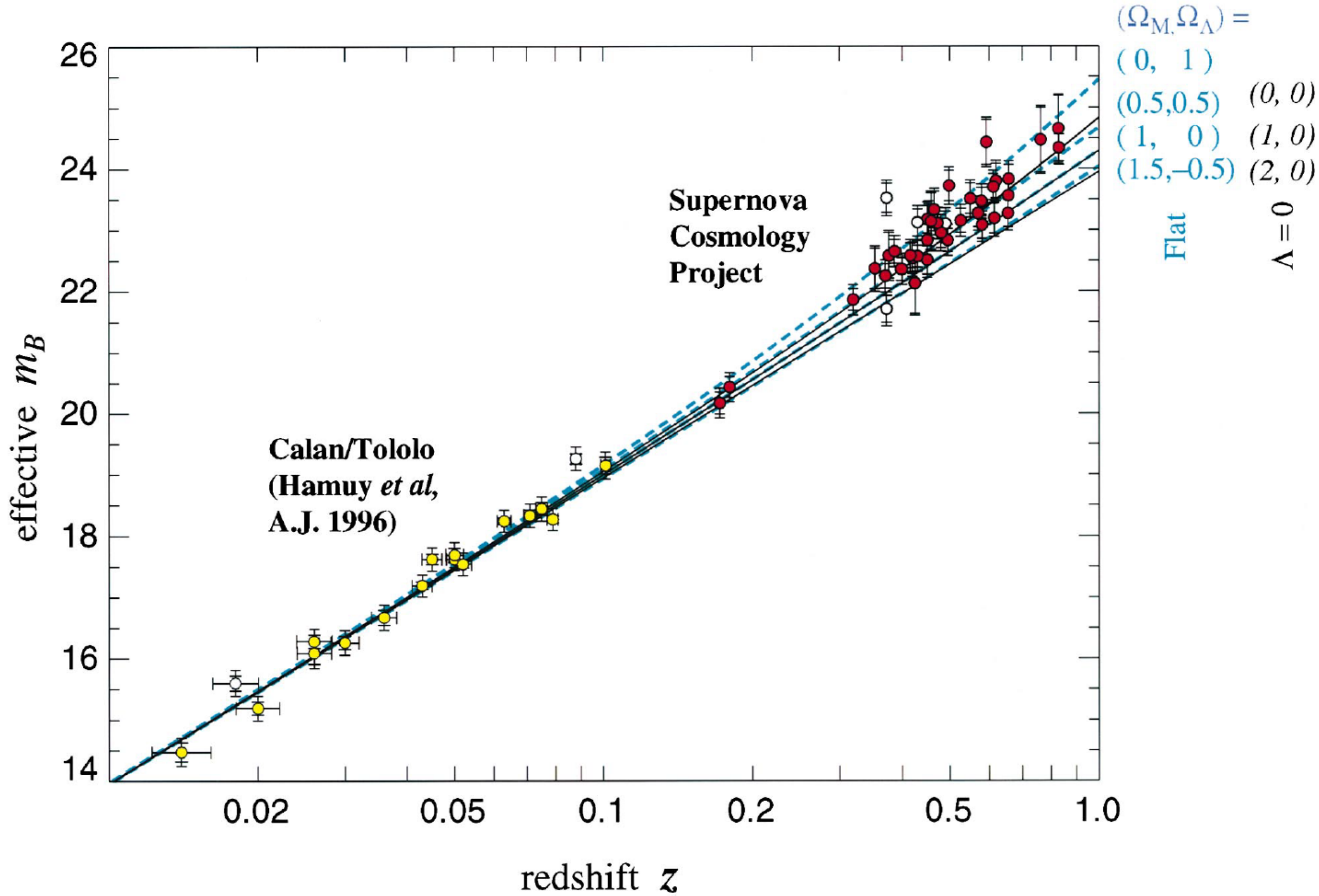


FIG. 1.—Hubble diagram for 42 high-redshift type Ia supernovae from the Supernova Cosmology Project and 18 low-redshift type Ia supernovae from the Calán/Tololo Supernova Survey after correcting both sets for the SN Ia light-curve width-luminosity relation. The inner error bars show the uncertainty due to measurement errors, while the outer error bars show the total uncertainty when the intrinsic luminosity dispersion, 0.17 mag, of light-curve–width-corrected type Ia supernovae is added in quadrature. The unfilled circles indicate supernovae not included in fit C. The horizontal error bars represent the assigned peculiar velocity uncertainty of 300 km s^{-1}. The solid curves are the theoretical $m_B^{\mathrm{eff}}(z)$ for a range of cosmological models with zero cosmological constant: $(\Omega_M, \Omega_\Lambda) = (0, 0)$ on top, (1, 0) in middle, and (2, 0) on bottom. The dashed curves are for a range of flat cosmological models: $(\Omega_M, \Omega_\Lambda) = (0, 1)$ on top, (0.5, 0.5) second from top, (1, 0) third from top, and (1.5, −0.5) on bottom.

Saul and I agreed (a theorist did not) that using g.o.f. to decide which values of $(\Omega_M,\Omega_\Lambda)$ to accept created dependence on data at $\mathcal{Z}$<0.1, which is irrelevant once model is accepted. (Also discussed by Fred James at 2002 Durham Stats mtg.)

We have compared the results of Bayesian and classical, "frequentist," fitting procedures. For the Bayesian fits, we have assumed a "prior" probability distribution that has zero probability for $\Omega_M < 0$ but otherwise has uniform probability in the four parameters Ω_M, Ω_Λ, α, and $\mathscr{M}_B$. For the frequentist fits, we have followed the classical statistical procedures described by Feldman & Cousins (1998) to guarantee frequentist coverage of our confidence regions in the physically allowed part of parameter space. Note that throughout the previous cosmology literature, completely unconstrained fits have generally been used that can (and do) lead to confidence regions that include the part of parameter space with negative values for Ω_M. The differences between the confidence regions that result from Bayesian and classical analyses are small. We present the Bayesian confidence regions in the figures, since they are somewhat more conservative; i.e., they have larger confidence regions in the vicinity of particular interest near $\Lambda = 0$.

Goodness-of-Fit Intervals and Regions (cont.)

N.B.: It is not well-known, but the *absolute* χ^2 used for g.o.f. is actually a log of a *likelihood ratio* $\Delta(-2\ln\mathcal{L})$, where the alternative hypothesis is the "saturated model", the model which is set each to the observed data.

See http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf

Classical Hypothesis Testing: Duality

"Test for $\theta=\theta_0$" $\leftrightarrow$ "Is θ_0 in confidence interval for θ"

Table 20.1 Relationships between hypothesis testing and interval estimation

Property of test	Property of corresponding confidence interval
Size $= \alpha$	Confidence coefficient $= 1 - \alpha$
Power = probability of rejecting a false value of $\theta = 1 - \beta$	Probability of not covering a false value of $\theta = 1 - \beta$
Most powerful	Uniformly most accurate
$\longleftarrow$	$\left\{ \begin{array}{c} \textit{Unbiased} \\ 1 - \beta \geq \alpha \end{array} \right\}$ $\longrightarrow$
Equal-tails test $\alpha_1 = \alpha_2 = \frac{1}{2}\alpha$	Central interval

"There is thus no need to derive optimum properties separately for tests and for intervals; there is a one-to-one correspondence between the problems as in the dictionary in Table 20.1" – Stuart99, p. 175.