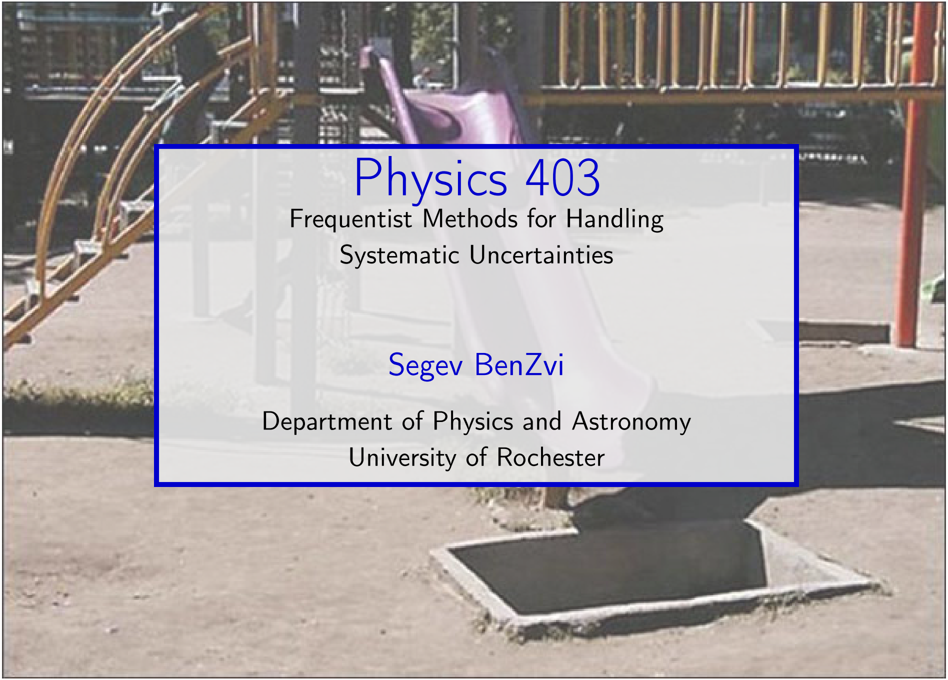


Physics 403

Frequentist Methods for Handling Systematic Uncertainties

Segev BenZvi

Department of Physics and Astronomy
University of Rochester

Table of Contents

Review of Systematic Effects

- When designing an experiment and taking data, you need to worry about systematic effects and offsets
- Systematics are not caused by faulty calibration or equipment; those are mistakes
- When taking data, test the robustness of results by varying the conditions: analysis cuts, techniques, etc.
- Worry about offsets and unexpected results, try to remove them
- Assign a systematic uncertainty when other options are exhausted. Knowing when to cut your losses comes with experience

An Example Error Budget

- Here is a systematic error budget for the energy scale of an air fluorescence detector, discussed in the last class:

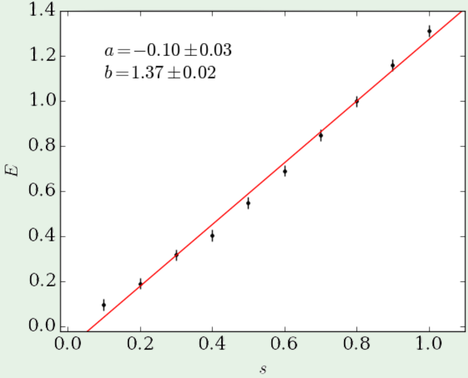
Source	Uncertainty
Fluorescence Yield Y	14%
p, T, e Effects on Y	7%
Calibration	9.5%
Atmosphere	4%
Reconstruction	10%
Invisible Energy	4%
Total	22%

- Note how the uncertainties are added in quadrature. What has been assumed here?
- If you were working on this experiment, which uncertainties would you try to minimize first? What is the best use of your time?

Case Study: Fitting an Inappropriate Function

Example

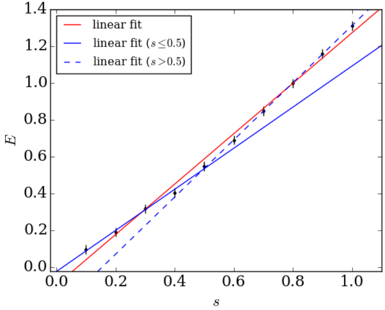
Suppose you have a calorimeter that gives you a signal s, which is related to energy by $E = s + 0.3s^2$ [1].



You take data and fit a straight line $E = a + b \cdot s$, and use the values $\hat{a}$ and $\hat{b}$ in your analysis.

Case Study: Fitting an Inappropriate Function

- You find that $\chi^2 = 16.94$ with 8 degrees of freedom, which is large but not unreasonable. (What is the approximate p-value?)
- So you stick with the linear fit, but as a check you calibrate (i.e., fit) the subranges $0 \leq s \leq 0.5$ and $0.5 < s \leq 1$ separately:



- Result: the slopes are 1.17 ± 0.03 and 1.57 ± 0.06, definitely not agreeing within statistical uncertainties.

Case Study: Fitting an Inappropriate Function

- You follow the procedure for dealing with systematic effects (check, re-check, worry) but fail to spot that the linear calibration is itself inadequate.
- Result: you incorporate a systematic uncertainty of $1.57 - 1.37 = 1.37 - 1.17 = 0.2$ into the slope b, reporting

$$b = 1.37 \pm 0.02 \pm 0.20$$

Is this reasonable?

Case Study: Fitting an Inappropriate Function

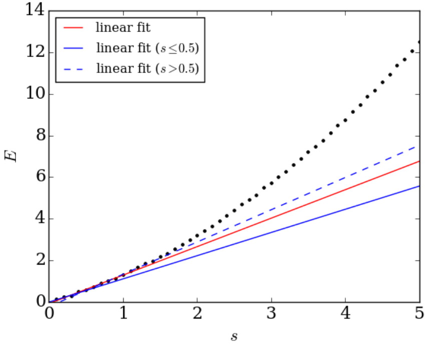
- You follow the procedure for dealing with systematic effects (check, re-check, worry) but fail to spot that the linear calibration is itself inadequate.
- Result: you incorporate a systematic uncertainty of $1.57 - 1.37 = 1.37 - 1.17 = 0.2$ into the slope b, reporting

$$b = 1.37 \pm 0.02 \pm 0.20$$

  Is this reasonable?
- In the region $0 \leq s \leq 1$ this systematic uncertainty seriously overstates the error.
- Look again at the fit. The slope 1.37 is a pretty reasonable description of the data

Case Study: Fitting an Inappropriate Function

- What happens if the "calibration" of $E(s)$ is extrapolated to $s = 5$?



- The linear extrapolation is clearly no good. Not only that, but the systematic uncertainty is worthless for describing the calibration offset
- Lesson: there is no correct procedure for incorporating a check that fails, but folding it into the systematics is probably wrong and should be avoided unless there is no alternative

Case Study: Superluminal Neutrinos

Example

- Recall the ν_μ time-of-flight anomaly measured by OPERA and discussed earlier in the semester [2]:

$$(v_\nu - c)/c = (2.48 \pm 0.28 \pm 0.30) \times 10^{-5}$$

- This result is in significant tension with Einstein's relativity. Later, a competitor experiment did not observe this effect [3]
- The OPERA collaboration carried out many checks of their analysis before making the announcement (Sep. 2011)
- Checks of the equipment were not redone until December 2011. In December they discovered that a partially unscrewed optical fiber was affecting the time-of-flight measurement
- **Question**: did the OPERA collaboration do the right thing by going public with their anomaly? What could/should they have done differently?

Table of Contents

Quantitative Approach to Handling Systematics

Suppose you've done your "homework" and identified and removed all the sytematic effects you can. You are left with some irreducible uncertainties:

- Calibration uncertainties
- Contributions from known sources of background with statistical uncertainties
- "Theory errors," e.g., a cross section calculated to some finite accuracy
- Inputs with measurement uncertainties, e.g., Hubble's constant H_0

If you're a Bayesian, you would propagate these kinds of uncertainties using marginalization (recall your homework problem about distance and recession velocity)

In the frequentist approach, there are also standard methods for handling systematics. This is what we will discuss today

The Δ Method

- The Δ or Shift Method [4] is based on the linear propagation of errors
- Given N nuisance parameters μ_i with uncorrelated Gaussian uncertainties σ_i and an estimator of the parameter of interest $f(\mu_1, \ldots, \mu_N)$, the linear approximation gives

$$\sigma_f^2 \approx \sum_{i=1}^{N} \left(\frac{\partial f}{\partial \mu_i} \right)^2 \sigma_i^2$$

- If f is roughly linear over the region $\mu_i \pm \sigma_i$, then

$$\frac{\partial f}{\partial \mu_i} \approx \frac{f(\mu_1, \ldots, \mu_i + \sigma_i, \ldots, \mu_N) - f(\mu_1, \ldots, \mu_i, \ldots, \mu_N)}{\sigma_i} = \frac{\Delta_i}{\sigma_i}$$

$$\therefore \sigma_f^2 \approx \sum_{i=1}^{N} \Delta_i^2$$

I.e., you just add all the 1σ shifts in quadrature.

Δ Method Example: Linear Fit with Distorting Systematics

- Nice example from Scott Oser: we have a model that predicts $y = a + bx$ and data $x_i, y_i, \sigma_i = 1$ giving

$$\hat{a} = 4.60 \pm 0.93, \qquad \hat{b} = 0.46 \pm 0.12$$

- Suppose the y_i are systematically biased by $\Delta y_i = \alpha x_i + \beta x_i^2$, where we believe that $\alpha = 0.00 \pm 0.05$ and $\beta = 0.00 \pm 0.01$
- Make a table with various permutations of the $\pm 1\sigma$ errors:

α	β	a	b	Δa	Δb
0	0	4.602	0.464	0.000	0.000
0.05	0	4.602	0.414	0.000	-0.050
-0.05	0	4.602	0.514	0.000	0.050
0	0.01	4.228	0.604	-0.374	0.140
0	-0.01	4.975	0.324	0.373	-0.140

Δ Method Example: Linear Fit with Distorting Systematics

- We treat the systematic uncertainties in the nuisance parameters α and β as uncorrelated and add them in quadrature:

$$\hat{a} = 4.60 \pm 0.93\ (\text{stat}) \pm 0.37\ (\text{sys}) = 4.60 \pm 1.00$$

$$\hat{b} = 0.46 \pm 0.12\ (\text{stat}) \pm 0.15\ (\text{sys}) = 0.46 \pm 0.19$$

- There is nothing in the Δ method that forces us to assume uncorrelated uncertainties. Given the full covariance matrix of the nuisance parameters $\boldsymbol{V}$ including correlations, we could write

$$\sigma_f^2 \approx \sum_{i=1}^{N} \sum_{j=1}^{N} \Delta_i \Delta_j \left(\frac{V_{ij}}{\sigma_i \sigma_j} \right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \Delta_i \Delta_j \rho_{ij}$$

where ρ_{ij} is the correlation coefficient for variables i and j

Monte Carlo Propagation of Systematics

If you can identify nuisance parameters and assign PDFs to them (not just Gaussians), Monte Carlo is a good way to propagate the uncertainties

1. Start by randomly sampling values of each nuisance parameter from its PDF
2. Analyze the data using the sampled values
3. Return to step 1 and repeat until you have enough statistics

Given the set of Monte Carlo samples you generated, you can plot the distribution of each parameter of interest. The width of the distribution gives the systematic uncertainty on the parameter

To identify the relative importance of each parameter, you can marginalize over the other parameters or rerun the Monte Carlo varying just one systematic uncertainty at a time

Monte Carlo Propagation of Systematics

Advantages of the Monte Carlo technique:

- Free of assumptions about the PDFs such as Gaussianity
- Considers effects of all systematics jointly
- Handles correlations between systematic uncertainties

Disadvantages of the Monte Carlo technique:

- The method does not allow the data to constrain the systematics; for that you use the pull method
- The technique does not allow you to identify the relative importance of each nuisance parameter unless you marginalize or vary the parameters one by one
- Monte Carlo can be a slow way to propagate uncertainties

Covariance Matrix Approach

Imagine measurements x_i which have independent statistical uncertainties σ_i and a common systematic uncertainty σ_s. E.g., the $\{x_i\}$ have a systematic additive offset $s \pm \sigma_s$, such that $x_i \to x_i + s$

$$\begin{aligned}
\text{var}\,(x_i) &= \langle x_i^2 \rangle - \langle x_i \rangle^2 \\
&= \langle (x_i + s)^2 \rangle - \langle x_i + s \rangle^2 \\
&= \langle x_i^2 \rangle + \langle s^2 \rangle + \cancel{\langle 2 s x_i \rangle} - \langle x_i \rangle^2 - \langle s \rangle^2 - \cancel{2 \langle x_i \rangle \langle s \rangle} \\
&= \sigma_i^2 + \sigma_s^2 \\
\text{cov}\,(x_i, x_j) &= \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \\
&= \langle (x_i + s)(x_j + s) \rangle - \langle x_i + s \rangle \langle x_j + s \rangle \\
&= [\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle] + [\langle x_i s \rangle - \langle x_i \rangle \langle s \rangle] + [\langle x_j s \rangle - \langle x_j \rangle \langle s \rangle] \\
&\quad + \langle s^2 \rangle - \langle s \rangle^2 \\
&= \cancel{\text{cov}\,(x_i, x_j)} + \cancel{\text{cov}\,(x_i, s)} + \cancel{\text{cov}\,(x_j, s)} + \text{var}\,(s) \\
&= \sigma_s^2
\end{aligned}$$

Covariance Matrix Approach

- Thus, the covariance matrix $\boldsymbol{V}$ is written

$$V_{ij} = \text{cov}\left(x_i, x_j\right) = \delta_{ij}\sigma_i^2 + \sigma_s^2$$
$$= \begin{pmatrix} \sigma_1^2 + \sigma_s^2 & \sigma_s^2 & \dots \\ \sigma_s^2 & \sigma_2^2 + \sigma_s^2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

  with the systematic uncertainty σ_s added in quadrature to the statistical uncertainties σ_i

- Note that if the systematic uncertainty were proportional to the measurement such that $\sigma_s = \epsilon x$, we could have written

$$V_{ij} = \begin{pmatrix} \sigma_1^2 + \epsilon^2 x_1^2 & \epsilon^2 x_1 x_2 & \dots \\ \epsilon^2 x_1 x_2 & \sigma_2^2 + \epsilon^2 x_2^2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Covariance Matrix Approach

- Generalizing, if there is another systematic uncertainty σ_T shared by x_1 and x_2 but not x_3, the covariance matrix becomes

$$V_{ij} = \begin{pmatrix} \sigma_1^2 + \sigma_s^2 + \sigma_T^2 & \sigma_s^2 + \sigma_T^2 & \sigma_s^2 \\ \sigma_s^2 + \sigma_T^2 & \sigma_2^2 + \sigma_s^2 + \sigma_T^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 & \sigma_3^2 + \sigma_s^2 \end{pmatrix}$$

- Once we have a covariance matrix, we can write down

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - \mu - \hat{s})V_{ij}^{-1}(x_j - \mu - \hat{s}).$$

if we have an existing estimator $\hat{s}$. Minimizing χ^2 will give a ML/LS estimator $\hat{\mu}$.

Handling a Systematic Offset

- Let $s = 2.0 \pm 0.4$, $\{x_i\} = (10.0, 10.0, 11.0, 12.0)$, and $\sigma_i = 1.0$
- What one would usually do is solve for the central value $\hat{\mu}$ given the estimator $\hat{s} = 2.0$:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{4} x_i - \hat{s} = 8.75$$

$$\text{var}\left(\hat{\mu}\right) = \frac{1}{\sum_i 1/\sigma_i^2} = \frac{1}{4}$$

$$\therefore \hat{\mu} = 8.75 \pm 0.5 \text{ (stat)}$$

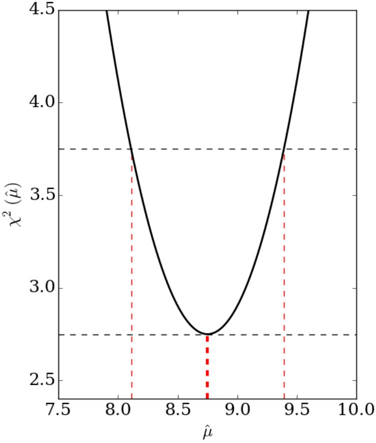
- There is also a systematic uncertainty on the parameter due to σ_s, which can be added in quadrature to the statistical uncertainty:

$$\hat{\mu} = 8.75 \pm 0.5 \text{ (stat)} \pm 0.4 \text{ (sys)} = 8.75 \pm \sqrt{0.5^2 + 0.4^2}$$
$$= 8.75 \pm 0.64$$

Solution using χ^2 Minimization

Just to demonstrate that it works, here is the result of minimizing

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - \mu - \hat{s})V_{ij}^{-1}(x_j - \mu - \hat{s})$$



- The χ^2 minimum is $\hat{\mu} = 8.75$
- Using $\Delta\chi^2 = 1$ to identify the 1σ uncertainty on $\hat{\mu}$ (exact this time because all errors are Gaussian) gives

$$\hat{\mu} = 8.75 \pm 0.64$$

- So the ML/LS methods we described last week can be applied in the presence of systematic uncertainties

Adding Constraints to the Likelihood

- Recall the definition of the posterior PDF given parameters θ and α:

$$p(\theta, \alpha|D, I) \propto p(D|\theta, \alpha, I)\ p(\theta|I)\ p(\alpha|I)$$

- The ML estimator $\ln\mathcal{L}(\theta) = \ln p(D|\theta, I)$ assumes a flat prior on θ
- This is easy to generalize to include a systematic in terms of a nuisance paramater α:

$$\ln\mathcal{L}(\theta, \alpha) = \ln\mathcal{L}(\theta|D, \alpha) + \ln p(\alpha)$$

- The first term is a regular likelihood. The second is a constraint or "penalty" term and behaves like a prior on α
- **Note**: $p(\alpha)$ can be any valid PDF, so this can handle non-Gaussian uncertainties

Adding Constraints to the Likelihood: Two Rulers

- We have discussed the example of measuring a length with two separate thermally expanding rulers made of different materials:

$$y_i = L_i + c_i(T - T_0),$$

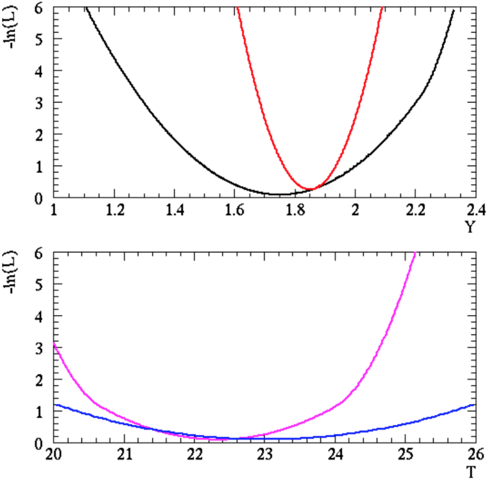
where y_i are the measured lengths, L_i are the lengths measured at T_0, and c_i are the coefficients of expansion

- We want to calculate the "true length" y considering T as a nuisance parameter $T = 23 \pm 2$:

$$-2\ln\mathcal{L}(y, T) = \sum_{i=1}^{2}\left(\frac{y - L_i - c_i(T - T_0)}{\sigma_L}\right)^2 + \left(\frac{T - 23}{2}\right)^2$$

- The first term is the usual Gaussian likelihood, and the second is a Gaussian constraint on T
- Procedure: marginalize over T to get the shape of the likelihood as a function of y

Constraint Term in the Likelihood



Top plot: $-\ln \mathcal{L}(y, T)$ after marginalizing T

- Red: fixed T
- Black: marginalize $-\ln \mathcal{L}$ as function of T in $y \pm 1\sigma$ range

Bottom: marginalizing y

- Blue: "a priori" constraint on $T = 23 \pm 2$
- Magenta: log likelihood after marginalization of y

The "Pull" Method

- A technique equivalent to the use of the covariance matrix, but easier to apply in practice, is the pull method
- Given observables y_i, predictions $f_i = f(x_i)$, and a covariance matrix V_{ij}, minimize
$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(y_i - f_i)V_{ij}^{-1}(y_j - f_j)$$
- In the pull method, factorize the uncertainties into uncorrelated errors u_i and correlated systematic uncertainties c_i^k (the shift of observable i by systematic error source k). Shift the difference $y_i - f_i$ by the amount $-c_i^k \xi_k$, where ξ_k is a Gaussian:
$$y_i - f_i \to (y_i - f_i) - \sum_{k=1}^{K} c_i^k \xi_k$$
$$\therefore \chi^2_{\text{pull}} = \min_{\{\xi_k\}} \left[\sum_{i=1}^{N}\left(\frac{y_i - f_i - \sum_k c_i^k \xi_k}{u_i}\right)^2 + \sum_{k=1}^{K} \xi_k^2\right]$$

The "Pull" Method

- Denote $\bar{\xi}_k$, the pulls of the sytematics, as the values of ξ_k at the minimim
- Define $\bar{x}_i$, the pulls of the observables as

$$\bar{x}_i = \frac{y_i - (f_i + \sum_k \bar{\xi}_k c_i^k)}{u_i}$$

- We can then split χ^2_{pull} into two diagonalized pieces:

$$\begin{aligned}\chi^2_{\text{pull}} &= \chi^2_{\text{obs}} + \chi^2_{\text{sys}} \\ &= \sum_{i=1}^{N} \bar{x}_i^2 + \sum_{k=1}^{K} \bar{\xi}_k^2\end{aligned}$$

I.e., we separate the χ^2 into contributions from the residuals of the observables x_i and from the systematics ξ_k

Example: "Pull" Method

- To illustrate the method in practice, let's go back to the example of data of form $y = a + bx$ with systematic offsets $\Delta y_i = \alpha x_i + \beta x_i^2$
- The pull χ^2 is minimized with respect to the nuisance parameters α and β:

$$\chi^2_{\text{pull}} = \sum_i \left(\frac{y_i - a - bx_i - \alpha x_i - \beta x_i^2}{1.0} \right)^2 + \left(\frac{\alpha - 0}{0.05} \right)^2 + \left(\frac{\beta - 0}{0.01} \right)^2$$

Case	a	b
fix α, β	4.64 ± 0.93	0.46 ± 0.12
min α, fix β	4.65 ± 0.93	0.45 ± 0.13
fix α, min β	4.65 ± 0.99	0.45 ± 0.18
min α, min β	4.65 ± 0.99	0.45 ± 0.19

Example: "Pull" Method

- By minimizing with respect to the nuisance parameters α and β, we are doing the frequentist equivalent of marginalization
- To break the systematic uncertainty out of the total uncertainty, calculate the quadrature difference of the statistical and total uncertainties:

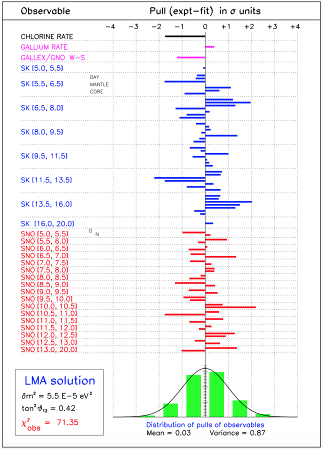
Case	a	b
fix α, β	4.64 ± 0.93	0.46 ± 0.12
min α, fix β	4.65 ± 0.93	0.45 ± 0.13
fix α, min β	4.65 ± 0.99	0.45 ± 0.18
min α, min β	4.65 ± 0.99	0.45 ± 0.19

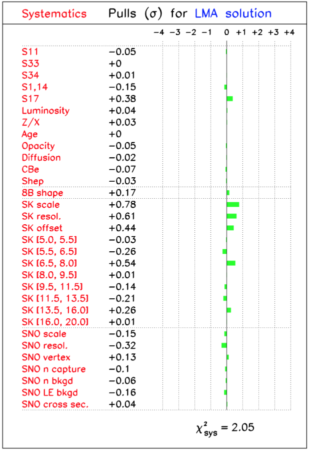
$$a = 4.65 \pm 0.93\ (\text{stat}) \pm \sqrt{0.99^2 - 0.93^2} = 4.65 \pm 0.93 \pm 0.34$$

$$b = 0.45 \pm 0.12\ (\text{stat}) \pm \sqrt{0.19^2 - 0.12^2} = 0.45 \pm 0.12 \pm 0.15$$

Plotting the Pulls

It is useful to plot the pulls $\bar{x}_i$ and $\bar{\xi}_k$ for the N parameters and K systematics, since it helps you to pick out which parts of the fit (if any) are dominating the disagreement with a model. For example [5],





Pull vs. Covariance Method

- The pull method puts nuisance parameters on the same footing as other parameters by adding penalty terms to the likelihood/χ^2
- The data are used to reject certain values of the nuisance parameters (α and β in our linear fit example) and keep their range reasonable
- So if you use a frequentist approach, your choices are to add constraint terms to the likelihood and minimize, or calculate the covariance matrix between all points and minimize that
- It is easier to work with pulls because the constraints on the systematics are more obvious
- Try not to use the Δ method if you don't have to. You won't get the same kinds of constraints from the data that you get using marginalization or the pull/covariance techniques

Table of Contents

Summary

- In frequentist statistics, it is convenient to separate uncertainties into values that depend on the statistics in the data and values that depend on systematic effects
- Don't get hung up on the idea that these uncertainties are completely different. As in the Bayesian world, you can combine the errors into a total uncertainty
- Systematics can be absorbed into ML/LS techniques by expressing them as correlated errors in a covariance matrix or using the pull method
- Systematics can be expressed in a likelihood as penalty terms, which you can marginalize or maximize to get a profile likelihood
- This approach lets you incorporate asymmetric uncertainties; see [6] for how to handle asymmetric error bars in published data

References I

[1] Roger Barlow. "Systematic Errors: Facts and Fictions". In: *Conf. on Adv. Stat. Techniques in Particle Physics*. Durham, England, 2002, pp. 134–144. arXiv: hep-ex/0207026 [hep-ex].

[2] T. Adam et al. "Measurement of the neutrino velocity with the OPERA detector in the CNGS beam". In: *JHEP* 1210 (2012), p. 093. arXiv: 1109.4897 [hep-ex].

[3] M. Antonello et al. "Measurement of the neutrino velocity with the ICARUS detector at the CNGS beam". In: *Phys.Lett.* B713 (2012), pp. 17–22. arXiv: 1203.3433 [hep-ex].

[4] Joel Heinrich and L. Lyons. "Systematic errors". In: *Ann.Rev.Nucl.Part.Sci.* 57 (2007), pp. 145–169.

[5] G.L. Fogli et al. "Getting the most from the statistical analysis of solar neutrino oscillations". In: *Phys.Rev.* D66 (2002), p. 053010. arXiv: hep-ph/0206162 [hep-ph].

References II

[6] Roger Barlow. "Asymmetric Errors". In: *PHYSTAT 2003*. SLAC, 2003, pp. 250–255.