Fits to combined datasets

Roger Barlow

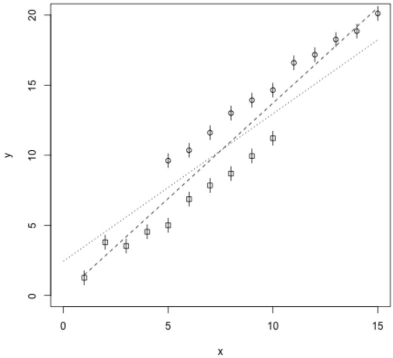
Manchester LHCb Group Meeting

17^{th} December 2020

The problem

Fitting a function to data which share a systematic error within experiments. (Or data-taking runs. Or tracker-modules. Or...)

Ignoring systematics gives wrong answer (dashed line)
Need to include them for correct (dotted) line



The puzzle

There are two different ways to do it. Which should you use?

The Question

Which of two Methods should you use?

Using i, j... to index the data and r to index the experiments
a represents the fitting parameters
Individual measurement errors σ_i and systematic errors S_r

Invert the Covariance matrix

Adjust a to minimise χ^2

$$\sum_{i,j}(y_i - f(x_i, a))\mathbf{V}_{ij}^{-1}(y_j - f(x_j, a))$$

$$V_{ij} = \delta_{ij}\sigma_i^2 + \delta_{r_i r_j} S_{r_i}^2$$

Need to invert a (very) large matrix

Introduce and fit deviations z_r

Adjust a and z to minimise χ^2

$$\sum_r \sum_{i \in r} \left(\frac{y_i - f(x_i, a) - z_r}{\sigma_i}\right)^2 + \left(\frac{z_r}{S_r}\right)^2$$

Minimisation space has (several) extra parameters

The Question

Which of two Methods should you use?

Using i, j... to index the data and r to index the experiments
a represents the fitting parameters
Individual measurement errors σ_i and systematic errors S_r

Invert the Covariance matrix

Adjust a to minimise χ^2

$$\sum_{i,j}(y_i - f(x_i, a))\mathbf{V}^{-1}_{ij}(y_j - f(x_j, a))$$

$$V_{ij} = \delta_{ij}\sigma_i^2 + \delta_{r_i r_j} S_{r_i}^2$$

Need to invert a (very) large matrix

Introduce and fit deviations z_r

Adjust a and z to minimise χ^2

$$\sum_r \sum_{i \in r} \left(\frac{y_i - f(x_i, a) - z_r}{\sigma_i} \right)^2 + \left(\frac{z_r}{S_r} \right)^2$$

Minimisation space has (several) extra parameters

The answer

It doesn't matter. They are equivalent

Inverting the matrix

$$\begin{pmatrix} \sigma_1^2+S_1^2 & S_1^2 & S_1^2 & S_1^2 & 0 & 0 & 0 & \dots \\ S_1^2 & \sigma_2^2+S_1^2 & S_1^2 & S_1^2 & 0 & 0 & 0 & \dots \\ S_1^2 & S_1^2 & \sigma_3^2+S_1^2 & S_1^2 & 0 & 0 & 0 & \dots \\ S_1^2 & S_1^2 & S_1^2 & \sigma_4^2+S_1^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \sigma_5^2+S_2^2 & S_2^2 & S_2^2 & \dots \\ 0 & 0 & 0 & 0 & S_2^2 & \sigma_6^2+S_2^2 & S_2^2 & \dots \\ 0 & 0 & 0 & 0 & S_2^2 & S_2^2 & \sigma_7^2+S_2^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

$\mathbf{V}$ is block diagonal, $diag\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 .., \}$ so $\mathbf{V}^{-1}$ is also block diagonal $diag\{\mathbf{v}_1^{-1}, \mathbf{v}_2^{-1}, \mathbf{v}_3^{-1} .., \}$

Consider case where all σ_i the same $\mathbf{v} = \sigma^2 \mathbf{I} + S^2 \mathbf{U}$

$\mathbf{U}$ is matrix of ones. Note $\mathbf{U}^2 = n\mathbf{U}$

$\mathbf{v}^{-1} = \frac{1}{\sigma^2}\mathbf{I} - \frac{S^2}{\sigma^2(\sigma^2+nS^2)}\mathbf{U}$

Now generalise $\mathbf{v}_{ij}^{-1} = \frac{\delta_{ij}}{\sigma_i^2} + \frac{S_r^2}{1+\sum_{k\in E_r}\frac{S_r^2}{\sigma_k^2}} \frac{1}{\sigma_i^2\sigma_j^2}$

Minimise $\chi^2 = \sum_r \sum_{i\in E_r} \frac{\Delta_i^2}{\sigma_i^2} - \sum_r \frac{S_r^2}{1+\sum_{k\in E_r}\frac{S_r^2}{\sigma_k^2}} \sum_{i\in E_r}\sum_{j\in E_r} \frac{\Delta_i\Delta_j}{\sigma_i^2\sigma_j^2}$

Fitting the extra parameters

Minimise $\chi^2 = \sum_r \sum_{i \in E_r} \left(\frac{\Delta_i - z_r}{\sigma_i}\right)^2 + \left(\frac{z_r}{S_r}\right)^2$

$\frac{\partial \chi^2}{\partial z_r} = 0 = -2 \sum_{i \in E_r} \frac{\Delta_i - z_r}{\sigma_i^2} + 2 \frac{z_r}{S_r^2}$

$z_r = \frac{\sum_{j \in E_r} \Delta_j / \sigma_j^2}{1/S_r^2 + \sum_{k \in E_r} \frac{1}{\sigma_k^2}}$

Put these back into the expression for χ^2 and get

$$\sum_r \sum_{i \in r} \left(\frac{\Delta_i - \frac{\sum_{j \in E_r} \Delta_j / \sigma_j^2}{1/S_r^2 + \sum_{k \in E_r} \frac{1}{\sigma_k^2}}}{\sigma_i} \right)^2 + S_r^2 \left(\frac{\sum_{j \in E_r} \Delta_j / \sigma_j^2}{1 + \sum_{k \in E_r} \frac{S_r^2}{\sigma_k^2}} \right)^2$$

Multiply out and collect terms in Δ_i^2 and $\Delta_i \Delta_j$ and get

$$\sum_r \sum_{i \in E_r} \frac{\Delta_i^2}{\sigma_i^2} - \sum_r \frac{S_r^2}{1 + \sum_{k \in E_r} \frac{S_r^2}{\sigma_k^2}} \sum_{i \in E_r} \sum_{j \in E_r} \frac{\Delta_i \Delta_j}{\sigma_i^2 \sigma_j^2}$$

We've seen that before

The two methods are equivalent

Maybe surprising. Maybe obvious.

Which to choose?

Doesn't matter.
Second method gives offsets, which may or may not be useful.
First method gives direct (non-iterative) solution for linear function.
Minimisation in second method requires care as there is a direction in parameter space adjusting c and all the z_r with small χ^2 dependence (Rosenbrock's Valley). Scrutinise 'solutions' from minimiser, or solve for z_r as you go.

Multiplicative Systematics

More common. Experiments have uncertainty in efficiency, or luminosity...

Matrix $V_{ij} = \delta_{ij}\sigma_i^2 + \xi_r^2 y_i y_j$
where ξ_r is shared fractional error
Or can minimise values scaled by z_r, and add $\left(\frac{z_r - 1}{\xi_r}\right)^2$ to χ^2
Ambiguity: do we scale y_i or $f(x_i)$?

Scaling the y_i gives the same results as the matrix inversion (similar algebra to last time)

But this has a problem...

"D'Agostini Bias"

Combining correlated results gives a biassed result

Simple illustration

Two measurements x_1, x_2 of the same quantity a with a shared error ξ

$$V = \begin{pmatrix} \sigma_1^2 + \xi^2 x_1^2 & \xi^2 x_1 x_2 \\ \xi^2 x_1 x_2 & \sigma_2^2 + \xi^2 x_2^2 \end{pmatrix}$$

$$\chi^2 \propto (x_1 - a, x_2 - a) \begin{pmatrix} \sigma_2^2 + \xi^2 x_2^2 & -\xi^2 x_1 x_2 \\ -\xi^2 x_1 x_2 & \sigma_1^2 + \xi^2 x_1^2 \end{pmatrix} \begin{pmatrix} x_1 - a \\ x_2 - a \end{pmatrix}$$

$$\hat{a} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2 + \xi^2 (x_1 - x_2)^2}$$

$$\langle \hat{a} \rangle < a$$

Bias is not due to correlation - but to multiplicative errors

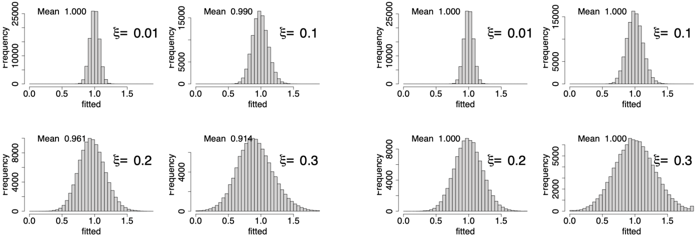
Sorted

Simple question

How should you average $90 \pm 10\%$ and $110 \pm 10\%$?

If you apply the z_r to the f_i (or use $V_{ij} = \xi_r^2 f(x_i) f(x_j)$) the bias is removed.

Toy MC results of averaging 2 correlated values using $\xi^2 y_i y_j$ (LEFT) and $\xi^2 f_i f_j$ (RIGHT)



Conclusions

Nucl. Instrum. and Meth. **A987** 164864 (2021) ,arXiv:1701.03701

You can use either obvious method for combining datasets

D'Agostini bias can be avoided, rather than corrected for

Conclusions

Nucl. Instrum. and Meth. **A987** 164864 (2021) ,arXiv:1701.03701

You can use either obvious method for combining datasets

D'Agostini bias can be avoided, rather than corrected for

Two early Christmas presents