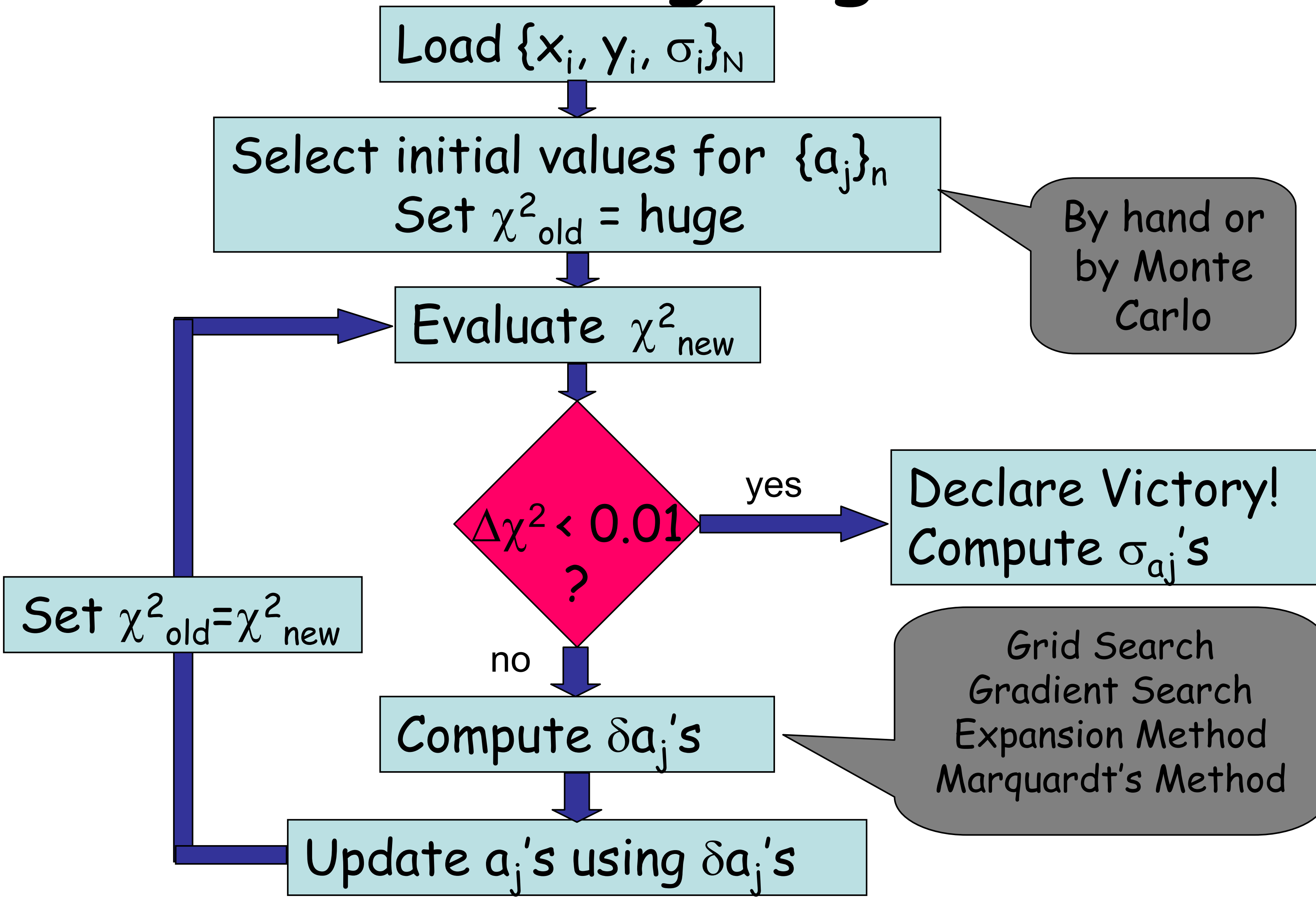
Week 9

More Issues in Non-linear Fitting

Generic Fitting Algorithm



Gradient Method pay-off:

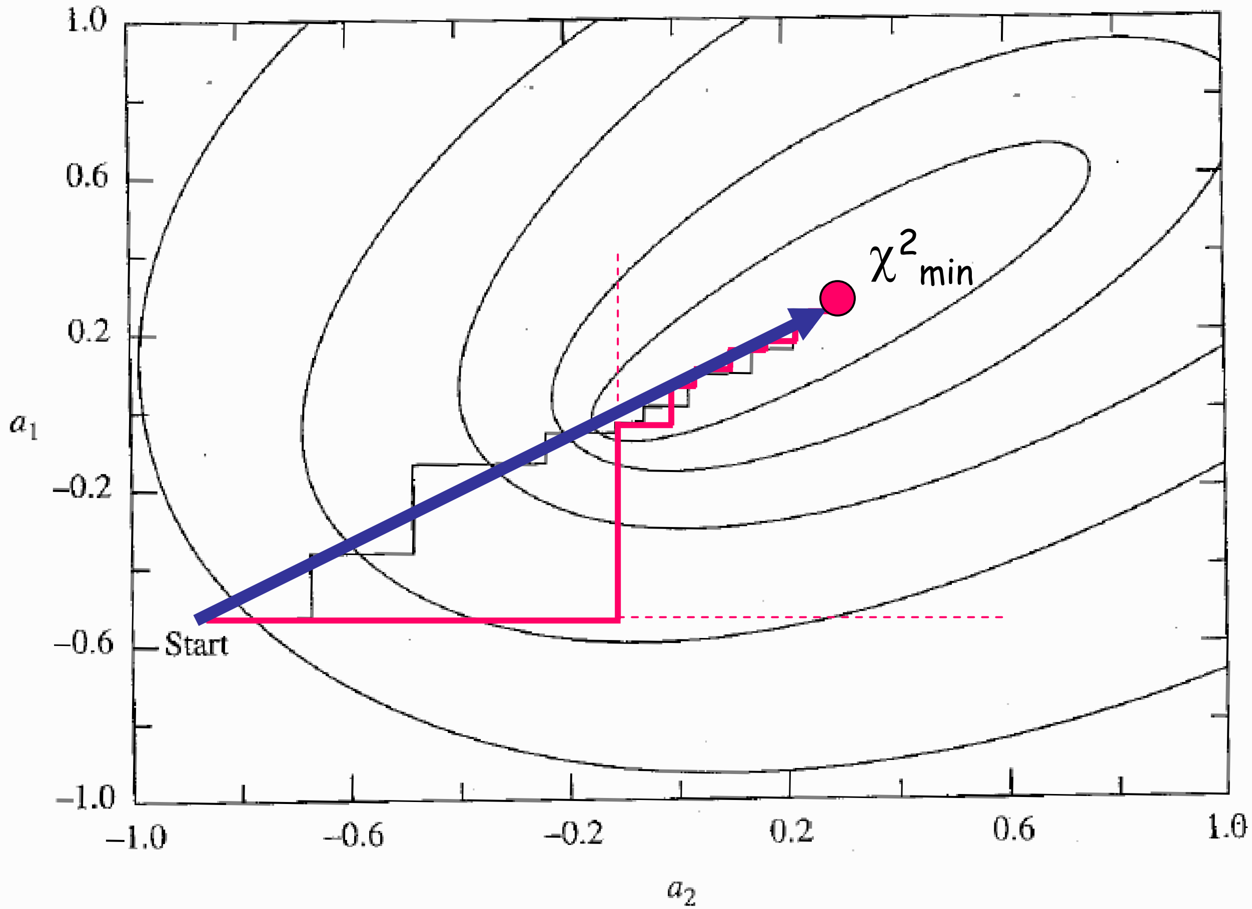


FIGURE 8.4
Contour plot of χ^2 as a function of two highly correlated variables. The zigzag line represents the search path approach to a local minimum by the grid-search method.

Gradient Search (Bevington 8.4; pp.153)

- Find the "steepest" descent down the valley of χ^2 by computing the normalized gradient
  - Gradient always computable numerically
  - Sometimes can be computed algebraically
- Pro's and Con's:
  - Fast approach to the minimum from far away
  - Very slow near the minimum where the gradient is near zero

Summary from last time

- Linear vs. non-linear fitting functions
  - Linear functions can be fitted in one algebraic iteration
  - Non-linear function require iterated solutions when fitting to data
- Grid search is simple but tends to be slow
- Gradient search is fast when far away from minimum but slow near the minimum
- We need a method that is fast near the minimum…next time

Expansion Method (cf. Bevington 8.5; pp156)

Exploit the quadratic nature of χ^2 near its minimum. Expand to 2nd order around whatever point in parameter space we are sitting, where $\chi^2 = \chi^2_0$:

$$\chi^2 \simeq \chi_0^2 + \sum_{j=1}^{m} \left\{ \frac{\partial \chi_0^2}{\partial a_j} \delta a_j \right\} + \frac{1}{2} \sum_{k=1}^{m} \sum_{j=1}^{m} \left\{ \frac{\partial^2 \chi_0^2}{\partial a_j \partial a_k} \delta a_j \delta a_k \right\} \qquad (8.22)$$

Minimize χ^2 with respect to the parameter increments δa_j, not the parameters themselves. Why? Because this linearizes the problem and it can be solved exactly.

$$\frac{\partial \chi^2}{\partial(\delta a_k)} = \frac{\partial \chi_0^2}{\partial a_k} + \sum_{j=1}^{m} \left\{ \frac{\partial^2 \chi_0^2}{\partial a_k \partial a_j} \delta a_j \right\} = 0 \quad k = 1, m$$

This is a set of m linear equations in m unknowns!

$$\beta_k \equiv -\frac{1}{2} \frac{\partial \chi_0^2}{\partial a_k}$$

$$\alpha_{jk} \equiv \frac{1}{2} \frac{\partial^2 \chi_0^2}{\partial a_j \partial a_k}$$

"Curvature Matrix" measures curvature of χ^2 hypersurface

Expansion Method

Matrix equation to solve: $\vec{\beta} = \delta\vec{a}\,\overleftrightarrow{\alpha}$

Invert the matrix α to get the matrix ε, the so-called "error matrix" $\underline{\varepsilon = \alpha^{-1}}$:

$$\boxed{\delta\vec{a} = \vec{\beta}\,\overleftrightarrow{\varepsilon}} \qquad \delta a_k = \sum_{j=1}^{m} (\epsilon_{kj}\beta_j)$$

How to compute the matrix α and the vector β?

$$\beta_k \equiv \sum \left\{ \frac{1}{\sigma_i^2} [y_i - y'(x_i)] \frac{\partial y'(x_i)}{\partial a_k} \right\} = -\frac{1}{2} \frac{\partial \chi_0^2}{\partial a_k}$$

$$\alpha_{jk} \equiv \sum \frac{1}{\sigma_i^2} \left\{ \frac{\partial y'(x_i)}{\partial a_j} \frac{\partial y'(x_i)}{\partial a_k} - [y_i - y'(x_i)] \frac{\partial^2 y'(x_i)}{\partial a_j \partial a_k} \right\}$$

$$= \frac{1}{2} \frac{\partial^2 \chi_0^2}{\partial a_j \partial a_k}$$

Product of first derivatives

Should average to zero; neglect

The "error matrix"

$$\alpha^{-1} = \varepsilon = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \cdots \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \\ \sigma_{31}^2 & \vdots & \sigma_{33}^2 & \\ \vdots & & & \ddots \end{pmatrix}$$

$$\sigma_{jj} = \text{"diagonal errors"} \equiv \sigma_{\mathrm{j}}$$

$$\sigma_{jk} = \text{"correlated errors"} \; ; \; \sigma_{jk}^2 = \text{"covariances"}$$

Example...

Marquardt(-Levenberg) Method

(cf. Bevington 8.6; pp161)

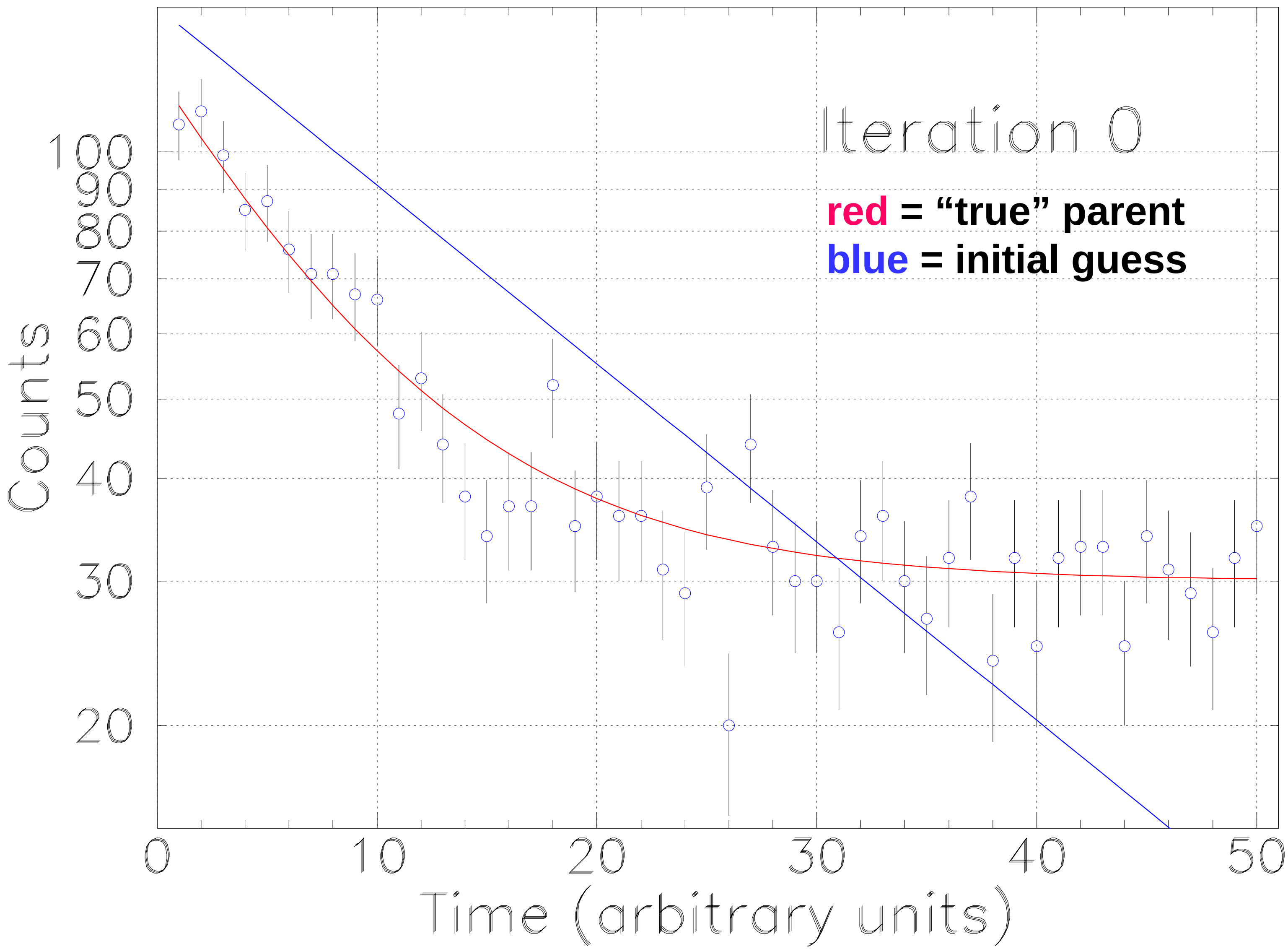
$$\beta = \delta a \, \alpha' \quad \text{with} \quad \alpha'_{jk} = \begin{cases} \alpha_{jk}(1+\lambda) & \text{for } j = k \\ \alpha_{jk} & \text{for } j \neq k \end{cases}$$

- For small λ it behaves like the expansion method.
- For large λ, fit behaves like the gradient method with small steps, since off-diagonals are negligible.
- Make λ just large enough such that χ^2 decreases.

The "recipe" for fast fit convergence:

1. Compute $\chi^2(a)$.
2. Start initially with $\lambda = 0.001$.
3. Compute δa and $\chi^2(a + \delta a)$ with this choice of λ.
4. If $\chi^2(a + \delta a) > \chi^2(a)$, increase λ by a factor of 10 and repeat step 3.
5. If $\chi^2(a + \delta a) < \chi^2(a)$, decrease λ by a factor of 10, consider $a' = a + \delta a$ to be the new starting point, and return to step 3, substituting a' for a.

Marquardt Method Fit

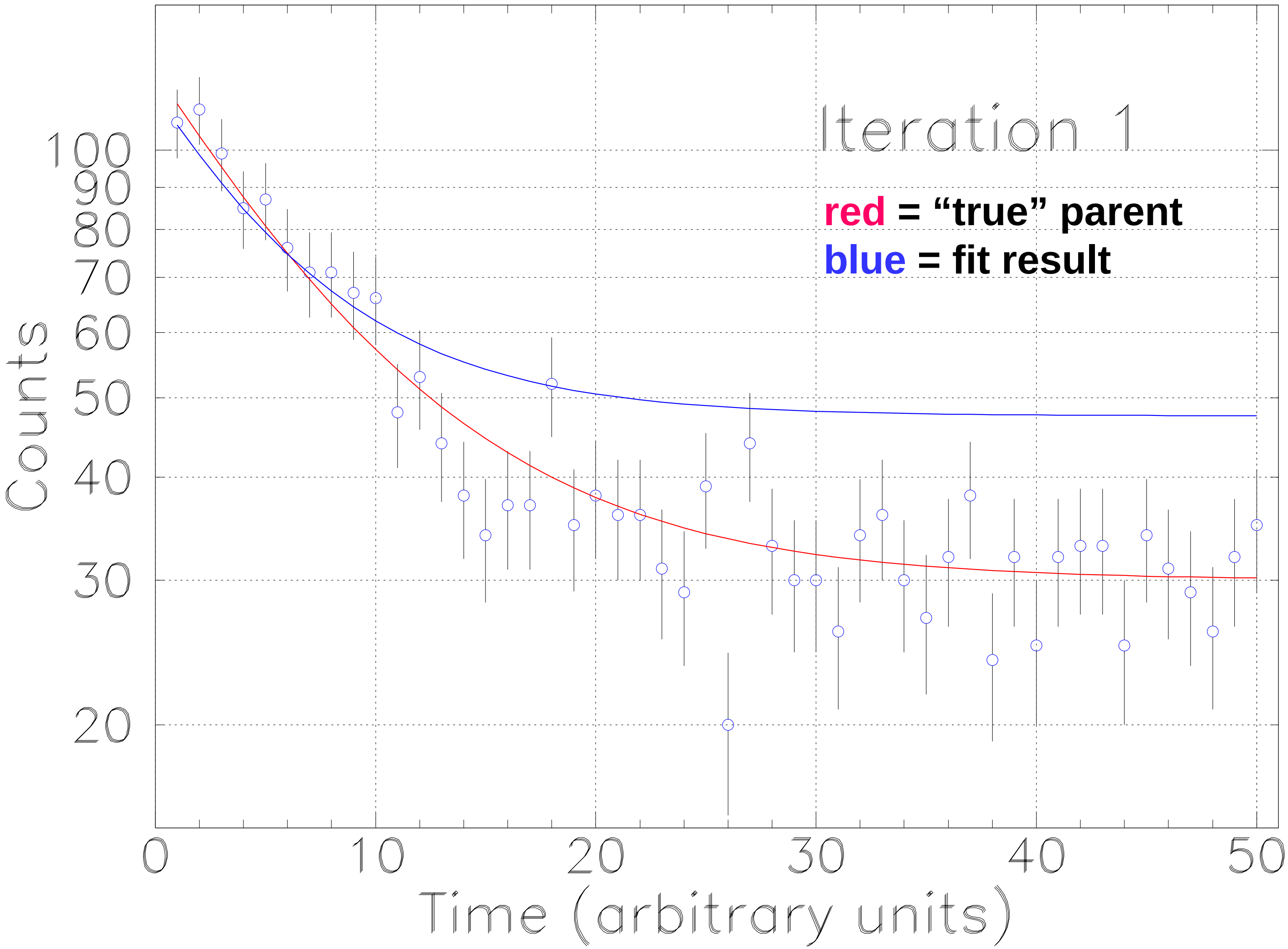


Numbers…

$$N(t) = N_0 e^{-t/\tau} + B$$

Trial	χ^2/d.o.f.	N_0 a_1	τ a_2	B a_3	
------	----------	------	------	------	
0	10.364	150.0	20.0	0.0	Value
1	8.030	69.9	6.4	47.5	Value
		4.4	1.8	4.3	Uncertainty
2	0.841	100.6	8.1	29.1	Value
		9.0	1.0	1.2	Uncertainty
3	0.796	103.4	7.2	29.3	Value
		7.6	0.8	1.3	Uncertainty
4	0.795	102.9	7.4	29.1	Value
		8.2	0.8	1.2	Uncertainty

Marquardt Method Fit

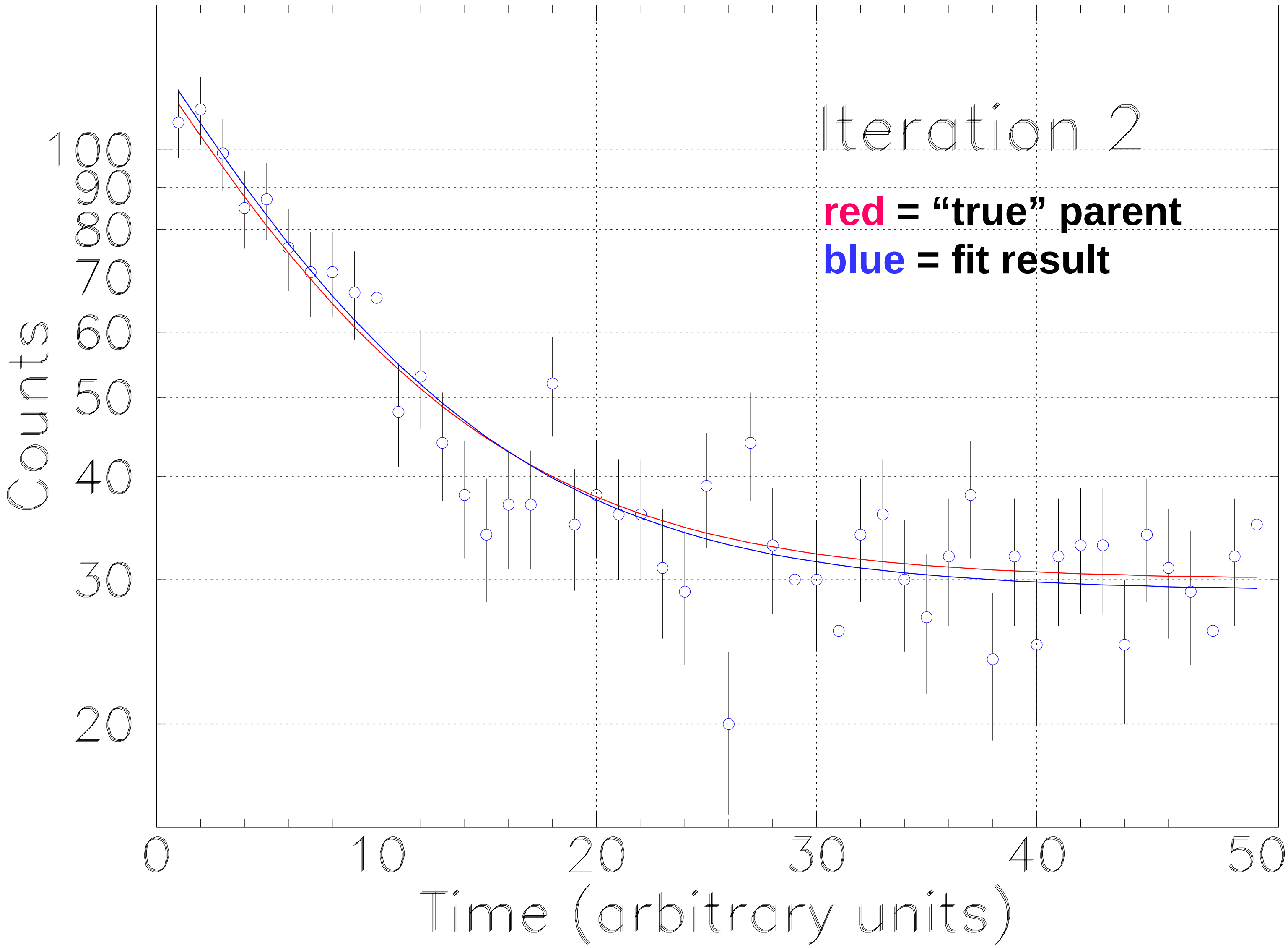


Numbers…

$$N(t) = N_0 e^{-t/\tau} + B$$

Trial	χ^2/d.o.f.	N_0 a_1	τ a_2	B a_3	
------	----------	------	------	------	
0	10.364	150.0	20.0	0.0	Value
1	8.030	69.9	6.4	47.5	Value
		4.4	1.8	4.3	Uncertainty
2	0.841	100.6	8.1	29.1	Value
		9.0	1.0	1.2	Uncertainty
3	0.796	103.4	7.2	29.3	Value
		7.6	0.8	1.3	Uncertainty
4	0.795	102.9	7.4	29.1	Value
		8.2	0.8	1.2	Uncertainty

Marquardt Method Fit

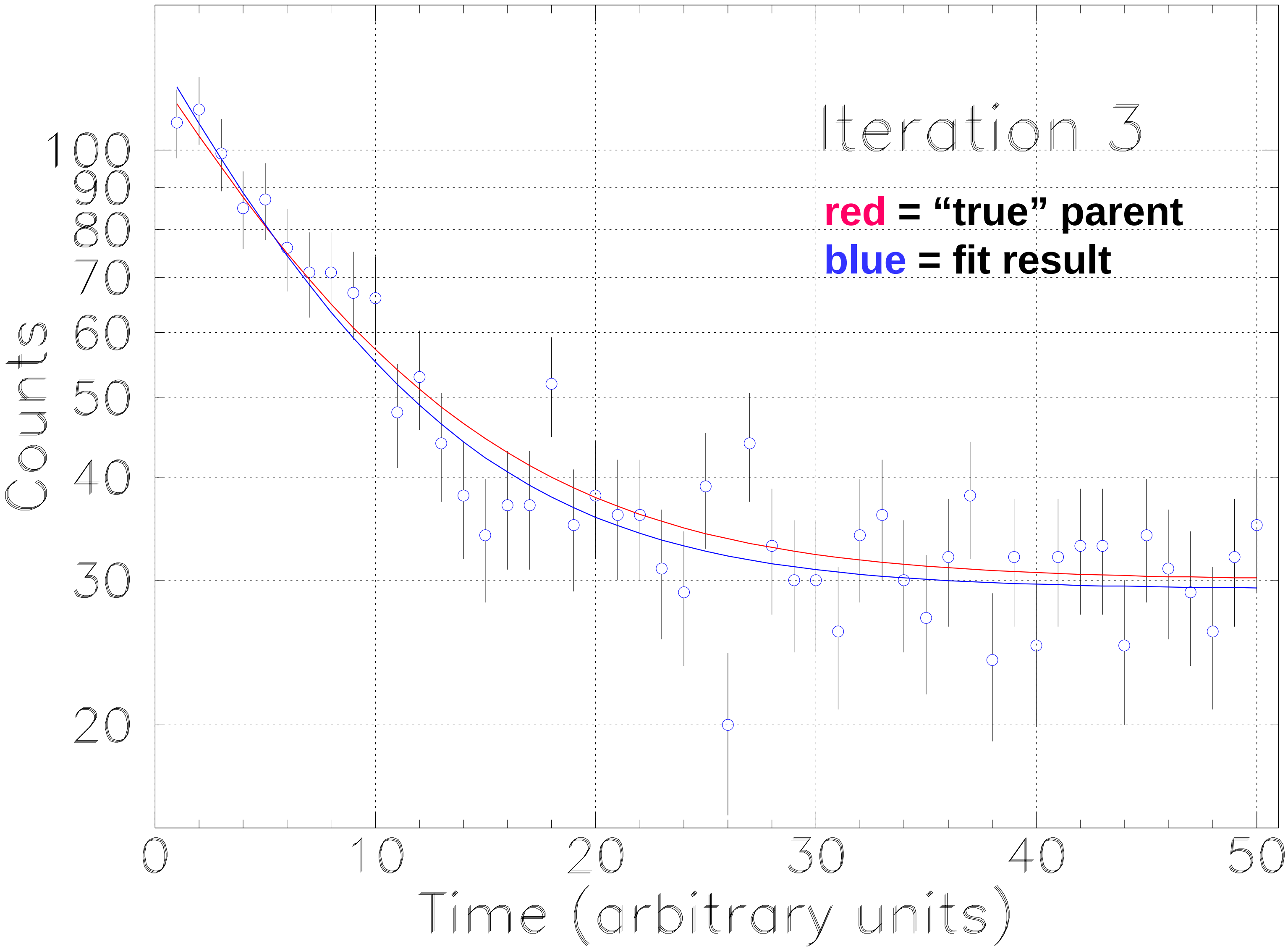


Numbers…

Trial	χ^2/d.o.f.	N_0 a_1	τ a_2	B a_3	
0	10.364	150.0	20.0	0.0	Value
1	8.030	69.9	6.4	47.5	Value
		4.4	1.8	4.3	Uncertainty
2	0.841	100.6	8.1	29.1	Value
		9.0	1.0	1.2	Uncertainty
3	0.796	103.4	7.2	29.3	Value
		7.6	0.8	1.3	Uncertainty
4	0.795	102.9	7.4	29.1	Value
		8.2	0.8	1.2	Uncertainty

$$N(t) = N_0 e^{-t/\tau} + B$$

Marquardt Method Fit

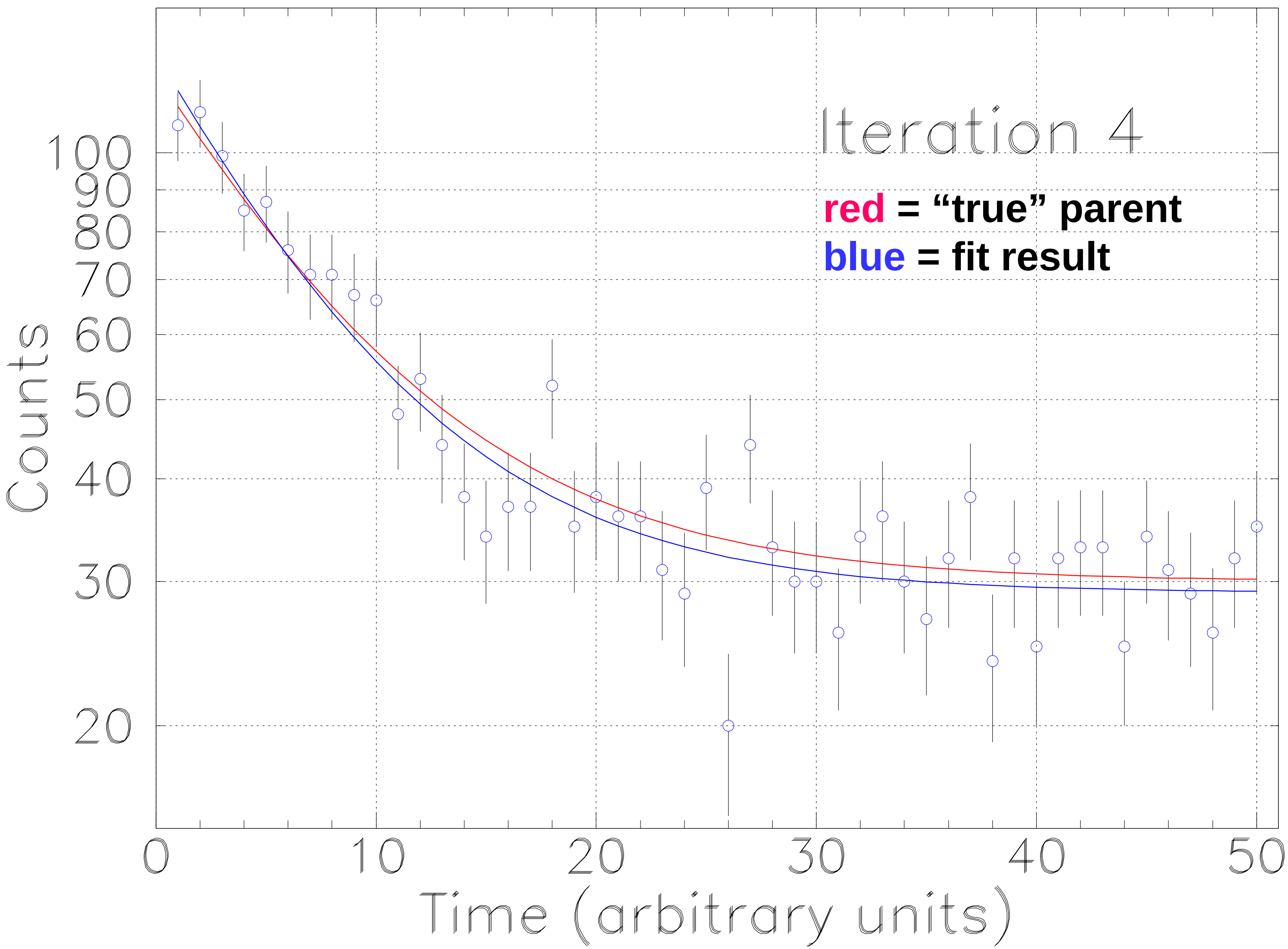


Numbers…

$$N(t) = N_0 e^{-t/\tau} + B$$

Trial	χ^2/d.o.f.	N_0 a_1	τ a_2	B a_3	
0	10.364	150.0	20.0	0.0	Value
1	8.030	69.9	6.4	47.5	Value
		4.4	1.8	4.3	Uncertainty
2	0.841	100.6	8.1	29.1	Value
		9.0	1.0	1.2	Uncertainty
3	0.796	103.4	7.2	29.3	Value
		7.6	0.8	1.3	Uncertainty
4	0.795	102.9	7.4	29.1	Value
		8.2	0.8	1.2	Uncertainty

Marquardt Method Fit



Numbers…

$$N(t) = N_0 e^{-t/\tau} + B$$

Trial	χ^2/d.o.f.	N_0 a_1	τ a_2	B a_3	
------	----------	------	------	------	
0	10.364	150.0	20.0	0.0	Value
1	8.030	69.9	6.4	47.5	Value
		4.4	1.8	4.3	Uncertainty
2	0.841	100.6	8.1	29.1	Value
		9.0	1.0	1.2	Uncertainty
3	0.796	103.4	7.2	29.3	Value
		7.6	0.8	1.3	Uncertainty
4	0.795	102.9	7.4	29.1	Value
		8.2	0.8	1.2	Uncertainty
		95	8	30	← The “real” parent numbers

Covariance Among Parameters

Covariance Matrix

		N_0 **a**$_1$	τ **a**$_2$	B **a**$_3$
		------	------	------
N_0	**a**$_1$)	**67.64**	**-4.35**	**1.84**
τ	**a**$_2$)	**-4.35**	**0.58**	**-0.59**
B	**a**$_3$)	**1.84**	**-0.59**	**1.55**

$$\alpha^{-1} = \varepsilon = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \cdots \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \\ \sigma_{31}^2 & \vdots & \sigma_{33}^2 & \\ \vdots & & & \ddots \end{pmatrix}$$

$$\sigma_{N_0} = \sqrt{67.64} = 8.2, \text{ etc...}$$

The off-diagonal terms specify the covariances or "correlations" among fit parameters.

Covariance Among Parameters

"Reduced" Covariance Matrix

$$
\begin{array}{ll|ccc}
 & & N_0 & \tau & B \\
 & & \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} \\
 & & \text{------} & \text{------} & \text{------} \\
N_0 & \mathbf{a_1)} & \mathbf{1.000} & & \\
\tau & \mathbf{a_2)} & \mathbf{-0.112} & \mathbf{1.000} & \\
B & \mathbf{a_3)} & \mathbf{0.018} & \mathbf{-0.664} & \mathbf{1.000}
\end{array}
$$

$$\varepsilon_{ij}^{'} = \frac{\varepsilon_{ij}}{\varepsilon_{ii}\varepsilon_{jj}}, i \neq j$$

We see that the lifetime and background rate are strongly correlated.

Summary

- Expansion method uses quadratic approx. of the χ^2 function near the minimum to zero in on the minimum
- Marquardt method combines the best features of the gradient and expansion methods to fit a function to data quickly
- Covariance matrix of parameter errors is one outcome of the expansion method algebra.

Backup Slides…