Data Analysis and Fitting: Errors and Goodness of Fit

Ashton S. Reimer

[1]Center for Geospace Studies
SRI International

July 2020

Topics

1. Errors
2. Goodness of Fit

Chi-Squared

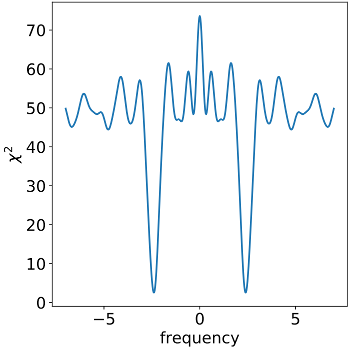
We can use least-squares to solve inverse problems:

$$\chi^2(\mathbf{p}) = [\mathbf{z} - f(\mathbf{p})]^T \, \mathbf{\Sigma_e}^{-1} \, [\mathbf{z} - f(\mathbf{p})]$$

where $\hat{\mathbf{p}}_{LS}$ are the "best-fit" model parameters, those that minimizes $\chi^2(\mathbf{p})$

Great! But:

- What are the errors in the fitted parameters $\hat{\mathbf{p}}_{LS}$?
- Is the fit meaningful? Does the model accurately reproduce the measurements?



Error Propagation (e.g. Linear Least-Squares)

For a linear forward model:

$$\mathbf{z} = f(\mathbf{p}) + \mathbf{e} \qquad f(\mathbf{p}) = H\mathbf{p}$$

The Least-Squares solution is:

$$\hat{\mathbf{p}}_{LS} = \left[H^T \boldsymbol{\Sigma}_e^{-1} H\right]^{-1} H^T \boldsymbol{\Sigma}_e^{-1} \mathbf{z}$$

Given that jointly Gaussian random variables have the following property:

$$\mathbf{Y} = A\mathbf{X} \quad \Rightarrow \quad \boldsymbol{\Sigma}_{\mathbf{Y}} = A \boldsymbol{\Sigma}_{\mathbf{X}} A^T$$

it can be shown that:

$$\boldsymbol{\Sigma}_{\hat{\mathbf{p}}_{\mathrm{LS}}} = \left[H^T \boldsymbol{\Sigma}_e^{-1} H\right]^{-1}$$

Error Propagation (e.g. Nonlinear Least Squares)

For a non-linear forward model, guess a $\mathbf{p}_i$, linearize, and step towards minimum:

$$\mathbf{z} = f(\mathbf{p}) + \mathbf{e} \qquad f(\mathbf{p}_i + \Delta\mathbf{p}) \approx f(\mathbf{p}_i) + \mathbf{J}_i \Delta\mathbf{p} \qquad \mathbf{J}_i = \frac{\partial f}{\partial \mathbf{p}_i}$$

$\mathbf{J}$ is known as the Jacobian:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_0}{\partial p_0} & \frac{\partial f_0}{\partial p_1} & \cdots & \frac{\partial f_0}{\partial p_{N-1}} \\ \frac{\partial f_1}{\partial p_0} & \frac{\partial f_1}{\partial p_1} & \cdots & \frac{\partial f_1}{\partial p_{N-1}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{M-1}}{\partial p_0} & \frac{\partial f_{M-1}}{\partial p_1} & \cdots & \frac{\partial f_{M-1}}{\partial p_{N-1}} \end{pmatrix}$$

$\mathbf{J}$ is $M \times N$ (tall and skinny)

Non-linear fitting process:

- iterate until $\mathbf{p}_{i+1} = \hat{\mathbf{p}}_{\mathrm{LS}}$: that which minimizes χ^2
- The covariance of $\hat{\mathbf{p}}_{\mathrm{LS}}$ is:

$$\mathbf{\Sigma}_{\hat{\mathbf{p}}_{\mathrm{LS}}} = \left[\mathbf{J}^T \mathbf{\Sigma}_e^{-1} \mathbf{J}\right]^{-1}$$

Note the similarity to the linear case!

Error Propagation

The covariance of the fitted parameters is the covariance of the input data propagated through the least-squares operation:

$$\boldsymbol{\Sigma}_{\hat{\mathbf{p}}_{\mathrm{LS}}} = \left[\mathbf{J}^T \boldsymbol{\Sigma}_e^{-1} \mathbf{J}\right]^{-1}$$

"Error bars" for fitted parameters:

- Assumption: measurement errors are **normally** distributed with covariance $\boldsymbol{\Sigma}_e$, denoted $\mathcal{N}(0, \boldsymbol{\Sigma}_e)$
- The "errors" in the fitted parameters are related to confidence intervals
- Confidence intervals are constructed from $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}_{\mathrm{LS}}}$
- $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}_{\mathrm{LS}}}$ may look reasonable, even if the fit is meaningless

Constructing Confidence Intervals: From Fitted Covariance

Error bars, δp_m, for a fitted parameter can be constructed from the covariance $\mathbf{\Sigma}_{\hat{\mathbf{p}}_{\mathrm{LS}}}$ and a $\Delta\chi^2$:

$$\delta p_m = \pm\sqrt{\Delta\chi^2}\sqrt{\Sigma_{mm}}$$

The value of $\Delta\chi^2$ selects the "significance level", α, for the error bars:

$$\alpha = \mathcal{P}\left(\frac{N}{2}, \frac{\Delta\chi^2}{2}\right)$$

where $\mathcal{P}$ is the Regularized Gamma Function (and CDF of χ^2 dist.):

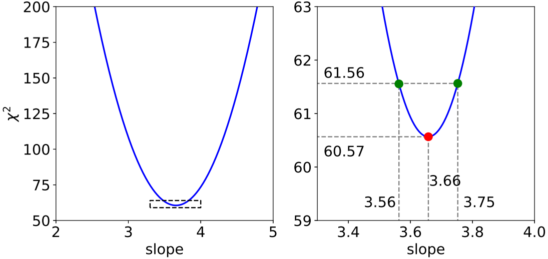
$$\mathcal{P}(s,x) = 1 - \Gamma(s,x)/\Gamma(s,0), \quad \Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$$

e.g. For a 68% significance for, $\alpha = 0.68$ and $N = 1$, corresponds to $\Delta\chi^2 = 1$. For 95.4%: $\Delta\chi^2 = 4$ and for 99.73%: $\Delta\chi^2 = 9$.

Constructing Confidence Intervals: From Chi-Squared

Equivalent method:

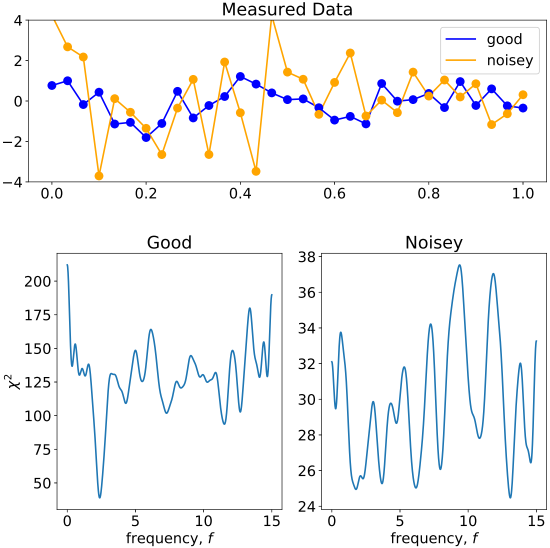
- Use $\chi^2(\mathbf{p})$ directly to construct $\delta\mathbf{p}$.
- In the figure, χ^2 for vs. the "slope" parameter



$$\delta p_{upper,lower} = |(p_m @ \chi^2_{min} + \Delta\chi^2) - (p_m @ \chi^2_{min})|$$

- $N = 1$, so $\Delta\chi^2 = 1$. Taking values the figure: $\delta p_{slope} \approx \pm 0.095$
- Using covariance from fit ($\Sigma_{mm} = 0.00905$):
  $\delta p_{slope} = \pm\sqrt{\Delta\chi^2}\sqrt{\Sigma_{mm}} \approx \pm 0.095$

Challenges With Constructing Confidence Intervals



Validity of Confidence Intervals

Only quantitatively valid when:

- measurement errors are Gaussian, and
  - the model $f(\mathbf{p})$ is linear in for all $\mathbf{p}$, or
  - measurement errors are small enough that $f(\mathbf{p})$ can be accurate approximated by a linear model in the region around $\mathbf{p}$

Otherwise, alternative fitting methods are required: Monte Carlo, Bayesian, etc.

Goodness of Fit

How do we know if the fit is even meaningful? The standard goodness of fit test involves computing the "reduced chi-squared":

$$\chi^2_\nu = \chi^2/(m - n + 1)$$

Then, typically:

- $\chi^2_\nu \approx 1$: a good fit
- $\chi^2_\nu << 1$: an "over fit"
- $\chi^2_\nu >> 1$: a poor fit

The χ^2_ν could also be slightly larger or smaller than 1 depending on how accurately one is able to estimate the input measurement errors.

Summary

Now we can answer the question: Are the fitted parameters meaningful?

- What is the uncertainty in the fitted parameters?
  - Error bars correspond to confidence intervals (CI)
  - CIs are constructed from covariance of the fitted parameters
  - For a 68% CI, interpretation is: “If we could hypothetically make and infinite set of new measurements and fit each of those, 68% of the time the 'true' value of the parameter would lie within the CI.”
- Is the fit good?
  - Compute the reduced chi-squared
  - $\chi^2_\nu \approx 1$: usually means the model accurately represents the data
- All of this error analysis depends on the assumption that $(z_m - f_m)/\sigma_m$ are $\mathcal{N}(0, 1)$