Lecture 9: Least squares uncertainties

Recap

In Lecture 8, we looked at some specific examples of least-squares fitting, specifically focused on setting up inversion problems. In this lecture, we'll expand our repertoire by considering additional constraints, starting by looking at the linear regression case when our dependent variable (e.g. time or position) has uncertainties.

Linear regression with uncertain variables

In class we examined results using "ordinary least squares" compared with results based on the "standard major axis" approach (see Bellacicco et al, 2019). In a classic least-squares problem we define a model:

$$\mathbf{y} = \mathbf{G}\mathbf{m} + \mathbf{n} \tag{1}$$

where

$$\hat{\mathbf{y}} = \mathbf{G}\mathbf{m} \tag{2}$$

where $\mathbf{n}$ is the vector of the noise or misfit, which we aim to minimize. We minimize the cost function:

$$\epsilon = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3}$$

to obtain the standard least-squares solution. For a linear regression that finds a constant and a slope, of the form $\hat{\mathbf{y}} = m_1 + m_2\mathbf{x}$, and the matrix $\mathbf{G}$ is:

$$\mathbf{G} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}. \tag{4}$$

The standard least-squares solution gives us

$$\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{y}, \tag{5}$$

which gives

$$m_1 = \langle \mathbf{y} \rangle - m_2 \langle \mathbf{x} \rangle \tag{6}$$

$$m_2 = \frac{C_{xy}}{C_{xx}}, \tag{7}$$

where C_{xy} is the covariance of x and y, in the form $\langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle$.

The standard major axis method assumes that both x and y have uncertainties so minimizes the area of a triangle between the data point y and the line defining $\hat{y}$. The produces:

$$m_1 = \langle \mathbf{y} \rangle - m_{2_{OLS}} \langle \mathbf{x} \rangle \tag{8}$$

$$m_2 = \sqrt{\frac{C_{yy}}{C_{xx}}} = \pm\frac{s_y}{s_x}, \tag{9}$$

where s_y is the estimated standard deviation of y and x_s is the estimated standard deviation in x. Check the appendix of Bellacicco et al (2019) for details.

Uncertainties in model parameters

All of the above discussion is a temporary digression. In general, when we compute a least squares fit, we probably want to know uncertainties for our fitted parameters m_i. If our data have a known covariance, we can define a weight matrix:

$$\mathbf{W} = \langle \mathbf{d}\mathbf{d}^T \rangle = \sigma^2 \mathbf{I}. \tag{10}$$

If we weight each line of our matrix equation by the uncertainty in the data, we have

$$\mathbf{d}\mathbf{W}^{-1/2} = \mathbf{W}^{-1/2}\mathbf{G}\mathbf{m} \tag{11}$$

As we noted before, this yields

$$\mathbf{m} = (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}^{-1}\mathbf{d}. \tag{12}$$

We can also estimate the covariance of $\mathbf{m}$:

$$\begin{aligned}
\langle \mathbf{m}\mathbf{m}^T \rangle &= \langle (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}^{-1}\mathbf{d}((\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}^{-1}\mathbf{d}\mathbf{d})^T \rangle & (13)\\
&= (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}^{-1}\langle \mathbf{d}\mathbf{d}^T \rangle \mathbf{W}^{-1}\mathbf{G}(\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1^T} & (14)\\
&= (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}^{-1}\mathbf{W}\mathbf{W}^{-1}\mathbf{G}(\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1^T} & (15)\\
&= (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}(\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})(\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1^T} & (16)\\
&= (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}(\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})(\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1^T} & (17)\\
&= (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}. & (18)
\end{aligned}$$

This is conveniently just the matrix that we were inverting, and it tells us that the inverted matrix, weighted appropriately by the uncertainties, will provide the uncertainties in $\mathbf{m}$.

Let's test this out in the simplest possible case. Consider the case where our inversion is simply used to find the mean of $\mathbf{d}$. We define an $N \times 1$ matrix:

$$\mathbf{G} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{19}$$

and

$$\mathbf{W} = \sigma^2 \mathbf{I} \tag{20}$$

The standard least-squares solution gives us

$$
\begin{aligned}
\mathbf{m} &= (\mathbf{G}^T\mathbf{W}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}^{-1}\mathbf{d} &(21)\\
&= \left(\begin{bmatrix}1 & 1 & \ldots & 1\end{bmatrix}\sigma^{-2}\mathbf{I}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}\right)^{-1}\begin{bmatrix}1 & 1 & \ldots & 1\end{bmatrix}\sigma^{-2}\mathbf{Id} &(22)\\
&= \left(\sum_{i=1}^{N}\frac{1}{\sigma^2}\right)^{-1}\sum_{i=1}^{N}\frac{d_i}{\sigma^2} &(23)\\
&= \overline{\mathbf{d}}, &(24)
\end{aligned}
$$

where the overbar indicates the mean. The uncertainty in this estimate comes from the inverted matrix:

$$
\begin{aligned}
\langle\mathbf{mm}^T\rangle &= \left(\sum_{i=1}^{N}\frac{1}{\sigma^2}\right)^{-1} &(25)\\
&= \left(\frac{N}{\sigma^2}\right) &(26)\\
&= \sigma^2 N. &(27)
\end{aligned}
$$

This wonderfully shows us that the uncertainty in our estimate m_1 of the mean is the standard error of the mean, $\sigma/\sqrt{N}$.

Example

In class, we considered one example.

1. *In an annual record, data* ($\mathbf{y}$) *collected in summer are more accurate than data collected in winter. How do we represent that?*
   To account for varying accuracy in our data, we'll want to adjust our weight matrix $\mathbf{W}_e$ to have differing values of σ_i^{-2} on the diagonals. Smaller σ_i in summer imply larger weights. Note that overweighting summer values could mess up our estimates of the annual mean and annual cycle, so we would need to scrutinize our results fairly carefully.

In class discussion, we noted that formally we would create a matrix $\mathbf{W}$ of the form:

$$
\mathbf{W} = \begin{bmatrix}
\sigma_i^2 & 0 & 0 & \ldots & 0\\
0 & \sigma_2^2 & 0 & \ldots & 0\\
0 & 0 & \sigma_i^2 & \ldots & 0\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
0 & 0 & 0 & \ldots & \sigma_N^2
\end{bmatrix} \qquad (28)
$$

In this case, σ_i should be large in the winter and small in the summer, and we talked about ways it could vary—either over time blocks or perhaps following a sinusoidal variation. Formally, we might want $\mathbf{W}$ to contain off-diagonal elements, but we often can assume data to be uncorelated, allowing us to omit the covariance terms ($\langle d_i d_j \rangle$), which is good, since they would make $\mathbf{W}$ difficult to invert. $\mathbf{W}$ is an $N \times N$ matrix. Although it's quick to invert, we still might not want to fill the full matrix, and we might be able to speed up the calculation by dividing each d_i by σ_i and each row of $\mathbf{G}$ (i.e. $G_{i,j}$) by σ_i.

The corresponding matrix $\mathbf{G}$ should at a minimum solve for a mean and a diurnal cycle:

$$\mathbf{G} = \begin{bmatrix} 1 & \cos(2\pi t_1) & \sin(2\pi t_1) \\ 1 & \cos(2\pi t_2) & \sin(2\pi t_2) \\ \vdots & \vdots & \vdots \\ 1 & \cos(2\pi t_N) & \sin(2\pi t_N) \end{bmatrix}. \tag{29}$$

where t_i is time in days. Depending on the data, we might also fit to additional functions to account for a linear trend or an annual cycle, for example. We only require $M < N$, so that the problem is overdetermined.

There are challenges intrinsic in this problem—notably that downweighting the winter data will bias the solution to summer conditions. You could address this by (a) solving separately for winter and summer, or (b) using the winter uncertainties for the entire record, or (c) tolerating the fact that the solution was based on summer conditions only.

Bellacicco, Marco, Vincenzo Vellucci, Michele Scardi, Marie Barbieux, Salvatore Marullo, and Fabrizio D'Ortenzio. 2019. Quantifying the Impact of Linear Regression Model in Deriving Bio-Optical Relationships: The Implications on Ocean Carbon Estimations, *Sensors* 19, no. 13: 3032. https://doi.org/10.3390/s19133032