Combining imperfect data, and an introduction to data assimilation

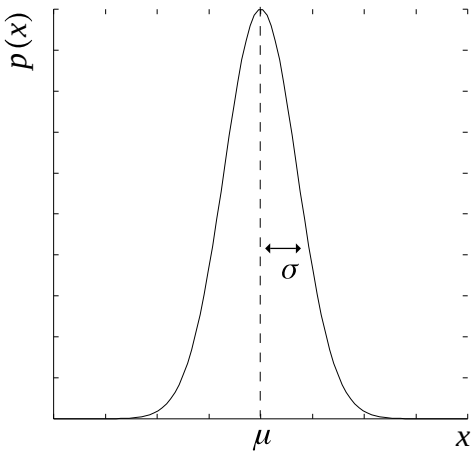
Ross Bannister, NCEO, September 2010
r.n.bannister@reading.ac.uk

The probability density function (PDF)

prob. that x lies between x and $x + dx$	$p(x)$
restriction on $p(x)$	$\int_{x=-\infty}^{+\infty} dx\, p(x) = 1$
expectation value of $f(x)$	$\int_{x=-\infty}^{+\infty} dx\, f(x) p(x) = \langle f(x) \rangle$
expectation value of x (the mean)	$\int_{x=-\infty}^{+\infty} dx\, x p(x) = \langle x \rangle$
jth moment of x around $\langle x \rangle$	$\int_{x=-\infty}^{+\infty} dx\, (x - \langle x \rangle)^j p(x) = \langle (x - \langle x \rangle)^j \rangle$
$j = 1$ moment	$\int_{x=-\infty}^{+\infty} dx\, (x - \langle x \rangle) p(x) = \langle x - \langle x \rangle \rangle = 0$
$j = 2$ moment (the variance)	$\int_{x=-\infty}^{+\infty} dx\, (x - \langle x \rangle)^2 p(x) = \sigma_x^2$

The Gaussian (or normal) distribution is a commonly used example of $p(x)$

$$p(x) = N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-\mu)^2}{2\sigma^2}$$



For these notes, x may be considered to be a measurement of some variable which is subject to a normally distributed error with standard deviation σ. If the measurement error is unbiased, then the mean, μ, is the true value.

The PDF for a number of imperfect observations

No measurement is exact, and so all measurements have error. The error is unmeasureable, but we assume that we know its statistics (the PDF). We wish to combine N unbiased, normally distributed measurements to estimate the true value, and its uncertainty.

Let the nth measurement be x_n, and let the possible true value be x. The PDF of this measurement is

$$p_n(x_n \mid x_e) = \frac{1}{\sigma_n\sqrt{2\pi}} \exp -\frac{(x_n - x)^2}{2\sigma_n^2}.$$

The notation $p_n(x_n \mid x)$ means the probability that measurement x_n lies between x_n and $x_n + dx_n$

given that the true value is x. The combined PDF for N measurements of the same quantity is

$$\begin{aligned} p(x_1, x_2, \ldots x_N \mid x) &= p_1(x_1 \mid x) p_2(x_2 \mid x) \ldots p_N(x_N \mid x), \\ &= \prod_{n=1}^{N} p_n(x_n \mid x), \\ &= \prod_{n=1}^{N} \frac{1}{\sigma_n \sqrt{2\pi}} \exp -\frac{(x_n - x)^2}{2\sigma_n^2}, \\ &= \frac{1}{(2\pi)^{N/2}} \left(\prod_{n=1}^{N} \frac{1}{\sigma_n} \right) \left(\prod_{n=1}^{N} \exp -\frac{(x_n - x)^2}{2\sigma_n^2} \right). \end{aligned}$$

When considered a function of x, this PDF is called a likelihood function. We wish to calculate the value of x that maximizes this likelihood (the maximum likelihood estimate, x_e).

The x_e that maximizes $p(x_1, x_2, \ldots x_N \mid x_e)$ is the same x_e that maximizes $\ln p(x_1, x_2, \ldots x_N \mid x_e)$

$$\ln p(x_1, x_2, \ldots x_N \mid x_e) = \ln \frac{1}{(2\pi)^{N/2}} - \sum_{n=1}^{N} \sigma_n - \sum_{n=1}^{N} \frac{(x_n - x_e)^2}{2\sigma_n^2}.$$

The x_e that maximizes $\ln p(x_1, x_2, \ldots x_N \mid x_e)$ is the same x that minimizes $-\ln p(x_1, x_2, \ldots x_N \mid x_e)$

$$\begin{aligned} -\ln p(x_1, x_2, \ldots x_N \mid x_e) &= -\ln \frac{1}{(2\pi)^{N/2}} + \sum_{n=1}^{N} \sigma_n + \sum_{n=1}^{N} \frac{(x_n - x_e)^2}{2\sigma_n^2}, \\ &= \text{constant} + I(x_e), \\ \text{where } I(x_e) &= \frac{1}{2} \sum_{n=1}^{N} \frac{(x_n - x_e)^2}{\sigma_n^2}. \end{aligned}$$

$I(x_e)$ is sometimes called a cost function. The maximum likelihood estimate of x_e is equivalent to solving the least squares problem above.

Minimizing the cost function

Differentiate $I(x_e)$ with respect to x_e

$$\frac{dI}{dx_e} = \sum_{n=1}^{N} \frac{x_e - x_n}{\sigma_n^2}.$$

Set to zero for the minimum (the function $I(x)$ is concave)

$$\sum_{n=1}^{N} \frac{x_e - x_n}{\sigma_n^2} = 0,$$

$$x_e = \frac{\sum_{n=1}^{N} x_n \sigma_n^{-2}}{\sum_{n=1}^{N} \sigma_n^{-2}}$$

The inverse variances as weights

This problem does allow for the fact that some measurements are more accurate than others (e.g. more accurate instrument).

$$\text{more accurate measurement} \iff \text{larger value of } \sigma_n^{-2}$$

Consider the case for two measurements

$$x_e = \frac{x_1 \sigma_1^{-2} + x_2 \sigma_2^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}.$$

If measurement 1 has much better accuracy than measurement 2, then $\sigma_1^{-2} \gg \sigma_2^{-2}$. Then

$$x_e \approx \frac{x_1\sigma_1^{-2}}{\sigma_1^{-2}} = x_1,$$

and so measurement 2 will not be considered very strongly by the procedure (automatically). If the two measurements have the same accuracy then the maximum likelihood estimate will be an arithmetic mean of the two

$$x_e = \frac{x_1 + x_2}{2}.$$

The variance of the maximum likelihood estimate

Calculating the variance of the maximum likelihood can be done without reverting to doing some difficult moment integrals. The error in the estimate is $x_e - x$

$$x_e - x = \frac{\sum_{n=1}^{N} x_n\sigma_n^{-2}}{\sum_{n=1}^{N} \sigma_n^{-2}} - x = \frac{\sum_{n=1}^{N} (x_n - x)\,\sigma_n^{-2}}{\sum_{n=1}^{N} \sigma_n^{-2}}.$$

The variance of the estimate, σ_e^2, is the mean-square of this error

$$\sigma_e^2 = \left\langle \left(\frac{\sum_{n=1}^{N} (x_n - x)\,\sigma_n^{-2}}{\sum_{n=1}^{N} \sigma_n^{-2}} \right)^2 \right\rangle,$$

$$= \frac{1}{\left(\sum_{n=1}^{N} \sigma_n^{-2}\right)^2} \left\langle \left(\sum_{n=1}^{N} (x_n - x)\,\sigma_n^{-2} \right) \left(\sum_{m=1}^{N} (x_m - x)\,\sigma_m^{-2} \right) \right\rangle,$$

$$= \frac{1}{\left(\sum_{n=1}^{N} \sigma_n^{-2}\right)^2} \sum_{nm} \sigma_n^{-2}\sigma_m^{-2} \left\langle (x_n - x)\,(x_m - x) \right\rangle.$$

The errors in each measurement are assumed to be uncorrelated, so $\left\langle (x_n - x)\,(x_m - x) \right\rangle = \delta_{nm}\sigma_n^2$

$$\sigma_e^2 = \frac{1}{\left(\sum_{n=1}^{N} \sigma_n^{-2}\right)^2} \sum_{n=1}^{N} \sigma_n^{-2} = \frac{1}{\sum_{n=1}^{N} \sigma_n^{-2}}.$$

Note that σ_e^2 has the property that it is smaller than (or equal to if there is just one observation) the variance of any of the individual observations

$$\sigma_e^2 \leqslant \sigma_n^2 \qquad \forall n.$$

Again, consider the case of two measurements

$$\sigma_e^2 = \frac{1}{\sigma_1^{-2} + \sigma_2^{-2}}.$$

If measurement 1 has much better accuracy than measurement 2, then $\sigma_1^{-2} \gg \sigma_2^{-2}$. Then

$$\sigma_e^2 \approx \sigma_1^2,$$

ie the estimate is the same as measurement 1 (result found before) and the variance of the estimate is the same as that of measurement 1. If the two measurements have the same accuracy then the variance of the estimate is halved

$$\sigma_e^2 = \frac{\sigma_1^2}{2}.$$

If all N measurements have the same accuracy then the following classical result is found

$$\sigma_e^2 = \frac{\sigma_1^2}{N}, \qquad \text{ie} \qquad \sigma_e = \frac{\sigma_1}{\sqrt{N}}.$$

Generalizations - introduction to data assimilation

The above example is limited in the following ways.

- One quantity, x, is estimated.
- Many observations are made.
- The observations are direct observations of the unknown quantity.
- The measurement errors are uncorrelated.

The problem can be generalized to deal with many quantities to be estimated, measurements which may observe the quantities indirectly and whose errors may be correlated.

An indirect observation is one that measures some function of the unknown quantities, instead of the quantities themselves. Some example are as follows.

- Measurements of wind speed and direction when the north/south, east/west wind components are required.
- Measurements of temperature and pressure when the potential temperature is required.
- Measurements of the temperature over a large region when the local temperatures are required.
- Measurements from space of the thermal radiation emitted by a column of the atmosphere when the vertical profile of temperature is required.

The following notation is used.

Symbol	Meaning	Reference
$\mathbf{y}$	Vector of p observations	Observation vector
$\mathbf{x}$	Vector of q unknown quantities	State vector
$\mathbf{h}(\mathbf{x})$	Simulated observations according to $\mathbf{x}$	Observation operator
$\mathbf{R}$	Matrix of observation error covariances	Observation error covariance matrix
$\mathbf{x}_\mathrm{b}$	Prior information about $\mathbf{x}$	Background or a-priori
$\mathbf{B}$	Matrix of error covariances of $\mathbf{x}_\mathrm{b}$	Background error covariance matrix

A least squares problem can be constructed along the same lines as the one for the single unknown quantity case

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{x}))^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x})).$$

$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ 1 \times 1 & 1 \times p & p \times p & p \times 1 \end{array}$$

The transpose operator turns the column vector into a row vector and the above evaluates to a scalar quantity. The problem is to minimize $J(\mathbf{x})$ to find $\mathbf{x}_\mathrm{e}$. This can be done only when there is enough information in the observation vector to determine the state vector. A necessary (but not sufficient condition) condition for this is $p \geqslant q$. If $\mathbf{h}(\mathbf{x})$ is a linear function then it may be represented as the $p \times q$ matrix $\mathbf{H}$. Then the cost function becomes

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}).$$

The cost function may be minimized by finding the gradient of J with respect to each element of $\mathbf{x}$. This is represented by the vector $\nabla_\mathbf{x} J$, which is the following q-element vector

$$\nabla_\mathbf{x} J = -\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}).$$

Setting the gradient to zero (to find the $\mathbf{x} = \mathbf{x}_e$ that minimizes J) gives rise to the so-called 'normal equations'

$$\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{x}_e = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y},$$

$$\mathbf{x}_e = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}.$$

$\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is a $q \times q$ matrix. The condition for this solution to exist lies in the properties of $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$. The condition is that $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ must be non-singular (e.g. have no zero eigenvalues).

The error covariance of $\mathbf{x}_e$, denoted $\mathbf{A}$, is found to be the following (not proven here)

$$\mathbf{A} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}.$$

In data assimilation, there are usually very many more unknowns in the state vector than there are observations in the observation vector ($p < q$). In this case, $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is singular and the best fit solution cannot be found. In this case extra information is required, which comes from prior information, $\mathbf{x}_b$. This is called the 'background state' or 'a-priori state' and comes from a numerical forecast of the current state of the atmosphere where this is available. Its error covariance is denoted $\mathbf{B}$. The new cost function fits to the data and to the a-priori simultaneously

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x})).$$

The minimum at $\mathbf{x} = \mathbf{x}_e$ is

$$\mathbf{x}_e = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}_b)),$$

where $\mathbf{H}$ is the linearization (Jacobian) of $\mathbf{h}$. The error covariance of $\mathbf{x}_e$ is

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1},$$

$$= (\mathbf{I} - \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\mathbf{H})\mathbf{B}.$$

References

- Kalnay E., *Atmospheric Modelling, Data Assimilation and Predictability*, Ch. 5.
- Daley R., *Atmospheric Data Analysis*, Ch.13.
- ECMWF, Data assimilation course handouts, `http://www.ecmwf.int/newsevents/training/lecture_notes/LN_DA.html`.