A CRASH COURSE IN CORRELATION

CONTENTS

Uncorrelated Data & Some Basic Definitions 2
Mean, Variance and Expected Value 2
Variances in Averages 2
Variances in Sums 4
Summary 6
Introducing Correlation & Covariance With Continuous Signals 7
Covariance 7
Correlation 8
The Correlation Function 9
Summary 10
Introducing The COvariance Matrix 11
The Covariance Matrix And Chi-Squared Values 12
Estimating the Average of Correlated Data 13
Using CoVariance For Prediction/estimation 14
Summary 15
Stochastic Processes: Bringing It All Together 16
Building the Covariance Matrix in Stochastic Signals 16
Rearranging thE Estiamte/Prediction to be More Efficient 17
Handling Measurement Uncertainty 18

UNCORRELATED DATA & SOME BASIC DEFINITIONS

Before we begin looking at correlation, it's worth re-visiting how uncorrelated data behaves to get a general understanding of how uncertainties interact with one another before complicated stuff.

MEAN, VARIANCE AND EXPECTED VALUE

Even if we aren't looking at normal distributions, it's common to summarise probability distributions in terms of their mean/average/expectation value and their variance:

$$\mu_x = \bar{x} = \langle x \rangle = \int xP(x)\,dx, \qquad \sigma_x^2 = \langle (x-\bar{x})^2 \rangle = Var(x) = \int x^2 P(x)\,dx$$

Strictly speaking, the standard deviation, σ, should only really be used when we know that Gaussian distributions are at play, but we'll use it anyway for simplicity's sake.

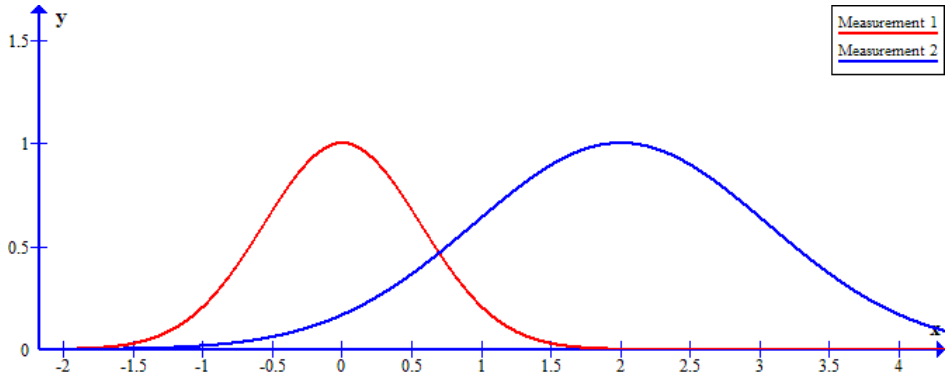
The variance is a measure of how broadly the variable tends to sway about its average. A useful thing to remember about these values is how they scale when their variable is multiplied by some constant:

$$\langle a \cdot x \rangle = a\langle x \rangle, \qquad \sigma_{a\cdot x}^2 = a^2 \cdot \sigma_x^2$$

An important thing to keep in mind is that variance is based on the **square** of differences/deviation. This idea of measuring variation in terms of squares shows up consistently: you can see it in the least squares method of model fitting, the R^2 and χ^2 values in statistics, and the standard deviation of Gaussian distributions in probability.

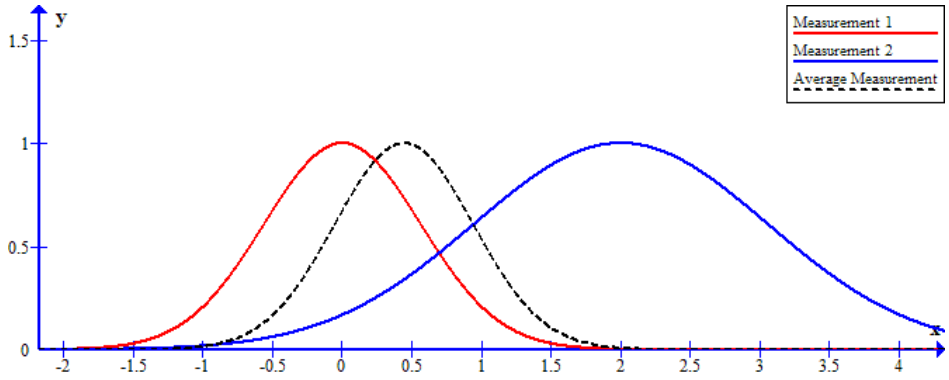
VARIANCES IN AVERAGES

Suppose we have two imprecise measurements, x_1 and x_2, both of the same variable, and we're trying to get a best guess for what that variable is. Obviously, we'd do this by averaging the two measurements somehow, but where exactly do the uncertainties come into play?



Each measurement is effectively a probability distribution for the true value, so it's pretty obvious that we can combine them by simply multiplying:

$$P(\bar{x} = x) \propto P(x_1 = x) \cdot P(x_2 = x)$$



If we go through the actual maths on this, we find that our combined distribution has properties:

$$\bar{x} = \frac{(\sigma_2^2 x_1 + \sigma_1^2 x_2)}{(\sigma_2^2 + \sigma_1^2)}, \qquad \sigma_{\bar{x}} = \sqrt{\frac{\sigma_1^2 \sigma_2^2}{(\sigma_2^2 + \sigma_1^2)}}$$

This looks a little messy, but fortunately it can be re-written in a much neater (and much more general) way:

$$\frac{\bar{x}}{\sigma_{\bar{x}}^2} = \frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}, \qquad \frac{1}{\sigma_{\bar{x}}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Looking at this, we can notice a pattern:

- The average is the inverse-variance weighted sum
- The inverse variance in the average is the sum of inverse variances

We might call the inverse variance, $\frac{1}{\sigma^2}$, something along the lines of "reliability", in which case we could say that we take the "reliability" weighted average, and that the "reliabilities" of the measurements add together.

It turns out that this is actually a general statement for **any** number of measurements:

$$\frac{\bar{x}}{\sigma_{\bar{x}}^2} = \sum_i \frac{x_i}{\sigma_i^2}, \qquad \frac{1}{\sigma_{\bar{x}}^2} = \sum_i \frac{1}{\sigma_i^2}$$

This concept of weighting by the inverse variance shows up even in correlated data, though the specifics become a little more complicated.

VARIANCES IN SUMS

Suppose we've got two variables, 'x' and 'y', with probability distributions 'X' and 'Y'. Now suppose we want to add them together to get 'z', with distribution 'Z':

$$Z = X + Y$$

Intuitively, it makes sense that the average of the sum is the sum of the averages:

$$\bar{z} = \bar{x} + \bar{y}$$

But the variance in z isn't quite so straightforward. First notice that the actual probability distribution of 'Z' looks like:

$$P(Z = z) = \int P(X = x) \cdot P(Y = z - x)\, dx$$

You may notice this as being a **convolution** of the two probability distributions:

$$P(Z = z) = P(X = x) \star P(Y = y)$$

In the specific case of normal distributions, this convolution has a pretty simple solution. First, we remember that convolution in the real domain is multiplication in the fourier domain:

$$F[P(X = x) \star P(Y = y)] = F[P(X = x)] \times F[P(Y = y)]$$

And that the variance of a normal distribution is inversely proportional to that of its fourier transform:

$$\sigma_F \propto \frac{1}{\sigma_x}$$

And use what we know about multiplication of two normal distributions to say that:

$$\frac{1}{\sigma^2_{F[z]}} = \frac{1}{\sigma^2_{F[x]}} + \frac{1}{\sigma^2_{F[y]}}$$

And revert back to the real domain:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

Or, in the simplest terms: **The variance of a sum is the sum of the variances**. More generally:

$$\sigma^2 = \sum_i \sigma_i^2$$

This ends up being the case even when we're not looking at normal distributions: variances stack linearly in sums.

Speaking *even more* generally, we can extend this to include any linear sum. If we have some linear sum:

$$z = \sum a_i x_i$$

Then the variance is found by:

$$Var(z) = \sum_i a_i^2 Var(x_i)$$

Using the inverse-variance weighting, we can actually use this to recover our formula for the variance of an average. Begin with the (normalized) inverse variance weighting:

$$a_i = \frac{\left(\frac{1}{Var(x_i)}\right)}{\sum_j \left(\frac{1}{Var(x_j)}\right)}$$

And using this with our formula for the variance of a linear sum:

$$\begin{aligned} Var(z) &= \sum_i \left(\frac{\left(\frac{1}{Var(x_i)}\right)}{\sum_j \left(\frac{1}{Var(x_j)}\right)} \right)^2 Var(x_i) \\ &= \frac{1}{\left[\sum_j \left(\frac{1}{Var(x_j)}\right)\right]^2} \sum_i \frac{1}{Var(x_i)} \\ &= \frac{1}{\sum_j \left(\frac{1}{Var(x_j)}\right)} \\ \frac{1}{Var(z)} &= \sum_j \left(\frac{1}{Var(x_j)}\right) \end{aligned}$$

Which is exactly what we had before.

SUMMARY

A random variable's probability distribution is typically summarized by its average and is variance:

$$\bar{x} = \langle x \rangle, \qquad Var(x) = \langle (x - \bar{x})^2 \rangle$$

The square root of the variance gives the standard deviation if the variable's distribution is Gaussian:

$$\sigma^2 = Var(x)$$

Averages scale linearly with the variable, while variance scales quadratically:

$$\overline{a \cdot x} = a \cdot \bar{x}, \qquad Var(ax) = a^2 \cdot Var(x)$$

If variables are added together, their variances add:

$$Var\left(\sum a_i x_i\right) = \sum_i a_i^2 Var(x_i)$$

When averaging, we weight by the inverse of variance:

$$\overline{\{x_i\}} = \sum \frac{x_i}{Var(x_i)}, \qquad Var(\bar{x}) = \left[\sum \frac{1}{Var(x_i)}\right]^{-1}$$

INTRODUCING CORRELATION & COVARIANCE WITH CONTINUOUS SIGNALS

Though terms like "correlation" and "covariance" are very general terms, it's easiest to get a handle on what they mean by looking at how they're used in the context of continuous signals. Though the underlying definitions given here are general, we'll spend this section looking at some more specific versions of the concepts.

COVARIANCE

Suppose we have two signals, 'x(t)' and 'y(t)', and we're interested in how much they "line up", i.e. how much the peaks and troughs of one signal line up with the peaks and troughs of the other. To describe this, we introduce the **covariance** between the signals. Strictly speaking, this is defined as:

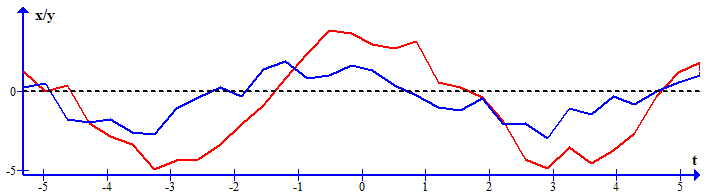
$$Covar(x, y) = \langle (x - \bar{x})(y - \bar{y}) \rangle$$

But, for continuous variables in time, it's a bit easier to think of it as:

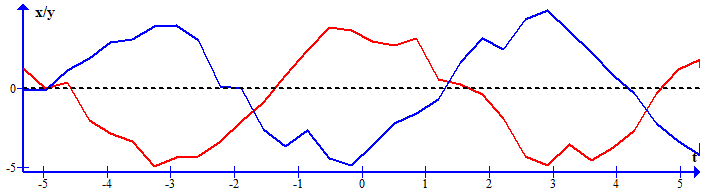
$$Covar(x, y) = \frac{1}{\Delta t} \int (x(t) - \bar{x}) \cdot (y(t) - \bar{y})\, dt$$

It's clear that this measures how much the signals move in-sync with one another:

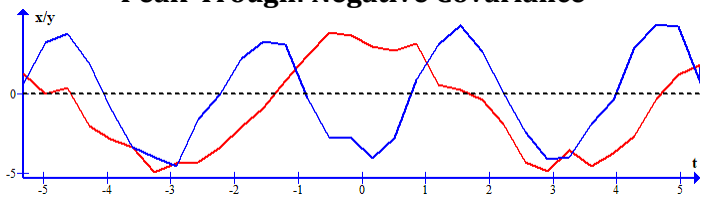
- If the ups line up with ups and the downs with downs, the signals are in-phase and the covariance is positive
- If the ups line up with the downs and vice-versa, the signals are anti-phase and the covariance is negative
- If the ups and downs don't line up in any consistent pattern at all, the covariance is zero



Peak-Peak: Positive Covariance



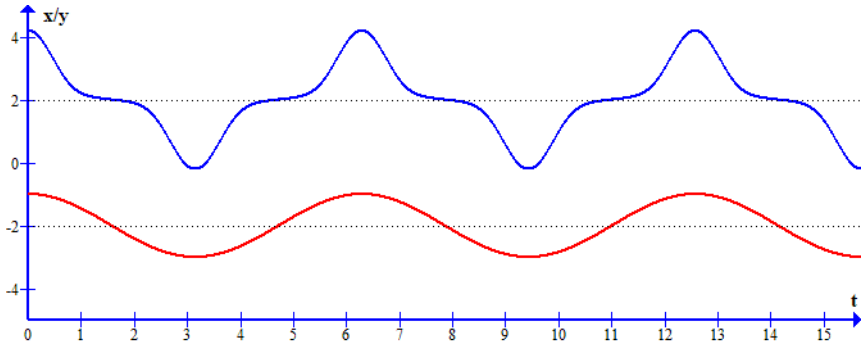
Peak-Trough: Negative Covariance



No Pattern: Zero Covariance

This bares a close resemblance to the "inner product" that shows up in least-squares function fitting, something that is far from a coincidence.

You may also notice that covariance doesn't assume anything about the shape of the signals, for example, the two signals below have a positive covariance, even though they aren't linearly dependant.



It's worth noting that a signal's covariance *with itself* is just its normal variance:

$$Covar(x, x) = Var(x)$$

This idea plays into the concept of a **covariance matrix**, a core feature of correlation in discrete data that discussed further on.

CORRELATION

You might notice that the definitions of covariance mean that the value itself isn't unitless: it scales in proportion to the signal magnitude, i.e.:

$$Covar(a \cdot x, y) = a \cdot Covar(x, y)$$

With this in mind, it might be worth our time to use a signal's variance as a sort of measure of its "width", and use this to normalize the covariance. Doing so gives us the **correlation** between two signals:

$$Corr(x, y) = \phi_{xy} = \frac{Covar(x, y)}{\sqrt{Var(x) \cdot Var(y)}} = \frac{Covar(x, y)}{\sigma_x \sigma_y}$$

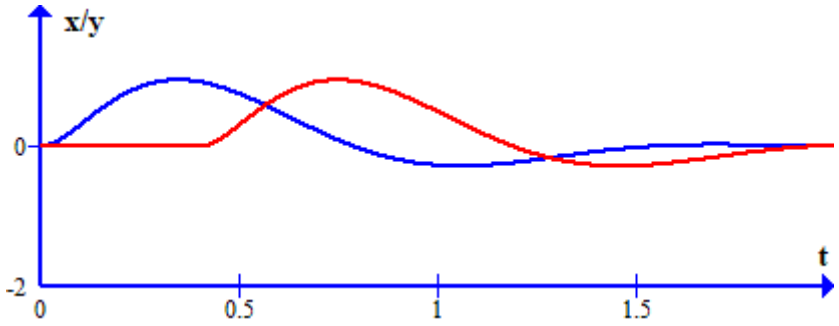
Covariance is useful in that it gives an intuitive description of how the signals relate to eachother. Correlation is a scale from -1 to 1, with -1 being perfect anti-phase correlation, 0 being no correlation at all, and 1 being perfect correlation.

In this way, correlation and covariance have a very simple relationship:

$$Covar(x, y) = \phi_{xy} \sigma_x \sigma_y$$

THE CORRELATION FUNCTION

Consider the following two signals:



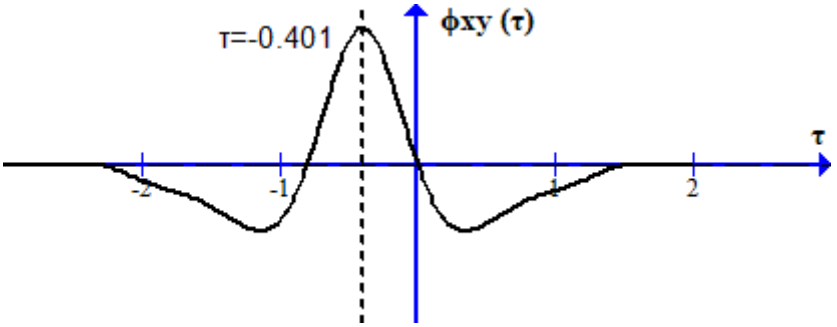
Even at a first glance, it's obvious that there's a relationship between them: one is just a slightly shifted copy of the other. However, if we were to calculate the covariance between them, we'd find that they have *zero correlation*.

This is where the **correlation function** comes in: we need a description of how correlation varies as we shift one of the functions left or right:

$$\begin{aligned}\phi_{xy}(\tau) &= Correl\big(x(t), y(t+\tau)\big) \\ &= \frac{Covar\big(x(t), y(t+\tau)\big)}{\sqrt{Va(x)Var(y)}} \\ &= \frac{1}{\sigma_x \sigma_y}\frac{1}{\Delta t}\int (x(t)-\bar{x})\cdot(y(t)-\bar{y})\,dt\end{aligned}$$

As you can see in the third description, the correlation function is mathematically *very* similar to the convolution of two functions, though the significance of the two ideas is very separate.

Staying with the above example, we find that their correlation function looks like this:



The correlation function has a lot of physical significance: measuring it can help us figure out if a signal is responding to another one, and if so how much of a delay there is between the effect and response. In the above example, the correlation function reaches a peak at $= -0.401$, which tells us that the first signal creates a response in the second after a delay time of 0.401.

When we take a correlation function of a signal *with itself*, we call that its **autocorrelation function**. The autocorrelation function is often used in random signals, where it tells us how closely related nearby points in the signal are.

SUMMARY

The **covariance** between two variables/signals is a measure of how much their variability syncs up:

$$Covar(x,y) = \langle (x-\bar{x})(y-\bar{y}) \rangle \approx \frac{1}{\Delta t}\int (x(t)-\bar{x})(y(t)-\bar{y})\,dt$$

If this is positive, an increase in one tends to line up with an increase in the other. If it is negative, an increase in one usually means a decrease in the other. If its zero, there's no consistent relationship between the two.

A variables covariance with itself is its regular variance:

$$Covar(x,x) = Var(x)$$

Correlation is the covariance normalized to the domain $[-1,1]$, found by dividing the covariance by the standard deviation/square root variances of the signals:

$$Corr(x,y) = \phi_{xy} = \frac{Covar(x,y)}{\sqrt{Var(x)\cdot Var(y)}} = \frac{Covar(x,y)}{\sigma_x \sigma_y}$$

The **correlation function** is a function describing how the correlation varies as we shift the signals left/right:

$$\begin{aligned}\phi_{xy}(\tau) &= Correl\big(x(t), y(t+\tau)\big) \\ &= \frac{Covar\big(x(t), y(t+\tau)\big)}{\sqrt{Va(x)Var(y)}} \\ &= \frac{1}{\sigma_x \sigma_y}\frac{1}{\Delta t}\int (x(t)-\bar{x})\cdot(y(t)-\bar{y})\,dt\end{aligned}$$

The correlation function acts as a measure of how response vs delay between two signals.

INTRODUCING THE COVARIANCE MATRIX

Suppose we have two random variables, 'x' and 'y', that are added together:

$$z = x + y$$

If they're uncorrelated, the variances just add together:

$$Var(x + y) = Var(x) + Var(y)$$

But what if the two variables are (positively) correlated? If 'x' is big, 'y' is likely to also be big, making it easier for 'z' to swing above its mean value. The same thing can be said for 'x' and 'y' being below average, meaning that 'z' will have a higher variance than if 'x' and 'y' were not correlated.

It turns out that the variance of this correlated sum adds **twice** the covariance:

$$Var(x + y) = Var(x) + Var(y) + 2Covar(x, y)$$

Now, notice that this can be written in a slightly different way:

$$\begin{aligned} Var(x + y) = {} & Covar(x, x) + Covar(x, y) \\ & + Covar(y, x) + Covar(y, y) \end{aligned}$$

In this shape, the variance almost looks like a linear system. We might then choose to write this system like:

$$Var(x + y) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} Covar(x, x) & Covar(x, y) \\ Covar(y, x) & Covar(y, y) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Following along with this idea for a general set of variables, $\{x_i\}$, we can define a "covariance matrix", C:

$$C_{ij} = covar(x_i, x_j) = \phi_{ij}\sigma_i\sigma_j$$

In this matrix, we've summarized **everything** about the variance and correlation of the system:

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\phi_{12} & \cdots \\ \sigma_1\sigma_2\phi_{12} & \sigma_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Why is this useful to us? Well, suppose we have some value which is a linear sum of the variables:

$$y = \sum_i a_i x_i$$

Its variance can be found right away using the covariance matrix and a little algebra:

$$Var(y) = \boldsymbol{a}^{\boldsymbol{T}} C \boldsymbol{a} = \vec{\boldsymbol{a}} \cdot C\vec{\boldsymbol{a}}$$

THE COVARIANCE MATRIX AND CHI-SQUARED VALUES

Consider the familiar chi-squared test statistic for uncorrelated data, the classic reliability weighted description of how data varies about the mean:

$$\chi^2 = \sum_i \frac{y_i^2}{\sigma_i^2}$$

If the data is uncorrelated, its covariance matrix (and inverse) is:

$$C = \begin{bmatrix} \sigma_1^2 & 0 & \dots \\ 0 & \sigma_2^2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \qquad C^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots \\ 0 & \frac{1}{\sigma_2^2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Which means we can write the chi-squared expression as:

$$\chi^2 = \vec{y} \cdot C^{-1}\vec{y}$$

As it happens, this is the most general expression for the chi-squared statistic, working even for correlated data.

> *Note: The above assumes $\langle y_i \rangle = 0 \; \forall i$, i.e. that each variable's average has already been subtracted away.*

When we're trying to fit an average or estimate some value, we do so by attempting to minimize this χ^2 value. Equipped with this expression of χ^2 for correlated data, a lot of avenues are opened up to us.

ESTIMATING THE AVERAGE OF CORRELATED DATA

Armed with our χ^2 expression, we can set about using it to estimate values of interest. The first thing we might try our hand at is getting the best guess for the average of a data set, $\bar{y}$.

In this case, the χ^2 value is:

$$\chi^2 = \vec{y} \cdot C^{-1}\vec{y} = \sum_j \sum_i C_{ij}^{-1}(y_i - \bar{y})(y_j - \bar{y})$$

Which we then optimize:

$$\frac{d\chi^2}{d\bar{y}} = 0 = -\sum_j \sum_i C_{ij}^{-1}\left[(y_i - \bar{y}) + (y_j - \bar{y})\right]$$

Rearranging:

$$2\bar{y}\sum_j \sum_i C_{ij}^{-1} = \sum_j \sum_i C_{ij}^{-1} y_i + \sum_j \sum_i C_{ij}^{-1} y_j$$

$$\bar{y} = \frac{\sum_j \sum_i C_{ij}^{-1} y_i}{\sum_j \sum_i C_{ij}^{-1}}$$

Notice that this is a more complicated version of the result for uncorrelated data: we weight each value oif y_i by the inverse of its variance and then normalize.

This is sometimes described by defining a matrix 'L', filled with 1's:

$$L = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$$

In which case we can write the average in the much more concise form:

$$\bar{y} = \frac{\boldsymbol{L^T C^{-1} L}}{\boldsymbol{L^T C^{-1} L}}$$

USING COVARIANCE FOR PREDICTION/ESTIMATION

Suppose we have some dataset, $\{y_i\}$, that we know the covariance matrix for, but we don't have an actual measurement for one element, y_*. Using our understanding of covariance, we can get a best-estimate for y_* by choosing a value that minimizes the χ^2 value.

> *Note: The following assumes $\langle y_i \rangle = 0 \; \forall i$, i.e. that each variable's average has already been subtracted away.*

Firs, remember that the χ^2 value is given by:

$$\chi^2 = \vec{y} \cdot C^{-1}\vec{y} = \sum_j \sum_i C_{ij}^{-1} y_i y_j$$

Now, optimize this value:

$$\frac{\partial \chi^2}{\partial y_i} = 0 = 2\sum_j C_{ij}^{-1} y_j$$

$$\sum_j C_{ij}^{-1} y_j = 0$$

Applying this equation to index $i = *$, and then rearranging:

$$y_* = -C_{**}^{-1} \sum_j C_{i*}^{-1} y_i$$

$$= \frac{\left[\boldsymbol{C_{i*}^{-1}} \cdot \vec{\boldsymbol{y}}\right]}{C_{**}^{-1}}$$

Here, C_{i*}^{-1} is the $*^{th}$ column/row of the C^{-1} matrix, excluding the $*^{th}$ element in that row/column, C_{**}^{-1}.

Notice that this estimate for y_* is a linear sum of remaining values:

$$y_* = \boldsymbol{w^T}\vec{y}$$

Where:

$$w^T = \frac{C_{i*}^{-1}}{C_{**}^{-1}}$$

We have an expression for the variance of a linear sum:

$$\boldsymbol{\Delta}\left(\boldsymbol{w^T}\vec{y}\right)^2 = \boldsymbol{w^T C w}$$

This is the variance in y_* that can be accounted for by the covariance. The actual uncertainty in the value is its initial variance, C_{**}, with this value subtracted away:

$$\Delta y_*^2 = C_{**} - \boldsymbol{w^T C w}$$

When predicting large numbers of variables at once, this formulation can be a little cumberson, as it requires us to invert the matrix 'C' every time we estimate a value. A more efficient rearrangement of this process is shown in the stochastic signals section.

SUMMARY

The **covariance matrix** contains the variances and covariances of a set of data:

$$C_{ij} = covar(x_i, x_j) = \phi_{ij}\sigma_i\sigma_j$$

In this matrix, we've summarized **everything** about the variance and correlation of the system:

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\phi_{12} & \cdots \\ \sigma_1\sigma_2\phi_{12} & \sigma_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

When we take a linear sum of the dataset, its variance can be described by adding together these covariances:

$$y = \sum_i a_i x_i \Rightarrow Var(y) = \sum_{ij} a_i a_j Covar(x_i, x_j)$$

Typically summarized in linear algebra form as:

$$Var(y) = \boldsymbol{a^T C a} = \vec{\boldsymbol{a}} \cdot C\vec{\boldsymbol{a}}$$

The χ^2 statistic of a dataset is found from:

$$\chi^2 = \boldsymbol{y^T C^{-1} y}$$

Defining the matrix '$\boldsymbol{L^T} = [1 \quad 1 \quad \cdots]$', the average of a correlated dataset is estimated with:

$$\bar{y} = \frac{\boldsymbol{L^T C^{-1} L}}{\boldsymbol{L^T C^{-1} L}}$$

And an estimate can be made for a value of an value in a dataset with:

$$y_* = \boldsymbol{w^T}\vec{y}$$

$$\Delta y_*^2 = C_{**} - \boldsymbol{w^T C w}$$

Where:

$$w^T = \frac{C_{i*}^{-1}}{C_{**}^{-1}}$$

STOCHASTIC PROCESSES: BRINGING IT ALL TOGETHER

An area that the correlation concepts discussed in this document become extremely useful is in analysing stochastic processes: signals that are partially random, but still have non-zero correlation between nearby values. This means that, when we have at least one measurement of the signal, we can use prior knowledge about the autocorrelation function of the signal to build up a covariance matrix and predict the signal's behaviour in the places we aren't observing it.

BUILDING THE COVARIANCE MATRIX IN STOCHASTIC SIGNALS

Stochastic signals aren't perfectly determinant :having perfect information about one point in the signal does not guarantee perfect information about one down the track due to the accumulated effect of the random elements. Instead, we say that nearby points are "correlated", i.e. that, because the random elements take time to exert there influence, points in the signal at similar times are likely to be similar to eachother.

This idea manifests in the autocorrelation function, which is often times already found analytically for a process. For example, a damped random walk of characteristic timescale τ_d has an autocorrelation function of:

$$\phi(\tau) = \exp\left(-\frac{|t|}{\tau}\right)$$

Meaning that, if we measure the signal at some time, we can make guesses about the signal at nearby times with exponentially decreasing reliability.

Because we know the correlation, we also know the covariance. If we have some set of data (i.e. measurements of the signal at certain times) we can use this known covariance to build up a covariance matrix between them:

$$C_{ij} = S_{ij} = \phi_{ij}\sigma_i\sigma_j$$

The matrix 'S' is the covariance *due to the stochastic nature of the signa*l. In instances where there is noise as well, we need to account for that as well.

REARRANGING THE ESTIAMTE/PREDICTION TO BE MORE EFFICIENT

In the covariance matrix section, it was said that the best estimate (i.e. mean) for a variable that already know all the covariances for is given by:

$$y_* = \boldsymbol{w^T}\vec{y}$$

With variance

$$\Delta y_*^2 = C_{**} - \boldsymbol{w^T} C \boldsymbol{w}$$

Where:

$$w^T = \frac{C_{i*}^{-1}}{C_{**}^{-1}}$$

This is a perfectly valid way of using covariance to predict a value and to put constraints on its uncertainty, but has some clear practical shortfalls: namely that we have to calculate C^{-1} every time we want to get a new y_*.

Fortunately, it turns out that (though not proven here) the weighting vector can be expressed in a *much* more convenient way:

$$\boldsymbol{w} = \frac{C_{i*}^{-1}}{C_{**}^{-1}} = \boldsymbol{S}_* \, C^{-1}$$

Where $\boldsymbol{S}_*$ is a column vector populated with the covariances of the data point we're predicting with the datapoints we've measured, as arrived at from the stochastic signals autocorrelation function, I,e:

$$[\boldsymbol{S}_*]_i = \langle s_i s_* \rangle$$

With this new description of the weighting vector, our expression becomes:

$$y_* = \boldsymbol{S_*^T} \, C^{-1} y, \qquad \Delta y_*^2 = \langle y_*^2 \rangle - \boldsymbol{S}_* \, C^{-1} \boldsymbol{S_*^T}$$

In this format, we only need to invert the data-data covariance matrix once to get C^{-1}, and we can re-use this for as many particular instances of y_* as we like.

In fact, the linearity of this formulation means that we can replace the estimate-data covariance vector with an estimate-data covariance matrix so as to estimate as many datapoints at once as we like.

> *Note: The above assumes $\langle y_i \rangle = 0 \; \forall i$, i.e. that each variable's average has already been subtracted away.*

HANDLING MEASUREMENT UNCERTAINTY

Measurement uncertainty is surprisingly easy to account for. Suppose your measurements have (1 STD) error $\{E_i\}$, leading to variances of E_i^2. We treat these as being the effect of a (usually uncorrelated) "noise" signal on top of the stochastic process:

$$y = s + n$$

With its own covariance matrix, 'N' For uncorrelated noise errors, this is just a simple diagonal matrix:

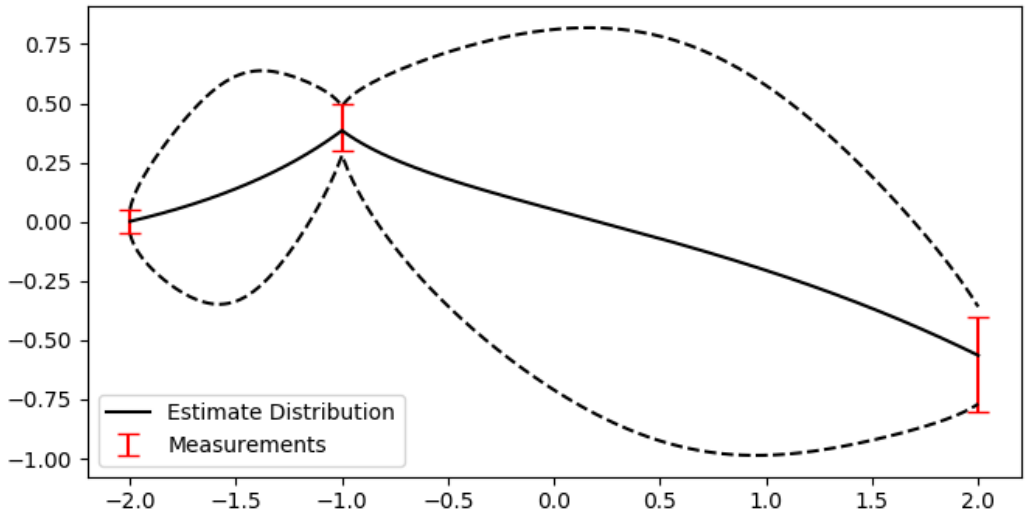
$$N = \begin{bmatrix} E_1^2 & 0 & \cdots \\ 0 & E_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Such that the overall covariance matrix is:

$$C = S + N$$

From here, there's nothing different from the cases explained above.

As an example, here is the described process applied to a damped random walk signal:



.