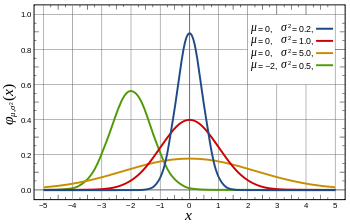
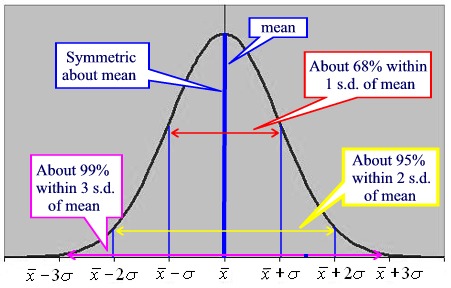
The Covariance Matrix

# Review of some prior knowledge

## Normal distribution

[](https://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg)

## Standard Deviation

[](http://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwjZrYi2ma_LAhWLF5oKHRx8CtwQjRwIBw&url=http://www.regentsprep.org/regents/math/algtrig/ats2/normallesson.htm&bvm=bv.116274245,d.bGQ&psig=AFQjCNFsqNDyK8VV_UXGf5E1zI5BSgOqww&ust=1457461229738514)

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwi0tpeFpa_LAhVFLZoKHfnGA34QjRwIBw&url=https://www.mathsisfun.com/data/standard-deviation-formulas.html&v6u=https://s-v6exp1-ds.metric.gstatic.com/gen_204?ip%3D82.74.10.246%26ts%3D1457378122022710%26auth%3Dpmrkit2qhbt3owzdjsxsairqidylmcoo%26rndm%3D0.7230980506850644&v6s=2&v6t=124256&bvm=bv.116274245,d.bGs&psig=AFQjCNHZqifVA__krxbWSafZFElL0cBbNQ&ust=1457464521971455) (Eqn 1)

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwi5muuTpa_LAhWjNJoKHUuyDroQjRwIBw&url=https://www.mathsisfun.com/data/standard-deviation-formulas.html&bvm=bv.116274245,d.bGs&psig=AFQjCNEHayiaQWkkdcBE81A8UbF6NV5PyQ&ust=1457464671969037) (Eqn 2)

Note : Standard deviation has the same units as the variable itself.

**Question** : Why divide by the factor N-1 rather than N

**{ See 2nd document}**

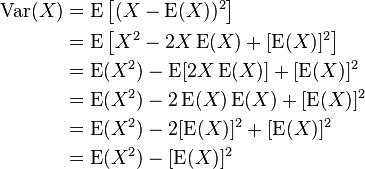
# What’s the Variance then?

***The*** variance is the square of the standard deviation.

Var(X) = ( σ(X))2  (Eqn 3)

This can be ‘translated’ into the

“Expected value of the {difference between X and the mean of X} squared”.

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwigs7qnp6_LAhVDJ5oKHdiLBOgQjRwIBw&url=https://en.wikipedia.org/wiki/Algebraic_formula_for_the_variance&psig=AFQjCNH24p8_-qn3C8WUnCc7RbkRj-4xHQ&ust=1457465075042133) (Eqn 4)

Both Variance and Standard Deviation are measures for the ‘range’ of a variable.

Variance has units ‘the square of the units of the variable’. ***The*** Variance can also be estimated by taking the square of the estimate of the standard deviation (i.e. ‘s’ instead of σ). The difference then is that we use a sample from the whole population, and divide by N-1 rather than N.

*Intermezzo* : more often than not we are talking about a sample, or a finite total population. Then the formulae above are all valid. In the case of an infinite population (so probably continuous) then the variance is given by

\operatorname{Var}(X) =\sigma^2 =\int (x-\mu)^2 \, f(x) \, dx\, =\int x^2 \, f(x) \, dx\, - \mu^2 (Eqn 5)

Where µ is the mean, and f(x) is the probability density function of the distribution. (This can have a number of different forms). We don’t need that here, but is useful to realise.

# Remember some basic rules for Variance:

**Rule 1**: The variance of a sum or difference of two **independent** random variables is the sum of the variances.

Variance of (X±Y) = Var(X) + Var (Y)

Example: X~N(av=5,var=20) Y~N(av=3,var=25) If D=X-Y. Then D~N(av=5-3=2,var = 20+25=45)

**Rule 2**: When a random variable is multiplied by a constant, its variance is multiplied by that constant squared

Variance of cX = c².Var(X)

Example (continued) If P=2X-3Y then P~N(av=2x5 – 3x3=1,var=4.20 + 9.25 = 305)

**Question** : If we apply rule 1 : Var(2X) = Var (X+X) = Var(X) + Var(X) = 2. Var(X) which is not the same as rule 2 (which gives 4.Var(X)) **Why the difference?**

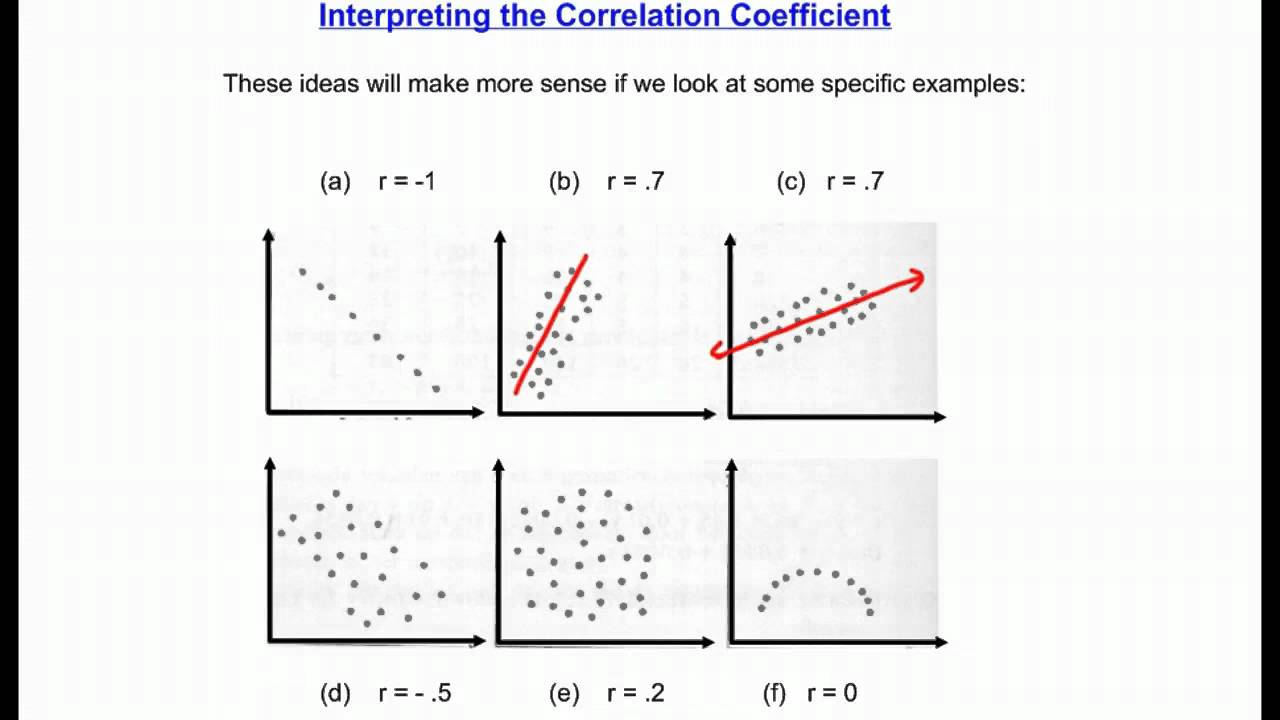
**Rule 3**: (follows from the other two) The variance of a mean of independent random variables, all with the same variance σ2 , is σ2 divided by the number of things being averaged.





# Correlation

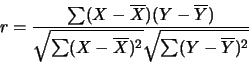
Correlation shows us (or at least, will help us decide) if there is a relation between two variables.

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwi5ke-Hqa_LAhXobZoKHZYnDyUQjRwIBw&url=https://www.youtube.com/watch?v%3Djf-SIOFUuEo&psig=AFQjCNH5ngJWqfwfkjN-HPtPWrQaanWZ9w&ust=1457465689258473)

But be careful about how you interpret this correlation value. The plots below all have a correlation coefficient of 0.



The formula for calculating the Pearson correlation coefficient between variables X and Y is :

[](http://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=0ahUKEwj0ufnOqq_LAhWhCJoKHU-RDzAQjRwIBw&url=http://www.stat.wmich.edu/s216/book/node122.html&bvm=bv.116274245,d.bGs&psig=AFQjCNFHXFNBM-lVV9Tz39NDI9j4jGe3qA&ust=1457466099730274) (Eqn 6)

If we modify this equation :

=

(Eqn 7)

So the correlation coefficient is simply the Covariance between two variables X and Y, scaled by the product of their respective standard deviations, always giving a value between -1 and +1.

Note : if we fill in the formula for correlation for Y=X (i.e. to find the correlation between X and X) we get…… 1.(as we would expect).

So we have defined the Covariance between X and Y as

(Eqn 8)

If we fill in Y=X, we get the formula for the Variance of X.

# The covariance matrix

In a system where there are a number of different variables, X1, X2, X3,….XN , we can define the covariance matrix as the matrix that gives us the covariance between the different variables

(Eqn 9)

Note that along the diagonal, we can replace with Var()

**See the MathCad document for an example where the covariance matrix between two different sensors can be calculated from test measurements.**

A method of calculating the covariance between two different parameters in a system

is to use the following definition of covariance:


\begin{align}
\operatorname{cov}(X,Y)
&= \operatorname{E}\left[\left(X - \operatorname{E}\left[X\right]\right) \left(Y - \operatorname{E}\left[Y\right]\right)\right] \\
&= \operatorname{E}\left[X Y - X \operatorname{E}\left[Y\right] - \operatorname{E}\left[X\right] Y + \operatorname{E}\left[X\right] \operatorname{E}\left[Y\right]\right] \\
&= \operatorname{E}\left[X Y\right] - \operatorname{E}\left[X\right] \operatorname{E}\left[Y\right] - \operatorname{E}\left[X\right] \operatorname{E}\left[Y\right] + \operatorname{E}\left[X\right] \operatorname{E}\left[Y\right] \\
&= \operatorname{E}\left[X Y\right] - \operatorname{E}\left[X\right] \operatorname{E}\left[Y\right].
\end{align}
 (Eqn 10)

For random vectors \mathbf{X} \in \mathbb{R}^mand \mathbf{Y} \in \mathbb{R}^n, the *m×n* covariance matrix is equal to


\begin{align}
    \operatorname{cov}(\mathbf{X},\mathbf{Y}) 
               & = \operatorname{E}
               \left[(\mathbf{X} - \operatorname{E}[\mathbf{X}])
                  (\mathbf{Y} - \operatorname{E}[\mathbf{Y}])^\mathrm{T}\right]\\
               & = \operatorname{E}\left[\mathbf{X} \mathbf{Y}^\mathrm{T}\right] - \operatorname{E}[\mathbf{X}]\operatorname{E}[\mathbf{Y}]^\mathrm{T},
\end{align}
 (Eqn 11)

These formulae may be useful if you know the standard error (= standard deviation) of some of the measurement tools. See the example below.

# Example for the simple KF

Going back to the simple KF document, bottom page 1 – top page 2:

w and z are the different noise parameters – w is the process noise (noise occurring naturally due to instabilities of the system), and z is the measurement noise (noise in the measuring system we are using).

Bottom page 3 – top page 4:

The Kalman Filter satisfies these two conditions., but assumes some conditions about the noise parameters.

(w is the process noise, and z is the measurement noise.)

We must assume that the average value of both w and z are zero. We must also assume that there is no correlation between w and z. I.e. – at any time ‘k’, wk and zk are independent, random variables.

### Measurement Noise

So for the measurement **noise (note** – we are talking about the noise, not the measurement itself) we have only one variable (the gps system), so the covariance matrix is a 1x1 matrix.

The Variance of the GPS noise is simply the square of the Standard Deviation of the GPS noise. We have taken this (see the problem description) to be 10 m. So the Variance is then 102 m2.

An alternative way of calculating the same thing is by using Eqn 4:

We see that Var(X) = E(X2) – (E(X))2

Now, the expected value of GPS noise is zero (see above). So we have Var(GPS noise) = E(GPSnoise2) =10 x 10 ( from the problem description).

### Process Noise

For the process noise, we have two variables in a vector .

If we look at the state equations for our system

The ‘state equation’

xk+1 = Axk + Buk + wk  (1)

And the Output equation

yk = Cxk + zk (2)

In our particular example this became

With T = 0.1 we have

We then see that position is proportional to 0.005 times acceleration, and speed proportional to 0.1 times acceleration 🡪 *so the noise has the same proportionality factors*.

(This comes from the part B.u in equation (1) from the KF document above) :

The covariance matrix we want is

Similar to above, variance is the square of the standard deviation. So for the “position **noise”** we get (0.005)2 x st.dev(accel)2 ; (the factor 0.005 **squared** comes from rule 2 on page 3) and for the “speed noise” we get (0.1)2 x st.dev(accel)2 . These are the two diagonal elements of the matrix.

For the covariance we can use eqn 11, which says that Cov(pN,sN) = E(pN.sNT) – E(pN).E(sN)T . Again, the E(pN) and E(sN) are both zero (**Why was that again….?),** so we get Cov(pN,sN) = E(pN.sNT)

And the expected values of the noise components are known via the state equation:

E(pN.sNT) = E( (0.005 x accelnoise) . (0.1 x accelnoise) )

This gives the covariance matrix as described in the KF1 document.

This is the matrix used in the KF document example.

**Now see the MathCad doc about making a covariance matrix for a sensor, and the 2nd assignment.**