# definitions & Equations

<https://www.youtube.com/watch?v=xBE8qdAAj3w> 34:47

<http://krasserm.github.io/2018/03/19/gaussian-processes/>

<https://www.youtube.com/channel/UCQITXIAgtKlUdfBVhAmTRQA/videos>

P(set of parameters | data) α P(set of parameters) \* prop(data| set of parameters)

P(set of parameters), P(θ) – **prior**: what do we think are sensible parameters, a distribution of likely parameters that we think are sensible. Probably gaussian between some sensible values N(mean, variance (or squared std dev)). The probability we think that the outcome will occur/be true given our prior understanding

P(data| set of parameters or model) - **likelihood**: P(D|M(parameters)) (or P(y|θ)) Model could map surface temperature(measurable) to core temperature (not measurable). Will have uncertainty. Probably a normal distribution. Normal(Model(parameter) – measure, variance). Very dependent on model if that’s wrong you won’t get a good answer

P(set of parameters| data) P(θ|y)– **posterior**: "*When the facts change, I change my mind. What do you do, sir?"* This is the probability given all known knowledge of the system. We do not discard the prior belief, but instead reweight our beliefs to make future guesses less wrong. If we have an infinite number of instances of evidence, the Bayesian function will tend to the frequentist result. For a small N frequentist is very unstable, but this is where Bayesian excels. By having a prior and returning probabilities we keep the uncertainties of our (in this case) small dataset.

**Covariance:** How much we know about y1 given y2

**Marginalisation:** When you have N elements of a big multivariate gaussian. In the most basic case, we use just x and y. But lets say we have a whole array of values N, and we are only interested in a few, we can just integrate out the rest. E.g. (µ - mean, – covariance matrix).

**Conditioning:** Given some value in our multi variate normal what do we expect the next (or some other) value to be?

**Gaussian process:** We generalise the multivariate normal to not just containg many elements, but infinite elements. Gaussian distribution – Y~N(µ,). Guassian process:

Rather than having a mean and a covariance function. We have a mean function and a covariance function. Functions are sort of generalisations of arrays (if we have infinite data points, better to map them to a function than save them all).

*“An infinite collection of random variables, any finite subset of which have a Gaussian distribution”*

Although Gaussian processes are non-parametric (no parameters), they actually have infinite parameters. We are now modelling functions.

**Covariance function:** Functions that generate covariance matrices

**Exponential Quadratic covariance function:** main parameter is the length scale. “how far you have to go between two parameters for them to be more or less independent”, Affects the auto correlation

**Cosine:** Modelling periodic data

**Mean functions:** We generally do not pay too much attention to this. Its just a function that generates mean vectors.. Typically m(x) = 0 or C or Ax+b. The posterior Gaussian process doesn’t involve the mean, which is weird. It is useful as prior info; it can be thought of as a prior guess. When there is no data, a Bayesian model just shrinks to the mean.

**Posterior predictive distribution:** The first ‘function’ in the integral is the likelihood and the second is the posterior. This is how we make predictions with a Bayesian model.