# Gaussian Processes

## Overview

A gaussian process is a sampling from a mvnormal distribution from which the mean and covariance matrix are just a function rather than set values. As an example, if we were sampling a seasonal time series we might make the mean the slope of the time series for each time point (calculated using the mean function). We might then take the covariance to represent seasonal correlations between points to allow the time series to vary in meaningful ways, letting us then predict out this time series into the future.

## Covariance Matrix

A covariance matrix is the thing that shows the relationship between different points we may sample. The way this works is that the covariance is the variance of one element times the variance of the second element times the correlation between the two points. This shows how the overall distribution may vary. In a simple 2D example we can angle the 2D potential mvnormal distribution. This means that we can say how different samples are more likely given what other past data we have.

## Marginal Implementation

If we want to sample just the likelihood of a gaussian process combination then we can add them together and take the marginal likelihood, it adds the noise we need and it is a very simple way to do this.

## Latent Implementation

This is a way where we extract the output of the gaussian process as a prior, here we would want to add some white noise or some jitter to make noise predictions, otherwise the lines would exactly hit every point, rather than fitting to the data in a more general manner.

## Hierarchical Implementations

A basic implementation is to have a global GP that represents the overall trend of the series. We could then created a new GP for each sub element that would vary off this GP. If we wanted we could also make the length scale and smoothing parameters (or other parameters depending on the covariance function) hierarchical in nature. This may cause them to be more intercorrelated than we would like and may perform worse than simply using the global GP on its own and having completely unrelated series that can learn their own things presented separately.

Another structure would be to have a GP that created the time varying mean of all the gaussian processes. This could then be used to be added to a unique mean function for each of the sub GPs that would be the mean function of the global mean plus their respective mean (constant in time) and given a unique scale parameter on the covariance function, however using the same correlation matrix (remember that covariance is variance A \* variance B \* correlation (A,B)). A non differentiable (ie something we are not like directly optimizing) like an AR(1) function of p^|I - j| given time steps I and j. Online there is a good example of what this means and how you build a covariance matrix for an AR process <https://stats.stackexchange.com/questions/295102/how-to-write-variance-covariance-matrix-of-ar1-process-in-r> We can achieve this rather easily by following this formula. This means that we would have a random walk like covariance. This is probably easier if seen in code but yeah. We could use a more complex time series model, but a very simple one is to have a p value that is a normal distribution representing the correlation. Then we would just take p to the power of the number of steps of different we have.

## Joint Hierarchical Gaussian Process

This is a way that we can join the GP so that the effects are modelled together more accurately. To explain this we have the probability of survival and a specific outcome occurring to be the probability of survival given the event time the probability of the outcome occurring given the event times the probability of the event. This means that say a patient gets cancer, the risk of it spreading is some distribution and the survival rate if it spreads is another distribution. This can be sued together to model outcomes of medical patients.

I haven’t worked with survival analysis much. In the future I may come back to this to learn some of it, but for now I will leave it here as it is relatively complicated to make the joint hierarchical process.

<https://arxiv.org/pdf/1408.4660.pdf#:~:text=A%20hierarchical%20Gaussian%20process%20model,continuous%20measurements%20and%20survival%20events>.

## Latent Hierarchical Gaussian Processes

<https://icml.cc/imls/conferences/2007/proceedings/papers/408.pdf>