Smoothing over (a) space.

Matthew Blackwell | Harvard University | blackwel@fas.harvard.edu

The Problem

When we have time series on multiple units, a reasonable prior is that similar units should have similar shocks. The difficulty here is twofold:

1. How to define similarity between two units?
2. How to translate this similarity into a Bayesian prior?

For example, pre-election polls measure state-level support for a candidate, yet most states are not polled on a given day. We typically use priors to smooth across time:

Obama Support Yesterday $\rightarrow$ Obama Support Today

The prior considered here is slightly more subtle, but equally valid:

Trend line in CT $\rightarrow$ Trend line in MA

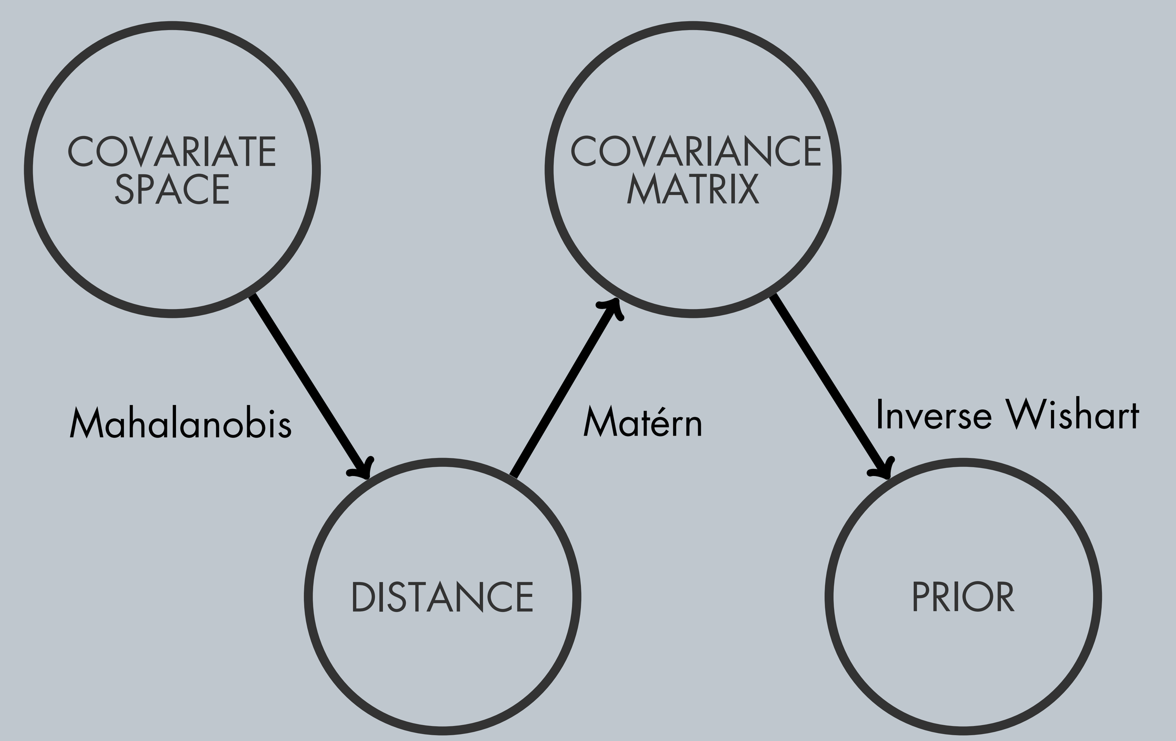
The Solution

kriging (n.) – Optimal spatial interpolation.

Closer points in space $\rightarrow$ Higher covariance.

But, unlike the environmental sciences, geographic distance is rarely the most important spatial consideration. We often care about political distance, economic distance or social distance. Thus, we will use covariates to define a space and use the distances in this space to form a covariance matrix.

1. Define measures of the political space.
2. Calculate distances (e.g. Euclidean, Mahalanobis) from these measures.
3. Feed these distances into a covariance function (e.g., Matérn family).
4. Use this covariance matrix as the scale matrix parameter of an Inverse Wishart prior distribution for a covariance matrix of random walk errors.



The modeler uses external data to set a complex prior. The work comes from choosing the covariates and tuning the covariance functions and prior to induce more or less smoothing.

The Model

polls

$$y_t \sim \mathcal{N}(\alpha_t, \Sigma_y)$$
$$\Sigma_y = \sigma^2_{it} I$$

latent support

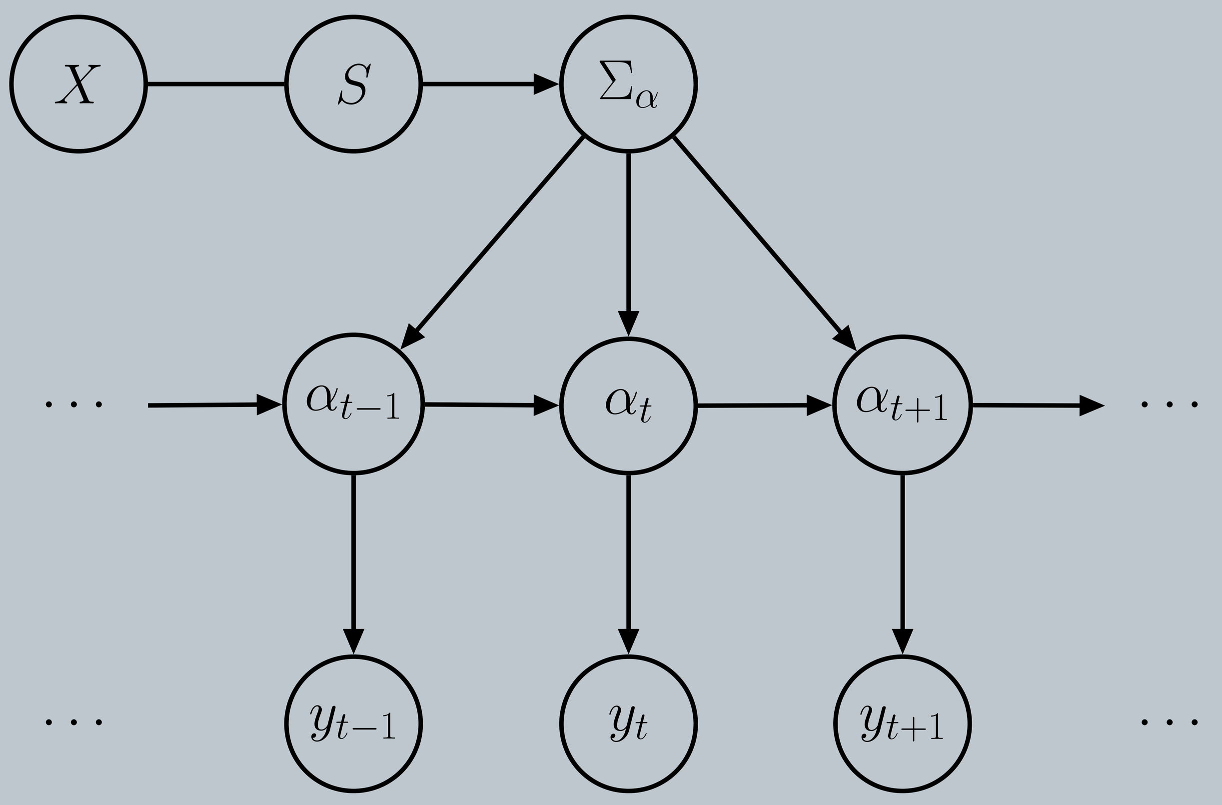
$$\alpha_t \sim \mathcal{N}(\alpha_{t-1}, \Sigma_\alpha)$$
$$\Sigma_\alpha \sim \mathcal{IW}_\nu(S)$$

covariance

$$[S]_{ij} = h\left(1 + \frac{\sqrt{3}d_{ij}}{m}\right)\exp\left(-\frac{\sqrt{3}d_{ij}}{m}\right)$$

distance

$$d_{ij} = \sqrt{(X_i - X_j)'W^{-1}(X_i - X_j)}$$



The Data

Every national and state-level poll conducted in the US Presidential election in 2008 after Hillary Clinton dropped out of the race. For these preliminary results, we construct a political space consisting of just three covariates:

$$X_s = \{\text{\% vote for Bush in 2004, \% black, \% Hispanic}\}$$

We use the Normal approximation to get the variance of the polls:

$$\sigma^2_{it} = \frac{y_{it}(1 - y_{it})}{n_{it}}$$

The Results

In the two plots below, we can see how the prior affects our posterior. These are plots of the α_t for the entire campaign. With our covariate-space smoothing prior, the trend lines of similar states track with each other, even if their means are different. Without these priors, each trend line has a very different path.

Note that geographic distance is not what is driving this as Iowa and Nebraska as the closest two states, but each of their trends track more closely with the other states that are physically farther away.