

# Theory Notes

Dav King

## Overview/Background

Variable selection is important in high-dimensional settings: We often expect that only a small handful of the predictors are actually associated with the outcome, but when we include many in the model, we have issues with computational complexity, sparse solutions, and potentially issues with finding variables significant which are not actually meaningful predictors.

Why is a Bayesian approach potentially better? Bayesian methods allow us to introduce prior information on the betas, which can help us introduce known structure into the variable selection setting and also stabilize inferences in high-dimensional settings (Lu & Lou, 2022).

## Priors

### Spike-and-Slab Priors

The spike-and-slab prior is a two-point mixture on the  $\beta_j$ , which forces some of the  $\beta_j$  to zero and estimates the coefficients of the others. The generic form of the spike-and-slab prior is

$$\beta_j | \gamma_j \sim \gamma_j \phi_1(\beta_j) + (1 - \gamma_j) \phi_0(\beta_j), \quad \gamma_j \sim \pi(\cdot)$$

In this case,  $\phi_1(\beta_j)$  is a diffuse “slab distribution” so that the  $\beta_j$  can reach their true coefficients, and  $\phi_0(\beta_j)$  is a concentrated “spike distribution” pulling effects to 0, and  $\gamma_j$  is a binary latent indicator representing the  $2^p$  possible models (Lu & Lou, 2022).

**Shrinkage Priors**

**Hybrid Priors**

**Other Methods**

**Bayesian Model Averaging**

**Best Subset Selection**

## References

- Lu, Z., & Lou, W. (2022). Bayesian approaches to variable selection: A comparative study from practical perspectives. *The International Journal of Biostatistics*, 18(1), 83–108.  
<https://doi.org/10.1515/ijb-2020-0130>