Bayesian Stability Selection

Mahdi Nouraie
joint work with Prof. Samuel Muller, and Dr. Connor Smith
December 2025

The Instability Issue

$$
\begin{array}{cc}
 & \text{variables} \\
\text{samples} & \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

✗

$$
\begin{array}{cc}
 & \text{variables} \\
\text{samples} & \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}
\end{array}
$$

✓

Stability Selection

- A base selection algorithm

$$\hat{\boldsymbol{\beta}}_{\mathbf{0}}(\lambda), \hat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta}_{\mathbf{0}} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \left(\|\boldsymbol{Y} - \boldsymbol{\beta}_{\mathbf{0}} - X\boldsymbol{\beta}\|_2^2 + \lambda \sum_{k=1}^{p} |\beta_k| \right)$$

- Draw B random sub-samples of observations without replacement
- For each λ, apply Lasso on each sub-sample

$$\hat{S}^{\text{stable}} := \{j \mid \max_{\lambda \in \Lambda}(f_j^{\lambda}) \geq \pi_{\text{thr}}\};\ j = 1, \ldots, p$$

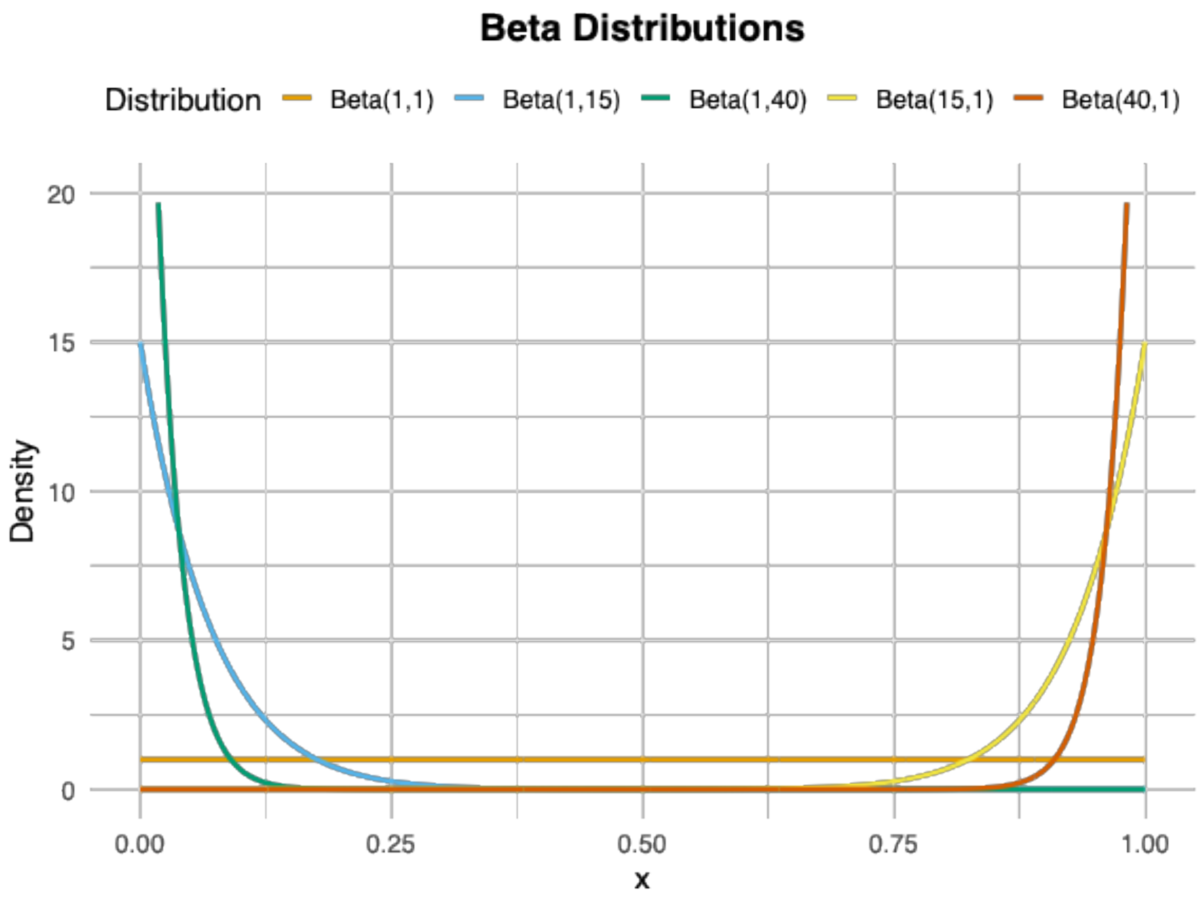
Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *72*(4), 417-473.

What about Inference?

- The authors treat selection frequencies as selection probabilities
- Valid inference requires incorporating prior knowledge
- Beta-Binomial framework

$$n_j^{\lambda} := \sum_{b=1}^{B} M(\lambda)_{bj} \sim \text{Binomial}(B, \Pi_j^{\lambda}); \quad j = 1, \ldots, p$$

Quick Refresher



From Priors to Posteriors

$$\pi(\Pi_j^\lambda | K) = \frac{(\Pi_j^\lambda)^{\alpha_j - 1}(1 - \Pi_j^\lambda)^{\beta_j - 1}}{B(\alpha_j, \beta_j)}; \quad j = 1, \ldots, p \quad \text{and} \quad \alpha_j, \beta_j \geq 1$$

$$\mathcal{L}(n_j^\lambda | \Pi_j^\lambda, K) = \binom{B}{n_j^\lambda} (\Pi_j^\lambda)^{n_j^\lambda} (1 - \Pi_j^\lambda)^{B - n_j^\lambda}; \quad j = 1, \ldots, p$$

$$\pi(\Pi_j^\lambda | n_j^\lambda, K) \propto (\Pi_j^\lambda)^{\alpha_j - 1 + n_j^\lambda} (1 - \Pi_j^\lambda)^{\beta_j - 1 + B - n_j^\lambda}$$

How to Set Priors?

- Existing approaches either fix the mean and variance of the distributions to determine their parameters, or
- Assume large β and small α values to enforce sparsity

Desiderata

- Transparently translating prior knowledge about variable relevance
- Managing the degree of subjectivity in the final results through considering observations vs. pseudo-observations

Prior Elicitation

MACQUARIE University SYDNEY·AUSTRALIA

Question 1 Considering that the final results are a synthesis of both your opinions and data-driven insights, what percentage of the final results for the jth variable would you prefer to be influenced by your prior knowledge, up to a maximum of 50%?

$$\tilde{\zeta}_j := \frac{\gamma_j}{\gamma_j + B}$$

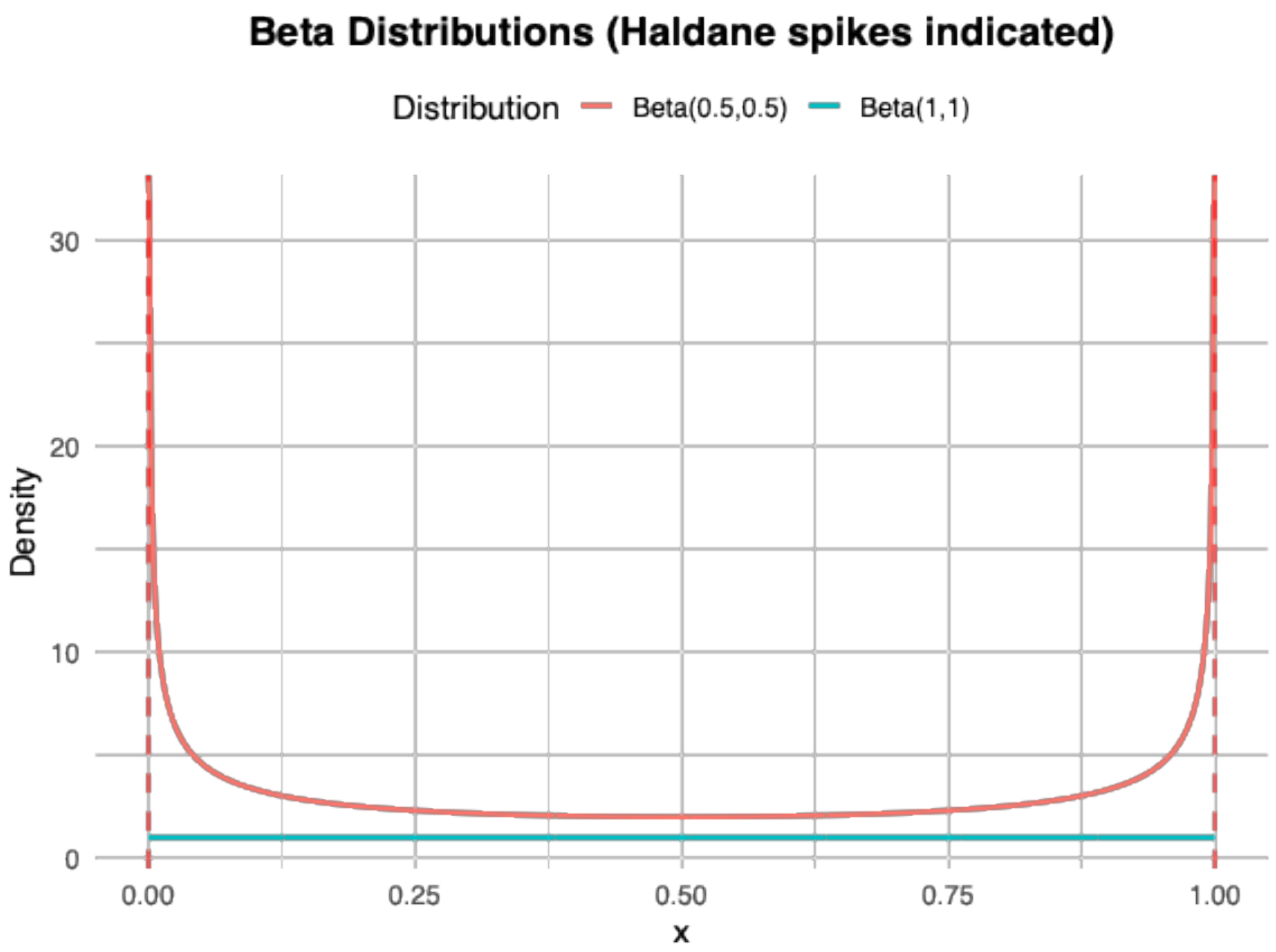
$$\gamma_j := \alpha_j + \beta_j$$

Prior Elicitation

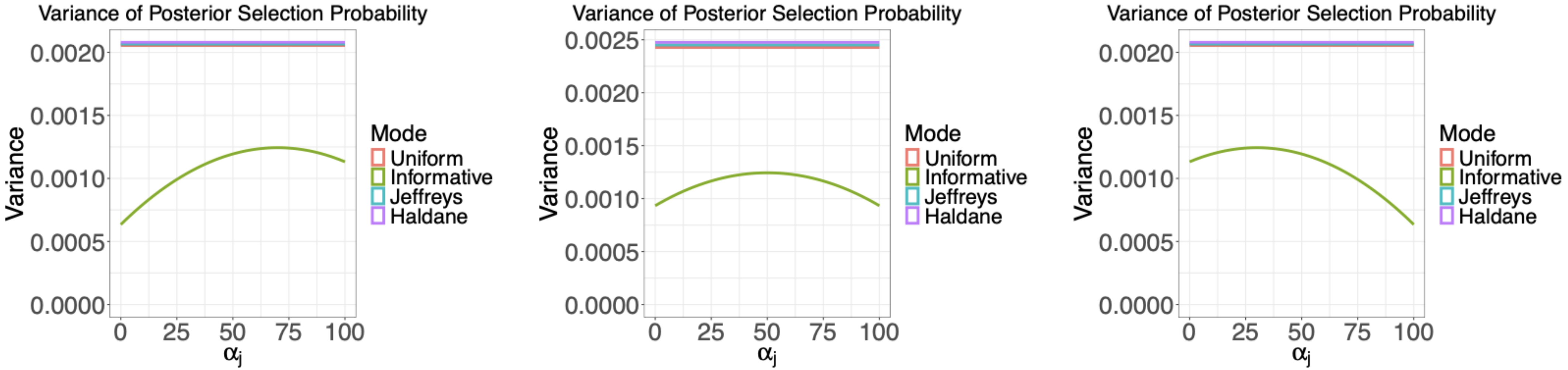
Question 2 Based on your knowledge and expertise, what percentage of the data-driven experiments do you expect to indicate that the jth variable is relevant to the response variable?

$$\alpha_j = \lfloor \tilde{\xi}_j \gamma_j \rfloor \quad \text{and} \quad \beta_j = \gamma_j - \alpha_j$$

Weakly Informative Priors



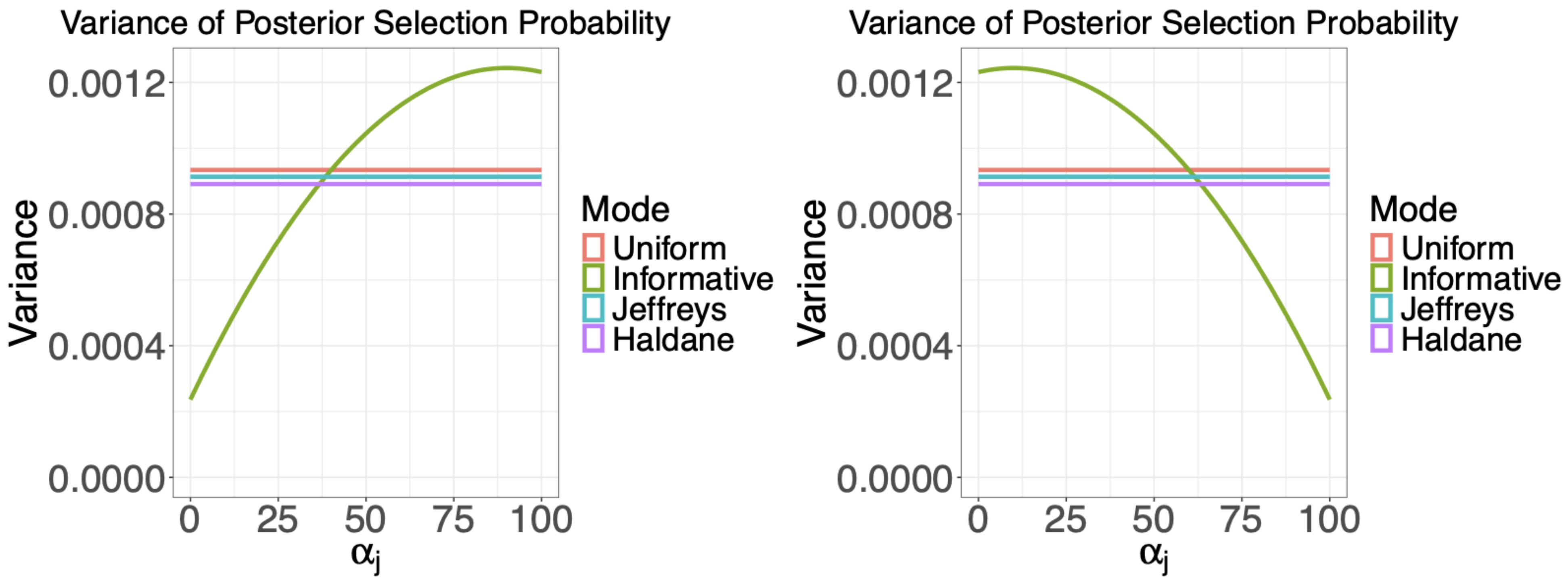
Selection Stability



(a) $n_j^\lambda = 30$ (b) $n_j^\lambda = 50$ (c) $n_j^\lambda = 70$

Variance of the posterior selection probability as a function of α_j when $B = 100$ and $n_j^\lambda \in \{30, 50, 70\}$

Selection Stability

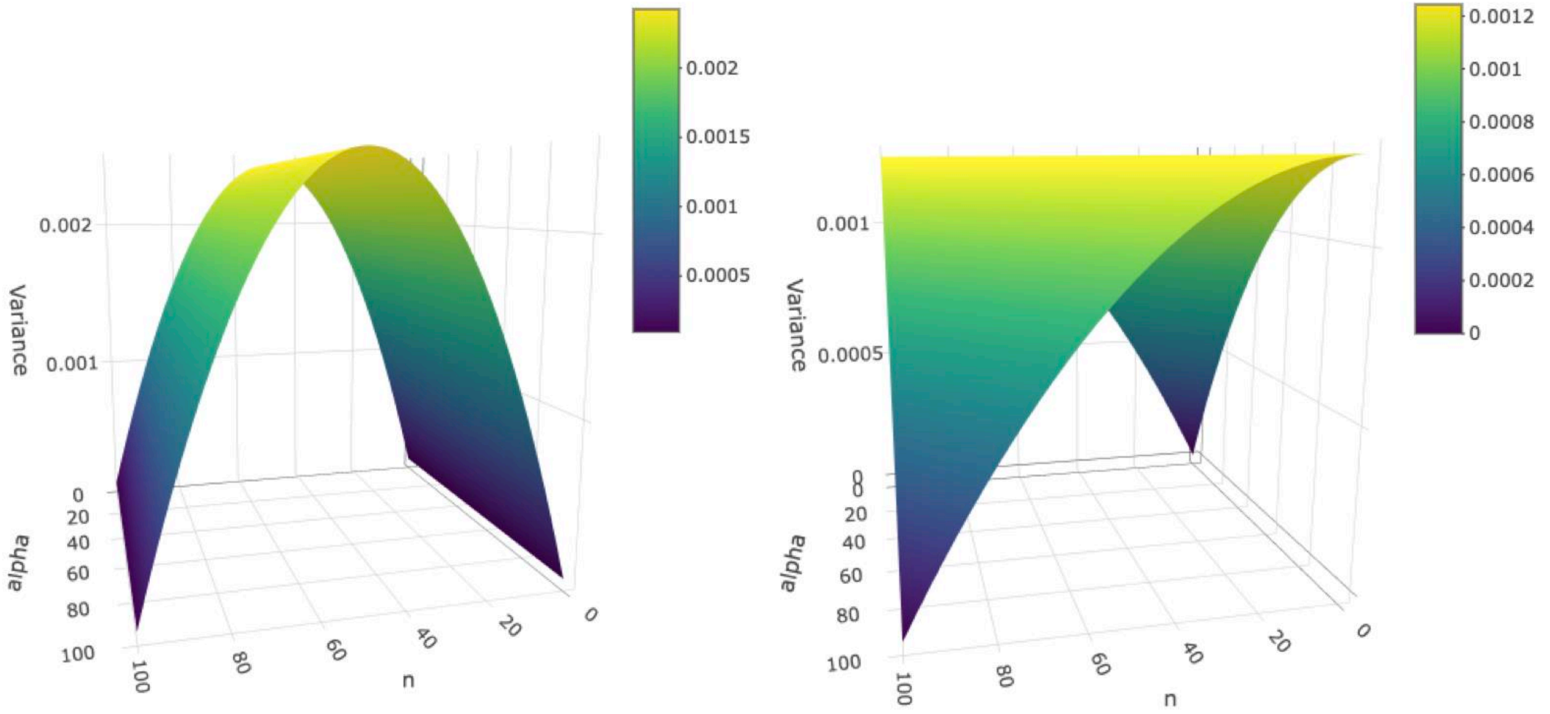


(a) $n_j^\lambda = 10$

(b) $n_j^\lambda = 90$

Variance of the posterior selection probability as a function of α_j when $B = 100$ and $n_j^\lambda \in \{10, 90\}$

Selection Stability



(a) Uniform prior (b) Informative prior

Variance of the posterior selection probability as a function of α_j and n_j^{λ} when $B = 100$

Selection Accuracy

$$\hat{S}^{\text{stableBayes}}(\lambda) := \{j \mid \mathbb{E}(\Pi_j^\lambda | n_j^\lambda, K) \geq \pi_{\text{thr}}\}; \quad j = 1, \ldots, p$$

$$\tilde{\zeta}_j = 50\% \text{ and } \tilde{\xi}_j = 100\%$$

$$f_j^\lambda \geq 2\pi_{\text{thr}} - 1$$

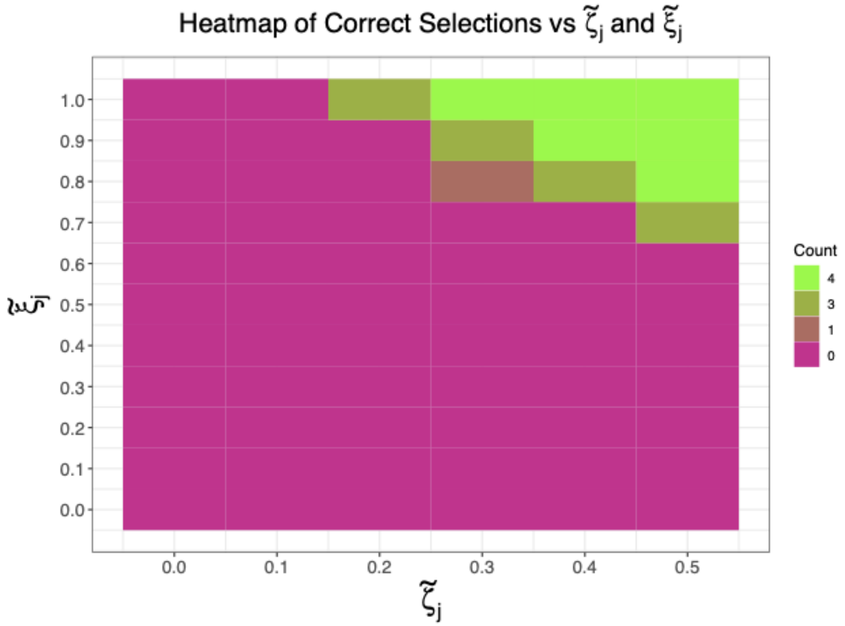
Selection Accuracy

Remark Assume $\tilde{\zeta}_j = 50\%$ and $\tilde{\xi}_j = 100\%$ for all $j \in N$, and let $V \coloneqq |N \cap \hat{S}^{\text{stableBayes}}(\lambda)|$ denote the number of irrelevant variables selected by Bayesian stability selection. Using the indicator function, we have $V = \sum_{j \in N} \mathbf{1}(j \in \hat{S}^{\text{stableBayes}}(\lambda))$, so that $\mathbb{E}(V|\lambda, K) = \sum_{j \in N} \Pr(f_j^{\lambda} \geq 2\pi_{\text{thr}} - 1|\lambda, K)$. If irrelevant variables are exchangeable, this simplifies to $\mathbb{E}(V|\lambda, K) = |N| \Pr(f_j^{\lambda} \geq 2\pi_{\text{thr}} - 1|\lambda, K)$. Suppose that there exists a sequence $\epsilon_{|N|} \to 0$ such that, uniformly over all $j \in N$, $\Pr(f_j^{\lambda} \geq 2\pi_{\text{thr}} - 1|\lambda, K) \leq \epsilon_{|N|}$. Then $\mathbb{E}(V|\lambda, K) \leq |N|\epsilon_{|N|}$. In particular, if $\epsilon_{|N|} = o(1/|N|)$, it follows that $|N|\epsilon_{|N|} \to 0$ as $|N| \to \infty$, implying that $\mathbb{E}(V|\lambda, K) \to 0$ asymptotically.

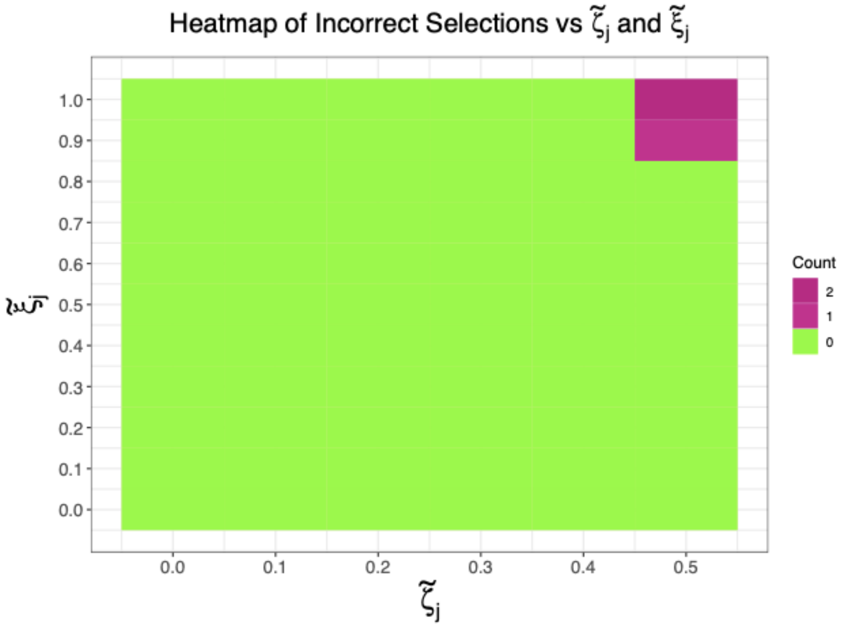
Selection Accuracy

- Frequentist upper-bounds on the number of falsely selected variables remain valid asymptotically under weakly informative priors

Numerical Results



(a) Relevant variables



(b) Irrelevant variables

Bayesian Stability Selection and Inference on Selection Probabilities

Mahdi Nouraie[1], Connor Smith[1], and Samuel Muller*[1,2]

[1]School of Mathematical and Physical Sciences, Macquarie University
[2]School of Mathematics and Statistics, The University of Sydney

Abstract

Stability selection is a versatile framework for structure estimation and variable selection in high-dimensional setting, primarily grounded in frequentist principles. In this paper, we propose an enhanced methodology that integrates Bayesian analysis to refine the inference of selection probabilities within the stability selection framework. Traditional approaches rely on selection frequencies for decision-making, often disregarding domain-specific knowledge. Our methodology uses prior information to derive posterior distributions of selection probabilities, thereby improving both inference and decision-making. We present a two-step process for engaging with domain experts, enabling statisticians to construct prior distributions informed by expert knowledge while allowing experts to control the weight of their input on the final results. Using posterior distributions, we offer Bayesian credible intervals to quantify uncertainty in the variable selection process. Furthermore, we demonstrate how the integration of prior knowledge reduces the variance of selection probabilities, thereby improving the stability of decision-making. Our approach preserves the versatility of stability selection and is suitable for a broad range of structure estimation challenges.

Keywords: Bioinformatics, Bayesian Inference, Feature Selection, Prior Elicitation, Structure Estimation, Variable Selection

Contact Information

mahdi.nouraie@hdr.mq.edu.au

https://mahdinouraie.quarto.pub