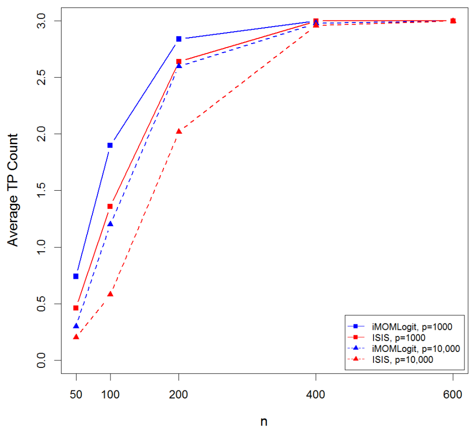
# Bayesian Variable Selection for Binary Outcomes in High Dimensional Genomic Studies Using Non-Local Priors

## Supplementary Data

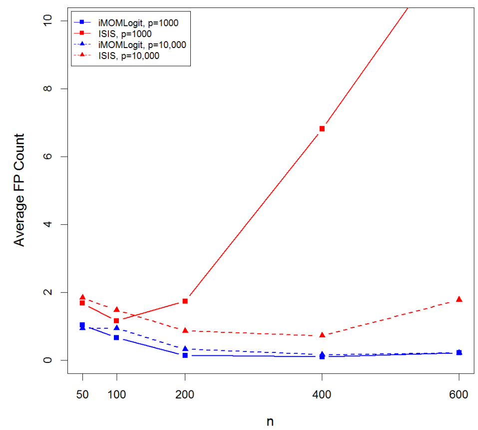
Amir Nikooienejad, Wenyi Wang and Valen E. Johnson

## 1 Simulation Studies

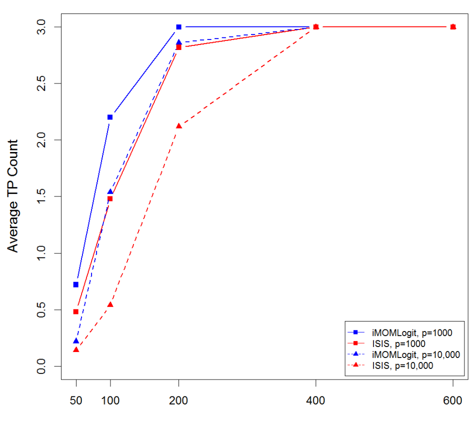
Average TP and FP for $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$ explained in simulation studies in the paper.



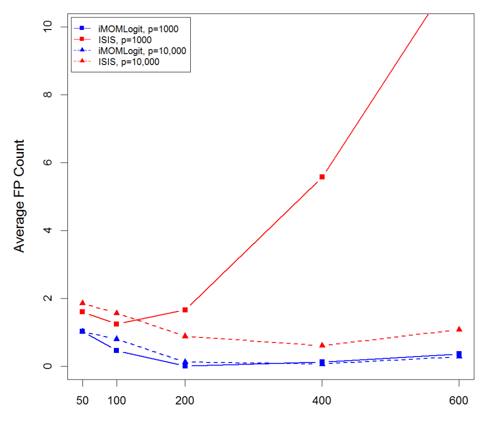

Average true positive count for $\boldsymbol{\beta_2}$




Average false positive count for $\boldsymbol{\beta_2}$



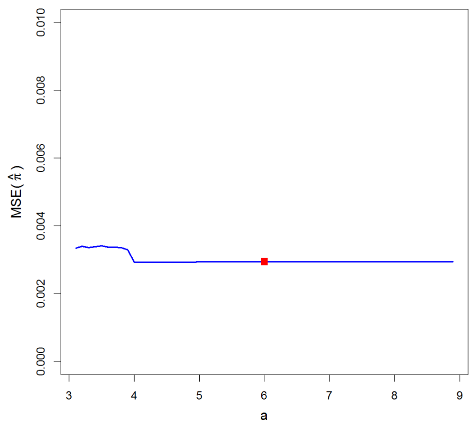

Average true positive count for $\boldsymbol{\beta_3}$




Average false positive count for $\boldsymbol{\beta_3}$

## 2 Sensitivity Analysis for Prior Parameters on Model Space

The figure below depicts MSE($\hat{\boldsymbol{\pi}}$) for different values of $a$ for $n = 200$, $p = 1000$ and $\boldsymbol{\beta} = [4, 5, 6]^T$. The output at nominal value of $a$ is indicated by a red square. As shown in this figure, the value of output does not change dramatically with changes in $a$.




Sensitivity analysis for parameters of prior on model space

## 3 Discussion on $1/\sqrt{p}$ Overlap

Our rationale for setting the overlap between the sampling distribution of the MLE and the prior density to be $p^{-1/2}$ can be explained as follows. For simplicity, we motivate our criterion in the context of a scalar-valued parameter $\theta$. Let $p(\theta)$ denote the prior density for $\theta$ under a nonlocal prior defining the alternative hypothesis, $H_1$, and let $f(\theta) = \prod_{i=1}^{n} f_i(x_i|\theta)$ denote the likelihood function, and let $i(\hat{\theta})$ denote the observed information evaluated at the MLE $\hat{\theta}$, i.e.,

$$i(\hat{\theta}) = -\left.\frac{\partial^2 \log f(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}}.$$

Under the null hypothesis, $\theta = 0$. The marginal likelihood function under the alternative hypothesis can be approximated using Laplace's method as

$$m_1(\hat{\theta}) \approx \sqrt{\frac{2\pi}{i(\hat{\theta})}} f(\hat{\theta}) p(\hat{\theta}),$$

while under the null model the marginal density of the data is simply

$$m_0 = f(0).$$

In large samples when the null hypothesis is true,

$$f(\hat{\theta}) \approx f(0) e^{\eta(\hat{\theta})/2},$$

where $\eta$ is a chi-squared random variable, which is bounded in probability. Also, for large $n$, the observed information $i(\hat{\theta})$ converges to Fisher's information, $I(0)$. Define $w$ to be

$$w = \sqrt{\frac{2\pi}{I(0)}}.$$

Now let $g(\hat{\theta})$ denote the sampling distribution of the maximum likelihood estimate under the null hypothesis. We assume that this sampling density is approximately normally distributed around 0 and let $\pm x$ denote the point at which the sampling density of the MLE and the non-local prior densities overlap. Under our constraint on the overlap between densities, the expected value of $m_1$ satisfies

$$\begin{aligned}
E_0[m_1(\hat{\theta})]/w &\approx \int_{|\hat{\theta}|<x} f(0)e^{\eta(\hat{\theta})/2} p(\hat{\theta}) g(\hat{\theta}) d\hat{\theta} + \int_{|\hat{\theta}|>x} f(0)e^{\eta(\hat{\theta})/2} p(\hat{\theta}) g(\hat{\theta}) d\hat{\theta} \\
&\leq \max[g(\hat{\theta})] \int_{|\hat{\theta}|<x} f(0)e^{\eta(\hat{\theta})/2} p(\hat{\theta}) + \max[p(\hat{\theta})] \int_{|\hat{\theta}|>x} f(0)e^{\eta(\hat{\theta})/2} g(\hat{\theta}) d\hat{\theta} \\
&\leq \max[g(\hat{\theta}), p(\hat{\theta})] \left[ \int_{|\hat{\theta}|<x} f(0)e^{\eta(\hat{\theta})/2} p(\hat{\theta}) + \int_{|\hat{\theta}|>x} f(0)e^{\eta(\hat{\theta})/2} g(\hat{\theta}) d\hat{\theta} \right] \\
&\approx \max[g(\hat{\theta}), p(\hat{\theta})] f(0) e^{\eta'/2} \frac{1}{\sqrt{p}}
\end{aligned}$$

for some random variable $\eta'$ that is bounded in probability. The Bayes factor in favor of the larger model is thus

$$BF_{10} < w \max[g(\hat{\theta}), p(\hat{\theta})] \exp(\eta'/2) \frac{1}{\sqrt{p}}.$$

For large $n$, the second term on the right hand side of the inequality is determined by the sampling distribution of the MLE and is $O_p(n^{1/2})$, while $w$ is $O(n^{-1/2})$. Thus, the average Bayes factor is $O_p(p^{-1/2})$, and combined with the beta-binomial prior on the model space (which imposes a penalty that is $O(1/p)$ on new variables), this suggests that the number of false positives under the null model of no effects will decrease to 0 as $p$ increases.