Illustrative simulations

Hanyu Song

February 22, 2017

1 Illustrative simulations

In this section, we present the empirical performance of **BayesDECO** via extensive simulations. In particular, we adopt the model structure used in Yang et al.'s (2015) work as the conventional Bayesian approach, and apply the same model to each subset of the data after partitioning as the BayesDECO method. We compare the performance of the proposed BayesDECO with conventional Bayesian approach and naive Bayesian with partitioning but NO decorrelation (**Bayesian partition with no DECO**). The synthetic datasets are from the linear regression model

$$Y = X\beta + \epsilon, \tag{1}$$

with $X \sim N(0, \Sigma)$ and $\epsilon \sim N(0, \sigma^2)$. The variance σ^2 is chosen so that $R^2 = \frac{var(\mathbf{X}\beta)}{var(Y)} \in \{0.5, 0.9\}$. For evaluation purposes, we consider five different structures of Σ, proposed by Wang et al. (2016), as follows:

1. Model (i) Independent predictors
   The support of β is $S = \{1, 2, 3, 4, 5\}$. We generate X_i from a standard multivariate normal distribution with independent components. The coefficients are specified as

$$t\beta_i = \begin{cases} (-1)^{Ber(0.5)} \left(|N(0,1)| + 5\sqrt{\frac{\log p}{n}} \right) & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}$$

2. Model (ii) Compound symmetry
   All predictors are equally correlated with correlation $\rho = 0.6$. The coefficients are the same as those in Model (i).

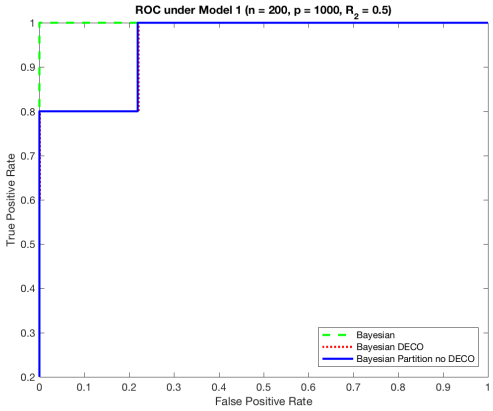
3. Model (iii) Group structure
   This example is Example 4 in Zou and Hastie (2005), for which we allocate the 15 true variables into three groups. Specifically, the predictors are generated as $x_{1+3m} = z_1 + N(0, 0.01), x_{2+3m} = z_2 + N(0, 0.01)$ and $x_{3+3m} = z_3 + N(0, 0.01)$, where $m = 0, 1, 2, 3, 4$ and $z_i \sim N(0, 1)$ are independent. The coefficients are set as $\beta_i = 3, i = 1, 2, \ldots, 15$; $\beta_i = 0, i = 16, \ldots, p$.

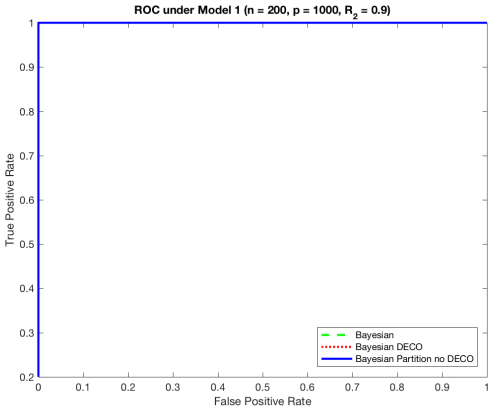
4. Model (iv) Factor models
   This model is considered in Meinshausen and Bhlmann (2010). Let $\phi_j, j = 1, 2, \ldots, k$ be independent standard normal variables. We set predictors as $x_i = \sum_{i=1}^{k} \phi_j f_{ij} + \eta_i$, where f_{ij} and η_i are independent standard normal random variables. The number of factors is chosen as $k = 5$ in the simulation while the coefficients are specified the same as in Model (i).

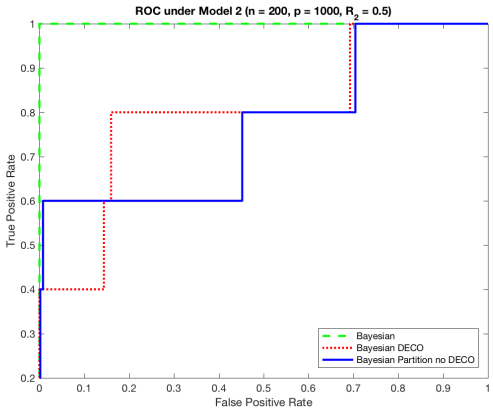
5. Model (v) l_1-ball
   This model takes the same correlation structure as Model (ii), with the coefficients drawn from Dirichlet distribution $\beta \sim Dir\,(1/p, 1/p, \ldots, 1/p) \times 10$. This model is to test the performance under a weakly sparse assumption on β, since β is non-sparse satisfying $||\beta||_1 = 10$.

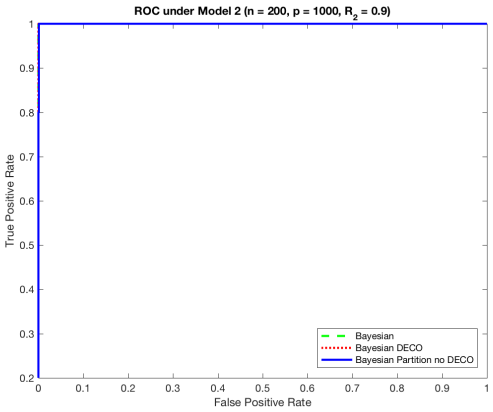
Throughout this section, the performance is evaluated in terms of two metrics: computational time (in seconds) and ROC curves of variable selection based on posterior marginal inclusion probability. For each method, the reported computational time is a trimmed average of four values across ten replications, excluding the two largest and the two smallest values. All the algorithms are coded and implemented in *Matlab* on computers with Intel Xeon Processor E5-2640 v2. For any embarrasingly parallel algorithm we report the preprocessing time plus the longest runtime of a single machine as its runtime.

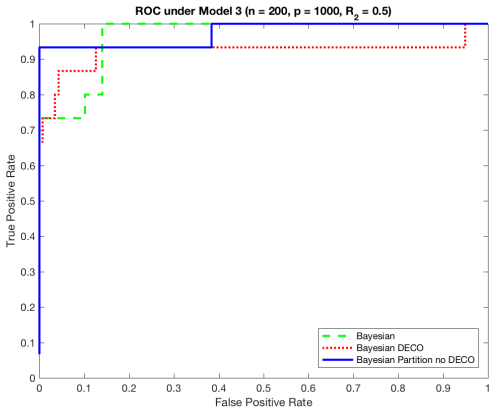
1.1 ROC curves

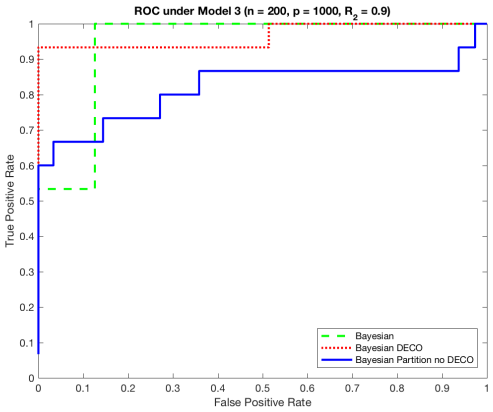


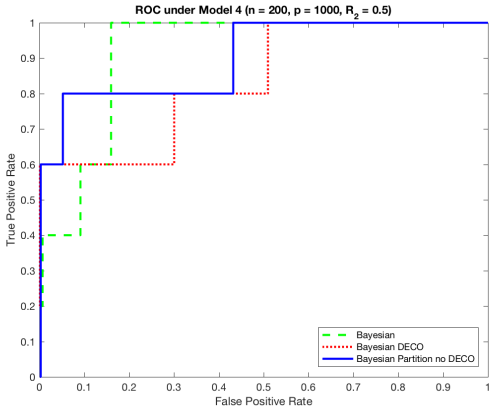


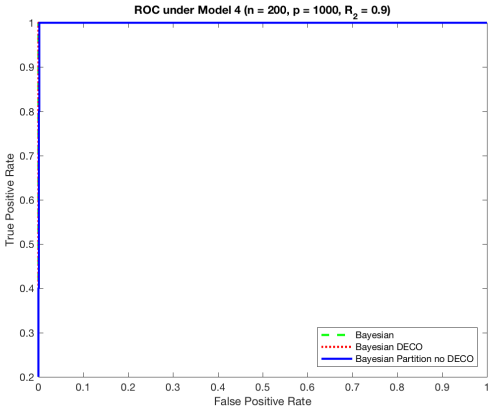


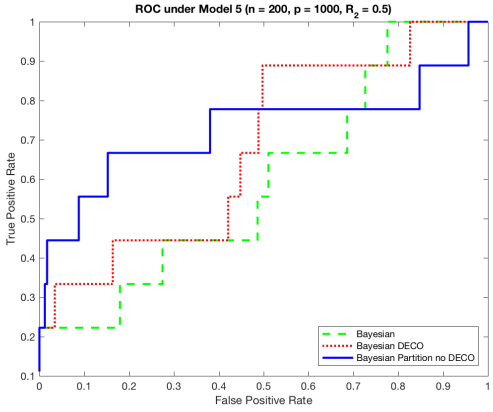


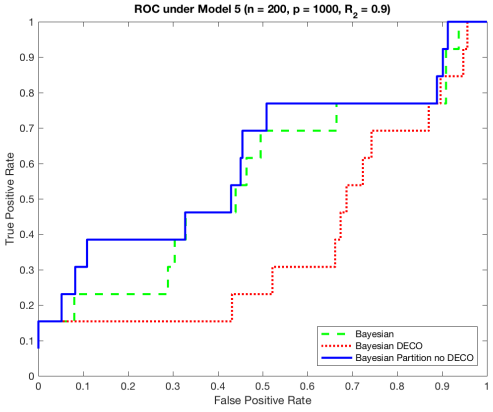


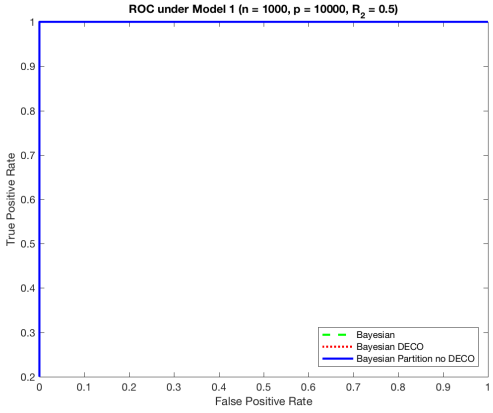


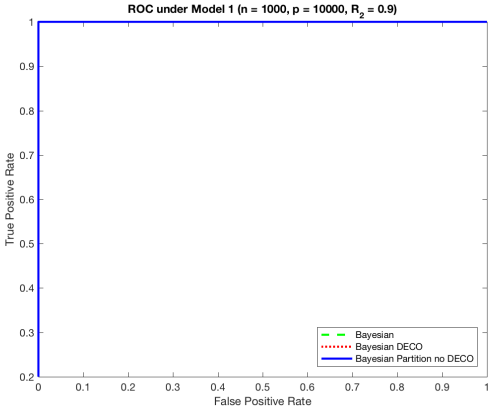

ROC under Model 4 (n = 200, p = 1000, R2 = 0.5)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

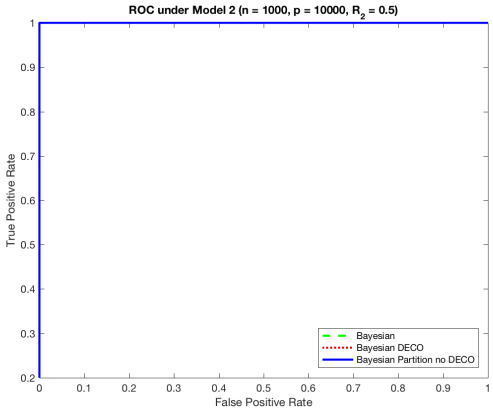

ROC under Model 4 (n = 200, p = 1000, R2 = 0.9)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

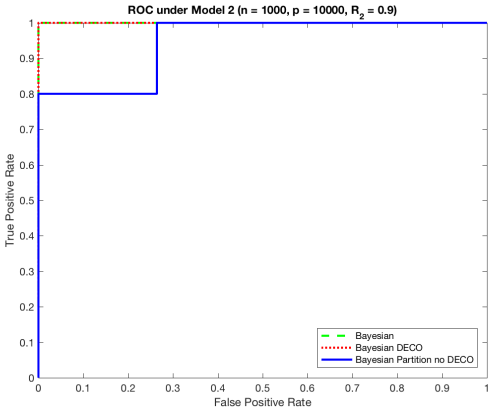

ROC under Model 5 (n = 200, p = 1000, R2 = 0.5)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

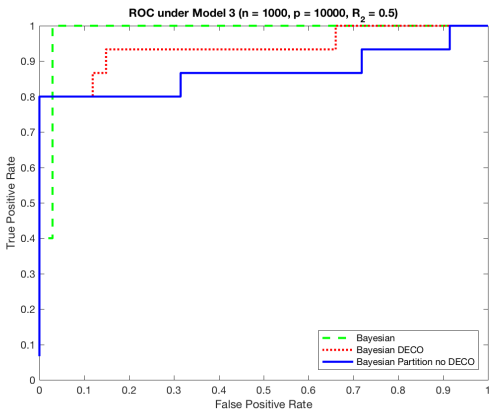

ROC under Model 5 (n = 200, p = 1000, R2 = 0.9)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

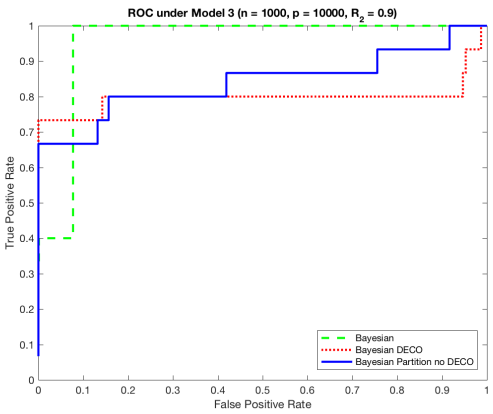

ROC under Model 1 (n = 1000, p = 10000, R2 = 0.5)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

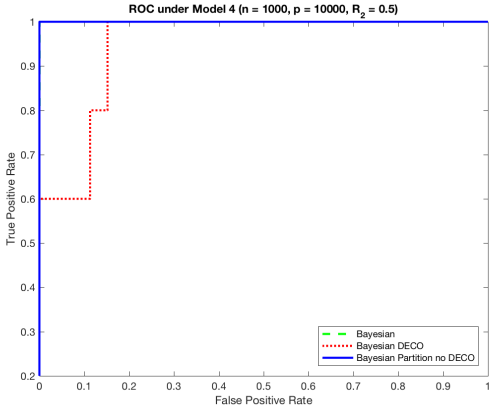

ROC under Model 1 (n = 1000, p = 10000, R2 = 0.9)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

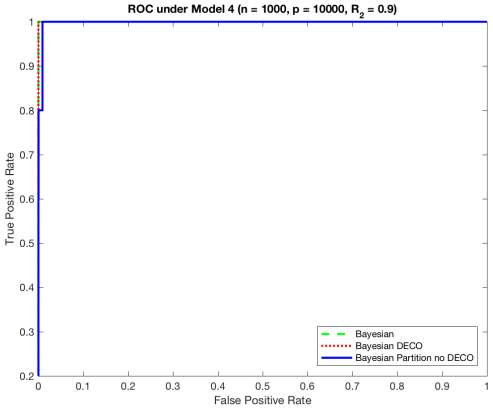

ROC under Model 2 (n = 1000, p = 10000, R2 = 0.5)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

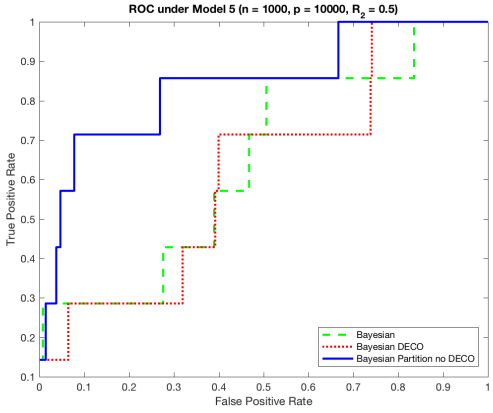

ROC under Model 2 (n = 1000, p = 10000, R2 = 0.9)
True Positive Rate
False Positive Rate
Bayesian
Bayesian DECO
Bayesian Partition no DECO

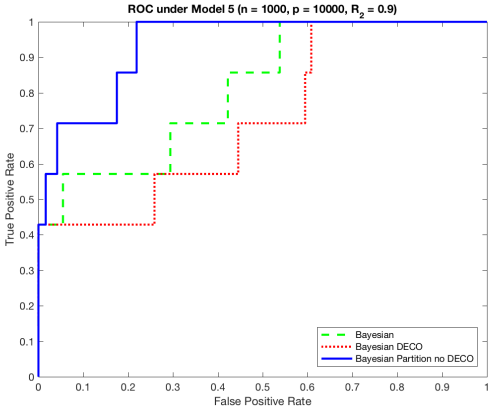












1.2 Runtime

Table 1: Runtime(s) for five models with $(n, p, R^2) = (200, 1000, 0.5)$

Model	Conventional Bayesian variable selection	Bayesian DECO	Bayesian partitioning no DECO
1	8.35	4.46	4.48
2	7.63	4.80	5.33
3	6.32	5.19	5.20
4	8.59	4.43	6.69
5	4.89	3.72	6.04

about 15 seconds to initialise Parallel Pool

Table 2: Runtime(s) for five models with $(n, p, R^2) = (1000, 10000, 0.5)$

Model	Conventional Bayesian variable selection	Bayesian DECO	Bayesian partitioning no DECO
1	223.65	71.92	79.20
2	227.89	78.12	187.38(*)
3	176.16	84.14	81.86
4	225.39	71.90	143.92
5	168.43	60.53	151.01

2 Appendix: Derivation of posterior inclusion probability

$$\text{Linear model} : Y = X_\gamma \beta_\gamma + w, w \sim N(0, \phi^{-1} I_n) \tag{2}$$

$$\text{Precision prior} : \pi(\phi) \propto 1/\phi \tag{3}$$

$$\text{Regression prior} : \beta_\gamma | \gamma \sim N(0, g\phi^{-1} \left(\mathbf{X}_\gamma^T \mathbf{X}_\gamma\right)^{-1}) \tag{4}$$

$$\text{Sparsity prior} : \pi(\gamma) \propto \left(\frac{1}{p}\right)^{\kappa|\gamma|} \mathbb{I}[|\gamma| \leq s_0], \tag{5}$$

where $g = n$, $s_0 = 100$, and $\kappa = 1$.

Under this setting, we can obtain a closed-form expression for the marginal likelihood of the indicator vector γ by integrating out β_γ and ϕ:

$$\mathcal{L}(Y|\gamma) := \pi(Y|\gamma) \tag{6}$$

$$= \frac{\Gamma(n/2)(1+g)^{n/2}}{\pi^{n/2}||Y||_2^n} \frac{(1+g)^{-|\gamma|/2}}{(1+g(1-R_\gamma^2))^{n/2}}, \tag{7}$$

where R_γ^2 is the coefficient of determination for the model $\mathbb{M}_\gamma$

$$R_\gamma^2 = \frac{Y^T \phi_\gamma Y}{||Y||_2^2}$$

with $\phi_\gamma = X_\gamma (X_\gamma^T X_\gamma)^{-1} X_\gamma^T$.
Let $\mathcal{M} := \{\gamma : |\gamma| \leq s_0\}$ denote the entire model space (which is a subset of the p- dimensional hypercube $\{0,1\}^p$). Then by Bayes' theorem the posterior probability of γ is given by

$$\pi_n(\gamma|Y) = C \cdot \frac{1}{p^{\kappa|\gamma|}} \cdot \frac{(1+g)^{-|\gamma|/2}}{(1+g(1-R_\gamma^2))^{n/2}} \mathbb{1}[\gamma \in \mathcal{M}], \tag{8}$$

where C is a normalization constant.

$$p(\gamma_j = 1|\gamma_{(-j)}, y) = \frac{p(\gamma_j = 1, \gamma_{(-j)}|y)}{p(\gamma_{(-j)}|y)} \tag{9}$$

$$= \frac{p(\gamma_j = 1, \gamma_{(-j)}|y)}{p(\gamma_j = 1, \gamma_{(-j)}|y) + p(\gamma_j = 0, \gamma_{(-j)}|y)} \tag{10}$$

$$= \frac{1}{1 + \frac{p(\gamma_j=0, \gamma_{(-j)}|y)}{p(\gamma_j=1, \gamma_{(-j)}|y)}} \tag{11}$$

$$= \frac{1}{1 + \frac{\pi(\gamma_j=0, \gamma_{(-j)})L(y;\gamma_j=0, \gamma_{(-j)})}{\pi(\gamma_j=1, \gamma_{(-j)}L(y;\gamma_j=1, \gamma_{(-j)})}} \tag{12}$$

$$= \left[1 + \left(\frac{(\frac{1}{p})^{\kappa|\gamma_{(-j)}|}\mathbb{1}[|\gamma_{(-j)}| \leq s_0](1+g)^{-|\gamma_{(-j)}|/2}(1 + g(1 - R^2_{\gamma_j=1,\gamma_{(-j)}}))^{n/2}}{(\frac{1}{p})^{\kappa|\gamma_{(-j)}|+1}\mathbb{1}[|\gamma_{(-j)}| \leq s_0 - 1](1+g)^{-(|\gamma_{(-j)}|+1)/2}(1 + g(1 - R^2_{\gamma_j=0,\gamma_{(-j)}}))^{n/2}}\right)\right]^{-1} \tag{13}$$

$$= \left[1 + \left(\frac{1 + g(1 - R^2_{\gamma_j=1,\gamma_{(-j)}})}{1 + g(1 - R^2_{\gamma_j=0,\gamma_{(-j)}})}\right)^{n/2}(1+g)^{1/2}p^{\kappa}\right]^{-1}\mathbb{1}[|\gamma_{(-j)}| \leq (s_0 - 1)] \tag{14}$$