Stochastic Matching Pursuit for Bayesian Variable Selection and Analysis of Supersaturated Design

Ray-Bing Chen

Institute of Statistics, National University of Kaohsiung

Joint with Ying Nian Wu (UCLA); Chi-Hsiang Chu, Te-You Lai and Jian-Zhong Weng (NUK)
1. Variable Selection Problems
1. Variable Selection Problems
2. Stochastic Variable Selection Methods
1. Variable Selection Problems
2. Stochastic Variable Selection Methods
3. Large $n$ Small $p$ Problems
1. Variable Selection Problems
2. Stochastic Variable Selection Methods
3. Large $n$ Small $p$ Problems
4. Small $n$ Large $p$ Problems
1. Variable Selection Problems
2. Stochastic Variable Selection Methods
3. Large $n$ Small $p$ Problems
4. Small $n$ Large $p$ Problems
5. Comparison with Conjugate Prior Assumption
1. Variable Selection Problems
2. Stochastic Variable Selection Methods
3. Large $n$ Small $p$ Problems
4. Small $n$ Large $p$ Problems
5. Comparison with Conjugate Prior Assumption
6. Analysis of and Supersaturated Design
Variable Selection

- Model: \( Y = \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon \)
  - \( Y \): the \( n \)-dimensional response vector
  - \( X_i \): the \( n \)-dimensional regressor vector
  - \( \varepsilon \): white noise

- Find the “promising” model:
  \[
  Y = \beta_1^* X_1^* + \cdots + \beta_q^* X_q^* + \varepsilon.
  \]

- \( n > p \) (Large \( n \) Small \( p \))
- \( p > n \) (Small \( n \) Large \( p \))
Variable Selection Methods

- How to find the “promising” variables, $X_1^*, \ldots, X_q^*$?
  - Stepwise procedures (Forward, Backward, Stepwise)
Variable Selection Methods

- How to find the “promising” variables, $X_1^*, \ldots, X_q^*$?
  - Stepwise procedures (Forward, Backward, Stepwise)
  - Cross-validation method (CV)
Variable Selection Methods

- How to find the “promising” variables, $X_1^*, \ldots, X_q^*$?
  - Stepwise procedures (Forward, Backward, Stepwise)
  - Cross-validation method (CV)
  - Information criteria (AIC, BIC, ...)

Stochastic MP
Variable Selection Methods

- How to find the “promising” variables, $X_1^*, \ldots, X_q^*$?
  - Stepwise procedures (Forward, Backward, Stepwise)
  - Cross-validation method (CV)
  - Information criteria (AIC, BIC, \ldots)
  - Lasso, Lars, Bayesian Lasso
Variable Selection Methods

- How to find the “promising” variables, $X_1^*, \ldots, X_q^*$?
  - Stepwise procedures (Forward, Backward, Stepwise)
  - Cross-validation method (CV)
  - Information criteria (AIC, BIC, \ldots)
  - Lasso, Lars, Bayesian Lasso
  - Two-stage method (Screening and Selection)
Variable Selection Methods

How to find the “promising” variables, $X_1^*, \ldots, X_q^*$?

- Stepwise procedures (Forward, Backward, Stepwise)
- Cross-validation method (CV)
- Information criteria (AIC, BIC, \ldots)
- Lasso, Lars, Bayesian Lasso
- Two-stage method (Screening and Selection)
- Stochastic Search Variable Selection
Bayesian Variable Selection

- **Stochastic Search Variable Selection (SSVS)**
  - George and McCulloch (1993, 1997)
  - Chipman (1996) and Chipman et al. (1997)
  - Smith and Kohn (1996)

- **Applications:**
  - Supersaturated design: Beattie et al. (2002)
  - Signal processing: Wolfe et al. (2004), and Févotte and Godsill (2006)
  - Gene selection: Lee et al. (2003)
  - ...
Statistical Model

- Model: 
  \[ Y = X\beta + \epsilon \]

- \( Y \) is an \( n \times 1 \) response vector.
- \( X = [X_1, \ldots, X_p] \) is an \( n \times p \) model matrix, and \( X_i \) is the \( i \)-th predictor variable or regressor.
- \( \beta = (\beta_1, \ldots, \beta_p)' \) is a \( p \times 1 \) vector of the unknown coefficients.
- \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)' \) is an \( n \times 1 \) noise vector that follows \( MN(0, \sigma^2 I_n) \)

- A \( p \times 1 \) vector of latent variables, \( \gamma = (\gamma_1, \ldots, \gamma_p)' \): 
  \[ \gamma_i = \begin{cases} 
  1, & X_i \text{ is selected;} \\
  0, & \text{otherwise.} 
  \end{cases} \]
Stochastic Search Variable Selection

George and McCulloch (1993)

Prior assumptions:

- $(\beta_i, \gamma_i), i = 1, 2, \ldots, p$ are assumed to be independent.
- $\gamma$: $P(\gamma_i = 0) = p_i$, and $P(\gamma_i = 1) = 1 - p_i$.
- $\beta$:
  
  
  $[\beta_i | \gamma_i = 0] \sim N(0, \nu_{0i})$, and $[\beta_i | \gamma_i = 1] \sim N(0, \nu_{1i})$.

  Usually set $\nu_{1i} = c_i \nu_{0i}$ and $c_i \gg 1$, i.e. $\nu_{1i} \gg \nu_{0i}$.
- $\sigma$: $\sigma^2 \sim IG(\nu/2, \nu \lambda/2)$. 

Stochastic MP
Stochastic Search Variable Selection

- Use Gibbs sampling scheme to sample from $[\beta, \sigma, \gamma|Y]$. 

$A_\gamma = (\sigma^{-2} X'X + D - 1 \gamma R - 1 D - 1)^{-1}$

$R$ is the prior correlation matrix.

$D - 2 \gamma = \text{diag}[(a_1 \nu_0)^{-1}, \ldots, (a_p \nu_0)^{-1}]$ with $a_i = 1$ if $\gamma_i = 0$, and $a_i = c_i$ if $\gamma_i = 1$. 

Stochastic MP
Stochastic Search Variable Selection

- Use Gibbs sampling scheme to sample from $[\beta, \sigma, \gamma|Y]$.
- Iteratively sample from $[\beta|\gamma, Y, \sigma]$, $[\gamma|\beta, Y, \sigma]$, and $[\sigma|Y, \beta, \gamma]$. 

The best subset of variables is selected according to the Monte Carlo samples of $\gamma$. 

The most costly step: $[\beta|\gamma, Y, \sigma] \sim \mathcal{N}(\sigma^{-2} A \gamma X' Y, A \gamma)$. 

$A \gamma = (\sigma^{-2} X' X + D - 1 \gamma R - 1 D - 1 \gamma)$.

$R$ is the prior correlation matrix. 

$D - 2 \gamma = \text{diag}[(a_{1, \nu} 0_1^{-1}, \ldots, a_{p, \nu} 0_p^{-1})]$ with $a_i = 1$ if $\gamma_i = 0$, and $a_i = c_i$ if $\gamma_i = 1$. 

Stochastic MP
Stochastic Search Variable Selection

- Use Gibbs sampling scheme to sample from $[\beta, \sigma, \gamma | Y]$
- Iteratively sample from $[\beta | \gamma, Y, \sigma]$, $[\gamma | \beta, Y, \sigma]$, and $[\sigma | Y, \beta, \gamma]$.
- The best subset of variables is selected according to the Monte Carlo samples of $\gamma$. 
Stochastic Search Variable Selection

- Use Gibbs sampling scheme to sample from $[\beta, \sigma, \gamma|Y]$.
- Iteratively sample from $[\beta|\gamma, Y, \sigma]$, $[\gamma|\beta, Y, \sigma]$, and $[\sigma|Y, \beta, \gamma]$.
- The best subset of variables is selected according to the Monte Carlo samples of $\gamma$.
- The most costly step:

$$
[\beta|\gamma, Y, \sigma] \sim N(\sigma^{-2} A_\gamma X'Y, A_\gamma),
$$

- $A_\gamma = (\sigma^{-2} X'X + D_\gamma^{-1} R^{-1} D_\gamma^{-1})^{-1}$.
- $R$ is the prior correlation matrix.
- $D_\gamma^{-2} = \text{diag}[(a_1 \nu_{01})^{-1}, \ldots, (a_p \nu_{0p})^{-1}]$ with $a_i = 1$ if $\gamma_i = 0$, and $a_i = c_i$ if $\gamma_i = 1$. 
Stochastic Search Variable Selection

- Speed up the stochastic search variable selection process:
  - Cholesky decomposition for $A_\gamma$.
  - Sample $\gamma$ componentwise, i.e. $[\gamma_i|\gamma_{-i}, Y]$.
  - When $\nu_0 = 0$, Geweke (1996) suggested to jointly draw $(\gamma_i, \beta_i)$.
  - Choose conjugate prior for $\beta$. Sample $[\gamma|Y]$ directly. (Smith and Kohn, 1996, and George and McCulloch, 1997)
Componentwise Gibbs Sampler

- The prior of $\beta$: $\beta_i | \gamma_i \sim (1 - \gamma_i)\delta_0 + \gamma_i N(0, \tau_i^2)$.
Componentwise Gibbs Sampler

- The prior of $\beta$: $\beta_i|\gamma_i \sim (1 - \gamma_i)\delta_0 + \gamma_i N(0, \tau^2_i)$.
- Sample $(\gamma_i, \beta_i)$ one at time conditioning on $(\gamma_{-i}, \beta_{-i})$. 

$\tau_i = \tau$. The key step:

$z_i = P(Y|\gamma_i = 1, \{\beta_k, \forall k \neq i\}) P(Y|\gamma_i = 0, \{\beta_k, \forall k \neq i\}) = \sqrt{\sigma^2 i^* \tau^2} \exp\{r^2_i \sigma^2 i^*\}$, where $\sigma^2 i^* = \sigma^2 \tau^2 X_i^t X_i \tau^2 + \sigma^2$, $r_i = R_i^t X_i \tau^2 \sigma^2 + X_i^t X_i \tau^2$, and $R_i = Y - \Sigma_k \neq i \beta_k X_k$. 

Set $p_i = \rho_i$. Then $P(\gamma_i = 1|\{\beta_k, \forall k \neq i\}, Y) = (1 - \rho_i)z_i \rho_i + (1 - \rho_i)z_i$. 

Stochastic MP
Componentwise Gibbs Sampler

- The prior of $\beta$: $\beta_i|\gamma_i \sim (1 - \gamma_i)\delta_0 + \gamma_i N(0, \tau_i^2)$.
- Sample $(\gamma_i, \beta_i)$ one at time conditioning on $(\gamma_{-i}, \beta_{-i})$.
- Assume $\tau_i = \tau$. 
Componentwise Gibbs Sampler

- The prior of $\beta$: $\beta_i | \gamma_i \sim (1 - \gamma_i) \delta_0 + \gamma_i N(0, \tau_i^2)$.
- Sample $(\gamma_i, \beta_i)$ one at time conditioning on $(\gamma_{-i}, \beta_{-i})$.
- Assume $\tau_i = \tau$.
- The key step:

$$z_i = \frac{P(Y | \gamma_i = 1, \{\beta_k, \forall k \neq i\})}{P(Y | \gamma_i = 0, \{\beta_k, \forall k \neq i\})} = \sqrt{\frac{\sigma_{i*}^2}{\tau^2}} \exp \left\{ \frac{r_i^2}{2\sigma_{i*}^2} \right\},$$

where $\sigma_{i*}^2 = \frac{\sigma^2 \tau^2}{X_i' X_i \tau^2 + \sigma^2}$, $r_i = \frac{R_i' X_i \tau^2}{\sigma^2 + X_i' X_i \tau^2}$, and $R_i = Y - \sum_{k \neq i} \beta_k X_k$. 

Stochastic MP
Componentwise Gibbs Sampler

- The prior of $\beta$: $\beta_i | \gamma_i \sim (1 - \gamma_i)\delta_0 + \gamma_i N(0, \tau_i^2)$.
- Sample $(\gamma_i, \beta_i)$ one at time conditioning on $(\gamma_{-i}, \beta_{-i})$.
- Assume $\tau_i = \tau$.
- The key step:

$$z_i = \frac{P(Y | \gamma_i = 1, \{\beta_k, \forall k \neq i\})}{P(Y | \gamma_i = 0, \{\beta_k, \forall k \neq i\})} = \sqrt{\frac{\sigma_i^{2*}}{\tau^2}} \exp \left\{ \frac{r_i^2}{2\sigma_i^{2*}} \right\},$$

where $\sigma_i^{2*} = \frac{\sigma^2 \tau^2}{X_i'X_i \tau^2 + \sigma^2}$, $r_i = \frac{R_i'X_i \tau^2}{\sigma^2 + X_i'X_i \tau^2}$, and $R_i = Y - \sum_{k \neq i} \beta_k X_k$.

- Set $p_i = \rho$. Then

$$P(\gamma_i = 1 | \{\beta_k, \forall k \neq i\}, Y) = \frac{(1 - \rho)z_i}{\rho + (1 - \rho)z_i}.$$
The componentwise Gibbs sampler for variable selection

(I) Randomly select a variable $X_i$.

(II) Compute
$$z_i = \frac{p(Y|\gamma_i = 1, \{\beta_k, \forall k \neq i\})}{p(Y|\gamma_i = 0, \{\beta_k, \forall k \neq i\})} = \sqrt{\frac{\sigma^2_i}{\tau^2}} \exp\left\{\frac{r_i^2}{2\sigma^2_i}\right\}.$$ Then evaluate the posterior probability
$$P(\gamma_i = 1|\{\beta_k, \forall k \neq i\}, Y) = \frac{(1 - \rho)z_i}{\rho + (1 - \rho)z_i}.$$ 

(III) Sample $\gamma_i$ from the above posterior probability. If $\gamma_i = 0$, then set $\beta_i = 0$, otherwise, sample $\beta_i \sim N(r_i, \sigma^2_{i*})$. Go back to (I).

(IV) After a number of iterations of the above steps, compute the current residual vector,
$$Res = Y - \sum_i \beta_i X_i.$$ Then sample
$$\sigma^2 \sim IG\left(\frac{n+\nu}{2}, \frac{Res'Res+\nu\lambda}{2}\right).$$ Go back to (I).
Componentwise Gibbs Sampler

- Similar to the search algorithm of Geweke (1996) with the truncated normal prior distribution.
Componentwise Gibbs Sampler

- Similar to the search algorithm of Geweke (1996) with the truncated normal prior distribution.
- Two possible problems:
Componentwise Gibbs Sampler

- Similar to the search algorithm of Geweke (1996) with the truncated normal prior distribution.
- Two possible problems:
  - The variables are highly correlated.
Componentwise Gibbs Sampler

- Similar to the search algorithm of Geweke (1996) with the truncated normal prior distribution.
- Two possible problems:
  - The variables are highly correlated.
  - The residual variance is small.
Matching Pursuit

Matching Pursuit (Mallat and Zhang, 1993): Suppose \( \|X_i\|^2 = 1 \). At each iteration,

- Select \( X_j \) such that
  \[
  j = \arg \max \{|\langle R, X_i \rangle|\}.
  \]
- Updated \( \beta_i \leftarrow \beta_i + \langle R, X_i \rangle \), and \( R \leftarrow R - \langle R, X_i \rangle X_i \).
Matching Pursuit

- Matching Pursuit (Mallat and Zhang, 1993): Suppose $\|X_i\|^2 = 1$. At each iteration,
  - Select $X_j$ such that
    
    $$j = \arg \max |\langle R, X_i \rangle|.$$ 
  - Updated $\beta_i \leftarrow \beta_i + \langle R, X_i \rangle$, and $R \leftarrow R - \langle R, X_i \rangle X_i$.

- Forward selection
Metropolized Matching Pursuit

- Metropolis scheme with a pair of reversible moves: addition and deletion moves based on
  
  \[ z_i = \frac{P(Y|\gamma_i = 1, \{\beta_k, \forall k \neq i\})}{P(Y|\gamma_i = 0, \{\beta_k, \forall k \neq i\})}. \]

  The larger \( z_i \) is, the more promising the variable \( X_i \) is.
Metropolized Matching Pursuit

- Metropolis scheme with a pair of reversible moves: addition and deletion moves based on
  \[ z_i = \frac{P(Y | \gamma_i = 1, \{\beta_k, \forall k \neq i\})}{P(Y | \gamma_i = 0, \{\beta_k, \forall k \neq i\})}. \]
- The larger \( z_i \) is, the more promising the variable \( X_i \) is.
Metropolized Matching Pursuit

- Metropolis scheme with a pair of reversible moves: addition and deletion moves based on
  \[ z_i = P(Y|\gamma_i = 1, \{\beta_k, \forall k \neq i\}) / P(Y|\gamma_i = 0, \{\beta_k, \forall k \neq i\}) \].
- The larger \( z_i \) is, the more promising the variable \( X_i \) is.
- Proposal for the next status:
Metropolized Matching Pursuit

- Metropolis scheme with a pair of reversible moves: addition and deletion moves based on
  \[ z_i = \frac{P(Y|\gamma_i = 1, \{\beta_k, \forall \ k \neq i\})}{P(Y|\gamma_i = 0, \{\beta_k, \forall \ k \neq i\})}. \]
- The larger \( z_i \) is, the more promising the variable \( X_i \) is.
- Proposal for the next status:
  - Add or delete a variable.
Metropolized Matching Pursuit

- Metropolis scheme with a pair of reversible moves: addition and deletion moves based on 
  \[ z_i = \frac{P(Y | \gamma_i = 1, \{\beta_k, \forall k \neq i\})}{P(Y | \gamma_i = 0, \{\beta_k, \forall k \neq i\})}. \]
- The larger \( z_i \) is, the more promising the variable \( X_i \) is.
- Proposal for the next status:
  - Add or delete a variable.
  - Addition proposal: Sample an inactive variable with probability proportional to \( z_i \).
Metropolized Matching Pursuit

- Metropolis scheme with a pair of reversible moves: addition and deletion moves based on
  \[ z_i = \frac{P(Y|\gamma_i = 1, \{\beta_k, \forall k \neq i\})}{P(Y|\gamma_i = 0, \{\beta_k, \forall k \neq i\})}. \]
- The larger \( z_i \) is, the more promising the variable \( X_i \) is.
- Proposal for the next status:
  - Add or delete a variable.
  - Addition proposal: Sample an inactive variable with probability proportional to \( z_i \).
  - Deletion proposal: Randomly select one active variable.
Acceptance probability for addition move:

\[
P_{\text{accept-add}} = \min \left[ 1, \frac{P(\gamma_i = 1|\{\beta_k, \forall k \neq i\}, Y)}{P(\gamma_i = 0|\{\beta_k, \forall k \neq i\}, Y)} \frac{p_{\text{delete}}}{p_{\text{add}}} \frac{1}{(A + 1)} \right] \frac{z_i}{\sum_{j: \gamma_j = 0} z_j}
\]

\[
= \min \left[ 1, \frac{(1 - \rho)}{\rho} \frac{p_{\text{delete}}}{p_{\text{add}}} \frac{\sum_{j: \gamma_j = 0} z_j}{(A + 1)} \right].
\]

(1)

Acceptance probability of the deletion move:

\[
P_{\text{accept-delete}} = \min \left[ 1, \frac{P(\gamma_i = 0|\{\beta_k, \forall k \neq i\}, Y)}{P(\gamma_i = 1|\{\beta_k, \forall k \neq i\}, Y)} \frac{p_{\text{add}}}{p_{\text{delete}}} \frac{z_i}{(\sum_{j: \gamma_j = 0} z_j + z_i)} \frac{1}{A} \right]
\]

\[
= \min \left[ 1, \frac{\rho}{(1 - \rho)} \frac{p_{\text{add}}}{p_{\text{delete}}} \frac{A}{\sum_{j: \gamma_j = 0} z_j + z_i} \right].
\]

(2)
Stochastic matching pursuit for variable selection

(I) Let $A$ be the number of active variables. With probability $p_{\text{add}}$, go to (II). With probability $p_{\text{delete}} = 1 - p_{\text{add}}$ go to (IV).

(II) With probability $p_{\text{accept-}\text{add}}$ calculated according to Eq. (1), go to (III), and with probability $1 - p_{\text{accept-}\text{add}}$ go back to (I).

(III) Among all the inactive variables $i$ with $\gamma_i = 0$, sample a variable $i$ with probability proportional to $z_i$, then let $\gamma_i = 1$ and sample $\beta_i$ as described in (III) of Algorithm 1. Go back to (I).
Stochastic matching pursuit for variable selection

(IV) If $A > 0$, then randomly select an active variable $i$ with $\gamma_i = 1$.

(V) With probability $p_{\text{accept–delete}}$ calculated according to Eq. (2), accept the proposal of deleting the variable $i$, i.e., set $\gamma_i = 0$, and $\beta_i = 0$. With probability $1 - p_{\text{accept–delete}}$, reject the proposal of deleting variable $i$, and sample $\beta_i$ as described in (III) of Algorithm 1. Go back to (I).

(VI) After a number of iterations of the above steps, compute the current residual vector, $Res = Y - \sum_i \beta_i X_i$, and then update $\sigma^2 \sim IG\left(\frac{n+\nu}{2}, \frac{Res' \cdot Res + \nu \lambda}{2}\right)$. Go back to (I).
Combine the strengths of the matching pursuit and the componentwise Gibbs sampler.

1. Pursue proposing variables.
2. Don’t need to compute the inverse of the large matrix.
Implementation Details

- After a burn-in period, use \( \{\gamma^{(i)}, i > T\} \) to estimate \( P(\gamma_j = 1|Y) \).
- Selection criteria:
  - The highest posterior probability: \( \max P(\gamma_1, \ldots, \gamma_p|Y) \).
  - The median probability criterion in Barbieri and Berger (2004): \( X_i \) is included in the model if
    \[
P(\gamma_i = 1|Y) \geq 1/2.
    \]
Implementation Details

- Tuning parameters, $p_{\text{add}}$ and $\rho$.
Implementation Details

- Tuning parameters, $p_{add}$ and $\rho$.
  - Set $p_{add} = 1/2 = p_{delete}$.
Implementation Details

- Tuning parameters, $p_{\text{add}}$ and $\rho$.
  - Set $p_{\text{add}} = 1/2 = p_{\text{delete}}$.
  - Set $\rho = 1/2$ (George and McCulloch, 1993 and 1997).
Implementation Details

- Tuning parameters, $p_{\text{add}}$ and $\rho$.
  - Set $p_{\text{add}} = 1/2 = p_{\text{delete}}$.
  - Set $\rho = 1/2$ (George and McCulloch, 1993 and 1997).
- Another tuning parameter, $\tau$
Implementation Details

- Tuning parameters, \( p_{add} \) and \( \rho \).
  - Set \( p_{add} = 1/2 = p_{delete} \).
  - Set \( \rho = 1/2 \) (George and McCulloch, 1993 and 1997).

- Another tuning parameter, \( \tau \)
  - \( z_i \) is a decreasing function of \( \tau \). Then
    \[
    P(\gamma_i = 1|\{\beta_k, k \neq i\}, Y) \text{ is smaller for larger } \tau.
    \]
Implementation Details

- Tuning parameters, $p_{\text{add}}$ and $\rho$.
  - Set $p_{\text{add}} = 1/2 = p_{\text{delete}}$.
  - Set $\rho = 1/2$ (George and McCulloch, 1993 and 1997).

- Another tuning parameter, $\tau$
  - $z_i$ is a decreasing function of $\tau$. Then $P(\gamma_i = 1|\{\beta_k, k \neq i\}, Y)$ is smaller for larger $\tau$.
  - Cross-Validation approach for selecting $\tau$:
    Use $K$-fold CV (or Monte Carlo CV) to choose “proper” value of $\tau$. Thus
    $$\hat{\tau} = \arg \min_{\tau} \sum_{k=1}^{K} \sum_{j} (y_{kj} - \hat{y}_{-kj}(\tau))^2.$$
Implementation Details

- Tuning parameters, $p_{\text{add}}$ and $\rho$.
  - Set $p_{\text{add}} = 1/2 = p_{\text{delete}}$.
  - Set $\rho = 1/2$ (George and McCulloch, 1993 and 1997).

- Another tuning parameter, $\tau$
  - $z_i$ is a decreasing function of $\tau$. Then
    \[ P(\gamma_i = 1|\{\beta_k, k \neq i\}, Y) \] is smaller for larger $\tau$.
  - Cross-Validation approach for selecting $\tau$:
    Use $K$-fold CV (or Monte Carlo CV) to choose “proper” value of $\tau$. Thus
    \[
    \hat{\tau} = \arg \min_{\tau} \sum_{k=1}^{K} \sum_{j} (y_{k,j} - \hat{y}_{-k,j}(\tau))^2.
    \]

- $\rho$ can also be selected by this CV approach.
Large $n$ Small $p$

Example 3.1: $(n, p) = (60, 5)$

- These five variables, $X_1, \ldots, X_5 \overset{iid}{\sim} N_{60}(0, I_{60})$.
- The response variable is generated by

$$Y = X_4 + 1.2X_5 + \epsilon,$$

where $\epsilon \sim N_{60}(0, I_{60})$.

- Set $(\rho, \tau) = (0.5, 10)$ for SMP and set $(\nu_0, c) = (0.01, 2500)$ for SSVS.

- Totally there are 1000 replications. Draw 3000 samples from posterior samples.
### Table 1: Variable selection results in Example 3.1

<table>
<thead>
<tr>
<th>method</th>
<th>Number of selected variables</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMP</td>
<td>$f_1$</td>
<td>0</td>
<td>0</td>
<td>997</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>0</td>
<td>0</td>
<td>997</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SSVS</td>
<td>$f_1$</td>
<td>0</td>
<td>0</td>
<td>997</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>0</td>
<td>0</td>
<td>997</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Large $n$ small $p$

Example 3.2: $(n, p) = (60, 10)$

- 10 variables, $X_1, \ldots, X_{10} \overset{iid}{\sim} N_{60}(0, I_{60})$.
- The true model is

$$Y = 2X_1 + 3X_2 + 4X_5 + 5X_6 + 6X_9 + 7X_{10} + \epsilon,$$

where $\epsilon \sim N_{60}(0, 2.5^2 I_{60})$.

- Set $(\rho, \tau) = (0.5, 15)$ for SMP and set $(\nu_0, c) = (0.01, 2500)$ for SSVS.

- Totally there are 1000 replications. In each replication, draw 3000 samples.
Large $n$ small $p$

Table 2: Variable selection results in Example 3.2

<table>
<thead>
<tr>
<th>method</th>
<th>Number of selected variables</th>
<th>( \leq 3 )</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>( \geq 8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMP</td>
<td>( f_1 )</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>961</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( f_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>961</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>SSVS</td>
<td>( f_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>934</td>
<td>64</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( f_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>934</td>
<td>64</td>
<td>1</td>
</tr>
</tbody>
</table>
Computational Cost

Table 3: CPU times (in seconds) of 10,000 iterations

<table>
<thead>
<tr>
<th>Description</th>
<th>SMP</th>
<th>SSVS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU time in Example 3.1 ((p = 5))</td>
<td>14.6s</td>
<td>9.3s</td>
</tr>
<tr>
<td>CPU time in Example 3.2 ((p = 10))</td>
<td>39.5s</td>
<td>20.8s</td>
</tr>
<tr>
<td>CPU time with (p = 100)</td>
<td>2016.6s</td>
<td>7266.4s</td>
</tr>
</tbody>
</table>
Small $n$ Large $p$

- The gene selection problem in microarray experiments: the number of candidate genes, $p >$ the number of available sample size, $n$. (Yi et al., 2003, and Lee et al., 2003)
- Overcomplete signal representation: the number of basis functions, $p >$ the size of the signal, $n$. (Wolf et al., 2004)
- Sparse assumption.
Simulations for Small $n$ Large $p$ Problem

- Shao and Chow (2007) studied the small $n$ large $p$ problem in microarray experiments.
- The ridge regression estimator for $\beta$ is
  \[
  \hat{\beta} = (X'X + h_n I_p)^{-1}X'Y = R_D X'Y,
  \]
  - $I_p$ is the $p \times p$ identity matrix.
  - $h_n$ is the ridge parameter.
  - $R_D = (X'X + h_n I_p)^{-1}$.
- Screen out $X_i$ if $|\hat{\beta}_i| \leq a_n$, and $a_n \to 0$ as $n \to \infty$.
- Their procedure is asymptotically consistent and their idea is similar to that of the Lasso method (Tibshirani, 1996).
Simulation

- $(n, p) = (50, 200)$ and $(100, 400)$.
- There are 5 true active variables, and
  \[ \beta = (3, -3.5, 4, -2.8, 3.2, 0, \ldots, 0)' \]
- The regressor $X_i$ is generated by
  \[ X_i = G_i + \lambda G, \]
  where $G_i$ and $G \sim N_n(0, I_n)$, and $\lambda = 0$ or 1.
- $\epsilon \sim N_n(0, I_n)$. 

Stochastic MP
Simulation

- Three methods are used here, Shao and Chow (2007), SMP and Lasso + CV.
- The screening method of Shao and Chow (2007): Set $h_n = n^{2/3}$ and $a_n = n^{-1/6}$.
- SMP:
  - $\tau$ is selected by 5-fold CV from $\{80, 120, 160, 220\}$ for $(n, p) = (50, 200)$ and from $\{100, 150, 200, 250\}$ for $(n, p) = (100, 400)$.
  - Draw 2000 posterior samples by taking every $p$th sample.
Simulation

- **Lasso + CV**: There is a Matlab implementation of the homotopy/LARS-LASSO algorithm for tracing the regularization path of the L1-penalized squared error loss (Rocha, 2006), and this tool-box is available at [http://www.stat.berkeley.edu/twiki/Research/YuGroup/Software](http://www.stat.berkeley.edu/twiki/Research/YuGroup/Software).

- **Stopping criterion**: 
  \[ \|X'(Y - X'\hat{\beta})\|_\infty < b. \]

- **lasso_cv**: Fit the parameters of a linear model by using the lasso and k-folds cross validation.
  - \( b \in \{2 \times 10^{-1}, 10^{-1}, 10^{-2}, \ldots, 10^{-5}, 10^{-8}\} \).
  - 10-folds CV is used here.
  - Identify \( \{i|\beta_i| > 0\} \).
Frequencies based on 100 replications with \((n, p) = (50, 200)\)

<table>
<thead>
<tr>
<th>(\lambda)</th>
<th>method</th>
<th>Number of selected variables</th>
<th>(\leq 2)</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>(\geq 11)</th>
<th># of sel.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>SC</td>
<td>(f_1)</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td>14</td>
<td>20</td>
<td>18</td>
<td>9</td>
<td>10</td>
<td>4</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>13</td>
<td>9</td>
<td>9</td>
<td>3</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Lasso</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>5</td>
<td>10</td>
<td>14</td>
<td>6</td>
<td>4</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>5</td>
<td>10</td>
<td>14</td>
<td>6</td>
<td>4</td>
<td>44</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>SMP</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>94</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>94</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>SC</td>
<td>(f_1)</td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>21</td>
<td>10</td>
<td>19</td>
<td>13</td>
<td>12</td>
<td>4</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Lasso</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>SMP</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
Frequencies based on 100 replications with \((n, p) = (100, 400)\)

<table>
<thead>
<tr>
<th>(\lambda)</th>
<th>method</th>
<th>Number of selected variables</th>
<th>(\leq 2)</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>(\geq 11)</th>
<th># of sel.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>SC</td>
<td>(f_1)</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>49</td>
<td>22</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>41</td>
<td>19</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Lasso</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>75</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>75</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>SMP</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>SC</td>
<td>(f_1)</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>35</td>
<td>28</td>
<td>13</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>28</td>
<td>13</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Lasso</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>67</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>SMP</td>
<td>(f_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>✓</td>
<td>100</td>
</tr>
</tbody>
</table>

Stochastic MP
An illustration in Image Representation

- Gabor regression model (Wolf et al., 2004) is

\[ f = \sum_{i} c_i g_i + \varepsilon, \]

where \( g_i \)'s are the Gabor basis functions.

- The Gabor basis function can be defined as

\[
g(u, v) = \exp \left[ -\frac{1}{2} (\sigma_u u^2 + \sigma_v v^2) \right] \cos \left[ \frac{2\pi u}{\lambda} + \varphi \right],
\]

\[
u = u_0 + x_1 \cos \theta - x_2 \sin \theta,
\]

\[
v = v_0 + x_1 \sin \theta - x_2 \cos \theta,
\]
Gabor Regression Function

Stochastic MP
Give a grid
\[ \mathcal{X} = \{(x_1, x_2) | x_1 \in \{22, 24, \ldots, 40\} \text{ and } x_2 \in \{7, 9, \ldots, 25\}\}. \]

Totally we have 200 Gabor basis functions on \( \mathcal{X} \) by setting \( \varphi = 0, \sigma_u = 1 \) and \( \theta \in \{0, 3/8\pi\} \).

The response is generated by
\[
Y = 7X_{17} - 7X_{71} + 7X_{161} - 7X_{177} + \varepsilon,
\]

SMP:
- \( \tau \) is chosen from \( A = \{50, 100, \ldots, 300\} \) by Monte Carlo cross validation with 100 replications.
- Draw 3000 samples by taking every \( p \)th sample.

<table>
<thead>
<tr>
<th>Selected Bases</th>
<th>SNR1</th>
<th>SNR2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_{17} )</td>
<td>0.9957</td>
<td></td>
</tr>
<tr>
<td>( X_{71} )</td>
<td>0.5243</td>
<td></td>
</tr>
<tr>
<td>( X_{73} )</td>
<td>0.6317</td>
<td></td>
</tr>
<tr>
<td>( X_{161} )</td>
<td>0.6933</td>
<td></td>
</tr>
<tr>
<td>( X_{177} )</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>SNR1</td>
<td>0.373</td>
<td>0.080</td>
</tr>
<tr>
<td>SNR2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conjugate Prior for $\beta$

- Smith and Kohn (1996): The prior of $\beta$ given $\gamma$ is $N(0, c\sigma^2(X_\gamma X_\gamma)^{-1})$.
- Obtain $[\gamma|Y]$ by integrating $\beta$ and $\sigma^2$ out.

\[
P(\gamma|Y) \propto (1 + c)^{-q_\gamma/2} S(\gamma)^{-n/2} \prod_{i=1}^{p} p_i^{\gamma_i} (1 - p_i)^{1-\gamma_i},
\]

where $q_\gamma$ is the number of selected variables and

\[
S(\gamma) = Y'Y - \frac{c}{1 + c} Y'X_\gamma (X_\gamma X_\gamma)^{-1} X_\gamma'Y.
\]

- Use Gibbs sampler to generate $\gamma_i|Y, \gamma_{-i}$.
- Need to prespecify the prior parameter $c$. When the norm of $X_i$ is equal to 1, $c \in [10, 1000]$. 

Stochastic MP
Simulations for the algorithm of Smith and Kohn (1996):

- Fix $p_i = \rho = 1/2$.
- Set $n = 50$ and $p = 20, 50, 100, 300$.
- The variables, $X_1, \ldots, X_p \overset{iid}{\sim} N_n(0, I_n)$.
- The response variable is generated by

$$Y = 3X_1 + 3X_2 + \cdots + 3X_{10} + \epsilon,$$

where $\epsilon \sim N_n(0, I_n)$.

- The median probability criterion

<table>
<thead>
<tr>
<th>$c$</th>
<th>$p = 20$</th>
<th>$p = 50$</th>
<th>$p = 100$</th>
<th>$p = 300$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>√</td>
<td>√</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>√</td>
<td>√</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1000</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- SMP with $\tau = 250$ works.
Summarization

- Stochastic matching pursuit + median probability criterion works for both the cases of large $n$ small $p$ and small $n$ large $p$.
- Tune the parameters, $\rho$ and $\tau$, via CV approach.
- “Full” Bayesian procedure
- CPU times: Componentwise Gibbs sampler < SMP < SSVS
- Window (or block) version
- Selection criterion
- Theoretical Properties
- Other applications
Analysis of Supersaturated Design

Supersaturate design:
- Investigates $p$ factors in only $n(< p + 1)$ experimental runs.
- Particularly useful in factor screening.

Analysis methods:
- Chipman (1996) and Chipman et al. (1997): Propose different priors for SSVS.
- Beattie et al. (2002): A two-stage method via SSVS.
Analysis Approach

- Use componentwise Gibbs sampler:
  - The sample correlations between the factors are not so high.
  - The variance would not be too small.
- Follow the pre-process in Phoa et al. (2009), standardize $Y$ and $X_i$’s are unit norm.
- Use leave-two-out cross-validation approach to choose the proper parameters, $\rho$ and $\tau$.
- Selection criterion: the median probability criterion and the highest posterior probability criterion.
### Example 1. Cast Fatigue Experiment

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>3</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>7</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>9</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>
Example 1. Cast Fatigue Experiment

- Consider main effect model. i.e. \( n = 12 \) and \( p = 7 \).
- Fix \( \rho = 1/2 \).
- Iterate \( 10000 \times p \) times and get 1000 samples from last \( 5000 \times p \) iterations.
- \( \tau \) is select from \( \mathcal{A} = \{1, 2, 3, 4, 5\} \). \( \hat{\tau} = 2 \).
- The marginal posterior probabilities

<table>
<thead>
<tr>
<th>Variable</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>0.350</td>
<td>0.353</td>
<td>0.341</td>
<td>0.553</td>
<td>0.292</td>
<td>0.899</td>
<td>0.279</td>
</tr>
</tbody>
</table>

- Wu and Hamada (2000) and Phoa et al. (2009): \([F(D)]\)
Example 1. Cast Fatigue Experiment

- Consider main effects + two-factor interactions. i.e. \( n = 12 \) and \( p = 28 \).
- \( \tau \) is select from \( A = \{40, 80, 120, 160, 200\} \). \( \hat{\tau} = 120 \).
- The marginal posterior probabilities

<table>
<thead>
<tr>
<th>Variable</th>
<th>F</th>
<th>FG</th>
<th>AE</th>
<th>AC</th>
<th>BD</th>
<th>BC</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>0.763</td>
<td>0.759</td>
<td>0.129</td>
<td>0.015</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
</tr>
</tbody>
</table>

- The highest posterior probability criterion: \([F \ FG]\).
- Same as Phoa et al. (2009) by mAIC.
Example 2. Blood Glucose Experiment

- Sample size, $n = 18$.
- $p = 15$: 1 two-level factors, $A$, 7 three-level factors, $B, \ldots, H$ and 7 quadratic contrasts of these seven three-level factors, $B^2, \ldots, H^2$.
- $\tau$ is select from $A = \{3, 4, 5, 6, 7, 8\}$. $\widehat{\tau} = 4$.
- The marginal posterior probabilities

<table>
<thead>
<tr>
<th>Variable</th>
<th>$F^2$</th>
<th>$E^2$</th>
<th>$C$</th>
<th>$B$</th>
<th>$G$</th>
<th>$F$</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>0.596</td>
<td>0.538</td>
<td>0.384</td>
<td>0.383</td>
<td>0.378</td>
<td>0.364</td>
<td>0.322</td>
</tr>
</tbody>
</table>

- Same as Wu and Hamada (2000) and Phoa et al. (2009)
Example 2. Blood Glucose Experiment

- Include two-factor interaction terms, $p = 113$.
- $\tau$ is select from $\mathcal{A} = \{40, 80, 120, 160, 200\}$. $\hat{\tau} = 80$.
- The marginal posterior probabilities

<table>
<thead>
<tr>
<th>Variable</th>
<th>$BH^2$</th>
<th>$B^2H^2$</th>
<th>$EG$</th>
<th>$AH^2$</th>
<th>$DE$</th>
<th>$BC$</th>
<th>$DE^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>0.821</td>
<td>0.748</td>
<td>0.578</td>
<td>0.496</td>
<td>0.154</td>
<td>0.147</td>
<td>0.145</td>
</tr>
</tbody>
</table>

- The highest posterior probability criterion:

<table>
<thead>
<tr>
<th>Model</th>
<th>Post. Prob.</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AH^2 BH^2 EG B^2H^2$</td>
<td>0.116</td>
<td>0.9568</td>
</tr>
<tr>
<td>$BH^2 B^2H^2$</td>
<td>0.027</td>
<td>0.7696</td>
</tr>
<tr>
<td>$BH^2 EG B^2H^2$</td>
<td>0.018</td>
<td>0.8737</td>
</tr>
<tr>
<td>$AH^2 BH^2 EG B^2H^2 E^2G^2$</td>
<td>0.017</td>
<td>0.9766</td>
</tr>
</tbody>
</table>

Stochastic MP
Example 3. An Example in Lin (1993)

- A supersaturated design with \( n = 14 \) and \( p = 23 \).
- Fix \( \rho = 1/2 \).
- Iterate 10000 times and get 1000 samples from last 5000 iterations.
- \( \tau \) is select from \( \mathcal{A} = \{20, 40, 60, 80, 100\} \). \( \hat{\tau} = 20 \).
- The marginal posterior probabilities

<table>
<thead>
<tr>
<th>Variable</th>
<th>14</th>
<th>12</th>
<th>19</th>
<th>4</th>
<th>10</th>
<th>11</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>0.967</td>
<td>0.574</td>
<td>0.561</td>
<td>0.444</td>
<td>0.099</td>
<td>0.069</td>
<td>0.063</td>
</tr>
</tbody>
</table>
Example 3. An Example in Lin (1993)

- The highest posterior probability criterion:

<table>
<thead>
<tr>
<th>Model</th>
<th>Post. Prob.</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 12 14 19</td>
<td>0.206</td>
<td>0.9548</td>
</tr>
<tr>
<td>14</td>
<td>0.133</td>
<td>0.6317</td>
</tr>
<tr>
<td>12 14 19</td>
<td>0.034</td>
<td>0.8706</td>
</tr>
<tr>
<td>12 14</td>
<td>0.031</td>
<td>0.7401</td>
</tr>
<tr>
<td>14 19</td>
<td>0.023</td>
<td>0.7225</td>
</tr>
</tbody>
</table>

- Phoa et al. (2009): [14]
Future Works

- Apply SMP when $p$ is large.
Future Works

- Apply SMP when $p$ is large.
- Select $(\rho, \tau)$ via CV approach.
Future Works

- Apply SMP when $p$ is large.
- Select $(\rho, \tau)$ via CV approach.
- Two-stage procedure via CGS: First screen out useless factors and then select the important factors.
Future Works

- Apply SMP when $p$ is large.
- Select $(\rho, \tau)$ via CV approach.
- Two-stage procedure via CGS: First screen out useless factors and then select the important factors.
- Other examples
Future Works

- Apply SMP when $p$ is large.
- Select $(\rho, \tau)$ via CV approach.
- Two-stage procedure via CGS: First screen out useless factors and then select the important factors.
- Other examples
- One-stage method or two-stage method?

Stochastic MP
Future Works

- Apply SMP when $p$ is large.
- Select $(\rho, \tau)$ via CV approach.
- Two-stage procedure via CGS: First screen out useless factors and then select the important factors.
- Other examples
- One-stage method or two-stage method?
- The idea of Chipman (1996) and Chipman et al. (1997)
Example 3. An Example in Lin (1993)

- Main effects + Two-factor interaction effects
- Totally 252 variables (23 + 229)
- Fix $\rho = 1/2$.
- Iterate 10000 times and get 1000 samples from last 5000 iterations.
- $\tau$ is select from $\mathcal{A} = \{150, 170, 190, 210, 230\}$. $\hat{\tau} = 170$.
- The marginal posterior probabilities

<table>
<thead>
<tr>
<th>Var.</th>
<th>14</th>
<th>7 × 15</th>
<th>13 × 20</th>
<th>6 × 10</th>
<th>3 × 5</th>
<th>7 × 19</th>
<th>9 × 22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>0.367</td>
<td>0.133</td>
<td>0.116</td>
<td>0.059</td>
<td>0.057</td>
<td>0.056</td>
<td>0.053</td>
</tr>
</tbody>
</table>
Example 3. An Example in Lin (1993)

- Select the variables whose marginal probabilities > 0.04.
- Totally 21 variables.
- Fix $\rho = 1/2$.
- Iterate 10000 times and get 1000 samples from last 5000 iterations.
- $\tau$ is select from $\mathcal{A} = \{5, 10, 15, 20, 25\}$. $\widehat{\tau} = 5$.
- The marginal posterior probabilities

<table>
<thead>
<tr>
<th>Var.</th>
<th>5 × 20</th>
<th>23</th>
<th>14</th>
<th>6 × 10</th>
<th>11</th>
<th>9 × 21</th>
<th>7 × 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>0.593</td>
<td>0.551</td>
<td>0.548</td>
<td>0.542</td>
<td>0.47</td>
<td>0.45</td>
<td>0.314</td>
</tr>
</tbody>
</table>

Stochastic MP