Handling factors in variable selection problems

arxiv.org/abs/1709.07238

Gonzalo Garcia-Donato[1] and Rui Paulo[2]

[1] Universidad de Castilla-La Mancha (Spain), [2] Universidad de Lisboa (Portugal)

Workshop Métodos Bayesianos'17. 8/11/17

Bayesian variable selection

Competing models

- Main uncertainty concerns which (numerical) variables of a given set

$$\{x_1, x_2, \ldots, x_k\}$$

explain a response variable y. Other variables are known to explain y (eg. the constant). Focus here is on linear models and the response y is Gaussian.

Bayesian variable selection

Competing models

• Main uncertainty concerns which (numerical) variables of a given set

$$\{x_1, x_2, \ldots, x_k\}$$

explain a response variable y. Other variables are known to explain y (eg. the constant). Focus here is on linear models and the response y is Gaussian.
• The formal Bayesian answer considers all possible models that arise when different combination of the potential variables are chosen.
• There are a total of 2^k models, that normally are notated through the use of a binary vector $\boldsymbol{\gamma}$.

$$M_\gamma : y_i = \alpha + \beta_1 x_{i1} + \beta_7 x_{i7} + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2),\ i = 1, \ldots, n,$$

or

$$M_{\gamma^\star} : y_i = \alpha + \beta_1 x_{i1} + \beta_4 x_{i4} + \epsilon_i,\ \epsilon_i \sim N(0, \sigma^2),\ i = 1, \ldots, n.$$

Bayesian variable selection

Competing models (cont')

- The simplest possible model (null model) is

$$M_{(0,\ldots,0)}: y_i = \alpha + \epsilon_i,\ \ \epsilon_i \sim N(0,\sigma^2),\ i = 1,\ldots,n.$$

and the most complex model (full model) is

$$M_{(1,\ldots,1)}: y_i = \alpha + \sum_{j=1}^{k} \beta_j\, x_{ij} + \epsilon_i,\ \ \epsilon_i \sim N(0,\sigma^2),\ i = 1,\ldots,n.$$

Bayesian variable selection

Competing models (cont')

- The simplest possible model (null model) is

$$M_{(0,\ldots,0)} : y_i = \alpha + \epsilon_i,\ \ \epsilon_i \sim N(0,\sigma^2),\ i = 1,\ldots,n.$$

and the most complex model (full model) is

$$M_{(1,\ldots,1)} : y_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i,\ \ \epsilon_i \sim N(0,\sigma^2),\ i = 1,\ldots,n.$$

- In general, model M_γ can be compactly expressed as

$$M_\gamma : \boldsymbol{y} = \mathbf{1}\alpha + \boldsymbol{X}_\gamma \boldsymbol{\beta}_\gamma + \boldsymbol{\epsilon},$$

where $\boldsymbol{X}_\gamma$ is the design matrix (assume columns centered) that has k_γ columns.

Bayesian variable selection

Competing models (cont')

- The simplest possible model (null model) is

$$M_{(0,\dots,0)} : y_i = \alpha + \epsilon_i,\ \ \epsilon_i \sim N(0,\sigma^2),\ i = 1,\dots,n.$$

and the most complex model (full model) is

$$M_{(1,\dots,1)} : y_i = \alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i,\ \ \epsilon_i \sim N(0,\sigma^2),\ i = 1,\dots,n.$$

- In general, model M_γ can be compactly expressed as

$$M_\gamma : \boldsymbol{y} = \mathbf{1}\alpha + \boldsymbol{X}_\gamma \boldsymbol{\beta}_\gamma + \boldsymbol{\epsilon},$$

where $\boldsymbol{X}_\gamma$ is the design matrix (assume columns centered) that has k_γ columns.

The Bayesian answer and the prior inputs

The Bayesian method then proceeds computing posterior probabilities of all 2^k the models:

$$Pr(M_\gamma \mid \boldsymbol{y}) \propto m_\gamma(\boldsymbol{y})\, Pr(M_\gamma),\;\; m_\gamma(\boldsymbol{y}) = \int M_\gamma(\boldsymbol{y} \mid \boldsymbol{\beta}_\gamma, \sigma, \alpha)\pi_\gamma(\boldsymbol{\beta}_\gamma, \sigma, \alpha)\, d\boldsymbol{\beta}_\gamma d\sigma d\alpha.$$

The Bayesian answer and the prior inputs

The Bayesian method then proceeds computing posterior probabilities of all 2^k the models:

$$Pr(M_\gamma \mid \boldsymbol{y}) \propto m_\gamma(\boldsymbol{y})\, Pr(M_\gamma), \;\; m_\gamma(\boldsymbol{y}) = \int M_\gamma(\boldsymbol{y} \mid \boldsymbol{\beta}_\gamma, \sigma, \alpha)\pi_\gamma(\boldsymbol{\beta}_\gamma, \sigma, \alpha)\, d\boldsymbol{\beta}_\gamma d\sigma d\alpha.$$

- The ratio $\frac{m_\gamma(\boldsymbol{y})}{m_0(\boldsymbol{y})} \equiv B_\gamma$ is called the Bayes factor of M_γ to the null model.

The Bayesian answer and the prior inputs

The Bayesian method then proceeds computing posterior probabilities of all 2^k the models:

$$Pr(M_\gamma \mid \boldsymbol{y}) \propto m_\gamma(\boldsymbol{y})\, Pr(M_\gamma),\;\; m_\gamma(\boldsymbol{y}) = \int M_\gamma(\boldsymbol{y} \mid \boldsymbol{\beta}_\gamma, \sigma, \alpha)\pi_\gamma(\boldsymbol{\beta}_\gamma, \sigma, \alpha)\, d\boldsymbol{\beta}_\gamma d\sigma d\alpha.$$

- The ratio $\frac{m_\gamma(\boldsymbol{y})}{m_0(\boldsymbol{y})} \equiv B_\gamma$ is called the Bayes factor of M_γ to the null model.
- The method is parsimonious since the Bayes factor automatically penalizes complexity.

The Bayesian answer and the prior inputs

The Bayesian method then proceeds computing posterior probabilities of all 2^k the models:

$$Pr(M_\gamma \mid \mathbf{y}) \propto m_\gamma(\mathbf{y})\, Pr(M_\gamma), \;\; m_\gamma(\mathbf{y}) = \int M_\gamma(\mathbf{y} \mid \boldsymbol{\beta}_\gamma, \sigma, \alpha)\pi_\gamma(\boldsymbol{\beta}_\gamma, \sigma, \alpha)\, d\boldsymbol{\beta}_\gamma d\sigma d\alpha.$$

- The ratio $\frac{m_\gamma(\mathbf{y})}{m_0(\mathbf{y})} \equiv B_\gamma$ is called the Bayes factor of M_γ to the null model.
- The method is parsimonious since the Bayes factor automatically penalizes complexity.

Of great importance is the specification of

- Prior for parameters under each model: $\pi_\gamma(\boldsymbol{\beta}_\gamma, \alpha, \sigma)$, and

The Bayesian answer and the prior inputs

The Bayesian method then proceeds computing posterior probabilities of all 2^k the models:

$$Pr(M_\gamma \mid \boldsymbol{y}) \propto m_\gamma(\boldsymbol{y})\, Pr(M_\gamma),\;\; m_\gamma(\boldsymbol{y}) = \int M_\gamma(\boldsymbol{y} \mid \boldsymbol{\beta}_\gamma, \sigma, \alpha)\pi_\gamma(\boldsymbol{\beta}_\gamma, \sigma, \alpha)\, d\boldsymbol{\beta}_\gamma d\sigma d\alpha.$$

- The ratio $\frac{m_\gamma(\boldsymbol{y})}{m_0(\boldsymbol{y})} \equiv B_\gamma$ is called the Bayes factor of M_γ to the null model.
- The method is parsimonious since the Bayes factor automatically penalizes complexity.

Of great importance is the specification of

- Prior for parameters under each model: $\pi_\gamma(\boldsymbol{\beta}_\gamma, \alpha, \sigma)$, and
- prior over the model space $Pr(M_\gamma)$.

About $\pi_\gamma(\boldsymbol{\beta}_\gamma, \alpha, \sigma)$ and the conventional approach

What we call ‘conventional’ approach are a family of priors of the form:

$$\pi_\gamma(\alpha, \boldsymbol{\beta}_\gamma, \sigma) = \sigma^{-1} \int N_{k_\gamma}(\boldsymbol{\beta} \mid \mathbf{0}, g\sigma^2(\boldsymbol{X}^t\boldsymbol{X})^{-1})\, h(g)\, dg,$$

where $h(g)$ is a mixing function and for the null

$$\pi_0(\alpha, \sigma) = \sigma^{-1}.$$

• Conventional priors follow the tradition of Jeffreys-Zellner-Siow (60‘s and 80’s), followed by many others (90’s and 00’s) and recently formally endorsed by Bayarri et al (2012).

About $\pi_\gamma(\boldsymbol{\beta}_\gamma, \alpha, \sigma)$ and the conventional approach

What we call 'conventional' approach are a family of priors of the form:

$$\pi_\gamma(\alpha, \boldsymbol{\beta}_\gamma, \sigma) = \sigma^{-1} \int N_{k_\gamma}(\boldsymbol{\beta} \mid \mathbf{0}, g\sigma^2(\boldsymbol{X}^t\boldsymbol{X})^{-1})\, h(g)\, dg,$$

where $h(g)$ is a mixing function and for the null

$$\pi_0(\alpha, \sigma) = \sigma^{-1}.$$

• Conventional priors follow the tradition of Jeffreys-Zellner-Siow (60's and 80's), followed by many others (90's and 00's) and recently formally endorsed by Bayarri et al (2012).

• You obtain

$$B_\gamma = \mathcal{B}\Big(\frac{SSE_\gamma}{SSE_0}, 1, k_\gamma + 1, n\Big),$$

where SSE_γ is the sum of squared errors and $\mathcal{B}$ is a univariate integral.

About $Pr(M_\gamma)$ and the multiplicity issue

About $Pr(M_\gamma)$ and the multiplicity issue

- The Bayes factor already penalizes complexity so this should not be done through $Pr(M_\gamma)$.

About $Pr(M_\gamma)$ and the multiplicity issue

- The Bayes factor already penalizes complexity so this should not be done through $Pr(M_\gamma)$.
- The standard default prior over the model space is

$$Pr(M_\gamma) = 1/2^k,$$

but this tends to favor models of dimension $\approx k/2$, particularly if k is large.

About $Pr(M_\gamma)$ and the multiplicity issue

- The Bayes factor already penalizes complexity so this should not be done through $Pr(M_\gamma)$.
- The standard default prior over the model space is

$$Pr(M_\gamma) = 1/2^k,$$

  but this tends to favor models of dimension $\approx k/2$, particularly if k is large.
- Our preferred prior is

$$Pr(M_\gamma \mid \tau) = \tau^{k_\gamma}(1-\tau)^{k-k_\gamma}, \; \tau \sim U(0,1),$$

  which was studied by Scott and Berger (2010), who argued adjusts for multiplicity.

About $Pr(M_\gamma)$ and the multiplicity issue

- The Bayes factor already penalizes complexity so this should not be done through $Pr(M_\gamma)$.
- The standard default prior over the model space is

$$Pr(M_\gamma) = 1/2^k,$$

  but this tends to favor models of dimension $\approx k/2$, particularly if k is large.
- Our preferred prior is

$$Pr(M_\gamma \mid \tau) = \tau^{k_\gamma}(1-\tau)^{k-k_\gamma},\ \tau \sim U(0,1),$$

  which was studied by Scott and Berger (2010), who argued adjusts for multiplicity.This adjustment effect becomes clear with the alternative form:

$$Pr(M_\gamma) = 1/\{\#\text{ of models of dimension } k_\gamma\}$$

  which has also the form

Factors as (potential) explanatory variables

Factors

A factor, Λ, is a categorical variable (eg. nationality, sex, etc) and for each sample unit takes only one of several, ℓ, categories or levels (eg. "Español/a", "Francés/a", "Argentino/a").

Factors as (potential) explanatory variables

Factors

A factor, Λ, is a categorical variable (eg. nationality, sex, etc) and for each sample unit takes only one of several, ℓ, categories or levels (eg. "Español/a", "Francés/a", "Argentino/a").

- In many applied problems, factors are considered as possible explanatory variables jointly with perhaps numerical variables.

Factors as (potential) explanatory variables

Factors

A factor, Λ, is a categorical variable (eg. nationality, sex, etc) and for each sample unit takes only one of several, ℓ, categories or levels (eg. "Español/a", "Francés/a", "Argentino/a").

- In many applied problems, factors are considered as possible explanatory variables jointly with perhaps numerical variables.

Main goal

"Repeat" the variable selection exercise but now selecting among certain set of potential covariates and/or factors

$$\{x_1, \ldots, x_k, \Lambda_1, \ldots, \Lambda_p\}$$

Factors as (potential) explanatory variables

Factors

A factor, Λ, is a categorical variable (eg. nationality, sex, etc) and for each sample unit takes only one of several, ℓ, categories or levels (eg. "Español/a", "Francés/a", "Argentino/a").

- In many applied problems, factors are considered as possible explanatory variables jointly with perhaps numerical variables.

Main goal

"Repeat" the variable selection exercise but now selecting among certain set of potential covariates and/or factors

$$\{x_1, \ldots, x_k, \Lambda_1, \ldots, \Lambda_p\}$$

In principle the problem is solved using dummies, but we will see that certain aspects are not well understood and may lead to unexpected results and accompanying challenges.

Factors as (potential) explanatory variables

Factors

A factor, Λ, is a categorical variable (eg. nationality, sex, etc) and for each sample unit takes only one of several, ℓ, categories or levels (eg. "Español/a", "Francés/a", "Argentino/a").

- In many applied problems, factors are considered as possible explanatory variables jointly with perhaps numerical variables.

Main goal

"Repeat" the variable selection exercise but now selecting among certain set of potential covariates and/or factors

$$\{x_1, \ldots, x_k, \Lambda_1, \ldots, \Lambda_p\}$$

In principle the problem is solved using dummies, but we will see that certain aspects are not well understood and may lead to unexpected results and accompanying challenges. These can be better understood in the simplest scenario with only one factor (ℓ levels) and no numerical covariates:

$$\{\Lambda\}$$

Issue 1: All levels or any levels?

The null model M_0 only contains the intercept, but

Question

What does it mean that the factor Λ is a relevant predictor?

Issue 1: All levels or any levels?

The null model M_0 only contains the intercept, but

Question

What does it mean that the factor Λ is a relevant predictor?

- Answer 1. All levels of the factor are important, or

Issue 1: All levels or any levels?

The null model M_0 only contains the intercept, but

Question

What does it mean that the factor Λ is a relevant predictor?

- Answer 1. All levels of the factor are important, or
- Answer 2. any level of the factor is important.

Issue 1: All levels or any levels?

The null model M_0 only contains the intercept, but

Question

What does it mean that the factor Λ is a relevant predictor?

- Answer 1. All levels of the factor are important, or
- Answer 2. any level of the factor is important.

Answer 1 implies comparing only two models:

$$M_0, \; M_1 : y_{ij} = \alpha + a_j + \epsilon_{ij}, \, j = 1, \ldots, \ell,$$

but has two severe drawbacks:

- If M_1 is accepted, we do not know which levels are relevant

Issue 1: All levels or any levels?

The null model M_0 only contains the intercept, but

Question

What does it mean that the factor Λ is a relevant predictor?

- Answer 1. All levels of the factor are important, or
- Answer 2. any level of the factor is important.

Answer 1 implies comparing only two models:

$$M_0,\ M_1 : y_{ij} = \alpha + a_j + \epsilon_{ij},\ j = 1, \ldots, \ell,$$

but has two severe drawbacks:

- If M_1 is accepted, we do not know which levels are relevant
- M_1 is highly penalized due to its complexity (particularly if $\ell >>$).

Issue 1: All levels or any levels?

The null model M_0 only contains the intercept, but

Question

What does it mean that the factor Λ is a relevant predictor?

- Answer 1. All levels of the factor are important, or
- Answer 2. any level of the factor is important.

Answer 1 implies comparing only two models:

$$M_0, \; M_1 : y_{ij} = \alpha + a_j + \epsilon_{ij}, \, j = 1, \ldots, \ell,$$

but has two severe drawbacks:

- If M_1 is accepted, we do not know which levels are relevant
- M_1 is highly penalized due to its complexity (particularly if $\ell >>$).

We prefer (and use in what follows) Answer 2, implying that there are 2^ℓ competing models:

$$M_0, \; M_\gamma : y_{ij} = \alpha + a_j \gamma_j + \epsilon_{ij}, \; \gamma \in \{0, 1\}^\ell,$$

we will see other advantages of this approach later.

Issue 2: Prior model probabilities

Question

How to specify $Pr(M_\gamma)$?

Issue 2: Prior model probabilities

Question

How to specify $Pr(M_\gamma)$?

The standard approaches imply that prior probability of M_0 is largely affected by ℓ. This effect is very severe for constant prior:

	$Pr(H_0)$	
	$\ell = 3$	$\ell = 7$
Constant	1/8	1/128
Scott-Berger	1/4	1/8

Issue 2: Prior model probabilities

Question

How to specify $Pr(M_\gamma)$?

The standard approaches imply that prior probability of M_0 is largely affected by ℓ. This effect is very severe for constant prior:

	$Pr(H_0)$	
	$\ell = 3$	$\ell = 7$
Constant	1/8	1/128
Scott-Berger	1/4	1/8

Our proposal is a two-stage (hierarchical) specification: $Pr(H_0) = Pr(\Lambda) = 1/2$ and then

$$Pr(M_\gamma \mid \Lambda) = \text{Constant},$$

Issue 2: Prior model probabilities

Question

How to specify $Pr(M_\gamma)$?

The standard approaches imply that prior probability of M_0 is largely affected by ℓ. This effect is very severe for constant prior:

	$Pr(H_0)$ $\ell = 3$	$\ell = 7$
Constant	1/8	1/128
Scott-Berger	1/4	1/8

Our proposal is a two-stage (hierarchical) specification: $Pr(H_0) = Pr(\Lambda) = 1/2$ and then

$$Pr(M_\gamma \mid \Lambda) = \text{Constant},$$

or (better) the Scott-Berger in this second stage:

$$Pr(M_\gamma \mid \Lambda) = \frac{1}{\ell \binom{\ell}{k_\gamma}} .$$

Which automatically controls for the multiplicity issue that arises due to the ℓ dummy variables used, a potential pitfall already observed by Chipman (1996).

Issue 3: Reparameterizations?

Perhaps someone has realized that the full model is rank deficient (many more models with more than one factor).

Issue 3: Reparameterizations?

Perhaps someone has realized that the full model is rank deficient (many more models with more than one factor).This is quite uncomfortable but also the conventional priors cannot be used.

Idea!

Issue 3: Reparameterizations?

Perhaps someone has realized that the full model is rank deficient (many more models with more than one factor).This is quite uncomfortable but also the conventional priors cannot be used.

Idea!

Reparametrize the problem from a full rank (eg. choosing one level as the baseline) expression of the full model.

We could have the illusory perception that parameterizations do not have either any impact in testing composed hypotheses. But this turned out to be quite wrong:

Issue 3: Reparameterizations?

Perhaps someone has realized that the full model is rank deficient (many more models with more than one factor).This is quite uncomfortable but also the conventional priors cannot be used.

Idea!

Reparametrize the problem from a full rank (eg. choosing one level as the baseline) expression of the full model.

We could have the illusory perception that parameterizations do not have either any impact in testing composed hypotheses. But this turned out to be quite wrong:

Real example 1

y is body mass index of $n = 1002$ obese children aged 3-11 (Zurriaga et al, 2011).

- Potential factor: Sports (coded 0 to 5).

Issue 3: Reparameterizations?

Perhaps someone has realized that the full model is rank deficient (many more models with more than one factor).This is quite uncomfortable but also the conventional priors cannot be used.

Idea!

Reparametrize the problem from a full rank (eg. choosing one level as the baseline) expression of the full model.

We could have the illusory perception that parameterizations do not have either any impact in testing composed hypotheses. But this turned out to be quite wrong:

Real example 1

y is body mass index of $n = 1002$ obese children aged 3-11 (Zurriaga et al, 2011).

- Potential factor: Sports (coded 0 to 5).

$Pr(H_0 \mid \mathbf{y})$	
$base = 0$	$base = 1$

Issue 3: Reparameterizations?

Perhaps someone has realized that the full model is rank deficient (many more models with more than one factor).This is quite uncomfortable but also the conventional priors cannot be used.

Idea!

Reparametrize the problem from a full rank (eg. choosing one level as the baseline) expression of the full model.

We could have the illusory perception that parameterizations do not have either any impact in testing composed hypotheses. But this turned out to be quite wrong:

Real example 1

y is body mass index of $n = 1002$ obese children aged 3-11 (Zurriaga et al, 2011).

- Potential factor: Sports (coded 0 to 5).

$Pr(H_0 \mid \mathbf{y})$	
$base = 0$	$base = 1$
0.440	0.002

Issue 3: Reparameterizations? (cont')

- Since parameterizations are so influential, the safest alternative is not doing any!

Issue 3: Reparameterizations? (cont')

- Since parameterizations are so influential, the safest alternative is not doing any!

How to use the conventional priors?

For rank-deficient models M_γ use a particular family of $(\boldsymbol{X}_\gamma^t \boldsymbol{X}_\gamma)^-$ which is regular.

Issue 3: Reparameterizations? (cont')

- Since parameterizations are so influential, the safest alternative is not doing any!

How to use the conventional priors?

For rank-deficient models M_γ use a particular family of $(\boldsymbol{X}_\gamma^t \boldsymbol{X}_\gamma)^-$ which is regular. Priors are not unique, but the Bayes factor is:

$$B_\gamma = \mathcal{B}\left(\frac{SSE_\gamma}{SSE_0}, 1, r_\gamma + 1, n\right),$$

where r_γ is the rank of $\boldsymbol{X}_\gamma$ ($r_\gamma < k_\gamma$).

full problem

Recall:

$$\{x_1, \ldots, x_k, \Lambda_1, \ldots, \Lambda_p\}.$$

full problem

Recall:

$$\{x_1, \ldots, x_k, \Lambda_1, \ldots, \Lambda_p\}.$$

Our proposal can be summarized as:

- Do not reparameterize,

full problem

Recall:

$$\{x_1, \dots, x_k, \Lambda_1, \dots, \Lambda_p\}.$$

Our proposal can be summarized as:

- Do not reparameterize,
- use hierarchical-SB prior:

$$P(\{x_{i_1}, \dots, x_{i_{m_1}}, \Lambda_{j_1}, \dots, \Lambda_{j_{m_2}}\}) = \left[(k+p+1)\binom{k+p}{m_1+m_2}\right]^{-1} \quad (1)$$

$$P(M_\gamma \mid \{x_{i_1}, \dots, x_{i_{m_1}}, \Lambda_{j_1}, \dots, \Lambda_{j_{m_2}}\}) = \left[\prod_{h=1}^{m_2} \ell_h \binom{\ell_h}{k_\gamma^h}\right]^{-1}, \quad (2)$$

where, in (2), $m_2 \geq 1$ (otherwise, it is equal to one), and $1 \leq k_\gamma^h \leq \ell_h$ is the number of levels of factor Λ_h active in M_γ.

Application: childhood Obesity

Real example

y is body mass index of $n = 1002$ obese children aged 3-11 (Zurriaga et al, 2011).

- 4 Fixed covariates: Intercept, WeightBorn, HeightBorn and Age;
- 2 Potential covariates: HrsScrDay and HrsSleep;
- 2 Potential factors: Sports (coded 0 to 5) and HealthyFood (0-2). In both cases smaller codes correspond to negative habits.

Application: childhood Obesity

Real example

y is body mass index of $n = 1002$ obese children aged 3-11 (Zurriaga et al, 2011).

- 4 Fixed covariates: Intercept, WeightBorn, HeightBorn and Age;
- 2 Potential covariates: HrsScrDay and HrsSleep;
- 2 Potential factors: Sports (coded 0 to 5) and HealthyFood (0-2). In both cases smaller codes correspond to negative habits.

Sports($\ell = 6$)	HealthyFood($\ell = 3$)	HrsScrDay	HrsSleep
0.995	0.998	0.999	0.622

Table: Inclusion probabilities of factors and covariates.

Application: childhood Obesity

Real example

y is body mass index of $n = 1002$ obese children aged 3-11 (Zurriaga et al, 2011).

- 4 Fixed covariates: Intercept, WeightBorn, HeightBorn and Age;
- 2 Potential covariates: HrsScrDay and HrsSleep;
- 2 Potential factors: Sports (coded 0 to 5) and HealthyFood (0-2). In both cases smaller codes correspond to negative habits.

Sports($\ell = 6$)	HealthyFood($\ell = 3$)	HrsScrDay	HrsSleep
0.995	0.998	0.999	0.622

Table: Inclusion probabilities of factors and covariates.

Sports						HealthyFood		
0	1	2	3	4	5	0	1	2
0.99	0.08	0.25	0.09	0.14	0.09	0.82	0.76	0.78

Table: Inclusion probabilities of levels of factors.

Thanks!