Bayesian Nonparametric Modeling of Suicide Attempts

Francisco J. R. Ruiz, Isabel Valera, Carlos Blanco, Fernando Perez-Cruz

[3]

Outline

Outline

Introduction

- Every year, more than 34,000 suicides occur and over 370,000 individuals are treated for self-inflicted injuries in emergency rooms in the U.S.
- First step to suicide prevention: Identification of the factors that increase the risk of attempting suicide [1].
- Challenging and complex task.
- We apply Bayesian nonparametric tools to find out the latent features that increase the risk of attempting suicide.

Goals

- Find out and analyze the latent causes of suicides attempts.
- NESARC database (*National Epidemiologic Survey on Alcohol and Related Conditions*):
  - Samples the U.S. population.
  - Nearly 3,000 questions regarding the way of life, medical conditions, depression and other mental disorders.
  - Mainly yes-or-no questions, and some multiple-choice and questions with ordinal answers.

Outline

Indian Buffet Process [2]

- The IBP places a prior distribution over binary matrices where the number of columns (features) $K \to \infty$.
- Matrix $\mathbf{Z}_{N \times K} \sim \mathrm{IBP}(\alpha)$ (α: concentration parameter).
- Each element $z_{nk} \in \{0, 1\}$ indicates whether the k^{th} feature contributes to the n^{th} data point.
- For a finite number of data points N, the number of non-zero columns K_+ is finite.

Culinary Metaphor

Generative Process:

- N customers enter sequentially a restaurant with an infinitely long buffet of dishes.
- The first customer takes a serving from a $\text{Poisson}(\alpha)$ number of dishes.
- The i-th customer:
  1. Moves along the buffet, sampling dishes in proportion to their popularity, i.e., with probability proportional to $m_k = \sum_{j<i} z_{jk}$.
  2. Having reached the end of all previous sampled dishes, the i-th customer then tries a $\text{Poisson}(\frac{\alpha}{i})$ number of new dishes.

Inference via Gibbs Sampling

- $N \times D$ observation matrix $\mathbf{X}$, where the n^{th} row contains a D-dimensional observation vector $\mathbf{x}_{n\cdot}$.
- The algorithm iteratively samples the value of each element z_{nk}:

$$p(z_{nk} = 1|\mathbf{X}, \mathbf{Z}_{\neg nk}) \propto p(\mathbf{X}|\mathbf{Z})p(z_{nk} = 1|\mathbf{Z}_{\neg nk}).$$

  Exchangeability property of the IBP:

$$p(z_{nk} = 1|\mathbf{Z}_{\neg nk}) = \frac{m_{-n,k}}{N},$$

  where $m_{-n,k} = \sum_{i \neq n} z_{ik}$.
- Then, the number of new columns is drawn from a distribution where the prior is $\mathrm{Poisson}(\frac{\alpha}{N})$.

Outline

Observation Model (1/2)

- Discrete observations: $x_{nd} \in \{1, \ldots, R\}$ ({'Yes', 'No', 'Blank', 'Unkown', ...} in the NESARC database).
- Multiple-logistic function:

$$p(x_{nd} = \text{'yes'} | \mathbf{z}_{n\cdot}, \mathbf{B}^d) \propto \exp\left(\mathbf{z}_{n\cdot} \mathbf{b}^d_{\cdot \text{yes}}\right),$$
$$p(x_{nd} = \text{'no'} | \mathbf{z}_{n\cdot}, \mathbf{B}^d) \propto \exp\left(\mathbf{z}_{n\cdot} \mathbf{b}^d_{\cdot \text{no}}\right),$$
$$\ldots$$

$$p(x_{nd} = r | \mathbf{z}_{n\cdot}, \mathbf{B}^d) = \frac{\exp\left(\mathbf{z}_{n\cdot} \mathbf{b}^d_{\cdot r}\right)}{\sum_{r'=1}^{R} \exp\left(\mathbf{z}_{n\cdot} \mathbf{b}^d_{\cdot r'}\right)}, \quad r = 1, \ldots, R,$$

where $\mathbf{b}^d_{\cdot r} \sim \mathcal{N}(0, \mathbf{\Sigma}_b = \sigma^2_B \mathbf{I})$ is the $r-$th column of matrix $\mathbf{B}^d_{K \times R}$.

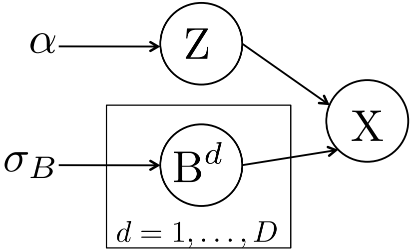
- The matrices $\mathbf{B}^d$ weight differently the contribution of every latent feature for every component d (the probability of every response to a question).

Observation Model (2/2)

Given the latent feature matrix $\mathbf{Z}$ and the weighting matrices $\mathbf{B}^d$, the elements in $\mathbf{X}$ are independent:

$$p(\mathbf{X}|\mathbf{Z}, \mathbf{B}^1, \ldots, \mathbf{B}^D) = \prod_{n=1}^{N} \prod_{d=1}^{D} p(x_{nd}|\mathbf{z}_{n\cdot}, \mathbf{B}^d).$$

Graphical model:



Outline

Inference

- To apply the *Gibbs sampling* algorithm, we need to integrate out $\mathbf{B}^d$, for which an approximation is required.
- The posterior of $\mathbf{B}^1, \ldots, \mathbf{B}^D$ factorizes as

$$\overbrace{p(\mathbf{B}^1, \ldots, \mathbf{B}^D | \mathbf{X}, \mathbf{Z})}^{\text{Non Gauss}} = \prod_{d=1}^{D} p(\mathbf{B}^d | \mathbf{x}_{\cdot d}, \mathbf{Z}) = \prod_{d=1}^{D} \frac{\overbrace{p(\mathbf{x}_{\cdot d} | \mathbf{B}^d, \mathbf{Z})}^{\text{Non Gauss}} \overbrace{p(\mathbf{B}^d)}^{\text{Gauss}}}{p(\mathbf{x}_{\cdot d} | \mathbf{Z})}.$$

- We define the function:

$$\psi(\boldsymbol{\beta}^d) = \log p(\mathbf{x}_{\cdot d} | \boldsymbol{\beta}^d, \mathbf{Z}) + \log p(\boldsymbol{\beta}^d),$$

  where $\boldsymbol{\beta}^d = \mathbf{B}^d(:)$.
- The likelihood term we want to compute is

$$p(\mathbf{x}_{\cdot d} | \mathbf{Z}) = \int \exp\left(\psi(\boldsymbol{\beta}^d)\right) d\boldsymbol{\beta}^d.$$

Laplace Approximation [4]

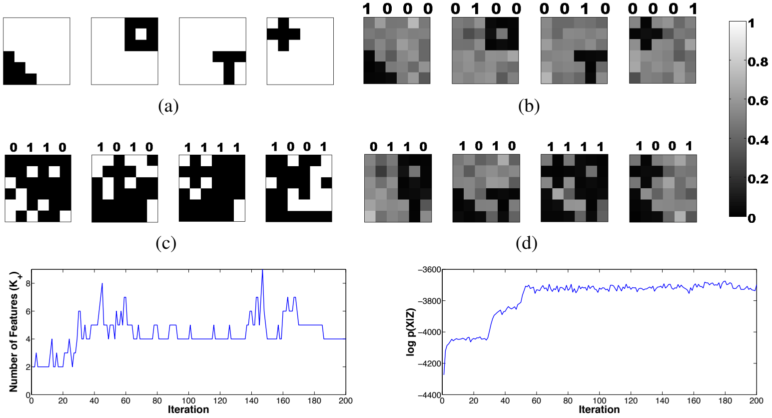
- Our objective is to approximate the integral

$$p(\mathbf{x}_{\cdot d}|\mathbf{Z}) = \int \exp\left(\psi(\boldsymbol{\beta}^d)\right) d\boldsymbol{\beta}^d.$$

- Approximate $\psi(\boldsymbol{\beta}^d)$ by its second-order Taylor series expansion $\equiv$ Approximate $p(\boldsymbol{\beta}^d|\mathbf{X},\mathbf{Z})$ by a Gaussian distribution.
- $\psi(\boldsymbol{\beta}^d)$ is a strictly concave function of $\boldsymbol{\beta}^d$ and therefore it has a unique maximum, which coincides with the maximum of $p(\boldsymbol{\beta}^d|\mathbf{X},\mathbf{Z})$.
- Then, $p(\boldsymbol{\beta}^d|\mathbf{X},\mathbf{Z}) = \mathcal{N}(\boldsymbol{\beta}^d_{MAP}, -\nabla\nabla\psi|_{\boldsymbol{\beta}^d_{MAP}})$.

Outline

Toy Example



(a) (b) (c) (d) (e) (f)

NESARC database (1/3)

- The *National Epidemiologic Survey on Alcohol and Related Conditions* was designed to determine the magnitude of alcohol use disorders.
- $43,093$ subjects representing the U.S. population.
- $2,991$ questions about alcohol and other drug consumption and abuse, medicine use, medical treatment, mental disorders, phobias, family history, etc.
- One question about having attempted suicide.
- We select the 20 questions that present the highest mutual information with the 'suicide attempt' question.
- **Goal:** Search for the latent variables related to the suicide attempt risk.

NESARC database (2/3)

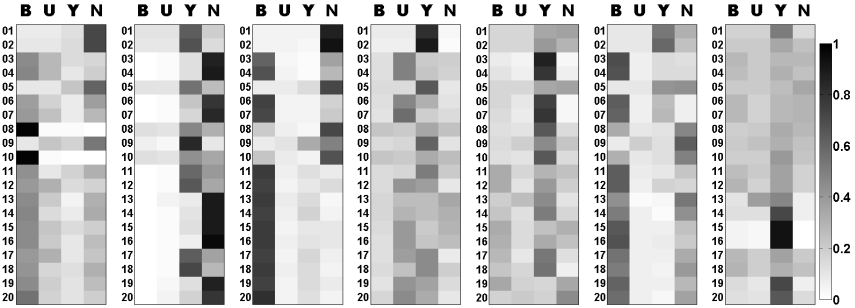


Figure: Probability of answering 'blank' (B), 'unknown' (U), 'yes' (Y) and 'no' (N) to each of the 20 selected questions, when only one of the seven latent features is active.

NESARC database (3/3)

- Suicide attempt probability in the whole database: $\sim 8\%$.

Latent features							Suicide attempt probability		Number of cases	
							Train	Test	Train	Test
1	-	-	-	-	-	-	6.74%	5.55%	430	8072
-	1	-	-	-	-	-	10.56%	11.16%	322	6083
-	-	1	-	-	-	-	**3.72%**	**4.60%**	457	8632
-	-	-	1	-	-	-	**25.23%**	**22.25%**	111	2355
-	-	-	-	1	-	-	8.64%	9.69%	301	5782
-	-	-	-	-	1	-	6.90%	7.18%	464	8928
-	-	-	-	-	-	1	14.29%	14.18%	91	1664
-	-	0	0	-	-	-	30.77%	28.55%	26	571
-	-	0	1	-	-	-	**82.35%**	**61.95%**	17	297
-	-	1	0	-	-	-	**0.83%**	**0.87%**	363	6574
-	-	1	1	-	-	-	14.89%	16.52%	94	2058
-	-	0	1	-	-	1	100.00%	69.41%	4	85
0	-	0	1	-	-	-	80.00%	66.10%	5	118
1	-	1	0	-	1	0	0.00%	0.25%	252	4739
-	-	1	0	-	-	0	0.33%	0.63%	299	5543
1	-	1	0	-	-	-	**0.32%**	**0.41%**	317	**5807**

Outline

Conclusions and Future Work

Conclusions

- We have developed a likelihood model for the IBP that allows dealing with discrete observations.
- For that purpose, we need to resort to Laplace approximation.
- We found that 7 latent features can model the factors that increase the risk of attempting suicide.

Future work

- Adding a constant term to the likelihood function, i.e., $p(x_{nd} = r | \mathbf{z}_{n\cdot}, \mathbf{B}^d) \propto \exp\left(\mathbf{z}_{n\cdot}\mathbf{b}^d_{\cdot r} + b^d_{0r}\right)$.
- An inference algorithm that allows working with more data points, e.g., variational inference or sequential Monte Carlo.
- Prior that enforces sparsity in the weighting matrices $\mathbf{B}^d$.

Outline

Summary of national strategy for suicide prevention: Goals and objectives for action, 2007.
Available at: http://www.sprc.org/library/nssp.pdf.

T. L. Griffiths and Z. Ghahramani.
The Indian Buffet Process: An introduction and review.
Journal of Machine Learning Research, 12:1185–1224, 2011.

F. J. R. Ruiz, I. Valera, C. Blanco, and F. Perez-Cruz.
Bayesian nonparametric modeling of suicide attempts.
Advances in Neural Information Processing Systems (NIPS), 2012.

C. K. I. Williams and D. Barber.
Bayesian classification with Gaussian Processes.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 20:1342–1351, 1998.