

Bayesian modeling with B-Course

Introduction

This learning material is about Bayesian modeling (see e.g., Bayes, 1763; Gill, 2002; Lindley, 1971). Specifically, we practice discrete non-linear modeling with nominal data. The software we use in this exercise is called B-Course. It is programmed by a well-known Finnish statistical computing team, Complex Systems Computation Group. Bayesian modeling has been applied to educational science research work for quite a long time (e.g., Nokelainen, Ruohotie, & Tirri, 1999; Nokelainen, Miettinen, Kurhila, Silander, & Tirri, 2002).

This learning material has two parts:

- First, we conduct Bayesian Classification Modeling (BCM) analysis to learn how predictor variables are used to construct a classification model. As a theoretical background, we use Howard Gardner's (1983) MI theory and Goleman's (1998) Emotional Intelligence framework.
- Second, we examine how Bayesian Dependency Modeling (BDM) constructs Bayesian Networks (BN) and how the results of an analysis are interpreted. We use forementioned MI theory and also Ruohotie's motivational part of the Abilities for Professional Learning model (Ruohotie & Nokelainen, 2000).

Introduction.....	1
Sample data.....	2
Table 1. MI_bayes.dat data variables.....	3
Table 2. APLQ_bayes.dat data variables.....	4
Table 3. Theoretical Structure of the Motivational Scale of APLQ.....	5
Table 4. LCCQ_bayes.dat data description.....	5
Getting started with B-Course.....	6
Bayesian classification modeling.....	7
Exercise 1.....	7
Exercise 2.....	9
Exercise 3.....	9
Exercise 4.....	10
Bayesian dependency modeling.....	10
Exercise 5.....	10
Exercise 6.....	13
Exercise 7.....	14
References.....	15

Sample data

MI_bayes.dat "Multiple Intelligences Questionnaire"

(see e.g., Tirri, K., Komulainen, Nokelainen, & Tirri, H., 2002)

The sample is collected in January 2001 with an online web form from Finnish Open University students. 269 respondents represent five different universities: Helsinki, Joensuu, Tampere, Oulu and Kuopio. The data includes 134 males and 132 females. See Table 1 for data description.

APLQ_bayes.dat "Abilities for Professional Learning Questionnaire"

(see e.g., Nokelainen & Ruohotie, 2002)

The sample consists of 459 second and third year students of a Finnish vocational high school. The sample was collected in March 2000. It represents 32 per cent of the total population ($N = 1436$) and contains 323 (70%) females and 134 (29%) males. See Tables 2 and 3 for data description.

LCCQ_bayes.dat "Leadership Competencies and Characteristics Questionnaire"

(see e.g., Nokelainen & Ruohotie, 2005)

The sample consists of 239 adult employees of a Finnish medium size industrial automation company. Respondents' age mean is 34.6 years ($SD = 10.9$). 194 of the respondents are males (81.2%) and 45 are females (18.8%). See Table 4 for data description.

Download all the datasets from:

https://homepages.tuni.fi/petri.nokelainen/mmx9120/4_bayes/

Table 1. MI_bayes.dat data variables

Variable	Description
GENDER	Male, Female
UNIVER	Hel = Helsinki, Joe = Joensuu, Tam = Tampere, Oul = Oulu, Kuo = Kuopio
MOTIVA	What is the level of your study motivation at the moment? (1 = very good, ..., 4 = very bad)
M01LOGI	At school I was good at mathematics, physics or chemistry.
M03INTR	I often think about my own feelings and sentiments and seek reasons for them.
M04LING	Writing is a natural way for me to express myself.
M05SPAT	At school, geometry and various kinds of assignments involving spatial perception were easier for me than solving equations.
M14MUSI	After hearing a tune once or twice I am able to sing or whistle it quite accurately.
M15MUSI	When listening to music, I am able to discern instruments or recognize melodies.
M16INTR	I am able to analyze my own motives and ways of action.
M17INTR	I spend time regularly reflecting on the important issues in life.
M22INTE	Even in strange company, I easily find someone to talk to.
M23INTE	I get along easily with different types of people.
M29BODK	I am handy.
M30LOGI	I can work with and solve complex problems.
M32INTE	I make contact easily with other people.
M33BODK	I can easily do something concrete with my hands (e.g.
M39LOGI	I am good at games and problem solving, which require logical thinking.
M40LING	I have recently written something that I am especially proud of, or for which I have received
M42INTR	I like to read psychological or philosophical literature to increase my self-knowledge.
M45BODK	I am good at showing how to do something in practice.
M48SPAT	It is easy for me to conceptualize complex and multidimensional patterns.
M53SPAT	I can easily imagine how a landscape looks from a bird's-eye view.
M54LOGI	Mental arithmetic is easy for me.
M55MUSI	I can easily keep the rhythm when drumming a melody.
M56LING	Metaphors and vivid verbal expressions help me learn efficiently.
M59INTE	In negotiations and group work, I am able to support the group to find a consensus.
M62MUSI	I notice immediately if a melody is out of tune.
M65SPAT	When I read, I form illustrative pictures or designs in my mind.
M67BODK	I was good at handicrafts at school.
M70LING	At school studies in native language or social studies were easier for me than mathematics, physics and chemistry.
LING	Sum of items M04,M40,M56,M70
LOGI	Sum of items M01,M30,M39,M54
MUS	Sum of items M14,M15,M55,M62
SPAT	Sum of items M05,M48,M53,M65
BODK	Sum of items M29,M33,M45,M67
INTE	Sum of items M22,M23,M32,M59
INTR	Sum of items M03,M16,M17,M42

NOTE. M01LOGI, ... , M70LING scale is Likert from 1 (totally disagree) to 7 (totally agree)

Table 2. APLQ_bayes.dat data variables

Variable	Description
A1	I prefer to study theoretical subjects and work tasks which are demanding and from which I can learn
A2	I am able to learn even the most difficult subjects if I use good study methods.
A3	During an exam I wonder how I am performing in comparison to other students.
A4	In my opinion craftsmanship can be acquired through practice without theory lessons.
A5	I believe that theory lessons will prove useful in practical training and later in working life.
A6	I expect to get excellent grades in my vocational/occupational studies.
A7	I am confident that I understand even the most difficult aspects of my studies.
A8	I want to receive as high grades as possible.
A9	When answering essay questions, I am also concerned about the other questions on the test that I
A10	It is my own fault if I fail to learn theory.
A11	It is important for me to learn theory related to my profession.
A12	I am confident that I will learn the skills related to my field.
A13	I am only interested in mastering learning tasks that are required in real working life.
A14	Nervousness during exams affects my performance.
A15	Studying feels often burdensome and/or frustrating.
A16	When taking part in a practical examination I am concerned about failing and what will happen as a
A17	I am confident that I will learn even the most difficult theoretical subjects and work tasks.
A18	I prefer to study theoretical subjects that interest me even if I find them difficult.
A19	I am very interested in my field of study as well as in the new information related to it.
A20	I will acquire the required professional skills if I work hard enough.
A21	I am really nervous in all test situations.
A22	I am confident that I will do well in my studies.
A23	I find it most rewarding when I can research a subject as thoroughly as possible.
A24	I believe that my vocational/occupational benefits will be of practical benefit to me.
A25	When given an opportunity I choose exercises and literature from which I can learn something new
A26	If I do not understand theory, it is because I am not trying hard enough.
A27	It is important for me to do well in my studies and show others (my family, friends, colleagues) what
A28	It is essential for me to understand the topics contained in my vocational/occupational studies.

NOTE. A1, ... , A28 scale is Likert from 1 (totally disagree) to 5 (totally agree)

Table 3. Theoretical Structure of the Motivational Scale of APLQ

Motivational Scale (A)	Factors	Items (28) in APLQ
<i>1. Value Components</i>		
1.1 Intrinsic Goal Orientation	MF1	A1, A18, A23, A25
1.2 Extrinsic Goal Orientation	MF2	A8, A13, A27
1.3 Meaningfulness of Study	MF3	A4 (rev.), A5, A11, A15, A19, A24, A28
<i>2. Expectancy Components</i>		
2.1 Control Beliefs	MF4	A2, A10, A20, A26
2.2 Self-Efficacy	MF5	A6, A7, A12, A17, A22
<i>3. Affective Components</i>		
3.1 Test Anxiety	MF6	A3, A9, A14, A16, A21

Table 4. LCCQ_bayes.dat data description

Variable	Description
GENDER	1 = Female, 2 = Male
AGE	1 = -> 29, 2 = 30 – 39, 3 = 40 – 49, 4 = 50 ->
EI_4_1	Inspirational Leadership
EI_4_2	Influence
EI_4_3	Developing Others
EI_4_4	Change Catalyst
EI_4_5	Conflict Management
EI_4_6	Teamwork and Collaboration

NOTE. EI_4_1, ... , EI_4_6 scale is Likert from 1 (totally disagree) to 5 (totally agree)

Getting started with B-Course

1. Install Docker according to instructions in <https://www.docker.com>.
2. Start Terminal and type:
`docker run --rm -d -p 80:80 --name bcourse tomsilander/bcourse`
3. Start web browser and type this address:
<http://localhost/>

Now B-Course should be ready for use!

4. To end (kill) the B-Course session type in Terminal:
`docker kill bcourse`

Bayesian classification modeling

Exercise 1

1. Choose "Classification" and "The real trail".
2. Click on "Data upload form" fields "Browse ..." button.
3. Search and select "**MI_bayes.dat**" file.
4. Type "My first classification exercise with MI data" into "Data set name:" field.
5. Click on "Send data" button.
6. Choose "GENDER" as a class variable and click on "Proceed" button.
7. Remove selection mark from the "Include" column of the "M01LOGI", ..., "M70LING" variables. Following variables remain selected for the analysis: "UNIVER", "MOTIVA", "LING", "LOOG", "MUS", "SPAT", "BODKIN", "INTER", "INTRA".
8. Click on "Go beyond the data" button.
9. Click on "Start search" button.

You may also *not* search for the best model by selecting "Do not search". That is usually a good choice when you need to compare classification accuracy of two specific models.

10. Click on "Next report" button. Note that the estimated classification accuracy changes as the variable selection proceeds.
11. Click on "Final report" button.

"Information about the search" tells the estimated classification accuracy, which is usually considered against the base line model (guessing that everybody belongs to the biggest group).

12. Click on "Java playground" link under "Playgrounds" heading.

If the **Java playground** link does not work, you may need to install JAVA into the computer. If that is not possible, you should use the **Vanilla playground**.

13. Click on the "Launch the Bayes Browser" link.
14. Click the variable name of each node (e.g., "GENDER") to see the conditional probability distributions.

15. Examine the behavior of the conditional probability distributions by "fixing" the different values of class variable (in this exercise it is "GENDER"). Try to answer the following questions:

- What university ("UNIVER") represents most male respondents in this sample?

- Which gender has more intrapersonal ("INTR") characteristics?

Exercise 2

1. Choose "Classification".
2. Choose "The real trail".
3. Click on "Data upload form" fields "Browse ..." button.
4. Search and select "MI_bayes.dat" file.
5. Type "My second classification exercise with MI data" into "Data set name:" field.
6. Click on "Send data" button.
7. Choose "MOTIVA" as the classification variable.
8. Remove the selection from "M01LOGI", ... , "M70LING" variables.
9. Carry out classification analysis without variable selection by choosing "Do not search" in the "Model search" window.
10. Answer the following questions:
 - What is the classification accuracy? It is _____ %.
 - Is it OK?
 - Yes / no compared to the
_____.
 - How about comparing it to the "traditional" levels of classification accuracy in educational science research?
13. Examine the model with Bayes Browser (go to "Playgrounds" and choose "Java playground").
 - How would you explain the changes in self-rated motivation related to self-rated logical-mathematical competence?

Exercise 3

1. Conduct a new classification dependency analysis with the same data, but this time choose "UNIVER" as a classification variable and all but "M01LOGI", ... , "M70LING" variables as predictors.
2. Answer the following question:

- Which university's student's rate mathematics as their strength?

- Which university's students rate them as linguistically competent?

Exercise 4

1. Choose "Classification modeling".
2. Choose "The real trail".
3. Click on "Data upload form" fields "Browse ..." button.
4. Search and select "LCCQ_bayes.dat" file.

As the variables in this sample represent a totally different theoretical model, you are encouraged to examine Table 4 at this point (see page 5).

5. Type "My first LCCQ classification analysis" into "Data set name:" field.
6. Click on "Send data" button. "Show some class" page appears.
7. Choose "GENDER" as the classification variable.
8. Choose all but "AGE" to be the predictors.
9. Carry out the classification analysis with variable selection.
10. Answer the following questions:
 - a. Which variables were selected for the model? _____
 - b. What is the classification accuracy of the model?
 - c. Give a free interpretation of the results regarding the classification variable "GENDER":

Bayesian dependency modeling

Exercise 5

1. Select "Dependency modeling".
2. Select "The real trail".
3. Click on "Data upload form" fields "Browse ..." button.
4. Search and select "MI_bayes.dat" file.
5. Write to the "Data set name:" field something like "My first MI dependency analysis".

6. Click on "Send data" button and "Your data as we see it" page will open.
7. Remove "Include" selection from the following variables: "GENDER", "UNIVER", "MOTIVA", "LING", "LOOG", "MUS", "SPAT", "BODKIN", "INTER", "INTRA". What you still have selected are the 28 MI variables ("M01LOGI", ... , "M70LING").
8. Click on "Go beyond the data" button.
9. Click on "Start search" button. The server begins the analysis of your data. You are able to observe how the best model evolves by clicking "Next report" button.
10. Click on "Final report" button when you are ready to see the final model.

It is not wise to Click on "Final report" button right away, as the model search algorithm needs to have some time for successful model search. It is not trivial to figure out how many models are out there, see "How many models are out there?" in "Library" of the B-Course site (<http://b-course.hiit.fi/howmanymodels.html>).

11. "Final report" page presents all the essential results of dependency modeling:
 - "Information about the search" tell you how many models were tested before the chosen one.

Compare this figure to the one that is in the table of "How many models are there" page. For example, with 20 variables you have at least $15,6 \times 10^{56}$ models!

- "Speculating about causalities" reminds you about what is and what is *not* a causal relationship.
12. Click on "Strength of dependencies (arcs)" link under "Dependency model" heading.

The picture shows the model with inter-variate strengths of dependencies. Dark colors indicate stronger dependencies.

13. Analyze the model on the basis of MI theory (refer to Table 1 for dimensional structures).

Here are some points to consider:

- Is the model homogeneous (are all the variables included in the model)?
- Is the model unidimensional according to the MI theory (are the nodes connected)?
- Do the strength of dependencies between variables make sense according to the MI theory?

More detailed information about the strengths of dependencies is available in "Importance ranking of the arcs in the final model" table.

14. Browse back to "Final report" page.
15. Choose "Java playground" link under the "Playgrounds" heading.

If the JAVA application does not appear, you may need to install JAVA support to the computer. If that is not possible, you should use the **Vanilla Playground** link.

- 16. Click the variable name of each node (e.g., "M01LOGI") to see the probability distributions (minimum value is "1" and maximum value is "7").
- 17. Examine how predictive distributions behave by clicking on distribution bars (that is, to simulate situation where the respondents have answered, "3"). That is called "fixing". It is cancelled by clicking the same bar again.

When you click something, it is always passive action. It is like saying "I know that this is true", or "What if everybody have selected this". It is something that has already happened. You are the expert, but you have only passive role at this point, as you are not able to actively intervene to the model. Thus, it is important not to confuse active and passive observation.

- 18. Make your own conclusions about inter-variable dependencies. You may find helpful to take a look at table 1 for variable descriptions. Research article 3 contains further information about the MI theory.
 - Nodes M_____ and M_____ are related in a such way that

 - Nodes M_____ and M_____ are related in a such way that

 - Nodes M_____ and M_____ are related in a such way that

 - Nodes M_____ and M_____ are related in a such way that

Exercise 6

1. Choose "Dependency modeling".
2. Choose "The real trail".
3. Click on "Data upload form" fields "Browse ..." button.
4. Search and select "**MI_bayes.dat**" file.
5. Type "My second MI dependency analysis" into "Data set name:" field.
6. Click on "Send data" button. "Your data as we see it" page appears.
7. Remove selection mark from the "Include" column of the following variables: "GENDER", "UNIVER", "MOTIVA", "M01LOGI" ... "M70LING". Seven summary variables ("LING", ..., "INTR") remain selected.
8. Click on "Go beyond the data" button.
9. Click on "Start search" button and after few minutes click on "Final report" button.
10. Click on "Strength of dependencies (arcs)" link that is situated under the "Dependency model" heading on the "Final report" page.
11. Examine the model on the basis of MI theory (refer to the research article 3):
 - Is the model homogeneous (are all the variables included in the model)?
 - Is the model unidimensional according to the MI theory (are the nodes connected)?
12. Browse back to the "Final report" page.
13. Choose "Java playground" (or "Vanilla playground") link under the "Playgrounds" heading.
14. Click the variable name of each node (e.g., "INTER") to see the probability distributions (minimum value is $1+1+1+1=4$ and maximum value is $7+7+7+7=28$). For further details, see table 2.
15. Examine how the predictive distributions behave by clicking on certain interesting values on the node's distribution bars.
16. Now it is time to play interactively with the model:
 - Examine how the value distributions react when you "fix" certain value or values.

Exercise 7

11. Choose "Dependency modeling".
12. Choose "The real trail".
13. Click on "Data upload form" fields "Browse ..." button.
14. Search and select "APLQ_bayes.dat" file.

As the variables in this sample represent a totally different theoretical model, you are encouraged to examine tables 2 and 3 at this point. Note that the research article 4 uses the same data.

15. Type "My first APLQ dependency analysis" into "Data set name:" field.
16. Click on "Send data" button. "Your data as we see it" page appears.
17. Allow all the variables remain selected and click on "Go beyond the data" button.
18. Try to figure out if the theoretical structure presented in table 3 is valid.

In practice you are interested to see, if the variables are closely situated to each other in a way that the theoretical six-dimensional model suggests.

References

- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53, 370–418.
- Gardner, H. (1983). *Frames of mind*. New York: Basic Books.
- Gill, J. (2002). *Bayesian methods. A Social and Behavioral Sciences Approach*. Boca Raton: Chapman & Hall/CRC.
- Goleman, D. (1998). What makes a leader? *Harvard Business Review*, November-December, 93-102.
- Lindley, D.V. (1971). *Making Decisions*. London: Wiley.
- Nokelainen, P. (2008). Modeling Professional Growth and Learning: Bayesian Approach.
- Nokelainen, P., Miettinen, M., Kurhila, J., Silander, T., & Tirri, H. (2002). Optimizing and profiling users online with Bayesian probabilistic modeling. In *Proceedings of the International Networked Learning Conference of Natural and Artificial Intelligence Systems Organization*, Berlin, Germany. Canada: ICSC-NAISO Academic Press.
- Nokelainen, P., & Ruohotie, P. (2002). Modeling Student's Motivational Profile for Learning in Vocational Higher Education. In H. Niemi & P. Ruohotie (Eds.), *Theoretical Understandings for Learning in the Virtual University* (pp. 177-206). Hämeenlinna, Finland: RCVE.
- Nokelainen, P., & Ruohotie, P. (2003a). Common Components Between Dimensions of Growth-oriented Atmosphere and Empowerment. In *Proceedings of International Research on Work and Learning (RWL) Conference* (pp. 176-185). Tampere, Finland.
- Nokelainen, P., & Ruohotie, P. (2003b). Modeling the Prerequisites of Empowerment. In B. Beairsto, M. Klein & P. Ruohotie (Eds.), *Professional Learning and Leadership* (pp. 147-176). Hämeenlinna, Finland: RCVE.
- Nokelainen, P., & Ruohotie, P. (2005). Investigating the Construct Validity of the Leadership Competence and Characteristics Scale. In *Proceedings of International Research on Work and Learning (RWL) Conference*. Sydney, Australia.
- Nokelainen, P., Ruohotie, P., & Tirri, H. (1999). *Professional Growth Determinants-Comparing Bayesian and Linear Approaches to Classification*. ERIC Document Reproduction Service No.ED442850.
- Nokelainen, P., & Silander, T. (2014). Using New Models to Analyze Complex Regularities of the World: Commentary on Musso et al. (2013). *Frontline Learning Research*, 2(3), 78-82.
- Nokelainen, P., & Tirri, H. (2004). Bayesian Methods that Optimize Cross-cultural Data Analysis. In J. R. Campbell, K. Tirri, P. Ruohotie, & H. Walberg (Eds.), *Cross-cultural Research: Basic Issues, Dilemmas, and Strategies* (pp. 141-158). Hämeenlinna, Finland: RCVE.
- Ruohotie, P., & Nokelainen, P. (2000). Modern Modeling of Student Motivation and Self-Regulated Learning. In P. R. Pintrich & P. Ruohotie (Eds.), *Conative Constructs and Self-Regulated Learning* (pp. 141-193). Hämeenlinna, Finland: RCVE.
- Ruohotie, P., Tirri, H., Nokelainen, P., & Silander, T. (1999). *Modern Modeling of Professional Growth, vol. 1*. Hämeenlinna, Finland: RCVE.
- Tirri, K., Komulainen, E., Nokelainen, P., & Tirri, H. (2002). Conceptual Modeling of Self-Rated Intelligence-Profile. In *Proceedings of the 2nd International Self-Concept Research Conference*. University of Western Sydney, Self Research Center.