Bayesian Thinking

Max Kochurov

Moscow State University

Lecture 1

Agenda

About Me

- Graduated from
  - BS - MSU EF (2018)
  - MS - Skoltech DS (2020)
- Core developer at PyMC
- Principle Data Scientist at PyMC Labs
- BayesGroup alumni
- Experience with
  - Large scale Deep Learning and Computer Vision
  - Differential geometry for Graph Neural Networks
  - Bayesian Methods for Deep Learning
  - Applied Bayesian Statistics for industry (e.g. AB testing, Bio-Informatics)

In this course

You'll learn...

- how to think critically about your model
- tools to check the validity of the results
- how to present your results
- non-parametric models for time series

Grading

The grade consists of

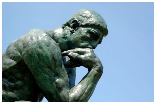
- 60% Homework
- 40% Group project

Grades will be assigned as

- 5 - 85%+
- 4 - 65%+
- 3 - 40%+
- 2 - $<$ 40%

Why learn Bayesian methods?

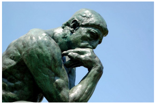
- Used in advanced research
  - CBR - Bayesian DSGE (link)
  - Papers using PyMC from google scholar is overwhelming (link)
- Used in top notch industry
  - Marketing at Indigo (link)
  - Drug development at Roche (link)
  - Portfolio Theory at Quantopian (link)
  - Financial Advisory at Everysk (link)
  - Conducting Surveys at Civiqs (link)
- Growth opportunity
  - Links many disciplines and career transitions
  - Hot non-boring job offerings in industry
  - Opens new research possibilities



$$\underbrace{p(\Theta|\mathcal{D})} = \frac{\overbrace{p(\mathcal{D}|\Theta)}\ \overbrace{p(\Theta)}}{p(\mathcal{D})}$$

$\mathcal{D}$ = Data $\qquad$ Θ = World State

Prior Distribution



$$p(\Theta|\mathcal{D}) = \frac{p(\mathcal{D}|\Theta)\ \overbrace{p(\Theta)}}{p(\mathcal{D})}$$

Case Study: Where do priors come from?

Authors: Marielle Zondervan-Zwijnenburg, Margot Peeters, Sarah Depaoli, Rens van de Schoot [**?**]. Bayesian Econometrics example

Policy question

Should we increase cannabis control for adolescents?

- Drugs long term influence on brain activity after early onset

Case Study: Where do priors come from?

Authors: Marielle Zondervan-Zwijnenburg, Margot Peeters, Sarah Depaoli, Rens van de Schoot [?]. Bayesian Econometrics example

Policy question

Should we increase cannabis control for adolescents?

- Drugs long term influence on brain activity after early onset
- Few to zero relevant prior research
  - no exact match for cannabis case and brain activity
  - additional developing diseases

Case Study: Where do priors come from?

Authors: Marielle Zondervan-Zwijnenburg, Margot Peeters, Sarah Depaoli, Rens van de Schoot [**?**]. Bayesian Econometrics example

Policy question

Should we increase cannabis control for adolescents?

- Drugs long term influence on brain activity after early onset
- Few to zero relevant prior research
  - no exact match for cannabis case and brain activity
  - additional developing diseases
- Scarce, hard to obtain data

Case Study: Where do priors come from?

Authors: Marielle Zondervan-Zwijnenburg, Margot Peeters, Sarah Depaoli, Rens van de Schoot [**?**]. Bayesian Econometrics example

Policy question

Should we increase cannabis control for adolescents?

- Drugs long term influence on brain activity after early onset
- Few to zero relevant prior research
  - no exact match for cannabis case and brain activity
  - additional developing diseases
- Scarce, hard to obtain data
- Classical econometrics fails (16 data points in the group)

Case Study: Where do priors come from?

Authors: Marielle Zondervan-Zwijnenburg, Margot Peeters, Sarah Depaoli, Rens van de Schoot [**?**]. Bayesian Econometrics example

Policy question

Should we increase cannabis control for adolescents?

- Drugs long term influence on brain activity after early onset
- Few to zero relevant prior research
  - no exact match for cannabis case and brain activity
  - additional developing diseases
- Scarce, hard to obtain data
- Classical econometrics fails (16 data points in the group)
- Expert knowledge feels important

Case Study: Where do priors come from?

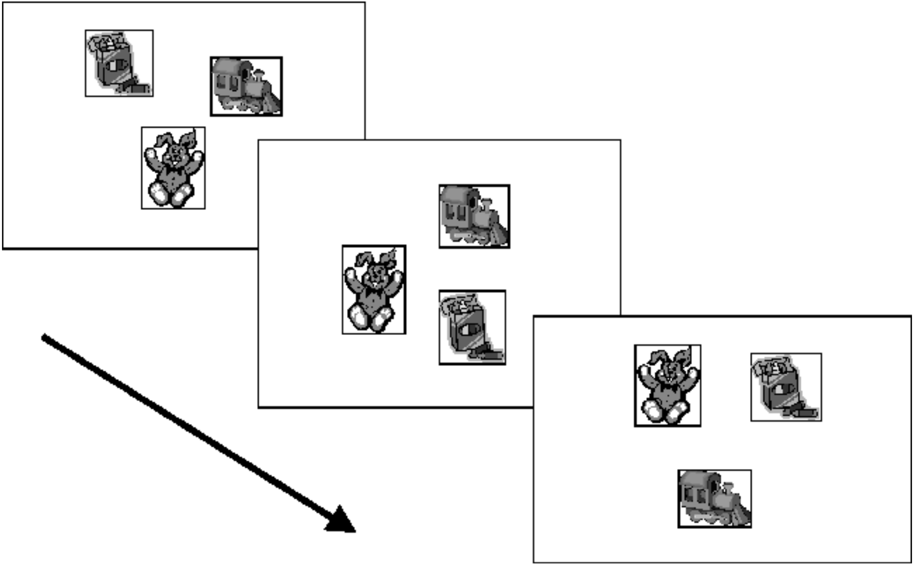
Authors: Marielle Zondervan-Zwijnenburg, Margot Peeters, Sarah Depaoli, Rens van de Schoot [**?**]. Bayesian Econometrics example

Policy question

Should we increase cannabis control for adolescents?

- Drugs long term influence on brain activity after early onset
- Few to zero relevant prior research
  - no exact match for cannabis case and brain activity
  - additional developing diseases
- Scarce, hard to obtain data
- Classical econometrics fails (16 data points in the group)
- Expert knowledge feels important
- Statistics is required for an informed decision

The Game



Quick intro

Measuring the existence and severity of cannabis usage

- You can measure brain development with a Game (Self Ordered Pointing Test [**?**])
- Cannabis use is checked for participants
- Adolescents pass the Game 2 times a year
- The results are compared between the **heavy** and **light** cannabis users

Leftover questions

- How does the amount of cannabis used affects the brain development?
- What age is sufficient to minimize the effect of usage?
- What policy should be used to minimize the effect?

Case Study: Prior Distribution

To develop a prior researchers combined many sources of information

1. Knowledge before seeing any data
2. Prior research results
3. Expert knowledge
4. Constraints
5. Model properties

Case Study: Prior Distribution

To develop a prior researchers combined many sources of information

1. Knowledge before seeing any data
   - Practical range for SOPT measure and growth rates
2. Prior research results

3. Expert knowledge

4. Constraints

5. Model properties

Case Study: Prior Distribution

To develop a prior researchers combined many sources of information

1. Knowledge before seeing any data
   - Practical range for SOPT measure and growth rates
2. Prior research results
   - effect size of cannabis use was mixed with other diseases or had missing information
   - "13 out of 693 articles yielded useful information"
3. Expert knowledge
4. Constraints
5. Model properties

Case Study: Prior Distribution

To develop a prior researchers combined many sources of information

1. Knowledge before seeing any data
   - Practical range for SOPT measure and growth rates
2. Prior research results
   - effect size of cannabis use was mixed with other diseases or had missing information
   - "13 out of 693 articles yielded useful information"
3. Expert knowledge
   - Prior research was reviewed by experts
   - Relation of diseases and behaviour to cannabis use
4. Constraints
5. Model properties



Case Study: Prior Distribution

To develop a prior researchers combined many sources of information

1. Knowledge before seeing any data
    - Practical range for SOPT measure and growth rates
2. Prior research results
    - effect size of cannabis use was mixed with other diseases or had missing information
    - "13 out of 693 articles yielded useful information"
3. Expert knowledge
    - Prior research was reviewed by experts
    - Relation of diseases and behaviour to cannabis use
4. Constraints
    - Slope (SOPT development rate) is more positive than negative
5. Model properties

Case Study: Prior Distribution

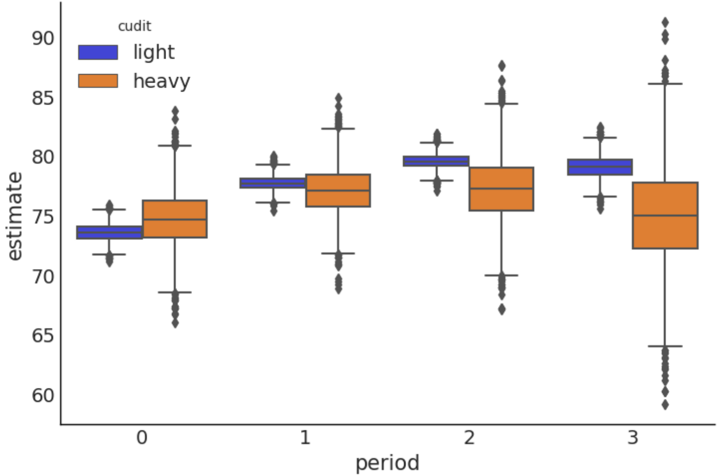
To develop a prior researchers combined many sources of information

1. Knowledge before seeing any data
   - Practical range for SOPT measure and growth rates
2. Prior research results
   - effect size of cannabis use was mixed with other diseases or had missing information
   - "13 out of 693 articles yielded useful information"

3. Expert knowledge
   - Prior research was reviewed by experts
   - Relation of diseases and behaviour to cannabis use
4. Constraints
   - Slope (SOPT development rate) is more positive than negative

5. Model properties
   - Sign of the quadratic term

Results



Common Issues

In the research the prior was defended

1. Prior is subjective
2. Prior specification is unclear
3. Prior is incorrectly specified

Common Issues

In the research the prior was defended

1. Prior is subjective
   - Informed prior was compared to uninformed one
2. Prior specification is unclear
3. Prior is incorrectly specified

Common Issues

In the research the prior was defended

1. Prior is subjective
   - Informed prior was compared to uninformed one
2. Prior specification is unclear
   - The log book was provided with a full description of the choice
   - https://osf.io/aw8fy/
3. Prior is incorrectly specified

Common Issues

In the research the prior was defended

1. Prior is subjective
   - Informed prior was compared to uninformed one
2. Prior specification is unclear
   - The log book was provided with a full description of the choice
   - https://osf.io/aw8fy/
3. Prior is incorrectly specified
   - There are still some issues with the analysis but it is out of the scope

Likelihood Distribution

$$p(\Theta|\mathcal{D}) = \frac{\overbrace{p(\mathcal{D}|\Theta)}^{\text{FACT}}\, p(\Theta)}{p(\mathcal{D})}$$

You sell nuts. You want sell more nuts! How to increase sales?

- increase purchase probability

- increase order size

Case Study: AB Test

You sell nuts. You want sell more nuts! How to increase sales?

- increase purchase probability
  - Create a banner about healthy food
  - Add a banner with recipes
  - Improve the layout
- increase order size

ТОВ "АТБ-Маркет"
МАГАЗИН "ПРОДУКТИ-556"
м.Чернігів,вул.50 років СРСР,5
АТБ - ЦІНИ РАДУЮТЬ !
email: in@atbmarket.com
ПН 304872104175
00005 Коса Я. І. 02
Касса 5 #
Чек 861621, Z 941 #
Пакет п/е 40х60 зі
святковою сим 1,45 А
- - - - - - - - - - - - - - - - - - - -
СУМА 1.45
- - - - - - - - - - - - - - - - - - - -
ПДВ А 20,00% 0,24
СУМА ПОДАТКУ 0,24
- - - - - - - - - - - - - - - - - - - -
ГОТІВКА 2,00
РЕШТА 0,55
ДЯКУЄМО ЗА ПОКУПКУ!
ГАРЯЧА ЛІНІЯ 0 800 500 415
1 АРТИКУЛ
0242196 0590204 17-01-2016 14:19:28
ЗН ИП00001405 ФН 2801001754
ФІСКАЛЬНИЙ ЧЕК
Екселліо

Case Study: AB Test

You sell nuts. You want sell more nuts! How to increase sales?

- increase purchase probability
  - Create a banner about healthy food
  - Add a banner with recipes
  - Improve the layout
- increase order size
  - Lower the price
  - Increase quality
  - Make better packaging

ТОВ "АТБ-Маркет"
МАГАЗИН "ПРОДУКТИ-556"
м.Чернігів,вул.50 років СРСР,5
АТБ - ЦІНИ РАДУЮТЬ !
email: in@atbmarket.com
ПН 304872104175
00005 Коса Я. І. 02
Касса 5 #
Чек 861621, Z 941 #
Пакет п/е 40x60 зі
святковою сим 1,45 А
СУМА 1.45
ПДВ А 20,00% 0,24
СУМА ПОДАТКУ 0,24
ГОТІВКА 2,00
РЕШТА 0,55
ДЯКУЄМО ЗА ПОКУПКУ!
ГАРЯЧА ЛІНІЯ 0 800 500 415
1 АРТИКУЛ
0242196 0590204 17-01-2016 14:19:28
ЗН ИП00001405 ФН 2801001754
ФІСКАЛЬНИЙ ЧЕК
Екселліо

Case Study: AB Test

You sell nuts. You want sell more nuts! How to increase sales?

- increase purchase probability
  - Create a banner about healthy food
  - Add a banner with recipes
  - Improve the layout
- increase order size
  - Lower the price
  - Increase quality
  - Make better packaging

What is better?

AB testing can answer the question

ТОВ "АТБ-Маркет"
МАГАЗИН "ПРОДУКТИ-556"
м.Чернігів,вул.50 років СРСР,5
АТБ - ЦІНИ РАДУЮТЬ !
email: in@atbmarket.com
ПН 304872104175
00005 Коса Я. І. 02
Касса 5 #
Чек 861621, Z 941 #
Пакет п/е 40x60 зі
святковою сим 1,45 А
СУМА 1.45
ПДВ А 20,00% 0,24
СУМА ПОДАТКУ 0,24
ГОТІВКА 2,00
РЕШТА 0,55
ДЯКУЄМО ЗА ПОКУПКУ!
ГАРЯЧА ЛІНІЯ 0 800 500 415
1 АРТИКУЛ
0242196 0590204 17-01-2016 14:19:28
ЗН ИП00001405 ФН 2801001754
ФІСКАЛЬНИЙ ЧЕК
Екселліо

Not all customers buy nuts

- A significant portion of data are just zeros
- A classical 2 sample t-test assumes normality, not our case
- Researchers admit t-test weaknesses in these cases [?]

Not all customers buy nuts

- A significant portion of data are just zeros
- A classical 2 sample t-test assumes normality, not our case
- Researchers admit t-test weaknesses in these cases [?]

Solution

Think of a non-normal likelihood

Zero inflation

Zero Inflation

Data property, when a significant portion of data is exactly zero

Examples:

- Wait times in a queue (no queue is zero)
- Defects on a production line (no defects is zero)
- Rain water level (sunny weather is 0 water level)
- Purchase order statement (no nuts is zero)

Zero Inflated Distribution

Wisdom

Any distribution can be made zero inflated

Example: Zero Inflated Gamma.
We'll have parameters α, β (from Gamma) and p - probability of non-zero
Sampling:

$$z \sim \text{Bernoulli}(p)$$
$$sample \sim \begin{cases} \text{Gamma}(\alpha, \beta) & , z = 1 \\ 0 & , z = 0 \end{cases}$$

Zero Inflated Distribution

Wisdom

Any distribution can be made zero inflated

Example: Zero Inflated Gamma.
We'll have parameters α, β (from Gamma) and p - probability of non-zero
Sampling:

$$z \sim \text{Bernoulli}(p)$$

$$sample \sim \begin{cases} \text{Gamma}(\alpha, \beta) & , z = 1 \\ 0 & , z = 0 \end{cases}$$

Log Probability Density Function

$$\log p(x \mid p, \alpha, \beta) = \begin{cases} \log(1 - p) & , x = 0 \\ \log(p) + \frac{x^{\alpha-1} e^{-\beta x} \beta^{\alpha}}{\Gamma(\alpha)} & , x > 0 \end{cases}$$

Zero Inflated Distribution

Wisdom

Any distribution can be made zero inflated

Example: Zero Inflated Gamma.
We'll have parameters α, β (from Gamma) and p - probability of non-zero
Sampling:

$$z \sim \text{Bernoulli}(p)$$

$$sample \sim \begin{cases} \text{Gamma}(\alpha, \beta) & , z = 1 \\ 0 & , z = 0 \end{cases}$$

Log Probability Density Function

$$\log p(x \mid p, \alpha, \beta) = \begin{cases} \log(1 - p) & , x = 0 \\ \log(p) + \frac{x^{\alpha-1} e^{-\beta x} \beta^{\alpha}}{\Gamma(\alpha)} & , x > 0 \end{cases}$$

Zero inflation as a mixture

Wisdom

Zero inflation is a special case of a Mixture

Components:

1. Constant(0)
2. Gamma(α, β)

Mixture probability is p

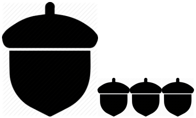
$$\text{ZI-Gamma}(p, \alpha, \beta) \equiv \text{Mixture}([1-p, p], [\text{Constant}(0), \text{Gamma}(\alpha, \beta)])$$

Back to the example

$$\text{make an order} \sim \text{Bernoulli}(p)$$

$$\text{order amount} \sim \begin{cases} \text{Gamma}(\alpha, \beta) & \text{, make an order} = 1 \\ 0 & \text{, make an order} = 0 \end{cases}$$






Takeouts

- Good likelihood helps to get better sense of the problem
  - split purchase probability and purchase amount
  - more possibilities over a classical t-test
- Understanding a problem is a first step to a good likelihood

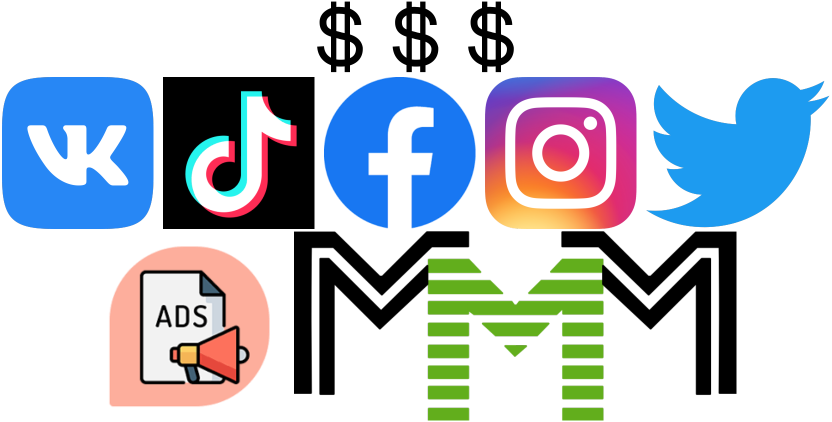
Posterior Distribution

$$\underbrace{p(\Theta|\mathcal{D})} = \frac{p(\mathcal{D}|\Theta)p(\Theta)}{p(\mathcal{D})}$$

Posterior Distribution

$$
\begin{aligned}
&p(\text{what you think}|\text{data})\\
\propto &p(\text{data}|\text{what you think})p(\text{what you think})
\end{aligned}
$$

Case Study: Marketing



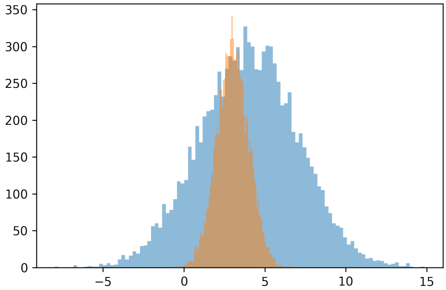
Media Mix Model

MMM - helps to evaluate marketing channels from historical data

Uncertainty

Media Mix Models (read more here)

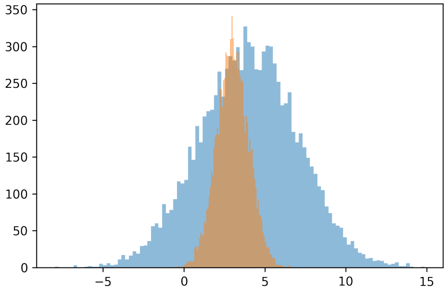
- How valuable is additional $1000 invested in VK? Or Facebook?
- How certain is the model estimation?
- High value, high uncertainty or low value low uncertainty?
- How to allocate money?



Uncertainty

Media Mix Models (read more here)

- How valuable is additional $1000 invested in VK? Or Facebook?
- How certain is the model estimation?
- High value, high uncertainty or low value low uncertainty?
- How to allocate money?



Takeout

Uncertainty helps to make more informed decision

Recap

In Bayesian framework you have:

- Prior = What I think the problem is
- Likelihood = What the facts I have
- Posterior = What the problem actually is *given priors and data*

The Model

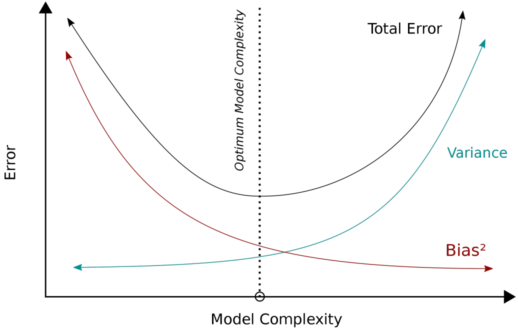
$$p_{\mathcal{M}}(\Theta|\mathcal{D}) = \frac{p_{\mathcal{M}}(\mathcal{D}|\Theta)p_{\mathcal{M}}(\Theta)}{p_{\mathcal{M}}(\mathcal{D})}$$

Treat the model as a "one of many" descriptions of the problem.

Bias - Variance trade-off

Getting to a good model

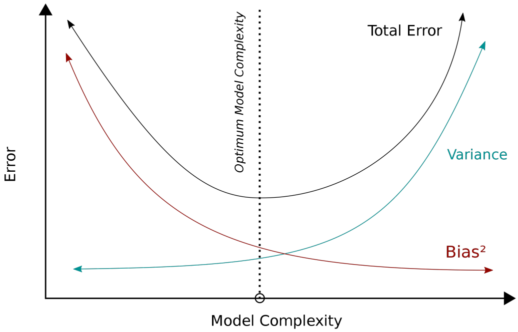
1. Start with a over-simplified model
2. Make sure it samples well
3. Increase the complexity
4. ...
5. Choose the best model using cross validation



Bias - Variance trade-off

Getting to a good model

1. Start with a over-simplified model
2. Make sure it samples well
3. Increase the complexity
4. ...
5. Choose the best model using cross validation



Common mistake

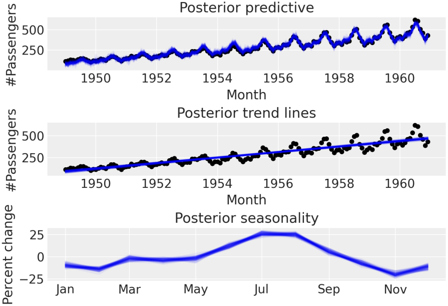
Starting with a complicated model make debugging painful

Case Study: generalized additive models

$$y(t) = \underbrace{g(t)}_{\text{trend}} + \underbrace{s(t)}_{\text{seasonality}} + \underbrace{r(X_t)}_{\text{regressors}} + \varepsilon_t$$

Adding more and more complexity

1. start with a simple trend model
2. add seasonality
3. add fine seasonality details, holidays or other features



When Bayes?

Bayesian modeling does not fit all the use cases at once

- It requires extra skills compared to classical machine learning
- You might not need the extra "uncertainty"

When Bayes?

Bayesian modeling does not fit all the use cases at once

- It requires extra skills compared to classical machine learning
- You might not need the extra "uncertainty"

Proposition

Bayesian modeling starts when fit-predict is useless

When Bayes?

Bayesian modeling does not fit all the use cases at once

- It requires extra skills compared to classical machine learning
- You might not need the extra "uncertainty"

Proposition

Bayesian modeling starts when fit-predict is useless

- You have interpretable confidence intervals
- You have flexibility and control over the model
- You have reliable low data applications

It comes with a price...

You ultimately HAVE to understand the model

PyMC

1. Pure Python!
2. Automated inference
3. No complicated formulas for MCMC!
4. Visualizations with ArviZ
5. Reproducible research
6. Used in industry applications
7. Huge community
8. Active development

https://github.com/pymc-devs/pymc

References I