UPPSALA
UNIVERSITET

- Course information
- Assignments
- Mini-project
- Practicalities
- Examination
- Course improvements

Bayesian Statistics and Data Analysis
Course information

Måns Magnusson
Department of Statistics, Uppsala University

UPPSALA UNIVERSITET

Who am I?

1. Associate professor in Statistics
2. Teaching: Machine Learning and Bayesian Statistics
3. Research: Machine Learning and Bayesian Statistics
   - 3.1 Bayesian Statistics and computational statistics
   - 3.2 Data-centric machine learning
   - 3.3 Statistical inference from textual data
   - 3.4 Applications: Law, sociology, political science, and medicine

UPPSALA UNIVERSITET

Section 1

Course information

UPPSALA UNIVERSITET

Course information

The aims of this course are that, after this course you should:

UPPSALA UNIVERSITET

Course information

The aims of this course are that, after this course you should:

1. have knowledge in basic concepts, philosophy, and perspectives in Bayesian Statistics,
2. derive posterior distributions in simple situations,
3. derive and use predictive distributions,
4. identify and formulate Bayesian probabilistic models for analysis and predictions,
5. estimate models using contemporary computer-based methods for posterior approximations,
6. understand and use basic principles for decisions under uncertainty,
7. have knowledge about and be able to use Bayesian methods for model comparisons,
8. be able to critically evaluate Bayesian methods,
9. report, orally and in writing, a Bayesian statistical analysis

UPPSALA UNIVERSITET

Pre-requisites

- Basic probability theory
  - probability, probability density, distribution
  - sum, product rule, and Bayes' rule
  - expectation, mean, variance, median

UPPSALA UNIVERSITET

Pre-requisites

- Basic probability theory
  - probability, probability density, distribution
  - sum, product rule, and Bayes' rule
  - expectation, mean, variance, median
- Basic linear algebra and calculus

UPPSALA UNIVERSITET

Pre-requisites

- Basic probability theory
  - probability, probability density, distribution
  - sum, product rule, and Bayes' rule
  - expectation, mean, variance, median
- Basic linear algebra and calculus
- Basic visualisation techniques (R or Python)
  - histogram, density plot, scatter plot
- *Note!* This is a masters course in Statistics.

UPPSALA UNIVERSITET

Pre-requisites

- Basic probability theory
  - probability, probability density, distribution
  - sum, product rule, and Bayes' rule
  - expectation, mean, variance, median
- Basic linear algebra and calculus
- Basic visualisation techniques (R or Python)
  - histogram, density plot, scatter plot
- *Note!* This is a masters course in Statistics.

First assignment is a recap.

UPPSALA UNIVERSITET

Who are you? (very brief)

1. Name
2. What is your expectation on this course? Why did you choose it?

UPPSALA UNIVERSITET

Course Outline

Two main parts:

- Core Content (9 lecture blocks)

UPPSALA UNIVERSITET

Course Outline

Two main parts:

- Core Content (9 lecture blocks)
- Assignments (8 individual assignments)

UPPSALA UNIVERSITET

Course Outline

Two main parts:

- Core Content (9 lecture blocks)
- Assignments (8 individual assignments)
- Mini-project: do your own Bayesian data analysis (1-3 students)

UPPSALA UNIVERSITET

Course Outline

Two main parts:

- Core Content (9 lecture blocks)
- Assignments (8 individual assignments)
- Mini-project: do your own Bayesian data analysis (1-3 students)

Exact dates and details; see the course page.

UPPSALA UNIVERSITET

Core Content

- Every week: lectures (approx. 2-4h)
  - Online video material and reading assignments (approx. 2-4h, 50-90 pages a week)
  - Lecture(s): present overall theory and content (overview)
  - Assignment(s): Computational and theoretical individual work. (approx. 12-16h). Deadline Sundays 23.59.
    Start monday morning every week!

UPPSALA UNIVERSITET

Core Content

- Every week: lectures (approx. 2-4h)
  - Online video material and reading assignments (approx. 2-4h, 50-90 pages a week)
  - Lecture(s): present overall theory and content (overview)
  - Assignment(s): Computational and theoretical individual work. (approx. 12-16h). Deadline Sundays 23.59. *Start monday morning every week!*
- Recommended workflow for each week
  - Do the reading assignments
  - Watch the videos (although, optional)
  - Do self-study exercises
  - Start with the assignment
  - Attend lecture (bring questions!)
  - Attend Zoom datalabs (bring questions! Helps with debugging Stan code)
  - Submit the assignment

UPPSALA UNIVERSITET

Section 2

Assignments

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!
3. All labs can be turned in a three times (See Studium for details), but only first time can give VG:
   - 3.1 The week of the assignment
   - 3.2 The last day of the course
   - 3.3 2-4 weeks after the course

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!
3. All labs can be turned in a three times (See Studium for details), but only first time can give VG:
   - 3.1 The week of the assignment
   - 3.2 The last day of the course
   - 3.3 2-4 weeks after the course
4. We will mark and return each assignment within 5 working days (max 10 working days).

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!
3. All labs can be turned in a three times (See Studium for details), but only first time can give VG:
   - 3.1 The week of the assignment
   - 3.2 The last day of the course
   - 3.3 2-4 weeks after the course
4. We will mark and return each assignment within 5 working days (max 10 working days).
5. *Important!* Do *not* write your name anywhere

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!
3. All labs can be turned in a three times (See Studium for details), but only first time can give VG:
   - 3.1 The week of the assignment
   - 3.2 The last day of the course
   - 3.3 2-4 weeks after the course
4. We will mark and return each assignment within 5 working days (max 10 working days).
5. *Important!* Do *not* write your name anywhere
6. *Important!* Do the assignment evaluation!

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!
3. All labs can be turned in a three times (See Studium for details), but only first time can give VG:
   - 3.1 The week of the assignment
   - 3.2 The last day of the course
   - 3.3 2-4 weeks after the course
4. We will mark and return each assignment within 5 working days (max 10 working days).
5. *Important!* Do *not* write your name anywhere
6. *Important!* Do the assignment evaluation!
7. Three one hour zoom seminars per week with individual help on assignments. *Note!* Don't send us e-mail, instead use the Zoom seminars.

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!
3. All labs can be turned in a three times (See Studium for details), but only first time can give VG:
   - 3.1 The week of the assignment
   - 3.2 The last day of the course
   - 3.3 2-4 weeks after the course
4. We will mark and return each assignment within 5 working days (max 10 working days).
5. *Important!* Do *not* write your name anywhere
6. *Important!* Do the assignment evaluation!
7. Three one hour zoom seminars per week with individual help on assignments. *Note!* Don't send us e-mail, instead use the Zoom seminars.
8. We supply grading information for all assignments. Although, they may change!

UPPSALA UNIVERSITET

Assignments

1. Core components and concepts and state-of-the-art methods
2. *Warning!* There might be bugs in the assignments!
3. All labs can be turned in a three times (See Studium for details), but only first time can give VG:
   - 3.1 The week of the assignment
   - 3.2 The last day of the course
   - 3.3 2-4 weeks after the course
4. We will mark and return each assignment within 5 working days (max 10 working days).
5. *Important!* Do *not* write your name anywhere
6. *Important!* Do the assignment evaluation!
7. Three one hour zoom seminars per week with individual help on assignments. *Note!* Don't send us e-mail, instead use the Zoom seminars.
8. We supply grading information for all assignments. Although, they may change!
9. We have quite strict formatting guidelines. Read carefully!
10. Language models (eg ChatGPT) is not allowed. I will include "white herrings". Only what you can read in the assignment is included.

UPPSALA UNIVERSITET

R vs Python and Colab

- We strongly recommend using R in the course as there are more packages for Stan and statistical analysis in general in R

UPPSALA UNIVERSITET

R vs Python and Colab

- We strongly recommend using R in the course as there are more packages for Stan and statistical analysis in general in R
- If you are already fluent in Python, but not in R, then using Python may be easier, but it can still be useful to learn R

UPPSALA UNIVERSITET

R vs Python and Colab

- We strongly recommend using R in the course as there are more packages for Stan and statistical analysis in general in R
- If you are already fluent in Python, but not in R, then using Python may be easier, but it can still be useful to learn R
- All data for the assignments are included in the R package. You can access the data `data` or `data-raw` in the R package folder in the repo if you need to get the data without using the `bsda` R package.

UPPSALA UNIVERSITET

R vs Python and Colab

- We strongly recommend using R in the course as there are more packages for Stan and statistical analysis in general in R
- If you are already fluent in Python, but not in R, then using Python may be easier, but it can still be useful to learn R
- All data for the assignments are included in the R package. You can access the data `data` or `data-raw` in the R package folder in the repo if you need to get the data without using the `bsda` R package.
- We supply a Google Colab template with everything pre-installed.

UPPSALA UNIVERSITET

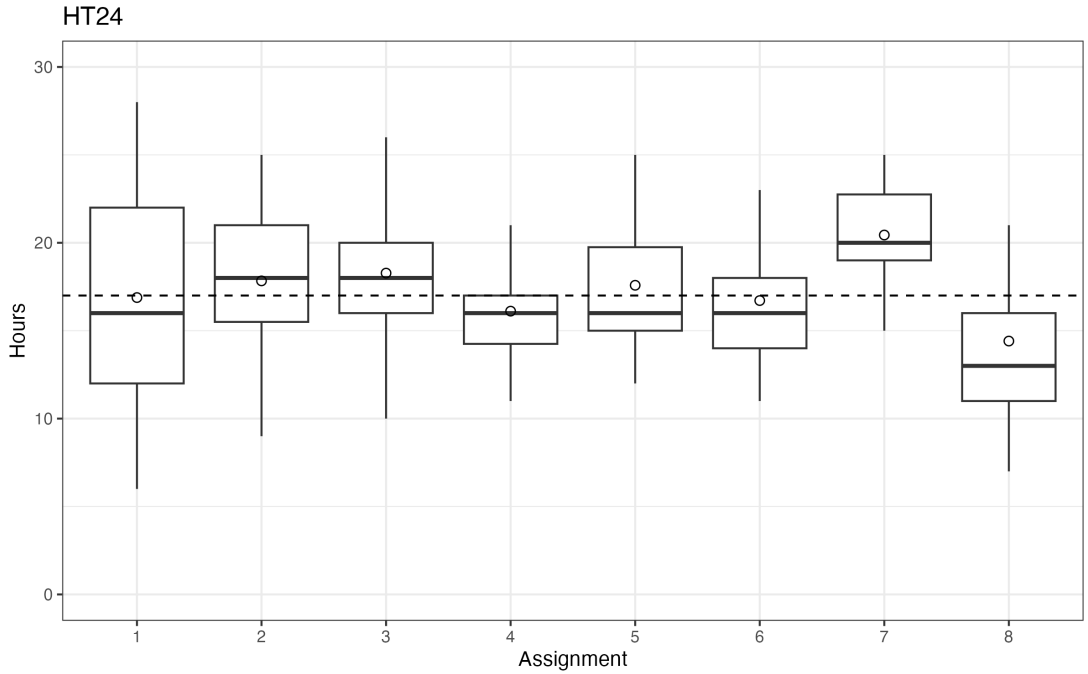
R vs Python and Colab

- We strongly recommend using R in the course as there are more packages for Stan and statistical analysis in general in R
- If you are already fluent in Python, but not in R, then using Python may be easier, but it can still be useful to learn R
- All data for the assignments are included in the R package. You can access the data `data` or `data-raw` in the R package folder in the repo if you need to get the data without using the `bsda` R package.
- We supply a Google Colab template with everything pre-installed.
- We also supply a knitR LaTeX template suitable for Overleaf.

UPPSALA UNIVERSITET

Course Workload

- Conclusions from 2024
  - Slightly too large workload in assignment 7



Student course workload

UPPSALA UNIVERSITET

Stan

- Stan is a probabilistic programming framework (PPF) and ecosystem
- 40+ developers, 100+ contributors, 100K+ users
- R, Python, Julia, Scala, Stata, Matlab, command line interfaces
- More than 120 R packages using Stan
- Many packages to support diagnostics and workflow
- Can be used for frequentist inference as well
- Alternative PPF exists, Turing (Julia), Pyro (PyTorch), etc.

mc-stan.org

UPPSALA UNIVERSITET

Section 3

Mini-project

UPPSALA UNIVERSITET

Mini-project

- See project instructions on webpage for details.
- Data analysis of choice on real data.
- 2-3 students.

UPPSALA UNIVERSITET

Mini-project

- See project instructions on webpage for details.
- Data analysis of choice on real data.
- 2-3 students.
- Supply a half-page project proposal of data and problem in the middle of the course.

UPPSALA UNIVERSITET

Mini-project

- See project instructions on webpage for details.
- Data analysis of choice on real data.
- 2-3 students.
- Supply a half-page project proposal of data and problem in the middle of the course.
- Ideally, use a model not presented in this course.
- Project will last two weeks (half time) - but start earlier.
- The project should use Stan.
- Approximate 40 hours of work *per student*.

UPPSALA UNIVERSITET

Mini-project

- See project instructions on webpage for details.
- Data analysis of choice on real data.
- 2-3 students.
- Supply a half-page project proposal of data and problem in the middle of the course.
- Ideally, use a model not presented in this course.
- Project will last two weeks (half time) - but start earlier.
- The project should use Stan.
- Approximate 40 hours of work *per student*.
- The project should result in a 4 page report (PDF) using the ICML LaTeX template (see course page).
- Project oral presentation (10-15 minutes)

UPPSALA UNIVERSITET

Mini-project

- See project instructions on webpage for details.
- Data analysis of choice on real data.
- 2-3 students.
- Supply a half-page project proposal of data and problem in the middle of the course.
- Ideally, use a model not presented in this course.
- Project will last two weeks (half time) - but start earlier.
- The project should use Stan.
- Approximate 40 hours of work *per student*.
- The project should result in a 4 page report (PDF) using the ICML LaTeX template (see course page).
- Project oral presentation (10-15 minutes)
- The first author is corresponding author

UPPSALA UNIVERSITET

Mini-project

- See project instructions on webpage for details.
- Data analysis of choice on real data.
- 2-3 students.
- Supply a half-page project proposal of data and problem in the middle of the course.
- Ideally, use a model not presented in this course.
- Project will last two weeks (half time) - but start earlier.
- The project should use Stan.
- Approximate 40 hours of work *per student*.
- The project should result in a 4 page report (PDF) using the ICML LaTeX template (see course page).
- Project oral presentation (10-15 minutes)
- The first author is corresponding author
- *This year*:
  - Test to implement complex models in Stan as project - reach out to me.
  - Mini-project discussions: after lectures and at Thursday Zoom sessions (first from next week.

UPPSALA UNIVERSITET

Master theses using Stan

- New thing this year: Joint supervision and work with researchers and students (PhD and masters)
- For students that want to work closely to research(ers)
- Headed by me and Sara Hamis at IT
- First step: a joint hackathon in mid november (morning + lunch).

UPPSALA UNIVERSITET

Master theses using Stan

- New thing this year: Joint supervision and work with researchers and students (PhD and masters)
- For students that want to work closely to research(ers)
- Headed by me and Sara Hamis at IT
- First step: a joint hackathon in mid november (morning + lunch).
- Six project (three from stats and three from IT) on Stan and Bayes:

UPPSALA UNIVERSITET

Master theses using Stan

- New thing this year: Joint supervision and work with researchers and students (PhD and masters)
- For students that want to work closely to research(ers)
- Headed by me and Sara Hamis at IT
- First step: a joint hackathon in mid november (morning + lunch).
- Six project (three from stats and three from IT) on Stan and Bayes:
- At Statistics:
  - Modelling of pollen data in Ancient greece (with Anton Bonnier)
  - Using the Gumbel Softmax to do missing data imputation with discrete parameters (with Jakob Torgander)
  - Modelling proteins in pharmacometrics

UPPSALA UNIVERSITET

Section 4

Practicalities

UPPSALA UNIVERSITET

Practicalities

- Course page/content: Github – please do a PR or submit an issue if something is wrong!
  https://github.com/MansMeg/BSDA/issues

UPPSALA UNIVERSITET

Practicalities

- Course page/content: Github – please do a PR or submit an issue if something is wrong!
  https://github.com/MansMeg/BSDA/issues
- Communication: Studium

UPPSALA UNIVERSITET

Practicalities

- Course page/content: Github – please do a PR or submit an issue if something is wrong!
  https://github.com/MansMeg/BSDA/issues
- Communication: Studium
- Schedule: Time Edit/Studium
- Assignments submissions: Studium

UPPSALA UNIVERSITET

Practicalities

- Course page/content: Github – please do a PR or submit an issue if something is wrong! https://github.com/MansMeg/BSDA/issues
- Communication: Studium
- Schedule: Time Edit/Studium
- Assignments submissions: Studium
- Acknowledgements: Aki Vehtari

UPPSALA UNIVERSITET

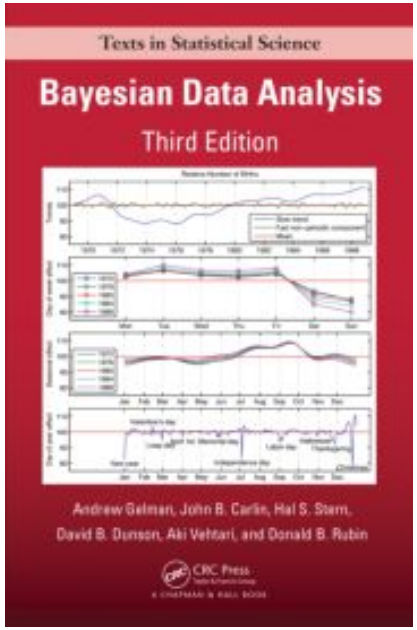
Practicalities

- Course page/content: Github – please do a PR or submit an issue if something is wrong! https://github.com/MansMeg/BSDA/issues
- Communication: Studium
- Schedule: Time Edit/Studium
- Assignments submissions: Studium
- Acknowledgements: Aki Vehtari
- Teaching assistant: Väinö Yrjänäinen

UPPSALA UNIVERSITET

Literature

- Book: Gelman, Carlin, Stern, Dunson, Vehtari & Rubin: Bayesian Data Analysis, Third Edition. (online pdf available)



- Additional articles and blog posts (see reading list per week)

UPPSALA UNIVERSITET

Section 5

Examination

UPPSALA UNIVERSITET

Examination

1. To pass (G): All labs, mini-project, and project review need to be passed

UPPSALA UNIVERSITET

Examination

1. To pass (G): All labs, mini-project, and project review need to be passed
2. To pass with distinction (VG): 7/10 VG points

UPPSALA UNIVERSITET

Examination

1. To pass (G): All labs, mini-project, and project review need to be passed
2. To pass with distinction (VG): 7/10 VG points
3. If everything is correct in an assignment (¿90%), 1 VG point is awarded *on the first submission deadline*.

UPPSALA UNIVERSITET

Examination

1. To pass (G): All labs, mini-project, and project review need to be passed
2. To pass with distinction (VG): 7/10 VG points
3. If everything is correct in an assignment (¿90%), 1 VG point is awarded *on the first submission deadline*.
4. The mini-project is worth 2 VG-points (if it is passed with distinction).
5. Ph.D. students: I suggest you get VG to pass the course. Make the project a potential paper.
6. Reassesment of grades (supply form to course admin)
7. Failing the course: You will need to redo all assignments and mini-project next year.
8. Large language models, e.g. chatGPT, are not allowed. "White herrings" are included!

UPPSALA UNIVERSITET

Section 6

Course improvements

UPPSALA UNIVERSITET

Course improvements since last time

- Better balance between Assignment 7 and 8.
- White herrings and Large Language Model use.
- New idea (test) of mini project.

UPPSALA UNIVERSITET

Section 6

Course improvements

UPPSALA UNIVERSITET

Comments from previous students

- Put in the work, it is worth it.

UPPSALA UNIVERSITET

Comments from previous students

- Put in the work, it is worth it.
- There is alot to do but it will be worth it if you really put in work.

UPPSALA UNIVERSITET

Comments from previous students

- Put in the work, it is worth it.
- There is alot to do but it will be worth it if you really put in work.
- Be prepared that the tempo and workload is very high.

UPPSALA UNIVERSITET

Comments from previous students

- Put in the work, it is worth it.
- There is alot to do but it will be worth it if you really put in work.
- Be prepared that the tempo and workload is very high.
- Attend the lectures and start with the assignments early. Because they are really fun if you get passed the time pressure aspect

UPPSALA UNIVERSITET

Comments from previous students

- Put in the work, it is worth it.
- There is alot to do but it will be worth it if you really put in work.
- Be prepared that the tempo and workload is very high.
- Attend the lectures and start with the assignments early. Because they are really fun if you get passed the time pressure aspect
- Start the assignments as soon as possible. It is perfectly reasonable to be done with the assignments before the weekend, so aim for that.

UPPSALA UNIVERSITET

Comments from previous students

- Put in the work, it is worth it.
- There is alot to do but it will be worth it if you really put in work.
- Be prepared that the tempo and workload is very high.
- Attend the lectures and start with the assignments early. Because they are really fun if you get passed the time pressure aspect
- Start the assignments as soon as possible. It is perfectly reasonable to be done with the assignments before the weekend, so aim for that.
- Do the course properly, it is worth the effort

UPPSALA UNIVERSITET

- Course information
- Assignments
- Mini-project
- Practicalities
- Examination
- Course improvements

Questions?

Questions?