Land Acknowledgement

The land on which the University of Toronto operates has for thousands of years been the traditional land of the Huron-Wendat, the Seneca, and most recently, the Mississaugas of the Credit. Today, this meeting place is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to live, work, learn, and grow on this land.

Applied Bayesian Statistics (STA365H1S LEC0101)

Winter 2024 / Prof. Scott Schwartz

- Quercus
  - All annoucements, email communications, and course material releases will be done through Quercus
- General public course related questions should be asked through piazza (sign-up)

Teaching Assistants Mandy Yao, Yovna Junglee, Jorge Arturo EF, Morris Greenberg

- TAs hold office hours but do not communicate with students thruogh emails or piazza

Summary

- Theoretical / Conceptual lectures and exams
- Practical coding assignments and course project
  - http://jupyter.utoronto.ca (assuming library versioning requirements are working)
  - http://colab.research.google.com (as a back up if UofT jupyterhub library versioning requirements fail)
  - Homework submissions will be links to notebooks hosted on personal https://github.com/ repos

Calendar

Date	Topics
Jan 10	Bayes Theorem, Beta-Binomial, Bayesian bandit
Jan 17	Normal-Normal, Conjugate and other Priors
Jan 24	Gibbs sampling Normal-Gamma, Probabilistic Programming with PyMC, and Autocorrelation

Date	Topics	
Jan 31	Metropolis-Hastings, Hamiltonian Monte Carlo, Diagnostics	
Feb 07	Bayesian Multiple Linear Regression, Inverse-Wishart and LKJ (Cholesky) priors and Multivariate Normal Distributions	maybe delay
Feb 14	Midterm	
Feb 21	Reading Week	
Feb 28	GLMs, Hierarchical Modeling, Multiplicity adjustment, Variable Selection, Shrinkage, Robust Regression	
	Project Teams must be formed by Mar 06 for 2 of 2 points *(find team members on piazza or in class, etc.)*	
Mar 06	Bayes Factors, Effective Model Size, Generalization/Overfitting, Model Selection (WAIC, LOO-CV, etc.), and Random Effects Models	
Mar 11	Last date to drop course	
Mar 13	(Nonparametric) Dirichlet Process Mixture Models and Latent/Missing Value Imputation	maybe delay
Mar 20	(Nonparametric) Gaussian Processes and Variational Inference	maybe omit
Mar 27	TBA	
	Project Videos must be submitted by Apr 02 -- late submissions MAY (or MAY NOT) be accepted after this deadline with a penalty	
	Project "Peer Evaluation" must be submitted by Apr 03 for 2 of 2 points -- late submissions MAY (or MAY NOT) be accepted for 1 of 2 points	
Apr 03	Couse Project Presentations	

Learning Objectives

- Ability to perform applied Bayesian analysis from simple to moderate complexity on real world data sets.
  - Bayesian analysis is generally considered the "gold standard" of uncertainty quantification since the Bayesian paradigm is predicated on careful specification of the data generating mechanism to fully capture the variability and dependency structure present within the data, and fully and coherrently propegating that uncertainty into parameter posterior inference.
  - Demonstrated capability in applied Bayesian analysis communicates to industry practitioners and academic researchers alike a strong understanding of and commitment to characterizing and respecting uncertainty during the data analysis process, and

subsequently utilizing data driven information about uncertainty as part of downstream decision making efforts.

- Ability to use Python and PyMC to implement Bayesian analyses, including creating specifications for unique data generating mechanisms and associated hierarchical modelling inference structures.
  - Python is an undeniably ubiquitous programming language present in both industrial and academic contexts, and can greatly faciliate the speed and scale at which data analyses can be carried out.
  - PyMC in particular is a highly popular probabilistic programming language which provides universal sampling of arbitrary data generating mechanisms and prior specifications that is generally effective and computationally efficient for many modeling and analysis purposes.
- Ability to evaluate, critque, and troubleshoot Bayesian modeling specifications and code producing Bayesian analyses, including assessing the influence of choices regarding priors and other modelling choices and decisions, interpreting model convergence and performance diagnostics, and providing posterior inference analysis and interpretation after modeling fitting and/or sampling and validation.
  - A true practitioner of Bayesian analysis must be able to assess, criticize, and utilize the inputs and outputs of a Bayesian posterior analysis since being unable to do so could result in unreliable and misleading statistical inference analysis and poor or wrong conclusions and decisions making.
- Gain familiarity with the historical development of Bayesian Analysis and exposure to its current advanced forms and manifestations, including modern Markov Chain Monte Carlo (MCMC) sampling methods such as Hamiltonian Monte Carlo (HMC), Langevin Monte Carlo (LMC), and Stochastic Gradient MCMC (SG-MCMC) as well as Bayesian Deep Learning (BDL).
  - While this is a course in "Applied" Bayesian analysis, some degree of familiarity with the more formal theoretrical aspects of the development of Bayesian analysis are expected of professional Bayesian practitioners, so familiarity and knowledge with the historical topics driving Bayesian analysis will be expected of analysts working in the field of Applied Bayesian analysis.
  - Many advanced modern Bayesian analysis methodologies such as HMC and the BDL techniques of "Bayes by Dropout" and "MC-Dropout" can actually be relatively quickly implemented and utilized, so having the ability to produce sophisticated cutting-edge data analyses by leveraging these tools affords a Bayesian practitioner the possibility to greatly extend the scope of their analysis capabilities in a relatively accessible manner.

Course Resources

The course is largely self-contained, with material for the course primarily provided through Quercus as jupyter notebook slide decks; however, some students may also be interested in the example vignettes,expositional demonstrations, more formal "textbooks" available at https://www.pymc.io/projects/docs/en/stable/learn.html

Weekly Schedule

Offset	Event	Day	Time	Location	Recorded?	Piazza Monitored?
0	Lecture	Wednesday	3:10-6 (15:10-18) PM ET	MB 128	Yes	No
			Bring paper and pencil...	...there is usually an in class quiz		
+1	Prof Hybrid? OH	Thursday	5-6:30 PM ET	Scott zoom + Loc TBD	Yes	No
+2-4		Fri-Sat-Sun				Yes
+5	TA Hybrid? OH	Monday	3:30-5 PM ET 5-6:30 PM ET	Morris zoom (wishart) + Loc TBD Mandy zoom (629921) + Loc TBD	No	Yes
+6	TA Hybrid? OH	Tuesday	3:30-5 PM ET 5-6:30 PM ET	Arturo zoom + Loc TBD Yovna zoom + Loc TBD	No	Yes
+7	Homework Due	Wednesday	11 AM ET	Submit a link on Quercus to a `.ipynb` notebook `github` repo		No

- OH all weeks except Reading Week
- TA OH subject to change

Grades

There are two marking schemes for the course

	In-Class Quizzes	Project Evaluations	Homework	Course Project	Midterm	Final	
	Participation	During Apr 03 Class	2/1.5/1/0.5/0	See Below	Marked	Marked	
	$\frac{1}{2} \times 10$	1	2×7	20	30	30	
	5	1	14	20	30	30	100 points total
or	0	0	0	20	40	40	100 points total

- Students may request the 40 point exam scheme at any time by emailing the course instructor through Quercus

Missed work

In-Class Quizzes	Occurs in class and must be completed in-person so no late submissions are accepted
Project Evaluations	Occurs in class and must be completed in-person so no late submissions are accepted

Homework	MAY be submitted late for a 1 point penalty so long as solutions have not yet been released
Course Project	Late submissions MAY (or MAY NOT) be accepted after this deadline with a penalty (see below for details)
Midterm	The points for a missed midterm exam will be reallocated to the final exam by emailing the course instructor (through Quercus) your ACORN Absence Declaration
Final	Students must petition FAS for a deferred final exam

Absence Declaration Policy

The course policies regarding absense conform to the new absence declaration policy. If special consideration appears necessary for a particular assignment beyond the policies indicated above, a review of the sitution can be initiated once per term using the Acorn Absence Declaration Tool (and providing the necessary documentation to the course instructor through a Quercus email).

Remarking

Remark requests must be made within one week of the release of the return of marks or solutions for the associated assignment

Late Enrollments

Students who miss the Jan 10 and/or Jan 17 lectures due to late enrolment may submit missed Homework assignments and In-Class Quizzes by Wendesday Jan 24

- Solutions for the Jan 10 Homework will have a delayed return on or after Jan 24

Religious Accommodations

If you anticipate missing a course activity due to a religious observance, please let us know at least three weeks ahead of your observance, and we will provide alternate assignment arrangements

Accessibility and Additional Accomodations

For course accomodations please contact Accessibility Services or your College Registrar and alert the course instructor through a Quercus email

If the policies above do not appear sufficient for your circumstances contact your college registrar (https://www.artsci.utoronto.ca/current/academic-advising-and-support/college-registrars-offices)

In-Class Quizzes (5 points)

These are completion marks for effort and are turned in as a sheet of paper with your name and email during class

- Quizzes take place at the beginning of class and students must be present in class to complete and submit the quiz; however, students are not required to stay for the lecture after completing the quiz to recieve credit for the quiz

Project Evaluations (1 point)

See **"Class Presentation" Video** in **Course Project** below

Homework (14 points)

Homework can be completed on UofT Jupyterhub or google collab and must be submitted on Quercus as a link to a github repo with a jupyter notebook

- Due the "next Wednesday" at 11 AM ET

> ***May be submitted late for a 1 point penalty so long as solutions have not yet been released***

2	Completed Fully	and correctly
1.5	Mostly Completed	with some errors
1	Good Effort	but not being done correctly
0.5	Minimal Work	with something submitted
0	No Work	with nothing submitted

Course Project (20 points)

	Team Formation	Class Video	Appendix Video	Peer Evaluation	Project Participation
Points	2	4	4	2	8

Team Formation (2 points): Due Mar 6

Each team must email course instructor (through Quercus) with your project team and ALL team member emails (by CC'ing all team members on the email)

- 4 students/team or 5 students/team if some team members are CR'ing the course

Points	Rubric Evaluation
2	Team formed by Mar 06
1	Team formed after Mar 06
0	Student assigned to a project team by course TA/Instructor

"Class Presentation" Video (4 points): Due Apr 2

> ***May not be submitted late***

Submit a link to a zoom recording (including the password to access the video) on Quercus of a 4-minute video (recorded on zoom) presenting data modeling analysis and introducing the

- data
- modeling
- findings

Evaluation will be A/B/C/D/F (1/0.75/0.5/0.25/0) for the following items

Item	Data	Modeling	Findings	Presentation
Points	1	1	1	1
Criteria	Explanation Clarity	Specification Clarity	Comprehensive Model Usage	Overall Quality
Details	what the data is and is useful for	data generating model prior, hierarchy details, + unique aspects noted without technical errors	fully leveraging model not missing analyses	verbal exposition and visualizations presented clearly

"Project Evaluations" During Apr 03 Class

> ***May not be submitted late***

Each student in attendance will score the above rubric and submit their scores for 1 point

- Submissions will be a formatted excel sheet submitted on Quercus

"Appendix" Video (4 points): Due Apr 2

> ***May be submitted late so long as marking has not yet completed***

Submit a 10-minute video (recorded on zoom) on Quercus as a link to a zoom recording (including the password to access the video)

- The first 4-minutes of this video must be the (possibly rerecorded) "Class Presentation" Video above
- The last 6-minutes of this video should go more deeply into clarifying and expanding the technical details of the analysis which (due to time limiations) could not be addressed in the "Class Presentation" Video

Evaluation will be A/B/C/D/F (4/3/2/1/0) based on a subjective overall assessment of the quality of the project work with respect to items such as

- Explanation of technical decisions and associated assumptions/implications
- Diagnostic analysis of the model fitting performance and model appropriateness
- Quality, appropriateness, and completeness of the visualizations presentations and explanations

Peer Evaluation (2 points): Due Apr 3

> ***May be submitted late for a 1 point penalty so long as marking has not yet completed***

You will submit an excel sheet on Quercus containing the following information with the requested formatting

Your Full Name	Team Member Full Name	Contribution Evaluation	Comments
(First then Last)	(First then Last)	(Choose from the options below)	(Please share relevant details or explanations)
A a	A a		
A a	B b		
A a	C c		
A a	D d		

Contribution Evalution Options

- "Above Average"
- "Fair/Average"
- "Unfair/Below Average"
- "Unfair/No Contribution"

Project Participation (8 points)

This will be based on an assessment of student contribution to project team work as evidenced by the above "Peer Evaluations"

- Roughly speaking, depending upon the details of feedback provided by all the students in their "Peer Evaluation" submission...
    - "Peer Evaluations" of "Unfair/Below Average" or "Unfair/No Contribution" could result in a loss of 1 or 2 marks, with unanimous evaluations of "Unfair/Below Average" possibly resulting in a 4/8 "Unfair/No Contribution" mark and unanimous evaluations of "Unfair/Below Average" possibly resulting in a 0/8 "Project Participation" mark
        - Students consistently evaluated as "Unfair/Below Average" by their project peers may not recieve "Appendix" Video (4 points) marks
        - Students consistently evaluated as "Unfair/No Contribution" by their project peers may not recieve "Appendix" Video (4 points) or "Class Presentation" Video (4 points) marks
- The following marks are NOT affected by "Peer Evaluations"
    - Team Formation (2 points)
    - Peer Evaluation (2 points)

Collaboration and Academic Integrity

Working with peers within the constraints of the academic integrity policies is highly encouraged

- for example participation in RSGs is strongly recommended: https://sidneysmithcommons.artsci.utoronto.ca/recognized-study-groups/

All students, faculty and staff are expected to follow the University's guidelines and policies on academic integrity: https://www.artsci.utoronto.ca/current/academic-advising-and-support/student-academic-integrity (and see also, e.g., http://academicintegrity.utoronto.ca and https://www.academicintegrity.utoronto.ca/smart-strategies/)

- violations will be processed through department and institutional SAI protocols and procedures

AI Support

Generative artificial intelligence large language models such as Bard or ChatGPT are allowed for this class

- however, I don't currently suspect they will be particularly helpful for this coures; so... it would be very helpful for me and would be greatly appreciated if
- when you submit work that relies on this kind of tool, you would please be so kind as to briefly describe how you used this resource ...this would help me determine if we might be able leverage and provide practice using these kinds of tools in future iterations of the course

TA Assignments

		Total		Notes	
Office Hours	12×1.5 hr	18			Preferably this should be hybrid in-person/online
HW Marking	7×2.5 hrs	18		~3.75 min/student (~40 students+regrades)	TAs select correct solution "Examples" to share with the class
Midterm	10 hrs	10		~15 min/student (~40 students+regrades)	Midterm is Feb 14
Final	10 hrs	10		~15 min/student (~40 students+regrades)	Final date TBA
Invigilation	6 hrs	6			Midterm is Feb 14 and Final date TBA
Project	10 hrs	3		~4 min/project (~40 projects)	Apr 03 in-class presentations
		1		~5 min/project (~10 projects)	Collect student scored "Class Presentation" Video rubrics
		3		~15 min/project (~10 projects)	Review Apr 02 project video submissions and assign marks
		3		~15 min/project (~10 projects)	Collect and review participation evaluations and assign participation marks
Other	4 hrs	4			
		76	total		