Department of Mathematics & Statistics
Concordia University

STAT 497 (MAST 679/MAST 881), Sec. P
Topics in Statistics & Probability
"High Dimensional Probability with Applications to Data Science'
Winter 2022

Instructor: Dr. S. Brugiapaglia, Office: LB 921-9 (SGW), Phone: 848-2424, Ext. 4250
Email: simone.brugiapaglia@concordia.ca

Lectures: Fridays, 14:45-17:30, LS 208 SGW (In person).

Office Hours: TBA

Description: This course is an introduction to high-dimensional probability techniques lying at the core of contemporary data science. Topics covered include (time permitting):

- **Concentration of sum of independent random variables**
  Hoeffding's inequality, Chernoff's inequality, sub-gaussian and sub-exponential random variables, Bernstein's inequality
- **Random vectors in high dimension**
  concentration of the norm, covariance matrices and PCA, high-dimensional sub-gaussian distributions
- **Non-asymptotic random matrix theory**
  nets, covering numbers, spectral bounds for sub-gaussian matrices
- **Concentration without independence**
  concentration of Lipschitz functions on the sphere, Johnson-Lindenstrauss lemma
- **Further high-dimensional probability tools**
  decoupling, symmetrization, random matrices with non-i.i.d. entries
- **Random processes**
  basic concepts and examples, Gaussian width, chaining, Dudley's inequality

Although the focus of the course will be theoretical, we will illustrate applications of high-dimensional probability theory in a variety of data science areas such as random graphs, community detection in networks, clustering, dimensionality reduction, statistical learning theory, and sparse recovery.

Main Textbook: Vershynin, R. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge University Press, 2018. E-book: https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html

Other References: Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*. Vol. 48. Cambridge University Press, 2019. E-book: https://www.cambridge.org/core/books/highdimensional-statistics/8A91ECEEC38F46DAB53E9FF8757C7A4E

Blum, A., Hopcroft, J., and Kannan, R. *Foundations of data science.* Cambridge University Press, 2020.
E-book: https://home.ttic.edu/~avrim/book.pdf

Pre-requisites: Working knowledge of linear algebra, probability theory, and analysis. General familiarity with basic notions about Hilbert and normed spaces, linear operators, and knowledge of measure theory are not essential but could be helpful.

Assignments: Homework will be assigned approximately every other week and will focus on theoretical problems.

Exams: There will be no exams. Instead, students will be required to complete an independent project on a topic approved by the instructor and based on recent research papers in the field. The project will be presented in the form of an oral presentation *and* submitted as a written report. The focus of the project can be theoretical or computational, based on student's background and interests.

Grading Scheme: 25% Homework assignments + 75% Project (proposal, oral presentation, written report).

PhD students will be required to do additional work (to be determined) compared with their MA/MSc classmates.

If the grading scheme for this course includes graded assignments, a reasonable and representative subset of each assignment may be graded. Students will not be told in advance which subset of the assigned problems will be marked and should therefore attempt all assigned problems.

Communication: Communication between the students and the instructor will take place in class, via Moodle announcements, or emails. Students are responsible for reading and taking note of all electronic communication from the instructor and the University.

Academic Integrity and the Academic Code of Conduct
This course is governed by Concordia University's policies on Academic Integrity and the Academic Code of Conduct as set forth in the Undergraduate Calendar and the Graduate Calendar. Students are expected to familiarize themselves with these policies and conduct themselves accordingly. "Concordia University has several resources available to students to better understand and uphold academic integrity. Concordia's website on academic integrity can be found at the following address, which also includes links to each Faculty and the School of Graduate Studies: concordia.ca/students/academic-integrity." *[Undergraduate Calendar, Sec 17.10.2]*

Use of Zoom
Note: Zoom is included as an institutionally-approved technology. This means we have been assured of the privacy protections needed to use freely within the classroom)

Zoom will be used in this course to facilitate learning at a distance. It may be used to record some or all of the lectures and/or other activities in this course. If you wish to ensure that your image is not recorded, speak to your instructor as soon as possible.

Also, please note that you may not share recordings of your classes and that the instructor will only share class recordings for the purpose of course delivery and development. Any other sharing may be in violation of the law and applicable University policies, and may be subject to penalties.

Behaviour
All individuals participating in courses are expected to be professional and constructive throughout the course, including in their communications.

Concordia students are subject to the Code of Rights and Responsibilities which applies both when students are physically and virtually engaged in any University activity, including classes, seminars, meetings, etc. Students engaged in University activities must respect this Code when engaging with any members of the Concordia community, including faculty, staff, and students, whether such interactions are verbal or in writing, face to face or online/virtual. Failing to comply with the Code may result in charges and sanctions, as outlined in the Code.

Intellectual Property
Content belonging to instructors shared in online courses, including, but not limited to, online lectures, course notes, and video recordings of classes remain the intellectual property of the faculty member. It may not be distributed, published or broadcast, in whole or in part, without the express permission of the faculty member. Students are also forbidden to use their own means of recording any elements of an online class or lecture without express permission of the instructor. Any unauthorized sharing of course content may constitute a breach of the Academic Code of Conduct and/or the Code of Rights and Responsibilities. As specified in the Policy on Intellectual Property, the University does not claim any ownership of or interest in any student IP. All university members retain copyright over their work.

Extraordinary circumstances
In the event of extraordinary circumstances and pursuant to the Academic Regulations the University may modify the delivery, content, structure, forum, location and/or evaluation scheme. In the event of such extraordinary circumstances, students will be informed of the change.