HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

BST 230: Probability I
Fall 2024
Tuesdays & Thursdays, 9:45 – 11:15 pm

Instructor Information

Faculty
Jeffrey W. Miller
Assistant Professor of Biostatistics
Office: 655 Huntington Avenue, Building 1, Room 419
Phone: 617-645-1068
Email: jwmiller@hsph.harvard.edu
Office hour: TBD

Teaching Fellow
Christian Testa, PhD student, Biostatistics department
ctesta@hsph.harvard.edu
Office hour: TBD

Course Description

Axiomatic foundations of probability, independence, conditional probability, joint distributions, transformations, moment generating functions, characteristic functions, moment inequalities, sampling distributions, modes of convergence and their interrelationships, laws of large numbers, central limit theorem, and stochastic processes.

Pre-requisites
You must be a Biostatistics student or have taken BST222 to register for this course.

Credits: 5 credits

Course Website

Canvas Course Website: https://canvas.harvard.edu/courses/142083

Course Readings

The lectures will primarily be based on my lecture notes, which will be available on the course website. The structure and content of the course will follow the following required textbooks:

- *Statistical Inference (Second Edition)*, by George Casella and Roger L. Berger. Cengage Learning, 2021.
- *Probability and Random Processes (Fourth Edition)*, by Geoffrey Grimmett and David Stirzaker. Oxford University Press, 2020.

Lectures/Labs

Lectures will be held on Tuesdays and Thursdays at 9:45-11:15am in FXB G13.
Labs will be held weekly, time and location TBD.

School holidays (no class): Nov 28 (Thanksgiving)

Grading, Progress, and Assessment

The final grade for this course will be based on:

- Homework (20%)
- Two midterm exams (25% each) - Tuesday, Oct 8 and Tuesday, Nov 12
- Final exam (30%) - Thursday, Dec 19

Each exam will focus on the material covered since the previous exam, but the content builds upon earlier concepts, so you will need to know all of the material up to that point.

Homework policies

Homework assignments will consist of problem sets and computer programming assignments. Assignments will be posted on the course website, and homework submission is mandatory. Homework will be graded by around one week after the due date.

Homework must be submitted online on the course website. Any handwritten material must be legible, otherwise no credit will be given. For programming exercises, you may use R, Python, or Julia, whichever you prefer. Include (a) plots and numerical results when appropriate, (b) discussion of the results when appropriate, (c) any supporting derivations, written out separately from the code, and (d) your source code (typed). The TAs will not run your code (e.g., to generate plots, output, etc.), so anything you want them to see must be included.

Policy on homework collaboration:

- Each student is required to come up with their own solutions for the homework.
- Students are allowed to discuss the problems in general terms (without sharing complete solutions) among themselves, or with the TAs or instructor. HOWEVER, when writing up their solutions, students are required to do this on their own, without copying from any other source.
- Students are forbidden from using solutions from any other source (such as solutions found online). The use of AI such as ChatGPT or similar technologies is not allowed.
- Violation of this policy will result in a score of zero for that assignment, and possible disciplinary action.

Late submission policy:

- Homework submissions will be timestamped, and late submissions will be penalized as follows: starting from the due time until 48 hours after the due time, a multiplicative factor starting at 1.0 and decreasing linearly to 0.0 will be applied to the score. So, for example, the score for an assignment submitted 12 hours late will be multiplied by 0.75 (75% credit), 24 hours late will be multiplied by 0.5 (50% credit), and 48 hours late or later will be multiplied by 0.0 (no credit).
- There will be no make-ups or extensions.

Background Reading

To fully understand the course material, you will need to be familiar with the basics of real analysis (particularly sequences, the topology of the real numbers, and continuous functions). A good set of lecture notes is available free for download here: Introduction to Analysis Notes (ucdavis.edu) and the complete pdf is in Files/Reading on the Canvas site.

We will also assume familiarity with the basics of linear algebra. Two sources of background on the linear algebra concepts we will use are in Files/Lectures/A-Linear-algebra-basics.pdf and Files/Readings/Bishop_LinearAlgebra_pp695-701.pdf on the Canvas site.

You will need to be able to write mathematical justifications and derivations in your homework and exams. While formal proofs are not the focus of this course, it is important to be able to read and write mathematical arguments. Here are two helpful resources:

https://opendsa-server.cs.vt.edu/ODSA/Books/Everything/html/Proofs.html
https://artofproblemsolving.com/blog/articles/how-to-write-a-solution

Outline of Topics / Learning Objectives

CB = Casella & Berger book
GS = Grimmett & Stirzaker book

Upon successful completion of this course, students should understand the following topics:

- Introduction (CB 1.1)
  - History of probability, Set theory basics
- Probability basics (CB 1.2 - 1.3)
  - Measure theory basics, Properties of probability measures, Combinatorics, Conditional probability and Independence
- Random variables (CB 1.4 - 1.6, CB 2.1)
  - Random variables, Change of variable formula, Probability integral transform
- Expectations of random variables CB 2.2 - 2.4)
  - Expected values, Properties, Existence, Moments, Differentiation of expectations
- Families of distributions (CB 3.1 - 3.5)
  - Discrete and continuous families, location-scale families, exponential families
- Multiple random variables (CB 4.1 - 4.2, 4.6)
  - Random vectors, joint distributions, marginal and conditional distributions
- Independence and conditional expectations (CB 4.2, 4.4, 4.6)
  - Independence, conditional expectations, mixture distributions
- Bivariate dependence (CB 4.3, 4.5)
  - Correlation, covariance, bivariate transformations, bivariate normal distribution
- Inequalities (CB 3.6, 3.8, 4.7, also Inequalities-ch15M.pdf in Files/Reading on Canvas)
  - Markov, Chebyshev, Jensen, AM-GM, Hölder, Cauchy-Schwarz, Minkowski
- Gaussian distributions (Bishop pp. 78-93, in Files/Reading on Canvas)
  - Multivariate normal, marginals and conditionals, linear-Gaussian model

- Statistics of a random sample (CB 5.1 - 5.4)
  - Sampling distributions, definition of a statistic, statistics of normal samples, order statistics and their friends
- Convergence of random variables (CB 5.5)
  - Modes of convergence, Limit theorems, Delta method
- Laws of large numbers (CB 5.5, GS 7.4 - 7.5)
  - Weak laws of large numbers, Strong laws of large numbers, Generalizations
- Central limit theorems (CB 5.5, GS 5.7 – 5.10)
  - Weak convergence, characteristic functions, central limit theorems
- Stochastics processes (GS 5.1, 5.3 - 5.4, 6.1, 6.3, 6.8, 7.7 – 7.8, 8.2, 9.5)
  - Generating functions, random walks, branching processes, Martingales, Markov chains, Poisson processes

Please note, session topics and activities may be subject to change during the course.