Mathematics & Statistics Department Writing Plan
Boston University
2023-2024 Academic Year

Conducted by
Writing in the Disciplines Program

Financially Supported by a Gift from Jennifer Simpson (CAS 2001)
July 31, 2024

Academic Unit Writing Plan

Writing Plan COVER PAGE

X__ 1st edition ____ Revision

Academic Unit Name College Mathematics and Statistics Department____

Department Chair Contact Information Glenn Stevens ghs@bu.edu____________

Writing Plan Ratified by Faculty

Date: May 14, 2024________

Votes: 6 (yes)/ 0 (no)/0 (absent)

Process by which Writing Plan was ratified within unit: (vote, consensus, other): consensus in person and via email

*****

Signatures

WID Faculty Consultant:

Department/Program Chair/Director: Glenn Stevens

Director Writing in the Disciplines: David Shawn, Assoc. Dir. WID

Associate Dean or Dean of College:

Unit Profile:

______ Professors	______ Lecturers
______ Assoc. Professors	______ Sr. Lecturers
______ Asst. Professors	______ Master Lecturers
______ Research/Clinical Professors	______ Other (e.g. Affiliated Faculty)

	Major Name Total # undergrad students enrolled in major as of Spring 2024	Total # undergrad students graduating w/major in Spring 2024
Pure Math		
Applied Math		
Statistics		
Economics and Math		
Math and CS		
Math and Philosophy		
Math and Math Ed.		
Statistics and CS		

Table of Contents

Writing Plan Narrative (Executive Summary) - For what reason(s) did this unit (department, school, college) become involved in this project? What key implementation activities are proposed in this edition of its Writing Plan and what, briefly, is the thinking behind these proposed activities?

The Department of Mathematics and Statistics has created this Writing Plan in an effort to better support our students' learning. Writing plays a critical role in any field. In mathematics and statistics, the role of writing is distinctive both in terms of the precision, clarity, and symbolic notation that is required, and also because of the role that writing plays in helping us to clarify our own thinking and understanding. By communicating these ideas more intentionally through our courses, we hope to better enable our students to not only be more effective communicators, but also to be more perceptive thinkers.

In this writing plan we first discuss discipline specific writing characteristics, including rigor, precision, clarity, and conciseness, as well as the idea of writing for understanding. We also discuss discipline specific writing tools, including LaTeX, the use of numerical computation, and the description and depiction of data-based and numerical results.

We then outline the ways that writing instruction is currently integrated into our undergraduate curriculum, discussing separately the lower division courses (MA 100-200 level) and upper division courses (MA 300-500 level). We also discuss the forms of assessment that are currently being used in our courses, particularly as they relate to student writing. Subsequently, we make recommendations for changes that could be incorporated into our existing coursework to better achieve the writing-related goals of our department. In addition, we also make some recommendations for structural changes to our course offerings that we feel would substantially improve our program.

A summary of our key pedagogical recommendations is as follows:

- Lower division courses: intentionally frame mathematics/statistics writing (including both pen-and-paper computations common in calculus/statistics, as well as code writing) as *writing*; discuss correct usage of common symbols ($=$, $\Rightarrow$); assign more language-intensive problems, eg from MyLab Math, including true/false questions; emphasize standard phrasing for discussing and interpreting numerical and statistical techniques.
- Upper division courses: continue to frame mathematics and statistics work as a form of writing, and emphasize that it serves not just as a way to communicate but also as a way to explore and clarify our own thinking; enlarge the discussion of common symbols to also include common words and phrases that appear in proof writing (e.g. there exists, for all, for some, assume, implies); incorporate more writing-focused assignments with appropriate levels of feedback; reinforce correct descriptions of numerical and statistical analysis, as well as ideas related to reproducible results.
- Create shared departmental resources that will support the above recommendations, including examples of writing-focused syllabi, rubrics, assignments, and assessments; an asynchronous module for learning LaTeX; frameworks for providing feedback on student writing; and examples of (anonymized) student writing.

A summary of our key curricular recommendations is as follows:

- The creation of a training for incoming teaching fellows that would help TFs better support student writing.
- The redesign of MA 225 and MA 242 into versions "for majors" that explicitly provide writing instruction to help students transition to upper-division courses.
- The development of a 400-level "Tools for Statistics" course to serve as a follow-up course for MA 213/214 and that would include numerous writing-focused aspects.

The rationale for the first two curricular recommendations is connected with the historical context of how teaching in our department has changed over the last few decades. The number of BU students who take our courses has been increasing, which has led to an increase in our department's reliance on large lecture courses that involve TF support. For this reason, it is important that we include our TFs in any plans for modifications to the feedback that we provide to our students. Moreover, to accommodate the increased enrollments in MA 242 (Linear Algebra), our department moved from small, 25-student sections, to large, 100+ student sections around AY 19-20 (see the spreadsheet entitled "Historical Enrollments in Selected Undergraduate Math&Stat Courses" in Appendix A). Although this felt necessary, given the departmental resources we had at the time, this was not in the best interests of our students. This course, along with MA 225, are key courses in our curriculum, and the second suggestion above is intended to reflect that, and to allow for corresponding levels of attention and support to students in these courses. The rationale for the third curricular recommendation stems from the increasing importance of computational skills such as fundamental programming skills (object-oriented programming, reproducibility, data visualization) and to properly prepare students for writing in upper-level course, including using LaTeX.

Section #1: Discipline Specific Writing Characteristics (what characterizes academic and professional communication in this discipline?)

Professional and academic mathematical and statistical writing is characterized by rigor, precision, clarity, and conciseness. Rigor means that arguments are completely supported by logical arguments, where at each stage logical steps/implications are clear (what this means will depend on the sophistication of the audience at which the writing/communication is aimed). Precision means that language is used precisely and consistently, that statements are unambiguous, that terms are properly defined, and that little or nothing is left "up to interpretation". Clarity means that the problems and objectives are precisely defined, and that an effort is made to structure arguments and statements in such a way as to make them as easy to understand (i.e., psychologically accessible) as possible. Furthermore, motivation, examples, counterexamples, diagrams, visualizations, and at times intuition is supplied where appropriate. Finally, conciseness means that only ideas serving to further the previous three aspects are presented, while unnecessary tangents, details, or prose should be avoided.

Writing in our disciplines should also be grammatically correct, at least to the extent that there is no ambiguity in the intended meaning. Here, grammar is understood in both the ordinary sense (i.e., Standard Academic English), as well as the mathematical sense, where in particular, this means that symbols and notation are used correctly. Furthermore, mathematical notation and symbols need to be correctly integrated within sentences/prose.

Another aspect is expository writing for a broader (sometimes non-technical) audience. Such writing requires translating technical terminology and ideas into accessible descriptions, and the skillful use of naive analogies to help convey the essence of these.

Additionally, math and statistics writing is often used to help clarify and synthesize one's own thinking. For example, writing about one's own ideas often helps to identify gaps or holes in the logic, and the act of summarizing and paraphrasing the ideas of others can help to solidify one's own understanding. Moreover, writing about topics that are only partially understood, or problems that are not yet solved, can help spark creativity and the formation of new ideas.

Statistical writing includes all the characteristics of mathematical writing, although at the undergraduate level there is less emphasis on proof writing. Good statistical writing also requires careful, precise interpretation of statistical analyses that is justified by the data and the particular methods used, including clear statements of any assumptions required and the checking of those assumptions when feasible.

Visualization of data and the results of analyses is also central to statistical writing. Visualization must provide easy-to-understand graphical summaries that communicate what is essential to answer the question or questions of interest to the relevant audiences and/or stakeholders. Good visualization must be paired with appropriate explanation and synthesis of what is being depicted.

Finally, mathematical and statistical writing also often includes numerical calculations and computer programming. For example, one must be able to implement statistical algorithms and perform statistical analyses, including data pre-processing and visualization. Also, any outputs of numerical simulations should be clearly described, explained, and interpreted. All code should be well documented, easy to follow, and reproducible.

Section #2: Desired Writing Abilities (With which writing abilities should students in this unit's majors graduate?) – Longitudinal approach to writing

Undergraduates in mathematics and statistics should develop strong problem solving skills and discipline-specific writing abilities throughout the course of the major. We divide these skills into four categories: Precision and Rigor; Exposition, Explanation, and Argumentation; Audience and Context; and Writing Tools. While most of the writing skills apply to all mathematics and statistics majors, several skills may only be encountered in specific tracks such as statistics or applied math.

Precision and Rigor

- ***Solve complex problems:*** Students should be able to make sense of a problem and determine the tools and techniques necessary to solve it. They should then be able to efficiently and accurately carry out the computations, follow the algorithms, or apply the proof techniques that are relevant.
- ***Use definitions and mathematical terminology correctly and precisely:*** Students should be able to use discipline-specific vocabulary and symbolic notation correctly, including mathematical definitions (e.g., "confidence interval" or "linearly independent") as well as mathematical quantifiers (e.g., "for all" or "there exists").
- ***Follow discipline-specific conventions for exposition:*** Students should use standard academic English grammar sufficiently well that there is no ambiguity in meaning, integrate mathematical expressions and equations into sentences, and follow standard proof-writing patterns.
- ***Support claims with valid statistical, numerical, and/or experimental evidence:*** Students should be able to rigorously formulate claims and support them by gathering or using data and analyzing that data using valid statistical methods.

Exposition, Explanation, and Argumentation

- ***Interpret computational results:*** Students should be able to interpret the meaning of a computation. For example, they should be able to state the significance of the confidence interval being a particular interval or the derivative having a particular value.
- ***Rigorously justify steps in a mathematical argument:*** Students should be able to write proofs that are clear and rigorous using a variety of proof techniques.
- ***Write for understanding:*** Students should be able to deepen and synthesize their own thinking and understanding by writing, including identifying gaps or holes in their logic. Students should appreciate writing as part of their mathematical process.
- ***Summarize and communicate computational results and statistical inferences:*** Students should be able to write reports that explain the full pipeline of data analysis, from data collection/sourcing through visualizations and recommendations for next steps, as well as the process through which any numerical results are produced and how they inform and supplement any theoretical results. They should be able to correctly interpret and explain statistical results, including limitations.

Audience and Context

- ***Provide audience and context specific background:*** Students should be able to adapt their communication style to their audience. They should give background information as appropriate, and adjust the level of technicality based on their audience's expertise.

- ***Explain technical subject matter to a general audience:*** Students should be able to communicate mathematical and statistical ideas to a broad audience, helping people to understand the significance and key points.

Writing Tools

- ***Typeset in LaTeX:*** Students should be able to typeset technical documents in LaTeX with formulas, mathematical symbols, and figures.
- ***Use relevant programming languages and mathematical software:*** Students should be able to use software and programming as appropriate to their specific needs for computational purposes. Students should be able to implement numerical methods, conduct statistical/numerical analyses, and visualize data & results in a reproducible way.

Section #3: Integration of Writing into Unit's Undergraduate Curriculum: (How is writing instruction and support currently positioned in this unit's undergraduate curriculum (or curricula)? What, if any, structural plans does this unit have for changing the way that writing and writing instruction are sequenced across its course offerings? With what rationale are changes proposed and what indicators will signify their impact?)

Current instruction and support of writing in our department

Current instruction and support of writing varies widely across courses and instructors. Many courses rarely mention "writing" and all but a few courses provide very little explicit instruction or feedback on writing. Several faculty have begun to provide more explicit writing instruction, such as providing rubrics, allowing for revisions, and providing feedback on ways to improve writing (e.g., in some sections of MA 111, 214, 293, 301, 341, 415, 511, 512, 586).

Our department's undergraduate curriculum is divided into two parts: lower division (100 & 200 level) and upper division (300+ level) courses. In the lower division courses, the writing typically consists of writing solutions to problems that are largely computational (in a pen-and-paper sense), in contexts such as weekly homework assignments, weekly in-class quizzes, and in-class exams. Additional examples of writing activities include online homework assignments that incorporate writing-focused questions (e.g., in MA 242), coding (e.g., in MA 213 & 214), lab reports (e.g., in MA 213 & 214), and projects (e.g., in MA 226). It is worth noting that there are a small number of faculty who teach lower-division courses in a way that addresses writing.

The upper-division courses exhibit much more variation in the types of writing instruction and assignments they contain. For example, many 500+ level math courses involve extensive amounts of proof-writing (e.g., MA 511, 512, 541, 542, 563, 564), while others continue to remain focused primarily on pen-and-paper computations (e.g., MA 561, 562, 573). In addition, there are courses that focus on numerical computation and analysis, and hence require extensive code writing (e.g., MA 539 and 565). Statistics courses contain a similar mix, as well as writing to convey the assumptions and interpretations of statistical analysis.

Beyond the support students receive during in-class time and from their instructors and TFs during office hours, students can also visit the Tutoring Room, a drop-in help center organized by our department and staffed with graduate student TFs. We note, however, that the Tutoring Room is not focused specifically on writing, and it is likely that the more writing-related aspects of student work are not emphasized in this setting.

One of the main aspects of our curriculum that should be noted within the context of writing is that the word "writing" is not often used when talking with students (except possibly when referring to writing proofs or writing code). We often do not refer to the pen-and-paper computations that students regularly do, for example, in calculus, as "writing". This is significant in that it makes students less likely to take writing-related skills they have developed in other contexts and transfer them to the context of their mathematics and statistics courses.

Recommendations for changes and supporting rationale

As a result of what is described above, a primary recommendation of the Writing Plan is to explicitly and intentionally discuss the work that students do in our courses *as writing*. This will come about through individual faculty interactions as they teach their courses, but also through messaging on our departmental website. A list of more specific recommendations follow.

Recommendations for 100-level courses:

1. Frame written work as a form of writing, specifically as very precise writing.
2. Correct usage of the equals (=) sign: discuss correct and incorrect usage in class; provide examples of correct and incorrect usage, ideally from past student work (appropriately anonymized; see Appendix B of this Writing Plan for examples); and allocate points on each exam for correct equals sign usage.
3. Choose problems for the online homework (e.g., in MyLab Math) that involve language (see Appendix E of this Writing Plan for examples).
4. Assess understanding through True/False questions that emphasize precision in writing.

Recommendations for 200-level courses:

1. Continue to frame written work as forms of writing, specifically as very precise writing. In addition, discuss the idea of using writing not just to communicate your ideas but as a way to understand and explore ideas, as well.
2. Correct usage of implication/assertion: discuss correct and incorrect usage of the arrow ($\Rightarrow$) symbol; provide examples of correct and incorrect usage, ideally from past student work (appropriately anonymized; see Appendix B of this Writing Plan for examples); and allocate points on each exam for correct arrow symbol usage.
3. Reinforce correct usage of the equals sign: this will be review for students who took 100-level MA courses at BU but could be new for incoming students who go straight to 200-level MA courses; continue to allocate points on each exam for correct equals sign usage.
4. Continue to choose problems for the online homework (e.g., in MathLab) that involve language, particularly true/false questions (see Appendix E of this Writing Plan for examples).
5. In MA 213 & 214, frame coding as writing and use coding to understand and explore ideas. Emphasize standard phrasing for discussing and interpreting statistical techniques such as confidence intervals and hypothesis tests.

Recommendations for upper-division courses:

1. Continue to frame assignments as forms of precise writing and continue to emphasize that mathematical and statistical writing serves not just as a way to communicate our ideas but also as a way to explore and clarify our own thinking.
2. Reinforce correct usage of the equals sign and arrow symbol, and discuss the ideas of equality, implication, and assertion more generally. Discuss the usage of words like assume, implies, therefore, there exists, for all, etc.
3. Include more intentional writing-focused activities and assignments, with details dependent on the course. For examples, see Appendix E of this Writing Plan.
4. Reinforce correct descriptions of statistical analyses such as point estimation, confidence intervals, and hypothesis tests and train students in how to write about more complex and sophisticated statistical techniques.

5. Provide instruction and assessments on tools for reproducible research and analysis such as RMarkdown, Quarto, and Jupyter notebooks.

Recommendations for departmental resources:

1. Asynchronous, flipped module that students can use to learn LaTeX; require students to complete this module independently before their first upper-division course.
2. Examples of good, and not so good, writing from students in our courses at various levels (appropriately anonymized).
3. Templates for including more writing into our courses; Examples that can be found in the Appendix E of this Writing Plan include
   - Example syllabi (MA 512).
   - Example rubrics for writing assignments (rubric for written proofs used in MA 511 and MA 512).
   - Example activity used to teach MA 512 students LaTeX.
4. Trainings for PhD student Teaching Fellows regarding all of the above, so that during discussion sections, office hours, and when working in the tutoring room, they can better support the writing-focused aspects of our students' learning.

Another recommendation that was discussed was to create versions of MA 225 and MA 242 for majors that provide writing instruction to explicitly help students transition to upper-division courses. It is not clear what the best title or label for these courses should be, so for now, please consider the term “for majors” to be temporary and only for the purposes of this document. This way of structuring the undergraduate curriculum is used at many other universities, such as Boston College, Carnegie Mellon, Cornell, UPenn, UC Irvine. Such courses would be required for our majors, with appropriate exceptions given to students who join the major late but may be open to non-majors as well. They could be more rigorous, more language-intensive, and intentionally introduce students to the concepts of proof and proof-writing.

A final, longer-term recommendation discussed was to create a new 400-level “Tools for Statistics” course that would be required for all statistics majors. This course would cover numerous aspects of writing, including programming, advanced visualization, and LaTeX.

Assessment of impact

In Section 5, we ask for support for faculty course releases and stipends to carry out the recommendations in this plan. Faculty who receive such support would be expected to conduct surveys of department faculty after their work is completed to determine the extent to which the recommendations are being implemented and the impacts faculty feel they are having in student learning. For example, the surveys could include questions regarding the extent to which shared resources are being utilized, the extent to which faculty teaching lower division courses are implementing the corresponding recommendations, and the extent to which faculty teaching upper division courses feel the students are prepared for those courses.

Section #4: Assessment of Student Writing (How does this unit currently communicate writing expectations – see sections #1/#2 – to undergraduate students? What do these expectations look like when they are translated into ratable criteria? How satisfied is the unit faculty that students are adequately familiar with these expectations? How satisfied is the unit faculty that student writers successfully meet identified expectations by the time they graduate? Why? If less than satisfied, what plans does the unit propose for closing the gap?)

Currently, individual faculty communicate the writing expectations for their courses to students with no department-wide coordination. Most math classes assess students based on (bi)weekly problem sets and several exams, and "writing" is not typically emphasized. These assignments can be graded by undergraduate graders, graduate student TFs, and faculty. Giving meaningful feedback is a challenge because of the lack of person-power (and undergraduate graders and even some graduate TFs lack of expertise). In addition, even if feedback is given, faculty have the sense that because most classes don't have a revision process, feedback is not internalized. Notably many upper-level statistics courses include a final project, but these tend not to include formal opportunities for feedback and revision.

We propose several strategies to address these challenges, mostly around shared department resources:

- Adopt a statement on writing for the department that lays out our goals and expectations of students in all levels of classes, from 100 to 500 level, including a policy on the use of AI.
- Create a shared class in Gradescope with sample rubrics for grading large lower-level classes.
- Create model rubrics for writing assignments in upper-level classes at various levels of specificity.
- Train Teaching Fellows on assessing mathematical writing and providing feedback on assignments.

Section #5: SUMMARY of IMPLEMENTATION PLANS and Requested Support: (Based on above discussions, what does the unit plan to implement during the period covered by this plan? What forms of instructional support does the unit request to help implement proposed changes? What are the expected outcomes of named support? What kinds of assessment support does this unit request to help assess the efficacy of this Writing Plan? What are the expected outcomes of this support?

To effectively teach mathematical writing, the department must make pedagogical and curricular changes that allow for more individual attention and feedback for our students. This section outlines the staffing and compensation support necessary to put the committee's recommendations into effect. We are unable to provide a timeline for these changes because they are not possible until we receive the requested support.

The rationale for the below recommendations is in part based on the extent to which our department has increased our reliance on large lecture courses in recent years, and the extent to which we are often understaffed in terms of having regular faculty who are available to teach these courses.

For example, in AY19/20, in order to accommodate increased enrollment in MA 242 (Linear Algebra), our department moved from small, 25-student sections, to large, 100+ student sections (see the spreadsheet "Historical Enrollments in Selected Undergraduate Math & Statistics Courses" in Appendix A). Although this change was necessary given the departmental resources we had at the time, it was not in the best interests of our students.

For these reasons, in order to carry out the recommendations from the preceding pages, we request support to help enable faculty to create shared departmental resources, to develop TF training materials so that our TFs can better support faculty in achieving writing-related goals in their courses, and to carry out course redesigns and development of a new course that collectively will substantially improve the teaching of mathematical and statistical writing skills.

Creation of shared departmental resources

The recommendations for the creation of shared resources are intended as an efficient way to positively affect many students' experiences while minimizing the increased workloads on individual faculty who teach large lecture courses, and while being accessible to anyone who teaches courses in our department, even if they are not regular faculty.

The shared resources that were mentioned on the previous pages of this Writing Plan include the following:

- *Statement on Writing:* This would be a public statement that would be posted on our departmental web page which *(i)* defines what "writing" means in our department; *(ii)* lists associated learning goals and expectations that we have for our undergraduate students at different stages of our majors; and *(iii)* lists resources for students to help support them in achieving these goals.
- *Genre models:* A list of examples of writing from a variety of genres that undergraduates might encounter during their time at BU. Examples include

- Written solutions to problems from lower division courses (calculus, statistics, differential equations, linear algebra)
- Written lab reports from statistics courses
- Written projects from differential equations courses
- Written projects from upper and lower-level statistics courses
- Examples of numerical code from upper-level applied mathematics courses
- Examples of proof writing from courses such as analysis and algebra
- Examples of senior theses

- *LaTeX Learning Module:* Asynchronous, flipped module that students can use to learn LaTeX; require students to complete this module independently before their first upper-division course.
- *Model rubrics:* A collection of rubrics from a variety of courses and at varying levels of specificity.
- *Shared Gradescope rubrics* through a Gradescope class that all instructors can join. These rubrics could be copied and modified for specific courses.
- *Example Syllabi:* from courses that explicitly incorporate writing-focused assignments and activities.
- *Frameworks for Providing Feedback on Writing:* Suggestions for division of labor between grader/TF/instructor when providing feedback on writing assignments.
- *Activities and assignments that explicitly involve writing:*
    - Discussions of essays like "Good Mathematical Writing" by Francis Su.
    - Opportunities for revision of written work, including frameworks for providing associated feedback.
    - Activities where students must spot the error in a proof, or more generally assess the quality of a proof.
- *Policy on AI*: Incoming first-year students are being told that each department has its own policy on the use of AI in student work. Our department should create such a policy; for example, we could choose to adopt or modify the Generative AI Assistance Policy (GAIA) that was created by CDS.

Support requested: In order to enable faculty members to develop these shared resources, we request one course release for each of two faculty members, one who will focus on resources related to statistics courses and one who will focus on resources related to mathematics courses.

Creation and implementation of TF trainings

The majority of undergraduates who take courses in our department do so in large, introductory courses where the faculty instructors are supported by graduate student teaching fellows. Therefore, any plans to incorporate more writing-related support in such courses must also include associated training for the TFs. Such training could be focused on how TFs use time in discussion sections, on how TFs provide feedback on written student work, and on how TFs assist instructors with grading and assessment. This could impact not just TF-lead discussion sections and grading, but also the departmental tutoring room, which is staffed by TFs.

The creation of such training materials is, in and of itself, a large undertaking. In addition, it must be accompanied by time to deliver and discuss such training with new TFs. For example, we imagine offering a TF training once per academic year during the fall semester. It could be more of a short course, consisting of several days of full-day workshops at the beginning of the semester, or it could be a 2-credit course that runs regularly throughout the duration of the fall semester. Such details would be determined by the faculty member who develops the training.

Support requested: In order to enable a faculty member to develop these training materials, we request one course release for one faculty member. In addition, we anticipate that ongoing resources would be needed to support the faculty member who runs the training each semester, for example a stipend or the incorporation of such a training into their teaching obligations.

Redesign of MA225 and MA242 into versions “for majors”

MA 242, along with MA 225, are key courses in our curriculum, and the course redesign we propose is intended to reflect that, and to allow for corresponding levels of attention and support to students in these courses.

Through talking with colleagues who teach upper-division courses, particularly those that involve extensive writing (including, but not limited to proof writing), we have heard that many students enter these courses underprepared. Our impression is that one major factor that contributes to this is that there is a significant gap in the writing-related expectations placed on students in lower division courses compared to those placed on them in upper division courses.

To address these challenges, we suggest redesigning two key 200-level courses: multivariable calculus (MA225) and linear algebra (MA242). These courses are often the lower division courses that students take just before moving to upper division courses. Moreover, they could be redesigned in a way that would help bridge the significant gap in writing-related training and expectations in our current curriculum.

The redesign of multivariable calculus and linear algebra into versions for majors would be an extensive undertaking. The rationale for doing so was discussed in Section 3 of this Writing Plan, and through this redesign students in the for majors sections would receive much more writing-focused instruction and feedback to prepare them for upper-level mathematics. We also note here again that it is not clear what the best title or label for these courses should be, so for now consider the term "for majors" to be provisional and for the purposes of this document only. Furthermore, this label is not meant to imply that these courses would only be open to majors. Rather, majors would be required to take these versions of the courses (with exceptions granted as needed, for example, if a student joins the major after taking another version of the course, or is a transfer student). Other students would be allowed to take them, with the understanding that these courses would include more writing-focused instruction, in addition to the usual content-based instruction.

The exact details of how such a redesign would take place should be the subject and work of a future committee that is focused specifically on this task. We include here some aspects of such a redesign that have been discussed among members of the Writing Plan committee and that members of any future redesign committee may wish to consider.

The committee might wish to do one course at a time – for example first focus on MA 242 and then, if things go well, subsequently redesign MA 225 – or do both at once, which could allow for the possibility of combining MA 225 and MA 242 into a two-course sequence that would be taken together and taught in a coordinated way. A combined sequence could be based on one of the many existing books that combine multivariable calculus and linear algebra in this way (see, for example, "Vector Calculus, Linear Algebra and Differential Forms: A Unified Approach," by Hubbard and Hubbard).

The versions of these courses for majors need not replace the current honors versions of courses, but we encourage future decisions regarding this to take into account equity issues regarding the amount of resources that are allocated to individual students. For example, under the current system, our honors courses tend to be quite small (e.g., MA 230 and MA 442 tend to have enrollments of 5-15 students), whereas the non-honors versions tend to be quite large (e.g., MA 225 and MA 242 tend to have enrollments of 125 students per section). This effectively means that the students who need the least support to achieve a reasonable level of understanding are given the most support in doing so. If versions of MA 225 and MA 242 for majors are created in the future, they could be small sections (e.g., 25 students each) and a concurrent mechanism could be developed for encouraging certain students to enroll in the honors section, so all of our majors receive more equitable levels of support in their learning processes.

Support Requested: We request one course release for each of 4 faculty members who will comprise the committee that will focus on redesigning these courses. We also request that three new teaching postdoctoral positions be created in our department. We envision these postdocs will teach two courses per semester and also contribute to departmental shared teaching resources.

The rationale for requesting three such postdoctoral positions is *(i)* so there is a critical mass of teaching postdocs to enable a sense of community and also to allow for collaboration and the sharing of ideas and resources; *(ii)* to support the additional course load due to increased enrollment numbers in MA 225 and MA 242 (see the spreadsheet Course Enrollments by Major in Appendix A); and *(iii)* to make our teaching of these subjects to our majors more equitable. Based on the data from Fall 2023 and Spring 2024, we expect to need places for 100 majors per semester in each of MA 225 and MA 242. In order to provide quality feedback to students in these courses, related to but not limited to their writing, we anticipate capping enrollments at 25 students per course. Therefore, we anticipate replacing one large section of each of MA 225 and MA 242 each semester with four 25-student sections (while keeping another large section as-is for non-majors). This would increase the number of associated faculty course assignments from 2 each semester to 8 each semester. The teaching postdocs would cover 6 of these assignments, whereas regular faculty would cover the other two, as well as serve as mentors for the teaching postdocs. (We also note this would free up two teaching fellows each semester from these large courses, who could then serve the needs of other courses, thus helping us address our current shortage of teaching fellows.) Teaching all of our majors in smaller sections will help to provide similar levels of support to all of our majors and hence be more equitable.

Develop 400-level "Tools for Statistics" course

The "Tools for Statistics" course (the name is tentative) would be modeled on MA 415 ("Data Science in R"). However, unlike MA 415, it would assume students have taken MA 213, MA 214, and, possibly, calculus and/or linear algebra. At the earliest, this new course would be developed starting in Summer 2026, after substantial work on the overhaul of MA 213/214 is complete. The MA 213/214 overhaul is being done through DL&I's Large Course Transformation program and is tentatively scheduled for

Summer 2024–Fall 2027. The overhauled versions of MA 213/214 will introduce students to R programming, so it is important to develop "Tools for Statistics" as a follow-on that appropriately augments what the new versions of MA 213/214 cover. While the exact topics for "Tools for Statistics" will be determined during the course development process, some likely topics include:

- More advanced R programming skills
- Introduction to object-oriented programming
- Principles and tools for reproducibility
- Advanced data visualization and data cleaning/pre-processing techniques
- Basics of scientific programming (e.g., floating point arithmetic, random number generation, numerical linear algebra)
- Communicating about statistical concepts to varying audiences
- Introduction to LaTeX

We expect this new course would have the additional benefit of alleviating over-enrollment in MA 415.

Support Requested: We request one course release for each of two faculty members who will develop the course materials. We also request that one new teaching postdoctoral position be created in our department to compensate for the addition of a new course, which would need to be taught twice per year to ensure enrollment is manageable. We envision that this postdoc will teach two courses per semester and also contribute to departmental shared teaching resources.

Total of all support requested

Course releases: 9 total courses, for the faculty members who develop the shared resources (2), the TF training (1), the new versions of 225 and 242 (4), and the new Tools for Statistics course (2).

Teaching postdoctoral positions (ongoing): 4 total, to staff the redesigned versions of 225 and 242 (3) and the new Tools for Statistics course (1).

Note: We anticipate needing additional ongoing compensation for whomever runs the TF training each year.

Section #6: Process used to create this Writing Plan (How and to what degree were stakeholders in this unit (faculty members, instructors, affiliates, teaching assistants, undergraduate students, others) engaged in providing, revising, and approving the content of this Writing Plan?)

The Math and Statistics Writing Plan was developed and written over the course of AY 2023/2024 through an intensive collaboration between the Math and Statistics Department and Writing in the Disciplines.

<u>Math and Statistics Faculty Team:</u>
Glenn Stevens (Department chair)
Margaret Beck (Math faculty)
Jonathan Huggins (Statistics faculty)
Li-Mei Lim (Math faculty)
Dan Sussman (Statistics faculty)
Matt Szczesny (Math faculty)

<u>Writing in the Disciplines Team:</u>
David Shawn (WID Associate Director)
Jessica Kent (WID Consultant)

These groups coordinated in numerous ways, including but not limited to the following: planning sessions within the WID team; planning and debriefing meetings with Math and Statistics Department chair Glenn Stevens, Math faculty member Margaret Beck, the WID team, and Interim Associate Dean of Undergraduate Academic Programs and Policies Tereasa Brainard; and regular meetings with the Math and Statistics faculty team and the WID consultant, Jessica Kent (see Appendix F for meeting dates).

This group met biweekly throughout the year, for a total of thirteen meetings. Throughout the fall, the Math and Statistics Faculty Team, with facilitation from the WID consultant, reflected on the qualities of strong math and statistics writing at the professional and undergraduate levels, reviewed their current writing teaching and assessment practices, identified opportunities for improvement, chose action items to address each one, and identified the support they would need for implementation. On February 20, 2024, the committee led a full faculty meeting to discuss their proposals and solicit feedback. The meeting was well attended, with about forty faculty members present, representing both math and statistics. In early Spring 2024, the team voted on which changes to implement and spent the remainder of the Spring semester drafting the Writing Plan. Each element of this Writing Plan was explored in depth during the various meetings, and committee members volunteered for teams responsible for composing each section. The group then met to revise each section draft collaboratively until they reached consensus.

This group sought input from other stakeholders in several ways, including a survey of all Math and Statistics undergraduate majors, a survey of all full-time Math and Statistics faculty, and the in-person spring full faculty meeting. See Appendix C for the findings of the student survey, Appendix D for finding of the faculty survey, and Appendix G for the faculty meeting slides. The students primarily provided insight into their needs and experiences, while faculty provided feedback on current writing instruction and intermediate drafts of the plan. Special thanks to Jennifer Balakrishnan, Luis Carvalho,

Uri Eden, Sam Isaacson, Ranjan Panth, David Rohrlich, and Konstantinos Spiliopoulos for their feedback on drafts.

Section #7: Briefly describe the ways that the ideas contained in this Undergraduate Writing Plan address the University's Writing Intensive Learning Outcomes?

Our goal is to provide our students with the tools to effectively communicate mathematical ideas and arguments to both expert and non-technical audiences.

Currently, MA 586 Stochastic Methods for Algorithms and MA 301 Writing in Mathematics are both designated as fulfilling a Writing Intensive Course HUB unit. Our current plan does not explicitly suggest adding the Writing Intensive Course HUB unit to other courses. However, this will be a strong consideration as we revise existing courses and develop new courses with more intensive writing instruction. Additionally, regardless of HUB status, this plan addresses all three writing-intensive learning outcomes, as outlined below.

Students will be able to craft responsible, considered, and well-structured written arguments, using media and modes of expression appropriate to the situation.

The concepts of precision and rigor in mathematical writing echo the idea of "responsible, considered, and well-structured written arguments". Using correct notation, terminology, and conventions ensures that the broad community of practitioners of various quantitative disciplines will understand a student's writing. By embracing document creation such as LaTeX, RMarkdown, and Jupyter notebooks, students can disseminate their writing in a large set of clear, reproducible formats used throughout various disciplines.

Students will be able to read with understanding, engagement, appreciation, and critical judgment.

Reading is a key component of all of our courses. To begin, assigned readings specifically about writing, such as "Some Guidelines for Good Mathematical Writing" by Francis Edward Su, will help the students understand what to look for when assessing and appreciating writing in the discipline.

In many courses, students will read mathematical proofs and, in class, discuss the formal and structural elements of the proofs, as well as the strengths and weaknesses of the proofs. Courses following our plan will incorporate the comparison of proofs from different students and sources to better understand how to incorporate concision and clarity into their own writing. The plan also recommends critical evaluations of peer work, including proofs, statistical analyses, and expository writings. Advanced courses also ask students to read academic literature. Many courses, especially statistics courses, ask students to evaluate statistical arguments or expository writing about mathematics in news articles and scientific reports.

Students will be able to write clearly and coherently in a range of genres and styles, integrating graphic and multimedia elements as appropriate.

The genres of writing in mathematics and statistics range from problem-set solutions to full academic articles to expository writings aimed at more general audiences. Most student's first exposure to writing in the discipline will be in writing solutions to problems where our plan emphasizes the use of correct notation, grammar, and terminology. Later, many students will work on writing shorter proofs of various results in courses such as analysis and algebra. Our plan also incorporates reflective writing to explain one's thought process and difficulties encountered in understanding new concepts and solving problems.

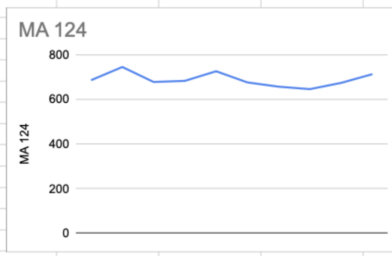
In many courses, students are asked to write reports that detail the data, methods, results, and conclusions of computational experiments and statistical analyses. This writing will often include either within the document or, as a supplemental, the code used to generate the result. Additionally, most of such writing will include visualizations that provide insight into methodology and concepts, and the results of the computations. Finally, these writing assignments and other expository writing assignments will train students to explain the roles of mathematics and statistics in real-world situations and other disciplines.

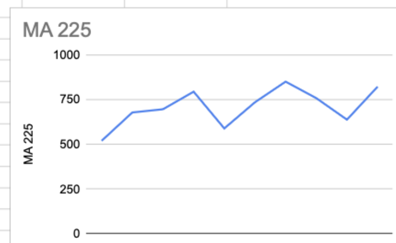
Appendix A: Mathematics and Statistics Departmental Data

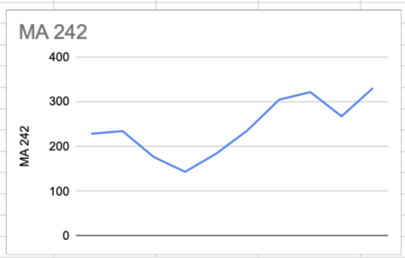
Contents:

Historical Course Enrollment Numbers

Course	AY 13-14	AY 14-15	AY 15-16	AY 16-17	AY 17-18	AY 18-29	AY 19-20	AY 20-21	AY 21-22	AY 22-23	Ave honors class size
MA 123	655	618	640	516	610	522	474	564	744	719	
MA 124	685	744	677	682	725	675	656	645	673	712	
MA 213	214	206	211	188	201	198	215	202	228	256	
MA 214	104	115	115	111	130	118	145	150	151	145	
MA 225	519	677	695	793	588	734	849	757	637	821	
MA 226	501	514	531	499	530	522	542	574	550	571	
MA 230	16	0	7	8	10	0	5	0	8	0	9
MA 231	13	10	10	0	15	9	14	15	11	8	11.66666667
MA 242	228	234	176	143	184	236	304	321	267	330	
MA 412	46	67	56	62	66	56	48	60	65	71	
MA 442	19	11	12	14	13	13	20	8	7	17	13
MA 511	32	41	32	42	36	55	27	39	42	48	
MA 541	21	0	28	0	0	38	16	30	22	31	
MA 575	51	81	74	91	105	124	126	146	111	141	
MA 581	78	169	178	160	179	202	183	205	204	180	
MA 582	45	61	84	88	93	171	94	131	108	127	
MA 583	45	59	50	54	56	61	63	50	58	59	
MA 585	38	23	44	39	59	34	64	36	31	58	
Notes							**Beginning of larger sections of 242**				





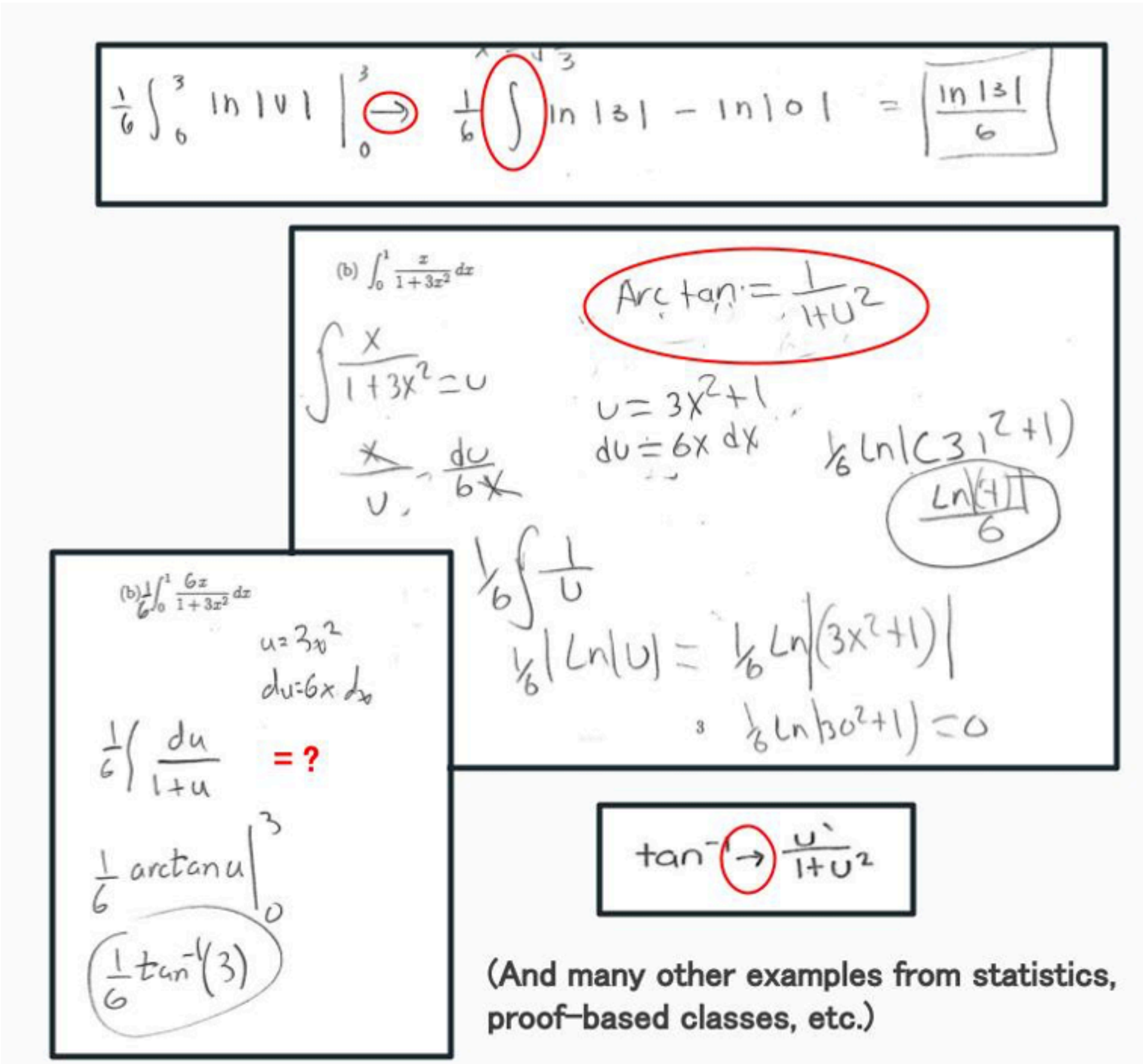


Course Enrollments by Major

2023/2024	Fall '23 MA225	Sprg '24 MA225	Fall '23 MA226	Sprg '24 MA226	Fall '23 MA242	Sprg '24 MA242
1) math/stats major (pure math, applied math, or stats)	25	31	14	37	23	32
2) joint major math + CS	23	18	1	4	16	16
3) joint major math + philosophy	1	1	0	0	1	1
4) joint major math + econ	39	21	15	24	27	37
5) joint major stats + CS	10	4	1	1	6	3
6) joint major math + math ed	0	1	1	0	0	0
7) joint major math + physics	2	2	3	2	0	0
Total Enrollment - MA Majors	100	78	35	68	73	89
Total Enrollment - Overall	512	290	215	337	174	180
Other college breakdowns:						
ENG	303	109	144	226	1	3
QST	6	14	3	9	12	6
SAR	1	1	0	0	1	2
BUA	2	0	6	3	10	6
CAS - not including math majors	70	52	20	13	49	59
CDS	5	1	0	1	4	1
CFA	1	2	0	0	1	0
CGS	22	22	3	16	16	7
COM	2	4	1	0	2	2
MET	0	3	0	0	2	3
Other	0	4	3	1	1	0

Appendix B: Student Writing Examples

Contents:


= ?
(And many other examples from statistics, proof-based classes, etc.)

MA 242: Linear Algebra Midterm

MA 242 – Linear Algebra
Midterm #1

[redacted]

Instructions: To receive full credit you must show all work. Explain your answers fully and clearly. You may refer to theorems/facts in the book or from class. No calculators, books or notes of any form are allowed. Please make sure your cell phone is turned off and put away. Good luck!

Question	Score	Out of
1		16
2		8
3		12
4		18
5		12
6		12
7		24
Total		102

1. **(16 points)**
Let

$$A = \begin{bmatrix} 1 & -2 & -1 & 1 \\ 2 & -2 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

$Ax = 0$

$x_0 \left[\vdots \right] + \dots + x_n \left[\vdots \right]$

- **(a) (8 points)** Find the general solution to the homogenous system $Ax = 0$ in parametric vector form.

$$\left(\begin{array}{cccc|c} 1 & -2 & -1 & 1 & 0 \\ 2 & -2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & -2 & -1 & 1 & 0 \\ 0 & 2 & 2 & -1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{array}\right)$$

$R2 \to R2 - 2R1$

$$R2 \to \tfrac{1}{2}R2 \sim \left(\begin{array}{cccc|c} 1 & -2 & -1 & 1 & 0 \\ 0 & 1 & 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & -2 & -1 & 1 & 0 \\ 0 & 1 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 3/2 & 0 \end{array}\right)$$

$R3 \to R3 - R2$

$$R3 \to \tfrac{2}{3}R3 \sim \left(\begin{array}{ccccc} 1 & -2 & -1 & 1 & 0 \\ 0 & 1 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

basic: x_1, x_2, x_4

free: x_3

general solution:

$X_1 = 2x_2 + X_3 - X_4$

$X_2 = -X_3 + \frac{1}{2}x_4$

$X_3 = X_3$

$X_4 = 0 + 0x_5$

$$X_1 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + X_2 \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} + X_3 \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + X_4 \begin{bmatrix} -1 \\ 1/2 \\ 0 \\ 0 \end{bmatrix}$$

- (b) **(4 points)** Are the columns of the matrix A from part (a) linearly independent ? If not, find an explicit dependence relation between them. Explain your reasoning.

$$\begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array}$$
$$\begin{bmatrix} 1 & -2 & -1 & 1 \\ 2 & -2 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

Because when this matrix was row reduced to echelon form, there was not a pivot in every column, the columns of this matrix are linearly dependent.

Explicit dependence relation:

$$\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$
$$V_1 + V_2 = V_3$$

- (c) **(4 points)** Let

$$T : \mathbb{R}^4 \to \mathbb{R}^3$$

be the linear transformation $T(x) = Ax$ where A is the matrix from (a). Is T onto ? Is T one-to-one ? Explain

— T is onto because the matrix from (a) has a pivot in every row which means all $b \in \mathbb{R}^3$ can be obtained (aka. columns of A span $\{\mathbb{R}^3\}$ under this transformation)

— T is not one-to-one because the trivial solution for $T(x)=0$ is not the only solution; moreover, there is not a pivot in every column.

2. **(8 points)**
Consider the matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 4 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & 0 \end{pmatrix}.$$

Find the products AB and BA.

$(2\times3)(3\times2)$ ✓ 2×2 $(3\times2)(2\times3)$ ✓ 3×3

AB:

$$\begin{pmatrix} 1 & 2 & 3 \\ -1 & 4 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1+4-3 & 1+6 \\ -1+8-1 & -1+12+0 \end{pmatrix} = \begin{pmatrix} 2 & 7 \\ 6 & 11 \end{pmatrix}$$

BA:

$$\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ -1 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1-1 & 2+4 & 3+1 \\ 2-3 & 4+12 & 6+3 \\ -1+0 & -2+0 & -3+0 \end{pmatrix} = \begin{pmatrix} 0 & 6 & 4 \\ -1 & 16 & 9 \\ -1 & -2 & -3 \end{pmatrix}$$

3. **(12 points)** Find the inverses of the following matrices

- (a) **(6 points)** Determine for what values of h the matrix

$$A = \begin{pmatrix} 1 & -4 \\ h & -2 \end{pmatrix}$$

is invertible.

$$\frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$\det(A) = ad - bc$

$\det(A) \neq 0$

$1(-2) - (-4h)$

$-2 + 4h = 0$

$4h = -2$

$h = -\frac{1}{2}$

h is invertible when

$h \neq -\frac{1}{2}$

- (b) **(6 points)** Find the inverse of the matrix

$$B = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}$$

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 3 & 1 & 0 & 0 & 1 \end{array}\right) \overset{R3 \to R3 - 3R2}{\sim} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -3 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -2 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -3 & 1 \end{array}\right)$$

$R1 \to R1 - 2R2$

$$B^{-1} = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix}$$

$B B^{-1} = I$

$$\begin{pmatrix} 1 & \underbrace{-2+2+0}_{=0} & 0 \\ 0 & 1 & 0 \\ 0 & \underbrace{3+-3}_{=0} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

m rows
n cols

4. **(18 points)** Find the standard matrix of each of the following linear transformations.

(a) **(6 points)** Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $T(x_1, x_2, x_3) = (x_1 + 2x_2 - x_3, 5x_1 - x_3)$

$T(e_1) = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

$T(e_2) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$T(e_3) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ $\quad T(x_1, x_2, x_3)$

$x_1 + 2x_2 - x_3$

$5x_1 - x_3$

$\mathbb{R}^2$

standard matrix $\begin{pmatrix} 1 & 2 & -1 \\ 5 & 0 & -1 \end{pmatrix}$

(b) **(6 points)** The linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$T(e_1 - e_2) = \begin{pmatrix} 4 \\ -2 \end{pmatrix} \quad \text{and } T(e_1 + e_2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \text{where } e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$T(e_1) - T(e_2) = \begin{pmatrix} 4 \\ -2 \end{pmatrix}$

$T(e_1) + T(e_2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

$\hookrightarrow T(e_1) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} - T(e_2)$

$\hookrightarrow \begin{pmatrix} 0 \\ 2 \end{pmatrix} - T(e_2) - T(e_2) = \begin{pmatrix} 4 \\ -2 \end{pmatrix}$

$\begin{pmatrix} 0 \\ 2 \end{pmatrix} - 2T(e_2) = \begin{pmatrix} 4 \\ -2 \end{pmatrix}$

$-2T(e_2) = \begin{pmatrix} 4 \\ -2 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

$-2T(e_2) = \begin{pmatrix} 4 \\ -4 \end{pmatrix}$

$T(e_2) = -\frac{1}{2}\begin{pmatrix} 4 \\ -4 \end{pmatrix}$

$T(e_2) = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$

$T(e_1) + T(e_2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

$T(e_1) + \begin{pmatrix} -2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

$T(e_1) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} - \begin{pmatrix} -2 \\ 2 \end{pmatrix}$

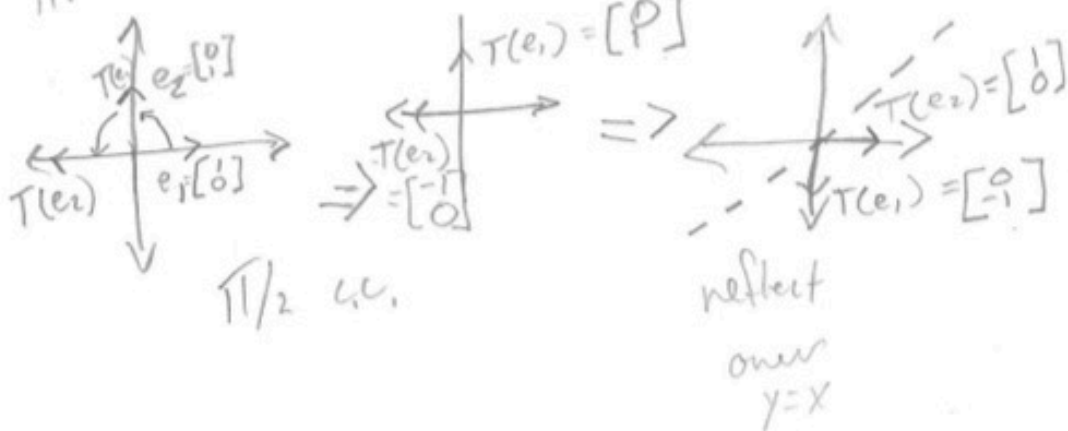
$T(e_1) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

standard matrix: $\begin{pmatrix} 2 & -2 \\ 0 & 2 \end{pmatrix}$

$= 90°$

(c) **(6 points)** The linear transformation $T : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ defined by first rotating by $\pi/2$ radians counterclockwise and then reflecting in the line $y = x$.

$\mathbb{R}^2$



Standard matrix: $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

$X_1 + X_2$
$X_2 + 3$

5. **(12 points)**

- (a) **(6 points)** Let

$$T : \mathbb{R}^2 \to \mathbb{R}^2$$

be given by the formula

$$T(x_1, x_2) = (x_1 + x_2, x_2 + 3)$$

Determine if T is a linear transformation. Explain your reasoning.

$T(e_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$T(e_2) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

~~T is linearly dependent because~~ standard matrix: $\begin{array}{cc} V_1 & V_2 \\ \end{array}$ $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$V_1 \, V_2$

T is not a linear transformation because T(x) does not pass through the origin. expression $X_2 + 3$ raises the line.

$$X_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + X_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 3$$

- (b) **(6 points)** Does there exist a 3×3 invertible matrix whose columns add up to 0 ? If yes, give an example of such a matrix, and if not, explain clearly why such a matrix does not exist.

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right]$$

Because the columns $V_1 + V_2 + V_3 = 0$ the matrix cannot be invertible. This is because values form zeros in columns as row operations are performed making it impossible to obtain the identity matrix on the left side.

6. **(12 points)**

(a) **(6 points)** Let A be a 2×2 matrix and $b \neq 0 \in \mathbb{R}^2$. Suppose that the solution set of the linear system $Ax = b$ consists of the line $2x_1 + x_2 = 3$

1) • Determine the number of pivots of A. Explain your reasoning.

2) • Give an explicit equation for the solution set of the homogenous system $Ax = 0$, and a vector that spans this solution set. Explain your reasoning.

$b \neq 0 \in \mathbb{R}^2$

$2x_1 + x_2 = 3$ solves $Ax = b$

1) Because $b \neq 0$, we know not all b can be obtained. Thus, we know at most 1 pivot exists. A cannot have 0 pivots otherwise $b = 0$ is a solution.

$\therefore$ A has 1 pivot

$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ or $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$

2) $A = [V_1\ V_2]$ $\quad x_1 \begin{bmatrix} 2 \end{bmatrix} + x_2 \begin{bmatrix} 1 \end{bmatrix} = 3$ $\quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$[2\ 1 \mid 3] \to V_h = V_1 - V_2$

$\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$ $\begin{bmatrix} 1 & 1 \\ -2 & 2 \end{bmatrix}$ $x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$V = V + t V_h$ $\quad V_h$ = homogeneous solution

Because $b \neq 0 \in \mathbb{R}^2$ $Ax = 0$ can't work → sort of confused

$2x_1 + x_2 = 3$

(b) **(6 points)** Let

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \text{ and } v_3 = \begin{pmatrix} 6 \\ h \\ 6 \end{pmatrix}$$

For what values of h is v_3 in the span of $\{v_1, v_2\}$? Explain.

$$\left[\begin{array}{cc|c} 1 & 1 & 6 \\ 2 & 0 & h \\ 1 & 2 & 6 \end{array}\right] \xrightarrow{R2 \to R2-2R1} \left(\begin{array}{cc|c} 1 & 1 & 6 \\ 0 & 0 & h-12 \\ 1 & 2 & 6 \end{array}\right) \xrightarrow{R3 \to R3 - R1} \left(\begin{array}{cc|c} 1 & 1 & 6 \\ 0 & 0 & h-12 \\ 0 & 1 & 0 \end{array}\right)$$

Swap $R2 \leftrightarrow R3$ $\sim \left(\begin{array}{cc|c} 1 & 1 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & h-12 \end{array}\right)$

$h - 12 = 0$ $\quad$ or $\quad 6V_1 = V_3$

$h = 12$ $\quad 6\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 12 \\ 6 \end{pmatrix}$

8

7. **(24 points)** Circle either TRUE or FALSE. No justification is needed.
Each question is worth 3 points

(a) If A is a 2×3 matrix, then the system $Ax = 0$ must have a non-trivial solution.

$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

☐ TRUE ☐ FALSE

(b) Every linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^4$ is one-to-one.

☐ TRUE ☐ FALSE

(c) If the set of vectors $\{v_1, v_2, v_3\}$ is linearly independent, then so is $\{v_1, v_3\}$

☐ TRUE ☐ FALSE

(d) If the set of vectors $\{v_1, v_2, v_3\}$ is linearly dependent, then v_3 is a linear combination of v_1 and v_2

it could be that V_1 is a combo of V_2 and V_3

does not need to be V_1 and V_2 form V_3

☐ TRUE ☐ FALSE

(e) If A, B are square matrices such that AB is invertible, then A and B are both invertible.

☐ TRUE ☐ FALSE

(f) If A is a 3×5 matrix, and the general solution of $Ax = 0$ has exactly two free variables, then $Ax = b$ is consistent for every $b \in \mathbb{R}^3$.

3 pivots in row

pivot in every row

☐ TRUE ☐ FALSE

(g) If A is a square matrix with two identical rows, then A is not invertible.

$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ $\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$

☐ TRUE ☐ FALSE

(h) There exists a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ that is one-to-one but not onto .

square $\therefore$ $1\text{-}1 \iff$ onto

☐ TRUE ☐ FALSE

MA 415/615 Final Project

Directions and a rubric for the final group project (which are revised and updated each semester) are found at https://sussmanbu.github.io/ma4615-final-project-quarto/.

Example Projects

- https://sussmanbu.github.io/final-project-group-13/
- https://sussmanbu.github.io/final-project-team5/
- https://sussmanbu.github.io/final-project-team9/

Mini project: Stochastic Gradient Descent

1 Introduction

Stochastic gradient descent (SGD) is a popular optimization algorithm applied in many fields. Under the assumption of strong convexity and smoothness of the loss function, the error analyses in Chapter 4 of the textbook establishes bounds that guarantee the accuracy of SGD and highlight the dependence of algorithm error on the hyperparameters. However, such assumptions are challenging to meet in practice. In this paper, we conduct empirical experiments to examine whether the Chapter 4 analyses provide an accurate guide to the SGD performance when the underlying assumptions are violated. We apply SGD in the context of a linear regression problem featuring a non-convex and non-smooth loss function, and then evaluate its performance while considering different choice of parameters including number of iterations, initialization, step size, gradient noise, and loss function. The experiments reveal that the Chapter 4 theorems still provide valuable insights into the relationship between the hyperparameters and SGD errors, and can be used as a general guide to understand the performance of the algorithm.

2 Background

Consider an optimization problem where we have a model with parameters x in D dimension, denoted as $x \in \mathbb{R}^D$. Our goal is to find a set of optimal parameters x_* that minimize the loss function $\mathcal{L}(x)$ of the model, which is in the form

$$x_* = \arg\min_x \mathcal{L}(x).$$

To reach the optimum, we update our initial parameter estimate x_0 towards the steepest descent of the loss function, which is the opposite direction of the gradient $\nabla\mathcal{L}$. This update process operates in a series of iterations. At each iteration k, the SGD algorithm uses a random subset of the dataset to estimate the overall loss function and computes the gradient of the loss function with respect to the parameters x_k. Then, with a step size η_k, the parameters are updated to minimize the loss following the rule

$$x_{k+1} \longleftarrow x_k - \eta_k \nabla\mathcal{L}_k(x_k).$$

The update process continues until the parameters x_k get "close" enough to the optimal parameters x_*. This closeness is measured in terms of the squared error $E_k := \|x_k - x_*\|_2^2$. Chapter 4 of the textbook discusses useful bounds for these SGD iterate errors. For SGD with constant step size $\eta_k = \eta$, Theorem 4.6.1 states that if the loss function is μ-strongly convex, the i.i.d. gradient estimates are L-co-coercive, and the variance at the optimum is bounded by σ^2, then the following bound holds for $\eta_k = \eta \in (0, \frac{1}{2L})$ and $\beta := 1 - 2\eta\mu(1 - \eta L)$:

$$\mathbb{E}(E_k) \leq \beta^k \|x_0 - x_*\|_2^2 + \frac{2\eta}{\mu}\sigma^2.$$

The theorem shows the expected squared error of iteration k is bounded by a term dependent on the initial squared distance to the optimum $\|x_0 - x_*\|_2^2$ plus an irreducible error. Because $\beta < 0$, as the number of iterations k increases, the dependence on initialization decreases exponentially. The components of the irreducible error term imply that a larger step size η, smaller μ, and higher noise level σ^2 can all lead to less accurate estimates of x_*. While the SGD algorithm is noisy, we can reduce the noise by taking the iterative average over the most recent 50% of iterations, denoted as $\bar{x}_{k/2:k}$. By additionally assuming the stochastic gradient is M-strongly smooth, Corollary 4.6.4 states that for $\rho := M/\mu$, the iterative average satisfies the bound that

$$\begin{aligned}\mathbb{E}(\|\bar{x}_{k/2:k} - x_*\|_2) \leq & \frac{\rho\eta\sigma^2}{\mu} + \frac{\sigma}{\mu k^{1/2}}(1 + 2^{3/2}\rho 1/2) + \frac{2\rho\sigma^2}{\mu^2 k} + \frac{3\rho^{1/2}\sigma}{\mu\eta k}(2 + \rho^{1/2}) \\ & + \beta^{k/2}\frac{\|x_0 - x_*\|_2}{\mu\eta k}\left\{2(1 + \rho^{1/2}) + \frac{\rho}{2\mu}\beta^{k/2}\|x_0 - x_*\|_2\right\}.\end{aligned}$$

The final term of the error bound depends on the initial distance to the optimum $\|x_0 - x_*\|_2$ and decays exponentially. The irreducible error $\frac{\rho\eta\sigma^2}{\mu}$ is scaled by the noise level σ^2 and step size η, which indicates noisier stochastic gradients and larger step size can lead to higher error in x_* estimates. Having discussed the bounds for SGDs with fixed step sizes, let us consider the case where the initial step size η decreases at a rate α. Theorem 4.7.1 addresses the bound on $\mathbb{E}(E_k)$ for SGD with decreasing step sizes $\eta_k = \eta/k^\alpha$: for $\alpha \in (1/2, 1)$,

$$\mathbb{E}(E_k) \leq exp\left\{-\frac{\mu\eta}{2}(k^{1-\alpha} - 1)\right\}\left(\|x_0 - x_*\|_2^2 + \frac{4\alpha\sigma^2\eta^2}{2\alpha - 1}\right) + \frac{2\sigma^2\eta}{\mu k^\alpha}.$$

As k increases, the term $exp\left\{-\frac{\mu\eta}{2}(k^{1-\alpha}-1)\right\}\|x_0 - x_*\|_2^2$ dependent on the initialization decays sub-exponentially. The term $\frac{2\sigma^2\eta}{\mu k^\alpha}$ is similar to the irreducible error term in theorem 4.6.1 but with step size η/k^α, and it slowly decreases with the choice of α. When α goes to one, the algorithm converges faster but the dependence on initialization no longer converges to zero.

Combining the information in Theorems 4.6.1 and 4.7.1 and Corollary 4.6.4, it becomes evident that the error bounds all contain 2 components: 1) an irreducible error term (which slowly decays for SGD with decreasing step size) associated with step size, gradient noise level, and the shape of the loss function, and 2) an exponentially decaying term dependent on initialization. These components are tied to our choice of hyperparameters, which raises the question of how changes in the hyperparameters influence the SGD performance. Moreover, the theorems assume strong convexity and smoothness of the loss function, and establish bounds for the expected values of the errors and hence do not guarantee consistent behavior for the entire SGD path. In the situation where the assumptions are not satisfied, whether the theorems can still be used to understand the general performance of SGD becomes uncertain. To address these questions, we will conduct empirical experiments.

3 Methods

To explore how SGD performance is affected by the violation of theorem assumptions and the choice of hyperparameters, we employ SGD in the context of a linear regression problem and evaluate its performance. Consider the observed data $\mathcal{D} = \{y_n, \boldsymbol{z}_n\}_{n=1}^N$ that has the following linear regression model: for response $y_n \in \mathbb{R}$, covariates $\boldsymbol{z}_n \in \mathbb{R}^D$, parameters $\boldsymbol{\beta} \in \mathbb{R}^D$, and i.i.d random errors $\varepsilon_1, ..., \varepsilon_N$,

$$y_n = \boldsymbol{z}_n^T \boldsymbol{\beta} + \varepsilon_n.$$

The errors follow a t-distribution with mean 0, scale e^ψ, and $\nu > 0$ degrees of freedom. As $\nu \to \infty$, the t-distribution converges to normal distribution $N(0, e^{2\psi})$ transforming the problem into a standard linear regression as a special case. Treating ν as fixed, we have the model parameters $\boldsymbol{x} = (\psi, \beta_1, ..., \beta_D) \in \mathbb{R}^{D+1}$. The conditional distribution of the response variable y_n can be written as $y_n \,|\, \boldsymbol{z}_n, \boldsymbol{x} \sim \mathcal{T}(\boldsymbol{z}_n^T\boldsymbol{\beta}, e^\psi, \nu)$, and its probability density function $p(y_n \mid \boldsymbol{z}_n^T\boldsymbol{\beta}, e^{\psi}, \nu)$ can be calculated using the following formula: for $\mathcal{T}(\hat{y}, s, \nu)$ with location parameter $\hat{y}$, scale parameter s and ν degrees of freedom

$$p(y|\hat{y}, s, \nu) = \frac{c(\nu)}{s}\left\{1 + \frac{1}{\nu}(\frac{y-\hat{y}}{s})^2\right\}^{-(\nu+1)/2}.$$

Then, we obtain the loss function of the n^{th} observation that

$$\begin{aligned} \ell_{(n)}(x) &= -\log p(y_n \mid \boldsymbol{z}_n^T\boldsymbol{\beta}, e^{\psi}, \nu) \\ &= \begin{cases} \frac{v+1}{2}\log\{1 + \frac{e^{-2\psi}}{\nu}(y - \boldsymbol{z}_n^T\boldsymbol{\beta})^2\} + \psi & \text{if } \nu < \infty, \\ \frac{e^{-2\psi}}{2}(y - \boldsymbol{z}_n^T\boldsymbol{\beta})^2 + \psi & \text{if } \nu = \infty. \end{cases} \end{aligned}$$

Setting the regularization term to be $\frac{1}{2}\|x\|_2^2$, we construct the overall loss

$$\mathcal{L}(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\ell_{(n)}(\boldsymbol{x}) + \frac{1}{2N}\|\boldsymbol{x}\|_2^2.$$

To examine the convexity and smoothness of the loss function, we define

$$g(\hat{y}, \psi, y) := \frac{v+1}{2}\log\{1 + \frac{e^{-2\psi}}{\nu}(y - \hat{y})^2\} + \psi,$$

hence the loss $\ell_{(n)}(\boldsymbol{x}) = g(\boldsymbol{z}_n^T\boldsymbol{\beta}, \psi, y_n)$. When y is fixed, the function $g(\hat{y}, \psi, y)$ is strongly convex for all values of ψ and when $\hat{y}$ lies in the range $y - \nu^{1/2}e^{\psi} \leq \hat{y} \leq y + \nu^{1/2}e^{\psi}$. Because the Hessian of $g(\hat{y}, \psi, y)$ is unbounded, the function is not strongly smooth. Therefore, the loss $g(\boldsymbol{z}_n^T\boldsymbol{\beta}, \psi, y_n)$ is convex for all values of ψ and for the set of $\boldsymbol{\beta}$ vectors where $y_n - \nu^{1/2}e^{\psi} \leq \boldsymbol{z}_n^T\boldsymbol{\beta} \leq y_n + \nu^{1/2}e^{\psi}$. We can then conclude that $\ell_{(n)}(x)$ is not strongly convex or strongly smooth everywhere for $\boldsymbol{x} = (\psi, \beta_1, ..., \beta_D)$. (See full proof in Appendix A.1) Because the overall loss $\mathcal{L}(\boldsymbol{x})$ is a linear combination of $\ell_{(n)}(\boldsymbol{x})$, it is not strongly convex and smooth.

Since the convexity and smoothness conditions do not hold, the Chapter 4 theorems cannot be applied exactly. However, since the loss is convex for some values of $\boldsymbol{\beta}$, there remains the possibility that the theorems can still be used to understand the performance of SGD to some extent. Therefore, we aim to use numerical experiments to evaluate SGD's behavior and its dependence on hyperparameter choices in this scenario of non-convexity and smoothness, and to assess whether the general results in theorems still hold.

To simulate the linear regression data, a total of N=10,000 data points are generated. We start by constructing multi-variate t-distributed covariates Z with dimension D=10 and 10 degrees of freedom, denoted as $Z \sim T(0, \Sigma)$ where Σ is a non-diagonal covariance. Then, we calculate the response Y by taking the dot product of the covariates Z and the true regression coefficients $\boldsymbol{\beta} = (1, 2, ..., D)$, and then adding t-distributed noise. We then center the response by subtracting its mean. As a result, the generated data has the true parameters $\boldsymbol{x} = (0, 1, 2, ..., D)$, and the optimal parameters x_* is determined by minimizing the overall loss function $\mathcal{L}(\boldsymbol{x})$. At each SGD iteration, a random batch of data points with size B is selected from the full dataset, and the parameter estimate x_k is updated using the gradient of the batch loss

$$\mathcal{L}_k(x) = \frac{1}{B}\sum_{b=1}^{B} \ell_{(n(k,b))}(x) + \frac{1}{2N}\|\boldsymbol{x}\|_2^2.$$

This stochastic estimate of the overall loss is unbiased in the sense that $\mathbb{E}\{\mathcal{L}_k(x)\} = \mathcal{L}(x)$.

4 Experiments

To determine whether the error analyses from Chapter 4 accurately infer the practical performance of SGD under the violation of theorem assumptions, we will conduct experiments with varying hyperparameters while considering three SGD algorithms respectively: last-iterate SGD with constant step size η, last-iterate SGD with decreasing step size η_0/k^α, and iterate-averaged SGD with constant step size η. We will examine how SGD accuracy is affected by the number of iterations, initialization, step size, gradient noise, and the loss function. For consistency purposes, the key parameters, unless otherwise specified, are set as follows: initial parameters $x_0 = \vec{0}$, batch size $B = 10$, degree of freedom $\nu = 5$, constant step size $\eta = 0.2$, initial step size $\eta_0 = 5$, and step size decay rate $\alpha = 0.51$.

4.1 Number of Iterations

The error bounds provided in theorems indicate that as the number of iterations k increases, the expected error decreases and approaches the irreducible errors. To investigate how the number of iterations affects the accuracy of SGDs, we run the algorithms for 10 epochs and measure the accuracy in terms of the squared error $\|x_k - x_*\|_2^2$.

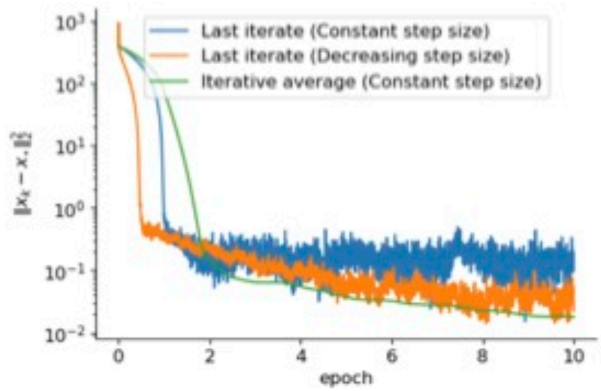


Figure 1: Accuracy across different iterations.

In Figure 1, we can see the accuracy of all three algorithms improves until they gradually converge, which meets the theorem's rule that the errors are bounded over iterations. The Last iterate SGD with constant step size reaches convergence at epoch 2 and starts oscillating. Compared to the last iterate SGD, the iterative average takes longer to achieve the same accuracy but continues to improve with more iterations. While the blue last iterate SGD path seems noisy, the green iterative average line approaches convergence more steadily and lies below the blue line as the number of iterations increases. Therefore, the performance of the iterative average is overall better than the last iterate SGD with a constant step size. The SGD with decreasing step size, on the other hand, decreases at the fastest speed of all three algorithms due to its initial large step sizes. Over the iterations, the step size decreases and the algorithm gradually converges. Because of the change in step size, the last iterate SGD with decreasing step size appears to be less noisy compared to the last iterate SGD with constant step size and hence is more preferable in the situation where we only run the algorithm for a few epochs. In the long run, the iterative average converges to a better accuracy.

4.2 Initialization

Intuitively, the further the initial parameters x_0 lie away from the optimal x_*, the longer it would take for the algorithm to converge in the sense that $\|x_k - x_*\|_2^2 < \varepsilon$. The error analyses in Chapter 4 state that the error's dependence on initialization $\|x_0 - x_*\|_2^2$ decays exponentially (or sub-exponentially) over iterations. To test if this rule still applies, we run 20 sets of initialization with different distances to the optimum and count the number of iterations required to converge. In Figure 1, the last iterate SGD reaches convergence at epoch 2. By taking the average of the squared norm starting from epoch 2 to epoch 10, we have the estimated expected error for the last iterate SGD with constant step size to be 0.15, and the plot also shows the error goes up to 0.3 after convergence. To be conservative, we define $\varepsilon = 0.3$.

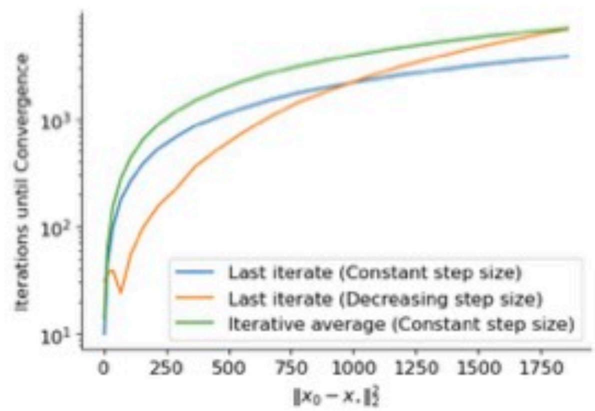


Figure 2: Initial distances vs number of iterations until convergence

All three algorithms exhibit similar patterns. From the plot, we can see that as the initial distance to the optimum $\|x_0 - x_*\|_2^2$ increases, the number of iterations required to reach convergence $\|x_k - x_*\|_2^2 < \varepsilon$ increases. For the SGDs with constant step size (with and without iterate averaging), Figure 2 shows a clear logarithmic trend. The decreasing step size SGD increases at a slower rate compared to the other two algorithms. These patterns correspond to the exponential and sub-exponential decaying dependence on the initialization addressed in the theorems.

4.3 Step Size

According to the theorem bounds on the expected squared error, taking a larger step size value increases the value of the irreducible error term. Meanwhile, as we have seen in the decreasing step SGD case, a larger step size may also take the algorithm to approach the optimum faster. For SGD with decreasing step size, the step size depends on both the initial step and the decay rate, therefore it will be evaluated separately from the constant step size SGDs. To visualize the influence of step size choices on SGD performance, we plot the squared errors $\|x_k - x_*\|_2^2$ over the number of epochs for different step sizes.

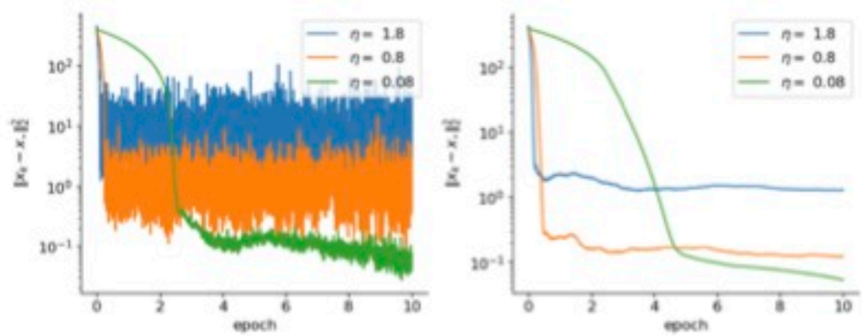


Figure 3: Last iterate (left) and iterate average (right); constant step size

With higher step sizes, both the last iterate and the iterate average converge faster. However, this improvement in convergence speed accompanies a lower convergence accuracy. For instance, with $\eta = 0.08$, the two algorithms all converge to squared errors below 0.1. On the other hand, with $\eta = 1.8$, the last iterate average SGD has a convergence error of 10 on average, and the iterate average converges to an error around 2. This pattern aligns with the Chapter 4 theorems which indicate a larger step size can increase the irreducible error term.

In the case of the last iterate SGD with decreasing step size, we will fix the decay rate to be $\alpha = 0.51$ while exploring the effect of changing step size, then keep the initial step size constant with $\eta_0 = 5$ to examine the effect of different decay rates.

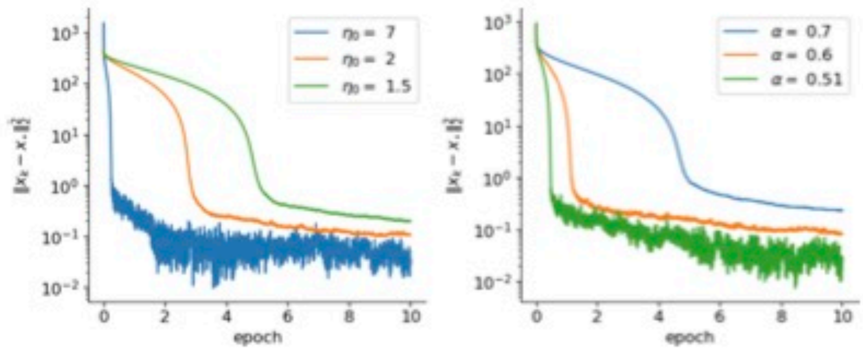


Figure 4: SGD with different initial step size(left) and decay rate (right)

Similar to the constant step case η, a larger initial step size η_0 also leads to faster convergence of the SGD with decreasing step size. Moreover, as shown in the left plot of Figure 4, a larger initial step leads to more accurate algorithm performance. When $\eta_0 = 7$, the algorithm converges at epoch 2 with an error below 0.1, while for $\eta_0 = 1.5$ it converges at epoch 6 with an error around 0.3. This could result from the fact that with a small initial step size, after running some iterations the step size becomes too small to decrease the error further.

The plot on the right shows when keeping the initial step size constant $\eta_0 = 5$, the algorithm with a higher decay rate α appears to have lower convergence accuracy. SGD with the highest decay rate $\alpha = 0.7$ converges to an error around 0.5 while the SGDs with $\alpha = 0.51$ and $\alpha = 0.6$ reach errors of 0.02 and 0.1 at convergence. One possible explanation for this behavior is that when the decay rate is high, the initial step size decreases rapidly and the step size becomes too small to approach the optimum further. Although the theorem 4.7.1 suggests a larger α makes the algorithm converge faster, we cannot conclude such a result from this experiment.

4.4 Gradient Noise

Recall that the gradient noise is equal to $\frac{\nu}{B}$ for some $\nu \geq 0$. By replacing the noise term σ^2 in the Ch. 4 theorems with $\frac{\nu}{B}$, it becomes clear that a higher batch size B reduces the value of the irreducible error term and hence can lead to more accurate algorithm performance. To test this rule, we run the

SGD algorithms for different batch sizes and examine their performance in terms of the error $\|x_k - x_*\|_2^2$ over the iterations.

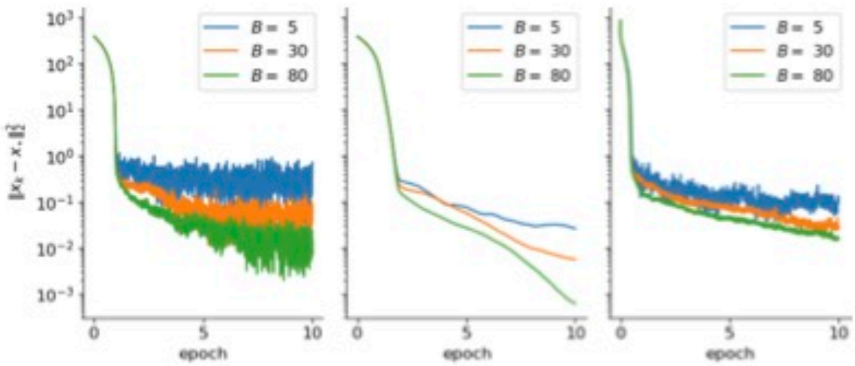


Figure 5: Last iterate with constant step size (Left); Iterative average with constant step size (Middle); Last iterate with decreasing step size (Right)

From the plots displayed on the left and right in Figure 5, we observe that for the two last iterate SGDs, the algorithms become less noisy with higher batch sizes. The improvement in stability is most obvious in the case of last iterate with decreasing step size. For all of the three SGDs, higher batch sizes appear to result in more accurate performance. In the three subplots, the green lines with batch size B=80 all converge to the lowest error compared to the blue and orange lines of batch sizes 5 and 30. Moreover, the accuracy difference between SGDs using batch sizes of B=5 and B=30 appears to be more substantial than the difference observed between batch sizes B=30 and B=80, which suggests a non-linear relationship. To further explore the relationship between the Batch size and the convergence error, we run the SGDs for 13 epochs using 10 different batch sizes, and then take the average squared error from 10 to 13 epochs to estimate the expected convergence error $\mathbb{E}\{E_k\}$.

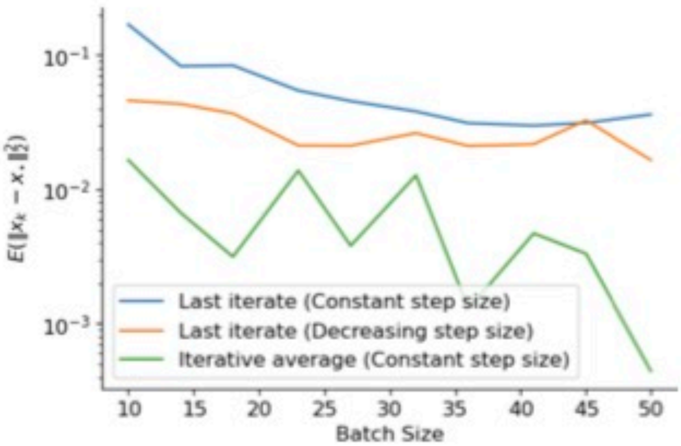


Figure 6: Convergence error for 10 different batch sizes

In Figure 6, as the batch size increases, the expected error of convergence decreases in an exponential pattern for the two last-iterate SGDs. When batch size increases from 10 to 20, the error on average decreases. However, after the batch size rises to an extent, the improvement in accuracy becomes less substantial. The error of the iterate average SGD also has a general decreasing pattern but does not show an exponential trend. Therefore, the exponential pattern generally aligns with the theorems.

4.5 Standard linear regression loss

When $\nu = \infty$, we have the loss function

$$\ell_{(n)}(x) = \frac{e^{-2\psi}}{2}(y - \boldsymbol{z}_n^T \boldsymbol{\beta})^2 + \psi.$$

Since the loss is the scaled squared error, it is more sensitive to outliers compared to the t distribution loss with $\nu = 5$. To investigate the effect of using the standard linear regression loss, we calculate the optimal parameters x_* based on the new loss and then compare the SGD performance to the case $\nu = 5$.

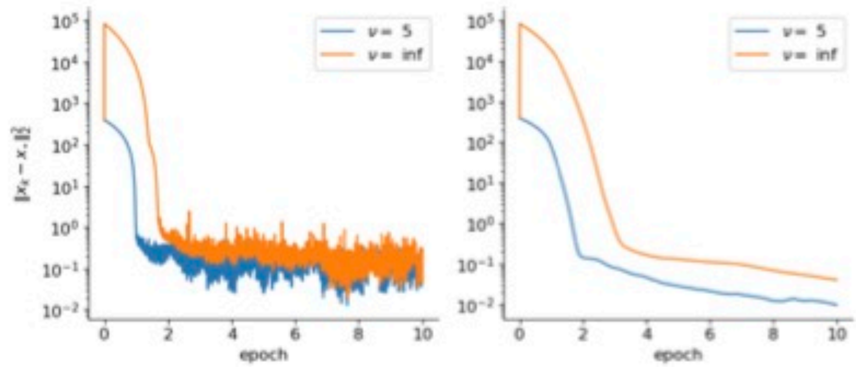


Figure 7: Last iterate with constant step size (left) and iterate average(right) with different loss function

In Figure 7, for the last iterate SGD and the iterate average with constant step sizes, setting $\nu = \infty$ results in a slightly less accurate but similar performance compared to using t distribution loss with $\nu = 5$. When $\nu = 5$, the error continues to decrease and gradually reaches convergence, whereas, for $\nu = \infty$, the error first goes up and then starts decreasing over iteration. Moreover, the algorithm appears to act noisier with the standard linear regression loss and this becomes more obvious in the case of decreasing step size SGD.

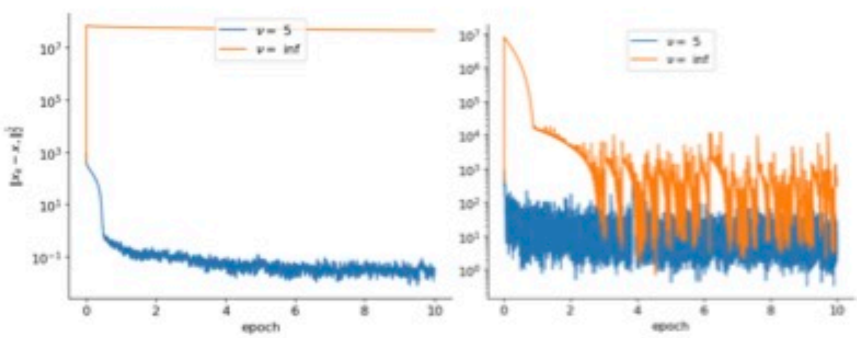


Figure 8: Decreasing step size SGD with decay rate 0.51 (left) and 0.1 (right)

The left plot in Figure 8 shows that for SGD with decreasing step size, at step size decay rate $\alpha = 0.51$, setting $\nu = \infty$ makes the error go up above 10^7. Theorem 4.7.1 requires the decay rate to satisfy $\alpha \in (1/2, 1)$. However, when $\nu = \infty$, having α in this range results in high errors. As shown by the orange line on the left side, the error goes up and never drops. In the constant step cases in Figure 7, the error also goes up in the first few iterations but soon decreases to better accuracy. The reason for this discrepancy in accuracy could be that in the decreasing step size case, the step size decreases too fast for the algorithm to improve the error further. Therefore, we set the decay rate smaller to see if the performance improves. In the right side plot where $\alpha = 0.1$, although the algorithm is still more noisy compared to using t distribution loss with $\nu = 5$, the error does go down over iterations. We can conclude that when using the standard linear regression loss, the algorithm seems to be more sensitive to step size choice and a smaller decay rate is more suitable. Overall, the SGD algorithms have less favorable performance using the standard linear regression loss compared to using t distribution loss.

5 Conclusion

In the numerical experiments, we discover that the SGD error still gradually decreases over the number of iterations until reaches convergence. Taking small constant step sizes and larger batch sizes can improve the algorithm's accuracy. More importantly, although with some exceptions, the observed relationship between the SGD performance and the hyperparameters appears to match with the theorems in Chapter 4. Therefore, we can conclude that the error analyses can provide a general guide to the performance of SGD algorithms even under the violation of the theorem assumptions. Because our conclusion is merely based on results from a linear regression problem, in future experiments, we can explore SGD performance in other types of regressions. Additionally, we may construct the theorem bound values to see whether they provide an accurate upper bound for the convergence error.

A Appendix

A.1 Convexity and Smoothness

Given the function that

$$g(\hat{y}, \psi, y) := \frac{v+1}{2} \log\{1 + \frac{e^{-2\psi}}{\nu}(y - \hat{y})^2\} + \psi,$$

treating ψ and y as fixed constants, we can calculate the first and second derivatives with respect to $\hat{y}$.

$$\begin{aligned}
\nabla g(\hat{y}) &= \frac{v+1}{2} \cdot [1 + \frac{e^{-2\psi}}{v}(y-\hat{y})^2]^{-1} \cdot \frac{2e^{-2\psi}}{v}(y-\hat{y})(-1) \quad \text{by chain rule} \\
&= \frac{(v+1)\frac{e^{-2\psi}}{v}(\hat{y}-y)}{1+\frac{e^{-2\psi}}{v}(y-\hat{y})^2} \quad \text{simplifying} \\
&= \frac{(v+1)(\hat{y}-y)}{\nu e^{2\psi}+(y-\hat{y})^2} \quad \text{multiplying } \nu e^{2\psi} \text{ to both parts of the fraction.}
\end{aligned}$$

By the quotient rule,

$$\begin{aligned}
\nabla_2\, g(\hat{y}) &= \frac{(\nu e^{2\psi}+(y-\hat{y})^2)\cdot(v+1) - (v+1)(\hat{y}-y)\cdot(-2)(y-\hat{y})}{(\nu e^{2\psi}+(y-\hat{y})^2)^2} \\
&= \frac{(\nu e^{2\psi}+(y-\hat{y})^2)\cdot(v+1) - (v+1)\cdot 2(y-\hat{y})^2}{(\nu e^{2\psi}+(y-\hat{y})^2)^2} \quad \text{by simplifying} \\
&= \frac{(v+1)(\nu e^{2\psi}-(y-\hat{y})^2)}{(\nu e^{2\psi}+(y-\hat{y})^2)^2} \quad \text{by simplifying}
\end{aligned}$$

When $\nabla_2\, g(\hat{y}) \geq 0$, the function $g(\hat{y},\psi,y)$ is convex. We can find the $\hat{y}$ range by solving

$$\begin{aligned}
\frac{(v+1)(\nu e^{2\psi}-(y-\hat{y})^2)}{(\nu e^{2\psi}+(y-\hat{y})^2)^2} &\geq 0 \\
\nu e^{2\psi}-(y-\hat{y})^2 &\geq 0 \qquad \text{because } \nu > 0 \\
\nu e^{2\psi} &\geq (y-\hat{y})^2 \\
\nu^{1/2}e^{\psi} &\geq |y-\hat{y}|.
\end{aligned}$$

Therefore, we have $\nu^{1/2}e^{\psi} \geq y-\hat{y} \geq -\nu^{1/2}e^{\psi}$, and by division and subtraction, the function $g(\hat{y},\psi,y)$ is convex for the $\hat{y}$ range that

$$y-\nu^{1/2}e^{\psi} \leq \hat{y} \leq y+\nu^{1/2}e^{\psi}.$$

Because $z_n^T\beta$ is a linear function and the composition of a convex function and a linear function is convex, we can conclude that $g(z_n^T\beta, \psi, y_n)$ is also convex for the set of β that satisfies $|y_n - z_n^T\beta| \leq \nu^{1/2}e^{\psi}$.

To evaluate convexity with respect to ψ, we treat $\hat{y}$ and y as fixed constants and calculate the first and second derivatives. By the chain rule,

$$\begin{aligned}
\nabla\, g(\psi) &= \frac{v+1}{2}\cdot[1+\frac{e^{-2\psi}}{v}(y-\hat{y})^2]^{-1}\cdot\frac{-2(y-\hat{y})^2}{v}\cdot e^{-2\psi}+1\\
&= 1-\frac{(\nu+1)(y-\hat{y})^2e^{-2\psi}}{\nu(1+\frac{e^{-2\psi}}{\nu}(y-\hat{y})^2)}\\
&= 1-\frac{(\nu+1)(y-\hat{y})^2e^{-2\psi}}{\nu+e^{-2\psi}(y-\hat{y})^2}.
\end{aligned}$$

By the quotient rule,

$$\begin{aligned}
\nabla_2\, g(\psi) &= -\frac{(\nu+e^{-2\psi}(y-\hat{y})^2)\cdot(\nu+1)(y-\hat{y})^2e^{-2\psi}(-2)-(\nu+1)(y-\hat{y})^4e^{-4\psi}(-2)}{(\nu+e^{-2\psi}(y-\hat{y})^2)^2}\\
&= -\frac{(-2)(\nu+1)(y-\hat{y})^2e^{-2\psi}(\nu+e^{-2\psi}(y-\hat{y})^2-(y-\hat{y})^2e^{-2\psi})}{(\nu+e^{-2\psi}(y-\hat{y})^2)^2}\\
&= \frac{2\nu(\nu+1)(y-\hat{y})^2e^{-2\psi}}{(\nu+e^{-2\psi}(y-\hat{y})^2)^2}.
\end{aligned}$$

For $\nabla_2\, g(\psi) \geq 0$, the function $g(\hat{y}, \psi, y)$ is convex. so we have

$$\begin{aligned}
\frac{2\nu(\nu+1)(y-\hat{y})^2e^{-2\psi}}{(\nu+e^{-2\psi}(y-\hat{y})^2)^2} &\geq 0\\
e^{-2\psi} &\geq 0 \qquad\qquad \text{because } \nu > 0.
\end{aligned}$$

Therefore, the function $g(\hat{y}, \psi, y)$ is convex for all $\psi \in \mathbb{R}$. Again, because $z_n^T\beta$ is a linear function, $g(z_n^T\beta, \psi, y_n)$ is also convex for all values of ψ. Taking into account that $\ell_{(n)}(x) = g(z_n^T\beta, \psi, y_n)$ is not convex for all values of β, the loss function $\ell_{(n)}$ is not convex everywhere.

If the function $\ell_{(n)}(x) = g(z_n^T\beta, \psi, y_n)$ is strongly smooth, then for $\hat{y} = z_n^T\beta$ there exist some constant $c_1 < \infty$ and $c_2 < \infty$, the second derivatives satisfy that

$$\|\frac{\partial^2 g}{\partial \psi^2}\|_2 < c_1 \quad \text{and} \quad \|\frac{\partial^2 g}{\partial \hat{y}^2}\|_2 < c_2.$$

We will look at these two cases separately.

$$\|\frac{\partial^2 g}{\partial \psi^2}\|_2 = |\frac{2\nu(\nu+1)(y-\hat{y})^2 e^{-2\psi}}{(\nu + e^{-2\psi}(y-\hat{y})^2)^2}|$$

As $\psi \to \infty$, the value $e^{-2\psi} \to 0$ and norm goes to zero. As $\psi \to -\infty$, the value $e^{-2\psi} \to \infty$ and the norm also goes to infinity. Therefore, there does not exist a constant $c_1 < \infty$ such that $|\frac{\partial^2 g}{\partial \psi^2}\|_2 < c_1$. Hence the loss is not strongly smooth with respect to ψ. For the

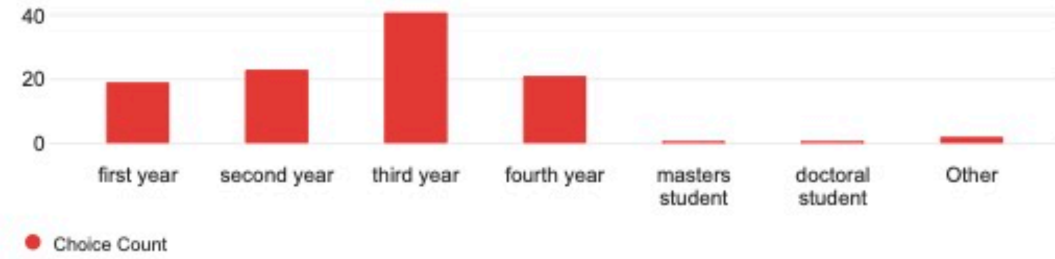
$$\|\frac{\partial^2 g}{\partial \hat{y}^2}\|_2 = |\frac{(v+1)(\nu e^{2\psi} - (y-\hat{y})^2)}{(\nu e^{2\psi} + (y-\hat{y})^2)^2}|.$$

As $|y-\hat{y}| \to \infty$, $\|\frac{\partial^2 g}{\partial \hat{y}^2}\|_2$ goes to $-\infty$, hence there exists a constant $c_2 < \infty$ such that $\|\frac{\partial^2 g}{\partial \hat{y}^2}\|_2 < c_2$. Therefore, the norm $\|\frac{\partial^2 g}{\partial \beta^2}\|_2$ is also bounded, and the loss function is strongly smooth with respect to β. Because the $g(z_n^T\beta, \psi, y_n)$ is not strongly smooth with respect to ψ, the loss $\ell_{(n)}(x)$ is not strongly smooth everywhere.

Appendix C: Student Survey Results

AY 23/24 Mathematics and Statistics Student Survey Results

Q4 - What is your current year at Boston University? - Selected Choice

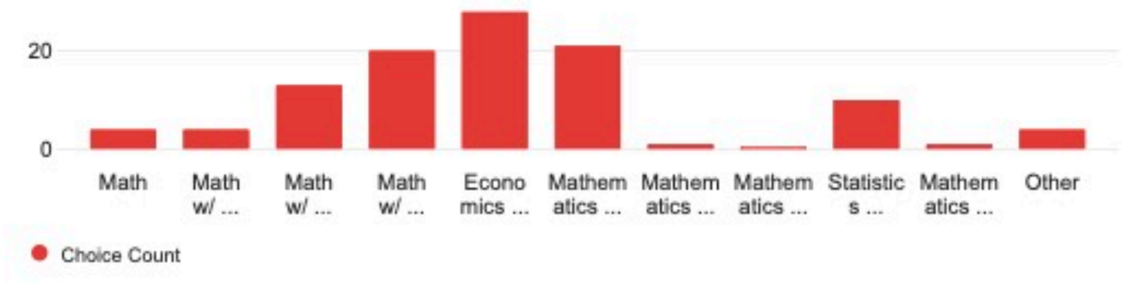


Q4_7_TEXT - Other - Text

Other - Text
Fifth year BA/MS
senior in BA/MS

Q5 - Did you transfer to Boston University from another institution? - Selected Choice

Q6 - What is your academic major? - Selected Choice

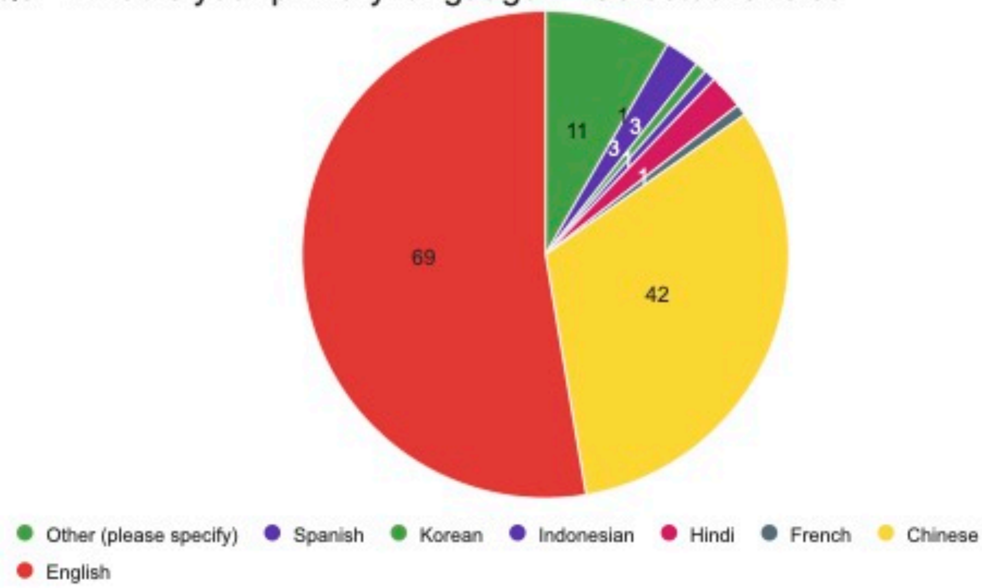


Q6_12_TEXT - Other - Text

Other - Text
Mathematics w specialty in Pure and Applied
Math w/ specialty in Pure & Applied Mathematics
Statistics and Data Science
Mathematics (pured/applied)

Q7 - Have you declared more than one major? - Selected Choice

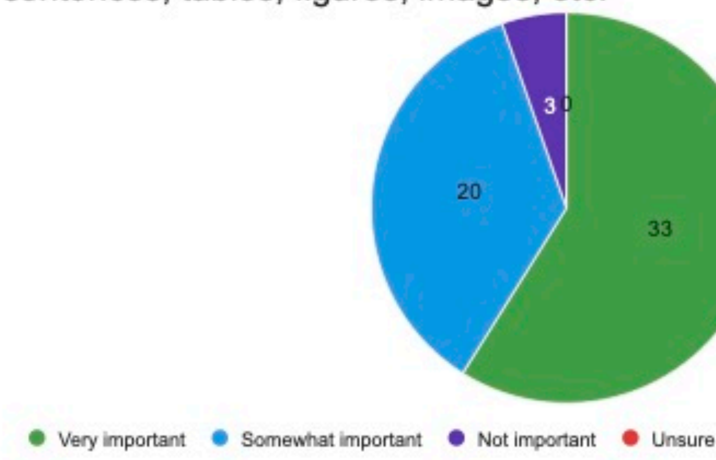
Q8 - What is your primary language? - Selected Choice



Q8_20_TEXT - Other (please specify) - Text

Other (please specify) - Text
Slovak
Italian
Czech
Vietnamese
Turkish
Vietnamese
Vletnamese
Vietnamese
Turkish
Turkish
Romanian

Q1 - How important is writing to the scholarly and professional work done in the mathematics and/or statistics discipline? "Writing" is defined as communication in which visual marks make meaning, such as words, sentences, tables, figures, images, etc.



Q2 - Describe in a sentence or two why you answered the previous question the way you did.

Describe in a sentence or two why you answered the previous question the way you did.
I have not done any sort of writing in my MA classes. In my EC204 class we wrote a 15 page research paper; however, this is an EC class, not an MA class.
The mathematics major can be applied to a variety of subjects.
many tables are confusing to understand, so using words can help show what you're trying to analyze
I have not written any essay outside of writing class since I started learning my corse.
I think writing sometimes help convey ideas, but for math cannot only use words to explain clearly.
There just not much writing, as much as problem solving.
Most proof-based math classes are all about writing. Problem sets, proofs, studying. The words add up, and some homework sets can be longer than English essays.
To be able to effectively write out scholarly and professional work in mathematics communicates a very deep level of understanding that cannot really be seen from simply solving equations.

5

My research over the summer was largely an exercise in concise writing, beyond our actual mathematical research. Even then, concise computation wasn't as important as explaining our proofs and motivations.

it is important to attain writing knowledge before going into the math world. i feel as if we need a writin in math prerequisite before we take other math classes

Writing is used to help explain the research and purpose of it and how it is applied.

Professional/Academic writing skills plays a big part in both writing and understanding scientific and mathematics papers and publications.

I find that learing math is alot easier when you can visualize what is going on.

I believe writing is essential to delivering a complete explanation of what you are trying to prove, explain, or create for any math concept. Without writing, it would be extremely difficult to do those things.

Proof in Math is extremely important and one of the best and most efficient ways for students to grasp the concepts is visually whether it be on a board or a presentation.

It's easier for others to understand the context and logic when solving math problems

Writing is indispensable to the scholarly and professional work in mathematics and statistics, as it serves as the primary means of communicating complex ideas, theorems, proofs, and research findings to both peers and the broader scientific community.

Math is best understood with tables and/or figures, but are even better when said in sentences what they mean.

We are urged to make figures, etc, very clear in order to convey our information correctly. This especially maters when considering figures, tables, data, etc.

I think it is important to convey your work in an understandable way.

Q10 - In your opinion, which of the following characteristics are particularly descriptive of writing in the mathematics and/or statistics discipline? (Select all that apply) - Selected Choice

Field	Choice Count
Expressive: emphasizing personal feelings and impressions	5
Interpretive and/or Evaluative of others' works and ideas	21
Descriptive: conveying processes, objects, data, environments, etc.	45
Analytical: emphasizing the logical examination of subjects	50

Persuasive: presenting and evidencing positions or claims	20
Exploratory: investigating and developing ideas using discovery-based writing	28
Visual: emphasizing visual components such as graphic presentation, sketches, drawings, videos, etc.	35
Explanatory: translating complex content into generally comprehensible definitions and/or instructions	39
Innovative: approaching subject in fresh and inventive ways	20
Collaboratively authored	18
Reflective: applying lessons to one's own life; metacognition	8
Multimodal: communicating in more than one modality (ex. visual, linguistic, spatial, aural, and gestural)	15
Unsure	2
Other	1

Q11 - In the previous question, you indicated characteristics that are particularly descriptive of writing in mathematics and/or statistics. Which three characteristics seem the most important when describing writing in this field?

In the previous question, you indicated characteristics that are particularly descriptive of writing in mathematics and/or statistics. Which three characteristics seem the most important when describing writing in this field?

Visual is the most important. Visualization is extremely helpful in helping to understand higher level concepts. Exploratory is the second most important. Going into depth in topics is very important. Descriptive is the third most important. Properly describing data, objects, and processes is pivotal in understanding higher level concepts.

Descriptive/Analytical/Reflective

i chose those three because they are most relevant when reading an academic article

Idk.

Analytical, explanatory, exploratory

Descriptive, analytical, persuasive

Writing is writing, mathematical or no. Each of the above boxes represents a facet of writing in mathematics, and as a consequence, writing for the math department and its students.

Explanatory, Visual, and Innovate

Analytical, Explanatory, Persuasive

interpretive, descriptive and analytical

Descriptive, analytical, explanatory

Interpretive, analytical, and explanatory

analytical, descriptive, explanatory

analytical, explanatory, and visual

Explanatory then visual then analytical

.

Three important characteristics when describing writing in the math field are precision, clarity, and rigor.

Analytical, exploratory, explanatory

Descriptive, Visual and Analytical—the goal is most often to convey information in the most clear way possible.
Descriptive, analytical, exploratory

Q3 - Which of the following writing abilities do faculty in your mathematics and/or statistics major expect you to demonstrate by the time you graduate? - Selected Choice

Field	Choice Count
Use field-specific terminology, organizational formats, and/or conventions	40
Argue a position using a central thesis and hypothesis and evidence	28
Create and incorporate visuals or presentation formats (figures, drawings, tables, photos, posters, slides)	26
Describe processes, objects, findings, environments, etc.	37
Summarize ideas, texts, or events	26
Analyze, interrogate, and/or evaluate ideas, texts, or events	40
Use correct grammar, spelling, mechanics (punctuation etc.), and notation	22
Propose innovative ideas or perspectives	15
Co-author texts with one or more writer	9
Report and explain complex data, findings, or figures, using examples as appropriate	24
Use writing to develop and deepen thinking	15
Synthesize disparate ideas and/or perspectives	11
Express feelings or impressions	5
Reflect upon experiences and/or assumptions	13
Solve complex problems	35
Integrate and correctly cite information from well-chosen sources	15
Apply theory, argument, or findings to real-world circumstances	33
Contextualize your argument or findings	18

9

Other (please specify)	3

Q3_19_TEXT - Other (please specify) - Text

Other (please specify) - Text
Reflect, solve , and create visuals
I'm unaware of the writing expectations
No writing

Q12 - Of the writing abilities identified in the previous question, which three are most critical for students graduating with a mathematics and/or statistics major?

Of the writing abilities identified in the previous question, which three are most critical for students graduating with a mathematics and/or statistics major?
Synthesize ideas, texts, or events. Describe processes, objects, findings, environments, etc. Creative and incorporate visuals or presentations.
Use field-speciric terminology/Argue a position/Analyze
i use these three when writing labs
Idk.
Solve problems, analysis, write to lead deep thinking
Apply theory
Again, they all do, and for the same reason.
Analyze/evaluate ideas and text, describe processes, and summarize ideas
Analysis, Concise Central Thesis, Contextualizing Findings
the three i chose!!
Contextualize, analyze, report/explain data
Summarize, Analyze, and Apply
solve complex problems, apply theory,use field specific terminology

Solve complex problems, propose innovative ideas, and describe processes
field specific terminology, presentation formats and report and explain complex data
Analyze, summarize and report
visual presentations, apply theory,describe processes
Apply theory, explain complex data, describe processes
Problem solving, application of concepts, and analysis
Apply into real world, describe, and argue

Q13 - Considering the writing you do for courses in mathematics and/or statistics...

Field	Min	Max	Mean	Standard Deviation	Variance	Responses	Sum
Use field specific terminology, organizational formats, and/or conventions	1.00	3.00	1.96	0.72	0.52	50	98.00
Argue a position using a central thesis or hypothesis and evidence	1.00	3.00	2.14	0.66	0.44	50	107.00
Create and incorporate visuals or presentation formats (figures, drawings, tables, photos, posters, slides, etc.)	1.00	3.00	1.98	0.67	0.45	51	101.00
Describe processes, objects, findings, environments, etc.	1.00	3.00	2.02	0.68	0.47	49	99.00
Summarize ideas, texts, or events	1.00	3.00	1.92	0.67	0.45	48	92.00
Analyze, interrogate, and/or evaluate ideas, texts, or events	1.00	3.00	2.04	0.75	0.57	49	100.00
Use correct grammar, spelling, mechanics (punctuation, etc.), and notation	1.00	3.00	1.74	0.72	0.51	50	87.00
Propose innovative ideas or perspectives	1.00	3.00	2.26	0.73	0.53	42	95.00
Co-author texts with one or more writers	1.00	3.00	2.29	0.68	0.47	38	87.00
Report and explain complex data, findings, or figures, using examples as appropriate	1.00	3.00	1.96	0.71	0.51	51	100.00

Use writing to develop and deepen thinking	1.00	3.00	2.20	0.68	0.46	46	101.00
Synthesize disparate ideas and/or perspectives	1.00	3.00	2.11	0.70	0.49	46	97.00
Express feelings or impressions	1.00	3.00	2.03	0.72	0.52	40	81.00
Reflect upon experiences and/or assumptions	1.00	3.00	2.00	0.65	0.42	43	86.00
Solve complex problems	1.00	3.00	1.92	0.71	0.50	51	98.00
Integrate and correctly cite information from well-chosen sources	1.00	3.00	1.93	0.76	0.58	46	89.00
Apply theory, argument, or findings to real-world circumstances	1.00	3.00	1.98	0.70	0.49	51	101.00
Contextualize your argument or findings	1.00	3.00	2.00	0.52	0.27	51	102.00

Q14 - Briefly describe one writing assignment from a mathematics and/or statistics course that has been particularly useful for you and explain why it has been useful.

Briefly describe one writing assignment from a mathematics and/or statistics course that has been particularly useful for you and explain why it has been useful.
In my EC204 class, which is essentially MA214, we wrote a 15 page research paper, and I found the semester long process of doing research and finding trends within data was extremely useful. It is a project that I include on my resume.
重点词汇 7/5000 传统翻译模型 通用场景 An experimental course in statistics
in my lab for ma213 we were able to create a project and go step by step into academic writing
Idk.
No
Idk
Each problem set is an adventure. A journey. The writing is useful in and of itself; a tool for learning.
I have not had many writing assignments from any of my mathematics or statistics courses.

Calculus 2 Proof Writing. It was my first exposure to higher math concepts and theorems, and it proved to be quite difficult to prove even elementary theorems.

MA 116, this class propelled me into analyzing and reading data in a formal and productive manner!

Research paper for MA 213 where we researched a topic and proposed various hypotheses. We calculated them off of that and explained our process and conclusion.

N/A

We barely ever use words in math. But It has made be better at using symbols and characters to solve real world problems

The writing of proofs. It has helped me break down concepts and explain to the reader and myself why I am taking particular steps

My problem sets in Statistics helps me contextualize my problems in real-world scenarios

The project in ma433 because it helps us to learn how to write math in academic way

n/a

Have not had big writing assignments so far

I found it quite interesting when I had to submit weekly differential equation homeworks in a format akin to an essay. It caused to me to have to truly understand the computations I was making rather than just cranking them out.

I haven't had one yet.

Q15 - What is the highest level Writing Program course you have completed at Boston University?

Q16 - Have you taken a first-year writing course or equivalent at another post-secondary institution or in high school? - Selected Choice

Q16_1_TEXT - Yes - Text

Yes - Text
AP Lang and AP Lit
Writing for Research, CUNY Baruch
UMass Boston
AP English Language
My courses aren't listed: Kilachand (KHC) writing seminars and WR 415 covered my HUB.
AP English Literature
UMass Boston
High school
IDK

Q17 - Which of the following interdisciplinary writing abilities were addressed in your first-year BU writing courses or equivalent? - Selected Choice

Field	Choice Count
Audience address: communicating ideas to specific audiences (e.g. peers, scholars in the field, individuals outside academia, etc.)	30
Purpose: focusing on specific writing purposes (persuading, informing, expressing, etc.)	35

Thesis: articulating a central idea or position	41
Organization: sequencing content logically	39
Evidence: using evidence to support an idea or position	42
Counterarguments: acknowledging and responding to opposing perspectives	32
Paragraphing: constructing cohesive, structured, and focused paragraphs	33
Drafting and revising: writing multiple drafts	37
Peer response: responding constructively to peers' drafts	36
Research: locating, evaluating, and using research material	36
Citation: citing sources using a consistent format	40
Grammar and usage: controlling such features as mechanics, sentence structure, and spelling	34
Entering a scholarly conversation: contributing to ongoing scholarship	17
Other (please specify)	4

Q17_14_TEXT - Other (please specify) - Text

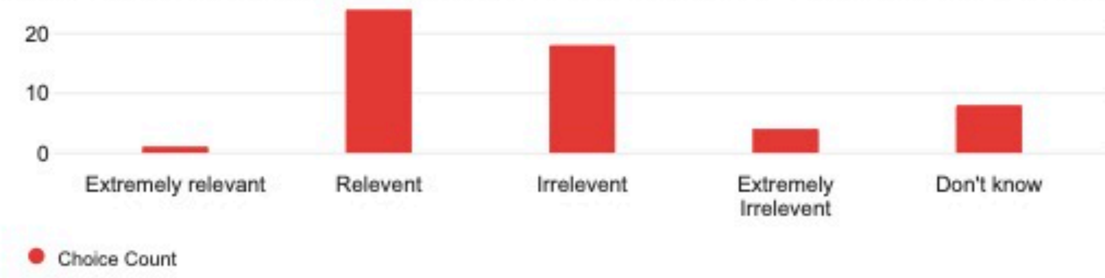
Other (please specify) - Text
n/a
Didn't do this class.
none, really; see next Q
Have not taken yet.

Q18 - Is there anything you would like to add about what you learned in your "First Year Writing" seminar and/or your "Research Writing" seminar?

Is there anything you would like to add about what you learned in your "First Year Writing" seminar and/or your "Research Writing" seminar?
no

no
No.
No
n/a
no
n/a
no
No
Academic research writing
no
N/A
N/A, they were quite cohesive.
nope
creativity, we made a website and podcast to convey our point
nope
Nothing much. I feel like my writing skills were ironed out in high school, so WR120/153 were a bit insignificant to me.
n/a
No
No, I think the points above covered everything.

Q19 - How relevant was the writing you did and the writing instruction you received in your "First Year Writing" and/or "Research Writing" seminars to the writing you do in mathematics and/or statistics courses?



Q20 - Is there anything you would like to add about how the writing you did in the "First Year Writing" and/or "Research Writing" seminars or did not relate to the writing you do in mathematics and/or statistics courses?

Is there anything you would like to add about how the writing you did in the "First Year Writing" and/or "Research Writing" seminars or did not relate to the writing you do in mathematics and/or statistics courses?
none
No.
No
n/a
no
n/a
N/A
No
Specific for math
no
In math we do more proof based writing, so I feel like the structure my writing class taught me and the structure I have done in my math classes so far are not the same
For me, the writing in those courses was quite disconnected to what we do in math/stats classes.

Those were more of follow an MLA format and stuff while math was more of converting a proof to words.
n/a
Most of the writing skills are correlated, but for maths, more figure-related stuff is needed. Also to be logical is important.
Proof writing is very different than expository writing, which is what's mainly expected from you from WR120/15X.
n/a
The writing courses I took in college relate to writing in math because they both involve the flow of thoughts, explanation, using supporting evidence, and stating a conclusion.
No
No

Q21 - During the past academic year, how frequently has your writing in mathemati...

Field	Min	Max	Mean	Standard Deviation	Variance	Responses	Sum
You received a number or letter grade without comments	1.00	4.00	2.73	1.02	1.04	52	142.00
You received a number or letter grade with comments	1.00	4.00	2.48	1.10	1.21	52	129.00
You received a grade and comments accompanied by a list of criteria (or a "grading rubric")	1.00	4.00	2.65	1.02	1.03	52	138.00
Your writing was evaluated as part of a grading contract and did not receive a specific number or letter grade	1.00	4.00	3.04	0.83	0.69	52	158.00
Your instructor asked you to evaluate your own writing as part of the grading formula	1.00	4.00	3.02	0.84	0.71	52	157.00
You were provided with and discussed grading criteria before the assignment was due	1.00	4.00	2.56	1.10	1.21	52	133.00
You received a grade on a portfolio of work you selected and assembled	1.00	4.00	2.92	0.92	0.84	52	152.00

Q22 - Were you satisfied with the methods used to respond to and/or evaluate your writing in mathematics and/or statistics?

Q23 - Is there anything you would like to add about the evaluation of and feedback on your writing in mathematics and/or statistics courses?

Is there anything you would like to add about the evaluation of and feedback on your writing in mathematics and/or statistics courses?
No
A writing credit for upper-level proof-based mathematics courses. It's frankly absurd. I took three courses last semester and wrote over 100,000 words total in the course of my problem sets.
I have not had much, if any, writing in my math courses yet.
n/a
N/A
I am a transfer this is my 1st semester
No
No
no
I have not taken higher level courses that require this, prereqs just have you write a paragraph explaining your answer at most
I think the grading of these assignments are less subjective so fall under stricter grading guidelines with less leeway.
no

n/a

It's kind of subjective for evaluating writing assignments. It is inevitable.

n/a

No

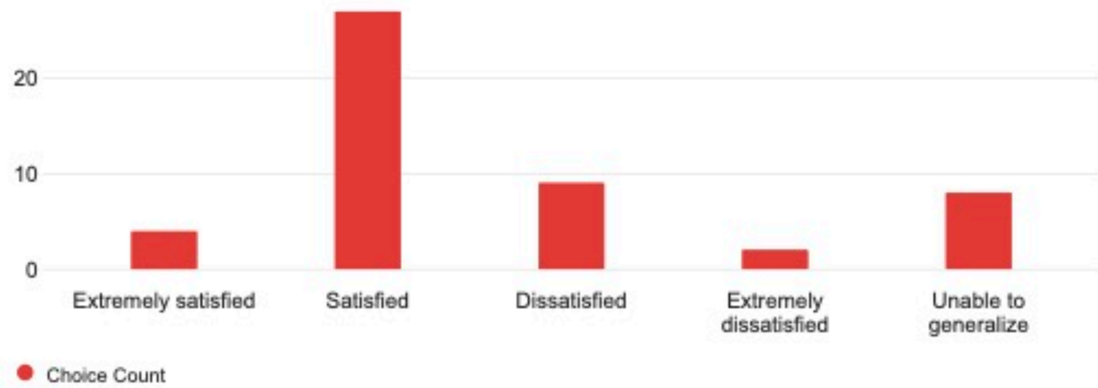
No

N/a

N/A

N/A

Q24 - To what degree are you satisfied with the overall quality of writing instruction you receive in your mathematics and/or statistics courses?



Q25 - Is there anything you would like to add about the overall quality of writing instruction in your mathematics and/or statistics courses?

Is there anything you would like to add about the overall quality of writing instruction in your mathematics and/or statistics courses?

n/a

no

unfufilling, it feels as if the only purpose is to pass, not learn.

writing is taught by the tas, never the instructors

N/A

I would appreciate it if they gave us a word limit or a minimum of how much we should write when making or writing proofs because at times I feel like I am writing too much or sometimes not enough

No

Have not had much writing instruction

no

I know there was a 2-credit 'writing in mathematics' course that I was interested in, but couldn't join since it was only for mathematics majors (im cs and maths). It could have been helpful, maybe it's worth offering it again.

Hard to find resources or help

I think anyone that wants to take an abstract math course or proof writing course should either be required to or recommended to take an intro level proof writing class. I took MA541 without a lot of proof writing experience and I realized it would've helped a lot if I had taken an intro level class before taking MA541.

No

No

With the exception of MA301 which only runs during the spring semester, there is virtually no instruction in writing at the underclassmen level--this seems flawed.

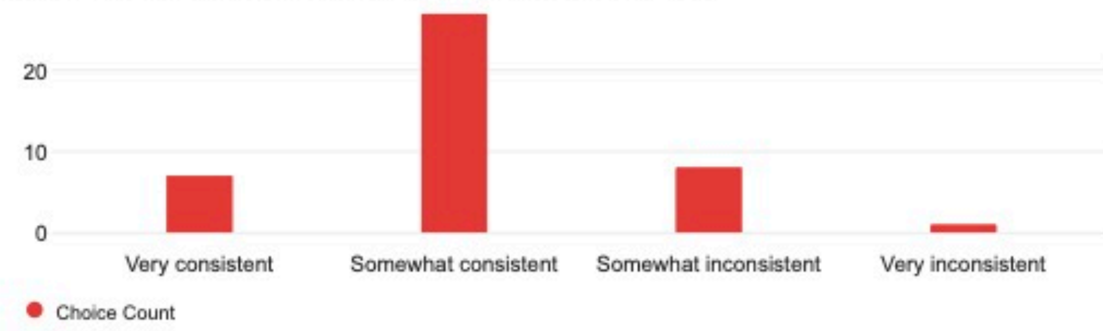
No

No

No

more!

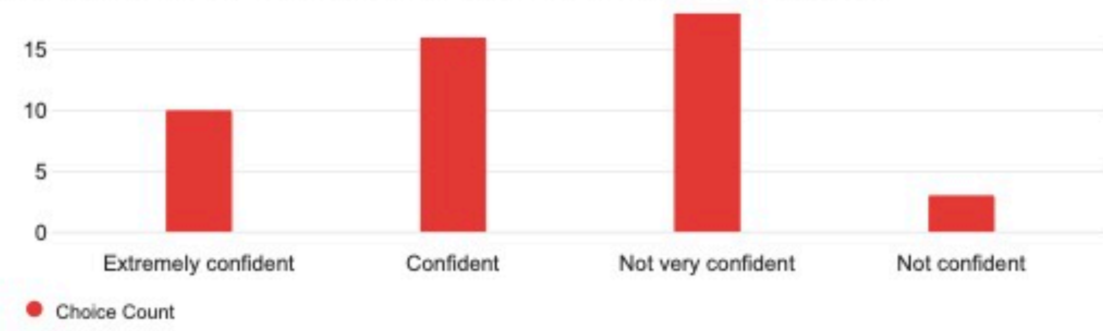
Q26 - How consistent is the approach to writing and writing instructions across mathematics and/or statistics courses?



Q27 - Is there anything you would like to add about the consistency or inconsistency of writing instruction across mathematics and/or statistics courses?

Is there anything you would like to add about the consistency or inconsistency of writing instruction across mathematics and/or statistics courses?
n/a
no
it's up to teacher discretion
N/A
Cannot think of anything
No
no
n/a
Unsure as my only experience so far is MA511/512.
No
Some classes are very much "Go to lecture and do the homework." I appreciate classes that let us think more broadly about how the concepts we're learning are part of an ongoing scholarly conversation
No
No
No
No
More focused at theory and solving problems other than writing
nope
Not sure
It is very consistently not there.

Q28 - How confident are you in your ability to write in ways that meet the expectations of courses in mathematics and/or statistics?



Q29 - Is there anything you would like to add about your confidence level?

Is there anything you would like to add about your confidence level?
n/a
Mainly because I have not had any experience with it yet
n/a
N/A
No
No
I feel like I am not bad, but could be better
I believe all classes should have some writing in math portion
Different professor have different ideas about what kind of writing is what they want, I can't say I am confident with every assignment.
No
No
No
No

No

No

just haven't had practice

No

I have never done it.

dk

Q30 - What other comments would you like to make about the role or importance of writing in mathematics and/or statistics?

What other comments would you like to make about the role or importance of writing in mathematics and/or statistics?

n/a

I believe writing in general is vital

n/a

N/A

N/a

No

I feel like it's pretty important when you have to convey your ideas, your thinking, how to get to an answer

I want to pursue further studies in mathematics and therefore I belive it is very important for students like me.

nope

No

None

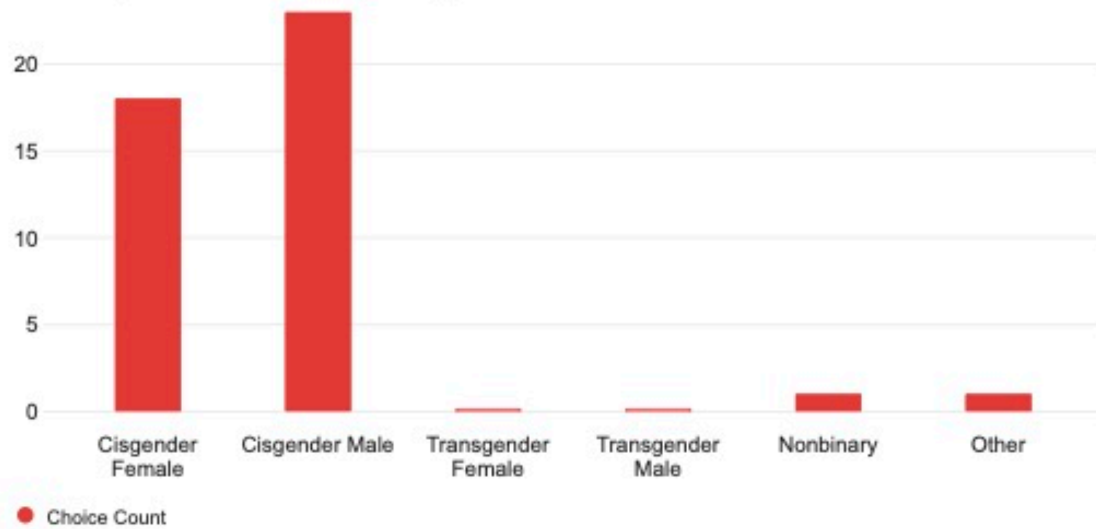
No

No

No

none
No
It is important, but it is not present in math courses at BU.
dk
None

Q33 - Optional: What is your gender? - Selected Choice

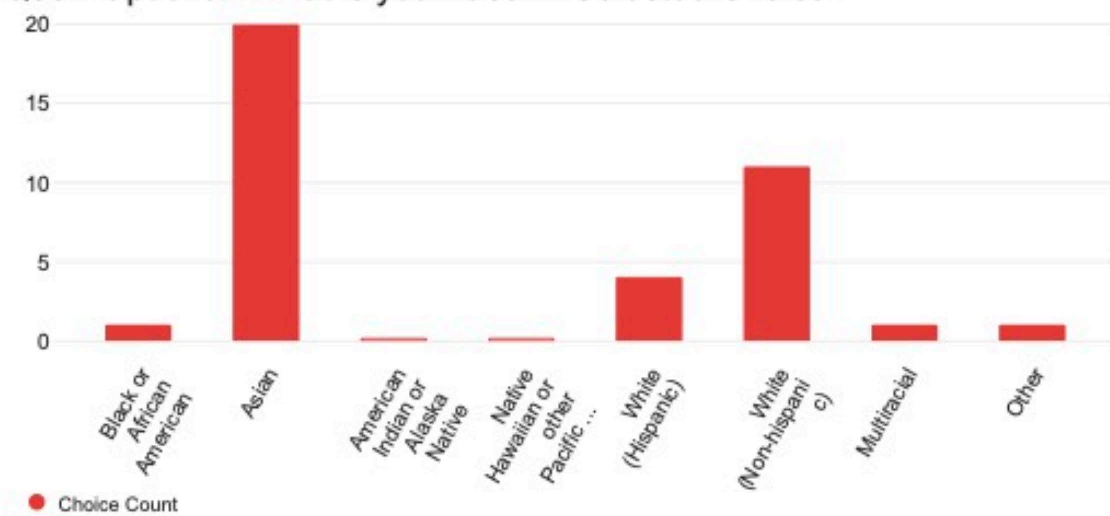


Q34 - Optional: What are your pronouns?

Optional: What are your pronouns?
He/him/his
He/Him
he/him/his
he,him,his
he him
she/her

She/her
He/Him
she
he/they
he/him
She/her/hers
He/Him
He/Him
he/him
she/her
she/her
She/her
Marco
She/ her

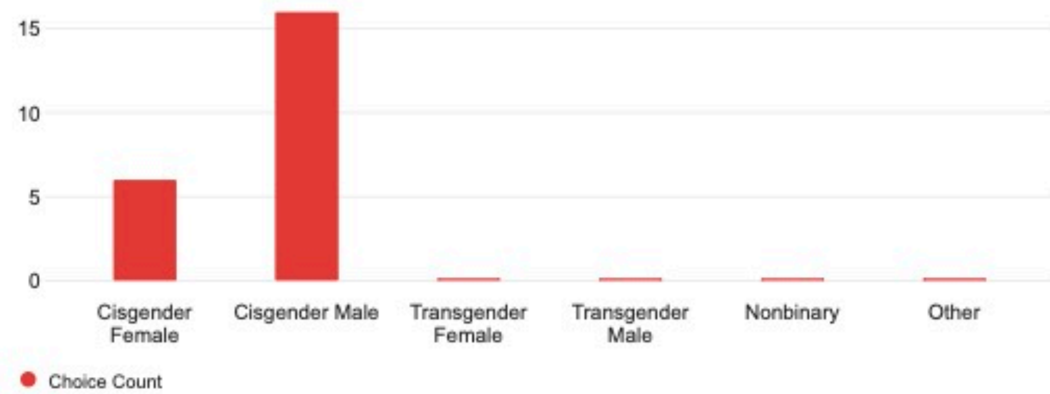
Q35 - Optional: What is your race? - Selected Choice



Appendix D: Faculty Survey Results

AY 23/24 Mathematics and Statistics Faculty Survey Results

Q36 - Gender - Selected Choice



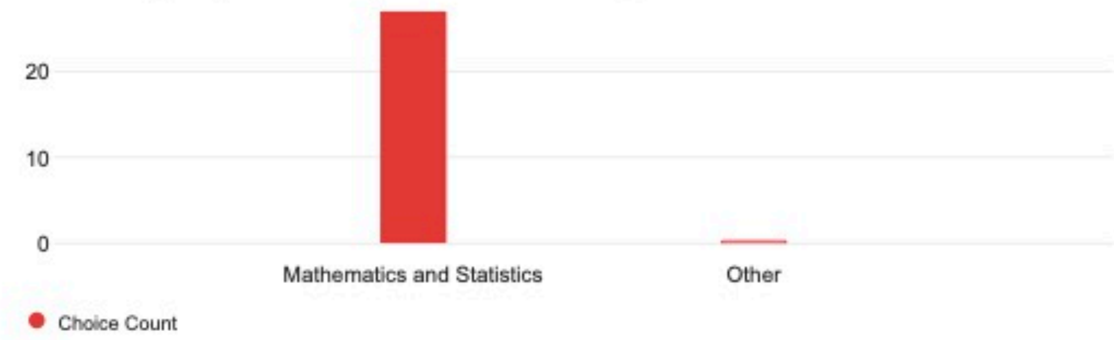
Q4 - What is your instructor title? - Selected Choice

Field	Choice Count
Assistant Professor	7
Associate Professor	5
Professor	13
Lecturer	0
Senior Lecturer	0
Master Lecturer	0
Affiliate/Visiting Professor	0
Part-time Lecturer or Adjunct Professor	0
Graduate Instructor (grad student as instructor of record)	0
Teaching Assistant or Teaching Fellow (graduate student)	0
Teaching Assistant or Teaching Fellow (undergraduate)	0
Other	2

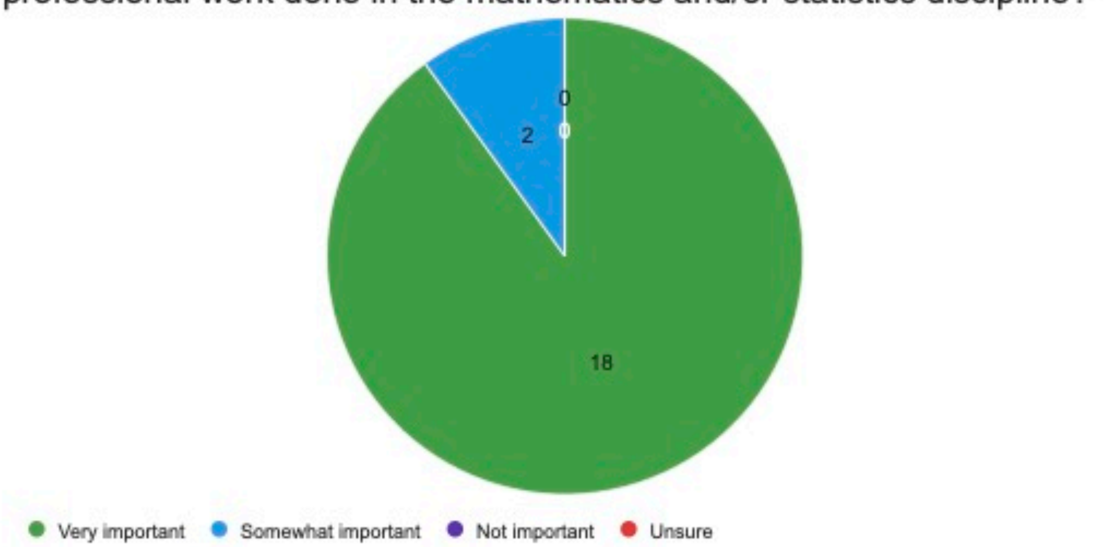
Q4_12_TEXT - Other - Text

Other - Text
Clinical Professor
Research Assistant Professor

Q5 - Please identify the academic unit at Boston University where you do the majority of undergraduate teaching. - Selected Choice



Q1 - How important is writing (as it is defined above) to the scholarly and professional work done in the mathematics and/or statistics discipline?



Q2 - Describe in a sentence or two why you answered the previous question the way you did.

Describe in a sentence or two why you answered the previous question the way you did.

Every product of "work" is written.

We make progress in our own understanding and further the field when we can communicate what we understand clearly to other people!

In statistics, the ability to express an idea properly is of utmost importance. It is as important as knowing the technical aspects of the subject.

I have found, throughout my own career, that the clear communication of mathematical results is nearly as important as the results themselves.

Grant and paper writing are the keys to success and require real skill.

Good proof writing is important

Writing is an important way to convey our findings to others.

Mathematics at its heart utilizes ideas, definitions, hypotheses, statements, and justifications, all of which require the ability to express such notions in a concise, logical, and precise manner in prose, often with picture, as well as equations.

Communicating well (writing included) is an essential skill to any professional, but it is often overlooked in math/stats.

Writing is a way to share your thinking with others, but it's also much more than that -- it's a way to improve your thinking and cut through the self-delusion we're all susceptible to that we understand things better that we actually do.

Written articles are a central, or perhaps the main, method of communicating mathematical and statistical methods and findings.

The archival literature, which is the record of our accumulated mathematical knowledge, should be carefully and thoughtfully written, or else it invites errors.

The purpose of statistics is to use data to understand the world better, and as such it is inseparable from the communication of that knowledge through writing

It is a daily requirement of all facets of my job.

Some people say Math *IS* a language.

Accurately and clearly community statistical concepts and results is crucial.

Math and statistics requires clear and precise communication of proofs, theoretical results, experiments using words and graphical representations of data and/or mathematical objects
Not only is mathematics communicated to others via writing, writing is an important tool for building ones own understanding.
Writing is one of the main ways that mathematicians communicate with each other, and it is extremely important for clarifying one's own thinking.
Effective mathematical writing requires a good balance between clarify/concision/intuition. The concepts and ideas are very abstract, so getting these across in an optimal way is no easy task.

Q10 - In your opinion, which of the following characteristics are particularly descriptive of writing in the mathematics and/or statistics discipline? (Select all that apply) - Selected Choice

Field	Choice Count
Expressive: emphasizing personal feelings and impressions	1
Interpretive and/or Evaluative of others' works and ideas	5
Descriptive: conveying processes, objects, data, environments, etc.	12
Analytical: emphasizing the logical examination of subjects or texts	18
Persuasive: presenting and evidencing positions or claims	9
Exploratory: investigating and developing ideas using discovery-based writing	9
Visual: emphasizing visual components such as graphic presentation, sketches, drawings, videos, etc.	13
Explanatory: restating complex content as generally comprehensible definitions and/or instructions	19
Innovative: approaching subject in fresh and inventive ways	12
Collaboratively authored	13
Reflective: applying lessons to one's own life; metacognition	2
Multimodal: communicating in more than one modality (ex. visual, linguistic, spatial, aural, and gestural)	8
Unsure	0

	5
Other	3

Q10_15_TEXT - Other - Text

Other - Text
All of the above
Um these terms are a bit weird.... but I think there is a lot of emotion involved as well... and it's great if we can make space for that, which usually we don't...
Note: "Collaboratively authored" is an option, often used in practice, but not a requirement.

Q11 - In the previous question, you indicated characteristics that are particularly descriptive of writing in mathematics and/or statistics. Which three characteristics seem the most important when describing writing in this field?

In the previous question, you indicated characteristics that are particularly descriptive of writing in mathematics and/or statistics. Which three characteristics seem the most important when describing writing in this field?
Descriptive, Analytical, Explanatory
Analytical, Persuasive, Explanatory
1. Descriptive, 2. Analytical, 3. Explanatory.
I think that the two I selected in the previous are the most important.
Analytical, descriptive, persuasive
Explain one's thought process in a clear way
Descriptive, analytical, visual
Analytical, explanatory, and descriptive
descriptive, analytical and explanatory
analytical, innovative, explanatory
Analytical, Explanatory, Descriptive
analytical, explanatory, innovative

Analytical, Visual, and Explanatory
Descriptive, interpretive, explanatory
All
Descriptive, Visual, Explanatory
analytical, visual, explanatory
Descriptive, Persuasive, Multimodal
Analytical, Explanatory, Innovative
descriptive/visual/explanatory

Q3 - Which of the following writing abilities should undergraduate students in mathematics and/or statistics be able to demonstrate by the time they graduate? (select all that apply) - Selected Choice

Field	Choice Count
Use field-specific terminology, organizational formats, and/or conventions	15
Argue a position using a central thesis or hypothesis and evidence	12
Create and incorporate visuals or presentation formats (figures, drawings, tables, photos, posters, slides)	12
Describe processes, objects, findings, environments, etc.	13
Summarize ideas, texts, or events	13
Analyze, interrogate, and/or evaluate ideas, texts, or events	10
Use correct grammar, spelling, mechanics (punctuation etc.), and notation	16
Propose innovative ideas or perspectives	9
Co-author texts with one or more writers	9
Report and explain complex data, findings, or figures, using examples as appropriate	14
Use writing to develop and deepen thinking	14
Synthesize disparate ideas and/or perspectives	8

Express feelings or impressions	1
Reflect upon experiences and/or assumptions	8
Solve complex problems	18
Integrate and correctly cite information from well-chosen sources	11
Apply theory, argument or findings to real-world circumstances	10
Contextualize your argument or findings	9
Other	1

Q3_19_TEXT - Other - Text

Other - Text
argue a position, develop and deepen thinking, solve complex problems

Q12 - Of the writing abilities identified in the previous question, which three are most critical for undergraduate students graduating with a mathematics and/or statistics major?

Of the writing abilities identified in the previous question, which three are most critical for undergraduate students graduating with a mathematics and/or statistics major?
.
Use correct grammar, argue a position, reflect on assumptions
1. Use correct grammar, 2. Argue a position, 3 Summarize ideas.
Solving complex problems, analyzing complex problems and ideas, and using correct grammar, spelling and notation.
Argue a position, summarize ideas, use correct grammar
Organize their ideas logically and be able to write reports reflecting that
Visuals, applications, argue a position ,
Use field-specific terminology, analyze ideas, and solve complex problems
synthesis, analysis, and depth of thinking

8

use writing to deepen thinking, analyze ideas, use correct notation
Solve complex problems, integrate and correctly cite information, summarize ideas
same as answer to the previous question. I'm not sure I understand the difference.
Using field-specific conventions, Report and explain complex findings, and create and incorporate visuals
Summarize, Describe, Report
All
Creating figures/tables, describing processes, applying theory
using correct terminology, report and explain complex data...; solve complex problems; apply theory, ... to real-world circumstances
Solve complex problems, Use writing to develop and deepen thinking, Argue a position (which I'm interpreting as proof-writing)
Use field-specific terminology....; Argue a position....; Solve complex problems....
Report and explain complex data .../solve complex problems/apply theory, argument ... to real-world circumstances

Q32 - Which of the following writing assignments have you incorporated in any of the undergraduate mathematics and/or statistics courses that you teach within the past year? (Select all that apply) - Selected Choice

Field	Choice Count
Logs, blogs, notebooks, or journals (paper or online)	4
Essays (personal, critical, analytical, argumentative)	6
Literature reviews or annotated bibliographies	1
Summaries or abstracts	4
Reports (lab, feasibility, progress, patient, etc., written by a single author)	8
Research papers	10
Professional communication (memos, correspondence, resumes, grant or conference proposals)	0
Literary work (poetry, fiction, drama, etc.)	0

	9
Brief, informal responses (written in or out of class)	8
Presentations (oral, PowerPoint, scientific poster, etc.)	11
Problem sets, equations, or proofs	19
Informational brochures or newsletters, etc.	1
Drawings, illustrations, technical specifications, etc.	2
Web pages	2
Other (please specify)	1

Q32_15_TEXT - Other (please specify) - Text

Other (please specify) - Text
I've used portfolios with reflective essays, have had students set their own goals for growing as a mathematician and keep data and report on them. I also like to do papers where students solve a problem and combine writing up a convincing solution and reflecting on their problem solving process. Not sure if this belongs on the list, but I've found peer editing really helpful here as well. I also like to have students reflect on an interesting mistake -- and in fact require that, to normalize making mistakes as part of the process of doing mathematics.

Q13 - Considering the writing students do for undergraduate courses in mathematic...

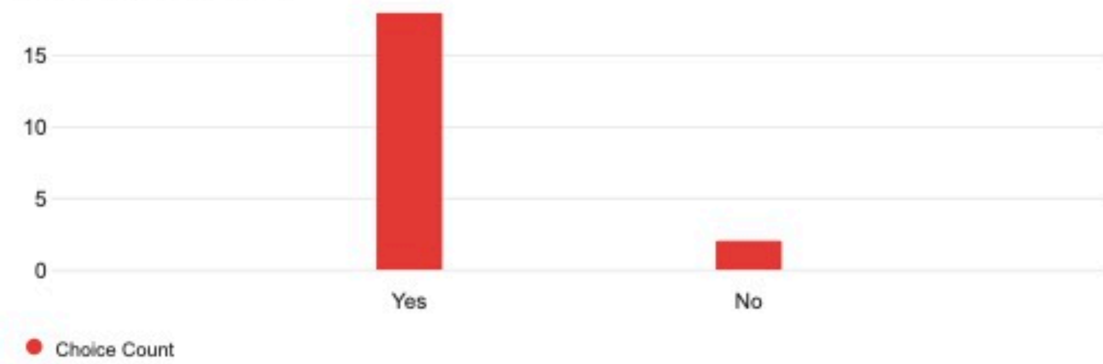
Field	Strong	Satisfactory	Weak	Don't know/Unable to Generalize	N/A
Use field specific terminology, organizational formats, and/or conventions	0	6	12	2	0
Argue a position using a central thesis or hypothesis and evidence	0	5	12	2	1
Create and incorporate visuals or presentation formats (figures, drawings, tables, photos, posters, PowerPoint, etc.)	1	6	6	5	2
Describe processes, objects, findings, environments, etc.	0	8	8	2	2
Summarize ideas, texts, or events	0	7	8	5	0

Analyze, interrogate, and/or evaluate ideas, texts, or events	0	5	7	6	2
Use correct grammar, spelling, and mechanics (punctuation, etc.)	0	11	7	2	0
Propose innovative ideas or perspectives	0	5	9	5	1
Co-author texts with one or more writers	0	6	5	5	4
Report and explain complex data or findings	0	7	9	3	1
Use writing to develop and deepen thinking	0	3	9	8	0
Synthesize disparate ideas and/or perspectives	0	3	7	9	1
Express feelings or impressions	0	4	2	8	6
Reflect upon experiences and/or assumptions	0	4	4	8	4
Solve complex problems	1	11	6	2	0
Integrate and correctly cite information from well-chosen sources	0	9	8	2	1

Q21 - During the past academic year, how frequently have you graded undergraduate...

Field	Always	Sometimes	Never	Don't Know
Assigned a number or letter grade without comments	0	6	12	2
Assigned a number or letter grade with comments	7	8	3	2
Assigned grade and made comments, using a list of criteria or grading rubric	1	11	6	2
Used a grading contract and did not not assign a specific number or letter grade	1	3	13	3
Assigned grades to a portfolio of student selected work	0	1	16	3
Incorporated students' self-assessment or reflection into grading formula	0	3	14	3
Provided and discussed grading criteria before assignment was due	6	8	4	2

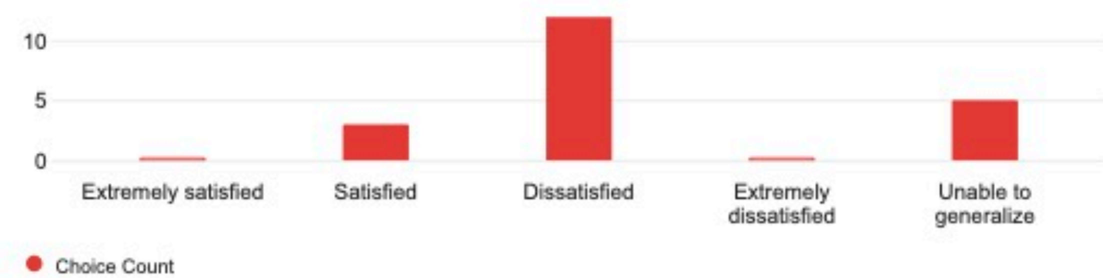
Q39 - Do you employ TAs, TFs, or graders in any of the undergraduate classes you teach?



Q40 - During the past academic year, how frequently have you graded undergraduate...

Field	Always	Sometimes	Never
Provided TAs with criteria with which to grade student writing	3	9	5
Held "norming sessions" with TAs as they prepared to grade student writing	1	5	11

Q22 - To what degree are you satisfied with the overall quality of undergraduate student writing in the courses you teach in mathematics and/or statistics?



Q23 - Is there anything you would like to add about the overall quality of undergraduate student writing?

Is there anything you would like to add about the overall quality of undergraduate student writing?

I haven't taught undergrad courses at BU yet, not sure if my answers are helpful!

It is impossible, in my opinion, to provide the detailed feedback needed to help students improve their writing in sections of 120-150 students.

Most in need of remediation is grammer and sentence structure

Most of undergraduates don't even use english when they write math, only symbols and equations...

I have graded together with TAs, where we went over graded reports together before finalizing grades. I think most students are able to convey their thoughts, although grammar is sometimes a problem and that might hinder them in future jobs.

In lower level classes, students are conditioned to manipulate formulas with little justification, logic, or analysis partly because they are not trained to write and think about the meaning and the domain of validity of the formulas.

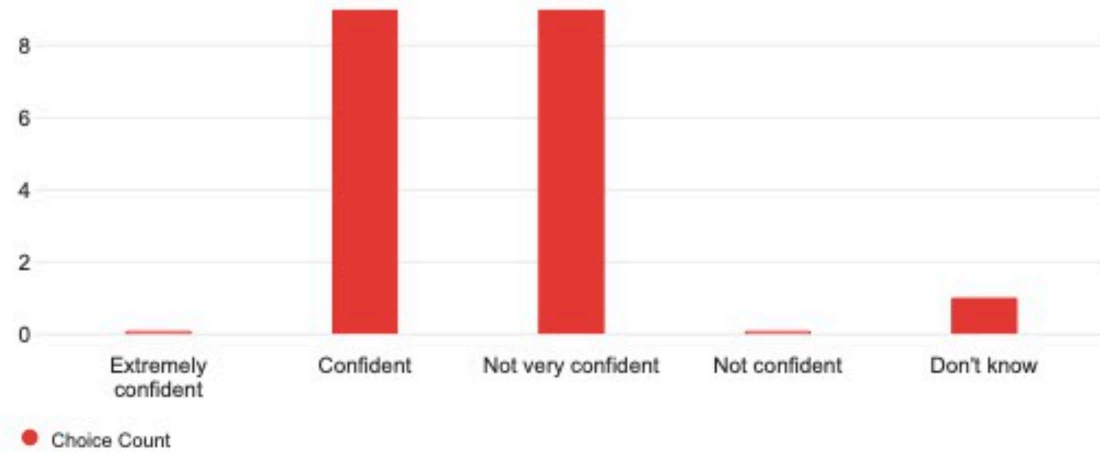
I find the above question hard to answer.... I see students improve a lot.... so i'm often satisfied at the end of the class, not so at the beginning

I am having a hard time taking these questions seriously, because the range of ability is huge. Some students do wonderfully well, others less so.

(While I lead MA213, which has a lab project, I am not the lab instructor and I don't deal with or grade the student writing)

I think there is a big difference between the abilities (and expectations) of students in intro classes, like MA 123, and upper level proof courses. Classes like MA 123 which have so many students make it hard to give substantive feedback on writing on a regular basis (and we've used automated online homework systems which don't require any writing to be submitted). In that context, I wonder how we can raise expectations while simultaneously giving students the support they need to grow and improve.

Q24 - How confident are you in your ability to help undergraduate students to meet writing expectations in mathematics and/or statistics?



Q29 - Is there anything you would like to add about your confidence level?

Is there anything you would like to add about your confidence level?
I wish I could do more to help undergrads to recognize "types" of math objects and how they fit together and use them in a gramatically correct sentence.
I believe that I can help my students improve their writing in small sections, but not in the large sections in which we currently teach calculus and some other undergraduate courses.
The answer depends on so many factors, I would think. If primary math classes were to consistently emphasize writing then it would be much easier than if one instructor were to emphasize writing, obviously. If we were able to get more support particularly for larger classes then that would also be helpful, if not necessary. In the absence of both, perhaps a smaller undergraduate course could successfully emphasize writing, particularly the more advanced proof based classes. How much work this would entail for the instructor is not easy to estimate.
You can lead a horse to water but you can't make him drink. I am not too confident of my ability to make horses drink.
I'm confident I could do this in advanced undergraduate classes where there are fewer than 30 students, it seems difficulty to do this effectively (except for requiring short explanations of solutions with complete sentences) for large lecture courses of 100+ students.
I simply don't know at this point how large the task is (i.e. how much effort it would take), and therefore find it difficult to speculate on my level of confidence

Q33 - Would you be interested in learning more about any of the following topics related to writing instruction in your field? - Selected Choice

Field	Choice Count
Incorporating brief in-class writing instruction into class activity	12
Designing effective, well-scaffolded, course-relevant writing assignments	12
Providing useful feedback on drafts	12
Organizing peer review activities	12
Working with multilingual writers	8
Grading in ways that are efficient and fair	9
Addressing grammar, usage, and mechanics	5
Teaching with writing in new media environments	3
Supervising Teaching Assistants/Fellows	8
Avoiding and detecting plagiarism	9
Using ungraded writing exercises to teach course content	8
Incorporating student's reflection on their own learning	8
Other	1

Q30 - What other comments would you like to make about the role or importance of writing in mathematics and/or statistics?

What other comments would you like to make about the role or importance of writing in mathematics and/or statistics?
As a profession, I think we have paid too little attention to this in the past, and I hope that we can improve going forward.

I think it's incredibly important and am glad you're doing this... if I weren't on sabbatical I'd be more involved. One thing I'd like to say is that I think it's really important not to just jump to formal theorem/proof style and to have other formats that incorporate more of process.... I think there's this gray zone where students want to say, "I see a pattern, therefore this is the answer," and it's a big cognitive shift for them to explain why this is the RIGHT pattern for this problem. Not sure how well this is even articulated in the literature, but I saw this at Wheelock a lot too.... I think something really valuable about writing is to help students get out of them memorizing/relying on authority/ jumping to conclusions phase and to see that they need to make arguments.... but if we start with theorem/proof then they can just try to memorize how to do that... not sure if this makes sense in this little field where I can't see what I've written.... but say in Mat 293 I see a lot of students who have learned induction in other classes, but what they've learned is that you have an equation and you add the same thing to both sides.... most have no idea that induction is about proving a sequence of statements.... but when I have students solve problems and explain why their patterns are the right ones for the problem then by the time we get to formalizing induction, they see that it's a way to organize their thinking for what they were trying to articulate before (well, some of them do)... of course, that's a particular type of problem, but perhaps the thought generalizes

It's underrated.

Appendix E: Faculty Resources

Content:

Some Guidelines for Good Mathematical Writing

By Francis Edward Su

President's Message

Communicating mathematics well is an important part of doing mathematics. Many of us know from writing papers or giving talks that communicating effectively not only serves our audience but also clarifies and structures our own thinking. There is an art and elegance to good writing that every writer should strive for. And writing, as a work of art, can bring a person great personal satisfaction.

Within the MAA, we value exposition and mathematical communication. In this column, I'm sharing the advice I give my students to help them write well. There are more extensive treatments (e.g., see Paul Halmos's *How to Write Mathematics*), but I wanted a shorter introduction. So I developed the guidelines below.

Basics

Know your audience. This is the most important consideration for writers. Put yourself in your reader's shoes. What background can we assume of the reader? What terminology should we define? What kind of "voice" do we want to project: casual or professional, serious or inviting, terse or loquacious?

If you are a student writing solutions for a homework set and your professor has not specified your audience, a good rule of thumb is to assume you are writing to another student in the course who has not yet done the assignment. Though you may assume that she has attended all the same lectures and has read the same textbook, it is standard courtesy to remind your readers of any relevant items that they have learned recently from the class or textbook, or things they should know but might have forgotten.

For instance, if the concept of a rational number was only recently learned in class, you might insert "Recall that a rational number can be expressed as a fraction" before saying "since x is rational, $x = m/n$ where m and n are integers."

Set an invitational tone. It is traditional to create an inviting atmosphere in one's mathematical writing. In effect, we invite readers to join us in our reasoning process by writing in the present tense, using the pronoun "we" instead of "I" (e.g., "we construct a tangent plane..."), and directing the reader with gentle commands (e.g., "let n be . . . ," "recall that . . . ," or "consider the set of . . .").

Use complete sentences. All mathematics should be written in sentences. Open any mathematics text and you'll see that this is true. Equations, even displayed ones, have punctuation that helps you see where it fits in the context of a larger sentence. Consider this piece of writing:

$$(x-2)^2 + (x-1)^2 = 5^2\ 5^2 = 25$$
$$(x-2)^2 = x^2 - 4x + 4 + x^2 - 2x + 1 = 25.$$
$$2x^2 - 6x - 20$$
$$2(x+2)(x-5)\ x = -2, 5\ x > 0\ x = 5$$

Can you figure out what the writer is doing? What's being assumed? What's being proved? Where does one thought end and another begin? What's the relationship between these phrases? Some phrases are dangling, and others, as statements, are not even true. The reader should not have to figure out what the writer was thinking.

Now consider the work of another writer attempting the same problem:

> **Problem.** Find a point on the line $y = x$ that is distance 5 from the point (2,1) and whose x-coordinate is positive.
>
> **Solution.** The desired point is (5,5). To see this, we solve $(x-2)^2 + (x-1)^2 = 5^2$, an equation obtained from the distance formula in the plane. A little algebra turns this equation into:
>
> $$2x^2 - 6x - 20 = 0.$$
>
> Factoring the left side, we obtain
>
> $$2(x+2)(x-5) = 0,$$
>
> whose solutions are $x = -2$ and $x = 5$. Since we assumed $x > 0$, we have (5,5) as the desired point on the line $y = x$.

Here, the writer has clearly stated the problem and described her path to a solution. She has set an invitational tone, and every thought is expressed in a complete sentence. Now it is clear that $x > 0$ is a condition, not a result. Notice the punctuation in equations: one ended with a period because her thought was complete, the other ended with a comma because she wanted to continue the thought.

Since she assumed her audience could do algebra, she didn't bore them with algebraic manipulation, which would obscure the thread of her arguments. But she did show the crucial and most interesting piece: the factoring and its result. And she made sure she answered the original question.

Use words to give context to equations. Consider the difference in meaning between these three statements: "Let $A = 5$." "Suppose $A = 5$." "Therefore $A = 5$." Then reflect on the ambiguity of the statement "$A = 5$."

Avoid shorthand in formal writing. The many types of mathematical writing can be loosely grouped into formal and informal writing. Informal writing includes writing on a blackboard during lecture, or explaining something to a friend on a piece of scratch paper. Formal writing includes the kind of writing expected on a homework assignment or in a paper. There are differences in what is acceptable. For instance, in informal writing, it is common to use shorthand for quantifiers and implications: symbols such as $\forall$, $\exists$, $\Rightarrow$, $\Leftrightarrow$, or abbreviations such as "iff" and "s.t."

However, in formal writing, such shorthand should generally be avoided. You should write out "for all," "there exists," "implies," "if and only if," and "such that." Most other symbols are acceptable in formal writing, after defining them where needed. The membership symbol $\in$ is traditionally acceptable in formal writing, as are relations (e.g., $<$, $+$, $\cup$, etc.), variable names (e.g., x, y, z), and symbols for sets (e.g., $\mathbb{R}$). Here is an acceptable use of symbols in formal mathematical writing:

> Let A and B be two subsets of $\mathbb{R}$. We say A *dominates* B if for every $x \in A$ there exists $y \in B$ such that $y > x$.

Learn the etiquette. The above example also illustrates two common conventions of mathematical etiquette. It is customary to avoid beginning sentences or phrases with a number or symbol because that can be confusing. It is also customary to emphasize unfamiliar words that we are about to define, such as by italicizing them. Other etiquette can be learned by observing the norms used in your area of study.

Toward Elegance

Decide what's important to say. Writing well does not necessarily mean writing more. If your solution is too wordy, it can sometimes obscure the points you are making.

A well-written solution will present just enough details and highlight the most interesting or unexpected parts of the argument. What theorems or axioms were crucial in getting your solution, and where were they used? Your role as a writer is not primarily to give details (though that can be important). Your primary role is to give insight.

Highlight structure. If your argument is going to be a long one, with lots of technical details, then try to help the reader by summarizing the outline of your argument at the beginning. Then, throughout your writing, help your reader see how you are progressing through your outline.

Use paragraphs to emphasize blocks of ideas that are related. The role of the first sentences of paragraphs is crucial: imagine a reader skimming your writing and reading only the first sentences. Will she see the flow of your argument? Similarly, you might want to display only the most important equations. Replacing an oft-repeated argument by a good lemma can streamline the flow as well as highlight a key idea.

Choose good examples. A difficult idea may be easier to digest if accompanied by an example. Choose one simple enough to follow, but interesting enough to retain the salient features. A proof of a very general idea could be preceded by an example in a specific context. A long exposition might benefit from a running example—one in which the same example is used multiple times in different contexts.

Avoid red herrings! Omit details that have no bearing on the solution of the problem, because they may throw the reader off. For example, if you say "we express the rational r as m/n where m and n have no common factors," you are leaving a clue that later you will use the "no common factors" idea. So if you never use that fact, you should omit saying it. It's extraneous. Red herrings may make mystery novels fun, but in mathematical writing, your goal is to dispel mystery!

Step back and simplify. After writing a proof, step back and ask: How can I simplify this argument? Did I use every tool I pulled out to solve this problem? Can I streamline this argument? For example, consider this apparent proof by contradiction:

> **Problem.** Show that if 4 divides an integer n, then n is even.
>
> *Proof.* Suppose n were not even.
> Since 4 divides n, we have $n = 4k$ for some integer k.
> Thus $n = 2(2k)$, which is even.
> This contradicts our hypothesis that n is not even. QED.

Do you see why this is not really a proof by contradiction? The contradictory hypothesis in the first sentence was never used! Strip away the first and last sentence, and you have an elegant, direct proof.

Refine, refine, refine. Good writing is a process of successive approximations. You should not expect your first draft to be perfect. You will find that when you review your writing, you will see ways to shorten an argument or say something in a better way. This is the

part of the writing process that will help clarify your own thinking as well.

Often, after completing a draft, a writer may notice that a particular choice of notation or definition was not optimal. A lazy writer would leave things as they are, but a thoughtful writer will take the time to go back and make changes.

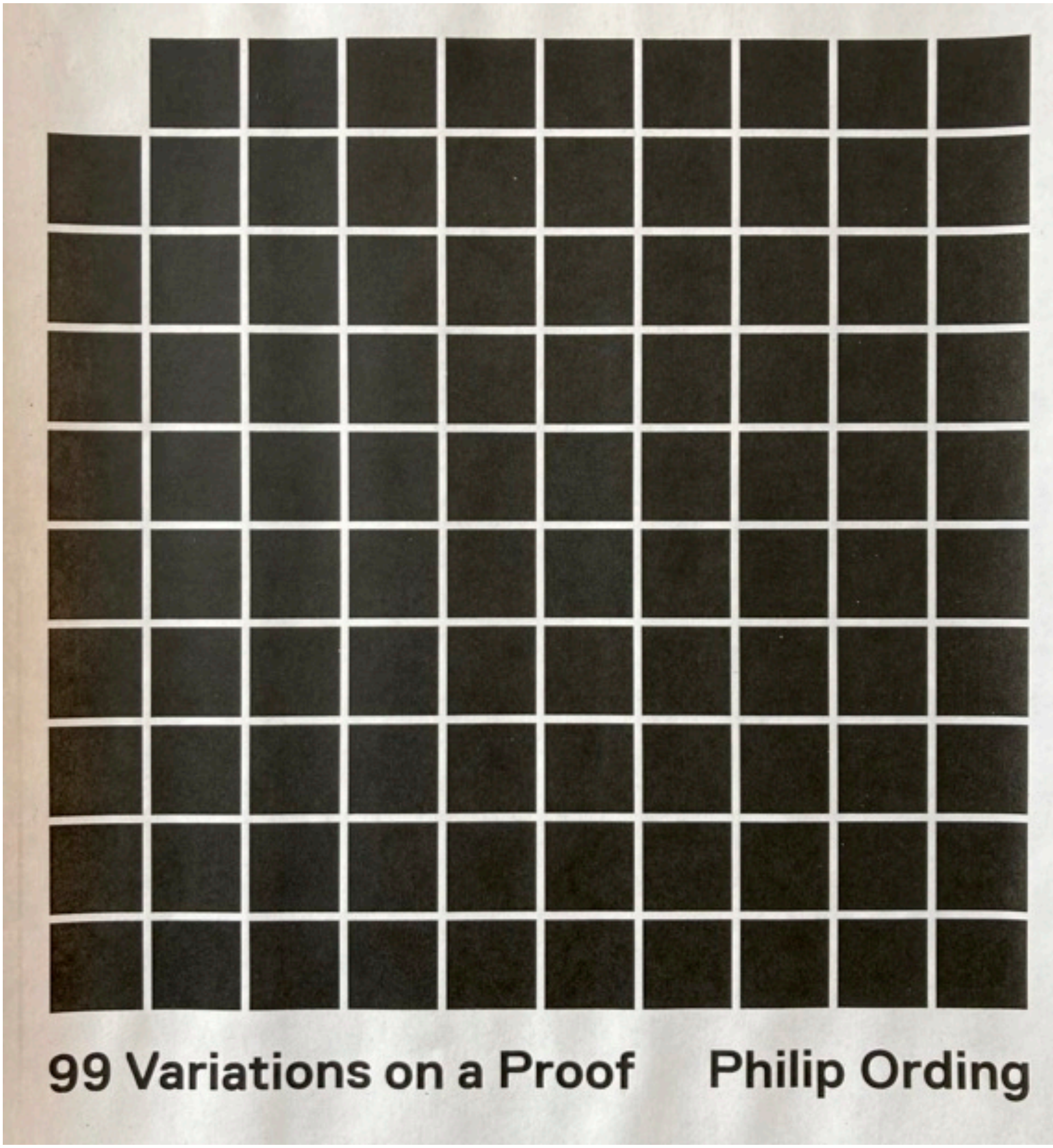
Observe the culture. Good communication is inseparable from the culture in which it takes place. This realization may unsettle budding mathematicians who are attracted to the logical absolutes of mathematics. But even these absolutes are expressed differently by mathematicians of different eras, as can be seen by comparing Newton's writings with any of today's calculus texts. The rules of mathematical etiquette have evolved.

Although these guidelines attempt to draw up some common principles for formal writing, there will always be exceptions—because some mathematical field may have a slightly different norm. The best way to get a sense of what is acceptable in your context is to browse several highly regarded texts or papers related to the document you are writing.

Enjoy the art of writing. Writing is an occasion to reflect on beautiful ideas and paint them on a paper canvas with great artistic care.

Acknowledgments. It is appropriate to acknowledge any support you received. My revisions of these guidelines benefited greatly from the helpful feedback of Jon Jacobsen and Lesley Ward.

Francis Edward Su is president of the MAA. His email is su@math.hmc.edu. And he is on Twitter: @mathyawp.


99 Variations on a Proof
Philip Ording

Theorem. *Let x be real. If $x^3-6x^2+11x-6=2x-2$, then $x=1$ or $x=4$.*

Proof. By subtraction, $x^3-6x^2+9x-4=0$, which factors as $(x-1)^2(x-4)=0$. □

1

One-Line

Hypothesis: $x^3-6x^2+11x-6=2x-2$, where x is a real number.
To prove: $x=1$ or 4.

	STATEMENT	REASON
1.	$x^3-6x^2+11x-6=2x-2$	Given.
2.	$x^3-6x^2+11x-6+2=2x-2+2$	Addition property of equations.
3.	$x^3-6x^2+11x-4=2x$	Addition.
4.	$x^3-6x^2+11x-4-2x=2x-2x$	Subtraction property of equations.
5.	$x^3-6x^2+9x-4=0$	Subtraction.
6.	$x^3-(1+5)x^2+(5+4)x-4=0$	Addition.
7.	$x^3-x^2-5x^2+5x+4x-4=0$	Distributive property.
8.	$x^2(x-1)-5x(x-1)+4(x-1)=0$	Factoring.
9.	$(x^2-5x+4)(x-1)=0$	Factoring.
10.	$[x^2-(1+4)x+4](x-1)=0$	Addition.
11.	$(x^2-x-4x+4)(x-1)=0$	Distributive property.
12.	$[x(x-1)-4(x-1)](x-1)=0$	Factoring.
13.	$[(x-4)(x-1)](x-1)=0$	Factoring.
14.	$x-1=0$ or $x-4=0$	Zero product property.
15.	$x-1+1=1$ or $x-4+4=4$	Addition property of equations.
16.	$x=1$ or $x=4$	Addition.
		QED

3

Illustrated

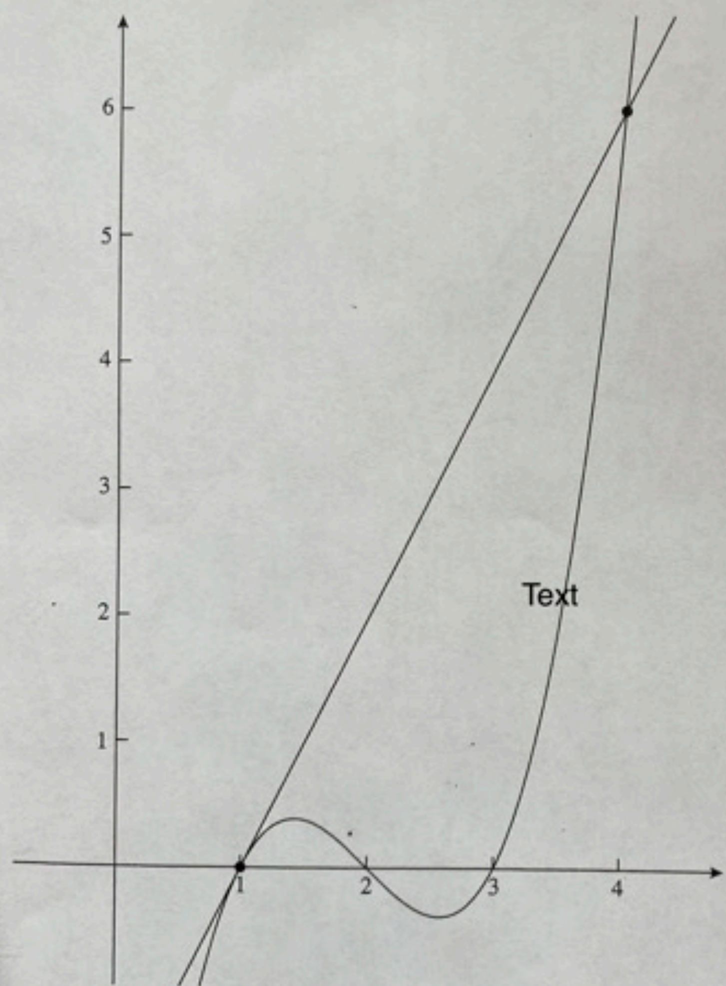


FIGURE. The two points of intersection of the cubic $y=x^3-6x^2+11x-6$ and the line $y=2x-2$ occur at $(1,0)$ and $(4,6)$.

Suppose that among four consecutive numbers, the product of the first three equals twice the third. What's the fourth number?

5

Puzzle

Answer: 4

9

Monosyllabic

Here is a fact:

If x is real and the cube of x less six times the square of x plus five times x plus six times x less six is twice x less two, then x must be one or four.

The proof goes like this:

See, the first three terms on the left side split as the square of x less five times x all times x less one. And more, the last two terms on the left side split as six times x less one, while the right side splits as two times x less one. Thus, if x were to be one, we have nought plus nought is nought, which is true. So, x may be one.

Else x is not one, and x less one is not nought. So we can times the whole thing by one on top of x less one to yield: the square of x less five times x plus six is two. Drop two from each side, and the square of x less five times x plus four is nought. Now this splits as x less four times x less one. Since we said x less one is not nought, x less four must be. So x is one or four, as was to be shown.

11

Exam

Instructions: Write your answer to the question directly on this page. All work should be written in blue or black pen. Clearly show all appropriate algebraic work. Scrap paper is not permitted, but you may use the reverse side of this page for scratch work. The use of any communications device is strictly prohibited when taking this examination. If you use any communications device, no matter how briefly, your examination will be invalidated, and no score will be calculated for you.

Solve $x^3-6x^2+11x-6=2x-2.$

$$x^3-6x^2+11x-6 = 2x-2$$
$$-2x+2 \quad -2x+2$$

$$x^3-6x^2+9x-4=0$$
$$\cancel{x^2(x-6)+9(x-4)=0}$$

$$\frac{x^3-6x^2+11x-6}{(x-1)} = \frac{2(x-1)}{(x-1)}$$

$$\begin{array}{r} x^2-5x+6 \\ x-1 \overline{)\, x^3-6x^2+11x-6} \\ -x^3+x^2 \\ \hline -5x^2+11x \\ +5x^2-5x \\ \hline 6x-6 \\ -6x+6 \\ \hline 0 \end{array}$$

$$x^2-5x+6=2$$
$$\cancel{(x-2)(x-3)=2}$$
$$x^2-5x+4=0$$
$$(x-1)(x-4)=0$$

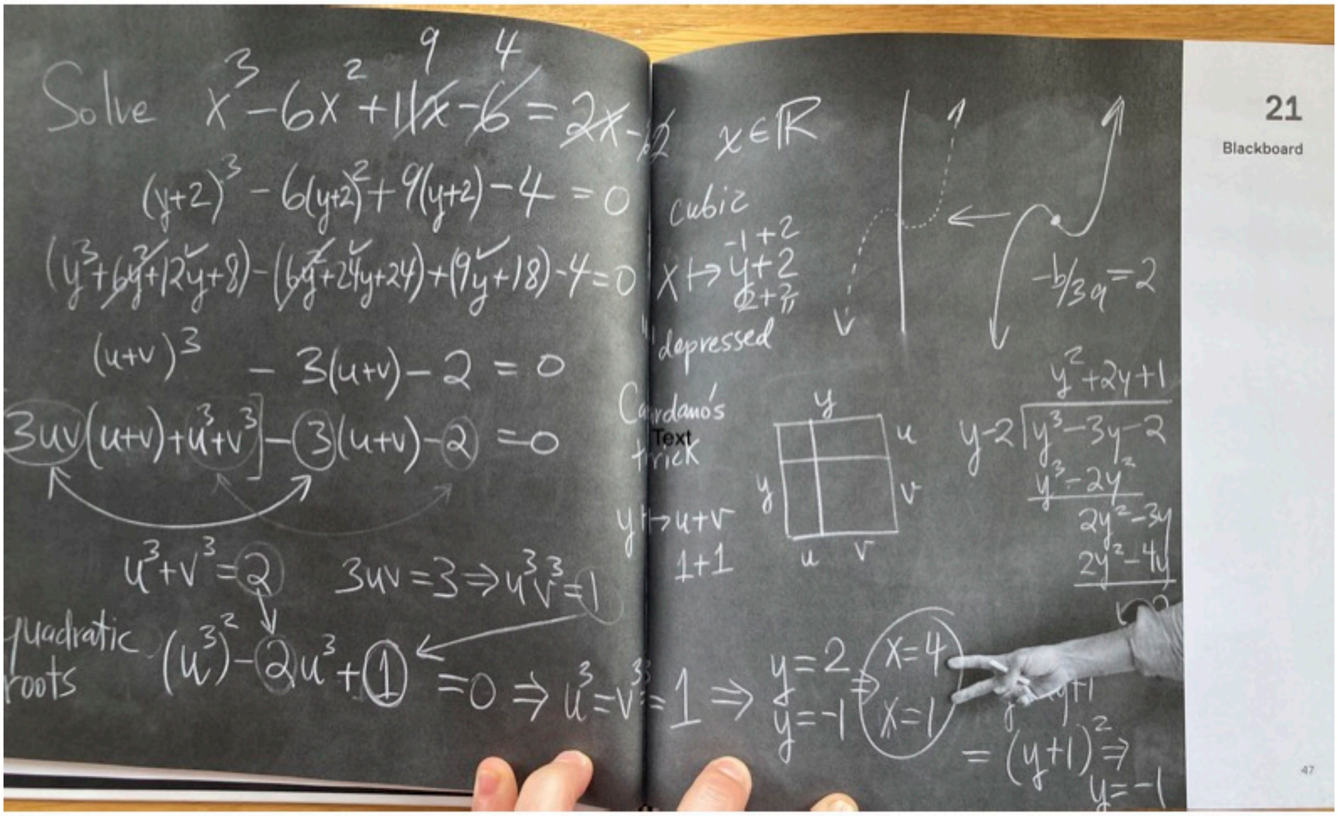
$$x=1$$
$$x=4$$

Theorem. *Let x be a real number. If $x^3 - 6x^2 + 11x - 6 = 2x - 2$, then $x = 1$ or $x = 4$.*

Proof. That $x = 1$ and $x = 4$ are solutions is readily verified. Suppose there is a third solution x such that $x \neq 1$ and $x \neq 4$. Then we may divide $x^3 - 6x^2 + 11x - 6 = 2x - 2$ by $x - 1$ to obtain $x^2 - 5x + 6 = 2$, that is, $x^2 - 5x + 4 = 0$. Now dividing by $x - 4$ leaves $x - 1 = 0$. Dividing again by $x - 1$ we conclude $1 = 0$, which is absurd. Hence $x = 1$ or $x = 4$ as required. □

13

Reductio ad Absurdum


21
Blackboard
Solve
cubic
depressed
Cardano's
Text
trick
quadratic
roots
y=2
y=-1
x=4
x=1
47

25

Open Collaborative

Problem: Prove that if $x \in \mathbb{R}$ and $x^3-6x^2+11x-6=2x-2$, then $x=1$ or $x=4$.

1. Just to get things going, it's perhaps worth noting that, in standard form, the polynomial equation is equivalent to

$$x^3-6x^2+9x-4=0.$$

Comment by Alpha, June 25 @ 5:03pm | Reply

1.1. You could also express it in Horner's (another "standard") form

$$((x-6)x+9)x-4=0.$$

Comment by Beta, June 25 @ 5:29pm | Reply

2. I'm sure others have already tried this, but I'm just wondering whether it can't be factored

$$x(x^2-6x+9)-4=0$$
$$x(x-3)^2=4.$$

Comment by Gamma, June 25 @ 5:35pm | Reply

2.1. I was just thinking along the same lines, but thought it would be cheating to use the factor theorem, which, knowing the solutions, gives

$$(x-1)(x-4)(x-4)=0.$$

Comment by Delta, June 25 @ 5:36pm | Reply

2.1.1. Sorry, but why are there three solutions? And isn't 1 the repeated root?
Comment by Epsilon, June 25 @ 5:45pm | Reply

2.1.2. That follows from the Fundamental Thm of Algebra, but using such a big hammer for this problem seems like another kind of cheating. Oops! You're right, should be $(x-1)(x-1)\,(x-4)=0$.
Comment by Delta, June 25 @ 5:50pm | Reply

3. One usually solves quadratics that don't obviously factor by "completing the square." Is there such a thing as "completing the cube"?? Not sure this is a real strategy as yet.
Comment by Zeta, June 25 @ 5:59pm | Reply

Open
Collaborative

3.1. The cubic binomial expansion is $(x+a)^3 = x^3 + 3ax^2 + 3a^2x + a^3$. Given that our quadratic coefficient is -6, we ought to take $a = -2$, so

$$x^3 - 6x^2 + 9x - 4 = (x-2)^3 - 3x + 4.$$

Seems like there's no guarantee that we can complete a cube, at least not by simply adding a constant.
Comment by Alpha, June 25 @ 6:19pm | Reply

4. Actually, I just noticed that the original equation factors very neatly

$$(x-1)(x-2)(x-3) = 2(x-1).$$

Comment by Gamma, June 25 @ 6:20pm | Reply

4.1. Canceling the common factor $(x-1)$ gives a quadratic with the desired solutions!
Comment by Epsilon, June 25 @ 6:22pm | Reply

5. I want to go back to Alpha's last comment. If you make the change of variable $y = x - 2$ then you do at least get a cubic without a square term,

$$y^3 - 3y - 2 = 0.$$

This seems important.
Comment by Zeta, June 25 @ 6:41pm | Reply

5.1. Tau calls this a "depressed cubic". Here's the relevant paper.
Comment by Eta, June 25 @ 6:44pm | Reply

5.1.1. Thanks Eta, I hadn't seen this before!
Comment by Zeta, June 25 @ 6:47pm | Reply

5.2. If you take $z = x + 1$, the linear and constant term drop out

$$z^3 - 3z = 0$$

and it's easily factored. Is this cheating?
Comment by Gamma, June 25 @ 6:51pm | Reply

5.3. I don't really have an idea here, but the 3 seems significant.

$$y^3 = 3y + 2$$

Can we use it to kill more terms?
Comment by Alpha, June 25 @ 6:52pm | Reply

Open
Collaborative

5.3.1. Maybe this factored form of the cubic binomial expansion helps?

$$(x+a)^3 = x^3 + 3xa(x+a) + a^3$$

Comment by Beta, June 25 @ 7:22pm | Reply

5.3.2. Ah, yes! Let's take $y = u + v$, then

$$(u+v)^3 = 3uv(u+v) + u^3 + v^3$$

looks very much like our equation indeed

$$y^3 = 3y - 2.$$

Now we need to solve the system

$$uv = 1$$
$$u^3 + v^3 = 2.$$

Comment by Alpha, June 25 @ 7:25pm | Reply

5.3.3. That's a quadratic in u^3. Substitute $v = 1/u$ into the second equation

$$u^3 + \frac{1}{u^3} = 2$$
$$u^6 - 2u^3 + 1 = 0$$

with solution $u^3 = 1$, which implies $v = 1$ and $y = 2$. We have a solution!
Comment by Zeta, June 25 @ 7:28pm | Reply

5.3.4. Since $y = x - 2$, that gives the root $x = 4$.
Comment by Alpha, June 25 @ 7:29pm | Reply

5.3.5. Dividing out the original cubic by $(x-4)$ leaves $(x-1)^2 = 0$.
Comment by Zeta, June 25 @ 7:31pm | Reply

. That's it, we have a proof!
Comment by Alpha, June 25 @ 7:32pm | Reply

33
Calculus

Theorem. *Define* $f:\mathbb{R}\to\mathbb{R}$ *to be the function* $f(x)=x^3-6x^2+9x-4$. *If* $f(x)=0$, *then* $x=1$ *or* $x=4$.

Proof. Let's construct the Taylor series of f about the first of the two purported roots, $x=1$. Compute the derivatives of f: $f'(x)=3x^2-12x+9$, $f''(x)=6x-12$, $f'''(x)=6$, and $f^{(n)}=0$ for $n\geq 4$. Hence

$$\begin{aligned} f(x) &= f(1)+\frac{f'(1)}{1!}(x-1)+\frac{f''(1)}{2!}(x-1)^2+\frac{f'''(1)}{3!}(x-1)^3+\cdots \\ &= 0+0-3(x-1)^2+(x-1)^3+0 \\ &= (x-1)^2(-3+x-1) \\ &= (x-1)^2(x-4). \end{aligned}$$

Thus, the roots of f are 1 and 4, as claimed.

40

Induction

Theorem. *If n is a natural number and $n^3-6n^2+11n-6=2n-2$, then $n=1$ or $n=4$.*

Proof. Direct calculation shows that the cubic equation is satisfied for $1\leq n\leq 4$ only if $n=1$ or $n=4$. It remains to show that $n^3-6n^2+11n-6\neq 2n-2$ for all $n\geq 5$. Let $P(n)$ be the proposition $n^3-6n^2+11n-6>2n-2$. We will show $P(n)$ by induction on $n\geq 5$.

The base case: For $n=5$, $n^3-6n^2+11n-6=24>8=2n-2$. Thus $P(5)$ is true.

The inductive step: Observe that

$$
\begin{aligned}
(n+1)^3-6(n+1)+11(n+1)-6(n+1) &= (n^3-6n^2+11n-6)+(3n^2-9n+6)\\
&= (n^3-6n^2+11n-6)+3n(n-3)+6\\
&> (n^3-6n^2+11n-6)+6 \quad \text{since } n>3\\
&> (2n-2)+6 \quad \text{by the inductive hypothesis}\\
&> 2(n+1)-2
\end{aligned}
$$

and thus $P(n+1)$ is true. Then, by induction, we conclude that $P(n)$ is true for every natural number $n\geq 5$. Hence, $n=1$ and $n=4$ are the only solutions, as was to be shown. □

44

Omitted with Condescension

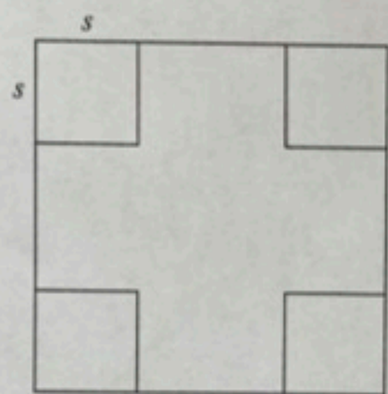
There is a simply beautiful theorem which provides all solutions of the equation $x^3-6x^2+11x-6=2x-2$. Alas, any further explanation would deny you the satisfaction of discovering it on your own....

68

Word Problem

A small box is to be made by cutting out four equal squares from a sheet of cardboard that measures 6 in by 6 in and then folding the resulting flaps. If the box is to have a volume of 16 in^3, show that the length of the sides of the cut-away squares must be 1 in.

Solution. Let s be the length measured in inches of the side of the cut-away square, as shown in the diagram. Our task is to show that $s=1$. After cutting away a square from each corner, the width of each flap is $6-2s$, since the sides of the cardboard sheet were originally 6 in before we removed a square from each of the two corners joined by a side. The depth of the box equals the depth of the flap, which is s.



Now we can write down an equation for the volume of the box in terms of the dimensions of the flap:

$$V = \text{length} \times \text{width} \times \text{depth} = (6-2s)(6-2s)s.$$

The problem stipulates that the volume V must equal 16, therefore, our task is to solve the equation

$$(6-2s)(6-2s)s = 16.$$

This equation is equivalent to the degree three equation

$$s^3 - 6s^2 + 9s - 4 = 0,$$

which has two distinct roots, $s=1$ and $s=4$. Notice however that a valid solution must satisfy the inequality $s<3$, otherwise the squares would cover the perimeter of the sheet and there would be no flaps to fold. Thus $s=1$ is the only valid solution to the problem.

Of course, if $x^3-6x^2+11x-6=2x-2$, then it follows from Euler that the real number in question must be 1 or 4.

94

Authority

Theorem. *Let x be a real number. If* $x^3 - 6x^2 + 11x - 6 = 2x - 2$, *then* $x = 1$ *or* $x = 4$.

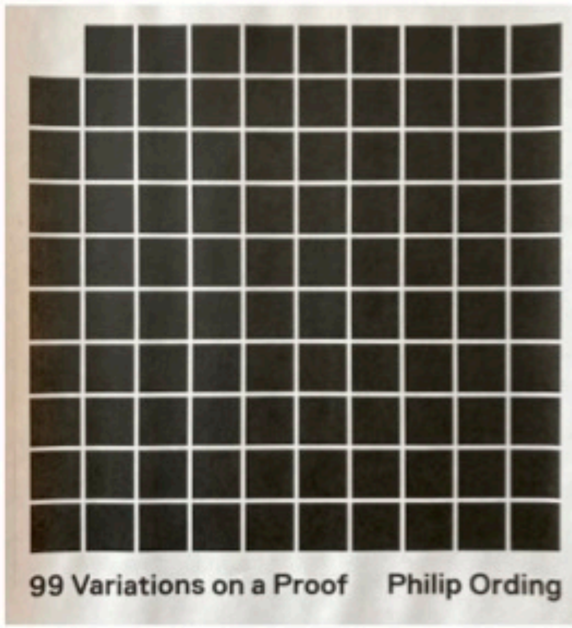
99

Prescribed

Proof. The proof is left to the reader. □

There are many ways to write a successful proof!

This activity is based on the book 99 *Variations on a Proof*, by Philip Ording, which is in turn based on Raymond Queneau's *Exercises in Style*.



Exercises in Style is a literary work from 1947 that presents in 99 different ways a story about a person who is first seen in a dispute on a bus, and then later discussing the position of a coat button with a friend. I hope this will help you view proof writing as creative writing and allow you to find and to use your own voice and style when writing your proofs.

(1) You've been given three proofs of the fact that $x^3 - 6x^2 + 11x - 6 = 2x - 2$ has exactly two solutions, $x = 1$ and $x = 4$.

(i) Which of these seem like successful proofs? In this context, "successful" can be understood to mean the proof provides a clear and convincing argument that the result is true.

(ii) What are the similarities and differences among the successful proofs?

(iii) What prevents the unsuccessful proofs from being successful? Can you modify them, keeping their original spirit and style, so as to make them successful?

(2) Pick your group's favorite of the three to share with the rest of the class.

Reading, Writing, and Evaluating Statistics and ML Papers

1 What makes for high-quality scientific and mathematical writing?

Writing high-quality scientific and mathematical content is a challenging, and requires an overlapping but not identical skillset to other forms of writing such as the types of essays you write in a humanities course. Some distinctive features of scientific and mathematical writing include the following:

1. The technical and mathematical focus requires different approaches to structuring arguments and introducing ideas and terminology.

2. Documents are typically structured into sections such as an introduction, background, methods, theory, empirical results, and related work.

3. The major types of evidence used are empirical experiments (usually numerical in statistics and ML) and mathematical theory.

Before you start writing, you should always ask yourself: Who is my audience? Who are audience can affect all aspects of what you write: what needs to be said, what knowledge can be assumed, what kinds of claims need justification.

Elements of high-quality writing: The big picture

1. Clear logical flow

    - Central claims and contributions of paper clearly stated (e.g., in abstract, introduction, conclusion)
    - Problem-solution framework
        - There should be a clear reason for each idea/result/method/etc. that is introduced
        - This framework operates at the paragraph, section and document level
    - Where to start? Depends on the context!

2. Claims are supported by appropriate evidence such as...
   - results from previous work
   - theoretical results
   - numerical results
3. The descriptions of the experiments should be detailed enough that the numerical results can be recreated based on paper alone. Often this will require looking at appendices and/or supplementary materials.[1]
4. Does not include extraneous or repetitive results
5. Acknowledges important limitations and directions for improvement such as...
   - unrealistic assumptions required for theoretical results
   - important problems or types of problems to which the theoretical or empirical results do not apply
   - additional experiments that could provide valuable insights
   - aspects of the method that could be improved (e.g., because it is slow, unreliable, doesn't scale to larger problems, is difficult to implement, ...)

Elements of high-quality writing: the details

1. Define terms and notation before they are used
   - What needs to be defined is context-dependent
2. Avoid forward references that require the reader to find that part of the paper to understand the current text. For example:
   - "Using Eq. (7) in the next section, we can conclude that..." [Introduce eq. (7) earlier if you need it at this point!]
   - "Since this relationship in linear (see Figure 2), we will..." [Show the figure here!]

[1] Best practices now require releasing the code used to run experiments. Ideally the code can be straightforwardly run to regenerate all results. But beware of bit rot.

3. Forward references as guidance are fine, and in fact can be very helpful. For example:
   - "We will generalize our result to the multidimensional case in Section 4.3."
   - "We will provide further empirical support for this claim in Section 5."
   - "We suggest some approaches to avoiding this numerical instability in the next section."
4. Figures should...
   - be high-resolution
   - have clearly labeled axes and legend (when appropriate)
   - use a readable font size
   - be colorblind friendly
   - use axis scales to improve clarity and readability
   - be information dense but not overwhelming
   - use consistent coloring and labeling schemes
5. Mathematical expressions should be integrated into the sentence, not stand by themself.
   - INCORRECT: "In fact, a is always positive, as shown by the following derivation.

     $$\begin{aligned} b - a &< c \\ a &> b - c \geq 0. \end{aligned}$$

     Where the final inequality follows from Eq. (2)."
   - CORRECT: "In fact, a is always positive since $b - a < c$ implies that $a > b - c \geq 0$, where the final inequality follows from Eq. (2)."

   Note that the incorrect example also ends with a sentence fragment, whereas complete sentences should always be used.

6. Tables should be professionally formatted using horizontal lines only where necessary and no vertical lines.[2]

7. Only include significant figures when reporting numerical results. For example:

   - You compute the p-values of 0.0196572, 0.0000312619002, and 0.007102031. Usually p-values less than 0.001 are considered very small, so it would suffice to report these p-values as 0.0197, 3×10^{-5}, and 7×10^{-3}.

   - You run a Monte Carlo simulation taking 100 i.i.d. samples to estimate an expectation $\mu = \mathbb{E}(X)$. Since $\mathrm{Var}(X) = 4$, the standard of the estimate is $\sqrt{4/100} = 0.2$. So, if you get an estimate of 3.1038023 for μ, it would suffice to it as 3.1 since the estimate could easily be off by ± 0.4 or more.

What to avoid

- Don't structure your writing to follow the logical flow of how you figured out the results or the order in which work was published

- Don't assume your audience knows everything that you do

- Don't include every conceivable numerical experiment and plot you can think of

2 Reading

There is no "right way" to read/understand a paper. It depends on personal preference, familiarity with the subject matter, the type of paper, and one's goals for reading the paper. But generally reading a paper linearly is not the best approach.

[2] In LaTeX professional tables can be created used the `booktabs` package. A good starting place to learn more is `https://en.wikibooks.org/wiki/LaTeX/Tables#Professional_tables`. The website `https://www.tablesgenerator.com` makes creating (professional) LaTeX tables much easier.

2.1 Before Reading

Find supplementary sources. books, book chapters, surveys; lecture/scribe notes from university courses; slides and videos of oral presentations; more recent papers that discuss the paper's contributions/limitations/etc. (can use Google Scholar to perform a reverse look-up)

Print the paper. the paper out or open it on a tablet device that enables annotating the text (these days I used my iPad). Having a "hard copy" can make a surprisingly big difference in your reading comprehensive.

2.2 One Approach

I'll outline my general approach to reading a methodologically focused paper. Methodology papers usually develop implementable algorithms for solving problems in statistics and machine learning. Oftentimes rigorous justification for the method is provided and various mathematical properties are characterized. For example, an optimization paper might contain theoretical results that give asymptotic conditions for convergence and quantitative bounds on solution accuracy.

(i) Read the paper through relatively quickly, focusing on the abstract, introduction, theorem statements (if central and/or decipherable), experiments, conclusion. It's fine to annotate as you go, but focus on getting the big picture and as much as possible avoid getting bogged down in confusing details.

(ii) Reread the paper more carefully but still skipping details that aren't core to understanding the major claims (e.g., practical implementation details, proofs, variants of the main method, supplementary theory). Make notes on questions for understanding and evaluating listed below

(iii) The next day or later, reread the paper again and make additional notes; read more of the details if they turn out to be important understanding and evaluating the paper

(iv) Use your notes to organize your evaluation in a way that will make it easy for someone who has not read the paper to follow.

(v) Write your evaluation of the paper. I often find I have to go back and reread parts of the paper as I write.

3 Understanding

When reading (or writing) a methodologically focused paper in statistics/ML/DS, there are a few major questions to keep in mind when trying to understand the paper (cf. Tim Roughgarden)

1. What problem is the paper trying to solve?
2. Why is the problem important?
3. What are the primary contributions?
4. What approach or approaches are used?
5. What are the key takeaways?

4 Evaluation

1. Do the authors provide sufficient evidence of their claims? *Claims are not just about the paper's contributions. They also concern motivation such as empirical claims about the performance of other methods.*
2. Do the authors compare their results (both theoretical and empirical) to appropriate baselines and alternatives?
3. Do the authors appropriately document the limitations of their approach? *Limitations can relate to methods, theory, and/or experiments.*
4. What are the biggest strengths of the paper? What are the greatest weaknesses of the paper? *Strengths and weaknesses could relate to any aspect of the paper: motivation, organization, utility of the results, correctness, strength of the evidence used to support the claim, ...*
5. What are the most important changes that would improve the paper? *Changes could be about any aspect of the paper: the writing, presentation of results (tables, figures), unsupported claims, problems with the method, ... However, they should be specific and not stray too far from the paper's current focus. So, a change that would not be appropriate to suggest is "develop a method that works better than this one." Change*

that would be appropriate to suggest include "run a specific new experiment" (e.g., check the accuracy for this dataset and model when varying the step step), "improve the formatting of this figure" (e.g., increase font size and use a log-linear scale), and "provide better motivation and more details of previous algorithms."

5 Looking Forward

1. What are the most important changes that would improve the paper?
2. Do the authors mention specific directions for future work or improvement?
3. Which directions for future work do you think are the most important? The most promising?

Working with Quantifiers

Quantifiers include phrases such as "for all," "for some," and "for one," as well as objects that provide bounds in proofs, such as those typically denoted by ϵ, δ, and N. It is very important to use the appropriate quantifiers for the situation in the appropriate order; accidentally swapping quantifiers can drastically change the meaning of a statement, and potentially destroy a proof.

As an example, consider how quantifiers appear in the definition of sequential convergence:

- **Correct:** If (x_n) converges to x in the metric space (X, d), given any $\epsilon > 0$ there exists an N such that $d(x_n, x) < \epsilon$ whenever $n \geq N$.
- **Incorrect:** If (x_n) converges to x in the metric space (X, d), there exists an N such that $d(x_n, x) < \epsilon$ for all $\epsilon > 0$ whenever $n \geq N$.
- **Incorrect:** If (x_n) converges to x in the metric space (X, d), there exists an N and $\epsilon > 0$ such that $d(x_n, x) < \epsilon$ whenever $n \geq N$.

Also, when multiple quantifiers are being used at different steps in a proof, it is important to choose them in the appropriate order, so that you don't end up with any unwanted dependence of your quantifiers on each other.

Suppose (X, d_X) and (Y, d_Y) are metric spaces, $f_n : X \to Y$ for all $n \in \mathbb{N}$, $f : X \to Y$, and $f_n \to f$ uniformly. Suppose f_n is continuous at $x_0 \in X$ for each n. Prove that f is continuous at x_0. Hint: utilize the fact that

$$d_Y(f(x), f(x_0)) \leq d_Y(f(x), f_n(x)) + d_Y(f_n(x), f_n(x_0)) + d_Y(f_n(x_0), f(x_0)).$$

Proof: Given $\epsilon > 0$, we must find a $\delta > 0$ such that $d_X(x, x_0) < \delta$ implies $d_Y(f(x), f(x_0)) < \epsilon$. Since each f_n is continuous at x_0, given $\epsilon > 0$ we can find a $\delta > 0$ such that $d_X(x, x_0) < \delta$ implies $d_Y(f_n(x), f_n(x_0)) < \epsilon/3$. Since $f^n \to f$ uniformly, there is an N such that $n \geq N$ implies $d_Y(f_n(x), f(x)) < \epsilon/3$ for all $x \in X$. As a result, if $d_X(x, x_0) < \delta$, whenever $n \geq N$ we have

$$\begin{aligned} d_Y(f(x), f(x_0)) &\leq d_Y(f(x), f_n(x)) + d_Y(f_n(x), f_n(x_0)) + d_Y(f_n(x_0), f(x_0)) \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence, f is continuous at x_0. $\square$

The issue with this incorrect proof is that the initial choice of δ depends on n: $\delta = \delta(n)$. Therefore, it is incorrect to say that the bound holds for all $n \geq N$. Given $\epsilon > 0$, one needs to find a single choice of δ that produces the required ϵ-bound. By choosing δ before fixing the n used in the hint, it is unclear that δ can be chosen to only depend on ϵ, and not also on n.

A Correct Proof: Given any $\epsilon > 0$, we must find a $\delta > 0$ such that $d_X(x, x_0) < \delta$ implies $d_Y(f(x), f(x_0)) < \epsilon$. Let an $\epsilon > 0$ be given. By the definition of uniform convergence, there exists an N such that $n \geq N$ implies $d_Y(f_n(x), f(x)) < \epsilon/3$ for all $x \in X$. This implies $d_Y(f_N(x), f(x)) < \epsilon/3$ for all $x \in X$ and, in particular, $d_Y(f_N(x_0), f(x_0)) < \epsilon/3$. Since f_N is continuous at x_0, there exists a $\delta > 0$ such that $d_X(x, x_0) < \delta$ implies $d_Y(f_N(x), f_N(x_0)) < \epsilon/3$. Therefore, whenever $d_X(x, x_0) < \delta$, we find

$$\begin{aligned} d_Y(f(x), f(x_0)) &\leq d_Y(f(x), f_N(x)) + d_Y(f_N(x), f_N(x_0)) + d_Y(f_N(x_0), f(x_0)) \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence, f is continuous at x_0. $\square$

Notice that, in this proof, $\delta = \delta(N)$, but $N = N(\epsilon)$ depends only on ϵ. Hence $\delta = \delta(N(\epsilon)) = \delta(\epsilon)$. It is clear that, once $\epsilon > 0$ is given, this fixes the choice of N, which in turn fixes the choice of δ.

Nested Quantifiers

We've now seen a variety of examples of proofs and proof techniques. Going forward, any given proof might use more than one method, or a combination of different strategies.

For example, we've seen two types of proofs involving quantifiers: the ϵ-N proofs used in sequential convergence, and the ϵ-δ proofs used for functional limits and continuity. Sometimes, a proof will require nested quantifiers. As an example, consider this HW problem.

Suppose that $h : (0,1) \to \mathbb{R}$ is uniformly continuous and let (x_n) be a Cauchy sequence in $(0,1)$. Prove that $(h(x_n))$ is also a Cauchy sequence.

Proof: Let $\epsilon > 0$ be given. Since h is uniformly continuous, there exists a $\delta > 0$ such that $|x-y| < \delta$ and $x, y \in (0,1)$ implies $|h(x) - h(y)| < \epsilon$. Since (x_n) is Cauchy, there is an N such that $n, m \geq N$ implies $|x_n - x_m| < \delta$. Then whenever $n, m \geq N$, we find $|h(x_n) - h(x_m)| < \epsilon$. $\square$

In this proof, it was important that we first identified the $\delta = \delta(\epsilon)$ from the definition of uniform continuity, and only after that did we identify the $N = N(\delta)$ from the definition of Cauchy sequence. As a result, $N = N(\delta) = N(\delta(\epsilon))$, and so our ultimate choice of N really only depends on ϵ.

Note that this result really does rely on uniform continuity. To see this, note that $f(x) = 1/x$ is a function that is continuous on $(0,1)$, but not uniformly continuous there. And $(x_n) = (1/n)$ is a Cauchy sequence for which $(f(x_n)) = (n)$ is not Cauchy.

Here's another proof involving nested quantifiers that is not quite correct. Can you figure out what's wrong with the proof?

Suppose that $g : B \to \mathbb{R}$ is uniformly continuous on a bounded set B and let (z_n) be a sequence in $g(B)$. Prove that (z_n) must be bounded.

Proof: The fact that $z_n \in g(B)$ means there is an $x_n \in B$ such that $g(x_n) = z_n$. Since g is uniformly continuous, for each $\epsilon > 0$ there is a $\delta > 0$ such that $|x - y| < \delta$ and $x, y \in B$ implies that $|g(x) - g(y)| < \epsilon$. Since $x_n \in B$ for all n, this means $|g(x_n) - g(x_m)| < \epsilon$ if $|x_n - x_m| < \delta$. Since $y_n = g(x_n)$ and $y_m = g(x_m)$ this means $|y_n - y_m| < \epsilon$. As a result (y_n) is a Cauchy sequence. Every Cauchy sequence is convergent, and every convergent sequence is bounded. Hence (y_n) is bounded. $\square$

The issues with the proof are (i) the assumption that B is bounded is not used anywhere, which is suspicious; and (ii) the claim is made that (y_n) is Cauchy, but there is no requirement that $n, m \geq N$ in order to conclude $|y_n - y_m| < \epsilon$. So the claim that (y_n) is Cauchy is not really justified.

Here's a correct version of the proof.

Suppose that $g : B \to \mathbb{R}$ is uniformly continuous on a bounded set B and let (z_n) be a sequence in $g(B)$. Prove that (z_n) must be bounded.

Proof: Suppose by way of contradiction that (z_n) is not bounded. This means there is some subsequence (z_{n_k}) such that $|z_{n_k}| \geq k$ for all $k \in \mathbb{N}$. The fact that $z_{n_k} \in g(B)$ means there is an $x_{n_k} \in B$ such that $g(x_{n_k}) = z_{n_k}$. Since (x_{n_k}) is in B and B is bounded, by Bolzano-Weierstrass it must have a convergent subsequence $(x_{n_{k_j}})$. Note this subsequence is also a Cauchy sequence because is it convergent. Pick $\epsilon = 1$ and let $\delta > 0$ be such that, by uniform continuity, $|g(x) - g(y)| < 1$ whenever $|x - y| < \delta$ and $x, y \in B$. Since the subsequence is Cauchy, there is an M such that whenever $j, l \geq M$ we have $|x_{n_{k_j}} - x_{n_{k_l}}| < \delta$. But then whenever $j \geq M$,

$$|z_{n_{k_j}}| = |g(x_{n_{k_j}})| = |g(x_{n_{k_j}}) - g(x_{n_{k_M}}) + g(x_{n_{k_M}})| \leq |g(x_{n_{k_j}}) - g(x_{n_{k_M}})| + |g(x_{n_{k_M}})| < 1 + |g(x_{n_{k_M}})|,$$

which is a contradiction, because $|(z_{n_{k_j}})| \geq n_{k_j} \to \infty$ as $j \to \infty$. $\square$

Note that it was important to first identify the δ from uniform continuity, and to then afterwards identify the M from the definition of Cauchy sequence.

Research Article Evaluation

Purpose

The goal of a Research Article Evaluation is to practice your critical reading skills and scientific writing skills. You will have the opportunity to:

- *Understand the goals, results, and limitations of methodological papers in statistics/machine learning/data science*
- *Structure your writing into a single cohesive and logical argument*
- *Think creatively about how to improve and/or extend a theoretical or methodological research contribution*

Task

For your assigned or selected research paper, write a single, cohesive evaluation that addresses the following questions:

1. **Understanding:**
   (a) What problem is the paper trying to solve?
   (b) Why is the problem important?
   (c) What are the primary contributions?
   (d) What approach or approaches are used?
   (e) What are the key takeaways?
2. **Evaluation:**
   (a) Do the authors provide sufficient evidence of their claims? *Claims are not just about the paper's contributions. They also concern motivation such as empirical claims about the performance of other methods.*
   (b) Do the authors compare their results (both theoretical and empirical) to appropriate baselines and alternatives?
   (c) Do the authors appropriately document the limitations of their approach? *Limitations can relate to methods, theory, and/or experiments.*

(d) What are the biggest strengths of the paper? What are the greatest weaknesses of the paper? *Strengths and weaknesses could relate to any aspect of the paper: motivation, organization, utility of the results, correctness, strength of the evidence used to support the claim, ...*

(e) What are the most important changes that would improve the paper? *Changes could be about any aspect of the paper: the writing, presentation of results (tables, figures), unsupported claims, problems with the method, ... However, they should be specific and not stray too far from the paper's current focus. So, a change that would not be appropriate to suggest is "develop a method that works better than this one." Changes that would be appropriate to suggest include "run a specific new experiment" (e.g., check the accuracy for this dataset and model when varying the step size), "improve the formatting of this figure" (e.g., increase font size and use a log-linear scale), and "provide better motivation and more details of previous algorithms."*

Deliverable

Initial submission. Submit your evaluation as a pdf via Blackboard. You must use the course LaTeX template. The maximum length is 5 pages. There is no strict minimum but it would be difficult to address all questions in less than a 2 or 3 pages.

Resubmission. You will have the chance to revise and resubmit if your initial submission does not receive a Pass. The requirements for the resubmission are the same as for the initial submission except that, in addition, you must include an appendix reflecting on how you incorporated the feedback your received for your initial submission.

Name:__________________________

	Pass	Almost Pass	Partial Pass	Not Yet

✓ = *always/completely* ✓− = *almost always/completely* ✗ = *not yet*

Formatting/Presentation

✓		✗	Write-up is a pdf that uses the course LaTeX template
✓	✓−	✗	Follows high-quality writing recommendations
✓	✓−	✗	Notation is consistent with the paper.

Content

✓	✓−	✗	Writing is well-organized with clear, easy-to-follow logical flow
✓	✓−	✗	Writing is professional (rather than informal) and constructive (rather than dismissive or unkind)
✓	✓−	✗	Provides reasonable answers to "Understanding" questions ☐ What problem is the paper trying to solve? ☐ Why is the problem important? ☐ What are the primary contributions? ☐ What approach or approaches are used? ☐ What are the key takeaways?
✓	✓−	✗	Provides reasonable answers to at least 4 of 5 "Evaluation" questions and supports them with concrete, specific evidence ☐ Sufficient evidence of their claims? ☐ Compare to appropriate baselines and alternatives? ☐ Appropriately document the limitations? ☐ Biggest strengths and weaknesses of the paper? ☐ Changes that would improve the paper?

Grading

To receive a **Pass:**

- A ✓ on all Formatting/Presentations rubric items
- A ✓ on all Content rubric items

To receive an **Almost Pass:**

- A ✓ on 2 out of 3 Formatting/Presentation rubric items and a ✓– or ✓ on the other one
- A ✓ on 3 out of 4 Content rubric items and a ✓– or ✓ on the other one

To receive a **Partial Pass:**

- A ✓ on 1 out of 3 Formatting/Presentation rubric items and a ✓– or ✓ on the other two
- A ✓ on 2 out of 4 Content rubric items and a ✓– or ✓ on the other two

Introduction to Evaluating Expository Mathematical Writing

Purpose

Become familiar with some of the writing standards we will be using throughout the semester and get a feel for the types of work we will be doing during class time.

Tasks

Complete the following tasks in order. I've included estimates of how much time each task should take.

1. (5 minutes) Read *Some Guidelines for Good Mathematical Writing* until the heading "Toward Elegance."
2. (5 minutes) Read the writing sample on the next page, which is meant to be a conceptual introduction to the central limit theorem for high school students who are familiar with basic probability theory (e.g., some common distributions) but may not have taken calculus or linear algebra.
3. (15 minutes) Now, reread paragraph a few more times and check whether it follows these guidelines:
   (a) writing is appropriate for audience
   (b) take invitational tone
   (c) uses complete sentences
   (d) uses words to give context to equations
   (e) avoids shorthands
   (f) avoids beginning sentences or phrases with a number or symbol
   (g) emphasizes unfamiliar words that we are about to define

   Whenever you see an error, circle it and label it with the letter of the guideline it doesn't follow. Are there other issues that make the text hard to read? If so, make a note of these.

4. (10 minutes) Compare notes with your partner. Did they notice something you didn't or vice versa? Do the two of you disagree about whether any of the guidelines was followed or not?

Writing Sample

There are infinitely many probability distributions. For example, binomial, Poisson, normal, and exponential distributions. All of these and more occur quite commonly in real life. But the normal distribution is special because of one of the most important results in probability theory: the central limit theorem. The central limit theorem tells us that if we have a sequence of mean-μ i.i.d. random variables $X_1, X_2, \ldots$ with finite second moment, then as $n \to \infty$, $\sqrt{n}(\bar{X}_n - \mu)$ convergences to a $\mathcal{N}(0, \sigma^2)$ random variable, where $\bar{X}_n := n^{-1}\sum_{i=1}^{n} X_i$ and $\sigma^2 := \mathrm{Var}(X_1)$. In other words, when I have n independent samples from the same distribution which has mean μ and variance σ^2, if n is sufficiently large the average of those samples will be approximately normal with mean μ and variance σ^2/n. This is an important and remarkable results! Why is that? Because it holds no matter how complicated or weird the distribution is, and we don't even need to know it. Thus, the central limit theorem is a powerful result with many statistical applications. For example, given some samples from a population (e.g., the heights of high school seniors), we can construct a *confidence interval* (a, b) that contains μ (e.g., the average height of high school seniors) with high probability. n does not even need to be very big since the width of the interval will be $c\sigma/\sqrt{n}$ for a small constant c (usually less than 6).

Quantifiers in the context of sequential convergence.

Consider the following (in)correct definitions of sequential convergence.

(a) **Correct:** The sequence (x_n) converges to x if given any $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $|x_n - x| < \epsilon$ whenever $n \geq N$.

(b) **Incorrect:** The sequence (x_n) converges to x if there exists an $N \in \mathbb{N}$ such that $|x_n - x| < \epsilon$ for all $\epsilon > 0$ whenever $n \geq N$.

(c) **Incorrect:** The sequence (x_n) converges to x if given any $\epsilon > 0$ and any $N \in \mathbb{N}$, there exists an $n \geq N$ such that $|x_n - x| < \epsilon$.

As a group, answer the following questions.

(1) What is the relationship between (a) and (b)? Can you find an explicit example sequence that converges in the sense of (a) but not (b)? What about one that converges in the sense of (b) but not (a)?

(2) What is the relationship between (a) and (c)? Can you find an explicit example sequence that converges in the sense of (a) but not (c)? What about one that converges in the sense of (c) but not (a)?

(3) Can you order the conditions (a), (b), and (c) from most to least restrictive?

(4) Since (b) and (c) are incorrect, in that they do not imply the sequence converges in the usual sense, can you explain, in words, what they do imply about the sequence?

Learning Overleaf and LaTeX for MA 512

Margaret Beck

December 18, 2023

1 Instructions

The goal of this exercise is for you to complete the assignment in Section 4 using LaTeX. In your document please also create a title with your name and the date.

2 Basic Set-up

Here are instructions for setting up an account in Overleaf and creating your first .tex file.

1. Create a Free Account
   - Plans & Pricing.
   - Create a free account.
   - Login to your account
2. Create your first "project":
   - Click on New Project → Blank Project. (You can also play around with different templates and customize various aspects of the style, but you need not.)
   - Name the project.
   - Starting adding content.
   - In order to see your changes in the output file, you must "Recompile" the document by clicking the associated green button at the top of the right half of the screen. There are also keyboard shortcuts for this, such as Command-S on a mac.
3. Learning more:
   - Go to Home (the house icon) → Help → Documentation.
   - Check out "Learn Latex in 30 minutes"
   - You can also run LaTeX locally on your computer. More info is here. I use MacTex, for example.
4. To download a pdf file of your project, next to the green "recompile button" you will see an icon that will allow you to "Download PDF" (hover over it to see these words).

3 Basic information to get you going

To write standard characters, you just type like you usually would. If you want to use special mathematical symbols, you need to know the LaTeX command for that symbol, and then put the command inside dollar signs, so LaTeX knows you are writing in "math mode". For example, to produce $f(x) = \nabla F(x)$, you need to type `$f(x) = \nabla F(x)$`; the command for the gradient symbol is "\nabla".

If you want to produce output that is not inline, meaning not within the standard paragraph format, you need to work in different environments. For example, to produce

$$f(x) = \sin(x)$$

you need to type

```
\begin{equation*}
f(x) = \sin(x)
\end{equation*}
```

This is an example of working in the equation environment. The asterisk after the word equation is telling LaTeX not to give that equation a number. You could equivalently write

```
\[
f(x) = \sin(x)
\]
```

Without the asterisk it produces

$$f(x) = \sin(x). \tag{1}$$

These equation numbers can be useful if you want to refer to things later. For example, I could write

```
\begin{equation}
f(x) = \sin(x) \label{def-f}
\end{equation}
```

to produce

$$f(x) = \sin(x) \tag{2}$$

and then use the \eqref{def-f} to refer to that equation, as in: We can see from equation (2) that f is defined to be the sine function. The red box around the equation number is a hyperlink. If you use this file to create a pdf file, then a person could click on the red box and be automatically taken to the page in the document where that equation appears. This is particularly useful in multi-page documents, where you might be referring to equations that appear on other pages.

When you first start working with LaTeX you'll probably need to do a lot of googling, or a lot of reading the Overleaf documentation, until you learn the various symbol commands and environments that LaTeX offers. (I still do this on a regular basis.) For the assignment below, you might want to google the align environment. It might feel slow and frustrating at first, but it will get easier and faster as you practice.

4 Assignment

Reproduce the following examples using your own file.

For this first example you will need to use bold text, the align environment, the ˆ symbol, and the commands \frac, \sin, and \cos.

Example: Here we compute the derivative of the function $f(x) = \sin(x^2)$ using the chain rule.

$$\begin{aligned} f'(x) &= \frac{d}{dx}[\sin(x^2)] \\ &= \cos(x^2)\frac{d}{dx}[x^2] \\ &= 2x\cos(x^2). \end{aligned}$$

For this second example you will additionally need the \left and \right commands, to make the outer parentheses larger than inner ones. Also note how this example is written in complete sentences, with punctuation at the end of each equation.

Example: The fundamental trig identity $\sin^2(x) + \cos^2(x) = 1$ can be used to derive other trig identities. For example, if we divide this equation by $\cos^2(x)$, we find

$$\frac{1}{\cos^2(x)}\left(\sin^2(x) + \cos^2(x)\right) = \frac{1}{\cos^2(x)},$$

or

$$\tan^2(x) + 1 = \sec^2(x).$$

MA 341 Number Theory
Portfolio Prompts

You may choose a prompt multiple times. Each portfolio entry should at least a few paragraphs, but no more than 2 pages long. Be sure to include enough context and examples to make the mathematical ideas clear.

- **Productive Struggle:** Choose a homework problem that you struggled with. Restate the problem, then describe the struggle and how you overcame it. How do you understand the problem now?
- **Mathematical Curiosity:** Give an example of a pattern you noticed or other mathematical topic from this class that you are curious about. Why is this topic interesting to you? How might you explore it further?
- **Making a Strategy:** Choose a problem from the class (either homework or classwork, or a related question that you came up with yourself) that you still aren't confident that you can solve. Discuss strategies for how you might tackle this problem, and try them out. Even if you do not feel that you have completed the problem, give any insights you have and identify the places where you feel your strategy doesn't work.
- **Proof Analysis:** Take a proof that you wrote earlier in the class, either one you're especially proud of or one you think could be improved. Critique the proof the way we critiqued the proof of division with remainder—explain what you did well and what choices you made to make the proof clear to the reader, and what might be less clear. If you think your proof can be significantly improved, write a new version incorporating your feedback.
- **Mathematical Ode:** Choose a topic or pattern you found that surprised or intrigued you, or that you find beautiful. What about this fact or topic is surprising/intriguing/beautiful? Briefly explain the topic in a way you would have understood at the beginning of the semester.
- **Going Beyond:** There are many topics that we didn't cover in this class or could have taken further, but which may have come up in various discussions. Some examples are: how to test whether a number is prime, mathematical induction vs. the well-ordering principle, which U_m have a generator (besides when m is prime). Explore one of these topics either by searching online, doing some of your own mathematical experimentation, or a combination. Report on the results of your explorations.
- **Miscellaneous:** Do you have your own idea of a 1-2 page thing you want to write about for the class? Propose an idea! Other topics can be done with approval.

Negations.

Discuss how to negate each of the following. To come up with a mathematical negation means to come up with a mathematically precise way of saying that the statement is not true.

Let's start with some non-mathematical statements.

(i) At every college in the United States, there is a student who is at least seven feet tall.

(ii) For all colleges in the United States, there exists a professor who gives every student a grade of either A or B.

(iii) There exists a college in the United States where every student is at least six feet tall.

(iv) All people who like the color purple like to eat pizza.

(v) Every person either likes the color blue or likes to eat tacos.

(vi) Everyone likes the color green and likes to eat carrots.

(vii) If all people like the color orange, then all people like to eat ice cream.

(viii) There exists a person who likes to eat sushi.

Now for some mathematical ones.

(a) **Sequential convergence.** Recall that $\lim x_n = x$ if, given any $\epsilon > 0$, there is an N such that $|x_n - x| < \epsilon$ whenever $n \geq N$. What does it mean to say that (x_n) does not converge to x?

(b) **Series convergence.** Recall that $\sum_{k=1}^{\infty} a_k$ converges to S if the sequence of partial sums (s_n), where $s_n = a_1 + \cdots + a_n$, converges to S: $\lim s_n = S$. What does it mean to say that $\sum_{k=1}^{\infty} a_k$ does not converge to S?

(c) **Open.** Recall that a set $O \subseteq \mathbb{R}$ is open if for every $x \in O$ there is an $\epsilon > 0$ such that $V_\epsilon(x) \subseteq O$. What does it mean to say that a set O is not open?

(d) **Closed.** Recall that a set $F \subseteq \mathbb{R}$ is closed if it contains its limit points. What does it mean to say that a set F is not closed?

(e) **Compact.** Recall that a set $K \subseteq \mathbb{R}$ is compact if every sequence (x_n) in K has a subsequence (x_{n_j}) that converges to a limit that is in K. What does it mean to say that a set K is not compact?

(f) **Connected** Recall that a set E is disconnected if it can be written as $E = A \cup B$ where A and B are nonempty sets such that $\bar{A} \cap B = A \cap \bar{B} = \emptyset$. We defined a set to be connected if it is not disconnected. What is a precise way of saying that a set is connected? (Do not just say it cannot be written as $E = A \cup B$ where A and B are nonempty sets such that $\bar{A} \cap B = A \cap \bar{B} = \emptyset$. Think about how to precisely say what it means for this to not be possible.)

Theorem. *(Cut Property) If A and B are nonempty, disjoint sets with $A \cup B = \mathbb{R}$ and $a < b$ for all $a \in A$ and all $b \in B$, then there exists a $c \in \mathbb{R}$ such that $a \leq c$ whenever $a \in A$ and $c \leq b$ whenever $b \in B$.*

Prove this theorem using the Axiom of Completeness, which states that any nonempty set of real numbers that is bounded from above has a least upper bound.

Proof. Assume I have A and B as in the statement of the Cut Property. Since A is nonempty, it has a least upper bound. Let $c \in \mathbb{R}$ be that least upper bound. Because c is an upper bound for A, $a \leq c$ for some $a \in A$. Since $a < b$ for all $a \in A$ and all $b \in B$, this means $c \leq b$ for all $b \in B$. Hence, c is the desired number that satisfies the cut property. $\square$

Theorem. *(Cut Property) If A and B are nonempty, disjoint sets with $A \cup B = \mathbb{R}$ and $a < b$ for all $a \in A$ and all $b \in B$, then there exists a $c \in \mathbb{R}$ such that $a \leq c$ whenever $a \in A$ and $c \leq b$ whenever $b \in B$.*

Prove this theorem using the Axiom of Completeness, which states that any nonempty set of real numbers that is bounded from above has a least upper bound.

Proof. Assume I have A and B as in the statement of the Cut Property. Since A is nonempty, it has a least upper bound. Let $c \in \mathbb{R}$ be that least upper bound. Because c is an upper bound for A, $a \leq c$ for some $a \in A$. Since $a < b$ for all $a \in A$ and all $b \in B$, we have $c \leq b$ for all $b \in B$. Hence, c is the desired number that satisfies the cut property. □

Overall thoughts/questions:

- Do you think the writer fully understands why the Theorem is true?
- Which key ideas of the proof does the writer mention accurately?
- Which key ideas of the proof does the writer leave out?

Comments by sentence:

- First sentence: mathematicians don't typically write in the first person. A more traditional wording would be "Let A and B be sets of real numbers that satisfy the assumptions of the theorem." Also, the sets A and B are not typeset. These are very minor things that don't affect the reader's ability to understand the proof.
- Second sentence: the existence of the least upper bound for A is not fully justified.
- Fourth sentence: this has a mathematically consequential error of word choice.
- Fifth sentence: the conclusion in this sentence is not fully justified.

Exercise 1.1.14 Prove Proposition 1.1.19, which states:

Proposition. *Let X be any set, let d_{disc} be the discrete metric on X, let $(x^{(n)})$ be a sequence in X, and let $x \in X$. Then $(x^{(n)})$ converges to x with respect to the discrete metric if and only if there exists an $N \in \mathbb{N}$ such that $x^{(n)} = x$ for all $n \geq N$.*

Proof. First, suppose that $(x^{(n)})$ converges to x with respect to the discrete metric. This means there is an $\epsilon > 0$ and $N \in \mathbb{N}$ such that $d_{disc}(x^{(n)}, x) < \epsilon$ for all $n \geq N$. This means we can choose $\epsilon = 1/2$ and find the corresponding N. By the definition of the discrete metric, $d_{disc}(x^{(n)}, x) < 1/2$ can only happen when $x^{(n)} = x$. Thus, $x^{(n)} = x$ when $n = N$.

Next, suppose that there exists an $N \in \mathbb{N}$ such that $x^{(n)} = x$ for all $n \geq N$. Then for $\epsilon = 1/2$ we find that $d_{disc}(x^{(n)}, x) = 0 < \epsilon$ for all $n \geq N$, and so $(x^{(n)})$ converges to x with respect to the discrete metric. □

Proposition. *Let X be any set, let d_{disc} be the discrete metric on X, let $(x^{(n)})$ be a sequence in X, and let $x \in X$. Then $(x^{(n)})$ converges to x with respect to the discrete metric if and only if there exists an $N \in \mathbb{N}$ such that $x^{(n)} = x$ for all $n \geq N$.*

Proof. First, suppose that $(x^{(n)})$ converges to x with respect to the discrete metric. This means there is an $\epsilon > 0$ and $N \in \mathbb{N}$ such that $d_{disc}(x^{(n)}, x) < \epsilon$ for all $n \geq N$. This means we can choose $\epsilon = 1/2$ and find the corresponding N. By the definition of the discrete metric, $d_{disc}(x^{(n)}, x) < 1/2$ can only happen when $x^{(n)} = x$. Thus, $x^{(n)} = x$ when $n = N$.

Next, suppose that there exists an $N \in \mathbb{N}$ such that $x^{(n)} = x$ for all $n \geq N$. Then for $\epsilon = 1/2$ we find that $d_{disc}(x^{(n)}, x) = 0 < \epsilon$ for all $n \geq N$, and so $(x^{(n)})$ converges to x with respect to the discrete metric. □

Overall thoughts/questions:

- The writer addressed both parts of the if and only if statement, and lots of the key ideas are there.
- But in multiple places the writer does not correctly use quantifiers.

More detailed comments:

- First paragraph: the only error is in the second sentence, where the quantifiers ϵ and N are not used correctly.
- Second paragraph: again there is an issue with quantifiers. The proof is only provided for one specific choice of ϵ. So it needs to be rewritten in order to handle an arbitrary $\epsilon > 0$.

A draft of a proof.

Suppose a classmate asks you the following questions regarding the below proof draft.

(1) Did I forget any cases?

(2) Did you notice any gaps in the logical argument?

(3) Is there anything I can do to improve the formatting in LaTeX?

With these questions in mind, carefully read their proof draft and write some feedback in response. You don't need to show anyone what you write. Once everyone has a chance to do this, we will discuss, as a class, what feedback might be useful for this person.

Lemma: For any two real numbers a, b $\in \mathbb{R}$, the triangle inequality holds, meaning that

$$|a+b| \leq |a| + |b|.$$

Proof: For this proof we will consider four cases: (i) when both a and b are both positive; (ii) when both a and b are both negative; (iii) when a is positive and b is negative; and (iv) when a is negative and b is positive.

For case (i), we assume a > 0 and b > 0. This implies $|a| = a$ and $|b| = b$. It also implies that a+b > 0, and so $|a+b| = a+b$. Therefore, $|a+b| = a+b = |a| + |b|$, and so $|a+b| \leq |a| + |b|$.

For case (ii), we assume a < 0 and b < 0. This implies $|a| = -a$ and $|b| = -b$. It also implies that a+b < 0, and so $|a+b| = -(a+b)$. Therefore, $|a+b| = -a-b = |a|+|b|$, and so $|a+b| \leq |a|+|b|$.

For case (iii), we assume a > 0 and b < 0. This implies $|a| = a$ and $|b| = -b$. It also implies that $|a+b| = |a| - |b|$. Since $b < 0$, $-|b| < -b$. Therefore, $|a+b| = |a| - |b| < a - b = |a| + |b|$.

For case (iv), we assume a < 0 and b > 0. This implies $|a| = -a$ and $|b| = b$. It also implies that $|a+b| = |b| - |a|$. Since $a < 0$, $-|a| < -a$. Therefore, $|a+b| = -|a| + |b| < -a + b = |a| + |b|$.

$\square$

Rubric for Portfolio Proofs

For the overall proof to be successful, each row of the table must be achieved.

	Description	Achieved	Not yet achieved
Mathematical correctness			
Mathematical reasoning	The logical steps provided clearly advance the argument and lead to a convincing proof of the stated assertion.		
Correct use of notation	Is all mathematical notation correct, clearly denoted, and appropriate for the situation?		
Correct use of mathematical terminology	Are all stated mathematical terms used correctly? Are any terms inaccurate, inappropriate, or misused?		
Clarity of writing			
Quality and quantity of details	Too few details make it difficult for the reader to follow the argument, while too many can distract the reader. The notions of "too few" and "too many" should be interpreted based on your intended audience.		
Coherence devices	The proof is written in complete sentences, with transition words, phrases, or sentences that help the reader follow your argument. Does the proof sound reasonable when read out loud?		
Care in proofreading	If the writing free from typos or grammatical errors that are significant enough to change the meaning of what's written or distract the reader?		

MA 512, Introduction to Analysis II
Syllabus, Spring 2024

Instructor: Margaret Beck (mabeck@bu.edu)	Meetings: TR 1230-145pm, PRB 146
Book: Analysis II (see also Analysis I)	Prerequisite: MA 511
Office Hours: M 2-330 (drop in) and R 330-5 (by appointment). I would love to see you there!	

Learning Goals and Objectives: In this course you will learn both content related to analysis and also skills connected with precise and logical thinking and communication. In particular:

- *Analysis Content:* You will extend your understanding of analysis from the context of real numbers to general metric spaces, as well as to topics beyond those discussed in MA 511, such as Fourier series, differential calculus in several variables, and Lebesgue integration.
- *Thinking and Communication:* You will expand your problem solving skills and your ability to communicate and to evaluate precise and logical mathematical arguments, both through writing and revising formal mathematical proofs and orally communicating mathematical ideas.

Assessment and Grading: Grades for this course will be determine as follows.

	Homework	Portfolio Proofs	Midterm and Final Oral Exams	Participation
A	All except 1	At least 6	At least 4 Points	Positive Contribution
A-	All except 1	At least 5	At least 3 Points	Positive Contribution
B+	All except 2	At least 4	At least 2 Points	Positive Contribution
B	All except 2	At least 3	At least 1 Point	Positive Contribution
B-	All except 2	At least 2	None needed	Positive Contribution
C	All except 3	At least 1	None needed	Positive Contribution

Other plus/minus grades (eg C+, C-) will be assigned to students who fall slightly short of the above grading standards. A grade of F may be assigned to any student who does not meet the requirements for a grade of C.

- **Participation**: Since we will regularly develop ideas together as a class, regular participation (which necessitates regular attendance) is essential. *If you need to miss a class, please email me to acknowledge your absence and to provide a brief explanation why.*

- **Homework:** Homework will be due on Monday of most weeks, via Blackboard. Your homework may be handwritten. (But if you want to type it in LaTeX, that's great!) Homework will be graded mostly for completion, and as long as you make a good-faith effort to solve all the problems in a given assignment, that assignment will be deemed complete. If there are N total assignments, then you need N-1 complete for a grade of A or A-, N-2 complete for a grade of B+, B, or B-, and N-3 complete for a grade of C. *You may submit homework late if needed, but each two late assignments will increase the number of required homeworks for a given grade by 1.* For example, if you submit one assignment late it will not impact the way your final grade is determined. But if you submit two assignments late, you will need N to be complete for a grade of A or A-, N-1 complete for a grade of B+, B, or B-, etc. *If you need to submit an assignment late, please email me to let me know and to provide a brief explanation why.*
- **Portfolio Proofs:** These will be opportunities for you to select an eligible problem from the homework assignments, revise it, and turn it into a final draft that you feel represents your best work. Portfolio proofs will be determined to be achieved or not yet achieved according to this rubric. Roughly speaking, a successful proof is one that provides a clearly written and logically convincing argument that the stated result is true. You may achieve at most one portfolio proof from any given homework assignment. You will have the opportunity to turn in up to two portfolio proofs on most Thursdays. You may revise and resubmit a given portfolio proof as many times as you want, subject to the constraint that at most two portfolio proofs may be submitted on each relevant Thursday. Portfolio proofs must be typeset in LaTeX, include one copy of the rubric for each problem submitted, and be submitted via Blackboard as a pdf file. *The deadlines for portfolio proof submission are firm. If you miss the deadline in a given week, you must wait until to the following week to submit your proofs, and thus lose two opportunities to submit proofs.*
- **Midterm and Final Oral Exams:** The midterm oral exams will take place during the weeks of March 18 and March 25, and the final oral exams will take place during May 1-8. You can earn 0, 1, or 2 points on each exam. These exams are optional, but they are required for anyone wishing to earn a grade of B or higher. Each exam will be 30 minutes long and audio recorded, and they will be scheduled at a time that is mutually convenient for me and the student. More details about the format of the exams will be given closer to the exam dates.

How to succeed in this course: Each of you is capable of succeeding in this course. Here are some suggestions for how to do that.

- **Embrace productive struggle.** The easy thing to do is to give up, but we rarely learn from things that are easy. Find ways to make the struggle productive. Do not immediately resort to asking someone, or the internet, for help. Your experience in the course will be much more meaningful if you figure things out on your own, and you will internatilize the knowledge much more deeply. It does not matter how quickly you learn something, but whether you ultimately get there in the end.

- **Collaborate.** Talk with your friends and classmates. Work together. Collaborating with someone who is approaching the problem in a different way can deepen your own understanding. If you are stuck, hearing someone else's perspective can often unstick you, even if they don't directly tell you the answer, and even if they are stuck too. And if you feel you understand something, it is not a waste of time to help someone else who is stuck. Teaching others is a way to check whether or not you truely understand something.
- **Ask for help when you need it.** It is ok to need help, especially after you have tried to embrace the productive struggle. We all need help sometimes. Ask your friends and classmates for help. Be kind to yourself and each other. Come to office hours; I want you to have a good experience in this course and I would be happy to help.
- **Don't wait until the last minute.** The material we discuss will build progressively throughout the semester. If you are confused about a topic we discuss, sort that confusion out soon, ideally before we move too far forward on to additional topics. Also, don't wait until the end of the semester to start submitting portfolio proofs, because then you lose valuable opportunities for feedback and revision.

What to do when you don't know what to do: If you are feeling stuck, here are some suggestions for working productively through your struggle.

- **Work out some examples**. If the problem is asking you to prove an abstract result, make it more concrete. What does the result look like in R? What about R^n? If the result is about a sequence of functions that satisfy some general assumptions, write down a specific example sequence that satisfies the assumptions. Why is the result true for that example sequence? Once you understand some examples, you can return to thinking about the general case or more abstract setting.
- **Have you see a similar problem?** If the result is asking you to prove something in R^n or in a general metric space, have you already proven a similar result for R? Maybe you did so in MA 511. Or maybe you have essentially thought about the problem in multivariable calculus, and now what you need to do is to make that calculus calculation rigorous.
- **Do some calculations; focus on your intuition.** Temporarily forget about writing a proof. Just think about why the result should (or shouldn't) be true. What steps have lead you to that conclusion? Once you have an outline of the key steps, you can go back and work on making them rigorous, and turn your outline into a formal proof (or into a counterexample).
- **Talk with someone.** Collaborate, and ask for help when you need it.

Classroom environment: I hope that all students in this course feel it is in an environment in which they can productively learn. Diversity of background (including, but not limited to: race, gender, ethnicity, sexual orientation, age, socioeconomic status, religion, ability) is an asset. Diversity of ideas makes our ability to do mathematics stronger. See this article, which shows that "Being around people who are different from us makes us more creative, more diligent and

harder-working." It is extremely important that all members of our classroom community feel welcomed and respected. If there are any ways I can help facilitate this, I welcome that feedback. **I hope that each student feels comfortable letting me know if they feel that their learning is being adversely affected by any experiences, inside or outside of class.**

With this in mind, I would like to acknowledge and emphasize:

- As Francis Su eloquently put it in The Lesson of Grace in Teaching, each of you is a valuable human, regardless of what your accomplishments may or may not be.
- Frederico Ardila's Axioms: Axiom 1) Mathematical potential is distributed equally among different groups, irrespective of geographic, demographic, and economic boundaries. Axiom 2) Everyone can have joyful, meaningful, and empowering mathematical experiences. Axiom 3) Mathematics is a powerful, malleable tool that can be shaped and used differently by various communities to serve their needs. Axiom 4) Every student deserves to be treated with dignity and respect.
- Labels like "good at math" are problematic. Each of us is capable of learning analysis in a deep and meaningful way. As explained in The Secret to Raising Smart Kids by Carol S. Dweck, "a focus on 'process' - not on intelligence or ability - is key to success."

You are always welcome to tell me your preferred name and/or pronouns at any time.

Accessibility Resources and Other Types of Support: A variety of support resources exist on BU's campus.

- **Accessibility:** BU's Disability and Access Services can provide services and support to ensure that students are able to access and participate in the opportunities available at Boston University. Please reach out to them if you need any additional support or accommodations. Please also feel free to reach out directly to me with requests for support and accommodation, regardless of whether or not you are in touch with the Disability and Access Services office.
- **Office hours**: Please come to office hours! I welcome the opportunity to get to know students outside of class. I want you to have a great experience in this course and I am happy to help.
- **Tutoring Room in the Department of Mathematics and Statistics**: The tutoring room is staffed by graduate students and provides drop-in help for students enrolled in any of our courses. Different tutors have different backgrounds and areas of expertise, so I recommend that you consult not only the schedule but also the list of tutoring expertise to find a time when a tutor with an appropriate background in analysis will be available.
- **Peer tutoring at the Educational Resource Center (ERC)**: The ERC Peer Tutoring program provides BU students an opportunity to meet with a fellow student and ask questions related to their course material. These services are free to all BU students.
- **Mental Health and Wellness:** Additional types of support that you may find helpful can be found on the Academic Help and Wellness page of the Department of Mathematics and Statistics.

Academic Integrity and Tech Policies: I trust that you are all aware of BU's academic conduct code. I hope and expect that you will all uphold it. I would like to highlight two key expectations I have for each student.

- Be honest with yourself and others about when you do not understand something. There is no shame in needing help with any aspect of this course.
- Do not present the work of anyone else (human or otherwise) as your own. It is natural in mathematics to collaborate with others, but you must acknowledge your collaborators, only write up ideas that you genuinely understand yourself, and those ideas should be written in your own words.

Regarding the use of AI, like ChatGPT, and the use of other online resources like Chegg, MathOverflow, etc: Using these services to directly complete your work for this course will significantly compromise your learning. However, there are productive ways to utilize online resources. Here are some examples:

- Looking up definitions, theorems, etc, if for some reason you are not able to access our book in that moment, or if you would like a complementary persepctive on the topic.
- Looking up LaTeX commands and debugging LaTeX code. You can even ask things like ChatGPT to generate LaTeX code for you. Just be aware that the code it generates might not be very good. But, nevertheless, it could be a good starting point for you.
- Asking things like ChatGPT to prove results that we have already proven in class, or results for which a proof exists in the book, and then trying to determine yourself if the proof is correct. This is a great way to practice your ability to detect errors or logical inconsistencies in proofs. Some proofs I've tried in this way come out essentially perfect, but sometimes there are unexpected, and often interesting, errors.

Regarding the use of cell phones, tablets, and laptops in class: Although there are many positive ways to use these devices (like note taking, typing up work in LaTeX, etc), there are also studies that show that using such devices for non-academic purposes leads to reduced long-term retention of in-class material, such as lower performance on exams. So I strongly encouage you to limit your device use to only academic purposes that are necessary for our class.

CAS MA 586 / CDS DS 522

Stochastic Methods for Algorithms

Table of Contents

Overview

Instructor	Dr. Jonathan Huggins (huggins@bu.edu)
Office	CDS 427
Drop-in Hours	Wednesday 12:30–2:00pm and Thursday 10:30–11:30am (CDS 427)
TF	Andrew Roberts (arober@bu.edu)
Drop-in Hours	Tuesday 4:00–5:00pm (CDS 1326)
Meetings	Tuesday & Thursday 2–3:15pm (MCS B37) Friday 9:05–9:55am (CAS 222) or 10:10–11:00am (IEC B12)
Textbook	*Stochastic Methods for Algorithms* (available on Perusall)
Course Content	Blackboard and Perusall
Communication	Perusall
Other software	Python, LaTeX editor (e.g., Overleaf, TeXShop, TeXstudio, or LyX)
Hub Units	Creativity/Innovation (CRI) and Writing-Intensive Course (WIN)
Prerequisites	See below

Welcome / About This Course

"There is nothing more practical than a good theory." – *Kurt Lewin*

This course concerns the use of stochastic processes for designing and analyzing algorithms, with a focus on applications in statistics and machine learning. You will learn about core concepts and results about stochastic processes, then use this machinery to (1) develop algorithms and (2) characterize their statistical and numerical properties. The two recurring applications will concern (large-scale) optimization using stochastic gradients and sampling from complex distributions such as Bayesian posterior distributions and energy-based models using Markov chain Monte Carlo algorithms. In addition to statistics and machine learning, the optimization and sampling algorithms derived and analyzed have diverse applications including to problems in computer science, physics, chemistry, ecology, biology, and operations research. As such, a strong emphasis is placed on the ability to describe and investigate the practical implications of the results.

Considering the application-oriented motivations, the stochastic processes material focuses on intuitive understanding of definitions and theoretical properties rather than their rigorous development. This approach will allow us to almost immediately begin to investigate algorithms and their properties. Toward the same ends, we will mostly forgo using advanced tools probability theory and functional analysis. Thus, we will tend not to dwell on regularity conditions and other (still important!) technicalities. However, throughout the course I will provide pointers to sources where rigorous treatments can be found.

When describing the practical implications of algorithm analyses and empirical results aimed at validating such analyses, clear communication is a must. Thus, you will do a significant amount of expository writing in this course. These writing opportunities will enable you to learn how to synthesize complex mathematics and sophisticated experiments in a clear way that can be understood by someone who is familiar with the underlying algorithms but not the mathematical theory.

Learning Outcomes

After successful completion of this course, you will be able to...

1. Explain how to use the theory of Markov chains and stochastic differential equations to analyze Markov chain Monte Carlo and stochastic optimization algorithms.

2. Evaluate the implications of theoretical analyses of Markov chain Monte Carlo and stochastic optimization algorithms, including law of large numbers, geometric ergodicity, central limit theorems, error analyses, and scaling limits.
3. Interpret results and arguments in the modern statistics and machine learning literature about the design and analysis of Markov chain Monte Carlo and stochastic optimization algorithms.

Prerequisites

The formal prerequisites are (recommended versions marked with a *):

- **Writing** *[undergraduate only]***:** First-Year Writing Seminar (e.g., WR 120)
- **Programming:** *CAS CS111, *CDS DS110, *ENG EK125, or equivalent.
- **Vector calculus:** *CAS MA225, *CAS CS235, *CDS DS122, or equivalent
- **Linear algebra:** *CAS MA242, *CAS CS132, *CDS DS121, or equivalent
- **Probability theory:** *CAS MA581, CAS CS237, ENG EK381, *ENG EK500, or equivalent.

However, there are some additional "soft" requirements/recommendations that do not correspond to a specific course:

- You must **have some experience writing scientific code**, preferably in Python (which will be used for all course assignments). For example, you may have taken CAS MA415, CAS MA615, CDS DS210, or a more advanced CS programming course.
- You must **have experience writing rigorous mathematical proofs**, with proofs by induction being particularly important. For example, you may have taken one or more 500-level proof-based math courses, or a course on analysis algorithms, or a computer science theory course.
- Having some **previous exposure to stochastic process theory** will be invaluable but is not strictly required. For example, you may have taken CAS MA583, CAS MA783, or ENG EC505.

- Having some **previous exposure to statistics and/or machine learning** – particularly regression, statistical models, Markov chain Monte Carlo, and/or (stochastic) optimization – is very helpful but not required.

Communication

Course content will primarily be on Perusall. Blackboard will be used for homework submission and recording grades.

All communication should be done via Perusall. I will typically respond within 24 hours during the week. However, I have three kids, so my weekends are very busy with family commitments. Therefore, you should not expect me to be responsive between approximately 5pm Friday and 9am Monday.

I encourage you to come to my drop-in hours sometime in the first month of class, as I would very much like to get to know you all. As the semester gets going, you can come to drop-in hours to discuss the class material – or math, statistics, and machine learning in general. If you can't make it during my regularly scheduled hours or if you need to talk to me privately, please email me and I will do my best to schedule a time for us to meet.

Textbook

The primary textbook will be the provided course notes. However, the following may be useful supplements with more detailed treatments of certain topics:

- Probability theory roughly at the same level as this course ("just short of measure theory"): *Probability and Random Processes* by Grimmett & Stirzaker

- Background on (probabilistic) models and algorithms: *Machine Learning: A Probabilistic Approach* by Murphy[1]
- Mathematical background: *Mathematics for Machine Learning* by Deisenroth, Faisal & Ong[2]
- Detailed but accessible treatment of stochastic differential equations: *Applied Stochastic Differential Equations* by Särkkä & Solin[3]

Equity/Inclusion

As your professor, I pledge to work to create an equitable learning environment where all students belong. Statistics can seem like an "objective" subject, but like all education, mathematical education is a cultural activity, and many aspects of who we are affect how we experience statistics classes. There is considerable research on how students' "sense of belonging" in their classes impacts their learning. Feeling like you don't belong in a class can impact your cognitive load and diminish your ability to focus on the mathematical and statistical content; so, for example, you might use your brain to worry whether the professor or your classmates will think you're stupid if you make a mistake; to deal with the impact of a racist,

[1] Available at https://probml.github.io/pml-book/book0.html

[2] Available at https://mml-book.github.io

[3] Available at https://users.aalto.fi/~asolin/sde-book/sde-book.pdf

sexist, classist, xenophobic, homophobic, transphobic, or ableist comments; or to wonder whether the other students are better than you.

Although frustration and struggle are also part of the educational enterprise, ultimately, I see my work as setting up situations where all students can experience the joy of mathematics and statistics, to feel a sense of belonging, and to use our brains for learning.

Assignments

There will be 6 types of assignments:

1. **Pre-class Assignments.** In addition to readings, there will often be exercises and questions to be submitted prior to class. These will help you learn the basics, reflect on new material, and prepare you for lectures and discussions during class time. There are 20 of these planned.
2. **Exit Tickets.** At the end of each week, you will complete an "exit ticket" survey through Blackboard that will ask you questions about how the past week went (What did you learn? What are you confused about?). These exit tickets facilitate meta-cognition (e.g., ask yourself questions like "Which topics/concepts do I understand?" and "Which ones am I having more difficulty with?"). They also help me to understand which topics or concepts I haven't done a good enough job teaching you. There are 12 of these planned.
3. **Problem Sets.** More challenging problems with a mix of derivations, proofs, and coding problems. These will assess your technical knowledge and auxiliary skills (e.g., LaTeX, scientific programming) and can be done in collaboration with other students. There are 7 of these planned.
4. **Mini Project.** This will be more cohesive and open-ended than problem sets. It will feature a mix of scientific programming, mathematical derivations, and some synthesis that provides a chance to develop writing skills. There will be one of these.

5. **Article Evaluations.** These will involve reading a published paper, then evaluating the writing and scientific content following a provided rubric. There are two of these planned.
6. **Final Project.** The final project will involve further investigation of a topic closely related to what is covered in the course. After critiquing a paper in your second Article Evaluation, you will make a novel contribution that is motivated by your critique. The final project is similar in scope to the Mini Project.

There is no final exam in this course.

Grades

Philosophy

Your grade in the course is earned by **demonstrating evidence of skill on the main concepts in the course** and by **showing appropriate engagement with the course**. This is done by completing the assignments outlined above, at a reasonably high level of quality. The class should be a learning community, where students support each other to increase everyone's learning. You should not be competing with other students for grades. You should have some choice in your trajectory through the course.

Therefore, in this course, **there are no points or percentages** on any items. Instead, the work you turn in will be evaluated against **quality standards** that will be made clear on each assignment. If your work meets the standard, then you will receive full credit for it. Otherwise, you will get helpful feedback and, on most items, the chance to reflect on the feedback, revise your work, and then resubmit it for regrading.

This feedback loop represents and supports the way that people learn: By trying things, making mistakes, reflecting on those mistakes, and then trying again. **You can make mistakes without penalty** if you *eventually* demonstrate evidence of skill.

Determining Your Course Grade

The individual kinds of assignments are marked as follows:

Assignment	How it's marked
Pre-class Assignments	**Pass** or **No Pass**
Exit Tickets	**Pass** or **No Pass**
Problem Sets	**Pass** or **Not Yet**
Projects	**Pass (P)**, **Almost Pass (AP)**, **Partial Pass (PP)**, or **Not Yet**
Article Evaluations	**Pass (P)**, **Almost Pass (AP)**, **Partial Pass (PP)**, or **Not Yet**

The criteria for each mark are explained below in the "Grading Criteria" section below.

Your final grade in the course is determined by the following table. Each grade has a *requirement* specified in its column in the table. **To earn a grade, you will need to meet *all* the requirements in the column for that grade.** Put differently, your grade is the **highest** grade level for which **all** the requirements in a column of the table have been met or exceeded.

	A	B	C	D
Pre-class Assignments + Exit Tickets passed (32)	28	22	16	8
Problem Sets passed (6)	5	4	3	1
Project minimal marks (2)	1 P + 1 AP	1 P + 1 PP or 2 AP	1 AP + 1 PP or 1 P	1 AP or 2 PP
Article Evaluation minimal marks (2)	1 P + 1 AP	1 P + 1 PP or 2 AP	1 AP + 1 PP or 1 P	1 AP or 2 PP

A grade of **F** is given if all the requirements for a **D** are not met.

Plus/minus grades: Plus/minus grades will be assigned at my discretion based on how close you are to the next higher grade level.

Grading Criteria

- **Pre-class Assignments** and **Exit Tickets:** A **Pass** mark is given if it is turned in before its deadline and if each item has a response that represents a good faith effort to be right. Mistakes are not penalized. A **No Pass** is given if an item is left blank (even accidentally), has an answer but it shows insufficient effort (including responses like "I don't know"), or if it is turned in late.
- **Problem Sets**, **Mini Projects**, **Article Evaluations**, and **Final Project**: Each type has its own standards which will be included with the assignment.

Revising and Resubmitting Work

Instead of earning partial credit, on most assignments you will have the opportunity to revise and resubmit your work based on feedback that I or the TF provide, if the work doesn't meet its standard for acceptability. **Mistakes, and work that does not meet the standard for acceptability, are typically not penalized.** Instead, if your work has enough room for improvement that it would benefit from redoing parts of it or the whole thing, you'll get the chance to do so. This again is because **human beings learn from making mistakes and fixing them with feedback and reflection.**

- **Pre-class Assignments** and **Exit Tickets** <u>may not</u> be revised or resubmitted. They are graded on completeness and effort only, and therefore can only be done once.
- Each **Problem Set** may be revised **twice**. One problem set revision may be submitted each week, where the week begins at midnight on Monday and ends at 11:59pm on Sunday. To revise, simply reflect on the feedback that's given, make corrections or rewrites to the

original, and upload the work again to Blackboard, then submit a regrade request using this form.

- The **Mini Project** and each **Article Evaluation** may be revised **once**. One such revision may be submitted each week, where the week begins at midnight on Monday and ends at 11:59pm on Sunday. To revise, simply reflect on the feedback that's given, make corrections or rewrites to the original, upload the work again to Blackboard, then submit a regrade request using this form.
- **Final Project:** The project may not be revised, but you will have the chance to get feedback on a draft and some partial credit.

Tokens

We all have things – good and bad – that come up in our lives that affect our ability to get work done on time. So, at the beginning of the semester, you will receive 5 "tokens," which you can use to break course rules in prescribed ways:

- You may spend **1 token** to convert a **No Pass** on a Pre-class Assignment or Exit Ticket to a **Pass**.
- You may spend **1 token** to submit an additional Problem Set for regrading in a week.
- You may spend **1 token** to receive a 5-day extension on an assignment with a flexible deadline. Assignments with flexible deadlines will be explicitly labeled as such.
- You may spend **2 tokens** to submit a Mini Project or Article Evaluation for a second regrade. However, note that you can still only submit one regrade per week.

In general, I will offer a 24-hour grace period for late submissions (excluding pre-class assignments and others explicitly labeled as not having a flexible deadline).

Use of Generative AI

I will follow the Faculty of Computing & Data Science's **Generative AI Assistance (GAIA) Policy**, which can be found in full at https://www.bu.edu/cds-faculty/culture-community/gaia-policy/. You must read the GAIA policy if you plan to use Generative AI tools. The intent of the policy is as follows:

> *Students should learn how to use AI text generators and other AI-based assistive resources (collectively, AI tools) to enhance rather than damage their developing abilities as writers, coders, communicators, and thinkers. Instructors should ensure fair grading for both those who do and do not use AI tools. The GAIA policy stresses transparency, fairness, and honoring relevant stakeholders such as students eager to learn and build careers, families who send students to the university, professors who are charged with teaching vital skills, the university that has a responsibility to attest to student competency with diplomas, future employers who invest in student because of their abilities and character, and colleagues who lack privileged access to valuable resources. To that end, the GAIA policy adopts a few commonsense limitations on an otherwise embracing approach to AI tools.*

If you have questions or aren't sure if something is allowed, ask me!

Collaboration and Academic Honesty

I strongly encourage you to collaborate with your classmates whenever it is allowed. However, realize that collaboration is not always allowed, and, in all cases, there are limitations on how you can collaborate. In particular:

- On **Exit Tickets**, **Article Evaluations**, and the **Final Project**, your work must represent your own understanding in your own words using your own code. You may not use solutions, directly or indirectly, from any sources not explicitly allowed – including other students, past students, online sources, or other textbooks.

- On **all other assignments**, you may collaborate with others, but you must contribute significantly to the assignment, and your work must represent your own understanding in your own words and using your own code.

You are responsible for understanding this policy and Boston University's Academic Conduct Code. Violations will result, at minimum, in a mark of **No Pass / Not Yet** on the assignment with no chance to resubmit. Serious or repeat violations of this policy will result in increasingly unfortunate consequences, including being barred from further submissions of the assignment, or even receiving an **F** in the course.

Hub Learning Outcomes

Creativity/Innovation

Students will earn a Creativity/Innovation credit by satisfy the learning outcomes as follows:

1. *Students will demonstrate understanding of creativity as a learnable, iterative process of imagining new possibilities that involves risk-taking, use of multiple strategies, and reconceiving in response to feedback, and will be able to identify individual and institutional factors that promote and inhibit creativity.*
2. *Students will be able to exercise their own potential for engaging in creative activity by conceiving and executing original work either alone or as part of a team.*
   Students will demonstrate creativity in structuring and creating original mathematical proofs and implementing numerical experiments to validate their results. Students will improve their solutions and results on assignments through an iterative process of peer and instructor feedback. Students will be asked to show increasing degrees of creativity in the progression from In-class Exercises to Problem Sets, Article Evaluations, and Mini Projects to the Final Project. Overall, students will exercise their own potential for

creativity in all assignment types, which will include both individual and small-group work.

Writing-Intensive Course

Students will earn a Writing-Intensive Course credit by satisfy the learning outcomes as follows:

1. *Students will be able to craft responsible, considered, and well-structured written arguments, using media and modes of expression appropriate to the situation.* Students will write formal mathematical proofs and expository explanations of mathematical and experimental results through In-class Exercises, Problem Sets (4–8 pages total), the Mini Project (10–12 pages total), Article Evaluations (8–10 pages total), and the Final Project (10–15 pages). They will improve the clarity and style of their writing through peer and instructor feedback, which they will incorporate into revisions that they will resubmit. The course grade will be almost completely determined by student performance on these assignment types.
2. *Students will be able to read with understanding, engagement, appreciation, and critical judgment.*
   By completing structured Problem Sets and homework readings from the primary literature, students will learn how to *(i)* read, understand, and deconstruct complex mathematical proofs and *(ii)* analyze and critically judge numerical experiments designed to support mathematical theory of algorithms. Readings will also be discussed during class in small groups and with the whole class. Students will also learn how to read and respond to each other's writing in ways that are beneficial and useful. As a class and in smaller groups, including instructors and students, students will learn how to give and respond to feedback on assignments.

3. *Students will be able to write clearly and coherently in a range of genres and styles, integrating graphic and multimedia elements as appropriate.*
   N/A

Tentative Schedule

The course will consist of 6 modules, each approximately 2 weeks long:

0. **Preliminaries (Weeks 1–2).** Course overview, review of background material.
1. **Probability theory and Markov chains (Weeks 2–3).** Review and extension of probability theory and introduction to Markov chains.
2. **Markov chains for stochastic optimization (Weeks 4–7).** Convex analysis, Taylor series error, and the application of Markov chain theory to analyzing the error of SGD
3. **Convergence and stationarity of Markov chains (Weeks 8–9).** Stationary distributions of Markov chains, convergence of SGD with constant step size, and introduction to Markov chain Monte Carlo (MCMC) algorithms.
4. **Markov chain limit theorems (Weeks 10–11).** Central limit theorems for Markov chains, including for averaged SGD and MCMC.
5. **Stochastic differential equations (Weeks 11–13).** Basics of SDEs including Itô calculus and generators. Application of Langevin diffusions and scaling limits to design and analyze SGD and MCMC algorithms.

I reserve right to make changes to the assessment system and to other aspects of the syllabus to better meet the needs of students in the class. If appropriate, students might have input into these changes. Any changes will be clearly documented with sufficient notice for students to adapt.

Community of Learning: Additional Class and University Policies

Accommodations for Students with Documented Disabilities

If you are a student with a disability or believe you might have a disability that requires accommodations, please contact the Office for Disability Services (ODS) at (617) 353-3658 or access@bu.edu to coordinate any reasonable accommodation requests. ODS is located at 19 Buick Street.

Attendance & Absences

Attendance is critical to your success in this course, and if you want to do well, you need to be present and prepared. In this class you'll be part of a learning community, and we will miss you when you aren't here. However, there are many reasons why people need to miss labs and discussions, such as illness, religious holidays, family emergencies and milestones, civic responsibilities (jury duty, citizenship ceremonies, etc.), dangerous commutes during bad weather, etc. If you know in advance that you're going to miss a lab or discussion section, please let the lab instructor or TF know. If you do miss a lab or discussion section, it's your responsibility to find out what you missed.

If there's something in your life that's interfering with your ability to engage in the course, please come talk to me about it (you only need to share the details you want to share). Note that you don't need to bring me doctor's notes; I don't read them. Also note that I have considerable experience with chronic illness and how it impacts academic life.

Incompletes

If you have health issues, an emergency, or find yourself in other difficult circumstances that affect your performance in the course, you may be eligible for an incomplete, where you would finish the work after the semester ends. Please feel free to talk to me about this possibility.

Acknowledgments

Parts of this syllabus are borrowed from / based on the syllabi of Debra Borkovitz (BU, CAS MA 293) and Robert Talbert (GVSU, MTH 350).

Calculus MA 123

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 62 of 472 | **Item #:** 2.2.33
Difficulty: Hard
Median time: 2m 44s
Correct on first try: 34.8%

Section 2.2 | **Objective:** Find limits from a graph.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Determine whether the following statements are true and give an explanation or counterexample. Complete parts (a) through (e).

d. $\lim_{x\to 0} \sqrt{x} = 0$ (Hint: Graph $y = \sqrt{x}$). Choose the correct answer below and, if necessary, fill in the answer box to complete your choice.

- A. This statement is false because the value of $\lim_{x\to 0} \sqrt{x} = 0$ does exist but is equal to the nonzero value ☐.
  (Simplify your answer.)
- B. This statement is true because because $\lim_{x\to 0^-} \sqrt{x} = 0$ and $\lim_{x\to 0^+} \sqrt{x} = 0$.
- ★ C. This statement is false because the value of $\lim_{x\to 0^-} \sqrt{x}$ does not exist if x is a real number.
- D. This statement is false because the value of $\lim_{x\to 0^+} \sqrt{x}$ does not exist if x is a real number.

e. $\lim_{x\to \frac{\pi}{2}} \cot x = 0$ (Hint: Graph $y = \cot x$). Choose the correct answer below and, if necessary, fill in the answer box to complete your choice.

- A. This statement is false because the value of $\lim_{x\to \frac{\pi}{2}} \cot x$ does exist but is equal to the nonzero value ☐.
  (Simplify your answer.)
- B. This statement is false because the value of $\lim_{x\to \frac{\pi}{2}^-} \cot x$ does not exist.
- C. This statement is false because the value of $\lim_{x\to \frac{\pi}{2}^+} \cot x$ does not exist.
- ★ D. This statement is true because $\lim_{x\to \frac{\pi}{2}^-} \cot x = 0$ and $\lim_{x\to \frac{\pi}{2}^+} \cot x = 0$.

Add instructor tip **Ask the publisher** Similar question

Show completed problem ○ Work problem as student ☐ Student to show work **Question points:** 1 Scoring options

Show Answer | Reload | Copy and Edit | ◂ Previous | Add | Next ▸

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 137 of 472 | **Item #:** 2.3.71

Section 2.3 | **Objective:** Answer conceptual questions involving techniques to compute limits.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Determine whether the following statements are true and give and explanation or counterexample. Assume a and L are finite numbers.

c. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = L$, then $f(a) = g(a)$. Choose the correct answer.

- A. The statement is true. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = L$, then $\lim_{x\to a} (f(x) - g(x)) = 0$. Because $\lim_{x\to a} (f(x) - g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x)$, it follows that $f(a) = g(a)$.
- ★ B. The statement is false. If $f(x) = \frac{x^3 - 8}{4(x-2)}$ and $g(x) = \frac{x^2 + 2x + 4}{4}$, then $\lim_{x\to 2} f(x) = \lim_{x\to 2} g(x) = 3$ and $g(2) = 3$, but $f(2)$ does not exist.
- C. The statement is true. Because limits describe behaviors of functions near points, it follows that functions that have the same limit at a point must also have the same function value there.
- D. The statement is false. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = L$, then $f(a) = g(a)$ if and only if $f(x)$ and $g(x)$ are both polynomial functions.

d. The limit $\lim_{x\to a} \frac{f(x)}{g(x)}$ does not exist if $g(a) = 0$. Choose the correct answer.

- A. The statement is true. The indicated limit only exists provided $\lim_{x\to a} g(x) \neq 0$. If $g(a) = 0$, then $\lim_{x\to a} g(x) = 0$. Therefore, if $g(a) = 0$, then $\lim_{x\to a} \frac{f(x)}{g(x)}$ does not exist.
- ★ B. The statement is false. If $f(x) = x^2 - 4$ and $g(x) = x - 2$, then $g(2) = 0$ but $\lim_{x\to 2} \frac{f(x)}{g(x)} = 4$.
- C. The statement is true. If $g(a) = 0$, then the expression $\frac{f(a)}{g(a)}$ is undefined. Because $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f(a)}{g(a)}$, it follows that if $g(a) = 0$, then $\lim_{x\to a} \frac{f(x)}{g(x)}$ does not exist.
- D. The statement is false. If $f(x) = x^2 - 4$ and $g(x) = x + 2$, then $g(2) = 0$ but $\lim_{x\to 2} \frac{f(x)}{g(x)} = 0$.

e. If $\lim_{x\to 1^+} \sqrt{f(x)} = \sqrt{\lim_{x\to 1^+} f(x)}$, it follows that $\lim_{x\to 1} \sqrt{f(x)} = \sqrt{\lim_{x\to 1} f(x)}$. Choose the correct answer.

- A. The statement is true. It follows from the limit law $\lim (f(x))^{n/m} = \left(\lim f(x)\right)^{n/m}$ with $a = 1$, $n = 1$, and $m = 2$.

Add instructor tip **Ask the publisher**

Show completed problem ○ Work problem as student ☐ Student to show work **Question points:** 1 Scoring options

Show Answer | Reload | Copy and Edit | ◂ Previous | Add | Next ▸

Linear Algebra MA 242

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 24 of 35 | **Item #:** 1.5.38
Difficulty: Easy
Median time: 1m 6s
Correct on first try: 80.2%

Section 1.5 | **Objective:** Determine if a matrix equation has a unique or nontrivial solution.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Suppose $A\mathbf{x} = \mathbf{b}$ has a solution. Explain why the solution is unique precisely when $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Choose the correct answer.

A. Since $A\mathbf{x} = \mathbf{b}$ is consistent, then the solution is unique if and only if there is at least one free variable in the corresponding system of equations. This happens if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

B. Since $A\mathbf{x} = \mathbf{b}$ is consistent, its solution set is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$. So the solution set of $A\mathbf{x} = \mathbf{b}$ is a single vector if and only if the solution set of $A\mathbf{x} = \mathbf{0}$ is a single vector, and that happens if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

C. Since $A\mathbf{x} = \mathbf{b}$ is inconsistent, then the solution set of $A\mathbf{x} = \mathbf{0}$ is also inconsistent. The solution set of $A\mathbf{x} = \mathbf{0}$ is inconsistent if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

D. Since $A\mathbf{x} = \mathbf{b}$ is inconsistent, its solution set is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$. For $A\mathbf{x} = \mathbf{b}$ to be inconsistent, $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Add instructor tip Ask the publisher Textbook

Show completed problem Work problem as student Student to show work **Question points:** 1 Scoring options

Show Answer Reload Copy and Edit Previous Add Next

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 15 of 35 | **Item #:** 1.5.27

Section 1.5 | **Objective:** Determine the validity of statements about solution sets of linear equations.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Determine whether the statement below is true or false. Justify the answer.

A homogeneous equation is always consistent.

Choose the correct answer below.

A. The statement is false. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one nontrivial solution. Thus a homogenous equation is always inconsistent.

B. The statement is false. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one solution, namely, $\mathbf{x} = \mathbf{0}$. Thus a homogenous equation is always inconsistent.

C. The statement is true. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one solution, namely, $\mathbf{x} = \mathbf{0}$. Thus a homogenous equation is always consistent.

D. The statement is true. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one nontrivial solution. Thus a homogenous equation is always consistent.

Add instructor tip Ask the publisher Textbook

Show completed problem Work problem as student Student to show work **Question points:** 1 Scoring options

Show Answer Reload Copy and Edit Previous Add Next

Appendix F: Meeting Dates

Committee Meetings (13)

WID Writing Consultant:	Jessica Kent
Math and Statistics Faculty:	Glenn Stevens Margaret Beck Jonathan Huggins Li-Mei Lim Dan Sussman Matt Szczesny

Friday, September 22, 2023
Friday, October 6, 2023
Friday, October 27, 2023
Friday, November 17, 2023
Friday, December 8, 2023
Friday, December 15, 2023
Wednesday, January 24, 2024
Wednesday, February 7, 2024
Wednesday, February 21, 2024
Wednesday, March 6, 2024
Wednesday, April 3, 2024
Wednesday, April 24, 2024
Tuesday, May 14, 2024

Statistics-focused Meeting (1)

Attendees: Kent, Beck, Huggins, Lim, Sussman

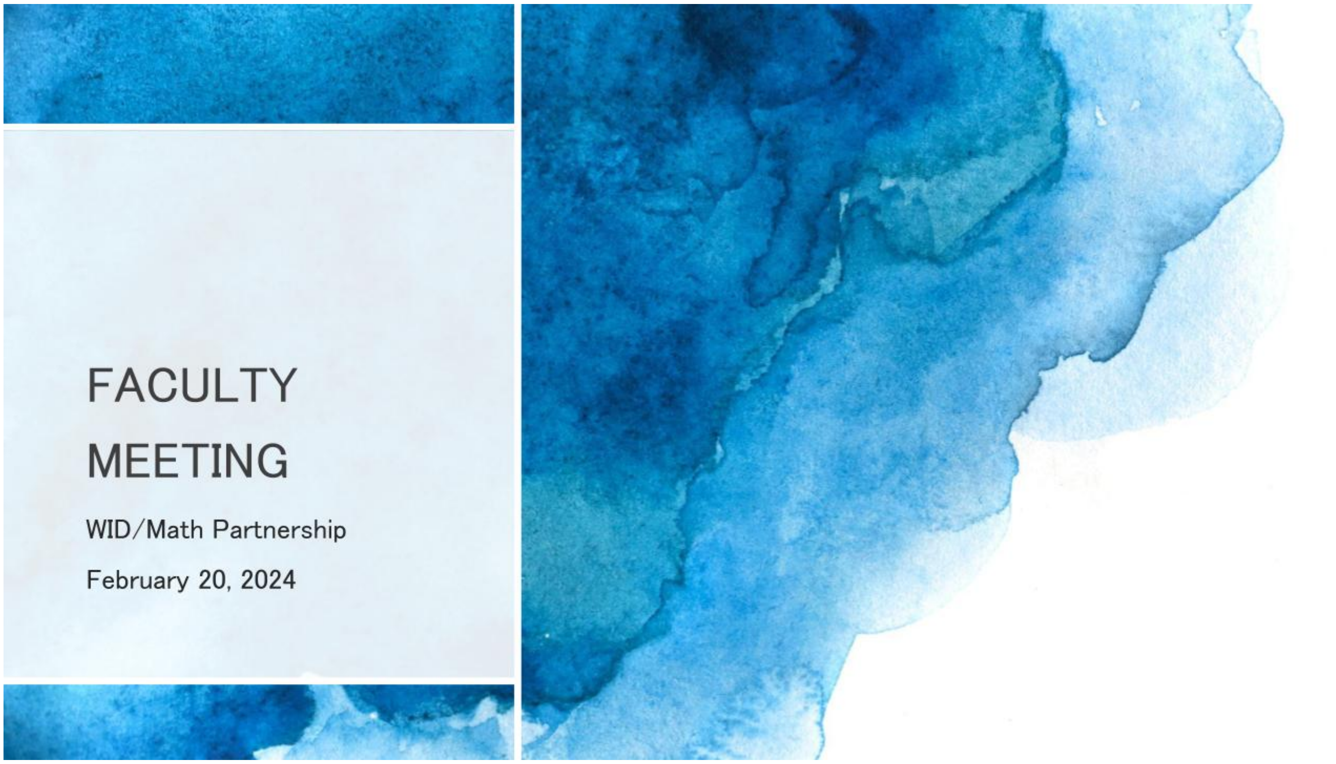
Friday, December 15, 2023 after the full committee meeting

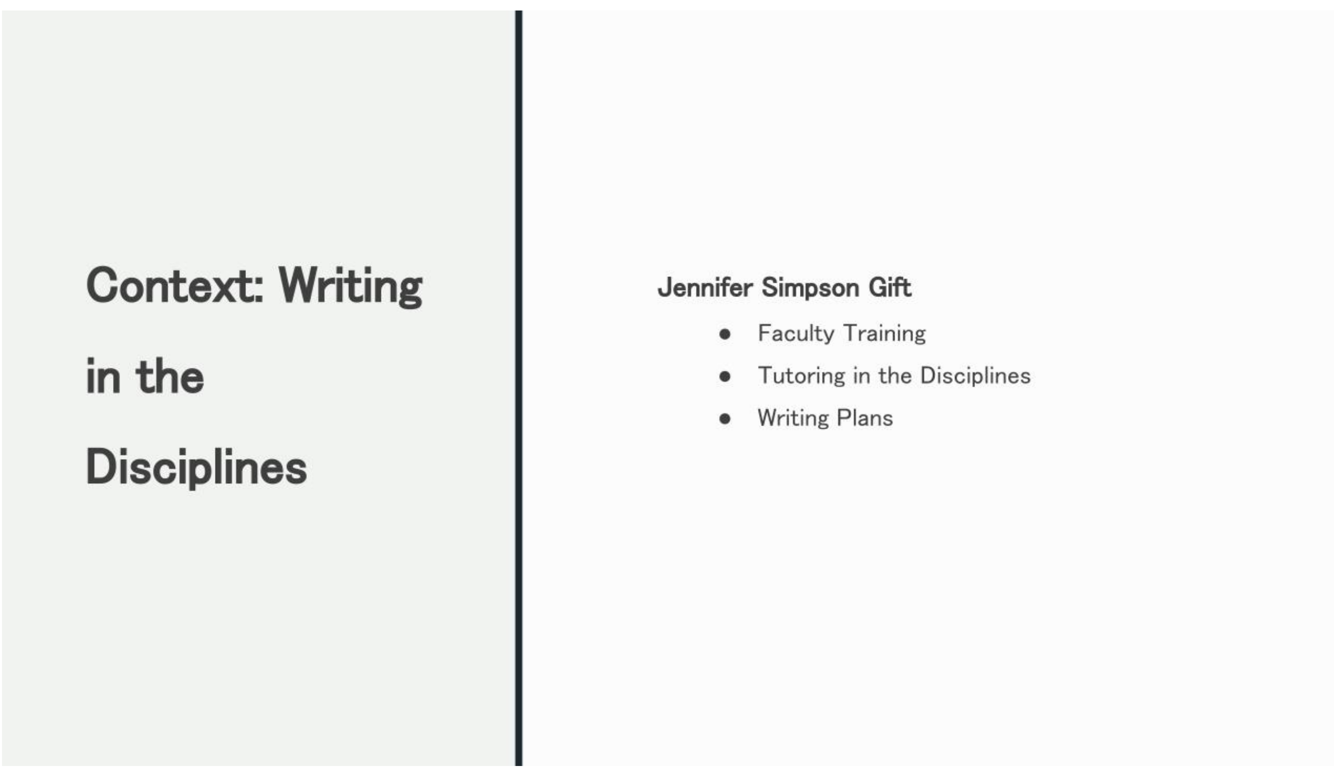
All-faculty Meeting (1)

Presenters: Kent, Beck, Huggins, Lim, Sussman, Szczesny

Tuesday February 20, 2024

Appendix G: Slides from February 20, 2024 Math/Statistics Faculty Meeting





Context: Writing in the Disciplines

Math Writing Plan Process

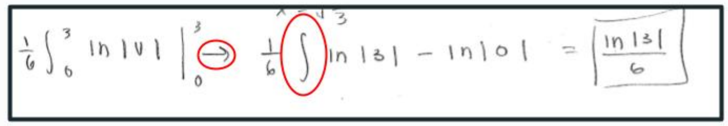
- Define “writing” in math
- Assess current practices
- Develop recommendations
- Write plan (including proposed timeline for implementation)

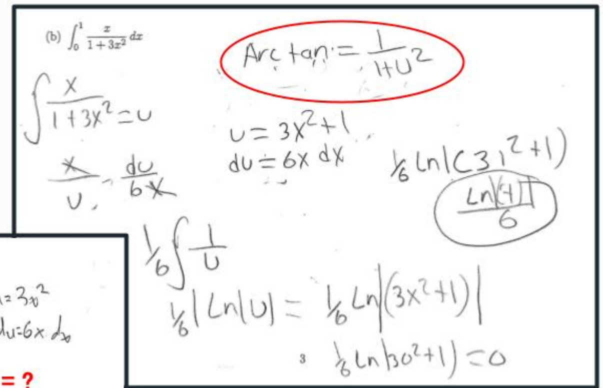
What is Writing in Math/Stats?

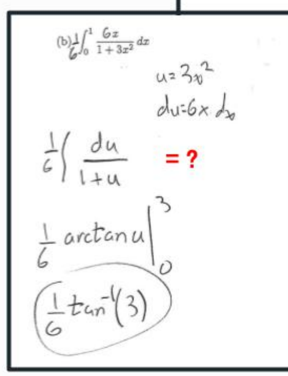
We are thinking broadly about writing, including:

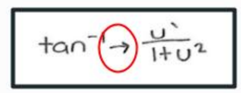
- Solutions to calculus or statistics problems
- Interpretation of calculations or procedures (e.g. “*f’(2) = 3 means….*” or “*calculating the confidence interval means…*”)
- Stating the problem you need to solve; drawing a picture of the setting of the problem; doing the calculations necessary to solve the problem
- Proof writing
- Lab reports and project reports
- Code writing
- Data visualization
- Expository writing explaining math/statistics ideas for a broader audience

What we want to address









(And many other examples from statistics, proof-based classes, etc.)

Why do this?

- Problems we see in MA 123 become bigger problems later, e.g., with using precise mathematical language in proofs, etc.
- Students are often overwhelmed when they take their first upper-level or proof-based class.
- General need for improved communication of mathematics and statistics concepts

Preliminary suggestions for 100-level courses

- Discuss correct usage of the equals sign =
- Give students examples of correct and incorrect = usage (see examples on "What we want to address" slide)
- Allocate points on each exam for correct = usage
- Choose MathLab problems that involve language (see examples on forthcoming slide)

Preliminary suggestions for 200-level courses

- Discuss the notion of implication or assertion
- Discuss correct usage of the arrow symbol ->
- Give students examples of correct and incorrect usage of -> (see examples on "What we want to address" slide)
- Reinforce correct usage of =
- Allocate points on each exam for correct = and -> usage
- Continue to choose MathLab problems that involve language (see examples on forthcoming slide)
- Assign true/false questions, particularly those of the form "Ture or False: A implies B."
- In MA213/214: frame coding as writing, specifically as very precise writing; use coding to understand and explore ideas

Preliminary suggestions for upper division courses

- LaTeX: create asynchronous, flipped, department-wide learning module that students must complete independently before their first upper-division course
- Reinforce correct usage of = and ->, and discuss the usage of words like "assume" "implies" "therefore" "there exists" "for all" etc.
- Include more intentional and writing-specific activities and assignments in courses, with details dependent on the course. Examples include
  - Reflection assignments
  - Student blogs
  - Opportunities for revisions of proofs
  - Reading/discussing essays like "Good Mathematical Writing" by Francis Su.
- Reinforce correct descriptions of statistical analyses

Create versions of 225/242 for majors?

- Examples of Universities that do this: Boston College, Carnegie Mellon, Cornell, Penn, UC Irvine, ...
- Courses would be required for math/stats majors
- Could be useful as transitional courses
- More rigorous/language-intensive versions
- Could replace current "honors" versions
- Introduce students to the concept of proofs
- Could combine into one year-long sequence (many books do this)

Create version of 415 for majors?

- Use as a transitional course for statistics majors that builds key writing skills needed for 500-level courses…
- Build toolkit for communicating about statistical concepts
- Learn more advanced visualization and data processing tools
- Introduction to LaTeX
  - Could be limited to small amounts within R markdown documents
- Gain scientific coding experience and skills
  - Currently CS111 only required programming course
  - This could change in the next few years with MA213/214 revamp introducing R, so "415 for majors" could build on that foundation
- For-majors version would focus more on communicating about advanced statistical concepts

Low on-ramp – what can be done in large lower-division courses ?

What can we do to get students thinking about the correct use of mathematical language in large lower-division courses?

- True/False problems
- Multiple choice problems focussed on correct reasoning

Some problems like these are even available in Pearson's MyMathLab question banks for Calculus/Linear Algebra

Examples of departmental resources that could be created

- Examples of good/bad writing at various levels to share with students in our courses
- Brief, asynchronous, flipped learning modules, for example on LaTeX/Overleaf
- Template for incorporating more writing into courses
  - Example syllabi
  - Rubrics for writing assignments
  - Division of labor between faculty/TF/grader for providing feedback on writing
  - Discussion activities such as essays like “Some Guidelines for Good Mathematical Writing” by Francis Su.

Linear Algebra – MA 242

[redacted]

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 15 of 35 | **Item #:** 1.5.27

Section 1.5 | **Objective:** Determine the validity of statements about solution sets of linear equations.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Determine whether the statement below is true or false. Justify the answer.

A homogeneous equation is always consistent.

Choose the correct answer below.

A. The statement is false. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one nontrivial solution. Thus a homogenous equation is always inconsistent.

B. The statement is false. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one solution, namely, $\mathbf{x} = \mathbf{0}$. Thus a homogenous equation is always inconsistent.

C. The statement is true. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one solution, namely, $\mathbf{x} = \mathbf{0}$. Thus a homogenous equation is always consistent.

D. The statement is true. A homogenous equation can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ always has at least one nontrivial solution. Thus a homogenous equation is always consistent.

Add instructor tip   Ask the publisher   Textbook

Show completed problem   Work problem as student   Student to show work   **Question points:** 1   Scoring options

Show Answer   Reload   Copy and Edit   ◂Previous   Add   Next ▸

Linear Algebra – MA 242

[redacted]

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 24 of 35 | **Item #:** 1.5.38
Difficulty: Easy
Median time: 1m 6s
Correct on first try: 80.2%

Section 1.5 | **Objective:** Determine if a matrix equation has a unique or nontrivial solution.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Suppose $A\mathbf{x} = \mathbf{b}$ has a solution. Explain why the solution is unique precisely when $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Choose the correct answer.

A. Since $A\mathbf{x} = \mathbf{b}$ is consistent, then the solution is unique if and only if there is at least one free variable in the corresponding system of equations. This happens if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

B. Since $A\mathbf{x} = \mathbf{b}$ is consistent, its solution set is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$. So the solution set of $A\mathbf{x} = \mathbf{b}$ is a single vector if and only if the solution set of $A\mathbf{x} = \mathbf{0}$ is a single vector, and that happens if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

C. Since $A\mathbf{x} = \mathbf{b}$ is inconsistent, then the solution set of $A\mathbf{x} = \mathbf{0}$ is also inconsistent. The solution set of $A\mathbf{x} = \mathbf{0}$ is inconsistent if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

D. Since $A\mathbf{x} = \mathbf{b}$ is inconsistent, its solution set is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$. For $A\mathbf{x} = \mathbf{b}$ to be inconsistent, $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Add instructor tip   Ask the publisher   Textbook

Show completed problem   Work problem as student   Student to show work   **Question points:** 1   Scoring options

Show Answer   Reload   Copy and Edit   ◂Previous   Add   Next ▸

Calculus – MA 123

[redacted]

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 137 of 472 | **Item #:** 2.3.71

Section 2.3 | **Objective:** Answer conceptual questions involving techniques to compute limits.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Determine whether the following statements are true and give and explanation or counterexample. Assume a and L are finite numbers.

c. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = L$, then $f(a) = g(a)$. Choose the correct answer.

- A. The statement is true. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = L$, then $\lim_{x\to a}(f(x) - g(x)) = 0$. Because $\lim_{x\to a}(f(x) - g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x)$, it follows that $f(a) = g(a)$.
- ★ B. The statement is false. If $f(x) = \dfrac{x^3 - 8}{4(x-2)}$ and $g(x) = \dfrac{x^2 + 2x + 4}{4}$, then $\lim_{x\to 2} f(x) = \lim_{x\to 2} g(x) = 3$ and $g(2) = 3$, but $f(2)$ does not exist.
- C. The statement is true. Because limits describe behaviors of functions near points, it follows that functions that have the same limit at a point must also have the same function value there.
- D. The statement is false. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = L$, then $f(a) = g(a)$ if and only if $f(x)$ and $g(x)$ are both polynomial functions.

d. The limit $\lim_{x\to a} \dfrac{f(x)}{g(x)}$ does not exist if $g(a) = 0$. Choose the correct answer.

- A. The statement is true. The indicated limit only exists provided $\lim_{x\to a} g(x) \neq 0$. If $g(a) = 0$, then $\lim_{x\to a} g(x) = 0$. Therefore, if $g(a) = 0$, then $\lim_{x\to a} \dfrac{f(x)}{g(x)}$ does not exist.
- ★ B. The statement is false. If $f(x) = x^2 - 4$ and $g(x) = x - 2$, then $g(2) = 0$ but $\lim_{x\to 2} \dfrac{f(x)}{g(x)} = 4$.
- C. The statement is true. If $g(a) = 0$, then the expression $\dfrac{f(a)}{g(a)}$ is undefined. Because $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{f(a)}{g(a)}$, it follows that if $g(a) = 0$, then $\lim_{x\to a} \dfrac{f(x)}{g(x)}$ does not exist.
- D. The statement is false. If $f(x) = x^2 - 4$ and $g(x) = x + 2$, then $g(2) = 0$ but $\lim_{x\to 2} \dfrac{f(x)}{g(x)} = 0$.

e. If $\lim_{x\to 1^+} \sqrt{f(x)} = \sqrt{\lim_{x\to 1^+} f(x)}$, it follows that $\lim_{x\to 1} \sqrt{f(x)} = \sqrt{\lim_{x\to 1} f(x)}$. Choose the correct answer.

- A. The statement is true. It follows from the limit law $\lim (f(x))^{n/m} = \left(\lim f(x)\right)^{n/m}$ with $a = 1$, $n = 1$, and $m = 2$.

Add instructor tip **Ask the publisher**

Show completed problem | Work problem as student | Student to show work | **Question points:** 1 | Scoring options

Show Answer | Reload | Copy and Edit | ◂Previous | Add | Next ▸

Calculus – MA 123

[redacted]

Preview and Add to Homework

Items in your Homework: 0
Preview Item: 62 of 472 | **Item #:** 2.2.33
Difficulty: Hard
Median time: 2m 44s
Correct on first try: 34.8%

Section 2.2 | **Objective:** Find limits from a graph.
Availability: Homework, Tests and Quizzes, Study Plan
Origin: Publisher

Determine whether the following statements are true and give an explanation or counterexample. Complete parts (a) through (e).

d. $\lim_{x\to 0} \sqrt{x} = 0$ (Hint: Graph $y = \sqrt{x}$). Choose the correct answer below and, if necessary, fill in the answer box to complete your choice.

- A. This statement is false because the value of $\lim_{x\to 0} \sqrt{x} = 0$ does exist but is equal to the nonzero value ☐.
  (Simplify your answer.)
- B. This statement is true because because $\lim_{x\to 0^-} \sqrt{x} = 0$ and $\lim_{x\to 0^+} \sqrt{x} = 0$.
- ★ C. This statement is false because the value of $\lim_{x\to 0^-} \sqrt{x}$ does not exist if x is a real number.
- D. This statement is false because the value of $\lim_{x\to 0^+} \sqrt{x}$ does not exist if x is a real number.

e. $\lim_{x\to \frac{\pi}{2}} \cot x = 0$ (Hint: Graph $y = \cot x$). Choose the correct answer below and, if necessary, fill in the answer box to complete your choice.

- A. This statement is false because the value of $\lim_{x\to \frac{\pi}{2}} \cot x$ does exist but is equal to the nonzero value ☐.
  (Simplify your answer.)
- B. This statement is false because the value of $\lim_{x\to \frac{\pi}{2}^-} \cot x$ does not exist.
- C. This statement is false because the value of $\lim_{x\to \frac{\pi}{2}^+} \cot x$ does not exist.
- ★ D. This statement is true because $\lim_{x\to \frac{\pi}{2}^-} \cot x = 0$ and $\lim_{x\to \frac{\pi}{2}^+} \cot x = 0$.

Add instructor tip **Ask the publisher** Similar question

Show completed problem | Work problem as student | Student to show work | **Question points:** 1 | Scoring options

Show Answer | Reload | Copy and Edit | ◂Previous | Add | Next ▸