# Ve401 Probabilistic Methods in Engineering

## Spring 2021 Term Project 1

**Date Due: 11:59 PM, Friday, the 26th of March 2020**

The goal of this term project is to apply your new-found knowledge of probability theory and statistics in extended tasks that are beyond the scope of ordinary assignments. **It is strongly recommended that you do not leave the entire project to the last minute** but rather commence work on individual parts as soon as you are able to do so.

## Group Work

You will be divided into groups of *4–5 students* each.

> Each group member must be familiar with and have contributed to each part of the project report. **You may not divide up the work in such a way that only certain members are involved with certain parts.** In the event of an Honor Code violation (plagiarism or other), all members of the group will be held equally responsible for the violation. Exceptions may only be made, at my discretion, in exceptional situations.

It is therefore all group members' duty to ensure that all collaborators' contributions are plausibly their own and to check on all collaborators' work progress and verify their contributions within reason.

## Project Report

The term project will be submitted **electronically only** as a typed report of **normally no more than 10 pages** in length. Handwritten submission will not be accepted! It is recommended that you use a professional type-setting program (such as LaTeX) for your report. Unless you are able to ensure a unified font size and style for formulas and text in Microsoft Word, use of Word is *not recommended.*

Your report should have the appearance, style and contents of a professional report. It should be comprehensible without reference to this document and should be comprehensible by any other student in this course. It is strongly suggested that all members of the project team proof-read the report before submission. **The report should not look like the solution to an assignment.** Do not structure the section titles as "Answer to Question i)" or similarly.

## Grading Policy

This term project accounts for 15% of the course grade; it will be scored based on

- **Form (3 points):** Does the report contain essential elements, such as a cover page (with title, date, list of authors), a synopsis (abstract giving the main conclusions of the project), table of contents, clear section headings, introduction, clear division into sections and appendices with informative titles and bibliography (if applicable)? Are the pages numbered? Are the text and formulas composed in a unified font? Are all figures (graphs and images) clearly labeled with identifiable source?
- **Language (3 points):** Is the style of english appropriate for a technical report? Do not treat the project as an assignment and simply number your results like part-exercises. Your text should be a single, coherent whole. The text should be a pleasant read for anyone wanting to find out about the subject matter.

  Errors in grammar and orthography (use a spell-checker!) will be penalized. Make sure that the report is interesting to read. Avoid simply repeating sentences by cut-and-paste.
- **Content (9 points):** Are the mathematical and statistical methods and deductions clearly exhibited and easy to follow? Are the conclusions well-supported by the mathematical analysis? It is important to not just copy calculations from elsewhere, but to fully make them your own, adding details and comments where necessary.

All group members will generally receive the same basic grade for the term project. Exceptions are possible in certain circumstances, such as a group member not contributing to the project. If circumstances permit, peer evaluations of the contribution of each group member will be implemented. These may have a further effect on the project grade.

## On Plagiarism

Study JI's Honor Code carefully. **Any** information from third parties (books, web sites, even conversations) that you use in your project must be accounted for in the bibliography, with a reference in the text. Follow the rules regarding the correct attribution of sources that you have learned in your English course (e.g., Vy100, Vy200). All members of a group are jointly responsible for the correct attribution of all sources in all parts of the project essay, i.e., any plagiarism will be considered a violation of the Honor Code by all group members. Every group member has a duty to confirm the origin of any part of the text.

The following list includes some specific examples of plagiarism:

- Use of any passage of three words or longer from another source without proper attribution. Use of any phrase of three words or more must be enclosed in quotation marks ("example, example, example"). This excludes set phrases (e.g., "and so on", "it follows that") and very precise technical terminology (e.g., "without loss of generality", "reject the null hypothesis") that cannot be paraphrased,
- Use of material from an uncredited source, making very minor changes (like word order or verb tense) to avoid the three-word rule.
- Inclusion of facts, data, ideas or theories originally thought of by someone else, without giving that person (organization, etc.) credit.
- Paraphrasing of ideas or theories without crediting the original thinker.
- Use of images, computer code and other tools and media without appropriate credit to their creator and in accordance with relevant copyright laws.

# Randomness and Pseudorandom Numbers

In many applications, it is desirable to have a pool of random numbers. However, it is very difficult to obtain truly random numbers in sufficient quantities. One important example is that of encryption: every time a website is accessed via the `https` protocol, an encrypted connection is opened using random numbers as a seed. Attempts can be made to obtain random numbers from the physical environment; these include the timestamp of user actions registered by a computer's processor, discarding the most significant digits (e.g., the number of milliseconds in the timestamp of a user clicking on a link), randomness found from atmospheric noise and other, more elaborate methods.

Truly random numbers generated in a verifiable and provably random way are of great importance; a recent project is the *League of Entropy* [1] that uses different physical sources of entropy to generate and publicize random numbers.

However, in practice and for most user-based applications, so-called *pseudo-random* numbers are used. These are not truly random, but rather obtained from some mathematical function in a way that (hopefully) appears random. This gives rise to two problems:

- Which functions might be used to obtain (generate) pseudo-random numbers?
- How to determine if a generator of pseudo-random number is sufficiently "random looking"?

For this project, give an overview of (possibly only some aspects) of these problems and describe and analyze in detail one approach to creating pseudo-random numbers. You may decide to describe one of the following:

- A feasible pseudo-random number generator (PRNG) algorithm;
- A test for randomness of such a PRNG;
- A known flaw in a PRNG;
- Any other interesting aspect in the generation of (pseudo-)random numbers.

You should write somewhere between 5 and 10 pages. Your report should be written in such a way that a typical classmate would be able to follow the contents.

Examples of PNRG algorithms include the linear congruential method [2] or the Mersenne twister [3] (the latter, however, is hard to explain clearly and you need to make an effort to explain it without relying on the reader being familiar with computer science jargon). Of course, there are many others!

Tests for randomness include techniques such as the cumulative sum test or the Wald-Wolfowitz runs test [4]. Again, there are many others, such as those given in [5]. You should explain (and derive, as far as possible) the basis for these tests and the criteria that they use. For example, the cumulative sums test is based on finding a certain limiting probability which was first calculated by Erdös and Kac in [6].

An example of a known flaw is that of the linear congruential generators found by Marsaglia [7].

So feel free to write about some aspect that you find interesting; don't blindly choose one of the examples above (although you may, if you wish).

# References

[1] D. Kozlov. League of entropy: Not all heroes wear capes. `https://blog.cloudflare.com/league-of-entropy/`, June 17, 2019. [Online; accessed 18-June-2019].

[2] Wikipedia contributors. Linear congruential generator — Wikipedia, the free encyclopedia, 2021. `https://en.wikipedia.org/w/index.php?title=Linear_congruential_generator&oldid=1002030177` [Online; accessed 7-March-2021].

[3] Wikipedia contributors. Mersenne twister — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Mersenne_Twister&oldid=1007306517`, 2021. [Online; accessed 7-March-2021].

[4] L. Bassham, III, A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, and S. Vo. *SP 800-22 Rev. 1a. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications.* National Institute of Standards & Technology, Gaithersburg, MD, United States, 2010. `https://csrc.nist.gov/publications/detail/sp/800-22/rev-1a/final`.

[5] Wikipedia contributors. Diehard tests — wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Diehard_tests`, 2018. [Online; accessed 11-February-2018].

[6] P. Erdös and M. Kac. On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.*, 52(4):292–302, 04 1946. `https://projecteuclid.org:443/euclid.bams/1183507847`.

[7] G. Marsaglia. Random numbers fall mainly in the planes. *Proceedings of the National Academy of Sciences of the United States of America*, 61(1):25–28, 1968. `http://www.jstor.org/stable/58853`.