

# A preparation online course for Probability and statistics applications for Modelling, Uncertainty, Data for Engineers

10 january 2023

## Summary

Probability and statistics applications for Modelling, Uncertainty, Data for Engineers (PSAM) is a collection of openly available materials, including examples and short exercises, for learning basic probabilistic concepts related to Civil Engineering and Geosciences (CEG) topics. The course materials have a particular emphasis on CEG real-world applications. This material collection was initially developed as an online course (OC) for MSc students from the faculty of Civil Engineering and Geosciences at TU Delft with a weak background in probability and statistics. The materials for this course have been made publicly available, and are hosted at: <https://tudelft-citg.github.io/learn-probability/>

The course materials are built upon the open-access Jupyter notebook framework, the PSAM notebooks consist of interactive documents that contain executable code, textbook-like explanations, mathematical formulas and plots. The course assumes some prior knowledge of probability and statistics. Although most of the example calculations illustrated in this course use Python, previous experience with this programming language is not necessary. By providing interactive and structured material we hope that the PSAM notebooks make students entangled with probability and statistics CEG applications.

The course materials are developed in the Python programming language and built as Jupyter notebooks. The self-assessment quizzes are built with JupyterQuiz (Shea 2021). We use the Jupyter Book tool (Executable Books Community 2020) to build the materials for online hosting at GitHub.

## Statement of Need

Uncertainty topics (i.e., concepts based on the axioms of probability theory) have been increasingly incorporated in civil engineering and geoscience fields over the

last century; this is supported by the inclusion of uncertainty as one of three key learning lines in the foundational Modelling, Uncertainty, Data for Engineers (MUDE) module in the new curriculum for the master programme at CEG TU Delft, and there are many examples of successful integration of these methods in engineering practice. However, many students still struggle with the basic concepts, implementing some concepts in their course assignments and thesis work, and to communicate about uncertainty topics effectively. Furthermore, despite being taught for decades, the adoption of some probabilistic methods still seems to be slow within some traditional branches of civil engineering practice. Incorporating open-access education resources in the new curriculum will not only provide students with a more motivating and accessible education regarding the uncertainty learning line but also encourage greater integration of these methods in practice through collaboration and accessibility via our future graduates.

Existing material about probability and statistics exists but tends to be very mathematical and is hard for students to apply within CEG applications when asked to self-study. Furthermore, examples with CEG applications tend to be simple (e.g., short arithmetic problems), and it is difficult for students to get a real sense of how these methods are used in practice until much later in the curriculum. The use of probabilistic methods in practice often relies on software that historically is not open access (paid or privately-developed software where source code is not shared), which can be addressed directly by using Python as the computational tool of choice in future educational resources. While there are many open-access Python packages available, open-access education materials that include the use of Python with probabilistic methods in CEG fields are very limited.

## Structure and content

This set of materials was originally created as a background probabilistic material provided as an online course of BSc-course material for students from TU Delft, and a basic manual for CEG MSc students in general with a weak background in probability and statistics. The OC was first deployed in the summer of 2022 and has over 200 students from 35 countries enrolled. The online course is organized in a modular set-up and consists of three sections. Each Section contains a number of modules set up with a consistent 3-unit format. The typical 3-unit format is the following:

1. The basic theory of a specific concept.
2. Example of how a specific concept is applied to a CEG problem, through step-by-step explanations and short python code blocks.
3. Introduction of a new problem, to practice using the method from unit II.

In the first section, we provide basic and general probability concepts such as events and sample space, intersection, union and total probability theorem.

Discrete random variables are presented in section two. Finally, in section three some continuous random variables are discussed. The examples and exercises presented are real civil engineering and geoscience applications such as flood hazards assessment, traffic loads, hurricanes and construction safety. All the source notebooks can be downloaded to be run locally.

## Conclusion

The PSAM course materials serve as a complement to many other resources regarding probability and statistics. Overall, the materials provide a consistent and accessible set of learning materials for uncertainty-related concepts focused on real world probabilistic concepts related to Civil Engineering and Geosciences subjects.

## Acknowledgements

We would like to thank the teaching assistants, course staff, colleagues and the numerous students who have contributed with ideas, time and work to this built the course and provided valuable feedback.

## References

- Executable Books Community. 2020. *Jupyter Book* (version v0.10). Zenodo.  
<https://doi.org/10.5281/zenodo.4539666>.  
Shea, John. 2021. *JupyterQuiz* (version 1.5).