Teaching performance analysis: essential skills and learning outcomes

Dieter Fiems
Ghent University, Department TELIN
St-Pietersnieuwstraat 41
B-9000 Gent, Belgium
Dieter.Fiems@UGent.be

ABSTRACT

In the age of machine learning, traditional performance analysis courses face challenges such as declining student interest and increasing competition from courses within the respective study programmes. At the same time, courses must accommodate increasingly heterogeneous groups of students, both in terms of background, interests and mathematical ability. In this paper, we present a personal perspective on teaching performance evaluation techniques. We argue that stochastic modelling should be the focus in a performance analysis course and that stochastic analysis techniques are a means to an end to solve performance problems, not the main focus.

1. INTRODUCTION

The definition of course content and the development of course materials varies depending on the context. Often there is a specific study programme the course must fit, perhaps there are already be some course notes of an earlier version of the course, one should consider the background of the students, there is an obvious influence of similar courses at other institutions, etc. In addition, the background of the lecturers responsible will undoubtedly influence the selection of topics covered. The combination of context and background leads to diverse course offerings on the same subject. This is particularly true for performance evaluation courses where the lecturers' backgrounds range from applied mathematics and statistics to computer science, engineering, operations research and operations management. The previous editions of TeaPACS provide a nice illustration of this diversity. While some have argued for significantly less time to be spent on theoretical foundations [1, 2] others have called for the introduction of advanced probabilistic concepts [3], albeit with less focus on mathematical rigour. The diversity is further increased by how the course aligns with the rest of the study programme. If there are no preparatory probability courses, a larger share of the course will be devoted to fundamentals. If the course is embedded in an applied degree programme, the course will be specifically focused on the application, etc.

This leads to the observation that there is no one-size-fits-all performance evaluation course. Regardless of the level of mathematical complexity or the particular application domain,"understanding" of systems that evolve randomly is for me the ultimate learning outcome of a performance evaluation course. Understanding means that (i) students recognise that system dynamics depend on the statistical properties of the random processes that drive the evolution of the system; (ii) that they have some means of characterising performance by means of metrics in terms of these properties; and (iii) that they can critically evaluate and iteratively refine their performance evaluation studies. This learning outcome is closely linked to an extended exposure to stochastic modelling techniques, which should lead to the realisation that there is often no single "good" model of a system. The level of detail of the modelling of a system depends very much on the purpose of the model. The search for a good model should be seen as an exercise in gaining understanding how the system evolves over time. Sometimes the models are overly detailed and the details can hide the key determinants of the dynamics of the system. Sometimes the models are oversimplified, so there is a risk that some key determinants are overlooked, leading to qualitative or quantitative errors.

The ideal course context involves inquisitive students with a strong probabilistic background and a deep understanding of the relevant application area. In a more realistic context, probabilistic reasoning and analysis skills (and sometimes programming skills) place considerable constraints on what can be achieved within the confines of a one-semester course with a given number of credits. Domain knowledge may also be limited, depending on how the course is scheduled within the degree programme. At the other end of the spectrum, the "performance evaluation course" is just a small part of another course added as an afterthought. But even then, for many systems you can gain a lot of understanding with garden-variety mathematics, back-of-the-envelope calculations and/or 10-line computer programs.

The remainder of this article contains some personal observations on various topics related to teaching performance. These observations come from teaching queueing analysis and simulation to computer engineering students, simulation in industrial engineering and operations research, and computer-intensive Bayesian statistics in the statistics programme. These observations then lead to some personal recommendations in Section 3.

2. OBSERVATIONS

The course on performance analysis (Queueing Theory and Simulation) at Ghent University combines a course on classical queueing theory with a course on the theory of

TeaPACS 2024 Venice, Italy

stochastic simulation. Typical reference works for this course are Mor Harchol-Balter's "Performance modelling and design of computer systems queueing theory in action" [4] and Sheldon M. Ross' "Simulation" [5]. The following comments come mainly from comparingthe teaching experience for this course with the experience of teaching other courses.

2.1 Probabilistic reasoning

Students often encounter great challenges with probabilistic reasoning. Translating the stochastic problem into a corresponding deterministic problem is a major hurdle. While solving the deterministic problem (e.g. solving sets of balance equations or ordinary differential equations) requires considerable technical skill, there are well-defined rules and methods that must be followed. In contrast, probabilistic reasoning often requires a deeper understanding of less intuitive concepts that are not as easy to grasp as the step-by-step procedures involved in solving systems of equations or differential equations. Therefore, the cognitive leap required for effective probabilistic reasoning can be a significant hurdle. For example, for many students, the biggest obstacle in solving (quasi-)birth-death queueing systems is finding the set of balance equations for a given, textually described queueing problem. The solution itself is the relatively easy part, while the interpretation of the results again poses problems. In particular, questions related to performance measures at different observation times such as arrival or departure times [6].

2.2 Clarity

The lack of clear procedures for analysing stochastic models makes probabilistic reasoning difficult. There are several analysis methods, and for each method there is a certain number of problems that can be efficiently handled by the method. For a given performance evaluation problem, it is not trivial to choose the right method, and different approaches may provide different insights. In a typical performance analysis course, the basic techniques are explored and instructive problems deepen the understanding of the methods. For clarity, problem sets are selected that are not overly technical and do support the basic understanding of the methodology. In a way, this is in contrast to the problem of model selection, where you have to explore the limits of the methods. A typical scenario where the boundaries are explored would include both approximate/exact/numerical approaches to a particular performance problem and then use simulation to explore further details that are difficult to capture by the analytical models. This is fine if simulation is used as a black box tool. If you also want to teach the simulation methodology (or improve the simulation approach with a variance reduction technique), the cognitive load increases significantly.

2.3 What have we learned?

Practically orientated students easily get lost in a forest of equations, even if the equations are only a means to an end. If we lose students along the way, spending a lot of time and effort explaining the end result intuitively will not bring them back. The same goes for illustrating the results by solving practical performance problems. On the other hand, introducing a lot of applied detail at the beginning can obscure the generality of the methodological approach. There is no simple solution to the mathematical difficulties, but we can alleviate the problems somewhat. We can start with the type of performance questions we want to tackle, somewhat in line with the 5 basic questions presented by Mor Harchol-Balter and Ziv Scully at the first TeaPACS workshop [7]. These questions can be quite basic at the beginning of the course and later orientated more towards problems that pique students' interest. Some basic questions are for example the following:

- Why can't we fully load our system?
- Do stochastic assumptions like distributions matter?
- Does it matter when I make measurements?
- How can I improve service?
- Does my solution scale well when I have more customers, servers, etc?

To answer these questions, we can look for the simplest model possible that can answer the question. Basically, this is not very different from teaching tools and introducing examples. However, the question up front makes it clear why the mathematics are needed, and is now positioned earlier in the presentation. Note that a question-based approach does not require the course material to be restructured. The same result is achieved by looking for questions that the specific (existing) material can answer.

2.4 Motivation

While motivation, or lack thereof, is fairly consistent across programmes over the years, I have noticed a significant difference in the motivation of students on these programmes. This is closely related to the perception of the course within each programme. Simulation is one of the main topics in the Industrial Engineering and Operations Research programme. Although the preparation of the mostly international students varies greatly, there is a consensus that this is an important topic that is worth the effort. Simulation is also part of the toolkit of professionals working in operations research, which strengthens the position of the course. To cater for different backgrounds, I provide material on the basics of probability theory, with an emphasis on solving practical problems to help students who do not have the necessary background knowledge. The material deals with issues such as calculating the distribution of a function of a random variable, working with expectations and variances, etc. Although a large percentage of students have difficulty understanding the various topics, the content of the course is never questioned. In the computer engineering programme, performance analysis is a mandatory mathematics-oriented course, which keeps them away from "fun" courses such as machine learning, blockchains or cybersecurity. Although students have already had a basic course and applied probability course before starting the performance course, too many students are unable or unwilling to make the effort to solve simple probabilistic problems, even though similar problems were solved in the applied probability course. Finally, the students in the statistics programme are amazing. They start without much mathematical skill and knowledge (historians, psychologists, biologists, etc.) and gain an understanding of Bayesian statistics, basic probability theory, simulation of Markov chains (including variance reduction), and Markov chain Monte Carlo and particle filtering, all in a 6 credit course. The course is perceived to be strenuous, but it is worth the effort and students feel like fully-fledged statisticians after completing the course. These students are usually more mature, having obtained an earlier Master's degree and spread their statistics studies over two years, in combination with a full-time or part-time job. As a rule, they lack time, but not motivation.

2.5 Applications

There are many good arguments in favour of an application-oriented approach. Following Giuliano Casale [1]'s suggestion to create a new course on performance evaluation, choose a popular application area and build the course around this topic. This could partially counteract the declining popularity of performance evaluation courses and facilitate positioning within degree programmes. In particular, programme committees can more easily integrate an application-based course into the degree programme as a mandatory course or preferred elective. However, there are also important arguments against such an approach. First, a performance evaluation course on a current hot or entertaining topic cannot compete with a course that deals exclusively with the same current topic. This is probably the most important counter-argument. Second, what is a current topic today may not be tomorrow. This means not only a lack of continuity, but also the need to constantly adjust the trade-off between the intended (performance evaluation) learning objectives and the hot application domain. The compromise may also lead to a loss of diversity in the topics that can be covered in the course. Finally, students who are interested in performance evaluation but not in the chosen application will no longer have the opportunity to study performance evaluation. Which of the (counter)arguments is decisive depends heavily on the context, including the presence of competing courses and the possibility of integrating the course into different degree programmes. Finally, note that it is not only possible to convert a performance evaluation course into a hot topic course. It is equally possible to complement a hot topic course with a performance evaluation module, a proposal explored by Cristina L. Abad in this edition of TeaPACS [8].

3. RECOMMENDATIONS

In the previous section, I made the case that the content of a performance evaluation course depends very much on the context. I therefore reluctantly formulate some general recommendations for a performance evaluation course that focuses mainly on queueing theory and the theory of stochastic simulation.

1. Clarity requires simplicity, but should not come at the expense of versatility. While our original course material included a discussion of the M/G/1 queue, we no longer include it and limit ourselves to Markovian queues. What do you lose by restricting analytical methods to methods for countable state-space Markov chains? Distributional effects can be studied by means of phase-type distributions, Markov arrival processes allow for burstiness. Various limiting results and approximations can be used (fluid and diffusion limits, mean field limits), albeit somewhat less rigorously. We also add transform methods that greatly simplify some calculations, as well as the discussion of some limit results.

2. Simulation methodology and analytical methods do not go well together. The mathematical arguments used to study and improve simulation experiments are quite different from the analytical arguments of queueing theory. Nevertheless, it is quite convenient to have some simulation methodology at one's disposal either to explore modelling options or evaluate proposed approximations. A clear separation can be achieved by first introducing simulation as a black box approach in the analytical part. A parallel simulation track can then deal with the simulation methodology and refer to analytical results as theorems if required.

3. It might be worthwhile, interesting, and/or necessary to expand the scope of the traditional "performance evaluation" course to at least retain the essential insights of a typical such course offering. Stochastic modelling and its methods are not limited to the assessment of congestion in various computer, communication or physical systems, but also apply, for example, to biological, ecological and financial processes, to name but a few. For example, "sustainability" could be an interesting, contemporary "fun" topic for a stochastic modelling course.

Finally, I find beauty in simplicity in both course design and stochastic modelling. Or, in the words of Antoine de Saint-Exupéry: "La perfection est atteinte, non pas lorsqu'il n'y a plus rien à ajouter, mais lorsqu'il n'y a plus rien à retirer" (Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away). This is an often used quote from Terre des Hommes (Wind, Sand and Stars). The quote refers to the design of aeroplanes, but has also been applied to design in general. A little later in the book, the protagonist also says "Il est également admirable que, dans son usage même, la machine peu à peu se fasse oublier" (It is also admirable that, in its very use, the machine gradually makes itself forgotten), a quote that fits just as well. Our machinery is the various tools and models for performance evaluation and our flight is the exploration and discovery of how a system under investigation performs in reality.

4. REFERENCES

[1] G. Casale. Performance evaluation teaching in the age of cloud computing. Proceedings of the Second International Workshop on Teaching Performance Analysis of Computer Systems (TeaPACS), 2023.

[2] G. Serrazzi. Updating the Content of Performance Analysis Textbooks. Proceedings of the First International Workshop on Teaching Performance Analysis of Computer Systems (TeaPACS), 2021.

[3] Z. Scully. The Role of Advanced Math in Teaching Performance Modeling. Proceedings of the Second International Workshop on Teaching Performance Analysis of Computer Systems (TeaPACS), 2023.

[4] M. Harchol-Balter. Performance Modeling and Design of Computer Systems: Queueing Theory in Action. Cambridge University Press, 2013.

[5] S. M. Ross. Simulation, 6th edition. Academic Press, 2022.

[6] J.-Y. Le Boudec. Performance Evaluation: A Preparation for Statistics and Data Science? Proceedings of the First International Workshop on Teaching Performance Analysis of Computer Systems (TeaPACS), 2021.
[7] M. Harchol-Balter, Z. Scully. The most common queueing theory questions asked by computer systems practitioners. Proceedings of the First International Workshop on Teaching Performance Analysis of Computer Systems (TeaPACS), 2021.
[8] C. L. Abad. How can we teach workload modeling in CS systems classes? Proceedings of the Third International Workshop on Teaching Performance Analysis of Computer Systems (TeaPACS), 2024.