40.305 Advanced topics in Stochastic Modeling Nov 13, 2024

Mini-project exercise: Choice I (a simulation exercise)

Due date: Dec 13, 11:59 pm

Note.

1) You are required to do only ONE of the three project choices provided. This pdf contains only a description of Project choice I, which is a concretely defined simulation exercise. Please see Choice II if you prefer a theoretical exercise (or) Choice III if you prefer to do a more open-ended simulation exercise of your choice.
2) Our purpose in this mini-project exercise is to help you perfect the skill in identifying regenerative structure and using it for output statistical analysis for managerial decision-making.

Choice I: Use of regenerative simulation method for staffing call centers

Motivation. The recent decades have witnessed an explosive growth in the number of companies that provide services via the telephone. According to some estimates, world-wide expenditure on call centers exceeds $300 billion. A central challenge in designing and managing a service operation in general, and a call center in particular, is to achieve a desired balance between operational efficiency and service quality.

Here we consider the staffing aspects of this problem, namely, having the right number of agents in place. "The right number" means, first of all, not too many, thus avoiding overstaffing. That is a crucial consideration because personnel costs (e.g., salaries of operators and spending on training) typically constitute about 70% of a call center's expenditure. "The right number," however, also means not too few, thus avoiding understaffing and consequent poor service quality. Indeed, understaffing would imply excessive customers' wait in tele-queues which is unpleasant in itself and, moreover, is likely to lead to abandonment of frustrated customers. (According to a Purdue University study, 63% of the customers name a negative call center experience as their main reason for stopping transactions with a company.) In this context, it is important to consider a model which takes customer abandonment into account.

A call-center model incorporating customer abandonment. Consider a moderate-size call center with an average of 20 arrivals per minute over a given time interval. Let the average service time per call be three minutes ($\mu = 1/3$). The arrivals are taken to be according to a Poisson process and the service times are i.i.d. exponentially distributed.

If an inbound call finds that all agents are busy, the call is placed to wait in a queue and is served on a first-come-first-serve (FCFS) basis. Suppose that each inbound caller has a generally distributed patience time T. If the caller ends up waiting in the queue more than T units of time, he/she simply abandons the call (thus leaving the queue); otherwise the caller gets served.

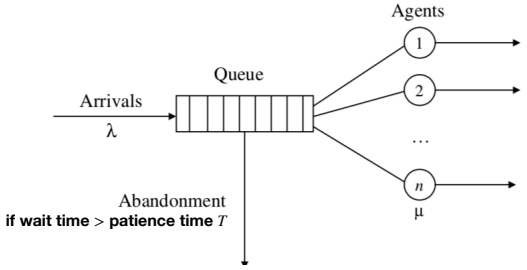

Figure 1: Call-center model

On meeting target quality of service. We consider the staffing problem subject to one of the following two target performance constraints: Suppose that the call center would like to maintain a high level of service. Two possible *steady-state* performance constraints to measure this quality of service are,

i) Probability to abandon should be less than 2%;
ii) Average wait should be less than five seconds.

For the given arrival and service rates, it is advisable to choose more than $\lambda/\mu = 60$ agents. However, it is not clear how many agents are precisely needed to meet target performance constraint of our choice.

Description of your tasks in staffing the call center. Our goal is to use regenerative method of simulation output analysis to identify the optimal staffing level (minimum number of call center agents) which meets a given target performance constraint, *with 95% confidence*, in the following situations:

a) Suppose that the customers of this call center have i.i.d. patience time T distributed uniformly over the interval $[0, 6]$ minutes. Find the minimum staffing levels that are suitable for meeting the target performance constraints i) and ii) above, with 95% confidence.
b) Suppose that the patience time T has the following hyperexponential distribution: T is Exponential with mean = 1 minute with probability 1/2 and T is an Exponential RV with mean = 5 minutes with probability 1/2 (In other words, the customers of this call center constitute a 50% - 50% mixture of impatient and relatively more patient customers: the impatient customers have patience time which is exponential with mean = 1 minutes and the relatively more patient customers have patience time which is exponential with mean = 5 minutes) Find the minimum staffing levels that are suitable for separately meeting the target performance constraints i) and ii) above, with 95% confidence.
c) How does the patience distribution in parts a) and b) affect the optimal staffing level?

Some guidelines

1) You must make use of regenerative method for performing statistical analysis of simulation outputs. We covered this topic in Week 9, Class 2.
2) This exercise is to be done in groups of size upto 2. If you wish to do it in a group of 3, it is recommended that you build on the proposed setup to make the workload proportional for 3 people.
3) It is sufficient to submit (i) your code and (ii) either a concise report (or) a markdown which clearly explains everything specific to your setup, the choice of regeneration times you have identified, and the results containing snapshot of your simulation estimates. In case you are submitting a markdown instead of a report, please make sure the markdown is also readable (say via html, or pdf) without having to suffer from compatibility issues.
4) Computer simulations can be carried out, for example, in Matlab, R, Python or a similar platform.
5) Please feel free to approach me with questions if you encounter difficulties.