Call center staffing

Introduction

You are in charge of the management of the call center of a credit card company. The call center receives various types of calls in different languages, and needs a number of employees capable of treating certain types of calls only in a given language.

The decisions that you have to make are:

- The number of operators to hire of each seniority level and language, and how to construct their shifts
- What call response strategy to employ.

The aim of the "Simulation Project" is to develop a discrete event simulation that represents the system and to evaluate the performance of two customer service designs.

During the "Optimization Project", the discrete event simulation is expanded, and the optimal solution in terms of service level, operation cost and shift structure is identified by an optimization algorithm.

Develop the discrete event simulation with a modular structure. It should be possible to modify the various components, such as shift start time, language and seniority level, during the "Optimization Project".

Problem

We consider the call center of a credit card company. It operates 24 hours a day and calls arrive following a non-homogenous poison process $\lambda(t)$ in terms of calls per hour. The arrival rates can be grouped into three homogeneous segments: (1) 20:00-9:00 and 13:00-14:00 with $\lambda(t) = 200$; (2) 9:00-13:00 and 18:00-20:00 with $\lambda(t) = 400$ and (3) 14:00-18:00 with $\lambda(t) = 300$. Each customer requires service in a specific language. 35% of them speaks French, 45% German and 20% Italian. The company also classifies the type of call as (1) a general inquiry (2) an accounting inquiry and (3) a lost-stolen report. Out of all received calls, on average 50% are general inquiries, 35% are accounting inquiries and 15% are lost-stolen reports. The distributions of the language and the type of call are independent. For each arrival, the language and the type of call are drawn according to the percentage specification above. The call duration follows a log-normal distribution with $\mu = 1.3$ (location) and $\sigma = 0.5$ (scale) which is independent of the type of call, language or agent. However, the duration of calls is limited to 20 min.

To maintain a service level, the company aims to answer all calls within a waiting time of 10 minutes. Afterwards, the company considers a monetary value of 1 CHF per minute for those who wait between 10 and 15 minutes, which is increased to 5 CHF per minute after 15 minutes. For example, if a customer waits for 17 minutes before being served, the company considers a penalty of $5 \times 1 + 2 \times 5 = 15$ CHF. If a call is routed between several agents, the total waiting time is considered for penalty.

The company wants to staff the call center with the following types of agents:

1. *Supervisors.* They can serve all types of calls and receive 250 CHF per day.
2. *Seniors.* They can serve general inquiries and accounting inquiries. They receive 180 CHF per day.

3. *Juniors.* They are only able to serve general inquiries. They receive 120 CHF per day.

According to company rules, employees work 8-hour shifts. In addition, an agent can only answer calls in one language, i.e. French, German or Italian.

Simulation

For the simulation project, you are requested to:

- Develop a discrete event simulation to represent the described project.
- Define the metrics used to quantify the quality of the service
    - Remember that extreme cases are important; evaluate other metrics in addition to the mean. Moreover, the waiting time for different call types could have different impact on how your service is perceived.
    - Report the mean square error of your estimation using bootstrapping when necessary.
    - Use variance reduction techniques to reduce the computational time.
- Assume that the call center operates three shifts: 8:00-16:00, 16:00-24:00 and 24:00-8:00. The call center should be staffed in the cheapest way that makes it possible, on average, to respond to the maximum arrival rate in each shift interval. For example, the maximum arrival rate for the shift 8:00-16:00 is 400. Multiplying this by the average call duration and the expected percentages for language and type of call, we can obtain the average agent time necessary to serve each call type for each language in one hour. Find the number of supervisors, senior and junior staff that can respond to this demand in the cheapest way. Remember that each agent can only work in one language. Simulate the following two scenarios:
    - Calls enter a primary queue and are assigned to the first available agent on a first-come, first-served (FCFS) basis. If the agent cannot handle the call due to language or seniority level mismatch, he/she reroutes the call to the secondary queue for the suitable pair of language and call type, which is also served on a FCFS basis by the suitable agents. The secondary queues take priority over the primary queue. That is, when an agent becomes available, he/she is assigned a call from the primary queue only if the secondary queue that pertains to him/her is empty.
    - Calls are first answered by an automatic machine and sorted by type of call and language. A queue is formed for each pair of language and call type, and handled on a FCFS basis by the suitable agents. Here, customers may make a mistake. On average, 20% of the time they do not select the correct call type. When this happens, the responding agent reroutes the call to the suitable pair of language and call type, where it joins at the end of the queue.
- Make any necessary assumptions.