CS exam

Alireza Ghahramani Kure

January 11, 2023

1. **Answer the following questions according to the topics taught.**

   (a) **What is a statistical model in simulation?**
   A statistical model in simulation is a mathematical representation of a real-world system that incorporates probability and randomness. It can be used to generate simulated outcomes that mimic the behavior of the real-world system, and to make predictions or estimate performance metrics based on those outcomes.

   (b) **What are the different types of statistical models used in simulation?**
   There are several different types of statistical models that can be used in simulation, including:

   - Discrete-event models, which simulate the flow of events in a system over time, and can be used to model systems such as manufacturing plants or transportation networks.
   - Agent-based models, which simulate the behavior of individual agents within a system and the interactions between them.
   - Stochastic process models, which simulate the evolution of a system over time, and can be used to model systems such as financial markets or biological populations.

   (c) **What is a queueing model?**
   A queueing model is a mathematical representation of a system that contains one or more waiting lines (queues) of customers or entities, and servers or resources that process or service these entities. The queueing model can be used to analyze the performance of the system, such as the number of customers in the queue, the waiting time for customers, or the utilization of the servers.

   (d) **What are the different types of queueing models?**
   There are several types of queueing models, including:

   - Single-server models, which simulate a system with one queue and one server.
   - Multi-server models, which simulate a system with one or more queues and multiple servers.
   - Multi-channel models, which simulate a system with multiple queues and multiple servers.

   (e) **What is the role of random number generation in simulation?**
   Random number generation plays a critical role in simulation because it is used to introduce randomness and uncertainty into the simulated outcomes. This is important because many real-world systems are inherently uncertain, and generating random outcomes allows us to model that uncertainty and make predictions about how the system will behave in different scenarios.

   (f) **What is the difference between pseudorandom number generation and true random number generation?**
   Pseudorandom number generation is a method of generating a sequence of numbers that appear random, but are actually determined by a mathematical algorithm. These numbers are generated using a pseudorandom number generator (PRNG), which starts with an initial value (called a seed) and generates a sequence of numbers based on a mathematical formula.
   True random number generation, on the other hand, relies on physical processes to generate numbers that are truly random. Example of such processes: radioactive decay, electronic noise. These numbers are typically generated using a hardware random number generator (HRNG).

(g) **What is the importance of a good random number generator for simulation?**
A good random number generator is important for simulation because it is used to introduce randomness and uncertainty into the simulated outcomes. If the random number generator is not of good quality, the simulated outcomes may not accurately reflect the real-world system, which can lead to incorrect conclusions or predictions.

(h) **What is the importance of seed value in Random Number Generation?**
The seed value is the initial value used to start a pseudorandom number generator. It is important because it determines the sequence of numbers that will be generated. Different seed values will result in different sequences of numbers, and using the same seed value will result in the same sequence of numbers.

(i) **Can you explain the concept of ”Random Variate Generation” in simulation?**
Random variate generation, also called inverse transform sampling, is a method of generating random numbers that are distributed according to a specific probability distribution. The method involves generating a uniformly distributed random number between 0 and 1, and then using the inverse cumulative distribution function (CDF) of the desired probability distribution to convert this number into a random variate (a value from the desired distribution).

(j) **How does one generate random numbers from a non-uniform probability distribution?**
There are several methods to generate random numbers from non-uniform probability distribution, such as:

- Inverse transform sampling, as described above.
- Rejection sampling, where random numbers are generated and then accepted or rejected based on whether they fall within the desired distribution.
- Composition method, where the distribution is generated as a combination of simpler distributions.
- Acceptance-Rejection method. using accept-reject method. where numbers are sampled from known distribution and then rejected or accepted based on the comparison of the unknown and known distribution.

These are only a few examples of methods for generating random numbers from non-uniform distributions. The choice of method will depend on the specific distribution and the desired properties of the generated random numbers.

2. **Suppose we have a linear congruential generator (LCG) with the following parameters:**

   - $a = 5, c = 3, m = 8$, and a seed value of $x_0 = 4$

   Calculate the first five random numbers generated by the LCG.
   A linear congruential generator (LCG) is a type of pseudorandom number generator that generates a sequence of numbers according to the following formula:

$$x_i = ax_i + c \mod m \tag{1}$$

   Given the parameters of the LCG provided, we can calculate the first five random numbers as follows:

$$\begin{aligned}
x_1 &= (5 * 4 + 3) \mod 8 = 7 \\
x_2 &= (5 * 7 + 3) \mod 8 = 6 \\
x_3 &= (5 * 6 + 3) \mod 8 = 1 \\
x_4 &= (5 * 1 + 3) \mod 8 = 0 \\
x_5 &= (0 * 0 + 3) \mod 8 = 3
\end{aligned} \tag{2}$$

So the first five random numbers generated by the LCG with the given parameters and seed value of $x_0 = 4$ are: $(7, 6, 1, 0, 3)$
Note that this is a simple example and in practical applications the LCG generators are not recommended to use, as they have poor statistical properties. There are other generators that have superior properties and can be used for simulation such as Mersenne Twister.

3. **A random number generator generates uniformly distributed random numbers between 0 and 1. Suppose we want to generate random numbers from a standard normal distribution (a normal distribution with mean 0 and standard deviation 1). How would we do this using the inverse transform sampling method?**
Inverse transform sampling method can be used to generate random numbers from a non-uniform distribution by using the inverse cumulative distribution function (CDF) of the desired distribution. To generate random numbers from a standard normal distribution, we use the inverse CDF of the standard normal distribution, which is given by the inverse of the cumulative distribution function of the standard normal distribution (also known as the standard normal probability integral or the error function) denoted as Φ^{-1}.
so for any random variable X with standard normal distribution, we can generate random variable Y uniformly distributed between 0 and 1, then $X = \Phi^{-1}(Y)$. so we can generate random numbers from a standard normal distribution by generating uniformly distributed random numbers between 0 and 1 and applying the inverse standard normal distribution function to each number. This can be done using a standard library function or a table look-up of pre-calculated values of the inverse standard normal distribution function.
Note that this method is not applicable to all the distributions, some distributions it is difficult to find the inverse of the cumulative distribution function(CDF) analytically and that's why other methods are used, such as Acceptance-Rejection method, composition method, etc.

4. **A simulation model of a manufacturing process generates 10,000 random samples of the time it takes to complete a unit of production. The sample mean is found to be 5.2 minutes with a standard deviation of 1.5 minutes. Using this information, what is the probability that the time to complete a unit of production will be less than 4 minutes?**
The simulation has generated a random sample of 10,000 observations of the time it takes to complete a unit of production, and we know the sample mean and standard deviation. Based on these data, we can assume that the underlying population of completion times follows a normal distribution, and we can use the mean and standard deviation from the sample to estimate the population mean and standard deviation.
Given a normal distribution with mean of 5.2 minutes and a standard deviation of 1.5 minutes we can use the cumulative distribution function of normal distribution (or Z-score) to find the probability that the time to complete a unit of production is less than 4 minutes. We calculate the Z-score (which is the number of standard deviations from the mean a data point is), using the following formula:

$$Z = (X - \mu)/\sigma \tag{3}$$

Where X is the value of interest (4 minutes), μ is the mean (5.2 minutes), and σ is the standard deviation (1.5 minutes).
Substituting the values we get, $Z = (4 - 5.2)/1.5 = -0.8$. Now we can use standard normal table or a calculator to get the probability of X being less than 4 minutes, that is $P(X < 4) = P(Z = 0.8) = 0.21$. So the probability of completing a unit of production in less than 4 minutes is 0.21, or 21%.

5. **A simulation of a computer system shows that the time between system failures follows an exponential distribution with a mean of 200 hours. What is the probability that a system failure will occur within the next 100 hours?**
The simulation data shows that the time between system failures follows an exponential distribution. The exponential distribution with parameter λ has the probability density function

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0 \tag{4}$$

and 0 otherwise. where λ is the rate parameter and it's the reciprocal of the mean. Given that the mean of the system failure is 200 hours, the rate parameter would be 1/200 = 0.005.
To find the probability that a system failure will occur within the next 100 hours, we need to calculate the cumulative distribution function (CDF) of the exponential distribution for the interval [0,100]. The CDF of the exponential distribution is given by $1 - e^{-\lambda x}$, where x is the time interval and λ is the rate parameter. So the probability of a system failure occurring within the next 100 hours is:

$$P(X < 100) = 1 - e^{-0.005*100} = 1 - e^{-0.5} = 1 - 0.6065 = 0.3935 \tag{5}$$

So the probability that a system failure will occur within the next 100 hours is 39.35%.

6. **A single-server queueing system has an average arrival rate of 6 customers per minute and an average service rate of 8 customers per minute. Using the Little's law, what is the average number of customers in the queue?**
The Little's law states that the average number of customers in a queueing system (L) is equal to the product of the average arrival rate λ and the average time spent in the system (W), $L = \lambda * W$. In this case we are given the average arrival rate (6 customers per minute) and service rate (8 customers per minute) and we are looking to find the average number of customers in the queue.
The average time spent in the system can be calculated as the ratio of the average number of customers in the system to the average departure rate. The average number of customers in the system is the sum of the average number of customers in the queue and the average number of customers being served, which is $L_q + L_s = L$. The average departure rate is the reciprocal of the average service rate (1/8 customers per minute). so we can calculate the average time spent in the system (W) as $W = L/(1/8)$.
now we can use the little's law to calculate the average number of customers in the queue

$$L_q = \lambda * W = 6 * (L/(1/8)) \tag{6}$$

Given that we don't know the value of L and we're trying to calculate it, it's a bit of a circular relationship. However, we can assume that there are no customers in the system, if that's the case $L_q = 0$ and we can get the value of L. Therefore, the average number of customers in the queue is Lq = 6 * (0 / (1/8)) = 0. This is the expected result if the system is operating in an equilibrium state, with the arrival and service rates balancing each other out, meaning that the number of customers waiting in the queue is 0 on average.

7. **A call center receives calls according to a Poisson process with an average arrival rate of 100 calls per hour. Each call is answered by one of the four call center agents, and service times are exponentially distributed with an average service rate of 4 minutes per call. The call center operates for 8 hours a day, and customers are placed on hold if all agents are busy. The time spent on hold follows an exponential distribution with an average rate of 2 minutes per call. What is the probability that a customer will have to wait for more than 5 minutes to speak with an agent?** In this question, we have a multi-server queueing system with customers arriving according to a Poisson

process with an average arrival rate of 100 calls per hour, and service times that are exponentially distributed with an average of 4 minutes per call. Customers are placed on hold if all agents are busy and the time spent on hold follows an exponential distribution with an average rate of 2 minutes per call. To find the probability that a customer will have to wait for more than 5 minutes to speak with an agent, we will use the Engset formula.

The Engset formula is used to calculate the probability that a customer will be blocked (or turned away) in a multi-server system with multiple sources of traffic arriving and service times are exponentially distributed. The Engset formula for the blocking probability is given by $Pb = 1 - (1 - s)^c$ where c is the number of sources of traffic, and s is the server utilization.

To calculate the server utilization we can use the formula: $s = \lambda/(c * \lambda)$ where λ is the arrival rate, μ is the service rate and c is the number of servers.

In this case, $\lambda = 100$ calls per hour, $\mu = 4$ minutes per call, and $c = 4$ servers, so the server utilization is $s = 100/(4 * 4) = 0.75$.

Then we can calculate the probability that a customer will be blocked as $Pb = 1-(1-0.75)^4 = 0.3164$. So the probability that a customer will be blocked, that is, unable to speak to an agent is 31.64%. To find the probability that a customer will have to wait more than 5 minutes, we need to use the cumulative distribution function of the exponential distribution of the holding time which is given by:

$$P(X > 5) = 1 - P(X < 5) = 1 - e^{-\lambda * 5} \tag{7}$$

Given that the average holding time is 2 minutes, the rate parameter λ is 1/2 = 0.5. So we have $P(X > 5) = 1 - e^{-0.5*5} = 1 - e^{-2.5} = 0.0668$. So the probability that a customer will have to wait more than 5 minutes to speak with an agent is 6.68%