STAT253/317 Lecture 18

Cong Ma

Section 7.7 The Inspection Paradox
Chapter 8 Queueing Models

Section 7.7 The Inspection Paradox

Given a renewal process $\{N(t), t \geq 0\}$ with interarrival times $\{X_i, i \geq 1\}$, the length of the current cycle,

$$X_{N(t)+1} = S_{N(t)+1} - S_{N(t)}$$

tend to be *longer* than X_i, the length of an ordinary cycle.

Precisely speaking, $X_{N(t)+1}$ is stochastically greater than X_i, which means

$$\mathrm{P}(X_{N(t)+1} > x) \geq \mathrm{P}(X_i > x), \quad \text{for all } x \geq 0.$$

Heuristic Explanation of the Inspection Paradox

Suppose we pick a time t uniformly in the range $[0, T]$, and then select the cycle that contains t.

- Possible cycles that can be selected: $X_1, X_2, \ldots, X_{N(T)+1}$
- These cycles are not equally likely to be selected.
  The longer the cycle, the greater the chance.

$$\mathrm{P}(X_i \text{ is selected}) = X_i/T, \quad \text{for } 1 \le i \le N(T)$$

- So the expected length of the selected cycle $X_{N(t)+1}$ is roughly

$$\sum_{i=1}^{N(T)} X_i \times \frac{X_i}{T} = \frac{\sum_{i=1}^{N(T)} X_i^2}{T} \to \frac{\mathbb{E}[X_i^2]}{\mathbb{E}[X_i]} \ge \mathbb{E}[X_i] \quad \text{as } T \to \infty.$$

- Last time we have shown that if F is non-lattice,

$$\lim_{t\to\infty} \mathbb{E}[Y(t)] = \lim_{t\to\infty} \mathbb{E}[A(t)] = \frac{\mathbb{E}[X_i^2]}{2\mathbb{E}[X_i]},$$

Since $X_{N(t)+1} = A(t) + Y(t)$, $\lim_{t\to\infty} \mathbb{E}[X_{N(t)+1}] = \dfrac{\mathbb{E}[X_i^2]}{\mathbb{E}[X_i]}$

Example: Waiting Time for Buses

- Passengers arrive at a bus station at Poisson rate λ
- Buses arrive one after another according to a renewal process with interarrival times X_i, $i \geq 1$, independent of the arrival of customers.
- If X_i is deterministic, always equals 10 mins, then on average passengers has to wait 5 mins
- If X_i is random with mean 10 min, then a passenger arrives at time t has to wait $Y(t)$ minutes. Here $Y(t)$ is the residual life of the bus arrival process. We know that

$$\mathbb{E}[Y(t)] \to \frac{\mathbb{E}[X_i^2]}{2\mathbb{E}[X_i]} \geq \frac{\mathbb{E}[X_i]}{2} = 5 \text{ min.}$$

Passengers on average have to weight more than half the mean length of interarrival times of buses.

Class Size in U of Chicago

University of Chicago is known for its small class size, but a majority of students feel most classes they enroll are big. Suppose U of Chicago have five classes of size

$$10, 10, 10, 10, 100$$

respectively.

- Mean size of the 5 classes: $(10 + 10 + 10 + 10 + 100)/5 = 28$.
- From students' point of view, only the 40 students in the first four classes feel they are in a small class, the 100 students in the big class feel they are in a large class.
  Average class size students feel

$$\frac{\overbrace{10 + \cdots + 10}^{40 \text{ students}} + \overbrace{100 + \ldots + 100}^{100 \text{ students}}}{140} = \frac{10 \times 40 + 100 \times 100}{140} \approx 74.3.$$

Proof of the Inspection Paradox

For $s > x$,

$$P(X_{N(t)+1} > x | S_{N(t)} = t - s) = 1 \geq P(X_i > x)$$

For $s < x$,

$$\begin{aligned}
& P(X_{N(t)+1} > x | S_{N(t)} = t - s) \\
&= P(X_1 > x | X_1 > s) \\
&= \frac{P(X_1 > x,\, X_1 > s)}{P(X_1 > s)} \\
&= \frac{P(X_1 > x)}{P(X_1 > s)} \\
&\geq P(X_1 > x)
\end{aligned}$$

Thus $P(X_{N(t)+1} > x | S_{N(t)} = t - s) \geq P(X_i > x)$ for all $N(t)$ and $S_{N(t)}$. The claim is validated

Limiting Distribution of $X_{N(t)+1}$

If the distribution F of the interarrival times is non-lattice, we can use an alternating renewal process argument to determine

$$G(x) = \lim_{t\to\infty} \mathrm{P}(X_{N(t)+1} \leq x).$$

We say the renewal process is ON at time t iff $X_{N(t)+1} \leq x$, and OFF otherwise. Thus in the ith cycle,

$$\text{the length of ON time is } \begin{cases} X_i & \text{if } X_i \leq x, \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

and hence

$$\begin{aligned} G(x) = \lim_{t\to\infty} \mathrm{P}(X_{N(t)+1} \leq x) &= \frac{\mathbb{E}[\text{On time in a cycle}]}{\mathbb{E}[\text{cycle time}]} \\ &= \frac{\mathbb{E}[X_i \mathbf{1}_{\{X_i \leq x\}}]}{\mathbb{E}[X_i]} = \frac{\int_0^x z f(z) dz}{\mu} \end{aligned}$$

Chapter 8 Queueing Models

A queueing model consists "customers" arriving to receive some service and then depart. The mechanisms involved are

- input mechanism: the arrival pattern of customers in time
- queueing mechanism: the number of servers, order of the service
- service mechanism: the time to serve one or a batch of customers

We consider queueing models that follow the most common rule of service: first come, first served.

Common Queueing Processes

It is often reasonable to assume

- the interarrival times of customers are i.i.d. (the arrival of customers follows a renewal process),
- the service times for customers are i.i.d. and are independent of the arrival of customers.

Notation: M = memoryless, or Markov, G = General

- $M/M/1$: Poisson arrival, service time $\sim Exp(\mu)$, 1 server $=$ a birth and death process with birth rates $\lambda_j \equiv \lambda$, and death rates $\mu_j \equiv \mu$
- $M/M/\infty$: Poisson arrival, service time $\sim Exp(\mu)$, ∞ servers $=$ a birth and death process with birth rates $\lambda_j \equiv \lambda$, and death rates $\mu_j \equiv j\mu$
- $M/M/k$: Poisson arrival, service time $\sim Exp(\mu)$, k servers $=$ a birth and death process with birth rates $\lambda_j \equiv \lambda$, and death rates $\mu_j \equiv \min(j, k)\mu$

Common Queueing Processes (Cont'd)

- $M/G/1$: Poisson arrival, General service time $\sim G$, 1 server
- $M/G/\infty$: Poisson arrival, General service time $\sim G$, ∞ server
- $M/G/k$: Poisson arrival, General service time $\sim G$, k server
- $G/M/1$: General interarrival time, service time $\sim Exp(\mu)$, 1 server
- $G/G/k$: General interarrival time $\sim F$, General service time $\sim G$, k servers
- $\dots$

Quantities of Interest for Queueing Models

Let

$$X(t) = \text{number of customers in the system at time } t$$
$$Q(t) = \text{number of customers waitng in queue at time } t$$

Assume that $\{X(t), t \geq 0\}$ and $\{Q(t), t \geq 0\}$ has a stationary distribution.

- $L =$ the average number of customers in the system

$$L = \lim_{t\to\infty} \frac{\int_0^t X(t)dt}{t};$$

- $L_Q =$ the average number of customers waiting in queue (not being served);

$$L_Q = \lim_{t\to\infty} \frac{\int_0^t Q(t)dt}{t};$$

- $W =$ the average amount of time, including the time waiting in queue and service time, a customer spends in the system;
- $W_Q =$ the average amount of time a customer spends waiting in queue (not being served).

Little's Formula

Let

$N(t) =$ number of customers enter the system at or before time t.

We define λ_a be the arrival rate of entering customers,

$$\lambda_a = \lim_{t \to \infty} \frac{N(t)}{t}$$

Little's Formula:

$$L = \lambda_a W$$
$$L_Q = \lambda_a W_Q$$