Queueing theory fundamentals

Stochastic Delays

Cathy Wu

1.041/1.200 Transportation: Foundations and Methods
Readings

Unit 2: Queuing systems

LAB 1: Build your own traffic jam
LAB 2: Build a queuing model for Seattle transit
LAB 3: Build an AI agent to optimize traffic
LAB 4: Solve the traveling salesman problem

Unit 2
Modeling
Stochastic

- Modeling mathematical programs
- Linear programs
- Simplex method
- Time-space diagrams
- Cumulative diagrams
- Uncertainty
- Poisson process
- Queueing models
- Facility dynamics
- Markov chains
- Discrete event simulation
- Value iteration
- Q-learning
- Integer programs
- Deep learning
- Deep Q Networks
- Branch-and-bound
- Traffic flow theory
- Vehicle dynamics
- Numerical integration
- Sequential decision problems
- Integer programs
- Markov decision processes
- Time-space diagrams
- Cumulative diagrams
- Vehicle dynamics
- Numerical integration
- Sequential decision problems
- Integer programs
- Markov decision processes
- Time-space diagrams
- Cumulative diagrams
Outline

1. Delays in transportation
2. Queueing theory fundamentals
3. Problem: timing of a pedestrian crossing light
Outline

1. Delays in transportation
2. Queueing theory fundamentals
3. Problem: timing of a pedestrian crossing light
Delays in transportation

- Delays are pervasive in transportation

- Today: Introduce stochastic models of delay
  - To diagnose and improve delay
  - Framework: Queueing theory
Local delays

<table>
<thead>
<tr>
<th>#</th>
<th>U.S. City</th>
<th>Hours Lost Per Year Per Driver</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Boston</td>
<td>164</td>
</tr>
<tr>
<td>2</td>
<td>Washington, D.C.</td>
<td>155</td>
</tr>
<tr>
<td>3</td>
<td>Chicago</td>
<td>138</td>
</tr>
<tr>
<td>4</td>
<td>Seattle</td>
<td>138</td>
</tr>
<tr>
<td>5</td>
<td>New York City</td>
<td>133</td>
</tr>
<tr>
<td>6</td>
<td>Los Angeles</td>
<td>128</td>
</tr>
</tbody>
</table>

Sources:
Boston.com
Boston Magazine
[February, 2019]
Delays in transportation

- Congestion/delays arises across all transportation modes
- Sources: Urban mobility report (TTI, 2009); Transportation Vision for 2030 (US DOT, 2008)
Delays in transportation

- Urban congestion:
  - “Urban congestion lead in 2007 to an estimated additional 4.2 billion hours of travel and 2.8 billion gallons of fuel with a cost of $87.2 billion across urban areas in the US, an increase of more than 50% over the previous decade”

- Highway traffic:
  - “Highway vehicle miles traveled are projected to grow 60%, from 2,952 billion miles traveled in 2005 to 4,733 billion miles traveled in 2030.”
Delays in transportation

- **Air traffic:**
  - “The airline industry’s on-time performance in the first seven months of 2007 was the worst on record, and nationally almost 30% of all flights are now cancelled or substantially delayed.”
  - “Aircraft travel is projected to nearly double, and current forecasts estimate over 1.5 billion air passengers annually by 2030. This will place unparalleled demand on the air system.”

- **Freight traffic:**
  - “The U.S. transportation system currently moves over 50 million tons of freight on the US transportation network. (...) By 2035, tons transported overall are expected to double to over 100 million, placing incomparable pressure on our domestic transportation network.”
Outline

1. Delays in transportation

2. Queueing theory fundamentals
   a. Exponential distribution
   b. Poisson process

3. Problem: timing of a pedestrian crossing light
Background for queuing theory

- To be useful, the assumed distributional form, should be:
  1. Sufficiently realistic (reasonable predictions)
  2. Sufficiently simple (mathematically tractable)

- Key tools
  - The Poisson process
  - The exponential distribution, which is intimately tied to the Poisson process
Exponential distribution

- Also called the negative exponential
- \( T \sim \text{Exp}(\lambda) \)
- \( \lambda \) is a parameter (a constant)
- PDF: \( f_T(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \)
- \( E[T] = \frac{1}{\lambda} \)
- \( \text{Var}[T] = \frac{1}{\lambda^2} \)
- CDF: The probability that \( T \) will be less than or equal to any particular constant value, \( t \), is equal to:

\[
F_T(t) = P(T \leq t) = \int_{x=0}^{t} \lambda e^{-\lambda x} \, dx = 1 - e^{-\lambda t}
\]
Exponential distribution: numerical example

Gap acceptance problem

- Suppose that $T$:
  - a) represents the times between successive passages (headways) of vehicles on a road; and
  - b) is described by an exponential pdf.

- A driver requires 5 sec to insert into mainstream traffic that flows at 400 veh/hr

- What is the probability that the next gap is acceptable?
Exponential dbn

- What are the implications of assuming $T \sim \text{Exp}(\lambda)$?

Main properties of the exponential dbn:

- Prop. 1) $f_T(t)$, the pdf, is a strictly decreasing function of $t$ ($t \geq 0$)
  - $P(0 \leq T \leq \Delta t) > P(t \leq T \leq t + \Delta t)$, $t > 0$, $\Delta t > 0$
  - $P\left(0 < T < \frac{E[T]}{2}\right) = 0.393$
  - $P\left(\frac{E[T]}{2} < T < \frac{3E[T]}{2}\right) = 0.38$
  - So, $T$ is more likely to be $< \frac{E[T]}{2}$ (small), than near $E[T]$, even though 2nd interval is twice as wide as the 1st.

- Implications in practice?
Exponential dbn

- Prop. 2) Lack of memory ("memoryless")
  \[ P(T > t + s | T \geq s) = P(T > t) \quad t > 0, \quad s > 0 \]

- Interpretation: The probability dbn of the remaining time until the occurrence of the next event (e.g. next vehicle arrival, next service completion) is always the same, regardless of how much time has already passed.
  - For inter-arrival times: the time until the next arrival is uninfluenced by when the last arrival occurred
  - For service times: if they differ from customer to customer, then "memoryless" may be a desirable property

- The exponential dbn is the only continuous dbn with this property.

- Also, a small interval is independent of the time:
  \[ P(T \leq t + \varepsilon | T > t) = \lambda \varepsilon + o(\varepsilon), \quad \varepsilon \ll 1 \]

- Implications in practice?
Example: Jitney rider

- Carla waits on the side of the road for a **jitney**, which will transport her to the next town. This jitney travels along a fixed route between the edge of two towns.
  - A jitney is a form of unlicensed taxi, often unscheduled. Jitney service (in various forms) is prevalent in developing countries.
  - Trivia: Technically, Uber is considered an illegal taxi service in jurisdictions with medallion systems that restrict the number of legal cabs in operation.

1. Suppose that the interarrival time of the jitney service is an **exponentially distributed random variable with a mean of 10 minutes**. Carla has **already waited 15.5 minutes**. What is the expected additional time (conditional mean) that she will have to wait?

2. Now assume the jitney interarrival time is **uniformly distributed between 2 to 18 minutes**. What is the expected additional time (conditional mean) that Carla will have to wait?
Example: Jitney rider

1. Suppose that the interarrival time of the jitney service is an exponentially distributed random variable with a mean of 10 minutes. Carla has already waited 15.5 minutes. What is the expected additional time (conditional mean) that she will have to wait?

2. Now assume the jitney interarrival time is uniformly distributed between 2 to 18 minutes. What is the expected additional time (conditional mean) that Carla will have to wait?
Exponential dbn

- Prop. 3) Minimum of Exp. R.V.s yields an Exp. R.V.
  
  \[ T_i \sim \text{Exp}(\lambda_i), \quad \forall i \in 1, 2, \ldots, n \quad \text{independent} \]
  
  \[ U = \min_{i=1,\ldots,n} T_i \]
  
  \[ U \sim \text{Exp}\left(\sum_{i=1}^{n} \lambda_i\right) \]

- Inter-arrivals of different types, then the time until the next arrival follows an exponential dbn.

- Example: \( n \) check-in counters currently attending customers, then the time until the next service completion follows an exponential dbn.
Exponential dbn

- Prop. 4) Relationship to the Erlang distribution.
- \( k \) independent r.v’s: \( T_i \sim \text{Exp}(\lambda), \forall i \in 1, 2, \ldots, k \)
  Then, \( E_k = \sum_{i=1}^{k} T_i \) follows a \( k^{th} \)-order Erlang dbn.
  \[
  f_{E_k}(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}, \quad x \geq 0, \quad k = 1, 2, \ldots
  \]
- \( E[E_k] = \frac{k}{\lambda} \)
- \( \text{Var}(E_k) = \frac{k}{\lambda^2} \)
- Example: total waiting time of passengers for a bus
- Note (sanity check): the exponential dbn coincides with \( k = 1 \).
Exponential dbn

- Prop. 5) Relationship to the Poisson process.
- Suppose we have a sequence of demand arrivals ("events") such that
  a) successive demand inter-arrival times are mutually independent and
  b) the demand inter-arrival times are all described by the same exponential pdf.
Then, the number of arrivals constitutes a "Poisson process".

- Stated the other way around:
The inter-arrival times between events in a Poisson process that occur
at the rate of $\lambda$ per unit of time are: (i) mutually independent; and (ii)
described by an exponential pdf with parameter $\lambda$. 

Outline

1. Delays in transportation

2. Queueing theory fundamentals
   a. Exponential distribution
   b. Poisson process

3. Problem: timing of a pedestrian crossing light
The Poisson process

- Count the number of events $N(t)$ that occur during time interval $[0, t]$
- Definition (Poisson process): If successive inter-event (inter-arrival) times are independent and identically distributed as $Exp(\lambda)$, then: $\{N(t)\}$ is a Poisson process with rate $\lambda$
- $P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$: probability that $n$ events take place during the time interval $t$
- $t$: time interval length
- $\lambda$: arrival rate (veh/unit time), also referred to as the “intensity” of the arrivals of Poisson events
- $E[N(t)] = \lambda t$, $Var(N(T)) = \lambda t$
Poisson process $\rightarrow$ Exponential dbn

- $T$ is the time between successive arrivals, then during a time interval of length $t \leq T$, no arrivals occurred, so for all such $t$:

$$P(t \leq T) = P(N(t) = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

- Cumulative dbn function:

$$F(t; \lambda) = P(0 \leq T \leq t) = 1 - e^{-\lambda t}$$

- Probability density function:

$$f(t; \lambda) = \frac{dF(t; \lambda)}{dt} = \lambda e^{-\lambda t}$$

- Inter-arrival times follow an exponential dbn, with parameter $\lambda$.

Inter-arrival times that are i.i.d. $Exp(\lambda)$ is equivalent to arrivals according to a Poisson process $P(\lambda)$
Fundamental properties of the Poisson process

1. For a sufficiently small $\Delta t$:
   \[
P(N(t + \Delta t) - N(t) = 1) \approx \lambda \Delta t \\
P(N(t + \Delta t) - N(t) = 0) \approx 1 - \lambda \Delta t \\
P(N(t + \Delta t) - N(t) \geq 2) \approx 0
\]
   where these probabilities represent, respectively:
   - the probability that exactly one event will occur in the next $\Delta t$
   - the probability that no event will occur in the next $\Delta t$
   - the probability that two or more events will occur in the next $\Delta t$

2. The number of events that occur in disjoint time intervals are mutually independent r.v.'s

3. The number of events that occur during any pre-specified interval of length $\Delta t$ does not depend on the “starting time” of the time interval or on the number of events recorded prior to the time interval
More properties of the Poisson process

Given the relationship between the Poisson process and the exponential dbn, we have the following properties:

1. (“Memoryless”) If we begin observing a Poisson process with rate $\lambda$ at $t = 0$, the pdf for the time, $X$, until the next arrival is given by $f_X(t) = \lambda e^{-\lambda t}, t \geq 0$ no matter how long before $t = 0$ the last arrival occurred.

2. “The sum of $K$ independent Poisson processes is a Poisson process with a rate equal to the sum of the $K$ rates”.

3. Given two independent Poisson processes, with rates $\lambda_1$ and $\lambda_2$, respectively. Let $X_1$ and $X_2$, respectively, be the time until the next arrival from each process. Then, the probability that an arrival from Process 1 will take place before an arrival from Process 2 is equal to:

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$
Poisson events and Bernoulli trials

- Note that by Property 3 in the previous slide, the probability that the next arrival from either process will be a process 1 arrival is given by:
  \[ p = \frac{\lambda_1}{\lambda_1 + \lambda_2} \]
  and the probability that it will be a process 2 arrival by:
  \[ q = 1 - p = \frac{\lambda_2}{\lambda_1 + \lambda_2} \]

- Thus, the type of arrival (from process 1 or from process 2) that will occur when an arrival takes place is determined by a Bernoulli trial ("coin flip").

- The probability that, out of \( n \) arrivals, \( m \) will be from process 1 is given by:
  \[ P(m \text{ successes in } n \text{ trials}) = C_m^n p^m (1 - p)^{n-m}, \quad 0 \leq m \leq n \]
Outline

1. Delays in transportation
2. Queueing theory fundamentals
3. Problem: timing of a pedestrian crossing light
Example: Timing of a pedestrian crossing light*

- Pedestrians approach from the left (resp. right) side of the crossing in a Poisson manner with arrival rate $\lambda_L$ (resp. $\lambda_R$) arrivals per minute. Pedestrians wait until light turns green, referred to as a “dump”. Assume:
  - all pedestrians cross instantaneously
  - the dump duration is zero
  - left and right arrival processes are independent.

- We consider three operating rules:
  - Dump every $t_1$ minutes
  - Dump whenever the total number of waiting pedestrians equals $n_0$
  - Dump when the first pedestrian to arrive, after the previous dump, has waited for $t_2$ minutes

- For each rule, determine:
  - 1. The expected number of pedestrians crossing left to right on any dump
  - 2. The probability that zero pedestrians cross left to right on any particular dump
  - 3. The pdf for the time between dumps

Example: Timing of a pedestrian crossing light

- Arrival rates $\lambda_L, \lambda_R$

- We consider three operating rules:
  - A. Dump every $t_1$ minutes
  - B. Dump whenever the total number of waiting pedestrians equals $n_0$
  - C. Dump when the first pedestrian to arrive, after the previous dump, has waited for $t_2$ minutes

For each rule, determine:
- 1. The expected number of pedestrians crossing left to right on any dump
- 2. The probability that zero pedestrians cross left to right on any particular dump
- 3. The pdf for the time between dumps

Example: Timing of a pedestrian crossing light

- Arrival rates $\lambda_L, \lambda_R$

- We consider three operating rules:
  - A. Dump every $t_1$ minutes
  - B. Dump whenever the total number of waiting pedestrians equals $n_0$
  - C. Dump when the first pedestrian to arrive, after the previous dump, has waited for $t_2$ minutes

- For each rule, determine:
  1. The expected number of pedestrians crossing left to right on any dump
  2. The probability that zero pedestrians cross left to right on any particular dump
  3. The pdf for the time between dumps

References


3. Slides adapted from Carolina Osorio