TCOM 501: Networking Theory & Fundamentals

Lectures 4 & 5
February 5 and 12, 2003
Prof. Yannis A. Korilis
Topics

- Markov Chains
- M/M/1 Queue
- Poisson Arrivals See Time Averages
- M/M/* Queues
- Introduction to Sojourn Times
The M/M/1 Queue

- Arrival process: Poisson with rate $\lambda$
- Service times: iid, exponential with parameter $\mu$
- Service times and interarrival times: independent
- Single server
- Infinite waiting room
- $N(t)$: Number of customers in system at time $t$ (state)
**Exponential Random Variables**

- **X**: exponential RV with parameter $\lambda$
- **Y**: exponential RV with parameter $\mu$
- **X, Y**: independent

Then:

1. $\min\{X, Y\}$: exponential RV with parameter $\lambda+\mu$
2. $P\{X<Y\} = \frac{\lambda}{\lambda+\mu}$

[Exercise 3.12]

**Proof:**

\[
P\{\min\{X, Y\} > t\} = P\{X > t, Y > t\} = P\{X > t\}P\{Y > t\} = e^{-\lambda t}e^{-\mu t} = e^{-(\lambda+\mu)t} \Rightarrow P\{\min\{X, Y\} \leq t\} = 1 - e^{-(\lambda+\mu)t}
\]

\[
P\{X < Y\} = \int_0^\infty \int_0^y f_{XY}(x, y) \, dx \, dy = \int_0^\infty \int_0^y \lambda e^{-\lambda x} \cdot \mu e^{-\mu y} \, dx \, dy = \int_0^\infty \mu e^{-\mu y} \left( \int_0^y \lambda e^{-\lambda x} \, dx \right) \, dy
\]

\[
= \int_0^\infty \mu e^{-\mu y} \left( 1 - e^{-\lambda y} \right) \, dy = \int_0^\infty \mu e^{-\mu y} \, dy - \frac{\mu}{\lambda+\mu} \int_0^\infty (\lambda + \mu)e^{-(\lambda+\mu)y} \, dy
\]

\[
= 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}
\]
M/M/1 Queue: Markov Chain Formulation

- Jumps of \( \{N(t): t \geq 0\} \) triggered by arrivals and departures
- \( \{N(t): t \geq 0\} \) can jump only between neighboring states

Assume process at time \( t \) is in state \( i: N(t) = i \geq 1 \)

- \( X_i \): time until the next arrival – exponential with parameter \( \lambda \)
- \( Y_i \): time until the next departure – exponential with parameter \( \mu \)
- \( T_i = \min \{X_i, Y_i\} \): time process spends at state \( i \)
- \( T_i \): exponential with parameter \( \nu_i = \lambda + \mu \)
- \( P_{i,i+1} = P \{X_i < Y_i\} = \frac{\lambda}{\lambda + \mu} \), \( P_{i,i-1} = P \{Y_i < X_i\} = \frac{\mu}{\lambda + \mu} \)
- \( P_{01} = 1 \), and \( T_0 \) is exponential with parameter \( \lambda \)

- \( \{N(t): t \geq 0\} \) is a continuous-time Markov chain with

\[
q_{i,i+1} = \nu_i P_{i,i+1} = \lambda, \quad i \geq 0
\]
\[
q_{i,i-1} = \nu_i P_{i,i-1} = \mu, \quad i \geq 1
\]
\[
q_{ij} = 0, \quad |i - j| > 1
\]
M/M/1 Queue: Stationary Distribution

- Birth-death process $\rightarrow$ DBE
  \[
  \mu p_n = \lambda p_{n-1} \Rightarrow \\
  p_n = \frac{\lambda}{\mu} p_{n-1} = \rho p_{n-1} = \ldots = \rho^n p_0
  \]

- Normalization constant
  \[
  \sum_{n=0}^{\infty} p_n = 1 \iff p_0 \left[ 1 + \sum_{n=1}^{\infty} \rho^n \right] = 1 \iff p_0 = 1 - \rho, \text{ if } \rho < 1
  \]

- Stationary distribution
  \[
  p_n = \rho^n (1 - \rho), \quad n = 0, 1, \ldots
  \]
The M/M/1 Queue

- Average number of customers in system
  \[ N = \sum_{n=0}^{\infty} np_n = (1 - \rho) \sum_{n=0}^{\infty} n \rho^n = (1 - \rho) \rho \sum_{n=0}^{\infty} n \rho^{n-1} \]
  \[ \Rightarrow N = \rho(1 - \rho) \frac{1}{(1 - \rho)^2} = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda} \]

- Little’s Theorem: average time in system
  \[ T = \frac{N}{\lambda} = \frac{1}{\lambda} \frac{\lambda}{\mu - \lambda} = \frac{1}{\mu - \lambda} \]

- Average waiting time and number of customers in the queue – excluding service
  \[ W = T - \frac{1}{\mu} = \frac{\rho}{\mu - \lambda} \quad \text{and} \quad N_Q = \lambda W = \frac{\rho^2}{1 - \rho} \]
The M/M/1 Queue

- $\rho = \frac{\lambda}{\mu}$: utilization factor
- Long term proportion of time that server is busy
- $\rho = 1 - p_0$: holds for any M/G/1 queue
- Stability condition: $\rho < 1$
  - Arrival rate should be less than the service rate
M/M/1 Queue: Discrete-Time Approach

- Focus on times 0, δ, 2δ, ... (δ arbitrarily small)
- Study discrete time process \( N_k = N(\delta k) \)
  \[
  \lim_{t \to \infty} P\{N(t) = n\} = \lim_{k \to \infty} P\{N_k = n\}
  \]
- Show that transition probabilities are
  \[
  \begin{align*}
  P_{00} &= 1 - \lambda \delta + o(\delta) \\
  P_{ii} &= 1 - \lambda \delta - \mu \delta + o(\delta), \quad i \geq 1 \\
  P_{i,i+1} &= \lambda \delta + o(\delta), \quad i \geq 0 \\
  P_{i,i-1} &= \mu \delta + o(\delta), \quad i \geq 0 \\
  P_{ij} &= o(\delta), \quad |i - j| > 1
  \end{align*}
  \]
- Discrete time Markov chain, omitting \( o(\delta) \)
**M/M/1 Queue: Discrete-Time Approach**

- **Discrete-time birth-death process → DBE:**

  \[
  [\mu\delta + o(\delta)]\pi_n = [\lambda\delta + o(\delta)]\pi_{n-1} \Rightarrow \\
  \pi_n = \frac{\lambda\delta + o(\delta)}{\mu\delta + o(\delta)}\pi_{n-1} = \cdots = \left[\frac{\lambda\delta + o(\delta)}{\mu\delta + o(\delta)}\right]^{n} \pi_0
  \]

- **Taking the limit \( \delta \to 0: \)**

  \[
  \lim_{\delta \to 0} \pi_n = \lim_{\delta \to 0} \left[\frac{\lambda\delta + o(\delta)}{\mu\delta + o(\delta)}\right]^{n} \lim_{\delta \to 0} \pi_0 \Rightarrow \, p_n = \left(\frac{\lambda}{\mu}\right)^n p_0
  \]

- **Done!**
Transition Probabilities?

- $A_k$: number of customers that arrive in $I_k=(k\delta, (k+1)\delta]$
- $D_k$: number of customers that depart in $I_k=(k\delta, (k+1)\delta]$
- Transition probabilities $P_{ij}$ depend on conditional probabilities:
  \[ Q(a,d \mid n) = P\{A_k=a, D_k=d \mid N_{k-1}=n\} \]
- Calculate $Q(a,d \mid n)$ using arrival and departure statistics
- Use Taylor expansion $e^{-\lambda \delta} = 1 - \lambda \delta + o(\delta)$, $e^{-\mu \delta} = 1 - \mu \delta + o(\delta)$, to express as a function of $\delta$
- Poisson arrivals: $P\{A_k \geq 2\} = o(\delta)$
- Probability there are more than 1 arrivals in $I_k$ is $o(\delta)$
  - Show: probability of more than one event (arrival or departure) in $I_k$ is $o(\delta)$
  - See details in textbook
**Example: Slowing Down**

- M/M/1 system: slow down the arrival and service rates by the same factor $m$
- Utilization factors are the same $\Rightarrow$ stationary distributions the same, average number in the system the same
- Delay in the slower system is $m$ times higher
- Average number in queue is the same, but in the 1st system the customers move out faster

\[
N = \frac{\rho}{1 - \rho} = \frac{\lambda / \mu}{1 - \lambda / \mu} = \frac{\lambda}{\mu - \lambda}
\]

\[
T = \frac{N}{\lambda} = \frac{1}{\mu - \lambda}
\]

\[
W = \frac{\rho}{\mu - \lambda} = \frac{\lambda / \mu}{\mu - \lambda}
\]

\[
N' = \frac{\rho'}{1 - \rho'} = \frac{\lambda / \mu}{1 - \lambda / \mu} = \frac{\lambda}{\mu - \lambda} = N
\]

\[
T' = \frac{N'}{\lambda / m} = \frac{m}{\mu - \lambda} = mT
\]

\[
W' = \frac{\rho'}{\mu / m - \lambda / m} = \frac{m(\lambda / \mu)}{\mu - \lambda} = mW
\]
Example: Statistical MUX-ing vs. TDM

- \( m \) identical Poisson streams with rate \( \lambda/m \); link with capacity 1; packet lengths iid, exponential with mean \( 1/\mu \)
- Alternative: split the link to \( m \) channels with capacity \( 1/m \) each, and dedicate one channel to each traffic stream
- Delay in each “queue” becomes \( m \) times higher
- Statistical multiplexing vs. TDM or FDM
- When is TDM or FDM preferred over statistical multiplexing?
“PASTA” Theorem

Markov chain: “stationary” or “in steady-state:”
- Process started at the stationary distribution, or
- Process runs for an infinite time $t \to \infty$

Probability that at any time $t$, process is in state $i$ is equal to the stationary probability

$$p_i = \lim_{t \to \infty} P\{N(t) = i\} = \lim_{t \to \infty} \frac{T_i(t)}{t}$$

Question: For an M/M/1 queue: given $t$ is an arrival time, what is the probability that $N(t) = i$?
Answer: Poisson Arrivals See Time Averages!
PASTA Theorem

- Steady-state probabilities:
  \[ p_n = \lim_{t \to \infty} P\{N(t) = n\} \]

- Steady-state probabilities upon arrival:
  \[ a_n = \lim_{t \to \infty} P\{N(t^-) = n \mid \text{arrival at } t\} \]

- Lack of Anticipation Assumption (LAA): Future inter-arrival times and service times of previously arrived customers are independent

**Theorem:** In a queueing system satisfying LAA:

1. If the arrival process is Poisson:
   \[ a_n = p_n, \quad n = 0, 1,... \]

2. Poisson is the only process with this property (necessary and sufficient condition)
PASTA Theorem

Doesn’t PASTA apply for all arrival processes?

- Deterministic arrivals every 10 sec
- Deterministic service times 9 sec
- Upon arrival: system is always empty $a_1=0$
- Average time with one customer in system: $p_1=0.9$

- “Customer” averages need not be time averages
- Randomization does not help, unless Poisson!
PASTA Theorem: Proof

- Define $A(t, t+\delta)$, the event that an arrival occurs in $[t, t+\delta)$
- Given that a customer arrives at $t$, probability of finding the system in state $n$:
  \[
P\{N(t^-) = n \mid \text{arrival at } t\} = \lim_{\delta \to 0} P\{N(t^-) = n \mid A(t, t+\delta)\}
  \]
- $A(t, t+\delta)$ is independent of the state before time $t$, $N(t^-)$
  - $N(t^-)$ determined by arrival times $< t$, and corresponding service times
  - $A(t, t+\delta)$ independent of arrivals $< t$ [Poisson]
  - $A(t, t+\delta)$ independent of service times of customers arrived $< t$ [LAA]

\[
a_n(t) = \lim_{\delta \to 0} P\{N(t^-) = n \mid A(t, t+\delta)\} = \lim_{\delta \to 0} \frac{P\{N(t^-) = n, A(t, t+\delta)\}}{P\{A(t, t+\delta)\}}
\]

\[
= \lim_{\delta \to 0} \frac{P\{N(t^-) = n\} P\{A(t, t+\delta)\}}{P\{A(t, t+\delta)\}} = P\{N(t^-) = n\}
\]

\[
a_n = \lim_{t \to \infty} a_n(t) = \lim_{t \to \infty} P\{N(t^-) = n\} = p_n
\]
PASTA Theorem: Intuitive Proof

- $t_a$ and $t_r$: randomly selected arrival and observation times, respectively.
- The arrival processes prior to $t_a$ and $t_r$, respectively, are stochastically identical.
  - The probability distributions of the time to the first arrival before $t_a$ and $t_r$ are both exponentially distributed with parameter $\lambda$.
  - Extending this to the 2nd, 3rd, etc. arrivals before $t_a$ and $t_r$ establishes the result.
- State of the system at a given time $t$ depends only on the arrivals (and associated service times) before $t$.
  - Since the arrival processes before arrival times and random times are identical, so is the state of the system they see.
Arrivals that Do not See Time-Averages

Example 1: Non-Poisson arrivals
- IID inter-arrival times, uniformly distributed between in 2 and 4 sec
- Service times deterministic 1 sec
- Upon arrival: system is always empty
- $\lambda = 1/3, T = 1 \rightarrow N = T/\lambda = 1/3 \rightarrow p_1 = 1/3$

Example 2: LAA violated
- Poisson arrivals
- Service time of customer $i$: $S_i = \alpha T_{i+1}$, $\alpha < 1$
- Upon arrival: system is always empty
- Average time the system has 1 customer: $p_1 = \alpha$
Distribution after Departure

- Steady-state probabilities after departure:
  \[ d_n = \lim_{t \to \infty} P\{X(t^+) = n \mid \text{departure at } t\} \]

- Under very general assumptions:
  - \( N(t) \) changes in unit increments
  - limits \( a_n \) and exist \( d_n \)
  - \( a_n = d_n, n=0,1,\ldots \)
  - In steady-state, system appears stochastically identical to an arriving and departing customer

- Poisson arrivals + LAA: an arriving and a departing customer see a system that is stochastically to the one seen by an observer looking at an arbitrary time
M/M/* Queues

- Poisson arrival process
  - Interarrival times: iid, exponential
- Service times: iid, exponential
- Service times and interarrival times: independent
- $N(t)$: Number of customers in system at time $t$ (state)
  - $\{N(t): t \geq 0\}$ can be modeled as a continuous-time Markov chain
  - Transition rates depend on the characteristics of the system
  - PASTA Theorem always holds
M/M/1/K Queue

- M/M/1 with finite waiting room
  - At most K customers in the system
  - Customer that upon arrival finds K customers in system is dropped
- Stationary distribution
  \[ p_n = \rho^n p_0, \quad n = 1,2,\ldots,K \]
  \[ p_0 = \frac{1-\rho}{1-\rho^{K+1}} \]
- Stability condition: always stable – even if \( \rho \geq 1 \)
- Probability of loss – using PASTA theorem:
  \[ P\{\text{loss}\} = P\{N(t) = K\} = \frac{\rho^K (1-\rho)}{1-\rho^{K+1}} \]
Exactly as in the M/M/1 queue:

\[ p_n = \rho^n p_0, \quad n = 1, 2, ..., K \]

Normalization constant:

\[
\sum_{n=0}^{K} p_n = 1 \Rightarrow p_0 \sum_{n=1}^{K} \rho^n = 1 \Rightarrow p_0 \frac{1-\rho^{K+1}}{1-\rho} = 1
\]

\[ \Rightarrow p_0 = \frac{1-\rho}{1-\rho^{K+1}} \]

**Generalize:** Truncating a Markov chain
Truncating a Markov Chain

- \( \{X(t): t \geq 0\} \) continuous-time Markov chain with stationary distribution \( \{p_i: i=0,1,...\} \)
- \( S \) a subset of \( \{0,1,...\} \): set of states; Observe process only in \( S \)
  - Eliminate all states not in \( S \)
  - Set \( \tilde{q}_{ji} = \tilde{q}_{ij} = 0, \ j \in S, i \not\in S \)

- \( \{Y(t): t \geq 0\} \): resulting truncated process; If irreducible:
  - Continuous-time Markov chain
  - Stationary distribution

\[
\tilde{p}_j = \begin{cases} 
  \frac{p_j}{\sum_{i \in S} p_i} & \text{if } j \in S \\
  0 & \text{if } j \not\in S
\end{cases}
\]

- Under certain conditions – need to verify depending on the system
Possible sufficient condition

\[ p_j \sum_{i \in S} q_{ji} = \sum_{i \in S} p_i q_{ij}, \quad j \in S \]

Verify that distribution of truncated process

1. Satisfies the GBE

\begin{align*}
    p_j \sum_{i \in S} q_{ji} &= \sum_{i \in S} p_i q_{ij} \\
    &\Rightarrow p_j \sum_{i \in S} q_{ji} = \sum_{i \in S} p_i q_{ij} \Rightarrow \frac{p_j}{p(S)} \sum_{i \in S} q_{ji} = \sum_{i \in S} \frac{p_i}{p(S)} q_{ij} \\
    &\Rightarrow \tilde{p}_j \sum_{i \in S} q_{ji} = \sum_{i \in S} \tilde{p}_i q_{ij} \Rightarrow \tilde{p}_j \sum_{i \in S} \tilde{q}_{ji} = \sum_{i \in S} \tilde{p}_i \tilde{q}_{ij}, \quad j \in S
\end{align*}

2. Satisfies the probability conservation law:

\[ \sum_{i \in S} \tilde{p}_i = \sum_{i \in S} \frac{p_i}{p(S)} = \frac{p(S)}{p(S)} = 1, \quad p(S) \equiv \sum_{i \in S} p_i \]

Another – even better – sufficient condition: DBE!

Relates to “reversibility”

Holds for multidimensional chains
M/M/1 Queue with State-Dependent Rates

- Interarrival times: independent, exponential, with parameter $\lambda_n$ when at state $n$
- Service times: independent, exponential, with parameter $\mu_n$ when at state $n$
- Service times and interarrival times: independent
- $\{N(t): t \geq 0\}$ is a birth-death process
- Stationary distribution:

$$p_n = p_0 \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}, \quad n \geq 1 \quad p_0 = \left[ 1 + \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}} \right]^{-1}$$
M/M/c Queue

- Poisson arrivals with rate \( \lambda \)
- Exponential service times with parameter \( \mu \)
- \( c \) servers
- Arriving customer finds \( n \) customers in system
  - \( n < c \): it is routed to any idle server
  - \( n \geq c \): it joins the waiting queue – all servers are busy
- Birth-death process with state-dependent death rates

\[
\mu_n = \begin{cases} 
  n\mu, & 1 \leq n \leq c \\
  c\mu, & n \geq c 
\end{cases}
\]

[Time spent at state \( n \) before jumping to \( n -1 \) is the minimum of \( B_n = \min\{n,c\} \) exponentials with parameter \( \mu \)]
**M/M/c Queue**

![Diagram of M/M/c Queue](image)

- **Detailed balance equations**

  \[ 1 \leq n \leq c: \quad p_n = \frac{\lambda}{n\mu} p_{n-1} = \cdots = \frac{\lambda}{n\mu (n-1)\mu} \cdots \frac{\lambda}{\mu} p_0 = \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n p_0 = \frac{(c\rho)^n}{n!} p_0, \quad \rho = \frac{\lambda}{c\mu} \]

  \[ n > c: \quad p_n = \left( \frac{\lambda}{c\mu} \right)^{n-c} p_c = \frac{1}{c!} \left( \frac{\lambda}{\mu} \right)^c \left( \frac{\lambda}{c\mu} \right)^{n-c} p_0 = \frac{c^c}{c!} \left( \frac{\lambda}{c\mu} \right)^n p_0 = \frac{c^c \rho^n}{c!} p_0 \]

- **Normalizing**

  \[ \sum_{n=0}^{\infty} p_n = 1 \Rightarrow p_0 = \left[ 1 + \sum_{k=1}^{c-1} \frac{(c\rho)^k}{k!} + \frac{(c\rho)^c}{c!} \sum_{k=c}^{\infty} \rho^{k-c} \right]^{-1} = \left[ \sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} + \frac{(c\rho)^c}{c!} \frac{1}{1-\rho} \right]^{-1} \]
M/M/c Queue

- Probability of queueing – arriving customer finds all servers busy

\[ P_Q = P\{\text{queueing}\} = \sum_{n=c}^{\infty} p_n = p_0 \frac{(c\rho)^c}{c!} \sum_{n=c}^{\infty} \rho^{n-c} = \frac{(c\rho)^c}{c!} \frac{1}{1 - \rho} p_0 \]

- Erlang-C Formula: used in telephony and circuit-switching
  - Call requests arrive with rate \( \lambda \); holding time of a call exponential with mean \( 1/\mu \)
  - \( c \) available circuits on a transmission line
  - A call that finds all \( c \) circuits busy, continuously attempts to find a free circuit – “remains in queue”

- M/M/c/c Queue: c-server loss system
  - A call that finds all \( c \) circuits busy is blocked
  - Erlang-B Formula: popular in telephony
M/M/c Queue

- Expected number of customers waiting in queue – not in service
  \[ N_Q = \sum_{n=c}^{\infty} (n-c) p_n = p_0 \frac{(c\rho)^c}{c!} \sum_{n=c}^{\infty} (n-c) \rho^{n-c} = p_0 \frac{(c\rho)^c}{c!} \frac{\rho}{(1-\rho)^2} \]
  \[ = P_0 (1-\rho) \frac{\rho}{(1-\rho)^2} = P_0 \frac{\rho}{1-\rho} \]

- Average waiting time (in queue)
  \[ W = \frac{N_Q}{\lambda} = P_0 \frac{\rho}{\lambda(1-\rho)} \]

- Average time in system (queued + serviced)
  \[ T = W + \frac{1}{\mu} = P_0 \frac{\rho}{\lambda(1-\rho)} + \frac{1}{\mu} \]

- Expected number of customers in system
  \[ N = \lambda T = P_0 \frac{\rho}{(1-\rho)} + c\rho \]
M/M/$\infty$ Queue: Infinite-Server System

- Infinite number of servers – no queueing
- Stationary distribution:
  \[ p_n = \frac{(\lambda / \mu)^n}{n!} e^{-\lambda / \mu}, \quad n = 0, 1, ... \]
  Poisson with rate $\lambda / \mu$
- Average number of customers & average delay:
  \[ N = \frac{\lambda}{\mu}, \quad T = \frac{N}{\lambda} = \frac{1}{\mu} \]

*The results hold for an M/G/$\infty$ queue*
M/M/c/c Queue: c-Server Loss System

- c servers, no waiting room
- An arriving customer that finds all servers busy is blocked
- Stationary distribution:

\[
p_n = \frac{(\lambda / \mu)^n}{n!} \left[ \sum_{k=0}^{c} \frac{(\lambda / \mu)^k}{k!} \right]^{-1}, \quad n = 0, 1, \ldots, c
\]

- Probability of blocking (using PASTA):

\[
p_c = \frac{(\lambda / \mu)^c}{c!} \left[ \sum_{k=0}^{c} \frac{(\lambda / \mu)^k}{k!} \right]^{-1}
\]

- Erlang-B Formula: used in telephony and circuit-switching

\textit{Results hold for an M/G/c/c queue}
M/M/∞ and M/M/c/c Queues (proof)

- **DBE:**

\[(n \mu) p_n = \lambda p_{n-1} \Rightarrow p_n = \frac{\lambda}{n \mu} p_{n-1} = \frac{\lambda}{n \mu (n-1) \mu} p_{n-2} = \cdots = \frac{\lambda \cdot \lambda \cdots \lambda}{n \mu \cdot (n-1) \mu \cdots \mu} p_0\]

\[\Rightarrow p_n = \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} p_0, \quad n = 0, 1, \ldots\]

- **Normalizing:**

\[p_0 = \left[ \sum_{k=0}^{\infty} \frac{(\lambda/\mu)^k}{k!} \right]^{-1}, \quad \text{for M/M/c/c} \]

\[p_0 = \left[ \sum_{k=0}^{\infty} \frac{(\lambda/\mu)^k}{k!} \right]^{-1} = e^{-\lambda/\mu}, \quad \text{for M/M/c/c} \]
Sum of IID Exponential RV’s

- \(X_1, X_2, \ldots, X_n\): iid, exponential with parameter \(\lambda\)
- \(T = X_1 + X_2 + \ldots + X_n\)
- The probability density function of \(T\) is:
  \[
  f_T(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0
  \]
  [Gamma distribution with parameters \((n, \lambda)\)]

- If \(X_i\) is the time between arrivals \(i-1\) and \(i\) of a certain type of events, then \(T\) is the time until the \(n^{th}\) event occurs
- For arbitrarily small \(\delta\):
  \[
  P\{n^{th}\ \text{arrival occurs in } [t, t + \delta)\} = \delta f_T(t) = \lambda \delta \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}
  \]
- Cumulative distribution function:
  \[
  P\{t_n \leq t\} = \int_0^t \frac{\lambda (\lambda s)^{n-1}}{(n-1)!} e^{-\lambda s} ds = 1 - P\{n^{th}\ \text{arrival occurs after } t\}
  \]
Sum of IID Exponential RV’s

Example 1: Poisson arrivals with rate \( \lambda \)

- \( \tau_1 \): time until arrival of 1st customer
- \( \tau_i \): \( i \)th interarrival time
- \( \tau_1, \tau_2, \ldots, \tau_n \): iid exponential with parameter \( \lambda \)
- \( t_n = \tau_1 + \tau_2 + \ldots, + \tau_n \): arrival time of customer \( n \)

\( t_n \) follows Gamma with parameters \( (n, \lambda) \).

\[
f(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0; \quad P\{t_n \leq t\} = \int_0^t \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} dt
\]

- For arbitrarily small \( \delta \):

\[
P\{n^{th} \text{ arrival occurs in } [t, t + \delta)\} = \delta f_T(t) = \lambda \delta \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}
\]
Sojourn Times in a M/M/1 Queue

- M/M/1 Queue – FCFS
- $T_i$: time spent in system (queueing + service) by customer $i$
- $T_i$: exponentially distributed with parameter $\mu - \lambda$
- Example of a sojourn time of a customer: describes the evolution of the queue together with the specific customer

Proof:
1. Direct calculation of probability distribution function
2. Moment generating functions
3. Intuitive: Exercise 3.11(b)
Proof 1: Let $t_i$ be the arrival time of customer $i$, and $N_i = N(t_i^-)$, the number of customers in the system right before the $i^{th}$ arrival.

\[
P\{T_i > t\} = \sum_{k=0}^{\infty} P\{T_i > t | N_i = k\} P\{N_i = k\}
\]

\[
= \sum_{k=0}^{\infty} P\{D(t_i + t) - D(t_i) \leq k\} p_k
\]

\[
= \sum_{k=0}^{\infty} \sum_{n=0}^{k} e^{-\mu t} \frac{(\mu t)^n}{n!} \cdot (1 - \rho) \rho^k
\]

\[
= e^{-\mu t} \sum_{n=0}^{\infty} \frac{(\mu t)^n}{n!} \sum_{k=n}^{\infty} (1 - \rho) \rho^k
\]

\[
= e^{-\mu t} \sum_{n=0}^{\infty} \frac{(\mu t)^n}{n!} \cdot \rho^n = e^{-\mu t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!}
\]

\[
= e^{-\mu t} e^{\lambda t} = e^{-(\mu - \lambda) t}
\]
M/M/1 Queue: Sojourn Times (proof)

Proof 1: Note that:

- Time customer $i$ stays in the system is greater than $t$, given that it finds $k$ customers in the system, iff the number of departures in interval $(t_i, t_i + t)$ are less than $k + 1$. The server is always busy during that interval, thus times between departures are iid, exponential with parameter $\mu$. Then:

$$P\{D(t_i + t) - D(t_i) = n\} = e^{-\mu t} \frac{(\mu t)^n}{n!}, \quad 0 \leq n \leq k$$

- $P\{N_i = k\} = p_k$, by PASTA theorem.

- Eq. (3) follows by changing order of summation.

- Eq. (4) uses:

$$\sum_{k=n}^{\infty} \rho^k = \sum_{k=0}^{\infty} \rho^k - \sum_{k=0}^{n-1} \rho^k = \frac{1}{1 - \rho} - \frac{1 - \rho^n}{1 - \rho} = \frac{\rho^n}{1 - \rho}$$
M/M/1 Queue: Sojourn Times (proof)

Proof 2:

- $N_i$: number of customers in system upon arrival of customer $i$
- $T_i^{(k)}$: sojourn time of customer $i$ when it finds $k$ customers in system

$$T_i^{(k)} = S_i + S_{i-1} + \ldots + S_{i-k+1} + R_{i-k}$$

$S_j$ is service time of customer $j$, and $R_{i-k}$ the residual service time of the customer in service.

- $S_i, \ldots, S_{i-k+1}$: iid, exponential with parameter $\mu$
- $R_{i-k}$: exponential with parameter $\mu$, independent of $S_i, \ldots, S_{i-k+1}$
- $T_i^{(k)}$ is the sum of $k$ iid exponential RV’s
- $T_i = T_i^{(N_i)}$ is the sum of a random number of iid exponential RV’s
- Use moment generating functions
M/M/1 Queue: Sojourn Times (proof)

Proof 2:

\[ M_{T_i}(t) = E[e^{tT_i}] = \sum_{k=0}^{\infty} E[e^{tT_i} | N_i = k] P\{N_i = k\} \]
\[ = \sum_{k=0}^{\infty} E[e^{tT_i^{(k)}}] p_k = \sum_{k=0}^{\infty} M_{T_i^{(k)}}(t) p_k \]

A RV is exponential with parameter \( \mu \) if and only if its moment generating function is \( \mu/(\mu - t) \).

\( T_i^{(k)} \) sum of \( k + 1 \) independent RV's:

\[ M_{T_i^{(k)}}(t) = M_{S_i}(t)M_{S_{i-1}}(t) \ldots M_{S_{i-k+1}}(t)M_{R_{i-k}}(t) = \left( \frac{\mu}{\mu - t} \right)^{k+1} \]

Then:

\[ M_{T_i}(t) = \sum_{k=0}^{\infty} \left( \frac{\mu}{\mu - t} \right)^{k+1} (1 - \rho) \rho^k = (1 - \rho) \frac{\mu}{\mu - t} \sum_{k=0}^{\infty} \left( \frac{\lambda}{\mu - t} \right)^k \]
\[ = \frac{\mu - \lambda}{\mu - t} \frac{1}{1 - \frac{\lambda}{\mu - t}} = \frac{\mu - \lambda}{(\mu - \lambda) - t} \]
Moment Generating Function

1. Definition: for any $t \in \mathbb{R}$:

$$M_X(t) = E[e^{tX}] = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f_X(x) \, dx, & X \text{ continuous} \\ \sum_{j} e^{tx_j} P\{X = x_j\}, & X \text{ discrete} \end{cases}$$

2. If the moment generating function $M_X(t)$ of $X$ exists and is finite in some neighborhood of $t = 0$, it determines the distribution of $X$ uniquely.

3. Fundamental Properties: for any $n \in \mathbb{N}$:

   (i) $\frac{d^n}{dt^n}M_X(t) = E[X^n e^{tX}]$

   (ii) $\frac{d^n}{dt^n}M_X(0) = E[X^n]$

4. Moment Generating Functions and Independence:

   $X, Y :$ independent $\Rightarrow M_{X+Y}(t) = M_X(t)M_Y(t)$

   The opposite $\not\Rightarrow$ true.