

Applications to Queueing Theory

Introduction to Stochastic Processes (Erhan Cinlar)
Ch. 6.5, 6.6

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης
(ΕΚΠΑ) - 2025

Applications to Queueing Theory: M/G/1 Queue

M/G/1:

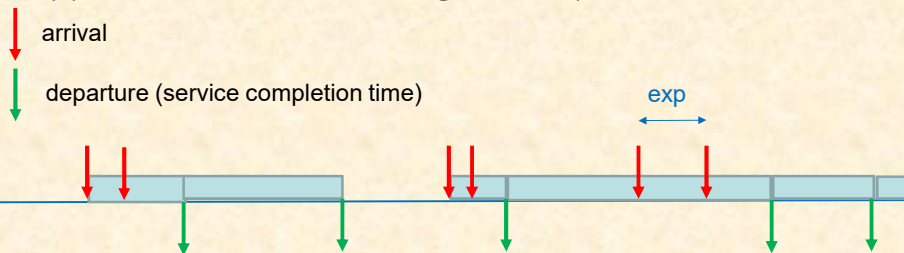
Arrival Process: **M**emoryless (Poisson arrival or exponential (geometric) interarrivals)

Service Process: **G**enerally-distributed service times

Number of servers: **1**

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queueing Theory: M/G/1 Queue



$X(t)$: Number of customers in the system (queue and under service)

Consider a specific subset of times $\{t_e\}$ only. That means that we embed $X(t)$ on times $\{t_e\}$. We do not look at $X(t)$ at times other than in $\{t_e\}$.

$X(t_e)$ is the process $X(t)$ embedded at times $\{t_e\}$.

$X(t)$ is not a MC. Why?

If $\{t_e\} = \{\text{times of customer departure}\}$, then $X(t_e)$ is a MC. Why?

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queueing Theory: M/G/1 Queue

$N_i(\omega)$: number of arrivals during the time interval $[0, t]$.

$Z_1(\omega), Z_2(\omega), \dots$: service times of customers who depart first, second, ...

$Y_i(\omega)$: number of customers in the system (waiting or being served at time t)

Assumptions:

♣ $N = \{N_i; i \geq 0\} \square P(a)$

♣ $Z_1, Z_2, \dots$ i.i.d. $\square \phi$

- Consider the future of Y from a time T of a departure onward.
- Define X_n as the number of customers in the system just after the instant of the n^{th} departure. (X_n is a SP embedded at departure times)

Theorem: X is a MC with the transition matrix

$$P = \begin{pmatrix} q_0 & q_1 & q_2 & q_3 & \cdots \\ q_0 & q_1 & q_2 & q_3 & \cdots \\ & q_0 & q_1 & q_2 & \cdots \\ & & q_0 & q_1 & \cdots \\ & & & q_0 & \cdots \\ & & & & \ddots \end{pmatrix}, \quad q_k = \int_0^\infty \frac{e^{-at}(at)^k}{k!} d\phi(t), \quad k = 0, 1, \dots$$

Distribution of arrivals over a service time

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Proof: We need to show $P\{X_{n+1} = j \mid X_0, \dots, X_n\} = P\{X_{n+1} = j \mid X_n\}$

$$P\{X_{n+1} = j \mid X_n = i\} = \begin{cases} q_j & i=0, j \geq 0 \\ q_{j+1-i} & i > 0, j \geq i-1 \\ 0 & \text{otherwise} \end{cases}$$

- Let T the time of the n^{th} departure.
- Let $Z = Z_{n+1}$ the service time of the $n+1$ customer.

$$\text{Then, } X_{n+1} = \begin{cases} X_n + (N_{T+Z} - N_T) - 1, & X_n > 0 \\ N_{S+Z} - N_S, & X_n = 0 \end{cases} \quad (\text{S: arrival time of the } n+1 \text{ customer})$$

Using Poisson properties: $P\{N_{T+Z} - N_T = k \mid X_0, \dots, X_n; T\} = P\{N_Z = k\}$

Distribution of arrivals over a service time

$$\rightarrow q_k = P\{N_Z = k\} = E[P\{N_Z = k \mid Z\}] = E\left[\frac{e^{-aZ}(aZ)^k}{k!}\right] = \int_0^\infty \frac{e^{-at}(at)^k}{k!} d\phi(t)$$

- $i = 0$ $P\{X_{n+1} = j \mid X_n = 0\} = P\{N_{S+Z} - N_S = j\} = P\{N_Z = j\} = q_j$
- $i > 0$ $P\{X_{n+1} = j \mid X_n = i\} = P\{N_{T+Z} - N_T = j + 1 - i\}$

$$= P\{N_Z = j + 1 - i\} = \begin{cases} q_{j+1-i}, & j \geq i-1 \\ 0, & j < i-1 \end{cases}$$

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queueing Theory: M/G/1 Queue

The MC X is irreducible and aperiodic. If

$$r = E[N_Z] = aE[Z] = ab \quad \text{Mean number of arrivals over a mean service time}$$

Then (intuitively based on queue evolution/growth, also rigorously proven)

- If $r > 1$ all states are transient
- If $r < 1$ all states are recurrent non-null.
- If $r = 1$ all states are recurrent null

Notation:

$r_k = 1 - q_0 - \dots - q_k$ (prob arrivals over a service time exceed k ; summing them we get r , next)

$$r = r_0 + r_1 + \dots = (q_1 + q_2 + q_3 + \dots) + (q_2 + q_3 + \dots) + (q_3 + \dots) + \dots$$

$= q_1 + 2q_2 + 3q_3 + \dots$ (this is the definition of the mean r of the distr of arrivals over a service time)

Proposition: The chain X is recurrent non-null aperiodic if and only if $r < 1$.

Proof: We need to show that

$$\pi = \pi \cdot P, \quad \pi \cdot 1 = 1$$

$$\left. \begin{aligned} \pi_0 &= \pi_0 q_0 + \pi_1 q_0 \\ \pi_1 &= \pi_0 q_1 + \pi_1 q_1 + \pi_2 q_0 \\ \pi_2 &= \pi_0 q_2 + \pi_1 q_2 + \pi_2 q_1 + \pi_3 q_0 \\ &\vdots \end{aligned} \right\} \Rightarrow \left. \begin{aligned} \pi_1 q_0 &= \pi_0 r_0 \\ \pi_2 q_0 &= \pi_0 r_1 + \pi_1 r_1 \\ \pi_3 q_0 &= \pi_0 r_2 + \pi_1 r_2 + \pi_2 r_1 \\ &\vdots \end{aligned} \right\}$$

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queueing Theory: M/G/1 Queue

Summing all equations ($q_0 = 1 - r_0$, $r = r_0 + r_1 + r_2 + \dots$)

$$(1 - r_0) \cdot \sum_{j=1}^{\infty} \pi_j = \pi_0 r + (r - r_0) \sum_{j=1}^{\infty} \pi_j$$

If $r < 1$, then we obtain
$$\sum_{j=1}^{\infty} \pi_j = \frac{r}{1-r} \pi_0 \Rightarrow \sum_{j=0}^{\infty} \pi_j = \frac{1}{1-r} \pi_0$$

The condition $\pi \cdot 1 = 1$ is satisfied with $\pi_0 = 1 - r$

Theorem: The limits $\pi(j) = \lim_{n \rightarrow \infty} P^n(i, j)$ exist $\forall j \in E$ and are independent of the initial state i .

- If $r \geq 1$, then $\pi(j) = 0$, $\forall j$.
- If $r < 1$, then

$$\pi(0) = 1 - r$$

$$\pi(1) = (1 - r) \frac{r_0}{q_0}$$

$$\vdots$$

$$\pi(j+1) = (1 - r) \sum_{k=1}^j \left(\frac{1}{q_0} \right)^{k+1} \sum_{\mathbf{a} \in S_{jk}} r_{a_1} r_{a_2} \dots r_{a_k}$$

where S_{jk} is the set of all k -tuples $\mathbf{a} = (a_1, \dots, a_k)$ of integers $a_i \geq 1$ with $a_1 + \dots + a_k = j$

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queueing Theory: G/M/1 Queue

G/M/1:

Arrival Process: **G**enerally-distributed arrival times

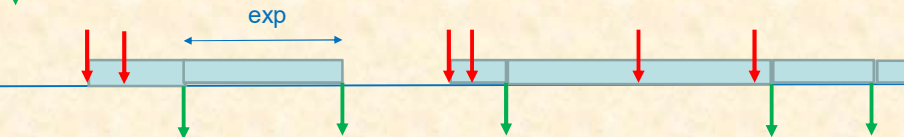
Service Process: **M**emoryless (exponential (geometric) service times)

Number of servers: **1**

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queuing Theory: $G/M/1$ Queue

arrival
departure (service completion time)



$X(t)$: Number of customers in the system (queue and under service)

Consider a specific subset of times $\{t_e\}$ only. That means that we embed $X(t)$ on times $\{t_e\}$. We do not look at $X(t)$ at times other than in $\{t_e\}$.

$X(t_e)$ is the process $X(t)$ embedded at times $\{t_e\}$.

$X(t)$ is not a MC.

If $\{t_e\} = \{\text{times of customer arrival}\}$, then $X(t_e)$ is a MC.

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΤΑ) - 2025

Applications to Queuing Theory: $G/M/1$ Queue

Exponentially distributed service times $\square \exp(a)$

i.i.d. interarrival times $\square \phi$.

In this case $q_n = \int_0^\infty \frac{e^{-at}(at)^n}{n!} d\phi(t)$ ← Distribution of services over an interarrival time

is the probability that the server completes exactly n services during an interarrival time (provided that there are that many customers).

Define: $r_n = q_{n+1} + q_{n+2} + \dots$

$$r = \sum_{n=1}^{\infty} nq_n = r_0 + r_1 + r_2 + \dots$$

r is the expected number of services which the server is capable of completing during an interarrival time. It can be proved that

- $r \geq 1$ Server can keep up with arrivals (recurrent)
- $r < 1$ Queue size increases to infinity (transient)

If X_n^e is the number of customers present in the system just before the time T_n of the n^{th} arrival, then

Theorem: $X^e = \{X_n^e; n \in N\}$ is a MC with $E = \{0, 1, 2, \dots\}$, $P^e =$

$$\begin{pmatrix} r_0 & q_0 & & & \\ r_1 & q_1 & q_0 & & \\ r_2 & q_2 & q_1 & q_0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΤΑ) - 2025

Applications to Queuing Theory: G/M/1 Queue

Proof: Let M_{n+1} be the number of services completed during the $n+1^{\text{th}}$ interarrival time $[T_n, T_{n+1})$. Then,

$$X_{n+1}^e = X_n^e + 1 - M_{n+1}$$

But M_{n+1} is conditionally independent of the past history before T_n given the present number X_n^e . If $Z = T_{n+1} - T_n$

$$P\{M_{n+1} = k \mid X_n^e, Z\} = \begin{cases} \frac{e^{-aZ} (aZ)^k}{k!} & X_n^e + 1 > k \\ \sum_{m=k}^{\infty} \frac{e^{-aZ} (aZ)^m}{m!} & X_n^e + 1 = k \quad (*) \\ 0 & \text{otherwise} \end{cases}$$

Taking expectations with respect to Z , which is independent of X_n^e , we obtain

$$P\{M_{n+1} = k \mid X_n^e = i\} = \begin{cases} q_k & k \leq i \\ r_{k-1} & k = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Equation $X_{n+1}^e = X_n^e + 1 - M_{n+1}$ and the previous one provide matrix P^e
M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queuing Theory: G/M/1 Queue

Theorem: X^e is recurrent non-null if and only if $r > 1$. If $r > 1$,

$$\pi^e(j) = \lim_{n \rightarrow \infty} P^n(i, j) = \lim_{n \rightarrow \infty} P^e\{X_n^e = j \mid X_0^e = i\}$$

and

$$\pi^e(j) = (1 - \beta)\beta^j, \quad j = 0, 1, 2, \dots$$

where β is the unique number satisfying

$$\beta = q_0 + q_1\beta + q_2\beta^2 + \dots$$

If $r \leq 1$ then $\pi^e(j) = 0$ for all j .

Proof: X^e is recurrent non-null if and only if

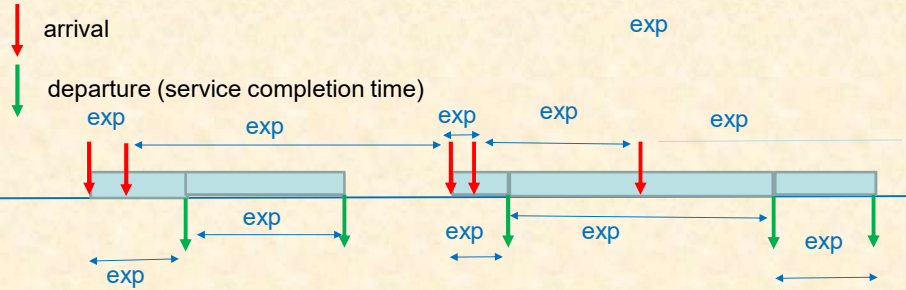
$$\nu = \nu \cdot P^e, \quad \nu \cdot 1 = 1$$

has a solution.

$$\begin{aligned} \nu_0 &= q_1\nu_0 + q_2\nu_0 + q_3\nu_0 + \dots \\ &\quad + q_2\nu_1 + q_3\nu_1 + \dots \\ &\quad + q_3\nu_2 + \dots \\ \nu_1 &= q_0\nu_0 + q_1\nu_1 + q_2\nu_2 + q_3\nu_3 + \dots \\ \nu_2 &= q_0\nu_1 + q_1\nu_2 + q_2\nu_3 + q_3\nu_4 + \dots \end{aligned}$$

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Applications to Queueing Theory: M/M/1 Queue



$X(t)$: Number of customers in the system (queue and under service)

$X(t)$ is a MC. Why?

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025

Special case M/M/1

We can consider this queue as a special case of M/G/1 or G/M/1. In the sequel we use G/M/1. Now the interarrival distribution is given by:

$$\phi(t) = 1 - e^{-\lambda t}, \quad t \geq 0$$

To compute the limiting distribution of X^e (queue size just before the n^{th} arrival, we find first β , where

$$\beta = \sum_{k=0}^{\infty} q_k \beta^k = \sum_{k=0}^{\infty} \beta^k \int_0^{\infty} \frac{e^{-at} (at)^k}{k!} \lambda e^{-\lambda t} dt = \int_0^{\infty} e^{-at(1-\beta)} \lambda e^{-\lambda t} dt = \frac{\lambda}{\lambda + a - a\beta}$$

The previous equation becomes $\beta = \frac{\lambda}{\lambda + a - a\beta}$ or $(1-\beta)(\lambda - a\beta) = 0$

with solutions $\beta = 1$ and $\beta = \frac{\lambda}{a}$. When $r = \frac{a}{\lambda} > 1$, the smallest solution is $\beta = \frac{\lambda}{a}$

So we have $\lim_{n \rightarrow \infty} P\{X_n^e = j\} = \left(1 - \frac{\lambda}{a}\right) \left(\frac{\lambda}{a}\right)^j, \quad j = 0, 1, \dots$

It turns out that $\lim_{t \rightarrow \infty} P\{Y_t = j\} = \left(1 - \frac{\lambda}{a}\right) \left(\frac{\lambda}{a}\right)^j, \quad j = 0, 1, \dots$ for the queue size Y_t at time t .

and $\lim_{n \rightarrow \infty} P\{X_n = j\} = \left(1 - \frac{\lambda}{a}\right) \left(\frac{\lambda}{a}\right)^j, \quad j = 0, 1, \dots$ for the queue size X_n just after the n^{th} departure.

M105 - Ανάλυση και Μοντελοποίηση Δικτύων - Ιωάννης Σταυρακάκης (ΕΚΠΑ) - 2025