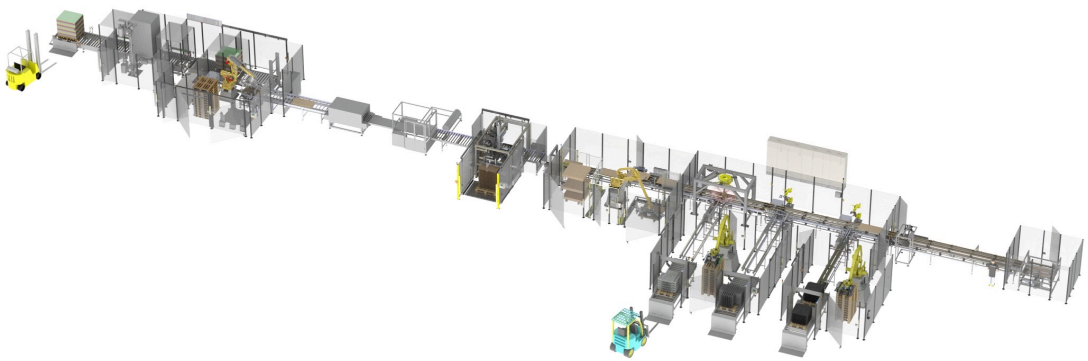


MANUFACTURING SYSTEMS AND SUPPLY CHAIN DYNAMICS

Chapter 4: Elements of Queueing Theory

EPFL, Master MT

Roger Filliger (BFH), Olivier Gallay (UniL)

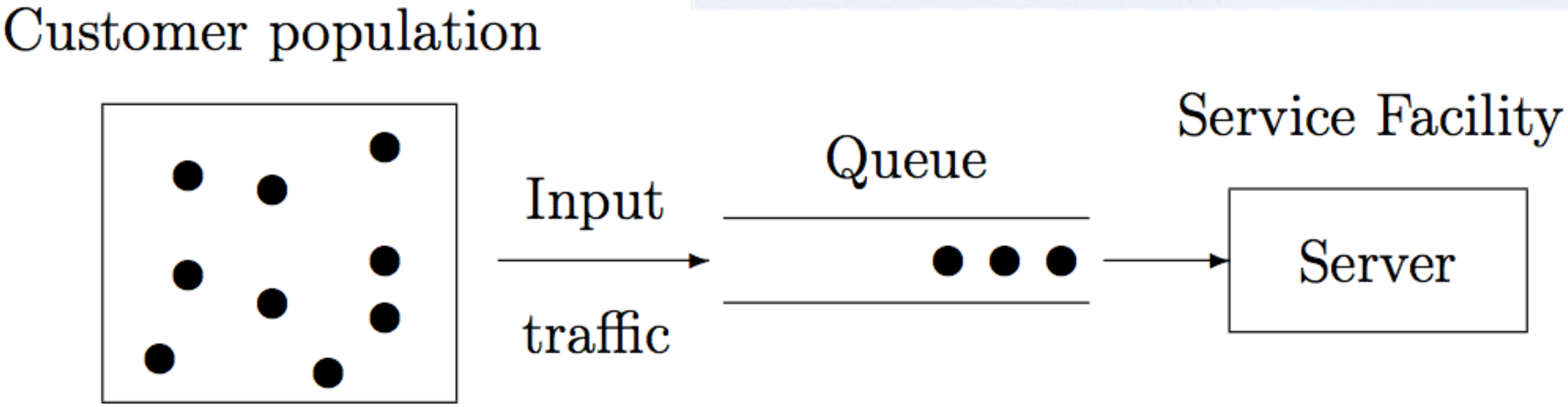
Course Content

1. *Introduction*
2. *Inventory Theory*
3. *Safety Stock in Manufacturing Systems*
4. ***Elements of Queueing Theory***
5. *Productions Flows*
6. *Production Dipole*
7. *Production Lines and Aggregation*
8. *Cooperative Flow Dynamics*
9. *Introduction to Queueing Networks*
10. *Supply Chain Analysis*
11. *Elements of Reliability Analysis*
12. *Maintenance Policies*

Queueing Systems

Natural theoretical framework to model manufacturing systems

- server -> workstation
- queue -> inventory
- customers -> items

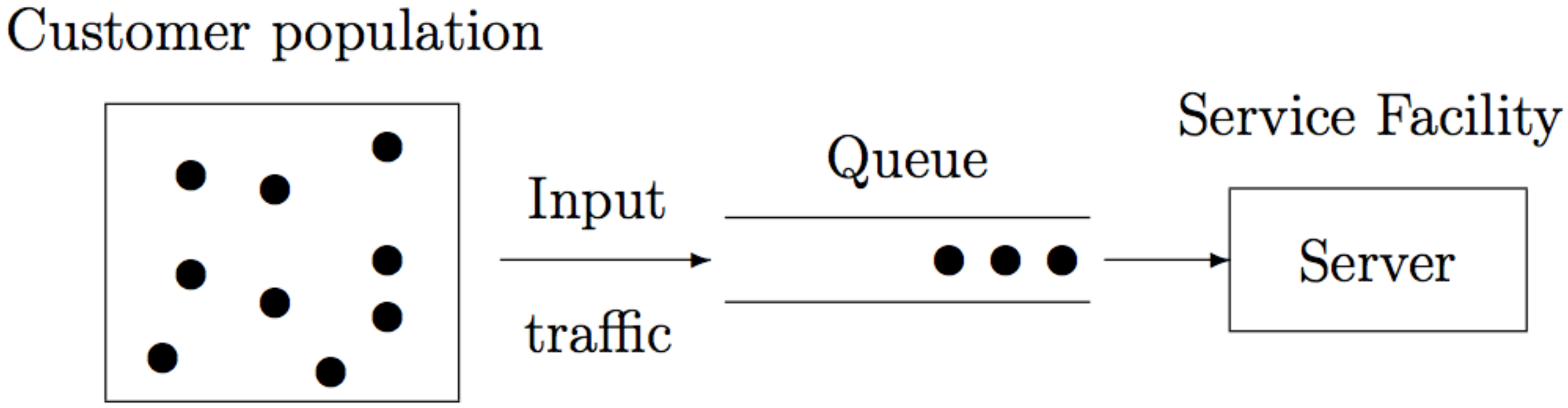


Queueing Systems

Goal: finding stationary distribution of the number of jobs in the queue

Important quantities for manufacturing systems:

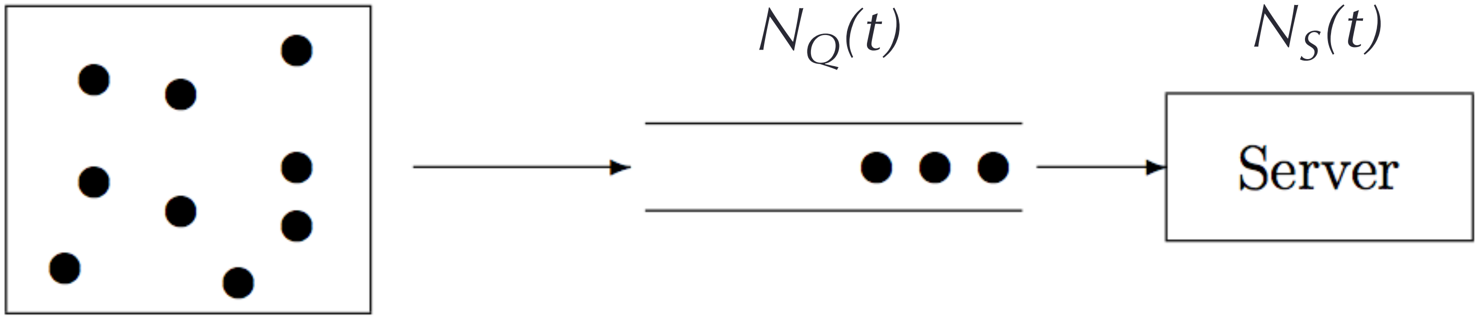
- mean time an item spends in a buffer
- mean number of items in a buffer



Queueing Systems

Modeling:

- $N_Q(t)$: **number of customers queueing** for service at time t
- $N_S(t)$: **number of customers receiving service** at time t
- N_t: **total number of customers** in the system at time t
- $N_Q(t) + N_S(t) = N_t$

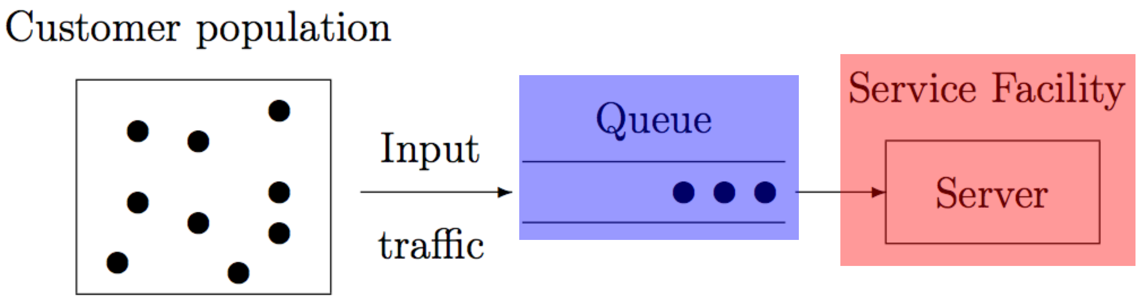


Queueing Systems

Modeling:

Y_n : **interarrival time between item customers *n* and *n + 1*,** S_n **service time of customer *n*** , **i.i.d.** with:

- Distribution function $F_A(t) := P(Y_n \le t) = P(Y_1 \le t)$ $F_S(t)$
- Probability density function $f_A(t) := \dfrac{dF_A(t)}{dt}$ $f_S(t)$
- Expectation $E(Y_1) = \int_0^\infty t f_A(t)dt = \dfrac{1}{\lambda}$ $E(S) = \int_0^\infty t f_S(t)dt = \dfrac{1}{\mu}$
- λ: **mean arrival rate** of items/customers into the system
- **μ: mean service rate** for an item/customer entering service



Characterization with Kendall Notation

Characterization for queueing systems: ***A/S/m/N – SD***

- *A* : **distribution of the interrarrival times** (item arrivals feeding a manufacturing system) -> *M* for Poisson, *D* for deterministic, *G* for general, etc.
- ***S*** : **distribution of the service times** (processing time of a workstation in a manufacturing system) -> *M, D, G,* etc.
- ***m*** : **number of servers** (workstations in a manufacturing system)
- *N* : **capacity** of the waiting room (of the buffer in a manufacturing system)
- ***SD*** : **service discipline** (order the items are taken from the buffer in a manufacturing system) -> *FIFO, LIFO,* etc. If ommited: *FIFO*.

Examples: ***M/M/1/∞***, ***M/G/1/∞ - LIFO***

Little's Law

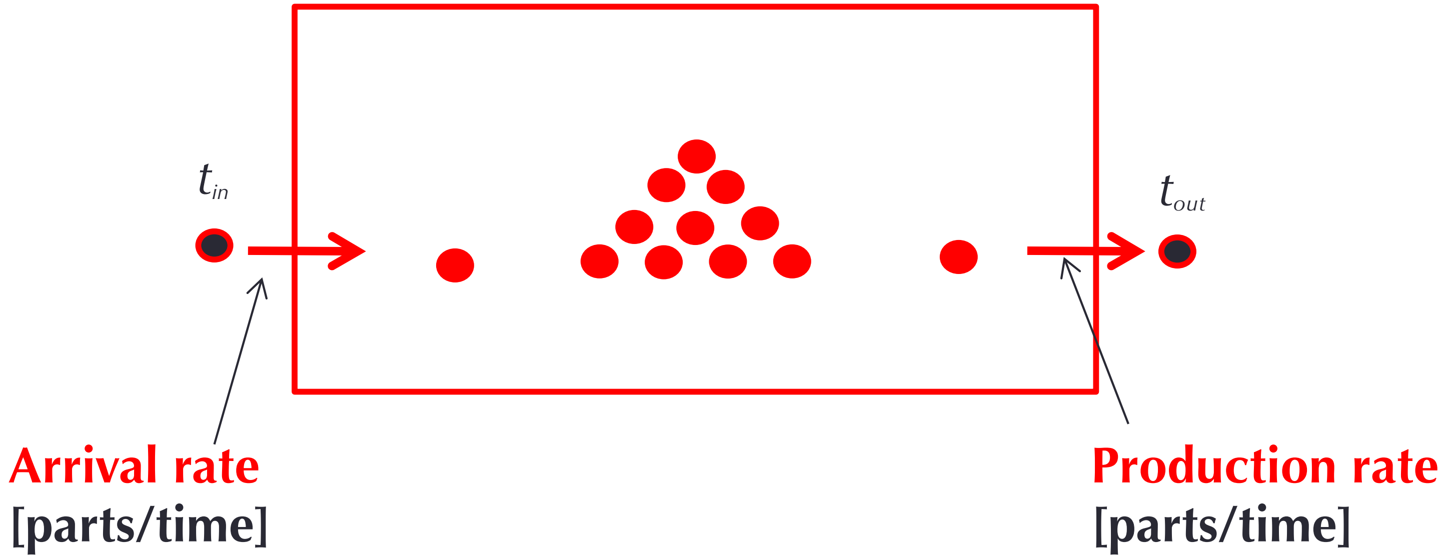
Valid for **mean values** (when mean values can be calculated)

L: **mean number of customers in the system**

W: **mean time spent in the system**

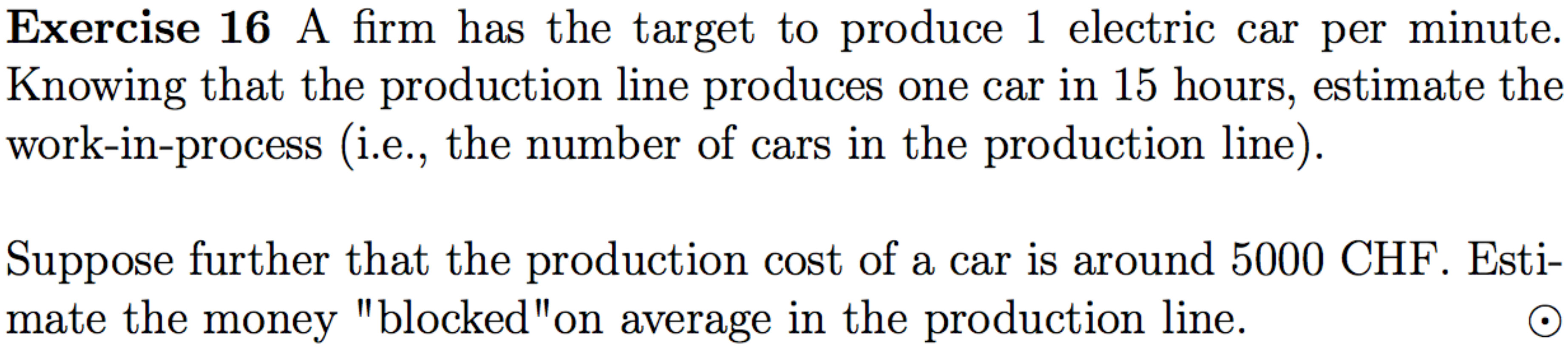
λ: **production rate**

$$\boxed{L = \lambda W}$$



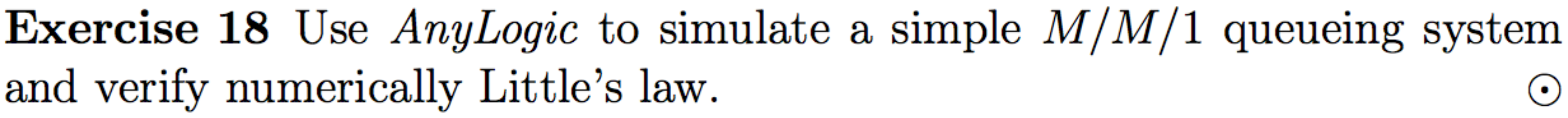
Application: a car factory produces 40cars/h. 1 car is produced in 25h.
-> Estimated inventory is 1000 (partially finished) cars.

Little's Law



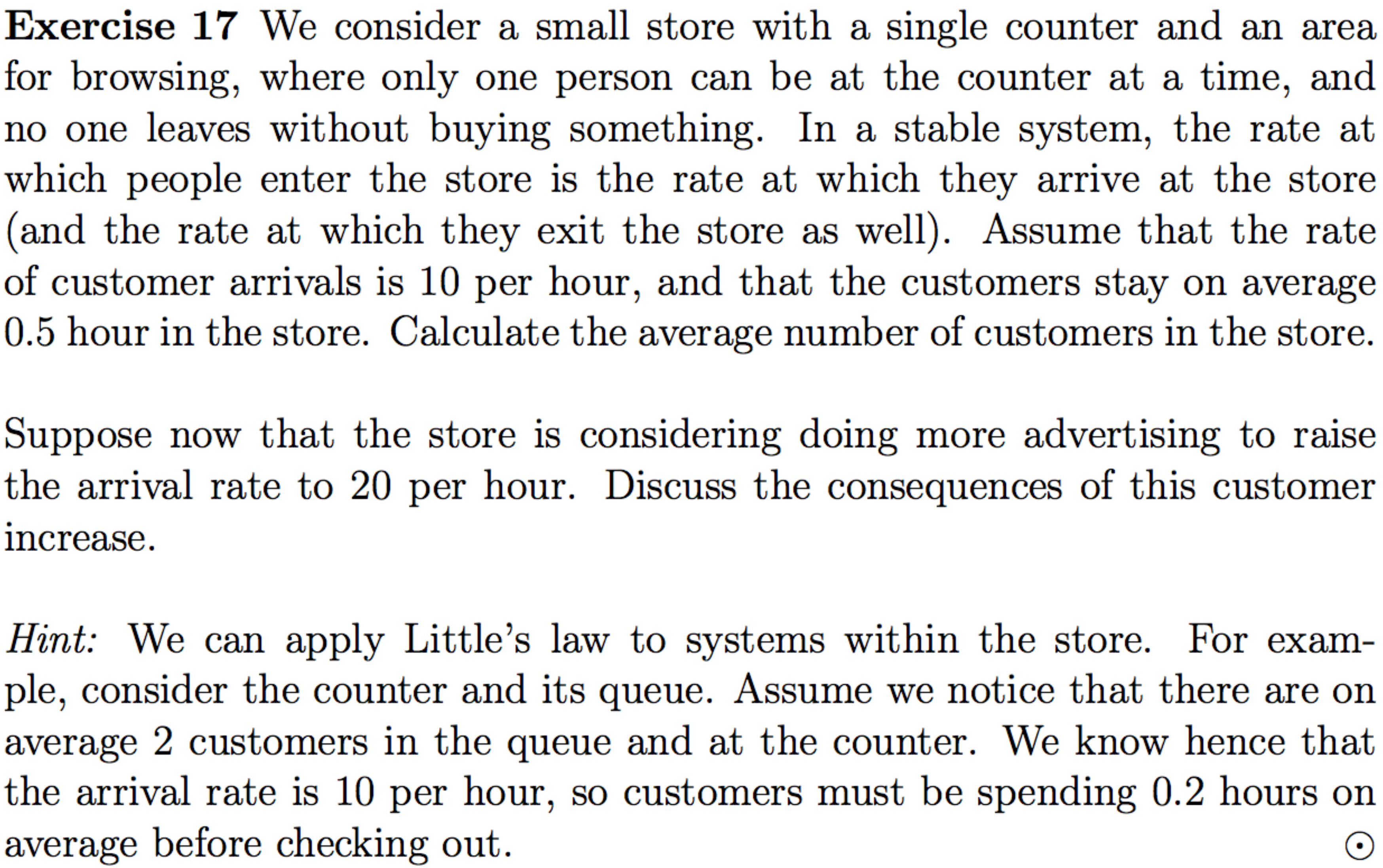
Exercise 16 A firm has the target to produce 1 electric car per minute. Knowing that the production line produces one car in 15 hours, estimate the work-in-process (i.e., the number of cars in the production line).

Suppose further that the production cost of a car is around 5000 CHF. Estimate the money "blocked"on average in the production line. ⊙



Exercise 18 Use *AnyLogic* to simulate a simple $M/M/1$ queueing system and verify numerically Little's law. ⊙

Little's Law



Exercise 17 We consider a small store with a single counter and an area for browsing, where only one person can be at the counter at a time, and no one leaves without buying something. In a stable system, the rate at which people enter the store is the rate at which they arrive at the store (and the rate at which they exit the store as well). Assume that the rate of customer arrivals is 10 per hour, and that the customers stay on average 0.5 hour in the store. Calculate the average number of customers in the store.

Suppose now that the store is considering doing more advertising to raise the arrival rate to 20 per hour. Discuss the consequences of this customer increase.

Hint: We can apply Little's law to systems within the store. For example, consider the counter and its queue. Assume we notice that there are on average 2 customers in the queue and at the counter. We know hence that the arrival rate is 10 per hour, so customers must be spending 0.2 hours on average before checking out. ⊙

M/M/1/∞ - FIFO

A : arrivals occur according to a Poisson process with rate λ (interrarrival time is **exponential** with mean 1/λ)

M : service time is **exponential** with mean 1/μ

m and ***N*** : **single buffer of infinite size**, positioned in front of a **single workstation**

Traffic intensity :ρ **=** $\boldsymbol{\lambda}$/μ (key performance indicator):

- $\rho < 1$: **stable** queue
- $\rho \geq 1$: **unstable** queue

Objective: finding the **stationary distribution** of the queueing system

M/M/1/∞ - FIFO : Stationary Distribution

Finding the stationary distribution:

- Interpret the number N_t of customers/items in the system as a **Markov chain with state space {0, 1, 2, ...}**

- Write down the **balance equations** :

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1 \\ (\lambda+\mu)\pi_n &= \lambda\pi_{n-1} + \mu\pi_{n+1}, \quad n = 1, 2, 3, \ldots \end{aligned}$$

π_i : stationary (in the long-run) probability that the queueing system is in state *i (i* customers in the system)

- Solving the balance equations with $\sum_{n=0}^{\infty} \pi_n = 1$, we obtain:

$$\pi_n = (1-\rho)\rho^n, \quad n = 0, 1, 2, \ldots$$

M/M/1/∞ - FIFO : Results

Mean number of customers in the system *L* = $\sum_{n=0}^{\infty} n\pi_n = \ldots = \frac{\lambda}{\mu-\lambda}$

Mean number of customers waiting for service *L_q* :

$$L_q = \sum_{n=1}^{\infty}(n-1)\pi_n = \ldots = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

Time spent in the system *W* = $\frac{1}{\lambda}L = \frac{1}{\mu-\lambda}$ (using Little's Law)

Time spent in the queue *W_q* = $\frac{1}{\lambda}L_q = \frac{\lambda}{\mu(\mu-\lambda)}$ (using Little's Law)

Distribution of the waiting time in the queue *$F_q(t)$*:

$$F_q(t) = P(T_q \leq t) = (1-\rho)e^{-\mu(1-\rho)t}.$$

Busy period *B* : $E(B) = \frac{1}{\mu-\lambda}$

Departure process: Poisson(**λ**)

M/M/1/∞ - FIFO : Example

Example 5 Jobs arrive at a (Poisson) rate of $15/h$ and it is assumed that (exponential) service takes 3 minutes on average. Under the $M/M/1/\infty$ assumption, we determine the following stationary **performance measures**:

(i) System utilization ρ: arrival rate $\lambda = 15/h$, service rate $\mu = 20/h$ therefore $\rho = \frac{3}{4}$.

(ii) Expected number of jobs waiting in the queue:

$$L_q = \frac{\rho^2}{1-\rho} = 2.25$$

(iii) Expected waiting time for service:

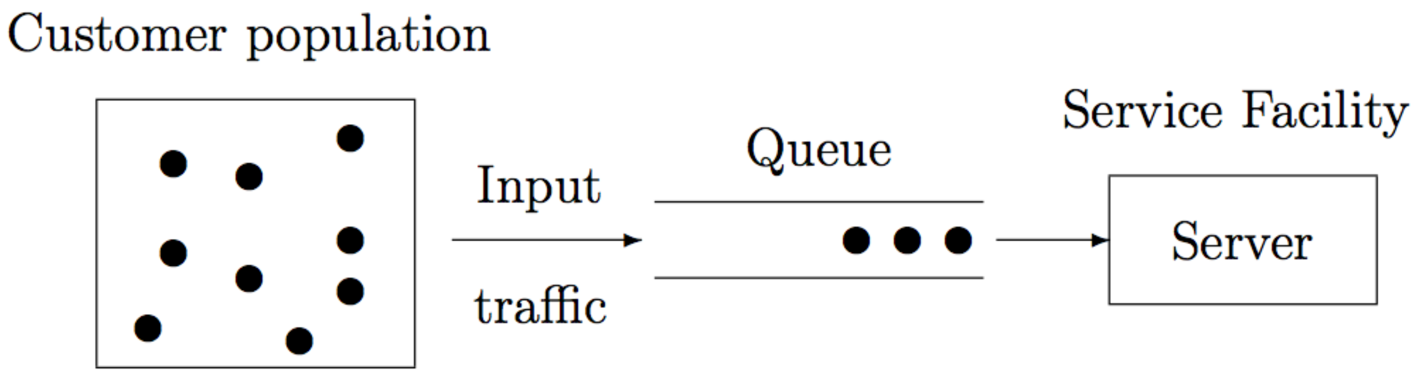
$$W_q = \frac{L_q}{\lambda} = \frac{9}{60} hour = 9 min$$

(iv) Probability that a job has to wait more than $5min$ (using Eq. (4.5), $P(T_q > t) = \rho e^{-\mu(1-\rho)t}$):

$$P(T_q > 5min) = \frac{3}{4} e^{-20(1-\frac{3}{4})5/60} = 0.494$$

M/M/1/∞ - FIFO : Derivation of the Results

Exercise 14 Queueing with a Single Server. We consider an $M/M/1/\infty$ queueing system with a single server, in which the customers (or items) arrive according to a Poisson process with rate λ, and the service times (or machine processing times) are independent exponential random variables, with mean equal to $1/\mu$. We suppose that the system capacity is infinite, as well as the population from which the customer or items come. Finally, the queue discipline is that of First-In First-Out (FIFO). The goal is to establish yourself the results stated in the above example.



M/M/1/∞ - FIFO : Derivation of the Results

Show that the number of clients N_t in the queueing system at time t has stationary probabilities π_n given by

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) \text{ for all } n \geq 0,$$

where we have assumed that $\lambda < \mu$. In particular:

- Write down the balance equations for this Markov chain and verify the above given stationary solution.

- Show that $L = E(X_\infty)$: the average number of customers in the system (at equilibrium) is given by:

$$L = \frac{\lambda}{\mu - \lambda}.$$

Note the explosion of this number if λ is close to μ!

M/M/1/∞ - FIFO : Derivation of the Results

(i) The probability that the server is busy is a performance measure for the queueing system. Convince yourself that this utilization factor is $1 - \pi_0 = \frac{\lambda}{\mu}\rho$.

(ii) If we denote by $L_q = E(X_\infty^Q)$ the mean number of customers waiting in the queue (excluding the one in service), show that

$$L_q = \sum_{n=1}^{\infty}(n-1)\pi_n = \ldots = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

Note that the average queue length L_q is not $E(X_\infty) - 1$ but $E(X_\infty) - \frac{\lambda}{\mu}$ because the server is not always busy.

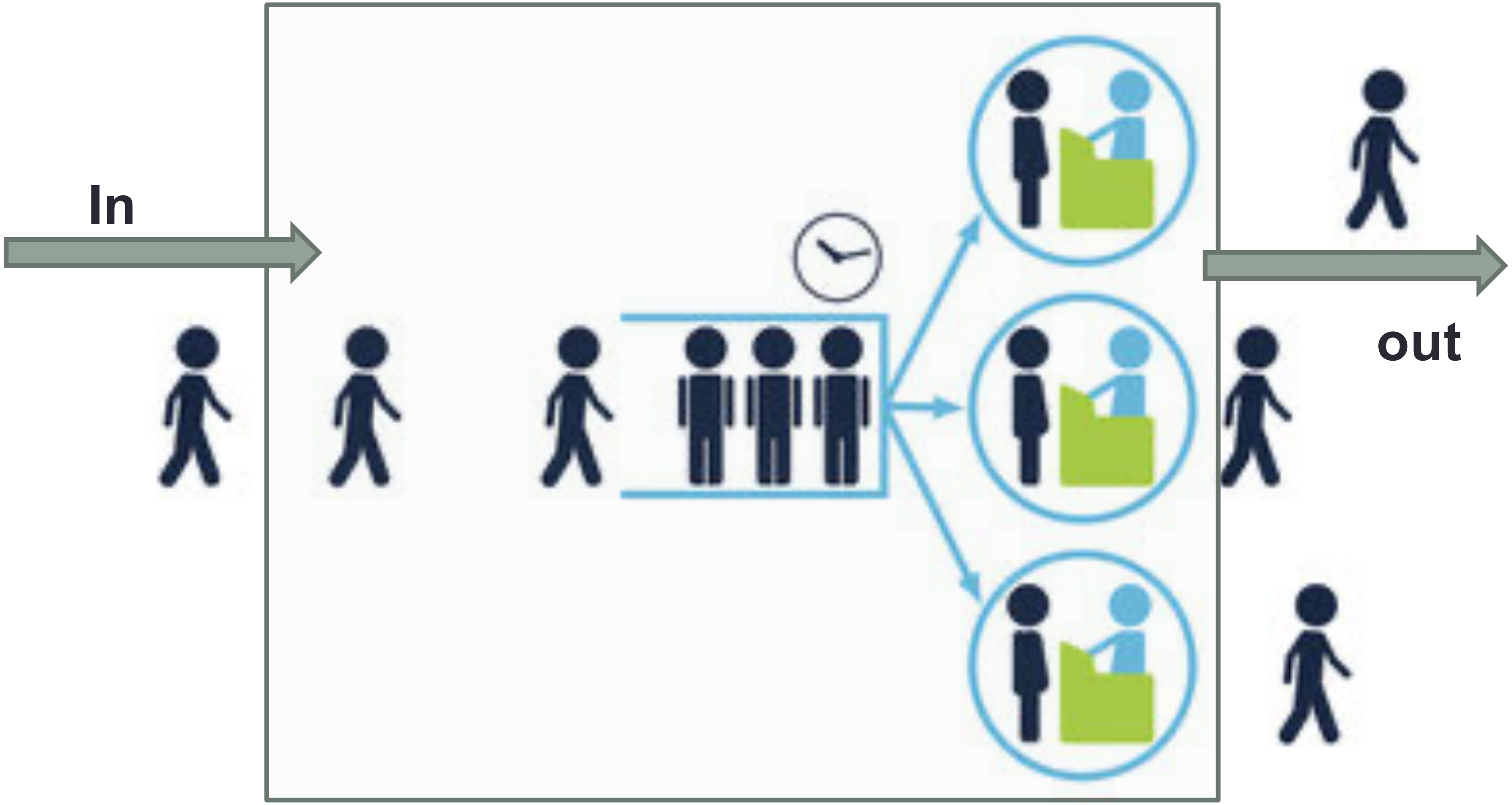
(iii) Use your preferred mathematical software to compute the (stationary) variance of customers in the system (answer: $V(X_\infty) = \frac{\lambda\mu}{(\mu-\lambda)^2}$)

⊙

M/M/1/∞ - FIFO : AnyLogic

Exercice: Simulate the behavior of the *M/M/1/∞* queueing system under both *FIFO* and *LIFO* service disciplines, and compare the distribution of the waiting time in the queue. Observe in particular the variance of the waiting time in the queue.

M/M/s/K - FIFO



M/M/s/K - FIFO

Finite queue : **capacity limit *K*** (arrivals denied when $N_t > K$)

Several workstations: ***s* servers** ($K > s$)

Arrival rate: λ

Service rate: *s*μ

The queueing system is **stable** if ρ = **λ**/*s*μ < 1

$$
\begin{aligned}
\lambda_n &= \lambda \quad n = 0, 1, 2, ..., K-1 \\
\mu_n &= n\mu \quad n = 1, 2, ..., s-1 \\
&= s\mu \quad n = s, s+1, ..., K.
\end{aligned}
$$

arrival rate when *n* customer in the queue

service rate when *n* customer in the queue

M/M/s/K – FIFO : Results

Stationary distribution:

$$\pi_0 = \Big[\sum_{r=0}^{s-1} \frac{\alpha^t}{r!} + \frac{\alpha^s}{s!} \sum_{r=s}^{K} \rho^{r-s}\Big]^{-1} \qquad \alpha = \frac{\lambda}{\mu}$$

$$\pi_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \pi_0, \quad n = 1, 2, 3, ..., s$$

$$\pi_n = \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s \left(\frac{\lambda}{s\mu}\right)^{n-s} \pi_0, \quad n = s, 2, 3, ..., K$$

Distribution of the waiting time in the queue *$F_q(t)$*:

$$F_q(t) = 1 - \frac{1}{1-\pi_K} \sum_{n=s}^{K-1} \pi_n \sum_{r=0}^{n-s} e^{s\mu t} \frac{(s\mu t)^r}{r!}$$

Time spent in the queue *W_q* = $\displaystyle \frac{1}{s\mu(1-\pi_K)} \sum_{n=s}^{K-1} (n-s+1)\pi_n$

Time spent in the system *W* = $\displaystyle W_q + \frac{1}{\mu}$

M/M/s/K – FIFO : Results

Mean number of customers in the system *L* = $\lambda(1-\pi_K)W$

Mean number of customers waiting for service L_q $= \lambda(1-\pi_K)W_q$

M/M/1/3 – FIFO : Example

Example 6 A small post office has one telephone line and a facility for call waiting for two additional customers. Orders arrive at the rate of 1 per minute and each order requires – on average – 2 minutes and 30 seconds. Model this system as an $M/M/1/3$ queue, and answer the following questions (FIFO service discipline is assumed):

(i) What is the expected number of calls waiting in the queue? What is the mean waiting time in the queue?

(i) Assuming that the arrivals form a Poisson process with rate 1 per minute and the service times are exponential with mean 2.5 minutes, we have $\rho = 2.5$. Also, $K = 3$. Using the above results we get with $s = 1$:

$$L_q \approx 1.48$$

and since $\lambda = 1$, the mean waiting time in queue is $W_q = 3.73\text{min}$.

M/M/1/3 – FIFO : Example

(ii) What is the probability that the call has to wait for more than 1.5 minutes before getting served?

(ii) We use the formula for $1 - F_q(t)$ with $t = 1.5$, $1/\mu = 2.5$ and $\rho = 2.5$. We get:

$$P(\text{wait in queue } > 1.5 \text{ min}) = 0.7$$

M/M/1/3 – FIFO : Example

(iii) Because of the excessive waiting time of customers, the business decides to use two telephone lines instead of one, keeping the same total capacity for the number in the system (namely, 3). What improvements result in the performance measures considered under (i) and (ii)?

(iii) With two lines, we have a $M/M/2/3$ queueing system. Accordingly, with $\alpha = 2.5$, $\rho = 1.25$, $s = 2$ and $K = 3$, we get:

$$\begin{array}{ll} p_0 = 0.0950 & p_1 = 0.2374 \\ p_2 = 0.2969 & p_3 = 0.3711 \end{array}$$

and, therefore, using the corresponding formulas for W_q and L_q, we obtain:

$$\begin{aligned} W_q &= 0.59\text{min} \\ L_q &= \lambda(1 - p_3)W_q = 0.371 \end{aligned}$$

and $P(\text{wait in queue} > 1.5\text{min}) = 1 - F_q(1.5) = 0.142$.

M/M/1/3 – FIFO : Example

(iv) What is the impact of increasing the capacity to four customers in the system?

(iv) Now, we have a $M/M/2/4$ queue. Using the formulas as above, we get:

$$\begin{array}{lll} p_0 = 0.0649 & p_1 = 0.1622 & \\ p_2 = 0.2028 & p_3 = 0.2535 & p_4 = 0.3169 \end{array}$$

and, therefore, using the corresponding formulas for W_q and L_q, we obtain:

$$\begin{aligned} W_q &= 1.30\text{min} \\ L_q &= \lambda(1 - p_4)W_q = 0.887 \end{aligned}$$

and $P(\text{wait in queue} > 1.5\text{min}) = 1 - F_q(1.5) = 0.3353$.

It is instructive to note that the performance has not improved from the viewpoint of customers! This is because the system now accepts more customers than before. From the management perspective, fewer customers are being denied access to the system ($p_4 = 0.317$ versus $p_3 = 0.371$). $\Diamond$

M/M/1/∞ - FIFO vs *M/M/2/∞ - FIFO*: AnyLogic

Exercice: Simulate the behavior of the *M/M/1/∞* and the *M/M/2/∞* queueing systems under *FIFO* service discipline, and compare the distribution of the waiting time in the queue. Observe in particular the variance of waiting time in the queue. Focus on the situation where the singler server of the *M/M/1/∞* queueing system is twice as fast of the two servers of the *M/M/2/∞* queueing system.

M/M/s/K – FIFO : AnyLogic

Exercise 15 Simulate the behavior of the above queues $M/M/1/3$, $M/M/2/3$ and $M/M/2/4$ using *AnyLogic* , and compare the findings. $\odot$