MIT 2.853/2.854

Introduction to Manufacturing Systems

Markov Processes and Queues

Stanley B. Gershwin
Laboratory for Manufacturing and Productivity
Massachusetts Institute of Technology
gershwin@mit.edu

Stochastic processes

- t is time.
- $X()$ is a *stochastic process* if $X(t)$ is a random variable for every t.
- t is a scalar — it can be discrete or continuous.
- $X(t)$ can be discrete or continuous, scalar or vector.

Stochastic processes

Markov processes

- A *Markov process* is a stochastic process in which the probability of finding X at some value at time $t + \delta t$ depends only on the value of X at time t.

- Or, let $x(s), s \leq t$, be the history of the values of X before time t and let A be a possible value of X.

$$P\{X(t + \delta t) = A | X(s) = x(s), s \leq t\} =$$
$$P\{X(t + \delta t) = A | X(t) = x(t)\}$$

Stochastic processes
Markov processes

- In words: if we know what X was at time t, we don't gain any more useful information about $X(t+\delta t)$ by *also* knowing what X was at any time earlier than t.

- *This is* ***ONLY*** *the definition of a class of mathematical models. It is NOT a statement about reality!!* That is, not everything is a Markov process.

Markov processes
Example

Example:

- I have \$100 at time $t=0$.
- At every time $t \geq 1$, I have $\$N(t)$.
  - $\star$ A (possibly biased) coin is flipped.
  - $\star$ If it lands with H showing, $N(t+1) = N(t) + 1$.
  - $\star$ If it lands with T showing, $N(t+1) = N(t) - 1$.

$N(t)$ is a Markov process. *Why?*

Discrete state, discrete time

States and transitions

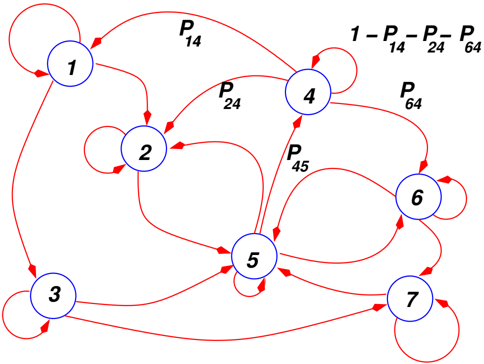
- States can be numbered 0, 1, 2, 3, ... *(or with multiple indices if that is more convenient)*.
- Time can be numbered 0, 1, 2, 3, ... *(or 0, Δ, 2Δ, 3Δ, ... if more convenient)*.
- The probability of a transition from j to i in one time unit is often written P_{ij}, where

  $$P_{ij} = P\{X(t+1) = i | X(t) = j\}$$

States and transitions

Transition graph

Transition graph

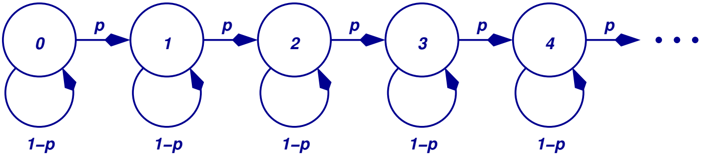


P_{ij} is a probability. Note that $P_{ii} = 1 - \sum_{m,m \neq i} P_{mi}$. This is the *self-loop* probability.

States and transitions
Transition graph

Example : $H(t)$ is the number of Hs after t coin flips.

Assume probability of H is p.

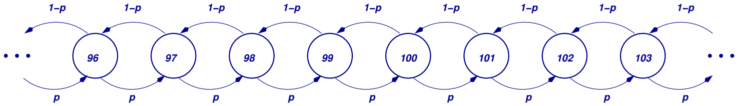


This is a system with an infinite state space.

States and transitions
Transition graph

Example : Coin flip bets on Slide 5.

Assume probability of H is p.



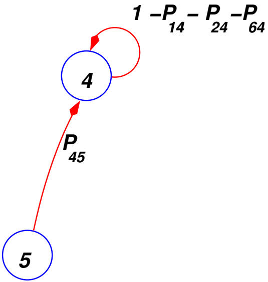
Markov processes

Notation

- $\{X(t) = i\}$ is the event that random quantity $X(t)$ has value i.
  - ⋆ *Example:* $X(t)$ is any state in the graph on Slide 7. i is a *particular* state.
- Define $\pi_i(t) = P\{X(t) = i\}$.
- Normalization equation: $\sum_i \pi_i(t) = 1$.

Markov processes
Transition equations

Transition equations: application of the law of total probability.



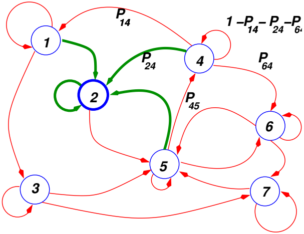
(Detail of graph on slide 7.)

$$\pi_4(t+1) = \pi_5(t)P_{45} + \pi_4(t)(1 - P_{14} - P_{24} - P_{64})$$

(Remember that
$P_{45} = P\{X(t+1) = 4 | X(t) = 5\}$,
$P_{44} = P\{X(t+1) = 4 | X(t) = 4\} = 1 - P_{14} - P_{24} - P_{64}$*)*

Markov processes

Transition equations



$$
\begin{aligned}
&P\{X(t+1)=2\} \\
&\quad = P\{X(t+1)=2|X(t)=1\}P\{X(t)=1\} \\
&\quad +P\{X(t+1)=2|X(t)=2\}P\{X(t)=2\} \\
&\quad +P\{X(t+1)=2|X(t)=4\}P\{X(t)=4\} \\
&\quad +P\{X(t+1)=2|X(t)=5\}P\{X(t)=5\}
\end{aligned}
$$

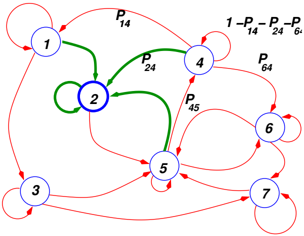
Markov processes
Transition equations

- Define $P_{ij} = P\{X(t+1) = i | X(t) = j\}$

- Transition equations: $\pi_i(t+1) = \sum_j P_{ij}\pi_j(t)$.
  An application of the *(Law of Total Probability)*

- Normalization equation: $\sum_i \pi_i(t) = 1$.

Markov processes
Transition equations



Therefore, since

$$P_{ij} = P\{X(t+1) = i | X(t) = j\} \text{ and}$$

$$\pi_i(t) = P\{X(t) = i\},$$

we can write

$$\pi_2(t+1) = P_{21}\pi_1(t) + P_{22}\pi_2(t) + P_{24}\pi_4(t) + P_{25}\pi_5(t).$$

Note that $P_{22} = 1 - P_{52}$.

Markov processes
Transition equations — Matrix-Vector Form

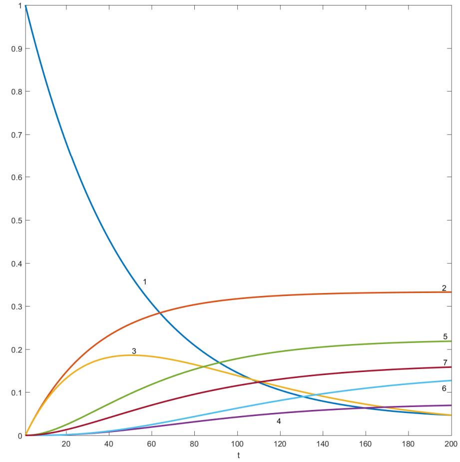
For an n-state system,

- Define

$$\pi(t) = \begin{bmatrix} \pi_1(t) \\ \pi_2(t) \\ \ldots \\ \pi_n(t) \end{bmatrix}, \quad P = \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1n} \\ P_{21} & P_{22} & \ldots & P_{2n} \\ & & \ldots & \\ P_{n1} & P_{n2} & \ldots & P_{nn} \end{bmatrix}, \quad \nu = \begin{bmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{bmatrix}$$

- Transition equations: $\pi(t+1) = P\pi(t)$
- Normalization equation: $\nu^T \pi(t) = 1$
- Other facts:
  - ⋆ $\nu^T P = \nu^T$ *(Each column of P sums to 1.)*
  - ⋆ $\pi(t) = P^t \pi(0)$

Markov processes
Steady state



◊ * ⊛

State probabilities vs. t for system in Slide 7

Markov processes

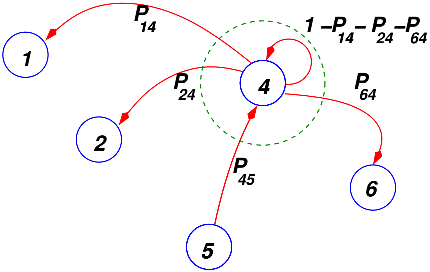
Steady state

* ⋆ ▲

- Steady state: $\pi_i = \lim_{t\to\infty} \pi_i(t) = \lim_{t\to\infty} P^t \pi(0)$, if it exists.
- Steady-state transition equations: $\pi_i = \sum_j P_{ij}\pi_j$.
- *Alternatively,* steady-state balance equations: $\pi_i \sum_{m,m\neq i} P_{mi} = \sum_{j,j\neq i} P_{ij}\pi_j$
- Normalization equation: $\sum_i \pi_i = 1$.

Markov processes

Balance equations



Balance equation:

$$(P_{14} + P_{24} + P_{64})\pi_4$$
$$= P_{45}\pi_5$$

in steady state *only*.

Intuitive meaning: The average number of transitions *into* the circle per unit time equals the average number of transitions *out of* the circle per unit time.

Markov processes
Steady state

How to calculate the steady-state probability distribution π

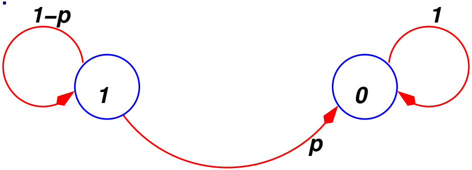
- Assume that the system has N states, where N is finite.
- Assume that there is a unique steady-state probability distribution.
- The transition equations form a set of N linear equations in N unknowns.
- The normalization equation is also a linear equation.
- *Problem?* We have $N+1$ equations in N unknowns.
- *No problem:* there is one redundant equation because each column sums to 1.
- Delete one transition equation and replace it with the normalization equation.
- Solve the system of N linear equations in N unknowns.

Markov processes
Steady state

- A system that has a unique steady-state solutions is called *ergodic* . The probability distribution approaches that limit no matter the initial probability distribution was.
- For systems that have more than one steady-state solution, the limiting distribution depends on the initial probability.
- The balance equations can be used to find the limiting distribution instead of the transition equations. As before, one equation has to be replaced by the normalization equation.
- If a system has an infinite number of states and it has a steady state probability distribution, there are two possibilties for finding it:
  - ⋆ It might be possible to solve the equations analytically. We will see an example of that.
  - ⋆ Truncate the system. That is, solve a system with a large but finite subset of the states. If you understand the system, you can guess which are the highest probability states. Keep those. This provides an approximate solution.

Markov processes
Geometric distribution

Consider a two-state system. The system can go from 1 to 0, but not from 0 to 1.



Let p be the conditional probability that the system is in state 0 at time $t+1$, given that it is in state 1 at time t. Then

$$p = P\left[\alpha(t+1) = 0 \middle| \alpha(t) = 1\right].$$

Markov processes

Geometric distribution — Transition equations

Let $\pi(\alpha, t)$ be the probability of being in state α at time t.

$$\begin{aligned} \pi(0, t+1) &= P\left[\alpha(t+1) = 0 \middle| \alpha(t) = 1\right] P[\alpha(t) = 1] \\ &+ P\left[\alpha(t+1) = 0 \middle| \alpha(t) = 0\right] P[\alpha(t) = 0], \end{aligned}$$

we have

$$\begin{aligned} \pi(0, t+1) &= p\pi(1, t) + \pi(0, t), \\ \pi(1, t+1) &= (1-p)\pi(1, t), \end{aligned}$$

and the normalization equation

$$\pi(1, t) + \pi(0, t) = 1.$$

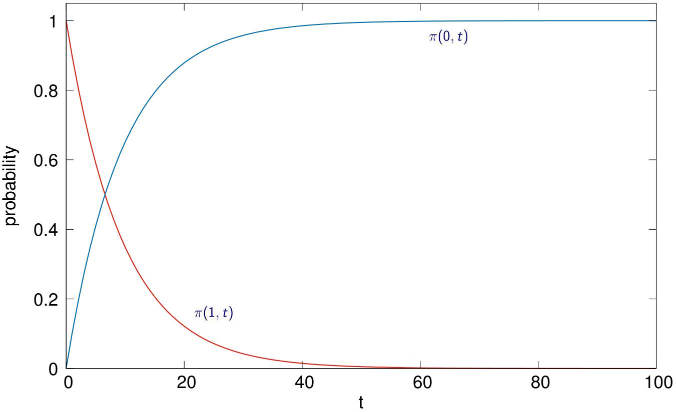
Markov processes
Geometric distribution — transient probability distribution

Assume that $\pi(1,0) = 1$. Then the solution is

$$\begin{aligned} \pi(0,t) &= 1 - (1-p)^t, \\ \pi(1,t) &= (1-p)^t. \end{aligned}$$

Markov processes

Geometric distribution — transient probability distribution



Markov processes
Geometric distribution

We have shown that the probability that the state goes from 1 to 0 at time t is

$$P(t) = (1-p)^{t-1}p$$

The mean time for the state to go from 1 to 0 is then

$$\bar{t} = \sum_{t=1}^{\infty} tP(t) = \sum_{t=1}^{\infty} t(1-p)^{t-1}p$$

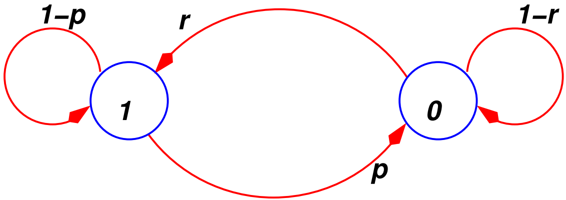
It is not hard to show that

$$\bar{t} = \frac{1}{p}$$

Markov processes

Unreliable machine

1=up; 0=down.



Mean up time = Mean time to fail = MTTF$= 1/p$

Mean down time = Mean time to repair = MTTR$= 1/r$

Markov processes

Unreliable machine — transient probability distribution

The probability distribution satisfies

$$\pi(0, t+1) = \pi(0, t)(1-r) + \pi(1, t)p,$$

$$\pi(1, t+1) = \pi(0, t)r + \pi(1, t)(1-p).$$

Markov processes

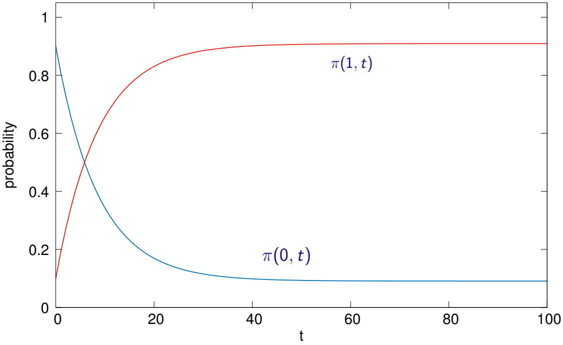
Unreliable machine — transient probability distribution

It is not hard to show that

$$\begin{aligned}\pi(0,t) &= \pi(0,0)(1-p-r)^t \\ &\quad + \frac{p}{r+p}\left[1-(1-p-r)^t\right],\end{aligned}$$

$$\begin{aligned}\pi(1,t) &= \pi(1,0)(1-p-r)^t \\ &\quad + \frac{r}{r+p}\left[1-(1-p-r)^t\right].\end{aligned}$$

Markov processes
Unreliable machine — transient probability distribution



Markov processes

Unreliable machine — steady-state probability distribution

As $t \to \infty$,

$$\begin{aligned} \pi(0,t) &\to \frac{p}{r+p}, \\ \pi(1,t) &\to \frac{r}{r+p} \end{aligned}$$

which is the solution of

$$\begin{aligned} \pi(0) &= \pi(0)(1-r) + \pi(1)p, \\ \pi(1) &= \pi(0)r + \pi(1)(1-p). \end{aligned}$$

Markov processes

Unreliable machine — efficiency

If a machine makes one part per time unit when it is operational, its average production rate is

$$\pi(1) = \frac{r}{r+p}$$

This quantity is the *efficiency* (e) of the machine. If the machine makes one part per τ time units when it is operational, its average production rate is

$$P = \frac{1}{\tau}\left(\frac{r}{r+p}\right)$$

Note that we can also write

$$e = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}$$

Discrete state, continuous time

States and transitions

- States can be numbered 0, 1, 2, 3, ... *(or with multiple indices if that is more convenient)*.
- Time is a real number, defined on $(-\infty, \infty)$ or a smaller interval.
- The probability of a transition from j to i during $[t, t + \delta t]$ is approximately $\lambda_{ij}\delta t$, where δt is small, and

  $$\lambda_{ij}\delta t \approx P\{X(t + \delta t) = i | X(t) = j\} \text{ for } i \neq j$$

Discrete state, continuous time

States and transitions

More precisely,

$$\lambda_{ij}\delta t = P\{X(t+\delta t) = i | X(t) = j\} + o(\delta t)$$
$$\text{for } i \neq j$$

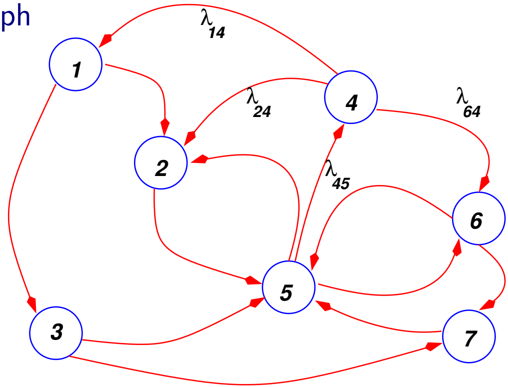
where $o(\delta t)$ is a function that satisfies $\lim_{\delta t \to 0} \frac{o(\delta t)}{\delta t} = 0$

This implies that for small δt, $o(\delta t) \ll \delta t$.

Discrete state, continuous time

States and transitions

Transition graph



λ_{ij} is a probability *rate.* $\lambda_{ij}\delta t$ is a probability.

Compare with the discrete-time graph.

* □

Discrete state, continuous time

States and transitions

One of the transition equations:

Define $\pi_i(t) = P\{X(t) = i\}$.

$$\pi_5(t + \delta t) \approx$$

$$(1 - \lambda_{25}\delta t - \lambda_{45}\delta t - \lambda_{65}\delta t)\pi_5(t) +$$

$$\lambda_{52}\delta t \pi_2(t) + \lambda_{53}\delta t \pi_3(t) + \lambda_{56}\delta t \pi_6(t) + \lambda_{57}\delta t \pi_7(t) +$$

Discrete state, continuous time

States and transitions

Or,

$$\pi_5(t + \delta t) \approx$$

$$\pi_5(t) - (\lambda_{25} + \lambda_{45} + \lambda_{65})\pi_5(t)\delta t$$

$$+(\lambda_{52}\pi_2(t) + \lambda_{53}\pi_3(t) + \lambda_{56}\pi_6(t) + \lambda_{57}\pi_7(t))\delta t$$

Discrete state, continuous time

States and transitions

Or,

$$\lim_{\delta t \to 0} \frac{\pi_5(t + \delta t) - \pi_5(t)}{\delta t} =$$

$$\frac{d\pi_5}{dt}(t) = -(\lambda_{25} + \lambda_{45} + \lambda_{65})\pi_5(t)$$

$$+\lambda_{52}\pi_2(t) + \lambda_{53}\pi_3(t) + \lambda_{56}\pi_6(t) + \lambda_{57}\pi_7(t)$$

Discrete state, continuous time

States and transitions

Define $\boxed{\textit{for convenience}}$

$$\lambda_{55} = -(\lambda_{25} + \lambda_{45} + \lambda_{65})$$

Then

$$\frac{d\pi_5}{dt}(t) = \lambda_{55}\pi_5(t) +$$

$$\lambda_{52}\pi_2(t) + \lambda_{53}\pi_3(t) + \lambda_{56}\pi_6(t) + \lambda_{57}\pi_7(t)$$

Discrete state, continuous time

States and transitions

- Define $\pi_i(t) = P\{X(t) = i\}$
- It is $\boxed{\textit{convenient}}$ to define $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$ $***$
- Transition equations: $\dfrac{d\pi_i(t)}{dt} = \sum_j \lambda_{ij}\pi_j(t)$.
- Normalization equation: $\sum_i \pi_i(t) = 1$.

$***$ *Often confusing!!!*

Discrete state, continuous time

Transition equations — Matrix-Vector Form

- Define $\pi(t), \nu$ as before *.

  Define $\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2n} \\ & & \dots & \\ \lambda_{n1} & \lambda_{n2} & \dots & \lambda_{nn} \end{bmatrix}$

- Transition equations: $\dfrac{d\pi(t)}{dt} = \Lambda\pi(t)$.

- Normalization equation: $\nu^T\pi = 1$.

- Other facts:

  - $\nu^T P = 0$ *(Each column of P sums to 0.)*
  - $\pi(t) = e^{\Lambda t}\pi(0)$

Discrete state, continuous time

Steady State

- *Steady state:* $\pi_i = \lim_{t \to \infty} \pi_i(t)$, if it exists.

- Steady-state transition equations: $0 = \sum_j \lambda_{ij} \pi_j$.

- *Alternatively,* steady-state balance equations:
  $\pi_i \sum_{m, m \neq i} \lambda_{mi} = \sum_{j, j \neq i} \lambda_{ij} \pi_j$

- Normalization equation: $\sum_i \pi_i = 1$.

Discrete state, continuous time

Steady State — Matrix-Vector Form

- *Steady state:* $\pi = \lim_{t \to \infty} \pi(t)$, if it exists.
- Steady-state transition equations: $0 = \Lambda \pi$.
- Normalization equation: $\nu^T \pi = 1$.

Discrete state, continuous time

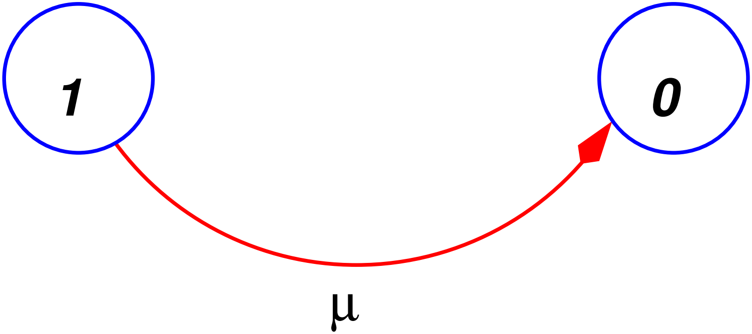
Sources of confusion in continuous time models

- $\boxed{\textit{Never}}$ Draw self-loops in continuous time Markov process graphs.
- $\boxed{\textit{Never}}$ write $1 - \lambda_{14} - \lambda_{24} - \lambda_{64}$. Write
  - $1 - (\lambda_{14} + \lambda_{24} + \lambda_{64})\delta t$, or
  - $-(\lambda_{14} + \lambda_{24} + \lambda_{64})$
- $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$ is $\boxed{\textit{NOT}}$ a rate and $\boxed{\textit{NOT}}$ a probability. It is $\boxed{\textit{ONLY}}$ a convenient notation.

Discrete state, continuous time

Exponential distribution

Exponential random variable T: the time to move from state 1 to state 0.



Discrete state, continuous time

Exponential distribution

$\pi(0, t+\delta t) =$

$$P\left[\alpha(t+\delta t) = 0 \middle| \alpha(t) = 1\right] P[\alpha(t) = 1] +$$

$$P\left[\alpha(t+\delta t) = 0 \middle| \alpha(t) = 0\right] P[\alpha(t) = 0].$$

or

$$\pi(0, t+\delta t) = \mu \delta t \pi(1, t) + \pi(0, t) + o(\delta t)$$

or

$$\frac{d\pi(0, t)}{dt} = \mu \pi(1, t).$$

Discrete state, continuous time

Exponential distribution

$\pi(1, t + \delta t) =$

$$P\left[\alpha(t + \delta t) = 1 \middle| \alpha(t) = 1\right] P[\alpha(t) = 1] +$$

$$P\left[\alpha(t + \delta t) = 1 \middle| \alpha(t) = 0\right] P[\alpha(t) = 0].$$

or

$$\pi(1, t + \delta t) = (1 - \mu\delta t)\pi(1, t) + (0)\pi(0, t) + o(\delta t)$$

or

$$\frac{d\pi(1, t)}{dt} = -\mu\pi(1, t).$$

Discrete state, continuous time

Exponential distribution

Transition equations
$$\begin{cases} \dfrac{d\pi(0,t)}{dt} = \mu\pi(1,t) \\[2ex] \dfrac{d\pi(1,t)}{dt} = -\mu\pi(1,t) \end{cases}$$

If $\pi(0,0) = 0, \ \pi(1,0) = 1$, then

$$\pi(1,t) = e^{-\mu t}$$

and

$$\pi(0,t) = 1 - e^{-\mu t}$$

Discrete state, continuous time

Exponential distribution

The probability that the transition takes place at some $T \in [t, t+\delta t]$ is

$$f(t)\delta t = P[\alpha(t+\delta t) = 0 \text{ and } \alpha(t) = 1]$$

$$\approx P[\alpha(t+\delta t) = 0 | \alpha(t) = 1] P[\alpha(t) = 1]$$

$$= (\mu \delta t)(e^{-\mu t})$$

The exponential density function is therefore $f(t) = \mu e^{-\mu t}$ for $t \geq 0$ and 0 for $t < 0$.

The time of the transition from 1 to 0 is said to be *exponentially distributed* with rate μ.

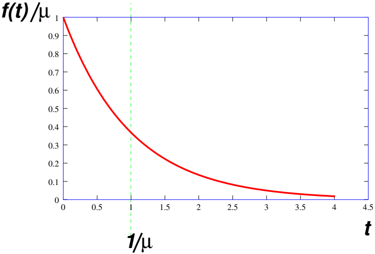
The expected transition time is $\dfrac{1}{\mu} = \displaystyle\int_0^\infty te^{-\mu t}$.

Discrete state, continuous time

Exponential distribution

- $f(t) = \mu e^{-\mu t}$ for $t \geq 0$; $f(t) = 0$ otherwise;
  $F(t) = 1 - e^{-\mu t}$ for $t \geq 0$; $F(t) = 0$ otherwise.

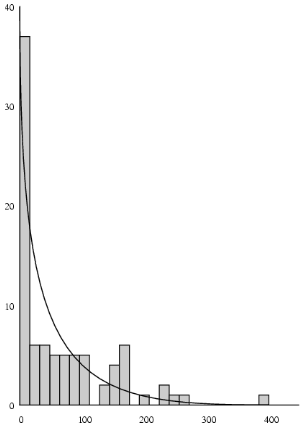
- $ET = 1/\mu, V_T = 1/\mu^2$. Therefore, $\sigma = ET$ so cv=1.



Markov processes
Exponential

Density function

Exponential density function and a small number of samples.



Discrete state, continuous time

Exponential distribution: some properties

- Memorylessness:

  $P(T > t + x | T > x) = P(T > t)$

- $P(t \leq T \leq t + \delta t | T \geq t) \approx \mu \delta t$ for small δt.

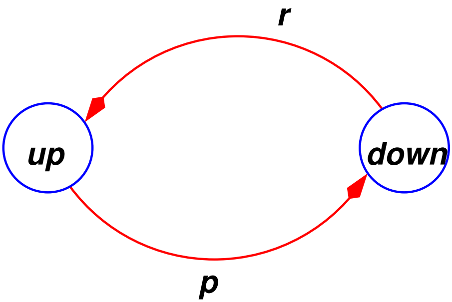
Discrete state, continuous time

Exponential distribution: some properties

- If $T_1, ..., T_n$ are independent exponentially distributed random variables with parameters $\mu_1 ..., \mu_n$, and
- $T = \min(T_1, ..., T_n)$, then
- T is an exponentially distributed random variable with parameter $\mu = \mu_1 + ... + \mu_n$.
- Consequently, the time that the system stays in any state is exponentially distributed.

Discrete state, continuous time
Unreliable machine

Continuous time unreliable machine.



Discrete state, continuous time

Unreliable machine

From the *Law of Total Probability:*

$P(\{\text{the machine is up at time } t+\delta t\})=$

$P(\{\text{the machine is up at time } t+\delta t \mid \text{the machine was up at time } t\}) \times$
$P(\{\text{the machine was up at time } t\}) +$

$P(\{\text{the machine is up at time } t+\delta t \mid \text{the machine was down at time } t\}) \times$
$P(\{\text{the machine was down at time } t\})$

$+o(\delta t)$

and similarly for $P(\{\text{the machine is down at time } t+\delta t\})$.

Discrete state, continuous time

Unreliable machine

Probability distribution notation and dynamics:

$\pi(1, t) =$ the probability that the machine is up at time t.

$\pi(0, t) =$ the probability that the machine is down at time t.

$$P(\text{the machine is up at time } t + \delta t | \text{ the machine was up at time } t) = 1 - p\delta t$$

$$P(\text{the machine is up at time } t + \delta t | \text{ the machine was down at time } t) = r\delta t$$

Discrete state, continuous time
Unreliable machine

Therefore

$$\pi(1, t + \delta t) = (1 - p\delta t)\pi(1, t) + r\delta t\pi(0, t) + o(\delta t)$$

Similarly,

$$\pi(0, t + \delta t) = p\delta t\pi(1, t) + (1 - r\delta t)\pi(0, t) + o(\delta t)$$

Discrete state, continuous time

Unreliable machine

or,

$$\pi(1, t + \delta t) - \pi(1, t) = -p\delta t \pi(1, t) + r\delta t \pi(0, t) + o(\delta t)$$

or,

$$\frac{\pi(1, t + \delta t) - \pi(1, t)}{\delta t} = -p\pi(1, t) + r\pi(0, t) + \frac{o(\delta t)}{\delta t}$$

Discrete state, continuous time

or,

$$\frac{d\pi(1,t)}{dt} = \pi(0,t)r - \pi(1,t)p$$

$$\frac{d\pi(0,t)}{dt} = -\pi(0,t)r + \pi(1,t)p$$

Markov processes
Unreliable machine

Solution

$$\begin{aligned}\pi(0,t) &= \frac{p}{r+p} + \left[\pi(0,0) - \frac{p}{r+p}\right] e^{-(r+p)t} \\ \pi(1,t) &= 1 - \pi(0,t).\end{aligned}$$

As $t \to \infty$,

$$\begin{aligned}\pi(0) &\to \frac{p}{r+p}, \\ \pi(1) &\to \frac{r}{r+p}\end{aligned}$$

Markov processes
Unreliable machine

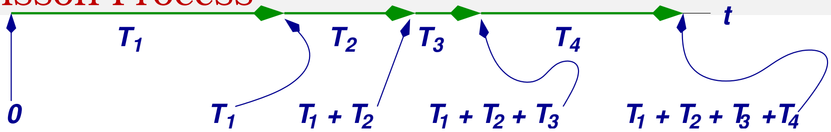
Note that MTTF$=1/p$; MTTR$=1/r$. Units are natural time units, not operation times.

If the machine makes μ parts per time unit on the average when it is operational, the steady-state average production rate is

$$\mu\pi(1) = \mu\frac{r}{r+p} = \mu\frac{\text{MTTF}}{\text{MTTF}+\text{MTTR}} = \mu e$$

Discrete state, continuous time

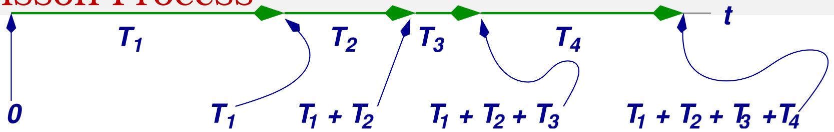
Poisson Process



- Let $T_i, i = 1, \ldots$ be a set of independent exponentially distributed random variables with parameter λ. Each random variable may represent the time between occurrences of a repeating event.
  - ⋆ Examples: customer arrivals, clicks of a Geiger counter
- Then $\sum_{i=1}^{n} T_i$ is the time required for n such events.

Discrete state, continuous time

Poisson Process



- *Informally:* $N(t)$ is the number of events that occur between 0 and t.
- *Formally:* Define

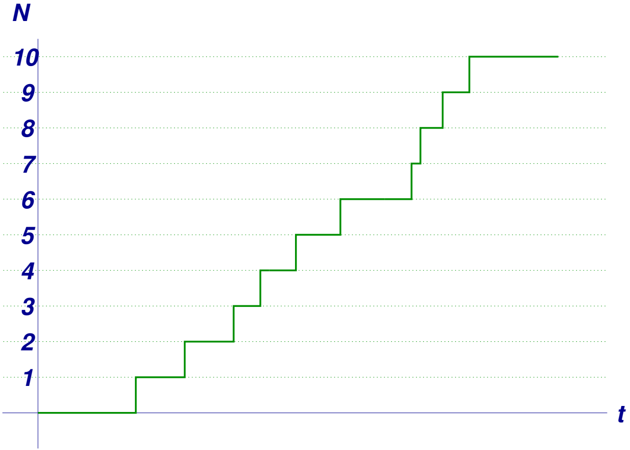
$$N(t) = \begin{cases} 0 \text{ if } T_1 > t \\ n \text{ such that } \sum_{i=1}^{n} T_i \le t, \quad \sum_{i=1}^{n+1} T_i > t \end{cases}$$

- Then $N(t)$ is a *Poisson process* with parameter λ.

Queueing theory

$M/M/1$ Queue

Number of events $N(t)$ during $[0, t]$

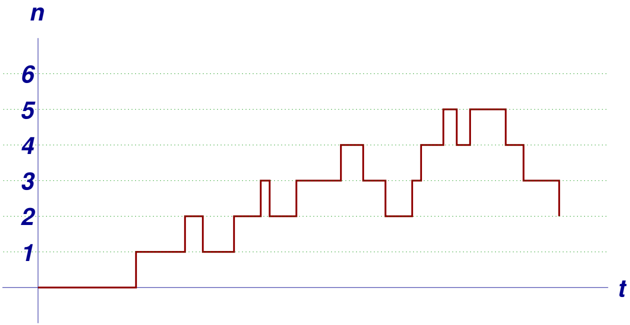


Queueing theory
$M/M/1$ Queue

- Simplest model is the $M/M/1$ queue:
  - ⋆ Exponentially distributed inter-arrival times — mean is $1/\lambda$; λ is *arrival rate* (customers/time). *(Poisson arrival process.)*
  - ⋆ Exponentially distributed service times — mean is $1/\mu$; μ is *service rate* (customers/time).
  - ⋆ The arrival and service processes are independent.
  - ⋆ 1 server.
  - ⋆ Infinite waiting area.
- Define the *utilization* $\rho = \lambda/\mu$.

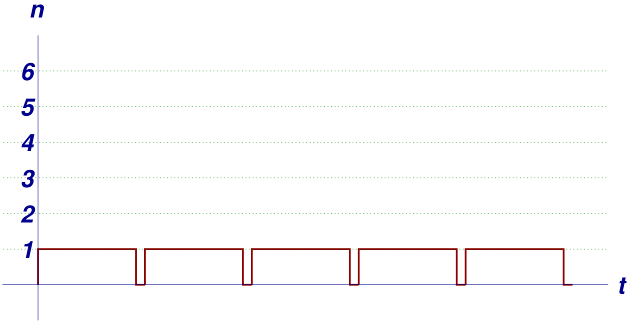
Queueing theory
$M/M/1$ Queue

Number of customers in the system as a function of time for a $M/M/1$ queue.



Queueing theory
$D/D/1$ Queue

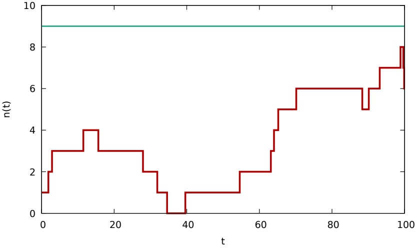
Number of customers in the system as a function of time for a $D/D/1$ queue.



Queueing theory
Sample path

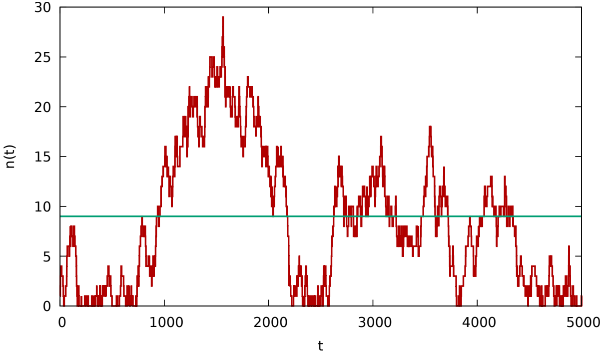
- Suppose customers arrive in a Poisson process with *average* inter-arrival time $1/\lambda = 1$ minute; and that service time is exponentially distributed with *average* service time $1/\mu = 54$ seconds.

  ⋆ The average number of customers in the system is *9*.



Queue behavior over a short time interval — initial transient

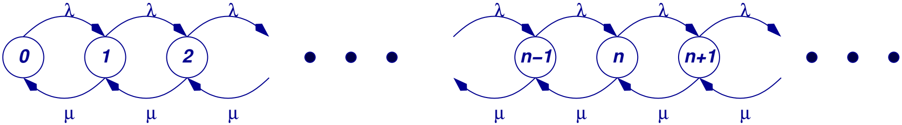
Queueing theory
Sample path



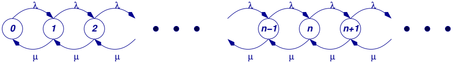
Queue behavior over a long time interval

Queueing theory
$M/M/1$ Queue

State space



Queueing theory
$M/M/1$ Queue



Let $\pi(n,t)$ be the probability that there are n parts in the system at time t. Then,

For $n > 0$,

$$\begin{aligned}\pi(n, t+\delta t) &= \pi(n-1,t)\lambda\delta t + \pi(n+1,t)\mu\delta t + \\ &\quad \pi(n,t)(1-(\lambda\delta t + \mu\delta t)) + o(\delta t)\end{aligned}$$

and

$$\pi(0, t+\delta t) = \pi(1,t)\mu\delta t + \pi(0,t)(1-\lambda\delta t) + o(\delta t).$$

Queueing theory
$M/M/1$ Queue

Or,

$$\begin{aligned}\frac{d\pi(n,t)}{dt} &= \pi(n-1,t)\lambda + \pi(n+1,t)\mu - \pi(n,t)(\lambda+\mu), \;\; n > 0\\ \frac{d\pi(0,t)}{dt} &= \pi(1,t)\mu - \pi(0,t)\lambda.\end{aligned}$$

If a steady state distribution exists, it satisfies

$$\begin{aligned}0 &= \pi(n-1)\lambda + \pi(n+1)\mu - \pi(n)(\lambda+\mu), n > 0\\ 0 &= \pi(1)\mu - \pi(0)\lambda.\end{aligned}$$

Why "if"?

Queueing theory
$M/M/1$ Queue – Steady State

Let $\rho = \lambda/\mu$. These equations are satisfied by

$$\pi(n) = (1-\rho)\rho^n, n \geq 0$$

if $\rho < 1$.

The average number of parts in the system is

$$\bar{n} = \sum_{n=0}^{\infty} n\pi(n) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}.$$

Queueing theory
Little's Law

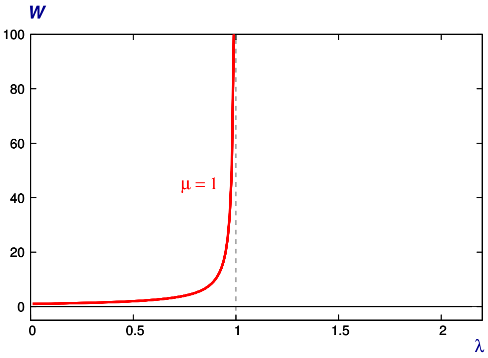
- True for most systems of practical interest *(not just $M/M/1$)*.
- Steady state only.
- $L =$ the average number of customers in a system.
- $W =$ the average delay experienced by a customer in the system.

$$L = \lambda W$$

In the $M/M/1$ queue, $L = \bar{n}$ and

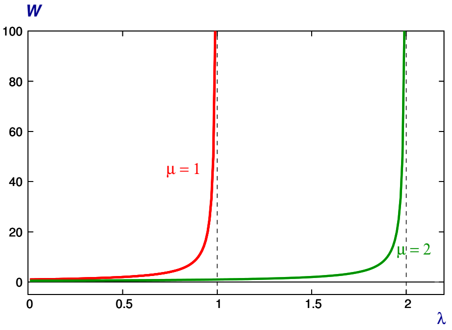
$$W = \frac{1}{\mu - \lambda}.$$

Queueing theory
$M/M/1$ Queue capacity



- μ is the *capacity* of the system.
- If $\lambda < \mu$, system is stable and waiting time remains bounded.
- If $\lambda > \mu$, waiting time grows over time.

Queueing theory
$M/M/1$ Queue capacity



- To increase capacity, increase μ.
- To decrease delay for a given λ, increase μ.

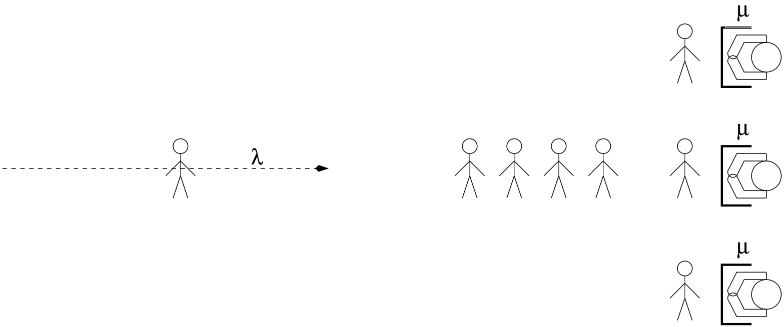
Queueing theory
Other Single-Stage Models

Things get more complicated when:

- There are multiple servers.
- There is finite space for queueing.
- The arrival process is not Poisson.
- The service process is not exponential.

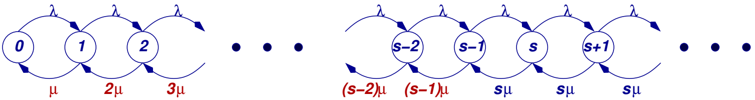
Closed formulas and approximations exist for some, but not all, cases.

Queueing theory
$M/M/s$ Queue



s-Server Queue, $s = 3$

Queueing theory
$M/M/s$ Queue



- The departure rate when there are $k > s$ customers in the system is $s\mu$ since all s servers are always busy.

- The departure rate when there are $k \leq s$ customers in the system is $k\mu$ since only k of the servers are busy.

Queueing theory
$M/M/s$ Queue

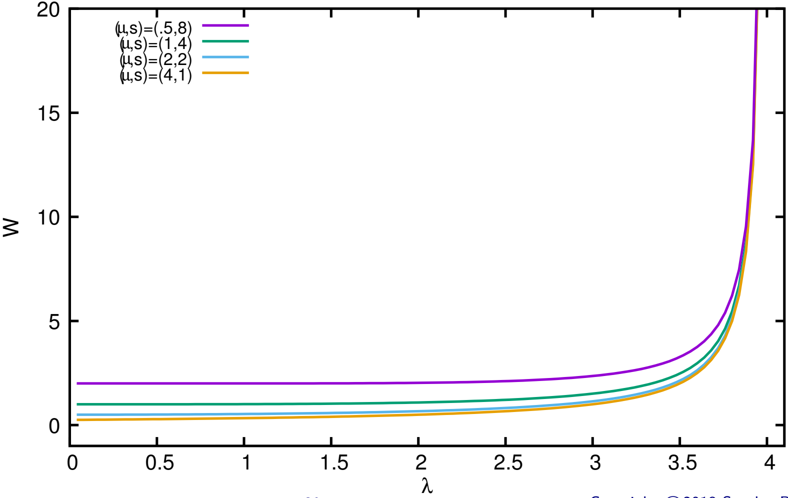
$$P(k) = \begin{cases} \pi(0)\dfrac{s^k \rho^k}{k!}, & k \le s \\[2ex] \pi(0)\dfrac{s^s \rho^k}{s!}, & k > s \end{cases}$$

where

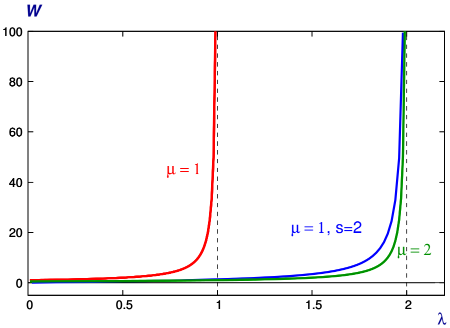
$$\rho = \frac{\lambda}{s\mu} < 1; \qquad \pi(0) \text{ chosen so that } \sum_k P(k) = 1$$

Queueing theory
$M/M/s$ Queue

W vs. λ; $s\mu = 4$



Queueing theory
$M/M/1$ Queue capacity



To increase capacity or reduce delay,

- increase μ, or
- add servers in parallel
  ... but that will not reduce delay as much.

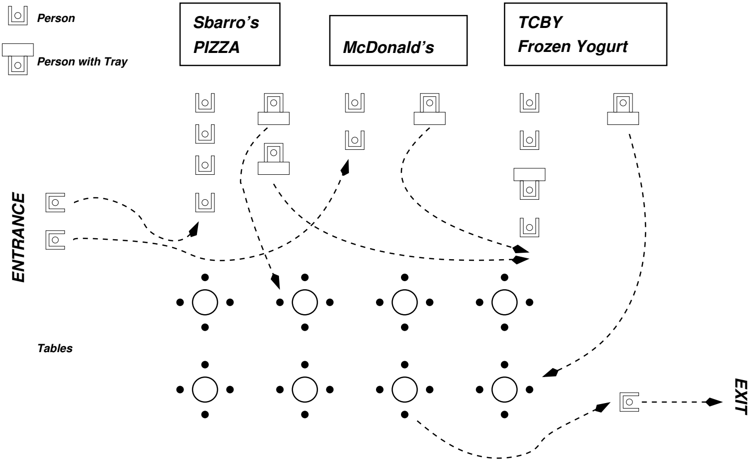
Queueing theory
Networks of Queues

- Set of queues where customers can go to another queue after completing service at a queue.
- *Open network:* where customers enter and leave the system. λ is known and we must find L and W.
- *Closed network:* where the population of the system is constant. L is known and we must find λ and W.

Queueing theory
Networks of Queues

Examples of open networks

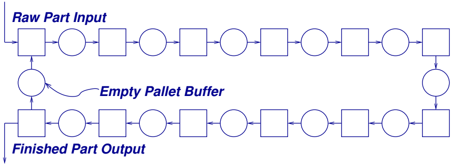
- internet traffic
- emergency room (*arrive*, triage, waiting room, treatment, tests, *exit* or *hospital admission*)
- food court
- airport (*arrive*, ticket counter, security, passport control, gate, *board plane*)
- factory with no *centralized* material flow control after material enters


Person
Person with Tray
Sbarro's PIZZA
McDonald's
TCBY Frozen Yogurt
ENTRANCE
Tables
EXIT

Queueing theory
Networks of Queues

Examples of closed networks

- factory with limited fixtures or pallets



- factory with material controlled by keeping the number of items constant (CONWIP)

Queueing theory
Jackson Networks

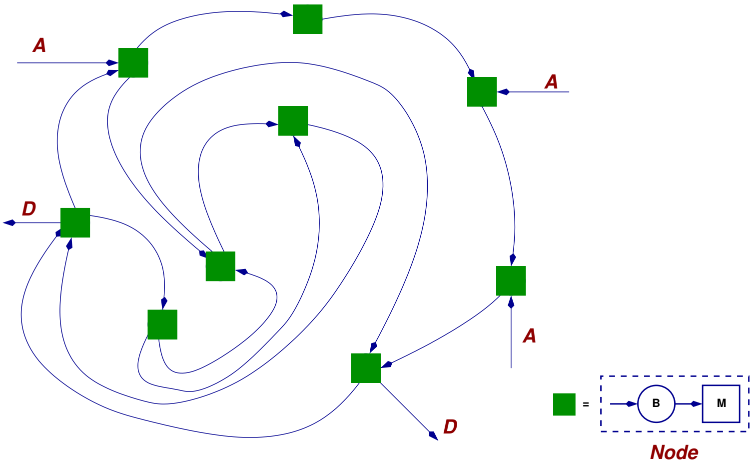
Queueing networks are often modeled as *Jackson networks.*

- Relatively easy to compute performance measures (capacity, average time in system, average queue lengths).
- Easily provides intuition.
- Easy to optimize and to use for design.
- Valid (or good approximation) for a large class of systems ...

Queueing theory
Jackson Networks

- ... but not all. Storage areas must be assumed to be infinite (which means blocking is assumed not to happen).
  - ⋆ This assumption leads to bad results for systems with bottlenecks at locations other than the first station.

Queueing theory
Open Jackson Networks



Queueing theory
Open Jackson Networks

- Items *arrive* from outside the system to node i according to a Poisson process with rate α_i.
- $\alpha_i > 0$ for at least one i.
- When an item's service at node i is finished, it goes to node j next with probability p_{ij}.
- If $p_{i0} = 1 - \sum_j p_{ij} > 0$, items *depart* from the network from node i.
- $p_{i0} > 0$ for at least one i.
- We will focus on the special case in which each node has a single server with exponential processing time. The service rate of node i is μ_i.

Queueing theory
Open Jackson Networks

Goals of analysis:

- to determine if the system is feasible
- to determine how much inventory is in this system (on the average) and how it is distributed
- to determine the average waiting time at each node and the average time a part spends in the system.

Queueing theory
Open Jackson Networks

- Define λ_i as the total arrival rate of items to node i. This includes items entering the network at i and items coming from all other nodes.
- $p_{ji}\lambda_j$ is the portion of the flow arriving at node j that goes to node i.
- Then $\lambda_i = \alpha_i + \sum_j p_{ji}\lambda_j$
- In matrix form, let λ be the vector of λ_i, α be the vector of α_i, and P be the matrix of p_{ij}. Then

$$\lambda = \alpha + \mathsf{P}^T \lambda$$

- Solving for λ,

$$\lambda = (I - \mathsf{P}^T)^{-1}\alpha$$

Queueing theory
Open Jackson Networks

Probability distribution:

- If $\lambda_i < \mu_i$ for each i, define $\rho_i = \lambda_i/\mu_i$ and

$$\pi_i(n_i) = (1 - \rho_i)\rho_i^{n_i}$$

- This is the solution of an M/M/1 queue with arrival rate λ_i calculated on the previous slide and service rate μ_i specified by the given problem data.

- If $\lambda_i \geq \mu_i$ for some i, *the demand is not feasible.* The system cannot handle the demand placed on it.

Queueing theory
Open Jackson Networks

Solution:

- Define $\pi(n_1, n_2, ..., n_k)$ to be the steady-state probability that there are n_i items at node i, $i = 1, ..., k$.
- Then the probability distribution for the entire system is

$$\pi(n_1, n_2, ..., n_k) = \prod_i \pi_i(n_i)$$

- At each node i

$$\bar{n}_i = En_i = \frac{\rho_i}{1 - \rho_i}$$

Queueing theory
Open Jackson Networks

- The solution is product form. It says that the probability of the system being in a given state is the product of the probabilities of the queue at each node being in the corresponding state.
- This exact analytic formula is the reason that the Jackson network model is of interest. It is relatively easy to use to calculate the performance of a complex system.
- The product form solution holds for some more general cases.
- However, it is restricted to models of systems with unlimited storage space. *Consequently, it cannot model blocking.*
    - ⋆ It is a good approximation for systems where blocking is rare, for example when the arrival rate of material is much less than the capacity of the system.
    - ⋆ It will not work so well where blocking occurs often.

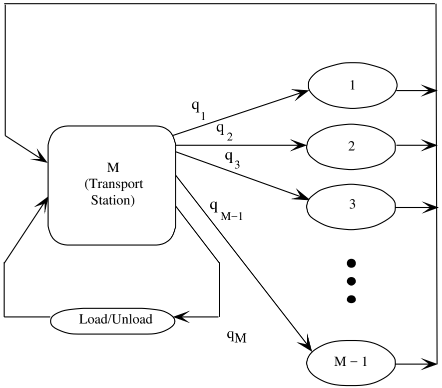
Queueing theory
Closed Jackson Networks

- Consider an extension in which
  - ⋆ $\alpha_i = 0$ for all nodes i.
  - ⋆ $p_{i0} = 1 - \sum_j p_{ij} = 0$ for all nodes i.
- Then
  - ⋆ Since nothing is entering and nothing is departing from the network, the number of items in the network is *constant* . That is, $\sum_i n_i(t) = N$ for all t.
  - ⋆ $\lambda_i = \sum_j p_{ji}\lambda_j$ does not have a unique solution.
  - ⋆ This means that a different solution approach is needed to analyze the system. It is used in the example that follows.

Queueing theory

Closed Jackson Network model of an FMS

Solberg's "CANQ" model.



Let $\{p_{ij}\}$ be the set of routing probabilities, as defined on Slide 89.

$p_{iM} = 1$ if $i \neq M$

$p_{Mj} = q_j$ if $j \neq M$

$p_{ij} = 0$ otherwise

Service rate at Station i is μ_i.

Queueing theory

Closed Jackson Network model of an FMS

- Input data: $M, N, q_j, \mu_j, s_j \quad (j = 1, ..., M)$
  - ⋆ M = number of stations, including transportation system
  - ⋆ N = number of pallets
  - ⋆ q_j = fraction of parts going from the transportation system to Station j
  - ⋆ μ_j = processing rate of machines at Station j
  - ⋆ s_j = number of machines at Station j

- Output data: $P, W, \rho_j \quad (j = 1, ..., M)$
  - ⋆ P = production rate
  - ⋆ W = average time a part spends in the system
  - ⋆ ρ_j = utilization per machine of Station j

Queueing theory

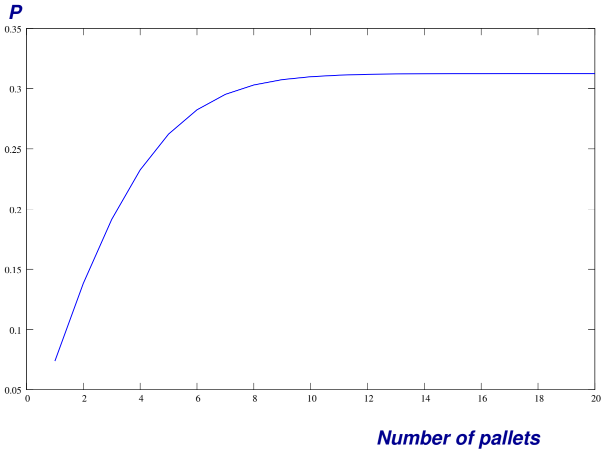
Closed Jackson Network model of an FMS

For the following graphs,

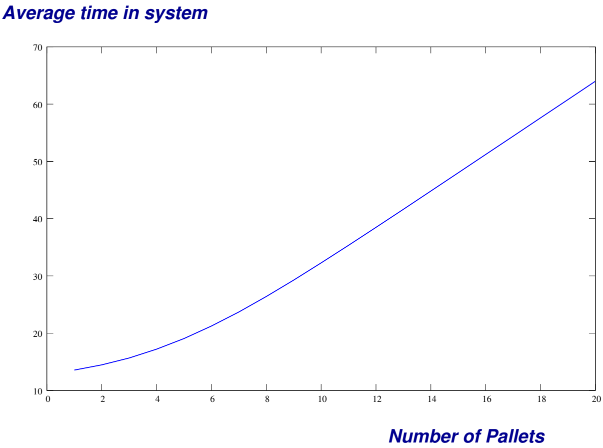
- Base input data: $M, N, q_j, \mu_j, s_j \quad (j = 1, ..., M)$
  - $M = 5$
  - $N = 10$
  - $q_j = .1, .2, .2, .25, .25$
  - $1/\mu_j = 3., 4., 3.44, 1.41, 5.$
  - $s_j = 2, 1, 2, 1, 15$

We see the effect of one of the variables on the performance measures in the following graphs.

Queueing theory
Closed Jackson Network model of an FMS

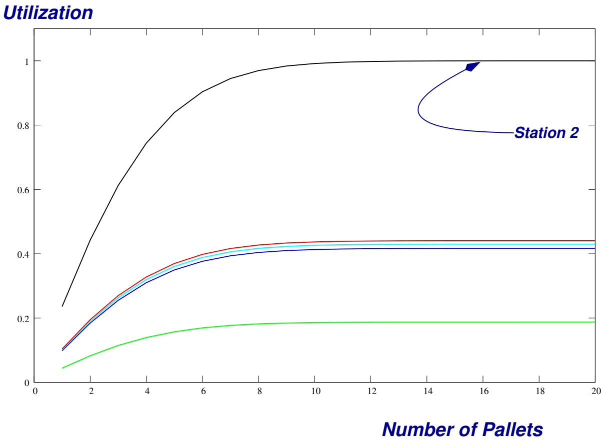


Queueing theory
Closed Jackson Network model of an FMS

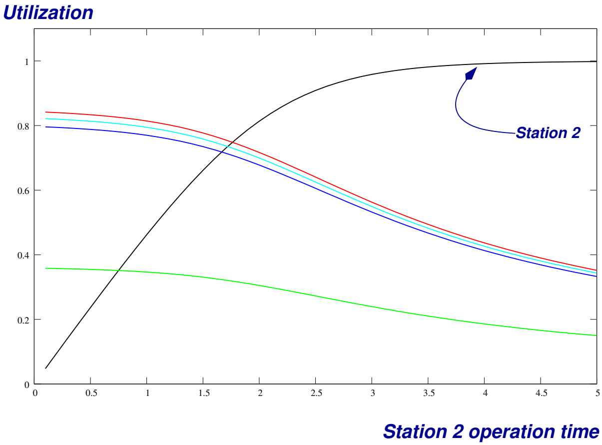


Queueing theory

Closed Jackson Network model of an FMS

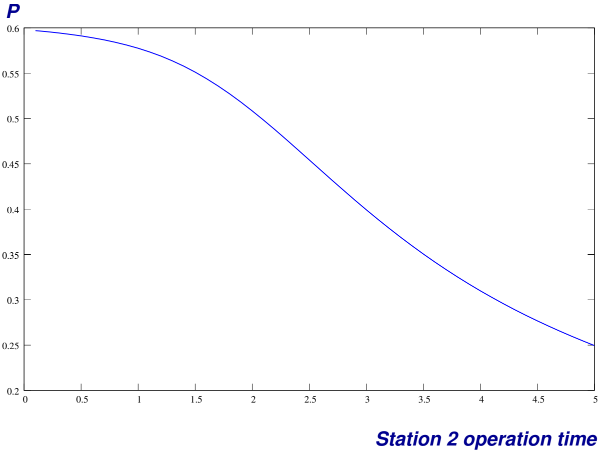


Queueing theory
Closed Jackson Network model of an FMS



Queueing theory

Closed Jackson Network model of an FMS



Queueing theory

Closed Jackson Network model of an FMS