Queues

Lecturer: Stanley B. Gershwin

Stochastic processes

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

Stochastic 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. Then
 - $\mathop{\mathrm{prob}} \{X(t+\delta t) = A | X(s) = x(s), s \leq t\} = \\ \mathop{\mathrm{prob}} \{X(t+\delta t) = A | X(t) = x(t)\}$

Stochastic 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 the definition of a class of mathematical models. It is <u>NOT</u> a statement about reality!! That is, not everything is a Markov process.

Example

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

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

Discrete state, discrete time

Markov processes

States and transitions

Transition graph



 P_{ij} is a probability. Note that $P_{ii} = 1 - \sum_{m,m
eq i} P_{mi}$.

Markov
processesDiscrete state, discrete timeStates and transitions

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

Assume probability of H is p.



Discrete state, discrete time

States and transitions

Example : Coin flip bets on Slide 5.

Assume probability of H is p.



Discrete state, <u>discrete</u> time

States and transitions

- Define $\pi_i(t) = \operatorname{prob}\{X(t) = i\}.$
- Transition equations: $\pi_i(t+1) = \sum_j P_{ij}\pi_j(t)$. (Law of Total Probability)
- Normalization equation: $\sum_i \pi_i(t) = 1$.

Markov
processesDiscrete state, discrete time
States and transitions



Discrete state, discrete time

Markov processes

States and transitions



 $= \operatorname{prob}\{X(t+1) = 2 | X(t) = 1\} \operatorname{prob}\{X(t) = 1\}$ $+ \operatorname{prob}\{X(t+1) = 2 | X(t) = 2\} \operatorname{prob}\{X(t) = 2\}$ $+ \operatorname{prob}\{X(t+1) = 2 | X(t) = 4\} \operatorname{prob}\{X(t) = 4\}$ $+ \operatorname{prob}\{X(t+1) = 2 | X(t) = 5\} \operatorname{prob}\{X(t) = 5\}$

Discrete state, discrete time

States and transitions



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

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

Discrete state, discrete time

States and transitions

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

Discrete state, discrete time

States and transitions



Balance equation:

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

 $=\pi_5 P_{45}$

in steady state only.

Markov
processesDiscrete state, discrete time
Geometric distribution

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

0

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

$$p = \text{ prob } [lpha(t+1) = 0 | lpha(t) = 1].$$

Markov processes Discrete state, <u>discrete</u> time

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

$$\begin{aligned} \pi(0,t+1) &= \text{prob} \; [\alpha(t+1) = 0 | \alpha(t) = 1] \; \text{prob} \; [\alpha(t) = 1] \\ &+ \text{prob} \; [\alpha(t+1) = 0 | \alpha(t) = 0] \; \text{prob} \; [\alpha(t) = 0], \end{aligned}$$

we have

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

and the normalization equation

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

Markov processes Discrete state, <u>discrete</u> time

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

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

 $\pi(1,t) = (1-p)^t.$



Geometric Distribution



Markov Discrete state, discrete time processes Unreliable machine

1=up; 0=down.



Discrete state, discrete time

Unreliable machine

The probability distribution satisfies

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

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

Discrete state, discrete time

Unreliable machine

It is not hard to show that

$$egin{aligned} \pi(0,t) &= \pi(0,0)(1-p-r)^t \ &+ rac{p}{r+p} \left[1-(1-p-r)^t
ight], \end{aligned}$$

$$egin{aligned} \pi(1,t) &= \pi(1,0)(1-p-r)^t \ &+rac{r}{r+p}ig[1-(1-p-r)^tig]\,. \end{aligned}$$

probability

Discrete state, discrete time

Unreliable machine

Discrete Time Unreliable Machine



Discrete state, discrete time

Unreliable machine

As $t \to \infty$,

$$egin{aligned} \pi(0) &
ightarrow rac{p}{r+p}, \ \pi(1) &
ightarrow rac{r}{r+p} \end{aligned}$$

which is the solution of

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

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

Discrete state, discrete time

Unreliable machine

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

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

Discrete state, <u>continuous</u> time

States and transitions

- States can be numbered 0, 1, 2, 3, ... (or with multiple indices if that is more convenient).
- \bullet 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

 $\lambda_{ij}\delta t pprox \operatorname{prob}\{X(t+\delta t)=i|X(t)=j\} ext{ for } i
eq j\}$

Markov
processesDiscrete state, continuous time
States and transitions

Transition graph



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

Markov
processesDiscrete state, continuous timeStates and transitions

Transition equation

Define $\pi_i(t) = \operatorname{prob}\{X(t) = i\}$. Then for δt small,

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

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

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

Markov
processesDiscrete state, continuous timeStates and transitions

Or,

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

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

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

Markov
processesDiscrete state, continuous timeStates and transitions

Or,

$$\lim_{\delta t o 0} rac{\pi_5(t+\delta t)-\pi_5(t)}{\delta t} = rac{d\pi_5}{dt}(t) = -(\lambda_{25}+\lambda_{45}+\lambda_{65})\pi_5(t)$$

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

Discrete state, <u>continuous</u> time

States and transitions

- Define $\pi_i(t) = \operatorname{prob}\{X(t) = i\}$
- It is convenient to define $\lambda_{ii} = -\sum_{j
 eq i} \lambda_{ji}$
- Transition equations: $rac{d\pi_i(t)}{dt} = \sum_j \lambda_{ij}\pi_j(t).$
- Normalization equation: $\sum_i \pi_i(t) = 1$.

Discrete state, <u>continuous</u> time

States and transitions

- 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: $\pi_i \sum_{m,m
 eq i} \lambda_{mi} = \sum_{j,j
 eq i} \lambda_{ij} \pi_j$
- Normalization equation: $\sum_i \pi_i = 1$.

Discrete state, <u>continuous</u> time

States and transitions

Sources of confusion in continuous time models:

• Never Draw self-loops in continuous time markov process graphs.

• Never write
$$1 - \lambda_{14} - \lambda_{24} - \lambda_{64}$$
. Write

$$egin{array}{lll} \star & 1-(\lambda_{14}+\lambda_{24}+\lambda_{64})\delta t, \, {
m or} \ \star & -(\lambda_{14}+\lambda_{24}+\lambda_{64}) \end{array}$$

• $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$ is NOT a rate and NOT a probability. It is ONLY a convenient notation.

Markov Discrete state, continuous time processes Exponential

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



Markov Discrete state, continuous time processes Exponential

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

Or

prob $[\alpha(t+\delta t)=0|\alpha(t)=1]$ prob $[\alpha(t)=1]+$ prob $[\alpha(t+\delta t)=0|\alpha(t)=0]$ prob $[\alpha(t)=0].$ or

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

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

Discrete state, <u>continuous</u> time

Exponential

٠

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

$$\frac{d\pi(1,t)}{dt} = -p\pi(1,t)$$
If $\pi(1,0) = 1$, then
 $\pi(1,t) = e^{-pt}$
and

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


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

prob
$$[lpha(t+\delta t)=0$$
 and $lpha(t)=1]=e^{-pt}p\delta t.$

The exponential density function is pe^{-pt} .

The time of the transition from 1 to 0 is said to be *exponentially distributed* with rate p. The expected transition time is 1/p. (*Prove it!*)

Discrete state, <u>continuous</u> time

Exponential

• $f(t) = \mu e^{-\mu t}$ for $t \ge 0$; f(t) = 0 otherwise; $F(t) = 1 - e^{-\mu t}$ for $t \ge 0$; F(t) = 0 otherwise. • $ET = 1/\mu, V_T = 1/\mu^2$. Therefore, cv=1.



Discrete state, <u>continuous</u> time

Exponential

• 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 .
- 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 $\mu = \mu_1 + ... + \mu_n$.

Discrete state, <u>continuous</u> time

Unreliable machine

Continuous time unreliable machine. MTTF=1/p; MTTR=1/r.





Let T_i , i = 1, ... be a set of independent exponentially distributed random variables with parameter λ that each represent the time until an event occurs. Then $\sum_{i=0}^{n} T_i$ is the time required for nsuch events.

Define
$$N(t) = \left\{egin{array}{l} 0 ext{ if } T_1 > t \ n ext{ such that } \sum_{i=0}^n T_i \leq t, \ \sum_{i=0}^{n+1} T_i > t \end{array}
ight.$$

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

Discrete state, <u>continuous</u> time

Markov processes

Poisson Distribution

$$P(N(t)=n)=e^{-\lambda t}rac{(\lambda t)^n}{n!}$$



 $\lambda t = 6$

Discrete state, <u>continuous</u> time

Poisson Distribution

- 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).
 - \star 1 server.
 - ★ Infinite waiting area.
- Define the *utilization* $\rho = \lambda/\mu$.

M/M/1 Queue

Sample path

Number of customers in the system as a function of time.

M/M/1 Queue

State Space

and

M/M/1 Queue

Performance of M/M/1 queue

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

$$P(n,t+\delta t) = P(n-1,t)\lambda\delta t + P(n+1,t)\mu\delta t$$

 $+P(n,t)(1-(\lambda\delta t+\mu\delta t))+o(\delta t)$
for $n>0$

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

M/M/1 Queue

Performance of M/M/1 queue

Or,

If a steady state distribution exists, it satisfies

 $egin{aligned} 0 &= P(n-1)\lambda + P(n+1)\mu - P(n)(\lambda+\mu), n > 0 \ 0 &= P(1)\mu - P(0)\lambda. \end{aligned}$ Why "if"?

M/M/1 Queue

Performance of M/M/1 queue

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

$$P(n)=(1-
ho)
ho^n,n\geq 0$$

if $\rho < 1$. The average number of parts in the system is

$$ar{n} = \sum_n n P(n) = rac{
ho}{1-
ho} = rac{\lambda}{\mu-\lambda}.$$

M/M/1 Queue

Little's Law

- True for most systems of practical interest.
- 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

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

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.

M/M/1 Queue

Capacity

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

M/M/1 Queue

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 cases.

M/M/s Queue

s-Server Queue, s = 3

M/M/s Queue

State Space

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

M/M/s Queue

Steady-State Probability Distribution

$$P(k) = egin{cases} P(0)rac{s^k
ho^k}{k!}, & k\leq s\ P(0)rac{s^s
ho^k}{s!}, & k>s \end{cases}$$

where

$$ho = rac{\lambda}{s\mu} < 1; \quad P(0)$$
 chosen so that $\sum_k P(k) = 1$

M/M/s Queue

Performance

M/M/s Queue

Performance

M/M/s Queue

Performance

• Why do the curves go to infinity at the same value of λ ?

 \bullet Why is the $(\mu,s)=(.5,8)$ curve the highest, followed by $(\mu,s)=(1,4),$ etc.?

- 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.

Examples

Open networks

- internet traffic
- emergency room
- food court
- airport (*arrive*, ticket counter, security, passport control, gate, *board plane*)
- factory with serial production system and no material control after it enters

Examples

Food Court

Closed Networks

- factory with material controlled by keeping the number of items constant (CONWIP)
- factory with limited fixtures or pallets

Benefits

Jackson Networks

Queueing networks are often modeled as *Jackson networks*.

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

Limitations

- ... but not everything. Storage areas must be infinite (i.e., blocking never occurs).
 - ***** This assumption *fails* for systems with bottlenecks.

 In Jackson networks, there is only one class. That is, all items are interchangeable. However, this restriction can be relaxed.

Goal of analysis: say something about how much inventory there is in this system and how it is distributed.

Open Jackson Networks

Assumptions

- 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 $p_{i0} = 1 \sum_{j} p_{ij} > 0$, then 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 .

 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.

• Then
$$\lambda_i = lpha_i + \sum_j p_{ji} \lambda_j$$

• 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

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

or

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

Open Jackson Networks

Product Form 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.
- Define $ho_i=\lambda_i/\mu_i;$ $\pi_i(n_i)=(1ho_i)
 ho_i^{n_i}.$
- Then

$$egin{aligned} \pi(n_1,n_2,...,n_k) &= \prod_i \pi_i(n_i) \ ar{n}_i &= En_i = rac{
ho_i}{1-
ho_i} \end{aligned}$$

Does this look familiar?

Open Jackson Networks

Product Form Solution

- This looks as though each station is an M/M/1 queue. But even though this is *NOT* in general true, the formula holds.
- The product form solution holds for some more general cases.
- This exact analytic formula is the reason that the Jackson network model is very widely used *sometimes where it does not belong!*

Closed Jackson Networks

Jackson Networks

• Consider an extension in which

$$st lpha_i = 0$$
 for all nodes i .
 $st p_{i0} = 1 - \sum_j p_{ij} = 0$ 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 .
 $\star \lambda_{i} = \sum_{j} p_{ji}\lambda_{j}$ does not have a unique solution:
If $\{\lambda_{1}^{*}, \lambda_{2}^{*}, ..., \lambda_{k}^{*}\}$ is a solution, then $\{s\lambda_{1}^{*}, s\lambda_{2}^{*}, ..., s\lambda_{k}^{*}\}$ is also a solution for any $s \geq 0$.

For some *s*, define

$$\pi^o(n_1,n_2,...,n_k) = \prod_i \left[(1-
ho_i)
ho_i^{n_i}
ight] = \left[\prod_i (1-
ho_i)
ight] \left[\prod_i
ho_i^{n_i}
ight]$$
 where

Closed Jackson Networks

$$ho_i = rac{s\lambda_i^*}{\mu_i}$$

This looks like the open network probability distribution, but it is a function of s.
Consider a closed network with a population of N. Then if $\sum_{i} n_{i} = N$,

$$\pi(n_1,n_2,...,n_k) = rac{\pi^o(n_1,n_2,...,n_k)}{\sum\limits_{m_1+m_2+...+m_k=N}\pi^o(m_1,m_2,...,m_k)}$$

Closed Jackson Networks

Since π^o is a function of s, it looks like π is a function of s. But it is not because all the s's cancel! There are nice ways of calculating

$$C(k,N) = \sum_{m_1+m_2+...+m_k=N} \pi^o(m_1,m_2,...,m_k)$$

Closed Jackson Networks

Application — Simple FMS model



Solberg's "CANQ" model.

Let $\{p_{ij}\}$ be the set of routing probabilities, as defined on slide 67.

$$p_{iM}=1$$
 if $i
eq M$

$$p_{Mj} = q_j$$
 if $j
eq M$

 $p_{ij}=0$ otherwise

Service rate at Station i is μ_i .

Closed Jackson Networks

Application — Simple FMS model

Let N be the number of pallets.

The production rate is

$$P = \frac{C(M, N-1)}{C(M, N)} \mu_m$$

and C(M, N) is easy to calculate in this case.

- ullet Input data: $M, N, q_j, \mu_j (j=1,...,M)$
- Output data: $P, W,
 ho_j (j = 1, ..., M)$

Closed Jackson Networks

Application — Simple FMS model



Number of pallets

Closed Jackson Networks

Application — Simple FMS model

Average time in system



Number of Pallets

Closed Jackson Networks

Application — Simple FMS model



Number of Pallets

Closed Jackson Networks

Application — Simple FMS model



Station 2 operation time

Closed Jackson Networks

Application — Simple FMS model

Average time in system



Station 2 operation time

Closed Jackson Networks

Application — Simple FMS model

Utilization



Station 2 operation time

2.854 / 2.853 Introduction to Manufacturing Systems Fall 2010

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.