Introduction to Markov Chains

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

1. Applications in Biology and Queueing Theory
   - Markov Chains in Biology
   - Queues and Queueing Networks

Subsection 1

Markov Chains in Biology

Modeling with a Branching Process

- Galton and Watson in the 1870s used a branching process while seeking a quantitative explanation for the phenomenon of the disappearance of family names, even in a growing population.
- Assume each male in a given family has a probability p_k of having k sons.
- The goal is to determine the probability that, after n generations, an individual had no male descendants.
- Suppose at time $n = 0$ there is one individual.
- He dies and is replaced at time $n = 1$ by a random number of offspring N.
- These offspring also die and are replaced at time $n = 2$, each independently, by a random number of further offspring, having the same distribution as N.

  $\vdots$

The Branching Process

- Take, for each $n \in \mathbb{N}$, a sequence of independent random variables $(N_k^n)_{k \in \mathbb{N}}$, each with the same distribution as N.
- Set $X_0 = 1$.
- Define inductively, for $n \geq 1$,

$$X_n = N_1^n + \cdots + N_{X_{n-1}}^n.$$

- Then X_n gives the size of the population in the n-th generation.
- The process $(X_n)_{n \geq 0}$ is a Markov chain on $I = \{0, 1, 2, \ldots\}$ with absorbing state 0.

Transience of the Branching Process

- We exclude the case where $\mathbb{P}(N=1)=1$.
- We have $\mathbb{P}(X_n=0|X_{n-1}=i)=\mathbb{P}(N=0)^i$.
  - Suppose $\mathbb{P}(N=0)>0$.
    Then state i leads to 0.
    Every state $i \geq 1$ is transient.
  - Suppose $\mathbb{P}(N=0)=0$.
    Then $\mathbb{P}(N \geq 2)>0$.
    So, for $i \geq 1$, i leads to j, for some $j>i$, and j does not lead to i.
- Hence, i is transient in any case.
- We deduce that with probability 1, one of the following happens:
  - $X_n=0$, for some n;
  - $X_n \to \infty$, as $n \to \infty$.

The Probability Generating Function

- Further information on $(X_n)_{n\geq 0}$ is obtained by exploiting the branching structure.
- The probability generating function, defined for $0 \leq t \leq 1$, is

$$\phi(t) = \mathbb{E}(t^N) = \sum_{k=0}^{\infty} t^k \mathbb{P}(N = k).$$

- Conditional on $X_{n-1} = k$, we have

$$X_n = N_1^n + \cdots + N_k^n.$$

- So

$$\mathbb{E}(t^{X_n} | X_{n-1} = k) = \mathbb{E}(t^{N_1^n + \cdots + N_k^n}) = \phi(t)^k.$$

The Probability Generating Function (Cont'd)

- It follows that

$$\mathbb{E}(t^{X_n}) = \sum_{k=0}^{\infty} \mathbb{E}(t^{X_n} | X_{n-1} = k) \mathbb{P}(X_{n-1} = k) = \mathbb{E}(\phi(t)^{X_{n-1}}).$$

- By induction, we find that

$$\mathbb{E}(t^{X_n}) = \phi^{(n)}(t),$$

where $\phi^{(n)}$ is the *n*-fold composition $\phi \circ \cdots \circ \phi$.

- In principle, this gives the entire distribution of X_n, though $\phi^{(n)}$ may be a rather complicated function.

Probability of Survival

- Suppose $\mu = \mathbb{E}(N)$.
- We have

$$\mathbb{E}(X_n) = \lim_{t \nearrow 1} \frac{d}{dt}\mathbb{E}(t^{X_n}) = \lim_{t \nearrow 1} \frac{d}{dt}\phi^{(n)}(t) = \left(\lim_{t \nearrow 1} \phi'(t)\right)^n = \mu^n.$$

- Moreover,

$$\mathbb{P}(X_n = 0) = \phi^{(n)}(0).$$

- But state 0 is absorbing.
- So we have

$$q = \mathbb{P}(X_n = 0 \text{ for some } n) = \lim_{n \to \infty} \phi^{(n)}(0).$$

Probability of Survival (Cont'd)

- $\phi(t) = \mathbb{E}(t^N)$ is a convex function with $\phi(1) = 1$.
- Set
$$r = \inf\{t \in [0,1] : \phi(t) = t\}.$$
- Then $\phi(r) = r$ by continuity.
- ϕ is increasing and $0 \leq r$.
- So we have $\phi(0) \leq r$.
- By induction,
$$\phi^{(n)}(0) \leq r, \quad \text{for all } n.$$
- It follows that $q \leq r$.

Probability of Survival (Cont'd)

- On the other hand

$$q = \lim_{n\to\infty} \phi^{(n+1)}(0) = \lim_{n\to\infty} \phi(\phi^{(n)}(0)) = \phi(q).$$

- So also $q \geq r$.
- We conclude that $q = r$.
- We consider two cases.
  - Suppose, first, $\phi'(1) > 1$.
    Then we must have $q < 1$.
  - Suppose, next, $\phi'(1) \leq 1$.
    Now either $\phi'' = 0$ or $\phi'' > 0$ everywhere in $[0, 1)$.
    So we must have $q = 1$.
- We have shown that the population survives with positive probability if and only if $\mu > 1$, where μ is the mean of the offspring distribution.

Branching Processes and Random Walks

- We explore a connection between branching processes and random walks.
- Suppose that in each generation we replace individuals by their offspring one at a time.
- So if $X_n = k$, then it takes k steps to obtain X_{n+1}.
- The population size then performs a random walk $(Y_m)_{m \geq 0}$ with step distribution $N - 1$.
- Define stopping times:
  - $T_0 = 0$;
  - $T_{n+1} = T_n + Y_{T_n}$, for $n \geq 0$.
- Observe that

$$X_n = Y_{T_n}, \quad \text{for all } n.$$

Branching Processes and Random Walks (Cont'd)

- $(Y_m)_{m\geq 0}$ jumps down by at most 1 each time;
- So $(X_n)_{n\geq 0}$ hits 0 if and only if $(Y_m)_{m\geq 0}$ hits 0.
- Moreover, we can use the Strong Markov Property and a variation of the argument of a previous example to see that if $q_i = \mathbb{P}(Y_m = 0 \text{ for some } m | Y_0 = i)$ then $q_i = q_1^i$, for all i.
- So

$$q_1 = \mathbb{P}(N = 0) + \sum_{k=1}^{\infty} q_1^i \mathbb{P}(N = i) = \phi(q_1).$$

- Each non-negative solution of this equation provides a non-negative solution of the hitting probability equations.
- So we deduce that q_1 is the smallest non-negative root of the equation $q = \phi(q)$.
- This agrees with the generating function approach.

Epidemics

- In an idealized population we might suppose that:
  - All pairs of individuals make contact randomly and independently at a common rate, whether infected or not.
- For an idealized disease we might suppose that:
  - On contact with an infective, individuals themselves become infective and remain so for an exponential random time, after which they either die or recover.
- This idealized model is unrealistic.
- However, it is the simplest mathematical model to incorporate the basic features of an epidemic.
- We explore the consequences for the progress of the epidemic.

Formalization of the Model

- Denote:
  - The number of susceptibles by S_t;
  - The number of infectives by I_t.
- In the idealized model,
$$X_t = (S_t, I_t)$$
performs a Markov chain on $(\mathbb{Z}^+)^2$ with transition rates:
  - $q_{(s,i)(s-i,i+1)} = \lambda si$, for some $\lambda \in (0, \infty)$;
  - $q_{(s,i)(s,i-1)} = \mu i$, for some $\mu \in (0, \infty)$.
- Since $S_t + I_t$ does not increase, we effectively have a finite state-space.

Features of the Model

- The states $(s, 0)$, for $s \in \mathbb{Z}^+$, are all absorbing.
- All the other states are transient.
- All the communicating classes are singletons.
- The epidemic must therefore eventually die out.
- The absorption probabilities give the distribution of the number of susceptibles who escape infection.

Behavior in a Large Population

- We analyze the behavior in a large population, of size N, say.
- Consider the proportions
$$s_t^N = \frac{S_t}{N} \quad \text{and} \quad i_t^N = \frac{I_t}{N}.$$
- Suppose that
$$\lambda = \frac{\nu}{N}.$$
where ν is independent of N.
- Consider a sequence of models as $N \to \infty$.
- Choose
$$s_0^N \to s_0 \quad \text{and} \quad i_0^N \to i_0.$$

Behavior in a Large Population (Cont'd)

- It can be shown that as $N \to \infty$ the process (s_t^N, i_t^N) converges to the solution (s_t, i_t) of the differential equations

$$\begin{aligned} \frac{d}{dt} s_t &= -\nu s_t i_t; \\ \frac{d}{dt} i_t &= \nu s_t i_t - \mu i_t, \end{aligned}$$

  starting from (s_0, i_0).
- This means that

$$\mathbb{E}\left[\left|(s_t^N, i_t^N) - (s_t, i_t)\right|\right] \to 0, \quad \text{for all } t \geq 0.$$

- We will not prove this result, but will give an example of another easier asymptotic calculation.

Spreading of a Rumor

- Consider the case where:
  - $S_0 = N - 1$;
  - $I_0 = 1$;
  - $\lambda = \frac{1}{N}$;
  - $\mu = 0$.
- This can be given an alternative interpretation.
- A rumor is begun by a single individual who tells it to everyone she meets.
- They in turn pass the rumor on to everyone they meet.
- We assume that each individual meets another randomly at the jump times of a Poisson process of rate 1.
- We look at how long it takes until everyone knows the rumor.

Spreading of a Rumor (Cont'd)

- Suppose i people know the rumor.
- Then $N - i$ people do not.
- The rate at which the rumor is passed on is

$$q_i = \frac{i(N-i)}{N}.$$

- The expected time until everyone knows the rumor is then

$$\begin{aligned} \textstyle\sum_{i=1}^{N-1} q_i^{-1} &= \textstyle\sum_{i=1}^{N-1} \frac{N}{i(N-i)} \\ &= \textstyle\sum_{i=1}^{N-1} (\frac{1}{i} + \frac{1}{N-i}) \\ &= 2\textstyle\sum_{i=1}^{N-1} \frac{1}{i} \sim 2\log N. \end{aligned}$$

- This is not a limit but, rather, an asymptotic equivalence.
- The fact that the expected time grows with N is related to the fact that we do not scale I_0 with N.
- When the rumor is known by very few or by almost all, the proportion of "infectives" changes very slowly.

The Wright-Fisher Model in Population Genetics

- This is the discrete-time Markov chain on $\{0, 1, \ldots, m\}$ with transition probabilities

$$p_{ij} = \binom{m}{j} \left(\frac{i}{m}\right)^j \left(\frac{m-i}{m}\right)^{m-j}.$$

- In each generation there are m alleles.
- Some are of type A and some of type a.
- The types of alleles in generation $n+1$ are found by choosing randomly (with replacement) from the types in generation n.

The Wright-Fisher Model (Cont'd)

- Let X_n denote the number of alleles of type A in generation n.
- Then $(X_n)_{n \geq 0}$ is a Markov chain with transition probabilities p_{ij}.
- This can be viewed as a model of inheritance for a particular gene with two alleles A and a.
- We suppose that each individual has two genes.
- So the possibilities are AA, Aa and aa.
- Let us take m to be even with $m = 2k$.

The Wright-Fisher Model (Cont'd)

- Suppose that:
  - Individuals in the next generation are obtained by mating randomly chosen individuals from the current generation;
  - Offspring inherit one allele from each parent.
- We allow that both parents may be the same.
- In particular, it is not required that parents be of opposite sex.
- E.g., assume generation n is

  $$\text{AA} \quad \text{aA} \quad \text{AA} \quad \text{AA} \quad \text{aa}.$$

  Then each gene in generation $n+1$ is, independently:
  - A with probability $\frac{7}{10}$;
  - a with probability $\frac{3}{10}$.

  We might, for example, get

  $$\text{aa} \quad \text{aA} \quad \text{Aa} \quad \text{AA} \quad \text{AA}.$$

- The structure of pairs of genes is irrelevant to $(X_n)_{n\geq 0}$.
- $(X_n)_{n\geq 0}$ counts the number of alleles of type A.

The Wright-Fisher Model (Absorbing and Transient States)

- The communicating classes of $(X_n)_{n\geq 0}$ are $\{0\}$, $\{1,\ldots,m-1\}$, $\{m\}$.
- States 0 and m are absorbing and $\{1,\ldots,m-1\}$ is transient.
- The hitting probabilities for state m (pure AA) are given by

$$h_i = \mathbb{P}_i(X_n = m \text{ for some } n) = \frac{i}{m}.$$

- This can be seen by noticing that $(X_n)_{n\geq 0}$ is a martingale.
- Alternatively one can check that

$$h_i = \sum_{j=0}^{m} p_{ij} h_j.$$

- According to this model, genetic diversity eventually disappears.

The Moran Model

- The Moran model is the birth-and-death chain on $\{0, 1, \ldots, m\}$ with transition probabilities

$$p_{i,i-1} = \frac{i(m-i)}{m^2}, \quad p_{ii} = \frac{i^2 + (m-i)^2}{m^2}, \quad p_{i,i+1} = \frac{i(m-i)}{m^2}.$$

- It has the following genetic interpretation.
- A population consists of individuals of two types, a and A.
- At time n:
  - We choose randomly one individual from the population;
  - We add a new individual of the same type;
  - Then we choose, again randomly, one individual from the population;
  - We remove the chosen individual.
- In this way, we obtain the population at time $n+1$.
- The same individual may be chosen to give birth and to die.
- In this case there is no change in the make-up of the population.

Differences and Similarities with Wright-Fisher

- Let X_n denote the number of type A individuals at time n.
- Then $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix P.
- There are some differences from the Wright-Fisher model.
    - The Moran model cannot be interpreted in terms of a species where genes come in pairs, or where individuals have more than one parent;
    - In the Moran model we only change one individual at a time, not the whole population.
- The basic Markov chain structure is the same.
    - The communicating classes are $\{0\}$, $\{1,\ldots,m-1\}$, $\{m\}$, absorbing states 0 and m and transient class $\{1,\ldots,m-1\}$;
    - The Moran model is reversible, and, like the Wright-Fisher model, is a martingale.
    - The hitting probabilities are given by $\mathbb{P}_i(X_n = m \text{ for some } n) = \frac{i}{m}$.

Mean Time of Absorption

- We can also calculate explicitly the mean time to absorption
$$k_i = \mathbb{E}_i(T),$$
where T is the hitting time of $\{0, m\}$.
- The simplest method is to:
  - Fix j;
  - Write equations for the mean time k_i^j spent in j, starting from i, before absorption.
$$\begin{aligned} k_i^j &= \delta_{ij} + (p_{i,i-1}k_{i-1}^j + p_{ii}k_i^j + p_{i,i+1}k_{i+1}^j), \ i = 1, \ldots, m-1; \\ k_0^j &= k_m^j = 0. \end{aligned}$$
- Then, for $i = 1, \ldots, m-1$
$$k_{i+1}^j - 2k_i^j + k_{i-1}^j = -\delta_{ij}\frac{m^2}{j(m-j)}.$$

Mean Time of Absorption (Cont'd)

- We found, for $i = 1, \ldots, m-1$,
$$k_{i+1}^j - 2k_i^j + k_{i-1}^j = -\delta_{ij}\frac{m^2}{j(m-j)}.$$
- This has solution
$$k_i^j = \begin{cases} \frac{i}{j}k_j^j & \text{for } i \leq j \\ \frac{m-i}{m-j}k_j^j & \text{for } i \geq j \end{cases}.$$

Mean Time of Absorption (Cont'd)

- k_j^j is determined by

$$\left(\frac{m-j-1}{m-j} - 2 + \frac{j-1}{j}\right) k_j^j = -\frac{m^2}{j(m-j)}.$$

- This gives

$$k_j^j = m.$$

- Hence,

$$k_i = \sum_{j=1}^{m-1} k_i^j = m\left\{\sum_{j=1}^{i} \frac{m-i}{m-j} + \sum_{j=i+1}^{m-1} \frac{i}{j}\right\}.$$

The Case of Large m and $i = pm$, $0 < p < 1$

- The main interest lies in the case where:
  - m is large;
  - $i = pm$, for some $p \in (0,1)$.
- Then, as $m \to \infty$,

$$\begin{aligned} \frac{k_{pm}}{m^2} &= (1-p)\sum_{j=1}^{mp} \frac{1}{m-j} + p\sum_{j=mp+1}^{m-1} \frac{1}{j} \\ &\to -(1-p)\log(1-p) - p\log p. \end{aligned}$$

- So, as $m \to \infty$,

$$\mathbb{E}_{pm}(T) \sim -m^2\{(1-p)\log(1-p) + p\log p\}.$$

The Case of Large m and $i = pm$, $0 < p < 1$

- We found
$$\mathbb{E}_{pm}(T) \sim -m^2\{(1-p)\log(1-p) + p\log p\}.$$
- For the Wright-Fisher model, one has
$$\mathbb{E}_{pm}(T) \sim -2m\{(1-p)\log(1-p) + p\log p\}.$$
- This has the same functional form in p and differs by a factor of $\frac{m}{2}$.
- This factor is partially explained by the fact that:
  - The Moran model deals with one individual at a time;
  - The Wright-Fisher model changes all m at once.

Subsection 2

Queues and Queueing Networks

Basic Elements of Queues

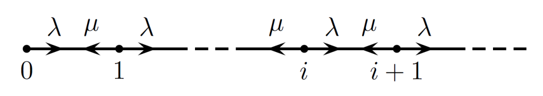
- Queues form in many circumstances and it is important to be able to predict their behavior.
- The basic mathematical model for queues runs as follows.
  - There is a succession of customers wanting service;
  - On arrival each customer must wait until a server is free, giving priority to earlier arrivals;
  - Probabilistically, it is assumed that:
    - The times between arrivals are independent random variables of the same distribution;
    - The times taken to serve customers are also independent random variables, of some other distribution.
- The main quantity of interest is the random process $(X_t)_{t\geq 0}$ recording the number of customers in the queue at time t.
- This is always taken to include both those being served and those waiting to be served.

Preview of Some Special Cases

- In cases where inter-arrival times and service times have exponential distributions, $(X_t)_{t\geq 0}$ turns out to be a *continuous-time Markov chain*.
- In this case many questions about the queue can be answered.
- If the inter-arrival times only are exponential, an analysis is still possible.
- One exploits:
  - The memorylessness of the *Poisson process* of arrivals;
  - A certain *discrete-time Markov chain* embedded in the queue.

M/M/1 Queue

- M/M/1 means *memoryless inter-arrival times*/*memoryless service times*/*one server*.
- Let us suppose that:
  - The inter-arrival times are exponential of parameter λ;
  - The service times are exponential of parameter μ.
- Then the number of customers in the queue $(X_t)_{t\geq 0}$ evolves as a Markov chain with the following diagram:



Markov Chain Justification

- Suppose at time 0 there are $i > 0$ customers in the queue.
- Denote by:
  - T the time taken to serve the first customer;
  - A the time of the next arrival.
- Then the first jump time J_1 is $A \wedge T$.
- This is exponential of parameter $\lambda + \mu$.
- Moreover,

$$X_{J_1} = \begin{cases} i-1, & \text{if } T < A, \\ i+1, & \text{if } T > A. \end{cases}$$

  These events are independent of J_1, with probabilities:

$$\mathbb{P}(T < A) = \frac{\mu}{\lambda + \mu} \quad \text{and} \quad \mathbb{P}(T > A) = \frac{\lambda}{\lambda + \mu}.$$

Markov Chain Justification (Cont'd)

- If we condition on $J_1 = T$, then $A - J_1$ is exponential of parameter λ and independent of J_1.
- The time already spent waiting for an arrival is forgotten.
- Similarly, conditional on $J_1 = A$, $T - J_1$ is exponential of parameter μ and independent of J_1.
- The case where $i = 0$ is simpler as there is no serving going on.
- Hence, conditional on $X_{J_1} = j$, $(X_t)_{t \geq 0}$ begins afresh from j at time J_1.
- It follows that $(X_t)_{t \geq 0}$ is the claimed Markov chain.

Average Number of Customers

- The M/M/1 queue evolves like a random walk, except that it does not take jumps below 0.
  - Suppose $\lambda > \mu$.
    Then $(X_t)_{t\geq 0}$ is transient, that is $X_t \to \infty$ as $t \to \infty$.
    Thus, if $\lambda > \mu$ the queue grows without limit in the long term.
  - Suppose, next, $\lambda < \mu$.
    Then $(X_t)_{t\geq 0}$ is positive recurrent with invariant distribution

$$\pi_i = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^i.$$

    So when $\lambda < \mu$, the average number of customers in the queue in equilibrium is given by

$$\mathbb{E}_\pi(X_t) = \sum_{i=1}^{\infty} \mathbb{P}_\pi(X_t \geq i) = \sum_{i=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^i = \frac{\lambda}{\mu - \lambda}.$$

Mean Continuously Busy Time

- The mean time to return to 0 is given by

$$m_0 = \frac{1}{q_0\pi_0} = \frac{1}{\lambda}\frac{1}{\frac{\mu-\lambda}{\mu}} = \frac{\mu}{\lambda(\mu-\lambda)}.$$

- So the *mean length of time that the server is continuously busy* is given by

$$m_0 - \frac{1}{q_0} = \frac{\mu}{\lambda(\mu-\lambda)} - \frac{1}{\lambda} = \frac{1}{\mu-\lambda}.$$

Mean Waiting Time

- Another quantity of interest is the *mean waiting time for a typical customer*, when $\lambda < \mu$ and the queue is in equilibrium.
- Conditional on finding a queue of length i on arrival, this is
$$\frac{i+1}{\mu}.$$
- So the overall mean waiting time is
$$\mathbb{E}_\pi \frac{X_t + 1}{\mu} = \frac{\frac{\lambda}{\mu-\lambda} + 1}{\mu} = \frac{\frac{\mu}{\mu-\lambda}}{\mu} = \frac{1}{\mu - \lambda}.$$

Mean Total Waiting Time

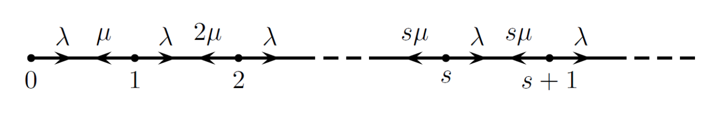
- A rough check is available as we can calculate in two ways the *expected total time spent in the queue* over an interval of length t.
  - We may multiply the average queue length by t.
  - We get
$$t \cdot \mathbb{E}_\pi(X_t) = \frac{\lambda t}{\mu - \lambda}.$$
  - We may multiply the mean waiting time by the expected number of customers λt.
    We get
$$\lambda t \cdot \mathbb{E}_\pi \frac{X_t + 1}{\mu} = \frac{\lambda t}{\mu - \lambda}.$$
- The first calculation is exact but the second has not been fully justified.

M/M/s Queue

- This is a variation on the last example where:
  - There is one queue;
  - There are s servers.
- We assume that:
  - The arrival rate is λ;
  - The service rate by each server is μ.

The Associated Markov Chain

- If i servers are occupied, the first service is completed at the minimum of i independent exponential times of parameter μ.
- The first service time is therefore exponential of parameter $i\mu$.
- The total service rate increases to a maximum $s\mu$ when all servers are working.
- Suppose the queue size includes the customers being served.
- Then the queue size $(X_t)_{t\geq 0}$ performs a Markov chain with the following diagram:



- So this time we obtain a birth-and-death chain.

Invariant Measures

- If $\lambda > s\mu$, the birth-and-death chain is transient; Otherwise, recurrent.
- To find an invariant measure we look at the detailed balance equations

$$\pi_i q_{i,i+1} = \pi_{i+1} q_{i+1,i}.$$

  - If $i \leq s$,

$$\frac{\pi_i}{\pi_0} = \frac{\pi_i}{\pi_{i-1}} \cdots \frac{\pi_1}{\pi_0} = \frac{\lambda}{i\mu} \cdots \frac{\lambda}{\mu} = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}.$$

  - Similarly, if $i > s$,

$$\frac{\pi_i}{\pi_0} = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{s^{i-s} s!}.$$

- So we have

$$\frac{\pi_i}{\pi_0} = \begin{cases} \frac{(\lambda/\mu)^i}{i!}, & \text{for } i = 0, 1, \ldots, s \\ \frac{(\lambda/\mu)^i}{s^{i-s} s!}, & \text{for } i = s+1, s+2, \ldots \end{cases}$$

- The queue is therefore positive recurrent when $\lambda < s\mu$.

Special Cases

- There are two special cases when the invariant distribution has a particularly nice form.
  - Suppose $s = 1$.
    Then we are back to the preceding example.
    The invariant distribution is geometric of parameter $\frac{\lambda}{\mu}$,
    $$\pi_i = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^i.$$
  - Suppose $s = \infty$.
    We normalize π by taking $\pi_0 = e^{-\lambda/\mu}$.
    Then
    $$\pi_i = e^{-\lambda/\mu}\frac{(\lambda/\mu)^i}{i!}.$$
    So the invariant distribution is Poisson of parameter $\frac{\lambda}{\mu}$.

Number of Arrivals and Departures

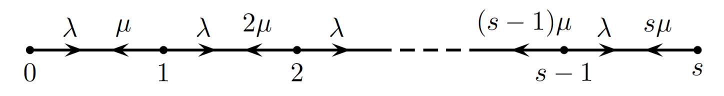
- The number of arrivals by time t is a Poisson process of rate λ.
- Each arrival corresponds to an increase in X_t.
- Each departure corresponds to a decrease in X_t.
- Suppose that $\lambda < s\mu$, so there is an invariant distribution.
- Consider the queue in equilibrium.
- The detailed balance equations hold.
- Moreover, $(X_t)_{t\geq 0}$ is non-explosive.
- So by a previous theorem, for any $T > 0$, $(X_t)_{0\leq t\leq T}$ and $(X_{T-t})_{0\leq t\leq T}$ have the same law.
- It follows that, in equilibrium, the number of departures by time t is also a Poisson process of rate λ.

Comments

- The fact that the number of departures by time t is also a Poisson process of rate λ is slightly counter-intuitive.
- One might imagine that the departure process runs in fits and starts depending on the number of servers working.
- But it turns out that the process of departures, in equilibrium, is just as regular as the process of arrivals.

A Telephone Exchange

- A variation on the M/M/s queue is to turn away customers who cannot be served immediately.
- This might serve as a simple model for a telephone exchange, where the maximum number of calls that can be connected at once is s.
- When the exchange is full, additional calls are lost.
- The maximum queue size or buffer size is s.
- We get the following modified Markov chain diagram:



A Telephone Exchange (Cont'd)

- We can find the invariant distribution of this finite Markov chain by solving the detailed balance equations.
- This time we get a **truncated Poisson distribution**

$$\pi_i = \frac{\frac{(\lambda/\mu)^i}{i!}}{\sum_{j=0}^{s} \frac{(\lambda/\mu)^j}{j!}}.$$

- The long run proportion of time that the exchange is full equals the long run proportion of calls that are lost.
- By the Ergodic Theorem, it is given by

$$\pi_s = \frac{\frac{(\lambda/\mu)^s}{s!}}{\sum_{j=0}^{s} \frac{(\lambda/\mu)^j}{j!}}.$$

- This is known as **Erlang's formula**.

Queues in Series

- Suppose that arriving customers have two service requirements.
  - They arrive as a Poisson process of rate λ;
  - They are seen first by server A;
  - They are seen then by server B.
- For simplicity, we assume that the service times are independent exponentials.
  - Service by A is an exponential of parameter α;
  - Service by B is an exponential of parameter β.
- We compute the average queue length at B.

Queues in Series (Cont'd)

- Let $(X_t)_{t\geq 0}$ be the queue length at A.
- Let $(Y_t)_{t\geq 0}$ be the queue length at B.
- Then $(X_t)_{t\geq 0}$ is simply an M/M/1 queue.
  - Suppose $\lambda > \alpha$.
    Then $(X_t)_{t\geq 0}$ is transient.
    So there is eventually always a queue at A.
    Moreover, departures form a Poisson process of rate α.
  - Suppose $\lambda < \alpha$.
    Then, by the reversibility argument of a previous example, the process of departures from A is Poisson of rate λ, provided queue A is in equilibrium.

Queues in Series (Cont'd)

- The question about queue length at B is not precisely formulated.
- One needs to specify that the queues should be in equilibrium.
- If $\lambda \geq \alpha$, there is no equilibrium.
- We may treat arrivals at B as a Poisson process of rate $\alpha \wedge \lambda$.
  - Suppose $\alpha \wedge \lambda < \beta$.
    By a previous example, the average queue length at B, when in equilibrium, is given by
    $$\frac{\alpha \wedge \lambda}{\beta - (\alpha \wedge \lambda)}.$$
  - Suppose $\alpha \wedge \lambda > \beta$.
    Then $(Y_t)_{t \geq 0}$ is transient.
    Now the queue at B grows without limit.

Queues in Series (Cont'd)

- There is an equilibrium for both queues if $\lambda < \alpha$ and $\lambda < \beta$.
- The fact that, in equilibrium, the output from A is Poisson greatly simplifies the analysis of the two queues in series.
- For example, the average time taken by one customer to obtain both services is given by

$$\frac{1}{\alpha - \lambda} + \frac{1}{\beta - \lambda}.$$

Closed Migration Process

- Consider a single particle in a finite state-space I which performs a Markov chain with irreducible Q-matrix Q.
- We know there is a unique invariant distribution π.
- The holding times of the chain may be thought of as service times, by a single server at each node $i \in I$.
- Suppose that there are N particles in the state-space.
- They move as before except that they must queue for service at every node.
- Suppose we do not care to distinguish between the particles.
- Then this is a new process $(X_t)_{t\geq 0}$ with state-space $\widetilde{I} = \mathbb{N}^I$.
- $X_t = (n_i : i \in I)$ if at time t there are n_i particles at state i.
- In fact, this new process is also a Markov chain.

Q-Matrix of the Closed Migration Process

- Suppose $(X_t)_{t\geq 0}$ has Q-matrix $\widetilde{Q}$.
- Define a function $\delta_i : \widetilde{I} \to \widetilde{I}$ by

$$(\delta_i n)_j = n_j + \delta_{ij}.$$

- Thus, δ_i adds a particle at i.
- Then, for $i \neq j$, the non-zero transition rates are given by

$$\widetilde{q}(\delta_i n, \delta_j n) = q_{ij}, \quad n \in \widetilde{I},\ i, j \in I.$$

- Observe that we can write the invariant measure equation $\pi Q = 0$ in the form

$$\pi_i \sum_{j \neq i} q_{ij} = \sum_{j \neq i} \pi_j q_{ji}.$$

Invariant Measure

- For $n = (n_i : i \in I)$ we set
$$\widetilde{\pi}(n) = \prod_{i \in I} \pi_i^{n_i}.$$
- Then
$$\begin{array}{rcl} \widetilde{\pi}(\delta_i n) \sum_{j \neq i} \widetilde{q}(\delta_i n, \delta_j n) & = & \prod_{k \in I} \pi_k^{n_k} (\pi \sum_{j \neq i} q_{ji}) \\ & = & \prod_{k \in I} \pi_k^{n_k} (\sum_{j \neq i} \pi_j q_{ji}) \\ & = & \sum_{j \neq i} \widetilde{\pi}(\delta_j n) \widetilde{q}(\delta_j n, \delta_i n). \end{array}$$
- Given $m \in \widetilde{I}$, we can set $m = \delta_i n$, whenever $m_i \geq 1$.
- On summing the resulting equations we obtain
$$\widetilde{\pi}(m) \sum_{n \neq m} \widetilde{q}(m, n) = \sum_{n \neq m} \widetilde{\pi}(n) \widetilde{q}(n, m).$$
- So $\widetilde{\pi}$ is an invariant measure for $\widetilde{Q}$.

Communicating Classes

- The total number of particles is conserved.
- So $\widetilde{Q}$ has communicating classes

$$C_N = \left\{ n \in \widetilde{I} : \sum_{i \in I} n_i = N \right\}.$$

- The unique invariant distribution for the N-particle system is given by normalizing $\widetilde{\pi}$ restricted to C_N.

Open Migration Process

- We consider a modification of the last example.
- We make the following assumptions.
  - New customers, or particles, arrive at each node $i \in I$ at rate λ_i.
  - Customers receiving service at node i leave the network at rate μ_i.
- In this setting, like in a shopping center:
  - Customers enter the network;
  - They move from queue to queue according to a Markov chain;
  - Eventually, they leave.
- This model includes:
  - The closed system of the last example;
  - The queues in series of a previous example.

Formalism

- Let
$$X_t = (X_t^i : i \in I),$$
where X_t^i denotes the number of customers at node i at time t.
- $(X_t)_{t \geq 0}$ is a Markov chain in $\widetilde{I} = \mathbb{N}^I$.
- The non-zero transition rates are given, for all $n \in \widetilde{I}$ and distinct states $i, j \in I$, by:
  - $\widetilde{q}(n, \delta_i n) = \lambda_i$;
  - $\widetilde{q}(\delta_i n, \delta_j n) = q_{ij}$;
  - $\widetilde{q}(\delta_j n, n) = \mu_j$.
- We shall assume that:
  - $\lambda_i > 0$, for some i;
  - $\mu_j > 0$, for some j.
- Then $\widetilde{Q}$ is irreducible on $\widetilde{I}$.

Invariant Measure

- The system of invariant measure equations for an invariant measure is replaced here by

$$\pi_i \left(\mu_i + \sum_{j \neq i} q_{ij} \right) = \lambda_i + \sum_{j \neq i} \pi_j q_{ji}.$$

- This system has a unique solution, with $\pi_i > 0$ for all i.
- This may be seen by considering the invariant distribution for the extended Q-matrix $\overline{Q}$ on $I \cup \{\partial\}$ with off-diagonal entries

$$\overline{q}_{\partial j} = \lambda_j, \quad \overline{q}_{ij} = q_{ij}, \quad \overline{q}_{i\partial} = \mu_i.$$

- Summing over $i \in I$, we find

$$\sum_{i \in I} \pi_i \mu_i = \sum_{i \in I} \lambda_i.$$

Invariant Measure (Cont'd)

- As in the last example, for $n = (n_i : i \in I)$, set

$$\widetilde{\pi}(n) = \prod_{i \in I} \pi_i^{n_i}.$$

- Transitions from $m \in \widetilde{I}$ may be divided into:
  - Those where a new particle is added;
  - For each $i \in I$ with $m_i \geq 1$, those where a particle is moved from i to somewhere else.
- For the first sort of transition

$$\begin{aligned} \widetilde{\pi}(m) &= \sum_{j \in I} \widetilde{q}(m, \delta_j m) \\ &= \widetilde{\pi}(m) \sum_{j \in I} \lambda_j \\ &= \widetilde{\pi}(m) \sum_{j \in I} \pi_j \mu_j \\ &= \sum_{j \in I} \widetilde{\pi}(\delta_j m) \widetilde{q}(\delta_j m, m). \end{aligned}$$

Invariant Measure (Cont'd)

- For the second sort,

$$\begin{aligned}&\widetilde{\pi}(\delta_i n)(\widetilde{q}(\delta_i n, n) + \textstyle\sum_{j\neq i} \widetilde{q}(\delta_i n, \delta_j n))\\ &= \textstyle\prod_{k\in I} \pi_k^{n_k}(\pi_i(\mu_i + \sum_{j\neq i} q_{ij}))\\ &= \textstyle\prod_{k\in I} \pi_k^{n_k}(\lambda_i + \sum_{j\neq i} \pi_j q_{ji})\\ &= \widetilde{\pi}(n)\widetilde{q}(n, \delta_i n) + \textstyle\sum_{j\neq i} \widetilde{\pi}(\delta_j n)\widetilde{q}(\delta_j n, \delta_i n).\end{aligned}$$

- On summing these equations, we obtain

$$\widetilde{\pi}(m)\sum_{n\neq m}\widetilde{q}(m,n) = \sum_{n\neq m}\widetilde{\pi}(n)\widetilde{q}(n,m).$$

- So $\widetilde{\pi}$ is an invariant measure for $\widetilde{Q}$.

Queue Lengths

- Suppose $\pi_i < 1$, for all i.
- Then $\widetilde{\pi}$ has finite total mass

$$\prod_{i \in I} (1 - \pi_i).$$

- Otherwise the total mass if infinite.
- Hence, $\widetilde{Q}$ is positive recurrent if and only if $\pi_i < 1$ for all i.
- In that case, in equilibrium, the individual queue lengths $(X_t^i : i \in I)$ are independent geometric random variables with

$$\mathbb{P}(X_t^i = n_i) = (1 - \pi_i)\pi_i^{n_i}.$$

M/G/1 Queue

- The service requirements have often observable distributions which are generally not exponential.
- A better model in this case is the M/G/1 queue, where G indicates that the service-time distribution is general.
- We can characterize the distribution of a service time T in one of two ways.
  - By its distribution function $F(t) = \mathbb{P}(T \leq t)$;
  - By its Laplace transform

$$L(w) = \mathbb{E}(e^{-wT}) = \int_0^\infty e^{-wt} dF(t).$$

- This integral is the Lebesgue-Stieltjes integral.
- When T has a density function $f(t)$ we can replace $dF(t)$ by $f(t)dt$.
- Then the mean service time μ is given by

$$\mu = \mathbb{E}(T) = -L'(0+).$$

Formalism

- Let X_n be the queue size immediately following the n-th departure.
- Let Y_n be the number of arrivals during the n-th service time.
- Then
$$X_{n+1} = X_n + Y_{n+1} - 1_{X_n > 0}.$$
- The case where $X_n = 0$ is different because then we get an extra arrival before the $(n+1)$-th service time begins.
- By the Markov Property of the Poisson process, $Y_1, Y_2, \ldots$ are independent and identically distributed.
- It follows that $(X_n)_{n \geq 0}$ is a discrete time Markov chain.
- Indeed, except for visits to 0, $(X_n)_{n \geq 0}$ behaves as a random walk with jumps $Y_n - 1$.

Service Intensity and Generating Function

- Let T_n denote the nth service time.
- Conditional on $T_n = t$, Y_n is Poisson of parameter λt.
- So
$$\mathbb{E}(Y_n) = \int_0^\infty \lambda t dF(t) = \lambda\mu.$$
- $\rho = \mathbb{E}(Y_n)$ is termed the **service intensity**.
- We can compute the probability generating function
$$\begin{aligned} A(z) &= \mathbb{E}(z^{Y_n}) \\ &= \int_0^\infty E(z^{Y_n} | T_n = t) dF(t) \\ &= \int_0^\infty e^{-\lambda t(1-z)} dF(t) \\ &= L(\lambda(1-z)). \end{aligned}$$

Positive Recurrence

- We set $\rho = \mathbb{E}(Y_n) = \lambda\mu$.
- Suppose $\rho < 1$.
- Let Z_n be the number of visits of X_n to 0 before time n.
- Then we have

$$X_n = X_0 + (Y_1 + \cdots + Y_n) - n + Z_n.$$

- So

$$\mathbb{E}(X_n) = \mathbb{E}(X_0) - n(1-\rho) + \mathbb{E}(Z_n).$$

- Take $X_0 = 0$.
- Since $X_n \geq 0$, for all n, we have $0 < 1 - \rho \leq \mathbb{E}(Z_n/n)$.
- By the Ergodic Theorem, as $n \to \infty$, $\mathbb{E}(Z_n/n) \to \frac{1}{m_0}$, where m_0 is the mean return time to 0.
- Hence, $m_0 \leq \frac{1}{1-\rho} < \infty$, showing that $(X_n)_{n\geq 0}$ is positive recurrent.

Equilibrium

- Suppose we start $(X_n)_{n\geq 0}$ with its equilibrium distribution π.
- Set
$$G(z) = \mathbb{E}(z^{X_n}) = \sum_{i=0}^{\infty} \pi_i z^i.$$
- Then
$$\begin{aligned} zG(z) &= \mathbb{E}(z^{X_{n+1}+1}) \\ &= \mathbb{E}(z^{X_n+Y_{n+1}+1_{X_n=0}}) \\ &= \mathbb{E}(z^{Y_{n+1}})(\pi_0 z + \textstyle\sum_{i=1}^{\infty} \pi_i z^i) \\ &= A(z)(\pi_0 z + G(z) - \pi_0). \end{aligned}$$
- So
$$(A(z) - z)G(z) = \pi_0 A(z)(1 - z).$$

Equilibrium (Cont'd)

- We obtained $(A(z)-z)G(z)=\pi_0 A(z)(1-z)$.
- Rewrite $\frac{A(z)-z}{1-z}=\frac{\pi_0 A(z)}{G(z)}$.
- Equivalently,

$$\frac{A(1)-z}{1-z}-\frac{A(1)-A(z)}{1-z}=\frac{\pi_0 A(z)}{G(z)}.$$

  - As $z\nearrow 1$, the left approaches $1-A'(1-)=1-\rho$.
  - As $z\nearrow 1$, since $G(1)=1=A(1)$, the right approaches π_0.
- So we must have:
  - $\pi_0=1-\rho$;
  - $m_0=\frac{1}{1-\rho}$;
  - $G(z)=(1-\rho)(1-z)\frac{A(z)}{A(z)-z}$.
- A is given explicitly in terms of the service time distribution.
- So we can now obtain, in principle, the full equilibrium distribution.

Mean Queue Length

- We now obtain the **mean queue length**.
- Start again with $(A(z)-z)G(z)=\pi_0 A(z)(1-z)$.
- Differentiate, recalling $\pi_0=1-\rho$,

$$(A(z)-z)G'(z)+(A'(z)-1)G(z)=(1-\rho)\{A'(z)(1-z)-A(z)\}.$$

- Substitute $G(z)=(1-\rho)(1-z)\frac{A(z)}{A(z)-z}$ to obtain

$$\begin{array}{rcl} G'(z) & = & -\frac{A'(z)-1}{A(z)-z}G(z)+(1-\rho)\frac{A'(z)(1-z)-A(z)}{A(z)-z} \\ & = & -(1-\rho)(1-z)\frac{A(z)(A'(z)-1)}{(A(z)-z)^2}+(1-\rho)\frac{A'(z)(1-z)-A(z)}{A(z)-z} \\ & = & (1-\rho)A'(z)\frac{1-z}{A(z)-z}-(1-\rho)A(z)\frac{(A'(z)-1)(1-z)+A(z)-z}{(A(z)-z)^2}. \end{array}$$

- Now note, using l'Hôpital's Rule, that:
  - $\lim_{z\nearrow 1}(1-\rho)A'(z)\frac{1-z}{A(z)-z}=(1-\rho)\rho\frac{1}{1-\rho}=\rho$;
  - $\lim_{z\nearrow 1}\frac{(A'(z)-1)(1-z)+A(z)-z}{(A(z)-z)^2}=\lim_{z\nearrow 1}\frac{A''(z)(1-z)}{2(A'(z)-1)(A(z)-z)}=\frac{A''(1-)}{2(1-\rho)^2}$.

Mean Queue Length (Cont'd)

- We found:
  - $G'(z) = (1-\rho)A'(z)\frac{1-z}{A(z)-z} - (1-\rho)A(z)\frac{(A'(z)-1)(1-z)+A(z)-z}{(A(z)-z)^2}$;
  - $\lim_{z\nearrow 1}(1-\rho)A'(z)\frac{1-z}{A(z)-z} = \rho$;
  - $\lim_{z\nearrow 1}\frac{(A'(z)-1)(1-z)+A(z)-z}{(A(z)-z)^2} = \frac{A''(1-)}{2(1-\rho)^2}$.
- Now we obtain

$$\begin{aligned}\mathbb{E}(X_n) &= G'(1-) \\ &= \rho + \frac{A''(1-)}{2(1-\rho)} \\ &= \rho + \lambda^2\frac{L''(0+)}{2(1-\rho)} \\ &= \rho + \lambda^2\frac{\mathbb{E}(T^2)}{2(1-\rho)}.\end{aligned}$$

- In the case of the M/M/1 queue, we have
  - $\rho = \frac{\lambda}{\mu}$;
  - $\mathbb{E}(T^2) = \frac{2}{\mu^2}$.

  Consequently, $\mathbb{E}(X_n) = \rho + \frac{(\lambda/\mu)^2}{1-(\lambda/\mu)} = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$.

Mean Queueing Time

- Consider the queue $(X_n)_{n\in\mathbb{Z}}$ in equilibrium.
- Suppose that the customer who leaves at time 0 has spent:
  - Time Q queueing to be served;
  - Time T being served.
- Note that the customers in the queue at time 0 are precisely those who arrived during the queueing and service times of the departing customer.
- So, conditional on $Q + T = t$, X_0 is Poisson of parameter λt.
- Hence,

$$G(z) = \mathbb{E}(e^{-\lambda(Q+T)(1-z)}) = M(\lambda(1-z))L(\lambda(1-z)),$$

where M is the Laplace transform $M(w) = \mathbb{E}(e^{-wQ})$.

Mean Queueing Time (Cont'd)

- We have
$$G(z) = M(\lambda(1-z))L(\lambda(1-z)),$$
where M is the Laplace transform $M(w) = \mathbb{E}(e^{-wQ})$.
- Recall that:
  - $A(z) = L(\lambda(1-z))$;
  - $G(z) = (1-\rho)(1-z)\frac{A(z)}{A(z)-z}$.
- Setting $w = \lambda(1-z)$, we obtain

$$M(w) = \frac{G(z)}{A(z)}$$
$$M(w) = \frac{(1-\rho)\frac{w}{\lambda}\frac{A(z)}{A(z)-(1-\frac{w}{\lambda})}}{A(z)}$$
$$M(w) = (1-\rho)\frac{w}{\lambda}\frac{1}{L(w)-(1-\frac{w}{\lambda})}$$
$$M(w) = (1-\rho)\frac{w}{w-\lambda(1-L(w))}.$$

Mean Queueing Time (Cont'd)

- We obtained

$$M(w) = (1-\rho)\frac{w}{w-\lambda(1-L(w))}.$$

- Differentiation and l'Hôpital's Rule yield the **mean queueing time**

$$\begin{aligned}
\mathbb{E}(Q) &= -M'(0+) \\
&= -(1-\rho)\lim_{w\to 0^+}\tfrac{w+\lambda L(w)-\lambda-w(1+\lambda L'(w))}{(w+\lambda L(w)-\lambda)^2} \\
&= -(1-\rho)\lim_{w\to 0^+}\tfrac{\lambda L(w)-\lambda-w\lambda L'(w)}{(w+\lambda L(w)-\lambda)^2} \\
&= -(1-\rho)\lim_{w\to 0^+}\tfrac{\lambda L'(w)-\lambda L'(w)-\lambda w L''(w)}{2(w+\lambda L(w)-\lambda)(1+\lambda L'(w))} \\
&= -(1-\rho)\lim_{w\to 0^+}\tfrac{-\lambda L''(w)-\lambda w L'''(w)}{2(1+\lambda L'(w))^2+2\lambda L''(w)(w+\lambda L(w)-\lambda)} \\
&= -(1-\rho)\lim_{w\to 0^+}\tfrac{-\lambda L''(w)}{2(1+\lambda L'(w))^2} \\
&= (1-\rho)\tfrac{\lambda L''(0+)}{2(1+\lambda L'(0+))^2} = \tfrac{\lambda\mathbb{E}(T^2)}{2(1-\rho)}.
\end{aligned}$$

Busy Period

- We turn to the busy period S.
- Consider the Laplace transform

$$B(w) = \mathbb{E}(e^{-wS}).$$

- Let T be the service time of the first customer in the busy period.
- Let N be the number of customers arriving while the first customer is served.
- This is Poisson of parameter λt.
- Conditional on $T = t$, we have

$$S = t + S_1 + \cdots + S_N,$$

  where $S_1, S_2, \ldots$ are independent, with the same distribution as S.

Busy Period (Cont'd)

- Now we have

$$\begin{aligned} B(w) &= \int_0^\infty \mathbb{E}(e^{-wS} | T = t) dF(t) \\ &= \int_0^\infty e^{-wt} e^{-\lambda t(1-B(w))} dF(t) \\ &= L(w + \lambda(1 - B(w))). \end{aligned}$$

- Using $B(w)$, we can obtain moments by differentiation.

$$\begin{aligned} \mathbb{E}(S) &= -B'(0+) \\ &= -L'(0+)(1 - \lambda B'(0+)) \\ &= \mu(1 + \lambda\mathbb{E}(S)). \end{aligned}$$

- So the **mean length of the busy period** is given by $\mathbb{E}(S) = \frac{\mu}{1-\rho}$.

M/G/∞ Queue

- Arrivals at this queue form a Poisson process, of rate λ, say.
- Service times are independent, with a common distribution function

$$F(t) = \mathbb{P}(T \leq t).$$

- There are infinitely many servers.
- So all customers receive service at once.
- Suppose there are no customers at time 0.
- Let X_t be the number of customers being served at time t.
- Let N_t be the number of arrivals by time t.
- This is a Poisson random variable of parameter λt.
- Condition on $N_t = n$.
- Label the times of the n arrivals randomly by $A_1, \ldots, A_n$.
- By a previous theorem, $A_1, \ldots, A_n$ are independent and uniformly distributed on the interval $[0, t]$.

$M/G/\infty$ Queue (Cont'd)

- For each of these customers, service is incomplete at time t with probability

$$p = \frac{1}{t}\int_0^t \mathbb{P}(T > s)ds = \frac{1}{t}\int_0^t (1 - F(s))ds.$$

- Hence, conditional on $N_t = n$, X_t is binomial of parameters n and p. Then

$$\begin{aligned}
\mathbb{P}(X_t = k) &= \textstyle\sum_{n=0}^{\infty} \mathbb{P}(X_t = k | N_t = n)\mathbb{P}(N_t = n) \\
&= \textstyle\sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
&= e^{-\lambda t} \frac{(\lambda p t)^k}{k!} \textstyle\sum_{n=k}^{\infty} \frac{(\lambda(1-p)t)^{n-k}}{(n-k)!} \\
&= e^{-\lambda t} \frac{(\lambda p t)^k}{k!} e^{\lambda(1-p)t} = e^{-\lambda p t} \frac{(\lambda p t)^k}{k!}.
\end{aligned}$$

- So we have shown that X_t is Poisson of parameter $\lambda \int_0^t (1 - F(s))ds$.

M/G/∞ Queue (Cont'd)

- We have shown that X_t is Poisson of parameter $\lambda \int_0^t (1 - F(s))ds$.
- Recall that

$$\int_0^\infty (1 - F(s))ds = \int_0^\infty \mathbb{E}(1_{T>t})dt = \mathbb{E}\int_0^\infty 1_{T>t}dt = \mathbb{E}(T).$$

- Assume $\mathbb{E}(T) < \infty$.
- Then the queue size has a limiting distribution, which is Poisson of parameter

$$\lambda \mathbb{E}(T).$$