Queueing Systems

M.Sc. Course in Computer/Communication Engineering

1995/96. 2nd semester

Jereb László, Telek Miklós

Department of Telecommunication

Contents

Chapter 1

Introduction

1.1 Important information

1. Lecturers

   - Dept. of Telecommunications
   - Jereb László, associate prof.,
     St. 304., phone: 20-96, e-mail: jereb@hit.bme.hu
   - Telek Miklós, assistant prof.,
     St. 305., phone: 20-84, e-mail: telek@hit.bme.hu

2. Project laboratory

   - available tasks: queueing systems, transmission network planning, ATM traffic simulation, ..
   - opportunity for thesis, publications, international cooperation

1.2 The subject

1. Aim of the subject

   - to attain the base concept, the base models and the computation methods of queueing theory
   - to introduce application examples of queueing theory to real computer/communication systems

2. Preliminaries

   - Probability theory, Stochastic processes, Information theory Communication/Computer systems

3. Content

   - Motivations
   - Summary and extension of mathematical background
   - Basic queueing theory
   - Case studies

4. Requirements

   - short tests during the semester
   - written examination

5. Lessons

   - questions, interruptions are encouraged anytime

1.3 Short test on the preliminaries

1. X_d is a discrete and X_c is a continuous random variable.
    - What are their distribution functions?
    - What is the probability density function of X_c?
    - What is the distribution of X_d?
    - What are their means?
    - What are their n-th moments?
    - What are their variances?
2. X_1 is a random variable. It is exponentially distributed with parameter λ.
    - What is the probability density function of X_1 ?
    - What is the distribution function of X_1 ?
    - What is the probability that X_1 less than c (i.e. $\mathbf{P}(X_1 < c)$) ?
    - What is the mean of X_1 ?
3. X_2 is a random variable. $\mathbf{P}(X_2 = c) = 1$
    - What is the distribution function of X_2 ?
    - What is the variance of X_2 ?
4. X_1 and X_3 are exponentially distributed with parameter λ and μ respectively.
    - What is the distribution function of $X_4 = min(X_1, X_3)$?
    - What is the probability of $X_1 < X_3$
5. X_1 is exponentially distributed with parameter λ and X_2 is deterministic ($\mathbf{P}(X_2 = c) = 1$).
    - What is the distribution function of $X_5 = min(X_1, X_2)$?
    - What is the mean of X_5 ?
6. What does memoryless property of the exponential distribution mean?

1.4 Motivations

1. Information society

   - increasing number of customers
   - increasing distence between customers
     (international companies and cooperation)
   - increasing amount of information
   - different kinds of information
   - increasing dependence on information

2. Customer preferences

   - to reach anybody — anywhere
   - at anytime — without delay and limitations
   - with sufficient quality and reliability
   - anyway but
     at a low price !!

3. Problems

   - random demand arrivals (heavyly loaded and silent periods, non-stationary behaviour)
   - big distances $\longrightarrow$ technical and traffic conflicts
   - different demands (voice, data, video, etc.), and on top of all of them *multimedia*
   - different relations
     - point-to-point (telephone, E-mail, ..)
     - multipoint-to-point (database applications, ..)
     - point-to-multipoint (tv, radio broadcast, ..)
     - multipoint-to-multipoint ((video)conference)

4. Need for communication networks

   - impossible to set up an independent connection for every pair of customers
     - low utilization, high cost
     - complicated customer equipment
       $\longrightarrow$ difficult to upgrade, technically impossible
     - limited resources
   - reasonable solution is a communication network (Tivadar Puskás)
     - standardized access interfaces (the 7 OSI layer)
     - concentrated individual (small) traffic components (switches)
     - distributed network resources with high utilization
   - customers and service providers
     - lots of customers and service providers
     - with different interests and goals
     - competitions and cooperation between service providers
     - state controlled market (in Hungary), national and international standards and recommendations

   Everybody is a customer, but some of you can become a service provider as well!

5. The conflict of customers' and service providers' interests

   - customer wants oversized resources, planned for peak load
     but *(it would be expensive)*
   - the service provider wants high utilization of the resources, he plans for average load, but *(it restrict the customer sometimes)*

6. economic technical compromise should be find **that is why:**

   - **we need effective resource sharing algorithms**
     $\longrightarrow$ *multiuser communication protocols*
   - **we should be able to characterize the properties of these systems**
     $\longrightarrow$ *queueing theory*

Chapter 2

SUMMARY OF MATHEMATICAL BACKGROUND

2.1 Essential probability theory

1. concepts considered to be known:
   - continuous, discrete random variables (r.v.)
   - mean and higher moments of continuous, discrete random variables,
   - variance and standard deviation of continuous, discrete random variables,
   - conditional distribution, conditional mean value,
   - independence of random variables, joint distribution,
   - distribution of sum of random variables, (convolutions)
2. well-known discrete distributions:
   - geometric distribution: $\eta =$ the first occurrence of an event,

     $$p_n = \mathbf{P}(\eta = n) = p^{n-1}(1-p), \quad n > 0$$

     where p is the occurrence probability of the event in every step.

     $$\mathbf{P}(\eta > n) = \sum_{k=n+1}^{\infty} p^{k-1}(1-p) = p^n(1-p)\sum_{j=0}^{\infty} p^j = p^n$$

memoryless property:
$\mathbf{P}(\eta = k + \Delta k \mid \eta > k) =$

$$\frac{\mathbf{P}(\eta = k + \Delta k, \eta > k)}{\mathbf{P}(\eta > k)} = \frac{\mathbf{P}(\eta = k + \Delta k)}{\mathbf{P}(\eta > k)} = \frac{p^{k+\Delta k-1}(1-p)}{p^k} = p^{\Delta k-1}(1-p)$$

- Bernoulli distribution:

  $p_0 = \mathbf{P}(\text{occurence of one event}) = p$

  $p_1 = \mathbf{P}(\text{no event occurs}) = 1 - p$

  $p_{n|n\geq 2} = \mathbf{P}(\text{more than one events occur}) = 0$
- binomial distribution:

  $p_k = \mathbf{P}(\text{the event occur } k \text{ times out of } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k}$
- Poisson distribution:
  the limit of binomial distribution as $n \to \infty$, $p \to 0$ and $np \to \lambda$

$$p_k = \mathbf{P}(\text{occurence of k events}) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

  When events occur with a constant "event rate" λ the number of events occurs during the interval $(0, t)$ is Poisson distributed with parameter λt.
- deterministic distribution:

  $\mathbf{P}(\eta = c) = 1$, where c is an integer, i.e. $p_c = 1$ and $p_{k|k \neq c} = 0$

3. well-known continuous distributions:

- uniform distribution:

$$F_\eta(t) = \mathbf{P}(\eta \leq t) = \begin{cases} 0 & \text{if } t < a \\ \dfrac{t-a}{b-a} & \text{if } a \leq t \leq b \\ 1 & \text{if } b < t \end{cases}$$

  and $f_\eta(t) = \dfrac{dF_\eta(t)}{dt} = \dfrac{1}{b-a}$ when $a \leq t \leq b$ otherwise $f(t) = 0$.

- deterministic distribution:

  $\mathbf{P}(\eta = a) = 1$, where a is an positive real number, i.e. $F_\eta(t) = 0$ if $t < a$ and $F_\eta(t) = 1$ if $t \geq a$.

- exponential distribution: η = time of occurence of an event,

  $$\mathbf{P}(\eta \leq t) = 1 - \mathrm{e}^{-\lambda t} \qquad \mathbf{P}(\eta > t) = \mathrm{e}^{-\lambda t}$$

  Memoryless property: $\mathbf{P}(\eta > t + \Delta t \mid \eta > t) =$

  $$\frac{\mathbf{P}(\eta > t + \Delta t, \eta > t)}{\mathbf{P}(\eta > t)} = \frac{\mathbf{P}(\eta > t + \Delta t)}{\mathbf{P}(\eta > t)} = \frac{\mathrm{e}^{-\lambda(t+\Delta t)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda \Delta t}$$

4. distribution of the minimum of random variables:

   - $\tau_1, \ldots, \tau_n$ are random variables,
   - with distribution function: $F_i(t) = \mathbf{P}(\tau_i \ < \ t), \quad i = 1, ..., n$
   - let τ be their minimum, i.e. $\tau = \min\limits_i \tau_i$,
   - What is the distribution of τ ($F_\tau(t) = \mathbf{P}(\tau \ < \ t)$) ?
   - consider the fact that when $\tau < t$ then all τ_i should be less than t as well: $F_\tau(t) = \mathbf{P}(\tau \ < \ t) = \mathbf{P}(\min_i \ \tau_i \ < \ t) =$

     $$1 - \mathbf{P}(\min_i \ \tau_i \ \geq \ t) = 1 - \mathbf{P}(\tau_i \ \geq \ t, \forall i)$$

     When $\tau_i, \ i = 1, ..., n$ are independent:

     $$F_\tau(t) = 1 - \prod_{i=1}^{n} \mathbf{P}(\tau_i \ \geq \ t) = 1 - \prod_{i=1}^{n} (1 - \mathbf{P}(\tau_i \ < \ t) = 1 - \prod_{i=1}^{n} (1 - F_i(t))$$

     from which

     $$1 - F_\tau(t) = \prod_{i=1}^{n} (1 - F_i(t))$$

   - If $F_i(t) = 1 - \mathrm{e}^{-\lambda_i t}$, i.e. τ_i are independent exponentially distributed random variables,
     then $F_\tau(t) = 1 - \prod_{i=1}^{n} \mathrm{e}^{-\lambda_i t} = 1 - \mathrm{e}^{-\lambda t}$ is exponentially distributed as well with parameter $\lambda = \sum\limits_{i=1}^{n} \lambda_i$.

- If $\mathbf{P}(\tau_i = k) = p_i^{k-1}(1-p_i)$, i.e. τ_i are independent geometrically distributed random variables
  $\mathbf{P}(\tau > k) = \left(\prod_{i=1}^{n} p_i\right)^k = p^k$, from which τ is geometrically distributed as well,
  with parameter $p = \prod_{i=1}^{n} p_i$

5. distribution of the maximum of random variables:

   - by considering independent random variables similar steps result in:

$$F_{max}(t) = \prod_{i=1}^{n} \mathbf{P}(\tau_i \le t) = \prod_{i=1}^{n} F_i(t) \tag{2.1}$$

6. z-transform (transform of discrete series)
   If $\exists a > 0$, that $\lim_{n\to\infty} \frac{f_n}{a^n} = 0$, and $|z| < 1$, then exists the infinite sum

$$F(z) = \sum_{n=0}^{\infty} f_n z^n$$

   which is, by definition, the z-transform of the series f_n.
   Properties of the z-transform:

   - infinite sum: $F(1) = \sum_{n=0}^{\infty} f_n$
     or at least $\lim_{z\to 1} F(z) = \sum_{n=0}^{\infty} f_n$
   - initial value: $F(0) = f_0$
   - limit theorem: $\lim_{z\to 0}(1-z)F(z) = \lim_{n\to\infty} f_n$
   - inverse transform: $\frac{1}{k!}\frac{d^k}{dz^k}F(z)\Big|_{z=0} = f_k$
   - convolution: $h_n = f_n \otimes g_n = \sum_{k=0}^{n} f_k g_{n-k} \longrightarrow H(z) = F(z)\,G(z)$

7. z-transform of discrete random variable (generator function)

   - η is a discrete r.v. with distribution $p_k = \mathbf{P}(\eta = k)$, $k = 0, 1, 2, \cdots$, hence $\sum_{k=0}^{\infty} p_k = 1$.

- the generator function is defined as: $P(z)=\sum_{k=0}^{\infty} p_k z^k$
- by the properties of the z-transform:
  $P(1)=1,\ P(0)=p_0,\ \frac{1}{k!}\frac{d^k}{dz^k}P(z)\Big|_{z=0}=p_k$
- Mean of η: $\mathbf{E}(\eta) = P'(z)|_{z=1} = \frac{d}{dz}P(z)\Big|_{z=1} = \sum_{k=0}^{\infty} kp_k$
- Higher moments: $P''(z)|_{z=1} = \frac{d^2}{dz^2}P(z)\Big|_{z=1} = \sum_{k=0}^{\infty} k(k-1)p_k = \sum_{k=0}^{\infty} k^2 p_k - \sum_{k=0}^{\infty} kp_k = \mathbf{E}(\eta^2) - \mathbf{E}(\eta)$
- sum of discrete r.v. (convolutions):
  η, ν are independent r.v. with distribution f_n, g_n and $\rho=\eta+\nu$.
  The distribution of ρ (h_n) is given by
  $h_n = f_n \otimes g_n \longrightarrow H(z) \ = \ F(z)\ G(z)$

8. application examples of z-transformation (generator function)

- 'discrete' deterministic distribution:
  - $P(z)=z^n$, $\mathbf{E}(\eta)=n$
- geometric distribution:
  - $P(z)=\sum_{k=1}^{\infty}(1-p)^{k-1}pz^k = pz\sum_{k=1}^{\infty}(1-p)^{k-1}z^{k-1} = \frac{pz}{1-(1-p)z}$
  - $\mathbf{E}(\eta) = \frac{p(1-(1-p)z)+pz(1-p)}{(1-(1-p)z)^2}\Big|_{z=1} = \frac{p^2+p-p^2}{p^2} = \frac{1}{p}$
- Bernoulli distribution:

  $P(z)=1-p+pz, \quad \mathbf{E}(\eta)=p$
- binomial distribution:
  - $P(z)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}z^k = (1-p+pz)^n$,

    the sum of n pieces of independent identical Bernoulli distributed r.v.
  - $\mathbf{E}(\eta) = np(1-p+pz)^{n-1}\Big|_{z=1} = np$

- Poisson distribution:

$$P(z) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t} z^k = \mathrm{e}^{(z-1)\lambda t}, \quad \mathbf{E}(\eta) = \lambda t \mathrm{e}^{(z-1)\lambda t}\Big|_{z=1} = \lambda t$$

9. Laplace-transform :
(transform of continuous non-negative functions)

$$F^*(s) = \int_{0^-}^{\infty} \mathrm{e}^{-st} f(t) dt$$

10. Properties of Laplace-transform:

- Integral property: $F^*(0) = \int_{0^-}^{\infty} f(t)dt$
- Initial value: $\lim_{s \to \infty} sF^*(s) = \lim_{t \to 0} f(t)$
- Limit theorem: $\lim_{s \to 0} sF^*(s) = \lim_{t \to \infty} f(t)$,
- Convolutions: $h(t) = f(t) \otimes g(t) = \int_{u=0^-}^{t} f(t)g(t-u)du$
$\longrightarrow H^*(s) = F^*(s)\ G^*(s)$
- $\dfrac{df(t)}{dt} \to sF^*(s) - f(0+); \quad \int_{0^-}^{\infty} f(u)du \to \dfrac{1}{s} F^*(s) + c$

11. Application of Laplace transform to non-negative continuous r.v. (characteristic function)

- $\eta \geq 0$ is a nonnegative r.v., and $f_\eta(t)$ is its probability density function, and $F_\eta(t)$ is its cummulated density function. The characteristic function of η is defined as:

$$F^*(s) = \int_{0^-}^{\infty} \mathrm{e}^{-st} f(t) dt = \int_{0^-}^{\infty} \mathrm{e}^{-st} dF(t)$$

- Mean value: $\mathbf{E}(\eta) = -\dfrac{d}{ds} F^*(s)\Big|_{s=0}$
- Higher moments: $\mathbf{E}\left(\eta^k\right) = (-1)^k \dfrac{d^k}{ds^k} F^*(s)\Big|_{s=0}$
- sum of continuous r.v. (convolutions):
η, ν are independent r.v. with PDF f_n, g_n and $\rho = \eta + \nu$.
The PDF of ρ (h_n) is given by
$h(t) = f(t) \otimes g(t) \longrightarrow H^*(s) = F^*(s)\ G^*(s)$

12. application example of Laplace transform

   - 'continuous' deterministic distribution ($a > 0$):

   $$F^*(s) = \int_0^\infty \mathrm{e}^{-st} dF(t) = \mathrm{e}^{-sa}, \quad \mathbf{E}(\eta) = a$$

   - uniform distribution:

   $$F^*(s) = \int_a^b \mathrm{e}^{-st} \frac{1}{b-a} dt = \frac{1}{s(b-a)} (\mathrm{e}^{-sa} - \mathrm{e}^{-sb}), \; \mathbf{E}(\eta) = \frac{a+b}{2}$$

   - exponential distribution:

   $$F^*(s) = \int_0^\infty \mathrm{e}^{-st} \lambda \mathrm{e}^{-\lambda t} dt = \frac{\lambda}{s+\lambda}, \quad \mathbf{E}(\eta) = \left. \frac{\lambda}{(s+\lambda)^2} \right|_{s=0} = \frac{1}{\lambda}$$

2.2 Summary of discrete time Markov chains

1. Markov property: "the future is independent of the past conditioned on the persent"

$$\mathbf{P}(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = \mathbf{P}(X_n = x_n | X_{n-1} = x_{n-1}). \quad (2.2)$$

2. Discrete time Markov chains (DTMC):
   $X_0, X_1, X_2, ... \in S$ serie of discrete random variables is a Discrete time Markov chain, when the Markov property holds for all X_i.

3. from which:

$$\begin{aligned} &\mathbf{P}(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = \\ &\quad \mathbf{P}(X_n = x_n | X_{n-1} = x_{n-1}) \; \mathbf{P}(X_{n-1} = x_{n-1} | X_{n-2} = x_{n-2}) \ldots \\ &\quad \ldots \; \mathbf{P}(X_1 = x_1 | X_0 = x_0) \; \mathbf{P}(X_0 = x_0) \end{aligned} \quad (2.3)$$

4. An interpretation of DTMCs:
   We have a stochastic system which "stays" in a state $j \in S$ of the state space S in every /discrete/ time epoch.

5. By introducing:

   - (transient) state probability vector:

   $$P^{(n)} = \left[p_0^{(n)}, p_1^{(n)}, p_2^{(n)}, \cdots \right], \quad p_i^{(n)} = \mathbf{P}(X_n = i),$$

- n-step state transition probability matrix:

$$\boldsymbol{\Pi}^{(k)}(m) = \left[p_{ij}^{(k)}(m)\right], \quad p_{ij}^{(k)}(m) = \mathbf{P}(X_{m+k} = j \mid X_m = i),$$

  $\forall\ i, j \in S,\ m = 0, 1, 2, \cdots,\ k = 1, 2, \cdots.$

- Chapman-Kolmogorov equality:

$$p_{ij}^{(k+n)}(m) = \sum_{l \in S} p_{il}^{(k)}(m) p_{lj}^{(n)}(m+k); \quad \boldsymbol{\Pi}^{(k+n)}(m) = \boldsymbol{\Pi}^{(k)}(m)\boldsymbol{\Pi}^{(n)}(m+k) \tag{2.4}$$

We have, in matrix form:

$$P^{(n)} = P^{(0)}\boldsymbol{\Pi}^{(n)}(0) = \ldots = P^{(n-1)}\boldsymbol{\Pi}^{(1)}(n-1) \tag{2.5}$$

6. for homogeneous Markov chains:

$$\boldsymbol{\Pi}^{(1)}(m) = [p_{ij}] = \boldsymbol{\Pi}$$

$\forall\ i, j \in S,\ \forall m = 0, 1, \cdots.$ In this case

$$\boldsymbol{\Pi}^{(k+n)}(m) = \boldsymbol{\Pi}^{(k+n)} = \boldsymbol{\Pi}^{(k)}\boldsymbol{\Pi}^{(n)}, \text{ and } P^{(n)} = P^{(0)}\boldsymbol{\Pi}^{n} \tag{2.6}$$

7. main proporeties of DTMCs:

- irreducibility: "each states are reachable from each others"
  $\forall i, j\ \ \exists n < \infty$, that $p_{ij}^{(n)} > 0$
- aperiodicity: the DTMC is not periodic, i.e.
  i state is aperiodic, if $\exists n_0 < \infty$, that $p_{ii}^{(n)} > 0\ \ \forall n \geq n_0$.
- recurrence: positive (*finit expected recurrence time*)
  for detailes see Kleinrock.

$$f_{ij}^{(n)} = \mathbf{P}(X_n = j, X_k \neq j, 1 \leq k < n \mid X_0 = i), \quad (n > 0,\ i, j \in S)$$

  - $i \in S$ recurrent, if: $\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$
  - $i \in S$ non recurrent, if: $\sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$
  - $i \in S$ positive recurrent, if: $m_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)} < \infty$

- heredity: the aperiodicity and the recurrency is heritaged in irreducibile DTMCs.

8. the limit distribution exists and unique — i.e. the DTMC is stabil, if
   - a finit (state space S) DTMC: irreducibil and aperiodic
   - an infinite DTMC: irreducibil, aperiodic and
   - a sufficient condition of stability of infinite DTMCs is the — Foster criteria
     An irreducibil, aperiodic DTMC is stabil, if exist
     $I \geq 0,\ C > 0,\ d > 0$ numbers, such that

     for $k \leq I$: $\mathbf{E}(X_{n+1} \mid X_n = k) \leq C$
     and for $k > I$: $\mathbf{E}(X_{n+1} \mid X_n = k) \leq k - d$

9. Consequences of stability
   - $\lim_{n\to\infty} p_{ij}^{(n)} = p_j > 0,\ \forall i, j$, where $p_j = \lim_{n\to\infty} \mathbf{P}(X_n = j),$ for $\forall j$
   - in this case there is only one solution of the equation $P = P\boldsymbol{\Pi}$ ($P = [p_0, p_1, p_2, \cdots]$) with the additional condition $\sum_{i=0}^{\infty} p_i = 1$ which is the steady state of the DTMC.

     If $P^{(0)} = P$ then $P^{(n)} = P,$ for $\forall n$.

2.3 Analysis of DTMCs

We always consider a finite state space $S = \{1, 2, \ldots, N\}$ without mentioning anything else.

1. steady state distribution
   - applying the above expression for state j we have:

$$P_j^{(n)} = \sum_{i=0}^{N} P_i^{(n-1)} p_{ij} = P_j^{(n-1)} p_{jj} + \sum_{i \neq j} P_i^{(n-1)} p_{ij} =$$

$$P_j^{(n-1)} (1 - \sum_{k \neq j} p_{jk}) + \sum_{i \neq j} P_i^{(n-1)} p_{ij}$$

from which

$$P_j^{(n)} - P_j^{(n-1)} = \sum_{i \neq j} P_i^{(n-1)} p_{ij} - P_j^{(n-1)} \sum_{k \neq j} p_{jk} \tag{2.7}$$

As n goes to infinity $\lim_{n\to\infty} P_j^{(n)} = \lim_{n\to\infty} P_j^{(n-1)} = p_j$, for $\forall j$

$$p_j - p_j = 0 = \sum_{i \neq j} p_i p_{ij} - p_j \sum_{k \neq j} p_{jk}$$

i.e. (*the probability of entering into state j equals to the probability of exiting from state j*)

$$\sum_{i \neq j} p_i p_{ij} = p_j \sum_{k \neq j} p_{jk} \tag{2.8}$$

- the same holds for state groups as well

$$\sum_{i \in U} p_i \sum_{j \in D} p_{ij} = \sum_{j \in D} p_j \sum_{k \in U} p_{jk} \tag{2.9}$$

2. random walks:

- interpretation

$$p_{ij} = \begin{cases} b_i & j = i+1,\ 0 \leq i < N \\ d_i & j = i-1,\ 0 < i \leq n \\ 1 - b_i - d_i & j = i,\ 0 < i < N \\ 1 - b_0 & j = i,\ i = 0 \\ 1 - d_N & j = i,\ i = N \\ 0 & \text{otherwise} \end{cases}$$

- steady state of random walks

$$p_{k-1} p_{k-1,k} + p_{k+1} p_{k+1,k} = p_k (p_{k,k-1} + p_{k,k+1}),\ 0 < k \leq N$$
$$p_{k-1} b_{k-1} + p_{k+1} d_{k+1} = p_k (b_k + d_k), \quad 0 < k \leq N$$
$$p_1 d_1 = p_0 b_0$$

from which

$$p_k d_k = p_{k-1} b_{k-1}, \quad 0 < k \leq N \tag{2.10}$$

Hence

$$p_k = \frac{b_{k-1}}{d_k} p_{k-1} = p_0 \prod_{j=1}^{k} \frac{b_{j-1}}{d_j}, \quad 0 < k \leq N \tag{2.11}$$

Since $\sum_{k \in S} p_k = 1$

$$\sum_{k \in S} p_k = p_0 + p_0 \sum_{k=1}^{N} \prod_{j=1}^{k} \frac{b_{j-1}}{d_j} = 1$$

from which

$$p_0 = \frac{1}{1 + \sum_{k=1}^{N} \prod_{j=1}^{k} \frac{b_{j-1}}{d_j}} \qquad (2.12)$$

- example: b, d are constant /they do not depend on the state/, the Foster criteria can be applied for the $N = \infty$ case.

3. state sojourn time distribution

   - consider a state i for which $p_{ii} < 1$:
   - the probability, that the process leaves state i at time n is:

$$\mathbf{P}(\tau_i = n) = p_{ii}^{n-1}(1 - p_{ii}) \qquad (2.13)$$

   (Note that the Markov property is utilized.)
   i.e. the sojourn time is geometrically distributed, from which

$$E(\tau_i) = \frac{1}{1 - p_{ii}} = \frac{1}{\sum_{j \neq i} p_{ij}} \qquad (2.14)$$

   Note the memoryless property of the geometric distribution!

4. The distribution of the consecutive state after state i

   - the probability that the system leaves state i for state j at time n is

$$\mathbf{P}(X_n = j \mid X_n \neq i,\ X_k = i,\ k < n) =$$

$$\frac{\mathbf{P}(X_n = j, X_n \neq i,\ X_k = i,\ k < n)}{\mathbf{P}(X_n \neq i,\ X_k = i,\ k < n)} = \frac{\mathbf{P}(X_n = j,\ X_k = i,\ k < n)}{\mathbf{P}(X_n \neq i,\ X_k = i,\ k < n)} =$$

$$\frac{p_{ii}^{n-1} p_{ij}}{p_{ii}^{n-1}(1 - p_{ii})} = \frac{p_{ij}}{\sum_{k \neq i} p_{ik}} \qquad (2.15)$$

   Recognize: the distribution of the consecutive state is independent of the time when state i is left. The consecutive state depends only on the tagged state. —- Markov property

5. z-transform domain analysis of DTMCs
   Define $P(z)$ as follows:

$$P(z) = \sum_{n=0}^{\infty} P^{(n)} z^n$$

Based on $P^{(n)} = P^{(n-1)}\boldsymbol{\Pi}$ we have:

$$\sum_{n=1}^{\infty} P^{(n)} z^n = \sum_{n=1}^{\infty} P^{(n-1)}\boldsymbol{\Pi} z^n$$

$$P(z) - P^{(0)} = z\sum_{n=1}^{\infty} P^{(n-1)}\boldsymbol{\Pi} z^{n-1} = zP(z)\boldsymbol{\Pi}$$

$$P(z) = P^{(0)}\left[\boldsymbol{I} - z\boldsymbol{\Pi}\right]^{-1} \tag{2.16}$$

$$\lim_{n\to\infty} P^{(n)} = \lim_{z\to 1} P^{(0)}(1-z)\left[\boldsymbol{I} - z\boldsymbol{\Pi}\right]^{-1} = P^{(0)}\lim_{z\to 1}(1-z)\left[\boldsymbol{I} - z\boldsymbol{\Pi}\right]^{-1}$$

(*What is the condition of stability based on this expression?*)

6. z-transform domain analysis of a two-state example

$$\boldsymbol{\Pi} = [p_{ij}] = \begin{bmatrix} .5 & .5 \\ .75 & .25 \end{bmatrix}$$

- steady state distribution:

$$p_0 = 0.5p_0 + 0.75p_1 \quad p_1 = 0.5p_0 + 0.25p_1 \quad p_0 + p_1 = 1$$

$$2p_0 = 3p_1 \quad \longrightarrow \quad p_0 = 0.6, \ \ p_1 = 0.4$$

- transient behaviour

$$\boldsymbol{I} - \boldsymbol{\Pi} z = \begin{bmatrix} 1-0.5z & -0.5z \\ -0.75z & 1-0.25z \end{bmatrix}$$

$$det(\boldsymbol{I} - \boldsymbol{\Pi} z) = (1-z)(1+0.25z)$$

$$(\boldsymbol{I} - \boldsymbol{\Pi} z)^{-1} = \frac{1}{(1-z)(1+0.25z)}\begin{bmatrix} 1-0.25z & 0.5z \\ 0.75z & 1-0.5z \end{bmatrix}$$

Partial fraction decomposition of $(\boldsymbol{I} - \boldsymbol{\Pi} z)^{-1}$:

$$\frac{\boldsymbol{A}}{1-z} + \frac{\boldsymbol{B}}{1+0.25z}$$

$$(\boldsymbol{I}-\boldsymbol{\Pi}z)^{-1}=\frac{1}{1-z}\begin{bmatrix}0.6 & 0.4\\ 0.6 & 0.4\end{bmatrix}+\frac{1}{1+0.25z}\begin{bmatrix}0.4 & -0.4\\ -0.6 & 0.6\end{bmatrix}$$

$$(\boldsymbol{I}-\boldsymbol{\Pi}z)^{-1}=\sum_{n=0}^{\infty}z^n\begin{bmatrix}0.6 & 0.4\\ 0.6 & 0.4\end{bmatrix}+\sum_{n=0}^{\infty}(\frac{-z}{4})^n\begin{bmatrix}0.4 & -0.4\\ -0.6 & 0.6\end{bmatrix}$$

From which:

$$P(z)=P^{(0)}\sum_{n=0}^{\infty}\boldsymbol{\Pi}^{(n)}z^n=P^{(0)}\boldsymbol{\Pi}(z)$$

$$\boldsymbol{\Pi}^{(n)}=\boldsymbol{\Pi}^n=\begin{bmatrix}0.6 & 0.4\\ 0.6 & 0.4\end{bmatrix}+(\frac{-1}{4})^n\begin{bmatrix}0.4 & -0.4\\ -0.6 & 0.6\end{bmatrix}$$

$$P_0^{(n)}=0.6+\left(\frac{-1}{4}\right)^n\left(0.4P_0^{(0)}-0.6P_1^{(0)}\right)$$

$$P_1^{(n)}=0.4+\left(\frac{-1}{4}\right)^n\left(-0.4P_0^{(0)}+0.6P_1^{(0)}\right)$$

2.4 Summary of continuous time Markov chains and their properties

1. $X(t)$ is a continuous time Markov chain (CTMC), if it enjoys the Markov the markov propoerty every time, i.e. if the following equation holds for $\forall n \geq 1$, $0 < t_0 < t_1 < \cdots < t_n$ and $x_0, x_1, \cdots, x_n \in S$

$$\begin{aligned}\mathbf{P}(X(t_n)=x_n|X(t_{n-1})=x_{n-1},\ldots,X(t_0)=x_0)=\\ \mathbf{P}(X(t_n)=x_n|X(t_{n-1})=x_{n-1}).\end{aligned} \tag{2.17}$$

2. Consequence of Markov property:

$$P(t+u)=P(t)\boldsymbol{\Pi}(\boldsymbol{t},\boldsymbol{u}) \tag{2.18}$$

where:

$$P(t)=[p_0(t),p_1(t),p_2(t),\cdots],\quad p_i(t)=\mathbf{P}(X(t)=i),$$

$$\boldsymbol{\Pi}(t,u)=[p_{ij}(t,u)],\quad p_{ij}(t,u)=\mathbf{P}(X(t+u)=j\mid X(t)=i),$$

$\forall\ i,\ j \in S, \forall t, u \geq 0$

Chapman-Kolmogorov equality:

$$p_{ij}(t,u+v)=\sum_{l\in S}p_{il}(t,u)p_{lj}(t+u,v);\quad \boldsymbol{\Pi}(t,u+v)=\boldsymbol{\Pi}(t,u)\boldsymbol{\Pi}(t+u,v) \tag{2.19}$$

3. homogeneous CTMC:

   $\boldsymbol{\Pi}(t, u+v) = \boldsymbol{\Pi}(u+v) = \boldsymbol{\Pi}(u)\boldsymbol{\Pi}(v); \quad p_{ij}(u) = \mathbf{P}(X(t+u) = j \mid X(t) = i),$

   $\boldsymbol{\Pi}(u) = [p_{ij}(u)], \quad \forall\ i,\ j \in S, \forall t, u \geq 0$

4. for state j we have:

$$P_j(t+u) = \sum_{i=0}^{n} P_i(t)p_{ij}(u) = P_j(t)p_{jj}(u) + \sum_{i \neq j} P_i(t)p_{ij}(u)$$

   By the Taylor serie of $p_{ij}(u)$:

$$P_j(t+u) - P_j(t) = \sum_{i \neq j} P_i(t)p_{ij}(u) - P_j(t)\sum_{k \neq j} p_{jk}(u) =$$

$$\sum_{i \neq j} P_i(t)\, q_{ij}\, u - P_j(t)\sum_{k \neq j} q_{jk}\, u + o(u),$$

   where $\boldsymbol{Q}$ rate matrix (infinitesimal generator) is defined as follows:
   $q_{ij} = \lim\limits_{u \to 0} \dfrac{p_{ij}(u) - \delta_{ij}}{u}, \quad \boldsymbol{Q} = [q_{ij}], \quad \sum\limits_{j \in S} q_{ij} = 0, \forall i \in S.$

   From which:

$$\frac{d}{dt}P_j(t) = \lim_{u \to 0}\frac{P_j(t+u) - P_j(t)}{u} = \sum_{i \neq j} P_i(t)q_{ij} - P_j(t)\sum_{k \neq j} q_{jk} \tag{2.20}$$

$$\frac{d}{dt}P(t) = P(t)\boldsymbol{Q} \tag{2.21}$$

   We suppose that for $\forall i\ q_{ii} > -\infty$ *— phisical meaning.*

5. properties of CTMCs

   - irreducibility: "each states are reachable from each others"
     for $\forall i, j\ \ \exists t < \infty$, that $p_{ij}(t) > 0$
   - recurrence: positive (*finit expected recurrence time*)
     see the discerete case
   - heredity: the recurrency is heritaged in irreducibile DTMCs.

6. existence of unique steady state — stability

- finite CTMC: irreducibility
- infinite CTMC: irreducibility and positive recurrency
- sufficient condition of stability of infinite CTMCs — Foster kriteria

Below we suppose that the sojourn time is positive for each states with probability 1, and the state space $S = \{1, 2, \ldots, N\}$ is finite, without mentioning anything else.

7. consequence of stability
   - $\lim_{t\to\infty} p_{ij}(t) = p_j > 0, \ \forall i, j \in S$, where $p_j = \lim_{n\to\infty} \mathbf{P}(X(t) = j)$
   - in this case there is only one solution of the equation $P\boldsymbol{Q} = 0$ ($P = [p_0, p_1, p_2, \cdots]$) with the additional condition $\sum_{i\in S} p_i = 1$ which is the steady state distribution of the CTMC.
     If $P(0) = P$ then $P(t) = P, \ \ \forall t$.

8. steady state distribution
   - since $\lim_{t\to\infty} P_j(t+u) = \lim_{t\to\infty} P_j(t) = p_j, \ \ \forall j \in S$

     $$\lim_{t\to\infty} \frac{d}{dt} P_j(t) = 0 = \sum_{i\neq j} p_i q_{ij} - p_j \sum_{k\neq j} q_{jk}$$

     which means that

     $$\sum_{i\neq j} p_i q_{ij} = p_j \sum_{k\neq j} q_{jk} \tag{2.22}$$

     (*Interpretation of this result.*)
   - The same holds for a group of states as well

     $$\sum_{i\in U} p_i \sum_{j\in D} q_{ij} = \sum_{j\in D} p_j \sum_{k\in U} q_{jk} \tag{2.23}$$

9. birth-death process
   - interpretation:

     $$q_{ij} = \begin{cases} \lambda_i & j = i+1, \ \ 0 \le i < N \\ \mu_i & j = i-1, \ \ 0 < i \le N \\ -\lambda_i - \mu_i & j = i, \ \ 0 < i < N \\ -\lambda_0 & j = i, \ \ i = 0 \\ -\mu_N & j = i, \ \ i = N \\ 0 & \text{otherwise} \end{cases}$$

- steady state distribution of birth-death process

$$p_{k-1}q_{k-1,k} + p_{k+1}q_{k+1,k} = p_k(q_{k,k-1} + q_{k,k+1}),\ 0 < i \leq N$$
$$p_{k-1}\lambda_{k-1} + p_{k+1}\mu_{k+1} = p_k(\lambda_k + \mu_k),\quad 0 < i \leq N$$
$$p_1\mu_1 = p_0\lambda_0$$

from which

$$p_k\mu_k = p_{k-1}\lambda_{k-1},\quad 0 < i \leq N \tag{2.24}$$

Hence

$$p_k = \frac{\lambda_{k-1}}{\mu_k}p_{k-1} = p_0\prod_{j=1}^{k}\frac{\lambda_{j-1}}{\mu_j},\quad 0 < i \leq N \tag{2.25}$$

Since $\sum\limits_{k\in S} p_k = 1$ we have

$$\sum_{k\in S} p_k = p_0 + p_0\sum_{k=1}^{N}\prod_{j=1}^{k}\frac{\lambda_{j-1}}{\mu_j} = 1$$

from which

$$p_0 = \frac{1}{1 + \sum\limits_{k=1}^{N}\prod\limits_{j=1}^{k}\frac{\lambda_{j-1}}{\mu_j}} \tag{2.26}$$

- state independent intensities (*M/M/1 queue*)

10. state sojourn time distribution

- for a given state i for which $q_{ii} < 0$:
- due to the Markov property the sojourn time in state i is characterized by the following differential equation:

$$\frac{d}{dt}P_i(t) = q_{ii}P_i(t),\quad P_i(0) = 1$$

which results in:

$$P_i(t) = \mathbf{P}(\tau_i > t) = \mathrm{e}^{q_{ii}t} \tag{2.27}$$

and

$$E(\tau_i) = \frac{1}{-q_{ii}} = \frac{1}{\sum_{j\neq i} q_{ij}} \tag{2.28}$$

since

$$F_{\tau_i}(t) = \mathbf{P}(\tau_i \leq t) = 1 - \mathrm{e}^{q_{ii}t} = 1 - \mathrm{e}^{-\sum_{j\neq i} q_{ij}t}$$

See the memoryless property of the exponential distribution!

11. distribution of the consecutive state after state i

    - the probability that the system enters into state j $(j \neq i)$ suppose that it leaves state i during the interval $(t, t + \Delta t)$ is

$$\mathbf{P}(X(t+\Delta t) = j \mid X(t+\Delta t) \neq i,\ X(u) = i,\ \forall 0 \leq u \leq t) =$$
$$\frac{\mathbf{P}(X(t+\Delta t) = j,\ X(t+\Delta t) \neq i,\ X(u) = i,\ \forall 0 \leq u \leq t)}{\mathbf{P}(X(t+\Delta t) \neq i,\ X(u) = i,\ \forall 0 \leq u \leq t)} =$$
$$\frac{\mathbf{P}(X(t+\Delta t) = j,\ X(u) = i,\ \forall 0 \leq u \leq t)}{\mathbf{P}(X(t+\Delta t) \neq i,\ X(u) = i,\ \forall 0 \leq u \leq t)} =$$
$$\frac{q_{ij}\, \Delta t\, \mathrm{e}^{q_{ii} t} + o(\Delta t)}{\sum_{k \neq i} q_{ik}\, \Delta t\, \mathrm{e}^{q_{ii} t} + o(\Delta t)}$$

    from which:

$$\lim_{\Delta t \to 0} \mathbf{P}(X(t+\Delta t) = j \mid X(t+\Delta t) \neq i,\ X(t) = i,\ \forall 0 \leq t < t + \Delta t) = \frac{q_{ij}}{\sum_{k \neq i} q_{ik}} \tag{2.29}$$

    The distribution of thhe consecutive state depends only on the persent state and independent of the time of the state transition — Markov property!

12. Discrete time process embedded into the state transition instances (embedded DTMC)

    - consider the states of a CTMC at the state transitions instances (X_n^a) *It is a DTMC. Why?*
    - Let $\boldsymbol{\Pi}^a$ is the one-step state transition matrix of the embedded DTMC
    - the elements of $\boldsymbol{\Pi}^a$ can be evaluated by 2.29 based on the infinitesimal generator of the CTMC:

$$p_{ij}^a = \frac{q_{ij}}{\sum_{k \neq i} q_{ik}}, \quad p_{ii}^a = 0, \quad \forall i$$

    - the transient $P_i^a(n)$ and the steady state p_i^a behaviour of the embedded DTMC can be evaluated based on $\boldsymbol{\Pi}^a$
    - the relation of the steady state of a CTMC (p_i) and the embedded DTMC (p_i^a) is characterized by the following equation:

$$p_i = \frac{p_i^a\ \mathbf{E}(\tau_i)}{\sum_{j \in S}\ p_j^a\ \mathbf{E}(\tau_j)} \tag{2.30}$$

where $\mathbf{E}(\tau_i) = \dfrac{1}{\sum_{j\neq i} q_{ij}} = \dfrac{1}{-q_{ii}}$ is the mean sojourn time in state i
The above consideration allows us to apply the Foster criteria to check the positive recurrence property of a CTMC

- example: random walk over a finite state space
  See the state transition graph

$$p^a_{ij} = \begin{cases} \lambda_i/(\lambda_i+\mu_i) & j=i+1,\ 0<i<N \\ \mu_i/(\lambda_i+\mu_i) & j=i-1,\ 0<i<N \\ 0 & j=i,\ 0\leq i\leq N \\ 1 & i=0,\ j=1 \\ 1 & i=N,\ j=N-1 \\ 0 & \text{otherwise} \end{cases}$$

hence $p^a_1 = \dfrac{\lambda_1+\mu_1}{\mu_1} p^a_0$ and

$$p^a_k = \frac{\lambda_{k-1}(\lambda_k+\mu_k)}{\mu_k(\lambda_{k-1}+\mu_{k-1})} p_{k-1} = \frac{\lambda_k+\mu_k}{\mu_1} p_0 \prod_{j=2}^{k} \frac{\lambda_{j-1}}{\mu_j}, \quad 1<k\leq N$$

from which

$$p^a_0 = \frac{1}{1+(\lambda_1+\mu_1)/\mu_1 + \sum_{k=2}^{N}((\lambda_k+\mu_k)/\mu_1)\prod_{j=2}^{k}(\lambda_{j-1}/\mu_j)}$$

and from 2.30 we have

$$p_k = \frac{(1/\mu_1)\prod_{j=2}^{k}(\lambda_{j-1}/\mu_j)}{1/\lambda_0 + 1/\mu_1 + (1/\mu_1)\sum_{k=2}^{N}\prod_{j=2}^{k}(\lambda_{j-1}/\mu_j)} = \frac{\prod_{j=1}^{k}(\lambda_{j-1}/\mu_j)}{1+\sum_{k=1}^{N}\prod_{j=1}^{k}(\lambda_{j-1}/\mu_j)}$$

and

$$p_0 = \frac{1/\lambda_0}{1/\lambda_0 + 1/\mu_1 + (1/\mu_1)\sum_{k=2}^{N}\prod_{j=2}^{k}(\lambda_{j-1}/\mu_j)} = \frac{1}{1+\sum_{k=1}^{N}\prod_{j=1}^{k}(\lambda_{j-1}/\mu_j)}$$

which is consistent with the former results.

13. Laplace transform domain analysis of CTMCs

$$\frac{d}{dt}P(t) = P(t)\boldsymbol{Q}$$

$$
\begin{aligned}
sP^*(s) - P(0) &= P^*(s)\boldsymbol{Q} \\
P^*(s) &= P(0)\,[s\boldsymbol{I} - \boldsymbol{Q}]^{-1} \qquad (2.31)\\
\lim_{t\to\infty} P(t) = \lim_{s\to 0} s\, P^*(s) &= P(0) \lim_{s\to 0} s\ \ [s\boldsymbol{I} - \boldsymbol{Q}]^{-1}
\end{aligned}
$$

(*What is the condition of stability based on the LT domain description?*)

14. Consider the following two-state example

$$
\boldsymbol{Q} = [q_{ij}] = \begin{bmatrix} -0.5 & 0.5 \\ 0.75 & -0.75 \end{bmatrix}
$$

- evaluation of the steady state distributrion

$$
0 = -0.5p_0 + 0.75p_1 \quad 0 = 0.5p_0 + -0.75p_1 \quad p_0 + p_1 = 1
$$

$$
2p_0 = 3p_1 \quad \longrightarrow \quad p_0 = 0.6, \ \ p_1 = 0.4
$$

- transient behaviour

$$
\boldsymbol{Y} = s\boldsymbol{I} - \boldsymbol{Q} = \begin{bmatrix} s+0.5 & -0.5 \\ -0.75 & s+0.75 \end{bmatrix}
$$

$$
det(\boldsymbol{Y}) = det(s\boldsymbol{I} - \boldsymbol{Q}) = s(s+1.25)
$$

introducing $(\boldsymbol{Y}_i(s)) = s\boldsymbol{I} - \boldsymbol{Q}_i$, where in $\boldsymbol{Q}_i$ the ith row of $\boldsymbol{Q}$ is substituted by $P(0)$ and applying the Cramer's role:

$$
P_i^*(s) = \frac{det(\boldsymbol{Y}_i(s))}{det(\boldsymbol{Y}(s))}
$$

Let $P(0) = [P_0(0) \ \ P_1(0)]$:

$$
det(\boldsymbol{Y}_0(s)) = \begin{bmatrix} P_0(0) & P_1(0) \\ -0.75 & s+0.75 \end{bmatrix} = P_0(0)(s+0.75) + P_1(0)0.75
$$

$$
det(\boldsymbol{Y}_1(s)) = \begin{bmatrix} s+0.5 & -0.5 \\ P_0(0) & P_1(0) \end{bmatrix} = P_0(0)0.5 + P_1(0)(s+0.5)
$$

$$
P_0(s) = P_0(0)\frac{s+0.75}{s(s+1.25)} + P_1(0)\frac{0.75}{s(s+1.25)}
$$

$$
P_1(s) = P_0(0)\frac{0.5}{s(s+1.25)} + P_1(0)\frac{s+0.5}{s(s+1.25)}
$$

$$P_0(t) = P_0(0)(\frac{0.75}{1.25} + \frac{0.5}{1.25}\ \mathrm{e}^{-1.25t}) + P_1(0)(\frac{0.75}{1.25} - \frac{0.75}{1.25}\ \mathrm{e}^{-1.25t})$$
$$P_0(t) = 0.6 + 0.4P_0(0)\mathrm{e}^{-1.25t} - 0.6P_1(0)\mathrm{e}^{-1.25t})$$
$$P_1(t) = P_0(0)(\frac{0.5}{1.25} - \frac{0.5}{1.25}\ \mathrm{e}^{-1.25t}) + P_1(0)(\frac{0.5}{1.25} + \frac{0.75}{1.25}\ \mathrm{e}^{-1.25t})$$
$$P_1(t) = 0.4 - 0.4P_0(0)\mathrm{e}^{-1.25t} + 0.6P_1(0)\mathrm{e}^{-1.25t})$$

If $P_0(0) = 1, \quad P_1(0) = 0$ then:

$$P_0(t) = 0.6 + 0.4\ \mathrm{e}^{-1.25t}, \qquad P_1(t) = 0.4 - 0.4\ \mathrm{e}^{-1.25t}$$

15. constant intensity birth process — *Poisson process*

- interpretation:

$$q_{ij} = \begin{cases} \lambda & j = i+1, \ \ i \geq 0 \\ -\lambda & j = i, \ \ i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  The exit rate of all the states are λ, hence the sojourn time in each states are exponentially distributed with parameter λ, i.e. the time between consecutive state transitions /arrivals/ are exponentially distributed with parameter λ.

- this process is not irreducible $\rightarrow$ no steady state
- the transient state probability distribution is characterized by:

$$\frac{d}{dt}P_j(t) = P_{j-1}(t)\lambda - P_j(t)\lambda, \quad j > 0, \quad \frac{d}{dt}P_0(t) = -P_0(t)\lambda$$

  Introducing the z-transform of the state probability distribution $P(z,t) = \sum_{j=0}^{\infty} P_j(t)z^j$
  we have:

$$\sum_{j=1}^{\infty} \frac{d}{dt}P_j(t)z^j = \sum_{j=1}^{\infty} P_{j-1}(t)\lambda z^j - \sum_{j=1}^{\infty} P_j(t)\lambda z^j$$
$$\frac{\partial}{\partial t}P(z,t) - \frac{d}{dt}P_0(t) = \lambda z P(z,t) - \lambda P(z,t) + \lambda P_0(t)$$
$$\frac{\partial}{\partial t}P(z,t) = -\lambda(1-z)P(z,t)$$

And now taking its Laplace transform by considering the initial condition $P_0(0) = 1$ (P(z,0) = 1) we get:

$$sP(z,s) - P(z,0) = -\lambda(1-z)P(z,s)$$

$$P(z,s) = \frac{1}{s+\lambda-\lambda z} = \frac{\frac{1}{s+\lambda}}{1-\frac{\lambda z}{s+\lambda}} = \frac{1}{s+\lambda}\sum_{j=0}^{\infty}(\frac{\lambda}{s+\lambda})^j z^j$$

from which $P_j(s) = \dfrac{\lambda^j}{(s+\lambda)^{j+1}}$, and finally

$$P_j(t) = \frac{(\lambda t)^j}{j!}e^{-\lambda t}, \tag{2.32}$$

which is the Poisson distribution with parameter λt.

- The expected number of state transitions (arrivals) up to time t, i.e. the mean of $X(t)$ is:

$$\mathbf{E}(X(t)) = \frac{\partial}{\partial z}P(z,t)\Big|_{z=1} = LT^{-1}\frac{\partial}{\partial z}P(z,s)\Big|_{z=1} =$$
$$LT^{-1}\frac{\lambda}{(s+\lambda(1-z))^2}\Big|_{z=1} = LT^{-1}\frac{\lambda}{s^2}$$

from which

$$\mathbf{E}(X(t)) = \lambda t \tag{2.33}$$

i.e. the mean number of arrivals is a linear function of the time, which means that the arrival rate is constant.

Consequence: the following descriptions defines the same stochastic process:

- *constant arrival intensity,*
- *independent identical exponentially distributed interarrival times,*
- *Poisson arrival process.*

- the variation of the number of arrivals up to time t:

$$\frac{\partial^2}{\partial z^2}P(z,t)\Big|_{z=1} = LT^{-1}\frac{\partial}{\partial z}P'(z,s)\Big|_{z=1} = LT^{-1}\frac{2(s+\lambda(1-z))\lambda^2}{(s+\lambda(1-z))^4}\Big|_{z=1} =$$
$$LT^{-1}\frac{2\lambda^2}{s^3}$$

hence

$$\mathbf{E}(X^2(t)) = LT^{-1}\frac{\partial^2}{\partial z^2}P(z,s)\bigg|_{z=1} + \mathbf{E}(X(t)) = (\lambda t)^2 + \lambda t \qquad (2.34)$$

and by 2.34 and 2.33 we have:

$$\sigma^2 = \mathbf{E}(X^2(t)) - (\mathbf{E}(X(t)))^2 = (\lambda t)^2 + \lambda t - (\lambda t)^2 = \lambda t \qquad (2.35)$$

- The superposition of independent Poisson processes is a Poisson process as well. Let $X_1(t), \cdots, X_n(t)$ be independent Poisson processes with parameter $\lambda_1, \cdots, \lambda_n$. Their superposition $X(t) = X_1(t) + \cdots, X_n(t)$ is a Poisson process with parameter $\lambda = \lambda_1 + \cdots + \lambda_n$.
  Proof: first we consider two processes
  Let $X(t,\tau) = X(\tau) - X(t)$ the number of arrivals during the interval $[t, \tau)$. In this case

$$\mathbf{P}(X(u,u+t) = n) = \sum_{i=0}^{n}\mathbf{P}(X_1(u,u+t) = i \cap X_2(u,u+t) = n-i) =$$

$$\sum_{i=0}^{n}\mathbf{P}(X_1(u,u+t) = i)\mathbf{P}(X_2(u,u+t) = n-i) =$$

$$\sum_{i=0}^{n}\frac{(\lambda_1 t)^i}{i!}e^{-\lambda_1 t}\frac{(\lambda_2 t)^{(n-i)}}{(n-i)!}e^{-\lambda_2 t} =$$

$$\frac{e^{-\lambda_1 t}e^{-\lambda_2 t}}{n!}\sum_{i=0}^{n}\binom{n}{i}(\lambda_1 t)^i(\lambda_2 t)^{(n-i)} = \frac{e^{-(\lambda_1+\lambda_2)t}}{n!}((\lambda_1+\lambda_2)t)^n$$

  from which the $n > 2$ cases can be prooved by induction.

  Alternative proof in z-transform domain

$$P_k(z,t) = \mathrm{e}^{(z-1)\lambda_k t}, \quad \forall k = 1, \cdots, n$$

$$P(z,t) = \prod_{k=1}^{n}\mathrm{e}^{(z-1)\lambda_k t} = \mathrm{e}^{(z-1)\sum_{k=1}^{n}\lambda_k t} = \mathrm{e}^{(z-1)\lambda t}$$

- random decomposition of a Poisson process is a Poisson process as well
  Let $q_1, \cdots, q_m$ ($\sum_i q_i = 1$) be the switching probabilities of the $1, \cdots, n$th kind of arrivals. In this case:

$$\mathbf{P}(X(\tau,\tau+t) = n_1 + \cdots + n_m = n) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}\frac{n!}{n_1!\cdots n_m!}q_1^{n_1}\cdots q_m^{n_m} =$$

$$\frac{(q_1\lambda t)^{n_1}}{n_1!}e^{-q_1\lambda t}\cdots\frac{(q_m\lambda t)^{n_m}}{n_m!}e^{-q_m\lambda t}=\prod_{k=1}^{n}\frac{(q_k\lambda t)^{n_k}}{n_k!}e^{-q_k\lambda t}$$

- an alternative definition of Poisson processes is:
  $X(t)$ is a Poisson process if:
  - $X(t)$ is Poisson distributed with parameter λt for $\forall t$, i.e.
    $X(0)=0,\ \ \mathbf{P}(X(t)=k)=\frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t},\ \ (t>0, k\in S)$
  - $X(t)$ has independent increments, i.e. $\forall n\geq 2$ and $0<t_1<\cdots<t_n$ the random varianbles $X(t_1)-X(t_0), X(t_2)-X(t_1),\cdots,X(t_n)-X(t_{n-1})$ are independent,
  - $X(t)$ has stacionar increment, i.e. $\forall u,t>0$ the distribution of $X(t)-X(0)$ and $X(t+u)-X(u)$ are identical.
- a consequence of the definition of Poisson processes:
  - $\mathbf{P}(X(t)=0)=1-\lambda t+o(t)$
  - $\mathbf{P}(X(t)=1)=\lambda t+o(t)$ *density condition*
  - $\mathbf{P}(X(t)\geq 2)=o(t)$ *rerity condition*

2.5 Other stochastic processes (Renewal process, Semi-Markov process)

1. Renewal process
   Let $T_1, T_2,\cdots$ be the time of occurence of random events and let $Z_i=T_i-T_{i-1},\ T_0=0,\ i>0$ the time delay between the consecutive occurences. If $Z_i,\ i>0$ are i.i.d. r.v.s then the serie of random times $T_i,\ i\geq 0$ is called renewal process.
   Properties:
   - the Markov property holds at time T_i for $\forall i$,
   - generally distributed sojourn times,
2. Semi-Markov process:
   a continuous time discrete state stochastic process ($X(t)\in S$) for which:
   - the Markov property holds at every state transition instance,
   - the states of the process at state transition instances form a DTMC,
   - generally distributed sojourn times (between state transitions),

- the sojourn times are state dependent,
- a useful result:

$$p_i = \frac{p_i^a \ \mathbf{E}(\tau_i)}{\sum_{j \in S} \ p_j^a \ \mathbf{E}(\tau_j)} \tag{2.36}$$

  where

  - $\mathbf{E}(\tau_i)$ is the mean sojourn time in state i
  - p_j^a is the steady state distribution of the embedded DTMC

Chapter 3

INTRODUCTION TO QUEUEING THEORY

3.1 Performance characterization of queueing systems

3.1.1 Queueing system as a black box

1. Queueing system:

    - server(s)
    - queue or buffer

2. Environment of queueing systems:

    - customers (requests, demands, jobs, etc.)
    - sources – submits customers (terminals, telephone sets, etc.)

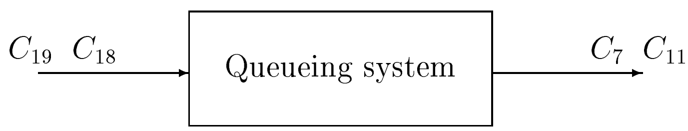

Figure 3.1: Queueing system as a black box

3. Queueing system as a black box:
   C_n: the nth customer arrives to the queueing system for service (see fig. 3.1)

4. Formal description of a single server queue: (see fig. 3.2)

3.1.2 Arrival process of customers

1. τ_n: the arrival time of the nth customer (r.v.)
2. $t_n = \tau_n - \tau_{n-1}$: the nth interarrival time (r.v.)
3. $A_n(t) = P(t_n < t)$: the distribution of the nth interarrival time, and its average $\bar{a_n}$
4. $\alpha(t)$: number of arrivals in $(0, t)$ – r.v.
5. mean arrival rate in $(0, t)$: $\lambda_t = \dfrac{\alpha(t)}{t}$ – r.v.
6. instantaneous arrival rate:

$$\lambda(t) = \mathbf{E}\left(\lim_{\Delta t \to 0} \frac{\alpha(t + \Delta t) - \alpha(t)}{\Delta t}\right)$$

7. special case: $\lambda(t) = \lambda$, $\forall\ t$, in this case $\lim\limits_{t\to\infty} \lambda_t = \lambda$ as well.

3.1.3 The departure process of (served) customers

1. x_n: the service time of the nth customer (r.v.)
2. $B_n(x) = P(x_n < x)$: its distribution, $\bar{b_n}$, $\bar{x_n}$
3. w_n: the waiting time of the nth customer (r.v.), i.e. the time while the customer waits for a free server in the queue, $\bar{w_n}$
4. $s_n = w_n + x_n$: the system time of the nth customer (r.v.), i.e. the time the nth customer spent in the system, $\bar{s_n}$
5. $\delta(t)$: number of departure in $(0, t)$ – r.v.
6. mean departure rate: $S_t = \dfrac{\delta(t)}{t}$ – r.v.

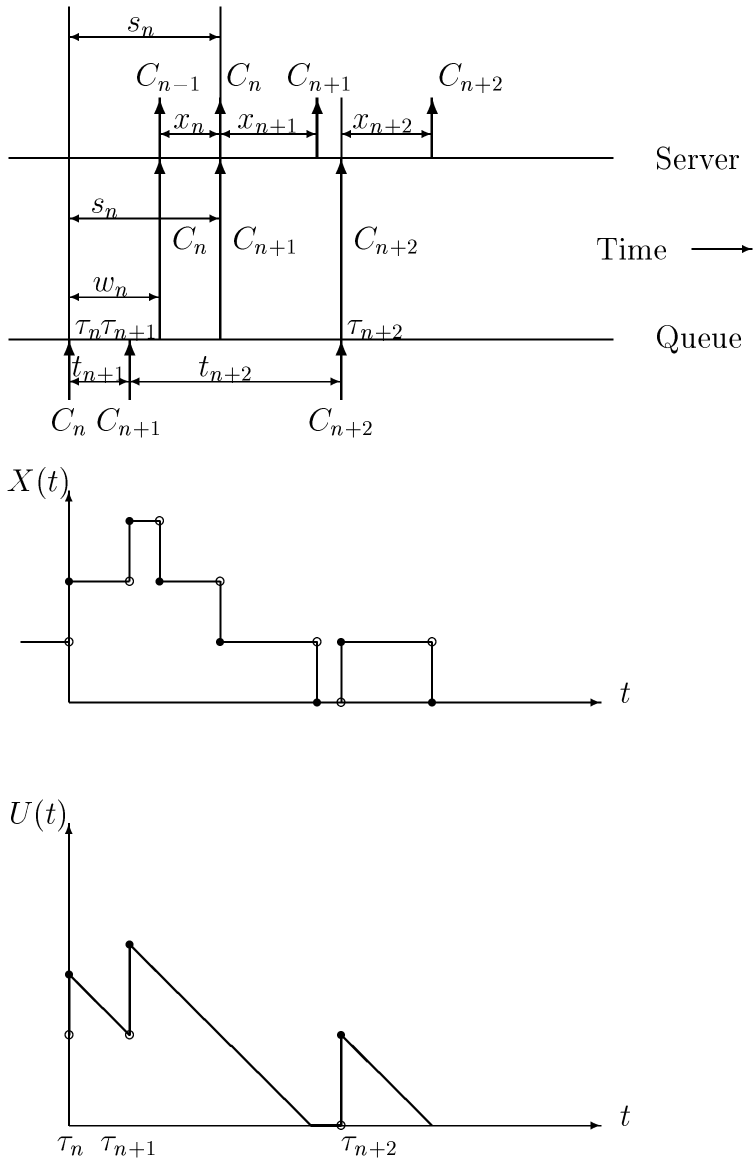

Figure 3.2: A realization of the time diagram of a single server queueing system

7. instantaneous departure rate:

$$S(t) = \mathbf{E}\left(\lim_{\Delta t \to 0} \frac{\delta(t+\Delta t) - \delta(t)}{\Delta t} \right)$$

8. special case: $S(t) = S$, $\forall\ t$, hence $\lim_{t\to\infty} S_t = S$

3.1.4 Parameters of queueing systems

1. $X(t)$: number of customers in the system at time t (r.v.)

2. $X_w(t)$: number of waiting customers in the system at time t (r.v.), i.e. the number of customers in the queue.

3. $X_s(t)$: number of customers under service at time t (r.v.) i.e. the number of customers in the server(s).

4. $U(t)$: remaining service time at time t (r.v.)

3.2 Performance parameters of queueing systems

3.2.1 Classification of performance parameters

1. Customer's point of view

   - throughput:
     - $\delta(t)$: number of departure in $(0, t)$,
     - S_t: mean departure rate,
     - $S(t)$: instantaneous departure rate,
   - s_n: service time (distribution, moments, mean)
   - probability of loss:
     $\mathbf{P}$(a customer is refused by the system | a customer arrived)

2. Service provider's point of view

   - number of customers in the system (distribution, moments): $X(t), \cdots$
   - utilization of server(s) (see ergodicity):
     $\rho = \mathbf{P}($ the server(s) is busy$)$

3.2.2 Evaluation of performance parameters

1. utilization:

   - of a single server queue

$$\rho = \mathbf{P}(X > 0) = 1 - p_0 = \bar{X}_s = \bar{\lambda}\ \bar{x} \tag{3.1}$$

   - of a multi (c) server queue

$$\rho = \frac{\bar{X}_s}{c} = \frac{(\sum_{k=0}^{c-1} k\ p_k + \sum_{k=c}^{\infty} c\ p_k)}{c} = \frac{\bar{\lambda}\ \bar{x}}{c} \tag{3.2}$$

2. probability of loss:
   in a c servers 0 buffer system

$$p_{loss} = \mathbf{P}(\text{a customer is refused} \mid \text{a customer arrives to the system})$$

$$= \frac{\mathbf{E}(\text{number of refused customers})}{\mathbf{E}(\text{number of customers arrived to the system})}$$

$$= \frac{\sum_{k=0}^{c} \mathbf{E}(\text{number of refused customers} \mid \text{there are } k \text{ customers in the system})}{\sum_{k=0}^{c} \mathbf{E}(\text{number of arrived customers} \mid \text{there are } k \text{ customers in the system})} \tag{3.3}$$

   if $\lambda(t) = \lambda$ and $p_i = \mathbf{P}(X = i)$:

$$p_{loss} = \frac{\lambda\ p_c}{\lambda \sum_{k=0}^{c} p_k} = \frac{\lambda\ p_c}{\lambda} = p_c \tag{3.4}$$

3.2.3 Little's rule

1. for a lossless, work-conservative queueing system, if exists
   $\bar{\lambda} = \lim_{t\to\infty} \lambda_t$ and $\bar{T} = \lim_{t\to\infty} T_t$, then

$$\bar{X} = \bar{\lambda}\bar{T} \tag{3.5}$$

2. proof: (Kleinrock)

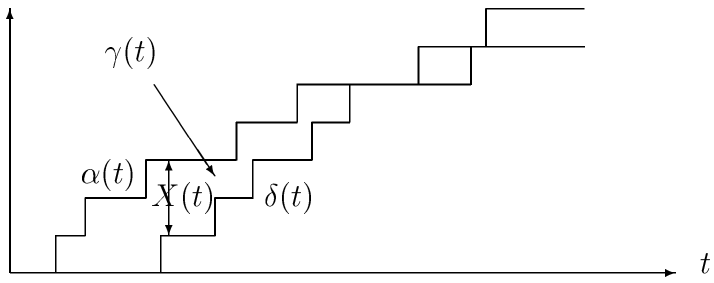

Figure 3.3: Number of arrivals and departures

- $\alpha(t)$: number of arrivals in $(0, t)$ (see above)
- $\delta(t)$: number of departures in $(0, t)$
- if $X(0) = 0$, then $X(t) = \alpha(t) - \delta(t)$
- $\lambda_t = \dfrac{\alpha(t)}{t}$: is the mean arrival intensity in $(0, t)$
- $\gamma(t)$: is integral of the difference between $\alpha(t)$ and $\delta(t)$ over the $(0, t)$ interval (*i.e. the cumulative time the customers spent in the system in* $(0, t)$) – r.v.
- T_t: the average time a customer spent in the system in $(0, t)$, *also referred to as delay*

$$T_t = \frac{\gamma(t)}{\alpha(t)}$$

- X_t: the average number of customers in the system in $(0, t)$

$$X_t = \frac{\gamma(t)}{t} = \frac{T_t\ \alpha(t)}{t} = \frac{T_t\ \lambda_t\ t}{t} = \lambda_t\ T_t$$

from which the Little's rule comes by considering the initial conditions.

3. Application of Little's rule for a part of the system:

- server(s):

$$\bar{X_s} = \bar{\lambda}\ \bar{x} \tag{3.6}$$

- queue (or buffer):

$$\bar{X_w} = \bar{\lambda}\ \bar{W} \tag{3.7}$$

- for a loss system:

$$\bar{X} = (1 - p_{loss})\ \bar{\lambda}\ \bar{T} \tag{3.8}$$

3.3 Standard notation of queueing systems (*by Kendall*)

1.
$$A/B/c/d/e-x$$

- A: the type of the interarrival time distribution.
  options:
  - M: memoryless (continuous time $\rightarrow$ exponential distr., discrete time $\rightarrow$ geometric distr.)
  - $Geom$: discrete time, memoryless, i.e. geometric distr.
  - D: deterministic
  - G: any general distr. (the consecutive interarrival times can be dependent)
  - GI: general i.i.d.
- B: the type of the service time distribution, the options are the same as for A.
- c: number of servers (finite or infinite).
- d: the capacity of the system, i.e. the number of servers plus the size of the buffer.
- e: number of sources submitting customers
- x: the order of service

FIFO (FCFS)	–	First In First Out (First Come First Served)
LIFO (LCLS)	–	Last In First Out (Last Come First Served)
RO	–	Random Order
RR	–	Round Robin
PS	–	Processor Sharing
Priority		

2. examples:

- $M/M/1 \quad \leftarrow \quad M/M/1/\infty/\infty - FIFO$
- $M/M/1 - LIFO \quad \leftarrow \quad M/M/1/\infty/\infty - LIFO$
- $M/G/2/3/10 \quad \leftarrow \quad M/G/2/3/10 - FIFO$
- $G/M/1//6 \quad \leftarrow \quad G/M/1/\infty/6 - FIFO$

3.4 The $M/M/1$ queue

1. special birth-death process (*infinite state space*)
   - arrival process: Poisson, with intensity λ
   - service time: exponentially distributed, with parameter μ
   - service order: $FIFO$
2. CTMC representation:
   elements of the generator matrix:

$$q_{ij} = \begin{cases} \lambda & j = i+1,\ i \geq 0 \\ \mu & j = i-1,\ i > 0 \\ -\lambda - \mu & j = i,\ i > 0 \\ -\lambda & j = i,\ i = 0 \\ 0 & \text{otherwise} \end{cases}$$

3. state transition graph
4. the system is stabil if $0 < \lambda < \mu < \infty$

$$p_{k-1}\lambda + p_{k+1}\mu = p_k(\lambda + \mu), \quad k > 0$$

$$p_1\mu = p_0\lambda$$

   from which

$$p_k = \frac{\lambda}{\mu}\ p_{k-1} = p_0\ (\frac{\lambda}{\mu})^k, \quad \forall k \geq 0 \tag{3.9}$$

   and

$$p_0 = \frac{1}{\sum_{k=0}^{\infty}(\frac{\lambda}{\mu})^k} = 1 - \frac{\lambda}{\mu} \tag{3.10}$$

5. the probability that the server is busy — utilization

   introduce $\rho = \lambda/\mu < 1$, hence

$$p_0 = 1 - \rho \qquad p_k = (1-\rho)\rho^k \qquad k \geq 0 \tag{3.11}$$

$$\mathbf{P}(\text{the server is busy}) = \sum_{k=1}^{\infty}\ p_k\ = 1 - p_0\ = 1 - (1-\rho)\ =\ \rho \tag{3.12}$$

On the other hand: $\mathbf{E}(X_s) = 0\ p_0 + 1\ (1 - p_0)$ from which

$$\mathbf{E}(X_s) = \rho \tag{3.13}$$

It is in keeping with the Little's rule. If $\rho < 1$ (i.e. the system is stabil)

$$\mathbf{E}(X_s) = \bar{\lambda}\bar{x} = \frac{\lambda}{\mu} = \rho \tag{3.14}$$

since $\bar{\lambda} = \lambda$ and $\bar{x} = \dfrac{1}{\mu}$.

6. the average number of customers in the system

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} k\ p_k\ = \sum_{k=0}^{\infty} k\ (1-\rho)\rho^k\ = \rho(1-\rho)\ \sum_{k=0}^{\infty} k\ \rho^{k-1}\ =$$

$$\rho(1-\rho)\ \sum_{k=0}^{\infty} \frac{d}{d\rho}\rho^k\ = \rho(1-\rho)\ \frac{d}{d\rho}\ \sum_{k=0}^{\infty} \rho^k\ = \rho(1-\rho)\ \frac{1}{(1-\rho)^2}$$

hence:

$$\mathbf{E}(X) = \frac{\rho}{1-\rho} \tag{3.15}$$

7. The average number of waiting customers

$$\mathbf{E}(X_w) = \sum_{k=1}^{\infty}\ (k-1)\ p_k = \sum_{k=1}^{\infty}\ k\ p_k - \sum_{k=1}^{\infty}\ p_k =$$

$$\sum_{k=1}^{\infty}\ k\ (1-\rho)\rho^k - (1-p_0) = \frac{\rho}{1-\rho} - \rho$$

from which

$$\mathbf{E}(X_w) = \frac{\rho^2}{1-\rho} \tag{3.16}$$

One can see that

$$X = X_w + X_s \tag{3.17}$$

and so

$$\mathbf{E}(X) = \mathbf{E}(X_w) + \mathbf{E}(X_s) \tag{3.18}$$

from which

$$\mathbf{E}(X_w) = \mathbf{E}(X) - \mathbf{E}(X_s) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho} \tag{3.19}$$

8. time parameters (T, W, x)

$$T = W + x \tag{3.20}$$

and so

$$\bar{T} = \bar{W} + \bar{x} \tag{3.21}$$

by the Little's rule:

$$\bar{T} = \frac{\mathbf{E}(X)}{\lambda} = \frac{1}{\mu(1-\rho)} \tag{3.22}$$

$$\bar{W} = \frac{\mathbf{E}(X_w)}{\lambda} = \frac{\rho}{\mu(1-\rho)} \tag{3.23}$$

$$\bar{x} = \frac{1}{\mu} \tag{3.24}$$

which satisfies (3.21).

An other way to evaluate $\bar{T}$:

$$\bar{T} = \frac{1}{\mu(1-\rho)} = \frac{1-\rho+\rho}{\mu(1-\rho)} = \frac{1}{\mu} + \frac{\rho}{\mu(1-\rho)} = \frac{1}{\mu} + \frac{1}{\mu}\frac{\rho}{1-\rho} = \frac{1}{\mu} + \sum_{k=0}^{\infty} k\ \frac{1}{\mu}\ p_k$$

(*Which means that the system state distribution is the same as the steady state distribution at the arrival instances of a Poisson arrival process.*)

9. (*the FIFO service order do not play role in the overall system behaviour only in the waiting time distribution of the customers.*)

10. the probability of more than k customers in the system:

$$\mathbf{P}(X \geq k) = \sum_{i=k}^{\infty}(1-\rho)\rho^i = (1-\rho)\rho^k \sum_{j=0}^{\infty} \rho^j = \rho^k \tag{3.25}$$

11. Example 1: There is a multiplexer which receives messages from a group of terminals and transmit them through a transmission link. The messages arrive according to a Poisson process with an average of $1 message/4ms$. The transmission time of the messages are exponentially distributed with a mean of $3ms$.

    What is the average number of messages at the multiplexer?

$$\rho = \frac{1}{4}3 = \frac{3}{4}$$

$$\mathbf{E}(X) = \frac{\rho}{1-\rho} = 3$$

What is the average time a message spends at the multiplexer?

$$\mathbf{E}(T) = \frac{\mathbf{E}(X)}{\lambda} = \frac{3}{1/4} = 12 msec$$

What increase of the arrival rate would result in a double average system time?

$$\mathbf{E}(T') = 24 = \frac{\mathbf{E}(x)}{1-\rho'} = \frac{3}{1-\rho'}$$

$$\rho' = 1 - \frac{1}{8} = 7/8$$

from which:

$$\lambda' = \rho'\mu = \frac{7}{8\,3} = \frac{7}{24}$$

i.e. a 17% increase in the arrival rate doubles the system time of the messages.

12. Example 2: There is a fast computer which serves the jobs with a service rate of $K\mu$, (i.e. the service time of the jobs are exponentially distributed with parameter $K\mu$). The jobs arrive according to a Poisson process with parameter $K\lambda$. There is another set of computers. Each of them has a service rate of μ. Suppose we submit jobs to each of them according to a Poisson process with parameter λ (*see the random decomposition of Poisson process*), we are interested in the performance parameters of the two computer systems.

    The single fast computer system:

$$\rho = \frac{K\lambda}{K\mu} = \frac{\lambda}{\mu}$$

$$\mathbf{E}(T) = \frac{\mathbf{E}(x)}{1-\rho} = \frac{1}{K\mu(1-\rho)}$$

    The system of K independent computers:

$$\rho = \frac{\lambda}{\mu}$$

    for each of the computers and

$$\mathbf{E}(T') = \frac{\mathbf{E}(x)}{1-\rho} = \frac{1}{\mu(1-\rho)} = K\mathbf{E}(T)$$

    i.e. the average time a job spends at a computer is K times higher.

3.5 The $M/M/m$ system

1. special birth-death process (*infinite state space*)
   - arrival process: Poisson with parameter λ
   - service time: exponentially distributed with parameter μ in each servers
   - service order: $FIFO$
2. CTMC interpretation:
   the elements of the generator matrix:
$$q_{ij} = \begin{cases} \lambda & j = i+1,\ i \geq 0 \\ i\mu & j = i-1,\ 0 < i \leq m \\ m\mu & j = i-1,\ m < i \\ -\lambda - i\mu & j = i,\ 0 \leq i \leq m \\ -\lambda - m\mu & j = i,\ i \geq m \\ 0 & \text{otherwise} \end{cases}$$
3. state transition graph
4. steady state distribution: exists (the system is stabil) if $0 < \lambda < m\mu < \infty$
$$p_{k-1}\lambda + p_{k+1}(k+1)\mu = p_k(\lambda + k\mu), \quad 0 < k < m$$
$$p_{k-1}\lambda + p_{k+1}m\mu = p_k(\lambda + m\mu), \quad k \geq m$$
$$p_1\mu = p_0\lambda$$
   from which
$$p_k = \frac{\lambda}{k\mu}p_{k-1} = \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}p_0 \quad k = 1, 2, \cdots, m \tag{3.26}$$
   and
$$p_k = \frac{\lambda}{m\mu}p_{k-1} \quad k = m+1, m+2, \cdots \tag{3.27}$$
   hence
$$p_{m+i} = \left(\frac{\lambda}{m\mu}\right)^i p_m \quad i \geq 1 \tag{3.28}$$
   To sum these up:
$$p_j = \begin{cases} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!}\, p_0 & j = 1, 2, \cdots, m \\ \left(\frac{\lambda}{\mu}\right)^j \frac{1}{m!}\frac{1}{m^{j-m}}\, p_0 & j > m \end{cases} \tag{3.29}$$

Expressing p_0:

$$p_0 = \frac{1}{1 + \sum_{j=1}^{m} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} + \sum_{j=m+1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{m!} \frac{1}{m^{j-m}}} \tag{3.30}$$

in which the second term can be written as:

$$\frac{m^m}{m!} \sum_{j=m}^{\infty} \left(\frac{\lambda}{m\mu}\right)^j = \frac{\left(\frac{\lambda}{m\mu}\right)^m}{\left(1 - \frac{\lambda}{m\mu}\right)} \frac{m^m}{m!} \tag{3.31}$$

5. Time parameters:

$$\bar{T} = \bar{x} + \bar{W} = \frac{1}{\mu} + \sum_{k=m}^{\infty} E[W \mid k]\ p_k^{(a)} \tag{3.32}$$

where $p_k^{(a)}$ is the state probability distribution at the arrival instances. From which:

$$E[T] = \frac{1}{\mu} + \sum_{k=m}^{\infty} \frac{k-m+1}{m\mu} p_k \tag{3.33}$$

and:

$$\bar{T} = \frac{1}{\mu} + \sum_{k=m}^{\infty} \frac{k-m+1}{m\mu} p_m \left(\frac{\lambda}{m\mu}\right)^{k-m} = \frac{1}{\mu} + \frac{p_m}{m\mu} \sum_{k=m}^{\infty} (k-m+1) \left(\frac{\lambda}{m\mu}\right)^{k-m} \tag{3.34}$$

since we have a Poisson arrival process. Further more:

$$\bar{T} = \frac{1}{\mu} + \frac{p_m}{m\mu} \sum_{i=1}^{\infty} i \left(\frac{\lambda}{m\mu}\right)^{i-1} = \frac{1}{\mu} + \frac{m\mu p_m}{(m\mu - \lambda)^2} \tag{3.35}$$

and applying the Little's rule:

$$\mathbf{E}(X) = \lambda \bar{T} = \frac{\lambda}{\mu} + \frac{m\lambda\mu p_m}{(m\mu - \lambda)^2} \tag{3.36}$$

6. Telecommunication systems with $M/M/m$ queue

   - waiting systems: (for example mobile telephony)

- the probability of waiting for service

$$P\{\text{waiting}\} = \sum_{k=m}^{\infty} p_k = \sum_{k=m}^{\infty} p_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{m!}\frac{m^m}{m^k} = \frac{m^m}{m!} p_0 \sum_{k=m}^{\infty} \left(\frac{\lambda}{m\mu}\right)^k$$

$$= p_0 \frac{m^m}{m!} \frac{\left(\frac{\lambda}{m\mu}\right)^m}{1 - \frac{\lambda}{m\mu}} = p_0 \frac{1}{m!} \left(\frac{\lambda}{\mu}\right)^m \frac{1}{1-\rho} = \frac{\frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{1}{1-\rho}}{\sum_{k=0}^{m-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{1}{1-\rho}}$$

where $\rho = \frac{\lambda}{m\mu}$.
This expression is wellknown as the $ErlangC$ formula with parameters: m and λ/μ, and the usual notation is $C(m, \lambda/\mu)$.

7. Example 3: There are 4 telephone lines between two buildings of a company. Calls arrive according to a Poisson process with an average of 1 call per 2 minits. The communication time is exponentially distributed with a mean of 4 minits. The calls are waiting if all of the 4 lines are busy. What is the probability that a new call should wait for a free line?

$$\lambda = 1/2, \quad 1/\mu = 4, \quad a = \lambda/\mu = 2, \quad \rho = a/m = 2/4 = 0.5$$

And so

$$p_0 = \frac{1}{1 + 2 + 2^2/2 + 2^3/6 + 16/24(1/(1-0.5))} = 3/23$$

hence

$$C(4,2) = \frac{2^4/4!}{1-0.5}\frac{3}{23} = 4/23 = 0.17$$

8. Example 4: Compare the performance parameter of the $M/M/1$ and the $M/M/2$ queueing systems if arrival rate is $\lambda = 1/2$ in both systems while $\mu_1 = 1$ and $\mu_2 = 1/2$ are the service rate of the $M/M/1$ and the $M/M/2$ queue respectively.
$M/M/1$ queue:

$$\rho = \frac{\lambda}{\mu_1} = \frac{1/2}{1} = 0.5$$

$$\mathbf{E}(W) = \frac{\rho/\mu}{1-\rho} = 1sec$$

$$\mathbf{E}(T) = \frac{1/\mu}{1-\rho} = 2sec$$

$M/M/2$ queue:

$$a = \frac{\lambda}{\mu_2} = \frac{1/2}{1/2} = 1,\ \rho = a/m = 1/2 = 0.5$$

$$p_0 = \frac{1}{1 + 2 + \frac{a^2/2}{1-0.5}} = 1/3$$

from which

$$C(2,1) = \frac{a^2/2}{1-0.5} p_0 = 1/3$$

$$\mathbf{E}(W') = \frac{1/\mu_2}{1-\rho} C(2,1) = 2/3$$

$$\mathbf{E}(T) = 2/3 + 1/\mu_2 = 8/3 sec$$

i.e. the system time is less in the $M/M/1$ queue, but the waiting time is more.

3.6 The $M/M/\infty$ queue

1. It is a special sace of the $M/M/m$ queue

$$\begin{cases} \lambda_i = \lambda & i \geq 0 \\ \mu_i = i\mu & i \geq 1 \end{cases} \tag{3.37}$$

hence

$$p_k = p_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} \qquad k \geq 1 \tag{3.38}$$

and so

$$p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}} = \frac{1}{\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}} = \mathrm{e}^{-\lambda/\mu} \tag{3.39}$$

if $0 < \lambda, \mu < \infty$ the CTMC is stabil and:

$$p_k = \frac{(\lambda/\mu)^k}{k!} \mathrm{e}^{-\lambda/\mu} \qquad k \geq 0 \tag{3.40}$$

the number of customers in the system is Poisson distributed with a mean of

$$E[X] = \bar{\lambda}\, \bar{T} = \lambda \frac{1}{\mu} = \frac{\lambda}{\mu} \tag{3.41}$$

3.7 The $M/M/m/m$ queue

1. It is an other special case of the $M/M/m$ queue

$$\begin{cases} \lambda_i = \lambda & 0 \leq i < m \\ \mu_i = i\mu & 0 < i \leq m \end{cases} \tag{3.42}$$

2. the steady state distribution is:

$$p_k = p_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} \quad k = 1, 2, \cdots, m \tag{3.43}$$

where

$$p_0 = \frac{1}{\sum_{k=0}^{m} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}} \tag{3.44}$$

The $M/M/m/m$ queue is stabil ($p_k > 0 \;\; \forall k$) if $0 < \lambda, \mu < \infty$ i.e. when the mean interarrival time and the mean service time are positive and finite.

3. The mean arrival rate (of the accepted customers)

$$\bar{\lambda} = \sum_{k=0}^{m-1} \lambda_k p_k = \lambda(1 - p_m) \tag{3.45}$$

4. the average number of customers in the system

$$E[X] = \frac{\lambda}{\mu}(1 - p_m) \tag{3.46}$$

(check the Little's rule.)

5. Traditional telephony — lossy system !!

$$P\{\text{loss}\} = p_m^{(a)} = p_m = \frac{\left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!}}{\sum_{k=0}^{m} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}} = B(m, \lambda/\mu) \tag{3.47}$$

which is the wellknown Erlang B formula. The first 80 years of the telephone traffic engineering was based on this expression.

6. Example 5: There are 4 telephone lines between two buildings of a company. Calls arrive according to a Poisson process with an average of 1 call per 2 minits. The communication time is exponentially distributed with a mean of 4 minits. The calls are lost if all of the 4 lines are busy.

   What is the probability that a new call is lost? Compare the results with Example 3. Where the difference comes from?

$$p_{loss} = B(4,2) = \frac{16/24}{1+2+2^2/2+2^3/6+16/24} = \frac{2/3}{5+4/3+2/3} = 2/21 = 0.095$$

   $0.095 < 0.17$, since the lost customers do not occupy the server, while the delayd customers do. i.e. $B(4,2) < C(4,2)$.

3.8 The $M/M/1//N$ queue

1. the elements of the infinitezimal generator

$$\lambda_i = (N-i)\lambda \quad 0 \le i \le N$$

$$\mu_i = \mu \qquad i \ge 1$$

2. the average arrival rate

$$\bar{\lambda} = \sum_{i=0}^{N} \lambda_i p_i = \sum_{i=0}^{N-1} (N-i)\,\lambda\, p_i \tag{3.48}$$

3. the CTMC stabil (i.e. $p_k > 0, \quad \forall 0 \le k \le N$) if the mean interarrival time and the mean service time are positive and finite.

4. limit distribution

$$p_k = p_0 \left(\frac{\lambda}{\mu}\right)^k [N(N-1)\cdots(N-k+1)] = p_0 \left(\frac{\lambda}{\mu}\right)^k \frac{N!}{(N-k)!} \quad k = 1,2,\cdots,N \tag{3.49}$$

   where

$$p_0 = \frac{1}{1+\sum_{j=1}^{N}\left(\frac{\lambda}{\mu}\right)^j \frac{N!}{(N-j)!}} \tag{3.50}$$

5. utilization: $\rho = 1 - p_0$

6. By the Little's rule:

$$\rho = \bar{\lambda}\mathbf{E}(x) = \bar{\lambda}/\mu$$

   Hence the average arrival rate is:

$$\bar{\lambda} = \frac{\rho}{\mathbf{E}(x)} = \mu\rho = \mu(1 - p_0) \tag{3.51}$$

7. An other way to evaluate $\bar{\lambda}$:
   every source submits a customer in every $(1/\lambda + \mathbf{E}(T))^{-1}$ time units in average. From which the mean arrival rate of the N sources:

$$\bar{\lambda} = \frac{N}{1/\lambda + \mathbf{E}(T)}$$

   and

$$\mathbf{E}(T) = \frac{N}{\bar{\lambda}} - \frac{1}{\lambda} \tag{3.52}$$

   by the Little's rule:

$$\mathbf{E}(X) = \bar{\lambda}\mathbf{E}(T) = N - \frac{\bar{\lambda}}{\lambda} \tag{3.53}$$

   and from $\mathbf{E}(T)$ we have $\mathbf{E}(W) = \mathbf{E}(T) - \dfrac{1}{\mu}$.

8. The probability of that a source is busy, i.e. it is waiting for the completion of the submitted customer:

$$\mathbf{P}(\text{a source is busy}) = \frac{1/\lambda}{1/\lambda + \mathbf{E}(T)}$$

9. Example 6: In a computer system 6 terminals are served by a central unit. Each of the terminals submits jobs after an exponentially distributed preparation time with parameter λ. The completion time of a job is exponentially distributed with parameter μ.

   Evaluate the throughput (the mean departure rate) and the mean system time (a job sends in the system) for the two extreme cases, when N is large, and when N is small.

- when N is small (i.e. waiting time ~ 0):

$$\mathbf{E}(T) = \frac{1}{\mu}$$

$$\bar{\lambda} = \frac{N}{1/\lambda + \mathbf{E}(T)} = \frac{N}{1/\lambda + 1/\mu}$$

- when N is large (the server is almost always busy):

$$\bar{\lambda} = \mu$$

$$\mathbf{E}(T) = \frac{N}{\mu} - \frac{1}{\lambda}$$

- diagrams

3.9 The $M/M/m/K/N$ queue

1. the elements of the generator matrix

$$q_{ij} = \begin{cases} (N-i)\lambda & j = i+1,\ 0 \le i < N \\ i\mu & j = i-1,\ 0 < i \le m \\ m\mu & j = i-1,\ m \le i \le K \\ -(N-i)\lambda - i\mu & j = i,\ 0 \le i \le m \\ -(N-i)\lambda - m\mu & j = i,\ m \le i < K \\ -m\mu & j = i = K, \\ 0 & \text{otherwise} \end{cases}$$

2. state transition graph

3. condition of stability

3.10 Queueing systems with associated general CTMC

1. limits of queues with associated birth-death process

   - exponentially distributed interarrival time
   - exponentially distributed service time

- different customers
- different servers
- batch arrival
- batch service

2. *Erlang* n distributed service (interarrival) time (E_n)
   can be considered as n consecutive exponential services

   - the PDF of one exponential service

$$f_{\gamma_i}(t) = n\mu \mathrm{e}^{-n\mu t} \tag{3.54}$$

   - the parameters of one exponential service

$$\mathbf{E}(\gamma_i) = \frac{1}{n\mu} \tag{3.55}$$

$$\sigma_i^2 = \mathbf{E}(\gamma_i^2) - (\mathbf{E}(\gamma_i))^2 = \frac{1}{(n\mu)^2} \tag{3.56}$$

$$C_i^2 \triangleq \frac{\sigma_i^2}{\mathbf{E}(\gamma_i)^2} = 1 \tag{3.57}$$

   where $\mathbf{E}(\gamma_i)$ denotes the mean of γ_i, σ_i^2 its standard deviation and C_i its (*coefficient of variation*)

   - Let

$$\gamma = \sum_{i=1}^{n} \gamma_i = n\gamma_i \tag{3.58}$$

   - state transition graph
   - the PDF and the related parameters of the time of n consecutive services

$$f_{\gamma}^{(n)}(t) = \frac{n\mu(n\mu t)^{n-1}}{(n-1)!}\mathrm{e}^{-n\mu t} \tag{3.59}$$

$$\mathbf{E}(\gamma) = \frac{1}{\mu} \quad \forall n \tag{3.60}$$

$$\sigma^2 = n\sigma_i^2 = \frac{1}{n\mu^2} \tag{3.61}$$

$$C_{\gamma}^2 \triangleq \frac{\sigma^2}{\mathbf{E}(\gamma)^2} = \frac{1}{n} \leq 1 \tag{3.62}$$

- diagram

3. paralell service (arrival) H_n nth order hiperexponential distribution

   - n paralell service units are available with service rate $\mu_i, \quad i = 1, \cdots, n$, the ith one is choosen with probability $\alpha_i, \quad i = 1, \cdots, n$, the overall PDF of the service time is:
   $$f_\eta(t) = \sum_{i=1}^{n} \alpha_i \; \mu_i \; \mathrm{e}^{-\mu_i t} \tag{3.63}$$
   where
   $$\sum_{i=1}^{n} \alpha_i = 1 \tag{3.64}$$

   - parameters of the service time
   $$\mathbf{E}(\eta) = \sum_{i=1}^{n} \frac{\alpha_i}{\mu_i} \tag{3.65}$$
   $$\mathbf{E}(\eta^2) = 2 \sum_{i=1}^{n} \frac{\alpha_i}{\mu_i^2} \tag{3.66}$$
   $$C_\eta^2 = \frac{2 \sum_{i=1}^{n} \frac{\alpha_i}{\mu_i^2}}{\sum_{i=1}^{n} \frac{\alpha_i}{\mu_i}} - 1 \tag{3.67}$$
   Cauchy-Schwarz-Bunyakovszkij showed that C_η^2 is greater than 1

   - example: two paralell secvice unit
   $$f_\eta(t) = \alpha \mu_1 \mathrm{e}^{-\mu_1 t} + (1-\alpha) \mu_2 \mathrm{e}^{-\mu_2 t} \tag{3.68}$$
   $$\mathbf{E}(\eta) = \frac{\alpha}{\mu_1} + \frac{1-\alpha}{\mu_2} \tag{3.69}$$
   $$\mathbf{E}(\eta^2) = \frac{2\alpha}{\mu_1^2} + \frac{2(1-\alpha)}{\mu_2^2} \tag{3.70}$$
   $$C_\eta^2 = \frac{2\left[\frac{\alpha}{\mu_1^2} + \frac{(1-\alpha)}{\mu_2^2}\right]}{\left(\frac{\alpha}{\mu_1} + \frac{1-\alpha}{\mu_2}\right)^2} - 1 \tag{3.71}$$

4. Phase-type distribution — their application and importance

   - (*approximation of general distributions*)
   - (*system description by CTMC*)

- (*drawback: state space expansion*)

5. Queues with associated general CTMC — state transition graph

6. Example 7: Jobs arrive to a computer system according to a Poisson process with parameter λ. Each jobs require two phases of service. Phase 1 (2) is exponentially distributed with parameter μ_1 (μ_2).

   Consider the following 4 cases:

   (a) there is only a single processor:
      i. there is no buffer in the system,
      ii. there is an infinite buffer in the system.

   (b) there are two processors associated to the two phases of service:
      i. there is no buffer in the system,
      ii. there are two infinite buffers accosiater to the two processors,

Excercises:

(a) What is the system utilization (ρ)?
cases: 6(a)i, 6(a)ii.

(b) What is the probability that the processors are busy?
cases: 6(b)i, 6(b)ii.

(c) When the first job is lost (in average) if the system started from the empty state.
case: 6(a)i.

(d) What the total system time and the waiting time in the buffer of the first processor is?
cases: 6(a)ii, 6(b)ii.

(e) What is the maximum arrival rate at which the loss is less than 1/3 and $1/\mu_1 = 5$ sec, $1/\mu_2 = 20$ sec.
case: 6(a)i.

(f) What is the value of (ρ) and ($\mathbf{E}(T)$) if the arrival is according to 6e.
case: 6(a)ii.

Solutions I.:

6a–6(a)i. Draw the state transition graph!

The steady state can be evaluated as:

$$\lambda p_0 = \mu_2 p_2 \quad (3.72)$$
$$\mu_1 p_1 = \lambda p_0 \quad (3.73)$$
$$\mu_1 p_1 = \mu_2 p_2 \quad (3.74)$$

Since $p_0 + p_1 + p_2 = 1$, we have:

$$\rho = 1 - p_0 = 1 - \frac{1}{1 + \frac{\lambda}{\mu_1} + \frac{\lambda}{\mu_2}}$$

6b–6(b)i.

$$\mathbf{P}(\text{both of the processors are busy}) = p_3 = \frac{\lambda^2 \mu_1}{\lambda^2 \mu_1 + (\mu_1 + \mu_2)(\lambda\mu_1 + \lambda\mu_2 + \mu_1\mu_2)}.$$

6e–6(a)i.

$$\mathbf{P}(\text{loss}) = \mathbf{P}(\text{the processor is busy}) = \rho$$

$$\rho = 1 - p_0 = 1 - \frac{1}{1 + \frac{\lambda}{\mu_1} + \frac{\lambda}{\mu_2}} < \frac{1}{3}$$

from which

$$\lambda < \frac{1}{50} 1/\text{sec}.$$

3.11 The $M/G/1$ queue

General considerations on the $M/G/1$ queue

- to avoid the state PH space expansion
- problem of general service time
  - $X(t)$ is not a CTMC, since it is not memoryless in general,
  - to characterize the future evolution of $X(t)$ not only the system state but the time for the last customer is under service is necessary.

3.11.1 Residual time

1. paradox: exponential interarrival times — random time instant — residual time to the next arrival (distribution, moments)

2. notations:

   - ξ: interarrival time r.v.
   - γ: residual time from a random time instant to the next arrival r.v.

3. main result: the nth moment of γ can be evaluated as

$$\mathbf{E}(\gamma^n) = \frac{\mathbf{E}(\xi^{n+1})}{(n+1)\mathbf{E}(\xi)}$$

   - mean residual time:

$$\mathbf{E}(\gamma) = \frac{\mathbf{E}(\xi^2)}{2\mathbf{E}(\xi)} = \frac{(\mathbf{E}(\xi))^2 + \sigma^2}{2\mathbf{E}(\xi)} = \mathbf{E}(\xi)\frac{1 + C_\xi^2}{2}$$

   - example, deterministic interarrival time: $\mathbf{E}(\xi) = m$, $\mathbf{E}(\xi^2) = m^2$, $C_\xi^2 = 0$

$$\mathbf{E}(\gamma) = \frac{m^2}{2m} = m/2$$

   - example, exponencial interarrival time: $\mathbf{E}(\xi) = m$, $\mathbf{E}(\xi^2) = 2m^2$, $C_\xi^2 = 1$

$$\mathbf{E}(\gamma) = \frac{2m^2}{2m} = m$$

4. Wald's equation
   For x_i i.i.d. r.v. and an η r.v.

$$\mathbf{E}(\textstyle\sum_{i=1}^{\eta} x_i) = \mathbf{E}(\eta)\mathbf{E}(x)$$

5. M/G/1 queue: mean waiting time ($\mathbf{E}(W)$)
   Let x r.v. the service time of a customer, γ' r.v. the residual service time of the customer under service when a new customer arrives. And by the Wald's equation we have

$$\mathbf{E}(W) = \mathbf{E}(\gamma') + \mathbf{E}(X_w)\mathbf{E}(x)$$

6. From the Little's rule: $\mathbf{E}(X_w) = \lambda\mathbf{E}(W)$, and so

$$\mathbf{E}(W) = \mathbf{E}(\gamma') + \lambda\mathbf{E}(W)\mathbf{E}(x) = \mathbf{E}(\gamma') + \rho\mathbf{E}(W)$$

7. Since $\mathbf{E}(\gamma') = 0\,\mathbf{P}(X(t) = 0) + \mathbf{E}(\gamma)\mathbf{P}(X(t) > 0)$:

$$\mathbf{E}(\gamma') = \frac{\mathbf{E}(x^2)}{2\mathbf{E}(x)}\lambda\mathbf{E}(x) = \frac{\lambda\mathbf{E}(x^2)}{2}$$

and from the above expressions:

$$\mathbf{E}(W) = \frac{\mathbf{E}(\gamma')}{1-\rho} = \frac{\lambda\mathbf{E}(x^2)}{2(1-\rho)} = \lambda(\mathbf{E}(x))^2\frac{1+C_x^2}{2(1-\rho)}$$

i.e.

$$\mathbf{E}(W) = \rho\mathbf{E}(x)\frac{1+C_x^2}{2(1-\rho)} = \frac{\rho}{1-\rho}\mathbf{E}(x)\frac{1+C_x^2}{2} \tag{3.75}$$

Which is the famous Pollaczek-Hincsin mean value formula

8. the mean system time: $\mathbf{E}(T)$

$$\mathbf{E}(T) = \mathbf{E}(x) + \mathbf{E}(W) = \mathbf{E}(x) + \frac{\rho}{1-\rho}\mathbf{E}(x)\frac{1+C_x^2}{2} \tag{3.76}$$

9. examples:

   - M/D/1 queue *(deterministic service time)*: $\mathbf{E}(x) = m$, $\mathbf{E}(x^2) = m^2$, $C_x^2 = 0$

   $$\mathbf{E}(W) = \frac{\rho}{2(1-\rho)}\mathbf{E}(x)$$

   - M/M/1 queue: $\mathbf{E}(x) = m$, $\mathbf{E}(x^2) = 2m^2$, $C_x^2 = 1$

   $$\mathbf{E}(W) = \frac{\rho}{1-\rho}\mathbf{E}(x)$$

3.11.2 Average number of customers in an M/G/1 queue

1. system state can be characterized by the pair of $(X(t), Z(t))$ (*where* $X(t)$ *is discrete and* $Z(t)$ *is continuous*)

$$X(t) \triangleq \text{the number of customers in the system at time } t$$

$$Z(t) \triangleq \text{the residual service time of the customer under service at time } t$$

Neither $(X(t), Z(t))$ nor $(X(t), Z(t))$ is a CTMC.

2. solution: look for an embedded DTMC !!!!
   (for example the start of the completion of sercives)
   We consider the embedded DTMC at departure times below

3. notation:

   $X_n \triangleq$ number of customers remained in the system at the departure of the nth customer

   $Y_n \triangleq$ number of customers arrived to the system during the service of the nth customer

4. the evolution rule of the embedded DTMC:

$$X_{n+1} = \begin{cases} X_n + Y_{n+1} - 1 & \text{ha } X_n > 0 \\ Y_{n+1} & \text{ha } X_n = 0 \end{cases} \tag{3.77}$$

$$X_{n+1} = X_n + Y_{n+1} - \chi(X_n) \tag{3.78}$$

   where

$$\chi(X) = \begin{cases} 0 & \text{if the system is empty} \\ 1 & \text{if the system is not empty} \end{cases} \tag{3.79}$$

   which shows that the $\{X_n, n = 0, 1, \cdots\}$ serie of r.v. is a DTMC.

5. the average number of customers in the system:

$$\mathbf{E}(X_{n+1}) = \mathbf{E}(X_n) + \mathbf{E}(Y_{n+1}) - \mathbf{E}(\chi(X_n)) \tag{3.80}$$

   in equilibrium

$$\lim_{n\to\infty} \mathbf{E}(X_n) = \lim_{n\to\infty} \mathbf{E}(X_{n+1}) = \mathbf{E}(X) \tag{3.81}$$

   further more $\mathbf{E}(Y_n) \to \mathbf{E}(Y)$ és $\mathbf{E}(\chi(X_n)) \to \mathbf{E}(\chi(X))$. And so

$$\mathbf{E}(X) = \mathbf{E}(X) + \mathbf{E}(Y) - \mathbf{E}(\chi(X)) \tag{3.82}$$

   from which

$$\mathbf{E}(Y) = \mathbf{E}(\chi(X)) = \rho \tag{3.83}$$

   since

$$\mathbf{E}(\chi(X)) = 0\mathbf{P}(X = 0) + 1\mathbf{P}(X > 0) = \mathbf{P}(\text{the server is busy}) = \rho \tag{3.84}$$

   Taking the square of equation 3.78 we get :

$$X_{n+1}^2 = X_n^2 + Y_{n+1}^2 + \chi(X_n) + 2X_nY_{n+1} - 2X_n\chi(X_n) - 2Y_{n+1}\chi(X_n) \tag{3.85}$$

and the mean of both sides is as follows:

$$\begin{aligned}\mathbf{E}\Big(X_{n+1}^2\Big) &= \mathbf{E}(X_n^2) + \mathbf{E}\Big(Y_{n+1}^2\Big) + \mathbf{E}(\chi(X_n)) + \\ &\quad 2\mathbf{E}(X_n)\mathbf{E}(Y_{n+1}) - 2\mathbf{E}(X_n) - 2\mathbf{E}(Y_{n+1})\mathbf{E}(\chi(X_n))\end{aligned}$$

the limit $n \to \infty$ gives:

$$0 = \mathbf{E}(Y^2) + \mathbf{E}(Y) + 2\mathbf{E}(X)\mathbf{E}(Y) - 2\mathbf{E}(X) - 2(\mathbf{E}(Y))^2 \tag{3.86}$$

$$2(1 - \mathbf{E}(Y))\mathbf{E}(X) = \mathbf{E}(Y^2) + \mathbf{E}(Y) - 2(\mathbf{E}(Y))^2 \tag{3.87}$$

$$\mathbf{E}(X) = \mathbf{E}(Y) + \frac{\mathbf{E}(Y^2) - \mathbf{E}(Y)}{2\,(1 - \mathbf{E}(Y))} \tag{3.88}$$

Using $\mathbf{E}(Y) = \rho = \lambda\mathbf{E}(x)$, we finally have:

$$\begin{aligned}\mathbf{E}(Y^2) &= \sum_{k=1}^{\infty} k^2\mathbf{P}(Y = k) = \sum_{k=1}^{\infty} k^2 \int_0^{\infty} \frac{(\lambda x)^k}{k!}\mathrm{e}^{-\lambda x}\, b(x)\, dx \\ &= \int_0^{\infty} \sum_{k=1}^{\infty} k^2 \frac{(\lambda x)^k}{k!}\mathrm{e}^{-\lambda x}\, b(x)dx \\ &= \int_0^{\infty} \sum_{k=1}^{\infty} (k\ (k-1) + k)\ \frac{(\lambda x)^k}{k!}\mathrm{e}^{-\lambda x}\, b(x)dx \\ &= \int_0^{\infty} \Big(\lambda^2 x^2 + \lambda x\Big)\ b(x)\, dx = \lambda^2\mathbf{E}(x^2) + \lambda\mathbf{E}(x)\end{aligned} \tag{3.89}$$

$$\mathbf{E}(X) = \rho + \frac{\lambda^2\mathbf{E}(x^2)}{2(1-\rho)} \tag{3.90}$$

considering the fact that

$$C_x^2 = \frac{\mathbf{E}(x^2) - (\mathbf{E}(x))^2}{(\mathbf{E}(x))^2} \tag{3.91}$$

we get:

$$\mathbf{E}(X) = \rho + \rho^2\frac{(1 + C^2)}{2(1-\rho)} \tag{3.92}$$

which is also known as the Pollaczek-Hincsin mean value formula.

6. time parameters:

$$\bar{T} = \frac{\mathbf{E}(X)}{\mathbf{E}(\lambda)} = \frac{\mathbf{E}(Y)}{\lambda} = \mathbf{E}(x) + \rho\mathbf{E}(x)\frac{1 + C^2}{2(1-\rho)} \tag{3.93}$$

This result can be checked for exponential service time as follows.

7. $M/M/1$ queue: $C^2 = 1$

$$\mathbf{E}(X) = \rho + \frac{\rho^2}{1-\rho} = \frac{\rho}{1-\rho} \tag{3.94}$$

8. $M/D/1$ queue: $C^2 = 0$

$$\mathbf{E}(X) = \rho + \frac{\rho^2}{2(1-\rho)} = \frac{\rho}{1-\rho} - \frac{\rho^2}{2(1-\rho)} \tag{3.95}$$

9. Example 7: **Solution II.:**

6a–6(a)ii. By the above results of $M/G/1$ queue:

$$\rho = \lambda \mathbf{E}(x) = \lambda(\frac{1}{\mu_1} + \frac{1}{\mu_2})$$

6d–6(a)ii. Further more:

$$T = \mathbf{E}(x) + \frac{\rho \mathbf{E}(x)(1 + C_b^2)}{2(1-\rho)}, \quad \mathbf{E}(X) = \lambda T,$$

where

$$\mathbf{E}(x) = \frac{1}{\mu_1} + \frac{1}{\mu_2} \text{ and } C_b^2 = \frac{\sigma_b^2}{(\mathbf{E}(x))^2} = \frac{\frac{1}{\mu_1^2} + \frac{1}{\mu_2^2}}{(\mathbf{E}(x))^2}.$$

6f–6(a)i.

$$\mathbf{E}(x) = \frac{1}{\mu_1} + \frac{1}{\mu_2} = 25 \text{ sec}$$

$$\rho = \lambda \mathbf{E}(x) = \lambda(\frac{1}{\mu_1} + \frac{1}{\mu_2}) = \frac{1}{50}(20 + 5) = \frac{1}{2},$$

$$C_b^2 = \frac{\sigma_b^2}{(\mathbf{E}(x))^2} = \frac{\frac{1}{\mu_1^2} + \frac{1}{\mu_2^2}}{(\mathbf{E}(x))^2} = 425/625 = 17/25.$$

$$T = \mathbf{E}(x) + \frac{\rho \mathbf{E}(x)(1 + C_b^2)}{2(1-\rho)} = 46 \text{ sec }.$$

3.11.3 Distribution of number of customers in M/G/1 queue

1. the evolution equation describes the dinamics of the system:

$$X_{n+1} = X_n + Y_{n+1} - \chi(X_n) \tag{3.96}$$

2. from which the state transition probability:

$$\mathbf{P}(X_{n+1} = j \mid X_n = i) = \mathbf{P}(Y_{n+1} - \chi(X_n) = j - i) \tag{3.97}$$

$\{X_n, n = 0, 1, \cdots\}$ is a DTMC with the following state transition probabilities:

$$\mathbf{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \mathbf{P}(Y_{n+1} - 1 = j - i) & \text{if } i > 0 \\ \mathbf{P}(Y_{n+1} = j - i) & \text{if } i = 0 \end{cases} \tag{3.98}$$

Conditioning on the service time we have:

$$\mathbf{P}(Y_n = k) = \int_0^\infty \mathbf{P}(Y_n = k \mid x_n = x)\, b(x)\, dx \tag{3.99}$$

and utilizing the Poisson arrival process:

$$\mathbf{P}(Y_n = k) = \int_0^\infty \frac{(\lambda x)^k}{k!} \mathrm{e}^{-\lambda x}\, b_n(x)\, dx \triangleq y_k \tag{3.100}$$

In matrix form:

$$\boldsymbol{\Pi} = \begin{bmatrix} y_0 & y_1 & y_2 & y_3 & \cdots & y_k & \cdots \\ y_0 & y_1 & y_2 & y_3 & \cdots & y_k & \cdots \\ 0 & y_0 & y_1 & y_2 & \cdots & y_{k-1} & \cdots \\ 0 & 0 & y_0 & y_1 & \cdots & y_{k-2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{3.101}$$

3. The equilibrium equations are:

$$p_0 = p_0 p_{00} + p_1 p_{10} = (p_0 + p_1) y_0$$

$$p_k = p_0 p_{0k} + \sum_{i=1}^{k+1} p_i p_{i,k} = p_0 y_k + \sum_{i=1}^{k+1} p_i y_{k+1-i}$$

Taking the z-transform of p_k and y_k:

$$P(z) = \sum_{k=0}^{\infty} p_k z^k, \quad Y(z) = \sum_{k=0}^{\infty} y_k z^k \tag{3.102}$$

and multiplying the kth equilibrium equation by z^k and summ them up we have:

$$P(z) = \sum_{k=0}^{\infty} p_k z^k = \sum_{k=0}^{\infty} p_0 y_k z^k + \sum_{k=0}^{\infty} \sum_{i=1}^{k+1} p_i y_{k+1-i} z^k = p_0 Y(z) + \sum_{i=1}^{\infty} p_i \sum_{k=i-1}^{\infty} y_{k+1-i} z^k =$$

$$p_0 Y(z) + z^{-1} \sum_{i=1}^{\infty} p_i z^i \sum_{l=0}^{\infty} y_l z^l = p_0 Y(z) + z^{-1}(P(z) - p_0) Y(z)$$

from which

$$zP(z) - Y(z)P(z) = p_0 Y(z)(z-1)$$

and

$$P(z) = \frac{p_0 Y(z)(1-z)}{Y(z) - z} \tag{3.103}$$

Considering that $\lim_{z \to 1} P(z) = 1$ and using the L'Hospital's rule:

$$1 = p_0 \frac{-Y(z) + (1-z)Y'(z)}{Y'(z) - 1} \, |_{z=1}$$

from which

$$p_0 = 1 - Y'(z)|_{z=1} = 1 - E(Y) = 1 - \rho \tag{3.104}$$

and

$$P(z) = \frac{(1-\rho) Y(z)(1-z)}{Y(z) - z} \tag{3.105}$$

4. evaluation of $Y(z)$:

$$Y(z) = \sum_{k=0}^{\infty} y_k z^k = \sum_{k=0}^{\infty} \int_{t=0}^{\infty} \mathbf{P}(Y = k \mid x = t) b(t) dt z^k =$$

$$\sum_{k=0}^{\infty} \int_{t=0}^{\infty} \frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t} b(t) dt z^k = \int_{t=0}^{\infty} \mathrm{e}^{\lambda(z-1)t} b(t) dt = B^*(\lambda(z-1))$$

hence

$$Y(z) = B^*(\lambda(z-1)) \tag{3.106}$$

and

$$P(z) = \frac{(1-\rho) B^*(\lambda(z-1))(1-z)}{B^*(\lambda(z-1)) - z} \tag{3.107}$$

which is known as transform Pollaczek-Hincsin formula.

5. Example: M/M/1 queue

$$b(x) = \mu e^{-\mu x} \quad B^*(s) = \frac{\mu}{s+\mu} \quad B^*(\lambda(z-1) = \frac{\mu}{\lambda z - \lambda + \mu}$$

$$P(z) = \frac{(1-\rho)\mu(1-z)}{\mu - \lambda z^2 + \lambda z - \mu z} = \frac{(1-\rho)\mu(1-z)}{(\mu - \lambda z)(1-z)} = \frac{1-\rho}{1-\rho z} \tag{3.108}$$

from which: $p_k = (1-\rho)\rho^k$

6. Example 8: Consider a multiplexer with N identical input links each of them with geometrically distributed interarrival time. There is an output buffer of size K. The output link is m times faster than the input links. Suppose the first m cells can be transmitted in the slot they arrived if the buffer was empty.

   Problems to solve:

   - the evolution equation of the system,
   - state transition probability matrix,

   (Evolution equation: $X_{n+1} = (X_n - m)^+ + Y_{n+1}$)

7. Furter related problems:

   - non-Poisson arrival, general service time (G/M/1, G/G/1)
   - queueing network (set of queues and links)
   - new chalanges:
     - multiplexing different classes of traffic (B-ISDN)
     - long term dependences
     - self-similarity
   - new models:
     - fluid model
     - fractal, kaos model