Part I – Stochastic variables and Markov chains

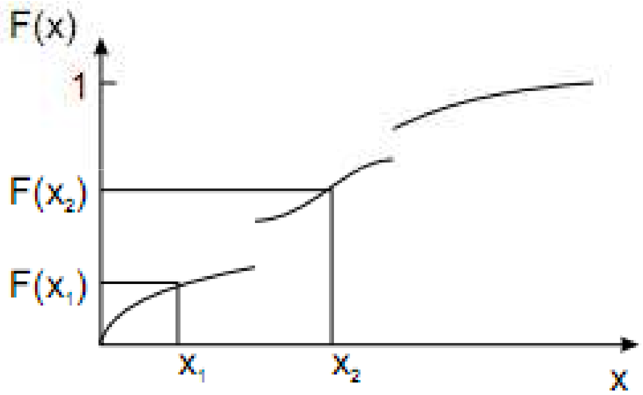
Random variables

- describe the behaviour of a phenomenon independent of any specific sample space

- **Distribution function** (cdf, cumulative distribution function)

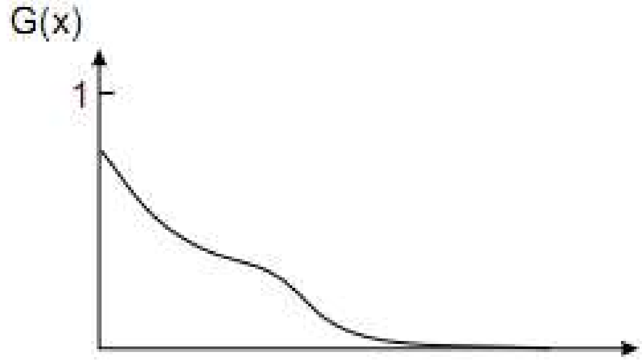
$$F(x) = \mathrm{P}\{X \leq x\}$$

$$\mathrm{P}\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$$



- **Complementary distribution function** (tail distribution)

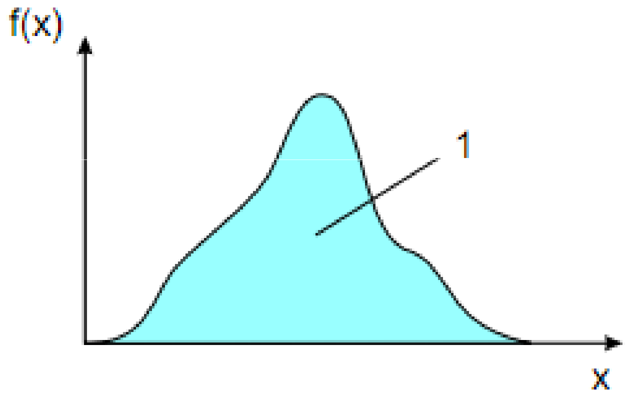
$$G(x) = 1 - F(x) = \mathrm{P}\{X > x\}$$



Continuous random variables

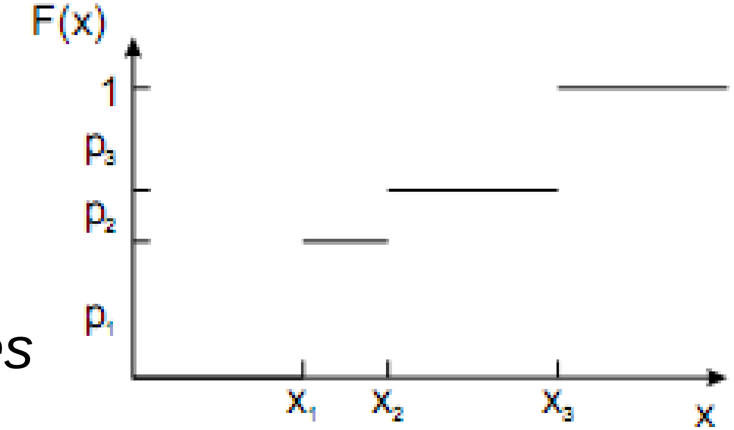
- can be described by means of the **probability density function** (pdf)

$$f(x) = \frac{dF(x)}{dx} = \lim_{dx \to 0} \frac{\mathrm{P}\{x < X \leq x + dx\}}{dx}$$



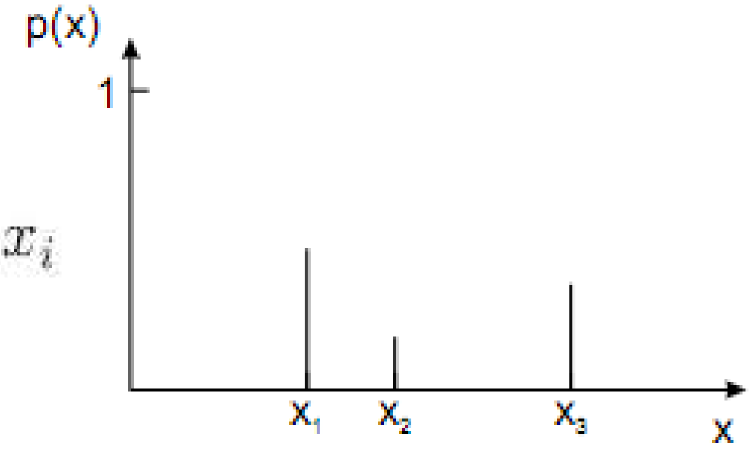
Discrete random variables

- The set of values a discrete random variable *X* can take is either *finite* or *countably infinite*,
  - $X \in \{ x_1, x_2, . . . \}$
- Associated a set of *point probabilities* with these values
  - $p_i = P\{ X = x_i \}$



- *Probability mass function* (pmf):

$$p(x) = \mathrm{P}\{X = x\} = \begin{cases} p_i & \text{when } x = x_i \\ 0, & \text{otherwise} \end{cases}$$



Independent random variables

- The random variables X and Y are independent if and only if the events {X ≤ x} and {Y ≤ y} are independent:

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$$

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

- (the two conditions are equivalent)

Parameters of distributions

- **Expectation**

denoted by $\mathrm{E}[X] = \bar{X}$

continuous distribution $\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$

discrete distribution $\mathrm{E}[X] = \sum_i x_i p_i$

in general $\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, dF(x)$

- Properties of Expectation

$\mathrm{E}[cX] = c\mathrm{E}[X]$ (c constant)

$\mathrm{E}[X_1 + \cdots X_n] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]$

$\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$ when X and Y are independent

Parameters of distributions

- **Variance**, denoted by *V[X]*, or *Var[X]*

$$\boxed{\mathrm{V}[X] = \mathrm{E}[(X - \bar{X})^2]} = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

- **Standard deviation**, denoted by σ:

$$\boxed{\sigma = \sqrt{V[X]}}$$

- **Coefficient of variation**, denoted by C_v:

$$\boxed{C_v = \frac{\sigma}{E[X]}}$$

- **Covariance**, denoted by *Cov[X,Y]*

$$\boxed{\mathrm{Cov}[X,Y] = \mathrm{E}[(X - \bar{X})(Y - \bar{Y})]} = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

- *Cov[X,X]* = *Var[X]*
- *Cov[X,Y]* = 0 if *X* and *Y* are independent

Parameters of distributions

- Properties of Variance

$$V[cX] = c^2 V[X]$$ (c constant)

$$V[X_1 + \cdots X_n] = \sum_{i,j=1}^{n} \mathrm{Cov}[X_i, X_j]$$

$$V[X_1 + \cdots X_n] = V[X_1] + \cdots + V[X_n]$$ when Xi are independent

- Properties of Covariance

$$\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$$

$$\mathrm{Cov}[X + Y, Z] = \mathrm{Cov}[X, Z] + \mathrm{Cov}[Y, Z]$$

Discrete variables: Bernoulli distribution

- X ~ Bernoulli (p)
- $p_0 = q \qquad p_1 = p = 1 - q$
- $E[X] = p$
- $V[X] = pq$

- Example: The cell stream arriving at an input port of an ATM switch
  – in a time slot (cell slot) there is a cell with probability p or the slot is empty with probability q

Discrete variables: Binomial distribution

- *X* ~ Bin (*n*,*p*)
  - the number of successes in a sequence of *n* independent Bernoulli(*p*) trials
  - Can be written as the sum of n Bernoulli variables

$$p_i = \mathrm{P}\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

$$\begin{aligned} \mathrm{E}[X] &= n\mathrm{E}[Y_i] = np \\ \mathrm{V}[X] &= n\mathrm{V}[Y_i] = np(1-p) \end{aligned}$$

Discrete variables: Geometric distribution

- *X* ~ Geom(*p*)
  - the number of trials in a sequence of independent Bernoulli(*p*) trials until the first success occurs

$$p_i = \mathrm{P}\{X = i\} = (1-p)^{i-1}p$$

$$\mathrm{E}[X] = \frac{1}{p}$$

$$\mathrm{E}[X^2] = \frac{1}{p} + \frac{2(1-p)}{p^2}$$

$$\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{1-p}{p^2}$$

Discrete variables: Geometric distribution

- probability that for the first success one needs more than n trials

$$\mathrm{P}\{X > n\} = \sum_{i=n+1}^{\infty} p_i = (1-p)^n$$

- Memoryless property:

$$\mathrm{P}\{X > i+j \,|\, X > i\} = \frac{\mathrm{P}\{X > i+j \cap X > i\}}{\mathrm{P}\{X > i\}} = \frac{\mathrm{P}\{X > i+j\}}{\mathrm{P}\{X > i\}}$$

$$= \frac{(1-p)^{i+j}}{(1-p)^i} = \mathrm{P}\{X > j\}$$

Discrete variables: Poisson distribution

- X ~ Poisson(a)

$$p_i = \mathrm{P}\{X = i\} = \frac{a^i}{i!} e^{-a} \qquad \mathrm{E}[X] = a \qquad \mathrm{V}[X] = a$$

- Poisson(λt) represents the number of occurrences of events (e.g. arrivals) in an interval of length t from a *Poisson process* (see later) with intensity λ
  - the probability of an event ('success') in a small interval dt is λdt
  - the probability of two simultaneous events is $O(\lambda dt)$
  - the number of events in disjoint intervals are independent

Properties of Poisson distribution

- The sum of two Poisson random variables $X_1 \sim Poisson(a_1)$ and $X_2 \sim Poisson(a_2)$ is Poisson distributed, and we have $X_1+X_2 \sim Poisson(a_1+a_2)$
- If the number N of elements in a set obeys Poisson distribution, $N \sim Poisson(a)$, and one makes a random independent selection with probability p, then the size of the selected set $K \sim Poisson(pa)$
- If the elements of a set with size $N \sim Poisson(a)$ are randomly assigned to one of two groups 1 and 2 with probabilities p_1 and $p_2 = 1-p_1$, then the sizes of the two sets, N_1 and N_2, are independent and distributed as $N_1 \sim Poisson(p_1a)$, $N_2 \sim Poisson(p_2a)$

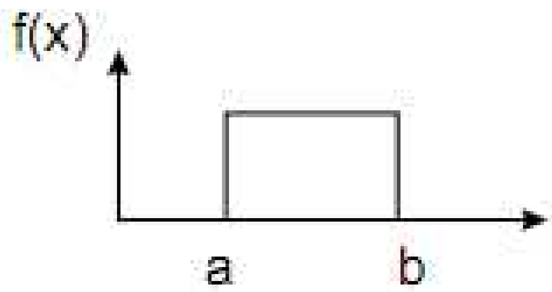
Continuous variables: Deterministic distribution

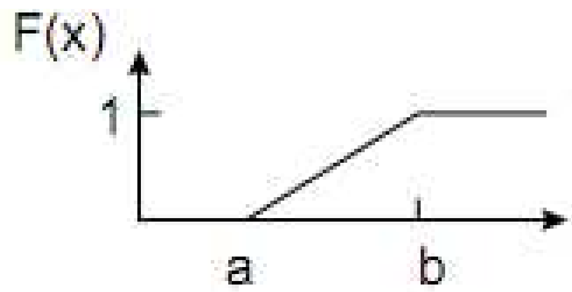
- Takes on a single value t_0
- pdf is impulsive
- Distribution is a step function

$$E[X] = t_0, \quad E[V] = 0, \quad C_v^{\ 2} = 0$$

Continuous variables: Uniform distribution

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & elsewhere \end{cases}$$





$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx = \frac{a+b}{2}$$

$$V[X] = \int_{-\infty}^{+\infty} \left(x - \frac{a+b}{2}\right)^2 f(x) dx = \frac{(b-a)^2}{12}$$

Continuous variables: Normal distribution

- $X \sim N(\mu, \sigma^2)$ $$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$ $$\begin{cases} E[X] = \mu \\ V[X] = \sigma^2 \end{cases}$$

- Property:

If $X \sim N(\mu, \sigma^2)$, then $Y = \alpha X + \beta \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$.

Proof: $$F_Y(y) = P\{Y \le y\} = P\{X \le \tfrac{y-\beta}{\alpha}\} = F_X(\tfrac{y-\beta}{\alpha})$$
$$= \int_{-\infty}^{(y-\beta)/\alpha} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx \qquad z = \alpha x + \beta$$
$$= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}(\alpha\sigma)} e^{-\frac{1}{2}(z-(\alpha\mu+\beta))^2/(\alpha\sigma)^2} dz$$

thus: $$Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \qquad (\alpha = 1/\sigma,\ \beta = -\mu/\sigma)$$

Continuous variables: Normal distribution

- X_i independent and identically distributed random variables, with means μ_i and standard deviations σ_i (μ_i, σ_i, and μ_i / σ_i finite)

  then

$$\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \rightarrow N(0,1)$$

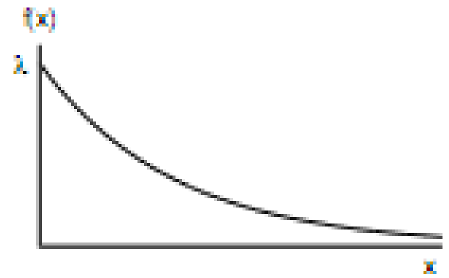
- as $n \rightarrow \infty$
- nothing assumed about X_i, but their arithmetic average tends to be normally distributed for large n
- useful in statistical evaluation of simulation results
- known as the **Central Limit Theorem**

Continuous variables: Exponential Distribution

- $X \sim Exp(\lambda)$

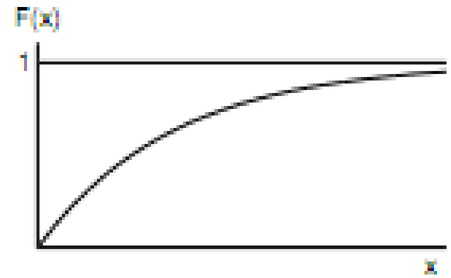
- probability density function:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$



- cumulative distribution function:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$



- mean and standard deviation:

$$\mathrm{E}[X] = \frac{1}{\lambda} \qquad \mathrm{E}[X^2] = \frac{2}{\lambda^2}$$

$$\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{1}{\lambda^2}$$

Continuous variables: Exponential Distribution

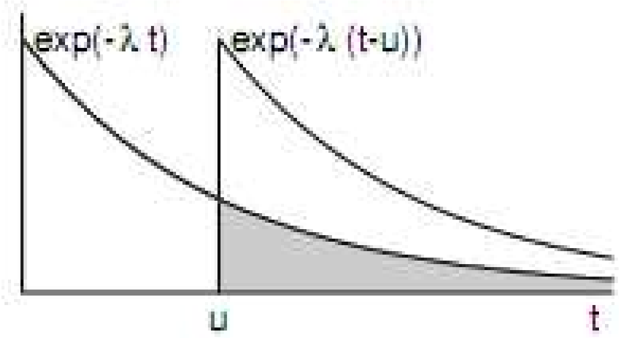
- A continuous R.V. *X* is exponentially distributed if and only if for $s, t \geq 0$

$$P\{X > s + t \mid X > t\} = P\{X > s\}$$

or equivalently

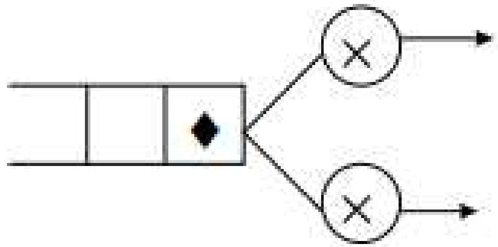
$$P\{X > s + t\} = P\{X > s\}P\{X > t\}$$

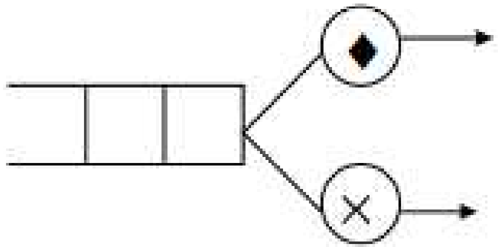
- A random variable with this property is said to be *memoryless*.



Continuous variables: Exponential Distribution

- A queueing system with two servers
- The service times are assumed to be exponentially distributed (with the same parameter)
- Upon arrival of a customer (♦), both servers are occupied (×) but there are no other waiting customers.





- What is the probability that the customer (♦) will be the last to depart from the system?

Continuous variables: Exponential Distribution

- minimum and maximum of exponentially distributed random variables $X_1 \sim X_2 \sim \ldots X_n \sim Exp(\lambda)$
- The minimum obeys the distribution $Exp(n\lambda)$, with ending intensity equal to $n\lambda$:

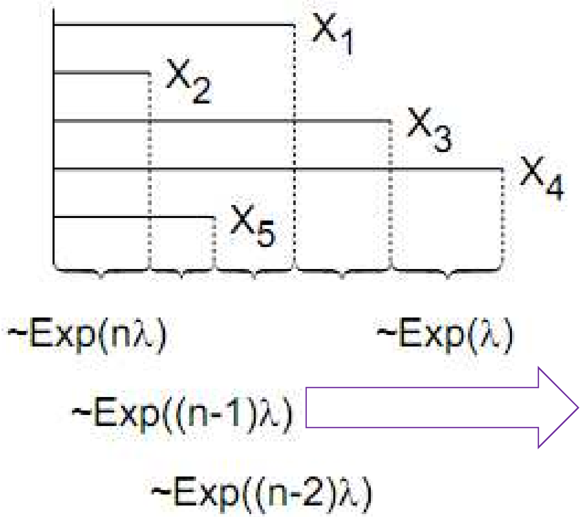
$$P\{\min(X_1, \ldots, X_n) > x\} = P\{X_1 > x\} \cdots P\{X_n > x\} = (e^{-\lambda x})^n = e^{-n\lambda x}$$

- The *cdf* of the maximum is

$$P\{\max(X_1, \ldots, X_n) \leq x\} = (1 - e^{-\lambda x})^n$$

- The expectation can be deduced by considering "*one exp at time*":

$$E[\max(X_1, \ldots, X_n)] = \frac{1}{n\lambda} + \frac{1}{(n-1)\lambda} + \cdots + \frac{1}{\lambda}$$

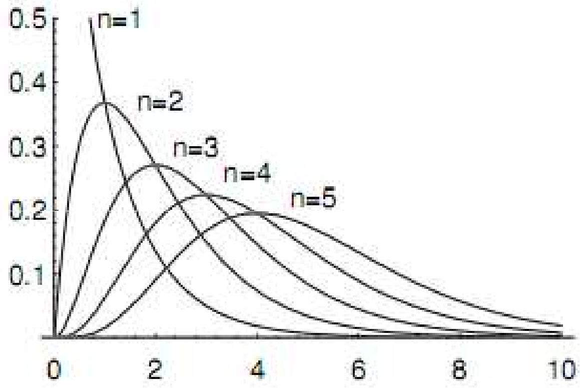


Continuous variables: Erlang Distribution

- $X \sim Erl(n,\lambda)$
- $X = X_1 + X_2 + \ldots + X_n$, where $X_i \sim Exp(\lambda)$ i.i.d.

$$f_X(t) = \begin{cases} 0, & for \quad t < 0 \\ e^{-\lambda t}\lambda^n t^{n-1}/(n-1)!, & for \quad t \geq 0 \end{cases}$$

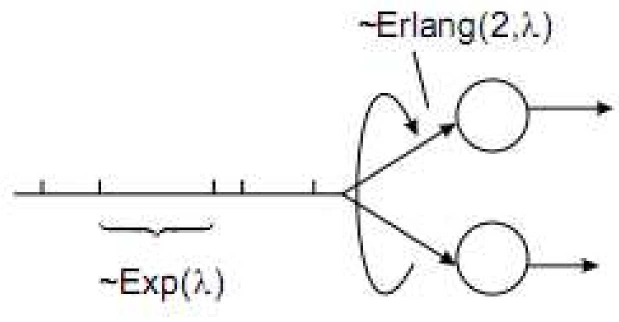
$$E[X] = \frac{n}{\lambda}, \quad V[X] = \frac{n}{\lambda^2}, \quad C_v^{\ 2} = \frac{1}{n}$$



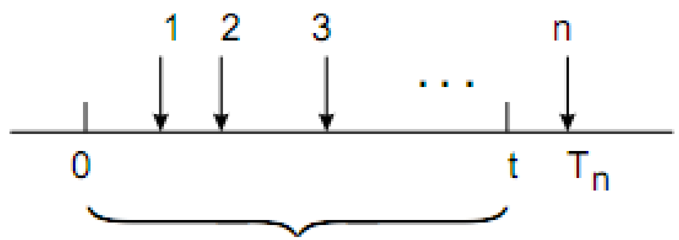
- An *Erlang-n* distribution is a series of *n* independent and identically distributed exponential distributions (*n-stage Erlang distribution*)
- n=1 >>> $Erl(n, \lambda)$ tends to be exponential
- $n \rightarrow inf.$, n/λ finite >>> $Erl(n, \lambda)$ tends to be deterministic

Erlang distribution: example

- A system consists of two servers. Customers arrive with *Exp(λ)* distributed interarrival times and are alternately sent to servers 1 and 2. The interarrival time distribution of customers arriving at a given server is ~ *Erl(2,λ)*



- Let N_t, the number of events in an interval of length t, obey the Poisson distribution: $N_t \sim Poisson(\lambda t)$. Then the time T_n from an arbitrary event to the nth event thereafter obeys the distribution *Erl(n,λ)*



Continuous variables: Gamma distribution

- *X ~ Gamma(a,λ)*
- generalizes the Erlang distribution

$$f_X(t) = \begin{cases} 0, & for \;\; t < 0 \\ e^{-\lambda t} \lambda^a t^{a-1} / G(a), & for \;\; t \geq 0 \end{cases}$$

$$\text{where} \quad G(a) = \int_0^{+\infty} s^{(a-1)} e^{-s} ds$$

$$E[X] = \frac{a}{\lambda}, \quad V[X] = \frac{a}{\lambda^2}, \quad C^2 = \frac{1}{a}$$

a: form parameter, *λ*: scale parameter

a < 1 → the distribution is more irregular

Continuous variables: Hypo-exponential distribution

- like Erlang distributions, is made of *n* successive exponential stages: $X_1 + X_2 + \ldots + X_n$
  - but here the exponential stages need not have the same mean

- With two exponential variables $X_1 \sim Exp(\lambda_1)$ and $X_2 \sim Exp(\lambda_2)$:

$$
\begin{aligned}
f_X(x) &= \frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}(e^{-\lambda_1 x} - e^{-\lambda_2 x}), \\
F_X(x) &= 1 - \frac{\lambda_2}{\lambda_2-\lambda_1}e^{-\lambda_1 x} + \frac{\lambda_1}{\lambda_2-\lambda_1}e^{-\lambda_2 x}
\end{aligned}
$$

Continuous variables: Hyper-exponential distribution

- a random choice between exponential variables
  - For example, let X_1 and X_2 be two independent exponentially distributed random variables with parameters λ_1 and λ_2
  - Let X be a variable that, with probability, p_1 is distributed as X_1 and with probability p_2 is distributed as X_2 (where $p_1+p_2 = 1$)
- X has a *2-stage hyperexponential* distribution:

$$f_X(x) = p_1\lambda_1 e^{-\lambda_1 x} + p_2\lambda_2 e^{-\lambda_2 x}$$

Continuous variables: Phase-type distributions

- *Phase-type distributions*
  – are formed by summing exponential distributions, or by probabilistically choosing among them
- Some examples include the distributions seen above:
  – Erlang distribution
  – Hypo-exponential distribution
  – Hyper-exponential distribution

Random variables in computer modelling

- In modelling of computer and communication systems
  - make assumptions about the involved job service times or the packet transmission times
  - often the exponential, Erlang, hypo- and hyper-exponential distribution are used
- Exponential distribution is especially advantageous to use
  - Memoryless property advantageous for mathematics
  - Useful in many practical cases:
  - for example, extensive monitoring of telephone exchanges has revealed that telephone call durations typically are exponentially distributed
  - the average time between successive call arrivals obeys the exponential distribution.

Stochastic processes

- Systems we consider evolve in time. We are interested in their dynamic behaviour, usually involving some randomness
  - the length of a queue
  - the temperature outside
  - the number of data packets in a network, etc.
- A stochastic process X_t,or $X(t)$, is a family of random variables indexed by a parameter t
  - Formally, it is a mapping from the sample space S to functions of t. Each element e of S is associated to $X_t(e)$, a function in time t
  - For a given value of t, X_t is a random variable
  - For a given value e, the function $X_t(e)$ is called the *realization* of the stochastic process (also trajectory or sample path).

Stochastic processes

- Stochastic processes can be either:
  - Discrete-time/Discrete-state
    - e.g., the number of jobs present in a computer system at the time of departure of the *k*th job
  - Continuous-time/Discrete-state
    - e.g., the number of jobs present in a computer system at time *t*
  - Discrete-time/Continuous-state
    - e.g., the time the *k*th job has to wait until the service starts
  - Continuous-time/Continuous-state
    - e.g., the total amount or service that needs to be done on all jobs present in the system at a given time *t*

Stochastic processes

- First-order distribution:

$$F(x,t) = P\{X(t) \le x\}$$

- *n*-th order distribution:

$$F(\underline{x},\underline{t}) = P\{X(t_1) \le x_1, \cdots, X(t_n) \le x_n\}$$

- The process is strictly stationary when

$$F(\underline{x},\underline{t}) = F(\underline{x},\underline{t}+\tau)$$

- Stationarity in wide sense
  - 1st and 2nd order statistics are translation invariant

- Independent process:

$$F(\underline{x},\underline{t}) = \prod_{i=1}^{n} F(x_i,t_i) = \prod_{i=1}^{n} P\{X(t_i) \le x_i)$$

Markov Process

- Next state taken on by a stochastic process only depends on the current state of the process
  - does not depend on states that were assumed previously
  - called first-order dependence or Markov dependence

$$P\{X(t_{n+1}) \le x_{n+1} \mid X(t_0) = x_0, \ldots, X(t_n) = x_n\} =$$
$$P\{X(t_{n+1}) \le x_{n+1} \mid X(t_n) = x_n\}$$

- Markov Process: discrete- or continuous- time

- Time-homogeneous Markov process:

$$P\{X(t) \le x \mid X(s) = x_s\} = P\{X(t-s) \le x \mid X(0) = x_s\}$$

Discrete-Time Markov Chain

- Discrete-time stochastic process $\{X_n \mid n = 0,1,2,\ldots\}$
- Takes values in $\{0,1,2,\ldots\}$
- Memoryless property:

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P\{X_{n+1} = j \mid X_n = i\}$$

$$P_{ij} = P\{X_{n+1} = j \mid X_n = i\}$$

- Transition probabilities P_{ij} $\quad P_{ij} \geq 0, \quad \sum_{j=0}^{\infty} P_{ij} = 1$
- Transition probability matrix $P=[P_{ij}]$
- It is a *stochastic matrix*:

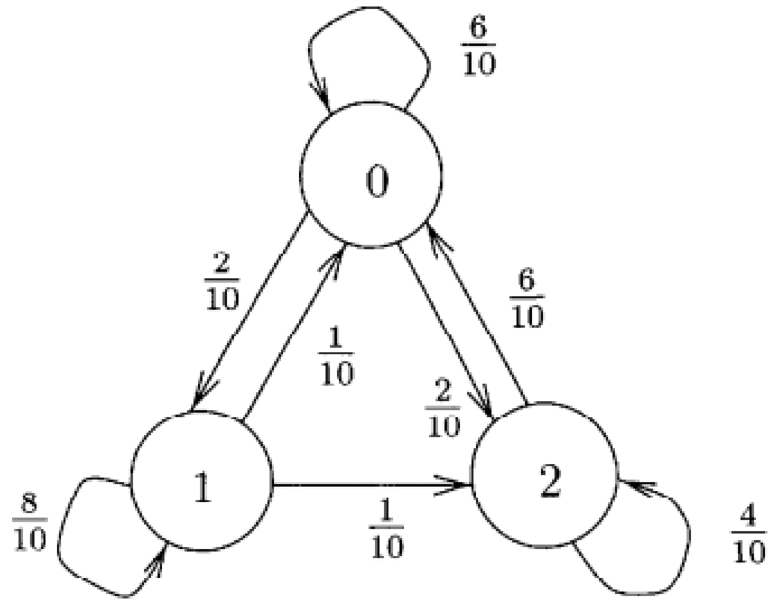
$$0 \leq P_{ij} \leq 1, \quad \sum_j P_{ij} = 1$$

$$\mathbf{P} = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & \cdots \\ p_{1,0} & p_{1,1} & p_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{matrix} \text{final state} \longrightarrow \\ \downarrow \text{initial state} \end{matrix}$$

Discrete-Time Markov Chain: an example

- Transition matrix and graphical representation
- (can both be used to represent a DTMC)

$$\mathbf{P} = \frac{1}{10}\begin{pmatrix} 6 & 2 & 2 \\ 1 & 8 & 1 \\ 6 & 0 & 4 \end{pmatrix}$$



Discrete-Time Markov Chain

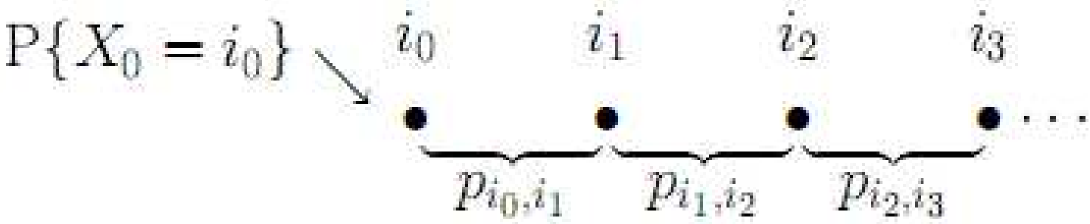
- Probability of a *path*:

$$P\{X_0 = i_0, \ldots, X_n = i_n\} = P\{X_0 = i_0\}\, p_{i_0,i_1}\, p_{i_1,i_2} \cdots p_{i_{n-1},i_n}$$

$$P\{X_0 = i_0, X_1 = i_1\} = \underbrace{P\{X_1 = i_1 \mid X_0 = i_0\}}_{p_{i_0,i_1}} P\{X_0 = i_0\}$$

$$P\{X_0 = i_0, X_1 = i_1, X_2 = i_2\} = \underbrace{P\{X_2 = i_2 \mid X_1 = i_1, X_0 = i_0\}}_{p_{i_1,i_2}} \underbrace{P\{X_1 = i_1, X_0 = i_0\}}_{p_{i_0,i_1} P\{X_0 = i_0\}}$$

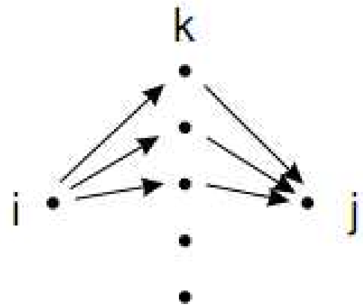
$$= P\{X_0 = i_0\}\, p_{i_0,i_1}\, p_{i_1,i_2}$$



Discrete-Time Markov Chain

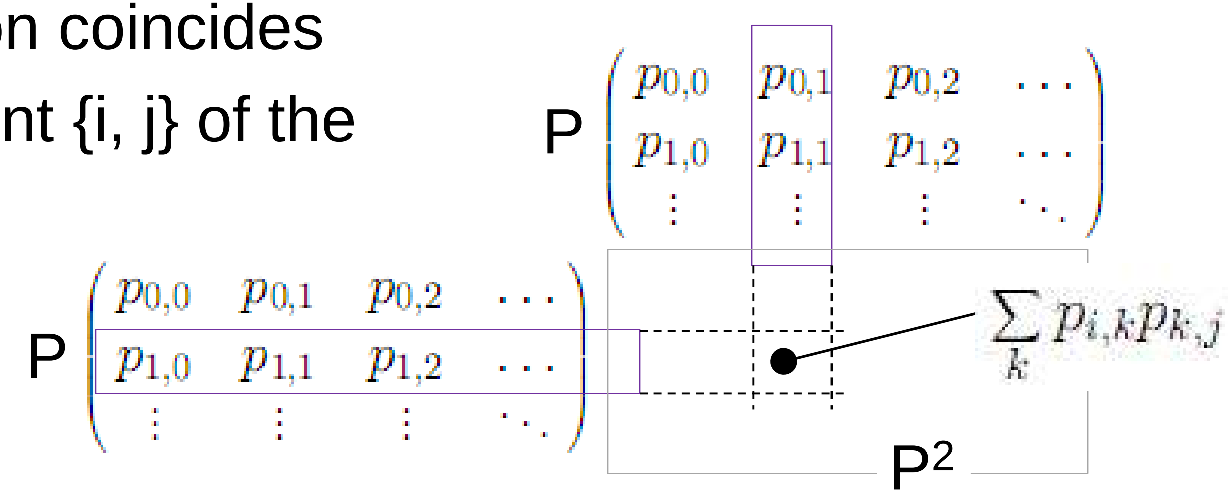
- The probability that the system, initially in state i, will be in state j after two steps is

$$\sum_k p_{i,k} p_{k,j}$$



(takes into account all paths via an intermediate state k)

- This summation coincides with the element {i, j} of the matrix P^2



State Probabilities – Stationary Distribution

- P^n describes how the probabilities change in n steps
- Define the probability distribution as

$$\pi_j^n = P\{X_n = j\}, \qquad \pi^n = (\pi_0^n, \pi_1^n, \ldots)$$

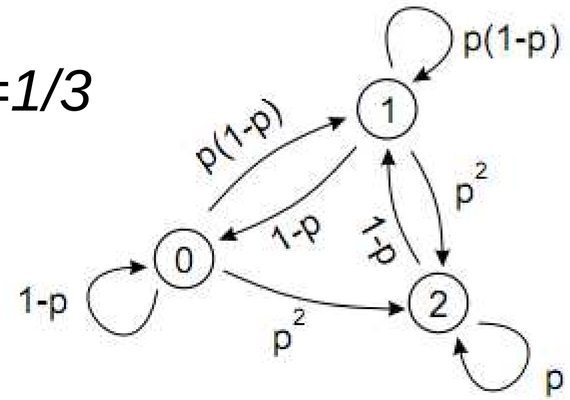
- the probability the process is in state j is $\pi^n{}_j$ for each j
- If the "state" of the process at time n is described by π^n then the new probabilities after k steps are simply calculated as $\pi^n P^k$
- If the π^n distribution converges to a limit for $n \rightarrow \infty$:

$$\pi = \lim_{n\to\infty} \pi^n \qquad \pi = \pi P$$

- π is called the *stationary*, or *equilibrium*, distribution
- Existence depends on the structure of Markov chain

State Probabilities – Stationary Distribution

- Example $\mathbf{P} = \begin{pmatrix} 1-p & p(1-p) & p^2 \\ 1-p & p(1-p) & p^2 \\ 0 & 1-p & p \end{pmatrix}$ $p=1/3$



$$\mathbf{P} = \frac{1}{9}\begin{pmatrix} 6 & 2 & 1 \\ 6 & 2 & 1 \\ 0 & 6 & 3 \end{pmatrix} = \begin{pmatrix} 0.6666 & 0.2222 & 0.1111 \\ 0.6666 & 0.2222 & 0.1111 \\ 0 & 0.6666 & 0.3333 \end{pmatrix}$$

$$\mathbf{P}^2 = \frac{1}{9^2}\begin{pmatrix} 48 & 22 & 11 \\ 48 & 22 & 11 \\ 36 & 30 & 15 \end{pmatrix} = \begin{pmatrix} 0.5926 & 0.2716 & 0.1358 \\ 0.5926 & 0.2716 & 0.1358 \\ 0.4444 & 0.3704 & 0.1852 \end{pmatrix}$$

$$\mathbf{P}^3 = \frac{1}{9^3}\begin{pmatrix} 420 & 206 & 103 \\ 420 & 206 & 103 \\ 396 & 222 & 111 \end{pmatrix} = \begin{pmatrix} 0.5761 & 0.2826 & 0.1413 \\ 0.5761 & 0.2826 & 0.1413 \\ 0.5432 & 0.3045 & 0.1523 \end{pmatrix}$$

$$\mathbf{P}^8 = \begin{pmatrix} 0.5714 & 0.2857 & 0.1429 \\ 0.5714 & 0.2857 & 0.1429 \\ 0.5714 & 0.2857 & 0.1429 \end{pmatrix}$$

Kolmogorov's theorem

- In an *irreducible*, *aperiodic* Markov chain there always exists the stationary, or equilibrium, distribution:

$$\pi_j = \lim_{n\to\infty} \pi_j^n$$

and this is *independent of the initial state*.

- If the limit probabilities (the components of the vector) π exist, they must satisfy the equation $\pi = \pi$ P, because

$$\pi = \lim_{n\to\infty} \pi^n = \lim_{n\to\infty} \pi^{n+1} = \lim_{n\to\infty} \pi^n P = \pi P$$

- π defines which proportion of time (steps) the system stays in state j.
  - *equilibrium* does not mean that nothing happens in the system, but merely that the information on the initial state of the system has been "forgot"

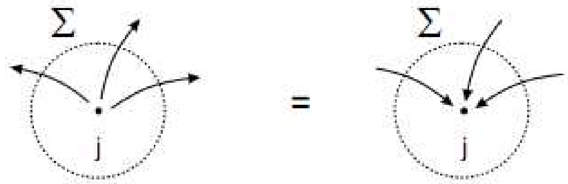
Global Balance Equations

- Markov chain with infinite number of states
- Global Balance Equations (GBE)

$$\pi_j = \pi_j \sum_{i=0}^{\infty} P_{ji} = \sum_{i=0}^{\infty} \pi_i P_{ij} \;\Leftrightarrow\; \pi_j \sum_{i \neq j} P_{ji} = \sum_{i \neq j} \pi_i P_{ij}, \quad j \geq 0$$

- $\pi_j P_{ji}$ is the frequency of transitions from j to i

$$\left(\begin{array}{c}\text{Frequency of} \\ \text{transitions out of } j\end{array}\right) = \left(\begin{array}{c}\text{Frequency of} \\ \text{transitions into } j\end{array}\right)$$



- Intuition: j visited infinitely often; for each transition out of j there must be a subsequent transition into j with probability 1

Solving the balance equations

- Assume here a finite state space with n states
- Write the balance condition for all but one of the states ($n - 1$ equations)
  - the equations fix the relative values of the equilibrium probabilities
  - the solution is determined up to a constant factor
- The last balance equation (automatically satisfied) is replaced by the normalization condition $\sum \pi_j = 1$

Solving the balance equations: example

$$\begin{pmatrix}\pi_0 & \pi_1 & \pi_2\end{pmatrix} = \begin{pmatrix}\pi_0 & \pi_1 & \pi_2\end{pmatrix}\begin{pmatrix} 1-p & p(1-p) & p^2 \\ 1-p & p(1-p) & p^2 \\ 0 & 1-p & p \end{pmatrix}$$

$$\begin{aligned}
\pi_0 &= (1-p)\pi_0 + (1-p)\pi_1 && \Rightarrow \quad \pi_0 = \tfrac{1-p}{p}\pi_1 \\
\pi_1 &= p(1-p)\pi_0 + p(1-p)\pi_1 + (1-p)\pi_2 \\
&= (1-p)^2\pi_1 + p(1-p)\pi_1 + (1-p)\pi_2 \\
&= (1-p)\pi_1 + (1-p)\pi_2 && \Rightarrow \quad \pi_1 = \tfrac{1-p}{p}\pi_2 \quad \Rightarrow \quad \pi_0 = \left(\tfrac{1-p}{p}\right)^2 \pi_2
\end{aligned}$$

or

$$\boldsymbol{\pi} = \begin{pmatrix}\left(\frac{1-p}{p}\right)^2 & \frac{1-p}{p} & 1\end{pmatrix}\pi_2 \,.$$

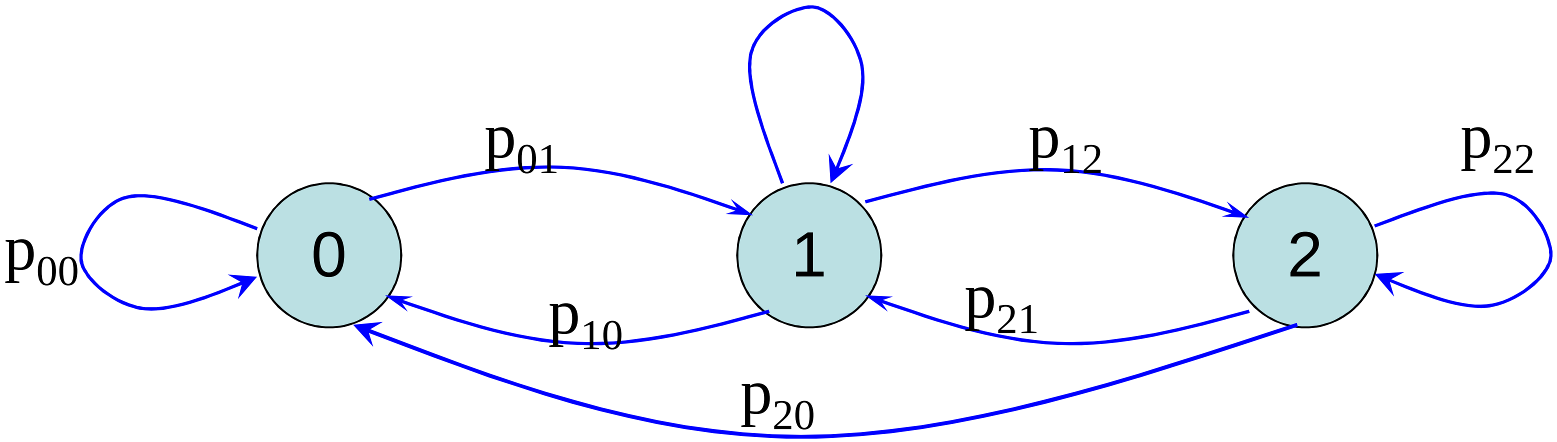
- by the normalization $\sum \pi_j = \pi_0 + \pi_1 + \pi_2 = 1$

$$\boldsymbol{\pi} = \begin{pmatrix}\frac{(1-p)^2}{1-p(1-p)} & \frac{p(1-p)}{1-p(1-p)} & \frac{p^2}{1-p(1-p)}\end{pmatrix} \qquad \text{With } p = \tfrac{1}{3}: \qquad \boldsymbol{\pi} = \begin{pmatrix}0.5714 & 0.2857 & 0.1429\end{pmatrix}$$

Discrete-Time Markov Chain: an example

- Consider a two-processor computer system where time is divided into *time slots*, operating as follows:
  - at most one job can arrive during any time slot
  - an arrival in any slot can always happen with probability $\alpha = 0.5$
  - jobs are served by whichever processor is available
  - if both are available then the job is given to processor 1
  - if both processors are busy, then the job is lost
  - when a processor is busy, it can complete the job with probability $\beta = 0.7$ during any one time slot
  - if a job is submitted during a slot when both processors are busy but at least one processor completes a job, then the job is accepted (departures occur before arrivals)
- What is the long-term *system efficiency*?
- What is the fraction of lost jobs?

Discrete-Time Markov Chain: an example

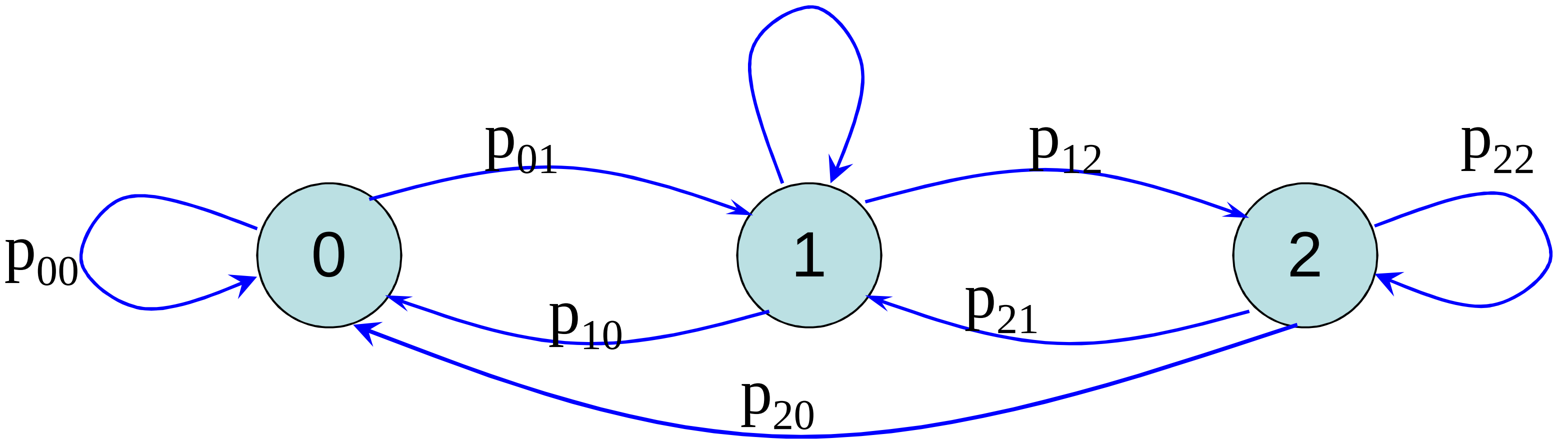


$$p_{00} = (1-\alpha) \qquad p_{01} = \alpha \qquad p_{02} = 0$$

$$p_{10} = \beta(1-\alpha) \qquad p_{11} = (1-\beta)(1-\alpha) + \alpha\beta \qquad p_{12} = \alpha(1-\beta)$$

$$p_{20} = \beta^2(1-\alpha) \qquad p_{21} = \beta^2\alpha + 2\beta(1-\beta)(1-\alpha) \qquad p_{22} = (1-\beta)^2 + 2\alpha\beta(1-\beta)$$

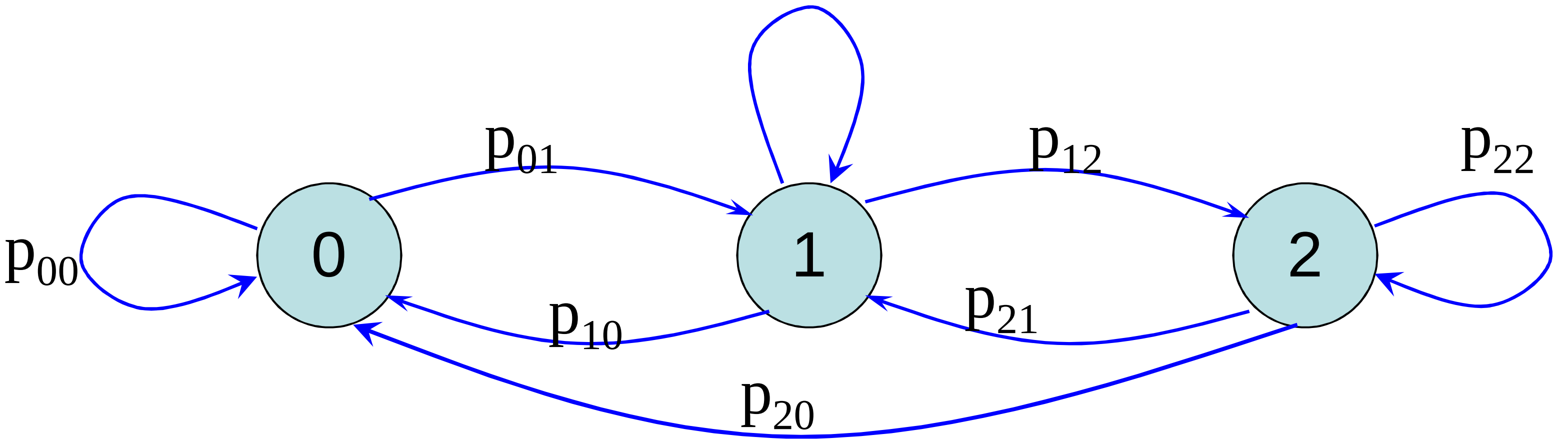
Discrete-Time Markov Chain: an example



with α = 0.5
and β = 0.7 :

$$\mathbf{P} = \left[p_{ij} \right] = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{bmatrix}$$

Discrete-Time Markov Chain: an example



$$\mathbf{P} = \left[p_{ij} \right] = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{bmatrix}$$

$$\pi = \pi P, \quad \sum \pi_j = 1$$

yields:

$$\pi = [\, 0.0278,\ 0.3094,\ \ 0.6628 \,]$$

Discrete-Time Markov Chain: an example

- What is the long-term *system efficiency*?

- What is the fraction of lost jobs?

Continuous-time Markov chain (CTMC)

- Index t is a real time
- By definition of Markov property, the probability distribution of residence time in state n *does* depend on state n, but *does not* depend on the time spent in that state and in the previous ones
  - the only memoryless continuous distribution is Exp()
  - associate a rate μ_i with each state: *out-transition rate* ~Exp(μ_i)
  - as in DTMCs, use a transition matrix Q to describe the state transition process behaviour
- CTMCs can be built from DTMCs by associating a state residence time distribution with each state

Continuous-time Markov chain (CTMC)

- $p_{i,j}$: as in DTMCs, the probability that, being in state i, the state after the next transition will be j
- $p_{i,i}=0$: a transition always changes the process state
- Matrix $[p_{i,j}]$ defines the *embedded* DTMC in the CTMC
- Define a *generator* matrix $Q = [\ q_{i,j}\ ]$
  - with $q_{i,j} = \mu_i\, p_{i,j}$ when $i \neq j$ and $q_{i,i} = -\sum_{j \neq i} q_{i,j} = -\mu_i$
  - the implication of defining $q_{i,i}$ as above will be clear soon
- For each $j \neq i$ there is an $F_{i \to j}(t) = 1 - e^{-q_{i,j}\, t}$
  - these distributions model the delay perceived in state i when going from i to j
  - the residence time is the minimum of n Exp: still exponentially distributed, with rate $\sum_{j \neq i} q_{i,j}$

Continuous-time Markov chain (CTMC)

- Matrix Q contains information about the transition probabilities of the embedded DTMC:
- given n random variables X_i~$Exp(\lambda_k)$ with rate λ_k, X_j is the minimum of them with probability $\lambda_j / \sum \lambda_i$
- so, being in state i, the shortest delay (i.e. the actual next transition) will lead to state j with probability

$$\frac{q_{i,j}}{\sum_{k \neq i} q_{i,k}} = \frac{p_{i,j}\mu_i}{\mu_i} = p_{i,j}$$

Continuous-time Markov chain (CTMC)

- Generator matrix $Q=[q_{ij}]$ also describes the infinitesimal transitions as follows:
  - $P\{ X(t+\Delta t)=j \mid X(t)=i \} = q_{ij}\, \Delta t + o(\Delta t), \quad i \neq j$
  - follows from the fact that state residence time are negative exponentially distributed
  - q_{ij} is the rate at which state i change to state j

- Denote $p_i(t) = P\{ X(t)=i \}$. Then

$$
\begin{aligned}
p_i(t+h) &= p_i(t)\,\Pr\{\text{do not depart from } i\} + \sum_{j \neq i} p_j(t)\,\Pr\{\text{go from } j \text{ to } i\} \\
&= p_i(t)\left(1 - \sum_{j \neq i} q_{i,j} h\right) + \left(\sum_{j \neq i} p_j(t) q_{j,i}\right) h + o(h).
\end{aligned}
$$

Continuous-time Markov chain (CTMC)

- using $q_{i,i} = -\sum_{j \neq i} q_{i,j}$:

$$p_i(t+h) = p_i(t) + \left(\sum_{j \in \mathcal{I}} q_{j,i} p_j(t) \right) h + o(h)$$

- rearranging the terms, dividing by h and taking the limit for $h \to 0$

$$p_i'(t) = \lim_{h \to 0} \frac{p_i(t+h) - p_i(t)}{h} = \sum_{j \in \mathcal{I}} q_{j,i} p_j(t)$$

- expressing the *transient behaviour*. In matrix notation:

$$\underline{p}'(t) = \underline{p}(t)\mathbf{Q},$$

- Using a Taylor series expansion:

$$\underline{p}(t) = \underline{p}(0) e^{\mathbf{Q}t} = \underline{p}(0) \left(\sum_{i=0}^{\infty} \frac{(\mathbf{Q}t)^i}{i!} \right)$$

Continuous-time Markov chain (CTMC)

- Steady-state probabilities: $p_i = lim_{t \to \infty} p_i(t)$
- That would mean $lim_{t \to \infty} p_i'(t) = 0$

  $\to$ p_i can be obtained by solving the following system of linear equations:

$$pQ=0 \quad \sum p_i=1$$

- Very similar to $\underline{v}=\underline{v}P$ for DTMCs, since it yields $\underline{v}(P\text{-}I)=\underline{0}$, so the matrix $P\text{-}I$ can be interpreted as a generator matrix in the discrete case

Continuous-time Markov chain (CTMC)

- It is also possible to calculate the p_i via the associated DTMC (the embedded DTMC)
- Build the discrete transition matrix from the $q_{i,j}$:

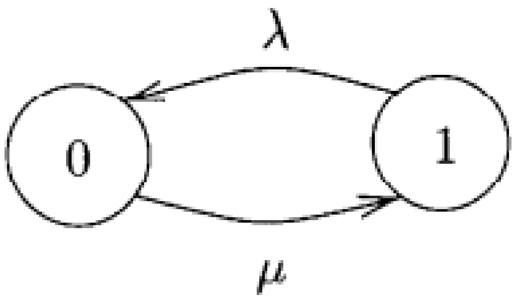
$$p_{i,j} = q_{i,j} / |q_{i,i}| \ (i \neq j) \quad p_{i,i} = 0$$

- Solve $\underline{v}=\underline{v}P$: the solutions v_i represent the probability that state i is visited, irrespective of the length of stay in this state
- Renormalize the probabilities according to the expected state residence time $1/q_i$ for every state i

$$p_i = \frac{v_i / q_i}{\sum_j v_j / q_j} \qquad \text{for all } i$$

CTMC: example

- Example
  - a resource (e.g. a processor) can be either available or busy
  - a computer system that can be either operational or not
  - time the system is available (operational) exponentially distributed ~Exp(λ) (i.e. average time is $1/\lambda$)
  - time busy (not operational) ~Exp(μ)



CTMC: example

- CTMC with generator matrix:

$$\mathbf{Q} = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}$$

- Assume initial propabilities *p(0)=(0,1)*
- steady-state probability vector

$$\underline{p}Q = \underline{0}, \quad \sum_{i \in S} p_i = 1 \quad \Rightarrow \quad \underline{p} = \left(\frac{\lambda}{\lambda + \mu}, \frac{\mu}{\lambda + \mu} \right)$$

CTMC: example

- Probability vector can also be computed via the embedded DTMC

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Rightarrow \underline{v} = \left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

- this indicated that both states are visited equally often
- Incorporating the average state residence times ($1/\lambda$ and $1/\mu$, respectively):

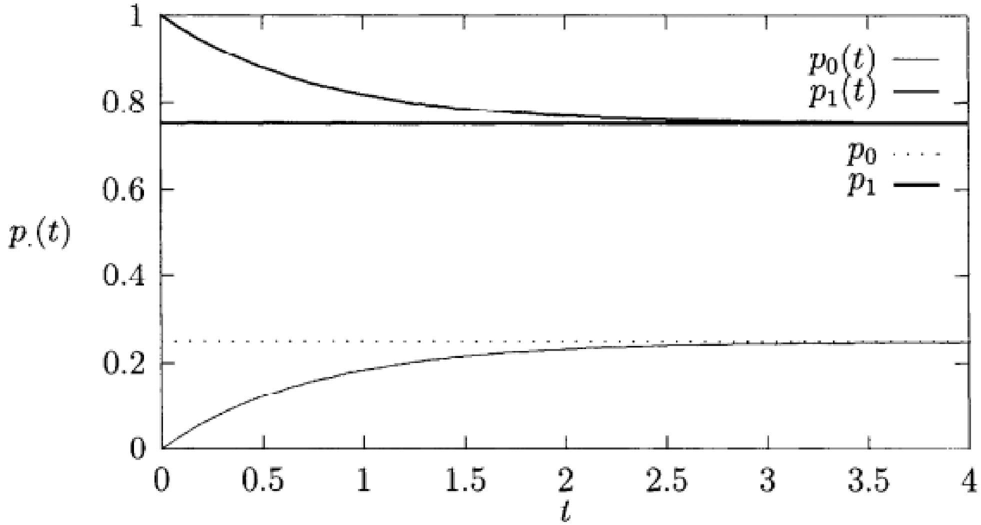
$$\underline{p} = \left(\frac{\frac{1}{2}\frac{1}{\mu}}{\frac{1}{2}\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)}, \frac{\frac{1}{2}\frac{1}{\lambda}}{\frac{1}{2}\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)} \right) = \left(\frac{\lambda}{\lambda+\mu}, \frac{\mu}{\lambda+\mu} \right)$$

CTMC: example

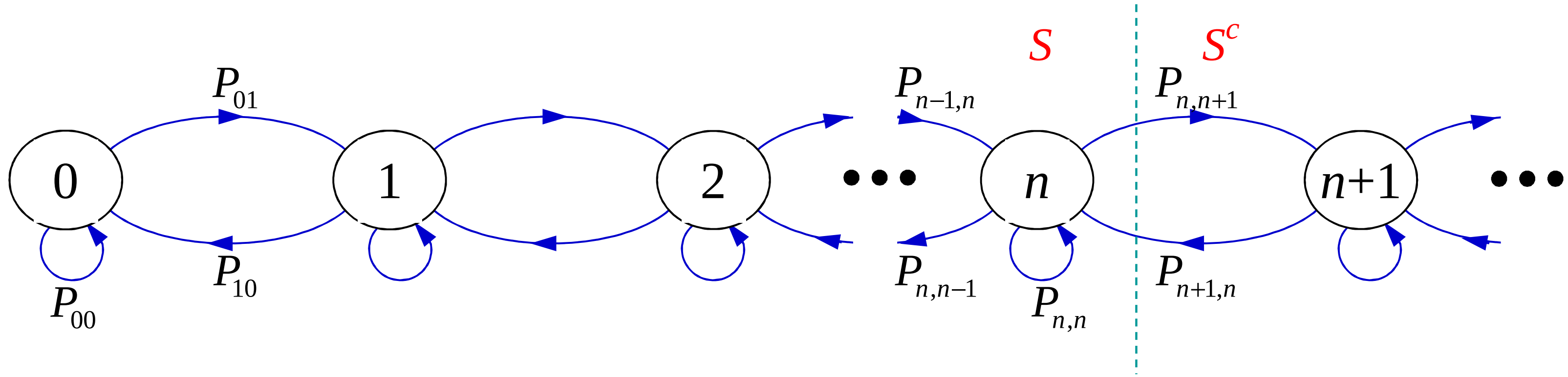
- Transient behaviour of the CTMC
- Can be solved explicitly from $\underline{p}(t)=\underline{p}(0)\,e^{\mathbf{Q}t}$:

$$p_0(t) = \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t},$$
$$p_1(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}.$$

- graphical representation of $p_i(t)$ for $3\lambda = \mu = 1$



Discrete time Birth-Death Process

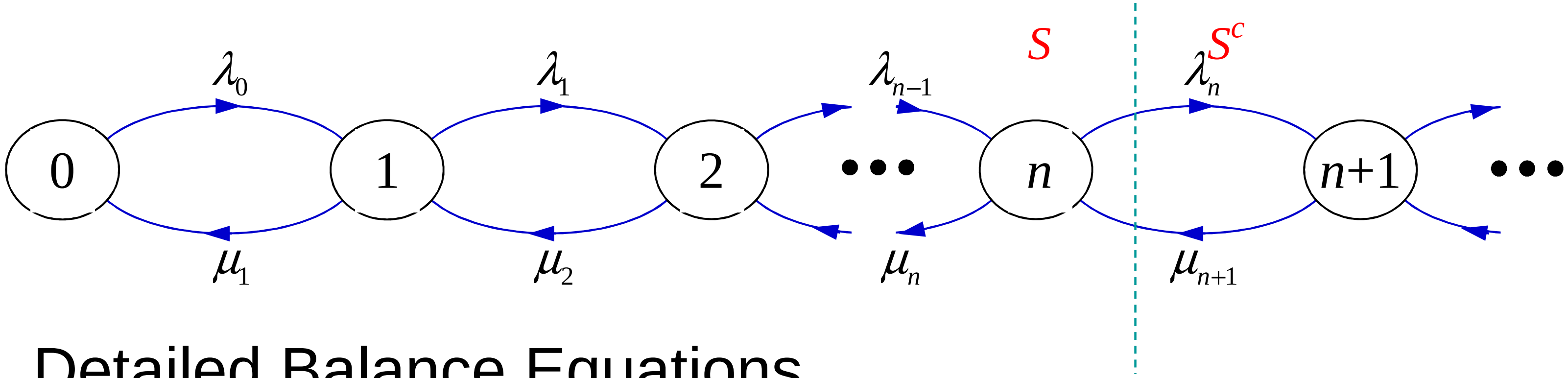


- One-dimensional DTMC with transitions only between neighboring states: $P_{ij}=0$, if $|i\text{-}j|>1$
- Detailed Balance Equations (DBE)

$$\pi_n P_{n,n+1} = \pi_{n+1} P_{n+1,n} \quad n = 0,1,\ldots$$

Continuous time Birth-Death Process

- One-dimensional CTMC:



- Detailed Balance Equations

$$q_{i,i+1} = \lambda_i, \quad q_{i,i-1} = \mu_i, \quad q_{ij} = 0, \ |i - j| > 1$$

$$\lambda_n p_n = \mu_{n+1} p_{n+1}, \quad n = 0,1,\ldots$$

Continuous time Birth-Death Process

- Use DBE to determine state probabilities as a function of p_0

$$\mu_n p_n = \lambda_{n-1} p_{n-1} \Rightarrow$$

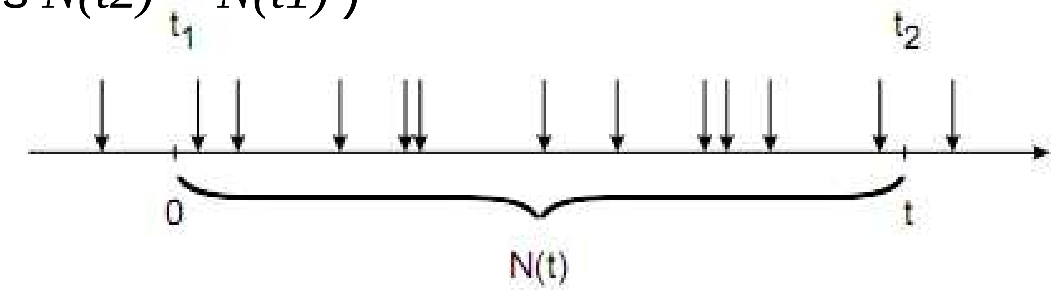
$$p_n = \frac{\lambda_{n-1}}{\mu_n} p_{n-1} = \frac{\lambda_{n-1}}{\mu_n}\frac{\lambda_{n-2}}{\mu_{n-1}} p_{n-2} = \ldots = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1} p_0 = p_0 \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}$$

- Use the probability conservation law to find p_0

$$\sum_{n=0}^{\infty} p_n = 1 \Leftrightarrow p_0\left[1+\sum_{n=1}^{\infty}\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}\right] = 1 \Leftrightarrow p_0 = \left[1+\sum_{n=1}^{\infty}\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}\right]^{-1}, \text{ if } \sum_{n=1}^{\infty}\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}} < \infty$$

Poisson Process

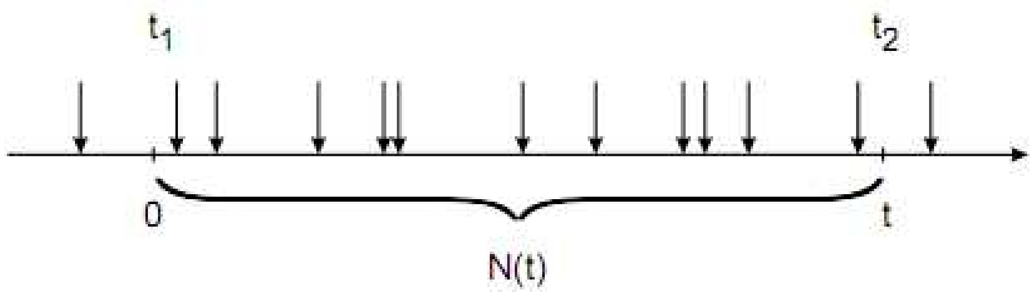
- One of the most important models used in queueing theory
- often suitable when arrivals originate from a large population of <u>independent</u> users
- $N(t)$: tells the number of arrivals that have occurred in the interval $(0, t)$, or more generally:
  - $N(t)$: number of arrivals in the interval $(0, t)$ (the stochastic process we consider)
  - $N(t1, t2)$: number of arrival in the interval $(t1, t2)$ (the increment process $N(t2) - N(t1)$)



Poisson Process: definition

- A Poisson process can be characterized in different ways:
  - Process of independent increments
  - Pure birth process:

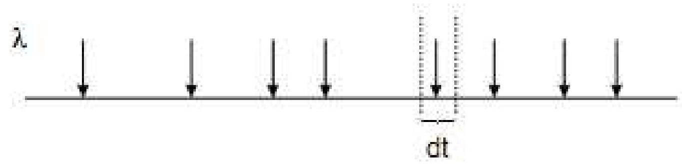
    the arrival intensity λ (mean arrival rate; probability of arrival per time unit
  - The "most random" process with a given intensity λ



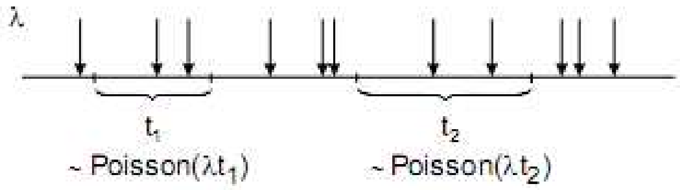
Poisson Process: definition

- Three equivalent definitions:

1) Poisson process is a pure **birth process**: In an infinitesimal time interval dt there may occur only one arrival. This happens with the probability λdt independent of arrivals outside the interval

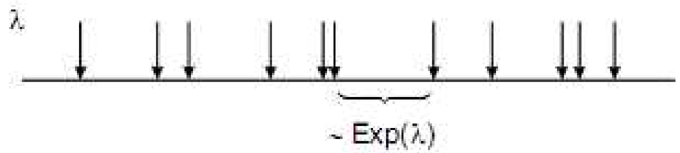


2) The **number of arrivals** $N(t)$ in a finite interval of length t obeys the Poisson(λt) distribution, $P\{N(t) = n\} = e^{-\lambda t} (\lambda t)^n / n!$ Moreover, the number of arrivals $N(t1, t2)$ and $N(t3, t4)$ in non-overlapping intervals ($t1 \leq t2 \leq t3 \leq t4$) are independent



3) The **interarrival times** are independent and obey the Exp(λ) distribution:

$P\{ \text{interarrival time} > t\} = e^{-\lambda}$

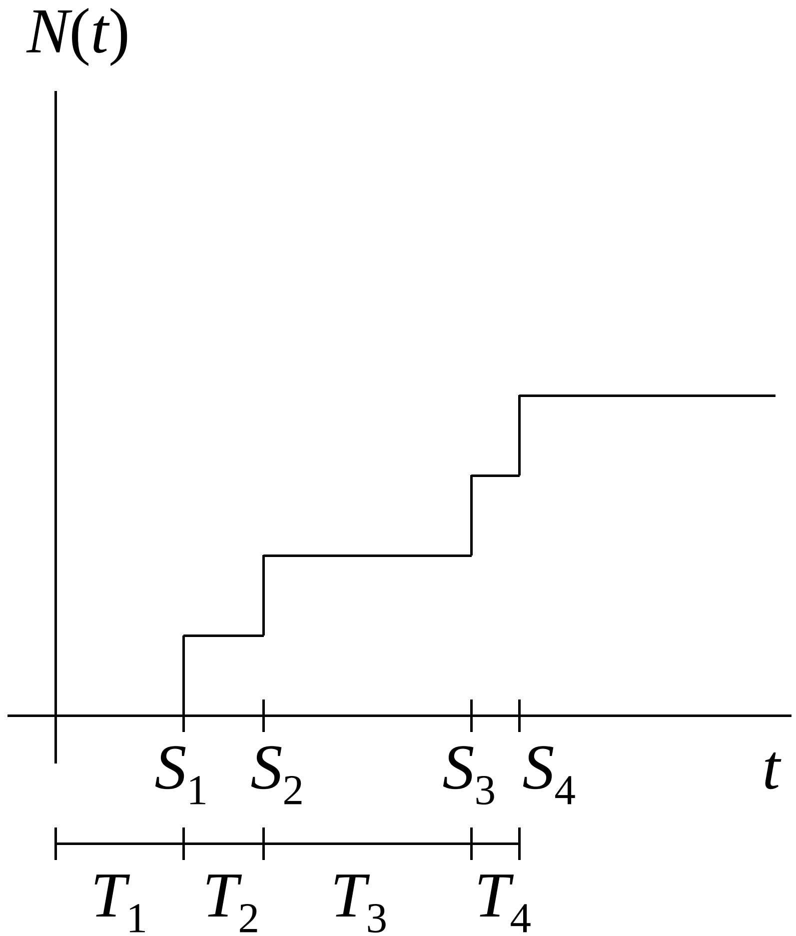


Interarrival and Waiting Times

- The times between arrivals $T_1, T_2, \ldots$ are independent exponential random variables with mean $1/\lambda$:

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

- The total waiting time until the *n*th event has a *Gamma* distribution



Poisson Process: properties

- Conditioning on the number of arrivals: Given that in the interval $(0, t)$ the number of arrivals is $N(t) = n$, these n arrivals are independently and uniformly distributed in the interval.
- Superposition: The superposition of two Poisson processes with intensities λ_1 and λ_2 is a Poisson process with intensity $\lambda = \lambda_1 + \lambda_2$.
- Random selection: If a random selection is made from a Poisson process with intensity λ such that each arrival is selected with probability p, independently of the others, the resulting process is a Poisson process with intensity $p\lambda$.

Poisson Process: properties

- Random split: If a Poisson process with intensity λ is randomly split into two subprocesses with probabilities p_1 and p_2, where $p_1 + p_2 = 1$, then the resulting processes are independent Poisson processes with intensities $p_1\lambda$ and $p_2\lambda$.
- PASTA: Poisson Arrivals See Time Averages. For instance, customers with Poisson arrivals see the system as if they came into the system at a random instant of time
  - despite they induce the evolution of the system

PASTA property

- Proof
  - consider a system with a number of users $X(t)$
    consider the event
    e = " there was at least one arrival in $(t\text{-}h, t]$ "
  - homogeneous Poisson process → $P\{e\} = P\{N(h)\geq 1\}$
    (*for a non Poisson process this wouldn't be true*!)
  - memoryless: $P\{\ N(h)\geq 1 \mid X_{t-h} = i\ \} = P\{\ N(h)\geq 1\ \}$, thus
    $P\{\ N(h)\geq 1 \cap X_{t-h} = i\ \} \ = \ P\{\ N(h)\geq 1\ \}\, P\{\ X_{t-h} = i\ \}$
  - Hence: $P\{\ X_{t-h} = i \mid N(h)\geq 1\ \} \ = \ P\{\ X_{t-h} = i\ \}$
    take the limit for $h \rightarrow 0$

PASTA property

- arrivals act as random observers (also called Random Observer Property, ROP)
  - follows from the memoryless property of the exponential distribution
  - the remaining time to the next arrival has the same exponential distribution irrespective of the time that has already elapsed since the previous arrival