Applied Probability Formulary

Ca' Foscari University of Venice
Department of Environmental Sciences, Informatics and Statistics

[CM0546] APPLIED PROBABILITY FOR COMPUTER SCIENCE
Academic Year 2022 - 2023

Student Nicola Aggio 880008

Contents

Chapter 1

Elementary Probability

- **De morgan laws**:
  - $(A \cap B)^c = A^c \cup B^c$
  - $(A \cup B)^c = A^c \cap B^c$
- **Disjoint events** : $A \cap B = \emptyset$
- **Exhaustive events** : $A \cup B = \Omega$
- **Partition**:
  - $\bigcap_{i=1}^{n} A_i = \emptyset$
  - $\bigcup_{i=1}^{n} A_i = \Omega$
  - collection of exhaustive and pairwise disjoint events
  - any event A can be expressed as: $A = \bigcup_{i=1}^{n} (A \cap B_i)$
- **Independence**: A and B are independent $\Leftrightarrow$P[A $\cap B$] $= P[A]P[B]$
- **Union**:
  - $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
  - $P[A \cup B] = P[A] + P[B]$ if A and B are disjoint
- **Conditional probability**:
  - $P[A|B] = \frac{P[A \cap B]}{P[B]}$
  - $P[A|B] = P[A]$ or $P[B|A] = P[B]$ if A and B are independent
  - **NOTE**: $P[\bar{A}|B] = 1 - P[A|B]$
- **Law of total probability**: $P[A] = \sum_{i=1}^{k} P[A|E_i]P[E_i]$
- **Bayes theorem**: $P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B]} = \frac{P[B|A]P[A]}{\sum_{i=1}^{k} P[B|C_i]P[C_i]}$

- **System reliability**:
  - *Sequential components:* works if **all** sequential components work $\rightarrow \cap$
  - *Parallel components:* it works if **at least** one of the components work $\rightarrow \cup$
  - **NOTE**:
    - if the events are **disjoint**, write things in terms of **union**, for example:

$$A \cap B = \overline{(\bar{A} \cup \bar{B})}$$

    - if the events are **independent**, write things in terms of **intersection**, for example:

$$A \cup B = \overline{(\bar{A} \cap \bar{B})}$$

- **Combinations**: possible selections of k indistinguishable objects from a set of n:
  - **combinations without replacement**: $C(n,k) = \binom{n}{k}$ = choose(n,k)
  - **combinations with replacement**: $C_r(n,k) = \binom{k+n-1}{k}$ = choose(k+n-1, k)
- **Permutations**: possible selections of k distinguishable objects from a set of n:
  - **permutations without replacement**: $P(n,k) = \frac{n!}{(n-k)!}$
  - **permutations with replacement**: $P_r(n,k) = \frac{n!}{k!(n-k)!}$

Chapter 2

Random Variables

2.1 Pmf versus pdf

Distribution	Discrete	Continuous
Definition	$P(x) = P\{X = x\}$ (pmf)	$f(x) = F'(x)$ (pdf)
Computing probabilities	$\boldsymbol{P}\{X \in A\} = \sum_{x \in A} P(x)$	$\boldsymbol{P}\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = \boldsymbol{P}\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$
Total probability	$\sum_x P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

2.1.1 Properties of pmf

1. $0 \leq P_x(x) \leq 1, \forall x$
2. $\sum_{x \in R} P_x(x) = 1$

2.1.2 Properties of pdf

1. $f_x(x) \geq 0, \forall x \in R$
2. $\int_R f(x)dx = 1$

2.1.3 Properties of cdf

1. $F(x) \geq 0, \forall x \in R$
2. $\lim_{x \to -\infty} F_x(x) = 0$
3. $\lim_{x \to \infty} F_x(x) = 1$
4. $F_x(x)$ is a non decreasing function

2.2 Joint and marginal distribution in discrete and continuous case

Distribution	Discrete	Continuous
Marginal distributions	$P(x) = \sum_y P(x,y)$ $P(y) = \sum_x P(x,y)$	$f(x) = \int f(x,y)dy$ $f(y) = \int f(x,y)dx$
Independence	$P(x,y) = P(x)P(y)$	$f(x,y) = f(x)f(y)$
Computing probabilities	$\boldsymbol{P}\{(X,Y) \in A\}$ $= \sum\sum_{(x,y)\in A} P(x,y)$	$\boldsymbol{P}\{(X,Y) \in A\}$ $= \iint_{(x,y)\in A} f(x,y)\,dx\,dy$

2.3 Moments for discrete and continuous distribution

Discrete	Continuous
$\mathbf{E}(X) = \sum_x xP(x)$	$\mathbf{E}(X) = \int xf(x)dx$
$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \int (x-\mu)^2 f(x)dx$ $= \int x^2 f(x)dx - \mu^2$
$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P(x,y)$ $= \sum_x \sum_y (xy)P(x,y) - \mu_x\mu_y$	$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \iint (x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy$ $= \iint (xy)f(x,y)\,dx\,dy - \mu_x\mu_y$

2.3.1 Properties of the expected value

- $\mathbb{E}[aX + bY + c] = a\mathbb{E} + b\mathbb{E} + c$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[X] = a\mathbb{E}[X]$
- **For independent** X **and** Y $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$

- $X_1, X_2, ..., X_n$ RVs, then $\mathbb{E}[X_1 + X_2 + ... + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + ... + \mathbb{E}[X_n] = \sum_{i=1}^{n} X_i$

2.3.2 Properties of the variance

- $\mathbb{VAR}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $\mathbb{VAR}[X + c] = \mathbb{VAR}[X] \qquad \mathbb{VAR}[cX] = c^2\mathbb{VAR}[X]s$
- $\mathbb{VAR}[aX + bY + c] = a^2\mathbb{VAR}[X] + b^2\mathbb{VAR}[Y] + 2ab\mathbb{COV}[X, Y]$
- **For independent** X **and** Y $\mathbb{VAR}[X + Y] = \mathbb{VAR}[X] + \mathbb{VAR}[Y]$

2.3.3 Properties of the covariance

- $\mathbb{VAR}[X] = COV[X, X]$
- $\mathbb{COV}[cX, Y] = c\mathbb{COV}[X, Y] \qquad \mathbb{COV}[X, cY] = c\mathbb{COV}[X, Y]$
- $\mathbb{COV}[X + Y, Z] = \mathbb{COV}[X, Z] + \mathbb{COV}[Y, Z]$
- $\mathbb{COV}[X, Y + Z] = \mathbb{COV}[X, Y] + \mathbb{COV}[X, Z]$
- $\mathbb{COV}[X, Y] = \mathbb{COV}[Y, X]$
- $\mathbb{COV}[X, c] = 0$
- **For independent** X **and** Y $\mathbb{COV}[X, Y] = 0$
- $\mathbb{COV}[X + Y, Z + W] = \mathbb{COV}[X, Y] + \mathbb{COV}[X, W] + \mathbb{COV}[Y, Z] + \mathbb{COV}[Y, W]$

2.3.4 Correlation and its properties

$$Cor[X, Y] = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

1. $-1 \leq Cor[X, Y] \leq 1$
2. $Cor[X, Y] = Cor[Y, X]$

An example of calculation of the correlation is provided in the R file **Joint distributions.R**!

2.4 Conditional probability

- *Discrete case*

$$P_{y|x}(y|x) = \frac{P_{xy}(x, y)}{P_{xx}(x)}$$

$$P_{x|y}(x|y) = \frac{P_{xy}(x, y)}{P_{yy}(y)}$$

If X and Y are independent, then

$$P_{y|x}(y|x) = \frac{P_{xy}(x, y)}{P_{xx}(x)} = \frac{P_{xx}(x) \cdot P_{yy}(y)}{P_{xx}(x)} = P_{yy}(y)$$

- *Continuous case*

$$f_{y|x}(y|x) = \frac{f_{xy}(x,y)}{f_{xx}(x)}$$

$$f_{x|y}(x|y) = \frac{f_{xy}(x,y)}{f_{yy}(y)}$$

If X and Y are independent, then

$$f_{y|x}(y|x) = \frac{f_{xy}(x,y)}{f_{xx}(x)} = \frac{f_{xx}(x) \cdot f_{yy}(y)}{f_{xx}(x)} = f_{yy}(y)$$

2.5 Chebyshev's inequality

$$P[|X - u| > \epsilon] \leq (\frac{\sigma}{\epsilon})^2$$

if and only if

$$P[|X - u| \leq \epsilon] \geq 1 - (\frac{\sigma}{\epsilon})^2$$

Chapter 3

Discrete RVs

3.1 Bernoulli distribution ($X \sim Ber(p)$)

Distribution that takes the value 1 (success) with probability p, and the value 0 (failure) with probability $1 - p$.

Good or defective components, parts that pass or fail tests, transmitted or lost signals, working or malfunctioning hardware, benign or malicious attachments, sites that contain or do not contain a keyword, girls and boys, heads and tails, and so on, are examples of Bernoulli trials.

$P[X = x]$	$p^x(1-p)^{1-x} = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$
$\mathbf{E}[X]$	p
$\mathbf{Var}[X]$	$p(1-p)$

3.2 Binomial distribution ($X \sim Bin(n, p)$)

Distribution that counts the number of success in a sequence of independent Bernoulli trials.

Examples of Binomilan variables are the number of defective computers in a shipment, the number of updated files in a folder, the number of girls in a family, the number of e-mails with attachments, etc..

n	number of Bernoulli trials
x	number of successes
p	probability of success
$P[X = x]$	$\binom{n}{x} p^x (1-p)^{n-x}$
$P[X \leq x]$	$\sum_{i=1}^{x} \binom{n}{i} p^i (1-p)^{n-i}$
$\mathbf{E}[X]$	np
$\mathbf{Var}[X]$	$np(1-p)$
$P[X = x]$	**dbinom**(x, n, p)
$P[X \leq x]$	**pbinom**(x, n, p)
$P[X < x]$	**pbinom**(x − 1, n, p)
$P[X > x]$	**pbinom**(x, n, p, **lower.tail** = FALSE)
$P[X \geq x]$	**pbinom**(x − 1, n, p, **lower.tail** = FALSE)
$P[X \leq x] = q$	**qbinom**(**q**, n, p) = x
r realizations of X	**rbinom**(r, n, p)

3.3 Geometric distribution ($X \sim Geom(p)$)

Distribution that counts the number of Bernoulli trials needed to get the first success at the x-th trial.

$P[X = x]$	$(1-p)^{x-1}p, x = 1, 2, ...$
$P[X \leq x]$	$p\sum_{i=0}^{x}(1-p)^i$
$P[X > x]$	$(1-p)^x$
$\mathbf{E}[X]$	$\frac{1}{p}$
$\mathbf{Var}[X]$	$\frac{1-p}{p^2}$
$P[X = x]$	**dgeom**(x−1, p)
$P[X \leq x]$	**pgeom**(x−1, p)
$P[X < x] = P[X \leq x-1]$	**pgeom**(x−2, p)
$P[X \geq x]$	1 − **pgeom**(x−2, p)
$P[X > x] = P[X \geq x+1]$	1 − **pgeom**(x−1, p)

3.3.1 Memoryless property

The Geometric distribution is the only discrete distribution with the **memoryless property**, i.e. if $X \sim Geom(p)$, then

$$P[X > x + y | X > y] = P[X > x] \tag{3.1}$$

We can derive this property from

$$P[X > x+y|X > y] = \frac{P[X > x+y, X > y]}{P[X > y]} = \frac{P[X > x+y]}{P[X > y]} = \frac{(1-p)^{x+y}}{(1-p)^y} = (1-p)^x = P[X > x] \tag{3.2}$$

3.4 Negative binomial distribution ($X \sim NBin(k, n, p)$)

Distribution that counts the number of Bernoulli trials needed to get the k-th success at the x-th trial.

k	number of desired successes
$P[X = x]$	$\binom{x-1}{k-1}(1-p)^{x-k}p^k \quad x = k, k+1, ...$
$\mathbf{E}[X]$	$\frac{k}{p}$
$\mathbf{Var}[X]$	$\frac{k(1-p)}{p^2}$
$P[X = x]$	**dnbinom**(x, k, p)
$P[X \leq x]$	**pnbinom**(vec, size, prob)

3.5 Multinomial distribution

Whereas the binomial distribution describes the number of successes in a Bernoulli process, for which each single test can provide only two results, the multinomial distribution describes the more general case in which each test can provide a finite number of results, each with the own probability.

$$P[X_1, X_2, ..., X_k] = \frac{n!}{x_1!...x_k!} p_1^{x_1} ... p_k^{x_k} \quad \text{where} \quad \sum_{i=1}^{n} p_1 = 1$$

3.6 Hypergeometric distribution

Distribution that is related to the problem of populations and sub-populations without replacement.

N	Total number of elements
K	Total number of elements with the characteristic
n	Number of chosen elements
k	Number of chosen elements with the characteristic
$P[X = k]$	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
$\mathbf{E}[X]$	$n\frac{K}{N}$
$\mathbf{Var}[X]$	$n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$
$P[X = k]$	**dhyper**(k, K, N–K, n)
$P[X \leq k]$	**phyper**(k, K, N–K, n)
$P[X \leq k] = q$	**qhyper**(**q**, K, N–K, n) = k

3.7 Poisson distribution ($X \sim Po(\lambda)$)

Distribution that describes the number of rare events occurring within a fixed period of time.

Arrivals of jobs, telephone calls, e-mail messages, traffic accidents, network blackouts, virus attacks, errors in software, floods, and earthquakes are examples of rare events.

λ	frequency, average number of events
p	probability of success
$P[X = x]$	$e^{-\lambda}\frac{\lambda^x}{x!} \quad x = 0, 1, 2, ...$
$\mathbf{E}[X]$	λ
$\mathbf{Var}[X]$	λ
$P[X = x]$	**dpois**(x, lambda)
$P[X \leq x]$	**ppois**(x, lambda)
$P[X < x]$	**ppois**(x − 1, lambda)
$P[X > x]$	**ppois**(x, lambda, **lower**.tail = FALSE)
$P[X \geq x]$	**ppois**(x − 1, lambda, **lower**.tail = FALSE)

3.7.1 Poisson approximation of Binomial distribution

Poisson distribution can be effectively used to approximate Binomial probabilities when the *number of trials* n is **large** and the *probability of success* p is **small**. In particular, this approximation is valid when $n \geq 30$ and $p \leq 0.05$.

$$\text{Binomial}(n, p) \approx \text{Poisson}(\lambda)$$

, where $n \geq 30$, $p \leq 0.05$ and $np = \lambda$. More formally, let's consider $X \sim Bin(n, p)$, then:

$$\lim_{n\to\infty, p\to 0, np\to\lambda} \binom{n}{x} p^x (1-p)^{n-x} = e^{-}\lambda\frac{\lambda^x}{x!}$$

3.7.2 Additivity

If $X \sim Pois(\lambda)$ and $Y \sim Pois(\mu)$ and they are **independent**, then we can say that:

$$W = X + Y \sim Pois(\lambda + \mu)$$

This property extends to countable additivity, so if $X_1, X_2, ..., X_n \sim Pois(\lambda_i)$, then $S = \sum_{i=1}^{n} X_i \sim Pois(\sum_i \lambda_i)$

3.7.3 Relation between Poisson and Multinomial distribution

Let $S_n = X_1 + X_2 + ... + X_n$ with $X_i \overset{iid}{\sim} Pois(\lambda_i)$, then:

$$(X_1, X_2, ..., X_n)|S_n \sim Mult\Big(\frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}, ..., \frac{\lambda_n}{\lambda}\Big) \quad \text{where} \quad \lambda = \sum_{i=1}^{n} \lambda_i$$

Chapter 4

Continuous RVs

4.1 Uniform distribution

The distribution describes an experiment where there is an arbitrary outcome that lies between certain bounds. The bounds are defined by the parameters, a and b, which are the minimum and maximum values.

(a, b)	range of values
$f(x)$	$\frac{1}{b-a} \quad a < x < b$
$P[X \leq x]$	$\frac{x-a}{b-a} \quad a < x < b$
$\mathbf{E}[X]$	$\frac{a+b}{2}$
$\mathbf{Var}[X]$	$\frac{(b-a)^2}{12}$

$P[X = x]$	**dunif**(x, a, b)
$P[X \leq x] = P[X < x]$	**punif**(x, a, b)
$P[X \geq x] = P[X > x]$	**punif**(x, a, b), **lower.tail** = FALSE

4.2 Exponential distribution

Exponential distribution used to model **time**. In a sequence of rare events, when the number of events is Poisson, the time between events is Exponential

Example of times are waiting time, interarrival time, hardware lifetime, failure time, time between telephone calls, etc..

λ	frequency parameter: number of events per time unit
$f(x)$	$\lambda e^{-\lambda x} \quad x > 0$
$P[X \leq x]$	$1 - e^{-\lambda x} \quad x > 0$
$P[X > x]$	$e^{-\lambda x} \quad x > 0$
$\mathbf{E}[X]$	$\frac{1}{\lambda}$
$\mathbf{Var}[X]$	$\frac{1}{\lambda^2}$
$P[X = x]$	**dexp**(x, lambda)
$P[X \leq x] = P[X < x]$	**pexp**(x, lambda)
$P[X \geq x] = P[X > x]$	**pexp**(x−1, lambda, **lower.tail** = FALSE)

4.2.1 Times between rare events are Exponential

Event: "the time T until the next event is greater than t" can be rephrased as: "zero events occur by the time t".

$$P_X(0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

Then the cdf of T is:

$$F_T[t] = 1 - P[T > t] = 1 - P[T = t] = 1 - e^{-\lambda t}$$

, which denotes an Exponential cdf.

4.2.2 Lack of memory property

The fact of having waited for t minutes gets "forgotten", and it does not affect the future waiting time.

$$P[T > t + x | T > t] = P[T > x] \quad \forall t, x > 0$$

4.2.3 Minimization

Consider a collection of $X_i \sim Exp(\lambda_i)$ with $i = 1, ..., n$ independent from each other we state that there exist a new random variable:

$$\min\{X_1, ..., X_n\} \sim Exp(\lambda) \quad \text{,with } \lambda = \sum_{j=1}^{n} \lambda_i$$

and we have that:

$$P[X_k = \min\{X_1, ..., X_n\}] = \frac{\lambda_k}{\lambda_1 + \lambda_2 + .. + \lambda_n}$$

4.3 Gamma distribution

When a certain procedure consist of α *independent steps*, and each step takes **Exponential(λ)** amount of time, then the total time has **Gamma distribution** with parameters α and λ.
In a process of rare events, with **Exponential** times between any two consecutive events, the time of the α-th events has **Gamma** distribution because it consists of α *independent* **Exponential** times.

α	shape parameter
λ	frequency parameter
$f(x)$	$\frac{\lambda^\alpha}{\rho(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad x > 0$
$\mathbf{E}[X]$	$\frac{\alpha}{\lambda}$
$\mathbf{Var}[X]$	$\frac{\alpha}{\lambda^2}$

$P[X = x]$	**dgamma**(x, alpha, rate = lambda)
$P[X \leq x] = P[X < x]$	**pgamma**(x, alpha, rate = lambda)
$P[X \geq x] = P[X > x]$	**pgamma**(x, alpha, rate = lambda, **lower.tail** = FALSE)

4.3.1 Gamma-Poisson formula

Computation of Gamma probabilities can be significantly simplified by thinking of a Gamma variable as the time between some rare events. Indeed, let T be a Gamma variable with an integer parameter α and some positive λ. This is a distribution of the time of the $\alpha - th$ rare event. Then, the event $T > t$ means that the $\alpha - th$ rare event occurs after the moment t, and therefore, fewer than α rare events occur before the time t.

For these reasons, for a variable $T \sim Gamma(\alpha, t)$ and a variable $X \sim Poisson(\lambda t)$,

$$P[T \geq t] = P[X \leq \alpha]$$
$$P[T \leq t] = P[X \geq \alpha]$$

4.4 Normal distribution

Besides sums, averages, and errors, Normal distribution is often found to be a good model for physical variables like weight, height, temperature, voltage, pollution level, and for instance, household incomes or student grades.

μ	expectation, location parameter
σ	standard deviation, scale parameter
$f(x)$	$\frac{1}{\sigma\sqrt{2\pi}} exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < x < \infty$
$F_x[X]$	$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} dz \quad -\infty < x < \infty$
$\mathbf{E}[X]$	μ
$\mathbf{Var}[X]$	σ^2
$P[X = x]$	**dnorm**(x, **mean**, **sd**)
$P[X \leq x] = P[X < X]$	**pnorm**(x, **mean**, **sd**)
$P[X \geq x] = P[X > X]$	**pnorm**(x, **mean**, **sd**, **lower.tail** = FALSE)

4.4.1 Standard normal distribution

Normal distribution with "standard parameters" $\mu = 0$ and $\sigma = 1$ is called **Standard Normal distribution**.

μ	expectation, location parameter
σ	standard deviation, scale parameter
Z	Standard Normal Random Variable
$\phi(x)$	$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ Standard Normal **pdf**
$\Phi(x)$	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ Standard Normal **cdf**

$P[X = x]$	**dnorm**(x, 0, 1)
$P[X \leq x] = P[X < x]$	**pnorm**(x, 0, 1)
$P[X \geq x] = P[X > x]$	**pnorm**(x, 0, 1, **lower.tail** = FALSE)

A Standard Normal can be obtained from a non standard Normal(μ, σ) random variable X by **standardizing**, which means *subtracting* the **mean** and *dividing* by the **standard deviation**:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Using the transformation, any Normal Random Variable can be obtained from a **Standard Normal Random Variable** Z**:**

$$F_x[x] = P[X \leq x] = P\Big[\frac{X' - \mu'}{\sigma'} \leq \frac{x - \mu}{\sigma}\Big] = P\Big[Z \leq \frac{x - \mu}{\sigma}\Big] = F_z\Big[\frac{x - \mu}{\sigma}\Big]$$

Linear Combination of Normal RVs are Normal

$$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$\text{and if} \quad a_i = \frac{1}{n} \quad \forall i \quad \text{or} \quad \frac{1}{n}\sum_{i=1}^{n} x_i \sim N(\mu, \sigma^2/n)$$

$$\sum_{i=1}^{n} a_i x_i \sim N\Big(\mu \sum_{i=1}^{n} a_i, \sigma^2 \sum_{i=1}^{n} a_i\Big)$$

4.5 Central limit theorem

$$X_1, X_2, ... \quad \text{independent RVs} \quad \mu = \mathbf{E}[X_i] \quad \sigma = \text{Std}[X_i]$$

$$S_n = \sum_{i=1}^{n} X_i = X_1 + ... + X_n$$

As $n \to \infty$ the standardized sum is

$$Z_n = \frac{S_n - \mathbf{E}[S_n]}{\text{Std}[S_n]} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a **Standard Normal Random Variable**

$$F_{Z_n}(z) = P\Big[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\Big] \to \Phi(z) \quad \forall z$$

Applied when $n \geq 30$ and for any distribution

4.5.1 Normal approximation to binomial distribution

Binomial Variables represent a special case of $S_n = X_1 + ... + X_n$, where all $X_i \sim Ber(p)$, moreover in case our n is **large** and for moderate values of $p : (0.05 \leq p \leq 0.95)$ we have the following approximation

$$\text{Binomial}(n, p) \approx Normal\Big(\mu = np, \sigma = \sqrt{np(1-p)}\Big)$$

4.5.2 Continuity correction

It is needed when we approximate a discrete distribution (like Binomial) by a continuous distribution (Normal). Since in the discrete case $P[X = x]$ could be positive in the continuous case it is always 0. This is way we introduce this correction.
We expand the interval by 0.5 units in each direction, then use the Normal approximation.

$$P_X[x] = P[X = x] = P[x - 0.5 < X < x + 0.5]$$

Chapter 5

Approximations

Given $X_1, X_2, ..., X_n$ a sequence of independent random variables with $S_n = \sum_{i=1}^{n} X_i$ if:

- $X_i \overset{iid}{\sim} \text{Bernulli}(p) \approx S_n \sim \text{Binomial}(n, p)$
- $X_i \overset{iid}{\sim} \text{Geometric}(p) \approx S_n \sim \text{NegativeBinomial}(n, p)$
- $X_i \overset{iid}{\sim} \text{Exponential}(\lambda) \approx S_n \sim \text{Gamma}(n, \lambda)$
- $X_i \overset{iid}{\sim} \text{Poisson}(\lambda) \approx S_n \sim \text{Poisson}(n, \lambda)$

Chapter 6

Stochastic processes

A **stochastic process** is a random variable $X(t, w)$ that also depends on **time**, thus it is a function of two arguments:

- $t \in \mathcal{T}$ is **time**, where $\mathcal{T}$ is the set of possible times;
- $w \in \Omega$ is an outcome of an experiment.

Moreover:

- at any time t we have a **random variable** $X_t(w)$, a function of a random outcome;
- at a given w we obtain a **function of time** $X_w(t)$, which is also called the **trajectory** of the stochastic process;
- the collection of all sample functions is called **ensembe** of the random process.

A stochastic process is classified as follows:

- **variable classification:**
  - $X(t, w)$ is a **discrete-state process** if X_t is a *discrete random variable* $\forall t$
  - $X(t, w)$ is a **continuous-state process** if X_t is a *continuous random variable* $\forall t$
- **time dimension classification:**
  - $X(t, w)$ is a **discrete-time process** if the set of time $\mathcal{T}$ is *discrete*
  - $X(t, w)$ is a **continuous-time process** if $\mathcal{T}$ is unbounded and thus *continuous*

6.1 Proprieties

6.1.1 Mean function

Given a stochastic process $X(t)$, the **mean function** of the process is given by:

$$\mu_x(t) = \mathbb{E}[X(t)]$$

,where $\mathbb{E}[X(t)]$ is the **expected value** of the random variable for the fixed *time point* t.

6.1.2 Variance function

Similarly, the **variance function** of $X(t)$ is given by:

$$\sigma_x^2(t) = \mathbf{Var}[X(t)] = \mathbb{E}[(X(t) - \mu_x(t))^2] = \mathbb{E}[X^2(t)] - [\mu_x(t)]^2$$

6.1.3 Standard deviation function

$$\sigma_x(t) = \sqrt{\mathbf{Var}[X(t)]} = \sqrt{\sigma_x^2(t)}$$

NOTE: these three functions are non-random functions, which means that they're numbers and not random quantities.

6.1.4 Autocovariance function

We've seen that both the mean function and the standard deviation function contain the information about the ensemble at each single point of time. If we consider two different times ts, then we get two random variables $X(t)$ and $X(s)$, which are in some way related:

$$\sigma_x(t,s) = C_{x,x} = \mathbf{Cov}[X(t), X(s)] = \mathbb{E}\Big[\Big(X(t) - \mu_x(t)\Big) \times \Big(X(s) - \mu_x(s)\Big)\Big]$$

$$= \mathbb{E}\Big[\Big(X(t) \times X(s)\Big) - \Big(\mu_x(t) \times \mu_x(s)\Big)\Big]$$

And it has the following proprieties:

- $C_{x,x}(t,s) = C_{x,x}(s,t)$
- $\sigma_x^2(t) = \mathbf{Var}[X(t)] = \mathbf{Cov}[X(t), X(t)] = C_{x,x}(t,t) = \mathbb{E}[X^2(t)] - u_x^2(t)$
- It is interpreted as the classic covariance, so it basically follows its properties:
  - when $C_{xx}(t,s) > 0$, above-average values of $X(t)$ tend to be associated with above-average values of $X(s)$, and vice-versa;
  - $\sigma_x(t,s) = 0$ does not automatically implies that $X(s)\text{and}X(t)$ are independent (they are if $\sigma_x(t,s) = 0\ \forall s \neq t$ and if $X(t)$ is a Gaussian process).

6.1.5 Autocorrelation function

$$\varphi_x(t,s) = \frac{\sigma_x(t,s)}{\sigma_x(t)\sigma_x(s)} = \frac{C_{x,x}(t,s)}{C_{x,x}(t)C_{x,x}(s)} = \frac{\text{autocovariance}}{\text{SD(s)} \cdot \text{SD(t)}}$$

In context of **signal processing** ad in **engineering literature** the **autocorrelation functon** is denoted as $R_{x,x}(t,s)$ and is defined as:

$$R_{x,x}(t,s) = \mathbb{E}[X(t)X(s)]$$

And it is equivalent to $\sigma_x(t,s)$ only when the mean $= 0$ and the variance $= 1$

6.2 Stationary and wide-sense stationary processes

6.2.1 Strongly or Strict-sense stationary processes

A stochastic process is called **strongly / strict-sense stationary** if:

- all its *statistical proprieties* are **invariant over time**
- for any points $t_1, ..., t_r$ and any value τ, the two following **joint distributions** are **equivalent**

$$X(t_1), ..., X(t_r) \equiv X(t_1 + \tau), ..., X(t_r + \tau)$$

It has the following proprieties:

- $X(t)$ and $X(t+\tau)$ have the *same distribution*, thus same mean, variance and standard deviation;
- since the **joint distribution** of $X(t_1)$ and $X(t_2)$ is invariant with respect to its statistical proprieties over time (can be shifted over time with no changes in proprieties) or, more shortly, it is **translation invariant**, also the **autocovariance** of $X(t)$ must be **translation invariant**;
- same invariance $\forall r \geq 1$.

6.2.2 Weakly ot Wide-Sense stationary processes

A stochastic process $X(t)$ is **weakly / wide-sense stationary** if the following two conditions holds:

1. the **mean function** of $X(t)$, $\mu_x(t)$ is *constant*
2. The **autocovariance function** of $X(t), C_{xx}(t,s)$ *depends* only on $(s-t)=\tau$. If we write $s=t+\tau$, then $C_{xx}(t,t+\tau)$ only depends on τ, not on t.

NOTE:

- neither $\mu_x(t)$ not $C_{xx}(t,t+\tau)$ depend on t, so we can arbitrarily fix time t_0 at the most convenient time;
- some processes are **asymptotically (weakly) stationary** if:
  - $\mu(t) \xrightarrow[t\to\infty]{} \mu$;
  - $\sigma(t,t+s) \xrightarrow[t\to\infty]{} \sigma(h)$

Chapter 7

Poisson Process

7.1 Definitions

7.1.1 Definition 1

A **Poisson process** with intensity λ is a *continuous-time counting process* $N = \{N(t) : t \geq 0\}$ taking values in $S = \{0, 1, 2, ...\} = \mathbb{N}$ such that:

1. $N(0) = 0$
2. **The increments are independent and stationary**, i.e. the distribution of $N(t + h) - N(t)$ depends only on h
3. $P[N(h) = 1] = \lambda h + o(h)$ and $P[N \geq 2] = o(h)$, which means that:

$$\lim_{h \to 0} P[N(h) = 1] = \lambda h$$

and

$$\lim_{h \to 0} P[N(h) \geq 2] = 0$$

.

7.1.2 Definition 2

A **Poisson process** with intensity λ is a *continuous-time counting process* $N = \{N(t) : t \geq 0\}$ such that:

1. $N(0) = 0$
2. For each non overlapping time intervals $(s_1, s_1+t_1]$ and $(s_2, s_2+t_2]$ for $t_1, t_2, s_1, s_2 \geq 0$, the increments $N(s_1 + t_1) - N(s_1)$ and $N(s_2 + t_2) - N(s_2)$ are **independent**
3. $\forall t, s \geq 0$ the **increment** $N(t + s) - N(s)$ has a **Poisson** (λt) distribution, i.e. $N(t + s) - N(s) \sim Po(\lambda t)$

7.1.3 Definition 3

A **Poisson process** with intensity λ is a *continuous-time counting process* $N = \{N(t) : t \geq 0\}$ s.t.

1. $N(0) = 0$
2. Let:
   - $T_0 = 0$;
   - $\forall n \geq 1\ T_n = inf\{t : N(t) = n\}$ be the n-th arrival time.

   Then, the **interarrival times** $X_n = T_n - T_{n-1}$ are *iid* **exponential random variables** with rate λ

7.2 Properties

- **marginal distribution**:

$$N(t) \sim Po(\lambda t) \quad t > 0$$

and

$$N(0) = 1$$

In this sense, the pmf of the poisson process is:

$$P[N(t) = x] = e^{-\lambda t}\frac{(\lambda t)^x}{x!}$$

- **mean function**: $\mu(t) = \mathbb{E}[N(t)] = \lambda t$
- **variance function**: $\sigma^2(t) = \mathbb{VAR}[N(t)] = \lambda t$
- **covariance function**: $\sigma(s,t) = Cov(N(s), N(t))$
  - If $s > t$ we split into $[0,s) = [0,t) \cup (t,s]$

$$\begin{aligned} Cov(N(t), N(s)) =& Cov(N(t), N(t) + N(s) - N(t))) \\ & Cov(N(t), N(t)) + Cov(N(t), N(s) - N(t)) \\ & \mathbb{VAR}(N(t)) + 0 = \lambda t \end{aligned}$$

  - If $t > s$ we split into $[0,t) = [0,s) \cup (s,t]$

$$\begin{aligned} Cov(N(s), N(t)) =& Cov(N(s) + N(t) - N(s), N(t))) \\ & Cov(N(s), N(s)) + Cov(N(t) - N(s), N(s)) \\ & \mathbb{VAR}(N(s)) + 0 = \lambda s \end{aligned}$$

- **transition probabilities**

$$\mathcal{P}_{m,n+m}(h) = \mathbb{P}[N(t+h) = n + m | N(t) = m] = \begin{cases} 1 - \lambda h + o(h) & \text{if } n = 0 \\ \lambda h + o(h) & \text{if } n = 1 \\ o(h) & \text{if } n > 1 \end{cases}$$

- **stationary increments**

$$\begin{aligned} \mathcal{P}_{m,n+m}(h) &= \mathbb{P}[N(t+h) = n + m | N(t) = m] \\ &= \mathbb{P}[N(t+h-t) = n + m - m | N(0) = 0] \\ &= \mathbb{P}[N(h) = n] \\ &= \sum_{x=0}^{n} \frac{e^{-\lambda t}(\lambda t)^x}{x!} \text{ used also for } \mathbb{P}[N(h) \leq n] \end{aligned}$$

NOTE: this means that, for example, $P[N(t) = 0]$ in the interval $(0, t)$ is the same as the $P[N(t) = 0]$ in the interval $(s, s + t)$. In other words:

$$N(t + s) - N(s) = N(t)$$

- **independent increments**

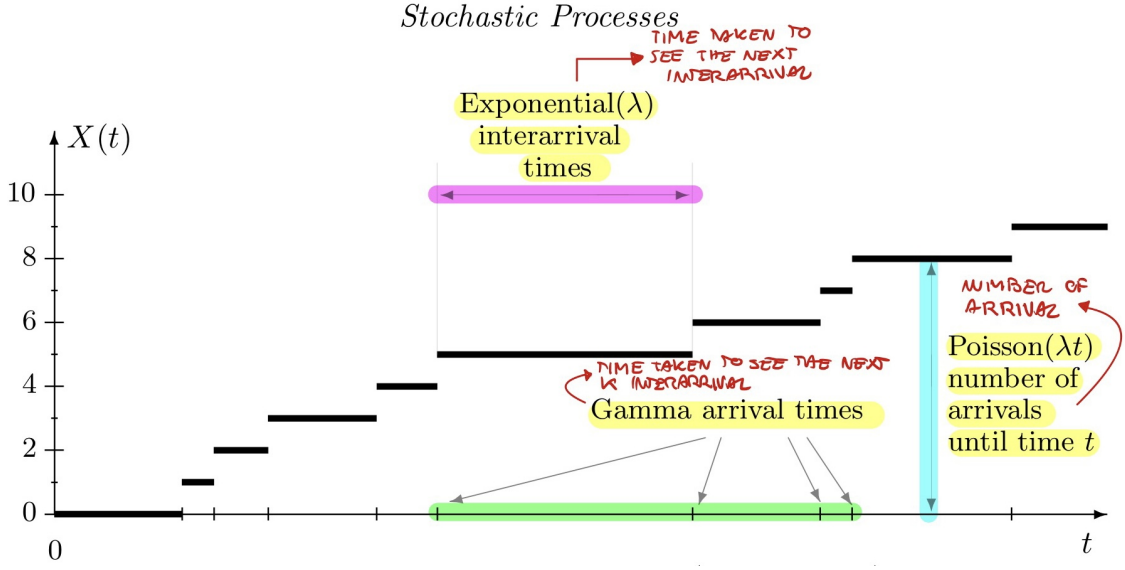
$$(s_1, t_1) \cap (s_2, t_2) = \emptyset \rightarrow N(t_1) - N(s_1) \perp\!\!\!\perp N(t_2) - N(s_2)$$

- **interarrival times** : $X_i \sim Exp(\lambda)$
- **arrival times**: $T_n = \sum_i X_i \sim Gamma(n, \lambda)$, so each arrival time is the sum of independent interarrival times, and they have the following properties:

$$P[T_n \leq t] = P[N(t) \geq n]$$

and

$$P[T_n > t] = P[N(t) < n]$$



- **superposition**:

$$N_i \overset{\text{iid}}{\sim} PP(\lambda_i) \rightarrow N = \sum_{i=1}^{\infty} N_i \sim PP(\lambda) \quad \text{where} \quad \lambda = \sum_{i=1}^{\infty} \lambda_i \text{ must be finite}$$

- **thinning**: If $N \sim PP(\lambda)$ and each arrival assigned to process N_i with probability p_i independently from all others for $p_i \in (0, 1)$ such that $\sum_{i=1}^{\infty} p_i = 1$, then:

$$N_i \overset{\text{ind}}{\sim} PP(\lambda \cdot p_i)$$

NOTE: an example of a simulation of a Poisson process is contained if the R file **PoissonProcessSimulation.r**!

Chapter 8

HCTMC

8.1 Markov property

A stochastic process $X = \{X(t) : t \geq 0\}$ is **Markov** if for any $t_1 < t_2 < ... < t_n < t$ and for any sets (events) $A, A_1, ..., A_n$:

$$\mathbb{P}\Big[X(t) \in A | X(t_1) \in A_1, ..., X(t_{n-1}) \in A_{n-1}, X(t_n) \in A_n\Big] =$$

$$\mathbb{P}\Big[X(t) \in A | X(t_n) \in A_n\Big]$$

$$\mathbb{P}\Big[future | past, present\Big] = \mathbb{P}\Big[future | present\Big]$$

The future given the present is **independent** from the past

For a **Markov process** the *conditional distribution* of $X(t)$ is the **same** under two **different conditions:**

1. Given observations of the process X at several moments in the past
2. Given only the present, so the latest observation of X.

8.2 Definition

A family $X = X(t) : t \geq 0$ of random variables taking values in a discrete space S is called a **continuous-time Markov chain** if it satisfies the Markov property:

$$P[X(t_n) = x_j | X(t_1) = x_i, .., X(t_{n-1} = x_{i_{n-1}})] = P[X(t_n) = x_j | X(t_{n-1} = x_{i_{n-1}})]$$

Let $p_{ij}(s,t) = P[X(t) = j | X(s) = i]$ be the **transition probabilities** of a CTMC X, then the chain is called **homogeneous** if, for each pair (i,j), the transition probabilities only depends on the difference $t - s$.

In other words: $p_{ij}(s,t) = p_{ij}(0, t-s) \rightarrow$ **homogeneous chain**.

The matrix P_t is a S x S matrix whose entries are $p_{ij}t = P[X(t) = j | X(0) = i]$, and in which:

- each row represents the current state (i.e. the conditioning state);
- each column represents the future state.

$$P_t = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \dots \\ \vdots & & \end{bmatrix}$$

8.3 Transition semigroup

The family $\{P_t : t \geq 0\}$ is the **transition semigroup** of the process.

The transition semigroup satisfies the following properties:

1. $P_0 = I$. This means that at time zero, the probability of staying where we are is equal to 1:

$$p_{ij}(0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

2. $\forall t > 0$ P_t is a *stochastic matrix*, i.e.

   - every entity is non negative
   - the sum of the elements of each row is 1 by (law of total probability)

3. $\forall s \geq 0$ and $\forall t \geq 0$,

$$P_{s+t} = P_s \cdot P_t$$

   This system of equations is called **Chapman-Kolmogorov Equations** and it can be written in the matrix form as follows:

$$p_{ij}(s+t) = \begin{bmatrix} p_{i1}(s) & p_{is}(s) & \dots \end{bmatrix} \cdot \begin{bmatrix} p_{1j}(s) \\ p_{2j}(s) \\ \vdots \end{bmatrix} = \sum_{k=1}^{\#S} p_{ik}(s) \cdot p_{kj}(t)$$

   This means that the **stochastic semigroup** $\{P_t\}$ together with the **distribution of the initial point** $X(0)$ determine the **behaviour of the process** X.

The semigroup $\{P_t\}$ is **standard** if in addition to the previous 3 properties we have also that:

$$P_t \to I \quad p_{ij}(t) \xrightarrow[t \to 0]{} \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

8.4 Generator

Suppose that at time $t \geq 0$ the chain X is in state i. Let $h > 0$ small and consider what could happened in the small time interval $(t, t+h)$

- *No change in state:* with probability $p_{ii}(h) + o(h)$;
- *The chain arrives at state j:* with probability $p_{ij}(h) + o(h)$.

There exist constants $\{g_{ij} : i, j \in S\}$ such that, for small h,:

$$p_{ij}(h) = \begin{cases} g_{ij}h & \text{if } i \neq j \\ 1 + g_{ii}h & \text{if } i = j \end{cases}$$

The matrix $\mathbf{G}$ with entries g_{ij} is called **generator**.

The generator takes the role of the **transition matrix** P over *discrete-time chains*, thus we can construct a **path** of X by moving it iteratively for sufficient small h:

$$X(t+h) = \begin{cases} i & \text{with probability } 1 + g_{ii}(h) + o(h) \\ j \neq i & \text{with probability } g_{ij} + o(h) \end{cases}$$

Moreover, assuming that h is small enough, we obtain:

$$1 = \sum_{j \in S} p_{ij}(h) \approx 1 + g_{ii}h + \sum_{j \neq i} g_{ij}h = 1 + h \sum_{j \in S} g_{ij} = \sum_{j \in S} g_{ij} = 0$$

This means that G is a matrix with **rows adding up to 0** and **non-negative value outside the diagonal**:

$$g_{ii} = -\sum_{j \neq i} g_{ij}$$

The **generator** is therefore a square matrix with:

- the **instantaneous exit rates** on the *diagonal*, controlling the random time that the process stays in each state before jumping;
- the **instantaneous transition rates** *outside the diagonal*, controlling where the process goes after the jump.

8.5 Forward and Backward Equations

Can we recover P_t from G? Yes using the Chapman-Kolmogorov equation, for sufficiently small h considering the transition probabilities for $(t + th)$ given $X(t)$ yields to:

- **Forward**

$$p_{ij}(t+h) = \sum_{k \in S} p_{ik}(t) \cdot p_{kj}(h)$$

$$\lim_{h \to 0} \frac{p_{ij}(t+h) - p_{ij}(t)}{h} \approx \lim_{h \to 0} \frac{p_{ij}(t) + h \sum_{k \in S} p_{ik}(t) g_{kj} - p_{ij}(t)}{h}$$

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(t) g_{kj}$$

  or, in matrix notation,

$$P'_t = P_t \cdot G$$

- **Backward**

$$p_{ij}(t+h) = \sum_{k \in S} p_{ik}(h) \cdot p_{kj}(t)$$

$$\lim_{h \to 0} \frac{p_{ij}(t+h) - p_{ij}(t)}{h} \approx \lim_{h \to 0} \frac{p_{ij}(t) + h \sum_{k \in S} p_{ik}(t) g_{kj} - p_{ij}(t)}{h}$$

$$p'_{ij}(t) = \sum_{k \in S} g_{ik} p_{kj}(t)$$

  or, in matrix notation,

$$P'_t = G \cdot P_t$$

Subject to the (boundary) condition $P_0 = I$ satisfied by the transition semigroup, both the forward and backward equations have the same solution:

$$P_t = e^{tG}$$

which can be solved using the following approaches:

- using the **exponential matrix**;
- using the **diagonalization**

8.5.1 Matrix exponential

Let A be a square matrix, the matrix exponential e^A is the square matrix of the same size as A given by:

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2} A^2 + \frac{1}{6} A^3 + \dots$$

8.5.2 Diagonalization

A square matrix A is **diagonalizable** if it can be rewritten as $A = VDV^{-1}$

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_3 \end{bmatrix}$$

Where:

- $\{\lambda_1, ..., \lambda_2\}$ are the **eigenvalues** of A
- V is an invertible matrix whose columns are the corresponding **eigenvectors**

The diagonalization can be used to efficiently compute the **power** of the matrix $A = VDV^{-1}$:

$$e^{tG} = V\left(\sum_{n=1}^{\infty} \frac{1}{n!} D^n\right) V^{-1} = Ve^D V^{-1}$$

, where e^D is the diagonal matrix containing $e^{td_{ii}}$ in the diagonal.

NOTE:

- in general, the solution for a 2-state process is:

$$P_t = \frac{1}{\mu + \lambda} \begin{bmatrix} \lambda + \mu e^{-t(\lambda+\mu)} & \mu - \mu e^{-t(\lambda+\mu)} \\ \lambda - \lambda e^{-t(\lambda+\mu)} & \mu + \lambda e^{-t(\lambda+\mu)} \end{bmatrix}$$

- an example of computation of the solutions of forward and backward equations using the matrix exponential and the diagonalization methods in contained in the R file **Transition semigroup.R**.

8.6 Holding time & transition probabilities

The **holding time** is defined as the further time until the chain changes its state, and in can be written as

$$U_i = inf\{t \geq 0 : x(s+t) \neq i\}$$

, and $U_i \sim Exp(g_i)$, where

$$g_i := \sum_{j \neq i} g_{ij} = -g_{ii}$$

The **transition probability** that a transition to state j occurs, given that a transition out of state i occurs is:

$$\tilde{p_{ij}} = \frac{g_{ij}}{-g_{ii}} = \frac{g_{ij}}{g_i}$$

$$g_{ij} = -g_{ii} \cdot \tilde{p_{ij}}$$

The matrix $\tilde{P}$ is the **transition probability matrix**, and it is a stochastic matrix with zeros in the diagonal.

NOTE: in order to calculate the probability that a certain jump will be in a given state, we must consider the transition matrix. An example of these calculations is provided in the R file **Transition probability matrix.R**

8.7 Irreducible HCTMC

A HCTMC is **irreducible** if the probability of reaching state j from state i in time t is **positive**, i.e.

$$p_{ij}(t) > 0$$

$\forall i, j \in S$ and $t > 0$.

It is possible to verify the irreducibility condition using:

- the **generator** G: let X be a HCTMC with generator G and transition semigroup $P_t = e^{tG}$; then, X is irreducible if and only if for any pair if states (i, j) there exists a sequence $k_1, k_2, .., k_n$ of states such that:

$$g_{i,k_1} g_{k_1,k_2}, .., g_{k_n,j} \neq 0$$

- the **graph representation** we just need to verify there is **at least** one **path** from i to j $\forall i \neq j$

8.8 Stationary distribution

The vector π is a **stationary distribution** of the chain X if:

- $\pi_j \geq 0 \quad \forall j \in S$
- $\sum_{i \in S} \pi_i = 1$
- $\pi^t = \pi^t \cdot P_t$

The stationary distribution is also called **steady state sistribution** because if the process starts at the stationary distribution, then it will stay in the stationary distribution

$$X(0) \sim \mu_0 = \pi \Rightarrow X(t) \sim \mu_t = \pi \cdot P_t = \pi$$

Let X and irreducible HCTMC with standard semigroup $\{P_t\}$. If there exists a stationary distribution π then it is **unique** $\forall i, j \in S$

$$p_{ij}(t) = P[X(t) = j | X(0) = i] \xrightarrow[t \to \infty]{} \pi_j$$

This means that if π exists, the chain will eventually reach the steady state, i.e. in the long run the marginal distribution of the chain will approach π, regardless of the initial distribution.

Example: if we have the following transition probability matrix:

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

, then the stationary distribution is calculated as:

$$\begin{cases} 0.8\pi_1 + 0.4\pi_2 = \pi_1 \\ 0.2\pi_1 + 0.6\pi_2 = \pi_2 \\ \pi_1 + \pi_2 = 1 \end{cases}$$

, from which we get that $\pi_1 = 2/3$ and $\pi_2 = 1/3$.

In order to calculate the stationary distribution of HCTMC, wee need to consider the global balance equations.

8.8.1 Global balance equations

A distribution π is the **stationary distribution** of a HCTMC with transition semigroup P_t if and only if:

$$\pi^t G = 0$$

The system of equations $\pi^T G = 0$ is called **global balance equations**, and it can be written as follows:

$$\pi \text{ is a stationary distribution} \leftrightarrow \pi_i \sum_{j \neq i} g_{ij} = \sum_{j \neq i} \pi_j \cdot g_{ji}$$

This result can be interpreted as **flux in** and **flux out** of a given state $j \in S$, meaning that in a steady state, the time that the process spends in a state i multiplied by the rate at which it leaves must equal the, on average, the sum of times the process spends in other states multiplied by the rate at which it returns to state i.

In general the global balance equations define an **irreducible** system of equations, so in order to obtain a **unique solution** we must consider the **normalization condition**:

$$\sum_{i \in S} \pi_i = 1$$

In the matrix form, this can be done by substituting one of the columns of the matrix G with a row of ones and placing a one in the corresponding pace of the row vector 0.

NOTE: Most of software systems used to compute the solution of linear systems in the form $Ax = b$, so it is convenient to write:

- $A = \tilde{G}^T$ where $\tilde{G}$ is obtained by substituting the **last column** of G with a column of **ones**;
- $b = e_N \in \mathbb{R}^n$ the n-th **canonical vector**, where only the **last element** is 1 and the others are 0;
- $x = \pi$.

NOTE: An example of solution of the global balance equations is contained in the R file **Stationary distribution.R**!

8.9 Simulating a HCTMC

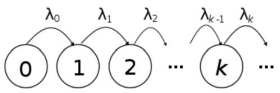
An example of a simulation of a HCTMC is contained in the R file **HCTMC simulation.r**!

8.10 Birth process

A **birth process** with intensity $\lambda_0, \lambda_1, \lambda_2, ...$ is a stochastic process $N = \{N(t) : t \geq 0\}$ taking value in $S = \{0, 1, 2, ...\}$ such that:

1. it is positive and non decreasing: $N(0) \geq 0$ and $N(s) \leq N(t)$ for $s < t$
2. $$P[N(t+h) = n + m | N(t) = n] = \begin{cases} \lambda_n + o(h) & \text{if } m = 1 \\ o(h) & \text{if } m > 1 \\ 1 - \lambda_n h + o(h) & \text{if } m = 0 \end{cases}$$
3. given $N(s)$, the increment $N(t) - N(s)$ is independent of all arrivals prior to s, $\forall s < t$.

A visual representation of the functioning of the birth process is provided in the following picture.



Properties 2 and 3 imply that N is an homogeneous continuous-time Markov chain. The **generator** of a birth process is the following:

$$G = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \dots \\ 0 & -\lambda_1 & \lambda_1 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

NOTE: the birth process is **not irreducible**, since we can never go back from a state!

8.11 Birth-death process

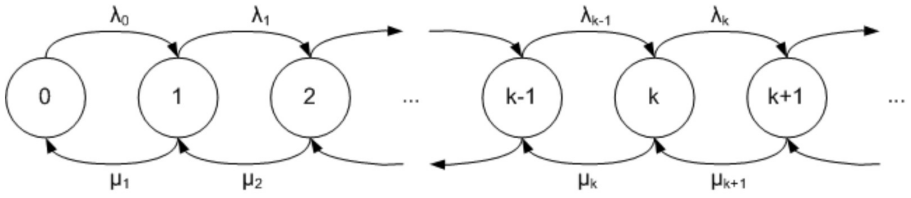
Suppose that the number $X(t)$ of individuals alive in some population at time t evolves in the following way:

1. X is a Markov chain taking values in $\{0, 1, 2, 3, ...\}$
2. The **infinitesimal transition probabilities** are given by:

$$P[N(t+h) = n+m | N(t) = n] = \begin{cases} \lambda_n h + o(h) & \text{if } m = 1 \\ \mu_n h + o(h) & \text{if } m = -1 \\ o(h) & \text{if } |m| > 1 \end{cases}$$

3. the birth rates $\lambda_i \geq 0$, the death rates $\mu_i \geq 0, \quad \mu_0 = 0$

A visual representation of the functioning of the birth-death process is provided in the following picture.



The **generator** is given by:

$$G = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

NOTE: the death-birth process is **irreducible** if $\lambda_i > 0$, $\forall i \geq 0$ and $\mu_i > 0$, $\forall i \geq 1$.

8.11.1 Global balance equations

If $\lambda_0 > 0$, the process is irreducible, so the stationary distribution π, if it exists, can be found as the unique solution to the global balance equations $\pi^T G = 0$. This is a system with an infinite number of equations:

$$\begin{aligned} -\lambda_0 \pi_0 + \mu_1 \pi_1 &= 0 \\ \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} &= 0 \text{ if } n \geq 1 \end{aligned}$$

In general, we have that:

$$\pi_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} \pi_0 \quad n \geq 1$$

However, the vector π is a stationary distribution if and only if $\sum_n \pi_n = 1$ which may happen if and only if:

$$1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} < \infty$$

, which, in turn, becomes:

$$1 + \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \frac{1}{1 - \frac{\lambda}{\mu}} < \infty$$

If it holds, then:

$$\pi_0 = \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}\right)^{-1}$$

, which becomes:

$$\pi_0 = 1 - \frac{\lambda}{\mu}$$

Using π_0 we can compute any π_i:

$$\pi_i = \frac{\lambda_0 \cdot \lambda_1 \cdot \dots \cdot \lambda_{i-1}}{\mu_1 \cdot \mu_2 \cdot \dots \cdot \mu_i} \pi_0$$