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 Zuliani Riccardo 875532

Contents

Chapter 1

Elementary Probability

- $A \cup B \quad A \cap B \quad A^c = 1 - A$
- $(A \cap B)^c = A^c \cup B^c \quad (A \cup B)^c = A^c \cap B^c$
- $A \setminus B = A \cap B^c$
- $P[A \cup B] = P[A] + P[B] - P[A \cap B] \quad P[A \cup B] = P[A] + P[B]$ if independent
- $P[A|B] = \frac{P[A \cap B]}{P[B]}$
  - If A and B are independent$P[A] = P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A] \cdot P[B]}{P[B]} = P[A]$
- Bayes rule:
$$P[B|A] = \frac{P[A|B] \cdot P[B]}{P[A]}$$
- System reliability:
  - *Series Circuit:* works if **all** sequential components work $\rightarrow \cap$
  - *Parallel Circuit:* it works if **at least** one of the components work $\rightarrow \cup$

Chapter 2

Random Variables

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	$\sum_x P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

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

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

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

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

Chapter 3

Function Properties

3.1 Mean

- $\mathbb{E}[aX + bY + c] = a\mathbb{E} + b\mathbb{E} + c$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[X] = a\mathbb{E}[X]$
- **For independent** X **and** Y $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$
- $X_1, X_2, ..., X_n$ RVs, then $\mathbb{E}[X_1 + X_2 + ... + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + ... + \mathbb{E}[X_n] = \sum_{i=1}^{n} X_i$

3.2 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 $\mathbb{VAR}[X + Y] = \mathbb{VAR}[X] + \mathbb{VAR}[Y]$

3.3 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 $\mathbb{COV}[X, Y] = 0$
- $\mathbb{COV}[X + Y, Z + W] = \mathbb{COV}[X, Y] + \mathbb{COV}[X, W] + \mathbb{COV}[Y, Z] + \mathbb{COV}[Y, W]$

Chapter 4

Discrete RVs

4.1 Bernoulli Distribution

Used whenever we have a 0 / 1 outcome, thus when we could have only two possible result in our experiment.

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

4.2 Binomial Distribution

Used whenever we consider a sequence of independent Bernoulli trials and count the number of success in it.

n	number of trials
p	probability of success
$P[x]$	$\binom{n}{x} p^x (1-p)^{n-x}$
$F[x]$	$\sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$
$\mathbf{E}[X]$	np
VAR$[X]$	$np(1-p)$

$P[X = x]$	**dbinom**$(\#success\ ,\ \ size\ ,\ \ prob\_success)$
$P[X \leq x]$	**pbinom**$(\#success\ ,\ \ size\ ,\ \ prob\_success)$

4.3 Geometric Distribution

Consider a sequence of independent Bernoulli trials, each trial results in a "success" or a "failure". The number of Bernoulli trials needed to get the first success has Geometric Distribution.

p	probability of success
$P[x]$	$(1-p)^{x-1}p, x = 1, 2, ...$
$F[x]$	$p\sum_{i=0}^{x}(1-p)^i$
$\mathbf{E}[X]$	$\frac{1}{p}$
VAR$[X]$	$\frac{1-p}{p^2}$
$P[X = x]$	**dgeom**$(\#failures\ ,\ \ prob\_success)$
$P[X \leq x]$	**pgeom**$(\#failures\ ,\ \ prob\_success)$

4.4 HyperGeometric Distribution

Describes the probability of k successes (random draws for which the object drawn has a specified feature) in n draws, without replacement, from a finite population of size N that contains exactly K objects with that feature, wherein each draw is either a success or a failure.

N	Is the population size
K	Is the number of success states in the population
n	Is the number of draws
k	Is the number of observed successes
$P[x]$	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
$\mathbf{E}[X]$	$n\frac{K}{N}$
$P[X = x]$	**dhyper**$(\#succ\ ,\ \#succ\_samp\ ,\ \#pop\_dim\ ,\ \#samp\_dim)$
$P[X \leq x]$	**phyper**$(\#succ\ ,\ \#succ\_samp\ ,\ \#pop\_dim\ ,\ \#samp\_dim)$

4.5 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.

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

4.6 Negative Binomial Distribution

In a sequence of independent Bernoulli trials, the number of trials needed to obtain k successes has Negative Binomial distribution.
In other words it counts the number of failures before obtaining a target number of successes (k).

k	number of success
p	probability of success
$P[x]$	$\binom{x-1}{k-1}(1-p)^{x-k}p^k \quad x = k, k+1, ...$
$\mathbf{E}[X]$	$\frac{k}{p}$
VAR$[X]$	$\frac{k(1-p)}{p^2}$
$P[X = x]$	**dnbinom**$(\#failures\ ,\ \#successes\ ,\ prob\_success)$ **dnbinom**$(\#trial-\#successes\ ,\ \#successes\ ,\ prob\_success)$
$P[X \leq x]$	**pnbinom**$(\#failures\ ,\ \#successes\ ,\ prob\_success)$ **pnbinom**$(\#trial-\#successes\ ,\ \#successes\ ,\ prob\_success)$

4.7 Poisson Distribution

Poisson distribution is related to **rare events**. It means that two events are extremely unlikely to occur simultaneously or within a very short period of time.
The number of rare events occurring within a fixed period of time has Poisson Distribution.

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

4.7.1 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**

$$n \geq 30 \quad p \leq 0.05$$

$$\text{Binomial}(n, p) \approx \text{Poisson}(\lambda)$$
$$\text{where} n \geq 30 \quad p \leq 0.05 \quad np = \lambda$$

4.7.2 Additivity

If

$$X \sim Pois(\lambda) \quad \text{and} \quad Y \sim Pois(\mu)$$

and they are **independent**, then we can say that:

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

4.7.3 Relation between Poisson and Multinomial Distribution

Let

$$S_n = X_1 + X_2 + ... + X_n \quad \text{with} \quad X_i \overset{iid}{\sim} Pois(\lambda_i)$$

given

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

Chapter 5

Continuous RVs

5.1 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
$f(x)$	$\frac{1}{b-a} \quad a < x < b$
$F_x(x)$	$\frac{x-a}{b-a} \quad a < x < b$
$\mathbf{E}[X]$	$\frac{a+b}{2}$
VAR$[X]$	$\frac{(b-a)^2}{12}$
$P[X = x]$	**dunif**(x, **min**, **max**)
$P[X \leq x]$	**punif**(x, **min**, **max**)

5.2 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

λ	frequency parameter, the number of events per time unit
$f(x)$	$\lambda e^{-\lambda x} \quad x > 0$
$F_x(x)$	$1 - e^{-\lambda x} \quad x > 0$
$\mathbf{E}[X]$	$\frac{1}{\lambda}$
VAR$[X]$	$\frac{1}{\lambda^2}$
$P[X = x]$	**dexp**(x, rate ($\hat{}-1$))
$P[X \leq x]$	**pexp**(x, rate ($\hat{}-1$))

If in the exercise it is not explicit the measure unit at $^{-1}$, we have to insert the **rate** parameter: $1/rate = rate^{-1}$. Like in the case that we know the mean, by the formula we can retrieve the exact value of λ.

5.2.1 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:

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

5.2.2 Memory Less Propriety

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

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

5.2.3 Minimization

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

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

It has the same propriety as a classical **Exponential Random Variable**

5.2.4 Maximization

$$\mathbb{P}[x \leq x] = \prod_{i=1}^{n}(1 - e^{-\lambda_i x})$$

$$\mathbb{E}[X] = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i}$$

5.3 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
$f(x)$	$\frac{\lambda^\alpha}{\rho(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad x > 0$
$\mathbf{E}[X]$	$\frac{\alpha}{\lambda}$
VAR$[X]$	$\frac{\alpha}{\lambda^2}$
$P[X = x]$	**dgamma**(x, alpha, rate (^−1))
$P[X \leq x]$	**pgamma**(x, alpha, rate (^−1))

If in the exercise it is not explicit the measure unit at $^{-1}$, we have to insert the **rate** parameter: $1/rate = rate^{-1}$. Like in the case that we know the mean, by the formula we can retrieve the exact value of λ.

5.4 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
VAR$[X]$	σ^2

5.4.1 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 − **mean**) / **sd**)
$P[X\leq x]$	**pnorm**((X − **mean**) / **sd**)
$\Phi^{-1}(x)$	**qnorm**(x)

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

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

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

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

5.5 Central Limit Theorem

To be used when in the exercise it is asked to find some king of probability given quantity of elements and relative mean and sd.

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

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

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 $n \geq 30$

5.5.1 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

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

5.5.2 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 6

Approximation

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

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

Chapter 7

Stochastic Processes

A **Stochastic Process** is a Random Variable that also depends on **time**, thus it is a function of two arguments $X(t, w)$ where:

- $t \in \mathcal{T}$ is **time**
- $w \in \Omega$

Moreover:

- At any time t we see a **random variable** $X_t(w)$: function of random income.
- At a given w we obtain a **function of time** $X_w(t)$

We have two classification of stochastic processes and are the following one:

- **Variable Classification:**
  - $X(t, w)$ is a **discrete-state process** if X_t is a *discrete rv* $\forall t$
  - $X(t, w)$ is a **continuous-state process** if X_t is a *continuous rv* $\forall t$
- **Time Dimension Classification:**
  - $X(t, w)$ is a **discrete-state process** if the set of time $\mathcal{T}$ is *discrete*
  - $X(t, w)$ is a **continuous-state process** if $\mathcal{T}$ is unbounded and thus *continuous*

7.1 Proprieties

7.1.1 Mean Function

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

Where $\mathbb{E}[X(t)]$ si the **expected value** of the rv for the fixed *time point* t

7.1.2 Variance Function

$$\sigma^2(t) = VAR[X(t)] = \mathbb{E}[(X(t) - \mu_x(t))^2] = \mathbb{E}[X^2(t)] - [\mu_x(t)]^2$$

7.1.3 Standard Deviation Function

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

7.1.4 Auto-Covariance Function

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

And it has the following proprieties:

- $C_{x,x}(t,s) = C_{x,x}(s,t)$
- $\sigma_x^2(t) = VAR[X(t)] = COV[X(t), X(t)] = C_{x,x}(t,t) = \mathbb{E}[X^2(t)] - u_x^2(t)$
- It is interpreted as the classic covariance

7.1.5 Auto-Correlation Function

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

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

7.2 Stationary and wide-sense stationary processes

7.2.1 Strongly / Strict-Sense Stationary

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 if 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, sd
- Since the **joint distribution** of $X(t_1)$ and $X(t_2)$ is invariant respect to its statistical proprieties over time (can be shifted over time with no changes in proprieties) or more shortly **translation invariant** also the **autocovariance** of $X(t)$ must be **translation invariant**
- Same invariance $\forall r \geq 1$

7.2.2 Weakly / Wide-Sense Stationary

A stochastic process $X(t)$ is **Weakly / Wide-Sense Stationary** if the following two conditions holds:

1. The **Mean Function** of $X(t), X(s)$ is *constant*, thus:

$$u(t) \xrightarrow[t\to\infty]{} \mu$$

2. The **Auto-Covariance Function** of $X(t), C_{xx}(t,s)$ *depends* only on $(s-t) = \tau$, thus:

$$\sigma(t, t+h) \xrightarrow[t\to\infty]{} \sigma(h) \quad \text{depends only on the } \textit{distance}$$

7.3 Markov Processes

A stochastic process $X = \{X(t) : t \geq 0\}$ is **Markov** 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

Markov Propriety

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

7.3.1 Markov Chain

The **Markov Chain** is a stochastic process with **discrete space** and the **Markov Propriety**. From now all processes we will study will be **Markov Chain (continuous time)** unless otherwise started.

- **Conditional Probability Mass Function**

$$\mathcal{P}_{(s,t)}(X_s, X_t) = \mathbb{P}[X(t) = X_t | X(s) = X_s]$$

- **Time Homogeneous MC**

$$\mathbb{P}[X(t) = X_t | X(s) = X_s] = \mathcal{P}_h(X_s, X_t)$$

  Depends only on $h = t - s$

- $\mathcal{P}_h(i, j) \to \mathcal{P}_{ij}(h) = \mathbb{P}[X(h) = j | X(0) = i]$
- Set of distribution for X in a time dependent matrix

$$\mathcal{P}_h = \begin{bmatrix} P_{11}(h) & P_{12}(h) & \dots & P_{1n}(h) \\ P_{21}(h) & P_{22}(h) & \dots & P_{2n}(h) \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1}(h) & P_{n2}(h) & \dots & P_{nn}(h) \end{bmatrix}$$

- Each $P_{ij}(h)$ represent the probability of going **from state** i **to state** j in **time** h

- It is a stochastic matrix so, the row has sum up to 1 and each element is greater or equal to 0 $\forall ij, h \geq 0$

A stochastic process X is **counting** if $X(t)$ is the *number of items / arrivals* counted by the time t

Chapter 8

Poisson Process

Poisson process is a continuous-time counting stochastic process obtained from a Binomial counting process when its frame size Δ decreases to 0 while the arrival rate λ remains constant.

- **Marginal Distribution**

$$N(t) \sim Po(\lambda t) \quad t > 0 \quad N(0) = 1$$

- **Transition Probabilities**

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

- **Transition Increments**

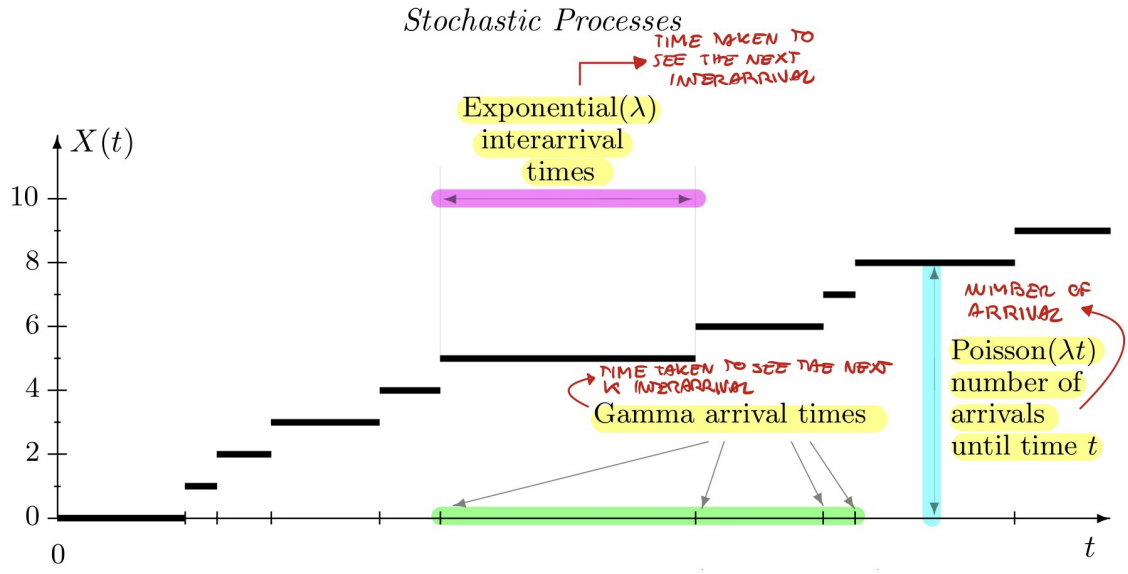
$$\begin{aligned} \mathcal{P}_{m,n+m}(h) &= \mathbb{P}[N(t+h) = n+m | N(t) = m] \\ &= \mathbb{P}[N(t+h-t) = n+m-m | N(0) = 0] \\ &= \mathbb{P}[N(h) = n] \\ &= \sum_{x=0}^{n} \frac{e^{-\lambda t}(\lambda t)^x}{x!} \text{ used also for } \mathbb{P}[N(h) \leq n] \end{aligned}$$

- **Independent Increments**

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

Poisson Process

- $X(t) = Poisson(\lambda t)$
- $T = Exponential(\lambda)$
- $T_k = Gamma(k, \lambda)$
- $\mathbb{P}[T_k \leq t] = \mathbb{P}[X(t) \geq k]$
- $\mathbb{P}[T_k > t] = \mathbb{P}[X(t) < k]$



8.1 Proprieties

8.1.1 Superposition

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

8.1.2 Thinning

$N \sim PP(\lambda)$ and each arrival assigned to process N_i with probability p_i independently from all others for $p_i \in (0,1)$ such that:

$$\sum_{i=1}^{\infty} p_i = 1 \qquad N_i \overset{\text{ind}}{\sim} PP(\lambda \cdot p_i)$$

8.1.3 Auto-covariance - $\sigma(s,t) = COV(N(s), N(t))$

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

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

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

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

Thus we obtain that the auto-covariance is equal to

$$\sigma(s,t) = COV(N(s), N(t)) = \lambda(min\{s,t\})$$

Chapter 9

CTMC

Transition Semi-group of the CTMC. The family $\{P_t : t \geq 0\}$ (set of matrices containing the conditional probability mass function) is the **transition semi-group** of the process.

1. $P_0 = I$
2. $\forall t > 0$ P_t is a *Stochastic Process*
   - Every entity is non negative
   - Sum of elements of each row is 1 by (law of total probability)
3. **Chapman-Kolmogorov Equation** $P_{s+t} = P_s \cdot P_t$
   - The **Stochastic Semi-group** $\{P_t\}$ with the **distribution of the initial point** $X(0)$ determine the **behaviour of the process** X

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

Standard Semi-Group. The **Semi-Group** $\{P_t\}$ standard if in addition to the previous 3 properties we have also that:

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

The Generator. Time $t \geq 0$ the chain X is in state i. Let $h \geq$ small and consider what could happened in the small time interval $(t, t+h)$

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

There exist constants $\{g_{ij} : i, j \in S\}$ st 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 **G** is the **Infinitesimal Generator Matrix** it takes the role of the **Transition Matrix** P over *discrete-time chains*, thus we can construct a **path** of X by moving iteratively for sufficient small h:

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

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

G is a matrix with **rows adding to 0** and **non-negative value outside the diagonal**

$$g_{ii} = -\sum_{j \ neqi} g_{ij}$$

The **Generator of a Process** is therefore a square matrix with:

- The **Instantaneous Exit Rates** on the *diagonal*
- The **Instantaneous Transition Rates** *outside the diagonal*

9.1 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 $S = \{0, 1, 2, ...\}$ st:

1. Is positive and non decreasing: $N(0 > 0)$ 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, $\forall s < t$

9.2 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**

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

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

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

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

- **Backward**

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

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

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

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

9.3 Matrix Exponential

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

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

Forward and **Backward** equations are the same:

$$P_t = e^{t \cdot G}$$

9.3.1 Matrix Exponential and Diagonalization

A Square matrix A is **Diagonalizable** it it can be rewritten as $A = V \cdot D \cdot V^{-1}$

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

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 $A = V \cdot D \cdot V^{-1}$

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

e^D has $e^{t \cdot d_{ii}}$ in the diagonal and **zeros** everywhere else

9.4 Holding time & Transition Probabilities

If we want to simulate a **path** given $X(t) = i$ we need to know **how long it stay** *(holding time)* and **where it goes once it jumps** *(transition probabilities)*

9.4.1 Holding time

Total exit rate from state i is:

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

9.4.2 Transition Probabilities

$$P[\text{the chain jumps from } i \text{ to } j | \text{the chain jumps}] = \frac{p_{ij}(h)}{1 - p_{ii}(h)}$$

Or

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

9.5 Irreducible HCTMC

A HCTMC is **irreducible** if the probability of reaching state j from state i in time t is **positive**, $\forall i, j \in S$ and $t > 0$. With the graph representation we just need to verify there is **at least** one **path** from i to j $\forall i \neq j$

9.6 Birth-Death Process

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. $\lambda_i \geq 0, \quad \mu_i \geq 0, \quad \mu_0 = 0$

9.7 Stationary Distribution

The vector π is a **Stationary Distribution** of the chain X if:

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

The Stationary Distribution is also called **Steady State Distribution** 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 semi-group $\{P_t\}$. If there exists a stationary distribution π then it is **unique** $\forall i, j \in S$

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

9.7.1 Global Balance Equations

A Distribution π is the **stationary distribution** of a HCTMC with transition semi-group P_t iff:

$$\pi^t \cdot G = 0$$

The result can be interpreted as **flux in** and **flux out** of a given state $j \in S$. So a distribution π is the **stationary distribution** of an HCTMC with transition semi-group P_t iff it satisfy the **global balance equations:**

$$\pi_i \sum_{j:j \neq i} g_{ij} = \sum_{j:j \neq i} \pi_j \cdot g_{ji}$$

- 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**

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

  Applied by substituting one of the columns of the, matrix G with a row of ones and placing a one i the corresponding pace of the row vector 0.

- $Ax = b$ where:
  - A, b are matrix and vector of known constants
  - x vector for which we wish to find

  It is therefor convenient to write:

  - $A = \tilde{G}^T$ where $\tilde{G}$ is obtained by substituting the **last column** of G with a column of **once**

- $b = e_N \in \mathbb{R}^n$ the n-th **canonical vector** where only the **last element** is 1 and the others are 0