

# Complementi di analisi e probabilità

Appunti delle lezioni di  
Paolo Alfano



## Note Legali

*Complementi di analisi e probabilità*

è un'opera distribuita con Licenza Creative Commons

Attribuzione - Non commerciale - Condividi allo stesso modo 3.0 Italia.

Per visionare una copia completa della licenza, visita:

<http://creativecommons.org/licenses/by-nc-sa/3.0/it/legalcode>

Per sapere che diritti hai su quest'opera, visita:

<http://creativecommons.org/licenses/by-nc-sa/3.0/it/deed.it>

### **Liberatoria, aggiornamenti, segnalazione errori:**

Quest'opera viene pubblicata in formato elettronico senza alcuna garanzia di correttezza del suo contenuto. Il documento, nella sua interezza, è opera di

Paolo Didier Alfano

e viene mantenuto e aggiornato dallo stesso, a cui possono essere inviate eventuali segnalazioni all'indirizzo *paul15193@hotmail.it*

Ultimo aggiornamento: 18 febbraio 2018

# Indice

<b>1</b>	<b>Introduzione ai segnali</b>	<b>6</b>
<b>2</b>	<b>Filtri</b>	<b>7</b>
2.1	Norme e continuità . . . . .	7
2.2	Numeri complessi . . . . .	10
2.3	Funzione di trasferimento . . . . .	12
2.4	Un esempio: il circuito RC . . . . .	13
2.5	Convoluzione . . . . .	14
2.6	Spazi coincidenti quasi ovunque . . . . .	17
2.7	Ancora su RC . . . . .	17
<b>3</b>	<b>Trasformata di Fourier</b>	<b>18</b>
3.1	Spazi di Banach e delta di Dirac . . . . .	18
3.2	Trasformata di Fourier . . . . .	21
3.3	Alcune proprietà riguardanti Fourier . . . . .	26
3.4	Convoluzione e trasformata di Fourier . . . . .	27
3.5	Realizzabilità di un filtro . . . . .	28
3.6	Serie di Fourier . . . . .	31
<b>4</b>	<b>Trasformata di Gabor</b>	<b>32</b>
4.1	Definizione e significato intuitivo . . . . .	32
4.2	Principio di indeterminazione e ampiezza della finestra . . . . .	33
<b>5</b>	<b>Introduzione alla probabilità</b>	<b>36</b>
5.1	Eventi . . . . .	36
<b>6</b>	<b>Variabili aleatorie discrete</b>	<b>38</b>
<b>7</b>	<b>Variabili aleatorie continue</b>	<b>40</b>
<b>8</b>	<b>Distribuzioni congiunte</b>	<b>44</b>
8.1	Funzione generatrice dei momenti . . . . .	48
<b>9</b>	<b>Attese condizionate</b>	<b>48</b>
<b>10</b>	<b>Catene di Markov discrete</b>	<b>55</b>
10.1	Processi stocastici . . . . .	55
10.2	Processi markoviani . . . . .	56
10.3	Equazione di Chapman Kolmogorov . . . . .	59
10.4	Rappresentare le catene di Markov . . . . .	60
10.5	Stabilizzazione . . . . .	64

<b>11 Distribuzione esponenziale</b>	<b>67</b>
11.1 Distribuzione esponenziale e memoria . . . . .	67
11.2 Proprietà della distribuzione esponenziale . . . . .	70
<b>12 Processi di conteggio e di Poisson</b>	<b>71</b>
12.1 Processi di conteggio . . . . .	71
12.2 Processi di Poisson . . . . .	73
12.3 Alcune proprietà su Poisson . . . . .	76
<b>13 Catene di Markov continue</b>	<b>77</b>
13.1 Definizioni . . . . .	77
13.2 Funzione probabilità di transizione . . . . .	81
13.3 Stabilizzazione . . . . .	84

## Struttura del documento

Il corso è strutturato in due parti, una di analisi e una di probabilità. Le lezioni a riguardo saranno separate in due parti distinte. Nella prima parte mostreremo il contenuto delle lezioni di analisi. Nella seconda parte il contenuto delle lezioni di probabilità.

# Complementi di Analisi

27 / 09 / 2017

## 1 Introduzione ai segnali

Nel seguito diremo che un segnale è una funzione del tipo  $y = f(x)$  (anche se sovente nel seguito una funzione verrà indicata con il simbolo  $x = x(t)$  e talvolta addirittura solo con  $x$ ).

Abbiamo tre tipi di segnale:

1. Segnale continuo: gli elementi  $x$  da considerare appartengono ad un intervallo  $I$  che è incluso in  $\mathbb{R}$ . Dunque:  $x \in I \subseteq \mathbb{R}$
2. Segnale discreto: gli elementi appartengono ad un dominio discreto. Dunque  $x \in I \subseteq \mathbb{Z}$  oppure  $x \in I \subseteq \mathbb{N}$
3. Segnale finito:  $x \in I$  dove  $I$  è un insieme finito. Questo è l'unico caso reale.

Nel seguito useremo molto spesso un operatore, detto *sistema* che presa una funzione in input, restituisce una funzione in output. Ovvero:

$$x = x(t) \rightarrow \text{SISTEMA } A \rightarrow y = y(t)$$

Solitamente faremo riferimento al sistema  $A$  con alcune notazioni equivalenti:

$$x \xrightarrow{A} y$$
$$y = A(x)$$

Riportiamo di seguito alcuni esempi di sistemi:

- Continui: alcuni esempi di sistemi continui sono
  - Amplificatore: il sistema dell'amplificatore è  $y(t) = kx(t)$  ovvero prende una funzione  $x$  e ne "amplifica" il valore di un fattore  $k$ .
  - Ritardatore: il sistema di un ritardatore (delayer) è  $y(t) = x(t - c)$  dove  $c > 0$
  - Differenziatore:  $y(t) = x'(t)$
- Discreti: vediamo invece alcuni esempi di sistemi discreti
  - Circuito RC:  $v'(t) + v(t) = x(t)$ . In questo caso bisogna risolvere l'equazione differenziale.

Notiamo che la trasformazione del segnale non è sempre reversibile.

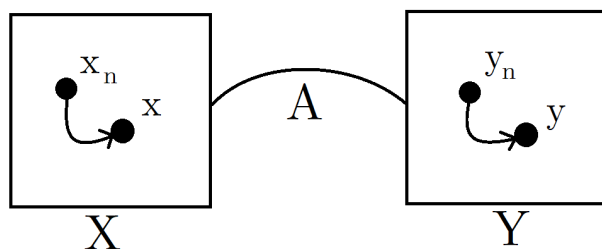


Figura 1: Due spazi di funzioni messi in comunicazione dall'operatore  $A$

03 / 10 / 2017

## 2 Filtri

Abbiamo visto nella sezione precedente l'operatore  $A$ . Tale operatore prende funzioni in input e restituisce funzioni in output. Cerchiamo di vederlo un po' più nel dettaglio.

$$A : x(t) \rightarrow y(t) \quad (1)$$

Può possedere le seguenti proprietà:

1. Linearità
2. Causalità: se  $x(t) = 0 \forall t \leq t_0 \Rightarrow (Ax)(t) = y(t) = 0 \forall t \leq t_0$   
Notiamo che l'ordine delle parentesi in  $(Ax)(t)$  è giusto poiché  $(Ax)$  restituisce una funzione che poi applicheremo a  $t$ .
3. Invarianza per traslazione:  $y = Ax \Leftrightarrow \tau_a y = A(\tau_a x)$ . Notiamo che anche  $\tau_a$  è un operatore come  $A$
4. Continuità: intuitivamente il concetto è che se gli input sono vicini, anche gli output devono esserlo. Ci dedicheremo con molta attenzione alla continuità tra poco.

Solitamente chiamiamo:

- Filtro: un operatore su cui valgano le proprietà 1) 3) 4)
- Filtro causale: un operatore su cui valgano le proprietà 1) 2) 3) 4)

### 2.1 Norme e continuità

Vediamo in maggiore dettaglio la continuità. Possiamo dare una descrizione qualitativa a riguardo. Supponiamo di avere i due spazi di funzioni mostrati in Figura 1.

Possiamo dire che se  $x_n \rightarrow x$  in  $X$  allora  $y_n \rightarrow y$  in  $Y$ . Dove il simbolo  $\rightarrow$  può essere letto come "si avvicina a", oppure "tende a"

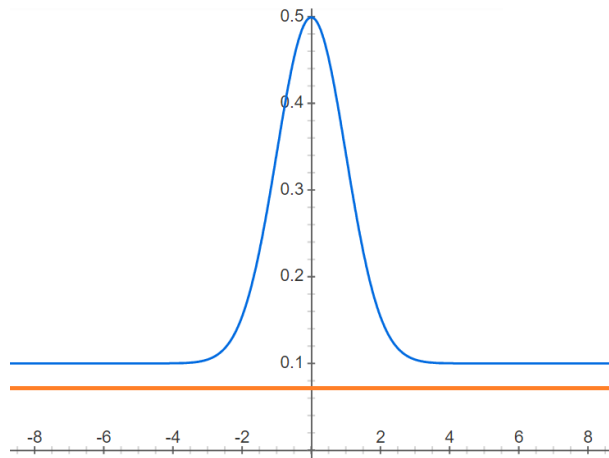


Figura 2: Le due funzioni dell'esempio 1

Dobbiamo stabilire cosa intendiamo per "avvicinarsi". La vicinanza può essere stabilita andando a definire un criterio per calcolare la distanza. Allora potremo dire che  $x_n$  si avvicina ad  $x$  in  $X$  se la loro distanza fissata in  $X$  tende a 0 per  $n$  che tende ad  $\infty$ . Allo stesso modo possiamo procedere nello spazio di funzioni  $Y$ .

Resta solo da definire quali "distanze" possiamo usare.

Sia  $I \subseteq \mathbb{R}$  un intervallo<sup>1</sup>. Sia  $X$  un certo spazio di funzioni, allora possiamo definire tre distanze:

1. Norma 1:  $\|x_1 - x_2\|_1 = \int_a^b |x_1(t) - x_2(t)|, dt$
2. Norma 2:  $\|x_1 - x_2\|_2 = (\int_a^b |x_1(t) - x_2(t)|^2, dt)^{1/2}$
3. Norma  $\infty$ :  $\|x_1 - x_2\|_\infty = \sup_{t \in I} |x_1(t) - x_2(t)|$

Possiamo notare che la norma 1 rappresenta l'area compresa tra le due funzioni, la norma 2 rappresenta l'area compresa tra le due funzioni ma elevata al quadrato, mentre la norma  $\infty$  rappresenta la massima distanza tra le due funzioni. Notiamo che questi tipi di "distanza" possono restituire valori molto diversi fra loro.

**Esempio 1** : consideriamo le due funzioni mostrate in Figura 2

In questo caso la norma 1 e la norma 2 restituiscono un valore più piccolo rispetto a quello fornito dalla norma infinito. Infatti l'area compresa è piccola ma nel punto di massima distanza le due funzioni distano di più.

<sup>1</sup>Gli intervalli che useremo più di frequente sono  $I = [a, b]$ ,  $I = [-\infty, b]$ ,  $I = [a, +\infty]$ ,  $I = [-\infty, +\infty] = \mathbb{R}$



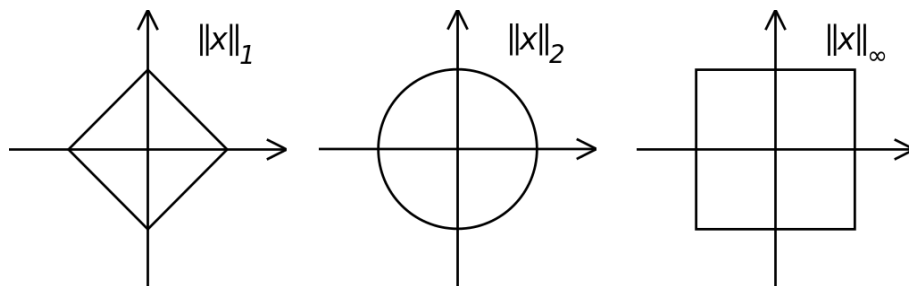


Figura 3: Rappresentazione in  $\mathbb{R}^2$  delle norme se eguagliate alla sfera unitaria

Notiamo che esiste una corrispondenza tra queste norme appena viste e alcuni concetti di geometria ben noti in uno spazio vettoriale come  $\mathbb{R}^2$ , ovvero il piano cartesiano. Ad esempio, ricordando che un integrale altro non è che una somma su un numero infinito di elementi, possiamo notare che la norma due corrisponde alla distanza euclidea.

Inoltre le norme possono esser rappresentate nel piano come in Figura 3 se le eguagliamo alla sfera unitaria.

Nel complesso diremo che, considerata come distanza la norma 1, se  $\lim_{n \rightarrow \infty} \|x_n - x\|_1 = 0$  allora diremo che  $x_n$  tende ad  $x$  per  $n$  che tende ad infinito. Ovviamente la regola che qui è definita per la norma 1 può anche essere definita per le altre norme.

Visto che abbiamo dato la definizione di distanza, possiamo dare la definizione di continuità.

Diremo che<sup>2</sup>

$$A : (X, \|\dots\|_X) \rightarrow (Y, \|\dots\|_Y)$$

è continua se vale che

$$\lim_{n \rightarrow \infty} \|x_n - x\|_X = 0 \Rightarrow \lim_{n \rightarrow \infty} \|y_n - y\|_Y = 0 \quad (2)$$

dove  $y_n = Ax_n$  e  $y = Ax$ .

Come verificare la validità della formula di continuità fornita poco fa? In generale non è facile, ma sfruttando le proprietà di  $A$ , in certi casi diventa facile.

Supponiamo di voler verificare la continuità del seguente operatore  $A$

$$A : (X, \|\dots\|_1) \rightarrow (Y, \|\dots\|_1)$$

Ovvero vogliamo verificare la continuità quando sia lo spazio di input che quello di output utilizzano come distanza la norma 1.

Grazie alle proprietà di  $A$  sappiamo che dobbiamo verificare

$$\exists M \geq 0 \forall x \in X : \|Ax\|_1 \leq M \|x\|_1$$

Cerchiamo di dimostrarlo.

<sup>2</sup>La notazione  $\|\dots\|_X$  indica la norma utilizzata nello spazio di funzioni  $X$

Dalle ipotesi sappiamo che  $\lim_{n \rightarrow \infty} \|x_n - x\|_1 = 0$ .

Dunque:

$$\begin{aligned} \lim_{n \rightarrow \infty} \|Ax_n - Ax\|_1 &= \lim_{n \rightarrow \infty} \|A(x_n - x)\|_1 \leq \\ &\leq \lim_{n \rightarrow \infty} M \|x_n - x\|_1 = M \underbrace{\lim_{n \rightarrow \infty} \|x_n - x\|_1}_0 \end{aligned}$$

Vediamo un esempio su cui verificare la continuità:

**Esempio 2** : sia  $Ax(t) = 5x(t)$

A è continuo? Ovvero, vale che:

$$A : (X, \|\dots\|_1) \rightarrow (Y, \|\dots\|_1)$$

Vale soltanto nel caso in cui:

$$\exists M \geq 0 \forall x \in X : \|Ax\|_1 \leq M \|x\|_1.$$

Partiamo sempre dalla norma dello spazio di arrivo:

$$\|Ax\|_1 = \int_{-\infty}^{+\infty} |5x(t)| dt = 5 \underbrace{\int_{-\infty}^{+\infty} |x(t)| dt}_{\|\dots\|_1} = 5 \|x\|_1$$

**Esempio 3** : dimostrare invece che una certa proprietà di continuità non vale è più difficile. Supponiamo ad esempio di voler verificare che:

$$A : (X, \|\dots\|_1) \rightarrow (Y, \|\dots\|_2)$$

Questa è una dimostrazione più complicata perché devo dimostrare che:

$$\forall M \geq 0 \exists x \in X : \|Ax\|_1 \geq M \|x\|_2$$

Notiamo che nel caso in cui le funzioni da considerare siano discrete possiamo ripetere le dimostrazioni in modo analogo, solo che al posto degli integrali avremo le sommatorie. Riportiamo a titolo di esempio la norma 1 per funzioni discrete:

$$\|x' - x\|_1 = \sum_{k \in I} \|x'_k - x_k\|$$

## 2.2 Numeri complessi

L'argomento della sezione successiva è la funzione di trasferimento. Per poterne parlare dobbiamo fare un veloce ripasso riguardo i numeri complessi.

Un numero complesso  $z$  è un punto avente due componenti:

$$z = (x, y) = x + iy$$

dove  $x, y \in \mathbb{R}$  mentre  $i = \sqrt{-1}$  viene detta *unità immaginaria*.

Sui numeri complessi possiamo definire alcune operazioni. Siano  $z = (x, y)$  e  $w = (u, v)$  con  $z, w \in \mathbb{C}$  allora vale che:

$$z + w = (x + u, y + v)$$

$$z \cdot w = (x + iy)(u + iv) = xu + ixv + iyu - yv \equiv (xu - yv, xv + yu)$$

La forma con cui abbiamo rappresentato i numeri complessi finora viene detta *forma algebrica*. Esistono altre due forme di rappresentazione dei numeri complessi.

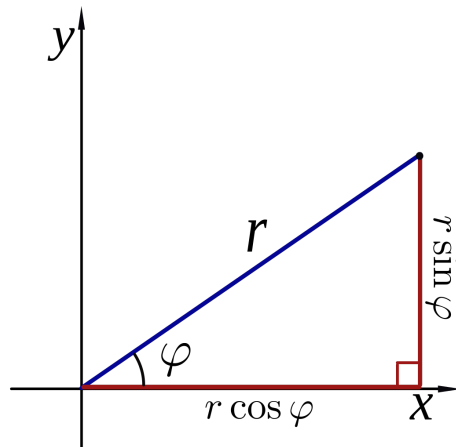


Figura 4: Rappresentazione trigonometrica di un numero complesso

La prima è la forma trigonometrica che da una visione geometrica dei numeri complessi. Essendo un numero complesso formato da due componenti, allora possiamo rappresentarlo nel piano come in Figura 4

Notiamo che questa rappresentazione coincide con quella delle *coordinate polari*. Infatti vale che, dato un numero complesso  $z = (x, y)$

$$z = \begin{cases} x = r \cos(\varphi) \\ y = r \sin(\varphi) \end{cases}$$

L'ultima forma di rappresentazione per i numeri complessi è la *forma esponenziale*. Un numero complesso può essere rappresentato come

$$e^{i\theta} = \cos\theta + i \sin\theta$$

infatti moltiplicando tutto per  $r$  otteniamo

$$r e^{i\theta} = r \cos\theta + r i \sin\theta$$

Notiamo che la forma esponenziale è molto comoda per calcolare prodotti tra numeri complessi. Infatti siano

$$z = (x, y) = r_1 e^{i\theta_1}$$

$$w = (u, v) = u + iv = r_2 e^{i\theta_2}$$

allora:

$$zw = r_1 e^{i\theta_1} r_2 e^{i\theta_2} = r_1 r_2 e^{i(\theta_1 + \theta_2)}$$

Quindi è sufficiente moltiplicare il modulo e sommare gli angoli di ogni numero per ottenerne il prodotto.

Notiamo che di conseguenza l'elevamento a potenza è molto comodo:

$$z^n = r_1^n e^{i n \theta_1}$$

I numeri complessi della forma

$$e_{\lambda}(t) = e^{2\pi i \lambda t} = \cos(2\pi \lambda t) + i \sin(2\pi \lambda t)$$

vengono detti *frequenze pure* perché oscillano secondo un multiplo dell'angolo giro.

Possiamo anche definire delle *funzioni complesse* della forma  
 $f : t \in \mathbb{R} \rightarrow x(t), y(t)$

04 / 10 / 2017

### 2.3 Funzione di trasferimento

Il nostro scopo è di costruire un *filtro passa basso*, ovvero una funzione che per basse frequenze mantiene un segnale o lo amplifica, mentre impedisce alle alte frequenze il passaggio. Quello che scopriremo è che applicando l'operatore  $A$  alle frequenze pure, otterremo un multiplo delle stesse. Ovvero:

$$Ae_\lambda(t) = H(\lambda)e_\lambda(t)$$

Supponendo che  $H(\lambda)$  sia un funzione del tipo:

$$H(\lambda) = \begin{cases} 3, & \text{se } \lambda < 10.000 \\ 0, & \text{altrimenti} \end{cases}$$

Allora quello che otteniamo è che per frequenze basse (ovvero  $\lambda$  ridotti, minori di 10.000) il valore di  $e_\lambda(t)$  viene "amplificato" ovvero moltiplicato per 3. Altrimenti, per frequenze alte, viene moltiplicato per 0. Questo è dunque il senso del filtro passa basso: basse frequenze vengono amplificate, alte frequenze vengono azzerate.

Per cercare di costruire tutto questo dobbiamo partire dall'assunzione che esista uno spazio di funzioni  $X$  e che valga che  $\forall \lambda e_\lambda \in X$

Consideriamo allora la costante complessa  $c_n$  e tutti gli  $x(t)$  tali che  $\forall x(t) \in X$  della forma:

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n e_\lambda(t)^n = \sum_{n=-\infty}^{+\infty} c_n e^{2\pi i n \lambda t}$$

In seguito vedremo che i segnali periodici possono essere scritti nella forma di  $x(t)$ . Dunque partiamo proprio da questa forma.

Consideriamo adesso un operatore  $A$  che riceve in input funzioni complesse e restituisce in output funzioni complesse. Allora consideriamo:

$$\begin{aligned} Ax(t) &= A\left(\sum_{n=-\infty}^{+\infty} c_n e^{2\pi i n \lambda t}\right) = A\left(\lim_{N \rightarrow +\infty} \sum_{n=-N}^N c_n e^{2\pi i n \lambda t}\right) = \\ &= \lim_{N \rightarrow +\infty} \sum_{n=-N}^N A(c_n e^{2\pi i n \lambda t}) = \lim_{N \rightarrow +\infty} \sum_{n=-N}^N c_n A(e^{2\pi i n \lambda t}) = \sum_{n=-\infty}^{+\infty} c_n A(e^{2\pi i n \lambda t}) \end{aligned}$$

Dopo tutti questi passaggi abbiamo scoperto che, quando applico la  $A$  ad un segnale periodico, quello che conta nell'applicazione di  $A$  è il valore della frequenza fondamentale  $\lambda$ .

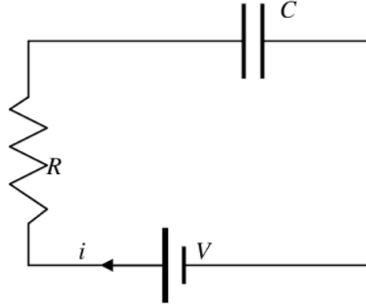


Figura 5: Circuito RC

Prima di procedere notiamo che

$$e_\lambda(t+u) = e^{2\pi i \lambda(t+u)} = e^{2\pi i \lambda t} \cdot e^{2\pi i \lambda u} = e_\lambda(t) e_\lambda(u)$$

Adesso poniamo:

$$f_\lambda(t) = A e_\lambda(t)$$

E applichiamo:

$$f_\lambda(t+u) = A(e_\lambda(t+u)) = A(e_\lambda(t)e_\lambda(u)) = e_\lambda(u)f_\lambda(t)$$

Consideriamo il caso in cui  $t = 0$ . Abbiamo che  $f_\lambda(t+u) = \underbrace{f_\lambda(0)}_{H(\lambda)} e_\lambda(u) \forall u$

Dunque il sistema si comporta così:  $f_\lambda = A e_\lambda = H(\lambda) e_\lambda$

Questo significa che le frequenze  $\lambda$  date in input vengono amplificate o smorzate a seconda di  $\lambda$ .

## 2.4 Un esempio: il circuito RC

Cerchiamo di applicare quanto visto nella sezione precedente ad un esempio, il circuito RC. Mostriamo in Figura 5 un tipico circuito RC.

Ovviamente con  $R$  si intende la resistenza, con  $C$  si intende il condensatore mentre con  $V$  si intende la tensione in input.

Questo esempio è particolarmente interessante perché ad un segnale dato in input al circuito, il circuito restituisce un segnale in output.

L'equazione differenziale che regola un circuito è la seguente:

$$RCv'(t) + v(t) = x(t)$$

Iniziamo col porre  $w(t) = v(t)e^{t/RC}$ . Ovvero

$$v(t) = w(t)e^{-t/RC}.$$

Adesso calcoliamo  $v'(t)$ :

$$v'(t) = w'(t)e^{-t/RC} + w(t)\left(\frac{-1}{RC}\right)e^{-t/RC} = e^{-t/RC}\left(w'(t) - \frac{1}{RC}w(t)\right)$$

Possiamo riprendere l'equazione del circuito RC e andare a sostituire

$$\begin{aligned} RC e^{-t/RC}\left(w'(t) - \frac{1}{RC}w(t)\right) + w(t)e^{-t/RC} &= x(t) \Leftrightarrow \\ \Leftrightarrow RC e^{-t/RC}w'(t) \underbrace{-e^{-t/RC}w(t) + w(t)e^{-t/RC}}_{=0} &= x(t) \Leftrightarrow \\ \Leftrightarrow w'(t) &= \frac{1}{RC}e^{t/RC}x(t) \end{aligned}$$

Ma allora:

$$w(t) = \int_{-\infty}^t \frac{1}{RC}e^{s/RC}x(s), ds$$

E poiché  $w(t) = v(t)e^{t/RC}$  allora:

$$v(t) = e^{-t/RC} \frac{1}{RC} \int_{-\infty}^t e^{s/RC}x(s), ds = \frac{1}{RC} \int_{-\infty}^t e^{(s-t)/RC}x(s), ds$$

A questo punto, andando a definire la funzione:

$$u(t-s) = \begin{cases} 1, & \text{se } -\infty < s < t \\ 0, & t \leq s \end{cases}$$

Allora possiamo scrivere che:

$$\frac{1}{RC} \int_{-\infty}^t e^{(s-t)/RC}x(s), ds = \frac{1}{RC} \int_{-\infty}^t u(t-s)e^{(s-t)/RC}x(s), ds$$

A questo punto possiamo introdurre la funzione  $g(t-s) = u(t-s)e^{(s-t)/RC}$ .  
Dunque abbiamo che:

$$\frac{1}{RC} \int_{-\infty}^t u(t-s)e^{(s-t)/RC}x(s), ds = \frac{1}{RC} \int_{-\infty}^t g(t-s)x(s), ds$$

In particolare,  $\int_{-\infty}^t g(t-s)x(s), ds$  viene detto *convoluzione*.

La convoluzione è indicata con il simbolo  $*$ .

## 2.5 Convoluzione

Cerchiamo quindi di riassumere. Che cosa abbiamo fatto? Fondamentalmente abbiamo espresso la tensione in uscita  $v(t)$  come prodotto della funzione di entrata  $x(t)$  e di una certa  $g$  avente caratteristiche da noi scelte.

Cerchiamo invece di dare una definizione più accurata di convoluzione.  
Diremo che la convoluzione fra  $h$  ed  $x$  è:

$$(h * x)(t) = \int_{-\infty}^{+\infty} h(t-s)x(s), ds$$

È abbastanza facile dimostrare che l'operazione di convoluzione possiede alcune proprietà come la commutatività.

Riprendiamo il circuito RC. Consideriamo la funzione di input e quella di output. Di fatto il circuito RC opera da filtro. Quindi:

$$x(t) \rightarrow RC \rightarrow v(t)$$

Possiamo chiederci se questo filtro sia continuo sotto la norma infinito in input e output. Ovvero:

$$A : (X, \| \dots \|_{\infty}) \rightarrow (Y, \| \dots \|_{\infty})$$

Di fatto è continuo se e solo se

$$\exists M > 0 \forall x \in X : \| Ax \|_{\infty} \leq M \| x \|_{\infty}$$

Per dimostrarlo partiamo dal fatto che  $v = Ax$ :

$$\begin{aligned} \| Ax \|_{\infty} &= \| v \|_{\infty} = \sup_{t \in \mathbb{R}} |v(t)| = \sup_{t \in \mathbb{R}} |e^{-t/RC} \frac{1}{RC} \int_{-\infty}^t e^{s/RC} x(s), ds| = \\ &= \frac{1}{RC} \sup_{t \in \mathbb{R}} |e^{-t/RC} \int_{-\infty}^t e^{s/RC} x(s), ds| \leq \frac{1}{RC} \sup_{t \in \mathbb{R}} e^{-t/RC} \int_{-\infty}^t |e^{s/RC} x(s), ds| = \\ &= \frac{1}{RC} \sup_{t \in \mathbb{R}} e^{-t/RC} \int_{-\infty}^t e^{s/RC} |x(s)|, ds \leq \\ &\leq \frac{1}{RC} \sup_{t \in \mathbb{R}} e^{-t/RC} \int_{-\infty}^t e^{s/RC} \| x \|_{\infty}, ds = \\ &= \| x \|_{\infty} \frac{1}{RC} \sup_{t \in \mathbb{R}} e^{-t/RC} \int_{-\infty}^t e^{s/RC}, ds = \| x \|_{\infty} \frac{1}{RC} \sup_{t \in \mathbb{R}} e^{-t/RC} \left[ \frac{e^{s/RC}}{1/RC} \right]_{-\infty}^t \end{aligned}$$

Ma poiché  $e^{-\infty/RC} = 0$

$$\begin{aligned} \| x \|_{\infty} \frac{1}{RC} \sup_{t \in \mathbb{R}} e^{-t/RC} \left[ \frac{e^{s/RC}}{1/RC} \right]_{-\infty}^t &= \\ = \| x \|_{\infty} \sup_{t \in \mathbb{R}} e^{-t/RC} (e^{t/RC} - 0) &= \\ = \| x \|_{\infty} \end{aligned}$$

05 / 10 / 2017

Riprendiamo la definizione di convoluzione data precedentemente. Consideriamo il filtro che segue

$$Ag : f \rightarrow f * g$$

Questo filtro dipende da  $g$ . Possiamo verificare che sia un filtro andando a verificare le solite tre proprietà:

- $Ag$  è lineare: lo verifichiamo andando a vedere che

$$\begin{aligned} Ag(f_1 + f_2) &= \int_{-\infty}^{+\infty} (f_1(t-s) + f_2(t-s))g(s), ds = \\ &= \int_{-\infty}^{+\infty} f_1(t-s)g(s) + f_2(t-s)g(s), ds = \\ &= \int_{-\infty}^{+\infty} f_1(t-s)g(s), ds + \int_{-\infty}^{+\infty} f_2(t-s)g(s), ds \end{aligned}$$

- $Ag$  è invariante per traslazione: per esserlo deve valere che

$$Ag[\tau_a]f(t) = \tau_a Ag[f](t) = Ag[f](t-a)$$

Verifichiamolo:

$$\begin{aligned} Ag[\tau_a]f(t) &= (\tau_a f * g)(t) = \int_{-\infty}^{+\infty} f(t-s-a)g(s), ds = \\ &= (f * g)(t-a) = \\ &= Agf(t-a) \end{aligned}$$

- $Ag$  è continua: supponiamo di verificarlo quando la norma dello spazio di partenza e di arrivo sia  $\|\dots\|_\infty$

Dunque vogliamo dimostrare che

$$\exists M > 0 \forall f \in X : \| Agf \|_\infty \leq M \| f \|_\infty$$

Verifichiamolo:

$$\begin{aligned} \| Agf \|_\infty &= \sup_{t \in \mathbb{R}} |Agf(t)| = \sup_{t \in \mathbb{R}} \left| \int_{-\infty}^{+\infty} f(t-s)g(s), ds \right| \leq \\ &\leq \sup_{t \in \mathbb{R}} \int_{-\infty}^{+\infty} |f(t-s)g(s)|, ds \leq \sup_{t \in \mathbb{R}} \int_{-\infty}^{+\infty} \| f \|_\infty |g(s)|, ds = \\ &= \| f \|_\infty \underbrace{\sup_{t \in \mathbb{R}} \int_{-\infty}^{+\infty} |g(s)|, ds}_{=\|\dots\|_1} = \| f \|_\infty \underbrace{\| g \|_1}_{=M} \end{aligned}$$

Più in generale tutti gli operatori di convoluzione sono continui e questo è mostrato (senza dimostrazione) nella formula che segue:

$$\| f * g \|_r \leq \| f \|_p \| g \|_q$$

dove  $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$

Notiamo che, una volta fissati  $p$  e  $q$ , automaticamente fissiamo anche il valore di  $r$ . Un caso particolarmente interessante è quello in cui  $p = 1$  e  $q = 1$  perché in tal modo anche  $r = 1$ .



## 2.6 Spazi coincidenti quasi ovunque

Nel seguito capiterà di utilizzare, per comodità, degli spazi di funzioni della forma seguente:

$$\{f : \int_I |f(t)|, dt = \|f\|_1 < \infty\}$$

le funzioni appartenenti a questo spazio vengono dette *coincidenti quasi ovunque*, ovvero coincidono su tutti i punti ad eccezione di un insieme di misura nulla.

Questi spazi verranno identificati con la notazione

$$L^1(I)$$

**Esempio 1** : consideriamo il seguente spazio di funzioni

$$L^\infty(I) = \{f : I \rightarrow \mathbb{C} : \|f\|_\infty = \sup_{t \in I} |f(t)| < \infty\}$$

Consideriamo allora  $L^\infty((0, +\infty))$ . Consideriamo anche la funzione

$f(t) = \frac{1}{t}$ . Questa funzione non appartiene a  $L^\infty$  poiché l'estremo superiore di  $1/t$  non è  $< \infty$

**Esempio 2** : consideriamo invece la funzione  $f(t) = \sin(t)$ . In questo caso  $f(t) \in L^\infty(t)$

**Esempio 3** : consideriamo la funzione a tratti:

$$f(t) = \begin{cases} \frac{1}{t}, & \text{se } t = \frac{1}{n} \\ \sin(t), & \text{altrimenti} \end{cases}$$

Anche in questo caso  $f(t) \in L^\infty(I)$

Una proposizione interessante è che: sia  $g \in L^1(\mathbb{R})$  allora l'operatore  $Ag : f \in L^\infty(\mathbb{R}) \rightarrow f * g \in L^\infty(\mathbb{R})$  è continuo.

## 2.7 Ancora su RC

Riprendiamo l'equazione del circuito RC:

$$RCv'(t) + v(t) = x(t)$$

Dalla sottosezione 2.4 avevamo scoperto che:

$$\begin{aligned} x(t) &= e_\lambda(t) = e^{2\pi i \lambda t} \\ f_\lambda(t) &= A e_\lambda(t) = H(\lambda) e_\lambda(t) = H(\lambda) e^{2\pi i \lambda t} \\ f'_\lambda(t) &= H(\lambda) e^{2\pi i \lambda t} 2\pi i \lambda \end{aligned}$$

da cui:

$$\begin{aligned}
& RC H(\lambda) e^{2\pi i \lambda t} 2\pi i \lambda + H(\lambda) e^{2\pi i \lambda t} = e^{2\pi i \lambda t} \Leftrightarrow \\
& \Leftrightarrow RC H(\lambda) e^{2\pi i \lambda t} 2\pi i \lambda + H(\lambda) e^{2\pi i \lambda t} = e^{2\pi i \lambda t} \Leftrightarrow \\
& \Leftrightarrow H(\lambda) = \frac{1}{\underbrace{1}_{=a} + i \underbrace{2\pi \lambda RC}_{=b}}
\end{aligned}$$

Notiamo che, il nostro obiettivo era comprendere se il valore di una certa frequenza viene amplificata o attenuata. Purtroppo però  $H(\lambda)$  restituisce un numero complesso, che dunque ci dà una misura dell'amplificazione certamente, ma in uno spazio complesso! Dunque preferiamo calcolare il modulo di  $H(\lambda)$  che è uno scalare.

Al denominatore abbiamo un numero complesso, della forma  $x = a + ib$ . Ma poiché  $|x| = \sqrt{a^2 + (ib)^2} = \sqrt{x\bar{x}}$ , allora il modulo di  $H$  è un numero della forma:

$$|H(\lambda)| = \frac{1}{\sqrt{1 + \underbrace{4\pi^2 R^2 C^2}_{=c} \lambda^2}} = \frac{1}{\sqrt{1 + c\lambda^2}}$$

11 / 10 / 2017

### 3 Trasformata di Fourier

#### 3.1 Spazi di Banach e delta di Dirac

Riconsideriamo gli spazi coincidenti quasi ovunque che abbiamo analizzato nella sottosezione 2.6. Questi spazi vengono anche detti *spazi di Banach*. Di questi spazi, quello  $L^2(I)$  (dunque relativo alla norma due) possiede almeno un paio di proprietà per cui vale la pena considerarlo:

1. Motivo fisico: come già detto, rappresenta l'energia di un segnale. Molto importante ma non è una questione che considereremo adesso.
2. Motivo matematico: è originato da un prodotto scalare. Vediamo nel seguito cosa intendiamo e quali conseguenze ha questo fatto.

Dobbiamo fare alcune considerazioni prima di procedere. Analizziamo nuovamente un valore complesso della forma

$$z = (x, y) = x + iy$$

Consideriamo anche il suo coniugato complesso

$$\bar{z} = x - iy$$

Calcoliamo:

$$z\bar{z} = (x + iy)(x - iy) = x^2 - ixy + iyx - i^2 y^2 = x^2 + y^2$$

Dunque

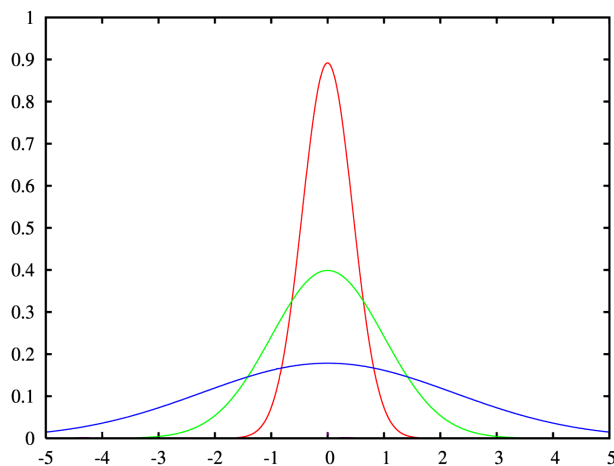


Figura 6: Tre possibili funzioni di densità di carica

$$|z| = \sqrt{x^2 + y^2} = \sqrt{z\bar{z}}$$

Notiamo che esiste una certa somiglianza tra il termine  $\sqrt{x^2 + y^2}$  e la norma 2. Infatti sono entrambe somme elevate al quadrato (la prima è una somma finita mentre la seconda è una somma infinita) poste sotto radice quadrata. Infatti il prodotto scalare in un campo complesso in cui si consideri la norma 2, viene definito come:

$$(f, g) = \left( \int_I f(t)g(\bar{t}), dt \right)^{1/2}$$

Il fatto che la  $|\dots|_2$  sia originata da un prodotto scalare, la rende la più maneggevole matematicamente e allo stesso tempo la rende la più significativa dal punto di vista fisico.

Consideriamo adesso un caso particolarmente interessante per introdurre il concetto della *delta di Dirac*.

Prendiamo in esame un filo la cui funzione di densità di carica  $\phi(x)$  sia qualcosa di simile a quanto mostrato in Figura 6

Notiamo che se concentriamo sempre più nella zona centrale la carica, otteniamo una funzione sempre più "concentrata" nell'origine. Il fatto che la carica rimane invariata può essere espresso tramite il seguente integrale:

$$Q = \int_{-\infty}^{+\infty} \phi(x), dx$$

Supponiamo di fare un'estremizzazione e andiamo a concentrare tutta la carica nell'origine. Di fatto otteniamo una funzione particolare  $\delta(x)$  come quella che segue:

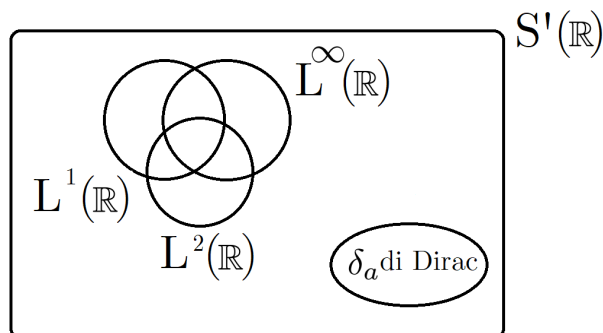


Figura 7: Rapporto tra gli spazi definiti finora

$$\delta(x) = \begin{cases} +\infty, & \text{se } x = 0 \\ 0, & \text{altrimenti} \end{cases}$$

Dunque abbiamo visto che per questa distribuzione di carica abbiamo dato una definizione  $\delta(x)$  e allo stesso tempo deve valere che la carica resta invariata. Dunque possiamo certamente affermare che  $\delta(x)$  non può corrispondere con la distribuzione di carica nulla.

Il problema è che per la definizione che abbiamo dato nella sottosezione 2.6 la funzione  $\delta(x)$  corrisponde proprio alla distribuzione di carica nulla visto che differisce soltanto su un insieme di misura nulla costituito dall'unico punto  $x = 0$ . Notiamo che otteniamo una situazione assurda, poiché in base ad una definizione abbiamo una normale distribuzione di carica ma invece in base ad un'altra definizione otteniamo la distribuzione di carica nulla.

Ne ricaviamo che l'oggetto  $\delta(x)$  è troppo particolare per essere associato alle funzioni. Infatti quella appena mostrata non è una funzione, bensì una *distribuzione*.

La distribuzione  $\delta(x)$  in particolare viene detta *delta di Dirac* quando è non nulla in zero. Altrimenti se è non nulla in un certo punto  $a$  viene detta:

$$\delta_a(x) = (\tau_a \delta)(x).$$

Lo spazio delle distribuzioni  $S'(\mathbb{R})$  è un sovrainsieme delle funzioni. Mostriamo questa situazione in Figura 7

Notiamo che sebbene esistano delle funzioni in comune tra i vari spazi  $L^1$ ,  $L^2$ ,  $L^\infty$ , la  $\delta$  di Dirac è del tutto disgiunta dagli spazi appena citati.

In generale possiamo considerare le distribuzioni come dei limiti di funzioni. Dobbiamo comunque considerare che essendo passati in uno spazio di maggiore generalità non è detto che le proprietà da noi dimostrate finora siano ancora valide. Fortunatamente è possibile estendere praticamente tutte le operazioni definite finora affinché siano valide anche in  $S'(\mathbb{R})$ . Le operazioni valide in  $S'(\mathbb{R})$  sono:

- Somma di funzioni e moltiplicazione di funzioni per costante
- Convoluzione sotto alcune condizioni

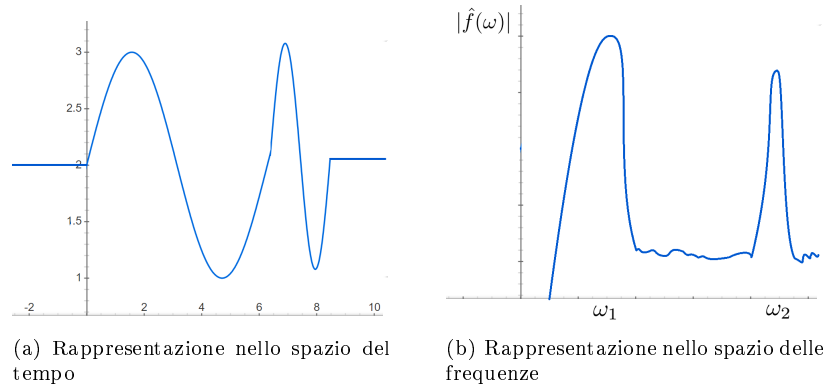


Figura 8: Applicazione della trasformata di Fourier

- Moltiplicazione puntuale di funzioni sotto alcune condizioni
- Traslazione
- Modulazione  $\mu_b$ . La modulazione è una funzione della forma  $\mu_a f(t) = e^{2\pi i a t} f(t)$
- Riflessione, ovvero  $\tilde{f}(t) = f(-t)$ . In particolare vale che  $\tilde{\delta}_a = \delta_{-a}$

### 3.2 Trasformata di Fourier

Andiamo a considerare finalmente la *trasformata di Fourier*. È un operatore che agisce su segnali della forma  $f(t)$ , dove con  $t$  intendiamo il tempo.

$$\mathcal{F}[f](\omega) = \hat{f}(\omega) = \int_{-\infty}^{+\infty} e^{-2\pi i t \omega} f(t), dt$$

con  $\omega \in \mathbb{R}$ .

Notiamo che il risultato dell'applicazione della trasformata di Fourier è una funzione complessa che porta dallo spazio di  $f(t)$  allo spazio delle frequenze  $\omega$ . Prima di procedere diamo una visione intuitiva della trasformata di Fourier tramite il seguente esempio:

**Esempio 1** : supponiamo di trovarci in una stanza in cui inizialmente abbiamo una situazione di silenzio. Successivamente viene emesso un suono avente frequenza  $\omega = \omega_1$ . Successivamente viene emesso un nuovo suono avente frequenza  $\omega = \omega_2$ . Infine, la situazione torna ad uno stato di quiete. Il grafico di questa situazione viene mostrato nella parte sinistra della Figura 8.

Quando applichiamo la trasformata di Fourier, otteniamo un grafico che presenta dei picchi in prossimità delle frequenze dei suoni emessi, ovvero

$\omega_1$  e  $\omega_2$ . È per questo che quando consideriamo la trasformata di Fourier diciamo che conduce dallo spazio dei tempi a quello delle frequenze.

Andiamo a vedere alcune proprietà di cui gode l'operatore  $\mathcal{F}$ :

- $\mathcal{F} : L^1(\mathbb{R}) \rightarrow L^\infty(\mathbb{R})$   
Abbiamo continuità e in particolare vale che  $\|\hat{f}\|_\infty \leq \|f\|_1$
- $\mathcal{F} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$   
Anche in questo caso vale la continuità, inoltre  $\mathcal{F}$  si comporta come un isomorfismo isometrico. Inoltre vale che  $\|\hat{f}\|_2 = \|f\|_2$  (uguaglianza di Plancherel). Questo non ci dovrebbe stupire particolarmente visto che abbiamo detto più volte in precedenza che la norma 2 può rappresentare l'energia di un segnale. E visto che la trasformata di Fourier semplicemente porta a una rappresentazione diversa del segnale (che però è sempre lo stesso!) l'energia del segnale dopo la trasformata non cambia.
- La trasformata è applicabile anche alle distribuzioni e non solo alle funzioni. Vale che:  
 $\mathcal{F} : S'(\mathbb{R}) \rightarrow S'(\mathbb{R})$   
dunque  $\mathcal{F}$  è funzione continua biettiva.

Delle proprietà che abbiamo precedentemente elencato, ci interesseremo in particolare della prima, visto che è l'unica che siamo in grado di trattare con gli strumenti attuali. Questo perché è più facile verificare proprietà di  $L^1$

Cerchiamo allora di verificarla:

$$\begin{aligned} \|\hat{f}\|_\infty &= \sup_{t \in \mathbb{R}} \left| \int_{-\infty}^{+\infty} e^{-2\pi i t \omega} f(t) dt \right| \leq \sup_{t \in \mathbb{R}} \int_{-\infty}^{+\infty} \underbrace{|e^{-2\pi i t \omega}|}_{=1} |f(t)| dt = \\ &= \sup_{t \in \mathbb{R}} \int_{-\infty}^{+\infty} |f(t)| dt = \|f\|_1 \end{aligned}$$

Un'altra proprietà interessante della trasformata è che quando viene applicata alla delta di Dirac, otteniamo la funzione uguale a 1. Vale anche il viceversa. Dunque vale che:

$$\begin{aligned} \hat{\delta} &= 1 \\ \hat{1} &= \delta \end{aligned}$$

Purtroppo non possiamo fornire una dimostrazione adeguata di questo poiché  $1 \in L^\infty$  e come dicevamo non abbiamo gli strumenti adeguati per trattare questo spazio. Ci accontentiamo delle due dimostrazioni intuitive che seguono:

$$\begin{aligned} \hat{1}(\omega) &= \int_{-\infty}^{+\infty} e^{2\pi i \omega t} f(t) dt = \int_{-\infty}^{+\infty} e^{2\pi i \omega t} 1 dt = \int_{-\infty}^{+\infty} \cos(2\pi \omega t) + i \sin(2\pi \omega t) dt = \\ &= \int_{-\infty}^{+\infty} \cos(2\pi \omega t) dt + i \int_{-\infty}^{+\infty} \sin(2\pi \omega t) dt \end{aligned}$$

Consideriamo i due casi che seguono: se  $\omega = 0$  il secondo integrale è nullo poiché  $\sin(0) = 0$ . Il primo non si annulla, e per essere precisi otteniamo:

$\int_{-\infty}^{+\infty} \cos(2\pi\omega t) dt + i \int_{-\infty}^{+\infty} \sin(2\pi\omega t) dt = \int_{-\infty}^{+\infty} 1 dt = +\infty$ . Nei casi dove invece  $\omega \neq 0$  possiamo accontentarci del fatto che essendo funzioni trigonometriche tutte le aree positive annullano tutte le aree negative (in modo molto informale) e dunque che:

$$\int_{-\infty}^{+\infty} \cos(2\pi\omega t) dt + i \int_{-\infty}^{+\infty} \sin(2\pi\omega t) dt = 0$$

Nel complesso vale che:

$$\hat{1} = \begin{cases} +\infty, & \text{se } \omega = 0 \\ 0, & \text{altrimenti} \end{cases}$$

Notiamo che questa è proprio la definizione di una  $\delta$  di Dirac centrata sull'origine. Verifichiamo invece che  $\hat{\delta} = 1$ .

Partiamo dal fatto che  $\hat{1} = \delta$  e andiamo a valutare  $\hat{\delta} = \widehat{\hat{1}}$ . Ma visto che  $\widehat{f(x)} = f(-x)$ , allora visto che in questo caso la nostra funzione è la funzione costante, allora vale che  $\hat{1} = 1$ . Dunque  $\hat{\delta} = 1$

Un'ulteriore proprietà essenziale è che  $\mathcal{F}$  commuta traslazioni in modulazioni. Ovvero:

- $\mathcal{F}[\tau_a f](\omega) = \mu_{-a} \hat{f}(\omega) \forall f \in S'(\mathbb{R})$
- $\mathcal{F}[\mu_a f](\omega) = \tau_a \hat{f}(\omega) \forall f \in S'(\mathbb{R})$

Proviamo a dimostrare la prima:

$$\begin{aligned} \mathcal{F}[\tau_a f](\omega) &= \int_{-\infty}^{+\infty} e^{-2\pi i t \omega} f(t-a), dt \\ \text{Applichiamo il cambio di variabile } s &= t-a \\ \int_{-\infty}^{+\infty} e^{-2\pi i t \omega} f(t-a), dt &= \int_{-\infty}^{+\infty} e^{-2\pi i (s+a) \omega} f(s), ds = e^{-2\pi i a \omega} \underbrace{\int_{-\infty}^{+\infty} e^{-2\pi i s \omega} f(s), ds}_{=\hat{f}(\omega)} \\ e^{-2\pi i a \omega} \hat{f}(\omega) &= \mu_{-a} \hat{f}(\omega) \end{aligned}$$

Consideriamo un altro aspetto interessante. Visto che le frequenze pure si concentrano in un solo valore di frequenza, ci aspettiamo che la trasformata di Fourier di una frequenza pura del tipo  $f(t) = e^{-2\pi i \lambda t}$  generi una  $\delta$  di Dirac traslata nella frequenza  $\lambda$ . Verifichiamolo:

$$\begin{aligned} f(t) &= e^{-2\pi i \lambda t} 1 = \mu_{\lambda} 1 \text{ dunque} \\ \hat{f}(u) &= \widehat{\mu_{\lambda} 1} = \tau_{\lambda} \hat{1} = \tau_{\lambda} \delta = \delta_{\lambda} \end{aligned}$$

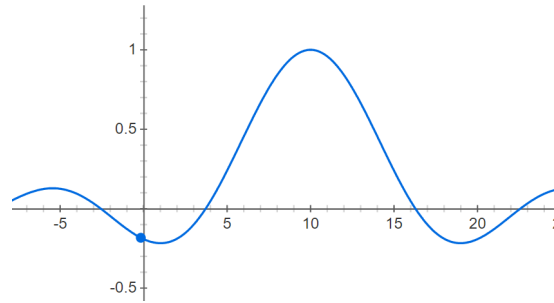


Figura 9: Rappresentazione nello spazio dei tempi del segnale  $e^{-t^2} + e^{2\pi iat}$

Dunque la trasformata di Fourier applicata ad una frequenza pura restituisce una delta di Dirac

18 / 10 / 2017

Generalizziamo questo concetto: sappiamo che un segnale può essere visto come somma di frequenze, ovvero:

$$f(t) = \sum_{k=1}^N c_k e^{-2\pi i a_k t}$$

Dunque applicandogli la trasformata di Fourier otteniamo:

$$\mathcal{F}(f(t)) = \mathcal{F}\left(\sum_{k=1}^N c_k e^{-2\pi i a_k t}\right) = \sum_{k=1}^N c_k \mathcal{F}(e^{-2\pi i a_k t}) = \sum_{k=1}^N c_k \delta_{a_k}$$

Dunque ogni segnale può essere visto come somma di  $N$  diverse delta di Dirac.

**Esempio 2** : supponiamo di avere il segnale mostrato in Figura 9

I segnale è dato dalla sovrapposizione dei due segnali:

$$\begin{matrix} e^{-t^2} \\ e^{2\pi iat} \end{matrix}$$

Essendo che il segnale è il prodotto di due parti, di cui una è una frequenza pura, ci aspettiamo che non venga proprio una delta di Dirac ma un qualcosa di più smussato, come quanto mostrato dalla rappresentazione della parte reale in Figura 10

Questo succede perché la frequenza pura è in qualche modo "rovinata" dalla presenza del segnale  $e^{-t^2}$

**Esempio 3** : andiamo a considerare il seguente segnale:



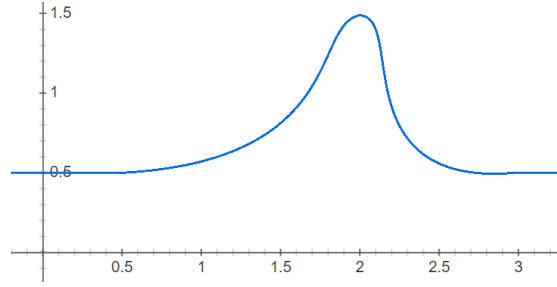


Figura 10: Rappresentazione nello spazio delle frequenze del segnale  $e^{-t^2} + e^{2\pi iat}$

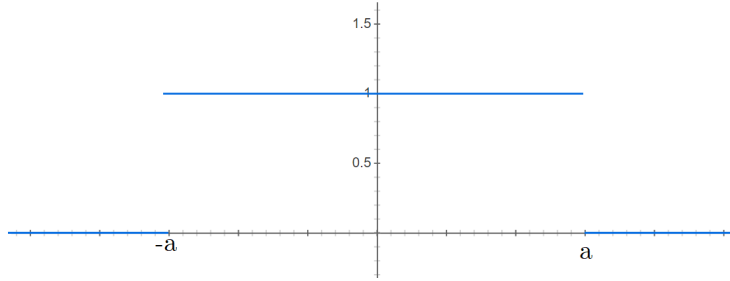


Figura 11: Rappresentazione nello spazio dei tempi del segnale  $\chi_{[-a,a]}$

$$\chi_{[-a,a]}(t) = \begin{cases} 1, & \text{se } t \in [-a, a] \\ 0, & \text{altrimenti} \end{cases}$$

la cui rappresentazione nello spazio dei tempi è mostrata in Figura 11

Assumiamo che  $\chi_{[-a,a]} \in L^1(\mathbb{R})$ .

$$\begin{aligned} \hat{\chi}_{[-a,a]}(\omega) &= \int_{-a}^a e^{-2\pi i \omega t} dt = \left[ \frac{e^{-2\pi i \omega t}}{-2\pi i \omega} \right]_{-a}^a = \frac{1}{-2\pi i \omega} (e^{-2\pi i \omega a} - e^{-2\pi i \omega (-a)}) = \\ &= \frac{1}{-2\pi i \omega} (\cos(-2\pi \omega a) + i \sin(-2\pi \omega a) - \cos(-2\pi \omega (-a)) - i \sin(-2\pi \omega (-a))) = \\ &= \frac{-2i \sin(2\pi \omega a)}{-2\pi i \omega} = \frac{\sin(2\pi \omega a)}{\pi \omega} \end{aligned}$$

La quantità trovata con l'esempio precedente viene solitamente indicata con il nome *seno cardinale*. Prendiamola come una definizione da ora in poi:

$$\text{sinc}_a(x) = \frac{\sin(\pi a x)}{\pi x}$$

Una sua rappresentazione con  $a = 1$  viene mostrata in Figura 12

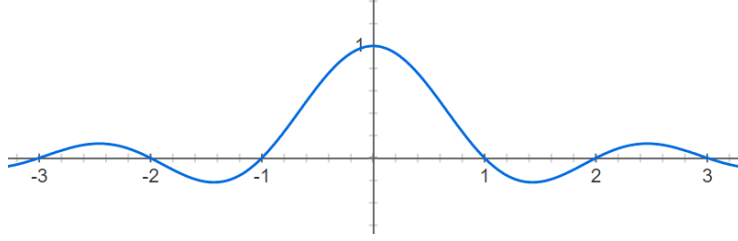


Figura 12: Rappresentazione del segnale seno cardinale per  $a = 1$

### 3.3 Alcune proprietà riguardanti Fourier

Consideriamo la seguente proposizione: se una certa funzione  $f(x) \in L^1(\mathbb{R})$  e  $\hat{f}(x) \in L^1(\mathbb{R})$  allora  $f(x)$  è uguale alla sua *antitrasformata*, ovvero:

$$f(t) = \int_{-\infty}^{+\infty} e^{-2\pi i \omega t} \hat{f}(\omega) d\omega = \int_{-\infty}^{+\infty} e^{-2\pi i \omega (-t)} \hat{f}(\omega) d\omega = \hat{f}(-t)$$

E dunque:

$$f(t) = \hat{\hat{f}}(-t)$$

Questa proprietà vale quasi ovunque per  $t \in \mathbb{R}$  e vale  $\forall f \in S'(\mathbb{R})$ .

Nel seguito useremo la notazione  $\mathcal{F}^{-1}$  per indicare l'antitrasformata e indicheremo con  $\mathcal{F}^n(f)(t)$  per indicare l'applicazione per  $n$  volte della trasformata di Fourier. Volendo possiamo riassumere la proprietà appena vista con la seguente notazione:

$$f(t) = \mathcal{F}^2(f)(-t)$$

Riprendiamo in considerazione la funzione  $\chi$  definita nell'esempio 3 della pagina precedente. Abbiamo che:

$$\widehat{\text{sinc}_{2a}}(t) = \hat{\chi}_{[-a,a]}(t) = \chi_{[-a,a]}(-t) = \chi_{[-a,a]}(t)$$

Dunque possiamo dire che:

$$\widehat{\text{sinc}_{2a}}(t) = \chi_{[-a,a]}(t)$$

Un'altra proprietà interessante è la *proprietà di scambio* della trasformata di Fourier. Infatti date due funzioni  $f$  e  $g$  vale che:

$$\widehat{f * g} = \hat{f} \cdot \hat{g}$$

e viceversa

$$\widehat{f \cdot g} = \hat{f} * \hat{g}$$

Ovvero la trasformata di una convoluzione da il prodotto puntuale delle trasformate. Questa proprietà vale  $\forall f, g \in S'(\mathbb{R})$  purché abbiano senso i prodotti puntuali (fatto non sempre vero in  $S'(\mathbb{R})$ )

### 3.4 Convoluzione e trasformata di Fourier

Consideriamo ancora una volta l'operatore di convoluzione

$$A : f \rightarrow f * h = Af$$

con  $h$  fissata. Andiamo a dimostrare la seguente e importantissima proposizione:

sia  $h \in L^1(\mathbb{R})$  e sia  $H$  la funzione di trasferimento precedentemente introdotta, allora vale che

$$\hat{h} = H$$

Ovvero stiamo dicendo che la funzione di trasferimento è la trasformata di Fourier della funzione  $h$  che definisce il filtro di convoluzione. Andiamo a dimostrarlo:

$$f * h = Af$$

ma per  $f = e_\lambda$  ho che

$$e_\lambda * h = Ae_\lambda = H(\lambda)e_\lambda$$

Consideriamo allora la trasformata dell'uguaglianza appena trovata:

$$\widehat{e_\lambda * h} = \widehat{H(\lambda)e_\lambda}$$

Consideriamo separatamente la parte sinistra dell'uguaglianza:

$$\widehat{e_\lambda * h} = \hat{e}_\lambda \cdot \hat{h} = \delta_\lambda \hat{h}$$

Consideriamo adesso la parte destra dell'uguaglianza, ricordandoci che  $H(\lambda)$  è di fatto una costante, sappiamo anche che

$$\widehat{H(\lambda)e_\lambda} = H(\lambda)\hat{e}_\lambda = H(\lambda)\delta_\lambda$$

Andiamo allora a considerare i due risultati ottenuti:

$$\delta_\lambda \hat{h} = H(\lambda)\delta_\lambda$$

Dunque:

$$(\hat{h} - H(\lambda))\delta_\lambda(\omega) = 0$$

Da cui (tramite un passaggio per noi non dimostrabile di cui ci fidiamo) otteniamo che

$$\hat{h} - H(\lambda) = 0$$

ovvero

$$\hat{h} = H$$

L'uguaglianza appena dimostrata è molto importante perché mette in relazione praticamente tutto quello che abbiamo visto finora, visto che lega la convoluzione con la funzione di trasferimento tramite la trasformata di Fourier.

Visto che la funzione  $h$  sembra avere una certa importanza, diamole un nome più significativo. Solitamente tale funzione viene detta *risposta impulsiva del filtro*. Questo nome è originato dal fatto che, avendo definito  $Af = f * h$ , considerando  $f = \delta$  otteniamo che:

$$A\delta = \delta * h = h$$

Notiamo che il passaggio  $\delta * h = h$  non è da noi dimostrabile, dobbiamo fidarci.

E visto che  $\delta$  è un impulso nell'origine, l'applicazione del filtro  $A$  a  $\delta$  ne determina la risposta, che viene detta per l'appunto risposta impulsiva del filtro.

Un'ultima proposizione interessante è quella che segue. Consideriamo:

$$Af = f * h = \mathcal{F}^{-1} \mathcal{F}(f * h) = \mathcal{F}^{-1}[\hat{f} \cdot \hat{h}] = \mathcal{F}^{-1}[\hat{f} \cdot H]$$

Consideriamo il termine  $\mathcal{F}^{-1}[\hat{f} \cdot H]$ . È particolarmente interessante perché ha una evidente corrispondenza intuitiva con i grafici mostrati in Figura 13.

In questa figura si suppone di aver applicato un filtro passa basso, dunque nella prima immagine mostriamo il valore di  $f$  nello spazio dei tempi. Nella seconda mostriamo  $\hat{f}$  nello spazio delle frequenze, nella terza abbiamo il segnale formato dal prodotto  $\hat{f}H$  sempre nello spazio delle frequenze, e infine torniamo nello spazio dei tempi con la quarta immagine in cui possiamo osservare che la frequenza maggiore è stata effettivamente tagliata dal filtro passa basso.

25 / 10 / 2017

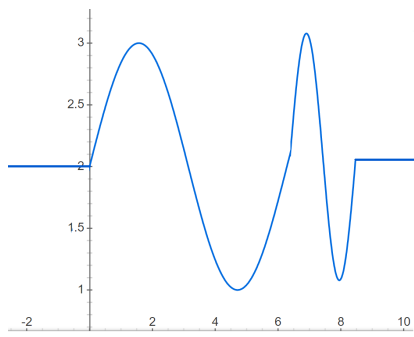
### 3.5 Realizzabilità di un filtro

Vediamo adesso un teorema che esplica la realizzabilità di un filtro causale.

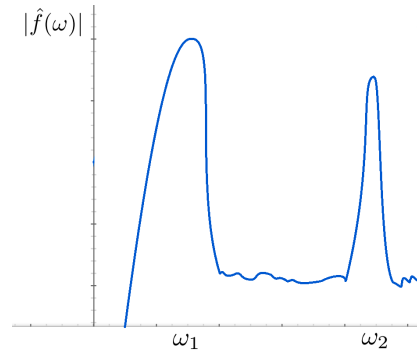
**Teorema 3.1.** *Il filtro  $Af = f * h$  è realizzabile (è causale) se e solo se  $\text{supp } h \subseteq [0, +\infty]$ , ovvero se  $h(t) = 0, \forall t < 0$*

*Dimostrazione.* ( $\Rightarrow$ ) sia  $A$  realizzabile. Supponiamo per assurdo che  $\exists t_0 < 0$  t.c.  $h(t_0) \neq 0$ . In particolare supponiamo che  $h(t_0) > 0$  (potremmo effettuare una dimostrazione analoga con  $h(t_0) < 0$ ). Inoltre ci limitiamo al caso in cui  $h$  sia continua in  $t_0$ .

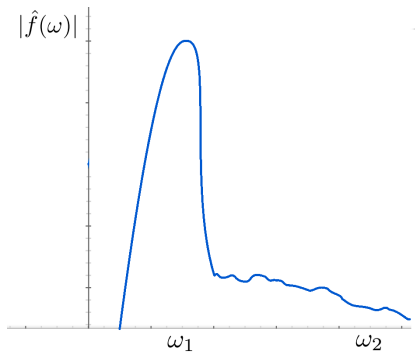
Per tutti questi motivi possiamo assumere che  $h(t)$  sia positiva in un intorno di



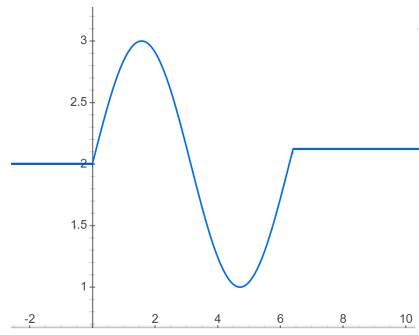
(a) Rappresentazione nello spazio del tempo di  $f$



(b) Rappresentazione nello spazio delle frequenze di  $\hat{f}$



(c) Rappresentazione nello spazio delle frequenze di  $\hat{f}H$ . La frequenza più alta è stata "eliminata" dal filtro



(d) Rappresentazione nello spazio del tempo di  $\mathcal{F}^{-1}(\hat{f} \cdot H)$ . La frequenza più alta è stata "eliminata" dal filtro

Figura 13: Le varie componenti della formula  $\mathcal{F}^{-1}[\hat{f} \cdot H]$

$t_0$ , in particolare in un intorno compreso tra gli estremi  $t \in [-b, -a]$  con  $a, b > 0$  e  $a < b$ .

Consideriamo allora la funzione

$$\chi_{[0, b-a]} = \begin{cases} 1, & \text{se } t \in [0, b-a] \\ 0, & \text{altrimenti} \end{cases}$$

$$\text{Allora } Af(t) = \int_{-\infty}^{+\infty} \chi_{[0, b-a]}(t-s)h(s)ds$$

Dunque in base alla definizione  $\chi_{[0, b-a]} = 1$  solo per  $0 \leq t-s \leq b-a$ , ovvero  $t-b+a \leq s$ . In ogni altro punto vale 0. Integriamo solo su tale intervallo allora:

$$Af(t) = \int_{t-b+a}^t h(s)ds$$

Ma per  $t = -a$

$$Af(t) = \int_{-b}^{-a} h(s)ds > 0$$

Ma questo è assurdo poiché avevamo supposto la realizzabilità del filtro che impone la condizione  $f(t) = 0, \forall t \leq t_0 \Rightarrow Af(t) = 0$   $\square$

*Dimostrazione.* ( $\Leftarrow$ ) sia che  $\text{supph} \subseteq [0, +\infty]$ , ovvero se  $h(t) = 0, \forall t < 0$ . Sia  $f(t) = 0, \forall t < t_0$ . Mostriamo che  $A$  è realizzabile, ovvero che  $Af(t) = 0, \forall t \leq t_0$ .

$$Af(t) = \int_{-\infty}^{+\infty} f(t-s)h(s)ds$$

Ma visto che nel semiasse negativo vale che  $h(t) = 0$ , allora:

$$Af(t) = \int_0^{+\infty} f(t-s)h(s)ds$$

Ma poiché  $s > 0$  allora  $t \leq t_0 \Leftrightarrow t-s \leq t_0-s \leq t_0$  dunque  $t-s \leq t_0$ .

Ma poiché  $f(t) = 0, \forall t < t_0$ , dunque  $f(t-s) = 0$ . Da cui ricaviamo che:

$$Af(t) = \int_0^{+\infty} 0 \cdot h(s)ds = 0 \quad \square$$

Grazie al risultato appena ottenuto ne ricaviamo che il filtro di convoluzione  $A; f \rightarrow f * \text{sinc}_{2a}$  passa basso ideale non è realizzabile. Infatti:

$$\begin{aligned} f * \text{sinc}_{2a} &= \mathcal{F}^{-1} \mathcal{F}[f * \text{sinc}_{2a}] = \mathcal{F}^{-1}[\hat{f} \cdot \widehat{\text{sinc}_{2a}}] = \mathcal{F}^{-1}[\hat{f} \cdot \widehat{\chi_{[-a, a]}}(t)] = \\ &= \mathcal{F}^{-1}[\hat{f} \cdot \chi_{[-a, a]}(-t)] = \mathcal{F}^{-1}[\hat{f} \cdot \chi_{[-a, a]}(t)] \end{aligned}$$

Dunque vediamo che applicando il filtro ideale a scalino  $\chi_{[-a, a]}(t)$  otteniamo una convoluzione basata sul seno cardinale, che non sarà mai a scalino, e dunque fornirà sempre una risultato in qualche modo "smussato".

Sempre in tema di realizzabilità è invece possibile mostrare che il circuito RC è un filtro di convoluzione realizzabile che approssima il filtro passa basso ideale.

Infatti se consideriamo il filtro

$A : f \rightarrow f * h$  con  $h(t) = \frac{1}{RC} e^{\frac{-t}{RC}} \chi_{[0, +\infty]}(t)$  allora abbiamo che il filtro è certamente realizzabile perché vale che  $\text{supp} h \subseteq [0, +\infty]$ , ovvero che  $h(t) = 0, \forall t < 0$ . La sua funzione di trasferimento è

$$H(\omega) = \hat{h}(\omega) = \frac{1}{1 + 2\pi i \omega RC}$$

e dunque

$$|H(\omega)| = \frac{1}{\sqrt{1 + 4\pi^2 R^2 C^2 \omega^2}}$$

### 3.6 Serie di Fourier

Un argomento di notevole interesse è la *serie di Fourier*.

Con questa serie matematica asseriamo<sup>3</sup> che ogni funzione in un certo intervallo può essere espressa come somma di seni e coseni, ovvero che le funzioni trigonometriche seno e coseno costituiscano una *base ortonormale* per lo spazio in cui la funzione è rappresentata. In particolare vale che:

$$f(t) = \frac{1}{2} a_0 \sum_{n=1}^{+\infty} a_n \cos\left(\frac{2\pi n}{T} t\right) + b_n \sin\left(\frac{2\pi n}{T} t\right) \quad (3)$$

con

$$a_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) \cos(2\pi n \omega_0 x) dx$$

$$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) \sin(2\pi n \omega_0 x) dx$$

$\forall n \in \mathbb{N}$

Notiamo che non forniamo alcun tipo di dettaglio sulla convergenza della serie di Fourier, ovvero non abbiamo detto come converge. La serie di Fourier può essere scritta, alternativamente, nella forma esponenziale come segue:

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{2\pi i n \frac{1}{T} t} \quad (4)$$

dove

$$c_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{2\pi i n \frac{1}{T} t} dt$$

Se andiamo a considerare un intervallo  $T$  sempre più ampio, nel limite  $T \rightarrow +\infty$

---

<sup>3</sup>Similmente alla serie di Taylor

otteniamo che il termine  $\frac{n}{T}$  presente nell'esponente di  $e$  diventa pari a  $\omega$ .  
 Notiamo che proprio con  $\frac{n}{T} = \omega$  otteniamo la trasformata di Fourier

## 4 Trasformata di Gabor

### 4.1 Definizione e significato intuitivo

Nelle pagine precedenti abbiamo visto che la trasformata di Fourier è uno strumento potente per comprendere la natura di un segnale ma anche la trasformata possiede delle limitazioni. Vediamone alcune:

**Esempio 1** : consideriamo un segnale  $f$  e lo stesso segnale traslato  $\tau_a f$ . Il modulo del primo (ovvero quello che riusciamo a rappresentare della trasformata) è pari a  $|\hat{f}(\omega)|$ . Consideriamo con maggiore attenzione il modulo della trasformata del secondo:

$$|\widehat{\tau_a f}(\omega)| = |\mu_{-a} \hat{f}(\omega)| = |e^{-2\pi a i \omega} \hat{f}(\omega)| = \underbrace{|e^{-2\pi a i \omega}|}_{=1} |\hat{f}(\omega)| = |\hat{f}(\omega)|$$

Dunque la trasformata di un segnale traslato è uguale alla trasformata del segnale non traslato

Quanto visto nell'esempio precedente è dovuto al fatto che con la trasformata di Fourier perdiamo ogni tipo di informazione temporale.

Per risolvere questo problema utilizziamo la *trasformata di Gabor di finestra  $g$*

$$V_g f(x, \omega) = \int_{-\infty}^{+\infty} f(t) e^{-2\pi i \omega t} \overline{g(t-x)} dt \quad (5)$$

Dove  $g(t)$  è una funzione graficamente simile alla funzione scalino, ma continua, dunque in qualche modo "smussata".

Intuitivamente che cosa stiamo facendo? Stiamo andando a moltiplicare la funzione da trasformare per una certa funzione  $g(t)$  più o meno a scalino. Questo implica che la trasformata di Gabor assumerà valori diversi da zero solo all'interno della "finestra" in cui la funzione  $g(t)$  è diversa da zero. Per questo motivo la trasformata di Gabor viene indicata come *trasformata di finestra  $g$* .

Solitamente, la rappresentazione del modulo di una trasformata di Gabor è qualcosa che assomiglia a quanto mostrato in Figura 14

Anche se più spesso viene rappresentato lo stesso grafico in due dimensioni  $x, \omega$  e il valore di  $|V_g f(x, \omega)|$  viene mostrato con una gradazione di colore.

Andiamo ad applicare la trasformata di Gabor ad un esempio per vederlo meglio

**Esempio 2** : supponiamo di avere a disposizione il solito segnale a due frequenze utilizzato finora e mostrato in Figura 8. La rappresentazione di  $|V_g f(x, \omega)|$  in due dimensioni viene mostrata in Figura 15

Notiamo che almeno in parte i segnali si sovrappongono perché la finestra utilizzata possiede una sua ampiezza. Quanto più è larga una finestra,



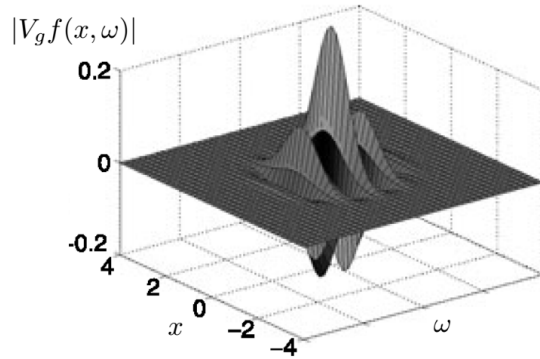


Figura 14: Rappresentazione di un generico  $|V_g f(x, \omega)|$

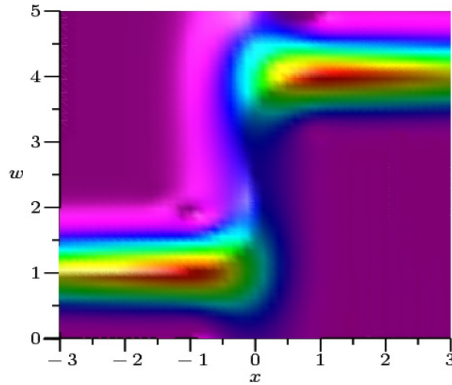


Figura 15: Rappresentazione di  $|V_g f(x, \omega)|$  relativamente al segnale di Figura 8

tanto più l'intervallo in cui la finestra copre entrambi i segnali è larga. Notiamo l'aspetto più interessante dell'intera questione: le due frequenze del segnale analizzato sono rappresentate nel tempo lungo l'asse  $x$ . Questo ci fornisce un collegamento con la vita di tutti i giorni: le partiture musicali altro non sono che la rappresentazione di una trasformata di Gabor. Notare come quanto mostrato in Figura 15 possa corrispondere a due note consecutive differenti (aventi diversa frequenza) nel tempo.

## 4.2 Principio di indeterminazione e ampiezza della finestra

In meccanica quantistica il famoso principio di indeterminazione di Heisenberg afferma che non è possibile determinare contemporaneamente con precisione arbitraria posizione e velocità di una particella. In teoria dei segnali vale un principio di indeterminazione che ha identica struttura matematica ma invece

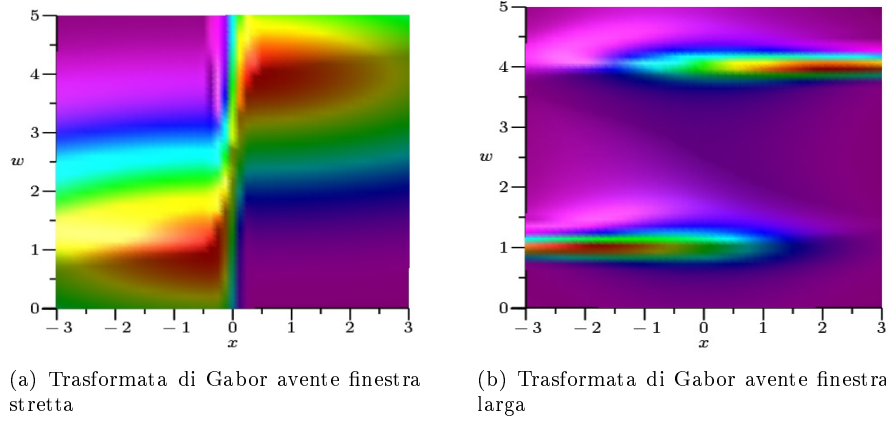


Figura 16: Valore di  $|V_g f(x, \omega)|$  al variare della larghezza della finestra

delle variabili posizione e velocità coinvolge le variabili durata e frequenza di un segnale.

Intuitivamente si può riassumere dicendo che maggiore è la "dispersione" nel tempo (ovvero la durata) di un segnale, minore è la sua "dispersione" nelle frequenze. Il concetto di dispersione è chiaramente legato al concetto di varianza. Infatti consideriamo la varianza del segnale e la varianza della sua trasformata:

$$\sigma_f^2 = \int_{-\infty}^{+\infty} t^2 |f(t)|^2 dt$$

$$\sigma_{\hat{f}}^2 = \int_{-\infty}^{+\infty} \omega^2 |\hat{f}(\omega)|^2 d\omega$$

Il principio di indeterminazione di Heisenberg afferma che

$$\sigma_f^2 \cdot \sigma_{\hat{f}}^2 \geq \frac{1}{4\pi} \quad (6)$$

Dunque abbiamo un'elevata varianza o nello spazio dei tempi, o nello spazio delle frequenze.

Questo rappresenta un problema perché una finestra troppo stretta non permette di individuare con precisione le frequenze appartenenti al segnale. Questa situazione è mostrata nella parte sinistra della Figura 16

D'altro canto, una finestra troppo larga porta ad avere un tempo abbastanza lungo in cui i segnali si sovrappongono come accennavamo in precedenza, dunque perdiamo parte dell'informazione temporale.

15 / 11 / 2017

Un'idea che potremmo avere è quella di considerare il prodotto di due trasforma-

ta di Gabor  $V_{g1}(f(x)) \cdot V_{g2}(f(x))$  dove  $V_{g1}$  usa una finestra molto stretta mentre  $V_{g2}$  usa una finestra molto larga. Questo farebbe sì che il risultato si annulli in ogni punto in cui almeno una delle due trasformate si annulla. Il problema di questo approccio è che otteniamo una rappresentazione tempo-frequenze fatta a griglia in cui compaiono molti segnali "fantasma".

Per concludere potremmo ricostruire un procedimento analogo a quello della trasformata di Fourier. Supponendo di avere un filtro della forma

$$Q(x, \omega) = \begin{cases} 0, & \text{se } (x, \omega) \in Q \\ 1, & \text{se } (x, \omega) \in \mathbb{R}^2 - Q \end{cases}$$

Supponendo di applicare il filtro ad un certo segnale da ripulire  $f(t)$ , otterremo un segnale filtrato  $\tilde{f}(t)$  della forma:

$$\tilde{f}(t) = \underbrace{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Q(x, \omega) \underbrace{V_g(f(x, \omega))}_{\text{analisi}} e^{2\pi i t \omega} g(t - x) dx \, d\omega}_{\text{ricostruzione}} \quad (7)$$

detto anche *filtro di Gabor*.

Una particolarità da ricordare è che il filtro lo potevamo già fare con le convoluzioni, ma in realtà un filtro che cambia nel tempo (fattibile scegliendo opportunamente  $Q(x, \omega)$ ) non può essere fatto dalla convoluzione.

## 5 Introduzione alla probabilità

### 5.1 Eventi

Consideriamo uno *spazio campione*  $S$  che contiene tutti i possibili esiti di un certo esperimento. Quando si verifica un certo *evento*  $E$ , lo indichiamo proprio con  $E$ . Gli eventi possono essere composti come segue:

$$E \cup F, E \cap F, \bigcup_{i=1}^{\infty} F_i, F^C$$

Rispettivamente le operazioni sopra indicate sono l'unione, l'intersezione che da ora in poi verrà indicata semplicemente con  $P(EF)$ , la somma infinita e il complemento.

Cosa è la probabilità dunque? È un valore compreso tra 0 e 1 che indica quanto è probabile che un certo evento si verifichi. Se una certa probabilità è influenzata da altri eventi, possiamo utilizzare la *probabilità condizionata*.

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Quando invece due eventi sono indipendenti, vale che  $P(E|F) = P(E)$  e dunque  $P(EF) = P(E)P(F)$ .

Vediamo un esempio che applica quanto detto finora:

**Esempio 1** : supponiamo che tre persone possiedano ognuna un cappello che lanciano in mezzo ad una stanza. Supponendo che ognuno riprenda casualmente un cappello, quale è la probabilità che ognuno prenda il cappello di un altro? Indichiamo questo evento con  $P^*$

Cominciamo a calcolare la probabilità di alcuni eventi più semplici. Chiamiamo  $E_i$  l'evento in cui l' $i$ -esima persona riprende il proprio cappello.

$$\text{Calcoliamo: } P(E_i) = \frac{1}{3}$$

$$P(E_i E_j) = \frac{1}{3} * \frac{1}{2} = \frac{1}{6}$$

$$P(E_1 E_2 E_3) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) = \frac{1}{3} * \frac{1}{2} * 1 = \frac{1}{6}$$

Ma noi volevamo calcolare la probabilità che nessuno riprendesse il proprio. Possiamo vedere

$$\begin{aligned} P^* &= 1 - P(E_1 \cup E_2 \cup E_3) = \\ &= 1 - (P(E_1) + P(E_2) + P(E_3) + P(E_1 E_2) + P(E_2 E_3) + P(E_1 E_3) + \\ &\quad P(E_1 E_2 E_3)) = 1/3 \end{aligned}$$

Affinché un insieme di eventi (prendiamone a titolo di esempio 3) siano indipendenti deve essere vero che:

$$\begin{aligned}
P(E_1 E_2) &= P(E_1)P(E_2) \\
P(E_2 E_3) &= P(E_2)P(E_3) \\
P(E_1 E_3) &= P(E_1)P(E_3) \\
P(E_1 E_2 E_3) &= P(E_1)P(E_2)P(E_3)
\end{aligned}$$

Notiamo che l'ultima equazione non è in alcun modo implicata dalla precedenti. Vediamolo con l'esempio che segue

**Esempio 2** : consideriamo quattro palline inserite in un'urna. Consideriamo i tre eventi che seguono:

$$E_1 = \{1, 2\}, E_2 = \{1, 3\}, E_3 = \{1, 4\}$$

dove  $E_1$  si verifica se estraiamo la pallina 1 o la 2.

Se andiamo a calcolare le tre probabilità che seguono:  $P(E_1 E_2) = \frac{1}{4} =$

$$P(E_1)P(E_2)$$

$$P(E_2 E_3 = \frac{1}{4}) = P(E_2)P(E_3)$$

$$P(E_1 E_3 = \frac{1}{4}) = P(E_1)P(E_3)$$

Abbiamo verificato che gli eventi  $E_1, E_2, E_3$  sono indipendenti a coppie.

Eppure, se andiamo a calcolare la probabilità

$$P(E_1 E_2 E_3) = \frac{1}{4} \neq P(E_1)P(E_2)P(E_3) = \frac{1}{8}$$

Consideriamo adesso l'evento:

$$A = \bigcup_i A_i$$

Possiamo ricavare il teorema delle probabilità totali il quale ci dice che:

$$P(A) = \sum_{i=1}^k P(A|B_i)P(B_i) = \sum_{i=1}^k P(A \cap B_i)$$

Un altro elemento importante nella probabilità è la formula di Bayes:

$$P(D|E) = \frac{P(E|D)P(D)}{P(E)}$$

che si ricava dalla definizione di probabilità condizionata e serve per identificare la probabilità di una certa causa partendo dalla probabilità dell'effetto.

**Esempio 3** : supponiamo di avere a disposizione un test che riconosce una malattia nel 95% dei casi. Purtroppo il test crea dei falsi positivi con probabilità  $p = 0.01$  nelle persone sane. Se sappiamo che lo 0.05% della popolazione è malata, quale è la probabilità che avendo ottenuto un esito positivo, una persona sia effettivamente malata? Andando a definire gli eventi:

$$D = \{\text{è malato}\}$$

$$E = \{\text{test positivo}\}$$

Allora vogliamo calcolare:

1		1	1			1		1			
---	--	---	---	--	--	---	--	---	--	--	--

Figura 17: Vettore di bit dell'esempio 1

$$P(D|E) = \frac{P(E|D)P(D)}{P(E)} = \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|\bar{D})P(\bar{D})} = 0.32$$

Quindi ricevere un esito positivo impone di effettuare altri controlli, visto che è comunque meno probabile essere malati che non esserlo!

## 6 Variabili aleatorie discrete

Una variabile aleatoria è una funzione misurabile da uno spazio campione  $S$  a uno spazio misurabile  $E$ , ovvero:

$$X : S \rightarrow E$$

Spesso indicheremo  $P(X = x) = P(\omega : X(\omega) = x)$

Esistono diversi tipi di variabili aleatorie discrete:

- Bernoulli:  $Ber : P(X = 1) = p$   
Notiamo che  $p = 1 - P(X = 0)$
- Binomiale:  $Bin(n, p) : P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$   
supponendo di avere  $n$  prove indipendenti, contiamo il numero di successi in queste prove. Per comprendere meglio la natura della distribuzione binomiale, considerare l'esempio 1. Notiamo che comunque la distribuzione binomiale considera il numero di successi di una serie di eventi che seguono la distribuzione di Bernoulli
- Geometrica:  $G : P(X = n) = (1 - p)^{n-1} p$   
questa distribuzione va a verificare quale sia la probabilità di avere un successo dopo  $n - 1$  insuccessi. Dunque abbiamo successo all' $n$ -esimo tentativo.

**Esempio 1:** supponiamo di avere un vettore di bit come quello mostrato in Figura 17

All'interno di questo vettore abbiamo cinque valori settati a 1 mentre gli altri sono settati a 0. La probabilità che si verifichi proprio questa configurazione (supponendo che un 1 si presenti con probabilità  $p$ ) è:

$p \cdot (1 - p) \cdot p \cdot p \cdot (1 - p) \cdot (1 - p) \cdot p \cdot (1 - p) \cdot p \cdot (1 - p) \cdot (1 - p) \cdot (1 - p) = p^5 (1 - p)^7$   
Ma questa è solo una delle possibili  $\binom{n}{k}$  configurazioni. Dunque la probabilità  $P(X = 5)$  è pari a:

$$P(X = 5) = \binom{12}{5} p^5 (1 - p)^7$$

Ognuna delle variabili aleatorie appena presentate possiede un proprio *valore atteso* solitamente noto come *media*. La definizione di valore atteso è:

$$E[X] = \mu = \sum_i x_i P(X = x_i)$$

Ovvero è pari alla somma del valore degli eventi per la probabilità che si verifichino proprio quegli eventi. Vediamo la media delle variabili discrete sopra elencate:

- Bernoulli:  $E_{Ber}[X] = 0 \cdot (1 - p) + 1 \cdot p = p$
- Binomiale:  $E_{Bin}[X] = \sum_i k \binom{n}{k} p^k (1 - p)^{n-k} = \dots$   
possiamo dimostrare quale sia la media secondo questa serie di passaggi matematici ma in effetti è più facile ricordarsi che una distribuzione binomiale è la sommatoria di n distribuzioni di Bernoulli. Dunque:  
 $E_{Bin}[X] = \sum_i E_{Ber}[X] = np$
- Geometrica:  $E_G[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p \sum_{n=1}^{\infty} \underbrace{n(1-p)^{n-1}}_{=-\frac{d(1-p)}{dp}}$

Ricordando che  $\sum_0^{\infty} q^n = \frac{1}{1-q}$ , otteniamo:

$$p \sum_{n=1}^{\infty} n(1-p)^{n-1} = -p \frac{d(\frac{1}{p} - 1)}{dp} = -p(-\frac{1}{p^2}) = \frac{1}{p}$$

Notiamo che l'operatore media è lineare, dimostriamolo:

$$\begin{aligned} E[X + Y + Z] &= \sum_{i,j,k} (x_i + y_j + z_k) P(X = x_i, Y = y_j, Z = z_k) = \\ &= \sum_i x_i \underbrace{\sum_{j,k} P(X = x_i, Y = y_j, Z = z_k)}_{=P(X=x_i)} + \sum_j y_j \underbrace{\sum_{i,k} P(X = x_i, Y = y_j, Z = z_k)}_{=P(Y=y_j)} + \\ &\sum_k z_k \underbrace{\sum_{i,j} P(X = x_i, Y = y_j, Z = z_k)}_{=P(Z=z_k)} = E[X] + E[Y] + E[Z] \end{aligned}$$

Un dato particolarmente importante oltre alla media è la dispersione intorno alla media, detta *varianza*. La definizione di varianza è=:

$$\begin{aligned} Var[X] &= E((X - E[X])^2) = \sum_x (X - E[X])^2 P(X = x) = \\ &= \sum_x (X - \mu)^2 P(X = x) \end{aligned}$$

Insieme alla varianza introduciamo la deviazione standard:

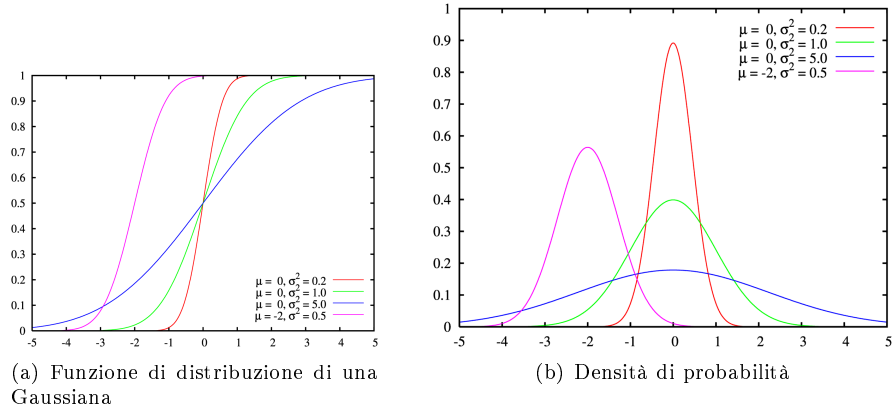


Figura 18: Confronto fra funzione di distribuzione e densità di probabilità

$$StDev(X) = \sqrt{Var(X)}$$

Notiamo che la relazione  $Var(\sum X_i) = \sum(Var X_i)$  è verificata solo se le variabili sono indipendenti.

17 / 10 / 2017

## 7 Variabili aleatorie continue

Nel dominio continuo non ha più senso considerare  $P(X = x)$  perché come vedremo abbiamo a che fare con gli integrali, e l'integrale in un certo punto è sempre uguale a 0.

Dunque ci interessa maggiormente considerare la *funzione di distribuzione cumulativa* o *funzione di ripartizione*

$$P(X \leq x) = F_X(x)$$

con  $x \in \mathbb{R}$ .

Un altro aspetto che ci interessa e da non confondere con quello appena illustrato è la definizione di *densità di probabilità*:

$$f_X(x) = \frac{d(F_X(x))}{dx}$$

Onde evitare di fare confusione mostriamo in Figura 18 entrambi i concetti descritti, relativamente ad una Gaussiana<sup>4</sup>

Notiamo che la funzione di distribuzione tende sempre ad 1 per  $x \rightarrow \infty$ . Questo succede perché deve sempre essere vero che la somma delle probabilità su tutti

<sup>4</sup>che definiremo in seguito ma che speriamo sia nota al lettore



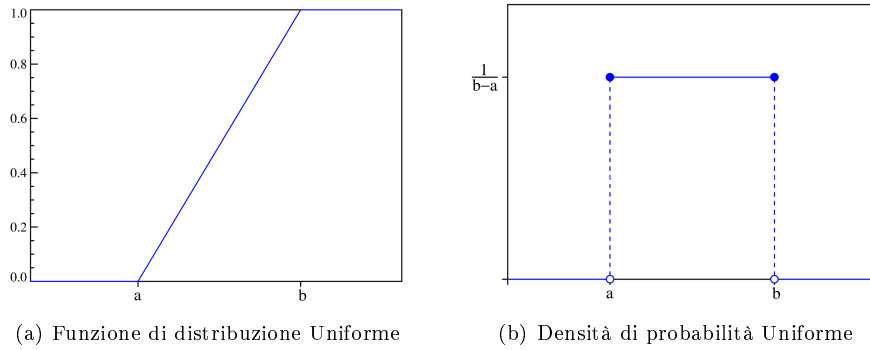


Figura 19: Variabile aleatoria Uniforme

gli eventi deve essere pari a 1.

Vediamo quali sono le più note variabili aleatorie continue

- Uniforme: la distribuzione uniforme possiede la seguente densità di probabilità:

$$f_U(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a, b] \\ 0, & \text{altrimenti} \end{cases}$$

La sua funzione di distribuzione è

$$F_X(x) = \int_a^x \frac{1}{b-a}$$

Mostriamo sia la funzione di distribuzione che la densità in Figura 19

La media della distribuzione è

$$E[f_U(x)] = \int_a^b x \frac{1}{b-a} dx = \frac{b+a}{2}$$

La varianza è invece:

$$Var[f_U(x)] = \frac{(b-a)^2}{12}$$

- Esponenziale: la densità di probabilità esponenziale è:

$$f_E(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x > 0 \\ 0, & \text{altrimenti} \end{cases}$$

mentre la funzione di distribuzione è:

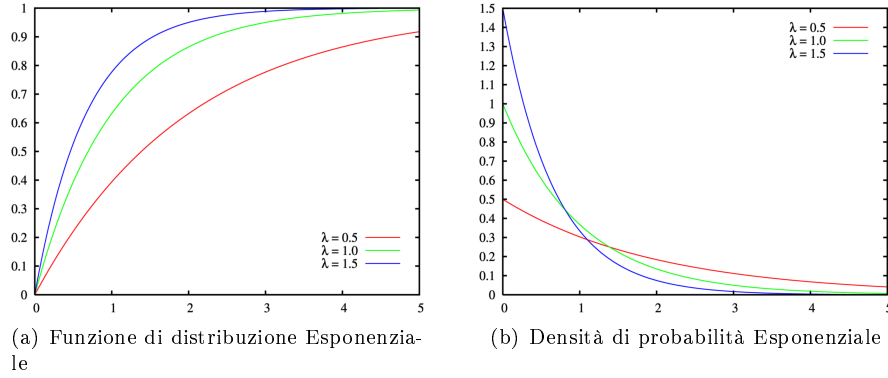


Figura 20: Variabile aleatoria Esponenziale al variare di  $\lambda$

$$F_E(x) = 1 - e^{-\lambda x} \text{ se } x > 0$$

Anche in questo caso mostriamo in Figura 20 la funzione di distribuzione e la densità di probabilità.

La media della distribuzione esponenziale è:

$$E[f_E(x)] = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} x e^{-\lambda x} dx = \frac{1}{\lambda}$$

Mentre la varianza è:

$$Var[f_E(x)] = \int_0^{+\infty} (X - E[X])^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

È interessante notare che sia media che varianza dipendano in qualche modo dall'inverso di  $\lambda$ . Questo ci dice che presa una serie di dati di cui vogliamo individuare la distribuzione  $X$ , se andando a calcolare media e varianza ottengo che la varianza è il quadrato della media, ci sono buone probabilità che la distribuzione  $X$  sia di tipo esponenziale.

- Gamma: la densità di probabilità della funzione Gamma è:

$$f_G(x) = \begin{cases} \lambda e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{se } x > 0 \\ 0, & \text{altrimenti} \end{cases}$$

Dove  $\Gamma(\alpha)$  è detta *gamma di Eulero* ed è una funzione la cui definizione è ricavata dalla funzione di distribuzione. Infatti, sapendo che  $\int_{-\infty}^{+\infty} F_G(x) = 1$  otteniamo che:

$$\Gamma(\alpha) = \int_0^{+\infty} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} dx$$

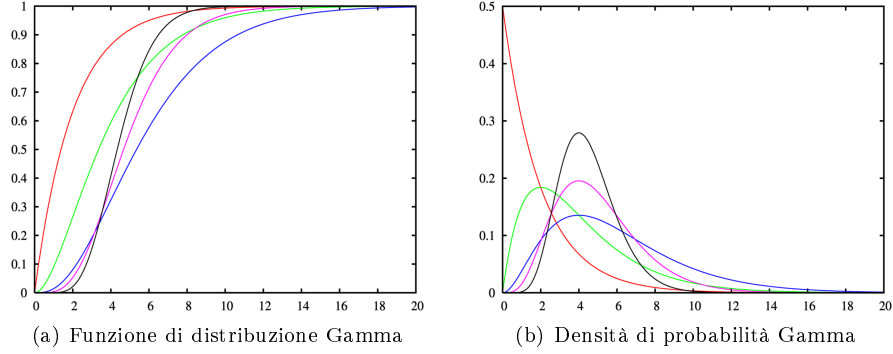


Figura 21: Variabile aleatoria Gamma al variare di  $\lambda$  e di  $\alpha$

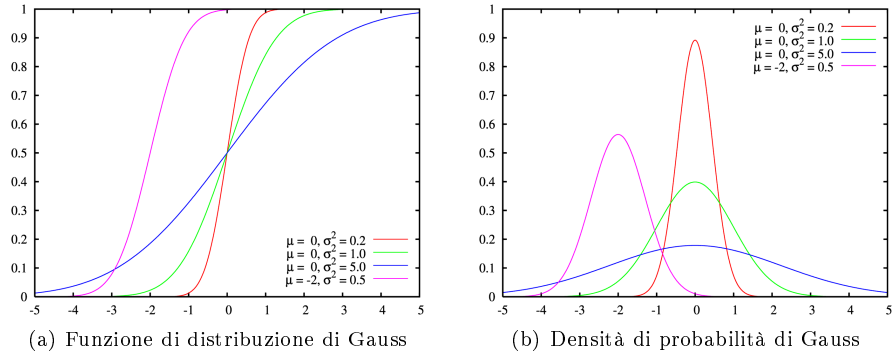


Figura 22: Variabile aleatoria di Gauss al variare di  $\mu$  e di  $\sigma$

Mostriamo in Figura 21 la funzione di distribuzione e la densità di probabilità:

In questo caso diamo senza eseguire il calcolo media e varianza:

$$E[f_G(x)] = \frac{\alpha}{\lambda}$$

$$Var[f_G(x)] = \frac{\alpha}{\lambda^2}$$

- Gaussiana: la densità di probabilità gaussiana è:

$$f_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Mostriamo in Figura 22 la funzione di distribuzione e la densità di probabilità:

In questo caso abbiamo che:

$$\begin{aligned} E[f_N(x)] &= \mu \\ Var[f_N(x)] &= \sigma^2 \end{aligned}$$

## 8 Distribuzioni congiunte

Andiamo a considerare le distribuzioni congiunte di probabilità. Siano  $X$  e  $Y$  due variabili aleatorie, allora:

$$P(X \leq x, Y \leq y) = F_{X,Y}(x, y)$$

Nei casi in cui  $X$  e  $Y$  siano indipendenti, possiamo dire che:

$$P(X \leq x, Y \leq y) = F_X(x)F_Y(y)$$

Osserviamo alcuni casi particolari di distribuzioni congiunte:

$$\begin{aligned} P(X < \infty, Y \leq y) &= F_Y(y) \\ P(X < -\infty, Y \leq y) &= 0 \\ P(X < \infty, Y < \infty) &= 1 \end{aligned}$$

La densità di probabilità per variabili congiunte è:

$$f_{XY} = \frac{\partial(F_{XY}(x, y))}{\partial x \partial y}$$

La media di una distribuzione congiunta è:

$$E[g(x, y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x, y) f_{X,Y}(x, y) dx, dy$$

La linearità dell'operatore media continua a valere. Dunque se le variabili sono indipendenti, vale che:

$$E[k(X)h(Y)] = E[k(X)]E[h(Y)]$$

Il momento di ordine secondo è invece detto *covarianza* e ci fornisce un valore che indica quanto le due variabili siano legate fra loro. Infatti abbiamo che:

$$\begin{aligned} Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]] = \\ &= E[XY] - E[XE[Y]] - E[YE[X]] + E[E[X]E[Y]] = E[XY] - E[X]E[Y] \end{aligned}$$

Dalla primissima uguaglianza possiamo notare una cosa importante: se i due termini sono di segno concorde la varianza è positiva, altrimenti è negativa.

Il significato intuitivo della covarianza è il seguente: se è molto positiva vuol

dire che tanto più cresce(decesce) una variabile, quanto più anche l'altra cresce(decesce). Se invece è molto negativa tanto più una variabile cresce(decesce), l'altra decresce(cresce).

Notiamo una cosa importante. Se due variabili sono indipendenti, la loro covarianza è zero. Il viceversa non è vero in generale. Infatti se la covarianza è nulla non è detto che le variabili siano indipendenti.

Un problema della covarianza si presenta al momento dell'analisi dimensionale. Infatti, supponendo che una variabile abbia la propria unità di misura in metri e l'altra in chilogrammi, otterremo una covarianza la cui unità di misura è  $m \cdot kg$ . Allora si preferisce usare il *coefficiente di correlazione*

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

con  $\rho_{X,Y} \in [-1, 1]$

Andiamo invece a definire l'*indicatore* o *funzione indicatrice* di un evento rispetto ad una certa variabile:

$$X = 1_A = \begin{cases} 1, & \text{se } A \text{ si verifica} \\ 0, & \text{altrimenti} \end{cases}$$

Allora consideriamo la covarianza di due indicatori:

$$\begin{aligned} Cov(1_A, 1_B) &= E[1_A, 1_B] - (E[1_A])(E[1_B]) = \\ &= 1 \cdot P(1_A = 1, 1_B = 1) - 1 \cdot P(1_A = 1) \cdot 1 \cdot P(1_B = 1) \end{aligned}$$

Vediamo cosa succede ponendo che la covarianza sia positiva. Dunque:

$$\begin{aligned} Cov(1_A, 1_B) > 0 &\Leftrightarrow P(1_A = 1, 1_B = 1) > P(1_A = 1)P(1_B = 1) \Leftrightarrow \\ &\Leftrightarrow \frac{P(1_A = 1, 1_B = 1)}{P(1_B = 1)} > P(1_A = 1) \Leftrightarrow P(1_A = 1 | 1_B = 1) > P(1_A = 1) \end{aligned}$$

Dunque essere a conoscenza che si verifichi B, aumenta la probabilità che si verifichi A.

Proviamo invece a calcolare la varianza di una somma di variabili aleatorie:

$$\begin{aligned} Var(X + Y) &= E[(X + Y - \underbrace{E[X]}_{=\mu_X} - \underbrace{E[Y]}_{=\mu_Y})^2] = E[((X - \mu_X) + (Y - \mu_Y))^2] = \\ &= E[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)] = \\ &= E[(X - \mu_X)^2] + E[(Y - \mu_Y)^2] + 2E[(X - \mu_X)(Y - \mu_Y)] = \\ &= Var(X) + Var(Y) + 2Cov(X, Y) \end{aligned}$$

Notiamo una cosa abbastanza interessante: se la  $Cov(X, Y) = 0$ , allora la varianza della somma è pari alla somma delle varianze. Invece, se la covarianza è positiva, la varianza della somma è maggiore della somma delle varianze. Ma il fatto più importante è che, potendo essere negativa la covarianza, allora in certi

casi la varianza della somma potrebbe invece essere addirittura minore della somma delle varianze!

## Esercizi di fine lezione

**Esercizio 1** : supponiamo di avere a disposizione due variabili aleatorie indipendenti che seguano una distribuzione binomiale, ovvero:

$$\begin{aligned} N &\sim Bi(n, p) \\ M &\sim Bi(m, p) \end{aligned}$$

Quale è il valore di  $P(M + N = j)$ ?

Intuitivamente, visto che  $N$  effettua  $n$  prove mentre  $M$  ne effettua  $m$ , e visto  $N$  ed  $M$  sono indipendenti, ci potremmo aspettare che la somma di queste variabili aleatorie sia  $K \sim Bi(n + m, p)$

Cerchiamo di dimostrarlo:

$$\begin{aligned} P(M + N = j) &= \sum_{k=0}^{\min(n, j)} P(M + N = j | N = k) \cdot P(N = k) = \\ &= \sum_{k=0}^{\min(n, j)} P(M = j - k | N = k) \cdot P(N = k) = \\ &= \sum_{k=0}^{\min(n, j)} P(M = j - k) \cdot P(N = k) = \\ &= \sum_{k=0}^{\min(n, j)} \binom{m}{j-k} p^{j-k} (1-p)^{m-j+k} \binom{n}{k} p^k (1-p)^{n-k} = \\ &= p^j (1-p)^{m+n-j} \sum_{k=0}^{\min(n, j)} \binom{m}{j-k} \binom{n}{k} = \\ &= Bi(n + m, p) \end{aligned}$$

Dunque la somma di due binomiali aventi stessa probabilità di successo  $p$  fornisce ancora una binomiale in cui eseguiamo un numero di prove pari alla somma delle prove delle due binomiali.

**Esercizio 2** : consideriamo, in modo analogo, due variabili di Poisson indipendenti. Quale è il risultato di  $P(X + Y = i)$ ?

$$\begin{aligned} P(X + Y = i) &= \sum_{j=0}^i P(X + Y = i | Y = j) \cdot P(Y = j) = \\ &= \sum_{j=0}^i P(X = i - j | Y = j) \cdot P(Y = j) = \sum_{j=0}^i \frac{\lambda^{i-j} e^{-\lambda}}{(i-j)!} \frac{\beta^j e^{-\beta}}{j!} = \\ &= (\lambda + \beta)^i \frac{e^{-(\lambda+\beta)}}{i!} \end{aligned}$$

Riprendiamo quanto detto sulle distribuzioni congiunte. Supponiamo di avere due risultati  $y_1, y_2$  dipendenti ognuno da due parametri  $x_1, x_2$ . Ovvero vale che:

$$y_1 = g_1(x_1, x_2) \text{ e } y_2 = g_2(x_1, x_2)$$

Supponendo di conoscere la distribuzione congiunta  $f_{x_1, x_2}(x_1, x_2)$  è possibile conoscere la distribuzione  $f_{y_1, y_2}(y_1, y_2)$ ?

In effetti sì, tramite la seguente uguaglianza:

$$f_{y_1, y_2}(y_1, y_2) = \{f_{x_1, x_2}(x_1, x_2) \left| \frac{1}{J(x_1, x_2)} \right| \} \quad (8)$$

dove  $J(x_1, x_2)$  è il determinante della *matrice di Jacobi* della forma:

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix}$$

Cerchiamo di comprendere meglio tramite un esempio

**Esempio 1** : consideriamo due variabili aleatorie  $X \sim \Gamma(n, \lambda)$  e  $Y \sim \Gamma(m, \lambda)$

Allora abbiamo che:

$$f_{X,Y}(x, y) = (\lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{\Gamma(n)}) (\lambda e^{-\lambda y} \frac{(\lambda y)^{m-1}}{\Gamma(m)})$$

Prendiamo come  $g_1(x, y)$  la somma

$$U = X + Y$$

e come  $g_2(x, y)$  la divisione

$$V = \frac{X}{X + Y}$$

Allora vediamo che possiamo scrivere  $x$  e  $y$  in una forma alternativa:

$$x = uv = h_1(u, v) \text{ e } y = u(1 - v) = h_2(u, v)$$

A questo punto abbiamo che:

$$J(x, y) = \begin{vmatrix} \frac{1}{y} & \frac{1}{-x} \\ \frac{1}{(x+y)^2} & \frac{-1}{(x+y)^2} \end{vmatrix} = \frac{-1}{x+y} = \frac{-1}{u}$$

Dunque, complessivamente abbiamo che:

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(uv, u(1-v)) \left| \frac{1}{J(x, y)} \right| = f_{X,Y}(uv, u(1-v)) \cdot u = \\ &= \underbrace{\frac{\lambda e^{-\lambda u} (\lambda u)^{n+m-1}}{\Gamma(m+n)}}_{\phi(u)} \cdot \underbrace{v^{n-1} (1-v)^{m-1} \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)}}_{\psi(v)} \end{aligned}$$

Notiamo che  $\phi(u)$  e  $\psi(v)$  sono funzioni indipendenti, dunque  $U$  e  $V$  lo

sono, fatto per niente scontato reso possibile soltanto dalle distribuzioni che abbiamo utilizzato.

## 8.1 Funzione generatrice dei momenti

Un breve parallelismo che intendiamo fare riguarda la *funzione generatrice dei momenti*. Questa è solitamente definita come

$$\phi_x(t) = E[e^{tx}] = \begin{cases} \sum_i e^{tx_i} P(X = x_i), & \text{nel caso discreto} \\ \int_{\mathbb{R}} e^{tx} f_X(x) dx, & \text{nel caso continuo} \end{cases} \quad (9)$$

Notiamo una cosa interessante: se andiamo ad utilizzare una funzione leggermente diversa della forma:

$$\varphi_x(t) = E[e^{-i\pi t X}] = \begin{cases} \sum_x e^{-2\pi i t x} p(x), & \text{nel caso discreto} \\ \int_{\mathbb{R}} e^{-2\pi i t x} f_X(x) dx, & \text{nel caso continuo} \end{cases} \quad (10)$$

abbiamo la trasformata di Fourier. Ad ogni modo ci concentreremo sulla versione basilare della funzione generatrice. L'idea interessante che sta dietro alla funzione appena definita è che la sua derivata fornisce il momento di ordine primo, ovvero la media. La derivata seconda fornisce il momento di ordine secondo, ovvero la varianza. Vediamolo:

$$\left. \frac{d\varphi}{dt} \right|_{t=0} = \left. \int x e^{tx} f_X(x) dx \right|_{t=0} = \int x f_X(x) dx = E[X]$$

## 9 Attese condizionate

Consideriamo la probabilità condizionata di due variabili aleatorie discrete:

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

Tale probabilità è una distribuzione. Per verificarlo devono essere dimostrate due proprietà

1.  $P(X = x_i | Y = y_j) \geq 0$ : verificata sempre assumendo che gli eventi siano compatibili
2.  $\sum_i P(X = x_i | Y = y_j) = 1$ : per verificarlo dobbiamo vedere che

$$\sum_x \frac{P(X = x, Y = y_j)}{P(Y = y_j)} = \frac{P(Y = y_j)}{P(Y = y_j)} = 1$$

Dunque abbiamo una distribuzione di cui possiamo calcolare un valore atteso:

$$E(X | Y = y_j) = \sum x_i P(X = x_i | Y = y_j) = \psi(y_j)$$



È importante notare che la media di una variabile condizionata da  $y$  è solamente funzione di  $y$ .

**Esempio 1** : consideriamo due variabili di Poisson indipendenti

$$X \sim P(\lambda_1)$$

$$Y \sim P(\lambda_2)$$

Quanto vale  $P(X = k | X + Y = n)$ ?

$$P(X = k | X + Y = n) = \frac{P(X = k, X + Y = n)}{P(X + Y = n)} = \frac{P(X = k, Y = n - k)}{P(X + Y = n)}$$

Avevamo però dimostrato in precedenza che una somma di due variabili di Poisson segue sempre una distribuzione di Poisson avente parametro  $\lambda$  pari alla somma delle  $\lambda$ . Dunque vale che:

$$P(X + Y = n) = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}$$

Dunque:

$$\begin{aligned} \frac{P(X = k, Y = n - k)}{P(X + Y = n)} &= \frac{e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}} = \\ &= \binom{n}{k} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k} = \\ &= \binom{n}{k} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left( 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{n-k} \end{aligned}$$

Notiamo che l'ultima uguaglianza rappresenta una binomiale. Dunque, ricordando che la media di una binomiale è  $np$  allora abbiamo che:

$$E[(X | X + Y = n)] = n \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Spostiamoci invece nel continuo. Ancora una volta, non ha senso chiedersi quanto valga  $P(X = k | \dots)$  perché varrebbe sempre zero. Preferiamo usare una probabilità su un intervallo del tipo.:

$$\begin{aligned} P(X \leq x | Y = y) &= \lim_{h \rightarrow 0} P(X \leq x | Y \in (y, y+h)) = \lim_{h \rightarrow 0} \frac{P(X \leq x, Y \in (y, y+h))}{P(Y \in (y, y+h))} = \\ &= \lim_{h \rightarrow 0} \frac{\int_{-\infty}^x du \int_y^{y+h} dv f_{X,Y}(u, v)}{\int_y^{y+h} dv f_Y(v)} = \frac{\int_{-\infty}^x du f_{X,Y}(u, y)}{f_Y(y)} = \int_{-\infty}^x du \underbrace{\frac{f_{X,Y}(u, y)}{f_Y(y)}}_{=f_{X|Y}(u|y)} \end{aligned}$$

Dunque:

$$P(X \leq x | Y = y) = F_{X|Y}(X|Y) \quad (11)$$

## Esercizi di fine lezione

**Esercizio 1** : consideriamo due variabili indipendenti che seguano una distribuzione binomiale:

$$X \sim B(n, p)$$

$$Y \sim B(n, p)$$

Quanto vale  $P(X = k | X + Y = m)$ ?

Similmente a quanto visto poco fa, vale che

$$\begin{aligned} P(X = k | X + Y = m) &= \frac{P(X = k, X + Y = m)}{P(X + Y = m)} = \frac{P(X = k, Y = m - k)}{P(X + Y = m)} = \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k}}{\binom{2n}{m} p^m (1-p)^{2n-m}} = \frac{\binom{n}{k} \binom{n}{m-k}}{\binom{2n}{m}} \end{aligned}$$

**Esercizio 2** : supponiamo di avere eseguito un esperimento con tre possibili esiti aventi probabilità  $p_1, p_2, p_3$ . Dunque vale che  $p_1 + p_2 + p_3 = 1$ . Ripetiamo l'esperimento  $n$  volte e indichiamo con  $X_i$  il numero di volte che l'esito  $i$ -esimo si è verificato. Quanto vale  $P(X_1 = k | X_2 = m)$ ?

$$P(X_1 = k | X_2 = m) = \frac{P(X_1 = k, X_2 = m)}{P(X_2 = m)}$$

Ma visto che, se eseguiamo  $n$  prove, di cui  $k$  danno esito 1,  $m$  danno esito 2, allora il numero di esiti di tipo 3 sarà pari a  $n - m - k$ . Possiamo aggiungere questa condizione alla congiunzione che resta equivalente. Dunque:

$$\begin{aligned} \frac{P(X_1 = k, X_2 = m)}{P(X_2 = m)} &= \frac{P(X_1 = k, X_2 = m, X_3 = n - m - k)}{P(X_2 = m)} = \\ &= \frac{\frac{n!}{k!m!(n-m-k)!} p_1^k p_2^m p_3^{n-m-k}}{\frac{n!}{m!(n-m)!} p_2^m (1-p_2)^{n-m}} = \binom{n-m}{k} \left( \frac{p_1}{1-p_2} \right)^k \left( \frac{p_3}{1-p_3} \right)^{n-m-k} \end{aligned}$$

Complessivamente la distribuzione seguita è ancora una binomiale del tipo  $B(n - m, \frac{p_1}{1-p_2})$

31 / 10 / 2017

Riprendiamo le attese condizionate tramite un esempio:

**Esempio 2** : consideriamo la seguente distribuzione congiunta

$$f_{X,Y}(x, y) = \begin{cases} 6xy(2 - x - y), & \text{se } x \in (0, 1), y \in (0, 1) \\ 0, & \text{altrimenti} \end{cases}$$

Quanto vale il valore atteso  $E[X|Y = y]$ ?

Dalla definizione di dipendenza condizionata sappiamo che

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Dunque per trovare  $f_{X|Y}(x,y)$  abbiamo bisogno della distribuzione marginale  $f_Y(y)$ . Per ottenerla dobbiamo integrare la distribuzione congiunta su tutti i valori di  $X$  come segue:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx = \int_0^1 f_{X,Y}(x,y) dx = \int_0^1 6xy(2-x-y) dx = \\ &= 6y - 2y - 3y^2 = 4y - 3y^2 = y(4-3y) \end{aligned}$$

dunque

$$f_{X|Y}(x|y) = \frac{6xy(2-x-y)}{y(4-3y)} = \frac{6x(2-x-y)}{4-3y}$$

Adesso che abbiamo la distribuzione di  $X$  e  $Y$  possiamo calcolarne il valore atteso:

$$\begin{aligned} E[X|Y = y] &= \int_0^1 x \frac{6x(2-x-y)}{4-3y} dx = \frac{1}{4-3y} \int_0^1 x(6x(2-x-y)) dx = \\ &= \frac{5/2 - 2y}{4-3y} \end{aligned}$$

Un fatto abbastanza importante da notare è che, come già detto in precedenza, la media di una distribuzione congiunta dipendente da una certa variabile  $Y$  è effettivamente una funzione di  $y$  del tipo  $\psi(y)$

**Esempio 3** : consideriamo stavolta la seguente distribuzione congiunta:

$$f_{X,Y}(x,y) = \begin{cases} 1/2ye^{-xy}, & \text{se } x > 0 \\ 0, & \text{altrimenti} \end{cases}$$

Quanto vale il valore atteso  $E[e^{X/2}|Y = 1]$

Anche in questo caso cominciamo andando a calcolare il valore della distribuzione marginale in  $Y$ :

$$f_Y(y) = \int_0^{+\infty} 1/2ye^{-xy} dx$$

Ma noi sappiamo che  $Y = 1$  dunque abbiamo che

$$f_Y(y = 1) = \int_0^{+\infty} 1/2e^{-x} dx = \frac{1}{2}$$

Allora il valore atteso è:

$$E[e^{X/2}|Y=1] = \int_0^{+\infty} e^{x/2} \cdot e^{-x} dx = \int_0^{+\infty} e^{-x/2} dx = 2$$

Abbiamo visto alcuni esempi relativi alle attese condizionate. Cerchiamo di generalizzare partendo da un piccolo esempio. Supponendo di avere un esperimento  $\varphi(m)$  con tre esiti possibili  $\varphi(m_1)$ ,  $\varphi(m_2)$ ,  $\varphi(m_3)$ .

L'attesa è condizionata dal risultato dell'esperimento. Ovvero vale che  $E[N|M=m] = \phi(m)$ .

La generalizzazione da fare è la seguente: se siamo effettivamente a conoscenza del valore assunto da  $M$  allora otteniamo un valore atteso che è funzione di  $m$ . Se invece generalizziamo ulteriormente e diciamo che non sappiamo quale valore assuma  $M$ , allora  $E[N|M]$  è invece una variabile aleatoria, e non più un valore o una funzione di  $m$ . Riassumiamo tutto quello che sappiamo riguardo alla "forma" del valore atteso in diversi casi tramite la Tabella 1

Forma	Risultato	Forma	Risultato
$E[N]$	valore	$E[X]$	valore
$E[N M=m]$	funzione di $m$	$E[X Y=y]$	funzione continua di $m$
$E[N M]$	variabile aleatoria	$E[X Y]$	variabile aleatoria

Tabella 1: Risultato del valore atteso in funzione della forma della variabile condizionante  $M$  nel caso discreto(sinistra) e continuo(destra)

Vediamo adesso un importante teorema

**Teorema 9.1** (della doppia attesa). *Date due variabili aleatorie  $X$  e  $Y$ , è verificata la seguente equazione*

$$E[X] = E[ E[X|Y] ] \quad (12)$$

*Dimostrazione.* consideriamo  $X, Y$  variabili aleatorie discrete. Sappiamo che, dalla definizione di valore atteso:

$$\begin{aligned} E[ E[X|Y] ] &= \sum_y E[X|Y=y]P(Y=y) = \\ &= \sum_y \sum_x x \cdot P(X=x|Y=y)P(Y=y) = \sum_x x \sum_y P(X=x, Y=y) = \\ &= \sum_x xP(X=x) = E[X] \end{aligned}$$

Analogamente possiamo dimostrarlo nel continuo:

$$\begin{aligned} E[ E[X|Y] ] &= \int_{\mathbb{R}} E[X|Y=y]f_Y(y)dy = \int_{\mathbb{R}} \int_{\mathbb{R}} xf_{X|Y}(x|y)f_Y(y)dx dy = \\ &= \int_{\mathbb{R}} x dx \int_{\mathbb{R}} f_{X,Y}(x,y)dy = \int_{\mathbb{R}} x \cdot f_X(x)dx = E[X] \end{aligned} \quad \square$$

Vediamo alcuni esempi su cui applicare il teorema della doppia attesa che abbiamo appena dimostrato:

**Esempio 4** : supponiamo di avere una variabile che segua una distribuzione geometrica di parametro  $p$ . Ovvero:

$$N \sim \text{Geom}(p).$$

Quanto vale  $E[N]$ ?

Normalmente potremmo rispondere a questa domanda con la classica definizione di media:  $E[N] = \sum_{n=1}^{\infty} np(1-p)^n = \frac{1}{p}$ .

Vogliamo invece utilizzare il teorema della doppia attesa. Introduciamo una variabile  $Y$  tale che:

$$Y = \begin{cases} 0, & \text{se } N \text{ fallisce} \\ 1, & \text{altrimenti} \end{cases}$$

Allora dal teorema della doppia attesa abbiamo che:

$$\begin{aligned} E[N] &= E[E[N|Y]] = \underbrace{E[N|Y=1]}_{=1} \underbrace{P(Y=1)}_{=p} + \underbrace{E[N|Y=0]}_{=1+E[N]} \underbrace{P(Y=0)}_{=1-p} = \\ &= p + (1-p)(1+E[N]) = p + 1 + E[N] - p - pE[N] = 1 + E[N] - pE[N] \end{aligned}$$

Dunque

$$E[N] = 1 + E[N] - pE[N] \iff pE[N] = 1 \iff E[N] = \frac{1}{p}$$

**Esempio 5** : supponiamo che un minatore sia in una miniera con tre porte.

La prima gli permette di raggiungere l'uscita in 2 ore. La seconda lo fa ritornare al punto di inizio dopo averlo fatto girare per 3 ore, mentre l'ultima lo fa tornare al punto di inizio dopo avere girato per 5 ore.

Quale è il tempo medio di uscita dalla miniera?

Introduciamo la variabile  $X = \{\text{tempo di salvezza}\}$  e aggiungiamo la variabile  $Y = \{\#\text{porta}\}$ . Anche in questo caso, tramite il teorema della doppia attesa abbiamo che:

$$\begin{aligned} E[X] &= E[E[X|Y]] = \\ &= E[X|Y=1]P(Y=1) + E[X|Y=2]P(Y=2) + E[X|Y=3]P(Y=3) \end{aligned}$$

Ma supponendo che le porte vengano scelte in modo equiprobabile allora:

$$\begin{aligned} E[X|Y=1] \frac{1}{3} + E[X|Y=2] \frac{1}{3} + E[X|Y=3] \frac{1}{3} &= \\ = (2) \frac{1}{3} + (3 + E[X]) \frac{1}{3} + (5 + E[X]) \frac{1}{3} \end{aligned}$$

Dunque:

$$E[X] = \frac{1}{3}(10 + 2E[X]) \iff E[X] = 10$$

### Esercizi di fine lezione

**Esercizio 3** : supponiamo di avere una variabile  $N$  con  $N \sim \text{Poisson}(\lambda)$

Quanto vale  $E[\sum_{i=1}^N X_i]$ ?

$$E[ E[\sum_{i=1}^n X_i | N = n] ]$$

Però sappiamo che  $E[\sum_{i=1}^N X_i | N = n] = E[\sum_{i=1}^n X_i] = nE[X_i]$

**Esercizio 4** : consideriamo una agenzia di polizze assicurative. Ogni persona che entra nell'agenzia sottoscrive una polizza con probabilità  $p$ . Se il numero di clienti in un giorno segue una distribuzione di Poisson, quale è la probabilità che nessuno sottoscriva una polizza in un giorno?

Consideriamo le due variabili  $X = \{\text{\#polizze vendute}\}$  e la variabile  $N = \{\text{\#clienti in un giorno}\}$ . Di fatto stiamo cercando la  $P(X = 0)$ . Andiamo a considerare:

$$\begin{aligned} P(X = 0) &= \sum_n P(X = 0 | N = n) P(N = n) = \sum_n (1-p)^n \lambda \frac{e^{-\lambda}}{n!} = \\ &= e^{-\lambda} \underbrace{\sum_n (1-p)^n \lambda \frac{1}{n!}}_{=e^{\lambda(1-p)}} = e^{-\lambda p} \end{aligned}$$

Se invece avessimo voluto che le polizze stipulate ogni giorno fossero  $k$ , avremmo avuto un calcolo simile ma leggermente più complesso:

$$\begin{aligned} P(X = k) &= \sum_{n \geq k} P(X = k | N = n) P(N = n) = \\ &= \sum_{n \geq k} \lambda^n \frac{e^{-\lambda}}{n!} p^k (1-p)^{n-k} \frac{n!}{k!(n-k)!} = \sum_{n \geq k} (\lambda p)^k \frac{(1-p)^{n-k}}{(n-k)!} \lambda^{n-k} \frac{e^{-\lambda}}{k!} = \\ &= (\lambda p)^k \frac{e^{-\lambda}}{k!} \underbrace{\sum_{n \geq k} \frac{(\lambda(1-p))^{n-k}}{(n-k)!}}_{=e^{\lambda(1-p)}} = \frac{(\lambda p)^k}{k!} e^{-\lambda p} \end{aligned}$$

07 / 11 /2017

Consideriamo ancora qualche esempio sulle attese condizionate

**Esempio 6** : supponendo che in una certa elezione, un candidato  $A$  prenda  $n$  voti mentre un candidato  $B$  prenda  $m$  voti, dove  $m < n$ . Denotiamo con

$P_{n,m}$  la probabilità che dal primo voto in poi,  $A$  sia sempre avanti a  $B$ . Quanto vale  $P_{n,m}$ ?

Per risolverlo andiamo a condizionare su una nuova variabile. Chiameremo  $A_{avanti}$  il fatto che  $A$  sia sempre avanti e  $L_A$  il fatto che  $A$  prenda l'ultimo voto (e lo stesso faremo per  $L_B$ ). Ma allora:

$$\begin{aligned} P_{n,m} &= P(A_{avanti}|L_A)P(L_A) + P(A_{avanti}|L_B)P(L_B) = \\ &= P_{n-1,m} \frac{n}{n+m} + P_{n,m-1} \frac{m}{n+m} \end{aligned}$$

Dunque potremmo proseguire iterando il procedimento

**Esempio 7** : supponiamo di lanciare una moneta un certo numero di volte. Quale è la probabilità che ad un certo punto si abbiano tante teste quante croci?

Indichiamo con  $P_{n,m}$  la probabilità che si siano verificate  $n$  teste e  $m$  croci. Indichiamo con  $I = n$  l'istante in cui ci troviamo, e con  $T_{a,b}$  il fatto che si siano verificate  $a$  teste nei primi  $b$  lanci. Quello che cerchiamo è:

$$P_{n,n}(I = 2n) = P(I = 2n|T_{n,2n})P(T_{n,2n}) = P_{n,n-1} \binom{2n}{n} p^n (1-p)^n$$

## 10 Catene di Markov discrete

### 10.1 Processi stocastici

Diremo che un *processo stocastico* è una famiglia di variabili aleatorie definite su uno spazio  $\Omega$  a valori in uno spazio  $S$ , anche se solitamente viene utilizzato  $\mathbb{R}$ . Dunque definiamo il processo stocastico come segue:

$$\{X_t\}, t \in T \quad (13)$$

Dunque per ogni  $t \in T$ ,  $X_t$  è una variabile aleatoria.

Se fissiamo l'esito del nostro esperimento, dunque imponiamo un esito  $\omega$ , allora notiamo che  $X_t(\omega)$  è di fatto una funzione del tempo.

Per comprendere cosa siano e come introduciamo i processi Markoviani andiamo a considerare il seguente esempio

**Esempio 1** : supponiamo di avere il processo stocastico  $\{X_n\}$  con  $n \in [0, 10]$ .

Per fare un'analisi completa dovremmo considerare tutte le variabili aleatorie  $X_1, \dots, X_{10}$ . Purtroppo non essendo necessariamente indipendenti tra loro dovremmo anche andare a considerare tutte le distribuzioni condizionate.

In generale un esempio vagamente complesso come quello appena mostrato rende molto difficile fare un'analisi del problema. Per questo introduciamo i processi markoviani

## 10.2 Processi markoviani

L'intuizione nei processi markoviani consiste nel semplificare l'intera questione. Intuitivamente diciamo che *il futuro è indipendente dal passato se mi è noto il presente*. L'idea è quindi quella di tenere in considerazione solo un limitato numero di eventi vicini al tempo presente, tralasciando tutti gli altri

**Esempio 1** : il valore in borsa di domani di un titolo certamente dipenderà in qualche modo anche dal valore che aveva sei mesi fa, ma tenderà a dipendere maggiormente dal valore che assunto negli ultimi giorni, estremizzando anche solo dal valore di oggi.

È chiaro che questa assunzione semplifica la realtà approssimandola in un qualche modo, ma ci permette di risolvere molto più velocemente problemi complessi commettendo un errore limitato.

Dal punto di vista probabilistico diremo che:

$$P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1}) \quad (14)$$

Andiamo ad applicare questa definizione ad una serie di eventi congiunti fra loro legati da variabili markoviane:

$$\begin{aligned} P(X_n = x_n, \dots, X_0 = x_0) &= \\ &= P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) P(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \end{aligned}$$

Ma dalla definizione di processo markoviano:

$$\begin{aligned} P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) P(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) &= \\ &= P(X_n = x_n | X_{n-1} = x_{n-1}) P(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \end{aligned}$$

Iterando il procedimento abbiamo che:

$$\begin{aligned} P(X_n = x_n, \dots, X_0 = x_0) &= \\ &= P(X_n = x_n | X_{n-1} = x_{n-1}) \dots P(X_1 = x_1 | X_0 = x_0) P(X_0 = x_0) \end{aligned}$$

Dunque possiamo generalizzare dicendo che, per i processi markoviani vale che  $P(X_n = j | X_{n-1} = i)$ .

Da queste definizioni possiamo ricavare un utile strumento, la *matrice di transizione*. Supponendo che la quantità  $P(X_n = j | X_{n-1} = i)$  indichi la probabilità di passare da uno stato  $n-1$  a valore  $i$  ad uno stato  $n$  a valore  $j$ , possiamo costruire una matrice dove, in posizione  $(i, j)$  inseriamo la probabilità di transire dallo stato  $i$  allo stato  $j$ . Cerchiamo di comprenderlo con l'esempio che segue:

**Esempio 1** : supponiamo di voler comprendere se domani piovierà basandoci solo sul tempo atmosferico di oggi. Ne possiamo ricavare una tabella riassuntiva, ovvero la Tabella 2. Dunque, ad esempio, se oggi piove, la probabilità che piovano domani è  $\alpha$ . Allo stesso modo, se oggi non piove, la



Oggi	Domani	Probabilità
piove	piove	$\alpha$
$\neg$ piove	piove	$\beta$

Tabella 2: Tabella relativa alla previsione del tempo, tenendo conto solo di oggi

probabilità che piova domani è  $\beta$ . Ne possiamo ricavare la seguente matrice di transizione:

$$P = \begin{bmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

Dobbiamo notare un paio di fatti importanti riguardo alla matrice di transizione. Visto che ogni elemento rappresenta una probabilità, ne segue che tutti gli elementi della matrice saranno  $P_{i,j} \geq 0$ . Inoltre possiamo notare che la somma su ogni riga della matrice è pari a 1. Dunque vale che:

$$\sum_j P_{ij} = 1$$

Consideriamo qualche altro esempio:

**Esempio 2** : supponiamo di avere un sistema di trasmissione di bit che mantiene inalterato(non distorce) il bit ricevuto con probabilità  $p$ . Se  $X_n$  è il bit ricevuto al tempo  $n$ -esimo, allora la matrice di transizione è:

$$P = \begin{bmatrix} p & 1 - p \\ 1 - p & p \end{bmatrix}$$

**Esempio 3** : supponiamo di riprendere l'esempio del tempo atmosferico e di voler complicare un po' le cose. Supponiamo che il tempo atmosferico dipenda sia dal tempo di oggi che da quello di ieri, la tabella dunque è

Stato	Ieri	Oggi	Domani	Probabilità
$y_n = 0$	piove	piove	piove	0.7
$y_n = 1$	$\neg$ piove	piove	piove	0.5
$y_n = 2$	piove	$\neg$ piove	piove	0.4
$y_n = 3$	$\neg$ piove	$\neg$ piove	piove	0.2

Tabella 3: Tabella relativa alla previsione del tempo tenendo conto del giorno precedente

La matrice di transizione in questo caso è:

$$P = \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix}$$

I valori mostrati in questa tabella necessitano di alcune spiegazioni ulteriori. Consideriamo la prima riga della matrice. La prima casella indica la possibilità di transire dallo stato 0 (ieri ha piovuto, oggi ha piovuto) allo stato 0 (oggi ha piovuto domani piove). Noi sappiamo che la probabilità che questo succeda è 0.7 dunque impostiamo il primo valore della tabella proprio a 0.7.

Consideriamo invece la seconda casella: la probabilità di transire dallo stato zero allo stato uno è nulla. Questo perché se ieri ha piovuto e oggi piove (stato 0) non è possibile che oggi non piova e domani piova, perché differiscono sul valore di oggi. Questo si verifica perché non esiste una compatibilità fra il valore di oggi dello stato zero e il valore di ieri dello stato uno.

Per quanto riguarda la terza colonna vale che passare dallo stato zero allo stato due, la probabilità è pari a  $1 - 0.7 = 0.3$ . Infatti notiamo che il valore di oggi dello stato zero coincide con il valore di ieri dello stato due. Quindi questo ci permette di transire senza problemi da uno stato all'altro. Non abbiamo incompatibilità. Invece abbiamo problemi a passare dallo stato zero allo stato tre per lo stesso motivo per cui avevamo problemi a passare dallo stato zero allo stato uno.

## Esercizi di fine lezione

**Esercizio 1** : consideriamo un processo stocastico della forma  $\{X_n\}$  con  $n \geq 0$   
 Supponiamo che lo spazio degli stati sia  $\{0, 1, 2\}$   
 Supponendo di considerare:

$$P(X_n = j | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \begin{cases} P_{i,j}^I, & \text{se } n \text{ pari} \\ P_{i,j}^{II}, & \text{se } n \text{ dispari} \end{cases}$$

Quella appena descritta è una catena di Markov?

La risposta è no perché non possiamo effettuare ragionamenti su  $X_{n-1} = i_{n-1}$  se non sappiamo se  $n$  sia pari o dispari.

Una soluzione al problema consiste nell'estendere lo spazio degli stati:

$$\{0^p, 1^p, 2^p, 0^d, 1^d, 2^d\}$$

Dove  $k^p$  indica un esito pari a  $k$  ottenuto ad un tempo pari, mentre  $k^d$  indica un esito ottenuto ad un tempo dispari. Costruiamo di conseguenza la matrice di transizione:

$$P = \begin{bmatrix} 0 & 0 & 0 & P_{00}^{II} & P_{01}^{II} & P_{02}^{II} \\ 0 & 0 & 0 & P_{10}^{II} & P_{11}^{II} & P_{12}^{II} \\ 0 & 0 & 0 & P_{20}^{II} & P_{21}^{II} & P_{22}^{II} \\ P_{00}^I & P_{01}^I & P_{02}^I & 0 & 0 & 0 \\ P_{10}^I & P_{11}^I & P_{12}^I & 0 & 0 & 0 \\ P_{20}^I & P_{21}^I & P_{22}^I & 0 & 0 & 0 \end{bmatrix}$$

Per comprendere meglio la matrice, consideriamo la prima riga. Partiamo dalla prima casella: transire da uno stato in cui abbiamo ottenuto uno zero al tempo pari, ad uno in cui otteniamo uno zero ad un tempo pari è impossibile perché nel mezzo deve necessariamente intercorrere un tempo dispari! Analogamente per le due caselle successive, se otteniamo uno zero al tempo pari non potremmo mai ottenere un uno o un due al tempo pari come stato successivo, perché nel mezzo deve intercorrere un tempo dispari. Invece transire da uno stato in cui otteniamo uno zero a tempo pari ad uno in cui otteniamo uno zero a tempo dispari è possibile! E succede esattamente con probabilità  $P_{01}^{II}$ . Lo stesso possiamo dire per le due caselle successive.

**Esercizio 2** : prendiamo una variabile aleatoria di Poisson il cui parametro è però una variabile aleatoria esponenziale. Dunque  $N \sim \text{Poisson}(\Lambda)$  dove  $\Lambda \sim \text{exp}(1)$  dunque  $f_\Lambda(x) = e^{-x}$ . Quanto vale  $P(N = n)$ ?

$$\begin{aligned} P(N = n) &= \int_0^\infty P(N = n | \Lambda = x) f_\Lambda(x) dx = \\ &= \int_0^\infty \underbrace{\frac{x^n}{n!} e^{-x}}_{=P(N=n|\Lambda=x)} e^{-x} dx \end{aligned}$$

08 / 11 / 2017

### 10.3 Equazione di Chapman Kolmogorov

Fino ad ora abbiamo indicato con  $P_{ij}$  la probabilità di transire da uno stato  $i$  ad uno stato  $j$  successivamente ad un certo passo. Possiamo generalizzare e considerare l'esecuzione di più passi e indicarla con  $P_{ij}^n$ . Come possiamo calcolare la probabilità in questo caso? L'equazione di Chapman Kolmogorov ce lo dice con il seguente teorema

**Teorema 10.1.**

$$P_{i,j}^{n+m} = \sum_k P_{i,k}^n P_{k,j}^m \quad (15)$$

$\forall n, m \geq 0$  e anche  $\forall i, j$

$$\begin{aligned} \text{Dimostrazione. } P_{i,j}^{n+m} &= P(X_{n+m} = j | X_0 = i) = \\ &= \sum_{k=0}^{+\infty} P(X_{n+m} = j, X_n = k | X_0 = i) = \\ &= \sum_{k=0}^{+\infty} P(X_{n+m} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i) = \sum_k P_{k,j}^m P_{i,k}^n \quad \square \end{aligned}$$

Se andiamo a denotare la matrice del passo  $n$ -esimo con il simbolo  $P^{(n)}$ , allora notiamo che  $P^{(n+m)} = P^{(n)} \cdot P^{(m)}$ . In particolare  $P^{(2)} = P \cdot P = P^2$ , e più in generale  $P^{(n)} = P^n$ . Andiamo a comprenderlo con l'esempio che segue:

**Esempio 1** : riprendiamo l'esempio del tempo atmosferico in cui teniamo conto solo del tempo di oggi per prevedere quello di domani. Questa volta vogliamo però prevedere quale sia la probabilità che, avendo piovuto oggi, piovra fra esattamente quattro giorni. Supponiamo che  $\alpha = 0.7$  e  $\beta = 0.4$  e ricostruiamo la matrice e otteniamo che:

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Per prevedere il tempo fra quattro giorni ci è sufficiente calcolare  $P^4$ . Ovvero:

$$P^4 = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix}$$

La probabilità che a noi interessa è  $P_{00}^4 = 0.5749$

Riprendiamo anche l'esempio in cui tenevamo di conto del giorno precedente. Sapendo che ha piovuto lunedì e martedì, quale è la probabilità che piovra giovedì? Per calcolare la probabilità che piovra mercoledì basterebbe la matrice di transizione  $P$ . Se invece vogliamo sapere riguardo a giovedì dobbiamo calcolare  $P^2$ . Dunque:

$$P^2 = \begin{bmatrix} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{bmatrix}$$

Sapendo che lunedì e martedì ha piovuto (stato 0), la probabilità che cerchiamo si trova nella prima riga. Ad ogni modo la probabilità che vogliamo è data da  $P_{00}^2 + P_{01}^2 = 0.49 + 0.12 = 0.61$

Questo perché se transiamo nello stato 0 dopo due passi, stiamo dicendo che giovedì pioverà e mercoledì ha piovuto. Ma d'altronde di cosa succede mercoledì non ci interessa, quindi anche transire nello stato 1 ci va bene visto che comunque nello stato 1 giovedì piove e non ci importa che mercoledì non abbia piovuto

## 10.4 Rappresentare le catene di Markov

Nelle prossime pagine vedremo un metodo per rappresentare le catene di Markov in modo da poter dedurre proprietà riguardo ad esse. Prima di farlo dobbiamo però fornire qualche definizione:

Diremo che uno stato  $j$  è *accessibile* da uno stato  $i$  se  $P_{i,j}^n > 0$ , ovvero se dopo un numero  $n$  di passi transiamo dallo stato  $i$  allo stato  $j$ . Indicheremo

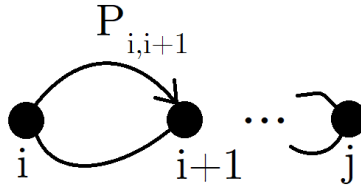


Figura 23: Rappresentazione del concetto di accessibilità

l'accessibilità con la notazione  $i \rightarrow j$ . Graficamente potremmo rappresentare la ;accessibilità come in Figura 23

Se  $j$  non è accessibile da  $i$  vale che:

$$P(X_n = j | X_0 = i) = P\left(\bigcup_{n=1}^{\infty} \{X_n = j\} | X_0 = i\right) \leq \sum P(X_n = j | X_0 = i) = \\ = \sum_n P_{i,j}^n = 0$$

Diremo che due stati sono comunicanti se  $i$  è accessibile da  $j$  e viceversa. Indicheremo questo fatto come segue  $i \leftrightarrow j$ . Notiamo che la comunicazione fra stati definisce una relazione matematica che gode delle seguenti proprietà:

- Riflessività: ogni stato comunica con sé stesso
- Simmetria: vale che  $i \leftrightarrow j \iff j \leftrightarrow i$
- Transitività: se  $i \leftrightarrow k$  e  $k \leftrightarrow j$  allora  $i \leftrightarrow j$

Le tre proprietà appena elencate definiscono una *classe di equivalenza*. Infatti diremo che due stati comunicanti apparterranno alla stessa classe. Nel caso in cui tutti gli stati comunichino con tutti, diremo che abbiamo una *catena irriducibile*. Vediamo come mettere a frutto le definizioni fornite. Cominciamo convertendo una matrice di transizione in un grafo

**Esempio 1** : supponiamo di avere la seguente matrice di transizione

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.25 & 0.25 \\ 0 & 0.33 & 0.67 \end{bmatrix}$$

La rappresentazione della matrice di transizione descritta è mostrata in Figura 24

Abbiamo indicato con 0,1,2 gli stati possibili e indicato su ogni arco la probabilità relativa. Notiamo che in questo caso abbiamo una catena irriducibile perché tutti gli elementi sono raggiungibili da un qualsiasi altro elemento. Dunque abbiamo una sola classe

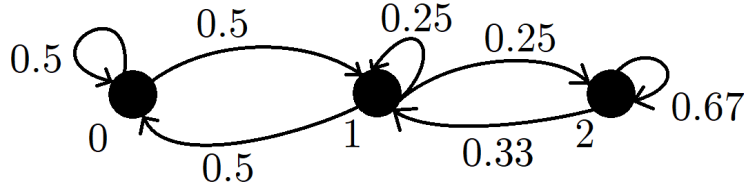


Figura 24: Rappresentazione della matrice di transizione dell'esempio 1

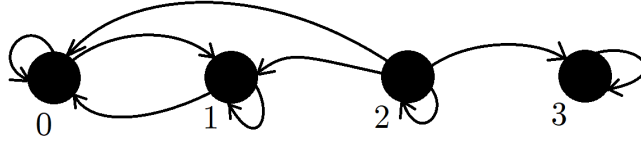


Figura 25: Rappresentazione della matrice di transizione dell'esempio 2

**Esempio 2** consideriamo la seguente matrice di transizione

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

La rappresentazione di questa matrice viene mostrata in Figura 25

Notiamo che questa catena possiede tre classi: la classe dei nodi  $\{0, 1\}$ , la classe contenente il nodo  $\{2\}$  e la classe contenente il nodo  $\{3\}$ .

Una distinzione che dobbiamo fare è quella tra stati *ricorrenti* e stati *transienti*. Consideriamo la probabilità  $f_i$  di tornare in un certo stato una volta che ci siamo passati. Diremo che:

$$f_i = \begin{cases} 1, & \text{se stato ricorrente} \\ 0, & \text{se stato transiente} \end{cases} \quad (16)$$

Dunque stiamo dicendo che se uno stato è ricorrente, vi passeremo infinite volte. Il caso di uno stato transiente non ci interessa particolarmente. Ci concentreremo sugli stati ricorrenti. Come fare a vedere se uno stato è ricorrente?

**Teorema 10.2.** *Se la serie  $\sum_n P_{i,i}^n$  diverge, allora lo stato  $i$  è uno stato ricorrente. Altrimenti è uno stato transiente*

*Dimostrazione.* Consideriamo una variabile aleatoria bernoulliana  $A_n$  tale che

$$A_n = \begin{cases} 1, & \text{se } X_n = i \\ 0, & \text{se } X_n \neq i \end{cases}$$

Consideriamo la serie  $\sum A_n = \# \text{volte che torno nello stato } i$   
Allora:

$$E\left[\sum_{n=1}^{\infty} A_n | X_0 = i\right] = \sum_{n=1}^{\infty} E[A_n | X_0 = i] = \sum_{n=1}^{\infty} P(X_n = i | X_0 = i) = \sum_n P_{i,i}^n \quad \square$$

Enunciamo due ulteriori teoremi di cui non forniamo una dimostrazione

**Teorema 10.3.** *Se  $i$  è uno stato ricorrente e comunica con  $j$ , allora anche  $j$  è ricorrente*

**Teorema 10.4.** *In una catena irriducibile tutti gli stati sono ricorrenti*

**Esempio 3 :** consideriamo i due grafi mostrati in Figura 24 e in Figura 25 e valutiamo quali stati siano transienti o ricorrenti. Partiamo dal primo grafo. Come abbiamo detto la catena è irriducibile, dunque abbiamo una sola classe e per il teorema appena mostrato tutti gli stati sono ricorrenti. Per quanto riguarda il secondo grafo abbiamo tre classi:  $\{0, 1\}$ ,  $\{2\}$  e  $\{3\}$ . La classe  $\{0, 1\}$  è ricorrente visto che ci permette di passare indefinitamente da uno stato all'altro. Per quanto riguarda la classe  $\{2\}$  abbiamo che è transiente mentre la classe  $\{3\}$  è ricorrente.

**Esempio 4 :** consideriamo il problema del random walk. Una persona può muoversi a destra con probabilità  $p$  o a sinistra con probabilità  $1 - p$ . Vogliamo studiare la probabilità che dopo  $2n$  passi siamo ritornati al punto di origine. Dunque, supponendo di avere a disposizione gli stati  $s = 0, \pm 1, \pm 2, \dots$  vogliamo studiare la serie

$$\sum_n P_{0,0}^{2n}$$

Sappiamo che

$$P_{0,0}^{2n} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!n!} p^n (1-p)^n$$

Ma dall'approssimazione di Stirling sappiamo che  $n! \sim \sqrt{2\pi n} e^{-n} n^{n/2}$  e dunque andando a sostituire

$$P_{0,0}^{2n} \sim \frac{[4p(1-p)]^n}{\sqrt{\pi n}}$$

Dunque abbiamo che

$$\sum_n P_{0,0}^{2n} \sim \sum_n \frac{[4p(1-p)]^n}{\sqrt{\pi n}}$$

Notiamo che  $4p(1-p) \leq 1$  e dunque questa serie diverge (dando luogo ad uno stato ricorrente) soltanto se  $p = 0.5$ , mostrandoci matematicamente un risultato intuitivo: se la probabilità di muoversi a destra e sinistra è

la stessa, allora lo stato iniziale è ricorrente. Ma se la probabilità è anche solo leggermente sbilanciata verso uno dei due movimenti, allora la serie convergerà e non ripasseremo dallo stato iniziale, che dunque è solo uno stato transiente

14 / 11 / 2017

Fino a questo momento abbiamo effettuato una distinzione tra stati transienti e ricorrenti. In realtà possiamo fare un'ulteriore distinzione fra varie tipologie di stati ricorrenti.

Indicando come  $T$  la variabile aleatoria che descriva il tempo di ritorno su di un certo stato, allora abbiamo due tipi di stati ricorrenti:

- Positivo ricorrenti: quelli per cui vale che  $E[T] < \infty$
- Nullo ricorrenti: quelli per cui vale che  $E[T] = \infty$

Ad esempio lo stato iniziale del random walk è uno stato nullo ricorrente

## 10.5 Stabilizzazione

Una questione interessante riguarda la stabilizzazione dei fenomeni. Sarebbe interessante se dopo un certo periodo di tempo un evento tendesse a stabilizzarsi. Vorrebbe dire che oltre un certo momento potremmo aspettarci una situazione senza mutazioni.

**Esempio 1** : nel caso del tempo atmosferico avevamo calcolato la matrice di transizione dopo quattro passi, ovvero  $P^4$

$$P^4 = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix}$$

Notiamo che i valori appartenenti ad una stessa colonna sono simili fra loro. Se andassimo a calcolare  $P^8$  si assomiglierebbero ancora di più.

Se esistesse potrebbe essere interessante andare a calcolare il limite:

$$\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n$$

Ovviamente questo limite può esistere per stati ricorrenti (visto che negli stati transienti non torniamo) e per calcolarlo dobbiamo prima essere sicuri di non avere funzioni periodiche come il seno e il coseno<sup>5</sup>.

Uno stato si dice di *periodo*  $d$  se  $P_{ij}^n = 0$  con  $n$  non è divisibile per  $d$ . Ovvero:

$$d_i = \text{MCD}\{n : P(X_n = i | X_0 = i) > 0\}$$

se  $d_i = 1$  la catena è aperiodica.

---

<sup>5</sup>Visto che il limite di funzioni periodiche non converge



Notiamo che se uno stato periodico comunica con un certo insieme di stati, anch'essi saranno periodici. Dunque la periodicità è una proprietà della classe.

Nel complesso ci interessano gli stati ricorrenti aperiodici e positivo ricorrenti. Questi stati formano all'interno del grafo una *catena ergodica*. In particolare consideriamo le catene irriducibili perché quando una catena non è irriducibile possiamo comunque spezzarla in più sotto-catene irriducibili. Per le catene ergodiche abbiamo un importante teorema

**Teorema 10.5.** *Una catena irriducibile ergodica ammette una distribuzione limite  $\{\pi_j\}$*

$$\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n \quad (17)$$

Inoltre  $\pi_j$  è l'unica soluzione non negativa del sistema  $\pi_j = \sum_i \pi_i P_{ij}$  con  $\sum_j \pi_j = 1$ , ovvero la somma per ogni riga è sempre uguale a 1

*Dimostrazione.* assumiamo che il limite esista<sup>6</sup>. Dunque

$$\exists \lim_{n \rightarrow \infty} P(X_{n+1} = j) = \pi_j.$$

Allora consideriamo:

$$P(X_{n+1} = j) = \sum_i P(X_{n+1} = j | X_n = i) P(X_n = i)$$

Visto che il limite esiste lo calcoliamo e otteniamo che:

$$\pi_j = \sum_i P_{ij} \pi_i$$

se vale che  $\sum_j \pi_j = 1$  □

Il teorema appena mostrato ci dice qualcosa di importante, ovvero che il valore di una certa  $P_{ij}$  dopo  $n$  iterazioni è una combinazione lineare dei  $P$  e dei  $\pi$ . Vediamone le potenzialità sull'esempio che segue

**Esempio 2 :** consideriamo nuovamente l'esempio della pioggia tenendo conto soltanto del tempo presente. inizialmente la matrice da noi introdotta era:

$$P = \begin{bmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

Per valutare il comportamento del sistema dopo  $n$  passi ci è sufficiente considerare il risultato del seguente sistema lineare:

$$\begin{cases} \pi_0 = \alpha \pi_0 + \beta \pi_1 \\ \pi_1 = (1 - \alpha) \pi_0 + (1 - \beta) \pi_1 \end{cases} \quad \begin{cases} \dots \\ \dots \end{cases} \quad \begin{cases} \pi_0 = \frac{\beta}{(1 + \beta - \alpha)} \\ \pi_1 = \frac{1 - \alpha}{1 + \beta - \alpha} \end{cases}$$

---

<sup>6</sup>Questa parte è complessa da dimostrare, la assumiamo vera

**Esempio 3** : un approccio possibile per studiare la società consiste nel valutare il benessere dei cittadini e suddividerli in tre classi: classe bassa, classe media e classe alta. Con la seguente matrice indichiamo la probabilità di passare da una classe ad un'altra da una generazione alla successiva:

$$P = \begin{bmatrix} 0.45 & 0.48 & 0.07 \\ 0.05 & 0.7 & 0.25 \\ 0.01 & 0.5 & 0.49 \end{bmatrix}$$

Dove ad esempio  $P[3][2]$  indica la probabilità di passare dalla classe alta alla classe media da una generazione alla successiva.

Assumendo che la catena originata sia aperiodica, otteniamo il sistema:

$$\begin{cases} \pi_0 = 0.45\pi_0 + 0.05\pi_1 + 0.01\pi_2 \\ \pi_1 = 0.48\pi_0 + 0.7\pi_1 + 0.52\pi_2 \\ \pi_2 = 0.07\pi_0 + 0.25\pi_1 + 0.49\pi_2 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases} \quad \begin{cases} \dots \\ \dots \\ \dots \\ \dots \end{cases} \quad \begin{cases} \pi_0 = 0.07 \\ \pi_1 = 0.62 \\ \pi_2 = 0.31 \end{cases}$$

Notiamo che della probabilità limite  $\pi_j$  possiamo dare tre interpretazioni diverse:

- Probabilità limite
- Probabilità stazionaria: infatti vale che

$$P(X_0 = j) = \pi_j \Rightarrow P(X_n = j) = \pi_j$$

Questo perché:

$$P(X_n = j) = \sum_i P(X_n = j | X_{n-1} = i) P(X_{n-1} = i) = \sum_i P_{ij} \pi_j = \pi_j$$

- Tempo speso in uno stato  $j$ : infatti, detto  $m_{jj}$  il numero di passi per tornare nello stato  $j$  partendo dallo stato  $j$ , allora osserviamo che la catena spende un quanto di tempo su  $m_{jj}$  nello stato  $j$ . Dunque

$$\pi_j = \frac{1}{m_{jj}}$$

## Esercizi di fine lezione

**Esercizio 1** : supponiamo che vengano sviluppati due farmaci per curare una malattia. Ogni farmaco ha una probabilità di cura  $P_i$ .

Supponiamo di considerare coppie di pazienti e di definire su di essi le seguenti variabili aleatorie:

$$X_j = \begin{cases} 1, & \text{se il primo paziente della coppia } j\text{-esima guarisce con il farmaco 1} \\ 0, & \text{altrimenti} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{se il secondo paziente della coppia } j\text{-esima guarisce con il farmaco 2} \\ 0, & \text{altrimenti} \end{cases}$$

Terminiamo la sperimentazione quando, per un  $M$  fissato a priori, vale una delle due seguenti condizioni:

$$\begin{aligned} X_1 + \dots + X_n - (Y_1 + \dots + Y_n) &= M \\ X_1 + \dots + X_n - (Y_1 + \dots + Y_n) &= M \end{aligned}$$

In questo caso, ogni volta che un paziente guarisce per il farmaco 1, mi sposto verso  $M$ , ogni volta che un altro guarisce grazie al farmaco 2, ci spostiamo verso  $-M$ . Esiste quindi una forte similitudine tra questo problema e quello del random walk.

21 / 11 / 2017

## 11 Distribuzione esponenziale

### 11.1 Distribuzione esponenziale e memoria

Andiamo a considerare una variabile aleatoria esponenziale di cui ricordiamo la definizione

$$f_E(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x > 0 \\ 0, & \text{altrimenti} \end{cases}$$

$$F_X(x) = 1 - e^{-\lambda x}$$

La variabile aleatoria esponenziale diremo che è *senza memoria*. Pensiamo intuitivamente ad un oggetto, sia una lampadina. Se il tempo di vita della lampadina segue una distribuzione esponenziale, potremmo chiederci quale è la probabilità che continui a funzionare per un tempo pari a  $t$  sapendo che è già vissuta per un tempo pari a  $s$ . Più formalmente vogliamo verificare che

$$P(X > s + t | X > s) = P(X > t) \quad (18)$$

Andiamo a verificarlo:

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \\ &= e^{-\lambda t} = P(X > t) \end{aligned}$$

Da quanto appena detto possiamo anche ricavare che

$$P(X > s + t) = P(X > s)P(X > t) \quad (19)$$

**Esempio 1** : supponiamo che un certo cliente  $X_3$  entri in un ufficio postale in cui servono due impiegate. Quando  $X_3$  entra, entrambe le impiegate stanno già servendo due clienti  $X_1$  e  $X_2$ . Il tempo di servizio di ogni

cliente segue una stessa variabile esponenziale di parametro  $\lambda$ . Quale è la probabilità che  $X_3$  sia l'ultimo cliente ad uscire?

Di fatto quando  $X_1$  o  $X_2$  (indifferentemente) lasciano la posta,  $X_3$  prende il posto che si è liberato. Quindi ci stiamo chiedendo quanto vale:

$$P(X_i > X_3)$$

Con  $i = 1, 2$ . Visto che tutti i clienti seguono una variabile esponenziale, quando  $X_3$  comincia ad essere servito, l'altro cliente  $X_i$  non ha alcuna memoria del fatto che è stato servito fino a quel momento. Dunque, per simmetria, la probabilità che finisca prima  $X_3$  o  $X_i$  è la stessa, ed è pari a 0.5

Dunque

$$P(X_i > X_3) = \frac{1}{2}$$

**Esempio 2** : supponiamo che i danni da pagare per un certo incidente seguano una variabile esponenziale di parametro  $\lambda = 1/1000$ .

L'assicurazione paga i danni solo se la cifra supera i 400 euro. Quanto paga in media l'assicurazione? E quanto vale la varianza su tale cifra?

Per risolvere il problema introduciamo una variabile  $Y = (X - 400)^+$  ovvero la parte positiva di  $X - 400$

Introduciamo quindi l'indicatore:

$$\mathbb{1} = \begin{cases} 1, & \text{se } X \geq 400 \\ 0, & \text{se } X < 400 \end{cases}$$

Andiamo allora a calcolare  $E[Y]$ :

$$E[Y] = E[E[Y|\mathbb{1}]] = E[Y|\mathbb{1} = 1] P(\mathbb{1} = 1) + E[Y|\mathbb{1} = 0] P(\mathbb{1} = 0)$$

Ma poiché in media in un incidente si ha un danno  $E[X] = \frac{1}{\lambda} = 1000$ , allora:

$$\begin{aligned} E[Y|\mathbb{1} = 1] &= 1000 \\ P(\mathbb{1} = 1) &= P(X > 400) = e^{-\frac{400}{1000}} = e^{-0.4} \\ E[Y|\mathbb{1} = 0] &= 0 \\ P(\mathbb{1} = 0) &= 1 - e^{-0.4} \end{aligned}$$

Dunque ne ricaviamo che  $E[Y] = 670,32$

Il fatto che la variabile aleatoria esponenziale è l'unica continua con assenza di memoria può essere verificato come segue.

Supponiamo che esista una variabile aleatoria  $X$  che gode della proprietà di

assenza di memoria. Allora per quanto detto finora abbiamo che

$$\bar{F}(X) = P(X > x)$$

Ma in base a quanto detto finora deve essere vero che

$$\bar{F}(s+t) = \bar{F}(s)\bar{F}(t)$$

Quindi  $\bar{F}(s+t)$  soddisfa l'equazione funzionale  $g(s+t) = g(s)g(t)$ . Quello che emerge è che l'unica soluzione continua a questa equazione è  $e^{-\lambda x}$ . Ovvero

$$\bar{F}(x) = e^{-\lambda x}$$

Ovvero abbiamo che:

$$F(X) = P(X \leq x) = 1 - e^{-\lambda x}$$

Che dunque segue una distribuzione esponenziale.

In realtà in precedenza avevamo già visto un'altra variabile (discreta) che possiede l'assenza di memoria. La variabile geometrica. Non è un caso che anche la geometrica possieda questa proprietà, esiste infatti un legame tra queste due distribuzioni. Vediamolo nel seguito.

Consideriamo una variabile aleatoria  $T = \frac{X}{n}$  dove  $X \sim Geo(p)$ .

$$P(X > i) = P(T > \frac{i}{n}) = (1-p)^i$$

Andando ad eseguire il cambio di variabile  $p = \lambda/n$  e  $i/n = t$  otteniamo

$$(1-p)^i = (1 - \frac{\lambda}{n})^{nt}$$

Andando ad eseguire il limite per  $n \rightarrow \infty$  otteniamo che

$$P(X > tn) = e^{-\lambda t}$$

E dunque anche in questo caso

$$P(X \leq tn) = 1 - e^{-\lambda t}$$

Intuitivamente, aumentando il valore di  $n$  abbiamo ridotto in infinitesimi i quanti di tempo in cui viene calcolata la distribuzione. Dunque possiamo dire che la distribuzione esponenziale è il corrispettivo nel continuo della distribuzione geometrica. E questo spiega come mai abbiamo entrambe l'assenza di memoria.

Consideriamo un ultimo esempio relativo all'assenza di memoria:

**Esempio 3** : consideriamo un oggetto che abbia un tempo di vita  $X$ , che ha funzionato per  $t$  ore. Vogliamo conoscere la probabilità che non sopravviva per un altro tempo  $dt$ . Supponendo che  $X$  abbia una funzione di distribuzione  $F$  e una densità  $f$  definiamo come il rate di fallimento la seguente quantità:

$$r(t) = \frac{f(t)}{1 - F(t)}$$

Nel caso in cui  $X$  sia esponenziale  $r(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$  e dunque il tasso è pari a una costante nel tempo.

## 11.2 Proprietà della distribuzione esponenziale

Verifichiamo adesso alcune proprietà della distribuzione esponenziale che ci serviranno nel seguito

- Cosa succede se avendo due variabili indipendenti  $X_1 \sim \exp(\lambda)$  e  $X_2 \sim \exp(\mu)$ , andiamo a calcolare  $P(X_1 < X_2)$ ? Questo calcolo è relativamente facile se andiamo a condizionare su  $X_1$

$$\begin{aligned} P(X_1 < X_2) &= \int_{-\infty}^{+\infty} P(X_1 < X_2 | X_1 = x) P(X_1 = x) dx = \\ &= \int_{-\infty}^{+\infty} P(X_1 < X_2 | X_1 = x) f_X(x) dx = \int_0^{+\infty} P(X_2 > x) \lambda e^{-\lambda x} dx = \\ &= \int_0^{+\infty} e^{-\mu x} \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{-(\lambda+\mu)x} dx = \frac{\lambda}{\lambda + \mu} \end{aligned}$$

Dunque la probabilità che una sia minore dell'altra dipende solo e soltanto dai parametri  $\lambda$  e  $\mu$

- Supponiamo di avere  $n$  variabili aleatorie esponenziali indipendenti  $X_1, \dots, X_n \sim \exp(\mu_i)$ . Quanto vale  $Y = \min\{X_1, \dots, X_n\}$ ?

Intuitivamente, se ogni variabile rappresenta un tempo di arrivo, il minimo fra i tempi d'arrivo è quello che "arriva per primo". Dunque:

$$\begin{aligned} P(\min\{X_1, \dots, X_n\} > x) &= P(X_1 > x, \dots, X_n > x) = \\ \prod_{i=1}^n P(X_i > x) &= \prod_{i=1}^n e^{-\mu_i x} = e^{-x \sum \mu_i} \end{aligned}$$

Dunque il minimo è ancora una variabile esponenziale il cui parametro è la somma dei parametri

- Sotto le condizioni del punto precedente, quanto vale invece la probabilità che  $X_i$  sia proprio il minimo? Ovvero  
 $P(X_i = \min\{X_1, \dots, X_n\}) = P(X_i < \min\{X_j\})$  con  $j \neq i$

$$P(X_i < \min\{X_j\}) = \frac{\mu_i}{\mu_i + \sum_{j \neq i} \mu_j}$$

## Esercizi di fine lezione

**Esercizio 1** : consideriamo una variabile  $T_i \sim \exp(\mu)$  che esprime il tempo di servizio dell' $i$ -esimo cliente servito. Ogni cliente, prima di abbandonare la coda senza essere servito trascorre in coda un tempo  $\Theta_j \sim \exp(\lambda)$ . Quale è la probabilità che il cliente  $i$ -esimo sia servito?

**Esercizio 2** : supponendo di avere a disposizione le due variabili aleatorie  $X \sim \exp(\lambda)$  e  $Y \sim \exp(\mu)$  chiamiamo  $M = \min\{X, Y\}$ . Quanto vale  $E[MX|M = X]$ ?

$$E[MX|M = X] = E[M^2|M = X] = E[M^2] = \frac{2}{(\lambda + \mu)^2}$$

Questo perché  $M \sim \exp(\lambda + \mu)$  e perché  $E[M] = \frac{1}{\lambda + \mu}$

Se invece ci chiedessimo quanto vale  $E[MX|M = Y]$ ?

In tal caso:

$$E[MX|M = Y] = E[M(M + X')|M = Y] = E[M^2] + E[MX'] = \frac{2}{(\lambda + \mu)^2} + EM \cdot EX' = \frac{2}{(\lambda + \mu)^2} + \frac{1}{\lambda + \mu} \cdot \frac{1}{\lambda}$$

28 / 11 / 2017

## 12 Processi di conteggio e di Poisson

Per convergere verso le catene di Markov a tempo continuo abbiamo la necessità di parlare dei processi di conteggio. Vediamo dapprima i processi di conteggio in generale e successivamente andiamo a vedere un particolare processo di conteggio detto processo di Poisson.

### 12.1 Processi di conteggio

Un *processo di conteggio* è un processo stocastico  $\{N(t)\}_{t \geq 0}$  relativo al tempo  $t$  per il quale vale che

1.  $N(0) = 0$   
Dunque all'inizio il conteggio è pari a zero
2.  $N(t) \geq 0$   
Possiamo contare solo in positivo
3.  $N(s) \leq N(t), s < t$   
Preso un istante di tempo  $t$  successivo ad un istante  $s$ , il conteggio dopo

che è passato il tempo  $t - s$  potrà essere solo aumentato o al più essere rimasto invariato

4.  $N(t) - N(s), s < t$   
Il valore  $N(t) - N(s)$  è il numero di eventi che si verificano nell'intervallo  $[s, t]$

Vediamo alcune proprietà che un processo di conteggio *può* possedere

- Incrementi indipendenti: diremo che un processo è a incrementi indipendenti se dati  $t_1 < t_2 < t_3 < t_4$ , la quantità  $N(t_2) - N(t_1)$  è indipendente dalla quantità  $N(t_4) - N(t_3)$
- Incrementi stazionari: un processo di conteggio è a incremento stazionario se vale che la quantità  $N(t + s) - N(t)$  non dipende dal tempo  $t$

Vediamo un paio di esempi reali per capire quando possiamo supporre che valgano queste proprietà

**Esempio 1** : consideriamo le tre seguenti quantità

- #persone che entrano in un negozio
- #nati dall'inizio dell'anno
- #gol di un giocatore fino ad un certo tempo  $t$  della sua carriera

Quali delle precedenti quantità sono a incrementi indipendenti? La prima in certi casi lo è. Ad esempio l'ingresso di una persona al supermercato non è regolato da quante altre persone siano già entrate. Anche se invece andassimo a considerare un banco di vendita di strada una folla intorno al banco attirerebbe un maggior numero di persone. Dunque non è vero per tutti i negozi. Anche la seconda quantità potrebbe essere -almeno in parte- a incrementi indipendenti, anche se è noto che esistono mesi in cui si ha una maggiore natalità. La terza quantità invece non è a incrementi indipendenti visto che il numero di gol segnati da un giocatore può dipendere da quanti gol ha già fatto in precedenza, fatto che influenza il morale del giocatore.

Quali delle tre precedenti quantità è a incrementi stazionari? La prima no perché l'ingresso al supermercato è influenzato dall'orario. La seconda sì perché i nati dall'inizio dell'anno non influenzano le nascite nel resto dell'anno. La terza invece no perché visto che andiamo a considerare l'intera carriera un giocatore ha maggiori probabilità di segnare nel fiore degli anni invece che in età più avanzata. Se invece ci fossimo limitati all'interno di un intervallo di tempo più breve come "il primo tempo di una partita" allora avremmo ottenuto un processo a incremento stazionario

Come abbiamo visto nella realtà non possiamo contare spesso sulle due proprietà da noi descritte. Dunque in generale la stazionarietà non esiste ma abbiamo



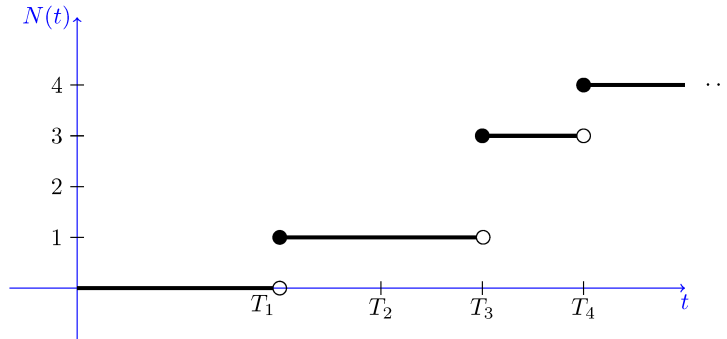


Figura 26: Grafico di un generico processo di conteggio

anche visto che andando a ridurre gli intervalli di tempo in certi casi è un'ipotesi accettabile.

Per concludere mostriamo in Figura 26 una rappresentazione di un processo di conteggio

## 12.2 Processi di Poisson

Un *processo di Poisson* è un particolare tipo di processo di conteggio per cui valgono le seguenti proprietà

1.  $N(0) = 0$
2. Indipendenza degli incrementi
3. Il numero di eventi in un certo intervallo  $t$  segue una distribuzione di Poisson di media  $\lambda t$ . Quindi  $\forall t, s > 0$   $P(N(t+s) - N(s) = k) = (\lambda t)^k \frac{e^{-\lambda t}}{k!}$

Notiamo che dalla terza proprietà abbiamo che

$$E[N(t)] = \lambda t$$

La particolarità del processo di Poisson inoltre è che l'incremento è sempre unitario.

Il problema con la definizione appena data è il seguente: quando, nella realtà, possiamo assimilare un processo ad un processo di Poisson? Non è facile dirlo e per questo vengono usati test statistici come il *test del  $\chi^2$* .

Per questo motivo sovente viene utilizzata una forma alternativa del processo di Poisson che è un po' più "operativa". Secondo questa definizione, per un processo di Poisson valgono le seguenti proprietà

1.  $N(0) = 0$

2. Indipendenza e stazionarietà degli incrementi
3.  $P(N(h) = 1) = \lambda h + o(h)$   
dove  $o(h)$  è la notazione *o piccolo* per gli infinitesimi. La proprietà appena descritta indica quindi che in un certo intervallo (solitamente molto piccolo)  $h$ , si verifica un solo incremento.
4.  $P(N(h) \geq 2) = o(h)$   
ovvero è infinitesimamente probabile che si verifichi più di un conteggio in un solo intervallo  $h$ . Dunque non si verificano due conteggi

Per poter utilizzare questa definizione dobbiamo verificare che le due definizioni di processo di Poisson siano equivalenti fra loro. Cominciamo col verificare che la prima definizione implichi la seconda.

**Prop. 1)** è identica in entrambe le definizioni

**Prop. 2)** Dalla prima definizione sappiamo che

$$P(N(t+s) - N(s) = k) = (\lambda t)^k \frac{e^{-\lambda t}}{k!}$$

e questo implica la stazionarietà

**Prop. 3)** Anche in questo caso dalla prima definizione sappiamo che

$$P(N(t+s) - N(s) = k) = (\lambda t)^k \frac{e^{-\lambda t}}{k!}$$

Supponiamo di porre i valori  $k = 1, s = 0, t = h$ . Da cui otteniamo che

$$P(N(t+s) - N(s) = k) = P(N(h) - \underbrace{N(0)}_{=0} = 1) = \lambda h e^{-\lambda h}$$

Ma dallo sviluppo in serie di  $e^{-\lambda h}$  otteniamo che

$$\lambda h e^{-\lambda h} = \lambda h \left( 1 - \lambda h + \frac{(\lambda h)^2}{2} - \dots \right) = \lambda h + o(h)$$

**Prop. 4)** Anche in questo caso dalla prima definizione sappiamo che

$$P(N(t+s) - N(s) = k) = (\lambda t)^k \frac{e^{-\lambda t}}{k!}$$

Supponiamo di porre  $s = 0, t = h, k = 0$ . Da cui otteniamo che:

$$P(N(h) = 0) = e^{-\lambda h} = 1 - \lambda h + \frac{(\lambda h)^2}{2} - \dots = 1 - \lambda h + o(h)$$

Quale è il significato intuitivo della seconda definizione del processo di Poisson? Che dovremmo sempre andare a considerare un intervallo  $h$  sufficientemente

stretto affinché un solo conteggio ricada nell'intervallo.

Fino ad ora abbiamo mostrato che la prima definizione implica la seconda. Dobbiamo ancora dimostrare che la seconda definizione implica la prima. Consideriamo la funzione

$$g(x) = E[e^{-uN(x)}]$$

Andando a considerare la funzione nel punto  $t + h$  che

$$\begin{aligned} g(t+h) &= E[e^{-uN(t+h)}] = E[e^{-u(N(t+h)-N(t)+N(t))}] = \\ &= E[e^{-u(N(t+h)-N(t))} \cdot e^{-uN(t)}] \end{aligned}$$

E visto che vale l'indipendenza degli incrementi abbiamo che:

$$E[e^{-u(N(t+h)-N(t))} \cdot e^{-uN(t)}] = E[e^{-u(N(t+h)-N(t))}] \cdot \underbrace{E[e^{-uN(t)}]}_{=g(t)}$$

Ma visto che  $N(t+h) - N(t) = N(h)$ , allora  $g(t+h) = g(t)E[e^{-uN(h)}]$ . Per completare andiamo a calcolare  $E[e^{-uN(h)}]$  partendo dalla definizione di valore atteso:

$$\begin{aligned} E[e^{-uN(h)}] &= \underbrace{1}_{N(h)=0} \cdot \underbrace{(1-\lambda h + o(h))}_{P(h=0)} + \underbrace{e^{-u}}_{N(h)=1} \underbrace{(\lambda h + o(h))}_{P(h=1)} + \underbrace{o(h)}_{N(h)>1 \cdot P(h>1)} = \\ &= 1 - \lambda h + \lambda h e^{-u} + o(h) \end{aligned}$$

Dunque

$$g(t+h) = g(t)(1 - \lambda h + \lambda h e^{-u} + o(h))$$

Se di questa quantità andiamo a calcolarne la derivata secondo la definizione, otteniamo che:

$$\lim_{h \rightarrow 0} \frac{g(t+h) - g(t)}{h} = \lim_{h \rightarrow 0} g(t) \frac{\lambda h(e^{-u} - 1) + o(h)}{h} = \lim_{h \rightarrow 0} g(t) (\lambda(e^{-u} - 1) + \frac{o(h)}{h})$$

Ovvero abbiamo che:

$$g'(t) = g(t)\lambda(e^{-u} - 1) \Leftrightarrow \frac{g'(t)}{g(t)} = \lambda(e^{-u} - 1)$$

Dunque, integrando in  $t$  da ambo le parti otteniamo che:

$$\ln(g(t)) = \lambda t(e^{-u} - 1) + c \Leftrightarrow g(t) = A e^{\lambda t(e^{-u} - 1)}$$

Dove sia  $c$  che  $A$  sono costanti derivanti dall'integrazione. Inoltre riguardo ad  $A$  ricordiamo che vale che  $A = g(0) = 1$  visto che  $g(t) = E[e^{-uN(t)}]$  e per

$t = 0$  abbiamo che  $g(0) = E[1] = 1$

### Esercizi di fine lezione

**Esercizio 1** : supponiamo di lanciare un dado e consideriamo la seguente variabile aleatoria

$X_n$  = valore più alto ottenuto al tempo  $n$

Come è fatta la matrice di transizione?

Abbiamo sei stati possibili, uno per ogni possibile valore del dado. Ovviamente quando passiamo da uno stato di "valore" inferiore a uno di valore superiore è impossibile tornare allo stato precedente. Infatti se dopo aver ottenuto  $k$  con un tiro di dado, otteniamo  $k + 1$  al tiro successivo, da ora in poi il tiro di dado più alto sarà  $k + 1$ .

Dunque la matrice di transizione è fatta come segue:

$$P = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 4/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 6/6 \end{bmatrix}$$

05 / 12 / 2017

### 12.3 Alcune proprietà su Poisson

Supponiamo di considerare due intervalli di tempo successivi  $T_1$  e  $T_2$ . Rispetto a un generico istante di tempo  $t$  vale che:

$$\begin{aligned} P(T_1 > t) &= P(N(t) = 0) = e^{-\lambda t} \\ P(T_1 < t) &= 1 - e^{-\lambda t} \Rightarrow T_1 \sim \exp(\lambda) \end{aligned}$$

Inoltre vale che:

$$\begin{aligned} P(T_2 > t) &= E[P(T_2 > t | T_1)] \\ P(T_2 > t | T_1 = s) &= P(N(t+s) - N(s) = 0 | T_1 = s) = P(N(t+s) - N(s) = 0) = e^{-\lambda t} \end{aligned}$$

Supponiamo invece di considerare due diversi tipi di eventi  $N_1(t)$  e  $N_2(t)$ . Andiamo a considerarne la somma:

$$N(t) = N_1(t) + N_2(t)$$

Abbiamo un teorema a riguardo

**Teorema 12.1.** *Se  $N(t)$  è un processo di Poisson avente parametro  $\lambda$  e se ciascun evento è indipendente dagli altri ed è del primo tipo con probabilità  $P$ , allora  $N_1(t)$  è un processo di Poisson di parametro  $\lambda p$  e  $N_2(t)$  è un processo di Poisson di parametro  $\lambda(1 - p)$ , indipendente da  $N_1(t)$ .*

**Esempio 1** : supponiamo di arrivare in ritardo di  $t$  minuti ad una partita allo stadio e ci viene detto che è già stato segnato un gol. Quale è la probabilità che il gol si sia verificato nei primi  $s$  minuti?

$$\begin{aligned} P(T_1 < s | N(t) = 1) &= \frac{P(T_1 < s, N(t) = 1)}{P(N(t) = 1)} = \\ &= \frac{P(N(s) = 1, N(t) - N(s) = 0)}{P(N(t) = 1)} = \frac{\lambda s e^{-\lambda s} e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t} \end{aligned}$$

con  $s \in [0, t]$ .

Dunque l'evento può essere avvenuto nei primi  $s$  minuti con probabilità proporzionale alla frazione di tempo trascorsa. Dunque non abbiamo alcuna informazione di rilievo.

Per concludere andiamo a considerare un particolare tipo di processo di Poisson: il *processo di Poisson non omogeneo*. In questi processi l'aumento dei conteggi al tempo  $t$  è funzione di  $t$  stesso. Valgono infatti le seguenti proprietà:

1.  $N(0) = 0$
2. Indipendenza e stazionarietà degli incrementi
3.  $P(N(t+h) - N(t) \geq 2) = o(h)$
4.  $P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$

Di fatto sarà vero che:

$$P(N(t+s) - N(t) = k) = \frac{(m(t+s) - m(t))^k}{k!} e^{-[m(t+s) - m(t)]}$$

## 13 Catene di Markov continue

### 13.1 Definizioni

In questa sezione andremo ad estendere i processi di Markov dal discreto al continuo. Anche in questo caso faremo l'assunzione che gli eventi futuri siano indipendenti dal passato se conosciamo il presente.

Una catena di Markov a tempo continuo è un processo stocastico  $\{X(t)\}_t$  che assume un valore sugli interi non negativi e tale per cui

$$\forall s, t > 0, u < s, \forall i, j, x(u)$$

$$P(X(s+t) = j | X(s) = i, X(u) = x(u), 0 \leq u < s) = P(X(t+s) = j | X(s) = i) \quad (20)$$

Quindi stiamo dicendo che la distribuzione dei futuri conteggi  $X(t+s)$  dato il conteggio al tempo presente  $X(s)$  e al tempo passato  $X(u)$ ,  $0 \leq u < s$  dipende in realtà soltanto dal tempo presente ed è indipendente dal passato. Inoltre se il termine

$$P(X(t+s) = j | X(s) = i)$$

è indipendente da  $s$ , allora la catena possiede la proprietà di probabilità di transizione omogenea.

Se un processo di Markov entra in uno stato  $i$  e vi resta per un certo periodo di tempo  $s$ , la probabilità che vi resti per un ulteriore tempo  $t$  è pari a:

$$P(T_i > s+t | T_i > s) = P(T_i > t)$$

Dunque stiamo dicendo che una catena di Markov a tempo continuo è priva di memoria e dunque segue necessariamente una distribuzione esponenziale. Dunque possiamo dire che una catena di Markov è un processo stocastico che, una volta entrato in uno stato  $i$

- Vi rimane per un tempo che segue una distribuzione esponenziale di media  $\frac{1}{v_i}$
- Quando lascia lo stato  $i$  entra in uno stato  $j$  con probabilità  $P_{ij}$  tale per cui:

$$\begin{aligned} P_{ii} &= 0, \\ \sum_j P_{ij} &= 1 \end{aligned}$$

- La quantità di tempo che il processo spende nello stato corrente e nello stato successivo, sono variabili aleatorie indipendenti

Consideriamo un paio di esempi

**Esempio 1 :** consideriamo un autolavaggio formato da due fasi. Prima la macchina passa nella fase uno di pulizia esterna, successivamente nella fase due di pulizia interni. Il tempo che impiega in ognuna delle fasi è dato da una variabile aleatoria esponenziale di media  $\mu_1$  per la prima fase e  $\mu_2$  per la seconda.

Supponiamo inoltre che i clienti arrivino secondo un processo di Poisson di frequenza  $\lambda$ . Supponiamo inoltre che un cliente entri nel sistema solo se entrambe le fasi sono libere.

Quando un cliente arriva esistono tre stati possibili:

1. Nessun cliente è presente nel negozio
2. Una macchina è nella fase uno
3. Una macchina è nella fase due

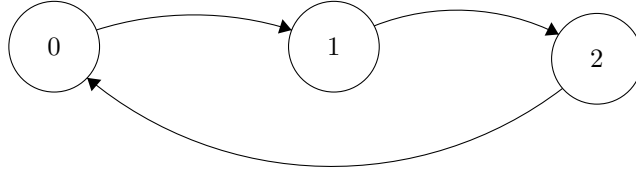


Figura 27: Diagramma di transizione dell'esempio 1

Questi tre stati sono mostrati in Figura 27

Poiché il valore atteso della prima fase è pari  $\mu_1$ , allora il tempo medio di permanenza nello stato  $v_1$  è pari proprio a  $\mu_1$  infatti sono entrambi distribuiti secondo una variabile esponenziale. Similmente anche per le altre variabili. Dunque:

$$v_0 = \lambda \quad v_1 = \mu_1 \quad v_2 = \mu_2$$

**Esempio 2** : supponiamo di avere a disposizione un sistema in cui da  $n$  persone nascono nuovi individui secondo un processo esponenziale di parametro  $\lambda_n$  e muoiono degli individui secondo un processo esponenziale di parametro  $\mu_n$ . Allora vale che:

$$v_0 = \lambda_0 \quad v_i = \min\{\lambda_i, \mu_i\} = \lambda_i + \mu_i$$

e per tale sistema le probabilità di transire negli altri stati sono pari a:

$$P_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}$$

$$P_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}$$

**Esempio 3** : andiamo a considerare una variante del problema appena visto. Consideriamo un modello lineare di crescita con immigrazione. Supponiamo infatti che in questo sistema valga che:

$$\mu_n = n\mu, n \geq 1$$

$$\lambda_n = n\lambda + \theta, n \geq 0$$

dove la popolazione cresce secondo una distribuzione esponenziale di parametro  $\lambda$  ma viene incrementata anche da un fattore esterno  $\theta$  detto fattore di immigrazione. Supponiamo che gli individui muoiano come in precedenza secondo una distribuzione esponenziale di parametro  $\mu_n = n\mu$ . Andiamo a considerare la quantità:

$$M(t) = E[X(t)]$$

E vediamo come cambia nel tempo. Per farlo dobbiamo considerare la quantità:

$$M(t+h) = E[X(t+h)] = E[E[X(t+h)|X(t)]]$$

Ma d'altronde sappiamo che:

$$X(t+h) = \begin{cases} X(t) + 1, & \text{con probabilità } (\theta + X(t)\lambda)h + o(h) \\ X(t) - 1, & \text{con probabilità } X(t)\mu h + o(h) \\ X(t), & \text{con probabilità } 1 - [\theta + X(t)\lambda + X(t)\mu]h + o(h) \end{cases}$$

Andando a considerare il termine della prima delle due medie:

$$E[X(t+h)|X(t)] = X(t) + (\theta + X(t)\lambda - X(t)\mu)h + o(h)$$

Dunque:

$$M(t+h) = M(t) + (\lambda - \mu)M(t)h + \theta h + o(h)$$

Oppure in modo equivalente:

$$\frac{M(t+h) - M(t)}{h} = (\lambda - \mu)M(t) + \theta + \frac{o(h)}{h}$$

Da cui, considerando il limite  $h \rightarrow 0$  otteniamo la seguente equazione differenziale:

$$\begin{aligned} M'(t) &= (\lambda - \mu)M(t) + \theta \\ M(0) &= i \end{aligned}$$

La cui soluzione è:

$$M(t) = \frac{\theta}{\lambda - \mu}(e^{(\lambda - \mu)t} - 1) + ie^{(\lambda - \mu)t}$$

13 / 12 / 2017

Riprendiamo in generale i processi di nascita e morte andando a studiare il tempo  $T_i$  con cui passiamo da uno stato  $i$  ad uno stato  $i+1$ . Andando a calcolare  $E[T_i]$ . Andiamo a distinguere due casi: se vale che  $i = 0$  allora:

$$E[T_0] = \frac{1}{\lambda_0}$$

Se invece  $i > 0$  allora:

$$I_i = \begin{cases} 0, & \text{se I manda in } i-1 \\ 1, & \text{se I manda in } i+1 \end{cases}$$



In questo caso allora

$$E[T_i] = E[E[T_i|I_i]]$$

Ma visto che sappiamo che

$$E[T_i|I_i = 1] = \frac{1}{\lambda_i + \mu_i}$$

$$E[T_i|I_i = 0] = \frac{1}{\lambda_i + \mu_i} + E[T_{i-1}] + E[T_i]$$

dunque

$$E[T_i] = \frac{1}{\lambda_i + \mu_i} \cdot \frac{\lambda_i}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i} \left( \frac{1}{\lambda_i + \mu_i} + E[T_{i-1}] + E[T_i] \right) = \dots = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E[T_{i-1}]$$

Quindi ad esempio:

$$E[T_1] = \frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_1 \lambda_0}$$

## 13.2 Funzione probabilità di transizione

Consideriamo la probabilità

$$P_{ij}(t) = P(X(t+s) = j | X(s) = i)$$

Definiamo  $X_k$  come il tempo che il processo trascorre in  $k$  prima di transire in  $k+1$ . Se andiamo a considerare solo le nascite, vale che

$$P(X(t) < j | X(0) = i) = P\left(\underbrace{\sum_{k=i}^{j-1} X_k}_{\text{ipoesponenziale}} > t\right)$$

Intuitivamente stiamo dicendo che se non ci sono ancora  $j$  individui, allora vuol dire che il nostro processo ha trascorso i precedenti quanti di tempo in uno stato  $k < j$  fino ad almeno il tempo  $t$ .

Dunque sarebbe interessante determinare la quantità  $P(X(t) < j | X(0) = i)$

Cerchiamo di costruire un'equazione differenziale che studia la variazione di tale quantità. Lo possiamo fare andando ad introdurre le seguenti quantità

- $P_{ij}(t) = P(X(t+s) = j | X(s) = i)$
- $v_i$  tasso con cui il processo effettua una transizione
- $q_{ij} = v_i P_{ij}(t)$  detta frequenza di transizione istantanea che mi dice con quale frequenza passi dallo stato  $i$  allo stato  $j$

Possiamo osservare che

- $\sum_j q_{ij} = \sum_j v_i P_{ij} = v_i \sum_j P_{ij} = v_i$   
Ovvero la somma su ogni "riga"  $i$  ci dà  $v_i$

- $P_{ij} = \frac{q_{ij}}{v_i}$

Avremo infine bisogno del seguente lemma che non dimostriamo. Vale che

$$\begin{cases} \lim_{h \rightarrow 0} \frac{1 - P_{ii}(h)}{h} = v_i \\ \lim_{h \rightarrow 0} \frac{P_{ij}(h)}{h} = q_{ij} \end{cases} \quad (21)$$

Da cui possiamo andare a riscrivere l'equazione di Chapman Kolmogorov nel continuo:

$$P_{ij}(t+s) = \sum_k P_{ik}(s)P_{kj}(t) \quad (22)$$

Allora in base a questo possiamo dire che

**Teorema 13.1** (Equazione backward di Kolmogorov). *Viene detta equazione backward di Kolmogorov la seguente equazione*

$$P'_{ij}(t) = \frac{dP_{ij}(t)}{dt} = \sum_{k \neq i} q_{ik}P_{kj}(t) - v_i P_{ij}(t)$$

**Esempio 1** : consideriamo un processo di pura nascita

$$q_{i,i+1} = \lambda_i$$

Andando ad applicare il teorema otteniamo che

$$P'_{ij}(t) = \lambda_i P_{i+1,j}(t) - \lambda_i P_{i,j}(t)$$

Notiamo che in questo caso la sommatoria è di un solo membro perché possiamo transire solo dallo stato  $i$  allo stato  $i+1$  a causa del processo che è di pura nascita

**Esempio 2** : consideriamo invece il caso generale di un processo di nascita e morte. Ricordando che  $q_{ik} = v_i P_{ik}$  poiché minimo tra due variabili esponenziali, e dunque pari alla somma:

$$\begin{cases} P'_{0j}(t) = \lambda_0 P_{0j}(t) - \lambda_0 P_{0j}(t), & \text{se } i = 0 \\ P'_{ij}(t) = (\lambda_i + \mu_i) \left( \frac{\lambda_i}{\lambda_i + \mu_i} P_{i+1,j}(t) + \frac{\mu_i}{\lambda_i + \mu_i} P_{i-1,j}(t) \right) - (\lambda_i + \mu_i) P_{ij}(t), & \text{se } i \neq 0 \end{cases}$$

Sicuramente il caso  $i \neq 0$  può essere riscritto semplificando il termine  $(\lambda_i + \mu_i)$ :

$$P'_{ij}(t) = (\lambda_i P_{i+1,j}(t) + \mu_i P_{i-1,j}(t)) - (\lambda_i + \mu_i) P_{ij}(t)$$

**Esempio 3** : applichiamo la formula del caso generale. Supponiamo che una certa macchina presenti un tempo di rottura che segue una variabile esponenziale. Dunque chiamiamo

$T$ : il tempo di rottura tale per cui  $T \sim \exp(\lambda)$   
 $\Theta$ : tempo di riparazione con  $\Theta \sim \exp(\mu)$

Inoltre daremo come risultato 0 se la macchina funziona e 1 se la macchina non funziona. Sapendo che la macchina è perfettamente funzionante al tempo 0, qual'è la probabilità che sia funzionante al tempo  $t = 10$ ?

Sapendo che allo stato iniziale la macchina è funzionante, vale che:  
 $\lambda_0 = \lambda$  e che  $\lambda_i = 0$   
 $\mu_1 = \mu$  e che  $\mu_i = 0$

Allora vale che:

$$\begin{aligned} P'_{00}(t) &= \lambda(P_{10}(t) - P_{00}(t)) \\ P'_{10}(t) &= \mu(P_{00}(t) - P_{10}(t)) \end{aligned}$$

Moltiplicando queste due equazione per  $\lambda$  e sommandole, otteniamo che

$$\mu P'_{00}(t) + \lambda P'_{10}(t) = 0$$

Che sottoposto a integrazione:

$$\mu P_{00}(t) + \lambda P_{10}(t) = c$$

Ma sapendo che la macchina è inizialmente funzionante, allora vale che  $P_{00}(0) = 1$  e  $P_{10}(0) = 0$  dunque  $c = \mu$  da cui

$$\lambda P_{10}(t) = \mu(1 - P_{00}(t))$$

Che sostituito nelle due equazioni iniziali relative a  $P'_{00}(t)$  e  $P'_{10}(t)$  ci fornisce

$$P_{00}(t) = \frac{\mu}{\lambda + \mu} - \frac{\mu}{\mu + \lambda} e^{-(\lambda + \mu)t} \quad P_{00}(10) = \frac{\lambda}{\lambda + \mu} e^{-10(\lambda + \mu)} + \frac{\mu}{\lambda + \mu}$$

Oltre all'equazione Kolmogorov backward esiste anche un'equazione di Kolmogorov forward

**Teorema 13.2** (Equazione forward di Kolmogorov). *Viene detta equazione forward di Kolmogorov la seguente equazione*

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t)$$

### 13.3 Stabilizzazione

Come già visto nel caso discreto, ci potremmo chiedere se dopo un certo periodo di tempo abbiamo una qualche forma di stabilizzazione. Dunque ci chiediamo se esista una distribuzione limite.

Andiamo a considerare:

$$\lim_{t \rightarrow \infty} P_{ij}(t) = P_j$$

dunque non dipende né dal tempo  $t$  né dallo stato iniziale. L'idea è che la distribuzione limite esiste se abbiamo una situazione di equilibrio tra flusso in ingresso e flusso in uscita. Per derivare delle equazioni per le  $P_j$  possiamo partire dalla forma forward di Kolmogorov vista poc'anzi, e andarne a calcolare il limite:

$$\lim_{t \rightarrow \infty} P'_{ij}(t) = \lim_{t \rightarrow \infty} \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) = \sum_{k \neq j} q_{kj} P_k - v_j P_j$$

Ma poiché la  $P_{ij}(t)$  è vincolata tra zero e uno, ne segue che se  $P'_{ij}(t)$  converge, lo fa convergendo al valore zero. Dunque  $\sum_{k \neq j} q_{kj} P_k - v_j P_j = 0$  da cui

$$v_j P_j = \sum_{k \neq j} q_{kj} P_k \quad (23)$$

Inoltre sappiamo che deve valere  $\sum_j P_j = 1$ .

Notiamo che esiste una precisa analogia con quanto visto nel discreto. Infatti valeva che:

$$\begin{cases} \pi_i = \sum_j \pi_j P_{ji} \\ \sum_i \pi_i = 1 \end{cases}$$

Visto che  $\pi_i$  rappresenta la proporzione di transizione che conducono il processo nello stato  $i$ , e visto che  $1/v_i$  è il tempo medio speso in uno stato  $i$ , sembrerebbe intuitivo dedurre che  $P_i$ , la proporzione di tempo spesa nello stato  $i$ , dovrebbe essere una media pesata dei vari  $\pi_i$  proporzionalmente a  $1/v_i$ . Dunque

$$P_i = \frac{\frac{\pi_i}{v_i}}{\sum_j \frac{\pi_j}{v_j}}$$

Andiamo a vedere se la quantità  $P_i$  che abbiamo definito ci porta a qualche contraddizione. Andiamo a sostituirla nell'equazione di bilanciamento vista poc'anzi:

$$v_j \frac{\frac{\pi_j}{v_j}}{\sum_l \frac{\pi_l}{v_l}} = \sum_{k \neq j} q_{kj} \frac{\frac{\pi_k}{v_k}}{\sum_l \frac{\pi_l}{v_l}} \iff \pi_j = \sum_{k \neq j} \frac{q_{kj}}{v_k} \pi_k = \sum_{k \neq j} \frac{v_k P_{kj}}{v_k} \pi_k = \sum_{k \neq j} P_{kj} \pi_k$$

**Esempio 1 :** quale è la distribuzione limite per un processo di nascita morte?

Consideriamo uno stato alla volta:  
 lo stato zero viene lasciato quando c'è una nascita. Dunque viene lasciato  
 con un tasso di uscita  $\lambda_0 P_0$ . Vi rientriamo dallo stato 1 con un tasso  $\mu_1 P_1$ .  
 Abbiamo una situazione di equilibrio se

$$\lambda_0 P_0 = \mu_1 P_1$$

Nello stato 1 invece la situazione di equilibrio si verifica quando

$$(\lambda_1 + \mu_1) P_1 = \mu_2 P_2 + \lambda_0 P_0$$

In generale, per lo stato  $n$  vale che

$$(\lambda_n + \mu_n) P_n = \mu_{n+1} P_{n+1} + \lambda_{n-1} P_{n-1}$$

Andando a sommare tutte le situazioni di equilibrio otteniamo che

$$P_n = \frac{\lambda_{n-1} \dots \lambda_0}{\mu_1 \dots \mu_n} P_0$$

Ma poiché  $\sum P_i = 1$  allora vale che  $P_0 = \frac{1}{1 + \sum_{n=1}^{+\infty} \frac{\lambda_{n-1} \dots \lambda_0}{\mu_1 \dots \mu_n}}$  e dunque

$$P_n = \frac{\lambda_{n-1} \dots \lambda_0}{\mu_1 \dots \mu_n (1 + \sum_{n=1}^{+\infty} \frac{\lambda_{n-1} \dots \lambda_0}{\mu_1 \dots \mu_n})}$$

Affinché si verifichi la convergenza della serie, deve essere vero che:

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