

Politecnico di Bari

Corso di Laurea in Ingegneria delle Telecomunicazioni

Appunti del corso di

TEORIA DEI SEGNALI

Pietro Guccione

Anno Accademico 2007-2008

Indice

Capitolo 1. Richiami principali ai segnali	5
1.1. Introduzione	5
1.2. Tipi di segnale	6
1.3. Segnali elementari	9
1.4. La Correlazione	15
Capitolo 2. La teoria delle probabilità	21
2.1. Esperimenti Aleatori	21
2.2. Le Basi della Teoria delle Probabilità	22
2.3. Variabili Aleatorie	28
2.4. Densita' di Probabilita'	30
2.5. Operazioni sulla Variabile Aleatoria	32
2.6. Parametri Statistici di una Variabile Aleatoria	33
2.7. Esempi di Variabili Aleatorie	36
2.8. Variabili Aleatorie Condizionate	45
2.9. Applicazioni notevoli	46
2.10. Sistemi di Variabili Aleatorie	50
2.11. Convergenza ed approssimazione	62
Capitolo 3. I Processi Stocastici	67
3.1. Definizione di Processi Stocastici	67
3.2. Parametri Statistici del 1° e 2° Ordine	70
3.3. Processi Stazionari	79
3.4. Filtraggio di un Processo Aleatorio	88
3.5. Analisi Spettrale di un Processo Aleatorio	92
3.6. Processi Aleatori Gaussiani	100
3.7. Processi Ergodici	104
3.8. Cenni sulle Catene di Markov	110
Capitolo 4. La trasmissione dei segnali	117
4.1. Introduzione	117
4.2. Generalita' sui Sistemi di Trasmissione	117
4.3. Trasmissione Analogica e Numerica	122
4.4. Il Campionamento	123
4.5. La Quantizzazione	129
4.6. Il Canale Binario	134

4.7. Teoria dell'Informazione	140
Capitolo 5. Il rumore	153
5.1. Introduzione	153
5.2. Caratteristiche Generali del Rumore	154
5.3. Fattore e Temperatura Equivalente di Rumore	160
Capitolo 6. La modulazione analogica	167
6.1. Introduzione	167
6.2. Rappresentazione complessa dei segnali	169
6.3. Sistemi di trasmissione con modulazione	176

CAPITOLO 1

Richiami principali ai segnali

1.1. Introduzione

La definizione di segnale parte dall'esperienza comune. Esempi di segnale nella vita quotidiana sono il segnale acustico che viene prodotto da uno strumento musicale, il segnale radio captato dall'antenna di un ricevitore, la rappresentazione del battito cardiaco attraverso un elettrocardiografo e così via.

Tutti gli esempi che si possono fare hanno una matrice comune: il segnale è una *grandezza fisica variabile a cui è associata una qualche forma di informazione*. Lo studio dei segnali quindi passa necessariamente attraverso lo studio delle funzioni matematiche di una o più variabili. Le grandezze fisiche rappresentate da un segnale sono le più svariate: l'intensità luminosa e il colore su uno schermo nel caso di un segnale televisivo, la variazione della pressione dell'aria nel caso di un segnale musicale, la tensione elettrica o la corrente nel caso di un segnale misurato su di un circuito elettrico, un'onda elettromagnetica nel caso di un segnale radio captato dallo spazio.

L'evoluzione di molti segnali monodimensionali (cioè dipendenti da una sola grandezza) avviene nel tempo: esempi sono il segnale musicale, la misura della tensione su un condensatore, la variazione dell'intensità luminosa del sole durante il giorno, eccetera. Tuttavia è possibile considerare dipendenze diverse di un segnale: ad esempio la sua variazione nello spazio. La misura dell'intensità dell'oscillazione di un terremoto ad uno stesso istante nelle varie località rappresenta un segnale di cui interessa la cui estensione spaziale e non la sua evoluzione temporale. Naturalmente è sempre possibile immaginare lo stesso tipo di informazione (l'intensità di un terremoto) in una data località e seguirne la sua evoluzione nel tempo.

Quest'ultimo esempio porta alla rappresentazione di segnali bidimensionali o anche multidimensionali, segnali cioè che variano in dipendenza della variazione di due o più grandezze. Il segnale televisivo bianco e nero è un esempio di segnale tridimensionale, dato che esso è dipendente da due coordinate spaziali (larghezza ed altezza dello schermo) e da una coordinata temporale (il susseguirsi delle scene sullo schermo).

Se consideriamo invece un segnale televisivo a colori esso è in realtà la sovrapposizione di tre segnali tridimensionali, dato che separatamente in ogni punto dello schermo è rappresentata la sovrapposizione dei tre colori fondamentali: rosso, verde, blu. Quindi un segnale televisivo a colori si può pensare come un segnale vettoriale (costituito cioè da tre componenti) a tre dimensioni, dipendente cioè da tre grandezze fisiche: $c(x, y, t) = [red(x, y, t), green(x, y, t), blue(x, y, t)]$.

1.2. Tipi di segnale

Una prima classificazione di segnale è stata già fatta differenziando i segnali monodimensionali da quelli multidimensionali, come anche quelli scalari da quelli vettoriali, costituiti cioè da più componenti.

Si possono inoltre differenziare i segnali in base ai valori assunti dalla variabile indipendente:

- *segnali a tempo continuo*: sono quelli per i quali il dominio della funzione ha la cardinalità dei numeri reali. La variabile indipendente (ad esempio il tempo) assume valori in modo continuo (ad esempio un segnale musicale emesso da uno strumento).
- *segnali a tempo discreto*: sono quelli per i quali il dominio della funzione ha la cardinalità dei numeri naturali. Per questi segnali la variabile indipendente assume valori in un insieme discreto. In tal caso la dipendenza del segnale dalla variabile indipendente è rappresentata mediante la successione dei valori assunti: $x(n)$ per indicare il valore del segnale x dall' n -simo valore della variabile indipendente. Esempio di un segnale tempo discreto è il segnale televisivo, dato che esso è rappresentato sullo schermo mediante la successione di 25 fotogrammi al secondo.

I segnali stessi possono assumere valori in un insieme non numerabile di valori (*segnali ad ampiezza continua*) o in un insieme numerabile di valori (*segnali ad ampiezza discreta*). Esempio di un segnale ad ampiezza continua è la misura della tensione su un condensatore così come essa è rappresentata su un oscilloscopio analogico; esempio di un segnale ad ampiezza discreta è invece lo stato di un semaforo: ad ogni istante esso può assumere solo due possibili valori: acceso o spento. I segnali ad ampiezza continua sono detti anche segnali analogici, quelli ad ampiezza discreta sono detti numerici. In figura (1.2.1) sono rappresentati i due tipi di segnale sinora visti.

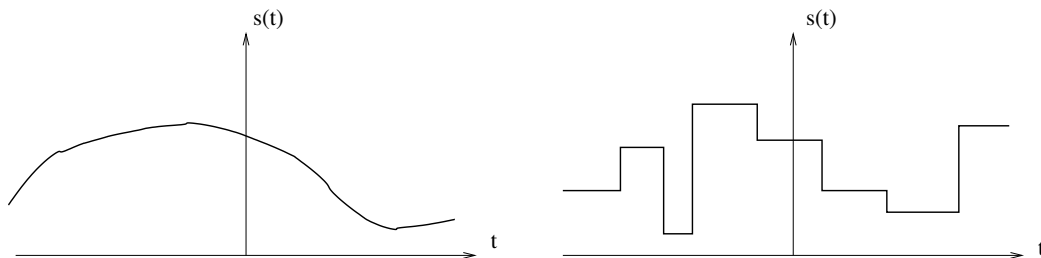


FIGURA 1.2.1. Differenza tra segnale ad ampiezza continua e segnale ad ampiezza discreta

Un'altra distinzione può essere fatta tra i segnali *periodici* e segnali *non periodici* (o aperiodici). Detto T un numero reale > 0 , un segnale $s(t)$ si dice periodico se $\forall n \in \mathbb{Z} : s(t) = s(t + nT)$. Un segnale periodico è quindi definito su tutto l'asse reale e per una sua descrizione completa è sufficiente la conoscenza all'interno di un periodo. Un segnale di durata finita è, quindi, aperiodico. Una combinazione lineare di segnali periodici di stesso periodo T o di periodo che è un sottomultiplo di T , cioè T/n è, a sua volta, periodica di periodo T .

I segnali inoltre possono essere suddivisi in base al loro comportamento energetico. Si dicono ad *energia finita* i segnali che verificano la seguente proprietà:

$$(1.2.1) \quad \int_{-\infty}^{+\infty} |s(t)|^2 dt < +\infty$$

dove la quantità a primo membro dell'espressione è detta *energia* del segnale. I segnali periodici non sono segnali ad energia finita, dato che, se $\int_{-T/2}^{+T/2} |s(t)|^2 dt$ è una quantità finita, l'integrale su tutto \mathfrak{R} risulterà sicuramente infinito. Tali segnali sono allora segnali a potenza finita, per i quali cioè risulta:

$$(1.2.2) \quad \lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-T/2}^{+T/2} |s(t)|^2 dt < +\infty$$

La quantità a primo membro è detta **potenza** del segnale. Per i segnali ad energia finita la potenza è nulla.

Per i segnali tempo discreti la definizione di energia e potenza è rispettivamente:

$$(1.2.3) \quad \sum_{n=-\infty}^{+\infty} |s(n)|^2$$

$$(1.2.4) \quad \lim_{N \rightarrow +\infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} |s(n)|^2$$

Infine altre distinzioni tra segnali possono essere fatte sulla base delle loro proprietà puramente matematiche: ad esempio si distinguono i segnali reali da quelli complessi, composti cioè di una parte reale e di una parte immaginaria: $s_c(t) = s_R(t) + js_I(t)$. Particolari simmetrie dei segnali possono permettere di distinguere i segnali pari, per i

quali risulta: $s(t) = s(-t)$, da quelli dispari, per i quali vale invece: $s(t) = -s(-t)$. Per un segnale che non gode di simmetria pari, nè dispari, si può sempre pensare di estrarne la sua parte pari:

$$(1.2.5) \quad s_e(t) = \frac{1}{2}[s(t) + s(-t)]$$

e la sua parte dispari

$$(1.2.6) \quad s_o(t) = \frac{1}{2}[s(t) - s(-t)]$$

1.2.1. Operazioni sui segnali. Vengono qui richiamate le principali operazioni che è possibile compiere sui segnali. Particolare interesse assumono le operazioni sulla variabile indipendente

1.2.1.1. *Traslazione.* La traslazione di un segnale è il suo spostamento sull'asse della variabile indipendente (o nel piano delle sue variabili indipendenti se dipende da due variabili): $s(t - t_o)$ è il segnale $s(t)$ spostato temporalmente nella posizione t_o . Se la variabile indipendente è il tempo, si dice anche che il segnale è ritardato di t_o secondi se $t_o > 0$ altrimenti è anticipato di t_o secondi, se risulta $t_o < 0$.

1.2.1.2. *Ribaltamento.* Il ribaltamento di un segnale corrisponde all'operazione: $s(t) \rightarrow s(-t)$, esso cioè viene descritto con l'asse della variabile indipendente riflesso rispetto all'asse delle ordinate. Questa operazione è utile per esaminare le proprietà di simmetria di un segnale (segnale pari o dispari).

1.2.1.3. *Scalatura dell'asse.* Considerato un numero reale $a > 0$, un segnale si dice che ha subito un cambiamento di scala se risulta la seguente trasformazione: $s(t) \rightarrow s(at)$. In particolare se $a > 1$ il segnale ha subito un *restringimento*, altrimenti, con $0 < a < 1$ il segnale subisce un'*espansione*. E' sempre possibile estendere il cambiamento di scala dell'asse della variabile indipendente ai casi in cui risulta $a < 0$, basta applicare separatamente le due operazioni di ribaltamento e di scalatura del segnale: $s(t) \rightarrow s(-t) \rightarrow s(-|a|t)$. Si ricordi che l'operazione di cambiamento di scala, come quella di ribaltamento – che si può considerare come un caso particolare con $a = -1$ – non commuta con quella di traslazione.

1.2.1.4. *Convoluzione tra segnali.* Dati due segnali $x(t)$ ed $h(t)$, si definisce il prodotto di convoluzione tra i due segnali come:

$$(1.2.7) \quad y(t) = x(t) \star h(t) = \int_{-\infty}^{+\infty} x(\tau)h(t - \tau)d\tau$$

La convoluzione gode delle seguenti proprietà:

- (1) La convoluzione è un'operazione commutativa: $x(t) \star h(t) = h(t) \star x(t)$
- (2) La convoluzione gode della proprietà associativa: $x(t) \star y(t) \star h(t) = (x(t) \star y(t)) \star h(t) = x(t) \star (y(t) \star h(t))$
- (3) La convoluzione è distributiva rispetto alla somma: $(x(t) + y(t)) \star h(t) = x(t) \star h(t) + y(t) \star h(t)$

1.3. Segnali elementari

Esiste una classe di segnali che, per la loro particolare semplicità, viene spesso utilizzata per schematizzare il comportamento dei segnali che si incontrano nei casi reali. A questi segnali si dà il nome di segnali elementari. Le proprietà viste precedentemente si applicano ovviamente anche ai segnali elementari.

1.3.1. Gradino unitario. Il gradino unitario è la funzione così definita:

$$(1.3.1) \quad u(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \end{cases}$$

Per $t = 0$ si assume che $s(0) = 0.5$.

1.3.2. Rampa. E' un segnale nullo per $t < 0$ e che, per $t > 0$, cresce proporzionalmente a t :

$$(1.3.2) \quad r(t) = \begin{cases} t, & t > 0 \\ 0, & t < 0 \end{cases}$$

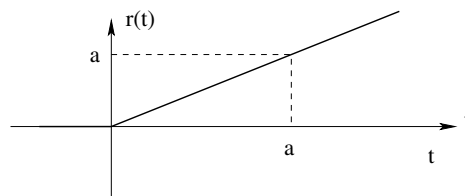


FIGURA 1.3.1. Rampa unitaria

Tale segnale può considerarsi come il risultato del passaggio dello scalino unitario attraverso un integratore:

$$(1.3.3) \quad r(t) = \int_{-\infty}^t u(\tau) d\tau$$

1.3.3. Parabola. La parabola (o rampa parabolica) è il segnale che si ottiene ri-applicando l'operatore di integrazione alla rampa:

$$(1.3.4) \quad p(t) = \int_{-\infty}^t r(\tau) d\tau = \frac{1}{2} t^2$$

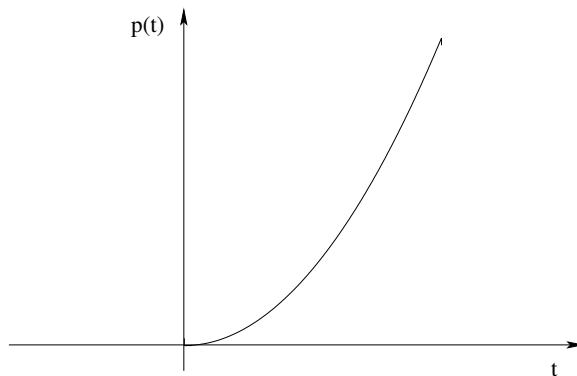


FIGURA 1.3.2. Rampa parabolica

1.3.4. Segnale rettangolare, onda quadra. Si chiama rettangolare un segnale che mantenga valore costante per tutta la sua durata limitata:

$$(1.3.5) \quad \text{rect} \left(\frac{t}{\tau} \right) = \begin{cases} 1, & |t| < \frac{\tau}{2} \\ 0, & |t| > \frac{\tau}{2} \end{cases}$$

È, chiaramente, un segnale di energia finita e la sua energia vale τ . La somma di segnali rettangolari ripetuti a distanza T dà luogo ad un segnale periodico, di periodo T :

$$(1.3.6) \quad sq(t) = \sum_{n=-\infty}^{+\infty} \text{rect}\left(\frac{t - nT}{\tau}\right)$$

che viene detto onda quadra.

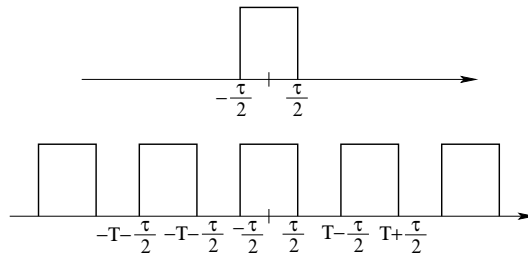


FIGURA 1.3.3. Rettangolo ed onda quadra

Se $\tau = T/2$ l'onda quadra si dice a duty cycle 50%. L'onda quadra (1.3.6) oscilla tra 0 e 1 ed ha valor medio τ/T . Un'onda quadra con duty cycle 50% che oscilla tra +1 e -1 ha valor medio nullo.

Si osservi infine che, a rigore, il segnale rettangolare (1.3.5) è discontinuo in $\pm\tau/2$ ed il suo valore in tali punti sarebbe indefinito. In un punto di discontinuità assumeremo che il segnale assuma il valore $s(t_o) = \frac{1}{2}[s(t_o^-) + s(t_o^+)]$

1.3.5. Delta di Dirac. Il Delta di Dirac non è in realtà una vera e propria funzione, ma una *distribuzione*. Essa, a rigore, dovrebbe essere definita solo all'interno di un integrale. La sua definizione parte dalla osservazione che la funzione:

$$(1.3.7) \quad \frac{1}{T} \text{rect}\left(\frac{t}{T}\right)$$

ha sempre area pari ad 1, qualunque sia il valore di T . Al tendere però di T a zero, il rettangolo diventa infinitamente stretto ed alto. Una definizione della funzione delta è allora la seguente:

$$(1.3.8) \quad \delta(t) = \lim_{T \rightarrow 0} \frac{1}{T} \text{rect}\left(\frac{t}{T}\right)$$

La funzione così definita ha valori sempre nulli tranne in $t = 0$ dove assume valore nominalmente infinito. La sua rappresentazione su di un grafico è quindi a rigore impossibile. La schematizzazione che si usa è quella riportata in fig. 1.3.4

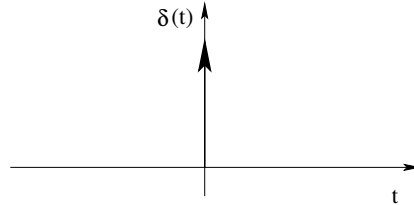


FIGURA 1.3.4. Rappresentazione grafica dell'impulso o delta di Dirac.

In base a quanto detto:

$$(1.3.9) \quad \int_{-\infty}^{+\infty} \delta(t) dt = 1$$

inoltre la funzione delta è pari: $\delta(-t) = \delta(t)$. La principale proprietà della funzione delta è la seguente:

$$(1.3.10) \quad \int_{-\infty}^{+\infty} s(t) \delta(t - t_o) dt = s(t_o)$$

essa cioè applicata ad una funzione all'interno di un integrale permette di *estrarre* il valore di quella funzione nel punto in cui il delta è applicato (sempre che la funzione $s(t)$ sia continua in $t = t_o$). Questa notazione è utilizzata per indicare l'estrazione di un *campione* da un segnale nella posizione in cui è posto l'impulso. La proprietà in (1.3.10) può essere vista anche nel modo seguente: l'impulso piazzato ad un dato istante τ e moltiplicato per una funzione $s(t)$ risulta pari all'impulso stesso ma con area uguale al valore che il segnale assume in quella posizione τ : $s(t) \delta(t - \tau) = s(\tau) \delta(t - \tau)$.

Un segnale può essere *rappresentato* mediante una successione infinita di impulsi delta infinitamente vicini tra loro e di valore pari al valore che il segnale assume in quel punto:

$$(1.3.11) \quad \int_{-\infty}^{+\infty} s(\tau) \delta(t - \tau) d\tau = s(t)$$

Il significato di quest'ultimo integrale è anche quello di una convoluzione tra il segnale $s(t)$ e la funzione delta.

Un cambiamento di scala della variabile indipendente influisce sul risultato:

$$(1.3.12) \quad \int_{-\infty}^{+\infty} x(t) \delta(at + b) dt = \int_{-\infty}^{+\infty} x\left(\frac{\zeta - b}{a}\right) \delta(\zeta) \frac{d\zeta}{|a|} = \frac{1}{|a|} x\left(-\frac{b}{a}\right)$$

Per l'impulso quindi un cambiamento di scala ed una traslazione comporta la variazione dell'area dell'impulso stesso:

$$(1.3.13) \quad \delta(at + b) = \frac{1}{|a|} \delta\left(t + \frac{b}{a}\right)$$

Ultima considerazione è quella relativa alle derivate dell'impulso. La derivata dell'impulso, indicata con $\delta'(t)$ è detta *doppio*:

$$(1.3.14) \quad \int_{-\infty}^{+\infty} x(t) \delta'(t - \tau) dt = -x'(\tau)$$

sempre che $x(t)$ sia dotata di derivata in $t = \tau$. La (1.3.14) si può ricavare dalla definizione dell'impulso (1.3.8) mediante integrazione per parti (ricordando che $\mathcal{D}(AB) = A\mathcal{D}(B) + B\mathcal{D}(A)$, dove $\mathcal{D}(\cdot)$ rappresenta l'operatore di derivazione):

$$(1.3.15) \quad \int_{-\infty}^{+\infty} x(t) \delta'(t - \tau) dt = x(t)\delta(t - \tau)|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} x'(t) \delta(t - \tau) dt = -x'(\tau)$$

Si osservi infine che l'integrale dell'impulso è lo scalino di ampiezza unitaria:

$$(1.3.16) \quad u(t) = \int_{-\infty}^t \delta(\tau) d\tau$$

infatti tale integrale vale zero finchè $t < 0$, ed 1 non appena $t > 0$. Dualmente, la derivata dello scalino unitario è l'impulso unitario: $\frac{d}{dt}u(t) = \delta(t)$

1.3.6. Funzioni sinusoidali. Una classe di funzioni molto utilizzate, soprattutto nell'ambito dell'analisi di funzioni periodiche sono le funzioni sinusoidali. Per la definizione di una funzione sinusoidale sono sufficienti tre elementi: ampiezza A , pulsazione ω_o e fase iniziale φ (cioè l'argomento della senoide per $t = 0$). L'ampiezza rappresenta l'escursione massima che la funzione assume, la frequenza il numero di cicli per unità di tempo che esegue:

$$(1.3.17) \quad A \sin(2\pi f t + \varphi)$$

La senoide si ripete uguale a se stessa ad una distanza temporale T tale che $\omega_o T = 2\pi$. Il periodo di una senoide di pulsazione ω_o è, perciò:

$$(1.3.18) \quad T = \frac{2\pi}{\omega_o}$$

$f = 1/T$ è la *frequenza*. Va da sé che una senoide di frequenza f è periodica di periodo $T = 1/f$ ma, anche, di periodo $2T, 3T, \dots, NT$. Una senoide con fase iniziale $\pi/2$ è chiamata *cosenoide* e vale la relazione $\sin(\omega t + \pi/2) = \cos(\omega t)$. La potenza media di una senoide di ampiezza unitaria vale:

$$(1.3.19) \quad P_m = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \sin^2(\omega t) dt = \frac{1}{2}$$

La sua potenza di picco è

$$(1.3.20) \quad P_p = \max_t \sin^2(\omega t) = 1$$

Il rapporto tra potenza di picco e potenza media è detto *fattore di picco* e, per una senoide vale 2.

1.3.7. Seno cardinale. Un'ultima funzione molto utilizzata è la funzione seno cardinale, così definita:

$$(1.3.21) \quad \text{sinc}(t) = \frac{\sin(\pi \frac{t}{T})}{\pi \frac{t}{T}}$$

e che assume valore pari ad 1 al limite per $t \rightarrow 0$. E' una funzione pari, in quanto rapporto di due funzioni dispari.

1.4. La Correlazione

Dato un segnale deterministico e non periodico, $s(t)$, di esso si può definire, come già visto l'energia:

$$(1.4.1) \quad E_s = \int_{-\infty}^{+\infty} |s(t)|^2 dt = \int_{-\infty}^{+\infty} |S(f)|^2 df$$

dove l'ultima uguaglianza discende dal **teorema di Parseval**, il quale afferma che l'energia del segnale, calcolabile nei due domini tempo e frequenza, non cambia.

Se il segnale passa attraverso un sistema lineare tempo invariante con funzione di trasferimento: $H(f)$:

$$Y(f) = S(f) \cdot H(f)$$

$$(1.4.2) \quad E_y = \int_{-\infty}^{+\infty} |S(f)|^2 \cdot |H(f)|^2 df$$

L'energia si può quindi ottenere conoscendo lo spettro del segnale (e $|S(f)|^2$ è detto **spettro di energia** del segnale) e la funzione di trasferimento del sistema.

1.4.1. Autocorrelazione per segnali ad energia finita. Sia ora $x(t)$ un segnale reale ad energia finita. Si definisce **autocorrelazione** di $x(t)$ la funzione che si ottiene dal seguente integrale:

$$(1.4.3) \quad R_x(\tau) = \int_{-\infty}^{+\infty} x(t)x(t - \tau)dt$$

Dalla definizione si osserva subito che: $R_x(\tau) = x(\tau) \star x(-\tau)$ (per dimostrarlo si

provi a porre $x(-\tau) = y(\tau)$ e ad eseguire l'integrale di convoluzione: $R_x(\tau) = \int_{-\infty}^{+\infty} x(t)y(\tau - t)dt$ e quindi che:

$$(1.4.4) \quad R_x(\tau) = \int_{-\infty}^{+\infty} |X(f)|^2 \cdot e^{j2\pi f\tau} df$$

cioè l'autocorrelazione di un segnale è anche l'antitrasformata del suo spettro di energia. Si ricordi che per un segnale reale, se ad $x(t) \rightarrow X(f)$, allora ad $x(-t) \rightarrow X(-f) = X^*(f)$, mentre per un segnale complesso si ha che se ad $x(t) \rightarrow X(f)$, allora ad $x(-t) \rightarrow X(-f)$, e ad $x^*(t) \rightarrow X^*(-f)$, infine ad $x^*(-t) \rightarrow X^*(f)$.

Poichè quest'ultima definizione vale sempre, allora se il segnale è complesso la definizione di autocorrelazione deve essere adeguatamente modificata:

$$(1.4.5) \quad R_x(\tau) = \int_{-\infty}^{+\infty} x(t)x^*(t - \tau)dt = x(\tau) \star x^*(-\tau)$$

1

Proprietà della funzione di autocorrelazione:

- (1) $R_x(0) = E_x$, cioè la funzione di autocorrelazione calcolata per $\tau = 0$ rappresenta l'energia del segnale
- (2) $R_x(\tau) = R_x(-\tau)$, cioè la funzione di autocorrelazione è una funzione pari ($R_x(\tau) = R_x^*(-\tau)$ per i segnali complessi)
- (3) $|R_x(\tau)| \leq R_x(0)$, cioè il massimo della funzione di autocorrelazione è localizzato in $\tau = 0$:

$$[x(t - \tau) - x(t)]^2 \geq 0 \Leftrightarrow$$

$$x(t - \tau)^2 + x(t)^2 - 2x(t - \tau)x(t) \geq 0$$

ed integrando da $-\infty$ a $+\infty$ si ha: $2E_x \geq 2R_x(\tau)$.

L'autocorrelazione di un segnale ha un'interessante interpretazione fisica. Essa rappresenta una misura del *grado di somiglianza del segnale con sè stesso*. Infatti quanto più un segnale somiglia a sè stesso tanto più è alto il valore dell'integrale in 1.4.3. Ecco quindi il motivo per cui la funzione di autocorrelazione assume valore massimo per $\tau = 0$: quando infatti il segnale è perfettamente sovrapposto a sè stesso il grado di somiglianza è massimo. Per valori di τ crescenti i segnali generalmente tendono

¹Su alcuni testi è riportata la relazione: $R_x(\tau) = \int_{-\infty}^{+\infty} x^*(t)(t - \tau)dt = x^*(\tau) \star x(-\tau)$.

a non somigliare più a sè stessi e quindi il valore dell'autocorrelazione diminuisce. Eccezione notevole a questa regola sono, come si vedrà più avanti, i segnali periodici.

1.4.2. Cross correlazione di due segnali. Dati due segnali $x(t)$ ed $y(t)$, si definisce la crosscorrelazione tra i due segnali come:

$$(1.4.6) \quad R_{xy}(\tau) = \int_{-\infty}^{+\infty} x(t)y(t - \tau)dt = x(\tau) \star y(-\tau)$$

ed anche:

$$(1.4.7) \quad R_{yx}(\tau) = \int_{-\infty}^{+\infty} y(t)x(t - \tau)dt = y(\tau) \star x(-\tau)$$

Per i segnali complessi la definizione è invece:

$$(1.4.8) \quad R_{xy}(\tau) = \int_{-\infty}^{+\infty} x^*(t)y(t - \tau)dt = x^*(\tau) \star y(-\tau)$$

$$(1.4.9) \quad R_{yx}(\tau) = \int_{-\infty}^{+\infty} y^*(t)x(t - \tau)dt = y^*(\tau) \star x(-\tau)$$

Si può facilmente dimostrare che: $R_{xy}(\tau) = R_{yx}^*(-\tau)$:

$$\begin{aligned} R_{xy}(\tau) &= \int_{-\infty}^{+\infty} x^*(t)y(t - \tau)dt = \int_{-\infty}^{+\infty} x^*(z + \tau)y(z)dz = \\ &= \left[\int_{-\infty}^{+\infty} y(z)x^*(z + \tau)dz \right]^{**} = \left[\int_{-\infty}^{+\infty} y^*(z)x(z - (-\tau))dz \right]^* = \\ &= R_{yx}^*(-\tau) \end{aligned}$$

Due segnali si dicono ortogonali se risulta che $R_{xy}(\tau) = 0, \forall \tau$. La cross correlazione dà una misura del grado di somiglianza tra due segnali, analogamente all'autocorrelazione di un segnale.

1.4.3. Segnali a potenza finita.

Per i segnali a potenza finita

$$(1.4.10) \quad P = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-T/2}^{+T/2} |s(t)|^2 dt$$

si può ancora definire una quantità che nel dominio delle frequenze ci dice come sono distribuite le potenze del segnale: la **densità spettrale di potenza** del segnale. Sia infatti: $s_T(t)$ la limitazione di $s(t)$ nell'intervallo: $[-T, T]$:

$$(1.4.11) \quad s_T(t) = \begin{cases} s(t) & |t| \leq T \\ 0 & \text{altrove} \end{cases}$$

Poichè quest'ultimo segnale è sicuramente ad energia finita, per esso si può dare la definizione di trasformata di Fourier e quindi la densità spettrale di energia: $s_T(t) \rightarrow S_T(f)$:

$$(1.4.12) \quad E_T = \int_{-\infty}^{+\infty} |s_T(t)|^2 dt = \int_{-\infty}^{+\infty} |S_T(f)|^2 df$$

Poichè la potenza di $s(t)$ è definita come limite dell'energia della sua limitazione, $s_T(t)$, al tendere dell'intervallo di limitazione all'infinito (e rapportando per l'intervallo di tempo stesso), la densità spettrale di potenza si può scrivere come:

$$(1.4.13) \quad P = \int_{-\infty}^{+\infty} \lim_{T \rightarrow +\infty} \frac{1}{2T} |S_T(f)|^2 df \Rightarrow$$

$$S_p(f) = \lim_{T \rightarrow +\infty} \frac{1}{2T} |S_T(f)|^2$$

La densità spettrale di potenza gode di proprietà simili a quelle della densità spettrale di energia: cioè è una funzione pari (per i segnali reali), è sempre non negativa e il suo integrale su tutto l'asse delle frequenze dà luogo alla potenza del segnale.

Analogamente a ciò che accade per i segnali ad energia finita, il passaggio di un segnale a potenza finita attraverso un sistema lineare tempo invariante dà luogo ad un segnale a potenza finita in uscita, la cui densità spettrale di potenza è pari a: $S_y(f) = S_x(f) \cdot |H(f)|^2$.

Troviamo ora la funzione del tempo che corrisponde alla funzione densità spettrale di potenza:

$$S_p(f) = \lim_{T \rightarrow +\infty} \frac{1}{2T} |S_T(f)|^2 = \lim_{T \rightarrow +\infty} \frac{1}{2T} S_T(f) \cdot S_T^*(f) \Rightarrow$$

antitrasformando:

$$\begin{aligned} &\Rightarrow \lim_{T \rightarrow +\infty} \frac{1}{2T} s_T(\tau) \star s_T(-\tau) = \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} s_T(t) s_T(t + \tau) dt \end{aligned}$$

A tale quantità diamo il nome di funzione di autocorrelazione:

$$(1.4.14) \quad R_g(\tau) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^{+T} s_T(t) s_T(t + \tau) dt$$

La funzione di autocorrelazione per i segnali a potenza finita è l'antitrasformata di Fourier della densità spettrale di potenza, nello stesso modo con cui nel caso di segnali ad energia finita essa è l'antitrasformata di Fourier della densità spettrale di energia.

La funzione di autocorrelazione dei segnali a potenza finita gode delle stesse proprietà della corrispondente funzione definita per i segnali ad energia finita. Inoltre è possibile dare una definizione analoga anche per la cross correlazione di segnali a potenza finita.

1.4.4. Segnali periodici. Sia dato un segnale periodico e la sua rappresentazione in serie di Fourier:

$$s(t) = s(t + n \cdot T)$$

$$(1.4.15) \quad s(t) = \sum_{n=-\infty}^{+\infty} c_n \cdot \exp(j2\pi \frac{n}{T} t)$$

Lo spettro d'ampiezza di un segnale periodico è uno spettro a righe:

$$(1.4.16) \quad S(f) = \sum_{n=-\infty}^{+\infty} c_n \cdot \delta(f - \frac{n}{T})$$

dove i c_n si possono calcolare in base alla trasformata di Fourier di una singola ripetizione del segnale:

$$(1.4.17) \quad c_n = \frac{1}{T} \int_{-T/2}^{+T/2} s(t) \cdot e^{-j2\pi \frac{n}{T} t} dt = \frac{1}{T} S_T(f) \Big|_{f=\frac{n}{T}}$$

I segnali periodici sono ovviamente segnali a potenza finita. La loro densità spettrale di potenza è anch'essa a righe e si può ricavare facilmente :

$$(1.4.18) \quad \begin{aligned} P &= \frac{1}{T} \int_{-T/2}^{+T/2} s(t) s^*(t) dt = \\ &= \frac{1}{T} \int_{-T/2}^{+T/2} \sum_n c_n \cdot e^{j2\pi \frac{n}{T} t} \left[\sum_m c_m \cdot e^{j2\pi \frac{m}{T} t} \right]^* dt = \\ &= \frac{1}{T} \sum_n \sum_m c_n c_m^* \int_{-T/2}^{+T/2} e^{j2\pi \frac{n}{T} t} e^{-j2\pi \frac{m}{T} t} dt = \sum_n |c_n|^2 \Rightarrow \\ S_p(f) &= \sum_{n=-\infty}^{+\infty} |c_n|^2 \cdot \delta\left(f - \frac{n}{T}\right) \end{aligned}$$

La corrispondente funzione di autocorrelazione, essendo un integrale di funzione periodica, è anch'essa periodica di periodo T e la sua definizione si può restringere ad un singolo periodo:

$$(1.4.19) \quad \begin{aligned} R_g(\tau) &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-\infty}^{+\infty} s_T(t) s_T(t + \tau) d\tau = \\ &= \frac{1}{T} \int_{-T/2}^{+T/2} s(t) s(t + \tau) d\tau \end{aligned}$$

CAPITOLO 2

La teoria delle probabilità

2.1. Esperimenti Aleatori

Nelle scienze sperimentali la verifica di una ipotesi di lavoro è affidata all'esperimento. L'esperimento quindi consiste nel controllare che, sotto alcune ipotesi, la teoria e la realtà sono equivalenti, cioè la teoria è descrittiva di un certo fenomeno della natura.

Esempio classico può essere la descrizione della caduta di un grave. Poiché esso segue la legge: $s = \frac{1}{2}gt^2$, si può facilmente determinare quanto tempo il grave impiega a cadere per terra a partire da una certa altezza s con velocità iniziale nulla. I dati raccolti in molte prove ripetute permetteranno di ridurre l'incertezza legata alla misura sperimentale, affetta sempre da una certa dose di errore. Un esperimento di questo tipo, oltre a verificare le ipotesi, ci dice anche un'altra cosa e cioè che se ci poniamo in certe condizioni (un grave cade da una altezza fissa, si riduce al minimo l'effetto della resistenza dell'aria in modo da ridurre l'incertezza della misura, e così via), la realtà *non può fare a meno di comportarsi seguendo determinate leggi*. L'esperimento condotto è cioè di tipo *deterministico*, segue una legge ben precisa e verificabile ogni volta che si desidera, a meno delle inevitabili incertezze dovute alle non perfette condizioni pratiche.

Si supponga ora di voler condurre un altro tipo di esperimento. Si vogliono misurare il numero di autovetture che attraversano un casello autostradale durante una giornata. In questo tipo di esperimento, come si capisce bene, una determinata ipotesi di lavoro come ad esempio che i giorni feriali sono più trafficati di quelli festivi, *non permette di prevedere l'esito dell'esperimento stesso*. La prova che si effettua inoltre darà un risultato diverso giorno per giorno. La prova si dice di tipo *aleatorio*. Per questa classe di esperimenti non è possibile quindi trovare una legge che permetta di predire l'esito dell'esperimento stesso. Tuttavia è possibile trovare una descrizione globale dell'esperimento che permetta cioè di predire, dopo numerose prove, che queste seguono comunque una certa *regolarità statistica*. Il risultato dell'esperimento singolo non è quindi mai prevedibile a priori, ma esso può essere inglobato in una teoria che, entro certi limiti, ne dà una previsione grossolana.

Si supponga, per maggiore chiarezza, di volere osservare i risultati del lancio di un dado. Questo tipo di esperimento appartiene alla classe ora vista, cioè dà luogo ad un risultato che non può essere previsto. Tuttavia dopo il lancio dello stesso dado mille volte, può essere abbastanza ragionevole supporre che la faccia con il numero 6 si sarà presentata all'incirca 167 volte ($\sim 1000/6$). Quindi se il risultato dell'esperimento dà

un valore che è ragionevolmente vicino a questo numero possiamo dire che questo risultato è prevedibile, e possiamo dire anche che il dado si è comportato seguendo le ipotesi iniziali, cioè che non fosse truccato e che tutte e sei le facce avessero la stessa probabilità di presentarsi.

La teoria alla base dei fenomeni della natura che seguono leggi aleatorie è la *teoria delle probabilità*. Questa teoria è stata sviluppata da fisici e matematici come Bernoulli, Pascal e Laplace, durante il XVII e il XVIII secolo e inizialmente fu utilizzata per quantificare le vincite ai tavoli da gioco da gestori di casinò e giocatori d'azzardo.

2.2. Le Basi della Teoria delle Probabilità

Vediamo ora come la teoria delle probabilità permette di modellare un esperimento aleatorio, in modo che si possano ricavare delle leggi applicabili all'esperimento stesso.

Un elemento fondamentale della teoria è quello di ricavare tutti i possibili risultati che l'esperimento stesso è in grado di produrre. Per il lancio di un dado questo è piuttosto facile, dato che lo **spazio campione** dell'esperimento è costituito dai numeri $\{1, 2, 3, 4, 5, 6\}$. In altre situazioni lo spazio campione è più difficile da ottenere. Nell'esperimento descritto precedentemente, delle automobili che transitano da un casello autostradale durante una giornata, si può dire che il risultato è sicuramente un numero intero, zero compreso. Tuttavia è piuttosto difficile indicare il limite superiore di questo intervallo se non intervengono altre ipotesi di lavoro (come ad esempio potrebbero essere il tempo medio di transito, la velocità media delle autovetture sull'autostrada, e così via).

PROPOSITION 2.2.1. Lo spazio campione Ω rappresenta l'insieme dei possibili risultati di un esperimento aleatorio.

Dato inoltre un certo esperimento, come quello delle auto al casello, possono interessare anche determinati gruppi di risultati. Ad esempio potrebbe essere interessante valutare il numero di automobili che transitano al casello in un'ora, oppure il numero di automobili che transita dalle 8.30 alle 11.30 e così via. Questi possibili risultati sono nient'altro che possibili sottoinsiemi dello spazio campione e sono detti **eventi**. Gli eventi devono però soddisfare determinate condizioni per potere essere definiti tali:

- se A è un evento, anche il suo complemento rispetto allo spazio campione, \bar{A} , è un evento;
- se A e B sono eventi, anche $A \cup B$ è un evento.

Utilizzando queste due condizioni si può dimostrare anche che:

- l'intersezione $A \cap B$ di due eventi arbitrari, A e B è un evento (infatti si ha che $A \cap B = \overline{(\bar{A} \cup \bar{B})}$);
- dato un evento A , anche $A \cup \bar{A}$ e $A \cap \bar{A}$ sono eventi. Il primo rappresenta tutto lo spazio campione Ω , il secondo rappresenta l'evento nullo detto anche evento impossibile.

Gli eventi di uno spazio campione costituiscono quindi una classe S cioè un insieme chiuso rispetto alle operazioni di unione e di intersezione.

Un esperimento aleatorio è completamente caratterizzato se sono dati i seguenti tre elementi: i) la descrizione del suo spazio campione Ω , ii) l'individuazione della classe degli eventi S , ed infine iii) la descrizione della **legge di probabilità** $P(\bullet)$, la legge che associa ad ogni evento di S la sua probabilità di presentarsi. La terna $\Omega, S, P(\bullet)$ è detta lo **spazio delle probabilità**. A volte l'esperimento aleatorio viene identificato con il suo spazio delle probabilità, cioè con la sua descrizione matematica astratta.

2.2.1. La probabilità. Varie definizioni ed interpretazioni sono state date alla probabilità. Secondo la teoria assiomatica moderna, dovuta al matematico Kolmogorov, dato un esperimento aleatorio con il suo spazio campione, la legge di probabilità è una corrispondenza che permette di associare ad ogni evento di S un numero reale che soddisfa i seguenti tre assiomi:

- la probabilità di un evento arbitrario è sempre non negativa: $P(A) \geq 0$;
- La probabilità dell'evento certo è pari ad 1: $P(\Omega) = 1$;
- Dati due eventi mutuamente esclusivi, la probabilità dell'evento unione è pari alla somma delle probabilità dei singoli eventi: $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Da questi assiomi si ricavano alcune proprietà (quindi teoremi che si possono dimostrare a partire dagli assiomi):

THEOREM 2.2.2. *Dato un evento A la probabilità dell'evento complementare \bar{A} è pari al complemento ad uno della probabilità di A : $P(\bar{A}) = 1 - P(A)$.*

THEOREM 2.2.3. *L'evento nullo ha probabilità zero di verificarsi: $P(\emptyset) = 0$.*

THEOREM 2.2.4. *La probabilità di un evento A è sempre un numero reale compreso tra zero ed 1: $0 \leq P(A) \leq 1$.*

THEOREM 2.2.5. *Dati due eventi, A e B , la probabilità dell'evento unione è espressa da: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

DIMOSTRAZIONE. $A \cup B = (A \cup B) \cap \Omega = (A \cup B) \cap (A \cup \bar{A}) = (A \cap A) \cup (A \cap \bar{A}) \cup (B \cap A) \cup (B \cap \bar{A}) = A \cup (B \cap \bar{A})$
 $P(A \cup B) = P(A \cup (B \cap \bar{A}))$. Tuttavia, essendo $B = B \cap \Omega = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A})$. Quindi: $P(B) = P(B \cap A) + P(B \cap \bar{A})$, da cui la tesi. \square

La probabilità intersezione di due eventi è anche detta probabilità congiunta, mentre le probabilità dei due eventi, prese separatamente, sono dette probabilità marginali. Data

una coppia di eventi, A e B con $P(B) \neq 0$, la *probabilità di A condizionata all'evento B* , indicata con $P(A/B)$ è definita dalla relazione:

$$(2.2.1) \quad P(A/B) = \frac{P(A \cap B)}{P(B)}$$

La probabilità di A , presa separatamente, è detta *probabilità a priori*, mentre la probabilità di A noto anche l'evento B , cioè $P(A/B)$ è detta *probabilità a posteriori*. L'evento B condiziona l'evento A e quindi ne modifica la sua probabilità, una volta che esso si sia verificato. Da questa osservazione nasce la definizione stessa nella quale l'evento congiunto è *rinormalizzato* per la probabilità di B che funge quindi da nuovo spazio campione (da definizione infatti: $P(B/B) = 1$).

EXAMPLE 2.2.6. Supponiamo di voler studiare l'esperimento aleatorio che modella il lancio di un dado non truccato. Lo spazio campione, costituito dall'insieme dei possibili risultati, è dato da: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ dove ω_i rappresenta il risultato della faccia i – *sima* al termine dell'esperimento. La classe S di tutti i possibili eventi è costituita da 2^6 possibili valori, compresi Ω e \emptyset . La legge di probabilità resta assegnata non appena si assegna una probabilità a ciascuno dei risultati dello spazio dei campioni ω_i . Poichè abbiamo ritenuto il dado non truccato e quindi è ragionevole supporre che in un lancio tutte le facce di un dado abbiano uguale possibilità di presentarsi, si può ritenere che:

$$(2.2.2) \quad P(\omega_i) = \frac{1}{6}$$

A questo punto è possibile definire un qualsiasi evento e trovare la sua probabilità di occorrenza. Si voglia ad esempio determinare la probabilità che lanciando il dado, appaiano numeri inferiori a 3. Questa probabilità è la probabilità che accada: $P(A) = P(\omega_1 \cup \omega_2)$. Poichè questi eventi sono disgiunti, la probabilità della loro unione è anche pari alla somma delle loro probabilità: $P(A) = P(\omega_1) + P(\omega_2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

In casi semplici come questo, dove lo spazio dei campioni è finito ed è simmetrico (cioè vi è *equiprobabilità* di tutti i possibili risultati dello spazio campione Ω), è possibile utilizzare la definizione classica di probabilità dovuta a Laplace. Questa definizione parte dall'osservazione dei casi favorevoli nell'insieme di tutti i casi possibili che si possono verificare. Detta allora N il numero di tutti i casi possibili ed N_A quelli favorevoli all'evento A , la probabilità cercata è data dal rapporto:

$$(2.2.3) \quad P(A) = \frac{N_A}{N}$$

L'ipotesi cruciale alla base di questa definizione sta nel fatto che tutti i risultati dello spazio campione hanno pari probabilità di verificarsi. Nell'ipotesi in cui non vi sia equiprobabilità dei risultati dello spazio campione la definizione precedente non è più adeguata e si ricorre allora ad un approccio di tipo sperimentale. Si supponga di effettuare un numero molto alto di lanci N e di collezionare il numero di volte che l'evento A si verifica, N_A . All'aumentare di N si comincia a notare una certa regolarità nella relazione che esiste tra il numero di lanci e il numero di volte che A si verifica. La frequenza relativa con cui si verifica A , cioè: N_A/N tende allora, per un numero di lanci molto elevato, alla probabilità, secondo la definizione di Von Mises:

$$(2.2.4) \quad P(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}$$

Questa definizione, seppure non corrispondente alla visione moderna (assiomatica) della teoria delle probabilità, ha il vantaggio di prescindere dalla simmetria (e quindi equiprobabilità) del problema in esame.

Si osservi che la definizione di Von Mises non è in contrasto con quella assiomatica di Kolmogorov, dato che il rapporto tra due numeri positivi è sempre positivo. Se inoltre A è un sottinsieme di Ω , accade sempre che $N_A \leq N$, e quindi che $0 \leq P(A) \leq 1$. Inoltre si può osservare che, detti A e B due eventi disgiunti, e dette N_A ed N_B le loro occorrenze su un numero totale di esperimenti pari ad N , la probabilità dell'evento unione:

$$(2.2.5) \quad P(A \cup B) = \lim_{N \rightarrow \infty} \frac{N_{A \cup B}}{N} = \lim_{N \rightarrow \infty} \frac{N_A + N_B}{N} = P(A) + P(B)$$

e quindi gli assiomi di Kolmogorov sono verificati.

PROPOSITION 2.2.7. *Due eventi A e B sono detti indipendenti se la probabilità marginale di A e la probabilità di A condizionata a B sono uguali, cioè se:*

$$(2.2.6) \quad P(A) = P(A/B)$$

Partendo dalla definizione della probabilità condizionata, questo significa che:

$$(2.2.7) \quad P(A) = P(A/B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A) \cdot P(B) = P(A \cap B)$$

I due eventi sono detti indipendenti quando la probabilità congiunta è pari al prodotto delle singole probabilità. L'indipendenza tra i due eventi è esplicitata nel fatto che la probabilità dell'evento A è uguale a priori ed a posteriori dell'evento B . L'evento B quindi non ha alcuna influenza su A , cioè i due eventi sono tra loro indipendenti.

Dalla definizione di probabilità condizionata nasce anche la seguente osservazione:

$$(2.2.8) \quad P(A/B) \cdot P(B) = P(B/A) \cdot P(A) \Rightarrow P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

nota anche con il nome di *teorema (o formula) di Bayes*. IL teorema di Bayes è noto anche con il nome di **teorema delle probabilità totali**.

Si consideri infatti una certa partizione dello spazio dei campioni Ω , fatto da N eventi disgiunti tra loro: B_1, B_2, \dots, B_N , con $B_i \cap B_j = \emptyset$ e $\bigcup_i B_i = \Omega$. La probabilità di un dato evento A si può allora calcolare in base alla conoscenza delle probabilità condizionate di A con le B_i :

$$(2.2.9) \quad P(A) = P(A \cap \Omega) = P(A \cap \bigcup_{i=1}^N B_i) = P(\bigcup_{i=1}^N (A \cap B_i)) = \sum_{i=1}^N P(A \cap B_i)$$

da cui si ricava, ricordando la relazione che esiste tra la probabilità congiunta e quella condizionata:

$$(2.2.10) \quad P(A) = \sum_{i=1}^N P(A/B_i) \cdot P(B_i)$$

2.2.2. Esperimento composto. Si considerino ora due esperimenti aleatori differenti tra loro e caratterizzati dagli spazi campione Ω_1 ed Ω_2 . Si può pensare un esperimento composto come la contemporanea osservazione dei due esperimenti. Lo spazio campione sarà allora il prodotto cartesiano dei due spazi campione: $\Omega_1 \times \Omega_2$ e gli elementi di questo spazio sono le coppie ordinate che si ottengono dalla combinazione di tutti i possibili risultati di Ω_1 con quelli di Ω_2 . I due esperimenti naturalmente possono fare riferimento a due esperienze uguali (ad esempio due lanci di dadi) o a due completamente differenti, come ad esempio il lancio di un dado e l'estrazione di una carta da un mazzo di 52 carte francesi.

Sia ora A_1 un evento del primo spazio campione ed A_2 un evento del secondo. Si voglia studiare la probabilità dell'evento composizione dei due eventi A_1 ed A_2 , cioè: $A = A_1 \times A_2$. Se i due eventi fossero indipendenti è evidente che la probabilità dell'evento A è pari al prodotto delle due probabilità: $P(A) = P(A_1) \cdot P(A_2)$. Se invece i due esperimenti sono tra loro in qualche modo legati è necessario valutare il grado di *correlazione* dei due eventi e quindi la probabilità non è più pari al prodotto delle due probabilità. E' ad esempio evidente che se si vuole stabilire la probabilità di un evento come l'estrazione di un numero dispari da un lancio di un dado e di un asso da un mazzo di carte, avremo: $P(A_{disp} \cap A_{asso}) = P(A_{disp}) \cdot P(A_{asso}) = \frac{1}{2} \cdot \frac{4}{52} = \frac{1}{26}$. Le considerazioni fatte per la composizione di due esperimenti si possono fare per la composizione di N qualunque esperimenti, ricordando però che in generale, dalla conoscenza delle leggi di probabilità dei singoli esperimenti non è possibile determinare la legge di probabilità dell'esperimento composto. In tale ambito ricade il problema delle *prove ripetute ed indipendenti*. Caso notevole è quello delle prove binarie ripetute ed indipendenti o *prove di Bernoulli*.

EXAMPLE 2.2.8. Formula di Bernoulli. Si supponga di voler indagare sull'esperimento composto da n esperimenti uguali tra loro ed indipendenti. Ciascuno degli esperimenti dà luogo ad uno spazio dei campioni con due soli possibili risultati: ω_o ed ω_1 , con $P(\omega_o) = p$ e $P(\omega_1) = 1 - p$. Un classico esempio è il lancio di n monete, o anche il lancio di una stessa moneta, purchè il risultato sia la composizione dei singoli lanci. Si costruisca ora l'evento $A = \omega_o$ si presenta k volte negli n esperimenti (o prove ripetute). La formula di Bernoulli (o binomiale) dice che:

$$(2.2.11) \quad P(A) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

ove il coefficiente binomiale vale: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

1

¹Si ricordi che il modo con cui possono essere disposti k oggetti in n differenti posizioni, distinguendo i gruppi anche per l'ordine, è dato dal numero $D_{n,k} = n \cdot (n-1) \cdot \dots \cdot (n-k+1)$, chiamato

2.3. Variabili Aleatorie

Si consideri l'esperimento aleatorio costituito dal lancio di un dado. Sappiamo già che il suo spazio campione è costituito da tutti i possibili valori che possono essere ottenuti, e cioè i numeri da 1 a 6. Questi stessi numeri li potremmo ottenere anche con altri esperimenti aleatori (ad esempio un qualche esperimento che consideri i giorni della settimana lavorativi).

Quello che si può osservare da un insieme di esperimenti di questo tipo è la comune cardinalità dello spazio campione, sebbene gli elementi dello spazio campione siano differenti. Se allora astraiamo i casi particolari che abbiamo ottenuto, è possibile numerare gli elementi (od i risultati) dello spazio campione, sino ad ottenere il valore associato a ciascuno dei possibili risultati. Quindi in questo modo l'esito del lancio di un dado diventa l'insieme dei numeri da 1 a 6, mentre l'esito di un qualche esperimento che coinvolga i giorni della settimana lavorativi diventa, ancora una volta, l'insieme dei numeri da 1 a 6.

Abbiamo costruito quindi una quantità variabile a seconda del risultato dell'esperimento. A questa quantità è dato il nome di *variabile aleatoria*.

Formalmente si può definire la variabile aleatoria come segue.

PROPOSITION 2.3.1. *Dato un esperimento aleatorio avente come spazio campione Ω , come classe degli eventi S e come legge di probabilità $P(\bullet)$, si definisce una corrispondenza che associ a ciascun risultato dello spazio Ω un unico numero reale. Tale corrispondenza tra l'asse reale e lo spazio Ω è detta variabile aleatoria se l'insieme dei risultati per i quali è verificata la disuguaglianza $X(\omega_i) \leq a$ è un evento, comunque si scelga il numero reale a .*

La variabile aleatoria si introduce ogni volta che il risultato di un esperimento aleatorio è un valore numerico, come ad esempio una misura. Per quanto preciso ed accurato possa essere lo strumento, ripetendo più volte un esperimento (anche deterministico!) si otterranno di volta in volta valori differenti, dovuti agli errori di misura. L'insieme delle misure ottenute rappresenta proprio una variabile aleatoria, per l'effetto di incertezza dovuto all'errore di misura.

Rimane ora il problema di come trasferire la legge di probabilità alle variabili aleatorie. Vogliamo cioè essere in grado di stabilire qual è la probabilità di un evento, quando questo sia definito sull'asse dei numeri reali e non nella classe degli eventi S . In particolare, dati due numeri reali a e b , con $a < b$, ha interesse determinare qual è

disposizioni di n oggetti in classe k .

Le disposizioni di n oggetti in classe n , cioè il modo con cui possono essere disposti n oggetti distinguendoli solo per l'ordine che assumono nelle n posizioni è detto **permutazioni** in classe n e vale: $P_n = n!$.

Infine si dicono **combinazioni** di n oggetti in classe k il modo con cui è disporre k oggetti in n differenti posizioni, non distinguendoli per l'ordine. E' quindi il numero di disposizioni $D_{n,k}$ diviso il numero delle permutazioni di k oggetti: $C_{n,k} = D_{n,k}/P_k = \frac{n!}{k!(n-k)!} = \binom{n}{k}$. Il numero $\binom{n}{k}$ è detto anche coefficiente binomiale.

la probabilità che la variabile aleatoria sia compresa tra a e b , cioè $P(a < X \leq b)$. Estendendo il linguaggio usato solo nell'ambito degli esperimenti aleatori, si definirà evento anche l'intervallo di valori sull'asse reale compreso tra a e b , dato che, per la definizione di variabile aleatoria, l'intervallo $]a, b]$ è associabile ad un dato evento di S .

Questa operazione di “determinazione” della legge di probabilità di un dato evento definito direttamente sull'asse reale diventa immediato se si introduce una funzione, la **funzione distribuzione di probabilità**: $F_X(x)$, definita come segue:

$$(2.3.1) \quad F_X(x) = P(X \leq x)$$

dove x è un numero reale ben definito. La funzione di distribuzione di probabilità è una funzione che associa ad ogni numero reale il valore della probabilità dell'evento identificato dall'intervallo $X \leq x$. Per $F_X(x)$ valgono le seguenti proprietà:

- (1) $0 \leq F_X(x) \leq 1$
- (2) Il suo valore limite, per $x \rightarrow +\infty$ vale 1: $\lim_{x \rightarrow +\infty} F_X(x) = F_X(+\infty) = P(X \leq +\infty) = 1$
- (3) Il suo valore limite per $x \rightarrow -\infty$ vale 0: $\lim_{x \rightarrow -\infty} F_X(x) = F_X(-\infty) = P(X \leq -\infty) = 0$
- (4) La funzione è monotona non decrescente, cioè se $x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$
- (5) La funzione è continua da destra, cioè $F_X(x) = \lim_{h \rightarrow 0^+} F_X(x + h)$
- (6) Se la funzione di distribuzione presenta una discontinuità di prima specie nel punto \bar{x} , allora la differenza tra il limite a destra e quello a sinistra è proprio il valore della probabilità dell'evento in $X = \bar{x}$: $P(X = \bar{x}) = \lim_{h \rightarrow 0^+} F_X(\bar{x} + h) - \lim_{h \rightarrow 0^-} F_X(\bar{x} + h)$
- (7) La probabilità dell'evento $a < X \leq b$ può essere calcolata tramite la relazione: $F_X(b) - F_X(a)$.

Le variabili aleatorie possono essere suddivise in tre classi: variabili aleatorie *continue*, variabili aleatorie *discrete* e variabili aleatorie *miste*. Una variabile aleatoria è detta discreta se la sua funzione di distribuzione è continua a tratti: $F_X(x) = \sum_k P(X = x_k) \cdot u(x - x_k)$. Tenendo conto delle ultime due proprietà viste precedentemente questo significa che la variabile aleatoria assume valore solo in un numero discreto (cioè con cardinalità pari a quella dei numeri naturali) di valori, e non continuo. Le posizioni in cui questo accade sono proprio le x_k . In queste posizioni la probabilità dell'evento è “concentrata” nel valore x_k : $p_k = P(X = x_k)$. Le p_k sono dette anche *masse* di probabilità.

Se invece abbiamo a che fare con una distribuzione di probabilità continua, allora l'insieme dei valori che può assumere la funzione $F_X(x)$ si distribuisce con continuità

sull'asse dei numeri reali. L'insieme degli eventi a cui è associata tale v.a. è un infinito di cardinalità pari a quello dei numeri reali, quindi la probabilità che la variabile aleatoria assuma un certo valore x è un infinitesimo, tende cioè a zero.

Una variabile aleatoria mista è una variabile aleatoria continua quasi ovunque, tranne che per un numero finito (o un'infinità numerabile) di punti per i quali presenta discontinuità.

2.4. Densità di Probabilità

Una descrizione alternativa di una variabile aleatoria è data anche della funzione densità di probabilità, $f_X(x)$, definita dalla relazione:

$$(2.4.1) \quad f_X(x) = \frac{dF_X(x)}{dx}$$

La relazione inversa è invece:

$$(2.4.2) \quad F_X(x) = \int_{-\infty}^x f_X(x) dx$$

La funzione densità di probabilità è ovviamente non negativa, discendendo dalla derivazione di una funzione monotona non decrescente, inoltre la sua area vale sempre 1:

$$(2.4.3) \quad \int_{-\infty}^{+\infty} f_X(x) dx = 1$$

Il nome di densità di probabilità discende dalla sua stessa definizione. Infatti si supponga di considerare un intervallino molto piccolo: $[x, x + \Delta x]$ e di voler calcolare la probabilità che X capiti in quell'intervallo: $P(x < X \leq x + \Delta x)$. Per definizione si ha:

$$(2.4.4) \quad \begin{aligned} P(x < X \leq x + \Delta x) &= \int_x^{x+\Delta x} f_X(x) dx \approx f_X(x) \cdot \Delta x \Rightarrow \\ f_X(x) &= \frac{P(x < X \leq x + \Delta x)}{\Delta x} \end{aligned}$$

cioè la funzione densità di probabilità in un punto rappresenta il valore della probabilità che si può calcolare in un intervallino nell'intorno di quel punto diviso l'ampiezza di quell'intervallino. La sua misura è quindi una *misura di densità*, cioè di come la probabilità si *addensa* attorno ai vari valori che la variabile aleatoria può assumere sull'asse reale.

Poichè la funzione distribuzione di probabilità può essere continua, discreta o mista, anche per la densità di probabilità dovremmo distinguere i vari casi. Quando la funzione di distribuzione è discreta o mista, essa è costituita da un insieme (anche infinito) di discontinuità di prima specie. Conseguentemente in questi punti la funzione non è, a rigore, derivabile e quindi non si potrebbe definire la densità di probabilità.

Tuttavia di una variabile aleatoria discreta è stata data una descrizione in termini di distribuzione di probabilità che introduceva l'uso dei gradini. Difatti il gradino dà informazione del "salto" di probabilità che è avvenuto in un certo punto a causa della presenza di una certa massa di probabilità. Una funzione di distribuzione di probabilità discreta è rappresentata in figura (2.4.1)

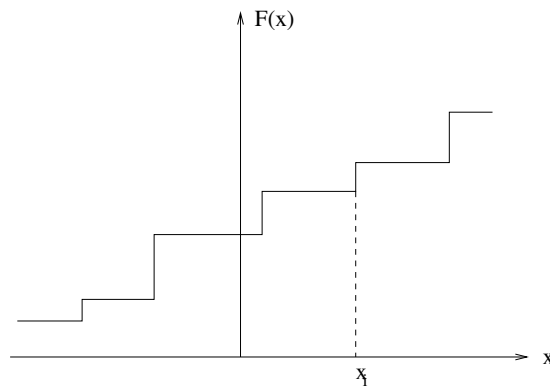


FIGURA 2.4.1. Distribuzione di probabilità di una variabile aleatoria discreta

Se allora si considera la descrizione per gradini è possibile introdurre, come densità di probabilità, una densità che sia costituita da impulsi nelle posizioni delle discontinuità e sia uguale a zero altrove. Gli impulsi infatti rappresentano, nella descrizione della densità di probabilità, un valore "concentrato" e non distribuito della probabilità, un valore cioè che assume una densità infinita, dovendo essere definita in un solo punto matematico (vedi figura (2.4.2)).

Da un punto di vista della rappresentazione matematica si ha:

$$(2.4.5) \quad F_X(x) = \sum_k P(X = x_k) \cdot u(x - x_k) \Rightarrow f_X(x) = \sum_k P(X = x_k) \cdot \delta(x - x_k)$$

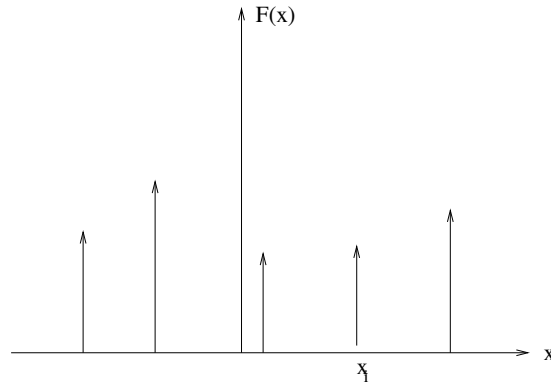


FIGURA 2.4.2. Densità di probabilità di una variabile aleatoria discreta

2.5. Operazioni sulla Variabile Aleatoria

Nei problemi che coinvolgono una variabile aleatoria può essere comune l'esigenza di dover effettuare alcune operazioni su di essa. In particolare, data una variabile aleatoria X , si pone il problema di come determinare le caratteristiche della variabile aleatoria ottenuta come $Y = g(X)$, dove $g(\bullet)$ è una funzione deterministica definita sull'asse reale (e dotata di determinate proprietà). Un esempio può essere dato dalla tensione di rumore ai capi di una resistenza. Questa quantità può essere descritta mediante una variabile aleatoria, X , dato che il fenomeno che sta alla base della tensione di rumore è un fenomeno di tipo statistico. Se ora si vuole misurare la potenza di rumore dissipata sul resistore, poichè la potenza su un resistore è sempre pari a $P_R = x^2/R$, sarà anch'essa una variabile aleatoria, ottenuta come prodotto di una costante (il valore della resistenza) per il quadrato di una quantità aleatoria. Se dunque X varia in modo imprevedibile, ma con una certa legge di probabilità, ci si può chiedere come varia la potenza P_R . Questa nuova variabile aleatoria si può ottenere trasformando la variabile aleatoria originaria.

Sia $y = g(x)$. Si vuole determinare: $F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$. Si devono allora prendere tutti i valori di x , per i quali risulta $g(x) \leq y$. Detto D_Y questo insieme: $D_Y = \{x \ni g(x) \leq y\}$, si ha che: $F_Y(y) = \int_{D_Y} f_X(x) dx$. Da questa si ricava poi la densità di probabilità: $f_Y(y) = \frac{dF_Y(y)}{dy}$.

Si supponga in particolare che la funzione $g(\bullet)$ sia monotona strettamente crescente. In tal caso è possibile definire la sua inversa: $g^{-1}(\bullet)$ ed è immediata la relazione per determinare la densità di probabilità di Y :

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) \Rightarrow$$

$$(2.5.1) \quad f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{dg^{-1}(y)}{dy} = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

se la funzione è monotona strettamente decrescente invece si ha:

$$(2.5.2) \quad f_Y(y) = -f_X(g^{-1}(y)) \cdot \frac{dg^{-1}(y)}{dy} = -\frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

La relazione generale si può quindi riassumere nella seguente formula:

$$(2.5.3) \quad f_Y(y) = \int_{d_Y} \frac{f_X(x)}{|g'(x)|} dx$$

dove d_Y è l'insieme di tutti i valori x che sono soluzioni dell'equazione $g(x) = y$. Naturalmente l'insieme delle soluzioni di $g(x) = y$ può anche essere l'insieme vuoto, nel qual caso si ha ovviamente: $f_Y(y) = 0$. Il caso in cui invece risulta: $g'(x) = 0$ è trattato differientemente a seconda che anche $f_X(x)$ sia nullo oppure no. Nel primo caso sono costanti sia $F_X(x)$ che $g(x)$ quindi risulterà: $P(Y = y) = P(X \in I)$ con I intervallo delle x in cui $g(x)$ assume valore costante. Nel secondo caso $f_Y(y)$ tenderà ad un valore infinito (cioè ad un impulso).

2.6. Parametri Statistici di una Variabile Aleatoria

Nelle situazioni reali non è sempre possibile avere a disposizione tutte le conoscenze necessarie per caratterizzare una variabile aleatoria. Il massimo di informazione che si può trarre da un esperimento aleatorio è la determinazione della sua funzione densità di probabilità. Quando questa funzione non si conosce è comunque possibile determinare alcuni parametri statistici che, seppure non permettono una conoscenza completa della variabile aleatoria, permettono di estrarne qualche proprietà.

Il più importante di questi parametri statistici è il valore atteso o media, μ_x , definito dalla seguente relazione:

$$(2.6.1) \quad \mu_X = \int_{-\infty}^{+\infty} x f_X(x) dx$$

e rappresenta una sorta di “baricentro” della funzione densità di probabilità (si confronti a tale proposito la media con le definizioni, meno note di *moda* e *mediana*). Se la variabile aleatoria è discreta la relazione precedente, a causa della presenza degli

impulsi, diventa una sommatoria:

$$(2.6.2) \quad \mu_X = \int_{-\infty}^{+\infty} x f_X(x) dx = \sum_k p_k \cdot \int_{-\infty}^{+\infty} x \delta(x - x_k) dx = \sum_k x_k p_k$$

L'operazione precedente di media può essere scritta molto più facilmente introducendo l'**operatore di aspettazione** (o di valor medio):

$$(2.6.3) \quad E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

che nel caso della media assume la semplice relazione: $\mu_X = E[X]$. L'operatore di valor medio gode della proprietà di linearità, dato che è definito attraverso un'operazione di integrazione: $E[a \cdot g(X) + b \cdot h(X)] = a \cdot E[g(X)] + b \cdot E[h(X)]$. Inoltre, si supponga di avere una variabile aleatoria Y ottenuta tramite trasformazione della v.a. X attraverso la funzione $y = g(x)$. Senza passare attraverso il calcolo (a volte difficoltoso) della densità di probabilità di Y nota quella di X è possibile determinare il valor medio di Y :

$$(2.6.4) \quad \mu_Y = E[Y] = E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

Questo risultato è noto con il nome di **teorema del valor medio**.

Due v.a. possono possedere lo stesso valor medio ed essere molto differenti tra loro. In particolare è possibile che le v.a. abbiano una densità di probabilità che sia in un caso molto "stretta", nell'altro molto "larga". Si confrontino le due densità in figura (2.6.1).

Questo fatto suggerisce che, seppure con una media uguale, le due v.a. hanno comportamenti molto differenti tra loro. Nel caso della v.a. con densità di probabilità molto larga è più probabile che capitino valori della v.a. lontani dal valor medio, cosa invece meno probabile nel secondo caso. E' possibile allora quantificare questo fatto statistico introducendo un nuovo parametro, la **varianza**, che è definita come segue:

$$(2.6.5) \quad \sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx$$

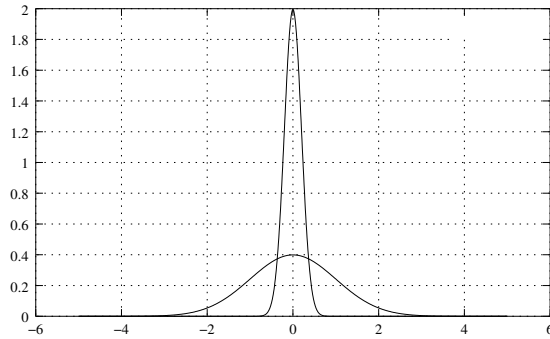


FIGURA 2.6.1. Confronto tra due densità di probabilità con la stessa media

La radice quadrata della varianza è detta **deviazione standard** e rappresenta una misura di quanto “dispersa” sia la densità di probabilità attorno alla media (più grande è la deviazione standard, maggiore la dispersione). Una v.a. che non presenti affatto dispersione attorno alla media (cioè con $\sigma_X = 0$) sarebbe tutta concentrata sulla media, cioè avrebbe una densità di probabilità pari ad un impulso di area unitaria posto sulla posizione della media (ovviamente in questo caso non si può parlare di densità di probabilità vera e propria, dato che i possibili valori collassano su unico valore certo).

Il valore quadratico medio (chiamato a volte anche **potenza**) è definito come segue:

$$(2.6.6) \quad m_X^2 = E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

L'operatore $E[\bullet]$ è un operatore lineare, quindi è possibile trovare la relazione che lega tra loro varianza e potenza:

$$(2.6.7) \quad \begin{aligned} \sigma_X^2 &= E[(X - \mu_X)^2] = E[X^2 - 2X\mu_X + \mu_X^2] = E[X^2] - 2E[X] \cdot \mu_X + \mu_X^2 = \\ &= m_X^2 - 2\mu_X^2 + \mu_X^2 = m_X^2 - \mu_X^2 \end{aligned}$$

2.7. Esempi di Variabili Aleatorie

2.7.1. Variabile aleatoria uniforme. Una variabile aleatoria uniforme presenta una densità di probabilità costante in tutto l'intervallo in cui è definita, $[a, b]$ e valore nullo al di fuori di questo. Conseguentemente, dato che l'area sottesa dalla densità di probabilità deve essere unitaria, l'altezza di tale valore costante è: $1/(b - a)$. La densità di probabilità si può quindi scrivere come:

$$(2.7.1) \quad f_X(x) = \frac{1}{b-a} \operatorname{rect}\left(\frac{x - \frac{b+a}{2}}{b-a}\right)$$

La v.a. non può assumere mai valori al di fuori dell'intervallo $[a, b]$, ma dentro di questo intervallo la probabilità di occorrenza di tutti i possibili valori è uguale (è come se fosse un dado "continuo", dotato cioè di infinite facce).

La funzione di distribuzione, essendo la funzione integrale della densità di probabilità avrà comportamento a "rampa" nell'intervallo in cui la funzione di densità è non nulla:

$$(2.7.2) \quad F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Gli andamenti della funzione di densità e di quella di distribuzione sono mostrati in figura (2.7.1).

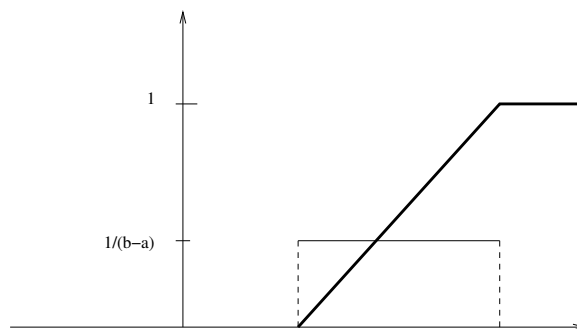


FIGURA 2.7.1. Densità e distribuzione della v.a. uniforme

Si possono calcolare facilmente i suoi parametri statistici:

$$(2.7.3) \quad \mu_X = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{b+a}{2}$$

$$\sigma_X^2 = \int_a^b \left(x - \frac{b+a}{2}\right)^2 \cdot \frac{1}{b-a} dx =$$

$$(2.7.4) \quad \frac{1}{b-a} \left(\frac{b^3 - a^3}{3} - \frac{(b+a) \cdot (b^2 - a^2)}{2} + \frac{(b^2 + a^2 + 2ab)(b-a)}{4} \right) = \frac{(b-a)^2}{12}$$

$$(2.7.5) \quad m_X^2 = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

2.7.2. Variabile aleatoria esponenziale. Una variabile aleatoria molto utilizzata è la cosiddetta variabile aleatoria continua esponenziale unilatera o semplicemente esponenziale, così definita:

$$(2.7.6) \quad f_X(x) = \frac{1}{\eta} \cdot \exp\left(-\frac{x}{\eta}\right) \cdot u(x)$$

dove $u(x)$ è il gradino unitario con discontinuità in $x = 0$. Il significato del parametro reale e positivo η sarà chiaro in seguito, quando si vedrà uno dei più comuni utilizzi della v.a. esponenziale, cioè nei problemi di affidabilità e calcolo del rischio.

La distribuzione di probabilità esponenziale vale:

$$(2.7.7) \quad F_X(x) = \int_0^x \frac{1}{\eta} \cdot \exp\left(-\frac{x}{\eta}\right) dx = \left[1 - \exp\left(-\frac{x}{\eta}\right)\right] \cdot u(x)$$

ed entrambe sono illustrate in figura (2.7.2).

I suoi parametri statistici valgono:

$$(2.7.8) \quad \mu_X = \int_0^{+\infty} x \cdot \frac{1}{\eta} \cdot \exp\left(-\frac{x}{\eta}\right) dx = \eta$$

$$(2.7.9) \quad m_X^2 = \int_0^{+\infty} x^2 \cdot \frac{1}{\eta} \cdot \exp\left(-\frac{x}{\eta}\right) dx = 2\eta^2$$

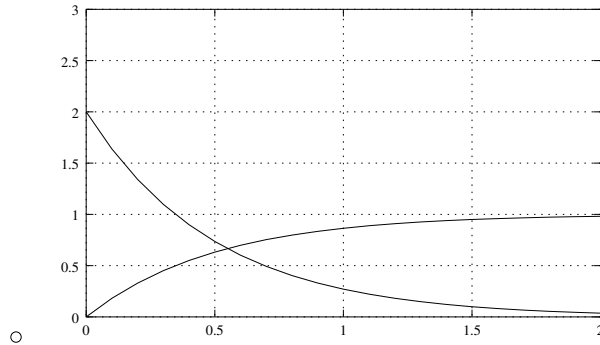


FIGURA 2.7.2. Densità e distribuzione della v.a. esponenziale

$$(2.7.10) \quad \sigma_X^2 = \int_0^{+\infty} (x - \eta)^2 \cdot \frac{1}{\eta} \cdot \exp\left(-\frac{x}{\eta}\right) dx = \eta^2$$

La v.a. esponenziale è spesso utilizzata (in ambito telecomunicazionistico) nella seguente forma:

$$(2.7.11) \quad f_X(x) = \lambda \cdot \exp(-\lambda x) \cdot u(x)$$

dove $\lambda = 1/\eta$ assume il significato di *rate* della v.a. esponenziale.

2.7.3. Variabile aleatoria di Poisson. La variabile aleatoria di Poisson è una v.a. discreta con densità di probabilità:

$$(2.7.12) \quad f_Z(z) = \sum_{n=0}^{+\infty} e^{-\Lambda} \frac{\Lambda^n}{n!} \delta(z - n)$$

dove il parametro Λ caratterizza la v.a. discreta. La v.a. di Poisson assume valori di probabilità (di massa) differenti da zero solo per valori interi e non negativi. La variabile aleatoria di Poisson e quella esponenziale sono in realtà legate tra loro, come si vedrà in seguito. Esse modellano bene fenomeni come il conteggio del numero di clienti che paga ad una cassa di un supermercato nell'unità di tempo o il numero di automobili che transita ad un casello autostradale o il numero di elettroni che transita attraverso una giunzione np .

La funzione di distribuzione essendo l'integrale della $f_Z(z)$ precedente, è molto semplice:

$$(2.7.13) \quad F_Z(z) = \sum_{n=0}^{+\infty} e^{-\Lambda} \frac{\Lambda^n}{n!} u(z-n)$$

dovendo integrare solo la variabile z . Un andamento della massa di probabilità per $\Lambda = 3$ è mostrato in figura (2.7.3).

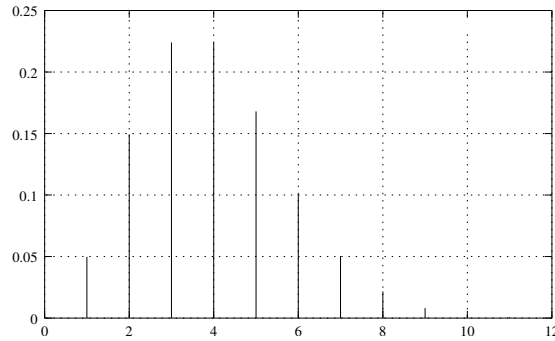


FIGURA 2.7.3. Densità e distribuzione della v.a. di Poisson

I suoi parametri statistici sono:

$$(2.7.14) \quad \mu_Z = \int_0^{+\infty} z \cdot \sum_{n=0}^{+\infty} e^{-\Lambda} \frac{\Lambda^n}{n!} \delta(z-n) dz = \sum_{n=0}^{+\infty} e^{-\Lambda} \frac{\Lambda^n}{n!} n = e^{-\Lambda} \cdot \sum_{n=1}^{+\infty} \frac{\Lambda^n}{n!} n = \Lambda$$

$$m_Z^2 = \int_0^{+\infty} z^2 \cdot \sum_{n=0}^{+\infty} e^{-\Lambda} \frac{\Lambda^n}{n!} \delta(z-n) dz = e^{-\Lambda} \cdot \sum_{n=0}^{+\infty} \frac{\Lambda^n}{n!} n^2 =$$

$$(2.7.15) \quad e^{-\Lambda} \cdot \sum_{n=1}^{+\infty} \frac{\Lambda^n}{(n-1)!} (n-1+1) = \Lambda e^{-\Lambda} \cdot \sum_{n=2}^{+\infty} \frac{\Lambda^{n-1}}{(n-1)!} (n-1) + e^{-\Lambda} \Lambda \cdot \sum_{n=2}^{+\infty} \frac{\Lambda^{n-1}}{(n-1)!} = \Lambda^2 + \Lambda$$

$$(2.7.16) \quad \sigma_Z^2 = m_Z^2 - \mu_Z^2 = \Lambda$$

Quindi per la v.a. di Poisson il parametro caratteristico Λ rappresenta sia il valor medio sia la varianza.

2.7.4. Variabile aleatoria di binomiale. Considerato un esperimento che conduce a due soli possibili risultati (successo, con probabilità p e insuccesso, con probabilità $1-p$), la variabile aleatoria binomiale (o di Bernoulli) conta il numero di successi accaduti in n esperimenti aleatori di questo tipo indipendenti tra loro:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, \dots, n$$

Questa v.a. è discreta, quindi hanno ovvia formulazione sia la distribuzione sia la densità di probabilità. La media vale:

$$\mu_X = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{n(n-1)!}{k(k-1)!(n-k)!} p^k (1-p)^{n-k} = np$$

la varianza vale invece:

$$\sigma_X^2 = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k (1-p)^{n-k} = np(1-p)$$

2.7.5. Variabile aleatoria geometrica. Considerati n esperimenti aleatori indipendenti di Bernoulli la v.a. geometrica conta qual è il numero di successi da osservare prima di registrare il primo insuccesso:

$$P(X = k) = p^k (1-p) \quad k = 0, \dots, \infty$$

La media vale:

$$\mu_X = \sum_{k=0}^{\infty} k p^k (1-p) = \frac{p}{1-p}$$

la varianza vale invece:

$$\sigma_X^2 = \sum_{k=0}^{\infty} \left(k - \frac{p}{1-p} \right)^2 p^k (1-p) = \frac{p}{(1-p)^2}$$

sebbene la determinazione attraverso la formula riportata risulti alquanto difficoltosa.

2.7.6. Variabile aleatoria binomiale negativa e ipergeometrica. La variabile aleatoria binomiale negativa o di Pascal conta il numero di successi che si devono collezionare in una serie di prove ripetute ed indipendenti di Bernoulli prima di osservare un numero di insuccessi complessivamente pari ad m , con m intero positivo, zero compreso:

$$P(X = n) = \binom{n+m-1}{m-1} p^n (1-p)^{m-1} (1-p)$$

Il valore medio è pari a: $\mu_X = m \frac{p}{1-p}$.

Infine la variabile aleatoria ipergeometrica si introduce in una particolare classe di esperimenti detti senza rimessa (o senza rimescolamento). Si supponga, per rendere

chiara l'idea con un esempio, di avere un lotto di N oggetti di cui D difettosi. Si supponga ora di pescare da questo lotto un numero di oggetti n senza rimessa (cioè senza rimetterli dentro dopo aver osservato di quale oggetto si tratti). Detti k gli oggetti difettosi tra gli n pescati, la v.a. ipergeometrica permette di valutare la probabilità di k (numero compreso tra $0, \dots, n$):

$$P(X = k) = \frac{\binom{D}{k} \binom{N - D}{n - k}}{\binom{N}{n}}$$

2.7.7. Derivazione e significato delle v.a. esponenziale e di Poisson. La v.a. esponenziale e quella di Poisson sono legate allo stesso significato fisico che è quello dell'attesa di un evento. In un processo di Poisson la casualità è affidata al tempo di arrivo di un certo evento. In generale nei processi di Poisson siamo interessati da vari fenomeni:

- (1) osservare il numero di eventi in un certo intervallo di tempo fissato;
- (2) il tempo di interarrivo, cioè il tempo che intercorre tra l'arrivo di due eventi successivi;
- (3) il tempo di attesa, cioè il tempo che occorre affinché arrivi il primo evento a partire da un istante iniziale di osservazione.

I tre tipi di fenomeni sono riassunti nella figura (2.7.4), dove le crocette rappresentano gli arrivi di un certo evento sull'asse temporale.



FIGURA 2.7.4. Rappresentazione grafica dei tre fenomeni descritti

Per poter ricavare la distribuzione di un processo poissoniano si fanno alcune ipotesi semplificative:

- (1) fissato un intervallo T e suddividendo questo intervallo in n (con n grande) intervallini piccoli di durata δT , $T = n \cdot \delta T$, la probabilità che un evento

capiti in un intervallino è pari ad una v.a. di Bernoulli:

$$\begin{cases} P(N(\delta T) = 1) = p \\ P(N(\delta T) = 0) = 1 - p \end{cases}$$

si esclude la probabilità che in un singolo intervallino capiti più di un evento

(2) Gli arrivi in intervallini diversi sono indipendenti tra loro.

Calcoliamo ora qual è la probabilità che in un dato intervallo finito T capitino k eventi: $P_n(N(T) = k)$. In base alle formule viste per la v.a. di Bernoulli si ha:

$P(N(T) = k) = \binom{n}{k} p^k (1-p)^{n-k}$ con n numero totale di intervallini in cui si può pensare suddiviso l'intervallo T . Sia ora Λ un parametro costante, tale che si possa scrivere: $\Lambda T = np = \alpha$, così che, quando il numero di intervallini tende ad infinito, la probabilità che un evento capiti in un dato intervallino vada a zero: $n \rightarrow \infty \Rightarrow p \rightarrow 0$. La probabilità diventa allora:

$$\begin{aligned} P(N(T) = k) &= \lim_{n \rightarrow \infty} P_n(N(T) = k) = \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \\ &= \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k} = \frac{\alpha^k}{k!} \cdot \lim_{n \rightarrow \infty} \frac{n \cdot (n-1) \cdot \dots \cdot (n-k)!}{n^k \cdot (n-k)!} \cdot \left(1 - \frac{\alpha}{n}\right)^n \cdot \left(1 - \frac{\alpha}{n}\right)^{-k} = \\ (2.7.17) \quad &= \frac{\alpha^k}{k!} \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\alpha}{n}\right)^n = \frac{\alpha^k}{k!} \cdot \exp(-\alpha) \end{aligned}$$

Si osservi che se si pone $T = 1$ allora la $P(N(1) = k)$ coincide con la distribuzione di Poisson trovata nel par. 2.7.3, che a questo punto rappresenta la probabilità che nell'unità di tempo capitino k eventi. La probabilità che nell'unità di tempo non capitino affatto eventi vale: $P(N(1) = 0) = \exp(-\Lambda)$.

Calcoliamo ora il *tempo di attesa*, cioè il tempo che bisogna attendere affinché capiti il primo evento a partire da un istante iniziale di osservazione. Se è x l'istante in cui si vuole valutare la v.a., distribuzione di probabilità della v.a. tempo di attesa può essere espressa anche come: $F_\tau(x) = P(\tau \leq x) = 1 - P(\tau > x)$. Ma $P(\tau > x)$ è anche la probabilità che sino ad x non sia capitato alcun evento: $P(\tau > x) = \exp(-\Lambda x)$. Quindi:

$$(2.7.18) \quad \begin{aligned} F_\tau(x) &= 1 - e^{-\Lambda x} \\ f_\tau(x) &= \Lambda e^{-\Lambda x} \end{aligned}$$

che, confrontata con le (2.7.6) e (2.7.7) dà significato alla v.a. esponenziale, purchè si ponga: $\Lambda = \frac{1}{\eta}$.

Si supponga ora che, a partire da un certo istante in cui è capitato un evento, si voglia determinare quale sarà la probabilità che sia τ il tempo di arrivo dell'evento successivo. Questa probabilità di arrivo, detta *tempo di interarrivo* si può calcolare facilmente a partire dalle considerazioni fatte precedentemente. Infatti, poichè gli eventi sono indipendenti tra loro, l'occorrere di un evento ad un certo istante (quello nel quale noi poniamo $t = 0$) non genera alcuna dipendenza futura sull'evento successivo. Ne consegue che la distribuzione e la densità di probabilità del tempo di interarrivo sono uguali a quelle calcolate per il tempo di attesa. La variabile aleatoria esponenziale esprime cioè la *mancaza di memoria* di un sistema.

2.7.8. Variabile aleatoria gaussiana. La variabile aleatoria di Gauss detta anche v.a. normale, o a campana, *emerge nell'esperienza dell'umanità come una delle più ampie generalizzazioni della filosofia naturale. Essa serve come strumento guida in ricerche della scienza, della medicina e dell'ingegneria. E' uno strumento indispensabile per l'analisi e l'interpretazione dei dati fondamentali ottenuti dall'osservazione e dall'esperimento.*²

Moltissimi fenomeni naturali si modellano statisticamente, in mancanza di altre informazioni, come se seguissero una variabile aleatoria gaussiana. Inoltre, come verrà dimostrato più avanti con il teorema del limite centrale, la v.a. gaussiana si può sempre considerare una generalizzazione di altre v.a. quando il numero di elementi presenti diventa molto grande.

La densità di probabilità della v.a. gaussiana è:

$$(2.7.19) \quad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

dove, come si può dimostrare, i parametri μ e σ^2 sono rispettivamente il valor medio e la varianza della v.a. La densità di probabilità gaussiana si estende su tutto l'asse dei numeri reali, ed è simmetrica rispetto al suo valor medio μ . La v.a. gaussiana è indicata anche con $\aleph(\mu, \sigma^2)$, dato che la media e la varianza sono sufficienti per caratterizzarla completamente. La gaussiana standard è quella con densità di probabilità $\aleph(0, 1)$, cioè:

$$(2.7.20) \quad f_{X_N}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

²J. Gleick: "Caos: la nascita di una nuova scienza", ed. Bur.

Essa è particolarmente importante poichè si può facilmente vedere che una gaussiana qualunque $\mathcal{N}(\mu, \sigma^2)$ può essere ottenuta come trasformazione lineare della gaussiana standard: $X = \sigma \cdot X_N + \mu$. Infatti:

$$f_X(x) = \frac{1}{\sigma} \cdot f_{X_N}\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

La funzione di distribuzione della gaussiana non può essere espressa in forma chiusa. A tale proposito si introduce la funzione di distribuzione della gaussiana standard:

$$(2.7.21) \quad \Phi_{X_N}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

Questa funzione è calcolata con metodi numerici e spesso si danno anche valori tabulati. Talvolta si usa anche la funzione $Q(x) = 1 - \Phi(x)$. Nota la funzione di distribuzione standard è possibile calcolare la funzione di distribuzione per una normale qualunque $\mathcal{N}(\mu, \sigma^2)$: $\Phi_X(x) = P(X \leq x) = P(\sigma X_N + \mu \leq x) = \Phi_{X_N}\left(\frac{x - \mu}{\sigma}\right)$. Quindi, ad esempio, se si vuole conoscere la probabilità che la variabile gaussiana assuma valori in un intervallo $[a, b]$, si ottiene:

$$(2.7.22) \quad P(a < x \leq b) = F_X(b) - F_X(a) = \Phi_N\left(\frac{b - \mu}{\sigma}\right) - \Phi_N\left(\frac{a - \mu}{\sigma}\right)$$

Molte volte nei calcolatori si ha a disposizione, direttamente implementata, la funzione di distribuzione standard. Quando questa non è presente, si hanno le funzioni errore ed errore complementare (*error function* e *complementary error function*):

$$(2.7.23) \quad \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz$$

$$(2.7.24) \quad \operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-z^2} dz$$

Quando si hanno a disposizione solo la funzione errore o la sua complementare si può ricavare la funzione di distribuzione standard da quest'ultima: $\Phi(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$, e la funzione $Q(x) = \frac{1}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)$. Da questa relazione si può ricavare facilmente la

probabilità che una gaussiana assuma valori nell'intervallo $[a, b]$: $P(a < x \leq b) = \frac{1}{2}[\operatorname{erf}(\frac{b-\mu}{\sqrt{2}\sigma}) - \operatorname{erf}(\frac{a-\mu}{\sqrt{2}\sigma})] = \frac{1}{2}[\operatorname{erfc}(\frac{a-\mu}{\sqrt{2}\sigma}) - \operatorname{erfc}(\frac{b-\mu}{\sqrt{2}\sigma})]$. Nelle figura (2.7.5) sono riportate la densità di probabilità gaussiana con la funzione di distribuzione e la $Q(x)$, in figura (2.7.6) è riportata invece la funzione errore e la sua complementare.

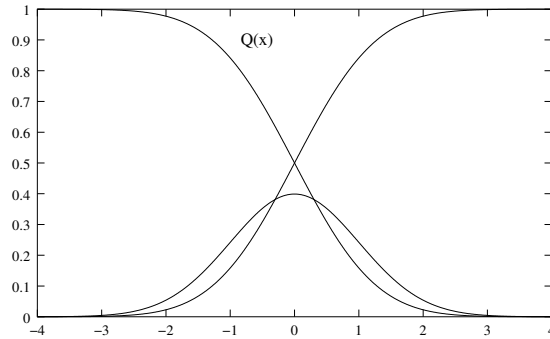


FIGURA 2.7.5. Densità, distribuzione e funzione $Q(x)$ per la v.a. gaussiana

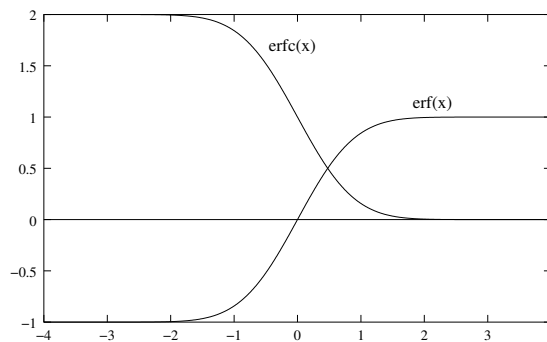


FIGURA 2.7.6. Funzione errore e funzione errore complementare

2.8. Variabili Aleatorie Condizionate

La funzione di distribuzione della probabilità, $F_X(x)$ passa attraverso la definizione di un evento, di cui la funzione ne rappresenta la probabilità: $F_X(x) = P(X \leq x) = P(A)$, dove l'evento A è l'evento che la v.a. assuma valori minori od uguali ad X . Il verificarsi di un evento però può essere anche influenzato dal verificarsi o meno di un altro evento B avente probabilità non nulla di accadere, $P(B)$. Ha quindi senso porsi il problema del calcolo di una funzione di distribuzione condizionata dall'occorrere dell'evento B . Tale funzione di distribuzione della v.a. X , indicata con $F_{X/B}(x/B)$,

vale ovviamente:

$$(2.8.1) \quad F_{X/B}(x/B) = \frac{P(A, B)}{P(B)} = \frac{P(X \leq x, B)}{P(B)}$$

da cui si può definire anche la densità di probabilità:

$$(2.8.2) \quad f_{X/B}(x/B) = \frac{dF_{X/B}(x/B)}{dx}$$

Le funzioni di distribuzione e di densità di probabilità godono di tutte le proprietà viste finora e valide per le funzioni e distribuzioni non condizionate.

2.9. Applicazioni notevoli

2.9.1. Trasformazione di una variabile aleatoria. Schematizzazione del guasto di un circuito elettrico. Si supponga di avere il semplice circuito elettrico riportato in figura (2.9.1). Il generatore di tensione sia collegato alla serie RC all'istante $t = 0$. Il resistore R abbia un tempo di guasto aleatorio X , in corrispondenza del quale esso interrompe il circuito. Questo tipo di fenomeno, cioè l'istante in cui interrompe il circuito, si può modellare (per quanto detto in par. 2.7.7) come una v.a. esponenziale con parametro (scelto arbitrariamente) pari a $2\alpha = 2RC$:

$$(2.9.1) \quad f_X(x) = \frac{1}{2\alpha} \exp\left(-\frac{x}{2\alpha}\right) \cdot u(x)$$

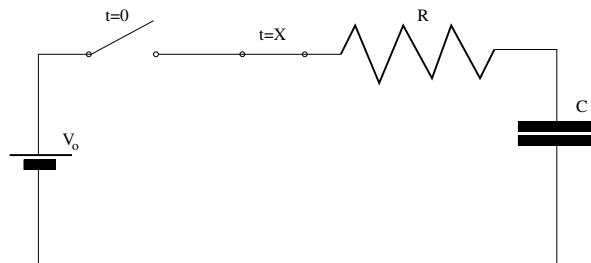


FIGURA 2.9.1. Schema del circuito RC con un guasto in $t = X$.

Si vuole determinare la densità di probabilità $f_V(v)$ della v.a. V che rappresenta la tensione ai capi del condensatore dopo che è avvenuto il guasto al resistore R . Il guasto al resistore si può schematizzare come l'interruzione del circuito e il conseguente mantenimento della tensione sul condensatore (qui supposto ideale). Poichè non si conosce l'istante in cui il guasto avverrà, anche la tensione che verrà mantenuta ai capi del condensatore è una quantità statistica, cioè ignota a priori, di cui però è possibile determinare la probabilità che assuma un certo valore. E' sufficiente a tale proposito determinare la legge che lega il tempo alla tensione ai capi del condensatore: $v(t) = V_o[1 - \exp(-t/\alpha)] \cdot u(t)$. Ponendo $t = X$, segue: $v(X) = V_o[1 - \exp(-X/\alpha)] \cdot u(X)$. Conosciamo quindi la legge di trasformazione e la densità di probabilità di X . Si deve quindi applicare quanto riportato nel par. 2.5:

$$(2.9.2) \quad f_V(v) = \frac{f_X(x)}{v'(x)}$$

dove x è la quantità che soddisfa l'equazione $v = v(x)$. Poichè la legge $v(t)$ è perfettamente invertibile nell'intervallo $[0, V_o]$, solo in questo intervallo avrà senso definire la densità di probabilità di $f_V(v)$. L'inversione della legge porta a:

$$(2.9.3) \quad v = v(x) \Rightarrow x = -\alpha \ln\left(1 - \frac{v}{V_o}\right)$$

poichè inoltre:

$$(2.9.4) \quad v'(x) = \frac{V_o}{\alpha} \exp(-t/\alpha)$$

si ha infine:

$$(2.9.5) \quad f_V(v) = \frac{1}{2V_o} \cdot \frac{1}{\sqrt{1 - \frac{v}{V_o}}}$$

2.9.2. Tempo di guasto dopo il rodaggio. Un altro problema interessante è quello del tempo di guasto dopo il rodaggio. Si abbia una serie di resistenze, tutte nominalmente uguali tra loro. Se queste resistenze si pongono sotto tensione, presto o tardi esse tenderanno a rompersi. La rottura di una singola resistenza è ovviamente un evento casuale, che è ben modellato da una variabile aleatoria esponenziale, con densità di probabilità data dalla (2.7.6). Il parametro η , che nella densità di probabilità esponenziale rappresenta il valor medio, è detto *tempo medio di guasto* o MTTF (Mean Time To Failure).

Effettuiamo ora un'operazione di rodaggio. Dato cioè un tempo prefissato a piacere, t_o , scartiamo le resistenze che si sono guastate sino a quell'istante. Quindi cominciamo, per istanti $t \geq t_o$, ad osservare le resistenze che non si sono ancora guastate. In base alla proprietà di mancanza di memoria della variabile aleatoria esponenziale, ci si aspetta che la densità di probabilità condizionata da questo evento non sia mutata. Verifichiamolo. Quello che vogliamo determinare è la densità di probabilità condizionata dall'evento B , con $B = \{t \geq t_o\}$.

Si calcola prima la distribuzione di probabilità $F_{X/B}(x/B)$. La probabilità dell'evento B è: $P(B) = P(X \geq t_o) = 1 - P(X < t_o) = 1 - F_X(t_o)$, dove $F_X(x)$ è la funzione di distribuzione della v.a. X . La probabilità congiunta dell'evento $P(X \leq x, B)$ si può determinare invece a partire dai due casi in cui $x > t_o$ oppure $x \leq t_o$:

$$\begin{aligned} P(X \leq x, B) &= P(X \leq x, X \geq t_o) = \begin{cases} F_X(x) - F_X(t_o) & x > t_o \\ 0 & \text{altrimenti} \end{cases} = \\ (2.9.6) \quad &= [F_X(x) - F_X(t_o)] \cdot u(x - t_o) \end{aligned}$$

Sostituendo nella definizione di distribuzione di probabilità condizionata da un evento:

$$\begin{aligned} F_{X/B}(x/B) &= \frac{P(X \leq x, B)}{P(B)} = \frac{[F_X(x) - F_X(t_o)] \cdot u(x - t_o)}{1 - F_X(t_o)} = \\ (2.9.7) \quad &= \frac{[F_X(x) - F_X(t_o)]}{1 - F_X(t_o)} \cdot u(x - t_o) \end{aligned}$$

da cui si ricava facilmente la densità di probabilità condizionata:

$$(2.9.8) \quad f_{X/B}(x/B) = \frac{dF_{X/B}(x/B)}{dx} = \frac{f_X(x)}{1 - F_X(t_o)} \cdot u(x - t_o)$$

Questa densità di probabilità spiega il comportamento delle resistenze quando si introduce il tempo di rodaggio: la probabilità che se ne guasti qualcuna per $x < t_o$ è ovviamente nulla, dato che si stanno considerando solo le resistenze sopravvissute all'istante $t = t_o$; inoltre la densità di probabilità è la stessa del caso in cui si cominci ad osservare il fenomeno per $t = 0$ (e quindi è verificato che il sistema è privo di memoria), tranne per il fattore di scala $\frac{1}{1-F_X(t_o)}$ che ha lo scopo di rinormalizzare la densità di probabilità in modo che la sua area sia sempre pari ad 1.

2.9.3. Generatori aleatori. Nei problemi di simulazione capita talvolta di richiedere, ai computer, di produrre dei numeri casuali, generati con una legge assegnata. La routine di sistema di un computer, basata sulle complesse relazioni esistenti tra i registri della macchina e il clock, è in grado spesso di fornire un numero casuale, ad aritmetica finita, compreso tra 0 ed 1 e distribuito in modo uniforme.

Il primo problema da risolvere per produrre numeri a caso con distribuzione assegnata, consiste nel costruire una funzione ϕ tale che se X è uniforme nell'intervallo $[0, 1]$, allora $\phi(X)$ abbia la distribuzione assegnata nell'intervallo assegnato. Il problema si formalizza così: data una v.a. X uniforme in $[0, 1]$, ed assegnata una densità di probabilità (continua) f , si deve trovare un'applicazione ϕ , tale che $Y = \phi(X)$ abbia densità di probabilità f .

Supponiamo che si voglia f non nulla all'interno di un intervallo assegnato $[a, b]$ e nulla al di fuori di esso. In tal caso la F , funzione cumulativa, sarà strettamente crescente e quindi invertibile in questo intervallo. Mostriamo che la scelta $\phi = F^{-1}$ risolve il nostro problema.

Anzitutto osserviamo che la F di una v.a. uniforme vale:

$$F(x) = x \quad 0 \leq x \leq 1$$

vale 0 per $x < 0$ e 1 per $x > 1$. Si ha allora che $\forall t, 0 \leq F(t) \leq 1$ e quindi che:

$$P(F^{-1}(X) \leq t) = P(X \leq F(t)) = F(t)$$

La v.a. $Y = \phi(X) = F^{-1}(X)$ risolve il problema, dato che avrà una funzione cumulativa pari ad F .

Supponiamo, ad esempio, di voler ottenere una legge esponenziale con parametro λ . Siccome la funzione cumulativa vale:

$$F(t) = 1 - \exp(-\lambda t), t \geq 0$$

essa è invertibile su \mathfrak{R}^+ e la sua inversa vale:

$$F^{-1}(x) = -\frac{1}{\lambda} \log(1-x)$$

Quindi se X è uniforme su $[0, 1]$, allora la funzione trasformata $Y = -\frac{1}{\lambda} \log(1-X)$ è esponenziale con parametro λ .

In altri casi esistono tecniche più raffinate (o più semplici, quando la funzione da invertire non è semplice) che, pur sfruttando il principio sopra esposto, permettono di aggirare le difficoltà del problema in esame.

2.10. Sistemi di Variabili Aleatorie

2.10.1. Sistemi di due variabili aleatorie. Nello studio di un esperimento aleatorio può avere senso associare due grandezze fisiche differenti a due risultati differenti dell'esperimento. Tuttavia le corrispondenti v.a. associate a queste grandezze, X ed Y , difficilmente forniranno risultati significativi all'esperimento stesso, se prese singolarmente.

Ad esempio si supponga di considerare un esperimento statistico in cui si misura l'altezza e il peso di una certa popolazione di persone. Sarà molto difficile trovare una persona molto alta e molto magra, sebbene la variabilità di peso e di altezza, prese singolarmente permettono escursioni ampie. Questo significa che, nell'esperimento aleatorio, le due grandezze forniscono informazione utile solo se prese insieme (informazione *congiunta*).

Data allora una coppia di variabili aleatorie (X, Y) si definisce la *funzione di distribuzione di probabilità congiunta*:

$$(2.10.1) \quad F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

che descrive in modo completo il comportamento statistico delle due v.a. In particolare, conoscendo $F_{XY}(x, y)$ è possibile avere informazioni sul comportamento statistico delle due v.a. prese separatamente (*probabilità marginali*). Le proprietà della funzione di distribuzione di probabilità congiunta sono molto simili a quelle viste per la funzione di distribuzione di una sola variabile:

- (1) la funzione $F_{XY}(x, y)$ assume valori compresi tra 0 ed 1;
- (2) Dato un valore fisso di y , $y = y_0$, la funzione $F_{XY}(x, y_0)$ è monotona non decrescente in x e continua da destra; analoga proprietà vale per l'altra variabile;
- (3) la funzione soddisfa le seguenti uguaglianze:

$$F_{XY}(-\infty, y) = P(X \leq -\infty, Y \leq y) = 0$$

$$F_{XY}(x, -\infty) = P(X \leq x, Y \leq -\infty) = 0$$

$$F_{XY}(-\infty, -\infty) = P(X \leq -\infty, Y \leq -\infty) = 0$$

(4) Le funzioni di distribuzione marginale si ricavano come:

$$F_X(x) = F_{XY}(x, +\infty)$$

$$F_Y(y) = F_{XY}(+\infty, y)$$

(5) $\lim_{x,y \rightarrow \infty} F_{XY}(x, y) = 1$

(6) La probabilità dell'evento rettangolare $R = \{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\}$ si calcola con la relazione:

(2.10.2)

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$$

In particolare l'ultima proprietà permette di determinare la funzione densità di probabilità congiunta quando l'ampiezza degli intervalli in x ed in y tende a diventare molto piccola:

$$P(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y) =$$

$$[F_{XY}(x + \Delta x, y + \Delta y) - F_{XY}(x, y + \Delta y)] - [F_{XY}(x + \Delta x, y) - F_{XY}(x, y)] =$$

$$= \frac{\partial F_{XY}(x, y + \Delta y)}{\partial x} \Delta x - \frac{\partial F_{XY}(x, y)}{\partial x} \Delta x = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \Delta x \Delta y$$

Se allora si definisce la funzione: $f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$, si ha:

$$(2.10.3) \quad P(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y) \cong f_{XY}(x, y) \cdot \Delta x \Delta y$$

La funzione definita è detta **densità di probabilità congiunta**. Essa è sempre non negativa ed integra ad 1 su tutto il piano:

$$(2.10.4) \quad \int \int_{\mathbb{R}^2} f_{XY}(x, y) dx dy = 1$$

Le densità di probabilità marginali si ricavano in base a quanto già visto per la distribuzione di probabilità congiunta:

$$(2.10.5) \quad f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$$

$$(2.10.6) \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

Dato un evento A , identificabile con un dominio sul piano $\mathfrak{R}^2 : (x, y) \in A$, la probabilità di quell'evento è:

$$(2.10.7) \quad \int \int_A f_{XY}(x, y) dx dy$$

Infine la funzione di distribuzione può essere ricavata dalla densità di probabilità mediante la relazione:

$$(2.10.8) \quad F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s, t) ds dt$$

Anche in questo caso si può definire la funzione di distribuzione e la densità di probabilità condizionata. Si supponga infatti che la v.a. X abbia assunto un certo valore $X = x$. In tal caso la probabilità marginale di Y sarà condizionata da questo cambiamento. Si definisce allora la funzione di distribuzione condizionata:

$$(2.10.9) \quad F_{Y/X}(y/x) = \frac{\int_{-\infty}^y f_{XY}(x, t) dt}{f_X(x)}$$

Da questa si ricava poi la densità di probabilità condizionata derivando rispetto ad y :

$$(2.10.10) \quad f_{Y/X}(y/x) = \frac{\partial F_{Y/X}(y/x)}{\partial y} = \frac{f_{XY}(x, y)}{f_X(x)}$$

Se il comportamento della variabile aleatoria Y è uguale sia sotto condizionamento per $X = x, \forall x$ sia senza condizionamento, cioè se $f_Y(y) = f_{Y/X}(y/x)$ allora questo significa che le v.a. sono indipendenti tra loro. In questo caso la densità di probabilità congiunta è pari al prodotto delle due densità di probabilità: $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$.

Come già visto nel caso di una sola v.a., anche nel caso di due v.a. si può effettuare una trasformazione: $Z = g(X, Y)$, dove $g(\bullet, \bullet)$ è una funzione reale di due variabili reali. La funzione definisce una nuova v.a. con funzione di distribuzione: $F_Z(z) = P(g(X, Y) \leq z)$. Il calcolo della $F_Z(z)$ può essere facilmente effettuato tramite:

$$(2.10.11) \quad F_Z(z) = \int \int_{R(Z)} f_{XY}(x, y) dx dy$$

dove il dominio $R(Z)$ indica la regione di piano in cui vale la relazione $g(X, Y) \leq z$. Nota la funzione di distribuzione, la densità di probabilità si ricava mediante semplice derivazione rispetto all'unica variabile z .

EXAMPLE 2.10.1. Somma di due variabili aleatorie. Questo esempio dà luogo ad un risultato notevole. Detta infatti D la regione di piano individuata dalla relazione $x + y \leq z$, si ha che $D = \{x, y \leq z - x\}, \forall x$. Quindi si ha:

$$F_Z(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_{XY}(x, y) dx dy = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-x} f_{XY}(x, y) dy \right] dx$$

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{+\infty} \frac{d}{dz} \left[\int_{-\infty}^{z-x} f_{XY}(x, y) dy \right] dx = \int_{-\infty}^{+\infty} f_{XY}(x, z-x) dx$$

Se poi le v.a. sono indipendenti si ottiene:

$$(2.10.12) \quad f_Z(z) = \int_{-\infty}^{+\infty} f_{XY}(x, z-x) dx = \int_{-\infty}^{+\infty} f_X(x) \cdot f_Y(z-x) dx = f_X(z) \star f_Y(z)$$

cioè date due v.a. sommate tra loro ed indipendenti, la densità di probabilità della v.a. somma è pari alla convoluzione delle densità di probabilità delle due v.a. di partenza.

Il risultato è facilmente generalizzabile alla somma di n variabili aleatorie indipendenti tra loro.

2.10.2. Correlazione e covarianza. Il comportamento statistico di una coppia di v.a. può essere descritto da alcuni parametri che le descrivono in modo congiunto. Tra questi parametri vi sono la *correlazione*:

$$(2.10.13) \quad r_{XY} = E[XY] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{XY}(x, y) dx dy$$

e la *covarianza*:

$$(2.10.14) \quad c_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) f_{XY}(x, y) dx dy$$

Si fa vedere facilmente che $c_{XY} = r_{XY} - \mu_X \mu_Y$.

La covarianza è un parametro statistico molto importante. Essa stabilisce se esiste un qualche tipo di *dipendenza lineare* tra le v.a. Cerca in ogni caso di misurare una dispersione congiunta intorno ai rispettivi valori medi. Se ad esempio la covarianza è positiva questo significa che, prevalentemente, le v.a. tendono a muoversi nella stessa direzione, cioè è più probabile che se una di esse è sopra la media lo sia anche l'altra (come ad esempio peso ed altezza di una persona). Una covarianza negativa indica invece il fenomeno contrario, cioè che prevalentemente le due v.a. si muovono statisticamente in direzioni opposte, come ad esempio età ed acuità visiva di una popolazione.

Supponiamo che tra le due v.a. X e Y esista una certa dipendenza lineare:

$$(2.10.15) \quad \begin{aligned} Y &= aX + b \\ \mu_Y &= a\mu_X + b \\ Y - \mu_Y &= a(X - \mu_X) \end{aligned}$$

La covarianza vale allora:

$$\begin{aligned} c_{XY} &= E[(X - \mu_X)(Y - \mu_Y)] = E[(X - \mu_X)a(X - \mu_X)] = a\sigma_X^2 \\ c_{XY} &= E[(X - \mu_X)(Y - \mu_Y)] = E[(Y - \mu_Y)(Y - \mu_Y)/a] = \sigma_Y^2/a \end{aligned}$$

da cui si ricava che:

$$(2.10.16) \quad c_{XY}^2 = \sigma_X^2 \sigma_Y^2$$

In questo caso il rapporto $\frac{c_{XY}}{\sigma_X \sigma_Y} = \pm 1$.

E' possibile generalizzare il discorso precedente e porre, per una data coppia di v.a. X e Y , la seguente definizione:

$$(2.10.17) \quad \rho = E \left[\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y} \right] = \frac{c_{XY}}{\sigma_X \sigma_Y}$$

che dà una misura del grado di correlazione o di dipendenza lineare tra le due v.a.

Si ponga ora il caso generale: $Y = aX + b + Z$, in cui la quantità Z è una v.a. Con questo modello abbiamo supposto che la dipendenza lineare tra X e Y sia dubbia o comunque non nota. Il problema che ci poniamo è quello di determinare la retta migliore possibile (cioè i coefficienti a e b) che permettano di formulare la migliore predizione lineare di Y in funzione di X .

La soluzione considerata ottima è quella che si ottiene imponendo che la media di Z sia nulla e che la sua varianza sia minima:

$$(2.10.18) \quad \begin{aligned} \mu_Z &= \mu_Y - a\mu_X - b = 0 \\ \sigma_Z^2 &= \sigma_Y^2 + a^2\sigma_X^2 - 2ac_{XY} = \min \end{aligned}$$

la seconda equazione va derivata e posta = 0:

$$(2.10.19) \quad \frac{\partial \sigma_Z^2}{\partial a} = 2a\sigma_X^2 - 2c_{XY} = 0$$

da cui si ricava abbastanza facilmente:

$$(2.10.20) \quad a = \frac{c_{XY}}{\sigma_X^2}$$

Sostituendo il valore determinato di a nella seconda equazione della (2.10.18) si ricava la varianza minima che deve assumere la v.a. Z :

$$(2.10.21) \quad \sigma_{Z_{\min}}^2 = \sigma_Y^2 + \frac{c_{XY}^2}{\sigma_X^4} \sigma_X^2 - 2 \frac{c_{XY}}{\sigma_X^2} c_{XY} = \sigma_Y^2 (1 - \rho^2)$$

Dalla precedente equazione si ricavano le seguenti osservazioni:

- (1) il valore del coefficiente di correlazione ρ è un numero in valore assoluto sempre minore di 1: $0 \leq |\rho| \leq 1$;
- (2) Tanto più $|\rho| \rightarrow 1$ tanto più sono linearmente dipendenti le v.a. X e Y . Se $|\rho| = 1$, X e Y dipendono linearmente tra loro.
- (3) Se $|\rho| = 0$ allora $c_{XY} = 0$, cioè le due v.a. sono incorrelate.

Il coefficiente di correlazione serve a normalizzare la covarianza che altrimenti potrebbe assumere valori anche molto differenti per diverse coppie di v.a. Esso permette quindi di confrontare diversi risultati tra loro, dato che l'intervallo in cui è definito è sempre $[-1, 1]$. Il coefficiente di correlazione si può anche vedere come una correlazione definita per la nuova coppia di v.a. normalizzate $(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y})$. Quanto più vicino è questo coefficiente ad 1 in modulo, tanto più le v.a. tendono a seguire una legge di variazione lineare.

Se poi le v.a. sono indipendenti tra loro la loro correlazione vale:

$$\begin{aligned}
 r_{XY} &= E[XY] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{XY}(x, y) dx dy = \\
 (2.10.22) \quad &= \int_{-\infty}^{+\infty} x f_X(x) dx \cdot \int_{-\infty}^{+\infty} y f_Y(y) dy = \mu_X \mu_Y
 \end{aligned}$$

cioè due v.a. indipendenti sono anche incorrelate. L'inverso di questa relazione non è sempre vera tuttavia: cioè due v.a. incorrelate possono anche essere dipendenti tra loro. L'indipendenza è una condizione *più restrittiva* della incorrelazione.

2.10.3. Metodo dei minimi quadrati. Questo metodo, strettamente correlato con il concetto di dipendenza lineare tra due v.a. è in realtà oggetto di studio della statistica e non della teoria delle probabilità. Infatti il metodo si introduce in un contesto in cui l'eventuale dipendenza lineare tra due v.a. è ignota ma si suppone esistente per ipotesi di lavoro; si suppone inoltre che le statistiche delle due v.a. non siano note. Si suppongono invece note una serie di misure delle due v.a., che in un esperimento aleatorio, si possono considerare due grandezze fisiche in qualche modo dipendenti tra loro (per esempio si potrebbe pensare ad un esperimento aleatorio che coinvolga spazio percorso da un oggetto che si muove di moto rettilineo uniforme e tempo trascorso. E' ragionevole supporre dipendenza lineare tra le due grandezze fisiche).

Siano x_i e y_i , con $i = 1, 2, \dots, N$ la serie di misure ottenute. La retta ottima consiste nel considerare la soluzione che minimizza la somma dei quadrati della relazione di dipendenza lineare:

$$\begin{aligned}
 z_i &= y_i - ax_i - b \\
 S &= \sum_i z_i^2 = \sum_i (y_i - ax_i - b)^2 \\
 \frac{\partial S}{\partial a} &= - \sum_i x_i y_i + a \sum_i x_i^2 + b \sum_i x_i = 0 \\
 (2.10.23) \quad \frac{\partial S}{\partial b} &= - \sum_i y_i + a \sum_i x_i + Nb = 0
 \end{aligned}$$

Le relazioni precedenti possono essere riarrangiate in un sistema di due equazioni in due incognite: a e b , dato che le restanti quantità sono note, essendo ricavabili dalle coppie di misure (x_i, y_i) .

Se si pongono le seguenti stime:

$$\begin{aligned}\widehat{\mu}_X &= \frac{1}{N} \sum_i x_i \\ \widehat{\mu}_Y &= \frac{1}{N} \sum_i y_i \\ \widehat{\sigma}_X^2 &= \frac{1}{N} \sum_i (x_i - \widehat{\mu}_X)^2 \\ \widehat{\sigma}_Y^2 &= \frac{1}{N} \sum_i (y_i - \widehat{\mu}_Y)^2 \\ \widehat{c}_{XY} &= \frac{1}{N} \sum_i (x_i - \widehat{\mu}_X)(y_i - \widehat{\mu}_Y)\end{aligned}$$

I valori ottimali di a e b nel senso dei minimi quadrati si possono riscrivere nel seguente modo:

$$(2.10.24) \quad \begin{aligned}a &= \frac{\widehat{c}_{XY}}{\widehat{\sigma}_X^2} \\ b &= \widehat{\mu}_Y - a\widehat{\mu}_X \\ \rho &= \frac{\widehat{c}_{XY}}{\widehat{\sigma}_X \widehat{\sigma}_Y}\end{aligned}$$

2.10.4. Sistemi di n variabili aleatorie. Quanto visto per due v.a. può essere facilmente generalizzato per n variabili aleatorie. La funzione di distribuzione di probabilità congiunta è definita come:

$$(2.10.25) \quad F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

e la relativa densità di probabilità è:

$$(2.10.26) \quad f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

Dalla densità di probabilità congiunta è possibile ricavare la densità di probabilità marginale rispetto a ciascuna delle variabili: è sufficiente integrare su tutto il dominio \mathfrak{R} per tutte le altre. Inoltre è possibile ricavare la densità di probabilità marginale di un qualunque sottoinsieme di v.a., sempre integrando in \mathfrak{R} rispetto a quelle che

devono mancare (se ad esempio si vuole la densità di probabilità marginale rispetto ad x_3, x_4, \dots, x_n si deve integrare $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ rispetto ad x_1, x_2).

In modo analogo si ricavano le densità di probabilità condizionate. Se si vuole determinare la densità di probabilità condizionata ad un qualunque sottoinsieme di v.a. è sufficiente dividere la densità di probabilità congiunta per la marginale ristretta a quel sottoinsieme. Se ad esempio si vuole determinare

$$(2.10.27) \quad f_{X_1, X_4, \dots, X_n / X_2, X_3}(x_1, x_4, \dots, x_n / x_2, x_3) = \frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{f_{X_2, X_3}(x_2, x_3)}$$

Le v.a. si dicono indipendenti tra loro se, preso un qualunque sottoinsieme di esse, condizionato da un qualunque altro sottoinsieme (distinto dal primo), la densità di probabilità condizionata è pari alla densità del primo sottoinsieme considerato non condizionato.

Per lo studio dei sistemi di v.a. si utilizza normalmente la notazione matriciale: $X = \{X_1, X_2, \dots, X_n\}$ dove X è un vettore aleatorio:

$$(2.10.28) \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1, X_2, \dots, X_n]^T$$

La funzione di distribuzione di probabilità congiunta e la funzione di densità di probabilità congiunta possono essere quindi indicate con notazione vettoriale: $F_X(X)$ ed $f_X(X)$. Anche i parametri statistici possono essere indicati con notazione vettoriale:

$$(2.10.29) \quad \mu_X = E[X] = [\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_n}]^T$$

Poichè la correlazione e la covarianza sono state definite per coppie di v.a. quando si hanno più di due v.a., ha senso definire la correlazione e la covarianza per tutte le possibili coppie di v.a. In tal caso tutte le correlazioni, come pure tutte le covarianze possono essere riunite in una matrice di dimensione $n \times n$ dette matrici di correlazione e di covarianza:

$$(2.10.30) \quad R_X = E[XX^T] = \begin{bmatrix} r_{X_1X_1} & r_{X_1X_2} & \cdots & r_{X_1X_n} \\ r_{X_2X_1} & r_{X_2X_2} & \cdots & r_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{X_nX_1} & r_{X_nX_2} & \cdots & r_{X_nX_n} \end{bmatrix}$$

$$(2.10.31) \quad C_X = E[(X - \mu_X)(X - \mu_X)^T] = \begin{bmatrix} c_{X_1X_1} & c_{X_1X_2} & \cdots & c_{X_1X_n} \\ c_{X_2X_1} & c_{X_2X_2} & \cdots & c_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{X_nX_1} & c_{X_nX_2} & \cdots & c_{X_nX_n} \end{bmatrix}$$

Le matrici di correlazione e di covarianza sono simmetriche, essendo

$$(2.10.32) \quad r_{X_iX_j} = r_{X_jX_i} \quad c_{X_iX_j} = c_{X_jX_i}$$

dalle loro stesse definizioni. I valori sulla diagonale di R_X sono i valori quadratici medi delle singole v.a. X_i : $r_{X_iX_i} = E[X_iX_i] = m_{X_i}^2$. I valori sulla diagonale della matrice di covarianza sono le varianze delle singole v.a. X_i :

$$(2.10.33) \quad c_{X_iX_i} = E[(X_i - \mu_{X_i})(X_i - \mu_{X_i})] = \sigma_{X_i}^2$$

La relazione tra la matrice di correlazione e quella di covarianza è pari alla relazione che esiste tra la correlazione e la covarianza per una coppia di v.a.: $C_X = R_X - \mu_X \mu_X^T$.

Anche per la trasformazione si possono fare considerazioni analoghe. Si consideri una funzione vettoriale di n variabili in n valori $g(\bullet, \bullet, \dots, \bullet) = g_1(\bullet), g_2(\bullet), \dots, g_n(\bullet)$, e si applichi tale funzione al vettore aleatorio X ottenendo un nuovo vettore aleatorio di n v.a.: $Y = g(X)$. Per determinare la densità di probabilità congiunta del nuovo vettore $f_Y(y)$ a partire da quella di X si può utilizzare la generalizzazione di quanto visto in par. 2.5:

$$(2.10.34) \quad f_Y(y) = \sum_i \frac{f_X(x_i)}{|\det(J(x_i))|}$$

dove x_i è il sottoinsieme di \mathbb{R}^n soluzione della relazione: $g(x_i) = y$ e dove $J(x_i)$ è la matrice jacobiana calcolata per tali valori.

Nell'ipotesi in cui il vettore aleatorio sia trasformato in un altro vettore di dimensioni differenti, m , è necessario passare prima attraverso il calcolo della funzione di distribuzione di probabilità congiunta, estesa al dominio che soddisfa la disuguaglianza data dalla trasformazione:

$$(2.10.35) \quad F_Z(z) = \int_{R_D} f_X(x) dx$$

dove R_D è l'insieme che soddisfa la relazione:

$$(2.10.36) \quad R_D = \left\{ (X_1, X_2, \dots, X_n) \ni g_1(X_1, X_2, \dots, X_n) \leq Z_1 \bigcap \bigcap g_2(X_1, X_2, \dots, X_n) \leq Z_2 \dots \bigcap g_m(X_1, X_2, \dots, X_n) \leq Z_m \right\}$$

Nota poi la funzione di distribuzione, si può determinare la funzione di densità di probabilità congiunta mediante derivazione.

Particolare interesse ha il caso in cui $m = 1$, cioè si voglia trasformare il vettore aleatorio in una sola variabile aleatoria. Facciamo il caso in cui la v.a. che si vuole ottenere è somma delle n v.a. di partenza: $Z = \sum_i X_i$.

Poichè si può scrivere: $Z = 1^T X$ con $1^T = [1 \ 1 \ \dots \ 1]$, si ha che il valore atteso vale:

$$(2.10.37) \quad \mu_Z = E[Z] = E[1^T X] = 1^T E[X] = 1^T \mu_X = \sum_i \mu_{X_i}$$

Per la varianza si ha:

$$(2.10.38) \quad \begin{aligned} \sigma_Z^2 &= E[(Z - \mu_Z)^2] = E[(Z - \mu_Z)^T (Z - \mu_Z)] = \\ &= E[(1^T X - 1^T \mu_X)^T (1^T X - 1^T \mu_X)] = E[(X - \mu_X)^T 11^T (X - \mu_X)] = \\ &= \sum_i \sum_j c_{x_i x_j} \end{aligned}$$

Se allora le v.a. componenti il vettore aleatorio X sono a due a due incorrelate (o addirittura indipendenti), cioè se $c_{x_i x_j} = 0 \forall i, j = 1, \dots, n$ ed $i \neq j$ allora la varianza della v.a. somma è pari alla somma delle varianze delle singole v.a. X_i .

2.10.5. Variabili aleatorie congiuntamente gaussiane. Particolare interesse assume la composizione di v.a. gaussiane. Si supponga di avere n v.a. gaussiane che costituiscono un vettore aleatorio $X = [X_1, X_2, \dots, X_n]^T$. Se le v.a. sono tutte indipendenti tra loro la densità di probabilità congiunta è pari al prodotto delle densità di probabilità di tutte le componenti del vettore:

$$(2.10.39) \quad f_X(x) = \prod_{i=1}^n f_{x_i}(x_i)$$

Inoltre, poichè si è supposta la gaussianità delle singole X_i , $\mathcal{N}(\mu_i, \sigma_i^2)$, la $f_X(x)$ diventa:

$$(2.10.40) \quad f_X(x) = \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right) = \frac{1}{\sqrt{(2\pi)^n \cdot \prod_i \sigma_i^2}} \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x - \mu_i)^2}{\sigma_i^2}\right)$$

dove si è posto, con notazione abbreviata, $\sigma_i^2 = \sigma_{X_i}^2$ e $\mu_i = \mu_{X_i}$. La densità di probabilità può essere riscritta sfruttando il vettore dei valori medi $\mu = [\mu_1, \mu_2, \dots, \mu_n]^T$ e la matrice di covarianza, che è peraltro diagonale data l'indipendenza delle n variabili:

$$(2.10.41) \quad C_X = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

$$(2.10.42) \quad \det C_X = \prod_{i=1}^n \sigma_i^2$$

La densità di probabilità congiunta diventa:

$$(2.10.43) \quad f_X(x) = \frac{1}{\sqrt{(2\pi)^n |\det C_X|}} \cdot \exp\left(-\frac{1}{2}(x - \mu_X)^T C_X^{-1}(x - \mu_X)\right)$$

La densità di probabilità congiunta, nel caso di n v.a. congiuntamente gaussiane e non indipendenti può essere ancora scritta come riportato nella equazione precedente, purchè si tenga in conto in fatto che, in tal caso, la matrice di covarianza non è più diagonale ma ha i valori tutti genericamente differenti da zero.

Un vettore gaussiano gode delle seguenti proprietà:

- (1) il suo comportamento statistico è univocamente determinato dal vettore dei valori medi μ e dalla matrice di covarianza C_X ;
- (2) se le v.a. gaussiane sono incorrelate a due a due allora la densità di probabilità congiunta si può esprimere come prodotto delle densità di probabilità marginali (poichè gli elementi fuori diagonale di C_X sono nulli). Cioè per le v.a. gaussiane la incorrelazione implica la indipendenza.
- (3) Un qualunque sottoinsieme di v.a. gaussiane è ancora un insieme di v.a. congiuntamente gaussiane.
- (4) Data una qualunque trasformazione di tipo lineare: $Y = aX + b$ il vettore aleatorio Y è ancora congiuntamente gaussiano, con vettore dei valori medi pari a $\mu_Y = a\mu_X + b$ e matrice di covarianza $C_Y = AC_X A^T$
- (5) Un qualunque sottogruppo di v.a. preso tra le n del vettore X , condizionato ad un qualunque altro sottogruppo (purchè formato da v.a. distinte da quelle considerate nel primo sottogruppo) è ancora congiuntamente gaussiano.

2.11. Convergenza ed approssimazione

Si è già accennato al fatto che la deviazione standard (radice quadrata della varianza) è in grado di dare una misura della dispersione di una v.a. attorno al suo valor medio. Valori di varianza grandi sono indice del fatto che c'è una significativa probabilità che valori casuali estratti dalla v.a. siano abbastanza lontani dal valor medio; viceversa per valori piccoli della varianza.

E' evidente, tuttavia, che la varianza non è in grado di dire quanto questa dispersione sia significativa, dato che differenti v.a. possono avere anche varianze uguali, pur essendo *disperse* attorno al valor medio in modi differenti.

Esiste tuttavia un teorema che è in grado di dare una misura quantitativa della dispersione in termini di probabilità e che utilizza proprio la varianza.

THEOREM 2.11.1. *Disuguaglianza di Chebyshev.*

Data una v.a. X , $\forall \eta > 0$ risulta che:

$$(2.11.1) \quad P(|X - E[X]| > \eta) \leq \frac{Var(X)}{\eta^2}$$

DIMOSTRAZIONE. Si consideri la v.a. Y che vale:

$$Y = \begin{cases} \eta^2 & , |X - E[X]| > \eta \\ 0 & , |X - E[X]| \leq \eta \end{cases}$$

E' allora chiaro che

$$(|X - E[X]|)^2 \geq Y$$

sempre, dato che se accade l'evento $|X - E[X]| > \eta$, si ha $Y = \eta^2 < (|X - E[X]|)^2$. Se invece accade l'evento $|X - E[X]| \leq \eta$, la v.a. Y vale 0, ma $|X - E[X]|$ è comunque un numero ≥ 0 .

Se ora si fa l'aspettazione di ambo i membri della relazione precedente si ha:

$$Var(X) = E[(|X - E[X]|)^2] \geq E[Y] = \eta^2 P(|X - E[X]| > \eta)$$

che dà il risultato cercato. □

La disuguaglianza di Chebyshev rende rigorosa l'interpretazione intuitiva di varianza come misura della dispersione: più $Var(X)$ è piccola più piccola è la probabilità che X prenda valori lontani dalla media.

Tuttavia la disuguaglianza di Chebyshev è spesso una maggiorazione grossolana della probabilità di $P(|X - E[X]| > \eta)$. Ad esempio si consideri la v.a. che assume i valori $-1, 1$ con probabilità rispettivamente di $1/2, 1/2$. Per questa v.a. la media è 0 e la varianza vale $Var(X) = 1$. Se si sceglie $\eta = 2$ si ha che $P(|X - E[X]| > \eta) = 0$ mentre $Var(X)/\eta^2 = 1/4$, ma se addirittura si prende un $\eta < 1$ si ha una maggiorazione con il valore $Var(X)/\eta^2 > 1$, cosa ovvia dato che una probabilità è sicuramente maggiorata da un numero maggiore di 1.

In molte circostanza tuttavia la disuguaglianza di Chebyshev si dimostra preziosa. E' infatti fondamentale per dimostrare e giustificare la cosiddetta **Legge dei grandi numeri**.

Partiamo prima con un esempio. Si supponga di lanciare n volte una moneta e sia k il numero di lanci in cui si ottiene testa. La quantità k/n è quindi la proporzione di teste ottenute in n lanci. Se la moneta è equilibrata l'intuizione suggerisce che tale proporzione non debba discostarsi troppo dal valore $1/2$. Tuttavia sarà difficile che la quantità k/n dia esattamente $1/2$, come anche è poco probabile (ma non impossibile) che il numero di teste sia molto piccolo (o addirittura nullo) o molto grande. Tuttavia empiricamente si può verificare che al crescere del numero di lanci, il fenomeno di *discostamento* dal valore $1/2$ dovrebbe sparire: cioè il numero di teste e croci tende a compensarsi sempre più man mano che cresce il valore di n . Formalizziamo allora quanto l'intuizione ci suggerisce. Il lancio di una moneta è rappresentabile da una v.a. di Bernoulli con $n = 1$ e $p = 1/2$; a tale v.a. facciamo assumere valore 1 quando si presenta una testa: $X_i = 1$, altrimenti 0. Il numero totale di teste ottenute negli n lanci

può essere dunque rappresentato dalla quantità

$$S_n = X_1 + X_2 + \dots + X_n$$

e la proporzione di teste negli n lanci dalla quantità

$$\bar{X}_n = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

Quanto osservato prima può essere quindi schematizzato dall'osservazione che, all'aumentare di n la quantità \bar{X}_n tende a discostarsi sempre meno da $1/2$. Quanto trovato corrisponde al vero, anzi tale risultato è formalizzato e generalizzato dalla cosiddetta Legge dei Grandi Numeri:

THEOREM 2.11.2. *Sia $(X_n)_n$ una successione di v.a. indipendenti ed aventi tutte la stessa legge, la stessa media μ e varianza σ^2 . Posto allora*

$$\bar{X}_n = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

si ha che, $\forall \eta > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \eta) = 0$$

DIMOSTRAZIONE. La v.a. \bar{X}_n ha anch'essa media μ :

$$E[\bar{X}_n] = \frac{1}{n} E[X_1 + X_2 + \dots + X_n] = \frac{1}{n} (\mu + \mu + \dots + \mu) = \mu$$

e varianza pari a:

$$\begin{aligned} \text{Var}(\bar{X}_n) &= \frac{1}{n^2} \text{Var}(X_1 + X_2 + \dots + X_n) = \\ &= \frac{1}{n^2} (\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)) = \frac{1}{n^2} \cdot n \cdot \text{Var}(X_1) = \frac{\sigma^2}{n} \end{aligned}$$

Ora, applicando la disuguaglianza di Chebyshev si ha la dimostrazione:

$$0 \leq P(|\bar{X}_n - \mu| > \eta) \leq \frac{\text{Var}(\bar{X}_n)}{\eta^2} = \frac{\sigma^2}{n\eta^2} \xrightarrow{n \rightarrow \infty} 0$$

□

Riprendiamo l'esempio introduttivo sul lancio della moneta. Supponiamo di non sapere a priori se la moneta sia equilibrata o no ($p = 1/2$). la legge dei grandi numeri fornisce uno strumento per stimare tale probabilità. Lanciamo la moneta n volte e stimiamo p tramite la quantità:

$$\frac{\text{\# teste in } n \text{ lanci}}{n}$$

Se infatti poniamo

$$X_i = \begin{cases} 1 & \text{lancio } i\text{-simo dà testa} \\ 0 & \text{altrimenti} \end{cases}$$

allora $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ e, per la Legge dei Grandi Numeri $\bar{X}_n \rightarrow p = E[X_i]$ per $n \rightarrow \infty$. Tuttavia, nella pratica, noi possiamo fare soltanto un numero finito di lanci e quindi occorre valutare l'errore che si commette stimando p con il valore di \bar{X}_n che verrà fuori da tale esperimento composto. Si può procedere allora in questo modo. Si fissi un numero $\eta > 0$ e si stimi la probabilità di commettere un errore nel valutare p maggiore di η . Si tratta di valutare quindi la quantità:

$$P(|\bar{X}_n - p| > \eta)$$

Naturalmente, siccome tale valutazione richiederebbe il calcolo della funzione di distribuzione (cumulativa) di una binomiale con n molto grande (quantità per la quale il calcolo è spesso lungo e non vi sono formule chiuse), è meglio limitarci a maggiorare quella probabilità con la disuguaglianza di Chebyshev:

$$P(|\bar{X}_n - p| > \eta) \leq \frac{Var(\bar{X}_n)}{\eta^2} = \frac{p(1-p)}{n} \cdot \frac{1}{\eta^2}$$

Questa disuguaglianza dipende ancora dalla incognita p (che è la quantità che vogliamo stimare), ma un semplice studio di funzione permette di stabilire che $p(1-p) \leq 1/4$, con $0 \leq p \leq 1$. Allora si ha:

$$P(|\bar{X}_n - p| > \eta) \leq \frac{1}{4n\eta^2}$$

Per $n = 100$ la probabilità che p disti da \bar{X}_n più di 0.1 è una quantità minore di 0.25. Tale valutazione, come si può vedere, è spesso grossolana, soprattutto per esperimenti semplici e per un numero n di prove piccolo. Esiste tuttavia un Teorema che permette di migliorare tale stima, ed è il **Teorema del Limite Centrale**, dovuto al matematico russo Lyapunov. Questo teorema vale sotto condizioni non particolarmente restrittive, sebbene la sua dimostrazione risulti difficoltosa nel caso più generale.

Si considerino n v.a. X_i indipendenti tra loro e tutte dotate della stessa densità di probabilità $f_{X_i}(x) = f_X(x)$ e quindi con stesso valor medio μ e stessa varianza σ^2 . Sappiamo che, se si considera la somma delle v.a. $S_n = \sum_i X_i$ questa avrà media pari alla somma dei valori medi e varianza pari alla somma delle varianze: $\mu_n = n \cdot \mu$ e $\sigma_n^2 = n \cdot \sigma^2$. Ovviamente, al crescere di n , sia il valor medio, sia la varianza tendono a divergere. Si può considerare in tal caso una v.a. normalizzata (nello stesso modo con cui si fa per la gaussiana):

$$(2.11.2) \quad Z_n = \frac{S_n - \mu_n}{\sigma_n} = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}$$

che, per qualunque valore di n , ha sempre valor medio nullo e varianza pari ad 1.

THEOREM 2.11.3. *Date n v.a. indipendenti e con la stessa densità di probabilità, al limite per n che tende ad infinito la variabile aleatoria somma normalizzata, Z_n , tende ad una gaussiana standard, cioè a media 0 e varianza 1:*

$$(2.11.3) \quad \lim_{n \rightarrow \infty} f_{Z_n}(x) = f_N(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

A prescindere dalla particolare distribuzione che possiedono le v.a. X_i la loro somma tende comunque a diventare gaussiana. Questo risultato è particolarmente utile per modellare numerosi fenomeni fisici quali il rumore termico.

Riprendiamo ora l'esempio del lancio ripetuto di una moneta. Si vuole stimare meglio la quantità

$$P(|\bar{X}_n - p| > \eta)$$

avendo posto $\eta = 0.1$ ed $n = 100$. Siccome la somma di $n = 100$ v.a. di Bernoulli si può ritenere con ottima approssimazione una gaussiana, allora si ha:

$$\begin{aligned} & P\left(\frac{\sqrt{n}}{\sigma} |\bar{X}_n - p| \leq \frac{\sqrt{n}}{\sigma} \cdot \eta\right) = \\ & = P\left(\left|\frac{S_n - np}{\sqrt{n}\sigma}\right| \leq \frac{\sqrt{n}}{\sigma} \cdot \eta\right) = \\ & \simeq P\left(|Z_N| \leq \frac{\sqrt{n}}{\sigma} \cdot \eta\right) = \Phi\left(\frac{\sqrt{n}}{\sigma} \eta\right) - \Phi\left(-\frac{\sqrt{n}}{\sigma} \eta\right) = 2\Phi\left(\frac{\sqrt{n}}{\sigma} \eta\right) - 1 \end{aligned}$$

avendo indicato con Z_N una v.a. gaussiana a media 0 e varianza 1. Per $\eta = 0.1$, $n = 100$ e $\sigma^2 \leq 1/4$ si ha:

$$P\left(|Z_N| \leq \frac{\sqrt{n}}{\sigma} \cdot \eta\right) \simeq 2\Phi(2) - 1$$

la quantità che volevamo stimare si determina facilmente dal risultato precedente:

$$P(|\bar{X}_n - p| > \eta) \simeq 1 - (2\Phi(2) - 1) = 0.0455$$

stima migliore della quantità 0.25 trovata precedentemente.