Appunti di Calcolo delle Probabilità

A cura di Claudio Guida e Giovanni Marchetti

5 aprile 2005

Indice

1 Spazi di probabilità

1.1 Esperimenti casuali, spazio campionario, eventi

Un esperimento casuale genera un evento con una certa probabilita'. Alcuni esempi di eventi casuali: il risultato di un lancio di un dado o di una moneta, il risultato di un trattamento sulla salute di un paziente, la durata di un dispositivo elettronico, la presenza di un disturbo in una immagine trasmessa. L'insieme dei possibili risultati di un esperimento casuale e' chiamato Ω. Per esempio lo spazio campionario associato al lancio di un dado è $\Omega = \{1, 2, 3, 4, 5, 6\}$. Un elemento dello spazio campionario $\omega \in \Omega$ e' chiamato evento elementare. Un sottoinsieme dello spazio degli eventi e' chiamato evento. Nell'esempio del lancio di un dado, l'evento "il risultato è pari" e' $A = \{2, 4, 6\}$, e l'evento "il risultato è minore di 4" e' $B = \{1, 2, 3\}$.

Si definiscono delle operazioni tra eventi A e B:

1. l'unione di A e B, $A \cup B$ e' l'evento che si verifica se si verifica uno almeno dei due eventi;

2. l'intersezione $A \cap B$ e' l'evento che si verifica se si verificano tutti e due;

3. la negazione A^c e' l'evento che si verifica se A non si verifica.

Lo spazio Ω e' detto anche evento certo. L'insieme vuoto $\emptyset$ e' detto evento impossibile. Due eventi A e B che sono disgiunti si dicono anche incompatibili. Se esiste una partizione $E_1, \ldots, E_k$ dello spazio campionario gli eventi componenti si dicono necessari e incompatibili.

In certi casi lo spazio campionario Ω non e' numerabile. Si pensi all'esperimento che consiste nel far funzionare un dispositivo finche' non si guasta. Il tempo fino al primo guasto e' una grandezza continua.

1.2 Spazi di probabilità

La probabilità di un evento è una funzione che associa un numero reale ad ogni evento.

Definizione 1.1. *Sia Ω lo spazio campionario e $\mathcal{A}$ un opportuno insieme di eventi definiti su Ω. La probabilità P è un'applicazione $P : \mathcal{A} \to \mathbb{R}^+$ tale che*

1. $P\left\{\Omega\right\} = 1$,

2. Se $(A_n)_n$ è una successione di eventi a due a due disgiunti allora

$$P\left\{\bigcup_{i=1}^{\infty} A_i\right\} = \sum_{i=1}^{\infty} P\left\{A_i\right\} \tag{1}$$

La terna costituita da Ω, $\mathcal{A}$ e $P(\cdot)$ si dice spazio di probabilita'.

1.3 Spazi di probabilita' uniformi

Se lo spazio campionario ha cardinalita' finita e si puo' assumere che ogni evento elementare abbia la stessa probabilita', la probabilita' di un evento si puo' calcolare facilmente.

Esempio 1. Torniamo al lancio del dado, dove $\Omega = \{1, 2, 3, 4, 5, 6\}$. E' inoltre ragionevole assegnare una stessa probabilità ad ogni evento elementare. Pertanto, poiche'

$$1 = P\left\{\Omega\right\} = P\left\{\{1\} \cup \cdots \cup \{6\}\right\} = P\left\{1\right\} + \cdots + P\left\{6\right\} = 6p \tag{2}$$

la probabilita' p di ogni faccia e' $p = 1/6$.

In generale, sia Ω un insieme a cardinalità finita. Si dice che uno spazio di probabilita' su Ω e' uniforme se ogni evento elementare ω ha la stessa probabilita' p. Ragionando in modo analogo al caso del dado otteniamo $p = 1/\#\Omega$. Di conseguenza per ogni evento $A \in \Omega$ otteniamo

$$P\left\{A\right\} = \frac{\#A}{\#\Omega}. \tag{3}$$

questa definizione di probabilità rappresenta la definizione popolare di probabilità e cioè il rapporto tra il numero di casi favorevoli $\#A$ e il numero di casi possibili. Si faccia attenzione però, che questa definizione è valida solo nell'ipotesi che gli eventi elementari siano equiprobabili.

1.4 Proprieta'

Vediamo alcune proprietà generali di uno spazio di probabilità. Gli eventi A e il complementare A^c sono necessari e incompatibili perche' $A \cup A^c = \Omega$. Quindi

$$P(A \cup A^c) = P(A) + P(A^c) = P(\Omega) = 1.$$

Percio' la probabilita' dell'evento complementare e' $P(A^c) = 1 - P(A)$.

Inoltre, dati due eventi A e B si puo' sempre scrivere

$$A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$$

come unione di tre insiemi disgiunti. Quindi,

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$$

da cui aggiungendo e togliendo $P(A \cap B)$ risulta che in generale la probabilita' dell'unione di due eventi e'

$$P\{A \cup B\} = P(A) + P(B) - P(A \cap B).$$

Questa formula può essere espressa anche quando gli eventi sono più di due, ma l'espressione si complica rapidamente all'aumentare del numero di eventi considerati.

Esempio 2. Tornando all'esempio del dado, vediamo quale è la probabilità di ottenere almeno un 6 lanciando due dadi. Lo spazio degli eventi e' $\Omega = \{(i,j) : i,j = 1,\dots 6\}$ e l'evento e' $A = \{(i,j) : i = 6 \text{ oppure } j = 6\}$ Per calcolare la probabilita' possiamo fare in due modi. Primo modo: notiamo che A e' il complementare di $B = \{(i,j) : i,j = 1,\dots,5\}$. Inoltre $P(B) = 25/36$ e quindi $P(A)1 - 25/36 = 11/36$.

Il secondo metodo consiste nello scrivere $A = A_1 \cup A_2$ dove $A_1 = \{(6,j) : j = 1\dots 6\}$ è l'insieme delle coppie aventi come primo elemento 6 e $A_2 = \{(i,6) : i = 1\dots 6\}$ è l'insieme delle coppie aventi come secondo elemento 6. Notiamo che $A_1 \cap A_2 = \{(6,6)\}$. Quindi:

$$P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 6/36 + 6/36 - 1/36 = 11/36.$$

1.5 Probabilità condizionale e indipendenza

La probabilità di un evento B sapendo che l'evento A si è verificato, cioè la probabilità di B condizionata ad A e' per definizione

$$P\{B|A\} = \frac{P\{A \cap B\}}{P\{A\}} \tag{4}$$

assumendo che $P(A) > 0$. In generale dunque la probabilita' del verificarsi congiunto di due eventi e'

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A).$$

Teorema 1.1. *(Delle probabilita' totali). Se $A_1,\dots,A_k$ e' una partizione dello spazio campionario e B e' un evento qualsiasi*

$$P(B) = P(B \mid A_1)P(A_1) + \dots + P(B \mid A_k)P(A_k).$$

Dimostrazione. Poiche' $B = (B \cap A_1) \cup \dots \cup (B \cap A_k)$ e l'unione e' disgiunta, risulta

$$P(B) = \sum_i P(B \cap A_i) = \sum_i P(B \mid A_i)P(A_i)$$

dove nell'ultima uguaglianza abbiamo applicato la regola del prodotto. □

Esempio 3. Se una popolazione è costituita dal 40% di fumatori (F) e dal 60% di non fumatori (N) e sapendo che tale popolazione è affetta da una malattia respiratoria cronica nelle misure del 25% per i fumatori 7% per i non fumatori vogliamo sapere quale è la probabilità che scegliendo una persona a caso tra la popolazione sia malato. Abbiamo $P\{F\} = 0.4$, $P\{N\} = 0.6$, $P\{M|F\} = 0.25$ e $P\{M|N\} = 0.07$ abbiamo

$$P\{M\} = P\{M \cap F\} + P\{M \cap N\} = P\{F\} P\{M|F\} + P\{N\} P\{M|N\} = 0.142$$

Teorema 1.2. *(Formula di Bayes). Se $A_1, \ldots, A_k$ e' una partizione di Ω e B e' un evento qualsiasi tale che $P(B) > 0$*

$$P\{A_i|B\} = \frac{P\{A_i\} P\{B|A_i\}}{P\{B\}} = \frac{P\{A_i\} P\{B|A_i\}}{\sum_{j=1}^{k} P\{A_j\} P\{B|A_j\}} \tag{5}$$

Dimostrazione. Innanzitutto

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{P(B)}$$

semplicemente applicando la regola di probabilita' condizionata. La formula segue applicando la regola delle probabilita' totali a $P(B)$. □

Definizione 1.2. *Gli eventi $A, B \in \mathcal{A}$ si dicono indipendenti se e solo se*

$$P\{A \cap B\} = P\{A\} P\{B\} \tag{6}$$

Generalmente si dice che $A_1, \ldots, A_n \in \mathcal{A}$ sono indipendenti se e solo se per ogni $k \leq n$ e per ogni scelta degli indici $i_1, \ldots, i_k$ diversi tra loro abbiamo

$$P\{A_{i_1} \cap \cdots \cap A_{i_k}\} = P\{A_{i_1}\} \cdots P\{A_{i_k}\} \tag{7}$$

La definizione precedente implica quindi che se A e B sono eventi indipendenti dobbiamo avere $P\{B|A\} = P\{B\}$ cioè l'informazione sul verificarsi o meno dell'evento A non porta informazioni utili sul verificarsi di B.

Esempio 4. Pensiamo a n lanci di una moneta truccata. Sia p la probabilità di testa e quindi $1 - p$ la probabilità di croce. Un evento elementare in questo esperimento casuale e' una n-upla di risultati. Ad esempio, se $n = 3$ ci sono 2^3 terne

$$\begin{gathered} T_1T_2T_3 \\ T_1T_2C_3 \\ T_1C_2T_3 \\ T_1C_2C_3 \\ C_1T_2T_3 \\ C_1T_2C_3 \\ C_1C_2T_3 \\ C_1C_2C_3 \end{gathered}$$

Se gli n lanci sono indipendenti la probabilita' di una n-upla e' il prodotto delle probabilita' dei risultati dei singoli lanci. Ad esempio, la probabilita' di un'n-upla con i primi k risultati testa e i rimanenti $n - k$ croce e'

$$P(T_1, \cdots T_k C_{k+1} \cdots C_n) = P(T_1) \times \cdots \times P(T_k) \times P(C_{k+1}) \times \cdots P(C_n) = p^k (1-p)^{n-k}.$$

Si noti qualsiasi n-upla con k teste ha la stessa probabilita'.

L'esperimento dei lanci ripetuti di una moneta e' un caso particolare delle cosiddette prove di Bernoulli. Le prove di Bernoulli sono prove indipendenti in cui ogni prova puo' dare come risultato due soli risultati chiamati convenzionalmente successo e insuccesso. Tipicamente la probabilita' di successo si assume che sia costante e viene denotata con p.

1.6 Calcolo combinatorio

Supporremo N e K degli insiemi rispettivamente di cardinalità n e k. Due insiemi hanno la stessa cardinalità se e solo se è possibile metterli in corrispondenza biunivoca.

Proposizione 1.1. *$N \times K$ ha cardinalità nk.*

Consideriamo C_k^n l'insieme di tutti i sottoinsiemi di N di cardinalità k. Gli elementi di C_k^n si chiamano anche combinazione di n elementi presi a k a k.

Proposizione 1.2. *La cardinalità di C_k^n è data da*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{8}$$

Esempio 5. Un esempio particolare e' costituito dalla legge ipergeometrica. Un'urna contiene b palline bianche e r rosse. Si estraggono n palline dall'urna senza rimetterle nell'urna. Si vuole calcolare la probabilità che k palline di queste siano rosse. Ω è l'insieme di tutte le combinazioni di $\{1, \ldots, b+r\}$. Supponiamo che le palline siano numerate e che le rosse siano quelle con indice da 1 a r. L'insieme A_k è l'insieme che contiene le combinazioni che hanno esattamente k palline rosse. La probabilità si ottiene tramite il quoziente tra la cardinalità di A_k e quella di Ω che è del tipo C_n^{b+r}. Un elemento dell'insieme A_k è assegnato fissando un sottoinsieme di cardinalità k di $\{1, \ldots, r\}$ e uno di cardinalità $n-k$ di $\{b+1, \ldots, b+r\}$, quindi A_k può essere messo in corrispondenza con $C_k^r \times C_{n-k}^b$, da cui si ottiene che

$$P\{A_k\} = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}} \tag{9}$$

che vale per $k < r$ e $n - k \leq b$. Questa probabilità di chiama ipergeometrica.

2 Modelli discreti

2.1 Variabili aleatorie e loro distribuzioni

Definizione 2.1. *Dato uno spazio di probabilità $\{\Omega, \mathcal{A}, P\}$ si definisce variabile aleatoria l'applicazione $X : \Omega \to \mathbb{R}$ tale che l'insieme $\{\omega; X(\omega) \leq t\}$ per ogni $t \in \mathbb{R}$ sia in $\mathcal{A}$.*

Una variabile aleatoria è quindi un funzione per cui abbia senso calcolare la probabilità

$$P\{X(\omega) \leq t\} \tag{10}$$

Più in generale interessa la probabilità $P\{X(\omega) \in A\}$ con $A \subset \mathbb{R}$. Indicheremo semplicemente con $X \leq x$ l'evento $\{\omega : X(\omega) \leq x\}$. Si definisce funzione di ripartizione di una variabile aleatoria la funzione

$$F(t) = P\{X \leq t\}. \tag{11}$$

2.2 Variabili aleatorie discrete

Le variabili aleatorie discrete sono variabili aleatorie che prendono al più un insieme discreto di valori reali $\{x = x_1, \ldots, x_n, \ldots\}$. L'insieme $\{X = x\}$ è un evento, per cui possiamo definire una funzione $f(x) = P\{X = x\}$. La funzione f si dice funzione di massa di probabilita' della variabilediscreta X e gode delle proprietà seguenti:

1. $f(x) \geq 0$ per ogni x,
2. $\sum_{i=1}^{\infty} f(x) = 1$.

Esempio 6. Riprendiamo l'esempio degli n lanci della moneta truccata e consideriamo la variabile X che conta il numero di teste in n lanci indipendenti della moneta. La probabilita' che $X = k$ e' data dalla probabilita' che capiti un'n-upla di risultati con k teste. Questo evento e' composto da $\binom{n}{k}$ n-uple disgiunte ciascuna delle quali ha probabilita' $p^k(1-p)^{n-k}$. Possiamo scrivere quindi che

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

La funzione di massa di probabilita' della variabile X e'

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, \ldots, n$$

Questa variabile e' detta *binomiale* di parametri n e $p \in [0,1]$ e si indica con il simbolo $Bin(n,p)$. Misura le probabilita' di ottenere x successi in n prove indipendenti ciascuna con probabilita' di successo p costante.

Esempio 7. Si consideri l'esperimento che consiste nel lanciare una moneta fino a che non esce testa. Si consideri la variabile X che conta il numero di lanci fino ad ottenere testa. La probabilita' che il numero di lanci sia uguale a k e' la probabilita' di ottenere $k-1$ volte croce e testa al k-esimo lancio:

$$P(X = k) = P(C_1 C_2 \ldots C_{k-1} T_k) = (1-p)^{k-1} p^k.$$

L'esempio precedente e' collegato alla cosiddetta variabile geometrica. La variabile geometrica di parametro $p \in [0,1]$ e' definita dalla funzione di probabilita'

$$f(x) = p(1-p)^{k-1}, \quad x = 1, 2, 3 \ldots \infty.$$

Misura la probabilita' che il numero di prove di Bernoulli necessarie per ottenere il primo successo sia uguale a x. Notare che mentre nella binomiale x indica il numero di successi in n prove fissate, nella geometrica x indica il numero di prove necessarie per ottenere 1 successo. Notare che x puo' assumere come valori tutti gli interi diversi da zero.

Si chiama variabile aleatoria di Poisson di parametro $\lambda > 0$ la distribuzione

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2 \ldots\ \infty$$

E' facile verificare che la somma delle probabilita' e'

$$e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{-\lambda} e^{\lambda} = 1.$$

La distribuzione di Poisson è usata tipicamente per modellare il numero di eventi in processi aleatori di punto caratterizzati da una certa intensita' λ. Per esempio il numero di incidenti su un tratto di strada in un dato intervallo di tempo. Si puo' derivare come approssimazione della distribuzione binomiale, per n che tende ad infinito p che tende a zero e np che tende a λ.

2.3 Leggi congiunte e indipendenza

E' utile a volte considerare variabili aleatorie multidimensionali.

Definizione 2.2. *Un'applicazione*

$$X = (X_1, \ldots, X_m) : \Omega \to \mathbb{R}^m \tag{12}$$

in cui $X_1, \ldots, X_m$ *siano variabili aleatorie reali discrete, si chiama variabile aleatoria* $m-$*dimensionale discreta.*

Il caso piu' semplice e' quello delle variabili aleatorie discrete doppie (X, Y). Definiamo un insieme di possibili modalita' (x, y) dell variabile e a ciascuna associamo la probabilita' $P(X = x, Y = y)$ che la variabile X assuma il valore x e la variabile Y assuma il valore y. Resta definita la funzione di probabilita' congiunta del vettore aleatorio (X, Y)

$$f_{XY}(x, y) = P(X = x, Y = y)$$

che sodisfa le proprieta'

1. $f_{XY}(x, y) \geq 0$ per ogni (x, y)
2. $\sum_x \sum_y f_{XY}(x, y) = 1$.

Data la funzione di probabilita' congiunta di (X, Y) la funzione di probabilita' delle variabili componenti X e Y si deduce usando le regole:

$$f_X(x) = \sum_y f_{XY}(x, y), \qquad f_Y(y) = \sum_x f_{XY}(x, y).$$

Esempio 8. Tipicamente queste funzioni di probabilita' si rappresentano mediante tavole dette tavole di contingenza. Ad esempio se $x = 1, 2$ e $y = 1, 2, 3$:

$x \setminus y$	1	2	3	*Totale*
1	$f_{XY}(1,1)$	$f_{XY}(1,2)$	$f_{XY}(1,3)$	$f_X(1)$
2	$f_{XY}(2,1)$	$f_{XY}(2,2)$	$f_{XY}(2,3)$	$f_X(2)$
Totale	$f_Y(1)$	$f_Y(2)$	$f_Y(3)$	1

Le distribuzioni delle variabili componenti X e Y della variabile doppia vengono dette distribuzioni marginali.

Data una variabile aleatoria doppia (X, Y) si definiscono le variabili condizionate di X dato $Y = y$ con funzione di probabilita'

$$f(x \mid Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

e di Y dato $X = x$ con funzione di probabilita'

$$f(y \mid X = x) = \frac{f_{XY}(x, y)}{f_X(x)}.$$

Esempio 9. Sia data la variabile doppia con distribuzione congiunta di probabilita'

$x \setminus y$	1	2	*Totale*
1	0.2	0.1	0.3
2	0.4	0.3	0.7
Totale	0.6	0.4	1

Le distribuzioni condizionate di $X \mid Y = 1$ e di $X \mid Y = 2$ sono le seguenti:

$x \setminus y$	1	2
1	0.33	0.25
2	0.67	0.75
Totale	1.00	1.00

Definizione 2.3. *Le variabili aleatorie X e Y definite su uno stesso spazio di probabilità si dicono indipendenti se e solo se*

$$f_{XY}(x, y) = f_X(x) f_Y(y) \text{ per ogni } (x, y).$$

Si noti che X e Y sono indipendenti se e solo se le distribuzioni condizionate di $X \mid Y = y$ non variano al variare di y (e viceversa).

Esempio 10. Si considerino la variabile doppia (X, Y) con distribuzione congiunta

$x \setminus y$	1	2	*Totale*
1	0.18	0.12	0.3
2	0.42	0.28	0.7
Totale	0.60	0.40	1

E' semplice verificare che X e Y sono indipendenti. Le distribuzioni condizionate di $X \mid Y = 1$ e di $X \mid Y = 2$ sono identiche fra loro e identiche alla distribuzione marginale di X:

$x \setminus y$	1	2
1	0.30	0.30
2	0.70	0.70
Totale	1.00	1.00

2.4 Trasformazioni di variabili aleatorie

Si definiscono variabili aleatorie trasformate funzioni di variabili aleatorie date.

Esempio 11. Se (X, Y) e' una variabile doppia $Q = X^2$, $S = X + Y$ o $Z = X + 10$ sono esempi di variabili trasformate.

Se $Z = g(X)$ e' una trasformazione di una variabile aleatoria discreta X con funzione di probabilita' $f_X(x)$, la funzione di probabilita' di Z e'

$$f_Z(z) = P(Z = z) = P(G(X) = z) = \sum_{x: g(x) = z} f_X(x).$$

Se (X, Y) e' una variabile doppia con funzione di probabilita' congiunta $f_{XY}(x, y)$, la variabile $Z = X + Y$ ha funzione di probabilita'

$$f_Z(z) = \sum_{(x,y): x+y=z} f_{XY}(x, y).$$

Esempio 12. Sia data la variabile doppia (X, Y) con funzione di probabilita' congiunta

$x \setminus y$	1	2	3	*Totale*
1	0.2	0.1	0.2	0.5
2	0.1	0.3	0.1	0.5
Totale	0.3	0.4	0.3	1

La variabile somma Z ha la seguente funzione di probabilita' (a fianco di ogni z e' riportato l'insieme delle coppie rispetto a cui si sono sommate le probabilita' congiunte):

z	$f_Z(z)$	(x, y)
2	0.2	$(1, 1)$
3	0.2	$(1, 2), (2, 1)$
4	0.5	$(1, 3), (2, 2)$
5	0.1	$(2, 3)$

2.5 Valore atteso

Sia X una variabile aleatoria discreta con funzione di probabilita' $f_X(x)$.

Definizione 2.4. *Il valore atteso (o valor medio, o media o speranza matematica) di X denotato con $\mu = E(X)$ e'*

$$E[X] = \sum_x x f_X(x) \tag{13}$$

ammesso che la somma sia ben definita. Con questa espressione si conviene denotare il fatto che

$$\sum_x |x| f(x) < \infty.$$

Il concetto di valore atteso rappresenta la media della variabile aleatoria X. Due variabili aleatorie aventi stessa funzione di probabilita' hanno, se esistono, lo stesso valore atteso.

Se abbiamo $U = g(X)$ e' una variabile aleatoria trasformata il valor medio di Y si puo' calcolare con la regola seguente.

Teorema 2.1. *La variabile trasformata $U = g(X)$ ammette valore atteso se $\sum_x |g(x)| f_X(x) < \infty$ e*

$$E[U] = \sum_x g(x) f_X(x).$$

Se (X, Y) e' una variabile aleatoria doppia con funzione di probabilita' congiunta $f_{XY}(x, y)$ e $U = g(X, Y)$ e' una trasformazione del vettore aleatorio si puo' calcolare il valore atteso di U con la regola seguente.

Teorema 2.2. *$U = g(X, Y)$ ammette valore atteso se $\sum_x \sum_y |g(x, y)| f_{XY}(x, y) < \infty$ e*

$$E[U] = \sum_x \sum_y g(x, y) f_{XY}(x, y).$$

Utilizzando la definizione di valore atteso e il teorema precedente si possono dimostrare i punti della seguente proposizione.

Proposizione 2.1. *Siano X e Y due variabili aleatorie indipendenti, avremo che*

1. $E(kX) = kE(X)$;
2. $E(X + k) = E(X) + k$;
3. $E[X + Y] = E[X] + E[Y]$.

Dimostrazione. La tecnica per la dimostrazione usa la linearita' dell'operatore valore atteso $E(kX) = \sum_x kxf_X(x) = k\sum_x xf_X(x) = kE(X)$. Inoltre

$$
\begin{aligned}
E(X+Y) &= \sum_x\sum_y (x+y)f_{XY}(x,y) \\
&= \sum_x\sum_y xf_{XY}(x,y) + \sum_y\sum_x yf_{XY}(x,y) \\
&= \sum_x x\sum_y f_{XY}(x,y) + \sum_y y\sum_x f_{XY}(x,y) \\
&= \sum_x xf_X(x) + \sum_y yf_Y(y) \\
&= = E(X) + E(Y).
\end{aligned}
$$

Il secondo risultato si dimostra analogamente al primo. □

Se due variabili sono indipendenti si puo' dimostrare il risultato seguente.

Teorema 2.3. *Date due variabili X e Y indipendenti con valore atteso finito $E(X)$ e $E(Y)$, il valore atteso del prodotto e'*

$$E(XY) = E(X)E(Y)$$

il prodotto dei valori attesi.

Dimostrazione. Risulta

$$
\begin{aligned}
E(XY) &= \sum_x\sum_y xyf_{XY}(x,y) \\
&= \sum_x\sum_y xyf_X(x)f_Y(y) \text{ (per l'indipendenza)} \\
&= \sum_x xf_X(x)\sum_y yf_Y(y) \\
&= E(X)E(Y).
\end{aligned}
$$

□

2.6 Momenti, varianza

Definizione 2.5. *Il momento di ordine k intero positivo di una variabile aleatoria discreta X e' il valore atteso, se esiste, di X^k. Il momento centrale di ordine k e' il valore atteso, se esiste, della variabile trasformata $(X - E[X])^k$. Useremo le notazioni*

$$\mu_k = E(X^k), \bar{\mu}_k = E[(X-\mu)^k]$$

rispettivamente per il momento di ordine k e per il momento centrale di ordine k.

Proposizione 2.2. *Siano X e Y due variabili aleatorie.*

1. *Se X ha momento di ordine k finito, il momento di ordine r è anch'esso finito se $r \leq k$,*
2. *Se X e Y hanno momento di ordine k finito anche $X+Y$ ha momento di ordine k finito.*

Un importante momento è il momento centrale del secondo ordine che prende il nome di varianza della variabile aleatoria:

$$\sigma_X^2 = \mathrm{var}\,(X) = E[(X - E(X))^2] = \sum_x (x-\mu)^2 f_X(x).$$

Mentre il valore atteso e' la media dei valori assunti dalla variabile aleatoria X, la varianza rappresenta invece la dispersione di X rispetto alla sua media μ. Se X assume valori molto distanti da μ allora la variabile aleatoria $X-\mu$ sarà grande, e lo sarà quindi anche $E\left[(X-\mu)^2\right]$. Al contrario, se X è costante, risultera' $P(X=\mu)=1$ e quindi la variabile aleatoria $X-\mu$ è nulla con probabilita' 1 e dunque la varianza risulta uguale a zero.

Esempio 13. Consideriamo X ed Y con funzione di probabilita' rispettivamente

x	$f_X(x)$	y	$f_Y(y)$
-1	$1/2$	-100	$1/2$
$+1$	$1/2$	$+100$	$1/2$

La media delle due variabili aleatorie è identica $E[X] = E[Y] = 0$. Invece, mentre la varianza di X e' uguale a 1, la varianza di Y e' 10000.

Questo aspetto della nozione varianza è messo in evidenza dalla seguente risultato.

Teorema 2.4. *1.* $\mathrm{var}\,(X) = E(X^2) - [E(X)]^2$.

2. $\mathrm{var}\,(kX) = k^2 \mathrm{var}\,(X)$

3. $\mathrm{var}\,(X+k) = \mathrm{var}\,(X)$.

Dimostrazione. Risulta subito $\mathrm{var}\,(X) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$ da cui segue la prima regola: la varianza e' uguale alla media dei quadrati meno il quadrato della media. La varianza di $Z = kX$ si ottiene notando che $E(Z^2) = k^2 E(X^2)$ e $[E(Z)]^2 = (kE(X))^2 = k^2[E(X)]^2$. Quindi per differenza si ottiene $k^2\mathrm{var}\,(()\,X)$. Questa proprietà è evidente anche pensando che la varianza esprime la dispersione della variabile X rispetto alla media. Scrivere $X+k$ significa spostare tutti i valori assunti da X, e di conseguenza anche la media di X, di k. Pertanto, la dispersione dei valori di $X+k$ rispetto alla media $\mu+k$ è la stessa. Infine $\mathrm{var}\,(X+k) = \sum_x (X+k-\mu-k)^2 f_X(x) = \mathrm{var}\,(X)$. □

La quantità $\sigma_X = \sqrt{\mathrm{var}\,(X)}$ viene detta deviazione standard di X o anche scarto quadratico medio di X.

2.7 Covarianza di due variabili aleatorie

La covarianza della variabile aleatoria doppia (X,Y) e' il valore atteso, se esiste, della trasformazione $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$:

$$\mathrm{cov}(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y) f_{XY}(x,y).$$

La covarianza e' un indice di associazione lineare delle due variabili. Infatti si dimostra il seguente risultato.

Teorema 2.5. *1. La covarianza e' compresa sempre nell'intervallo* $-\sigma_X\sigma_Y \leq \sigma_{XY} \leq +\sigma_X\sigma_Y$.

2. *La covarianza coincide con* $\sigma_X\sigma_Y$ *se* $P(Y = a + bX) = 1$ *dove* $b > 0$*, cioe' se* X *e' funzione lineare crescente di* Y.

3. *La covarianza coincide con* $\sigma_X\sigma_Y$ *se* $P(Y = a + bX) = 1$ *dove* $b < 0$*, cioe' se* X *e' funzione lineare decrescente di* Y.

In generale la covarianza e' positiva se in media valori sopra la media di X sono associati a valori sopra la media di Y e valori sotto la media di X sono associati valori sotto la media di Y. E' negativa se scarti dalla media di segno positivo per X sono associati a scarti di segno opposto di Y.

Proposizione 2.3. *La covarianza si puo' calcolare come* $\text{cov}\,(X, Y) = E(XY) - E(X)E(Y)$.

Dimostrazione. Dimostrare per esercizio. □

Proposizione 2.4. *La covarianza e' un operatore bilineare simmetrico. Pertanto*

- $\text{cov}\,(X, Y) = \text{cov}\,(Y, X)$*;*
- $\text{cov}\,(a_1X_1 + a_2X_2, Y) = a_1\text{cov}\,(X_1, Y) + a_2\text{cov}\,(X_2, Y)$.

Dimostrazione. Dimostrare per esercizio. □

La covarianza viene utilizzata come misura dell'associazione lineare tra due variabili aleatorie. Una misura piu' conveniente di associazione e' il coefficiente di correlazione che e' una covarianza normalizzata.

Definizione 2.6. *Il coefficiente di correlazione tra* X *e* Y *e'*

$$\text{corr}(X, Y) = \rho_{X,Y} = \frac{\text{cov}\,(X, Y)}{\sqrt{\text{var}\,(X)\,\text{var}\,(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

Proposizione 2.5. *Il coefficiente di correlazione è invariante ai cambiamenti di scala, cioè se* $a, b > 0$ *allora* $\rho_{aX,bY} = \rho_{X,Y}$*. Inoltre risulta sempre* $-1 \leq \rho_{XY} \leq +1$.

Dimostrazione. Dimostrare per esercizio usando le proprieta' della covarianza. □

Definizione 2.7. *Se due variabili* X *e* Y *hanno correlazione (o equivalentemente covarianza) zero si dicono incorrelate.*

Se due variabili sono incorrelate non esiste una relazione lineare tra di loro. E' importante la relazione tra incorrelazione e indipendenza.

Proposizione 2.6. *Se due variabili* X *e* Y *sono indipendenti sono anche incorrelate.*

Dimostrazione. Se X e Y sono indipendenti $E(XY) = E(X)E(Y)$ e quindi $\sigma_{XY} = E(XY) - E(X)E(Y) = 0$. □

L'incorrelazione e' una forma debole di indipendenza: indica infatti l'indipendenza lineare. Si noti percio' che se due variabili sono incorrelate esse non sono necessariamente indipendenti!

Esempio 14. Si consideri la variabile doppia

$x \setminus y$	1	2	3	*Totale*
1	0.0	0.6	0.0	0.6
2	0.2	0.0	0.2	0.4
Totale	0.2	0.6	0.2	1

E' evidente che X e Y non sono indipendenti poiche' $f_{XY}(1,1) = 0 \neq f_X(1)f_Y(1) = 0.2 \times 0.6 = 0.12$. Tuttavia e' semplice verificare (esercizio!) che $\sigma_{XY} = 0$.

La covarianza e' utile anche per calcolare la varianza di una somma di variabili aleatorie.

Proposizione 2.7. *Risulta sempre che* $\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y) \pm 2\text{cov}(X,Y)$. *Se* X *e* Y *sono incorrelate:*

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y).$$

Dimostrazione. Infatti risulta

$$\begin{aligned}
\text{var}(X+Y) &= E\left[(X+Y-E[X+Y])^2\right] \\
&= E\left[((X-E[X]) + (Y-E[Y]))^2\right] \\
&= E\left[(X-E[X])^2\right] + E\left[(Y-E[Y])^2\right] + +2E[(X-E[X])(Y-E[Y])] \\
&= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y).
\end{aligned}$$

Inoltre,

$$\begin{aligned}
\text{var}(X-Y) &= \text{var}(X+(-1)Y) \\
&= \text{var}(X) + \text{var}((-1)Y) + \text{cov}(X,(-1)Y) \\
&= \text{var}(X) + (-1)^2\text{var}(Y) + (-1)\text{cov}(X,Y) \\
&= \text{var}(X) + \text{var}(Y) - 2\text{cov}(X,Y).
\end{aligned}$$

L'ultima relazione segue subito ponendo $\text{cov}(X,Y) = 0$. □

3 Modelli continui

3.1 Definizioni

Sia X una variabile aleatoria, l'insieme $\{X \le t\}$ è un evento, per cui ha senso calcolare la probabilità $P\{X \le t\}$. Come fatto per le v.a. discrete si definisce funzione di ripartizione la funzione

$$F_X(t) = P\{X \le t\}$$

Se si conosce F_X si conosce anche la quantità

$$P\{a \le X \le b\} = P\{X \le b\} - P\{X \le a\} = F_X(b) - F_X(a)$$

Definizione 3.1. *Una variabile aleatoria di dice continua se esiste una funzione* f_X *detta funzione di densita' di probabilita' di* X*, tale che*

1. $f_X(x) \ge 0$ *per ogni* x

2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$

3. *Per ogni* $a \leq b$: $P(a < X < b) = \int_a^b f_X(x)dx$.

Proposizione 3.1. *La funzione di ripartizione di una variabile continua* X *con funzione di densita'* f_X *e'*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

ed e' la primitiva della funzione di densita'. Pertanto $F'_X(x) = f_X(x)$ *per ogni* x.

Dimostrazione. Il risultato segue dal fatto che $P(a < X < b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$. □

Avvertenza: La densità f_X non è univoca, infatti se una funzione g differisce da f_X per un insieme di misura nulla l'integrale di g è identico a quello di f, ma f_X e g sono differenti.

Osservazione: A differenza di quanto accada per le v.a. discrete abbiamo che $P\{X = x\} = 0$. Percio' le v.a. continue

$$P\{a < X < b\} = P\{a \leq X \leq b\} = P\{a \leq X < b\} = P\{a < X \leq b\}$$

sono tutte equivalenti.

Le definizioni date si possono estendere alle variabili aleatorie doppie (X, Y).

Definizione 3.2. *Una funzione* $f_{XY}(x, y)$ *si dice funzione di densita' di probabilita' congiunta delle variabili* (X, Y) *se*

- $f_{XY}(x, y) \geq 0$ *per ogni* (x, y)
- $\int -\infty^{infty} \int -\infty^{infty} f_{XY}(x, y) = 1$
- *Per ogni insieme* $A \subseteq \mathbb{R}^2$ *del piano*

$$P(X \in A) = \int \int_A f_{XY}(x, y)dxdy.$$

3.2 Valore atteso e varianza

Supponiamo di avere una variabile aleatoria X con funzione di ripartizione $F_X(x)$ monotona crescente. La funzione inversa $x = F_X^{-1}(p)$ si chiama funzione quantile.

Definizione 3.3. *Il quantile di ordine* p *e' quel valore* x *tale che* $P(X \leq x) = p$. *Il quantile di ordine 1/2 si dice mediana della variabile aleatoria.*

Definizione 3.4. *Il valore atteso di una variabile continua* X *e'*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$$

assumendo che l'integrale sia ben definito cioe' $\int |x| f_X(x)dx < \infty$.

La regola usuale per trovare il valor medio di una trasformazione $g(X)$ di una variabile si estende anche alle variabili continue.

Proposizione 3.2. *Se $Z = g(X)$ e' una funzione di una variabile aleatoria allora $E(Z) = E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$. Analogamente se $Z = g(X,Y)$ il suo valor medio e'*

$$E(Z) = E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{XY}(x,y)dxdy.$$

La varianza e la covarianza di una variabile continua si definiscono in modo analogo al caso discreto.

Definizione 3.5. *La varianza di una variabile continua X avente media μ e'*

$$\mathrm{var}\,(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$$

Definizione 3.6. *La covarianza di due variabili continue (X,Y) aventi funzione di densita' congiunta f_{XY} e medie rispettive μ_X e μ_Y e'*

$$\mathrm{cov}\,(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f_{XY}(x,y)dxdy.$$

Tutte le proprieta' del valore atteso, della varianza e della covarianza valide nel caso discreto si dimostrano anche nel caso continuo.

3.3 Alcuni esempi di variabili continue

Definizione 3.7. *Una variabile X ha densita' continua uniforme in un intervallo finito $[a,b]$ se ha funzione di densita'*

$$f(x) = \frac{1}{b-a}, \ \textit{se } x \in [a,b]$$

e $f_U(u) = 0$ altrove. Si e' soliti scrivere $X \sim U(a,b)$

Proposizione 3.3. *Il valore atteso e la varianza di $X \sim U(a,b)$ sono rispettivamente:*

$$E(X) = \frac{a+b}{2}, \quad \mathrm{var}\,(X) = \frac{(b-a)^2}{12}.$$

Dimostrazione. Verificare per esercizio. □

Definizione 3.8. *Una variabile aleatoria X si dice esponenziale di parametro $\beta > 0$ se ha funzione di densita'*

$$f(x) = \frac{1}{\beta}e^{-x/\beta} \ \textit{se } x \geq 0$$

e zero altrove. Si scrive $X \sim Exp(\beta)$.

La variabile esponenziale e' usata per modellare le durate dei dispositivi elettronici e i tempi di attesa tra eventi rari. Il parametro β e' la durata media tra gli eventi.

Proposizione 3.4. *Se X ha densita' esponenziale $X \sim Exp(\beta)$ la sua funzione di ripartizione e'*

$$P(X \leq x) = 1 - e^{x/\beta}.$$

La media e la varianza sono rispettivamente

$$E(X) = \beta, \quad \mathrm{var}\,(X) = \beta^2.$$

3.4 Leggi normali

Definizione 3.9. *Si dice che una variabile Z e' Gaussiana (o normale) standard se ha funzione di densita'*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad per \; -\infty < z < \infty$$

Si scrive $Z \sim N(0, 1)$

Proposizione 3.5. *Il valor medio e la varianza della variabile Gaussiana standard sono $E(Z) = 0$ e $\operatorname{var}(X) = 1$.*

La Gaussiana standard e' il prototipo di un'importante classe di funzioni densità ottenute per cambiamento di posizione e scala.

Definizione 3.10. *Si dice che X e' una variabile normale di parametri μ e σ^2 se $X = \mu + \sigma Z$ dove $Z \sim N(0, 1)$. Si scrive $X \sim N(\mu, \sigma^2)$.*

La funzione di densita' generale si determina trasformando la variabile.

Proposizione 3.6. *La funzione di densita' di $X \sim N(\mu, \sigma^2)$ e'*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}, \quad per \; -\infty < z < \infty$$

Il valor medio e la varianza di X sono $E(X) = \mu$ e $\operatorname{var}(X) = \sigma^2$.

Il grafico della densità normale è una campana simmetrica (funzione pari) con asse di simmetria $x = \mu$ e il cui massimo è in corrispondenza di μ. La curva ha due flessi in corrispondenza di $\mu \pm \sigma$ e due asintoti per $x \to \infty$. Le code sono trascurabili per $|x| > 3\sigma$.

La funzione di ripartizione di $Z \sim N(0, 1)$, indicata solitamente con $\Phi(z)$ non e' calcolabile in modo esplicito. Viene calcolata con metodi numerici e tabulata. Qualsiasi probabilita' normale relativa a un intervallo $[a, b]$ si calcola usando il risultato seguente.

Proposizione 3.7. *Se $X \sim N(\mu, \sigma^2)$ la probabilita' relativa a un intervallo si puo' calcolare come*

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

dove $\Phi(z)$ e' la funzione di ripartizione della normale standard.

Dimostrazione. Infatti, $X = \mu + \sigma Z$ e quindi $Z = (X - \mu)/\sigma$ e quindi

$$P(X \leq b) = P\left(Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right).$$

La stessa cosa si puo' fare per a. Il risultato segue dato che $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$. $\square$