ELEMENTI DI PROBABILITA' E STATISTICA.

§1. PROBABILITA'.

1.1. COS'E' LA PROBABILITA'. PROBABILITA' CLASSICA.

1.1.1. Considerazioni introduttive. I giochi d'azzardo.

Spesso, sia nella ricerca scientifica che in varie attivit pratiche, si ha a che fare con eventi le cui cause non sono note, o non sono controllabili, o anche in situazioni in cui, pur essendo le cause note e controllabili, almeno in linea di principio, conviene, per ragioni di tempo o di economia, rinunciare ad indagarle.

Il caso classico quello dei giochi d'azzardo, come i giochi di carte, il lotto, la roulette, i dadi, etc. In tutti questi giochi c' un insieme di singoli risultati di una certa azione ("eventi elementari"), che sono "equivalenti", possono cio, per quanto se ne sa, verificarsi indifferentemente. Non si ha, cio, alcuna ragione che permetta di prevedere un risultato piuttosto che un altro. Questo il senso che spesso si attribuisce all'espressione "scelta a caso".

Nei giochi d'azzardo questa indifferenza, o equivalenza, dei risultati, una condizione ricercata: il gioco "equo" se garantita la scelta a caso, mentre considerato "truccato" se con qualche artificio (o "trucco") si cerca di favorire un risultato scelto in precedenza.

Un'altra condizione che spesso richiesta nei giochi d'azzardo che il rapporto tra le vincite e le somme puntate o scommesse sia tale che il gioco sia "alla pari", cio nessuno dei partecipanti sia favorito a priori.

Questa richiesta comporta gi considerazioni di carattere matematico, che nei casi semplici sono elementari. Se per esempio, consideriamo il caso di un giocatore che gioca contro un banco scommettendo sul risultato del lancio di un dado il problema ha una facile soluzione. Se punta su un singolo valore, ad esempio "3", e vince, dovr avere sei volte la posta, perch sei sono i risultati possibili. (Assumiamo che il giocatore paghi la posta e poi incassi la vincita: la vincita netta data dalla differenza.) Se invece scommetter sul risultato "pari", dovr, in caso di vincita, ricevere solo $6/3 = 2$ volte la posta, perch l'evento su cui ha puntato corrisponde a tre casi elementari.

Per giochi pi complessi il calcolo si complica. Se per esempio, come si fa di solito, si gioca con due dadi, puntando sul risultato della somma, i risultati possibili sono undici (e precisamente $2, 3, \dots, 11, 12$), ma non sono equivalenti. Infatti si vede facilmente che assai pi facile ottenere 7 che non 12. I risultati elementari equivalenti sono le coppie (ω_1, ω_2), dove ω_1 il risultato del primo dado e ω_2 del secondo, e sono in tutto $6 \times 6 = 36$. Quindi, se il gioco alla pari, puntando sul "2" si dovr avere 36 volte la posta, perch $\omega_1 + \omega_2 = 2$ corrisponde al solo evento elementare $(1, 1)$ su 36 possibili, mentre puntando sul 10 si dovr avere solo 12 volte la posta, perch $\omega_1 + \omega_2 = 10$ corrisponde ai tre eventi elementari $(6, 4), (4, 6), (5, 5)$.

Da simili considerazioni si sviluppata, tra il Quattrocento e il Settecento, la teoria matematica della Probabilit.

1.1.2. Breve nota storica.

Gli inizi della trattazione matematica della probabilit, legata, come si detto, a problemi di giochi d'azzardo, si possono far risalire agli italiani Pacioli (1445-1514) e Tartaglia (1499-1557). Partendo dai loro risultati, si svilupp poi nel secolo XVII, in ambiente prevalentemente francese, la teoria classica della probabilit, ad opera soprattutto di Pascal (1623-1662), di Fermat (1601-1665) e J. Bernoulli (1654-1705).

La teoria matematica moderna della probabilit, che la ricollega ad altri settori della matematica, si sviluppata essenzialmente a partire dagli anni trenta del secolo XX, ed oggi uno strumento essenziale di molte discipline scientifiche.

1.1.3. Eventi elementari equivalenti. Probabilit classica o uniforme.

l primo modello matematico generale di probabilit apparve nel secolo XVII, e prende oggi il nome di "probabilit classica" o "uniforme". Il modello, che ora passiamo a descrivere, si applica ogni volta che si ha a che fare con un insieme finito di eventi elementari equivalenti, cio che possono verificarsi indifferentemente.

L' insieme dei possibili risultati Ω prende il nome di **"spazio degli eventi"**. I suoi punti $\omega \in \Omega$ sono gli **"eventi elementari"**. Un **"evento"** un qualsivoglia sottoinsieme $A \subset \Omega$ dello spazio degli eventi, inclusi lo stesso Ω, che prende il nome di "evento certo", e l'insieme vuoto $\emptyset$, detto anche "evento impossibile".

Se i punti sono equivalenti, la possibilit di verificarsi di un evento sar proporzionale al numero dei suoi punti. Ad un evento A naturale pertanto attribuire un numero, detto **"probabilit di A"**, che una misura della possibilit del suo verificarsi:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{numero dei casi favorevoli}}{\text{numero dei casi possibili}}. \tag{1.1.1a}$$

(Qui in seguito, se A un insieme, indichiamo con $|A| = \text{card}\{A\}$ la sua cardinalit, cio il numero dei suoi punti.)

Questa la formula della probabilit classica, che detta anche "uniforme", perch le probabilit degli eventi elementari sono tutte eguali. Il modello della probabilit classica tuttora di importanza fondamentale per l'intuizione. E' in questo contesto che la parola "probabilit" (="dimostrabilit"), che nel sec. XVII indicava la possibilit di dimostrare, sotto certe ipotesi, la correttezza di un'affermazione, diventa un numero che caratterizza la sua "verificabilit".

La probabilit data dalla (1.1.1a) il raporto trala posta e l'incasso in caso di vincita in un gioco equo: pi l'evento probabile, meno si vince. L'espressione storica "numero dei casi favorevoli" fa riferimento al giocatore che scommette sull'evento A.

La (1.1.1a) attribuisce, come si detto, ad ogni evento elementare la stessa probabilit $p(\omega) = \frac{1}{n}$, dove $n = |\Omega|$ il numero delle possibilit. La (1.1.1a) si pu riscrivere nella forma

$$P(A) = \sum_{\omega \in A} p(\omega). \tag{1.1.1b}$$

Per la probabilit degli eventi elementari, che consideriamo una funzione definita su Ω, useremo d'ora in poi la lettera minuscola p, riservando la lettera maiuscola P alla probabilit degli eventi, definita sui sottoinsiemi di Ω. Le due nozioni sono collegate dalla (1.1.1b).

1.1.4. Operazioni con gli eventi.

Si tratta delle operazioni con gli insiemi, che nel contesto della probabilit acquistano nuovi significati.

i) **Unione.** Se $A, B \subseteq \Omega$ sono eventi, la loro unione $A \cup B$ un nuovo evento costituito dagli elementi $\omega \in \Omega$ che appartengono ad A, o a B (compresi gli elementi comuni).

ii) **Intersezione.** Se $A, B \subseteq \Omega$ sono eventi, la loro intersezione $A \cap B$ un nuovo evento costituito dagli elementi $\omega \in \Omega$ che sono comuni ad A e a B.

iii) **Complemento.** Se $A \subseteq \Omega$ un evento, $A^c = \Omega \setminus A$ il suo complemento o evento complementare, costituito da tutti gli elementi $\omega \in \Omega$ che non sono in A.

Due eventi $A, B \subseteq \Omega$ si diconono **incompatibili** se $A \cap B = \emptyset$, cio se non hanno elementi in comune, e non possono quindi essere realizzati contemporaneamente. A ed A^c sono sempre incompatibili: $A \cap A^c = \emptyset$.

E' facile verificare le seguenti relazioni elementari:

$$A \subseteq A \cup B, \quad B \subseteq A \cup B, \quad A \cap B \subseteq A, \quad A \cap B \subseteq B, \quad A \cup A^c = \Omega. \tag{1.1.2}$$

Nel seguito gli eventi specificati da una condizione o dall'elenco degli eventi elementari costituenti li indicheremo di regola con parentesi graffe $\{\cdot\}$.

Esempio 1. Si lancia due volte una moneta. Lo spazio degli eventi il prodotto cartesiano $\Omega = \{T, C\} \times \{T, C\}$, che per convenzione si indica anche $\{T, C\}^2$. Se ω_1 il risultato del primo lancio e ω_2 del secondo, il generico evento elementare sar $\omega = (\omega_1, \omega_2)$.

Dati gli eventi $A = \{\omega_1 = T\}$, $B = \{$ esce C almeno una volta$\}$, abbiamo: $A^c = \{\omega_1 = C\}$, $B^c = \{TT\}$, $A \cap B = \{TC\}$, $A \cup B = \Omega$.

Esercizio 1. Se A, B sono come nell' esempio di sopra, si individuino gli insiemi $A \cap B^c$, $A^c \cap B$, $A^c \cap B^c$, $A^c \cup B^c$.

1.1.5. Propriet della probabilit.

Dalle (1.1.1a,b), aiutandosi con le relazioni (1.1.2), si deducono facimente tre propriet fondamentali della propriet.

i) Se $A \subseteq B$, allora $P(A) \leq P(B)$;

ii) $P(\Omega) = 1$;

iii) Se A, B sono incompatibili si ha $P(A \cup B) = P(A) + P(B)$, e, in particolare, essendo $A \cup A^c = \Omega$, $P(A^c) = 1 - P(A)$.

Esempio 1 (cont.) Torniamo all'esempio 1 visto sopra e calcoliamo le probabilit di $A, B, A^c, B^c, A \cap B$ e $A^c \cap B$.

Abbiamo $A = \{TT, TC\}$, $B = \{CC, CT, TC\}$, per cui $P(A) = \frac{2}{4} = \frac{1}{2}$ e $P(B) = \frac{3}{4}$. Per i complementari abbiamo quindi $P(A^c) = \frac{1}{2}$ e $P(B^c) = \frac{1}{4}$. Inoltre $A \cap B = \{TC\}$, da cui $P(A \cap B) = \frac{1}{4}$, mentre $A^c \cap B = A^c$.

Esercizio 2. Si lancia un dado, e sia $\omega \in \{1, 2, \ldots, 6\}$ il risultato. Dati gli eventi $A = \{\omega \text{ pari}\}$, $B = \{\omega > 3\}$, si calcolino le probabilit $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \cup B)$, $P(A^c)$.

Suggerimento. Si ha $A = \{2, 4, 6\}$, $B = \{4, 5, 6\}$ *e quindi* $A \cap B = \ldots$.

1.2. ELEMENTI DI CALCOLO COMBINATORIO.

La probabilit classica comporta in molti problemi pratici il calcolo della cardinalit di insiemi che corrispondono a diversi modi di ordinare o di raggruppare gli elementi dello spazio degli eventi Ω. Questo l'oggetto del **Calcolo Combinatorio**, di cui diamo qui di seguito alcune nozioni elementari.

1.2.1. Disposizioni complete (o permutazioni) di n oggetti.

Sono i modi di disporre, o ordinare, n oggetti. Si pu pensare di avere n caselle numerate da 1 a n in cui disporre gli oggetti.

Si vuol calcolare il numero D_n di tutti i modi possibili di collocare gli oggetti nelle caselle. E' facile vedere che D_n dato dal prodotto di tutti gli interi da 1 a n:

$$D_n = n! := 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n. \tag{1.2.1}$$

(Il prodotto di tutti gli interi da 1 a n si denota con $n!$ e prende in nome di "n fattoriale".)

Per la dimostrazione della (1.2.1) notiamo innanzitutto che evidente per $n = 1$ e per $n = 2$. Per $n = 3$ si tratta di disporre tre oggetti in tre caselle. Il primo (quello che va, per intendenrci, nella casella 1) pu essere scelto in tre modi, e, per ogni scelta del primo, i due rimanenti si collocano in due modi. Per cui $D_3 = 2 \cdot 3 = 6$.

Si pu procedere con D_4, etc., ma per risolvere il problema per ogni n si usa il metodo dell'induzione matematica: se dimostriamo che la validit della (1.2.1) per $n \le n_0$ implica, qualsiasi sia n_0, la sua validit per $n = n_0 + 1$, abbiamo dimostrato la (1.2.1).

Difatti, se $n = n_0 + 1$ possiamo scegliere l'oggetto da collocare nella prima casella in $n_0 + 1$ modi. Per ogni scelta del primo rimangono n_0 oggetti da disporre nelle n_0 caselle rimaste, e per l'ipotesi induttiva questo si fa in $n_0!$ modi. In totale $D_{n_0+1} = (n_0 + 1)n_0! = (n_0 + 1)!$.

1.2.2. Disposizioni (senza ripetizioni) di n oggetti di classe $k < n$.

Si tratta di disporre non tutti gli n oggetti, ma solo $k < n$ di loro. Abbiamo, cio, solo k caselle in cui disporre gli oggetti.

E' chiaro che il primo oggetto pu essere scelto in n modi, per ogni scelta del primo il secondo oggetto scelto in $n - 1$ modi, etc. Quindi, se $D_{n;k}$ il numero delle disposizioni richieste, abbiamo

$$D_{n;k} = n(n-1)\ldots(n-k+1) \; = \; \frac{n!}{(n-k)!}. \tag{1.2.2}$$

Si noti che per $k = n$ si ritrova la (1.2.1).

Nella seconda eguaglianza bisogna usare la convenzione $0! = 1$, che adotteremo d'ora in poi.

1.2.3. Rapida crescita del fattoriale. Formula di Stirling.

Un fatto che ha conseguenze pratiche e scientifiche rilevantissime la rapida crescita di $n!$ al crescere di n. Se infatti $5! = 120$, per cui provare tutti i modi di ordinare cinque oggetti non difficile, gi con 10 oggetti l'impresa assai onerosa, perch $10! = 3.628.800$.

Per fare un esempio pratico, supponiamo che una squadra di 12 giocatori voglia fotografarsi su un podio in tutte le disposizioni possibili. Se anche per cambiare di posto e scattare impiegassero un solo minuto, la procedura durerebbe 12!=479.001.600 minuti, cio 911 anni e alcuni mesi!

Numeri cos grossi non sono comodi da trattare, e per grandi n si ricorre ad una formula approssimata, detta **"formula di Stirling"**:

$$n! = n^n e^{-n} \sqrt{2\pi n}\, e^{\frac{\theta_n}{n}}, \tag{1.2.3}$$

dove $e = 2,7182\ldots$ la base dei logaritmi naturali (numero di Nepero), e $|\theta_n| < \frac{1}{12}$. Il termine $e^{\frac{\theta_n}{n}}$ chiaramente un termine correttivo che vale praticamente 1 per n anche moderatamente grandi, e quindi in genere si pu trascurare.

Applicando questa formula si pu vedere che il numero dei possibili risultati del mescolamento di un mazzo di carte napoletane (40 carte) $40! \approx 10^{46}$. Per cui anche se una popolazione dell'ordine di 10 miliardi giocasse perennemente a briscola per un periodo dell'ordine dell'et dell'universo (10 miliardi di anni), impiegando un quarto dora per ogni partita, si realizzerebbero non pi di 10^{24} partite, e quindi solo una frazione infinitesima di tutte le possibilit.

Infine si noti che 70! supera il numero dei nucleoni dell'universo (stimato a circa 10^{80}).

1.2.4. Combinazioni di n oggetti di classe k .

Il termine "combinazione" indica che si prescinde dall'ordinamento: sono i sottoinsiemi costituiti da k elementi tratti dal dato insieme.

Sia $C_{n;k}$ il numero di tali combinazioni tratte da un insieme di n oggetti. E' ovvio che se $k = n$ prendiamo tutto l'insieme, e c' quindi una sola scelta: $C_{n;n} = 1$. Se invece $k = 1$ le scelte saranno tante quante sono gli oggetti a disposizione: $C_{n;1} = n$.

Nel caso generale si noti che per ogni data combinazione di k oggetti si hanno esattamente $k!$ disposizioni diverse di classe k, ottenute disponendo i k oggetti in tutti i modi possibili. Quindi $D_{n;k} = k! C_{n;k}$, ovvero

$$C_{n;k} = \frac{D_{n;k}}{k!} = \frac{n!}{k!(n-k)!} := \binom{n}{k}. \tag{1.2.4}$$

Il simbolo $\binom{n}{k}$ detto "coefficiente binomiale", e si noti che $\binom{n}{k} = \binom{n}{n-k}$. Esso compare nella formula dello sviluppo del binomio:

$$(a+b)^n = a^n + \binom{n}{1} a b^{n-1} + \binom{n}{2} a^2 b^{n-2} + \ldots + \binom{n}{n-1} a^{n-1} b + b^n,$$

che, usando la convenzione $0! = 1$ si scrive nella forma compatta

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{1.2.5}$$

Esempio 1. Si lancia una moneta cinque volte. Calcolare quante sono le successioni risultanti in cui "testa" (T) appare esattamente tre volte.

Rappresentiamo la generica successione risultante come $\omega = (\omega_1, \dots, \omega_5)$, con $\omega_i \in \{T, C\}$, $i = 1, \dots, 5$. Le successioni con tre T (e due C) sono tante quanti sono i modi di scegliere i tre posti per le T (i tre indici i per cui $\omega_i = T$), che devono essere tanti quanti sono i modi di scegliere i due posti per le C (i due indici i per cui $\omega_i = C$).

Sono quindi $\binom{5}{3} = \binom{5}{2} = 10$.

Osservazione. L'eguaglianza $\binom{n}{k} = \binom{n}{n-k}$, si verifica facilmente dalla (1.2.4) per ogni possibile scelta di n e $0 \leq k \leq n$, ed una conseguenza del fatto che scegliere un sottoinsieme di k oggetti tra n la stessa cosa che scegliere il sottoinsieme complementare di $n - k$ oggetti.

Esempio 2. Il gioco del lotto consiste nell' estrazione di cinque numeri tra i primi novanta. Per le giocate che non dipendono dall'ordine di estrazione (come il "terno a lotto"), il numero totale delle possibili estrazioni del lotto su una ruota quindi $\binom{90}{5} = \frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}{5 \cdot 4 \cdot 3 \cdot 2}$.

Eercizio 1. Si lancia una moneta quattro volte. Qual' la probabilit di avere esttamente due T?

Esercizio 2. Un'urna contiene quattro sfere nere e sei bianche. Si estraggono contemporaneamente due sfere. Qual' la probabilit che siano entrambe bianche?

1.2.5. Disposizioni con ripetizioni di n oggetti di classe k .

Si tratta di collocare in k caselle n oggetti. Grazie alle ripetizioni, pu ben essere $k > n$, nel qual caso le ripetizioni sono inevitabili. Per l' intuizione, roprio per il fatto che ci sono ripetizioni possibili, meglio parlare di simboli, piuttosto che di oggetti.

L'esempio pi noto quello una colonna del totocalcio, dove $n = 3$ il numero dei simboli, che sono 1,2,X, e $k = 13$ il numero delle caselle.

Detto $\tilde{D}_{n;k}$ il numero di tutte le possibili tali disposizioni si vede facilmente che

$$\tilde{D}_{n;k} = n^k. \tag{1.2.6}$$

Infatti il primo simbolo si sceglie in n modi, e, fissato il primo, anche il secondo si sceglie in n modi, e cos gli altri, fino al k-esimo.

Esempio 3. Se si lancia una moneta cinque volte, il risultato una successione di cinque simboli, ciascuno dei quali pu essere T o C. Tutti i possibili casi sono quindi $2^5 = 32$.

Esempio 4. Nella colonna del totocalcio con tre simboli si devono riempire tredici casello: il numero dei possibili modi di farlo quindi 3^{13}.

1.2.6. Numero di eventi di uno spazio degli eventi finito.

Ci si pu chiedere quale sia il numero di tutti i possibili eventi per uno spazio degli eventi Ω finito, e fatto di $n = |\Omega|$ elementi.

Si tratta del numero di tutti i sottoinsiemi distinti, incluso Ω e l'insieme vuoto $\emptyset$. Per ogni $k < n$ vi sono, per la (1.2.4), $\binom{n}{k}$ sottoinsiemi di k elementi. Quindi il numero totale

, per la (1.2.5),

$$\sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n.$$

Esempio 5. Si vuol calcolare quanti sono tutti i possibili modi di indovinare esattamente 10 risultati compilando una colonna del totocalcio.

Pensando di aver fissato il risultato incognito delle 13 partite, si tratta di calcolare il numero delle possibili colonne con 10 risultati giusti e 3 errati. Possiamo scegliere le 10 partite indovinate (o le 3 sbagliate) in $\binom{13}{10} = \binom{13}{3}$ modi, e siccome una partita si pu sbagliare in due modi, il numero totale $2^3\binom{13}{10} = 2288$.

1.3. PROBABILITA' DISCRETA.

1.3.1. Spazio di probabilit finito: caso generale.

Torniamo all'esempio 1 del §1.1, e sia N_T il numero di "teste" (T) ottenuto nei due lanci della moneta. Si ha

$$P(\{N_T = 0\}) = \frac{1}{4}, \qquad P(\{N_T = 1\}) = \frac{1}{2}, \qquad P(\{N_T = 2\}) = \frac{1}{4}.$$

Se ci interessa solo il valore di N_T possiamo prendere uno spazio degli eventi con tre soli elementi $\Omega = \{0, 1, 2\}$. I tre elementi per non hanno probabilit eguale: infatti $P(0) = P(2) = \frac{1}{4}$, mentre $P(1) = \frac{1}{2}$. Siamo dunque usciti dallo schema classico, anche se le nuove probabilit sono state calcolate in un modello di probabilit classica "sottostante".

In generale, uno **spazio di probabilit finito** dato da una coppia (Ω, p), costituita da un insieme finito Ω, detto "spazio degli eventi", e da una funzione non negativa su di esso $p(\omega) \geq 0, \omega \in \Omega$, che d la probabilit degli eventi elementari (detta anche "densit discreta"), ed tale che $\sum_{\omega \in \Omega} p(\omega) = 1$.

Gli **eventi** sono i sottoinsiemi di Ω, e la probabilit di un evento $A \subseteq \Omega$ data dalla somma delle probabilit dei suoi eventi elementari:

$$P(A) = \sum_{\omega \in A} p(\omega). \tag{1.3.1}$$

Dalla (1.3.1) segue che non pu mai essere $P(A) > 1$, per nessun evento A, perch $P(A)$ data da una somma di probabilit non negative $p(\omega)$, che non pu mai superare la somma su tutti gli elementi $\omega \in \Omega$.

1.3.2. Propriet fondamentali della probabilit.

Le propriet della probabilit sono gi viste nel §1.1.4 valgono anche nel caso generale. Le ripetiamo qui insieme ad alcune immediate conseguenze.

i) Se $A \subseteq B$, allora $P(A) \leq P(B)$;

ii) $P(\Omega) = 1$;

iii) Se A, B sono incompatibili si ha $P(A \cup B) = P(A) + P(B)$, e, in particolare, $P(A^c) = 1 - P(A)$.

Dalla i), usando la iii), si vede che, se $A \subseteq B$, allora $P(B) = P(A) + P(B \setminus A)$.

La iii) si estende a tre o pi eventi. Nel caso di tre eventi, se A, B, C sono incompatibili, cio $A \cap B = A \cap C = B \cap C = \emptyset$, allora $P(A \cup B \cup C) = P(A) + P(B) + P(C)$. Infatti, basta prendere $B' = B \cup C$, verificare che $A \cap B' = \emptyset$, e applicare la iii) due volte. Iterando la procedura si arriva alla conclusione per un qualunque numero di eventi incompatibili.

Per l'unione di due eventi, nel caso generale, si ha la seguente formula

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) = P(A) + P(B) - P(A \cap B). \qquad (1.3.2)$$

Per la dimostrazione si pu considerare che nella somma $P(A)+P(B)$ i punti dell'intersezione $A \cap B$ sono contati due volte, e quindi bisogna sottrarre $P(A \cap B)$. Oppure si pu ragionare cos: $A \cup B$ l'unione di tre eventi incompatibili $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$, e poi si osserva che $A = (A \setminus B) \cup (A \cap B)$ e similmente per B.

Esempio 1. Nel lancio di due dadi, detto (ω_1, ω_2) l'evento elementare, si vuole calcolare la probabilit dell'evento $B = \{\omega_1 \cdot \omega_2 \text{ multiplo di } 3\}$.

Invece del conteggio esplicito, introduciamo gli eventi $B_1 = \{\omega_1 \text{ multiplo di } 3\}$, e $B_2 = \{\omega_2 \text{ multiplo di } 3\}$. Si ha $B = B_1 \cup B_2$, perch se $\omega_1 \cdot \omega_2$ multiplo di 3 almeno uno dei due numeri ω_1, ω_2 deve esserlo. Per la (1.3.2), essendo $P(B_1) = P(B_2) = \frac{2 \cdot 6}{36} = \frac{1}{3}$, e $P(B_1 \cap B_2) = \frac{2 \cdot 2}{36} = \frac{1}{9}$ abbiamo $P(B) = \frac{2}{3} - \frac{1}{9} = \frac{5}{9}$.

Esercizio. Nel lancio di due dadi dell'esempio precedente, si considerino gli eventi $A = \{\omega_1 + \omega_2 \text{ dispari}\}$ e $C = \{\omega_2 \text{ pari}\}$. Si calcolino le probabilit degli eventi A, C, $A \cup C$, $A \cap C$, $(A \cap C)^c$, e $C \cap B$, dove B l'evento considerato nell'esempio.

1.3.3. Il caso di un'infinit (numerabile) di eventi elementari.

La costruzione vista sopra rimane essenzialmente la stessa se si considera uno spazio degli eventi che ha un'infinit numerabile di eventi elementari. Ricordiamo che un insieme infinito "numerabile" se i suoi punti si possono numerare, cio si pu attribuire a ciascun elemento un numero d'ordine. Come noto, vi sono insiemi infiniti, come i punti di un segmento della retta, per cui ci non possibile.

Se Ω numerabile la (1.3.1) pu rappresentare una somma infinita, o "serie", che d per un risultato finito, anzi ≤ 1. Questo fatto si esprime dicendo che la somma infinita, o serie, "converge". Rimandiamo alle lezioni sulle successioni e serie per maggiore informazione.

Esempio 2. Si lancia un dado finch non esce per la prima volta "6". Qual' la probabilit che il gioco si fermi a tre lanci? E qual' la probabilit che il gioco non si fermi mai?

Per la prima domanda consideriamo lo spazio degli eventi per tre lanci, il cui evento elementare denotiamo con $\omega = (\omega_1, \omega_2, \omega_3)$. Detto τ il tempo di arresto, l'arresto al terzo lancio l'evento $\{\tau = 3\} = \{\omega_1 \neq 6, \omega_2 \neq 6, \omega_3 = 6\}$. I casi totali sono 6^3, quelli favorevoli $5 \cdot 5 \cdot 1 = 25$, e la probabilit quindi $P(\{\tau = 3\}) = (\frac{5}{6})^2 \frac{1}{6}$.

Per l'arresto al lancio di numero k si vede analogamente che

$$P(\{\tau = k\}) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}.$$

Applicando la formula armonica

$$1 + a + a^2 + \ldots + a^{n-1} = \frac{1 - a^n}{1 - a},$$

che valida per ogni $a \neq 1$, al caso $a = \frac{5}{6}$ otteniamo

$$P(\{\tau = 1\}) + P(\{\tau = 2\}) + \ldots + P(\{\tau = n - 1\}) = 1 - \left(\frac{5}{6}\right)^n .$$

Per $n \to \infty$ otteniamo la somma della serie, che rappresenta la probabilit che 6 esca prima o poi. Questa serie, come si vede facendo il limite, si somma a 1. Quindi la probabilit del complementare, cio dell'evento che 6 non appare mai, nulla.

Se prendiamo i numeri interi $\Omega = \{1, 2, \ldots\}$ con le probabilit $p(k) = P(\{\tau = k\})$ abbiamo uno spazio di probabilit non pi finito, ma numerabile.

Per **probabilit discreta** si intende che lo spazio degli eventi Ω finito o numerabile. Come vedremo in seguito, si parla invece di **probabilit continua** quando Ω un intervallo della retta, o un sottoinsieme di $\mathbb{R}^n$ che ha la potenza del continuo.

1.4. PROBABILITA' CONDIZIONATA.

1.4.1. Probabilit condizionata o condizionale.

Sia (Ω, p) uno spazio di probabilit discreto, e $A, B \subseteq \Omega$ due eventi con $P(B) > 0$. Si dice **"probabilit condizionata (o condizionale)** di A rispetto a B", o anche "probabilit di A sotto la condizione B" la quantit

$$P(A|B) := \frac{P(A \cap B)}{P(B)}. \tag{1.4.1a}$$

Si pu dire anche che $P(A|B)$ la probabilit di A assumendo che "accade B".

Per B fissato, la (1.4.1a) una nuova probabilit, che denotiamo P_B, su Ω:

$$P_B(A) := P(A|B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} = \sum_{\omega \in A} P_B(\omega). \tag{1.4.1b}$$

L'evento B, come ovvio, diviene certo: $P_B(B) = 1$.

Se $P(A|B) > P(A)$ vorr dire che l'evento A "favorito" da B, e sar "sfavorito" se $P(A|B) < P(A)$. Chiaramente A sfavorito al massimo se incompatibile con B, cio se $A \cap B = \emptyset$, cio $A \subseteq B^c$: in questo caso $P(A|B) = 0$.

Come si vede dalle (1.4.1a,b), per la probabilit $P(A|B)$ conta solo la parte di A che si trova in B, cio $A \cap B$. Il resto, cio $A \cap B^c$, come tutti gli eventi contenuti in B^c, ha probabilit condizionata nulla.

Esempio 1. Si lancia un dado e sia $\omega \in \{1, \ldots, 6\}$ il risultato. Dati $A = \{\omega \leq 3\}$, $B = \{\omega \text{ dispari }\}$, vogliamo calcolare $P(A|B)$.

Abbiamo $P(B) = \frac{1}{2}$, e $P(A \cap B) = P(\{1, 3\}) = \frac{2}{6} = \frac{1}{3}$. Quindi $P(A|B) = \frac{2}{3}$.

Si ha $P(A|B) > P(A) = \frac{1}{2}$, cio, come ovvio, B favorisce A.

Esercizio 1. Si lanciano tre monete, e siano N_T, N_C il numero di teste e croci. Dati gli eventi $A = \{N_T < N_C\}$ e $B = \{N_T \text{ dispari }\}$, calcolare $P(A)$ e $P(A|B)$.

Nota. Storicamente la probabilit condizionata nata dal problema di dividere la posta, in caso di interruzione di un gioco, tenendo conto dei risultati ottenuti.

Supponiamo che due giocatori, G_1 e G_2, lancino a turno per due volte una moneta: chi ottiene il maggior numero di "teste" incassa tutta la posta, e in caso di pareggio la si divide a met. Se al primo turno G_1 ottiene "testa" e G_2 "croce", e poi il gioco si interrompe (p.es., la moneta cade in un tombino), come si calcolano le nuove probabilit per dividere la posta?

Lo spazio di probabilit $\Omega = \{T, C\}^4$ (quattro lanci di moneta), con 16 elementi, e sia l'evento elementare si pu scrivere $\omega = (\omega_1^{(1)}, \omega_2^{(1)}; \omega_1^{(2)}, \omega_2^{(2)})$. Detti $N_T^{(1)}, N_T^{(2)}$ i numeri di teste realizzati dai due giocatori, la probabilit di pareggio

$$P(N_T^{(a)} = N_T^{(b)}) = P(N_T^{(a)} = N_T^{(b)} = 0) + P(N_T^{(a)} = N_T^{(b)} = 1) + P(N_T^{(a)} = N_T^{(b)} = 2).$$

Si ha $P(N_T^{(1)} = N_T^{(2)} = i) = P(N_T^{(1)} = i)\ P(N_T^{(2)} = i) = (P(N_T^{(1)} = i))^2$, $i = 0, 1, 2$, per l'indipendenza delle prove. Sommando si ottiene la probabilit di pareggio che $\frac{3}{8}$. La probabilit che uno dei due giocatori vinca perci $1 - \frac{3}{8} = \frac{5}{8}$, e poich i due sono sullo stesso piano , ciascuno ha probabilit $\frac{5}{16}$ di vincere.

Pertanto se $A = \{G_1 \text{ vince}\}$ e $D = \{\text{pareggio }\}$, abbiamo $P(A) = \frac{5}{16}$ e $P(D) = \frac{3}{8}$.

Sotto la condizione $B = \{\omega_1^{(1)} = T, \omega_1^{(2)} = C\}$ il giocatore G_2 non pu vincere. Abbiamo $D \cap B = \{T, C; C, T\}$, e siccome $P(B) = \frac{1}{4}$, si ha $P(D|B) = \frac{1}{4}$, e $P(A|B) = \frac{3}{4}$.

Quindi tre quarti della posta vanno ad G_1 e il rimanente quarto diviso a met.

1.4.2. Formula della probabilit totale e formula di Bayes.

Formula della probabilit totale. Se lo spazio degli eventi Ω si decompone in modo naturale in parti (eventi), $A_1, \dots, A_n$, caratterizzati da certe propriet, può essere conveniente esprimere le probabilit tramite le corrispondenti probabilit condizionate agli eventi A_k, $k = 1, \dots, n$.

Gli eventi A_k, $k = 1, \dots, n$ devono costituire una "partizione" (o "decomposizione") di Ω, cioè, non devono avere punti in comune (devono essere incompatibili), $A_j \cap A_k = \emptyset$ se $j \neq k$, e devono coprire tutto lo spazio Ω, cioè $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$. L'esempio più semplice di partizione è quello costituito da un evento A e dal suo complementare A^c.

Data una partizione, per ogni evento $B \subset \Omega$, vale la **formula della probabilit totale**:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n). \qquad (1.4.2)$$

La dimostrazione immediata. Infatti gli eventi $B_k := B \cap A_k$ costituiscono la parte di B che sta in A_k, e poich gli A_k ricoprono tutto Ω, si ha $B = B_1 \cup B_2 \cup \ldots \cup B_n$, con i B_k incompatibili. Dunque $P(B) = \sum_{k=1}^{n} P(B_k)$, ma per la definizione di probabilit condizionata, $P(B_k) = P(B \cap A_k) = P(B|A_k)P(A_k)$, e questo dimostra la (1.4.2).

Esempio 2. Tre bacini, A,B,C contengono 100 carpe ciascuno, delle quali $N_A = N_B = 30$ possiedono un certo carattere genetico g, nei bacini A e B, e $N_C = 90$ nel bacino C. Si lancia una moneta: se viene testa si sceglie il bacino A, e se viene croce sceglie un bacino a caso tra B e C. Dal bacino scelto si cattura un pesce a caso e lo si esamina. Qual' la probabilit che possieda il carattere g?

Lo spazio degli eventi Ω è fatto di 300 elementi (ev. elementare il pesce catturato), e si divide in tre parti, a seconda del bacino di provenienza, che indichiamo ancora con A, B, C. Si tratta di una partizione con $P(A) = \frac{1}{2}$ e $P(B) = P(C) = \frac{1}{4}$.

Detto G l' evento che interessa, abbiano $P(G|A) = P(G|B) = 0,3$ e $P(G|C) = 0,9$, e applicando la (1.4.2) si trova $P(G) = \frac{0,3}{2} + \frac{1}{4}(0,3 + 0,9) = 0,45$.

Esercizio 2. Due urne, A e B, contengono, rispettivamente, 3 sfere bianche e 6 nere, e 6 sfere bianche e tre nere. Si lancia un dado, e, detto x il risultato, si prende l'urna A se $x > 4$, la B altrimenti, e si estrae una sfera a caso dall'urna scelta. Qual' la probabilit che sia bianca?

Formula di Bayes. Dato uno spazio degli eventi con una partizione, come per la formula della probabilit totale, può accadere che, accertato un certo evento B, si voglia calcolare la probabilit che esso sia dovuto alle varie parti $A_1, \ldots, A_n$ in cui diviso Ω. Si tratta quindi di determinare le probabilit condizionate degli eventi $A_1, \ldots, A_n$, sotto la condizione che si verificato un certo evento B, cio le probabilit $P(A_k|B)$, $k = 1, \ldots, n$, che sono dette probabilit "a posteriori", mentre le $P(A_k)$ sono dette probabilit "a priori".

La formula che d le probabilit "a posteriori" in funzione delle probabilit "a priori" si deriva dalla formula della probabilit totale (1.4.2), ed detta "formula di Bayes":

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}, \qquad k = 1, \ldots, n. \tag{1.4.3}$$

Infatti il numeratore è $P(A_k \cap B)$ e il denominatore è la probabilità $P(B)$ espressa con la (1.4.2). Quindi per la (1.41a) il membro di destra esprime la probabilità condizionata $P(A_k|B)$.

Esempio 3. Tornando all'esempio 1 di sopra, supponiamo noto che la carpa esaminata ha il carattere g. Qual' la probabilit "a posteriori" che venga dal bacino C?

Sappiamo che $P(G) = 0,45$ e $P(C) = 0,25$, e quindi $P(C|G) = P(G|C)P(C)/P(G) = \frac{0,9 \cdot 0,25}{0,45} = \frac{1}{2}$.

Esercizio 3. Continuando l'esercizio 2 di sopra, si chiede qual' la probabilit "a posteriori", noto che la sfera estratta bianca, che essa provenga dall'urna B.

1.4.3 Eventi indipendenti.

Se $P(A|B) = P(A)$ si dice che **"A indipendente da B"**. Se questo accade, moltiplicando entrambi i membri della (1.4.1a) per $P(B)$ otteniamo la relazione

$$P(A \cap B) = P(A) \cdot P(B). \tag{1.4.4}$$

Dividendo entrambi i membri per $P(A)$ (supponiamo $P(A) > 0$) troviamo $P(B|A) = P(B)$.

Quindi, se A indipendente da B, allora B indipendente da A. La condizione di indipendenza tra eventi simmetrica, e conviene prendere la relazione simmetrica (1.4.4) come sua definizione.

Nella (1.4.4) i due eventi possono anche essere $\emptyset$ o Ω, e si vede che questi eventi sono indipendenti da tutti gli altri. Sono infatti eventi che non danno nessuna "informazione".

Osservazione. Se gli eventi A e B sono indipendenti, lo sono anche gli eventi A e B^c, A^c e B, A^c e B^c.

Infatti, nel primo caso $A \cap B^c$ l'insieme dei punti che sono in A e non in B, quindi bisogna togliere ad A l'insieme $A \cap B$: $A \cap B^c = A \backslash (A \cap B)$, e i due eventi sono incompatibili, per cui $P(A \cap B^c) = P(A) - P(A \cap B)$. Se A e B sono indipendenti $P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$.

Le altre relazioni seguono applicando la prima.

Esempio 4. Torniano all'esempio 1 del §1.1 (due lanci di una moneta), e siano $A_1 = \{\omega_1 = T\}$ e $A_2 = \{\omega_2 = C\}$. Abbiamo $P(A_1 \cap A_2) = P(\{TC\}) = \frac{1}{4}$, e $P(A_1) = P(A_2) = \frac{2}{4} = \frac{1}{2}$. Gli eventi sono indipendenti.

Esempio 5. (Schema di urne con restituzione.) Da un'urna con 60 sfere rosse e 40 bianche si effettuano in successione due estrazioni con restituzione. Consideriamo gli eventi $A = \{1^a \text{ estratta rossa }\}$, $B = \{2^a \text{ estratta bianca }\}$. Il numero delle possibili (doppie) estrazioni $N = (100)^2$, per cui $P(A) = \frac{60 \cdot 100}{N} = 0,6$, $P(B) = \frac{100 \cdot 40}{N} = 0,4$, e $P(A \cap B) = \frac{60 \cdot 40}{N} = 0,24 = P(A) \cdot P(B)$. Gli eventi sono indipendenti.

Esercizio 4. Da un mazzo di carte napoletane (40 carte) se ne estrae una a caso. Si considerino gli eventi $A = \{\text{ un asso}\}$, $B = \{\text{ denari}\}$, $C = \{\leq 5\}$. Vi sono coppie di eventi indipendenti tra le tre coppie possibili A e B, B e C, A e C?

Indipendenza per pi di due eventi.

Se abbiamo tre eventi A, B, C si dice che sono indipendenti, se sono indipendenti a due a due e inoltre $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Nel caso di pi eventi, $A_1, A_n, \ldots, A_n$, sono indipendenti se comunque se ne prendono due o pi, $A_{i_1}, \ldots, A_{i_k}$, $k \geq 2$, si ha

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})\ P(A_{i_2}) \ldots P(A_{i_k}). \tag{1.4.5}$$

Come mostra il seguente esempio non basta che gli eventi siano indipendenti a due a due per essere indipendenti (globalmente). smallskip **Esempio 6.** Si abbiano quattro carte, una bianca, una nera, una rossa, e una quarta con tutti e tre i colori. Se ne estrae una a caso e si consideramo gli eventi $A = \{\text{la carta ha il colore bianco}\}$, $B = \{\text{la carta ha il colore nero}\}$ e $C = \{\text{la carta ha il colore rosso}\}$.

Si ha $P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}$. Inoltre $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$. Quindi i tre eventi sona a due a due indipendenti, ma non sono indipendenti globalmente perch $P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C)$.

1.4.4. Prove indipendenti.

Nella ricerca scientifica e in varie attivit pratiche si ha spesso a che fare con i risultati di prove che sono tra loro indipendenti, nel senso che il risultato di ciascuna prova non influenza i risultati delle altre prove. Cos se si ripete pi volte una prova (lancio di moneta o di dado, test clinico, estrazione del lotto) in condizioni identiche.

Se abbiamo prove indipendenti, eventi che dipendono da risultati di prove diverse devono essere indipendenti. Per esempio, se lanciamo pi volte una moneta o un dado nelle stesse condizioni i corrispondenti risultati sono indipendenti.

Nello schema classico il fatto che tutti i risultati sono equivalenti implica che le diverse prove avvengono nelle stesse condizioni. Cos nell'esempio 1 precedente abbiamo visto che gli eventi $A_1 = \{\omega_1 = T\}$ e $A_2 = \{\omega_2 = C\}$ sono indipendenti.

Consideriamo in generale il caso di una prova ripetuta pi volte in condizioni identiche. Sia Ω lo spazio di probabilit di una singola prova e $p(\omega)$, $\omega \in \Omega$ la probabilit degli eventi elementari. Se ripetiamo la prova n volte, lo spazio degli eventi sar Ω^n (il prodotto cartesiano di Ω per se stesso n volte). Il generico evento elementare $\omega \in \Omega^n$, si pu rappresentare come $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, dove le variabili ω_k, per $k = 1, \ldots, n$, rappresentano il risultato della k-esima prova, $k = 1, 2, \ldots$

Ogni particolare evento elementare $\bar{\omega} = (\bar{\omega}_1, \ldots, \bar{\omega}_n)$, cio ogni scelta concreta dei risultati, $\omega_k = \bar{\omega}_k$ pu essere considerato un evento (consistente in un solo punto di Ω^n), intersezione di n eventi ciascuno dei quali dipende da un solo ω_k:

$$\{\omega = \bar{\omega}\} = \{\omega_1 = \bar{\omega}_1\} \cap \{\omega_2 = \bar{\omega}_2\} \cap \ldots \cap \{\omega_n = \bar{\omega}_n\}.$$

Se le prove sono indipendenti, e p_n denota la probabilit degli eventi elementari in Ω^n, si ha

$$\begin{aligned} p_n(\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_n) &= P\left(\{\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_n\}\right) = P\left(\{\omega_1 = \bar{\omega}_1\} \cap \{\omega_2 = \bar{\omega}_2\} \ldots \cap \{\omega_n = \bar{\omega}_n\}\right) \\ &= P\left(\{\omega_1 = \bar{\omega}_1\}\right) P\left(\{\omega_2 = \bar{\omega}_2\}\right) \ldots P\left(\{\omega_n = \bar{\omega}_n\}\right) = p(\bar{\omega}_1) p(\bar{\omega}_2) \ldots p(\bar{\omega}_n). \end{aligned}$$

Si vede quindi che nel caso di n prove ripetute indipendenti, **la probabilit p_n sullo spazio di n prove Ω^n data dal prodotto delle probabilit p su Ω.**

Pi in generale le prove indipendenti possono essere diverse, con diversi spazi di singola prova $\Omega_1, \ldots, \Omega_n$, e diverse probabilit degli eventi elementari $p^{(1)}, \ldots, p^{(n)}$. Lo spazio delle n prove sar $\Omega^{(n)} = \Omega_1 \times \ldots \times \Omega_n$, e su di esso avremo la **probabilit prodotto**:

$$p_n(\omega_1, \omega_2, \ldots, \omega_n) = p^{(1)}(\omega_1) p^{(2)}(\omega_2) \ldots p^{(n)}(\omega_n). \tag{1.4.6}$$

Riassumendo, possiamo dire che la probabilit degli eventi elementari sullo spazio degli eventi $\Omega^{(n)}$ di n prove indipendenti data dal prodotto delle corrispondenti probabilit sugli spazi di singola prova, secondo la formula (1.4.6).

Nota. Dalla (1.4.6) si ricava facilmente l'indipendenza di eventi generici che dipendono da prove diverse. Supponiamo per semplicit che $n = 3$, e sia $A = \{(\omega_1, \omega_2) \in \mathcal{A}\}$, e $B = \{\omega_3 \in \mathcal{B}\}$. Al variare di $\mathcal{A} \subset \Omega_1 \times \Omega_2$ e $\mathcal{B} \subset \Omega_3$, A e B rappresentano i generici eventi dipendenti dalle prime due prove e dalla terza, rispettivamente. Applicando la (1.4.6) troviamo

$$P(A) = \sum_{(\omega_1, \omega_2) \in \mathcal{A}} p^{(1)}(\omega_1) p^{(2)}(\omega_2) \sum_{\omega_3 \in \Omega_3} p^{(3)}(\omega_3) = \sum_{(\omega_1, \omega_2) \in \mathcal{A}} p^{(1)}(\omega_1) p^{(2)}(\omega_2)$$

$$P(B) = \sum_{(\omega_1,\omega_2)\in\Omega_1\times\Omega_2} p^{(1)}(\omega_1)p^{(2)}(\omega_2) \sum_{\omega_3\in\mathcal{B}} p^{(3)}(\omega_3) = \sum_{\omega_3\in\mathcal{B}} p^{(3)}(\omega_3)$$

e quindi

$$P(A\cap B) = \sum_{(\omega_1,\omega_2)\in\mathcal{A}} p^{(1)}(\omega_1)p^{(2)}(\omega_2) \sum_{\omega_3\in\mathcal{B}} p^{(3)}(\omega_3) = P(A)P(B).$$

Nel caso generale il ragionamento simile.

Dunque, **la probabilit prodotto l'espressione matematica dell'indipendenza delle prove.**

Esempio 7. Si effettuano tre estrazioni con restituzione da un'urna con tre sfere nere e sette bianche. Si calcoli la probabilit che la successione delle estratte sia (b, n, b) (b indica sfera bianca e n sfera nera).

In una singola estrazione le probabilit sono $P(\{b\}) = \frac{3}{10}, P(\{n\}) = \frac{7}{10}$. Pertanto la probabilit richiesta $P(\{b, n, b\}) = (0,3)^2\ 0,7$.

Formula binomiale.

Consideriamo n prove identiche indipendenti con due soli risultati, che chiamiamo convenzionalmente "successo" e "insuccesso" e indichiamo rispettivamente con 1 e 0. Lo spazio degli eventi $\Omega = \{0,1\}^n$, e supponiamo inoltre che in ciascuna prova sia $p(1) = p$, e quindi $p(0) = 1 - p$, con $p \in (0,1)$. Un caso particolare è dato dall'esempio 7 appena visto, una volta stabilito quale dei due risutati (b o n) sia il "successo".

La probabilit di un evento elementare $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ il prodotto delle probabilit, e quindi se $N(\omega)$ il numero degli ω_j per cui $\omega_j = 1$ avremo

$$p(\omega_1, \dots, \omega_n) = p^{N(\omega)}(1-p)^{n-N(\omega)}. \tag{1.4.7}$$

Se ora vogliamo calcolare la probabilit che il numero N di successi prenda un valore fissato k, dovremo sommare le probabilit di tutti gli eventi elementari per i quali $N(\omega) = k$. Per la (1.4.7) la probabilit di ogni tale ω $p^k(1-p)^{n-k}$. Si tratta solo di contare quanti sono gli ω per cui k degli ω_j sono 1 (e quindi $n-k$ sono 0), non importa in quale ordine. E' chiaro che sono tanti quanti i modi di collocare i k simboli 1 nelle n posizioni che corrispondono alle singole prove, e dunque $\binom{n}{k}$. Pertanto

$$P(\{N = k\}) \ = \ \binom{n}{k} p^k (1-p)^{n-k}, \qquad\qquad k = 0, 1, \dots, n. \tag{1.4.8}$$

Osservazione: evento pi probabile. Le probabilit (1.4.8) variano con k, e ci si pu domandare qual' l'evento pi probabile, cio qual'è il numero k di "successi" che ha la massima probabilit.

Supponendo che $p \le \frac{1}{2}$, il minimo si ha per $k = 0$ (se è $p \ge \frac{1}{2}$ si ha per $k = n$): $P(\{N = 0\}) = p^n$. La condizione $P(\{N = k\}) \ge P(\{N = k-1\})$, implica, come si vede facilmente dalla (1.4.8) $p\binom{n}{k} \ge q\binom{n}{k-1}$, che, semplificando i fattoriali, si riduce alla condizione $\frac{k}{n+1} \le p$.

Quindi se k cresce la probabilit aumenta fino a giungere ad un massimo, che per n grandi si ha per $\frac{k}{n} \approx p$, e poi calare.

Esempio 8. Su una popolazione di 100 conigli, di cui 70 di razza a e 30 di razza b, si effettua per 5 volte un test che consiste nel prendere un coniglio a caso, registrarne la razza e reintrodurlo nella popolazione.

Si calcolino le probabilit degli eventi $A = \{$i 5 catturati sono della stessa razza$\}$, e $B = \{$almeno uno dei 5 catturati di razza b$\}$.

Si ha $A = A_1 \cup A_2$, con $A_1 = \{$i 5 sono di razza a$\}$ e $A_2 = \{$i 5 sono di razza b$\}$. $A_1 \cap A_2 = \emptyset$, per cui $P(A) = P(A_1) + P(A_2) = (0,7)^5 + (0,3)^5$.

Poich $B = A_1^c$ si ha $P(B) = 1 - (0,7)^5$.

Esercizio 5. Si lancia tre volte una moneta. Qual' la probabilit che venga due volte di seguito T?

Esercizio 6. Si piantano due semi, uno di germinabilit (=probabilit di germinare) $p_a = 0,9$ e l'altro di germinabilit $p_b = 0,6$. Qual' la probabilit che non germini nessuno dei due? (Si assume indipendenza delle due prove.)

Esercizio 7. Da un'urna con quattro sfere nere e sei bianche se ne estrae una a case per tre volte con restituzione. Calcolare la probabilit dei segenti eventi $A = \{$le sfere estratte sono tutte nere$\}$, $B = \{$si estrae almeno una sfera nera$\}$, $C = \{$si estraggono due sfere bianche e una nera$\}$.

1.5. VARIABILI ALEATORIE.

1.5.1. Variabili aleatorie e loro distribuzioni.

Definizione. Una "variabile aleatoria" sullo spazio di probabilità discreto (Ω, p) è una funzione su Ω a valori reali: $F : \Omega \to \mathbb{R}$.

Di variabili aleatorie ne abbiamo già viste molte: ad esempio, lo sono il numero di "teste" N_T nel doppio lancio di una moneta dell' esempio 1 del §1.1, o la somma $S = \omega_1 + \omega_2$ nel lancio di due dadi.

L'insieme dei valori assunti da F è il codominio, o immagine di F che indichiamo con $X_F \subset \mathbb{R}$:

$$X_F := F(\Omega) = \{x \in \mathbb{R} : F(\omega) = x \text{ per } \text{ qualche } \omega \in \Omega\}.$$

Quindi per ogni elemento $x \in X_F$ si trova almeno un $\omega \in \Omega$ tale che $F(\omega) = x$.

La probabilit su Ω induce allora una probabilit P_F su X_F: l'evento elementare $x \in X_F$ avr probabilit

$$p_F(x) = P(\{F(\omega) = x\}) = P(\{\omega \in \Omega : F(\omega) = x\}) = \sum_{\omega \in F^{-1}(x)} p(\omega). \tag{1.5.1}$$

L'insieme $F^{-1}(x) = \{\omega \in \Omega : F(\omega) = x\} \subset \Omega$, che per brevità possiamo anche denotare con $\{F(\omega) = x\}$, è l'immagine invers di x. Le immagini inverse di punti diversi non hanno ovviamente punti in comune, e la loro unione, per tutti gli $x \in X_F$, l'intero spazio

Ω. Quindi gli eventi $F^{-1}(x)$ costituiscono una partizione e la somma su tutti gli x della (1.5.1) d la probabilit totale:

$$\sum_{x \in X_F} p_F(x) = \sum_{\omega \in \Omega} p(\omega) = 1.$$

Dunque (X_F, p_F) un nuovo spazio di probabilit.

Osservazione 1. Se $F^{-1}(x)$ consta sempre di un solo punto di Ω, allora la F invertibile, e il nuovo spazio di probabilit in pratica identico al precedente: al posto di ogni punto $\omega \in \Omega$ c' un nuovo punto $x \in X_F$ con la stessa probabilit.

Esempio 1. Consideriamo il lancio di un dado, con $\Omega = \{1, 2, \ldots\}$ e la probabilit uniforme, e sia $F(\omega) = 1$ se ω dispari e $F(\omega) = 0$ se ω pari.

Si ha $X_F = \{0, 1\}$ e si vede subito che $p_F(1) = p_F(0) = \frac{1}{2}$.

Esempio 2. Consideriamo il lancio di una moneta, con $\Omega = \{T, C\}$ e la probabilit uniforme, e sia $F(T) = 1$ e $F(C) = 0$.

E' chiaro che (X_F, p_F) la stessa coppia che nell'esercizio precedente.

(Si noti che in questo caso F invertibile.)

Osservazione 2. Il caso di una funzione F che prende valori non numerici, come sono ad esempio i simboli T, C nel caso del lancio di una moneta, non è sostanzialmente diverso, perchè, se siamo nell'ambito della probabilità discreta i valori presi da F si possono numerare o comunque rappresentare con dei numeri.

Definizione. Lo spazio di probabilit (X_F, p_F) prende il nome di **distribuzione** della variabile aleatoria F.

Osservazione 3. Come mostrano gli esempi 1 e 2 appena visti, variabili aleatorie definite su spazi di probabilit diversi possono avere la stessa distribuzione. Di regola quando si parla di variabile aleatoria ci si riferisce alla distribuzione, e non ad una particolare variabile aleatoria con la data distribuzione. In pratica le variabili aleatorie con la stessa distribuzione vengono identificate, o meglio, si passa alla classe di equivalenza delle variabili aleatorie che hanno una data distribuzione.

Esempio 3. Il numero di successi N in n prove indipendenti considerato nel §1.4.4 è una variabile aleatoria e la sua distribuzione data dalla (1.4.8). Tale distribuzione prende il nome di **distribuzione binomiale**, espressione che, per quanto detto nell'osservazione 3, è equivalente all'espressione **variabile aleatoria binomiale .**

La distribuzione binomiale, per n prove con probabilità $p \in (0, 1)$ di successo indicata con il simbolo $B(n, p)$.

Osservazione 4. E' chiaro che ogni funzione di una o pi variabili aleatorie ancora una variabile aleatoria.

Esercizio 1. Nel caso del lancio di tre monete, si trovi la distribuzione della variabile aleatoria N = numero massimo di T consecutive.

Esercizio 2. Si lancia un dado per tre volte, e sia N il numero dei lanci che danno risultati multipli di tre. Si calcoli la distribuzione della variabile aleatoria N^2.

1.5.2. Valor medio di una variabile aleatoria.

Sia data una variabile aleatoria F su uno spazio di probabilit discreto (Ω, p).

Il **valor medio** o **valore aspettato** di F, indicato con il simbolo $M(F)$, è la media dei valori assunti da F pesata con le probabilità con cui vengono assunti:

$$M(F) = \sum_{x \in X_F} x \, p_F(x). \tag{1.5.2}$$

Facendo uso della definizione di distribuzione di F (1.5.1), ed osservando che $\omega \in F^{-1}(x)$ equivale a $F(\omega) = x$, trovimo che il valor medio si può anche scrivere in termini della probabilità sullo spazio di partenza (Ω, p):

$$M(F) = \sum_{x \in X_F} \sum_{\omega \in F^{-1}(x)} F(\omega) \, p(\omega) = \sum_{\omega \in \Omega} F(\omega) p(\omega) \tag{1.5.3}$$

dove si è usato il fatto che gli insiemi $F^{-1}(x)$, per $x \in X_F$, sono una partizione di Ω.

Osservazione 5. Osservazione . Se lo spazio X_F, che è sempre discreto se lo è Ω, ha infiniti elementi, il valor medio $M(F)$ nella (1.5.2) non è una semplice somma, ma una serie, che potrebbe non convergere. Se la serie data dalla (1.5.2) non converge assolutamente, cioè se non converge la serie $\sum_{x \in X_F} |x| \; p_F(x)$, si dice che il valor medio $M(F)$ non esiste.

Un esempio di variabile aleatoria che non ha valor medio si ricava dall'esempio 2 del §1.3. Si consideri la variabile aleatoria $F(\omega) = 2^{\tau(\omega)}$. Il suo valor medio è dato dalla serie

$$\sum_{k=1}^{\infty} 2^k P(\{\tau = k\}) = \frac{1}{6} \sum_{k=1}^{\infty} 2^k \left(\frac{5}{6} \right)^{k-1} = \frac{1}{5} \sum_{k=1}^{\infty} \left(\frac{5}{3} \right)^k = +\infty.$$

Il valor medio un parametro numerico che corrisponde al "centro" dei valori di F, o meglio, come si vede dalla (1.5.2), ad una specie di "baricentro" dei valori assunti dalla F, a ciascuno dei quali viene dato un "peso" pari alla sua probabilit.

Propriet del valor medio.

i) Somma di variabili aleatorie. Se $F = F_1 + F_2$ abbiamo

$$M(F) = \sum_{\omega} [F_1(\omega) + F_2(\omega)] p(\omega) = \sum_{\omega} F_1(\omega) p(\omega) + \sum_{\omega} F_2(\omega) p(\omega) = M(F_1) + M(F_2).$$

Quindi il valor medio della somma di due variabile aleatorie la somma dei loro valori medi, e lo stesso avviene naturalmente per tre o pi variabili.

ii) Moltiplicazione di una variabile aleatoria per un numero. Se k un numero reale, si vede immediatamente dalla (1.5.2) che $M(kF) = kM(F)$.

iii) Variabile aleatoria degenere. Se $F(\omega)$ una costante "quasi ovunque", cio per qualche numero c $F(\omega) = c$ per ogni $\omega \in \Omega$, tranne al più per i punti di un insieme A tale che $P(A) = 0$ (e quindi $p(\omega) = 0$ per ogni $\omega \in A$), allora $M(F) = c$.

Esempio 3. Si lancia nove volte un dado, e sia N il numero di lanci che d un multiplo di tre. Calcolare il valor medio $M(N)$.

N è rappresentabile come una somma di variabili aleatorie $N = \sum_{j=1}^{9} N_j$ dove $N_j = 1$ se il j-simo lancio del dado d 3 o 6, e $N_j = 0$ altrimenti. Per ogni j si ha $M(N_j) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \frac{1}{3}$. Quindi $M(N) = M(N_1) + \ldots + M(N_9) = \frac{9}{3} = 3$.

Disuguaglianze elementari.

Se, per qualche numero a si ha $F(\omega) \geq a$, per ogni $\omega \in \Omega$, allora anche $M(F) \geq a$. Infatti $M(F) = \sum_{\omega \in \Omega} F(\omega)p(\omega) \geq a \sum_{\omega \in \Omega} p(\omega) = a$. Analogamente si vede che se $F(\omega) \leq b$, per ogni $\omega \in \Omega$, allora anche $M(F) \leq b$.

Quindi il valor medio sempre compreso tra il minimo e il massimo di F:

$$\min_{\omega} F(\omega) \leq M(F) \leq \max_{\omega} F(\omega). \tag{1.5.4}$$

Variabile centrata. Se F una variabile aleatoria, la nuova variabile aleatoria $\widehat{F} = F - M(F)$ la variabile aleatoria "centrata" corrispondente ad F. Poich $M(F)$ una costante, il valor medio di $\widehat{F}$ $M(\widehat{F}) = M(F - M(F)\ ) = M(F) - M(F) = 0$.

Quindi una variabile aleatoria centrata ha media nulla.

Esercizio 3. Si lanciano due dadi, e sia N_3 il numero dei "3" che risultano dal lancio. Calcolare il valor medio $M(N_3)$.

1.5.3. Varianza o dispersione di una variabile aleatoria.

Si dice "varianza" o "dispersione" di una variabile aleatoria F, definita sullo spazio di probabilit (Ω, p), la quantit

$$\text{Var }(F) = M\left((F - M(F))^2\right) = M\left(\widehat{F}^2\right), \tag{1.5.5}$$

dove $\widehat{F}$ la variabile centrata relativa ad F definita dianzi. Sviluppando il quadrato si giunge ad una nuova espressione per la varianza:

$$\begin{aligned} \text{Var }(F) &= M\left(F^2 - 2FM(F) + (M(F))^2\ \right) = M(F^2) - 2(M(F))^2 + (M(F))^2 \\ &= M(F^2) - (M(F))^2, \end{aligned} \tag{1.5.6}$$

dove nel calcolare il valor medio si di nuovo usato il fatto che $M(F)$ una costante .

Propriet della varianza.

i) Positivit. Si ha Var $(F) \geq 0$, e Var $(F) = 0$ se e solo se F degenere.

Infatti per la (1.5.5) si ha Var $(F) = \sum_{\omega} p(\omega)\widehat{F}^2(\omega)$. Poichè $\widehat{F}^2(\omega) \geq 0$, se la somma è nulla vuol dire che può essere $\widehat{F}(\omega) \neq 0$ solo se $p(\omega) = 0$, quindi $F(\omega) = M(F)$ a meno di un insieme di misura nulla, e quindi F una costante quasi ovunque, cioè una variabile degenere.

ii) Variazione per cambiamento di scala. Se $G = \alpha F + \beta$, dove α e β sono due numeri fissi, allora Var $(G) = \alpha^2$Var (F).

Infatti per le propriet del valor medio i) e ii), la costante β si cancella nel calcolo della variabile centrata di G, e si ha $\widehat{G} = \alpha\widehat{F}$. Quindi la ii) segue dalla (1.5.5).

Osservazione 6. La trasformazione lineare $G = \alpha F + \beta$ corrisponde da un cambiamento di scala, cio possiamo pensare che F e G sono le misure della stessa grandezza con scale di misura diverse. L'origine $F = 0$ corrisponde al punto $G = \beta$ e l'unit di misura che d il risultato F $|\alpha|$ volte maggiore di quella che d il risultato G. Le due scale hanno lo stesso verso se $\alpha > 0$, cio F e G crescono e calano insieme, mentre hanno verso opposto per $\alpha < 0$, cio se F cresce G cala e viceversa.

Dalla (1.5.6) si vede che il valor medio del quadrato di una variabile aleatoria sempre maggiore od eguale del quadrato del valor medio, ed eguale se e solo se la variabile degenere.

Osservazione 7. Dalla propriet ii) si vede che aggiungendo ad una variabile aleatoria una costante la varianza non cambia (mentre cambia il valore medio). Quindi la varianza indifferente al centro della distribuzione, ma misura piuttosto quanto la distribuzione "sparpagliata" o "dispersa" attorno al valor medio.

Deviazione standard. La deviazione standard o "errore standard" di una variabile aleatoria F la quantit

$$\sigma_F = \sqrt{\mathrm{Var}\ (F)}. \tag{1.5.7}$$

Essa rappresenta, in un certo senso, la distanza media dal centro (cio dal valor medio) dei valori di F.

Esercizio 4. Si lanciano du monete, siano N_T e N_C il numero di teste e croci risultanti, e si consideri la variabile aleatoria $F = N_T - N_C$. Calcolare il valor medio e la varianza di F.

1.6. DISTRIBUZIONE DI PIU' VARIABILI ALEATORIE.

1.6.1. Il caso di due variabili aleatorie.

Supponiamo di avere due variabili aleatorie F_1, F_2 su uno spazio di probabilit discreto (Ω, p). Possiamo considerarle un vettore aleatorio $F(\omega) = (F_1(\omega), F_2(\omega))$, che avr valori nell'insieme di coppie di numeri $X_F = X_{F_1} \times X_{F_2}$, dove $X_{F_1} = F_1(\Omega)$ e $X_{F_2} = F_2(\Omega)$ sono gli insiemi dei valori assunti dalle due variabili.

Detti x_1, x_2 i punti generici di X_{F_1}, X_{F_2}, rispettivamente, la **distribuzione congiunta delle due variabili** F_1, F_2, la probabilit su X_F con densit discreta

$$p_F(x_1, x_2) = P\left(\{F_1 = x_1, F_2 = x_2\}\right). \tag{1.6.1}$$

Se $A \subseteq X_F = X_{F_1} \times X_{F_2}$ un evento qualsiasi, sommando otteniamo

$$P_F(A) := P(\{F \in A\}) = \sum_{(x_1,x_2)\in A} p_F(x_1, x_2). \tag{1.6.2}$$

Conoscendo la distribuzione congiunta si pu calcolare il valor medio di ogni variabile aleatoria esprimibile come funzione di F_1, F_2. Se infatti $G = g(F_1, F_2)$, essa assumer i valori $g(x_1, x_2)$ con probabilit $P(\{F_1 = x_1, F_2 = x_2\}) = p_F(x_1, x_2)$, per cui

$$M(G) = M\left(g(F_1, F_2)\right) = \sum_{(x_1,x_2)\in\Omega_F} g(x_1, x_2) p_F(x_1, x_2). \tag{1.6.3}$$

Le distribuzioni di F_1 ed F_2 (dette "distribuzioni marginali" della distribuzione congiunta) si ottengono facilmente dalla (1.6.1): se specifichiamo che $F_1 = x$ mentre F_2 pu assumere qualsiasi valore, abbiamo

$$p_{F_1}(x) = P(\{F_1 = x\}) = P_F(\{x\} \times X_{F_2}) = \sum_{y \in X_{F_2}} p_F(x, y), \tag{1.6.4}$$

e analogamente

$$p_{F_2}(x) = P(\{F_2 = x\}) = P_F(X_{F_1} \times \{x\}) = \sum_{y \in X_{F_1}} p_F(y, x). \tag{1.6.5}$$

Esempio 1. Si lanciano due dadi, e siano N_3, N_6 il numero, rispettivamente, di "3" e "6" risultanti dal lancio. Si trovi la loro distribuzione congiunta.

E' chiaro che $N_3 \in \{0, 1, 2\}$ e lo stesso vale per N_6, per cui la distribuzione del vettore $F = (N_3, N_6)$ su $X_F = \{0, 1, 2\}^2$. L'evento $\{F_1 = F_2 = 0\}$ corrisponde a lanci che danno (x_1, x_2) con $x_i \notin \{3, 6\}$, $i = 1, 2$. Tali coppie sono in tutto $4 \cdot 4 = 16$, per cui $p_F(0, 0) = P(\{F_1 = 0, F_2 = 0\}) = \frac{16}{36} = \frac{4}{9}$. Similmente si trova $p_F(1, 0) = p_F(0, 1) = \frac{2}{9}$, $p_F(1, 1) = \frac{1}{18}$ e $p_F(2, 0) = p_F(0, 2) = \frac{1}{36}$.

Variabili aleatorie indipendenti.

Le variabili aleatorie F_1, F_2 definite su uno spazio di probabilit (Ω, p) si dicono indipendenti se, per qualunque scelta della coppia di valori $(x_1, x_2) \in X_F = X_{F_1} \times X_{F_2}$, abbiamo

$$p_F(x_1, x_2) = P(\{F_1 = x_1, F_2 = x_2\}) = P(\{F_1 = x_1\})\ P(\{F_2 = x_2\}) = p_{F_1}(x_1) \cdot p_{F_2}(x_2). \tag{1.6.6}$$

In altre parole, gli eventi $\{F_1 = x_1\}$, $\{F_2 = x_2\}$ sono indipendenti qualunque siano $x_1 \in X_{F_1}$, $x_2 \in X_{F_2}$.

1.6.2. Tre o più variabili aleatorie.

Il caso di $n \geq 3$ variabili aleatorie analogo. Posto $F = (F_1, \ldots, F_n)$, la distribuzione congiunta sullo spazio $\Omega_F = \Omega_{F_1} \times \ldots \times \Omega_{F_n}$ sar

$$p_F(x_1, \ldots, x_n) = P(\{F_1 = x_1, \ldots, F_n = x_n\}). \tag{1.20}$$

Le marginali di ciascuna variabile si otterranno sommando sui valori di tutte le altre.

1.7.5. Variabili aleatorie indipendenti.

Per l'indipendenza nel caso di pi variabili avremo

$$p_F(x_1, \ldots, x_n) = P(\{F_1 = x_1, \ldots, F_n = x_2\}) = p_{F_1}(x_1) \ldots p_{F_n}(x_n). \tag{1.21b}$$

Le (1.21a,b) mostrano che la distribuzione p_F sullo spazio $\Omega_F = \Omega_{F_1} \times \Omega_{F_2}$ (o sullo spazio $\Omega_F = \Omega_{F_1} \times \Omega_{F_2} \cdots \times \Omega_{F_n}$) nient'altro che una probabilit prodotto del tipo (1.12) gi vista nel caso delle prove indipendenti.

Ne segue, come spiegato nella nota successiva alla (1.12), che comunque si prendano eventi specificati da variabili diverse, essi sono indipendenti. In particolare sono indipendenti eventi del tipo $\{F_1 \in A_1\}, \ldots \{F_n \in A_n\}$, con $A_1 \in \Omega_{F_1}, \ldots, A_n \in \Omega_{F_n}$.

Valor medio del prodotto di variabili aleatorie indipendenti.

Se F_1 ed F_2 sono indipendenti, ed esistono i valori medi $M(F_1), M(F_2)$, allora si ha $M(F_1 \cdot F_2) = M(F_1) \cdot M(F_2)$. In altre parole, il valor medio del prodotto pari al prodotto dei valori medi.

Infatti possiamo applicare la (1.18c) alla variabile aleatoria $G = F_1 F_2$: tenendo conto dell'indipendenza abbiamo

$$M(F_1 F_2) = \sum_{x \in \Omega_{F_1}, y \in \Omega_{F_2}} x\, y\, p_F(x,y) = \sum_{x \in \Omega_{F_1}} x\, p_{F_1}(x) \sum_{y \in \Omega_{F_2}} y\, p_{F_2}(y) = M(F_1)M(F_2). \tag{1.22}$$

1.7.6. Covarianza o correlazione di due variabili aleatorie.

Date due variabili aleatorie F_1, F_2, su uno spazio di probabilit (Ω, p), si definisce loro covarianza o correlazione la quantit

$$\text{Cov } (F_1, F_2) = M(F_1 \cdot F_2) - M(F_1)M(F_2) = M(\widehat{F}_1 \cdot \widehat{F}_2), \tag{1.23a}$$

dove le $\widehat{F}_i$ sono le variabili centrate. La seconda eguaglianza deriva dal fatto che

$$M(\widehat{F}_1 \widehat{F}_2) = M(F_1F_2 - F_1M(F_2) - F_2M(F_1) + M(F_1)M(F_2)\ ) = M(F_1F_2) - M(F_1)M(F_2).$$

La covarianza di grandissima importanza nelle applicazioni. Se $\text{Cov } (F_1, F_2) > 0$ si dice che F_1, F_2 sono "positivamente correlate", se $\text{Cov } (F_1, F_2) < 0$ si dice che sono "negativamente correlate", e se $\text{Cov } (F_1, F_2) = 0$ si dice che sono "scorrelate".

Dalla (1.22) segue che **variabili indipendenti sono scorrelate**.

Ma il viceversa non vero. Variabili scorrelate possono essere fortemente dipendenti, come mostra il seguente esempio.

Esempio 5. Si consideri il lancio di due monete, e siano N_T, N_C definiti come al solito. Le variabili aleatorie $F_1 = N_T - N_C$ e $F_2 = (N_T - N_C)^2$ non sono indipendenti, perch $F_2 = F_1^2$, ma sono scorrelate.

Infatti $M(F_1) = M(N_T) - M(N_C) = 1 - 1 = 0$, quindi $M(F_1)M(F_2) = 0$. Inoltre $M(F_1F_2) = M((F_1)^3)$, e poich F_1 assume i valori $-2, 0, 2$, abbiamo

$$\text{Cov } (F_1, F_2) = M((F_1)^3) = 2^3 \frac{1}{4} - 2^3 \frac{1}{4} = 0.$$

Coefficiente di correlazione. E' chiaro che la covarianza (1.23a), come la varianza (per la propriet ii) vista sopra) dipende dalla scala di misura delle grandezze.

Conviene allora introdurre una quantit che non ne dipende, e questa data dal coefficiente di correlazione:

$$\rho(F_1, F_2) = \frac{\text{Cov } (F_1, F_2)}{\sqrt{\text{Var } (F_1) \text{Var } (F_2)}}. \tag{1.23b}$$

Infatti se aggiungiamo delle costanti a F_1, F_2, ρ non cambia perch sia Cov (F_1, F_2) che le varianze dipendono solo dalle variabili centrate. Se poi moltiplichiamo F_1, F_2 per α_1, α_2, rispettivamente (che supponiamo entrambe positive), allora sia il numeratore che il denominatore della (1.23b) vengono moltiplicati per $\alpha_1\alpha_2$ e di nuovo ρ non cambia.

Si noti che $\rho(F_1, F_2) \in [-1, 1]$. Per vederlo, basta notare che, preso comunque un numero x, dobbiamo avere

$$0 \leq M\left((\widehat{F}_1 + x\widehat{F}_2)^2\right) = \text{Var }(F_1) + x^2\text{Var }(F_2) + 2x\text{Cov }(F_1, F_2).$$

Come noto, la condizione affinch questa espressione non possa diventare negativa per nessun x che il discriminante sia negativo:

$$\Delta = (\text{Cov }(F_1, F_2))^2 - \text{Var }(F_1)\text{Var }(F_2) \leq 0.$$

Dividendo entrambi i membri per Var (F_1)Var (F_2) e confrontando con la (1.23b) si vede che deve essere $\rho^2 \leq 1$.

E' chiaro che $\rho^2 = 1$ se e solo se $\Delta = 0$. In tal caso x deve essere tale che $\widehat{F}_1 + x\widehat{F}_2 = 0$. Quindi F_1 una funzione lineare di F_2: $F_1 = -x\widehat{F}_2 + M(F_1)$, cio le due grandezze differiscono solo per un cambiamento di scala.

1.7.7. Varianza della somma di variabili aleatorie. A differenza di quanto avviene per il valor medio, la varianza della somma di due variabili aleatorie non in generale data dalla somma delle varianze. Infatti se $F = F_1 + F_2$ la corrispondente variabile centrata $\widehat{F} = F_1 + F_2 - M(F_1) - M(F_2) = \widehat{F}_1 + \widehat{F}_2$, e per le (1.5.5), (1.22a) abbiamo

$$\text{Var }(F_1 + F_2) = M\left(\widehat{F}_1^2 + \widehat{F}_2^2 + 2\widehat{F}_1\widehat{F}_2\right) = \text{Var }(F_1) + \text{Var }(F_2) + 2\text{Cov }(F_1, F_2). \quad (1.24a)$$

Nel caso di pi variabili aleatorie si ottiene facilmente la formula generale

$$\text{Var }(F_1 + \ldots + F_n) = \sum_{i=1}^{n} \text{Var }(F_i) + 2\sum_{\{i,j\}} \text{Cov }(F_i, F_j), \quad (1.24b)$$

dove con $\sum_{\{i,j\}}$ si intende la somme sulle coppie di indici distinti $i \neq j$.

Osservazione 7. Se le variabili sono scorrelate, in particolare se, come visto nel paragrafo precedente, sono indipendenti, allora le covarianze sono nulle e la varianza della somma di variabili aleatorie eguale alla corrispondente somma delle varianze.

1.7.8. Variabili aleatorie discrete notevoli. Descriviamo qui le variabili aleatorie discrete pi importanti.

1. Variabile aleatoria binomiale. La variabile aleatoria binomiale $B(n; p)$, relativa ad $n = 1, 2\ldots$ prove e con probabilit di "successo" $p \in (0, 1)$ in ciascuna prova, stata gi introdotta nel §1.7.1, e la sua distribuzione data dalla (1.13b). Vogliamo ora calcolarne il valor medio e la varianza.

Indicando per semplicit (come abbiamo gi fatto) la variabile con N, invece che con $B(n;p)$, possiamo scrivere $N = F_1 + F_2 + \ldots + F_n$, dove le F_i rappresentano il "numero di successi" nella singola prova i-esima, $i = 1, \ldots, n$. Questo numero pu essere 0 oppure 1, cio $F_i \in \{0,1\}$, per cui $F_i^2 = F_i$ (eguaglianza che vale infatti sia per $F_i = 0$ che per $F_i = 1$). Inoltre $P(\{F_i = 1\}) = p$, per cui

$$M(F_i) = 1 \cdot p = p, \qquad \text{Var}\,(F_i) = M(F_i^2) - (M(F_i))^2 = M(F_i) - p^2 = p - p^2 = p(1-p).$$

Il valor medio quindi $M(N) = np$. Per la varianza si osservi che, siccome le F_i sono indipendenti, la varianza la somma delle varianze, per cui Var $(N) = np(1-p)$.

2. Variabile aleatoria Poissoniana. La variabile aleatoria poissoniana, o di Poisson, approssima la variabile binomiale quando n molto grande, con p molto piccolo, in modo che il valor medio np sia finito. La variabile poissoniana si ottiene nel limite $n \to \infty$, se la probabilit dipende da n (diciamo $p = p_n$) e cambia in modo che np_n converge ad un limite: $np_n \to \rho > 0$.

Consideriamo il caso semplice in cui np_n costante: $np_n = \rho$, e quindi $p_n = \frac{\rho}{n}$. Per un dato k la formula (1.13b) ci d

$$P(\{N = k\}) \;=\; \binom{n}{k}\left(\frac{\rho}{n}\right)^k (1 - \frac{\rho}{n})^{n-k} = \frac{\rho^k}{k!}\,\frac{n(n-1)\ldots(n-k+1)}{n^k}(1-\frac{\rho}{n})^{-k}(1-\frac{\rho}{n})^n.$$

Poich k fisso e $\frac{\rho}{n} \to 0$ per $n \to \infty$, abbiamo $(1 - \frac{\rho}{n})^{-k} \to 1$. Inoltre, ricordando la successione che definisce il numero e, base dei logaritmi naturali, abbiamo $(1 - \frac{\rho}{n})^n \to e^{-\rho}$. Infine, per ogni $j = 0, \ldots, k-1$ si ha, per $n \to \infty$, $\frac{n-j}{n} \to 1$, per cui $\frac{n(n-1)\ldots(n-k+1)}{n^k} to 1$.

Il limite di queste probabilit d la distribuzione di una nuova variabile aleatoria F, che pertanto

$$P(\{F = k\}) = e^{-\rho}\,\frac{\rho^k}{k!}, \qquad k = 0, 1, 2, \ldots. \tag{1.25}$$

E' detta "variabile aleatoria poissoniana" con valor medio ρ. Si tratta di una variabile discreta, ma con un numero infinito di valori, perch abbiamo preso il limite $n \to \infty$.

Si potrebbe dimostrare che $\sum_{k=1}^{\infty} \frac{\rho^k}{k!} = e^{\rho}$, per cui la probabilit totale , come deve, pari a 1. Il valor medio , come abbiamo visto, ρ. La dispersione anche il limite della dispersione della binomiale:

$$\text{Var(F)} \;= \lim_{n\to\infty} np_n(1 - p_n) = \lim_{n\to\infty} n\frac{\rho}{n}(1 - \frac{\rho}{n}) = \rho.$$

Osservazione 8. La distribuzione poissoniana si applica ad eventi che non accadono quasi mai in una singola prova, ma il cui numero medio non trascurabile a causa del gran numero di prove. Un esempio pu essere quello delle chiamate al 113 in un dato giorno in una grande citt. (La singola prova il comportamento di un dato abitante: positiva se chiama il 113.) Se ciascuno ha probabilit $p = 10^{-5}$ di chiamare il 113, e la citt ha un milione di abitanti, la media $\rho = np = 10$. Un esempio ancora migliore quello del decadimento radioattivo. Una massa di materiale radioattivo pu essere dell'ordine di 10^{23}

atomi, ma ciascun atomo decade con probabilit cos piccola che il conteggio al Geiger in un'ora di qualche unit.

1.7.9. Legge dei grandi numeri. La legge dei grandi numeri uno dei risultati fondamentali della teoria della Probabilit, se non addirittura il pi importante.

La sua formulazione semplice. Consideriamo una **successione infinita di variabili aleatorie indipendenti** $F_1, F_2, \ldots,$ tutte con la stessa distribuzione. Questa la situazione che abbiamo quando ripetiamo delle prove in condizioni identiche. Il valore medio e la dispersione delle variabile (che sono le stesse per tutte, perch la distribuzione la stessa) le indichiamo con $M(F_i) = m$ e $\mathrm{Var}\,(F_i) = \sigma^2$, $i = 1, 2 \ldots.$

Vale allora il seguente teorema, detto "Legge dei Grandi numeri".

Teorema. Per ogni scelta del numero $\epsilon > 0$ comunque piccolo, posto $S_n = F_1 + F_2 + \ldots + F_n$, si ha

$$\lim_{n\to\infty} P\left(\left\{\left|\frac{S_n}{n} - m\right| \;>\; \epsilon\right\}\right) \;=\; 0. \tag{1.26}$$

Il teorema afferma che la media aritmetica $\frac{S_n}{n} = \frac{1}{n}(F_1 + F_2 + \ldots + F_n)$, con probabilit che tende ad 1 per $n \to \infty$, vicina quanto si vuole alla media probabilistica m.

Dimostrazione . La dimostrazione molto semplice. Osserviamo che $M(S_n) = n \cdot n$, per cui la probabilit nella (1.26) si pu scrivere, motiplicando per n entrambi i membri della disuguaglianza, come $P(\{|\widehat{S}_n| > n\epsilon\})$,con $\widehat{S}_n = S_n - M(S_n)$.

Ora, se F una variabile aleatoria non negativa, e $a > 0$ un numero, facile vedere che vale la disuguaglianza (detta "disuguaglianza di Chebyshev")

$$P\left(\{F > a\}\right) \;\leq\; \frac{M(F)}{a}. \tag{1.27}$$

Infatti, usando la distribuzione di F, si ha

$$M(F) = \sum_x p_F(x)x \leq \sum_{x>a} p_F(x)x \leq a \sum_{x>a} p_F(x) = aP(\{F > a\}).$$

(Il primo $\leq$ deriva dal fatto che sommiamo solo su $x > a$ e il secondo dal fatto che sotituiamo $x > a$ con a.)

La probabilit (1.26) si stima con la (1.27) ponendo $F = \widehat{S}_n^2$ e $a = n^2\epsilon^2$. Infatti

$$P(\{|\widehat{S}_n| > n\epsilon\}) = P(\{\widehat{S}_n^2 > n^2\epsilon^2\}) \leq \frac{\mathrm{Var}\,(\widehat{S}_n)}{n^2\epsilon^2},$$

dove abbiamo usato il fatto che $M(\widehat{S}_n^2) = \mathrm{Var}\,(S_n)$. Essendo le variabili indipendenti (Osservazione 7), la varianza della somma la somma delle varianze, per cui $\mathrm{Var}\,(S_n) = n\sigma^2$. In conclusione si ha

$$P\left(\left\{\left|\frac{S_n}{n} - m\right| \;>\; \epsilon\right\}\right) \;\leq\; \frac{\sigma^2}{n\epsilon^2}. \tag{1.28}$$

L'espressione a destra tende a 0 per $n \to \infty$, e il teorema dimostrato.

Applicazione della Legge dei grandi numeri alla binomiale e conseguenze. Se consideriamo la distribuzione binomiale discussa al punto 1 del §1.7.8, abbiamo visto che il numero di successi N si pu scrivere come $\sum_{i=1}^{n} F_i$, con le F_i indipendenti e $M(F_i) = p$ (la probabilit di successo in ogni singola prova). Quindi N ha le stesse propriet della S_n del teorema e possiamo applicare la (1.26), ottenendo

$$\lim_{n\to\infty} P\left(\left\{\left|\frac{N}{n} - p\right| \;>\; \epsilon\right\}\right) \;=\; 0. \tag{1.29}$$

Il rapporto $\frac{N}{n}$ la "frequenza" dei successi nelle n prove. Si pu quindi dire che, per la Legge dei grandi numeri, **la frequenza dei successi in n prove indipendenti tende per $n \to \infty$ alla probabilit p.**

E' chiaro che la (1.29) fornisce uno strumento per il calcolo, almeno approssimato, di probabilit incognite tramite le frequenze.