

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 attività 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 è quello dei giochi d'azzardo, come i giochi di carte, il lotto, la roulette, i dadi, etc. In questi giochi c'è un insieme di singoli risultati di una certa azione (che chiameremo "eventi elementari"), che sono "equivalenti", possono cioè, per quanto se ne sa, verificarsi indifferentemente. Ovvero, non c'è alcuna ragione che permetta di prevedere un risultato piuttosto che un altro. Un meccanismo o una procedura che garantiscano questa indifferenza prendono spesso il nome di "scelta a caso".

Nei giochi d'azzardo questa indifferenza, o equivalenza, dei risultati, è una condizione ricercata: se non è garantita la scelta a caso, come accade se con qualche artificio si cerca di favorire un risultato determinato in precedenza, si parla di gioco "truccato".

Una condizione che è spesso richiesta nei giochi d'azzardo in cui si gioca per denaro, è che il rapporto tra le vincite e le somme puntate o scommesse sia tale che il gioco sia "equo", o "alla pari", cioè che non ci siano partecipanti favoriti a priori. Nei casi semplici il rapporto "equo" si determina facilmente. Se, per esempio, chi lancia un dado scommette sul risultato "3", dovrà ottenere in caso di vincita sei volte la posta, perchè sei sono i risultati possibili. Se invece scommette sul risultato "pari", dovrà ricevere solo $6/3 = 2$ volte la posta, perchè l'evento "pari" corrisponde a tre casi elementari.

Ma già un gioco appena più complesso, come il tradizionale gioco con due dadi, in cui si punta sulla somma dei due risultati, richiede qualche calcolo. La somma risultante dei due dadi può assumere undici valori, $2, 3, \dots, 11, 12$, che però non sono equivalenti. I risultati elementari equivalenti sono le coppie (ω_1, ω_2) , dove ω_1 è il risultato del primo dado e ω_2 del secondo, e sono in tutto $6 \times 6 = 36$. Quindi, se il gioco è alla pari, puntando sul "2" si dovrà avere 36 volte la posta, perchè $\omega_1 + \omega_2 = 2$ corrisponde al solo evento elementare $(1, 1)$ su 36 egualmente possibili, mentre puntando sul 10 si dovrà avere solo 12 volte la posta, perchè $\omega_1 + \omega_2 = 10$ corrisponde ai tre eventi elementari $(6, 4), (4, 6), (5, 5)$.

L'esempio visto è un caso elementare delle considerazioni che hanno portato, tra il Quattrocento e il Settecento, agli inizi della teoria matematica della probabilità. E' da notare che la prima definizione precisa della probabilità è stata proprio data come rapporto tra posta e vincita in un gioco equo.

1.1.2. Breve nota storica.

Gli inizi della trattazione matematica della probabilità, legata, come si è detto, ai problemi dei giochi d'azzardo, si fanno risalire agli italiani Luca Pacioli (1445-1514) e Niccolò Fontana, detto il Tartaglia (1499-1557). Partendo dai loro risultati, si sviluppò poi nel secolo XVII, in ambiente prevalentemente francese, la teoria classica della probabilità, ad opera soprattutto di Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665) e Jacques Bernoulli (1654-1705).

Il termine "probabilità" era una parola dotta usata nel Seicento con il senso di dimostrabilità (dal latino *probare*) di una certa asserzione di carattere giuridico o morale, sulla base delle norme o dei principi. L'esame (e il conteggio) dei modi in cui un evento si può realizzare spiega la sua estensione, usata già nel Seicento, al senso quantitativo attuale, di misura, cioè, della possibilità che un evento ha di accadere in un certo contesto.

La teoria matematica moderna della probabilità, che la ricollega ad altri settori della matematica, si è sviluppata essenzialmente a partire dagli anni trenta del secolo XX, ad opera soprattutto del matematico russo A.N. Kolmogorov, ed è oggi uno strumento essenziale di molte discipline scientifiche.

1.1.3. Eventi elementari equivalenti. Probabilità classica o uniforme.

Il primo modello matematico di carattere generale della probabilità apparve nel secolo XVII. Si tratta della "probabilità classica" o "uniforme", che si applica ogni volta che si ha a che fare con un insieme finito di eventi elementari equivalenti, che possono, cioè, verificarsi indifferentemente, come detto sopra. Il modello della probabilità classica è tuttora d'importanza fondamentale per l'intuizione.

L'insieme dei possibili risultati prende il nome di "**spazio degli eventi**", e lo indicheremo in generale con Ω . I suoi elementi $\omega \in \Omega$ sono detti "**eventi elementari**". Un "**evento**" è un qualsiasi sottoinsieme $A \subset \Omega$ dello spazio degli eventi, inclusi lo stesso Ω e l'insieme vuoto \emptyset , che prendono il nome di "evento certo", ed "evento impossibile".

Se i punti sono equivalenti, la possibilità di verificarsi di un evento non può che essere proporzionale al numero dei suoi punti. Ad un evento A è naturale pertanto attribuire, come misura della possibilità del suo verificarsi il numero, detto "**probabilità di A** ",

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

L'espressione storica "numero dei casi favorevoli" fa riferimento alle scommesse. (Se A è un insieme, $|A| = \text{card}\{A\}$ indica la sua cardinalità, cioè il numero dei suoi punti.)

La (1.1.1a) è la formula della probabilità classica, che è detta anche "uniforme", perchè gli eventi elementari hanno tutti la stessa probabilità $p(\omega) = \frac{1}{n}$, dove $n = |\Omega|$. La (1.1.a) si può riscrivere nella forma

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

Notazione. Per la probabilità degli eventi elementari, che è una funzione definita su Ω , usiamo la lettera minuscola p , riservando la lettera maiuscola P alla probabilità 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 probabilità acquistano nuovi significati.

i) **Unione.** Se $A, B \subseteq \Omega$, la loro unione $A \cup B$ è l'evento costituito dagli elementi $\omega \in \Omega$ che sono in A oppure in B .

ii) **Intersezione.** Se $A, B \subseteq \Omega$, la loro intersezione $A \cap B$ è l'evento costituito dagli elementi $\omega \in \Omega$ che sono sia in A che in B , cioè dagli elementi comuni ad A e B .

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

Due eventi $A, B \subseteq \Omega$ si dicono **incompatibili** se $A \cap B = \emptyset$, cioè 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:

$$A \subseteq A \cup B, \quad B \subseteq A \cup B, \quad A \cap B \subseteq A, \quad A \cap B \subseteq B, \quad A \cup A^c = \Omega. \quad (1.1.2)$$

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

1.1.5. Proprietà della probabilità.

Dalle (1.1.1a,b), aiutandosi con le relazioni (1.1.2), si deducono facilmente tre proprietà semplici, ma fondamentali, della probabilità.

- i) $P(\Omega) = 1$;
- ii) se $A \subseteq B$, allora $P(A) \leq P(B)$;
- 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. Si lancia due volte una moneta. Lo spazio degli eventi è il prodotto cartesiano $\Omega = \{T, C\} \times \{T, C\}$, che per convenzione si denota anche $\{T, C\}^2$. Se ω_1 è il risultato del primo lancio e ω_2 del secondo, il generico evento elementare è $\omega = (\omega_1, \omega_2)$.

Consideriamo gli eventi $A = \{\omega_1 = T\}$, $B = \{\text{esce } C \text{ almeno una volta}\}$. Calcoliamo le probabilità di A, B , e degli eventi $A^c, B^c, A \cap B, A \cup B$.

In termini degli eventi elementari abbiamo $A = \{TT, TC\}$, $B = \{CC, CT, TC\}$, per cui $P(A) = \frac{1}{2}$ e $P(B) = \frac{3}{4}$. Per gli altri eventi si ha $A^c = \{\omega_1 = C\} = \{CT, CC\}$, $B^c = \{TT\}$, $A \cap B = \{TC\}$, $A \cup B = \Omega$, e le probabilità sono $P(A^c) = \frac{1}{2}$, $P(B^c) = \frac{1}{4}$, $P(A \cap B) = \frac{1}{4}$.

Esercizi.

1. Continuando l'esempio 1 qui sopra, calcolare le probabilità degli eventi $A \cap B^c$, $A^c \cap B$, $A^c \cap B^c$, $A^c \cup B^c$.

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

1.2. ELEMENTI DI CALCOLO COMBINATORIO.

La probabilità classica comporta in molti problemi pratici il calcolo della cardinalità di insiemi che corrispondono a diversi modi di ordinare o di raggruppare gli elementi dello spazio degli eventi Ω . Questo è 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 può pensare di avere n caselle numerate da 1 a n in cui disporre gli oggetti.

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

$$D_n = n! := 1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n. \quad (1.2.1)$$

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

Infatti la (1.2.1) è evidente per $n = 1$ e per $n = 2$. Per $n = 3$ si tratta di disporre tre oggetti in tre caselle. L'oggetto che va nella prima casella può essere scelto in tre modi, e, per ogni tale scelta, i due rimanenti si collocano in due modi. Per cui $D_3 = 2 \cdot 3 = 6$.

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

Difatti, supponiamo che la (1.2.1) valga fino ad $n = n_0$. Se ora $n = n_0 + 1$ possiamo scegliere l'oggetto da collocare nella prima casella in $n_0 + 1$ modi. Per ogni tale scelta dobbiamo disporre n_0 oggetti nelle n_0 caselle rimaste, e per l'ipotesi induttiva questo si fa in $n_0!$ modi. Quindi $D_{n_0+1} = (n_0 + 1)n_0! = (n_0 + 1)!$.

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, cioè, solo k caselle in cui disporre gli oggetti.

Come prima, si considera che il primo oggetto può essere scelto in n modi, per ogni scelta del primo il secondo oggetto è scelto in $n - 1$ modi, etc. Quindi, se $D_{n;k}$ è il numero delle disposizioni richieste, abbiamo

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

Per $k = n$, usando la convenzione $0! = 1$, si ritrova la (1.2.1).

1.2.3. Rapida crescita del fattoriale. Formula di Stirling.

Un fatto che ha conseguenze scientifiche relevantissime è la rapida crescita del fattoriale $n!$ al crescere di n . Se $5! = 120$, per cui provare tutti i modi di ordinare cinque oggetti non è difficile, già con 10 oggetti l'impresa è assai onerosa. Infatti $10! = 3.628.800$.

Per valutare la rilevanza pratica della rapida crescita di $n!$, supponiamo che una squadra di 12 giocatori voglia fotografarsi su un podio con dodici posti in tutte le disposizioni possibili. Se anche per cambiare di posto e scattare si impiegasse un solo minuto, la procedura durerebbe $12! = 479.001.600$ minuti, cioè 911 anni e alcuni mesi!

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

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

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

Applicando questa formula si vede che i possibili risultati del mescolamento di un mazzo di carte napoletane (40 carte) sono $40! \approx 10^{46}$. Anche se una popolazione di 10 miliardi giocasse perennemente a briscola per un periodo dell’ordine dell’età dell’universo (10 miliardi di anni), impiegando un quarto d’ora per ogni partita, si realizzerebbero non più di 10^{24} partite, solo una frazione infinitesima di tutte le possibilità.

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. Una combinazione di classe k di n oggetti è un sottoinsieme dell’insieme di n oggetti costituito da k elementi.

Sia $C_{n;k}$ il numero di tali combinazioni. E’ ovvio che se $k = n$ prendiamo tutti gli elementi dell’insieme, e c’è quindi una sola scelta: $C_{n;n} = 1$. Se invece $k = 1$ le scelte saranno tante quante sono gli oggetti a disposizione: $C_{n;1} = n$.

Si potrebbe proseguire in questo modo, ma per far prima 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 $k!$ modi possibili. Quindi $D_{n;k} = k!C_{n;k}$, ovvero

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

Il simbolo $\binom{n}{k}$ è detto ”coefficiente binomiale”, perchè compare nella formula dello sviluppo del binomio:

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

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

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

Esempio 1. Si lancia una moneta cinque volte. Calcolare quanti sono i risultati possibili in cui ”testa” (T) appare esattamente tre volte.

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

Osservazione. Dalla (1.2.4) si vede facilmente che per ogni scelta di n e $0 \leq k \leq n$ si ha $\binom{n}{k} = \binom{n}{n-k}$. Infatti identificando un sottoinsieme di k oggetti tra n si identifica anche il sottoinsieme complementare di $n - k$ oggetti.

Esempio 2. Il gioco del lotto consiste nell'estrazione di cinque numeri tra 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 è quindi $\binom{90}{5} = \frac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}{5 \cdot 4 \cdot 3 \cdot 2}$.

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

Si tratta di collocare in k caselle n oggetti, che possono essere ripetuti, cioè un oggetto può essere presente in più caselle. Grazie alle ripetizioni, può ben essere $k > n$, nel qual caso le ripetizioni sono inevitabili. Per l'intuizione, dato che gli oggetti in generale non si moltiplicano, è forse meglio parlare di simboli, piuttosto che di oggetti.

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

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

$$\tilde{D}_{n;k} = n^k. \quad (1.2.6)$$

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

Esempio 3. Se si lancia una moneta cinque volte, il risultato è una successione di cinque simboli, ciascuno dei quali può 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 caselle: il numero dei possibili modi di farlo è quindi $3^{13} = 1.594.323$.

Applicando la probabilità classica si vede quindi che giocando una colonna a caso la probabilità di "fare 13" è $1/3^{13}$.

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

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

Si tratta del numero di tutti i sottoinsiemi, inclusi Ω e \emptyset . Per ogni $k \leq n$ vi sono, per la (1.2.4), $\binom{n}{k}$ sottoinsiemi di k elementi. Quindi il numero totale è, per la (1.2.5),

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

Esempio 5. Qual'è la probabilità di indovinare esattamente 10 risultati compilando a caso una colonna del totocalcio?

Si tratta di calcolare il numero delle possibili colonne con 10 simboli giusti (eguali a quelli della colonna vincente) 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 può sbagliare in due modi, il numero cercato è $2^3 \binom{13}{10} = 2288$.

La probabilità di fare 10 giocando una colonna a caso è quindi pari al rapporto tra il numero dei casi favorevoli e il numero totale: $2288/3^{13} \approx 1/700$.

1.2.7. Scelta "a caso" di uno o più oggetti.

Per "scelta a caso" di un oggetto tra n oggetti di un certo insieme Ω si intende un procedimento per cui ogni oggetto di Ω può essere scelto indifferentemente. Si usa anche parlare di "estrazione a caso". Si tratta, come si vede, della probabilità uniforme su Ω . L'analisi statistica richiede spesso la ripetizione di un certo numero di scelte a caso.

Estrazioni con e senza restituzione. Se nelle successive estrazioni l'oggetto estratto viene ogni volta reinserito nell'insieme Ω si parla di "estrazioni con restituzione", mentre se non c'è reinserimento si parla di "estrazioni senza restituzione". Nel primo caso ogni estrazione avviene in condizioni identiche, e le probabilità saranno anche identiche, e quindi indipendenti dal risultato delle estrazioni precedenti, mentre nel secondo l'insieme da cui si estrae cambia ad ogni prova, e le probabilità delle estrazioni anche cambiano.

Per un riferimento concettuale si usano spesso gli "schemi di urne". Si immagina cioè che ci sia un contenitore ("urna") da cui si estraggono gli oggetti. Se, per esempio, abbiamo un'urna con 7 sfere, ed estraiamo due volte una sfera, con restituzione, lo spazio degli eventi Ω_2 ha $7 \cdot 7 = 49$ elementi, mentre se effettuiamo due estrazioni senza restituzione avremo uno spazio $\bar{\Omega}_2$ con $7 \cdot 6 = 42$ elementi.

Se nel caso precedente 3 delle sfere sono nere e 4 bianche, la probabilità di estrarre due nere nel caso con restituzione è $9/49 \approx 0,1836$, mentre nel caso senza restituzione è $3 \cdot 2/42 = 1/7 \approx 0,14285 \dots$. Nel caso senza restituzione è minore, perchè la probabilità di estrarre una seconda sfera nera è diminuita dal fatto che è stata eliminata una nera.

Se aumentiamo il numero degli oggetti estraibili, per esempio se abbiamo un'urna di 700 sfere, di cui 300 nere e 400 bianche, la probabilità di avere due nere in due estrazioni con restituzione è sempre $\frac{9 \cdot 10^4}{49 \cdot 10^4} = \frac{9}{49}$, mentre la probabilità dell'evento senza restituzione è $\frac{3 \cdot 299}{7 \cdot 699} \approx 0,1833$, ed è quindi molto più vicina a $9/49$.

Come si capisce dall'ultimo esempio, se le estrazioni sono poche rispetto al numero degli oggetti, le probabilità delle estrazioni successive con o senza restituzione sono simili.

Esercizi.

1. Si lancia una moneta quattro volte. Qual'è la probabilità di avere due T e due C ?
2. Un'urna contiene quattro sfere nere e sei bianche. Si estraggono contemporaneamente due sfere. Qual'è la probabilità che siano entrambe bianche?
3. Qual'è la probabilità che, compilando una colonna del totocalcio a caso, non si azzechi neppure una risultato?

1.3. PROBABILITA' DISCRETA.

1.3.1. Probabilità su uno spazio discreto.

Estenderemo ora la nozione di probabilità ad un qualunque spazio degli eventi discreto, cioè finito oppure infinito numerabile.

L'estensione della nozione di probabilità al di là dello schema classico, cioè della probabilità uniforme, si presenta in modo del tutto naturale.

Consideriamo, per esempio, il caso del lancio di due monete (esempio 1 del §1.1). Sia N_T il numero di "teste" (T) ottenuto nei due lanci della moneta. Lo spazio degli eventi è $\Omega = \{T, C\}^2 = \{TT, TC, CT, CC\}$, e N_T può prendere i valori 0, 1, 2, con probabilità

$$P(\{N_T = 0\}) = \frac{1}{4}, \quad P(\{N_T = 1\}) = \frac{1}{2}, \quad P(\{N_T = 2\}) = \frac{1}{4}. \quad (1.3.1)$$

Infatti Ω ha un solo elemento con $N_T = 0$ e con $N_T = 2$ e ne ha due con $N_T = 1$.

Se ci interessa solo il valore di N_T possiamo considerare la probabilità (1.3.1) dei valori assunti da N_T . N_T è chiaramente una funzione su Ω a valori numerici, e in terminologia matematica lo spazio dei valori assunti si chiama il **codominio** od anche **immagine** di N_T e si indica con $N_T(\Omega)$. E' uno spazio con tre soli elementi $N_T(\Omega) = \{0, 1, 2\}$ e la (1.3.1) è una probabilità su $N_T(\Omega)$, detta **probabilità indotta** da N_T .

La probabilità indotta da N_T su $N_T(\Omega)$ non assegna però ai tre elementi probabilità eguale, e siamo quindi usciti dallo schema classico.

La definizione generale è la seguente.

Uno **spazio di probabilità discreto** è dato da una coppia (Ω, P) , costituita da un insieme finito o infinito numerabile Ω ("spazio degli eventi"), e da una funzione non negativa P ("probabilità") sui sottoinsiemi di Ω ("eventi").

La probabilità di un evento $A \subset \Omega$ è data dalla somma delle probabilità degli elementi ("eventi elementari") che lo costituiscono:

$$P(A) = \sum_{\omega \in A} p(\omega). \quad (1.3.2)$$

La funzione $p(\omega)$ che dà la probabilità degli eventi elementari (detta anche "densità discreta") è una funzione su Ω con le seguenti proprietà:

- i) non può essere negativa, cioè $p(\omega) \geq 0$, per ogni $\omega \in \Omega$;
- ii) è tale che $\sum_{\omega \in \Omega} p(\omega) = 1$.

Se Ω è infinito numerabile la (1.3.2) può rappresentare una somma infinita, o "serie", che dà però un risultato finito, anzi minore, o al più eguale, ad 1. Quindi le serie che intervengono nel calcolo delle probabilità per uno spazio degli eventi infinito numerabile sono sempre convergenti.

Dalle proprietà i) e ii) di $p(\omega)$ e dalla (1.3.2) segue che non può mai essere $P(A) > 1$, per nessun evento A , perchè $P(A)$ è data dalla somma delle probabilità $p(\omega)$ degli elementi di A , che non può mai superare la somma su tutti gli elementi di Ω .

Nota. Ricordiamo che usiamo la p minuscola per la probabilità degli eventi elementari, che è una funzione su Ω per distinguerla dalla probabilità degli eventi, che è invece

una funzione sui sottoinsiemi di Ω . Ovviamente gli eventi elementari sono anche eventi, e quindi si potrà usare anche la maiuscola.

Oltre alla probabilità discreta, assegnata su uno spazio degli eventi Ω discreto, vedremo in seguito anche la **probabilità continua** assegnata su spazi degli eventi infiniti che hanno la potenza del continuo, e quindi non sono numerabili, come sono ad esempio la retta reale \mathbb{R} o i suoi segmenti.

1.3.2. Proprietà fondamentali della probabilità.

Le proprietà della probabilità sono già 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 più eventi. Nel caso di tre eventi, se A, B, C sono incompatibili, cioè $A \cap B = A \cap C = B \cap C = \emptyset$, poniamo $B' = B \cup C$. Allora $A \cup B \cup C = A \cup B'$ e $A \cap B' = \emptyset$. Applicando la iii) due volte otteniamo $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

Iterando la procedura si arriva ad una simile conclusione per un qualunque numero di eventi incompatibili A_1, A_2, \dots, A_n , tali cioè che $A_i \cap A_j = \emptyset$ se $i \neq j$:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

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

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) = P(A) + P(B) - P(A \cap B). \quad (1.3.3)$$

Infatti $A \cup B$ è l'unione di tre eventi incompatibili $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$, quindi, applicando la proprietà vista sopra si ottiene la prima eguaglianza (1.3.3). Per la seconda si osserva che $A = (A \setminus B) \cup (A \cap B)$ e quindi $P(A) = P(A \cap B) + P(A \setminus B)$ e analogamente per B .

Intuitivamente per la (1.3.3) si può considerare che nella somma $P(A) + P(B)$ i punti di $A \cap B$ sono contati due volte, e quindi bisogna sottrarre $P(A \cap B)$.

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

Sia $B_1 = \{\omega_1 \text{ multiplo di } 3\}$, e $B_2 = \{\omega_2 \text{ multiplo di } 3\}$. Si ha $B = B_1 \cup B_2$, perchè se $\omega_1 \cdot \omega_2$ è multiplo di 3 almeno uno dei due numeri ω_1, ω_2 deve esserlo. Per la (1.3.3), 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 $P(B) = \frac{2}{3} - \frac{1}{9} = \frac{5}{9}$.

Esempio 2. Da un'urna con cinque sfere nere e quattro bianche si effettuano due estrazioni con restituzione. Si vuol calcolare la probabilità degli eventi $A = \{\text{le sfere sono dello stesso colore}\}$, $B = \{\text{tra le due ce n'è almeno una bianca}\}$, nonchè di A^c e di $A \cup B$.

Abbiamo $A = A_n \cup A_b$, $A_n = \{\text{le due sfere sono nere}\}$, $A_b = \{\text{le due sfere sono bianche}\}$. Si trova $P(A) = P(A_n) + P(A_b)$ con $P(A_n) = \frac{5 \cdot 5}{9 \cdot 9} = (\frac{5}{9})^2$ e similmente $P(A_b) = (\frac{4}{9})^2$. Inoltre $B^c = A_n$, e $A \cup B = \Omega$, per cui $P(B) = 1 - P(A_n)$ e $P(A \cup B) = 1$.

Esercizi.

1. Nel lancio di due dadi dell'esempio 1, si considerino gli eventi $A = \{\omega_1 + \omega_2 \text{ dispari}\}$ e $C = \{\omega_2 \text{ pari}\}$. Si calcolino le probabilità degli eventi $A, C, A \cup C, A \cap C, (A \cap C)^c$, e $C \cap B$, dove B è l'evento considerato nell'esempio 1.

2. Si calcolino le probabilità degli eventi A, B, B^c e $A \cup B$ dell'esempio 2 nel caso di estrazioni senza restituzione.

3. Da un mazzo di carte napoletane si estrae una carta a caso. Si considerino gli eventi $A = \{\text{è denari}\}$, $B = \{\text{è } \geq 4\}$, $C = \{\text{è una figura}\}$. Calcolare le probabilità degli eventi A, B, C e degli eventi $A \cup B$ e $B \cap C$.

1.4. PROBABILITÀ CONDIZIONATA. INDIPENDENZA.

1.4.1. Probabilità condizionata o condizionale.

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

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

Si può dire anche che $P(A|B)$ è la probabilità di A assumendo che "accade B ".

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

$$P_B(A) := P(A|B) = \sum_{\omega \in A \cap B} \frac{p(\omega)}{P(B)} = \sum_{\omega \in A} p_B(\omega), \quad (1.4.1b)$$

dove si è posto $p_B(\omega) = p(\omega)/P(B)$ per $\omega \in B$, e $p_B(\omega) = 0$ per $\omega \notin B$.

Come si vede facilmente applicando la definizione si ha $P_B(B) = 1$, cioè per la probabilità condizionata l'evento B , come è ovvio, diviene certo.

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

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

Esempio 1. Si lancia un dado e sia $\omega \in \{1, \dots, 6\}$ 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 $P(A|B) > P(A) = \frac{1}{2}$, cioè, come è ovvio, B favorisce A .

Nota. Storicamente la probabilità condizionata è nata dal problema di dividere la posta, in caso di interruzione di un gioco, tenendo conto dei risultati ottenuti, come illustra il seguente esempio.

Esempio 2. Due giocatori, a e b , lanciano a turno una moneta per due volte: chi ottiene più "T" incassa tutta la posta, e in caso di pareggio si divide a metà.

Supponiamo che al primo turno di lancio a ottiene "testa" e b "croce", e poi il gioco si interrompe (p.es., la moneta cade in un tombino). Come si deve dividere la posta?

Si tratta di dividerla in base alle probabilità di vincita condizionate al risultato del primo turno già effettuato. Lo spazio di probabilità è $\Omega = \{T, C\}^4$ (quattro lanci di moneta), e l'evento elementare si può scrivere $\omega = (\omega_1^{(a)}, \omega_2^{(a)}; \omega_1^{(b)}, \omega_2^{(b)})$. Il risultato ottenuto al primo turno diventa l'evento condizionante $E = \{\omega_1^{(a)} = T, \omega_1^{(b)} = C\}$.

Sotto la condizione E il giocatore b non può vincere, ma solo pareggiare. Detto $D = \{\text{pareggio}\}$ tale evento, e $A = \{\text{vince } a\}$ l'altro evento possibile, abbiamo $D \cap E = \{T, C; C, T\}$, e siccome $P(E) = \frac{1}{4}$, si ha $P(D|E) = \frac{1}{4}$, e pertanto $P(A|E) = \frac{3}{4}$.

Quindi tre quarti della posta vanno ad a e il rimanente quarto è diviso a metà, ovvero $7/8$ vanno ad a e $1/8$ a b .

Esempio 3. Calcoliamo le probabilità di vincita e di pareggio (non condizionate) per il gioco dell'esercizio precedente. Detti $N_T^{(a)}, N_T^{(b)}$ i numeri di teste realizzati dai due giocatori, la probabilità di pareggio è

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

Si ha $\{N_T^{(a)} = N_T^{(b)} = 0\} = (C, C; C, C)$, e similmente $\{N_T^{(a)} = N_T^{(b)} = 2\} = (T, T; T, T)$, quindi entrambi questi eventi hanno probabilità $1/16$, perchè $|\Omega| = 16$, mentre l'evento $\{N_T^{(a)} = N_T^{(b)} = 1\}$ è fatto di quattro eventi elementari perchè ciascun giocatore può realizzare T, C o C, T . Sommando si ottiene $P(D) = \frac{3}{8}$.

La probabilità che uno dei due giocatori vinca è perciò $1 - \frac{3}{8} = \frac{5}{8}$, e poichè i due sono sullo stesso piano, ciascuno ha probabilità $\frac{5}{16}$ di vincere.

1.4.2. Formula della probabilità totale e formula di Bayes.

Spesso lo spazio degli eventi Ω si divide in modo naturale in un certo numero di parti, gli eventi A_1, \dots, A_n . Le diverse parti non devono naturalmente avere punti in comune, cioè devono essere eventi disgiunte (incompatibili), ovvero $A_j \cap A_k = \emptyset$ se $j \neq k$, e coprire tutto lo spazio, cioè $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$. Se questo accade gli eventi $A_k, k = 1, \dots, n$ si dicono una "partizione" di Ω .

L'esempio più semplice di partizione è quello costituito da un evento A e dal suo complementare A^c .

Un esempio pratico di spazio di probabilità di questo tipo è quello di una popolazione costituita da un certo numero di sottopopolazioni da cui si estrae un individuo a caso.

Supponiamo che le probabilità $P(A_k), k = 1, \dots, n$, siano note, e si voglia calcolare la probabilità di un evento B , di cui sono note le probabilità condizionate $P(B|A_k)$, per ogni $k = 1, \dots, n$. La probabilità $P(B)$ si può allora esprimere tramite la "formula della probabilità totale" che ora passiamo a descrivere.

Dividiamo l'evento B in parti: per ogni $k = 1, \dots, n$ l'evento $B_k = B \cap A_k$ è la parte di B che sta nell'elemento A_k della partizione. Poichè $\cup_{k=1}^n A_k = \Omega$, si ha $B = \cup_{k=1}^n B_k$, e siccome i B_k sono disgiunti, perchè lo sono gli A_k , abbiamo

$$P(B) = \sum_{k=1}^n P(B_k) = \sum_{k=1}^n P(B \cap A_k).$$

Considerando che $P(B \cap A_k) = P(B|A_k)P(A_k), k = 1, \dots, n$, otteniamo la **formula della probabilità totale**:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n). \quad (1.4.2)$$

Esempio 4. In un negozio un cliente acquista a caso un frigorifero da una partita di 100, di cui 60 provengono da una fabbrica a e 40 da un'altra fabbrica b . Si sa che i

frigoriferi della fabbrica a sono difettosi al 10%, mentre quelli della fabbrica B lo sono al 40%. Si chiede quale sia la probabilità che il frigorifero acquistato sia difettoso.

Lo spazio degli eventi (frigoriferi) si divide in due parti, che indichiamo con A (quelli di provenienza a) e B (quelli di provenienza b). Si tratta di una partizione: $A \cup B = \Omega$. Se D è l'evento che ci interessa (frigorifero difettoso), sappiamo che $P(D|A) = 0,1$ e $P(D|B) = 0,4$, mentre $P(A) = 60/100 = 0,6$ e $P(B) = 0,4$. Quindi

$$P(D) = P(D|A)P(A) + P(D|B)P(B) = 0,06 + 0,16 = 0,22.$$

Oltre al calcolo della probabilità di un evento B tramite le probabilità condizionate $P(B|A_k)$ si può porre il problema, una volta realizzato l'evento B , di calcolare le probabilità condizionate $P(A_k|B)$, che sono dette le probabilità "a posteriori" delle componenti A_k della partizione di Ω . (Le probabilità "a priori" sono le $P(A_k)$.)

Per esempio, nel caso dell'esempio 4 visto sopra, una volta accertato che il frigorifero acquistato è difettoso, si può chiedere quale sia la probabilità che provenga dalla fabbrica b o dalla fabbrica a .

Dalla definizione di probabilità condizionata abbiamo

$$P(A_k|B) = P(A_k \cap B)/P(B) = P(B|A_k)P(A_k)/P(B).$$

Usando la formula della probabilità totale (1.4.2) otteniamo le probabilità a posteriori in termini delle probabilità a priori e delle probabilità condizionate $P(B|A_k)$, cioè la "formula di Bayes":

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

Esempio 5. Tornando all'esempio 4, vediamo che la probabilità che il frigorifero difettoso provenga dalla fabbrica a è $P(A|D) = P(D|A)P(A)/P(D) = 0,06/0,22 = 0,27$.

1.4.3 Eventi indipendenti.

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

$$P(A \cap B) = P(A) \cdot P(B). \quad (1.4.4)$$

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

Quindi, se A è indipendente da B , allora B è indipendente da A . La condizione di indipendenza tra eventi è simmetrica, e conviene prendere la relazione in forma 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 .

Basterà mostrarlo per A e B^c , perchè le altre relazioni si ottengono scambiando i ruoli di A e A^c e di B e B^c , e ricordando che per ogni evento A si ha $(A^c)^c = A$.

Poichè B, B^c sono una partizione di Ω , se A, B sono indipendenti troviamo

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

Quindi portando $P(A)P(B)$ al primo membro, e ricordando che $P(B^c) = 1 - P(B)$, si ha

$$P(A)(1 - P(B)) = P(A)P(B^c) = P(A \cap B^c).$$

Esempio 6. Torniamo 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 7. (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 è $N = (100)^2$, 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.

Indipendenza per più 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 generale, gli eventi A_1, A_2, \dots, A_n , si dicono indipendenti se, comunque se ne prendono due o più, A_{i_1}, \dots, A_{i_k} , $k \geq 2$, si ha

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k}). \quad (1.4.5)$$

Come mostra il seguente esempio non basta che gli eventi siano indipendenti a due a due per essere indipendenti (globalmente).

Esempio 8. 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 considerano 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 sono a due a due indipendenti, ma non sono indipendenti globalmente perchè $P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C)$.

Esercizi.

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 $P(A|B)$.

2. Tre bacini, a, b, c contengono 100 carpe ciascuno, e si sa che 30 carpe del bacino a, 30 del bacino b e 90 del bacino c possiedono un certo carattere genetico g . 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'è la probabilità che possieda il carattere g ?

3. 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'è la probabilità che sia bianca?

4. Riprendendo l'esercizio 2 di sopra, supponiamo noto che la carpa esaminata ha il carattere g . Qual'è la probabilità "a posteriori" che venga dal bacino c ?

5. Riprendendo l'esercizio 3 di sopra, si chiede qual'è la probabilità "a posteriori", noto che la sfera estratta è bianca, che essa provenga dall'urna b .

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

1.5. SUCCESSIONI DI PROVE INDIPENDENTI.

Le successioni di prove indipendenti, come vedremo più in dettaglio in seguito, sono un modello fondamentale della probabilità e della statistica. Abbiamo già visto esempi di prove indipendenti, come il lancio ripetuto di una moneta o di un dado, e le estrazioni successive, con restituzione, di un oggetto da un'urna. Esempi di maggior interesse pratico sono le estrazioni settimanali del lotto, o il prelievo di un campione da una popolazione.

L'indipendenza delle diverse prove è stabilita in genere sulla base del fatto che, per quanto se ne sa, esse avvengono nelle stesse condizioni, e quindi il risultato di ciascuna prova non influenza le altre.

Consideriamo il caso di due lanci di una moneta (esempio 6 del precedente paragrafo). Abbiamo visto che

$$P(\{\omega_1 = T\} \cap \{\omega_2 = C\}) = P(\{\omega_1 = T\})P(\{\omega_2 = C\}) = 1/4,$$

quindi gli eventi $\{\omega_1 = T\}$ e $\omega_2 = C\}$ sono indipendenti. L'indipendenza c'è, ovviamente, per ogni scelta dei risultati delle prove $\bar{\omega}_i \in \{T, C\}$, $i = 1, 2$ fissata *a priori*, cioè per ogni evento elementare, la cui probabilità è

$$p(\bar{\omega}_1, \bar{\omega}_2) = P(\{\omega_1 = \bar{\omega}_1\} \cap \{\omega_2 = \bar{\omega}_2\}) = P(\{\omega_1 = \bar{\omega}_1\})P(\{\omega_2 = \bar{\omega}_2\}) = 1/4. \quad (1.5.1)$$

In questa formula $\omega_i, i = 1, 2$, sono variabili mentre gli $\bar{\omega}_i$ si riferiscono alla scelta fissata. $\{\omega_1 = \bar{\omega}_1\} \cap \{\omega_2 = \bar{\omega}_2\}$ è un modo di scrivere l'evento elementare $(\bar{\omega}_1, \bar{\omega}_2)$ dello spazio degli eventi $\Omega = \{T, C\} \times \{T, C\}$. Quindi, se $\bar{p}(\omega)$ è la probabilità degli eventi elementari nello spazio di singolo lancio $\bar{\Omega} = \{T, C\}$ la (1.5.1) si scrive

$$p(\bar{\omega}_1, \bar{\omega}_2) = \bar{p}(\bar{\omega}_1)\bar{p}(\bar{\omega}_2). \quad (1.5.2)$$

Consideriamo in generale una successione di prove indipendenti costituite da una prova ripetuta più volte in condizioni di incertezza identiche. Sia $\bar{\Omega}$ lo spazio di singola prova e $\bar{p}(\omega)$, $\omega \in \bar{\Omega}$, la probabilità degli eventi elementari di $\bar{\Omega}$. Lo spazio degli eventi della prova ripetuta n volte è il prodotto cartesiano $\Omega = \bar{\Omega}^n$, e il generico evento elementare $\omega \in \Omega$, si può rappresentare come $\omega = (\omega_1, \omega_2, \dots, \omega_n)$, dove le variabili $\omega_k \in \bar{\Omega}$ rappresentano il risultato della k -esima prova per $k = 1, 2, \dots, n$

Se $p(\cdot)$ è la funzione che dà la probabilità degli eventi elementari in $\Omega = \bar{\Omega}^n$, per l'indipendenza delle prove, ripetendo il ragionamento che ha portato alla (1.5.2), otteniamo

$$p(\omega_1, \omega_2, \dots, \omega_n) = \bar{p}(\omega_1)\bar{p}(\omega_2) \dots \bar{p}(\omega_n). \quad (1.5.3a)$$

Più in generale le prove indipendenti possono essere diverse, con diversi spazi di singola prova $\Omega_1, \dots, \Omega_n$, e diverse probabilità degli eventi elementari $p^{(1)}, \dots, p^{(n)}$. Lo spazio delle n prove sarà $\Omega = \Omega_1 \times \dots \times \Omega_n$. Avremo ancora, per l'indipendenza, un probabilità prodotto su Ω :

$$p(\omega_1, \omega_2, \dots, \omega_n) = p^{(1)}(\omega_1)p^{(2)}(\omega_2) \dots p^{(n)}(\omega_n). \quad (1.5.3b)$$

In conclusione, si ha che **nel caso di n prove indipendenti, la probabilità sullo spazio di n prove è data dal prodotto delle probabilità di singola prova**, secondo la formula (1.5.3b), di cui la formula (1.5.3a) è un caso particolare.

Esempio 1. Si effettuano tre estrazioni con restituzione da un'urna con tre sfere nere e sette bianche. Si calcoli la probabilità che la successione delle estratte sia (b, n, b) (b indica sfera bianca e n sfera nera).

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

Osservazione. Dalla (1.5.3a,b) si può vedere, come è logico aspettarsi, che **eventi dipendenti da prove diverse sono indipendenti**, e le loro probabilità si possono calcolare come se le prove da cui gli eventi considerati non dipendono non ci fossero.

Chiariamo il significato dell'osservazione con un esempio.

Esempio 2. Consideriamo il lancio di quattro monete. Lo spazio degli eventi di singola prova è $\bar{\Omega} = \{T, C\}$ e lo spazio degli eventi per le quattro prove è $\Omega = \bar{\Omega}^4$. Sia $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)$ il generico evento elementare di Ω e consideriamo gli eventi $A = \{\text{c'è un solo } T \text{ tra i primi due risultati}\}, B = \{\omega_4 = T\}$.

Per l'osservazione precedente A e B sono indipendenti, perchè A dipende dalle prime due prove e B dipende dalla quarta. Verifichiamo questo fatto con un calcolo esplicito. La probabilità che ci sia un solo T in due lanci è $1/2$ (casi favorevoli TC, CT), e ovviamente $P(B) = 1/2$. Quindi concludiamo che $P(A \cap B) = P(A)P(B) = 1/4$.

Si noti che il risultato del terzo lancio, per l'indipendenza, non gioca alcun ruolo, è come se non ci fosse.

Possono però nascere dei dubbi sul procedimento precedente, per il fatto che gli eventi A e B sono stati introdotti come eventi dello spazio $\Omega = \bar{\Omega}^4$, mentre noi abbiamo calcolato le probabilità come se A fosse un evento di $\bar{\Omega}^2$ e B un evento di $\bar{\Omega}$.

In realtà A è logicamente distinto dal corrispondente evento nello spazio del lancio di due monete $\bar{\Omega}^2 = \{T, C\}^2$, il quale è $\mathcal{A} = \{TC, CT\}$, e lo stesso accade per B . Se vogliamo rappresentare A e B come eventi di $\Omega = \bar{\Omega}^4$ abbiamo

$$A = \{\omega \in \Omega : (\omega_1, \omega_2) \in \mathcal{A}\}, \quad B = \{\omega \in \Omega : \omega_4 = T\}$$

$$A \cap B = \{\omega \in \Omega : (\omega_1, \omega_2) \in \mathcal{A}, \omega_4 = T\}$$

Per l'indipendenza, che è espressa dalla formula (1.3.5a), abbiamo

$$\begin{aligned} P(A \cap B) &= \sum_{(\omega_1, \omega_2, \omega_3, \omega_4) \in A \cap B} \bar{p}(\omega_1) \bar{p}(\omega_2) \bar{p}(\omega_3) \bar{p}(\omega_4) = \\ &= \sum_{(\omega_1, \omega_2) \in \mathcal{A}} \bar{p}(\omega_1) \bar{p}(\omega_2) \sum_{\omega_3} \bar{p}(\omega_3) \bar{p}(T) = \frac{1}{2} \sum_{(\omega_1, \omega_2) \in \mathcal{A}} \bar{p}(\omega_1) \bar{p}(\omega_2) = \frac{1}{2} P(\mathcal{A}) = \frac{1}{4}. \end{aligned}$$

Si vede dunque che la terza prova sparisce perchè $\sum_{\omega_3} \bar{p}(\omega_3) = 1$ e le probabilità si calcolano negli spazi corrispondenti alle sole prove da cui dipendono.

Non è difficile generalizzare l'esempio precedente al caso generale.

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 è $\Omega = \{0, 1\}^n$, e supponiamo inoltre che in ciascuna prova sia $p(1) = p$, e quindi $p(0) = 1 - p$, con $p \in (0, 1)$.

L'esempio 1 appena visto rientra in questo schema, una volta stabilito quale dei due risultati (b o n) sia il "successo".

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

$$p(\omega_1, \dots, \omega_n) = p^{N(\omega)}(1 - p)^{n - N(\omega)}. \quad (1.5.4)$$

Se ora vogliamo calcolare la probabilità che il numero N di successi prenda un valore fissato k , dovremo sommare le probabilità di tutti gli eventi elementari per i quali $N(\omega) = k$. Per la (1.5.4) la probabilità di ogni tale ω è $p^k(1 - p)^{n - k}$. Rimane quindi solo da 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

$$P(\{N = k\}) = \binom{n}{k} p^k (1 - p)^{n - k}, \quad k = 0, 1, \dots, n. \quad (1.5.5)$$

Osservazione: evento più probabile. Le probabilità (1.5.5) variano con k , e ci si può domandare qual'è l'evento più probabile, cioè qual'è il numero k di "successi" che ha la massima probabilità.

Supponendo che $p \leq \frac{1}{2}$, il minimo si ha per $k = 0$, ed è $P(\{N = 0\}) = p^n$ (se è $p \geq \frac{1}{2}$ si ha per $k = n$). La condizione $P(\{N = k\}) \geq P(\{N = k - 1\})$, implica, come si vede facilmente dalla (1.5.5) $p \binom{n}{k} \geq q \binom{n}{k - 1}$, che, semplificando i fattoriali, si riduce alla condizione $\frac{k}{n + 1} \leq p$.

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

Esempio 2. 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 probabilità degli eventi $A = \{\text{i 5 catturati sono della stessa razza}\}$, e $B = \{\text{almeno uno dei 5 catturati è di razza b}\}$.

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

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

Esempio 3 (Tempo di attesa di un giocatore). Consideriamo un giocatore che giochi ai dadi, alla roulette, al lotto, o altro gioco d'azzardo con prove ripetute identiche e indipendenti, e supponiamo che punti su un evento di probabilità $p > 0$, ripetendo lo stesso gioco in caso di insuccesso fino alla prima vincita. Qual'è la probabilità che non vinca mai?

In pratica stiamo considerando una serie di prove con due possibili risultati: la vincita del giocatore (“successo”), che ha probabilità $1/6$ e l’“insuccesso”, che ha probabilità $q = 1 - p$. Come spazio degli eventi abbiamo $\{0, 1\}^n$, dove 1 indica successo, e n dovrà essere arbitrariamente grande.

Conviene prendere direttamente $n = \infty$ e considerare lo spazio $\Omega_\infty = \{0, 1\}^\infty$ delle successioni infinite dei simboli 0 e 1. Un evento elementare $\omega \in \Omega_\infty$ sarà una successione di risultati del gioco che rappresentiamo come $\omega = (\omega_1, \omega_2, \dots)$ dove $\omega_j \in \{0, 1\}$: $\omega_j = 1$ indica successo al turno j .

Il numero di turni che deve attendere il giocatore fino alla prima vincita è la funzione

$$\tau(\omega) = \min\{k : \omega_k = 1\}.$$

L’evento $\{\tau = k\}$ corrisponde a $k-1$ insuccessi seguiti da un successo, cioè $\{\tau = k\} = \{\omega_1 = 0, \dots, \omega_{k-1} = 0, \omega_k = 1\}$. Poichè si ha indipendenza e quindi la probabilità prodotto, otteniamo dalla formula (1.5.3)

$$P(\tau = k) = p q^{k-1}, \quad k = 1, 2, 3, \dots \quad (1.5.6)$$

$\{\tau > K\}$ vuol dire che al turno K il giocatore non ha ancora vinto, e quindi $\{\tau > K\} = \{\omega_1 = \omega_2 = \dots = \omega_K = 0\}$, per cui abbiamo

$$P(\tau > K) = q^K, \quad K = 1, 2, 3, \dots \quad (1.5.7)$$

Se il giocatore non vince mai vuol dire che $\tau > K$ per ogni K , non importa quanto grande, e quindi, poichè $0 < q < 1$, la probabilità di non vincere mai è

$$\lim_{K \rightarrow \infty} P(\tau > K) = \lim_{K \rightarrow \infty} q^K = 0.$$

Calcolando la probabilità totale abbiamo $\sum_{k=1}^K P(\tau = k) + P(\tau > K) = 1$, e poichè $P(\tau > K) \rightarrow 0$, per $K \rightarrow \infty$, la serie dei termini (1.5.6) converge ad 1

Il risultato si può interpretare dicendo che *in una successione di prove identiche indipendenti ogni evento di probabilità p comunque piccola ma positiva, accade con certezza prima o poi.*

Con l’esempio precedente abbiamo ottenuto uno spazio di probabilità discreto ma infinito, con spazio degli stati dato dai numeri naturali $\mathbb{N} = \{1, 2, \dots\}$, mentre la probabilità di ogni elemento k è data dalla (1.5.6).

Esercizi.

1. Si lancia tre volte una moneta. Qual’è la probabilità che venga due volte di seguito T ?

2. Si piantano due semi, uno di germinabilità (=probabilità di germinare) $p_a = 0,9$ e l’altro di germinabilità $p_b = 0,6$. Qual’è la probabilità che non germini nessuno dei due? (Si assume indipendenza delle due prove.)

3. Da un’urna con quattro sfere nere e sei bianche se ne estrae una a caso per tre volte con restituzione. Calcolare la probabilità dei seguenti eventi $A = \{\text{le sfere estratte sono tutte nere}\}$, $B = \{\text{si estrae almeno una sfera nera}\}$, $C = \{\text{si estraggono due sfere bianche e una nera}\}$.

1.6. VARIABILI ALEATORIE.

1.6.1. Variabili aleatorie e loro distribuzioni.

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

Di variabili aleatorie ne abbiamo già viste molte: 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 , spesso denotato con $F(\Omega)$, che è un insieme di numeri reali. Nel nostro contesto preferiamo la notazione Ω_F a $F(\Omega)$, per non fare confusione con i valori della F :

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

E’ chiaro che per ogni $x \in \Omega_F$ si trova almeno un $\omega \in \Omega$ tale che $F(\omega) = x$.

La probabilità P su Ω induce quindi una probabilità P_F su Ω_F : l’evento elementare $x \in \Omega_F$ ha probabilità

$$p_F(x) = P(\{F(\omega) = x\}) = P(F^{-1}(x)) = \sum_{\omega \in F^{-1}(x)} p(\omega). \quad (1.6.1)$$

L’evento (sottoinsieme di Ω) $F^{-1}(x) = \{\omega \in \Omega : F(\omega) = x\}$, che per brevità si denota anche $\{F(\omega) = x\}$, è l’immagine inversa di x , cioè l’insieme dei punti $\omega \in \Omega$ in cui F assume il dato valore x . Le immagini inverse di due punti diversi $x_1 \neq x_2$ non hanno elementi in comune perchè per nessun $\omega \in \Omega$ può accadere che $F(\omega) = x_1$ e $F(\omega) = x_2$. Inoltre l’unione, per tutti i possibili $x \in \Omega_F$, degli eventi $F^{-1}(x)$ è l’intero spazio Ω .

Quindi gli eventi $F^{-1}(x)$, per $x \in \Omega_F$, costituiscono una partizione, e sommando la (1.6.1) su tutti gli x si ottiene la probabilità totale:

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

La funzione p_F su Ω_F definita dalla (1.6.1) è una densità discreta di probabilità, e dunque (Ω_F, P_F) è un nuovo spazio di probabilità.

Definizione. Lo spazio di probabilità (Ω_F, P_F) prende il nome di **distribuzione** della variabile aleatoria F .

La distribuzione di una variabile aleatoria F è dunque una probabilità. Gli eventi elementari sono i valori assunti da F e le loro probabilità sono date dalla (1.6.1).

Osservazione 1. Se $F^{-1}(x)$ contiene, per ogni $x \in \Omega_F$, sempre di un solo punto di Ω , allora si dice che F è “invertibile”, e F^{-1} è una funzione da Ω_F a Ω , detta “funzione inversa”. Il nuovo spazio di probabilità è in pratica identico al precedente: ad ogni punto di Ω corrisponde un solo punto di Ω_F con la stessa probabilità.

Esempio 1. Consideriamo il lancio di un dado, con $\Omega = \{1, 2, \dots, 6\}$ e la probabilità classica. Si consideri la variabile aleatoria definita dalle relazioni $F(\omega) = 1$ se ω è dispari e $F(\omega) = 0$ se ω è pari. Si ha $\Omega_F = \{0, 1\}$ e inoltre $F^{-1}(1) = \{1, 3, 5\}$ e $F^{-1}(0) = \{2, 4, 6\}$.

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 sia $F(T) = 1$ e $F(C) = 0$. E' chiaro che Ω_F e p_F sono gli stessi dell'esempio precedente.

Si noti che in questo caso F è invertibile, e la nuova probabilità differisce dalla vecchia solo perchè i due punti sono denotati 1, 0 invece che T, C .

Osservazione 2. Gli esempi 1 e 2 appena visti mostrano un caso di due variabili aleatorie definite su spazi di probabilità diversi che hanno la stessa distribuzione. In realtà quando si parla di variabile aleatoria ci si riferisce spesso alla sola distribuzione, e non ad una realizzazione in uno spazio di probabilità definito. 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.

Distribuzione binomiale. Il numero di successi N in n prove indipendenti considerato nel §1.4.4 è una variabile aleatoria e la sua distribuzione è data dalla (1.5.5).

Tale distribuzione prende il nome di **distribuzione binomiale** con n prove e probabilità di successo p . Una variabile aleatoria con questa distribuzione è detta **variabile aleatoria binomiale** con n prove e probabilità di successo p .

La distribuzione binomiale con n prove con probabilità di successo p è indicata con il simbolo $B(n, p)$, e questo simbolo, per l'osservazione 2, può anche indicare la totalità (la classe di equivalenza) delle variabili aleatorie con quella distribuzione.

Esempio 3. Da un'urna con sei sfere bianche e tre nere si effettuano tre estrazioni con restituzione, registrando il colore della sfera estratta. Si consideri la variabile aleatoria $F(\omega) = N_b(\omega) - N_n(\omega)$, dove $N_b(N_n)$ è il numero delle estrazioni che danno una sfera bianca (nera). Vogliamo trovare la distribuzione delle variabili aleatorie F e $G = F^2$.

Si ha $\Omega_F = F(\Omega) = \{-3, -1, 1, 3\}$, e $\Omega_G = G(\Omega) = \{1, 9\}$. Inoltre poichè $N_b + N_n = 3$ abbiamo $F = 2N_b - 3$ e quindi

$$P(F = 3) = \frac{8}{27}, \quad P(F = 1) = \frac{4}{9}, \quad P(F = -1) = \frac{2}{9}, \quad P(F = -3) = \frac{1}{27}.$$

Infine, dalla relazione $G = F^2$ troviamo

$$P(G = 1) = \frac{2}{3}, \quad P(G = 9) = \frac{1}{3}.$$

1.6.2. Valor medio di una variabile aleatoria.

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

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

$$M(F) = \sum_{\omega \in \Omega} F(\omega)p(\omega) = \sum_{x \in \Omega_F} x p_F(x). \quad (1.6.2)$$

Questa formula consta in realtà di due diverse formule, che interpretano in modo diverso l'espressione "media dei valori di F pesata con le probabilità." Infatti nella formula a

sinistra si tratta della probabilità nello spazio di partenza (Ω, P) e nella formula a destra, invece, della distribuzione di F , sullo spazio di arrivo (Ω_F, P_F) .

Per vedere che le due formule danno lo stesso risultato partiamo dalla (1.6.1) che dà la probabilità degli eventi elementari della distribuzione di F . Ricordando che gli eventi $F^{-1}(x)$, per $x \in \Omega_F$ costituiscono una partizione di Ω , possiamo eseguire la somma nella (1.6.2) sommando prima sugli ω che appartengono a ciascun elemento della partizione e poi sugli elementi della partizione:

$$M(F) = \sum_{x \in \Omega_F} x p_F(x) = \sum_{x \in \Omega_F} \sum_{\omega \in F^{-1}(x)} F(\omega) p(\omega) = \sum_{\omega \in \Omega} F(\omega) p(\omega). \quad (1.6.3)$$

Osservazione 3. Se F assume infiniti valori, e quindi Ω_F ha infiniti elementi, il valor medio $M(F)$ nella (1.6.2) diventa una serie, che può non convergere. Se la serie dei valori assoluti $\sum_{x \in \Omega_F} |x| p_F(x)$ non converge si dice che il valor medio $M(F)$ non esiste.

Un esempio di variabile aleatoria che non ha valor medio si ricava dall'esempio 3 del paragrafo §1.5. 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 è un parametro numerico che corrisponde al "centro" dei valori di F , o meglio ad una specie di "baricentro" di tali valori, ciascuno dei quali ha un "peso" pari alla sua probabilità.

Proprietà 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 è la somma dei loro valori medi, e lo stesso avviene naturalmente per tre o più variabili.

ii) Moltiplicazione di una variabile aleatoria per un numero. Se k è un numero reale, si vede immediatamente dalla (1.6.2) che $M(kF) = kM(F)$.

iii) Variabile aleatoria degenera. Se $F(\omega)$ è una costante, cioè per qualche numero c si ha $F(\omega) = c$ per ogni $\omega \in \Omega$, allora $M(F) = c$. Lo stesso accade se l'eguaglianza $F(\omega) = c$ vale tranne che per i punti $\omega \in A$ dove A è un evento di probabilità nulla: $P(A) = 0$. In tal caso si dice che F è una costante "quasi ovunque".

Una variabile aleatoria costante, ovunque o quasi ovunque, è detta "variabile aleatoria degenera".

Esempio 4. Si lancia nove volte un dado, e sia N il numero di lanci che dà 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 dà 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 $M(N) = M(N_1) + \dots + M(N_9) = \frac{9}{3} = 3$.

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 è sempre compreso tra il minimo e il massimo di F :

$$\min_{\omega} F(\omega) \leq M(F) \leq \max_{\omega} F(\omega). \quad (1.6.4)$$

Scarto dalla media. Se F è una variabile aleatoria, la nuova variabile aleatoria $\widehat{F} = F - M(F)$ è detta "scarto dalla media", o anche "variabile centrata" corrispondente ad F . Poichè $M(F)$ è una costante, il valor medio di \widehat{F} è $M(\widehat{F}) = M(F - M(F)) = M(F) - M(F) = 0$.

Il nome "variabile centrata" deriva dal fatto che il suo "centro" (la media) è lo zero.

Disuguaglianza di Chebyshev. Se F è non negativa, cioè $F(\omega) \geq 0$ per ogni $\omega \in \Omega$, vale la disuguaglianza di Chebyshev, che è una stima della probabilità che F assuma grandi valori: per ogni $a > 0$ si ha

$$P(F > a) \leq \frac{M(F)}{a}. \quad (1.6.5)$$

La dimostrazione è molto semplice. Abbiamo

$$M(F) = \sum_{\omega \in \Omega} p(\omega)F(\omega) \geq \sum_{\omega: F(\omega) > a} p(\omega)F(\omega),$$

perchè nella somma a destra abbiamo ommesso gli ω per cui $F(\omega) \leq a$, e quindi, essendo sempre $F \geq 0$, la somma a destra non può essere più grande di quella a sinistra. Inoltre nella somma a destra $F(\omega) > a$ per cui

$$\sum_{\omega: F(\omega) > a} p(\omega)F(\omega) \geq a \sum_{\omega: F(\omega) > a} p(\omega) = a P(F > a).$$

In conclusione quindi abbiamo $M(F) \geq aP(F > a)$, cioè la disuguaglianza (1.6.5).

1.6.3. Varianza o dispersione di una variabile aleatoria.

Si dice "varianza" o "dispersione" di una variabile aleatoria F , definita sullo spazio di probabilità (Ω, P) , la quantità

$$\text{Var}(F) = M((F - M(F))^2) = M(\widehat{F}^2), \quad (1.6.6a)$$

dove \widehat{F} è lo scarto dalla media, o variabile centrata relativa ad F definita sopra. Sviluppando il quadrato si giunge ad una nuova espressione per la varianza:

$$\begin{aligned} \text{Var}(F) &= M(F^2 - 2FM(F) + (M(F))^2) = M(F^2) - 2(M(F))^2 + (M(F))^2 \\ &= M(F^2) - (M(F))^2. \end{aligned} \quad (1.6.6b)$$

Nel calcolo si è usato il fatto che $M(F M(F)) = (M(F))^2$ perchè $M(F)$ è una costante .

Dalla (1.6.6b) si vede che il valor medio del quadrato di una variabile aleatoria è sempre maggiore od eguale del quadrato del valor medio: $M(F^2) \geq (M(F))^2$. L'eguaglianza vale se e solo se la varianza è nulla, e la variabile è quindi degenera.

Proprietà della varianza.

i) Positività. Si ha $\text{Var}(F) \geq 0$, e $\text{Var}(F) = 0$ se e solo se F è degenera.

Infatti per la (1.6.6a) $\text{Var}(F) = \sum_{\omega} p(\omega) \widehat{F}^2(\omega)$. 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 probabilità nulla, e dunque, secondo la definizione data sopra, F è una variabile aleatoria degenera.

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

Infatti per le proprietà 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 (1.6.6a).

Osservazione 4. La trasformazione lineare $G = \alpha F + \beta$ corrisponde da un cambiamento di scala. Possiamo pensare che F e G sono le misure della stessa grandezza con diverse scale di misura. L'origine $F = 0$ corrisponde al punto $G = \beta$ e l'unità di misura che dà il risultato F è $|\alpha|$ volte maggiore di quella che dà il risultato G . Le due scale hanno lo stesso verso se $\alpha > 0$, cioè F e G crescono e calano insieme, mentre hanno verso opposto per $\alpha < 0$, cioè se F cresce G cala e viceversa.

Un esempio di cambiamento di scala di questo tipo è quello tra gradi Celsius e gradi Fahrenheit.

Osservazione 5. Dalla proprietà ii) si vede che aggiungendo ad una variabile aleatoria una costante la varianza non cambia (mentre cambia il valore medio). Quindi la varianza è indifferente al centro della distribuzione, ma misura piuttosto quanto la distribuzione è "sparpagliata" o "dispersa" attorno al valor medio.

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

$$\sigma_F = \sqrt{\text{Var}(F)}. \quad (1.6.7)$$

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

1.6.4. Variabili aleatorie discrete notevoli.

Descriviamo brevemente alcune variabili aleatorie discrete di particolare importanza. In questo paragrafo, come accennato in precedenza, identifichiamo le variabili aleatorie con le loro distribuzioni.

1. Variabile binomiale.

La distribuzione di una variabile aleatoria binomiale relativa ad $n = 1, 2, \dots$ prove e con probabilità di "successo" $p \in (0, 1)$ in ciascuna prova, cioè di una variabile aleatoria nella classe $B(n; p)$, è data dalla (1.5.5).

Vogliamo ora calcolarne il valor medio. Come al solito, consideriamo la variabile binomiale definita nello spazio $\Omega = \{0, 1\}^n$, con probabilità di "successo" (rappresentato da

1) data da $p \in (0, 1)$. Il numero dei successi N è distribuito secondo $B(n; p)$, e indicando il generico risultato $\omega \in \Omega$ delle n prove con $\omega = (\omega_1, \omega_2, \dots, \omega_n)$, si ha $N = \omega_1 + \omega_2 + \dots + \omega_n$. Ciascun ω_i , $i = 1, \dots, n$, è una variabile aleatoria che prende due valori $\omega_i \in \{0, 1\}$, e il suo valor medio è

$$M(\omega_i) = 1P(\omega_i = 1) + 0P(\omega_i = 0) = p.$$

Ne segue $M(N) = \sum_{i=1}^n M(\omega_i) = n p$.

Per la varianza di N rimandiamo al prossimo paragrafo.

2. Variabile aleatoria poissoniana. La variabile aleatoria poissoniana, o di Poisson, è il limite di una successione di variabili binomiali quando n cresce all'infinito, e p va a zero con n , in modo che il valor medio np tende ad un limite ρ . Quindi è una variabile che si usa con buona approssimazione quando n è molto grande e p molto piccolo.

Supponiamo di avere una successione di binomiali con n prove e probabilità di successo p_n , tali che, posto $\rho_n = np_n$, si ha $\lim_{n \rightarrow \infty} \rho_n \rightarrow \rho > 0$. Vogliamo considerare il limite della distribuzione per $n \rightarrow \infty$.

Fissiamo un $k = 0, 1, \dots$. Per ogni $n \geq k$, dato che $p_n = \frac{\rho_n}{n}$, la formula (1.5.5) ci dà

$$P(N_n = k) = \binom{n}{k} \left(\frac{\rho_n}{n}\right)^k \left(1 - \frac{\rho_n}{n}\right)^{n-k} = \frac{\rho_n^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k \left(1 - \frac{\rho_n}{n}\right)^k} \left(1 - \frac{\rho_n}{n}\right)^n.$$

Poichè k è fisso, e $\rho_n \rightarrow \rho$, per $n \rightarrow \infty$, abbiamo

$$\frac{n(n-1)\dots(n-k+1)}{n^k \left(1 - \frac{\rho_n}{n}\right)^k} = \frac{n(n-1)\dots(n-k+1)}{(n - \rho_n)^k} \rightarrow 1,$$

e inoltre $\left(1 - \frac{\rho_n}{n}\right)^n = e^{n \ln\left(1 - \frac{\rho_n}{n}\right)} \rightarrow e^{-\rho}$, $\rho_n^k \rightarrow \rho^k$. Per $n \rightarrow \infty$ otteniamo pertanto la distribuzione limite

$$P(N = k) = e^{-\rho} \frac{\rho^k}{k!}, \quad k = 0, 1, 2, \dots \quad (1.6.8)$$

Questo limite definisce una nuova distribuzione, detta "variabile aleatoria poissoniana" con valor medio ρ . Si tratta di una variabile discreta, ma con un numero infinito di valori.

Poichè $\sum_{k=0}^{\infty} \frac{\rho^k}{k!} = e^\rho$, la serie dei termini (1.6.8) è, come deve, pari a 1. Per il valor medio non è difficile controllare che è $\rho = \lim_{n \rightarrow \infty} \rho_n$.

Per la dispersione, vedremo nel prossimo paragrafo.

Osservazione 6. La distribuzione poissoniana si applica ad eventi che non accadono quasi mai in una singola prova, ma il cui numero medio non è trascurabile a causa del gran numero di prove. Un esempio può essere quello delle chiamate ad un call center in un data ora. Si può supporre che gli utenti possibili siano dell'ordine dei milioni, ciascuno dei quali però ha una probabilità così bassa di chiamare nella data ora che il numero medio delle chiamate è di qualche unità.

Un esempio ancora più preciso è quello del decadimento radioattivo. Consideriamo una massa di uranio U238 fatta di un numero di atomi dell'ordine di 10^{13} , che è in realtà na massa assai piccola. Il numero di atomi è enorme, ma ciascun atomo decade con probabilità così piccola che il conteggio al Geiger in un'ora è di qualche unità.

3. Variabile aleatoria geometrica. La distribuzione della variabile geometrica è quella del tempo d'attesa del giocatore visto nel paragrafo 1.5, e dipende da un parametro $p \in (0, 1)$, che nel caso visto è la probabilità di vincita. Detta τ tale variabile, che prende valori negli interi $\mathbb{N} = 1, 2, \dots$, ha distribuzione

$$P(\tau = k) = p q^{k-1}, \quad q = 1 - p, \quad k = 1, 2, \dots$$

Questa distribuzione è detta “geometrica” con probabilità p . Il suo valor medio è $M(\tau) = 1/p$. Infatti

$$M(\tau) = \sum_{k=1}^{\infty} k P(\tau = k) = \sum_{k=1}^{\infty} p k q^{k-1} = p \frac{d}{dq} \sum_{h=0}^{\infty} q^h = p \frac{d}{dq} \frac{1}{1-q} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Persistenza della sfortuna. Un giocatore che ha già effettuato un numero K molto grande di giocate puntando sullo stesso risultato, senza vincere mai, sarà più vicino alla vincita di quanto lo era all'inizio? Così pensa chi gioca al lotto puntando sui “ritardi”.

Ma in realtà non è così. Per vederlo dobbiamo calcolare la probabilità che il giocatore che ha già passato senza successo K turni, debba attendere ancora altri r turni, per $r = 1, 2, \dots$, cioè la probabilità condizionata $P(\tau = K + r | \tau > K)$.

E' chiaro che $\{\tau = K + r\} \cap \{\tau > K\} = \{\tau = K + r\}$, e, come s'è visto, $P(\tau > K) = q^K$. Pertanto

$$P(\tau = K + r | \tau > K) = \frac{P(\tau = K + r)}{P(\tau > K)} = \frac{p q^{K+r-1}}{q^K} = p q^{r-1} = P(\tau = r).$$

La probabilità che debba attendere r turni è quindi la stessa che aveva all'inizio del gioco. Il “ritardo” causato dall'aver perso per K turni non avvicina il momento della vincita.

Esercizi.

1. Nel caso di tre lanci di una moneta, si trovi la distribuzione della variabile aleatoria $N =$ numero massimo di T consecutive.

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 .

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

4. Si lanciano due 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 .

5. In una popolazione con il 20% di individui appartenenti al gruppo sanguigno $Rh-$ e l' 80% al gruppo $Rh+$ si esaminano tre individui presi a caso (prove ripetute indipendenti). Qual'è il numero medio di $Rh-$ nei tre casi esaminati? E qual'è la media della differenza tra il numero di $Rh+$ e di $Rh-$ nei tre casi?

1.7. DISTRIBUZIONE DI PIU' VARIABILI ALEATORIE.

1.7.1. Distribuzione congiunta di due variabili aleatorie.

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

Detti x_1, x_2 i punti generici di $\Omega_{F_1}, \Omega_{F_2}$, rispettivamente, la **distribuzione congiunta delle due variabili** F_1, F_2 , è la probabilità su Ω_F con densità discreta

$$p_F(x_1, x_2) = P(\{F_1 = x_1, F_2 = x_2\}). \quad (1.7.1)$$

Se $A \subseteq \Omega_F = \Omega_{F_1} \times \Omega_{F_2}$ è un evento qualsiasi, sommando otteniamo

$$P_F(A) := P(\{F \in A\}) = \sum_{(x_1, x_2) \in A} p_F(x_1, x_2). \quad (1.7.2)$$

Conoscendo la distribuzione congiunta si può trovare la distribuzione di ogni variabile aleatoria esprimibile come funzione di F_1 e di F_2 . Sia infatti $G(\omega) = g(F_1(\omega), F_2(\omega))$, e sia Ω_G l'insieme dei valori assunti da G . Per ogni $y \in \Omega_G$ si ha

$$p_G(y) = P(G = y) = \sum_{\substack{(x_1, x_2) \in \Omega_F \\ g(x_1, x_2) = y}} p_F(x_1, x_2). \quad (1.7.3)$$

Le distribuzioni di F_1 ed F_2 (dette "distribuzioni marginali" della distribuzione congiunta) si ottengono facilmente dalla (1.7.1): se specifichiamo che $F_1 = x$ mentre F_2 può assumere qualsiasi valore, abbiamo

$$p_{F_1}(x) = P(\{F_1 = x\}) = P_F(\{x\} \times \Omega_{F_2}) = \sum_{y \in \Omega_{F_2}} p_F(x, y), \quad (1.7.4a)$$

e analogamente

$$p_{F_2}(x) = P(\{F_2 = x\}) = P_F(\Omega_{F_1} \times \{x\}) = \sum_{y \in \Omega_{F_1}} p_F(y, x). \quad (1.7.4b)$$

Esempio 1. Si lancia due volte un dado, e sia F_i , il numero di volte che esce il risultato i , $i = 1, \dots, 6$. Ciascun risultato può uscire non più di due volte in due lanci, e quindi $\Omega_{F_i} = \{0, 1, 2\}$ per ogni i .

Vogliamo trovare la distribuzione congiunta di F_1 e F_2 .

Abbiamo $\Omega_F = \{0, 1, 2\}^2$, ma siccome la somma non può superare 2, gli eventi elementari 12, 21, 22 hanno probabilità nulla, per cui basta assegnare la distribuzione su $\Omega'_F = \{00, 01, 11, 10, 02, 20\}$.

L'evento $\{F_1 = F_3 = 0\}$ corrisponde a lanci che danno (ω_1, ω_2) con $\omega_i \notin \{1, 3\}$, $i = 1, 2$. Tali coppie sono in tutto $4 \cdot 4 = 16$, per cui $p_F(0, 0) = P(\{F_1 = 0, F_3 = 0\}) = \frac{16}{36} = \frac{4}{9}$. 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 probabilità discreto (Ω, P) si dicono indipendenti se, per ogni scelta della coppia di valori $(x_1, x_2) \in \Omega_F = \Omega_{F_1} \times \Omega_{F_2}$, abbiamo

$$P(\{F_1 = x_1, F_2 = x_2\}) = P(\{F_1 = x_1\}) P(\{F_2 = x_2\}). \quad (1.7.5a)$$

La condizione è quindi che gli eventi $\{F_1 = x_1\}$, $\{F_2 = x_2\}$ siano indipendenti per ogni $x_1 \in \Omega_{F_1}$ e $x_2 \in \Omega_{F_2}$. La distribuzione, che è una probabilità su $\Omega_F = \Omega_{F_1} \times \Omega_{F_2}$, è pertanto la probabilità prodotto delle distribuzioni marginali

$$p_F(x_1, x_2) = p_{F_1}(x_1) \cdot p_{F_2}(x_2). \quad (1.7.5b)$$

Osservazione 1. Se le variabili aleatorie F_1, F_2 sono indipendenti, e f_1, f_2 sono due funzioni qualsiasi, definite su Ω_{F_1} e Ω_{F_2} , rispettivamente, allora sono anche indipendenti le variabili aleatorie $G_1 = f_1(F_1)$, $G_2 = f_2(F_2)$.

Infatti $\Omega_{G_1} = f_1(\Omega_{F_1})$, $\Omega_{G_2} = f_2(\Omega_{F_2})$ sono gli insiemi di valori da loro assunti, e per ogni $y_1 \in \Omega_{G_1}$ abbiamo

$$P(G_1 = y_1) = P(f_1(F_1) = y_1) = \sum_{x_1: f_1(x_1)=y_1} P(F_1 = x_1),$$

e similmente per G_2 . Per la distribuzione congiunta di G_1, G_2 abbiamo

$$P(G_1 = y_1, G_2 = y_2) = \sum_{x_1: f_1(x_1)=y_1} \sum_{x_2: f_2(x_2)=y_2} P(F_1 = x_1, F_2 = x_2).$$

Se F_1, F_2 sono indipendenti $P(F_1 = x_1, F_2 = x_2) = P(F_1 = x_1)P(F_2 = x_2)$ e l'espressione a destra diventa

$$\sum_{x_1: f_1(x_1)=y_1} P(F_1 = x_1) \sum_{x_2: f_2(x_2)=y_2} P(F_2 = x_2) = P(G_1 = y_1)P(G_2 = y_2).$$

Quindi G_1 e G_2 sono indipendenti.

Esempio 2. Si lancia quattro volte una moneta, e sia $\omega = (\omega_1, \omega_2, \omega_3, \omega_4) \in \{T, C\}^4$ il generico evento elementare. Consideriamo le variabili aleatorie $F_1(\omega) = N_T(\omega_1, \omega_2)$, il numero di teste nei primi due lanci, e il numero di teste negli ultimi due $F_2(\omega) = N_T(\omega_3, \omega_4)$. Abbiamo quindi $\Omega_{F_1} = \Omega_{F_2} = \{0, 1, 2\}$.

È facile vedere che F_1 ed F_2 sono indipendenti: infatti abbiamo, per esempio $P(F_1 = 2) = P(F_2 = 2) = 1/4$ e

$$P(F_1 = 2, F_2 = 2) = P(TTTT) = \frac{1}{16} = P(F_1 = 2)P(F_2 = 2),$$

e la relazione di indipendenza si verifica facilmente anche per gli altri valori.

Il risultato non è sorprendente, perchè F_1 ed F_2 sono funzioni determinate da prove indipendenti (per F_1 il primo e il secondo lancio, e per F_2 il terzo e il quarto).

Esempio 3. Le variabili F_1, F_3 dell'esempio 1 non sono invece indipendenti. Infatti $p_F(2, 0) = P(F_1 = 2, F_3 = 0) = \frac{1}{36}$ mentre $P(F_1 = 2) = \frac{1}{36}$ e $P(F_3 = 0) = \frac{25}{36}$, quindi $P(F_1 = 2, F_3 = 0) \neq P(F_1 = 2)P(F_3 = 0)$.

Lo stesso vale per le altre coppie possibili di variabili aleatorie F_i, F_j , $i, j = 1, \dots, 6$, $i \neq j$ dell'esempio 1.

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 di due variabili indipendenti è pari al prodotto dei valori medi.*

Per dimostrarlo, possiamo considerare il prodotto $F_1 F_2$ come una variabile aleatoria sullo spazio $\Omega_F = \Omega_{F_1} \times \Omega_{F_2}$ e applicare la (1.7.5a) o la (1.7.5b): tenendo conto dell'indipendenza abbiamo

$$\begin{aligned} M(F_1 F_2) &= \sum_{x_1 \in \Omega_{F_1}} \sum_{x_2 \in \Omega_{F_2}} x_1 x_2 p_F(x_1, x_2) \\ &= \sum_{x_1 \in \Omega_{F_1}} x_1 p_{F_1}(x_1) \sum_{x_2 \in \Omega_{F_2}} x_2 p_{F_2}(x_2) = M(F_1)M(F_2). \end{aligned} \quad (1.7.6)$$

1.7.2. Covarianza o correlazione di due variabili aleatorie.

Date due variabili aleatorie F_1, F_2 , su uno spazio di probabilità (Ω, P) , si definisce loro covarianza o correlazione la quantità

$$\text{Cov}(F_1, F_2) = M(\widehat{F}_1 \cdot \widehat{F}_2), \quad (1.7.7a)$$

dove le $\widehat{F}_i = F_i - M(F_i)$, $i = 1, 2$, sono le variabili centrate. E' facile vedere, come per la varianza che si ha

$$\text{Cov}(F_1, F_2) = M(F_1 \cdot F_2) - M(F_1)M(F_2). \quad (1.7.7b)$$

Infatti sviluppando il prodotto $\widehat{F}_1 \widehat{F}_2$ e prendendo il valor medio si vede che

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

La covarianza è di grande 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 (1.7.6) e dalla (1.7.7b) segue che **variabili indipendenti sono scorrelate.**

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

Esempio 4. 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, perchè $F_2 = F_1^2$, 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 $M(F_1 F_2) = M((F_1)^3)$, e poichè 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. La covarianza (1.7.7a,b), così come la varianza (per la proprietà ii) vista sopra) dipende dalla scala di misura delle grandezze. Conviene allora introdurre una quantità che non ne dipende, e questa è data dal coefficiente di correlazione:

$$\rho(F_1, F_2) = \frac{\text{Cov}(F_1, F_2)}{\sqrt{\text{Var}(F_1)\text{Var}(F_2)}}. \quad (1.7.8)$$

Per vedere che $\rho(F_1, F_2)$ è invariante per cambiamenti di scala (che conservano il verso) osserviamo che aggiungendo delle costanti a F_1, F_2 , ρ non cambia perchè sia $\text{Cov}(F_1, F_2)$ 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, altrimenti cambierebbe il verso della scala), allora sia il numeratore che il denominatore della (1.7.8) 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 , la varianza di $\widehat{F}_1 + x\widehat{F}_2$ non può essere negativa e quindi dobbiamo avere

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

Questa espressione è senz'altro positiva per x grandi. Infatti in questa regione domina il termine $x^2 \text{Var}(F_2)$ che è positivo. Come è noto, la condizione affinché non cambi mai segno è che il discriminante sia negativo:

$$\Delta = (\text{Cov}(F_1, F_2))^2 - \text{Var}(F_1)\text{Var}(F_2) \leq 0. \quad (1.7.9)$$

In questo caso infatti l'equazione di secondo grado $ax^2 + 2bx + c = 0$, con $a = \text{Var}(F_2)$, $b = \text{Cov}(F_1, F_2)$, e $c = \text{Var}(F_1)$ non ha soluzioni.

Dividendo entrambi i membri della (1.7.9) per $\text{Var}(F_1)\text{Var}(F_2)$ e confrontando con la (1.7.8) si vede che deve essere $\rho^2 < 1$.

E' chiaro che $\rho^2 = 1$ se e solo se $\Delta = 0$. In tal caso esiste un valore $x = x_0$ tale che la varianza di $\widehat{F}_1 + x_0\widehat{F}_2$ è nulla e quindi, a meno di insiemi di probabilità nulla, $\widehat{F}_1 + x_0\widehat{F}_2 = 0$. Quindi F_1 è una funzione lineare di F_2 : $F_1 = -x\widehat{F}_2 + M(F_1)$, cioè le due grandezze differiscono solo per un cambiamento di scala.

1.7.3. Distribuzione congiunta di tre o più variabili aleatorie.

Il caso di più di due variabili aleatorie F_1, F_2, \dots, F_n , con $n > 2$, è analogo. Detti $\Omega_{F_i}, i = 1, \dots, n$ gli insiemi dei valori assunti dalle variabili F_i , e posto $F = (F_1, \dots, F_n)$, la distribuzione congiunta sullo spazio $\Omega_F = \Omega_{F_1} \times \dots \times \Omega_{F_n}$ è

$$p_F(x_1, \dots, x_n) = P(\{F_1 = x_1, \dots, F_n = x_n\}). \quad (1.7.10)$$

Le marginali di ciascuna variabile si otterranno sommando sui valori di tutte le altre. Avremo, per esempio

$$p_{F_1}(x) = \sum_{x_2 \in \Omega_{F_2}} \dots \sum_{x_n \in \Omega_{F_n}} p_F(x, x_2, \dots, x_n),$$

e similmente per le altre.

La condizione d'indipendenza per più di due variabili è

$$p_F(x_1, \dots, x_n) = P(\{F_1 = x_1, \dots, F_n = x_n\}) = p_{F_1}(x_1) \dots p_{F_n}(x_n). \quad (1.7.11)$$

Quindi, come nel caso di due variabili, anche per più variabili indipendenti la distribuzione p_F sullo spazio $\Omega_F = \Omega_{F_1} \times \Omega_{F_2} \dots \times \Omega_{F_n}$ è il prodotto delle distribuzioni marginali. Ne segue, come visto nell'osservazione 1 sopra, che eventi o variabili aleatorie dipendenti da variabili indipendenti diverse sono indipendenti. In particolare sono indipendenti eventi del tipo $\{F_1 \in A_1\}, \dots, \{F_n \in A_n\}$, con $A_1 \in \Omega_{F_1}, \dots, A_n \in \Omega_{F_n}$.

Varianza della somma di variabili aleatorie. A differenza di quanto avviene per il valor medio, la varianza della somma di due variabili aleatorie non è in generale data dalla somma delle varianze. Infatti se $F = F_1 + F_2$ la corrispondente variabile centrata è $\widehat{F} = F_1 + F_2 - M(F_1) - M(F_2) = \widehat{F}_1 + \widehat{F}_2$, e per le (1.6.6a), (1.7.7a,b) abbiamo

$$\text{Var}(F) = M(\widehat{F}^2) = M(\widehat{F}_1^2 + \widehat{F}_2^2 + 2\widehat{F}_1\widehat{F}_2) = \text{Var}(F_1) + \text{Var}(F_2) + 2\text{Cov}(F_1, F_2). \quad (1.7.12a)$$

Nel caso di più variabili aleatorie, per la varianza di $F = F_1 + F_2 + \dots + F_n$ si ottiene la formula generale

$$\text{Var}(F) = \sum_{i=1}^n \text{Var}(F_i) + 2 \sum_{\{i,j\}} \text{Cov}(F_i, F_j), \quad (1.7.12b)$$

dove con $\sum_{\{i,j\}}$ si intende la somma sulle coppie di indici distinti $i \neq j$. Infatti basta sviluppare il quadrato: $\widehat{F}^2 = \sum_{i=1}^n \widehat{F}_i^2 + 2 \sum_{\{i,j\}} \widehat{F}_i \widehat{F}_j$.

Osservazione 2. 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 è eguale alla corrispondente somma delle varianze.

L'osservazione 2 consente di calcolare facilmente la varianza della binomiale.

Varianza delle variabili binomiale e di Poisson. Come visto nel §1.6.4, se N è una variabile binomiale nella classe $B(n; p)$, corrispondente cioè a n prove con probabilità

di successo $p \in (0, 1)$, allora si può scrivere come somma di variabili indipendenti: $N = \sum_{i=1}^n \omega_i$, dove le ω_i rappresentano il "numero di successi" nella singola prova i -esima.

Si ha $\omega_i \in \{0, 1\}$, per cui $\omega_i^2 = \omega_i$, e inoltre $P(\{\omega_i = 1\}) = p$. Si ottiene $M(\omega_i) = p$, come si è visto, e per la varianza

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

Il valor medio, come sappiamo, è $M(N) = np$, e la varianza, per l'indipendenza è la somma delle varianze, per cui $\text{Var}(N) = np(1 - p) = npq$.

La variabile poissoniana è sostanzialmente una variabile binomiale con n molto grande, $p = p_n$ molto piccolo, e corrisponde al limite $n \rightarrow \infty$ sotto la condizione $np_n \rightarrow \rho > 0$. Per ogni n la varianza è, come si è visto $np_n(1 - p_n)$, e siccome $np_n \rightarrow \rho$ per $n \rightarrow \infty$, il che implica che $p_n \rightarrow 0$, vediamo che il limite di tale varianza è pari a ρ . La varianza della poissoniana è quindi eguale al valor medio.

Esempio 5. Si piantano tre semi di una semente a con germinabilità (probabilità di ciascun seme di germogliare) del 90% e due semi di una semente b con germinabilità del 60%. Considerando che la germinazione di ogni seme sia indipendente, si trovino il valor medio e la varianza del numero dei semi germinati.

Il numero dei semi germinati si scrive come $N = N_a + N_b$ dove N_a, N_b si riferiscono ai semi germinati di tipo a e b . Sono entrambi variabili binomiali, con distribuzione, rispettivamente, $B(3; 0, 9)$ e $B(2; 0, 6)$. Quindi $M(N) = M(N_a) + M(N_b) = 3 \cdot 0,9 + 2 \cdot 0,6$. Poichè sono indipendenti, si ha $\text{Var}(N) = \text{Var}(N_a) + \text{Var}(N_b) = 3 \cdot 0,9 \cdot 0,1 + 2 \cdot 0,6 \cdot 0,4$.

Esercizi.

1. Si lancia tre volte una moneta, e sia $N_T^{(1)}$ il numero di teste ottenute nei primi due lanci e $N_T^{(2)}$ il numero di teste al secondo e terzo lancio. Trovare la distribuzione congiunta di $N_T^{(1)}$ e $N_T^{(2)}$.

2. Trovare la covarianza di $N_T^{(1)}$ e $N_T^{(2)}$ nel caso del precedente esercizio.

3. Da un'urna con sette sfere nere e tre bianche si effettuano tre estrazioni con restituzione. Si tolgono poi due sfere nere e si effettuano altre due estrazioni con restituzione. Sia N_b il numero di sfere bianche estratte nelle cinque prove. Si calcolino il valor medio e la varianza di N_b . *Suggerimento: si confronti con l'esempio 5.*

4. Un call center in una certa ora riceve chiamate dai bacini di utenza a e b . Il numero delle chiamate nella data ora dai due bacini è descritto da poissoniane indipendenti con medie, rispettivamente, $\rho_a = 2$ e $\rho_b = 3$.

Trovare. i) la probabilità che non ci sia alcuna chiamata al call center nella data ora; ii) il valor medio e la varianza del numero delle chiamate.

1.8. PROBABILITA' CONTINUA.

Per una migliore comprensione di alcuni risultati fondamentali, come il teorema centrale del limite, è opportuno dare dei cenni alla probabilità continua e alle variabili aleatorie continue. Tratteremo solo il caso di probabilità continua con densità.

Il presente paragrafo richiede una buona conoscenza almeno dei fatti elementari della teoria dell'integrazione. Gli aspetti generali non sono diversi dal caso della probabilità discreta, e la principale variazione consiste nel fatto che nel calcolare le probabilità le somme sono sostituite da integrali.

1.8.1. Probabilità continua su \mathbb{R} con densità.

Se $p(x)$, $x \in \mathbb{R}$, è una funzione non negativa su \mathbb{R} , cioè tale che $p(x) \geq 0$ per ogni x , è integrabile e si ha

$$\int_{-\infty}^{\infty} p(x)dx = 1, \quad (1.8.1)$$

allora $p(x)$ definisce una probabilità su \mathbb{R} , cioè uno spazio di probabilità (\mathbb{R}, P) , dove \mathbb{R} è lo spazio degli eventi, e la probabilità P è definita su ogni intervallo $I = [a, b]$ da un integrale

$$P(I) = P([a, b]) = \int_a^b p(x)dx. \quad (1.8.2)$$

Se $b = a$ l'intervallo degenera, si riduce ad un solo punto $[a, a] = \{a\}$, e dalla (1.8.2) vediamo che $P(\{a\}) = 0$. Quindi ogni punto $x \in \mathbb{R}$ ha probabilità nulla. In ogni punto x è però definita la densità $p(x)$ e, interpretando l'integrale (1.8.1) come una specie di somma, si può dire che ogni punto contribuisce con una "probabilità simbolica" $p(x)dx$. La (1.8.1) afferma quindi che la "somma", come sempre in probabilità, dà 1.

Segue da quanto detto che l'intervallo I ha la stessa probabilità (1.8.2) sia che sia aperto, o chiuso o semiaperto.

La probabilità si può estendere ad insiemi più generali degli intervalli. Nel caso di un'unione di intervalli disgiunti $A = I_1 \cup I_2 \cup \dots \cup I_m$, con $I_j = [a_j, b_j]$, $j = 1, \dots, m$ e $a_1 < b_1 < a_2 < b_2 < \dots < a_m < b_m$, avremo

$$P(A) = \sum_{j=1}^m P(I_j) = \sum_{j=1}^m \int_{a_j}^{b_j} p(x)dx =: \int_A p(x)dx, \quad (1.8.3)$$

dove la notazione standard \int_A sta ad indicare l'integrazione su tutto l'insieme A .

L'estensione al caso di infiniti intervalli disgiunti non presenta problemi, grazie al fatto che la serie degli integrali è sempre convergente: infatti è crescente, perchè gli integrali sui singoli intervalli non possono essere negativi, ed è limitata per la (1.8.1).

Se abbiamo un'unione di intervalli non che sono disgiunti, basta osservare che l'unione si può sempre rappresentare come unione di intervalli disgiunti, chiusi, aperti o semiaperti. Infatti l'intersezione di due intervalli è l'insieme vuoto oppure un intervallo.

Si possono prendere in considerazione anche intervalli infiniti, cioè tutto \mathbb{R} e semirette del tipo $(-\infty, a]$ e $[a, +\infty)$, e le loro varianti aperte.

La classe \mathcal{B} degli eventi rappresentati da unioni finite o numerabili di intervalli (finiti o infiniti) gode della proprietà che le intersezioni, unioni e complementi dei suoi elementi sono anche unioni finite o numerabili di intervalli. Infatti per l'unione abbiamo già visto, e i complementi di intervalli, finiti o infiniti, sono fatti di intervalli finiti o infiniti.

La classe \mathcal{B} è sufficiente per i nostri scopi. Per la probabilità su questa classe di eventi valgono le proprietà fondamentali della probabilità del paragrafo §1.3.2.

La probabilità così definita sulla classe \mathcal{B} dei sottoinsiemi di \mathbb{R} è detta probabilità continua su \mathbb{R} con "densità" data dalla funzione $p(x)$.

La probabilità si estende alle unioni di intervalli disgiunti, finite o infinite, la cui probabilità sarà data da una somma (o serie) di probabilità di intervalli.

Per esempio se l'evento A è un'unione finita di intervalli disgiunti, $A = I_1 \cup I_2 \cup \dots \cup I_m$, con $I_j = [a_j, b_j]$, $j = 1, \dots, m$, con $a_1 < b_1 < a_2 < b_2 < \dots < a_m < b_m$, abbiamo

$$P(A) = \sum_{j=1}^m P(I_j) = \sum_{j=1}^m \int_{a_j}^{b_j} p(x) dx =: \int_A p(x) dx, \quad (1.8.3)$$

dove introduciamo la notazione standard \int_A per indicare l'integrazione su un'unione di integrali disgiunti.

Abbiamo in questo modo definito lo spazio di probabilità continuo (\mathbb{R}, P) con densità di probabilità $p(x)$.

In conclusione si può dire che la probabilità continua ora definita è del tutto analoga alla probabilità discreta, con la differenza che abbiamo integrali al posto di somme.

Se $p(x)$ è nulla al di fuori di un qualche intervallo I (che può essere infinito), allora si ha $P(I) = 1$, cioè la probabilità "è concentrata su I ", e lo spazio di probabilità potremo anche denotarlo (I, P) .

Presentiamo ora le probabilità continue su \mathbb{R} con densità di maggiore rilevanza.

1. Probabilità uniforme su un intervallo. La probabilità continua uniforme sull'intervallo $I = [a, b]$ è specificata dalla densità $p(x) = \frac{1}{b-a}$, $x \in I$. E' concentrata su I , ma la si può considerare come probabilità su \mathbb{R} ponendo $p(x) = 0$ per $x \notin I$.

La densità è costante su I , e la costante è l'inverso della lunghezza cioè $\frac{1}{b-a}$, che garantisce la condizione $\int_a^b \frac{dx}{b-a} = 1$.

Si tratta dell'analogo continuo della probabilità classica o uniforme sugli spazi con un numero finito di elementi. Infatti la probabilità di un intervallo $(\alpha, \beta) \subset [a, b]$ è pari alla sua lunghezza divisa per la lunghezza totale:

$$P((\alpha, \beta)) = \frac{\beta - \alpha}{b - a}$$

e lo stesso vale per un'unione di intervalli) disgiunti in I che ha probabilità pari alla somma delle lunghezze divise per $b - a$.

Se $F(x)$ è una variabile casuale, cioè una funzione definita su I a valori reali, la sua media rispetto alla probabilità continua coincide con la nozione di media integrale:

$$M(F) = \frac{1}{b - a} \int_a^b F(x) dx.$$

2. Probabilità esponenziale. E' la probabilità concentrata sull'intervallo infinito $[0, \infty)$ con densità $p(x) = ke^{-kx}$, per $k > 0$. Questa funzione è infatti una densità di probabilità:

$$\int_0^{\infty} ke^{-kx} dx = - \int_0^{\infty} \frac{d}{dx} e^{-kx} dx = 1.$$

3. Probabilità gaussiana standard. E' la probabilità su \mathbb{R} che ha per densità la funzione $g(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$. Si può dimostrare, anche se la dimostrazione non è elementare, che

$$\int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = 1.$$

Esempio 1. Data la probabilità continua uniforme sull'intervallo $I = [0, 1]$, si calcoli la probabilità degli eventi $A = [0, \frac{3}{4}]$ e $B = [\frac{1}{4}, 1]$. Stabilire se sono indipendenti.

Abbiamo $P(A) = P(B) = \frac{3}{4}$, e $P(A \cap B) = P([\frac{1}{4}, \frac{3}{4}]) = \frac{1}{2}$. Quindi i due eventi non sono indipendenti.

Esempio 2. Data la probabilità esponenziale con densità $2e^{-2x}$, si consideri, per $s > 0$, l'evento $A_s = [s, \infty)$. La sua probabilità è

$$P(A_s) = \int_s^{\infty} 2e^{-2x} dx = - \int_s^{\infty} \frac{d}{dx} e^{-2x} dx = e^{-2s}.$$

Presi due numeri positivi R, r , calcoliamo la probabilità condizionata $P(A_{R+r}|A_R)$. Abbiamo $P(A_R) = e^{-2R}$ e $P(A_{R+r}) = e^{-2(R+r)}$. Inoltre è chiaro che $A_{R+r} \cap A_R = A_{R+r}$, e pertanto si ha

$$P(A_{R+r}|A_R) = \frac{P(A_{R+r})}{P(A_R)} = e^{-2r} = P(A_r).$$

1.8.2. Variabili aleatorie continue con densità.

Una variabile aleatoria F , definita su uno spazio di probabilità (Ω, P) , si dice variabile aleatoria continua con densità se esiste una funzione integrabile $p_F(x)$ tale che, comunque si prende un intervallo $I = [a, b] \subset \mathbb{R}$ si ha

$$P(F \in [a, b]) = \int_I p_F(x) dx = \int_a^b p_F(x) dx. \quad (1.8.4)$$

La funzione p_F si chiama la densità (continua) della distribuzione di F . Deve essere $p_F(x) \geq 0$, perchè le probabilità sono positive, e inoltre

$$\int_{-\infty}^{\infty} p_F(x) dx = 1. \quad (1.8.5)$$

Come nel caso discreto, la relazione (1.8.4) definisce una probabilità P_F , continua, su \mathbb{R} , che è detta "distribuzione della variabile aleatoria F ". Questa probabilità è naturalmente concentrata sul codominio $F(\Omega) \subset \mathbb{R}$ di F , che indichiamo anche come Ω_F .

Si noti che lo spazio Ω su cui è definita F non può essere discreto, altrimenti la F assumerebbe un insieme discreto di valori e non potrebbe avere una distribuzione continua.

Come nel caso discreto, vi sono più, anzi infinite, variabili aleatorie con una data distribuzione P_F , definite su diversi spazi di probabilità che però, per l'osservazione di sopra, non possono essere discreti.

Inoltre, dato che \mathbb{R} e i suoi sottoinsiemi sono fatti di numeri, la funzione identica $F_0(x) = x$ è anch'essa una variabile aleatoria, la cui distribuzione è data dalla probabilità di partenza. Questo spiega come mai le probabilità continue su \mathbb{R} sono spesso anche chiamate variabili aleatorie. Infatti sia P una probabilità continua su \mathbb{R} con densità $p(x)$, e consideriamo la distribuzione P_{F_0} della variabile aleatoria data dalla funzione identica $F_0(x) = x$. Il codominio di F_0 coincide con lo spazio di partenza (il dominio) $F_0(\mathbb{R}) = \mathbb{R}$, e per ogni $A \in \mathcal{B}$ si ha $P_{F_0}(A) = P(A)$. Infatti, se per esempio $A = (\alpha, \beta)$ si ha

$$P_F(A) = P(F_0(x) \in (\alpha, \beta)) = P((\alpha, \beta)) = \int_{\alpha}^{\beta} p(x) dx$$

e lo stesso accade se A è un'unione finita o numerabile di intervalli, per cui $P_{F_0} = P$.

Esempio 3. Data la distribuzione uniforme sull'intervallo $I = [-1, 1]$ e la variabile aleatoria $F(x) = (1+x)^2$ su I , si vuol calcolare la probabilità dell'evento $\{F(x) > \frac{1}{4}\}$.

Poichè $1+x \geq 0$ per ogni $x \in I$, l'evento corrisponde a $1+x > 1/2$ ovvero $x > -1/2$ per cui la probabilità richiesta è

$$\frac{1}{2} \int_{-1/2}^1 dx = \frac{3}{4}.$$

Valor medio e dispersione. Il valor medio di una variabile aleatoria con distribuzione continua ha un'espressione analoga a quella vista nel caso discreto, più precisamente analoga alla seconda formula a destra nella (1.6.2), con la sola differenza che al posto della somma c'è un integrale:

$$M(F) = \int_{\mathbb{R}} x p_F(x) dx. \quad (1.8.6)$$

In alcuni casi l'integrale, che è esteso in generale su tutto \mathbb{R} , può non esistere, anche se $p_F(x)$ è una funzione limitata, perchè la funzione integranda non va abbastanza rapidamente a zero per $|x| \rightarrow \infty$. In tal caso il valor medio non esiste.

La varianza è definita come nel caso generale $\text{Var}(F) = M((F - M(F))^2)$ ed è naturalmente calcolata come un integrale

$$\text{Var}(F) = \int_{\mathbb{R}} (x - M(F))^2 p_F(x) dx. \quad (1.8.7)$$

Esempio 4. Calcoliamo valor medio e dispersione della variabile aleatoria del precedente esempio 3. Il valor medio è

$$M(F) = \frac{1}{2} \int_{-1}^2 (1+x)^2 dx = \frac{1}{2} \left[\frac{(1+x)^3}{3} \right]_{-1}^1 = \frac{4}{3},$$

e per la dispersione basta calcolare la media del quadrato

$$M(F^2) = \frac{1}{2} \int_{-1}^2 (1+x)^4 dx = \frac{1}{2} \left[\frac{(1+x)^5}{5} \right]_{-1}^2 = \frac{32}{5},$$

per cui si ha $\text{Var}(F) = M(F^2) - (M(F))^2 = \frac{128}{45}$.

Esercizio 1. Si consideri la probabilità continua uniforme sull'intervallo $I = [-2, 1]$. Calcolare media e varianza della variabile aleatoria $F(x) = x + \frac{1}{2}$.

Variabili aleatorie continue notevoli. Le variabili aleatorie notevoli che qui tratteremo sono quelle le cui distribuzioni sono state introdotte nel precedente paragrafo, e cioè la variabile uniforme, la variabile esponenziale e la variabile gaussiana.

Come nel caso discreto, parlando di variabile aleatoria ci riferiamo alla sua distribuzione, o, se si vuole, alla classe di equivalenza di tutte le variabili aleatorie che hanno quella distribuzione. Così una variabile aleatoria, ovunque sia definita, è detta uniforme (su un intervallo I) se ha per distribuzione la probabilità continua uniforme su I , e in modo simile si parla di variabili esponenziali e gaussiane.

Calcoliamo ora il valor medio e la varianza per le variabili aleatorie notevoli introdotte nel paragrafo precedente.

1. La variabile con distribuzione uniforme su $I = [a, b]$ ha come valor medio il punto di mezzo $\frac{a+b}{2}$ di I :

$$M(F_0) = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

La varianza, cambiando variabile ponendo $y = x - \frac{a+b}{2}$, è data da

$$\text{Var}(F_0) = \frac{1}{b-a} \int_{\frac{a-b}{2}}^{\frac{b-a}{2}} y^2 dy = \frac{2}{b-a} \int_0^{\frac{b-a}{2}} y^2 dy = \frac{(b-a)^2}{12}.$$

2. Per trovare il valor medio della variabile esponenziale con parametro $k > 0$ si integra per parti:

$$M(F_0) = \int_0^{\infty} k x e^{-kx} dx = - \int_0^{\infty} x \frac{d}{dx} e^{-kx} dx = \int_0^{\infty} e^{-kx} dx = \frac{1}{k}.$$

Inoltre con un integrale simile è facile vedere che $\text{Var}(F_0) = \frac{1}{k^2}$.

3. La variabile gaussiana standard, che è tradizionalmente indicata con Z , ha valor medio nullo, perchè la funzione $x e^{-x^2/2}$ è dispari, e il suo integrale su \mathbb{R} è quindi nullo. Per la varianza abbiamo

$$\text{Var}(Z) = \int_{-\infty}^{\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 2 \int_0^{\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -2 \int_0^{\infty} x \frac{d}{dx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Integrando per parti quest'ultimo integrale si trova

$$\text{Var}(Z) = 2 \int_0^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = 1.$$

Spesso si considera una variabile gaussiana generale, $Z_{m,\sigma}$, con densità

$$g_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad m \in \mathbb{R}, \sigma > 0. \quad (1.8.8)$$

La variabile $Z_{m,\sigma}$ ha valor medio m e varianza σ^2 , ed inoltre è facile vedere che $Z_{m,\sigma}$ e la variabile standard Z sono collegate da semplici trasformazioni

$$Z_{m,\sigma} = \sigma Z + m, \quad Z = \frac{Z_{m,\sigma} - m}{\sigma}. \quad (1.8.9)$$

Infatti per la prima delle (1.8.9) basta dimostrare che la distribuzione di $\sigma Z + m$ è data dalla (1.8.8). Infatti

$$P(\{\sigma Z + m \in (a, b)\}) = P\left(\left\{Z \in \left(\frac{a-m}{\sigma}, \frac{b-m}{\sigma}\right)\right\}\right) = \int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx.$$

Cambiando variabile ponendo $x = \frac{y-m}{\sigma}$ vediamo che l'integrale a destra è pari a

$$\int_a^b g_{m,\sigma}(x) dx,$$

come volevasi dimostrare.

La seconda delle (1.8.9) si vede in modo analogo.

Poichè Z ha media nulla e varianza 1, dalla prima delle (1.8.9) si vede che $Z_{m,\sigma}$ ha media m e varianza σ^2 .

1.9. SOMME DI VARIABILI ALEATORIE INDIPENDENTI. RISULTATI FONDAMENTALI.

Questo paragrafo è dedicato all'esposizione di due risultati fondamentali sul comportamento di somme di un gran numero di variabili aleatorie indipendenti: la legge dei grandi numeri e il teorema centrale del limite (o teorema del limite centrale).

Questi risultati sono anche alla base dei principali metodi di analisi statistica.

1.9.1. Legge dei grandi numeri.

La legge dei grandi numeri è uno dei risultati fondamentali della teoria della Probabilità, se non addirittura il più importante. Il primo a formularla in termini generali fu Jakob Bernoulli nella sua opera postuma *Ars Conjectandi* (1712).

Consideriamo una successione infinita di variabili aleatorie indipendenti F_1, F_2, \dots , tutte con la stessa distribuzione. E' la situazione che abbiamo quando ripetiamo delle prove in condizioni identiche. I valori assunti da F_1, F_2, \dots possiamo considerarli come risultati delle successive prove.

La media aritmetica S_n/n , dove $S_n = F_1 + F_2 + \dots + F_n$ è la somma dei risultati, è una variabile aleatoria, detta anche "media empirica", perchè si può pensare come la media di risultati ottenuti da chi ripete n volte la prova.

Poichè le variabili F_i hanno tutte la stessa distribuzione, hanno anche lo stesso valor medio $M(F_i) = m$. Pertanto $M(S_n) = nm$, e la media (empirica) S_n/n ha valor medio (probabilistico) m : $M(S_n/n) = m$.

La "legge dei grandi numeri" afferma che la media empirica S_n/n , quando n è molto grande, è vicina quanto si vuole alla media probabilistica m , con probabilità che tende ad 1 per $n \rightarrow \infty$.

Detto in altro modo, per grandi n il rapporto S_n/n prende, con grande probabilità, valori molto vicini alla media probabilistica m , diventa, cioè, "quasi una costante".

La formulazione precisa è data dal teorema seguente, detto "legge dei grandi numeri".

Teorema (Legge dei Grandi Numeri). Nelle ipotesi dette sopra, per ogni scelta del numero $\epsilon > 0$ comunque piccolo, si ha

$$\lim_{n \rightarrow \infty} P \left(\left\{ \left| \frac{S_n}{n} - m \right| > \epsilon \right\} \right) = 0. \quad (1.9.1)$$

Dimostrazione. La dimostrazione è molto semplice. La probabilità nella (1.9.1) si scrive anche come $P(|\widehat{S}_n| > n\epsilon)$, dove $\widehat{S}_n = S_n - M(S_n) = S_n - nm$ è lo scarto dalla media, e quindi anche come $P(\widehat{S}_n^2 > n^2\epsilon^2)$. Possiamo applicare la disuguaglianza di Chebyshev (1.6.5) alla variabile aleatoria non negativa \widehat{S}_n^2 , ottenendo

$$P(\widehat{S}_n^2 > n^2\epsilon^2) \leq \frac{M(\widehat{S}_n^2)}{n^2\epsilon^2}. \quad (1.9.2)$$

Come è noto, $M(\widehat{S}_n^2) = \text{Var}(S_n)$, e, siccome S_n è una somma di variabili indipendenti, $\text{Var}(S_n)$ è la somma delle varianze delle singole variabili (osservazione 2 del paragrafo precedente), le quali sono tutte eguali, perchè le F_i sono identicamente distribuite:

$\text{Var}(F_i) = \sigma^2$. Quindi $\text{Var}(S_n) = n\sigma^2$, e sostituendo nella (1.9.2) si ha

$$P\left(\left\{\left|\frac{S_n}{n} - m\right| > \epsilon\right\}\right) \leq \frac{\sigma^2}{n\epsilon^2}. \quad (1.9.3)$$

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

Applicazione della Legge dei grandi numeri alla binomiale e conseguenze.

Consideriamo la distribuzione binomiale $B(n; p)$, discussa al punto 1 del §1.6.4. Il numero di "successi" N , come abbiamo visto, si scrive come $N = \sum_{i=1}^n F_i$, dove le F_i indipendenti e identicamente distribuite, con valori in $\{0, 1\}$ e con $M(F_i) = p$, $\text{Var}(F_i) = p(1-p)$. Quindi N è un caso particolare delle somme S_n considerate nella legge dei grandi numeri. Applicando la (1.9.1) otteniamo

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{N}{n} - p\right| > \epsilon\right) = 0. \quad (1.9.4)$$

Il rapporto $\frac{N}{n}$ è detto anche "frequenza empirica" dei successi. La legge dei grandi numeri quindi afferma che con probabilità che tende ad 1 per $n \rightarrow \infty$, **la frequenza empirica dei successi in n prove indipendenti è vicina quanto si vuole alla probabilità di successo in una singola prova p .**

La (1.9.4) suggerisce la possibilità di calcolare approssimativamente una probabilità di successo p incognita effettuando una serie di esperimenti indipendenti e calcolando la frequenza empirica. Questo fatto è alla base di gran parte della statistica matematica.

La (1.9.4) fornisce anche la possibilità di definire probabilità non definibili in altro modo come limite delle frequenze empiriche. Questa è in sostanza la cosiddetta "definizione frequentista" delle probabilità, di cui si fa grande uso nelle scienze applicate.

1.9.2. Teorema centrale del limite.

Legge di De Moivre-Laplace. Vogliamo ora considerare il comportamento della distribuzione del numero di successi N in n prove identiche indipendenti, quando il numero delle prove n è molto grande. Si tratta quindi del comportamento della distribuzione binomiale $B(n; p)$, discussa al punto 1 del §1.6.4, per grandi valori di n .

Si noti che, contrariamente al caso di Poisson, la probabilità p rimane costante.

Teorema. (Legge di de Moivre-Laplace). Se N è una variabile aleatoria con distribuzione binomiale $B(n; p)$, e $\sigma^2 = pq$ designa la dispersione di singola prova, si ha, per ogni k intero tale che $|k - np| < a_n n^{2/3}$, con $a_n \rightarrow 0$ per $n \rightarrow \infty$

$$P(N = k) = \binom{n}{k} p^k q^{n-k} = \frac{1}{\sqrt{2\pi n \sigma^2}} e^{-\frac{(k-np)^2}{2n\sigma^2}} (1 + r_n(k)) \quad (1.9.5)$$

con $\lim_{n \rightarrow \infty} \max_{k: |k-np| < a_n n^{2/3}} |r_n(k)| = 0$.

Osservazione 1. La legge di de Moivre-Laplace afferma dunque che nella zona centrale, in cui k non dista dalla media np più di una quantità dell'ordine di $n^{2/3}$, la

probabilità binomiale $P(N = k)$ scritta in termini della variabile normalizzata $x = \frac{N-np}{\sqrt{npq}}$ è approssimativamente proporzionale alla funzione “gaussiana” $e^{-\frac{x^2}{2}}$.

Dimostrazione (schema). La dimostrazione fa uso della formula di Stirling (1.2.3), che fornisce una buona approssimazione di $n!$ per n grandi. Nelle nostre ipotesi non solo n , ma anche k e $n - k$ vanno all'infinito per $n \rightarrow \infty$: infatti k è vicino ad np , mentre $n - k$ è vicino a $n(1 - p) = nq$. Quindi possiamo usare la formula di Stirling per tutti e tre i fattoriali, e, trascurando il termine correttivo, abbiamo

$$\frac{n!}{k!(n-k)!} p^k q^{n-k} \approx \frac{n^n e^{-n} \sqrt{np}^k q^{n-k}}{\sqrt{2\pi k(n-k)} k^k e^{-k} (n-k)^{n-k} e^{-n+k}},$$

dove \approx indica l'eguaglianza approssimata, tanto più precisa, quanto più n è grande. Ponendo $p_* = \frac{k}{n}$, $q_* = \frac{n-k}{n} = 1 - p_*$, troviamo

$$\frac{n^n e^{-n} \sqrt{np}^k q^{n-k}}{\sqrt{2\pi k(n-k)} k^k e^{-k} (n-k)^{n-k} e^{-n+k}} = \frac{\left(\frac{p}{p_*}\right)^k \left(\frac{q}{q_*}\right)^{n-k}}{\sqrt{2\pi p_* q_*}} = \frac{e^{-nH(p_*)}}{\sqrt{2\pi p_* q_*}},$$

dove abbiamo portato il numeratore all'esponente, e la funzione H è data dalla relazione

$$H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}.$$

Nelle nostre ipotesi $|p_* - p| = \left|\frac{k}{n} - p\right| \leq a_n n^{-\frac{1}{3}} \rightarrow 0$ per $n \rightarrow \infty$, quindi possiamo scrivere lo sviluppo di Taylor della funzione $H(p_*)$ intorno al valore $p_* = p$. Per le derivate troviamo

$$H'(x) = \ln \frac{x}{p} - \ln \frac{1-x}{1-p}, \quad H''(x) = \frac{1}{x} + \frac{1}{1-x}, \quad H'''(x) = -\frac{1}{x^2} + \frac{1}{(1-x)^2}.$$

Dunque $H(p) = H'(p) = 0$, $H''(p) = \frac{1}{p} + \frac{1}{q} = \frac{1}{pq}$. Scrivendo lo sviluppo di Taylor al secondo ordine e moltiplicando per n troviamo

$$nH(p_*) = n \frac{(p_* - p)^2}{2pq} + \frac{H'''(\bar{p})}{3!} n (p_* - p)^3,$$

per qualche \bar{p} intermedio tra p_* e p .

Ora $n|p_* - p|^3 = \frac{|k-np|^3}{n^2} \leq a_n^3 \rightarrow 0$ per $n \rightarrow \infty$, mentre $n \frac{(p_* - p)^2}{2pq} = \frac{(k-np)^2}{2npq}$. A questo punto non sarebbe difficile dimostrare che il resto $r_n(k)$ tende a 0, uniformemente nella zona centrale.

Osservazione 2. Come si vede dalla (1.9.5), il massimo della probabilità è si ha quando $k \approx np$, perchè la funzione $e^{-\frac{x^2}{2}}$ ha il massimo per $x = 0$. Pertanto il massimo delle probabilità tende a zero per $n \rightarrow \infty$:

$$\max_k P(N = k) \approx \frac{1}{\sqrt{2n p q}} \rightarrow 0, \quad n \rightarrow \infty.$$

Oltre all'andamento asintotico delle singole probabilità, la legge di de Moivre-Laplace ci permette anche di dimostrare che il limite per $n \rightarrow \infty$ della distribuzione della variabile normalizzata $\frac{N-np}{\sqrt{npq}}$ è la probabilità gaussiana su \mathbb{R} vista nel precedente paragrafo. Questo è il senso del **teorema centrale del limite**, che si formula al seguente modo.

Teorema. (Teorema centrale del limite per la binomiale.) Nelle ipotesi del precedente teorema, per ogni scelta dei numeri reali $A < B$ si ha

$$\lim_{n \rightarrow \infty} P\left(\frac{N-np}{\sqrt{npq}} \in (A, B)\right) = \lim_{n \rightarrow \infty} \sum_{k: A \leq \frac{k-np}{\sqrt{npq}} \leq B} P(N=k) = \int_A^B \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \quad (1.9.6)$$

e la relazione vale anche per $A = -\infty$, o $B = +\infty$.

Dimostrazione (schema). Supponiamo A e B finiti e poniamo $z_k = \frac{k-np}{\sqrt{2npq}}$, e $g(x) = e^{-\frac{x^2}{2}}/\sqrt{2\pi}$. La distanza tra i punti z_k tende a 0 per $n \rightarrow \infty$: $z_{k+1} - z_k = \frac{1}{\sqrt{2npq}}$. Considerando che per i k tali che $A \leq z_k \leq B$ vale la legge di de Moivre-Laplace (1.9.5), otteniamo una somma integrale (somma di Riemann) dell'integrale definito tra A e B della funzione $g(x)$, che, come è noto, converge all'integrale

$$\sum_{k: A \leq \frac{k-np}{\sqrt{npq}} \leq B} P(N=k) \approx \sum_{k: A \leq z_k \leq B} \frac{e^{-\frac{z_k^2}{2}}}{\sqrt{2\pi npq}} \rightarrow \int_A^B g(x) dx.$$

Il segno \approx indica un'eguaglianza approssimativa tanto più esatta quanto più n è grande.

Non è difficile estendere il risultato anche agli intervalli infiniti.

La validità del teorema centrale del limite non è limitata alla binomiale, ma vale per una somma generale di variabili aleatorie indipendenti identicamente distribuite. Proprio per questa ragione esso acquista un'importanza fondamentale in probabilità e statistica.

Consideriamo il caso di una successione di variabili aleatorie F_1, F_2, \dots come quella vista nel paragrafo §1.9.1 per la legge dei grandi numeri, e cioè tale che le F_i siano indipendenti ed egualmente distribuite, con comune valor medio $M(F_i) = m$ e $\text{Var}(F_i) = \sigma^2$, $i = 1, 2, \dots$. Poniamo come sopra $S_n = F_1 + F_2 + \dots + F_n$.

Il teorema centrale del limite per una tale successione si formula nel modo seguente.

Teorema. (Teorema centrale del limite nel caso generale.) Nelle ipotesi precedenti, per ogni scelta dei numeri reali $A < B$ si ha

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \in (A, B)\right) = \int_A^B \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx. \quad (1.9.7)$$

e la relazione vale anche per $A = -\infty$, o $B = +\infty$.

La dimostrazione di questo teorema non è complicata, ma richiede strumenti matematici che esulano dal programma del corso.

1.9.3. Applicazioni del teorema centrale del limite alla statistica.

In conclusione vogliamo mostrare alcune semplici applicazioni del teorema centrale del limite alla statistica. Discuteremo solo il caso di un test statistico per un'ipotesi di distribuzione binomiale e la determinazione su un campione di una media incognita con varianza nota.

Quantili della gaussiana. Una particolare importanza in statistica hanno i quantili della distribuzione gaussiana standard. Dato $\alpha \in (0, 1)$ il quantile della gaussiana z_α è definito come il numero tale che

$$P(Z < z_\alpha) = \alpha. \quad (1.9.11)$$

E' chiaro che $z_{\frac{1}{2}} = 0$, e che $z_\alpha \rightarrow +\infty$ quando $\alpha \rightarrow 1$.

Esercizio 1. Il lettore faccia il grafico della densità gaussiana e riporti i valori di z_α per qualche valore di α .

Esercizio 2. Si lancia per 10.000 volte una moneta. Assumendo valida l'approssimazione gaussiana per la distribuzione binomiale del numero di teste, calcolare la probabilità che il numero di teste sia superiore a 5.100.

Test statistico su binomiale. Si vuole testare l'ipotesi se i due sessi siano egualmente distribuiti tra i pazienti affetti da una certa malattia M , e si raccoglie un campione estratto a caso di 1000 pazienti affetti dalla malattia, di cui 540 risultano maschi e 460 femmine.

Si tratta cioè di accettare o rigettare l'ipotesi che il risultato sia dovuto ad una distribuzione binomiale $B(1000; 1/2)$. Se l'ipotesi in questione (detta, in statistica, "ipotesi nulla"), fosse vera, detto N il numero dei maschi, avremmo $M(N) = 500$ e $Var(N) = \frac{1000}{4} = 250$. Il numero delle prove è certamente abbastanza grande da poter applicare il teorema del limite centrale (TLC) e quindi possiamo porre, trascurando la correzione,

$$\frac{N - np}{\sqrt{np(1-p)}} = \frac{N - 500}{\sqrt{250}} \approx Z \quad (1.9.10)$$

dove Z è la gaussiana standard.

E' chiaro che la variabile Z può assumere tutti i valori di \mathbb{R} . Bisogna perciò stabilire un "criterio di rigetto": se il valore trovato cade nella zona di bassa probabilità delle code della gaussiana l'ipotesi viene rigettata. Infatti nel caso che fosse vera, il valore da trovato sarebbe "eccezionale", cadrebbe in una zona di bassa probabilità in cui cade molto raramente.

Il criterio consiste nello scegliere un "livello di fiducia" $p_* \in (0, 1)$. I valori più frequenti nella pratica statistica sono $p_* = 0,01$, $p_* = 0,05$ e $p_* = 0,1$.

L'ipotesi nulla viene rigettata se il valore trovato della variabile (1.9.10) è a destra del quantile $z_{1-\frac{p_*}{2}}$ o a sinistra di $-z_{1-\frac{p_*}{2}}$, le due aree di coda ciascuna delle quali ha probabilità $\frac{p_*}{2}$. Per cui si ha

$$P(|Z| > z_{1-\frac{p_*}{2}}) = p_* \quad (1.9.12)$$

Il livello di fiducia p_* rappresenta quindi la probabilità di rigettare l'ipotesi nel caso che sia vera. L'ipotesi viene invece accettata se il valore cade nell'intervallo $(-z_{1-\frac{p_*}{2}}, z_{1-\frac{p_*}{2}})$.

E' chiaro che chi sceglie p_* molto piccolo è restio a rigettare l'ipotesi, mentre chi lo sceglie grande è maggiormente disposto a farlo.

2. Determinazione di media incognita con varianza nota. Supponiamo di avere un campione che proviene da n prove ripetute indipendenti, i cui risultati sono realizzazioni delle variabili casuali F_1, \dots, F_n . Supponiamo che la varianza $\text{Var}(F_j) = \sigma^2$ sia nota, ma la media m sia ignota.

Siano x_1, \dots, x_n i valori registrati. Sulla base della Legge dei Grandi Numeri del paragrafo 1.9.1 prendiamo come stima di m la media campionaria $\bar{m} = \frac{x_1 + \dots + x_n}{n}$, che è una realizzazione della variabile aleatoria $\frac{S_n}{n}$ dove

$$S_n = F_1 + \dots + F_n.$$

$\frac{S_n}{n}$ ha media m e varianza $\frac{\sigma^2}{n}$, e per il TLC sappiamo che

$$\frac{S_n}{n} - m = \frac{S_n - nm}{n} \approx \frac{\sigma}{\sqrt{n}}Z, \quad (1.9.13)$$

dove Z indica ancora la variabile gaussiana standard, e di nuovo supponiamo che n sia così grande che si può prendere l'eguaglianza.

A questo punto fissiamo, come nel caso precedente un livello di fiducia p_* , e calcoliamo il corrispondente quantile $z_{1-\frac{p_*}{2}}$, dato dalla (1.9.12). Sulla base della (1.9.13), e del valore \bar{m} di $\frac{S_n}{n}$ registrato, possiamo fare la seguente affermazione circa il valore incognito m della media.

Con livello di fiducia (o di significanza) p_ si ha $|m - \bar{m}| < z_{1-\frac{p_*}{2}} \frac{\sigma}{\sqrt{n}}$, cioè il valore vero m si trova nell'intervallo di fiducia*

$$\bar{m} - z_{1-\frac{p_*}{2}} \frac{\sigma}{\sqrt{n}} < m < \bar{m} + z_{1-\frac{p_*}{2}} \frac{\sigma}{\sqrt{n}}. \quad (1.9.14)$$

Se vogliamo rendere piccola la probabilità di sbagliare nel caso che l'ipotesi sia vera, cioè vogliamo rendere p_* piccolo, l'intervallo cresce. Se si vuole essere assolutamente sicuri, cioè per $p_* \rightarrow 0$, si ha $z_{1-\frac{p_*}{2}} \rightarrow +\infty$ e quindi l'intervallo diventa infinito, e il valore di m del tutto indeterminato.

Senza variare il livello di fiducia p_* la stima, come si vede dalla (1.9.14), diventa sempre più precisa, cioè l'intervallo sempre più piccolo, al crescere di n .