1 Breve introduzione alla probabilità elementare: approccio intuitivo

1.1

È usuale che in molte situazioni che si presentano concretamente ci sia a priori incertezza su ciò che accadrà nel futuro: il calcolo delle probabilità studia i modelli matematici delle cosidette "situazioni di incertezza".
Fanno parte del linguaggio quotidiano espressioni del tipo: "è probabile che domani piova", "è poco probabile che sarò promosso", "il test diagnostico è affidabile al 90 %".
Queste frasi evidenziano l'incertezza sul futuro legata a molte situazioni contingenti, contestualmente mettono in evidenza che la mancanza di certezza non significa mancanza totale di informazione.
Nelle situazioni concrete difficilmente si ha certezza sull'esito, ma neppure totale incertezza.
Non si può equiparare le affermazioni precedenti a dichiarazioni di assoluta mancanza di conoscenza circa gli esiti di un certo evento. Esse veicolano una qualche informazione: vogliono comunicare una qualche misura dell'incertezza.
Dicendo "è probabile che domani piova" intendiamo in qualche modo dire che ci sembra più plausibile per domani una situazione di pioggia piuttosto che una di sole, dicendo "il test diagnostico è affidabile al 90 % " ci spingiamo oltre: diamo una misura numerica dell'affidabilità del test, cioè della possibilità che il test dia un risultato corretto.

Esempio 1 *Facciamo un esempio più tecnico: supponiamo di avere una moneta con la scritta 1 sulla prima faccia e la scritta 2 sulla seconda faccia. Siamo interessati per qualche motivo a lanciarla e a fare una previsione sul fatto che la faccia che vedremo dopo il lancio sia quella con la scritta 1.*
Adesso supponiamo di avere un dado e di essere interessati a lanciarlo e a fare una previsione sul fatto che dopo il lancio la faccia superiore del dado sia quella con scritto 1.

In entrambi i casi ci troviamo di fronte a un esperimento con esito incerto, ma chiaramente l'incertezza non è completa: basta un attimo di riflessione per rendersi conto che il fatto che esca 1 è "più facile" (sulla base di queste informazioni) nel primo caso che nel secondo.

Per parlare di situazioni che ci troviamo spesso a dover considerare: nessuno di noi può essere certo di non contrarre una malattia polmonare, ma chi fuma è maggiormente a rischio. È possibile in qualche modo dare una misura di questo rischio?

La teoria della probabilità si propone per prima cosa di dare un modello formale per descrivere queste situazioni.

Essa è nata per applicazioni ai giochi d'azzardo ma oggi viene usata in molti settori: emissioni di elettroni, chiamate telefoniche, controllo di qualità, meccanica statistica, ereditarietà, teoria dei giochi, ecc....

Per prima cosa analizziamo la situazione classica più semplice: supponiamo di avere un esperimento con un numero finito di esiti possibili ed "equiprobabili". Questa è una situazione che si presenta per esempio quando non si sa nulla dell'esperimento e a priori l'unica cosa che si può pensare è che tutti i risultati hanno "la stessa dignità di essere veri".

Definizione 1 CLASSICA *Supponiamo che un certo esperimento abbia N casi possibili incompatibili tra di loro (il fatto che se ne verifichi uno esclude che possa verificarsene un altro) "equiprobabili" e che n di questi casi siano favorevoli al verificarsi di un certo evento. In tal caso si definisce probabilità dell'evento il rapporto n/N tra il numero n dei casi favorevoli al verificarsi dell'evento e il numero N dei casi possibili.*

Esempio 2 *Supponiamo di avere un dado che a priori pensiamo non truccato e di lanciarlo. La probabilità che lanciandolo esca 1 è 1/6. Questo numero si ottiene dividendo il numero dei possibili esiti del lancio che danno 1 per il numero totale dei possibili esiti. La probabilità che lanciandolo esca un numero maggiore di 2 è 4/6. Questo numero si ottiene dividendo il numero dei possibili esiti del lancio che danno un numero maggiore di 2 per il numero dei possibili esiti.*

Esempio 3 *Supponiamo di avere una moneta non truccata e ci chiediamo quale è la probabilità che lanciandola esca "TESTA". Ovviamente per simmetria tale probabilità è 1/2.*

Esempio 4 *Calcoliamo la probabilità che sabato prossimo il primo estratto sulla ruota di Napoli sia il numero 57. Poichè il primo estratto sulla ruota di Napoli è scelto tra 90 numeri, se non ci sono trucchi (tutti i numeri sono ugualmente possibili) i casi sono 90 e quello favorevole è 1, dunque la probabilità risulta essere* $\frac{1}{90}$.

Esempio 5 *Calcoliamo la probabilità che sabato prossimo il secondo estratto sulla ruota di Napoli sia il numero 57. Il ragionamento analogo a quello usato nell'esempio precedente porta a concludere che la probabilità è ancora* $\frac{1}{90}$.

Esempio 6 *Calcoliamo la probabilità che sabato prossimo il secondo estratto sulla ruota di Napoli sia il numero 57, sapendo che il primo estratto non è il numero 57. In questo caso dopo la prima estrazione i numeri disponibili non sono più 90, bensì 89 e quindi la probabilità è* $\frac{1}{89}$.

Osservazione 1 *I due esempi precedenti mostrano come* **la probabilità dipende in modo essenziale dall'informazione di chi la calcola***: infatti in entrambi si tratta di calcolare la probabilità dell'evento "il secondo estratto sulla ruota di Napoli è il numero 57", ma nel secondo esempio si ha una informazione in più.*

Esempio 7 *Calcoliamo la probabilità che sabato prossimo il secondo estratto sulla ruota di Napoli sia il numero 57, sapendo che il primo estratto è il numero 57. In questo caso dopo la prima estrazione il numero 57 non è più disponibile e quindi la probabilità è 0.*

Esempio 8 *Calcoliamo la probabilità che in una famiglia con due figli ci siano esattamente 0, 1, 2 maschi. I casi possibili sono* $\{MM, MF, FM, FF\}$ *dove M sta per "maschio" , F sta per "femmina" e per esempio MF significa "Il primo è maschio e il secondo è femmina". Allora i casi possibili sono 4 e dunque la probabilità che ci siano 0 maschi è* $\frac{1}{4}$*, esattamente come la probabilità che ci siano 2 maschi, mentre la probabilità che ci sia un maschio e una femmina è* $\frac{1}{2}$.

Esempio 9 *Calcoliamo la probabilità che lanciando contemporaneamente due dadi la somma delle uscite sia 5. I casi possibili sono tanti quante le coppie ordinate di numeri naturali compresi tra 1 e 6, cioè 36, mentre i casi favorevoli sono tutte le coppie che hanno per somma 5 e cioè (1,4), (2,3), (3,2), (4,1). Quindi la probabilità richiesta è* $\frac{4}{36}$ *cioè* $\frac{1}{9}$.

La critica più frequente alla definizione classica di probabilità è il fatto che viene data usando il concetto di eventi equiprobabili, quindi in un certo senso usando la definizione stessa di probabilità. D'altra parte non sembra così necessario conoscere già la probabilità, ma basta far riferimento al principio di indifferenza o di simmetria.

Una impostazione alternativa è **la definizione "frequentista".**

Definizione 2 FREQUENTISTA *Secondo questa impostazione la probabilità di un evento, per esempio l'uscita di testa nel lancio di una moneta, è il limite a cui tende il rapporto* n/N *dove* n *è il numero di teste e* N *è il numero di lanci totali, quando* N *tende all'infinito.*

La definizione frequentista presenta il problema che, essendo una definizione sostanzialmente "pratica" basata su un concetto di limite, per il calcolo effettivo presuppone infiniti esperimenti (fatto impossibile). Si potrebbe pensare che infinito è "un numero molto grande", ma effettuando un numero grande di lanci n/N difficilmente è proprio 1/ 2.
Vale la pena di provare effettivamente a lanciare una moneta per esempio 50 volte e contare il numero delle teste.(Sembra decisamente più facile che lanciando una moneta due volte compaiano una testa e una croce piuttosto che 25 teste e 25 croci lanciandola 50 volte).
Inoltre spesso nei casi concreti un esperimento non e ripetibile. D'altra parte ovvie considerazioni di simmetria ci portano a pensare che quella probabilità **deve** essere 1/2.

Un'altra possibile definizione è quella di **probabilità soggettiva**.

Definizione 3 SOGGETTIVA *Considerato un evento A, la probabilità $p(A)$ che un soggetto attribuisce all'evento A è un numero reale che misura il grado di fiducia che un individuo "coerente" attribuisce, secondo le sue informazioni e opinioni, all'avverarsi di A.*

Vale la pena di osservare che proprio questa è la definizione di probabilità a cui più spesso ricorriamo nelle nostre considerazioni quotidiane ("domani pioverà", "questa volta passerò l'esame", ecc.),anche se resta aperto il problema di "come" scegliere effettivamente il numero che rappresenta la probabilità di un evento.

Se siamo in situazione di simmetria una possibilità ragionevole è la scelta che abbiamo chiamato "classica" di cui l'impostazione soggettiva è una generalizzazione. Si può infatti pensare all'ambito classico come a una situazione soggettivamente ragionevole in molti casi di completa simmetria o di assenza di informazione.

Cerchiamo ora nel contesto soggettivo di precisare cosa si intende per "coerenza": in altre parole, accettando questo'approccio assumiamo che la probabilità sia in qualche modo una opinione personale, soggetta ad alcuni vincoli (di coerenza appunto) che si possono formalizzare. Una possibile formalizzazione, è l'assiomatica di Kolmogorov. Si tratta di descrivere le proprietà da richiedere perchè una misura del grado di credenza sia una probabilità: è poi interessante notare che è possibile dimostrare che anche la probabilità "frequentista" soddisfa gli assiomi di Kolmogorov.

Vediamo di descrivere brevemente l'idea.
Una situazione di incertezza dà luogo ad un insieme di possibili eventi: per esempio nel caso del lancio di un dado i possibili esiti danno l'insieme:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

che rappresenta tutte le possibili uscite del lancio di un dado, mentre nel caso del lancio di una moneta con una testa e una croce i possibili esiti danno l'insieme

$$\Omega = \{TESTA, CROCE\}$$

che rappresenta tutte le possibili uscite dopo il lancio della moneta.
Indichiamo in generale con Ω, che chiameremo evento certo, l'insieme di tutti i possibili esiti dell'esperimento : è infatti certo che dopo l'esperimento accada uno dei possibili esiti dell'esperimento stesso.
È naturale attribuire all'evento certo probabilità 1: in effetti quando si lancia un dado è certo che uscirà uno dei sei numeri che compaiono sulle facce e quando si lancia una moneta si è certi che uscirà o TESTA o CROCE.
I sottoinsiemi di Ω vengono invece interpretati come eventi aleatori: tra questi l'insieme vuoto è l'evento impossibile ed è naturale attribuirgli probabilità 0. Tutti i sottoinsiemi di Ω vengono detti eventi e ciascuno di essi ha una probabilità che è un numero compreso tra 0 e 1.
Cerchiamo allora di formalizzare la **definizione assiomatica** di probabilità, almeno nel caso di un insieme finito.

Definizione 4 ASSIOMATICA *Sia Ω un insieme finito, che chiameremo spazio degli eventi. Consideriamo poi l'insieme di tutti i sottoinsiemi di Ω. Si dice probabilità su Ω una funzione p che associa a ogni sottoinsieme A di Ω un numero reale che soddisfi i seguenti assiomi:*

- *1)* $0 \leq p(A) \leq 1$ *per ogni* $A \subseteq \Omega$,
- *2)* $p(\Omega) = 1, \quad p(\emptyset) = 0.$
- *3)* $p(A_1 \cup A_2) = p(A_1) + p(A_2)$ *per ogni coppia di sottoinsiemi* A_1 A_2 *contenuti in* Ω *tali che* $A_1 \cap A_2 = \emptyset$.

Definizione 5 *Si dice evento un qualunque sottoinsieme di Ω, in particolare Ω si dice evento certo.*

Definizione 6 *Si dice evento elementare ogni sottoinsieme di Ω formato da un unico elemento.*

Esempio 10 *Se lanciamo un dado l'evento "esce 2" è un evento elementare mentre l'evento "esce un numero pari" che corrisponde all'insieme $\{2, 4, 6\}$ non è un evento elementare.*

Definizione 7 *Due eventi A_1 e A_2 si dicono incompatibili se $A_1 \cap A_2 = \emptyset$.*

Definizione 8 *Un evento A_2 si dice complementare di A_1 se è il complementare di A_1 in Ω, cioè se A_2 è l'insieme di tutti gli elementi di Ω che non appartengono ad A_1..*

Esercizio 1 *In un'urna ci sono 4 palline nere e 5 bianche. Ne estraiamo una: qual'è la probabilità che sia bianca?*

Soluzione: Siano $N_1,\ N_2,\ N_3,\ N_4$ le 4 palline nere e $B_5,\ B_6,\ B_7,\ B_8,\ B_9$ le cinque palline bianche.

$$\Omega = \{N_1, ..., N_4,\ B_5, ...,\ B_9\}$$

Poiche non c'e trucco:

$$p(N_1) = p(N_2) = ... = p(B_5) = ...p(B_9) = \frac{1}{9}$$

Noi dobbiamo calcolare

$$p\{B_5,\ ..., B_9\} = p(B_5) + ... + p(B_9) = \frac{5}{9}$$

Quindi la probabilità è 5/9.

Esercizio 2 *Nelle ipotesi dell'esercizio precedente estraiamo dapprima dall'urna una pallina senza guardarla. Successivamente ne estraiamo un'altra, sempre senza guardare la prima. Qual'è la probabilità che la seconda pallina sia bianca?*

Soluzione: In questo caso possiamo scegliere come Ω l'insieme di tutte le coppie ordinate di palline. Ω contiene $9 \cdot 8$ elementi e, sempre per simmetria, ciascun elemento di Ω ha esattamente probabilità $\frac{1}{72}$. L'insieme di cui dobbiamo trovare la probabilità è quello formato dalle coppie con il secondo elemento bianco e quindi contiene $4 \cdot 5 + 5 \cdot 4 = 40$ elementi, dunque la probabilità e $\frac{1}{72} \cdot 40 = \frac{5}{9}$
Avremmo potuto trovare lo stesso risultato anche pensando che l'operazione effettuata prima della seconda estrazione è assolutamente ininfluente. In questo ordine di idee lo spazio degli eventi è

$$\Omega' = \{N_1, ..., N_4,\ B_5, ...,\ B_9\}$$

con probabilità $\frac{1}{9}$ per ogni elemento e l'evento di cui vogliamo calcolare la probabilità è $\{B_5, ...,\ B_9\}$

Osservazione 2 *Nei due metodi di risoluzione dell'esercizio* 2 *abbiamo usato due insiemi* Ω *e* Ω' *diversi per arrivare poi allo stesso risultato. Questo non deve meravigliare, perché lo spazio degli eventi non è dato a priori dal problema, ma fa parte del modello costruito da noi. In generale è lecito aspettarsi che più modelli possano essere conformi allo stesso problema oggetto di studio.*

Esercizio 3 *Nelle ipotesi dell'esercizio precedente, supponiamo questa volta di guardare la prima pallina estratta e scoprire che è bianca. Qual'è la probabilità che la seconda pallina sia bianca?*

Soluzione: In questo caso lo spazio degli eventi contiene 4 palline bianche e 4 nere e quindi la probabilità è $\frac{4}{8} = \frac{1}{2}$.

Osservazione 3 *Ancora possiamo osservare che la probabilità dipende essenzialmente dall'informazione di chi la calcola, infatti nei due esperimenti precedenti l'unica cosa che abbiamo variato è un dato di informazione.*

2 Probabilità condizionale, indipendenza, formula di Bayes

2.1 Probabilità condizionale

Ci preoccupiamo ora di studiare come varia la probabilità al variare della conoscenza, ovvero delle informazioni in possesso di chi la calcola. Non è la stessa cosa la probabilità che entro sera piova, e la probabilià che entro sera piova, sapendo che il cielo è completamente coperto. Possiamo in qualche modo pensare che la probabilità che un giorno fissato piova, se accettiamo una impostazione frequentista, sia ragionevolmente calcolata con il rapporto tra il numero delle giornate di pioggia di un anno e il numero (365) dei giorni dell'anno, mentre la probabilità che piova sapendo che il cielo è coperto, nella stessa impostazione, è rappresentata dal rapporto tra il numero dei giorni di pioggia in cui il cielo si presenta coperto e il numero dei giorni in cui il cielo si presenta coperto.

Facciamo un altro esempio: supponiamo di sapere che un'urna contiene due palline bianche, tre gialle e quattro rosse. In assenza di altre informazioni, la probabilità di estrarre una pallina gialla dall'urna è $\frac{3}{9}$. Ma qual'è la probabilità di estrarre una pallina gialla se sappiamo che la pallina è colorata? È come se noi estraessimo sì la pallina dall'urna che ne contiene 9, ma già sapendo che la pallina è o gialla o rossa e quindi come se estraessimo

da una ipotetica “sottourna” contenente 7 palline di cui 3 gialle e 4 rosse. In questo caso la probabilità diventa $\frac{3}{7}$, ed è diversa da quella che avevamo valutato prima. Questa nuova probabilità è la probabilità di estrarre una pallina gialla, sotto la condizione che la pallina estratta è colorata.
Cerchiamo di generalizzare questo esempio.
Supponiamo dapprima che tutti gli elementi di Ω siano equiprobabili, cioè supponiamo di essere nella situazione classica di simmetria. In questa ipotesi si ha che la probabilità che si verifichi A sotto la condizione che si sia già verificato B, che d'ora in poi indicheremo con $p(A/B)$ è dato dal numero degli elementi di $A \cap B$, che indicheremo con c$(A \cap B)$ diviso il numero degli elementi di B che indicheremo con $c(B)$. Se c(Ω) è il numero degli elementi di Ω si ha :

$$p(A/B) = \frac{c(A \cap B)}{c(B)} = \frac{\frac{c(A\cap B)}{c(\Omega)}}{\frac{c(B)}{c(\Omega)}} = \frac{p(A \cap B)}{p(B)}.$$

La definizione di probabilità condizionata può essere data anche nel caso più generale in cui non c'è simmetria e precisamente nel modo seguente:
Sia Ω lo spazio degli eventi che supporremo sempre finito, e sia p una misura di probabilità definita su Ω.

Definizione 9 *Si dice probabilità condizionale dell'evento A dato l'evento B (supposto che la probabilità di B sia diversa da* 0*) la quantità*

$$*) \;\; p(A/B) = \frac{p(A \cap B)}{p(B)}.$$

Osservazione 4 *La formula non è simmetrica in* A *e* B*, infatti se anche la probabilità di* A *è diversa da* 0 *può accadere che la probabilità di* A *dato* B*sia diversa dalla probabilità di* B *dato* A.

Dalla formula *) si ottiene subito l'identità

$$**) p(A \cap B) = p(A/B) \cdot p(B)$$

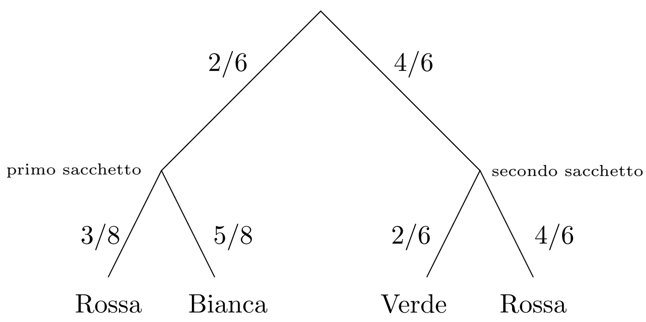
che è utile in molti casi pratici.
Nei casi concreti infatti è frequente che si conosca la probabilità di A dato B e la probabilità di B e sia interessante calcolare la probabilità che A e B accadano contemporaneamente, cioè la probabilità di $A \cap B$.

Esempio 11 *Supponiamo di dover calcolare la probabilità che in una estrazione della tombola sia uscito il numero 9, sapendo che è uscito un multiplo di 3.*

Poichè in una uscita della tombola i casi a priori possibili sono 90, l'informazione che il numero uscito è un multiplo di tre ci porta a dover restringere i casi possibili ai multipli di 3 tra 1 e 90, che sono esattamente 30, e dunque abbiamo un caso farovevole all'evento su 30 possibili ed equiprobabili, dunque la probabilità è 1/30.

Esempio 12 *Due sacchetti contengono il primo 5 palline bianche e 3 rosse, il secondo 4 palline rosse e 2 verdi. Tiriamo un dado: se esce 1 o 2 prendiamo una pallina dal primo sacchetto, se esce un altro numero la prendiamo dall'altro. Qual'è la probabilità che la pallina estratta sia bianca? sia rossa? sia verde?*

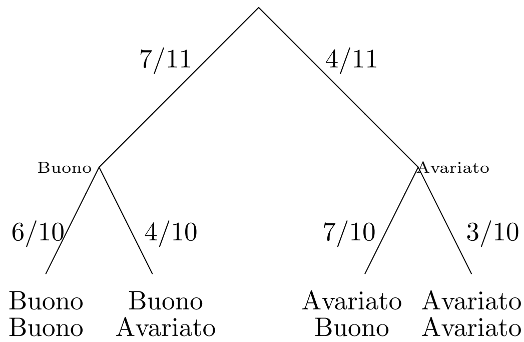
Anche in questo caso possiamo rappresentare la situazione con il grafo:



e dunque la probabilità che la pallina sia bianca è $\frac{2}{6} \cdot \frac{5}{8}$, quella che sia verde è $\frac{2}{6} \cdot \frac{4}{6}$ e quella che sia rossa è $\frac{2}{6} \cdot \frac{3}{8} + \frac{4}{6} \cdot \frac{4}{6}$.

Esempio 13 *Una scatola contiene quattro cioccolatini avariati e sette buoni. Quale è la probabilità che estraendone due a caso almeno uno di questi sia buono? E la probabilità che siano esattamente due buoni?*

Rappresentiamo l'esperimento come se scegliessimo successivamente i due cioccolatini. Nella prima scelta abbiamo due possibilità: con probabilità $\frac{4}{11}$ il cioccolatino è avariato e con probabilità $\frac{7}{11}$ è buono.
Se il primo cioccolatino scelto è avariato, nel scegliere il secondo abbiamo le due probabilità condizionate all'evento "il primo cioccolatino è avariato" rappresentate rispettivamente da $\frac{3}{10}$ per il caso che anche il secondo sia avariato e $\frac{7}{10}$ per il caso che il secondo sia buono. Se il primo cioccolatino è buono, abbiamo rispettivamente $\frac{4}{10}$ e $\frac{6}{10}$.
Tutto questo si rappresenta con più semplicità con il grafo:
Osserviamo che in ognuno dei vertici Avariato, Buono relativi all'estrazione del primo cioccolatino, la somma delle probabilità segnate sui rami che escono da quel vertice è 1. Questo fatto è ovvio se si pensa che la somma di



queste probabilità è la somma delle probabilità di tutti gli eventi possibili, ammesso che sia avvenuto l'evento che corrisponde al vertice.
Nel nostro caso, se vogliamo calcolare la probabilità dell'evento "almeno un cioccolatino è buono" , dobbiamo sommare le probabilità degli eventi corrispondenti e quindi otterremo:

$$\frac{4}{11} \cdot \frac{7}{10} + \frac{7}{11} \cdot \frac{4}{10} + \frac{7}{11} \cdot \frac{6}{10} \ldots\ldots\ldots$$

Non ci conviene fare questo conto perché l'evento "almeno un cioccolatino è buono" è complementare dell'evento "i cioccolatini sono tutti avariati", la cui probabilità è facilmente calcolabile:

$$\frac{4}{11} \cdot \frac{3}{10}$$

e dunque, poiché la probabilità del complementare di un evento si calcola con la formula:
$p(\text{complementare di } (A)) = 1 - p(A)$ la probabilità che stiamo cercando vale

$$1 - \frac{4}{11} \cdot \frac{3}{10}$$

Osserviamo che il grafo è molto comodo per descrivere le situazioni di probabilità condizionale.

2.2 Indipendenza

Definizione 10 **Si dice che due eventi A, B sono due eventi indipendenti se la probabilità che accadano entrambi è il prodotto della probabilità che accada A per la probabilità che accada B, cioè se vale la relazione:**

$$p(A \cap B) = p(A) \cdot p(B).$$

Per definizione di probabilità condizionale, due eventi A e B (se la probabilità di B è diversa da 0) sono indipendenti se e solo se la probabilità di A condizionata al verificarsi di B è uguale alla probabilità di B.

$$p(A/B) = p(A).$$

Osserviamo che la definizione di eventi indipendenti è simmetrica in A e B.

Dire che A e B sono indipendenti significa che sapere se B si è verificato o no non dà alcuna informazione che modifichi la previsione del verificarsi di A e analogamente sapere che A si è verificato non dà alcuna informazione che modifichi la previsione del verificarsi di B.

Esempio 14 *Qualunque sia A, l'insieme vuoto e A sono eventi indipendenti.*

Esempio 15 *Qualunque sia A l'insieme Ω e A sono indipendenti.*

Esempio 16 *Consideriamo il lancio di un dado . Gli eventi :*

- *esce un numero primo*
- *esce un numero divisibile per 3*

sono eventi indipendenti in quanto
p (esce un numero primo) =1/2
p (esce un numero divisibile per 3)=1/3.
L' unico numero primo divisibile per 3 è 3 e si ha : $p\{3\} = 1/6 = 1/2 \cdot 1/3$.

Osservazione 5 *Il concetto di eventi incompatibili è diverso da quello di eventi indipendenti. Se due eventi sono incompatibili e indipendenti, almeno uno di essi ha probabilità 0.*

Infatti per due eventi A e B indipendenti e incompatibili si ha:
$0 = p(\emptyset) = p(A \cap B) = p(A) \cdot p(B)$ e quindi almeno una tra $p(A)$ e $p(B)$ è 0.

2.3 Formula di Bayes

È esperienza di tutti i giorni variare il grado di fiducia nelle persone e nelle istituzioni a seconda dei riscontri nella realtà quotidiana.
Se a priori pensiamo che molto probabilmente un amico non è bugiardo, la terza volta che scopriamo che ci ha mentito, saremo certamente meno disposti a credergli per il futuro.

Ci poniamo il problema di come varia la probabilità assegnata a un certo evento, al crescere delle informazioni assunte: abbiamo già più volte osservato che nella valutazione della probabilità l' informazione gioca un ruolo fondamentale . La formula di Bayes dà una risposta a questo problema.

Facciamo un esempio tecnico: supponiamo di avere due urne uguali all'esterno, ma la prima (U_1) contiene 9 palline bianche e 1 nera, mentre la seconda (U_2) ne contiene 9 nere e 1 bianca. A priori possiamo scegliere una delle due urne in una situazione dall'esterno (per noi che ne possiamo vedere il contenuto) simmetrica.
Supponiamo di scegliere una delle due urne (con probabilità $\frac{1}{2}$ ciascuna). A questo punto estraiamo una pallina e la guardiamo: è bianca.

Domanda 1: siamo ancora disposti a pensare che ci troviamo di fronte all'urna U_1 con probabilità $\frac{1}{2}$?

E ancora. Estraiamo un'altra pallina dalla stessa urna senza reimbussolare la prima. È ancora bianca.

Domanda 2: siamo ancora disposti a pensare che ci troviamo di fronte all'urna U_1 con probabilità $\frac{1}{2}$?

La risposta alla domanda 2 è semplicissima: poichè l'urna U_2 contiene una sola pallina bianca, e invece sono state estratte due palline bianche, U_2 non è certamente stata scelta e quindi a posteriori U_2 ha probabilità 0 che vuol dire che è stata scelta certamente U_1.

La risposta alla domanda 1 è leggermente più complessa. Questa volta l'avere estratto una pallina bianca non dirime la questione, perché da entrambe le urne è possibile estrarre una pallina bianca, ma è anche chiaro che da U_1 è molto probabile estrarre una pallina bianca ($p = \frac{9}{10}$) mentre da U_2 è poco probabile ($p = \frac{1}{10}$).
Quindi il fatto che abbiamo osservato la pallina bianca fa "pendere la bilancia" verso l'urna U_1. Quanto?
Se le palline della prima urna fossero marcate con il numero 1 e quelle della seconda urna con il numero 2, estraendo una pallina e osservando che è bianca potremmo sapere da quale urna proviene. E con quale probabilità la pallina sarebbe marcata con 1? Le palline bianche sono 10 e quelle marcate con 1 sono 9. Dunque la probabilità che sulla pallina ci sia scritto il numero 1 è $\frac{9}{10}$, che coincide con la probabilità a posteriori (dopo aver osservato la pallina bianca) che l'urna fosse U_2.

Naturalmente occorre formalizzare meglio questi discorsi intuitivi.

Per far questo ci riferiamo come al solito allo spazio Ω degli eventi che supponiamo per ora sempre finito e consideriamo due eventi A e B contenuti in Ω, entrambi con probabilità positiva.

Per definizione di probabilità condizionata si ha:

$p(A \cap B) = p(A/B) \cdot p(B) = p(B/A) \cdot p(A)$ e dunque:

(**FORMULA DI BAYES**) $\quad p(A/B) = \dfrac{p(B/A) \cdot p(A)}{p(B)}$

Questa formula è detta **Formula di Bayes**

Per esempio nel caso delle due urne, noi conoscevamo dai dati del problema la probabilità di estrarre una pallina bianca se fosse stata scelta l'urna U_1 e invece volevamo calcolare la probabilità che l'urna fosse U_1 sapendo che era stata estratta una pallina bianca.

$$p(U_1/Bianca) = \frac{p(Bianca/U_1) \cdot p(U_1)}{p(Bianca)} = \frac{\frac{9}{10} \cdot \frac{1}{2}}{\frac{10}{20}} = \frac{9}{10}$$

Vogliamo ora scrivere in maniera un po' diversa la formula di Bayes, perché spesso $p(B)$ non è nota. Ci occorre dapprima un risultato

Osservazione 6 *Supponiamo che lo spazio degli eventi Ω sia unione di due sottospazi A_1 e A_2 disgiunti, entrambi con probabilità diversa da zero. Allora per ogni $B \subseteq \Omega$ con $p(B) > 0$ si ha:*
$p(B) = p(B/A_1) \cdot p(A_1) + p(B/A_2) \cdot p(A_2)$

Per dimostrare ciò scriviamo dapprima
$B = (B \cap A_1) \cup (B \cap A_2)$
sapendo che
$(B \cap A_1) \cap (B \cap A_2) = \emptyset$
Dunque, per le proprietà delle misure di probabilità e per la definizione di probabilità condizionale, si ha

$$p(B) = p(B \cap A_1) + p(B \cap A_2) = p(B/A_1) \cdot p(A_1) + p(B/A_2) \cdot p(A_2)$$

La formula di Bayes allora si può scrivere così:

$$p(A_1/B) = \frac{p(B/A_1) \cdot p(A_1)}{p(B/A_1) \cdot p(A_1) + p(B/A_2) \cdot p(A_2)}$$

ESEMPI di applicazione della formula di Bayes :

Esempio 17 *Supponiamo di giocare a dadi con una persona sconosciuta. Noi vinciamo se esce dispari e perdiamo se esce pari.*
A priori ci fidiamo abbastanza della persona con cui giochiamo, nel senso che attribuiamo al fatto che possa aver truccato il dado a suo favore (tutti numeri pari) probabilità $\frac{1}{100}$.
Dopo 10 *lanci è uscito sempre pari. Il nostro grado di fiducia nell'altro giocatore resta sempre lo stesso?*

In altre parole, come cambia la probabiltà $\frac{1}{100}$ di fronte a un ripetersi di fatti che farebbero propendere per una probabilità di trucco maggiore ?
Indichiamo con T e NT rispettivamente gli eventi "la moneta è truccata" e "la moneta non è truccata" e con $10P$ l'evento "esce 10 volte pari" . Applicando la formula di Bayes otteniamo:

$$p(T/10P) = \frac{p(10P/T) \cdot p(T)}{p(10P/T) \cdot p(T) + p(10P/NT) \cdot p(NT)} = \frac{\frac{1}{100}}{\frac{1}{100} + (\frac{1}{2})^{10} \cdot 99 \cdot \frac{1}{100}} =$$

$= \dfrac{1}{1 + 99 \cdot (\frac{1}{2})^{10}}$ che vale circa $\frac{91}{100}$

Ciò significa che dopo due uscite a noi sfavorevoli la probabilità $\frac{1}{100}$ che attribuivamo al fatto che lo sconosciuto barasse è diventata $\frac{91}{100}$.

3 Variabili aleatorie su spazi di probabilità finiti

3.1

Viene lanciata una moneta. Se esce testa vinco 100 euro, se esce croce non vinco niente. Quale è il valore della mia vincita?
Osserviamo che il valore della vincita dipende dal risultato dell'esperimento "lancio della moneta". Se esce testa, la vincita è 100, mentre se esce croce la vincita è 0. Ciò significa che la vincita non è un numero determinato, ma una funzione che assume un certo valore in corrispondenza dell'uscita testa e un altro valore in corrispondenza dell'uscita croce:
V(testa)= 100
V(croce)= 0,
dove V sta per vincita.

Vediamo altre situazioni di giochi in cui la vincita o il capitale dipendono dal risultato di un esperimento casuale.

Esempio 18 *Possiedo* 100 *euro e decido di giocarli alla roulette puntandoli sul numero* 7. *Quale è il valore del mio capitale dopo questa operazione? Quale è il valore della mia vincita?*

Naturalmente il valore del capitale non è noto a priori, ma dipende dal numero uscito alla roulette. Possiamo descrivere questa situazione con l'insieme Ω delle possibili uscite nel gioco della roulette, che sono i numeri tra 0 e 36:

$$\Omega = \{0, 1, 2, ..., 36\}$$

In caso di vittoria, e qundi se esce 7, il capitale sarà $100 \cdot 36 = 3600$. In caso di sconfitta, cioè se esce uno qualunque degli altri numeri, il capitale sarà 0. Il valore del capitale è quindi una funzione definita su Ω a valori reali che ad ogni numero diverso da 7 associa 0 e a 7 associa 3600.
Indicando tale funzione con C si ha:

$$C : \Omega \to \mathbf{R}$$

con $C(i) = \begin{cases} 0 & se \quad i \neq 7 \\ 3600 & se \quad i = 7 \end{cases}$
La vincita si ottiene invece sottraendo al capitale finale il capitale giocato.
Indicando la vincita con V si ha: $V(i) = \begin{cases} -100 & se \quad i \neq 7 \\ 3500 & se \quad i = 7 \end{cases}$.

Esempio 19 *Viene lanciato un dado, se il numero che compare sulla faccia superiore del dado è* 1 *mi vengono consegnati* 100 *euro, altrimenti me ne vengono consegnati* 10. *Quanti euro mi vengono consegnati dopo il lancio del dado?*

Anche questa volta il capitale che mi viene consegnato non è noto a priori ma dipende dal numero che è uscito sulla faccia superiore del dado. Questa volta

$$\Omega = \{0, 1, 2, 3, 4, 5, 6\}$$

e il capitale C vale $C(i) = \begin{cases} 100 & se \quad i = 1 \\ 10 & se \quad i \neq 1 \end{cases}$.

Esempio 20 *Giochiamo un euro sull'uscita di un ambo sulla ruota di Torino. Se l'ambo esce viene pagata* 250 *volte la posta, cioè* 250 *euro, se l'ambo non esce perdo semplicemente l'euro che ho puntato inizialmente. Quanto vinco?*

La vincita vale -1 se l'ambo non esce e vale 249 se l'ambo esce.

Esempio 21 *Due giocatori giocano a testa o croce con una moneta equilibrata. Vince chi per primo totalizza due vincite. Quanto dura la partita?*

Un insieme Ω adatto a modellizzare questa situazione è l'insieme dei possibili risultati delle prime 3 partite (osserviamo che dopo 3 partite il gioco è comunque finito). Un possibile spazio Ω degli eventi è dunque:

$$\Omega = \{TTT,\ TCT,\ TCC,\ TTC,\ CTT,\ CCT,\ CTC,\ CCC\}$$

con la misura di probabilità che, per ragioni di simmetria, vale $\frac{1}{8}$ su ogni evento elementare. La durata del gioco è una funzione definita su Ω che vale 2 nel caso TTC,TTC,CCT,CCC e 3 nei casi rimanenti.

$$Y : \Omega \to \mathbf{R}$$

con $Y(TTC) = 2$, $Y(TCT) = 3$,....

Definizione 11 *Dato uno spazio di probabilità Ω finito, si dice variabile aleatoria (brevemente v.a.) una funzione definita su Ω, cioè una funzione che dipende dal caso. D'ora in poi, faremo riferimento a v.a. a valori reali.*

Nel primo esempio la variabile aleatoria è il valore del capitale, nel secondo è il numero di euro che mi vengono consegnati, nel terzo è la vincita al lotto, nel quarto è la durata della partita.
Una variabile aleatoria si indica usualmente con una lettera maiuscola, per esempio C, V o X, e i suoi possibili valori (ricordiamo che abbiamo fatto l'ipotesi che Ω sia finito) con $x_1, x_2, ...x_n$.

Se x è un numero reale, si indica con :
$(X = x) = \{\omega \in \Omega \text{ tali che } X(\omega) = x\}$.
$(X \leq x) = \{\omega \in \Omega \text{ tali che } X(\omega) \leq x\}$.
Naturalmente :
$A_1 = (X = x_1), A_2 = (X = x_2), ..., A_n = (X = x_n)\}$
sono eventi .

Indichiamo con $p_1, p_2, ..., p_n$ le rispettive probabilità. Naturalmente $p_1 + p_2 + ... + p_n = 1$, infatti $p_1 + p_2 + ... + p_n$ è la probabilità che X assuma uno dei suoi possibili valori, cioè la probabilità dell'evento certo.

Definizione 12 La legge o distribuzione di una variabile aleatoria è l'insieme formato dai suoi valori $x_1,\ x_2, ...\ x_n$ e le corrispondenti probabilità $p_1, p_2, ...p_n$.

Osservazione 7 *Rappresentiamo nel modo seguente le leggi delle variabili aleatorie del primo e del quarto esempio:*

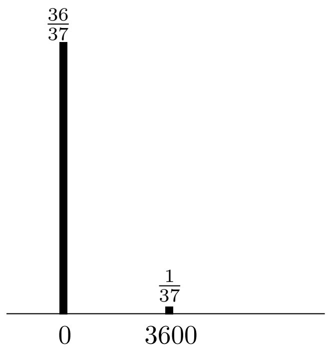


Figura 1: legge di X

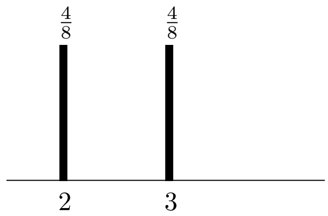


Figura 2: legge di Y

Osservazione 8 *La legge di una variabile aleatoria di fatto induce una misura di probabilità sull'insieme*

$$\{x_1, x_2, ...x_n\},$$

definita come $p_X(E) = p\{\omega \in \Omega \;\; t.c. \;\; X(\omega) \in E\}$.

In sostanza quindi una v.a. "trasporta" la struttura di spazio di probabilità da Ω a $\mathbf{R}$

Definizione 13 *Due variabili aleatorie che hanno la stessa legge si dicono* **equidistribuite.**

Definizione 14 **Si dice speranza matematica o media o valore atteso** *della variabile aleatoria X il numero*

$$E(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + ...x_n \cdot p_n.$$

Ovviamente variabili aleatorie equidistribuite hanno la stessa media.

Osservazione 9 *La media di una v.a. non è altro che una "media pesata" dei valori che essa ssume, dove i pesi sono le probabilità con cui un valore viene assunto.*

Esercizio 4 *Dimostrare che se* a *è un numero reale*

$$E(a \cdot X) = a \cdot E(x)$$

Infatti se X assume i valori $x_1, x_2, ..., x_n$ con probabilità rispettivamente $p_1, p_2, ..., p_n$, aX assume i valori $ax_1, ax_2, ..., ax_n$ con le stesse probabilità e dunque:

$$E(aX) = p_1ax_1 + p_2ax_2 + ... + p_nax_n = a(p_1x_1 + p_2x_2 + ... + p_nx_n) = aE(X)$$

Esercizio 5 *Analogamente si potrebbe dimostrare che se* X *e* Y *sono due variabili aleatorie*

$$E(X + Y) = E(X) + E(Y)$$

Esercizio 6 *Calcoliamo la media della variabile aleatoria del primo esempio. Si ha:*

$$E(X) = 0 \cdot 36/37 + 3600 \cdot 1/37 = 3600/37$$

Esercizio 7 *Calcoliamo la media della variabile aleatoria del secondo esempio:*

$$E(C) = 100 \cdot 1/6 + 10 \cdot 5/6 = 150/6 = 25$$

Esercizio 8 *Calcoliamo la media della variabile aleatoria del terzo esempio: La probabilità di indovinare un ambo su una ruota giocando due numeri (si potrebbe calcolare) è circa* $\frac{1}{400}$, *dunque la media è:*

$$249 \cdot 1/400 + (-1) \cdot 399/400 = -0,375$$

Osserviamo che non giocando si ha una vincita certa che vale 0 *e che quindi ha media* 0.

Esercizio 9 *Calcoliamo ora la media della variabile aleatoria del quarto esempio:*

$$E(Y) = 2 \cdot 4/8 + 3 \cdot 4/8 = 5/2.$$

Osservazione 10 *Osserviamo che in generale la media di una variabile aleatoria non coincide con alcuno dei valori che la variabile aleatoria assume, come accade in tutti gli esempi che abbiamo studiato.*

Definizione 15 Si dice varianza della variabile aleatoria X il numero

$$V(X) = E((X - E(X)^2)$$

Si può calcolare la varianza con la formula seguente:

$$V(X) = E(X^2) - (E(X))^2$$

La varianza misura quanto una variabile aleatoria è dispersa intorno alla sua media.

Esempio 22 *Se una variabile aleatoria è costante, la sua media è uguale al valore costante assunto dalla variabile aleatoria e la sua varianza è uguale a 0.*

Infatti se X è costante, assume solo un valore x_1 con probabilità 1 e dunque $E(X) = x_1$, X^2 assume solo il valore x_1^2 con probabilità 1 e dunque $V(X) = E((X - E(X)^2) = x_1^2 - x_1^2$.

Esempio 23 *Calcoliamo la varianza delle variabili aleatorie del primo e del quarto esempio*

$V(X) = \frac{(3600)^2}{37} - (\frac{3600}{37})^2$
$V(Y) = \frac{1}{4}$

Definizione 16 *Sia X una variabile aleatoria.*
Si dice **Funzione di ripartizione di** *X la funzione reale di variabile reale:*

$$F_X : R \longrightarrow R$$

così definita:

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

$F_X(x)$ è la probabilità che X sia minore o uguale di x.

La funzione di ripartizione F_X di una variabile aleatoria X possiede le seguenti proprietà:

a) per ogni $t \in \mathbf{R}$ si ha $0 \leq F(t) \leq 1$;

b) $\lim_{x\to-\infty} F_X(x) = 0 \quad \lim_{x\to+\infty} F_X(x) = 1$

c) se $x_1 \leq x_2 \quad F_X(x_1) \leq F_X(x_2)$ (è monotona debolmente crescente)

d) $F_X(x) = \lim_{t\to x+} F_X(t)$ (è continua a destra)

Disegnamo le funzioni di ripartizione delle v.a. del primo e del quarto esempio:

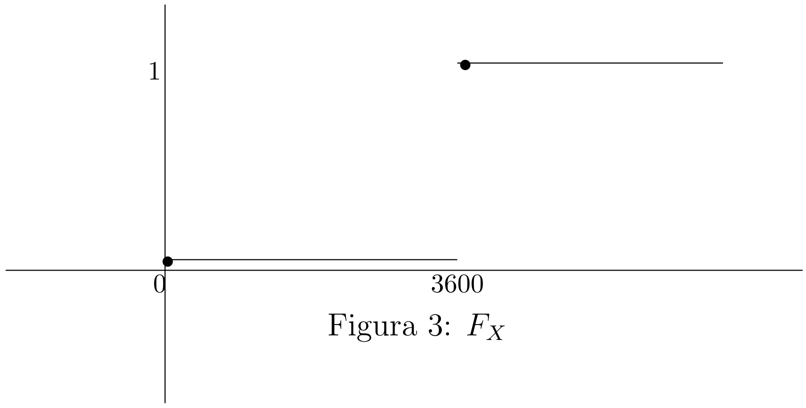


Figura 3: F_X

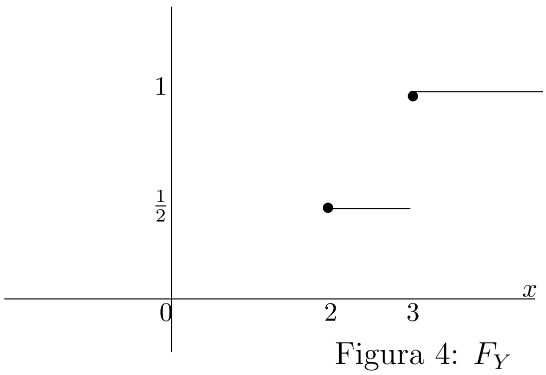


Figura 4: F_Y

Esempio 24 *Lancio una moneta non necessariamente equilibrata, in cui cioè p(T)=p e p(C)=q=1-p e considero la variabile aleatoria* X *definita su* $\Omega = \{T, C\}$ *con* $X(T) = 1$ *e* $X(C) = 0$*. In questo caso*

$$F_X(x) = \begin{cases} 0 & se \quad x < 0 \\ q & se \quad 0 \leq x < 1 \\ 1 & se \quad x \geq 1 \end{cases}$$

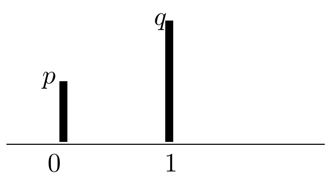


Figura 5: legge di X

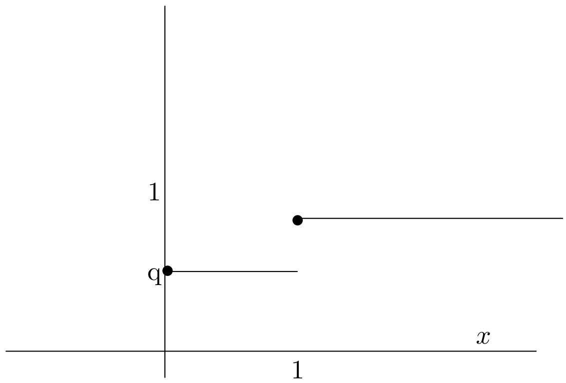


Figura 6: F_X

Esempio 25 *Lancio un dado e considero la v.a. definita su $\Omega = \{1, 2, ..., 6\}$ con $X(i) = i$, cioè X è semplicemente l'uscita del dado. Si ha*

$$F_X(x) = \begin{cases} 0 & se \quad x < 0 \\ \frac{1}{6} & se \quad 0 \le x < 1 \\ \frac{2}{6} & se \quad 1 \le x < 2 \\ \frac{3}{6} & se \quad 2 \le x < 3 \\ \dots\dots\dots & \\ \dots\dots\dots & \\ 1 & se \quad x \ge 1 \end{cases}$$

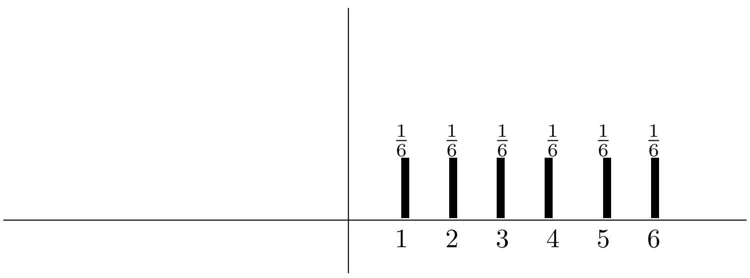


Figura 7: legge di X

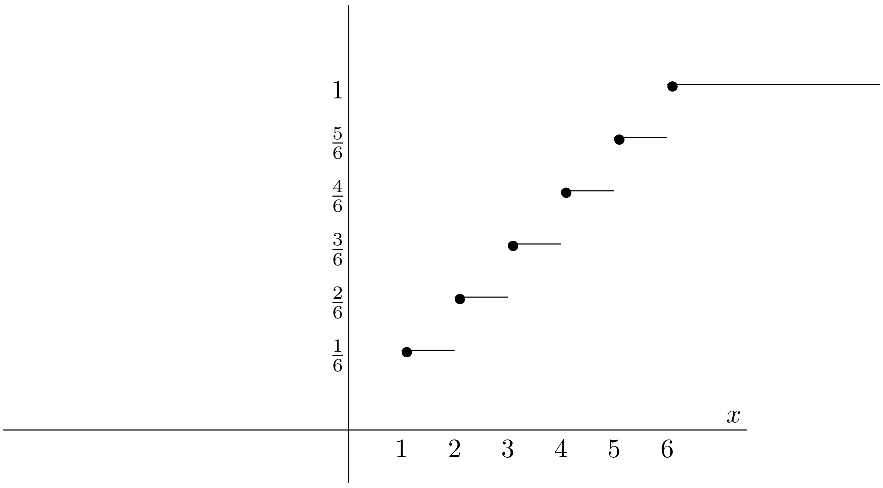


Figura 8: F_X

Le due situazioni di simmetria descritte nei due esempi precedenti danno luogo a variabili aleatorie che si dicono avere una:

Distribuzione di probabilità uniforme.

In tale situazione la funzione di ripartizione è una funzione a scala in cui l'altezza degli scalini è sempre uguale ed è uguale alla probabilità che X sia uguale all'ascissa dello scalino.

3.2 Alcuni esempi e applicazioni

Esempio 26 Paradosso di Simpson
*Si tratta di un esempio che mette in luce come è facile trarre conclusioni errate rispetto alla probabilità: è sempre necessario analizzare con attenzione i dati prima di trarre conclusioni. L'esempio prende il nome di "***paradosso di Simpson***" dal nome dello statistico che lo descrisse per la prima volta nel 1951*

In una stanza A ci sono 2 scatole, una bianca e una nera. Nella scatola bianca ci sono 90 caramelle di menta e 10 di liquirizia. Nella scatola nera ci sono 720 caramelle di menta e 180 di liquirizia. Nella stanza accanto (stanza B) ci sono altre 2 scatole, una bianca e una nera. Nella scatola bianca ci sono 160 caramelle di menta e 640 di liquirizia. Nella scatola nera ci sono 20 caramelle di menta e 180 di liquirizia. Gigi è un bambino goloso di caramelle di menta e odia la liquirizia. Viene portato nella stanza A e gli si dice di pescare una caramella da una delle due scatole (non si può vedere dentro le scatole), sapendo che è più facile pescare una caramella di menta dalla scatola *bianca (90% menta) che dalla scatola nera (80% menta). Naturalmente Gigi pesca la caramella dalla scatola bianca. Poi viene portato nella stanza B e si ripete lo stesso discorso. Di nuovo Gigi pesca la caramella dalla scatola bianca (20% menta) e non dalla scatola nera (10% menta). Fin qui niente di strano. Ma adesso Gigi viene portato in un'altra stanza (stanza C) dove ci sono altre due scatole, una bianca e una nera. Nella scatola bianca vengono versate tutte le caramelle delle 2 scatole bianche delle stanze A e B (rimpiazzando le due caramelle prese da Gigi) e nella scatola nera vengono versate tutte le caramelle delle altre due scatole nere. Senza dire altro si dice a Gigi di pescare di nuovo una caramella. da quale scatola la pescherà? Naturalmente dalla scatola bianca. Ma fa malissimo, perchè adesso la percentuale di caramelle di menta è molto più alta nella scatola nera (740 su 1100 circa il 67,3%) che nella scatola bianca (250 su 900 circa il 27,8%)! La cosa sembra paradossale, ma i numeri parlano chiaro.*

Il paradosso di Simpson è molto importante in statistica medica quando si deve indagare l'efficacia di nuovi farmaci. Infatti se un farmaco viene usato per curare una certa malattia e si testa la sua efficacia in 2 diversi campioni di pazienti, per esempio pazienti giovani-pazienti vecchi, può accadere che sia efficace sia per i pazienti giovani che per i pazienti vecchi, ma nei due campioni in percentuale diversa.. Mettendo insieme i dati di tutti i pazienti, giovani e vecchi, sembra che il farmaco non solo sia inefficace, ma addirittura sia dannoso.

Osservazione 11 *Quando si fanno esperimenti per testare l'efficacia di un farmaco, è necessario che nel gruppo di pazienti trattati le variabili che possono influenzare notevolmente la guarigione (nel nostro esempio l'età) siano distribuite in modo molto simile a come sono distribuite nei pazienti non trattati. Ciò equivale a dire che bisogna somministrare il farmaco ai giovani e ai vecchi nella stessa percentuale di casi. Questo può essere difficile da realizzare.*

Esempio 27 Rischio
Supponiamo di trovarci in una situazione in cui l'accadere di un evento **E** *comporta delle conseguenze negative per la popolazione: catastrofi naturali, emissione di gas nocivi, ... Supponiamo di conoscere la probabilità dell'evento (per esempio potremmo calcolarla con una definizione di tipo frequentistico come il numero delle volte in cui l'evento si è verificato diviso per il numero di giorni nei qualli abbiamo fatto l'osservazione). Supponiamo anche di conoscere la gravità delle conseguenze di* **E***, questa volta misurata come il numero m (magnitudo) delle conseguenze indesiderate che si verificano quando accade l'evento (potremmo calcolarlo come media aritmetica dei risultati*

ottenuti in tutte le osservazioni fatte). Lo spazio degli eventi di riferimento è $\Omega = \{\mathbf{E}\ ,\ non\ \mathbf{E}\ \}$.
Il rischio che comporta l'evento E si può definire come il prodotto del numero m delle conseguenze indesiderate per la probabilità di E e nel linguaggio probabilistico non è altro che il valore atteso della variabile aleatoria X con $X($ E $)=m$ e $X($non E $)=0$

Calcoliamo il rischio di morte in Italia derivante dall'esplosione di gas domestico. Sapendo che ci sono circa 200 esplosioni e che si ha circa un morto ogni 20 esplosioni e che la popolazione italiana è di 50.000.000 di individui si ha:

$$p = \frac{200}{50 \cdot 10^6} \qquad m = \frac{1}{20} \qquad R = \frac{200}{20 \cdot 50 \cdot 10^6} = \frac{2}{10^7}$$

cioè circa due morti ogni dieci milioni di abitanti. Naturalmente la parte più difficilmente quantificabile tra i parametri è la probabilità dell'evento, mentre la magnitudo è meno soggetta a errori.

Osservazione 12 *La percezione del rischio da parte del pubblico spesso non è legata esclusivamente a quello che abbiamo definito come rischio, ma è più legata al fattore m. A parità di uno stesso livello razionale di rischio il pubblico è maggiormente allarmato da eventi catastrofici con manifestazione singola e istantanea rara (incidenti aerei, esplosioni, ...) che da eventi cronici (incidenti sul lavoro, malattie da fumo , da alcoolismo,....). Si noti inoltre che la probabilità di un rischio in qualche modo ricorrente come l'esplosione di gas domestico si può ragionevolmente calcolare con un approccio frequentista, mentre per eventi più catastrofici e occasionali la frequenza è di difficile calcolo e quindi subentra una stima soggettiva anch'essa soggetta a errore e legata all'emotività.*

3.3 Distribuzioni di v.a. su spazi finiti

L'esempio più semplice di distribuzione di una v.a. è la cosidetta distribuzione di Bernoulli. Abbiamo già visto questa distribuzione studiando il lancio di una moneta non necessariamente equilibriata e la variabile aleatoria X definita su $\Omega = \{T, C\}$ come $X(T) = 1$ e $X(C) = 0$, in cui X rappresenta il numero delle teste osservate dopo il lancio. Una variabile aleatoria X che assume solo i valori 0 e 1 con $p(X = 1) = p$ si dice che ha una

Distribuzione di Bernoulli di parametro p

e si indica con $B(p)$.
Siamo interessati a calcolare la media e la varianza di una variabile aleatoria con distribuzione $B(p)$. Si ha:
$E(X) = p\ ,\ V(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)$
Siamo ora interessati a studiare altre distribuzioni di probabilità su spazi finiti

Esempio 28 *Lancio una moneta non necessariamente equilibrata, in cui cioè p(T)=p e p(C)=q=1-p n volte e considero come spazio degli eventi l'insieme di tutte le possibili n-ple di uscite, per esempio nel caso di un lancio avremo:*

$$\Omega = \{T, C\}$$

nel caso di due lanci

$$\Omega = \{TT, TC, CT, CC\}$$

nel caso di tre lanci

$$\Omega = \{TTT, TCT, CTT, CCT, TTC, TCC, CTC, CCC\}$$

e così via per il caso di più lanci. Osserviamo che nel caso di n lanci lo spazio Ω contiene esattamente 2^n elementi.

Questa è la classica situazione di indipendenza di eventi nel senso che l'evento "esce testa al k-esimo lancio" e "esce testa all' h-esimo lancio" se $h \neq k$ sono eventi indipendenti. (Osserviamo che l'evento "esce testa al k-esimo lancio" è rappresentato dal sottoinsieme di Ω formato da tutte le sequennze che hanno testa al k-esimo posto.) L'indipendenza ci permette di calcolare la probabilità dell'intersezione degli eventi come il prodotto delle probabilità dei singoli eventi. Per esempio nel caso dei 3 eventi la probabilità che esca la sequenza TCT è uguale al prodotto della probabilità che esca la prima volta testa per la probabilità che esca la seconda volta croce per la probabilità che esca la terza volta testa cioè $p \cdot q \cdot p = p^2 \cdot q$.
Siamo ora interessati a sapere quanto vale la probabilità che esca due volte testa e una volta croce in un ordine qualsiasi, cioè alla probabilità dell'evento:

$$\{TCT, CTT, TTC\}.$$

Per simmetria e per l'additività della misura di probabilità abbiamo:

$$p(\{TCT, CTT, TTC\}) = 3p^2 \cdot q.$$

Vogliamo ora generalizzare questo risultato, cioè stabilire in n lanci consecutivi di una moneta:

1) qual'è la probabilità che esca una ben determinata sequenza
2) qual'è la probabilità che esca esattamente k volte testa e quindi $n-k$ volte croce, in un ordine qualsiasi.
Il primo problema si risolve osservando che gli eventi corrispondenti agli n lanci sono tra loro indipendenti e quindi la probabilità dell'intersezione è uguale al prodotto delle probabilità: quindi se nella sequenza ci sono k teste e $n-k$ croci la probabilità dell'evento è $p^k \cdot q^{n-k}$
Per il secondo è necessario contare quante sono le sequenze distinte che contengono k teste e $n-k$ croci. Questo è un risultato di calcolo combinatorio:

Teorema 1 *Il numero delle sequenze di k teste e n-k croci è esattamente* $\binom{n}{k} = \frac{n!}{k!\cdot(n-k)!}$ *dove:*

$$k! = \begin{cases} 1 & se \quad k = 0 \\ k\cdot(k-1)\cdot ... \cdot 2\cdot 1 & se \quad k \neq 0 \end{cases}$$

In definitiva la probabilità che escano esattamente k teste e $n-k$ croci risulta essere:

$$\frac{n!}{k!\cdot(n-k)!}p^k q^{n-k}$$

Definizione 17 La variabile aleatoria X definita sullo spazio delle sequenze di n lanci di una moneta che associa a ogni lancio il numero delle teste uscite in quel lancio con

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

si dice distribuzione binomiale di parametro p e si indica con $B(n,p)$.

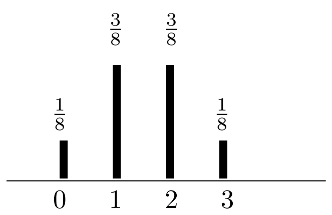


Figura 9: legge $B(3, \frac{1}{2})$

La distribuzione di probabilità binomiale si adatta a descrivere tutte quelle situazioni in cui un esperimento viene ripetuto n volte"indipendentemente l'una dall'altra“. Facciamo alcuni esempi:

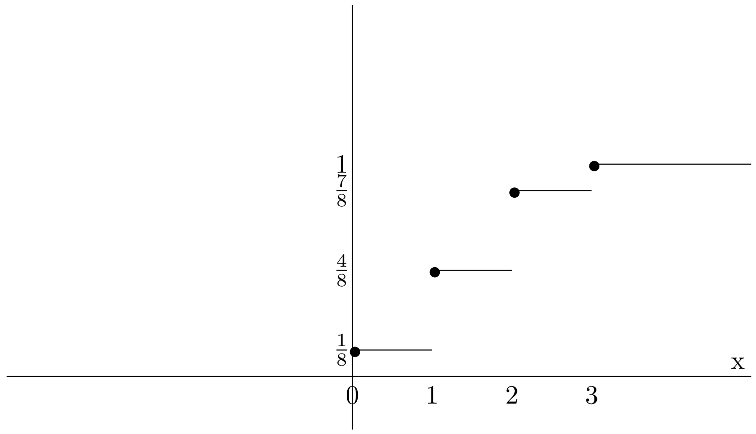


Figura 10: Funzione di ripartizione di$B(3, \frac{1}{2})$

Esempio 29 *Supponiamo di avere un'urna contenente n palline di cui r bianche e n-r nere. Estraiamo una pallina e registriamo se è bianca o nera, la reimbussoliamo e procediamo a una nuova estrazione. Registriamo se la nuova estrazione ha dato una pallina bianca o nera e procediamo sempre in questo modo sino ad aver effettuato n estrazioni. La distribuzione di probabilità relativa al numero di palline bianche estratte è una binomiale con* $p = \frac{r}{n}$ *e* $q = \frac{n-r}{n}$

Esempio 30 *La situazione analoga a quella dell'esempio precedente in cui la pallina non viene di volta in volta reimbussolata non dà luogo a una distribuzione binomiale, in quanto gli eventi corrispondenti alle successive estrazioni non sono più indipendenti.*

Calcoliamo ora la media di una variabile aleatoria $X = B(n, p)$. Basta osservare che X è la somma di n variabili aleatorie bernoulliane di parametro p indipendenti, ciascuna delle quali ha media p. Poiché la media di una somma è la somma delle medie, si ha $E(X) = np$.

3.4 Variabili aleatorie indipendenti

Come per le coppie di eventi, si può dare una analoga definizione di indipendenza per coppie di variabili aleatorie.

Definizione 18 *Siano* X *una variabile aleatoria definita su* Ω *a valori in* W_1 *e* Y *una variabile aleatoria definita su* Ω *a valori in* W_2. *Si dice che* X *e* Y *sono indipendenti se per ogni* $a \in W_1$ *e* $b \in W_2$ *si ha*
$p((X = a) \cap (Y = b)) = p(X = a) \cdot p(Y = b)$

In altre parole, l'indipendenza di v.a. non è altro che l'indipendenza di tutti gli eventi corrispondenti al fatto che la prima e la seconda v.a. assumono valori fissati.

Esempio 31 *Facciamo un esempio di due variabili aleatorie indipendenti Da un'urna contenente 6 palline numerate da 1 a 6 se ne estraggono 2 con rimpiazzo. Se indichiamo rispettivamente con X e Y i risultati delle due estrazioni, le due v.a. X e Y sono indipendenti.*

Esempio 32 *Facciamo ora un esempio di due variabili aleatorie che non sono indipendenti. Consideriamo la stessa estrazione dell'esempio precedente, ma senza rimpiazzo. Indichiamo rispettivamente con X_1, Y_1 i risultati delle due estrazioni. X_1 e Y_1 non sono indipendenti.*

3.5 Variabili aleatorie definite su un prodotto, leggi marginali

Vogliamo estendere il concetto di variabile aleatoria a funzioni a valori più in generale su $\mathbf{R^2}$ o ancora più in generale su uno spazio prodotto:

Definizione 19 - *La funzione*

$$Z = (X, Y): \quad \Omega \longrightarrow \mathbf{R^2}$$

è una variabile aleatoria se ciascuna delle sue componenti X, Y è una variabile aleatoria.
Analogamente la funzione

$$Z = (X, Y): \quad \Omega \longrightarrow W = W_1 \times W_2$$

è una variabile aleatoria se ciascuna delle sue componenti è una variabile aleatoria. Nel nostro caso, in cui Ω è finito, una variabile aleatoria è dunque una qualunque funzione definita su Ω a valori nello spazio prodotto $W_1 \times W_2$.

Se $Z = (X, Y)$ è una variabile aleatoria a valori in $\mathbf{R^2}$ e $z = (x, y) \in \mathbf{R^2}$, si ha :

$$\{Z = z\} = \{X = x\} \cap \{Y = y\}$$

Indichiamo con $z^{(1)}, z^{(2)}, ..., z^{(n)}$ i valori assunti da Z e con

$$p(z^{(k)}) = p\{Z = z^{(k)}\}.$$

Come abbiamo già visto nel caso di variabili aleatorie a valori reali, possiamo vedere W come uno spazio di probabilità. Più precisamente, abbiamo che $(W, P(W)), p_X)$ è uno spazio di probabilità. Basta definire, dato $E \subseteq W$:

$$p_X(E) = p\{\omega \in \Omega : X(\omega) \in E\}.$$

Si verifica facilmente che p_X è effettivamente una misura di probabilità.
Naturalmente, una variabile aleatoria a valori su $\mathbf{R}$ o su $\mathbf{R^2}$ induce su questi insiemi una misura di probabilità.
Consideriamo una variabile aleatoria Z a valori in un prodotto $W_1 \times W_2$. Disinteressan doci un attimo del fatto che $W_1 \times W_2$ è un prodotto, è evidente che possiamo usare la solita definizione per trovare la legge di Z. Si ha semplicemente

$$p_Z(E) = p\{\omega \in \Omega : Z(\omega) \in E\}.$$

Tuttavia, essendo $W_1 \times W_2$ un prodotto, assegnare Z equivale ad assegnare due funzioni. la prima a valori in W_1 e la seconda a valori in W_2:

$$X \ : \Omega \longrightarrow W_1$$

e

$$Y \ : \Omega \longrightarrow W_2$$

È allora legittimo chiedersi se ci sia una relazione tra le leggi di queste tre funzioni X, Y, Z.

Definizione 20 *Data Z , le leggi di X e Y vengono dette leggi marginali della legge di Z.*

Affrontiamo ora il problema se ci sia una relazione tra le leggi di X, Y e Z. Dobbiamo fare attenzione al fatto che p_X e p_Y sono leggi definite risp. su W_1 e W_2, mentre p_Z è definita su $W_1 \times W_2$. È evidente però che $p_X(E_1) = p_Z(E_1 \times W_2)$ e analogamente $p_Y(E_2) = p_Z(W_1 \times E_2)$.
Risulta allora evidente che, data p_Z, possiamo calcolarci facilmente p_X e p_Y. Che non sia altrettanto facile la strada a rovescio lo si può vedere già dal fatto che p_Z è definita su tutti i sottoinsiemi di $W_1 \times W_2$ e non tutti i sottoinsiemi di $W_1 \times W_2$ sono del tipo $E_1 \times E_2$.

Esempio 33

$$W_1 = \{1, 2\} \quad W_2 = \{1, 2\} \quad E = \{(1, 1), (2, 2)\}$$

E non è del tipo $E_1 \times E_2$. Infatti dovrebbe essere $E_1 = \{1, 2\}$ e $E_2 = \{1, 2\}$, dunque $E_1 \times E_2 = W_1 \times W_2$ che è falso.
Tuttavia questa è solo una difficoltà apparente e possiamo agevolmente superarla nel contesto semplice nel quale ci siamo messi e cioè con Ω finito. In ogni caso, anche qualora prendessimo $E = E_1 \times E_2$ non possiamo ricostruire $p_Z(E)$ dalla conoscenza di $p_X(E_1)$ e $p_Y(E_2)$.

Esempio 34 *Facciamo un esempio di due variabili aleatorie con leggi marginali uguali che danno luogo alla stessa legge congiunta. Da un'urna contenente 6 palline numerate da 1 a 6 se ne estraggono 2 con rimpiazzo. Se indichiamo rispettivamente con X e Y i risultati delle due estrazioni e calcoliamo la variabile aleatoria $Z = (X, Y)$, i possibili valori di Z sono coppie (i, j) dove i e j possono prendere i valori interi da 1 a 6. Si tratta di 36 valori possibili e ciascuno è assunto con probabilità 1/36. Le leggi marginali X e Y sono leggi uniformi su $\{1, 2, ..., 6\}$.*
Il successivo esperimento è la stessa estrazione ma senza rimpiazzo. Indichiamo rispettivamente con X_1, Y_1 i risultati delle due estrazioni e calcoliamo le leggi di

$$X_1,\ Y_1,\ e\ Z_1 = (X_1, Y_1)$$

I valori di Z non sono gli stessi, perchè ad esempio il risultato (1,1) non è più possibile. Sono quindi solo 30 i valori assunti da Z, tutti equiprobabili, e ciascuno ha probabilità 1/30. D'altra parte anche X_1, Y_1 hanno la legge uniforme .
Siamo dunque in presenza di due leggi congiunte diverse aventi le stesse marginali.

A questo punto osserviamo che se ci poniamo il problema di costruire p_Z conoscendo p_X e p_Y con l'ulteriore condizione che X e Y siano indipendenti, il problema ha un'unica soluzione (lo garantisce la definizione stessa di indipendenza) Infatti in tal caso

$$p_Z(E) = p_Z(E_1 \times E_2) = p_Z((E_1 \times Y) \cap (X \times E_2)) =$$

$$= p_Z(E_1 \times Y) \cdot p_Z(X \times E_2) = p_X(E_1) \cdot p_Y(E_2).$$

Esempio 35 *Sia $\Omega = \{a, b, c\}$ con $p(\omega) = 1/3\ \forall \omega \in \Omega$. Siano poi $W_1 = \{T, B\}$ e $W_2 = \{L, R\}$, $W = W_1 \times W_2$ e:*

$$Z :\ \Omega \longrightarrow W$$

così definita:

$$Z(a) = (T, L) \quad Z(b) = (T, R)\ Z(c) = (B, L).$$

Chi sono $p_Z, \quad p_X, \quad p_Y$*? X e Y sono indipendenti? Trovare* X_1 *e* Y_1 *con le stesse marginali ma indipendenti.*

Esempio 36 *Sia* $\Omega = \{a, b\}$ con $p(\omega) = \frac{1}{2} \quad \forall \omega \in \Omega$. *Siano* $X(a) = 1, \ X(b) = 2, \ Y(a) = 3, Y(b) = 4$. *Trovare* p_Z *assumendo che* X *e* Y *siano indipendenti.*