Il calcolo delle probabilità

La razionalizzazione delle scelte di fronte ai fenomeni casuali

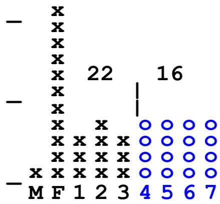
1. *Dalla statistica alla probabilità*
2. *Che cos'è la probabilità: misure di probabilità, eventi e variabili casuali*
3. *Leggi di distribuzione. Il generatore di numeri pseudocasuali*
4. *Modi per calcolare probabilità*
5. *Eventi (probabilisticamente) dipendenti e indipendenti*
6. *Usi (e limiti) del calcolo delle probabilità*
7. *Esercizi*

➜ *Sintesi*

1. Dalla statistica alla probabilità

Riprendiamo gli esempi discussi nel ➜ §4 della scheda *La matematica tra gioco e realtà.*
Abbiamo considerato la situazione:

(A) Sto giocando a *sette e mezzo* e sono il primo di mano. Ho **1**♦ e 3♣. Mi conviene chiedere carta?



I valori delle 38 carte rimanenti (tutte meno due) sono distribuiti secondo l'istogramma a fianco ("M" è la matta, Q♥, e "F" indica il valore 1/2 attribuito alle figure diverse da Q♥). Le carte che mi fanno sballare sono le 16 carte di valore maggiore o uguale a 4. La percentuale delle carte di valore maggiore o uguale a 4 è: 16 / 38 = 42.1%.

Questa è la probabilità che il valore C della nuova carta sia maggiore o uguale a 4. In simboli:

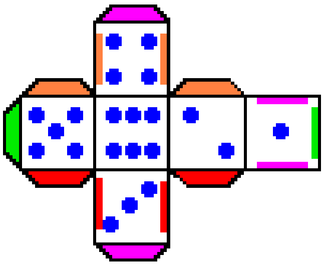
$$\Pr(C \geq 4) = 42.1\%$$

Questa è anche la probabilità di sballare. È minore del 50%. Quindi è più probabile che non sballi.

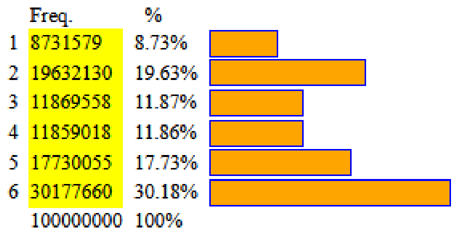
Qui trovi uno studio sperimentale della probabilità dell'evento effettuata con uno script: viene ***simulato*** il fenomeno e ripetuto per un certo numero di volte, e viene calcolata la frequenza relativa con cui si verifica l'evento. Prova ad eseguire questa simulazione (per un numero crescente di volte), confrontando il valore ottenuto con la probabilità ottenuta sopra.

Abbiamo poi affrontato questa situazione:

(B) Un amico mi dà un dado. Lo lancio. Esce 6. Posso ritenere che, se lo rilancio, l'uscita più probabile sia un altro 6?



Abbiamo discusso il problema alla luce del seguente esperimento: abbiamo costruito un dado utilizzando lo sviluppo a cui si può accedere cliccando **qui**, e che vedi riprodotto in piccolo a sinistra. Abbiamo effettuato molti lanci, ne abbiamo raccolto gli esiti in un'unica tabella e abbiamo, quindi, costruito il relativo istogramma, che abbiamo ottenuto abbastanza simile a quello raffigurato a destra, ottenuto con lo script **Dado**.



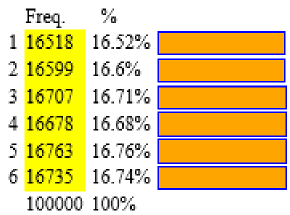
Indichiamo con U l'uscita del dado. L'istogramma ci ha fatto ritenere che per il nostro dado U=6 è più probabile di U=1, U=2, …, U=5, e quindi che il dado non sia ***equo***.

Di fronte a (**B**):
(1) C'è chi ha sostenuto che essendo già uscito 6 l'uscita successiva più probabile è di nuovo 6.
(2) C'è chi ha sostenuto che essendo già uscito 6 l'uscita successiva più probabile non è più 6.
(3) C'è chi ha sostenuto che ogni lancio fa storia a sé, per cui il fatto che sia già uscito 6 non mi dà alcuna informazione sulla maggior probabilità di un'uscita rispetto a un'altra.

L'opinione (3) ha qualche fondamento: effettivamente ogni lancio fa storia a sé, cioè la nuova uscita non è influenzata dalla precedente. In particolare se è uscito 6 può, con la stessa probabilità del lancio precedente, uscire nuovamente 6. Quindi l'opinione (2) è errata. Ma la seconda parte dell'opinione (3) («per cui …») non è giusta: l'uscita 6 mi dà qualche informazione. Infatti ad esempio mi permette di concludere che il dado non è truccato in modo tale che il 6 non esca mai.

L'opinione (3) sarebbe corretta anche nella sua conclusione se si sapesse che il dado è equo, cioè che le sue facce hanno tutte la stessa probabilità di uscita. *Come si fa a stabilire l'equità di un dado*?

Per costruire un dado equo occorre impiegare un materiale omogeneo, dargli forma regolare, …. Il nostro dado non lo era in quanto la presenza di linguette e pezzi di scotch più su certe facce che su altre faceva sì che esso non fosse equilibrato. Per assicurarsi dell'equità di un dado occorre tuttavia fare anche una verifica sperimentale: sottoporre il dado a moltissimi lanci e controllare se l'istogramma di distribuzione tende ad assumere una forma simile a quella a lato, cioè tende ad avere 6 colonne di eguale altezza. **DadoEquo** è uno script che simula il lancio di un dado equo.



In ogni caso non potremo avere una certezza assoluta dell'equità del dado. Anche se il dado è realmente equo, noi, essendo in grado di effettuare solo una quantità finita di prove, non potremo mai concludere che l'istogramma sperimentale ***tende*** effettivamente ***a stabilizzarsi*** su un insieme di colonne di altezza uguale. Potremo solo valutare la probabilità che ciò accada.

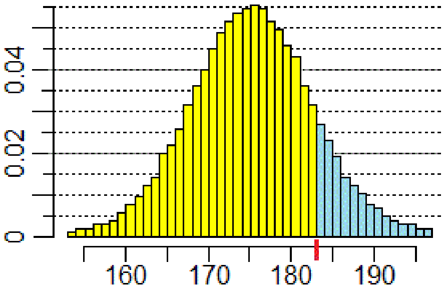
Ad esempio, se il dado fosse equo, con 1 milione di lanci potremmo concludere che al 99% la probabilità che esca 2 è compresa tra 0.166 e 0.168, con 100 milioni potremmo concludere che al 99% è compresa tra 0.1666 e 0.1668, … , ma non arriveremo mai né alla certezza, né a tutte le cifre di 1/6 = 0.1666…

Nei prossimi anni studierai concetti che ti permetteranno di esprimere meglio ciò che qui si è descritto con «tende a stabilizzarsi» e a valutare la probabilità degli scarti tra valori teorici e valori sperimentali.

Consideriamo una terza situazione:

(C) Siamo nello stato XX nell'anno AA. Un grande mobilificio, rinnovando la gamma dei suoi prodotti, vuole adattare il "formato" di alcuni mobili (letti, poltroncine da scrivania, …) alle caratteristiche fisiche della popolazione attuale e, a tal fine, si avvale della consulenza della società statistica SifanStat. Questa utilizza gli esisti di una recente indagine sulle altezze dei maschi ventenni di XX, riportati nell'istogramma seguente; le altezze sono classificate in intervalli ampi 1 cm; sull'asse verticale sono indicate le frequenze relative (cioè la densità di frequenza). Per avere un'idea del tipo di studi che fa la SifanStat consideriamo un problema semplice.

I letti matrimoniali prodotti finora dal mobilificio sono (internamente) lunghi 190 cm, adatti alla sistemazione comoda (tenendo conto dello spazio per il cuscino e per la rimboccatura inferiore) di un uomo alto meno di 183 cm. Utilizzando i dati sulle altezze citati, la SifanStat vuole valutare la probabilità che un potenziale cliente ventenne (maschio) trovi i letti matrimoniali troppo corti.



In assenza di ulteriori informazioni, la SifanStat assume che la probabilità che l'altezza cada in un dato intervallo coincida con la frequenza relativa di tale intervallo. Per es. assume che la probabilità che un cliente ventenne sia alto 183 cm (misura troncata agli interi) coincida con la frequenza 0.027 = 2.7% di tale altezza che si può leggere sull'istogramma.

Dall'istogramma posso trovare che la percentuale di coloro che sono alti 183 cm o più è, circa: 0.027 + 0.023 + 0.019 + 0.014 + 0.012 + 0.010 + 0.007 + 0.006 + 0.005 + 0.004 + 0.002 + 0.002 + 0.001 + 0.001 = 0.133 ≈ 13%.

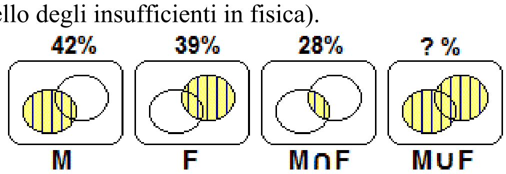
Questa ipotesi presuppone, ad es., che i clienti non provengano da particolari regioni del paese: se nella regione AA i maschi adulti sono mediamente più alti rispetto all'intero paese, e se i clienti venissero tutti da AA, le valutazioni basate sui nostri dati sarebbero sballate.

Anche nel caso delle altezze di una popolazione di persone, quando non si conoscano tutti i dati, si ricorre a uno studio sperimentale della distribuzione analogo a quello visto per il dado: si effettua un ***campionamento*** e si esamina la distribuzione delle altezze del campione. Più il campione è numeroso più (e se è fatto con opportuni criteri) l'istogramma si avvicina a (cioè: meglio approssima) l'istogramma dell'intera popolazione. A differenza della distribuzione delle uscite del dado, il cui l'istogramma "limite" è puramente teorico (i lanci effettuabili sono infiniti, almeno se il dado non si consuma, se …), nel caso delle altezze l'istogramma limite corrisponde all'istogramma "sperimentale" che otterrei esaminando l'intera popolazione (esame in via di principio realizzabile in quanto la popolazione è finita).

Un altro caso di uso di dati sperimentali per valutare delle probabilità è quello in cui si ricorre a dati relativi a come un fenomeno si è realizzato in tempi precedenti per prevedere come esso si realizzerà in un tempo futuro. Consideriamo ad esempio la seguente situazione:

(D) Conosco le percentuali degli studenti con *insufficienze* in matematica (42%), in fisica (39%) e in entrambe le materie (28%) del primo quadrimestre dell'anno passato nella scuola KK. Voglio valutare la probabilità che quest'anno uno studente debba essere coinvolto in corsi di recupero nell'area matematico-fisica, cioè valutare la probabilità che S∈M **or** S∈F (ho indicato con S un generico studente, con M l'insieme di quelli insufficienti in matematica e con F quello degli insufficienti in fisica).

Rispondi alla domanda posta aiutandoti con i diagrammi a lato.



Consideriamo ora una valutazione probabilistica in cui non ha neanche senso porsi il problema di uno studio sperimentale:

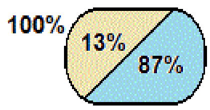
(E) Sta per disputarsi la partita Roma-Lazio. Gigi ritiene che la Roma 25 su 100 vincerà e 40 su 100 pareggerà. Qual è la probabilità per Gigi che vinca la Lazio?

Rispondi alla domanda.

(**E**) è una situazione diversa dalle precedenti: non abbiamo a che fare con una frequenza, anche se ci si esprime con il linguaggio "delle frequenze" («30 su 100», …). Infatti Gigi valuta la probabilità di vittoria in base alle sue valutazioni sullo stato di forma delle due squadre, sulle condizioni di salute dei giocatori, … e in base alle sue speranze. Avrebbe senso ricorrere a una valutazione basata sulle frequenze solo se non si disponesse di informazioni sull'andamento del campionato, non ci si basasse sulle proprie aspettative,… e se si avessero a disposizione i risultati degli incontri precedenti tra le due squadre.

2. Che cos'è la probabilità: misure di probabilità, eventi e variabili casuali

Riprendiamo, in breve, alcuni degli esempi visti; indicheremo con **Pr(E)** la probabilità che si verifichi l'evento **E**.



• Nella situazione ➔ **C** abbiamo visto che, indicata con H l'altezza di un cliente ventenne, Pr(H ≥ 183) = 13%.
Qual è la probabilità che un cliente sia più basso di 183 cm?
Pr(H < 183) = 100% − Pr(H ≥ 183) = 100%−13% = 87%.

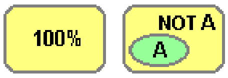
• La situazione ➔ **E**, indicato con R il risultato della partita ("1", "2" o "X"), può essere sintetizzata così:
Poiché Pr(R = "1") + Pr(R = "2") + Pr(R = "X") = 100%,
Pr(R = "2") = 100% − Pr(R = "1") − Pr(R = "X") = 100% − 25% − 40% = 35%.

• Nel caso del lancio di un dado (➔ **B**) ritenere che il dado sia ***equo*** significa supporre che l'uscita U abbia uguale probabilità di essere 1, 2, … o 6: Pr(U=1) = Pr(U=2) = … = Pr(U=6). Sia P il valore di questa probabilità. Poiché Pr(U=1) + Pr(U=2) + … + Pr(U=6) = 100% = 1, ho P + P + P + P + P + P = 6P = 1, da cui P = 1/6.
Per trovare la probabilità che l'uscita sia pari faccio: Pr(U è pari) = Pr(U=2) + Pr(U=4)+ Pr(U=6) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2.

Nei primi due esempi ho associato ad alcuni eventi *A* un numero compreso *tra 0 e 1* (=100%) come **Pr**(*A*) (*probabilità* di *A*). Nel terzo ho fissato delle condizioni sulla funzione *A* → **Pr**(*A*): ho supposto che Pr(U=1) = Pr(U=2) = ….

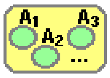
In tutti i casi ho poi dedotto le probabilità relative ad altri eventi applicando a **Pr** alcune delle ***proprietà*** che si erano già usate per le *frequenze percentuali*.

Rivediamo più sistematicamente queste proprietà.



- Pr(**not A**) = 100% – Pr(**A**)
- Pr(**A or not A**) = 100% =1
- Pr(**A and not A**) = 0

Esempio: Pr(H < 183) = Pr(**not** H ≥ 183) = 100% – Pr(H ≥ 183)



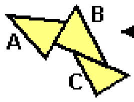
- Pr($\mathbf{A}_1$ **or** $\mathbf{A}_2$ **or** $\mathbf{A}_3$ **or** …) = Pr($\mathbf{A}_1$) + Pr($\mathbf{A}_2$) + Pr($\mathbf{A}_3$) + …

se $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, … sono tra loro ***incompatibili***, cioè se due qualunque eventi $\mathbf{A}_i$ e $\mathbf{A}_j$ non possono essere veri contemporaneamente.

Esempio: Pr(U è pari) = Pr(U=2 **or** U=4 **or** U=6) = Pr(U=2)+Pr(U=4)+Pr(U=6)

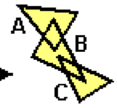
Quest'ultima proprietà è nota come ***proprietà additiva***.

Una proprietà analoga vale per le aree: se unisco dei poligoni l'area della figura risultante è la somma delle loro aree solo se essi non sono sovrapposti:



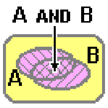
Area(A **U** B **U** C) = Area(A) + Area(B) + Area(C)

Area(A **U** B **U** C) ≠ Area(A) + Area(B) + Area(C)



Nel caso **D** per valutare Pr(S∈M **or** S∈F) non posso usare la proprietà additiva e fare Pr(S∈M)+Pr(S∈F) poiché si può essere insufficienti in entrambe le materie, ma devo fare:

Pr(S∈M **or** S∈F) = Pr(S∈M) + Pr(S∈F) – Pr(S∈M **and** S∈F) = 42%+39%–28% = 53%



Più in generale, di fronte a valutazioni del tipo Pr(**A** OR **B**) con **A** e **B** eventi non incompatibili, si usa la proprietà:

- Pr(**A or B**) = Pr(**A**) + Pr(**B**) – Pr(**A and B**)

Naturalmente, a seconda di come si scelgono le *valutazioni iniziali*, per la stessa situazione si possono ottenere ***diverse misure di probabilità***. Le valutazioni iniziali possono essere dedotte dall'esperienza o da considerazioni di tipo fisico o da propri convincimenti o …. Devono comunque essere tali da *non condurre a contraddizioni*: a partire da esse, applicando ripetutamente le ***proprietà*** sopra elencate, non posso ottenere valutazioni diverse per uno stesso evento, non posso ottenere probabilità negative o superiori al 100%, … (ad es. non posso valutare 60% la probabilità che nella prossima partita Roma-Lazio vinca la Roma e 50% che pareggino; verrebbe contraddetta la prima proprietà). Si osservi che il ruolo delle valutazioni iniziali mostra come anche in questo caso, come in altri discussi in altre voci, ***le conoscenze matematiche non sono di per sé sufficienti*** per modellizzare o risolvere "razionalmente" un problema.

Precisiamo meglio ***che cosa sono gli eventi***.

Chiamo ***fenomeno*** (o *esperimento*) ***casuale*** un fenomeno determinato da molti ***fattori*** alcuni dei quali li so valutare (so individuarli, ho gli strumenti per misurarli, non è troppo dispendioso rilevarli, … – nel caso del lancio del dado potrebbero essere le caratteristiche fisiche del dado), altri li ritengo ***casuali*** (nel caso del dado: l'impulso che gli dò, la rugosità e l'inclinazione della superficie della tavola, la presenza di correnti d'aria, …). La distinzione tra fattori casuali e non casuali è soggettiva, dipende dallo stato di conoscenze, dal tempo e dalle risorse che voglio dedicare all'analisi del problema, … : disponendo di strumenti sofisticati potrei misurare l'impulso che dò al dado, valutare la presenza di correnti d'aria, ….

Chiamo ***condizioni*** l'insieme dei fattori che riesco a valutare. A parità di condizioni (ad es. usando sempre un dado della stessa forma, dello stesso materiale, …), un fenomeno casuale può realizzarsi diversamente; in altre parole, più prove dell'esperimento possono dar luogo a risultati diversi.

Chiamo ***deterministico*** un fenomeno che non dipende da fattori casuali.

Un ***evento*** è un fatto che riguarda un fenomeno casuale; ogni volta che, a parità di condizioni, il fenomeno si realizza, l'evento può verificarsi o no: nel contesto del lancio di un dado, un evento può essere, lanciato un dado, "esce la faccia del dado con 4 pallini"; nel contesto delle "insufficienze" un esempio di evento è, preso uno studente alla fine del 1° quadrimestre, "lo studente è insufficiente in matematica e non in fisica".

Passando al ***modello matematico*****:**

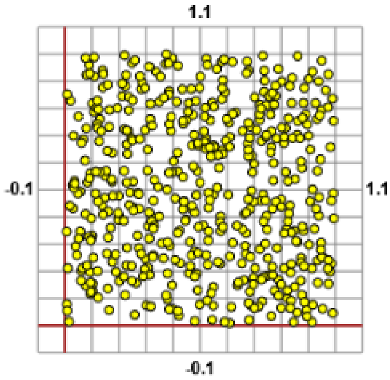
- rappresentiamo il fenomeno individuando uno o più *oggetti matematici* che rappresentino le grandezze o gli aspetti attraverso cui si manifesta il fenomeno; li indichiamo con dei *nomi*, così come si fa per le variabili nelle formule che descrivono fenomeni deterministici; per questo essi sono chiamati ***variabili casuali***;
- gli ***eventi*** vengono rappresentati mediante ***formule***, eventualmente combinate con *operatori logici* (**or**, **and**, ...), in cui compaiono variabili casuali riferite al fenomeno in questione:

– nel caso del lancio di una coppia di dadi posso considerare i *numeri* U1 e U2 pari alle quantità dei pallini che compaiono sulle facce superiori dopo il lancio e rappresentare il fatto che "venga 7" con U1+U2 = 7

– nel caso delle insufficienze posso indicare con S lo studente e scrivere S∈M **and not** S∈F per indicare l'evento "lo studente è insufficiente in matematica e non in fisica"

3. Leggi di distribuzione. Il generatore di numeri pseudocasuali

Quasi tutti i linguaggi di programmazione e le applicazioni matematiche sono dotate di un "fenomeno casuale". Più precisamente sono dotati di un comando, con nome "runif" o "random" o …, che ogni volta che viene chiamato genera "a caso" un "numero reale" tra 0 ed 1. Nel caso di JavaScript ecco, a fianco, che cosa si ottiene generando 500 punti con ascisse e ordinate pari a `Math.random()`; vedi lo script random.



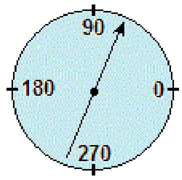
I numeri tra 0 ed 1 non vengono solo generati a caso, ma sembra che si distribuiscano in modo ***uniforme***: non viene privilegiata alcuna parte particolare degli intervalli tra 0 ed 1 e, quindi, alcuna zona particolare del quadrato raffigurato.

Anche gli istogrammi tracciati con gli script "Dado" e "DadoEquo" usano questo ***generatore di numeri casuali***. Ad esempio in "DadoEquo" vengono generati molti numeri casuali tra 0 ed 1 (100 mila nel caso illustrato), vengono moltiplicati per 6 (si ottengono numeri tra 0 e 6), viene aggiunto 1 (si ottengono numeri tra 1 e 7) e vengono troncati agli interi (si ottengono interi tra 1 e 6). E le frequenze relative di ciascuno di queste uscite tendono ad essere uguali (le colonne dell'istogramma tendono ad avere la stessa lunghezza). Tutto ciò è possibile perchè le uscite del "generatore" non privilegiano alcuna parte dell'intervallo (0,1).

$\boxed{1}$ Qual è la probabilità che un valore generato da `Math.random()`, assuma un valore compreso tra 0.5 ed 1? e tra 0.4 e 0.6?

L'istogramma relativo all'analisi statistica di un certo numero di valori assunti da una grandezza casuale viene chiamato anche *istogramma di distribuzione* ***sperimentale***, per distinguerlo dall'istogramma in cui al posto delle frequenze relative si considerano le probabilità, che viene chiamato *istogramma di distribuzione* ***teorico***. Analogamente si parla di media, mediana, moda, percentili, … *sperimentali* quando ci si riferisce alle frequenze percentuali ricavate da un'analisi di dati e di media, mediana, moda, percentili,… *teorici* quando si considerano al posto delle frequenze percentuali le corrispondenti probabilità.

Consideriamo un ago imperniato su un chiodo verticale. Diamo un colpo all'ago e vediamo qual è la direzione finale in cui si ferma.



Suppongo, in base a considerazioni di tipo fisico (verticalità del chiodo, bilanciamento dell'ago, …) e a verifiche sperimentali, che non vi siano direzioni privilegiate, nel senso che, comunque prenda due direzioni, la direzione finale dell'ago possa con la stessa probabilità essere più vicina all'una o all'altra.

Posso indicare la direzione finale dell'ago con la variabile casuale D. Associo a D [0,360) come intervallo di possibili valori e doto D di una legge di distribuzione uniforme: D ha una ***legge di distribuzione uniforme*** su [0,360).

$\boxed{2}$ Sia Pr la misura di probabilità associata alla variabile casuale D dell'esempio ora descritto.
Quanto vale $\Pr(0 \le D < 90)$? Quanto vale $\Pr(90 \le D < 270)$? Quanto vale $\Pr(60 \le D < 90)$?

Nel caso del lancio di un dado equo, se ne rappresento l'uscita con una variabile casuale U, devo:
− associare a U come insieme di valori possibili l'insieme finito {1,2,…,6};
− considerare la misura di probabilità Pr così definita per ogni h in {1,2,…,6}: $\Pr(U = h) = 1/6$.

Più in breve, si dice che ho dotato U di una ***legge di distribuzione uniforme*** sull'insieme {1,2…,6}.

Più in generale, se U è una variabile casuale a cui è associato come insieme di valori su cui può variare un insieme di *n* oggetti matematici {A1,A2,…,A*n*}, dotare U di una legge di distribuzione uniforme vuol dire considerare una misura di probabilità Pr che associ lo stesso valore di probabilità, 1/*n*, a tutti gli *n* eventi A1, A2, …, A*n*.

$\boxed{3}$ Considera una moneta equa, cioè una moneta per cui si possa ritenere che, lanciandola, l'uscita di "testa" e quella di "croce" siano egualmente possibili. Utilizzando la variabile casuale M con valori possibili "T" e "C", modellizza questa situazione definendo opportunamente una misura di probabilità Pr.

$\boxed{4}$ Considera il lancio di due dadi equi. Rappresentiamo le uscite dei due dadi con le variabili casuali U1 e U2 con distribuzione uniforme in {1,2,…,6}. Per rappresentare l'"uscita complessiva" consideriamo la variabile casuale U definita come U = U1+U2. Qual è l'insieme di valori che può assumere U? U ha distribuzione uniforme su tale insieme? Perché?

Se U è una variabile casuale a cui si è associato, come insieme di valori su cui può variare, un insieme di oggetti matematici {A1,A2,A3,…}, dotare U di una legge di distribuzione vuol dire descrivere come calcolare i valori di una misura di probabilità Pr per gli eventi U = A*i*. Nel paragrafo 4 studieremo la legge di distribuzione della variabile casuale considerata nel quesito 6.

Se invece a U è associato, come insieme di valori su cui può variare, un intervallo I di numeri, dotare U di una legge di distribuzione vuol dire descrivere come calcolare i valori di una misura di probabilità Pr per gli eventi del tipo $U \in J$ con J sottointervallo di I.

$\boxed{5}$ Uno sportello di un ufficio pubblico ha come orario di apertura l'intervallo [8.5, 13] (dalle 8 e mezza all'una). Indichiamo con la variabile casuale T l'ora in cui arriva un utente. Ha senso, secondo te, rappresentare questa situazione dotando T di una legge di distribuzione uniforme?

$\boxed{6}$ Prova ad eseguire lo script [TC](TC), di cui sotto è riportato il testo e una possibile uscita. U tende a distribuirsi in modo uniforme?

```
<script>
U=""; for(i=1; i<=50; i=i+1) if(Math.random()<0.5) {U=U+"T"} else {U=U+"C"}
document.write(U)
</script>
```

`CTCCCTTTTCTTTTCCCTTTCTTCTCTCTCCCTTCTCCCC`

Possiamo impiegare "random" anche nella nostra grande CT. Prova a mettere `trunc(random()*6+random()*6)+2` nella finestra sotto ad **A** e clicca più volte [=]. Che cosa ottieni?

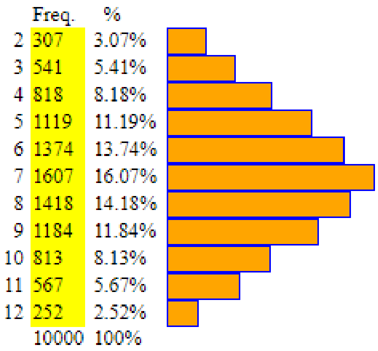
La legge di distribuzione uniforme è solo un modello matematico del comportamento di `random` ma non lo rappresenta esattamente:

- `random` non può assumere come valori tutti i numeri reali compresi tra 0 ed 1 ma solo un sottoinsieme dell'insieme dei numeri macchina contenuti in tale intervallo, e questi sono in quantità finita in quanto sono descrivibili con sequenze di bit di lunghezza fissata;
- inoltre i valori che assume `random` inevitabilmente (stiamo usando un computer!) non si susseguono in modo "casuale", ma sono

generati in un dato ordine da un algoritmo che, dopo una quantità finita N di valori – parecchi miliardi in questo caso –, ritorna a generare gli stessi valori; per questo motivo **random** (o, meglio, l'algoritmo che ne calcola i valori) viene chiamato anche ***generatore di numeri pseudocasuali***.

4. Modi per calcolare probabilità

Voglio studiare come si distribuiscono le uscite del lancio di due dadi equi, cioè individuare la legge di distribuzione di U definita con U = U1+U2 dove U1 e U2 hanno distribuzione uniforme in {1,2,…,6}. Per eseguire molte prove e non procedere a mano, con due dadi veri, posso ricorrere allo script **2dadi**. **Prova** ad usarlo effettuando 100, 1000, 10000 lanci.



Se esplori il testo dello script vedi che la generazione delle uscite è effettuata dal seguente comando:

```
for(i=1;i<=N;i=i+1) {k=
Math.floor(Math.random()*6)+Math.floor(Math.random()*6)+2; dat[k-1]= dat[k-1]+1}
```

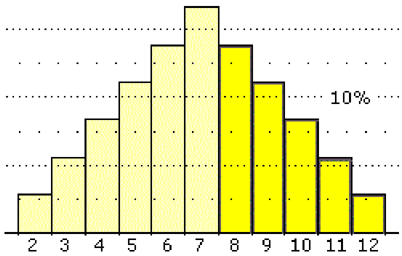
Math.floor(Math.random()*6) genera un numero intero tra 0 e 5; a **k** assegno la somma di due interi così generati a cui aggiungo 1+1 in modo che l'esito corrisponda al lancio di due dadi; poi incremeno di 1 la variabile **dat[k-1]**; il totale delle uscite uguali a 2, a 3, … lo metto in dat[1], in dat[2], … (queso è il motivo per cui uso [k-1]).

7 Usa la piccola CT per valutare la probabilità che un'uscita sia maggiore di 7 (calcola 14.18+11.84+8.13+5.67+2.52; quanto fa?).

In questo caso era facile determinare la probabilità anche per via teorica: le "possibili" coppie di uscite, tra loro equiprobabili se i dadi sono equi, sono 6·6 = 36 – (1,1), (1,2), …, (1,6), (2,1), …, (2,6), …, (6,6) – mentre quelle "favorevoli", che danno luogo ad un numero maggiore di 7, sono 1+2+3+4+5 = 15 – (2,6), (3,5), (3,6), (4,4), (4,5), (4,6), …, (6,6) – per cui la probabilità è 15/36 = 5/12 = 41.666…%, quasi uguale al valore 42.34% prima ottenuto.

La tabella sottostante riporta le uscite di U1+U2 al variare di U1 e U2. L'istogramma illustra la legge di distribuzione di U1+U2. L'uscita più probabile è 7, in quanto è quella formabile in più modi. Le parti in giallo scuro corrispondono all'evento U1+U2>7. Nel caso dell'istogramma la parte scura è i 5/12, ossia il 41.7%, dell'intera superficie.

	U_2=1	U_2=2	U_2=3	U_2=4	U_2=5	U_2=6
U_1=1	2	3	4	5	6	7
U_1=2	3	4	5	6	7	**8**
U_1=3	4	5	6	7	**8**	**9**
U_1=4	5	6	7	**8**	**9**	**10**
U_1=5	6	7	**8**	**9**	**10**	**11**
U_1=6	7	**8**	**9**	**10**	**11**	**12**



Come faccio, in questo caso, a calcolare **Pr(U = 4)**? Nel caso in questione le uscite sono 6·6 = 36, sono tutte equiprobabili, i modi in cui posso fare 4 sono 3, come si capisce dalla tabella precedente o ragionando nel modo seguente: U1 può essere 1, 2 o 3, ed U2 è 4 − U1.

Quindi la probabilità cercata è la somma di tre probabilità ciascuna uguale a 1/36, ed è quindi: 1/36+1/36+1/36 = 3/36 = 1/12.

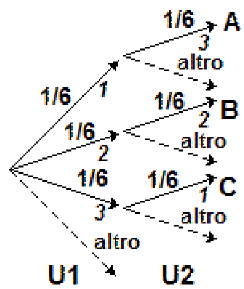
Posso, in altermativa, procedere con un ➔ ***grafo ad albero***:

− rappresento con successive diramazioni i diversi esiti possibili per U1 e per U2 (eventualmente raggruppando gli esiti "sfavorevoli" in un'unica diramazione);
− associo agli archi che corrispondono a esiti "favorevoli" la relativa probabilità;
− calcolo, per ogni percorso (dal nodo iniziale a un nodo finale) costituito solo da archi "favorevoli", il prodotto delle probabilità associate ai vari archi e lo scrivo a fianco del nodo finale;
− sommo i valori così calcolati.

I percorsi favorevoli sono U1 = 1, U2 = 3; U1 = 2, U2 = 2; U1 = 3, U2 = 1.

Nella figura a fianco sono indicati **A**, **B** e **C**, e corrispondono ciascuno alla probabilità 1/6·1/6 = 1/36.

Complessivamente, 1/36 + 1/36 + 1/36 = 1/12.



8 Dispongo di due dadi, uno equo con uscita U, l'altro con uscita V con la seguente legge di distribuzione:
Pr(V=1)=10%, Pr(V=2)=19%, Pr(V=3) = Pr(V=4) = 14%, Pr(V=5)=15%, Pr(V=6)=28%.
Per calcolare Pr(U+V=4) posso usare entrambi i metodi "teorici" visti sopra ("casi favorevoli"/"casi possibili"; grafo ad albero)?
Calcola Pr(U+V=4).

5. Eventi (probabilisticamente) dipendenti e indipendenti

Usando il "metodo dei grafi" per motivare il prodotto delle probabilità associate agli archi di un percorso abbiamo fatto riferimento al prodotto di percentuali. Consideriamo un esempio per approfondire questo aspetto.

Nella città XX il 21% degli abitanti sono biondi; il 52% degli abitanti sono femmine; il 9% degli abitanti sono abbonati "allo stadio", *per una o l'altra delle due squadre cittadine. Posso stimare la percentuale di abitanti che sono femmine bionde? Posso stimare la percentuale di abitanti che sono femmine abbonate?*

Affrontiamo la prima domanda.

Suppongo, ragionevolmente, che l'essere biondo (B) e l'essere femmina (F) siano caratteristiche *indipendenti,* cioè che i biondi si distribuiscano tra i maschi allo stesso modo che tra le femmine: la percentuale di biondi tra i maschi e quella tra le femmine siano uguali a quella che c'è nel complesso degli abitanti.

Quindi dalla tabella sottostante a sinistra, che rappresenta le informazioni iniziali, posso ottenere la tabella a destra riproducendo la riga in basso nelle righe soprastanti modificandola proporzionalmente mediante i fattori moltiplicativi 48% e 52%:

	B	**not** B	tot		B	**not** B	tot
F			52%		21%·52%	79%·52%	52%
M			48%	→	21%·48%	79%·48%	48%
tot	21%	79%	100%		21%	79%	100%

In formula: FrequenzaRelativa(B **and** F) = FrequenzaRelativa(B)·FrequenzaRelativa(F)

Se indico con la variabile *Sesso* a valori in {"m", "f"} il sesso di un generico abitante di XX e con la variabile *Biondo* a valori in {"sì","no"} il suo essere o no biondo, posso esprimere la supposizione fatta inizialmente dicendo che *Sesso* e *Biondo* sono variabili casuali "indipendenti". *Passando dalla frequenza alla probabilità,*

posso scrivere: Pr(*Biondo*="sì" **and** *Sesso*="f") = Pr(*Biondo*="sì")·Pr(*Sesso*="f")

Quindi se estraggo a caso dall'anagrafe di XX il nominativo di un abitante, la probabilità che sia una femmina bionda è: Pr(*Biondo*="sì")·Pr(*Sesso*="f") = 21%·52% = 11% (ho arrotondato a 2 cifre il risultato 0.1092 di 0.21·0.52 in quanto i fattori erano arrotondati a 2 cifre).

9 Si può procedere allo stesso modo (52%·9%=…) per valutare la probabilità che il nominativo estratto sia di una abbonata allo stadio?

10 *Provate* a rispondere alle seguenti domande: (**a**) *Qual* è la probabilità che *alzando* 2 volte un mazzo (nuovo) di carte da scopa ottenga sempre una carta di denari? (**b**) *Qual* è la probabilità che *estraendo* 2 carte dal mazzo queste siano entrambe di denari?

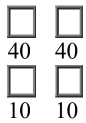
Mentre è ragionevole supporre che l'essere biondo e l'essere femmina siano caratteristiche indipendenti (ciò, in effetti, è confermato dalle statistiche: le percentuali di biondi tra i maschi e tra le femmine sono presso che identiche), sicuramente vi sono più abbonati tra i maschi che tra le femmine. Quindi è ***sbagliato*** rispondere al quesito 9 ritenendo applicabile la formula:

Pr(Abbonato = "sì" **and** Sesso = "f") = Pr(Abbonato = "sì")·Pr(Sesso = "f")

Senza altre informazioni non si può valutare la probabilità di estrarre il nominativo di una femmina abbonata.

Vedremo ora che le situazioni del quesito 10 presentano analogie con le situazioni appena discusse.

- Nel caso della ***alzata***, avendo supposto il mazzo nuovo (e non truccato e mescolato bene) posso ritenere che, tagliandolo, le carte compaiano con distribuzione *uniforme.* Quindi posso fare:



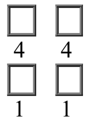
Probabilità = NumeroCasiFavorevoli / NumeroCasiPossibili

I casi possibili sono 40·40: 40 carte possibili per la prima alzata e 40 per la seconda.

I casi favorevoli sono 10·10: sono 10 le carte di denari.

La probabilità che esca una carta di denari entrambe le volte è: **(10·10)/(40·40)** = 1/16 = 6.25%

Posso anche considerare direttamente la variabile casuale "seme", a valori in {♥, ♦, ♣, ♠}, che, per gli stessi motivi di prima, posso ritenere con distribuzione uniforme: l'uscita di una carta di denari ha la stessa probabilità di quella di una di fiori o … Quindi posso usare:

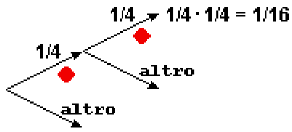


Probabilità = NumeroCasiFavorevoli / NumeroCasiPossibili.

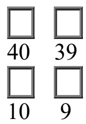
I casi possibili sono 4·4 (4 semi possibili), quelli favorevoli 1·1 (1 è il seme che mi interessa).

La probabilità cercata dunque è: **(1·1)/(4·4)** = 1/16 = 6.25%

Avrei anche potuto utilizzare il *grafo ad albero* raffigurato a lato.



- Anche nel caso della ***estrazione*** posso ritenere equiprobabili le carte del mazzo.

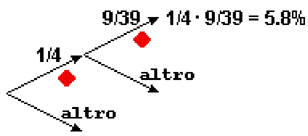


I casi possibili sono 40·39 (alla seconda estrazione dispongo di una carta in meno)

I casi favorevoli sono 10·9 (se nella prima estrazione è uscito denari, nel mazzo sono rimaste 9 carte di denari).

La probabilità cercata è: **(10·9)/(40·39)** = 9/4/39 = 5.8% (arrotondamento)

Anche in questo caso sarei potuto ricorrere a un grafo ad albero.



Indichiamo con le variabili casuali S1 e S2 il seme della prima uscita (alzata o estrazione) e quello della seconda. Nel caso della ***alzata*** S1 e S2 *sono indipendenti:* qualunque seme abbia la 1ª carta, la probabilità che la 2ª abbia un certo seme è sempre la stessa. Ciò corrisponde al fatto che il grafo relativo all'alzata si riproduce allo stesso modo passando da una diramazione alla successiva. Per calcolare Pr(S1=♦ **and** S2=♦) posso fare direttamente Pr(S1=♦)·Pr(S2=♦) = 1/4·1/4 = 1/16.

Nel caso della ***estrazione*** S1 e S2 *non sono indipendenti:* ad es. Pr(S2=♦) (la probabilità che la 2ª carta sia di ♦) dipende dal valore assunto da S1 (cioè dal seme della 1ª carta). Ciò corrisponde al fatto che il grafo relativo alla estrazione non si riproduce allo stesso modo passando da una diramazione alla successiva: al primo arco "♦" è associata la probabilità 1/4, al secondo arco "♦" è associata la probabilità 9/39.

11 *Provate* a rispondere alle seguenti domande: (**a**) *Qual* è la probabilità che *alzando* 2 volte un mazzo (nuovo) di carte da scopa ottenga sempre lo stesso colore? (**b**) *Qual* è la probabilità che *estraendo* 2 carte dal mazzo queste abbiano lo stesso colore?

Riassumendo, due ***eventi*** **A** e **B** vengono detti ***indipendenti*** se l'ipotesi che un evento si verifichi non modifica la valutazione che l'altro possa verificarsi. In questo caso: Pr(**A and B**) = Pr(**A**)·Pr(**B**). Altrimenti vengono detti ***dipendenti***. Esempi:

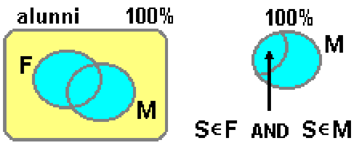
− "essere biondo" ed "essere femmina" sono indipendenti; "uscire denari alla 1ª alzata" e "uscire denari alla 2ª alzata sono indipendenti;
− "essere abbonato allo stadio" ed "essere femmina" sono dipendenti; "uscire denari alla 1ª estrazione" ed "uscire denari alla 2ª estrazione" sono dipendenti.

Due ***variabili casuali*** X e Y sono ***indipendenti*** se sono indipendenti gli eventi **A** e **B** comunque prenda **A** evento relativo a X (condizione in cui compare solo la variabile X) e **B** evento relativo a Y (condizione in cui compare solo variabile Y): conoscere qualcosa su come si manifesta X non modifica le mie aspettattive sui modi in cui può manifestarsi Y, e viceversa. Altrimenti sono ***dipendenti***. Esempio:

− sapere qualcosa a proposito del seme della 1ª carta estratta cambia le mie valutazioni sul seme che potrebbe avere la 2ª carta estratta: il seme della 1ª estrazione e quello della 2ª sono variabili casuali dipendenti.

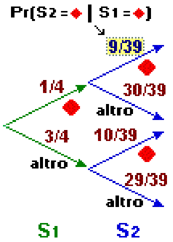
Nota. Il concetto di ***dipendenza*** ora introdotto è diverso da quello impiegato per esprimere il legame tra due grandezze quando una varia *in funzione* dell'altra. Se A e L sono area e lato di un generico quadrato, A *dipende* da L nel senso che ad ogni valore di L corrisponde un unico valore di A. Invece nel caso dell'estrazione dire che le variabili S1 e S2 sono *dipendenti* non significa che l'una è funzione dell'altra: l'uscita di ♦ alla 1ª estrazione non determina univocamente l'uscita della 2ª estrazione.
Quando è il caso, per distinguere i due tipi di dipendenza, si parla, rispettivamente, di *dipendenza* ***funzionale*** (o *deterministica*) e di *dipendenza* ***probabilistica***.

Facendo riferimento all'esempio ➜ (**D**) già considerato all'inizio della scheda, supponiamo di voler trovare la *probabilità che un alunno insufficiente in matematica sia insufficiente anche in fisica,* valore che indichiamo con l'espressione: Pr(S∈F | S∈M) (leggendo: «probabilità dell'evento S∈F *nella condizione* che si verifichi l'evento S∈M»).



La popolazione a cui riferire l'essere insufficiente in fisica non è il totale degli alunni, ma solo il sottoinsieme di quelli insufficienti in matematica. La probabilità cercata è dunque uguale a:

$$\frac{\texttt{Pr(S∈F and S∈M)}}{\texttt{Pr(S∈M)}} = \texttt{28\%/42\% = 67\%}$$



Analogamente, per calcolare Pr(S2=♦ **and** S1=♦) nel caso della ➜ *estrazione* ho fatto:
Pr(S2=♦ **and** S2=♦) = Pr(S1=♦) · Pr(S2=♦ | S1=♦) = 1/4 · 9/39
[Pr(S2=♦ | S1=♦): probabilità che la seconda estratta sia ♦ *nota l'informazione* che la prima estratta è stata ♦]

Il concetto di ***probabilità condizionata*** introdotto nei due esempi precedenti può essere generalizzato.
Dati due eventi **A** e **B** indichiamo Pr(**A**|**B**) (probabilità che si verifichi **A** a condizione che si verifichi **B**, o, più in breve, probabilità di "**A** sotto la condizione **B**") il rapporto definito nel modo sotto a sinistra:

$$\texttt{Pr(A|B)} = \frac{\texttt{Pr(A and B)}}{\texttt{Pr(B)}} \qquad\qquad \texttt{Pr(A and B) = Pr(B)·Pr(A|B)}$$

Nel secondo esempio (estrazione) ho usato la formula equivalente indicata sopra a destra. Tutte le volte che si calcolano probabilità con il "*metodo dei grafi*" di fatto si applica questa formula (che viene chiamata anche "della ***moltiplicazione*** delle probabilità").

Nel caso in cui **A** e **B** sono *indipendenti*, ovviamente le formule diventano Pr(**A**|**B**) = Pr(**A**) [il verificarsi o meno di **B** non condiziona la valutazione di Pr(**A**)] e Pr(**B and A**) = Pr(**B**)·Pr(**A**) (la formula considerata dopo il ➜ ques. 11).

12 Utilizzando la tabella a lato, *valutate* se sesso e settore di attività (classificato in "agricoltura", "industria", "altre attività") in cui una persona (nell'anno X, nello stato Y) era occupata possono essere ritenute (probabilisticamente) indipendenti e, in caso negativo, se, almeno, l'occupazione in agricoltura o l'occupazione nell'industria o quella in "altre attività" è indipendente dal sesso.

settore	*M*	*F*
agricoltura	1.17	0.66
industria	5.26	1.66
altre attiv.	7.68	5.18

Nota. Abbiamo visto che `Math.floor(Math.random()*6)+Math.floor(Math.random()*6)+2` in **JS** (e comandi simili in altri programmi) simula l'esito del lancio di due dadi equi. Affinché la simulazione sia corretta non basta che `random` abbia distribuzione uniforme: l'uscita di un dado deve essere indipendente da quella dell'altro. In altre parole due comparse di `random` si comportano come due distinte variabili casuali. Ciò accade in tutti i casi in cui compare `random` (quindi posso usare `random` per simulare non solo il lancio di dadi, ma anche lanci di monete, colpi ad aghi, …).

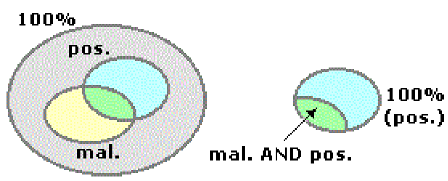
6. Usi (e limiti) del calcolo delle probabilità

Consideriamo qualche esempio d'uso significativo del calcolo delle probabilità in contesti "seri".

Un certo test sanitario per valutare la presenza (esito positivo) o assenza (esito negativo) della malattia X ha attendibilità del 95% (in caso di presenza c'è il 95% di probabilità che l'esito sia positivo , in caso di assenza il 95% di probabilità che sia negativo). Si sa da statistiche serie che l'1% della popolazione è affetta dalla malattia X. *Se per una persona il test dà esito positivo, qual è la probabilità che essa sia realmente malata?*

Devo determinare Pr("essere malato" | "risultare positivo"):

$$\frac{\texttt{Pr("essere malato" and "risultare positivo")}}{\texttt{Pr("risultare positivo")}}$$



Per calcolare il rapporto devo trovare il valore dei "?" della seguente ***tabella a 2 entrate*** (il valore di ogni casella dipende da due input: la condizione rispetto al test - riga - e quella rispetto alla malattia - colonna): metto i dati sulla popolazione (prime 2 colonne dell'ultima riga), poi utilizzo il dato sull'attendibilità del test (per ottenere le prime 2 colonne della 1ª riga: 1·95% = 0.95, 99·5% = 4.95), infine completo la 1ª riga (0.95+4.95 = 5.90).

	malati	sani	totale
positivi	**?**		**?**
negativi			
totale	1	99	100

→

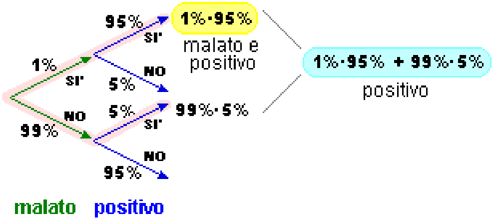
	malati	sani	totale
positivi	**0.95**	4.95	?
negativi			
totale	1	99	100

→

	malati	sani	totale
positivi	0.95	4.95	**5.90**
negativi			
totale	1	99	100

La probabilità cercata è dunque 0.95% / 5.90% = 16%, molto meno di 95%, come si sarebbe potuto pensare.

In alternativa all'uso della tabella, per trovare Pr("risultare positivo") potevo usare un grafo ad albero, come si è fatto a lato.



Questo esempio evidenzia il ruolo del *calcolo delle probabilità* nella ***razionalizzazione*** delle situazioni "incerte".

13 Una vetrina di un negozio viene infranta e viene prelevata parte della merce esposta. Un testimone afferma che il ladro era arabo. Quando gli inquirenti gli ripropongono scene simili in analoghe condizioni di luce, distanza, … il testimone identifica correttamente la razza (arabo, non arabo) del ladro nel 75% dei casi. È attendibile la testimonianza se nella località considerata il 12% dei furti sono opera di ladri di razza araba? [per rispondere calcola la probabilità che il ladro sia effettivamente un arabo]

Il calcolo delle probabilità ***non*** è tuttavia ***sempre sufficiente*** a stabilire qual è la scelta più conveniente:

14 X, noto professionista, arriva alla stazione per prendere il treno per un importante viaggio di lavoro quando si accorge di aver dimenticato la carta di credito e di aver nel portafoglio solo 20 euro, insufficienti per il biglietto: mancano 15 euro. Allora X decide di scommettere 15 euro (con un'altra persona) che giocando a bim-bum-bam esca un numero multiplo di 3. Secondo voi la decisione di X è conveniente? E per l'altra persona?

In tutte le *lotterie* si è sempre di fronte a situazioni simili a quella del quesito precedente: dal punto di vista probabilistico giocare è sempre sconveniente: • solo una parte dell'incasso derivante dalla vendita dei biglietti viene utilizzata per pagare le vincite, per cui la vincita *media* è inferiore al prezzo del biglietto, • ma nella decisione di acquistare un biglietto la valutazione probabilistica può essere controbilanciata da altre valutazioni di tipo economico, di tipo edonistico, … (un discorso a parte meriterebbero, poi, le lotterie "elettroniche", che qualcuno considera "più sicure" perché non interviene direttamente un operatore nell'estrazione: abbiamo visto come il generatore elettronico di numeri casuali può essere facilmente programmmato …).

Per un altro esempio si pensi ai *vaccini*, che a volte hanno una certa probabilità di causare l'insorgere delle malattie stesse. Per decidere se rendere obbligatoria una vaccinazione non basta trovare che tale probabilità è bassa rispetto alla diffusione della malattia: imporre a chi potrebbe rimanere sano una vaccinazione che può causare una malattia comporta valutazioni anche di tipo morale.

Vi sono anche casi in cui si ricorre a valutazioni probabilistiche ***erronee*** perché basate su campioni mal scelti o per altri difetti metodologici: tipico è l'esempio di un sondaggio telefonico che può avere come risposta A o B in cui si tenga conto solo di chi accetta di rispondere senza considerare il fatto che coloro che non vogliono rispondere potrebbero, per la natura della questione, essere più inclini a una delle due risposte.

E vi sono casi in cui se ne fa un uso ***improprio***, ad es. quando si confonde la presenza di una relazione di ***dipendenza probabilistica*** con la presenza di un legame di ***causa-effetto***:

15 Un gruppo di medici, utilizzando indagini statistiche, ha concluso che la malattia X (che fino ad allora si riteneva fosse causata dal tipo di alimentazione) è essenzialmente dovuta a fattori genetici. Infatti dalle statistiche ha dedotto che:
Pr("avere X" | "avere genitori o fratelli con X") > Pr("avere X" | **not** "avere genitori o fratelli con X").
Vi sembra corretta questa conclusione? Perché?

7. Esercizi

e1 Se in una situazione simile al quesito ➔ **C** di §1, indicata con H l'altezza di un cliente ventenne, sapessi che Pr(H<183) = 87% e che Pr(H<170) = 25%, che cosa potrei dire di Pr(170≤H≤183)?

e2 Un dado truccato dà 1, 2, 3, 4, 5, 6 con, in ordine, le probabilità (arrotondate) 0.125, 0.170, 0.240, 0.170, 0.125, 0.170. Qual è la probabilità che, lanciatolo, si ottenga un numero multiplo di 3?

e3 Devo ritenere più facile che esca il 10 sulla ruota di Genova o l'11 su quella di Torino, sapendo che il 10 non esce sulla ruota di Genova da 30 estrazioni e l'11 non esce su quella di Torino da 86 estrazioni?

e4 Ogni 4 ore di funzionamento una particolare macchina di una fabbrica viene controllata; al momento del controllo la macchina può trovarsi in uno dei seguenti stati:
– con probabilità 0.65, in perfetto funzionamento,
– con probabilità 0.08, in funzionamento leggermente irregolare, che richiede 15 min di riparazione,
– con probabilità 0.12, con alcuni pezzi da sostituire, il che richiede 2 ore di riparazione,
– con probabilità 0.15, in stato tale da richiedere una revisione completa, della durata di 24 ore.
Qual è la probabilità che la macchina possa essere rimessa in funzione entro 3 ore dal momento dell'interruzione per il controllo?

e5 Lanciando due dadi "equi" (con le facce numerate da 1 a 6), l'evento meno probabile fra i seguenti è:

A: escono due numeri maggiori di 3 **B:** escono due numeri uguali **C:** escono due numeri la cui somma è 9 **D:** escono un numero pari e uno dispari

e6 Faccio il seguente esperimento: lancio una moneta ripetutamente fino a che esce testa e annoto la quantità N di lanci che ho effettuato. Quali sono i valori che può assumere la variabile casuale N? Calcola Pr(N>1).

e7 Qual è la probabilità che in una classe di 25 alunni (nati in un anno non bistestile) ve ne siano almeno due nati nello stesso giorno?

e8 Da un'indagine statistica risulta che, negli USA, tra il 1980 e il 1985, 1 adulto su 3 fumava, 1 su 1500 morì di tumore entro un anno, 1 su 2000 fumava e morì di tumore entro un anno. Sulla base di queste informazioni valuta la probabilità che uno morì per problemi cardiaci se fumava e quella che uno morì per problemi cardiaci se non fumava.

e9 Sul *The Times* di Londra, nel 2004, viene riferito che lo skipper islandese E.O., un anno dopo aver urtato col suo peschereccio lo yacht dell' inglese J.H. provocando un danno di 30 mila dollari, lo ha riurtato, causando danni per 40 mila dollari: avvistato lo yacht si era diretto verso di esso per scusarsi della collisione precedente …. Ciò contraddice il fatto che la probabilità che si ripeta due volte un evento abbastanza raro come questo dovrebbe essere molto bassa? ossia se A è l'evento che E.O. nell'arco della sua vita urti lo yacht di J.H, e se, ad es. Pr(A) = 0.1% = 0.001 (1 su mille), la probabilità che questo evento si ripeta non è 0.1%·0.1% = 0.0001% (1 su 1 milione)? Discuti la cosa.

e10 Consideriamo → l'esempio relativo a un test sanitario considerato all'inizio di §6. Se, risultati positivi al test, si risultasse positivi anche a una seconda effettuazione dello stesso test (supponendo questa indipendente dalla prima), quale sarebbe la probabilità di essere malati?

1) Segna con l'evidenziatore, nelle parti della scheda indicate, frasi e/o formule che descrivono il significato dei seguenti termini:
fenomeno casuale, deterministico (§2), *evento, variabile casuale* (§2), *istogr. di distribuzione* (§3), *legge di distribuz.* (§3), *eventi e variabili casuali (in)dipendenti* (§5), *prob. condizionata* (§5)

2) Su un foglio da "quadernone", nella prima facciata, esemplifica l'uso di ciascuno dei concetti sopra elencati mediante una frase in cui esso venga impiegato.

3) Nella seconda facciata riassumi in modo discorsivo (senza formule, come in una descrizione "al telefono") il contenuto della scheda (non fare un elenco di argomenti, ma cerca di far capire il "filo del discorso").

script: piccola CT grande CT isto isto con % boxplot striscia 100 ordina Grafici divisori Indet distanza Triang eq.polinomiale eq.nonPolin sistemaLin moltPolin sempliciEq divisori fraz/mcd opFraz SumPro dadoCart 7e1/2 DadoEquo random TC 2dadi