Dissero di lei:

- TUTTI I MODELLI SONO SBAGLIATI, QUALCUNO E' UTILE. (COX)
- IL MODELLO DEVE SEGUIRE I DATI E NON VICEVERSA. (BENZECRI')
- DIO E LA STATISTICA SONO RISPOSTE CHE GODONO DI OTTIMA SALUTE

  (DANIEL PENNAC)
- SE I DATI VENGONO TORTURATI ALLA FINE CONFESSANO (ANONIMO)
- IL MIO NEMICO E' LA VARIANZA (ANONIMO)
- L'UOMO PUO' CREDERE ALL'IMPOSSIBILE, MA NON CREDERA' MAI ALL'IMPROBABILE (OSCAR WILDE)

Levine, Krehbiel, Berenson

Statistica

Capitolo 4

Probabilità

Insegnamento: Statistica

Corso di Laurea Triennale in Economia

Dipartimento di Economia e Management,

Università di Ferrara

Docenti: Prof. S. Bonnini, Dott.ssa A. Grassi

Argomenti

- I concetti di base della probabilità
  - Spazi campionari ed eventi
  - Tabelle di contingenza e diagrammi di Venn
  - Probabilità marginale e congiunta
  - Probabilità di un unione di eventi
- La probabilità condizionata
- Teorema di Bayes

I concetti di base della probabilità

La **probabilità** può essere definita come un numero che esprime la possibilità, il grado di verosimiglianza con cui un evento è destinato a verificarsi.

Si parla così della probabilità di pescare una carta nera da un mazzo di carte, della probabilità che un individuo preferisca un prodotto a un altro o della probabilità che un nuovo prodotto abbia successo sul mercato.

La probabilità è una proporzione o frazione che varia tra i valori 0 e 1, estremi inclusi. Associamo il valore zero a un evento che non ha nessuna possibilità di verificarsi (**evento impossibile**) e il valore uno a un evento che si verificherà sicuramente (**evento certo**).

I concetti di base della probabilità

Ci sono tre diversi approcci e interpretazioni del concetto di probabilità:

- Classico a priori $\Rightarrow$ conoscenza preliminare a priori
- Empirico classico $\Rightarrow$ conteggio delle frequenze osservate
- Soggettivo o bayesiano $\Rightarrow$ valutazione soggettiva da parte di un individuo

Secondo l'**approccio classico a priori**, nel semplice caso che ciascun risultato sia ugualmente probabile, la probabilità che un evento si verifichi è definita nel seguente modo:

Probabilità che l'evento si verifichi = X / T **(4.1)**

dove X = numero di risultati favorevoli all'evento

T = numero di risultati possibili

I concetti di base della probabilità

Nel caso dell'**approccio empirico** il risultato è lo stesso anche se il calcolo viene fatto in basi ai dati osservati. Questa probabilità potrebbe ad esempio riferirsi alla proporzione di individui in un campione che acquistano effettivamente un televisore, che preferiscono un certo candidato politico o che conciliano gli studi con un lavoro part-time.

Il cosiddetto **approccio soggettivo** definisce la probabilità associata ad un evento combinando l'esperienza passata dell'individuo, la sua opinione personale e l'analisi del particolare contesto di riferimento, ed è particolarmente utile quando la probabilità dell'evento non può essere determinata.

Spazi campionari ed eventi

Consideriamo un esperimento aleatorio, cioè un esperimento il cui esito è incerto. Gli elementi di base della teoria della probabilità sono i possibili risultati del fenomeno oggetto dell’analisi. Il concetto di probabilità è strettamente collegato al concetto di evento.

Chiamiamo **evento** ogni possibile risultato dell’esperimento aleatorio. Un **evento elementare** può essere descritto da una *singola caratteristica*. Ad esempio un evento elementare è l’esito testa o croce nel lancio di una moneta o la carta estratta in un mazzo di 52 carte.

Un **evento congiunto** è un evento definito da due o più caratteristiche (ad esempio estrarre 2 carte rosse).

Spazi campionari ed eventi

Il **complementare** di un evento *A*, indicato con il simbolo A', comprende tutti gli eventi elementari che non fanno parte di *A*.

Ad esempio il complementare dell'evento "estrarre un carta di picche" è l'evento "estrarre una carta di fiori, o di cuori, o di quadri".

L'insieme di tutti i possibili eventi elementari è detto **spazio campionario**.

Se estraiamo una carta da un mazzo standard, lo spazio campionario è costituito da tutte le carte che potremmo estrarre (quindi tutti i numeri e le figure di tutti i semi).

Spazi campionari ed eventi

Esempio: campione di 1000 intervistati rispetto a due variabili di interesse: l'acquisto pianificato e l'acquisto effettivo di un televisore.

	Acquisto effettivo		
Intenzione di acquisto	Sì	No	Totale
Sì	200	50	250
No	100	650	750
Totale	300	700	1000

Lo spazio campionario consiste nelle intenzioni e attuazioni di acquisto di ognuno dei 1000 individui intervistati. Gli eventi elementari sono "intenzione di acquisto", "intenzione di non acquisto", "acquisto effettivo", "non acquisto effettivo". L'evento complementare di "intenzione di acquisto" è "intenzione di non acquisto". L'evento "intenzione di acquisto e acquisto effettivo" è un evento congiunto.

Tabelle di contingenza e diagrammi di Venn

Ci sono diversi modi per interpretare e descrivere uno spazio campionario: una **tabella di contingenza** (slide precedente) oppure un **diagramma di Venn**. Questo diagramma rappresenta graficamente gli eventi come "unioni" ($A \cup B$) e "intersezioni" ($A \cap B$) di cerchi.

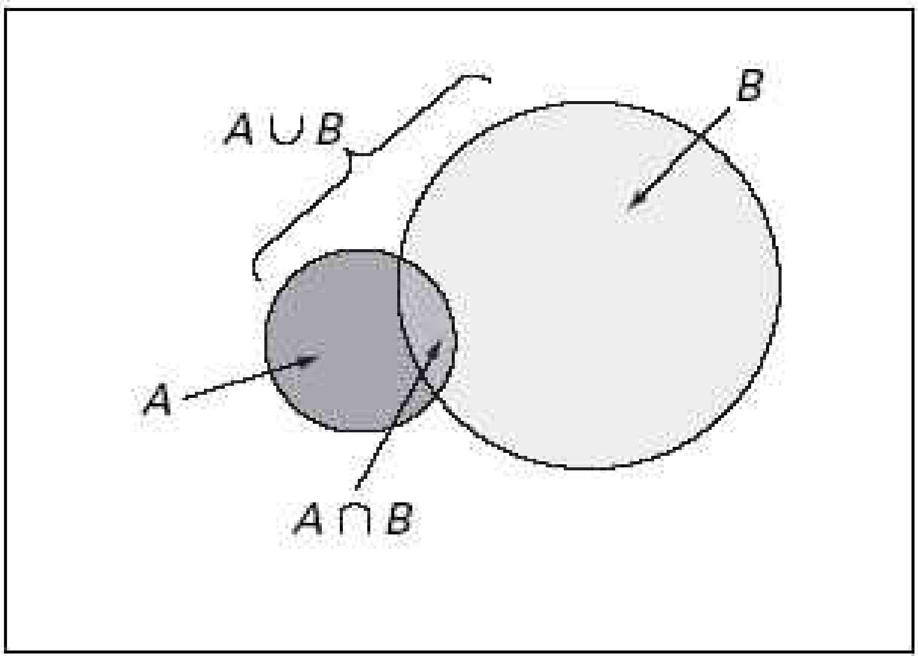


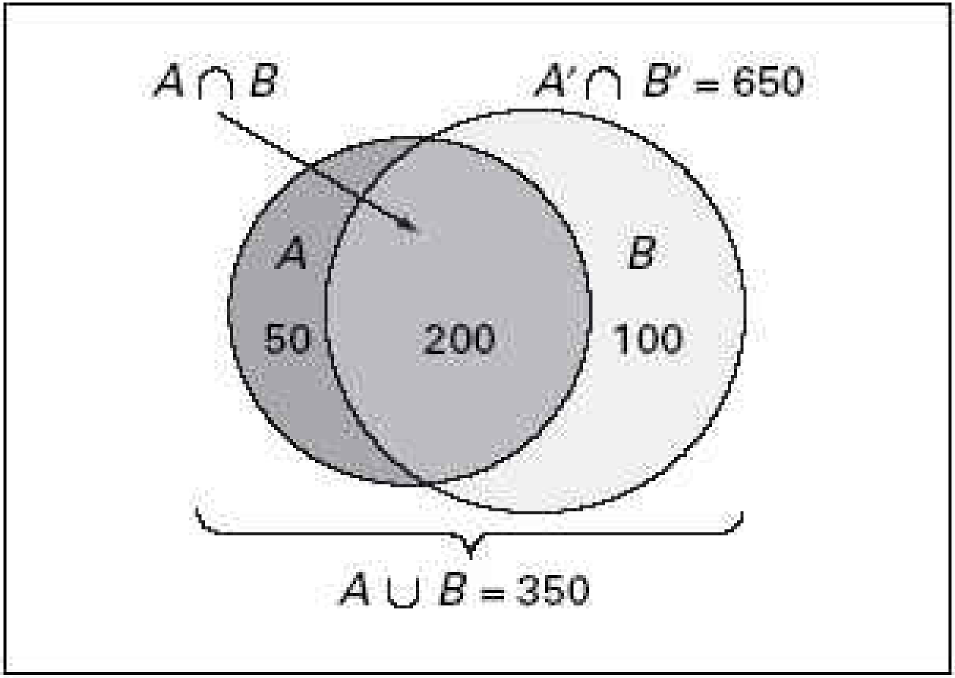
Tabelle di contingenza e diagrammi di Venn

Con riferimento all'esempio precedente, definiti gli eventi

A ="intenzione di acquisto", A' = "intenzione di non acquisto"

B = "acquisto effettivo", B' = "non acquisto effettivo"

è possibile rappresentare l'esempio come segue:



Probabilità marginale

Chiamiamo **probabilità marginale** la probabilità, *P(A)*, del verificarsi di un evento semplice, per esempio la probabilità che venga pianificato l'acquisto di un nuovo televisore.

La probabilità di estrarre a caso un soggetto che ha pianificato l'acquisto di un televisore è data da (equazione 4.1):

$$P(\text{acquisto pianificato}) = \frac{\text{numero di soggetti che hanno pianificato l'acquisto}}{\text{numero totale di soggetti}}$$

$$= \frac{250}{1000} = 0.25$$

Le probabilità considerate sono definite “marginali” perché possono essere ottenute marginalizzando opportunamente la tabella di contingenza.

Probabilità congiunta

Le **probabilità congiunte** (per esempio la probabilità che l'acquisto del televisore venga pianificato ed effettivamente realizzato) si riferiscono al verificarsi simultaneo di due o più eventi.

La probabilità che un soggetto pianifichi l'acquisto di un televisore e lo acquisti effettivamente entro dodici mesi si ottiene rapportando il valore che compare nella prima riga e nella prima colonna della tabella al numero totale dei soggetti nel campione.

$$P(\text{acquisto pianificato } e \text{ acquisto effettivo}) =$$

$$= \frac{\text{numero di soggetti che hanno pianificato } e \text{ realmente effettuato l'acquisto}}{\text{numero totale di soggetti}}$$

$$= \frac{200}{1000} = 0.20$$

Probabilità marginale in base alla probabilità congiunta

Una volta introdotto il concetto di probabilità congiunta, la probabilità marginale può essere interpretata in modo diverso.

Per esempio, dati due eventi B_1 e B_2, sotto certe condizioni la probabilità dell'evento *A*, *P(A)*, può essere calcolata sommando la probabilità congiunta di *A* e B_1 alla probabilità congiunta di *A* e B_2.

Calcolo della probabilità marginale

$$P(A) = P(A \text{ e } B_1) + P(A \text{ e } B_2) + \cdots + P(A \text{ e } B_k) \qquad (4.2)$$

dove $B_1, B_2, \ldots B_k$ sono eventi mutuamente incompatibili e collettivamente esaustivi.

Probabilità marginale in base alla probabilità congiunta

Due eventi si dicono **incompatibili** se non possono verificarsi entrambi. Un insieme di eventi è **collettivamente esaustivo** se almeno uno degli eventi si verifica sicuramente. Ad esempio "*estrarre una carta bianca* **e** *una carta nera*" da un mazzo di carte sono due eventi incompatibili e collettivamente esaustivi.

Utilizziamo ora l'equazione (4.2) per calcolare la probabilità marginale che un soggetto scelto a caso abbia pianificato l'acquisto di un televisore. Si ha:

$$\begin{aligned} P(\text{acquisto pianificato}) &= P(\text{acquisto pianificato } e \text{ acquisto effettivo}) \\ &\quad + P(\text{acquisto pianificato } e \text{ non acquisto effettivo}) \\ &= \frac{200}{1000} + \frac{50}{1000} \\ &= \frac{250}{1000} = 0.25 \end{aligned}$$

La probabilità di un'unione di eventi

Dobbiamo ancora individuare una regola che ci permetta di calcolare la probabilità dell'evento "*A* o *B*", evento che si realizza quando è vero l'evento *A* *oppure* l'evento *B* oppure sono veri entrambi.

Come si calcola la probabilità che un soggetto abbia pianificato l'acquisto di un televisore o lo acquisti effettivamente?

$$
\begin{aligned}
P(\text{acquisto pianificato } o \text{ acquisto effettivo}) &= P(\text{acquisto pianificato } e \text{ acquisto effettivo}) \\
&\quad + P(\text{acquisto non pianificato } e \text{ acquisto effettivo}) \\
&\quad + P(\text{acquisto pianificato } e \text{ non acquisto effettivo}) \\
&= \frac{50}{1000} + \frac{100}{1000} + \frac{200}{1000} = \frac{350}{1000}
\end{aligned}
$$

La probabilità di un'unione di eventi

La probabilità dell'unione di eventi può essere calcolata ricorrendo ad una regola generale.

Regola generale per il calcolo della probabilità di un'unione

La probabilità dell'evento "*A o B*" si ottiene sommando la probabilità di *A* e di *B* e sottraendo la probabilità dell'evento "*A e B*".

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ e } B) \tag{4.3}$$

Applicando questa regola all'esempio otteniamo il seguente risultato:

$$\begin{aligned} P(&\text{acquisto pianificato} \\ &\textit{o}\ \text{acquisto effettivo}) = P(\text{acquisto pianificato}) + P(\text{acquisto effettivo}) \\ &\qquad - P(\text{acquisto pianificato}\ \textit{e}\ \text{acquisto effettivo}) \\ &\qquad = \frac{250}{1000} + \frac{300}{1000} - \frac{200}{1000} = \frac{350}{1000} = 0.35 \end{aligned}$$

La probabilità di un'unione di eventi

La regola generale che abbiamo dato per il calcolo di "*A o B*", consiste nel sommare le probabilità dei due eventi e poi sottrarre la probabilità dell'evento "*A e B*" ("intersezione") che è stata inclusa due volte nel calcolo delle probabilità di *A* e di *B*. In alcuni casi nel calcolo della probabilità di un'unione la probabilità congiunta, essendo nulla, non deve essere sottratta.

Quindi, nel caso di eventi incompatibili, la regola per il calcolo della probabilità dell'unione si riduce a:

Regola per il calcolo della probabilità dell'unione di due eventi incompatibili

Se due eventi A e B sono incompatibili, la probabilità di "A o B" è si ottiene sommando la probabilità di A alla probabilità di B

$$P(A \text{ o } B) = P(A) + P(B) \tag{4.4}$$

Probabilità condizionata

Come si modifica il calcolo di una probabilità quando si dispone di un'informazione a priori sugli eventi coinvolti? Si dice **probabilità condizionata**, e si indica con il simbolo *P(A | B)*, la probabilità che si verifichi l'evento *A* sapendo che l'evento *B* si è verificato.

Probabilità condizionata

La probabilità di A dato B si ottiene rapportando la probabilità di A e B alla probabilità di B.

$$P(A \mid B) = \frac{P(A \text{ e } B)}{P(B)} \tag{4.5a}$$

La probabilità di B dato A si ottiene rapportando la probabilità di A e B alla probabilità di A.

$$P(B \mid A) = \frac{P(A \text{ e } B)}{P(A)} \tag{4.5b}$$

dove

$P(A \text{ e } B)$ = probabilità congiunta di A e B

$P(A)$ = probabilità marginale di A

$P(B)$ = probabilità marginale di B

Probabilità condizionata

Supponiamo di voler calcolare la probabilità che un televisore venga acquistato da un soggetto che ne ha pianificato l'acquisto, vale a dire

$$P\,(\text{acquisto effettivo} \mid \text{acquisto pianificato}).$$

Sapendo che il soggetto ha pianificato l'acquisto di un televisore, lo spazio campionario si riduce a soli 250 soggetti.

Tabella 4.1 *Acquisto di televisori di grandi dimensioni*

	Acquisto effettivo		
Acquisto pianificato	**Sì**	**No**	**Totale**
Sì	200	50	250
No	100	650	750
Totale	300	700	1000

Dei 250 soggetti in questione, 200 hanno effettivamente acquistato il televisore.

Quindi la probabilità condizionata che un soggetto acquisti un televisore dato che ne ha pianificato l'acquisto è pari a:

Probabilità condizionata

$$P(\text{acquisto effettivo} \mid \text{acquisto pianificato})$$

$$= \frac{\text{numero di soggetti che hanno pianificato l'acquisto } e \text{ lo hanno realizzato}}{\text{numero di soggetti che hanno pianificato l'acquisto}}$$

$$= \frac{200}{250} = 0.80$$

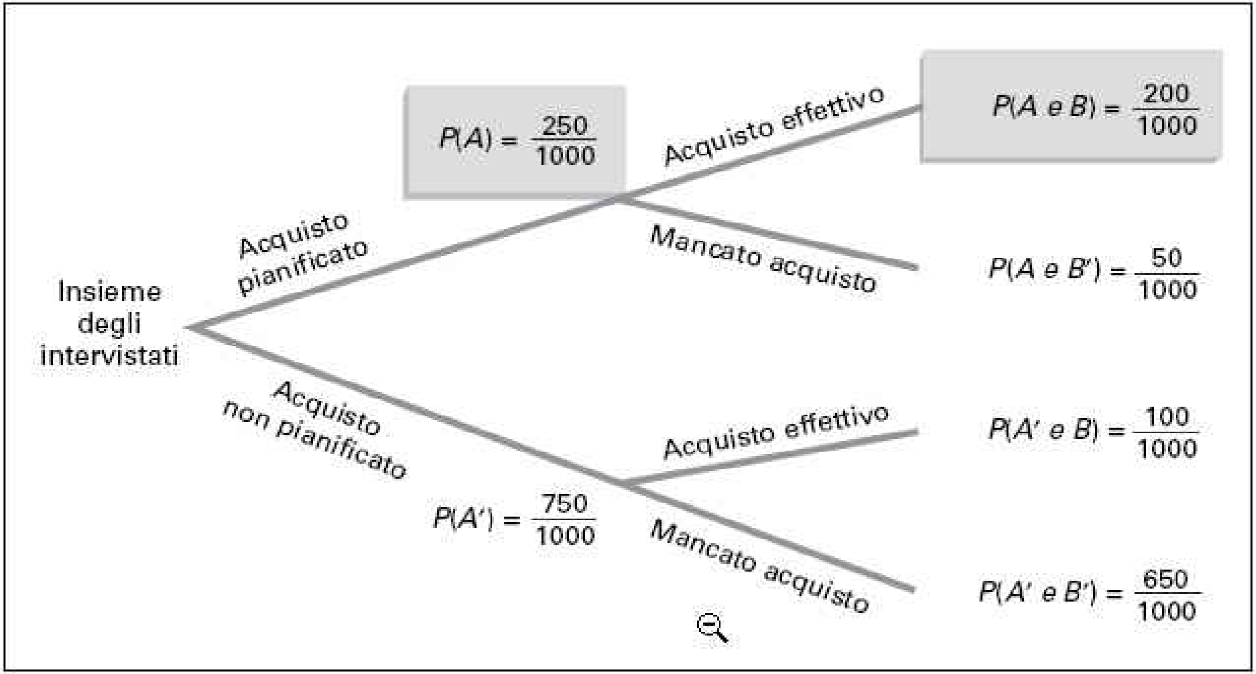
Si arriva allo stesso risultato applicando l'equazione (4.5b) e ponendo A = "acquisto pianificato" e B = "acquisto effettivo".

$$P(\text{acquisto effettivo} \mid \text{acquisto pianificato}) = \frac{200/1000}{250/1000} = \frac{200}{250} = 0.80$$

Albero delle decisioni

Una rappresentazione alternativa alla tabella di contingenza è data dal cosiddetto **albero delle decisioni**.

Nella Figura 4.3 è rappresentato l'albero delle decisioni relativo all'esempio citato.



Indipendenza

Quando il verificarsi di un evento non influenza la probabilità che se ne verifichi un altro, si dice che i due eventi sono statisticamente indipendenti.

L'**indipendenza** si definisce quindi come segue:

Indipendenza statistica

$$P(A \mid B) = P(A) \tag{4.6}$$

dove

$P(A \mid B)$ = probabilità condizionata di A dato B

$P(A)$ = probabilità marginale di A

Nel nostro esempio quindi gli eventi "acquisto pianificato" e "acquisto effettivo" non sono statisticamente indipendenti, poiché la conoscenza dell'uno influenza la probabilità che si verifichi l'altro.

La probabilità composta

Dalla definizione di probabilità condizionata si può ottenere una regola che permette di calcolare la probabilità congiunta di due eventi conoscendo la probabilità condizionata.

Data l'equazione (4.5a):

$$P(A \mid B) = \frac{P(A \text{ e } B)}{P(B)}$$

risolvendo rispetto alla probabilità congiunta $P(A \text{ e } B)$ otteniamo la **regola della probabilità composta**.

Regola della probabilità composta

La probabilità dell'evento congiunto A e B si ottiene moltiplicando la probabilità condizionata di A dato B per la probabilità marginale di B:

$$P(A \text{ e } B) = P(A \mid B)P(B) \tag{4.7}$$

La regola della probabilità composta per eventi indipendenti si ottiene sostituendo $P(A)$ a $P(A \mid B)$ nell'equazione (4.7).

La probabilità composta

Se A e B sono statisticamente indipendenti, la probabilità di A e B si ottiene dal prodotto delle due probabilità marginali.

$$P(A \text{ e } B) = P(A)P(B) \tag{4.8}$$

Abbiamo quindi a disposizione due modi per valutare l'indipendenza tra due eventi:

1. Gli eventi A e B sono statisticamente indipendenti se e solo se $P(A \mid B) = P(A)$.
2. Gli eventi A e B sono statisticamente indipendenti se e solo se $P(A \text{ e } B) = P(A)P(B)$.

Abbiamo definito la probabilità marginale (equazione 4.2):

$$P(A) = P(A \text{ e } B_1) + P(A \text{ e } B_2) + \ldots + P(A \text{ e } B_k)$$

Utilizzando la regola della probabilità composta è possibile calcolare la probabilità marginale.

La probabilità composta

Probabilità marginale

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k) \qquad (4.9)$$

essendo $B_1, B_2, \ldots B_k$ eventi incompatibili e collettivamente esaustivi.

La formula può essere applicata ai dati riportati nella Tabella 4.1. Per esempio, se vogliamo calcolare la probabilità che un soggetto abbia pianificato l'acquisto di un televisore, possiamo procedere nel seguente modo:

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2)$$

con

$P(A)$ = probabilità che sia stato pianificato l'acquisto

$P(B_1)$ = probabilità che l'acquisto venga effettuato

$P(B_2)$ = probabilità che l'acquisto non venga effettuato

ottenendo così:

$$P(A) = \left(\frac{200}{300}\right)\left(\frac{300}{1000}\right) + \left(\frac{50}{700}\right)\left(\frac{700}{1000}\right)$$
$$= \frac{200}{1000} + \frac{50}{1000} = \frac{250}{1000} = 0.25$$

Il teorema di Bayes

Quando calcoliamo una probabilità condizionata, utilizziamo l'informazione circa il verificarsi di un evento per determinare la probabilità che un altro evento si verifichi.

Un'estensione di questo concetto ci permetterà di aggiornare una probabilità sulla base di nuove informazioni e di calcolare la probabilità che un certo effetto sia il risultato di una particolare causa.

La procedura che illustreremo va sotto il nome di **teorema di Bayes**, si ottiene, a partire dalla definizione di probabilità condizionata, applicando la regola della probabilità composta.

Il teorema di Bayes

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k)} \quad (4.10)$$

dove B_i è l'i-esimo di k eventi a due a due incompatibili e collettivamente esaustivi.

Applicazione del teorema di Bayes

Il responsabile marketing di una società che produce giocattoli sta analizzando le chance sul mercato di un nuovo gioco.

Nel passato della compagnia solo il 40% dei nuovi giocattoli ha avuto successo di mercato, mentre il restante 60% non ha ottenuto un riscontro positivo.

L'80% dei giocattoli di successo avevano ricevuto un previo giudizio positivo da parte degli esperti di marketing della società, contro il solo 30% dei giocattoli poi rivelati fallimentari sul mercato.

Qual è la probabilità che il nuovo giocattolo sarà premiato dal mercato, sapendo che gli esperti di marketing della società lo hanno valutato positivamente?

Applicazione del teorema di Bayes

Tabella 4.2 *Applicazione del teorema di Bayes a un problema di marketing*

Evento S_i	Probabilità a priori $P(S_i)$	Probabilità condizionata $P(F \mid S_i)$	Probabilità congiunta $P(F \mid S_i)P(S_i)$	Probabilità aggiornata $P(S_i \mid F)$
S = giocattolo di successo	0.40	0.80	0.32	$0.32/0.50 = 0.64 = P(S/F)$
S' = giocattolo non di successo	0.60	0.30	0.18	$0.18/0.50 = 0.36 = P(S'/F)$
			0.50	

F = "giudizio positivo del marketing"; F'= "giudizio negativo del di marketing"

$$P(S \mid F) = \frac{P(F \mid S)P(S)}{P(F \mid S)P(S) + P(F \mid S')P(S')}$$

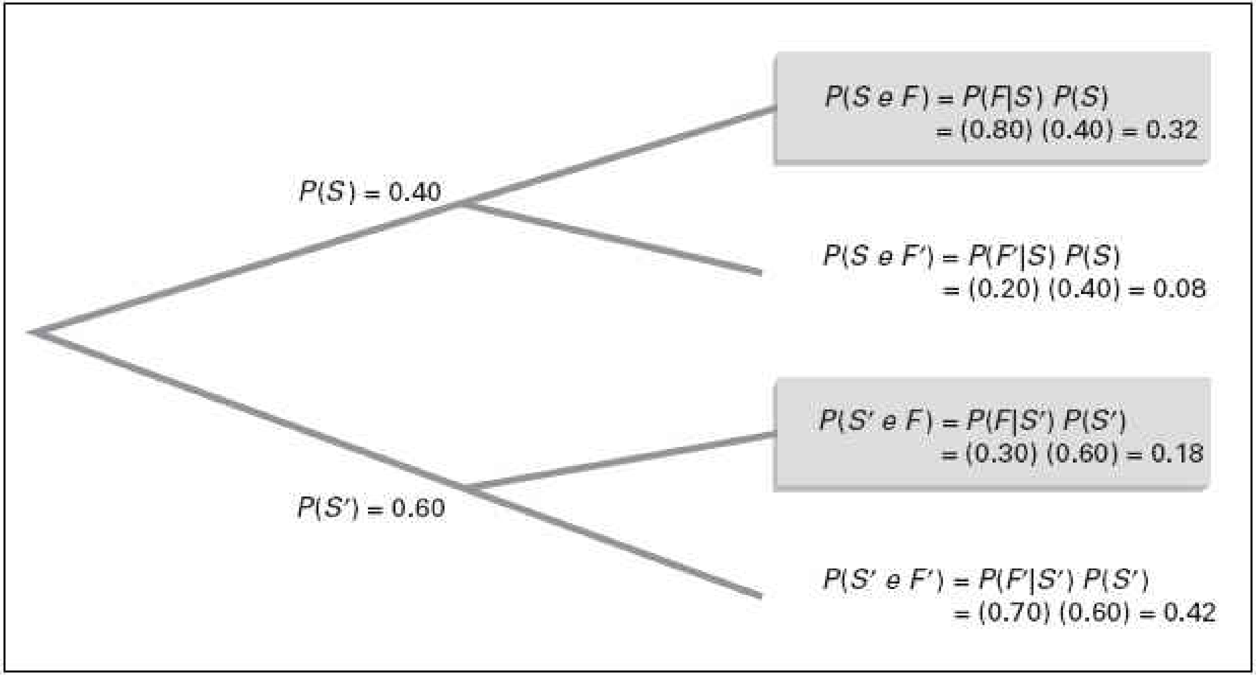
$$= \frac{(0.80)(0.40)}{(0.80)(0.40) + (0.30)(0.60)}$$

$$= \frac{0.32}{0.32 + 0.18} = \frac{0.32}{0.50} = 0.64$$

Applicazione del teorema di Bayes

Applicazione del teorema di Bayes a un problema di marketing

Evento S_i	Probabilità a priori $P(S_i)$	Probabilità condizionata $P(F \mid S_i)$	Probabilità congiunta $P(F \mid S_i)P(S_i)$	Probabilità aggiornata $P(S_i \mid F)$
S = televisore di successo	0.40	0.80	0.32	$0.32/0.50 = 0.64 = P(S \mid F)$
S' = televisore non di successo	0.60	0.30	0.18	$0.18/0.50 = 0.36 = P(S' \mid F)$
			0.50	



Applicazione del teorema di Bayes

Applicazione del teorema di Bayes a un problema di diagnosi medica

Evento D_i	Probabilità a priori $P(D_i)$	Probabilità condizionata $P(T \mid D_i)$	Probabilità congiunta $P(T \mid D_i)P(D_i)$	Probabilità aggiornata $P(D_i \mid T)$
D = Soggetto malato	0.03	0.90	0.0270	$0.0270/0.0464 = 0.582 = P(D \mid T)$
D' = Soggetto sano	0.97	0.02	0.0194	$0.0194/0.0464 = 0.418 = P(D' \mid T)$
			0.0464