15 Imparare dall'esperienza

La regola di Bayes svolge un ruolo centrale nelle applicazioni della probabilità personale, perché ci offre un modo per rappresentare il cambiamento razionale di credenza alla luce dell'emergere di nuovi dati.

LA REGOLA DI BAYES

Quando si hanno in mente le probabilità intese come frequenze la regola di Bayes è di scarsissimo interesse: è solo una regola come molte altre (a pag. 103 l'abbiamo derivata in poche righe dalla definizione di probabilità condizionale). Per molti problemi (i paraurti, le tarantole, i bambini con un avvelenamento da metalli pesanti, i taxi) la regola consente di rendere i calcoli più veloci, ma dal punto di vista frequentistico questo è tutto.

La regola è invece molto importante per le probabilità personali e per qualsiasi altro tipo di probabilità intesa come credenza, tanto che oggi gli approcci che trattano le probabilità come credenze vengono spesso chiamati "bayesiani". Se sentite uno statistico parlare di analisi bayesiana di un problema, potete essere certi che ha in mente una qualche variante delle idee che discuteremo in questo capitolo, solo che le varianti, dalla personale alla logicista, sono numerose: un bayesiano che sa ragionare in modo autonomo, Irving John Good (*vedi* pag. 240), ha calcolato che possono esservi 46 656 modi di essere bayesiani!

LE IPOTESI

Nella vita quotidiana abbiamo continuamente a che fare con ipotesi, e la maggior parte delle nostre decisioni dipende da un confronto fra i dati a favore di ipotesi diverse.

Albert dovrebbe ritirarsi da questo corso? La data ultima per lasciarlo è domani, e finora lui è andato male. Il corso diventerà più difficile, come spesso accade (ipotesi 1), o resterà al livello attuale, circostanza che gli consentirebbe di superare l'esame finale (ipotesi 2)? E io dovrei o non dovrei parcheggiare in sosta vietata? Tornate a pag. 143. Quali erano le ipotesi? E le probabilità?

Dovremmo investire nell'energia nucleare o in polverizzatori migliori per le centrali a carbone? Quali sono le ipotesi? Questo è un problema più complicato, perché bisogna capire, appunto, quali ipotesi sono in gioco. Riuscite a proporne? E ad assegnare loro probabilità?

Louise deve iscriversi a un corso di laurea in farmacia? Per un verso il lavoro del farmacista le piace, è brava in biochimica e ha uno zio che possiede una farmacia bene avviata; per l'altro, ultimamente si sono laureati in farmacia moltissimi studenti, quindi potrebbe esserci troppa concorrenza nel settore. Quali sono le ipotesi? Riuscite a formularne qualcuna, con le probabilità corrispondenti?

Spesso è difficile enunciare un'ipotesi con precisione, ma più la si formula con chiarezza, più chiaramente si può ragionare su essa. Possiamo riflettere sulle ipotesi anche chiedendoci quanto è probabile ciascuna secondo noi, e a volte riusciamo perfino a usare i numeri per rappresentare queste nostre probabilità personali.

I NUOVI DATI

Quando pensiamo alla probabilità personale che un'ipotesi può avere per noi la riferiamo alle nostre conoscenze, alle nostre credenze, ai nostri pregiudizi ecc., ma non ci fermiamo a questo punto, perché impariamo sempre cose nuove. A meno che i nostri pregiudizi siano veramente robusti, i nuovi dati dovrebbero pur avere un qualche effetto su ciò che crediamo, e dunque sulle nostre probabilità personali.

Possiamo verificare quanto appena detto in tutti gli esempi che abbiamo citato. Albert può sentirsi dire che il corso diventa più difficile verso la fine ogni volta che è quel tale professore a tenerlo; io vengo a sapere che il vigile ha "battuto" la strada ieri sera, e non passa quasi mai due sere di seguito; i ricercatori che lavorano nei settori dell'ambiente e dell'energia e sono interessati al confronto fra energia nucleare e carbone ricevono ogni giorno nuove informazioni che devono "digerire" e inserire nei loro giudizi probabilistici; Louise voleva laurearsi in farmacia, ma viene a sapere che con i tagli al bilancio decisi dal governo vi saranno meno fondi per la sanità pubblica e i medici che prescriveranno troppe medicine, e troppo costose, ai loro pazienti finiranno nei guai; quindi il giro d'affari delle farmacie potrebbe contrarsi.

Può essere necessario rivedere o aggiornare un giudizio probabilistico in qualsiasi momento. Esistono regole per farlo? Secondo i bayesiani, la risposta è sì.

L'IDEA BAYESIANA

Ricordiamo la regola di Bayes: questa vale per insiemi di ipotesi esaustive e reciprocamente esclusive H_i (gli insiemi di questo tipo sono detti "partizioni"). Per qualsiasi H_j di una partizione $(H_1, H_2, ..., H_n)$, la regola di Bayes dice:

$$\Pr(H_j/E) = \frac{\Pr(H_j)\Pr(E/H_j)}{\sum[\Pr(H_i)\Pr(E/H_i)]}.$$

L'idea bayesiana consiste nel pensare E come un insieme di nuovi dati e le H_i come ipotesi rivali. A questo punto è chiaro che per qualsiasi H_j

$\Pr(H_j)$ è la nostra probabilità iniziale, prima dell'acquisizione di nuovi dati;
E rappresenta i nuovi dati;
$\Pr(H_j/E)$ dovrebbe essere la nostra probabilità personale alla luce di questi nuovi dati.

Alla base c'è un'idea di apprendimento dall'esperienza, e finalmente possiamo capire perché la coerenza è così importante. Il nostro argomento è:

(1) le probabilità personali devono essere coerenti;
(2) probabilità personali coerenti soddisfano le regole di base della probabilità;
(3) perciò soddisfano la regola di Bayes;
(4) quindi ci permettono di dare forma all'idea di apprendimento dall'esperienza.

PROBABILITÀ A PRIORI

In ognuno dei nostri esempi, da quello su Albert che medita di ritirarsi dal corso a quello su Louise che pensa alla possibilità di laurearsi in farmacia, sono in gioco ipotesi rivali.

Supponete che vi sia chiaro quali sono le ipotesi rivali e di sapere come organizzarle in modo che siano esaustive e reciprocamente esclusive rispetto alle vostre conoscenze di sfondo e rispetto alle vostre credenze attuali. Chiamiamo le ipotesi rivali:

$H_1, H_2, ..., H_n$.

Ora supponiamo che rappresentiate i vostri attuali gradi di credenza in queste ipotesi come probabilità personali: le vostre probabilità a priori, basate su un primo esame delle ipotesi e delle informazioni di cui disponete. Qualsiasi probabilità da voi invocata in risposta alle domande fatte sopra sarebbe un esempio di probabilità a priori.

Supponiamo dunque che le vostre probabilità personali a priori siano

$\Pr(H_1), \Pr(H_2), ..., \Pr(H_n)$.

ESEMPIO: I PARAURTI (1)

Riprendiamo l'esempio di pagg. 82-83. Una fabbrica di automobili ha stipulato un contratto per acquistare paraurti da due fornitori, la Bolt&Co. e la Acme Inc.; la Bolt le fornisce il 40% dei paraurti, la Acme il 60%. Ogni pezzo viene sottoposto a un controllo di qualità e quelli che lo superano sono dichiarati affidabili.

Il 96% dei paraurti della Acme risulta affidabile, mentre la Bolt ha avuto alcuni problemi alla catena di montaggio e ultimamente è risultato affidabile solo il 72% dei suoi pezzi.

Arriva un camion con una consegna, e vogliamo sapere da quale azienda proviene. Le ipotesi sono:

H_1 = A = la consegna arriva dalla Acme;
H_2 = B = la consegna arriva dalla Bolt.

Le probabilità a priori sono:

$\Pr(H_1) = \Pr(A) = 0{,}6 \qquad \Pr(H_2) = \Pr(B) = 0{,}4.$

LA VEROSIMIGLIANZA

Avete alcune probabilità a priori e venite a sapere qualcosa di nuovo, cioè venite a conoscenza di nuovi dati E. Volete dunque rivedere i vostri gradi di credenza per tenere conto di E. Come dovreste procedere?

Supponiamo che possiate stabilire quanto sarebbe verosimile E, caso per caso, se le varie ipotesi fossero corrette, cioè che possiate assegnare un valore alle vostre probabilità condizionali personali

$\Pr(E/H_1), \Pr(E/H_2), ..., \Pr(E/H_n)$.

Queste grandezze sono molto importanti, e risulterà comodo disporre di un nome apposito per designarle; il nome in questione è stato scelto da R.A. Fisher (*vedi* pag. 290), forse lo statistico più innovativo del Novecento. Sfortunatamente il nome da lui scelto confonde molto le idee: Fisher le ha infatti battezzate "verosimiglianze" delle ipotesi $H_1, H_2, ..., H_n$ alla luce dei dati E, e la funzione che ce le fornisce per un insieme di ipotesi dato è detta "funzione di verosimiglianza".

Questa terminologia genera confusione perché nel linguaggio comune "probabilità" e "verosimiglianza" sono sinonimi in numerosi contesti: spesso capita che abbiano significati pressoché uguali; ma Fisher usa "verosimiglianza" in un senso tecnico che non è strettamente uguale a quello di "probabilità".

Naturalmente una qualsiasi verosimiglianza data, per esempio $Pr(E/H_1)$, non è che una probabilità; ma va osservato che un insieme di probabilità rispetto a una partizione in ipotesi non obbedisce alla regola dell'additività: quindi le verosimiglianze rispetto alle diverse ipotesi di una partizione data non si sommano a formare una nuova verosimiglianza, mentre si sommano le probabilità degli elementi della partizione, perché l'insieme delle verosimiglianze associate a una partizione non è un insieme di probabilità degli elementi della partizione stessa. Ma facciamo un esempio di questo uso poco chiaro della parola "verosimiglianza".

I PARAURTI (2)

Supponiamo di prendere a caso uno dei paraurti dell'ultima consegna e, una volta esaminatolo, di dichiararlo affidabile. A pag. 83 abbiamo chiamato R questo nuovo dato.

$$Pr(E/H_1) = Pr(R/A) = 0{,}96.$$
$$Pr(E/H_2) = Pr(R/B) = 0{,}72.$$

Le verosimiglianze delle due ipotesi alla luce dei dati E sono queste, e la loro somma è uguale a 1,68, non a 1. (1,68 non significa assolutamente nulla.) Se i dati sono noti, l'insieme delle verosimiglianze di una partizione è ben definito, ma non è un insieme di probabilità delle ipotesi della partizione stessa rispetto ai dati.

Perché qui usiamo questa terminologia, che non contribuisce certo a fare chiarezza? Perché è assolutamente fissa e invariabile per tutti gli statistici fin da quando è stata introdotta da Fisher tre quarti di secolo fa.

LA PROBABILITÀ A POSTERIORI

Volete rivedere le vostre probabilità a priori alla luce di nuovi dati. Tali probabilità rivedute sono dette "a posteriori". "A priori" significa anteriore ai nuovi dati, "a posteriori" significa successivo ai nuovi dati. Le probabilità a posteriori di una partizione sono

$$Pr(H_1/E), Pr(H_2/E), ..., Pr(H_n/E).$$

Dunque la regola di Bayes mette in relazione probabilità a priori, probabilità a posteriori e verosimiglianze:

$$Pr(H_j/E) = \frac{Pr(H_j)Pr(E/H_j)}{\sum[Pr(H_i)Pr(E/H_i)]}.$$

I PARAURTI (3)

Le probabilità a posteriori sono:

$$Pr(H_1/E) = Pr(A/R) \qquad Pr(H_2/E) = Pr(B/R).$$

Applicando la regola di Bayes calcoliamo le probabilità a posteriori che la consegna venga dalla Acme o dalla Bolt:

$$Pr(H_1/E) = Pr(A/R) = 2/3 \quad Pr(H_2/E) = Pr(B/R) = 1/3.$$

LA FORMA CONDENSATA DELLA REGOLA DI BAYES

Torniamo alla regola. Il suo denominatore è

$$\sum[Pr(H_i)Pr(E/H_i)],$$

ma questa è una costante determinata dalla partizione; di fatto coincide con $Pr(E)$. Perciò per ogni H_j

$Pr(H_j/E)$ è proporzionale a $Pr(H_j) \times Pr(E/H_j)$,

ovvero, usando $\propto$ come abbreviazione di "è proporzionale a",

probabilità a posteriori $\propto$ probabilità a priori $\times$ verosimiglianza.

APPLICAZIONI RIPETUTE DELLA REGOLA DI BAYES

Inizialmente disponiamo di probabilità a priori definite su una partizione di ipotesi, poi acquisiamo nuovi dati E. Usando la regola di Bayes deriviamo le probabilità a posteriori e ne facciamo le nostre nuove probabilità personali per quella partizione.

Successivamente acquisiamo ancora altri dati F e possiamo usare le nuove probabilità personali come componente a priori della regola di Bayes, calcolando un secondo insieme di probabilità a posteriori.

Tutto ciò equivale (detto in forma abbreviata) a:

Prima probabilità a posteriori di H $\propto$ *[(probabilità a priori di H)* $\times$ *(verosimiglianza di H alla luce di E)].*

Seconda probabilità a posteriori di H $\propto$ *[(prima probabilità a posteriori di H)* $\times$ *(verosimiglianza di H&E alla luce di F)].*

Sia data una partizione in *n* ipotesi ($H_1, H_2, ..., H_n$) e partiamo dalle probabilità a priori $Pr(H_j)$.

Quando veniamo a conoscenza dei nuovi dati E, cominciano a interessarci le prime probabilità a posteriori $Pr(H_j/E)$, che possiamo calcolare con la regola di Bayes.

Quando acquisiremo ulteriori dati F, ci interesseranno le seconde probabilità a posteriori $Pr(H_j/E\&F)$, e potremo calcolare anche queste con la regola di Bayes, prendendo come probabilità a priori (rispetto a F) le prime probabilità a posteriori.

Otteniamo così una forma condizionale della regola di Bayes

$$Pr(H_j/E\&F) = \frac{Pr(H_j/E)Pr(F/H_j\&E)}{\sum[Pr(H_j/E)Pr(F/H_j\&E)]}.$$

I PARAURTI (4)

Abbiamo controllato un paraurti scelto a caso da una certa partita appena consegnata ed è risultato affidabile; questi sono i dati E.

Ora ne prendiamo un altro, sempre a caso, lo esaminiamo, e non risulta affidabile; questi sono i nuovi dati F.

Poiché entrambi i pezzi sono stati scelti a caso e gli esiti sono indipendenti,

$$Pr(F/A\&E) = Pr(\sim R/A) = 0{,}04 \qquad Pr(F/B\&E) = Pr(\sim R/B) = 0{,}28.$$

Le nostre probabilità personali dato E erano

$$Pr(A/E) = 2/3 \qquad Pr(B/E) = 1/3.$$

Applicando la regola di Bayes

$$Pr(A/E\&F) = \frac{Pr(A/E)Pr(F/A\&E)}{Pr(A/E)Pr(F/A\&E) + Pr(B/E)Pr(F/B\&E)}.$$

Dunque $Pr(A/E\&F) = 2/9$.

I PARAURTI (5)

L'ordine di acquisizione dei dati è importante? Cambierebbe qualcosa nel caso che il primo paraurti venisse ritenuto affidabile, e il secondo invece no? Stando alla regola di Bayes, non cambierebbe nulla: se scoprissimo prima F, cioè che un certo paraurti scelto a caso non è affidabile, calcoleremmo

$$Pr(A/F) = 3/17 \qquad Pr(B/F) = 14/17,$$

dopodiché, usando di nuovo la regola di Bayes, otterremmo

$$Pr(A/E\&F) = 2/9.$$

CAPOVOLGERE L'ORDINE

Questo risultato illustra una proprietà generale molto elegante della regola di Bayes: l'ordine di acquisizione dei dati non modifica la probabilità a posteriori finale.

L'ESEMPIO DELL'APPENDICITE

Nell'esempio dei paraurti le probabilità a priori e le verosimiglianze sono determinate da fatti oggettivi di tipo frequentistico relativi alle due aziende fornitrici, e qualsiasi persona alla quale fossero state date le stesse informazioni contenute nel testo sarebbe giunta alla stesse conclusioni; ora passiamo a una probabilità di tipo più soggettivo, quella che una persona abbia l'appendicite.

Stadio 1: un a priori molto generico

Ho scelto un nome a caso dall'elenco telefonico. Prima ho preso un cognome, "Waukey", che mi era completamente sconosciuto, poi ho composto il numero e mi ha risposto una persona di sesso maschile apparentemente molto giovane che mi ha detto di chiamarsi Richard Waukey. Non so quasi nulla di questa persona, a parte il fatto che è un maschio, probabilmente giovane, abita in una certa parte del paese e risponde al telefono (ma potrebbe anche avermi mentito sul suo nome). Qual è la probabilità che Richard abbia l'appendicite proprio in questo momento? Se dovessi dirlo ora ipotizzerei un valore molto basso, diciamo 8 milionesimi. Sia A l'evento "Richard ha l'appendicite"; potrei proporre un numero arbitrariamente basso (ma maggiore di zero) per rappresentare la mia opinione, per esempio:

$$Pr(A) = 0{,}000008.$$

Stadio 2: un a priori più utilizzabile
Ho scritto un numero così piccolo perché oggi l'appendicite non è per nulla frequente tra gli adulti, e nella maggioranza dei casi le persone i cui nomi compaiono sugli elenchi telefonici sono adulte.

Ora però vengo a conoscenza di un nuovo dato E: Richard ha quattordici anni. Mi dicono anche che in questa parte del paese, in un momento dato qualsiasi solo sei maschi giovani su centomila hanno l'appendicite. Così adesso posso formulare un giudizio significativo:

$$Pr(A/E) = 0{,}00006.$$

Stadio 3: una verosimiglianza
Vengo anche a sapere che Richard ha i brividi, è irritabile e di cattivo umore, ha un po' di mal di testa e dopo l'ultimo pasto ha vomitato; chiameremo V (per "vomito") questi sintomi, frequenti nei pazienti affetti da appendicite. Giudico che la verosimiglianza del vomito, date l'appendicite e l'informazione di fondo, sia

$$Pr(V/A\&E) = 0{,}8.$$

Stadio 4: una partizione
Ma il ragazzo non potrebbe avere qualche altra malattia? Mi vengono in mente altre due ipotesi:

B: fondamentalmente Richard sta bene. Il suo è solo un disturbo passeggero, causato da qualcosa che ha mangiato.
C: ha un'intossicazione alimentare.

Supponiamo per semplicità che siano in gioco solo le tre ipotesi citate.
Giudico che $Pr(C/E) = 0{,}00009$; perciò

$$Pr(B/E) = 1 - 0{,}00009 - 0{,}00006 = 0{,}99985.$$

Stadio 5: altre verosimiglianze
Giudico che i sintomi di Richard siano più simili a quelli di un'intossicazione alimentare che a quelli di un'appendicite. D'altronde non è molto verosimile che un ragazzo ragionevolmente sano stia vomitando proprio adesso, quindi giudico, personalmente, che

$$Pr(V/C\&E) = 0{,}9.$$
$$Pr(V/B\&E) = 0{,}00001.$$

Stadio 6: la malattia si sviluppa
Il malessere peggiora rapidamente. Richard lamenta forti dolori all'addome, a destra e appena sopra l'inguine, e per non comprimere questa parte sta sdraiato con la coscia destra lievemente piegata; quando la si tocca con un dito la parte resiste alla pressione. Questi sintomi non sono sempre presenti in un'appendicite accompagnata da vomito, ma sono molto più rari quando ci sono vomito e un'intossicazione alimentare e pressoché ignoti nei ragazzi con una grave indigestione ma che sostanzialmente stanno bene. Indicando questa sintomatologia con D, valuto così le verosimiglianze:

$$Pr(D/A\&V\&E) = 0{,}6$$
$$Pr(D/C\&V\&E) = 0{,}02$$
$$Pr(D/B\&V\&E) = 0{,}00001.$$

TIRIAMO LE SOMME DELLE INFORMAZIONI DISPONIBILI

Intuitivamente, allo stadio 1 diremmo che non c'è alcuna probabilità che Richard abbia l'appendicite.

Forse allo stadio 2 c'è una piccola probabilità, ma ancora abbastanza trascurabile.

Allo stadio 4 c'è una probabilità reale, ma non si può ancora affermare che questa non è un'intossicazione alimentare, e praticare un'appendicectomia su una persona che non ha l'appendicite ma solo un'intossicazione significa mettersi in guai seri; ma anche non operare un'appendicite significa mettersi nei guai.

Così, giunti allo stadio 5 ci troviamo di fronte a un serio dilemma; ma alla fine, allo stadio 6, siamo abbastanza sicuri che Richard ha l'appendicite.

Quanto sicuri? La regola di Bayes ci permette di calcolarlo.

USARE LA REGOLA DI BAYES ALLO STADIO 4

Vogliamo stabilire $Pr(A/V\&E)$.

$$Pr(A/V\&E) = [Pr(A/E) \times Pr(V/A\&E)] \div K,$$
dove la costante di proporzionalità K è uguale a
$$[Pr(A/E) \times Pr(V/A\&E)] + [Pr(C/E) \times Pr(V/C\&E)] + [Pr(B/E) \times Pr(V/B\&E)].$$

Sappiamo già quali valori numerici dobbiamo inserire in questa formula per giungere al risultato che ci interessa:

$$\begin{aligned} &Pr(A/V\&E) \approx 48/139 \approx 0{,}35,\\ &Pr(C/V\&E) \approx 81/139 \approx 0{,}58,\\ &Pr(B/V\&E) \approx 10/139 \approx 0{,}07. \end{aligned}$$

Che cosa ci ha insegnato il nuovo dato, cioè che Richard sta vomitando? Innanzitutto che sta male; la probabilità a priori che stesse "sostanzialmente bene", o quantomeno non avesse un'intossicazione alimentare né un'appendicite, era praticamente uguale a 1, ma adesso è crollata al 7%.

Anche quel 7% può sembrare un valore troppo alto: se sta vomitando, non dovrebbe senz'altro avere qualcosa di serio? Non è detto. Questo valore deriva dai tassi di base: nella zona geografica in questione non sono frequenti le intossicazioni alimentari né le appendiciti, ma il fatto che una persona vomiti non è raro. Nel gruppo rilevante, gli adolescenti, la buona salute è talmente diffusa che il sintomo potrebbe ben essere dovuto, per esempio, al fatto che Richard si è ingozzato di ciliege.

Quel 7% è anche un esempio di come la regola di Bayes ci costringa all'onestà intellettuale: molti di noi salterebbero alla conclusione "Richard Waukey sta male" senza rispettare le nostre stesse valutazioni a priori di verosimiglianze e probabilità.

E il dilemma "intossicazione alimentare o appendicite"? La probabilità a posteriori della prima è aumentata rispetto alla seconda, ma non di molto. Il medico non può ancora formulare una diagnosi sicura.

USARE LA REGOLA DI BAYES ALLO STADIO 5

Idealmente, l'apprendimento è un processo cumulativo: si impara una cosa e si va avanti, imparando qualcos'altro. Ora, è sempre possibile sostituire la regola di Bayes con un calcolo diretto che parta dai primi princìpi, dalle regole di base della probabilità; ma una volta trovata la probabilità a posteriori allo stadio 5, quando si comincia ad analizzare una nuova informazione (il dolore), questa probabilità a posteriori 5 può essere usata come nuova probabilità a priori allo stadio 6:

$$Pr(A/D\&V\&E) = [Pr(A/V\&E) \times Pr(D/A\&V\&E)] \div K^*.$$

La costante di proporzionalità K^* è

$$[Pr(A/V\&E) \times Pr(D/A\&V\&E)] + [Pr(C/V\&E) \times Pr(D/C\&V\&E)] + [Pr(B/V\&E) \times Pr(D/B\&V\&E)].$$

Inserendo i valori numerici otteniamo, grossomodo, che:

$$\begin{aligned} &Pr(A/D\&V\&E) \approx 160/169 \approx 0{,}947,\\ &Pr(C/D\&V\&E) \approx 9/169 \approx 0{,}053,\\ &Pr(B/D\&V\&E) \text{ è trascurabile.} \end{aligned}$$

Le nuove informazioni sui dolori addominali di Richard hanno creato valori numerici completamente nuovi; ora sono praticamente certo che il ragazzo abbia qualcosa di peggio di un'indigestione, e la mia opinione ben meditata è che sia estremamente probabile, sebbene non assolutamente certo, che Richard abbia l'appendicite.

NON SMETTETE DI PENSARE!

Oggi esistono software diagnostici fondati sulla regola di Bayes, ed è stato scoperto che quando si verifica un disaccordo fra la diagnosi prodotta con uno di questi software e quella di un medico, posti di fronte a un paziente con gli stessi sintomi, o sintomi molto simili, gli altri medici, studiando il caso, danno ragione più spesso al software che al collega. Ma non usate dei calcoli meccanici come scusa per non pensare!

La coerenza è una forma di non contraddittorietà induttiva. Dopo aver raggiunto lo stadio 6 un medico può rivedere le proprie probabilità e verosimiglianze personali, e la regola di Bayes può garantirgli solo che il suo sistema complessivo di gradi di credenza non contenga contraddizioni. Una volta raggiunto lo stadio 6, un medico prudente potrebbe pensare: "Forse ho sottovalutato la possibilità che Richard accusi questi dolori in seguito a un'intossicazione alimentare, è meglio che ci ripensi"; potrebbe rivedere qualcuna delle probabilità e delle verosimiglianze usate fino a quel momento e raggiungere, a quello stadio, una probabilità a posteriori diversa.

I RAPPORTI FRA LE VEROSIMIGLIANZE

Che cosa ci dicono in realtà i dati? Se una persona riesce a rappresentare i propri gradi di credenza con probabilità personali coerenti, la regola di Bayes le suggerisce una risposta: individui diversi hanno probabilità a priori diverse, ma alla luce di nuovi dati E aggiornano le proprie probabilità tutti nello stesso modo, cioè moltiplicandole per le verosimiglianze. Alla luce di E un'ipotesi con una maggiore verosimiglianza relativa sale nella scala delle credenze di tutti, mentre una con una verosimiglianza relativa minore scende; sono i rapporti fra le verosimiglianze a determinare l'azione dei nuovi dati sulle credenze. È anche per questo che Fisher considerava tanto importanti le verosimiglianze delle ipotesi rispetto ai dati disponibili; è un peccato che abbia scelto un nome così equivoco per un'idea di tale importanza!

Un'altra ragione per cui questa idea è importante è che in genere, alla luce dei dati di cui disponiamo, ci mettiamo d'accordo molto più rapidamente sulle verosimiglianze relative di ipotesi diverse che sulle loro probabilità.

L'esempio della diagnosi di appendicite illustra bene il valore dei rapporti fra le verosimiglianze. Queste ultime sono state precisate in dettaglio a tutti gli stadi dell'esame del caso di Richard Waukey, e dati i loro rapporti gli stadi successivi del calcolo potevano essere eseguiti in modo meccanico.

In generale, quando usiamo la regola di Bayes:

- partiamo dalle probabilità a priori;
- otteniamo nuovi dati e cominciamo a interessarci a qualche probabilità a posteriori;
- otteniamo queste ultime (come ci mostra la forma condensata della regola di Bayes) moltiplicando le probabilità a priori per le verosimiglianze.

I rapporti fra verosimiglianze condensano tutte le informazioni (pertinenti alle ipotesi in esame) fornite dai nuovi dati.

> Dal punto di vista bayesiano i rapporti fra le verosimiglianze di ipotesi rivali ricapitolano il valore probatorio di ogni nuova informazione.

Un matematico esprimerebbe questa idea per mezzo di una funzione di verosimiglianza. La funzione di verosimiglianza di certi dati E (determinati) è quella funzione che calcola le verosimiglianze delle ipotesi alla luce di E.

UN MODELLO DI APPRENDIMENTO?

Alcuni filosofi affermano: «Ora che avete scoperto che Richard ha quattordici anni, ha vomitato e soffre dei dolori descritti sopra, dovreste assegnare alla vostra probabilità personale che abbia l'appendicite un valore di 94,5%». Ci è stato mostrato, dunque, il modo esatto in cui dovremmo elaborare le nuove informazioni; inoltre le verosimiglianze relative delle ipotesi rivali costituiscono un buon indicatore della significatività dei nuovi dati.

Per alcuni filosofi, dunque, questa lezione ha un valore del tutto generale: abbiamo scoperto la chiave di accesso a un'idea ragionevole di apprendimento dall'esperienza. Questa conclusione deriva, lo ripeto ancora una volta, dalla combinazione di tre idee diverse:

Idea 1: Collegare gradi personali di credenza e quotazioni di scommessa personali è plausibile.
Idea 2: Esiste un argomento plausibile in base al quale le quotazioni di scommessa devono soddisfare le regole di base della probabilità (e quindi la regola di Bayes).
Idea 3: La regola di Bayes può essere usata per modificare i giudizi di probabilità iniziali alla luce dei nuovi dati.

Secondo alcuni filosofi la regola di Bayes è la chiave dell'apprendimento dall'esperienza ed è utile perfino per affrontare il problema humeano dell'induzione; discuteremo questa applicazione della regola nel capitolo 21, nel quale metteremo anche in discussione l'uso della regola di Bayes come modello rigido di apprendimento dall'esperienza.

LA REGOLA DI JEFFREY

Ma c'è un problema: anche un uso liberale della regola di Bayes può non essere sufficiente per analizzare l'apprendimento dall'esperienza. Perché?

Il fatto è che la regola postula che l'apprendimento proceda attraverso passi molto precisi:

impariamo una certa verità E;
applichiamo la regola di Bayes;
impariamo una certa verità F;
applichiamo la regola di Bayes;
e così via, aggiornando costantemente le nostre probabilità a posteriori.

Talvolta è effettivamente questo il modo in cui impariamo, per esempio quando lavoriamo in laboratorio o quando effettuiamo campionature raccogliendo osservazioni di un tipo ben definito, ma nella vita reale spesso cambiamo opinione sulla base di osservazioni o esperienze molto meno precise. Ricordiamo, tanto per fare un esempio, il caso di Louise, che si sta chiedendo se sia opportuno iscriversi a un corso di laurea in farmacia. A un certo punto raccoglie alcune "voci": si mormora che ci saranno tagli al bilancio e meno fondi destinati alla sanità pubblica e che i medici che prescriveranno troppe medicine, e troppo costose, ai loro pazienti finiranno nei guai; quindi il giro d'affari delle farmacie potrebbe contrarsi. Queste sono voci molto utili, ma non c'è nulla di definito, nessuna verità consolidata, nessun fatto accertato su cui Louise possa fare assegnamento; solo voci.

Il logico induttivo Richard Jeffrey ha ampliato la regola di Bayes in modo da renderla applicabile anche a queste situazioni. La sua idea, grossomodo, è la seguente: siano le ipotesi della partizione di Louise L e ~L, dove

L = buone possibilità di lavoro per i laureati in farmacia nel giro di quattro anni.
Sia inoltre T (per "tagli") = ci saranno meno fondi per la sanità pubblica e i medici che prescriveranno troppe medicine, e troppo costose, ai loro pazienti finiranno nei guai.

Louise calcola che se T è vera, è una brutta notizia per i giovani laureati in farmacia.

$$\Pr(L/T) = 0{,}1.$$

Se invece è falsa, le possibilità di trovare un lavoro ben pagato sono discrete, anche se potrebbe esserci una certa penuria di posti per ragioni che non hanno a che vedere con i tagli:

$$\Pr(L/{\sim}T) = 0{,}8.$$

Louise ha sentito le voci, ma non ha saputo che T è vera: ha saputo solo che è probabilmente vera. Indichiamo con Pr* la probabilità alla luce di queste dicerie; in altre parole,

Pr*(L) sta per la probabilità di L alla luce delle voci che circolano sui tagli al bilancio.

Ma qual è il valore di Pr*(L)? Sulla base delle voci che ha raccolto, Louise immagina che tale valore sia 3/4:

$$\Pr{}^*(T) = 0{,}75 \qquad \Pr{}^*({\sim}T) = 0{,}25.$$

Non possiamo applicare la regola di Bayes, perché le voci non ci danno informazioni certe E da inserire nelle nostre formule; esiste però una revisione della regola, dovuta a Jeffrey, che ricorda un'altra regola, quella della probabilità totale, e per una partizione in due ipotesi L e ~L questa versione rivista della regola recita:

$$\Pr{}^*(L) = \Pr(L/T)\Pr{}^*(T) + \Pr(L/{\sim}T)\Pr{}^*({\sim}T).$$

La regola di Jeffrey consiglia a Louise di concludere che Pr*(L) = 0,275. Prima di sentire le voci Louise considerava molto probabile che una laurea in farmacia assicurasse un buon lavoro, ma se ragiona attenendosi alla regola di Jeffrey le sue prospettive di impiego appaiono sconfortanti.

I BAYESIANI LOGICISTI

Finora in questo capitolo ci siamo occupati di giudizi probabilistici puramente personali, ma esiste un intero spettro di tipi di probabilità intese come credenze; a un estremo ci sono quelle integralmente personali, all'altro c'è l'idea di probabilità logica.

Pensiamo una probabilità logica come una relazione logica fra un'ipotesi e i dati a suo favore; abbiamo già detto che l'economista John Maynard Keynes e (molto più tardi) il filosofo Rudolf Carnap sono stati fautori di una teoria logica della probabilità, ed entrambi hanno proposto un'assiomatizzazione di questa idea. Le probabilità come vengono concepite da Keynes e Carnap soddisfano le regole di base della probabilità, perciò soddisfano la regola di Bayes; dunque anch'esse riportano a questa regola l'apprendimento dall'esperienza.

Il contributo più importante a quello che potremmo chiamare "bayesianesimo logico" si deve però a uno scienziato della natura, il geofisico inglese Harold Jeffreys, la cui *Teoria della probabilità* (1939) contiene un gran numero di applicazioni della regola di Bayes a problemi scientifici pratici ed è ancora oggi una ricchissima fonte di idee per il programma di ricerca bayesiano.

IL REQUISITO DELL'EVIDENZA TOTALE

Una probabilità personale è la rappresentazione del grado di credenza di una persona. La nostra notazione per la probabilità categorica, Pr(A), si adatta perfettamente alle probabilità personali; ma la probabilità logica, concettualmente, è una relazione fra una proposizione e i dati di cui disponiamo su questa, quindi per rappresentarla ci serve Pr(H/E), la notazione che usiamo per la probabilità condizionale.

Le probabilità personali rappresentano i gradi di credenza, e abbiamo un'idea precisa di come combinare questi ultimi con le utilità in modo da calcolare le utilità attese e fare uso della teoria della decisione; queste probabilità hanno dunque un'interpretazione pratica diretta.

Ma qual è l'interpretazione pratica della probabilità logica? L'idea di base è che persone razionali dovrebbero massimizzare le loro attese condizionali usando probabilità logiche condizionate a tutti i dati disponibili; Carnap ha chiamato questo criterio "requisito dell'evidenza totale".

LOGICA E OPINIONE

Per la teoria della probabilità personale i giudizi di probabilità di una persona sono una questione privata nel senso che spettano al singolo, ma ciò non vuol dire che siano arbitrari. Esistono infatti alcune restrizioni cui tali giudizi devono sottostare: le probabilità personali devono essere induttivamente non contraddittorie, ovvero coerenti; se lo sono soddisfano le regole di base della probabilità.

Fatti salvi quelli appena citati, non vi sono vincoli sulla probabilità personale; a ogni modo, potremmo considerare un vincolo quello che a pagg. 183-184 abbiamo chiamato "principio della frequenza": se tutto ciò che sappiamo a proposito di un'ipotesi è una probabilità come frequenza, dovremmo rendere le nostre probabilità personali uguali alle frequenze.

La teoria logicista della probabilità afferma anche che esistono vincoli molto più forti della semplice coerenza, e che ci sono valutazioni razionali e corrette univoche della probabilità di un'ipotesi alla luce dei dati. Rudolf Carnap indebolì questa idea sostenendo l'esistenza di un continuum di "metodi induttivi", ciascuno dei quali corrisponde a un insieme di probabilità logiche a priori, ma il sistema da lui pubblicato in *Fondamenti logici della probabilità* (1950) e in *Continuum dei metodi induttivi* (1952) è troppo semplice, e non sembra avere applicazioni pratiche nelle scienze né nella vita quotidiana.

LEONARD J. SAVAGE

La teoria della probabilità personale fu inventata indipendentemente da Frank P. Ramsey e Bruno De Finetti, ma il suo successo si deve allo statistico americano Leonard J. Savage (1917-1971), che fece chiarezza sull'importanza dell'idea di probabilità personale (combinata con la regola di Bayes); il lavoro più famoso di Savage si intitola *Fondamenti dell'inferenza statistica* (1954). Qui però dobbiamo registrare un fatto singolare, una connessione accidentale (?) fra i primi vagiti, rispettivamente, della filosofia bayesiana e dell'informatica.

Durante la Seconda guerra mondiale Savage era capo assistente "statistico" di John von Neumann, il grande matematico che costruì i primi calcolatori elettronici e rese possibile l'èra dei computer e dell'informatica. Ma subito dopo la guerra era attivo anche un altro sostenitore delle idee bayesiane, il teorico della probabilità inglese Irving John Good, assistente di Alan M. Turing, ovvero di colui che definì l'idea di computer ideale e dimostrò il teorema fondamentale della computazione ideale (è per questo che oggi parliamo di "macchine di Turing"). È come se la versione contemporanea delle idee bayesiane fosse un sottoprodotto dell'età dei computer.

ESERCIZI

1 *Verosimiglianze.* Dite quali sono la probabilità a priori, la probabilità a posteriori e la verosimiglianza nel
(a) problema dei ragni (pag. 105). La somma delle verosimiglianze di G e H (rispetto a T) è uguale a 1?
(b) problema dei taxi (pagg. 105-106). La somma delle verosimiglianze di V e B (rispetto a T_b) è uguale a 1?

2 *Gli appunti perduti.* Immaginate di aver perso gli appunti del corso di logica induttiva. Formulate due ipotesi e non sapete quale sia quella vera.

B Ho lasciato gli appunti in biblioteca.
A Li ho lasciati nell'aula dove si tengono le lezioni del corso.

(a) Assegnate a B e A probabilità a priori che vi sembrano sensate. Usate numeri facili, ma assicuratevi che i valori assegnati siano abbastanza realistici in tutto il corso del problema.
Raccogliete un dato:

E Un'amica dice di avere visto gli appunti di qualcuno in biblioteca, ma non ha ragione di pensare che fossero i vostri.

(b) Assegnate le vostre verosimiglianze personali a B e A alla luce di E.
(c) Calcolate la probabilità a posteriori di B e A alla luce di E.
In seguito raccogliete un nuovo dato:

F Un altro amico dice di aver visto in biblioteca qualcosa con sopra il vostro nome.

(d) Assegnate verosimiglianze a B&E e A&E alla luce di F.
(e) Calcolate le nuove probabilità a posteriori di B e A a partire da quelle precedenti e dalle verosimiglianze.

3 *I funghi.* È notte alta, e nell'ospedale in cui lavorate viene ricoverato Jack Waukey. Vomita, è scosso dai brividi e accusa dolori al ventre. Secondo voi Pr(A), la probabilità che abbia l'appendicite, è 0,8.

Poi suo fratello Richard vi dice che al tramonto Jack ha raccolto un po' di funghi (i due erano in campeggio) e li ha mangiati per cena. Gli fate qualche domanda su questi funghi e Richard riesce a descriverne la forma abbastanza bene: quasi certamente si tratta di amanite, ma di che colore? L'*Amanita Cesarea*, color arancione chiaro, è un bocconcino delizioso, ma l'*Amanita Phalloides*, bianca, è mortale, tanto che il suo nome non scientifico è Berretto della morte.

I funghi sono stati raccolti al crepuscolo, e né Richard né Jack sono sicuri del colore. Quando sono ancora teneri i Berretti della morte sono bianchi all'interno, mentre l'*Amanita Cesarea*, molto simile per tutto il resto ai Berretti della morte, ha un colore giallastro; ma su questo non ci sono dati precisi. Richard dice: «Be', probabilmente sul giallo». Voi siete sotto pressione e valutate Pr(G) = 0,7.

Usate la regola di Jeffrey per trovare la nuova probabilità Pr*(A) che Jack abbia l'appendicite.

Badate che la decisione è veramente critica. In caso di avvelenamento è essenziale una lavanda gastrica e un'appendicectomia sarebbe fatale, mentre con un'appendice perforata la lavanda gastrica sarebbe pericolosissima ed è necessario operare subito.

4 *Voci di corridoio.* Poco prima dell'esame finale Joy, una studentessa di questo corso, ha sentito una voce di corridoio: il professore avrebbe detto a non si sa chi che nel test non ci saranno domande sulla fallacia del giocatore. Siano

G = Ci sarà una domanda sulla fallacia del giocatore.
P = Il professore ha detto a uno studente che non ci saranno domande sulla fallacia del giocatore.

Prima di aver sentito la voce in questione Joy, convinta che il professore non avrebbe lasciato trapelare nulla sul contenuto dell'esame finale, pensava che ci fosse una discreta probabilità (0,3) che nel test fosse inserita una domanda sulla fallacia del giocatore. D'altra parte considera il professore abbastanza attendibile, quindi per lei Pr(~G/P) = 0,9, e considera probabilmente giusta anche la voce che circola fra gli studenti (la sua Pr(P) è 0,8).

(a) Determinate la probabilità che Joy assegna a G dopo aver sentito la voce, enunciando e nominando tutte le regole che usate per aggiornare le probabilità.
(b) Secondo Joy, se un argomento sarà oggetto d'esame l'utilità di ripassarlo è 20, mentre quella di studiarne uno che non sarà oggetto d'esame è –5. Calcolate l'utilità attesa che ha per lei la decisione di ripassare la fallacia del giocatore prima di sentire la voce in questione e dopo averla sentita. Joy studia un argomento solo se l'utilità attesa dello studio è positiva: ripasserà la fallacia del giocatore dopo aver sentito la voce?

5 *Gli autobus.* State visitando Gotterdam, una grande città straniera che non conoscete per nulla; nel paese in cui vi trovate c'è solo un'altra grande città, Shamsterdam. Gli autobus sono ottimi in entrambe, ma non sapete nient'altro.

All'aeroporto trovate due tabelle con gli orari e i percorsi degli autobus, ma a entrambe manca una parte, ed è proprio quella con il nome della città: restano solo le ultime sei lettere, "terdam". Una di queste città che finiscono per "terdam" ha 20 linee di autobus, l'altra ne ha 200, e poiché Gotterdam è più grande di Shamsterdam voi immaginate che sia la prima ad avere 200 linee e la seconda ad averne 20. Siano

[200] = La proposizione che Gotterdam ha duecento linee di autobus.
[20] = La proposizione che Gotterdam ha solo venti linee di autobus.

(a) Assegnate le vostre probabilità personali Pr([200]) e Pr([20]).
(b) Siete arrivati la sera tardi e vi hanno portato a casa di un amico. Ora è mattina, state per uscire per una passeggiata e per quanto ne sapete vi trovate in una parte "tipica" della città. Enunciate la vostra probabilità personale a priori Pr(n) che il primo autobus che vedrete avrà il numero n.
(c) Qual è la vostra Pr($1 \le n \le 20$) personale? E la Pr($21 \le n \le 200$)?
(d) È mattina, uscite per fare una passeggiata e il primo autobus che vedete è il 19. Pensate: "È molto più probabile vedere un autobus con un numero basso se ci sono solo autobus con numeri bassi che se ci sono duecento linee".
Qual è la verosimiglianza (la vostra personale) di [200] alla luce di questo nuovo dato?
Qual è la verosimiglianza di [20]?
In quale rapporto sta la verosimiglianza di [20] con quella di [200]?
(e) Quali sono le vostre probabilità condizionali personali Pr([200]/19) e Pr([20]/19)?
(f) Fate una lunghissima camminata senza meta in tutte le zone di Gotterdam e incontrate cento autobus, tutti numerati da 1 a 20. Qual è la vostra probabilità condizionale personale che Gotterdam abbia duecento linee di autobus?
(g) Questa storia illustra l'idea di "apprendimento dall'esperienza"? In che modo?

6 *La fine del mondo.* Il professor John Leslie ha usato un ragionamento apparentemente simile per raggiungere una conclusione stupefacente: ha sfruttato la regola di Bayes per dimostrare che probabilmente il genere umano avrà presto fine! Prima ha considerato un problema in cui

> esistono una probabilità del 2% che un'urna nella quale c'è il mio nome contenga dieci nomi in tutto e una del 98% che ne contenga mille; queste probabilità a priori sono mie stime personali, fatte prima di qualsiasi estrazione. Se fra i primi tre estratti c'è il mio nome, la regola di Bayes mi dice che la probabilità a posteriori che nell'urna ci siano solo dieci nomi è __________.

(a) Riempite lo spazio bianco. In questo caso qual è la probabilità a posteriori?

Sarete d'accordo con Leslie, che prosegue:

> Una probabilità stimata di appena due su cento è stata riveduta e portata a due su tre,

e poi propone

di applicare un calcolo abbastanza simile alla nostra posizione nel tempo. Supponiamo, semplificando molto, che le sole alternative siano (I) che il genere umano abbia fine prima del 2150 e (II) che sopravviva per molte altre centinaia di migliaia di anni; e stabiliamo, di nuovo semplificando, che la probabilità che un umano sia in vita nella prima metà del XXI secolo sia 1/10 nel caso che il genere umano sia destinato a estinguersi presto, e solo 1/1000 nel caso che il genere umano non sia destinato a estinguersi presto.

Da dove emerge questo valore di 1/10? Leslie valuta che circa un decimo di tutti coloro che sono vissuti fino al 2150 fosse in vita negli anni novanta del Novecento, il decennio in cui scrive, mentre se gli esseri umani continueranno a esistere e moltiplicarsi nello stesso decennio era in vita solo 1/1000 del genere "destinato a durare a lungo", e prosegue:

> Supponiamo allora che, inizialmente, si consideri la probabilità che il genere umano abbia fine entro il 2150 pari a solo l'1%.

Fin qui accettiamo pure tutti i postulati del professor Leslie, il quale aggiunge poi che ognuno di noi sa di essere che *lui* (*lei*) è in vita nel XXI secolo. Prendiamo questo dato (si faccia il confronto con l'autobus 19 o con la possibilità che il proprio nome venga estratto dall'urna di cui sopra): usando la regola di Bayes, sostiene Leslie, la nuova stima riveduta della probabilità che la fine del mondo sia imminente è __________.

(b) Completate il calcolo di Leslie come se avesse senso usare la regola di Bayes qui come nel problema degli autobus o nell'esempio della lotteria.

L'argomento è molto stabile: non dipende in modo essenziale dalle probabilità personali che possiamo assegnare (se aumentiamo o diminuiamo un po' quell'1/10, quell'1/1000 e quell'1% la conclusione è la stessa). La probabilità che il genere umano abbia fine molto presto e quella che non l'abbia sono uguali; anzi, l'argomento funziona ancora meglio con l'ipotesi che il mondo sia destinato a finire di qui a pochi minuti!

(c) Spiegate perché.

Secondo Leslie la conclusione non è affatto sorprendente, considerati gli immensi arsenali nucleari di cui disponiamo, i virus assassini in via di preparazione, l'inquinamento globale, l'effetto serra, l'inverno nucleare, gli asteroidi vaganti e via dicendo; e queste deprimenti riflessioni ci aiutano a mostrare che la conclusione (fine del genere umano entro il 2150) è più probabile di quanto pensassimo a prima vista.

(d) Ma si può dire che rafforzano l'argomento costruito da Leslie per mezzo della regola di Bayes?

Ora torniamo agli autobus di Gotterdam: quando tutti quelli che incontriamo hanno un numero non superiore al 20, giungiamo alla virtuale certezza che a Gotterdam ci siano solo venti linee, e il professor Leslie sa di milioni e milioni di esseri umani, tutti nati prima dell'anno prossimo!

(e) Usando l'argomento di Leslie sulla fine del mondo qual è, con tutti questi dati, la probabilità che il genere umano abbia fine nel giro di un anno?

7 *La fallacia.* La vostra risposta a (e) è una *reductio ad absurdum* di questa forma di argomentazione: l'abbiamo usata per derivare una conclusione assurda. Non è affatto certo che il genere umano sia sul punto di estinguersi. Ma dove sta l'errore?
L'argomento dell'autobus funziona, mentre quello di Leslie pare folle a quasi tutti. Cercate di scoprirvi almeno una fallacia.

8 *L'evidenza totale.* Discutete brevemente l'argomento di Leslie dal punto di vista della probabilità logica e del requisito dell'evidenza totale.

PAROLE CHIAVE DEL RIPASSO

Probabilità a priori	Apprendere dall'esperienza
Probabilità a posteriori	Rapporto di verosimiglianza
Verosimiglianza	Regola di Jeffrey
Bayesiani	Bayesiani logicisti

16 La stabilità

In questo capitolo si illustrano alcune connessioni fra le regole di base della probabilità, la probabilità come frequenza e la stabilità statistica. Tali connessioni costituiscono il fondamento del ragionamento probabilistico. Il capitolo si conclude con l'enunciazione del teorema di Bernoulli, che è uno degli aspetti più importanti della probabilità.

Nei tre capitoli precedenti abbiamo assunto il punto di vista per cui la probabilità va intesa come credenza; in questo capitolo e nei tre successivi assumiamo invece il punto di vista frequentistico. Nei capitoli 13 e 14 abbiamo visto perché le probabilità intese come credenze devono soddisfare certe regole di base, nel 15 abbiamo mostrato come applicare questo risultato all'"apprendimento dall'esperienza"; nei capitoli 16-19 faremo qualcosa di simile dal punto di vista frequentistico.

IL PROGRAMMA

- Questo capitolo illustra alcune connessioni deduttive fra le regole della probabilità e la nostra nozione intuitiva di "frequenza stabile".
- Il capitolo 17 estende tali connessioni.
- Il capitolo 18 presenta un'idea centrale dell'inferenza induttiva di tipo frequentistico, quella di significatività.
- Il capitolo 19 presenta una seconda idea centrale dell'inferenza induttiva di tipo frequentistico, quella di fiducia; tale idea spiega, fra l'altro, il modo in cui oggi vengono riportati i sondaggi d'opinione, ma anche perché possiamo concepire l'uso delle statistiche come una forma di comportamento induttivo.

Tutti i risultati enunciati in questo capitolo sono dedotti dalle regole di base della probabilità; vengono solo enunciati, e non provati, perché il livello di difficoltà delle dimostrazioni supera quello del materiale presentato nel resto del volume.

CREDENZA E FREQUENZA A CONFRONTO

Le regole di base valgono per qualsiasi tipo di probabilità, ma le concezioni di quest'ultima come credenza o come frequenza mettono in risalto due tipi fondamentalmente diversi di conseguenze di tali regole.

Per la probabilità intesa come credenza l'idea più importante è la regola di Bayes, che può essere dedotta dalle regole di base che valgono per tutti i tipi di probabilità, ma da un punto di vista frequentistico la regola di Bayes non presenta particolare interesse, e l'idea più importante è quella secondo cui, man mano che il numero delle prove aumenta, le frequenze si stabilizzano. Il cosiddetto "teorema di Bernoulli" può aiutarci a capire questa idea. Esso può infatti essere dedotto, come la regola di Bayes, dalle regole di base della probabilità (anche se la sua dimostrazione è molto più difficile), e quindi vale in generale; ma se si intende la probabilità come credenza non presenta particolare interesse.

Questo teorema è solo il primo di un gran numero di importanti leggi dei grandi numeri e teoremi del limite centrale che sono stati dimostrati da generazioni di matematici, ma a mio parere è il risultato originario dello studioso svizzero a contenere l'idea centrale che si è rivelata tanto utile in tanti modi diversi. Potremmo dire che come il punto di vista secondo cui la probabilità va intesa come credenza si concentra in una sola proprietà logica fondamentale delle regole della probabilità, quella che corrisponde alla regola di Bayes, quello della frequenza si concentra in un'altra di tali proprietà, che corrisponde alle leggi dei grandi numeri. Ma entrambe le proprietà sono state ricchissime di conseguenze, sul piano sia matematico sia filosofico.

VINO NUOVO IN BOTTIGLIE VECCHIE

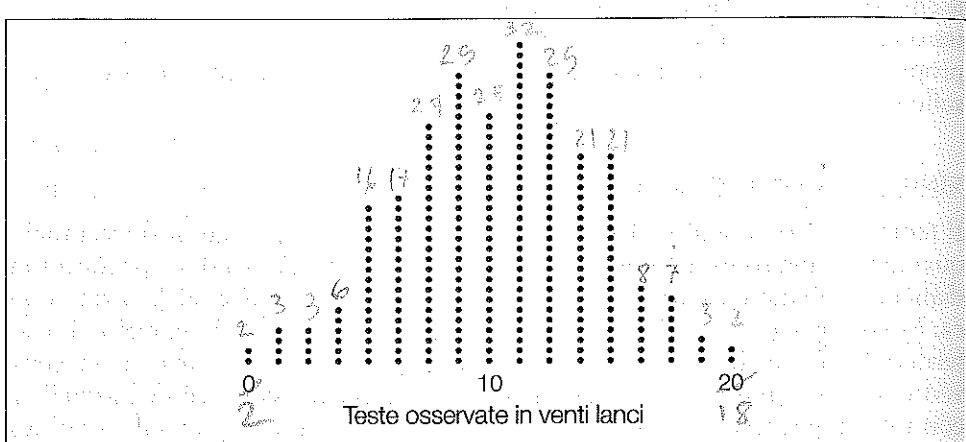
Il teorema di Bernoulli ha trecento anni, la regola di Bayes quasi duecentocinquanta, ma non sono superati: ciascuno è al centro di uno dei due approcci fondamentali all'inferenza induttiva che contano numerosi seguaci nella nostra epoca. Molti dogmatici sostengono che uno solo di questi approcci è corretto; noi affermiamo invece che sono utili entrambi, non solo nella pratica ma anche, come vedremo nei capitoli 21 e 22, per affrontare il problema filosofico dell'induzione con idee nuove. Hume diede a questo problema la sua formulazione definitiva, circa trent'anni dopo il risultato di Bernoulli e trenta prima che venisse pubblicato quello di Bayes, ma la sua analisi non è affatto superata e il "suo" problema continua a creare difficoltà ai filosofi.

LA STABILITÀ STATISTICA

Mettiamoci al lavoro. Una delle idee probabilistiche fondamentali è che nei tempi lunghi le frequenze relative si stabilizzano, ma abbiamo già precisato che l'idea di frequenza nei tempi lunghi è piuttosto oscura. Nei tempi lunghi saremo tutti morti, eppure siamo quasi tutti persuasi che, in certi tipi di scenario, più aumenta il numero delle prove effettuate, più la frequenza relativa si stabilizza. Dobbiamo dimostrare che è possibile spiegare questa idea mediante le regole di base della probabilità.

UN ESPERIMENTO

Prima, però, cerchiamo di appurare se con l'aumentare del numero delle prove otteniamo davvero la stabilità. Lanciate una moneta bilanciata venti volte e annotate il numero delle teste. Ci si aspètterebbe che le teste siano circa dieci, ma naturalmente in certi casi saranno pochissime, in altri molte più di dieci. Una volta ho dato una moneta a ogni studente di una classe molto numerosa, chiedendo a ciascuno di lanciarla per venti volte, e poi ho preparato un grafico che illustra 250 risultati. Nessuno degli studenti ha ottenuto 0, 1, 19 o 20 teste; solo ~~10~~ studenti 25 ne hanno ottenute esattamente 10, 32 ne hanno ottenute 11, 29 ne hanno ottenute 9 e altri 29 ne hanno ottenute 12.



Nonostante qualche irregolarità, se intorno a queste colonne tracciamo una curva regolare possiamo osservare che il tracciato assume grossomodo l'aspetto di una "campana"; questo è un esempio di "curva a campana sperimentale".

Ci aspettavamo che se ogni studente avesse fatto venti lanci il diagramma sarebbe apparso molto più piatto, e che quindi l'approssimazione alla "campana" sarebbe sfuggita all'occhio. E se avessimo usato solo monete bilanciate e tutti avessero fatto solo lanci regolari (cosa molto difficile), la curva avrebbe avuto un picco a 10.

LA MEDIA CAMPIONE

Il nostro esperimento consisteva di 250 miniesperimenti, ciascuno con un risultato k (il numero delle volte che la moneta ha dato testa). Possiamo chiederci quale sia il valore "medio" di k in queste 250 prove; per rispondere è necessario fare diverse addizioni, e chi vuole controllare appurerà che il valore medio è molto vicino a 10, per l'esattezza 9,76.

Ciascuno studente ha eseguito un miniesperimento lanciando la moneta venti volte, e possiamo pensare il nostro esperimento in grande, composto da 250 miniesperimenti, come un campione di un insieme molto più grande (anzi indefinitamente grande) di questi miniesperimenti. Il valore di k è detto "media campione" o "media sperimentale" del numero delle teste.

Se gli studenti avessero annotato k/n, la frequenza relativa (o proporzione) delle teste, la media campione sarebbe stata di 9,76/20 = 0,49, un valore molto vicino a 1/2.

LA DEVIAZIONE STANDARD DEL CAMPIONE

Ma la media campione è, appunto, solo una media, e quindi non dice tutto. Se ogni studente del corso avesse lanciato la sua moneta cinquanta volte ci si dovrebbe aspettare un addensamento molto più marcato intorno al valore medio, e quindi una curva a campana molto più "stretta": avremmo uno sparpagliamento meno marcato, una minore dispersione dei risultati.

Vorremmo misurare la dispersione intorno alla media X, un numero d che ci permettesse di dire, in buona sostanza, che la maggioranza dei risultati è compresa fra $X - d$ e $X + d$. C'è un modo naturale per misurare la dispersione: annotare la differenza fra la media X e ogni singolo risultato osservato X_j. Nel nostro caso ci sono 250 proporzioni k/n osservate e la media X = 0,49; ma se ci limitassimo a prendere le differenze $(X_j - X)$ e a sommarle alcune sarebbero positive e altre negative, quindi si eliderebbero. Così, il modo più usato per misurare la dispersione consiste nel sommare i quadrati di queste differenze, che sono tutti positivi, e poi estrarre la radice quadrata della somma. Otteniamo in tal modo la cosiddetta "deviazione standard del campione", abbreviato con DS:

$$\text{DS} = \sqrt{[\sum (X_j - X)^2/n]}.$$

Nel nostro esempio la deviazione standard del campione è di circa 2,9. Ma torniamo al grafico: non è difficile notare che 162 risultati, cioè il 65% dei risultati, sono compresi fra (10 – 3) e (10 + 3). Se, per assurdo, conducessimo un esperimento con diecimila studenti, ci aspetteremmo una deviazione standard minore. Vedremo subito che dovrebbe essere inferiore a 2,3.

LA DOMANDA STRANA NUMERO 1

La nostra idea è che in una lunga successione di prove molte frequenze tendono a stabilizzarsi, e anzi più la successione è lunga e maggiore è la stabilità, più grande è il numero delle prove e meno sono probabili grandi deviazioni dalla media. È l'idea alla base della domanda strana numero 1 (pag. 15), che abbiamo formulato così:

> Maschi e femmine nascono in numero pressappoco uguale. Nell'ospedale del capoluogo ogni settimana nascono moltissimi bambini; a Cornwall, una cittadina di provincia, c'è invece una piccola clinica dove in una settimana ne nascono pochi.
>
> Sono normali le settimane in cui sono femmine dal 45% al 55% dei neonati. Sono insolite le settimane in cui le femmine, oppure i maschi, superano il 55% dei neonati.
>
> Quale delle seguenti affermazioni è vera?
>
> ______ (a) Le settimane insolite sono altrettanto frequenti all'ospedale del capoluogo che a Cornwall.
>
> ______ (b) Le settimane insolite sono più frequenti all'ospedale del capoluogo che a Cornwall.
>
> ______ (c) Le settimane insolite sono più frequenti a Cornwall che all'ospedale del capoluogo.

Le grandi deviazioni (in questo caso le settimane insolite) sono più comuni in una piccola popolazione, come l'ospedale di Cornwall, che in una grande come l'ospedale del capoluogo; per questa ragione la risposta giusta è (c), la quale postula implicitamente che il sesso di un neonato dipenda dal caso (il che, d'altronde, sembra essere vero).

In quasi tutte le popolazioni, dal passato più lontano fino a oggi, la probabilità che un nato vivo sia maschio è pari a circa 0,51, quella che sia femmina è pari a 0,49; dunque se assumiamo come modello teorico delle nascite una moneta lievemente sbilanciata (c) è sicuramente la risposta giusta.

Naturalmente il modello potrebbe essere errato. Oggi è facile determinare il sesso di un futuro neonato sin dai primi mesi di gravidanza, e poiché presso certe popolazioni i figli maschi sono ritenuti molto più desiderabili delle femmine, molti feti destinati a diventare femmina vengono uccisi con l'aborto durante i primi tre mesi e la proporzione dei maschi rispetto al totale dei nati vivi aumenta.

LE PROVE BERNOULLIANE

In questo capitolo ci occuperemo di prove ripetute, indipendenti l'una dall'altra, con un esito possibile E che ha una probabilità costante p; spesso si parla, a questo proposito, di "prove bernoulliane", e qui abbiamo già presentato come prove bernoulliane i lanci di una moneta, i giri di una roulette e le estrazioni (con reimmissione) da un'urna.

Ora riassumiamo, senza usare formule, i teoremi fondamentali che discuteremo più avanti in questo capitolo. Chiediamoci:

> Se la probabilità dell'evento E è p, qual è la frequenza relativa con cui si presenta E in una lunga successione di prove bernoulliane?

Tutti i teoremi discussi in questo capitolo rispondono, in una forma o nell'altra, proprio a questo genere di domanda.

I LIMITI

Le regole di base della probabilità implicano che la frequenza relativa sarà quasi sempre vicina a p.

> In quasi tutte le lunghe successioni di prove bernoulliane la frequenza relativa degli E sarà vicina a p.

Inoltre:

> Più lunga è una successione di prove, maggiore è la probabilità che la frequenza relativa degli E sia vicina a p.

Ma possiamo formulare anche un'affermazione più forte: dato un qualsiasi *margine di errore* ε, per quanto piccolo, via via che il numero delle prove cresce illimitatamente, diventa sempre più probabile che la frequenza relativa E si approssimi a p con uno scarto non superiore a ε.

Useremo il nome "probabilità della precisione" per la probabilità che la frequenza relativa di E si approssimi a p entro un margine di errore ε:

> Per ogni margine di errore ε dato, all'aumentare del numero delle prove la probabilità della precisione si approssima a 1.

Questa è una conseguenza del teorema di Bernoulli e implica, chiaramente, che la probabilità di grandi deviazioni della frequenza relativa da p diminuisce quando il numero delle prove aumenta.

(Queste sono enunciazioni qualitative, ma per alcune probabilità della precisione useremo un'approssimazione quantitativa.)

Abbiamo formulato tutti i risultati che dobbiamo discutere nel corso del capitolo; ora è necessario riformularli in modo più preciso e fornire qualche esempio pratico del loro significato.

UN MODELLO URNA

Immaginiamo un'urna ideale e un meccanismo di campionatura con reimmissione. Le singole estrazioni sono indipendenti, e non c'è alcuno sbilanciamento: a ogni estrazione la probabilità di estrarre una pallina qualunque è identica a quella di estrarne qualsiasi altra, quindi le estrazioni sono prove bernoulliane.

L'urna contiene palline rosse e palline verdi in una proporzione p determinata, corrispondente alla probabilità di estrarne una verde.

ESTRAZIONI RIPETUTE

Consideriamo ora una prova composta di un tipo per noi nuovo:

> Facciamo n estrazioni indipendenti con reimmissione, cioè n prove bernoulliane, e annotiamo il numero k delle palline verdi estratte. Gli esiti possibili di una prova di questo tipo sono 0, 1, 2, ..., n.

Ora presentiamo alcuni semplici risultati sulle prove composte ripetute.

IL NUMERO PIÙ PROBABILE

Qual è il numero di palline verdi "più probabile da ottenere" estraendo n palline con reimmissione?

Naturalmente, se p è piccolo ce ne aspettiamo poche, se è grande ce ne aspettiamo molte; ma possiamo anche dare una risposta esatta a questa domanda, calcolando la probabilità di ottenere 0, 1, 2, ..., n palline verdi.

Fra i numeri da 0 a n, quali hanno la probabilità più elevata? Il numero con la probabilità più elevata è il numero più probabile di palline verdi. Ora,

> il numero più probabile di palline verdi è, approssimativamente, pn.

Ma oltre al numero più probabile di palline verdi su n prove, possiamo considerare anche la loro proporzione o frequenza relativa più probabile, e

> la frequenza relativa più probabile di palline verdi è, approssimativamente, p.

Perché "approssimativamente"? Per due ragioni. Innanzitutto pn non è in generale un intero: se $p = 1/2$ e $n = 5$, $pn = 2(1/2)$, che è un numero frazionario; ma quando facciamo cinque estrazioni da un'urna non estraiamo mai due palline verdi e mezzo. In secondo luogo, talvolta può accadere che ci siano due numeri più probabili di tutti gli altri, che però hanno entrambi la stessa probabilità: se $p = 1/2$ e $n = 5$, i due numeri più probabili (di palline verdi estratte) sono 2 e 3.

Negli esercizi 7 e 8 alla fine di questo capitolo mostriamo che il numero più probabile è in effetti l'intero (o la coppia di interi) "più vicino" a pn, quello compreso fra $pn - (1 - p)$ e $pn + p$ (a meno che questi ultimi siano essi stessi due interi).

LA FREQUENZA RELATIVA PIÙ PROBABILE

Sia k_0 il numero più probabile nel corso di n prove. In questo caso la frequenza relativa più probabile delle palline verdi è k_0/n, che è una frazione molto vicina a p; in effetti, gli esercizi alla fine del capitolo confermeranno che

$$(p - (1 - p)/n) \leq k_0/n \leq (p + p/n).$$

All'aumentare del numero n delle prove, p/n e $(1 - p)/n$ si approssimano entrambi a zero, quindi i due numeri rispettivamente sulla sinistra e sulla destra della formula si avvicinano molto (e molto rapidamente) a p, e se n è grande p è, sostanzialmente, la frequenza relativa più probabile.

> Teorema: Per un numero molto grande di prove la frequenza relativa più probabile k_0/n di palline verdi coincide sostanzialmente con p.

Quella appena illustrata è la prima connessione fra probabilità e frequenza nei tempi lunghi.

Confronto con il nostro esperimento. Nell'esperimento dei 250 studenti che lanciano ciascuno venti volte una moneta la frequenza relativa osservata più spesso non è 1/2: 32 studenti hanno ottenuto 11 teste in 20 lanci, quindi la frequenza relativa osservata più spesso non è stata 10/20 = 0,5, bensì 11/20 = 0,55.

NUMERO ATTESO DI PALLINE VERDI ESTRATTE IN n ESTRAZIONI

Possiamo anche chiederci quale sarà il numero medio delle palline verdi estratte in n estrazioni con reimmissione. Immaginiamo di fare moltissime volte n estrazioni, qualche volta senza estrarre nemmeno una pallina verde, qualche volta estraendone una sola, molto più spesso estraendone k_0 e così via; possiamo calcolare la media di tutti questi risultati e ottenere così il numero atteso di palline verdi in n estrazioni. È molto facile dimostrare, partendo dalle regole di base della probabilità, che il numero atteso è pn.

FREQUENZA RELATIVA ATTESA DELLE PALLINE VERDI IN n ESTRAZIONI

La frequenza relativa attesa non è che il numero atteso diviso per n. Vale dunque il

> Teorema: La frequenza relativa attesa delle palline verdi è p.

Questa è la seconda connessione fra probabilità e frequenza relativa nei tempi lunghi.

Confronto con il nostro esperimento. Nel nostro esperimento dei 250 studenti che lanciano ciascuno venti volte una moneta la frequenza relativa effettiva delle teste è molto vicina a 1/2, per l'esattezza 0,49.

CONVERGENZA E STABILITÀ

Posto che Pr(verde) = p, ci aspettiamo di ottenere circa pn palline verdi da una campionatura. Ma quanto deve essere grande un campione perché la proporzione k/n osservata sia prossima (molto probabilmente) a p? Più esattamente: dato un piccolo margine di errore ε, qual è la probabilità che su n estrazioni la frequenza relativa delle palline verdi non si allontani più di ε da p?

Jacques Bernoulli dimostrò che al crescere di n questa probabilità si approssima a 1.

MARGINE DI ERRORE E PROBABILITÀ DELLA PRECISIONE

Ora facciamo uso di una formula: dato un certo margine di errore ε, ci interessa la probabilità di ottenere in n prove un numero k di palline verdi tale che

$$p - \varepsilon \le k/n \le p + \varepsilon.$$

Questo è un "evento" insolitamente complicato in cui p ed n sono fissati arbitrariamente da chi pone la domanda, mentre k è il risultato dell'osservazione di n prove. Supponiamo per esempio che $p = 0{,}3$; se poniamo $n = 100$ estrazioni e prendiamo un margine di errore $\varepsilon = 0{,}01$, l'evento

$$p - \varepsilon \le k/n \le p + \varepsilon$$

consiste nell'ottenere un k tale che

$$(0{,}3 - 0{,}01) \le k/100 \le (0{,}3 + 0{,}01)$$

il che si verifica quando estraiamo k = 29, 30 o 31 palline verdi.

Qui ci interessa la probabilità di eventi di questo tipo, e per un margine di errore dato la chiameremo "probabilità della precisione". Per un margine di errore ε

> la probabilità della precisione è la probabilità che k non differisca da np per più di εn.

Esprimendoci in simboli, la probabilità della precisione diventa

$$\Pr[(np - \varepsilon n) \le k \le (np + \varepsilon n)].$$

Ma la probabilità della precisione è anche la probabilità che k/n non differisca da p per più di ε, ovvero, in simboli,

$$\Pr[(p - \varepsilon) \le k/n \le (p + \varepsilon)].$$

> TEOREMA:
> Al crescere del numero delle prove la probabilità della precisione si approssima a 1. Le frequenze relative tendono a convergere alle probabilità.

Quella appena illustrata è la terza connessione fra probabilità e frequenza nei tempi lunghi.

IL TEOREMA DI BERNOULLI

Di solito questo teorema viene enunciato in una forma leggermente diversa: al limite, k/n converge a p con una probabilità che si approssima sempre più a 1.

Per enunciare il risultato con precisione, vogliamo che la differenza fra p e la frequenza relativa sia piccola a piacere, cioè, come abbiamo già visto, che k/n non differisca da p per più di un piccolo margine di errore ε.

In secondo luogo, vogliamo che la probabilità della precisione, cioè la probabilità che k/n non differisca da p per più di ε, si approssimi a 1 quanto vogliamo; chiediamo cioè che la differenza x fra questa probabilità e 1 sia piccola quanto desideriamo, sempre che eseguiamo abbastanza prove. Mettendo insieme tutte queste idee giungiamo a enunciare il teorema di Bernoulli:

per ogni piccolo errore ε
e ogni piccola differenza x
esiste un numero di prove N
tale che
per ogni $n > N$
è maggiore di $1 - x$ la probabilità che
la proporzione k/n delle palline verdi non differisca da p per più di ε.

Vale la pena di "incorniciare" questo risultato:

> TEOREMA DI BERNOULLI:
> Per ogni piccolo errore ε
> e ogni piccola differenza x
> esiste un numero di prove N
> tale che
> per ogni $n > N$
> $\Pr[(p - \varepsilon) \leq k/n \leq (p + \varepsilon)] > (1 - x)$.

Questa è la quarta, nonché la più fondamentale, connessione fra probabilità e frequenza nei tempi lunghi. Ma resta ancora un altro nesso essenziale, che illustreremo nel prossimo capitolo.

JACQUES BERNOULLI

Il teorema di Bernoulli prende il nome dal matematico svizzero Jacques Bernoulli (1654-1705), zio di Daniel (*vedi* pag. 132), la cui principale opera sulla probabilità, l'*Ars coniectandi*, fu scritta poco dopo il 1690 ma pubblicata postuma solo nel 1714. L'*Ars coniectandi* è un libro importante per la matematica della probabilità, ma anche per la sua filosofia: nella quarta e ultima parte, infatti, l'autore illustra il nucleo centrale dell'approccio "fiduciario" all'apprendimento dall'esperienza che presenteremo nel capitolo 19.

ESERCIZI

1 *Clara l'affamata.* Clara, una neonata allattata artificialmente, fa sette poppate in ventiquattro ore. Questa la quantità di latte che ha ingerito in ciascuna poppata:

160 g, 100 g, 180 g, 100 g, 50 g, 150 g, 100 g.

(a) Qual è la quantità media di latte ingerita da Clara a ogni poppata? (b) Qual è la deviazione standard?

2 *Sam il malatino.* Sam è malato e mangia in modo irregolare. I genitori cercano di dargli sette poppate, ma lui mangia abbondantemente una sola volta al giorno; nelle varie poppate ingerisce:

80 g, 30 g, 60 g, 60 g, 200 g, 30 g, 100 g.

(a) Qual è la quantità media di latte ingerita da Sam a ogni poppata?
(b) Qual è la deviazione standard?
(c) Perché la media di Sam è più bassa di quella di Sara, mentre la deviazione standard è maggiore?

3 *Il reddito mediano.* Il reddito medio aritmetico di un gruppo è pari alla somma dei redditi dei membri del gruppo divisa per il numero dei membri. Si può dire che "sta nel mezzo", ma possiamo anche prendere in considerazione il reddito che divide la popolazione in due, con altrettante persone al di sopra e al di sotto, ovvero la cosiddetta "mediana", la quale viene fissata a metà strada fra il reddito più basso di coloro che stanno sopra la mediana e quello più alto di coloro che stanno sotto la mediana.

Una piccola impresa appena avviata impiega in tutto sei addetti, A, B, C, D, E ed F, compresa la donna delle pulizie, che lavora di notte e part-time. Qui di seguito elenchiamo i loro redditi:

A: 31 000$	B: 16 000$	C: 85 000$
D: 38 000$	E: 122 000$	F: 74 000$.

Quali sono (a) il reddito medio e (b) il reddito mediano?

4 *I redditi.* Nell'esercizio (3) la media e la mediana sono diverse. Dipende dal fatto che abbiamo preso in considerazione una popolazione molto piccola? No. Ecco il reddito medio e il reddito mediano del Canada, calcolati sulla base dei dati più recenti.

Tipo di famiglia	Reddito medio	Reddito mediano
Famiglia "standard"	57 146$	50 136$
Una persona sola	25 005	18 856
Tutte le famiglie	46 556	37 979

La differenza fra media e mediana vi suggerisce qualche considerazione circa la distribuzione dei redditi?

5 *La linea della povertà.* Esistono molte definizioni di "linea della povertà". Di solito negli Stati Uniti si parla, a questo proposito, di "soglia della povertà", e in Canada di "limite del basso reddito". Nelle grandi città del Canada la linea della povertà, per una famiglia di quattro persone, è fissata a 34 000 dollari canadesi (più o meno 23 000 dollari americani), cioè circa 6000 dollari (americani) più dell'equivalente americano, sempre per una famiglia di quattro persone.

Per i confronti internazionali si prendono di solito linee della povertà uguali (a) al 50% del reddito mediano della popolazione o (b) al 50% del reddito medio. Calcolate le linee della povertà canadesi per una famiglia "standard" usando entrambe le misure.

6 *Soluzioni rapide.* Il premier britannico ha annunciato che prima delle prossime elezioni il suo governo farà in modo che il 30% di coloro che oggi vivono sotto la linea della povertà riescano a superarla. (a) Quale potrebbe essere un modo poco costoso per farlo, supponendo che la linea della povertà sia definita in termini di reddito medio? (Non limitatevi a costringere un ugual numero di persone oggi sopra questa linea a scendere sotto!) (b) Il vostro metodo funzionerebbe se la linea della povertà fosse definita in termini di reddito mediano?

7 *I numeri più probabili.* È raro che il numero più probabile sia proprio pn, perché deve trattarsi di un numero intero (come 5 o 1097), mentre pn potrebbe essere una frazione o un numero decimale. Per esempio, se $p = 0{,}3$ e $n = 13$, $pn = 3{,}9$, che non è un numero intero. Inoltre, è possibile che i numeri (interi) più probabili siano due.

Considerate cinque estrazioni con reimmissione da un'urna con un numero uguale di palline verdi e rosse, per cui $p = 1/2$. Pr(0) è la probabilità che non venga estratta alcuna pallina verde, Pr(1) la probabilità che ne venga estratta una sola e così via.

(a) In quanti modi distinguibili e ugualmente probabili possiamo estrarre cinque palline dall'urna?

(b) Quali sono i valori di Pr(0) e Pr(5)?
(c) Quali sono i valori di Pr(1) e Pr(4)?
(d) Quali sono i valori di Pr(2) e Pr(3)?
(e) Quali sono i numeri più probabili?
(f) Sia k_0 il numero di palline verdi più probabile in n prove bernoulliane con probabilità p. Definite k_0 come quell'intero o quella coppia di interi tale che

$$np - (1 - p) \leq k_0 \leq np + p.$$

Confermate che nel nostro esempio con $p = 1/2$ e $n = 5$ i numeri più probabili sono 2 e 3.

8 *Il numero più probabile.* Qual è il numero più probabile quando $p = 0{,}3$ e $n = 13$?

9 *Il successo.*
(a) Trovate il numero più probabile dei successi (S) in undici prove bernoulliane con Pr(S) = 0,3.
(b) Qual è il numero atteso dei successi?

10 *La pioggia a Victoria.* In seguito a osservazioni svolte ogni anno a partire dal 1843 si è scoperto che la probabilità che a Victoria, nella British Columbia, il 1° luglio piova è di 4/17. Trovate il numero più probabile dei 1° luglio piovosi nei prossimi cinquant'anni.

11 *Gli acceleratori di particelle.* In un esperimento di fisica si stanno studiando particelle di un certo tipo. Nelle condizioni sperimentali A si osservano in media sessanta particelle al secondo, e per ognuna di queste c'è una probabilità di 0,7 che abbia una velocità maggiore di v; quando invece si passa alle condizioni B si osservano solo cinquanta particelle al secondo, che però tendenzialmente sono più veloci: la probabilità che la loro velocità superi v è 0,8.
(a) Qual è, rispettivamente, il numero più probabile delle particelle più veloci di v per secondo nelle condizioni A e B?
(b) I numeri medi delle particelle rapide (velocità > v) sono diversi da quelli più probabili? Si tratta di un fatto abituale o insolito?
(c) Se voleste studiare le particelle veloci preferireste lavorare nelle condizioni A o nelle condizioni B?

PAROLE CHIAVE DEL RIPASSO

Media campione
Deviazione standard
Prove bernoulliane
Numero più probabile
Teorema di Bernoulli
Numero atteso

17 Le approssimazioni normali

Abbiamo visto che le frequenze relative convergono alle probabilità teoriche. Con quale velocità? Quando possiamo cominciare a trattare una frequenza relativa osservata come una stima attendibile di una probabilità? Questo capitolo fornisce alcune risposte, in termini un po' più tecnici di quelli usati nel resto del libro; ma ai fini pratici vi basta sapere come usare i tre fatti normali che sono messi in evidenza nel testo.

CURVE A CAMPANA SPERIMENTALI

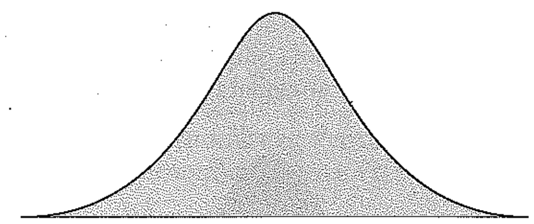
A pag. 248 abbiamo parlato del risultato di un esperimento di lancio di una moneta, e il grafico presentava, grossomodo, la forma di una campana. Fra le distribuzioni che è possibile osservare, ve ne sono molte che esibiscono questa proprietà.

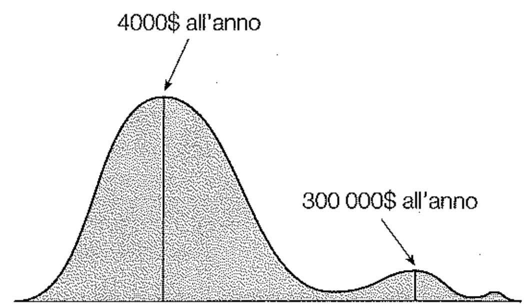
Esempio: i redditi. Oggi nei paesi industrializzati moderni ci si aspetta che la distribuzione dei redditi somigli più o meno alla curva 1 rappresentata alla pagina successiva, con poche persone incredibilmente ricche all'estremità destra del grafico; ma ai tempi del feudalesimo, per esempio, non esisteva una classe media, quindi ci si sarebbe dovuta aspettare una distribuzione del reddito "bimodale", cioè con due picchi, come la curva 2.

Esempio: gli errori. Non è possibile eseguire misurazioni perfettamente precise, e qualsiasi successione di misurazioni "esatte" della stessa quantità presenterà qualche variazione. Per questa ragione, spesso si prende una media dei risultati, che possiamo pensare come una media campione: un buon apparato di misura produrrà risultati che si addenseranno intorno alla media con una piccola deviazione standard, mentre un apparato scadente produrrà risultati che variano in modo abbastanza folle e quindi presentano una forte deviazione standard.

UNA CURVA A CAMPANA IDEALE

Esiste una formula matematica per un'intera famiglia di curve a campana perfettamente simmetriche. Qui non ci occuperemo di tale formula, limitandoci a enunciare alcune utili proprietà delle curve in questione, che sono dette "normali", perché si presentano empiricamente nel corso normale degli eventi, o "gaussiane", perché il matematico tedesco Carl Friedrich Gauss (1777-1855) vi dedicò studi approfonditi.


Curva 1



Curva 2

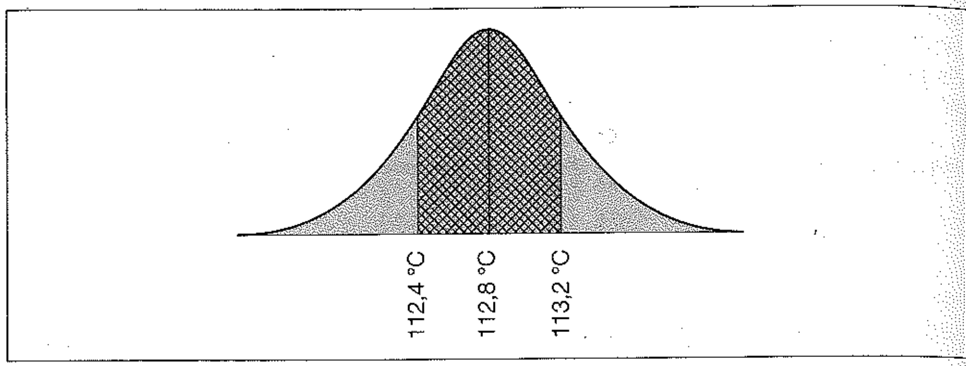
Ogni curva di questa famiglia è definita da due numeri. Il primo è la cosiddetta "media" (teorica), indicata da μ (la "m" minuscola dell'alfabeto greco), che corrisponde al valore della curva al picco; il secondo è la deviazione teorica standard σ (la "s" minuscola dell'alfabeto greco), che misura la larghezza della curva.

Quando si prende un campione di una popolazione con una distribuzione normale la media X del campione tende alla media teorica μ e la deviazione standard DS tende alla deviazione teorica standard σ.

SIGNIFICATO DELLA CURVA

Nel grafico dell'esperimento del lancio della moneta ci sono venti colonne, e la n-esima (da sinistra) rappresenta il numero degli studenti che in venti lanci hanno ottenuto n volte testa. La distribuzione normale è continua, quindi la matematica necessaria per descriverla è più difficile, ma il suo significato è molto semplice: invece che a una o più colonne dobbiamo pensare all'area compresa in un segmento della curva.

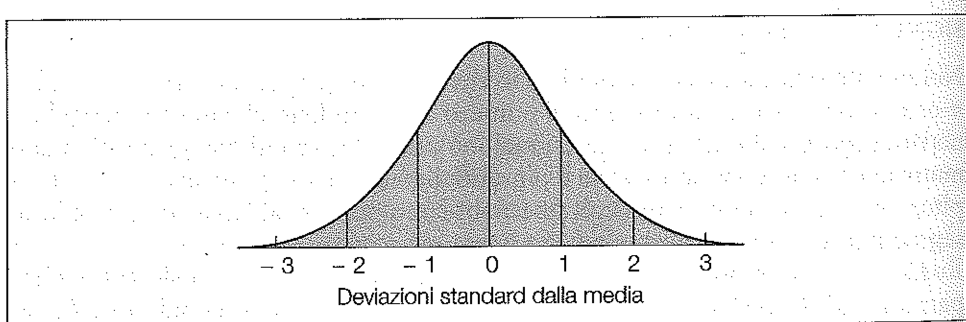
Esempio. Supponiamo di voler misurare il punto di fusione dello zolfo, che è pari a 112,8 °C. Se lavoriamo bene, le nostre misurazioni si distribuiranno in modo normale intorno a questo valore, che è la media della distribuzione. Possiamo chiederci, per esempio, se divergeranno al massimo di 0,4 °C dal vero punto di fusione. La percentuale delle misurazioni comprese fra 112,4 °C e 113,2 °C sarà rappresentata dal rapporto fra l'area tratteggiata della curva e la sua area totale.



APPLICAZIONE: FATTO NORMALE I

Supponiamo che una variabile (per esempio l'errore di una misurazione, o il reddito, o la statura) abbia una distribuzione normale, con media μ e deviazione standard σ. Se osserviamo un valore E di questa variabile,

> la probabilità che E differisca da pn non più di σ è di circa 0,68;
> la probabilità che E differisca da pn non più di 2σ è di circa 0,95;
> la probabilità che E differisca da pn non più di 3σ è di circa 0,99.



APPLICAZIONE: FATTO NORMALE II

Bernoulli dimostrò il suo teorema nell'ultimo decennio del Seicento. Pochi anni dopo (e prima che il risultato di Bernoulli venisse reso pubblico) Abraham de Moivre fece una scoperta ancora più importante: la distribuzione normale è un'eccellente approssimazione alle prove bernoulliane. Pensiamo alla probabilità di ottenere k volte l'evento E in n prove bernoulliane, dove Pr(E), la probabilità di E, è uguale a p; sia cioè

> $b(k; n, p)$ = la probabilità di ottenere k volte l'evento E in n prove quando $\Pr(E) = p$.

Questa è una cosiddetta "distribuzione binomiale", perché riguarda un binomio, cioè una coppia di termini, come "testa" e "croce". Quando n è grande lo sviluppo in dettaglio della funzione b è spaventosamente complesso, ma esiste una scorciatoia molto semplice: $b(k; n, p)$ è approssimata da una distribuzione normale a meno che p sia molto vicina a uno degli estremi 0 e 1 (nel qual caso avremo bisogno di un'approssimazione diversa).

> Una distribuzione binomiale $b(k; n, p)$ è approssimata da una distribuzione normale con $\mu = pn$ e $\sigma = \sqrt{[(1 - p)pn]}$.

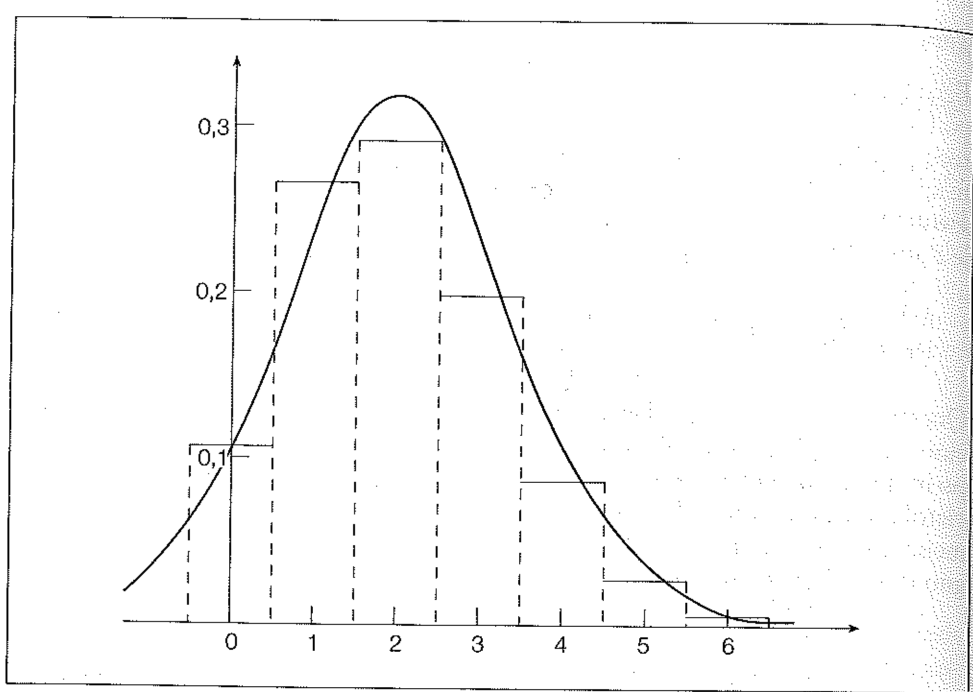
QUANTO È BUONA UN'APPROSSIMAZIONE?

Il grafico alla pagina successiva mostra quanto è buona l'approssimazione normale anche solo per dieci lanci di una moneta e con Pr(testa) molto bassa; va osservato che se Pr(testa) vale solo 1/5, su dieci lanci è improbabile ottenere più di cinque o sei volte testa.

L'altezza delle sbarre sull'asse delle Y rappresenta la probabilità di ottenere esattamente k teste. La curva è normale, con

$$\mu = np = 10 \times 0{,}2 = 2$$
$$\sigma = \sqrt{[(1 - p)pn]} = \sqrt{[(1 - 0{,}2)(0{,}2 \times 10)]} = 1{,}6.$$

È importante osservare con quanta precisione collima con le colonne.



QUANTO È RAPIDA LA CONVERGENZA DELLA FREQUENZA RELATIVA ALLA PROBABILITÀ?

Rispondiamo mettendo insieme i fatti normali I e II (il secondo ci dice che la distribuzione binomiale è approssimata da una distribuzione normale).

FATTO NORMALE III

Stiamo considerando una successione di prove bernoulliane con una probabilità p di testa, e il numero delle teste in n lanci è k; vogliamo conoscere la probabilità che k sia molto vicino a pn. Diamo le risposte, che seguono direttamente dai fatti normali I e II, per tre casi specifici:

> la probabilità che k diverga da pn per non più di σ è circa 0,68;
> la probabilità che k diverga da pn per non più di 2σ è circa 0,95;
> la probabilità che k diverga da pn per non più di 3σ è circa 0,99.

Dunque c'è, per esempio, una probabilità di 0,99 che k sia compresa fra $pn - 3\sigma$ e $pn + 3\sigma$. Questa è la quinta connessione (dal punto di vista pratico, la più interessante) fra probabilità e frequenza nei tempi lunghi.

LE LEGGI DEI GRANDI NUMERI

Abbiamo spiegato queste cinque connessioni tra probabilità e frequenza nei tempi lunghi per il caso più semplice, che per noi è anche il più familiare, perché fin qui abbiamo usato come esempi il lancio di una moneta o l'estrazione di una pallina da un'urna; tuttavia le situazioni pratiche con risposta SÌ/NO e strutturate come prove bernoulliane sono numerose. Per esempio, consideriamo i sondaggi d'opinione (almeno i più semplici). L'intervistatore chiede: «Lei è favorevole alla nuova legge sull'aborto?», e da un lato mette i "SÌ", dall'altro tutte le altre risposte. Analogamente, dividiamo i paraurti prodotti da una certa fabbrica in "OK" e "Sotto gli standard"; il modello dell'urna non è poi così lontano dalla vita reale.

Dopo Bernoulli e de Moivre sono stati dimostrati molti teoremi-limite ancora più profondi, sottili e complicati, chiamati spesso "leggi dei grandi numeri"; d'altronde lo stesso teorema di Bernoulli viene spesso denominato "legge debole dei grandi numeri", mentre per altri risultati, più generali e comprensivi, si usa l'espressione "teoremi del limite centrale". Ma per capire a un livello accettabile quello che è in gioco bastano il teorema di Bernoulli e i fatti normali enunciati sopra.

LE LAMPADINE

La VisioPerfect fabbrica lampadine di molti tipi diversi; il 96% è "a lunga durata" (oltre ottomila ore di luce), il 4% ha durata più breve. Queste lampadine vengono messe in vendita in confezioni da sei e la VisioPerfect spedisce ai suoi clienti almeno quattrocento confezioni, cioè 2400 lampadine. La distribuzione delle lampadine a lunga e a breve durata che ne risulta appare casuale, quindi possiamo usare il modello dell'insieme bernoulliano:

> In una spedizione di 2400 lampadine il numero atteso di quelle a lunga durata è $0,96 \times 2400 = 2304$.

Non accade spesso di ispezionare una partita e trovare esattamente 2304 lampadine a lunga durata, nonostante questo sia il numero più probabile, nonché il numero atteso. Ma quanto è verosimile che lo approssimiamo? La grandezza σ è

$$\sigma = \sqrt{[p(1-p)n]} = \sqrt{[0,96 \times 0,04 \times 2400]} = 9,6.$$

Perciò la probabilità che in una partita di 2400 lampadine il numero di quelle a lunga durata

sia compreso fra 2294 e 2314 è migliore di 2/3,
sia compreso fra 2284 e 2324 è molto migliore di 0,95,
sia compreso fra 2274 e 2334 è migliore di 0,99.

Nella prima riga abbiamo scritto "migliore" perché il margine di dieci intorno al numero atteso è maggiore di $\sigma = 9{,}6$; nella seconda riga abbiamo scritto "molto migliore" perché il margine di venti intorno al numero atteso è molto maggiore di $2\sigma = 19{,}2$; nella terza riga abbiamo scritto "migliore" per ragioni analoghe.

UN PRIGIONIERO ANNOIATO

James, un agente di cambio, è finito dietro le sbarre: lo hanno condannato a diversi anni di galera perché si è scoperto che derubava l'impresa presso la quale era impiegato. Non ha molto da fare e decide di lanciare diecimila volte una moneta, che per quanto ne sa lui è bilanciata.

Il numero atteso delle teste è 5000.
$\sigma = \sqrt{[p(1-p)n]} = \sqrt{[1/2 \times 1/2 \times 10\,000]} = 50.$

Perciò la probabilità che il numero delle teste osservate in diecimila lanci

sia compreso fra 4950 e 5050 è circa 2/3;
sia compreso fra 4900 e 5100 è circa 0,95;
sia compreso fra 4850 e 5150 è circa 0,99.

UNA PROBABILITÀ DI 1/2 È LO SCENARIO PEGGIORE

James ha lanciato la sua moneta diecimila volte, cioè ha fatto una serie di lanci più che quadrupla della serie di 2400 lampadine della VisioPerfect, eppure, per la stessa probabilità della precisione, i margini di errore sono molto più ampi.

Ciò significa che nei suoi lanci c'è una variabilità potenziale molto superiore a quella della produzione della fabbrica, ma supponiamo, per rendere il concetto ancora più chiaro, che la moneta sia stata lanciata esattamente 2400 volte. In tal caso il numero atteso delle teste è 1200 e

$$\sigma = \sqrt{[p(1-p)n]} = \sqrt{[1/2 \times 1/2 \times 2400]} = 24{,}5$$

contro il 9,6 della VisioPerfect; ciò significa, per esempio, che con una probabilità della precisione del 95% il margine di errore di colui che lancia una moneta è pari a 50, mentre per la VisioPerfect è pari a circa 20.

Possiamo giungere al medesimo risultato seguendo un'altra strada. In una partita di diecimila lampadine il numero atteso di quelle a lunga durata è 9600 e

$$\sigma = \sqrt{[p(1-p)n]} = \sqrt{[0{,}96 \times 0{,}04 \times 10\,000]} = 19{,}6$$

contro 50 per James. Ciò significa, per esempio, che con una probabilità della precisione del 99% il margine di errore è 150 per James e 60 circa per la VisioPerfect; cioè questo margine, data la stessa probabilità della precisione, è circa due volte e mezzo più grande per James che effettua lanci della sua moneta in prigione che per il produttore di lampadine.

La sola differenza numerica fra il caso delle lampadine e quello dell'agente di cambio è la probabilità.

Lampadine: $p = 0{,}96$ (probabilità che una lampadina sia di lunga durata).
Lanci di una moneta: $p = 0{,}5$ (probabilità che esca testa).

Il margine di errore è massimo quando $p = 1/2$, ma questo era prevedibile, perché il margine dipende da $\sigma = \sqrt{[p(1-p)n]}$, e per un n dato il massimo valore di $p(1-p)$ è $1/2 \times 1/2 = 1/4$.

Potete controllare questa affermazione lavorando con altri valori. Per esempio, se $p = 0{,}3$, $p(1-p) = 0{,}21 < 0{,}25$.

Nell'esercizio 6 vi si chiede di dimostrare che una probabilità di 1/2 è "lo scenario peggiore", nel senso che per probabili precisioni come 0,95 o 0,99 il margine di errore è massimo quando $p = 1/2$.

ABRAHAM DE MOIVRE

Abraham de Moivre (1667-1754), che stabilì l'approssimazione normale alle prove bernoulliane, apparteneva a una famiglia francese rifugiata in Inghilterra per motivi religiosi, e per molti anni si guadagnò da vivere facendo il precettore di allievi che gli venivano procurati da Isaac Newton.

La sua *Doctrine of Chances* rimase il più importante manuale di teoria della probabilità scritto in inglese per oltre cinquant'anni; dopo averla completata de Moivre scrisse anche il primo e fondamentale testo di teoria delle rendite a scadenza annuale.

ESERCIZI

1 *Bimodale.* Trovate qualche altro esempio di distribuzione bimodale.

2 *Normale.* Secondo voi, quali delle seguenti distribuzioni sono grossomodo normali, quali grossomodo bimodali e quali né una cosa né l'altra?
(a) La distribuzione delle stature in un gruppo etnicamente omogeneo di maschi.
(b) La distribuzione delle stature in un gruppo etnicamente omogeneo di individui.
(c) La distribuzione della pioggia nella città in cui vivete, settimana per settimana, durante l'anno.
(d) La quantità di latte prodotta da una mucca in una mandria di Jersey.

3 *Prodotti scadenti.* Supponiamo che solo 3/4 delle lampadine prodotte dalla VisioPerfect siano a lunga durata e che il processo produttivo possa essere rappresentato da un insieme bernoulliano come nel testo.
(a) Qual è il numero atteso di lampadine a lunga durata in una partita di 4800 pezzi?
(b) Qual è la probabilità che il numero delle lampadine a lunga durata di questa partita sia compreso fra 3540 e 3660?

4 *La doccia dopo l'allenamento.* Gli studenti maschi delle superiori usano le docce della loro scuola alla fine di un allenamento svolto all'interno dell'istituto? Una volta lo facevano quasi tutti, ma a quanto pare oggi lo fa solo la metà dei ragazzi mentre gli altri, be', puzzano.

Supponiamo di avere un campione casuale di 3136 studenti maschi delle scuole superiori che fanno sport a livello scolastico; ci aspetteremmo che circa 1568 facciano la doccia subito dopo ogni allenamento. Individuate un numero x tale che la probabilità di trovare più di $1568 + x$ ragazzi, o meno di $1568 - x$ ragazzi, che in questa situazione non fanno la doccia sia inferiore all'1%.

5 *Altre regioni.* Le abitudini possono variare da una regione all'altra. Supponiamo di prendere campioni di 3136 studenti in distretti scolastici molto lontani gli uni dagli altri, urbani e rurali, nel Nord, nel Sud e nel Centro del paese: trovate la x di cui alla domanda precedente in quattro distretti in cui la proporzione dei ragazzi che fanno la doccia dopo l'allenamento è:

(a) 0,3 (b) 0,7 (c) 0,1 (d) 0,9.

In ogni singolo caso x deve essere un numero intero tale che la probabilità che il campione devii dal numero atteso per più di x non superi l'1%.

6 *Il caso peggiore.* Dimostrate che con una probabilità della precisione pari allo 0,95 si ha il massimo margine di errore in un insieme bernoulliano quando $p = 1/2$.

PAROLE CHIAVE DEL RIPASSO

Distribuzione normale
Distribuzione binomiale
Approssimazione normale
Media teorica
Deviazione teorica standard