Modelli analitici descrittivi

Lezione 3: calcolo delle probabilità

1 Esercizio 3-1:

Si vuole studiare la distribuzione dei risultati degli scontri diretti fra due squadre per uno sport dove non sia previsto il pareggio (per es., la pallavolo) e il campionato preveda un girone di andata e uno di ritorno. Ogni scontro diretto, cioè, è definito come due partite, nelle quali le due squadre si scambiano il ruolo di squadra di casa e squadra in trasferta.

- Si definisca lo spazio campionario di questo studio, elencando gli esiti possibili.
- Si definiscano come sottoinsiemi di esiti gli eventi $A =$ "la squadra di casa vince la partita di andata" e $B =$ "la stessa squadra vince entrambe le partite".
- Si definiscano come sottoinsiemi di esiti gli eventi $C =$ "la squadra di casa vince la partita di andata o la stessa squadra vince entrambe le partite" e $D =$ "la squadra di casa vince la partita di andata e la stessa squadra vince entrambe le partite".

Soluzione

- Lo spazio campionario si può indicare con $\Omega = \{CC, CT, TC, TT\}$ dove, la prima lettera di ogni coppia indica la squadra che vince la partita di andata e la seconda la squadra che vince la partita di ritorno; inoltre, C indica la squadra di casa e T quella in trasferta.
- $A = \{CC, CT\}$ e $B = \{CT, TC\}$ (si noti che le due squadre si scambiano i ruoli passando da un girone all'altro).
- $C = \{CC, CT, TC\}$ e $B = \{CT\}$ (si noti che le due squadre si scambiano i ruoli passando da un girone all'altro).

2 Esercizio 3-2:

Si consideri un esperimento casuale avente come risultati possibili le lettere a, b, c, d ed e, con probabilità pari, rispettivamente, a 0.1, 0.2, 0.4, 0.2, 0.1. Si determini lo spazio campionario. Si descrivano poi i seguenti eventi determinandone la probabilità:

- $A =$ "lettera strettamente precedente la d";
- $B =$ "vocale";
- $C =$ "consonante diversa da b".

Soluzione Lo spazio campionario è $\Omega = \{a, b, c, d, e\}$. Per quanto riguarda gli eventi:

- $A = \{a, b, c\}$ ha probabilità $p(A) = 0.1 + 0.2 + 0.4 = 0.7$;
- $B = \{a, e\}$ ha probabilità $p(B) = 0.1 + 0.1 = 0.2$;
- $C = \{c, d\}$ ha probabilità $p(C) = 0.4 + 0.2 = 0.6$.

3 Esercizio 3-3:

Un'urna contiene tre palline: una rossa, una verde e una blu. Dall'urna si estrae prima una pallina e poi un'altra, senza rimettere la prima nell'urna. Qual è la probabilità che

- la prima pallina sia blu?
- la seconda pallina non sia verde?
- la prima pallina sia rossa oppure la seconda non sia blu?

Soluzione Se rappresentiamo l'esito di ogni singola estrazione con una lettera maiuscola (R per l'estrazione di una pallina rossa, V per quella di una pallina verde e B per quella di una pallina blu), e rappresentiamo l'esito dell'intero esperimento con due lettere associate nell'ordine alle due estrazioni, lo spazio campionario è $\Omega = \{RV, RB, BR, BV, VR, VB\}$. Sono sei esiti, che è ragionevole considerare equiprobabili, per cui ognuno ha probabilità $1/6$.

- la probabilità che la prima pallina sia blu è $2/6$ perché i casi favorevoli sono due (BR e BV) su sei;
- la probabilità che la seconda pallina non sia verde è $4/6$ perché i casi favorevoli sono quattro (RB, BR, VR e VB) su sei, oppure perché questo evento è esattamente complementare all'estrazione di una seconda pallina verde, evento la cui probabilità è $2/6$ (due casi favorevoli, RV e BV, su sei);
- la probabilità che la prima pallina sia rossa oppure la seconda non sia blu è $5/6$ perché i casi favorevoli sono cinque (RB, RV, BR, VR e VB) su sei, oppure per il teorema della somma: "prima pallina rossa" ($2/6$, identico per simmetria a "prima pallina blu") unito a "seconda pallina non blu" ($4/6$, identico per simmetria a "seconda pallina non verde"), meno l'intersezione ($1/6$, perché l'unico esito è RV) dunque $2/6 + 4/6 - 1/6 = 5/6$.

4 Esercizio 3-4:

Sono note le probabilità che un ordine a un negozio online richieda:

- un prodotto alimentare (evento A): $p(A) = 0.4$
- un prodotto in promozione (evento B, per "bonus"): $p(B) = 0.5$
- un prodotto in consegna rapida (evento R): $p(R) = 0.3$
- un prodotto che sia contemporaneamente A e B: $p(A \cap B) = 0.35$
- un prodotto che sia contemporaneamente A e R: $p(A \cap R) = 0.2$
- un prodotto che sia contemporaneamente B e R: $p(B \cap R) = 0.25$

- un prodotto che sia contemporaneamente A, B e R: $p\left(A \cap B \cap R\right) = 0.15$

Qual è la probabilità che il prossimo ordine...

1. non richieda un prodotto alimentare?
2. richieda un prodotto alimentare non in promozione?
3. richieda almeno una delle tre caratteristiche?
4. non richieda alcuna delle tre caratteristiche?

Soluzione

1. La probabilità di non richiedere un prodotto alimentare è
$$p\left(\bar{A}\right) = 1 - p\left(A\right) = 0.6$$
2. La probabilità di richiedere un prodotto alimentare non in promozione è
$$p\left(A \cap \bar{B}\right) = p\left(A \setminus \left(A \cap B\right)\right) = p\left(A\right) - p\left(A \cap B\right)) = 0.4 - 0.35 = 0.05$$
3. Per calcolare la probabilità di richiedere almeno una delle tre caratteristiche scomponiamo il problema va scomposto in due fasi, per usare la formula della somma:
$$p\left(A \cup B \cup C\right) = 0.4 + 0.5 + 0.3 - 0.35 - 0.2 - 0.25 + 0.15 = 0.55$$
4. Non richiedere alcuna delle tre caratteristiche è l'evento complementare del precedente:
$$p\left(\bar{A} \cap \bar{B} \cap \bar{C}\right) = 1 - p\left(\overline{A \cup B \cup C}\right) = 0.45$$

5 Esercizio 3-5:

Partendo dal file `Es2-3.txt`, che riporta i ritardi in arrivo registrati in una stazione, si valutino le seguenti probabilità:

- di arrivare con al massimo 30 minuti di ritardo,
- di arrivare con almeno 60 minuti di ritardo,
- di arrivare con un ritardo compreso fra 80 e 140 minuti,

applicando la definizione classica oppure i teoremi su eventi unione, intersezione, ecc...

Soluzione Il calcolo della frequenza cumulata è particolarmente utile per eventi che corrispondono a un intervallo $t \leq t_2$, $t > t_1$ o $t_1 < t \leq t_2$ di valori ammissibili per il ritardo t.

6 Esercizio 3-6:

Un magazzino riceve ogni sera le scorte per il giorno seguente. La domanda giornaliera è incognita, e anche la sua distribuzione è sconosciuta. Tuttavia, si conoscono la media $\mu = 800$ e lo scarto quadratico $\sigma = 15$. Supponendo di avere il magazzino vuoto, quanti prodotti conviene ordinare in modo da garantire che almeno il 90% dei clienti saranno soddisfatti il giorno dopo? È una stima precisa?

Se invece si conoscessero i quantili, come si potrebbe procedere?

Soluzione La formula di Čebyšëv stabilisce che, qualsiasi forma abbia la distribuzione

$$P\left[\mu - k\sigma \leq x \leq \mu + k\sigma\right] \geq \left(1 - \frac{1}{k^2}\right)$$

Se ordiniamo $\bar{x}$ unità di prodotto, vogliamo avere almeno il 90% di probabilità che la domanda sia $\leq \bar{x}$. Questo non è esattamente quel che dice la disuguaglianza, ma si possiamo arrivare per approssimazioni successive. Per prima cosa, possiamo scegliere $\bar{x} = \mu + k\sigma$, dove k è fissabile a piacere, per cui non abbiamo perso nessun grado di libertà nella scelta. Ne deriva che $k = (\bar{x} - \mu)/\sigma$. Ora la disuguaglianza dice che abbiamo almeno probabilità $\left(1 - 1/k^2\right)$ che la domanda sia compresa in $[\mu - k\sigma; \bar{x}]$. Ma se la domanda sta in quell'intervallo, è anche certamente $\leq \bar{x}$. Ora se il primo evento (domanda nell'intervallo limitato) è incluso nel secondo (domanda non superiore alla scorta), il secondo ha probabilitá superiore. Quindi la probabilità di non superare la scorta è $\geq \left(1 - 1/k^2\right)$. Ora basta sostituire k con $(\bar{x} - \mu)/\sigma$ e imporre che questo numero sia uguale al 90%:

$$\left(1 - \frac{\sigma^2}{(\bar{x} - \mu)^2}\right) = 0.9$$

da cui si può ricavare con semplici passaggi algebrici $\bar{x} = 848$, supponendo che la quantità del prodotto debba essere intera. Questa stima è approssimata per eccesso, cioè potrebbe bastare una scorta anche molto inferiore, dato che

1. la disuguaglianza vale per qualsiasi distribuzione, cioè anche nel caso pessimo: ma con un po' di fortuna si ottengono probabilità anche molto migliori;

2. la disuguaglianza garantisce che nel 90% dei casi la domanda sarà compresa fra $\mu - k\sigma = 752$ e $\mu + k\sigma = 848$, ma noi saremmo soddisfatti anche se fosse inferiore a 752, per cui i casi accettabili sono in effetti più del 90% (ma non possiamo migliorare la scelta se non sappiamo quanti).

Conoscendo i quantili, basterebbe utilizzare il nono decile, ovvero novantesimo percentile, che corrispondono esattamente al valore rispetto al quale la domanda è non superiore nel 90% dei casi.