Probabilità e Statistica (Informatica) 2021/22, Foglio II

1° novembre 2021 (aggiornato al 23 novembre)

Semplice definizione di sp. di probabilità con gli eventi P(B|A)=99% e P(B|A^C)=0.5

Esercizio 1. Per l'esempio del test clinico visto a lezione (incontro del 4 ottobre 2021), si trovi uno spazio di probabilità discreto che descriva il corrispondente esperimento aleatorio e si definiscano gli eventi d'interesse come sottoinsiemi dello spazio campionario scelto.

Si adotta come modello l'estrazione senza reinserimento (distr. ipergeometrica) e si risolve con i dati già presenti

Esercizio 2. Immaginiamo di voler stimare il numero di carpe (adulte) che vivono in un laghetto. Supponiamo che vi siano delle carpe. Indichiamo con X questo numero ignoto; possiamo interpretare X come una variabile aleatoria a valori in $\mathbb{N}$. Ora catturiamo r carpe, le marchiamo, e le liberiamo. Aspettiamo che i pesci marchiati si siano mescolati tra i loro compagni. Poi ne catturiamo n esemplari e contiamo il numero di carpe marchiate. Denotiamo con $A_{n,i}$ l'evento di aver trovato esattamente i carpe marchiate tra le n carpe prese alla seconda cattura. Per $r, n, N \in \mathbb{N}$ con $r \vee n \leq N$, $i \in \{0, \ldots, r \wedge n\}$, si calcoli la probabilità condizionale

$$c(N) \doteq \mathbf{P}\left(A_{n,i} \mid X = N\right).$$

Nel caso $r = 10$, $n = 20$, $i = 7$, si trovi numericamente il numero N che massimizzi la probabilità condizionale $c(N)$. Si dia infine una formula per stimare N in termini di r, n, i.

Verifica analitica considerando la binomiale e il principio di induzione.
Soluzione: https://ibb.co/2PKJQ03

Esercizio 3. Siano $N \in \mathbb{N}$, $q \in (0,1)$. Sia $p_{Bin(N,q)}$ la densità discreta della distribuzione binomiale di parametri N e q, quindi

$$p_{Bin(N,q)}(k) = \begin{cases} \binom{N}{k} \cdot q^k \cdot (1-q)^{N-k} & \text{se } k \in \{0, \ldots, N\}, \\ 0 & \text{altrimenti.} \end{cases}$$

Si mostri che, per $k \in \{0, \ldots, N-1\}$,

$$p_{Bin(N,q)}(k+1) = \frac{q}{1-q} \cdot \frac{N-k}{k+1} \cdot p_{Bin(N,q)}(k).$$

Si verifichi inoltre che, per $k \in \{0, \ldots, N\}$,

$$p_{Bin(N,q)}(k) = p_{Bin(N,1-q)}(N-k).$$

Dato l'esercizio precedente, si mette come parametro k e 1/2 nel pezzo precedente, calcolandolo numericamente quindi

Esercizio 4. In un ballottaggio tra due candidati, A e B, vota un milione di persone: $N \doteq 998000$ elettori sono completamente indecisi e votano a caso, con uguale preferenza per A e per B, mentre una minoranza di $M \doteq 2000$ persone sostiene il candidato A. Vogliamo trovare la probabilità che vinca A.

Come visto in aula, possiamo descrivere il comportamento elettorale delle N persone indecise tramite una variabile aleatoria S_N con distribuzione binomiale di parametri N e $1/2$, definita su un opportuno spazio discreto $(\Omega, \mathbf{P})$; S_N rappresenta il numero di voti che il candidato A riceve dal gruppo delle persone indecise. La probabilità che vinca A è allora data da

$$\mathbf{P}\left(S_N + M > \frac{N+M}{2}\right) = \mathbf{P}\left(S_N > \frac{N-M}{2}\right) = \sum_{k=\lfloor \frac{N-M}{2} \rfloor + 1}^{N} p_{Bin(N,1/2)}(k).$$

Si calcoli numericamente questa probabilità. [Suggerimento: usare l'Esercizio 3.]

Per la formula delle probabilità totali, possiamo approssimarle entrambe a Poisson e, essendo due distribuzioni, sarà di parametro 1/2 lambda, da cui consegue il risultato

Esercizio 5 (Esercizio 1.41 in CD)**.** "Si consideri il seguente modello di distribuzione dei figli nei nuclei familiari. La probabilità che un nucleo familiare scelto a caso abbia (esattamente) n figli, con $n \geq 0$, vale $e^{-\lambda}\frac{\lambda^n}{n!}$, dove λ è un parametro fissato, e ciascun figlio è maschio con probabilità $1/2$, indipendentemente da tutti gli altri. Consideriamo l'evento

$$A_k \doteq \text{"il nucleo familiare scelto (a caso) ha esattamente } k \text{ figli maschi"},$$

per $k \in \mathbb{N}_0$. Si mostri che la probabilità di A_k è uguale a $e^{-\lambda/2}\frac{(\lambda/2)^k}{k!}$."

Esercizio 6. Sia $q \in (1/2, 1]$. Per $n \in \mathbb{N}$, sia $(\Omega_n, \mathbf{P}_n)$ uno spazio di probabilità discreto con eventi $A_0^n, A_1^n, \ldots, A_n^n$ indipendenti e tali che $\mathbf{P}_n(A_0^n) = 1/2$, mentre $\mathbf{P}_n(A_i^n) = q$ per ogni $i \in \{1, \ldots, n\}$. Definiamo variabili aleatorie $X_i^n : \Omega_n \to \{-1, 1\}$, $i \in \{0, 1, \ldots, n\}$ ponendo

$$X_0^n(\omega) \doteq \begin{cases} 1 & \text{se } \omega \in A_0^n, \\ -1 & \text{altrimenti,} \end{cases} \qquad \omega \in \Omega_n,$$

e, per $i \in \{1, \ldots, n\}$, procedendo tramite ricorsione:

$$X_i^n(\omega) \doteq \begin{cases} X_{i-1}^n(\omega) & \text{se } \omega \in A_i^n, \\ -X_{i-1}^n(\omega) & \text{altrimenti,} \end{cases} \qquad \omega \in \Omega_n.$$

Poniamo infine $C_i^n \doteq \{\omega \in \Omega_n : X_i^n(\omega) = 1\}$, $i \in \{0, \ldots, n\}$.

a) Si calcoli $\mathbf{P}(C_i^n)$ per ogni $i \in \{0, \ldots, n\}$.

b) Si decida se gli eventi C_0^n e C_n^n sono indipendenti o meno.

c) Si calcoli il limite di $\mathbf{P}(C_n^n)$ per $n \to \infty$.

Questo esercizio, presente nella cartella Esercizi, è una verifica analitica abbastanza lunga da commentare.
Si guardi direttamente quella.

In poche parole, unendo le singole probabilità e i complementari di queste, la serie fa comunque venire fuori una probabilita data dall'insieme delle due, che risulterà 1/2 perché uniformi ed indipendenti.
Al limite sarà 1/2 e -1/2

Esercizio 7. Consideriamo l'esperimento del lancio di tre dadi regolari. Sia $\mathbf{P}$ la distribuzione uniforme sullo spazio campionario $\Omega \doteq \{1, \ldots, 6\}^3$. Per $i \in \{1, 2, 3\}$ poniamo

$$X_i(\omega) \doteq \omega_i, \quad \omega = (\omega_1, \omega_2, \omega_3) \in \Omega.$$

Di conseguenza, X_i può essere vista come il punteggio segnato dal dado i-esimo. Poniamo inoltre

$$Y_1 \doteq (X_1, X_2), \qquad Y_2 \doteq (X_2, X_3).$$

a) Si mostri che le variabili aleatorie X_1, X_2, X_3 sono indipendenti, e si trovino le loro distribuzioni.

Le variabili non sono indipendenti perché sono tra loro legate come insieme (si vede pure) e le loro distribuzioni sono, partendo dalla 1/omega uniforme date dai valori per le probabilità discrete, quindi X1=P(X=1), ecc.

b) Si mostri che Y_1, Y_2 *non* sono indipendenti.

c) Sia D la diagonale in $\{1, \ldots, 6\}^2$, cioè $D \doteq \{(i, i) : i \in \{1, \ldots, 6\}\}$. Si decida se gli eventi $\{Y_1 \in D\}$ e $\{Y_2 \in D\}$ sono indipendenti o meno.

Gli eventi non sono indipendenti come detto sopra e danno luogo alla stessa controimmagine (che umanamente significa che entrambi daranno la stessa probabilità quando sotto certi valori di P(X)

Esercizio 8. Siano X, Y, ξ variabili aleatorie definite sullo stesso spazio di probabilità discreto $(\Omega, \mathbf{P})$. Supponiamo che X e Y siano a valori nello stesso insieme non-vuoto E, ξ a valori in $\{0, 1\}$ con distribuzione di Bernoulli di parametro $p \in [0, 1]$, e che ξ e (X, Y) siano indipendenti. Poniamo

$$Z(\omega) \doteq \begin{cases} X(\omega) & \text{se } \xi(\omega) = 1, \\ Y(\omega) & \text{altrimenti.} \end{cases} \qquad \omega \in \Omega.$$

Essendo una bernoulliana e dovendo calcolare la densità discreta, essa sarà data da la probabilità p*P(x)+P(Y)*(1-p)

Si determini la distribuzione di Z in termini di $\mathbf{P}_X$, $\mathbf{P}_Y$, p.

Esercizio 9 (cf. Esempio 3.99 in [CD])**.** Immaginiamo di avere un'urna con N palline numerate da 1 a N. Estraiamo le palline a caso una dopo l'altra senza reinserimento, e indichiamo con X_i il numero dell'i-esima pallina estratta. Interpretiamo X_i, $i = 1, \ldots, N$, come variabili aleatorie a valori in $\{1, \ldots, N\}$ definite su un opportuno spazio di probabilità discreto $(\Omega, \mathbf{P})$.

Qui per determinare la distribuzione, essendo caso senza reinserimento, è data dalla serie delle misure per la loro probabilità, quindi come sempre per il discreto.

a) Si determini la distribuzione (congiunta) del vettore $(X_1, \ldots, X_N)$. Si determinino poi le distribuzioni (marginali) delle X_i.

b) Si mostri che, per $(i, j) \in \{1, \ldots, N\}^2$, X_i e X_j *non* sono indipendenti.

c) Sia σ una permutazione di $\{1, \ldots, N\}$, cioè $\sigma : \{1, \ldots, N\} \to \{1, \ldots, N\}$ è biiettiva. Si mostri che allora $(X_1, \ldots, X_N)$ e $(X_{\sigma(1)}, \ldots, X_{\sigma(N)})$ hanno la stessa distribuzione.

La b) e la c) sono legate, in quanto entrambi gli insiemi diminuiscono e vale sempre lo stesso risultato in termini di covarianza.

Nel caso di b), si può calcolare il valore medio come E[X]E[X-1], vedendo Y=X-1 che per la linearità del valor medio, darà sempre E(X)E(Y)+E(X).

Valendo quanto descritto, per (c) devono essere identicamente distribuite ed avere la stessa legge/distribuzione. Ciò si vede perché entrambi assumono esattamente gli stessi valori al variare di N e sarà la uniforme discreta (1/A)

Esercizio 10. Sia X una variabile aleatoria reale definita su $(\Omega, \mathbf{P})$ discreto. Si dimostrino le seguenti implicazioni:

Parliamo qui di variabili aleatorie reali e quindi si devono usare gli integrali. Sapendo che X[omega] <= C per P-quasi certamente (limite di X[omega] nell'insieme n di distribuzione tende ad 1), per la proprietà di normalizzazione l'integrale deve essere compreso tra due estremi, nel qual caso -C e C per poter valere l'implicazione.
Tante parole, ma si scrive in una riga, è dimostrata così.

a) Se esiste $C \in (0, \infty)$ tale che $|X(\omega)| \leq C$ per $\mathbf{P}$-quasi ogni $\omega \in \Omega$, allora X ammette valor medio finito e $-C \leq \mathbf{E}[X] \leq C$.

b) Se $X(\omega) \geq 0$ per $\mathbf{P}$-quasi ogni $\omega \in \Omega$ e $\mathbf{E}[X] = 0$, allora $X(\omega) = 0$ per $\mathbf{P}$-quasi ogni $\omega \in \Omega$.

Per lo stesso motivo di prima, ma qui se il valore atteso vale 0, necessariamente anche la legge/distribuzione di omega deve valere 0, lo dice la definizione di valor medio per var. aleat. reale. Quindi dimostrata così.

Esercizio 11. Siano X, Y variabili aleatorie discrete con densità discreta congiunta $p_{(X,Y)}$ data dalla seguente tabella:

x \ y	2	4	6	8
1	$\frac{0}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$
-3	$\frac{5}{64}$	$\frac{0}{64}$	$\frac{4}{64}$	$\frac{1}{64}$
5	$\frac{1}{64}$	$\frac{7}{64}$	$\frac{3}{64}$	$\frac{6}{64}$
-7	$\frac{2}{64}$	$\frac{9}{64}$	$\frac{8}{64}$	$\frac{11}{64}$

Si calcolino valor medio, varianza e covarianza di X e Y, e si decida se X e Y sono indipendenti o meno. Qui sono presenti tutti i dati, basta applicare le formule

Esercizio 12 (*KidsUniversity*)**.** Immaginiamo di avere una scacchiera da 20 righe (numerate da zero a 19 dal basso in alto) per 12 colonne (numerate da 1 a 12 da sinistra a destra). Disponiamo dodici pedine nelle caselle della riga zero. Ora lanciamo due dadi regolari da sei facce. Spostiamo di una casella in alto la pedina che si trova nella colonna il cui numero è uguale alla somma dei punteggi segnati dai due dadi. Continuiamo a lanciare i due dadi e a spostare pedine, muovendo sempre la pedina che si trova nella colonna del numero corrispondente alla somma dei punteggi segnati, finché la prima pedina non sarà arrivata alla riga 19. A questo punto il gioco si ferma. Si calcolino, analiticamente o solo numericamente, le probabilità dei seguenti eventi:

Si devono prendere tutte le probabilità marginali che diano 7 come congiunta (quindi sulle pedine 6,1 - 5,2 - 3,4 - 4,3 - 2,5 - 1,6 ed arrivando per prima, quindi tutta la prob. congiunta deve essere maggiore di 20). La stessa cosa vale per la b)

a) La pedina della colonna sette arriva per prima.

b) La pedina della colonna due arriva per prima.

c) La pedina della colonna sette arriva prima di quella della colonna otto.

Per il terzo si ha P(X = 7) > P (X=8), quindi anche qui è lo stesso ragionamento.

d) Il gioco ha durata di 100 mosse.

Per la quarta, si ha che il gioco dura 100 mosse se almeno 1 delle pedine fa 20 mosse per vincere e le altre 80 sono ricombinabili tra le 12 pedine.

Contatto: Markus Fischer (fischer@math.unipd.it)