La distribuzione binomiale

1. Che cos'è un numero casuale

Stiamo per lanciare un dado. Fermiamo la situazione un attimo prima che il dado cada e mostri la faccia superiore. Finché è in aria esso costituisce un numero causale, cioè un numero che può assumere certi valori con certe probabilità.
I valori che può assumere sono 1, 2, 3, 4, 5, 6. Se non abbiamo motivi per ritenere che il dado sia truccato, le corrispondenti probabilità sono tutte uguali. E poiché la somma di tutte le probabilità deve dare 1, ciascuna di esse è uguale a 1/6.
Possiamo sintetizzare un numero casuale mediante una tabella la cui prima riga descrive i valori che esso può assumere e la seconda riga descrive le corrispondenti probabilità.

$$\textbf{dado} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

Un numero casuale si chiama anche variabile aleatoria: "variabile" nel senso che può assumere diversi valori; "aleatoria" nel senso che ad ogni valore è assegnata una probabilità ("alea" sta per "dadi" in latino).
Nel caso del dado regolare, le probabilità sono tutte uguali; si dice che il corrispondente numero casuale ha una distribuzione uniforme.

Consideriamo un altro esempio. Gioco alla roulette e punto 100 euro sul rosso. La mia vincita, fino ad un attimo prima che la pallina si fermi, è un numero casuale, sintetizzato dalla seguente tabella.

$$\textbf{rosso} = \begin{bmatrix} -100 & 100 \\ \frac{19}{37} & \frac{18}{37} \end{bmatrix}$$

Infatti la roulette contempla 37 numeri, tutti i numeri naturali tra 0 e 36, e lo 0 non è né rosso né nero; qui la distribuzione delle probabilità non è uniforme, è maggiore la probabilità di perdere che quella di vincere.
Quando si gioca 100 sul rosso si perde 100 con probabilità circa del 51%

$$\frac{19}{37} \cong .513514$$

e si vince 100 con probabilità circa del 49%.

$$\frac{18}{37} \qquad \cong .486486$$

Questo significa che la strategia di giocare ripetutamente 100 sul rosso è in media perdente: in media si perde circa 19 volte su 37 e si vince circa 18 volte su 37. Se giochiamo n volte (supponiamo che n sia grande) ci aspettiamo di perdere circa

$$\frac{19}{37} \cdot n$$

volte, e di vincere circa

$$\frac{18}{37} \cdot n$$

volte. Vediamo il bilancio alla fine: abbiamo vinto

$$\frac{18}{37} \cdot n \cdot 100$$

euro e abbiamo perso

$$\frac{19}{37} \cdot n \cdot 100$$

euro. Il bilancio totale su n giocate è

$$\frac{18}{37} \cdot n \cdot 100 - \frac{19}{37} \cdot n \cdot 100$$

$$\frac{-100 \cdot n}{37}$$

cioè si perdono, in media, circa

$100/37$

2.70

euro a giocata. Se riprendiamo la tabella di questo numero casuale

$$\begin{bmatrix} -100 & 100 \\ \frac{19}{37} & \frac{18}{37} \end{bmatrix}$$

ci accorgiamo, in definitiva, che 100/37 è stato ottenuto addizionando i prodotti di ciascun valore assunto dalla variabile aleatoria per la corrispondente probabilità. Più precisamente:

$$-100 \cdot \left(\frac{19}{37}\right) + 100 \cdot \left(\frac{18}{37}\right)$$

-2.70

Questo numero si chiama <u>valore atteso</u> della variabile aleatoria.
Quindi il valore atteso della vincità è ... una perdita di 2.7 euro!

In generale il <u>valore atteso</u> di un numero casuale

$$\begin{bmatrix} x_{[1]} & x_{[2]} & \therefore & x_{[n]} \\ p_{[1]} & p_{[2]} & \therefore & p_{[n]} \end{bmatrix}$$

è

$$x_{[1]} \cdot p_{[1]} + x_{[2]} \cdot p_{[2]} + \therefore . + x_{[n]} \cdot p_{[n]}$$

oppure, utilizzando il simbolo di sommatoria:

$$\sum_{k=1}^{n} \left(x_{[k]} \cdot p_{[k]} \right)$$

Problema. Si lancia un dado; si vince 100 se esce 1, 2 o 3; si vince 300 se esce 4 o 5 e si perde 900 se esce 6. Conviene giocare?

La nostra risposta

Naturalmente la domanda è ambigua. La risposta potrebbe dipendere dalla propensione o avversione al rischio di ciascuno di noi. Possiamo precisarla in questo senso: qual è il valore atteso della vincita?
La vincita a questo gioco è la variabile aleatoria

$$\begin{bmatrix} 100 & 300 & -900 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{bmatrix}$$

e il valore atteso è 0:

$$100 \cdot \left(\frac{1}{2}\right) + 300 \cdot \left(\frac{1}{3}\right) - 900 \cdot \left(\frac{1}{6}\right)$$

$$0$$

Si tratta di un gioco equo: in media non si vince e non si perde.

2. Esprimenti con numeri casuali

Ogni sistema di calcolo possiede un generatore di numeri casuali. Ti-Interactive! possiede diversi comandi per la generazione di numeri casuali; il comando randint(a,b) genera un numero casuale intero compreso tra a e b.

randint(1, 6)

2.

Il comando rand(), senza argomenti, genera un numero casuale compreso tra 0 e 1.

rand()

.376756

Il comando rand(n) genera n numeri casuali compresi tra 0 e 1.

rand(10)

{.162659, .613134, .395658, .44127, .536837, .168852, .264546, .015933, .992974, .524638}

Con un generatore di numeri casuali è possibile svolgere esperimenti e simulazioni. In particolare è possibile simulare una variabile aleatoria.

Lavoriamo ora con Excel; in questo programma un numero casuale compreso tra 0 e 1 è generato dal comando
=casuale()
Il comando
=6*casuale()
genera allora un numero decimale compreso tra 0 e 6. Il comando
=tronca(6*casuale())
genera un numero casuale intero compreso tra 0 e 5. Infine il comando
=tronca(6*casuale())+1
genera un numero casuale intero compreso tra 1 e 6, cioè simula il lancio di un dado.
Il file Dadi simula 3000 lanci di un dado, di due dadi e di tre dadi (in tre fogli diversi). Nel caso di due e tre dadi siamo interessati alla somma delle facce. Provate ad aprire il file.
Per calcolare le frequenze assolute delle diverse uscite è stato usato il comando
=conta.se(intervallo;cella)
che calcola, nell'intervallo indicato, il numero di celle che contengono lo stesso

numero della cella indicata.
Ogni volta che si preme F9, oppure se si cambia un dato, l'intero foglio viene ricalcolato.
E' interessante osservare come si distribuiscono le frequenze (assolute e relative) delle diverse uscite; mentre con un dado le frequenze relative sono abbastanza uniformi e vicine a 1/6, con due dadi e tre dadi ciò non accade: la distribuzione sembra simmetrica rispetto al valore centrale e assume la tipica forma "a campana". Questo si spiega semplicemente: per esempio con due dadi l'uscita più probabile è il 7, perché si può ottenere con il maggior numero di combinazioni.

Problema. Quali sono le probabilità delle uscite 2, 3, ..., 12 nel lancio di due dadi? Scrivere la tabella della corrispondente variabile aleatoria. Suggerimento: osserva la seguente tabella.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

La nostra risposta

Confrontate le vostre risposte con le frequenze relative del file Dadi .

La variabile casuale che descrive la somma delle facce di due dadi regolari è la seguente.

$$\begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{bmatrix} \rightarrow due\_dadi$$

Approssimando le probabilità:

L1
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.

Si osservi che la lista L2 è stata costruita con il comando (6-abs(7-L1))/36.

Dovreste aver verificato, con il file Dadi , che queste probabilità non sono molto distanti dalle frequenze relative.

Problema. Qual è la probabilità di ottenere 13 lanciando 4 dadi? Suggerimento: non è un problema banale e per ora non è alla nostra portata. In prima approssimazione possiamo cercare una soluzione sperimentale: create un nuovo foglio nel file Dadi copiando il foglio "3 dadi" e aggiungendo una colonna per il quarto dado.

La nostra risposta

Dovreste aver trovato un risultato vicino a 11%; il risultato simbolico (che per ora non sapete calcolare) è

$$\frac{140}{6^4}$$

.108025

Vogliamo fare un altro esperimento: supponiamo di avere una moneta truccata, in cui T (TESTA) esce con probabilità p=70% e C (CROCE) esce con probabilità q=1−p=30%. Si dice che la moneta è "di trucco 0.7".
Se lanciamo questa moneta 2 volte, qual è la probabilità che esca T almeno una volta?
Il file di Excel Monete simula il lancio di 2 monete 4000 volte.
Nelle colonne B e C vengono generati due numeri casuali compresi tra 0 e 1.
Nelle colonne D e E, con i comandi
=SE(B2<0.7;1;0)
=SE(C2<0.7;1;0)
si genera 1 se esce T (cioè se il numero casuale è minore di 0.7), 0 se esce C.
Nella colonna F si addizionano i contenuti della colonna D e E; infine le colonne G, H e I elencano le possibili uscite di T su due lanci (0, 1, 2), le corrispondenti frequenze assolute e relative.
L'esperimento svolto ci fornisce una stima, mediante simulazione, della probabilità di avere almeno una T su due lanci con una moneta di trucco 0.7: addizionando infatti le frequenze relative a 1 e 2 uscite, risulta che la frequenza relativa del nostro esperimento è circa 91%, valore che non dovrebbe essere molto distante dalla probabilità teorica.

Problema. Qual è circa la probabilità di avere almeno 3 T su 5 lanci di una moneta truccata di trucco 0.7? Non fate alcun calcolo, date semplicemente una risposta "a naso". Sarà interessante, poi, confrontare la vostra risposta con la soluzione che daremo più avanti.

La nostra risposta

Ora modificate opportunamente il file Monete per simulare 4000 lanci di 5 monete di trucco 0.7. Addizionate le frequenze relative di 3, 4, 5 teste. Provate a confrontare la risposta "a naso" con il risultato sperimentale.

La nostra risposta

Dovreste aver trovato un risultato vicino a 83-84%; il risultato simbolico (che per ora non sapete calcolare) è 20923/25000 ≈ 83.7%.

3. La distribuzione binomiale

L'ultimo problema che abbiamo risolto è un trampolino di lancio per l'approccio ad una distribuzione di probabilità molto importante, la distribuzione binomiale.

Supponiamo che una moneta presenti T con probabilità t=0.7 e C con probabilità c=1−t=0.3. Lanciamola due volte. Gli eventi possibili sono illustrati dal grafo della figura contenuta nel file " due monete ". Puoi aprirla cliccando sul link.

Ogni volta che lanciamo la moneta, T e C hanno sempre la stessa probabilità di presentarsi (rispettivamente 0.7 e 0.3), indipendentemente da che cosa sia successo nei lanci precedenti; i lanci sono indipendenti e l'uscita di T al primo lancio non influenza il verificarsi di T anche al secondo lancio. Per questo motivo la probabilità che, per esempio, esca T sia al primo che al secondo lancio è uguale al prodotto delle probabilità: *t*t* = 0.7^2 = 0.49 = 49%. Possiamo immaginare che, se al primo lancio esce T il 70% delle volte, uscirà T anche al secondo lancio il 70% del 70% delle volte, cioè il 49%. Analogamente la probabilità che esca prima T e poi C è *tc* = 0.7*0.3 = 0.21 = 21%.
Sull'ultima riga riquadrata della figura contenuta nel file "due monete" abbiamo scritto i prodotti delle probabilità di ciascun ramo del grafo.
Poiché non siamo interessati all'*ordine* con cui escono T o C, possiamo concludere che la probabilità che lanciando due volte una moneta di trucco 0.7 esca:

due volte T è ....

una volta T e una volta C è ....

due volte C è ....

Possiamo riassumere le informazioni nella seguente tabella:

T	C	prob
2	0	
1	1	
0	2	

Domande di controllo
1) Quale dovrebbe essere la somma di tutte le probabilità? Tale condizione è rispettata?

La nostra risposta

2) Il ragionamento
"Gli eventi possibili sono tre:
- **due volte T**
- **una volta T e una volta C**
- **due volte C**

quindi ciascuno ha probabilità 1/3"
è giusto? Perché?

La nostra risposta

3) La probabilità di ottenere C <u>almeno</u> una volta è ....
4) La probabilità di ottenere T <u>almeno</u> una volta è ....

Proviamo a generalizzare quanto abbiamo visto, chiamando *t* la probabilità che esca TESTA e *c* la probabilità che esca CROCE ad ogni lancio. Il grafo degli eventi è rappresentato dalla figura2

Osservando il grafo della figura 2 provate a rispondere alle seguenti domande.

1) Se una moneta è truccata ed esce T con probabilità *t*=90%, qual è la probabilità che su due lanci esca T due volte? una volta? nessuna volta?

La nostra risposta

2) Qual è, se t=90%, la probabilità che esca T almeno una volta? E che esca C almeno una volta?

La nostra risposta

3) Quanto deve valere t se si vuole che esca T almeno una volta con probabilità 50%? Suggerimento: si tratta di risolvere l'equazione di 2° grado

$$t^2+2tc = 0.5$$

La nostra risposta

C'è una curiosa analogia tra l'ultima riga della figura 2 e il prodotto del binomio (t+c) per sé stesso:

$$(t+c)(t+c) = t^2 + tc + ct + c^2 = t^2 + 2tc + c^2$$

Che cosa c'è sotto?
In effetti lanciare due volte una moneta è in relazione con l'elevare al quadrato un binomio: in entrambi i casi ci sono quattro possibili esiti. Quando si moltiplica un binomio per un altro binomio occorre calcolare 4 prodotti: è come se ogni volta dovessimo scegliere una lettera nel primo binomio e una lettera nel secondo binomio. Ci sono quattro diverse possibilità: t^2, tc, ct, c^2. Se non consideriamo l'ordine dei lanci due di queste quattro possibilità (tc e ct) sono uguali.

Supponiamo ora di lanciare la moneta 3 volte. Osservate il grafo della figura3 ,

che conduce a 8 = 2^3 esiti differenti.
In analogia con il grafo della figura 1, completate la riga sotto il grafo della figura 3 con le probabilità dei vari eventi.

Compilate la tabella seguente.

T	C	prob
3	0	
2	1	
1	2	
0	3	

Domande di controllo
1) Quale dovrebbe essere la somma di tutte le probabilità?
Tale condizione è rispettata?

La nostra risposta

2) Qual è la probabilità che T esca almeno una volta? Almeno due volte?

La nostra risposta

3) Qual è la probabilità che C esca almeno una volta? Almeno due volte?

La nostra risposta

La figura 4 mostra il grafo del lancio di tre monete in generale, con probabilità *t* e *c*.

Completate l'ultima riga della fig. 4, in analogia con quanto fatto con la fig. 2.

Al lancio di due monete abbiamo associato il polinomio

$$t^2 + 2tc + t^2$$

che presenta tutti i termini di 2° grado che si possono costruire con le due lettere *t* e *c*: l'esponente di *t* decresce da 2 a 0, l'esponente di *c* cresce da 0 a 2. I coefficienti sono rispettivamente 1, 2, 1.
Nel lancio di tre monete si ottiene il polinomio

$$t^3 + 3t^2c + 3tc^2 + c^3 = (t+c)^3$$

I coefficienti, ordinati secondo le potenze decrescenti di *t* sono 1, 3, 3, 1.
Dovremmo riconoscere, nelle sequenze
1 2 1
1 3 3 1
le sequenze dei coefficienti binomiali per n=2 e n=3 rispettivamente; come sappiamo, i coefficienti dello sviluppo di (a+b)^n, oppure di (t+c)^n, come nel nostro caso, sono i coefficienti binomiali cb(n,k) per k da 0 a n.

Lanciamo una moneta equa 4 volte: T può uscire 4, 3, 2, 1, 0 volte. Vogliamo sapere quanto vale ciascuna corrispondente probabilità. Costruire il grafo inizia a diventare faticoso; però possiamo, in analogia con quanto fin qui, ipotizzare che le diverse probabilità siano descritte dal polinomio $(t+c)^4$.
Calcoliamo lo sviluppo di $(t+c)^4$.

$\mathrm{expand}\left((t+c)^4\right)$

$$t^4 + 4 \cdot c \cdot t^3 + 6 \cdot c^2 \cdot t^2 + 4 \cdot c^3 \cdot t + c^4$$

Osserviamo che i coefficienti sono ora 1, 4, 6, 4, 1: come sappiamo è la sequenza dei coefficienti binomiali per n=4. Ciascun addendo rappresenta una probabilità; più precisamente:
t^4 è la probabilità che esca 4 volte T (e 0 volte C)
4t^3*c è la probabilità che esca 3 volte T (e 1 volta C)
6t^2*c^2 è la probabilità che esca 2 volte T (e 2 volte C)
4t*c^3 è la probabilità che esca 1 volta T (e 3 volte C)
c^4 è la probabilità che esca 4 volte C (e 0 volte T)

Provate a completare la tabella seguente, per t=0.7 e c=0.3.

T	C	prob
4	0	
3	1	
2	2	
1	3	
0	4	

Verificate che la somma di tutte le probabilità sia 1.

Dunque: se conosciamo lo sviluppo del binomio (*t*+*c*)^*n*, possiamo calcolare le probabilità che, lanciando una moneta (con probabilità *t* e *c*) *n* volte, esca T *n* volte, *n*-1 volte, ..., 0 volte.

Per esempio: qual è la probabilità che, lanciando una moneta truccata 5 volte, esca T esattamente 3 volte? Sviluppiamo (*t*+*c*)^5.

$$\mathrm{expand}\left((t+c)^5\right)$$

$$t^5 + 5 \cdot c \cdot t^4 + 10 \cdot c^2 \cdot t^3 + 10 \cdot c^3 \cdot t^2 + 5 \cdot c^4 \cdot t + c^5$$

L'addendo che ci interessa è $10t^3c^2$: se *t*=0.7 allora la probabilità richiesta è circa il 31%:

$$10 \cdot 0.7^3 \cdot 0.3^2$$

$$.3087$$

Domande di controllo
1) Si lancia una moneta (con *t*=0.7) 6 volte. Qual è la probabilità che esca TESTA 4 volte?

La nostra risposta

2) Qual è la probabilità che esca TESTA almeno 4 volte?

La nostra risposta

3) Si lancia una moneta (con *t*=0.7) 10 volte. Qual è la probabilità che esca TESTA 9 volte?

La nostra risposta

4) Qual è la probabilità che esca TESTA almeno 9 volte?
La nostra risposta

5) Si lancia una moneta (con *t*=0.7) 10 volte. Qual è la probabilità che esca CROCE 9 volte?
La nostra risposta

6) Qual è la probabilità che esca CROCE almeno 9 volte?

La nostra risposta

Problema. Si lancia una moneta (con *t*=0.7) 10 volte, qual è la probabilità che CROCE esca almeno 5 volte? Non svolgete alcun calcolo, rispondete "a naso". Poi confronteremo la vostra risposta con quella corretta, verificando così la vostra "sensibilità probabilistica".

La nostra risposta

Ora svolgiamo i calcoli. Sviluppiamo (*t*+*c*)^10.

$\mathrm{expand}\left((t+c)^{10}\right)$

$$t^{10} + 10 \cdot c \cdot t^{9} + 45 \cdot c^{2} \cdot t^{8} + 120 \cdot c^{3} \cdot t^{7} + 210 \cdot c^{4} \cdot t^{6} + 252 \cdot c^{5} \cdot t^{5} + 210 \cdot c^{6} \cdot t^{4} + 120 \cdot c^{7} \cdot t^{3} + 45 \cdot c^{8} \cdot t^{2} + 10 \cdot c^{9} \cdot t + c^{10}$$

Ci interesano tutti gli addendi con *c* = 5, 6, 7, 8, 9, 10.

$c := 0.3 \qquad t := 0.7252 \cdot c^5 \cdot t^5 + 210 \cdot c^6 \cdot t^4 + 120 \cdot c^7 \cdot t^3 + 45 \cdot c^8 \cdot t^2 + 10 \cdot c^9 \cdot t + c^{10}$

$.3 \qquad .7 \qquad .150268$

Ecco la risposta. La probabilità che lanciando 10 volte una moneta di trucco 0.7 esca CROCE almeno 5 volte è circa il 15%. La risposta si accorda con la vostra congettura? Siete soddisfatti oppure no del vostro "naso" probabilistico?

Aggiungiamo ora ciò che già sappiamo: i coefficienti degli addendi dello sviluppo di (t+c)^n sono i coefficienti binomiali
cb(n,n), cb(n,n-1), ..., cb(n,2), cb(n,1), cb(n,0)

Abbiamo preso, come pretesto per l'introduzione della distribuzione binomiale, il problema della moneta truccata di trucco p.
Ma il problema che abbiamo risolto può essere formulato in modo più generale. Supponiamo che un certo esperimento, che si può ripetere quante volte si vuole, sempre nelle stesse condizioni, abbia soltanto due esiti, che chiameremo per comodità <u>successo</u> e <u>insuccesso</u> (nel caso della moneta T o C) e che la probabilità di un successo sia sempre la stessa, indichiamola con p (di conseguenza la probabilità di un insuccesso è q=1−p).
Se l'esperimento viene ripetuto n volte allora la probabilità di avere k successi ($0 \le k \le n$**) è**

$$\mathrm{ncr}(n, k) \cdot p^k \cdot (1 - p)^{(n-k)}$$

dove ncr(n,k) è il coefficiente binomiale n su k. La sequenza di queste probabilità si chiama <u>distribuzione binomiale di parametri n e p</u>, dove n è il numero di prove e p è la probabilità di successo.
Per esempio, la probabilità di successo in n=5 prove di probabilità p=0.7 è la seguente.

$\mathrm{seq}\left(\mathrm{ncr}(5, k) \cdot 0.7^k \cdot 0.3^{(5-k)}, k, 0, 5\right)$

$\{.002, .028, .132, .309, .36, .168\}$

il che significa che in 5 prove si hanno:

*** 0 successi con probabilità 0.2%**
*** 1 successo con probabilità 2.8%**
*** 2 successi con probabilità 13.2%**
*** 3 successi con probabilità 30.9%**
*** 4 successi con probabilità 36.0%**
*** 5 successi con probabilità 16.8%**

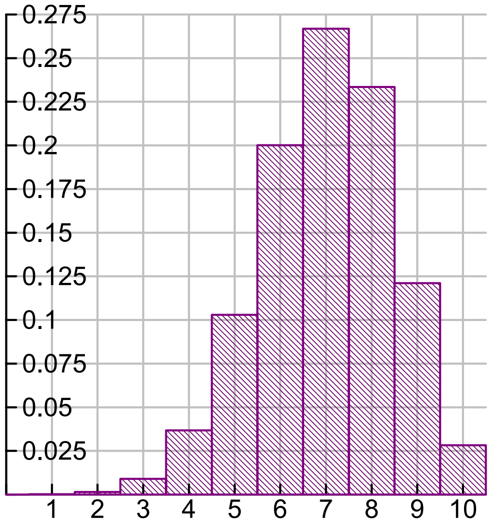
Ovviamente la somma delle probabilità è 1.

$$\sum_{k=0}^{5}\left(\mathrm{ncr}(5,k)\cdot 0.7^{k}\cdot 0.3^{(5-k)}\right)$$

1.

Con Ti-Interactive! è possibile creare la tabella della distribuzione binomiale (usare l'ambiente List) e il corripondente grafico mediante istogramma. Nella figura seguente è stata calcolata la variabile aleatoria con distribuzione binomiale di parametri n=10 e p=0.7.

L1	L2
0.	6E+009
1.	137781
2.	1E+015
3.	9E+015
4.	4E+015
5.	1E+015
6.	2E+015
7.	3E+015
8.	2E+015
9.	1E+015
10.	3E+015



Problema. Qual è la probabilità di avere almeno 7 successi in 10 prove di probabilità 0.7?

La nostra risposta

Ci interessano, della distribuzione binomiale di parametri n=10 e p=0.7, le probabilità corrispondenti a k=7, 8, 9, 10. Utilizzando il simbolo di sommatoria possiamo risolvere il problema nel seguente modo.

$$\sum_{k=7}^{10}\left(\mathrm{ncr}(10,k)\cdot 0.7^{k}\cdot 0.3^{(10-k)}\right)$$

.65

La probabilità di almeno 7 successi in 10 prove di probabilità 0.7 è il 65%.

Problema. Una fabbrica produce cuscinetti a sfera. La probabilità che un cuscinetto prodotto superi il controllo di qualità è 90%. Qual è la probabilità che su 10 pezzi prodotti ce ne siano più di due difettosi?

La nostra risposta

Nel problema precedente si chiede in sostanza quale sia la probabilità di avere 8, 9 o 10 successi in 10 prove di probabilità 0.9. La risposta è

$$\sum_{k=8}^{10} \left(\mathrm{ncr}(10, k) \cdot 0.9^{k} \cdot 0.1^{(10-k)}\right)$$

.93

Dunque la probabilità che il numero di successi sia minore di 8 (cioè ci siano più di 2 pezzi difettosi) è circa del 7%.

Problema. Si lancia una moneta regolare 100 volte. Qual è la probabilità che esca TESTA un numero di volte compreso tra 45 e 55?

La nostra risposta

Si tratta di addizionare i valori della distribuzione binomiale di parametri n=100 e p=0.5 per k da 45 a 55. La risposta è circa il 73%.

$$\sum_{k=45}^{55} \left(\mathrm{ncr}(100, k) \cdot 0.5^{k} \cdot 0.5^{(100-k)}\right)$$

.729

Un problema un po' più difficile: nel problema precedente, quanto deve valere p

affinché la probabilità che esca TESTA dalle 45 alle 55 volte su 100 lanci sia del 20%?
Si vuole in definitiva risolvere l'equazione in p:

$$\sum_{k=45}^{55} \left(\mathrm{ncr}(100, k) \cdot p^k \cdot (1 - p)^{(100 - k)}\right) = 0.2$$

Si tratta di un'equazione assai complicata; risolviamola, come è giusto che sia, per tentativi. Definiamo la funzione f(p) che fornisce, in funzione di p, la somma delle probabilità per un numero di successi compreso tra 45 e 55.

$$\sum_{k=45}^{55} \left(\mathrm{ncr}(100, k) \cdot p^k \cdot (1 - p)^{(100 - k)}\right) \to f(p)$$

"Done"

Ora costruiamo la sequenza di valori di f(p) per p compreso tra 30% e 50%, con passo 1%.

$\mathrm{seq}(\{p, f(p)\}, p, 0.3, 0.5, 0.01)$

.3	.001
.31	.002
.32	.004
.33	.008
.34	.015
.35	.025
.36	.04
.37	.061
.38	.091
.39	.13
.4	.178
.41	.236
.42	.302
.43	.373
.44	.448
.45	.521

La tabella mostra che la probabilità richiesta è compresa tra 40% e 41%. Esploriamo questo intervallo con passo 0.1%.

$$\mathrm{seq}(\{p, f(p)\}, p, 0.4, 0.41, 0.001)$$

$$\begin{bmatrix} .4 & .178 \\ .401 & .183 \\ .402 & .189 \\ .403 & .194 \\ .404 & .2 \\ .405 & .206 \\ .406 & .212 \\ .407 & .218 \\ .408 & .224 \\ .409 & .23 \end{bmatrix}$$

Si osserva che il valore di p richiesto è circa 40.4%.

Un problema analogo al precedente: se una moneta è regolare (p=0.5), lanciandola 100 volte uscirà TESTA un numero di volte non troppo lontano da 50, diciamo un numero di volte compreso tra 50–h e 50+h, con h abbastanza piccolo. Per quale valore di h la probabilità di aver un numero di TESTA compreso tra 50-h e 50+h è almeno del 90%? del 95%? del 99?
Questa volta l'incognita compare negli estremi del simbolo di sommatoria. Definiamo la funzione g(h) che fornisce la probabilità di avere, su 100 lanci, un numero di teste compreso tra 50–h e 50+h.

$$\sum_{k=(50-h)}^{50+h} \left(\mathrm{ncr}(100, k) \cdot 0.5^{100}\right) \rightarrow g(h)$$

"Done"

Dobbiamo ora risolvere l'equazione g(h)=0.9. A diferenza del problema precedente, l'incognita h è un numero naturale. Tabuliamo i valori di g(h), per h compreso tra 5 e 15.

$\text{seq}(\{h, g(h)\}, h, 5, 15)$

$$\begin{bmatrix} 5 & .729 \\ 6 & .807 \\ 7 & .867 \\ 8 & .911 \\ 9 & .943 \\ 10 & .965 \\ 11 & .979 \\ 12 & .988 \\ 13 & .993 \\ 14 & .996 \\ 15 & .998 \end{bmatrix}$$

Come si vede la probabilità è almeno del 90% per h uguale a 8: la probabilità che esca TESTA un numero di volte compreso tra 42 e 58 è del 90%. La probabilità è almeno del 95% per h=10, cioè per un numero di teste comprese tra 40 e 60. La probabilità è almeno del 99% per h=13, cioè per un numero di teste compreso tra 37 e 63. Se si allarga l'intervallo tale probabilità, come è giusto aspettarsi, cresce.