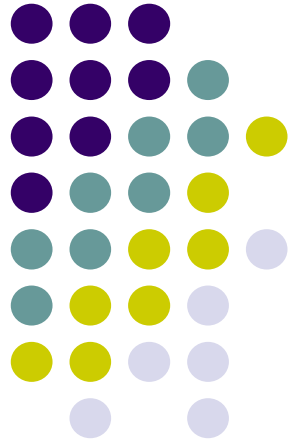
Levine, Krehbiel, Berenson

Statistica

Casa editrice: Pearson

Capitolo 5

Variabili aleatorie discrete notevoli



Insegnamento: Statistica
Corso di Laurea Triennale in Economia

Dipartimento di Economia e Management, Università di Ferrara
Docenti: Prof. Stefano Bonnini, Dott.ssa Valentina Mini

Argomenti

- La distribuzione di probabilità di una variabile aleatoria discreta
  - valore atteso
    di una variabile aleatoria discreta
  - varianza e scarto quadratico medio
    di una variabile aleatoria discreta
- La distribuzione binomiale
- La distribuzione di Poisson

Distribuzioni di probabilità

- Una distribuzione di probabilità è un modello matematico, uno schema di riferimento, che ha caratteristiche note e che può essere utilizzato per rispondere a delle domande derivate da problemi reali
- Ad esempio un sistema informativo aziendale per la gestione degli ordini ha il compito di individuare errori o informazioni incomplete. Se la probabilità di che un generico ordine non sia corretto è 0.1 qual è la probabilità che in una giornata in cui vengono sottoposti quattro ordini nessuno di questi venga segnalato come errato?
- Il processo informativo potrebbe essere approssimato da un modello descritto da una distribuzione di probabilità

La distribuzione di probabilità di una variabile aleatoria discreta

- Una *variabile aleatoria quantitativa* è stata definita come un fenomeno le cui modalità sono espresse da numeri (numero di ordini sottomessi o tempo impiegato per controllare un ordine)
- Le variabili aleatorie quantitative sono state classificate in *variabili discrete* e *variabili continue*, a seconda che scaturiscano da un processo di conteggio oppure da una misurazione su scala continua
- Con **variabile aleatoria** si intende che non è possibile conoscere a priori quale sarà la modalità della variabile che osserveremo
- Per caratterizzare questo tipo di variabile, dobbiamo introdurre la distribuzione di probabilità di una variabile aleatoria discreta

La distribuzione di probabilità di una variabile aleatoria discreta

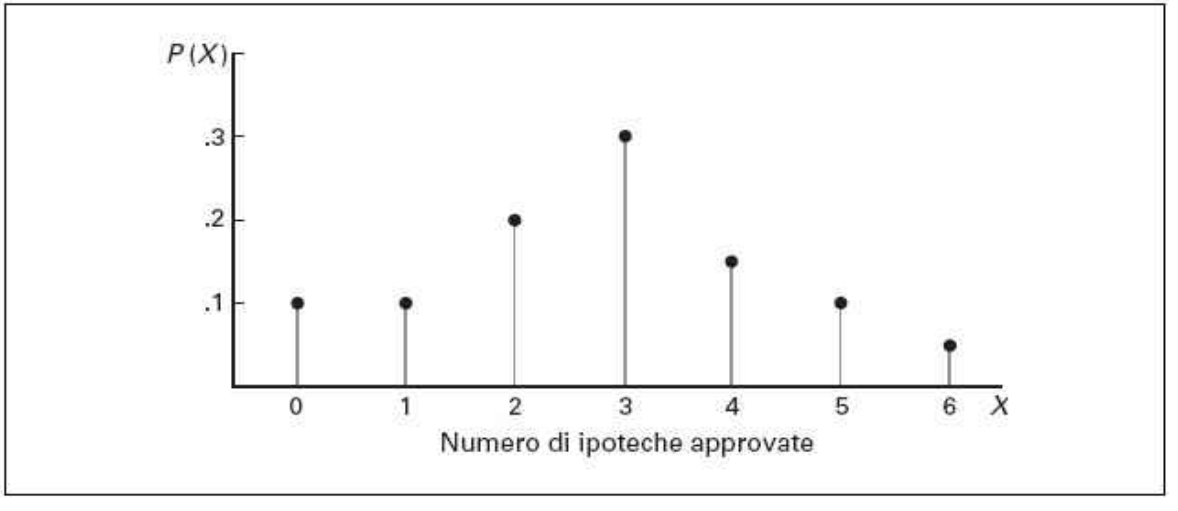
La **distribuzione di probabilità di una variabile aleatoria discreta** è rappresentata dall'elenco delle modalità che la variabile può assumere (modalità incompatibili e mutuamente esclusive), a ciascuna delle quali è associata la relativa probabilità

Consideriamo ad esempio la distribuzione del numero di ipoteche approvate settimanalmente da parte di una banca:

Numero di ipoteche approvate	Probabilità
0	0.10
1	0.10
2	0.20
3	0.30
4	0.15
5	0.10
6	0.05

La distribuzione di probabilità di una variabile aleatoria discreta

Naturalmente, poiché le modalità elencate sono mutuamente esclusive e collettivamente esaustive, le probabilità sommano a 1. La distribuzione può essere rappresentata anche graficamente



Un modo per sintetizzare una distribuzione di probabilità discreta consiste nel calcolarne le principali misure di sintesi: il valore atteso e lo scarto quadratico medio

La distribuzione di probabilità di una variabile aleatoria discreta

La media μ di una distribuzione di probabilità si dice **valore atteso** della variabile aleatoria

Il **valore atteso di una variabile aleatoria discreta** è una media ponderata delle modalità assunte dalla variabile, dove i coefficienti di ponderazione sono rappresentati dalle probabilità associate a ciascuna modalità

Valore atteso di una variabile aleatoria discreta

$$\mu = E(X) = \sum_{i=1}^{N} X_i P(X_i) \qquad \textbf{(5.1)}$$

dove X_i = *i*-esima modalità della variabile aleatoria X

$P(X_i)$ = probabilità associata alla modalità X_i

La distribuzione di probabilità di una variabile aleatoria discreta

Il valore atteso del numero di ipoteche approvate settimanalmente della banca si calcola come

	Probabilità	
Numero di ipoteche approvate in una settimana	$P(X_i)$	$X_i\,P(X_i)$
0	0.10	(0)(0.10) = 0.0
1	0.10	(1)(0.10) = 0.1
2	0.20	(2)(0.20) = 0.4
3	0.30	(3)(0.30) = 0.9
4	0.15	(4)(0.15) = 0.6
5	0.10	(5)(0.10) = 0.5
6	0.05	(6)(0.05) = 0.3
	1.00	$\mu = E(X) = 2.8$

Si noti che al valore atteso della variabile “numero di ipoteche approvate in una settimana” non può essere attribuito un significato letterale, visto che il numero effettivo di ipoteche approvate deve essere un valore intero

La distribuzione di probabilità di una variabile aleatoria discreta

La **varianza σ^2 di una variabile aleatoria discreta** è definita come la media *ponderata* dei quadrati delle differenze tra ciascuna modalità e il valore atteso della variabile, dove i coefficienti di ponderazione sono rappresentati dalle probabilità associate a ciascuna modalità

Varianza di una variabile aleatoria discreta

$$\sigma^2 = \sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i) \qquad \textbf{(5.2)}$$

dove X_i = *i*-esima modalità della variabile aleatoria X

$P(X_i)$ = probabilità associata alla modalità X_i

Lo **scarto quadratico medio σ di una variabile aleatoria discreta** è dato dalla radice quadrata della varianza: $\sigma=\sqrt{\sigma^2}$

La distribuzione di probabilità di una variabile aleatoria discreta

La varianza e lo scarto quadratico medio del numero di ipoteche approvate settimanalmente della banca si calcolano come

	Probabilità		
Numero di ipoteche approvate in una settimana	$P(X_i)$	$X_iP(X_i)$	$[X_i - E(X)]^2P(Xi)$
0	0.10	$(0)(0.10) = 0.0$	$(0 - 2.8)^2(0.10) = 0.784$
1	0.10	$(1)(0.10) = 0.1$	$(1 - 2.8)^2(0.10) = 0.324$
2	0.20	$(2)(0.20) = 0.4$	$(2 - 2.8)^2(0.20) = 0.128$
3	0.30	$(3)(0.30) = 0.9$	$(3 - 2.8)^2(0.30) = 0.012$
4	0.15	$(4)(0.15) = 0.6$	$(4 - 2.8)^2(0.15) = 0.216$
5	0.10	$(5)(0.10) = 0.5$	$(5 - 2.8)^2(0.10) = 0.484$
6	0.05	$(6)(0.05) = 0.3$	$(6 - 2.8)^2(0.05) = 0.512$

$$\sigma^2 = \sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i) = 2.46$$

$$\sigma = 1.57$$

La distribuzione binomiale

Un modello probabilistico è un'espressione matematica che rappresenta la distribuzione di probabilità di una variabile di interesse.

Uno dei modelli probabilistici più utilizzati è la **distribuzione binomiale** caratterizzata da quattro essenziali proprietà:

- Si considera un numero prefissato di *n* osservazioni
- Ciascuna osservazione può essere classificata in due categorie incompatibili ed esaustive, chiamate per convenzione *successo* e *insuccesso*
- La probabilità di ottenere un successo, *p*, è costante per ogni osservazione, così come la probabilità che si verifichi un insuccesso, (1 – *p*).
- Il risultato di un'osservazione, successo o insuccesso, è indipendente dal risultato di qualsiasi altra. …

La distribuzione binomiale

- … Per assicurare l'indipendenza, le osservazioni possono essere ottenute con due diversi metodi di campionamento: un *campionamento da una popolazione infinita senza reimmissione* oppure un *campionamento da una popolazione finita con reimmissione*

Con riferimento all'esempio del sistema informativo aziendale per la gestione degli ordini, supponiamo di osservare il seguente risultato in un campione di 4 ordini

Primo ordine	Secondo ordine	Terzo ordine	Quarto ordine
Segnalato	Segnalato	Non segnalato	Segnalato

Qual è la probabilità di ottenere questa particolare sequenza di successi e insuccessi in un campione di quattro ordini?

Primo ordine	Secondo ordine	Terzo ordine	Quarto ordine
$p = 0.10$	$p = 0.10$	$(1 - p) = (1 - 0.10) = 0.9$	$p = 0.10$

La distribuzione binomiale

Poiché le osservazioni sono indipendenti, la probabilità di ottenere questa particolare sequenza è pari a:

$$p \cdot p \cdot (1 - p) \cdot p = p^3 \cdot (1 - p) = (0.10)^3 \cdot (0.90)^1 = 0.0009$$

Tuttavia il valore trovato rappresenta la probabilità di ottenere tre successi in un campione di quattro ordini nell'*ordine specificato*. Volendo calcolare il numero di sequenze con tre successi su quattro ordini (in generale il numero di modi in cui si possono selezionare *X* oggetti da un campione di *n*, *indipendentemente dall'ordine*) dobbiamo affidarci alla **regola delle combinazioni**.

Combinazioni

$$_nC_X = \frac{n!}{X!(n-X)!} \qquad \textbf{(5.4)}$$

dove $n! = n \times (n-1) \times \ldots \times 2 \times 1$ è detto *n* fattoriale, e 0!=1

La distribuzione binomiale

Quindi con n=4 e X=3, il numero delle possibili sequenze è

$$ {}_4C_3 = \frac{4!}{3!(4-3)!} = \frac{4\times 3\times 2\times 1}{(3\times 2\times 1)\times 1} = 4 $$

Le quattro possibili sequenze sono

$p\cdot p\cdot p\cdot(1-p)=0.0009$ $p\cdot p\cdot(1-p)\cdot p=0.0009$

$p\cdot(1-p)\cdot p\cdot p=0.0009$ $(1-p)\cdot p\cdot p\cdot p=0.0009$

Di conseguenza la probabilità di osservare tre ordini non corretti (segnalati) su quattro ordini è uguale a:

(N. di possibili sequ.)×(probab. di una particolare sequ.)=

$$= 4 \times 0.0009 = 0.0036$$

Allo stesso modo possono essere derivate le probabilità degli altri quattro possibili risultati della variabile aleatoria: 0, 1, 2 e 4 ordini scorretti.

La distribuzione binomiale

Tuttavia, all'aumentare di *n*, questo calcolo diventa piuttosto laborioso e conviene elaborare un appropriato modello matematico. In generale la distribuzione binomiale è la legge della variabile aleatoria che rappresenta il numero di successi della variabile *X* = "numero di successi" quando i due parametri sono pari a *n* = numero di osservazioni e *p* = probabilità di successo in ciascuna osservazione.

Distribuzione binomiale

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X} \qquad \textbf{(5.5)}$$

dove $P(X_i)$ = probab. di ottenere *X* successi dati *n* e *p*
n = ampiezza campionaria
p = probabilità di successo
1−*p* = probabilità di insuccesso
X = numero di successi nel campione (*X*=0,1,2,…,*n*)

La distribuzione binomiale

Notiamo che l'equazione (5.5) non è altro che una formalizzazione di quanto già derivato intuitivamente. La variabile aleatoria *X* può assumere soltanto i valori interi compresi fra 0 e *n*. Nell'equazione (5.5) il prodotto

$$p^X (1-p)^{n-X}$$

rappresenta la probabilità di ottenere una *particolare sequenza* di *X* successi su *n* osservazioni. Il termine

$$\frac{n!}{X!(n-X)!}$$

rappresenta invece il numero di possibili sequenze di X successi su n osservazioni. Quindi, possiamo determinare la probabilità di osservare X successi in un ordine qualsiasi nel seguente modo:

P(*X*)=(N. di possibili sequ.)×(probab. di una particolare sequ.)

La distribuzione binomiale

Caratteristiche della distribuzione binomiale

- **Forma**: una distribuzione binomiale può essere simmetrica o asimmetrica in base ai valori assunti dai parametri. Per qualsiasi valore di *n* la distribuzione binomiale è simmetrica se *p* = 0.5 e asimmetrica per valori di p diversi da 0.5. L'asimmetria diminuisce all'avvicinarsi di *p* a 0.5 e all'aumentare del numero di osservazioni n.
- **Il valore atteso**: si ottiene moltiplicando fra loro i due parametri *n* e *p*.

Il valore atteso di una distribuzione binomiale

Il valore atteso μ di una distribuzione binomiale è uguale al prodotto tra l'ampiezza del campione n e la probabilità di successo p.

$$\mu = E(X) = np \qquad \textbf{(5.6)}$$

La distribuzione binomiale

In media, a lungo andare, cioè considerando un elevato numero di estrazioni di blocchi di 4 ordini, ci possiamo aspettare un numero medio di ordini non corretti pari a $\mu = E(X) = (4 \times 0.1) = 0.4$.

- **Lo scarto quadratico medio**: si calcola applicando la formula:

Scarto quadratico medio di una distribuzione binomiale

$$\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)} = \sqrt{np(1-p)} \qquad \textbf{(5.7)}$$

Nel nostro esempio, lo scarto quadratico medio della variabile che rappresenta il numero di ordini scorretti è data da:

$$\sigma = \sqrt{4(0.1)(0.9)} = 0.6$$

Si tratta dello stesso risultato che otterremmo applicando l'espressione generale che definisce lo scarto quadratico medio data dall'equazione (5.3).

La distribuzione binomiale

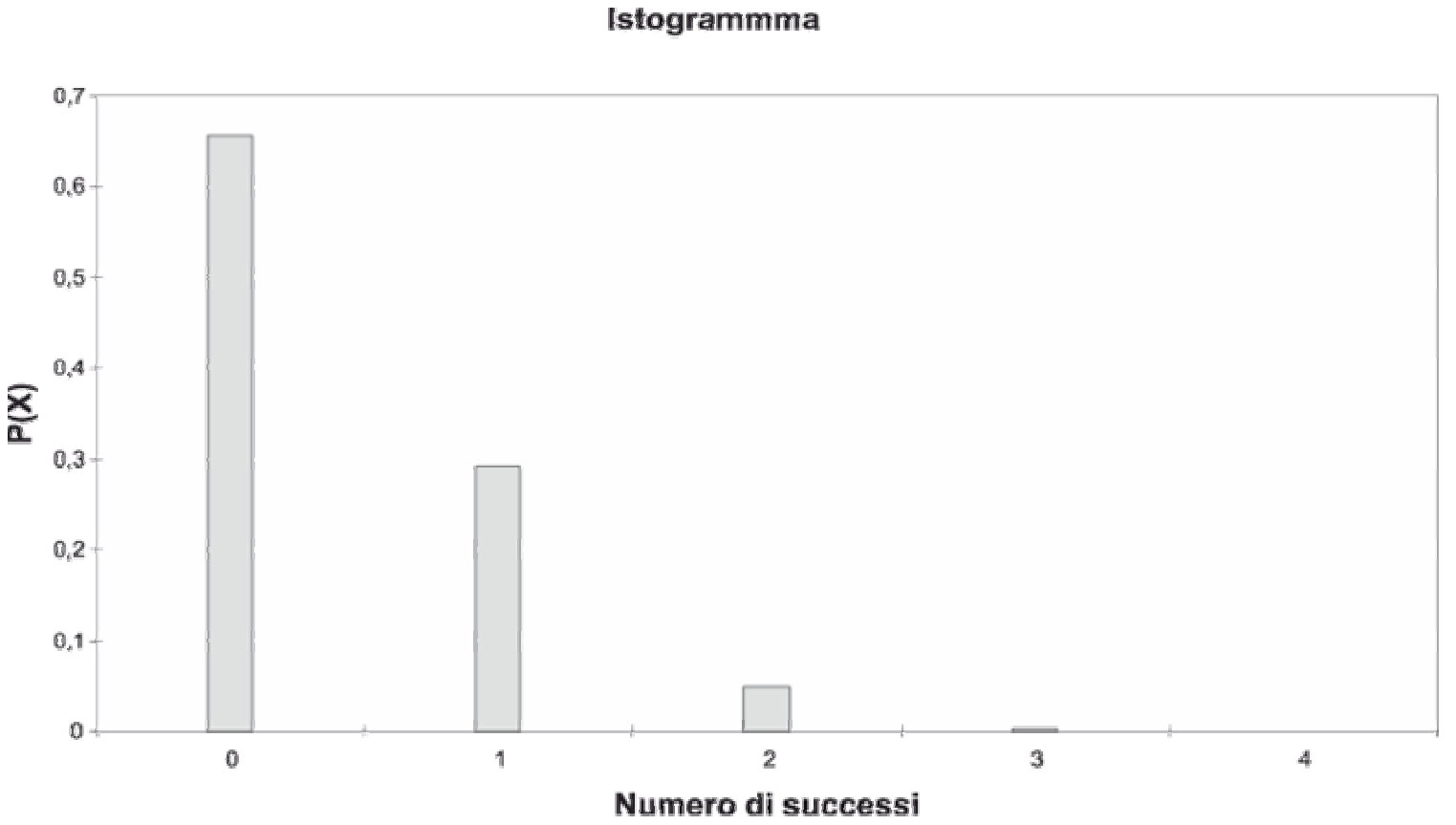
In questo paragrafo abbiamo presentato la distribuzione binomiale e descritto una sua applicazione a un problema aziendale.

Come vedremo nel dettaglio nei Capitoli 8 e 9, questo modello riveste un ruolo molto importante anche nell'ambito della statistica inferenziale, in particolare riguardo alla stima e alla verifica di ipotesi su una proporzione.

La distribuzione binomiale

Figura 5.3

Distribuzione binomiale per *n*=4 e *p*=0.1 realizzata utilizzando Microsoft Excel



La distribuzione di Poisson

In molte applicazioni si è interessati a contare il numero di volte in cui si osserva la realizzazione di un evento in una certa *area di opportunità*.

Un'area di opportunità è un intervallo continuo quale un tempo, una lunghezza, una superficie, o in generale un'area nella quale un certo evento può verificarsi più volte.

Esempi possono essere il numero di difetti su uno sportello di un frigorifero, il numero di telefonate che arrivano in un centralino in un certo periodo di tempo o ancora il numero di persone che entrano in un grande magazzino in un pomeriggio.

La distribuzione di Poisson

Quando si considerano aree di opportunità si può ricorrere alla **distribuzione di Poisson** se sono soddisfatte quattro condizioni:

- si è interessati a contare il numero di volte in cui un certo evento si realizza in una certa area di opportunità
- la probabilità che in una certa area di opportunità si osservi un certo evento è la stessa in tutte le aree di opportunità
- il numero di volte in cui un evento si realizza in una certa area di opportunità è indipendente dal numero di volte in cui un l'evento si è verificato in un'altra area
- la probabilità che in una certa area di opportunità l'evento di interesse si verifichi più di una volta diminuisce al diminuire dell'area di opportunità

La distribuzione di Poisson

Supponiamo di esaminare il numero di clienti che raggiungono una banca in un minuto. L'arrivo di un cliente è l'evento di interesse e l'area di opportunità è l'intervallo temporale di un minuto. Dato che le quattro condizioni sono soddisfatte possiamo ricorrere alla distribuzione di Poisson per determinare la probabilità con cui in un certo intervallo di tempo si presenti in banca un certo numero di clienti.

La distribuzione di Poisson è caratterizza dal parametro λ, che rappresenta il numero atteso di volte (che varia da zero ad infinito) in cui l'evento si verifica nell'area di opportunità considerata. Il numero di volte in cui si verifica un evento X in un certo intervallo temporale varia da zero a infinito (per numeri interi).

La distribuzione di Poisson

L'espressione matematica della distribuzione di Poisson per il numero di eventi *X*, dato che il numero atteso di eventi è pari a λ è dato da

Distribuzione di Poisson

$$P(X) = \frac{e^{-\lambda}\lambda^{X}}{X!} \qquad \textbf{(5.5)}$$

dove $P(X_i)$ = probabilità di ottenere *X* dato λ

λ = numero atteso di successi nell'area di opportunità

e = costante matematica approssimata da 2.71828

X = numero di successi per area di opportunità (X=0,1,2,…)

La distribuzione di Poisson

Riprendiamo l'esempio dell'arrivo di clienti presso una banca e supponiamo che in un minuto arrivano in media tre clienti. Qual è la probabilità che in un certo minuto arrivino esattamente due clienti? Qual è la probabilità che arrivino più di due clienti?

SOLUZIONE

Applicando l'equazione (5.8) si ha:

$$P(X = 2) = \frac{e^{-3.0}(3.0)^2}{2!} = \frac{9}{(2.71828)^3(2)} = 0.2240$$

Per rispondere alla seconda domanda osserviamo che:

$$P(X > 2) = P(X = 3) + P(X = 4) + \cdots + P(X = \infty)$$

Evidentemente, è più agevole il calcolo dell'evento complementare $P(X \leqslant 2)$. Quindi, poiché $P(A) = 1 - P(A')$, la probabilità richiesta può essere ottenuta nel seguente modo:

$$P(X > 2) = 1 - P(X \leqslant 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

La probabilità che in un minuto arrivino allo sportello della banca al più due clienti è pari a 0.423 e la probabilità che ne arrivino più di due è pari a $1 - 0.423 = 0.577$.

La distribuzione di Poisson

Per evitare molti conti, le probabilità relative alla distribuzione di Poisson possono essere ottenute a partire dalla Tavola E.7.

Probabilità per una variabile aleatoria di Poisson
Calcolo di $P(X=2)$ con $\lambda=3$

			λ	
X	2.1	2.2	....	3.0
0	.1225	.1108	....	.0498
1	.2572	.2438	....	.1494
2	.2700	.2681	....	.2240
3	.1890	.1966	....	.2240
4	.0992	.1082	....	.1680
5	.0417	.0476	....	.1008
6	.0146	.0174	....	.0504
7	.0044	.0055	....	.0216
8	.0011	.0015	....	.0081
9	.0003	.0004	....	.0027
10	.0001	.0001	....	.0008
11	.0000	.0000	....	.0002
12	.0000	.0000	....	.0001

Fonte: Tavola E.7