

Corso di Statistica Facoltà di Economia

a.a. 2010-2011

francesco mola

Lezione n° 14

Sommario

- Variabili Casuali Discrete
 - *Uniforme*
 - *Bernoulli*
 - *Binomiale*
 - *Ipergeometrica*
 - *Poisson*
 - *Geometrica*

VARIABILI CASUALI DISCRETE

- X discreta
- X continua discretizzata

Variabile Casuale Uniforme

$$X \sim Ud(n)$$

Esempio: lancio di un dado

n individui con una stessa caratteristica

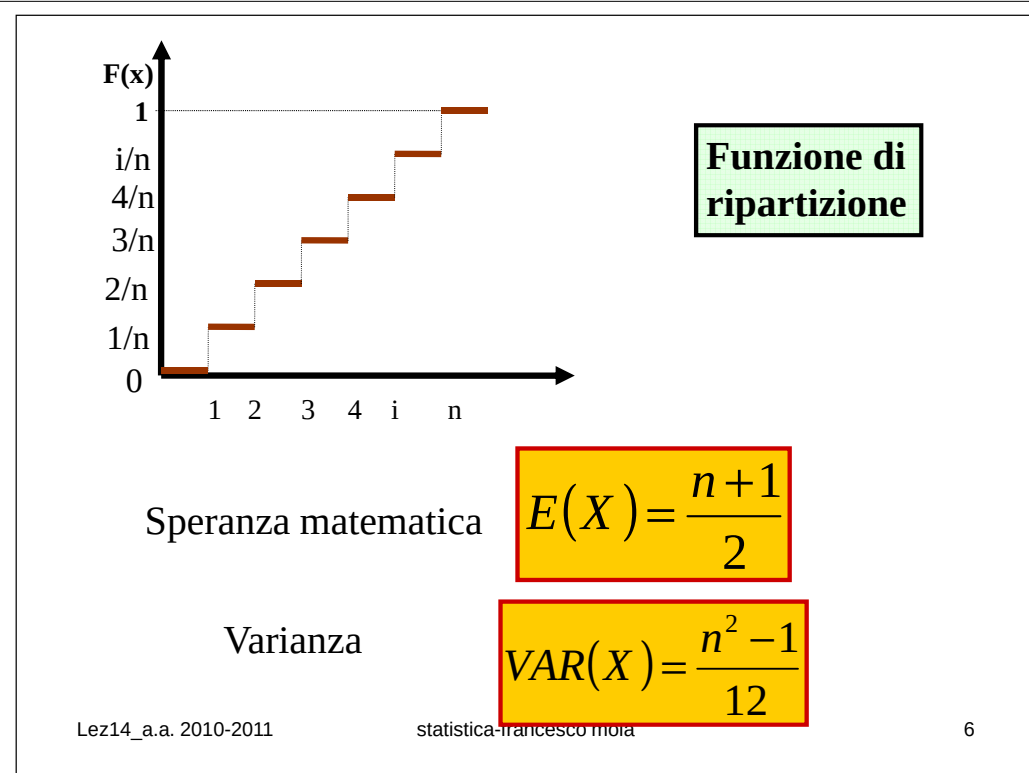
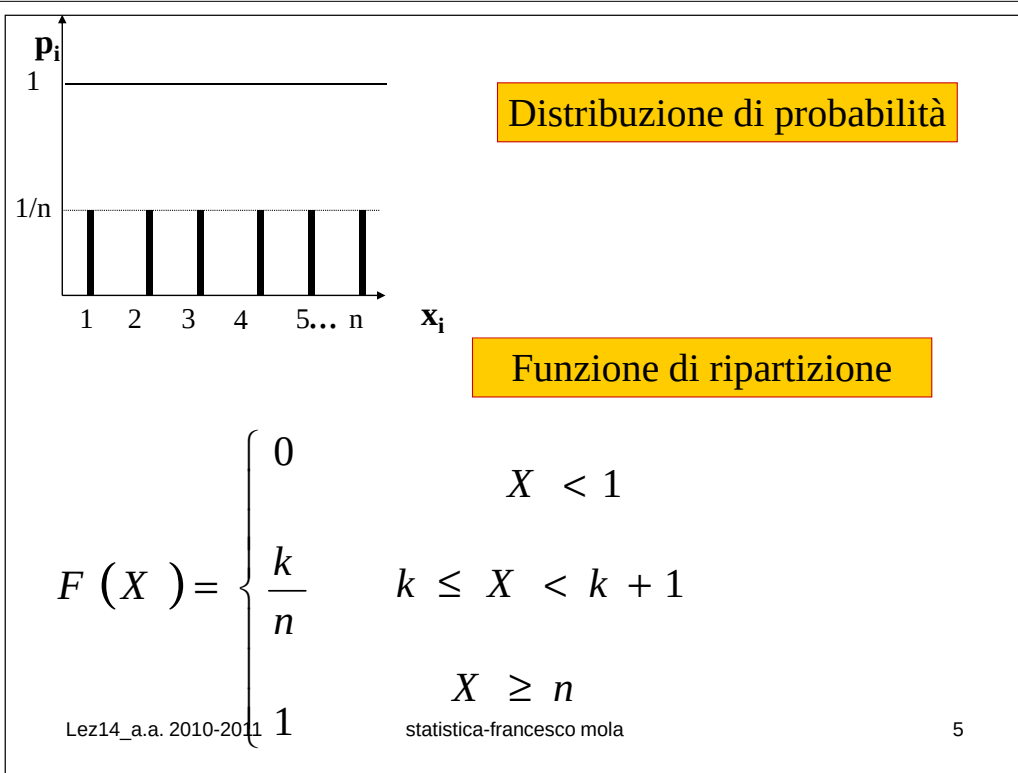
La prova può essere associata all'estrazione di una pallina da un'urna con n palline

$$x_1 = 1 \quad x_2 = 2 \quad x_3 = 3 \dots x_n = n$$

Funzione di probabilità

$$P(X = x) = \frac{1}{n} > 0 \quad \text{Probabilità costante}$$

$$\sum_{i=1}^n P(X = x) = \sum_i \frac{1}{n} = \frac{n}{n} = 1$$



Variabile Casuale di Bernoulli

$$X \sim Ber(p)$$

La prova consiste nel considerare un evento E che può o non può verificarsi

E si verifica l'evento

$\overline{E}$ non si verifica l'evento

Lez14_a.a. 2010-2011 statistica-francesco mola 7

Esempio: lancio di una moneta

Espressione di un voto referendario

Lancio di un dado(pari - dispari)

Possiamo quindi avere ad ogni prova un successo o un insuccesso

$$X : \begin{cases} 0 = \text{insuccesso} = \overline{E} \\ 1 = \text{successo} = E \end{cases}$$

$P(X=1) = p$
 $P(X=0) = 1 - p = q$

Lez14_a.a. 2010-2011 statistica-francesco mola 8

Distribuzione di probabilità

$$P(X = x) = p^x (1-p)^{1-x} \quad x = 0,1$$

$$\sum_{x=0,1} p^x (1-p)^{1-x} = p^0 (1-p)^{1-0} + p^1 (1-p)^{1-1} =$$

$$= 1 - p + p = 1$$

$$x = 0 \Rightarrow p^0 (1-p)^{1-0} = 1 - p$$

$$x = 1 \Rightarrow p^1 (1-p)^{1-1} = p$$

Funzione di ripartizione

$$F(X) = \begin{cases} 0 & X < 0 \\ 1 - p & 0 \leq X < 1 \\ p & X \geq 1 \end{cases}$$

$$E(X) = \sum x_i p_i = 0(1-p) + 1(p) = p$$

$$\begin{aligned} \text{Var}(X) &= \sum (x_i - \mu)^2 p_i = (0 - p)^2 (1-p) + (1 - p)^2 p = \\ &= p(1-p)[p + 1 - p] = p(1-p) \end{aligned}$$

$$E(X) \equiv \text{con la probabilità !!!!}$$

$$\text{VAR}(X) \text{ è massima se } p = q = \frac{1}{2} = 0.5$$

Variabile Casuale Binomiale

$$X \sim \text{Bin}(n, p)$$

Si ha quando abbiamo **n** prove **bernoulliane**

Esempi:

Estrazione con ripetizione di **n** palline da un'urna;
n interviste con risposte (SI/NO)

NOTA:

Le **n** sottoprobe sono a probabilità costanti!

Se **n=1** implica che **Bin(n,p)** coincide con **Ber(p)**

Variabile Casuale Binomiale

$$X = 0$$

Significa che in n sottoprove non si è verificato nemmeno un successo!

$$X = 1$$

Significa che in n sottoprove si è verificato solo un successo!

$\vdots$

$\vdots$

$$X = n$$

Significa che in n sottoprove si sono verificati n successi!

Se abbiamo x successi, abbiamo anche $n-x$ insuccessi .

Poichè le sottoprove sono indipendenti, una generica sequenza avrà probabilità

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

Quante sono tutte le possibili sequenze costituite dagli elementi di $\Omega(S)$???

$$2^n$$

Quante sono i modi di combinarsi di una specifica sequenza ???

$$\binom{n}{x} \text{ cioè } n \text{ elementi presi } x \text{ ad } x$$

Qual è la probabilità di ciascuna

delle $\binom{n}{x}$ sequenze? $p^x (1-p)^{n-x}$

Possiamo quindi in generale dire che :

$$X \sim \text{Bin}(n, p)$$

con distribuzione di probabilità

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Siccome le n sottoprove sono indipendenti, abbiamo:

$$1) P(X = x) \geq 0$$

$$\sum_{x=0}^n P(X = x) = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = (p + q)^n = 1^n = 1$$

$$F(X) = \begin{cases} 0 & X < 0 \\ \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n & 0 \leq X < 1 \\ \binom{n}{1} p^1 (1-p)^{n-1} & 1 \leq X < 2 \\ \vdots & \vdots \\ \binom{n}{n} p^n (1-p)^{n-n} = p^n & n \leq X < \infty \end{cases}$$

$$X \sim \text{Bin}(n, p)$$

$$X = X_1 + X_2 + X_3 + \dots + X_n$$

$X_i =$ Bernoulliane indipendenti

$$i = 1, 2, \dots, n$$

$$\begin{aligned} E(X) &= E(X_1 + X_2 + X_3 + \dots + X_n) = \\ &= E(X_1) + E(X_2) + \dots + E(X_n) = \\ &= p + p + \dots + p = np \end{aligned}$$

$$\text{VAR}(X) = \text{VAR}(X_1 + X_2 + X_3 + \dots + X_n) =$$

$$\begin{aligned} p(1-p) + p(1-p) + \dots + p(1-p) = \\ np(1-p) = npq \end{aligned}$$

Al crescere di n crescono media e varianza della v.c. Binomiale

esempio

Il proprietario di un negozio sa che la probabilità per un consumatore di acquistare un determinato prodotto è pari a 0.2.

Nel negozio in quel momento ci sono 9 clienti.

- 1) Qual è la probabilità che nessun cliente acquisti quel prodotto?
- 2) Qual è la probabilità che almeno 2 clienti acquistino quel prodotto?
- 3) Qual è il numero di clienti che il negoziante si aspetta acquistino quel prodotto?

Soluzione:

la prima cosa da fare è individuare lo schema di riferimento e quindi di conseguenza qual è la distribuzione di probabilità che dobbiamo considerare.

Alcune considerazioni preliminari:

il prodotto può o non può essere acquistato, quindi gli eventi sono dicotomici e la distribuzione potrebbe essere di tipo bernoulliana con parametro $p=0.2$.

Questo vale per un singolo cliente, ma noi abbiamo $n=9$ clienti, quindi 9 bernoulliane. Non essendo interessati al fatto che un cliente possa acquistare più di un prodotto e non essendo interessati altresì al fatto che il comportamento di acquisto di un cliente possa influenzare un altro cliente, possiamo considerare queste n bernoulliane come indipendenti tra loro.

la distribuzione di riferimento è allora una Binomiale con parametri $n=9$ e $p = 0.2$

$$X \sim \text{Bin}(n = 9, p = 0.2)$$

1) Probabilità che nessun cliente acquisti quel determinato prodotto

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(X = 0) = \binom{9}{0} 0.2^0 (1-0.2)^{9-0} =$$

$$= 1 * 1 * (0.8)^9 = 1 * 1 * 0.134 = 0.134$$

2) Probabilità che ALMENO 2 clienti acquistano quel determinato prodotto

Almeno 2 clienti significa nella pratica : 2 clienti , 3 clienti, 4 clienti.... 9 clienti.

Dovremmo quindi calcolare la stessa probabilità di prima per $x=2$, per $x=3, \dots$, per $x=9$ e poi sommarle.

Ma sappiamo che la somma delle probabilità è sempre pari ad 1! Quindi Possiamo fare meno calcoli considerando il complemento ad 1 della probabilità

$$1 - [P(X = 0) + P(X = 1)]$$

$$P(X = 0) = 0.134 \quad (\text{calcolato precedentemente})$$

$$P(X = 1) = \binom{9}{1} 0.2^1 (1-0.2)^{9-1} =$$

$$= 9 * 0.2^1 * (0.8)^8 = 9 * 0.2 * 0.167 = 9 * 0.034 = 0.30$$

$$1 - [P(X = 0) + P(X = 1)] = 1 - [0.134 + 0.3] = 0.566$$

E' noto che:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Casi particolari:

$$x = 0 \Rightarrow \binom{n}{0} = 1$$

$$x = 1 \Rightarrow \binom{n}{1} = n$$

$$x = n - 1 \Rightarrow \binom{n}{n-1} = n$$

$$x = n \Rightarrow \binom{n}{n} = 1$$

Si noti inoltre che:

$$x = k \Rightarrow \binom{n}{k} = \frac{n * (n-1) * (n-2) * \dots * (n-k+1) * \cancel{(n-k)!}}{k! * \cancel{(n-k)!}} = \frac{n * (n-1) * (n-2) * \dots * (n-k+1)}{k!}$$

3) Qual è il numero di clienti che il negoziante si aspetta acquistino quel prodotto?

Si tratta di calcolare il valore atteso $E(X)$ che in una binomiale è uguale a:

$$E(X) = np = 9 * 0.2 = 1.8$$

Il numero di persone che ci si aspetta acquistino il prodotto è circa 2

E se arrivano contemporaneamente 50 persone?

$$E(X) = np = 50 * 0.2 = 10$$

Il numero di persone che ci si aspetta acquistino il prodotto è 10

Nel primo caso la varianza è:

$$VAR(X) = np(1-p) = 9 * 0.2 * 0.8 = 1.44$$

Nel secondo caso la varianza è:

$$VAR(X) = np(1-p) = 50 * 0.2 * 0.8 = 8$$

Al crescere di n aumenta la variabilità!

v.c. ipergeometrica

Problema

Consideriamo un libro di esercizi di statistica che ha 120 esercizi di cui 40 di statistica descrittiva e 80 di inferenza.

Uno studente decide un pomeriggio di svolgere 5 esercizi scelti casualmente dal libro, assicurandosi ovviamente di non selezionare più di una volta lo stesso esercizio

(tecnicamente si chiama schema **senza reintroduzione**).

Ci si chiede:

$P(2 \text{ esercizi di descrittiva sui } 5 \text{ esercizi estratti}) = ?$

Hp(1): non siamo interessati all'ordine di estrazione degli esercizi di descrittiva estratti.

Possiamo quindi accettare come equiprobabili tutte le combinazioni di 120 esercizi presi 5 alla volta; questi sono quindi i "CASI POSSIBILI"

$$\binom{120}{5}$$

Hp(2): essendo interessati all'evento "2 esercizi di descrittiva e 3 di inferenza", diciamo che:

1) i 2 esercizi di descrittiva possono essere scelti in

$$\binom{40}{2} \text{ Diversi modi}$$

2) i 3 esercizi di inferenza possono essere scelti in

$$\binom{80}{3} \text{ Diversi modi}$$

Quindi: $\binom{40}{2} \binom{80}{3} = \text{Numero dei casi favorevoli}$



P(2 esercizi di descrittiva su 5 esercizi estratti)=

$$\frac{\binom{40}{2} \binom{80}{3}}{\binom{120}{5}} = \frac{780 \times 82160}{190578024} = 0.3363$$

generalizzando

Distribuzione Ipergeometrica

N= # elementi totale

n=# elementi estratti

x=# elementi a cui siamo interessati

M= # totale elementi aventi le stesse caratteristiche degli **x** elementi

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \cong Hg(N, M, n)$$

$$E(X) = np$$

$$VAR(X) = npq \left(\frac{N-n}{N-1} \right) \quad \text{dove } p = \frac{M}{N}$$

Nota: Se $N \rightarrow \infty$ si dimostra che:

$$\lim_{N \rightarrow \infty} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \cong \binom{n}{x} p^x (1-p)^{n-x}$$

Per il nostro esempio abbiamo:

N=120
n=5
M=40
x=2

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

$$P(X = 2) = \frac{\binom{40}{2} \binom{120-40}{5-2}}{\binom{120}{5}} = 0.3363$$

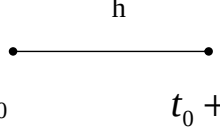
Distribuzione di Poisson

- Limite della Binomiale
- Si fissa un intervallo (tempo, spazio) costante e si contano quanti eventi di interesse si verificano nell'intervallo

Processo poissoniano o di Poisson

$X \cong P(\lambda)$ = Una serie di prove ripetute nelle stesse condizioni

1) la probabilità che in un intervallo h si verifichi **un solo** evento è proporzionale a λh



The diagram shows a horizontal line segment representing an interval of length h. The left endpoint is labeled t_0 and the right endpoint is labeled $t_0 + dx$.

2) la probabilità che in h si ha più di un evento = 0

3) $P[\text{eventi in } (t_0, t_0 + dx) \cap \text{eventi in } (t_1, t_1 + dx)]$
 $= P[\text{eventi in } (t_0, t_0 + dx)] \bullet P[\text{eventi in } (t_1, t_1 + dx)]$

Quindi sono indipendenti!! (il verificarsi di un evento non è condizionato da un evento già verificatosi e non condiziona un evento che potrà verificarsi)

Consideriamo una v.c. $Bin(n, p)$ con $np = \lambda = \text{costante}$

e quindi $p = \frac{\lambda}{n}$ ottenendo

$$\begin{aligned} & \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \\ & = \frac{n(n-1)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \end{aligned}$$

passando al limite per $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \left[\frac{n(n-1) \dots (n-x+1)}{x!} \left(\frac{\lambda}{n} \right)^x \left(1 - \frac{\lambda}{n} \right)^{n-x} \right] =$$

$$\lim_{n \rightarrow \infty} \left[\frac{n(n-1) \dots (n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n} \right)^n \left(1 - \frac{\lambda}{n} \right)^{-x} \right] =$$

$$\frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \left[\frac{n(n-1) \dots (n-x+1)}{n^x} \right] \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^n \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^{-x} =$$

$$\frac{\lambda^x}{x!} \times 1 \times e^{-\lambda} \times 1 =$$

$$\frac{\lambda^x e^{-\lambda}}{x!}$$

Si ricorda che

$$\lim_{h \rightarrow \infty} \left(1 + \frac{1}{h} \right)^h = e$$

quindi

$$\Rightarrow \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^n = e^{-\lambda}$$

$$P(X) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{con} \quad \begin{matrix} x = 0, 1, 2, \dots \\ \lambda > 0 \end{matrix}$$

$$\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Sviluppo in serie di e

$$E(X) = \lambda$$

$$VAR(X) = \lambda$$

Se n è molto grande e p è piccolo, la v.c. di Poisson è una buona approssimazione della Binomiale

Distribuzione Geometrica

Data una successione di prove Bernoulliane indipendenti con probabilità costante p di successo, la v.c. che rappresenta il numero di prove necessarie affinché si presenti una prima volta l'evento "successo" è:

$$X \sim Geo(p)$$

Con distribuzione di probabilità pari a:

$$p(1-p)^{x-1} \quad \text{con } x = 1, 2, 3, \dots$$

Essendo $p > 0$ la distribuzione di probabilità è sempre positiva e inoltre la somma delle probabilità è 1 per la proprietà della convergenza di serie geometriche.

Distribuzione Geometrica

Il valore atteso e la varianza di una distribuzione geometrica sono :

$$E(X) = \frac{1}{p} \quad \text{e} \quad \text{Var}(X) = \frac{1-p}{p^2}$$

Proprietà:

1) E' una v.c. che si ritrova nelle operazioni campionarie ripetute sino a trovare una data caratteristica (successo).

2) È l'unica v.c. discreta senza memoria, cioè il ritardo in k sottoprove indipendenti non modifica la probabilità che esso si possa verificare in sottoprove successive!

Andamento del valore atteso e della varianza di una v.c. geometrica al variare di p

