Università di Firenze - Corso di Laurea in Statistica

Statistica

A.A. 2010/2011

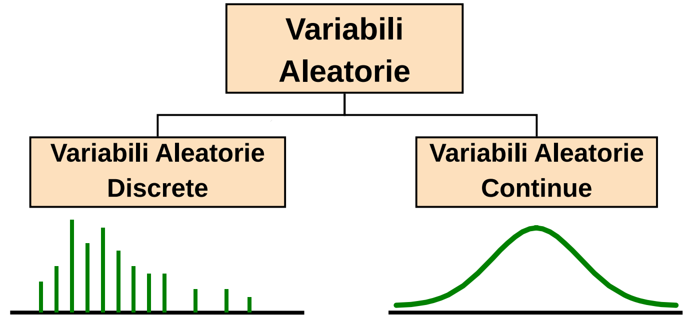
Variabili casuali discrete & alcuni modelli probabilistici

Cicchitelli: parte del cap 13 e parte del cap 14

A cura di Leonardo Grilli, Carla Rampichini

Variabili aleatorie /1

- Variabile casuale o variabile aleatoria
  - Rappresenta un possibile valore numerico prodotto dall'esperimento aleatorio



Variabili aleatorie /2

- Una variabile aleatoria (v.a.) è un numero ignoto (associato all'esito di un esperimento aleatorio) di cui sono specificati:
  - l'insieme dei valori che può assumere: **supporto**
  - come la probabilità si distribuisce su tali valori: **distribuzione di probabilità**
- Se il supporto ha cardinalità finita o infinita numerabile (es. i numeri naturali)

  → **v.a. discreta** (tipica dei processi di conteggio)
- Se il supporto ha cardinalità infinita non numerabile, es. $[0,1]$, $[0,\infty)$, $(-\infty, \infty)$

  → **v.a. continua** (tipica dei processi di misurazione)

Variabili aleatorie discrete /1

- Una v.a. discreta può assumere valori appartenenti ad un insieme numerabile

Esempi:

- Lancia due volte un dado
  - Sia X il numero di volte che si ottiene 4

    → X può essere 0, 1, o 2 volte
- Lancia una moneta 5 volte
  - Sia X il numero complessivo di teste

    → X = 0, 1, 2, 3, 4, o 5

Variabili aleatorie discrete /2

- Consideriamo un urna contenente 100 palline:
  - 20 palline con il numero 1
  - 30 palline con il numero 2
  - 50 palline con il numero 3
- Il numero che si presenterà alla prossima estrazione è ignoto, quindi è una v.a.
  - Quali valori può assumere? (**supporto**)
  - Con quali probabilità? (**distribuzione di probabilità**)
- Il supporto è solitamente facile da determinare, in questo esempio è S={1,2,3} → v.a. discreta
- Specificare la distribuzione di probabilità è invece un'operazione delicata perché richiede la formulazione di alcune ipotesi (spesso non verificabili)

Variabili aleatorie discrete /3

Nel caso dell'urna per determinare le probabilità possiamo usare la ***definizione classica*** → assumendo ***equiprobabilità*** (= le palline hanno la stessa probabilità di essere estratte) le probabilità coincidono con le frequenze relative

x	*f(x)*
1	0.2
2	0.3
3	0.5

Il prossimo numero che si presenterà è ignoto, ma possiamo dire che sarà 1 con probabilità 0.2, oppure sarà 2 con probabilità 0.3 oppure sarà 3 con probabilità 0.5

L'ipotesi di equiprobabilità non è sempre adeguata. Ad esempio, non lo è se l'estrazione è fatta da un soggetto che preferisce i numeri pari e che può vedere le palline e decidere quale estrarre

Variabili aleatorie discrete /4

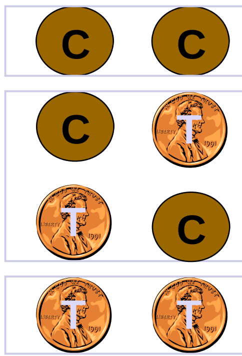
- Nel caso l'estrazione venga **ripetuta** *in modo indipendente e in identiche condizioni* (nell'esempio dell'urna ciò accade quando prima di ogni estrazione si rimette nell'urna la pallina precedentemente estratta) per lo stesso problema si può usare l'**approccio frequentista**: la probabilità di un certo elemento del supporto è stimata con la frequenza relativa con cui quell'elemento si è presentato nelle prove effettuate
- Al crescere del numero di prove le frequenze relative convergono alle "vere" probabilità
- Le "vere" probabilità coincidono con quelle dell'approccio classico se l'ipotesi di equiprobabilità è vera (in effetti, la ripetizione dell'esperimento è l'unico modo per verificare empiricamente l'ipotesi di equiprobabilità)

Funzione di (massa di) probabilità /1

Esperimento: lancia 2 monete bilanciate. Sia X = # di teste.

Calcola f(x) , cioè P(X = x) , per tutti i valori di x:

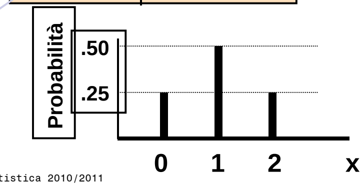
4 possibili risultati



Distribuzione di Probabilità

Valori di x	Probabilità
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25

Funzione di probabilità



Funzione di (massa di) probabilità /2

- Due proprietà necessarie della funzione di probabilità:
  - $f(x) \geq 0$ per ogni valore di x
  - Le singole probabilità sommano a 1:

$$\sum_x f(x) = 1$$

(La notazione indica che la sommatoria si estende a tutti i possibili valori di x)

Funzione di ripartizione

- La **funzione di ripartizione**, indicata con $F(x_0)$, esprime la probabilità che X non superi il valore x_0

$$F(x_0) = P(X \leq x_0)$$

- In altre parole,

$$F(x_0) = \sum_{x \leq x_0} f(x)$$

Sintesi delle distribuzioni

- In analogia con quanto visto per le distribuzioni di frequenze, anche per le distribuzioni di probabilità è utile avere degli indici di sintesi
  - Indici di posizione: media (nel caso delle v.a. chiamato anche ***valore atteso***)
  - Indici di variabilità: **varianza** e **deviazione standard**

Valore atteso /1

- Il **valore atteso** di una distribuzione discreta è la media dei possibili valori pesata con le rispettive probabilità

$$\mu = E(X) = \sum_x x f(x)$$

- Esempio: Lancia 2 monete,
  X = # di teste,
  calcoliamo il valore atteso di X:

x	f(x)
0	.25
1	.50
2	.25

E(X) = (0 x .25) + (1 x .50) + (2 x .25)
= 1.0

Valore atteso /2

- è uno scalare (riduzione di dimensione)
- è una costante (cioè, non è aleatorio)
- è l'equivalente certo di una somma aleatoria

Esempio roulette (18 rossi, 18 neri, 1 verde):

X = somma disponibile a seguito della puntata di 1 euro sul rosso

X = 2 se esce rosso, X = 0 se <u>non</u> esce rosso,

P(*X* = 2) = 18/37, P(*X* = 0) = 19/37,   E(*X*) = 36/37 = 0.973

L'equivalente certo della scommessa è 0.973 euro → con questo costo il gioco sarebbe equo: poiché la puntata costa 1 euro il gioco è a favore del banco

Valore atteso /3

- Un esperto finanziario afferma: "Penso che la prossima settimana il mercato salga, ma la mia previsione è negativa"
- Contraddittorio? Non necessariamente.
  - Scenario A - il mercato sale: prob 0.6, guadagno 1000 euro
  - Scenario B - il mercato scende: prob 0.4, perdita 5000 euro
  - Evento più probabile: il mercato sale, valore atteso = 1000*0.6+(-5000)*0.4 = -1400 euro
- I rendimenti delle attività finanziarie sono asimmetrici!

Cfr. N.N. Taleb (2003) Giocati dal caso – Il ruolo della fortuna nella finanza e nella vita. Il Saggiatore.

Varianza e Deviazione standard

- **Varianza** di una variabile aleatoria discreta X

$$\sigma^2 = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x)$$

- **Deviazione standard** (o Scarto quadratico medio) di una variabile aleatoria discreta X

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_x (x-\mu)^2 f(x)}$$

la Dev.Std. è un indice di variabilità (= dispersione attorno alla media), e quindi di incertezza. Dev.std. grande → media poco "affidabile"

Esempio: deviazione standard

- Esempio: Lancia 2 monete, X = # di teste,
- Calcoliamo la deviazione standard

$$\sigma = \sqrt{\sum_x (x-\mu)^2 f(x)}$$

- Poiché μ = 1 si ottiene

$$\sigma = \sqrt{(0-1)^2(.25)+(1-1)^2(.50)+(2-1)^2(.25)} = \sqrt{.50} = .707$$

Numero possibile di teste (0, 1, o 2)

Funzioni di variabili aleatorie

- X v.a. discreta, g( ) funzione, Y=g(X) è una nuova v.a.
- Valore atteso di Y

$$E[Y] = E[g(X)] = \sum_x g(x) f(x)$$

- Esempio: Lancia 2 monete, X = # di teste
  - Se il guadagno Y in euro è pari al quadrato del numero di teste, il valore atteso del guadagno è

$$E(Y) = E(X^2) = \sum_x x^2 f(x)$$

$$= 0^2(.25) + 1^2(.50) + 2^2(.25) = 1.5$$

Proprietà del valore atteso

- Per una ***variabile aleatoria degenere*** che certamente assume il valore a → valore atteso a e varianza 0

$$E(a) = a \qquad Var(a) = 0$$

- ***Cambiamento di scala***: se una v.a. viene moltiplicata per una costante b, il valore atteso risulta moltiplicato per b e la varianza per b^2

$$E(bX) = bE(X) \qquad Var(bX) = b^2 Var(X)$$

Combinazione lineare di variabile aleatoria

- Sia X una variabile aleatoria con media μ_X e varianza σ^2_X
- Siano a e b due costanti.
- Sia $Y = a + bX$
- Allora la media e varianza di Y sono

$$\mu_Y = E(a + bX) = a + b\mu_X$$

$$\sigma^2_Y = Var(a + bX) = b^2\sigma^2_X$$

- La deviazione standard di Y è allora

$$\sigma_Y = |b|\sigma_X$$

Sono le stesse proprietà che valgono per media, varianza e dev. std. calcolati su un insieme di dati

Famiglie parametriche di v.a. /1

- Nelle applicazioni è utile assumere che la funzione di probabilità di una v.a. sia una funzione analitica dei punti del supporto e che dipenda da uno o più parametri
- Una **famiglia parametrica di v.a.** (detta anche **modello probabilistico**) è una collezione di funzioni di probabilità dipendenti da un parametro

$$\{P(X = x) = f(x;\theta): \quad \theta \in \Theta\}$$

- P è una funzione di probabilità analitica (cioè una formula matematica in cui x e θ entrano come argomenti)
- θ è uno scalare o un vettore (detto **parametro**) che individua in modo univoco il membro della famiglia
- Θ è l'insieme dei valori assumibili da θ (detto **spazio parametrico**) e definisce l'ampiezza della famiglia

Famiglie parametriche di v.a. /2

- Es. lancio di un dado truccato (sappiamo che il dado è truccato, ma non conosciamo le probabilità di uscita dei vari numeri)
  - Il supporto è indubbiamente S= {1, 2, 3, 4, 5, 6}, ma la v.a. che assegna probabilità 1/6 a ogni faccia non è adeguata
  - Quale v.a. usare?
- Se si assume che il dado sia sbilanciato in favore dei valori 1 e 6, ma non si sa quanto sbilanciato, si può usare la seguente famiglia parametrica

$$P(X=x;\theta)=\begin{cases}\dfrac{1}{6}+2\theta & x\in\{1,6\}\\[2mm] \dfrac{1}{6}-\theta & x\in\{2,3,4,5\}\end{cases}$$

Parametro

Spazio parametrico

$$\theta\in\left[0,\frac{1}{6}\right]$$

Famiglie parametriche di v.a. /3

$$P(X=x;\theta)=\begin{cases}\dfrac{1}{6}+2\theta & x\in\{1,6\}\\[2mm] \dfrac{1}{6}-\theta & x\in\{2,3,4,5\}\end{cases}$$

Parametro

Spazio parametrico

$$\theta\in\left[0,\frac{1}{6}\right]$$

- Per ogni valore di θ tra 0 e 1/6 si ha un membro della famiglia, cioè una specifica v.a.

theta	*0.000*
1	0.167
2	0.167
3	0.167
4	0.167
5	0.167
6	0.167
	1.000

theta	*0.021*
1	0.208
2	0.146
3	0.146
4	0.146
5	0.146
6	0.208
	1.000

theta	*0.042*
1	0.250
2	0.125
3	0.125
4	0.125
5	0.125
6	0.250
	1.000

theta	*0.167*
1	0.500
2	0.000
3	0.000
4	0.000
5	0.000
6	0.500
	1.000

Famiglie parametriche di v.a. /4

- In una famiglia parametrica le probabilità sono "legate" tra di loro da una formula: scelto il parametro, si calcolano tutte le probabilità
- Derivare tutte le probabilità da una formula ha vantaggi e svantaggi:
  - PRO: semplifica la struttura. Nell'esempio precedente invece di dichiarare le 6 probabilità è sufficiente dichiarare il valore di θ
  - CONTRO: pone delle restrizioni. Nell'esempio precedente non è possibile attribuire alla faccia 2 una probabilità superiore a quella della faccia 1

Famiglie parametriche che tratteremo

Famiglie parametriche di v.a.
(o distribuzioni di probabilità)
(o modelli probabilistici)

- V.a. **Discrete**
  - Bernoulli e Binomiale
  - Ipergeometrica
  - Poisson
- V.a. **Continue**
  - Uniforme
  - Normale
  - Chi-quadrato

Distribuzione Bernoulli /1

- La distribuzione Bernoulli è la famiglia parametrica (l'unica possibile) per le variabili binarie, cioè variabili che indicano se un certo evento E è vero o falso
  - X = 1 ⇔ l'evento E si è verificato (cosiddetto SUCCESSO)
  - X = 0 ⇔ l'evento E non si è verificato (cosiddetto INSUCCESSO)
- Esempio: favorevole/contrario, soddisfatto/insoddisfatto, buono/difettoso
- La definizione di E e quindi di ciò che debba intendersi per SUCCESSO è arbitraria, es. in un problema di controllo di qualità si possono scegliere le due alternative
  - E = {pezzo buono} → SUCCESSO (X = 1) denota un pezzo buono
  - E = {pezzo difettoso} → SUCCESSO (X = 1) denota un pezzo difettoso

Distribuzione Bernoulli /2

- Indichiamo con p∈(0,1) la probabilità di SUCCESSO, cioè P(X=1) = p
- Di conseguenza la probabilità di INSUCCESSO è 1-p, cioè P(X=0) = 1-p
- In forma compatta si può dunque scrivere

$$P(X=x) = p^x(1-p)^{1-x} \qquad x \in \underbrace{\{0,1\}}_{\text{supporto}} \qquad p \in \underbrace{(0,1)}_{\text{spazio parametrico}}$$

sostituendo x con 0 → 1-p

sostituendo x con 1 → p

Notazione: X~Be(p)

(si legge: la v.a. X ha una distribuzione Bernoulli con probabilità di successo p)

Distribuzione Bernoulli /3

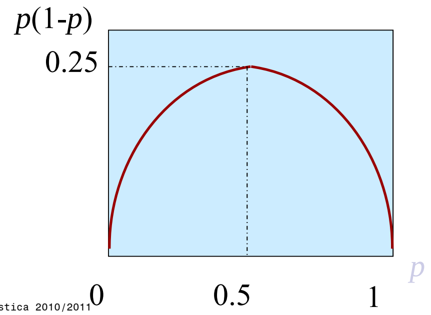
- Media (valore atteso) e varianza:

$$\mu = E(X) = 0(1-p) + 1p = p$$

$$\sigma^2 = Var(X) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)$$

Per un qualunque membro della famiglia Bernoulli la media coincide con la probabilità di successo p

Qual è la relazione tra varianza e media?



Distribuzione Bernoulli /4

- Si riportano alcune sequenze di 20 numeri 0-1:
  - tre sequenze sono casuali, generate da prove bernoulliane con p=0.5
  - due sequenze non sono casuali, ma scritte da una persona
- Quali sono verosimilmente le due sequenze non casuali?
- Che caratteristiche hanno le tre sequenze casuali?

A	1	1	0	1	1	0	0	1	1	0	1	1	0	1	0	1	1	0	0	1
B	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
C	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
D	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	1	1
E	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	0	1	0	0

D. binomiale: definizione /1

- La distribuzione binomiale è un modello per il **numero di successi** in n prove <u>identiche</u> e <u>indipendenti</u>
- L'ordine in cui si presentano i successi è irrilevante, conta solo il totale
- Ogni prova ha come esito successo/insuccesso e quindi ha distribuzione Bernoulli

D. binomiale: definizione /2

- Formalmente:
  - sia $X_1, X_2, \ldots, X_n$ un insieme di n v.a Bernoulli
  - sia $X = \Sigma_i X_i$ la loro somma.
  - poiché ogni X_i vale 1 in caso di successo e 0 altrimenti, X è interpretabile come il numero totale di successi sulle n prove (e quindi assume come valori gli interi tra 0 e n)
- <u>Se</u> le n v.a. Bernoulli sono iid-p (che si legge: indipendenti e identicamente distribuite con probabilità di successo p) <u>allora</u> la v.a. X ha distribuzione binomiale di parametri n e p, in simboli X~B(n, p)

D. binomiale: definizione /3

- La variabile somma $X = \Sigma_i X_i$ è sempre interpretabile come numero totale di successi sulle n prove, ma la sua distribuzione è binomiale quando le prove soddisfano i due requisiti menzionati:
  - <u>Indipendenza</u>: il risultato di una prova non dipende in alcun modo dal risultato delle altre prove

$$P(X_i = 1 \cap X_j = 1) = P(X_i = 1) \times P(X_j = 1) \quad \forall i \neq j$$

  - <u>Prove identiche</u>: la probabilità di successo è la stessa in tutte le prove

$$P(X_i = 1) = P(X_j = 1) \quad \forall i \neq j$$

D. binomiale: definizione /4

- I requisiti di indipendenza e prove identiche sono soddisfatti quando si lancia ripetutamente una moneta oppure quando si effettuano estrazioni da un urna **con reintroduzione**
  - Il requisito di indipendenza è violato quando la probabilità di successo in una prova dipende dal risultato delle altre, ad es. quando si effettuano estrazioni da un urna **senza reintroduzione** oppure quando si valuta la decisione di acquisto da parte di un insieme di clienti che si influenzano a vicenda
  - Il requisito di prove identiche è violato quando la probabilità di successo cambia nel tempo, ad es. perché le prove sono effettuate da un soggetto che impara con l'esperienza

D. binomiale: probabilità /1

- Scriviamo la funzione di massa della distribuzione binomiale, ovvero l'espressione di P($X = k$) nel caso di n prove indipendenti e con identica probabilità di successo p
- $X = k$ indica che su n prove si sono verificati *esattamente* k successi e, di conseguenza, $n - k$ insuccessi
- Per le ipotesi fatte la probabilità che le prime k prove siano successi e le rimanenti n-k prove siano insuccessi è $p^k(1-p)^{(n-k)}$

$$\begin{array}{c} 1 \quad 1 \ \ldots \ 1 \qquad 0 \qquad 0 \quad \ldots \quad 0 \\ \underbrace{p\times p\times\ldots p}_{k\text{ volte}}\times\underbrace{(1-p)\times(1-p)\times\ldots(1-p)}_{n-k\text{ volte}} \end{array}$$

D. binomiale: probabilità /2

- Cambiando la posizione dei successi (cioè mescolando gli 0 e gli 1) si ottengono altre sequenze con *esattamente k* successi, ognuna delle quali ha la stessa probabilità $p^k(1-p)^{(n-k)}$
- Poiché siamo interessati al numero totale di successi, l'ordine in cui i successi si presentano è irrilevante

Ad es. con n=5 prove le sequenze con k=3 successi sono 10

11100	10011
11010	01110
11001	01101
10110	01011
10101	00111

Ogni sequenza ha probabilità $p^3(1-p)^2$

→ la probabilità di ottenere k=3 successi su n=5 prove è 10 $p^3(1-p)^2$

Ad es. con p=0.4 questa probabilità è 0.35

D. binomiale: probabilità /3

- Se le prove sono poche è facile scrivere, e quindi contare, tutte le sequenze con un numero fissato di successi
- Per fortuna, in generale non occorre scrivere tutte le sequenze perché il loro numero è pari alle combinazioni di n elementi di classe k

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

Binomio di Newton

$n! = n \times(n\text{-}1)\times(n\text{-}2) \times \ldots 3\times 2 \times 1$

D. binomiale: probabilità /4

- Vi sono dunque $\binom{n}{k}$ sequenze di k successi su n prove, ognuna delle quali ha probabilità $p^k(1\text{-}p)^{(n\text{-}k)}$. Pertanto la probabilità di ottenere k successi su n prove in qualunque ordine possibile è

$$P(X = k\,;n,p) = \binom{n}{k} p^k(1-p)^{n-k}$$

$n \in \{0,1,2\ldots,\}$

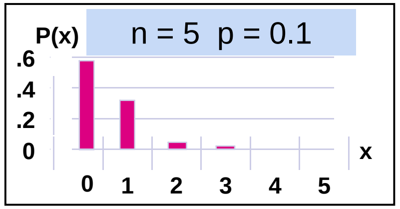
$k \in \{0,1,2,\ldots,n\}$ supporto

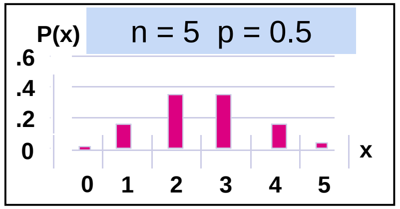
$p \in (0,1)$ spazio parametrico

Nelle applicazioni che vedremo n è un valore noto e quindi non viene considerato come parametro; l'unico parametro è p

D. binomiale: forma

- La forma della distribuzione binomiale dipende dai valori di *p* e *n*
  - $n = 5$ e $p = 0.1$ → asimmetrica positiva
  - $n = 5$ e $p = 0.5$ → simmetrica





D. binomiale: media e varianza

- Media e varianza possono essere calcolate applicando le rispettive formule (anche se il calcolo non è immediato)

$$E(X) = np$$ *n* volte la media Bernoulli

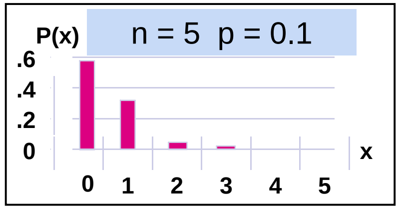
$$Var(X) = np(1-p)$$ *n* volte la varianza Bernoulli

D. binomiale: media e varianza /2

Esempi

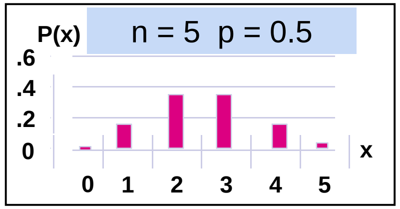
$$\mu = np = (5)(0.1) = 0.5$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{(5)(0.1)(1-0.1)} = 0.6708$$



$$\mu = np = (5)(0.5) = 2.5$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{(5)(0.5)(1-0.5)} = 1.118$$



D. binomiale: calcolare le probabilità /1

- Calcolo delle probabilità di ottenere un numero di successi maggiore di, minore di, compreso tra …
  - $P(X<2) = P(X=0)+P(X=1)$
  - $P(X\leq 2) = P(X=0)+P(X=1)+P(X=2)$
  - $P(X>2) = 1- P(X\leq 2)$
  - $P(1\leq X\leq 2) = P(X=1)+P(X=2)$
  - $P(1<X\leq 2) = P(X=2)$
  - $P(1<X<2) = P(\varnothing)= 0$
- Alcuni casi notevoli:
  - Zero successi: $P(X=0) = (1-p)^n$
  - Almeno un successo: $P(X>0) = 1-P(X=0) = 1-(1-p)^n$
  - Tutti successi: $P(X=n) = p^n$

D. binomiale: calcolare le probabilità /2

- Quando il numero di prove n è piccolo i calcoli si possono fare agevolmente con una calcolatrice tascabile
- Esistono tavole delle probabilità binomiali (non utilizzabili per la prova scritta)
- Microsoft Excel ha una funzione per calcolare le probabilità binomiali (la macro PHStat2 fornisce anche un'interfaccia grafica)

D. binomiale: esempio /1

- Si effettuano n = 4 prove indipendenti e con identica probabilità di successo p = 0.1 → $X \sim B(n = 4, p = 0.1)$

Calcolare la media (numero atteso di successi) e la varianza:

$$E(X) = np = 4 \times 0.1 = 0.4$$

$$Var(X) = np(1-p) = 4 \times 0.1 \times 0.9 = 0.36$$

Calcolare la probabilità di 3 successi, P(X=3):

$$P(X = 3) = \frac{4!}{3!(4-3)!} 0.1^3 (1-0.1)^{4-3}$$

$$= \frac{4 \times \cancel{3 \times 2 \times 1}}{\cancel{3 \times 2 \times 1} \times 1} \times 0.1^3 \times 0.9^1$$

$$= 4 \times 0.001 \times 0.9 \quad = \quad 0.0036$$

D. binomiale: esempio /2

Calcolare la probabilità di almeno 3 successi, Pr($X \geq 3$):

Poiché i due eventi sono incompatibili

$$P(X \geq 3) = P(X = 3 \cup X = 4) = P(X = 3) + P(X = 4)$$

$$= \frac{4!}{3!(4-3)!} 0.1^3 (1-0.1)^{4-3} + \frac{4!}{4!(4-4)!} 0.1^4 (1-0.1)^{4-4}$$

$$= \frac{4 \times \cancel{3 \times 2 \times 1}}{\cancel{3 \times 2 \times 1} \times 1} \times 0.1^3 \times 0.9^1 + \frac{\cancel{4 \times 3 \times 2 \times 1}}{\cancel{4 \times 3 \times 2 \times 1} \times 1} \times 0.1^4 \times 0.9^0$$

$$= 4 \times 0.001 \times 0.9 + 1 \times 0.0001 \times 1$$

$$= 0.0036 + 0.0001 = 0.0037$$

Perché 0! = 1 per definizione

D. binomiale: esempio /3

Calcolare la probabilità di meno di 3 successi, P(X < 3):

$$P(X < 3) = P(X = 0 \cup X = 1 \cup X = 2)$$

$$= P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X = 0) = \frac{4!}{0!(4-0)!} 0.1^0 (1-0.1)^{4-0} = 0.6561$$

$$P(X = 1) = \frac{4!}{1!(4-1)!} 0.1^1 (1-0.1)^{4-1} = 0.2916$$

$$P(X = 2) = \frac{4!}{2!(4-2)!} 0.1^2 (1-0.1)^{4-2} = 0.0486$$

$$P(X < 3) = 0.6561 + 0.2916 + 0.0486 = 0.9963$$

D. binomiale: esempio /4

Calcolare la probabilità di meno di 3 successi, P(X < 3):

Soluzione alternativa:

$$\begin{aligned}P(X<3) &= 1-P(X\geq 3)=1-P(X=3\cup X=4)\\ &=1-\big(P(X=3)+P(X=4)\big)\\ &=1-\big(0.0036+0.0001\big)\\ &=0.9963\end{aligned}$$

In questo caso la soluzione alternativa è conveniente perché richiede di calcolare la probabilità per 2 esiti (X=3 e X=4) anziché per 3 esiti (X=0, X=1 e X=2)

Proporzione di successi /1

- La media degli 0 e 1 è la proporzione di successi

$$\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{X}{n}=\frac{\text{numero di successi}}{\text{numero di prove}}$$

- Le n prove possono essere interpretate come un campione da una più ampia popolazione → si parla di media campionaria o proporzione campionaria di successi
- La proporzione di successi ha una distribuzione strettamente collegata alla binomiale poiché il numeratore $X=\Sigma_i X_i$ ha distribuzione B(n, p) mentre il denominatore n è una costante

Proporzione di successi /2

- La proporzione di successi è una v.a. con supporto S={0, 1/n, 2/n, … ,1} costituito da n+1 elementi
- Le probabilità si calcolano ricorrendo alla binomiale

X~B(n, p)

$$P(\bar{X}\leq c)=P\left(\frac{X}{n}\leq c\right)=P(X\leq cn)$$

Se cn non è intero va opportunamente arrotondato: es. $P(X \leq 2.7) = P(X \leq 2)$

- Valore atteso e varianza derivano dalla binomiale:

$$E(\bar{X})=E\left(\frac{X}{n}\right)=\frac{1}{n}E(X)=\frac{1}{n}np=p$$

$$Var(\bar{X})=Var\left(\frac{X}{n}\right)=\frac{1}{n^2}Var(X)=\frac{1}{n^2}np(1-p)=\frac{p(1-p)}{n}$$

Proporzione di successi /3

$$E(\bar{X})=p \qquad Var(\bar{X})=\frac{p(1-p)}{n}$$

- La proporzione di successi ha
  - valore atteso pari alla probabilità di successo p *qualunque sia il numero di prove*
  - varianza che dipende dalla probabilità di successo p tramite il numeratore p(1-p) e che (aspetto fondamentale!) dipende inversamente da n, cosicché *diminuisce al crescere del numero di prove*

LEGGE DEI GRANDI NUMERI: in una sequenza di prove bernoulliane, al crescere del numero di prove la proporzione osservata di successi converge alla probabilità di successo (e quindi diviene una stima sempre più precisa di tale probabilità)

Legge dei grandi numeri /1

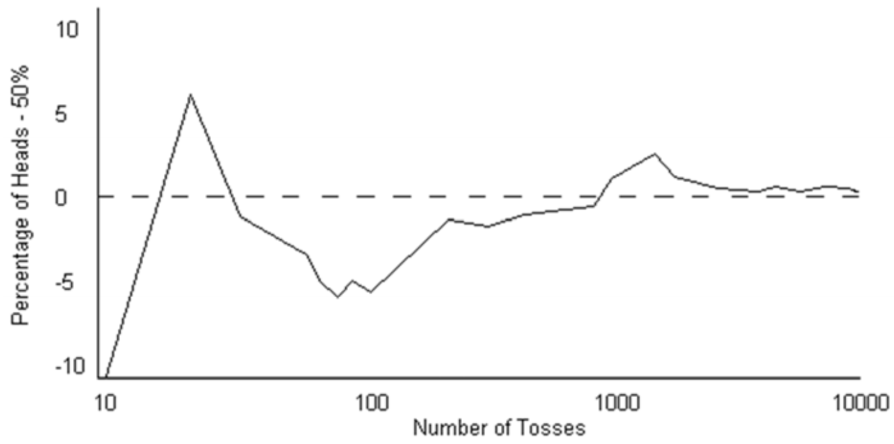
- Al crescere del numero di prove la distribuzione della proporzione di successi tende ad assumere una forma campanulare [per la precisione, vedremo più avanti che tende alla distribuzione normale] → per la regola empirica quasi tutta la probabilità è compresa tra $\mu-3\sigma$ e $\mu+3\sigma$
- Ad es. se p=0.5 si ha $\mu = 0.5$ e $\sigma^2 = 0.5\times0.5/n$ e quindi $\sigma = 0.5/\sqrt{n}$ → è quasi certo che la proporzione osservata di successi su n prove sia compresa in

$$(0.5 - 1.5/\sqrt{n}\ ,\ \ 0.5 + 1.5/\sqrt{n}\ )$$

  - n=100 (0.3500, 0.6500)
  - n=1000 (0.4526, 0.5474)
  - n=10000 (0.4850, 0.5150)
  - n=100000 (0.4953, 0.5047)

Legge dei grandi numeri /2

Legge empirica del caso per la probabilità di successo

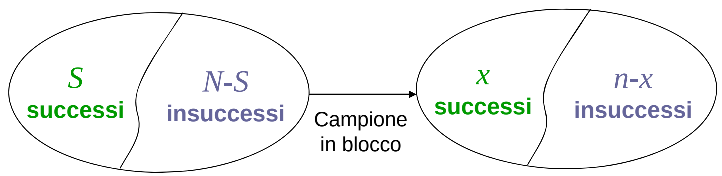


Legge dei grandi numeri /3

- L'esito di una specifica prova è imprevedibile, che sia la 1ª, la 10ª o la 100000ª (infatti le prove sono indipendenti e quindi quello che è successo prima è irrilevante)
- Tuttavia al crescere del numero di prove la proporzione di successi diventa sempre più prevedibile
- Dunque, mentre *a livello individuale* è difficile fare previsioni, *a livello aggregato* si possono fare previsioni molto accurate. Questo è il motivo per cui
  - il risultato delle singole giocate è imprevedibile, mentre il risultato del casinò è prevedibile con un piccolo margine di errore
  - il risultato del singolo assicurato è imprevedibile, mentre il risultato della compagnia di assicurazione è prevedibile con un piccolo margine di errore

La distribuzione Ipergeometrica /1

- n prove in un campione estratto da una popolazione finita di dimensione N
- Campione estratto **senza reintroduzione (in blocco)**
  → I risultati delle prove sono dipendenti
- Riguarda il calcolo della probabilità di "X" successi nel campione quando ci sono "S" successi nella popolazione



Popolazione (N elementi) Campione (n elementi)

La distribuzione Ipergeometrica /2

- La distribuzione ipergeometrica concerne la probabilità di x successi in un campione di dimensione n estratto (in modo casuale e senza reintroduzione) da una popolazione di dimensione N di cui S successi

$$P(X = x\,;N,S,n) = \frac{C_x^S C_{n-x}^{N-S}}{C_n^N} = \frac{\binom{S}{x}\binom{N-S}{n-x}}{\binom{N}{n}}$$

- Parametri:
  - N dimensione della popolazione
  - S numero di "successi" nella popolazione
  - n dimensione del campione
- Supporto: {0,1,…,min(n,S)}

La distribuzione Ipergeometrica /3

- Esempio: In un dipartimento ci sono 10 computer, 3 dei quali vengono selezionati per un controllo. Sappiamo che su 4 dei 10 computer sono stati installati dei programmi illegali. Qual è la probabilità che 2 dei 3 computer selezionati contengano programmi illegali?

N = 10 n = 3

S = 4 x = 2

$$P(X=2) = \frac{C_x^S C_{n-x}^{N-S}}{C_n^N} = \frac{C_2^4 C_1^6}{C_3^{10}} = \frac{(6)(6)}{120} = 0.3$$

La probabilità che 2 dei 3 computer selezionati contengano programmi illegali è 0.30, o 30%.

La distribuzione Ipergeometrica /4

- Denotando con p=S/N la *proporzione di successi nella popolazione* si ha

$$E(X) = np \qquad V(X) = np(1-p)\frac{N-n}{N-1}$$

- La v.a. ipergeometrica ha la stessa media della v.a. binomiale, ma varianza inferiore. Il fattore di riduzione della varianza è

$$\frac{N-n}{N-1} \cong 1 - \frac{n}{N}$$

dove n/N è la *frazione di campionamento* (se piccola, diciamo <5%, la contrazione della varianza è trascurabile → si può ignorare l'effetto del campionamento in blocco)

Distribuzione congiunta di due variabili aleatorie discrete

Distribuzione di prob. congiunta /1

- Una **distribuzione di probabilità congiunta** viene usata per esprimere la probabilità che
  - X assuma il valore x e, contemporaneamente,
  - Y assuma il valore y.
- E' una funzione in due variabili x e y

$$f_{XY}(x,y) = P(X = x \cap Y = y)$$

Distribuzione di prob. congiunta /2

- Esempio. Consideriamo una popolazione di 1000 studenti universitari con due caratteri binari X e Y
  - X=1 → "lo studente ha la maturità scientifica"
  - Y=1 → "lo studente ha superato l'esame di Analisi matematica entro un anno"

$X \setminus Y$	**0**	**1**	
0	90	660	750
1	10	240	250
	100	900	1000

Distribuzione di prob. congiunta /3

- (Segue esempio). Se estraiamo uno studente a caso le probabilità congiunte coincidono con le frequenze congiunte relative

$X \setminus Y$	**0**	**1**	
0	0.09	0.66	0.75
1	0.01	0.24	0.25
	0.10	0.90	1

$$P(X = 0 \cap Y = 1) = P(0,1) = 0.66$$

Distribuzione di prob. marginali /1

- Le **distribuzioni di probabilità marginali** riguardano una delle due variabili, ignorando l'altra

$$f_X(x) = P(X = x) \qquad f_Y(y) = P(Y = y)$$

- Si calcolano sommando le probabilità congiunte rispetto all'altra variabile

$$f_X(x) = \sum_y f_{XY}(x,y) \qquad f_Y(y) = \sum_x f_{XY}(x,y)$$

Distribuzione di prob. marginali /2

- (Segue esempio). La distribuzione marginale di Y si legge nei totali di colonna

Supporto della distribuzione marginale di Y

$X \setminus Y$	**0**	**1**	
0	0.09	0.66	0.75
1	0.01	0.24	0.25
	0.10	0.90	1

Probabilità marginali di Y

Distribuzioni di prob. condizionate /1

- Le distribuzioni di probabilità condizionate della variabile aleatoria Y esprimono le probabilità di Y condizionatamente ad uno specifico valore x di X:

$$\boxed{f_{Y|X}(y\,|\,x) = P(Y = y \,|\, X = x)}$$

Una distribuzione per ogni possibile valore x di X

- Si calcola dividendo la probabilità congiunta per la probabilità marginale di X (ricordarsi la definizione di probabilità condizionata!):

$$\boxed{f_{Y|X}(y\,|\,x) = \frac{f_{XY}(x,y)}{f_X(x)}}$$

In modo simile si definisce e si calcola la distribuzione di probabilità congiunta di X dato Y=y

Distribuzioni di prob. condizionate /2

(Segue esempio)

$X \setminus Y$	**0**	**1**	
0	0.12	0.88	1
1	0.04	0.96	1

Distribuzione di probabilità di aver superato l'esame di Analisi Matematica per gli studenti che …

… *non* hanno la maturità scientifica

Distribuzione di Y condizionata a X=0

0	**1**	
0.12	0.88	1

… hanno la maturità scientifica

Distribuzione di Y condizionata a X=1

0	**1**	
0.04	0.96	1

Indipendenza

- Due variabili aleatorie X e Y sono **indipendenti** se e solo se la loro distribuzione di probabilità congiunta è uguale al prodotto delle loro distribuzioni di probabilità marginali:

$$\boxed{f_{XY}(x,y) = f_X(x)\, f_Y(y)}$$

per tutte le possibili coppie di valori di x e y

- Estensione: k variabili aleatorie sono indipendenti se e solo se

$$\boxed{f_{X_1 X_2 \cdots X_k}(x_1, x_2, \cdots, x_k) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_k}(x_k)}$$

Covarianza

- La **covarianza** di due v.a. X e Y è il valore atteso del prodotto degli scarti dai rispettivi valori attesi

$$\boxed{Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]}$$

- Per variabili aleatorie discrete l'espressione è

$$\boxed{Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y) f_{XY}(x,y)}$$

- Un'espressione equivalente è

$$\boxed{Cov(X,Y) = E(XY) - \mu_x\mu_y = \sum_x \sum_y xy f_{XY}(x,y) - \mu_x\mu_y}$$

Covarianza e Indipendenza

- La covarianza misura la forza della ***relazione lineare*** tra due variabili aleatorie
  - Covarianza nulla ⇔ assenza di relazione lineare
- Se due variabili aleatorie sono statisticamente indipendenti non esiste alcuna relazione, né lineare né di altro tipo → la loro covarianza è nulla
  - Però se la covarianza è nulla non è detto che le variabili siano indipendenti

X e Y indipendenti $\Rightarrow$ Cov(X,Y) = 0

X e Y indipendenti $\not\Leftarrow$ Cov(X,Y) = 0

Correlazione

- La **correlazione** tra due v.a. X e Y è

$$\boxed{\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}}$$

- $\rho = 0 \Rightarrow$ non c'è relazione lineare tra X e Y
- $\rho > 0 \Rightarrow$ relazione lineare positiva tra X e Y
  - quando X assume valori alti (bassi) allora anche Y probabilmente assume valori alti (bassi)
  - $\rho = +1 \Rightarrow$ dipendenza lineare perfetta positiva
- $\rho < 0 \Rightarrow$ relazione lineare negativa tra X e Y
  - quando X assume valori alti (bassi) allora Y probabilmente assume valori bassi (alti)
  - $\rho = -1 \Rightarrow$ dipendenza lineare perfetta negativa

Combinazioni lineari: esempio

- Esempio: valutazione di un portafoglio titoli composto da
  - a azioni del titolo A
  - b azioni del titolo B
- X = quotazione del titolo A
- Y = quotazione del titolo B
- W = quotazione del portafoglio titoli
- La quotazione W del portafoglio è data dalla combinazione lineare

$$\boxed{W = aX + bY}$$

Combinazioni lineari: media e varianza

$$W = aX + bY$$

- La media di W è

$$\mu_W = E[W] = E[aX + bY] = a\mu_X + b\mu_Y$$

- La varianza di W è

$$\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X,Y)$$

- oppure, usando il coefficiente di correlazione

$$\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCorr(X,Y)\sigma_X\sigma_Y$$

Combinazioni lineari: casi speciali

- Siano X e Y due v.a. (dipendenti o indipendenti)

$$E(X+Y) = E(X) + E(Y) \qquad (a = 1, b = 1)$$
$$E(X-Y) = E(X) - E(Y) \qquad (a = 1, b = -1)$$

- Siano X e Y due v.a. indipendenti

$$Var(X+Y) = Var(X) + Var(Y) \qquad (a = 1, b = 1)$$
$$Var(X-Y) = Var(X) + Var(Y) \qquad (a = 1, b = -1)$$

Esempio: rendimento investimenti /1

Rendimento in migliaia di dollari per due investimenti

		Rendimento ($)	
P(x,y)	**Condiz. economica**	**Titolo prudente X**	**Titolo aggressivo Y**
0.2	Recessione	- 25	- 200
0.5	Economia Stabile	+ 50	+ 60
0.3	Economia in Espansione	+ 100	+ 350

$$\mu_X = (-25)(0.2) + (50)(0.5) + (100)(0.3) = 50$$
$$\mu_Y = (-200)(0.2) + (60)(0.5) + (350)(0.3) = 95$$
$$\sigma_X = \sqrt{(-25-50)^2(0.2) + (50-50)^2(0.5) + (100-50)^2(0.3)} = 43.30$$
$$\sigma_Y = \sqrt{(-200-95)^2(0.2) + (60-95)^2(0.5) + (350-95)^2(0.3)} = 193.71$$

Esempio: rendimento investimenti /2

		Rendimento ($)	
P(x,y)	**Condiz. economica**	**Titolo prudente X**	**Titolo aggressivo Y**
0.2	Recessione	- 25	- 200
0.5	Economia Stabile	+ 50	+ 60
0.3	Economia in Espansione	+ 100	+ 350

$$\mu_X = 50 \quad \mu_Y = 95 \quad \sigma_X = 43.30 \quad \sigma_Y = 193.71$$

$$Cov(X,Y) = (-25-50)(-200-95)(0.2) + (50-50)(60-95)(0.5) + (100-50)(350-95)(0.3) = 8250$$

$$Corr(X,Y) = \frac{8250}{43.3 \times 193.71} = 0.984$$

Esempio: rendimento investimenti /3

- Il fondo aggressivo ha un rendimento atteso più alto, ma molto più rischioso

$$\mu_Y = 95 \quad > \quad \mu_X = 50$$
$$\sigma_Y = 193.71 \quad > \quad \sigma_X = 43.30$$

- La correlazione 0.984 indica che tra i due investimenti c'è una forte relazione positiva e tendono a variare nella stessa direzione

Esempio: rendimento investimenti /4

- Immaginiamo un portafoglio W costituito per il 40% dal titolo X e per il 60% dal titolo Y: $\boxed{W = 0.4X + 0.6Y}$

$$\mu_X = 50 \quad \mu_Y = 95 \quad \sigma_X = 43.30 \quad \sigma_Y = 193.71 \quad \sigma_{XY} = 8250$$

$$\boxed{\mu_W = (0.4)(50) + (0.6)(95) = 77}$$

$$\sigma_W = \sqrt{(0.4)^2(43.30)^2 + (0.6)^2(193.21)^2 + 2(0.4)(0.6)(8250)} = 133.04$$

  - Il rendimento del portafoglio W ha media e varianza con valori intermedi rispetto a quelli dei titoli X e Y presi singolarmente