

STATISTICA

I modelli probabilistici

I modelli probabilistici

Dal nostro

11. Sapete che un'urna contiene

$X : \Omega \rightarrow \{0,1,2, \dots, 100\}$

Più che l'esatto risultato dell'esperimento (cioè, l'esatta sequenza dei colori delle 100 palline estratte) ci interessa una sua **funzione**, cioè il **numero di palline rosse** (trascurando la loro posizione nella sequenza).

Variabile casuale (o aleatoria) : una funzione degli esiti elementari dell'esperimento.

Distribuzione di una variabile casuale : l'elenco dei valori che la funzione può assumere con le relative probabilità.

$P(kR e (100 - k)N) = \binom{100}{k} \times \left(\frac{8}{10}\right)^k \times \left(\frac{2}{10}\right)^{100-k}$

Altri esempi $X : \Omega \rightarrow \{1,2,3,4,5,6\}$

- Numero di volte in cui esce T in n lanci
- Numero di lanci prima che esce T

- Somma del punteggio
- Uscita delle coppie di 6
- Uscita di punteggi uguali

Altri esempi $X : \Omega \rightarrow \{1,2,3,4,5,6\}$

- Numero di volte in cui esce T in n lanci
- Numero di lanci prima che esce T

- Somma del punteggio
- Uscita delle coppie di 6
- Uscita di punteggi uguali

- PIU
- Numero di morti ferro causati dal PIU

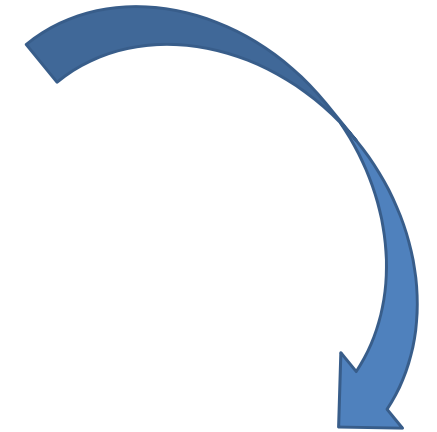
Altri esempi $X : \Omega \rightarrow \{1,2,3,4,5,6\}$

- Numero di volte in cui esce T in n lanci
- Numero di lanci prima che esce T

- Somma del punteggio
- Uscita delle coppie di 6
- Uscita di punteggi uguali

- PIU
- Numero di morti ferro causati dal PIU

- Percentuale di clienti soddisfatti



Altri esempi $X : \Omega \rightarrow [0,100] \%$

Sono la proprietaria della salumeria di Montecarlo, più piccola piazza d'Italia, soltanto 30 ...

GLIELO CHIEDO A TUTTI!!!

- Percentuale di clienti soddisfatti

Altri esempi $X : \Omega \rightarrow [0,100] \%$

Sono la proprietaria di un B&B a Roma ...

GLIELO CHIEDO A TUTTI!!!!!!

- Percentuale di clienti soddisfatti

I modelli probabilistici



- Numero di volte in cui esce T in n lanci
- Numero di lanci *prima* che esca T

- Somma del punteggio
- Uscita della coppia di 6
- Uscita di punteggi uguali



MONETA/DADO
EQUILIBRATI



DISTRIBUZIONE
NOTA

I modelli probabilistici



- Numero di volte in cui esce T in n lanci
- Numero di lanci *prima* che esca T

- Somma del punteggio
- Uscita della coppia di 6
- Uscita di punteggi uguali

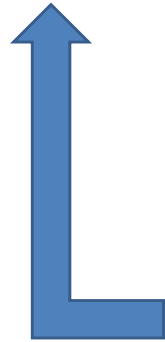


MONETA/DADO
FORSE NON
EQUILIBRATI



DISTRIBUZIONE
NON NOTA

I modelli probabilistici

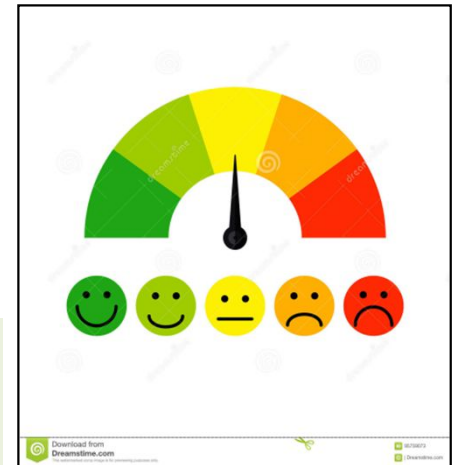


DISTRIBUZIONE
NON NOTA

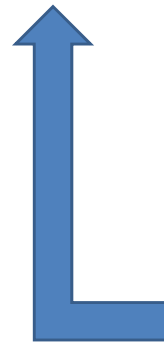


- PM_{10}
- Numero di morti l'anno causati dal PM_{10}

- Percentuale di clienti soddisfatti



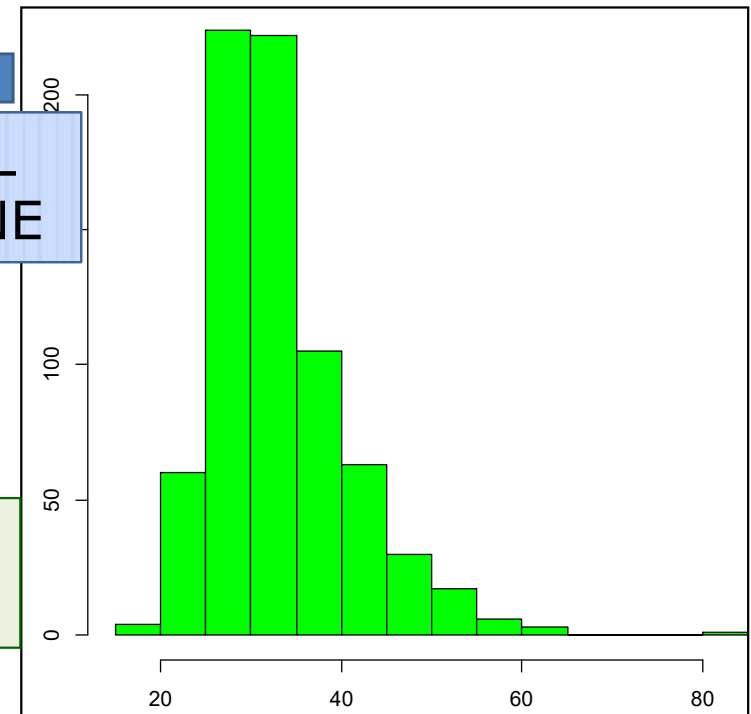
I modelli probabilistici



INFO DAL
CAMPIONE

DISTRIBUZIONE
NON NOTA

```
> summary(x)
  Min. 1st Qu. Median Mean 3rd Qu. Max.
 17.60 27.80 32.00 33.21 36.60 87.90
```



- PM_{10}
- Numero di morti l'anno causati dal PM_{10}

- Percentuale di clienti soddisfatti



I modelli probabilistici



PARAMETRICA!

Si fanno ipotesi sulla forma della distribuzione (per esempio, che sia Gaussiana), **forma** che è completamente **specificata da alcuni** (pochi) **parametri** (e.g, la media e la varianza).

Detective, romanziere o *statistico*?

Per risolvere i casi complicati bisogna essere capaci di **costruire una storia**, partendo dagli **indizi disponibili**, che contenga una **spiegazione plausibile** di tutti gli elementi che abbiamo. Ci vuole una certa dose di fantasia ed è un lavoro simile a quello di uno scrittore. Una volta costruita questa storia che è, in sostanza, un'**ipotesi** su come potrebbero essersi svolti i fatti, bisogna andare alla ricerca delle **conferme**.

G. Carofiglio
Una mutevole verità

Detective, romanziere o statistico?

LINGUAGGIO
della
PROBABILITA'

Per risolvere i casi complicati bisogna essere capaci di **costruire una storia**, partendo dagli **indizi disponibili**, che contenga una **spiegazione plausibile** di tutti gli elementi che abbiamo. Chi vuole una certa dose di fantasia ed è un po' simile a quello di uno scrittore. Una volta costruita questa storia che è, in sostanza, un **INFERENZA** come potrebbero essersi svolti i fatti, bisogna andare alla ricerca delle **conferme**.

DATI

MODELLI

INFERENZA

(G. Carofiglio)

Detective, romanziere e statistico?

**LINGUAGGIO
della
PROBABILITA'**

« Per risolvere i casi complessi bisogna essere

capaci di costruire

indizi disponibili

plausibile di tutti gli elementi che abbiamo

vuole una certa dose di fantasia ed è un

simile a quello di uno scrittore. Una volta costruita

questa storia che è, in sostanza, un' **INFERENZA**

come potrebbero essersi svolti i fatti, bisogna

andare alla ricerca delle **conferme**.»

**DAL CAMPIONE (I DATI) ALLA
POPOLAZIONE, TRAMITE I
MODELLI**

DATI

MODELLI

INFERENZA

Il campione

Sottoinsieme della **popolazione** di interesse, usato per trarre informazioni sull'intera popolazione.

⇒ deve essere ***rappresentativo*** della popolazione

Il campione

Sottoinsieme della **popolazione** di interesse, usato per trarre informazioni sull'intera popolazione.

⇒ deve essere ***rappresentativo*** della popolazione

tutte le parti della popolazione devono avere **uguale probabilità** di fare parte del campione.

Esempio dell'età: scelgo tra i nominativi sulla guida telefonica della città?

Il campione

Sottoinsieme della **popolazione** di interesse, usato per trarre informazioni sull'intera popolazione.

⇒ deve essere ***rappresentativo*** della popolazione

tutte le parti della popolazione devono avere **uguale probabilità** di fare parte del campione.

Esempio dell'età: scelgo tra i nominativi sulla guida telefonica della città? Meglio dall'anagrafe, per non escludere i bambini, per esempio.

Il campione

Sottoinsieme della **popolazione** di interesse, usato per trarre informazioni sull'intera popolazione.

⇒ deve essere ***rappresentativo della popolazione***

tutte le parti della popolazione devono avere **uguale probabilità** di fare parte del campione.

Esempio dell'età: scelgo tra i nominativi sulla guida telefonica della città? Meglio dall'anagrafe, per non escludere i bambini, per esempio.

scelta **completamente casuale** dalla popolazione:
Campione casuale (semplice) di k membri se questi sono scelti in modo che tutte le possibili scelte dei k membri dalla popolazione di interesse sono ugualmente probabili.

Il campione

Sottoinsieme della **popolazione** di interesse, usato per trarre informazioni sull'intera popolazione.

⇒ deve essere rappresentativa della popolazione

tutte le
p

sono avere **uguale**
el campione.

Esempio del
telefonica de
escludere i b

inativi sulla guida
nagrafe, per non



scelta **completamente casuale** dalla popolazione:
Campione casuale (semplice) di k membri se questi sono scelti in modo che tutte le possibile scelte dei k membri dalla popolazione di interesse sono ugualmente probabili.

Il campione

Scegliamo a caso...



**con o senza
reimmissione?**

A meno che non sia diversamente indicato,

quando diciamo che scegliamo **un campione a caso** da una popolazione, faremo come se le estrazioni fossero **con reimmissione**.

Le estrazioni in questo modo sono **indipendenti**: sapere l'esito della prima estrazione non influenza l'esito della seconda, etc.

Dal campione **alla popolazione**

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. **Di 14 coppie scelte a caso** su cui è stata messa in atto, 13 hanno avuto una femmina.

Possiamo affermare che tale tecnica aumenta la probabilità di avere una femmina per *qualunque coppia*?

Dal campione **alla popolazione**

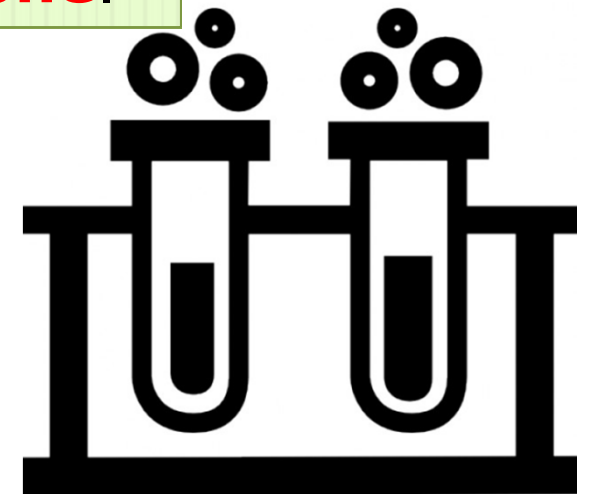
Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata, in 5 punti dell'acquedotto, la quantità presente nell'acqua di arsenico, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\text{g/L}$.

Sapendo che la legge impone una soglia massima per l'arsenico di 10 $\mu\text{g/L}$, **si può affermare che *l'acqua potabile di Milano* non è a norma di legge?**

Dal campione **alla popolazione**



Ci serve un modo per capire se quello che vediamo sul **campione** vale anche sull'intera **popolazione**.



Dal campione **alla popolazione**



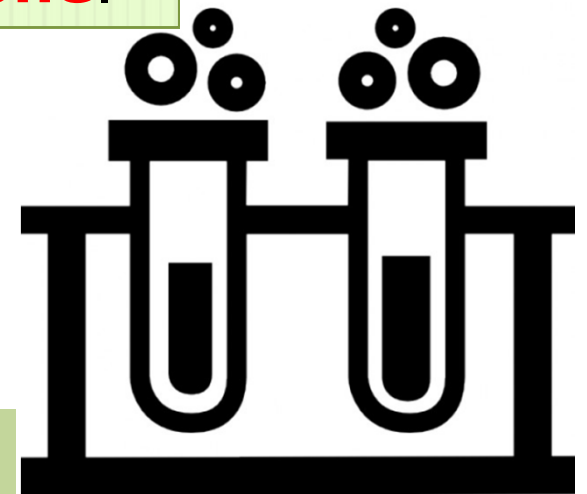
Tutti
i possibili
risultati

modello probabilistico

Ci serve un modo per capire se
quello che vediamo sul **campione**
vale anche sull'intera **popolazione**.

modello probabilistico

Uno dei tanti
possibili risultati



Dal campione **alla popolazione**



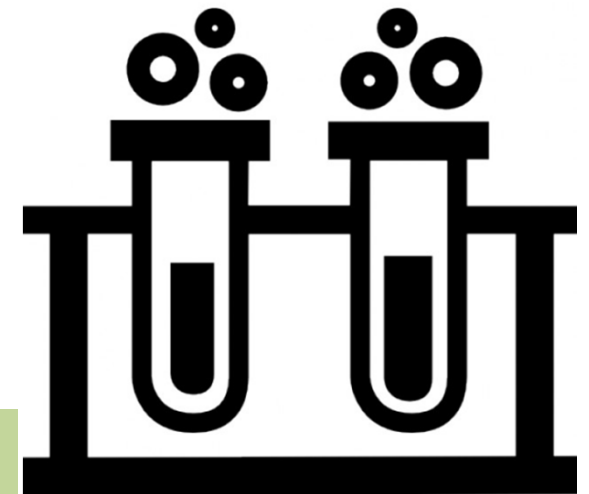
Tutti
i possibili
risultati

modello probabilistico

Il **modello** è una descrizione del meccanismo con cui si sviluppa il fenomeno (e, quindi, con cui sono prodotti i dati nel campione):

modello probabilistico

Uno dei tanti
possibili risultati



Dal campione **alla popolazione**



Tutti
i possibili
risultati

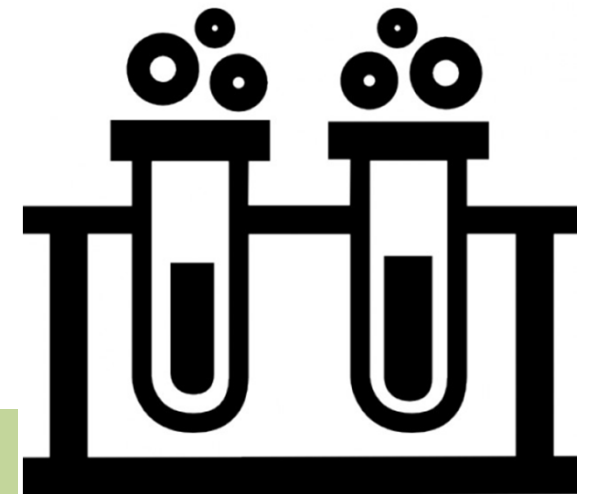
modello probabilistico

Il **modello** è una descrizione del meccanismo con cui si sviluppa il fenomeno (e, quindi, con cui sono prodotti i dati nel campione):

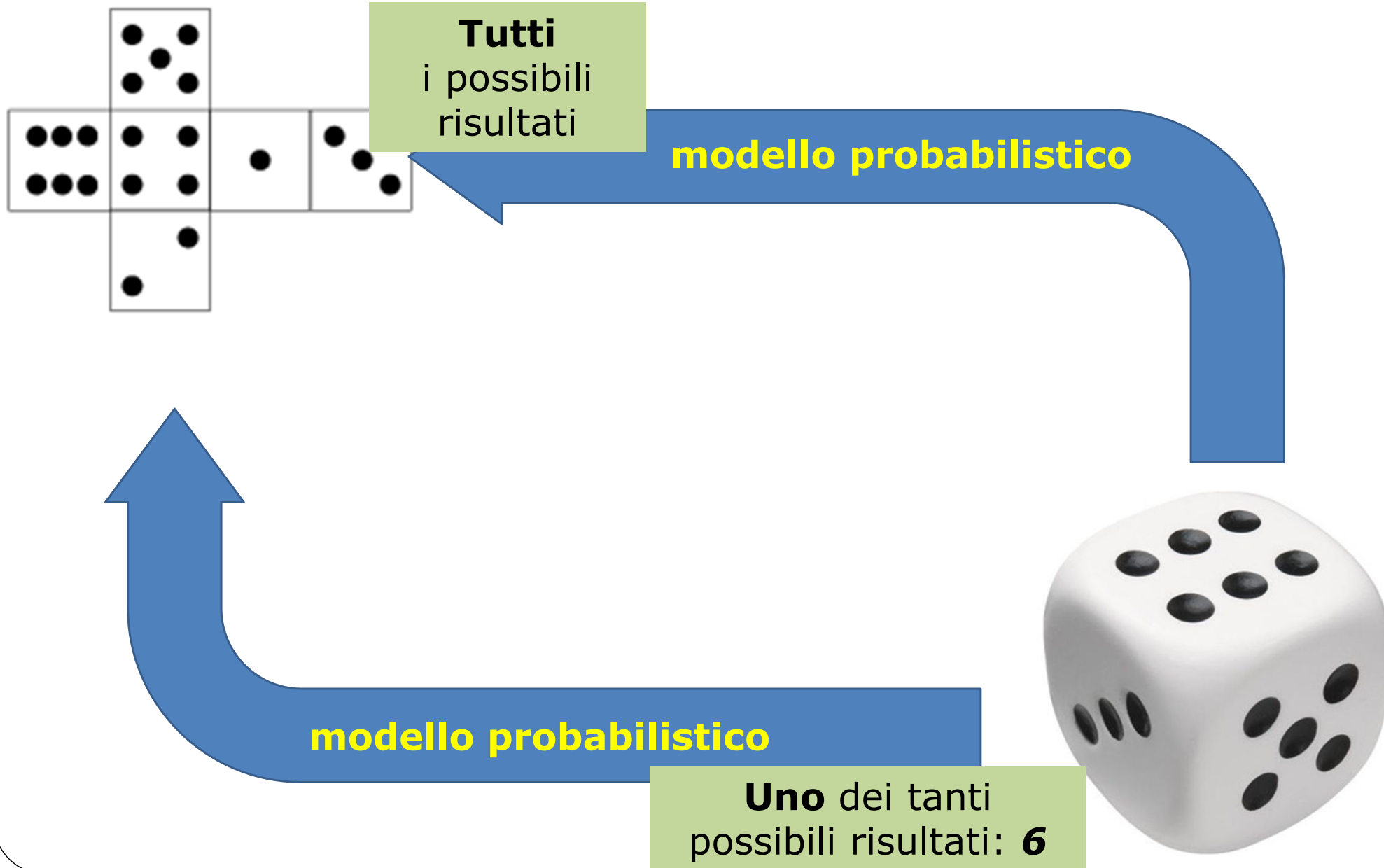
è in grado di darci informazioni sulle osservazioni *prima* di estrarre il campione (**osservazioni potenziali**)

modello probabilistico

Uno dei tanti
possibili risultati



Dal campione **alla popolazione**



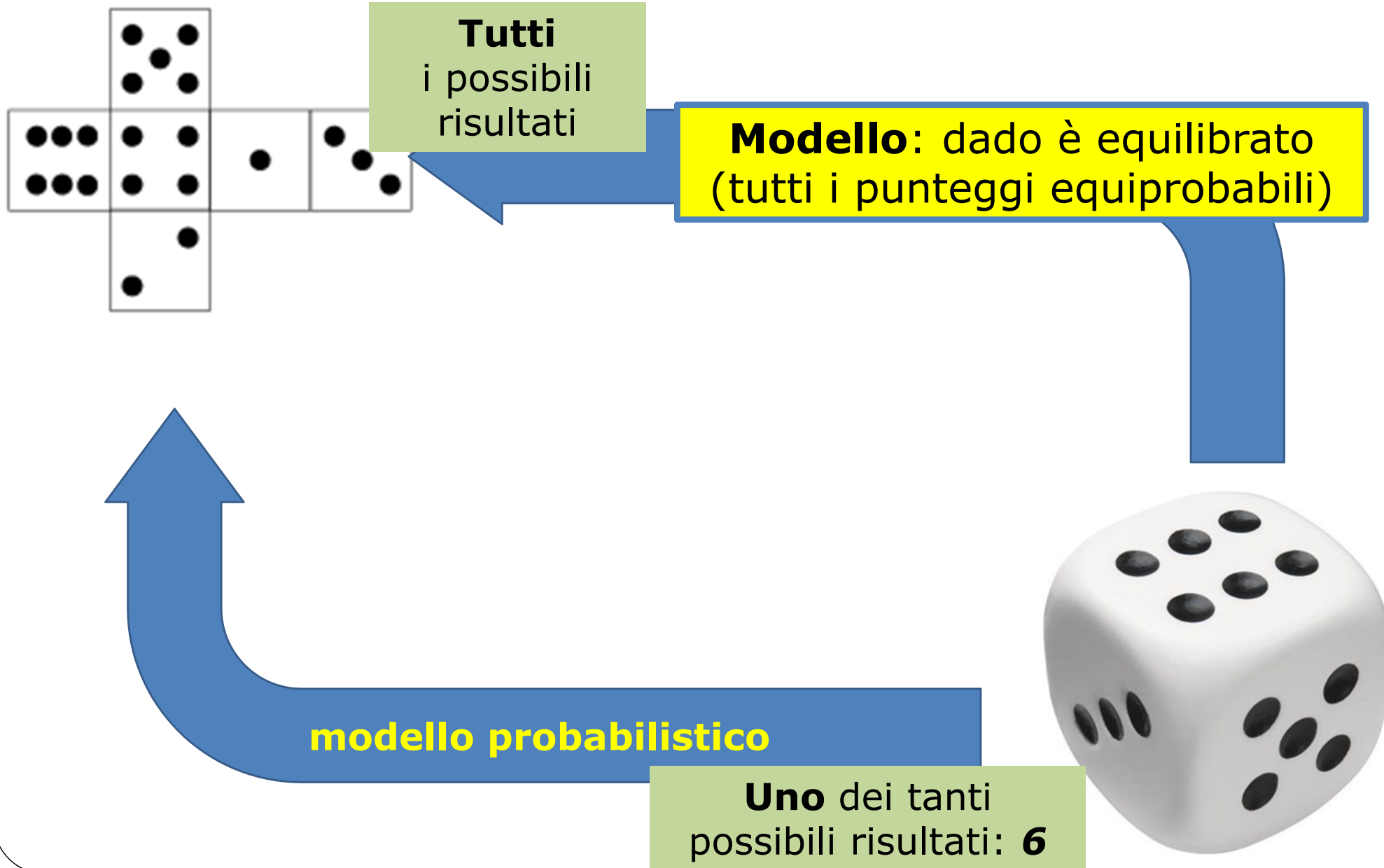
Tutti
i possibili
risultati

modello probabilistico

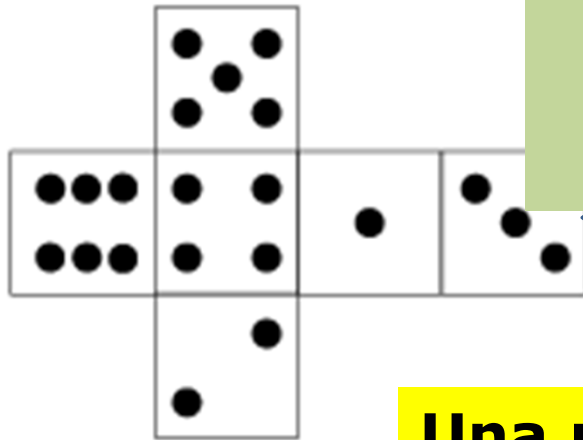
modello probabilistico

Uno dei tanti
possibili risultati: **6**

Dal campione **alla popolazione**



Dal campione **alla popolazione**



Tutti
i possibili
risultati

Modello: dado è equilibrato
(tutti i punteggi equiprobabili)

Una potenziale osservaz.: X
variabile casuale che assume i valori
1, 2, 3, 4, 5, 6 e tale che

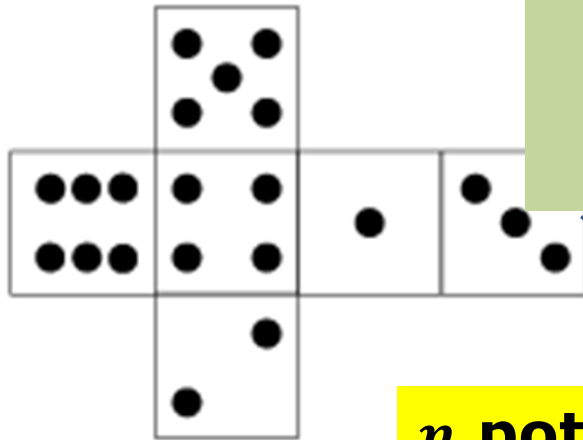
$$P(X = k) = \frac{1}{6}, \quad k = 1, \dots, 6$$

modello probabilistico

Uno dei tanti
possibili risultati: **6**



Dal campione **alla popolazione**



Tutti
i possibili
risultati

Modello: dado è equilibrato
(tutti i punteggi equiprobabili)

n potenziali osservaz.: X_1, X_2, \dots, X_n
indipendenti (come i lanci) **e** tutte
identiche con distribuzione

$$P(X_i = k) = \frac{1}{6}, \quad k = 1, \dots, 6$$

modello probabilistico

Uno dei tanti
possibili risultati: **6**



Dal campione **alla popolazione**

Modello: dado è equilibrato (tutti i punteggi equiprobabili)

Una potenziale osservaz.: X variabile casuale che assume i valori 1, 2, 3, 4, 5, 6 e tale che

$$P(X = k) = \frac{1}{6}, \quad k = 1, \dots, 6$$



n potenziali osservaz.: X_1, X_2, \dots, X_n indipendenti (come i lanci) e tutte identiche con distribuzione

$$P(X_i = k) = \frac{1}{6}, \quad k = 1, \dots, 6$$

Campione casuale osservato : esito di n lanci: $\{x_1, x_2, \dots, x_n\} = \{1, 3, 2, 1, 1, 6, 2, 3, 4, 4, \dots, 6\}$

Variabili casuali discrete

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{x_1, x_2, \dots, x_N\}, \quad N \leq +\infty$$

distribuzione di X :

$$P(X = x_i) \text{ per tutti gli } x_i$$

$$0 \leq P(X = x_i) \leq 1$$

$$\sum_{i=1}^N P(X = x_i) = 1$$

Variabile casuale **uniforme**

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{1, 2, 3, 4, 5, 6\}$$

distribuzione di X :

$$P(X = k) = 1/6, \quad k = 1, \dots, 6$$

$$0 \leq P(X = x_i) \leq 1$$

$$\sum_{i=1}^N P(X = x_i) = 1$$

Variabile casuale **bernoulliana**

**Modello
probabilistico
discreto:**

variabile casuale X :

$X : (\dots) \rightarrow \{0, 1\}$ (numero di *successi*
in una singola prova)

distribuzione di X :

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

$$0 \leq P(X = x_i) \leq 1$$

$$\sum_{i=1}^N P(X = x_i) = 1$$

Variabile casuale **bernoulliana**

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{0, 1\}$$

(numero di *successi*
in una singola prova)

distribuzione di X :

$$P(X = 1) = 0.5, \quad P(X = 0) = 0.5$$



Variabile casuale **bernoulliana**

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{0, 1\}$$

(numero di *successi*
in una singola prova)

distribuzione di X :

$$P(X = 1) = 1/6, \quad P(X = 0) = 5/6$$



La variabile bernoulliana

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. Di 14 coppie scelte a caso su cui è stata messa in atto, 13 hanno avuto una femmina.

X variabile casuale che dice se il neonato è F: 1=sì, 0=no

$$P(X = 1) = P(X = 0) = \frac{1}{2}, \textit{normalmente}$$

X_1, X_2, \dots, X_{14} variabili casuali indipendenti ed identiche a X che dicono se il neonato di ciascuna coppia è (1) o meno (0) F

Campione casuale osservato: esito di 14 parti: $\{x_1, x_2, \dots, x_{14}\} = \{1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1\}$

La variabile bernoulliana

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. Di 14 coppie scelte a caso su cui è stata messa in atto, 13 hanno avuto una femmina.

X variabile casuale che dice il sesso del neonato: 1=F, 0=M

$$P(X = 1) = p, \quad P(X = 0) = q = 1 - p, \quad \text{con qualche tecnica}$$

X è una variabile bernoulliana di parametro $p \in (0,1)$

$X_1 + X_2 + \dots + X_{14}$ conta il numero di F (*successi*) nei 14 parti (*prove*)

$$(1 + 1 + \dots + 1 + 0 = 13)$$

$S_n = X_1 + X_2 + \dots + X_n$ è una variabile **Binomiale**: $Bin(n, p)$

se possiamo supporre che le variabili siano **indipendenti ed identiche**

Variabile casuale **Binomiale**

$S_n = X_1 + X_2 + \dots + X_n$ è una variabile **Binomiale**: $Bin(n, p)$

**Modello
probabilistico
discreto:**

$S_n: (\dots) \rightarrow \{0, 1, 2, \dots, n\}$ (numero di *successi* in n prove indipendenti ed identiche)

distribuzione di S_n :

$$P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$P(S_n = k) = \frac{n!}{k! (n - k)!} p^k (1 - p)^{n-k}$$

Variabile casuale **Binomiale**

$S_n = X_1 + X_2 + \dots + X_n$ è una variabile **Binomiale**: $Bin(n, p)$

**Modello
probabilistico
discreto:**

$S_n: (\dots) \rightarrow \{0, 1, 2, \dots, n\}$ (numero di *successi* in n prove indipendenti ed identiche)

distribuzione di S_n :

$$P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$P(S_n = k) = \frac{n!}{k! (n - k)!} p^k (1 - p)^{n-k}$$

Se $p = 0.5, n = 14 \Rightarrow P(S_{14} = 13) = \frac{14!}{13!(14-13)!} 0.5^{13} (1 - 0.5)^{14-13} = 14 \times 0.5^{14} =$
 $0.0008544922 \approx 0.001$

Variabile casuale **Binomiale**

$S_n = X_1 + X_2 + \dots + X_n$ è una variabile **Binomiale**: $Bin(n, p)$

**Modello
probabilistico
discreto:**

$S_n: (\dots) \rightarrow \{0, 1, 2, \dots, n\}$ (numero di *successi* in n prove indipendenti ed identiche)

distribuzione di S_n :

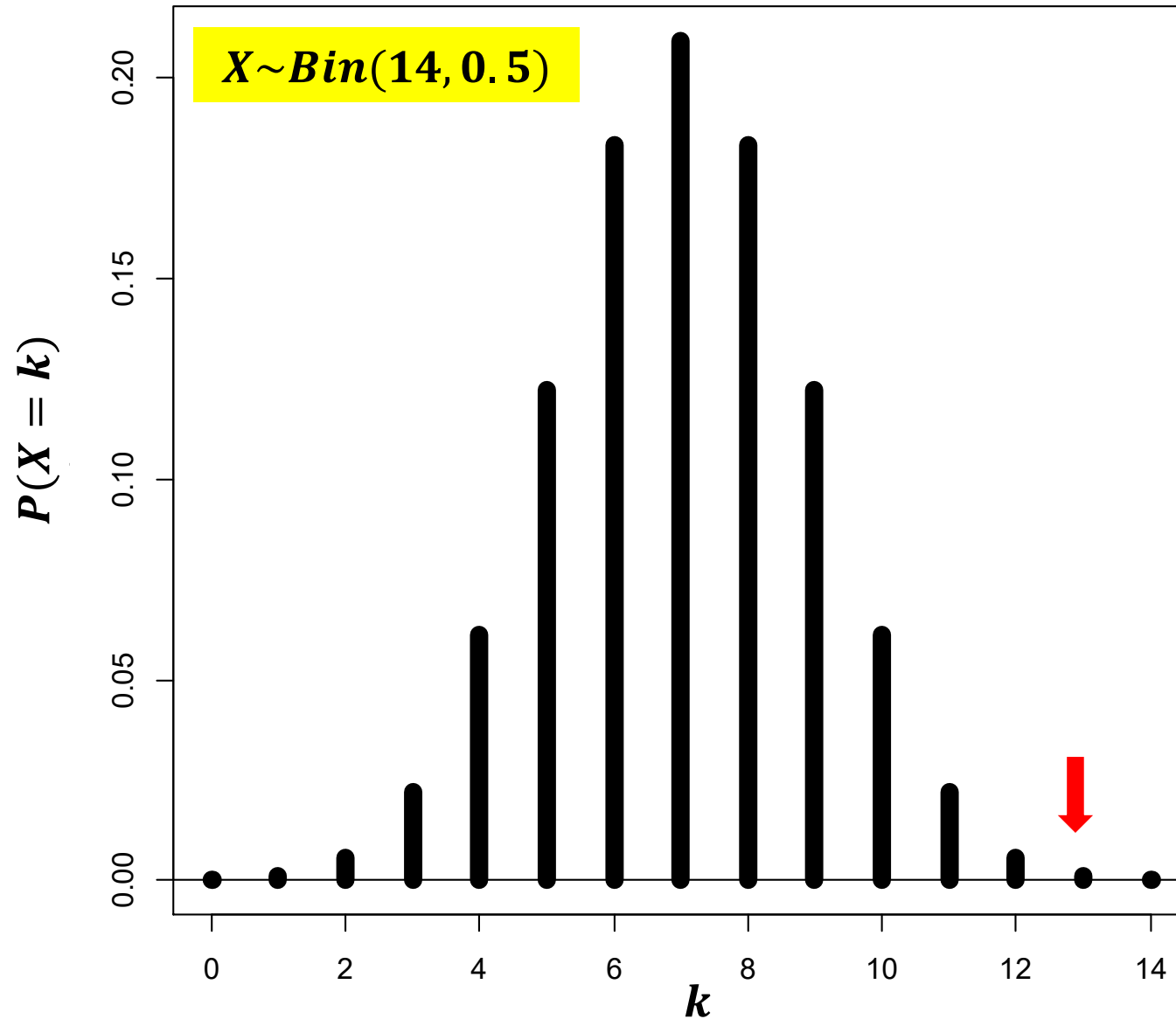
$$P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$P(S_n = k) = \frac{n!}{k! (n - k)!} p^k (1 - p)^{n-k}$$

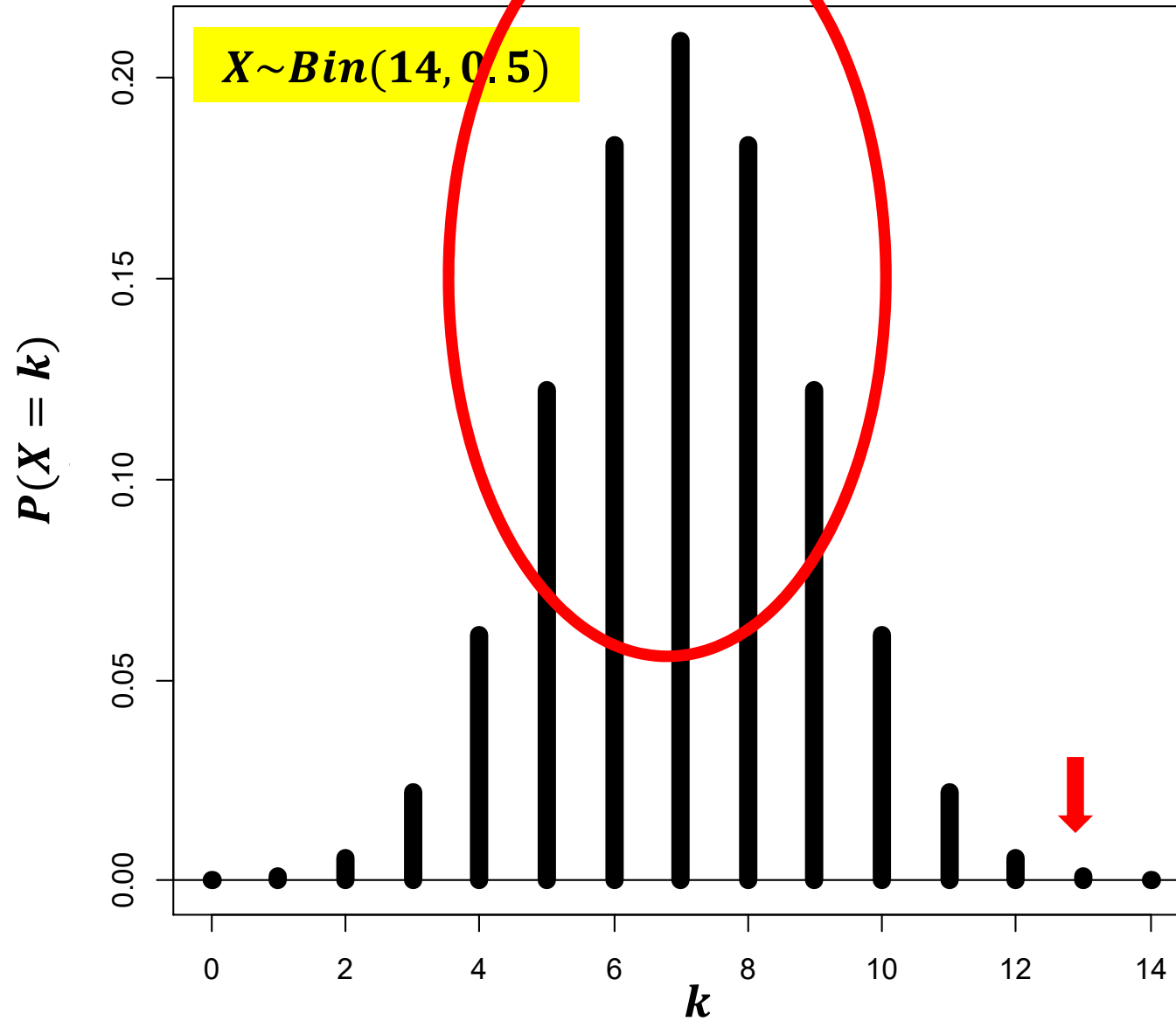
$$\text{Se } p = 0.5, n = 14 \Rightarrow P(S_{14} = 13) = \frac{14!}{13!(14-13)!} 0.5^{13} (1 - 0.5)^{14-13} = 14 \times 0.5^{14} = 0.0008544922 \approx 0.001$$

Se la tecnica fosse inefficace, sarebbe altamente improbabile avere 13 F su 14 parti!

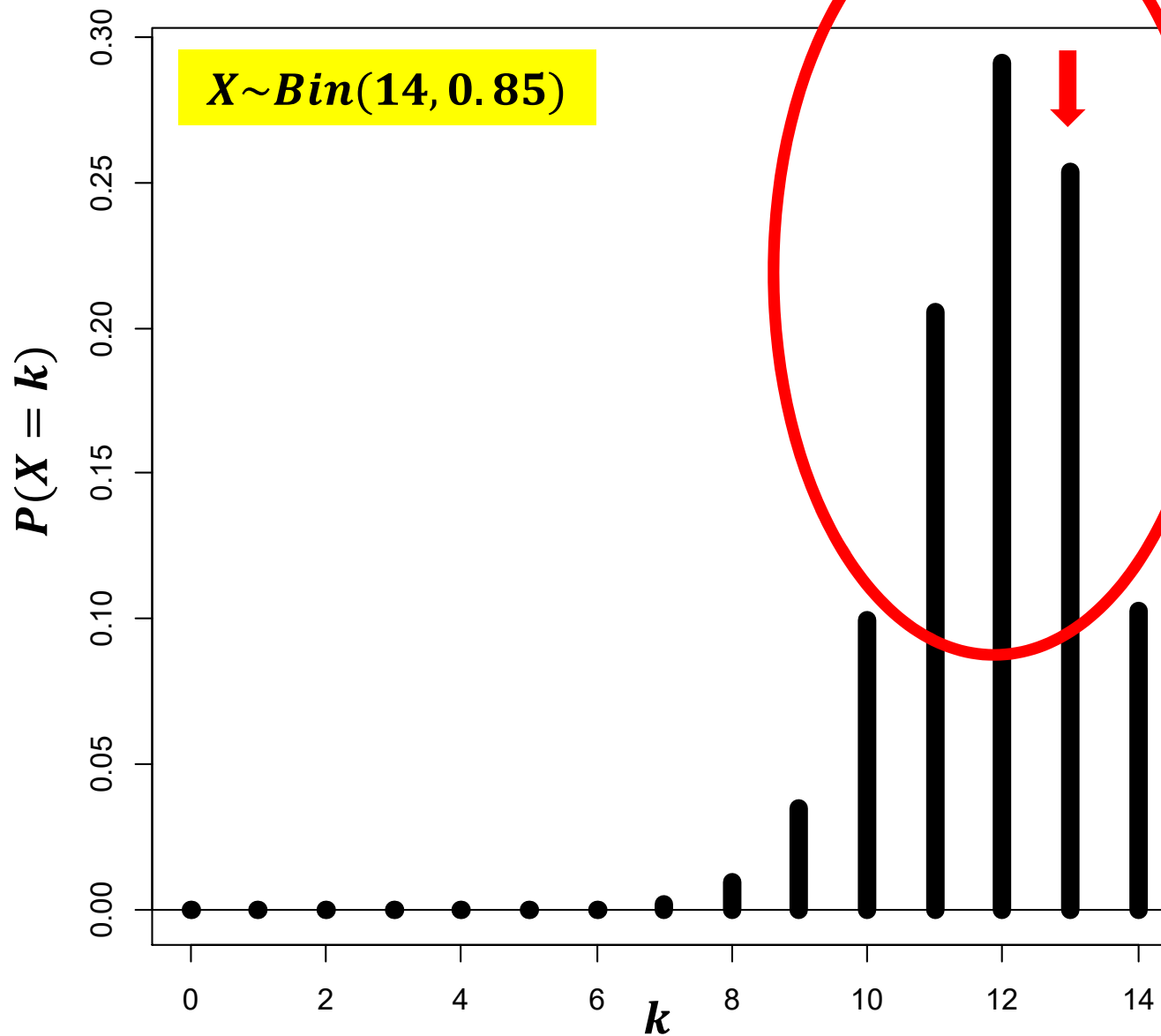
Binomiale(14, 0.5)



Binomiale(14, 0.5)



Binomiale(14, 0.85)



Esempio 1

Un test di ammissione ad un corso molto selettivo consiste di 15 domande a risposta multipla, con 3 risposte ciascuna di cui 1 sola corretta. Si è ammessi al corso solo rispondendo correttamente ad almeno 13 domande. Calcolare la probabilità di essere ammessi al corso rispondendo a caso a ciascuna domanda.

Esempio 1

Un test di ammissione ad un corso molto selettivo consiste di 15 domande a risposta multipla, con 3 risposte ciascuna di cui 1 sola corretta. Si è ammessi al corso solo rispondendo correttamente ad almeno 13 domande. Calcolare la probabilità di essere ammessi al corso rispondendo **a caso** a ciascuna domanda.

4. Lanciate due volte una moneta non truccata: dopo aver visto l'esito del primo lancio, potete prevedere l'esito del secondo?

SI	NO	NON SAPREI
----	----	------------

9. Sapete che un'urna contiene 10 palline: 5 rosse e 5 nere, di uguali dimensioni e materiale. Viene estratta, scegliendo a caso, una pallina: la pallina estratta è rossa. Lasciate fuori dall'urna la pallina rossa estratta, e vi accingete a fare un'altra estrazione casuale. Allora:

E' più probabile che la seconda pallina sia nera	E' più probabile che la seconda pallina sia rossa	La seconda pallina potrà essere rossa o nera con uguale probabilità.
--	---	--

Esempio 1

Un test di ammissione ad un corso molto selettivo consiste di 15 domande a risposta multipla, con 3 risposte ciascuna di cui 1 sola corretta. Si è ammessi al corso solo rispondendo correttamente ad almeno 13 domande. Calcolare la probabilità di essere ammessi al corso rispondendo **a caso** a ciascuna domanda.

X = numero di risposte esatte

risposte **indipendenti**,
la prob. di indovinare è la stessa ad ogni domanda

4. Lanciate due volte una moneta non truccata. dopo aver visto l'esito del primo lancio, potete prevedere l'esito del secondo?

SI	NO	NON SAPREI
----	----	------------

9. Sapete che un'urna contiene 10 palline: 5 rosse e 5 nere, di uguali dimensioni e materiale. Viene estratta, scegliendo a caso, una pallina: la pallina estratta è rossa. Lasciate fuori dall'urna la pallina rossa estratta, e vi accingete a fare un'altra estrazione casuale. Allora:

E' più probabile che la seconda pallina sia nera	E' più probabile che la seconda pallina sia rossa	La seconda pallina potrà essere rossa o nera con uguale probabilità.
--	---	--

Esempio 1

Un test di ammissione ad un corso molto selettivo consiste di 15 domande a risposta multipla, con 3 risposte ciascuna di cui 1 sola corretta. Si è ammessi al corso solo rispondendo correttamente ad almeno 13 domande. Calcolare la probabilità di essere ammessi al corso rispondendo **a caso** a ciascuna domanda.

X = numero di risposte esatte

$$P(X \geq 13) = ?$$

risposte **indipendenti**,
la prob. di indovinare è la stessa ad ogni domanda

Esempio 1

Un test di ammissione ad un corso molto selettivo consiste di 15 domande a risposta multipla, con 3 risposte ciascuna di cui 1 sola corretta. Si è ammessi al corso solo rispondendo correttamente ad almeno 13 domande. Calcolare la probabilità di essere ammessi al corso rispondendo **a caso** a ciascuna domanda.

X = numero di risposte esatte

$$P(X \geq 13) = ?$$

$$X \sim \text{Binomiale}\left(15, \frac{1}{3}\right)$$

risposte **indipendenti**,
la prob. di indovinare è la stessa ad ogni domanda

$$P(X \geq 13) = P(X = 13) + P(X = 14) + P(X = 15) = 1 - P(X \leq 12)$$



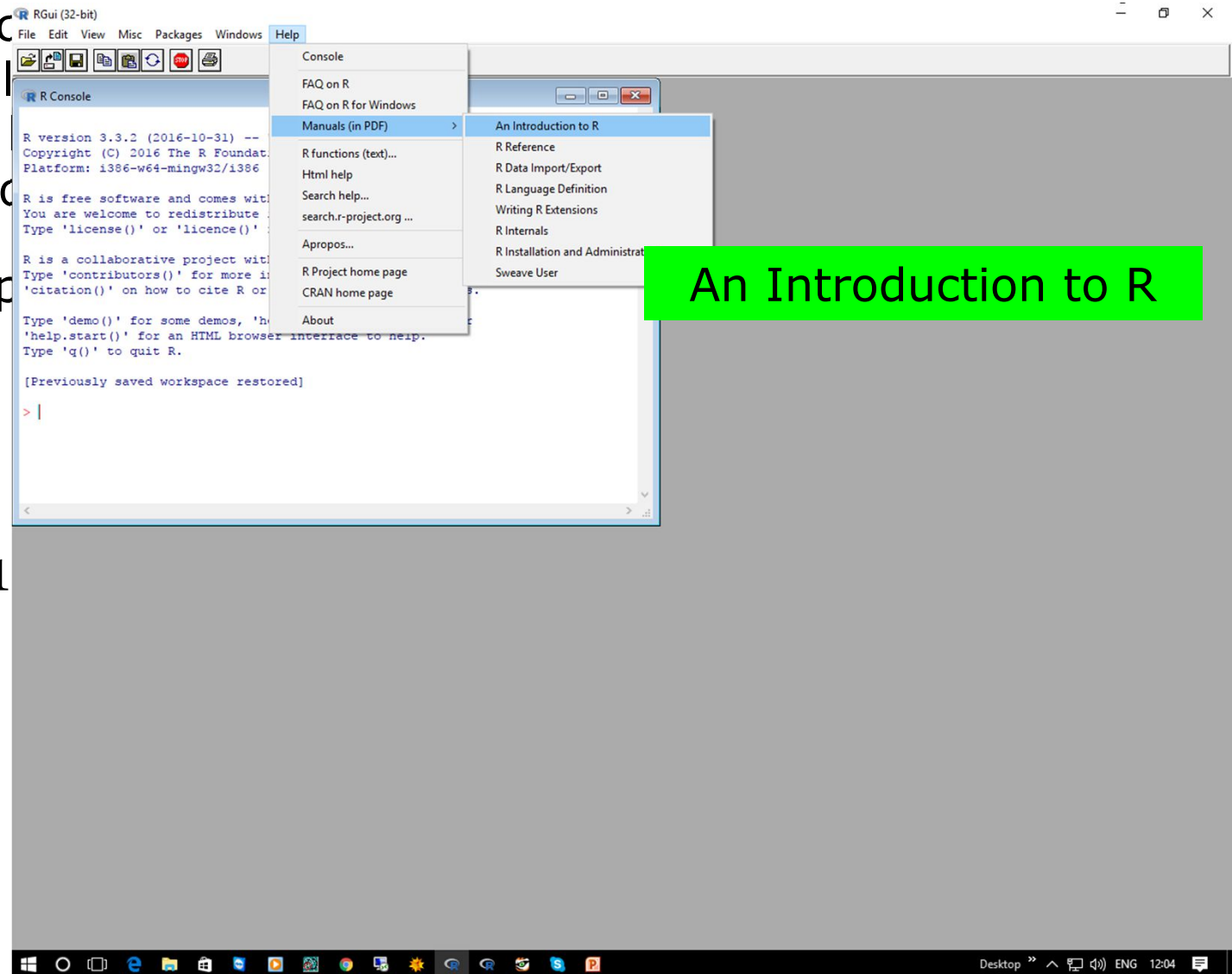
Esempio 1

Un test di ammissione ad un corso molto selettivo consiste di 15 domande a risposta multipla, con 3 alternative per ciascuna domanda. Solo i candidati ammessi al corso sono ammessi a rispondere a ciascuna domanda. Calcolare la probabilità che un candidato ammesso al corso solo a caso a ciascuna domanda

X = numero di risposte corrette

$$X \sim \text{Binomiale}\left(15, \frac{1}{3}\right)$$

$$P(X \geq 13) = P(X = 13) + P(X = 14) + P(X = 15)$$

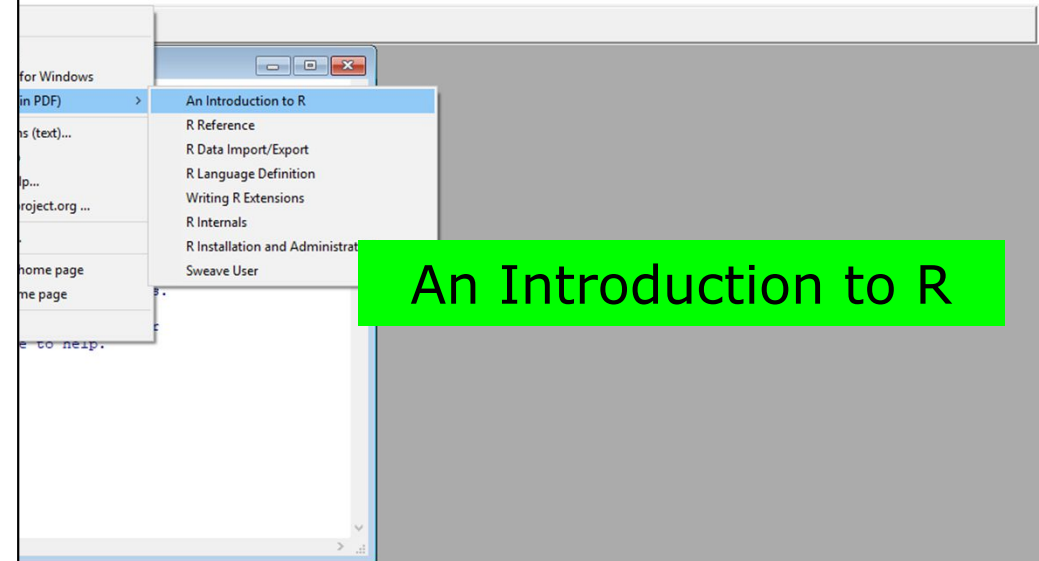




U
a
a
d
a

	ii
5.7.2 Linear equations and inversion	22
5.7.3 Eigenvalues and eigenvectors	23
5.7.4 Singular value decomposition and determinants	23
5.7.5 Least squares fitting and the QR decomposition	23
5.8 Forming partitioned matrices, <code>cbind()</code> and <code>rbind()</code>	24
5.9 The concatenation function, <code>c()</code> , with arrays	24
5.10 Frequency tables from factors	25
6 Lists and data frames	26
6.1 Lists	26
6.2 Constructing and modifying lists	27
6.2.1 Concatenating lists	27
6.3 Data frames	27
6.3.1 Making data frames	27
6.3.2 <code>attach()</code> and <code>detach()</code>	28
6.3.3 Working with data frames	28
6.3.4 Attaching arbitrary lists	28
6.3.5 Managing the search path	29
7 Reading data from files	30
7.1 The <code>read.table()</code> function	30
7.2 The <code>scan()</code> function	31
7.3 Accessing builtin datasets	31
7.3.1 Loading data from other R packages	31
7.4 Editing data	32
8 Probability distributions	33
8.1 R as a set of statistical tables	33
8.2 Examining the distribution of a set of data	34
8.3 One- and two-sample tests	36
9 Grouping, loops and conditional execution	40
9.1 Grouped expressions	40
9.2 Control statements	40
9.2.1 Conditional execution: <code>if</code> statements	40
9.2.2 Repetitive execution: <code>for</code> loops, <code>repeat</code> and <code>while</code>	40
10 Writing your own functions	42
10.1 Simple examples	42
10.2 Defining new binary operators	43
10.3 Named arguments and defaults	43
10.4 The <code>'...'</code> argument	44
10.5 Assignments within functions	44
10.6 More advanced examples	44
10.6.1 Efficiency factors in block designs	44
10.6.2 Dropping all names in a printed array	45
10.6.3 Recursive numerical integration	45
10.7 Scope	46
10.8 Customizing the environment	48
10.9 Classes, generic functions and object orientation	48

molto selettivo consiste di 15 domande



An Introduction to R

8.1 R as a set of statistical tables

One convenient use of R is to provide a comprehensive set of statistical tables. Functions are provided to evaluate the cumulative distribution function $P(X \leq x)$, the probability density function and the quantile function (given q , the smallest x such that $P(X \leq x) > q$), and to simulate from the distribution.

Distribution	R name	additional arguments
beta	beta	shape1, shape2, ncp
binomial	binom	size, prob
Cauchy	cauchy	location, scale
chi-squared	chisq	df, ncp
exponential	exp	rate
F	f	df1, df2, ncp
gamma	gamma	shape, scale
geometric	geom	prob
hypergeometric	hyper	m, n, k
log-normal	lnorm	meanlog, sdlog
logistic	logis	location, scale
negative binomial	nbinom	size, prob
normal	norm	mean, sd
Poisson	pois	lambda
signed rank	signrank	n
Student's t	t	df, ncp
uniform	unif	min, max
Weibull	weibull	shape, scale
Wilcoxon	wilcox	n, x

Prefix the name given here by 'd' for the density, 'p' for the CDF, 'q' for the quantile function and 'r' for simulation (random deviates). The first argument is x for $dxxx$, q for $pxxx$, p for $qxxx$ and n for $rxxx$ (except for $rhyper$, $rsignrank$ and $rwilcox$, for which it is nn). In not quite all cases is the non-centrality parameter ncp currently available: see the on-line help for details.



domande



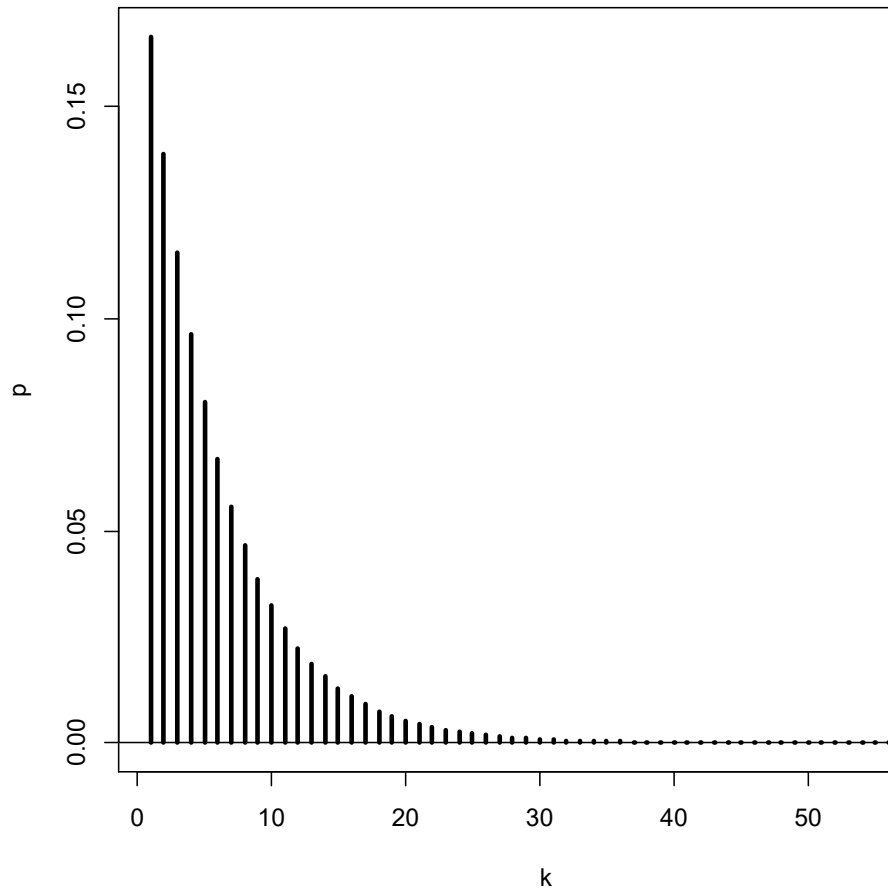
n to R

script2.R

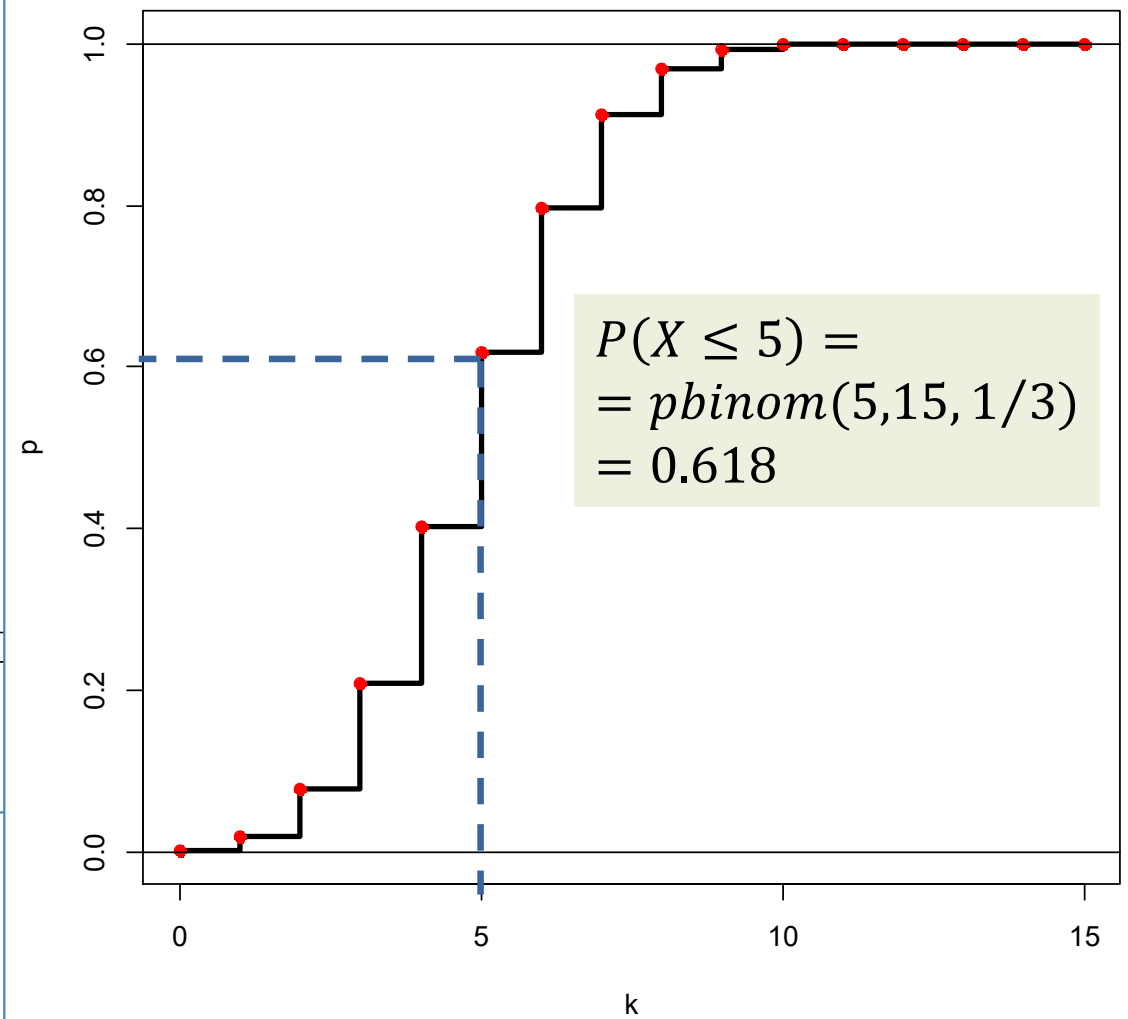
Facciamo un salto in



Geom(1/6)



CDF Bin(15,1/3)



Dal campione **alla popolazione**

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{x_1, x_2, \dots, x_N\}, \quad N \leq +\infty$$

distribuzione di X :

$$P(X = x_i) \text{ per tutti gli } x_i$$

Dal campione **alla popolazione**

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{x_1, x_2, \dots, x_N\}, \quad N \leq +\infty$$

distribuzione di X :

$P(X = x_i)$ per tutti gli x_i

dal censimento:

$$\bar{x}_n = \frac{x_1 + \dots + x_n}{n} =$$
$$= \sum \left(x_i \times \frac{1}{n} \right)$$

$$\mu = E(X) = \sum_{i=1, \dots, N} x_i P(X = x_i)$$



Media e varianza

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{x_1, x_2, \dots, x_N\}, \quad N \leq +\infty$$

distribuzione di X :

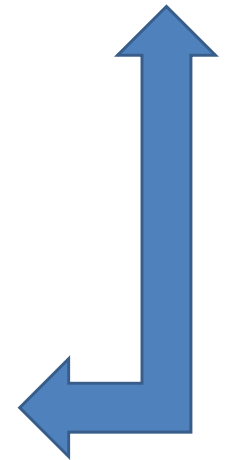
$$P(X = x_i) \text{ per tutti gli } x_i$$

dal censimento:

$$\sigma^2 = \sum \frac{1}{n} (x_i - \bar{x})^2$$

$$\mu = E(X) = \sum_{i=1, \dots, N} x_i P(X = x_i)$$

$$\sigma^2 = Var(X) = \sum_{i=1, \dots, N} (x_i - \mu)^2 P(X = x_i)$$



Media e varianza

**Modello
probabilistico
discreto:**

variabile casuale X :

$$X : (\dots) \rightarrow \{x_1, x_2, \dots, x_N\}, \quad N \leq +\infty$$

distribuzione di X :

$$P(X = x_i) \text{ per tutti gli } x_i$$

$$\mu = E(X) = \sum_{i=1, \dots, N} x_i P(X = x_i)$$

$$\sigma^2 = \text{Var}(X) = \sum_{i=1, \dots, N} (x_i - \mu)^2 P(X = x_i)$$

$$\sigma^2 = \left[\sum_{i=1, \dots, N} x_i^2 P(X = x_i) \right] - \mu^2$$

Esercizio "di compito"

La variabile discreta X ha la distribuzione di frequenza data nelle seguente tabella:

x	0	1	2	10	20
$P(X = x)$	0.55	0.05	0.05	0.05	0.30

- Rappresentare la distribuzione con un opportuno grafico
- Calcolare $P(X \geq 10)$ e $P(X > 5)$
- Calcolare $E(X)$ e $Var(X)$

Soluzioni parziali: b) 0.35, 0.35; c) 6.65, 81.0275

Media e varianza

**Modello
probabilistico
binomiale:**

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

$X : (\dots) \rightarrow \{0, 1, \dots, n\}, \quad N \leq +\infty$

distribuzione di X :

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

$$\mu = E(X) = np$$

$$\sigma^2 = \text{Var}(X) = np(1 - p) = npq$$

Media e varianza

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

$$\mu = E(X) = np$$

$$\sigma^2 = \text{Var}(X) = np(1 - p) = npq$$

Se $p = 0.5$, il numero medio di F che mi aspetto in 14 nascite è $14 \times 0.5 = 7$

Se $p = \frac{1}{3}$, il numero medio di risposte esatte che mi aspetto in 15 domande a risposta multipla, con 3 risposte possibili, è $15 \times \frac{1}{3} = 5$

Dal nostro *test*

5. Lanciate 100 volte una moneta non truccata: quante monete vi aspettate diano esito testa?

CIRCA 50

NON SI PUO' SAPERE

NON SAPREI

Dal nostro *test*

5. Lanciate 100 volte una moneta non truccata: quante monete vi aspettate diano esito testa?

CIRCA 50

NON SI PUO' SAPERE

NON SAPREI

$$X \sim \text{Binom}(100, 0.5) \Rightarrow E(X) = 100 \times 0.5 = 50$$

Dal nostro *test*

5. Lanciate 100 volte una moneta non truccata: quante monete vi aspettate diano esito testa?

CIRCA 50

NON SI PUO' SAPERE

NON SAPREI

$$X \sim \text{Binom}(100, 0.5) \Rightarrow E(X) = 100 \times 0.5 = 50$$

7. Lanciate 120 volte un dado non truccato: quante volte vi aspettate che esca il numero 2?

CIRCA 20

NON SI PUO' SAPERE

NON SAPREI

$$X \sim \text{Binom}(120, 1/6) \Rightarrow E(X) = 120 \times \frac{1}{6} = 20$$

Esercizio

Per i guidatori tra i 20 ed i 24 anni di età c'è un tasso annuo del 34% di incidenti d'auto (National Safety Council). Un investigatore assicurativo trova che in un gruppo di 500 guidatori scelti a caso a NY, tra i 20 e i 24 anni, il 42% ha avuto un incidente lo scorso anno.

a) quanti dei 500 hanno avuto un incidente?

Esercizio

Per i guidatori tra i 20 ed i 24 anni di età c'è un tasso annuo del 34% di incidenti d'auto (National Safety Council). Un investigatore assicurativo trova che in un gruppo di 500 guidatori scelti a caso a NY, tra i 20 e i 24 anni, il 42% ha avuto un incidente lo scorso anno.

a) quanti dei 500 hanno avuto un incidente?

Tra i 500 di NY, $500 \times 0.42 = 210$ hanno avuto un incidente

Esercizio

Per i guidatori tra i 20 ed i 24 anni di età c'è un tasso annuo del 34% di incidenti d'auto (National Safety Council). Un investigatore assicurativo trova che in un gruppo di 500 guidatori scelti a caso a NY, tra i 20 e i 24, il 42% ha avuto un incidente lo scorso anno.

b) Calcolare valore atteso e deviazione standard del numero di incidenti in un campione di 500 guidatori (20-24), secondo il NSC.

Tra i 500 di NY, $500 \times 0.42 = 210$ hanno avuto un incidente

$$X \sim \text{Bin}(500, 0.34) \Rightarrow$$

$$E(X) = 500 \times 0.34 = 170, \text{Var}(X) = 500 \times 0.34 \times 0.66 = 112.2$$

$$\Rightarrow \sqrt{\text{Var}(X)} = 10.59$$

Esercizio

Per i guidatori tra i 20 ed i 24 anni di età c'è un tasso annuo del 34% di incidenti d'auto (National Safety Council). Un investigatore assicurativo trova che in un gruppo di 500 guidatori scelti a caso a NY, tra i 20 e i 24 anni, il 42% ha avuto un incidente lo scorso anno.

c) Il risultato di NY sembra inusuale rispetto ai valori NSC? Si potrebbero giustificare assicurazioni più care per i Newyorkesi?

Tra i 500 di NY, $500 \times 0.42 = 210$ hanno avuto un incidente

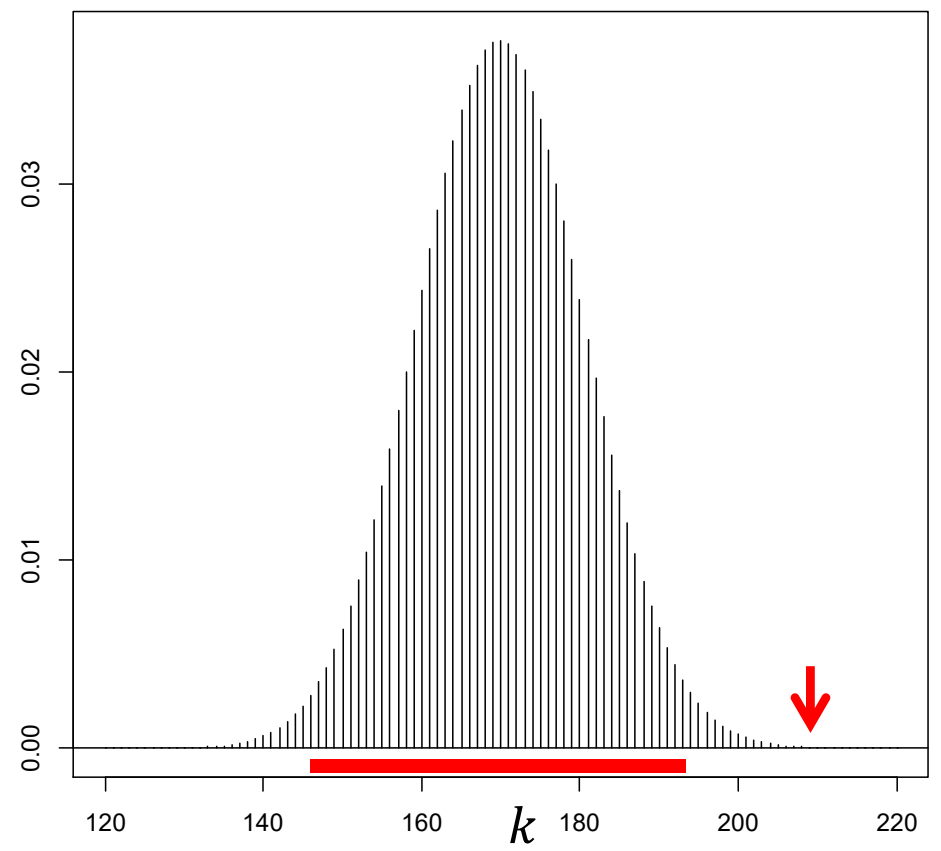
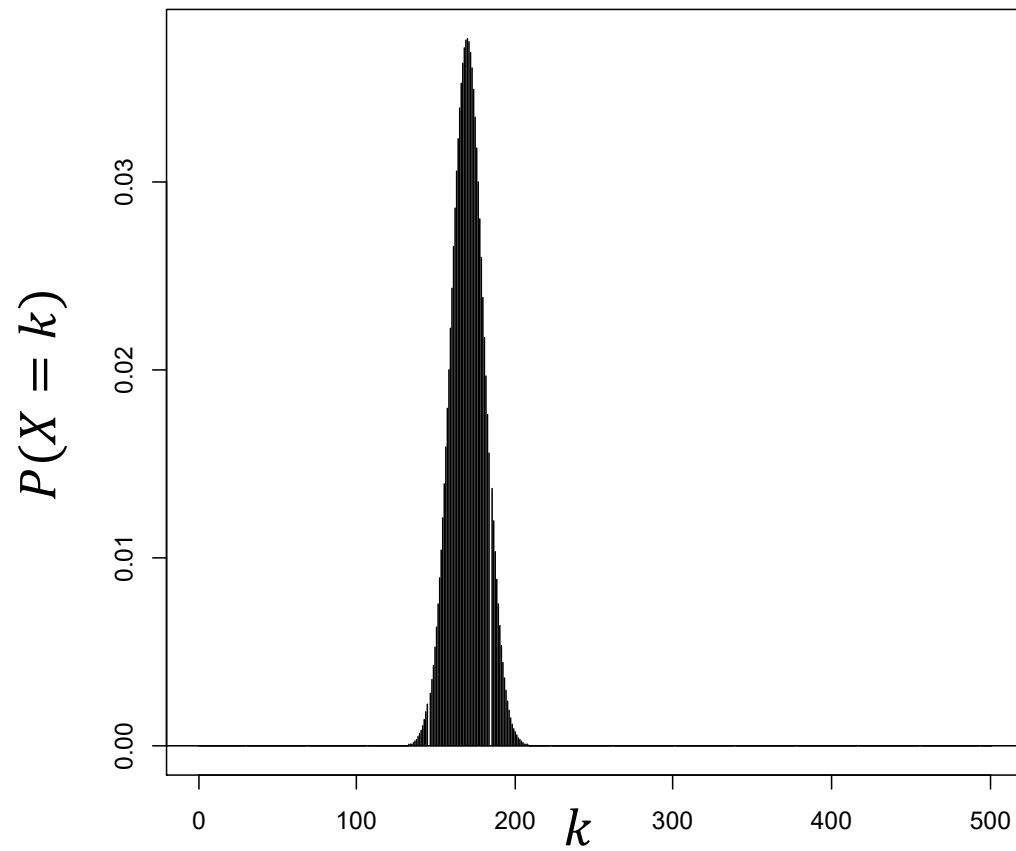
$X \sim \text{Bin}(500, 0.34) \Rightarrow$

$E(X) = 500 \times 0.34 = 170, \text{Var}(X) = 500 \times 0.34 \times 0.66 = 112.2$

$\Rightarrow \sqrt{\text{Var}(X)} = 10.59$

Esercizio

$Bin(500, 0.34)$



Esercizio

Per i guidatori tra i 20 ed i 24 anni di età c'è un tasso annuo del 34% di incidenti d'auto (National Safety Council). Un investigatore assicurativo trova che in un gruppo di 500 guidatori scelti a caso a NY, tra i 20 e i 24 anni, il 42% ha avuto un incidente lo scorso anno.

c) Il risultato di NY sembra inusuale rispetto ai valori NSC? Si potrebbero giustificare assicurazioni più care per i Newyorkesi?

Tra i 500 di NY, $500 \times 0.42 = 210$ hanno avuto un incidente

$$P(X \geq 210) = 0.00012$$

$$\text{e } P(149 \leq X \leq 191) = 0.96$$

Esercizio

Per i guidatori tra i 20 ed i 24 anni di età c'è un tasso annuo del 34% di incidenti d'auto (National Safety Council). Un investigatore assicurativo trova che in un gruppo di 500 guidatori scelti a caso a NY, tra i 20 e i 24 anni, il 42% ha avuto un incidente lo scorso anno.

c) Il risultato di NY sembra inusuale rispetto ai valori NSC? Si potrebbero giustificare assicurazioni più care per i Newyorkesi?

Tra i 500 di NY, $500 \times 0.42 = 210$ hanno avuto un incidente

$$P(X \geq 210) = 0.00012$$

$$\text{e } P(149 \leq X \leq 191) = 0.96$$

$$E(X) = 170, \sqrt{\text{Var}(X)} = 10.59 \Rightarrow \\ (149, 191) \approx (\mu \mp 3\sigma)$$

Dal nostro *test*

6. Lanciate 100 volte una moneta che, vi hanno detto, non è truccata, e segnate il risultato di ogni lancio. Alla fine dei 100 lanci vedete che in 88 lanci è uscita testa: il risultato può farvi dubitare della correttezza della moneta?

SI, perché	NO, perché	NON SAPREI

Dal nostro *test*

6. Lanciate 100 volte una moneta che, vi hanno detto, non è truccata, e segnate il risultato di ogni lancio. Alla fine dei 100 lanci vedete che in 88 lanci è uscita testa: il risultato può farvi dubitare della correttezza della moneta?

SI, perché

Dal nostro *test*

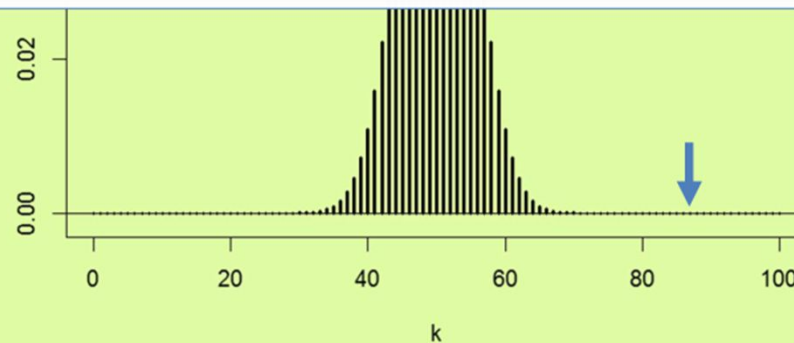
6. Lanciate 100 volte una moneta che, vi hanno detto, non è truccata, e segnate il risultato di ogni lancio. Alla fine dei 100 lanci vedete che in 88 lanci è uscita testa: il risultato può farvi dubitare della correttezza della moneta?

SI, perché

NO, perché

NON
SAPREI

Se la moneta fosse equilibrata il risultato avrebbe una prob. bassissima **a confronto** con gli altri risultati possibili!



modi: ora siamo in
equilibrate si ha

8.286361×10^{-16}

Dal nostro *test*

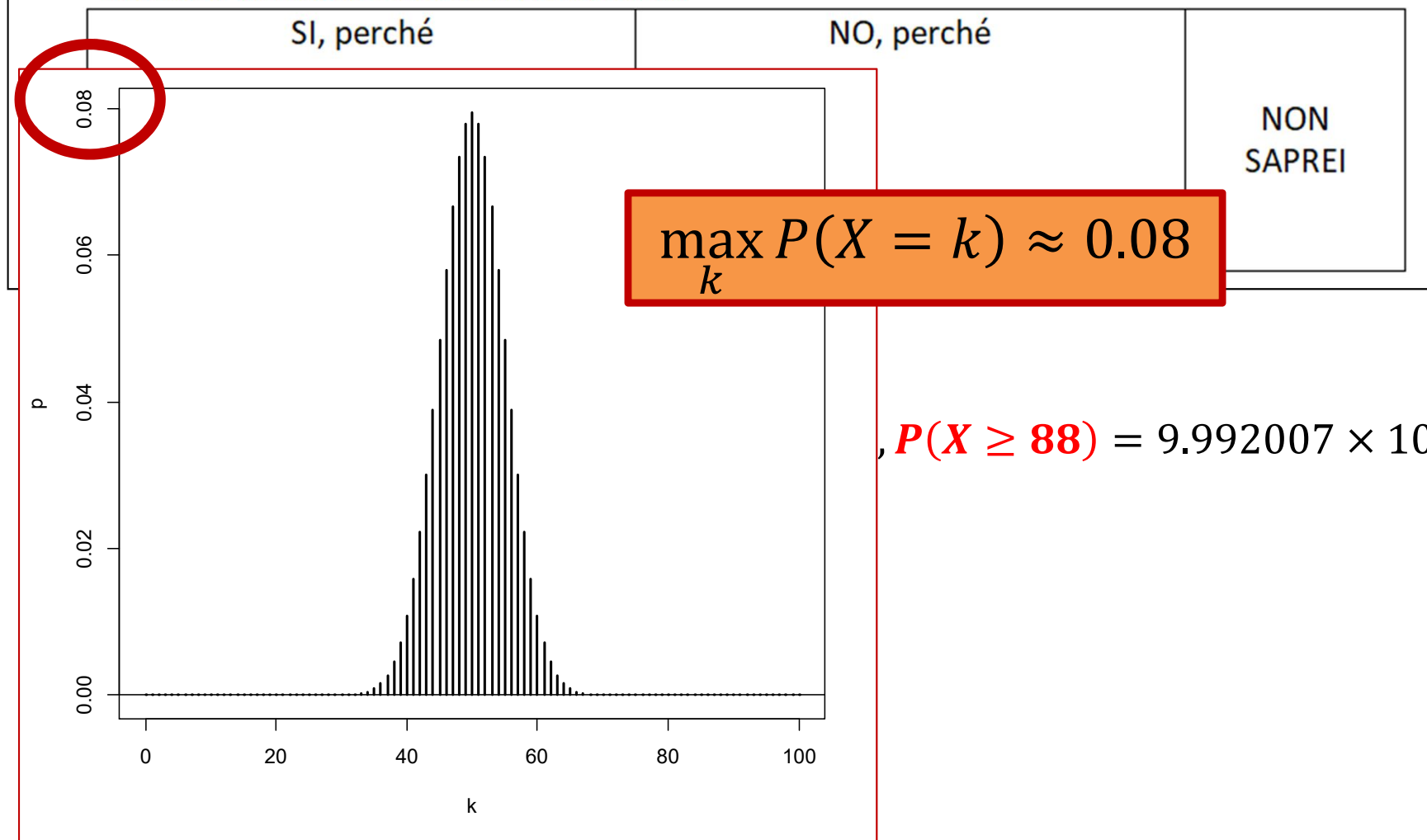
6. Lanciate 100 volte una moneta che, vi hanno detto, non è truccata, e segnate il risultato di ogni lancio. Alla fine dei 100 lanci vedete che in 88 lanci è uscita testa: il risultato può farvi dubitare della correttezza della moneta?

SI, perché	NO, perché	NON SAPREI
------------	------------	---------------

$$X \sim \text{Binom}(100, 0.5) \Rightarrow E(X) = 50 (\ll 88!), \mathbf{P(X \geq 88)} = 9.992007 \times 10^{-16}$$

Dal nostro *test*

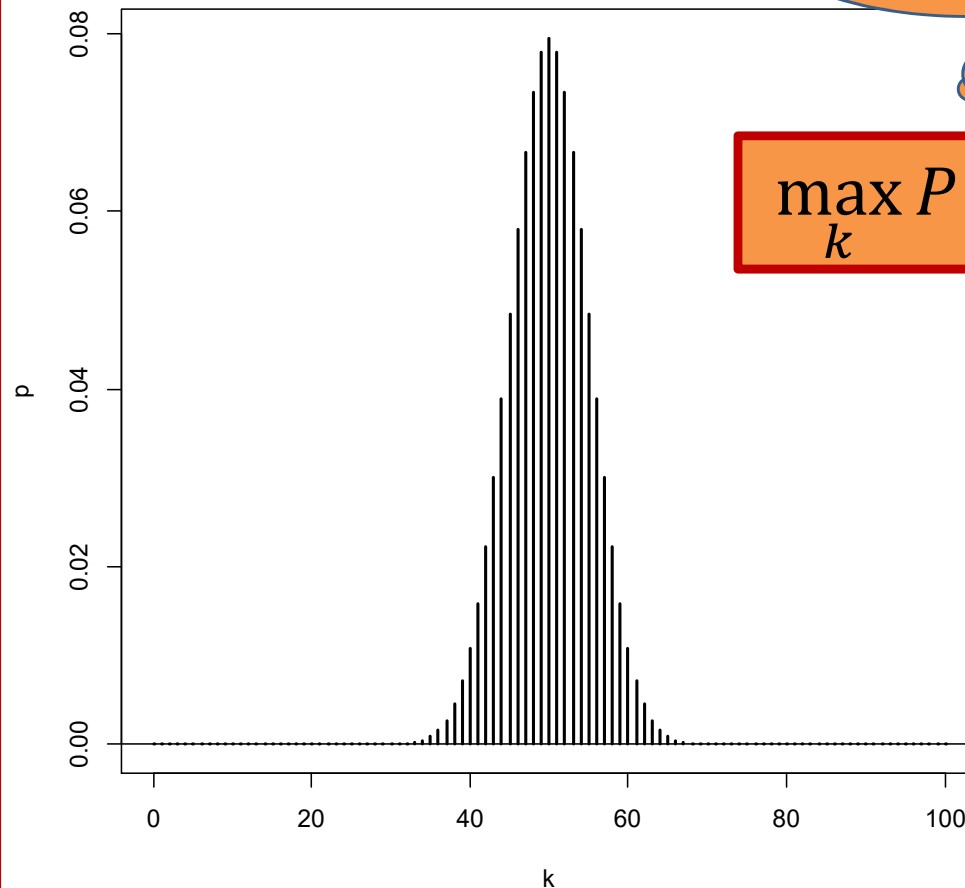
6. Lanciate 100 volte una moneta che, vi hanno detto, non è truccata, e segnate il risultato di ogni lancio. Alla fine dei 100 lanci vedete che in 88 lanci è uscita testa: il risultato può farvi dubitare della correttezza della moneta?



Dal nostro *test*

6. Lanciate 100 volte una moneta che è giusta. Alla fine dei 100 lanci, se il numero di teste è maggiore di 80, dubitate della correttezza della moneta.

SI, perché



$$\max_k P(X = k) \approx 0.08$$

OGNI risultato ha una probabilità bassa!! Perché n è grande.

NON
SAPREI

$$P(X \geq 88) = 9.992007 \times 10^{-16}$$

Cfr è
sempre
con 1!

Dal nostro *test*

6. Lanciate 100 volte una moneta che, vi hanno detto, non è truccata, e segnate il risultato di ogni lancio. Alla fine dei 100 lanci vedete che in 88 lanci è uscita testa: il risultato può farvi dubitare della correttezza della moneta?

SI, perché	NO, perché	NON SAPREI
------------	------------	---------------

POSTILLA :

$$X \sim \text{Binom}(100, 0.5) \Rightarrow E(X) = 50 (\ll 88!), \mathbf{P(X \geq 88)} = 9.992007 \times 10^{-16}$$

$$\mathbf{P(X \geq 88) = 1 - P(X \leq 87) =}$$
$$1 - \text{pbinom}(87, 100, 0.5) \text{ con R}$$

Attenzione che calcolato con R-studio viene un numero leggermente diverso, ma dello stesso ordine (8.881784×10^{-16})!