Parte 5

Modelli di probabilità per variabili casuali discrete

Vi sono **esperimenti casuali**, e le **variabili casuali** ad essi associate,
che presentano *caratteristiche comuni* secondo uno **schema standard** di riferimento

Le v.c. che rientrano in uno **schema** noto *definiscono* una **famiglia di v.c.**

Per *differenziare* le v.c. (i membri) appartenenti alla **famiglia**,
essa viene **parametrizzata** da *una* **costante** numerica denominata **parametro**
(o da un *vettore* di **parametri**) dai *molteplici valori*

Il particolare **valore** *attribuito* al **parametro** (o ai **parametri**)
individua una *specifica* v.c. all'interno della famiglia di v.c.

Famiglia parametrica di v.c.	X v.c. appartenente alla famiglia parametrica
$X \sim f(x; \boldsymbol{\theta})$ x valore di X	La v.c. X ha distribuzione descritta da $f(x; \boldsymbol{\theta})$

$\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$ vettore di $k \geq 1$ parametri (parametro vettoriale)
appartenente ad un insieme denominato **spazio parametrico** $\Theta \subseteq \mathbb{R}^k$

$f(x; \boldsymbol{\theta})$ funzione di **probabilità** (se X v.c. discreta) o funzione di **densità** (se X v.c. continua) **parametrica**	Modello di probabilità **parametrico**

$X \sim f(x;\boldsymbol{\theta})$	x valore di X	$\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$	$\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$ vettore di parametri

Modello di probabilità **uguale** nella *forma funzionale*
per **tutte** le v.c. della **famiglia parametrica**

Le **singole** v.c. (membri della famiglia parametrica)
e i loro **specifici** modelli di probabilità *si distinguono* esclusivamente
per i **valori numerici** attribuiti ai parametri del vettore $\boldsymbol{\theta}$

Insieme delle $f(x;\boldsymbol{\theta})$	**Famiglia parametrica** di modelli di probabilità

Famiglia parametrica di funzioni di **probabilità** (se X v.c. discreta)
o di funzioni di **densità** (se X v.c. continua)

Se θ singolo parametro	$X \sim f(x;\theta)$	x valore di X	$\theta \in \Theta \subseteq \mathbb{R}$

Famiglia parametrica (monoparametrica) di v.c.

Per talune famiglie parametriche, i **parametri** assumono specifiche denominazioni
(ad esempio $N, p, n, \mu, \sigma^2, g$)

Variabile casuale **Uniforme discreta**

Esperimento casuale *assimilabile* all'estrazione di una pallina da un'urna contenente N palline **identiche** (indistinguibili) e **numerate** da 1 a N, con la relativa osservazione del **numero** della pallina estratta

Variabile casuale discreta *associata* all'esperimento casuale	$X =$ "**Numero** della pallina estratta"
$x \in \{1,2,\cdots,N\}$	$\{1,2,\cdots,N\}$ **supporto** della v.c.

$N \geq 1$ (numero intero) è il **parametro** che *caratterizza* tale v.c.

L'**insieme** delle v.c. *generato* da *tutti i possibili* valori di N (sono una *infinità numerabile*) è la **famiglia parametrica** di v.c.

Funzione/Modello di **probabilità** parametrica/o della v.c. X

$f(x;N) = \dfrac{1}{N}$	$x \in \{1,2,\cdots,N\}$		$N \in \{1,2,\cdots,\infty\} \subset \mathbb{R}$	$X \sim Ud(N)$
	$f(x;N) = 0$	altrove	Spazio parametrico	

La v.c. X ha distribuzione di probabilità **Uniforme discreta** di parametro N

$$f(x;N)=\frac{1}{N} \qquad x\in\{1,2,\cdots,N\} \qquad N\in\{1,2,\cdots,\infty\}$$

Proprietà soddisfatte dalla funzione di probabilità

$f(x;N)=\frac{1}{N}>0$	$\sum_x f(x;N)=\sum_x \frac{1}{N}=N\frac{1}{N}=1$

L'**insieme** delle $f(x;N)$ *generato* dai valori di N è la **famiglia parametrica** di funzioni/modelli di **probabilità** di tipo **Uniforme discreto** di parametro N

Per alcuni valori del parametro N, si riconoscono v.c. **Uniformi discrete**

Esperimento casuale	Lancio di un dado **equilibrato**"	Esempio 5
$X=$ "**Numero** ottenuto (sulla faccia in alto)"		$X\sim Ud(6)$
Esperimento casuale	Estrazione di un numero su una ruota nel gioco del Lotto	
$X=$ "**Numero** estratto"		$X\sim Ud(90)$

Modello **Uniforme discreto** comune nei *giochi di sorte* (ma non solo)

Media e **varianza** di una v.c. $X \sim Ud(N)$

$$f(x;N) = \frac{1}{N} \qquad x \in \{1,2,\cdots,N\} \qquad N \in \{1,2,\cdots,\infty\}$$

$$E(X) = \sum_x x\frac{1}{N} = \frac{1}{N}\sum_x x = \frac{1}{N} \times \frac{N \times (N+1)}{2} = \frac{N+1}{2}$$

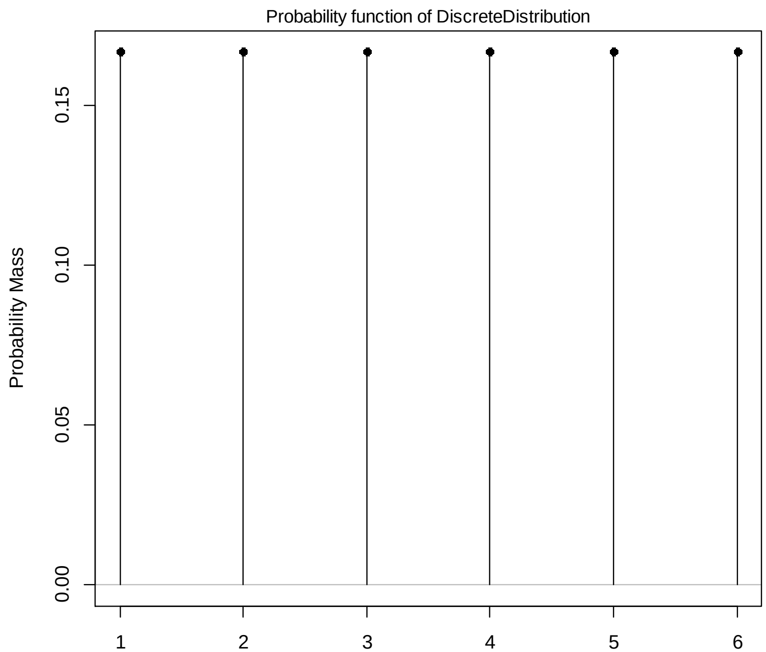
$$V(X) = \frac{(N-1)(N+1)}{12} = \frac{N^2-1}{12}$$

Media e varianza **funzioni** del parametro *N*

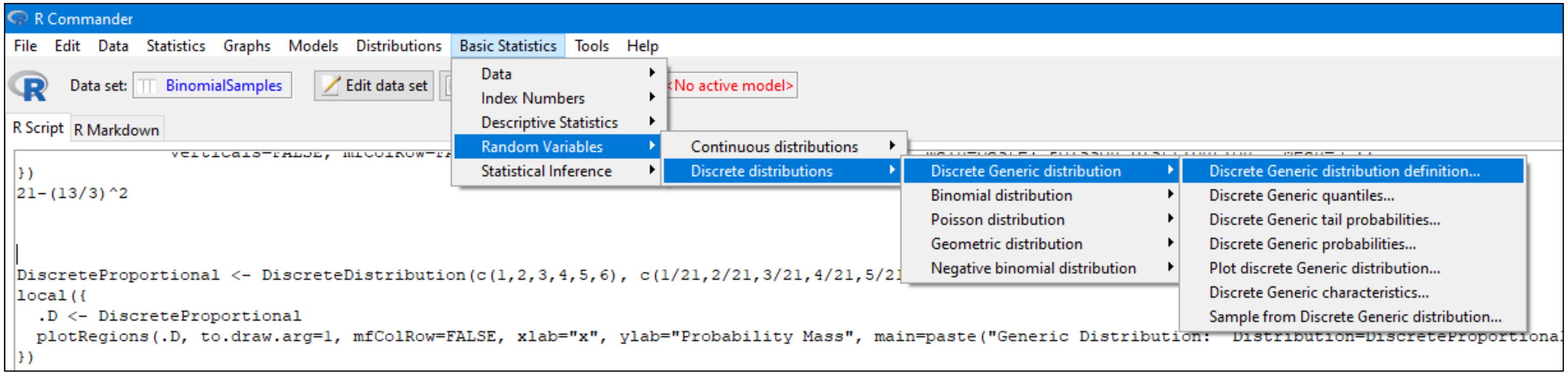
Notazione rigorosa	$E(X;N)$	$V(X;N)$

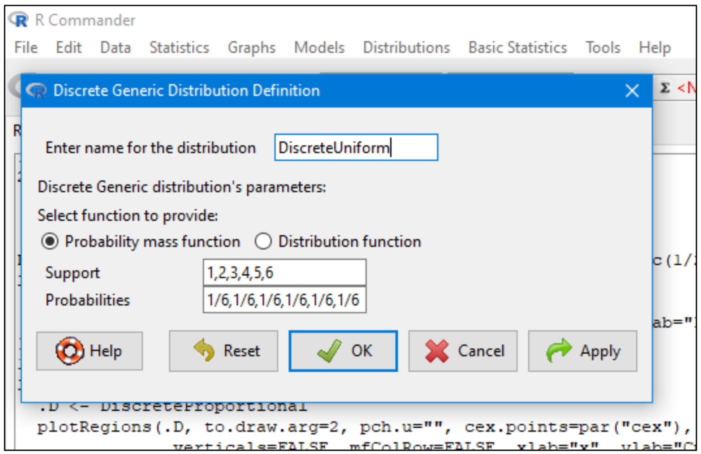
Accade (in generale) per tutti i **momenti** e
per ogni altro **valore di sintesi** di una v.c. *X*

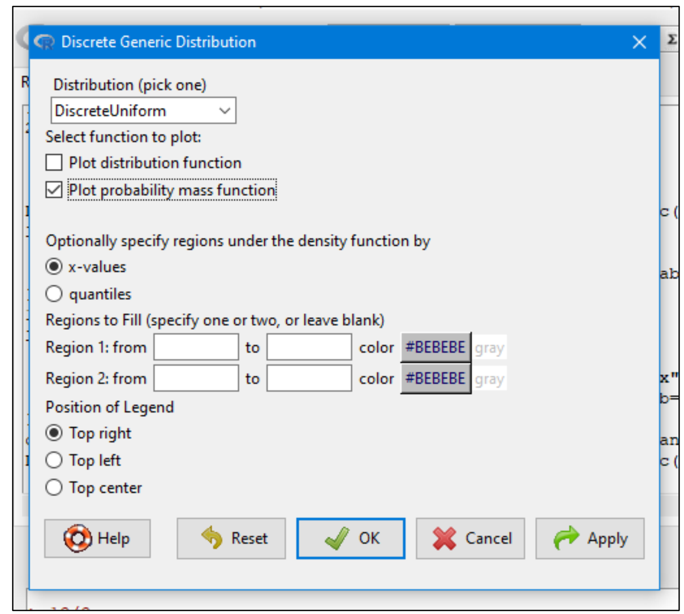
Esempio 5	Lancio di un dado equilibrato	
$X =$ "**Numero** ottenuto (sulla faccia in alto)"		$X \sim Ud(6)$
$f(x;6) = f(x) = \frac{1}{6}$	$x \in \{1,2,\cdots,6\}$	$f(x) = 0$ altrove
$E(X) = \frac{7}{2} = 3.5$	$V(X) = \frac{5 \times 7}{12} = \frac{5}{6} 3.5 = 2.917 < E(X)$	

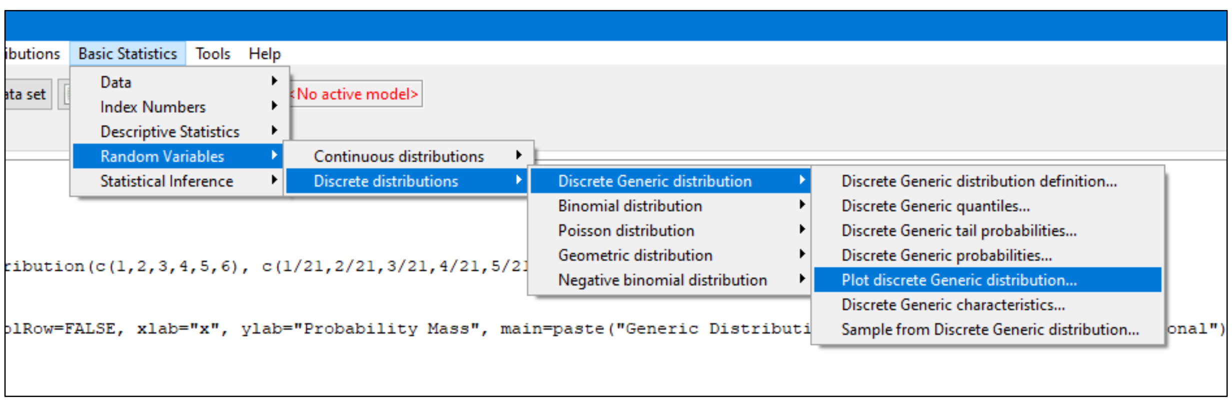


Distribuzione simmetrica	$E(X) = Me$
	$E(Z^3) = 0$


R Commander
File Edit Data Statistics Graphs Models Distributions Basic Statistics Tools Help
Data set: BinomialSamples
Edit data set
<No active model>
R Script R Markdown
Data
Index Numbers
Descriptive Statistics
Random Variables
Statistical Inference
Continuous distributions
Discrete distributions
Discrete Generic distribution
Binomial distribution
Poisson distribution
Geometric distribution
Negative binomial distribution
Discrete Generic distribution definition...
Discrete Generic quantiles...
Discrete Generic tail probabilities...
Discrete Generic probabilities...
Plot discrete Generic distribution...
Discrete Generic characteristics...
Sample from Discrete Generic distribution...
})
21-(13/3)^2
DiscreteProportional <- DiscreteDistribution(c(1,2,3,4,5,6), c(1/21,2/21,3/21,4/21,5/21
local({
.D <- DiscreteProportional
plotRegions(.D, to.draw.arg=1, mfColRow=FALSE, xlab="x", ylab="Probability Mass", main=paste("Generic Distribution: Distribution=DiscreteProportional
})


R Commander
File Edit Data Statistics Graphs Models Distributions Basic Statistics Tools Help
Discrete Generic Distribution Definition
Enter name for the distribution DiscreteUniform
Discrete Generic distribution's parameters:
Select function to provide:
Probability mass function
Distribution function
Support 1,2,3,4,5,6
Probabilities 1/6,1/6,1/6,1/6,1/6,1/6
Help
Reset
OK
Cancel
Apply
.D <- DiscreteProportional
plotRegions(.D, to.draw.arg=2, pch.u="", cex.points=par("cex"),


Discrete Generic Distribution
Distribution (pick one)
DiscreteUniform
Select function to plot:
Plot distribution function
Plot probability mass function
Optionally specify regions under the density function by
x-values
quantiles
Regions to Fill (specify one or two, or leave blank)
Region 1: from to color #BEBEBE gray
Region 2: from to color #BEBEBE gray
Position of Legend
Top right
Top left
Top center
Help
Reset
OK
Cancel
Apply


Basic Statistics Tools Help
<No active model>
Data
Index Numbers
Descriptive Statistics
Random Variables
Statistical Inference
Continuous distributions
Discrete distributions
Discrete Generic distribution
Binomial distribution
Poisson distribution
Geometric distribution
Negative binomial distribution
Discrete Generic distribution definition...
Discrete Generic quantiles...
Discrete Generic tail probabilities...
Discrete Generic probabilities...
Plot discrete Generic distribution...
Discrete Generic characteristics...
Sample from Discrete Generic distribution...

Variabile casuale **di Bernoulli** (o **Bernoulliana**)	Jacob Bernoulli (1655 – 1705)
Esperimento casuale che consiste nella **osservazione dell'esito** dello stesso, tra i suoi **due soli** possibili esiti (eventi elementari) denominati *convenzionalmente* **insuccesso** e **successo**	Esperimento Bernoulliano (o dicotomico)

Insuccesso e **successo** *numerizzati* (*codificati numericamente*) nei rispettivi valori 0 e 1

Variabile casuale discreta *associata* all'esperimento casuale	$X =$ "**Esito numerico** (0 o 1) osservato"
$x \in \{0,1\}$	$\{0,1\}$ **supporto** (binario) della v.c.

Sia p la **probabilità di successo**	$f(1) = P(X = 1) = p$	$f(0) = 1 - p$

$0 \leq p \leq 1$ (numero reale) è il **parametro** che *caratterizza* tale v.c.

L'**insieme** delle v.c. *generato* da *tutti i possibili* valori di p (sono una *infinità non numerabile*) è la **famiglia parametrica** di v.c.

<table>
<tr><td colspan="6">Funzione/Modello di **probabilità** parametrica/o della v.c. X</td></tr>
<tr><td rowspan="2">$f(x;p) = p^x(1-p)^{1-x}$</td><td colspan="2">$x \in \{0,1\}$</td><td>$p \in (0,1) \subset \mathbb{R}$</td><td rowspan="2">$X \sim B(p)$</td></tr>
<tr><td>$f(x;p) = 0$</td><td>altrove</td><td>Spazio parametrico</td></tr>
<tr><td colspan="6">La v.c. X ha distribuzione di probabilità **di Bernoulli** di parametro p</td></tr>
<tr><td colspan="6">**Proprietà** soddisfatte dalla funzione di probabilità</td></tr>
<tr><td rowspan="2">$f(x;p) = p^x(1-p)^{1-x}$</td><td>$= 1-p$</td><td rowspan="2">> 0</td><td rowspan="2" colspan="2">$\sum_x f(x;p) = (1-p) + p = 1$</td></tr>
<tr><td>$= p$</td></tr>
<tr><td colspan="6">L'**insieme** delle $f(x;p)$ *generato* dai valori di p è la **famiglia parametrica**
di funzioni/modelli di **probabilità** di tipo **Bernoulliano** di parametro p</td></tr>
<tr><td colspan="6">v.c. X *definita* anche per $p = 0$ e $p = 1$</td></tr>
<tr><td colspan="2">$p = 0$</td><td colspan="2">$p = 1$</td></tr>
<tr><td>X v.c. degenere</td><td>$X \equiv 0$ con $f(0) = 1$</td><td>X v.c. degenere</td><td>$X \equiv 1$ con $f(1) = 1$</td></tr>
</table>

Media e **varianza** di una v.c. $X \sim B(p)$

$$f(x;p) = p^x(1-p)^{1-x} \qquad x \in \{0,1\} \qquad p \in (0,1)$$

$$E(X) = \sum_x x\, p^x(1-p)^{1-x} = 0 \times (1-p) + 1 \times p = p$$

La media **coincide** con il parametro p che caratterizza tale v.c.

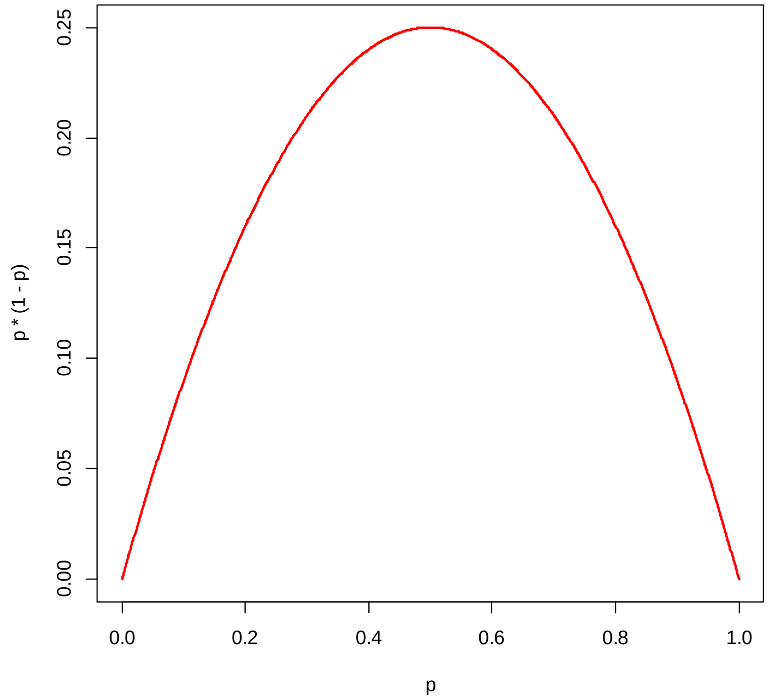
$$V(X) = \sum_x (x-p)^2\, p^x(1-p)^{1-x} = (0-p)^2(1-p) + (1-p)^2 p =$$

$$= p^2(1-p) + p(1-p)^2 = p(1-p)(p+1-p) = p(1-p)$$

$V(X) = p(1-p) = (1-p)E(X)$	$p = 0$	$V(X) = 0$	$E(X) = 0$
	$0 < p < 1$	$V(X) < E(X)$	
	$p = 1$	$V(X) = 0$	$E(X) = 1$

Andamento della varianza $p(1-p)$ in *funzione* della media p

```
> curve (p*(1-p), 0, 1, n=1000, xname = "p", col="red", lwd=2)
```



Andamento *simmetrico* rispetto al valore $p = 0.5$		
Valori di varianza *uguali* per valori di p *l'uno l'opposto dell'altro*		
$V(X) \leq 0.25$	$p = 0.5 \Longrightarrow V(X) = 0.25$	Massima *incertezza* tra i valori 0 e 1

La v.c. di Bernoulli ricorre *frequentemente* come v.c. **di popolazione**

Esempi di v.c. di popolazione con *dicotomia strutturale*

- Sesso (M/F) di un individuo scelto a caso da una popolazione finita
- Sopravvivenza (SI/NO) di un malato terminale entro un certo tempo
- Occupazione (SI/NO) di un individuo tra la forza lavoro di una data regione
- Propensione al voto (SI/NO) di un elettore per un dato partito
- Diagnosi di una malattia (SI/NO) per un paziente ambulatoriale
- Esito di un test diagnostico (POSITIVO/NEGATIVO)
- Superamento (SI/NO) di un esame universitario (concorso pubblico, ecc.)

Esempi di v.c. di popolazione (discrete o continue) opportunamente *dicotomizzate*

- Durata di una lampada (minore o uguale/maggiore di 20000 ore)
- Livello di colesterolo (minore o uguale/maggiore della soglia di riferimento)
- Precipitazione media mensile (minore o uguale/maggiore di 200 mm/cm^2)
- Numero di auto in attesa al casello (minore o uguale/maggiore di 30 al minuto)

In questi casi (ripetendo n volte, in modo **indipendente**, l'esperimento Bernoulliano per ottenere un **campione casuale**) il parametro p del modello (**probabilità** o proporzione **di successo**) è oggetto di **inferenza statistica**

Altri impieghi della v.c. di Bernoulli

Giochi di sorte

- Esito (TESTA/CROCE) nel lancio di una moneta
- Numero (minore o uguale/maggiore di 4) nel lancio di un dado
- Esito (tripla testa/altra terna) nel lancio di tre monete
- Numero (doppio 6/altra coppia) nel lancio di due dadi
- Carta (asso/altra carta) nell'estrazione da un mazzo di carte
- Numero estratto (13/altro numero) sulla ruota di Napoli nel gioco del Lotto

Esperimenti Bernoulliani *assimilabili* all'estrazione di una pallina da un'urna contenente palline **colorate**, di cui *un certo numero* di un **dato colore** (es. bianco) e *le rimanenti* di un **altro colore** (es. nero)

In questi casi, il parametro p del modello è **noto**
(valore *assegnato* in base alla concezione **classica** o **frequentista** di probabilità)

Assimilazione all'estrazione da un'urna anche nei casi di esperimenti Bernoulliani condotti per una **popolazione finita** (cfr esempi di pag. 13)

Esempio 1	Campionamento casuale ($n = 3$) da una classe di 50 studenti		
X_1 = "**Sesso** (0/1 per M/F) del **primo** studente estratto"			$X_1 \sim B(p)$
X_2 = "**Sesso** (0/1 per M/F) del **secondo** studente estratto"			$X_2 \sim B(p)$
X_3 = "**Sesso** (0/1 per M/F) del **terzo** studente estratto"			$X_3 \sim B(p)$
p è la probabilità di **successo** (lo studente estratto è **femmina**)			
Le v.c. X_1, X_2, X_3 sono **indipendenti** e **identicamente distribuite** (v.c. IID)			
In virtù dell'estrazione **con riposizione** dalla classe dei 50 studenti			
X_1, X_2, X_3 *denominazioni* diverse (funzionali all'estrazione di riferimento) della v.c. di popolazione X			
X v.c. di popolazione di tipo Bernoulliano		Popolazione Bernoulliana	
$f(x_i; p) = p^{x_i}(1-p)^{1-x_i}$	$x_i \in \{0,1\}$	$f(x_i, p) = 0$ altrove	$i = 1,2,3$
Logica inferenziale			
Parametro p *valutato* sulla base del **campione osservato**		Ad esempio $(1,0,1)$	

Logica probabilistica

Valore del parametro p *assegnato* usando la definizione **classica** di probabilità

Composizione numerica della classe		$p = P(X=1) = \frac{30}{50} = 0.6$
$n(0) = 20$ maschi	$n(1) = 30$ femmine	$1 - p = 0.4$

$X_1 \sim B(0.6)$	$X_2 \sim B(0.6)$	$X_3 \sim B(0.6)$

$f(x_i; 0.6) = f(x_i) = 0.6^{x_i} 0.4^{1-x_i}$	$x_i \in \{0,1\}$	$f(x_i) = 0$ altrove	$i = 1,2,3$

$$E(X_1) = E(X_2) = E(X_3) = p = 0.6$$

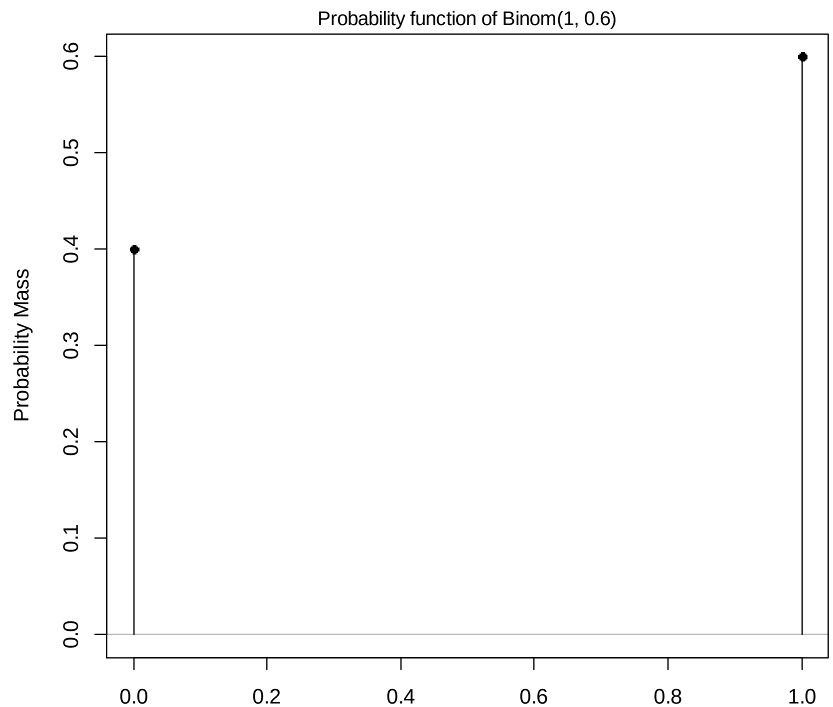
$$V(X_1) = V(X_2) = V(X_3) = p(1-p) = 0.6 \times 0.4 = 0.24 < 0.6$$

Distribuzioni di probabilità delle v.c. X_1, X_2 e X_3

x_1	$f(x_1)$
0	0.4
1	0.6

x_2	$f(x_2)$
0	0.4
1	0.6

x_3	$f(x_3)$
0	0.4
1	0.6



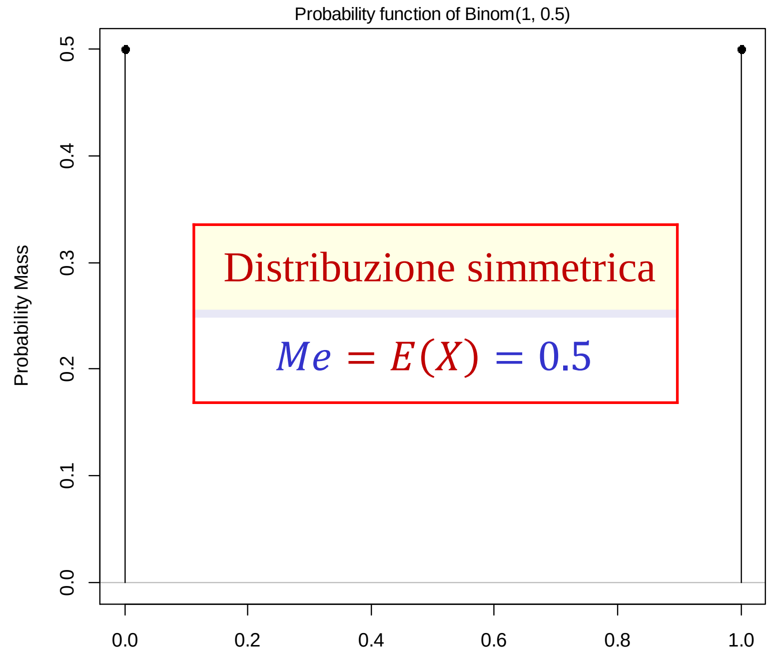
Per una generica v.c. $X \sim B(p)$

$p > 0.5$	$E(Z^3) < 0$	Asimmetria negativa
$p = 0.5$	$E(Z^3) = 0$	Simmetria
$p < 0.5$	$E(Z^3) > 0$	Asimmetria positiva

Esempio 8	Lancio di una moneta **bilanciata** (non truccata)

X = "**Faccia** ottenuta (0/1 per C/T) nel lancio della moneta"	$X \sim B(0.5)$
$p = 0.5$ è la probabilità di **successo** (esce **testa**)	Moneta bilanciata

$f(x; 0.5) = f(x) = 0.5^x 0.5^{1-x}$	$x \in \{0,1\}$	$f(x) = 0$ altrove
$E(X) = p = 0.5$	$V(X) = p(1-p) = \mathbf{0.25} < 0.5$	$SD(X) = \mathbf{0.5} = p = E(X)$



Distribuzione simmetrica
$Me = E(X) = 0.5$

Se la moneta è **truccata**	$p \neq 0.5 \neq 1 - p$

Variabile casuale **Binomiale**

Esperimento Bernoulliano **ripetuto** n volte nelle **medesime condizioni**
(dunque le n esecuzioni ripetute sono tra di esse **indipendenti**)

Probabilità di successo p **uguale** negli n esperimenti Bernoulliani

Esperimento Binomiale	**Esperimento casuale** *composto* da n sottoprove **indipendenti** di tipo Bernoulliano (con la **medesima** probabilità di successo p)

A *ciascuna* sottoprova Bernoulliana è *associata* una v.c. di Bernoulli di parametro p

$X_i \sim B(p)$	$x_i \in \{0,1\}$	$i = 1,2,\cdots,n$	v.c. IID di tipo $B(p)$

L'esperimento binomiale *genera* una sequenza di valori 0/1 (sequenza binomiale)

Variabile casuale discreta *associata* all'esperimento binomiale

$X = X_1 + X_2 + \cdots + X_n$	"**Numero di successi** osservati nella sequenza binomiale"

La v.c. X "conta" i successi ottenuti nell'esperimento (nelle n sottoprove Bernoulliane)

$x \in \{0,1,\cdots,n\}$	$\{0,1,\cdots,n\}$ **supporto** della v.c.

$n \geq 1$ (numero intero) e $0 \leq p \leq 1$ (numero reale) sono i due **parametri** che *caratterizzano* tale v.c.	$\boldsymbol{\theta} = (n, p)$
$\boldsymbol{\theta} = (n, p) \in \Theta = \{(n, p)\colon n \geq 1, 0 \leq p \leq 1\} \subset \mathbb{R}^2$	Θ spazio parametrico
L'**insieme** delle v.c. *generato* da *tutti i possibili* valori di n e p (sono una *infinità non numerabile*) è la **famiglia parametrica** di v.c.	

Funzione/Modello di **probabilità** parametrica/o della v.c. X		
$f(x; n, p) = P(X = x) = ?$	$x \in \{0,1, \cdots, n\}$	L'evento $(X = x)$ si verifica **se**
La sequenza binomiale contiene x *successi* ed $n - x$ *insuccessi* (in *tutti* i modi possibili)		
Sia $(x_1, x_2, \cdots, x_n)$ *una specifica* sequenza con x *successi* ed $n - x$ *insuccessi*		
Probabilità di tale *specifica* sequenza binomiale (è una probabilità congiunta)		
$f(x_1, x_2, \cdots, x_n) = f(x_1) \times f(x_2) \times \cdots \times f(x_n) = p^x (1-p)^{n-x}$		
Probabilità **uguale** per tutte le sequenze che generano il *medesimo* valore x della v.c. X		

Quante sono le sequenze binomiali nelle quali, *combinandosi* in ogni modo possibile, vi sono x successi (valori pari a 1) tra gli n valori?

Il loro numero è quello delle **combinazioni** di n elementi ad x ad x (gruppi *non ordinati*, o **sottoinsiemi**, di x elementi scelti tra gli n)

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

Si legge "coefficiente binomiale n su x (da n scegli x)"

Esempio		Quante sono le sequenze di 5 valori con 3 successi (valori unitari)?	$\binom{5}{3} = \frac{5!}{3!\,2!} = \frac{5\times 4}{2} = 10$
$n=5$	$x=3$		

(1,1,1,0,0)	(1,1,0,1,0)	(1,1,0,0,1)	(1,0,1,1,0)	(1,0,1,0,1)
(1,0,0,1,1)	(0,1,1,1,0)	(0,1,1,0,1)	(0,1,0,1,1)	(0,0,1,1,1)

Le $\binom{n}{x}$ sequenze con x successi sono tra di esse **incompatibili**

Il numero **totale** di sequenze è 2^n

Si tratta di **eventi elementari** in quanto **realizzazioni** della v.c. multipla $(X_1, X_2, \cdots, X_n)$

In definitiva, l'evento $(X = x)$ coincide con l'evento

Unione delle $\binom{n}{x}$ sequenze binomiali con x *successi* ed $n - x$ *insuccessi*,
tra di esse **incompatibili** e con la **medesima probabilità** $p^x(1-p)^{n-x}$

$P(X = x) =$	**Somma** delle probabilità di tali $\binom{n}{x}$ sequenze binomiali *equiprobabili*

Funzione/Modello di **probabilità** parametrica/o della v.c. X

$f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, 1, \cdots, n\}$	$f(x; n, p) = 0$ altrove

$(n, p) \in \Theta = \{(n, p) \colon n \geq 1, 0 < p < 1\} \subset \mathbb{R}^2$	Θ spazio parametrico
$X \sim Bin(n, p)$	La v.c. X ha distribuzione di probabilità **Binomiale** di parametri n e p

Caso speciale $n = 1$	$X \sim Bin(1, p)$	La v.c Binomiale **coincide** con la v.c. di Bernoulli	$X \sim B(p)$

Proprietà soddisfatte dalla funzione di probabilità

$$f(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x} > 0$$

$$\sum_x f(x;p) = \sum_x \binom{n}{x} p^x (1-p)^{n-x} = \big(p + (1-p)\big)^n = 1$$

Formula del binomio di Newton

La v.c. Binomiale *deve il suo nome* al fatto che la sua *funzione di probabilità* è il *termine generico* nella formula del binomio applicata al binomio $p + (1-p)$

L'**insieme** delle $f(x;n,p)$ *generato* dai valori di n e p è la **famiglia parametrica** di funzioni/modelli di **probabilità** di tipo **Binomiale** di parametri n e p

v.c. X *definita* anche per $p = 0$ e $p = 1$

$p = 0$		$p = 1$	
X v.c. degenere	$X \equiv 0$ con $f(0) = 1$	X v.c. degenere	$X \equiv n$ con $f(n) = 1$

<table>
<tr><td colspan="5">Media e varianza di una v.c. $X \sim Bin(n,p)$</td></tr>
<tr><td colspan="2">$X = X_1 + X_2 + \cdots + X_n$</td><td colspan="2">$E(X_i) = p$</td><td>$V(X_i) = p(1-p)$</td></tr>
<tr><td colspan="2">X somma di v.c. IID di tipo $B(p)$</td><td colspan="3">$i = 1,2,\cdots,n$</td></tr>
<tr><td colspan="5">$E(X) = E(X_1) + E(X_2) + \cdots E(X_n) = np$</td></tr>
<tr><td colspan="5">In virtù della identica distribuzione delle v.c.</td></tr>
<tr><td colspan="5">$V(X) = V(X_1) + V(X_2) + \cdots V(X_n) = np(1-p)$</td></tr>
<tr><td colspan="5">In virtù della indipendenza e della identica distribuzione delle v.c.</td></tr>
<tr><td colspan="5">Media e varianza dipendono da entrambi i parametri n e p che caratterizzano tale v.c.</td></tr>
<tr><td colspan="2" rowspan="2">$V(X) = np(1-p) = (1-p)E(X)$</td><td>$p = 0$</td><td>$V(X) = 0$</td><td>$E(X) = 0$</td></tr>
<tr><td>$0 < p < 1$</td><td colspan="2">$V(X) < E(X)$</td></tr>
<tr><td>$V(X) \leq 0.25n$</td><td>$p = 0.5 \Longrightarrow V(X) = 0.25n$</td><td>$p = 1$</td><td>$V(X) = 0$</td><td>$E(X) = n$</td></tr>
</table>

Altra **variabile casuale** discreta *associata* all'esperimento binomiale

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{X}{n}$$

$$X = X_1 + X_2 + \cdots + X_n \sim Bin(n,p)$$

"**Frequenza relativa** (proporzione) **di successi** nella sequenza binomiale"

La v.c. $\overline{X}$ "conta" i successi ottenuti nell'esperimento binomiale
e li *rapporta* al numero n di sottoprove Bernoulliane

$$\overline{x} = \frac{x}{n} \in \left\{\frac{0}{n}, \frac{1}{n}, \cdots, \frac{n}{n}\right\} = \left\{0, \frac{1}{n}, \cdots, 1\right\}$$

$\left\{0, \frac{1}{n}, \cdots, 1\right\}$ **supporto** della v.c.

Evento $\left(\overline{X} = \overline{x}\right)$ **equivalente** all'evento $(X = x)$

$$P\left(\overline{X} = \overline{x}\right) = P(X = x)$$

$$f\left(\overline{x} = \frac{x}{n}; n, p\right) = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$x \in \{0, 1, \cdots, n\}$$

$$\overline{X} \sim \frac{1}{n} Bin(n,p)$$

La v.c. $\overline{X}$ ha distribuzione $Bin(n,p)$ *scalata* del fattore $\frac{1}{n}$

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

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

Impiego della v.c. Binomiale nella **logica inferenziale**			
X v.c. di popolazione di tipo Bernoulliano		Popolazione Bernoulliana	
Campionamento casuale di ampiezza n			
$X_1, X_2, \cdots, X_n$ v.c. **indipendenti** e **identicamente distribuite** (v.c. IID)			
$Y = X_1 + X_2 + \cdots + X_n$	$\overline{X} = \frac{Y}{n}$	$Y \sim Bin(n, p)$	$\overline{X} \sim \frac{1}{n} Bin(n, p)$
Y e $\overline{X}$ rispettivamente **numero**, e **proporzione**, **di successi** nel campione			
Le v.c. Y e $\overline{X}$ dipendono esclusivamente dal parametro p (parametro n **noto**)			
Parametro p (probabilità di successo) oggetto di **inferenza**			
Campionamento casuale di ampiezza n da **popolazione finita**			
Esperimento binomiale *assimilabile* all'estrazione **con riposizione** di n palline da un'urna contenente palline **colorate**, di cui *un certo numero* di un **dato colore** (es. bianco) e *le rimanenti* di un **altro colore** (es. nero)			

Altri impieghi della v.c. Binomiale

Giochi di sorte (cfr esempi di esperimenti Bernoulliani di pag. 13)

- Lanci **ripetuti** n volte di monete/dadi *non truccati*
- Estrazioni **ripetute** n volte da un mazzo di carte *non truccate*
- Altri giochi di sorte **ripetuti** n volte in *condizioni simili*

Esperimenti binomiali *assimilabili* all'estrazione **con riposizione** di n palline da un'urna contenente palline **colorate**, di cui *un certo numero* di un **dato colore** (es. bianco) e *le rimanenti* di un **altro colore** (es. nero)

In questi casi (oltre ad n) il parametro p del modello è **noto**
(valore *assegnato* in base alla concezione **classica** o **frequentista** di probabilità)

Interesse per il **numero** (o la **proporzione**) **di successi** nella sequenza binomiale

Esempio 1	Campionamento casuale ($n = 3$) da una classe di 50 studenti

Popolazione Bernoulliana	$X =$ "Sesso" $\sim B(p)$	X_1, X_2, X_3 v.c. IID $B(p)$

p è la probabilità (o proporzione) di **successo** (lo studente estratto è **femmina**)

$f(x_i; p) = p^{x_i}(1-p)^{1-x_i}$	$x_i \in \{0,1\}$	$f(x_i, p) = 0$ altrove	$i = 1,2,3$

Logica inferenziale

$Y = X_1 + X_2 + X_3 \sim Bin(3, p)$	$\overline{X} = \frac{Y}{3} \sim \frac{1}{3} Bin(3, p)$
$y \in \{0,1,2,3\}$	$\overline{x} = \frac{y}{3} \in \left\{0, \frac{1}{3}, \frac{2}{3}, 3\right\}$
Numero **campionario** di successi	Proporzione **campionaria** di successi

Y e $\overline{X}$ sono v.c. **campionarie** (somma campionaria e media campionaria)

Parametro p *valutato* sulla base del **campione osservato**

Ad esempio (1,0,1)	$y = 2$	$\overline{x} = 2/3$	**Stima puntuale** di p

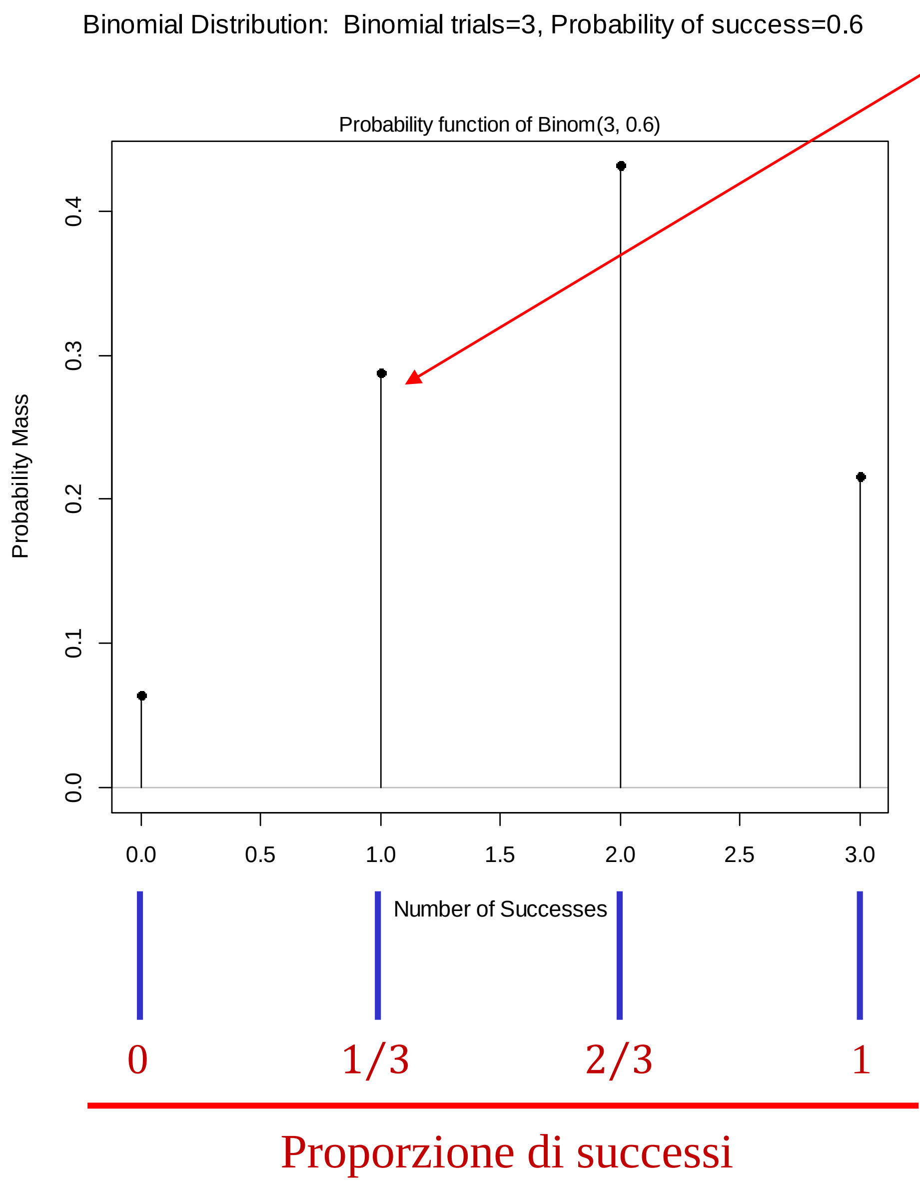
Logica probabilistica

Valore del parametro p *assegnato* usando la definizione **classica** di probabilità

$p = P(X = 1) = 0.6$ $1 - p = 0.4$	X_1, X_2, X_3 v.c. IID $B(0.6)$	Possibile uso della definizione **frequentista** di probabilità

$f(x_i; 0.6) = f(x_i) = 0.6^{x_i} 0.4^{1-x_i}$	$x_i \in \{0,1\}$	$f(x_i) = 0$ altrove	$i = 1,2,3$
$E(X_i) = p = 0.6$	$V(X_i) = p(1-p) = 0.24$		

Numero campionario di successi		**Proporzione** campionaria di successi	
$Y = X_1 + X_2 + X_3 \sim Bin(3, 0.6)$		$\overline{X} = \frac{Y}{3} \sim \frac{1}{3} Bin(3, 0.6)$	
$y \in \{0,1,2,3\}$		$\bar{x} = \frac{y}{3} \in \left\{0, \frac{1}{3}, \frac{2}{3}, 3\right\}$	
$f(y; 3, 0.6) = f(y) = \binom{3}{y} 0.6^y 0.4^{3-y}$		$f\left(\bar{x} = \frac{y}{3}\right) = f(y) = \binom{3}{y} 0.6^y 0.4^{3-y}$	
$E(Y) = 1.8$	$V(Y) = 0.72$	$E(\overline{X}) = 0.6$	$V(\overline{X}) = 0.08$



Evento $\{Y = 1\}$

Numero di femmine nel campione è **1**

$$P(Y = 1) = f(1) = \binom{3}{1} 0.6^1 0.4^{3-1} = 0.288$$

Cfr Parte 2, pagg. 43 e 58

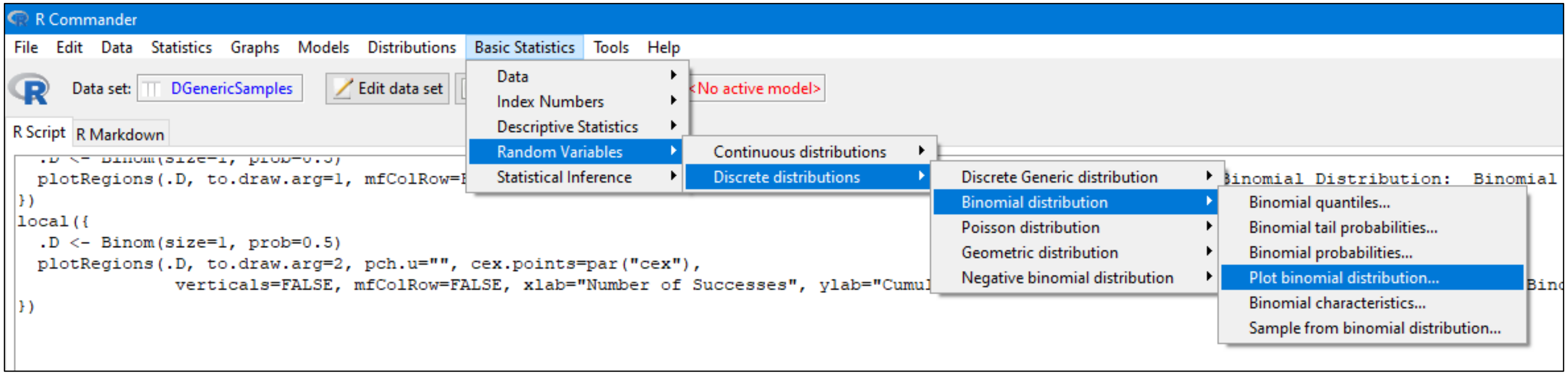
È anche la probabilità che la **proporzione** campionaria di femmine sia $1/3$

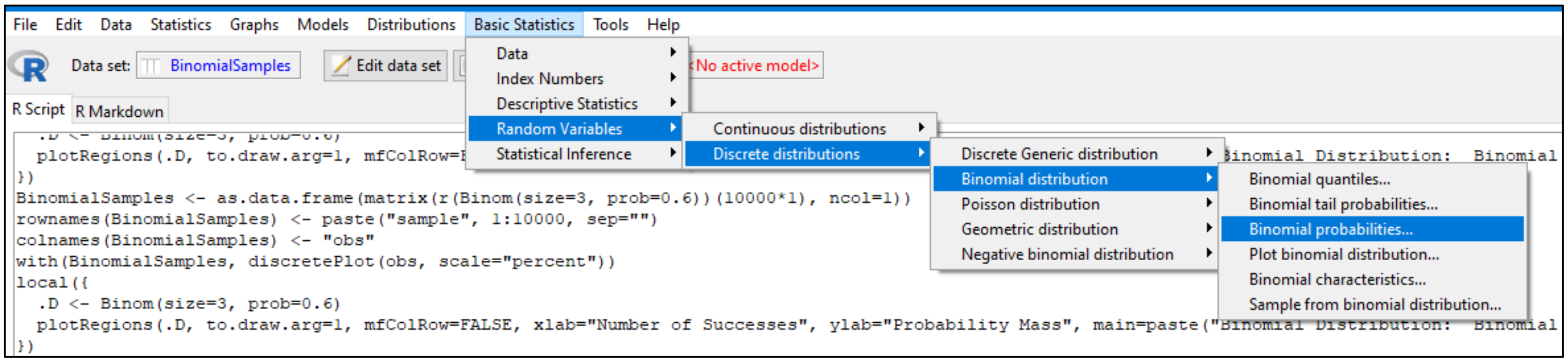
Calcolo di probabilità
esteso ad altri eventi di interesse

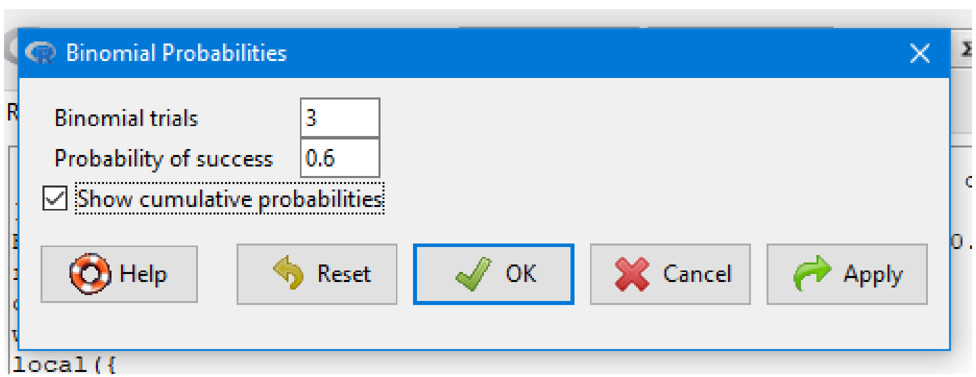
Per una generica v.c. $X \sim Bin(n, p)$

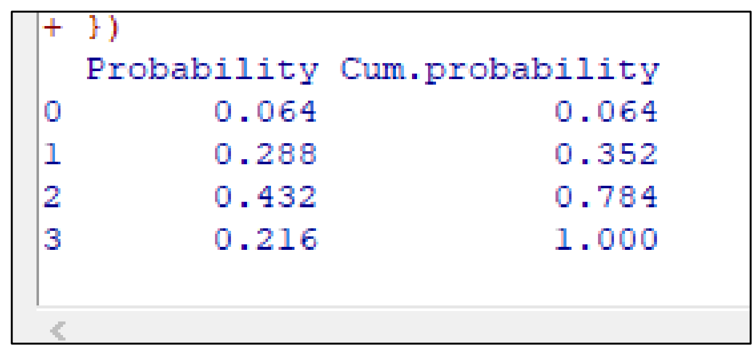
$p > 0.5$	$E(Z^3) < 0$	Asimmetria negativa
$p = 0.5$	$E(Z^3) = 0$	Simmetria
$p < 0.5$	$E(Z^3) > 0$	Asimmetria positiva

Al crescere di n la distribuzione tende ad essere **simmetrica**





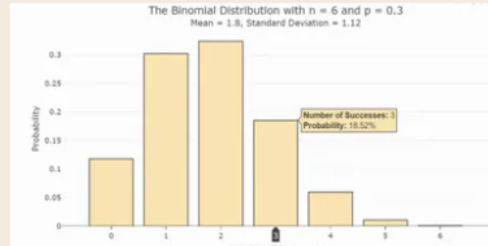




	Probability	Cum.probability
0	0.064	0.064
1	0.288	0.352
2	0.432	0.784
3	0.216	1.000

Distributions: B

Binomial Distribution



Explore how the shape of the Binomial distribution depends on the parameter n (the sample size) and p (the probability of success in a Bernoulli trial). Find and visualize probabilities of various kinds.

Nor

See
distr
stan
and
(criti

The Binomial Distribution | Explore | Find Probabilities | Formulas and Properties | Find Quantiles

The binomial distribution gives probabilities for the *number of successes* observed in *n* Bernoulli trials, each with success probability *p*.

Change the values of *n* and *p* to see how the shape of the distribution changes.

Number of Bernoulli Trials (n):

3

Probability of Success (p):

0.6

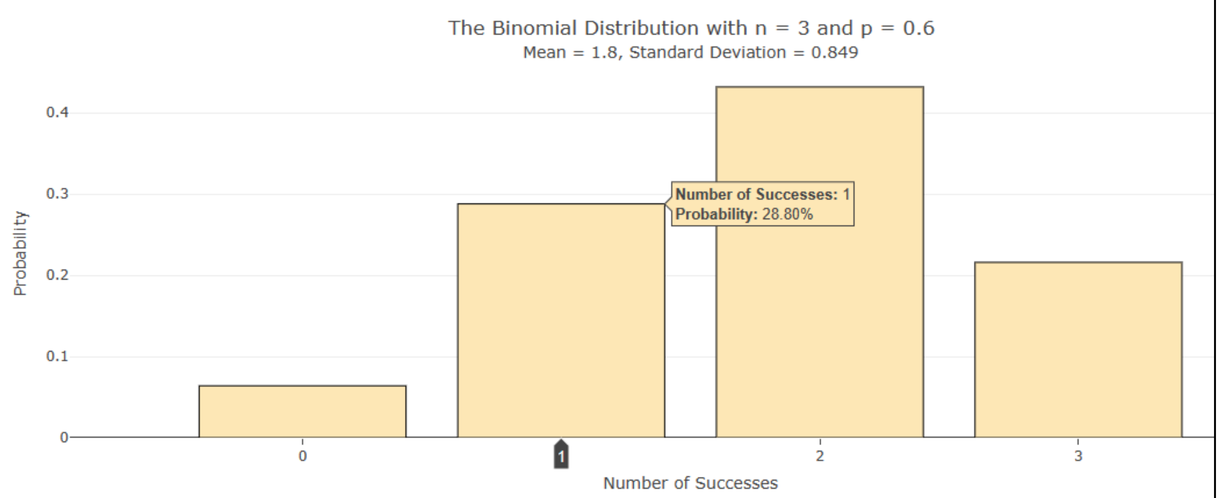
☑ Enter Numbers for n and p

☐ Overlay Normal Distribution

Hover over a bar in the graph to see the probability for the corresponding number of successes, or consult the table below.

Probability Table:

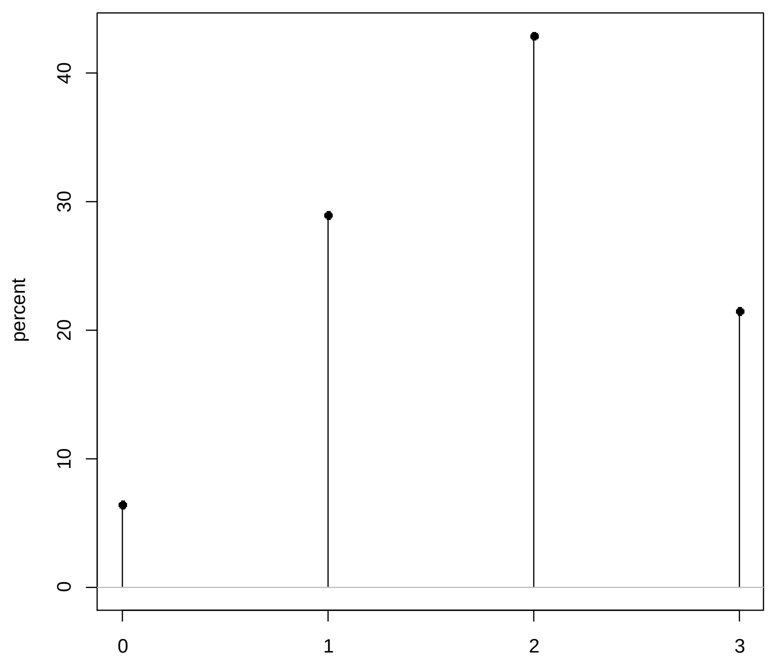
x	P(X = x)
0	0.0640
1	0.2880
2	0.4320
3	0.2160

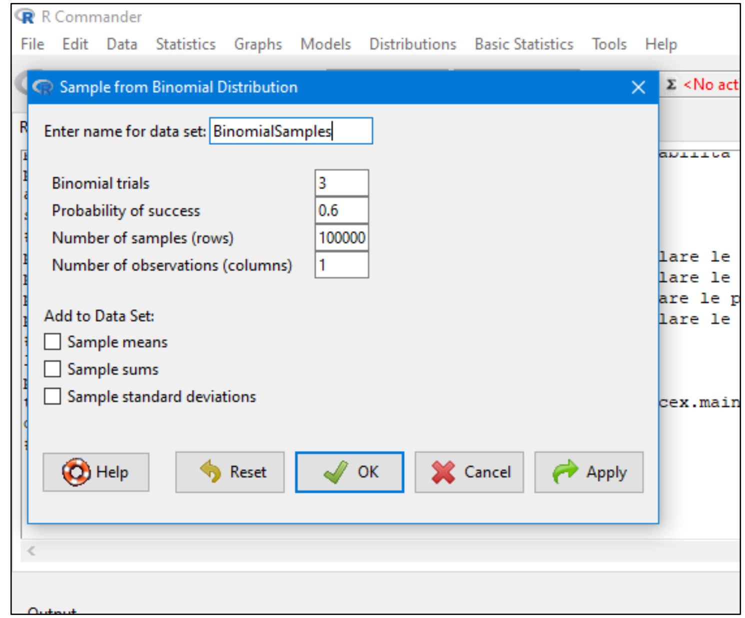


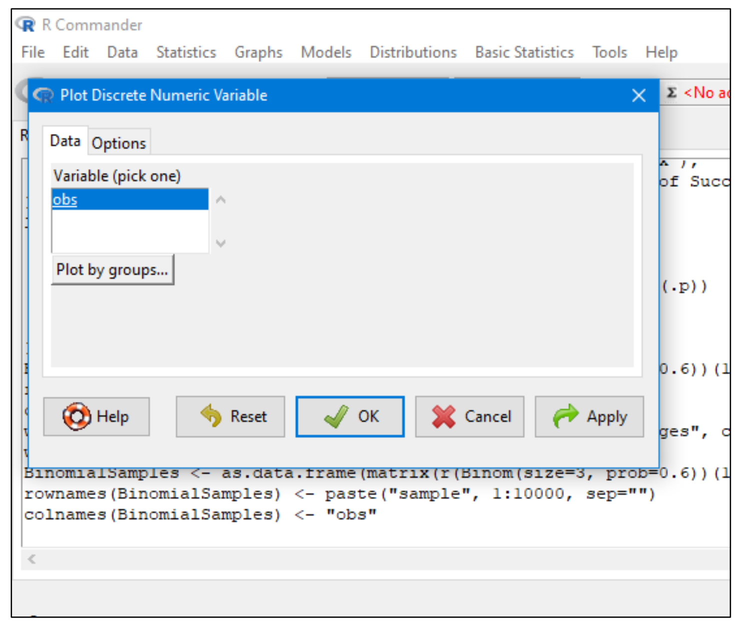
Logica frequentista

Distribuzione empirica della v.c. *Y*

Grafico ad aste dei valori di *Y*



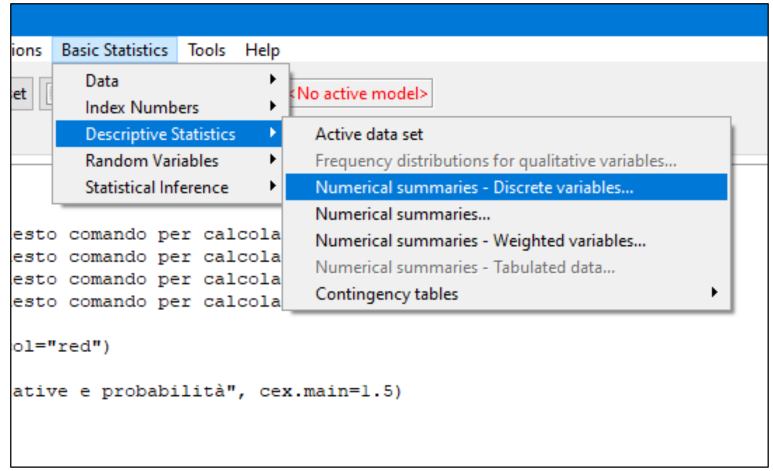




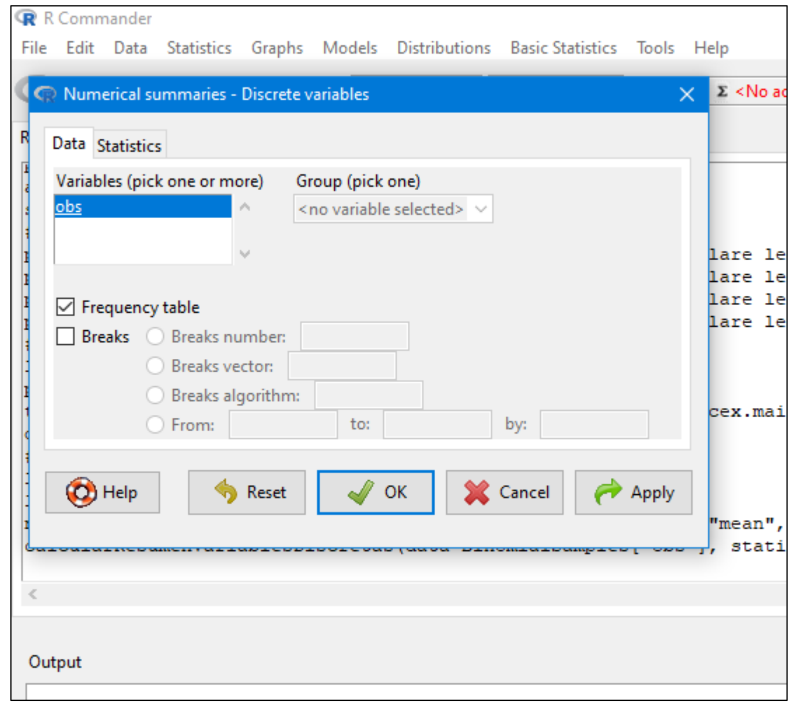
Distribuzione empirica della v.c. *Y*

Sintesi numeriche

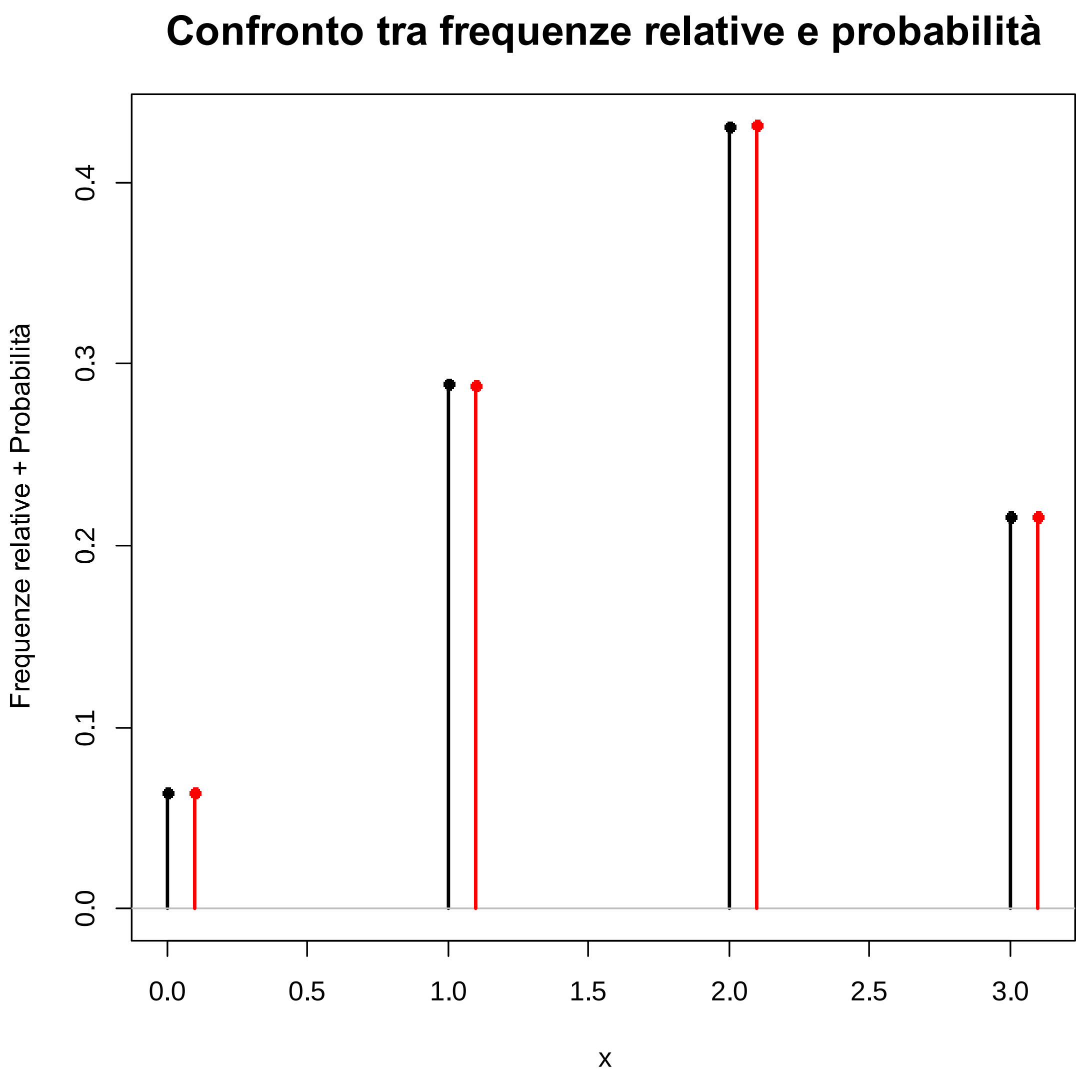
Confronto con le corrispondenti
probabilità e **sintesi probabilistiche**




```
------------------------------

Numerical summary:
    mean        sd IQR   skewness 0% 25% 50% 75% 100%      n
 1.79937 0.8487431   1 -0.2333682  0   1   2   2    3 100000

------------------------------

Frequency distribution for discrete variables:

Variable: obs
     ni    fi     Ni    Fi
0  6395 0.064   6395 0.064
1 28877 0.289  35272 0.353
2 43124 0.431  78396 0.784
3 21604 0.216 100000 1.000
N= 100000
```



```
#
# Distribuzione empirica vs distribuzione di probabilità - Modelli discreti
#
# Generazione in Rcmdr di m campioni di ampiezza 1 da una popolazione discreta
# Ad esempio, 100000 campioni di ampiezza 1 da una popolazione Binom(3,0.6)
#
BinomialSamples <- as.data.frame(matrix(r(Binom(size=3, prob=0.6))(100000*1), ncol=1))
rownames(BinomialSamples) <- paste("sample", 1:100000, sep="")
colnames(BinomialSamples) <- "obs"
#
data <- BinomialSamples    # inserire il nome del data frame
attach(data)
y <- as.vector(table(obs))
y <- y/sum(y)
x <- sort(unique(obs))
plot(x, y, type="h", lwd=2.5, ylab="Frequenze relative + Probabilità", xlim=c(min(x), max(x)+0.11), ylim=c(0, max(y)))
points(x, y, pch = 16)
abline(h = 0, col = "gray")
shift=0.1
#
p <- dbinom(x,3,0.6)      # lanciare questo comando per calcolare le probabilità - Binomiale
#
lines( x+shift, p, type="h", lwd=2.5, col="red")
points(x+shift, p, pch = 16, col="red")
title(main="Confronto tra frequenze relative e probabilità", cex.main=1.5)
detach(data)
#
```


Confronto tra frequenze relative e probabilità
Frequenze relative + Probabilità
0.0
0.1
0.2
0.3
0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x

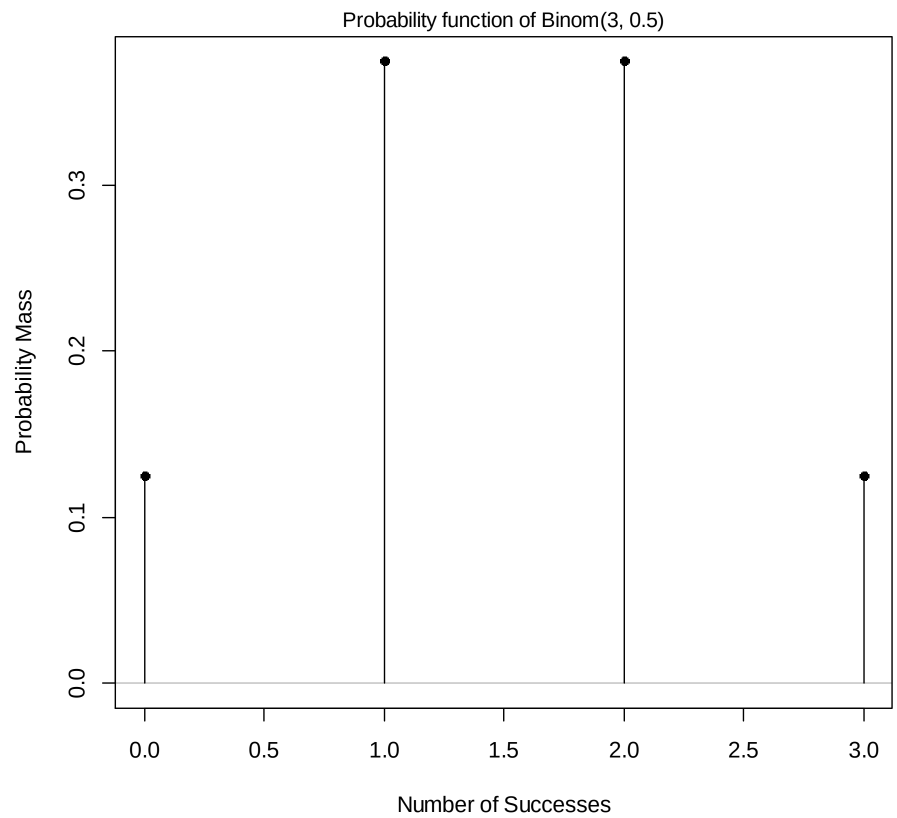
Esempio 9	Triplice lancio ($n = 3$) di una moneta **bilanciata** (o lancio di tre monete **bilanciate**)

v.c. X di *riferimento*	**Faccia** ottenuta (0/1 per C/T) in un lancio	$X \sim B(0.5)$
X_i = "**Faccia** ottenuta (0/1 per C/T) nel lancio $i-esimo$		$i = 1,2,3$

X_1, X_2, X_3 v.c. IID $B(0.5)$	**Logica probabilistica** (parametro p **noto**)
Numero di teste *nei tre lanci*	**Proporzione** di teste *nei tre lanci*
$Y = X_1 + X_2 + X_3 \sim Bin(3, 0.5)$	$\bar{X} = \frac{Y}{3} \sim \frac{1}{3} Bin(3, 0.5)$
$y \in \{0,1,2,3\}$	$\bar{x} = \frac{y}{3} \in \left\{0, \frac{1}{3}, \frac{2}{3}, 3\right\}$
$f(y) = \binom{3}{y} 0.5^y 0.5^{3-y}$	$f\left(\bar{x} = \frac{y}{3}\right) = f(y) = \binom{3}{y} 0.5^y 0.5^{3-y}$

$E(Y) = 1.5$	$V(Y) = \mathbf{0.75}$	$E(\bar{X}) = 0.5$	$V(\bar{X}) = \mathbf{0.08\overline{3}}$

Binomial Distribution: Binomial trials=3, Probability of success=0.5



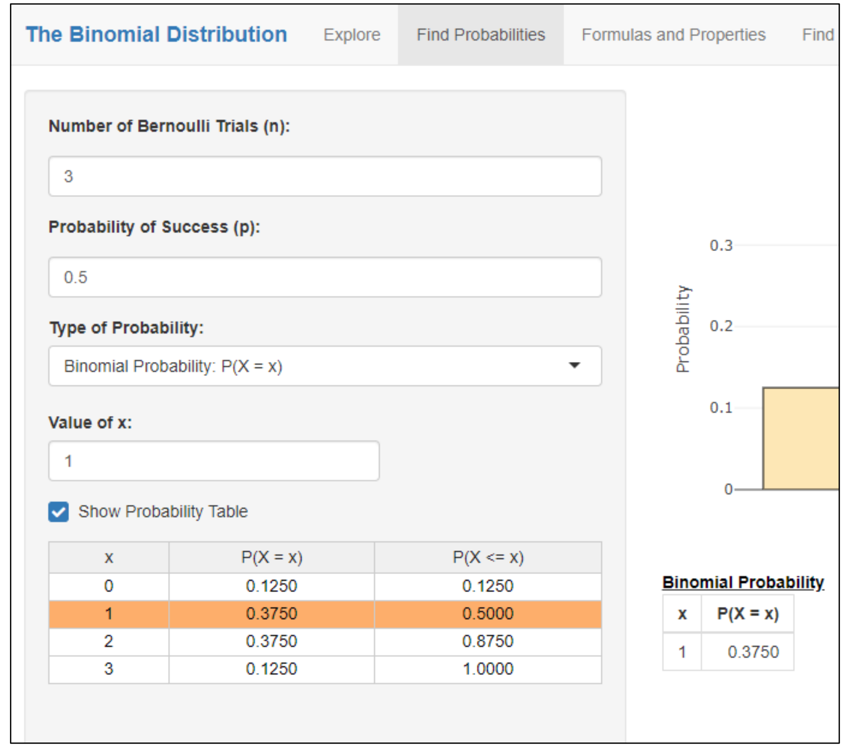
Distribuzione simmetrica
$Me = E(X) = 1.5$

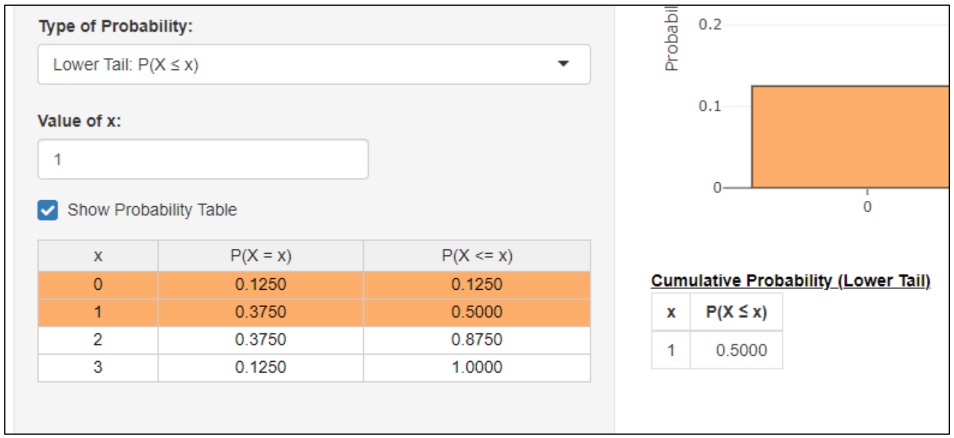
Se la moneta è **truccata**	$p \neq 0.5 \neq 1 - p$

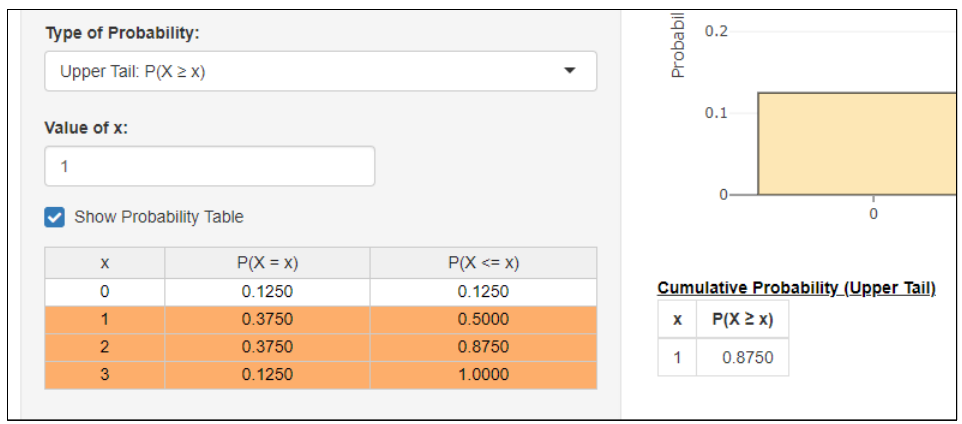
Eventi di interesse	**Numero** di teste nei tre lanci	
$C = \{Y = 1\}$	$D = \{Y \leq 1\}$	$E = \{Y \geq 1\}$

Cfr Parte 2, pag. 37

Eventi *equivalenti* in termini di **proporzione**







Modelli di probabilità per variabili casuali discrete			Quadro riassuntivo			
	Uniforme		**Bernoulli**		**Binomiale**	
Notazione	$X \sim Ud(N)$		$X \sim B(p)$		$X \sim Bin(n,p)$	
Funzione di probabilità	$\frac{1}{N}$		$p^x(1-p)^{1-x}$		$\binom{n}{x} p^x (1-p)^{n-x}$	
Supporto	$\{1,2,\cdots,N\}$		$\{0,1\}$		$\{0,1,\cdots,n\}$	
Parametri	N		p		n, p	
Spazio parametrico	$\{1,2,\cdots,\infty\}$		$[0,1]$		$\{(n,p)\}$ $n \geq 1 \quad p \in [0,1]$	
Media e varianza	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	p	$p(1-p)$	np	$np(1-p)$
Impiego principale	**Giochi di sorte**		**v.c. di popolazione**		**v.c. campionaria**	