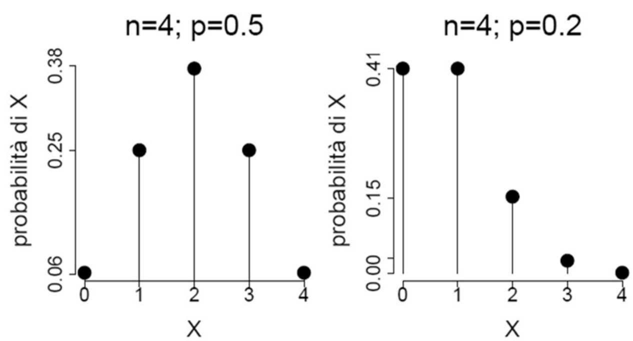
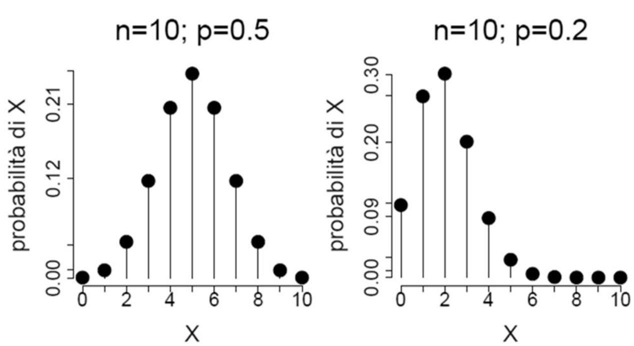
- Le distribuzioni binomiali sono una famiglia di distribuzioni di probabilità.
- Una variabile aleatoria binomiale X si osserva nelle seguenti condizioni:
  - c'è un numero fisso di n osservazioni;
  - le osservazioni sono *indipendenti* le une dalle altre;
  - ciascuna osservazione corrisponde ad uno di due esiti possibili, convenzionalmente chiamati "**successo**" e "**insuccesso**";
  - la probabilità di un successo, denotata da p, rimane costante per tutte le osservazioni.

- La variabile X corrisponde al numero di successi in n prove e dunque è un numero compreso tra 0 e n.
- La distribuzione binomiale con parametri n e p fornisce l'elenco delle probabilità associate a ciascun possibile valore X.
- Alcune distribuzioni binomiali sono rappresentate graficamente nelle figure seguenti.



Numero di successi in $n = 4$ esperimenti



Numero di successi in $n = 10$ esperimenti

- Consideriamo l'esperimento che consiste nel lanciare due monete oneste e nel contare il numero X di eventi "testa".
- Lo spazio campione dell'esperimento è $\Omega = \{TT, TC, CT, CC\}$ e ciascun evento nello spazio campione ha probabilità $1/4$.
- La distribuzione di probabilità di X è

x_i	p_i
0	$1/4$
1	$1/2$
2	$1/4$
somma	1

- **Questa è la distribuzione binomiale per** $n = 2$ **e** $p = 0.5$.

- **Consideriamo ora la distribuzione binomiale con** $n = 4$ **e** $p = 0.2$. [cfr. Figura]
- Possiamo immaginare questa distribuzione come quella che descrive l'esperimento in cui vengono lanciate 4 monete *disoneste*, tali per cui la probabilità di osservare "testa" è 0.2.
- Come in precedenza, contiamo il numero di eventi "testa" effettivamente osservati.

- Per calcolare la distribuzione di X potremmo elencare tutti gli eventi dello spazio campione Ω, contando gli eventi (sul totale) che soddisfano ciascuna modalità di X.
  - In questo caso lo spazio campione contiene $2 \times 2 \times 2 \times 2 = 2^4 = 16$ eventi.
  - Questa strategia (*punto campione*), però, diventa difficilmente praticabile al crescere di n.
  - Possiamo invece usare un metodo diverso. Consideriamo, ad esempio, $P(X = 2)$, ovvero la probabilità dell'evento che consiste nell'osservare due esiti "testa" in 4 lanci

- Due teste vengono osservate, per esempio, nella sequenza $TTCC$. Dato che i lanci sono indipendenti, tale evento ha una probabilità di

$$P(TTCC) = 0.2 \times 0.2 \times 0.8 \times 0.8 = 0.2^2 \times 0.8^2 = 0.0256$$

.

- Due teste vengono osservate, però, anche nella sequenza $TCTC$, a cui associamo la stessa probabilità:

$$P(TCTC) = 0.2 \times 0.8 \times 0.2 \times 0.8 = 0.2^2 \times 0.8^2 = 0.0256$$

- Qualunque sequenza contenente 2 teste e 2 croci, non importa in quale ordine, avrà probabilità

$$0.2^2 \times 0.8^2 = 0.0256.$$

- In quanti modi diversi si possono ottenere 2 teste in 4 lanci?
- Dato che lo spazio campione è piccolo, possiamo semplicemente contare tutte le combinazioni:

$$TTCC \;\; CTTC \;\; CCTT \; TCCT \; TCTC \;\; CTCT$$

- In conclusione, la probabilità di ottenere due teste in 4 lanci è

$$\begin{aligned} P(X=2) = \quad & P\big(TTCC \cup CTTC \cup CCTT \cup TCCT \cup TCTC \cup CTCT\big) \\ = \quad & 0.0256 + 0.0256 + 0.0256 + 0.0256 + 0.0256 + 0.0256 \\ = \quad & 6 \times 0.0256 \\ = \quad & \text{combinazioni} \times \text{prob. di ogni sequenza} = 0.1536 \end{aligned}$$

- Esaminiamo ora una regola generale, basata sul calcolo combinatorio, che consente di trovare le probabilità della distribuzione binomiale.
- Il numero di possibili combinazioni di k successi e $n-k$ insuccessi è data dal coefficiente binomiale:

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

dove $n! = n \times (n-1) \times \ldots \times 3 \times 2 \times 1$. Per convenzione $0! = 1$; quindi

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

- La probabilità di ciascuna sequenza di k successi e $n-k$ insuccessi in n prove è:

$$p^k \left(1-p\right)^{n-k}$$

  - si ricordi che $a^0 = 1$ per ciascun numero $a \neq 0$.
  - Se k o $n-k$ sono uguali a 0, dunque, il fattore corrispondente p^0 o $(1-p)^0$ è 1.

- Utilizzando i due risultati precedenti, otteniamo la formula per calcolare le probabilità della distribuzione binomiale:

$$\begin{aligned} P(X=k) &= \binom{n}{k} p^k \left(1-p\right)^{n-k} \\ &= \frac{n!}{k!(n-k)!} p^k \left(1-p\right)^{n-k} \end{aligned}$$

Distribuzione binomiale

- Per esempio, la probabilità di ottenere $k = 2$ successi in $n = 4$ prove con probabilità di successo $p = 0.2$ sarà:

$$\begin{aligned} P(X=2) = & \binom{4}{2} 0.2^2 (1-0.2)^{4-2} \\ = & \frac{4!}{2!(4-2)!} 0.2^2 0.8^2 \\ = & \frac{4 \times 3 \times 2 \times 1}{(2 \times 1)(2 \times 1)} 0.2^2 0.8^2 \\ = & 0.1536 \end{aligned}$$

- La distribuzione binomiale con parametri $n = 4$ e $p = 0.2$ diventa quindi

k	$P(X = k)$
0	0.4096
1	0.4096
2	0.1536
3	0.0256
4	0.0016
somma	1.0000

Media e varianza

- La distribuzione di Bernoulli o *bernoulliana* è un caso speciale della distribuzione binomiale, dove $n = 1$.
- Simbolicamente, $X \sim B(1, p)$ ha lo stesso significato di $X \sim Bern(p)$.
- Similmente, ciascuna distribuzione binomiale, $B(n, p)$, è la distribuzione della somma di n prove di bernoulli tra loro indipendenti, ciascuna con la stessa probabilità p di *successo*.

- La variabile aleatoria bernoulliana X_bern assume quindi due possibili esiti, 0 (nessun successo) e 1 (successo), con

$$
\begin{aligned}
P(X_{Bern} = 1) = &\quad p \\
P(X_{Bern} = 0) = &\quad 1 - p.
\end{aligned}
$$

- la cui media e varianza si ricavano facilmente come

$$
\begin{aligned}
E(X_{Bern}) = &\quad 1 \times p + 0 \times (1 - p) = p \\
Var(X_{Bern}) = &\quad (1 - p)^2 p + (0 - p)^2 (1 - p) \\
= &\quad (1 - p)\big[(1 - p)p + p^2\big] \\
= &\quad (1 - p)\big[p - p^2 + p^2\big] = p(1 - p)
\end{aligned}
$$

- Come detto, una **variabile aleatoria binomiale** $X \sim B(n, p)$, con parametri n e p, è un processo di Bernoulli, ossia una serie di n variabili aleatorie indipendenti X_i di uguale distribuzione di Bernoulli ($Bern(p)$), dette appunto prove di Bernoulli.
- Le modalità di $X \sim B(n, p)$ sono la somma degli esiti di "successo" (o 1 o 0) in ciascuna prova $X_i \sim Bern(p)$:

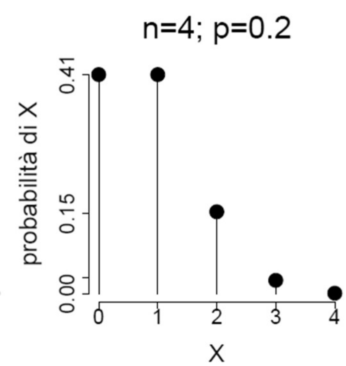
$$
X_B = X_{Bern\,1} + X_{Bern\,2} + \ldots + X_{Bern\,n}.
$$

Media e varianza

- Perciò, la media e varianza della variabile binomiale X sono presto ricavate:

$$\begin{aligned}\mathbf{E}(\mathbf{X_B}) = & \quad E(X_{Bern\ 1}) + E(X_{Bern\ 2}) + \ldots + E(X_{Bern\ n}) = \mathbf{np} \\ \mathbf{Var}(\mathbf{X_B}) = & \quad Var(X_{Bern\ 1}) + Var(X_{Bern\ 2}) + \ldots + Var(X_{Bern\ n}) \\ = & \quad \mathbf{np(1-p)}\end{aligned}$$

- La distribuzione binomiale con parametri $n = 4$ e $p = 0.2$ avrà quindi



$$\mu = 4 \times 0.2 = 0.8$$
$$\sigma^2 = 4 \times 0.2 \times 0.8 = 0.64$$