STATISTICA E ANALISI DEI DATI

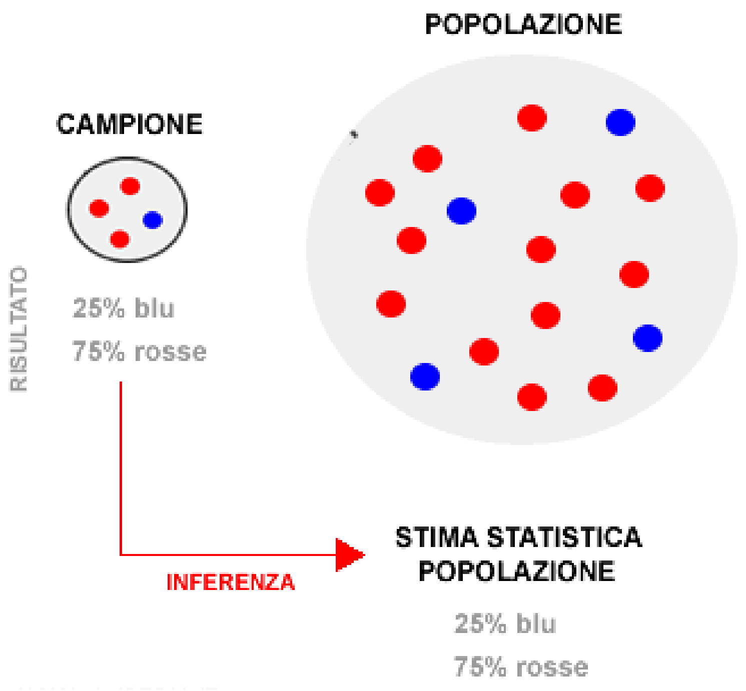
Seconda Parte

Dott. Stefano Cirillo
Dott. Luigi Di Biasi

a.a. 2025-2026

INTRODUZIONE

- L'indagine statistica è sempre effettuata su un insieme di entità (individui, oggetti,...) su cui si manifesta un fenomeno che si studia
- L'insieme di entità è detto **popolazione** o **universo** e può essere costituito da:
  - un numero finito di unità: **popolazione finita**
    - Si osserva la totalità delle entità della popolazione oppure un sottoinsieme di questa, detto **campione** estratto, per studiare le caratteristiche di una popolazione finita
  - un numero non finito di unità: **popolazione illimitata**
    - Data la dimensione, può invece essere studiata soltanto tramite un **campione** estratto



- **Spazio campionario** (Ω): insieme di tutti i possibili campioni
  - Può essere discreto o continuo a seconda della natura (discreta o continua) della variabile X

INFERENZA STATISTICA

- L'**inferenza statistica** ha lo scopo di estendere le misure ricavate dall'esame di un campione all'intera popolazione da cui il campione è stato estratto

1. **Problema dell'inferenza statistica:**
   - Si considera una popolazione descritta da una **variabile aleatoria** X, che rappresenta un fenomeno osservabile (come l'altezza delle persone, il risultato di un esperimento, ecc.)
   - La distribuzione di X è **nota nella forma**, ma contiene un parametro ϑ (ad esempio, la media o la deviazione standard) che non è noto e che desideriamo stimare.
2. **Spazio dei Parametri:**
   - Il parametro sconosciuto ϑ appartiene a uno **spazio dei parametri** Θ, che rappresenta l'insieme di tutti i valori possibili per ϑ
     - Ad esempio, se ϑ è la probabilità di successo in un processo binomiale, Θ sarà l'intervallo [0,1]
   - Il valore reale di ϑ **non è noto**, e il nostro obiettivo è stimarlo o fare inferenze su di esso.
   - Si desidera studiare una popolazione descritta da una **variabile aleatoria** osservabile X la cui funzione di distribuzione ha una forma nota ma contiene un parametro $\vartheta \in \Theta$ (spazio dei parametri) non noto

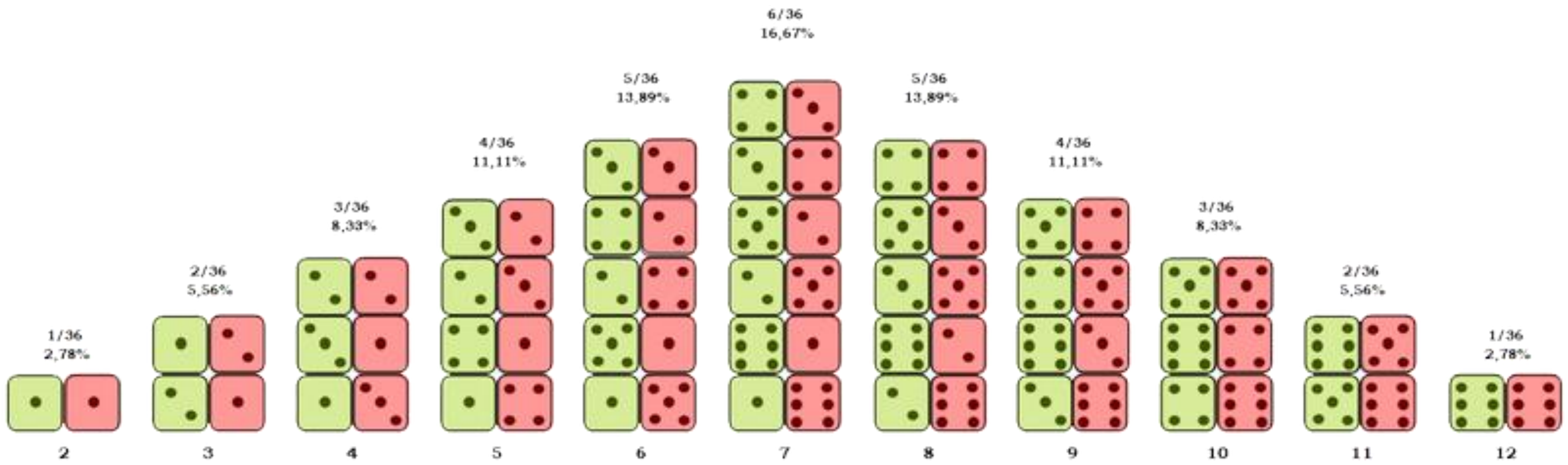
INFERENZA STATISTICA

3. **Osservabilità**:

- Osservabile: si possono osservare i valori assunti dalla variabile aleatoria X (ad esempio, eseguendo un esperimento casuale) e quindi il parametro non noto è presente soltanto nella **legge di probabilità** (funzione di distribuzione, funzione di probabilità, …)

- Possiamo osservare i valori assunti dalla variabile aleatoria X, ad esempio **raccogliendo dati** attraverso un esperimento o un campionamento dalla popolazione

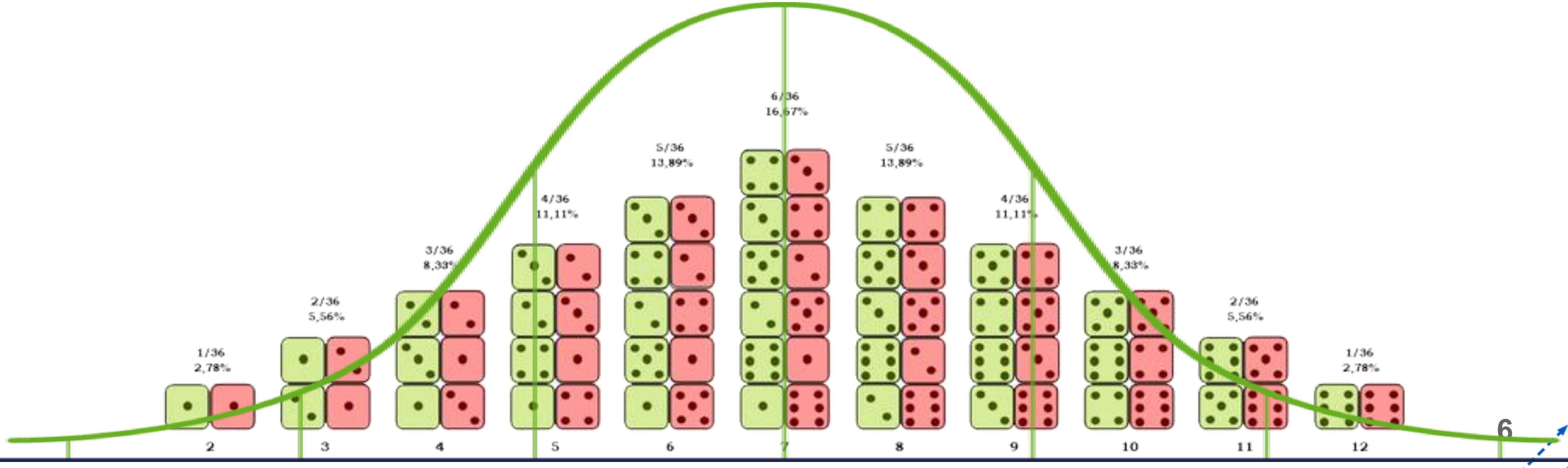
    - Tuttavia, il parametro ϑ non è direttamente osservabile; è presente solo nella forma della distribuzione di probabilità di X (es., la funzione di densità o la funzione di probabilità di X), ed è questa relazione che vogliamo sfruttare per stimare ϑ



INFERENZA STATISTICA

4. **Distribuzione Completamente Specificata**:

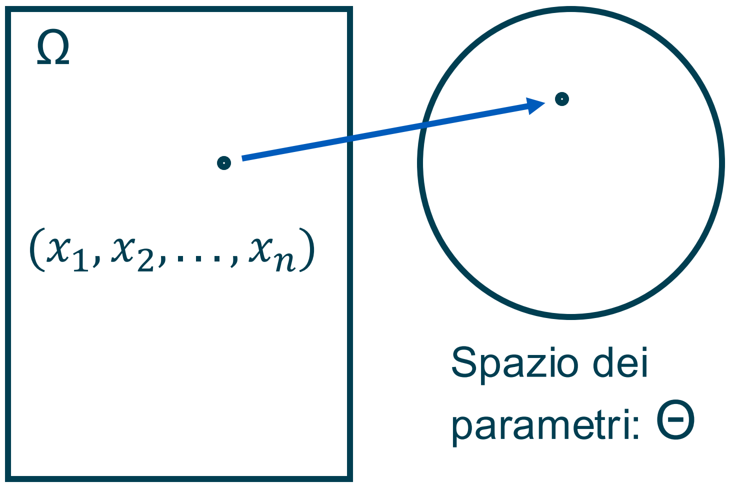
- Se il valore di ϑ fosse noto, la distribuzione di probabilità di X sarebbe completamente determinata, e potremmo descrivere esattamente la variabilità di X all'interno della popolazione.
- Tuttavia, poiché ϑ è **sconosciuto**, per ottenere informazioni sul parametro non noto ϑ della popolazione, si può fare uso dell'inferenza statistica considerando un **campione** $(x_1, x_2, ..., x_n)$ estratto dalla popolazione e effettuando su tale campione delle opportune **misure**



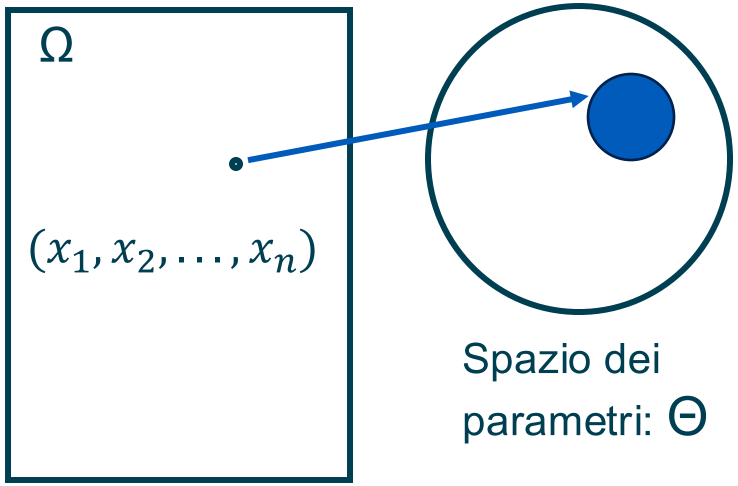
STIMA DEI PARAMETRI

- L'inferenza statistica si basa su due metodi fondamentali di indagine:

1. **Stima dei parametri**: sulla base del campione si assegna al parametro di interesse (ϑ) un valore (**stima puntuale**) o un insieme di valori (**stima per intervallo**)
   - **Scopo**: Fornire una migliore approssimazione possibile del parametro sconosciuto ϑ che caratterizza la popolazione



Stima puntuale



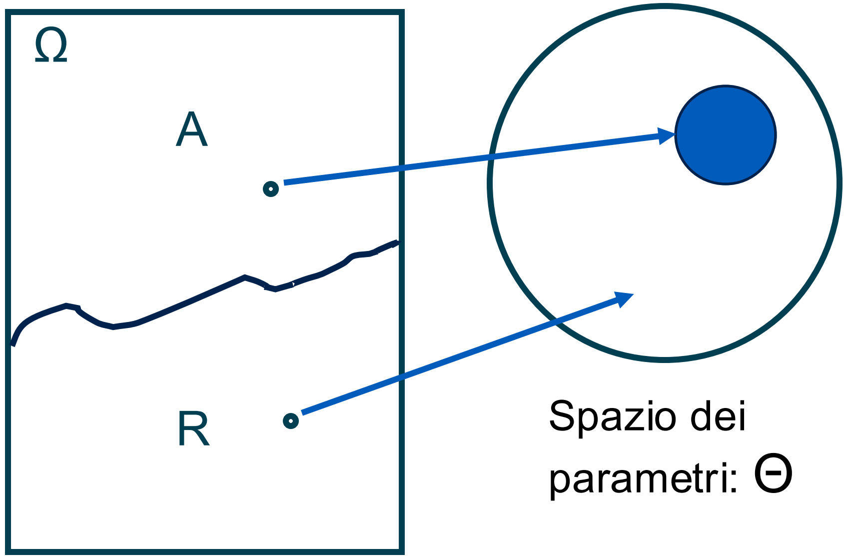
Stima per intervallo (di confidenza)

- Esempio: sulla base di un campione di soggetti si stima che l'altezza media degli italiani è pari a 175,36cm o è compresa nell'intervallo (173; 178)

VERIFICA DELLE IPOTESI

2. **Verifica delle ipotesi**: si formula un'ipotesi, o congettura, sul parametro di interesse (ϑ) e si verifica, sulla base del campione, se tale ipotesi è o meno accettabile (o plausibile)

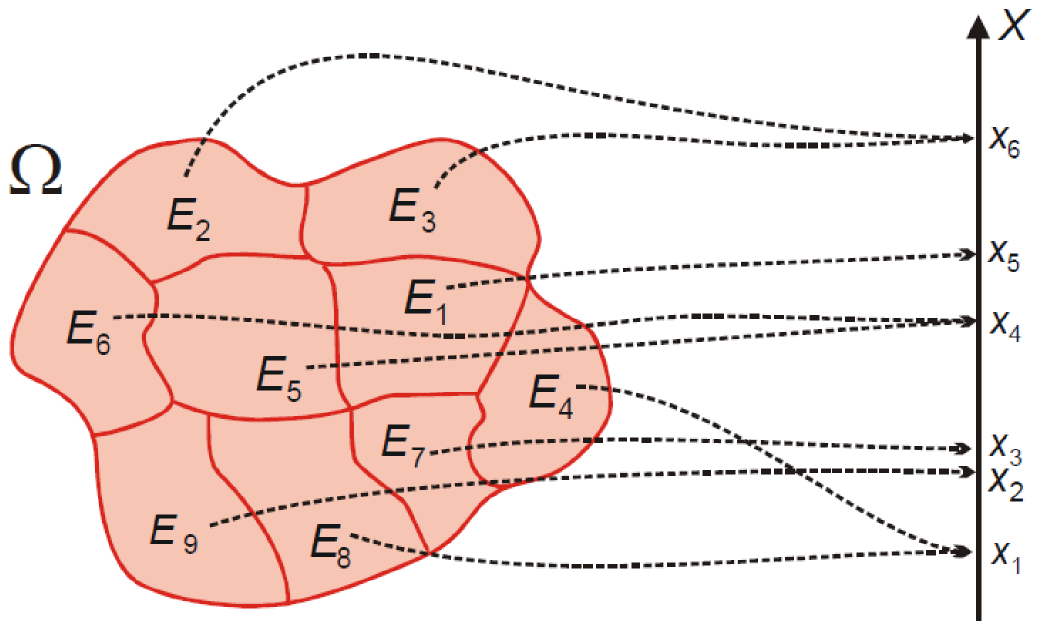
- **Scopo:** Determinare, sulla base di un campione, se una certa affermazione (ipotesi) riguardo al parametro sconosciuto può essere ritenuta **plausibile**
- **Problema della verifica delle ipotesi** consiste allora nel suddividere, mediante opportuni criteri, l'insieme dei possibili campioni in due sottoinsiemi, un sottoinsieme <u>A di accettazione dell'ipotesi</u> nulla e un sottoinsieme <u>R di rifiuto dell'ipotesi</u> nulla
- Lo spazio dei parametri Θ si suddivide in due sottoinsiemi disgiunti Θ_0 e $\Theta_1$tali che $\Theta = \Theta_1 \cup \Theta_2$
  - L'ipotesi H_0 soggetta a verifica su ϑ consiste nell'affermare che $\vartheta \in \Theta_0$ ed è detta ipotesi nulla, mentre nell'ipotesi alternativa H_1 si assume invece che $\vartheta \in \Theta_1$



- <u>Esempio</u>: si formula l'ipotesi che l'altezza media degli italiani sia pari a 175cm e sulla base di 10 soggetti estratti casualmente si decide se tale ipotesi è plausibile o meno

VARIABILI ALEATORIE (O VARIABILI CASUALI)

- Una **variabile aleatoria** X è una funzione che associa a ogni elemento E dello **spazio campionario** Ω un unico numero reale possibile che è l'**esito** di un esperimento casuale
- Formalmente, una variabile aleatoria può essere definita come una funzione X che mappa ogni elemento E dello spazio degli esiti Ω (lo spazio campionario) in un numero reale X_E
- Esempi:
  - Somma dei punteggi nel lancio di due dadi
  - Numero di pezzi difettosi in un lotto
  - Variazione giornaliera nel rendimento di un titolo
  - Numero di prodotti di un certo tipo venduti giornalmente in un particolare punto vendita



Variabili Aleatorie (o Variabili Casuali)

- Una variabile aleatoria può essere classificata come:
  - **Discreta:** può assumere un insieme discreto (finito o numerabile) di numeri reali
  - **Continua**: può assumere tutti i valori compresi in un intervallo reale

$$\Omega \text{ Discreto} \longrightarrow \text{Variabile Aleatoria Discreta}$$

$$\Omega \text{ Continuo} \longrightarrow \text{Variabile Aleatoria Continua}$$

- Le variabili aleatorie sono descritte dalle loro **funzioni di distribuzione**, che specificano la probabilità di ciascun possibile valore (nel caso discreto) o la probabilità che la variabile assuma valori entro un certo intervallo (nel caso continuo)

VARIABILE ALEATORIA DISCRETA

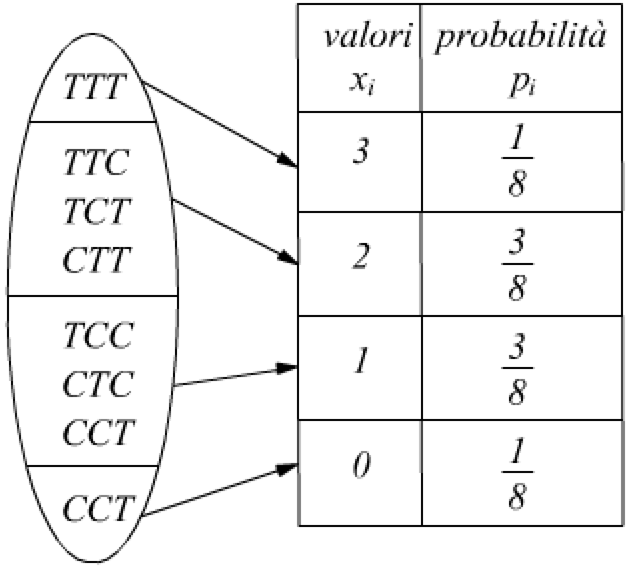
- Una variabile aleatoria **discreta** X assume un **numero finito** o al più numerabile di valori $(x_1, x_2, \dots, x_n)$
- La **funzione di probabilità** di una variabile aleatoria discreta X è una funzione che associa a ciascun valore possibile di X la probabilità che X assuma quel valore
  - Se denotiamo con $p(x_i)$ la funzione di probabilità di X, abbiamo che:
  $$p(x_i) = P(X = x_i)$$
  dove $P(X = x_i)$ è la probabilità che la variabile aleatoria X assuma il valore x_i
  - Quindi si ha che ad ogni valore x_i è associata la corrispondente probabilità $P(X = x_i)$, cioè
  $$(p_X(x_1),\ p_X(x_2), \dots, p_X(x_n))$$
- La **funzione di probabilità** deve verificare le due **proprietà**:
  - $P(x_i) \geq 0, \forall i$
  - La somma delle probabilità per tutti i valori possibili è: $\sum_i P(X_i) = 1$

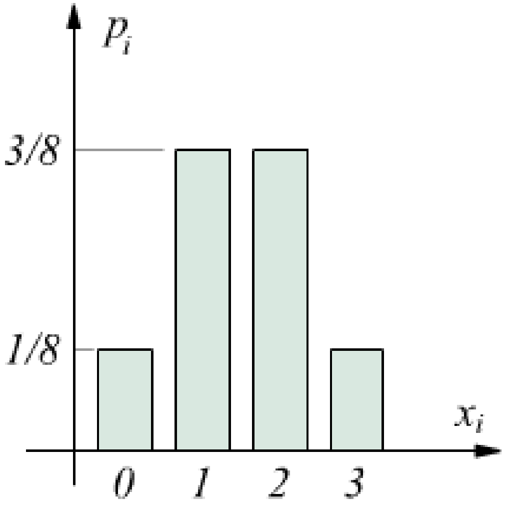
x_i	x_1	x_2	x_3	$\dots$	x_n
$P(X = x_i)$	$p(x_1)$	$p(x_2)$	$p(x_3)$	$\dots$	$p(x_n)$

Funzione di Probabilità (Esempio)

- <u>Esempio</u>:
  - Supponiamo di eseguire tre lanci successivi di una moneta e di voler rappresentare la distribuzione di probabilità della variabile
  - X = numero di teste che si possono presentare
  - L'insieme universo U è costituito da 8 eventi:

    U = {TTT, TTC, TCT, TCC, CTT, CTC, CCT, CCC}
  - come si vede, le cose da fare sono poche, bisogna solo determinare:
    - Quali valori può assumere la variabile
    - Con quale probabilità può assumere tali valori





Funzione di Probabilità (Esempio)

- Esempio:
  - In un sistema con una cache che memorizza temporaneamente i dati più richiesti, possiamo definire una **variabile aleatoria discreta** X che rappresenta l'accesso a specifici dati
  - Supponiamo che abbiamo un insieme di dati $x_1, x_2, \dots, x_n$ e che ogni dato abbia una probabilità diversa di essere richiesto, basata su dati storici di accesso
  - La funzione di probabilità $P(X = x_i)$ rappresenta la probabilità che il dato x_i sia richiesto dal sistema. Ad esempio:
    - Dato x_1: 30% delle richieste
    - Dato x_2: 25% delle richieste
    - Dato x_3: 20% delle richieste
    - Dato x_4: 15% delle richieste
    - Dato x_5: 10% delle richieste

Funzione di Probabilità (Esempio)

- Esempio:
  - In un sistema con una cache che memorizza temporaneamente i dati più richiesti, possiamo definire una **variabile aleatoria discreta** X che rappresenta l'accesso a specifici dati
  - Supponiamo che abbiamo un insieme di dati $x_1, x_2, \ldots, x_n$ e che ogni dato abbia una **probabilità diversa di essere richiesto**, basata su dati storici di accesso
  - La funzione di probabilità $P(X = x_i)$ rappresenta la probabilità che il dato x_i sia richiesto dal sistema. Ad esempio:
    - Dato x_1: 30% delle richieste
    - Dato x_2: 25% delle richieste
    - Dato x_3: 20% delle richieste
    - Dato x_4: 15% delle richieste
    - Dato x_5: 10% delle richieste

Questa funzione di probabilità potrebbe essere utilizzata per **ottimizzare le operazioni di cache**:
- Allocando maggiore spazio a x_1 e x_2, che hanno le probabilità più alte
- Aggiornando o rimpiazzando i dati nella cache con minore probabilità, come x_5

Funzione di Distribuzione

- La **funzione di distribuzione (o funzione di distribuzione cumulativa** (FDC)) $F(X)$ di una variabile aleatoria discreta X è una funzione che, dato un valore x_i, restituisce la probabilità che X assuma il un valore **minore o uguale** a x_i

$$F(x_i) = P(X \leq x_i) = \sum_i p(x_i)$$

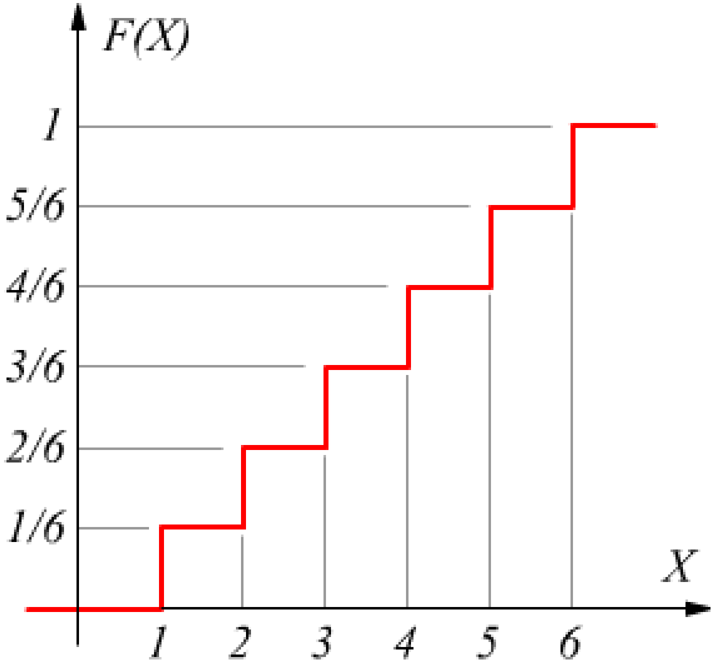
x_i	x_1	x_2	x_3	…	x_n
$P(X = x_i)$	$p(x_1)$	$p(x_2)$	$p(x_3)$	…	$p(x_n)$
$F(x_i)$	$p(x_1)$	$p(x_1) + p(x_2)$	$p(x_1) + p(x_2) + p(x_3)$	…	1

- **Monotona non decrescente**:
  - $F(x)$ è una funzione che non diminuisce mai al crescere di x, cioè se $x_1 < x_2$, allora $F(x_1) \leq F(x_2)$
- **Salti**:
  - Per una variabile aleatoria discreta, la funzione di distribuzione ha "salti" nei punti in cui la variabile può assumere valori specifici.

Funzione di Distribuzione (Esempio)

- Esempio: Quando si lancia un dado, tutte e 6 le facce hanno la stessa probabilità di presentarsi

x_i	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



- La funzione di distribuzione è pertanto una diretta conseguenza della distribuzione di probabilità

Funzione di Distribuzione (Esempio)

- Esempio: Supponiamo di monitorare i tempi di risposta di un server, misurati in secondi, per un campione di richieste
  - Definiamo una variabile aleatoria discreta X che rappresenta il **tempo di risposta** del server, e registriamo i seguenti tempi di risposta:
    - $X = 1$: 10% delle richieste sono servite entro 1 secondo
    - $X = 2$: 30% delle richieste sono servite entro 2 secondi
    - $X = 3$: 50% delle richieste sono servite entro 3 secondi
    - $X = 4$: 10% delle richieste sono servite entro 4 secondi
- La funzione di distribuzione è pertanto una diretta conseguenza della distribuzione di probabilità

x_i	1	2	3	4
$P(X = x_i)$	0,1	0,3	0,5	0,1
$F(x_i)$	0,1	0,4	0,9	1

C'è una probabilità del 40% che il tempo di risposta sia al massimo 2 secondi

C'è una probabilità del 90% che il tempo di risposta sia al massimo 3 secondi

VALORE ATTESO

- Supponiamo di avere una variabile aleatoria discreta X assume un numero finito o al più numerabile di valori $(x_1, x_2, \ldots, x_n)$ che associa ad ognuno dei valori x_i la corrispondente probabilità $P(X = x_i)$, cioè $(p_X(x_1), p_X(x_2), \ldots, p_X(x_n))$
- Chiamiamo **valore atteso** o **valor medio** o **speranza matematica** di una variabile aleatoria discreta X la sommatoria dei prodotti tra x_i e la rispettiva probabilità $p_X(x_i)$:

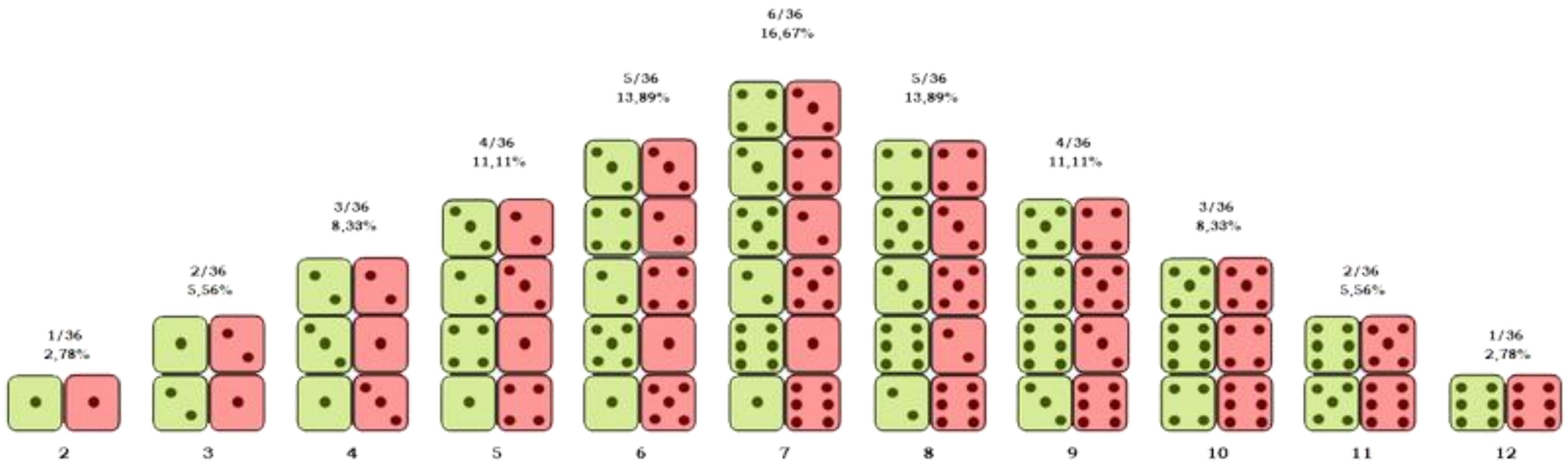
$$E(X) = \sum_{i=1}^{n} x_i * P(X = x_i) = \sum_{i=1}^{n} x_i * p_X(x_i)$$

  - Il valore atteso è una misura della sua **media ponderata** dei valori possibili, tenendo conto delle probabilità con cui ciascun valore si verifica
  - In altre parole, rappresenta il valore medio (o **tendenza centrale dei dati**) che ci si aspetta di ottenere quando la variabile aleatoria viene **ripetutamente** misurata o campionata un gran numero di volte

VALORE ATTESO

- Esempio: Supponiamo che la variabile aleatoria X indichi la somma dei punti ottenuti con il lancio di due dadi. Calcolare il punteggio totale atteso

$$\Omega = \{(i,j) \mid i,j = 1,2,3,4,5,6\} = \left\{\begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array}\right\}$$



VALORE ATTESO

- Esempio: Supponiamo che la variabile aleatoria X indichi la somma dei punti ottenuti con il lancio di due dadi. Calcolare il punteggio totale atteso

$$\Omega = \{(i,j) \mid i,j = 1,2,3,4,5,6\} = \left\{\begin{matrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{matrix}\right\}$$

x_i	2	3	4	5	6	7	8	9	10	11	12
$p_X(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$E(X) = \sum_{i=2}^{12} x_i * p_X(x_i) = 2 * \frac{1}{36} + 3 * \frac{2}{36} + \cdots + 12 * \frac{1}{36} = 7$$

VARIANZA

- La **varianza** $Var(X)$ è una misura di **dispersione** che indica quanto i valori di una variabile aleatoria si **discostano** dal valore atteso (media)
- Per una variabile aleatoria X con valore atteso $E[X] = \mu$, la varianza $Var(X)$ è definita come il valore atteso del quadrato della differenza tra X e la sua media:

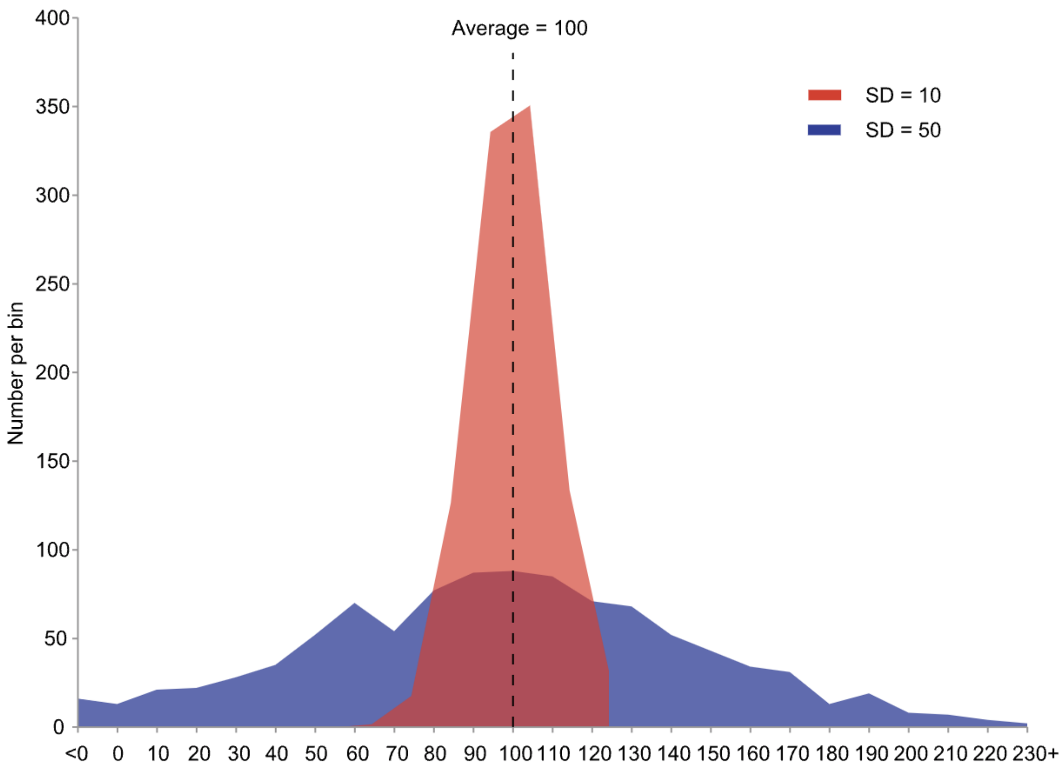
$$Var(X) = E[(X-\mu)^2]$$

  - Che può essere riscritta come:

$$Var(X) = E[X^2] - (E[X])^2$$

- Chiamiamo **varianza** di una **variabile aleatoria discreta** X la sommatoria dei prodotti tra x_i, la rispettiva probabilità $p_X(x_i)$ ed il valore atteso μ:

$$Var(X) = \sum_{i=1}^{n} p_X(x_i) * (x_i - \mu)^2$$



Wikipedia

Esempio di campioni da due popolazioni con la stessa media ma diversa varianza. La popolazione rossa ha media 100 e varianza 100 (SD=10), invece la popolazione blu ha media 100 e varianza 2500 (SD=50).

VARIANZA

$$Var(X) = \sum_{i=1}^{n} p_X(x_i) * (x_i - \mu)^2$$

- **Interpretazione dei valori della Varianza**
  - **Varianza piccola**: i valori della variabile sono **molto vicini alla media**. Questo significa che la variabile aleatoria o il dataset ha bassa variabilità
    - Ad esempio, se un server ha un tempo di risposta con una varianza molto bassa, significa che il tempo di risposta è piuttosto consistente e prevedibile.
  - **Varianza grande**: i valori sono **ampiamente distribuiti attorno alla media** e mostrano alta variabilità
    - Ad esempio, in un sistema informatico, un tempo di risposta con elevata varianza indica che il tempo di risposta del server è molto variabile, e potrebbe non essere affidabile o prevedibile

QUANTILI

- I quantili in una variabile aleatoria sono valori che suddividono la distribuzione in percentuali specifiche
- Questi valori sono utili per comprendere la posizione relativa di un dato all'interno della distribuzione
- I **quantili** di una variabile aleatoria sono ottenuti utilizzando la definizione di quantile per una **distribuzione di frequenza**
  - Il quantile (percentile) $z * 100\text{–}esimo$ di una variabile aleatoria X è definito come il più piccolo numero reale x, assunto dalla variabile aleatoria X, tale che:

$$F_X(x) = P(X \leq x) \geq z, \quad 0 \leq z \leq 1.$$

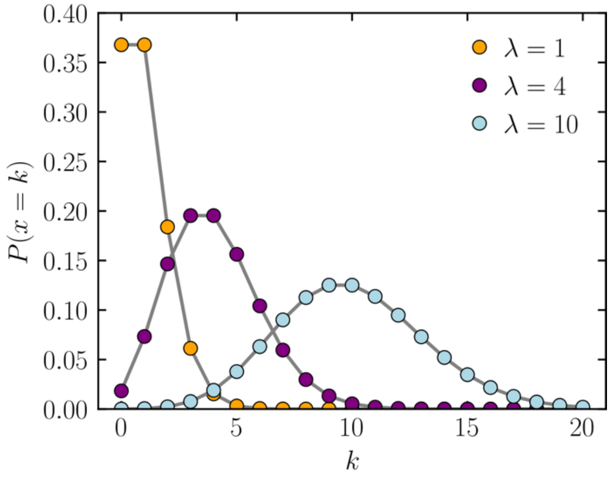
    cioè la funzione di distribuzione $F_X(x) = P(X \leq x)$ assuma valori maggiori o uguali di z
  - Per calcolare i **quartili** di una variabile aleatoria X basta porre:
    - $z = 0.25$ (primo quartile o 25–esimo percentile)
    - $z = 0.50$ (secondo quartile o mediana di una distribuzione di frequenza o 50–esimo percentile)
    - $z = 0.75$ (terzo quartile o 75–esimo percentile)

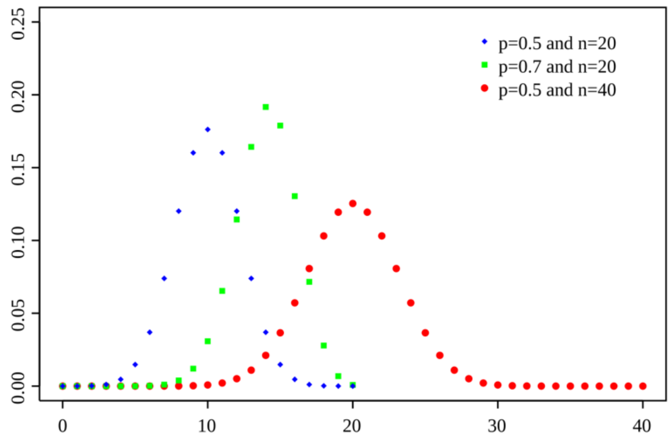
VARIABILI ALEATORIE DISCRETE

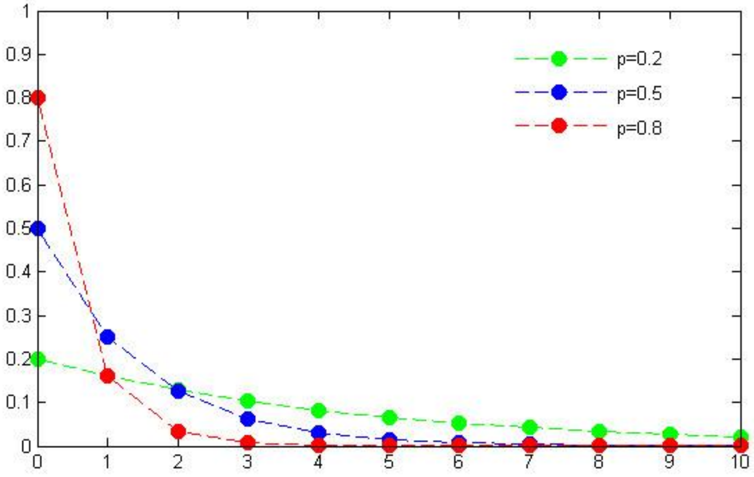
- **R** mette a disposizione per ciascuna delle principali variabili aleatorie discrete:
  - la funzione di probabilità:
  - la funzione di distribuzione;
  - la funzione per calcolare i quantili;
  - la funzione che simula la variabile aleatoria mediante la **generazione** di sequenze di numeri pseudocasuali
- I nomi delle funzioni iniziano con una lettera dell'alfabeto che indica il tipo di funzione a cui fa riferimento:
  - ***d()*** calcola la funzione di probabilità di una variabile aleatoria in uno specifico punto o in un insieme di punti (**density mass**);
  - ***p()*** calcola la funzione di distribuzione di una variabile aleatoria in uno specifico punto o in un insieme di punti (**probability distribution**);
  - ***q()*** calcola i **quantili**;
  - ***r()*** simula una variabile aleatoria generando una sequenza di numeri pseudocasuali.

VARIABILI ALEATORIE DISCRETE

- Nelle prossime lezioni discuteremo delle seguenti distribuzioni discrete:
  - Distribuzione di Bernoulli;
  - Distribuzione binomiale;
  - Distribuzione geometrica;
  - Distribuzione geometrica modificata;
  - Distribuzione binomiale negativa;
  - Distribuzione binomiale negativa modificata;
  - Distribuzione di Poisson;
  - Distribuzione ipergeometrica.







STATISTICA E ANALISI DEI DATI

Distribuzione di Bernoulli

Distribuzione di Bernoulli

- Una prova di Bernoulli è un esperimento casuale caratterizzato da due soli possibili risultati, interpretabili l'uno come **successo** e l'altro come **insuccesso**
- Si verificano rispettivamente con probabilità p e $1 - p$, con $0 < p < 1$
- La variabile aleatoria X che descrive il risultato di una prova di Bernoulli assume soltanto due valori:
  - 1 (indicante il successo) con probabilità p
  - 0 (indicante l'insuccesso) con probabilità $1 - p$.
- Esempio: Un tipico esempio è il lancio di una moneta in cui la probabilità di ottenere testa è p
  - Se codifichiamo testa con 1 e croce con 0, il risultato del lancio della moneta può essere descritto da una variabile aleatoria di Bernoulli con parametro p

DISTRIBUZIONE DI BERNOULLI

- La distribuzione di Bernoulli prende il nome da un matematico svizzero Jacob Bernoulli (1654–1705) che scoprì non solo tale distribuzione ma anche la **distribuzione binomiale**
- Con la notazione $X \sim B(1, p)$ intenderemo che X è una variabile aleatoria avente distribuzione di Bernoulli di parametro p, che chiameremo anche **variabile di Bernoulli**
- Una variabile aleatoria X di **funzione di probabilità**:

$$p_X(x) = P(X = x_i) = \begin{cases} 1 - p & se = 0 \\ p & se\ x = 1 \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$ è detta avere distribuzione di Bernoulli di parametro p

DISTRIBUZIONE DI BERNOULLI

- Una variabile aleatoria X di **funzione di probabilità**:

$$p_X(x) = P(X = x_i) = \begin{cases} 1-p & se = 0 \\ p & se\ x = 1 \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$ è detta avere distribuzione di Bernoulli di parametro p

- Quindi una Variabile aleatoria con distribuzione di Bernoulli:
  - Può assumere due valori $X = (0,1)$ con $x_1 = 0$, $x_2 = 1$
  - $P(X = 1) = p$
  - $P(X = 0) = 1 - p$

DISTRIBUZIONE DI BERNOULLI

- Una variabile aleatoria X di **funzione di probabilità**:

$$p_X(x) = P(X = x_i) = \begin{cases} 1-p & se = 0 \\ p & se\ x = 1 \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$ è detta avere distribuzione di Bernoulli di parametro p

- Quindi una Variabile aleatoria con distribuzione di Bernoulli:
  - Può assumere due valori $X = (0,1)$ con $x_1 = 0$, $x_2 = 1$
  - $P(X = 1) = p$
  - $P(X = 0) = 1 - p$
  - Il **valore atteso**

$$E[X] = \sum_{i=0}^{2} x_i P(X = x_i) = 0 * (1-p) + 1 * p = p$$

Distribuzione di Bernoulli

- Una variabile aleatoria X di **funzione di probabilità**:

$$p_X(x) = P(X = x_i) = \begin{cases} 1-p & se = 0 \\ p & se\ x = 1 \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$ è detta avere distribuzione di Bernoulli di parametro p

- Quindi una Variabile aleatoria con distribuzione di Bernoulli:
  - Può assumere due valori $X = (0,1)$ con $x_1 = 0$, $x_2 = 1$
  - $P(X = 1) = p$
  - $P(X = 0) = 1 - p$
  - Il valore atteso $E[X] = \sum_{i=0}^{2} x_i P(X = x_i) = 0 * (1-p) + 1 * p = p$
  - Varianza $\mathrm{Var}[X] = E[X^2] - [E[X]]^2$

$$E[X^2] = \sum_{i=0}^{2} x_i^2 P(X = x_i) = 0^2 * P(X = 0) + 1^2 * P(X = 1) = 0^2 * (1-p) + 1^2 * p = p$$

$$[E[X]]^2 = p^2$$

DISTRIBUZIONE DI BERNOULLI

- Una variabile aleatoria X di **funzione di probabilità**:

$$p_X(x) = P(X = x_i) = \begin{cases} 1-p & se = 0 \\ p & se\ x = 1 \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$ è detta avere distribuzione di Bernoulli di parametro p

- Quindi una Variabile aleatoria con distribuzione di Bernoulli:
  - Può assumere due valori $X = (0,1)$ con $x_1 = 0$, $x_2 = 1$
  - $P(X = 1) = p$
  - $P(X = 0) = 1 - p$
  - Il valore atteso $E[X] = \sum_{i=0}^{2} x_i P(X = x_i) = 0 * (1-p) + 1 * p = p$
  - Varianza $\mathrm{Var}[X] = E[X^2] - [E[X]]^2 = p - p^2 = p(1-p)$

$$E[X^2] = \sum_{i=0}^{2} x_i^2 P(X = x_i) = 0^2 * P(X = 0) + 1^2 * P(X = 1) = 0^2 * (1-p) + 1^2 * p = p$$

$$[E[X]]^2 = p^2$$

DISTRIBUZIONE DI BERNOULLI

- Una variabile aleatoria X di **funzione di probabilità**:

$$p_X(x) = P(X = x_i) = \begin{cases} 1-p & se = 0 \\ p & se\ x = 1 \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$ è detta avere distribuzione di Bernoulli di parametro p

- Per una variabile aleatoria di Bernoulli si ha:
  - **Valore atteso**: $E[X] = p$ e $E[X^2] = p$
  - **Varianza**: $Var(X) = E[X^2] - \left[E[X]\right]^2 = p(1-p)$
- La **funzione di distribuzione** di X è:

$$F_x(x) = \begin{cases} 0 & se\ x < 0 \\ 1-p & se\ 0 \le x < 1 \\ 1 & se\ x \ge 1 \end{cases}$$

Distribuzione di Bernoulli (Esempio)

- Immaginiamo che in una fabbrica si producano pezzi con una probabilità di difettosità pari a $p = 0.05$ (quindi c'è una probabilità del 5% che un pezzo sia difettoso)
- Definiamo la variabile casuale X come:
  - $X = 1$ se il pezzo è difettoso (successo, perché stiamo cercando difetti)
  - $X = 0$ se il pezzo non è difettoso (insuccesso)
- Poiché stiamo controllando un singolo pezzo, siamo di fronte a una singola prova $(n = 1)$, il che rende la distribuzione di X una distribuzione di Bernoulli $X \sim B(1, p)$ con parametro $p = 0.05$
  - Probabilità che il pezzo sia **difettoso**: $P(X = 1) = p = 0.05$
  - Probabilità che il pezzo **NON** sia difettoso: $P(X = 0) = 1 - p = 0.95$
  - **Valore atteso**: $E[X] = p = 0.05$
  - **Varianza**: $Var(X) = E[X^2] - \left[E[X]\right]^2 = p(1 - p) = 0.05 * 0.95 = 0.0475$
  - **Funzione di distribuzione:** $F_x(x) = \begin{cases} 0 & se\ x < 0 \\ 1 - 0.05 & se\ 0 \leq x < 1 \\ 1 & se\ x \geq 1 \end{cases}$

ESEMPIO CANALE DI COMUNICAZIONE

- Un sistema trasmette **bit** (0 o 1) attraverso un canale rumoroso
  - C'è una probabilità $p = 0.002$ che il bit venga ricevuto con **errore**
  - Analizziamo **un singolo bit** trasmesso, abbiamo che possiamo definire X come:

$$X = \begin{cases} 1 & \text{se il bit viene ricevuto CON errore} \\ 0 & \text{se il bit viene ricevuto CORRETTAMENTE} \end{cases}$$

  - Probabilità che il bit sia ricevuto con **errore**: $P(X = 1) = p = 0.002$
  - Probabilità che il bit sia **corretto**: $P(X = 0) = 1 - p = 0.998$
  - **Valore atteso**: $E[X] = p = 0.05$
  - **Varianza**: $Var(X) = E[X^2] - \left[E[X]\right]^2 = p(1-p) = 0.002 * 0.998\ oppure\ 0.002 - 0.000004 = 0.001996$

DISTRIBUZIONE DI BERNOULLI

- Esempi: Altri eventi che hanno una distribuzione Bernoulliana sono:
  - Il lancio di un dado, se ti interessa sapere se esce 4 o un numero diverso da 4
  - Il manifestarsi di una malattia, se ti interessa sapere se il soggetto ha o non ha quella malattia
  - La difettosità di un pezzo, se ti interessa sapere se un pezzo è difettoso oppure no
  - La nascita di un bambino, se ti interessa sapere se è maschio o femmina
  - L'estrazione ad una lotteria, in cui il biglietto può essere vincente oppure no
  - L'entrata di un cliente in un negozio, se ti interessa sapere se farà acquisti oppure no
  - La direzione da seguire, se ti interessa sapere se andare a destra oppure sinistra
  - L'invito ad uscire, se ti interessa sapere se la persona a cui l'hai chiesto risponderà "sì" oppure "no".
  - L'offerta di cambiare operatore telefonico, che può essere accettata o rifiutata
  - Il voto ad un referendum, che può essere a favore o contro
  - …

STATISTICA E ANALISI DEI DATI

Distribuzione Binomiale

Distribuzione Binomiale

- Consideriamo l'esperimento consistente in n prove di Bernoulli indipendenti ed effettuate tutte in condizioni identiche
- Assumiamo che in ogni prova i risultati di interesse siano sintetizzabili nel verificarsi dei seguenti due eventi necessari ed incompatibili:
  - A: interpretabile come successo
  - $\bar{A}$: interpretabile come insuccesso

  con $P(A) = p(0 < p < 1)$
- Questo esperimento si dice esperimento si dice costituito da n prove ripetute indipendenti di Bernoulli
  - **La variabile aleatoria Binomiale X rappresenta il numero di volte in cui si verifica l'evento A nelle n prove**

DISTRIBUZIONE BINOMIALE

- La distribuzione binomiale è utile per modellare:
  - il numero di successi in una sequenza di lanci indipendenti di una moneta;
  - il numero di processori che sono attivi in un sistema multiprocessore;
  - il numero di computer difettosi in una spedizione;
  - il numero di articoli in un lotto che hanno determinate caratteristiche
  - …

- Esempio:
  - Se analizziamo ad esempio il lancio di un dado regolare e prendiamo in considerazione l'evento **esce il numero 3** possiamo attribuire facilmente la probabilità associata pari 1/6
  - Questa probabilità resta uguale per ogni lancio, indipendentemente dal numero di lanci che effettuiamo

DISTRIBUZIONE BINOMIALE

- Una variabile aleatoria X di **funzione di probabilità**:

$$p_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,\dots,n \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$

  - n è il numero di prove
  - x numero di successi

- La notazione della variabile binomiale è: $X \sim B(n, p)$ intenderemo che X è una variabile aleatoria con distribuzione binomiale di parametri n e p

DISTRIBUZIONE BINOMIALE

- Una variabile aleatoria X di **funzione di probabilità**:

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,\dots,n \\ 0 & altrimenti \end{cases}$$

con $0 < p < 1$

- n è il numero di prove
- x numero di successi

- La notazione della variabile binomiale è: $X \sim B(n,p)$ intenderemo che X è una variabile aleatoria con distribuzione binomiale di parametri n e p
- <u>Esempio</u>: Se vogliamo calcolare la probabilità che lanciando <u>5 volte</u> un dado regolare a sei facce otteniamo <u>esattamente 3 volte</u> il numero 4 avremo:

$$p(3) = \binom{5}{3} * \left(\frac{1}{6}\right)^3 * \left(1 - \frac{1}{6}\right)^{5-3} = 10 * 0.0046 * 0.36 = 0.01656$$

Distribuzione Binomiale

- Esempio: Siamo interessati al numero di uscite di testa in 10 lanci. Qual è la probabilità di ottenere 6 uscite del lato testa?
  - Abbiamo che $n = 10$
  - Abbiamo che $x = 6$
- Consideriamo prima una prefissata sequenza di teste e croci

T C C T T T C C T T

Distribuzione Binomiale

- Esempio: Siamo interessati al numero di uscite di testa in 10 lanci. Qual è la probabilità di ottenere 6 uscite del lato testa?
  - Abbiamo che $n = 10$
  - Abbiamo che $x = 6$
- Consideriamo prima una prefissata sequenza di teste e croci

T C C T T T C C T T

Essendo i lanci indipendenti, la probabilità cercata è data da:

$$p * (1-p) * (1-p) * p * p * p * (1-p) * (1-p) * p * p \ = p^6(1-p)^{10-6}$$

DISTRIBUZIONE BINOMIALE

- Esempio: Siamo interessati al numero di uscite di testa in 10 lanci. Qual è la probabilità di ottenere 6 uscite del lato testa?
  - Abbiamo che $n = 10$
  - Abbiamo che $\mathrm{x} = 6$
- Consideriamo prima una prefissata sequenza di teste e croci

T C C T T T C C T T

Essendo i lanci indipendenti, la probabilità cercata è data da:

$$p * (1-p) * (1-p) * p * p * p * (1-p) * (1-p) * p * p \; = p^{6}(1-p)^{10-6}$$

- Se ora volessimo calcolare la probabilità di tutte le possibili sequenze di 10 lanci in cui ci siano 6 teste dobbiamo stabilire quante sono le sequenze che presentano 6 teste e 4 croci
- Il problema è equivalente a contare tutti i modi in cui è possibile scegliere di mettere le 6 teste nei 10 lanci, cioè: $\binom{10}{6} = \frac{10!}{6!(10-6)!}$

Essendo i lanci indipendenti, la probabilità cercata è data da: $p(6) = \binom{10}{6}(p)^{6}(1-p)^{10-6}$

DA BINOMIALE A BERNOULLI

- Se $n = 1$, la funzione di probabilità:

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

si semplifica ottenendo:

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

- Se $x = 0$ (insuccesso):

$$P(X = 0) = p^0 (1-p)^{1-0} = (\mathbf{1} - \boldsymbol{p})$$

- Se $x = 1$ (successo):

$$P(X = 1) = p^1 * (1-p)^{1-1} = \boldsymbol{p}$$

DISTRIBUZIONE BINOMIALE

- La **funzione di distribuzione** di X è poi immediatamente ottenibile:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \sum_{i=0}^{k} \binom{n}{i} p^i * (1-p)^{n-i} & k \le x < k+1 \;\; (k = 0,1,\dots,n-1) \\ 1 & x \ge n \end{cases}$$

- Per una variabile aleatoria binomiale si ha:

$$X = X_1 + X_2 + \cdots + X_n$$

dove $X_1, X_2, \dots, X_n$ sono variabili aleatorie di Bernoulli indipendenti ed identicamente distribuite

VARIANZA, VALORE ATTESO, DEVIAZIONE STANDARD

- In una variabile aleatoria binomiale, il valore atteso e la varianza sono indicatori statistici importanti che forniscono informazioni sulla distribuzione della variabile
  - il **valore atteso** di una variabile aleatoria binomiale rappresenta il numero medio di **successi** attesi in n tentativi
    - **Valore atteso**: $E(X) = np$
  - La **varianza** fornisce una misura di quanto i risultati si discostino da questo valore atteso
    - **Varianza**: $Var(X) = np(1-p)$
  - La **deviazione standard** è una misura della dispersione dei valori attorno alla media
    - una deviazione standard più piccola indica una maggiore coerenza dei dati intorno alla media
    - una deviazione standard più grande indica una maggiore dispersione o variabilità dei dati
    - **Deviazione standard**: $SD(X) = \sqrt{Var(X)}$
- Se $p = 1/2$ la funzione di probabilità binomiale è simmetrica rispetto al suo valore medio $E(X) = n/2$

DISTRIBUZIONE BINOMIALE (ESEMPIO)

- Esempio: la probabilità che il portiere del PSG pari un rigore è del 30%
  - Se durante una stagione intera la squadra subisce 10 rigori, qual è la probabilità che il portiere ne pari esattamente 4?

DISTRIBUZIONE BINOMIALE (ESEMPIO)

- I dati che abbiamo sono: $p = 0{,}30,\ n = 10,\ x = 4$

$$P(0) = \binom{10}{0} * 0{,}30^{0} * (1 - 0{,}30)^{10-0}$$

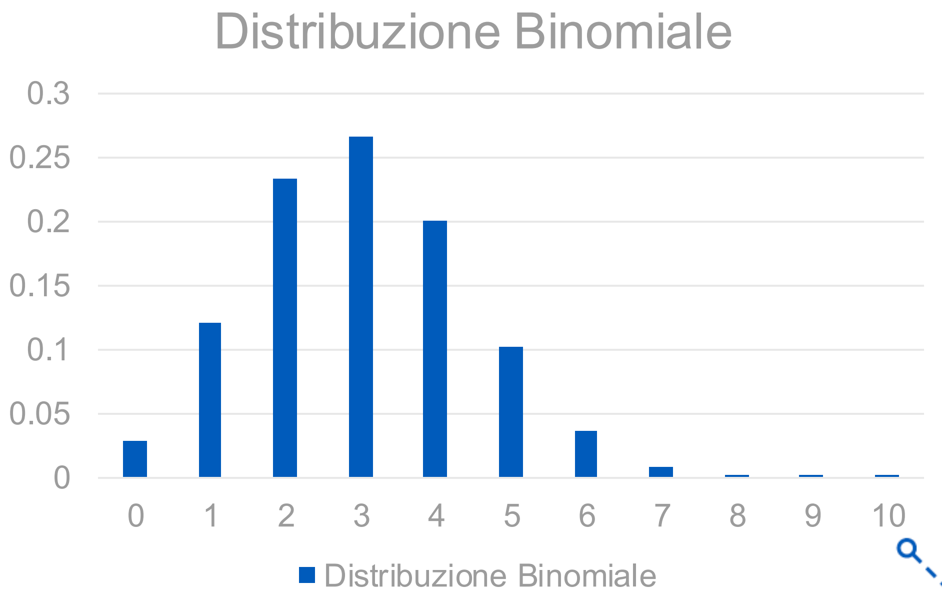
$$P(1) = \binom{10}{1} * 0{,}30^{1} * (1 - 0{,}30)^{10-1}$$

$$\dots$$

$$P(10) = \binom{10}{10} * 0{,}30^{10} * (1 - 0{,}30)^{10-10}$$

- Si nota subito che la funzione tocca il suo massimo nel valore medio, ovvero il 3

X	P(X)
0	0,02825
1	0,12106
2	0,23347
3	0,26683
4	0,20012
5	0,10292
6	0,03676
7	0,009
8	0,00145
9	0,00014
10	5,9E-06



DISTRIBUZIONE BINOMIALE (ESEMPIO)

- <u>Esempio</u>: la probabilità che il portiere del PSG pari un rigore è del 30%
  - Se durante una stagione intera la squadra subisce 10 rigori, qual è la probabilità che il portiere ne pari esattamente 4?
- I dati che abbiamo sono:
  - $p = 0{,}30$
  - $n = 10$
  - $x = 4$

$$p(4) = \binom{n}{x} p^x * (1-p)^{n-x} \binom{10}{4} 0{,}30^4 * (1-0{,}30)^{10-4} = 0{,}20$$

Distribuzione Binomiale (Esempio)

- Esempio: la probabilità che il portiere del PSG pari un rigore è del 30%
  - Se durante una stagione intera la squadra subisce 10 rigori, qual è la probabilità che il portiere ne pari esattamente 4?
- I dati che abbiamo sono:
  - $p = 0{,}30$
  - $n = 10$
  - $x = 4$

$$p(4) = \binom{n}{x} p^x * (1-p)^{n-x} \binom{10}{4} 0{,}30^4 * (1-0{,}30)^{10-4} = 0{,}20$$

$$E[X] = np = 10 * 0{,}30 = 3$$

DISTRIBUZIONE BINOMIALE (ESEMPIO)

- Esempio: la probabilità che il portiere del PSG pari un rigore è del 30%
  - Se durante una stagione intera la squadra subisce 10 rigori, qual è la probabilità che il portiere ne pari esattamente 4?
- I dati che abbiamo sono:
  - $p = 0{,}30$
  - $n = 10$
  - $x = 4$

$$p(4) = \binom{n}{x} p^x * (1-p)^{n-x} \binom{10}{4} 0{,}30^4 * (1-0{,}30)^{10-4} = 0{,}20$$

$$E[X] = np = 10 * 0{,}30 = 3$$

$$Var(X) = np(1-p) = 10 * 0{,}30(1-0{,}30) = 2{,}1$$

Distribuzione Binomiale (Esempio)

- Esempio: la probabilità che il portiere del PSG pari un rigore è del 30%
  - Se durante una stagione intera la squadra subisce 10 rigori, qual è la probabilità che il portiere ne pari esattamente 4?
- I dati che abbiamo sono:
  - $p = 0{,}30$
  - $n = 10$
  - $x = 4$

$$p(4) = \binom{n}{x} p^x * (1-p)^{n-x} \binom{10}{4} 0{,}30^4 * (1-0{,}30)^{10-4} = 0{,}20$$

$$E[X] = np = 10 * 0{,}30 = 3$$

$$Var(X) = np(1-p) = 10 * 0{,}30(1-0{,}30) = 2{,}1$$

$$SD(X) = \sqrt{Var(X)} = 1.449$$

DISTRIBUZIONE BINOMIALE (ESEMPIO)

- Esempio: la probabilità che il portiere del PSG pari un rigore è del 30%
  - Se durante una stagione intera la squadra subisce 10 rigori, qual è la probabilità che il portiere ne pari esattamente 4?
- I dati che abbiamo sono:
  - $p = 0{,}30$
  - $n = 10$
  - $x = 4$

$$p(4) = \binom{n}{x} p^x * (1-p)^{n-x} \binom{10}{4} 0{,}30^4 * (1-0{,}30)^{10-4} = 0{,}20$$

$$E[X] = np = 10 * 0{,}30 = 3$$

$$Var(X) = np(1-p) = 10 * 0{,}30(1-0{,}30) = 2{,}1$$

$$SD(X) = \sqrt{Var(X)} = 1.449$$

- Possiamo dire che abbiamo una forte probabilità che il portiere pari dai $3 - 1{,}45$ a $3 + 1{,}45$

Distribuzione Binomiale (R)

- Per il calcolo in R delle probabilità binomiali si utilizza la funzione:

dbinom(x, size, prob)

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria binomiale;
- **size** è il numero complessivo delle prove;
- **prob** è la probabilità di successo in ciascuna prova

- Ad esempio, se $n = 5$ e $p = 0.95$ le probabilità binomiali possono essere così valutate:

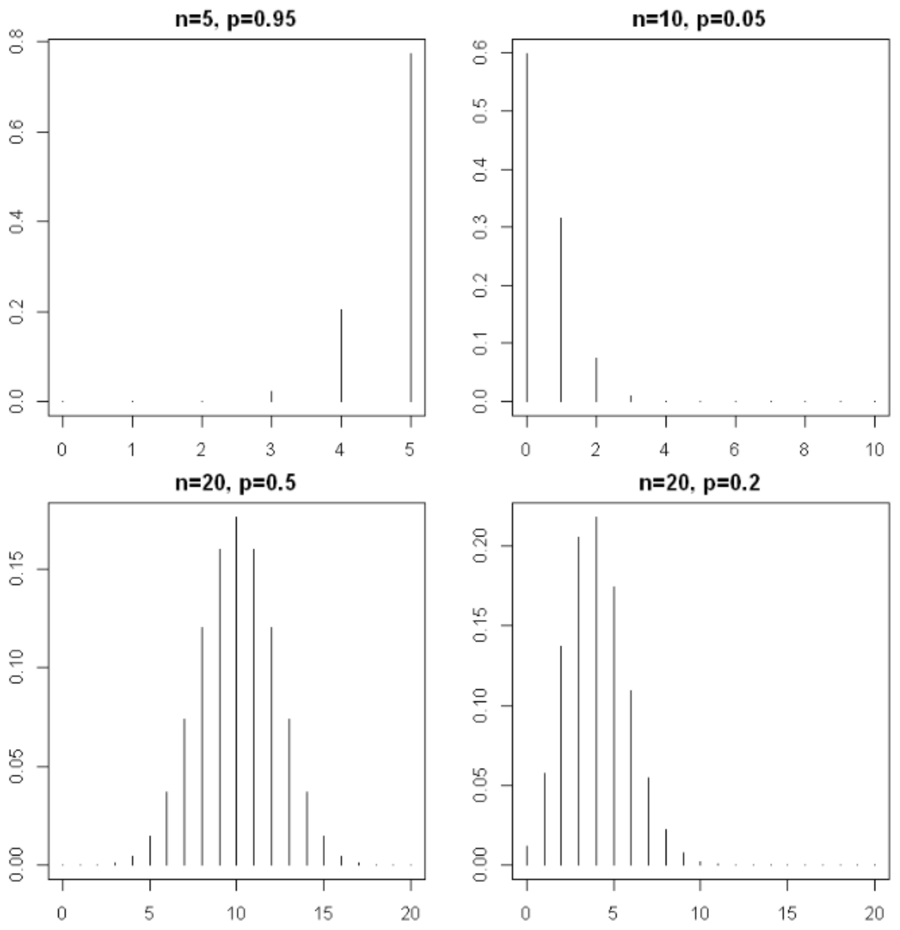
```
n <- 5
p <- 0.95
x <- 0:5
dbinom(x, size=n, prob=p)
```

3.12500000000001e-07 · 2.96875000000001e-05 · 0.001128125 · 0.021434375 · 0.2036265625 · 0.7737809375

DISTRIBUZIONE BINOMIALE (R)

- Creazione dei plot con diverse distribuzioni di probabilità binomiali:

```
#Creo una griglia per i grafici 2x2
par(mfrow=c(2,2))

x<-0:5
plot(x,dbinom(x,size=5,prob=0.95),xlab="x",ylab="P(X=x)",type="h",main="n=5, p=0.95")

x<-0:10
plot(x,dbinom(x,size=10,prob=0.05),xlab="x",ylab="P(X=x)",type="h",main="n=10, p=0.05")

x<-0:20
plot(x,dbinom(x,size=20,prob=0.5),xlab="x",ylab="P(X=x)",type="h",main="n=20, p=0.5")

x<-0:20
plot(x,dbinom(x,size=20,prob=0.2),xlab="x",ylab="P(X=x)",type="h",main="n=20, p=0.2")
```



DISTRIBUZIONE BINOMIALE (R)

- Per il calcolo in R della funzione di distribuzione binomiale si utilizza la funzione:

pbinom(x, size, prob, lower.tail = TRUE)

dove

- **x** è il valore assunto (o i valori assunti) dalla variabile aleatoria binomiale;
- **size** è il numero complessivo delle prove;
- **prob** è la probabilità di successo in ciascuna prova
- **lower.tail** se tale parametro è TRUE (caso di default) calcola $P(X \leq x)$, mentre se tale parametro è FALSE calcola $P(X > x)$

DISTRIBUZIONE BINOMIALE (R)

- Ad esempio, se n=5 e p=0.95 la funzione di distribuzione binomiale può essere così valutata:

```
n <- 5
p <- 0.95
x <- 0:5
pbinom(x, size=n, prob=p)
```

3.12500000000001e-07 · 3.00000000000001e-05 · 0.001158125 · 0.0225925 · 0.2262190625 · 1

- I cui risultati sono le probabilità:

$$P(X \leq x) = \sum_{n=0}^{x} P(X = x) \qquad x = 0,1,\dots,5$$

- Impostando il parametro *<u>lower.tail = FALSE</u>*

```
n <- 5
p <- 0.95
x <- 0:5
pbinom(x, size=n, prob=p, lower.tail=FALSE)
```

0.9999996875 · 0.99997 · 0.998841875 · 0.9774075 · 0.7737809375 · 0

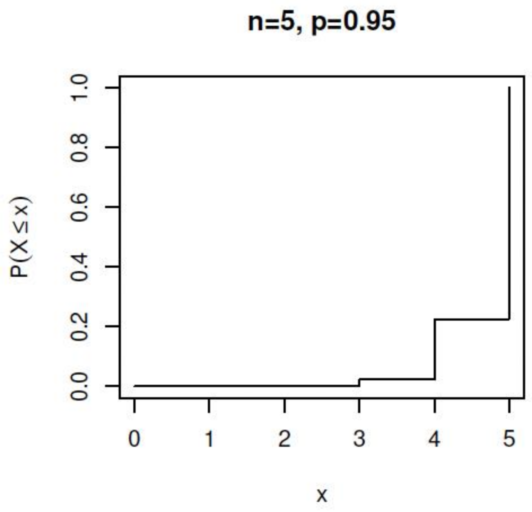
- Mostrano le probabilità:

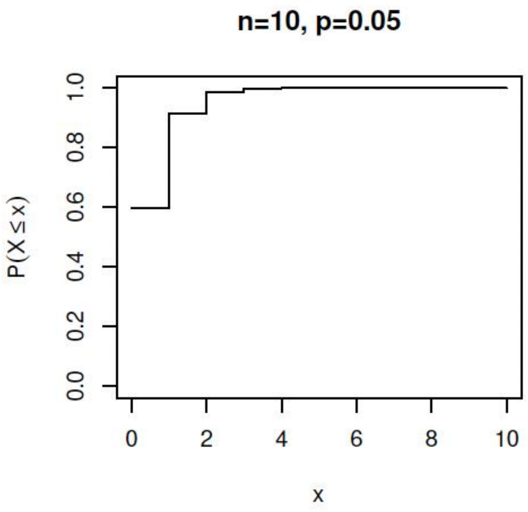
$$P(X > x) = 1 - P(X \leq x) \qquad x = 0,1,\dots,5$$

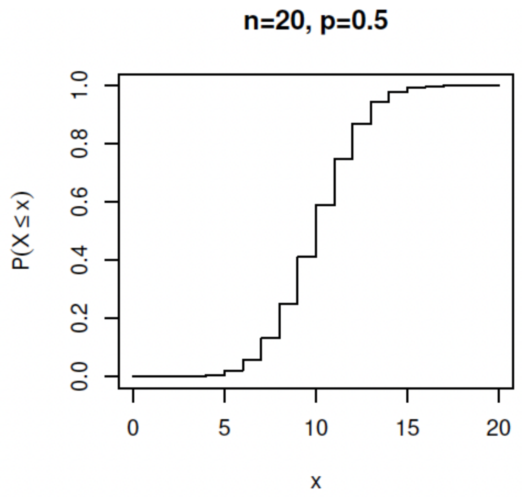
DISTRIBUZIONE BINOMIALE (R)

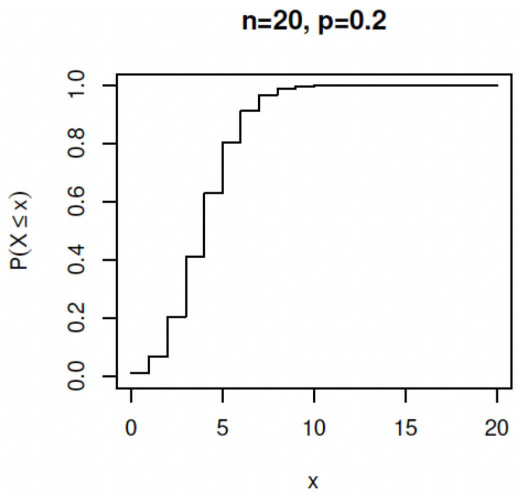
- Creazione dei plot con diverse funzioni di distribuzione:

```
par(mfrow=c(2,2))
x<-0:5
plot(x,pbinom(x,size=5,prob=0.95),xlab="x",ylab=expression(P(X <=x)),ylim=c(0,1),type="s",main="n=5,p=0.95")

x<-0:10
plot(x,pbinom(x,size=10,prob=0.05),xlab="x",ylab=expression(P(X <=x)),ylim=c(0,1),type="s",main="n=10,p=0.05")

x<-0:20
plot(x,pbinom(x,size=20,prob=0.5),xlab="x",ylab=expression(P(X <=x)),ylim=c(0,1),type="s",main="n=20,p=0.5")

x<-0:20
plot(x,pbinom(x,size=20,prob=0.2),xlab="x",ylab=expression(P(X <=x)),ylim=c(0,1),type="s",main="n=20,p=0.2")
```









ALTRI VALORI IN R

- In valutare il **valore medio**, la **varianza**, la **deviazione standard** e il **coefficiente di variazione** della distribuzione binomiale:

- <u>Esempio</u>: Consideriamo *n = 20* e probabilità *p = 0.2:*
  - **Valore atteso**: $E[X] = np = 20 * 0.2 = 4$
  - **Varianza**: $Var(X) = np(1 - p) = 20 * 0.2 * 0.8 = 3.2$
  - **Deviazione Standard**: $\sqrt{Var(X)} = 1.78885 \ldots$
  - **Coefficiente di Variazione**: $\mathrm{CV}(X) = \frac{\sqrt{Var(X)}}{E[X]} = 0.44721 \ldots$

```
x <- 0:20
n <- 20
p <- 0.2

M1 <- sum(x*dbinom(x,size=n, prob=p))
M2 <- sum(x^2*dbinom(x,size=n, prob=p))
V <- M2 - M1^2

c(M1,V,sqrt(V),sqrt(V)/M1)
```

4 · 3.2 · 1.78885438199983 · 0.447213595499958

ALTRI VALORI IN R

- Per calcolare i **Quantili** della distribuzione binomiale si utilizza la funzione:

  qbinom(z, size, prob)

  dove

  - **z** è il valore assunto (o i valori assunti) dalle probabilità relative al percentile z·100-esimo;
  - **size** è il numero complessivo delle prove;
  - **prob** è la probabilità di successo in ciascuna prova

- Esempio: se $n = 20$ e $p = 0.2$ le seguenti linee di codice forniscono i quartili Q_0, Q_1, Q_2, Q_3, Q_4:

```
z <- c(0,0.25,0.5,0.75,1)
qbinom(z,size=20, prob=0.2)
```

0 · 3 · 4 · 5 · 20

- Il risultato della funzione è il percentile $z \cdot 100 - esimo$, ossia il più piccolo numero intero k assunto dalla variabile aleatoria binomiale X tale che:

$$F_X(x) = P(X \leq k) \geq z \qquad k = 0,1, \dots, n$$

CANALE DI COMUNICAZIONE (ESEMPIO)

- Consideriamo un canale di comunicazione binario che trasmette parole di codice di n bits
- Assumiamo che la probabilità di successo nella trasmissione di un singolo bit sia p e che la probabilità di un errore sia $q = 1 - p$
  - Se assumiamo che la trasmissione di bit successivi avviene indipendentemente, siamo interessati a <u>calcolare la probabilità di commettere al più</u> n_e <u>errori nella trasmissione</u> di una parola di codice
- Denotiamo con N la variabile aleatoria che descrive il numero di errori commessi nella trasmissione di una parola di codice
  - la probabilità di commettere k errori è:

$$P(N = k) = \binom{n}{k} p^k q^{n-k} \qquad k = 0,1,\dots,n$$

  - E quindi la probabilità di commettere al più n errori è:

$$P(N \le n_e) = \sum_{k=0}^{n_e} \binom{n}{k} p^k q^{n-k} \qquad n_e = 0,1,\dots,n$$

CANALE DI COMUNICAZIONE (ESEMPIO)

- Consideriamo un canale di comunicazione binario che trasmette parole di codice di n bits
- Assumiamo che la probabilità di successo nella trasmissione di un singolo bit sia p e che la probabilità di un errore sia $q = 1 - p$
  - Se assumiamo che la trasmissione di bit successivi avviene indipendentemente, siamo interessati a <u>calcolare la probabilità di commettere al più</u> n_e <u>errori nella trasmissione</u> di una parola di codice
- <u>Esempio</u>: la probabilità di commettere al più un errore in una parola di codice di lunghezza 10 con probabilità di errore q = 0.1 è calcolabile in R:

```r
x <- 1
n <- 10
p <- 0.1
pbinom(x, n, p)
```

0.7360989291

REGOLA DI DECISIONE A MAGGIORANZA (ESEMPIO)

- Supponiamo di effettuare n lanci $(n = 3, 5, \ldots)$ indipendenti di una moneta
  - Supponiamo che sia p la probabilità di successo e $1 - p$ di insuccesso in ogni singola prova
- Sia X la variabile aleatoria che descrive il **numero di successi in n prove**
  - Desideriamo calcolare la probabilità Q_n che il numero di successi sia maggiore del numero di insuccessi nelle n prove
    - Quando si verifica ciò?
      - si verifica se $X > n - X$, ossia $X > \frac{n}{2}$ o equivalentemente $X \geq \frac{n+1}{2}$
    - Essendo le prove indipendenti, X ha distribuzione binomiale di parametri n e p;
      - Per $n = 3, 5, \ldots$ si ha:

$$Q_n = P\left(X > \frac{n}{2}\right) = P\left(X \geq \frac{n+1}{2}\right) = \sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k} n^k q^{n-k}$$

Regola di decisione a maggioranza (Esempio)

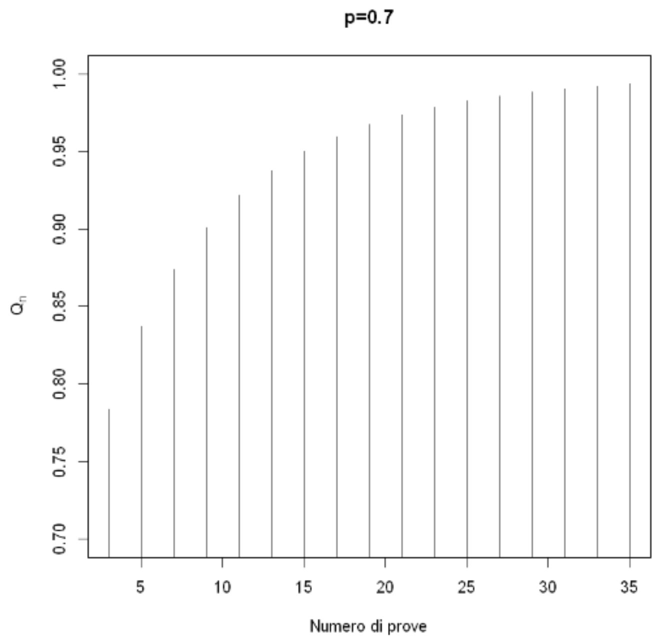
- Si può mostrare graficamente con R che se:
  - $p > 0.5$ e $n = 3, 5, \dots$ la probabilità Q_n <u>è una funzione crescente in</u> n
  - $p < 0.5$ e $n = 3, 5, \dots$ la probabilità Q_n <u>è una funzione decrescente in</u> n
  - $p = 0.5$ risulta $Q_n = \frac{1}{2}$ per $n = 3, 5, \dots$
- In R, il codice per calcolare la probabilità Q_n per $p = 0.5$ e $n = 3, 5, \dots, 35$:

```r
n <- seq(3,35,2)
Qn <- pbinom(n/2, n, 0.7, lower.tail=FALSE)
round(Qn,4)
```

0.784 · 0.8369 · 0.874 · 0.9012 · 0.9218 · 0.9376 · 0.95 · 0.9597 · 0.9674 · 0.9736 · 0.9786 · 0.9825 · 0.9857 · 0.9883 · 0.9905 · 0.9922 · 0.9936

Regola di decisione a maggioranza (Esempio)

- Si può mostrare graficamente con R che se:
  - $p > 0.5$ e $n = 3, 5, \ldots$ la probabilità Q_n è una funzione crescente in n
  - $p < 0.5$ e $n = 3, 5, \ldots$ la probabilità Q_n è una funzione decrescente in n
  - $p = 0.5$ risulta $Q_n = \frac{1}{2}$ per $n = 3, 5, \ldots$
- In R, il codice per calcolare la probabilità Q_n per $p = 0.5$ e $n = 3, 5, \ldots, 35$:




```
n <- seq(3,35,2)
Qn <- pbinom(n/2, n, 0.7, lower.tail=FALSE)
round(Qn,4)
```

0.784 · 0.8369 · 0.874 · 0.9012 · 0.9218 · 0.9376 · 0.95 · 0.9597 · 0.9674 · 0.9736 · 0.9786 · 0.9825 · 0.9857 · 0.9883 · 0.9905 · 0.9922 · 0.9936

```
plot(n,Qn, xlab="Numero di prove", ylab= expression('Q'['n']),ylim=c(0.7,1),type="h",main="p=0.7")
```

REGOLA DI DECISIONE A MAGGIORANZA (ESEMPIO)

- In R, il codice per calcolare la probabilità Q_n per $p = 0.3$ e $n = 3, 5, \ldots, 35$:

```
n <- seq(3,35,2)
Qn <- pbinom(n/2, n, 0.3, lower.tail=FALSE)
plot(n,Qn, xlab="Numero di prove", ylab= expression('Q'['n']),ylim=c(0,0.25),type="h",main="p=0.3")
```