



Last updated April 26, 2016

Probabilità e Distribuzione Normale

Lezione 3

G. Bacaro

Statistica
CdL in Scienze e Tecnologie per l'Ambiente e la Natura
I anno, II semestre

Statistica=scienza dell'incertezza

PROBABILITÀ ALLA BASE DELL'INFERENZA



**Qui entra in gioco la
probabilità**

Introduzione alla probabilità

La probabilità è un valore numerico che rappresenta la possibilità che un particolare evento si verifichi

È una proporzione che varia fra 0 e 1 (0-100%)

Probabilità di un evento = X / T

X = numero di casi nei quali l'evento si verifica

T = numero totale di casi

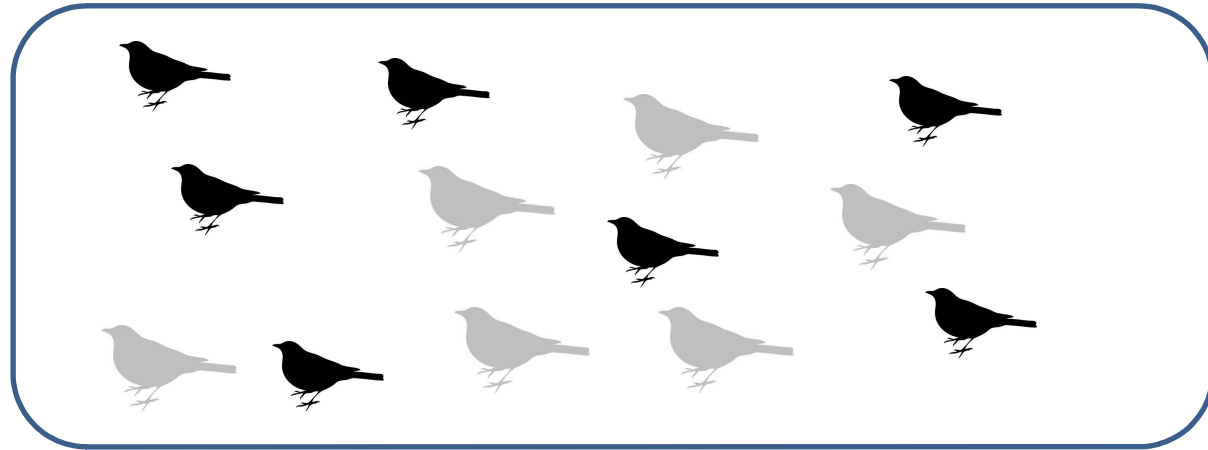
Introduzione alla probabilità

Esempio

$N=13$

$M=7$

$F=6$



Qual è la probabilità di estrarre un maschio?



$P=$

Evento  Probabilità

Introduzione alla probabilità

Esempio: probabilità di superare l'esame di statistica=0.8

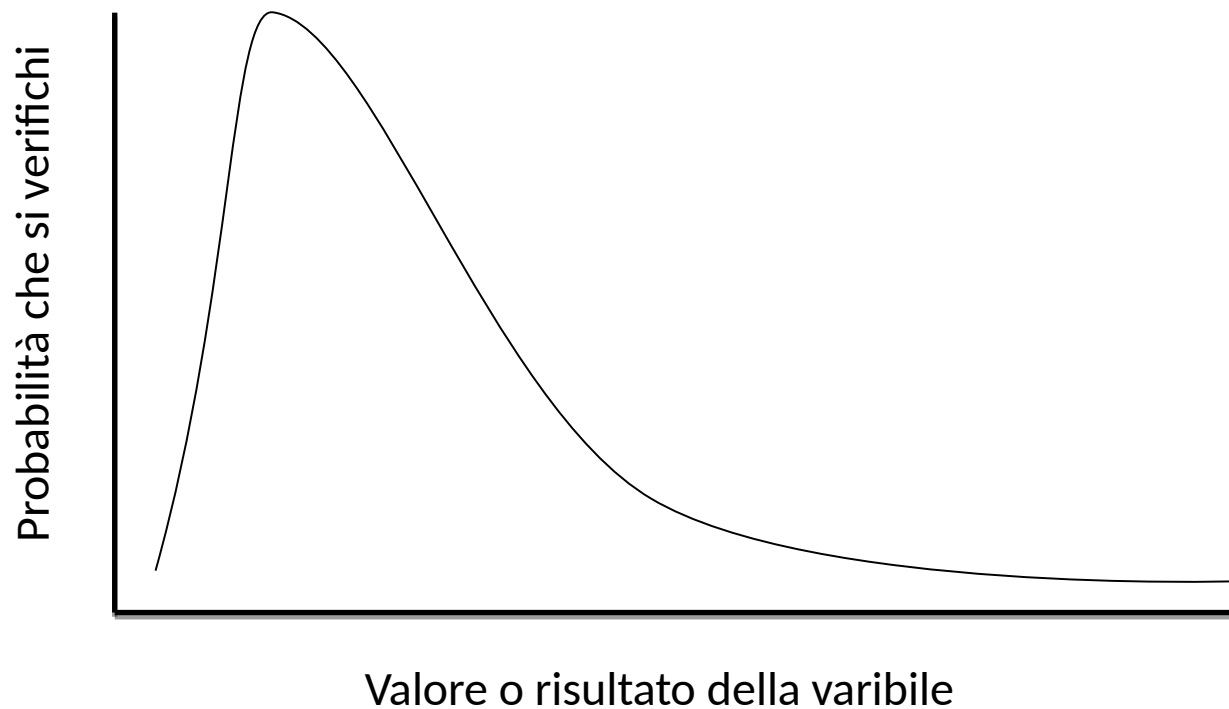
N=120 studenti iscritti

Quanti studenti supereranno l'esame al primo colpo?

Probabilità  Evento

Distribuzione di probabilità

Spesso ci interessa costruire una distribuzione di probabilità che ci permetta di calcolare la probabilità esatta che un qualsiasi risultato della variabile si verifichi



Distribuzione di probabilità

Esistono molte distribuzioni ma solo alcune vengono utilizzate frequentemente in statistica

Variabili discrete:

Binomiale

Poisson

Variabili continue:

Normale

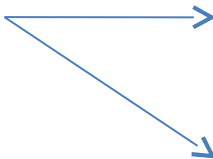
t

F

Distribuzione di probabilità discrete: binomiale

Descrive il numero di successi in un campione di n osservazioni

1. Numero finito di osservazioni (n)

2. Osservazione  1 (successo)
0 (insuccesso)

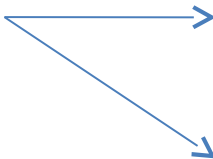
3. Probabilità nota e costante che si verifichi un successo

4. Indipendenza della osservazioni

Distribuzione di probabilità discrete: binomiale

Numero di studenti maschi in una classe di 50 studenti (25 F-25 M)

1. Numero finito di osservazioni= n

2. Studente  maschio (insuccesso)
femmina (successo)

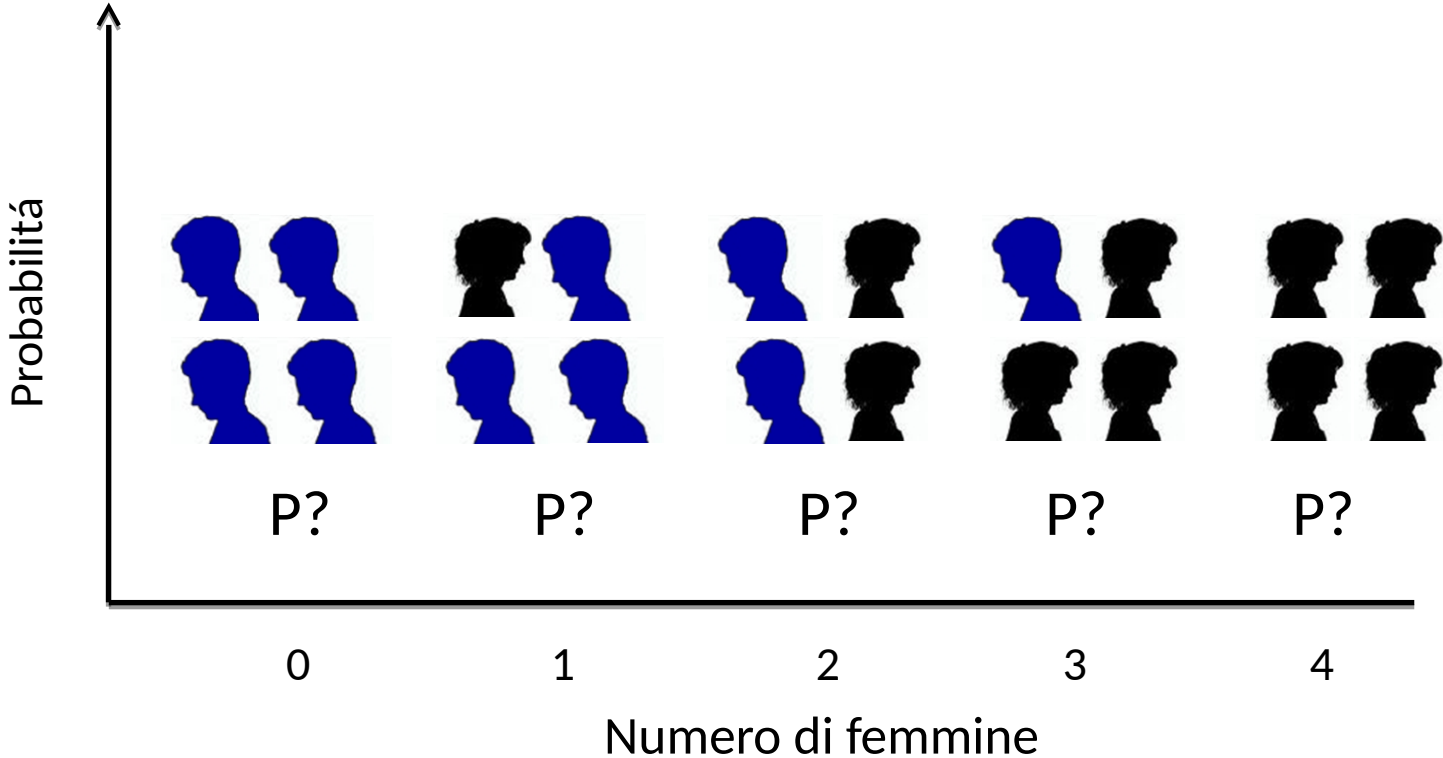
3. Probabilità nota e costante che si verifichi un successo= $25/50$

4. Indipendenza della osservazioni: campionamento casuale

Distribuzione di probabilità discrete: binomiale

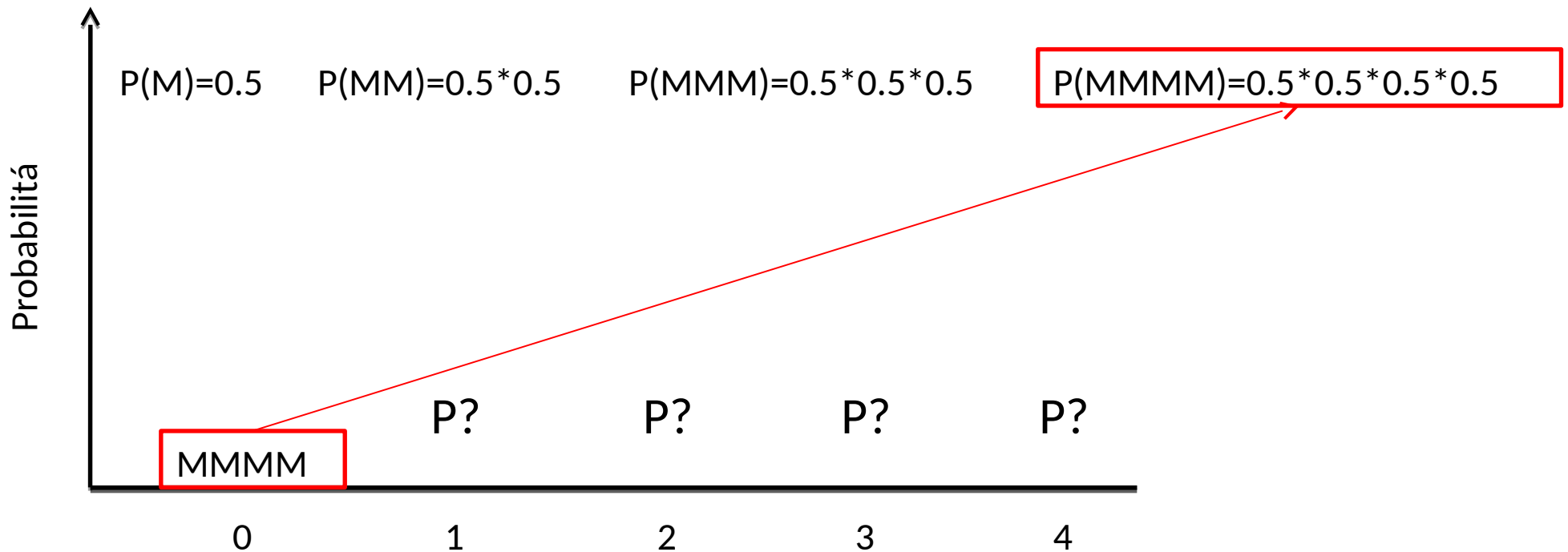
Estrazione di 4 studenti ($n=4$)

Probabilità di estrarre una femmina=0.5



Distribuzione di probabilità discrete: binomiale

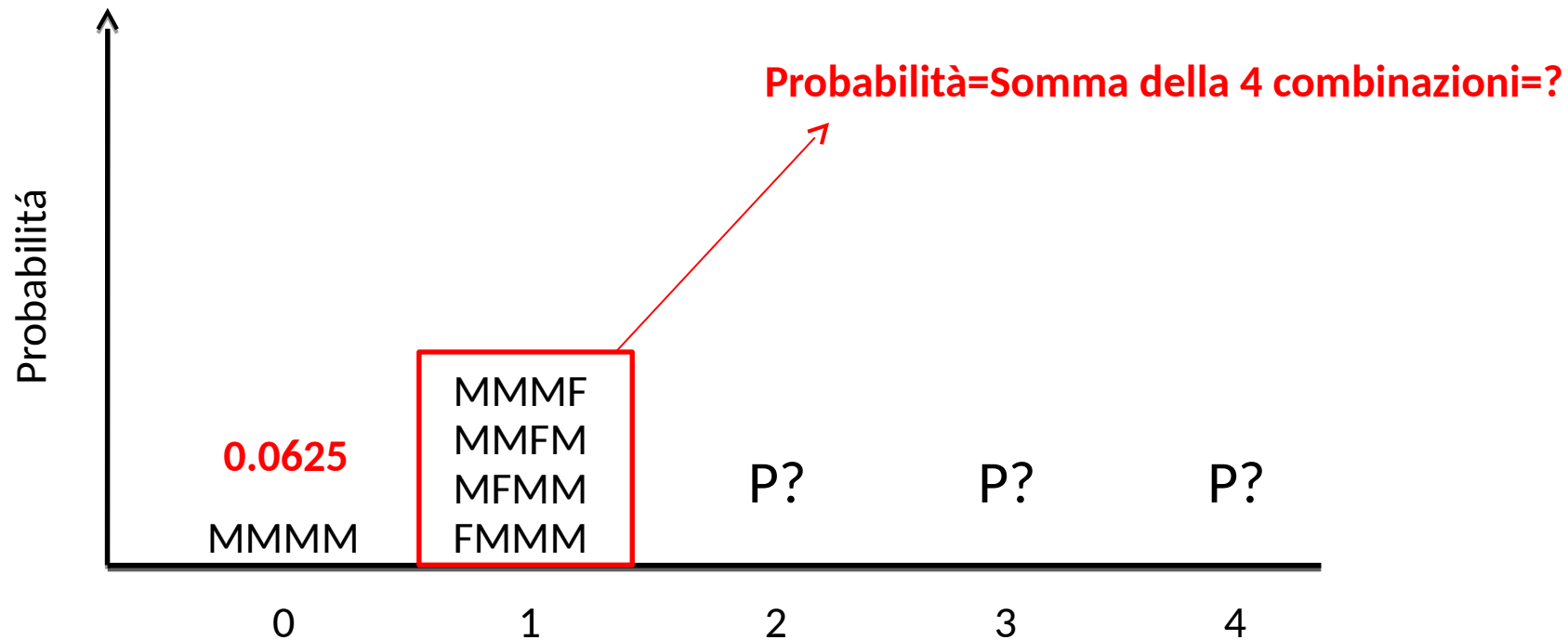
Estrazione di 4 studenti (n=4)



$P(XXXX)$ =prodotto delle probabilità dei singoli eventi in una combinazione

Distribuzione di probabilità discrete: binomiale

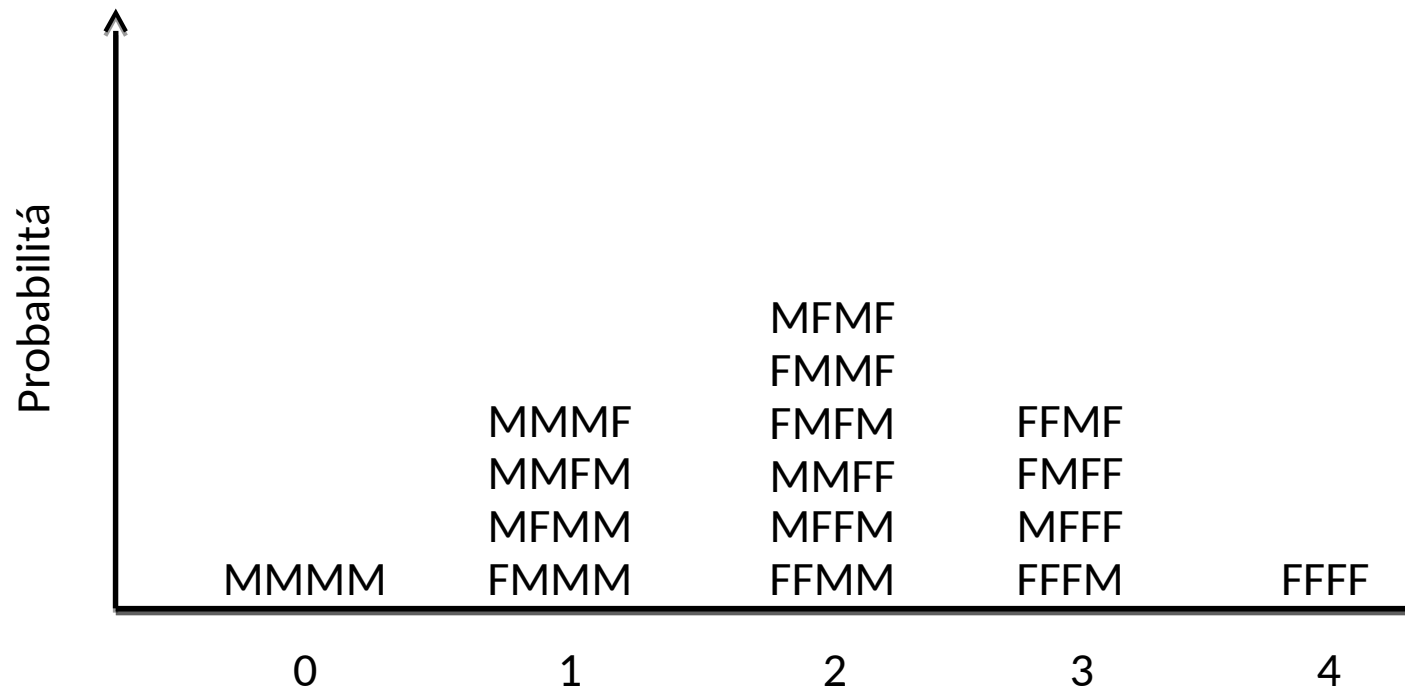
Estrazione di 4 studenti (n=4)



$P(\text{XXXX}) = \text{Somma delle probabilità delle varie combinazioni}$

Distribuzione di probabilità discrete: binomiale

Estrazione di 4 studenti (n=4)



La somma di tutte queste probabilità è 1

Calcolo di Probabilità: Permutazioni Semplici

Dato un insieme di n oggetti differenti $a_1, a_2, a_3, \dots, a_n$, si chiamano permutazioni semplici tutti i sottoinsiemi che si possono formare, collocando gli n elementi in tutti gli ordini possibili.

Il numero di permutazioni di n elementi è

$$P = n!$$

dove $n!$ (n fattoriale) è il prodotto degli n elementi: $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$.

Esempio: le permutazioni di 3 elementi abc $P=??$

PER DEFINIZIONE: $0! = 1$ e $1! = 1$

Calcolo di Probabilità: Permutazioni Semplici

Tabella dei fattoriali di interi (per facilitare i calcoli).

n	$n!$	n	$n!$
1	1	26	4.03291×10^{26}
2	2	27	1.08889×10^{28}
3	6	28	3.04888×10^{29}
4	24	29	8.84176×10^{30}
5	120	30	2.65253×10^{32}
6	720	31	8.22284×10^{33}
7	5040	32	2.63131×10^{35}
8	40320	33	8.68332×10^{36}
9	362880	34	2.95233×10^{38}
10	3.62880×10^6	35	1.03331×10^{40}
11	3.99168×10^7	36	3.71993×10^{41}
12	4.79002×10^8	37	1.37638×10^{43}
13	6.22702×10^9	38	5.23023×10^{44}
14	8.71783×10^{10}	39	2.03979×10^{46}
15	1.30767×10^{12}	40	8.15915×10^{47}
16	2.09228×10^{13}	41	3.34525×10^{49}
17	3.55687×10^{14}	42	1.40501×10^{51}
18	6.40327×10^{15}	43	6.04153×10^{52}
19	1.21645×10^{17}	44	2.65827×10^{54}
20	2.43290×10^{18}	45	1.19622×10^{56}
21	5.10909×10^{19}	46	5.50262×10^{57}
22	1.12400×10^{21}	47	2.58623×10^{59}
23	2.58520×10^{22}	48	1.24139×10^{61}
24	6.20448×10^{23}	49	6.08282×10^{62}
25	1.55112×10^{25}	50	3.04141×10^{64}

Calcolo di Probabilità: Disposizioni Semplici

Dato un insieme di n oggetti differenti $a_1, a_2, a_3, \dots, a_n$ si chiamano disposizioni semplici i sottoinsiemi di p elementi che si diversificano almeno per un elemento o per il loro ordine. Le disposizioni delle 4 lettere a, b, c, d , raggruppate 3 a 3 sono: $abc, acb, bac, dba, bda, abd$, ecc.

Il numero di disposizioni semplici di n elementi p a p è

$$D_n^p = \frac{n!}{(n-p)!}$$

Esempio 1: le disposizioni di 4 elementi 3 a 3 sono:

$$D_4^3 = \frac{4!}{(4-3)!} = \frac{24}{1} = 24$$

Derivato dalla semplificazione di questa formula, un altro modo per calcolare le disposizioni semplici di n elementi p a p è:

$$D_n^p = n(n-1)(n-2)\dots(n-p+1)$$

Calcolo di Probabilità: Combinazioni Semplici

Dato un insieme di n oggetti differenti $a_1, a_2, a_3, \dots, a_n$, si chiamano combinazioni semplici di n elementi p a p i sottoinsiemi che si diversificano almeno per un elemento, ma non per il loro ordine.

Le combinazioni semplici delle 4 lettere a, b, c, d , 3 a 3 sono: abc, abd, acd, bcd .

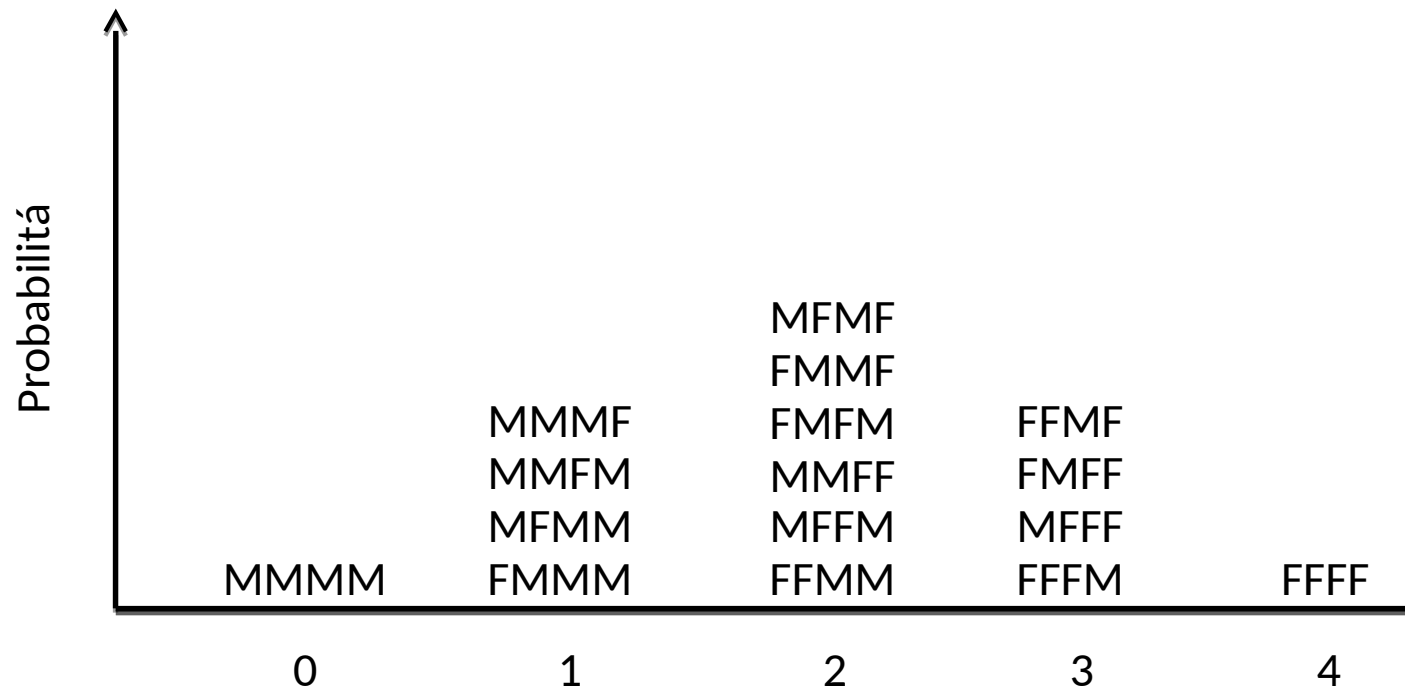
Il numero di combinazioni semplici di n elementi p a p è

$$C_n^p = \frac{n!}{(n-p)! p!}$$

Sotto l'aspetto del calcolo e dal punto di vista concettuale, il numero di combinazioni di n elementi p a p corrisponde al rapporto tra il numero di disposizioni di n elementi p a p ed il numero di permutazioni di p elementi.

Distribuzione di probabilità discrete: binomiale

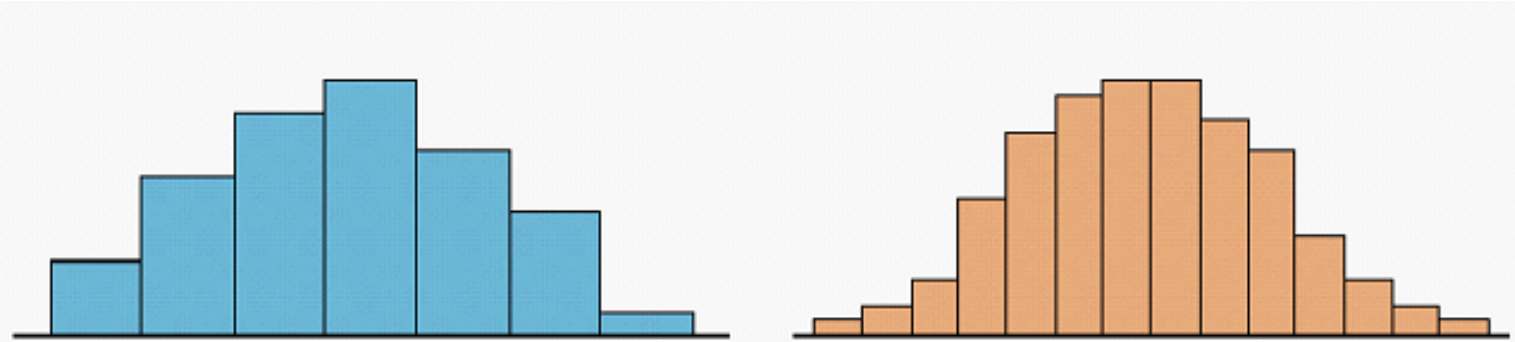
Estrazione di 4 studenti (n=4)



La somma di tutte queste probabilità è 1

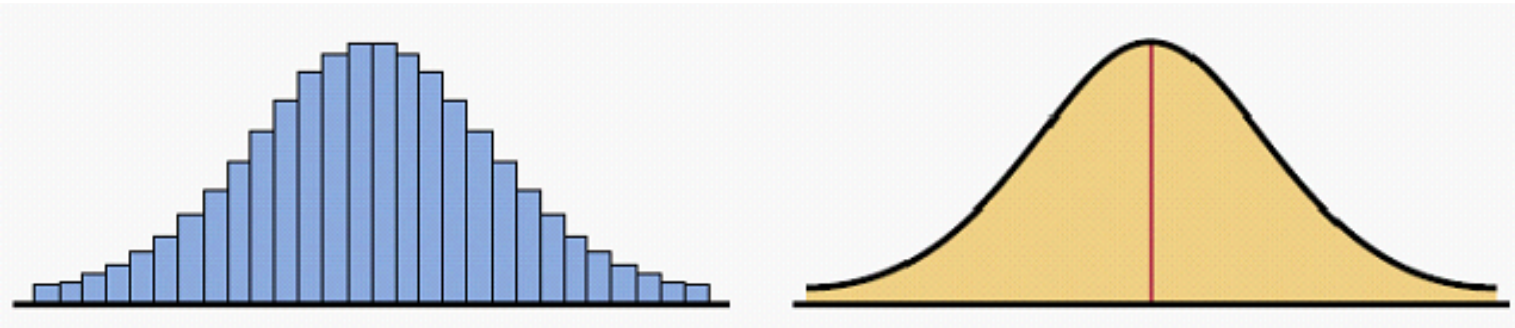
Distribuzione normale

Istogramma della distribuzione delle altezze



(a) 100 studenti

(b) 1000 studenti



(c) 10000 studenti

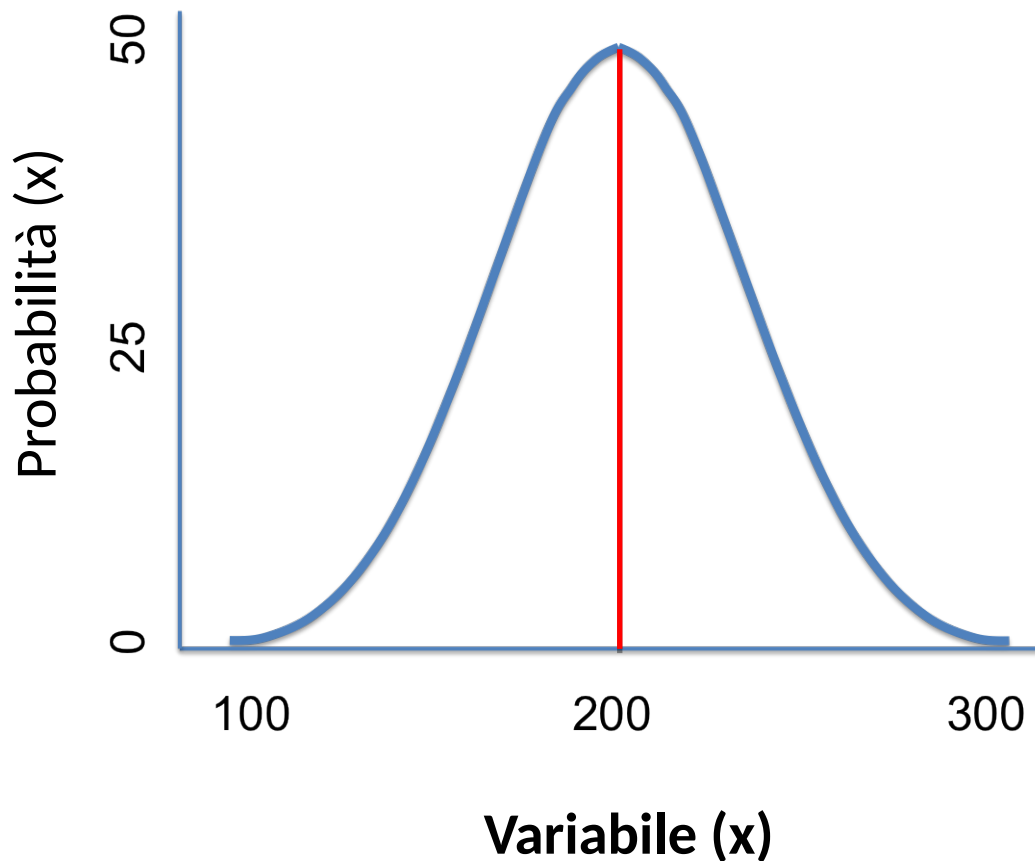
(d) Intera popolazione

Molte variabili continue seguono questa distribuzione!!!

La frequenza diventa una probabilità

Distribuzione normale

Per ogni valore della variabile ho una probabilità: una variabile continua spesso assume la distribuzione normale

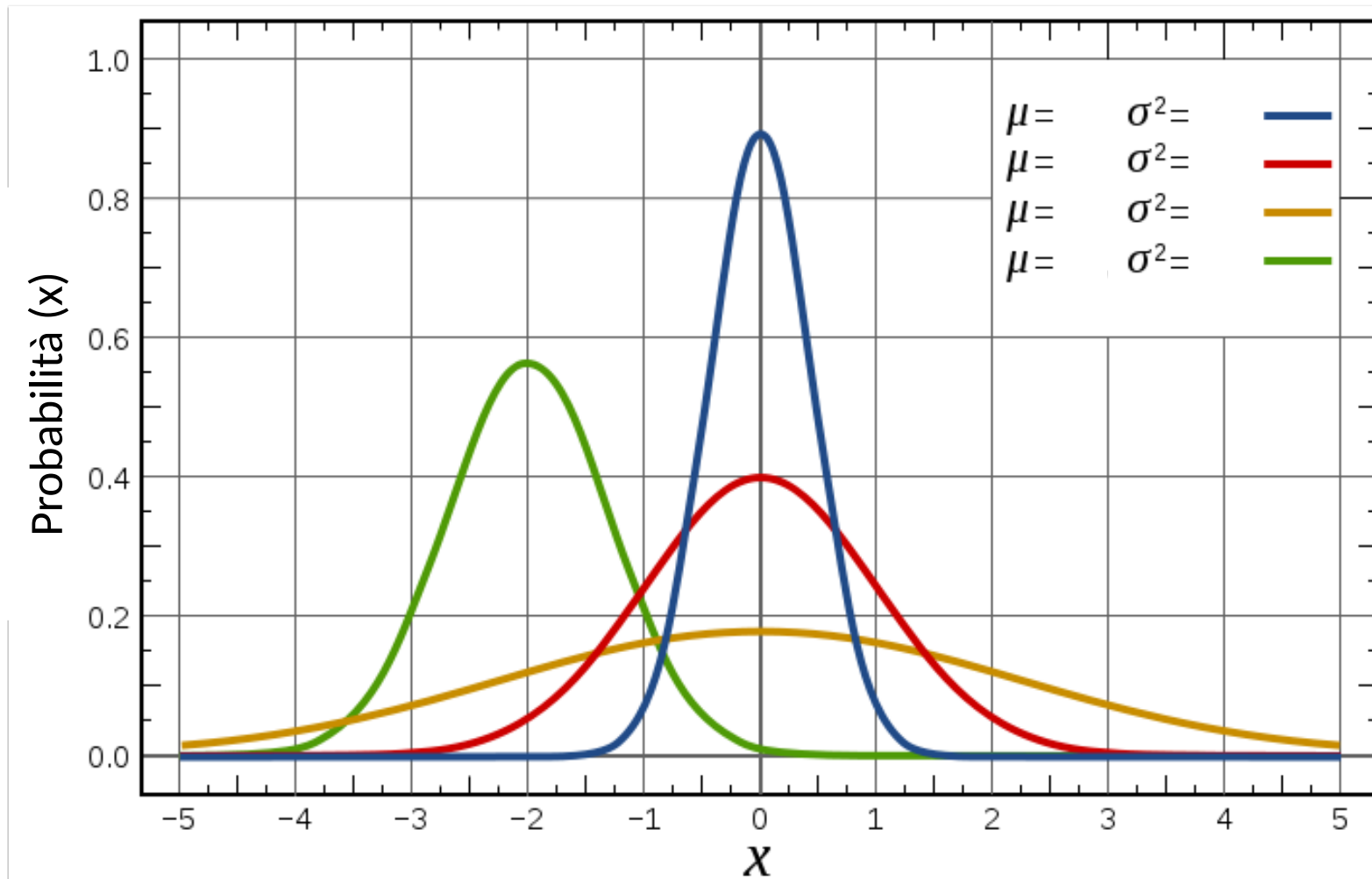


$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

1. Distribuzione simmetrica attorno alla media
2. La “larghezza” dipende dalla variabilità della popolazione

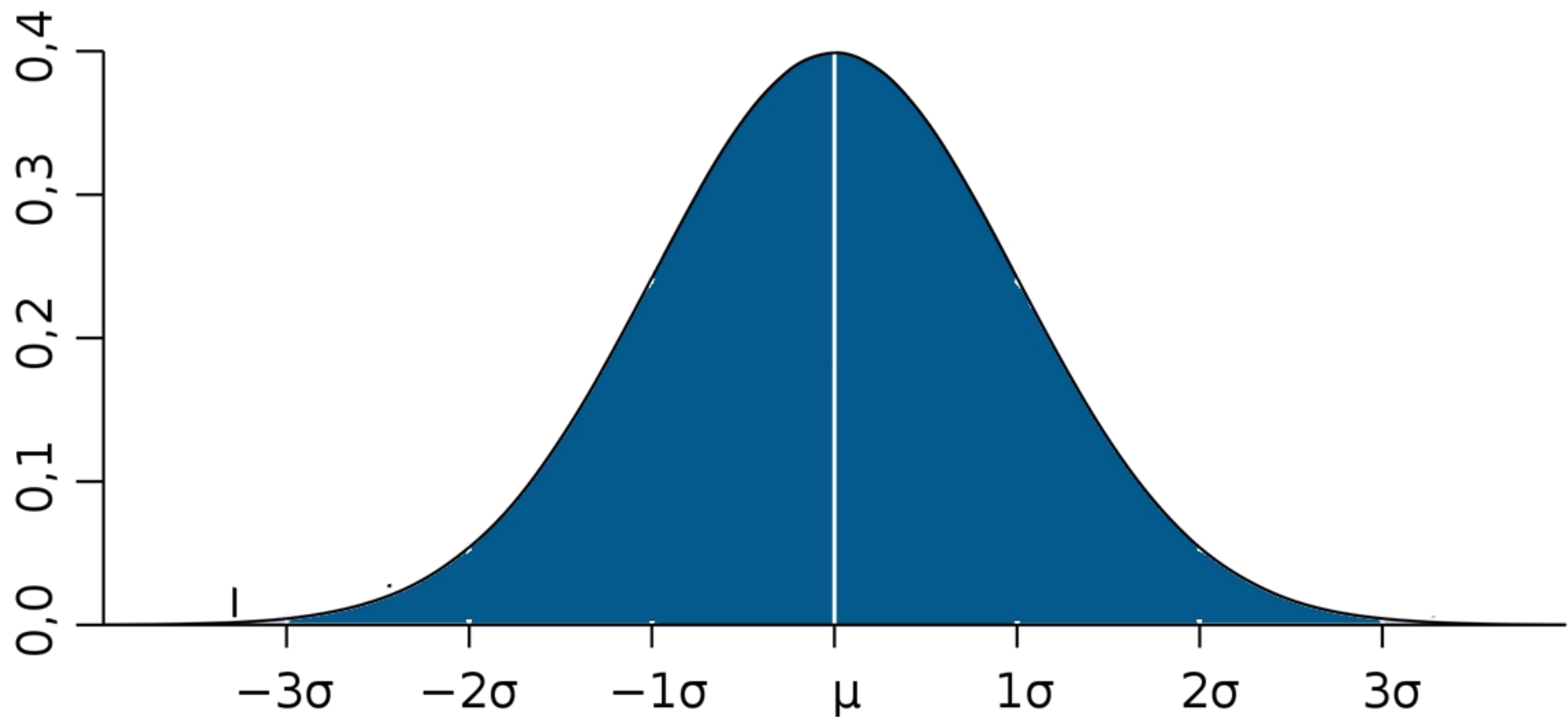
Distribuzione normale: i due parametri chiave

Esistono infinite distribuzioni a seconda della media e varianza



Distribuzione normale

L'area sotto la curva è sempre pari a 1 (100% probabilità)



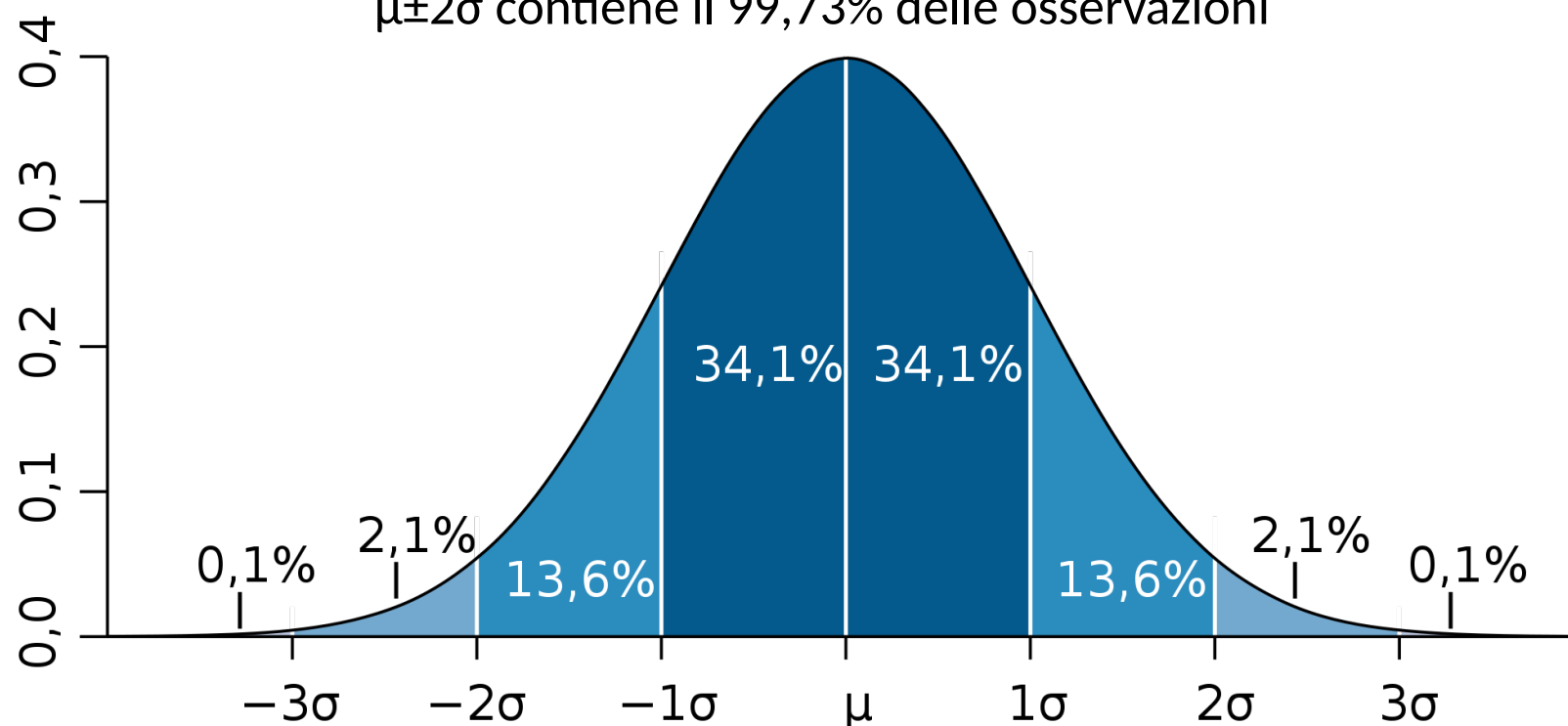
Distribuzione normale

Media e unità di SD definiscono degli intervalli di probabilità

$\mu \pm \sigma$ contiene il 68,27% delle osservazioni

$\mu \pm 2\sigma$ contiene il 95,45% delle osservazioni

$\mu \pm 2\sigma$ contiene il 99,73% delle osservazioni



Distribuzione di probabilità continue: Normale

Media e unità di SD definiscono degli intervalli di valori determinati

$\mu \pm 0.647\sigma$ contiene il 50% delle osservazioni

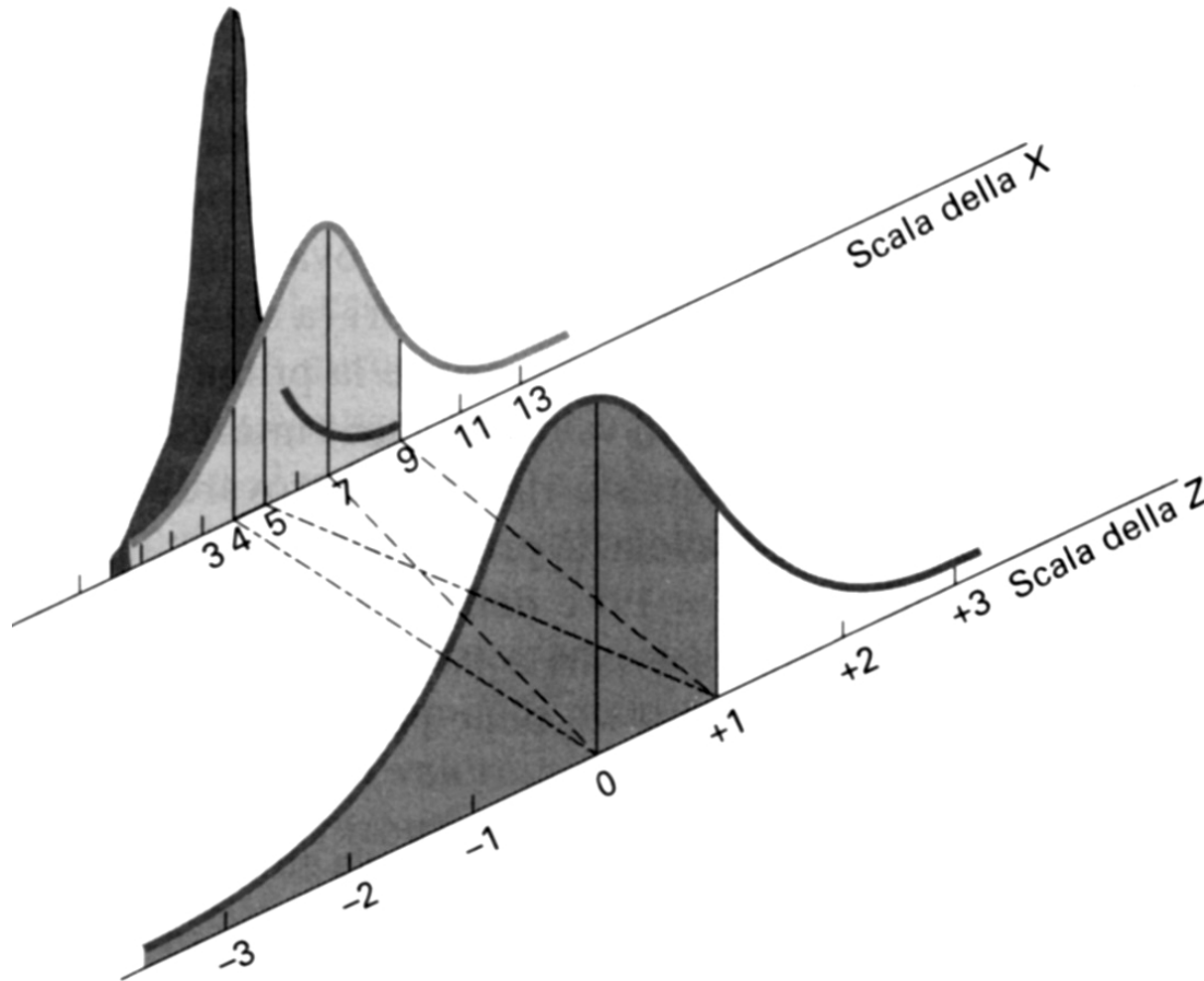
$\mu \pm 1.960\sigma$ contiene il 95% delle osservazioni

$\mu \pm 2.576\sigma$ contiene il 99% delle osservazioni

Come ottenere questi valori?

Curva normale standardizzata z ($\mu=0$ e $\sigma=1$)

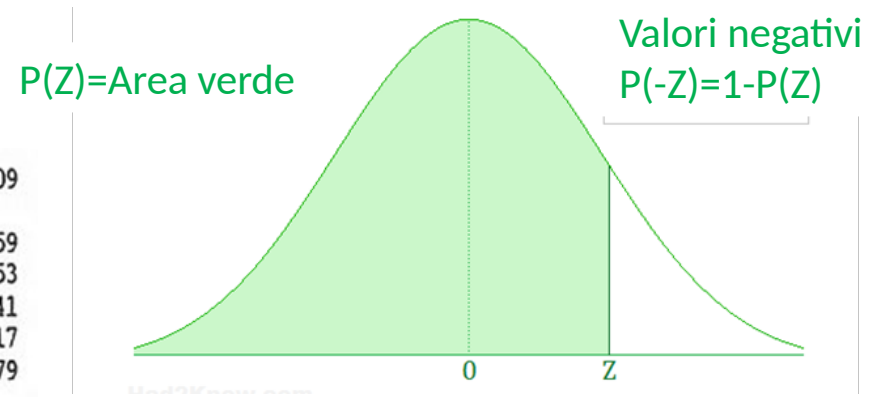
Ogni curva normale con media μ e deviazione standard σ può essere trasformata in una curva normale standardizzata z ($\mu = 0$ e $\sigma = 1$)



$$z = \frac{X - \mu}{\sigma}$$

Curva normale standardizzata z ($\mu=0$ e $\sigma=1$)

Curva normale standard z ($\mu=0$ e $\sigma=1$)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

$$P(1.06) = 0.8554$$

$$P(-1.06) = 1 - 0.8554 = 0.1446$$

La distribuzione normale: Usi

1. Testare semplici ipotesi sulla probabilità di estrarre da una popolazione osservazioni con determinati valori della variabile (es. qual è la probabilità di estrarre uno studente che beve più di 10 birre alla settimana?)

2. Molti test e analisi statistiche richiedono, per essere applicati, che le variabili seguano una distribuzione normale (statistica inferenziale)
(vedi prossime lezioni)



Valutare la normalità di una distribuzione (assunzioni)

Esempio 1

Qual è la probabilità di estrarre uno studente che beva più di 5 birre alla settimana?



Procedimento

1. Ci serve media e dev. st. della popolazione ($\mu = 3$ e $\sigma = 2$)

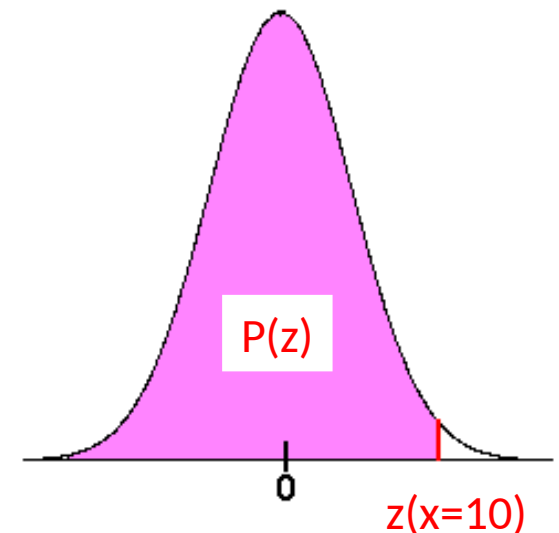


2. Trasformare X in z

3. Cercare il valore di $z(x=5)$

4. Trovare la valore di probabilità $P(z)$ di $z(x=5)$

5. Area a destra o sinistra del valore?



Esempio 2

Qual è la probabilità che uno studente beva da 1 a 3 birre alla settimana?