Note su Probabilità

LA STATISTICA

Trilussa

Carlo Alberto **Salustri**, più conosciuto con lo pseudonimo di **Trilussa** - **anagramma del suo cognome** – è nato a Roma il 26 ottobre 1871 ed è morto a Roma il 21 dicembre 1950. Senatore a vita della R.I. da 1/12/1950 a 21/12/1950.

Il pollo di Trilussa

Sai ched'è la statistica? È na' cosa
che serve pe fà un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che spósa.

Ma pè me la statistica curiosa
è dove c'entra la percentuale,
pè via che, lì,la media è sempre eguale
puro co' la persona bisognosa.

Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:

e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perch'è c'è un antro che ne magna due.

Trilussa, con la sua produzione poetica, ha commentato circa **cinquant'anni di cronaca romana e italiana, dall'età giolittiana agli anni del fascismo e a quelli del dopoguerra**. La corruzione dei politici, il fanatismo dei gerarchi, gli intrallazzi dei potenti sono alcuni dei suoi bersagli preferiti.

Ai tempi di Trilussa mangiare pollo era considerata "una cosa da ricchi". Oggi la situazione è diversa, ma non cambia il **significato del ragionamento**.

Ci si dimentica che **la "media", da sola, è un dato poco significativo** se non sappiamo a che cosa si riferisce, su quale base è calcolata, con quale criterio è definita.

La scienza statistica, peraltro, dispone di strumenti che permettono di tenere conto di questa variabilità. Il tema del "pollo di Trilussa" esemplifica bene la **sovrapposizione che si fa a livello popolare tra la statistica in generale**, che contiene tra l'altro delle misure di dispersione, **e la media statistica**, che è una misurazione tanto nota ed esaltata da essere spesso confusa con la statistica stessa.

Con questa poesia Trilussa anticipa un tema che è diventato assai attuale con la **diffusione dell'informazione statistica per fini di promozione** politica, economica e non solo.

Probabilità:

P = rapporto tra il numero dei casi favorevoli rispetto al numero totale dei casi possibili

$0 \le P \le 1$ ← certezza!

- Lancio di 1 moneta:

  $\# tot = 2 \quad (T \text{ o } C)$

  $p(T) = \frac{1}{2} = 50\%$

  $p(C) = \frac{1}{2} = 50\%$

  $p(T \text{ o } C) = \frac{2}{2} = 100\%$

- Lancio di 1 dado:

  $\# tot = 6 \quad (1, 2, 3, 4, 5, 6)$

  $p(2) = \frac{1}{6} \simeq 16.7\%$

  $p(\text{pari}) = \frac{3}{6} = 50\%$

  $p(1 \div 6) = \frac{6}{6} = 100\%$

probabilità per eventi composti da più eventi semplici:

1) eventi incompatibili o mutuamente esclusivi ovvero basta che se ne verifichi almeno uno di essi

$$\boxed{P = \sum_k P_k}$$

- lancio di 1 dado:

$$P(2 \text{ o } 3) = \frac{2}{6} \simeq 33.3\%$$

$$P(2 \text{ o } 3) = P(2) + P(3) =$$

$$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \simeq 33.3\%$$

2) eventi indipendenti: si chiede che si verifichino simultaneamente

$$\boxed{P = \prod_k P_k}$$

- lancio di 2 dadi:

$$P(\underset{\text{sul }1°}{2} \text{ e } \underset{\text{sul }2°}{3}) = \frac{1}{36} \simeq 2.8\%$$

$$P(\underset{\text{sul }1°}{2} \text{ e } \underset{\text{sul }2°}{3}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \simeq 2.8\%$$

- 13 al TOTOCALCIO ...... mettendo 1, 2, X puramente a caso!

$$P(13) = \underbrace{\frac{1}{3} \times \frac{1}{3} \times \dots \times \frac{1}{3}}_{13 \text{ volte}} = \left(\frac{1}{3}\right)^{13} \sim 6 \cdot 10^{-7}$$

$\downarrow$

$1.6 \cdot 10^{6}$ "colonne"

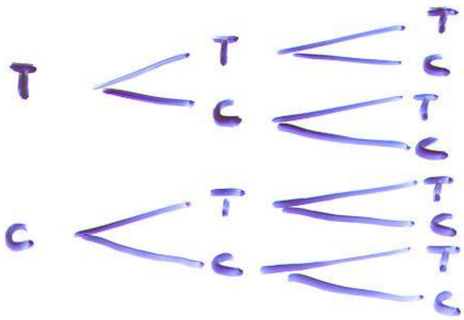
- TERNO su di una "ruota"

$$P = \frac{1}{90} \times \frac{1}{89} \times \frac{1}{88} \times 5 \times 4 \times 3 \sim 8.5 \times 10^{-5}$$

$\downarrow$

11748 "cartelle"

- lancio 3 monete



$\longrightarrow$

TTT
TTC
TCT
TCC
CTT
CTC
CCT
CCC

$P(0T) = P(3C) = \frac{1}{8} = 12.5\%$

$P(1T) = P(1T \,\&\, 2C) = P(TCC) + P(CTC) + P(CCT) =$
$= \frac{3}{8} = 37.5\%$

$P(2T) = P(2T \,\&\, 1C) = P(TTC) + P(TCT) + P(CTT) =$
$= \frac{3}{8} = 37.5\%$

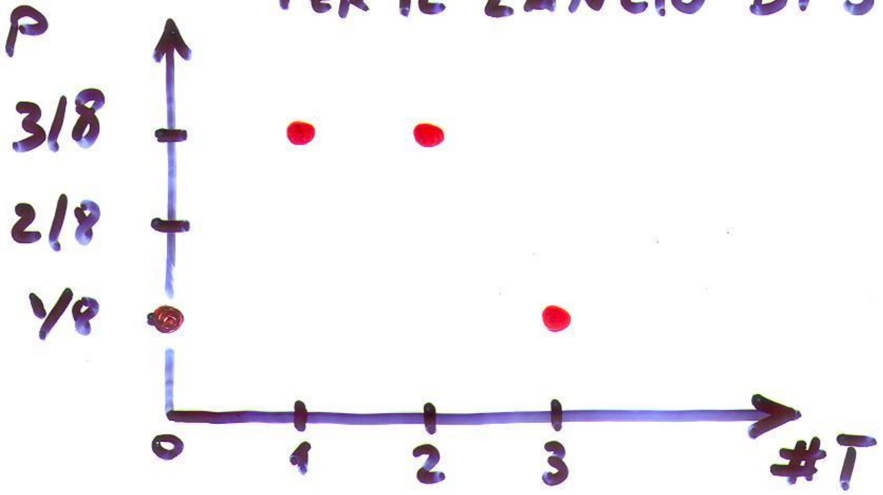
%.

$$P(3T) = P(0C) = \frac{1}{8} = 12.5\%$$

La somma di tutte le probabilità è 1, cioè è **normalizzata**

DISTRIBUZIONE DI PROBABILITÀ

PER IL LANCIO DI 3 MONETE:



$$\rightarrow \sum_{k=0}^{k=3} P_k = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$$

- Posso stimare la probabilità che si verifichi un evento di un certo tipo ripetendo più volte un certo esperimento calcolando la frequenza

$$\boxed{f_{exp} = \frac{\text{numero di volte che si verifica l'evento che mi interessa}}{\text{numero totale delle prove effettuate}}}$$

ebbene, la frequenza misurata (f_{exp}) si avvicinerà alla probabilità teorica (p_{th}) al crescere del numero delle prove indipendenti (N)

Legge dei grandi numeri o
Legge empirica del caso
(Teorema di Bernoulli)

scelto un $\varepsilon > 0$ comunque piccolo, allora:

$$\lim_{N \to \infty} p\left(\left| f_{exp} - p_{th} \right| < \varepsilon \right) = \underbrace{1}_{\uparrow \text{ certezza!}}$$

$$\cancel{\lim_{N \to \infty} f_{exp}} = p_{th}$$

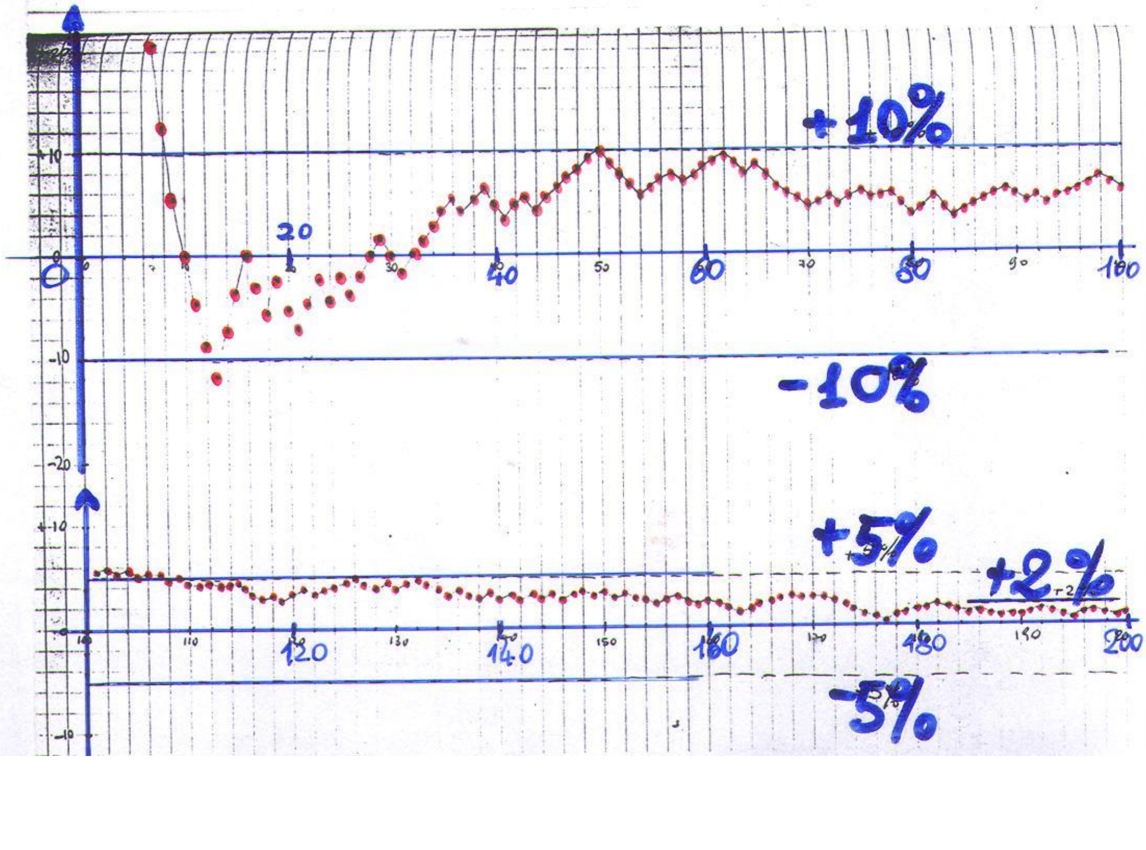
⇒ Determinazione sperimentale della p_{th}

- ESPERIMENTO DI STATISTICA: lancio di N monete o di una moneta per N volte

$$P_{th}(T) = \frac{1}{2} = 50\%$$

$$f_{exp}(T) = \frac{n(T)}{N}$$

$$\boxed{(P_{th} - f_{exp}) \text{ vs. } N}$$



- Campione di una popolazione

trovato sperimentalmente osservando un numero finito N di prove ripetute

P(n): distribuzione teorica di probabilità per l'evento "n"

F(n): numero di volte che si è riscontrato l'evento "n" su N prove

N·P(n): numero delle volte atteso di trovare nelle N prove l'evento "n"

$$\boxed{\frac{F(n) - NP(n)}{N \cdot P(n)} \searrow \text{ per } N \nearrow}$$

Al crescere di N ci aspettiamo un "avvicinamento" relativo, non assoluto

→ servono dei criteri per accettare o meno la differenza che si trova tra il campione finito e la distribuzione teorica ..... χ^2....

→ parametri globali rappresentativi del campione misurato
valore medio, deviazione standard
quali approssimazioni dei parametri della distribuzione teorica.

- **valore medio** : $E(n)$

$$\langle n\rangle = \sum_{0}^{N}{}_{n}\, n\cdot P(n) = \sum_{1}^{N}{}_{n}\, n\cdot P(n)$$

$[n - E(n)]$ $\qquad$ $E[n - E(n)]$

- **scarto dalla media** e **scarto medio**:

$$\langle (n - \langle n\rangle)\rangle = \sum_{0}^{N}{}_{n} \left(n - \langle n\rangle\right)\cdot P(n) =$$

$$= \sum_{1}^{N}{}_{n}\, n\cdot P(n) - \sum_{0}^{N}{}_{n} \langle n\rangle\cdot P(n) =$$

$$= \langle n\rangle - \langle n\rangle = 0$$

- **varianza** $\quad E\left[(n - E(n))^2\right] = \sigma_n^2$

$$\langle (n - \langle n\rangle)^2\rangle = \sum_{0}^{N}{}_{n}\left(n - \langle n\rangle\right)^2\cdot P(n) =$$

$$= \sum_{0}^{N}{}_{n}\left(n^2 + \langle n\rangle^2 - 2\langle n\rangle n\right)\cdot P(n) =$$

$$= \sum_{1}^{N}{}_{n}\, n^2\cdot P(n) + \langle n\rangle^2 \underbrace{\sum_{0}^{N}{}_{n} P(n)}_{1} - 2\langle n\rangle \underbrace{\sum_{1}^{N}{}_{n}\, n\cdot P(n)}_{\langle n\rangle = E(n)}$$

$$= \langle n^2\rangle - \langle n\rangle^2$$

..... applicando le definizioni precedenti al caso del lancio di 3 monete :

(↑@p 5 e 6)

$P(0T) = 0.125$
$P(1T) = 0.375$
$P(2T) = 0.375$
$P(3T) = 0.125$

↪ condizione di normalizzazione :

$$\sum_{n=0}^{3} P(nT) = 0.125 + 0.375 + 0.375 + 0.125 = 1.$$

↪ valore atteso :

$$\sum_{n=0}^{3} n \cdot P(nT) = \sum_{n=1}^{3} n \cdot P(nT)$$

$$= 0 \times 0.125 + 1 \times 0.375 + 2 \times 0.375 + 3 \times 0.125 = 1.5$$

↪ valore atteso dello scarto dal valore atteso

$$\sum_{n=0}^{3} (n - \langle n \rangle) \cdot P(nT) =$$

$$= (0 - \cancel{1.5}) \times 0.125 + (1 - \cancel{1.5}) \times 0.375 +$$

$$+ (2 - \cancel{1.5}) \times 0.375 + (3 - \cancel{1.5}) \times 0.125 = 0$$

deviazione standard o radice quadrata della varianza:

$$\sqrt{\sum_{0}^{3}{}_{n} (n-\langle n\rangle)^2 \cdot P(n,T)} =$$

$$= \left[(0.-1.5)^2\times 0.125 + (1-1.5)^2 \times 0.375 + (2-1.5)^2\times 0.375 + (3-1.5)^2 \times 0.125\right]^{1/2} =$$

$$= \sqrt{0.28125 + 0.09375 + 0.09375 + 0.28125} =$$

$$= \sqrt{0.75} = 0.866025 \to \sigma_n = 0.87$$

ATTENZIONE AGLI ESTREMI DELLA $\sum(\dots)$!

$$\boxed{\sigma_n = \sqrt{\langle n^2 \rangle - \langle n \rangle^2}}$$

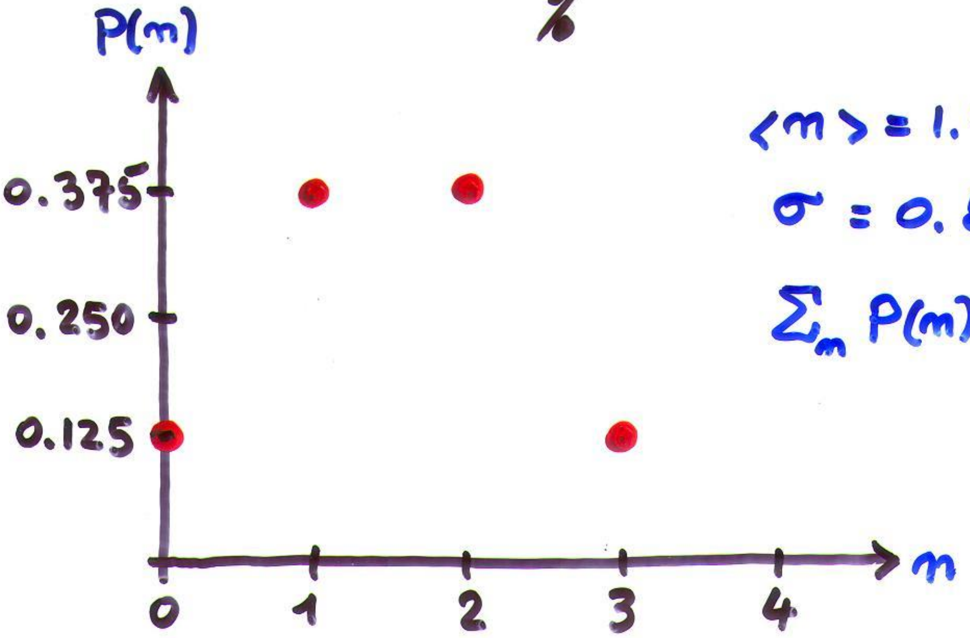
$$\langle n^2 \rangle = \sum_{0}^{3}{}_{n} \, n^2 \cdot P(nT) =$$

$$= 0 \times 0.125 + 1 \times 0.375 + 4 \times 0.375 + 9 \times 0.125 =$$

$$= 3$$

$$\hookrightarrow \sqrt{\langle n^2 \rangle - \langle n \rangle^2} = \sqrt{3 - (1.5)^2} = \sqrt{0.75} = 0.87$$

%



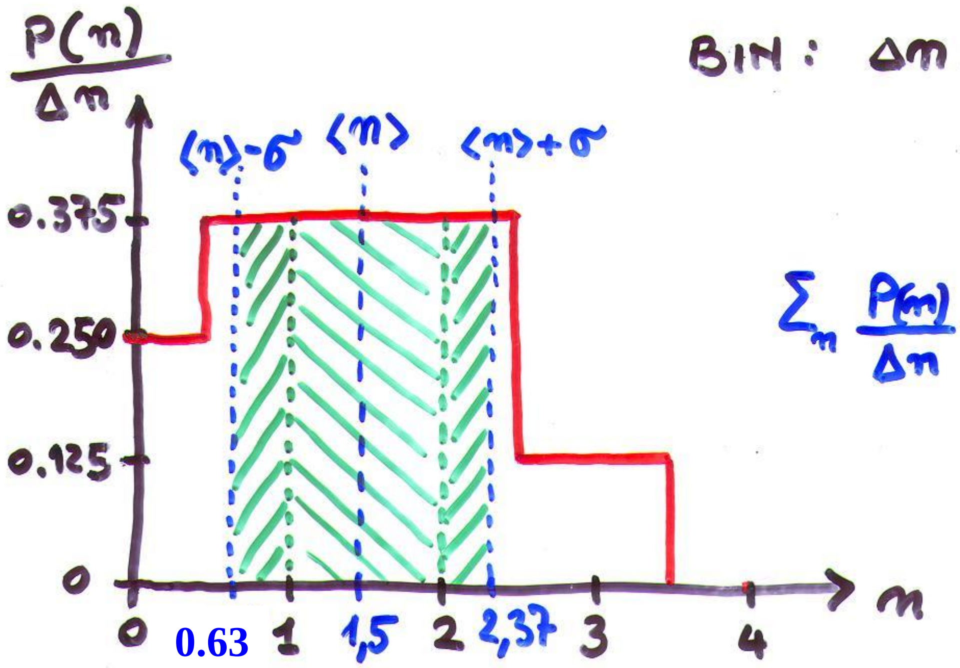
$\langle n \rangle = 1.50$

$\sigma = 0.87$

$\sum_n P(n) = 1$

$P(n = 1 \text{ o } n = 2) = 0{,}75$

BIN : $\Delta n = 1$



$\sum_n \frac{P(n)}{\Delta n} \cdot \Delta n = 1$

$$P(0{,}63 \leq n \leq 2{,}37) = 0{,}375 \times \left[(1 - 0{,}63) + 1 + (2{,}37 - 2) \right] =$$

$$= 0{,}6525 = 65{,}25\%$$

- **deviazione standard**

$$\sqrt{\langle (n - \langle n \rangle)^2 \rangle} = \ldots = \sqrt{\langle n^2 \rangle - \langle n \rangle^2}$$

→ utilizzando il campione finito della popolazione relativo alle N prove ripetute si calcolano gli estimatori dei parametri globali:

- $$\langle n \rangle = \sum_{1}^{N} n \cdot \frac{F(n)}{N}$$

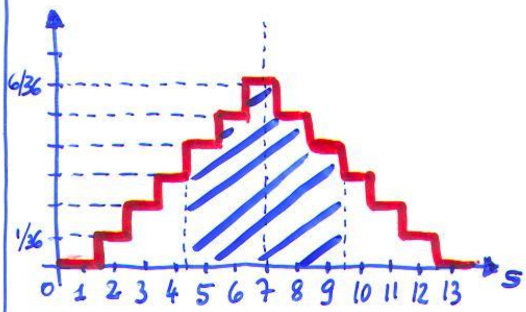
- $$\sigma = \sqrt{\sum_{0}^{N} \left[(n - \langle n \rangle)^2 \frac{F(n)}{(N-1)} \right]}$$

r. m. s.

• lancio di 2 dadi : $2 \leq S \leq 12$

S		P (S)	
2	1+1	$1 \cdot \frac{1}{36}$	2,8 %
3	1+2 2+1	$2 \cdot \frac{1}{36}$	5.6 %
4	1+3 2+2 3+1	$3 \cdot \frac{1}{36}$	8.3 %
5	1+4 2+3 3+2 4+1	$4 \cdot \frac{1}{36}$	11.1 %
6	1+5 2+4 3+3 4+2 5+1	$5 \cdot \frac{1}{36}$	13.9 %
7	1+6 2+5 3+4 4+3 5+2 6+1	$6 \cdot \frac{1}{36}$	16.7 %
8	2+6 3+5 4+4 5+3 6+2	$5 \cdot \frac{1}{36}$	13.9
9	3+6 4+5 5+4 6+3	$4 \cdot \frac{1}{36}$	11.1 %
10	4+6 5+5 6+4	$3 \cdot \frac{1}{36}$	8.3 %
11	5+6 6+5	$2 \cdot \frac{1}{36}$	5.6 %
12	6+6	$1 \cdot \frac{1}{36}$	2.8 %
			100.0 %

distribuzione di probabilità simmetrica intorno al valore $S=7$



$$\langle S \rangle \equiv \sum_{k=2}^{k=12} k \cdot P(k) =$$

$$= 2 \cdot P(2) + 3 \cdot P(3) + 4 \cdot P(4) + 5 \cdot P(5) + 6 \cdot P(6) + 7 \cdot P(7) + 8 \cdot P(8) + 9 \cdot P(9) + 10 \cdot P(10) + 11 \cdot P(11) + 12 \cdot P(12) =$$

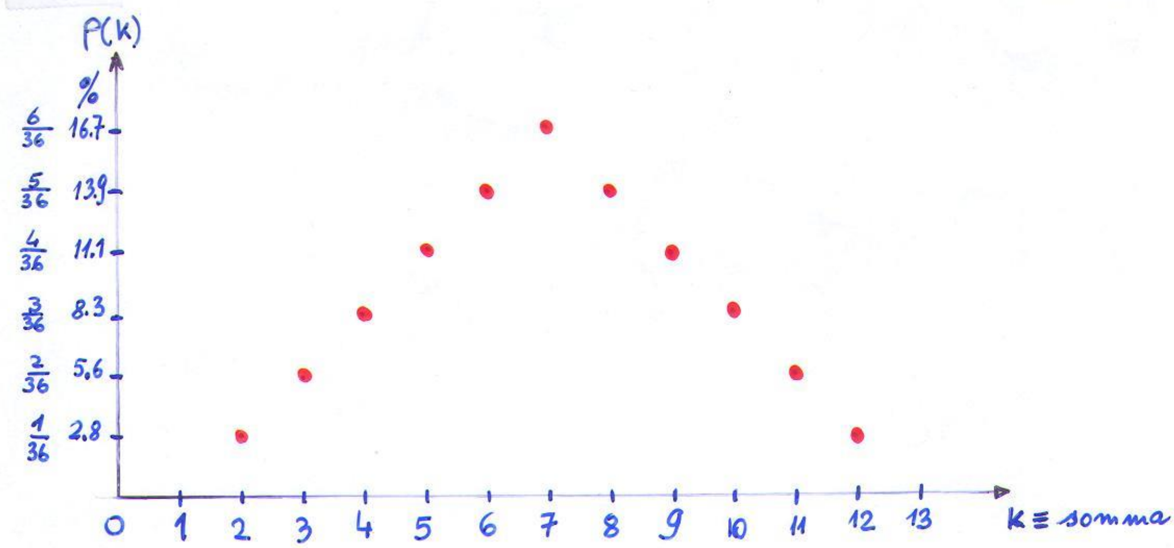
$$= \frac{2}{36} + \frac{6}{36} + \frac{12}{36} + \frac{20}{36} + \frac{30}{36} + \frac{42}{36} + \frac{40}{36} + \frac{36}{36} + \frac{30}{36} + \frac{22}{36} + \frac{12}{36} =$$

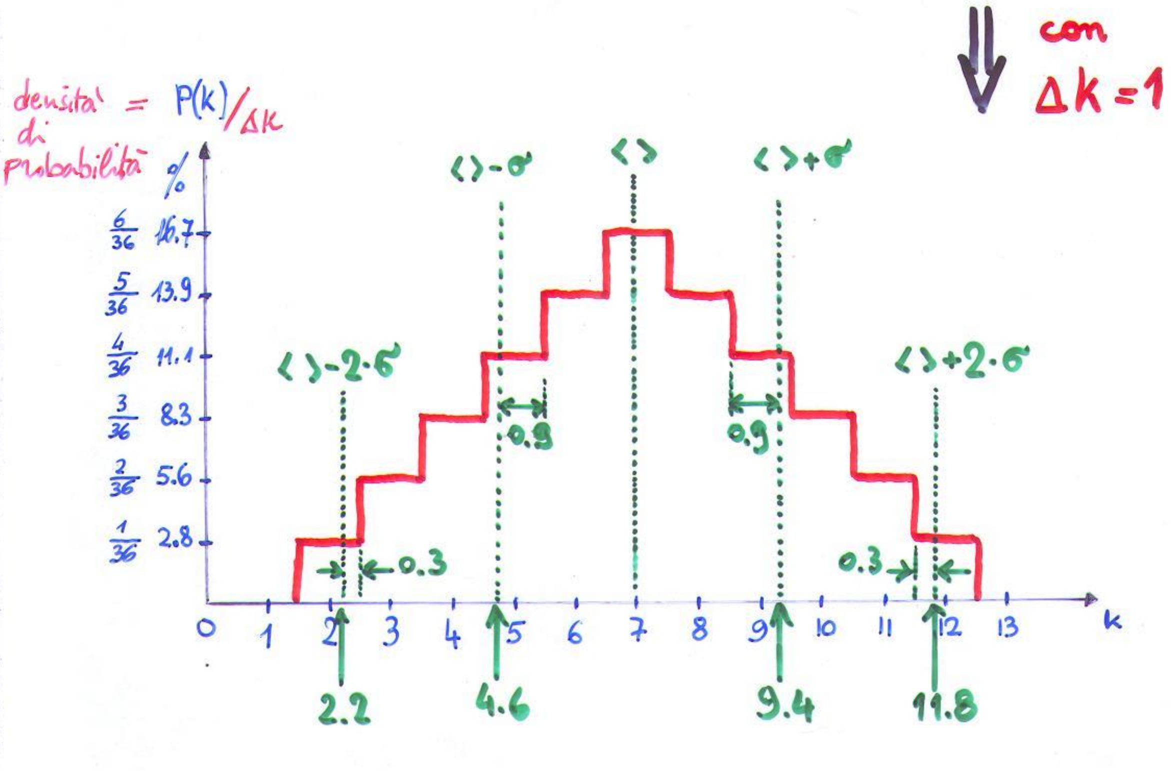
$$= \frac{252}{36} = 7$$

$$\sigma \equiv \sqrt{\sum_{2}^{12}{}_{k} \left(k - \langle S \rangle\right)^2 \cdot P(k)} =$$

$$= \ldots = \sqrt{\frac{205}{36}} \approx 2,4$$

$$\sum_{5}^{9}{}_{k} P(k) = \frac{4+5+6+5+4}{36} = \frac{24}{36} \approx 66,7\,\%$$





$$P\,(4{,}6 < s < 9{,}4) =$$

$$= 11{,}1\% \times 0{,}9 + 13{,}9\% \times 1 + 16{,}7\% \times 1 + 13{,}9\% \times 1 + 11{,}1\% \times 0{,}9 =$$

$$= 64{,}48\%$$

$$\sum_{k=2}^{k=12} p(k) = \frac{1}{36}+\frac{2}{36}+\frac{3}{36}+\frac{4}{36}+\frac{5}{36}+\frac{6}{36}+\frac{5}{36}+\frac{4}{36}+\frac{3}{36}+\frac{2}{36}+\frac{1}{36} = \frac{36}{36} = 1$$

$$\langle k \rangle = \sum_{k=2}^{k=12} k \cdot p(k) = 2\cdot p(2) + 3\cdot p(3) + 4\cdot p(4) + 5\cdot p(5) +$$

$$+ 6\cdot p(6) + 7\cdot p(7) + 8\cdot p(8) + 9\cdot p(9) +$$

$$+ 10\cdot p(10) + 11\cdot p(11) + 12\cdot p(12) = \dots = \frac{252}{36} = 7$$

$$\sum_{k=2}^{k=7} p(k) = \frac{1}{36}+\frac{2}{36}+\frac{3}{36}+\frac{4}{36}+\frac{5}{36}+\frac{1}{2}\times\frac{6}{36} = 0.5$$

$$\sum_{k=7}^{k=12} p(k) = \dots\dots = 0.5$$

$$\sigma = \sqrt{\sum_{k=2}^{k=12} (k-\langle k\rangle)^2 \cdot P(k) = ..} = \sqrt{\frac{205}{36}} \simeq 2.4$$

$$p\big((\langle k\rangle - \sigma) < k < (\langle k\rangle + \sigma)\big) = p(4.6 < k < 9.4) =$$

$$= \frac{4}{36}\times 0{,}9 + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36}\times 0{,}8 = \frac{23{,}2}{36} \simeq 64{,}4\%$$

$$p\big((\langle k\rangle - 2\sigma) < k < (\langle k\rangle + 2\sigma)\big) = p(2.2 < k < 11.8) =$$

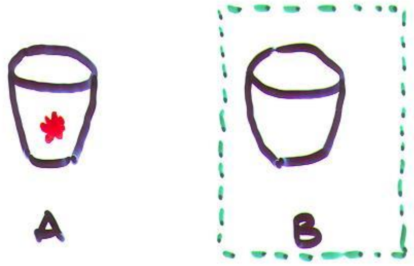
$$= \frac{1}{36}\times 0.3 + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} +$$

$$+ \frac{2}{36} + \frac{1}{36}\times 0.3 = \frac{34.6}{36} \simeq 96.1\%$$

$$p\big((\langle k\rangle - 3\sigma) < k < (\langle k\rangle + 3\sigma)\big) = p(\underset{0}{\cancel{-0.2}} < k < 14.2) = 100\%$$

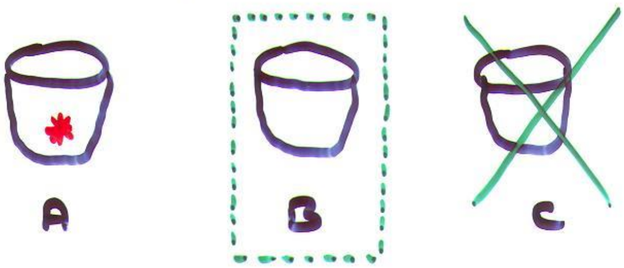
Strategia di gioco vincente :

1) Il BANCO mette di nascosto un oggetto in A



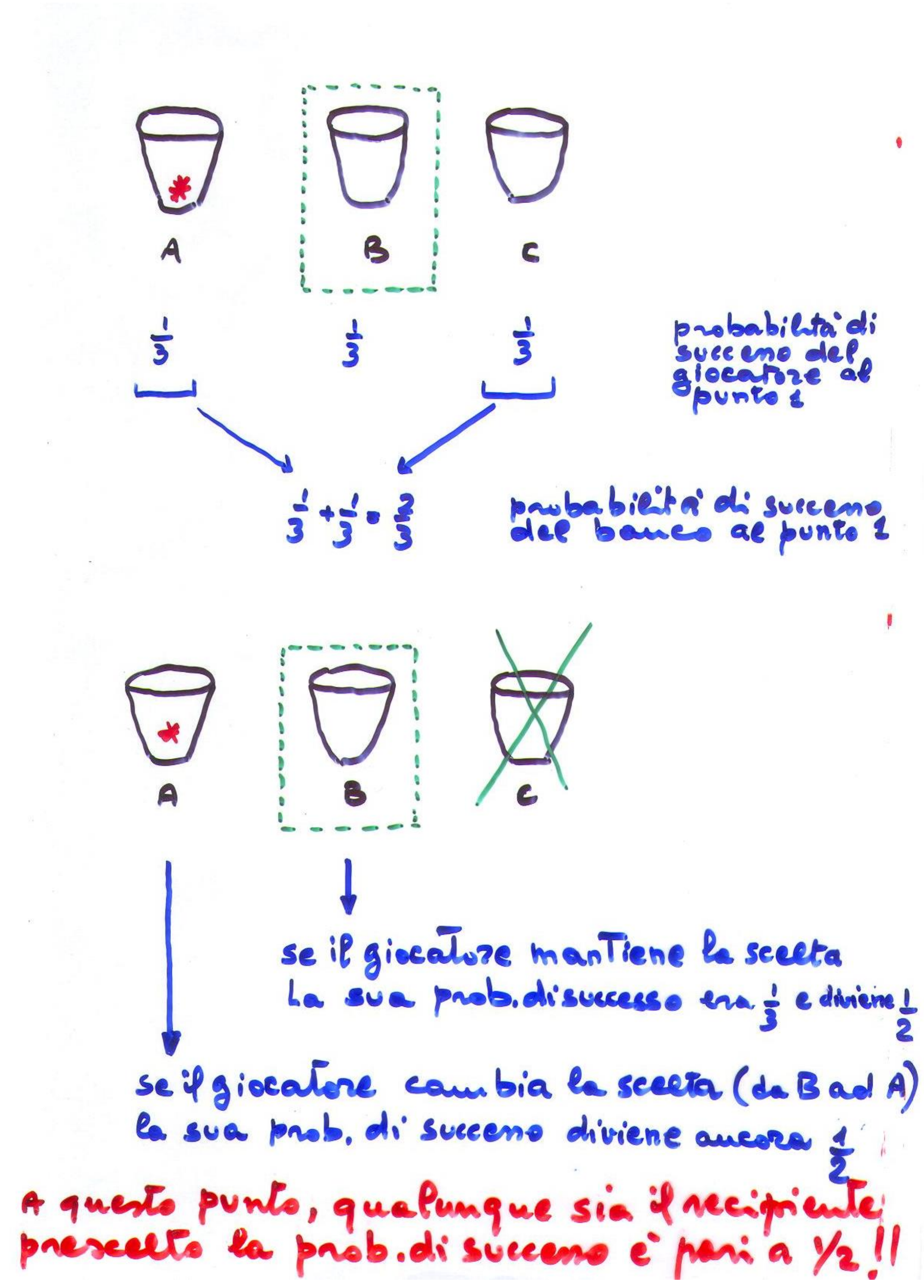
Il GIOCATORE sceglie casualmente uno dei tre recipienti (p. es. B)

2) Il BANCO mostra uno dei due recipienti che sa essere vuoto e non scelto dal giocatore (per es. C)



3) Il GIOCATORE può cambiare il recipiente prescelto o tenerlo ......

⇒ Quale strategia è vincente per il giocatore?



<u>ATTENZIONE</u>:
Conclusione da rivedere criticamente sia dal punto di vista del dilemma del **gioco Monty Hall**, sia alla luce del **Teorema di Bayes.**

- **Permutazioni**: numero di modi con i quali si possono collocare N oggetti distinguibili in N caselle

N caselle disponibili per il 1° da inserire
N−1 " " " " 2° meno il 1°
N−2 " " " " 3° meno i primi 2
⋮
2 " " " " penultimo
1 " " " l'ultimo da inserire

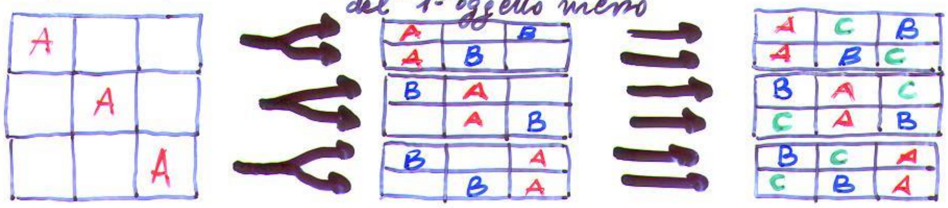
$$\boxed{N\cdot(N-1)\cdot(N-2)\cdot\ldots\cdot 2\cdot 1 \equiv N!}$$

$N=3$: $3! = 6$
(A, B, C)

per l'inserzione del 1° oggetto ho 3 possibilità

per l'inserzione del 2° oggetto ho 2 possibilità per ogni modalità di inserzione del 1° oggetto messo

per l'inserzione del 3° oggetto ho una sola possibilità



Formula di Stirling per il calcolo approssimato del fattoriale

$$\boxed{x! \cong \sqrt{2\pi x}\left(\frac{x}{e}\right)^x}$$

$\pi \simeq 3.14159$
$e \simeq 2.71828$

$\sqrt{2\pi x}\cdot(x/e)^x$	x	$x\cdot(x-1)\cdot(x-2)\cdots3\cdot2\cdot1$
118.0192	5	120
3.5987×10^6	10	3.6288×10^6
3.0363×10^{64}	50	3.0414×10^{64}
1.7092×10^{98}	69	1.7112×10^{98}

Differenza relativa:

5	1.65 %
10	0.83 %
50	0.17 %
69	0.12 %

• **Disposizioni**: numero di modi con i quali si possono collocare m oggetti distinguibili in N caselle

$(N \geq m)$

N caselle disponibili per il 1°

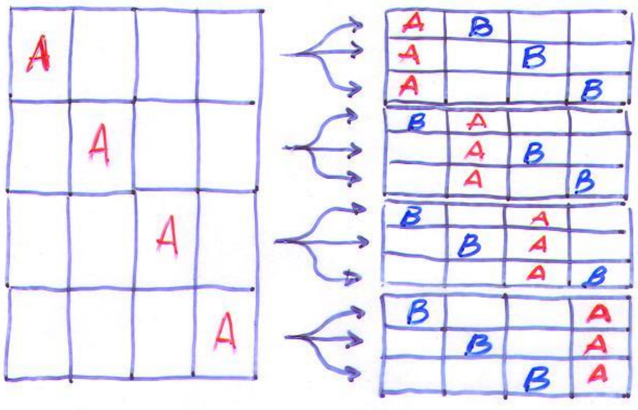
$N-1$ " " " " 2°

$N-2$ " " " " 3°

⋮

$N-m+1$ " " " m-esimo

$$\boxed{\begin{aligned} &N\cdot(N-1)\cdot(N-2)\cdot\ldots\ldots\cdot(N-m+1) = \\ &= N\cdot(N-1)\cdot(N-2)\cdot\ldots\cdot(N-m+1)\cdot\frac{(N-m)\cdot(N-m-1)\cdot\ldots\cdot2\cdot1}{(N-m)\cdot(N-m-1)\cdot\ldots\cdot2\cdot1} = \\ &= \frac{N!}{(N-m)!} \end{aligned}}$$



$m = 2 \quad (A;B)$

$N = 4$

$\Downarrow$

$$\frac{4!}{(4-2)!} = \frac{4\cdot3\cdot2}{2} = 12$$

P.S.:

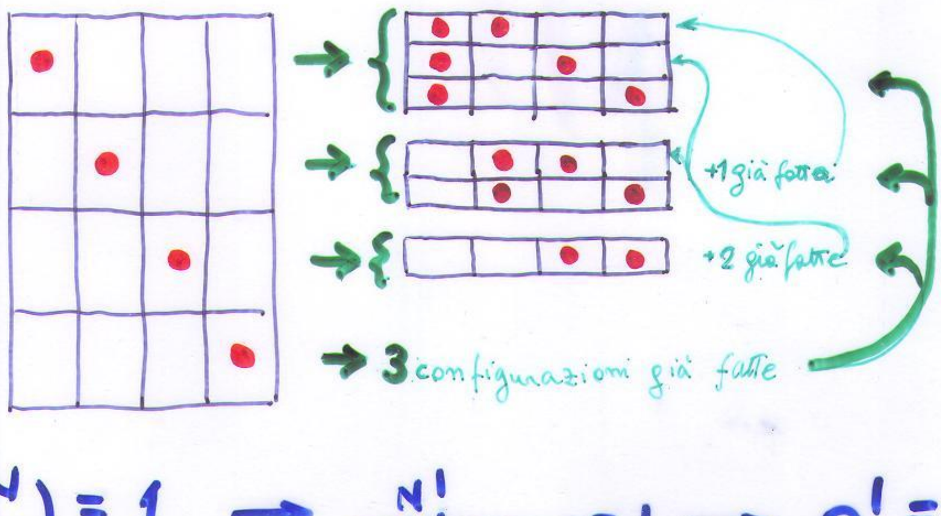
NATURALMENTE, SE $N = m$ DEVO RIOTTENERE LE PERMUTAZIONI $\Rightarrow \dfrac{N!}{(0)!} = N! \Rightarrow 0! = 1$

- **Combinazioni**: numero di modi con i quali si possono collocare **m oggetti identici in N caselle** ($N \geq m$) **indipendentemente dall'ordine**

$$\frac{\text{DISPOSIZIONI DI } m \text{ oggetti in } N \text{ caselle}}{\text{PERMUTAZIONI DI } m \text{ oggetti}} = \frac{N!}{m!\,(N-m)!} = \binom{N}{m}$$

$m = 2$, $N = 4$: $\binom{4}{2} = \dfrac{4!}{2!\,(4-2)!} =$

$$= \frac{4 \cdot 3 \cdot 2}{2 \cdot 2} = 6$$



P.S.: $\binom{N}{N} = 1 \longrightarrow \dfrac{N!}{N!\,(N-N)!} = 1 \longrightarrow 0! = 1$