PROBABILITÀ E VARIABILI CASUALI

Frequenza relativa e probabilità

Mediante le **probabilità** si descrivono i fenomeni che possono essere "pensati" come un "*esperimento*" il cui risultato sia soggetto a cambiamento al ripetersi dell'esperimento stesso (pur mantenendo le medesime condizioni operative).

Esempio:

Esperimento: Lancio "casuale" di un dado (ogni volta in modo leggermente diverso)

Risultato: Numero sulla faccia superiore del dado

Insieme dei possibili risultati (elementari): {1,2,3,4,5,6}

Evento: qualsiasi sottoinsieme dell'insieme dei risultati A={1,2}; B={2,4,6}; ecc.

Se si esegue un numero N di prove sufficientemente elevato, sia l'esperienza sia la teoria della probabilità mostrano che la frequenza relativa f_k dei singoli risultati ($k=1,2,3,4,5,6$) (o di un qualsiasi evento) è prossima alla loro probabilità:

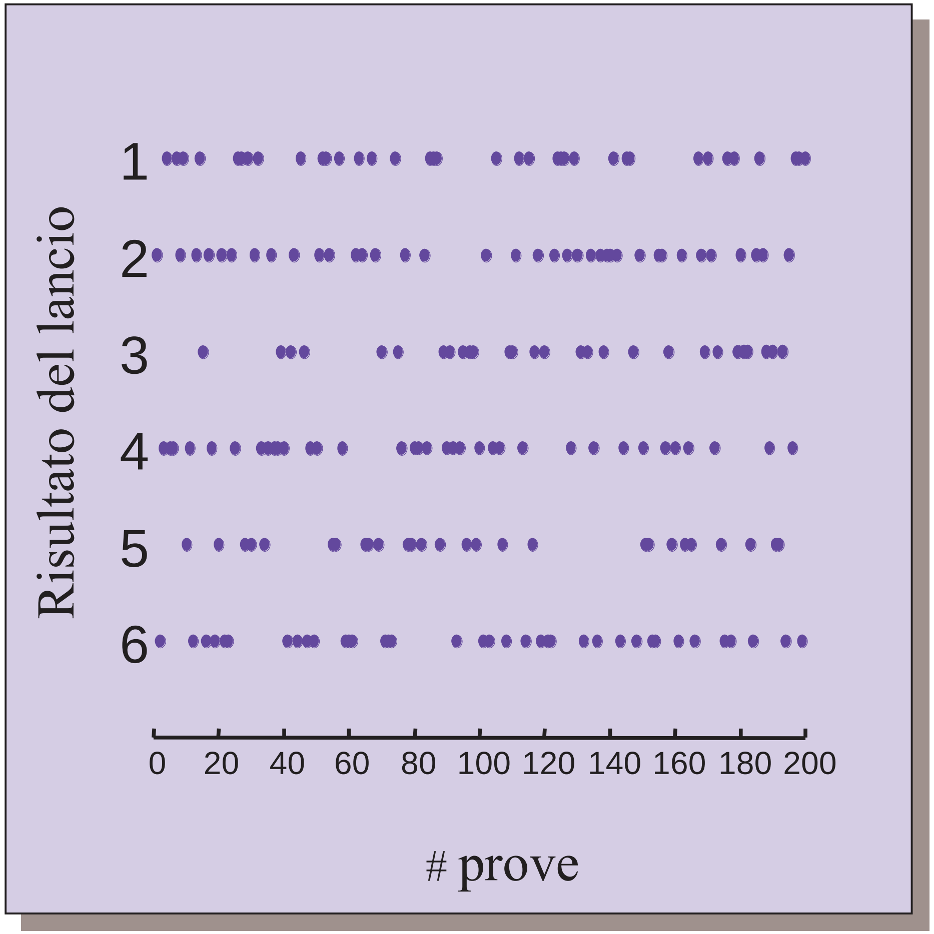
$$f_k = \frac{N_k}{N} \approx P(k) \qquad f_A = \frac{N_A}{N} \approx P(A) \qquad f_B = \frac{N_B}{N} \approx P(B) \quad \ldots$$

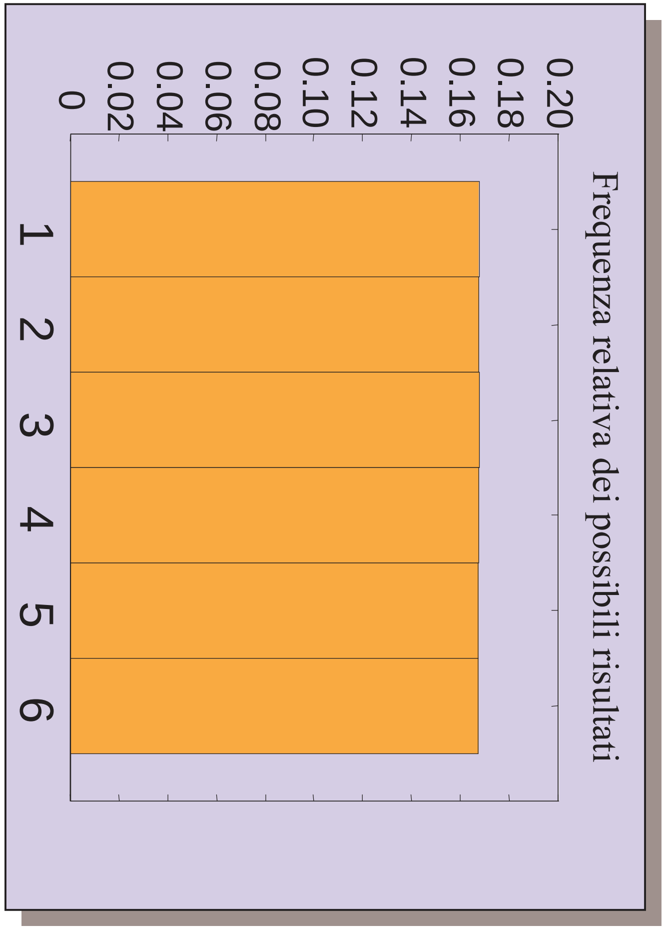
ATTENZIONE: $\boxed{0 \le f_A \le 1}$ $\boxed{0 \le P(A) \le 1}$

Istogramma dei risultati

L'istogramma dei risultati è il grafico delle frequenze relative

Lancio di un dado non truccato, esito di una serie di prove





Commento: questo istogramma è sospetto! è troppo regolare!!

Cenni di teoria della probabilità (1)

S = spazio degli eventi, cioè insieme di tutti i risultati elementari
A,B,C,... = eventi (sottoinsiemi di *S*, inclusi lo stesso *S* e l'insieme vuoto)
$A \cup B$ = unione di *A* e *B*
$A \cap B$ = intersezione di *A* e *B*
Nota: la probabilità di $A \cup B$ è spesso indicata con *P(A+B)*.
Nota: la probabilità di $A \cap B$ è indicata con *P(A,B)* e detta probabilità congiunta.

Assiomi della teoria della probabilità (proprietà delle probabilità):

1. Per ogni *A* esiste (cioè è definita) $P(A) \geq 0$
2. $P(S)=1$
3. Se *A* e *B* sono mutuamente esclusivi (hanno intersezione nulla) $P(A+B)=P(A)+P(B)$

Nota: non sono altro che le proprietà elementari della frequenza relativa

Conseguenza (facilmente dimostrabile): $P(A+B)=P(A)+P(B)-P(A,B)$

bisogna contare solo una volta l'intersezione di *A* e *B*!

Si attribuiscono alla probabilità le proprietà della frequenza relativa, perché i risultati del calcolo delle probabilità siano a loro volta interpretabili come frequenze relative.

Cenni di teoria della probabilità (2)

Esempio: lancio di un dado (ipotesi: dado non truccato ==> risultati equiprobabili)
A={1,2,3} (l'evento A si verifica se il risultato elementare è contenuto in A)
B={2,4,6} (l'evento B si verifica se il risultato è un numero pari)

Per calcolare la probabilità di un evento basta contare i risultati che lo compongono!
$P(A)=n_A/n$ dove n_A è il numero di risultati contenuti in A e n il numero totale di risultati.

$P(A)=P(B)=3/6$
$P(A+B)=P(\{1,2,3,4,6\})=5/6$
(o anche $P(A+B)=P(A)+P(B)-P(A,B)=3/6+3/6-1/6=5/6$, ma in questo caso non conviene)

$P(A)=n_A/n$ può essere una definizione generale di probabilità? NO perché esistono anche i dadi truccati (intenzionalmente o no); potrebbe essere $P(1)=0.5$ e $P(2)=...=P(6)=0.1$.
In questo caso si avrebbe $P(A)=0.7$, $P(B)=0.3$ e $P(A+B)=0.9$.

Regola generale: nel caso dei risultati equiprobabili, il calcolo delle probabilità richiede solo di saper contare; se i risultati non sono equiprobabili occorre saper sommare.

N.B.: il numero di termini da sommare può essere enorme, o addirittura infinito!

Indipendenza statistica

Esempio: lancio di due dadi (non truccati)

A={1 nel primo lancio (e risultato qualsiasi nel secondo lancio)}
B={3 o 4 nel secondo lancio (e risultato qualsiasi nel primo lancio)}
P(A,B)=*P*({1 nel primo lancio, 3 o 4 nel secondo lancio})=2/36
(infatti vi sono 36 coppie di risultati equiprobabili: le coppie (1,3) e (1,4) costituiscono l'evento congiunto)

In questo caso risulta *P(A,B) =P(A)P(B)*, cioè la probabilità congiunta è uguale al prodotto delle probabilità: si dice che gli eventi *A* e *B* sono statisticamente indipendenti (o indipendenti). Effettivamente i lanci sono indipendenti, a meno che si voglia credere che il dado ha memoria!!

Si assume *a priori* l'indipendenza statistica, e quindi si usa la regola *P(A,B)=P(A)P(B)*, quando *A* e *B* sono eventi relativi a esperimenti indipendenti. Esempio tipico: ripetizione di uno stesso esperimento, cioè "prove ripetute" (dette anche "prove di Bernoulli").

N.B.: nel caso di esperimenti indipendenti vale la regola *P(A,B)=P(A)P(B)* anche se i risultati elementari non sono equiprobabili (dado truccato).

Variabile casuali discrete (distribuzione di probabilità)

Si dice **variabile casuale** un numero reale associato al risultato dell'esperimento.

Se i possibili risultati sono numerabili la variabile casuale è detta discreta.

Ad esempio all'esperimento "lancio del dado" (non truccato) associamo la variabile casuale *x* che può assumere i valori interi compresi tra 1 e 6 con probabilità 1/6. Nota: se invece vogliamo indicare le facce del dado con *a,b,c,d,e,f* non definiamo una variabile casuale.

Si dice distribuzione di probabilità (o talvolta densità discreta di probabilità) della variabile casuale *x* la funzione *P(a)* (o talvolta *p(a)*), che rappresenta con quale probabilità la variabile casuale *x* assume il valore *a.*

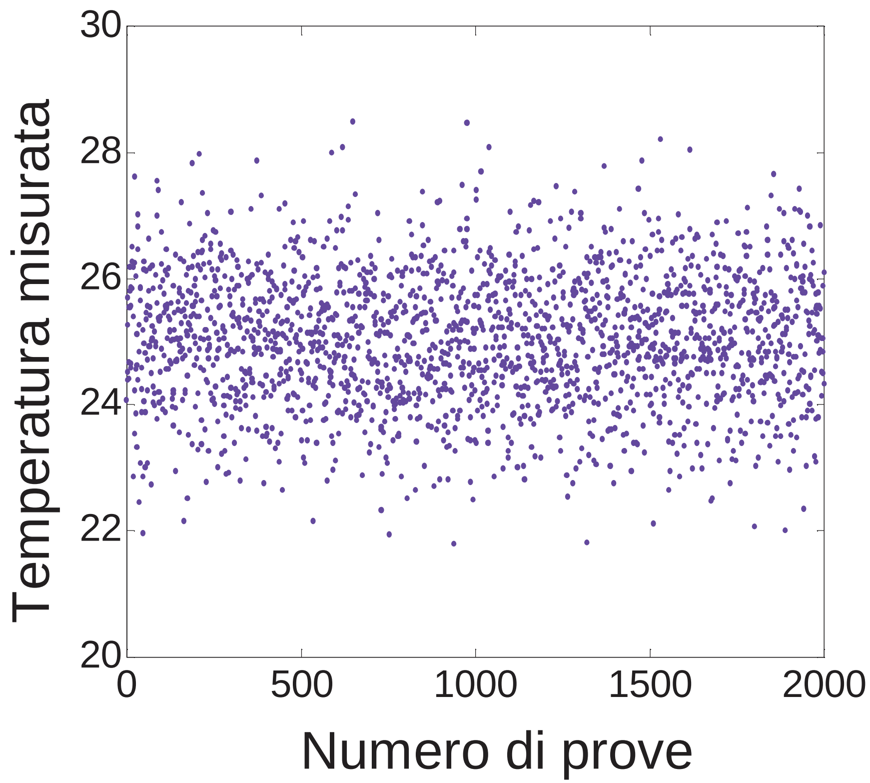
Se i risultati sono in numero finito si tratta di una rappresentazione del tutto equivalente ad una tabella contenente le probabilità P(a).

Nell'esperimento "lancio del dado" la distribuzione di probabilità della variabile casuale *x* vale *1/6* per i valori di *a* interi compresi tra *1 e 6,* e zero altrove.

Variabili casuali continue

Le variabili casuali sono continue quando possono assumere un insieme continuo di valori (e quindi i possibili risultati sono in numero infinito).

Esempio: la temperatura di una stanza misurata ad un istante di tempo casuale (con precisione infinita! ottenendo cioè un numero reale).



Il concetto di frequenza relativa viene recuperato approssimando l'insieme continuo di valori con un numero finito di intervallini di misura (discretizzazione).

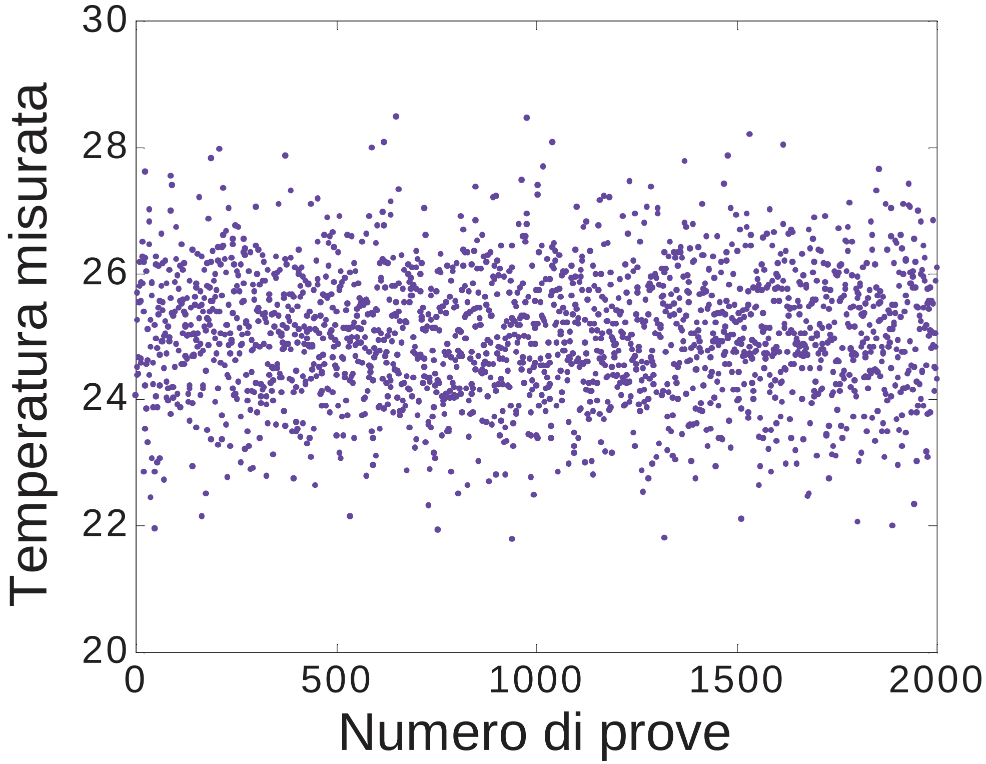
Ad esempio, se la temperatura della stanza può variare con continuità tra 20 e 30 gradi, non commettiamo un grosso errore approssimando l'intervallo continuo con 50 intervallini contigui di 0.2 gradi ciascuno.

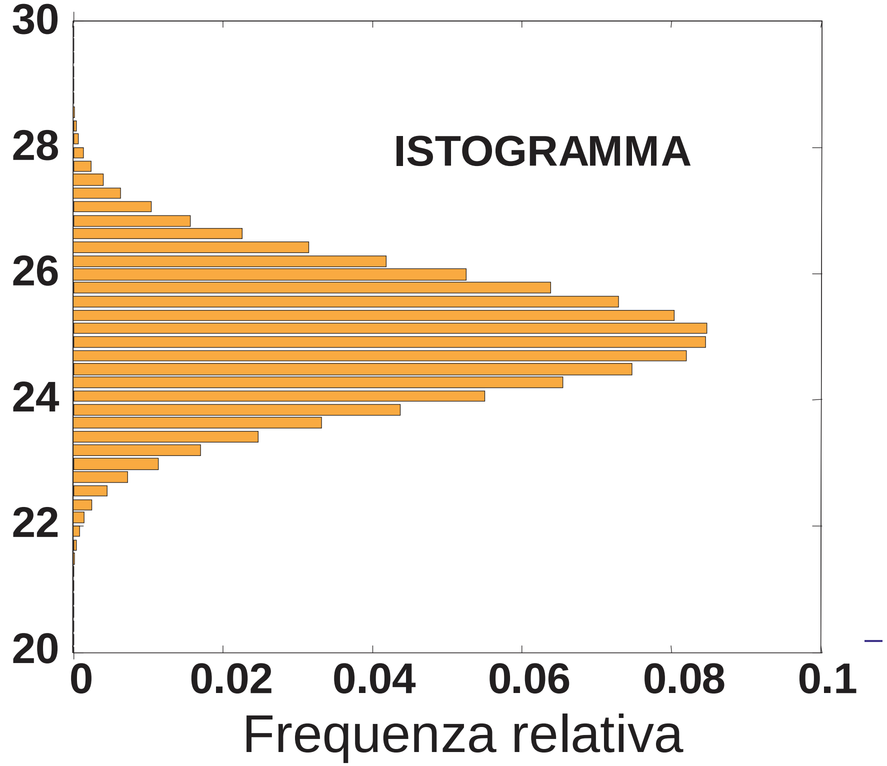
La variabile casuale è diventata discreta (ci sono 50 possibili risultati dell'esperimento) e possiamo approssimare la probabilità come limite della frequenza relativa per *N* elevato.

Istogramma

Anche per le variabili casuali continue, una volta “discretizzate”, è possibile tracciare l’istogramma come grafico della frequenza relativa dei risultati in ogni intervallino in cui si è suddiviso l’insieme continuo dei risultati.

ATTENZIONE: i valori dell’istogramma per le variabili casuali continue, una volta “discretizzate”, dipendono dalla dimensione dell’intervallino scelto: più è piccolo l’intervallo più sono bassi i valori dell’istogramma.





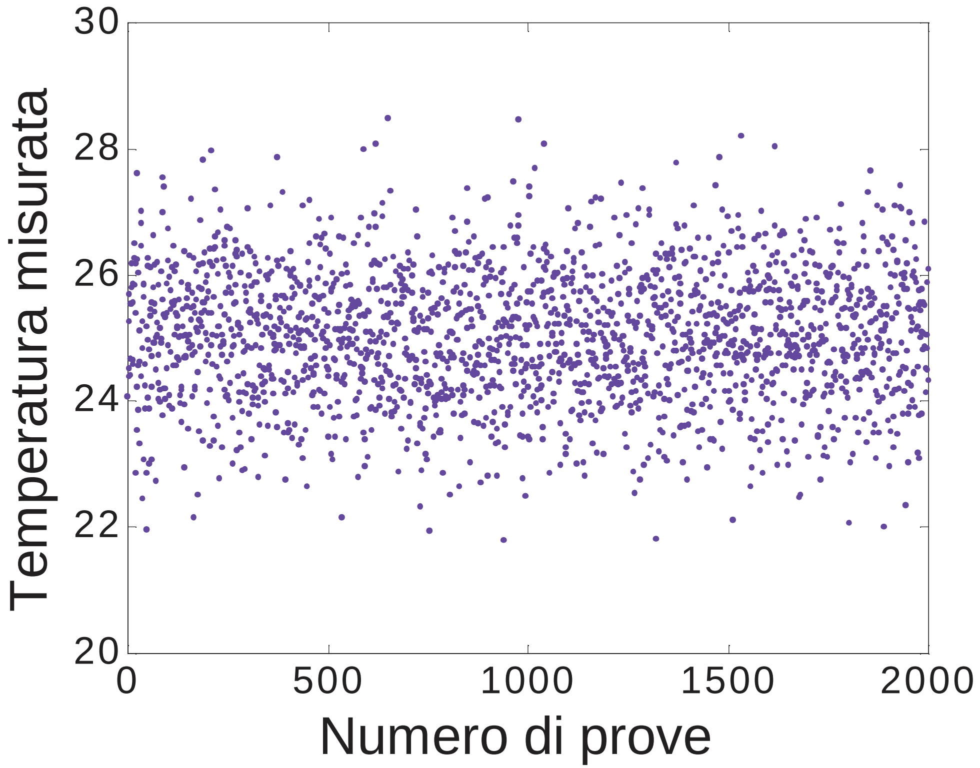
Densità di probabilità (ddp)

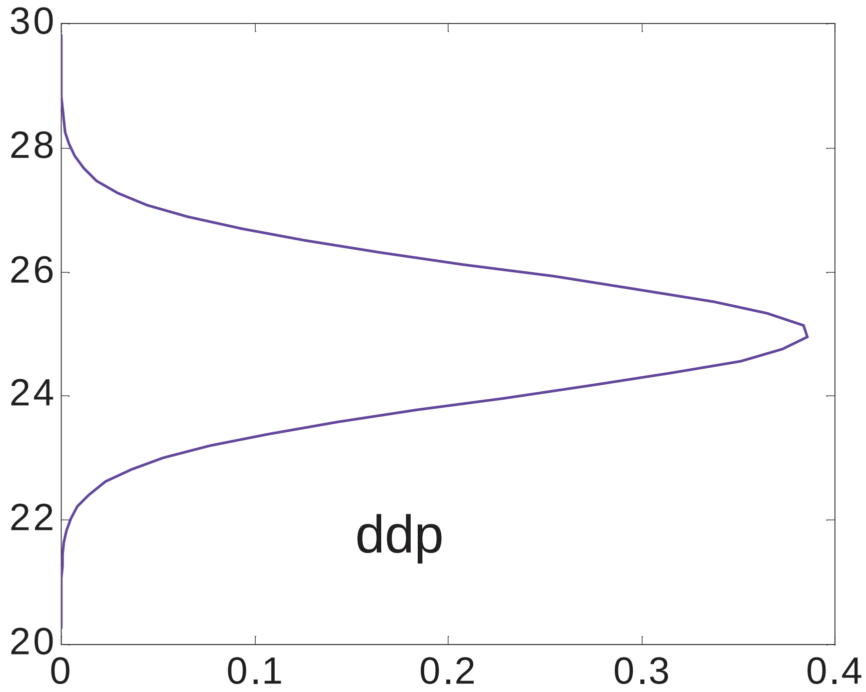
Per introdurre il concetto di densità di probabilità *p(a)* di una variabile casuale continua a partire dall'istogramma occorrono i seguenti passi:

1 - Utilizzare intervallini piccoli, così da poter ritenere la ddp costante al loro interno

2 - Dividere il valore dell'istogramma per la dimensione dell'intervallino (in modo che il risultato sia indipendente dalla dimensione dell'intervallino)

3 - Utilizzare un numero molto elevato di prove (tanto più elevato quanto più piccolo è l'intervallino) in modo che frequenze relative e probabilità quasi coincidano





Uso della densità di probabilità di una v.c. continua

La densità di probabilità $p(a)$ di una variabile casuale continua è dunque definibile come

$$p(a) = \lim_{da \to 0} \frac{P(a < x \leq a + da)}{da}$$

N.B.: se non è evidente di quale variabile casuale si sta parlando si scrive $p_x(a)$

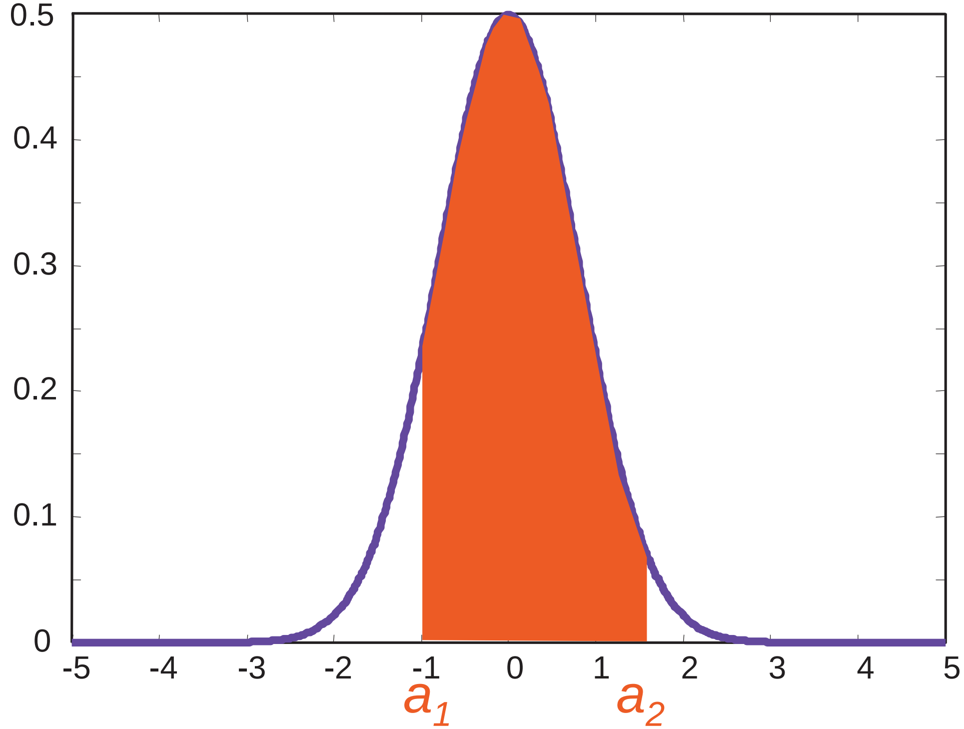
Dalla densità di probabilità $p(a)$ è facile calcolare la probabilità che la variabile casuale x assuma un valore compreso in un intervallo a_1, a_2. Basta sommare! si ottiene l'area sottesa dalla ddp nell'intervallo d'interesse.

$$P(a_1 < x \leq a_2) = \int_{a_1}^{a_2} p(a)\,da$$

Si noti che

$$P(-\infty < x < \infty) = \int_{-\infty}^{\infty} p(a)\,da = 1$$

Dunque l'area sottesa dalla ddp di una qualunque variabile casuale è unitaria.



Densità di probabilità congiunta

In modo del tutto analogo si definisce la ddp di due (o più) variabili casuali (densità di probabilità congiunta):

$$p_{xy}(a,b) = \lim_{\substack{da\to 0\\ db\to 0}} \frac{P(a < x \le a+da, b < y \le b+db)}{da\,db}$$

La densità di probabilità congiunta $p_{xy}(a,b)$ è utilizzata per calcolare la probabilità che le variabili casuali x e y assumano (congiuntamente) valori compresi in una regione del piano. Basta integrare nella regione d'interesse (integrale doppio).

Le variabili casuali x e y sono dette statisticamente indipendenti se

$$p_{xy}(a,b) = p_x(a)\,p_y(b) \qquad \text{per ogni } a \text{ e } b$$

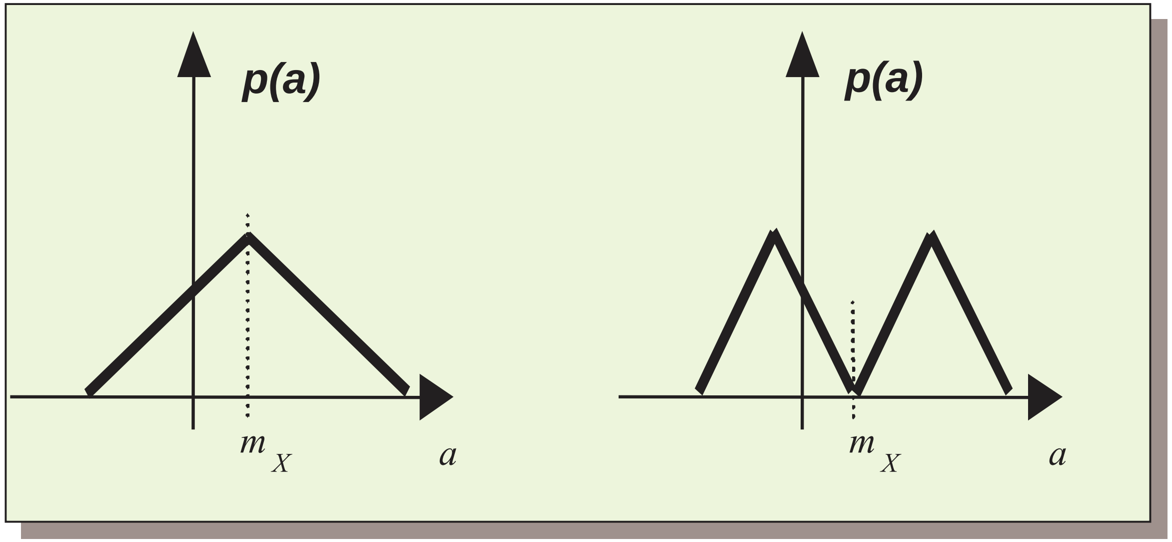
Si assume *a priori* che le variabili casuali x e y siano statisticamente indipendenti se ottenute da esperimenti svolti in condizioni indipendenti (esempio: prove ripetute).

Valor medio di una variabile casuale

Il valor medio m_x , detto anche valore atteso $E[x]$ o momento (statistico) di ordine uno, di una variabile casuale $\boldsymbol{x}$ è definito come segue. Se l'esperimento viene eseguito N volte (N grande) m_x è interpretabile approssimativamente come media aritmetica dei risultati:

$$m_x = E[x] = \int_{-\infty}^{\infty} a\, p(a)\, da \approx \frac{1}{N} \sum_{i=1}^{N} x_i$$

Il valor medio di una variabile casuale è l'ascissa del "baricentro" dell'area sottesa dalla densità di probabilità.



Proprietà del valor medio (1)

La proprietà fondamentale del valor medio è la seguente. Se dalla variabile casuale *x* si ottiene una nuova variabile casuale *y* attraverso la funzione *y=f(x)*, dove *f(x)* è una funzione prefissata (in tal caso si dice che *y* è funzione di variabile casuale), il calcolo del valor medio di *y* non richiede di determinarne la ddp (cosa che potrebbe essere difficile). Si può invece procedere nel seguente modo, mediante la ddp della variabile *x*:

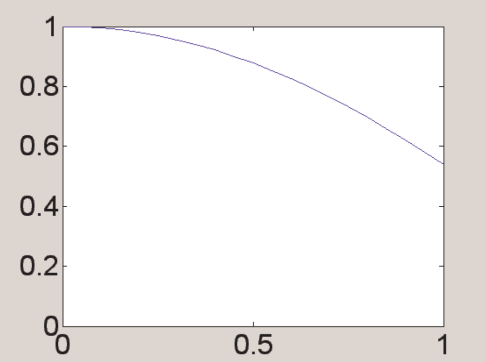
$$E[y] = \int_{-\infty}^{\infty} f(a)\, p_x(a)\, da$$

Analogo risultato vale per una funzione di più variabili casuali. La dimostrazione di questa importante proprietà non è affatto banale. Tuttavia il risultato non sorprende, se si pensa all'interpretazione del valor medio come media aritmetica di un gran numero *N* di risultati:

$$E[y] \approx \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

Esempio: la variabile casuale *x* ha ddp "uniforme" (cioè costante) nell'intervallo (0,1) e nulla altrove. La variabile casuale *y* è definita come *y=cos(x)*. Il valor medio di *y* è

$$E[y] = \int_{0}^{1} \cos(a)\, p_x(a)\, da = \int_{0}^{1} \cos(a)\, da = \sin(1) = 0.84$$



Proprietà del valor medio (2)

Dalla proprietà fondamentale del valor medio si ottengono immediatamente le seguenti proprietà, di uso frequentissimo:

Il valor medio della somma $x+y$ di variabili casuali è la somma dei valori medi.

Se a e b sono costanti $E[ax+b] = a\,E[x] + b$.

Se x e y sono variabili casuali indipendenti e $f(x)$ e $g(y)$ sono funzioni arbitrarie, $E[f(x)g(y)] = E[f(x)]\,E[g(y)]$.

In particolare, se x e y sono variabili casuali indipendenti si ha $E[xy] = m_x\, m_y$.

Variabili casuali x e y tali che sia $E[xy] = m_x\, m_y$ sono dette incorrelate.
N.B.: due variabili casuali possono essere incorrelate anche senza essere indipendenti.
Variabili casuali indipendenti sono invece sempre incorrelate.

Valore quadratico medio e varianza

Il valor quadratico medio $E[x^2]$, detto anche potenza statistica o momento (statistico) di ordine 2, di una variabile casuale ***x*** è definito come segue. Approssimativamente, è la media aritmetica di un numero molto elevato di risultati di altrettanti esperimenti:

$$E[x^2] = \int_{-\infty}^{\infty} a^2 \, p(a) da \approx \frac{1}{N} \sum_{i=1}^{N} x_i^{\,2}$$

La varianza σ_x^2 (detta anche momento centrale di ordine 2) di una variabile casuale ***x*** è il valore quadratico medio della differenza tra ***x*** e il suo valor medio m_x

$$\sigma_x^2 = E[(x - m_x)^2] = E[x^2] - m_x^{\,2} \approx \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \frac{1}{N} \sum x_i \right)^2$$

N.B.: dimostrare che

$$E[(x - m_x)^2] = E[x^2] - m_x^{\,2}$$

richiede un piccolo calcolo.

Deviazione standard

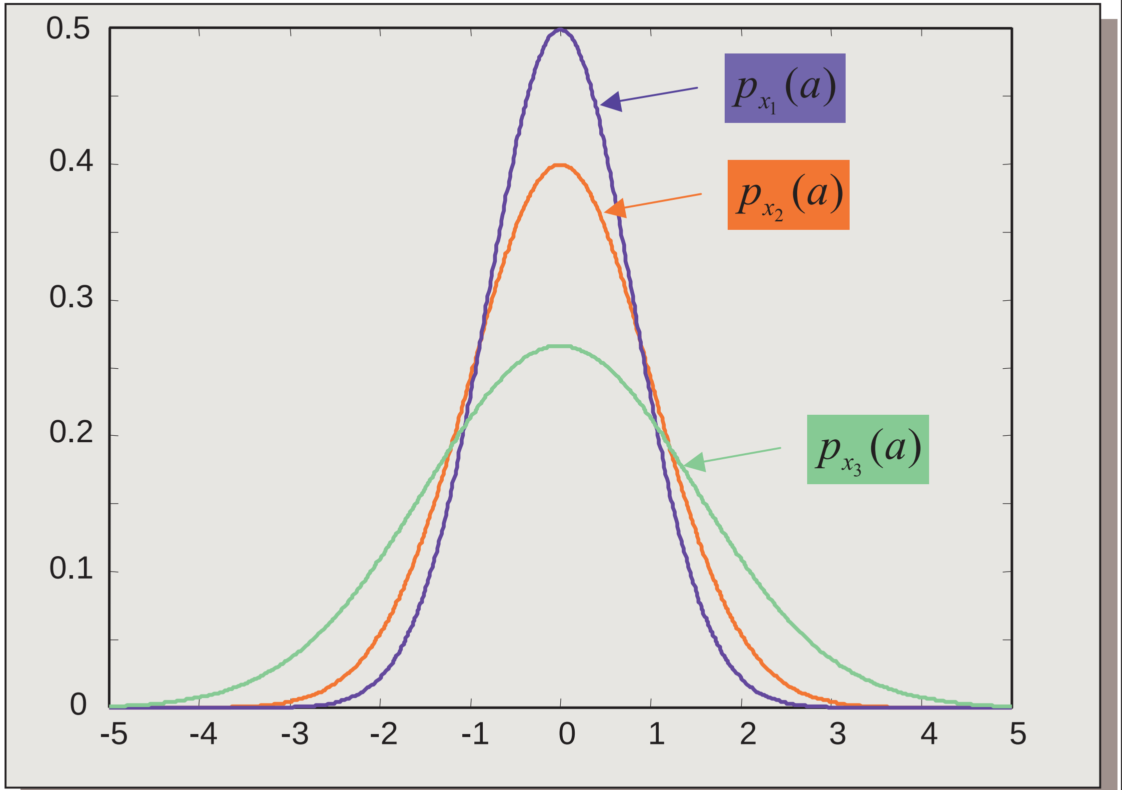
La radice quadrata della varianza è detta deviazione standard (o scarto quadratico medio) della variabile casuale ***x***

$$\sigma_x = \sqrt{\sigma_x^2}$$

La deviazione standard è una misura della dispersione, rispetto al valor medio, dei valori assunti nei vari esperimenti dalla variabile casuale ***x***.

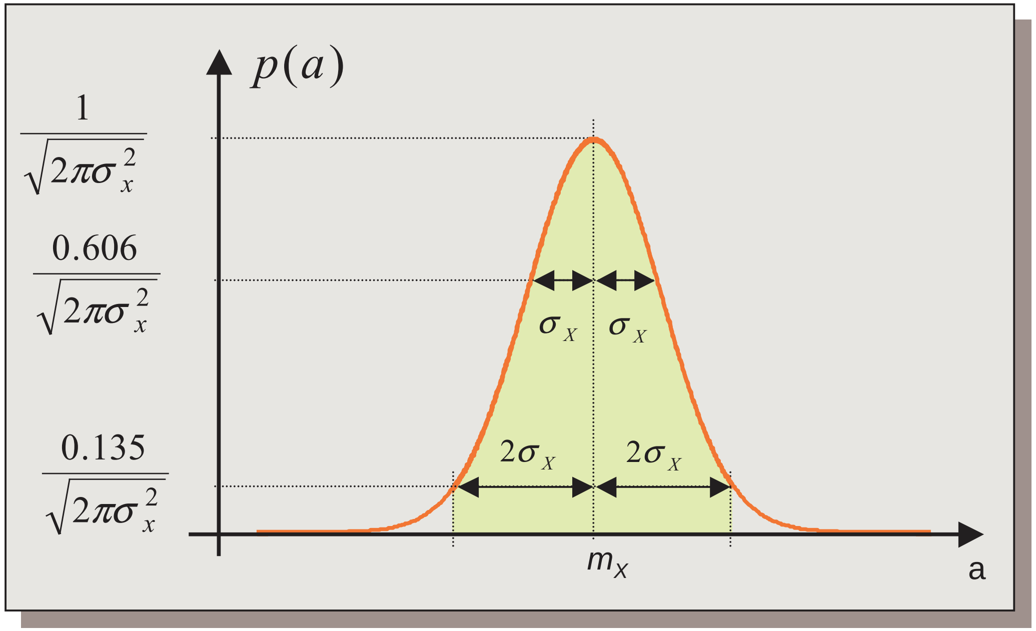
Più è elevata la deviazione standard più i risultati sono dispersi rispetto al valor medio e la ddp è “larga”.

$$\sigma_{x_3} > \sigma_{x_2} > \sigma_{x_1}$$



Densità di probabilità gaussiana

$$p(a)=\frac{1}{\sqrt{2\pi\sigma_x^2}}\exp\left(-\frac{(a-m_x)^2}{2\sigma_x^2}\right)$$

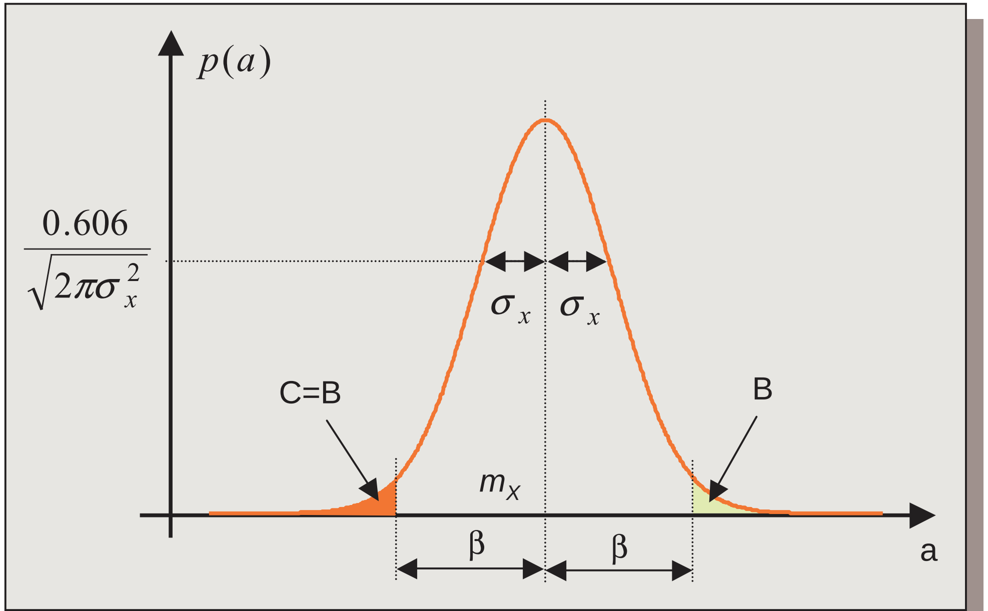


$$P(m_x-\sigma_x<x\le m_x+\sigma_x)=\int_{m_x-\sigma_x}^{m_x+\sigma_x}\frac{1}{\sqrt{2\pi\sigma_x^2}}\exp\left(-\frac{(a-m_x)^2}{2\sigma_x^2}\right)da\approx 0.683$$

$$P(m_x-2\sigma_x<x\le m_x+2\sigma_x)=\int_{m_x-2\sigma_x}^{m_x+2\sigma_x}\frac{1}{\sqrt{2\pi\sigma_x^2}}\exp\left(-\frac{(a-m_x)^2}{2\sigma_x^2}\right)da\approx 0.954$$

$$P(m_x-3\sigma_x<x\le m_x+3\sigma_x)=\int_{m_x-3\sigma_x}^{m_x+3\sigma_x}\frac{1}{\sqrt{2\pi\sigma_x^2}}\exp\left(-\frac{(a-m_x)^2}{2\sigma_x^2}\right)da\approx 0.997$$

Funzione Q e funzione errore complementare (*erfc*)



$$B = \int_{m_x+\beta}^{\infty} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(a-m_x)^2}{2\sigma_x^2}\right) da =$$

$$= Q\left(\frac{\beta}{\sigma_x}\right) = \frac{1}{2} erfc\left(\frac{\beta}{\sqrt{2}\sigma_x}\right)$$

t	Q(t)	t	Q(t)
0,00	5,000E-01	0,8	2,119E-01
0,05	4,801E-01	1,0	1,587E-01
0,10	4,602E-01	1,2	1,151E-01
0,15	4,404E-01	1,4	8,080E-02
0,20	4,207E-01	1,6	3,806E-01
0,25	4,013E-01	1,8	3,590E-02
0,30	3,821E-01	2,0	2,280E-02
0,35	3,622E-01	2,4	8,200E-03
0,40	3,446E-01	2,8	2,600E-03
0,45	3,264E-01	3,2	6,871E-04
0,50	3,085E-01	3,6	1,591E-04
0,60	2,743E-01	4,0	3,167E-05

s	erfc(s)	s	erfc(s)
0,0	1,000E+00	1,6	2,370E-02
0,1	8,875E-01	1,8	1,090E-02
0,2	7,730E-01	2,0	4,700E-03
0,3	6,714E-01	2,2	1,900E-03
0,4	5,716E-01	2,4	6,885E-04
0,5	4,795E-01	2,6	2,360E-04
0,6	3,961E-01	2,8	7,502E-05
0,7	3,222E-01	3,0	2,209E-05
0,8	2,579E-01	3,3	3,057E-06
1,0	1,573E-01	3,7	1,671E-07
1,2	9,700E-02	4,0	1,542E-08
1,4	4,770E-02	5,0	1,537E-12

$$\text{per } t > 3 \quad Q(t) \approx \frac{\exp(-t^2/2)}{\sqrt{2\pi}\, t}$$

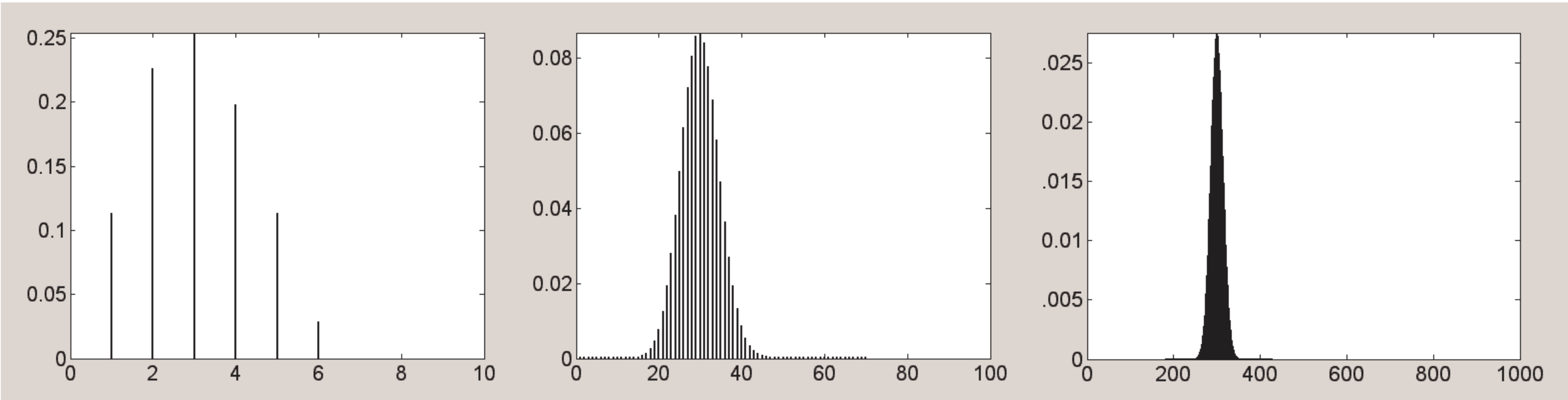
$$\text{per } s > 2 \quad erfc(s) \approx \frac{\exp(-s^2)}{\sqrt{\pi}\, s}$$

Prove ripetute (1)

Esempio: *N* lanci (indipendenti) di una moneta truccata, che dà testa con probabilità *p*. Consideriamo la variabile casuale *k* = numero di teste totali (non ci interessa l'ordine).

Si possono ottenere *k* teste in *N* prove in $\binom{N}{k}$ modi distinti, ciascuno avente probabilità $p^k(1-p)^{N-k}$ (prodotto delle probabilità), e quindi $P(k) = \binom{N}{k} p^k (1-p)^{N-k}$.

p=0.3 *N*=10 | *p*=0.3 *N*=100 | *p*=0.3 *N*=1000



E' evidente che all'aumentare di *N* la frequenza relativa si discosta sempre meno da *p* (**legge dei grandi numeri**)

Prove ripetute (2)

Varianza del numero k di "successi" in N prove indipendenti: se p è la probabilità di successo nella singola prova si può dimostrare che la varianza del numero di successi è

$$\sigma_k^2 = N\,p\,(1-p)$$

e quindi la varianza della frequenza relativa $f_k = k/N$ è $p\,(1-p)/N$ e tende a zero per N tendente a infinito.

Gli scarti quadratici medi sono dati rispettivamente da $\sqrt{N\,p\,(1-p)}$ e $\sqrt{p\,(1-p)/N}$.

Esempio: p = 0.3. $N =$ 10 100 1000

	$N = 10$	100	1000
$\sqrt{N\,p\,(1-p)} =$	1.45	4.58	14.5
$\sqrt{p\,(1-p)/N} =$	0.145	0.458	0.0145

Come si vede lo scarto quadratico medio del numero di successi aumenta (ma più lentamente di N), mentre lo scarto quadratico medio della frequenza relativa diminuisce.

Si comprende come sia possibile in pratica "misurare" una probabilità, eseguendo l'esperimento un numero sufficiente di volte (secondo la precisione desiderata).

Somma di variabili casuali

Si possono dimostrare molte notevoli proprietà:

Il valor medio della variabile casuale $z=x+y$ è pari alla somma dei valori medi.

Se x e y sono variabili casuali indipendenti, la variabile casuale $z=x+y$ ha come ddp la convoluzione delle due ddp:

$$p_z(a) = p_x(a) * p_y(a)$$

Se x e y sono variabili casuali indipendenti, la variabile casuale $z=x+y$ ha varianza pari alla somma delle varianze:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

La somma di un numero N grande di variabili casuali indipendenti x_i ha ddp prossima alla gaussiana, indipendentemente dalle singole densità! (teorema limite centrale)

$$p_y(a) = p_{x_1}(a) * p_{x_2}(a) * \ldots * p_{x_N}(a) \approx \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(a-m_y)^2}{2\sigma_y^2} \right)$$

La ddp può essere prossima alla gaussiana anche per N relativamente piccolo (N=5÷10).