Probabilità/2 – legge esponenziale

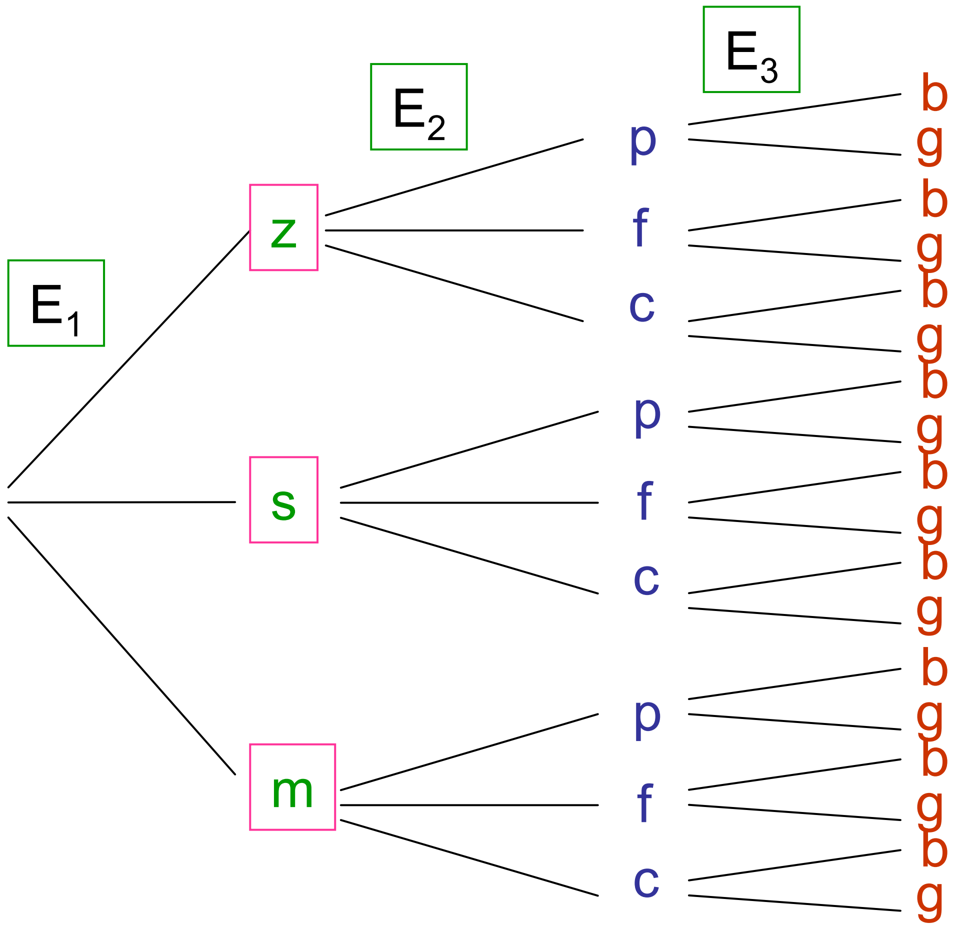
Statistica per Tossicologia dell'ambiente

AA2007/08

Eventi complessi – analisi combinatoria

- Se l'evento E1 si presenta in n1 modi, E2 in n2 etc., l'evento E1·E2 si presenterà in n1·n2 modi etc.
- Ad es. se E1 è il 1° piatto di un pranzo (n1=3, zuppa, succo d'arancia, melone), E2 è il 2° (n2=3, pesce, formaggio, carne), E3 è il 3° (n3=2, budino, gelato), avremo n1·n2·n3=3x3x2=18 possibilità (pranzi possibili): zpb, zpg, zfb, zfg ... che possono essere organizzati in un diagramma ad albero [vedi più avanti] e la probabilità di un dato pranzo sarà 1/18 – ovviamente se considero ordini diversi delle portate: zpb, zbp, pbz, pzb, bpz, bzp, in generale 123, 132, 231, 213, 312, 321, avrò in totale 18x6=108 possibilità che a 6 a 6 hanno gli stessi contenuti con solo l'ordine diverso

Diagramma ad albero



nell'ordine
1° 2° 3°

P(zpb) = P(zpg)
=⅓ ·⅓ ·½=1/18
&tc.

1° piatto (3) x 2° piatto (3) x 3° piatto (2) = 18 pranzi possibili

Disposizioni e permutazioni

- i) ad es. 3 biglie Rossa, Blu, Verde (m) prese a 2 per volta: RR, RB, BR, RV ... 9 possibilità (**lo stesso elemento può essere preso anche 2 volte**). In genere le **disposizioni** di m elementi presi a n per volta sono

  $D(m,n) = m^n$

- ii) ad es. 3 biglie R, B, V (m) e 3 posizioni (n)

  posizione 1: 3 possibilità R,B,V (scelta)

  “ 2: 2 “ R, V

  “ 3: 1 “ V

  => 3x2x1= 3! possibilità – in genere si ha (**ogni elemento max una volta**) per il n.o di **permutazioni**

  $P(m,n) = m\cdot(m-1)\cdot(m-2)\cdots(m-n+1) = m!/(m-n)!$

- 6 biglie (R, B, V, Ciano, Magenta, Giallo) e 3 posizioni

  $P(6,3) = 6x(6-1)x(6-2) = 120$

  $[= 6!/(6-3)! = 720/6 =120]$

Note

- Diagrammi ad albero – calcolo di probabilità (conteggio del numero di possibilità) di eventi complessi: in generale ogni ramo può essere pesato per una probabilità
- **Fattoriale** - definizione

  $n! = n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1$ con $1!=1$ e $0!=1$

  $(n+1)! = (n+1)\cdot n!$
- m oggetti selezionati n alla volta: la posizione 1 può essere riempita in m modi, la 2 in (m-1), la 3 in (m-2) e così via fino alla n in (m-n+1). La formula risultante può essere manipolata moltiplicando per *$1 = (m-n)!/(m-n)!$*

  => $\mathbf{m\cdot(m-1)\cdots(m-n+1)}\cdot[(m-n)\cdot(m-n-1)\cdots 3\cdot 2\cdot 1]/[(m-n)\cdot(m-n+1)\cdots 3\cdot 2\cdot 1] = m!/(m-n)!$

Combinazioni

- Voglio contare le possibilità indipendentemente dall'ordine – se non conta l'ordine RBV, RVB, BRV &tc. sono la stessa cosa/configurazione – ci sono n! modi di permutare le n posizioni quindi il n.o di possibilità o di **combinazioni** è ridotto rispetto alle permutazioni

$C(m,n) = P(m,n)/n! = m!/[n!(m-n)!]$

- Se $n = m$

$C(m,m) = m!/[m!(m-m)!] = m!/m! = 1$

- Se $n = 1$

$C(m,1) = m!/(m-1)!] = m$

- es.: $C(6,3) = P(6,3)/3! = 6!/(3!\cdot 3!) = 720/36 = 20$

La distribuzione binomiale

- Prendiamo 5 palline in una scatola (3R, 2B) ed estriamo 3 volte con reinserimento dopo ogni estrazione – ad es. P(RBB) = P(R)·P(B/R)·P(B/RB) = 3/5·2/5·2/5 = 12/125 - questa volta si ha P(R) = **p** = 3/5 = 0.6 e P(B) = **q = (1-p)** = 2/5 = 0.4
- P(0R), P(1R), P(2R), P(3R) = 8, 36, 54, 27/125
- **Distribuzione binomiale** – discreta

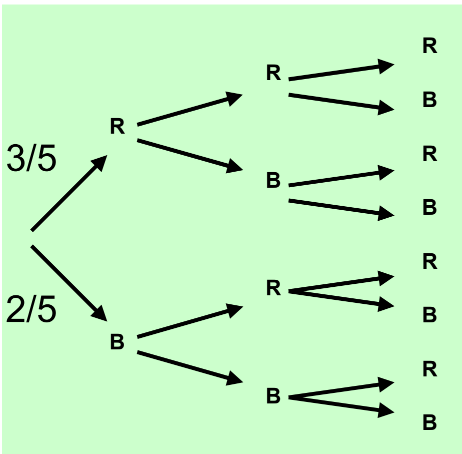
$$\mathbf{P(x)} = C(n,x)p^x q^{n-x} \mathbf{= n!/[x!(n-x)!]p^x q^{n-x}}$$

con n numero di tentativi e

$$\Sigma^n_{x=0} P(x) = 1$$

La distribuzione binomiale/2

diagramma ad albero



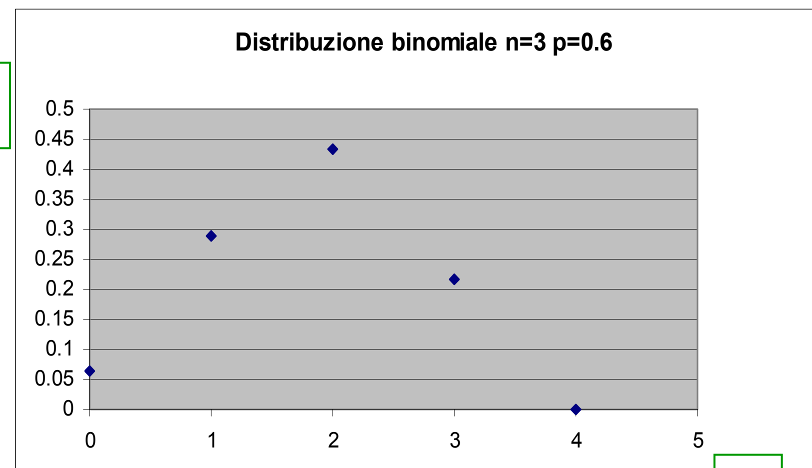
$3/5 \cdot 3/5 \cdot 3/5 = 27/125$

rami non equiprobabili

Risultato	Probabilità
RRR	27/125
RRB	18/125
RBR	18/125
RBB	12/125
BRR	18/125
BRB	12/125
BBR	12/125
BBB	8/125

1a 2a 3a
estrazione

P(x)



x

$\Sigma^n_{x=0} P(x) = 1$

x – numero di R
p – prob. di R
n – num. di prove

Proprietà della distribuzione binomiale

- Media

  $\mu = E(x) = \Sigma^n{}_{x=0} xP(x) = \Sigma^n{}_{x=1} x \cdot n!/[x!(n-x)!]p^x(1-p)^{n-x} =$

  $= np\{\Sigma^n{}_{x=1}(n-1)!/[(x-1)!(n-1-x+1)!]p^{x-1}(1-p)^{n-1-x+1}\} = np$

- Varianza

  $\sigma^2 = \Sigma^n{}_{x=0}(x-\mu)^2 \cdot n!/[x!(n-x)!]p^x(1-p)^{n-x} = \Sigma^n{}_{x=0} x^2P(x) - \mu^2 =$

  $= npq = np(1-p)$

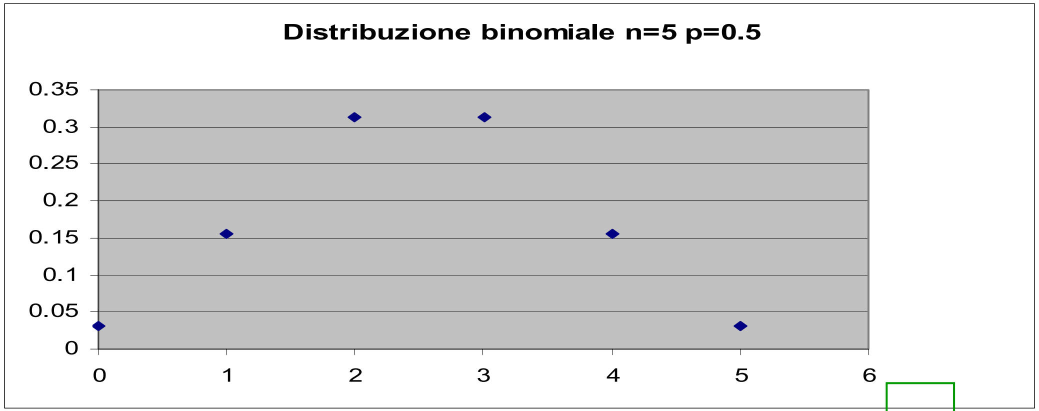
- Deviazione standard

  $\sigma = \sqrt{npq} = \sqrt{np(1-p)}$

- [Nota: per dimostrare la formula per μ basta porre $y=x-1$ e $k=n-1$ => la { } = $\Sigma^k{}_{y=0} k!/[y!(k-y)!]p^y(1-p)^{k-y} = 1$, in modo analogo si ottiene la formula per σ^2]

Binomiale/esempio – lancio di monete: risultato T o C con p uguale a ½

P(x)



x = n. T (oppure C)

$\mu = np = 2.5$

P(0T,5C) = $(1/2)^5$ = 1/32 = 0.03125
P(1T,4C) = $(1/2)^5 \cdot 5$ = 5/32
&tc.

$\sigma^2 = npq = 1.25$
$\sigma = 1.12$

Sarebbe sensato calcolare Binomiale(1.5T,3.5C)? **NO**

Relazione fra le principali distribuzioni di probabilità

Binomiale

Discreta
p fissa, n tentativi es. lancio di monete, dadi, nascita di una figlia femmina &tc.

p finita
n -> ∞

Gauss o normale

Continua
fenomeni aleatori continui (cause multiple e indipendenti) es. errori di misura casuali, h di un campione di donne &tc.

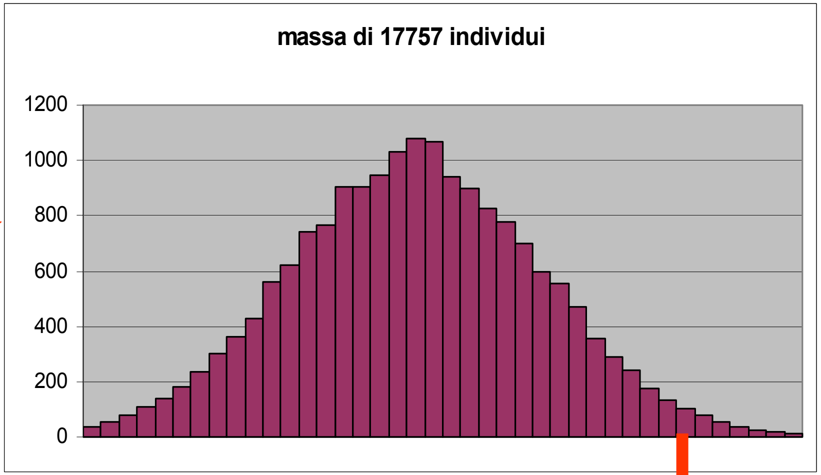
p -> 0
n -> ∞
np finito

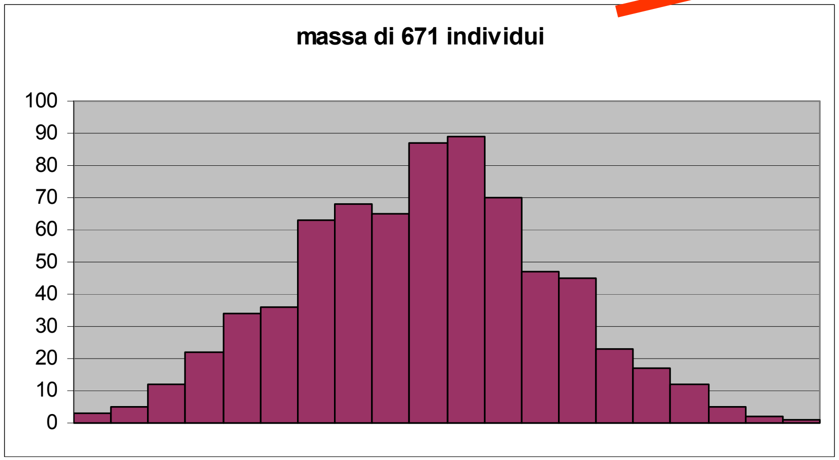
Poisson

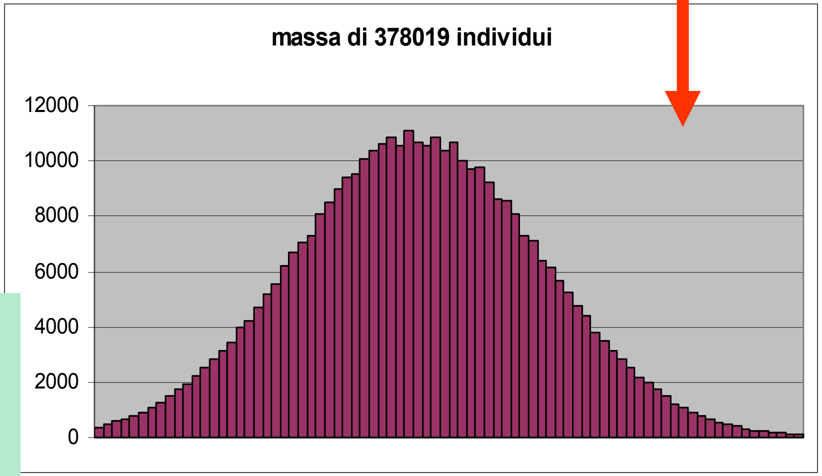
Discreta
conteggi di eventi rari (radiazioni, file di attesa, n.o di bambini nati in un ospedale in 3 gg &tc.)

Distribuzione normale/preliminare

Distribuzione discreta della massa di n individui



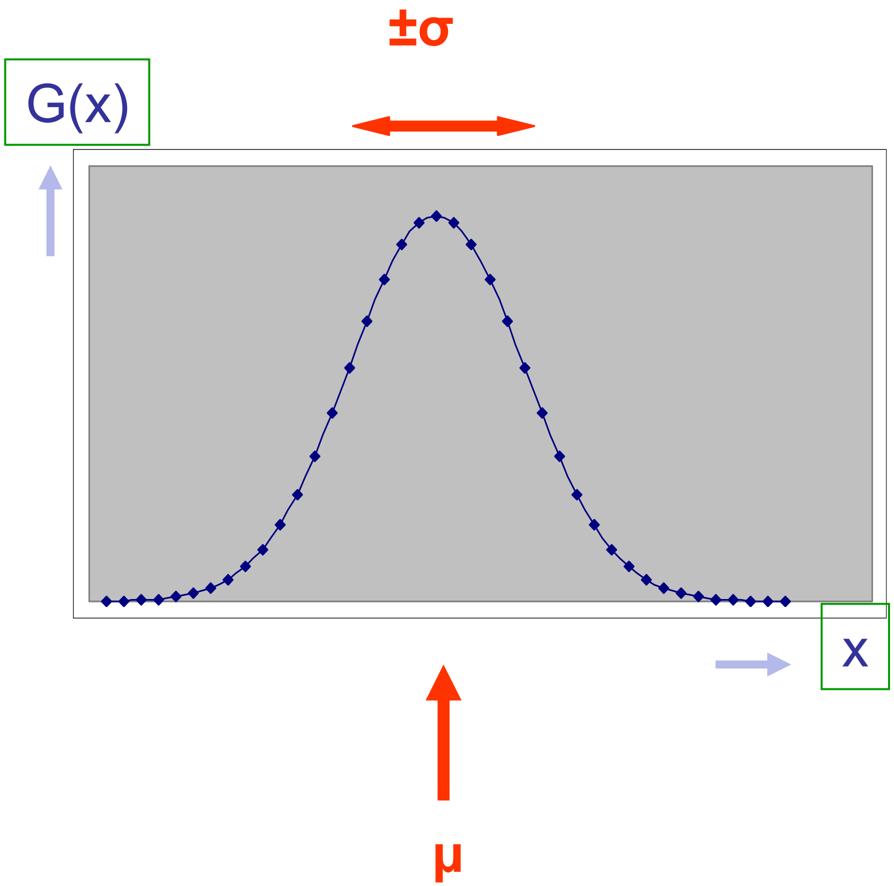




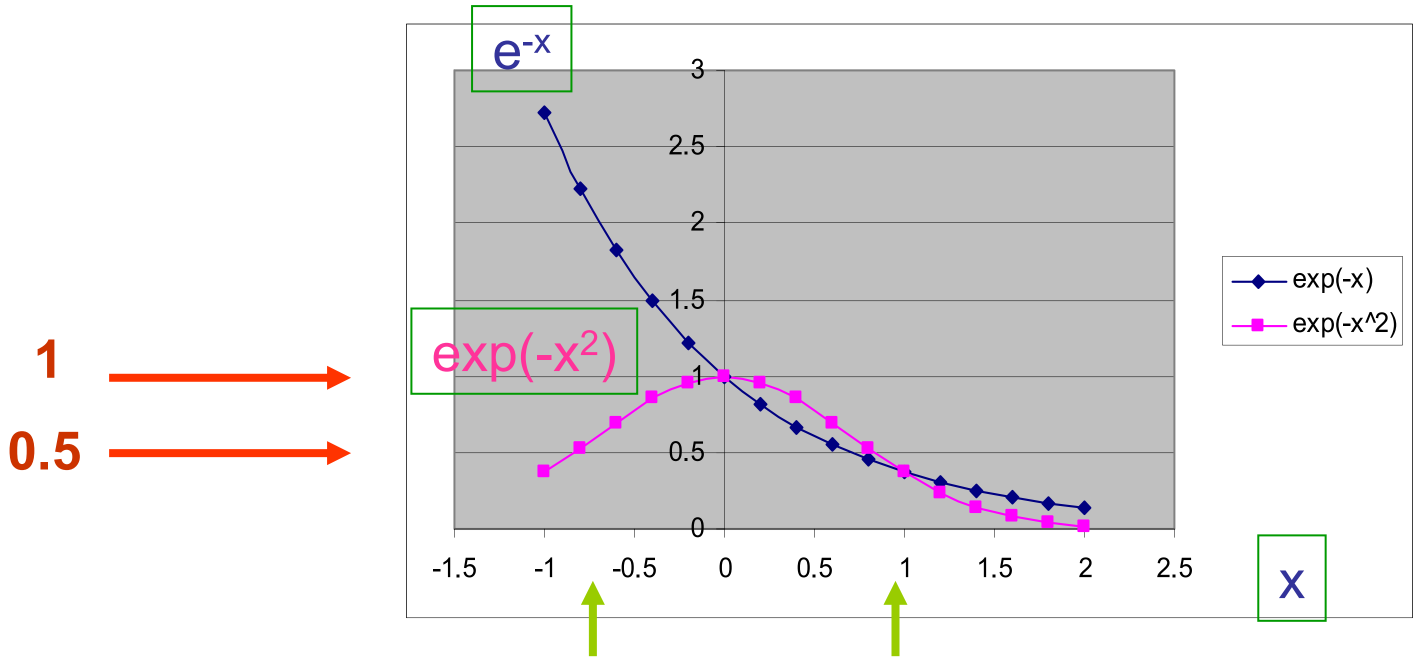
crescendo la dimensione del Campione
=> si può diminuire la larghezza dei canali,
=> variabile aleatoria continua

Distribuzione normale o di Gauss

- **$P(x)$ o $G(x) = 1/(\sqrt{(2\pi)}\,\sigma) \cdot$**
  $\cdot\ \exp[-(x-\mu)^2/2\sigma^2]$
- variabile normale standardizzata
  $z = (x-\mu)/\sigma$
- **$G(z) = 1/\sqrt{(2\pi)}\ \exp[-z^2/2]$**
- media = $E(x) = \mu$
  $E(z) = 0$
- varianza = $E((x-\mu)^2) = \sigma^2$
  $E(z^2) = 1$
- dev. stand. = σ
  $E(\sqrt{z^2}) = 1$

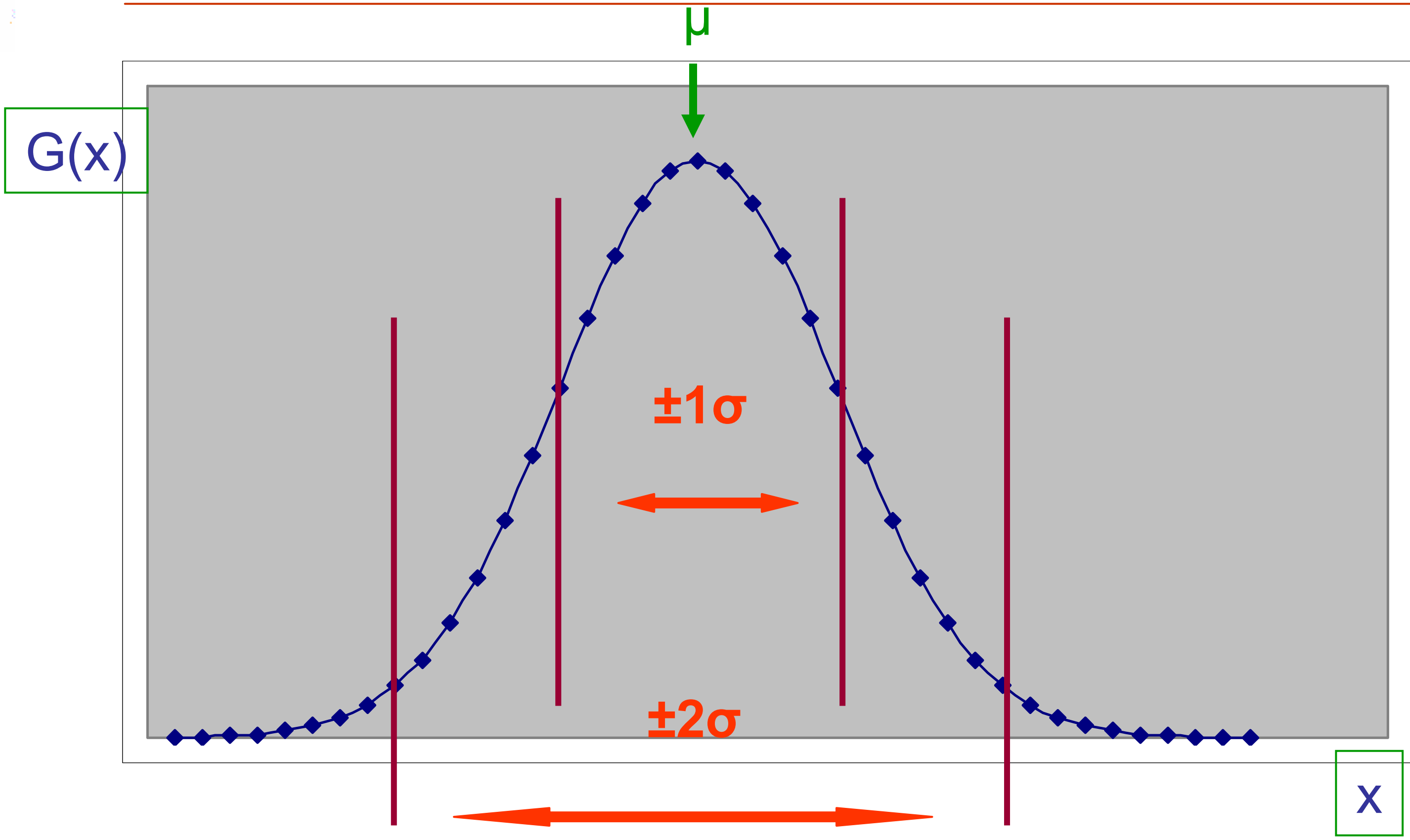


$exp(-x^2)$, FWHM etc.



FullWidthHalfMaximum = 2.36 σ corrisponde alla differenza fra le ascisse per cui $G(x_{50}) = 0.5\, G(\mu)$ – in figura è riportata $exp(-x^2)$ cioè $\mu=0$, $\sigma=1/\sqrt{2}$ e HalfWHM = 1.18σ = 0.834

Distribuzione normale/2



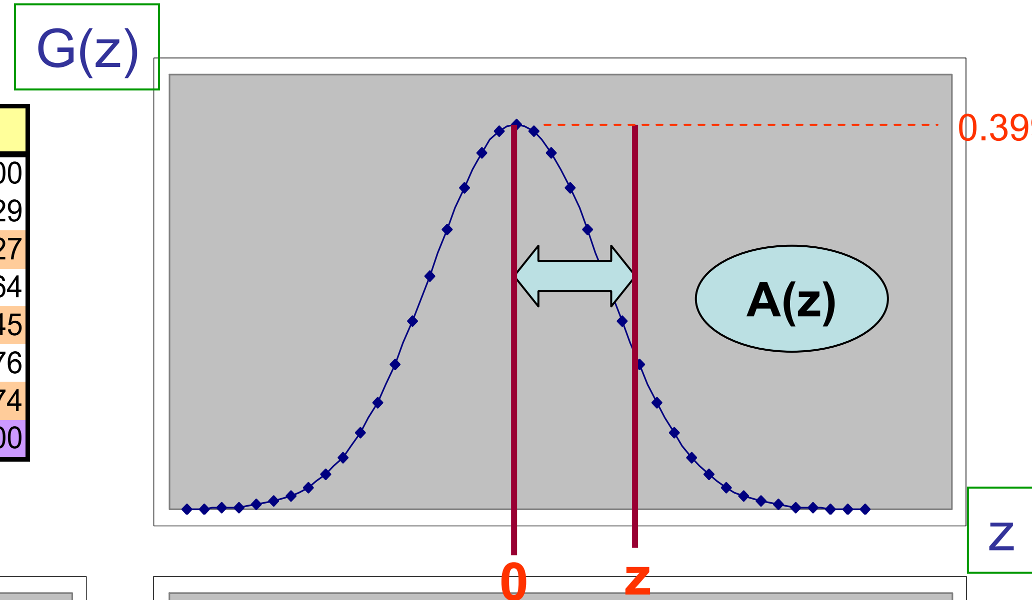
Fra μ-σ e μ+σ	è compreso il	68.27%	dell'area	sotto la	gaussiana
“ μ-2σ e μ+2σ	“	95.45%	“	“	“
“ μ-3σ e μ+3σ	“	99.74%	“	“	“

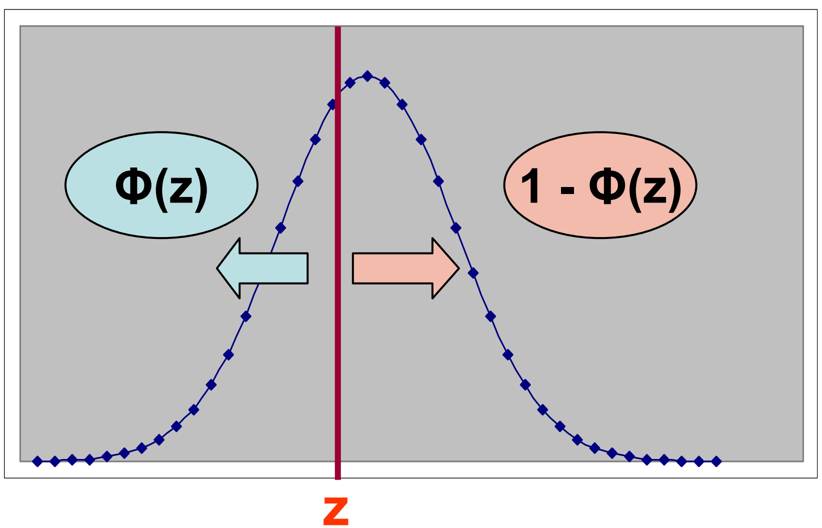
Integrali della distribuzione normale e probabilità

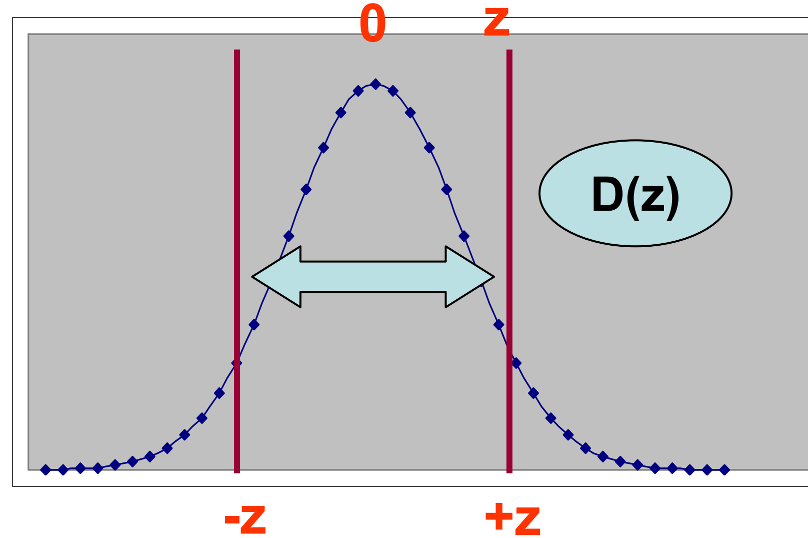
- Tavole di gaussiana e integrali nella forma **$G(z) = 1/\sqrt{2\pi} \exp[-z^2/2]$**
- $\Phi(z) = \int_{-\infty}^{z} G(z)dz$
- $D(z) = \int_{-z}^{+z} G(z)dz$
- $A(z) = D(z)/2$ con $A(-z) = A(z)$
- $\Phi(z) = \Phi(0) + z/|z| \cdot A(z)$
- **$p(-a<z<a) = D(a) = 2A(a)$**
- **$p(a<z<b) = \Phi(b) - \Phi(a)$**

Integrali della distribuzione normale/2 & probabilità

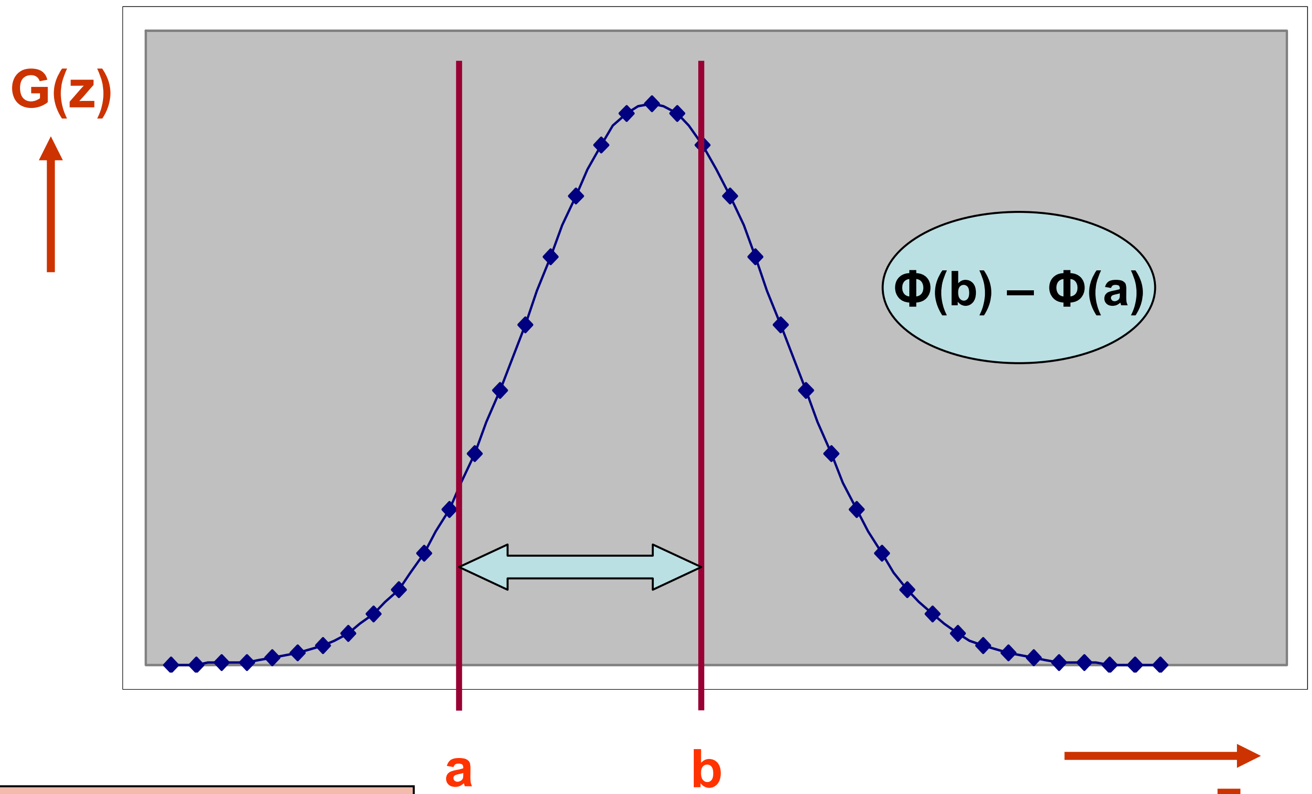
z	G(z)	Φ(z)	D(z)
0	0.3989	0.5000	0.0000
0.5	0.3521	0.6915	0.3829
1	0.2420	0.8413	0.6827
1.5	0.1295	0.9332	0.8664
2	0.0540	0.9772	0.9545
2.5	0.0175	0.9938	0.9876
3	0.0044	0.9987	0.9974
1.645	0.1031	0.9500	0.9000





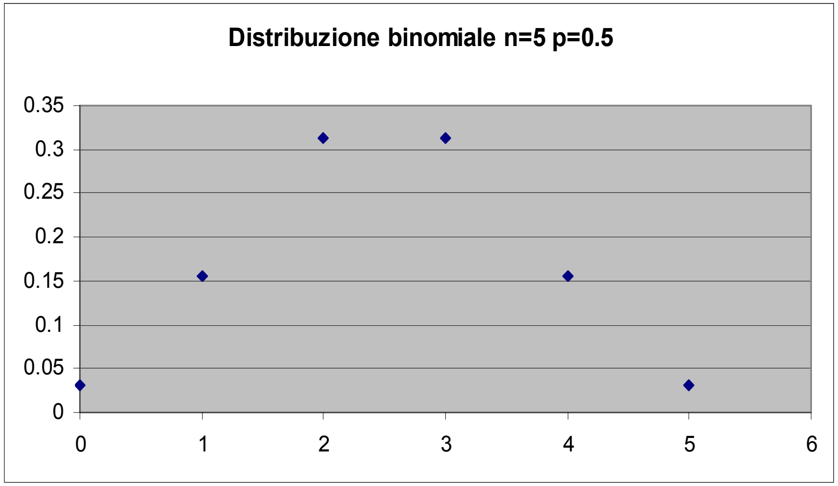


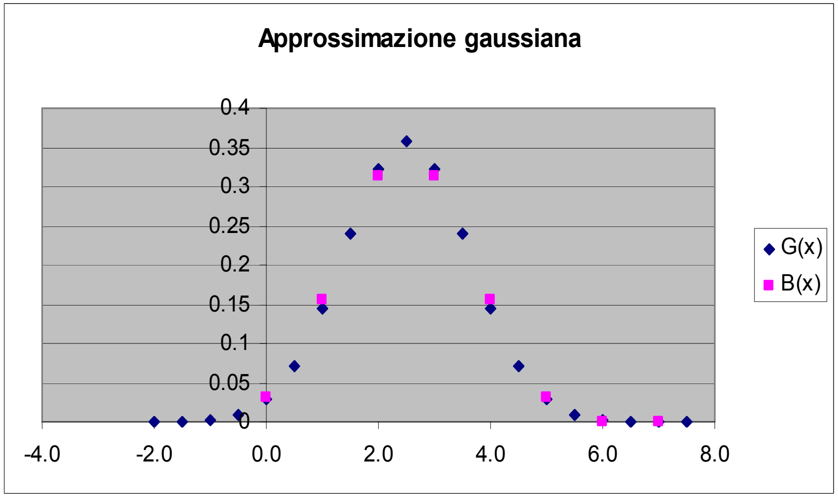
Integrali della distribuzione normale/3 & probabilità



p(a<z<b) = Φ(b) – Φ(a)

Approssimazione gaussiana della binomiale





p = 0.5, n = 5 => μ = np, σ = √np(1-p)

In genere una gaussiana è più facile da calcolare di una binomiale: l'approssimazione per p ~ 0.5 è buona anche con n piccolo, se p è vicino a 0 (o a 1) n deve essere grande per una buona approssimazione

Esercizio sulla binomiale

- Elezioni politiche in Australia, exit poll,10h, sabato 24/11/07.
- Le elezioni riguardavano 13.6x10^6 votanti. Un campionamento di 2700 votanti in 31 circoscrizioni all'uscita dei seggi, ha dato Rudd al 53% e Howard al 47%.
- Se il campione era rappresentativo, senza bias, era giusto il titolo del Sidney Morning Herald: "Howard ha bisogno di un miracolo (per vincere le elezioni)"?

Soluzione

- Usiamo la distribuzione binomiale. Due possibilità, R e H; $n = 2700$, $p = 0.53$, $q = 0.47$,

  $R+H = n$; $R = np = 1431$; $H = nq = 1269$

  $R-H = R-n+R = 2R-n = 162$
- la dev. stand. di R-H è (n è fisso, costante, R e H non sono indipendenti)

  $\sigma(R-H) = \sigma(2R) = 2\sigma(R) = 2\sqrt{(npq)} = 51.87$
- usiamo ora l'approssimazione gaussiana: R-H differisce da 0 per $162/51.87 = 3.12\sigma$ ossia la probabilità che, in un altro campionamento, possa essere $R-H \leq 0$ è $p=8.9 \times 10^{-4}$ (conta solo una coda della gaussiana) $\rightarrow$ la risposta è sì
- calcolando la coda della distribuzione di H centrata ad $nq=1269$ con dev. stand. $\sqrt{(npq)}=25.93$ oltre $n/2=1350$ si ottiene lo stesso risultato
- meglio, v. prossima lez.: se assumo $H \geq 1350$ (H vince o pareggia), la prob. di osservare nel campionamento $nq \leq 1269$ risulta inferiore a $p=9.1 \times 10^{-4}$

Distribuzione di Poisson

- **Discreta** - modello utile per distribuzioni in cui la variabile è il n.o di volte che si presenta un evento in un intervallo di tempo o di spazio. Es.: n.o di clienti che entrano in un negozio in 1 h, n.o di reclami ricevuti in 1 g da una compagnia di assicurazioni, n.o di lombrichi in una data area di terreno, n.o di batteri per ml di un certo liquido &tc.
- Dalla binomiale con $n \to \infty$, $p \to 0$: $np \to \mu$ costante ($q \to 1$)
- $\mathbf{P(x) = e^{-\mu}\mu^x/x!}$ $x = 0,1,2 \ldots\infty$ (intero +vo)
- **media** $E(x) = \boldsymbol{\mu}$
- **varianza** $E((x-\mu)^2) = \boldsymbol{\sigma^2 = \mu}$ dev. stand. $\sigma = \sqrt{\mu}$

  (si possono derivare dalla poissoniana ricordando che $\boldsymbol{\Sigma_0^\infty P(x) = 1}$)
- Conteggio di eventi – processo aleatorio con probabilità costante
- => errore statistico nei conteggi di N eventi

  $\sigma = \sqrt{N}$

  $\sigma/N = 1/\sqrt{N}$ $\quad\to 0$ per $N \to \infty$
- Variabile standardizzata $z = (x-\mu)/\sqrt{\mu}$

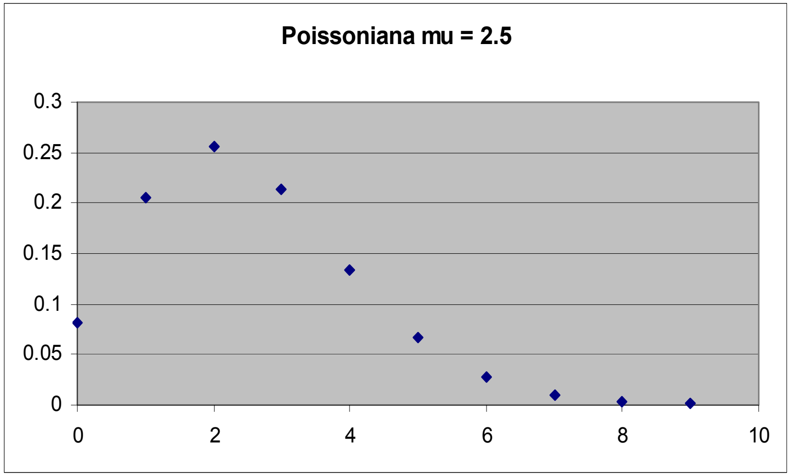
Distribuzione di Poisson/2

- Altri es. di variabili poissoniane
  - emissioni di un preparato o campione radioattivo in un dato intervallo di tempo
  - radiazioni assorbite in un dato spessore di materia
  - batteri o cellule/mm^2 su un vetrino o coltura &tc.
- Approx. gaussiana con μ piccola: ad es.

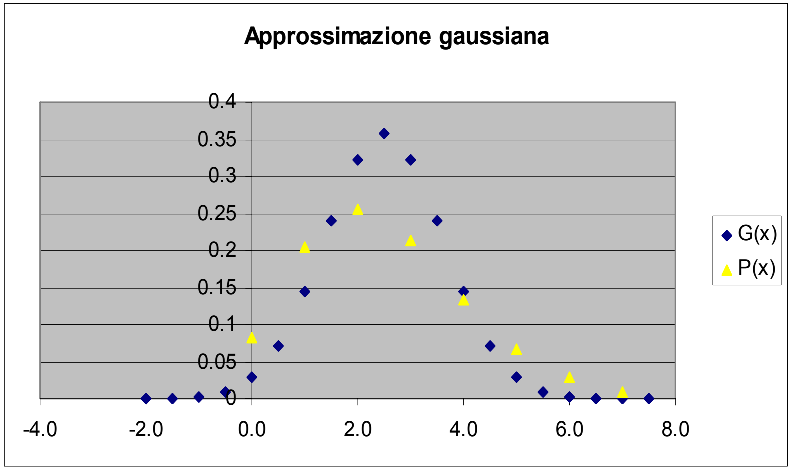
  $P(x) = e^{-2.5}2.5^x/x!$

  x può assumere tutti i valori interi >= 0

  (=> con μ = 2.5 l'approx gauss. è così così, grossolana)



x



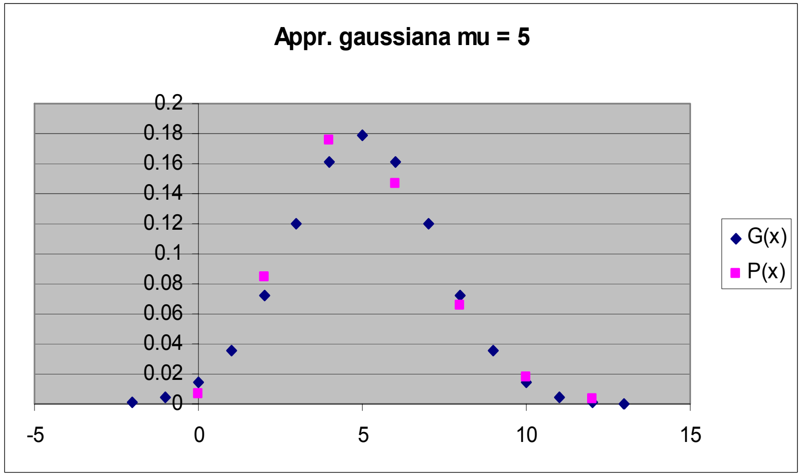
x

Approssimazione gaussiana della distribuzione di Poisson

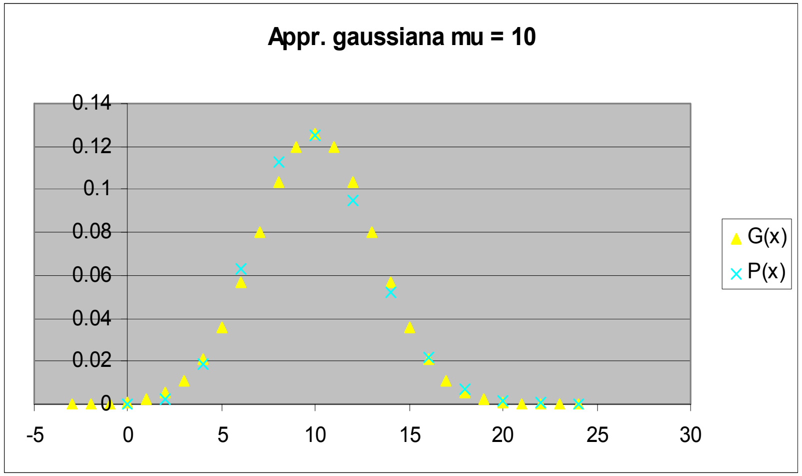
- Prendiamo invece μ = 5 o ancora meglio μ = 10: ora le approssimazioni gaussiane sono decisamente migliori

  (e più semplici da maneggiare, per es.

  se voglio integrare / calcolare una probabilità per μ grande)



x



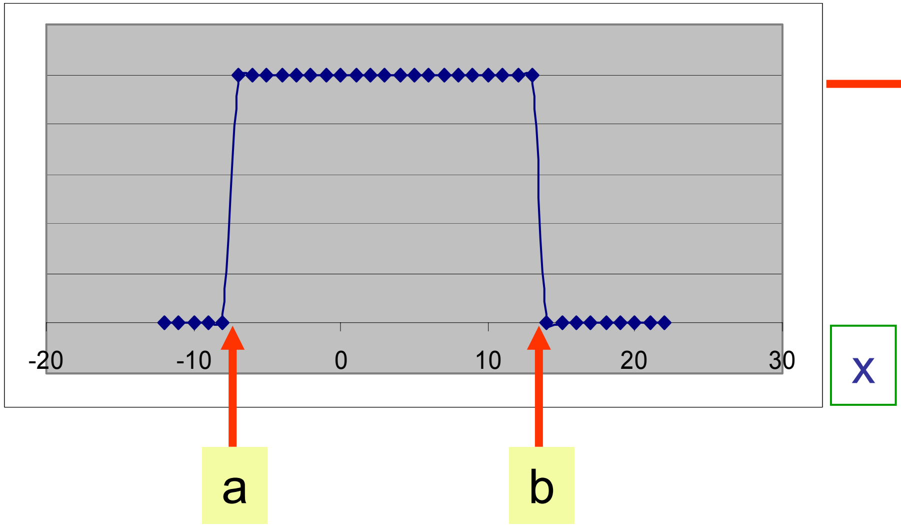
x

Sarebbe sensato calcolare Poissoniana(0.39)? NO

Distribuzione uniforme

- **P(x) = 1/(b-a)** per a<x<b **continua**

  = 0 per x<a oppure x>b
- $\int_a^b$ **P(x)dx = 1/(b-a)** $\int_a^b$ **dx = 1**
- R(x) uniforme fra 0 e 1: RAND() in Excel, RND sulla calcolatrice (numeri di 3 cifre 0.xyz) etc.
- es. liste di randomizzazione, si associa R(x) al nome del paziente (se R<0.5 → placebo, se R≥0.5 → farmaco)

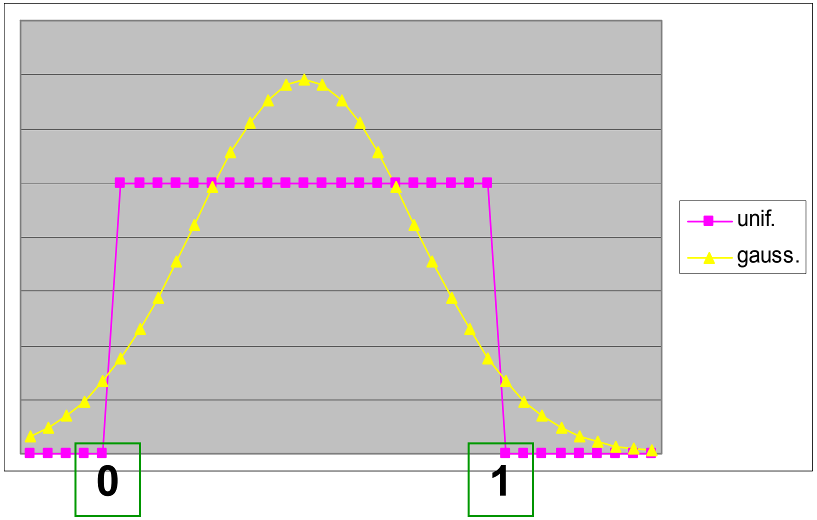


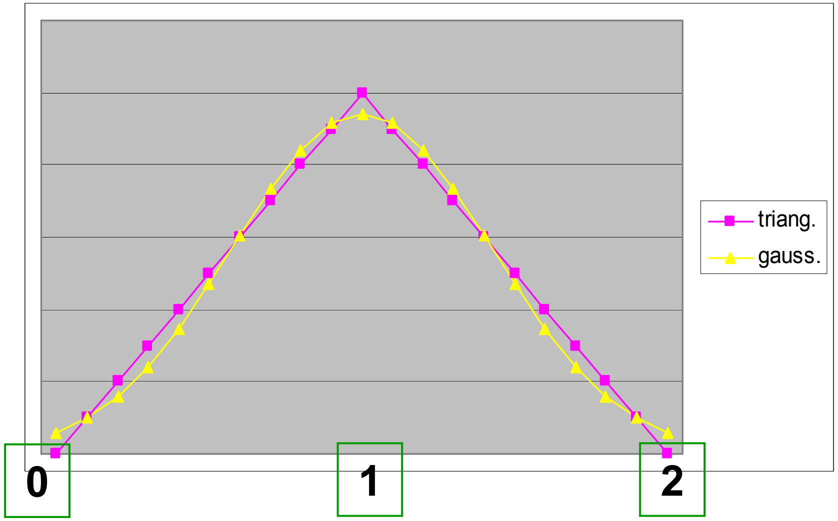
P(x) = 1/(b-a)

(Una **distribuzione uniforme discreta** è ad es. data dalla prob. di ottenere un numero compreso fra 1 e 6 quando si lancia un dado, p =1/6 cost.)

Distribuzione uniforme/2 **- facoltativa**

- μ = (a+b)/2
- σ = (b-a)/√12
- l'approssimazione gaussiana è scarsa
- se però sommo due distrib. unif. per es R1 e R2, distrib. fra 0 e 1, ottengo una distr. a triangolo fra 0 e 2, che è già molto vicina a una gaussiana con μ = 1 e σ = 0.424 &tc. (al limite sommando R1+R2+ ...+RN+ ... si ottiene proprio una gaussiana!)
- (una distr. discreta triangolare è quella dei numeri compresi fra 2 e 12 che si ottengono col lancio simultaneo di due dadi: la somma 7 [P=6/36] è 6 volte più prob. della somma 2 o 12 [P=1/36] etc.)





Legge esponenziale (decadimento, assorbimento) **- facoltativa**

- Fenomeni aleatori di tipo poissoniano: assorbimento della luce, o di radiazioni nella materia [F(x)], decadimento radioattivo [F(t)], metabolismo di un farmaco [F(t)]

  => **legge esponenziale**

- Probabilità - che sia avvenuto un certo numero di “eventi” (discreto) in un certo “intervallo” [Δx,Δt] - se P è costante si rientra nella statistica di Poisson
- Probabilità [ΔN/N] ~ intervallo (piccolo); tipo di processo (una costante che denota processi + o - probabili: “fisica”, “chimica”, “fisiologia”)

Legge esponenziale/2 **- facoltativa**

- Ad es. decadimento (t) - probabilità ~ dN/N

  t=0 N(0) = N_0 iniziali (atomi*, molecole*, nuclei)

  dt intervallo (di tempo)

  N(t) a t generico (sopravvissuti)

  dN variazione di N(t) dovuta al passare da t a t+dt

  dN = N(t+dt)-N(t)

- dN ~ dt; N(t); λ(costante di tempo, fisica del fenomeno);

  - (segno meno, decrescita)

----

* per atomi, molecole eccitati il decadimento consisterà nell'emissione di luce, UV, IR, RX – per i nuclei instabili si tratterà di radiazioni $\alpha, \beta^{\pm}, \gamma$

Legge esponenziale/3 **- facoltativa**

- Complessivamente

  $dN = -\lambda N(t)dt$ [t variab. indip.; N dipendente]

  separando le variabili

  $dN/N = -\lambda dt$ => $d(\ln N) = -\lambda d(t)$

  integro $\int_0^t$ => $\ln N(t) - \ln N(0) = -\lambda(t-0)$

  $\ln[N(t)/N_0] = -\lambda t$ => **$N(t) = N_0 e^{-\lambda t}$**

  (passando ai numeri)

  dimostrazione

- $[\lambda] = [T^{-1}]$ => **$1/\lambda = \tau$ (vita media)**

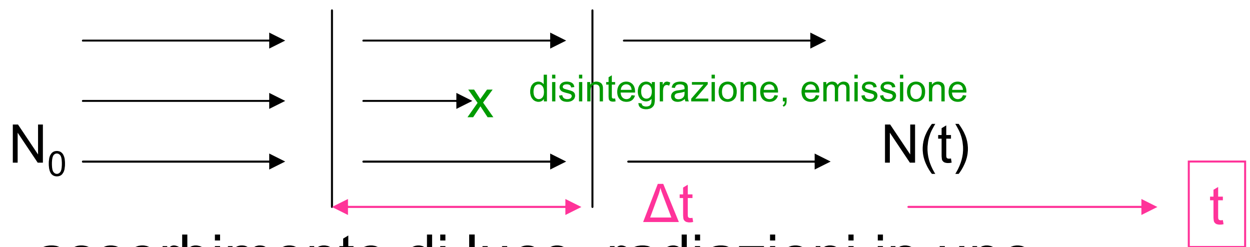
  dopo $t = \tau$: $N(\tau) = N_0 e^{-1} = N_0/2.718$

  tempo di dimezzamento $T_{1/2}$ $= \ln 2\tau = 0.693\tau$

  $(N_0/2 = N_0 e^{-0.693})$

Legge esponenziale/4 - **facoltativa**

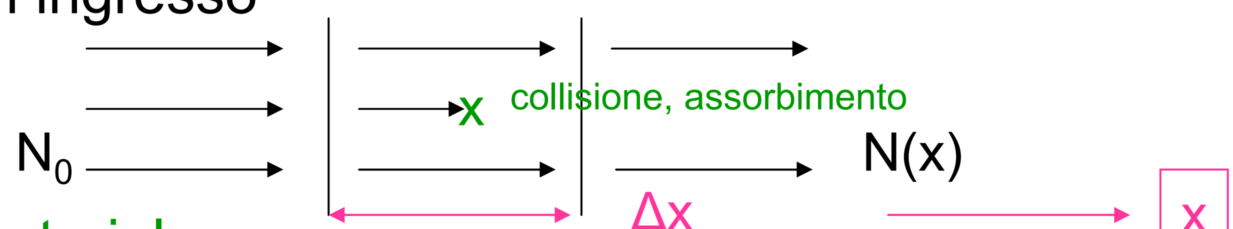
- pittoricamente



- analogamente – assorbimento di luce, radiazioni in uno spessore x di materiale – si ha

$$N(x) = N_0 e^{-\mu x}$$

con N_0 numero in ingresso

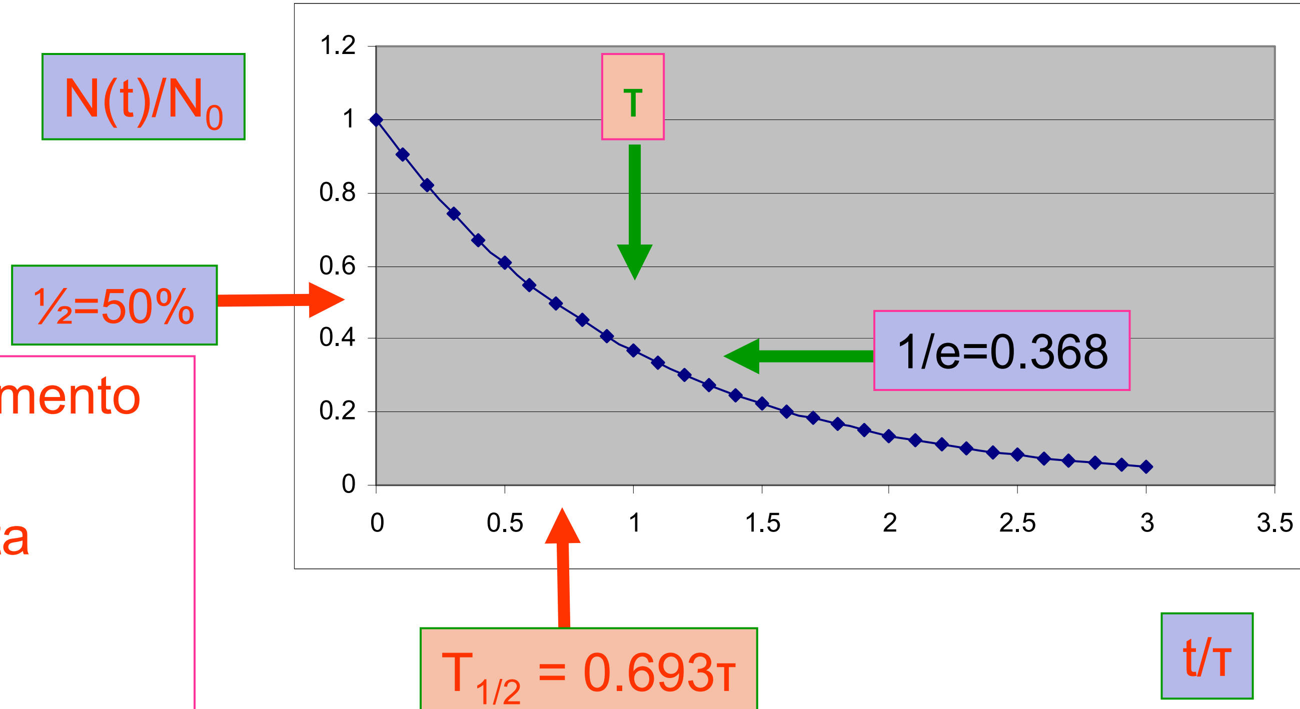


| materiale
μ dipende da | tipo di radiazione
| energia della radiazione

$[\mu] = [L^{-1}]$ si misura in m^{-1}, cm^{-1}, &tc.

coefficiente di assorbimento = 1/cammino libero medio

Legge esponenziale/5 - **facoltativa**



la legge del decadimento

$$N(t) = N_0 e^{-t/\tau}$$

può essere graficata come

$$N(t)/N_0 = e^{-t/\tau}$$

in funzione della variabile adimensionale t/τ

(allo stesso modo si può graficare la legge dell'assorbimento in funzione di μx)

dopo 3 vite medie rimane ~5% della popolazione iniziale