Nota sulle variabili casuali

E. Iacopini

26 gennaio 2023

Sommario

In questa nota tratteremo alcuni aspetti relativi alle variabili casuali, come la distribuzione binomiale, poissoniana e gaussiana.
Ci occuperemo anche di come definire operazioni algebriche con distribuzioni continue e, per quanto possibile, su quelle discrete.

1 Le variabili casuali

Una variabile casuale (*v.c.*) X è una variabile che assume valori non predicibili a priori ma per i quali è comunque possibile assegnare una probabilità definita. La variabile casuale può essere *discreta* o *continua*.

Nel caso discreto, i valori possibili di X saranno elementi dell'insieme[1] $\Omega_X \equiv \{x_n\}$, per ciascuno dei quali è definita una probabilità $P(x_n) > 0$ a cui richiediamo di soddisfare la condizione di normalizzazione[2]

$$\sum_{x_n} X(x_n) = 1 \tag{1}$$

Se X è una v.c. continua, i possibili valori di x per cui la probabilità sarà non nulla, apparterranno, per definizione, al dominio Ω_X di X, il quale sarà un sottoinsieme continuo della retta reale. In questo caso la v.c. è individuata da una p.d.f. (probability density function) $X(x)$ definita da Ω_X in R^+ e la condizione di normalizzazione diventa[3]

$$\int_{\Omega_x} dx\, X(x) = 1 \tag{2}$$

Data una v.c., attraverso la sua p.d.f., si possono definire alcune quantità fra cui, le due più importanti, sono la *media* $< X >$, detta anche *valore di aspettazione* $E(X) \equiv< X >$, e la *varianza*, cioè il valor medio del quadrato della differenza fra X e la sua media, ovvero la quantità $\sigma_X^2 \equiv (X- < X >)^2$. Più esplicitamente si ha

$$\begin{aligned} < X > &= \sum_n X(x_n)\, x_n \quad \textit{per v.c. discreta} && (3)\\ &= \int dx\, X(x)\, x \quad \textit{per v.c. continua} && (4)\\ \sigma_X^2 &= \sum_n X(x_n)\, (x_n- < X >)^2 \quad \textit{per v.c. discreta} && (5)\\ &= \int dx X(x)\, (x- < X >)^2 \quad \textit{per v.c. continua} && (6) \end{aligned}$$

Si dimostra poi facilmente che risulta

$$\begin{aligned} \sigma_X^2 &= \sum_n X(x_n)(x_n)^2- < X >^2 \quad \textit{per v.c. discreta} && (7)\\ &= \int dx X(x)\, x^2- < X >^2 \quad \textit{per v.c. continua} && (8) \end{aligned}$$

[1] L'insieme Ω_X è chiamato *dominio* della variabile casuale X: in generale si tratta di un sottoinsieme di numeri reali e potrà essere discreto o continuo.

[2] La somma è fatta sugli elementi x_n del dominio: per brevità di notazione, indicheremo questa somma anche mediante l'indice n.

[3] L'integrale è fatto sul dominio Ω_X ma, per brevità, nel seguito, ometteremo di scriverlo immaginando così una estensione del dominio della p.d.f. a tutta la retta reale, per valori nulli.

La radice quadrata della varianza, cioè la quantità σ_X, viene chiamata anche *scarto quadratico medio* (root mean square r.m.s.) oppure *deviazione standard*.

Definiremo adesso tre variabili casuali molto interessanti che sono, rispettivamente, la gaussiana, la poissoniana e la binomiale.

1.1 Distribuzione gaussiana

Una v.c. continua G è gaussiana (o di Gauss) quando la sua p.d.f. è *normale*, ovvero se accade che

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{9}$$

dove μ risulta essere il valor medio mentre σ^2 è la varianza, associati alla variabile casuale data.

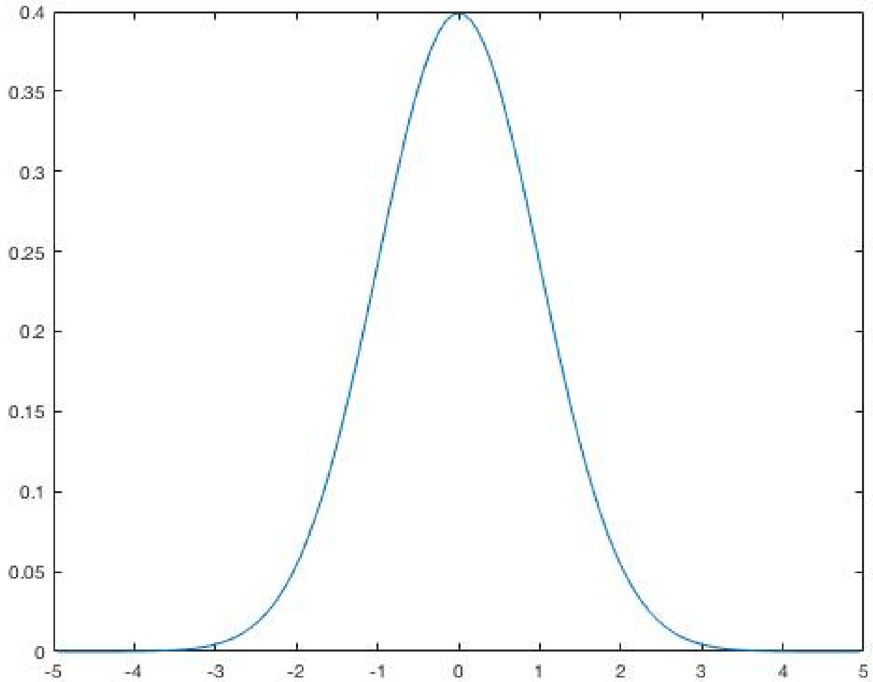


Figura 1: *Distribuzione gaussiana con media $\mu = 0$ e rms $\sigma = 1$*

Iniziamo dimostrando che la distribuzione $G(x)$, definita dalla (9), è normalizzata: si ha infatti

$$\begin{aligned}\int dx\, G(x) &= \frac{1}{\sigma\sqrt{2\pi}}\int dx\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}\int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}} = \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int \frac{d\xi}{\sigma\sqrt{2}}\, \sigma\sqrt{2}\, e^{-\frac{\xi^2}{2\sigma^2}} = \frac{1}{\sqrt{\pi}}\int dt\, e^{-t^2} = \frac{1}{\sqrt{\pi}}\sqrt{\pi} = 1 \end{aligned} \tag{10}$$

dove abbiamo posto $\xi = x - \mu$ e $t = \frac{\xi}{\sigma\sqrt{2}}$.

Veniamo ora al valor medio della distribuzione. Risulta

$$\begin{aligned}< G > &\equiv \int dx\; G(x)\, x = \frac{1}{\sigma\sqrt{2\pi}}\int dx\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, x = \frac{1}{\sigma\sqrt{2\pi}}\int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}}\, (\xi + \mu) = \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}}\; \xi + \mu \frac{1}{\sigma\sqrt{2\pi}}\int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}} \end{aligned} \tag{11}$$

ma il primo termine è nullo perché la funzione integranda è dispari, mentre il secondo termine, per quanto visto a proposito della normalizzazione, vale proprio μ; dunque è così provato che

$$< G >= \mu \tag{12}$$

Passiamo ora alla varianza. Si ha

$$\begin{aligned} < G^2 > &\equiv \int dx\ G(x)\, x^2 = \frac{1}{\sigma\sqrt{2\pi}} \int dx\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ x^2 = \frac{1}{\sigma\sqrt{2\pi}} \int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}}\, (\xi+\mu)^2 = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \left[\int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}}\, \xi^2 + 2\mu \int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}}\, \xi + \mu^2 \int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}} \right] \end{aligned} \tag{13}$$

Come già osservato, il secondo termine in parentesi quadra è nullo perchè la funzione integranda è dispari. Quanto al terzo termine, per quanto visto in precedenza, evidentemente esso vale $\mu^2\, \sigma\sqrt{2\pi}$.
Occupiamoci dunque del primo termine entro la parentesi quadra. Si ha

$$\int d\xi\, e^{-\frac{\xi^2}{2\sigma^2}}\, \xi^2 = \int \frac{d\xi}{\sigma\sqrt{2}}\, \sigma\sqrt{2}\ e^{-\frac{\xi^2}{2\sigma^2}}\ \frac{\xi^2}{2\sigma^2}\, 2\sigma^2 = 2\sigma^3\sqrt{2} \int dt\, e^{-t^2}\ t^2 \tag{14}$$

Osserviamo adesso che

$$\int dt\, e^{-t^2}\ t^2 = -\frac{d}{d\alpha} \int dt\, e^{-\alpha t^2} \Big|_{\alpha=1} \tag{15}$$

ma

$$\int dt\, e^{-\alpha t^2} = \frac{1}{\sqrt{\alpha}}\, \sqrt{\pi} \quad \Rightarrow \quad \int dt\, e^{-t^2}\ t^2 = \frac{1}{2}\, \sqrt{\pi} \tag{16}$$

e dunque

$$\begin{aligned} < G^2 > &= \frac{1}{\sigma\sqrt{2\pi}} \left[2\sigma^3\sqrt{2}\frac{1}{2}\, \sqrt{\pi} + \mu^2\, \sigma\sqrt{2\pi} \right] = \frac{1}{\sigma\sqrt{2\pi}}\ \sigma\sqrt{2\pi}\, \left[\sigma^2 + \mu^2\right] = \\ &= \sigma^2 + \mu^2 \end{aligned} \tag{17}$$

da cui, infine

$$\sigma_G^2 \equiv < G^2 > - < G >^2 = \sigma^2 \tag{18}$$

1.2 Distribuzione poissoniana

Una v.c. discreta P è poissoniana (o di Poisson) quando la sua p.d.f. è data da

$$P(n) = e^{-\mu}\,\frac{\mu^n}{n!} \tag{19}$$

dove $\mu > 0$ rappresenta sia il valor medio della distribuzione, come la sua varianza.

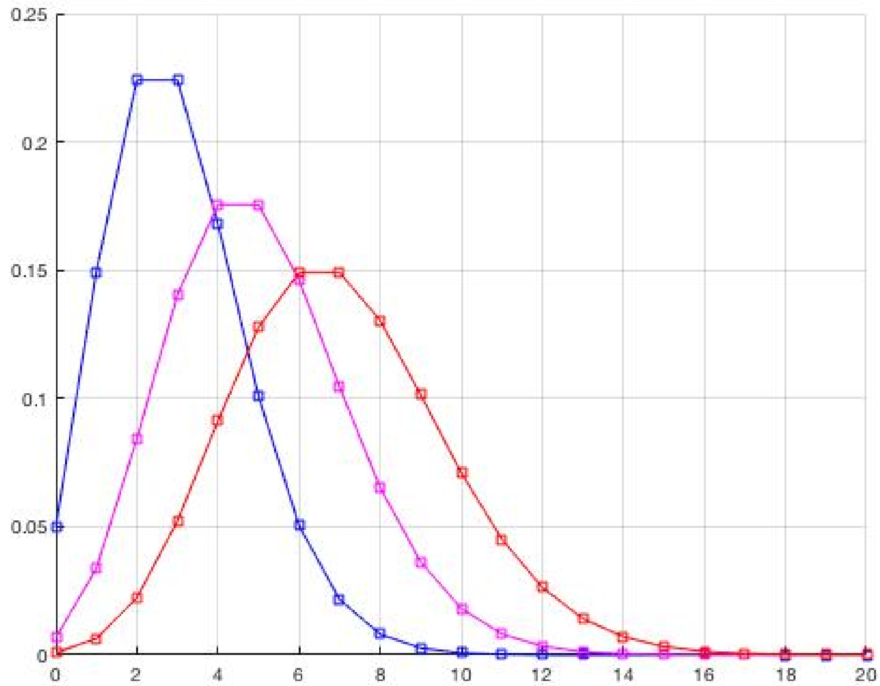


Figura 2: *Distribuzione di Poisson con medie pari a* 3, 5 *e* 7.

Iniziamo osservando che questa distribuzione di probabilità è normalizzata. Abbiamo infatti che

$$\sum_{n=0}^{\infty} P(n) = \sum_{n=0}^{\infty} e^{-\mu}\,\frac{\mu^n}{n!} = e^{-\mu}\sum_{n=0}^{\infty}\frac{\mu^n}{n!} = e^{-\mu}\,e^{\mu} = 1 \tag{20}$$

Determiniamone adesso il valor medio che abbiamo già anticipato essere uguale a μ. Si ha

$$\begin{aligned} < P > &\equiv \sum_{n=0}^{\infty} P(n)\,n = \sum_{n=0}^{\infty} e^{-\mu}\,\frac{\mu^n}{n!}\,n = e^{-\mu}\sum_{n=1}^{\infty}\frac{\mu^n}{(n-1)!} = e^{-\mu}\,\mu\sum_{n=1}^{\infty}\frac{\mu^{(n-1)}}{(n-1)!} = \\ &= \mu\,e^{-\mu}\sum_{m=0}^{\infty}\frac{\mu^m}{m!} = \mu\,e^{-\mu}\,e^{\mu} = \mu, \end{aligned} \tag{21}$$

Veniamo ora alla varianza, anch'essa, come anticipato, uguale a μ. Si ha

$$\begin{aligned}
< P^2 > \;&\equiv\; \sum_{n=0}^{\infty} P(n)\, n^2 = \sum_{n=0}^{\infty} e^{-\mu}\, \frac{\mu^n}{n!}\, n^2 = e^{-\mu} \sum_{n=1}^{\infty} \frac{\mu^n}{(n-1)!}\, n = \\
&= e^{-\mu}\, \mu \sum_{n=1}^{\infty} \frac{\mu^{(n-1)}}{(n-1)!}\, n = \mu\, e^{-\mu} \sum_{m=0}^{\infty} \frac{\mu^m}{m!} (m+1) = \\
&= \mu\, e^{-\mu} \left[\sum_{m=1}^{\infty} \frac{\mu^m}{(m-1)!} + \sum_{m=0}^{\infty} \frac{\mu^m}{m!} \right] = \mu\, e^{-\mu} \left[\mu \sum_{p=0}^{\infty} \frac{\mu^p}{p!} + \sum_{m=0}^{\infty} \frac{\mu^m}{m!} \right] = \\
&= \mu\, e^{-\mu} \left[\mu\, e^{\mu} + e^{\mu} \right] = \mu^2 + \mu
\end{aligned} \tag{22}$$

da cui ricaviamo infine

$$\sigma_p^2 \equiv < P^2 > - < P >^2 = \mu \tag{23}$$

1.3 Dalla distribuzione poissoniana a quella gaussiana

Tornando alla distribuzione poissoniana, come si può intuire anche dalla figura (2), al crescere del valor medio μ, essa tende alla distribuzione gaussiana (normale). Vediamone una dimostrazione.
Per questo utilizzeremo l'approssimazione di Stirling del fattoriale che stabilisce che, per grandi valori di n, risulta

$$n! \approx \sqrt{2\pi\, n}\, \left(\frac{n}{e}\right)^n = e^{-n}\, n^n\, \sqrt{2\pi n} \tag{24}$$

Si ha allora che

$$\begin{aligned} P(n) &= \frac{\mu^n}{n!}\, e^{-\mu} \approx \frac{e^{-\mu}\mu^n}{e^{-n}\, n^n\, \sqrt{2\pi n}} = \frac{e^{-\mu}\mu^n\sqrt{\mu}}{e^{-n}\, n^n\, \sqrt{2\pi n}\sqrt{\mu}} = \\ &= e^{n-\mu}\left(\frac{\mu}{n}\right)^{n+\frac{1}{2}} \frac{1}{\sqrt{2\pi\mu}} \end{aligned} \tag{25}$$

Definiamo ora la variabile ξ seguente

$$\xi \equiv \frac{n-\mu}{\sqrt{\mu}} \quad \Rightarrow \quad \frac{\xi}{\sqrt{\mu}} = \frac{n-\mu}{\mu} \quad \Rightarrow \quad n-\mu = \xi\sqrt{\mu} \tag{26}$$

Abbiamo

$$\begin{aligned} \left(\frac{\mu}{n}\right)^{n+\frac{1}{2}} &= \left(\frac{n}{\mu}\right)^{-(n+\frac{1}{2})} = e^{\log\left[\left(\frac{n}{\mu}\right)^{-(n+\frac{1}{2})}\right]} = e^{-(n+\frac{1}{2})\,\log\left(\frac{n}{\mu}\right)} = \\ &= e^{-(n+\frac{1}{2})\,\log\left(1+\frac{\xi}{\sqrt{\mu}}\right)} = e^{-\left(\mu+\xi\sqrt{\mu}+\frac{1}{2}\right)\,\log\left(1+\frac{\xi}{\sqrt{\mu}}\right)} \end{aligned} \tag{27}$$

Riguardo all'esponente che compare nella (27), possiamo sviluppare il logaritmo[4] in serie di Taylor, ottenendo

$$-\left(\mu+\xi\sqrt{\mu}+\frac{1}{2}\right)\,\log\left(1+\frac{\xi}{\sqrt{\mu}}\right) = -\left(\mu+\frac{1}{2}+\xi\sqrt{\mu}\right)\left[\frac{\xi}{\sqrt{\mu}} - \frac{\xi^2}{2\mu} + \frac{\xi^3}{3\mu^{3/2}} - \frac{\xi^4}{4\mu^2} + \ldots\right] \tag{29}$$

Se adesso supponiamo che μ sia abbastanza grande da poter trascurare i termini che lo vedono al denominatore, restiamo solamente con

$$-\xi\sqrt{\mu} + \frac{\xi^2}{2} - \xi^2 = -\xi\sqrt{\mu} - \frac{\xi^2}{2} \tag{30}$$

Dunque, per $\mu >> 1$ ed $n \approx \mu$, abbiamo

$$\begin{aligned} P(n) &\approx \frac{1}{\sqrt{2\pi\mu}}\, e^{n-\mu}\left(\frac{\mu}{n}\right)^{n+\frac{1}{2}} = \frac{1}{\sqrt{2\pi\mu}}\, e^{\xi\sqrt{\mu}}\, e^{-\xi\sqrt{\mu}-\frac{\xi^2}{2}} = \frac{1}{\sqrt{2\pi\mu}}\, e^{-\frac{\xi^2}{2}} = \\ &= \frac{1}{\sqrt{2\pi\mu}}\, e^{-\frac{(n-\mu)^2}{2\mu}} \end{aligned} \tag{31}$$

---

[4]Ricordiamo che

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots = \sum_{n=1}^{+\infty} (-1)^{n+1}\, \frac{x^n}{n} \tag{28}$$

1.4 Distribuzione binomiale

Si tratta di una v.c. discreta, avente un dominio Ω compreso fra 0 e un qualche numero naturale $N \geq 1$, a priori qualsiasi.
Per capire a che tipo di situazione si riferisca, immaginiamo di avere un processo A il quale abbia una probabilità p (con $0 < p < 1$) di avvenire con successo (o di avvenire, tout court) e dunque una probabilità $q = 1 - p$ di avvenire con insuccesso (o di non avvenire).
Supponiamo ora di considerare una successione di n processi indipendenti di tipo A: ogni volta potremo avere un risultato di successo oppure di insuccesso. La probabilità che una data successione contenga k successi e quindi $n-k$ insuccessi sarà, evidentemente $p^k\, q^{n-k}$. Però il numero di combinazioni in cui possono essere disposti i k successi negli n tentativi fatti, è dato dal coefficiente binomiale $\begin{pmatrix} n \\ k \end{pmatrix}$, dunque, la probabilità di ottenere k successi, comunque disposti nella sequenza degli n tentativi, è data da

$$P(n|k) = p^k\, q^{n-k} \begin{pmatrix} n \\ k \end{pmatrix} \tag{32}$$

dove $0 \leq k \leq n$.
Iniziamo dimostrando che la distribuzione $P(n|k)$ è normalizzata. Si ha

$$\sum_{k=0}^{n} P(n|k) = \sum_{k=0}^{n} p^k\, q^{n-k} \begin{pmatrix} n \\ k \end{pmatrix} = (p+q)^n = 1^n = 1 \tag{33}$$

dove abbiamo fatto uso della formula del binomio di Newton.

Determiniamo adesso il valor medio della distribuzione di probabilità binomiale data. Risulta

$$\begin{aligned}
< P > &\equiv \sum_{k=0}^{n} P(n|k)\, k = \sum_{k=0}^{n} p^k\, q^{n-k} \begin{pmatrix} n \\ k \end{pmatrix} k = \sum_{k=0}^{n} p^k\, q^{n-k} \frac{n!}{(n-k)!k!}\, k = \\
&= \sum_{k=1}^{n} p^k\, q^{n-k} \frac{n!}{(n-k)!(k-1)!} = \sum_{k=1}^{n} p^k\, q^{n-k} \frac{(n-1)!\, n}{[(n-1)-(k-1)]!(k-1)!} = \\
&= n \sum_{s=0}^{n-1} p^{s+1}\, q^{n-s-1} \frac{(n-1)!}{[(n-1)-s]!\, s!} = np \sum_{s=0}^{m} p^s\, q^{m-s} \frac{m!}{[m-s]!\, s!} = \\
&= np
\end{aligned} \tag{34}$$

dove abbiamo posto $s = k - 1$ e $m = n - 1$.

Calcoliamo ora la varianza. Si ha

$$
\begin{aligned}
< P^2 > \quad &\equiv \quad \sum_{k=0}^{n} P(n|k)\, k^2 = \sum_{k=0}^{n} p^k\, q^{n-k} \left(\begin{array}{c} n \\ k \end{array} \right) k^2 = \sum_{k=0}^{n} p^k\, q^{n-k}\, \frac{n!}{(n-k)!k!}\, k^2 = \\
&= \quad \sum_{k=1}^{n} p^k\, q^{n-k}\, \frac{n!}{(n-k)!(k-1)!}\, k = \sum_{k=1}^{n} p^k\, q^{n-k}\, \frac{(n-1)!\, n}{[(n-1)-(k-1)]!(k-1)!}\, k = \\
&= \quad n \sum_{s=0}^{n-1} p^{s+1}\, q^{n-s-1}\, \frac{(n-1)!}{[(n-1)-s]!\, s!}(s+1) = np \sum_{s=0}^{m} p^s\, q^{m-s}\, \frac{m!}{[m-s]!\, s!}(s+1) = \\
&= \quad np \left\{ \sum_{s=0}^{m} p^s\, q^{m-s}\, \frac{m!}{[m-s]!\, s!}\, s + \sum_{s=0}^{m} p^s\, q^{m-s}\, \frac{m!}{[m-s]!\, s!} \right\} = \\
&= \quad np\{mp+1\} = np\{(n-1)p+1\} = np(1-p+np) = \\
&= \quad npq + (np)^2
\end{aligned} \tag{35}
$$

e dunque ricaviamo infine che

$$\sigma^2 =< P^2 > - < P >^2 = npq \tag{36}$$

Nella figura (3) di cui sotto, riportiamo la distribuzione binomiale che si ottiene quando $p = 0.4$ e $n = 15$.

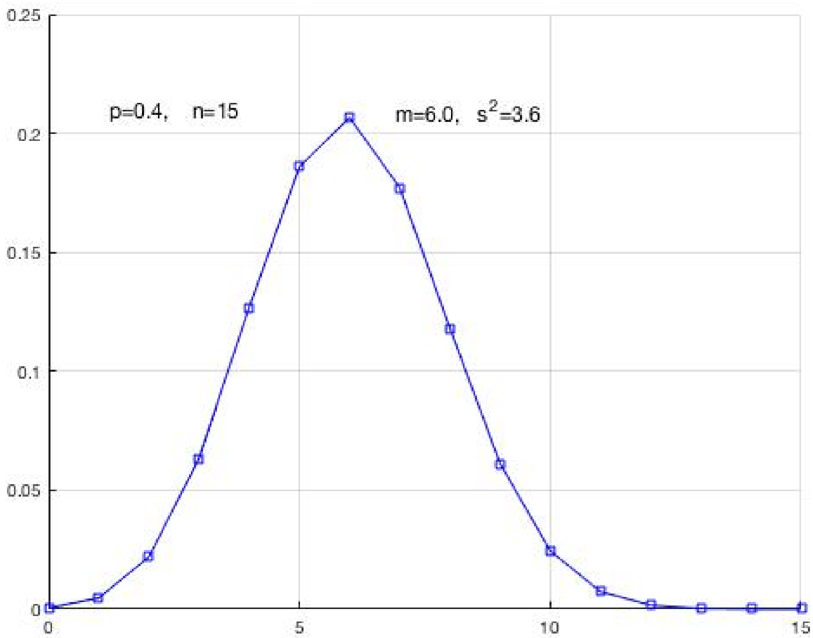


Figura 3: *Esempio di distribuzione binomiale*

1.5 Dalla distribuzione binomiale a quella poissoniana

Vogliamo dimostrare che, quando $n \to +\infty$ con $np = \mu$ fissato, la distribuzione binomiale tende alla distribuzione di Poisson con media $\mu = np$.
Poiché la distribuzione binomiale non ha la varianza uguale alla media come accade invece per quella poissoniana, potrebbe sembrare a priori impossibile che ci possa essere la convergenza di cui sopra, fra le due distribuzioni.
Ricordiamo che la varianza della binomiale è pari a $npq = np(1-p)$: nelle condizioni definite dal limite di cui sopra, cioè $p = \frac{\mu}{n}$, si traduce in $\mu(1-\frac{\mu}{n})$, che, a μ fissato, nel limite in cui $n \to +\infty$, tende, ovviamente a μ e dunque dà luogo a una varianza uguale alla media ...

Ma veniamo adesso alla dimostrazione.
Partiamo dunque dalla distribuzione binomiale e occupiamoci della probabilità di avere, con n prove totali, k risultati positivi, ciascuno avente probabilità p di manifestarsi: sappiamo che

$$P(n|k) = p^k\, q^{n-k} \begin{pmatrix} n \\ k \end{pmatrix} \tag{37}$$

Poniamo adesso

$$p = \frac{\mu}{n} \tag{38}$$

per cui abbiamo che

$$\begin{aligned} P(n|k) &= \left(\frac{\mu}{n}\right)^k \left(1-\frac{\mu}{n}\right)^{n-k} \begin{pmatrix} n \\ k \end{pmatrix} = \frac{\mu^k}{n^k} \left(1-\frac{\mu}{n}\right)^{n-k} \frac{n!}{(n-k)!\,k!} = \\ &= \frac{\mu^k}{k!} \left(1-\frac{\mu}{n}\right)^{n} \left(1-\frac{\mu}{n}\right)^{-k} \frac{1}{n^k} \frac{n!}{(n-k)!} \end{aligned} \tag{39}$$

Vogliamo dimostrare che, nel limite in cui $n \to +\infty$, risulta appunto che

$$P(n|k) \;\; \to \;\; e^{-\mu}\, \frac{\mu^k}{k!} \tag{40}$$

D'altronde, nel limite in cui $n \to +\infty$, $\left(1-\frac{\mu}{n}\right)^n \to e^{-\mu}$ quindi la tesi sarà dimostrata se proveremo che, nello stesso limite, risulta

$$R \equiv \left(1-\frac{\mu}{n}\right)^{-k} \frac{1}{n^k} \frac{n!}{(n-k)!} \;\; \to \;\; 1 \tag{41}$$

Per procedere alla dimostrazione, faremo uso, adesso, dell'approssimazione di Stirling del fattoriale, di cui abbiamo già parlato, che stabilisce che

$$\lim_{n \to +\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1 \quad \Rightarrow \quad per\ n >> 1 : \, n! \, \approx \, \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{42}$$

Abbiamo dunque

$$\begin{aligned}
R &= \frac{n^k}{(n-\mu)^k}\frac{1}{n^k}\;\frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi(n-k)}\left(\frac{n-k}{e}\right)^{n-k}} = \frac{1}{(n-\mu)^k}\sqrt{\frac{n}{n-k}}\left(\frac{n}{e}\right)^n\left(\frac{e}{n-k}\right)^{n-k} = \\
&= \sqrt{\frac{n}{n-k}}\;\frac{n^n}{(n-k)^{n-k}}\frac{e^{n-k}}{e^n}\;\frac{1}{(n-\mu)^k} = \sqrt{\frac{n}{n-k}}\;\frac{n^{n-k}\,n^k}{(n-k)^{n-k}}\,e^{-k}\;\frac{1}{(n-\mu)^k} = \\
&= \sqrt{\frac{n}{n-k}}\left(\frac{n}{n-k}\right)^{n-k} e^{-k}\left(\frac{n}{n-\mu}\right)^k = \\
&= \sqrt{\frac{n}{n-k}}\left(\frac{n-k+k}{n-k}\right)^{n-k} e^{-k}\left(\frac{n-\mu+\mu}{n-\mu}\right)^k = \\
&= \sqrt{\frac{n}{n-k}}\left(1+\frac{k}{n-k}\right)^{n-k} e^{-k}\left(1+\frac{\mu}{n-\mu}\right)^k
\end{aligned} \tag{43}$$

ma, fissato k, ecco che

$$\lim_{n\to+\infty}\sqrt{\frac{n}{n-k}} = 1 \tag{44}$$

$$\lim_{n\to+\infty}\left(1+\frac{k}{n-k}\right)^{n-k} = e^k \tag{45}$$

$$\lim_{n\to+\infty}\left(1+\frac{\mu}{n-\mu}\right)^k = 1 \tag{46}$$

e dunque è dimostrato che $\lim_{n\to+\infty} R = 1$ ovvero che, per grandi n e con $np = \mu$ fissato, la distribuzione binomiale tende a quella poissoniana.

Come esempio, possiamo considerare quello di una sostanza radioattiva costituita da $n >> 1$ radionuclidi di vita media τ. Se fissiamo una finestra temporale di osservazione lunga $T << \tau$, avremo che, durante questo tempo, un radionuclide avrà la probabilità $p = \frac{T}{\tau} << 1$ di decadere e dunque la probabilità $q = 1-p$ molto prossima a 1 di non farlo. In media, nel tempo T, osserveremo np decadimenti e la distribuzione che dà origine a questa media è, appunto, quella binomiale: in ogni finestra di osservazione T, partendo sempre da n radionuclidi iniziali, conteremo k decadimenti, con probabilità

$$P(n|k) = p^k\,q^{n-k}\begin{pmatrix} n \\ k \end{pmatrix} \quad\approx\quad e^{-np}\,\frac{(np)^k}{k!} \tag{47}$$

1.6 Dalla distribuzione binomiale a quella gaussiana

Vogliamo dimostrare che, per p fissato, nel limite di grandi valori di n, la distribuzione binomiale tende a quella gaussiana con media np e varianza npq, ovvero

$$P(n|k) = p^k\, q^{n-k} \left(\begin{array}{c} n \\ k \end{array} \right) \quad \rightarrow \quad \frac{1}{\sqrt{2npq}}\, e^{-\frac{(k-np)^2}{2npq}} \tag{48}$$

La dimostrazione utilizzerà l'approssimazione di Stirling del fattoriale che abbiamo già citato (cfr.(42)), secondo la quale, per grandi[5] valori di n, abbiamo

$$n! \approx \sqrt{2\pi\, n} \left(\frac{n}{e}\right)^n \tag{49}$$

Poichè la zona di interesse della probabilità è vicino al suo valor medio, ovvero per $k \approx np$, quando $n \to \infty$, anche i k di interesse divengono molto grandi e così pure $n-k \approx nq$, a meno di casi limite in cui p è vicinissimo allo zero oppure all'unità.
Questo garantisce la corretta applicabilità dell'approssimazione di Stirling. Abbiamo dunque

$$\begin{aligned}
P(n|k) &= p^k\, q^{n-k} \frac{n!}{(n-k)!k!} \approx \\
&\approx p^k\, q^{n-k} \sqrt{2\pi\, n} \left(\frac{n}{e}\right)^n \frac{1}{\sqrt{2\pi (n-k)}} \left(\frac{e}{n-k}\right)^{n-k} \frac{1}{\sqrt{2\pi\, k}} \left(\frac{e}{k}\right)^k = \\
&= p^k\, q^{n-k} \sqrt{\frac{n}{2\pi (n-k)\, k}}\, \frac{n^n}{(n-k)^{n-k}\, k^k} = \\
&= \sqrt{\frac{n}{2\pi (n-k)\, k}}\;\; \frac{p^k\;\, q^{n-k}\;\, n^k\;\, n^{n-k}}{(n-k)^{n-k}\;\, k^k} = \\
&= \sqrt{\frac{n}{2\pi (n-k)\, k}}\;\; \left(\frac{p\, n}{k}\right)^k \left(\frac{q\, n}{n-k}\right)^{n-k} \equiv Q(n,k)
\end{aligned} \tag{50}$$

dove $Q(n,k)$ è il risultato che si ottiene da $P(n|k)$ usando l'approssimazione di Stirling per i fattoriali. Passando ai logaritmi, si ha

$$\log\left(Q(n,k)\right) \;\approx\; \frac{1}{2}\log\left(\frac{n}{2\pi (n-k)\, k}\right) + k\,\log\left(\frac{p\, n}{k}\right) + (n-k)\log\left(\frac{q\, n}{n-k}\right) \tag{51}$$

Iniziamo dall'ultimo termine della (51): risulta

$$\begin{aligned}
\log\left(\frac{q\, n}{n-k}\right) &= \log\left(\frac{(1-p)\, n}{n-k}\right) = \log\left(\frac{n}{n-k} - \frac{np}{n-k}\right) = \\
&= \log\left(\frac{n-k}{n-k} + \frac{k}{n-k} - \frac{np}{n-k}\right) = \log\left(1 + \frac{k-np}{n-k}\right)
\end{aligned} \tag{52}$$

[5]Già per n=10 l'errore percentuale che commettiamo usando l'approssimazione di Stirling al posto del fattoriale è inferiore all'uno percento.

Definiamo adesso la differenza d dalla media, ovvero poniamo

$$d \equiv k - np \quad \Rightarrow \quad \log\left(\frac{q\,n}{n-k}\right) = \log\left(1 + \frac{d}{n-k}\right) \tag{53}$$

Per $k \approx np$, poichè $\frac{d}{n-k}$ sarà piccolo rispetto a uno, potremo sviluppare il logaritmo in serie di Taylor e, fermandoci al secondo ordine, si ha

$$\log\left(\frac{q\,n}{n-k}\right) \approx \frac{d}{n-k} - \frac{1}{2}\left(\frac{d}{n-k}\right)^2 \tag{54}$$

e dunque il terzo addendo della (51) potrà essere riscritto come

$$(n-k)\log\left(\frac{q\,n}{n-k}\right) \approx d - \frac{1}{2}\,\frac{d^2}{n-k} \tag{55}$$

Consideriamo ora il secondo addendo della (51): risulta

$$\begin{aligned} k\,\log\left(\frac{n\,p}{k}\right) &= k\,\log\left(\frac{k-d}{k}\right) = k\,\log\left(1 - \frac{d}{k}\right) \approx \\ &\approx k\left[\left(-\frac{d}{k}\right) - \frac{1}{2}\left(-\frac{d}{k}\right)^2\right] = -d - \frac{1}{2}\,\frac{d^2}{k} \end{aligned} \tag{56}$$

Sostituendo nella (51) sia la (55) che la (56), abbiamo

$$\begin{aligned} \log\left(Q(n,k)\right) &\approx \frac{1}{2}\,\log\left(\frac{n}{2\pi(n-k)\,k}\right) + d - \frac{1}{2}\,\frac{d^2}{n-k} - d - \frac{1}{2}\,\frac{d^2}{k} = \\ &= \frac{1}{2}\,\log\left(\frac{n}{2\pi(n-k)\,k}\right) - \frac{1}{2}d^2\left(\frac{1}{n-k} + \frac{1}{k}\right) = \\ &= \frac{1}{2}\,\log\left(\frac{1}{2\pi}\,\frac{n}{(n-k)\,k}\right) - \frac{1}{2}d^2\,\frac{n}{(n-k)k} \end{aligned} \tag{57}$$

Osserviamo adesso che, essendo $d = k - np$, risulta $k = d + np$ e dunque

$$\begin{aligned} \frac{k(n-k)}{n} &= \frac{(d+np)(n-d-np)}{n} = \frac{(d+np)(nq-d)}{n} = \frac{n^2pq + dn(q-p) - d^2}{n} = \\ &= npq + d(q-p) - \frac{d^2}{n} \end{aligned} \tag{58}$$

per cui, in prossimità di $k = np$, ovvero per piccoli valori di d rispetto a uno

$$\frac{k(n-k)}{n} \approx npq \;\Rightarrow\; \log\left(Q(n,k)\right) \;\approx\; \frac{1}{2}\,\log\left(\frac{1}{2\pi\,npq}\right) - \frac{d^2}{2\,npq} \tag{59}$$

e dunque, passando dai logaritmi ai numeri, otteniamo finalmente il risultato cercato, cioè

$$P(n|k) \approx Q(n,k) \approx \frac{1}{\sqrt{2\pi\,npq}}\;e^{-\frac{d^2}{2npq}} = \frac{1}{\sqrt{2\pi\,npq}}\;e^{-\frac{(k-np)^2}{2npq}} \tag{60}$$

1.7 Variabili casuali indipendenti

Supponiamo che X e Y siano due variabili casuali che, per concretezza, possiamo immaginare continue. Indichiamo con $P(x,y)$ la densità di probabilità di ottenere il valore x per la X e il valore y per la Y. Per ipotesi

$$\int dx\, dy\ P(x,y) = 1 \tag{61}$$

Integrando la densità $P(x,y)$ nella variabile y (oppure nella variabile x), determiniamo la densità di probabilità associata alla sola v.c. X (oppure Y), senza altre condizioni:

$$X(x) = \int dy\, P(x,y) \quad \Rightarrow \quad < X >= \int dx\, dy\, P(x,y) \cdot x \tag{62}$$

$$Y(y) = \int dx\, P(x,y) \quad \Rightarrow \quad < Y >= \int dx\, dy\, P(x,y) \cdot y \tag{63}$$

Definiamo ora la funzione di correlazione[6] $C(x,y)$ fra le due v.c. come

$$C(x,y) \equiv \int dx\, dy\, P(x,y)\ (x- < X >)\, (y- < Y >) \tag{64}$$

Essa è nulla se e solo se X e Y sono indipendenti.
Un'altra densità di probabilità interessante è la *probabilità condizionata* $P(x|y)$ di ottenere il valore x riguardo a X, **essendo dato** il valore y per Y.
Per definizione[7] si ha

$$P(x|y) \equiv \frac{P(x,y)}{Y(y)} \tag{65}$$

Nel caso in cui le due variabili casuali X e Y siano *indipendenti* fra loro, la probabilità condizionata $P(x|y)$ non dipende dal valore di y e avremo

$$P(x,y) = X(x) \cdot Y(y) \tag{66}$$

ovvero, per v.c. inipendenti, la densità $P(x,y)$ fattorizza semplicemente nel prodotto delle due densità di probabilità che si riferiscono a X e Y.

[6] Talvolta la definizione di questa funzione prevede che sia divisa per il prodotto $\sigma_X \cdot \sigma_Y$.

[7] Come si vede, al denominatore della (65) è presente la probabilità $Y(y)$. Poiché abbiamo imposto che y sia dato, questa probabilità non può essere nulla ...

2 Operazioni con variabili casuali

2.1 Somma di una v.c. con un numero reale

Sia data una variabile casuale (continua) X, con p.d.f. $X(x)$: definiamo la v.c. X' *somma di* X *con una costante reale* a in modo che la p.d.f. di X' sia tale che

$$X'(y) = X(y-a) \tag{67}$$

Evidentemente anche $X'(y)$ risulta normalizzata.
Quanto al valor medio, si ha

$$\begin{aligned} < X' > &\equiv \int dy\, X'(y)\, y = \int dy\, X(y-a)\, y = \int d\xi\, X(\xi)(\xi + a) = \\ &= < X > +a \end{aligned} \tag{68}$$

dove si è posto $\xi = y - a$.
Venendo ora al calcolo della varianza, si ha

$$\begin{aligned} < (X')^2 > &\equiv \int dy\, X'(y)\, y^2 = \int dy\, X(y-a)\, y^2 = \int d\xi\, X(\xi)(\xi + a)^2 = \\ &= \int d\xi\, X(\xi)(\xi^2 + 2a\xi + a^2) =< X^2 > +2a < X > +a^2 \end{aligned} \tag{69}$$

Ma la varianza è definita come $< (X')^2 > - < X' >^2$, dunque

$$\begin{aligned} < (X')^2 > - < X' >^2 &= < X^2 > +2a < X > +a^2 - (< X >^2 +2a < X > +a^2) = \\ &= < X^2 > - < X >^2 \end{aligned} \tag{70}$$

ovvero le due v.c. X e X' hanno la stessa varianza (come era ovvio che dovesse essere ...) e dunque anche la stessa r.m.s. .

2.2 Prodotto di una v.c. con un numero reale

Sia data una variabile casuale continua X, con p.d.f. $X(x)$: definiamo la v.c. $\hat{X}$ *prodotto di* X *con una costante reale* $a \neq 0$ in modo che la p.d.f. di $\hat{X}$ sia tale che (la pdf di una v.c. è necessariamente non negativa ...)

$$\hat{X}(y) = \frac{1}{|a|} X\left(\frac{y}{a}\right) \tag{71}$$

Iniziamo verificando che $\hat{X}$ sia normalizzata. Se $a > 0$, risulta

$$\int dy\, \hat{X}(y) = \frac{1}{a} \int dy\, X\left(\frac{y}{a}\right) = \int d\xi\, \hat{X}(\xi) = 1 \tag{72}$$

dove abbiamo posto $\xi = \frac{y}{a}$ e quindi y e ξ hanno lo stesso segno.
Se $a < 0$, poiché y e ξ hanno segno discorde, abbiamo

$$\int_{-\infty}^{+\infty} dy\, \hat{X}(y) = -\int_{-\infty}^{+\infty} \frac{dy}{a}\; X\left(\frac{y}{a}\right) = -\int_{+\infty}^{-\infty} d\xi\, \hat{X}(\xi) = \int_{-\infty}^{+\infty} d\xi\, \hat{X}(\xi) = 1 \tag{73}$$

Determiniamo adesso il valor medio di $\hat{X}$: si ha[8]

$$\begin{aligned} <\hat{X}> \;&\equiv\; \int_{-\infty}^{+\infty} dy\, \hat{X}(y)\; y = \frac{1}{|a|} \int_{-\infty}^{+\infty} dy\, X\left(\frac{y}{a}\right)\; y = \\ &= \frac{\pm 1}{a} \int_{-\infty}^{+\infty} dy\, X\left(\frac{y}{a}\right)\; y = \pm \frac{a^2}{a} \int_{-\infty}^{+\infty} \frac{dy}{a} X\left(\frac{y}{a}\right)\; \frac{y}{a} = \\ &\pm\; a \int_{\mp\infty}^{\pm\infty} d\xi\, X(\xi)\; \xi = a \int_{-\infty}^{+\infty} d\xi\, X(\xi)\; \xi = a <X> \end{aligned} \tag{74}$$

Veniamo quindi alla varianza di $\hat{X}$. Abbiamo

$$\begin{aligned} <\hat{X}^2> \;&\equiv\; \int_{-\infty}^{+\infty} dy\, \hat{X}(y)\; y^2 = \frac{1}{|a|} \int_{-\infty}^{+\infty} dy\, X\left(\frac{y}{a}\right)\; y^2 = \\ &= \frac{\pm 1}{a} \int_{-\infty}^{+\infty} dy\, X\left(\frac{y}{a}\right)\; y^2 = \pm \frac{a^3}{a} \int_{-\infty}^{+\infty} \frac{dy}{a} X\left(\frac{y}{a}\right)\; \frac{y^2}{a^2} = \\ &\pm\; a^2 \int_{\mp\infty}^{\pm\infty} d\xi\, X(\xi)\; \xi^2 = a^2 \int_{-\infty}^{+\infty} d\xi\, X(\xi)\; \xi^2 = a^2 <X^2> \end{aligned} \tag{75}$$

e dunque

$$<\hat{X}^2> - <\hat{X}>^2 = a^2 \left(<X^2> - <X>^2\right) \tag{76}$$

ovvero la varianza della v.c. $\hat{X}$ è quella della v.c. X, moltiplicata per a^2 e dunque la r.m.s associata a $\hat{X}$ è uguale a quella associata a X, moltiplicata per la costante a.

[8] Il segno positivo o negativo nelle espressioni seguenti è scelto concordemente con il segno di a.

2.3 Somma di due variabili casuali continue

Siano date due v.c. X e Y continue e indipendenti fra loro.
Definiamone la v.c. *somma*[9] delle due, ovvero $S \equiv X \oplus Y$.
Come sappiamo, occorre definirne la p.d.f.: abbiamo

$$\begin{aligned} S(s) &\equiv \int dx\, dy\, P(x,y)\, \delta(x+y-s) = \int dx\, dy\, X(x)\, Y(y)\, \delta(x+y-s) = \\ &= \int dx\, X(x)\, Y(s-x) \qquad (77) \end{aligned}$$

Segue dunque che, quanto alla media, risulta

$$\begin{aligned} < S > &= \int ds\ S(s)\, s = \int ds\, s \int dx\, dy\, X(x)\, Y(y)\, \delta(x+y-s) = \\ &= \int dx\, dy\, X(x)\, Y(y)\, (x+y) = < X > + < Y > \qquad (78) \end{aligned}$$

mentre, per quanto riguarda la varianza, abbiamo

$$\begin{aligned} \sigma_S^2 + < S >^2 &\equiv \int ds\, S(s)\, s^2 = \int ds\, s^2 \int dx\, dy\, X(x)\, Y(y)\, \delta(x+y-s) = \\ &= \int dx\, dy\, X(x)\, Y(y)\, (x+y)^2 = \\ &= \int dx\, x^2\, X(x) + \int dy\, y^2\, Y(y) + 2 \int dx\, dy\, X(x)\, Y(y)\, x\, y = \\ &= (\sigma_X^2 + < X >^2) + (\sigma_Y^2 + < Y >^2) + 2 < X >< Y > = \\ &= \sigma_X^2 + \sigma_Y^2 + (< X > + < Y >)^2 = \sigma_X^2 + \sigma_Y^2 + < S >^2 \qquad (79) \end{aligned}$$

e dunque

$$\sigma_S^2 = \sigma_X^2 + \sigma_Y^2 \qquad (80)$$

Prima di continuare, riproduciamo per questa strada il caso della somma di una v.c. con una **costante** reale. Partiamo, ovviamente, dalla (77) e supponiamo che la distribuzione $Y(y)$ abbia valor medio a e abbia un r.m.s. trascurabilmente piccolo: come condizione limite, immaginiamo dunque che

$$Y(y) = \delta(y-a) \qquad (81)$$

Abbiamo allora che

$$S(s) = \int dx\, X(x)\, Y(s-x) = \int dx\, X(x)\, \delta(s-x-a) = X(s-a) \qquad (82)$$

in accordo con quanto ottenuto per la somma della v.c. X con la costante reale a (cfr.(67)).

[9] Per indicare questa operazione, useremo il simbolo $\oplus$ per evitare possibili fraintendimenti. Non si tratta, infatti, della somma delle due p.d.f. (che non avrebbe senso, non foss'altro per la condizione di normalizzazione da rispettare), bensì della somma degli elementi del dominio dove le p.d.f. sono definite, ovvero dei valori che X e Y possono assumere.

Consideriamo ora, come primo esempio, la **somma di due distribuzioni piatte** tali che

$$
\begin{aligned}
X(x) &= \frac{1}{L_a} \;\; per \;\; a \le x \le a + L_a; \qquad X(x) = 0 \;\; altrove \\
Y(y) &= \frac{1}{L_b} \;\; per \;\; b \le y \le b + L_b; \qquad Y(y) = 0 \;\; altrove
\end{aligned} \tag{83}
$$

dove L_a e L_b sono quantità positive e, senza perdita di generalità, assumiamo che $L_a \le L_b$. Esplicitiamo la p.d.f. $S(s)$. Abbiamo

$$
S(s) \equiv \int_a^{a+L_a} dx\, X(x) \cdot Y(s-x) \tag{84}
$$

Ma nell'intervallo di integrazione, cioé fra a e $a + L_a$, la funzione $X(x)$ vale comunque sempre $\frac{1}{L_a}$, quindi possiamo riscrivere la pdf come

$$
S(s) = \frac{1}{L_a} \int_a^{a+L_a} Y(s-x) \tag{85}
$$

Per come è fatta la funzione Y la funzione S potrà essere diversa da zero solo nell'intervallo $[s_m, s_M]$ i cui estremi sono definiti da

$$
s_m = a + b \tag{86}
$$

$$
s_M = a + b + L_a + L_b \tag{87}
$$

Come mostra la figura 4, si distinguono facilmente tre zone

$$
\#1 : a + b \le s \le a + b + L_a \tag{88}
$$

$$
\#2 : a + b + L_a \le s \le a + b + L_b \tag{89}
$$

$$
\#3 : a + b + L_b \le s \le a + b + L_a + L_b \tag{90}
$$

Nelle zone #1 e #3, fissato s, l'intervallo di integrazione (dato dalla proiezione sulle ascisse del segmento definito dall'intersezione della retta $x + y = s$ con il rettangolo dei valori possibili per le due VC) non coincide con l'intero segmento $a \le x \le a + L_a$ e abbiamo così

$$
a + b \le s \le a + b + L_a : \qquad S(s) = \frac{1}{L_a L_b}(s - a - b) \tag{91}
$$

$$
a + b + L_a \le s \le a + b + L_b : \qquad S(s) = \frac{1}{L_b} \tag{92}
$$

$$
a + b + L_b \le s \le a + b + L_a + L_b : \qquad S(s) = \frac{1}{L_a L_b}(a + b + L_a + L_b - s) \tag{93}
$$

La funzione $S(s)$ è trapezoidale e diventa triangolare se $L_b = L_a$.
Nella figura 5 sono mostrate le pdf delle due distribuzioni piatte scelte, insieme alla pdf associata alla loro somma.

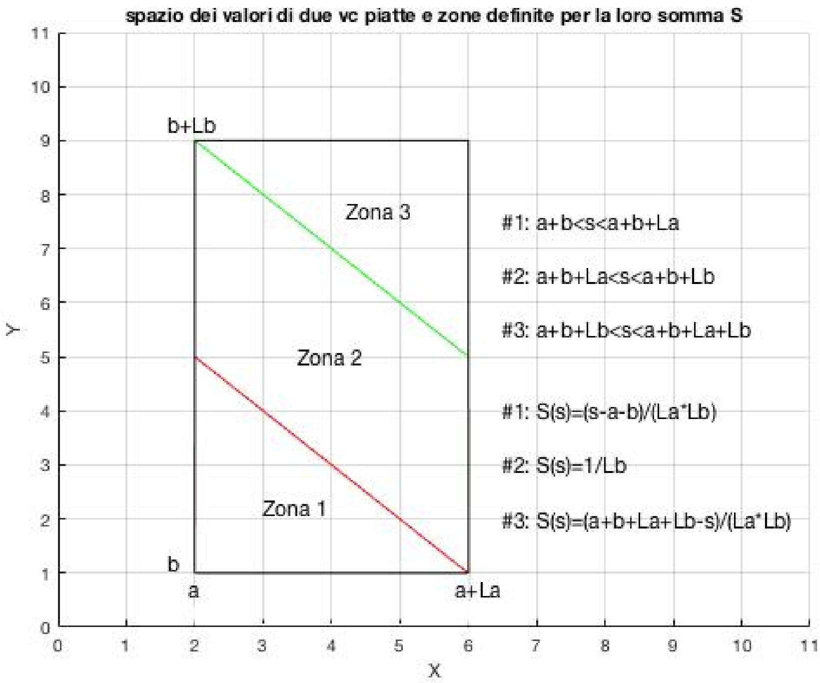


Figura 4:

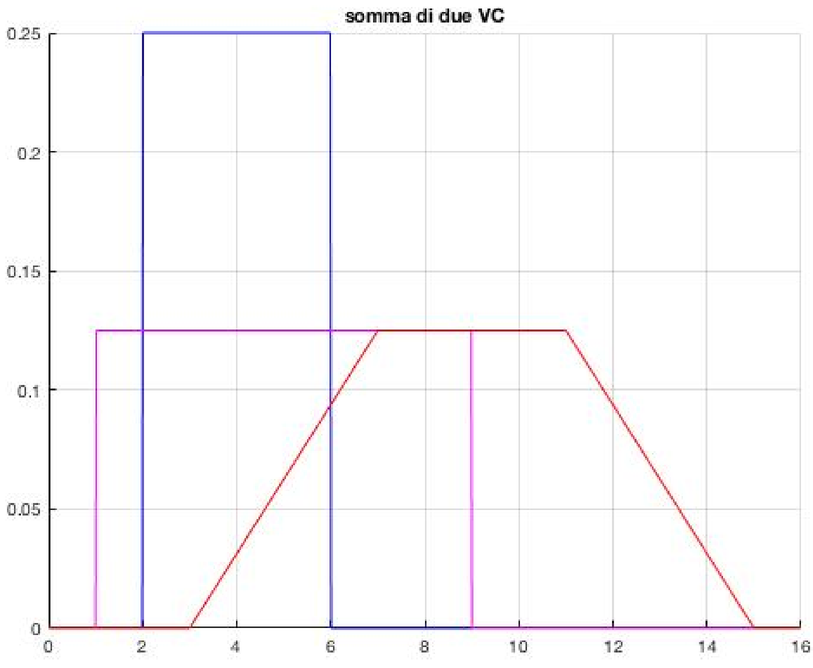


Figura 5:

Passiamo ora a considerare la **somma di due distribuzioni gaussiane**. Abbiamo

$$X(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \tag{94}$$

$$Y(y) = \frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} \tag{95}$$

Per quanto abbiamo visto (cfr. (77)), la pdf assocoata alla somma delle due v.c. risulterà data da

$$\begin{aligned} S(s) &= \int dx\, X(x)\, Y(s-x) = \frac{1}{2\pi\sigma_x\sigma_y}\int dx\, e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \cdot e^{-\frac{(s-x-\mu_y)^2}{2\sigma_y^2}} = \\ &= \frac{1}{2\pi\sigma_x\sigma_y}\int dx\, e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(s-x-\mu_y)^2}{2\sigma_y^2}} \end{aligned} \tag{96}$$

Consideriamo l'esponente dell'esponenziale che indicheremo con Z. Abbiamo

$$\begin{aligned} Z &\equiv -\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(s-x-\mu_y)^2}{2\sigma_y^2} = \frac{-\sigma_y^2(x-\mu_x)^2 - \sigma_x^2(s-x-\mu_y)^2}{2\sigma_x^2\sigma_y^2} = \\ &= -\frac{\sigma_x^2[x^2 - 2x\mu_x + \mu_x^2] + \sigma_x^2[x^2 + (s-\mu_y)^2 - 2x(s-\mu_y)]}{2\sigma_x^2\sigma_y^2} = \\ &= -\frac{x^2(\sigma_x^2+\sigma_y^2) - 2x[\mu_x\sigma_y^2 + (s-\mu_y)\sigma_x^2] + \mu_x^2\sigma_y^2 + (s-\mu_y)^2\sigma_x^2}{2\sigma_x^2\sigma_y^2} \equiv \\ &\equiv -\alpha^2x^2 + 2\alpha\beta\, x - \gamma \end{aligned} \tag{97}$$

dove abbiamo definito

$$\alpha^2 \equiv \frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2} \quad \Rightarrow \quad \alpha = \frac{1}{\sigma_x\sigma_y}\sqrt{\frac{\sigma_x^2+\sigma_y^2}{2}} \tag{98}$$

$$\beta \equiv \sqrt{\frac{1}{2(\sigma_x^2+\sigma_y^2)}}\,\frac{\mu_x\sigma_y^2 + (s-\mu_y)\sigma_x^2}{\sigma_x\sigma_y} \tag{99}$$

$$\gamma \equiv \frac{\mu_x^2\sigma_y^2 + (s-\mu_y)^2\sigma_x^2}{2\sigma_x^2\sigma_y^2} \tag{100}$$

Possiamo dunque riscrivere Z come

$$-\alpha^2x^2 + 2\alpha\beta\, x - \gamma = -(\alpha x - \beta)^2 + \beta^2 - \gamma \tag{101}$$

e dunque

$$\begin{aligned} S(s) &= \frac{1}{2\pi\sigma_x\sigma_y}\int dx\, e^{-(\alpha x-\beta)^2+\beta^2-\gamma} = \frac{1}{2\pi\sigma_x\sigma_y}e^{\beta^2-\gamma}\int dx e^{-(\alpha x-\beta)^2} = \\ &= \frac{1}{2\pi\sigma_x\sigma_y}e^{\beta^2-\gamma}\,\frac{1}{\alpha}\,\sqrt{\pi} = \frac{1}{2\alpha\,\sigma_x\sigma_y\sqrt{\pi}}\; e^{\beta^2-\gamma} \end{aligned} \tag{102}$$

Ma

$$\frac{1}{2\alpha\,\sigma_x\sigma_y\sqrt{\pi}} = \frac{1}{2\sigma_x\sigma_y\sqrt{\pi}}\sigma_x\sigma_y\sqrt{\frac{2}{\sigma_x^2+\sigma_y^2}} = \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}} \tag{103}$$

mentre

$$\begin{aligned}
\beta^2-\gamma \quad &= \quad \frac{1}{2(\sigma_x^2+\sigma_y^2)}\frac{[\mu_x\sigma_y^2+(s-\mu_y)^2\sigma_x^2]^2}{\sigma_x^2\sigma_y^2} - \frac{\mu_x^2\sigma_y^2+(s-\mu_y)^2\sigma_x^2}{2\sigma_x^2\sigma_y^2} = \\
&= \frac{\frac{[\mu_x\sigma_y^2+(s-\mu_y)^2\sigma_x^2]^2}{\sigma_x^2+\sigma_y^2} - \mu_x^2\sigma_y^2-(s-\mu_y)^2\sigma_x^2}{2\sigma_x^2\sigma_y^2} = \\
&= \frac{\mu_x^2\sigma_y^4+(s-\mu_y)^2\sigma_x^4+2\mu_x(s-\mu_y)\sigma_x^2\sigma_y^2-\mu_x^2\sigma_x^2\sigma_y^2-\mu_x^2\sigma_y^4-(s-\mu_y)^2\sigma_x^4-(s-\mu_y)^2\sigma_x^2\sigma_y^2}{2\sigma_x^2\sigma_y^2(\sigma_x^2+\sigma_y^2)} = \\
&= \frac{\sigma_x^2\sigma_y^2[2\mu_x(s-\mu_y)-\mu_x^2-(s-\mu_y)^2]}{2\sigma_x^2\sigma_y^2(\sigma_x^2+\sigma_y^2)} = \frac{-2\mu_x(\mu_y-s)-\mu_x^2-(\mu_y-s)^2}{2(\sigma_x^2+\sigma_y^2)} = \\
&= -\frac{(\mu_x+\mu_y-s)^2}{2(\sigma_x^2+\sigma_y^2)}
\end{aligned} \tag{104}$$

dunque, in definitiva abbiamo

$$S(s) \quad = \quad \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}}\; e^{-\frac{(\mu_x+\mu_y-s)^2}{2(\sigma_x^2+\sigma_y^2)}} \tag{105}$$

che mostra appunto come la v.c. che risulta dalla somma di due v.c. gaussiane indipendenti fra loro, è ancora una v.c. gaussiana, avente valor medio la somma dei due valori medi e varianza la somma delle due varianze.

2.4 Medie di variabili casuali

Media aritmetica

Data una v.c. X avente media μ e deviazione standard σ, ecco che se x è un suo generico valore, in assenza di ulteriori informazioni potremo assumere questo valore come una stima della media μ ovvero del valore di aspettazione $E(X) \equiv < X >$, con una indeterminazione data dalla deviazione standard σ.
Se vogliamo ridurre l'indeterminazione nella stima di μ, un modo è quello di effettuare la media aritmetica (detta anche semplicemente media) di n valori indipendenti di X.
Vediamo meglio come questa operazione realizzi quanto vogliamo ottenere.
Iniziamo definendo la variabile casuale *Somma* seguente

$$S \equiv \sum_{j=1}^{n} X_j \tag{106}$$

dove le v.c. X_j sono tutte copie indipendenti della v.c. X da cui siamo partiti.
Generalizzando quanto abbiamo visto per la somma di due v.c. indipendenti, possiamo senz'altro concludere che

$$< S >= \sum_{j=1}^{n} < X_j >= n\,\mu; \qquad\qquad \sigma_S^2 = n\,\sigma^2 \;\;\Rightarrow\;\; \sigma_S = \sigma\,\sqrt{n} \tag{107}$$

Definiamo adesso la *media* (aritmetica) delle X_j come

$$M \equiv \frac{1}{n}\;S = \frac{1}{n}\;\sum_{j=1}^{n} X_j \tag{108}$$

Per quanto abbiamo visto circa il prodotto di una variabile casuale per una costante reale, ecco che

$$< M >= \frac{1}{n}\;< S >= \mu; \qquad\qquad \sigma_M^2 = \frac{1}{n^2}\sigma_S^2 = \frac{1}{n}\,\sigma^2 \;\;\Rightarrow\;\; \sigma_M = \frac{1}{\sqrt{n}}\,\sigma \tag{109}$$

Come si vede, la nuova v.c. M ha lo stesso valor medio della v.c. X da cui siamo partiti, ma una deviazione standard ridotta del fattore $1/n$. Questo significa che un valore di M costituirà una migliore stima di μ di quanto non possa esserlo un valore di X.
C'è poi un altro fatto importante circa M che è legato al *teorema del limite centrale*.
Il teorema del limite centrale afferma che, la somma di n variabili casuali indipendenti aventi identica distribuzione è una variabile casuale che tende a distribuirsi normalmente, cioè secondo la distribuzione gaussiana, qualsiasi

sia la tipologia della distribuzione di partenza. Avremo quindi che, a prescindere dalla pdf di X, quanto a M, asintoticamente essa sarà distribuita in modo che

$$P(m) = \frac{1}{\sigma_M \sqrt{2\pi}} \, e^{-\frac{(m-\mu)^2}{2\sigma_M^2}} \quad dove \quad \sigma_M = \frac{\sigma}{\sqrt{n}} \tag{110}$$

Questo fatto ci consente di poter affermare, per esempio, che la probabilità che un generico valore m di M disti dal valore di aspettazione μ meno di una deviazione standard $\sigma_M = \frac{\sigma}{n}$ vale[10] il 68.3%.

Media pesata

La media aritmetica considerata sopra è un caso particolare di *media pesata*, in cui i pesi sono tutti uguali.
Ma vediamo la definizione generale di *media pesata.*
Immaginiamo di avere n variabili casuali X_i. Per ognuna di esse supponiamo di poter stabilire, con qualche criterio che considereremo in seguito, un *peso* w_i che consiste in un opportuno numero reale positivo. Poniamo quindi

$$M_w \equiv \frac{\sum_{i=1}^n w_i \, X_i}{\sum_{i=1}^n w_i} \tag{112}$$

Come si vede, se i pesi sono tutti uguali a 1 (in realtà basta che siano tutti uguali fra loro ...), ritroviamo la media aritmetica discussa in precedenza.
In generale, per quanto già visto, abbiamo

$$< M_w > \; = \; \frac{\sum_{i=1}^n w_i \, \mu_i}{\sum_{i=1}^n w_i} \tag{113}$$

$$\sigma_{M_w}^2 \; = \; \left(\frac{1}{\sum_{i=1}^n w_i} \right)^2 \sum_{i=1}^n w_i^2 \, \sigma_i^2 \tag{114}$$

La ragione per introdurre i pesi risiede nel fatto che si ritiene, a priori, che alcune v.c. che entrano nella media, forniscano informazioni più precise di altre e si vuole, quindi, che abbiano, appunto, un peso maggiore.
Supponiamo che le X_i abbiano tutte lo stesso valore di aspettazione (media) $\mu_i = \mu$, ma che abbiano deviazioni standard σ_i differenti, magari anche in modo rilevante. Non sembra dunque che, allo scopo di migliorare la stima di μ, la migliore strategia sia quella di determinare semplicemente la loro

[10] Abbiamo infatti che

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-\sigma}^{\mu+\sigma} dx \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\sigma}^{+\sigma} d\xi \, e^{-\frac{\xi^2}{2\sigma^2}} = \frac{1}{\sqrt{\pi}} \int_{-\frac{1}{\sqrt{2}}}^{+\frac{1}{\sqrt{2}}} d\tau \, e^{-\tau^2} =$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\frac{1}{\sqrt{2}}} d\tau \, e^{-\tau^2} = Erf\left(\frac{1}{\sqrt{2}}\right) = 0.683 \tag{111}$$

media aritmetica. Le v.c. con deviazione standard più piccola dovrebbero entrare nella media con un peso maggiore ... Solitamente, in questi casi, si usa come peso proprio l'inverso della deviazione standard, cioè $w_i = \frac{1}{\sigma_i}$.
Abbiamo allora che, per questa particolare media pesata, in cui, lo ricordiamo, $\mu_i = \mu$ e $w_i = \frac{1}{\sigma_i}$, risulta[11]

$$M_w = \frac{\sum_{i=1}^n \frac{X_i}{\sigma_i}}{\sum_{i=1}^n \frac{1}{\sigma_i}} \tag{115}$$

e dunque

$$< M_w > = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i}} \sum_{i=1}^n \frac{\mu}{\sigma_i} = \mu \tag{116}$$

$$\sigma^2_{M_w} = \left(\frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i}} \right)^2 \sum_{i=1}^n \frac{1}{\sigma_i^2} \sigma_i^2 = n \left(\frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i}} \right)^2 \tag{117}$$

Come si vede, il valore di aspettazione di M_w è uguale a μ, mentre la varianza[12] è data, adesso, dalla (117).

Supponiamo ora, per esempio, che le σ_i siano equamente distribuite fra σ_0 e $\sigma_0(1+t)$ con $t \geq 0$ e dunque con pdf piatta pari a $\frac{1}{\sigma_0 \, t}$.
Nel caso della **media aritmetica** abbiamo

$$\begin{aligned} \sigma^2_M &= \frac{1}{n^2} \sum_{I=1}^n \sigma_i^2 = \frac{1}{n} \frac{1}{n} \sum_{I=1}^n \sigma_i^2 \approx \frac{1}{n} \int_{\sigma_0}^{\sigma_0(1+t)} d\sigma \frac{1}{\sigma_0 \, t} \sigma^2 = \\ &= \frac{1}{n} \frac{1}{\sigma_0 t} \frac{1}{3} \left[\sigma_0^3 (1+t)^3 - \sigma_0^3 \right] = \\ &= \frac{1}{n} \frac{1}{\sigma_0 t} \frac{\sigma_0^3}{3} \left[1 + 3t + 3t^2 + t^3 - 1 \right] = \frac{\sigma_0^2}{n} \left(1 + t + \frac{t^2}{3} \right) \end{aligned} \tag{118}$$

dove abbiamo approssimato la media $\frac{1}{n} \sum_{I=1}^n \sigma_i^2$ come se la distribuzione delle deviazioni standard fosse continua e uniforme fra σ_0 e $\sigma_0(1+t)$.
Veniamo adesso alla **media pesata** con $w_i = \frac{1}{\sigma_i}$. Abbiamo

$$\sigma^2_{M_w} = \frac{n}{\left(\sum_{i=1}^n \frac{1}{\sigma_i} \right)^2} = \frac{n}{\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_i} \right)^2 n^2} = \frac{1}{n} \frac{1}{\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_i} \right)^2} \tag{119}$$

e, passando di nuovo al continuo, abbiamo

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_i} \approx \int_{\sigma_0}^{\sigma_0(1+t)} d\sigma \frac{1}{\sigma_0 \, t} \frac{1}{\sigma} = \frac{1}{\sigma_0 \, t} \log(1+t) \tag{120}$$

[11] Si osservi che le v.c. $\frac{X_i}{\sigma_i}$ hanno tutte la stessa deviazione standard, pari a 1.

[12] Se le σ_i sono tutte uguali a uno stesso valore σ, ritroviamo quanto già ottenuto nel caso della media aritmetica, ovvero una varianza uguale a $\frac{\sigma^2}{n}$.

per cui risulta infine che

$$\sigma_{M_w}^2 \quad = \quad \frac{1}{n} \frac{1}{\left(\frac{1}{\sigma_0 t} \log(1+t)\right)^2} = \frac{\sigma_0^2}{n} \frac{t^2}{(\log(1+t))^2} \tag{121}$$

Nella figura (6) che segue, riportiamo il rapporto $\frac{\sigma_M^2}{\sigma_{M_w}^2}$ per $0 \leq t \leq 1$.

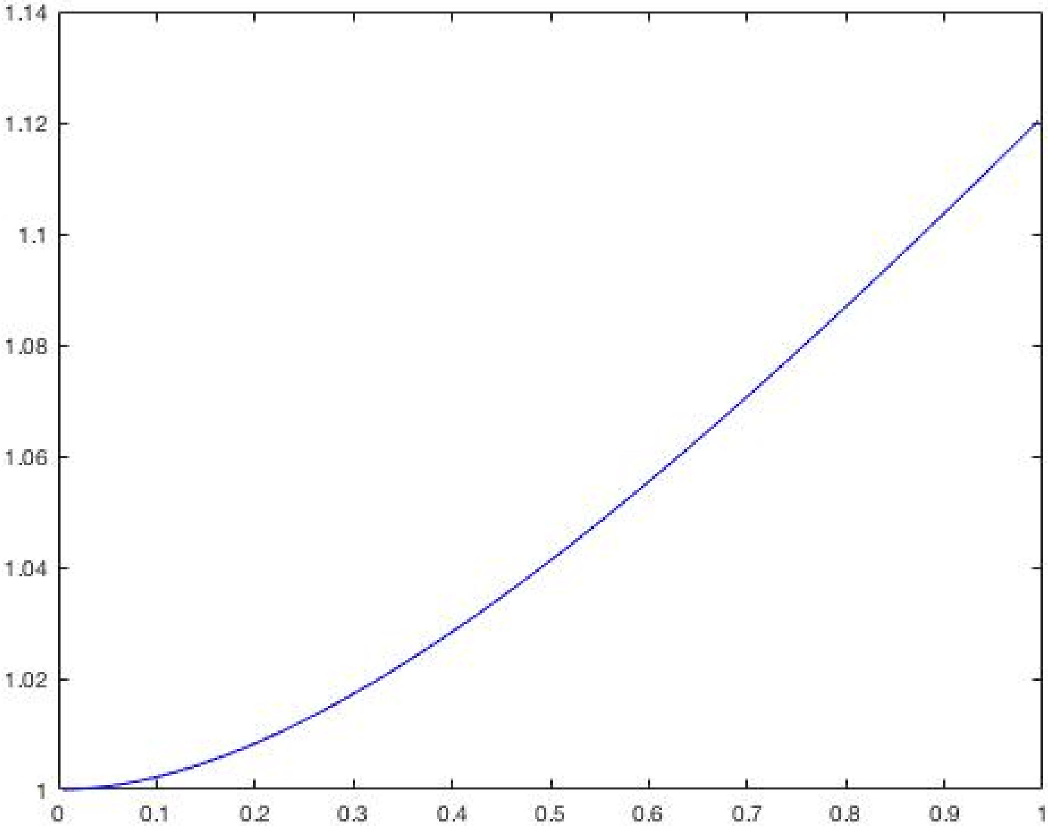


Figura 6:

2.5 Prodotto di due variabili casuali continue

Siano date due v.c. X e Y continue e indipendenti fra loro.
Definiamo la v.c. *prodotto* delle due, ovvero $P \equiv X \otimes Y$ come quella v.c. avente, in analogia con quanto già visto per la somma, la pdf definita come

$$P(p) \equiv \int dx\, dy\, X(x) \cdot Y(y)\ \delta(p - xy) \tag{122}$$

Iniziamo osservando che la definizione data rispetta la condizione di normalizzazione, infatti

$$\begin{aligned}&\int dp\, P(p) = \int dp \int dx\, dy\, X(x) \cdot Y(y)\ \delta(p - xy) = \int dx\, dy\, X(x) \cdot Y(y) = \\ &= \int dx X(x) \cdot \int dy\, Y(y) = 1 \end{aligned} \tag{123}$$

Passiamo a calcolare il valor medio associato a $P(p)$. Risulta

$$\begin{aligned} < P > &\equiv \int dp\, P(p)\, p = \int dp\, p \int dx\, dy\, X(x) \cdot Y(y)\ \delta(p - xy) = \\ &= \int dx\, dy\, X(x) \cdot Y(y)\ xy = \int dx X(x)\, x \cdot \int dy\, Y(y)\, y = \\ &= < X >< Y > \end{aligned} \tag{124}$$

Dunque, il valor medio della variabile casuale P è semplicemente il prodotto dei valori medi delle due variabili casuali X e Y.
Veniamo alla varianza. Si ha

$$\begin{aligned} \sigma_P^2 + < P >^2 &\equiv \int dp\, P(p)\, p^2 = \int dp\, p^2 \int dx\, dy\, X(x) \cdot Y(y)\ \delta(p - xy) = \\ &= \int dx\, dy\, X(x) \cdot Y(y)\ (xy)^2 = \int dx X(x)\, x^2 \cdot \int dy\, Y(y)\, y^2 = \\ &= \left(\sigma_X^2 + < X >^2\right) \left(\sigma_Y^2 + < Y >^2\right) \end{aligned} \tag{125}$$

e poiché, come abbiamo visto, $< P >=< X >< Y >$, risulta

$$\begin{aligned} \sigma_P^2 &= \sigma_X^2 \sigma_Y^2 + \sigma_X^2 < Y >^2 + \sigma_Y^2 < X >^2 + < X >^2 < Y >^2 - < X >^2 < Y >^2 = \\ &= \sigma_X^2 \sigma_Y^2 + \sigma_X^2 < Y >^2 + \sigma_Y^2 < X >^2 \end{aligned} \tag{126}$$

Come si vede, rispetto a quanto accadeva per la somma, adesso c'é una novità. La varianza del prodotto **non** è data solo dal prodotto delle varianze delle due v.c. che sono moltiplicate, ma occorre aggiungere anche i termini $\sigma_X^2 < Y >^2 + \sigma_Y^2 < X >^2$.
Vediamone meglio il significato.

Supponiamo, per questo, che la v.c. Y abbia valor medio $a \neq 0$ e un r.m.s. trascurabilmente piccolo, ovvero immaginiamo che si possa porre

$$Y(y) = \delta(y - a) \tag{127}$$

Abbiamo allora che

$$\begin{aligned} P(p) &= \int dx\, dy\, X(x)\, Y(y)\, \delta(p - xy) = \int dx\, dy\, X(x)\, \delta(y - a)\, \delta(p - xy) = \\ &= \int dx\, X(x)\, \delta(p - ax) = \frac{1}{|a|}\, X\left(\frac{p}{a}\right) \end{aligned} \tag{128}$$

ovvero otteniamo, per questa strada, quanto già ottenuto quando abbiamo trattato il caso della moltiplicazione di una v.c. per una costante reale (cfr.(71), ovvero $P = X\, Y \rightarrow a\, X$. Usando le formule di cui sopra, verifichiamo quanto già noto, ovvero che

$$< P > \;=\; < X >< Y >= a < X > \tag{129}$$

$$\sigma_P^2 \;=\; \sigma_X^2 \sigma_Y^2 + \sigma_X^2 < Y >^2 + \sigma_Y^2 < X >^2 = a^2 \sigma_X^2 \tag{130}$$

e, come si vede, nel quadrato dell'r.m.s., i termini legati ai valori medi delle due distribuzioni sono essenziali.

Tornando al prodotto di due distribuzioni casuali, indipendenti, allo scopo di concretizzare meglio il risultato ottenuto, consideriamo come primo esempio quello delle due v.c. piatte già considerate nel caso della somma ovvero tali che

$$\begin{aligned} X(x) &= \frac{1}{L_a} \;\; per \;\; a \leq x \leq a + L_a; \qquad X(x) = 0 \;\; altrove \\ Y(y) &= \frac{1}{L_b} \;\; per \;\; b \leq y \leq b + L_b; \qquad Y(y) = 0 \;\; altrove \end{aligned} \tag{131}$$

dove assumeremo che a, L_a e b, L_b siano quantità positive e che $b\, L_a \leq a\, L_b$. Esplicitiamo la p.d.f. $P(p)$. Abbiamo

$$P(p) \equiv \int_a^{a+L_a} dx \int_b^{b+L_b} dy\, X(x)\, Y(y)\, \delta(xy - p) \tag{132}$$

Integriamo la (132) nella variabile y: abbiamo così[13]

$$P(p) \;=\; \int_a^{a+L_a} dx\, X(x) \cdot Y\left(\frac{p}{x}\right) \frac{1}{|x|} \tag{134}$$

[13]Ricordiamo che

$$\int dy F(y)\, \delta[g(y)] = \sum_i \int dy F(y)\, \delta(y - y_i)\, \frac{1}{|\frac{dg}{dy}|_{y_i}} = \sum_i F(y_i)\, \frac{1}{|\frac{dg}{dy}|_{y_i}} \tag{133}$$

dove gli y_i individuano gli zeri della funzione $g(y)$.

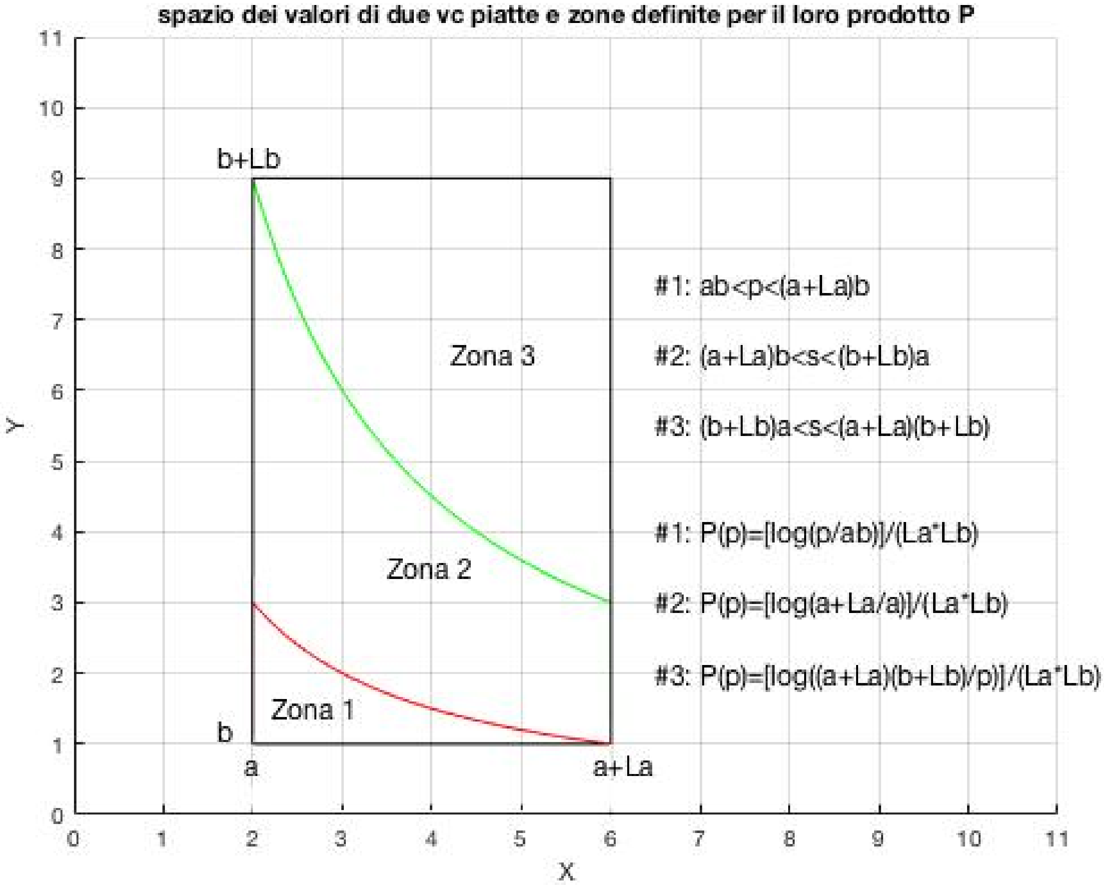


Figura 7:

Ma nell'intervallo di integrazione, cioé fra a e $a + L_a$, la funzione $X(x)$ vale comunque sempre $\frac{1}{L_a}$, quindi possiamo riscrivere la pdf come

$$P(p) = \frac{1}{L_a} \int_a^{a+L_a} dx \, Y\left(\frac{p}{x}\right) \frac{1}{x} \tag{135}$$

Osserviamo adesso che, date le due distribuzioni piatte, potremo avere un valore di $P(p)$ non nullo solo nell'intervallo $s_m \le p \le s_M$ dove

$$s_m = a \cdot b; \qquad s_M = (a + L_a)(b + L_b) \tag{136}$$

e in questo intervallo, come mostra la figura 7, dobbiamo di nuovo distinguere tre zone. Nelle zone #1 e #3, a differenza di quanto accade nella zona #2, l'intervallo di integrazione nella variabile x non è completo perché il tratto di iperbole equilatera $xy = p$ nel rettangolo dei possibili valori di X e Y ha una proiezione sull'asse delle ascisse che non coincide con tutto il segmento

$[a, a+L_a]$. Abbiamo così che

$$
\begin{aligned}
&zona\#1: ab \le p \le (a+L_a)b && a \le x \le \frac{p}{b} && (137)\\
&zona\#2: (a+L_a)b \le p \le (b+L_b)a && a \le x \le a+L_a && (138)\\
&zona\#3: (b+L_b)a \le p \le (a+L_a)(b+L_b) && \frac{p}{b+L_b} \le x \le a+L_a && (139)
\end{aligned}
$$

Ma, negli intervalli sopra specificati, la funzione Y è comunque pari a $1/L_b$, quindi l'integrale (134) si riduce alla differenza dei logaritmi negli estremi, diviso per L_b. Abbiamo dunque

$$\#1: P(p) = \frac{1}{L_a L_b}\left(\log[p/b] - \log[a]\right) = \frac{1}{L_a L_b}\log\frac{p}{ab} \tag{140}$$

$$\#2: P(p) = \frac{1}{L_a L_b}\left(\log[a+L_a] - \log[a]\right) = \frac{1}{L_a L_b}\log\frac{a+L_a}{a} \tag{141}$$

$$\#3: P(p) = \frac{1}{L_a L_b}\left(\log[a+L_a] - \log[p/(b+L_b)]\right) = \frac{1}{L_a L_b}\log\frac{(a+L_a)(b+L_b)}{p} \tag{142}$$

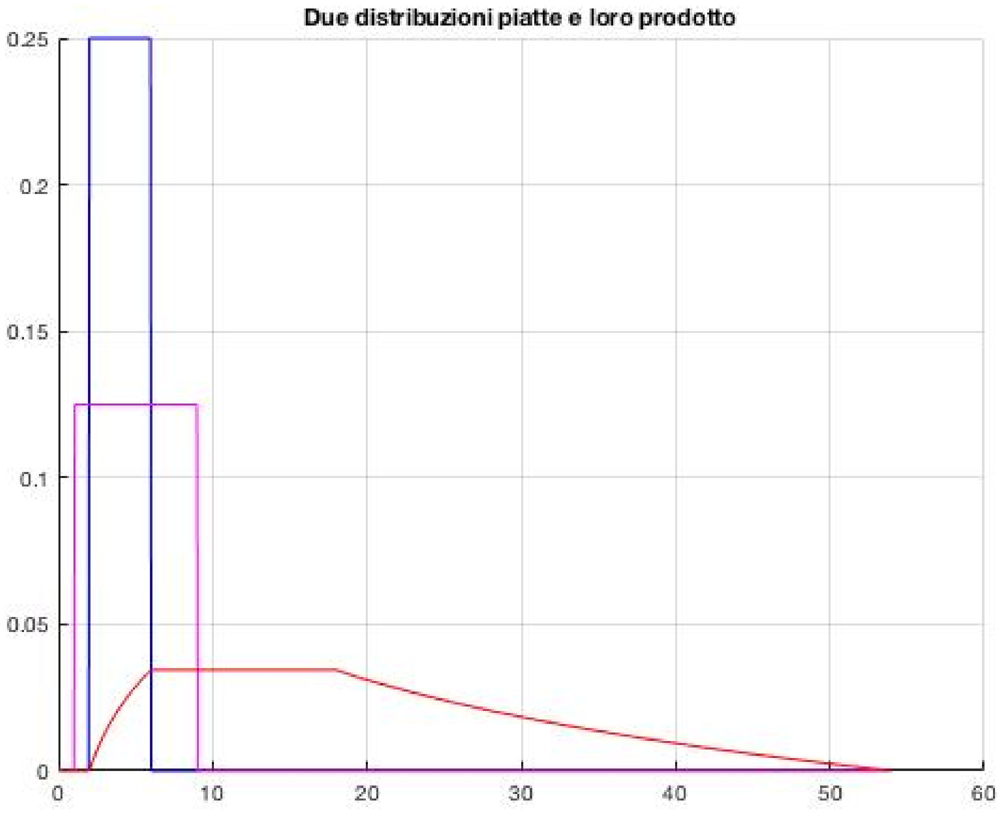


Figura 8:

Nella figura 8 sono riportate due distribuzioni piatte, la prima fra 2 e 6, la seconda fra 1 e 9, insieme alla distribuzione prodotto a cui esse danno origine, non nulla fra 2 e 54 con transizione dalla zona#1 alla zona#2 in 6 e dalla zona#2 alla zona#3 in 18.

Osserviamo infine che, nel limite in cui $L_a \to 0$, sia la zona #1 (compresa fra ab e $ab + bL_a$) che la zona #3 (compresa fra $a(b + L_b)$ e $L_a(b + L_b)$) tendono ad annullarsi. Resta solo la zona #2, compresa fra ab e $a(b + L_b)$, dove $P(p) = \frac{1}{L_a L_b} \log[(a + L_a)/a] \approx \frac{1}{L_a L_b}(L_a/a) = \frac{1}{a\, L_b}$.
Si tratta del caso, di cui abbiamo parlato prima, in cui la v.c. X tende ad assumere sempre lo stesso valore a ...

Assumiamo adesso che entrambe le v.c. indipendenti considerate siano gaussiane, ovvero che sia

$$X(x) = \frac{1}{s_x \sqrt{2\pi}}\, e^{-\frac{(x-\mu_x)^2}{2s_x^2}} \tag{143}$$

$$Y(y) = \frac{1}{s_y \sqrt{2\pi}}\, e^{-\frac{(y-\mu_y)^2}{2s_y^2}} \tag{144}$$

e consideriamo il loro prodotto. Come si è visto

$$P(p) = \int dx\, dy\, X(x)\, Y(y)\, \delta(p - xy) \tag{145}$$

e integrando la funzione delta si ottiene

$$\begin{aligned} P(p) &= \int dx\, X(x)\; Y\left(\frac{p}{x}\right) \frac{1}{|x|} = \\ &= \int dx\, \frac{1}{|x|}\, \frac{1}{2\pi\, s_x\, s_y}\, e^{-\frac{(x-\mu_x)^2}{2s_x^2}}\, e^{-\frac{(\frac{p}{x}-\mu_y)^2}{2s_y^2}} \end{aligned} \tag{146}$$

L'insieme dei possibili valori di p coincide con l'insieme dei numeri reali.
Fissato un valore non nullo di p, l'integrazione in x avviene sulla proiezione sull'asse delle ascisse dei due rami di iperbole equilatera $(x, p/x)$: se $p > 0$ i rami dell'iperbole sono nel primo e nel terzo quadrante, mentre se $p < 0$ i rami sono nel secondo e quarto quadrante. In ogni caso, l'integrazione avviene su tutto l'asse reale privato dello zero.
Il caso in cui $p = 0$ richiede che l'integrazione avvenga su tutto l'asse reale, zero compreso. Abbiamo infatti

$$\begin{aligned} P(0) &= \int dx\, dy\, X(x)\, Y(y)\, \delta(xy) = \int dx\, X(x) \int dy\, Y(y)\, \frac{1}{|x|}\delta(y) = \\ &= \int dx\, \frac{1}{|x|}\, X(x)\, Y(0) = \int dx\, \frac{1}{|x|}\, \frac{1}{2\pi\, s_x\, s_y}\, e^{-\frac{(x-\mu_x)^2}{2s_x^2}}\, e^{-\frac{\mu_y^2}{2s_y^2}} = \\ &= \frac{1}{2\pi\, s_x\, s_y}\, e^{-\frac{\mu_y^2}{2s_y^2}} \int dx\, \frac{1}{|x|}\, e^{-\frac{(x-\mu_x)^2}{2s_x^2}} \end{aligned} \tag{147}$$

Purtroppo né l'integrale (147), né, tantomeno, l'integrale (146), si riescono ad esplicitare analiticamente. I tre grafici della figura (9) riportano la distribuzione di $P(p)$ nel caso in cui $\mu_x = 0$, $\mu_y = 1$, $s_x = s_y = 1$ (blu), $\mu_x = 0$, $\mu_y = 3$, $s_x = s_y = 1$ (magenta) e $\mu_x = 2$, $\mu_y = 3$, $s_x = s_y = 1$ (rosso). Come si vede, la loro forma è abbastanza differente ...

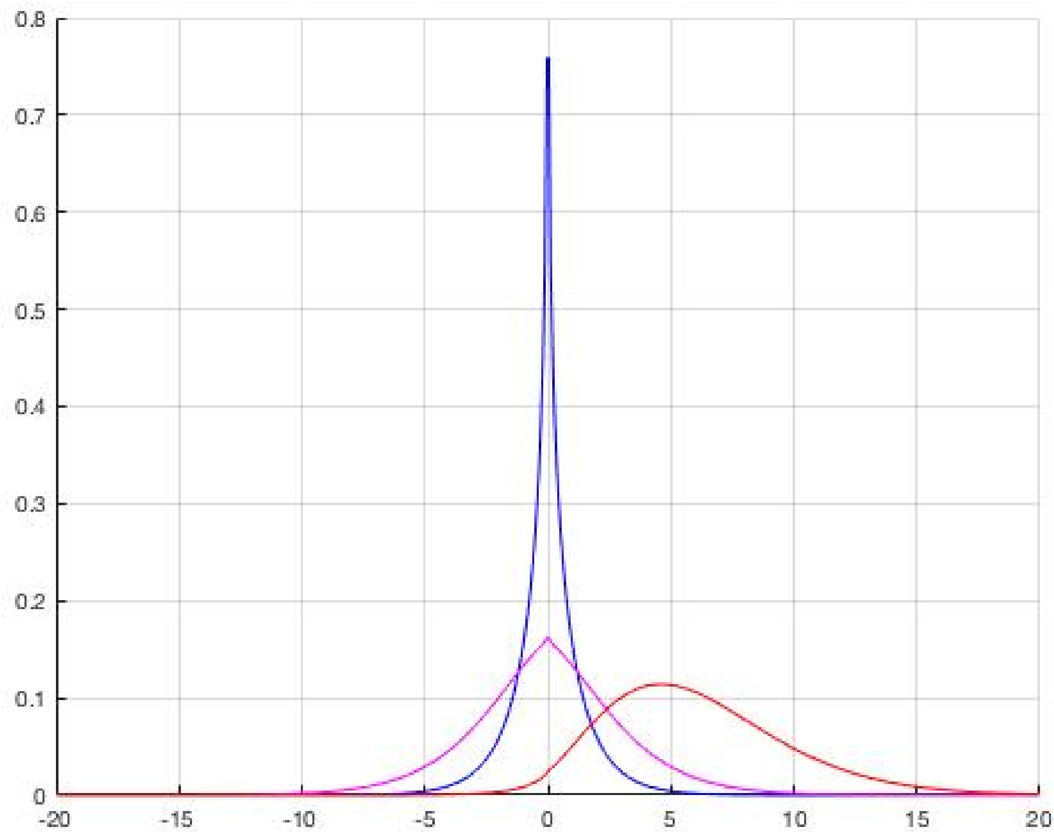


Figura 9:

2.6 Quoziente di due variabili casuali continue

Siano date due v.c. X e Y continue e indipendenti fra loro.
Definiamo la variabile casuale *rapporto* $R \equiv X \div Y$ come quella v.c. avente, in analogia con quanto già visto in precedenza, la pdf così fatta

$$R(r) \equiv \int dx\, dy\, X(x) \cdot Y(y)\; \delta\left(r - \frac{x}{y}\right) \tag{148}$$

Iniziamo verificando che la definizione (148) rispetta la condizione di normalizzazione, infatti si ha

$$\begin{aligned} \int dr\, R(r) = \int dr \int dx\, dy\, X(x) \cdot Y(y)\; \delta\left(r - \frac{x}{y}\right) = \int dx\, dy\, X(x) \cdot Y(y) = \\ = \int dx X(x) \cdot \int dy\, Y(y) = 1 \end{aligned} \tag{149}$$

Veniamo ora alla determinazione del valor medio associato a $R(r)$. Risulta

$$\begin{aligned} < R > \;&\equiv\; \int dr\, R(r)\, r = \int dr\, r \int dx\, dy\, X(x) \cdot Y(y)\; \delta\left(r - \frac{x}{y}\right) = \\ &= \int dx\, dy\, X(x) \cdot Y(y)\; \frac{x}{y} = \int dx X(x)\, x \cdot \int dy\, Y(y) \frac{1}{y} = \\ &= \; < X > \int dy\, Y(y) \frac{1}{y} \end{aligned} \tag{150}$$

Poiché, in generale l'integrale $\int dy\, Y(y) \frac{1}{y}$ **non** è uguale a $\frac{1}{<Y>}$, ecco che il valor medio della variabile casuale R non è il rapporto dei valori medi delle due variabili casuali X e Y.

Iniziamo osservando che, affinché R possa essere definita, è necessario che il dominio della v.c. Y, definito come l'insieme dei valori y per cui $Y(y) \neq 0$, non contenga lo zero, o, in altri termini, che $Y(0) = 0$.
Come primo esempio, iniziamo di nuovo da due v.c. piatte.
Supponiamo dunque che

$$X(x) \;=\; \frac{1}{L_a} \;\; per \;\; -a \leq x \leq -a + L_a; \qquad X(x) = 0 \;\; altrove \tag{151}$$

$$Y(y) \;=\; \frac{1}{L_b} \;\; per \;\; b \leq y \leq b + L_b; \qquad Y(y) = 0 \;\; altrove \tag{152}$$

dove assumeremo, per semplicità, che a, L_a, b, L_b siano quantità positive.
Esplicitiamo adesso la p.d.f. $R(r)$: abbiamo

$$R(r) \equiv \int dx \int dy\, X(x)\, Y(y)\, \delta\left(\frac{x}{y} - r\right) \tag{153}$$

Procediamo integrando nella variabile y, che si riferisce alla v.c. che compare al denominatore. Come sappiamo, detto y_0 il valore di y che annulla l'argomento della delta, ovvero posto $y_0 = \frac{x}{r}$ risulta

$$\begin{aligned} \delta\left(\frac{x}{y} - r\right) dy &= \frac{1}{\left|\frac{d\left(\frac{x}{y}\right)}{dy}\right|_{y=y_0}} \delta(y - y_0)\, dy = \frac{1}{\left|\frac{x}{y_0^2}\right|} \delta(y - y_0)\, dy = \\ &= \frac{|x|}{r^2}\, \delta(y - y_0)\, dy \end{aligned} \tag{154}$$

e dunque

$$\begin{aligned} R(r) &= \int dx \int X(x)\, Y(y)\, \frac{|x|}{r^2}\, \delta(y - y_0)\, dy = \int dx\ X(x)\, Y(y_0)\, \frac{|x|}{r^2} = \\ &= \frac{1}{r^2} \int dx\ X(x)\, Y\left(\frac{x}{r}\right) |x| \end{aligned} \tag{155}$$

Prima di procedere, consideriamo due casi limite.
Iniziamo assumendo che la v.c. Y al denominatore abbia media $a \neq 0$ e un r.m.s. trascurabilmente piccolo, talché si possa porre

$$Y(y) = \delta(y - a) \tag{156}$$

Questo corrisponde a dire che la Y è sempre costante e uguale ad a.
Dalla (155) si ottiene

$$R(r) = \frac{1}{r^2} \int dx\, X(x)\, \delta\left(\frac{x}{r} - a\right) |x| \tag{157}$$

Ma

$$\delta\left(\frac{x}{r} - a\right) dx = \delta(x - x_0)\, \frac{1}{\left|\frac{1}{r}\right|}\, dx \tag{158}$$

dove $x_0 = a\, r$. Dunque

$$R(r) = \frac{1}{r^2} \int dx\, X(x)\, \delta(x - x_0)\, |r|\, |a\, r| = |a|\, X(ar) \tag{159}$$

che è proprio quanto ci aspettiamo circa il prodotto della v.c. X per la costante $\frac{1}{a}$ (cfr(71)).

Più interessante è certamente il secondo caso che vogliamo considerare, cioè quello in cui è la v.c. X al numeratore ad avere dispersione nulla, ovvero

$$X(x) = \delta(x - a) \tag{160}$$

e supponiamo altresì che $a = 1$. Abbiamo

$$\begin{aligned} R(r) &= \int dx\, dy\, X(x)\, Y(y)\, \delta\left(r - \frac{x}{y}\right) = \int dx\, dy\, \delta(x - 1)\, Y(y)\, \delta\left(r - \frac{x}{y}\right) = \\ &= \int dy\ Y(y)\, \delta\left(r - \frac{1}{y}\right) \end{aligned} \tag{161}$$

Ma

$$\delta\left(r-\frac{1}{y}\right)\,dy=\delta(y-y_0)\,\frac{1}{\left|\frac{1}{y_0^2}\right|}\,dy \tag{162}$$

dove $y_0=\frac{1}{r}$. Abbiamo così che

$$\delta\left(r-\frac{1}{y}\right)\,dy=\delta\left(y-\frac{1}{r}\right)\,\frac{1}{r^2}\,dy \tag{163}$$

e quindi

$$R(r)=\int Y(y)\,\delta\left(y-\frac{1}{r}\right)\,\frac{1}{r^2}\,dy=\frac{1}{r^2}\,Y\left(\frac{1}{r}\right) \tag{164}$$

Questa p.d.f. definisce la v.c. che indicheremo con il simbolo Y^{-1}, tale che

$$Y^{-1}(z)\equiv\frac{1}{z^2}\,Y\left(\frac{1}{z}\right) \tag{165}$$

che è detta reciproca (inversa) della Y ed è facile dimostrare che

$$X\div Y=X\otimes Y^{-1} \tag{166}$$

Sotto è riportato un esempio di distribuzione piatta e sua inversa.

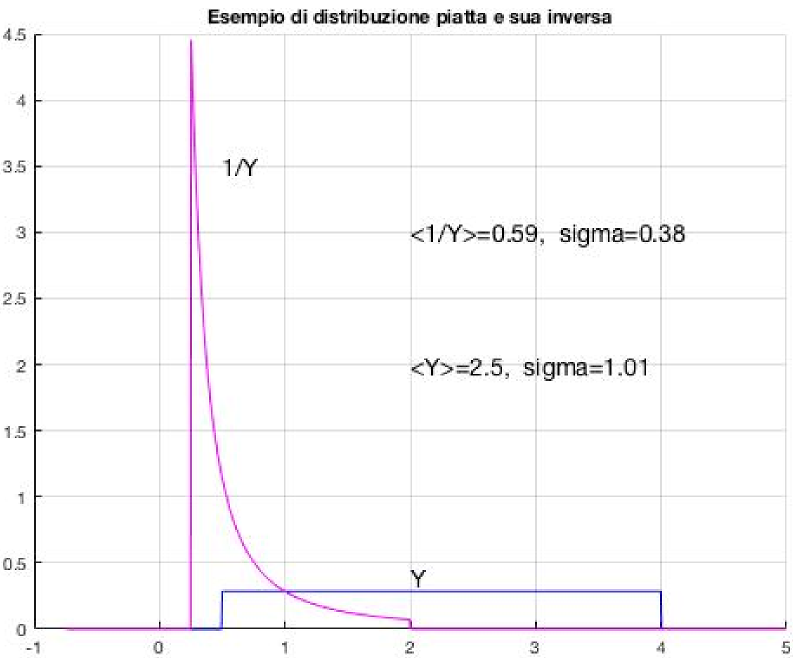


Figura 10:

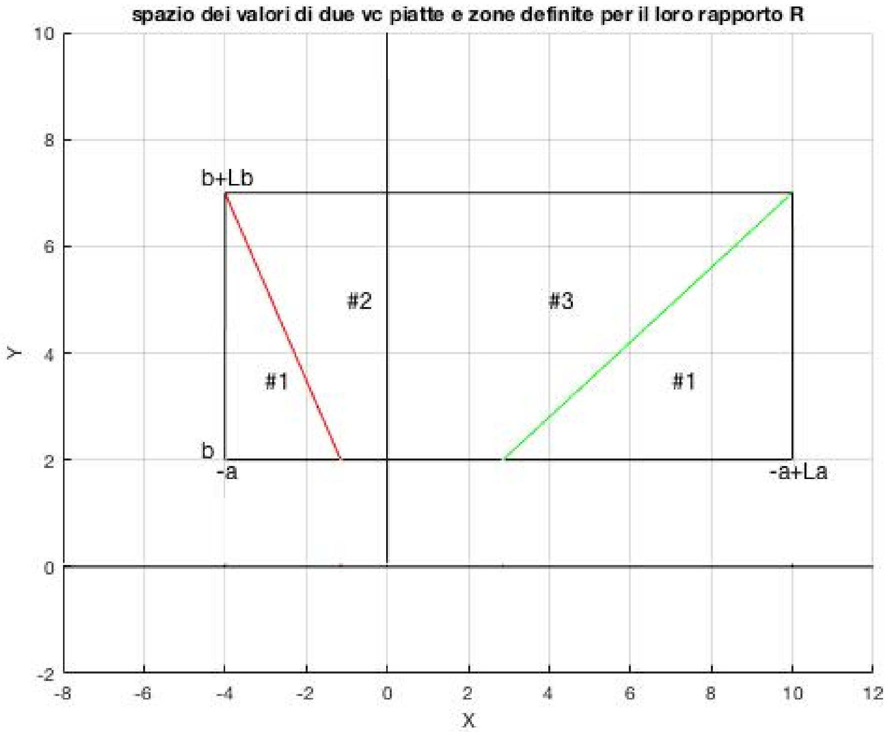


Figura 11:

Come esempio abbiamo scelto il caso in cui $a = 4$, $L_a = 14$; $b = 2$, $L_b = 5$: la figura (11) mostra lo spazio dei parametri utile, dove la funzione integranda non è nulla e vale $\frac{1}{L_a L_b}$. Essa mostra che la p.d.f. $R(r)$ sarà diversa da zero solo fra r_{min} ed r_{max}, dove

$$r_{min} = -\frac{a}{b}; \qquad\qquad r_{max} = \frac{L_a - a}{b} \tag{167}$$

Per fare il calcolo della $R(r)$ è opportuno dividere lo spazio dei parametri in quattro zone $\#1, ..., \#4$ come in figura, dove

$$r_1 = \frac{a}{b + L_b}; \qquad\qquad r_2 = \frac{L_a - a}{b + L_b} \tag{168}$$

Riportiamo di seguito la definizione delle varie zone, i limiti di integrazione nella variabile x e $R(r)$ (si ricordi che, nello spazio dei parametri, $X(x)\,Y\left(\frac{x}{r}\right)$

è costante e pari a $\frac{1}{L_a\, L_b}$), il cui plot, con quello di X e Y è riportato sotto.

$$\#1: r_{min} \leq r \leq r_1; \qquad b\, r_{min} \leq x \leq r\, b \qquad R(r) = \frac{a^2 - r^2 b^2}{2r^2\, L_a L_b} \tag{169}$$

$$\#2: r_1 \leq r \leq 0; \qquad r(b + L_b) \leq x \leq r\, b \qquad R(r) = \frac{L_b^2 + 2b\, L_b}{2r^2\, L_a L_b} \tag{170}$$

$$\#3: 0 \leq r \leq r_2; \qquad r\, b \leq x \leq r(b + L_b) \qquad R(r) = \frac{L_b^2 + 2b\, L_b}{2r^2\, L_a L_b} \tag{171}$$

$$\#4: r_2 \leq r \leq r_{max}; \qquad r\, b \leq x \leq r_{max}\, b \qquad R(r) = \frac{(L_a - a)^2 - r^2 b^2}{2r^2\, L_a L_b} \tag{172}$$

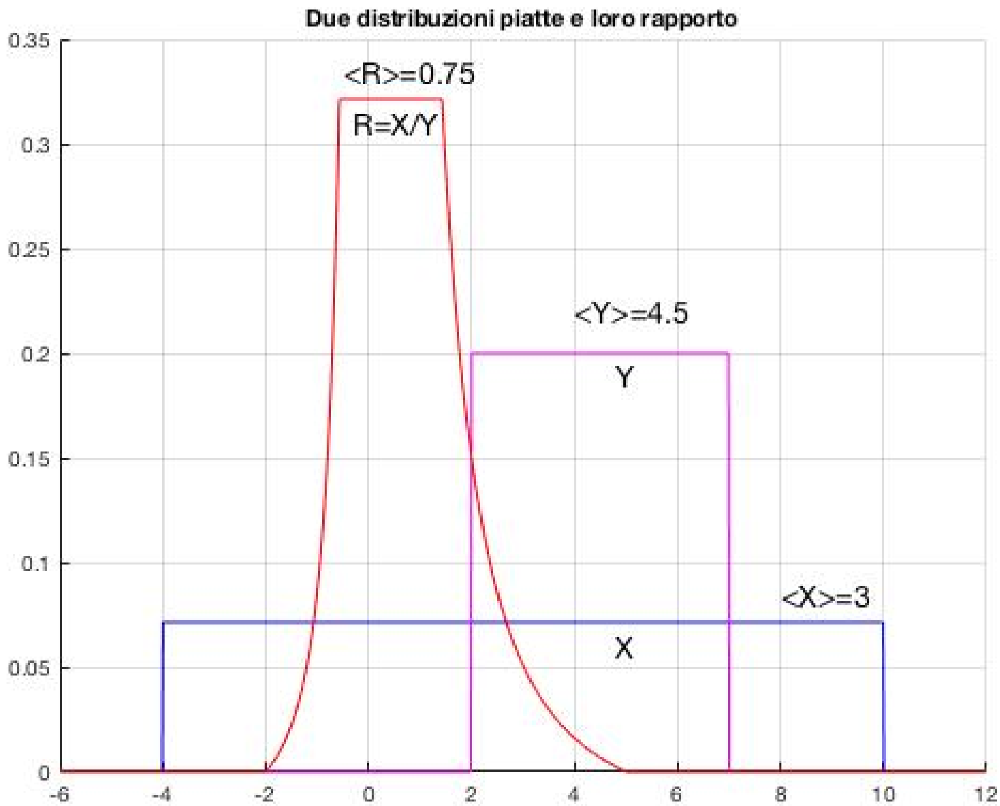


Figura 12:

Un altro caso interessante, che aiuta a capire il significato del rapporto fra distribuzioni di probabilità, è quello in cui entrambe le variabili casuali, pur essendo indipendenti fra loro, sono entrambe piatte e definite nello stesso dominio compreso fra $a > 0$ e $a + L > a$, ovvero tali che

$$X(x) = \frac{1}{L} \; per \;\; a \leq x \leq a + L; \quad X(x) = 0 \; altrove \tag{173}$$

$$Y(y) = \frac{1}{L} \; per \;\; a \leq y \leq a + L; \quad Y(y) = 0 \; altrove \tag{174}$$

Lo spazio dei parametri (x, y), dove le due v.c. sono non nulle, è rappresentato in figura (13) per $a = 2$ e $L = 5$.
Nella figura sono anche evidenziate le due zone per le quali

$$\#1 : \frac{a}{a+L} \equiv r_m \leq r \leq 1 : \quad a \leq x \leq r(a+L) \tag{175}$$

$$\#2 : 1 \leq r \leq \frac{a+L}{a} \equiv r_M : \quad ra \leq x \leq a+L \tag{176}$$

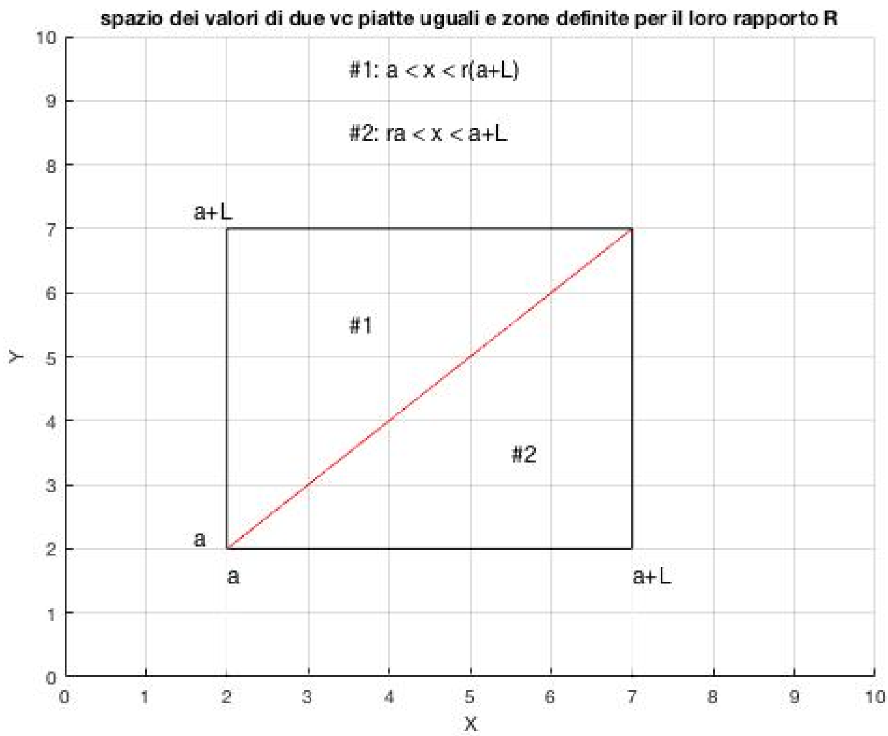


Figura 13:

Per quanto abbiamo già visto (cfr. eq.155), abbiamo

$$R(r) = \frac{1}{r^2} \int dx \; X(x) \, Y\left(\frac{x}{r}\right) \; |x| \tag{177}$$

e dunque

$$\frac{a}{a+L} \leq r \leq 1 : \qquad R(r) = \frac{1}{2L^2 r^2} \left[r^2 (a+L)^2 - a^2 \right] \tag{178}$$

$$1 \leq r \leq 1 \leq \frac{a+L}{a} : \qquad R(r) = \frac{1}{2L^2 r^2} \left[(a+L)^2 - a^2 r^2 \right] \tag{179}$$

La figura (14) seguente mostra sia le p.d.f. (coincidenti) delle due v.c. piatte, come la p.d.f. del loro rapporto.

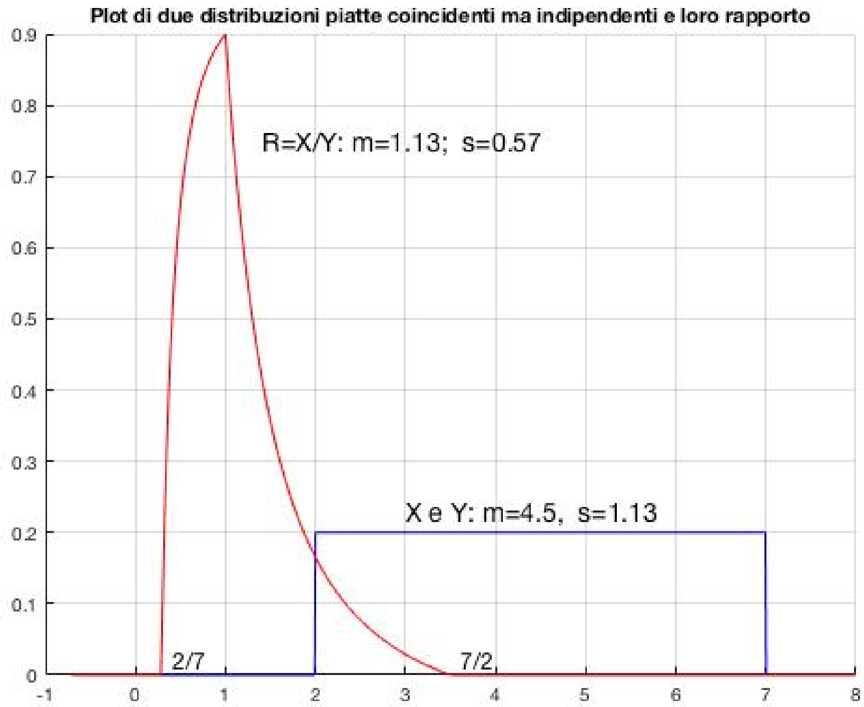


Figura 14:

Passiamo ora a considerare il caso in cui X **sia una v.c. gaussiana** mentre Y sia **una distribuzione piatta**, che assumeremo, come nel caso precedente, avente valori fra b e $b+L_b$ entrambi positivi. Dunque abbiamo

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \tag{180}$$

$$Y(y) = \frac{1}{L_b} \; per \;\; b \leq y \leq b+L_b; \qquad\qquad Y(y) = 0 \;\; altrove \tag{181}$$

mentre, come abbiamo visto con la (155), sarà

$$R(r) = \frac{1}{r^2} \int dx \; X(x)\, Y\left(\frac{x}{r}\right) \, |x| \tag{182}$$

dove l'integrazione sarà evidentemente limitata ai valori di x per cui $Y\left(\frac{x}{r}\right) \neq 0$, ovvero $b \leq \frac{x}{r} \leq b+L_b$.

Poiché il dominio dove la distribuzione G è non nulla è fatto da tutti i numeri reali, lo spazio dei parametri nel piano (x,y) è semplicemente costituito dalla striscia orizzontale definita da $-\infty < x < +\infty$; $\; b \leq y \leq b+L_b$ e

i possibili valori del rapporto r vanno anch'essi da $-\infty$ a $+\infty$.
Per un assegnato valore di r, i valori di x per cui la p.d.f. $Y\left(\frac{x}{r}\right)$ è diversa da zero (e uguale a $\frac{1}{L_b}$) sono i seguenti

$$\begin{aligned} r < 0: & \qquad r(b+L_b) \leq x \leq r\,b & (183)\\ r > 0: & \qquad r\,b \leq x \leq r(b+L_b) & (184)\end{aligned}$$

Partendo dunque dalla (182), abbiamo

$$\begin{aligned} r < 0: \quad R(r) &= -\frac{1}{r^2 L_b}\int_{r(b+L_b)}^{rb} dx\, x\, G(x) & (185)\\ r > 0: \quad R(r) &= \frac{1}{r^2 L_b}\int_{rb}^{r(b+L_b)} dx\, x\, G(x) & (186)\end{aligned}$$

D'altronde, ponendo al solito $\xi = x - \mu$ e $t = \frac{\xi}{\sigma\sqrt{2}}$, risulta

$$\begin{aligned}\int_\alpha^\beta dx\; x\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} &= \int_{\alpha-\mu}^{\beta-\mu} d\xi\;(\xi+\mu)\, e^{-\frac{\xi^2}{2\sigma^2}} = \int_{\alpha-\mu}^{\beta-\mu}\frac{d\xi}{\sigma\sqrt{2}}\,\sigma\sqrt{2}\left(\frac{\xi}{\sigma\sqrt{2}}\sigma\sqrt{2}+\mu\right)\, e^{-\frac{\xi^2}{2\sigma^2}} =\\ &= \sigma\sqrt{2}\int_{\frac{\alpha-\mu}{\sigma\sqrt{2}}}^{\frac{\beta-\mu}{\sigma\sqrt{2}}} dt\left(\sigma\sqrt{2}\;t+\mu\right)\, e^{-t^2} =\\ &= 2\sigma^2\int_{\frac{\alpha-\mu}{\sigma\sqrt{2}}}^{\frac{\beta-\mu}{\sigma\sqrt{2}}} dt\; e^{-t^2}\, t \;+\; \sigma\mu\sqrt{2}\int_{\frac{\alpha-\mu}{\sigma\sqrt{2}}}^{\frac{\beta-\mu}{\sigma\sqrt{2}}} dt\; e^{-t^2} & (187)\end{aligned}$$

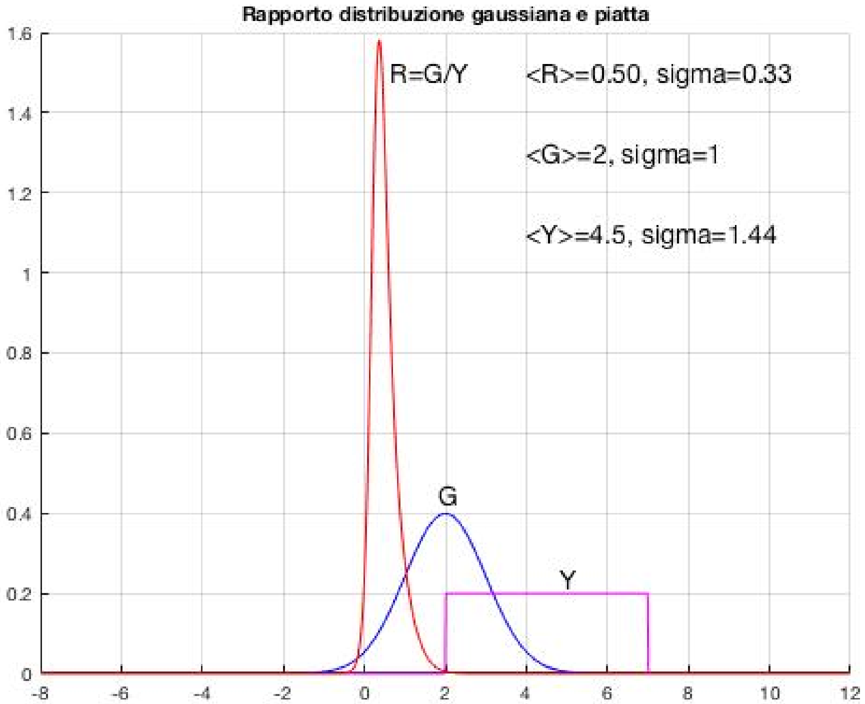
Ma

$$\begin{aligned}\int_A^B dt\, e^{-t^2}\, t &= \frac{1}{2}\left(e^{-A^2}-e^{-B^2}\right) & (188)\\ \int_A^B dt\, e^{-t^2} &= \frac{\sqrt{\pi}}{2}\left[erf(B)-erf(A)\right] & (189)\end{aligned}$$

per cui, in definitiva, otteniamo

$$\begin{aligned}\forall r \in R: R(r) &= \frac{\sigma}{r^2\, L_b\sqrt{2\pi}}\left(e^{-\left(\frac{rb-\mu}{\sigma\sqrt{2}}\right)^2} - e^{-\left(\frac{rb+rL_b-\mu}{\sigma\sqrt{2}}\right)^2}\right) +\\ &+ \frac{\mu}{2r^2 L_b}\left(erf\left(\frac{rb+rL_b-\mu}{\sigma\sqrt{2}}\right) - erf\left(\frac{rb-\mu}{\sigma\sqrt{2}}\right)\right) & (190)\end{aligned}$$

La figura riportata di seguito mostra sia le due p.d.f. di partenza G (gaussiana con media 2 e r.m.s. 1) e Y (distribuzione piatta fra 2 e 7), come il risultato del loro rapporto $R = G/Y$.


Rapporto distribuzione gaussiana e piatta
R=G/Y
<R>=0.50, sigma=0.33
<G>=2, sigma=1
<Y>=4.5, sigma=1.44
G
Y
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-8
-6
-4
-2
0
2
4
6
8
10
12
Figura 15:

2.7 Somma e prodotto di una v.c. discreta con una quantità reale

Abbiamo trattato il problema della somma e del prodotto con una quantità reale di una variabile casuale continua. Vediamo adesso se e cosa cambia se la v.c. è discreta.
Iniziamo dalla **somma**.
Sia dunque $\Omega_X \equiv \{x_k; k = 1, ..., n\}$ il dominio di una variabile casuale discreta X e sia dunque $X(x_n)$ la distribuzione di probabilità (normalizzata) a essa associata. Sulla base di quanto visto nel caso di variabili continue, la somma Y della v.c. X con una costante reale a sarà tale che

$$Y(y_n) \equiv X(y_n - a) \;\; con \;\; y_n = x_n + a \;\; e \;\; x_n \in \Omega_X \tag{191}$$

Dalla definizione abbiamo quindi che il dominio Ω_Y della variabile traslata Y è fatto dagli elementi di Ω_X, traslati, ciascuno, della costante a. Il fatto che il dominio delle due v.c. non resti generalmente il solito, appare, in questo caso, molto evidente. Però, anche nel caso delle v.c. continue, a meno che il dominio di X non fosse coinciso con l'intera retta reale, era $\Omega_Y \neq \Omega_X$, poiché $\Omega_Y = \{y : \; y = x + a; \; x \in \Omega_X\}$.
Tornando alla (191), è immediato che $Y(y_n)$ definisce sul dominio Ω_Y una distribuzione di probabilità che risulta normalizzata.
Quanto al valor medio di Y, abbiamo

$$\begin{aligned} < Y > &= \sum_{\Omega_Y} Y(y_n)\, y_n = \sum_n Y(y_n) \cdot y_n = \sum_n X(y_n - a) \cdot y_n = \\ &= \sum_n X(y_n - a) \cdot [(y_n - a) + a] = \sum_n X(x_n) \cdot [x_n + a] = \\ &= < X > +a \end{aligned} \tag{192}$$

Analogamente per la varianza si ha

$$\begin{aligned} < Y^2 > &= \sum_{\Omega_Y} Y(y_n)\, y_n^2 = \sum_n Y(y_n) \cdot y_n^2 = \sum_n X(y_n - a) \cdot y_n^2 = \\ &= \sum_n X(y_n - a) \cdot [(y_n - a) + a]^2 = \sum_n X(x_n) \cdot \left[x_n^2 + 2ax_n + a^2\right] = \\ &= < X^2 > +2a < X > +a^2 \end{aligned} \tag{193}$$

e dunque

$$\begin{aligned} \sigma_Y^2 &= < Y^2 > - < Y >^2 = < X^2 > +2a < X > +a^2 - (< X > +a)^2 = \\ &= < X^2 > - < X >^2 = \sigma_X^2 \end{aligned} \tag{194}$$

ovvero ritroviamo che la varianza della v.c. traslata Y coincide con quella della v.c. X da cui siamo partiti.

Passiamo adesso a considerare il **prodotto** Y di una v.c. X per una costante reale arbitraria $a \neq 0$. Poniamo per questo

$$Y(y_n) \equiv X\left(\frac{y_n}{a}\right) \quad con \;\; y_n = a\,x_n \;\; e \;\; x_n \in \Omega_X \tag{195}$$

Evidentemente, quanto ai dominii, si ha

$$\Omega_Y = \{y_n : \; y_n = a\,x_n, \; x_n \in \Omega_X\} \tag{196}$$

e dunque Ω_Y risulta essere uguale a Ω_x, *scalato* della quantità a.
Si osservi, a questo riguardo, che a può anche essere negativa: nel caso in cui $a = -1$ avremo che[14]

$$Y = -X \;\;\Rightarrow\;\; Y(y_n) \equiv (-X)(y_n) = X(-x_n) \tag{197}$$

e dunque che il dominio di $-X$ è costituito dall'insieme degli opposti degli elementi costituenti il dominio di X. Poiché

$$\sum_n Y(y_n) = \sum_n \left(\frac{y_n}{a}\right) = \sum_n X(x_n) \tag{198}$$

è evidente come la normalizzazione della funzione $X(x_n)$ implichi quella della $Y(y_n)$. Veniamo alla determinazione del valor medio. Si ha

$$\begin{aligned} < Y > \;&\equiv\; \sum_n Y(y_n)\cdot y_n = \sum_n X\left(\frac{y_n}{a}\right)\cdot y_n = \sum_n X(x_n)\cdot a\,x_n = \\ &= \; a < X > \end{aligned} \tag{199}$$

Quanto alla varianza. abbiamo che

$$\begin{aligned} < Y^2 > \;&\equiv\; \sum_n Y(y_n)\cdot y_n^2 = \sum_n X\left(\frac{y_n}{a}\right)\cdot y_n^2 = \sum_n X(x_n)\cdot (a\,x_n)^2 = \\ &= \; a^2 < X^2 > \end{aligned} \tag{200}$$

e dunque

$$\begin{aligned} \sigma_Y^2 \;&=\; < Y^2 > - < Y >^2 = a^2 < X^2 > -(a < X >)^2 = a^2(< X^2 > - < X >^2) = \\ &= \; a^2\sigma_X^2 \end{aligned} \tag{201}$$

nel caso del prodotto, la media $< Y >$ scala proporzionalmente alla costante a e lo scarto quadratico medio σ_Y scala con $|a|$.

[14]Deve essere ben chiaro che i valori della v.c. $Y = -X$, essendo probabilità, sono comunque quantità non negative: il cambiamento di segno riguarda il punto dove la funzione di probabilità viene calcolata.

2.8 Somma di due v.c. discrete

Veniamo adesso a trattare la somma di due (o più) variabili casuali. Non considereremo il caso generale perché non molto utile, ma ci occuperemo invece di alcuni casi specifici.
Cominciamo dal caso in cui la v.c. X abbia come dominio $\Omega = \{0, 1\}$, dove 0 corrisponde a *insuccesso* e ha probabilità di verificarsi pari a $q > 0$, mentre 1 corrisponde a *successo*, con probabilità di accadere $p > 0$ (ovviamente, essendo i due casi mutuamente esclusivi, è $p+q=1$). Questa v.c. potrebbe, per esempio, rappresentare da un punto di vista probabilistico l'esito del lancio di una moneta, con $0 = \textit{testa}$ e $1 = \textit{croce}$ (se la moneta non ha asimmetrie, allora $p = q = 0.5$).
Sia adesso[15]

$$X_2 = X \oplus X \tag{202}$$

dove, per ipotesi, le due funzioni di probabilità $X(x_n)$ si riferiscono a test diversi e indipendenti.
Evidentemente, quanto al dominio di X_2, abbiamo $\Omega_2 = \{0,\ 1,\ 2\}$ e i tre valori corrispondono ai seguenti risultati dei due test indipendenti che definiscono X_2:

$$\begin{aligned} (0,0) \quad &\Rightarrow \quad 0; \\ (1,0)\, o\, (0,1) \quad &\Rightarrow \quad 1; \\ (1,1) \quad &\Rightarrow \quad 2 \end{aligned}$$

Evidentemente il valore 0 si può ottenere solo se entrambi i test danno luogo al risultato 0 e dunque con probabilità $X_2(0) = q^2$, il valore 1 si ottiene se uno e uno solo dei due test dà risultato 1, ovvero in due casi, ciascuno con probabilità pq, da cui $X_2(1) = 2pq$, infine il valore 2 si otterrà solo se entrambi i test forniscono il risultato 1 e dunque con probabilità $X_2(2) = p^2$.
In conclusione abbiamo che

$$\textit{per } k = 0, 1, 2, 3 : \ X_2(k) = p^k\, q^{2-k} \begin{pmatrix} 2 \\ k \end{pmatrix} \tag{203}$$

Abbiamo dunque che X_2 risulta essere una binomiale definita dalla probabilità di *successo* uguale a p e da $n = 2$. Si osservi che anche la distribuzione di partenza X è una binomiale (la più semplice possibile ...), con probabilità di *successo* pari a p e $n = 1$. Quello che abbiamo trovato è che la somma di due binomiali con la stessa probabilità di successo p ed entrambe con $n = 1$, risulta una binomiale con probabilità di successo p e $n = 2$.

Questo risultato è un caso particolare del caso generale della somma di binomiali, che vediamo di seguito.

[15] Va da sé che $X_2 \neq 2X$ ...

2.9 Somma di distribuzioni binomiali con la stessa p

Supponiamo che siano date le due distribuzioni binomiali indipendenti seguenti

$$B_1(k) = p^k q^{n-k} \begin{pmatrix} n \\ k \end{pmatrix}; \quad \Omega_1 = \{0, 1, ..., n\} \tag{204}$$

$$B_2(j) = p^j q^{m-j} \begin{pmatrix} m \\ j \end{pmatrix}; \quad \Omega_2 = \{0, 1, ..., m\} \tag{205}$$

Dati i due dominii Ω_1 e Ω_2, la somma delle due distribuzioni $B = B_1 \oplus B_2$ avrà come dominio $\Omega = \{0, 1, ..., m+n\}$, poichè dalla prima distribuzione potremo ricavare un numero di *successi* compreso fra 0 ed n, dalla seconda fra 0 e m e quindi, sommando le due, i *successi* potranno andare da 0 a $m+n$. Però il generico numero di *successi* s, con $0 \leq s \leq n+m$, in genere, potrà essere ottenuto con varie combinazioni di *successi* da B_1 e B_2, ciascuna con la propria probabilità. Più esplicitamente avremo che la probabilità di ottenere s successi con B sarà

$$B(s) = \sum_{k=0}^{n} \sum_{j=0}^{m} B_1(n,k) \cdot B_2(m,j) \cdot \delta_{s,k+j} \tag{206}$$

dove $\delta_{r,t}$ è la funzione delta di Kronecker, che vale 1 se i due indici coincidono e vale 0 negli altri casi. Dunque

$$\begin{aligned} B(s) &= \sum_{k=0}^{n} \sum_{j=0}^{m} p^k q^{n-k} \begin{pmatrix} n \\ k \end{pmatrix} \cdot p^j q^{m-j} \begin{pmatrix} m \\ j \end{pmatrix} \cdot \delta_{s,k+j} = \\ &= \sum_{k=0}^{n} \sum_{j=0}^{m} p^{k+j} q^{n+m-k-j} \begin{pmatrix} n \\ k \end{pmatrix} \begin{pmatrix} m \\ j \end{pmatrix} \cdot \delta_{s,k+j} = \\ &= \sum_{s=0}^{m+n} p^s q^{n+m-s} \sum_{k=0}^{n} \sum_{j=0}^{m} \begin{pmatrix} n \\ k \end{pmatrix} \begin{pmatrix} m \\ j \end{pmatrix} \cdot \delta_{s,k+j} \end{aligned} \tag{207}$$

Ma

$$\sum_{k=0}^{n} \sum_{j=0}^{m} \begin{pmatrix} n \\ k \end{pmatrix} \begin{pmatrix} m \\ j \end{pmatrix} \cdot \delta_{s,k+j} = \begin{pmatrix} n+m \\ s \end{pmatrix} \tag{208}$$

Infatti, se consideriamo il seguente prodotto di potenze del binomio $(1+x)$

$$(1+x)^{n+m} = (1+x)^n (1+x)^m \tag{209}$$

si ha

$$(1+x)^{n+m} = \sum_{s=0}^{n+m} x^s \binom{n+m}{s} \tag{210}$$

$$= (1+x)^n(1+x)^m = \sum_{k=0}^{n} x^k \binom{n}{k} \cdot \sum_{j=0}^{m} x^j \binom{m}{j} =$$

$$= \sum_{k=0}^{n} \sum_{j=0}^{m} x^{k+j} \binom{n}{k} \cdot \binom{m}{j} \tag{211}$$

e, raccogliendo a fattor comune i termini a cui corrisponde la stessa potenza di x, abbiamo

$$(1+x)^n(1+x)^m = \sum_{s=0}^{n+m} x^s \sum_{k=0}^{n} \sum_{j=0}^{m} \binom{n}{k} \cdot \binom{m}{j} \cdot \delta_{s,k+j} \tag{212}$$

Perciò, per le note proprietà dei polinomi, il confronto fra la (210) e la (212), consente di poterci far concludere che

$$\sum_{k=0}^{n} \sum_{j=0}^{m} \binom{n}{k} \cdot \binom{m}{j} \cdot \delta_{s,k+j} = \binom{n+m}{s} \tag{213}$$

Dunque, tornando alla distribuzione $B(s)$ di cui alla (207), abbiamo infine che

$$B(s) = \sum_{s=0}^{m+n} p^s q^{n+m-s} \binom{n+m}{s} \tag{214}$$

ovvero resta così dimostrato che la somma di due distribuzioni binomiali definite dalla *stessa* probabilità di singolo *successo* p, relative a un numero di casi uguali, rispettivamente a n e a m, risulta essere una binomiale definita dalla stessa probabilità p, riferita a un numero di casi uguale a $n+m$.
Evidentemente questo si traduce anche nel fatto che, essendo

$$< B_1 > = np; \quad \sigma_1^2 = npq \tag{215}$$

$$< B_2 > = mp; \quad \sigma_2^2 = mpq \tag{216}$$

risulta, quanto alla *somma*, che

$$< B > = (n+m)p =< B_1 > + < B_2 > \tag{217}$$

$$\sigma_B^2 = (n+m)pq = \sigma_1^2 + \sigma_2^2 \tag{218}$$

2.10 Somma di distribuzioni di Poisson

Consideriamo adesso due distribuzioni di Poisson P_1 e P_2, indipendenti fra loro e caratterizzate dal valor medio μ e ν, rispettivamente. Per ipotesi, abbiamo

$$
\begin{aligned}
P_1(k) &= e^{\mu}\,\frac{\mu^k}{k!} && (219)\\
P_2(j) &= e^{\nu}\,\frac{\nu^j}{j!} && (220)
\end{aligned}
$$

Per definizione entrambe hanno come dominio l'insieme dei numeri naturali compreso lo zero. La loro somma $P = P_1 \oplus P_2$ avrà, naturalmente, lo stesso dominio essendo l'insieme dei numeri naturali invariante sotto la somma.
Al valore s della somma, corrisponderà una probabilità $P(s)$ legata alle probabilità con cui s può essere ottenuta attraverso P_1 e P_2.
Evidentemente avremo che

$$
\begin{aligned}
P(s) &= \sum_{r=0}^{s} P_1(r)\,P_2(s-r) = e^{-\mu}\,e^{-\nu}\sum_{r=0}^{s}\frac{\mu^r}{r!}\,\frac{\nu^{s-r}}{(s-r)!} = \\
&= e^{-(\mu+\nu)}\,\frac{1}{s!}\sum_{r=0}^{s}\mu^r\,\nu^{s-r}\,\frac{s!}{r!(s-r)!} = e^{-(\mu+\nu)}\,\frac{1}{s!}\sum_{r=0}^{s}\mu^r\,\nu^{s-r}\begin{pmatrix} s \\ r \end{pmatrix} = \\
&= e^{-(\mu+\nu)}\,\frac{(\mu+\nu)^s}{s!} && (221)
\end{aligned}
$$

Ovvero la somma P di due poissoniane P_1 e P_2 è anch'essa una poissoniana e risulta

$$
\begin{aligned}
< P > &= < P_1 > + < P_2 > = \mu + \nu && (222)\\
\sigma_P^2 &= \sigma_1^2 + \sigma_2^2 = \mu + \nu && (223)
\end{aligned}
$$

2.11 Binomiale di una binomiale ...

Consideriamo il processo di decadimento radioattivo di una sostanza a vita media τ molto lunga rispetto al tempo di osservazione T.
Indichiamo con N il numero dei radionuclidi costituenti la sostanza al tempo $t = 0$: ogni radionuclide avrà una probabilità $p = \frac{T}{\tau} << 1$ di decadere durante la finestra temporale di osservazione T e, per ciascuno di essi, quindi, p rappresenta quella che altrove abbiamo chiamato *probabilità di successo*, mentre $q = 1-p$ è la probabilità che, nel tempo T, il radionuclide sia rimasto stabile. Poichè i vari processi di decadimento sono indipendenti fra loro, il test, nel tempo T, di *successo* o meno relativo agli N radionuclidi conduce a una probabilità che nel tempo T avvengano k decadimenti, la quale è data dalla distribuzione binomiale (rappresentata di seguito con le sue possibili approssimazioni), ovvero

$$\begin{aligned} P_e(k|N,p) &= p^k\,(1-p)^{N-k} \begin{pmatrix} N \\ k \end{pmatrix} && (224) \\ &\approx e^{-\mu}\,\frac{\mu^k}{k!} && (225) \\ &\approx \frac{1}{\sqrt{2\pi\mu}}\,e^{-\frac{(k-\mu)^2}{2\mu}} && (226) \end{aligned}$$

dove $\mu \equiv Np = \frac{NT}{\tau}$ e $0 \leq k \leq N$.
Ma supponiamo adesso che l'osservazione dell'avvenuto decadimento abbia una efficienza $0 < \epsilon < 1$.
Quale sarà la distribuzione dei decadimenti che *osserveremo* ?
Nella finestra temporale T avverranno k decadimenti con la probabilità $P_e(k|N,p)$ data dalla (224). Ognuno di questi decadimenti, però, ha una probabilità di poter essere realmente *visto* che è pari a ϵ, dunque, dei k decadimenti avvenuti noi ne osserveremo solo un numero $s \leq k$, con una probabilità $P_v(s|k,\epsilon)$ data da

$$P_v(s|k,\epsilon) = \epsilon^s\,(1-\epsilon)^{k-s} \begin{pmatrix} k \\ s \end{pmatrix} \qquad (227)$$

La probabilità di osservare s decadimenti *a prescindere da quanti ne sono realmente avvenuti*, sarà quindi pari a

$$P_o(s|N,p,\epsilon) = \sum_{k\geq s}^{N} P_e(k|N,p) \times P_v(s|k,\epsilon) \qquad (228)$$

Prima di procedere al calcolo dell'espressione di cui sopra, vediamo di capire che cosa dobbiamo aspettarci circa questa probabilità.
Ripartendo dunque dall'inizio, possiamo dire che ognuno degli N radionuclidi ha, nel tempo T, una probabilità $p = \frac{T}{\tau}$ di decadere e ogni decadimento

avvenuto, ha una probabilità ϵ di essere realmente osservato.
Possiamo dunque concludere che la probabilità, per ciascauno degli N radionuclidi della sostanza radioattiva, di dar luogo, nel tempo T, a un decadimento *osservato* sarà uguale a $p\,\epsilon$.
Ci aspettiamo perciò che la probabilità di *osservare* un numero s di decadimenti nel tempo T sia data dalla binomiale seguente

$$P_{out}(s|N, p\epsilon) = (p\epsilon)^s \, (1-p\epsilon)^{N-s} \begin{pmatrix} N \\ s \end{pmatrix} \tag{229}$$

Vediamo. Ripartiamo dunque dalla (228). Abbiamo

$$\begin{aligned} P_o(s|N,p,\epsilon) &= \sum_{k\geq s}^{N} P_e(k|N,p) \times P_v(s|k,\epsilon) = \\ &= \sum_{k\geq s}^{N} p^k \, (1-p)^{N-k} \begin{pmatrix} N \\ k \end{pmatrix} \times \epsilon^s \, (1-\epsilon)^{k-s} \begin{pmatrix} k \\ s \end{pmatrix} = \\ &= \sum_{k\geq s}^{N} p^k \, (1-p)^{N-k} \; \epsilon^s \, (1-\epsilon)^{k-s} \, \frac{N!}{k!(N-k)!} \, \frac{k!}{s!(k-s)!} = \\ &= \frac{N!}{s!} \sum_{k\geq s}^{N} p^k \, (1-p)^{N-k} \; \epsilon^s \, (1-\epsilon)^{k-s} \, \frac{1}{(N-k)!(k-s)!} \end{aligned} \tag{230}$$

Poniamo adesso $k = s + j$: la condizione di somma $N \geq k \geq s$ implica dunque, quanto all'indice j, che $N - s \geq j \geq 0$. Si ha dunque che

$$\begin{aligned} P_o(s|N,p,\epsilon) &= \frac{N!}{s!} \sum_{k\geq s}^{N} p^k \, (1-p)^{N-k} \; \epsilon^s \, (1-\epsilon)^{k-s} \, \frac{1}{(N-k)!(k-s)!} = \\ &= \frac{N!}{s!} \sum_{j\geq s}^{N-s} p^s \, p^j \, (1-p)^{N-s} \, (1-p)^{-j} \; \epsilon^s \, (1-\epsilon)^j \; \frac{1}{(N-s-j)!j!} = \\ &= (p\,\epsilon)^s \, (1-p)^{N-s} \frac{N!}{s!} \sum_{j\geq s}^{N-s} p^j (1-p)^{-j} \, (1-\epsilon)^j \, \frac{1}{(N-s-j)!j!} = \\ &= (p\,\epsilon)^s \, (1-p)^{N-s} \frac{N!}{s!} \sum_{j\geq s}^{N-s} \left(\frac{p(1-\epsilon)}{1-p}\right)^j \, \frac{1}{(N-s-j)!j!} = \\ &= (p\,\epsilon)^s \, (1-p)^{N-s} \frac{N!}{(N-s)!\,s!} \sum_{j\geq s}^{N-s} \left(\frac{p(1-\epsilon)}{1-p}\right)^j \, \frac{(N-s)!}{(N-s-j)!j!} = \\ &= (p\,\epsilon)^s \, (1-p)^{N-s} \frac{N!}{(N-s)!\,s!} \sum_{j\geq s}^{N-s} \left(\frac{p(1-\epsilon)}{1-p}\right)^j \begin{pmatrix} N-s \\ j \end{pmatrix} \end{aligned} \tag{231}$$

Ma, dallo sviluppo della potenza del binomio, sappiamo che

$$(1+x)^n = \sum_{j=0}^{n} x^j \left(\begin{array}{c} n \\ j \end{array} \right) \tag{232}$$

e dunque

$$\begin{aligned} P_o(s|N,p,\epsilon) &= (p\,\epsilon)^s\,(1-p)^{N-s}\frac{N!}{(N-s)!\,s!}\,\sum_{j\geq s}^{N-s}\left(\frac{p(1-\epsilon)}{1-p}\right)^j\left(\begin{array}{c} N-s \\ j \end{array} \right) = \\ &= (p\,\epsilon)^s\,(1-p)^{N-s}\frac{N!}{(N-s)!\,s!}\,\left(1+\frac{p(1-\epsilon)}{1-p}\right)^{N-s} = \\ &= (p\,\epsilon)^s\,(1-p)^{N-s}\,\left(\frac{1-p\epsilon}{1-p}\right)^{N-s}\,\frac{N!}{(N-s)!\,s!} = \\ &= (p\,\epsilon)^s\,(1-p\,\epsilon)^{N-s}\left(\begin{array}{c} N \\ s \end{array} \right) \end{aligned} \tag{233}$$

in accordo con la previsione (229).

Questo risultato ci dice anche che, in termini di distribuzioni di Poisson, se la distribuzione dei decadimenti *avvenuti* nel tempo di osservazione T, visto che $p << 1$ e $N >> \frac{1}{p}$, è descrivibile con una poissoniana di media $\mu_a = pN$, allora è possibile descrivere anche la distribuzione dei decadimenti *osservati* con una poissoniana di media $\mu_o = \mu_a\epsilon = p\epsilon N$.

2.12 Stima dell'inverso di una poissoniana

Sia data una poissoniana avente media μ, per la quale è dunque

$$P(n) = e^{-\mu}\,\frac{\mu^n}{n!} \tag{234}$$

Abbiamo visto che se moltiplichiamo questa v.c. per una costante reale non nulla, oppure se le sommiamo una quantità reale arbitraria, il suo valore di aspettaqzione definito a partire dai possibili valori della distribuzione nel dominio trasformato, cambia in modo lineare.
Ma che succede se ci interessa il valor medio della distribuzione inversa ?
in altri termini, quanto vale $< P^{-1} >$ se la distribuzione è la (234) ?
Quanto è diverso questo valore da $\frac{1}{<P>} \equiv \frac{1}{\mu}$?
Procediamo nel calcolo di P^{-1}, escludendo naturalmente il caso $n = 0$, per cui sarà

$$< P^{-1} > \equiv \sum_{n=1}^{+\infty} P(n) \cdot \frac{1}{n} = \sum_{n=1}^{+\infty} e^{-\mu}\,\frac{\mu^n}{n!}\,\frac{1}{n} \tag{235}$$

Come riportato a pag.229 dell'Handbook of Mathematical Functions di M.Abramowitz e I.A.Stegun, quanto alla serie abbiamo

$$Ei(x) = \gamma + \log(x) + \sum_{n=1}^{+\infty} \frac{x^n}{n!\,n}, \qquad\qquad per\ x > 0 \tag{236}$$

dove γ è la costante di Eulero Mascheroni ($\gamma = 0.5772156649...$) e la funzione $Ei(x)$ è la *exponential integral*, definita per $x > 0$ (attraverso la parte principale di Cauchy nello zero) come

$$Ei(x) = -\,⨍_{-x}^{+\infty} dt\,\frac{e^{-t}}{t} = ⨍_{-\infty}^{x} dt\,\frac{e^{t}}{t} \tag{237}$$

Abbiamo quindi che

$$< P^{-1} > = e^{-\mu}\,[Ei(\mu) - \gamma - \log\mu] \tag{238}$$

La figura (16) mostra il confronto fra questo valor medio (in verde) e $\frac{1}{\mu}$ (in rosso). Come si può vedere, per $x > 2$ risulta che $< P^{-1} >$ è sempre maggiore di $\frac{1}{\mu}$ e si può verificare che per $\mu > 10$, risulta

$$\frac{< P^{-1} > - \frac{1}{\mu}}{\frac{1}{\mu}} \approx \frac{1}{\mu} \tag{239}$$

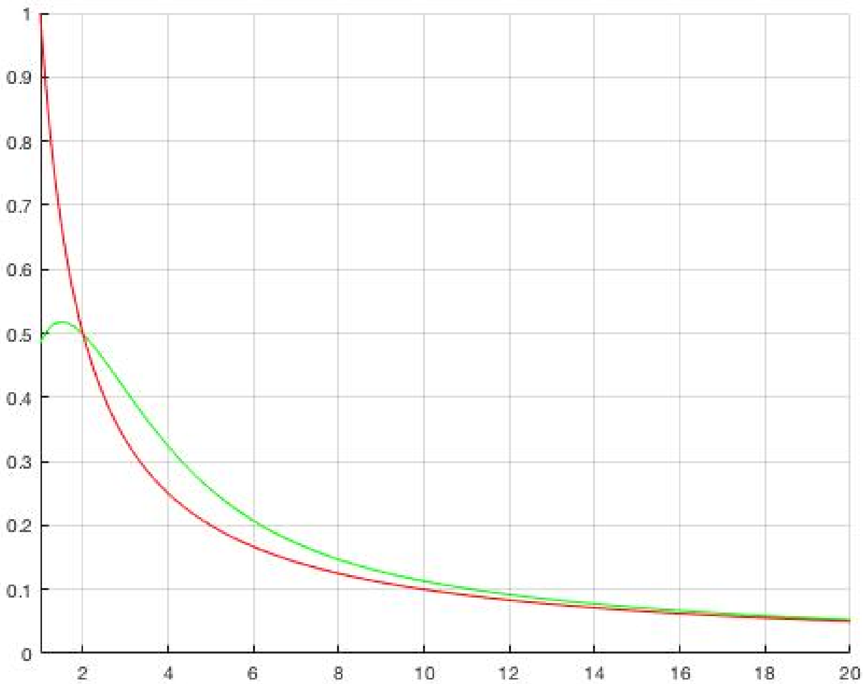


Figura 16:

Ma perché ci stiamo ponendo il problema del confronto fra $< P^{-1} >$ e $\frac{1}{<P>}$? Lo chiariremo adesso trattando un caso concreto.
Nell'esperimento $NA48$ del CERN si trattava di determinare l'esistenza o meno della violazione diretta della simmetria CP nel decadimento dei K neutri attraverso la misura del doppio rapporto

$$R = \frac{\Gamma(K_L \to \pi^0\pi^0)}{\Gamma(K_S \to \pi^0\pi^0)} \frac{\Gamma(K_S \to \pi^+\pi^-)}{\Gamma(K_L \to \pi^+\pi^-)} = \tag{240}$$

legato alla violazione diretta secondo la relazione

$$\Re\left(\frac{\epsilon'}{\epsilon}\right) = \frac{1-R}{6} \tag{241}$$

Per aver ragione della possibilità che le condizioni sperimentali mutino nel tempo, i quattro canali di decadimento venivano collezionati insieme. In questo modo i flussi legati al fascio di protoni incidente sulla targhetta, alle efficienze di identificazione dei K_S e K_L si cancellavano immediatamente, così come le inefficienze del detector, del trigger e le perdite dovute ad accidentali. Ogni burst di protoni costituiva quindi una misura del parametro di interesse attraverso il doppio rapporto del numero dei conteggi associati ai quattro tipi di decadimento. Però, da quanto precede risulta che, se P_1 e P_2 descrivono due distribuzioni poissoniane con media μ_1 e μ_2 rispettivamente, allora

$$< P_1 \div P_2 > = < P_1 \otimes P_2^{-1} > = < P_1 > < P_2^{-1} > \neq \frac{\mu_1}{\mu_2} \tag{242}$$

poiché $< P_2^{-1} > \neq \frac{1}{m_2}$.
Se si ignora questo fatto e si procede facendo semplicemente la media dei rapporti fra i conteggi n_1 associati alla v.c. P_1 e i conteggi n_2 associati a P_2, si introduce un errore sistematico (bias) nel risultato che, percentualmente su $\frac{1}{\mu_2}$, come si è visto è ancora dell'ordine di $\frac{1}{\mu_2}$. Occorre dunque, a partire dai conteggi distribuiti secondo la poissoniana P_2, definire un algoritmo che consenta una migliore stima di $\frac{1}{\mu_2}$. Poniamo per questo

$$A = \sum_{n=1}^{+\infty} P_2(n) \, \frac{1}{n+1} \tag{243}$$

Rispetto alla media di P_2^{-1} abbiamo sostituito il fattore $\frac{1}{n}$ corrispondente alla probabilità $P_2(n)$ con il fattore $\frac{1}{n+1}$. Risulta

$$\begin{aligned} A &= \sum_{n=1}^{+\infty} P_2(n) \, \frac{1}{n+1} = \sum_{n=1}^{+\infty} e^{-\mu_2} \, \frac{\mu_2^n}{n!} \, \frac{1}{n+1} = \frac{e^{-\mu_2}}{\mu_2} \sum_{n=1}^{+\infty} \frac{\mu_2^{n+1}}{(n+1)!} = \\ &= \frac{e^{-\mu_2}}{\mu_2} \sum_{k=2}^{+\infty} \frac{\mu_2^k}{k!} = \frac{e^{-\mu_2}}{\mu_2} \left(-1 - \mu_2 + \sum_{k=0}^{+\infty} \frac{\mu_2^k}{k!} \right) = \frac{e^{-\mu_2}}{\mu_2} \, (-1 - \mu_2 + e^{\mu_2}) = \\ &= \frac{1}{\mu_2} - \frac{\mu_2 + 1}{\mu_2} \, e^{-\mu_2} \end{aligned} \tag{244}$$

Anche questa stima *non* fornisce esattamente $\frac{1}{\mu_2}$, però la differenza è sostanzialemte pari a $e^{-\mu_2}$ e quindi tende a essere rapidamente una quantità trascurabile.
Nel rapporto fra conteggi poissoniani, se vogliamo ottenere una quantità che coincide il più possibile con il rapporto dei valori di aspettazione concernenti separatamente il numeratore e il denominatore, si può quindi aggiungere sistematicamente una unità ai conteggi al denominatore.

Indice