



below

PROBABILITÀ, STATISTICA E DISTRIBUZIONE DI GAUSS.

di Leonardo **Rubino**

Maggio 2024

Abstract: Probabilità, Statistica e Distribuzione di Gauss.

Indice

Indice

Concetti introduttivi

- 1) La Legge Binomiale di Bernoulli (Prove ripetute)
 - 2) La Legge (o Distribuzione) di Poisson
 - 3) Richiami di calcolo combinatorio
 - 4) La Distribuzione Multinomiale
 - 5) La Distribuzione Ipergeometrica
 - 6) La Distribuzione di Gauss (Gaussiana)
 - 7) La regola del 3σ
 - 8) Un segnale della Natura: la Gaussiana in Meccanica Quantistica
 - 9) La funzione Gamma di Eulero e la Formula di Stirling
 - 10) Dimostrazione più rigorosa della Distribuzione Normale
 - 11) Eventi condizionati
 - 12) Grandezze empiriche (dalla teoria alla pratica)
 - 13) N PASSI IN DIREZIONI CASUALI (Random Walk e monete truccate)
- APPENDICE: Supplementi di statistica

Concetti introduttivi

La statistica è probabilmente la scienza alla base di tutto l'Universo, in quanto la stessa meccanica quantistica si basa su di essa. Nel presente file si vuole dare una rara dimostrazione della distribuzione di Gauss, visto che la quasi totalità dei testi disponibili (opinione di chi scrive) la enuncia soltanto, senza darne prova estesa. Si vuole poi qui fare una raccomandazione che vale comunque in qualsiasi ambito scientifico, ossia quella del non fidarsi ciecamente dell'intuizione, in campo statistico e probabilistico. E' vero che, forse, 95 volte su 100 l'intuizione porta a conclusioni corrette, e chi scrive è un forte fruitore dell'intuizione, alla base della quale stanno tantissime scoperte scientifiche, ma è anche vero che, ogni tanto, l'intuizione inganna. A tal proposito, voglio riportare un caso inerente ad una trasmissione televisiva chiamata Let's Make a Deal, in cui vi erano tre porte chiuse e dietro una di esse stava un premio; dopo che il concorrente sceglieva una delle tre porte, il conduttore andava ad aprire una delle due porte non scelte dal concorrente, per mostrare a quest'ultimo cosa vi è (o non vi è) dietro ad essa. Veniva ovviamente aperta una porta che non nascondeva il premio. A questo punto viene chiesto al concorrente se vuole ora tenere la porta da lui scelta originariamente o se vuole switchare (cambiare) con l'altra, tra le due non scelte, che è rimasta chiusa. A prima vista, affidandosi all'intuizione, sembrerebbe che cambiare porta non è necessariamente più conveniente che tenere quella già scelta, in quanto si tratta del banale caso di due porte ancora chiuse, dietro una delle quali vi è il premio, dunque si avrebbe $\frac{1}{2}$ (il 50%) di probabilità di avere il premio sia che si tenga la porta attualmente selezionata che switchando sull'altra. Bene, questo è il classico caso in cui L'INTUIZIONE TRADISCE, in quanto l'ultimo ragionamento fatto è **COMPLETAMENTE ERRATO!**

La scelta più conveniente è **SEMPRE QUELLA DI ABBANDONARE LA PORTA PRECEDENTEMENTE SCELTA** e di prendere l'altra! Vediamo perché:

chiamiamo le tre porte A, B e C e supponiamo che tu, concorrente, abbia scelto la porta A; ora:

1-se il premio è dietro B, il conduttore ti aprirà la porta C e dunque, se tu passi a B vinci.

2-se invece il premio è dietro C, il conduttore aprirà e ti mostrerà B e, se tu cambi con C, pure vinci.
 3-se invece ancora il premio è dietro A, ossia dietro la porta da te originariamente scelta, il conduttore aprirà indifferentemente B o C e tu, se cambi perdi, mentre se tieni A, vinci.
 Potete facilmente verificare che il cambiare (switchare) porta conduce alla vittoria del premio in ben due casi su tre (1 e 2), mentre porta a perdere solo in uno su tre (caso 3).
 NON E' DUNQUE VERO CHE TENERSI LA PORTA CHE GIA' SI E' SCELTA E' INDIFFERENTE, ma bensì è sempre conveniente cambiare!

Sulla Probabilità di un evento e dintorni.

Se su n prove possibili un evento si può verificare favorevolmente k volte, si dice che la sua probabilità p è:

$$p = \frac{k}{n} \quad (0 \leq p \leq 1); p=1 \text{ (100\%)} \text{ è l'evento certo, mentre } p=0 \text{ è l'evento impossibile.}$$

La probabilità dell'evento contrario è q :

$$q = \frac{n-k}{n}.$$

Ovviamente: $p+q=1$.

Se si hanno più di due eventi, ossia m eventi, si può dire che:

$$\sum_{i=1}^m p_i = 1.$$

Esempio: nel lanciare un dado, qual è la probabilità che esca una certa faccia?

$$q = \frac{k}{n} = \frac{1}{6}.$$

La Frequenza di un evento.

Se eseguiamo m prove e l'evento B si verifica m_B volte, la frequenza dell'evento B è:

$$\nu = \frac{m_B}{m}; 0 \leq \nu \leq 1 \text{ e } \sum_{i=1}^m \nu_i = 1.$$

C'è una similitudine col concetto di probabilità.

Il concetto di probabilità è di natura astratta (matematica), mentre quello di frequenza è di natura empirica (fisica).

La legge dei grandi numeri, o legge empirica del caso, o legge di G. Bernoulli:

(questo è un postulato di natura sperimentale)

Al crescere del numero delle prove, la frequenza di un evento tende a diventare uguale alla sua probabilità.

In modo più volgare: con un gran numero di prove, la pratica tende alla teoria.

La Probabilità totale:

Se C si verifica quando si verifica uno qualsiasi dei k fenomeni $B_1, \dots, B_k$, escludentisi a vicenda, e ciascuno con probabilità singole $p_1, \dots, p_k$, la probabilità di C è:

$$p_C = \sum_{i=1}^k p_i$$

Esempio: lancio un dado; qual è la probabilità che esca l'uno o il due? (basta che esca uno dei due)

$$p_C = \sum_{i=1}^2 p_i = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

La somma di probabilità è una probabilità maggiore.

La Probabilità composta:

Se C si verifica quando si verificano tutti gli eventi $B_1, B_2, \dots, B_k$, ognuno di probabilità $p_1, \dots, p_k$, allora

$$p_C = p_1 \cdot p_2 \cdot \dots \cdot p_k = \prod_{i=1}^k p_i$$

Esempio: lancio tre dadi; qual è la probabilità che esca il cinque su tutti e tre contemporaneamente?

$$p_C = \prod_{i=1}^3 p_i = \frac{1}{6} \frac{1}{6} \frac{1}{6} = \frac{1}{216}$$

Il prodotto di probabilità è una probabilità minore.

Sul concetto di distribuzione:

L'insieme di tutte le probabilità di tutti i valori che la variabile casuale x può assumere è detta la distribuzione delle probabilità di x .

Valor medio di x :

$$\hat{x} = \frac{\sum x}{n}$$

E, per una media pesata, ossia se ogni valore x_i di x ha una probabilità sua p_i :

$$\hat{x} = \sum_{i=1}^n x_i p_i$$

Nel passaggio al continuo, non si hanno singoli valori discreti x_i , ma infiniti valori, come quelli che x può assumere in un certo intervallo, e le sommatorie, notoriamente, diventano integrali:

se dp è la probabilità che x sia compreso tra ξ e $\xi + d\xi$, allora:

$dp = f(\xi)d\xi$, con $f(\xi)$ che è la densità di probabilità.

Molto intuitivamente, la probabilità che x stia tra a e b è:

$P_{a,b} = \int_a^b f(\xi)d\xi$; inoltre, per la definizione di probabilità massima:

$\int_X f(\xi)d\xi = 1$, dove X è il campo di variabilità di x .

La FUNZIONE DI DISTRIBUZIONE $F(t)$ corrispondente ad $f(\xi)$ è:

$F(t) = \int_{-\infty}^t f(\xi)d\xi$, sicchè: $P_{a,b} = \int_a^b f(\xi)d\xi = F(b) - F(a)$.

Sui libri, $F(t)$ è tabulata.

Sempre nel caso del continuo, il valor medio di x è:

$\hat{x} = \int_X \xi \cdot f(\xi)d\xi$ (media di ξ pesata con i pesi proporzionali a $f(\xi)$)

E per il valor medio di una funzione $g(x)$:

$\hat{g}(x) = \sum_{i=1}^n g(x_i) \cdot p_i$ (nel discreto)

$\hat{g}(x) = \int_X g(\xi) \cdot f(\xi)d\xi$ (nel continuo)

Valor medio del quadrato (caso del discreto):

$vm_di_x^2 = \sum_{i=1}^n x_i^2 \cdot p_i$

Valor quadratico medio:

$vqm = \sqrt{vm_di_x^2}$

Lo scarto: $\Delta x_i = s = x_i - \hat{x}$

La varianza:

essa è il valor medio dei quadrati degli scarti: $\sigma^2(x) = \sum_{i=1}^n (x_i - \hat{x})^2 \cdot p_i$. Nel continuo, si ha:

$$\sigma^2(x) = \int_X (\xi - \hat{x})^2 \cdot f(\xi) d\xi$$

Una proprietà della varianza:

$$\sigma^2(x) = \text{vm}_{di} [(x - \hat{x})^2] = \text{vm}_{di} [x^2 - 2x\hat{x} + (\hat{x})^2] = (\text{vm}_{di} x^2) - 2\hat{x}\hat{x} + (\hat{x})^2 = (\text{vm}_{di} x^2) - (\hat{x})^2 \quad (1.1)$$

Lo scarto quadratico medio:

$$\sigma(x) = \sqrt{\sigma^2(x)} \quad (\text{è la radice della varianza})$$

1) La Legge Binomiale di Bernoulli (Prove ripetute)

esponiamola tramite un esempio: si lancia un dado 5 volte. Qual è la probabilità che esca due volte una data faccia?

($p=1/6$, $q=5/6$, $n=5$, $k=2$)

$$P_k = \binom{n}{k} p^k q^{n-k} \quad (1.1)$$

$$\left(\binom{n}{k} = \frac{n!}{k!(n-k)!} \right) \quad P_2 = \frac{5!}{2!3!} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = 0,16075 = 16,075\%$$

Dimostrazione:

Calcoliamo P_0 ($k=0$, ossia l'evento non si verifica); l'evento contrario, che ha probabilità $(1-p)=q$ si presenta in tutte le n prove con probabilità $(1-p)^n=q^n$, dunque: $P_0 = q^n = \binom{n}{0} p^0 q^{n-0}$

Calcoliamo ora P_1 , cioè l'evento si verifica nella r -esima prova con probabilità p^1 , mentre nelle $(n-1)$ prove rimanenti si verifica l'evento contrario con probabilità $q^{(n-1)}$; inoltre, r può essere scelto in n modi diversi; si ha:

$$P_1 = np q^{n-1} = \binom{n}{1} p^1 q^{n-1}$$

Calcoliamo ora P_2 , ossia quando l'evento si verifica nella r -esima e nella s -esima prova con probabilità p^2 , mentre gli eventi contrari hanno probabilità $q^{(n-2)}$; tuttavia, i due numeri r ed s possono essere scelti tra n numeri ed in $\frac{n(n-1)}{2}$ modi diversi, cioè in $\binom{n}{2}$ modi diversi; infatti, il primo lo scelgo in n modi, mentre per il secondo restano $(n-1)$ modi ed essendo gli eventi indistinguibili, bisogna dividere per due, poiché uno varrebbe l'altro. Dunque:

$$P_2 = \frac{n(n-1)}{2} p^2 q^{n-2} = \binom{n}{2} p^2 q^{n-2}$$

Nel caso di P_k , la probabilità dell'evento favorevole è p^k e quella dell'evento contrario è $q^{(n-k)}$ e le k prove possono essere prese in $\binom{n}{k}$ combinazioni diverse, da cui l'asserto.

Valor medio di k :

$$\hat{k} = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np \quad (1.2)$$

(si diano ad n dei valori diversi e si vedrà che questa equazione è sempre verificata)

per dimostrarla, si tenga conto innanzitutto che la sommatoria può partire anche da $k=1$, in quanto per $k=0$ si ha un valore nullo, dunque omissibile. Inoltre, tenendo conto della espressione per il

binomio di Newton per $(a+b)^n$, si ha: (binomio di N.: $(p+q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+1-p)^n = 1$)

$$\hat{k} = \sum_{k=1}^n \frac{n(n-1)!}{(k-1)!(n-k)!} p p^{k-1} q^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} = np(p+q)^{(n-1)} = np$$

Il valore più probabile di k è dunque np , o l'intero più vicino ad esso.

Inoltre, sempre per il binomio di Newton, si ha che: $\sum_{k=0}^n P_k = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = 1$

varianza di k :

$$\sigma^2(k) = vm_di_[(k-\hat{k})]^2 = vm_di_[(k-np)]^2 = \sum_{k=0}^n (k-\hat{k})^2 \cdot P_k = npq ;$$

infatti,

$$vm_di_ (k^2) = \sum_{k=0}^n \frac{n!}{k!(n-k)!} k^2 p^k q^{n-k} = \sum_{k=0}^n \frac{n!}{k!(n-k)!} [k(k-1) + k] p^k q^{n-k} = \sum_{k=0}^n [k(k-1)] \frac{n!}{k!(n-k)!} p^k q^{n-k} +$$

$$+ \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} =$$

(ora, nell'ultimo membro, la prima sommatoria può partire da $k=2$, mentre la seconda già la calcolammo, quindi:)

$$= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} + np = n(n-1)p^2 (p+q)^{n-2} + np = n(n-1)p^2 + np =$$

$$= n^2 p^2 - np^2 + np = vm_di_ (k^2) \text{ e dunque, considerando che:}$$

$$\sigma^2(k) = vm_di_ [(k-\hat{k})^2] = (vm_di_ k^2) - (\hat{k})^2, \text{ discende che:}$$

$$\sigma^2(k) = (vm_di_ k^2) - (\hat{k})^2 = (n^2 p^2 - np^2 + np) - n^2 p^2 = np(1-p) = \boxed{npq}$$

La radice di tale quantità sarà ovviamente lo scarto quadratico medio di k :

$$\sigma(k) = \sqrt{\sigma^2(k)} = \sqrt{npq} . \quad (1.3)$$

2) La Legge (o Distribuzione) di Poisson

Partiamo dalla Legge di Bernoulli $P_k = \binom{n}{k} p^k (1-p)^{n-k}$, dove $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ e poi $\hat{k} = np$; e

ricordiamo altresì la Formula di Stirling per il fattoriale $n! = \sqrt{2\pi n} n^n e^{-n}$ da cui:

$$\ln n! = \ln[\sqrt{2\pi n} n^n e^{-n}] = n \ln n - n + \frac{1}{2} \ln(2\pi n) \quad (2.1)$$

$$\text{La Legge di Bernoulli diviene } P_k = \binom{n}{k} \left(\frac{\hat{k}}{n}\right)^k \left(1 - \frac{\hat{k}}{n}\right)^{n-k} . \quad (2.2)$$

Per la Formula di Stirling, si ha:

$$\lim_{n \rightarrow \infty} \binom{n}{k} = \lim_{n \rightarrow \infty} \frac{\sqrt{2\pi n} n^n e^{-n}}{k! \sqrt{2\pi(n-k)} (n-k)^{n-k} e^{-(n-k)}} = \lim_{n \rightarrow \infty} \frac{\sqrt{n} n^n e^{-n}}{k! \sqrt{(n-k)} (n-k)^{n-k} e^{-(n-k)}}$$

(dove la stessa non è stata per il momento applicata a $k!$, in quanto più piccolino, ma solo ad $n!$ ed $(n-k)!$)

Notiamo ora che per grandi n , $\sqrt{n}$ si riduce con $\sqrt{(n-k)}$, n^n si riduce con $(n-k)^n$, $(n-k)^{-k}$ diviene $(n)^{-k}$ e e^{-n} si riduce con $e^{-(n-k)}$, da cui:

$$\lim_{n \rightarrow \infty} \binom{n}{k} = \lim_{n \rightarrow \infty} \frac{n^k}{k!} \quad (2.3)$$

Osservando altresì che, notoriamente, $\lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n = e$, segue che $\lim_{n \rightarrow \infty} (1 + \frac{1}{n/\hat{k}})^{n/\hat{k}} = e$,

ossia: $\lim_{n \rightarrow \infty} (1 + \frac{1}{n/\hat{k}})^n = \lim_{n \rightarrow \infty} (1 + \frac{\hat{k}}{n})^n = e^{\hat{k}}$ ossia ancora: $\lim_{n \rightarrow \infty} (1 + \frac{(-\hat{k})}{n})^n = \lim_{n \rightarrow \infty} (1 - \frac{\hat{k}}{n})^n = e^{-\hat{k}}$

e infine: $\lim_{n \rightarrow \infty} (1 - \frac{\hat{k}}{n})^{n-k} = \lim_{n \rightarrow \infty} \frac{(1 - \frac{\hat{k}}{n})^n}{(1 - \frac{\hat{k}}{n})^k} = \frac{e^{-\hat{k}}}{1} = e^{-\hat{k}}$ e così la (2.2) diviene, per grandi n:

$$P_k = \binom{n}{k} (\hat{k}/n)^k (1 - \hat{k}/n)^{n-k} = \frac{n^k}{k!} (\frac{\hat{k}}{n})^k e^{-\hat{k}} = \frac{\hat{k}^k}{k!} e^{-\hat{k}} = P_k,$$

nota come Equazione di Poisson $P_k = \frac{\hat{k}^k}{k!} e^{-\hat{k}}$. (2.4)

Per il valor medio di k (vedi più sopra-(1.2)), si aveva:

$$\hat{k} = np \quad (2.5)$$

e per lo scarto quadratico medio (vedi sempre più sopra-(1.3)):

$$\sigma(k) = \sqrt{\hat{k}} = \sqrt{npq}, \quad (2.6)$$

e si ha anche, con i nuovi p e q della (2.2) e grandi n:

$$\sigma(k) = \lim_{n \rightarrow \infty} \sqrt{n \frac{\hat{k}}{n} (1 - \frac{\hat{k}}{n})} = \sqrt{\hat{k}} = \sqrt{npq} \quad (2.7)$$

$$\text{Ovviamente: } \sum_{k=0}^n P_k = \sum_{k=0}^n \frac{\hat{k}^k}{k!} e^{-\hat{k}} = e^{-\hat{k}} \sum_{k=0}^n \frac{\hat{k}^k}{k!} = e^{-\hat{k}} e^{\hat{k}} = 1$$

3) Richiami di calcolo combinatorio

combinazione di n elementi presi a gruppi di k: per quanto detto a pagina 6 sul modo di scegliere due numeri r ed s tra n, si aveva che tali modi erano $\binom{n}{2}$.

Dunque, la combinazione di n elementi presi a gruppi di k è: $C_{n,k} = \binom{n}{k}$

Ora, se i gruppi di k elementi possono differire anche per l'ordine nel quale compaiono, si parla di "disposizione" di n elementi a gruppi di k:

$$D_{n,k} = k! C_{n,k}$$

Infatti, riguardo le diverse disposizioni di k elementi, è noto che con k colori si possono fare k! bandiere, da cui la "permutazione" di n elementi presi a gruppi di n:

$$P_{n,n} = D_{n,n} = \frac{n!}{0!} = n!$$

4) La Distribuzione Multinomiale

Se si eseguono n prove e si vuole che l'evento 1, con probabilità p_1 , si verifichi k_1 volte, l'evento 2, con probabilità p_2 , si verifichi k_2 volte e l'evento n, con probabilità p_n , si verifichi k_n volte, la probabilità che ciò accada è:

$$P_{(k_1, k_2, \dots, k_n)} = \frac{n!}{k_1! k_2! \dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}$$

Esempio: si ha una scatola con 50 palline, 20 bianche, 10 rosse, 5 nere e 15 gialle; calcolare la probabilità P che estraendo 10 palline, queste siano 3_R, 2_B, 4_N e 1_G (dopo ogni estrazione, la pallina viene rimessa nella scatola):

$$P_{(3_R, 2_B, 4_N, 1_G)} = \frac{10!}{3! 2! 4! 1!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 \left(\frac{1}{10}\right)^4 \left(\frac{3}{10}\right)^1$$

5) La Distribuzione Ipergeometrica

Se da una scatola con N palline, di cui k rosse, si esegue un'estrazione di x palline ($x \leq k$) ed ogni pallina estratta non viene riposta nella scatola, bisogna abbandonare la Legge di Bernoulli e ricorrere alla distribuzione ipergeometrica; la possibilità di estrarre in sequenza x palline rosse ed (n-x) palline di altro colore è ovviamente (il prodotto di tutte le probabilità delle estrazioni):

$$\begin{aligned} P^i(x) &= \frac{k}{N} \cdot \frac{k-1}{N-1} \dots \frac{k-(x-1)}{N-(x-1)} \cdot \frac{N-k}{N-x} \cdot \frac{N-k-1}{N-x-1} \dots \frac{N-k-(n-x-1)}{N-(n-1)} = \\ &= \frac{k!}{(k-x)!} \cdot \frac{(N-n)!}{N!} \dots \frac{(N-k)!}{(N-k-n+x)!} = \frac{k!}{N!} \cdot \frac{(N-n)!}{(k-x)!} \cdot \frac{(N-k)!}{(N-k-n+x)!}; \end{aligned}$$

il numero complessivo di queste sequenze di palline è il numero di combinazioni di n oggetti a x a x:

$$C(n, x) = \binom{n}{x} = \frac{n!}{x!(n-x)!}, \text{ dunque, la probabilità totale di avere x palline rosse è:}$$

$$P(x) = C(n, x) P^i(x) = \frac{C(k, x) C(N-k, n-x)}{C(N, n)} \quad (\text{legge ipergeometrica}), \text{ con:}$$

k=n. di palline rosse

n=n. di estrazioni

N=n. totale di palline

x=n. di palline rosse che si vogliono estrarre (variabile ipergeometrica)

6) La Distribuzione di Gauss (Gaussiana)

Prendiamo il logaritmo della (2.4): $\ln P_k = \ln \left[\frac{\hat{k}^k}{k!} e^{-\hat{k}} \right] = k \ln \hat{k} - \ln k! - \hat{k}$ e, per la (2.1) si ha:

$$\ln P_k = \ln \left[\frac{\hat{k}^k}{k!} e^{-\hat{k}} \right] = k \ln \hat{k} - k \ln k + k - \ln \sqrt{2\pi k} - \hat{k}, \quad (6.1)$$

da cui, prendendone la derivata:

$$\frac{d}{dk} \ln P_k = \ln \hat{k} - \ln k - k \frac{1}{k} + 1 - \frac{1}{\sqrt{2\pi k}} \frac{1}{2} \frac{1}{\sqrt{2\pi k}} 2\pi = \ln \hat{k} - \ln k - \frac{1}{2k} \quad (6.2)$$

Ponendo la (6.2) uguale a zero si ottiene il valore di k per cui si ha un massimo per $\ln P_k$:

$$\frac{d}{dk} \ln P_k = \ln \hat{k} - \ln k - \frac{1}{2k} = 0, \quad (6.3)$$

da cui, per grandi k: $\ln \hat{k} - \ln k = 0$, ossia: $k = \hat{k}$. Sviluppiamo ora in serie di Taylor $\ln P_k$ proprio nell'intorno di $\hat{k}$, per cui si ha il massimo di $\ln P_k$:

$$\begin{aligned} \ln P_k &= \ln P(k) = \ln P(\hat{k}) + (k - \hat{k}) \left[\frac{d}{dk} \ln P(k) \right]_{k=\hat{k}} + \frac{(k - \hat{k})^2}{2!} \left[\frac{d^2}{dk^2} \ln P(k) \right]_{k=\hat{k}} + \dots = \\ &= \ln P(k) = \ln P(\hat{k}) + \frac{(k - \hat{k})^2}{2!} \left[\frac{d^2}{dk^2} \ln P(k) \right]_{k=\hat{k}} + \dots, \quad (6.4) \end{aligned}$$

in quanto il termine con la derivata prima è nullo, per la (6.3). E dalla (6.2) si ha che:

$\frac{d^2}{dk^2} \ln P(k) = -\frac{1}{k} + \frac{1}{2k^2}$ e, grazie a quest'ultima equazione, la (6.4) fornisce (per grandi k):

$$\ln P(k) = \ln P(\hat{k}) + \frac{(k - \hat{k})^2}{2!} \left[-\frac{1}{\hat{k}} + \frac{1}{2\hat{k}^2} \right] = \ln P(\hat{k}) - \frac{(k - \hat{k})^2}{2\hat{k}} + \frac{(k - \hat{k})^2}{4\hat{k}^2} \quad (6.5)$$

Ricordando ora che per la (6.1) si ha: $\ln P(\hat{k}) = \hat{k} \ln \hat{k} - \hat{k} \ln \hat{k} + \hat{k} - \ln \sqrt{2\pi\hat{k}} - \hat{k} = -\ln \sqrt{2\pi\hat{k}}$, allora la (6.5) fornisce:

$$\ln P(k) = -\ln \sqrt{2\pi\hat{k}} - \frac{(k - \hat{k})^2}{2\hat{k}} + \frac{(k - \hat{k})^2}{4\hat{k}^2} \approx -\ln \sqrt{2\pi\hat{k}} - \frac{(k - \hat{k})^2}{2\hat{k}}, \quad (6.6)$$

avendo trascurato il termine con $\hat{k}^2$ a denominatore, essendo in regime di k grandi (e, dunque, anche di grandi $\hat{k}$). E, per la (6.6), finalmente:

$$P(k) = \frac{1}{\sqrt{2\pi\hat{k}}} e^{-\frac{(k - \hat{k})^2}{2\hat{k}}} \quad (\text{Distribuzione Normale o Gaussiana}) \quad (6.7)$$

Per la (2.7) (utilizzata nei denominatori della (6.7)) e la (2.5) (utilizzata al numeratore dell'esponenziale della (6.7), in quanto lì il confronto, non quadratico, è con k), si ha:

$$P_k = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k - np)^2}{2npq}} \quad (6.7b)$$

(purtroppo le approssimazioni sono state tante, ma, del resto, tante sono anche le condizioni cui la Distribuzione Gaussiana deve sottostare, come i grandi n e le grandi k)

Definendo ora la funzione $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(k - np)^2}{2npq}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, con $z = \frac{(k - np)}{\sqrt{npq}}$, si ha che $\varphi(z)$ è

appunto la “legge normale della probabilità” o “curva degli errori di Gauss”.

Ricordando poi le espressioni che ottenemmo per lo scarto semplice ($s = |k - \hat{k}| = |k - np|$) e per lo

scarto q.m. di k ($\sigma(k) = \sqrt{\sigma^2(k)} = \sigma_k = \sqrt{npq}$), si ha:

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{k - \hat{k}}{\sigma_k} \right)^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{s}{\sigma_k} \right)^2} \quad (6.8)$$

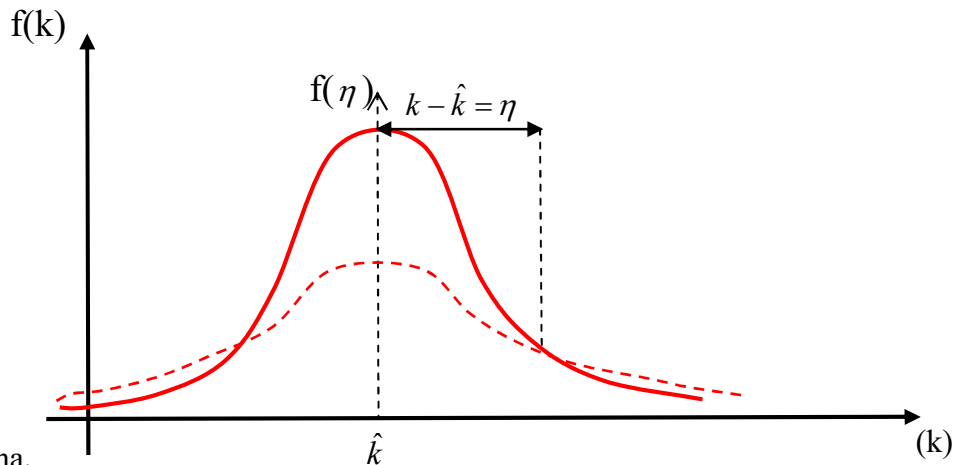


Fig. 6.1: Gaussiana.

Valutiamo ora la probabilità che l'evento si verifichi un numero di volte compreso tra k_α e k_β :

$$P_{k_\alpha, k_\beta} = \int_{k_\alpha}^{k_\beta} P_k dk = \int_{k_\alpha}^{k_\beta} \frac{1}{\sigma_k \sqrt{2\pi}} e^{\left[-\frac{(k-np)^2}{2npq}\right]} dk ; \text{ passando ora alla variabile d'integrazione } z:$$

$$P_{k_\alpha, k_\beta} = \int_{z_\alpha}^{z_\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{z_\alpha}^{z_\beta} \phi(z) dz .$$

Consideriamo ora la funzione:

$$\phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^t \phi(z) dz ; \text{ si ha che:}$$

$$P_{k_\alpha, k_\beta} = \phi(t_\beta) - \phi(t_\alpha) = \phi\left(\frac{k_\beta - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_\alpha - np}{\sqrt{npq}}\right) . \text{ Esistono tabelle di valori della funzione } \phi(t) .$$

7) La regola del 3σ

Riprendiamo la (6.8):

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\hat{k}}{\sigma_k}\right)^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} = P(s)$$

ora, a livello differenziale, la probabilità dp che lo scarto s sia compreso entro l'intervallo infinitesimo ds è:

$$dp = P(s)ds = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} ds \text{ e, quindi, la probabilità che lo scarto } s \text{ sia compreso tra i valori } s_1$$

ed s_2 è P :

$$P_{1,2} = \int_{s_1}^{s_2} dp = \int_{s_1}^{s_2} P(s)ds = \int_{s_1}^{s_2} \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} ds \quad (7.1)$$

$P_{1,2}$ è l'area sottesa dalla Gaussiana tra s_1 ed s_2 ; si calcola facilmente, tramite l'integrale della (7.1), che:

$-0,68\sigma < s < 0,68\sigma$	$\rightarrow P=50\%$
$-\sigma < s < \sigma$	$\rightarrow P=68,3\%$
$-1,64\sigma < s < 1,64\sigma$	$\rightarrow P=90\%$
$-2\sigma < s < 2\sigma$	$\rightarrow P=95,9\%$
$-2,58\sigma < s < 2,58\sigma$	$\rightarrow P=99\%$
$-3\sigma < s < 3\sigma$	$\rightarrow P=99,7\%$

Allora, la probabilità che lo scarto dal valore medio di una misura sia compreso tra -3σ e $+3\sigma$ è del 99,7%, ossia un valore piuttosto accettabile, da cui la frequente adozione della regola del tre sigma appunto.

8) Un segnale della Natura: la Gaussiana in Meccanica Quantistica

Relazioni di indeterminazione:

$$\Delta x \cdot \Delta p \geq \frac{\hbar}{2} \quad \text{e} \quad \Delta E \cdot \Delta t \geq \frac{\hbar}{2} . \quad (8.1)$$

Ci chiediamo in quale caso particolare valga il segno di uguaglianza, ossia in che caso il prodotto delle due indeterminazioni (ossia dei due delta) sia il più basso possibile, ossia il meno incerto possibile, ossia il più certo possibile. Per le definizioni dei vari Δ_F, Δ_G e $\Delta F, \Delta G$ si faccia riferimento a quanto segue:

$\langle F \rangle_i = f_i$ (numero) ; F sarebbe una "osservabile", evidentemente. Ricordiamo poi la definizione di scarto quadratico medio ΔF per F e richiediamo che esso si annulli:

$\Delta F = \sqrt{\langle F^2 \rangle_i - \langle F \rangle_i^2} = 0$. Definiamo altresì lo "scarto semplice" Δ_F : $\Delta_F = F - \langle F \rangle_i$. Si ha allora che: $\langle \Delta_F^2 \rangle_i = \langle (F - \langle F \rangle_i)^2 \rangle_i = \langle F^2 \rangle_i + \langle F \rangle_i^2 - 2\langle F \rangle_i \langle F \rangle_i = \langle F^2 \rangle_i - \langle F \rangle_i^2 = (\Delta F)^2$. Ora, la richiesta qui sopra, secondo cui: $\Delta F = 0$, si traduce in ciò:

$\langle \Delta_F^2 \rangle_i = 0 = (\Psi_i, \Delta_F^2 \Psi_i) = 0$. Essendo poi F un'osservabile, dunque hermitiana ($F^* = F$) anche Δ_F sarà hermitiano, e allora potremo scrivere:

$\langle \Delta_F^2 \rangle_i = (\Psi_i, \Delta_F^2 \Psi_i) = (\Delta_F \Psi_i, \Delta_F \Psi_i) = \int |\Delta_F \Psi_i|^2 d\xi = 0$, da cui: $\Delta_F \Psi_i = 0$, ossia: $F \Psi_i = f_i \Psi_i$, che è l'equazione agli autovalori per F.

Introduciamo l'operatore $O_\lambda = \lambda \Delta_F + i \Delta_G$ ($\lambda \in \mathbb{R}$); esso è palesemente non hermitiano, in quanto:

$O_\lambda^+ = \lambda \Delta_F - i \Delta_G \neq O_\lambda$; la norma di $O_\lambda \Psi$ è, per definizione di norma, positiva ($O_\lambda \Psi, O_\lambda \Psi \geq 0$).

Essendo poi $(O_\lambda \Psi, O_\lambda \Psi) = (\Psi, O_\lambda^+ O_\lambda \Psi)$ ed esplicitando, si ha:

$$(O_\lambda^+ O_\lambda = \Delta_F^2 + \Delta_G^2 + i\lambda \Delta_F \Delta_G - i\lambda \Delta_G \Delta_F)$$

$$\lambda^2 < \Delta_F^2 > + \lambda < i[\Delta_F, \Delta_G] > + < \Delta_G^2 > \geq 0, \text{ ossia:}$$

$\lambda^2 (\Delta F)^2 + \lambda < i[F, G] >_\Psi + (\Delta G)^2 \geq 0$; il discriminante di tale forma quadratica è negativo o nullo, visto il tipo di diseuguaglianza, dunque:

$(< i[F, G] >_\Psi)^2 - 4(\Delta F)^2 (\Delta G)^2 \leq 0$, da cui le relazioni di indeterminazione (8.1) di cui sopra, ossia

$\Delta F \cdot \Delta G \geq \frac{1}{2} |< i[F, G] >_\Psi|$. Vale quindi il segno di eguaglianza quando il discriminante è nullo ed in corrispondenza di ciò, si ottiene:

$$\lambda_{\min} = -\frac{< i[F, G] >_\Psi}{2(\Delta F)^2} ; \text{ in corrispondenza di tale valore, si ha: } (O_{\lambda_{\min}} \Psi, O_{\lambda_{\min}} \Psi) = 0, \text{ cioè:}$$

$O_{\lambda_{\min}} \Psi = 0$. Esplicitando, in quest'ultima relazione, $O_{\lambda_{\min}}$, si ha:

$$\left[\frac{< -i[F, G] >}{2(\Delta F)^2} (F - < F >) + i(G - < G >)] \Psi = 0. \quad (8.2)$$

Nel caso in cui $F=x$ e $G=p_x = -i\hbar \frac{\partial}{\partial x}$, si ha: $-i[F, G] = -i[x, p_x] = \hbar$, e la (8.2) si può così scrivere:

$$\left[\frac{x - < x >}{2(\Delta x)^2} + \frac{\partial}{\partial x} - i \frac{p_x - < p_x >}{\hbar} \right] \Psi = 0, \text{ ossia:}$$

$$\frac{d\Psi}{\Psi} = \left[-\frac{x - < x >}{2(\Delta x)^2} + i \frac{p_x - < p_x >}{\hbar} \right] dx, \text{ la cui soluzione è:}$$

$$\Psi(x) = N e^{-\frac{(x - < x >)^2}{4(\Delta x)^2}} e^{i \frac{< p_x >}{\hbar} x}, \text{ che è un pacchetto d'onde di forma gaussiana centrato in } < x >.$$

9) La funzione Gamma di Eulero e la Formula di Stirling

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \quad (\text{Gamma di Eulero})$$

Naturalmente, $\Gamma(1) = 1$. Poi, $\Gamma(n+1) = n\Gamma(n)$; infatti, dopo una integrazione per parti:

$$\Gamma(z) = e^{-t} \frac{t^z}{z} \Big|_0^\infty + \frac{1}{z} \int_0^\infty e^{-t} t^z dt = 0 + \frac{1}{z} \Gamma(z+1) \text{ ed iterando la } \Gamma(n+1) = n\Gamma(n), \text{ otteniamo:}$$

$$\Gamma(n+1) = n! .$$

Formula di Stirling:

$\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = \int_0^\infty e^{n \ln x - x} dx$ e la funzione $n \ln x - x$ ha un massimo in $x=n$, come si può vedere ponendo a zero la sua derivata. Ora, ponendo $x=n+y$, abbiamo:

$$\Gamma(n+1) = e^{-n} \int_{-n}^\infty e^{n \ln(n+y) - y} dy = e^{-n} \int_{-n}^\infty e^{n \ln n + n \ln(1+y/n) - y} dy = n^n e^{-n} \int_{-n}^\infty e^{n \ln(1+y/n) - y} dy$$

Ora, in accordo con Taylor, abbiamo che: $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$; poniamo dapprima $x=y/n$ e

$$\text{poi } y = \sqrt{n}v, \text{ così: } \Gamma(n+1) = n^n e^{-n} \int_{-n}^\infty e^{n \ln(1+y/n) - y} dy = n^n e^{-n} \int_{-n}^\infty e^{-y^2/2n + y^3/3n^2 - \dots} dy =$$

$$= n^n e^{-n} \sqrt{n} \int_{-\sqrt{n}}^\infty e^{-v^2/2 + v^3/3\sqrt{n} - \dots} dv, \text{ e se } n \text{ è grande abbastanza, il secondo termine nell'esponentiale si annulla e: } \Gamma(n+1) = \sqrt{2\pi n} n^n e^{-n} \text{ (Formula di Stirling), siccome l'integrale di Gauss valeva } \sqrt{2\pi}.$$

Integrale di Gauss

Consideriamo i due integrali (uguali tra loro):

$$I = \int_0^\infty e^{-\alpha x^2} dx \quad I = \int_0^\infty e^{-\alpha y^2} dy$$

Moltiplicandoli tra loro: $I^2 = \int_0^\infty \int_0^\infty e^{-\alpha(x^2+y^2)} dx dy$ e, in coordinate polari:

($x^2 + y^2 = r^2$, $dS = dx dy = r dr d\theta$), si ha:

$$I^2 = \int_0^{\pi/2} \int_0^\infty e^{-\alpha r^2} r dr d\theta = \frac{\pi}{2} \int_0^\infty e^{-\alpha r^2} r dr = -\frac{\pi}{4\alpha} \left| e^{-\alpha r^2} \right|_0^\infty = \frac{\pi}{4\alpha}$$

perciò: $I = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}}$ e allora, per l'integrale più sopra, nella dimostrazione della Formula di Stirling, si ha ($\alpha = 1/2$ ed n grande):

$$\Gamma(n+1) = n! = n^n e^{-n} \sqrt{n} \int_{-\infty}^\infty e^{-v^2/2} dv = n^n e^{-n} \sqrt{n} [2 \int_0^\infty e^{-v^2/2} dv] = \sqrt{2\pi n} n^n e^{-n}$$

10) Dimostrazione più rigorosa della Distribuzione Normale

Ripartendo ora dalla distribuzione di Bernoulli:

$$P_x = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x} \text{ e dalla Formula di Stirling: } x! \approx (\sqrt{2\pi}) x^{x+\frac{1}{2}} e^{-x}$$

e ricordando che avevamo ottenuto np come espressione per il valor medio di k , definiamo un $v = np - x$, da cui: $x = np - v$ e sostituiamo tale x nella espressione di Bernoulli, qui sopra:

$$P_x = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x} = \frac{n!}{(n+v-np)!(np-v)!} p^{(np-v)} q^{(n+v-np)} \cong (\text{per Stirling})$$

[e poi anche perchè: $(1-p)=q$]

$$\cong \frac{\sqrt{2\pi n}^{n+\frac{1}{2}} e^{-n}}{\sqrt{2\pi}(nq+v)^{(nq+v+\frac{1}{2})} e^{(-nq-v)} \sqrt{2\pi}(np-v)^{(np-v+\frac{1}{2})} e^{(-np+v)}} p^{(np-v)} q^{(nq+v)} =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{n^{\frac{n+1}{2}} p^{(np-v)} q^{(nq+v)}}{[nq(1+\frac{v}{nq})]^{nq+v+\frac{1}{2}} [np(1-\frac{v}{np})]^{np-v+\frac{1}{2}}} = \frac{1}{\sqrt{2\pi npq}} \frac{1}{(1+\frac{v}{nq})^{nq+v+\frac{1}{2}} (1-\frac{v}{np})^{np-v+\frac{1}{2}}}$$

Ricordando ora che a^x si può scrivere anche come: $a^x = e^{x \ln a}$ e ricordando anche che, per gli Sviluppo in Serie di Taylor, si ha:

$$\ln(1+x) \cong x - \frac{1}{2}x^2 + \dots \approx x, \text{ da cui: } \ln(1+\frac{v}{nq}) \approx \frac{v}{nq} \text{ e } \ln(1-\frac{v}{np}) \approx -\frac{v}{np}, \text{ si ha:}$$

$$\begin{aligned} P_x &= \frac{n!}{(n-x)!x!} p^x q^{n-x} \cong \frac{1}{\sqrt{2\pi npq}} \exp[-(nq+v+\frac{1}{2})\ln(1+\frac{v}{nq})] \exp[-(np-v+\frac{1}{2})\ln(1-\frac{v}{np})] \cong \\ &\cong \frac{1}{\sqrt{2\pi npq}} \exp[-(nq+v+\frac{1}{2})(\frac{v}{nq} - \frac{v^2}{2n^2q^2})] \exp[-(np-v+\frac{1}{2})(-\frac{v}{np} - \frac{v^2}{2n^2p^2})] \cong \\ &\cong \frac{1}{\sqrt{2\pi npq}} \exp[-\frac{v^2}{2npq} + \frac{v(p-q)}{2npq}] \end{aligned} \quad (10.1)$$

Se ora poniamo $p \cong q \cong \frac{1}{2}$ e ricordiamo poi come definimmo v , otteniamo:

$$P_x = \frac{n!}{(n-x)!x!} p^x q^{n-x} \cong \frac{1}{\sqrt{2\pi npq}} \exp(-\frac{v^2}{2npq}) = \frac{1}{\sqrt{2\pi npq}} \exp[-\frac{(x-np)^2}{2npq}], \text{ ossia:}$$

$$P_x = \frac{1}{\sqrt{2\pi npq}} e^{\left[-\frac{(x-np)^2}{2npq}\right]} \quad (\text{Distribuzione di Gauss})$$

Se invece la condizione precedente $p \cong q \cong \frac{1}{2}$ non è pienamente soddisfatta, il secondo addendo nell'esponentiale della (10.1) genererà una asimmetria intorno al valore $v=0$, o, ciò che è lo stesso, intorno ad $x=np$, ma tale asimmetria scompare per $n \rightarrow \infty$, in quanto si ha che:

$$\lim_{n \rightarrow \infty} \frac{v(p-q)}{2npq} = 0.$$

11) Eventi condizionati

Supponiamo che l'evento Y possa verificarsi solo se prima si è verificato l'evento X ed indichiamo con $p(Y/X)$ la probabilità (condizionata) che si verifichi appunto Y (ovviamente dopo X). Y/X è invece l'evento Y con X già avvenuto. Per ultimo, $p(Y)$ e $p(X)$ siano le probabilità non condizionate. Possiamo dire che $p(Y)=p(XY)$, con XY inteso come evento X e Y insieme; infatti, se si presenta Y è perché si è già presentato anche X. Allora, per la probabilità composta (moltiplicazione delle probabilità di eventi simultanei) si ha:

$$p(Y)=p(XY)=p(X)p(Y/X), \quad (11.1)$$

$$\text{da cui: } p(Y/X)=p(XY)/p(X). \quad (11.2)$$

Scrivendo ora la (11.1) con X ed Y scambiati, si ha: $p(YX)=p(Y)p(X/Y) [=p(XY)]$, da cui:

$$p(XY)=p(YX) \rightarrow p(X)p(Y/X)=p(Y)p(X/Y) \rightarrow$$

$$p(X/Y) = p(Y/X) \frac{p(X)}{p(Y)} \quad (\text{Teorema di Bayes}) \quad (11.3)$$

Adesso, per amor di generalità, supponiamo che Y si possa verificare non solo se si verifica X, ma bensì se si verifica uno dei k eventi X_i , che si escludono a vicenda, ossia se si sta verificando X_1 , allora non si stanno verificando gli altri X_i (con $i \neq 1$). Bene, allora per la probabilità composta (più favorevole, ossia data dalla somma, poiché vi sono varie X_i che autorizzano Y) si ha:

$$p(Y) = p(X_1)p(Y/X_1) + p(X_2)p(Y/X_2) + \dots + p(X_k)p(Y/X_k) = \sum_{i=1}^k p(X_i)p(Y/X_i) = p(Y) \quad (11.4)$$

che è il Teorema della probabilità totale per eventi condizionati.

$$\text{E analogamente: } p(X) = \sum_{i=1}^k p(Y_i)p(X/Y_i) \quad (11.5)$$

ed il Teorema di Bayes è un caso particolare di questo teorema più generale.

Esempio: si abbiano 8 scatole di tre tipi diversi: X_1, X_2 ed X_3 e ciascuna di queste 8 scatole contiene 80 palline, delle quali sono rosse solo 60, 45 e 20 (se rispettivamente dei tipi X_1, X_2 ed X_3). Si abbiano dunque 1 scatola del tipo X_1 , 3 del tipo X_2 e 4 del tipo X_3 . Scegliendo a caso una scatola (tutte ed 8 le scatole abbiano la stessa probabilità di essere scelte) ed estraendo da questa una pallina, qual è la probabilità che essa sia rossa?

Soluzione: tale evento di estrazione di una pallina rossa sia Y . Prima però avviene la scelta di una scatola, tra le 8 disponibili e le probabilità per tipo (di scatola) sono ovviamente:

$p(X_1)=1/8$, $p(X_2)=3/8$ e $p(X_3)=4/8$, mentre le probabilità condizionate di estrazione di una pallina rossa da ciascuna delle scatole sono: $p(Y/X_1)=60/80$, $p(Y/X_2)=45/80$ e $p(Y/X_3)=20/80$.

Il risultato allora è: $p(Y)=(1/8)(60/80)+(3/8)(45/80)+(4/8)(20/80)=275/640=0,43$

12) Grandezze empiriche (dalla teoria alla pratica)

Nei concetti introduttivi, a pagina 3, abbiamo parlato delle seguenti grandezze:

$$\text{Il valor medio: } \hat{x} = \frac{\sum x}{n} \quad (12.1)$$

$$\text{Lo scarto: } \Delta x_i = s = x_i - \hat{x} \quad (12.2)$$

La varianza: essa è il valor medio dei quadrati degli scarti:

$$\sigma^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2 \quad (12.3)$$

Lo scarto quadratico medio:

$$\sigma(x) = \sqrt{\sigma^2(x)} \quad (\text{è la radice della varianza}) \quad (12.4)$$

Inoltre, ricordiamo che il valor medio (matematico, teorico) $\hat{x}$ è di natura matematica, teorica, mentre se effettuiamo n misure di una variabile x , ossia $x_1, x_2, \dots, x_n$, possiamo valutare un parametro empirico detto “valor medio empirico” $\bar{x}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n} \quad (12.5)$$

Interpretiamo ora tali grandezze in forma operatoriale:

$$\text{Il valor medio: } = \frac{\sum x}{n} = E(x) \quad (\text{operatore } E()) \quad (12.6)$$

ed ecco anche alcune proprietà (perlopiù ovvie) del valor medio: $E(x+y) = E(x) + E(y)$,

$E(xy) = E(x)E(y)$, $E(ax+b) = aE(x) + b$, $E(\frac{x}{a}) = \frac{\hat{x}}{a}$. Definendo poi lo scarto:

$\eta = x - \hat{x} = x - E(x)$, si ha: $E(\eta) = \hat{\eta} = E(x - \hat{x}) = \hat{x} - \hat{x} = 0$.

La varianza: essa è il valor medio dei quadrati degli scarti:

$$\sigma^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2 = Var(x) = E(x_i - \hat{x})^2 \quad (\text{operatore } \sigma^2()) \quad (12.7)$$

Ecco alcune proprietà della varianza:

$\sigma^2(x_i) = \sigma^2(x)$ (così come la media di x è la media di qualsiasi x , ossia anche di un x_i , allora lo stesso vale anche per la varianza)

$$\sigma^2(x) = E(x - \hat{x})^2 = E(x^2 - 2x\hat{x} + (\hat{x})^2) = Ex^2 - 2\hat{x}Ex + (\hat{x})^2 = Ex^2 - 2\hat{x}\hat{x} + (\hat{x})^2 = Ex^2 - (\hat{x})^2 = Ex^2 - (Ex)^2.$$

Poi, se x ed y sono due variabili casuali indipendenti, allora:

$$\sigma^2(x \pm y) = \sigma^2(x) + \sigma^2(y) ; \text{ infatti:}$$

$$\sigma^2(x \pm y) = E[(x_i \pm y_i) - (\hat{x} \pm \hat{y})]^2 = E[(x_i - \hat{x}) \pm (y_i - \hat{y})]^2 = E(x_i - \hat{x})^2 + E(y_i - \hat{y})^2 \pm 2E[(x_i - \hat{x})(y_i - \hat{y})] = \sigma^2(x) + \sigma^2(y) \pm 2E(x_i - \hat{x})E(y_i - \hat{y}) = \sigma^2(x) + \sigma^2(y) \pm 2 \cdot 0 \cdot 0 = \sigma^2(x) + \sigma^2(y)$$

$$\sigma^2\left(\frac{x}{a}\right) = E\left(\frac{x}{a} - \frac{\hat{x}}{a}\right)^2 = \frac{1}{a^2} \sigma^2(x) , \quad \sigma^2(x + c) = \sigma^2(x) ,$$

$$\sigma^2(\eta) = E(\eta - \hat{\eta})^2 = E(\eta - 0)^2 = E\eta^2 = E(x - \hat{x})^2 = \sigma^2(x)$$

$$\text{Il valor medio empirico: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n} = M(x) \quad (\text{operatore } M()) \quad (12.8)$$

Il valor medio empirico avrà un suo valor medio; calcoliamolo:

$$E(\bar{x}) = E(M(x)) = E\left(\frac{x_1}{n}\right) + E\left(\frac{x_2}{n}\right) + \dots + E\left(\frac{x_n}{n}\right) = \frac{\hat{x}}{n} + \dots + \frac{\hat{x}}{n} = \hat{x}$$

Il valor medio empirico avrà una sua varianza; calcoliamola:

$$\sigma^2(\bar{x}) = \sigma^2\left(\frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n}\right) = \sigma^2\left(\frac{x_1}{n}\right) + \dots + \sigma^2\left(\frac{x_n}{n}\right) = \frac{\sigma^2(x_1)}{n^2} + \dots + \frac{\sigma^2(x_n)}{n^2} = \frac{\sigma^2(x)}{n^2} + \dots + \frac{\sigma^2(x)}{n^2} = n \frac{\sigma^2(x)}{n^2} = \frac{\sigma^2(x)}{n} , \text{ quindi:}$$

$$\boxed{\sigma^2(\bar{x}) = \frac{\sigma^2(x)}{n}} \quad (12.9)$$

$$\text{da cui ancora: } \sigma(\bar{x}) = \sqrt{\sigma^2(\bar{x})} = \frac{\sigma(x)}{\sqrt{n}} \quad (12.10)$$

La varianza empirica:

Se la varianza è: $\sigma^2()$, la varianza empirica la chiameremo: $s^2()$. La varianza come l'abbiamo conosciuta fino a questo momento si riferiva ad un parametro teorico $\hat{x}$ mentre nella pratica si dispone facilmente del parametro empirico $\bar{x}$; ed una siffatta nuova varianza (empirica) la chiameremo appunto $s^2(x)$.

Consideriamo allora la seguente espressione: $x_i - \bar{x} = (x_i - \hat{x}) - (\bar{x} - \hat{x})$; quadriamo e sommiamo su

$$i: \sum_i (x_i - \bar{x})^2 = \sum_i [(x_i - \hat{x}) - (\bar{x} - \hat{x})]^2 = \sum_i (x_i - \hat{x})^2 - 2 \sum_i [(x_i - \hat{x})(\bar{x} - \hat{x})] + \sum_i (\bar{x} - \hat{x})^2$$

Analizziamo ora gli ultimi tre addendi:

$$\text{-) } \sum_i (x_i - \hat{x})^2 = \sum_i (x_i - \hat{x})^2 \quad (\text{resta così})$$

$$\begin{aligned} \text{-) } 2 \sum_i [(x_i - \hat{x})(\bar{x} - \hat{x})] &= 2(\bar{x} - \hat{x}) \sum_i [(x_i - \hat{x})] = 2(\bar{x} - \hat{x}) \sum_i x_i - 2(\bar{x} - \hat{x}) \sum_i \hat{x} = \\ &= 2n\bar{x}(\bar{x} - \hat{x}) - 2n\hat{x}(\bar{x} - \hat{x}) = 2n(\bar{x} - \hat{x})(\bar{x} - \hat{x}) = 2n(\bar{x} - \hat{x})^2 \end{aligned}$$

$$\text{-) } \sum_i (\bar{x} - \hat{x})^2 = \sum_i \text{const} = n \cdot \text{const} = n(\bar{x} - \hat{x})^2$$

$$\text{dunque: } \sum_i (x_i - \bar{x})^2 = \sum_i (x_i - \hat{x})^2 - n(\bar{x} - \hat{x})^2 .$$

Riguardo i due termini a secondo membro dell'ultima equazione: -) $\sum_i (x_i - \hat{x})^2 = n\sigma^2(x)$

-) $n(\bar{x} - \hat{x})^2 = nE(\bar{x} - \hat{x})^2$, poiché abbiamo visto che il valor medio di una quantità costante è la costante stessa; inoltre: $E(\bar{x} - \hat{x})^2 = \sigma^2(\bar{x})$ (definizione di varianza applicata ad $\bar{x}$ invece che ad x generica, ossia la varianza di $\bar{x}$); ma più sopra abbiamo calcolato appunto la varianza del valor medio empirico $\bar{x}$ ed ottenemmo: $\sigma^2(\bar{x}) = \frac{\sigma^2(x)}{n}$, da cui: $n(\bar{x} - \hat{x})^2 = nE(\bar{x} - \hat{x})^2 = n \frac{\sigma^2(x)}{n} = \sigma^2(x)$, o, in alternativa, possiamo osservare che $(\bar{x} - \hat{x})$ è, per definizione, lo scarto che si ha nel fornire $\bar{x}$, ossia, per definizione, $\sqrt{\sigma^2(\bar{x})} = \sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{n}}$ e, quindi, di nuovo: $n(\bar{x} - \hat{x})^2 = n(\frac{\sigma(x)}{\sqrt{n}})^2 = n \frac{\sigma^2(x)}{n} = \sigma^2(x)$

Riassumendo: $\sum_i (x_i - \bar{x})^2 = n\sigma^2(x) - \sigma^2(x) = (n-1)\sigma^2(x)$ (12.11)

ossia: $\sigma^2(x) = \frac{n}{n-1} \frac{1}{n} \sum_i (x_i - \bar{x})^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$ e infine:

$$\sigma(x) = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} = s(x) \quad \left(\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \right) \quad (12.12)$$

che possiamo dunque usare come miglior risultato per disporre della varianza, che chiameremo, a questo punto, empirica ($s^2(x)$), essendo essa calcolata (ripetiamolo) usufruendo di un parametro empirico $\bar{x}$ in luogo di un parametro teorico $\hat{x}$. La (12.12) è considerabile come l'errore empirico di una misurazione.

Usando ora la (12.12) nella (12.11), si ottiene: $(n-1)s^2(x) = (n-1)\sigma^2(x)$, da cui: $s^2(x) = \sigma^2(x)$, fatto, questo, che ci incoraggia ancor di più ad utilizzare la varianza empirica s^2 come valutazione approssimata della varianza σ^2 .

Per ultimo, le (12.9) e (12.10) ci portano a scrivere: $s^2(\bar{x}) = \frac{s^2(x)}{n}$ e $s(\bar{x}) = \frac{s(x)}{\sqrt{n}}$ ossia:

$$\frac{s^2(x)}{n} = s^2(\bar{x}) = \frac{\sum_i (x_i - \bar{x})^2}{n(n-1)} \quad \text{e:} \quad (12.13)$$

$$s(\bar{x}) = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n(n-1)}} \quad (12.14)$$

La (12.14) è considerabile come l'errore medio empirico della media aritmetica.

Nella pratica:

Nella pratica, in un esperimento, come valore più probabile di x si assume: $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$.

Poi, si calcola $s(\bar{x})$ con la (12.14) e si fornisce infine: $x = \bar{x} \pm s(\bar{x})$ e l'errore relativo percentuale

sarà: $\varepsilon_r \% = \frac{s(\bar{x})}{\bar{x}} 100$.

Propagazione degli errori:

Si abbia $z=f(x,y)$; si ha: $\hat{z} = f(\hat{x}, \hat{y})$ e per una singola misura: $z_i = f(x_i, y_i)$ e, a livello di scarti:

$$\Delta z = z_i - \hat{z} = f(x_i, y_i) - f(\hat{x}, \hat{y}) \quad \Delta x = x_i - \hat{x} \quad \Delta y = y_i - \hat{y}$$

Sappiamo poi dal calcolo differenziale che: $dz = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$ e con le d che diventano Δ :

$$z_i - \hat{z} = \frac{\partial f}{\partial x} (x_i - \hat{x}) + \frac{\partial f}{\partial y} (y_i - \hat{y}) ; \text{facciamone il quadrato:}$$

$$(z_i - \hat{z})^2 = \left(\frac{\partial f}{\partial x}\right)^2 (x_i - \hat{x})^2 + \left(\frac{\partial f}{\partial y}\right)^2 (y_i - \hat{y})^2 + 2 \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} (x_i - \hat{x})(y_i - \hat{y})$$

$$\text{ed ora prendiamo la media: } E(z_i - \hat{z})^2 = \left(\frac{\partial f}{\partial x}\right)^2 E(x_i - \hat{x})^2 + \left(\frac{\partial f}{\partial y}\right)^2 E(y_i - \hat{y})^2, \quad (12.15)$$

in quanto $E(x_i - \hat{x}) = E(y_i - \hat{y}) = 0$. La (12.15) è: $\sigma^2(z) = \left(\frac{\partial f}{\partial x}\right)^2 \sigma^2(x) + \left(\frac{\partial f}{\partial y}\right)^2 \sigma^2(y)$ e infine:

$$\sigma(z) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma^2(x) + \left(\frac{\partial f}{\partial y}\right)^2 \sigma^2(y)}$$

Riassunto: si assume: $z = \bar{z} = f(\bar{x}, \bar{y})$ e: $s(\bar{z}) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 s^2(\bar{x}) + \left(\frac{\partial f}{\partial y}\right)^2 s^2(\bar{y})}$ e, come al solito:

$$z = \bar{z} \pm s(\bar{z}) \text{ ecc.}$$

In modo più semplicistico, ricordando che l'errore assoluto è $\varepsilon_a = V_{\text{mis}} - V_{\text{vero}}$ e l'errore relativo è $\varepsilon_r = (V_{\text{mis}} - V_{\text{vero}}) / V_{\text{vero}}$, si ha:

-) Con $z=x+y$ ed ε_{ax} , ε_{ay} , ε_{az} gli errori assoluti e ε_{rx} , ε_{ry} , ε_{rz} gli errori relativi, allora:

$z \pm \varepsilon_{az} = x \pm \varepsilon_{ax} + y \pm \varepsilon_{ay} = (x+y) \pm (\varepsilon_{ax} + \varepsilon_{ay})$ da cui: $\varepsilon_{az} = \pm (\varepsilon_{ax} + \varepsilon_{ay})$ e per l'errore relativo:

$$\varepsilon_{rz} = \varepsilon_{az} / z = \pm (\varepsilon_{ax} + \varepsilon_{ay}) / (x+y) = \pm [(\varepsilon_{ax}/x)x + (\varepsilon_{ay}/y)y] / (x+y) = \pm (\varepsilon_{rx}x + \varepsilon_{ry}y) / (x+y).$$

-) Analogamente, con $z=x-y$, si ha: $\varepsilon_{az} = \pm (\varepsilon_{ax} + \varepsilon_{ay})$ e $\varepsilon_{rz} = \pm (\varepsilon_{rx}x + \varepsilon_{ry}y) / (x-y)$.

-) Con $z=xy$, si ha: $z \pm \varepsilon_{az} = (x \pm \varepsilon_{ax})(y \pm \varepsilon_{ay}) = xy \pm x\varepsilon_{ay} \pm y\varepsilon_{ax} \pm \varepsilon_{ax}\varepsilon_{ay} \sim xy \pm x\varepsilon_{ay} \pm y\varepsilon_{ax}$, da cui:

$$\varepsilon_{az} = \pm (x\varepsilon_{ay} + y\varepsilon_{ax}) = \pm [xy(\varepsilon_{ay}/y) + xy(\varepsilon_{ax}/x)] = \pm xy(\varepsilon_{rx} + \varepsilon_{ry}) \text{ e } \varepsilon_{rz} = \pm (\varepsilon_{rx} + \varepsilon_{ry})$$

-) Con $z=x/y$, si ha: $z \pm \varepsilon_{az} = (x \pm \varepsilon_{ax}) / (y \pm \varepsilon_{ay}) = x(1 \pm \varepsilon_{rx}) / y(1 \pm \varepsilon_{ry}) = z(1 \pm \varepsilon_{rx}) / (1 \pm \varepsilon_{ry}) =$

$$= z(1 \pm \varepsilon_{rx} \pm \varepsilon_{ry} \pm \varepsilon_{rx}\varepsilon_{ry}) / (1 - \varepsilon_{ry}^2) \sim z[1 + (\pm \varepsilon_{rx} \pm \varepsilon_{ry})] / (1 - 0) = z \pm z(\varepsilon_{rx} + \varepsilon_{ry}), \text{ da cui: } \varepsilon_{az} = \pm z(\varepsilon_{rx} + \varepsilon_{ry}) \text{ ed infine:}$$

$$\varepsilon_{az} / z = \varepsilon_{rz} = \pm (\varepsilon_{rx} + \varepsilon_{ry}).$$

13) N PASSI IN DIREZIONI CASUALI (Random Walk e monete truccate)

Random Walk:

Sono in piedi al centro di un salone; ad un certo punto decido di compiere $N=10$ passi, ognuno dei quali in direzione casuale. Per semplicità, supponiamo che ogni passo è $r=1$ metro; dopo i dieci passi a che distanza mi trovo dal punto di partenza? Io penso che mi troverò a $R = \sqrt{10} = 3,16m$.

Dopo $N-1$ passi, il vettore che dal punto di partenza indicherà la posizione raggiunta, lo chiameremo $\vec{R}^{N-1}$. Se poi compio l'ennesimo passo $\vec{r}$, si avrà che:

$$\vec{R}^N = \vec{R}^{N-1} + \vec{r} \quad \text{ed il prodotto scalare di } \vec{R}^N \text{ con se stesso fornisce:}$$

$$\vec{R}^N \cdot \vec{R}^N = (R^N)^2 = (R^{N-1})^2 + 2\vec{R}^{N-1} \cdot \vec{r} + r^2 ; \text{prendendo ora la media:}$$

$$\langle (R^N)^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle 2\vec{R}^{N-1} \cdot \vec{r} \rangle + \langle r^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle r^2 \rangle, \quad (13.1)$$

visto che $\langle 2\vec{R}^{N-1} \cdot \vec{r} \rangle = 0$, dal momento che $\vec{r}$ può essere orientato in modo casuale su 360° (se trattasi di un piano), o su 4π sr (steradiani, se mi muovo nello spazio) e dunque un vettore che media con esso, come nella espressione precedente, fornisce un valore nullo.

Riscriviamo allora la (13.1): $\langle (R^N)^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle r^2 \rangle$ e procedendo, su di essa, per induzione, dal momento che (sostituendo N con N-1 e così via):

$$\langle (R^{N-1})^2 \rangle = \langle (R^{N-2})^2 \rangle + \langle r^2 \rangle, \text{ e poi: } \langle (R^{N-2})^2 \rangle = \langle (R^{N-3})^2 \rangle + \langle r^2 \rangle \text{ ecc, si ottiene:}$$

$$\langle (R^N)^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle r^2 \rangle = \langle (R^{N-2})^2 \rangle + 2\langle r^2 \rangle = \dots = 0 + N\langle r^2 \rangle = N\langle r^2 \rangle, \text{ cioè:}$$

$$\langle (R^N)^2 \rangle = N\langle r^2 \rangle, \text{ da cui, estraendo la radice di entrambi i membri:}$$

$$\sqrt{\langle (R^N)^2 \rangle} = R = \sqrt{N} \sqrt{\langle r^2 \rangle} = \sqrt{N} \cdot r, \text{ e cioè:}$$

$$R = \sqrt{N} \cdot r \quad !!! \quad (13.2)$$

Allora, per $N=10$ ed $r=1\text{m}$, si ha $R=3,16\text{m}$.

Random Walk e monete truccate:

Immaginiamo ora un cammino casuale (random walk) del tipo destra sinistra, lungo x:

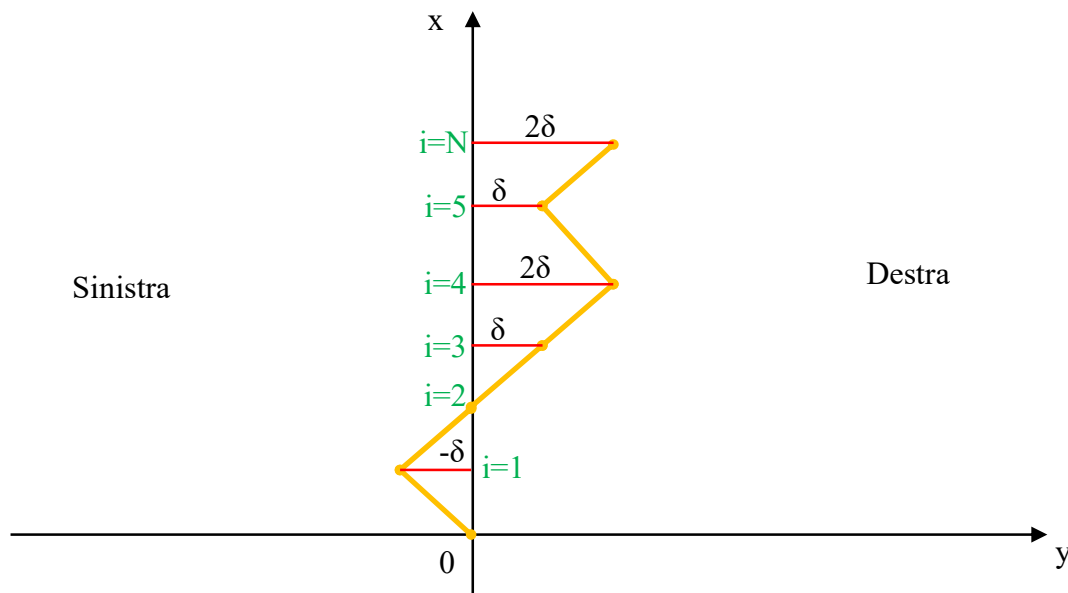


Fig. 13.1: Random Walk destra/sinistra.

Il discostamento rispetto all'asse x pesa, per ogni passo, di una quantità δ (passo p). La media di δ_i è zero: $\langle \delta_i \rangle = 0$, poiché ogni passo può contribuire per $p=-\delta$ o $p=+\delta$, da cui lo zero di media.

La probabilità di compiere un passo verso destra è $P(\delta)=1/2$ e quella di compiere un passo verso sinistra pure è del 50% : $P(-\delta)=1/2$.

Visto che i passi possono essere (a caso) positivi o negativi, per valutare l'impatto della serie di passi compiuti come entità assolute, mi propongo di considerare il quadrato del discostamento i-

esimo dall'asse x, che è sempre positivo, e lo è ad ogni punto/momento "i": $\langle \delta_i^2 \rangle > 0$; ma giunti al punto N, quanto vale $\langle \delta_N^2 \rangle$? Osserviamo innanzitutto che in ogni punto ho un δ_i e $\delta_i = \pm \delta$ (e $\delta_i^2 = \delta^2 \rightarrow \langle \delta_i^2 \rangle = \delta^2$), ovviamente. E posso poi dire che, con metodo induttivo:

$$\delta_N = \delta_{N-1} \pm \delta, \quad (13.3)$$

poiché, ad ogni step, posso andare a destra o a sinistra, con egual probabilità.

Valutiamo il quadrato della (13.3): $\delta_N^2 = \delta_{N-1}^2 + \delta^2 \pm 2\delta_{N-1}\delta$, ossia due valori che, nel computo della media, danno:

$$\langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \langle \delta^2 \rangle \pm 2\delta \langle \delta_{N-1} \rangle, \text{ ossia, mediamente: } \langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \langle \delta^2 \rangle = \langle \delta_{N-1}^2 \rangle + \delta^2; \text{ ribadiamo:}$$

$$\langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \delta^2. \quad (13.4)$$

Dicevamo "mediamente", ma il termine con $\langle \delta_{N-1} \rangle$ sarebbe scomparso comunque, in quanto già stabilimmo che: $\langle \delta_i \rangle = 0$. Allora:

$$\langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \delta^2 = \langle \delta_{N-2}^2 \rangle + \delta^2 + \delta^2 = \langle \delta_{N-2}^2 \rangle + 2\delta^2 = \langle \delta_{N-3}^2 \rangle + 3\delta^2 = \langle \delta_{N-N}^2 \rangle + N\delta^2 = 0 + N\delta^2 = N\delta^2$$

Per avere ora una stima di $\langle \delta_N^2 \rangle = N\delta^2$ estraggo la radice:

$$\delta_N = \sqrt{\langle \delta_N^2 \rangle} = \sqrt{N\delta^2} = \sqrt{N}\delta. \quad (13.5)$$

Si noti la similitudine con la (13.2).

$\sqrt{N}\delta$ significa $\sqrt{N}$ eventi δ , ossia $\sqrt{N}$ eventi.

Ponendo ora, per semplicità, $\delta=1$, possiamo parlare semplicemente di eventi.

Riassumendo: dopo N passi, mi trovo a distanza $\sqrt{N}$, cioè ho deviato di $\sqrt{N}$. Instaurando una analogia con il lancio di una moneta, dopo N lanci, ho ottenuto N_T teste (ed ovviamente $N - N_T = N_C$ croci). Di quanto abbiamo deviato dalla teoria? La teoria dice che devo avere N/2 teste ed N/2 croci, ma ciò è assolutamente vero per un numero infinito di lanci. Nel nostro caso, dove i lanci sono N, l'entità della deviazione è data dai due delta, che agiscono in senso opposto, cioè se ho più teste, allora ho meno croci, e viceversa:

(teste ottenute N_T meno quelle teoriche che avremmo dovuto ottenere, ossia N/2) e

(croci teoriche che avremmo dovuto ottenere, ossia N/2, meno croci ottenute N_C):

$\Delta = (N_T - N/2) + (N/2 - N_C) = N_T - N_C$, e instaurando l'analogia testa/croce con il caso passo a destra/passa a sinistra, abbiamo ottenuto che, per la statistica, questa deviazione Δ deve essere dell'ordine della radice dei passi effettuati $\sqrt{N}$, ossia dell'ordine della radice dei lanci effettuati:

$\Delta = N_T - N_C \rightarrow \sqrt{N}$, ma la somma delle teste e delle croci dà N: $N_T + N_C = N$, da cui: $N_C = N - N_T$ e per la Δ finalmente si ha: $\Delta = N_T - N_C = N_T - (N - N_T) = 2N_T - N \rightarrow \sqrt{N}$; riscriviamola:

$\Delta = 2N_T - N \rightarrow \sqrt{N}$, da cui, per lo scarto verso le sole teste: $N_T - (N/2) \rightarrow \sqrt{N}/2 \rightarrow$

$$N_T = (N/2) \pm \sqrt{N}/2 \quad (13.6)$$

Passando ad un caso concreto, con $N=4900$ lanci, $N_T = (4900/2) \pm \sqrt{4900}/2 = 2450 \pm 35$.

Se il numero di teste si discosta da 2450 per più di 35 casi, allora forse la moneta è truccata.

APPENDICE: SUPPLEMENTI DI STATISTICA

Indice dell'Appendice:

1-SULLA DISTRIBUZIONE NORMALE (GAUSSIANA)

2-STIMA DELLA MEDIA SU TUTTA LA POPOLAZIONE N_p , MA BASATA SU UN CAMPIONE N PICCOLO-LA DISTRIBUZIONE t DI STUDENT

3-DIMOSTRAZIONE DELLA DISTRIBUZIONE t DI STUDENT

4-DISTRIBUZIONE DEL CHI-QUADRO (χ^2) DI HELMERT

5-FUNZIONE GENERATRICE DEI MOMENTI

6-CORRELAZIONE LINEARE

1-Sulla Distribuzione Normale (Gaussiana).

Come già sappiamo e come riportato più sopra, su un campione con un N elevato vige sicuramente la Distribuzione Normale:

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{k - \hat{k}}{\sigma_k} \right)^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{s}{\sigma_k} \right)^2} \quad (\text{Distribuzione Normale o Gaussiana}) \quad (1.1)$$

Essendo la media $\hat{k} = np$ e lo scarto $s = |k - \hat{k}| = |k - np|$ e sapendo che per lo scarto q.m. di k si ha:

$\sigma(k) = \sqrt{\sigma^2(k)} = \sigma_k = \sqrt{npq}$, segue che:

$$P_k = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k - np)^2}{2npq}} \quad (1.2)$$

E definendo la funzione $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(k - np)^2}{2npq}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}$, con $z = \frac{(k - np)}{\sqrt{npq}}$ che è il Coefficiente di Confidenzialità, si ha che $\varphi(z)$ è appunto la “legge normale della probabilità” o “curva degli errori di Gauss”.

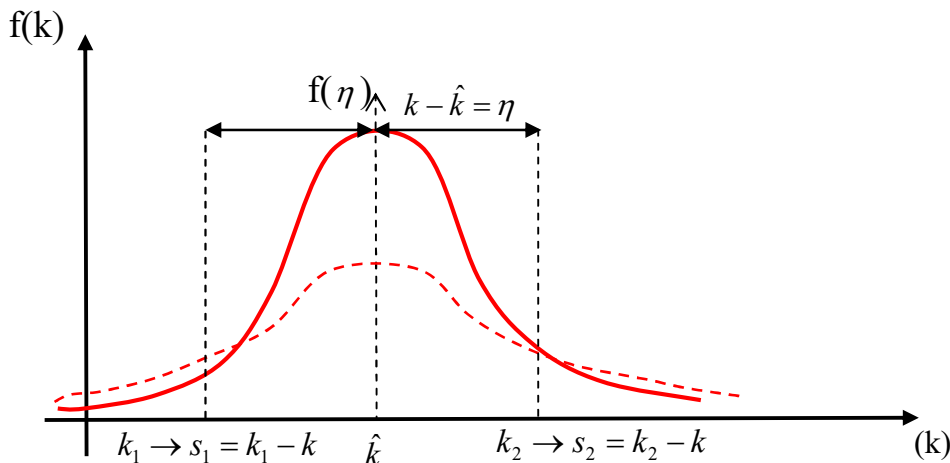


Fig. 1.1: Gaussiana.

Valutiamo ora la probabilità che un evento si verifichi un numero di volte compreso tra k_α e k_β :

$$P_{k_\alpha, k_\beta} = \int_{k_\alpha}^{k_\beta} P_k dk = \int_{k_\alpha}^{k_\beta} \frac{1}{\sigma_k \sqrt{2\pi}} e^{\left[-\frac{(k-np)^2}{2npq}\right]} dk ; \text{ passando ora alla variabile d'integrazione } z:$$

$$P_{k_\alpha, k_\beta} = \int_{z_\alpha}^{z_\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{z_\alpha}^{z_\beta} \phi(z) dz . \text{ Consideriamo ora la funzione:}$$

$$\phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^t \phi(z) dz ; \text{ si ha che:}$$

$$P_{k_\alpha, k_\beta} = \phi\left(\frac{k_\beta - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_\alpha - np}{\sqrt{npq}}\right) . \text{ Esistono tabelle di valori della funzione } \phi(t) .$$

-La regola del 3 σ

Riprendiamo la (1.1):

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\hat{k}}{\sigma_k}\right)^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} = P(s)$$

ora, a livello differenziale, la probabilità dp che lo scarto s sia compreso entro l'intervallo infinitesimo ds è:

$$dp = P(s)ds = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} ds \text{ e, quindi, la probabilità che lo scarto s sia compreso tra i valori } s_1$$

ed s_2 è P:

$$P_{1,2} = \int_{s_1}^{s_2} dp = \int_{s_1}^{s_2} P(s)ds = \int_{s_1}^{s_2} \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} ds \quad (1.3)$$

$P_{1,2}$ è l'area sottesa dalla Gaussiana tra s_1 ed s_2 ; si calcola facilmente, tramite l'integrale della (1.3), che:

$-0,68\sigma < s < 0,68\sigma$	$\rightarrow$	$P=50\%$
$-\sigma < s < \sigma$	$\rightarrow$	$P=68,3\%$
$-1,64\sigma < s < 1,64\sigma$	$\rightarrow$	$P=90\%$
$-2\sigma < s < 2\sigma$	$\rightarrow$	$P=95,9\%$
$-2,58\sigma < s < 2,58\sigma$	$\rightarrow$	$P=99\%$
$-3\sigma < s < 3\sigma$	$\rightarrow$	$P=99,7\%$

Allora, la probabilità che lo scarto dal valore medio di una misura sia compreso tra -3σ e $+3\sigma$ ($\hat{k} - 3\sigma < k < \hat{k} + 3\sigma$) è del 99,7%, ossia un valore piuttosto accettabile, da cui la frequente adozione della regola del tre sigma appunto. In questo caso, il coefficiente di confidenzialità z vale

3 ($z_{0,997} = 3$). Dall'espressione di z, ossia: $z = \frac{(x - \bar{x})}{\sigma}$, abbiamo che:

$$x = \bar{x} \pm z\sigma . \quad (1.4)$$

Se ci riferiamo ora non al σ normale, ma al $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ della media (vedere più sopra), si ha:

$x = \bar{x} \pm z\sigma_{\bar{x}}$, dove N è il numero di elementi del campione prelevato. Se il campione non è prelevato da una popolazione infinita ($N_p = \infty$), ma bensì da una popolazione finita di N_p elementi,

$$\text{allora si ha: } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}} \quad (\text{vedere dimostrazione qui sotto}) \quad (1.5)$$

da cui: $x = \bar{x} \pm z \sigma_{\bar{x}} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}}$; per una popolazione infinita si ha ovviamente che:

$$\sqrt{\frac{(N_p - N)}{(N_p - 1)}} = \sqrt{\frac{(1 - \frac{N}{N_p})}{(1 - \frac{1}{N_p})}} = 1 \quad \text{e: } x = \bar{x} \pm z \sigma_{\bar{x}} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}}$$

Dimostriamo la (1.5):

abbiamo un set di 5 valori $\{3,4,5,6,7\}$ (la cui media ovviamente è $\mu=5$) ed estraiamo un campioncino di due valori ($N_p=5$, $N=2$); tutti i possibili campioncini sono:

(3,4), (3,5), (3,6), (3,7), (4,5), (4,6), (4,7), (5,6), (5,7), (6,7).

I rispettivi valori medi di questi possibili 10 campioncini sono: 3,5|4|4,5|5|4,5|5|5,5|5,5|6|6,5 e la

loro media è: $\mu_{\bar{x}} = \frac{3,5+4+4,5+5+4,5+5+5,5+5,5+6+6,5}{10} = 5$ e la rispettiva deviazione

standard della media è: $\sigma_{\bar{x}} = \sqrt{\frac{(3,5-5)^2 + (4-5)^2 + \dots + (6,5-5)^2}{10}} = \pm 0,866$, mentre la deviazione

standard σ riferita a tutta la popolazione sarebbe:

$$\sigma = \sqrt{\frac{(3-5)^2 + (4-5)^2 + \dots + (7-5)^2}{5}} = \pm 1,414; \text{ notiamo così che la (1.5) è rispettata:}$$

$$\sigma_{\bar{x}} = \pm 0,866 = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}} = \frac{\pm 1,414}{\sqrt{2}} \sqrt{\frac{(5-2)}{(5-1)}}. \text{ E notiamo inoltre che: } \mu_{\bar{x}} = \mu \cdot \text{cvd}$$

2-Stima della media su tutta la popolazione N_p , ma basata su un campione N piccolo- La Distribuzione t di Student.

Un campione inizia ad essere non più piccolo se $N \geq 30$. Infatti, per $N \geq 30$, la distribuzione normale (gaussiana) è ben seguita. Lo stesso non si può dire per appunto $N < 30$.

Quando dunque il campione è piccolo, la statistica seguita non è più quella della distribuzione

normale (1.1), cioè quella del coefficiente di confidenzialità $z = \frac{(x - \bar{x})}{\sigma}$, ed in luogo di

quest'ultimo si usa un'equazione simile, ma a parametro t, basata sulla Distribuzione di Student (di W.S. Gosset) che, come dicevamo, non è quella Normale, essendo N piccolo:

$$t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} \quad (2.1)$$

dove $\bar{x}$ è la media del campioncino di N elementi, μ è la media di tutta la popolazione N_p , σ_N è la deviazione standard del campioncino (N). Sulla base di ciò, in luogo della (1.4), ossia della

$$x = \bar{x} \pm z \sigma, \text{ si ha: } x = \bar{x} \pm \frac{t \sigma_N}{\sqrt{(N-1)}} \quad (2.2)$$

ovviamente, poiché qui lo scopo è stimare la media μ di tutta la popolazione e nella (2.2) x gioca appunto il ruolo di μ nella (2.1).

Si farà questa volta riferimento ad una nuova tabella di valori di t_p ; ad esempio, $t_{0,95}$ fa riferimento ad un livello di confidenzialità del 95%.

In tabella (di Student) i vari t_p sono forniti in funzione del parametro v (nu) che è il numero di gradi

di libertà, ossia, nel nostro caso tipico, per la (2.1) si ha: $t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{v}$, poiché

σ_N e $\bar{x}$ ce li calcoliamo tramite i valori degli N elementi del campioncino, mentre l'incognita è una sola, ossia μ , ossia il valor medio di tutta la popolazione, dunque $v=N-n^\circ \text{ incognite}=N-1$.
Dunque, in generale, $v \cong N$, dacché, in tabella t si forniscono i valori t_p per tutti i v da 1 a 30, poiché, come dicevamo, tale statistica è per $N < 30$, mentre sopra i 30 si forniscono eventualmente solo i t per i v 40, 60, 120 ed ∞ , poiché per i v elevate la statistica torna ad essere quella normale (gaussiana).

3-Dimostrazione della Distribuzione t di Student.

Premesse:

ricordiamo qui il modo di valutare, tramite media pesata, i valori medi $\langle x \rangle$ ed $\langle y \rangle$ e poi anche le varianze $\sigma^2(x)$ e $\sigma^2(y)$, con una distribuzione (di probabilità) $f(x,y)$:

$$\langle x \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy ; \quad \langle y \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy$$

$$\sigma^2(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 f(x,y) dx dy ; \quad \sigma^2(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \langle y \rangle)^2 f(x,y) dx dy$$

Se adesso ho un cambio di coordinate, $x,y \rightarrow u,v$, con: $u=u(x,y)$ e $v=v(x,y)$ (e $x=x(u,v)$ e $y=y(u,v)$), allora, notoriamente:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(u,v) f[x(u,v), y(u,v)] |J| du dv \quad (3.1)$$

dove $|J|$ è il determinante dello Jacobiano (del classico cambio di coordinate in un integrale multiplo):

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}, \text{ essendo: } \begin{cases} x = x(u,v) \\ y = y(u,v) \end{cases} \rightarrow \begin{cases} d\bar{x} = \left(\frac{\partial x}{\partial u}\right) d\bar{u} + \left(\frac{\partial x}{\partial v}\right) d\bar{v} \\ d\bar{y} = \left(\frac{\partial y}{\partial u}\right) d\bar{u} + \left(\frac{\partial y}{\partial v}\right) d\bar{v} \end{cases}; \text{ e infatti in } (x,y) \text{ l'areola è } dx dy,$$

mentre in (u,v) essa è l'area del parallelogramma $d\bar{x} \times d\bar{y}$:

$$d\bar{x} \times d\bar{y} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} du dv = |J| du dv.$$

A volte capita di mediare (integrare) in anticipo su una delle due variabili (ad esempio y) e allora la (3.1) ci lascia con una funzione integranda in sol dx, del tipo:

$$F(y) = \int_{-\infty}^{\infty} f(x,y) dx ; \text{ anche qui, con un cambio di variabili del tipo } (x,y) \rightarrow (x,z/x) \text{ (cioè: } y=z/x \text{ e}$$

$$z=xy), \text{ si ha: } F(z) = \int_{-\infty}^{\infty} f(x, \frac{z}{x}) |J| dx, \text{ con: } \begin{cases} x = x(x,z) \\ y = y(x,z) \end{cases} \rightarrow \begin{cases} d\bar{x} = \left(\frac{\partial x}{\partial x}\right) d\bar{x} + \left(\frac{\partial x}{\partial z}\right) d\bar{z} \\ d\bar{y} = \left[\frac{\partial(z/x)}{\partial x}\right] d\bar{x} + \left[\frac{\partial(z/x)}{\partial z}\right] d\bar{z} \end{cases} \rightarrow ;$$

$$\rightarrow \begin{cases} d\bar{x} = 1d\bar{x} + 0d\bar{z} \\ d\bar{y} = -\frac{z}{x^2} d\bar{x} + \frac{1}{x} d\bar{z} \end{cases} \rightarrow |J| = \begin{vmatrix} 1 & 0 \\ -\frac{z}{x^2} & \frac{1}{x} \end{vmatrix} = \frac{1}{x} \rightarrow F(z) = \int_{-\infty}^{\infty} f(x, \frac{z}{x}) \frac{1}{|x|} dx \text{ e se } x \text{ ed } y \text{ sono variabili}$$

indipendenti, con densità di distribuzione f_1 ed f_2 , allora: $F(z) = \int_{-\infty}^{\infty} f_1(x) f_2\left(\frac{z}{x}\right) \frac{1}{|x|} dx$.

Analogamente, se il cambio di coordinate è: $(x,y) \rightarrow (zy,y)$ (cioè: $z=x/y$), allora:

$$|J| = \begin{vmatrix} \frac{\partial(yz)}{\partial y} & \frac{\partial(yz)}{\partial z} \\ \frac{\partial y}{\partial y} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} z & y \\ 1 & 0 \end{vmatrix} = -y \rightarrow F(z) = \int_{-\infty}^{\infty} f(yz, y) |y| dy \quad e$$

$$F(z) = \int_{-\infty}^{\infty} f_1(yz) f_2(y) |y| dy \quad (3.2)$$

Distribuzione Gamma:

Con il passaggio dalla variabile x alla variabile x^2 si passa dalla Distribuzione Normale alla Distribuzione Gamma ($x \rightarrow y = x^2$). Da ciò discende che $x = \pm\sqrt{y}$ (due valori) e questo significa due cose, ossia che vale solo $y > 0$ e però, nella trasformazione, devo moltiplicare per 2. Ora, la Distribuzione Normale con $\bar{x} = 0$ è:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (3.3)$$

Visto che la probabilità totale deve essere la stessa con entrambe le distribuzioni, deve essere:

$$\int_0^{\infty} f(x) dx = \int_0^{\infty} g(y) dy \rightarrow f(x) dx = g(y) dy \rightarrow f(x) = g(y) \frac{dy}{dx} \quad e \quad g(y) = f(x) \frac{dx}{dy} \quad e:$$

$$\frac{dy}{dx} = 2x = 2\sqrt{y} \quad (\rightarrow y > 0), \text{ da cui (più il coefficiente } x^2 \text{ di cui sopra...)}$$

$$g(y) = 2f(x) \frac{dx}{dy} = 2f[\sqrt{y}] \frac{dx}{dy} = 2\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}}\right) \frac{1}{2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} y^{-\frac{1}{2}} = g(y) \quad (3.4)$$

Tale $g(y)$ è normalizzata; infatti, effettuando l'ulteriore cambio di variabile $z = y/2\sigma^2$, si ha:

$$\left(\frac{dz}{dy} = \frac{1}{2\sigma^2} \quad e \quad dy = (2\sigma^2) dz \right)$$

$$g(z) dz = g(y) dy \rightarrow g(z) = g(y) \frac{dy}{dz} = \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} y^{-\frac{1}{2}}\right) (2\sigma^2) = \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z2\sigma^2}{2\sigma^2}} z^{-\frac{1}{2}} (2\sigma^2)^{-\frac{1}{2}}\right] (2\sigma^2) =$$

$$= \frac{1}{\sqrt{\pi}} e^{-z} z^{-\frac{1}{2}} = g(z) \quad ; \text{ notiamo, dunque, che tale } g(z) \text{ è già normalizzata, ossia ha conservato la}$$

normalizzazione, giustificando peraltro anche il coefficiente 2 aggiunto; infatti:

$$\int_0^{\infty} g(z) dz = \int_0^{\infty} \frac{1}{\sqrt{\pi}} e^{-z} z^{-\frac{1}{2}} dz = \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-z} z^{-\frac{1}{2}} dz = \frac{1}{\sqrt{\pi}} \left(\int_0^{\infty} e^{-z} z^{t-1} dz\right)_{t=1/2} = \frac{\Gamma(1/2)}{\sqrt{\pi}} = 1$$

($\Gamma(t) = \int_0^{\infty} e^{-z} z^{t-1} dz$ è la Funzione Gamma di Eulero; vedere più sopra)

Generalizzando la (3.4) ($g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} y^{-\frac{1}{2}}$), si ha:

$$g(p, \sigma, y) = \frac{1}{(2\sigma^2)^p \Gamma(p)} e^{-\frac{y}{2\sigma^2}} y^{p-1} \quad , (y > 0) \quad (3.5)$$

e la (3.4) ne è un caso particolare ($p=1/2$).

t-Student:

abbiamo visto che col passaggio da x ad x^2 , la distribuzione diventa Gamma; ora, essendo:

$$\sigma_N = \frac{\sqrt{\sum (x_i)^2}}{\sqrt{N}} = \frac{S}{\sqrt{N}} \quad (3.6)$$

ed essendo S una forma quadratica, essa è distribuita come Gamma, ossia con il $g(p, \sigma, y)$ della (3.5). Vogliamo ora capire la distribuzione $h(\sigma_N)$ di $\sigma_N = \frac{S}{\sqrt{N}}$; effettuiamo allora un cambio di

variabile nella (3.5), da y ad $N\sigma_N^2$, ossia $y = N\sigma_N^2$ da cui:

$$dy = 2N\sigma_N d\sigma_N \text{ e } h(\sigma_N) d\sigma_N = g(p, \sigma, y) dy \rightarrow$$

$$h(\sigma_N) = g(p, \sigma, y) \frac{dy}{d\sigma_N} = g(p, \sigma, N\sigma_N^2) 2N\sigma_N \quad (3.7)$$

Inoltre, valutiamo la distribuzione $r(Y)$ di $Y\sqrt{N}$, dove: $Y = (\frac{1}{N} \sum_j x_j) - \mu = \bar{x} - \mu$; essa è la distribuzione normale (di cui godono gli x_j), con varianza (della media-vedere più sopra) σ^2/n e consideriamoci sempre centrati sulla media pari a zero:

$$(\mu=0 \rightarrow Y = (\frac{1}{N} \sum_j x_j)) \quad (3.8)$$

$$r(Y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{NY^2}{2\sigma^2}} \text{ che è la (3.3) in } x. \text{ Procediamo ulteriormente per valutare ora la statistica di}$$

$$\frac{Y\sqrt{N}}{\sigma_N}; \text{ la cosa ci interessa poiché dalla (2.2) si ha } \mu = \bar{x} \pm \frac{t\sigma_N}{\sqrt{(N-1)}}, \text{ da cui:}$$

$$\bar{x} - \mu = Y = \pm \frac{t\sigma_N}{\sqrt{(N-1)}} \cong \frac{t\sigma_N}{\sqrt{N}} \text{ e dunque la quantità } \frac{Y\sqrt{N}}{\sigma_N} \text{ è proprio il } t \text{ introdotto con la (2.2)}$$

appunto. E ricordiamo che la (2.2) è interessante per la sua similitudine con la (1.4). ($Y \propto \sigma_N \cdot t$)

Quindi, utilizziamo la (3.2) che compone le due distribuzioni in yz ed in y , ossia $f_1(x)=f_1(yz)$ ed $f_2(y)$, e nel nostro caso presente, il rapporto $z=x/y$ della (3.2) corrisponde appunto al rapporto

$$\frac{Y\sqrt{N}}{\sigma_N} \text{ e dove la statistica } f_1(\text{di } x=yz) \text{ sarebbe quella di } Y\sqrt{N}, \text{ ossia la } r(Y)$$

$$(r(Y) = r(\sigma_N \cdot t) \rightarrow f_1(yz)),$$

mentre la $f_2(y)$ è la statistica di ciò che corrisponde appunto ad y , ossia σ_N ($y \rightarrow \sigma_N$)

$$(f_2(y) \rightarrow h(\sigma_N)); \text{ applichiamo dunque la (3.2): } f(t) = \int_0^\infty r(\sigma_N \cdot t) \cdot h(\sigma_N) | \sigma_N | d\sigma_N, \text{ ossia:}$$

$$(Y \cong \frac{t\sigma_N}{\sqrt{N}}, p=N/2 \text{ e } y = N\sigma_N^2)$$

$$f_p(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-\frac{NY^2}{2\sigma^2}} g(p, \sigma, N\sigma_N^2) 2N\sigma_N^2 d\sigma_N = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-\frac{\sigma_N^2 t^2}{2\sigma^2}} g(\frac{N}{2}, \sigma, N\sigma_N^2) 2N\sigma_N^2 d\sigma_N =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{(2\sigma^2)^p \Gamma(p)} \int_0^\infty e^{-\frac{\sigma_N^2 t^2}{2\sigma^2}} e^{-\frac{y}{2\sigma^2}} y^{p-1} 2N\sigma_N^2 d\sigma_N =$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{(2\sigma^2)^{\frac{N}{2}} \Gamma(\frac{N}{2})} \int_0^\infty e^{\frac{-\sigma_N^2 t^2}{2\sigma^2}} e^{\frac{-y}{2\sigma^2}} y^{\frac{N}{2}-1} 2N\sigma_N^2 d\sigma_N = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{(2\sigma^2)^{\frac{N}{2}} \Gamma(\frac{N}{2})} \cdot \\
&\cdot \int_0^\infty e^{\frac{-\sigma_N^2 t^2}{2\sigma^2}} e^{\frac{-N\sigma_N^2}{2\sigma^2}} (N\sigma_N^2)^{\frac{N}{2}-1} 2N\sigma_N^2 d\sigma_N = \boxed{\frac{2}{\sqrt{N\pi}\sigma^{N+1}\Gamma(\frac{N}{2})} \left(\frac{N}{2}\right)^{\frac{N}{2}+\frac{1}{2}} \int_0^\infty e^{\frac{-\sigma_N^2(t^2+N)}{2\sigma^2}} \sigma_N^N d\sigma_N = f_N(t)} \quad (3.9)
\end{aligned}$$

e considerando l'ultimo integrale, facciamo in esso il cambio di variabile $z = \frac{\sigma_N^2(t^2 + N)}{2\sigma^2}$:

$$\begin{aligned}
&(((\rightarrow dz = \frac{2(t^2 + N)}{2\sigma^2} \sigma_N d\sigma_N, \quad d\sigma_N = \frac{2\sigma^2}{2(t^2 + N)} \frac{1}{\sigma_N} dz, \quad \sigma_N^2 = \frac{2\sigma^2}{(t^2 + N)} z \quad \text{e} \\
&\sigma_N^N d\sigma_N = \sigma_N^N \left[\frac{2\sigma^2}{2(t^2 + N)} \frac{1}{\sigma_N} dz \right] = \{\sigma_N^{N-1}\} \left[\frac{2\sigma^2}{2(t^2 + N)} dz \right] = \left\{ \left(\frac{2\sigma^2}{(t^2 + N)} z \right)^{\frac{N-1}{2}} \right\} \left[\frac{2\sigma^2}{2(t^2 + N)} dz \right] = \\
&= \frac{1}{2} \left(\frac{2\sigma^2}{t^2 + N} \right)^{(N+1)/2} z^{(N-1)/2} dz)))) \\
&\int_0^\infty e^{\frac{-\sigma_N^2(t^2+N)}{2\sigma^2}} \sigma_N^N d\sigma_N = \frac{1}{2} \left(\frac{2\sigma^2}{t^2 + N} \right)^{(N+1)/2} \int_0^\infty z^{(N-1)/2} e^{-z} dz = \frac{1}{2} \left(\frac{2\sigma^2}{t^2 + N} \right)^{(N+1)/2} \Gamma\left(\frac{N+1}{2}\right) \quad , \quad \text{sicché la (3.9)}
\end{aligned}$$

diventa:

$$f_N(t) = \frac{\Gamma(\frac{N+1}{2})}{\sqrt{\pi N} \Gamma(\frac{N}{2})} (t^2 / N + 1)^{-(N+1)/2} \quad (\text{che è dunque la distribuzione del t di Student}). \quad (3.10)$$

Usiamo dunque la (2.2) per avere una valutazione della media μ di tutta la popolazione, con una confidenzialità data da $f_N(t)$ (eq. (3.10)).

$N=N_{\text{Sample}} - 1 = v$ è il n° di gradi di libertà.

La (3.10) ha la funzione di ripartizione, così come ce l'aveva la distribuzione normale (pag. 20), ed è la seguente:

$$\phi_N(t) = \int_{-\infty}^t f_N(\theta) d\theta = \frac{\Gamma(\frac{N+1}{2})}{\sqrt{\pi N} \Gamma(\frac{N}{2})} \int_{-\infty}^t (\theta^2 / N + 1)^{-(N+1)/2} d\theta .$$

La probabilità che un dato valore t^* di Student sia compreso tra $-t$ e $+t$, ossia che $-t \leq t^* \leq +t$ è:

$$W_N(t) = \int_{-t}^t f_N(\theta) d\theta = 2\phi_N(t) - 1 \quad (3.11)$$

Infatti, $W_N(t)$ è tabulata. Vi è una tabella che dà i vari valori di t in funzione di W_N (al v , o N) e un'altra tabella che dà i vari valori W_N in funzione di t (t e v , o N).

$f_N(t)$ è una funzione pari, ossia $f_N(\theta) = f_N(-\theta)$ e inoltre, per la (3.11) si ha:

$$\begin{aligned}
W_N(t) &= \int_{-t}^t = 2 \int_{-\infty}^t - 1 = 2 \int_{-\infty}^t - \int_{-\infty}^{\infty} = \left(\int_{-\infty}^t + \int_{-\infty}^{\infty} \right) - \int_{-\infty}^{\infty} = \int_{-\infty}^t + \left[\int_{-\infty}^t - \int_{-\infty}^{\infty} \right] = \\
&= \int_{-\infty}^t + \left[\int_{-\infty}^t - \left(\int_{-\infty}^{-t} + \int_{-t}^t + \int_t^{\infty} \right) \right] = \int_{-\infty}^t + \left[\int_{-\infty}^t - \left(\int_{-t}^t + 2 \int_t^{\infty} \right) \right] = \int_{-\infty}^t + \left[\int_{-\infty}^t - \int_{-t}^t - 2 \int_t^{\infty} \right] = \\
&= \int_{-\infty}^t + \left[\int_{-\infty}^{-t} - 2 \int_t^{\infty} \right] = \int_{-\infty}^t + \left[(-2 \int_t^{\infty}) + \int_{-\infty}^{-t} \right] = \int_{-\infty}^t + \left[\left(-\int_t^{\infty} - \int_t^{\infty} \right) + \int_{-\infty}^{-t} \right] = \int_{-\infty}^t + \left[-\int_t^{\infty} \right] =
\end{aligned}$$

$$= \left[\int_{-\infty}^t \right] - \int_t^{\infty} = \left[\int_{-\infty}^{-t} + \int_{-t}^t \right] - \int_t^{\infty} = \int_{-t}^t .$$

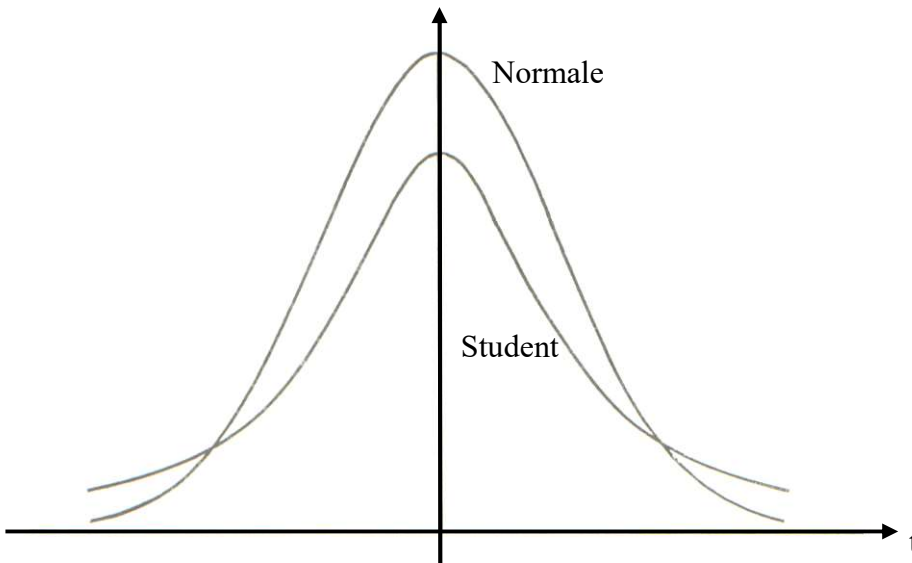


Fig. 3.1-Confronto tra Distribuzioni Normale e di Student.

Come si può vedere dalla Fig. 3.1, la distribuzione di Student è più schiacciata e più slargata; più schiacciata nell'origine poiché i dati maneggiati sono meno probabili, visto l'esiguo numero di elementi N , e più slargata ai lati (ossia lontani dai valori statisticamente meno errati/più probabili) poiché agli errori più grandi viene comunque riconosciuta più probabilità, sempre a causa dell'esiguo N ($N < 30$). Al crescere di N , le due distribuzioni vanno a sovrapporsi (già con $N=40$).

Per concludere, la probabilità che valga la (2.2):

$$\mu \rightarrow x = \bar{x} \pm \frac{t\sigma_N}{\sqrt{(N-1)}}$$

con un determinato t è pari appunto al $W_N(t)$ della (3.11).

Se invece si impone la probabilità (W_N), dalle tabelle di W_N si può evincere per quale t ciò sia vero e da lì, tramite la (2.2), si ha una valutazione della media μ della popolazione (ma con quel W_N di probabilità) e si decide dunque se accettare e proseguire (ad esempio con dei cicli produttivi), o se invece rifiutare ed intervenire (ad esempio con un controllo dei macchinari di produzione).

Tra parentesi, si ha l'Errore di Tipo I quando (erroneamente) rifiuto invece di accettare e si ha l'Errore di Tipo II quando (erroneamente) accetto invece di rifiutare.

Analogamente al caso della Distribuzione Normale, basata sul parametro z , anche nel caso del parametro t di Student, con il campione di N elementi che è ovviamente molto inferiore alla popolazione N_p (se esso è noto) ($N < 30$), si può usare la formula:

$$\mu = \bar{x} + \frac{|t|\sigma_N}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}} . \text{ Quando media e deviazione standard dell'intera popolazione non sono}$$

note, ecco la formula che si può usare come legame tra la deviazione standard del campione σ_N e quella di tutta la popolazione σ :

$$\sigma_N = \sigma \sqrt{\frac{N-1}{N}} , \quad (3.12)$$

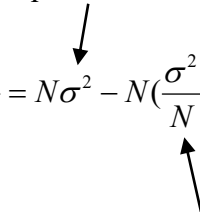
con il fattore $\frac{N-1}{N}$ che è noto come Fattore di Correzione di Bessel.

Dimostrazione (della (3.12)):

siano $\bar{x}$ la media del campioncino e μ quella della popolazione; abbiamo:

$$\begin{aligned}
x_j - \bar{x} &= (x_j - \mu) - (\bar{x} - \mu) \text{ da cui: } (x_j - \bar{x})^2 = (x_j - \mu)^2 - 2(x_j - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2 \text{ e:} \\
\sigma_N^2 &= \sum_{j=1}^N (x_j - \bar{x})^2 = \sum (x_j - \mu)^2 - 2(\bar{x} - \mu) \sum (x_j - \mu) + \sum (\bar{x} - \mu)^2 = \\
&= \sum_{j=1}^N (x_j - \mu)^2 - 2(\bar{x} - \mu) [\sum x_j - N\mu] + N(\bar{x} - \mu)^2 = \sum_{j=1}^N (x_j - \mu)^2 - 2(\bar{x} - \mu) [N\bar{x} - N\mu] + N(\bar{x} - \mu)^2 = \\
&= \sum_{j=1}^N (x_j - \mu)^2 - 2N(\bar{x} - \mu)^2 + N(\bar{x} - \mu)^2 = \sum_{j=1}^N (x_j - \mu)^2 - N(\bar{x} - \mu)^2 \rightarrow \\
&\quad \text{(Calcolato con } \mu \text{ della popolazione, dunque esso è il } \sigma \text{ della popolazione)}
\end{aligned}$$

$$\rightarrow \left\langle \sum_{j=1}^N (x_j - \bar{x})^2 \right\rangle = \left\langle \sum_{j=1}^N (x_j - \mu)^2 \right\rangle - N \left\langle (\bar{x} - \mu)^2 \right\rangle = N\sigma^2 - N \left(\frac{\sigma^2}{N} \right) =$$



 (Il $\sigma_{\bar{x}}^2$ della media, che notoriamente vale $\frac{\sigma^2}{N}$)

allora, proseguendo con i calcoli: $= (N-1)\sigma^2 = N\sigma_N^2$, da cui: $\sigma_N^2 = \frac{(N-1)}{N}\sigma^2$, cvd.

Dunque, una stima di σ dà: $\sigma = \sigma_N \sqrt{\frac{N}{N-1}}$ e per $N \sim 30 \rightarrow \sigma = \sigma_N$ (3.13)

Ma tale (3.13) non può essere usata per campioni piccoli, ovviamente (Student) e si userà comunque la σ data dalla (3.12). Inoltre, $\mu = \bar{x} + \frac{|t|\sigma_N}{\sqrt{N}} \sqrt{\frac{N}{N-1}}$ e per popolazioni non infinite, si

$$\text{ha: } \mu = \bar{x} + \frac{|t|\sigma_N}{\sqrt{N}} \sqrt{\frac{N}{N-1}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}}.$$

Confronto tra due medie aritmetiche sperimentali:

1° provino, n items, media $\bar{x}$, dev.std. σ_x

2° provino, n items, media $\bar{y}$, dev.std. σ_y

Domanda: la differenza $\bar{x} - \bar{y}$ è reale e significativa, oppure è dovuta semplicemente al caso? Cioè, in quel caso, μ_x e μ_y , o $\bar{x}$ ed $\bar{y}$ sono davvero diversi o la loro differenza è dovuta solo al caso?

Come σ_m (medio) considereremo ovviamente:

$$\sigma_m = \sqrt{\frac{(n-1)\sigma_x^2 + (m-1)\sigma_y^2}{(n-1) + (m-1)}} \text{ e } \sigma_m^2 = \sigma_m^2 \left(\frac{1}{n} + \frac{1}{m} \right).$$

A questo punto si fissa la sicurezza statistica W e dalle tabelle si va a vedere per quale valore di t^* essa sussista. Ad esempio, si sceglie un $W=0,05 \rightarrow 5\%$ e si cerca in tabella quel W di 0,05 corrispondente ad un $v=n+m-2$. Ottenuto il t^* , per quanto già sappiamo sulla statistica della distribuzione di Student, si ha: $\mu = \bar{x} + |t|\sigma_{\bar{m}}$ genericamente e, nello specifico nostro:

$$|\mu_x - \mu_y| = |\bar{x} - \bar{y}| + |t|\sigma_{\bar{m}}; \quad (3.14)$$

Visto che l'ipotesi $\mu_x = \mu_y$ nella (3.14) fa sì che: $|t| = \frac{|\bar{x} - \bar{y}|}{\sigma_{\bar{m}}}$, allora l'ipotesi $\mu_x = \mu_y$ è accettata se:

$$\frac{|\bar{x} - \bar{y}|}{\sigma_{\bar{m}}} < t^*, \text{ mentre si ha il rifiuto e } \mu_x \neq \mu_y \text{ se } \frac{|\bar{x} - \bar{y}|}{\sigma_{\bar{m}}} \geq t^*.$$

Nel caso in cui siano invece note le deviazioni standard della popolazione, allora si prende:

$\sigma_D = \sqrt{\sigma_x^2 + \sigma_y^2} = \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$, e in tal caso non ci si rifà più a t, ma a z, poiché il parametro z segue la distribuzione normale e non quella di Student, avendo noi i valori delle σ della popolazione; in tal caso, allora $z = \frac{\bar{x} - \bar{y}}{\sigma_D} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$ e fissato $W(z)$, si ottiene la z e $z > \frac{\bar{x} - \bar{y}}{\sigma_D}$ determina l'accettazione dell'ipotesi $\mu_x = \mu_y$ (proprio come nel caso precedente), mentre il rifiuto è rappresentato dalla situazione: $z \leq \frac{\bar{x} - \bar{y}}{\sigma_D}$.

ESEMPIO 1 di uso della Distribuzione di Student:

Un macchinario ha prodotto rondelle con uno spessore medio di 0,05cm ($\mu=0,05$ cm). Per valutare se il macchinario sta lavorando bene, preleviamo un campione di $N=10$ rondelle, ne misuriamo gli spessori e valutiamo la media, che è, ad esempio, $\bar{x}=0,053$ cm; valutiamo altresì la deviazione standard di questo campioncino, che è:

$$\sigma_N = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_{10} - \bar{x})^2}{(N-1)}} \text{ ed abbiamo, ad esempio, } \sigma_N = 0,003 \text{ cm}.$$

Valutiamo dunque l'affidabilità del macchinario, prima per un livello di significatività del 5% (0,05) e poi dell'1% (0,01).

A)0,05: $N < 30$, dunque non usiamo la distribuzione normale, ma quella di Student:

$$t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} = \frac{(0,053 - 0,05)}{0,003} \sqrt{10-1} = 3,00.$$

Adesso, volendo stare dentro il 5% (0,05), ed essendo che il diametro può discostarsi sia in più che in meno, ho un +2,5% ed un -2,5% di pezzi out, portando il range di accettazione a 97,5% $-(100\% - 2,5\%)$ e $+(100\% - 2,5\%)$, cioè tra -0,975 e +0,975; vado allora in tabella (di Student) e cerco il $t_{0,975}$ corrispondente a $v=N-1=9$ e trovo il valore $t_{0,975}=2,26$; visto che il nostro t è 3, esso sta fuori dal range $(-2,26 \div +2,26)$, dunque rifiuto e faccio controllare il macchinario.

B)0,01: viceversa, tenendomi sull'1%, posso sconfinare in alto o in basso dello 0,5%, dunque ho un range di accettazione di pezzi ok al 99,5% $-(100\% - 0,5\%)$ e $+(100\% - 0,5\%)$ cioè tra -0,995 e +0,995 e allora vado in tabella (di Student) e cerco il $t_{0,995}$ in corrispondenza di un $v=N-1=9$ e trovo $t_{0,995}=3,25$; in conclusione, visto che il nostro $t=3$ sta dentro al range $(-3,25 \div +3,25)$, accetto

lo stato del macchinario (all'1%) e: $\bar{x} - t_{0,995} \frac{\sigma_N}{\sqrt{(N-1)}} < \mu < \bar{x} + t_{0,995} \frac{\sigma_N}{\sqrt{(N-1)}}$.

ESEMPIO 2 di uso della Distribuzione di Student:

Si producono cavi d'acciaio; prelevo un campioncino di $N=6$ cavi e la media del carico di rottura di tali 6 cavi è $\bar{x} = 7750$ kg (ma pretendo un valore di 8000kg), e con una deviazione standard σ_N pari a 145kg.

$$(\sigma_N = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_6 - \bar{x})^2}{(N-1)}} = 145 \text{ kg})$$

Possiamo, sulla base dei risultati fornitici dal nostro campioncino di 6 cavi, tranquillizzare il produttore ad un livello di significatività di A)0,05 (5%) e B)0,01 (1%) ?

In questo caso il produttore è insoddisfatto solo quando $x < 8000$ kg, perché se $x > 8000$ kg, tanto meglio; dunque posso non sconfinare solo in basso (-):

$$t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} = \frac{(7750 - 8000)}{145} \sqrt{6-1} = -3,86 .$$

A) 5% (0,05) Cerco poi in tabella di Student il $t_{0,95}$ in corrispondenza di un $v=6-1=5$ e trovo $t_{0,95}=+2,01$; $(-t_{0,95}=-2,01)$; visto che $t=-3,86 < -t_{0,95}=-2,01$, allora rifiuto e NON rassicuro il produttore.

B) 1% (0,01) $t_{0,99 (v=5)}=+3,36 \rightarrow (-t_{0,99}=-3,36)$; siccome il nostro $t=-3,86 < -t_{0,99}=-3,36$, rifiuto.

4-Distribuzione del CHI-Quadro (χ^2) di Helmert.

Con il lancio di una moneta di $n=100$ volte, la probabilità di testa è $p=1/2$ e dunque la frequenza con cui si presenta testa è $np=50$ volte; l'evento contrario (croce) ha probabilità $q=(1-p)=(1-1/2)=1/2$ e frequenza $nq=50$. Tutto ciò teoricamente; nella pratica potremmo ottenere situazioni un po' diverse.

Ricordiamo il parametro z di pagina 19: $z = \frac{(x - \bar{x})}{\sigma} = \frac{(x - np)}{\sqrt{npq}}$ (4.1)

quadrandolo otteniamo: $z^2 = \frac{(x - np)^2}{npq}$ (4.2)

Notiamo ora subito che la (4.2) può essere così riscritta:

$$z^2 = \frac{(x_1 - np)^2}{np} + \frac{(x_2 - nq)^2}{nq} \quad (4.3)$$

dove x_1 è la variabile casuale associata all'evento desiderato x ed x_2 è la variabile casuale associata all'evento avverso $x_2=n-x_1$; dimostriamo ora la (4.3):

$$\begin{aligned} z^2 &= \frac{(x - np)^2}{np} + \frac{[(n - x) - n(1 - p)]^2}{n(1 - p)} = \frac{(1 - p)(x^2 + n^2 p^2 - 2np x) + p(n^2 p^2 + x^2 - 2np x)}{np(1 - p)} = \\ &= \frac{x^2 + n^2 p^2 - 2np x}{np(1 - p)} = \frac{(x - np)^2}{npq} \quad \text{c.v.d.} \end{aligned}$$

Ora la (4.3) ci suggerisce di considerare come parametro di valutazione della discrepanza tra le frequenze attese (np, nq) / ($np_1, \dots, np_i, \dots, np_n$) e le frequenze osservate ($x_1, x_2, \dots, x_i, \dots, x_n$), il parametro della (4.5) che ora andiamo ad esplicitare:

(frequenza attesa=expected frequency= f_e) (frequenza osservata=observed frequency= f_o) (e ovviamente la somma di tutte le frequenze osservate deve dare n , ossia il totale, ossia: $x_1 + x_2 + \dots + x_n = n \rightarrow \sum_j x_j = n$) (4.4)

$$\chi^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} + \dots + \frac{(x_n - np_n)^2}{np_n} = \sum_{j=1}^n \frac{(x_j - np_j)^2}{np_j} = \sum \frac{(f_o - f_e)^2}{f_e} \quad (4.5)$$

La (4.5) è una forma quadratica. Una forma equivalente della (4.5) è la seguente:

$$\chi^2 = \sum_{j=1}^n \frac{x_j^2}{np_j} - n \quad (4.6)$$

(anch'essa una forma quadratica)

dimostrazione:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^n \frac{(x_j - np_j)^2}{np_j} = \sum_j \left(\frac{x_j^2}{np_j} + \frac{n^2 p_j^2}{np_j} - \frac{2np_j x_j}{np_j} \right) = \sum_j \left(\frac{x_j^2}{np_j} + np_j - 2x_j \right) = \sum_j \frac{x_j^2}{np_j} - \sum_j (2x_j - np_j) = \\ &= \sum_j \frac{x_j^2}{np_j} - \sum_j [(2x_j - 2np_j) + np_j] = \sum_j \frac{x_j^2}{np_j} - 2 \sum_j (x_j - np_j) - \sum_j np_j = \sum_j \frac{x_j^2}{np_j} - 2 \cdot 0 - n = \\ &= \sum_j \frac{x_j^2}{np_j} - n \quad \text{c.v.d.} \quad \text{Infatti, } \sum_j np_j = n \sum_j p_j = n \cdot 1, \text{ poich\'e la somma di tutte le probabilit\'a } p_j\end{aligned}$$

deve dare la probabilit\'a massima, cio\'e 1, mentre: $2 \sum_j (x_j - np_j) = 2 \sum_j x_j - 2 \sum_j np_j$, ma per la (4.4) si ha che $\sum_j x_j = n$, da cui: $2 \sum_j x_j - 2 \sum_j np_j = 2n - 2n \sum_j p_j = 2n - 2n \cdot 1 = 0$.

Dalla (4.5) si evince che se frequenze osservate e frequenze attese coincidono, allora $\chi^2 = 0$, mentre se $\chi^2 > 0$ non coincidono.

Siamo partiti dalla (4.2) e considerando la variabile A: $A = (x - np)$, si ha: $z^2 = \frac{A^2}{\sigma^2}$ e dunque una quantit\'a come quella della (4.5), ossia del tipo:

$$\chi^2 = \sum_{j=1}^v \frac{A_j^2}{\sigma^2} = \gamma, \quad (\gamma > 0, \langle A_j \rangle = 0 \text{ e } \sigma_{A_j}^2 = \sigma^2) \quad (4.7)$$

ha una densit\'a di probabilit\'a Gamma (γ) del tipo della (3.5) (con $p=v/2$ e $y=\gamma\sigma^2$), essendo essa una forma quadratica (σ \u00e8 qui un parametro e non una variabile):

$$d_v(\gamma) = \frac{1}{2^{v/2} \Gamma(v/2)} \gamma^{v/2-1} e^{-\gamma/2} \quad (d_v(\gamma) = 0 \text{ per } \gamma \leq 0) \quad (4.8)$$

infatti, per la (3.5), si ha (con $p=v/2$ e $y=\gamma\sigma^2$):

$$g(p, \sigma, y) = \frac{1}{(2\sigma^2)^p \Gamma(p)} e^{-\frac{y}{2\sigma^2}} y^{p-1} = g\left(\frac{v}{2}, \sigma, \gamma\sigma^2\right) = e^{-\frac{\gamma\sigma^2}{2\sigma^2}} (\gamma^{v/2-1} \sigma^{2(v/2-1)}) \frac{1}{(2\sigma^2)^{v/2} \Gamma(v/2)} |\sigma^2| = d_v(\gamma)$$

($|\sigma^2|$ \u00e8 il determinante dello Jacobiano, dovuto al cambio di variabile $y=\gamma\sigma^2$, come da (3.2))

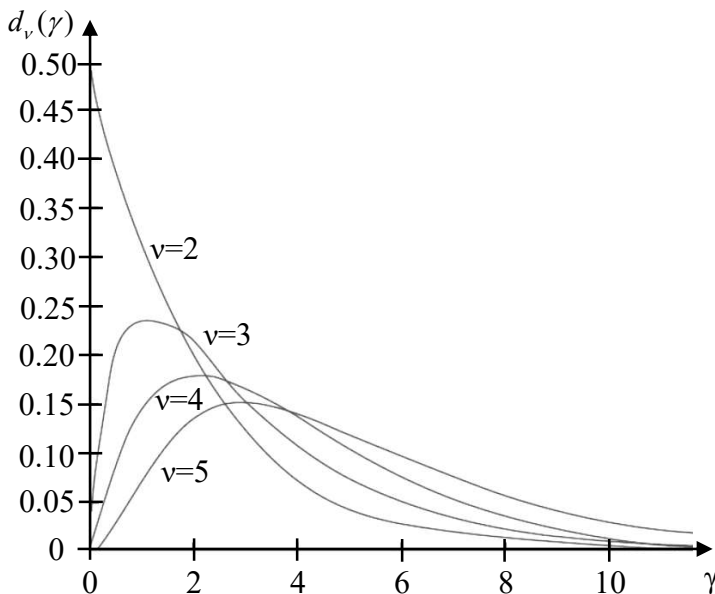


Fig. 4.1: Distribuzione $d_v(\gamma)$ del χ^2 .

Valor medio di χ^2 :

$$\overline{\chi^2} = \bar{\gamma} = \nu \quad (\text{dimostrazione a pag. 33}) \quad (4.9)$$

$$\text{Inoltre: } \sigma_\gamma^2 = \sigma_{\chi^2}^2 = \langle (\gamma - \langle \gamma \rangle)^2 \rangle = \langle (\chi^2 - \langle \chi^2 \rangle)^2 \rangle = 2\nu \rightarrow \quad (4.10)$$

$$\rightarrow \sigma_\gamma = \sqrt{2\nu} \quad (\text{dimostrazione a pag. 33}) \quad (4.11)$$

Al crescere di ν , $d_\nu(\gamma)$ tende alla Distribuzione Normale; infatti, nelle tabelle di $d_\nu(\gamma)$, ν non supera i 30/40. Ovviamente, la probabilità che γ non superi un tal valore γ_k ($\gamma_k < \gamma$) è:

$$D_\nu(\gamma_k) = \int_0^{\gamma_k} d_\nu(\gamma) d\lambda = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \int_0^{\gamma_k} \gamma^{\nu/2-1} e^{-\gamma/2} d\lambda$$

$$\text{La probabilità che } \gamma > \gamma_k \text{ è ovviamente: } P_\nu(\gamma_k) = 1 - D_\nu(\gamma_k) = \int_{\gamma_k}^\infty d_\nu(\gamma) d\lambda = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \int_{\gamma_k}^\infty \gamma^{\nu/2-1} e^{-\gamma/2} d\lambda$$

Saggiare una ipotesi di varianza σ^2 con la Distribuzione del χ^2 :

ipotesi: $\sigma^2 = \sigma_{0x}^2$; poi, ovviamente, si stabilisce anche la probabilità $D_\nu(\gamma_k)$ corrispondente con cui si desidera l'ipotesi. Premettiamo però che ordinariamente: $\sigma_0^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$ (varianza del

campione) e $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (media del campione) e quest'ultimo va ad approssimare μ (della popolazione), mentre la σ^2 (della popolazione) è appunto data dall'ipotesi. Ora, la variabile (vedi anche la (4.7)):

$$\chi^2 = \frac{(n-1)\sigma_0^2}{\sigma^2} = \frac{\nu\sigma_0^2}{\sigma_{0x}^2} \quad (4.12)$$

(σ è della popolazione, σ_0 è del campioncino n e σ_{0x} è quello che suppongo io essere σ)

ha la forma di $\sum_i (x_i - \bar{x})^2$ (per la (4.7)) ed è quindi una forma quadratica ed ha quindi (sempre per la (4.7)) la distribuzione del χ^2 (con $\nu=n-1$). Allora, si cercano in tabella χ^2 i due $\gamma_{1,2}$ (al tal ν)

corrispondenti alla probabilità $D_\nu(\gamma_k)$ e se risulta $\gamma_1 < \frac{\nu\sigma_0^2}{\sigma_{0x}^2} < \gamma_2$, allora l'ipotesi è accettata, altrimenti è rifiutata.

ESEMPIO 1 di uso della Distribuzione del χ^2 :

Su $n=200$ lanci di una moneta, si hanno $x_1=115$ teste e $x_2=85$ croci, ossia due gruppi di risultati ($k=2$). Valutare se la moneta è truccata oppure no, ai soliti livelli di significatività 0,05 (5%) e 0,01 (1%): applichiamo la (4.5):

$$\chi^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} = \frac{(115 - 100)^2}{100} + \frac{(85 - 100)^2}{100} = 4,50.$$

Poi, avendo due gruppi di risultati ($k=2$), il n. di gradi di libertà è: $\nu=k-1=1$.

Caso 1) (0,05 – 5%): ($1-0,05=0,95$) si cerca in tabella χ^2 il $\chi_{0,95}^2$ corrispondente a $\nu=1$ e si trova: $\chi_{0,95}^2 = 3,84$; ora, siccome $4,50 > 3,84$, allora il nostro χ^2 è oltre il $\chi_{0,95}^2$ più verosimile e fornito dalla statistica e dunque il nostro χ^2 più elevato andrebbe ad intercettare una probabilità minore, cioè insufficiente a giustificare l'ipotesi; dunque, respingiamo l'ipotesi che la moneta non sia truccata (5%).

Caso 2) (0,01 – 1%): (1-0,01=0,99) dalla tabella si ha: $\chi^2_{0,99(v=1)} = 6,63$ e $4,50 < 6,63$, allora il nostro χ^2 non è oltre il $\chi^2_{0,99}$ più verosimile e fornito dalla statistica, cioè non lo scavalca, dunque il nostro χ^2 più basso andrebbe ad intercettare una probabilità maggiore, cioè sufficiente a giustificare l'ipotesi; dunque dobbiamo accogliere l'ipotesi che la moneta sia buona (1%).
In totale, traiamo dai due casi il fatto che la, complessivamente, la moneta è probabilmente truccata.

ESEMPIO 2 di uso della Distribuzione del χ^2 :

In passato una macchina impacchettatrice riempiva i pacchetti con 40g ed una deviazione standard $\sigma=0,25$ g. Oggi viene prelevato un campione di $n=20$ pacchetti ($v=n-1=20-1=19$), che mostra una deviazione standard di $\sigma_0=0,32$ g. Ipotesi H_1 : tale aumento di variabilità è significativo (e dunque preoccupante) ai livelli di 5% (0,05) e 1% (0,01)? (oppure, ipotesi H_0 , esso è solo dovuto al caso?)

Con riferimento alla (4.12) si ha: $\chi^2 = \frac{n\sigma_0^2}{\sigma^2} = \frac{20 \cdot 0,32^2}{0,25^2} = 32,8$

(qui, a denominatore abbiamo usato la σ più affidabile che già avevamo dal passato, nonché un utilizzo di n a numeratore, e non $(n-1)$, perché la dev. std. del campione l'abbiamo valutata non con la media scaturente dal campione stesso, ma con la μ più attendibile del passato (40g))

Ora, utilizzeremo un test unilaterale (ad una coda), poiché ci preoccupiamo solo di deviazioni standard maggiori, ossia sconfinanti verso l'alto; le minori, invece, ben vengano.

-Allora, riguardo il livello di significatività del 5%, andiamo sulle tabelle a valutare il $\chi^2_{0,95}$, che è di 30,1 (ad un $v=19$); essendo dunque, per il nostro χ^2 , che: $\chi^2 > \chi^2_{0,95}$, rifiutiamo l'ipotesi H_0 e dunque ci preoccupiamo. Infatti qui il nostro χ^2 sarebbe oltre il $\chi^2_{0,95}$ più verosimile e fornito dalla statistica ed andrebbe dunque a corrispondere ad una probabilità minore, dunque insufficiente.

-Invece, al livello di significatività dell'1% (0,01), si ha che $\chi^2_{0,99(v=19)} = 36,2 > \chi^2 = 32,8$, quindi non rifiutiamo l'ipotesi H_0 e dunque qui ci preoccupiamo di meno.

Un riesame del macchinario, quindi, in generale, andrebbe rifatto.

5-Funzione Generatrice dei Momenti.

Tale funzione generatrice permette ad esempio di calcolare velocemente le medie $\langle x \rangle$ ed $\langle x^2 \rangle$.

Se ho una funzione di distribuzione $f(x)$, allora facendo la media pesata (di peso $f(x) dx$) della funzione e^{tx} , si ha il relativo valor medio: $\langle e^{tx} \rangle = \int e^{tx} f(x) dx = 1 + t\langle x \rangle + \frac{t^2}{2!}\langle x^2 \rangle + \dots$, (5.1)

dopo aver considerato anche lo sviluppo di Taylor di $\langle e^{tx} \rangle$. Ne discende banalmente che:

$$\langle x \rangle = \left. \frac{d}{dt} \langle e^{tx} \rangle \right|_{t=0}, \quad \langle x^2 \rangle = \left. \frac{d^2}{dt^2} \langle e^{tx} \rangle \right|_{t=0}, \quad \dots, \quad \langle x^n \rangle = \left. \frac{d^n}{dt^n} \langle e^{tx} \rangle \right|_{t=0}. \quad (5.2)$$

Abbiamo ora 4 palline in un'urna ed ogni pallina arreca il proprio numero (1 o 2, o 3 o 4). Ora, facendo delle estrazioni di due palline e sommando i relativi numeri, le sei somme che è possibile ottenere sono: 3,4,5 (due volte), 6 e 7; il 5 può essere ottenuto in due modi: 2+3 e 1+4.

Allora, le relative probabilità di comparsa di tali coppie sono: $P(3)=1/6$, $P(4)=1/6$, $P(5)=2 \times 1/6=1/3$, $P(6)=1/6$, $P(7)=1/6$. La Funzione Generatrice dei Momenti allora è, secondo la descrizione (5.1) che abbiamo dato della stessa:

$$M(t) = \frac{1}{6}e^{3t} + \frac{1}{6}e^{4t} + \frac{1}{6} \cdot 2 \cdot e^{5t} + \frac{1}{6}e^{6t} + \frac{1}{6}e^{7t} = \frac{1}{6}(e^{3t} + e^{4t} + 2e^{5t} + e^{6t} + e^{7t})$$

Riguardo le derivate:

$$M'(t) = \frac{1}{6}(3e^{3t} + 4e^{4t} + 10e^{5t} + 6e^{6t} + 7e^{7t}) \quad M''(t) = \frac{1}{6}(9e^{3t} + 16e^{4t} + 50e^{5t} + 36e^{6t} + 49e^{7t})$$

e per le (5.2), si ha: $\langle x \rangle = M'(0) = 5$ e $\langle x^2 \rangle = M''(0) = \frac{80}{3}$.

Avendo invece voluto calcolare $\langle x \rangle$ e $\langle x^2 \rangle$ col modo classico, avremmo ottenuto:

$$\langle x \rangle = \frac{3+4+5+5+6+7}{6} = 5 \text{ (idem)} \quad \text{e} \quad \langle x^2 \rangle = \frac{3^2+4^2+5^2+5^2+6^2+7^2}{6} = \frac{80}{3} \text{ (idem)}$$

Diamo ora una forma a e^{tx} : se la distribuzione è quella del χ^2 , allora si ha che $\langle e^{tx} \rangle = \frac{1}{(1-2t)^{\nu/2}}$;

infatti, la (4.8) ($d_\nu(\gamma) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \gamma^{\nu/2-1} e^{-\gamma/2}$) nella (5.1) fornisce:

$$\langle e^{t\gamma} \rangle = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty e^{t\gamma} \gamma^{\nu/2-1} e^{-\gamma/2} d\gamma = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \gamma^{\nu/2-1} e^{-(1-2t)\frac{\gamma}{2}} d\gamma ; \text{ ponendo ora: } u = (1-2t)\frac{\gamma}{2}, \text{ si ha:}$$

$$\langle e^{t\gamma} \rangle = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \left(\frac{2u}{1-2t}\right)^{(\nu-2)/2} e^{-u} \frac{2du}{(1-2t)} = \frac{(1-2t)^{-\nu/2}}{\Gamma(\nu/2)} \int_0^\infty u^{(\nu-2)/2} e^{-u} du = \frac{(1-2t)^{-\nu/2}}{\Gamma(\nu/2)} \Gamma(\nu/2) = \frac{1}{(1-2t)^{\nu/2}}$$

cvd. Finalmente, applicando le (5.2) si ha: $\langle \gamma \rangle = \langle \chi^2 \rangle = \frac{d}{dt} [(1-2t)^{-\nu/2}] \Big|_{t=0} = \nu$ ossia la (4.9) e

$$\langle \gamma^2 \rangle = \langle (\chi^2)^2 \rangle = \frac{d^2}{dt^2} [(1-2t)^{-\nu/2}] \Big|_{t=0} = \nu(\nu+2), \text{ da cui, per la (4.10):}$$

$$\begin{aligned} \sigma_\gamma^2 &= \sigma_{\chi^2}^2 = \langle (\gamma - \langle \gamma \rangle)^2 \rangle = \langle (\chi^2 - \langle \chi^2 \rangle)^2 \rangle = \langle (\chi^2)^2 \rangle + \langle \langle \chi^2 \rangle^2 \rangle - 2\langle \chi^2 \rangle \langle \chi^2 \rangle = \\ &= \langle (\chi^2)^2 \rangle - 2\langle \chi^2 \rangle^2 = \nu(\nu+2) - \nu^2 = 2\nu \text{ (che è la (4.10) appunto).} \end{aligned}$$

Tra parentesi, se x_1 e x_2 sono due variabili casuali indipendenti, distribuite come χ^2 e con gradi di libertà ν_1 e ν_2 , allora, se $x = x_1 + x_2$, $\rightarrow$

$$\rightarrow \langle e^{tx} \rangle = \langle e^{t(x_1+x_2)} \rangle = \langle e^{tx_1} \rangle \langle e^{tx_2} \rangle = (1-2t)^{-\nu_1/2} (1-2t)^{-\nu_2/2} = (1-2t)^{-(\nu_1+\nu_2)/2}.$$

6-Correlazione lineare.

Mettendo in relazione su un grafico gli scarti $\mathcal{E}(x)$ della variabile x e quelli $\mathcal{E}(y)$ della variabile y , si possono avere almeno questi tre casi:

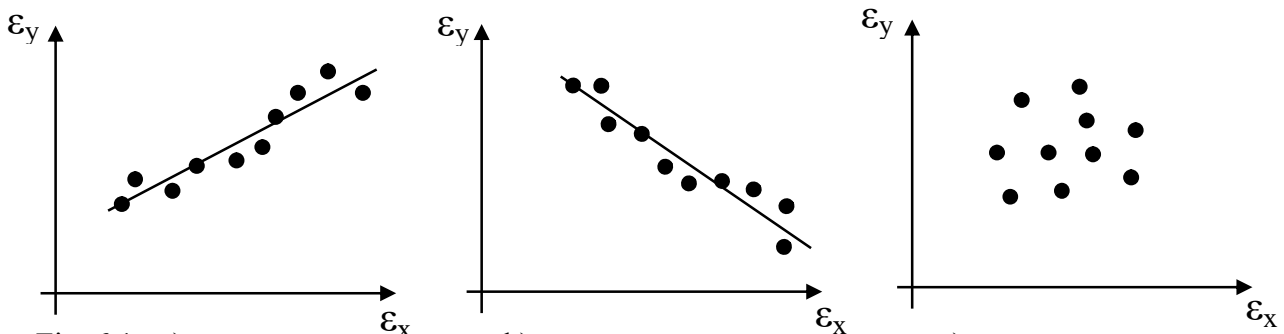


Fig. 6.1 a)

a) Correlazione lineare positiva

b)

b) Correlazione lineare negativa

c)

c) Nessuna correlazione

-Coefficiente di correlazione:

$$r = \frac{\sum [\varepsilon(x)\varepsilon(y)]}{\sqrt{\sum \varepsilon^2(x)}\sqrt{\sum \varepsilon^2(y)}} \quad (r \text{ sta tra } -1 \text{ e } +1) \quad (6.1)$$

Si nota che se i valori si addensano come in a), allora $r \rightarrow +1$ e $r \rightarrow -1$ nel caso b).
 $r \rightarrow 0$ come in c) se non vi è correlazione.

a) e b) ci dicono che correlazione significa equazione lineare:

$$(\varepsilon(y) = a\varepsilon(x) + b) \quad \text{retta di regressione} \quad (6.2)$$

Si ha: $\varepsilon(x) = x - \langle x \rangle$ e $\varepsilon(y) = y - \langle y \rangle$; se la funzione di distribuzione di $\varepsilon(x)$ ed $\varepsilon(y)$ congiunti è $f(\varepsilon(x), \varepsilon(y))$ ed essa è scomponibile/fattorizzabile in $f(\varepsilon(x)) \cdot f(\varepsilon(y))$, ciò che testimonierebbe indipendenza, ossia non correlazione, allora, in tal caso, ci si attende un $r \rightarrow 0$; infatti:

$\varepsilon(x)\varepsilon(y) = [x - \langle x \rangle][y - \langle y \rangle]$ e definendo la COVARIANZA: $\text{cov}(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle$, allora:

$$\begin{aligned} \text{cov}(x, y) &= \int [\varepsilon(x)\varepsilon(y)] f(\varepsilon(x), \varepsilon(y)) d\varepsilon(x) d\varepsilon(y) = \left[\int \varepsilon(x) f(\varepsilon(x)) d\varepsilon(x) \right] \left[\int \varepsilon(y) f(\varepsilon(y)) d\varepsilon(y) \right] = \\ &= \left[\int (x - \langle x \rangle) f(x) dx \right] \left[\int (y - \langle y \rangle) f(y) dy \right] = (\langle x \rangle - \langle x \rangle)(\langle y \rangle - \langle y \rangle) = 0, \text{ e passando dal continuo al } \end{aligned}$$

discreto ($\int \rightarrow \sum$), si ha:

$$\text{cov}(x, y) = \frac{\sum [\varepsilon(x)\varepsilon(y)]}{n} = \frac{\sum [(x - \langle x \rangle)(y - \langle y \rangle)]}{n} = 0 \quad (6.3)$$

(dunque $\text{cov}(x, y) \propto r$, ossia: $r \rightarrow 0$, cioè assenza di correlazione); e continuando con la (6.3):

$$\begin{aligned} \text{cov}(x, y) &= \frac{\sum [\varepsilon(x)\varepsilon(y)]}{n} = \frac{\sum [(x - \langle x \rangle)(y - \langle y \rangle)]}{n} = 0 = \frac{\sum (xy)}{n} - \frac{\sum (x\langle y \rangle)}{n} - \frac{\sum (\langle x \rangle y)}{n} + \frac{\sum (\langle x \rangle \langle y \rangle)}{n} = \\ &= \underline{\langle xy \rangle} - \langle y \rangle \frac{\sum x}{n} - \langle x \rangle \frac{\sum y}{n} + \langle x \rangle \langle y \rangle \frac{\sum 1}{n} = \underline{\langle x \rangle \langle y \rangle} - \langle y \rangle \langle x \rangle - \langle x \rangle \langle y \rangle + \langle x \rangle \langle y \rangle = 0 \end{aligned}$$

(indipendenza $\rightarrow \langle xy \rangle = \langle x \rangle \langle y \rangle$)

La (6.1), ossia l'espressione di r , è una normalizzazione di $\text{cov}(x, y)$ e vale dunque la seguente:

$$r = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} \quad (6.4)$$

$$\text{infatti, ricordando che: } \sigma(x) = \sqrt{\frac{\sum_{i=1}^n (x_i - \langle x \rangle)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (\varepsilon_{x_i})^2}{n}} \rightarrow n\sigma^2(x) = \sum_{i=1}^n (x_i - \langle x \rangle)^2 =$$

$$= \sum_{i=1}^n x_i^2 - 2\langle x \rangle \sum_{i=1}^n x_i + n\langle x \rangle^2 = n\langle x^2 \rangle - 2n\langle x \rangle \langle x \rangle + n\langle x \rangle^2 = n[\langle x^2 \rangle - \langle x \rangle^2] \rightarrow$$

$$\rightarrow \sigma^2(x) = \langle x^2 \rangle - \langle x \rangle^2 \quad (6.5)$$

e allora, per la (6.4):

$$r = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} = \frac{\langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle}{\sqrt{\frac{\sum_{i=1}^n (\varepsilon_{x_i})^2}{n}} \sqrt{\frac{\sum_{i=1}^n (\varepsilon_{y_i})^2}{n}}} = \frac{\sum [\frac{(x - \langle x \rangle)(y - \langle y \rangle)}{n}]}{\sqrt{\frac{\sum \varepsilon_{x_i}^2}{n}} \sqrt{\frac{\sum \varepsilon_{y_i}^2}{n}}} = \frac{\sum [\frac{\varepsilon(x)\varepsilon(y)}{n}]}{\sqrt{\frac{\sum \varepsilon^2(x)}{n}} \sqrt{\frac{\sum \varepsilon^2(y)}{n}}} =$$

$$= \frac{\sum [\varepsilon(x)\varepsilon(y)]}{\sqrt{\sum \varepsilon^2(x)} \sqrt{\sum \varepsilon^2(y)}} = r \quad , \text{ ossia appunto la (6.1).}$$

E proponiamoci altresì di dimostrare che: $-1 \leq r \leq 1$, ossia:

$$-1 \leq \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} \leq 1 \quad (6.6)$$

infatti, se consideriamo la quantità positiva (o, al massimo, nulla):

$$Q = \langle [a(x - \langle x \rangle) + c(y - \langle y \rangle)]^2 \rangle \quad , \text{ si ha che:}$$

$$Q = \langle [a(x - \langle x \rangle) + c(y - \langle y \rangle)]^2 \rangle = \langle a^2(x - \langle x \rangle)^2 + 2ac(x - \langle x \rangle)(y - \langle y \rangle) + c^2(y - \langle y \rangle)^2 \rangle =$$

$$= a^2 \langle (x - \langle x \rangle)^2 \rangle + 2ac \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle + c^2 \langle (y - \langle y \rangle)^2 \rangle = a^2 \sigma^2(x) + 2ac \cdot \text{cov}(x, y) + c^2 \sigma^2(y) \geq 0$$

ed il discriminante è (la variabile è “a”):

$$\Delta = 4c^2 [\text{cov}(x, y)]^2 - 4\sigma^2(x)c^2\sigma^2(y) = 4c^2 \{ [\text{cov}(x, y)]^2 - \sigma^2(x)\sigma^2(y) \} \text{ e quest'ultimo è zero quando}$$

la parabola inizia a stare sopra l'ascissa, da cui: $[\text{cov}(x, y)]^2 - \sigma^2(x)\sigma^2(y) = 0 \rightarrow$

$$\rightarrow \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} = \pm 1 = r \quad \text{cvd.}$$

Inoltre, con la (6.2) intuivamo che: $\varepsilon(y) = a\varepsilon(x) + b$; e infatti adesso ci proponiamo di dimostrare che la probabilità che ciò sia vero è massima; ma $\varepsilon(y) = a\varepsilon(x) + b \rightarrow y - \langle y \rangle = a(x - \langle x \rangle) + b \rightarrow$

$\rightarrow y = ax + [b + \langle y \rangle - a\langle x \rangle] \equiv ax + b' \equiv ax + b$, ossia:

$$y = ax + b \quad (6.7)$$

e vediamo, più semplicemente, che la probabilità che la (6.7) sia vera è massima (P=1):

-innanzitutto osserviamo che: (tenendo conto che $P(y=ax+b)=1$)

$$\langle y \rangle = \langle ax + b \rangle = a\langle x \rangle + b \quad , \text{ da cui, per la (6.5):}$$

$$\sigma^2(y) = \langle y^2 \rangle - \langle y \rangle^2 = \langle (ax + b)^2 \rangle - (a\langle x \rangle + b)^2 = a^2 \langle x^2 \rangle + 2ab\langle x \rangle + b^2 - (a^2 \langle x \rangle^2 + 2ab\langle x \rangle + b^2) =$$

$$= a^2 (\langle x^2 \rangle - \langle x \rangle^2) = a^2 \sigma^2(x) \quad , \text{ da cui:}$$

$$\sigma(y) = \pm a \sigma(x) \quad (6.8)$$

$$\text{inoltre: } \text{cov}(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle = \langle (x - \langle x \rangle)[(ax + b) - (a\langle x \rangle + b)] \rangle =$$

$$= a \langle x^2 + \langle x \rangle^2 - 2x\langle x \rangle \rangle = a (\langle x^2 \rangle - \langle x \rangle^2) = a \sigma^2(x) = [a \sigma(x) \sigma(x)] = \pm \sigma(y) \sigma(x) \rightarrow \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} = \pm 1$$

cvd.



PROBABILITY, STATISTICS AND GAUSSIAN DISTRIBUTION

Leonardo Rubino

May 2024

Abstract: Probability, Statistics and Gaussian Distribution.

Contents:

Contents

General view

- 1) The Bernoulli's Binomial Law (repeated attempts)
 - 2) The Poisson's Law (or Poisson's Distribution)
 - 3) Revising the Combinatory Calculation
 - 4) The Multinomial Distribution
 - 5) The Hypergeometric Distribution
 - 6) The Gaussian Distribution
 - 7) The 3σ Rule
 - 8) A sign from Nature: the Gaussian in Quantum Mechanics
 - 9) The Euler's Gamma Function and the Stirling's Formula
 - 10) A deeper proof of the Gaussian Distribution
 - 11) Conditional events
 - 12) Empirical quantities (from theory to practice)
 - 13) N STEPS IN RANDOM DIRECTIONS (Random Walk and unfair coins)
- APPENDIX: Supplements of statistics

General view

Statistics is likely the base for our Universe, as quantum mechanics itself is based on it. In this paper I give a rare proof of the Gaussian Distribution, as most of available books (in my opinion) just show it, without supplying any satisfactory proof. Moreover I want to give you a suggestion which is useful in every scientific field: do not absolutely trust the intuition, in the statistics and in the probability field. For sure 95 times out of 100 the intuition leads you to correct results and I am a strong user of the intuition; based on the intuition are many scientific discoveries, but every now and then the intuition is misleading. On this purpose, I bring here a case from a TV show called Let's Make a Deal, where there were three closed doors and behind one of them there was a prize; after that the guest chose one of those three doors, the anchorman opened one of the two doors left, in order to show the guest that behind one of those two doors (of course) there wasn't any prize. Of course he opened the door with no prize. Now the guest is asked if he wants to switch to the left unchosen door or if he wants to stay with the first door originally chosen. After a first glance, by relying on one's intuition, it seems that by making the switch no benefit is gotten, as it's all about the choice between two doors, where the success is at 50%. Well, this is the classic case where THE INTUITION IS MISLEADING, as the reasoning just carried out is wrong!

The best choice is always that of leaving the first door you chose and switch! Let's find out why:

let's name as A, B and C the three doors and let's suppose that the guest just chose door A; now:

1-if the prize is in B, the guy will open door C and so, if you switch to B, you'll win.

2-if, on the contrary, the prize is in C, the guy will show you what's behind B and if you switch you will win as well.

3-if the prize is in A, which is the door originally chosen, the guy will open B or C and if you switch you lose.

You can easily understand that the switch brings you to win in two cases over three (1 and 2), while it makes you lose just in one case over three (case 3).

Therefore, it's not true that keeping the first door originally chosen makes no difference, but it's always better to change!

Over the probability of an event and around it.

If on n possible trials an event can happen k times, we say that its probability p is:

$$p = \frac{k}{n} \quad (0 \leq p \leq 1); p=1 \text{ (100\%)} \text{ is the certain event, while } p=0 \text{ is the impossible one. .}$$

The probability of the opposite event is q : $q = \frac{n-k}{n}$. Obviously: $p+q=1$.

If we have more than two events, that is m events, we can say that: $\sum_{i=1}^m p_i = 1$.

Example: in casting a dice, what's the probability that a certain side comes out? $q = \frac{k}{n} = \frac{1}{6}$.

The Frequency of an event.

If we carry out m trials and the event B happens m_B times, the frequency of the event B is:

$$\nu = \frac{m_B}{m}; 0 \leq \nu \leq 1 \text{ and } \sum_{i=1}^m \nu_i = 1.$$

There is a similarity with the probability. The meaning of probability has an abstract (mathematical) origin, while that of frequency has an empirical nature (physical).

The Law of Large Numbers, or Empirical law of Large Numbers or Bernoulli's Law:

(this statement has an experimental nature)

As the number of trials increases, the frequency of an event tends to match its probability.

In easier words: with a large number of trials, the practice tends to the theory. .

The total probability:

If C happens only when even just one of the k events $B_1, \dots, B_k$ happens, excluding one another, and

each with single probabilities $p_1, \dots, p_k$, the probability of C is: $p_C = \sum_{i=1}^k p_i$

Example: I cast a dice; what's the probability that side 1 or side 2 comes out? (just one of them is enough)

$$p_C = \sum_{i=1}^2 p_i = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

The summation of probabilities is a bigger probability.

The composed probability:

If C happens when all events $B_1, B_2, \dots, B_k$ happen, each with probability $p_1, \dots, p_k$, then:

$$p_C = p_1 \cdot p_2 \cdot \dots \cdot p_k = \prod_{i=1}^k p_i$$

Example: I cast three dices; what's the probability that side 5 comes out over all the three dices simultaneously?

$$p_C = \prod_{i=1}^3 p_i = \frac{1}{6} \frac{1}{6} \frac{1}{6} = \frac{1}{216} \text{ The multiplication of probabilities is a smaller probability.}$$

On the meaning of distribution: The set of all probabilities of all values the random variable x can have is the distribution of probability of x .

Mean value of x: $\hat{x} = \frac{\sum x}{n}$. And for a weighed average, that is if every value x_i of x has its own

probability p_i : $\hat{x} = \sum_{i=1}^n x_i p_i$

In the continuum, we do not have single discrete values x_i , but infinite values, like those x can have in a certain range, and summations, of course, become integrals:

if dp is the probability that x lies between ξ and $\xi + d\xi$, then:

$dp = f(\xi)d\xi$, where $f(\xi)$ is the density of probability. Very intuitively, the probability that x lies between a and b is: $P_{a,b} = \int_a^b f(\xi)d\xi$; moreover, as a maximum probability:

$\int_X f(\xi)d\xi = 1$, where X is the field of variability of x .

The DISTRIBUTION FUNCTION $F(t)$ referring to $f(\xi)$ is:

$F(t) = \int_{-\infty}^t f(\xi)d\xi$, so: $P_{a,b} = \int_a^b f(\xi)d\xi = F(b) - F(a)$.

On books there are tables of values of $F(t)$. Still about the continuum, the mean value of x is:

$\hat{x} = \int_X \xi \cdot f(\xi)d\xi$ (average of ξ weighed with weights proportional to $f(\xi)$).

And for the mean function of a function $g(x)$:

$\hat{g}(x) = \sum_{i=1}^n g(x_i) \cdot p_i$ (in the discrete case)

$\hat{g}(x) = \int_X g(\xi) \cdot f(\xi)d\xi$ (in the continuum)

Mean value of the square (discrete case):

$vm_di_x^2 = \sum_{i=1}^n x_i^2 \cdot p_i$

Mean square value:

$vqm = \sqrt{vm_di_x^2}$

The deviation: $\Delta x_i = s = x_i - \hat{x}$

The variance: it is the mean value of the square deviations: $\sigma^2(x) = \sum_{i=1}^n (x_i - \hat{x})^2 \cdot p_i$. In the

continuum, we have: $\sigma^2(x) = \int_X (\xi - \hat{x})^2 \cdot f(\xi)d\xi$

A peculiarity of the variance:

$\sigma^2(x) = vm_di_[(x - \hat{x})^2] = vm_di_ [x^2 - 2x\hat{x} + (\hat{x})^2] = (vm_di_x^2) - 2\hat{x}\hat{x} + (\hat{x})^2 = (vm_di_x^2) - (\hat{x})^2$

The standard deviation:

$\sigma(x) = \sqrt{\sigma^2(x)}$ (it's the root of the variance)

1) The Bernoulli's Binomial Law (Repeated trials)

Let's introduce it through an example: we cast a dice 5 times. What's the probability that a side shows up twice?

($p=1/6$, $q=5/6$, $n=5$, $k=2$)

$$P_k = \binom{n}{k} p^k q^{n-k} \quad (1.1)$$

$$\left(\binom{n}{k} = \frac{n!}{k!(n-k)!} \right) \quad P_2 = \frac{5!}{2!3!} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = 0,16075 = 16,075\%$$

Proof:

We calculate P_0 ($k=0$, that is the event which doesn't happen); the opposite event, whose probability is $(1-p)=q$ happens in all n trials with probability $(1-p)^n=q^n$, so: $P_0 = q^n = \binom{n}{0} p^0 q^{n-0}$

Let's calculate now P_1 , that is the event which happens in the r -th trial with probability p^1 , while in the $(n-1)$ trials left the opposite event happens with probability $q^{(n-1)}$; moreover, r can be chosen out of n different ways; we have:

$$P_1 = np q^{n-1} = \binom{n}{1} p^1 q^{n-1}$$

Let's calculate now P_2 , that is when the event happens in the r -th and in the s -th trial with probability p^2 , while the opposite events have probability $q^{(n-2)}$; but both numbers r and s can be chosen among n numbers and in $\frac{n(n-1)}{2}$ different ways, that is $\binom{n}{2}$ different ways; in fact, the former is chosen in n ways, while for the latter there are $(n-1)$ ways left and as the events are undistinguishable, we have to divide by two, as one could be swapped with the other. Therefore:

$$P_2 = \frac{n(n-1)}{2} p^2 q^{n-2} = \binom{n}{2} p^2 q^{n-2}$$

About P_k , the probability of the positive event is p^k and that of the opposite event is $q^{(n-k)}$ and the k trials can be taken in $\binom{n}{k}$ different combinations, so the proof.

Mean value of k :

$$\hat{k} = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np \quad (1.2)$$

(give n different values and you will see this equation is always true)

In order to prove it, we observe that the summation can start from $k=1$, as for $k=0$ we have a zero value, so it can be removed. Moreover, considering the expression for the Newton's binomial for

$(a+b)^n$, we have: (N.'s binomial.: $(p+q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+1-p)^n = 1$)

$$\hat{k} = \sum_{k=1}^n \frac{n(n-1)!}{(k-1)!(n-k)!} p p^{k-1} q^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} = np(p+q)^{(n-1)} = np$$

The most probable value of k is so np , or the integer closest to it.

Moreover, still in force of the Newton's binomial, we have: $\sum_{k=0}^n P_k = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = 1$

Variance of k:

$$\sigma^2(k) = \text{vm_di_}[(k - \hat{k})]^2 = \text{vm_di_}[(k - np)]^2 = \sum_{k=0}^n (k - \hat{k})^2 \cdot P_k = npq ;$$

in fact,

$$\begin{aligned} \text{vm_di_}(k^2) &= \sum_{k=0}^n \frac{n!}{k!(n-k)!} k^2 p^k q^{n-k} = \sum_{k=0}^n \frac{n!}{k!(n-k)!} [k(k-1) + k] p^k q^{n-k} = \sum_{k=0}^n [k(k-1)] \frac{n!}{k!(n-k)!} p^k q^{n-k} + \\ &+ \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} = \end{aligned}$$

(now, in the last side of the equation, the first summation can start from $k=2$, while the second has been calculated, already; so:)

$$= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} + np = n(n-1)p^2 (p+q)^{n-2} + np = n(n-1)p^2 + np =$$

$$= n^2 p^2 - np^2 + np = \text{vm_di_}(k^2) \text{ and so, after considering that:}$$

$$\sigma^2(k) = \text{vm_di_}[(k - \hat{k})^2] = (\text{vm_di_}(k^2) - (\hat{k})^2), \text{ we have:}$$

$$\sigma^2(k) = (\text{vm_di_}(k^2) - (\hat{k})^2) = (n^2 p^2 - np^2 + np) - n^2 p^2 = np(1-p) = \boxed{npq}$$

The square root of that is, of course, the standard deviation of k:

$$\sigma(k) = \sqrt{\sigma^2(k)} = \sqrt{npq} . \quad (1.3)$$

2) The Poisson's Law (or Distribution)

Let's start from Bernoulli's Law $P_k = \binom{n}{k} p^k (1-p)^{n-k}$, where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and then $\hat{k} = np$;

and let's consider the Stirling's Formula for the factorial $n! = \sqrt{2\pi n} n^n e^{-n}$, from which:

$$\ln n! = \ln[\sqrt{2\pi n} n^n e^{-n}] = n \ln n - n + \frac{1}{2} \ln(2\pi n) \quad (2.1)$$

$$\text{Bernoulli's Law becomes } P_k = \binom{n}{k} \left(\frac{\hat{k}}{n}\right)^k \left(1 - \frac{\hat{k}}{n}\right)^{n-k} . \quad (2.2)$$

From Stirling's Formula we have:

$$\lim_{n \rightarrow \infty} \binom{n}{k} = \lim_{n \rightarrow \infty} \frac{\sqrt{2\pi n} n^n e^{-n}}{k! \sqrt{2\pi(n-k)} (n-k)^{n-k} e^{-(n-k)}} = \lim_{n \rightarrow \infty} \frac{\sqrt{n} n^n e^{-n}}{k! \sqrt{(n-k)} (n-k)^n (n-k)^{-k} e^{-(n-k)}}$$

(it's not been used on $k!$, so far, as it's smaller, but only on $n!$ and $(n-k)!$)

Now, we notice that for large n , $\sqrt{n}$ can be canceled with $\sqrt{(n-k)}$, n^n cancels with $(n-k)^n$, $(n-k)^{-k}$ becomes $(n)^{-k}$ and e^{-n} cancels with $e^{-(n-k)}$, so:

$$\lim_{n \rightarrow \infty} \binom{n}{k} = \lim_{n \rightarrow \infty} \frac{n^k}{k!} \quad (2.3)$$

We also notice that (of course), $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$, and it follows that $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{\frac{n}{\hat{k}}}\right)^{\frac{n}{\hat{k}}} = e$,

$$\text{so: } \lim_{n \rightarrow \infty} \left(1 + \frac{1}{\frac{n}{\hat{k}}}\right)^n = \lim_{n \rightarrow \infty} \left(1 + \frac{\hat{k}}{n}\right)^n = e^{\hat{k}}, \text{ so: } \lim_{n \rightarrow \infty} \left(1 + \frac{(-\hat{k})}{n}\right)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{\hat{k}}{n}\right)^n = e^{-\hat{k}}$$

and finally: $\lim_{n \rightarrow \infty} (1 - \frac{\hat{k}}{n})^{n-k} = \lim_{n \rightarrow \infty} \frac{(1 - \frac{\hat{k}}{n})^n}{(1 - \frac{\hat{k}}{n})^k} = \frac{e^{-\hat{k}}}{1} = e^{-\hat{k}}$ and so (2.2) becomes, for large n:

$$P_k = \binom{n}{k} (\frac{\hat{k}}{n})^k (1 - \frac{\hat{k}}{n})^{n-k} = \frac{n^k}{k!} (\frac{\hat{k}}{n})^k e^{-\hat{k}} = \frac{\hat{k}^k}{k!} e^{-\hat{k}} = P_k ,$$

known as the Poisson's Equation $P_k = \frac{\hat{k}^k}{k!} e^{-\hat{k}}$. (2.4)

About the mean value of k (see above (1.2)), we had:

$$\hat{k} = np \quad (2.5)$$

and for the mean square value (see again above, (1.3)):

$$\sigma(k) (= \sqrt{\hat{k}}) = \sqrt{npq} , \quad (2.6)$$

and we also have, with new p and q in (2.2) and big n :

$$\sigma(k) = \lim_{n \rightarrow \infty} \sqrt{n \frac{\hat{k}}{n} (1 - \frac{\hat{k}}{n})} = \sqrt{\hat{k}} = \sqrt{npq} \quad (2.7)$$

Of course: $\sum_{k=0}^n P_k = \sum_{k=0}^n \frac{\hat{k}^k}{k!} e^{-\hat{k}} = e^{-\hat{k}} \sum_{k=0}^n \frac{\hat{k}^k}{k!} = e^{-\hat{k}} e^{\hat{k}} = 1$

3) Revising the combinatorial calculus

Combination of n elements taken by groups of k : according to what has been said on page 6 on the way we choose two numbers r and s among n , we said those ways were $\binom{n}{2}$.

Therefore, the combination of n elements taken by groups of k is: $C_{n,k} = \binom{n}{k}$

Now, if those groups of k can differ even by the order they appear, we talk about “disposition” of n elements by groups of k :

$D_{n,k} = k! C_{n,k}$ In fact, about all the possible displacements of k elements, we know that by k colours we can have $k!$ flags, by which we have the “permutation” of n elements taken by groups of n :

$$P_{n,n} = D_{n,n} = \frac{n!}{0!} = n!$$

4) The Multinomial Distribution

If we carry out n trials and we claim the event 1, with probability p_1 , happens k_1 times, event 2, with probability p_2 , happens k_2 times and event n , with probability p_n , happens k_n times, the probability that all this happens is:

$$P_{(k_1, k_2, \dots, k_n)} = \frac{n!}{k_1! k_2! \dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}$$

Example: we have a box with 50 balls inside, 20 white, 10 red, 5 black and 15 yellow; calculate the probability P that by extracting 10 balls, they are 3_R, 2_W, 4_B and 1_Y (after any extraction, the ball is put back into the box):

$$P_{(3_R, 2_W, 4_B, 1_Y)} = \frac{10!}{3! 2! 4! 1!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 \left(\frac{1}{10}\right)^4 \left(\frac{3}{10}\right)^1$$

5) The Hypergeometric Distribution

If from the box with N balls, where k of them are red, we extract x balls ($x \leq k$) and every ball extracted is not put back in the box, we have to get rid of the Bernoulli's Law and use the hypergeometric distribution; the possibility to extract in sequence x red balls and $(n-x)$ of a different colour is obviously (the product of all probabilities of extractions):

$$P'(x) = \frac{k}{N} \cdot \frac{k-1}{N-1} \cdots \frac{k-(x-1)}{N-(x-1)} \cdot \frac{N-k}{N-x} \cdot \frac{N-k-1}{N-x-1} \cdots \frac{N-k-(n-x-1)}{N-(n-1)} =$$

$$= \frac{k!}{(k-x)!} \cdot \frac{(N-n)!}{N!} \cdots \frac{(N-k)!}{(N-k-n+x)!} = \frac{k!}{N!} \cdot \frac{(N-n)!}{(k-x)!} \cdot \frac{(N-k)!}{(N-k-n+x)!};$$

the total number of those sequence of balls is the number of combinations of n objects taken x by x :

$$C(n, x) = \binom{n}{x} = \frac{n!}{x!(n-x)!}, \text{ so the total probability to have } x \text{ red balls is:}$$

$$P(x) = C(n, x)P'(x) = \frac{C(k, x)C(N-k, n-x)}{C(N, n)} \quad (\text{hypergeometric law}), \text{ where:}$$

$k=n$. of red balls

$n=n$. of extractions

$N=n$.total of balls

$x=n$. of red balls we want to extract (hypergeometric variable)

6) Gaussian Distribution

Let's take the logarithm of (2.4): $\ln P_k = \ln \left[\frac{\hat{k}^k}{k!} e^{-\hat{k}} \right] = k \ln \hat{k} - \ln k! - \hat{k}$ and according to (2.1), we

$$\text{have: } \ln P_k = \ln \left[\frac{\hat{k}^k}{k!} e^{-\hat{k}} \right] = k \ln \hat{k} - k \ln k + k - \ln \sqrt{2\pi k} - \hat{k}, \quad (6.1)$$

and from that, after taking the derivative:

$$\frac{d}{dk} \ln P_k = \ln \hat{k} - \ln k - k \frac{1}{k} + 1 - \frac{1}{\sqrt{2\pi k}} \frac{1}{2} \frac{1}{\sqrt{2\pi k}} 2\pi = \ln \hat{k} - \ln k - \frac{1}{2k}$$

(6.2)

By making (6.2) equal to zero, we get the value of k where we have a maximum for $\ln P_k$:

$$\frac{d}{dk} \ln P_k = \ln \hat{k} - \ln k - \frac{1}{2k} = 0, \quad (6.3)$$

from which, for large k : $\ln \hat{k} - \ln k = 0$, that is: $k = \hat{k}$. After developing, with Taylor, $\ln P_k$ right around $\hat{k}$, where we had a maximum for $\ln P_k$:

$$\ln P_k = \ln P(k) = \ln P(\hat{k}) + (k - \hat{k}) \left[\frac{d}{dk} \ln P(k) \right]_{k=\hat{k}} + \frac{(k - \hat{k})^2}{2!} \left[\frac{d^2}{dk^2} \ln P(k) \right]_{k=\hat{k}} + \dots =$$

$$= \ln P(k) = \ln P(\hat{k}) + \frac{(k - \hat{k})^2}{2!} \left[\frac{d^2}{dk^2} \ln P(k) \right]_{k=\hat{k}} + \dots, \quad (6.4)$$

as the term with the first derivative is zero, for the (6.3). And from (6.2) we have:

$$\frac{d^2}{dk^2} \ln P(k) = -\frac{1}{k} + \frac{1}{2k^2} \quad \text{and because of that, (6.4) yields (for large } k):$$

$$\ln P(k) = \ln P(\hat{k}) + \frac{(k - \hat{k})^2}{2!} \left[-\frac{1}{\hat{k}} + \frac{1}{2\hat{k}^2} \right] = \ln P(\hat{k}) - \frac{(k - \hat{k})^2}{2\hat{k}} + \frac{(k - \hat{k})^2}{4\hat{k}^2} \quad (6.5)$$

Now, considering that, according to (6.1), we have:

$$\ln P(\hat{k}) = \hat{k} \ln \hat{k} - \hat{k} \ln \hat{k} + \hat{k} - \ln \sqrt{2\pi \hat{k}} - \hat{k} = -\ln \sqrt{2\pi \hat{k}}, \text{ then (6.5) gives:}$$

$$\ln P(k) = -\ln \sqrt{2\pi\hat{k}} - \frac{(k-\hat{k})^2}{2\hat{k}} + \frac{(k-\hat{k})^2}{4\hat{k}^2} \approx -\ln \sqrt{2\pi\hat{k}} - \frac{(k-\hat{k})^2}{2\hat{k}}, \quad (6.6)$$

after having neglected the term with $\hat{k}^2$ in the denominator, as we are dealing with large k (so, also large $\hat{k}$). And according to (6.6), finally:

$$P(k) = \frac{1}{\sqrt{2\pi\hat{k}}} e^{-\frac{(k-\hat{k})^2}{2\hat{k}}} \quad (\text{Normal or Gaussian Distribution}) \quad (6.7)$$

According to (2.7) (used in the denominators of (6.7)) and the (2.5) (used in the numerator of the exponent of (6.7)), as there the comparison is not square like, but rather with k , we have:

$$P_k = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \quad (6.7b)$$

(unfortunately the approximations have been a lot, but all in all a lot as well are the conditions which must be matched by the Gaussian Distribution, such as large n and k)

Let's define now the function $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(k-np)^2}{2npq}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, where $z = \frac{(k-np)}{\sqrt{npq}}$; we have that $\varphi(z)$ is indeed the "normal law of probability" or "Gauss' curve of errors".

After recalling the formulas we got on the deviation ($s = |k - \hat{k}| = |k - np|$) and on the mean square value of k ($\sigma(k) = \sqrt{\sigma^2(k)} = \sigma_k = \sqrt{npq}$), we have:

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\hat{k}}{\sigma_k}\right)^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} \quad (6.8)$$

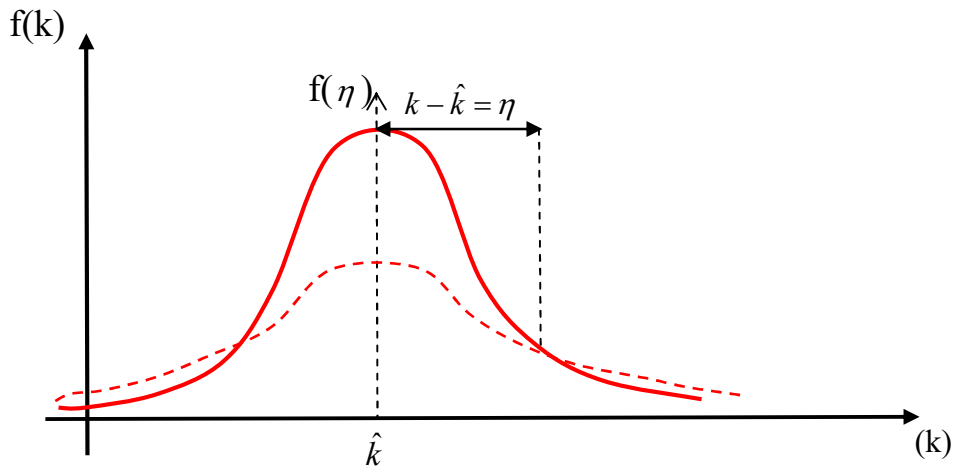


Fig. 6.1: Gaussian.

Let's calculate now the probability that an event happens a number of times between k_α and k_β :

$$P_{k_\alpha, k_\beta} = \int_{k_\alpha}^{k_\beta} P_k dk = \int_{k_\alpha}^{k_\beta} \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(k-np)^2}{2npq}} dk ; \text{ by jumping to the new integration variable } z:$$

$$P_{k_\alpha, k_\beta} = \int_{z_\alpha}^{z_\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{z_\alpha}^{z_\beta} \varphi(z) dz . \text{ Let's analyze now the function:}$$

$$\phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^t \varphi(z) dz ; \text{ we have: } P_{k_\alpha, k_\beta} = \phi(t_\beta) - \phi(t_\alpha) = \phi\left(\frac{k_\beta - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_\alpha - np}{\sqrt{npq}}\right) .$$

There exist tables of values for the function $\phi(t)$.

7) The 3 σ rule

Let's start now from the (6.8):

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{k-\bar{k}}{\sigma_k})^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{s}{\sigma_k})^2} = P(s)$$

now, in a differential form, the probability dp that the deviation s is between the infinitesimal ds is:

$$dp = P(s)ds = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{s}{\sigma_k})^2} ds \text{ and so the probability that the deviation s is between } s_1 \text{ and } s_2 \text{ is P:}$$

$$P_{1,2} = \int_{s_1}^{s_2} dp = \int_{s_1}^{s_2} P(s)ds = \int_{s_1}^{s_2} \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{s}{\sigma_k})^2} ds \quad (7.1)$$

$P_{1,2}$ is the surface covered by the Gaussian curve between s_1 and s_2 ; it's easy to calculate through the integral of (7.1), that:

$$\begin{aligned} -0,68\sigma < s < 0,68\sigma & \rightarrow P=50\% \\ -\sigma < s < \sigma & \rightarrow P=68,3\% \\ -1,64\sigma < s < 1,64\sigma & \rightarrow P=90\% \\ -2\sigma < s < 2\sigma & \rightarrow P=95,9\% \\ -2,58\sigma < s < 2,58\sigma & \rightarrow P=99\% \\ -3\sigma < s < 3\sigma & \rightarrow P=99,7\% \end{aligned}$$

Then, the probability that the deviation of the mean value of a measurement is between -3σ and $+3\sigma$ is of 99,7%, that is a rather acceptable value, from which the usual adoption of the three sigma rule, indeed.

8) A sign from Nature: the Gaussian in Quantum Mechanics

Uncertainty relations:

$$\Delta x \cdot \Delta p \geq \frac{\hbar}{2} \quad \text{and} \quad \Delta E \cdot \Delta t \geq \frac{\hbar}{2} . \quad (8.1)$$

We want to know in which particular case the equality sign holds, that is in which case the product of the two uncertainties (i.e. the two deltas) is the lowest, so the least uncertain, that is the most certain. To have the meaning of all various Δ_F, Δ_G and $\Delta F, \Delta G$ please refer to what follows:

$\langle F \rangle_i = f_i$ (number) ; F would be an "observable". And here is the definition of mean square deviation ΔF for F and we claim it becomes zero:

$$\Delta F = \sqrt{\langle F^2 \rangle_i - \langle F \rangle_i^2} = 0 . \text{ Let's define also the "simple deviation" } \Delta_F: \Delta_F = F - \langle F \rangle_i . \text{ Then, we}$$

have: $\langle \Delta_F^2 \rangle_i = \langle (F - \langle F \rangle_i)^2 \rangle_i = \langle F^2 \rangle_i + \langle F \rangle_i^2 - 2\langle F \rangle_i \langle F \rangle_i = \langle F^2 \rangle_i - \langle F \rangle_i^2 = (\Delta F)^2$. Now, the above request, according to which: $\Delta F = 0$, becomes:

$$\langle \Delta_F^2 \rangle_i = 0 = (\Psi_i, \Delta_F^2 \Psi_i) = 0 . \text{ Moreover, as F is an observable and so hermitian (F* = F), } \Delta_F \text{ as well will be hermitian; then, we can write:}$$

$$\langle \Delta_F^2 \rangle_i = (\Psi_i, \Delta_F^2 \Psi_i) = (\Delta_F \Psi_i, \Delta_F \Psi_i) = \int |\Delta_F \Psi_i|^2 d\xi = 0 , \text{ from which: } \Delta_F \Psi_i = 0 , \text{ that is: } F \Psi_i = f_i \Psi_i , \text{ which is the eigenvalue equation for F.}$$

Let's introduce the operator $O_\lambda = \lambda \Delta_F + i \Delta_G$ ($\lambda \in \Re$); it is obviously non hermitian, as:

$$O_\lambda^\dagger = \lambda \Delta_F - i \Delta_G \neq O_\lambda ; \text{ the modulus of } O_\lambda \Psi \text{ is, by definition of modulus, positive } (O_\lambda \Psi, O_\lambda \Psi) \geq 0 . \text{ Moreover, as } (O_\lambda \Psi, O_\lambda \Psi) = (\Psi, O_\lambda^\dagger O_\lambda \Psi) \text{ and after developing that, we have:}$$

$$(O_\lambda^\dagger O_\lambda = \Delta_F^2 + \Delta_G^2 + i\lambda \Delta_F \Delta_G - i\lambda \Delta_G \Delta_F) , \quad \lambda^2 < \Delta_F^2 > + \lambda < i[\Delta_F, \Delta_G] > + < \Delta_G^2 > \geq 0 , \text{ that is:}$$

$\lambda^2(\Delta F)^2 + \lambda \langle i[F, G] \rangle_\Psi + (\Delta G)^2 \geq 0$; the discriminant of such a square form is negative or zero, because of that kind of inequality, so: $(\langle i[F, G] \rangle_\Psi)^2 - 4(\Delta F)^2(\Delta G)^2 \leq 0$, from which we get the above uncertainty relations (8.1), so: $\Delta F \cdot \Delta G \geq \frac{1}{2} \left| \langle i[F, G] \rangle_\Psi \right|$.

The equality sign holds when the discriminant is zero and when this happens, we get:

$$\lambda_{\min} = -\frac{\langle i[F, G] \rangle_\Psi}{2(\Delta F)^2}; \text{ at that value, we have: } (O_{\lambda_{\min}} \Psi, O_{\lambda_{\min}} \Psi) = 0, \text{ that is: } O_{\lambda_{\min}} \Psi = 0.$$

By developing, in this equation, $O_{\lambda_{\min}}$, we have:

$$\left[\frac{\langle -i[F, G] \rangle}{2(\Delta F)^2} (F - \langle F \rangle) + i(G - \langle G \rangle) \right] \Psi = 0. \quad (8.2)$$

In case $F=x$ and $G=p_x = -i\hbar \frac{\partial}{\partial x}$, we have: $-i[F, G] = -i[x, p_x] = \hbar$, and (8.2) can be so written:

$$\left[\frac{x - \langle x \rangle}{2(\Delta x)^2} + \frac{\partial}{\partial x} - i \frac{\langle p_x \rangle}{\hbar} \right] \Psi = 0, \text{ or: } \frac{d\Psi}{\Psi} = \left[-\frac{x - \langle x \rangle}{2(\Delta x)^2} + i \frac{\langle p_x \rangle}{\hbar} \right] dx, \text{ whose solution is:}$$

$$\Psi(x) = N e^{-\frac{(x - \langle x \rangle)^2}{4(\Delta x)^2}} e^{i \frac{\langle p_x \rangle}{\hbar} x}, \text{ which is a pack of Gaussian waves centered in } \langle x \rangle.$$

9) The Euler's Gamma Function and the Stirling's Formula

$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$. Of course $\Gamma(1) = 1$. Then, $\Gamma(n+1) = n\Gamma(n)$; in fact, after an integration by

parts: $\Gamma(z) = e^{-t} \frac{t^z}{z} \Big|_0^\infty + \frac{1}{z} \int_0^\infty e^{-t} t^z dt = 0 + \frac{1}{z} \Gamma(z+1)$ and by iterating the $\Gamma(n+1) = n\Gamma(n)$, we get:

$$\Gamma(n+1) = n!.$$

Stirling's Formula:

$\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = \int_0^\infty e^{n \ln x - x} dx$ and the function $n \ln x - x$ has a maximum in $x=n$, as you can see by putting to zero its derivative. Now, after saying $x=n+y$, we have:

$$\Gamma(n+1) = e^{-n} \int_{-n}^\infty e^{n \ln(n+y) - y} dy = e^{-n} \int_{-n}^\infty e^{n \ln n + n \ln(1+y/n) - y} dy = n^n e^{-n} \int_{-n}^\infty e^{n \ln(1+y/n) - y} dy$$

Now, according to Taylor, we have that: $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$; we put first $x=y/n$ and later

$$y = \sqrt{n}v, \text{ so: } \Gamma(n+1) = n^n e^{-n} \int_{-n}^\infty e^{n \ln(1+y/n) - y} dy = n^n e^{-n} \int_{-n}^\infty e^{-y^2/2n + y^3/3n^2 - \dots} dy =$$

$$= n^n e^{-n} \sqrt{n} \int_{-\sqrt{n}}^\infty e^{-v^2/2 + v^3/3\sqrt{n} - \dots} dv, \text{ and if } n \text{ is large enough, the second term in the exponent vanishes,}$$

so: $\Gamma(n+1) = \sqrt{2\pi n} n^n e^{-n}$ (Stirling's Formula), as the left Gauss integral was $\sqrt{2\pi}$.

Gauss Integral

Let's take the following two integrals (identical):

$$I = \int_0^\infty e^{-\alpha x^2} dx \quad I = \int_0^\infty e^{-\alpha y^2} dy$$

By multiplying them together: $I^2 = \int_0^\infty \int_0^\infty e^{-\alpha(x^2+y^2)} dx dy$ and in polar coordinates:

($x^2 + y^2 = r^2$, $dS = dx dy = r dr d\theta$), we have:

$$I^2 = \int_0^{\pi/2} \int_0^\infty e^{-\alpha r^2} r dr d\theta = \frac{\pi}{2} \int_0^\infty e^{-\alpha r^2} r dr = -\frac{\pi}{4\alpha} \left| e^{-\alpha r^2} \right|_0^\infty = \frac{\pi}{4\alpha}$$

so: $I = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}}$ and therefore, in the above integral in the proof of the Stirling's Formula, we have

$$(\alpha = 1/2 \text{ and large } n): \Gamma(n+1) = n! = n^n e^{-n} \sqrt{n} \int_{-\infty}^\infty e^{-v^2/2} dv = n^n e^{-n} \sqrt{n} [2 \int_0^\infty e^{-v^2/2} dv] = \sqrt{2\pi n} n^n e^{-n}$$

10) A deeper proof of the Gaussian Distribution

Starting back from the Bernoulli's Distribution:

$$P_x = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x} \text{ and from the Stirling's Formula: } x! \approx (\sqrt{2\pi})^x x^{x+\frac{1}{2}} e^{-x}$$

Now, after recalling that we got np as mean value for k , let's define a $v = np - x$, from which: $x = np - v$ and let's insert such an x in the above Bernoulli's Formula:

$$P_x = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x} = \frac{n!}{(n+v-np)!(np-v)!} p^{(np-v)} q^{(n+v-np)} \cong (\text{for Stirling})$$

[and also because: $(1-p)=q$]

$$\begin{aligned} &\cong \frac{\sqrt{2\pi n}^{n+\frac{1}{2}} e^{-n}}{\sqrt{2\pi}(nq+v)^{(nq+v+\frac{1}{2})} e^{-(nq+v)} \sqrt{2\pi}(np-v)^{(np-v+\frac{1}{2})} e^{-(np+v)}} p^{(np-v)} q^{(nq+v)} = \\ &= \frac{1}{\sqrt{2\pi}} \frac{n^{n+\frac{1}{2}} p^{(np-v)} q^{(nq+v)}}{[nq(1+\frac{v}{nq})]^{nq+v+\frac{1}{2}} [np(1-\frac{v}{np})]^{np-v+\frac{1}{2}}} = \frac{1}{\sqrt{2\pi npq}} \frac{1}{(1+\frac{v}{nq})^{nq+v+\frac{1}{2}} (1-\frac{v}{np})^{np-v+\frac{1}{2}}} \end{aligned}$$

After considering that a^x can be also written as: $a^x = e^{x \ln a}$ and also after considering that, according to Taylor Developments, we have:

$$\ln(1+x) \cong x - \frac{1}{2}x^2 + \dots \approx x, \text{ from which: } \ln(1+\frac{v}{nq}) \approx \frac{v}{nq} \text{ and } \ln(1-\frac{v}{np}) \approx -\frac{v}{np}, \text{ we have:}$$

$$\begin{aligned} P_x &= \frac{n!}{(n-x)!x!} p^x q^{n-x} \cong \frac{1}{\sqrt{2\pi npq}} \exp[-(nq+v+\frac{1}{2})\ln(1+\frac{v}{nq})] \exp[-(np-v+\frac{1}{2})\ln(1-\frac{v}{np})] \cong \\ &\cong \frac{1}{\sqrt{2\pi npq}} \exp[-(nq+v+\frac{1}{2})(\frac{v}{nq} - \frac{v^2}{2n^2q^2})] \exp[-(np-v+\frac{1}{2})(-\frac{v}{np} - \frac{v^2}{2n^2p^2})] \cong \\ &\cong \frac{1}{\sqrt{2\pi npq}} \exp[-\frac{v^2}{2npq} + \frac{v(p-q)}{2npq}] \end{aligned} \quad (10.1)$$

If now we put $p \cong q \cong \frac{1}{2}$ and recall how we define v , we get:

$$P_x = \frac{n!}{(n-x)!x!} p^x q^{n-x} \cong \frac{1}{\sqrt{2\pi npq}} \exp(-\frac{v^2}{2npq}) = \frac{1}{\sqrt{2\pi npq}} \exp[-\frac{(x-np)^2}{2npq}], \text{ that is:}$$

$$P_x = \frac{1}{\sqrt{2\pi npq}} e^{\left[-\frac{(x-np)^2}{2npq}\right]} \quad (\text{Gaussian Distribution}). \quad \text{If, on the contrary, the previous condition}$$

$p \cong q \cong \frac{1}{2}$ is not fully matched, the second term in the exponent of the (10.1) will generate an

asymmetry around the value $v = 0$, or, which is the same, around $x = np$, but such an asymmetry will disappear for $n \rightarrow \infty$, as we have: $\lim_{n \rightarrow \infty} \frac{v(p-q)}{2npq} = 0$.

11) Conditional events

We suppose the event Y can happen only if the event X has happened first and we call $p(Y/X)$ the (conditional) probability that Y happens indeed (after X, of course). Y/X is the event Y with X that has already happened. At last, $p(Y)$ and $p(X)$ are the non conditional probabilities. We can state that $p(Y) = p(XY)$, where XY is intended as an X and Y events together; in fact, if Y happens it's because X, too, has already happened. Then, as per the combined probability (multiplication of probabilities of simultaneous events), we have:

$$p(Y) = p(XY) = p(X)p(Y/X), \quad (11.1)$$

$$\text{from which: } p(Y/X) = p(XY)/p(X). \quad (11.2)$$

If now we write the (11.1) with X and Y swapped, we get: $p(YX) = p(Y)p(X/Y) [=p(XY)]$, so:

$$p(XY) = p(YX) \rightarrow p(X)p(Y/X) = p(Y)p(X/Y) \rightarrow$$

$$p(X/Y) = p(Y/X) \frac{p(X)}{p(Y)} \quad (\text{Bayes' Theorem}) \quad (11.3)$$

Now, out of having a larger generality, we suppose that Y can happen not only if X does, but even if just one of the k events X_i does, (all the X_i exclude one another), that is if X_1 is going to happen, then all the other X_i (with $i \neq 1$) are not happening. Well, now as a combined probability (more favourable, that is given by the summation, as we have various X_i that allow Y) we have:

$$p(Y) = p(X_1)p(Y/X_1) + p(X_2)p(Y/X_2) + \dots + p(X_k)p(Y/X_k) = \sum_{i=1}^k p(X_i)p(Y/X_i) = p(Y) \quad (11.4)$$

which is the Theorem of the total probability for conditional events.

$$\text{Similarly: } p(X) = \sum_{i=1}^k p(Y_i)p(X/Y_i) \quad (11.5)$$

and Bayes' Theorem is a particular case of this more general theorem.

Example: we have 8 boxes of three different kinds: X_1, X_2 and X_3 and every box contains 80 little balls and just 60, 45 and 20 are red (if they are of the kinds X_1, X_2 and X_3 , respectively). So we have 1 box of kind X_1 , 3 of kind X_2 and 4 of kind X_3 . By randomly picking a box (all the 8 boxes have the same probability to be picked) and extracting a small ball from it, what is the probability that such a small ball is red?

Solution: let such an extraction event of a red ball be Y. First of all, a box is picked among the 8 available, and the probabilities per kind (of box) are, of course:

$p(X_1) = 1/8$, $p(X_2) = 3/8$ and $p(X_3) = 4/8$, while the conditional probabilities of extraction of a small red ball from every box, are: $p(Y/X_1) = 60/80$, $p(Y/X_2) = 45/80$ and $p(Y/X_3) = 20/80$.

So, the result is: $p(Y) = (1/8)(60/80) + (3/8)(45/80) + (4/8)(20/80) = 275/640 = 0,43$

12) Empirical quantities (from theory to practice)

In the introduction, on page 3, we talked about the following quantities:

$$\text{Mean value: } \hat{x} = \frac{\sum x}{n} \quad (12.1)$$

$$\text{Deviation: } \Delta x_i = s = x_i - \hat{x} \quad (12.2)$$

Variance: it's the mean value of the square deviations:

$$\sigma^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2 \quad (12.3)$$

Root mean square value:

$$\sigma(x) = \sqrt{\sigma^2(x)} \quad (\text{it is the root of the variance}) \quad (12.4)$$

Moreover, we point out that the mean value (mathematical, theoretical) $\hat{x}$ has a mathematical foundation, theoretical, while if we carry out n measurements of a variable x , that is $x_1, x_2, \dots, x_n$, then we can evaluate an empirical parameter called “empirical mean value” $\bar{x}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n} \quad (12.5)$$

Let's consider all this quantities as operators:

$$\text{Mean value: } = \frac{\sum x}{n} = E(x) \quad (\text{operator } E()) \quad (12.6)$$

and here we have some properties (most of them are obvious) on the mean value:

$$E(x+y) = E(x) + E(y) \quad , \quad E(xy) = E(x)E(y) \quad , \quad E(ax+b) = aE(x) + b \quad , \quad E\left(\frac{x}{a}\right) = \frac{\hat{x}}{a} \quad . \quad \text{Then, after}$$

defining the deviation:

$$\eta = x - \hat{x} = x - E(x) \quad , \quad \text{we have: } E(\eta) = \hat{\eta} = E(x - \hat{x}) = \hat{x} - \hat{x} = 0 \quad .$$

Variance: it is the mean value of the squared deviations:

$$\sigma^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2 = \text{Var}(x) = E(x_i - \hat{x})^2 \quad (\text{operator } \sigma^2()) \quad (12.7)$$

Here are some properties of the variance:

$\sigma^2(x_i) = \sigma^2(x)$ (as well as the mean x is any x 's mean, so also of x_i , then the same stands for the variance)

$$\begin{aligned} \sigma^2(x) &= E(x - \hat{x})^2 = E(x^2 - 2x\hat{x} + (\hat{x})^2) = Ex^2 - 2\hat{x}Ex + (\hat{x})^2 = Ex^2 - 2\hat{x}\hat{x} + (\hat{x})^2 = Ex^2 - (\hat{x})^2 = \\ &= Ex^2 - (Ex)^2 \quad . \quad \text{So, if } x \text{ and } y \text{ are two random and independent variables, then:} \end{aligned}$$

$$\sigma^2(x \pm y) = \sigma^2(x) + \sigma^2(y) \quad ; \quad \text{in fact:}$$

$$\begin{aligned} \sigma^2(x \pm y) &= E[(x_i \pm y_i) - (\hat{x} \pm \hat{y})]^2 = E[(x_i - \hat{x}) \pm (y_i - \hat{y})]^2 = E(x_i - \hat{x})^2 + E(y_i - \hat{y})^2 \pm \\ &\pm 2E[(x_i - \hat{x})(y_i - \hat{y})] = \sigma^2(x) + \sigma^2(y) \pm 2E(x_i - \hat{x})E(y_i - \hat{y}) = \sigma^2(x) + \sigma^2(y) \pm 2 \cdot 0 \cdot 0 = \sigma^2(x) + \sigma^2(y) \end{aligned}$$

$$\sigma^2\left(\frac{x}{a}\right) = E\left(\frac{x}{a} - \frac{\hat{x}}{a}\right)^2 = \frac{1}{a^2} \sigma^2(x) \quad , \quad \sigma^2(x+c) = \sigma^2(x) \quad ,$$

$$\sigma^2(\eta) = E(\eta - \hat{\eta})^2 = E(\eta - 0)^2 = E\eta^2 = E(x - \hat{x})^2 = \sigma^2(x)$$

$$\text{Empirical mean value: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n} = M(x) \quad (\text{operator } M()) \quad (12.8)$$

The empirical mean value must have its own mean value; let's find it out:

$$E(\bar{x}) = E(M(x)) = E\left(\frac{x_1}{n}\right) + E\left(\frac{x_2}{n}\right) + \dots + E\left(\frac{x_n}{n}\right) = \frac{\hat{x}}{n} + \dots + \frac{\hat{x}}{n} = \hat{x}$$

The empirical mean value must have its own variance; let's find it out:

$$\begin{aligned} \sigma^2(\bar{x}) &= \sigma^2\left(\frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n}\right) = \sigma^2\left(\frac{x_1}{n}\right) + \dots + \sigma^2\left(\frac{x_n}{n}\right) = \frac{\sigma^2(x_1)}{n^2} + \dots + \frac{\sigma^2(x_n)}{n^2} = \frac{\sigma^2(x)}{n^2} + \dots + \frac{\sigma^2(x)}{n^2} = \\ &= n \frac{\sigma^2(x)}{n^2} = \frac{\sigma^2(x)}{n} \quad , \quad \text{so:} \end{aligned}$$

$$\boxed{\sigma^2(\bar{x}) = \frac{\sigma^2(x)}{n}} \quad (12.9)$$

$$\text{and finally: } \sigma(\bar{x}) = \sqrt{\sigma^2(\bar{x})} = \frac{\sigma(x)}{\sqrt{n}} \quad (12.10)$$

Empirical variance:

If the variance is: $\sigma^2()$, then the empirical variance will be: $s^2()$. The variance we have met so far referred to a theoretical parameter $\hat{x}$ while, in practice, we easily handle the empirical parameter $\bar{x}$; and such a new (empirical) variance is $s^2(x)$, indeed.

So, we consider the following: $x_i - \bar{x} = (x_i - \hat{x}) - (\bar{x} - \hat{x})$; we square it and carry out the summation

$$\text{over } i: \sum_i (x_i - \bar{x})^2 = \sum_i [(x_i - \hat{x}) - (\bar{x} - \hat{x})]^2 = \sum_i (x_i - \hat{x})^2 - 2 \sum_i [(x_i - \hat{x})(\bar{x} - \hat{x})] + \sum_i (\bar{x} - \hat{x})^2$$

Let's inspect the last three terms:

$$\rightarrow \sum_i (x_i - \hat{x})^2 = \sum_i (x_i - \hat{x})^2 \quad (\text{it stays unchanged})$$

$$\begin{aligned} \rightarrow 2 \sum_i [(x_i - \hat{x})(\bar{x} - \hat{x})] &= 2(\bar{x} - \hat{x}) \sum_i [(x_i - \hat{x})] = 2(\bar{x} - \hat{x}) \sum_i x_i - 2(\bar{x} - \hat{x}) \sum_i \hat{x} = \\ &= 2n\bar{x}(\bar{x} - \hat{x}) - 2n\hat{x}(\bar{x} - \hat{x}) = 2n(\bar{x} - \hat{x})(\bar{x} - \hat{x}) = 2n(\bar{x} - \hat{x})^2 \end{aligned}$$

$$\rightarrow \sum_i (\bar{x} - \hat{x})^2 = \sum_i \text{const} = n \cdot \text{const} = n(\bar{x} - \hat{x})^2$$

$$\text{so: } \sum_i (x_i - \bar{x})^2 = \sum_i (x_i - \hat{x})^2 - n(\bar{x} - \hat{x})^2.$$

Regarding the two terms at the right side of the last equation:

$$\rightarrow \sum_i (x_i - \hat{x})^2 = n\sigma^2(x)$$

$\rightarrow n(\bar{x} - \hat{x})^2 = nE(\bar{x} - \hat{x})^2$, as we saw that the mean value of a constant is the constant itself; moreover: $E(\bar{x} - \hat{x})^2 = \sigma^2(\bar{x})$ (definition of variance applied to $\bar{x}$ instead of a generic x , that is the variance of $\bar{x}$); but we have above calculated the variance of the empirical mean value $\bar{x}$, indeed, and we got: $\sigma^2(\bar{x}) = \frac{\sigma^2(x)}{n}$, from which: $n(\bar{x} - \hat{x})^2 = nE(\bar{x} - \hat{x})^2 = n \frac{\sigma^2(x)}{n} = \sigma^2(x)$, or, as an alternative, we can acknowledge that $(\bar{x} - \hat{x})$ is, by definition, the deviation we have when we supply $\bar{x}$, that is, by definition, $\sqrt{\sigma^2(\bar{x})} = \sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{n}}$, so again: $n(\bar{x} - \hat{x})^2 = n \left(\frac{\sigma(x)}{\sqrt{n}} \right)^2 = n \frac{\sigma^2(x)}{n} = \sigma^2(x)$

$$\text{Let's sum up: } \sum_i (x_i - \bar{x})^2 = n\sigma^2(x) - \sigma^2(x) = (n-1)\sigma^2(x) \quad (12.11)$$

$$\text{that is: } \sigma^2(x) = \frac{n}{n-1} \frac{1}{n} \sum_i (x_i - \bar{x})^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1} \quad \text{and finally:}$$

$$\sigma(x) = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} = s(x) \quad \left(\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \right) \quad (12.12)$$

and we can use this as the best result in order to have a variance; a variance we will call “empirical” ($s^2(x)$), as it was calculated (let's repeat it) by using an empirical parameter $\bar{x}$ in place of a theoretical one ($\hat{x}$). The (12.12) can be held as the empirical error of a measurement.

Now, by using the (12.12) into the (12.11), we get: $(n-1)s^2(x) = (n-1)\sigma^2(x)$, so: $s^2(x) = \sigma^2(x)$, and this will support us once more to use the empirical variance s^2 as an estimate of the variance σ^2 .

At last, the (12.9) and (12.10) lead us to write: $s^2(\bar{x}) = \frac{s^2(x)}{n}$ and $s(\bar{x}) = \frac{s(x)}{\sqrt{n}}$, that is:

$$\frac{s^2(x)}{n} = s^2(\bar{x}) = \frac{\sum_i (x_i - \bar{x})^2}{n(n-1)} \quad \text{and:} \quad (12.13)$$

$$s(\bar{x}) = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n(n-1)}} \quad (12.14)$$

The (12.14) can be held as the empirical mean error of the arithmetic mean.

In practice:

In practice, in an experiment, we assume as the most probable value for x : $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$.

Then, we calculate $s(\bar{x})$ through the (12.14) and finally we supply: $x = \bar{x} \pm s(\bar{x})$ and the relative error in percentage will be: $\varepsilon_r \% = \frac{s(\bar{x})}{\bar{x}} 100$.

Propagation of the errors:

We have $z=f(x,y)$; and we have: $\hat{z} = f(\hat{x}, \hat{y})$ and for a single sampling: $z_i = f(x_i, y_i)$ and about the deviations: $\Delta z = z_i - \hat{z} = f(x_i, y_i) - f(\hat{x}, \hat{y})$ $\Delta x = x_i - \hat{x}$ $\Delta y = y_i - \hat{y}$

We know from calculus that: $dz = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$ and with all the d becoming Δ :

$$z_i - \hat{z} = \frac{\partial f}{\partial x} (x_i - \hat{x}) + \frac{\partial f}{\partial y} (y_i - \hat{y}) ; \text{ let's square it:}$$

$$(z_i - \hat{z})^2 = \left(\frac{\partial f}{\partial x}\right)^2 (x_i - \hat{x})^2 + \left(\frac{\partial f}{\partial y}\right)^2 (y_i - \hat{y})^2 + 2 \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} (x_i - \hat{x})(y_i - \hat{y})$$

$$\text{and now we get the mean value out of it: } E(z_i - \hat{z})^2 = \left(\frac{\partial f}{\partial x}\right)^2 E(x_i - \hat{x})^2 + \left(\frac{\partial f}{\partial y}\right)^2 E(y_i - \hat{y})^2, \quad (12.15)$$

as $E(x_i - \hat{x}) = E(y_i - \hat{y}) = 0$. The (12.15) is: $\sigma^2(z) = \left(\frac{\partial f}{\partial x}\right)^2 \sigma^2(x) + \left(\frac{\partial f}{\partial y}\right)^2 \sigma^2(y)$ and finally:

$$\sigma(z) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma^2(x) + \left(\frac{\partial f}{\partial y}\right)^2 \sigma^2(y)}$$

Summary: we assume: $z = \bar{z} = f(\bar{x}, \bar{y})$ and: $s(\bar{z}) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 s^2(\bar{x}) + \left(\frac{\partial f}{\partial y}\right)^2 s^2(\bar{y})}$ and, as usual:

$z = \bar{z} \pm s(\bar{z})$ etc.

In a more simple way, after considering that the absolute error is $\varepsilon_a = V_{\text{meas}} - V_{\text{true}}$ and the relative error is $\varepsilon_r = (V_{\text{meas}} - V_{\text{true}}) / V_{\text{true}}$, we have:

-) With $z=x+y$ and ε_{ax} , ε_{ay} , ε_{az} as the absolute errors and ε_{rx} , ε_{ry} , ε_{rz} as the relative errors, then: $z \pm \varepsilon_{az} = x \pm \varepsilon_{ax} + y \pm \varepsilon_{ay} = (x+y) \pm (\varepsilon_{ax} + \varepsilon_{ay})$ from which: $\varepsilon_{az} = \pm (\varepsilon_{ax} + \varepsilon_{ay})$ and as a relative error:

$$\varepsilon_{rz} = \varepsilon_{az} / z = \pm (\varepsilon_{ax} + \varepsilon_{ay}) / (x+y) = \pm [(\varepsilon_{ax}/x)x + (\varepsilon_{ay}/y)y] / (x+y) = \pm (\varepsilon_{rx}x + \varepsilon_{ry}y) / (x+y).$$

-) Similarly, with $z=x-y$, we have: $\varepsilon_{az} = \pm (\varepsilon_{ax} + \varepsilon_{ay})$ and $\varepsilon_{rz} = \pm (\varepsilon_{rx}x + \varepsilon_{ry}y) / (x-y)$.

-) With $z=xy$, we have: $z \pm \varepsilon_{az} = (x \pm \varepsilon_{ax})(y \pm \varepsilon_{ay}) = xy \pm x\varepsilon_{ay} \pm y\varepsilon_{ax} \pm \varepsilon_{ax}\varepsilon_{ay} \sim xy \pm x\varepsilon_{ay} \pm y\varepsilon_{ax}$, from which:

$$\varepsilon_{az} = \pm (x\varepsilon_{ay} + y\varepsilon_{ax}) = \pm [xy(\varepsilon_{ay}/y) + xy(\varepsilon_{ax}/x)] = \pm xy(\varepsilon_{rx} + \varepsilon_{ry}) \text{ and } \varepsilon_{rz} = \pm (\varepsilon_{rx} + \varepsilon_{ry})$$

-) With $z=x/y$, we have: $z \pm \varepsilon_{az} = (x \pm \varepsilon_{ax}) / (y \pm \varepsilon_{ay}) = x(1 \pm \varepsilon_{rx}) / y(1 \pm \varepsilon_{ry}) = z(1 \pm \varepsilon_{rx}) / (1 \pm \varepsilon_{ry}) =$

$= z(1 \pm \varepsilon_{rx} \pm \varepsilon_{ry} \pm \varepsilon_{rx}\varepsilon_{ry}) / (1 - \varepsilon_{ry}^2) \sim z[1 + (\pm \varepsilon_{rx} \pm \varepsilon_{ry} \pm 0)] / (1 - 0) = z \pm z(\varepsilon_{rx} + \varepsilon_{ry})$, from which: $\varepsilon_{az} = \pm z(\varepsilon_{rx} + \varepsilon_{ry})$ and finally: $\varepsilon_{az} / z = \varepsilon_{rz} = \pm (\varepsilon_{rx} \pm \varepsilon_{ry})$.

13) N STEPS IN RANDOM DIRECTIONS (Random Walk and unfair coins)

Random Walk:

I'm standing in the middle of a large room; then, I decide to take $N=10$ steps, each of them in a random direction. Out of simplicity, we suppose that every step is $r=1$ metre long; after those ten steps, what is the most probable distance from the starting point I will be? I think I will be, on an average, at $R = \sqrt{10} = 3,16m$. Is it really so? Let's carry out a reasoning about.

After $N-1$ steps, the vector from the starting point and pointing the reached position will be called $\vec{R}^{N-1}$. If then I take the n -th step $\vec{r}$, we have:

$\vec{R}^N = \vec{R}^{N-1} + \vec{r}$ and the scalar product of $\vec{R}^N$ with itself yields:

$\vec{R}^N \cdot \vec{R}^N = (R^N)^2 = (R^{N-1})^2 + 2\vec{R}^{N-1} \cdot \vec{r} + r^2$; now we take the mean value:

$$\langle (R^N)^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle 2\vec{R}^{N-1} \cdot \vec{r} \rangle + \langle r^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle r^2 \rangle, \quad (13.1)$$

as $\langle 2\vec{R}^{N-1} \cdot \vec{r} \rangle = 0$, because $\vec{r}$ can be oriented randomly over 360° , or over 4π sr (steradian, if we move in the space), so a vector averaging with it, as in the previous equation, yields zero.

So we rewrite the (13.1): $\langle (R^N)^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle r^2 \rangle$ and proceeding on it by induction:

(by replacing N with $N-1$ and so on):

$\langle (R^{N-1})^2 \rangle = \langle (R^{N-2})^2 \rangle + \langle r^2 \rangle$, and then: $\langle (R^{N-2})^2 \rangle = \langle (R^{N-3})^2 \rangle + \langle r^2 \rangle$ etc, we get:

$\langle (R^N)^2 \rangle = \langle (R^{N-1})^2 \rangle + \langle r^2 \rangle = \langle (R^{N-2})^2 \rangle + 2\langle r^2 \rangle = \dots = 0 + N\langle r^2 \rangle = N\langle r^2 \rangle$, that is:

$\langle (R^N)^2 \rangle = N\langle r^2 \rangle$, from which, by taking the square roots of both sides:

$\sqrt{\langle (R^N)^2 \rangle} = R = \sqrt{N} \sqrt{\langle r^2 \rangle} = \sqrt{N} \cdot r$, that is:

$$R = \sqrt{N} \cdot r \quad !!! \quad (13.2)$$

Therefore, for $N=10$ and $r=1m$, we have $R=3,16m$.

Random Walk and unfair coins:

We have now a random walk left/right like, along x :

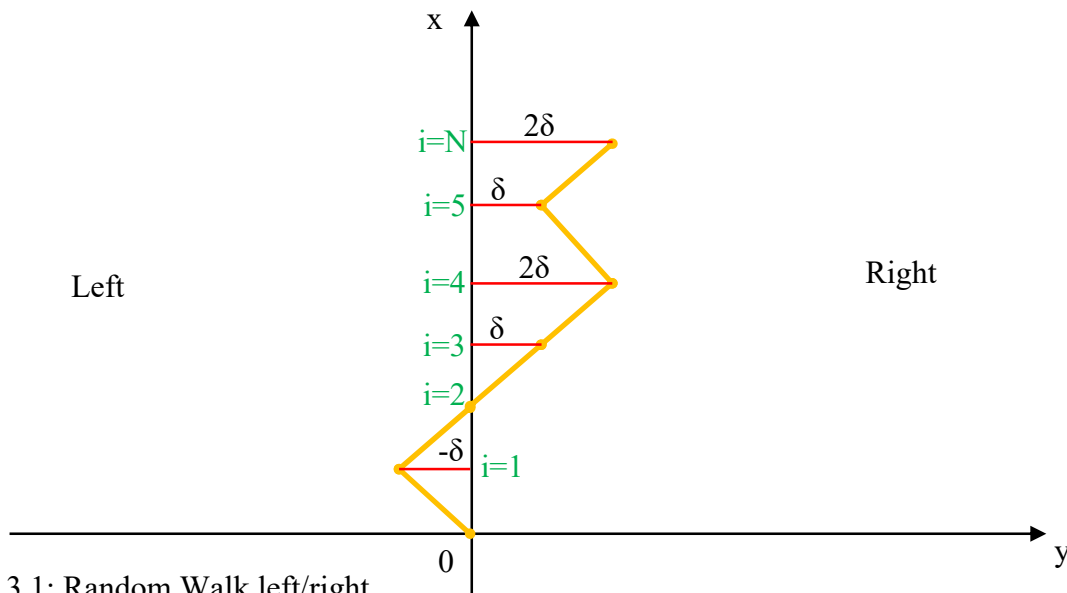


Fig. 13.1: Random Walk left/right.

The deviation with respect to the x axis is, for every step, about an amount δ (step p). The average value of δ_i is zero: $\langle \delta_i \rangle = 0$, as every step can give a contribution about $p = -\delta$ or $p = +\delta$, from which we get that zero.

The probability to take a step to the right is $P(\delta) = 1/2$ and that to take a step to the left is of 50% as well: $P(-\delta) = 1/2$.

As the steps can be (randomly) positive or negative, in order to evaluate the size of the series of steps taken as absolute entities, we will take into account the square value of the i-th deviation from the x axis, which is always positive, and it is so at every point/time step “i”: $\langle \delta_i^2 \rangle > 0$; but once at the N point, how much is $\langle \delta_N^2 \rangle = ?$ First of all, we acknowledge that in every point we have a δ_i and $\delta_i = \pm \delta$ (and $\delta_i^2 = \delta^2 \rightarrow \langle \delta_i^2 \rangle = \delta^2$), of course. And we can say that, through an inductive method:

$$\delta_N = \delta_{N-1} \pm \delta, \quad (13.3)$$

as at every step we can go to the left or to the right by the same probability.

Let's figure out the square value of the (13.3): $\delta_N^2 = \delta_{N-1}^2 + \delta^2 \pm 2\delta_{N-1}\delta$, that is, two values that, in taking the mean value, give:

$$\langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \langle \delta^2 \rangle \pm 2\delta \langle \delta_{N-1} \rangle, \text{ that is, in the average: } \langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \langle \delta^2 \rangle = \langle \delta_{N-1}^2 \rangle + \delta^2;$$

let's recall it: $\langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \delta^2$. (13.4)

We said “in the average”, but the term with $\langle \delta_{N-1} \rangle$ would have disappeared anyway, as we already said that: $\langle \delta_i \rangle = 0$. So:

$$\langle \delta_N^2 \rangle = \langle \delta_{N-1}^2 \rangle + \delta^2 = \langle \delta_{N-2}^2 \rangle + \delta^2 + \delta^2 = \langle \delta_{N-2}^2 \rangle + 2\delta^2 = \langle \delta_{N-3}^2 \rangle + 3\delta^2 = \langle \delta_{N-N}^2 \rangle + N\delta^2 = 0 + N\delta^2 = N\delta^2$$

In order to have an estimation of $\langle \delta_N^2 \rangle = N\delta^2$ we take the square root:

$$\delta_N = \sqrt{\langle \delta_N^2 \rangle} = \sqrt{N\delta^2} = \sqrt{N}\delta. \quad (13.5)$$

Please, notice the similarities with the (13.2). $\sqrt{N}\delta$ means $\sqrt{N}$ events δ , that is $\sqrt{N}$ events.

If now we put, out of simplicity, $\delta=1$, then we can simply talk about events.

In order to sum up things: after N steps, we are at a distance $\sqrt{N}$, that is we have had a deviation about $\sqrt{N}$. Now, if we set up a similarity with a coin tossing, after N tosses we get N_H heads (and of course $N - N_H = N_T$ tails). About how much have we got away from the theory? Theory says we must have $N/2$ heads and $N/2$ tails, but that is absolutely true for an infinite number of tosses. In this case, as we have N tosses, the size of the deviation is given by the two deltas acting in opposite directions one another; in other words, if we have more heads, then we also have less tails, and vice versa. (obtained heads N_H minus those theoretical ones we would have obtained, that is $N/2$) and (theoretical number of tails we would have obtained, that is $N/2$, minus the obtained tails N_T):

$\Delta = (N_H - N/2) + (N/2 - N_T) = N_H - N_T$, and after setting up a similarity between a coin tossing and the case of the random walk, we got from the latter one that, according to the statistics, such a deviation Δ must have the size of the square root of the steps taken $\sqrt{N}$, that is the size of all the coin tosses:

$\Delta = N_H - N_T \rightarrow \sqrt{N}$, but the summation of all heads and tails yields N: $N_H + N_T = N$, so: $N_T = N - N_H$

and because of the Δ , finally we get: $\Delta = N_H - N_T = N_H - (N - N_H) = 2N_H - N \rightarrow \sqrt{N}$; let's write it again:

$\Delta = 2N_H - N \rightarrow \sqrt{N}$, from which, about the deviation of just the heads: $N_H - (N/2) \rightarrow \sqrt{N}/2 \rightarrow$

$$N_H = (N/2) \pm \sqrt{N}/2 \quad (13.6)$$

Talking about a practical case, with $N=4900$ tosses, $N_H = (4900/2) \pm \sqrt{4900}/2 = 2450 \pm 35$.

If the number of heads gets far from 2450 about more than 35 cases, then maybe the coin is unfair.

APPENDIX: SUPPLEMENTS OF STATISTICS

Contents of the Appendix:

1-ON THE NORMAL DISTRIBUTION (GAUSSIAN)

2-ESTIMATION OF THE MEAN VALUE OVER THE WHOLE POPULATION N_P , BUT BASED ON A SMALL SAMPLE N – THE **STUDENT’S t** DISTRIBUTION

3-PROOF OF THE STUDENT’S t DISTRIBUTION

4-THE **CHI-SQUARE** (χ^2) DISTRIBUTION (OF HELMERT)

5-MOMENT GENERATING FUNCTION

6-LINEAR CORRELATION

1-On the Normal Distribution (Gaussian).

As we already know and as is above reported, over a sample with a high N , the Normal Distribution is in force:

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{k - \hat{k}}{\sigma_k} \right)^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{s}{\sigma_k} \right)^2} \quad (\text{Normal Distribution or Gaussian}) \quad (1.1)$$

As the mean value is $\hat{k} = np$ and the deviation $s = |k - \hat{k}| = |k - np|$ and as we know that for the mean square value of k we have: $\sigma(k) = \sqrt{\sigma^2(k)} = \sigma_k = \sqrt{npq}$, then we have:

$$P_k = \frac{1}{\sqrt{2\pi npq}} e^{\left[-\frac{(k - np)^2}{2npq} \right]} \quad (1.2)$$

Let’s define now the function $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{\left[-\frac{(k - np)^2}{2npq} \right]} = \frac{1}{\sqrt{2\pi}} e^{\left[-\frac{1}{2} z^2 \right]}$, where $z = \frac{(k - np)}{\sqrt{npq}}$ is the

Normal Std. Variate Value (Confidence Coefficient); we have that $\varphi(z)$ is indeed the “normal law of probability” or “Gauss’ curve of errors”.

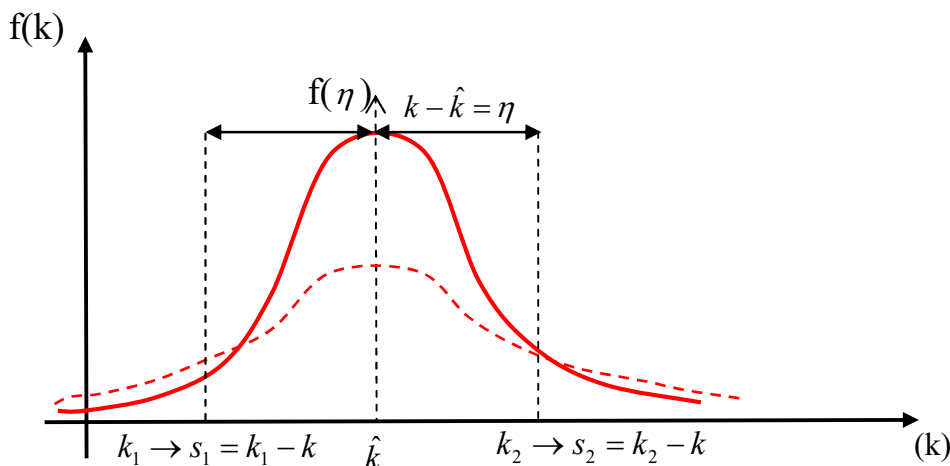


Fig. 1.1: Gaussian.

Let's calculate now the probability that an event happens a number of times between k_α and k_β :

$$P_{k_\alpha, k_\beta} = \int_{k_\alpha}^{k_\beta} P_k dk = \int_{k_\alpha}^{k_\beta} \frac{1}{\sigma_k \sqrt{2\pi}} e^{\left[-\frac{(k-np)^2}{2npq}\right]} dk ; \text{ by jumping to the new integration variable } z:$$

$$P_{k_\alpha, k_\beta} = \int_{z_\alpha}^{z_\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{z_\alpha}^{z_\beta} \phi(z) dz . \text{ Let's analyze now the function:}$$

$$\phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^t \phi(z) dz ; \text{ we have: } P_{k_\alpha, k_\beta} = \phi(t_\beta) - \phi(t_\alpha) = \phi\left(\frac{k_\beta - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_\alpha - np}{\sqrt{npq}}\right) .$$

There exist tables of values for the function $\phi(t)$.

-The 3 σ rule

Let's start now from the (1.1):

$$P_k = f(k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\hat{k}}{\sigma_k}\right)^2} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} = P(s)$$

now, in a differential form, the probability dp that the deviation s is between the infinitesimal ds is:

$$dp = P(s)ds = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} ds \text{ and so the probability that the deviation } s \text{ is between } s_1 \text{ and } s_2 \text{ is } P:$$

$$P_{1,2} = \int_{s_1}^{s_2} dp = \int_{s_1}^{s_2} P(s)ds = \int_{s_1}^{s_2} \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s}{\sigma_k}\right)^2} ds \quad (1.3)$$

$P_{1,2}$ is the surface covered by the Gaussian curve between s_1 and s_2 ; it's easy to calculate through the integral of (1.3), that:

$$\begin{aligned} -0,68\sigma < s < 0,68\sigma & \rightarrow P=50\% \\ -\sigma < s < \sigma & \rightarrow P=68,3\% \\ -1,64\sigma < s < 1,64\sigma & \rightarrow P=90\% \\ -2\sigma < s < 2\sigma & \rightarrow P=95,9\% \\ -2,58\sigma < s < 2,58\sigma & \rightarrow P=99\% \\ -3\sigma < s < 3\sigma & \rightarrow P=99,7\% \end{aligned}$$

Then, the probability that the deviation of the mean value of a measurement is between -3σ and $+3\sigma$ is of 99,7%, that is a rather acceptable value, from which the usual adoption of the three sigma rule, indeed.

In this case, the confidence coefficient z is 3 ($z_{0,997} = 3$). From the z function, that is: $z = \frac{(x - \bar{x})}{\sigma}$,

we get:

$$x = \bar{x} \pm z\sigma . \quad (1.4)$$

If now we refer no more to the normal σ , but to the $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ of the mean value (see above), we

have: $x = \bar{x} \pm z\sigma_{\bar{x}}$, where N is the number of elements of the sample. If the sample is not taken from an infinite population ($N_p = \infty$), but rather from a finite population with N_p elements, then we

$$\text{have: } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}} \quad (\text{see the proof below}) \quad (1.5)$$

so: $x = \bar{x} \pm z \sigma_{\bar{x}} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}}$; for an infinite population we obviously have:

$$\sqrt{\frac{(N_p - N)}{(N_p - 1)}} = \sqrt{\frac{(1 - \frac{N}{N_p})}{(1 - \frac{1}{N_p})}} = 1 \quad \text{and: } x = \bar{x} \pm z \sigma_{\bar{x}} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}} = \bar{x} \pm z \frac{\sigma}{\sqrt{N}}$$

Proof of the (1.5):

we have a set of 5 values {3,4,5,6,7} (whose mean value, of course, is $\mu=5$) and we extract a small sample of two values ($N_p=5$, $N=2$); all the possible small samples are:

(3,4), (3,5), (3,6), (3,7), (4,5), (4,6), (4,7), (5,6), (5,7), (6,7).

The relevant mean values of all those 10 possible small samples are: 3,5|4|4,5|5|4,5|5|5,5|5,5|6|6,5

and their mean value is: $\mu_{\bar{x}} = \frac{3,5 + 4 + 4,5 + 5 + 4,5 + 5 + 5,5 + 5,5 + 6 + 6,5}{10} = 5$ and the relevant

standard deviation of the mean value is: $\sigma_{\bar{x}} = \sqrt{\frac{(3,5-5)^2 + (4-5)^2 + \dots + (6,5-5)^2}{10}} = \pm 0,866$, while

the standard deviation σ referred to the whole population would be:

$$\sigma = \sqrt{\frac{(3-5)^2 + (4-5)^2 + \dots + (7-5)^2}{5}} = \pm 1,414; \text{ so we notice that the (1.5) is respected:}$$

$$\sigma_{\bar{x}} = \pm 0,866 = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}} = \frac{\pm 1,414}{\sqrt{2}} \sqrt{\frac{(5-2)}{(5-1)}}. \text{ And we also notice that: } \mu_{\bar{x}} = \mu. \text{ qed}$$

2-Estimation of the mean value over the whole population N_p , but based over a small sample N – The Student's t Distribution.

A sample starts being no more small if $N \geq 30$. In fact, when $N \geq 30$, the normal distribution (gaussian) is well respected. The same cannot be stated when $N < 30$, indeed.

Therefore, when a sample is small, the statistics is no more the normal one (1.1), that is that of the confidence coefficient $z = \frac{(x - \bar{x})}{\sigma}$, and in place of it we use a similar equation, but whose

parameter is t, based on the Student's t Distribution (of W.S. Gosset) which is, as we said, no more the Normal one, being the sample small:

$$t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} \quad (2.1)$$

where $\bar{x}$ is the mean value of the small sample with N elements, μ is the mean value of the whole population N_p , σ_N is the standard deviation of the small sample (N). On the base of all that, in place

of the (1.4), that is the $x = \bar{x} \pm z \sigma$, we have: $x = \bar{x} \pm \frac{t \sigma_N}{\sqrt{(N-1)}}$ (2.2)

of course, as the purpose here is that of estimating the mean value μ of the whole population and in the (2.2) x plays the role of μ in the (2.1), indeed.

This time we will refer to a new table of values of t_p ; for instance, $t_{0,95}$ refers to a confidentiality level of 95%.

In the table (of Student) all the t_p are supplied as a function of the parameter ν (nu) which is the number of degrees of freedom, that is, in this typical case, for the (2.1) we have:

$$t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{\nu}, \text{ as } \sigma_N \text{ and } \bar{x} \text{ are calculated through the values of the } N$$

elements of the small sample, while the unknown value is just one, that is μ , that is the mean value of the whole population, so $v=N-n$. of unknown values= $N-1$.

Therefore, , in general, $\nu \cong N$, and so in the table of the t values they supply the values t_p for all the ν values, from 1 to 30, because as well as we already said, such statistics is for $N < 30$, while over the 30 just the t values for the ν 40, 60, 120 and ∞ are supplied, as for high ν values, the statistics is back the normal one (gaussian).

3-Proof of the Student's t Distribution.

Preamble:

We start from the way we evaluate, through a weighed average computation, the mean values $\langle x \rangle$ and $\langle y \rangle$ and also the variances $\sigma^2(x)$ and $\sigma^2(y)$, with a probability distribution $f(x,y)$:

$$\langle x \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy ; \quad \langle y \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy$$

$$\sigma^2(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 f(x,y) dx dy ; \quad \sigma^2(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \langle y \rangle)^2 f(x,y) dx dy$$

If now we carry out a change of coordinates, $x,y \rightarrow u,v$, where: $u=u(x,y)$ and $v=v(x,y)$ (and $x=x(u,v)$ and $y=y(u,v)$), then, as we know:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(u,v) f[x(u,v), y(u,v)] |J| du dv \quad (3.1)$$

where $|J|$ is the determinant of the Jacobian (of the classical change of coordinates in a multiple integral):

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}, \text{ being: } \begin{cases} x = x(u,v) \\ y = y(u,v) \end{cases} \rightarrow \begin{cases} d\bar{x} = (\frac{\partial x}{\partial u}) d\bar{u} + (\frac{\partial x}{\partial v}) d\bar{v} \\ d\bar{y} = (\frac{\partial y}{\partial u}) d\bar{u} + (\frac{\partial y}{\partial v}) d\bar{v} \end{cases} ; \text{ in fact, in } (x,y) \text{ the small surface is}$$

$dx dy$, while in (u,v) it is the surface of the parallelogram $d\bar{x} \times d\bar{y}$:

$$d\bar{x} \times d\bar{y} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} du dv = |J| du dv .$$

Sometimes one has to integrate/average in advance over one of the two variables (such as y) and so the (3.1) is left with an integrand function on just x , such as:

$$F(y) = \int_{-\infty}^{\infty} f(x,y) dx ; \text{ here, too, through a change of variables like } (x,y) \rightarrow (x,z/x) \quad (\text{that is: } y=z/x$$

and $z=xy$), we have: $F(z) = \int_{-\infty}^{\infty} f(x, \frac{z}{x}) |J| dx$, where:

$$\begin{aligned} \begin{cases} x = x(x,z) \\ y = y(x,z) \end{cases} &\rightarrow \begin{cases} d\bar{x} = (\frac{\partial x}{\partial x}) d\bar{x} + (\frac{\partial x}{\partial z}) d\bar{z} \\ d\bar{y} = [\frac{\partial(z/x)}{\partial x}] d\bar{x} + [\frac{\partial(z/x)}{\partial z}] d\bar{z} \end{cases} \rightarrow ; \\ &\rightarrow \begin{cases} d\bar{x} = 1 d\bar{x} + 0 d\bar{z} \\ d\bar{y} = -\frac{z}{x^2} d\bar{x} + \frac{1}{x} d\bar{z} \end{cases} \rightarrow |J| = \begin{vmatrix} 1 & 0 \\ -\frac{z}{x^2} & \frac{1}{x} \end{vmatrix} = \frac{1}{x} \rightarrow F(z) = \int_{-\infty}^{\infty} f(x, \frac{z}{x}) \frac{1}{|x|} dx \quad \text{and if } x \text{ and } y \text{ are} \end{aligned}$$

independent variables, whose distribution densities are f_1 and f_2 , then: $F(z) = \int_{-\infty}^{\infty} f_1(x) f_2(\frac{z}{x}) \frac{1}{|x|} dx$.

Similarly, if the change of coordinates is: $(x,y) \rightarrow (zy,y)$ (that is: $z=x/y$), then:

$$|J| = \begin{vmatrix} \frac{\partial(yz)}{\partial y} & \frac{\partial(yz)}{\partial z} \\ \frac{\partial y}{\partial y} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} z & y \\ 1 & 0 \end{vmatrix} = -y \rightarrow F(z) = \int_{-\infty}^{\infty} f(yz, y)|y|dy \text{ and}$$

$$F(z) = \int_{-\infty}^{\infty} f_1(yz)f_2(y)|y|dy \quad (3.2)$$

Gamma Distribution:

Through the change from the variable x to the variable x^2 we go from the Normal Distribution to the Gamma one ($x \rightarrow y=x^2$). From this, it comes out that $x = \pm\sqrt{y}$ (two values) and this means two things, that is: it can be only $y>0$, but in such a change we have to multiply by 2.
Now, the Normal Distribution with $\bar{x} = 0$ is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (3.3)$$

As the total probability must be the same with both the distributions, it must be:

$$\int_0^{\infty} f(x)dx = \int_0^{\infty} g(y)dy \rightarrow f(x)dx = g(y)dy \rightarrow f(x) = g(y)\frac{dy}{dx} \text{ and } g(y) = f(x)\frac{dx}{dy} \text{ and:}$$

$$\frac{dy}{dx} = 2x = 2\sqrt{y} \quad (\rightarrow y>0), \text{ from which (plus the coefficient } x \text{ 2 as above said...):}$$

$$g(y) = 2f(x)\frac{dx}{dy} = 2f[x(y)]\frac{dx}{dy} = 2\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}}\right) \frac{1}{2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} y^{-\frac{1}{2}} = g(y) \quad (3.4)$$

Such a $g(y)$ is normalized; in fact, let's carry out a further change of variable $z = y/2\sigma^2$; we get:

$$\left(\frac{dz}{dy} = \frac{1}{2\sigma^2} \text{ and } dy = (2\sigma^2)dz\right)$$

$$g(z)dz = g(y)dy \rightarrow g(z) = g(y)\frac{dy}{dz} = \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} y^{-\frac{1}{2}}\right)(2\sigma^2) = \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z2\sigma^2}{2\sigma^2}} z^{-\frac{1}{2}} (2\sigma^2)^{-\frac{1}{2}}\right](2\sigma^2) = \frac{1}{\sqrt{\pi}} e^{-z} z^{-\frac{1}{2}} = g(z) ; \text{ at last, as we said, we notice that such a } g(z) \text{ is already normalized, that is it}$$

has preserved the normalization, so justifying also the added coefficient 2; in fact:

$$\int_0^{\infty} g(z)dz = \int_0^{\infty} \frac{1}{\sqrt{\pi}} e^{-z} z^{-\frac{1}{2}} dz = \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-z} z^{-\frac{1}{2}} dz = \frac{1}{\sqrt{\pi}} \left(\int_0^{\infty} e^{-z} z^{t-1} dz\right)_{t=1/2} = \frac{\Gamma(1/2)}{\sqrt{\pi}} = 1$$

($\Gamma(t) = \int_0^{\infty} e^{-z} z^{t-1} dz$ is the Gamma Function of Euler; see above)

After generalizing the (3.4) ($g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} y^{-\frac{1}{2}}$), we have:

$$g(p, \sigma, y) = \frac{1}{(2\sigma^2)^p \Gamma(p)} e^{-\frac{y}{2\sigma^2}} y^{p-1}, \quad (y>0) \quad (3.5)$$

and the (3.4) is a particular case ($p=1/2$).

Student's t:

we saw that through the passage from x to x^2 , the distribution becomes a Gamma one; now, being:

$$\sigma_N = \frac{\sqrt{\sum (x_i)^2}}{\sqrt{N}} = \frac{S}{\sqrt{N}} \quad (3.6)$$

and also being S a quadratic form, it is distributed like Gamma, so like the $g(p, \sigma, y)$ of the (3.5).

Now, we like to understand the distribution $h(\sigma_N)$ of $\sigma_N = \frac{S}{\sqrt{N}}$; then, let's carry out a change of variable in the (3.5), from y to $N\sigma_N^2$, that is: $y = N\sigma_N^2$, from which:

$$dy = 2N\sigma_N d\sigma_N \text{ and } h(\sigma_N) d\sigma_N = g(p, \sigma, y) dy \rightarrow$$

$$h(\sigma_N) = g(p, \sigma, y) \frac{dy}{d\sigma_N} = g(p, \sigma, N\sigma_N^2) 2N\sigma_N \quad (3.7)$$

Moreover, let's evaluate the distribution $r(Y)$ of $Y\sqrt{N}$, where: $Y = (\frac{1}{N} \sum_j x_j) - \mu = \bar{x} - \mu$; it is the normal distribution (that followed by all the x_j), whose variance is (of the mean value-see above) σ^2/n and let's go on staying centered over a zero mean value:

$$(\mu=0 \rightarrow Y = (\frac{1}{N} \sum_j x_j)) \quad (3.8)$$

$$r(Y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-NY^2}{2\sigma^2}} \text{ which is the (3.3) in } x. \text{ Let's go on further, in order to evaluate now the}$$

statistics of $\frac{Y\sqrt{N}}{\sigma_N}$; we are interested in that, as from the (2.2) we have $\mu = \bar{x} \pm \frac{t\sigma_N}{\sqrt{(N-1)}}$, from

which: $\bar{x} - \mu = Y = \pm \frac{t\sigma_N}{\sqrt{(N-1)}} \cong \frac{t\sigma_N}{\sqrt{N}}$, so **the quantity** $\frac{Y\sqrt{N}}{\sigma_N}$ **is really the t** introduced through

the (2.2), indeed. And let's also recall that the (2.2) is interesting for its similarities with the (1.4). ($Y \propto \sigma_N \cdot t$)

So, let's use the (3.2) which joins together those two distributions in yz and in y , that is $f_1(x)=f_1(yz)$ and $f_2(y)$, and in this current case, the ratio $z=x/y$ of the (3.2) corresponds to the ratio $\frac{Y\sqrt{N}}{\sigma_N}$,

indeed, where the statistics of f_1 (of $x=yz$) would be that of $Y\sqrt{N}$, that is the $r(Y)$

$$(r(Y) = r(\sigma_N \cdot t) \rightarrow f_1(yz)),$$

while the $f_2(y)$ is the statistics of what corresponds to y , indeed, that is σ_N ($y \rightarrow \sigma_N$)

$$(f_2(y) \rightarrow h(\sigma_N)) ; \text{ so, let's use the (3.2): } f(t) = \int_0^\infty r(\sigma_N \cdot t) \cdot h(\sigma_N) | \sigma_N | d\sigma_N, \text{ so:}$$

$$(Y \cong \frac{t\sigma_N}{\sqrt{N}}, p=N/2 \text{ and } y = N\sigma_N^2)$$

$$\begin{aligned} f_p(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{\frac{-NY^2}{2\sigma^2}} g(p, \sigma, N\sigma_N^2) 2N\sigma_N^2 d\sigma_N = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{\frac{-\sigma_N^2 t^2}{2\sigma^2}} g(\frac{N}{2}, \sigma, N\sigma_N^2) 2N\sigma_N^2 d\sigma_N = \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{(2\sigma^2)^p \Gamma(p)} \int_0^\infty e^{\frac{-\sigma_N^2 t^2}{2\sigma^2}} e^{\frac{-y}{2\sigma^2}} y^{p-1} 2N\sigma_N^2 d\sigma_N = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{(2\sigma^2)^{\frac{N}{2}} \Gamma(\frac{N}{2})} \int_0^\infty e^{\frac{-\sigma_N^2 t^2}{2\sigma^2}} e^{\frac{-y}{2\sigma^2}} y^{\frac{N}{2}-1} 2N\sigma_N^2 d\sigma_N = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{(2\sigma^2)^{\frac{N}{2}} \Gamma(\frac{N}{2})} \cdot \\
&\cdot \int_0^\infty e^{\frac{-\sigma_N^2 t^2}{2\sigma^2}} e^{\frac{-N\sigma_N^2}{2\sigma^2}} (N\sigma_N^2)^{\frac{N}{2}-1} 2N\sigma_N^2 d\sigma_N = \boxed{\frac{2}{\sqrt{N\pi}\sigma^{N+1}\Gamma(\frac{N}{2})} \left(\frac{N}{2}\right)^{\frac{N}{2}+\frac{1}{2}} \int_0^\infty e^{\frac{-\sigma_N^2(t^2+N)}{2\sigma^2}} \sigma_N^N d\sigma_N = f_N(t)} \quad (3.9)
\end{aligned}$$

and about the last integral, let's carry out a change of variable in it: $z = \frac{\sigma_N^2(t^2 + N)}{2\sigma^2}$:

$$\begin{aligned}
&(((\rightarrow dz = \frac{2(t^2 + N)}{2\sigma^2} \sigma_N d\sigma_N, \quad d\sigma_N = \frac{2\sigma^2}{2(t^2 + N)} \frac{1}{\sigma_N} dz, \quad \sigma_N^2 = \frac{2\sigma^2}{(t^2 + N)} z \quad \text{and} \\
&\sigma_N^N d\sigma_N = \sigma_N^N \left[\frac{2\sigma^2}{2(t^2 + N)} \frac{1}{\sigma_N} dz \right] = \{\sigma_N^{N-1}\} \left[\frac{2\sigma^2}{2(t^2 + N)} dz \right] = \left\{ \left(\frac{2\sigma^2}{(t^2 + N)} z \right)^{\frac{N-1}{2}} \right\} \left[\frac{2\sigma^2}{2(t^2 + N)} dz \right] = \\
&= \frac{1}{2} \left(\frac{2\sigma^2}{t^2 + N} \right)^{(N+1)/2} z^{(N-1)/2} dz)))) \\
&\int_0^\infty e^{\frac{-\sigma_N^2(t^2+N)}{2\sigma^2}} \sigma_N^N d\sigma_N = \frac{1}{2} \left(\frac{2\sigma^2}{t^2 + N} \right)^{(N+1)/2} \int_0^\infty z^{(N-1)/2} e^{-z} dz = \frac{1}{2} \left(\frac{2\sigma^2}{t^2 + N} \right)^{(N+1)/2} \Gamma\left(\frac{N+1}{2}\right) \quad , \quad \text{so the} \quad (3.9)
\end{aligned}$$

becomes:

$$f_N(t) = \frac{\Gamma(\frac{N+1}{2})}{\sqrt{\pi N} \Gamma(\frac{N}{2})} (t^2 / N + 1)^{-(N+1)/2} \quad (\text{which is therefore the Student's t distribution}). \quad (3.10)$$

So, let's use the (2.2) to get an evaluation of the mean value μ of the whole population, with a confidence level given by $f_N(t)$ (eq. (3.10)).

$N = N_{\text{Sample}} - 1 = v$ is the n. of degrees of freedom.

The (3.10) has a distribution integral as well as the normal distribution (page 19), and it is:

$$\phi_N(t) = \int_{-\infty}^t f_N(\theta) d\theta = \frac{\Gamma(\frac{N+1}{2})}{\sqrt{\pi N} \Gamma(\frac{N}{2})} \int_{-\infty}^t (\theta^2 / N + 1)^{-(N+1)/2} d\theta.$$

The probability that a given value t^* of Student is between $-t$ and $+t$, that is: $-t \leq t^* \leq +t$ is:

$$W_N(t) = \int_{-t}^t f_N(\theta) d\theta = 2\phi_N(t) - 1 \quad (3.11)$$

In fact, $W_N(t)$ values are gathered into tables. There is a table which gives all the t values as a function of W_N (at v , or N) and another table which gives all the W_N values as a function of t (t and v , or N).

$f_N(t)$ is an even function, that is: $f_N(\theta) = f_N(-\theta)$; moreover, from the (3.11), we have:

$$\begin{aligned}
W_N(t) &= \int_{-t}^t f_N(\theta) d\theta = 2 \int_{-\infty}^t f_N(\theta) d\theta - 1 = 2 \int_{-\infty}^t f_N(\theta) d\theta - \int_{-\infty}^{\infty} f_N(\theta) d\theta = \left(\int_{-\infty}^t f_N(\theta) d\theta + \int_{-\infty}^t f_N(\theta) d\theta \right) - \int_{-\infty}^{\infty} f_N(\theta) d\theta = \int_{-\infty}^t f_N(\theta) d\theta + \left[\int_{-\infty}^t f_N(\theta) d\theta - \int_{-\infty}^{\infty} f_N(\theta) d\theta \right] = \\
&= \int_{-\infty}^t f_N(\theta) d\theta + \left[\int_{-\infty}^t f_N(\theta) d\theta - \left(\int_{-\infty}^{-t} f_N(\theta) d\theta + \int_{-t}^t f_N(\theta) d\theta + \int_t^{\infty} f_N(\theta) d\theta \right) \right] = \int_{-\infty}^t f_N(\theta) d\theta + \left[\int_{-\infty}^t f_N(\theta) d\theta - \left(\int_{-\infty}^{-t} f_N(\theta) d\theta + 2 \int_t^{\infty} f_N(\theta) d\theta \right) \right] = \\
&= \int_{-\infty}^t f_N(\theta) d\theta + \left[\int_{-\infty}^{-t} f_N(\theta) d\theta - 2 \int_t^{\infty} f_N(\theta) d\theta \right] = \int_{-\infty}^t f_N(\theta) d\theta + \left[(-2 \int_t^{\infty} f_N(\theta) d\theta) + \int_{-\infty}^{-t} f_N(\theta) d\theta \right] = \int_{-\infty}^t f_N(\theta) d\theta + \left[\left(-\int_t^{\infty} f_N(\theta) d\theta - \int_t^{\infty} f_N(\theta) d\theta \right) + \int_{-\infty}^{-t} f_N(\theta) d\theta \right] = \int_{-\infty}^t f_N(\theta) d\theta + \left[-\int_t^{\infty} f_N(\theta) d\theta \right] =
\end{aligned}$$

$$= \left[\int_{-\infty}^t \right] - \int_t^{\infty} = \left[\int_{-\infty}^{-t} + \int_{-t}^t \right] - \int_t^{\infty} = \int_{-t}^t .$$

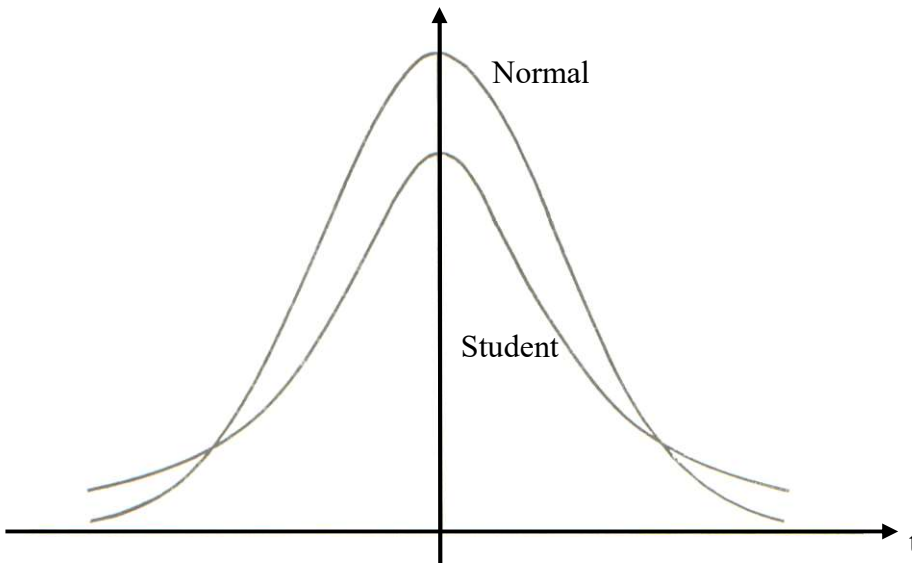


Fig. 3.1-Comparison between Normal and Student's t Distributions.

AS we can see in Fig. 3.1, the Student's t Distribution is lower and larger; lower at the origin of axes as handled data are less probable, as N is low, and it's larger at sides (that is when we get far from statistically less wrong/more probable values) as larger errors are typically acknowledged more probability, still because of the low N ($N < 30$). As N is increased, those two distributions start matching (already with $N=40$).

At last, the probability that the (2.2) holds:

$$\mu \rightarrow x = \bar{x} \pm \frac{t\sigma_N}{\sqrt{(N-1)}}$$

with a certain t is $W_N(t)$ of the (3.11), indeed.

If, on the contrary, we fix the probability (W_N), then from the tables of W_N we can find out the t value under which that is true and from there, through the (2.2), we have an estimate of the mean value μ of the population (but under that W_N probability value) and then one can decide if accepting and going on (for instance, with production processes), or, on the contrary, if refusing and take actions (for instance, by a check of the production machines).

Besides, we have the Type I Error when we (by mistake) refuse instead of accepting, and we have the Type II Error when (by mistake) we accept instead of refusing.

As well as the case of the Normal Distribution, based over the z parameter, also with the case of the Student t parameter, with a sample of N elements, which is, of course, much lower than the population N_P (if it is known) ($N < 30$), we can use the following formula:

$$\mu = \bar{x} + \frac{|t|\sigma_N}{\sqrt{N}} \sqrt{\frac{(N_P - N)}{(N_P - 1)}} . \text{ When mean value and standard deviation of the whole population are}$$

not known, here is the formula that can be used as a link between the standard deviation σ_N of the sample and that of the whole population σ :

$$\sigma_N = \sigma \sqrt{\frac{N-1}{N}} , \quad (3.12)$$

with the factor $\frac{N-1}{N}$ which is known as the Bessel's Correction Factor.

Proof (of the (3.12)):

let $\bar{x}$ be the mean value of the small sample and μ that of the population; we have:

$x_j - \bar{x} = (x_j - \mu) - (\bar{x} - \mu)$ so: $(x_j - \bar{x})^2 = (x_j - \mu)^2 - 2(x_j - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2$ and:

$$\begin{aligned}\sigma_N^2 &= \sum_{j=1}^N (x_j - \bar{x})^2 = \sum (x_j - \mu)^2 - 2(\bar{x} - \mu) \sum (x_j - \mu) + \sum (\bar{x} - \mu)^2 = \\ &= \sum_{j=1}^N (x_j - \mu)^2 - 2(\bar{x} - \mu) [\sum x_j - N\mu] + N(\bar{x} - \mu)^2 = \sum_{j=1}^N (x_j - \mu)^2 - 2(\bar{x} - \mu) [N\bar{x} - N\mu] + N(\bar{x} - \mu)^2 = \\ &= \sum_{j=1}^N (x_j - \mu)^2 - 2N(\bar{x} - \mu)^2 + N(\bar{x} - \mu)^2 = \sum_{j=1}^N (x_j - \mu)^2 - N(\bar{x} - \mu)^2 \rightarrow \\ &\quad \text{(Calculated with the } \mu \text{ of the population, so it's the } \sigma \text{ of the population)}\end{aligned}$$

$$\rightarrow \left\langle \sum_{j=1}^N (x_j - \bar{x})^2 \right\rangle = \left\langle \sum_{j=1}^N (x_j - \mu)^2 \right\rangle - N \left\langle (\bar{x} - \mu)^2 \right\rangle = N\sigma^2 - N \left(\frac{\sigma^2}{N} \right) =$$

(The $\sigma_{\bar{x}}^2$ of the mean value, which is known to be $\frac{\sigma^2}{N}$)

then, going on with calculations: $= (N-1)\sigma^2 = N\sigma_N^2$, so: $\sigma_N^2 = \frac{(N-1)}{N} \sigma^2$, qed.

Therefore, an estimation of σ yields: $\sigma = \sigma_N \sqrt{\frac{N}{N-1}}$ and with $N \sim 30 \rightarrow \sigma = \sigma_N$ (3.13)

But such a (3.13) cannot be used with small samples, of course (Student) and the σ given by the

(3.12) will be used. Moreover, $\mu = \bar{x} + \frac{|t|\sigma_N}{\sqrt{N}} \sqrt{\frac{N}{N-1}}$ and for non infinite populations, we have:

$$\mu = \bar{x} + \frac{|t|\sigma_N}{\sqrt{N}} \sqrt{\frac{N}{N-1}} \sqrt{\frac{(N_p - N)}{(N_p - 1)}}.$$

Comparison between two experimental mean values:

1° sample, n items, mean $\bar{x}$, dev.std. σ_x

2° sample, n items, mean $\bar{y}$, dev.std. σ_y

Question: is the difference $\bar{x} - \bar{y}$ real and with significance or, on the contrary, it's simply due to chance? In other words, in that case, are μ_x and μ_y , or $\bar{x}$ and $\bar{y}$, really different or their difference is due just to chance?

As σ_m (mean) we consider, of course:

$$\sigma_m = \sqrt{\frac{(n-1)\sigma_x^2 + (m-1)\sigma_y^2}{(n-1) + (m-1)}} \quad \text{and} \quad \sigma_m^2 = \sigma_m^2 \left(\frac{1}{n} + \frac{1}{m} \right).$$

At this point, we fix the statistical security threshold W and from the tables we find out for which t^* value this holds. As an example, we choose a $W=0,05 \rightarrow 5\%$ and we look up in the table that W of 0,05 corresponding to a $v=n+m-2$. Once we get a t^* , according to what we already know on the Student's t Distribution statistics, we have: $\mu = \bar{x} + |t|\sigma_m$ generically and, on our specific case:

$$|\mu_x - \mu_y| = |\bar{x} - \bar{y}| + |t|\sigma_m; \quad (3.14)$$

As the supposition $\mu_x = \mu_y$ in the (3.14) causes: $|t| = \frac{|\bar{x} - \bar{y}|}{\sigma_m}$, then the supposition $\mu_x = \mu_y$ is accepted

if $\frac{|\bar{x} - \bar{y}|}{\sigma_m} < t^*$, while it's refused and $\mu_x \neq \mu_y$ if $\frac{|\bar{x} - \bar{y}|}{\sigma_m} \geq t^*$.

Conversely, in case we know the standard deviations of the population, then we consider:

$\sigma_D = \sqrt{\sigma_x^2 + \sigma_y^2} = \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$, and in this case we refer no more to t, but to z, as the z parameter follows the normal distribution and not the Student's one, as we have the σ values of the population; therefore, in such a case: $z = \frac{\bar{x} - \bar{y}}{\sigma_D} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$ and once $W(z)$ is fixed, we get the z and

$z > \frac{\bar{x} - \bar{y}}{\sigma_D}$ allows the acceptance of the supposition $\mu_x = \mu_y$ (just like the previous case), while the

refusal is represented by the situation: $z \leq \frac{\bar{x} - \bar{y}}{\sigma_D}$.

EXAMPLE 1 of use of the Student's t Distribution:

A machine has produced washers with a mean thickness of 0,05cm ($\mu=0,05\text{cm}$). In order to understand if the machine is working properly, we sample $N=10$ washers, then we measure their thickness and figure out the mean value, which results to be, say, $\bar{x} = 0,053\text{cm}$; we also figure out the standard deviation of such a sample, which is:

$$\sigma_N = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_{10} - \bar{x})^2}{(N-1)}} \text{ and we get, say, } \sigma_N = 0,003\text{cm}.$$

Then, we evaluate the reliability of the machine, at the beginning under a significance level of 5% (0,05) and then of 1% (0,01).

A)0,05: $N < 30$, so we do not use the normal distribution, but rather the Student's t one:

$$t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} = \frac{(0,053 - 0,05)}{0,003} \sqrt{10-1} = 3,00.$$

Now, if we want to stay within the 5% (0,05), and being that the diameter deviates either upwards or downwards, then we have a +2,5% and a -2,5% of samples out, so bringing the acceptance range to 97,5% $(-(100\% - 2,5\%) \text{ and } +(100\% - 2,5\%))$, that is between -0,975 and +0,975; then, we look up in the Student's table the $t_{0,975}$ corresponding to a $v=N-1=9$ and we find the value $t_{0,975}=2,26$; as our t is 3, it is out of the range $(-2,26 \div +2,26)$, so we refuse and have the machine checked.

B)0,01: conversely, if we keep over the 1%, then we can go over upwards and downwards about 0,5%, so we have an acceptance range of good samples at 99,5% $(-(100\% - 0,5\%) \text{ and } +(100\% - 0,5\%))$ that is between -0,995 and +0,995 and so we look up in the Student's table the $t_{0,995}$ corresponding to a $v=N-1=9$ and find $t_{0,995}=3,25$; finally, as our $t=3$ stays within the range $(-3,25 \div +3,25)$, then we accept the status of the machine (at 1%) and:

$$\bar{x} - t_{0,995} \frac{\sigma_N}{\sqrt{(N-1)}} < \mu < \bar{x} + t_{0,995} \frac{\sigma_N}{\sqrt{(N-1)}}.$$

EXAMPLE 2 of use of the Student's t Distribution:

Steel cables are produced; we sample $N=6$ cables and the mean value of the breaking load of such 6 cables is $\bar{x} = 7750\text{kg}$ (but we claim a value of 8000kg), and with a standard deviation σ_N of 145kg.

$$(\sigma_N = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_6 - \bar{x})^2}{(N-1)}} = 145\text{kg})$$

Can we, on the basis of the results coming from the 6 cable sample, reassure the producer at a significance level of A)0,05 (5%) and B)0,01 (1%) ?

In this case, the producer is not satisfied only when $x < 8000\text{kg}$, as if $x > 8000\text{kg}$, so much the better; so, we can not to go just downwards (-):

$$t = \frac{(\bar{x} - \mu)}{\sigma_N} \sqrt{(N-1)} = \frac{(7750 - 8000)}{145} \sqrt{6-1} = -3,86 .$$

A) 5% (0,05) Then, we look up in the Student's table the $t_{0,95}$ corresponding to a $v=6-1=5$ and find $t_{0,95}=+2,01$; $(-t_{0,95}=-2,01)$; as $t=-3,86 < -t_{0,95}=-2,01$, then we refuse and DO NOT reassure the producer.

B) 1% (0,01) $t_{0,99 (v=5)}=+3,36 \rightarrow (-t_{0,99}=-3,36)$; as our $t=-3,86 < -t_{0,99}=-3,36$, we refuse.

4-The CHI-Square (χ^2) Distribution (of Helmer).

With a coin tossing of $n=100$ times, the probability to get head is $p=1/2$, so the frequency of head is $np=50$ times; the opposite event (tail) has got a probability $q=(1-p)=(1-1/2)=1/2$ and a frequency $nq=50$. All this stands just theoretically; in practice, we could get a bit different situations.

We recall the z parameter on page 18: $z = \frac{(x - \bar{x})}{\sigma} = \frac{(x - np)}{\sqrt{npq}}$ (4.1)

and by squaring it, we get: $z^2 = \frac{(x - np)^2}{npq}$ (4.2)

Now we immediately notice the (4.2) can be rewritten like that:

$$z^2 = \frac{(x_1 - np)^2}{np} + \frac{(x_2 - nq)^2}{nq} \quad (4.3)$$

where x_1 is the random variable corresponding to the desired event x and x_2 is the random variable corresponding to the opposite event $x_2=n-x_1$; now, let us prove the (4.3):

$$\begin{aligned} z^2 &= \frac{(x - np)^2}{np} + \frac{[(n - x) - n(1 - p)]^2}{n(1 - p)} = \frac{(1 - p)(x^2 + n^2 p^2 - 2np x) + p(n^2 p^2 + x^2 - 2np x)}{np(1 - p)} = \\ &= \frac{x^2 + n^2 p^2 - 2np x}{np(1 - p)} = \frac{(x - np)^2}{npq} \quad \text{qed.} \end{aligned}$$

Now, the (4.3) suggests to consider as an evaluation parameter for the discrepancy between the expected frequencies $(np, nq) / (np_1, \dots, np_i, \dots, np_n)$ and the observed ones $(x_1, x_2, \dots, x_i, \dots, x_n)$, the parameter of the (4.5) we are going to show now:

(expected frequency= f_e) (observed frequency= f_o) (and of course the summation of all the observed frequencies must yield n , that is the total value:

$$x_1 + x_2 + \dots + x_n = n \rightarrow \sum_j x_j = n \quad (4.4)$$

$$\chi^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} + \dots + \frac{(x_n - np_n)^2}{np_n} = \sum_{j=1}^n \frac{(x_j - np_j)^2}{np_j} = \sum \frac{(f_o - f_e)^2}{f_e} \quad (4.5)$$

The (4.5) is a quadratic form. A form which is the same as the (4.5) is the following one:

$$\chi^2 = \sum_{j=1}^n \frac{x_j^2}{np_j} - n \quad (4.6)$$

(this is a quadratic form as well)

proof:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^n \frac{(x_j - np_j)^2}{np_j} = \sum_j \left(\frac{x_j^2}{np_j} + \frac{n^2 p_j^2}{np_j} - \frac{2np_j x_j}{np_j} \right) = \sum_j \left(\frac{x_j^2}{np_j} + np_j - 2x_j \right) = \sum_j \frac{x_j^2}{np_j} - \sum_j (2x_j - np_j) = \\ &= \sum_j \frac{x_j^2}{np_j} - \sum_j [(2x_j - 2np_j) + np_j] = \sum_j \frac{x_j^2}{np_j} - 2 \sum_j (x_j - np_j) - \sum_j np_j = \sum_j \frac{x_j^2}{np_j} - 2 \cdot 0 - n = \\ &= \sum_j \frac{x_j^2}{np_j} - n \quad \text{qed.} \quad \text{In fact, } \sum_j np_j = n \sum_j p_j = n \cdot 1, \text{ as the summation of all the probabilities } p_j\end{aligned}$$

must yield the highest probability, that is 1, while: $2 \sum_j (x_j - np_j) = 2 \sum_j x_j - 2 \sum_j np_j$, but because of the (4.4), we have: $\sum_j x_j = n$, so: $2 \sum_j x_j - 2 \sum_j np_j = 2n - 2n \sum_j p_j = 2n - 2n \cdot 1 = 0$.

From the (4.5) we understand that if observed frequencies and expected ones match, then $\chi^2 = 0$, while if $\chi^2 > 0$, then they do not match.

We started from the (4.2) and by considering the variable A: $A = (x - np)$, we have: $z^2 = \frac{A^2}{\sigma^2}$, so it is a quantity like that of the (4.5), that is like:

$$\chi^2 = \sum_{j=1}^v \frac{A_j^2}{\sigma^2} = \gamma, \quad (\gamma > 0, \langle A_j \rangle = 0 \text{ e } \sigma_{A_j}^2 = \sigma^2) \quad (4.7)$$

which has a Gamma (γ) probability density like the (3.5) (with $p=v/2$ and $y=\gamma\sigma^2$), as it is a quadratic form (σ here is a parameter and not a variable):

$$d_v(\gamma) = \frac{1}{2^{v/2} \Gamma(v/2)} \gamma^{v/2-1} e^{-\gamma/2} \quad (d_v(\gamma) = 0 \text{ with } \gamma \leq 0) \quad (4.8)$$

in fact, according to the (3.5), we have (with $p=v/2$ and $y=\gamma\sigma^2$):

$$g(p, \sigma, y) = \frac{1}{(2\sigma^2)^p \Gamma(p)} e^{-\frac{y}{2\sigma^2}} y^{p-1} = g\left(\frac{v}{2}, \sigma, \gamma\sigma^2\right) = e^{-\frac{\gamma\sigma^2}{2\sigma^2}} (\gamma^{v/2-1} \sigma^{2(v/2-1)}) \frac{1}{(2\sigma^2)^{v/2} \Gamma(v/2)} |\sigma^2| = d_v(\gamma)$$

($|\sigma^2|$ is the determinant of the Jacobian, coming from the variable change $y=\gamma\sigma^2$, as per the (3.2))

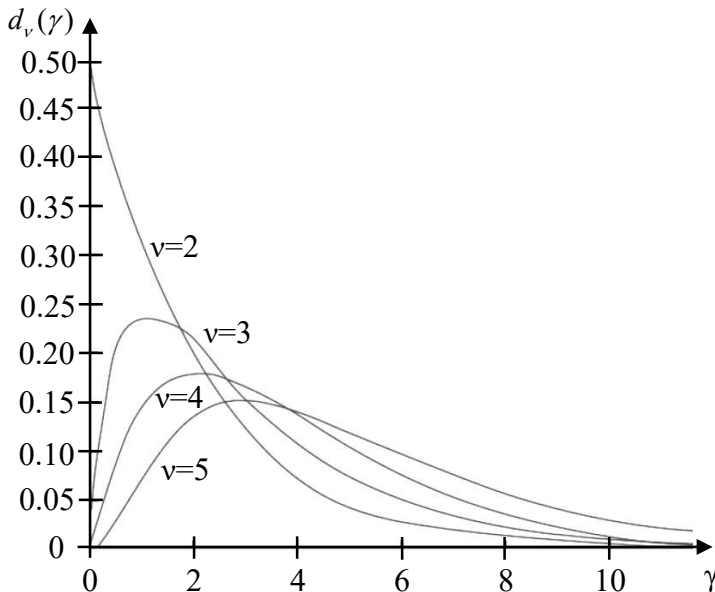


Fig. 4.1: Distribution $d_v(\gamma)$ of the χ^2 .

Mean value of χ^2 :

$$\overline{\chi^2} = \bar{\gamma} = \nu \quad (\text{proof at page 32}) \quad (4.9)$$

$$\text{Moreover: } \sigma_\gamma^2 = \sigma_{\chi^2}^2 = \langle (\gamma - \langle \gamma \rangle)^2 \rangle = \langle (\chi^2 - \langle \chi^2 \rangle)^2 \rangle = 2\nu \rightarrow \quad (4.10)$$

$$\rightarrow \sigma_\gamma = \sqrt{2\nu} \quad (\text{proof at page 32}) \quad (4.11)$$

As ν is increased, $d_\nu(\gamma)$ tends to the Normal Distribution; in fact, in the tables for $d_\nu(\gamma)$, ν is never over 30/40. Of course, the probability that γ is not higher than a given value γ_k ($\gamma_k < \gamma$) is:

$$D_\nu(\gamma_k) = \int_0^{\gamma_k} d_\nu(\gamma) d\lambda = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \int_0^{\gamma_k} \gamma^{\nu/2-1} e^{-\gamma/2} d\lambda$$

$$\text{The probability that } \gamma > \gamma_k \text{ of course is: } P_\nu(\gamma_k) = 1 - D_\nu(\gamma_k) = \int_{\gamma_k}^{\infty} d_\nu(\gamma) d\lambda = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \int_{\gamma_k}^{\infty} \gamma^{\nu/2-1} e^{-\gamma/2} d\lambda$$

Testing a σ^2 variance supposition through the χ^2 Distribution:

supposition: $\sigma^2 = \sigma_{0x}^2$; then, of course, one fixes also the corresponding probability $D_\nu(\gamma_k)$ at which one claims the supposition. We say in advance that, of course, normally:

$$\sigma_0^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ (variance of the sample) and } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ (mean v. of the sample) and it}$$

will represent the μ (of the population), while the σ^2 (of the population) is given by the supposition, indeed. Now, the variable (see also the (4.7)):

$$\chi^2 = \frac{(n-1)\sigma_0^2}{\sigma^2} = \frac{\nu\sigma_0^2}{\sigma_{0x}^2} \quad (4.12)$$

(σ is about the population, σ_0 is about the sample n and σ_{0x} is the one we suppose to represent σ)

has the following form $\sum_i (x_i - \bar{x})^2$ (from the (4.7)) and so it's a quadratic form and so it has (still from the (4.7)) the χ^2 Distribution (with $\nu=n-1$). Then, we look up in the χ^2 table those two $\gamma_{1,2}$

(at a given ν) corresponding to the probability $D_\nu(\gamma_k)$ and if $\gamma_1 < \frac{\nu\sigma_0^2}{\sigma_{0x}^2} < \gamma_2$, then the supposition

is accepted; conversely, it is refused.

EXAMPLE 1 on the use of the χ^2 Distribution:

Over $n=200$ tosses of a coin, we have $x_1=115$ heads and $x_2=85$ tails, that is two sets of results ($k=2$). We want to understand if the coin is unfair or not, at the usual significance levels 0,05 (5%) and 0,01 (1%): let's use the (4.5):

$$\chi^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} = \frac{(115 - 100)^2}{100} + \frac{(85 - 100)^2}{100} = 4,50.$$

Then, as we have two sets of results ($k=2$), then the n . of degrees of freedom is: $\nu=k-1=1$.

Case 1) (0,05 – 5%): ($1-0,05=0,95$) we look up in the χ^2 table the $\chi_{0,95}^2$ corresponding to $\nu=1$ and we find: $\chi_{0,95}^2 = 3,84$; now, as $4,50 > 3,84$, then our χ^2 is over the more reliable $\chi_{0,95}^2$ value, given by the statistics, then our higher χ^2 would match a lower probability value which is not enough to justify the supposition; therefore, we reject the supposition according to which the coin is not unfair (5%).

Case 2) (0,01 – 1%): (1-0,01=0,99) from the table we have: $\chi^2_{0,99(v=1)} = 6,63$ and $4,50 < 6,63$, then, our χ^2 is not over the more reliable $\chi^2_{0,99}$ value given by the statistics, so our lower χ^2 would match a higher probability value, which is enough to justify the supposition; so we have to accept the supposition that the coin is fair (1%).

To sum up things, we understand from those two cases, that, in the end, the coin is probably unfair.

EXAMPLE 2 of use of the χ^2 Distribution:

In the past a packing machine filled packs with 40g and with a standard deviation $\sigma=0,25$ g. Today, an $n=20$ packs sample is taken ($v=n-1=20-1=19$), and it shows a standard deviation $\sigma_0=0,32$ g. Supposition H_1 : is such an increase of variability significant (and so it worries us) at levels of 5% (0,05) and 1% (0,01)? (or, on the contrary, supposition H_0 holds, that is: that change in the std. dev. is merely due to chance?)

With reference to the (4.12), we have: $\chi^2 = \frac{n\sigma_0^2}{\sigma^2} = \frac{20 \cdot 0,32^2}{0,25^2} = 32,8$

(here, in the denominator we have used the most reliable σ we already had, and we have used n in the numerator, in place of $(n-1)$, as the std. dev. of the sample hasn't been evaluated through the mean value of the sample itself, but rather through the more reliable μ from the past (40g))

Now we will use a one-sided test, as we will be worried only by the higher std. devs., so the ones which exceed upwards; conversely, about the smaller ones, so much the better.

-So, about the significance level of 5%, we look up the value $\chi^2_{0,95}$, which is 30,1 (at a $v=19$); therefore, being about our χ^2 , that: $\chi^2 > \chi^2_{0,95}$, then we refuse the H_0 supposition and we worry about. In fact, here our χ^2 would be over the more reliable $\chi^2_{0,95}$ supplied by the statistics and so it would match a lower probability, which is not enough.

-Conversely, at the significance level of 1% (0,01), we have that $\chi^2_{0,99(v=19)} = 36,2 > \chi^2 = 32,8$, so we do not refuse the supposition H_0 , so we worry less.

Hence, a check of the machine, in general should be carried out.

5-Moment Generating Function.

Such a generating function allows us (for instance) to fastly calculate the mean values $\langle x \rangle$ and $\langle x^2 \rangle$. If we have a distribution function $f(x)$, then by calculating the weighed mean value (whose weight is $f(x) dx$) of the function e^{tx} , we get the relevant mean value:

$$\langle e^{tx} \rangle = \int e^{tx} f(x) dx = 1 + t \langle x \rangle + \frac{t^2}{2!} \langle x^2 \rangle + \dots \quad (5.1)$$

after that we have also considered the Taylor's development of $\langle e^{tx} \rangle$. It logically comes out that:

$$\langle x \rangle = \left. \frac{d}{dt} \langle e^{tx} \rangle \right|_{t=0}, \quad \langle x^2 \rangle = \left. \frac{d^2}{dt^2} \langle e^{tx} \rangle \right|_{t=0}, \quad \dots \quad \langle x^n \rangle = \left. \frac{d^n}{dt^n} \langle e^{tx} \rangle \right|_{t=0} \quad (5.2)$$

Now, we have 4 balls in an urn and every ball has its own number (1 or 2, or 3 or 4). Now, after carrying out extractions of two balls at a time and after summing the relevant numbers, the six summations we can get are: 3, 4, 5 (twice), 6 and 7; the 5 can be obtained in two ways: 2+3 and 1+4. Then, the relevant probabilities of appearing of such pairs of numbers are: $P(3)=1/6$, $P(4)=1/6$, $P(5)=2 \times 1/6=1/3$, $P(6)=1/6$, $P(7)=1/6$. So, the Moment Generating Function is, according to the description given by the (5.1) we gave about:

$$M(t) = \frac{1}{6} e^{3t} + \frac{1}{6} e^{4t} + \frac{1}{6} \cdot 2 \cdot e^{5t} + \frac{1}{6} e^{6t} + \frac{1}{6} e^{7t} = \frac{1}{6} (e^{3t} + e^{4t} + 2e^{5t} + e^{6t} + e^{7t})$$

Regarding the derivatives:

$$M'(t) = \frac{1}{6}(3e^{3t} + 4e^{4t} + 10e^{5t} + 6e^{6t} + 7e^{7t}) \quad M''(t) = \frac{1}{6}(9e^{3t} + 16e^{4t} + 50e^{5t} + 36e^{6t} + 49e^{7t})$$

and according to the (5.2), we have: $\langle x \rangle = M'(0) = 5$ and $\langle x^2 \rangle = M''(0) = \frac{80}{3}$.

Conversely, if we wanted to calculate $\langle x \rangle$ and $\langle x^2 \rangle$ through the classical way, we would have

$$\text{gotten: } \langle x \rangle = \frac{3+4+5+5+6+7}{6} = 5 \text{ (idem) and } \langle x^2 \rangle = \frac{3^2+4^2+5^2+5^2+6^2+7^2}{6} = \frac{80}{3} \text{ (idem)}$$

Now, let's give e^{tx} a shape: if the distribution is that of the χ^2 , then we have: $\langle e^{tx} \rangle = \frac{1}{(1-2t)^{\nu/2}}$; in

fact, the (4.8) ($d_\nu(\gamma) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \gamma^{\nu/2-1} e^{-\gamma/2}$) in the (5.1) yields:

$$\langle e^{t\gamma} \rangle = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty e^{t\gamma} \gamma^{\nu/2-1} e^{-\gamma/2} d\gamma = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \gamma^{\nu/2-1} e^{-(1-2t)\frac{\gamma}{2}} d\gamma ; \text{ now, by putting: } u = (1-2t)\frac{\gamma}{2},$$

we have:

$$\langle e^{t\gamma} \rangle = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \left(\frac{2u}{1-2t}\right)^{(\nu-2)/2} e^{-u} \frac{2du}{(1-2t)} = \frac{(1-2t)^{-\nu/2}}{\Gamma(\nu/2)} \int_0^\infty u^{(\nu-2)/2} e^{-u} du = \frac{(1-2t)^{-\nu/2}}{\Gamma(\nu/2)} \Gamma(\nu/2) = \frac{1}{(1-2t)^{\nu/2}}$$

qed. Finally, after using the (5.2), we have: $\langle \gamma \rangle = \langle \chi^2 \rangle = \frac{d}{dt}[(1-2t)^{-\nu/2}] \Big|_{t=0} = \nu$, that is the (4.9),

and $\langle \gamma^2 \rangle = \langle (\chi^2)^2 \rangle = \frac{d^2}{dt^2}[(1-2t)^{-\nu/2}] \Big|_{t=0} = \nu(\nu+2)$, so, according to the (4.10):

$$\begin{aligned} \sigma_\gamma^2 &= \sigma_{\chi^2}^2 = \langle (\gamma - \langle \gamma \rangle)^2 \rangle = \langle (\chi^2 - \langle \chi^2 \rangle)^2 \rangle = \langle (\chi^2)^2 \rangle + \langle \langle \chi^2 \rangle^2 \rangle - 2\langle \chi^2 \rangle \langle \chi^2 \rangle = \\ &= \langle (\chi^2)^2 \rangle - 2\langle \chi^2 \rangle^2 = \nu(\nu+2) - \nu^2 = 2\nu \text{ (which is the (4.10), indeed).} \end{aligned}$$

Besides, if x_1 and x_2 are two independent random variables, and distributed like χ^2 and with degrees of freedom ν_1 and ν_2 , then, if $x = x_1 + x_2$, $\rightarrow$

$$\rightarrow \langle e^{tx} \rangle = \langle e^{t(x_1+x_2)} \rangle = \langle e^{tx_1} \rangle \langle e^{tx_2} \rangle = (1-2t)^{-\nu_1/2} (1-2t)^{-\nu_2/2} = (1-2t)^{-(\nu_1+\nu_2)/2}.$$

6-Linear Correlation.

If we put in relation on a graph the deviations $\mathcal{E}(x)$ of the variable x and those $\mathcal{E}(y)$ of the variable y , we can have at least the following three cases:

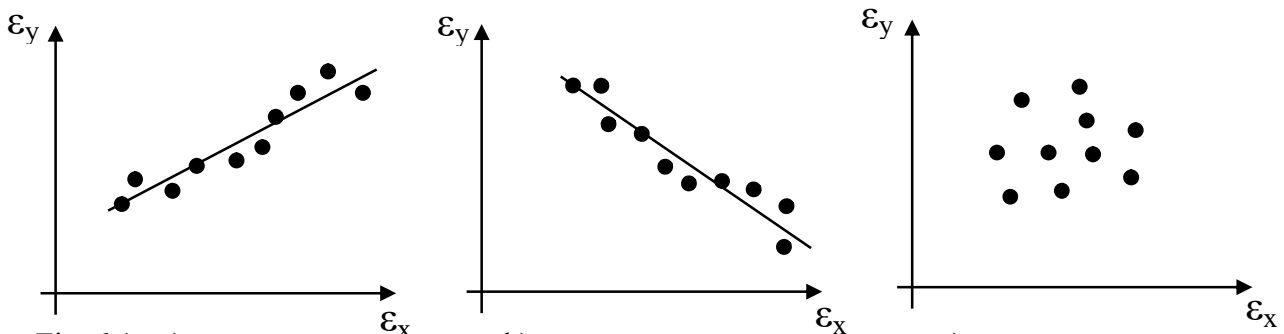


Fig. 6.1 a)

a) Positive Linear Correlations

b)

b) Negative Linear Correlations

c)

c) No Correlation

-Correlation coefficient:

$$r = \frac{\sum [\varepsilon(x)\varepsilon(y)]}{\sqrt{\sum \varepsilon^2(x)}\sqrt{\sum \varepsilon^2(y)}} \quad (r \text{ is between } -1 \text{ and } +1) \quad (6.1)$$

We notice that if the values concentrate like in a), then $r \rightarrow +1$ and $r \rightarrow -1$ in the case b).
 $r \rightarrow 0$ as in c) if there is no correlation.

a) and b) tell us that correlation means linear equation:

$$(\varepsilon(y) = a\varepsilon(x) + b) \quad \text{regression line} \quad (6.2)$$

We have: $\varepsilon(x) = x - \langle x \rangle$ and $\varepsilon(y) = y - \langle y \rangle$; if the distribution function of $\varepsilon(x)$ and $\varepsilon(y)$ together is $f(\varepsilon(x), \varepsilon(y))$ and it can be decomposed/factorized in $f(\varepsilon(x)) \cdot f(\varepsilon(y))$, which is what shows independency, that is no correlation, then, in such a case, we expect an $r \rightarrow 0$; in fact:

$\varepsilon(x)\varepsilon(y) = [x - \langle x \rangle][y - \langle y \rangle]$ and after defining the COVARIANCE: $\text{cov}(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle$ then:

$$\begin{aligned} \text{cov}(x, y) &= \int [\varepsilon(x)\varepsilon(y)] f(\varepsilon(x), \varepsilon(y)) d\varepsilon(x) d\varepsilon(y) = \left[\int \varepsilon(x) f(\varepsilon(x)) d\varepsilon(x) \right] \left[\int \varepsilon(y) f(\varepsilon(y)) d\varepsilon(y) \right] = \\ &= \left[\int (x - \langle x \rangle) f(x) dx \right] \left[\int (y - \langle y \rangle) f(y) dy \right] = (\langle x \rangle - \langle x \rangle)(\langle y \rangle - \langle y \rangle) = 0, \text{ and passing now from the} \\ &\text{continuum to the discrete case } (\int \rightarrow \sum), \text{ we have:} \end{aligned}$$

$$\text{cov}(x, y) = \frac{\sum [\varepsilon(x)\varepsilon(y)]}{n} = \frac{\sum [(x - \langle x \rangle)(y - \langle y \rangle)]}{n} = 0 \quad (6.3)$$

(so $\text{cov}(x, y) \propto r$, that is: $r \rightarrow 0$, so it is an absence of correlation); and after proceeding with the (6.3):

$$\begin{aligned} \text{cov}(x, y) &= \frac{\sum [\varepsilon(x)\varepsilon(y)]}{n} = \frac{\sum [(x - \langle x \rangle)(y - \langle y \rangle)]}{n} = 0 = \frac{\sum (xy)}{n} - \frac{\sum (x\langle y \rangle)}{n} - \frac{\sum (\langle x \rangle y)}{n} + \frac{\sum (\langle x \rangle \langle y \rangle)}{n} = \\ &= \underline{\langle xy \rangle} - \langle y \rangle \frac{\sum x}{n} - \langle x \rangle \frac{\sum y}{n} + \langle x \rangle \langle y \rangle \frac{\sum 1}{n} = \underline{\langle x \rangle \langle y \rangle} - \langle y \rangle \langle x \rangle - \langle x \rangle \langle y \rangle + \langle x \rangle \langle y \rangle = 0 \end{aligned}$$

$$(\text{independence} \rightarrow \langle xy \rangle = \langle x \rangle \langle y \rangle)$$

The (6.1), that is the expression of r , is a normalization of $\text{cov}(x, y)$ and the following formula holds:

$$r = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} \quad (6.4)$$

$$\text{In fact, after recalling that: } \sigma(x) = \sqrt{\frac{\sum_{i=1}^n (x_i - \langle x \rangle)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (\varepsilon_{x_i})^2}{n}} \rightarrow n\sigma^2(x) = \sum_{i=1}^n (x_i - \langle x \rangle)^2 =$$

$$= \sum_{i=1}^n x_i^2 - 2\langle x \rangle \sum_{i=1}^n x_i + n\langle x \rangle^2 = n\langle x^2 \rangle - 2n\langle x \rangle \langle x \rangle + n\langle x \rangle^2 = n[\langle x^2 \rangle - \langle x \rangle^2] \rightarrow$$

$$\rightarrow \sigma^2(x) = \langle x^2 \rangle - \langle x \rangle^2 \quad (6.5)$$

so, because of the (6.4):

$$r = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} = \frac{\langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle}{\sqrt{\frac{\sum_{i=1}^n (\varepsilon_{x_i})^2}{n}} \sqrt{\frac{\sum_{i=1}^n (\varepsilon_{y_i})^2}{n}}} = \frac{\sum [\frac{(x - \langle x \rangle)(y - \langle y \rangle)}{n}]}{\sqrt{\frac{\sum \varepsilon_{x_i}^2}{n}} \sqrt{\frac{\sum \varepsilon_{y_i}^2}{n}}} = \frac{\sum [\frac{\varepsilon(x)\varepsilon(y)}{n}]}{\sqrt{\frac{\sum \varepsilon^2(x)}{n}} \sqrt{\frac{\sum \varepsilon^2(y)}{n}}} =$$

$$= \frac{\sum [\varepsilon(x)\varepsilon(y)]}{\sqrt{\sum \varepsilon^2(x)} \sqrt{\sum \varepsilon^2(y)}} = r, \text{ that is the (6.1), indeed.}$$

And let's ask ourselves to show that: $-1 \leq r \leq 1$, that is:

$$-1 \leq \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} \leq 1 \quad (6.6)$$

in fact, if we consider the positive quantity (or equal to zero, at the least):

$Q = \langle [a(x - \langle x \rangle) + c(y - \langle y \rangle)]^2 \rangle$, we have that:

$$Q = \langle [a(x - \langle x \rangle) + c(y - \langle y \rangle)]^2 \rangle = \langle a^2(x - \langle x \rangle)^2 + 2ac(x - \langle x \rangle)(y - \langle y \rangle) + c^2(y - \langle y \rangle)^2 \rangle =$$

$$= a^2 \langle (x - \langle x \rangle)^2 \rangle + 2ac \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle + c^2 \langle (y - \langle y \rangle)^2 \rangle = a^2 \sigma^2(x) + 2ac \cdot \text{cov}(x, y) + c^2 \sigma^2(y) \geq 0$$

and the discriminant is (the variable here is "a"):

$\Delta = 4c^2[\text{cov}(x, y)]^2 - 4\sigma^2(x)c^2\sigma^2(y) = 4c^2\{[\text{cov}(x, y)]^2 - \sigma^2(x)\sigma^2(y)\}$ and this is zero when the parabola starts staying over the abscissa, so: $[\text{cov}(x, y)]^2 - \sigma^2(x)\sigma^2(y) = 0 \rightarrow$

$$\rightarrow \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} = \pm 1 = r \quad \text{qed.}$$

Moreover, through the (6.2) we noticed that: $\varepsilon(y) = a\varepsilon(x) + b$; in fact, now we ask ourself to prove that the probability that this is true is the highest; but $\varepsilon(y) = a\varepsilon(x) + b \rightarrow y - \langle y \rangle = a(x - \langle x \rangle) + b$

$\rightarrow y = ax + [b + \langle y \rangle - a\langle x \rangle] \equiv ax + b' \equiv ax + b$, that is:

$$y = ax + b \quad (6.7)$$

and we notice, more simply, that the probability that the (6.7) is true is the highest (P=1):

-first of all, we see that: (also taking into account that $P(y=ax+b)=1$)

$\langle y \rangle = \langle ax + b \rangle = a\langle x \rangle + b$, so, because of the (6.5):

$$\sigma^2(y) = \langle y^2 \rangle - \langle y \rangle^2 = \langle (ax + b)^2 \rangle - (a\langle x \rangle + b)^2 = a^2 \langle x^2 \rangle + 2ab\langle x \rangle + b^2 - (a^2 \langle x \rangle^2 + 2ab\langle x \rangle + b^2) =$$

$$= a^2(\langle x^2 \rangle - \langle x \rangle^2) = a^2 \sigma^2(x), \text{ from which:}$$

$$\sigma(y) = \pm a\sigma(x) \quad (6.8)$$

moreover: $\text{cov}(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle = \langle (x - \langle x \rangle)[(ax + b) - (a\langle x \rangle + b)] \rangle =$

$$= a \langle x^2 + \langle x \rangle^2 - 2x\langle x \rangle \rangle = a(\langle x^2 \rangle - \langle x \rangle^2) = a\sigma^2(x) = [a\sigma(x)\sigma(x)] = \pm \sigma(y)\sigma(x) \rightarrow \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} = \pm 1$$

qed.