Ottavio Serra

Statistica descrittiva e Calcolo delle probabilità.

1. Elementi di statistica descrittiva.

La statistica ha due scopi principali:

I) Ricavare da un insieme di dati, troppo numerosi per essere esaminati singolarmente e proficuamente, alcune informazioni significative per il particolare problema da studiare.

II) Fornire metodi che servano ad *imparare dall'esperienza*, giustificando, **fin dove è possibile**, il passaggio da osservazioni particolari a leggi generali.

Nel primo caso si parla di statistica descrittiva, nel secondo di statistica induttiva o inferenziale. Quest'ultima ha bisogno di nozioni di calcolo delle probabilità, mentre la statistica descrittiva non ne fa uso esplicitamente. Partiamo perciò da questa, dando alcuni elementi di statistica inferenziale dopo aver introdotto concetti e tecniche di calcolo delle probabilità.

Nella statistica descrittiva si parte da una popolazione che può essere di vario tipo: le molecole di un gas, una coltura batterica, gli studenti di una scuola, i professori di una classe. Di questa popolazione si studiano alcuni caratteri o attributi e si ripartisce la popolazione in classi, a seconda dei caratteri. Un carattere può essere

(a) qualitativo, come per esempio il mezzo di locomozione usato dagli studenti per andare a scuola;

(b) ordinale come quando si fa una scala di preferenze o di simpatia;

(c) numerico. In tal caso **l'attributo** (o **carattere**) viene detto ***variabile.***

Per un carattere di tipo (a) c'è poco da dire, l'unico dato significativo è quello di *classe modale*.

a) Moda o classe modale di una popolazione è la classe più numerosa. Per esempio, rispetto alla popolazione scolastica di una scuola, la Moda è la classe dei pedoni, se questi sono i più numerosi rispetto agli studenti che usano altri mezzi di locomozione.

b) Mediana. Se i valori della popolazione X sono ordinati, la *mediana* di X è quel valore (o quei valori) rispetto al quale X ammette tanti valori minori quanti sono i maggiori. Per esempio nella sequenza ordinata 44, 46, 50, 51, 59, la mediana è 50. (Potrebbe essere, in migliaia di euro, il reddito di 5 gruppi di persone: il reddito mediano è 50). La classe modale sarebbe il gruppo di persone col reddito più alto.

Supponiamo ora che in una popolazione S è assegnato un carattere espresso da una variabile numerica reale. In tal caso è determinata una **funzione a valori reali X: S→*R*.** S potrebbe essere un gruppo di n operai classificati secondo le ore lavorate in un anno. Se x_i sono le ore dell'i^{mo} operaio (la variabile X assume i valori x_1, x_2, .. x_n), si è solito chiamare la quantità

$$M(X) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

, ***media aritmetica*** di X.

c) Valore medio (o **atteso** o **speranza matematica**): Se una variabile X può assumere i valori x_1, x_2, ..x_n con *pesi statistici* $p_1=m_1/m$, $p_2=m_2/m$,...$p_n=m_n/m$ ($p_1+p_2+p_n=1$), la media pesata di X è M(X)= $\bar{x} = \sum_{i=1}^{n} p_i x_i$. Se $p_i=1/n$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (media aritmetica). Però rispetto alla media i valori di X possono essere più o meno concentrati o sparpagliati. Una misura della *concentrazione* può essere data dal valore medio dello scarto assoluto $s = \sum_{i=1}^{n} p_i |x_i - \bar{x}|$. Non ha senso il valor medio dello scarto, perché questo è sempre zero: $\sum_{i=1}^{n} p_i(x_i - \bar{x}) = \sum_{i=1}^{n} p_i x_i - \sum_{i=1}^{n} p_i \bar{x} = \bar{x} - \bar{x}\sum_{i=1}^{n} p_i = \bar{x} - \bar{x}.1 = 0$. Ma lo scarto assoluto è disagevole nei calcoli, perché il valore assoluto non è una funzione *liscia* (non è derivabile), perciò si usa lo scarto quadratico medio (o deviazione standard) **σ** definito come la radice quadrata della varianza: $Var(X)=\sum_{i=1}^{n} p_i(x_i-\bar{x})^2=\sum_{i=1}^{n} p_i x_i^2-2\bar{x}\sum_{i=1}^{n} p_i x_i+\bar{x}^2=M\left(x^2\right)-\bar{x}^2$, da cui $\sigma=\sqrt{Var(X)}$. Si noti che $M(x^2)$ è sempre maggiore o uguale a $\bar{x}^2$, perché? …

Esempio. Due fisici, Aldo e Bruno, misurano la massa del bosone di Higgs. Aldo fa due misure e trova i valori 125 e 127 GeV; Bruno fa pure due misure e trova 122 e 130 GeV. Calcolare i valori medi e le deviazioni standard. Chi, dei due fisici, è più affidabile?

2. Spazi di probabilità discreti.

Sia S un insieme,finito o numerabile, detto *Spazio degli Eventi;* A, B sottoinsiemi, detti *Eventi*; introdotta una *misura* m su S, chiameremo probabilità di un evento A il rapporto tra m(A) e m(S). p(A)=m(A)/m(S).

La probabilità di un evento è un numero p compreso tra 0 e 1, che gode delle seguenti proprietà:

1° $p(A\cup B) = p(A)+p(B)- p(A\cap B)$; Spiegare a livello intuitivo.
Se A e B sono disgiunti, $A\cap B=\Phi$, allora $p(A\cup B) = p(A)+p(B)$.

2° $p(A/B) = \dfrac{p(A\cap B)}{p(B)}$ La probabilità di A condizionata al verificarsi di B è la probabilità della parte di A inclusa in B relativamente a B, come se lo spazio degli eventi si *contraesse* in B: p(A/B)= $\dfrac{m(A\cap B)}{m(B)}=\dfrac{m(A\cap B)/m(S)}{m(B)/m(S)}=\dfrac{p(A\cap B)}{p(B)}$.

Per simmetria $p(B/A) = \dfrac{p(A\cap B)}{p(A)}$ e quindi

$$p(A\cap B)=p(B).p(A/B)=p(A).p(B/A).$$

A si dice indipendente <u>stocasticamente</u> da B (indipendenza <u>in senso probabilistico</u>) se p(A/B)=p(A); in tal caso $p(A\cap B)$=p(A).p(B), da cui deriva che p(B/A)=p(B): Se A è indipendente da B, B è indipendente da A. Perciò si dice che A e B sono (stocasticamente) indipendenti.

Esempi

(1) Sia S lo spazio degli eventi nel lancio di un dado: S={1,2,3,4,5,6}, A l'evento "esce un numero primo"={2,3,5}. La probabilità P(A)=m(A)/m(S)=3/6=1/2. Sia B "esce un numero pari"={2,4,6}. P(A/B)=P(A∩B)/P(B)=m(A∩B)/m(B)=1/3. In questo caso l'informazione B ha fatto diminuire la probabilità di A.

(2) Sia A="estrarre una figura da un mazzo di carte napoletane"; P(A)=12/40=3/10=0,3. Sia B l'evento "la carta estratta vale più di 5", cioè è 6 o 7 o una figura. P(B)=20/40=5/10=0,5. Calcolo ora P(A/B)=12/20=6/10=0,6. Questa volta l'informazione B ha fatto aumentare la probabilità di A.

(3) Sia A l'evento "estrarre un "Re" da un mazzo di carte napoletane, B l'evento "carta di danari". P(A)=4/40=/10, P(B)=10/40, P(A/B)= P(A∩B)/P(B)= P("Re di danari")/P("carta di danari")=(1/40)/(10/40)=1/10. Questa volta A è stocasticamente indipendente da B.

3° Se A_1, A_2, … A_n,…è una successione di eventi incompatibili (cioè a due a due disgiunti), allora

$$p(\bigcup_{k\in N} A_k)=\sum_{k=1}^{\infty} p(A_k)$$

(additività numerabile)

La probabilità dell'unione numerabile di insiemi <u>disgiunti</u> (cioè di eventi <u>incompatibili</u>) è la somma della serie delle singole probabilità. E' chiaro che la serie deve convergere a un numero <u>finito</u> (non negativo e) non maggiore di 1.

4° Se due eventi A e B sono incompatibili, cioè se il verificarsi di uno esclude il verificarsi dell'altro, e inoltre sono complementari (uno dei due si deve verificare), allora p(A)+p(B)=1.

[p+q=1]

Esempio1. In un'urna ci siano 7 gettoni bianchi e 3 neri. Se riteniamo che ogni gettone ha la stessa probabilità di essere estratto, (distribuzione ***uniforme***) allora p=P(bianco)=7/10 e q=P(nero)=3/10. (p+q=1). Se però nell'urna ci sono anche 8 gettono rossi, p+q= 7/18+3/18 = 10/18 <1.

Esempio2. Un tiratore ha probabilità p=0,7 di colpir il bersaglio: che probabilità ha di colpirlo almeno una volta in 3 tiri?

I Metodo: detta q la probabilità di non colpire (p+q=1 → q= 1-0,7 = 0.3), si ha

P(almeno una volta su 3) = (pqq+qpq+qqp)+(ppq+pqp+qpp)+(ppp) = $3(pq^2)+3(p^2q)+p^3$ =0,973

II Metodo: P(almeno una volta su 3) = 1 – P(nessuna volta) = 1- $(0,3)^3$ = 0,973.

(**Meglio il secondo metodo**!).

Esempio3. Lanciando due dadi i risultati vanno da 2 (=1+1) a 12 (=6+6). La somma 7 è la più probabile: come mai? Il primo a dare una giustificazione fu Galilei. (Si intende, dadi non truccati).

Esempio4. Due amici ugualmente bravi a briscola (o scopa) puntano somme uguali con l'accordo che vince l'intera posta chi per primo arriva a 5 vittorie. Le mogli però li interrompono quando il primo ha vinto 4 partite e il secondo 3. Come devono dividersi la posta? (Problema proposto dal cavaliere de Mèray a Pascal). Suggerimento: il primo vince il torneo se impedisce al secondo di vincere due partite consecutive: P(primo) = p+qp, P(secondo) = qq. Siccome p=q=1/2,
P(primo) =3/4 e P(secondo) = ¼, perciò la posta va divisa nella proporzione di 3 ad 1.
Generalizzare per p e q generiche e nell'ipotesi che al 1° manchino x partite e al 2° ne manchino y.

Esempio5. Qual'é la probabilità di fare ambo alla ruota di Napoli? I numeri sono 90 e se ne estraggono 5, perciò lo spazio degli eventi (lo spazio di tutte le cinquine) ha misura $C_{90,5} = \binom{90}{5} = \frac{90.89.88.87.86}{1.2.3.4.5}$. I casi favorevoli sono dati da quelle cinquine che contengono i numeri puntati, tolti i quali gli altri 3 possono essere qualunque, dunque la misura dell'evento è data dai casi favorevoli che sono $C_{88,3} = \binom{88}{3} = \frac{88.87.86}{1.2.3}$ e

P(ambo)= $\frac{88.87.86}{1.2.3} . \frac{1.2.3.4.5}{90.89.88.87.86} = \frac{4.5}{90.89} = \frac{2}{9.89} = \frac{2}{801}$.

Esercizio. Calcolare le probabilità P(terno), P(quaterna), P(cinquina).

Esempio6. Totocalcio. Supponendo di trascurare il fattore campo e la forza delle squadre, siccome ogni partita ha 3 risultati : 1, x, 2, la probabilità di azzeccare un risultato è p=1/3 e di sbagliarlo è perciò q= 2/3. La probabilità di imbroccare k risultati su 13 sarà P(k su 13)= $\binom{13}{k}(1/3)^k(2/3)^{13-k}$.

Verificare che $\sum_{k=0}^{13}\binom{13}{k}(1/3)^k(2/3)^{13-k} = 1$.

Esercizio. Qual è la probabilità di fare almeno un punto sulla schedina? E di fare meno di 3 punti?

Esempio7. Lanciando una moneta 10 volte (è lo stesso che 10 monete simultaneamente), qual'é la probabilità di ottenere esattamente 3 Teste? E quella di ottenere almeno 3 Teste?.
E' più probabile ottenere 10 Teste su 10 lanci, oppure 5 Teste seguite da 5 Croci?
Qual è la probabilità di ottenere 5 Teste e 5 Croci, se si prescinde dall'ordine?

Osservazione. In generale, se la probabilità di evento è p(E) = p e dell'evento contrario è q=1-p, la probabilità che E si verifichi k volte in una sequenza di n prove (**prove ripetute**) è

$$P(n,k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Distribuzione binomiale o di Bernoulli.

In particolare, se p=q=1/2, si ottiene

$$P(n,k) = \frac{n!}{k!(n-k)!} p^k q^{n-k} = \frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^n \quad (0 \leq k \leq n).$$

In tal caso si dimostra che P(n,k) ha un massimo per k=n/2 se n è pari o due massimi consecutivi per k=(n-1)/2 e per k=(n+1)/2 se n è dispari.
In generale si ha il massimo (o i due massimi) per k compreso tra np-q ed np+p (estremi inclusi).
Se n è grande, la distribuzione di probabilità assume la forma a campana di Gauss e possiamo dire che la probabilità massima spetta all'evento che si presenti un numero di volte pari ad np (se np è intero, altrimenti all'intero più vicino) e di conseguenza che l'evento contrario si presenti nq volte.
In una sequenza di prove ripetute la media dei successi è proporzionale a p (e la media degli insuccessi è proporzionale a q).
Esempio8. Lanciando una moneta 10 volte, l'evento più probabile è 5 Teste (e 5 Croci) **5=10.1/2.**

P(10;5)= $\frac{10!}{5!5!}\frac{1}{2^{10}}$ =3628800/(120)^{2}1/1024=252/1024=**0,246.**

P(10;4)=P(10;6)= $\frac{10!}{4!6!}\frac{1}{2^{10}}$ =3628800/(24x720)1/1024=210/1024=**0,205** < **0,246**.

Però la probabilità massima decresce al crescere di n; per esempio, lanciando la moneta 20 volte si ha **P(20;10)**=20!/(10!)^{2}1/2^{20}=184756/1048576=0,176 < **P(10;5)**; ma appena ci si allontana dal caso più probabile le probabilità sono ancora più piccole e decrescono rapidamente.
Però il calcolo esatto dei fattoriali, anche per numeri non troppo grandi, è molto laborioso. In tal caso si approssima n! con la formula di Stirling: **log(n!)=nlog(n)-n**, da cui $n! = Ae^{-n}n^n$, essendo A un fattore correttivo che vale circa $\sqrt{2\pi n}$.
Esempio9. Già conosciamo i concetti di valore medio e varianza. Basta riprendere quanto detto nella 1^a sezione (Statistica descrittiva) e interpretare i pesi statistici come probabilità. Lanciando un dado, il punteggio va da 1 a 6; si chiede il punteggio medio. La variabile aleatoria X assume i valori 1, 2,… 6 con probabilità costante p=1/6; perciò

$$\bar{x} = \frac{1}{6}(1+2+3+4+5+6) = \frac{1}{6}\left(\frac{6*7}{2}\right) = \frac{1}{6}\left(\frac{42}{2}\right) = 3,5.$$

Il valor medio del quadrato è

$$M(X^2) = \frac{1}{6}(1^2+2^2+...+6^2) = \frac{91}{6} \approx 15,167$$

perciò la varianza

$$Var(X) = \sigma^2 = M(X^2) - \bar{x}^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \approx 2,917$$

e lo scarto quadratico medio è $\sigma = \sqrt{2,917} \approx 1,708$.
Esempio10. Si voglia calcolare il numero medio di tiri necessari per colpire un bersaglio, essendo p la probabilità costante di colpirlo in ciascun tiro.
Posto q=1-p, si ha: $\bar{x} = 1.p + 2qp + 3q^2p + 4q^3p + ... + nq^{n-1}p + ... = p\sum_{n=1}^{\infty} nq^{n-1}$. Questa serie converge ad 1/(1-q)^2 =1/p^2 (*credetemi sulla parola !*), perciò $M(x) \equiv \bar{x} = p\frac{1}{p^2} = \frac{1}{p}$. (abbastanza intuitivo, col senno di poi!).
Osserviamo ancora che M(λX)=λ.M(X) e che Var(λX)=λ^{2}Var(X) da cui segue $\sigma(\lambda X) = |\lambda|\sigma(X)$.
UN CASO IMPORTANTE: valor medio e varianza della variabile **X = successo con probabilità p:**

P(x=1)=p, P(x=0)=q, [p+q=1]. $M(X) \equiv \bar{x} = p.1 + q.0 = p.$

$Var(X) = p(1-\bar{x})^2 + q(0-\bar{x})^2 = p(1-p)^2 + q(0-p)^2 = pq^2 + qp^2 = pq(p+q) = pq.$

Lo scarto quadratico medio sarà $\sigma = \sqrt{pq}$.

Prove ripetute. Supponiamo ora di avere n variabili indipendenti $X_1, X_2, \ldots X_n$ con la stessa legge: stesso valore medio $\bar{x}$ e stessa varianza σ^2. Posto $Y = X_1+X_2+\ldots X_n$, avremo $M(Y) = n\bar{x}$ e $Var(Y) = Var(X_1 + \ldots X_n) = n\sigma^2$, quindi $\sigma_Y = \sigma_X \sqrt{n}$. Se le n variabili contano i successi in n prove, M(Y) = np e $\sigma_Y = \sqrt{npq}$. Come si vede, la dispersione cresce indefinitamente con n. Però, se consideriamo la variabile Y/n, che conta la percentuale dei successi, cioè la frequenza relativa dei successi: ν/n, il suo valor medio è np/n = p e Var (Y/n)= σ^2/n =pq/n; perciò

$\sigma_Y = \sqrt{\frac{pq}{n}}$, (*Deviazione standard relativa*) **diminuisce al divergere di n**.

Teorema di Cebicev e legge dei grandi numeri di Jakob Bernoulli.

Sia Y la variabile aleatoria (ν/n –p) e $y_1, y_2, \ldots y_n$ i suoi valori; il valor medio dei quadrati è

$$M(Y^2) = \sigma_r^2 = \sum_{i=1}^{n} p_i y_i^2 .$$

Fissato un arbitrario numero reale positivo ε, alcuni degli y_i, diciamo i primi k, sono in modulo minori di ε, i rimanenti maggiori o uguali. Perciò

$$\varepsilon^2 \sum_{i=k+1}^{n} p_i \le \sigma_r^2$$

. Segue

$$\sum_{i=k+1}^{n} p_i \equiv P\left(\left|\frac{\nu}{n} - p\right| \ge \varepsilon\right) \le \frac{\sigma_r^2}{\varepsilon^2}$$

da cui $P\left(\left|\frac{\nu}{n} - p\right| < \varepsilon\right) > 1 - \frac{\sigma_r^2}{\varepsilon^2}, quindi$

$$P\left(\left|\frac{\nu}{n} - p\right| < \varepsilon\right) > 1 - \frac{pq}{n\varepsilon^2} .$$

Questa disuguaglianza, trovata da Cebicev (1821-1894), **giustifica** la legge dei grandi numeri di Bernoulli (1654-1705):

La probabilità che in una serie di prove ripetute lo scarto tra la frequenza relativa e la probabilità a priori p sia minore in valore assoluto di un arbitrario numero reale positivo ε tende ad 1 (alla certezza) al crescere del numero n delle prove:

$$\lim_{n\to\infty} P\left(\left|\frac{\nu}{n} - p\right| < \varepsilon\right) = 1$$

Ciò non significa che ν/n → p per n → ∞, ma solo che tende alla certezza la probabilità di uno scarto arbitrariamente piccolo.

Esempio 11. Quante volte devo lanciare una moneta perché la frequenza relativa dell'evento "Testa" sia compresa tra 0,49 e 0,51 con una probabilità maggiore del 90%? Si ha p=q=0,5, ε = 0,01; perciò $1 - \frac{1}{4n10^{-4}} = \frac{9}{10} \Rightarrow \frac{1}{4n10^{-4}} = \frac{1}{10} \Rightarrow n = 25000$. (Ci vogliono 25000 lanci).

Vedremo in un'altra lezione che si possono ottenere valutazioni più stringenti (valori di n più piccoli).

NOTA IMPORTANTE. Non crediate che il calcolo delle probabilità serve solo per giocare a carte o per lanciare dadi e monete. La fisica è imbevuta di calcolo delle probabilità, dalla teoria della misura, alla termodinamica, alla meccanica quantistica.

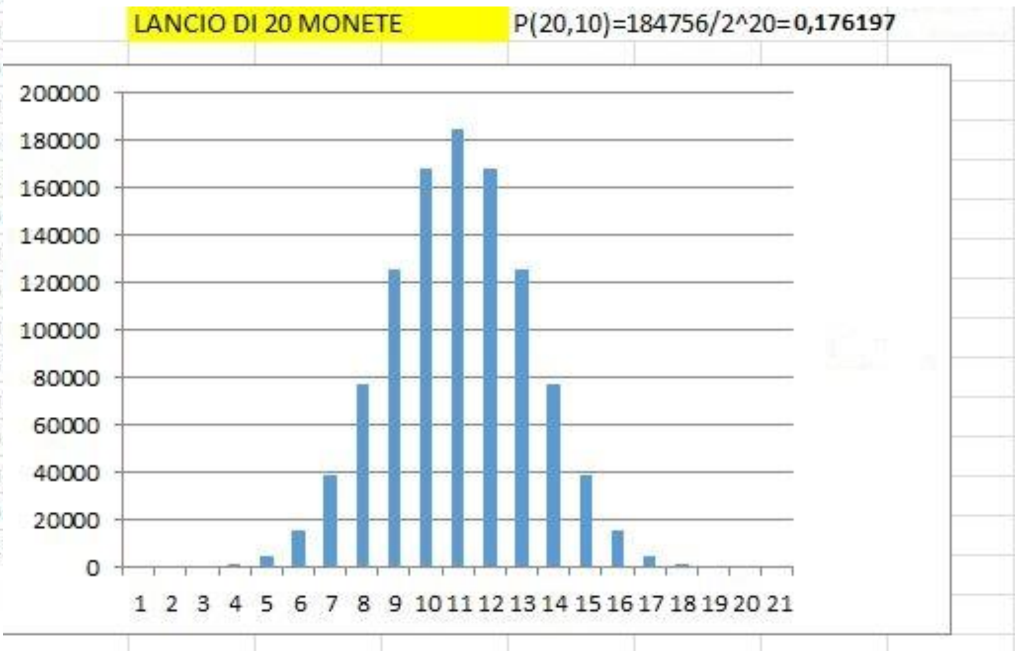
Un esempio elementare. In una scatola di 1 cm^3 ci siano 6 molecole. La distribuzione delle molecole nelle due metà destra e sinistra segua la distribuzione uniforme, cioè per ciascuna molecola sia ½ la probabilità di trovarsi a sinistra (e lo stesso a destra). Si tratta di valutare la probabilità che a sinistra ci siano zero molecole, una, … sei. Formalmente è come lanciare sei monete e calcolare la probabilità di k teste (o di k molecole a sinistra; k da 0 a 6). Ecco il grafico:

k	C(6,k)
0	1
1	6
2	15
3	20
4	15
5	6
6	1

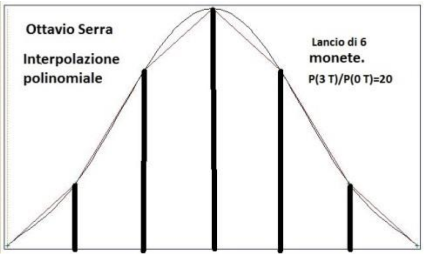
k=0 1 2 3 4 5 6

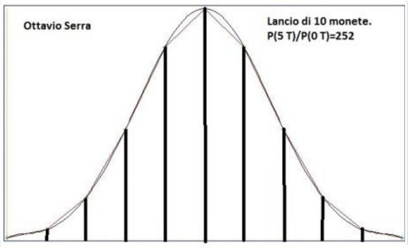
P(6,0)=1.(1/2)^6, P(6,1)=6.(1/2)^6, P(6,2)=15.(1/2)^6, **P(6,3)=20.(1/2)^6=0,3125 (Max)**, poi decrescono.
Il caso medio (3 molecole a sinistra o 3 Teste) è 20 volte più probabile di un caso estremo.
Lanciamo ora 20 monete, o mettiamo nella scatola 20 molecole: ora la probabilità massima si verifica per 10 Teste (e 10 croci) o, se preferite, per 10 molecole a sinistra (e 10 a destra). Però ora la probabilità massima è 0,176197, poco più della metà del caso precedente. Ma la cosa importante che ora il caso medio (10 molecole a sinistra) è più di 180000 volte più probabile di un caso estremo, il ché fa sì che l'istogramma è più concentrato e slanciato e scarti grandi dal valore medio sono molto improbabili. (**Vedi figura a fine pagina**).
Immaginate che succede in un cm^3 d'aria; in condizioni normali di temperatura e pressione ci sono 10^{19} molecole. La probabilità di uno scarto relativo di un miliardesimo dal valore medio (5.10^{18} molecole a sinistra e altrettante a destra) è praticamente nulla e si giustifica così come mai un gas abbandonato a se stesso finisce (in un tempo brevissimo) per raggiungere uno stato di densità uniforme in ogni punto del recipiente.

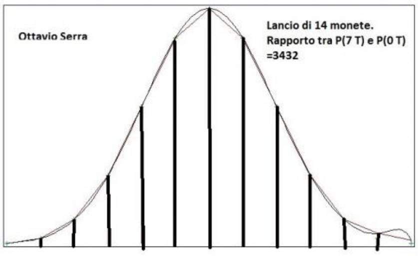
k	C(20,k)
0	1
1	20
2	190
3	1140
4	4845
5	15504
6	38760
7	77520
8	125970
9	167960
10	184756
11	167960
12	125970
13	77520
14	38760
15	15504
……………	
20	1



k=0, 1, … 9, 10, 11,… 19, 20







Dagli istogrammi della pagina precedente, nonché dalle interpolazioni polinomiali sovrastanti, si ricava un'informazione molto importante.

In primo luogo, la probabilità che, per esempio, Testa esca 5 o 6 o 7 volte su 14 lanci, è la somma delle singole probabilità:

P(5T o 6T o 7T)= P(5T)+P(6T)+P(7T), ma ciò è ovvio: gli eventi sono incompatibili e la probabilità dell'unione è la somma delle probabilità dei singoli eventi, come vuole la legge della distribuzione binomiale di Bernouilli: se in una sequenza di n prove il successo ha probabilità costante p,

$$P(n;\mathrm{m}_1 \le \mathrm{k} \le \mathrm{m}_2) = \sum_{k=m_1}^{m_2} \binom{n}{k} p^k (1-p)^{n-k} .$$

In secondo luogo, tale probabilità è all'incirca uguale all'*area* sottesa dalla curva di interpolazione polinomiale (**curva a "campana**", tra le *ascisse* 5 e 7.

L'approssimazione è tanto migliore quanto più fitto è l'istogramma, cioè quanto più numerose sono le monete (o i lanci) o quante più molecole ci sono nel recipiente. Questa osservazione giustificherà il passaggio a una distribuzione **continua** di probabilità e alla sostituzione di una **sommatoria** con un **integrale**. (Dalla distribuzione binomiale di Bernouilli alla distribuzione normale di Gauss).