Distribuzioni teoriche

A. DISTRIBUZIONI DISCRETE

- Distribuzione binomiale
- Distribuzione multinomiale
- Distribuzione di Poisson
- Distribuzione ipergeometrica
- Distribuzione uniforme

B. DISTRIBUZIONI CONTINUE

- Distribuzione normale o di Gauss
- Distribuzione rettangolare
- Distribuzione esponenziale negativa
- Le curve di Karl Pearson
- Distribuzione Gamma
- Distribuzione Beta
- Distribuzione chi-quadrato
- Distribuzioni F di Fisher e t di Student

Distribuzione binomiale

Momento terzo rispetto all'origine

$$M_3 = \sum_{i=0}^{i=n} x_i^3 . P(x_i) = \sum_{x=0}^{x=n} x^3 . \frac{n!}{x!(n-x)!} . p^x . q^{n-x}$$

e poiché $x^3 = x[x+x(x-1)]$,

$$M_3 = np + 3n(n-1)p^2 + n(n-1)(n-2)p^3$$

Momento terzo rispetto alla media

$$\mu_3 = M_3 - 3.M_2.M_1 + 2.M_1^3 = npq(q-p).$$

Asimmetria

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{npq(q-p)}{npq\sqrt{npq}} = \frac{q-p}{\sqrt{npq}}.$$

Coefficiente di asimmetria del Pearson

$$\beta_1 = \alpha_3^2 = \frac{\mu_3^2}{\mu_2^3} = \frac{(q-p)^2}{npq}.$$

10. **Appiattimento o Kurtosis o disnormalità**

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} = \frac{1}{(npq)^2}\left[3n^2p^2q^2 + npq(1-6pq)\right] = 3 + \frac{(1-6pq)}{npq}$$

***Coefficiente di appiattimento del Pearson*.**

$$\beta_2 = \alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} = 3 + \frac{(1-6pq)}{npq}.$$

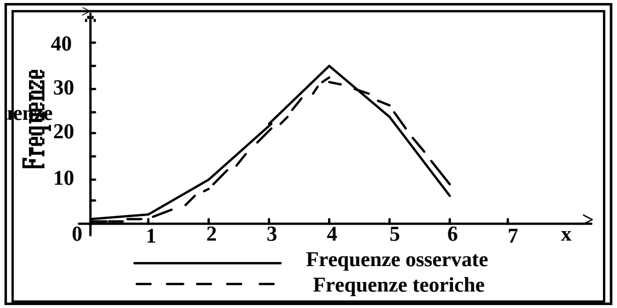
La curva binomiale è più aguzza della normale
ossia è leptokurtica o ipernormale
(Per la curva normale è $\beta_2 = 3$)

Esempio In un villaggio si esegue un'inchiesta sul numero dei campi gestiti direttamente dal proprietario. Da precedenti esperienze si sa che la proporzione dei campi gestiti direttamente dal proprietario è pari a 2/3. Estratti 100 campioni di 6 campi ciascuno i risultati sono:

Numero di campi gestiti dal proprietario sui 6 campi esaminati	Proporzione dei campi gestiti dal proprietario	Numero di campioni che danno tale risultato
0	0	1
1	1/6	2
2	2/6	10
3	3/6	22
4	4/6	35
5	5/6	24
6	6/6	6
		100

Adattiamo a tali dati una distribuzione binomiale e confrontiamo le frequenze teoriche calcolate con i risultati della rilevazione campionaria. Inoltre, troviamo media e deviazione standard di questa distribuzione e tracciamo un grafico di confronto tra frequenze osservate e teoriche.

In questo esempio, $p = 2/3$, $q = 1/3$, mentre $n = 6$ ed $N = 100$. La tavola che segue mostra le frequenze teoriche trovate calcolando i termini nell'espansione del binomio:



Avremo, quindi la tavola di confronto seguente:

x	Frequenze osservate	Frequenze teoriche
0	1	0,137
1	2	1,646
2	10	8,230
3	22	21,948
4	35	32,922
5	24	26,338
6	6	8,779
Totale	100	100,000

x	f	xf	x^2f
0	1	0	0
1	2	2	2
2	10	20	40
3	22	66	198
4	35	140	560
5	24	120	600
6	6	36	216
	100	384	1616

per cui sarà: $M = 384/100 = 3,84$; $M_2 = 1616/100 = 16,16$

$s^2 = M_2 - M^2 = 16,16 - (3,84)^2 = 16,16 - 14,75 = 1,41$

$s = \sqrt{1,41} = 1,187$.

La media e la deviazione standard dei valori teorici, invece, saranno: $M = np = 6.\ 2/3 = 4$,

$s = \sqrt{(6.2/3.1/3)} = \sqrt{4/3} = 1,153$.

Distribuzione multinomiale

Estensione della binomiale per k eventi indipendenti di probabilità p1, p2, ..., pi, ..., pn, la cui somma è uguale ad 1, che possono comparire nel corso di N prove indipendenti, successive o simultanee.

La probabilità della combinazione di eventi n1, n2, ..., nk in N prove è determinata dal multinomio:

$$\boxed{P_{(n_1, n_2, \ldots, n_k)} = \frac{N!}{n_1!\, n_2! \ldots . n_k!} p_1^{n_1} p_2^{n_2} \ldots . p_k^{n_k}}$$

Per k qualunque, 0<=p<=1 e $\sum_{i=1}^{i=k} p_i = 1$.

Distribuzione di Poisson

Per n tendente all'infinito (cioè il numero dei dati è molto grande) e p prossimo a zero (cioè la probabilità che l'evento ha di verificarsi è molto piccola), ma tali che n.p rimanga costante, una buona approssimazione alla binomiale è data dalla *funzione di distribuzione di Poisson*:,

$$\boxed{P(x) = \frac{e^{-M} . M^{x}}{x!}} \text{ se } \boxed{\begin{cases} n \to \infty \\ p \to 0 \\ n.p = \cos t. \end{cases}}.$$

La funzione di distribuzione di Poisson fornisce adattamenti molto buoni in problemi riferiti ad un *evento raro*

(Es.manifestazioni di patologie o eventi catastrofici rari)

La binomiale come approssimazione della poissoniana

$$\boxed{P(x) = \frac{n!}{x!(n-x)!} . p^{x} . q^{n-x}} \qquad (1)$$

Se np = a, è p = a/n e sostituendo p con a/n e q con (1 - a/n):

$$\boxed{P(x) = \frac{n!}{x!(n-x)!} . \left(\frac{a}{n}\right)^{x} . \left(1 - \frac{a}{n}\right)^{n-x} =}$$

$$\boxed{= \frac{n(n-1)(n-2) \ldots . (n-x+1)}{n^{x}} . \frac{a^{x}}{x!} . \left(1 - \frac{a}{n}\right)^{n-x}}$$

Poiché $\boxed{\underset{n \to +\infty}{\text{Limite}} \frac{n(n-1)(n-2) \ldots . (n-x+1)}{n^{x}} = 1}$ **e** $\boxed{\underset{n \to +\infty}{\text{Limite}} \left(1 - \frac{a}{n}\right)^{n-x} = e^{-a}}$

Sarà $\boxed{P(x) = \frac{a^{x}}{x!} . e^{-a}}$

Caratteristiche della distribuzione di Poisson

La distribuzione di Poisson è una *distribuzione teorica discreta* che dipende o è totalmente *definita dal solo parametro a*, cioè la *media aritmetica* della distribuzione di Poisson

L'approssimazione di Poisson o *legge dei piccoli numeri* si considera buona quando è p<0,03 o quando np < 5.

In questi casi sono assai frequenti le classi con zero o con pochi eventi rispetto a quelle con numerosi eventi.

Momenti della distribuzione di Poisson

Media o momento primo rispetto all'origine

$$\boxed{M = \sum_{x_i=0}^{x_i=+\infty} x_i . P(x_i) = \sum_{x=0}^{x=+\infty} x \frac{a^x}{x!} e^{-a} =}$$

$$= e^{-a}\left(0 + a + 2\frac{a^2}{2!} + 3\frac{a^3}{3!} + 4\frac{a^4}{4!} + \ldots\ldots\right) =$$

$$\boxed{= a.e^{-a}\left(1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \frac{a^4}{4!} + \ldots\ldots\right) = a.e^{-a}.e^{a} = a}$$

La funzione di distribuzione di Poisson dipende da un solo parametro che coincide con la media aritmetica della distribuzione.

Momento secondo rispetto alla media

$$\boxed{\mu_2 = \sigma^2 = M_2 - M_1^2 = a + a^2 - a^2 = a = M}$$

La varianza della distribuzione di Poisson è esattamente uguale alla media della distribuzione stessa

La deviazione standard della distribuzione di Poisson è uguale alla radice quadrata della media aritmetica della distribuzione.

Momento terzo e misura dell'asimmetria

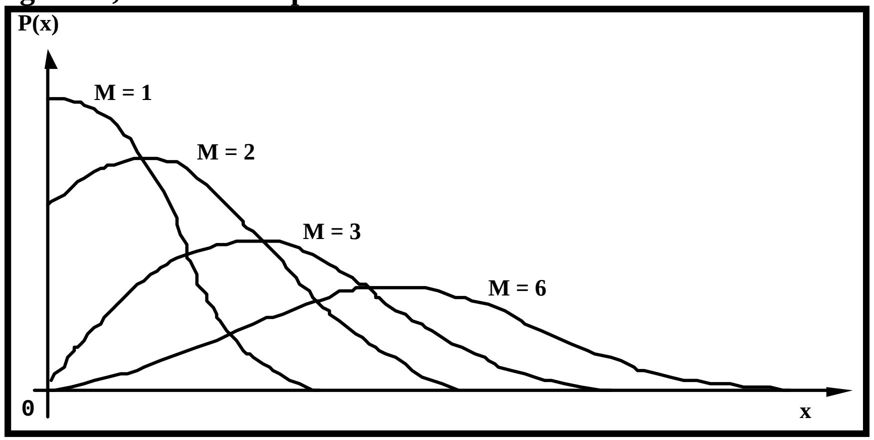
Stesso procedimento già mostrato per il calcolo del momento secondo. Risulta μ_3=M. Per p tendente a zero, np = M, q tendente all'unità ed n tendente all'infinito si hanno le seguenti misure di asimmetria:

$$\boxed{\alpha_3 = \lim_{n\to+\infty} \frac{q-p}{\sqrt{npq}} = \frac{1}{\sqrt{M}}} \quad e \quad \boxed{\beta_1 = \alpha_3^2 = \frac{1}{M}}$$

La forma della poissoniana è molto asimmetrica per valori piccoli della media aritmetica (inferiori a 3)

Forma generale del poligono di Poisson

Cambia al variare di M e di x. Per bassi valori di M il poligono assume una forma molto asimmetrica, ma man mano che M diventa più grande, essa diviene più simmetrica.



Esempio IV.2 Si riporta la distribuzione di frequenza dei decessi di mucche per una malattia rara in 50 provincie durante un periodo di 10 anni:

Numero di mucche decedute	Numero di provincie in 10 anni
0	240
1	150
2	60
3	25
4	17
5	8
Totale	500

Su questi dati calcoliamo le frequenze della distribuzione di Poisson che ha la stessa media e compariamo i risultati con le frequenze osservate.

Si ha:

x	f	xf
0	240	0
1	150	150
2	60	120
3	25	75
4	17	68
5	8	40
	500	453

np=M=453/500=0,906;

$E^{-0,906} = 0,4044$

$P_0 = 0,4044$ $\quad$ $P_1 = 0,4044*0,906 = 0,3664$

$P_2 = 0,3664*0,906/2 = 0,1660$ $\quad$ $P_3 = 0,1660*0,906/3 = 0,0501$

$P_4 = 0,0501*0,906/4 = 0,0113$ $\quad$ $P_5 = 0,0113*0,906/5 = 0,0020$

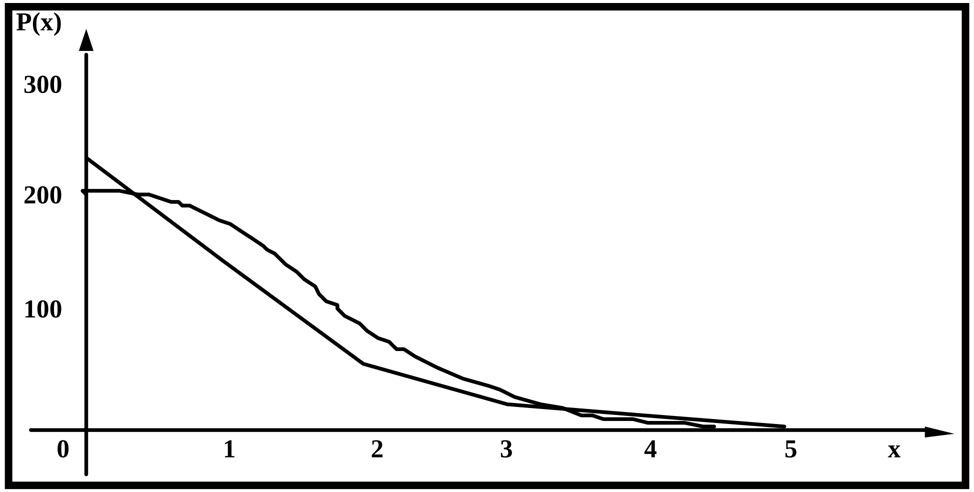
X	F	XF	P(X)	N.P(X)
0	240	0	0,4044	202,20
1	150	150	0,3664	183,20
2	60	120	0,1660	83,00
3	25	75	0,0501	25,05
4	17	68	0,0113	5,65
5	8	40	0,0020	1,00
	500	453		

Se si vuole misurare l'adattamento della curva poissoniana ai dati osservati e, nello stesso tempo, si vuole valutarne l'attendibilità, si può calcolare una media quadratica degli scarti tra i valori osservati ed i valori teorici ottenuti, operando nel modo indicato nella tabella che segue:

x	f	N.P(x)	Δ=[N.P(x)-f]	Δ^2	Δ^2/N.P(x)
0	240	202,20	- 37,8	1428,84	7,0665
1	150	183,20	+ 33,2	1102,24	6,0166
2	60	83,00	+ 23,0	529,00	6,3735
3	25	25,05	0	0	0
4	17	5,65	- 11,4	129,96	23,2071
5	8	1,00	- 7,0	49,00	49,0000
	500	500,00	0	3239,04	91,6637

Come si vede, se si considera il test del χ^2 per misurare la bontà dell'adattamento, si ha un valore di $\chi^2 = 91,6637$ cioè un valore che può essere superato per effetto del caso solo con una probabilità molto piccola. Ciò vuol dire che l'adattamento ottenuto con la funzione di Poisson non è molto buono e che si dovrebbe procedere alla selezione di una funzione di distribuzione di altro tipo per rappresentare il fenomeno studiato.

Rappresentando in un grafico i risultati ottenuti per i valori teorici e, insieme ad essi, anche i dati osservati si ottiene la distribuzione di Poisson adattata al diagramma di frequenza dato, che si riporta nel grafico seguente:



Distribuzione ipergeometrica

Nella distribuzione binomiale la *probabilità* di un evento si mantiene sempre costante. Quando essa *varia* in funzione degli eventi precedenti, come succede nell'estrazione *senza ripetizione* di alcuni oggetti da un campione di piccole dimensioni, si ha la *distribuzione ipergeometrica*.

Ad esempio, qual'è la probabilità che sia un re la seconda carta estratta da un mazzo di 40 carte? Se il gioco avviene con reimmissione della carta già estratta, la probabilità di estrarre un re è costantemente pari a 4/40. Ma se il gioco si svolge senza reintroduzione nel mazzo della carta estratta per prima, la probabilità che la seconda carta sia un re varia in rapporto all'estrazione della prima: se la prima era un re, la probabilità per la seconda estrazione è pari a 3/39; se la prima era una carta diversa, la probabilità che la seconda sia un re è 4/39.

La distribuzione ipergeometrica permette di calcolare la probabilità di una data combinazione di eventi quando le probabilità dei vari eventi sono variabili da prova a prova

$$P(N,n_1;n,r)=\frac{C(n_1,r).C(N-n_1,n-r)}{C(N,n)}=\frac{\binom{n_1}{r}\binom{N-n_1}{n-r}}{\binom{N}{n}}$$

N = unità della popolazione;
n = unità del campione, al massimo pari a N;
n_1 = unità della popolazione che hanno la caratteristica in esame, al massimo pari a N;
r = unità del campione estratte, che hanno la caratteristica in esame, al massimo pari a n.

La distribuzione ipergeometrica è definita da *tre parametri* (N, n_1, n: totale unità della popolazione, unità di popolazione che hanno il carattere in esame, numero di unità estratte) *in funzione di r* (numero di unità estratte che hanno il carattere in esame).
Momenti della distribuzione ipergeometrica

Ponendo p=n1/N e q = (N-n1)/N si ha:

$$\boxed{P(N,Np;n,r)=\frac{\binom{Np}{r}\binom{Nq}{n-r}}{\binom{N}{n}}}$$

Media aritmetica e varianza:

$$\boxed{M=np} \quad e \quad \boxed{\sigma^2=\frac{npq(N-n)}{N-1}}.$$

La media della distribuzione ipergeometrica è uguale a quella della distribuzione binomiale corrispondente, mentre la varianza è inferiore.

La distribuzione ipergeometrica, per N tendente ad infinito, ossia per N grande rispetto ad n, tende alla distribuzione binomiale.

Esempio IV.3 **Un'urna contiene N biglie, delle quali n_1 bianche e N-n_1 nere. Si estraggono dall'urna n biglie (con n <= N) senza reimmissione; si vuole determinare la probabilità che delle n biglie estratte r siano bianche (con r <= n). Si ha:**

$$\boxed{P(N,n_1;n,r)=\frac{\binom{n_1}{r}\binom{N-n_1}{n-r}}{\binom{N}{n}}=}$$

$$=\frac{n_1!(N-n_1)!n!(N-n)!}{r!(n_1-r)!(n-r)!(N-n_1-n+r)!N!}$$