Distribuzione Binomiale

La distribuzione (finita, discreta) binomiale si origina dall'osservazione ripetuta (n volte) di una *prova di Bernoulli*, caratterizzata da due esiti che chiameremo "successo" e "insuccesso" con probabilità p e (1-p) rispettivamente. La probabilità di successo p non si altera ad ogni successiva osservazione, che viene quindi definita come indipendente; il conteggio di successi in n sequenze di osservazioni determina la *variabile aleatoria binomiale*.

La probabilità di osservare k = 4 successi in n = 10 prove indipendenti, con p = 0.5 si determina come

$$P(k=4|n=10,p=0.5)=\frac{n!}{k!\,(n-k)!}\cdot p^k\cdot(1-p)^{(n-k)}$$
$$=\frac{10\times 9\times 8\times 7\times \cancel{6\times 5\times 4\times 3\times 2}}{4\times 3\times 2\times \cancel{6\times 5\times 4\times 3\times 2}}\cdot 0.5^{10}=0.2051$$

Naturalmente, essendo p = 0.5, la probabilità coincide con la quantità simmetrica k = 6 successi,

$$P(k=6|n=10,p=0.5)=\frac{10\times 9\times 8\times 7\times \cancel{6\times 5\times 4\times 3\times 2}}{\cancel{6\times 5\times 4\times 3\times 2}\times 4\times 3\times 2}\cdot 0.5^{10}=0.2051$$

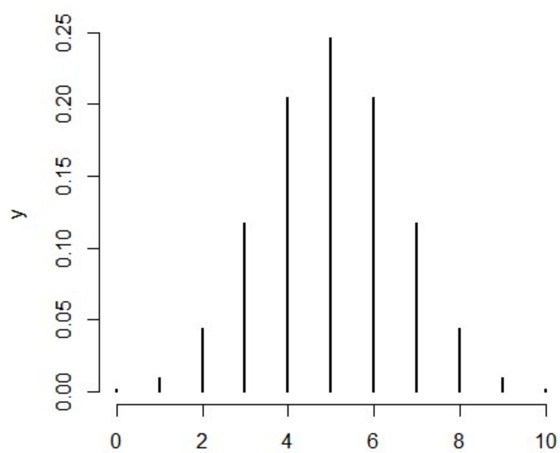


Figura 1 *Distribuzione binomiale per n = 10, p = 0.5. Si noti la simmetria della distribuzione per valori attorno al valore di k = 5.*

Se modifichiamo il valore di p = 0.17, rendendo (MOLTO) meno probabile il successo, allora la probabilità di osservare k = 4 successi dovrà necessariamente essere più bassa:

$$P(k=4|n=10,p=0.17)=\frac{10!}{4!\times 6!}\cdot 0.17^4\cdot 0.83^6=\ 0.0573$$

La probabilità di osservare k = 6 successi non sarà più uguale, infatti la distribuzione non è simmetrica attorno a k = 5,

$$P(k=6|n=10,p=0.17)=\frac{10!}{6!\times 4!}\cdot 0.17^6\cdot 0.83^4=0.0024$$

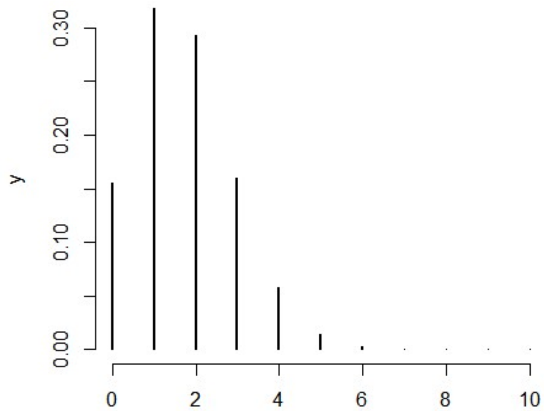


Figura 2 Distribuzione di probabilità discreta binomiale per n = 10, p = 0.17. Si noti l'asimmetria della distribuzione.

Vediamo come si modifica la binomiale osservando n = 30 ed n = 100 prove indipendenti di Bernoulli, sempre con p = 0.17.

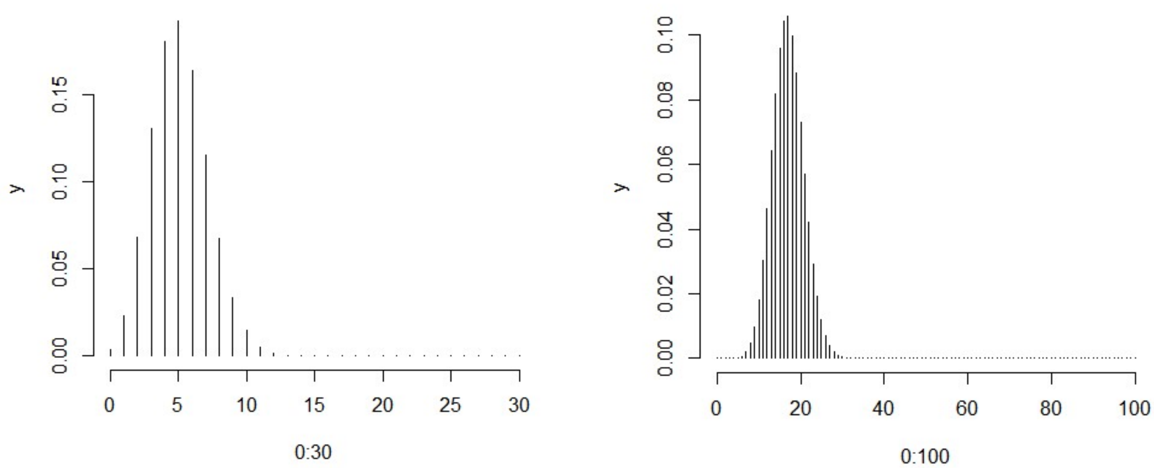


Figura 3 N =30 e N = 100. La distribuzione tende alla normalità.

Approssimazione Normale

La variabile aleatoria binomiale è la somma di n variabili indipendenti di Bernoulli, con probabilità di successo p,

$$b_{(n,p)} = \sum_{i=1}^{n} Bern_{(p)}$$

Secondo il Teorema del Limite Centrale, la somma di valori aleatori indipendenti e identicamente distribuiti tende a $N(\mu, \sigma)$ al crescere della grandezza campionaria (asintoto). La variabile binomiale, come somma di n valori campionari indipendenti (con sequenze di 0 e 1), tenderà quindi a distribuirsi secondo una variabile aleatoria normale, formalmente

$$b_{(n,p)} \approx N(E(b), \sqrt{Var(b)})$$

Consideriamo il caos particolare di n = 1, quindi una singola prova di Bernoulli:

$$E\left(Bern_{(p)}\right) = \sum_{i=1}^{2}(Bern_i \times prob_i) = 1 \times p + 0 \times (1-p) = p$$

$$Var\left(Bern_{(p)}\right) = \sum_{i=1}^{2}\left(Bern_i - E(Bern)\right)^2 \times prob_i =$$

$$= (1-p)^2 p + (-p)^2(1-p)$$

$$= p - 2p^2 + p^3 + p^2 - p^3$$

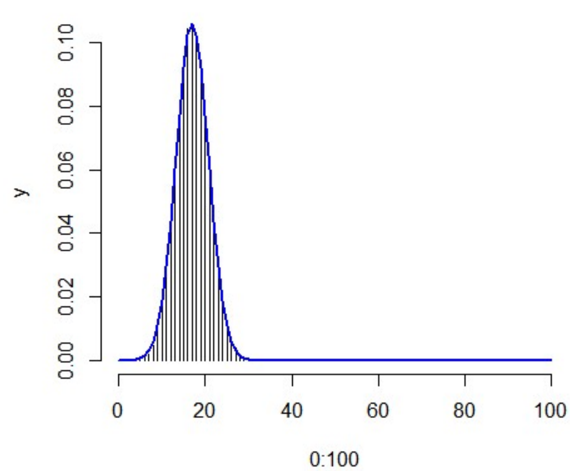
$$= p - p^2 = p(1-p)$$

Se consideriamo n prove indipendenti di Bernoulli avremo quindi

$$E(b_{(n,p)}) = E\left(\sum_{i=1}^{n} Bern_i\right) = \sum_{i=1}^{n} E(Bern_i) = np$$

$$Var(b_{(n,p)}) = Var\left(\sum_{i=1}^{n} Bern_i\right) = \sum_{i=1}^{n} Var(Bern_i) = np \cdot (1-p)$$

Ad esempio, per n = 100 osservazioni indipendenti di Bernoulli e probabilità di successo p = .17, il valore atteso è uguale a 17 e la varianza è pari a 14.11. Per il teorema del limite centrale, estendendo la sommatoria di successi fino ad n molto elevati (n = 100, come nel nostro caso), la forma matematica della distribuzione binomiale approssima alla distribuzione gaussiana, con opportuna sostituzione $\mu = np$ e $\sigma = \sqrt[2]{np \cdot (1-p)}$.



N 100 successi	Probabilità binomiale	normale
3	0,000	0,000
4	0,000	0,000
5	0,000	0,001
6	0,001	0,002
7	0,002	0,003
8	0,005	0,006
9	0,010	0,011
10	0,018	0,019
11	0,031	0,030
12	0,046	0,044
13	0,064	0,060
14	0,082	0,077
15	0,096	0,092
16	0,104	0,103
17	0,106	0,106
18	0,100	0,103
19	0,088	0,092
20	0,073	0,077
21	0,057	0,060
22	0,042	0,044
23	0,029	0,030
24	0,019	0,019
25	0,012	0,011
26	0,007	0,006
27	0,004	0,003
28	0,002	0,002
29	0,001	0,001
30	0,001	0,000
31	0,000	0,000

Stima di Massima Verosimiglianza: la definizione operativa di p.

Chiediamo a N = 20 persone se voteranno Si (1) oppure No (0) alle prossime consultazioni popolari, contando i "Si" in k = 5 persone. Come possiamo stimare la probabilità di "Si" nella Popolazione?

La sequenza registrata nel nostro campione potrebbe essere, ad esempio

$$X = \{1,0,0,1,0,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0\},$$

con funzione di probabilità (binomiale)

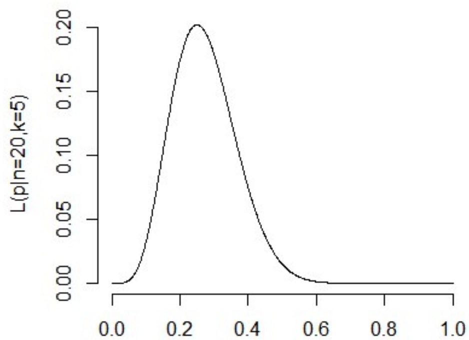
$$P(k = 5|n = 20, p) = \frac{20!}{5!\,(20-5)!} \cdot p^5 \cdot (1-p)^{(10-5)}.$$

Come stimare p a partire dai dati campionari? Cercando la definizione operativa di p che rende massima la probabilità di ottenere i dati osservati nel campione, la cui funzione è

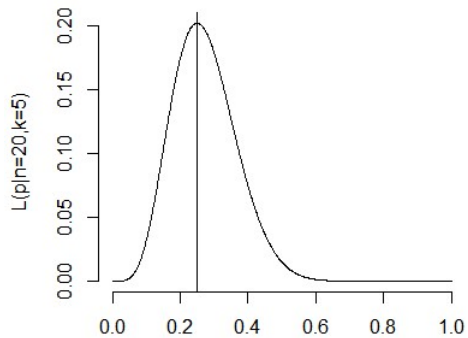
$$L(p|n = 20, k = 5) = \frac{20!}{5!\,(20-5)!} \cdot p^5 \cdot (1-p)^{(20-5)},$$

detta funzione di verosimiglianza (*likelihood*), e che si legge come: *la verosimiglianza di una probabilità p per il "Si" nella popolazione, dato che su venti soggetti intervistati cinque hanno risposto "Si"*.

Qual è il valore di p che rende massima la funzione di verosimiglianza? Iniziamo con il definire tutti i possibili valori di *p (0, 0.1, 0.2,…,1)* e vediamo come varia L.



La *verosimiglianza* di osservare 5 "Si" in 20 interviste raggiunge il valore massimo quando p = 0.25.



Detto altrimenti, se p (ignota) fosse effettivamente 0.25, ciò corrisponderebbe al valore massimo di probabilità di osservare 5 successi ("Si") in 20 prove. Valori superiori e inferiori a questo massimo sono meno *verosimili*, dati i risultati del campione intervistato. Il valore del parametro p a cui corrisponde il massimo della funzione L è la *stima di massima verosmiglianza di p (MLE)*, dati i risultati nel campione.

La soluzione analitica per la stima di massima verosimiglianza si ottiene ponendo uguale a zero la derivata prima della funzione L, considerando opportunamente il suo logaritmo naturale (ln) per semplificarne i calcoli.

Ricordiamo la nostra specifica sequenza "Si-No" osservata nel campione di n = 20 soggetti intervistati:

$$X = \{1,0,0,1,0,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0\},$$

con probabilità,

$$P(dati|p) = p^5 \cdot (1-p)^{(10-5)}.$$

Il coefficiente binomiale viene volutamente ignorato in quanto costante e non dipendente da p. La funzione di verosimiglianza si può quindi scrivere come

$$L(p|dati) = p^k \cdot (1-p)^{(n-k)},$$

ed esprimere nel suo logaritmo naturale, detto **log-verosmiglianza**, con la seguente funzione:

$$\ln L(p|dati) = \mathrm{k} \cdot \ln(p) + (n-k) \cdot \ln(1-p) \quad .$$

A questo punto si calcola la derivata prima in '*p*' della funzione di log-verosimiglianza, ponendola uguale a zero e risolvendo. Possiamo utilizzare a tal fine il calcolatore, per evitare errori:

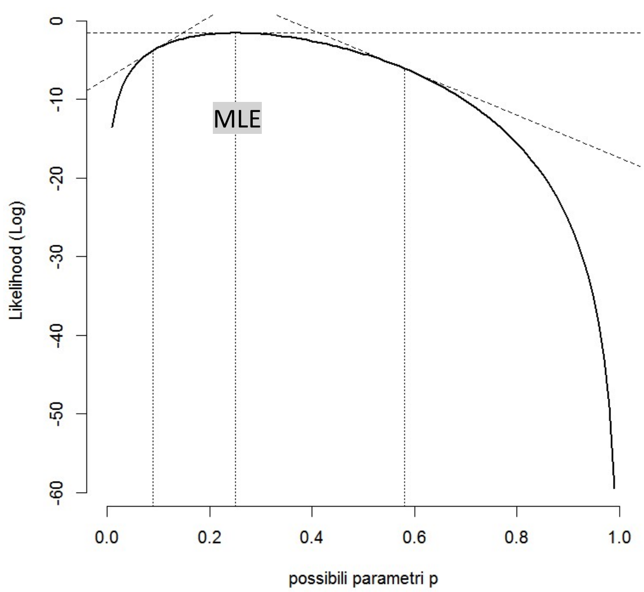
```
> D(expression(k*log(p)+(n-k)*log(1-p)),'p')
```

```
> k * (1/p) - (n - k) * (1/(1 - p))
```

Scrivendo quindi:

$$\frac{\delta}{\delta p}(\ln L(p|n,k)) = k \cdot \frac{1}{p} - (n-k) \cdot \frac{1}{(1-p)}$$

La derivata prima è una nuova funzione e corrisponde all'inclinazione della retta tangente alla funzione di log-verosmiglianza. La soluzione in p della retta tangente con inclinazione uguale a zero definisce il massimo della funzione:



Poniamo la derivata a zero e risolviamo per p:

$$k \cdot \frac{1}{p} - (n-k) \cdot \frac{1}{(1-p)} = 0$$

$$k = (n-k) \cdot \frac{p}{(1-p)}$$

$$k \cdot (1-p) = p \cdot (n-k)$$

$$p = \frac{k}{n}$$

ottenendo il valore k/n = 5/20 = 0.25 come stima di massima verosimiglianza della probabilità di “Si” nella popolazione. La **stima di massima verosimiglianza di una probabilità di successo** nella popolazione si ottiene mediante il rapporto tra i k successi e le n prove indipendenti:

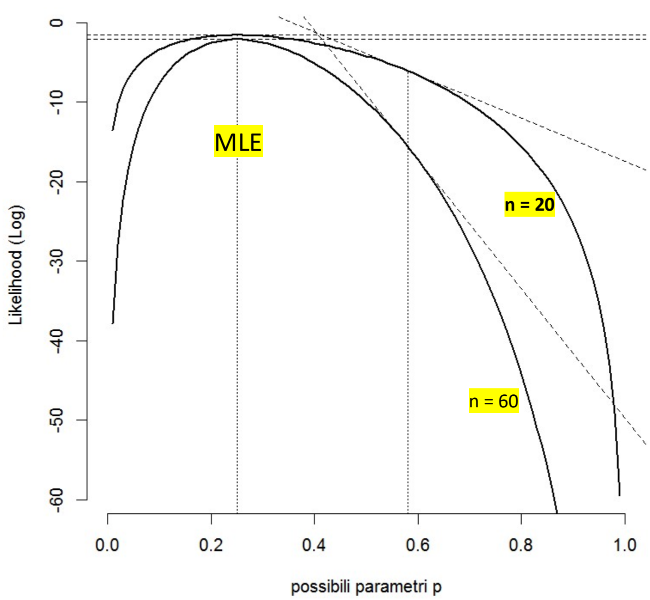
$$\boldsymbol{\hat{p} = \frac{k}{n}}$$

Se intervistassimo altre persone la stima p=k/n sarebbe certamente diversa, per l’effetto della variabilità campionaria, ma immaginiamo di ottenere proprio lo stesso valore di prima, con sequenza

$$X\begin{Bmatrix}1,0,0,1,0,1,1,1,1,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,1,1,1,1,\\0,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0\end{Bmatrix},$$

e quindi k = 15 successi in n = 60 prove, ovvero la stessa stima di massima verosimiglianza: MLE = k/n= 0.25. Di quale delle due stime ci fideremo di più? Intuitivamente, ci fideremo di quella proporzione basata su più osservazioni, perché più “fondata”.

Calcoliamo nuovamente i possibili valori delle funzioni di verosimiglianza per i dati dei due esperimenti, $\ln L(p|n=60,k=15) = 15\ln(p) + (60-15)\ln(1-p)$ e $\ln L(p|n=20,k=5) = 5\ln(p) + (20-5)\ln(1-p)$,e confrontiamoli mediante il diagramma precedente.



In entrambi gli esperimenti (n=60, n=20) la stima di massima verosimiglianza coincide con p = k/n = 0.25 Concettualmente, **la varianza campionaria della stima è legata alla curvatura della funzione di verosimiglianza**, nel senso che all’aumentare di n aumenterà la curvatura attorno alla stima di massima verosimiglianza, rendendo repentinamente meno verosimili valori superiori e inferiori. La precisione della stima MLE è maggiore nel campione di maggiori dimensioni.

In particolare, è possibile indagare la curvatura della funzione di verosimiglianza calcolando la derivata seconda della funzione di verosimiglianza, ossia la derivata prima della derivata prima in 'p', calcolata nel punto di MLE, e quindi sostituendo k=pn, dal momento che p=k/n.

```
> D(expression(k * (1/p) - (n - k) * (1/(1 - p))),'p')

-(k * (1/p^2) + (n - k) * (1/(1 - p)^2))
```

$$\frac{\delta}{\delta p}\left(k\frac{1}{p}-(n-k)\frac{1}{(1-p)}\right)=-k\frac{1}{p^2}-(n-k)\frac{1}{(1-p)^2}$$

sostituiamo la stima di massima verosmiglianza k=pn

$$=-\frac{pn}{p^2}-(n-pn)\frac{1}{(1-p)^2}$$

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

$$=-\left[\frac{n-pn+pn}{p(1-p)}\right]=-\frac{n}{p(1-p)}$$

Fisher ha dimostrato che il negativo dell'inversa della derivata seconda parziale della funzione di verosimiglianza, valutata alla MLE, rappresenta la **MLE della varianza del parametro ignoto p**

$$\boldsymbol{Var(\hat{p})=-\left[-\frac{n}{p(1-p)}\right]^{-1}=\frac{p(1-p)}{n}}.$$

Nei due esperimenti considerati, essendo diverso il denominatore, avremo che la varianza della stima di MLE = .25 sarà inferiore nel campione di n = 60 soggetti (0.003125) rispetto a quello di n = 20 (0.009375) soggetti.

La stima MLE p = k/n tende ad assumere una distribuzione *Normale* al crescere di n. Questa è una proprietà asintotica che si realizza anche in campioni finiti.

Trattandosi di una sommatoria di esiti di successo rapportata alla grandezza del campione, questo risultato è diretta conseguenza del Teorema del Limite Centrale per la media campionaria.

La quantità

$$z=\frac{\hat{p}-p_{H_0}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

verrà considerata come una variabile normale standardizzata, sotto H0.