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

$$P(k=4|n=10,p=0.5)=\frac{n!}{k!\,(n-k)!}\cdot p^k\cdot(1-p)^{(n-k)}$$

$$=\frac{10\times 9\times 8\times 7\times 6\times 5\times 4\times 3\times 2}{4\times 3\times 2\times 6\times 5\times 4\times 3\times 2}\cdot 0.5^{10}=0.2051$$

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

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

*($P(A\cap B)=P(A)P(B)$), nel caso di A e B eventi (non disgiunti) indipendenti.

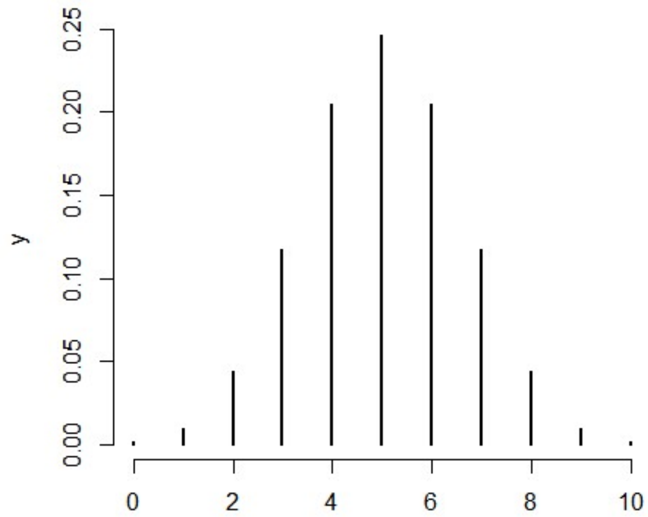


Figura 1 *Distribuzione binomiale per n = 10, p = 0.5. Si noti la simmetria della distribuzione per valori attorno al valore di k = 5.*

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:

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

*($P(A\cap B)=P(A)P(B)$), nel caso di A e B eventi (non disgiunti) indipendenti.

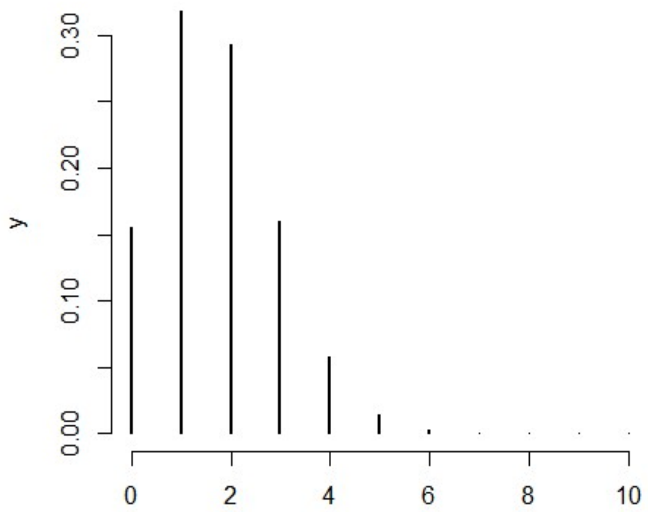


Figura 2 *Distribuzione di probabilità discreta binomiale per n = 10, p = 0.17. Si noti l'asimmetria della distribuzione.*

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

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

*$(P(A \cap B) = P(A)P(B))$, nel caso di A e B eventi (non disgiunti) indipendenti.

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

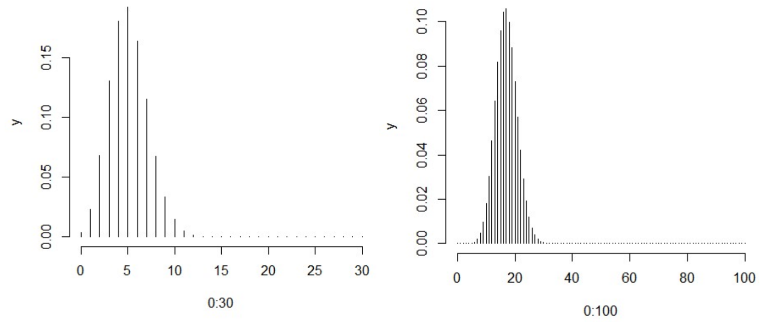


Figura 3 *N =30 e N = 100. La distribuzione tende alla normalità.*

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

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

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:

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

Consideriamo il caso particolare di n = 1, quindi una singola prova di Bernoulli: dalla definizione di valore atteso e varianza di una variabile aleatoria discreta,

$$E(Y) = \sum_{i=1}^{k} (y_i \times prob_i)$$
$$Var(Y) = \sum_{i=1}^{k} \big(y_i - E(Y)\big)^2 \times prob_i$$

abbiamo che:

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

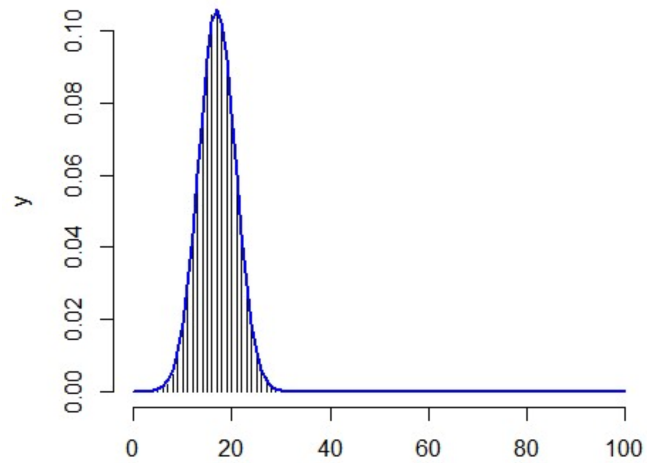
$$Var\big(Bern_{(p)}\big) = \sum_{i=1}^{2} \big[Bern_i - E\big(Bern_{(p)}\big)\big]^2 \times prob_i =$$
$$= (1-p)^2 p + (-p)^2 (1-p)$$
$$= p - 2p^2 + p^3 + p^2 - p^3$$
$$= p - p^2 = p(1-p)$$

*($P(A \cap B) = P(A)P(B)$), nel caso di A e B eventi (non disgiunti) indipendenti.

Se consideriamo n prove indipendenti di Bernoulli: avremo quindi

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

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 $\sigma = \sqrt[2]{np(1-p)}$.



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

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

dove ciascun elemento è una variabile casuale di Bernoulli con funzione di probabilità

$$f(y_i; p) = p^{y_i}(1-p)^{(1-y_i)},$$

Così, se

$$y_i = 1;\ p^1(1-p)^{(1-1)} = p;$$

$$y_i = 0;\ p^0(1-p)^{(1-0)} = 1-p.$$

La MLE di *p* si ottiene trovando il *massimo* della funzione:

$$l(p; y_i) = \ln\left[\prod_{i=1}^{n} f(y_i; p)\right] =$$

$$= \ln\left[\prod_{i=1}^{n} p^{y_i}(1-p)^{(1-y_i)}\right]$$

$$= \ln\left[p^{\sum y_i}(1-p)^{\sum(1-y_i)}\right] = \ln\left[p^{k}(1-p)^{(n-k)}\right]$$

$$= k\ln(p) + (n-k)\ln(1-p) \qquad \boldsymbol{Eq(9)}$$

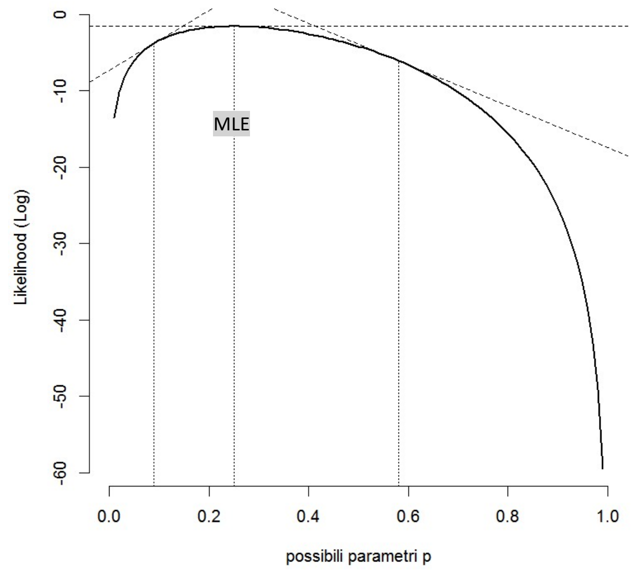
A questo punto si calcola la derivata prima in ‘*p*’ della funzione di log-verosimiglianza, ponendola uguale a zero e risolvendo. Possiamo utilizzare a tal fine il calcolatore,

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

Scrivendo quindi:

$$\frac{\delta}{\delta p} l(p; y_i) = k \cdot \frac{1}{p} - \underbrace{(n-k) \cdot \frac{1}{(1-p)}}_{\frac{d}{dy}\ln f(y) = \frac{1}{f(y)} \times \frac{df(y)}{dy}}$$

La derivata prima è una nuova funzione e corrisponde all’inclinazione della retta tangente alla funzione di log-verosimiglianza. 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*:

$$k \cdot \frac{1}{p} - (n-k) \cdot \frac{1}{(1-p)} = 0 \qquad \boldsymbol{Eq}(\mathbf{10})$$

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

$$k(1-p) = p(n-k) \Leftrightarrow p = \frac{k}{n} \qquad \boldsymbol{Eq}(\mathbf{11})$$

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:

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

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

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

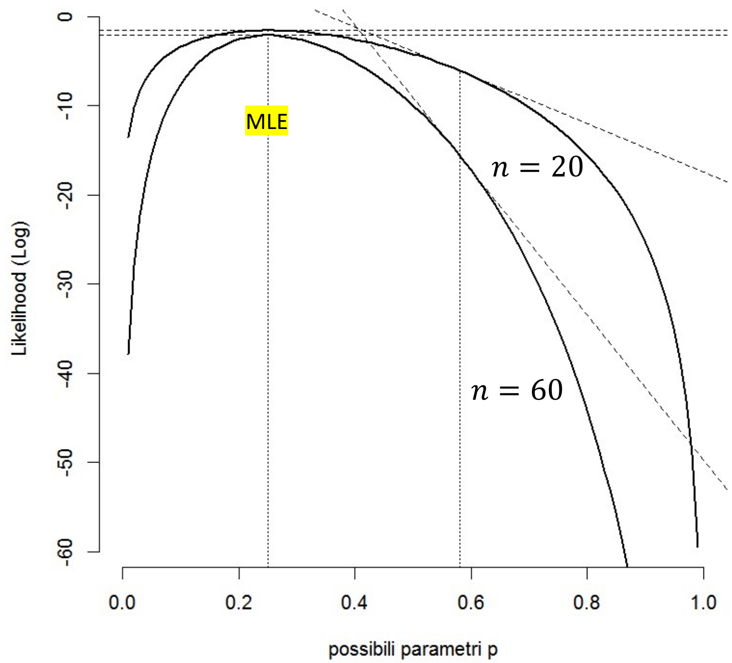
e quindi k = 15 successi in n = 60 prove, ovvero la stessa stima di massima verosimiglianza: **MLE** = *k/n*= 0.25. Di quale delle due stime ci fideremo di più? Intuitivamente, ci fideremo di quella proporzione basata su più osservazioni, perché più "fondata".

Calcoliamo nuovamente i possibili valori delle funzioni di verosimiglianza (***Eq(9)***) per i dati dei due esperimenti,

$$l(p; k=5, n=20) = 5\ln(p) + (20-5)\ln(1-p)$$

$$l(p; k=15, n=60) = 15\ln(p) + (60-15)\ln(1-p)$$

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 (**Eq(10)**) 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 verosimiglianza k=pn $\boldsymbol{Eq(12)}$

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

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

$$=-\left[\frac{n-pn+pn}{p(1-p)}\right]=-\frac{n}{p(1-p)}<0 \qquad \boldsymbol{Eq(13)}$$

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 soggetti (0.009375).
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

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

verrà considerata come una variabile normale standardizzata, sotto H0.

```
####### R

P_est<-seq(0,1,by=0.01)

lnL<-log(factorial(20))-log(factorial(5))-
log(factorial(15))+5*log(P_est)+15*log(1- P_est)

lnL60<-log(factorial(60))-log(factorial(15))-log(factorial(60-
15))+15*log(P_est)+(60-15)*log(1- P_est)

-------------------------------------------------------------

plot(y=exp(lnL),x=P_est)

plot(y=lnL,x=P_est)

-------------------------------------------------------------

DERIVATA<-function(x,k,n){ k * (1/x) - (n - k) * (1/(1 - x))}

b20 <-DERIVATA(x= P_est,5,20)

b60 <-DERIVATA(x= P_est,15,60)

a20<-lnL-b20*P_est

a60<-lnL60-b60*P_est

plot(x=P_est,y=lnL,type="l",bty="n",ylab="Likelihood
(Log)",xlab="possibili parametri p",lwd=2)

abline(a=a20[10],b=b20[10],lty="dashed")

segments(x0=P_est[10],y0=-70,x1=P_est[10],y1=lnL[10],lty="dotted")

abline(a=a20[26],b=b20[26],lty="dashed")

segments(x0= P_est[26],y0=-70,x1= P_est[26],y1=lnL[26],lty="dotted")

abline(a=a20[59],b=b20[59],lty="dashed")

segments(x0= P_est[59],y0=-70,x1= P_est[59],y1=lnL[59],lty="dotted")

 -------------------------------------------------------------

plot(x=P_est,y=lnL,type="l",bty="n",ylab="Likelihood
(Log)",xlab="possibili parametri p",lwd=2)

points(x=P_est,lnL60,type="l",lwd=2)

segments(x0= P_est[26],y0=-70,x1= P_est[26],y1=lnL[26],lty="dotted")

abline(a=a20[26],b=b20[26],lty="dashed")

abline(a=a60[26],b=b60[26],lty="dashed")

abline(a=a20[59],b=b20[59],lty="dashed")

abline(a=a60[59],b=b60[59],lty="dashed")

segments(x0= P_est[59],y0=-70,x1= P_est[59],y1=lnL[59],lty="dotted")
```