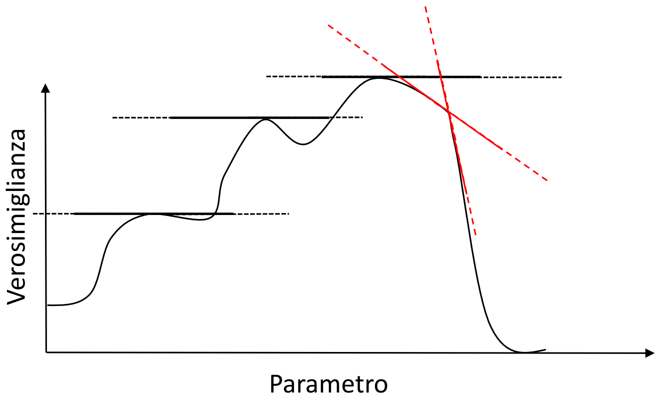
Newton-Raphson Method: metodo iterativo per trovare il massimo di una funzione. Inizia con una stima solo approssimativa della possibile soluzione nei parametri ignoti. L'aggiustamento successivo si basa sulle derivate di primo ordine (vettore gradiente - *u*) e sulle derivate di secondo ordine, organizzate in una matrice quadrata simmetrica (Hessiana - *H*).

Il metodo genera stime in successione fino ad un criterio numerico di convergenza, secondo il quale stime successive (MLE) non apportano un sostanziale miglioramento alla massimizzazione della funzione.

Il valore negativo delle derivate seconde, calcolate alle stime MLE individuate, assicura l'individuazione di un massimo e non di un massimo solo «locale».



Illustreremo il procedimento risolvendo due problemi di massimizzazione per i quali conosciamo già la soluzione formale:

1. Massimo della log-verosimiglianza basata sul conteggio di successi in *n* prove indipendenti di Bernoulli, distribuito secondo una *binomiale*. [***Eq(9)***]
2. Massimo della log-verosimiglianza basata su *n* osservazioni indipendenti distribuite secondo una *normale*. [***Eq.(1)***]

1. Log-verosimiglianza [**Eq(9)**]

$$l(p; y_i) = k\ln(p) + (n-k)\ln(1-p)$$

Derivata prima [**Eq(10)**]

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

Derivata seconda [**Eq(12)**]

$$H = -[k/p^2 + (n-k)/(1-p)^2]$$

In rosso viene indicato il parametro da stimare, la cui prima (*t*-esima, secondo la nomenclatura usata in Agresti) «approssimazione» potrebbe essere $p^{(1)} = .10$, ma qualsiasi altro valore plausibile va bene come stima iniziale (purché naturalmente compreso nei limiti [0,1]).

Il valore della Log-verosimiglianza sostituendo $p^{(t)}$:

$$l\left(p^{(t)}; y_i\right) = k\ln\left(p^{(t)}\right) + (n-k)\ln(1-p^{(t)})$$

verrà confrontato con quello della stima successiva $p^{(t+1)}$,

$$l\left(p^{(t+1)}; y_i\right) = k\ln\left(p^{(t+1)}\right) + (n-k)\ln(1-p^{(t+1)})$$

ottenuto mediante la seguente funzione generale («Newton-Raphson step»),

$$p^{(t+1)} = p^{(t)} - \left(H^{(t)}\right)^{-1} u^{(t)} \qquad \boldsymbol{Eq(14)}$$

$$= p^{(t)} + \left[\frac{k}{(p^{(t)})^2} + \frac{n-k}{(1-p^{(t)})^2}\right]^{-1} \frac{k-np^{(t)}}{p^{(t)}(1-p^{(t)})}. \qquad \boldsymbol{Eq(15)}$$

Applicazione pratica. Ad esempio, osservando k = 15 successi in n = 60 prove di Bernoulli, avremo che la stima MLE è data da $p = \frac{k}{n} = 0.25$. Iniziando dalla stima $p^{(1)} = .10$, l'algoritmo converge in poche iterazioni.

	Iterazione	1	2	3	4	5	6	7
	k	15	15	15	15	15	15	15
	n	60	60	60	60	60	60	60
	p(t)	0,100	0,164	0,225	0,248	0,250	0,250	0,250
	ln[p;(t)]	-39,280	-35,168	-33,848	-33,741	-33,740	-33,740	-33,740
Newton-Raphson	MLE	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*
	u	100,000	37,458	8,720	0,603	0,003	0,000	0,000
	H	-1555,556	-620,197	-371,994	-323,244	-320,016	-320,000	-320,000
Eq. 15	t+1	0,164	0,225	0,248	0,250	0,250	0,250	0,250
	ln[p;(t+1)]	-35,168	-33,848	-33,741	-33,740	-33,740	-33,740	-33,740
	Criterio	-4,112	-1,321	-0,107	-0,001	0,000	0,000	0,000

Il valore della Log-verosimiglianza dopo la settima iterazione è rimasto invariato raggiungendo il valore MLE $p^{(7)} = \frac{15}{60} = .25$. La varianza della stima MLE si ottiene mediante l'equazione:

$$Var(\hat{p}) = -\left(H^{(t)}\right)^{-1} = -\left(-\frac{1}{320.000}\right) = .003125$$

che coincide naturalmente con il risultato:

$$Var(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{.25\ (1-.25)}{60} = .003125.$$

Fisher scoring. Se nell'equazione ***Eq(15)*** sostituiamo il valore atteso di H, dal risultato noto della binomiale $E(k) = np$ abbiamo che:

$$E(H) = -\left[E(k)\frac{1}{p^2} + E(n-k)\frac{1}{(1-p)^2}\right]$$

$$= -\left[np\frac{1}{p^2} + (n-np)\frac{1}{(1-p)^2}\right] = -\left[\frac{n}{p(1-p)}\right]. \qquad \boldsymbol{Eq(16)}$$

Analogamente a quanto visto in precedenza, la stima dell'algoritmo di «Fisher scoring» procede secondo la formula:

$$p^{(t+1)} = p^{(t)} - \left(E\left(H^{(t)}\right)\right)^{-1} u^{(t)} = \qquad \boldsymbol{Eq(17)}$$

$$= p^{(t)} + \left[\frac{n}{p^{(t)}(1-p^{(t)})}\right]^{-1} \frac{k-np^{(t)}}{p^{(t)}(1-p^{(t)})}$$

$$= p^{(t)} + \frac{k-np^{(t)}}{n} = \frac{np^{(t)} + k - np^{(t)}}{n} = \frac{k}{n}.$$

Risultato in una singola iterazione, stabile in tutte quelle successive!!!

	Iterazione	1	2	3	4	5	6	7
	k	15	15	15	15	15	15	15
	n	60	60	60	60	60	60	60
	p(t)	0,100	0,250	0,250	0,250	0,250	0,250	0,250
	ln[p;(t)]	-39,280	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740
Fisher scoring	MLE	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*
	u	100,000	0,000	0,000	0,000	0,000	0,000	0,000
Eq. 16	H	-666,667	-320,000	-320,000	-320,000	-320,000	-320,000	-320,000
Eq. 17	t+1	0,250	0,250	0,250	0,250	0,250	0,250	0,250
	ln[p;(t+1)]	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740
	Criterio	-5,540	0,000	0,000	0,000	0,000	0,000	0,000

	Iterazione	1	2	3	4	5	6	7
	k	15	15	15	15	15	15	15
	n	60	60	60	60	60	60	60
	p(t)	0,999	0,250	0,250	0,250	0,250	0,250	0,250
	ln[p;(t)]	-310,864	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740
Fisher scoring	MLE	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*
	u	-44984,985	0,000	0,000	0,000	0,000	0,000	0,000
Eq. 16	H	-60060,060	-320,000	-320,000	-320,000	-320,000	-320,000	-320,000
Eq. 17	t+1	0,250	0,250	0,250	0,250	0,250	0,250	0,250
	ln[p;(t+1)]	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740
	Criterio	-277,124	0,000	0,000	0,000	0,000	0,000	0,000

	Iterazione	1	2	3	4	5	6	7
	k	15	15	15	15	15	15	15
	n	60	60	60	60	60	60	60
	p(t)	0,001	0,250	0,250	0,250	0,250	0,250	0,250
	ln[p;(t)]	-103,661	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740
Fisher scoring	MLE	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*	*0,250*
	u	14954,955	0,000	0,000	0,000	0,000	0,000	0,000
Eq. 16	H	-60060,060	-320,000	-320,000	-320,000	-320,000	-320,000	-320,000
Eq. 17	t+1	0,250	0,250	0,250	0,250	0,250	0,250	0,250
	ln[p;(t+1)]	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740	-33,740
	Criterio	-69,921	0,000	0,000	0,000	0,000	0,000	0,000

Risultato in una singola iterazione, stabile in tutte quelle successive!!!

2. Log-verosimiglianza **[Eq(1)]**

$$l(\mu, \sigma^2; y_i) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

Vettore delle derivate prime (*gradiente*) **[Eq(4, 7)]**

$$\boldsymbol{u} = \begin{bmatrix} \dfrac{\partial l(\mu, \sigma^2; y_i)}{\partial \mu} \\ \dfrac{\partial l(\mu, \sigma^2; y_i)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2}\left(\sum_{i=1}^{n} y_i - n\mu\right) \\ \dfrac{1}{2\sigma^2}\left[-n + \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right] \end{bmatrix}$$

Matrice (Hessiana) delle derivate seconde **[Eq(8.1-8.4)]**

$$\boldsymbol{H} = \begin{bmatrix} -\dfrac{n}{\sigma^2} & -\dfrac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n} y_i - n\mu\right) \\ -\dfrac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n} y_i - n\mu\right) & \dfrac{n}{2(\sigma^2)^2} - \dfrac{2\sigma^2}{2(\sigma^2)^4}\sum_{i=1}^{n}(y_i - \mu)^2 \end{bmatrix}$$

Vettore dei parametri ignoti (2x1):

$$\underset{(2\times1)}{\boldsymbol{\varphi}} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$$

Il valore della Log-verosimiglianza sostituendo $\boldsymbol{\varphi}^{(t)}$:

$$l\left(\boldsymbol{\varphi}^{(t)}; y_i\right) =$$
$$-\frac{n}{2}ln(2\pi) - \frac{n}{2}ln\left(\sigma^{2(t)}\right) - \frac{1}{2\sigma^{2(t)}}\sum_{i=1}^{n}\left(y_i - \mu^{(t)}\right)^2$$

verrà confrontato con quello della stima successiva $\boldsymbol{\varphi}^{(t+1)}$,

$$l\left(\boldsymbol{\varphi}^{(t+1)}; y_i\right) =$$
$$-\frac{n}{2}ln(2\pi) - \frac{n}{2}ln\left(\sigma^{2(t+1)}\right) - \frac{1}{2\sigma^{2(t+1)}}\sum_{i=1}^{n}\left(y_i - \mu^{(t+1)}\right)^2$$

ottenuto mediante la funzione generale («Newton-Raphson step»),

$$\underbrace{\boldsymbol{\varphi}^{(t+1)}}_{2\text{x}1} = \underbrace{\boldsymbol{\varphi}^{(t)}}_{2\text{x}1} - \underbrace{\left(\boldsymbol{H}^{(t)}\right)^{-1}\boldsymbol{u}^{(t)}}_{2\text{x}1}$$

Nota bene. Il calcolo richiede l'inversione di una matrice 2X2 ed il prodotto interno con un vettore 2X1 (vedi tutorial R Lezione 7).

Per una più rapida progressione è possibile utilizzare il valore atteso della matrice Hessiana:

$$E[\boldsymbol{H}] = E\begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n} y_i - n\mu\right) \\ -\frac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n} y_i - n\mu\right) & \frac{n}{2(\sigma^2)^2} - \frac{2\sigma^2}{2(\sigma^2)^4}\sum_{i=1}^{n}(y_i - \mu)^2 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{(\sigma^2)^2}E\left(\sum_{i=1}^{n} y_i - n\mu\right) \\ -\frac{1}{(\sigma^2)^2}E\left(\sum_{i=1}^{n} y_i - n\mu\right) & \frac{n}{2(\sigma^2)^2} - \frac{2\sigma^2}{2(\sigma^2)^4}E\left[\sum_{i=1}^{n}(y_i - \mu)^2\right] \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n} E(y_i) - n\mu\right) \\ -\frac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n} E(y_i) - n\mu\right) & \frac{n}{2(\sigma^2)^2} - \frac{2\sigma^2}{2(\sigma^2)^4}\sum_{i=1}^{n}E(y_i - \mu)^2 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{(\sigma^2)^2}(n\mu - n\mu) \\ -\frac{1}{(\sigma^2)^2}(n\mu - n\mu) & \frac{n}{2(\sigma^2)^2} - \frac{2\sigma^2}{2(\sigma^2)^4}n\sigma^2 \end{bmatrix}$$

$$E[\boldsymbol{H}] = \begin{bmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & -\frac{n}{2(\sigma^2)^2} \end{bmatrix}$$

La stima successiva $\boldsymbol{\varphi}^{(t+1)}$ verrà calcolata mediante la formula generale («Fisher scoring»):

$$\boldsymbol{\varphi}^{(t+1)} = \boldsymbol{\varphi}^{(t)} - \left(E\left(\boldsymbol{H}^{(t)}\right)\right)^{-1}\boldsymbol{u}^{(t)}$$

dati	Iterazione	1		2		3		4		5		6		7	
2	$\mu(t)$	*5,000*		3,965		4,080		4,475		4,325		4,362		4,357	
3	$\sigma^2(t)$	*2,000*		3,221		0,949		1,351		1,723		1,981		2,077	
3	$\ln[\mu,\sigma^2;(t)]$	-26,467170													
3	μ	4,357		4,357		4,357		4,357		4,357		4,357		4,357	
4	σ^2	2,087		2,087		2,087		2,087		2,087		2,087		2,087	
5		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*	
6		*9,000*		*3,860*		*4,328*		*6,123*		*5,405*		*5,579*		*5,555*	
7		*4,000*		*0,931*		*1,167*		*2,174*		*1,755*		*1,855*		*1,841*	
5		*4,000*		*0,931*		*1,167*		*2,174*		*1,755*		*1,855*		*1,841*	
6		*4,000*		*0,931*		*1,167*		*2,174*		*1,755*		*1,855*		*1,841*	
4		*1,000*		*0,001*		*0,006*		*0,225*		*0,106*		*0,131*		*0,127*	
4		*0,000*		*1,072*		*0,846*		*0,276*		*0,456*		*0,407*		*0,414*	
3		*1,000*		*4,142*		*3,685*		*2,327*		*2,806*		*2,683*		*2,700*	
6		*4,000*		*9,213*		*8,524*		*6,378*		*7,157*		*6,959*		*6,986*	
		0,000		*1,072*		*0,846*		*0,276*		*0,456*		*0,407*		*0,414*	
«Newton-		*1,000*		*4,142*		*3,685*		*2,327*		*2,806*		*2,683*		*2,700*	
Raphson»		*1,000*		*0,001*		*0,006*		*0,225*		*0,106*		*0,131*		*0,127*	
		1,000		*0,001*		*0,006*		*0,225*		*0,106*		*0,131*		*0,127*	
		4,000		*0,931*		*1,167*		*2,174*		*1,755*		*1,855*		*1,841*	
		1,000		*4,142*		*3,685*		*2,327*		*2,806*		*2,683*		*2,700*	
	u	-4,500		1,706		4,083		-1,216		0,262		-0,034		0,002	
		0,875		-0,661		9,442		2,872		0,859		0,188		0,016	
	H	-7,000	-2,250	-4,347	0,530	-14,755	4,304	-10,360	-0,900	-8,124	0,152	-7,067	-0,017	-6,742	0,001
		-2,250	-2,625	0,530	-0,264	4,304	-27,677	-0,900	-8,084	0,152	-3,355	-0,017	-1,974	0,001	-1,639
	H-Inversa	-0,197	0,169	-0,304	-0,610	-0,071	-0,011	-0,097	0,011	-0,123	-0,006	-0,142	0,001	-0,148	0,000
		0,169	-0,526	-0,610	-5,009	-0,011	-0,038	0,011	-0,125	-0,006	-0,298	0,001	-0,507	0,000	-0,610
	$\mu(t+1)$	3,965		4,080		4,475		4,325		4,362		4,357		4,357	
	$\sigma^2(t+1)$	3,221		0,949		1,351		1,723		1,981		2,077		2,087	
		-26,46717	$\ln[\mu,\sigma^2;(t+1)]$	-25,92228		-28,45716		-25,85367		-25,15539		-25,02404		-25,0144	
	Criterio			0,545		-2,535		2,603		0,698		0,131		0,010	

dati	Iterazione	1		2		3		4		5		6		7	
2	$\mu(t)$	*5,000*		4,357		4,357		4,357		4,357		4,357		4,357	
3	$\sigma^2(t)$	*2,000*		2,500		2,087		2,087		2,087		2,087		2,087	
3	$\ln[\mu,\sigma^2;(t)]$	-26,46717													
3	μ	4,357		4,357		4,357		4,357		4,357		4,357		4,357	
4	σ^2	2,087		2,087		2,087		2,087		2,087		2,087		2,087	
5		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*		*Scarti ^2*	
6		*9,000*		*5,556*		*5,556*		*5,556*		*5,556*		*5,556*		*5,556*	
7		*4,000*		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*	
5		*4,000*		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*	
6		*4,000*		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*	
4		*1,000*		*0,128*		*0,128*		*0,128*		*0,128*		*0,128*		*0,128*	
4		*0,000*		*0,413*		*0,413*		*0,413*		*0,413*		*0,413*		*0,413*	
3		*1,000*		*2,699*		*2,699*		*2,699*		*2,699*		*2,699*		*2,699*	
6		*4,000*		*6,985*		*6,985*		*6,985*		*6,985*		*6,985*		*6,985*	
		0,000		*0,413*		*0,413*		*0,413*		*0,413*		*0,413*		*0,413*	
«Fisher		*1,000*		*2,699*		*2,699*		*2,699*		*2,699*		*2,699*		*2,699*	
		1,000		*0,128*		*0,128*		*0,128*		*0,128*		*0,128*		*0,128*	
scoring»		*1,000*		*0,128*		*0,128*		*0,128*		*0,128*		*0,128*		*0,128*	
		4,000		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*		*1,842*	
		1,000		*2,699*		*2,699*		*2,699*		*2,699*		*2,699*		*2,699*	
	u	-4,500		0,000		0,000		0,000		0,000		0,000		0,000	
		0,875		-0,463		0,000		0,000		0,000		0,000		0,000	
	H	-7,000	0,000	-5,600	0,000	-6,709	0,000	-6,709	0,000	-6,709	0,000	-6,709	0,000	-6,709	0,000
		0,000	-1,750	0,000	-1,120	0,000	-1,608	0,000	-1,608	0,000	-1,608	0,000	-1,608	0,000	-1,608
	H-Inversa	-0,143	0,000	-0,179	0,000	-0,149	0,000	-0,149	0,000	-0,149	0,000	-0,149	0,000	-0,149	0,000
		0,000	-0,571	0,000	-0,893	0,000	-0,622	0,000	-0,622	0,000	-0,622	0,000	-0,622	0,000	-0,622
	$\mu(t+1)$	4,357		4,357		4,357		4,357		4,357		4,357		4,357	
	$\sigma^2(t+1)$	2,500		2,087		2,087		2,087		2,087		2,087		2,087	
		-26,46717	$\ln[\mu,\sigma^2;(t+1)]$	-25,12203		-25,01434		-25,01434		-25,01434		-25,01434		-25,0143	
	Criterio			1,345		0,108		0,000		0,000		0,000		0,000	