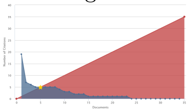
Qualche risultato asintotico relativo all'indice di Hirsch

Luca Pratelli,[1]
Alberto Baccini,[2] Lucio Barabesi,[2] Marzia Marcheselli[2]

[1]Gruppo Matematiche, Accademia Navale, Livorno
[2]Dipartimento di Economia Politica e Statistica, Università di Siena

8 Giugno 2015



Gli argomenti e i risultati di questa presentazione sono tutti contenuti in

- Pratelli, L., Baccini, A., Barabesi, L. and Marcheselli, M. (2012) *Statistical analysis of the Hirsch index*, Scandinavian Journal of Statistics 39, 681-694
- Baccini, A., Barabesi, L., Marcheselli, M. and Pratelli, L. (2012) *Statistical inference on the h-index with an application to top-scientist performance*, Journal of Infometrics 6, 721-728

Introduzione

Jorge Eduardo Hirsch, fisico argentino, il 3 Agosto 2005 pubblica su arXiv.org un articolo dove introduce quello che oggi è comunemente noto come h-index. Egli afferma che:

Introduzione

Jorge Eduardo Hirsch, fisico argentino, il 3 Agosto 2005 pubblica su arXiv.org un articolo dove introduce quello che oggi è comunemente noto come h-index. Egli afferma che:

A scientist has index h if h of his/her N_p papers have at least h citations each, and the others $N_p - h$ papers have no more than h citations each.

In nemmeno tre settimane appaiono due articoli; il primo su *Nature* di Bell e il secondo su *Science* di Bhattacharjee dove l'indice presentato da Hirsch viene definito come 'transparent, unbiased and very hard to rig' e viene applicato, per la prima volta, per ottenere una classifica di sette fisici di alto livello scientifico.

In nemmeno tre settimane appaiono due articoli; il primo su *Nature* di Bell e il secondo su *Science* di Bhattacharjee dove l'indice presentato da Hirsch viene definito come 'transparent, unbiased and very hard to rig' e viene applicato, per la prima volta, per ottenere una classifica di sette fisici di alto livello scientifico.

A fine settembre del 2005. l'articolo di Hirsch viene pubbblicato sui *Proceedings of the National Academy of Sciences.*

Questi fatti non possono essere considerati di secondaria importanza per spiegare il successo incontrato dall'indice di Hirsch.

L'immediato appoggio di due delle riviste più prestigiose e la pubblicazione quasi simultanea sui *Proceedings* sono stati un segnale forte e autorevole per l'intera comunità accademica alla ricerca di un modo semplice e utile per caratterizzare la produzione scientifica non solo di un singolo ricercatore ma anche di gruppi di ricerca, di dipartimenti, di istituzioni.

L'immediato appoggio di due delle riviste più prestigiose e la pubblicazione quasi simultanea sui *Proceedings* sono stati un segnale forte e autorevole per l'intera comunità accademica alla ricerca di un modo semplice e utile per caratterizzare la produzione scientifica non solo di un singolo ricercatore ma anche di gruppi di ricerca, di dipartimenti, di istituzioni.

Nonostante i non trascurabili svantaggi nell'adottare l'indice di Hirsch (forte dipendenza dall'età 'scientifica' del ricercatore e dalle auto-citazioni) il successo arrivò immediato e tre anni dopo il lavoro di Hirsch era già citato in ben 753 articoli ad esso collegati.

A differenza del successo nel campo della scientometria, fino al 2010 in ambito statistico non sono apparsi molti lavori incentrati a studiare le proprietà di un indice di Hirsch teorico e, conseguentemente, del suo stimatore empirico.

Soltanto Glanzel nel 2006 si muove in questa direzione deducendo alcune proprietà dell' h-index nel caso di un modello paretiano applicato alle citazioni.

A differenza del successo nel campo della scientometria, fino al 2010 in ambito statistico non sono apparsi molti lavori incentrati a studiare le proprietà di un indice di Hirsch teorico e, conseguentemente, del suo stimatore empirico.

Soltanto Glanzel nel 2006 si muove in questa direzione deducendo alcune proprietà dell' h-index nel caso di un modello paretiano applicato alle citazioni.

Seguendo una corretta prospettiva statistica, un notevole passo avanti viene fatto nel (2010) da Beirlant e Einmahl che vedono l'indice di Hirsch come uno stimatore empirico di un opportuno funzionale statistico dipendente dalla distribuzione del numero di citazioni.

Tuttavia, Beirlant e Einmahl assumono che la distribuzione del numero di citazioni sia assolutamente continua; ipotesi ovviamente irreale in quanto la distribuzione è concentrata sull'insieme dei numeri naturali.

Tuttavia, Beirlant e Einmahl assumono che la distribuzione del numero di citazioni sia assolutamente continua; ipotesi ovviamente irreale in quanto la distribuzione è concentrata sull'insieme dei numeri naturali.

Inoltre, il risultato asintotico e i corollari da loro ottenuti riguardanti l'indice di Hirsch h non sono applicabili direttamente alla stima di h in quanto fortemente legati alla conoscenza della distribuzione del numero di citazioni che, praticamente, non è mai nota e comunque difficile da modellizzare con i dati reali.

Infine, nel lavoro di Beirlant e Einmahl e più in generale nei lavori sull'indice di Hirsch, non campare uno stimatore dell'errore quadratico medio per campioni finiti.

Indice teorico di Hirsch

Sia X una variabile aleatoria a valori nell'insieme dei naturali. X rappresenta il numero di citazioni che riceve un articolo scritto da un assegnato ricercatore.

Indice teorico di Hirsch

Sia X una variabile aleatoria a valori nell'insieme dei naturali. X rappresenta il numero di citazioni che riceve un articolo scritto da un assegnato ricercatore.

- Seguendo motivazioni date in scientometria, $S(x) > 0, \forall x$

Indice teorico di Hirsch

Sia X una variabile aleatoria a valori nell'insieme dei naturali. X rappresenta il numero di citazioni che riceve un articolo scritto da un assegnato ricercatore.

- Seguendo motivazioni date in scientometria, $S(x) > 0, \forall x$
- Posto $S_{-}(x) = P(X \geq x)$ e detto n il numero di articoli pubblicati dal ricercatore, si definisce l'indice (teorico di Hirsch) h_n come

$$h_n = \sup \{x > 0 : nS_{-}(x) \geq x\}$$

Indice teorico di Hirsch

Sia X una variabile aleatoria a valori nell'insieme dei naturali. X rappresenta il numero di citazioni che riceve un articolo scritto da un assegnato ricercatore.

- Seguendo motivazioni date in scientometria, $S(x) > 0, \forall x$
- Posto $S_-(x) = P(X \geq x)$ e detto n il numero di articoli pubblicati dal ricercatore, si definisce l'indice (teorico di Hirsch) h_n come

$$h_n = \sup \{x > 0 : nS_-(x) \geq x\}$$

Si noti che h_n è un numero naturale minore o eguale a n che cresce all'aumentare del numero n di pubblicazioni.

Inoltre, $\lim_n h_n = \infty, \lim_n h_n/n = 0$. L'indice di Hirsch h_n si può riscrivere come

$$h_n = \max\{j \geq 0 : nS(j-1) \geq j\} = \sum_{j=1}^{n} I_{[j/n,1]}(S(j-1))$$

Inoltre, $\lim_n h_n = \infty, \lim_n h_n/n = 0$. L'indice di Hirsch h_n si può riscrivere come

$$h_n = \max\{j \geq 0 : nS(j-1) \geq j\} = \sum_{j=1}^{n} I_{[j/n,1]}(S(j-1))$$

La definizione introdotta da Beirlant e Einmahl si riduce a prendere X variabile aleatoria positiva e diffusa: in questo modo h_n risulta essere l'unico valore (reale) tale che $nS(h_n) = h_n$.

In altre parole, la parte intera dell'indice di Hirsch introdotto da Beirlant e Einmahl non è nient'altro che l'indice di Hirsch h_n relativo alla parte intera di X.

Indice empirico di Hirsch

Siano $X_1, \ldots, X_n$ indipendenti e con la stessa legge di X. Le variabili aleatorie X_i rappresentano il numero di citazioni ricevute dall'i-esimo articolo.

Indice empirico di Hirsch

Siano $X_1, \ldots, X_n$ indipendenti e con la stessa legge di X. Le variabili aleatorie X_i rappresentano il numero di citazioni ricevute dall'i-esimo articolo.

Si definisce indice (empirico di Hirsch) la variabile aleatoria

$$\widetilde{H}_n = \sum_{j=1}^{n} I_{[j/n,1]}(\widehat{S}_n(j-1))$$

dove

$$\widehat{S}_n(j-1) = \frac{1}{n}\sum_{k=1}^{n} I_{\{X_k > j-1\}} = \frac{1}{n}\sum_{k=1}^{n} I_{\{X_k \geq j\}}$$

In effetti: $\widetilde{H}_n = h$ se e solo se h articoli hanno almeno h citazioni ciascuno e gli altri $n-h$ articoli non hanno ognuno più di h citazioni.

In modo analogo al caso teorico, si può scrivere la relazione

$$\widetilde{H}_n = \sup\{x \geq 0 : n\widehat{S}_{n-}(x) \geq x\}$$

che rende naturale e 'giustifica' la definizione di h_n.

In modo analogo al caso teorico, si può scrivere la relazione

$$\widetilde{H}_n = \sup\{x \geq 0 : n\widehat{S}_{n-}(x) \geq x\}$$

che rende naturale e 'giustifica' la definizione di h_n.

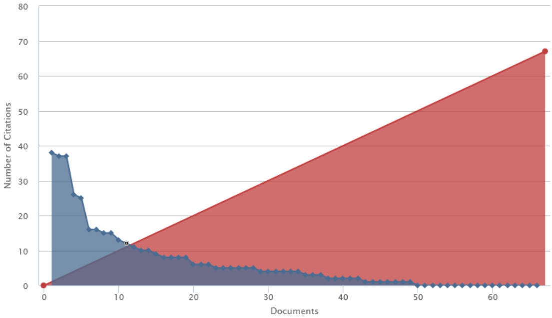


Figure: La visualizzazione su *Scopus* dell'indice di Hirsch di Barabesi

Nel seguito si pone

- $Y_{j,n} = I_{[j/n,1]}(\widehat{S}_n(j-1)), \quad \widetilde{H}_n = \sum_{j=1}^{n} Y_{j,n}$

Nel seguito si pone

- $Y_{j,n} = I_{[j/n,1]}(\widehat{S}_n(j-1)), \quad \widetilde{H}_n = \sum_{j=1}^{n} Y_{j,n}$
- $p_{j,n} = E[Y_{j,n}] = \sum_{m=j}^{n} \binom{n}{m} S(j-1)^m (1-S(j-1))^{n-m}$

Nel seguito si pone

- $Y_{j,n} = I_{[j/n,1]}(\widehat{S}_n(j-1)), \quad \widetilde{H}_n = \sum_{j=1}^{n} Y_{j,n}$
- $p_{j,n} = E[Y_{j,n}] = \sum_{m=j}^{n} \binom{n}{m} S(j-1)^m (1 - S(j-1))^{n-m}$
- $\mathrm{Var}[\widetilde{H}_n] = \sum_{j=1}^{n} r_{j,n}(1 - p_{j,n})$ dove

$$r_{j,n} = p_{j,n} + 2 \sum_{l=j+1}^{n} p_{l,n}.$$

Errore quadratico

Per determinare le proprietà statistiche dello stimatore $\widetilde{H}_n$ di h_n, occorre valutare il comportamento asintotico di $\widetilde{H}_n/h_n$. In generale, abbiamo provato che

- $\widetilde{H}_n/h_n$ converge in media quadratica verso 1.

Errore quadratico

Per determinare le proprietà statistiche dello stimatore $\widetilde{H}_n$ di h_n, occorre valutare il comportamento asintotico di $\widetilde{H}_n/h_n$. In generale, abbiamo provato che

- $\widetilde{H}_n/h_n$ converge in media quadratica verso 1.
- Se $\lim_n \mathrm{Var}[\widetilde{H}_n] = \infty$ allora

$$(E[\widetilde{H}_n] - h_n)^2 = o(\mathrm{Var}[\widetilde{H}_n])$$

Errore quadratico

Per determinare le proprietà statistiche dello stimatore $\widetilde{H}_n$ di h_n, occorre valutare il comportamento asintotico di $\widetilde{H}_n/h_n$. In generale, abbiamo provato che

- $\widetilde{H}_n/h_n$ converge in media quadratica verso 1.
- Se $\lim_n \mathrm{Var}[\widetilde{H}_n] = \infty$ allora

$$(E[\widetilde{H}_n] - h_n)^2 = o(\mathrm{Var}[\widetilde{H}_n])$$

In particolare, l'errore quadratico $E[\widetilde{H}_n - h_n)^2]$ può essere sostituito dalla varianza di $\widetilde{H}_n$ quand'essa tende a ∞.

Inoltre, si è mostrato che il comportamento della varianza di $\widetilde{H}_n$ viene individuato da:

- $\mathrm{Var}[\widetilde{H}_n] \geq \sum_{j=1}^{\lfloor 2h_n \rfloor} \tilde{r}_{j,n}(1-p_{j,n})$, dove

$$\tilde{r}_{j,n} = p_{j,n} + 2 \sum_{l=j+1}^{\lfloor 2h_n \rfloor} p_{l,n}$$

Inoltre, si è mostrato che il comportamento della varianza di $\widetilde{H}_n$ viene individuato da:

- $\mathrm{Var}[\widetilde{H}_n] \geq \sum_{j=1}^{\lfloor 2h_n \rfloor} \tilde{r}_{j,n}(1-p_{j,n})$, dove

$$\tilde{r}_{j,n} = p_{j,n} + 2\sum_{l=j+1}^{\lfloor 2h_n \rfloor} p_{l,n}$$

- Se $\lim_n \mathrm{Var}[\widetilde{H}_n] = \infty$ allora

$$\mathrm{Var}[\widetilde{H}_n] \sim \sum_{j=1}^{\lfloor 2h_n \rfloor} \tilde{r}_{j,n}(1-p_{j,n})$$

Occorre allora introdurre una condizione che assicuri la divergenza della varianza di $\widetilde{H}_n$. A questo scopo, si suppone che

$$\lim_n \frac{\sqrt{n}P(X=n)}{S(n)} = 0 \qquad (\star)$$

Occorre allora introdurre una condizione che assicuri la divergenza della varianza di $\widetilde{H}_n$. A questo scopo, si suppone che

$$\lim_n \frac{\sqrt{n}P(X=n)}{S(n)} = 0 \qquad (\star)$$

Sotto la condizione $(\star)$, si è provato che la varianza di $\widetilde{H}_n$ tende a ∞ e lo 'stimatore'

$$V_n = \sum_{j=1}^{\lfloor 3\widetilde{H}_n \rfloor \wedge n} p_{j,n}(1-p_{j,n}) + 2 \sum_{l=2}^{\lfloor 3\widetilde{H}_n \rfloor \wedge n} p_{l,n} \sum_{j=1}^{l-1} (1-p_{j,n})$$

è coerente ossia $V_n/\mathrm{Var}[\widetilde{H}_n]$ converge in probabilità verso 1.

Stima della varianza e teorema limite

Rafforzando leggermente la condizione ($\star$), si prova un teorema limite centrale e la consistenza di uno stimatore non parametrico di $\mathrm{Var}[\widetilde{H}_n]$. Più precisamente, sotto la condizione

Per ogni $M > 0$*, posto* $D_{M,n}$ *l'insieme dei numeri naturali contenuti in* $[n - M\sqrt{n}, n + M\sqrt{n}]$ *vale*

$$\lim_n \Big(\sup_{j \in D_{M,n}} \Big| \frac{P(X=j)}{P(X=n)} - 1 \Big| \Big) = 0 \qquad (\star\star)$$

si ha

Stima della varianza e teorema limite

Rafforzando leggermente la condizione $(\star)$, si prova un teorema limite centrale e la consistenza di uno stimatore non parametrico di $\mathrm{Var}[\widetilde{H}_n]$. Più precisamente, sotto la condizione

Per ogni $M > 0$, posto $D_{M,n}$ l'insieme dei numeri naturali contenuti in $[n - M\sqrt{n}, n + M\sqrt{n}]$ vale

$$\lim_n \Big(\sup_{j \in D_{M,n}} \Big| \frac{P(X=j)}{P(X=n)} - 1 \Big| \Big) = 0 \qquad (\star\star)$$

si ha

- (Equivalenza) $\qquad \mathrm{Var}[\widetilde{H}_n] \sim \frac{h_n}{(1+nP(X=h_n))^2}$

- (Coerenza) Posto

$$\widehat{V}_n = \sum_{j=1}^{\lfloor 3\widetilde{H}_n \rfloor \wedge n} \widehat{p}_{j,n}(1 - \widehat{p}_{j,n}) + 2 \sum_{l=2}^{\lfloor 3\widetilde{H}_n \rfloor \wedge n} \widehat{p}_{l,n} \sum_{j=1}^{l-1} (1 - \widehat{p}_{j,n})$$

dove

$$\widehat{p}_{j,n} = \sum_{m=j}^{n} \binom{n}{m} \widehat{S}(j-1)^m (1 - \widehat{S}(j-1))^{n-m}$$

il rapporto $\widehat{V}_n / \mathrm{Var}[\widetilde{H}_n]$ converge in probabilità a 1

- (Coerenza) Posto

$$\widehat{V}_n = \sum_{j=1}^{\lfloor 3\widetilde{H}_n \rfloor \wedge n} \widehat{p}_{j,n}(1-\widehat{p}_{j,n}) + 2 \sum_{l=2}^{\lfloor 3\widetilde{H}_n \rfloor \wedge n} \widehat{p}_{l,n} \sum_{j=1}^{l-1}(1-\widehat{p}_{j,n})$$

  dove

$$\widehat{p}_{j,n} = \sum_{m=j}^{n} \binom{n}{m} \widehat{S}(j-1)^m (1-\widehat{S}(j-1))^{n-m}$$

  il rapporto $\widehat{V}_n / \mathrm{Var}[\widetilde{H}_n]$ converge in probabilità a 1

- (Normalità) Le variabili aleatorie

$$\frac{\widetilde{H}_n - h_n}{\sqrt{\widehat{V}_n}}$$

  convergono in legge verso $\mathcal{N}(0,1)$.

Tecniche utilizzate

Il Lemma (opportuna applicazione di un risultato di Osipov) e l'osservazione che hanno permesso di ottenere i risultati fin qui presentati sono:

- Lemma. Esiste $A > 0$ tale che, per ogni n e per ogni j, risulta

$$|p_{j,n} - \Phi(x_{j,n})| \leq A \frac{v_{j,n}^2 + 1}{v_{j,n}^4 (1 + |x_{j,n}|)^6}$$

dove

$$x_{j,n} = (j - nS(j-1))/v_{j,n}, \quad v_{j,n}^2 = nS(j-1)(1 - S(j-1))$$

Tecniche utilizzate

Il Lemma (opportuna applicazione di un risultato di Osipov) e l'osservazione che hanno permesso di ottenere i risultati fin qui presentati sono:

- Lemma. Esiste $A > 0$ tale che, per ogni n e per ogni j, risulta

$$|p_{j,n} - \Phi(x_{j,n})| \leq A \frac{v_{j,n}^2 + 1}{v_{j,n}^4 (1 + |x_{j,n}|)^6}$$

dove

$$x_{j,n} = (j - nS(j-1))/v_{j,n}, \quad v_{j,n}^2 = nS(j-1)(1 - S(j-1))$$

- (Martingalità) $\widetilde{H}_n - h_n$ è la somma di martingale differenze e coincide con $\sum_{j=1}^n \left(\frac{Y_{j,n}}{p_{j,n}} - \frac{Y_{j-1,n}}{p_{j-1,n}} \right) (p_{j,n} + \cdots + p_{n,n})$

Confronto con normalità asintotica di B. et E.

Beirlant ed Einmahl nel loro lavoro del 2010 ottengono un risultato di normalità asintotica per $\widetilde{H}_n$ sotto l'ipotes che la densità f di $X(P)$ verifichi una condizione per molti versi equivalente a dire che $\lfloor X \rfloor$ soddisfi $(\star\star)$. In altre parole,

Confronto con normalità asintotica di B. et E.

Beirlant ed Einmahl nel loro lavoro del 2010 ottengono un risultato di normalità asintotica per $\widetilde{H}_n$ sotto l'ipotes che la densità f di $X(P)$ verifichi una condizione per molti versi equivalente a dire che $\lfloor X \rfloor$ soddisfi $(\star\star)$. In altre parole,

$$W_n = \frac{1 + nf(h_n)}{\sqrt{h_n}}(\widetilde{H}_n - h_n)$$

converge in legge verso $\mathcal{N}(0, 1)$.

Da un punto di vista teorico, grazie ai risultati precedentemente esposti, si ha che W_n ha lo stesso comportamento asintotico di

$$(\widetilde{H}_n - h_n)/\sqrt{\mathrm{Var}[\widetilde{H}_n]}$$

Tuttavia, da un punto di vista applicativo, il risultato di Beirl-nt ed Einmahl non è molto utilizzabile se non viene 'stimato' il valore di $\frac{1+nf(h_n)}{\sqrt{h_n}}$.

Tentando di ovviare a questo problema Beirlant ed Einmahl si limitano a considerare distribuzioni Paretiane e di tipo Weibull in quanto permettono di avere rispttivamente

$$\frac{1+nf(h_n)}{\sqrt{h_n}} = \frac{c_1}{\sqrt{h_n}}, \qquad \frac{1+nf(h_n)}{\sqrt{h_n}} = \frac{1+c_2\log(n/h_n)}{\sqrt{h_n}}$$

con c_1 e c_2 opportune costanti legate ai parametri di $X(P)$. In altre parole, per queste classi di distribuzioni, considerano gli stimatori $\frac{c_1}{\sqrt{\widetilde{H}_n}}$ e $\frac{1+c_2\log(n/\widetilde{H}_n)}{\sqrt{\widetilde{H}_n}}$.

Nonostante questa limitazione ad uno specifico contesto parametrico, dalle simulazioni si evidenzia che:

1. Lo stimatore non parametrico $\widehat{V_n}$, anche per valori piccoli di n, è tale che $\mathrm{Var}[\widetilde{H}_n] - \widehat{V_n}$ abbia varianza bassa e poca distorsione

Nonostante questa limitazione ad uno specifico contesto parametrico, dalle simulazioni si evidenzia che:

1. Lo stimatore non parametrico $\widehat{V_n}$, anche per valori piccoli di n, è tale che $\mathrm{Var}[\widetilde{H}_n] - \widehat{V_n}$ abbia varianza bassa e poca distorsione

2. Lo stimatore $c_1/\sqrt{\widetilde{H}_n}$ (risp. $(1 + c_2 \log(n/\widetilde{H}_n))/\sqrt{\widetilde{H}_n}$) può essere decisamente dissimile da $\mathrm{Var}[\widetilde{H}_n]$ anche per valori elevati di n.

Nonostante questa limitazione ad uno specifico contesto parametrico, dalle simulazioni si evidenzia che:

1. Lo stimatore non parametrico $\widehat{V_n}$, anche per valori piccoli di n, è tale che $\mathrm{Var}[\widetilde{H}_n] - \widehat{V_n}$ abbia varianza bassa e poca distorsione
2. Lo stimatore $c_1/\sqrt{\widetilde{H}_n}$ (risp. $(1 + c_2 \log(n/\widetilde{H}_n))/\sqrt{\widetilde{H}_n}$) può essere decisamente dissimile da $\mathrm{Var}[\widetilde{H}_n]$ anche per valori elevati di n.
3. La copertura, a livello $(1-\alpha)$, degli intervalli di confidenza

$$C_n = \{\llbracket \widetilde{H}_n - z_{1-\alpha/2}\sqrt{\widehat{V_n}} \rrbracket, \cdots, \llbracket \widetilde{H}_n + z_{1-\alpha/2}\sqrt{\widehat{V_n}} \rrbracket\}$$

è praticamente quella nominale anche per valori bassi di n.

Distribuzione di Weibull discretizzata $S(n) = e^{-n^{c_2}}$

c_2	n	h_n	$E[\widetilde{H}_n]$	$\mathrm{Var}[\widetilde{H}_n]$	$\frac{h_n}{(1+c_2 \log(n/h_n))^2}$	$E[\widehat{V}_n]$	Coper.
0.01	30	10	10.77	6.77	9.78	6.56	0.94
	50	17	17.86	11.25	16.64	11.05	0.96
	100	35	35.47	22.43	34.28	22.25	0.96
	150	52	52.98	33.57	50.92	33.39	0.96
	200	70	70.44	44.70	68.55	44.52	0.95
0.10	30	8	8.63	4.93	6.24	4.90	0.95
	50	13	13.60	7.83	10.10	7.82	0.95
	100	25	25.09	14.59	19.28	14.67	0.95
	150	35	35.83	20.95	26.67	21.12	0.95
	200	46	46.07	27.05	34.97	27.20	0.95

Tabelle

La classifica dei "top-ten" più citati in Statistica e Probabilità nel periodo 1985-2010 (citazioni prese da *Web-of-Science*).

	n	h_n	C_n
Hall, P.G.	418	46	$\{42, \ldots, 50\}$
Rubin, D.B.	104	39	$\{32, \ldots, 46\}$
Carroll, R.J.	198	38	$\{33, \ldots, 43\}$
Tibshirani, R.	104	37	$\{31, \ldots, 43\}$
Fan, J.	114	36	$\{30, \ldots, 42\}$
Marron, J.S.	107	36	$\{31, \ldots, 41\}$
Hastie, T.J.	77	34	$\{27, \ldots, 41\}$
Lin, D.Y.	93	32	$\{26, \ldots, 38\}$
Raftery, A.E.	88	31	$\{25, \ldots, 37\}$
Wei, L.J.	88	31	$\{26, \ldots, 36\}$

L'analisi della tabella precedente conduce a due riflessioni

1. Come si poteva facilmente immaginare, l'indice di Hirsch tende ad eguagliare il 'rendimento scintifico 'degli studiosi considerati, nonostante Peter Hall e Raymond Carroll abbiano un notevole numero di pubblicazioni rispetto agli altri colleghi.

L'analisi della tabella precedente conduce a due riflessioni

1. Come si poteva facilmente immaginare, l'indice di Hirsch tende ad eguagliare il 'rendimento scintifico 'degli studiosi considerati, nonostante Peter Hall e Raymond Carroll abbiano un notevole numero di pubblicazioni rispetto agli altri colleghi.

2. La maggioranza degli intervalli di confidenza (livello 95%) si intersecano simultaneamente (nella tabella sono 9) così che una stretta graduatoria può non essere determinabile.

Secondo noi, i 'professionisti delle valutazioni' farebbero bene a tener conto di queste considerazioni quando confrontano le produzioni scientifiche basandosi sull'h-index.

La classifica delle medaglie Fields dal 2002 al 2010 (citazioni prese da *Scopus*)

	n	h_n	C_n
Tao, T. (2006)	164	29	$\{25, \ldots, 33\}$
Villani, C. (2010)	55	21	$\{16, \ldots, 26\}$
Okounkov, A. (2006)	48	18	$\{16, \ldots, 20\}$
Werner, W. (2006)	39	16	$\{12, \ldots, 20\}$
Lindenstrauss, E. (2010)	26	8	$\{5, \ldots, 11\}$
Smirnov, S. (2010)	24	8	$\{6, \ldots, 10\}$
Bao Chau, N. (2010)	9	7	$\{5, \ldots, 9\}$
Voevodsky, V. (2002)	12	6	$\{3, \ldots, 9\}$
Lafforgue, L. (2002)	5	2	$\{0, \ldots, 4\}$
Perelman, G. (2006)	2	1	$\{0, \ldots, 2\}$

Recenti premi Nobel in Medicina (citazioni prese da *Scopus*)

	n	h_n	C_n
Steinmann, R.M. (2011)	412	110	$\{102, \ldots, 118\}$
Beutler, B.A. (2011)	308	77	$\{69, \ldots, 85\}$
Hoffmann, J.A. (2011)	203	71	$\{61, \ldots, 81\}$
Smithies, O. (2007)	297	66	$\{57, \ldots, 75\}$
Capecchi, M. (2007)	190	62	$\{54, \ldots, 70\}$
Blackburn, E.H. (2009)	227	59	$\{51, \ldots, 67\}$
Greider, C.W (2009)	97	58	$\{51, \ldots, 65\}$
Szostak, J.W. (2009)	203	56	$\{48, \ldots, 64\}$
zur Hauser, H. (2008)	338	54	$\{48, \ldots, 60\}$
Barr-Sinoussi, F. (2008)	251	45	$\{40, \ldots, 50\}$
Evans, M.J. (2007)	141	44	$\{37, \ldots, 51\}$
Montagnier, L. (2008)	311	42	$\{36, \ldots, 48\}$
Edwards, R.G. (2010)	316	41	$\{35, \ldots, 47\}$

Recenti premi Nobel in Economia
(citazioni prese da *Scopus*)

	n	h_n	C_n
Ostrom, E. (2009)	97	29	$\{25, \ldots, 33\}$
Krugman, P. (2008)	71	29	$\{22, \ldots, 36\}$
Sargent, T. (2011)	85	25	$\{21, \ldots, 29\}$
Diamond, P.A. (2010)	49	19	$\{15, \ldots, 23\}$
Maskin, E.S. (2007)	55	19	$\{14, \ldots, 24\}$
Myerson, R.B. (2007)	47	19	$\{15, \ldots, 23\}$
Pissarides, C.A. (2010)	37	17	$\{12, \ldots, 22\}$
Sims, C.A. (2011)	36	15	$\{10, \ldots, 20\}$
Mortensen, D.A. (2010)	28	12	$\{8, \ldots, 16\}$
Hurwicz, L. (2007)	20	7	$\{5, \ldots, 9\}$