

## Procedure di ricampionamento per la validazione di modelli di regressione dendro-auxometrici

Marziliano PA

*Dipartimento di Gestione dei Sistemi Agrari e Forestali (GESAF), Università Mediterranea di Reggio Calabria, Località Feo di Vito, I-89060 Reggio Calabria (RC - Italy) - \*Email: Pasquale Marziliano (pasquale.marziliano@unirc.it).*

**Abstract:** *Resampling procedures to validate dendro-auxometric regression models.* Regression analysis has a large use in several sectors of forest research. The validation of a dendro-auxometric model is a basic step in the building of the model itself. The more a model resists to attempts of demonstrating its groundlessness, the more its reliability increases. In the last decades many new theories, that quite utilizes the calculation speed of the calculators, have been formulated. Here we show the results obtained by the application of a bootstrap resampling procedure as a validation tool.

**Keywords:** Bootstrap, Regression, Model validation, Tree measurement

*Received: Nov 25, 2008; Accepted: Jan 19, 2009*

**Citation:** Marziliano PA, 2009. Procedure di ricampionamento per la validazione di modelli di regressione dendro-auxometrici. *Forest@* 6: 100-106 [online: 2009-03-25] URL: <http://www.sisef.it/forest@/>.

### Introduzione

In ambito forestale, soprattutto a partire dagli anni sessanta, si è assistito a una propensione sempre più marcata verso l'elaborazione e l'applicazione di modelli matematici (Corona 1996, Scarascia Mugnozza 1997). Tra questi, in ambito dendro-auxometrico, particolare importanza hanno assunto i modelli di regressione. Studiati e quantificati su un campione, i risultati dei modelli di regressione vengono poi estesi all'intero popolamento a prefissati livelli di precisione e di attendibilità.

La validazione di un modello è una fase fondamentale nella costruzione del modello stesso. Fondamentalmente vi sono due diversi modi di considerare le relazioni tra un modello di regressione e l'esperienza: si può cercare la conferma o la confutazione (Corona 1995). Popper (1972) afferma che i controlli scientifici sono tentativi di confutazione. Quanto più un modello resiste a tentativi di dimostrarne l'infondatezza, tanto più accresce la sua affidabilità.

Esistono diversi modi per corroborare un modello di regressione. Corona (1995) ne indica diversi, quali ad esempio l'analisi dei residui forniti dal modello sulla base dei dati mediante i quali sono stati stimati

i coefficienti del modello stesso, alcuni indicatori grafici e numerici (la media algebrica dei residui semplici, la media algebrica dei residui assoluti), ecc. Per ulteriori approfondimenti si rimanda a Corona (1995).

Caratteristica comune di questi modi per corroborare e validare un modello di regressione è la relativa semplicità di calcolo. È possibile operare anche con una semplice calcolatrice tascabile e un foglio millimetrato. In effetti ciò è abbastanza naturale se si pensa che la gran parte di queste analisi sono state sviluppate nel corso del 1800 e nei primi decenni del 1900. In quegli anni si era riusciti ad articolare una teoria statistica che sfruttava appieno gli strumenti di calcolo allora disponibili.

Nella gran parte dei casi, oggi, quando ci si appresta ad elaborare un qualsiasi modello di regressione, le operazioni che vengono effettuate sono quelle note già da un secolo. L'unica differenza è che vengono effettuate da un calcolatore elettronico, e dunque il tempo impiegato è decisamente minore.

In effetti gli strumenti di calcolo attualmente disponibili, permettono di effettuare operazioni a velocità milioni di volte più elevate rispetto ad un tempo.

Ed è proprio per questo che negli ultimi anni si è assistito alla formulazione di molte nuove teorie e di molti nuovi metodi statistici che sfruttano largamente il calcolatore ad alta velocità.

Tra i metodi statistici sviluppati facendo un largo uso del calcolatore elettronico, si possono citare quello del *jackknife*, quello della *convalida incrociata*, quello della *replica ripetuta equilibrata*, quello del *bootstrap*.

Ciascuno di questi procedimenti genera insiemi di dati fittizi a partire dai dati iniziali e valuta la variabilità di una proprietà statistica a partire dalla sua variabilità su tutti gli insiemi di dati fittizi. I metodi differiscono l'uno dall'altro per il modo in cui vengono generati gli insiemi di dati fittizi.

Quello che si ottiene in cambio è la liberazione di due fattori limitativi che hanno dominato la statistica teorica sin dai suoi inizi: il presupposto che i dati si distribuiscano secondo una curva a campana e la necessità di concentrarsi su misure statistiche le cui proprietà teoriche possano essere analizzate matematicamente. Ad esempio è pratica comune, prima di procedere all'analisi statistica, effettuare alcune assunzioni. Tra queste assunzioni vi è spesso quella della distribuzione normale. Ne consegue che per insiemi di dati che non soddisfano i presupposti gaussiani, non è correttamente possibile determinare la precisione di alcune proprietà dei risultati, come ad esempio nella stima dei coefficienti di regressione.

I metodi che fanno largo uso del calcolatore possono risolvere la maggior parte dei problemi senza presupporre che i dati abbiano una distribuzione gaussiana.

In questa nota si prenderà in esame il metodo del *bootstrap* e la sua possibile applicazione in ambito dendro-auxometrico.

La procedura *bootstrap*, poco applicata in Italia in ambito dendro-auxometrico, è citata da Corona (1995) come possibile metodo per la validazione del modello ed è stata applicata da Scrinzi (1988) per analizzare la variabilità di alcuni coefficienti stimati (varianze d'errore) in un lavoro sulla precisione degli inventari assestamentali per cavallettamento totale.

### Impostazione del problema

Come esempio di applicazione del metodo *bootstrap* si farà riferimento a un lavoro sulla modellizzazione dendroauxometrica per piantagioni di pioppo euroamericano (Marziliano 1998, Corona et al. 2002).

Gli Autori, sulla base di rilievi annuali effettuati all'interno di 4 aree di saggio, hanno messo a punto

più funzioni di sviluppo dell'area basimetrica: a livello di popolamento, a livello di classe diametrica e a livello di singolo individuo.

A livello di popolamento, l'incremento di area basimetrica ( $\Delta G$ ) è stato stimato dalla seguente equazione (eqn. 1):

$$\Delta G_t = \gamma \cdot \left( \frac{G_t}{E_t} \right) (SI^\theta - \ln G_t)$$

dove  $\Delta G_t$  è l'incremento dell'area basimetrica del popolamento tra l'età  $t$  e l'età  $t+1$  ( $m^2 ha^{-1} anno^{-1}$ );  $G_t$  = area basimetrica del popolamento all'età  $t$  ( $m^2 ha^{-1}$ );  $E_t$  = Età del popolamento (anni);  $SI$  = altezza media del popolamento all'età di 10 anni (m).

A livello di classe diametrica, l'incremento di area basimetrica ( $\Delta g$ ) è stato stimato dalle seguenti equazioni (eqn. 2, eqn. 3):

$$\Delta g_i = \frac{\Delta G_t}{N_t} + (g_i - \hat{g}_t) \cdot \beta_t$$

$$\beta_t = \beta_1 + \beta_2 \cdot SI^2 \cdot E_t^{-2}$$

dove  $\Delta g_{it}$  è l'incremento individuale di area basimetrica tra l'età  $t$  e l'età  $t+1$  di un albero di classe diametrica  $i$ -esima (in  $m^2 anno^{-1}$ );  $\Delta G_t$  = incremento stimato di area basimetrica a livello di popolamento (in  $m^2 ha^{-1} anno^{-1}$ );  $N_t$  = numero di piante ad ettaro;  $(g_i - \hat{g}_t)$  = differenza tra l'area basimetrica individuale di un albero di classe diametrica  $i$ -esima ( $g_i$ ) e l'area basimetrica media all'età  $t$  nel popolamento considerato ( $\hat{g}_t$ ) in  $m^2$ ;  $E_t$  = età del popolamento (in anni);  $SI$  = indice di fertilità (in m);  $\beta_t$ ,  $\beta_1$ ,  $\beta_2$  = coefficienti da stimare.

Per ogni ulteriore approfondimento del lavoro sopraccitato si rimanda a Marziliano (1998) e a Corona et al. (2002).

L'equazione (3) si basa su 16 osservazioni (4 anni considerati per quattro parcelle sperimentali), ed è proprio su questa equazione che in questa nota verrà applicato il metodo *bootstrap*.

I valori dei coefficienti stimati dell'equazione 3 sono i seguenti:  $\beta_1 = 0.02695$ ;  $\beta_2 = 0.01371$ . Questi valori, immessi nell'equazione (3), danno la stima di  $\beta_t$ . A sua volta  $\beta_t$ , inserito nell'equazione 2, fornisce gli incrementi di area basimetrica a livello di classe diametrica. Dunque, per avere valori affidabili, la corretta stima di  $\beta_1$  e  $\beta_2$  è fondamentale.

Con quale fondamento però si può pensare che i valori di  $\beta_1$  e  $\beta_2$  siano davvero affidabili?

In effetti il campione potrebbe essere fortemente atipico (nel senso che non soddisfa i presupposti gaussiani) e solo la legge dei grandi numeri ci assicu-

ra da questo punto di vista. Ma un campione di 16 osservazioni non è un campione di grandi dimensioni, e pertanto si ha bisogno di qualche misura che valuti la precisione statistica dei valori di  $\beta_1$  e  $\beta_2$  dati dal campione. E per insiemi di dati che non soddisfanno i presupposti gaussiani, i risultati di metodi statistici che si basano su tali presupposti sono ovviamente meno affidabili. Il metodo del *bootstrap* è stato formulato proprio per dare soluzione a problemi di questo tipo. Questi metodi possono risolvere la maggior parte dei problemi senza presupporre che i dati abbiano una distribuzione gaussiana.

Risulta d'altra parte evidente (ed è utile sottolinearlo) che le tecniche di ricampionamento creano artificiosamente l'universo dei dati: se la base di dati raccolti non è tale da coprire la reale variabilità del vero universo di interesse l'applicazione di queste tecniche conduce comunque a risultati teoricamente del tutto inadeguati. O anche, se il campione è preso in maniera sbagliata, ad esempio in una popolazione arborea con campo di variabilità di diametri tra 10 e 50 cm con media intorno a 25 cm, gli alberi campionati presentano diametri solo da 25 cm in su, qualsiasi tecnica di ricampionamento dei dati non può che condurre che a risultati comunque sbagliati.

In questi casi non è il *bootstrap* che non funziona, è la procedura di campionamento che è stata errata!!! Da qui dunque l'importanza fondamentale di un corretto campionamento delle unità statistiche. E questo vale sia per applicazione di metodi statistici classici che per procedure di ricampionamento quali il *bootstrap*.

È utile altresì dire, più in generale, che non è detto che il *bootstrap* o altri metodi di ricampionamento non si debbano applicare anche se il campione è di grandi dimensioni. Questi metodi sono utili, e a volte necessari, in tutte quelle situazioni in cui non si è in grado di assegnare analiticamente la precisione alla stima del parametro che interessa.

Dunque, tramite il *bootstrap* si vuole determinare la precisione statistica dei coefficienti stimati nell'equazione (3).

### La validazione del modello

Si supponga che siano disponibili i dati di altri campioni formati da 16 osservazioni, i quali derivano da altre aree di saggio realizzate nello stesso popolamento.

Per ciascuno di questi campioni si potrebbe impostare l'equazione (3). Applicando il metodo dei minimi quadrati si specifica quindi una curva il più possibile vicina a tutte le osservazioni disponibili. In

questo modo è possibile descrivere il grado di variazione delle curve interpolatrici, attraverso i coefficienti che li generano.

Se, ad esempio, il 99% delle curve generate per gli ipotetici campioni si trovassero concentrate intorno a quello stimato con i dati originali, allora si potrebbe attribuire un'elevata precisione alla stima eseguita col campione originario. Se invece i valori di  $\beta_1$  e  $\beta_2$  fossero distribuiti omogeneamente, la stima data dal campione originale non avrebbe alcuna precisione e pertanto non sarebbe di alcuna utilità. Dunque, la precisione statistica di un valore stimato dipende dall'ampiezza dell'intervallo, attorno al valore stimato, che comprende una determinata percentuale dei valori di tutti i campioni (Hall 1992, Lepage 1992, Dixon 1993, Efron & Tibshirani 1993, Hjorth 1994, Shao & Tu 1995).

Ma gli unici dati disponibili sono solamente quelli del campione originario formato da 16 osservazioni. In effetti, se fossero disponibili altri dati, si potrebbero utilizzare per dare una stima migliore dei valori trovati.

Non è neanche possibile effettuare la validazione del modello suddividendo i valori osservati in due insiemi: su un insieme viene condotta la calibrazione del modello e sull'altro la validazione. I dati a disposizione sono troppo pochi.

Dunque, in situazioni del genere, e in ambito dendro-auxometrico non sono le eccezioni, i metodi statistici tradizionali non offrono soluzioni.

### Il metodo statistico *bootstrap*

In casi come quello esposto il procedimento *bootstrap* si rivela molto utile. Il metodo può essere usato come uno strumento per stimare la precisione statistica dei coefficienti stimati a partire dai dati relativi a un unico campione.

La procedura *bootstrap* è stata elaborata da B. Efron del Dipartimento di statistica della Stanford University nel 1977. Benché concettualmente molto semplice, l'applicazione del *bootstrap* richiede un uso intensivo e prolungato di risorse di calcolo elettronico.

L'idea è di simulare il processo di selezione di molti campioni di grandezza 16, al fine di trovare la probabilità che i valori dei coefficienti cadano all'interno di vari intervalli. I campioni sono generati a partire dai dati del campione originale.

I campioni di *bootstrap* sono generati nel modo seguente (Efron & Tibshirani 1993):

- si fa un enorme numero di copie (anche fino a 1 miliardo) dei dati della prima osservazione e si fa lo stesso numero di copie per i dati di ciascuna al-

- tra osservazione;
- i miliardi di copie che risultano dopo la copiatura vengono mescolati completamente;
- si scelgono a caso campioni di grandezza 16 e si calcolano, per ciascuno di questi campioni, i coefficienti dell'equazione.

Con un calcolatore i passaggi della copiatura, del mescolamento e della selezione dei nuovi insiemi di dati vengono effettuati mediante un procedimento molto veloce. I campioni artificiali così generati vengono chiamati campioni di *bootstrap*. La distribuzione dei coefficienti stimati per i campioni di *bootstrap* può poi essere trattata come una distribuzione costruita a partire da campioni reali e fornisce una stima della precisione statistica dei valori di  $\beta_1$  e  $\beta_2$ , calcolati per il campione di partenza.

Il metodo del *bootstrap*, dunque, può stimare, a partire dall'unico insieme di 16 osservazioni campionarie, il grado di variabilità che i coefficienti stimati mostrerebbero se si potessero mettere a confronto molti insiemi di 16 osservazioni campionare.

Il procedimento fonda le sue basi su due principi che sono alla base della statistica stessa: il primo è quello detto della *sostituzione o principio plug-in*, e cioè di sostituire alla funzione di ripartizione teorica  $F$  della popolazione, il suo equivalente empirico  $F_n$ , cioè la funzione di ripartizione del campione, ottenuta costruendo una distribuzione di frequenza di tutti i valori che esso può assumere in quella situazione sperimentale. Il secondo principio non è altro che la legge dei grandi numeri, la quale ci assicura che la media campionaria delle  $n$  repliche converge quasi certamente al suo valore teorico al tendere di  $n$

all'infinito.

### Applicazione del *bootstrap*

I dati a disposizione (16 osservazioni per ogni variabile) sono stati moltiplicati 20000 volte, così da ottenere un *dataset* formato da 320000 osservazioni per ogni variabile. Da questo *dataset* sono stati generati 2000 campioni di *bootstrap* e per ognuno di essi sono stati calcolati i coefficienti dell'equazione (3).

Riassumendo, applicando il metodo del *bootstrap* ai dati, sono stati generati gruppi di dati fittizi. Con il metodo dei minimi quadrati applicato a ciascun gruppo fittizio si sono generate altrettante curve interpolatrici. La variabilità delle curve generate dal metodo *bootstrap* indica la precisione statistica dei coefficienti stimati con i dati originali.

I risultati scaturiti dall'applicazione del *bootstrap* sono presentati in Tab. 1. Le Fig. 1 e Fig. 2 evidenziano le distribuzioni di frequenza per i coefficienti  $\beta_1$  e  $\beta_2$  per 2000 campioni di *bootstrap*.

La tabella mostra che il coefficiente  $\beta_1$  può assumere valori che variano da -0.01887 a 0.07132. Il valore stimato con il campione originale è uguale a 0.02695, mentre il valore medio di tutti i coefficienti stimati con procedimento *bootstrap* è uguale a 0.02677 (valore vicino a quello stimato con il campione originale). La differenza tra i campioni *bootstrap* e il campione originario (*bias*) è di -0.00018. Una misura ampiamente accettata della precisione dei coefficienti stimati è data dalla fascia pari al 68% (1 deviazione standard) dell'area delimitata dalla distribuzione nel suo complesso (Efron & Tibshirani 1993, Hjorth 1994). La tabella mostra che il 68% dei valori sono

**Tab. 1** - Analisi del *bootstrap* per la funzione (3) - vedi testo. Output generato dal software SAS (*Bootstrap Method*).

Name	Observed Statistic	Bootstrap Mean	Approximate Bias	Approximate Standard Error	Approximate Lower Confidence Limit	Bias-Corrected Statistics	Approximate Upper Confidence Limit
$\beta_1$	0.02695	0.02677	-0.00018	0.01229	0.01492	0.02714	0.03937
$\beta_2$	0.01371	0.01374	0.00003	0.00121	0.01248	0.01369	0.01489
RMSE	0.03402	0.03023	-0.00379	0.00998	0.02788	0.03780	0.04773
RSQ	0.88891	0.90251	0.01360	0.05764	0.81799	0.87531	0.93261
Name	Confidence Level (68%)	Method for Confidence Interval	Minimum Resampled Estimate	Maximum Resampled Estimate	Number of Resamples		
$\beta_1$	68	Bootstrap normal	-0.01887	0.07132	2000		
$\beta_2$	68	Bootstrap normal	0.01023	0.01883	2000		
RMSE	68	Bootstrap normal	0.00994	0.05578	2000		
RSQ	68	Bootstrap normal	0.66702	0.99028	2000		

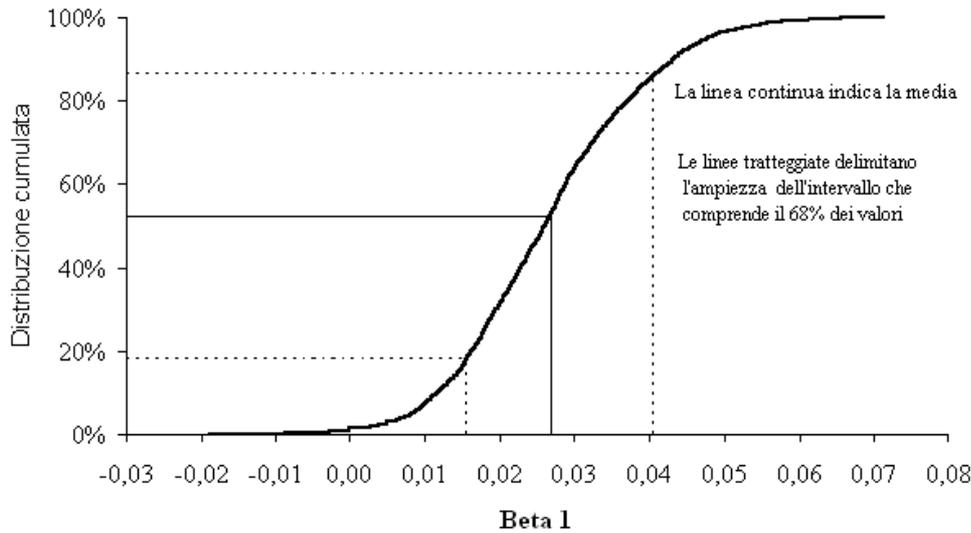


Fig. 1 - Distribuzione di frequenza per il coefficiente  $\beta_1$ .

compresi tra 0.01492 e 0.03937. In Fig. 1 sono riportate le distribuzioni di frequenza per il coefficiente  $\beta_1$  per 2000 campioni di *bootstrap*. L'ampiezza dell'intervallo pari alla fascia che contiene il 68% di tutti i valori stimati è 0.02501. La metà dell'ampiezza di tale intervallo, cioè 0.0125, costituisce una buona stima dello scostamento medio fra il valore di  $\beta_1$  osservato per un campione e il suo valore più probabile. Inoltre Efron & Tibshirani (1993) affermano che il rapporto tra il Bias e l'errore standard deve essere uguale o minore di 0.25 (*rule of thumb*). Nel caso di  $\beta_1$  tale rapporto è uguale a -0.015, valore molto piccolo e questo indica che la leggera differenza tra il valore originario stimato e il valore stimato con procedura *bootstrap* può essere tranquillamente ignorata.

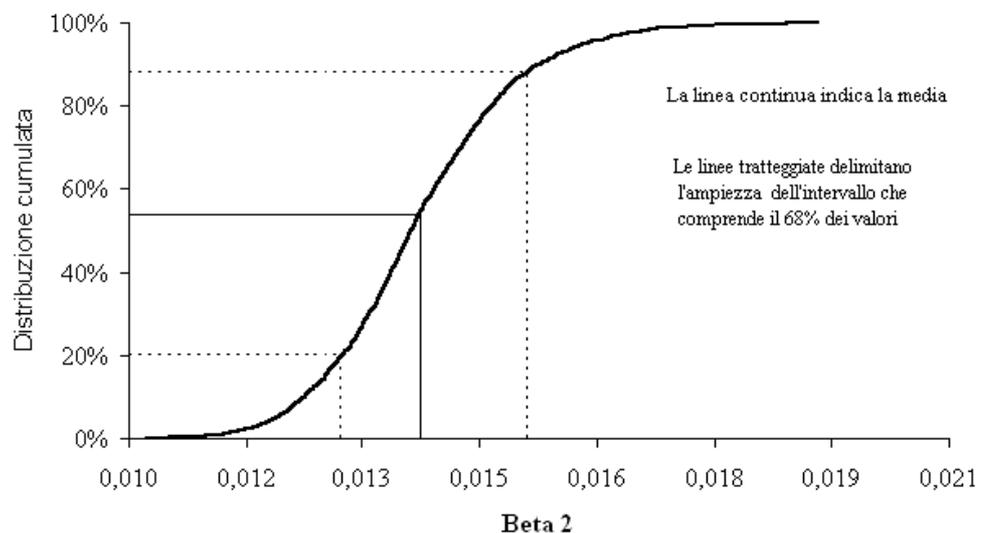
Per quanto riguarda  $\beta_2$ , il coefficiente può assumere valori che variano da 0.01023 a 0.01883. Il valore stimato con il campione originale è uguale a 0.01371,

mentre il valore medio di tutti i coefficienti stimati con procedimento *bootstrap* è uguale a 0.01374 (molto vicino a quello stimato con il campione originale). La differenza tra i campioni *bootstrap* e il campione originario (Bias) è di 0.00003. L'ampiezza dell'intervallo la cui area è pari al 68% dell'area delimitata dalla distribuzione nel suo complesso è uguale a 0.00237. La metà dell'ampiezza dell'intervallo, 0.00118, è la stima di *bootstrap* della distanza media tra il valore osservato di  $\beta_2$  e il suo valore più probabile. Anche in questo caso il rapporto tra il Bias e l'errore standard è molto piccolo (0.025) e dunque anche per  $\beta_2$  la leggera differenza tra il valore originario stimato e il valore stimato con procedura *bootstrap* può essere tranquillamente ignorata.

È possibile concludere questa fase d'analisi con le seguenti considerazioni:

- il valore dei coefficienti stimati con il campione

Fig. 2 - Distribuzione di frequenza per il coefficiente  $\beta_2$ .



originale è molto simile al valore medio dei coefficienti stimati con la procedura *bootstrap*;

- la stima di *bootstrap*, pur presentando una certa variabilità, non è tale da rendere privi di valore i coefficienti stimati;
- la stima di *bootstrap* è inferiore agli errori standard dei coefficienti stimati con il campione originale;
- alla luce delle procedure di ricampionamento effettuate, i coefficienti stimati con il campione originale, possono essere considerati attendibili.

Sin qui il procedimento *bootstrap* è stato utilizzato per valutare la variabilità di grandezze già stimate. Non è questo il solo uso possibile del *bootstrap*. Il metodo può essere utilizzato, ad esempio in un modello di regressione multipla, anche come strumento per pervenire alla determinazione delle variabili più importanti nella definizione dell'equazione.

Si supponga che in un popolamento di douglasia l'incremento di area basimetrica possa essere rappresentato da una funzione del tipo:

$$\Delta G = f \left( Et\dot{a}; Et\dot{a}^2; \frac{1}{Et\dot{a}}; \frac{1}{Et\dot{a}^2}; G; \frac{G}{Et\dot{a}}; NP; SI; \frac{1}{SI^2} \right)$$

In statistica è pratica comune controllare ogni singola variabile, prima di procedere a costruire un modello, con l'obiettivo di escludere tutte le variabili tranne quelle che presentano più significatività (*Step-wise Regression*). Si supponga che dopo l'analisi preliminare il modello sia definito da:

$$\Delta G = \gamma_0 + (\gamma_1 \cdot SI^{-2}) + (\gamma_2 \cdot Et\dot{a}^{-2}) + (\gamma_3 \cdot G)$$

Si programmi ora un calcolatore in modo da copiare molte volte l'insieme delle 9 variabili di partenza associate a ciascuna osservazione. Gli insiemi di dati sono poi mescolati e vengono estratti casualmente campioni di *bootstrap*. Per ognuno dei campioni di *bootstrap* vengono eseguite le stesse analisi che erano state fatte per il campione originale, allo scopo di individuare le variabili più importanti. Se, ad esempio, nel 50% o più dei campioni di *bootstrap*, vengono escluse alcune variabili che erano state identificate come significative con il campione originario, allora bisogna concludere che le variabili identificate nell'analisi originale devono essere considerate con molta cautela. Viceversa se le variabili identificate come significative nella gran parte dei campioni di *bootstrap* sono identiche a quelle identificate con il campione originale, il modello ne esce rafforzato.

## Discussione e conclusioni

Il procedimento potrebbe prestarsi a qualche obiezione, la principale delle quali è: i campioni *bootstrap*

sono dati fittizi, cioè sono stati generati dai dati originali; se si fossero usati campioni reali l'ampiezza dell'intervallo sarebbe stato lo stesso?

La risposta implica ragionamenti complessi basati sulle leggi statistico-probabilistiche. Studi teorici effettuati da R.J. Beran, P.J. Bickel e D.A. Freedman dell'Università della California a Berkeley, da K. Singh della Rutgers University, da P. Diaconis e B. Efron della Stanford University dimostrano che per un'ampia gamma di coefficienti statistici (tra i quali i coefficienti stimati in un'analisi di regressione), l'intervallo associato con la distribuzione di *bootstrap* e l'intervallo associato con la distribuzione dei campioni reali di solito hanno ampiezza quasi uguale.

Il procedimento *bootstrap* fornisce una buona immagine della precisione della stima nella maggior parte dei casi (è comunque importante dire che questo è vero solo e solo se il campione è stato estratto in modo del tutto casuale, o comunque oggettivo, cioè con campionamento *design based* e non *model based*, come invece avviene nella gran parte dei casi per le stime di regressione) (Corona, comunicazione personale). Ciò nondimeno esistono sempre alcuni campioni per i quali il procedimento non funziona. La limitazione non è tanto un difetto del procedimento, quanto una riconferma delle condizioni di incertezza sotto le quali qualunque analisi statistica deve procedere (Lepage 1992, Diaconis & Efron 1997).

Concludendo, se avessimo diversi campioni estratti dalla stessa popolazione potremmo calcolare il valore dello stimatore su ogni campione e poi osservarne la distribuzione oppure calcolarne la variabilità, ma questo caso si verifica raramente. Molto spesso si ha a disposizione un solo campione e di conseguenza un solo valore dello stimatore, il che non permette di sapere nulla sulla sua precisione. Procedure di ricampionamento quali il *bootstrap* permettono di stimare la distribuzione campionaria di una qualsiasi statistica sulla base dei soli dati empirici. Il vantaggio della procedura risiede nel fatto che attraverso di essa diviene possibile stimare i parametri di distribuzioni non note e quindi verificare ipotesi statistiche sia quando non siano disponibili le distribuzioni campionarie di riferimento sia quando le distribuzioni campionarie note non siano utilizzabili perché non sono rispettate le assunzioni alla loro base.

L'idea alla base del *bootstrap*, dunque, è quello di ricavare dalla distribuzione empirica del campione, che è l'unica informazione in genere disponibile, numerosi campioni con una procedura di ricampionamento. In questo modo si ottengono diverse stime del parametro che interessa, con le quali, grazie all'a-

iuto del computer e senza utilizzare formule matematiche particolarmente complicate, è possibile ottenere misure di variabilità dello stimatore quali errore standard, distorsione e intervalli di confidenza e quindi di usare le stime *bootstrap* per ogni metodo statistico.

Il metodo del *bootstrap* ha tutte le caratteristiche per trovare ampia applicazione in ambito forestale e può essere usato in moltissime situazioni operative e in tutti quei casi in cui si sia estratto un campione, calcolata una statistica e in cui si desideri conoscere quanto precisa sia quella statistica come misura per la popolazione dalla quale il campione proviene. Quello che si ottiene è la liberazione dell'ipotesi semplificante della distribuzione normale dei dati.

Ciò che più sorprende del procedimento *bootstrap* è la facilità d'uso. È un metodo nato nell'epoca del *Personal Computer*, senza il quale non sarebbe materialmente possibile realizzare tutto il procedimento.

In arboricoltura da legno, in dendrometria, e specificatamente nelle previsioni di stima di produzioni legnose il metodo può trovare larga applicazione, potendosi così affiancare ad altri metodi già comunemente usati.

## Bibliografia

- Corona P (1995). Criteri di corroborazione dei modelli di regressione. *Italia Forestale e Montana* 4: 390-403.
- Corona P (1996). Mathematical modelling in forestry: problems and perspectives. *Annali Accademia Italiana di Scienze Forestali XLV*: 41-61.
- Corona P, Marziliano PA, Scotti R (2002). Top-down growth modelling: a prototype for poplar plantations in Italy. *Forest Ecology and Management* 161: 65-73.
- Diaconis P, Efron B (1997). Statistica al calcolatore. *Le Scienze quaderni* 98: 72-84.
- Dixon PM (1993). The bootstrap and the jackknife: describing the precision of ecological indices. In: "Design and analysis of ecological experiments" (Scheiner SM, Gurevitch J eds). Chapman & Hall, New York, pp. 290-318.
- Efron B, Tibshirani RJ (1993). An introduction to the bootstrap. Chapman & Hall, New York, USA.
- Hall P (1992). The bootstrap and edgeworth expansion. Springer-Verlag, New York, USA.
- Hjorth JSU (1994). Computer intensive statistical methods. Chapman & Hall, London, UK.
- Lepage R (1992). Exploring the limits of bootstrap. Lynne Billard Editions, New York Wiley, pp. 426.
- Marziliano PA (1998). Metodologie di stima della crescita e della produzione in piantagioni da legno a densità definitiva. Tesi del dottorato di ricerca in Arboricoltura da legno, Università della Basilicata.
- Popper KR (1972). The logic of scientific discovery. Hutchinson, London, UK.
- Scarascia Mugnozza G (1997). Processi e meccanismi funzionali negli ecosistemi forestali: integrazioni di scala e moderne tecnologie sperimentali. *SISEF Atti I, Legnaro (PD)*, pp. 269-272.
- Scrinzi G (1988). Precisione degli inventari assestamentali per cavallettamento totale: un'analisi di trent'anni di esperienza applicativa in Trentino. *Annali dell'Istituto Sperimentale per l'Assestamento Forestale e per l'Alpicoltura XI*: 143-186.
- Shao J, Tu D (1995). The Jackknife and Bootstrap. Springer-Verlag, New York, USA.