林木直径成長データを用いた成長傾向に対する変化点探索法の構築

Detecting Changing Points of Tree Diameter Growth Process

二宮 嘉行・吉本 敦

Ninomiya, Y. and Yoshimoto, A.

キーワード: 直径成長、測樹学、変化点、非線形成長関数、統計的検定

要約

隣接木との競合や除間伐、自然枯死などによる周辺環境の変化の影響を考慮すると、林木の成長はある傾向的な成長過程においてなんらかの変化を伴っているものと言える。本研究では、過去の施業履歴がない林分から採取した林木の直径成長データを用いて、傾向的な成長過程に変化が発生しているか否かを統計的に判断できる方法を提示する。この方法は、ある非線形成長曲線モデルに対し、「成長に変化がなかった」という帰無仮説と「変化があった」という対立仮説とによる統計的検定問題を基にするものである。

Abstract:

Considering the competition among individual trees and some artificial effects on surrounding tree growth environments by pre-commercial, commercial thinning and self-thinning, each tree grows not only following its own growth trend but also with some structural change in growth pattern. In this paper, we propose a statistical method to investigate whether a structural change occurs in the growth of a tree. Using tree diameter growth data without any record on the management treatments, we present how the proposed method can be used. The proposed method is based on statistical testing under the null hypothesis supposing no change against the alternative hypothesis supposing the existence of the change.

Key Words: Diameter growth, Forest biometrics, Change-point, Non-linear growth function, Statistical testing

はじめに

時系列解析では，時系列に内在する情報に基づき確率モデルを用いて将来予測をすることがあるが、その際、時系列 (データ) を生成している構造を的確に反映することが重要となる。一般に長い期間に渡り時系列の構造が不変であることは少なく、構造に変化があるかどうかを判断 (変化検知) し、変化がある場合は変化した時点 (変化点) を考慮すること(変化点推定)が時系列に対するモデリングにおいて基本的な分析の一つとなる．例えば構造がなんらかの原因である時点に変化した場合、変化した時点の前後で確率モデルを変えるという変化点を考慮したモデル（変化点モデル）が必要になる。

このような変化点探索ではその推定上の特異性から通常の統計理論を適用できないことが多く、現在でも様々な変化点探索に関わる理論研究が行われている。例えば、変化点の推定量にはデータを増やせば真の変化点に近づくという性質 (一致性) があるが、その示し方は一般の方法 (例えば Wald 1949) と異なる。この点に関しては、これまで Chernoff & Rubin (1956)、Carlstein (1988)、Loader (1996) などにより、議論がなされている。さらに、変化の有無を調べるための尤度比検定統計量も通常のように漸近的に χ^2 分布に従わないため、やはり長い間その性質が議論されている (James et al. 1987、Horváth 1993、Davis et al. 1995 など)。その他詳しい研究の流れについては Csörgő & Horváth (1996) や Chen & Gupta (2000) といった本を参照されたい。

上記の変化点探索に関する統計的解析は変化点解析と呼ばれ、計量経済や品質管理といった研究分野で様々な応用がなされている。その例としては、ナイル河の流量に関する研究 (Cobb 1978)、炭鉱事故件数に関する研究 (Jarrett 1979)、株価の収益率に関する研究 (Chen & Gupta 1997) などがある。森林科学の分野においても森林の成長過程において変化点を観察することができる。林分内における各林木の成長は、例えば隣接木との競合や除間伐、自然枯死などによる周辺環境の変化の影響を受けている。そして、その成長過程

はそうした事象が発生した後、ある変化点を経て過去の成長構造とは別のものに変化するものと考えられる。本研究では、そうした森林の成長過程における変化点探索のための統計分析方法を提示することを目的とする。ここでは過去の施業履歴がない林分から採取した林木の直径成長データを用いて、傾向的な成長過程に変化が発生しているか否かを統計的に判断できる方法を提示する。これまで森林資源量予測のための成長モデル構築では、ある成長関数を仮定し、データへのあてはめによりパラメータが決定され、成長予測が行われてきたが、上記のような隣接木との競合などの影響を考慮すると、その成長は仮定した成長関数のみにより表現することは困難である。仮に成長過程に変化があれば、より正確な予測を可能とするために、その変化を取り入れた新たな成長モデルの構築が必要不可欠となる。その結果、より確かな予測に基づいた資源管理において適切な戦略が考案できるものと考えられる。

1．変化検知のための検定の一般論

成長過程において「変化はない」という帰無仮説と「変化はあった」という対立仮説を考える。他の統計解析同様、変化点解析でも、仮説の採否を判断するための検定統計量として尤度比統計量を用いることが多い。ここでは最も簡単な例として、図4.1のような平均が変化する独立観測の正規系列 $\{X_t;\ t=1,\dots,m\}$ を考え、変化点解析のための検定統計量について説明する。ここで t は時刻を意味する。系列 $\{X_t;\ t=1,\dots,m\}$ は独立に $N(\mu,1)$ に従うという帰無仮説と、時刻 k までは独立に $N(\mu^{(1)},1)$ に、時刻 $k+1$ 以降は独立に $N(\mu^{(2)},1)$ に従うという対立仮説とをたてる。なお μ, $\mu^{(1)}$, $\mu^{(2)}$, k は未知とする。$\sum_{t=i+1}^{j} X_t$ を $\bar{X}_i^j$ と記すことにすれば、帰無仮説のもとでの最大対数尤度 L_0 は

(1) $$L_0 = -\frac{m}{2}\log(2\pi) - \sum_{t=1}^{m}\frac{1}{2}(X_t - \bar{X}_0^m)^2,$$

対立仮説のもとでの最大対数尤度 L_1 は

(2) $$L_1 = -\frac{m}{2}\log(2\pi) - \max_{1\le k\le m-1}\{\sum_{t=1}^{k}\frac{1}{2}(X_t - \bar{X}_0^k)^2 - \sum_{t=k+1}^{m}\frac{1}{2}(X_t - \bar{X}_k^m)^2\}$$

であり、従って尤度比統計量 LR (最大対数尤度の差の 2 倍) は

(3) $$\mathrm{LR} = 2 \max_{1\le k\le m-1} \{(\bar{X}_0^k - \bar{X}_0^m)^2 + (\bar{X}_k^m - \bar{X}_0^m)^2\}$$

となる。仮にどこかで平均に変化があればこの統計量の実現値が大きくなるのは自明であり、直感的に変化点の有無を判断しやすい指標といえる。なぜならば、もし時刻 $t=k$ で変化があるとき、時刻 k までの平均 $\bar{X}_0^k$ や時刻 $k+1$ 以降の平均 $\bar{X}_k^m$ と、全体の平均 $\bar{X}_0^m$ との差は大きくなるからである。ちなみに、変化ありとみなした時、右辺の max を与える k が変化点位置推定量となる。

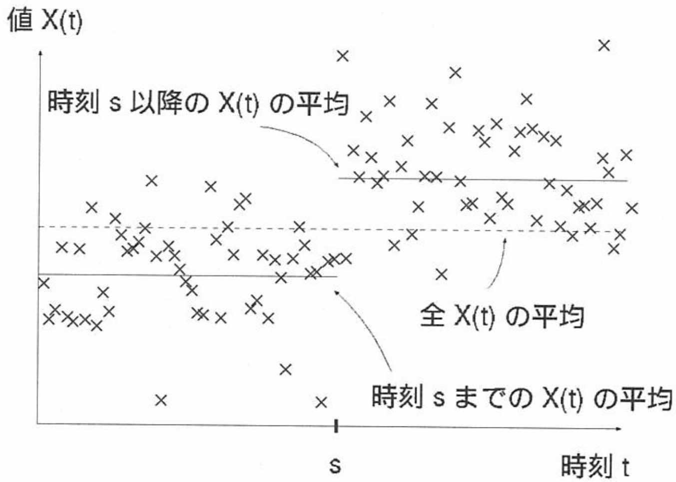

図 4.1 平均が変化する系列

変化点の存在に対する検定において、帰無仮説のもとでの検定統計量 T の分布関数を $F(x) \equiv \mathrm{P}(T \le x)$ 、検定統計量の実現値を x_T とし、p 値 $1-F(x_T)$ が 5 (あるいは 10) %以下ならば帰無仮説を棄却する、という一般的な方法をここでは用いる。なお、p 値 $1-F(x_T)$ を厳密に求められない場合は漸近論を用いた評価や上限を代用するのが一般的であるが、上限を代用

する場合は、帰無仮説の棄却は慎重にするべきであるという発想を基にするため保守的な検定の構成と言える。また、通常の統計モデルでは尤度比統計量は漸近的に χ^2 分布に従うのでそれに基づいて p 値が評価できるが、変化点モデルではそうはならず、その評価が問題となる。上の例では尤度比統計量はある Brownian ブリッジの最大値に分布収束することが知られており (Gombay & Horváth 1990)、それを用いた p 値の評価がなされる。

2．成長曲線のモデルあてはめ

本稿で用いたデータを図4. 2 に示す。これは宮崎県西郷村にある 38 年生の杉から４本採取した林木の直径成長データである。11、12 年目あたりに成長の促進があるように観察できるが、本稿で提示した方法により、これが本当に成長の変化なのか、あるいはノイズ（偶然変動）なのか、の判断を与えることができる。

これらの成長データに対し、変化点を調べる前に、変化点のない帰無仮説のモデルをあてはめる。ここでは、Richards (Richards 1958) の成長曲線を考慮し、直径成長データの差分 $\{Y(t);\ t=1,\ldots,m\}$ に対して

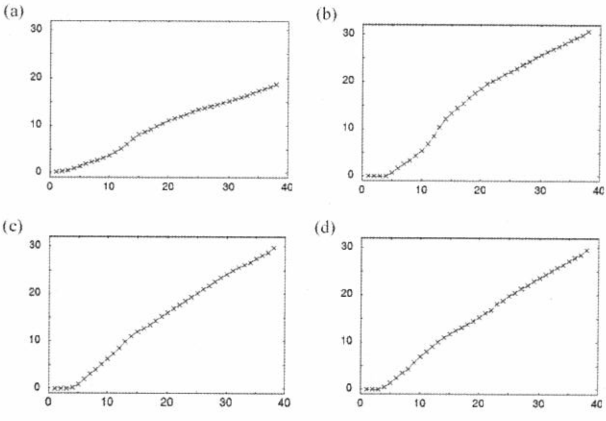

図4.2 林木直径成長データ（横軸：年、縦軸：cm）

$$(4) \qquad Y(t) = g(t \mid a,b,c) + \varepsilon(t), \quad \varepsilon(t) \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$$

なるモデルを仮定する。ここで

$$(5) \qquad g(t \mid a,b,c) \equiv a\exp(-bt)\{1-\exp(-bt)\}^c$$

である。元データを直接用いると傾向的なノイズの発生、すなわち一度正のノイズが出たらしばらく正のノイズが続くため、差分データに変換してノイズに独立性を仮定する。なお、ただ成長曲線をあてはめるだけであれば、ノイズの独立性あるいは相関はそれほど結果に影響を与えないが、変化点検知においてはノイズの相関によって変化点の有無が説明されてしまうため、ノイズの独立性の仮定は必要不可欠である。

図 4.3は差分データに対し、最尤法を用いてモデルをあてはめ、ノイズを除去して成長曲線（の差分）を得たものである。この図からわかるように、仮定したモデルではデータの動向を捉えきれていないことがわかる。

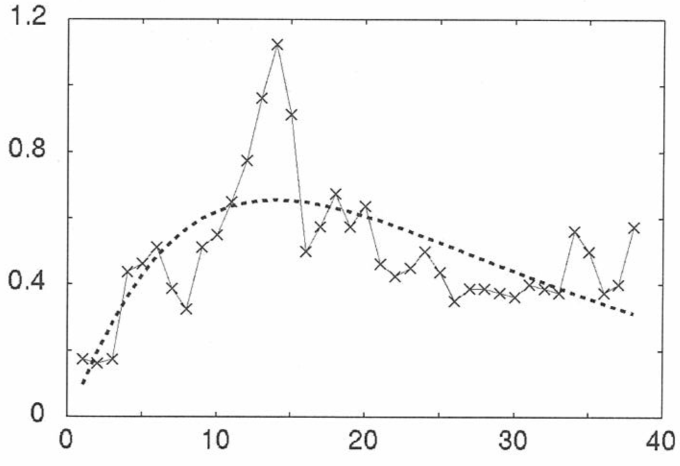

図4. 3 差分データに対するモデルのあてはめ（横軸：年、縦軸：cm）

3．成長曲線に対する変化検知法

林木の成長データの変化は、前述の例としてあげた正規系列の平均シフトとは性質が異なるため、検知方法も工夫する必要がある。仮になんらかの変化をもたらす事象があったとき、林木は事象発生後それまでの過程とは異なる成長をし、場合によっては徐々にその効果はなくなることが考えられる。そうした場合、得られるデータに対して成長曲線モデルをあてはめると、モデルとしての柔軟性により、図 4.4 のようにある程度うまくあてはまることが予想できる。例えば成長が促進するような変化が起こったとき、変化点前からデータは徐々にモデルより下回っていき、変化点後に今度は上の方に外れ、そして徐々にモデルに従うようになることが予想できる。つまり平均シフトの例とは異なり、変化点から遠く離れたデータは変化検知のための情報をもたないと考えられる。

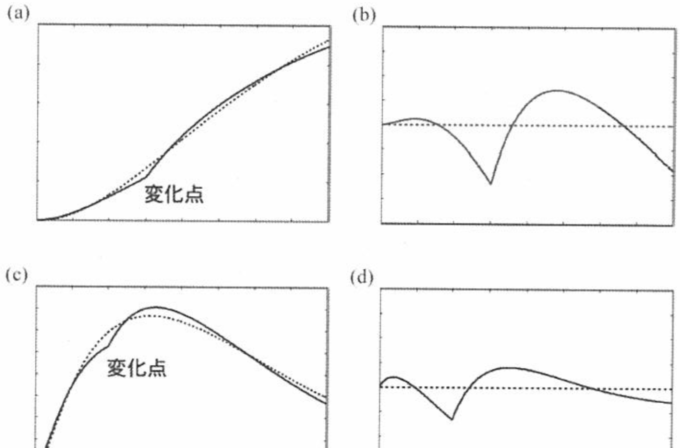

図4. 4　成長曲線モデルの変化

もともとの成長曲線、あるいはその差分に対し、変化点モデル（実線）とそれにあてはめた変化なしのモデル（点線）を描いたのが (a) と (c)、二つの曲線の差を描いたのが (b) と (d)

以上のことをふまえ、全データに対して変化なしの成長曲線モデルをあてはめ、生成される誤差の系列に対して図4.5のようなある種の信号が入っているかどうかを判断することにより変化の有無を分析する方法を次に提示する。この方法は、ある意味で信号検知の方法ともいえる。なお、変化点探索において、全データの中のある場所 (一部) だけみてモデルをたて、変化の有無を調べる、ということを様々な場所で行うことは自然に考えられる方法である。このとき検定統計量が複数個 $T_1, T_2, \ldots$ 得られることになるが、p値 $1-\mathrm{P}(T_1 \le x_1, T_2 \le x_2, \ldots)$ の理論的な評価が不可能となるため、ここではそういう方法を考慮していない。

ここで提示する検定方法を定式化する。s は整数値をとり、$f(s)$ は信号を意味する関数とする。ここで、ある $\underline{s}(<0)$, $\bar{s}(>0)$ に対して、$s<\underline{s}$, $s>\bar{s}$ ならば $f(s)=0$ が成り立つとする。図4.5の例では $-\underline{s}=\bar{s}=4$ 、$-4 \le s \le -1$ のとき $f(s)=-5-s$ 、$s=0$ のとき $f(s)=0$ 、$1 \le s \le 4$ のとき $f(s)=5-s$ となっている。ある未知のパラメータ k と信号の大きさを意味するパラメータ $d>0$ に対して

$$(6)\qquad Y(t) = g(t \mid a,b,c) + d \cdot f(t-k) + \varepsilon(t), \quad \varepsilon(t) \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$$

が成り立つという仮説を変化ありの対立仮説とし、(6) で $d \cdot f(t-k)$ をとったもの、つまり (4) が成り立つという仮説を変化なしの帰無仮説とする。k は正確な変化点の位置であるとはいえないが、ある程度それを定めるので、以降変化点パラメータと呼ぶことにする。そして、信号の形や統計的な意味を考え、$\underline{s}+1 \le k \le m-\bar{s}$ が満たされるものとする。

まず a, b, c, σ が既知のケースを考える。$\max(A,0)$ を A^+ と記せば、変化点を k としたときの d の最尤推定量は

$$(7)\qquad \hat{d}_k = \Big[\sum_{s=\underline{s}}^{\bar{s}} f(s)\{Y(k+s) - g(k+s \mid a,b,c)\} \Big/ \sum_{s=\underline{s}}^{\bar{s}} f(s)^2\Big]^+$$

であるから、尤度比統計量は

$$
\begin{aligned}
(8)\qquad & \max_{\underline{s}+1\le k\le m-\bar{s}}[-\sum_{t=1}^{m}\{Y(t)-g(t\,|\,a,b,c)-\hat{d}_k f(t-k)\}^2/\sigma^2]+\sum_{t=1}^{m}\{Y(t)-g(t\,|\,a,b,c)\}^2/\sigma^2 \\
& =(\max_{\underline{s}+1\le k\le m-\bar{s}}[\sum_{s=\underline{s}}^{\bar{s}} f(s)\{Y(k+s)-g(k+s|a,b,c)\}]^+)^2/\{\sigma^2\sum_{s=\underline{s}}^{\bar{s}} f(s)^2\}
\end{aligned}
$$

となる。帰無仮説のもとで、この統計量は

$$
(9)\qquad \{\max_{\underline{s}+1\le k\le m-\bar{s}} Z(k)^+\}^2 \;\equiv\; [\max_{\underline{s}+1\le k\le m-\bar{s}}\{\sum_{s=\underline{s}}^{\bar{s}} f(s)\varepsilon(k+s)\}^+]^2/\{\sigma^2\sum_{s=\underline{s}}^{\bar{s}} f(s)^2\}
$$

を意味する。$\sum_s f(s)\varepsilon(t+s)$ は信号がなければ 0 に近い値をとり、$t=k$ に信号があれば明らかに $\sum_s f(s)\varepsilon(k+s)$ が大きくなると考えられるので、直感的にも妥当な統計量といえる。

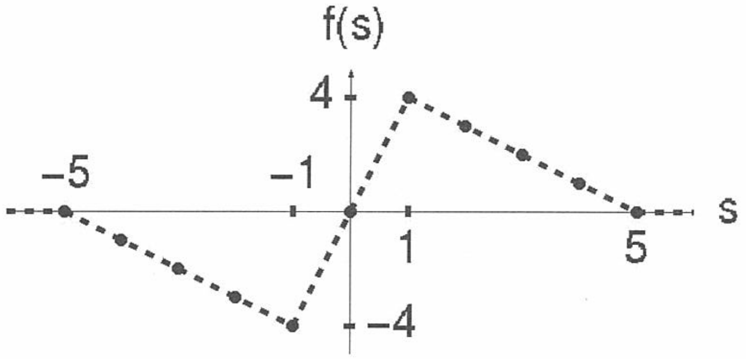

図4. 5　ノイズに含まれる信号

a, b, c, σ が未知のとき、漸近論を考えれば、尤度比統計量は $\{\max_{\underline{s}+1\le k\le m-\bar{s}} Z(k)^+\}^2$ に分布収束する。従ってその場合も p 値の評価をする際はその分布を用いればよいことになる。$Z(k)$ は正規確率過程であり、その最大値の分布関数のタイトな上限は Worsley (1982) の改良 Bonferroni 法により得られる。これより、尤度比統計量の実現値 z_T に対する p 値は

$$
(10) \quad \begin{aligned} & \mathrm{P}[\{\max_{\underline{s}+1\le k\le m-\bar{s}} Z(k)^{+}\}^2 > z_T] \;=\; \mathrm{P}\{\max_{\underline{s}+1\le k\le m-\bar{s}} Z(k) > \sqrt{z_T}\} \\ & \le\; (m+\underline{s}-\bar{s})\{1-\Phi(\sqrt{z_T})\} - (m+\underline{s}-\bar{s}+1)\int_{\sqrt{z_T}}^{\infty}\int_{\sqrt{z_T}}^{\infty} h(x,y\,|\,r)\mathrm{d}x\mathrm{d}y \end{aligned}
$$

と評価できる。ここで、

$$
(11) \qquad r \;\equiv\; \sum_{s=\underline{s}}^{\bar{s}-1} f(s)f(s+1)/\sum_{s=\underline{s}}^{\bar{s}} f(s)^2, \qquad h(x,y\,|\,r) \;\equiv\; \frac{1}{2\pi\sqrt{1-r^2}}\exp\{\frac{x^2-2rxy+y^2}{2(r^2-1)}\}
$$

であり、$\Phi(z)$ を標準正規分布の分布関数としている。これより保守的な検定が構成できる。

図4. 2 のデータに対してここで提示した変化検知の検定統計量の p 値を評価する (タイトな上限を求める) と、(a) のデータでは 0.95%、(b) のデータでは 0.11% となり、変化の存在が強く示唆された。変化点パラメータ k の最尤推定量はともに 11 となるため、この年の数年前に変化をもたらすなんらかの影響があったものと考えられる。一方、(c) のデータでは p 値が 86.3% (k の推定量は 34)、(d) のデータでは p 値が 80.0% (k の推定量は 20) と評価されたため、変化があるとはいえないことになる。つまり 4 本通して考えると変化はすべての林木に共通して発生しているわけではないということが分かる。施業履歴がないために断言はできないが、変化の時期や変化の発生の仕方を考えると変化が観察された林木の近くで除間伐があったことが予想される。

結論

以上の結果より、ここで提示した統計検定方法を用いれば、原因は何であれ、成長曲線における変化点の有無・時期の探索が可能になることがわかる. 現在、我が国においては、森林の保育の遅れから間伐の促進、整備が急務となっているが、管理の立場からすれば、将来の森林の状態を見据えるためにも、そうした人為的な行為がどのように森林の成長に影響を与えるのかを統計的に明らかにする必要がある. そのためにまず、除間伐の行為により成長過程に変化が発生しているのか否かを統計的に捉え、次の段階として、除間

伐の強度と変化の量との関係を明らかにして行く必要がある．それら一連の分析により、森林管理を目的とした新たな成長モデルの構築が可能となるものと期待できる．また、そこで得られるであろう定量的な情報は、森林資源による炭素固定量の把握など様々な場面で有効に利用できるものと考える．

引用文献

Carlstein, E. (1988). Nonparametric change-point estimation. *Ann. Statist.*, 16:188-197.

Chen, J. & Gupta, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *J. Amer. Statist. Assoc.* , 92:739-747.

Chen, J. & Gupta, A. K. (2000). *Parametric Statistical Change Point Analysis.* Birkhäuser.

Chernoff, H. & Rubin, H. (1956). The estimation of the location of a discontinuity in density. *Proc. 3rd Berkley Symp. Math. Statist. Probab.* , 1:19-37.

Cobb, G. W. (1978). The problem of the nile: conditional solution to a change-point problem. *Biometrika*, 62:243-251.

Csörgő, M. & Horváth, L. (1996). *Limit Theorems in Change-Point Analysis.* John Wiley & Sons.

Davis, R. A. , Huang, D. & Yao, Y. C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist.* , 23:282-304.

Gombay, E. & Horváth, L. (1990). Asymptotic distributions of maximum likelihood tests for change in the mean. *Biometrika*, 77:411-414.

Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.* , 21:671-680.

James, B. , James, K. L. & Siegmund, D. (1987). Tests for a change-point. *Biometrika*, 74:71-83.

Jarrett, R. G. (1979). Time intervals between coal mining disasters. *Biometrika*, 66:191-193.

Loader, C. R. (1996). Change point estimation using nonparametric regression. *Ann.*

Statist. , 24:1667-1678.

Richards, F. J. (1958). A flexible growth function to empirical use. *J. Exp. Bot.* , 10:290-300.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* , 29:595-601.

Worsley, K. J. (1982). An improved Bonferroni inequality and applications. *Biometrika*, 70:455-464.