

(vi)

$$y = h(x)$$

$$y_{t+\Delta t} = \phi(y_t)y_t + \theta(y_t)w_{t+\Delta t}$$

$w_t$  は離散時間単位白色雑音.

$$\phi(y_t) = e^{K(y_t)\Delta t}$$

$$\theta(y_t) = \sqrt{\{(e^{2K(y_t)\Delta t} - 1) / 2K(y_t)\}}$$

$$K(y_t) = \frac{1}{\Delta t} \log[1 + \{e^{J(y_t)\Delta t} - 1\} \alpha(y_t) / \{J(y_t)y_t\}]$$

$$J(y_t) = \left( \frac{\partial \alpha(y)}{\partial y} \right)_{y=y_t}$$

で (v) を近似する.

局所ガウス化法の一致性や安定性の詳細な議論は Ozaki (1983), Ozaki (1984), Ozaki (1985) を参照.

### 参 考 文 献

- Ozaki, T. (1983). Non-linear time series models and dynamical systems, in *Handbook of Statistics*, Vol. 5 (E.J. Hannan et al ed.) North-Holland.
- Ozaki, T. (1984). Local Gaussian time series modelling of stochastic dynamical systems, to appear in *J. Appl. Prob.*
- Ozaki, T. (1985). Statistical identification of storage models with application to stochastics hydrology, to appear in *Water Resources Bulletin*, also in AWARA Monograph, *Time Series Analysis in Water Resources*.

### 継続調査による誤差の推定

統計数理研究所 柏 木 宣 久・岸 野 洋 久

内閣支持, 政党支持等, 社会的意識の動向を, 支持率といった形で知ろうとする標本調査が良く行なわれている. これ等の調査結果の解析では——解釈が複雑になるとの危惧から誤差が全く無視されるのも稀ではない——単純標本抽出という仮定の下での標本誤差のみ評価され, 実際の標本抽出法に対応した標本誤差, あるいは観測誤差等は評価しないのが通常である. 「繰り返しのない実験によっては誤差を評価し得ない」とした R.A. Fisher の指摘どおり, 繰り返し調査の行なわれない社会調査に於いては, 仮定 (モデル) だけに頼った決定論的な誤差評価しか考慮されてこなかったのも無理からぬ所ではある.

とはいえ, こうした現状に甘んじるわけにはいかない. 誤差の過少評価が導く危険について, 今更説明の必要もなからう. 推計学的なデータ解析を志すのであれば, データによって誤差を評価したいと望むのは自然の成り行きといえよう.

誤差をデータから評価する為には, Fisher の指摘どおり, 調査を繰り返し行なえば良い. 無論, 問題がある. その第一は, 繰り返し調査を行なう為の費用の問題である. 支払う費用に見合った情報が得られない限り, 余分な調査の為に資金は提供されないのが通常である. 資金を引き出すに足る情報を獲得できるよう, 調査を設計しなければならない. これは次の第二の問題と深く関係する.

第二の問題は, いかに繰り返しを実現したら良いかという点である. 一言で “繰り返し” と

は言っても、その実現には二つの方法がある。第一が、英語で表現すれば、repetitionであり、第二がreplicationである。まず、第一の方法について考えてみよう。

説明するまでもないが、repetitionとは同一の環境の下での繰り返しである。環境の一様性が不可欠である。ところで、社会調査で取り扱う問題では、時間が変化すれば環境も変化すると考えられる場合が多い。それ故、repetitionを実現するには同一時点に於いてのみ可能となる。但し、それが故にrepetitionによって得られるのは、調査技術に付随した観測誤差に関する情報のみとなる。無論、観測誤差の評価はそれだけで十分重要的な仕事ではある。観測誤差の評価と、従ってその除去の為に、今まで多くの努力が払われてきたし、今後とも努力が為されていくであろう。

一方、こうした努力によって除去できるかもしれない誤差とは別に、いかなる努力によっても除去し得ない誤差も存在する。それは必然だけでは説明つかない環境の揺らぎに由来する誤差である。

環境の揺らぎを説明するのは難しい。神による必然（気紛れ）をどこまで容認するかによって、その存在すら否定する立場も考えられる。多分に哲学的な問題を含んでいる。但し、人間の立場で考えるならば、総てを必然として把握できるはずがない。環境は揺らいでいると考えるほうが実用的である。

環境の揺らぎに由来する誤差を知るには、環境の繰り返しが必要である。この繰り返しの実現について考えてみよう。まず、環境の一様性を必要とするrepetitionによっては実現できないのはすぐ分かる。また、同一時点に於ける環境の繰り返しの実現は、社会調査の場合、ほとんど非現実的である。そうすると結局残された手段は、時間軸方向での繰り返しと、そしてreplicationとなる。

実は、時間軸方向での調査の繰り返しは、社会調査では継続調査として既の実現されている。但し、その目的はあくまで必然としての環境の変化の把握であり、誤差をも評価しようとする考えはないようである。それ故当然ながら、継続調査とreplicationを結び付ける考えもないし、結果の尤もらしさを測る道具立ても用意されてこなかった。

本報告では、近接した時点の調査結果は似た傾向を示し易いという先験情報を継続調査に持ち込み、これをベイズモデルによって記述し、環境の揺らぎに由来する誤差をも含めた総体的な誤差評価を可能にする方法を提案する。こうした誤差評価が可能になるのは、先験情報の導入により、継続調査にreplicationを見出したからに他ならない。そして更に誤差評価ばかりでなく、要因配置実験の例からも明らかのように、replicationは獲得し得る情報の増大をもたらす。これが第一の費用の問題に貢献しないはずがない。

揺らぎの大きさと、各調査会社の相対的な調査特性の推定の為のモデルを、以下に簡単に記述する。

## 1. 揺らぎの大きさの推定の為のモデル

$$Y_i(T_k) = X_i(T_k) + V_i(T_k),$$

$$i = 1, 2, \dots, M, k = 1, 2, \dots, T,$$

$Y_i(T_k)$ :  $T_k$  時点に於ける  $i$  番目のカテゴリの占有率のデータ,

$X_i(T_k)$ :  $T_k$  時点に於ける  $i$  番目のカテゴリの占有率のトレンド,

$V_i(T_k)$ : 不規則成分,

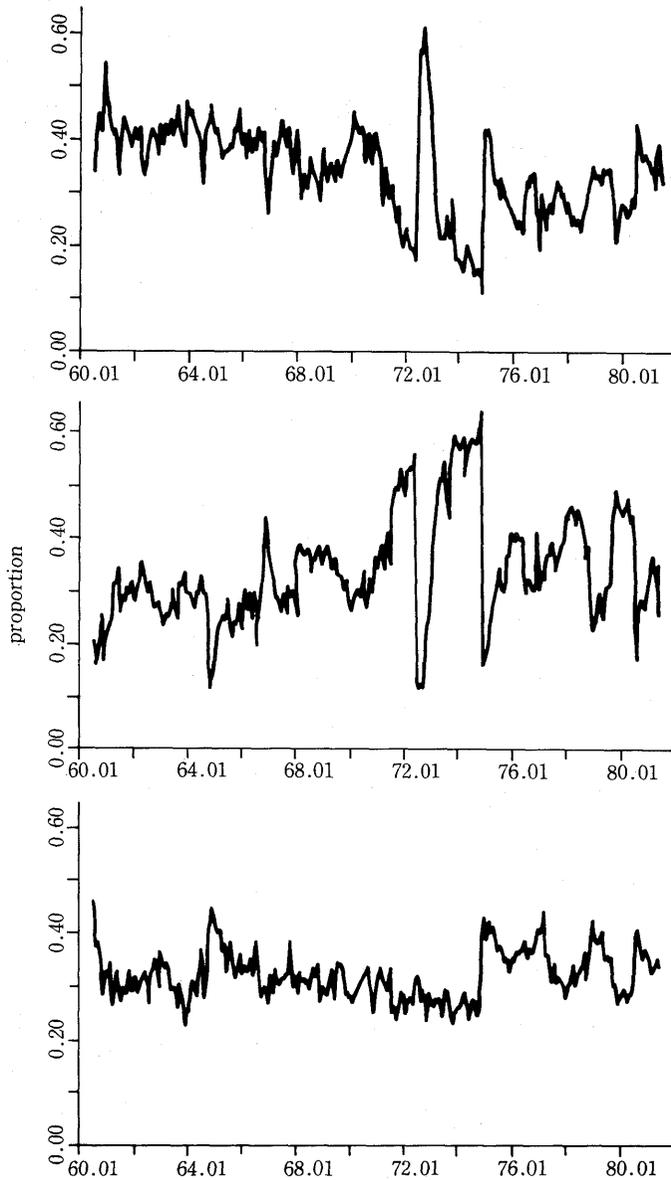


図1. Time-series data.

$$\begin{aligned}
 E[V_i(T_k)] &= 0, \\
 E[V_i^2(T_k)] &= \alpha X_i(Y_k)(1 - X_i(Y_k))/n(T_k), \\
 E[V_i(T_k)V_j(T_k)] &= -\alpha X_i(T_k)X_j(T_k)/n(T_k), \quad (i \neq j) \\
 E[V_i(T_k)V_j(T_l)] &= 0, \quad (k \neq l) \\
 i, j &= 1, 2, \dots, M, \quad k, l = 1, 2, \dots, T, \\
 \alpha &: \text{揺らぎ乗数}, \\
 n(T_k) &: T_k \text{ 時点に於ける標本の大きさ},
 \end{aligned}$$

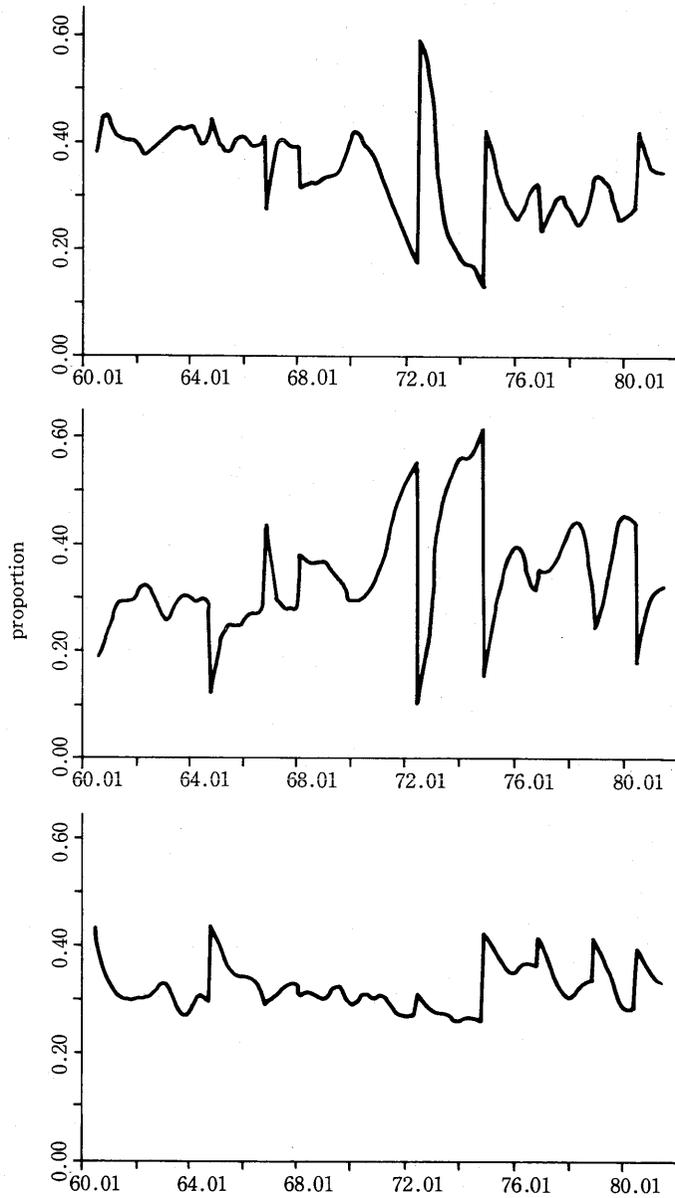


図2. Estimated trend.

$$X_i(T_k) = 2X_i(T_{k-1}) - X_i(T_{k-2}) + W_i(T_k),$$

$$i = 1, 2, \dots, M, k = 3, 4, \dots, T,$$

$W_i(T_k)$ : 不規則成分 (分散共分散構造は省略).

## 2. 相対的調査特性の推定の為のモデル

$$Y_{ci}(T_k) = X_i(T_k) + \beta_{ci} + V_{ci}(T_k),$$

$$\begin{aligned}
 &c=1, 2, \dots, S, i=1, 2, \dots, M, k=1, 2, \dots, T, \\
 &Y_{ci}(T_k): T_k \text{ 時点に於いて } c \text{ 番目の調査会社の得た } i \text{ 番目のカテゴリの占有} \\
 &\quad \text{率のデータ,} \\
 &\beta_{ci} \quad : c \text{ 番目の調査会社の } i \text{ 番目のカテゴリに対する相対的な調査特性,} \\
 &V_{ci}(T_k): \text{不規則成分 (分散共分散構造は省略),} \\
 &X_i(T_k)=2X_i(T_{k-1})-X_i(T_{k-2})+W_i(T_k), \\
 &i=1, 2, \dots, M, k=3, 4, \dots, T.
 \end{aligned}$$

本報告では、不規則成分の分散共分散構造として、単純標本抽出とした時の標本誤差に比例した構造を採用した。それは、単純標本抽出を基準として、揺らぎの大きさを測ろうとしたからである。本報告の分散共分散構造は一つの例に過ぎず、原則として、退化した分布の分散共分散構造であれば何でも構わないのは言うまでもない。本報告の本質的な部分は、揺らぎ乗数を導入し、揺らぎの大きさをフレキシブルに推定しようとする所にある。

以上のモデルのパラメータの推定及びトレンドの予測は、赤池のベイズ手順に沿って行なう。参考として、揺らぎの大きさの推定の為のモデルの適用例を、図にて掲載しておく。図1はデータで、上から、内閣を支持する、支持しない、わからないと答えた者の占有率の時系列を表わす。そして、図2が予測されたトレンドを表わす。この時、 $\alpha$ の推定値は3.49であった。

## 参 考 文 献

- Akaike, H. (1980). Likelihood and the Bayes procedure, *Trab. Estadist.*, **31**, 143-166.
- Kashiwagi, N. and Kishino, H. (1983). Estimation of the transition probabilities from the movement of the shares obtained by random sampling data, *Res. Memo.*, No. 254, The Inst. Statist. Math., Tokyo.
- Kashiwagi, N. and Kishino, H. (1984). Estimation of the trend and fluctuation of time-series data collected with sample surveys, *Res. Memo.*, No. 282, The Inst. Statist. Math., Tokyo.

## 生存時間分布関数の推定とベイズモデル

統計数理研究所 鎌 倉 稔 成

### 1. は じ め に

生存時間分析は、医学・薬学および信頼性工学の分野において、近年、急速に脚光を浴びつつある統計的手法の1つである。とりわけ、Cox (1972)の開発した比例ハザードモデルは、背景因子などの共変量と寿命分布の関係を評価する標準的な方法になりつつある。統計解析の大きな目的の1つは、解析の対象とする集団になんらかの意味で同質性を仮定して、集団全体を支配している確率法則である寿命分布に対する情報を得ることにある。寿命分布の推定は通常の分布関数の推定と同様に行われるが、大きく異なるのは、打ち切りを含む標本を対象とする点にある。経験分布に打ち切りデータを含む形で一般化したものにKaplan and Meier (1958)のProduct-limit estimateがある。これは普通Kaplan-Meierの推定量と呼ばれ、生存時間分析の基本となるものである。

今、寿命データが $n$ 個観測され、 $n$ 個のペア $(x_i, \delta_i)$  ( $i=1, \dots, n$ )で表されるものとすれ