研究開発
レター

欠損データ推定に基づく不等間隔サンプリングデータに対するシステム同定法

正　員　足　立　修　一　（宇都宮大学）
非会員　丹　治　雅　之　（宇都宮大学）

A System Identification Method for Nonuniform Sampling Data Based on Missing Data Estimation

Shuichi Adachi, Member, Masayuki Tanji, Non-member (Utsunomiya University)

In this letter, a system identification method for nonuniform sampling data is proposed. First, nonuniform sampling data is transformed to uniform sampling data whose sampling period is smaller than the original data. Second, a system identification method with missing data estimation is applied to the transformed data, since the transformed data contains a lot of missing data, Finally, the effectiveness of the proposed method is examined though a numerical example.

キーワード：システム同定，不等間隔サンプリング，欠損データ推定

1. はじめに

ディジタル信号処理やシステム同定では，通常，等間隔でサンプリングされたディジタル信号を用いて処理が行われることを想定している。しかしながら，実問題では，等間隔ではなく不等間隔でサンプリングされたデータしか利用できない場合が多々ある。そこで，本研究では対象システムの入出力データが不等間隔でサンプリングされた場合に対するシステム同定法として，欠損データの推定に基づく実用的な方法を提案し，その有効性を数値実験例を通して明らかにする。

2. 欠損データ推定に基づく不等間隔サンプリングデータに対するシステム同定法

図 1 の上図のような不等間隔サンプリングデータが計測された場合を想定する。もし，それぞれのサンプリング周期が，最も短いサンプリング周期（T とする）の整数倍であると仮定できるなら，不等間隔サンプリングデータを図 1 の下図のように，サンプリング周期が T より長い部分に欠損データが含まれているデータであるとみなすことができる。厳密にこの T を決定することは難しいが，実用上は T の整数倍がサンプリング周期の公称値になるように選択することで対処できる。このように，本研究では不等間隔サンプリングデータを，欠損データを含む等間隔サンプリングデータであるとみなすことを提案する。

欠損データの取り扱いに関しては統計学の分野で研究されてきたが [1]，ここではシステム同定理論の枠組みで以下のように欠損データを推定する [2]。まず，同定対象として LTI (Linear Time Invariant) システム

$$y(t) = G(q)u(t) + H(q)w(t) \quad \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

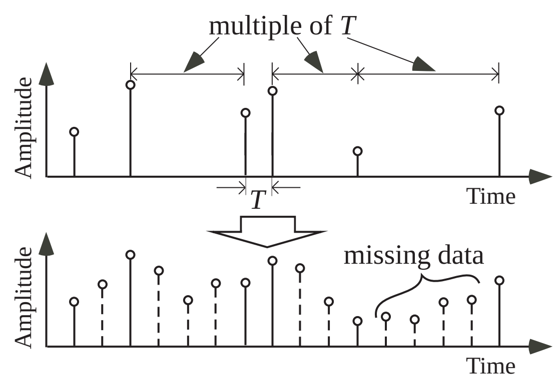

図 1　不等間隔サンプリングデータ（上図）を欠損データを含む等間隔サンプリングデータ（下図）とみなす。

Fig. 1.　Nonuniform sampling data (upper) can be considered that it contains missing data (lower).

を考える。ただし，$u(t)$ は入力，$y(t)$ は出力，$w(t)$ は白色雑音であり，$G(q), H(q)$ はそれぞれシステムと雑音の伝達関数である。

いま，時刻 $t_{u1}, t_{u2}, \cdots$ で入力が欠損し，時刻 $t_{y1}, t_{y2}, \cdots$ で出力が欠損すると仮定する。ここで，欠損データを未知パラメータとして扱うために，つぎのベクトルを定義する。

$$\boldsymbol{\eta} = [y(t_{y1}), y(t_{y2}), \cdots, u(t_{u1}), u(t_{u2}), \cdots]^T \quad \cdots (2)$$

すると，出力の 1 段先予測値 $\hat{y}(t|\boldsymbol{\theta})$ は，$G(q), H(q)$ のインパルス応答 g, h を用いて次式のように表される [2]。

$$\hat{y}(t|\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{k \in K_u} g(t-k, \boldsymbol{\theta}) u(k)$$

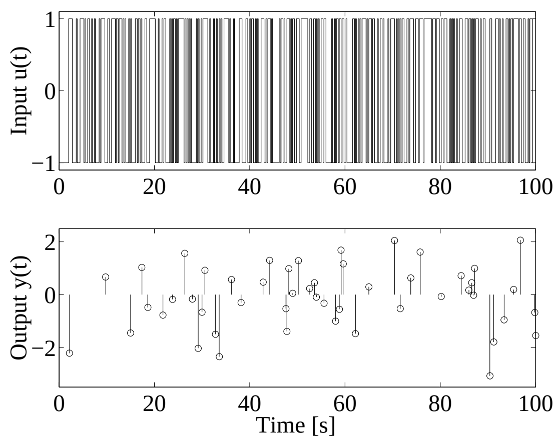

図 2　入出力データ

Fig. 2.　Input-Output data

$$+\sum_{k\in K_y} h(t-k,\boldsymbol{\theta})y(k)+\boldsymbol{\varphi}^T(t,\boldsymbol{\theta})\boldsymbol{\eta} \quad (3)$$

ただし，$\boldsymbol{\theta}$ はモデルの未知パラメータベクトルである。ここで，$k\in K_u, k\in K_y$ はそれぞれ入力，出力が欠損しなかった時刻 k の集合である。また，回帰ベクトル $\boldsymbol{\varphi}(t,\boldsymbol{\theta})$ は，入出力データが欠損した時刻 $t_{y1},t_{y2},\cdots,t_{u1},t_{u2},\cdots$ で作られる。ここで，(3) 式は，$\boldsymbol{\theta}$ が固定されたとき，欠損データ $\boldsymbol{\eta}$ に対して線形回帰となることに注意する。

以上より，$\boldsymbol{\theta}$ が一定ならば，最小 2 乗法によって $\boldsymbol{\eta}$ を推定できる。また，$\boldsymbol{\eta}$ が一定のときは，欠損データがない場合と同様の手順によって $\boldsymbol{\theta}$ を推定できる。したがって，$\boldsymbol{\eta}$ と $\boldsymbol{\theta}$ は，以下のような繰り返し手順で推定できる。

Step 1：欠損データ $\boldsymbol{\eta}$ の初期値を設定する。

Step 2：$\boldsymbol{\eta}$ を固定した状態で，最小 2 乗法によりモデルの未知パラメータ $\boldsymbol{\theta}$ を求める。

Step 3：求めたモデルから欠損データ $\boldsymbol{\eta}$ を最小 2 乗法で推定する（このとき $\boldsymbol{\theta}$ は固定しておく）。

Step 4：$\boldsymbol{\theta}$ と $\boldsymbol{\eta}$ が収束するまで，Step 2, 3 を繰り返す。

なお，この手順の収束性は理論的には保証されていないが，数値実験例では，適切な初期値から始めれば，ほとんどの場合収束することを確認している。

提案法の有効性を数値実験により検証しよう。まず，真のシステムを 12 次の FIR (Finite Impulse Response) モデル

$$y(t)=\sum_{j=1}^{12} b_j u(t-j)+w(t) \quad \cdots\cdots\cdots\cdots\cdots\cdots (4)$$

とする。ただし，$b_j=\exp(-(j-1)/4)$ とおいた。

このシステムに，サンプリング周期 $T_s=0.2$ s の擬似ランダム二値信号を入力したとき，出力データのサンプリング周期が不等間隔であったとする。ここで，入力は等間隔とする。本実験では，不等間隔サンプリングの出力データは，もともと 0.2 s の周期で得られた出力データの 90 % を欠損させて作成した。また，雑音 $w(t)$ は，$N(0,0.1^2)$ である正規性白色雑音とする。このときの入出力データを図 2 に示す。また，出力のサンプリング周期の度数分布を図 3 に示す。

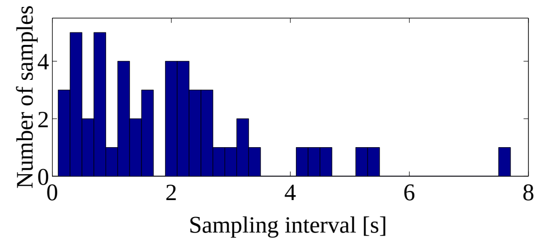

図 3　出力のサンプリング周期のヒストグラム

Fig. 3.　Histgram of output sampling intervals

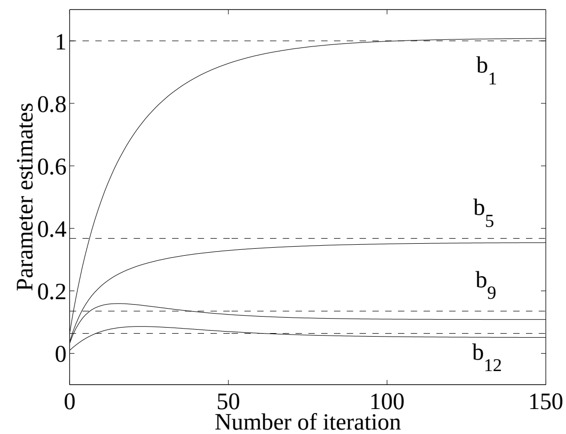

図 4　パラメータ推定値の収束の様子（実線：推定値，破線：真値）

Fig. 4.　Convergence property of unknown parameters : estimated value (solid line) and true value (dashed line)

提案法による未知パラメータ b_1,b_5,b_9,b_{12} の収束過程を図 4 に示す。図において，破線はパラメータの真値，実線は推定値を示す。これより，出力の 90 % が欠損しているにも関わらず，推定値は真値と近い値に収束していることがわかる。この結果から，提案法によって，不等間隔サンプリングデータからモデルを構築できることがわかった。

3. おわりに

本報告では，不等間隔サンプリングデータに対するシステム同定法として，欠損データ推定に基づく方法を提案し，その有効性を数値例によって明らかにした。提案法はヒューリスティックな方法であるが，数値実験結果から，90 % 近くのデータが欠損しても有効に動作することが確認されたことより，実問題に対しては有効な同定法になると考えられる。

（平成 14 年 2 月 21 日受付）

文　献

（1） R.Little and D.B.Rubin : Statistical Analysis with Missing Data (1987) John Woley & Sons

（2） L.Ljung : System Identification – Theory for the User (2nd Edition) (chapter 14) (1999) Prentice Hall