Akademie věd české republiky
ÚSTAV TEORIE INFORMACE A AUTOMATIZACE

Academy of Sciences of the Czech Republic
INSTITUTE OF INFORMATION THEORY AND AUTOMATION

RESEARCH REPORT

Ludvík Tesař

Identification of Parameters of AR(1)-process with Missing Data

No. 1847 August 1995

ÚTIA AV ČR P. O. Box 18, 182 08 Prague, Czech Republic
Telex: 122018 atom c, Fax: (+42)(2)6641 4903
Tel: (42)(2)66052337, E-mail: tesar@utia.cas.cz

Introduction

An autoregressive process is one of the most important processes studied in control design. It is used to model the most easy processes and it is also used for testing new estimators and controllers. To be able to analyze estimator, it is useful to know exactly the posterior probability density function.

When the whole sample of data is available, the solution is well known. The problem has a finite sufficient statistic and it is possible to compute the posterior probability density function by simple expression that uses this statistic. But when some data are missing, and there are some "gaps" in data sample, very little can be said about posterior p.d.f. in general.

This paper tries to deal with the problem of identification of posterior probability density function of parameters of an AR-process from incomplete data sample. The AR(1)-case is completely solved here. A sufficient statistic is constructed (limited by size of largest "gap" in data) that makes it possible to find an expression for posterior probability density function that is analogous to the complete data case.

In Chapter 2 the possibility of extending the presented method to AR(n) processes is discussed.

Chapter 1

AR(1) process

1.1 Assumptions

Assume an AR(1) process

$$X_t = \theta X_{t-1} + \varepsilon_t \tag{1.1}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, and X_t for $t = 1, 2, \ldots, T$ are real ($X_t \in \mathcal{R}$) random variables. We shall denote by x_t the realizations of them.

The problem solved here is computation of the posterior probability density function (p.d.f.) of the unknown parameter θ. We consider only certain data to be observed while the other "missing".

Indices of observed data are $\mathrm{OBS} \subset \{2, \ldots, T\}$, while $\mathrm{MIS} \subset \{2, \ldots, T\}$ are indices of missing data. We denote by x_{OBS} and x_{MIS} the observed and missing data respectively. The initial value x_1 is not included in data and its value is given. We assume, without loss of generality, that the last data x_T is observed ($T \in \mathrm{OBS}$).

Let us make clear that inference is here based on the Bayesian point of view. It means that θ is a random variable and we are interested in the posterior distribution

$$p(\theta \mid x_{\mathrm{OBS}}, x_1) \tag{1.2}$$

The standard method to estimate maximum posterior probability $\hat{\theta}$ is the EM-algorithm. We don't consider it here, because it provides only point estimate of θ and it can be rather slow, especially if more data are missing.

For the sake of completeness, the indices of observed data are assumed to be random and independent on value of data. This missing-data mechanism is called "missing at random". Further discussion on missing data mechanisms is in [2].

1.2 Main results

By Bayes rule we have:

$$p\left(\theta \mid x_{\mathrm{OBS}}, x_1\right) \propto p\left(\theta\right) p\left(x_{\mathrm{OBS}} \mid \theta, x_1\right) \tag{1.3}$$

where $\propto$ denotes proportionality without respect to normalization constant. $p(\theta)$ is prior information about θ. Under the assumption of independence of OBS and x_t as stated in previous section the likelihood function $p(x_{\mathrm{OBS}} \mid \theta, x_1)$ follows

$$p(x_{\mathrm{OBS}} \mid \theta, x_1) = \int_{\mathcal{R}^{|\mathrm{MIS}|}} p(x_{\mathrm{OBS}}, x_{\mathrm{MIS}}, x_1 \mid \theta, x_1)\, dx_{\mathrm{MIS}} \tag{1.4}$$

Applying the chain rule to $p(x_{\mathrm{OBS}}, x_{\mathrm{MIS}} \mid \theta, x_1)$ we obtain

$$p(x_{\mathrm{OBS}}, x_{\mathrm{MIS}} \mid \theta, x_1) = \prod_{k=2}^{T} p(x_k \mid x_{k-1}, x_{k-2}, \ldots, x_1, \theta) \tag{1.5}$$

Under the AR(1) assumption

$$p(x_k \mid x_{k-1}, \ldots, x_2, x_1, \theta) = p(x_k \mid x_{k-1}, \theta) \tag{1.6}$$

the equation (1.5) implies

$$p(x_{\mathrm{OBS}}, x_{\mathrm{MIS}} \mid \theta, x_1) = \prod_{k=2}^{T} p(x_k \mid x_{k-1}, \theta) \tag{1.7}$$

It follows from (1.1) that

$$p(x_k \mid x_{k-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_k - \theta x_{k-1})^2} \tag{1.8}$$

The following theorem states main results of the report

Theorem 1.1 *Denote by $\mathcal{D}$ the set $\mathrm{OBS} \cup \{1\} \setminus \{T\}$. Let $\mathrm{GAP}(i)$ be the length of a missing-data sequence after x_i, $i \in \mathcal{D}$ ($x_i, x_{i+\mathrm{GAP}(i)+1} \in \mathrm{OBS}$ and $x_{i+1}, \ldots, x_{i+\mathrm{GAP}(i)} \in \mathrm{MIS}$). Let $\mathcal{G}$ be the set of all sizes of missing-data sequences for every i. Literally*

$$\mathcal{G} = \{\ell \in \mathcal{R} \mid \exists i \in \mathcal{D} : \ell = \mathrm{GAP}(i)\} \tag{1.9}$$

Let $\mathcal{D}_\ell$ be the subset of $\mathcal{D}$

$$\mathcal{D}_\ell = \{i \in \mathcal{D} \mid \ell = \mathrm{GAP}(i)\} \tag{1.10}$$

We introduce the statistic S_ℓ as follows

$$S_\ell = \sum_{j \in \mathcal{D}_\ell} \begin{bmatrix} x_{j+\mathrm{GAP}(j)+1}^2 & x_{j+\mathrm{GAP}(j)+1} x_j \\ x_{j+\mathrm{GAP}(j)+1} x_j & x_j^2 \end{bmatrix} \tag{1.11}$$

Then the likelihood is

$$p(x_{\mathrm{OBS}} \mid \theta, x_1) \propto \prod_{\ell \in \mathcal{G}} \left\{ \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-\frac{\ell |\mathcal{D}_\ell|}{2}} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-1} \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix}^T S_\ell \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix} \right] \right\} \tag{1.12}$$

where $|\,.\,|$ *denotes number of elements of a set. Consequently the posterior p.d.f. is*

$$p(\theta \mid x_{\mathrm{OBS}}, x_1) \;\propto\; p(\theta) \prod_{\ell \in \mathcal{G}} \left\{ \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-\frac{\ell|\mathcal{D}_\ell|}{2}} \exp\left[-\frac{1}{2\sigma^2} \right. \right.$$
$$\left. \left. \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-1} \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix}^T S_\ell \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix} \right] \right\} \tag{1.13}$$

Proof.
The outline of the the proof is as follows.

1. The product $\prod_{k=2}^{T} p(x_k \mid x_{k-1}, \theta)$ in equation (1.7) will be splitted into products $\prod_{k=j}^{j+\ell} p(x_k \mid x_{k-1}, \theta)$ where $x_{j-1}, x_{j+\ell}$ are observed and all data between are missing

$$p(x_{j+\ell}, x_{j+\ell-1}, \ldots, x_j \mid x_{j-1}, \theta) \;=\; \prod_{k=j}^{j+\ell} p(x_k \mid x_{k-1}, \theta) \tag{1.14}$$

2. The expression (1.14) will be integrated according to $x_j, x_{j+1}, \ldots, x_{j+\ell-1}$ in order to obtain

$$p(x_{j+\ell} \mid x_{j-1}, \theta) \;=\; \int_{\mathcal{R}^\ell} \prod_{k=j}^{j+\ell} p(x_k \mid x_{k-1}, \theta) dx_j \; dx_{j+1} \ldots dx_{j+\ell-1} \tag{1.15}$$

3. Next we put integrated "pieces" together. The densities (1.15) define all together (1.4)

$$p(x_{\mathrm{OBS}} \mid \theta, x_1) \;=\; \prod_{j \in \mathcal{D}} \left[p(x_{j+\mathrm{GAP}(j)} \mid x_{j-1}, \theta) \right] \tag{1.16}$$

   for every $j \in \mathcal{D}$.

4. Finally by putting together pieces of the same size of "gap" ℓ there will be created statistics S_ℓ for every ℓ.

$$p(x_{\mathrm{OBS}} \mid \theta, x_1) \;=\; \prod_{\ell \in \mathcal{G}} \prod_{j \in \mathcal{D}_\ell} \left[p(x_{j+\ell} \mid x_{j-1}, \theta) \right] \tag{1.17}$$

<u>*Step 1*</u>
In case of AR(1) process the expression (1.14) can be then converted to

$$p(x_{j+\ell}, x_{j+\ell-1}, \ldots, x_j \mid x_{j-1}, \theta) \;=\; \exp \sum_{k=j}^{j+\ell} \log p(x_k \mid x_{k-1}, \theta) \tag{1.18}$$

Next we deal only with the sum. After substitution (1.8) in (1.18).

$$\begin{aligned}
\sum_{k=j}^{j+\ell} \log p(x_k \mid x_{k-1}, \theta) &= \\
&= \sum_{k=j}^{j+\ell} \left[-\frac{1}{2\sigma^2} (x_k - \theta x_{k-1})^2 - \frac{1}{2} \log(2\pi\sigma^2) \right] \qquad (1.19) \\
&= -\frac{1}{2\sigma^2} \sum_{k=j}^{j+\ell} \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}^T \begin{bmatrix} -1 \\ \theta \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix}^T \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix} - \frac{\ell+1}{2} \log(2\pi\sigma^2)
\end{aligned}$$

Introducting $(\ell+2) \times (\ell+1)$ matrix

$$L_\ell = \underbrace{\begin{bmatrix} -1 & & & & \\ \theta & -1 & & & \\ & \theta & -1 & & \\ & & & \ddots & \\ & & & & -1 \\ & & & & \theta \end{bmatrix}}_{\ell+1} \qquad (1.20)$$

we can rewrite (1.19)

$$\begin{aligned}
\sum_{k=j}^{j+\ell} \log p(x_k \mid x_{k-1}, \theta) &= \\
&= -\frac{1}{2\sigma^2} \begin{bmatrix} x_{j+\ell} \\ \vdots \\ x_{j-1} \end{bmatrix}^T L_\ell L_\ell^T \begin{bmatrix} x_{j+\ell} \\ \vdots \\ x_{j-1} \end{bmatrix} - \frac{\ell+1}{2} \log(2\pi\sigma^2) \qquad (1.21)
\end{aligned}$$

Now we permute the $\begin{bmatrix} x_{j+\ell} \\ \vdots \\ x_{j-1} \end{bmatrix}$ vector so to put $x_{j+\ell}$ and x_{j-1} at the beginning of a transformed vector. We denote by M_ℓ corresponding matrix permutation of L_ℓ.

$$M_\ell = \underbrace{\begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \theta \\ \theta & -1 & & & & \\ & \theta & -1 & & & \\ & & & \ddots & & \\ & & & & -1 & \\ & & & & \theta & -1 \end{bmatrix}}_{\ell+1} \qquad (1.22)$$

As a result we have

$$\sum_{k=j}^{j+\ell} \log p(x_k \mid x_{k-1}, \theta) =$$

$$= \quad -\frac{1}{2\sigma^2}\begin{bmatrix} x_{j+\ell} \\ x_{j-1} \\ x_{j+\ell-1} \\ x_{j+\ell-2} \\ \vdots \\ x_j \end{bmatrix}^T M_\ell M_\ell^T \begin{bmatrix} x_{j+\ell} \\ x_{j-1} \\ x_{j+\ell-1} \\ x_{j+\ell-2} \\ \vdots \\ x_j \end{bmatrix} - \frac{\ell+1}{2}\log(2\pi\sigma^2) \qquad (1.23)$$

The quadratic form can be put into a block-form $\begin{bmatrix} A_\ell \\ B_\ell \end{bmatrix}^T \begin{bmatrix} C_\ell \\ D_\ell \end{bmatrix} \begin{bmatrix} C_\ell \\ D_\ell \end{bmatrix}^T \begin{bmatrix} A_\ell \\ B_\ell \end{bmatrix}$ where

$$A_\ell = \begin{bmatrix} x_{j+\ell} \\ x_{j-1} \end{bmatrix} \qquad (1.24)$$

$$B_\ell = \begin{bmatrix} x_{j+\ell-1} \\ \vdots \\ x_j \end{bmatrix} \qquad (1.25)$$

$$C_\ell = \underbrace{\begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \theta \end{bmatrix}}_{\ell+1} \qquad (1.26)$$

$$D_\ell = \underbrace{\begin{bmatrix} \theta & -1 & & & & \\ & \theta & -1 & & & \\ & & & \ddots & & \\ & & & & -1 & \\ & & & & \theta & -1 \end{bmatrix}}_{\ell+1} \qquad (1.27)$$

The equation (1.23) with using substitutions (1.26) to (1.25) is

$$\sum_{k=j}^{j+\ell} \log p(x_k \mid x_{k-1}, \theta) = \qquad (1.28)$$

$$= \quad -\frac{1}{2\sigma^2}\left\{\begin{bmatrix} A_\ell \\ B_\ell \end{bmatrix}^T \begin{bmatrix} C_\ell \\ D_\ell \end{bmatrix} \begin{bmatrix} C_\ell \\ D_\ell \end{bmatrix}^T \begin{bmatrix} A_\ell \\ B_\ell \end{bmatrix}\right\} - \frac{\ell+1}{2}\log(2\pi\sigma^2) \qquad (1.29)$$

We apply *Lemma A.1* to (1.29)

$$\sum_{k=j}^{j+\ell} \log p(x_k \mid x_{k-1}, \theta) =$$

$$= \quad -\frac{1}{2\sigma^2}\left\{\left(B_\ell - \hat{B}_\ell\right)^T D_\ell D_\ell^T \left(B_\ell - \hat{B}_\ell\right)\right. \qquad (1.30)$$

$$\left. + A_\ell^T \left(C_\ell C_\ell^T - C_\ell D_\ell^T (D_\ell D_\ell^T)^{-1} D_\ell C_\ell^T\right) A_\ell\right\} - \frac{\ell+1}{2}\log(2\pi\sigma^2)$$

Step 2
The expression (1.14) in terms of (1.30) is integrated according to $x_2, \ldots, x_{\ell-1}$ in order to obtain (1.15). Using *Lemma A.3* it holds

$$p(x_{j+\ell} \mid x_{j-1}, \theta) = \int_{\mathcal{R}^\ell} \exp \sum_{k=j}^{j+\ell} \log p(x_k \mid x_{k-1}, \theta)\, dx_j \ldots dx_{j+\ell-1} \propto \tag{1.31}$$
$$\propto \det(D_\ell D_\ell^T)^{-\frac{\ell}{2}} \exp\left[-\frac{1}{2\sigma^2} A^T \left(C_\ell C_\ell^T - C_\ell D_\ell^T (D_\ell D_\ell^T)^{-1} D_\ell C_\ell^T\right) A\right]$$

Let us focus on the kernel of the quadratic form $C_\ell C_\ell^T - C_\ell D_\ell^T (D_\ell D_\ell^T)^{-1} D_\ell C_\ell^T$, and denote it by F_ℓ. After substitutions

$$P_\ell = \underbrace{\begin{bmatrix} -1 & & & & \\ \theta & -1 & & & \\ & & \ddots & & \\ & & & -1 & \\ & & & \theta & -1 \end{bmatrix}}_{\ell} \tag{1.32}$$

$$V_\ell = \left.\begin{bmatrix} \theta \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right\} \ell \tag{1.33}$$

we can write

$$F_\ell = C_\ell C_\ell^T - C_\ell \begin{bmatrix} V_\ell & P_\ell \end{bmatrix}^T (P_\ell P_\ell^T + V_\ell V_\ell^T)^{-1} \begin{bmatrix} V_\ell & P_\ell \end{bmatrix} C_\ell^T \tag{1.34}$$
$$= C_\ell C_\ell^T - C_\ell \begin{bmatrix} V_\ell & P_\ell \end{bmatrix}^T P_\ell^{-T} (I + (P_\ell^{-1} V_\ell)(P^{-1} V_\ell)^T)^{-1} \times$$
$$\times P_\ell^{-1} \begin{bmatrix} V_\ell & P_\ell \end{bmatrix} C_\ell^T \tag{1.35}$$

Since

$$C_\ell C_\ell^T = \begin{bmatrix} 1 & 0 \\ 0 & \theta^2 \end{bmatrix} \tag{1.36}$$

$$P_\ell^{-1} V_\ell = \begin{bmatrix} -\theta \\ -\theta^2 \\ -\theta^3 \\ \vdots \\ -\theta^\ell \end{bmatrix} \tag{1.37}$$

$$P_\ell^{-1} \begin{bmatrix} V_\ell & P_\ell \end{bmatrix} C_\ell^T = \begin{bmatrix} & 0 \\ & 0 \\ -P_\ell^{-1} V_\ell & \vdots \\ & 0 \\ & \theta \end{bmatrix} \tag{1.38}$$

it holds

$$F_\ell = \left[\begin{array}{cc} 1 - (P_\ell^{-1} V_\ell)^T H_\ell (P_\ell^{-1} V_\ell) & -(P_\ell^{-1} V_\ell)^T H_\ell \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \theta \end{array} \right] \\ -\left[\begin{array}{cccc} 0 & \cdots & 0 & \theta \end{array} \right] H_\ell (P_\ell^{-1} V_\ell) & \theta^2 - \left[\begin{array}{cccc} 0 & \cdots & 0 & \theta \end{array} \right] H_\ell \left[\begin{array}{c} 0 \\ \cdots \\ 0 \\ \theta \end{array} \right] \end{array} \right] \tag{1.39}$$

where $H_\ell = \left(I + (P_\ell^{-1} V_\ell)(P_\ell^{-1} V_\ell)^T\right)^{-1}$. The elements of the matrix

$$F_\ell = \left[\begin{array}{cc} F_{\ell(1,1)} & F_{\ell(1,2)} \\ F_{\ell(2,1)} & F_{\ell(2,2)} \end{array} \right] \tag{1.40}$$

should be furtherly simplified. The matrix H_ℓ, which occures in every element of F_ℓ

$$H_\ell \quad = \quad I - (P_\ell^{-1} V_\ell) \left[1 + (P_\ell^{-1} V_\ell)^T (P^{-1} V_\ell)\right]^{-1} (P_\ell^{-1} V_\ell)^T \tag{1.41}$$

Taking into account

$$1 + (P_\ell^{-1} V_\ell)^T (P_\ell^{-1} V_\ell) \quad = \quad \sum_{k=0}^{\ell} \theta^{2k} \tag{1.42}$$

we have

$$H_\ell \quad = \quad I - (P_\ell^{-1} V_\ell) \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-1} (P_\ell^{-1} V_\ell)^T \tag{1.43}$$

Next we simplify $H_\ell \, (P_\ell^{-1} V_\ell)$

$$\begin{aligned} H_\ell \, (P_\ell^{-1} V_\ell) \quad &= \quad (P_\ell^{-1} V_\ell) - (P_\ell^{-1} V_\ell) \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-1} \left(\sum_{k=1}^{\ell} \theta^{2k} \right) & (1.44) \\ &= \quad (P_\ell^{-1} V_\ell) \frac{1}{\sum\limits_{k=0}^{\ell} \theta^{2k}} & (1.45) \end{aligned}$$

The value of $F_{\ell(2,1)}$ is from (1.39), (1.45) and (1.37):

$$\begin{aligned} F_{\ell(2,1)} \quad &= \quad -\theta^{\ell+1} \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-1} & (1.46) \\ F_{\ell(1,2)} \quad &= \quad F_{\ell(2,1)} & (1.47) \end{aligned}$$

$F_{\ell(2,2)}$ follows from (1.43), (1.37) and (1.39)

$$F_{\ell(2,2)} \quad = \quad \theta^2 \theta^\ell \left(\sum_{k=0}^{\ell} \theta^{2k} \right)^{-1} \theta^\ell \tag{1.48}$$

$F_{\ell(1,1)}$ is (from *Lemma A.2*):

$$F_{\ell(1,1)} = 1 - \left(P_\ell^{-1} V_\ell\right)^T H_\ell \left(P_\ell^{-1} V_\ell\right) = \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \tag{1.49}$$

Summing up

$$\begin{aligned} F_\ell &= \begin{bmatrix} \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} & -\theta^{\ell+1}\left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \\ -\theta^{\ell+1}\left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} & \theta^{2(\ell+1)}\left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \end{bmatrix} \\ &= \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix} \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix}^T \end{aligned} \tag{1.50}$$

It is necessary to compute the determinant $\det(D_\ell D_\ell^T)$ to be the (1.31) completely computed. We assume

$$\det(D_\ell D_\ell^T) = \sum_{k=0}^{\ell} \theta^{2k} \tag{1.51}$$

and make proof by mathematical induction. The matrix $D_\ell D_\ell^T$ is $\ell \times \ell$, which can be easily computed from (1.27) is

$$D_\ell D_\ell^T = \begin{bmatrix} 1+\theta^2 & -\theta & & & & \\ -\theta & 1+\theta^2 & -\theta & & & \\ & -\theta & 1+\theta^2 & -\theta & & \\ & & & \ddots & & \\ & & & -\theta & 1+\theta^2 & -\theta \\ & & & & -\theta & 1+\theta^2 \end{bmatrix} \tag{1.52}$$

Validity of (1.51) for $\ell = 1$ and $\ell = 2$ is clear. Induction step can be done by computing determinant $\det(D_\ell D_\ell^T)$ across the first collumn and first row

$$\det(D_\ell D_\ell^T) = (1+\theta^2)\det(D_{\ell-1} D_{\ell-1}^T) - \theta^2 \det(D_{\ell-2} D_{\ell-2}^T) \tag{1.53}$$

Substituing to (1.53) for $\det(D_{\ell-1} D_{\ell-1}^T)$ and $\det(D_{\ell-2} D_{\ell-2}^T)$ we obtain (1.51) Finally from (1.31) and (1.50) and (1.51) it follows:

$$\begin{aligned} p(x_{j+\ell} \mid x_{j-1}, \theta) &\propto \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-\frac{\ell}{2}} \exp\Bigg\{ -\frac{1}{2\sigma^2} \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \\ &\quad \begin{bmatrix} x_{j+\ell} \\ x_{j-1} \end{bmatrix}^T \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix} \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix}^T \begin{bmatrix} x_{j+\ell} \\ x_{j-1} \end{bmatrix} \Bigg\} \\ &= \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-\frac{\ell}{2}} \exp\Bigg\{ -\frac{1}{2\sigma^2} \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \\ &\quad \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix}^T \begin{bmatrix} x_{j+\ell} \\ x_{j-1} \end{bmatrix} \begin{bmatrix} x_{j+\ell} \\ x_{j-1} \end{bmatrix}^T \begin{bmatrix} 1 \\ -\theta^{\ell+1} \end{bmatrix} \Bigg\} \end{aligned} \tag{1.54}$$

$$= \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-\frac{\ell}{2}} \exp\Bigg\{ -\frac{1}{2\sigma^2} \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \left[\begin{array}{c} 1 \\ -\theta^{\ell+1} \end{array}\right]^T \left[\begin{array}{cc} x_{j+\ell}^2 & x_{j+\ell}x_{j-1} \\ x_{j+\ell}x_{j-1} & x_{j-1}^2 \end{array}\right] \left[\begin{array}{c} 1 \\ -\theta^{\ell+1} \end{array}\right] \Bigg\}$$

<u>*Step 3 & 4*</u>
Now we can compute the (1.4) from (1.17) and (1.54):

$$p(x_{\mathrm{OBS}} \mid \theta, x_1) \propto \tag{1.55}$$
$$\propto \prod_{\ell \in \mathcal{G}} \left\{ \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-\frac{\ell|\mathcal{D}_\ell|}{2}} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1} \left[\begin{array}{c} 1 \\ -\theta^{\ell+1} \end{array}\right]^T S_\ell \left[\begin{array}{c} 1 \\ -\theta^{\ell+1} \end{array}\right] \right] \right\}$$

Now the equation (1.13) can be easily derived from (1.55) and Bayes rule (1.3).

1.3 Remarks

Remark 1 It can be seen that derived equation (1.13) is formally similar to classic posterior p.d.f. with complete data. In (1.13) there is product for every "gap"-size ℓ. Data matrix is replaced by matrices S_ℓ for every ℓ Every member of product includes expression $\left(\sum_{k=0}^{\ell} \theta^{2k}\right)^{-1}$, which equals to 1 for $\ell = 0$ and for $\ell \neq 0$ it distorts probability density function so that it is not Gaussian.

Remark 2 Size of resulting statistic depends on number of sizes of "gaps" (types of missing data pattern in subsequence). It is straightforward to pronounce this hypothesis: Is it possible to omit statistic S_ℓ for ℓ bigger then some limit and get precious estimates of posterior p.d.f.? The size of statistic would be then $4\times$limit. If this hypothesis is right, we would be able to compute estimate of posterior p.d.f. recursively without holding of whole sequence of data.

Chapter 2

Extension to AR(n) process

Equations (1.13) and (1.11) expresses posterior p.d.f. for AR(1) process. Let's now try to extend used method for obtaining posterior p.d.f. to AR(n) processes. Equation (1.7) would have changed so that:

$$p(x_{\mathrm{OBS}}, x_{\mathrm{MIS}} \mid \theta) = \prod_{k=n+1}^{T} p(x_k \mid x_{k-1}, \ldots, x_{k-n}, \theta) p(x_1 \mid \theta) \tag{2.1}$$

Definition of AR(n) process is:

$$p(x_k \mid x_{k-1}, \ldots, x_{k-n}, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_k - \sum\limits_{m=1}^{n} \theta_m x_{k-m})^2} \tag{2.2}$$

and now θ_m for $m \in \{1, \ldots, n\}$ are parameters. In AR(1) case we splitted the expression (1.7) into pieces (1.14) such that after integrating through missing data each such piece had disjunct piece of missing data. In general case of AR(n) process it is impossible. It seems that we must have an assumption that every piece of observed data must be of size at least n. Otherwise integrals could not be separated and separated statistic could not be constructed. For AR(1) process the sum

$$\sum_{k=2}^{T} \log p(x_k \mid x_{k-1}, \theta) \tag{2.3}$$

into (1.18). By integrating of exp of (1.18) through missing data we obtained $p(x_{j+\ell} \mid x_{j-1}, \theta)$. But in AR(2) case we need something like $p(x_{j+\ell+1}, x_{j+\ell} \mid x_{j-1}, x_{j-2}, \theta)$. I have not yet succeeded in simplifying of matrix expressions even in in AR(2) case so that statistic similar as (1.11) could be created. It seems that matrix similar like F_ℓ in (1.50) composed of multiplication of the vector and the same transposed vector doesn't exist.

Conclusion

In this research work was derived and proved expression for posterior probability-density function for sample of AR(1) process with missing data. Similar expression for AR(n) process was not proved. The possibility was pronounced to use limited statististic for recursive estimation of posterior prabability density function.

Appendix A

Lemmas for Matrices

In this appendix are several well known lemmas used to manipulating with matrices.

Lemma A.1 *For any corresponding matrices A, B, X, Y, Z such that matrix Z is square and invertible:*

$$\begin{bmatrix} A \\ B \end{bmatrix}^T \begin{bmatrix} X & Y^T \\ Y & Z \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = (B-\hat{B})^T Z (B-\hat{B}) + + A^T(X - Y^T Z^{-1} Y)A \tag{A.1}$$

where $\hat{B} = -Z^{-1}YA$. It can be rewritten as

$$\begin{bmatrix} A \\ B \end{bmatrix}^T \begin{bmatrix} C \\ D \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix}^T \begin{bmatrix} A \\ B \end{bmatrix} = = \left\{ \left(B-\hat{B}\right)^T DD^T \left(B-\hat{B}\right) \right\} + A^T \left(CC^T - CD^T(DD^T)^{-1}DC^T\right) A$$

for any suitable C, D.

Lemma A.2 ***Matrix inversion lemma***
For any rectangular matrix A and suitable square matrix B is equal

$$I - A^T(I + ABA^T)^{-1}A = (I + A^T BA)^{-1} \tag{A.2}$$

assuming that matrix inversions are possible and I is suitable square identity-matrix. Similarly for any rectangular matrix A and suitable square matrix B is

$$I + A^T(I - ABA^T)^{-1}A = (I - A^T BA)^{-1} \tag{A.3}$$

at the same assumptions.

Proves of both lemmas are in [3].

Lemma A.3 *Let A be $n \times n$ matrix, X and $\hat{X}$ column vectors (of size n), σ the real number. Then*

$$\int_{R^n} \exp\left[-\frac{1}{2\sigma^2}\left(X-\hat{X}\right)^T A \left(X-\hat{X}\right)\right] dx_1 \dots dx_n = = \left[(2\pi\sigma^2)\det(A)\right]^{-\frac{n}{2}} \tag{A.4}$$

Proof of the last lemma is also in [3].

Bibliography

[1] R. Kulhavý. Bayesian estimation, large deviations, and incomplete data. In *IEEE Conference on Decision and Control*, Orlando, FL, 1994. Submitted.

[2] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, New York, 1986.

[3] V. Peterka. Bayesian approach to system identification. In P. Eykhoff, editor, *Trends and Progress in System Identification*, chapter 8, pages 239–304. 1981.