# Parsimonious models of multivariate binary time series: statistical estimation and forecasting

Siarhei Shibalko
*Research Institute for Applied Problems of Mathematics and Informatics,*
Minsk, Belarus
shibalko2003@bk.ru

Yuriy Kharin
*Research Institute for Applied Problems of Mathematics and Informatics,*
Minsk, Belarus
kharin@bsu.by

***Abstract*—This paper is devoted to parsimonious models of multivariate binary time series. Consistent asymptotically normal statistical estimators for the parameters of proposed parsimonious models are constructed. Algorithms for statistical estimation of model parameters and forecasting of future states of time series are presented. Results of computer experiments on simulated and real statistical discrete-valued data are given.**

***Keywords*—binary time series, multivariate data, statistical estimation, parsimonious models, statistical forecasting**

## I. Introduction

The digitalization of the economy and the entire surrounding world leads to an increase of datasets in a discrete state space with discrete time. To mathematically describe such data, discrete, including binary, time series are used. Binary time series are used in modeling and data analysis of many economic and social processes. Examples of applied problems in statistical analysis of binary time series: in economics and finance, genetic sequence analysis, analysis of data flows in computer information security systems. Therefore, statistical analysis of multivariate discrete time series is an urgent task in mathematical and applied statistics [1, 2].

An universal model for description of high depth dependencies in discrete time series is a homogeneous Markov chain. Let $X_t$ be a *N*-dimensional homogeneous binary Markov chain (*N*-BMC) of order $s \geq 1$, defined on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$:

$$X_t = \begin{pmatrix} x_{t1} \\ \vdots \\ x_{tN} \end{pmatrix} \in V^N, t \in Z,$$

where $x_{ti} \in V = \{0,1\}$ − binary random variable specifying the *i*-th component at time *t*, $i = 1, \dots, N$.

## II. Parsimonious models

### A. Case of conditionally independent components

Consider the case of conditionally independent components under fixed prehistory:

$$\mathrm{P}\{X_t = J_t | X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\} =$$
$$\prod_{i=1}^{N} \mathrm{P}\{x_{ti} = j_{ti} | X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\},$$
$$J_t = (j_{ti}) \in V^S,$$

where the conditional probability distribution of the *i*-th bit under fixed prehistory can be represented as:

$$\mathrm{P}\{x_{ti} = j_{ti} | X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\} =$$
$$\begin{cases} p_i(J_{t-s}, \dots, J_{t-1}), & j_{ti} = 1, \\ 1 - p_i(J_{t-s}, \dots, J_{t-1}), & j_{ti} = 0. \end{cases}$$

Introduce a parsimonious model based on basis functions:

$$p = p(J_{1:s}) = F(\textstyle\sum_{k=1}^{m} b_k \psi_k(J_{1:s})), J_{1:s} \in V^{Ns}, \quad (1)$$

where $F(\cdot)$ − some given absolutely continuous distribution function, $B = (b_k) \in R^m$ − a column vector $m \leq 2^{Ns}$ unknown *N*-BMC coefficients, $\{\psi_k(J_{1:s})\}$ − basis functions, $J_{1:s} = (J_1', \dots, J_s')' \in V^{Ns}$ − composite column vector specifying the s-prehistory.

Introduce some assumptions on the function $F(\cdot)$:

- $0 < F(\cdot) < 1$,
- $F(\cdot)$ and $F^{-1}(\cdot)$ are twice continuously differentiable, (2)
- $F'(\cdot) \in (0; +\infty)$.

### B. Case of probability dependent components

Consider the case of probability dependent components under fixed prehistory $C_t = \{X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\}$ :

$$\mathrm{P}\{X_t = J_t | C_t\} = \mathrm{P}\{x_{t1} = j_{t1} | C_t\} \cdot \mathrm{P}\{x_{t2} = j_{t2} | x_{t1} = j_{t1}, C_t\}$$
$$\cdot \dots \cdot \mathrm{P}\{x_{tN} = j_{tN} | x_{t1} = j_{t1}, \dots, x_{t,N-1} = j_{t,N-1}, C_t\},$$

where the conditional distribution of the *i*-th bit under fixed prehistory can be represented in the following form:

first component:

$$\mathrm{P}\{x_{t1} = j_{t1} | C_t\} =$$
$$\begin{cases} p_1(J_{t-s}, \dots, J_{t-1}), & j_{t,1} = 1, \\ 1 - p_1(J_{t-s}, \dots, J_{t-1}), & j_{t,1} = 0. \end{cases}$$

other components:

$$\mathrm{P}\{x_{ti} = j_{ti} | x_{t1} = j_{t1}, \dots, x_{ti-1} = j_{ti-1}, C_t\} =$$
$$= \begin{cases} p_i(x_{t1}, \dots, x_{ti-1}, J_{t-s}, \dots, J_{t-1}), & j_{ti} = 1, \\ 1 - p_i(x_{t1}, \dots, x_{ti-1}, J_{t-s}, \dots, J_{t-1}), & j_{ti} = 0. \end{cases}$$

Similarly to (1), introduce a parsimonious model based on basis functions:

$$p = p(J_{1:s}) = F(\textstyle\sum_{k=1}^{m} b_k \psi_k(J_{1:s})), J_{1:s} \in V^{Ns}, \quad (3)$$

where $F(\cdot)$ − some fixed absolutely continuous distribution function, $B = (b_k) \in R^m$ − a column vector $m \leq 2^{Ns}$ unknown *N*-BMC coefficients, $\{\psi_k(J_{1:s})\}$ − basis functions, $J_{1:s} = (J_1', \dots, J_s')' \in V^{Ns}$ − composite column vector specifying the s-prehistory.

*C. Using artificial neural nets for basis function approximation*

In this case artificial neural nets are used to approximate basis function $\{\psi_k(J_{1:s})\}$ in models (1) and (3):

$$p = p(J_{1:s}) = F(\sum_{i=1}^{m} b_i F(\sum_{k=1}^{s} a_{ik} j_k)), J_{1:s} \in V^{Ns}, \quad (4)$$

where $F(\cdot), F_1(\cdot), \dots, F_m(\cdot)$ are some fixed absolutely continuous distribution functions (activation functions), $B = (b_i) \in R^m$ – a column vector $m$ unknown parameters, $A = (b_{ik}) \in R^{m*s}$.

This neural model can be realized as a 2-layer neural net with $s$ inputs, 1 output, $m$ neurons on the first layer and 1 neuron on the second layer [4].

## III. STATICAL ESTIMATION OF PARAMETRES

Let the observed time series of length $T$ is

$$X_{1:T} = (X_1, \dots, X_T) \in V^{TN}.$$

Using FBE (Frequencies-Based Estimation) method proposed in [5], we construct a statistical estimator for the parameter vector $B$ of models (1) and (3) based on the observed implementation $X_{1:T}$.

Construct consistent (at $T \to +\infty$) statistical estimators for transition probabilities:

$$\hat{p}(J_{1:s}) = \begin{cases} \dfrac{T-s}{T-s+1} \cdot \dfrac{v_{s+1}^T(J_{1:s};1)}{v_s^T(J_{1:s})}, J_{1:s} \in \boldsymbol{J}^{(s)}, \\ \dfrac{1}{2}, \qquad J_{1:s} \notin \boldsymbol{J}^{(s)}, \end{cases}$$

1) *Case of conditionally independent components*

$$v_s^T(J_{1:s}) = \sum_{t=s}^{T} \mathbf{1}\{X_t = J_t, \dots, X_{t-s} = J_{t-s}\},$$

$$v_{s+1}^T(J_{1:s};1) \sum_{t=s}^{T} \mathbf{1}\{x_{t+1,i} = 1, X_t = J_t, \dots, X_{t-s} = J_{t-s}\}, -$$

$s$-tuple frequencies $J_{1:s}$ and $(J_{1:s};1)$,

$$\boldsymbol{J}^{(s)} = \{J_{1:s} \in V^{Ns} : v_s^T(J_{1:s}) > 0\} \subseteq V^{Ns} -$$

subset of $s$-tuples with non-zero frequencies in $X_{1:T}$, $\mathbf{1}\{C\}$-indicator function of event $C$.

2) *Case of probability dependent components*

First component similarly to case 1), other components can be represented in the following form:

$$v_s^T(J_{1:s}) = \sum_{t=s}^{T-1} \mathbf{1}\left\{ \begin{matrix} x_{t+1,i-1} = j_{t+1,i-1}, \dots, x_{t+1,1} = j_{t+1,1}, \\ X_T = J_s, \dots, X_{t-s-1} = J_1 \end{matrix} \right\},$$

$$v_{s+1}^T(J_{1:s};1) \sum_{t=s}^{T} \mathbf{1}\left\{ \begin{matrix} x_{t+1,i} = 1, x_{t+1,i-1} = j_{t+1,i-1}, \dots, \\ x_{t+1,1} = j_{t+1,1}, X_T = J_s, \dots, X_{t-s-1} = J_1 \end{matrix} \right\}.$$

Introduce the following notation:

$$\hat{u}(J_{1:s}) = F^{-1}(\hat{p}(J_{1:s})) \in R^1,$$

$$D = \sum_{J_{1:s} \in J^{(s)}} \Psi(J_{1:s}) \Psi^T(J_{1:s}) \in R^{m \times m},$$

$$\Psi(J_{1:s}) = \{\psi_k(J_{1:s})\} \in R^{m \times 1},$$

$$E = \sum_{J_{1:s} \in J^{(s)}} \hat{u}(J_{1:s}) \Psi(J_{1:s}) \in R^{m \times 1}.$$

The idea of FBE method is to find $\hat{B}$ such that the function $p(J_{1:s})$ is close to $\hat{p}(J_{1:s})$ in $l_2$-metrics:

$$W(b) = \left\| \hat{u}(J_{1:s}) - \sum_{k=1}^{m} b_k \psi_k(J_{1:s}) \right\|^2 \to \min_b$$

Using the gradient:

$$\nabla W(b) = \sum_{J_{1:s} \in J^{(s)}} (-2\, \hat{u}(J_{1:s}) \Psi(J_{1:s}) + 2\, \Psi(J_{1:s}) \Psi^T(J_{1:s}) B)$$

we get the FBE-estimator:

$$\hat{B} = \left( \sum_{J_{1:s} \in J^{(s)}} \Psi(J_{1:s}) \Psi^T(J_{1:s}) \right)^{-1} \cdot \left( \sum_{J_{1:s} \in J^{(s)}} \hat{u}(J_{1:s}) \Psi(J_{1:s}) \right) = D^{-1} E. \quad (5)$$

**Theorem 1.** If the $N$-BMC is ergodic and the determinant of the matrix $D$ defined by (5) is $|D| \neq 0$, then the FBE estimate for models (1) and (3) has the form:

$$\hat{B} = (\widehat{b_k}) = D^{-1} E$$

and for $T \to +\infty$ is consistent, i.e. converges in probability to the true value $B^0$.

For the model A the bias and the variance of the estimator (5) are:

$$E\{\hat{B} - B^0\} = E\{\hat{B}\} - B^0 \xrightarrow[T\to\infty]{} 0,$$

$$E\left\{T(\hat{B} - B^0)(\hat{B} - B^0)'\right\} \xrightarrow[T\to\infty]{} D^{-1} \Psi \bar{F} (\Sigma_p)^{-1} \bar{F}' \Psi' (D^{-1})',$$

$$\Sigma_p = diag(p_i(1-p_i)) \in R^{2^s \times 2^s},$$

$$\bar{F} = \left(F^{-1'}(p_i)\right) \in R^{2^s}.$$

For the model B the bias and the variance of the estimator (5) are:

$$E\{\hat{B} - B^0\} = E\{\hat{B}\} - B^0 \xrightarrow[T\to\infty]{} 0,$$

$$E\left\{T(\hat{B} - B^0)(\hat{B} - B^0)'\right\} \xrightarrow[T\to\infty]{} D^{-1} \Psi \bar{F} (\Sigma_p)^{-1} \bar{F}' \Psi' (D^{-1})',$$

$$\Sigma_p = diag(p_i(1-p_i)) \in R^{2^{s+(i-1)} \times 2^{s+(i-1)}},$$

$$\bar{F} = \left(F^{-1\prime}(p_i)\right) \in R^{2^{s+(i-1)}}.$$

**Theorem 2.** Under the conditions of Theorem 1 and assumptions (2), the FBE estimator $\hat{B}$ for models (1) and (3) has an asymptotically normal distribution:

$$\sqrt{T}(\hat{B} - B) \xrightarrow[T\to\infty]{} \mathcal{N}_m(0, \Sigma),$$

$$\Sigma = D^{-1}\Psi\bar{F}\Sigma_p\bar{F}'\Psi'(D^{-1})'.$$

## IV. Statistical Forecasting

The substitution algorithm for optimal forecasting for one step is determined by the explicit expression:

$$\hat{x}_{ti} = \mathbf{1}\left\{F\left(\sum_{k=1}^{m}\widehat{b_k}\psi_k(X_{t-s}^{t-1})\right) - \frac{1}{2} > 0\right\} \quad (6)$$

Forecasting of $\hat{x}_{t+1\,i}$ for the next step is similarly to (6), only the fragment $X_{t-s}^{t-1} = (X_{t-1}, \dots, X_{t-s})$ is replaced by $(\hat{X}_t, \dots, X_{t-s+1})$, etc.

## V. Results of Computer Experiments

### A. Experiments with simulated data

We estimate the dependence of root mean square error for estimation of probability $p^0$ on number of observations $T$, number of basis functions $m$ and length of history $s$:

$$\widehat{\Delta_{p^0}}(T) = \frac{1}{M}\sum_{v=1}^{M}\left(\sum_{i=1}^{N}\sum_{J_{1:s}\in J^{(s)}}\left(p^0(J_{1:s}) - F\left(\sum_{k=1}^{m}\hat{b}_k\cdot\psi_k(J_{1:s})\right)\right)^2\right),$$

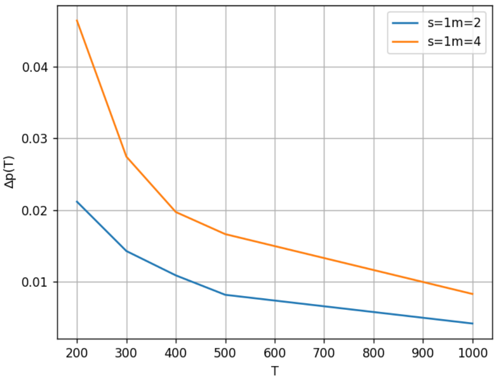


Fig. 1. Depence of root mean square error for estimation of probability $p^0$on number of observations $T$, number of basis functions $m$ and length of history $s$.

where $M$ is the number of simulated time series realizations. The results are illustrated by Fig. 1.

### B. Experiments with economic data

We used the exchange rate data $x_t = (x_{ti})$ for two currencies ($N = 2$): Russian ruble ($i = 1$) and U.S. dollar ($i = 2$) against the Belarusian ruble from 01/01/2020 to 01/12/2022 ($T$=1064). We analyzed increments in exchange rates: if the exchange rate increased by compared with yesterday, then the value of $x_{ti} = 1$ if the rate fell or remained the same, then the value $x_{ti} = 0$ ($i = 1, 2$).

We assessed the probability of correctness of the 1-step forecasting algorithm (6):

$$\hat{p}_i = \frac{1}{T-s-1}\sum_{t=s+1}^{T}\mathbf{1}\{\widehat{x_{ti}} = x_{ti}\}. \quad (7)$$

The results are illustrated by Fig. 2.

### C. Experiments with genome data

We took the complete Drosophila melanogaster genome of the length $T$=4000. Each nucleotide is represented as follows: A - ($x_{t1} = 0, x_{t2} = 0$), C - ($x_{t1} = 0, x_{t2} = 1$), G - ($x_{t1} = 1, x_{t2} = 0$), T - ($x_{t1} = 1, x_{t2} = 1$).

We assessed the probability of correctness of the 1-step forecasting algorithm (6):

$$\hat{p} = \frac{1}{T-s-1}\sum_{t=s+1}^{T}\mathbf{1}\{\widehat{x_{t1}} = x_{t1}, \widehat{x_{t2}} = x_{t2}\}.$$

The results are illustrated by Fig. 3.

### D. Experiments with medical data

We used a database $x_t = (x_{ti})$ of cases of notifiable diseases, and confirmation of pathogens, reported under the German 'Act on the Prevention and Control of Infectious Diseases in Man' (Infektionsschutzgesetz, IfSG) SurvStat@RKI 2.0 for two regions of Germany ($N = 2$): Berlin ($i = 1$) and Bavaria ($i = 2$) from 2001 to 2020 ($T$=1007). We examined the increase in the number of cases: if the number of cases increased in compared with yesterday, then the value $x_{ti} = 1$, if the number of cases has decreased or remained the same, then the value $x_{ti} = 0$ ($i = 1,2$).

We assessed the probability of correctness (7). The results are illustrated by Fig. 4.

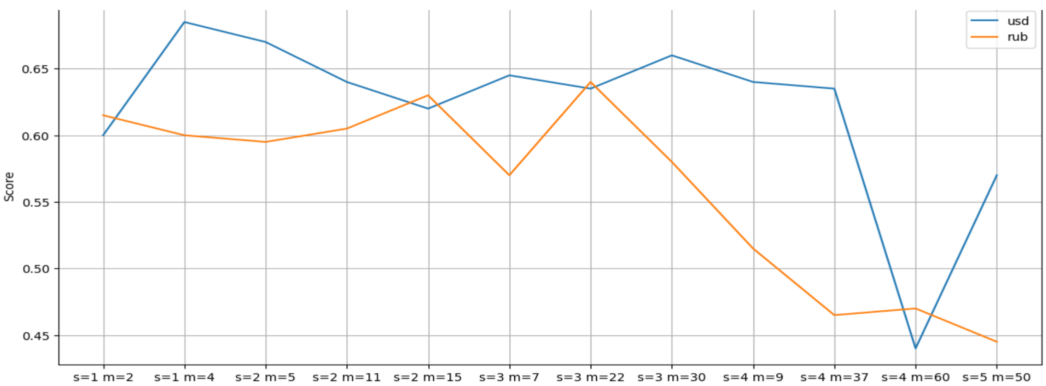


Fig. 2. Estimation of the probability of correctness of a 1-step forecast of exchange rates.

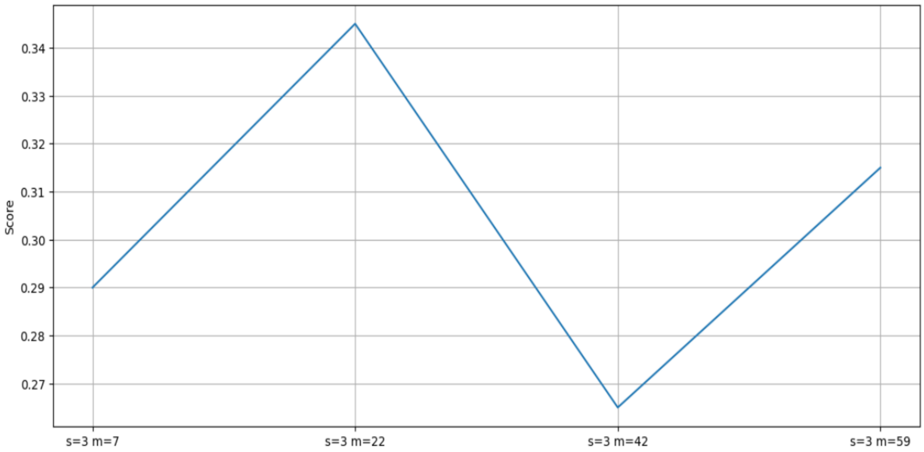


Fig.3. Estimation of the probability of correctness of a 1-step forecast of genetic sequence

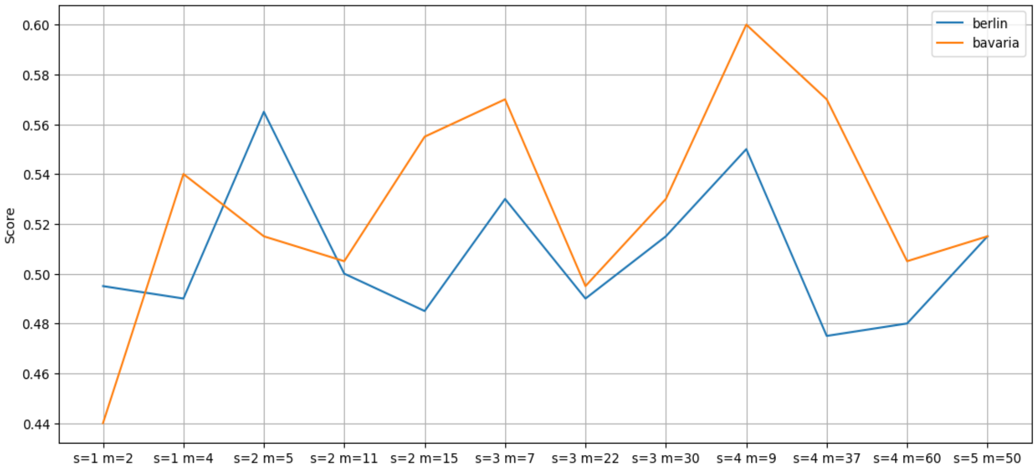


Fig.4. Estimation of the probability of correctness of a 1-step forecast of number of cases

## VI. Conclusion

The following results are obtained in the paper:

1) three parsimonious models of multivariate binary time series are proposed;

2) consistent asymptotically normal statistical estimators of the parameters for the proposed models are constructed;

3) asymptotic bias and variance for the parameters estimators are given;

4) algorithms of computer data analysis and experiments on simulated and real statistical discrete-valued data are presented.

The results of the paper can be used to solve the applied problems of statistical analysis of discrete-valued time series in economics, genetics and other fields.

## References

[1] Yu. Kharin, "Robustness in statistical forecasting," N. Y.: Springer, 2013.

[2] K. Fokianos, R. Fried, Yu. Kharin and V. Voloshko, "Statistical analysis of multivariate discrete-valued time series," Journal of Multivariate Analysis, 2021, vol. 186, art. 104805.

[3] Yu. Kharin and V. Voloshko, "Robust estimation for Binomial conditionally nonlinear autoregressive time series based on multivariate conditional frequencies," Journal of Multivariate Analysis, 2021, vol. 185, art. 104777.

[4] Yu. Kharin, "Neural net models of binomial time series in data analysis," Reports of National Academy of Sciences of Belarus, 2021, Vol. 65, No. 6, pp. 654-660.

[5] Yu. Kharin, V. A. Voloshko and E. A. Medved, "Statistical Estimation of Parameters for Binary Conditionally Nonlinear Autoregressive Time Series," Mathematical Methods of Statistics, 2018, Vol. 26. No. 2. pp. 103-118.