PLISKA
STUDIA MATHEMATICA

ПЛИСКА
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. **27** (2017), 23–36

PLISKA
STUDIA MATHEMATICA

STATISTICAL FORECASTING BASED ON BINOMIAL CONDITIONAL AUTOREGRESSIVE MODEL OF SPATIO-TEMPORAL DATA

Yuriy Kharin, Maryna Zhurak

Binomial conditional autoregressive model of spatio-temporal data is presented. Asymptotic properties of the maximum likelihood estimators of parameters for the binomial conditionally autoregressive model of spatio-temporal data are studied. Statistical tests on the values of true unknown parameters are constructed. Results of computer experiments on simulated and real data are given.

1. Introduction

Studying the probabilistic models of spatio-temporal data is a new topical scientific direction. Statistical analysis and modeling of spatio-temporal data is a challenging task [7]–[9]. Such models allows to model adequately processes taking into account both the dependence on time and the dependence on space.

Models based on spatio-temporal data become widely used for solving practical problems in meteorology, ecology, economics, medicine and other fields. In [10] spatio-temporal model is used to analyse daily precipitation for 71 meteorological stations over 60 years in Austria. Bayesian spatio-temporal model is applied to predict cancer cases in [2]. In [8] spatio-temporal models used to analyse forest stand data. In [9] authors present results of simulation studies and demonstrate the practical application of spatio-temporal processes in a study of radiation anomaly data.

2010 *Mathematics Subject Classification:* 62-07, 62M20, 62M30.

Key words: spatio-temporal data; vector Markov chain; maximum likelihood estimator; statistical hypotheses testing; optimal forecasting statistic.

2. Binomial conditional autoregressive model

Introduce the notation: $(\Omega, F, \mathbf{P})$ is the probability space; $\mathbf{N}$ is the set of positive integers; $\mathbf{Z}$ is the set of integers; $S = \{1, 2, \dots, n\}$ is the set of indexed spatial regions or space locations (let us agree to call them sites), into which the analyzed spatial area is partitioned; n is number of sites; $t \in \mathbf{Z}$ is discrete time; T is the length of observation period; $x_{s,t} \in A = \{0, \dots, N\}$ is a discrete random variable at time point t at site s; $F_{<t} = \sigma\{x_{u,\tau} : u \in S, \tau \leq t-1\} \subset F$ is the σ-algebra generated by the indicated in braces random variables; $z_{j,t} \in R^1 (j = 1, \dots, m)$ is an observed (known) level of the j-th exogenous factor at time point t which influences $x_{s,t}$; $L(\xi)$ means the probability distribution law of random variable ξ; $E\{\cdot\}$, $\mathbf{D}\{\cdot\}$, $\mathbf{cov}\{\cdot\}$ are symbols of expectation, variance, covariance of random variables; $Bi(\cdot; N, p)$ is the binomial probability distribution law of random variable ξ with the parameters $N \in \mathbf{N}$, $0 \leq p \leq 1$:

$$\mathbf{P}\{\xi = l\} = Bi(l; N, p) ::= C_N^l p^l (1-p)^{N-l}, l \in A; L\{\xi\} = Bi(\cdot; N; p). \tag{2.1}$$

We construct the binomial conditional autoregressive model for spatio-temporal data $\{x_{s,t}\}$ similar to [4], [7]. Provided that prehistory $\{x_{s,\tau} : s \in S, \tau \leq t-1\}$ is fixed, random variables $x_{1,t}, \dots, x_{n,t}$ are assumed to be conditionally independent and

$$L\{x_{s,\, t} | F_{<t}\} = Bi(\cdot; N; p_{s,\, t}), \tag{2.2}$$

$$\ln \frac{p_{s,t}}{1 - p_{s,t}} = \sum_{i=1}^{n} a_{s,i} x_{i,t-1} + \sum_{j=1}^{m} b_{s,j} z_{j,t}, t \in \mathbf{Z}, s \in S, \tag{2.3}$$

where $a_s = (a_{s,1}, \dots, a_{s,n})' \in R^n$, $b_s = (b_{s,1}, \dots, b_{s,m})' \in R^m$, $s \in S$, $\theta_s = \left(a_s', b_s'\right)' \in R^{n+m}$, $\theta = \left(\theta_1', \dots, \theta_n'\right)' \in R^{n(n+m)}$ is the composed vector of the parameters of the model; $p_{s,t}$ can be calculated as follows:

$$p_{s,t} = p_s\left(X_{t-1}, Z_t\right) ::= \exp\left\{\theta_s' Y_t\right\} \left(1 + \exp\left\{\theta_s' Y_t\right\}\right)^{-1},\ s \in S,\ t \in \mathbf{Z}, \tag{2.4}$$

where $Z_t = (z_{1,t}, z_{2,t}, \dots, z_{m,t})' \in R^m$ is the column vector specifying exogenous factors at time point t; $X_t = (x_{1,t}, x_{2,t}, \dots, x_{n,t})' \in A^n$ is the column vector specifying the time slice of the process under consideration at time point $t \in \mathbf{Z}$; $Y_t = \left(X_{t-1}', Z_t'\right)' \in R^{n+m}, t \in \mathbf{Z}$.

Let $L = \left\{l_j = (l_{1,j}, \dots, l_{n,j})' \in A^n :\ j = 1, 2,\ \dots, (N+1)^n\right\}$ be the ordered set of all admissible values of the vector X_t; $|L| = \nu = (N+1)^n$.

Theorem 2.1. *For the model* (2.2), (2.3) *the observed vector process X_t is the n-dimensional nonhomogeneous Markov chain with the finite state space L*

and the one-step transition probability matrix $Q(t) = (q_{I,J}(\theta, t))$, $I = (i_s)$, $J = (j_s) \in L$*:*

$$q_{I,J}(\theta, t) = \prod_{s=1}^{n} \frac{C_N^{j_s} \left(\exp\left\{a_s' I + b_s' Z_{t-1}\right\}\right)^{j_s}}{\left(1 + \exp\left\{a_s' I + b_s' Z_{t-1}\right\}\right)^N}. \tag{2.5}$$

Proof. Let us prove that vector process X_t for the model (2.2), (2.3) is the n-dimensional nonhomogeneous Markov chain. Using model assumptions (2.2), (2.3) and conditional independence of random variables $x_{1,t}, ..., x_{n,t} \in A$ in case of fixed σ-algebra $F_{<t} = \sigma\{x_{u,\tau} : u \in S, \tau \leq t-1\} \subset F$ we get

$$\mathbf{P}\left\{X_t = I_t \middle| X_{t-1} = I_{t-1}, \ldots, X_1 = I_1\right\} =$$

$$= \mathbf{P}\left\{x_{1,t} = i_{1,t}, \ldots, x_{n,t} = i_{n,t} \middle| X_{t-1} = I_{t-1}\right\} =$$

$$= \prod_{s=1}^{n} \mathbf{P}\left\{x_{s,t} = i_{s,t} \middle| X_{t-1} = I_{t-1}\right\} = \mathbf{P}\left\{X_t = I_t \middle| X_{t-1} = I_{t-1}\right\}.$$

Thus, the Markov property is valid, and X_t is the n-dimensional nonhomogeneous Markov chain. The one-step transition probability matrix $Q(t) = (q_{IJ}(t))$ is determined by taking into account (2.1), (2.3) and the model property of conditional independence:

$$q_{I,J}(t) = \mathbf{P}\left\{X_t = J \middle| X_{t-1} = I\right\} = \prod_{s=1}^{n} \mathbf{P}\left\{x_{s,\,t} = j_s \middle| X_{t-1} = I\right\} =$$

$$= \prod_{s=1}^{n} C_N^{j_s} p_s^{j_s}(I, Z_t)\left(1 - p_s(I, Z_t)\right)^{N - j_s}$$

$$= \prod_{s=1}^{n} C_N^{j_s} \left(\frac{p_s(I, Z_t)}{1 - p_s(I, Z_t)}\right)^{j_s} \left(1 - p_s(I, Z_t)\right)^N.$$

By condition (2.4) we have:

$$q_{I,J}(\theta, t) = \prod_{s=1}^{n} \frac{C_N^{j_s} \left(\exp\left\{a_s' I + b_s' Z_{t-1}\right\}\right)^{j_s}}{\left(1 + \exp\left\{a_s' I + b_s' Z_{t-1}\right\}\right)^N}.$$

that coincides with (2.5). □

Corollary 2.1. *Under conditions of Theorem* 2.1*, the matrix of transition probabilities* $H(t_1, t_2) = (h_{I,J}(t_1, t_2))$*,* $h_{I,J}(t_1, t_2) = \mathbf{P}\{X_{t_2} = J | X_{t_1} = I\}, I, J \in L$*, for* $t_2 - t_1$ *steps from time point* t_1 *to time point* t_2 $(t_1 < t_2, t_1, t_2 \in \mathbf{Z})$ *is:*

$$H(t_1,\, t_2) = Q(t_1 + 1)Q(t_1 + 2) \dots Q(t_2). \tag{2.6}$$

Corollary 2.2. *Under conditions of Theorem* 2.1*, if vector of exogenous factors* $Z_t = Z = (z_1, \dots, z_m)' \in R^m$ *does not depend on* t*, then the one-step transition probability matrix does not depend on* t*, and Markov chain* X_t *is homogeneous:*

$$Q = (q_{I,J}(\theta)) \in [0,1]^{\nu\times\nu}, \quad I, J \in L, \tag{2.7}$$

$$q_{I,J}(\theta) = \prod_{s=1}^{n} C_N^{j_s} \left(\exp\left\{a_s' I + b_s' Z\right\}\right)^{j_s} \left(1 + \exp\left\{a_s' I + b_s' Z\right\}\right)^{-N}, \tag{2.8}$$

the n*-dimensional Markov chain is ergodic, and single stationary probability distribution* $\pi = (\pi_I) \in [0,1]^\nu$ *exists:*

$$Q'\pi = \pi, \ \sum_{I \in A^n} \pi_I = 1.$$

Proofs of Corollaries 2.1, 2.2 of Theorem 2.1 are presented in [5].

Lemma 2.1. *For the model* (2.2)*,* (2.3) *the observed vector process* X_t *is nondegenerate process for any finite coefficients values* $\{\theta_s\}$ *and finite* $\{z_{i,t}\}$*, that means in* (2.4)*:*

$$0 < p_{s,t} < 1, \quad s \in S, t \in \mathbf{Z}.$$

Lemma 2.2. *For the model* (2.2)*,* (2.3) *in case of any finite values of coefficients* $\{\theta_s\}$ *and finite values* $\{z_{i,t}\}$ *the covariance matrix* $\mathbf{cov}\{X_t, X_t\}$ *is positively defined and takes the form:*

$$\mathbf{cov}\{X_t, X_t\} = N diag\left\{p_i(X_{t-1}, Z_t)(1 - p_i(X_{t-1}, Z_t))\right\} + D \in R^{n\times n},\ i \in S,$$

$$D = (d_{ij}) \in R^{n\times n}, d_{ij} = N^2 \mathbf{cov}\left\{(1 + \exp(-\theta_i Y_t))^{-1}, (1 + \exp(-\theta_j Y_t))^{-1}\right\}.$$

Proofs of Lemmas 2.1, 2.2 are presented in [6].

3. Statistical estimation of parameters

Theorem 3.1. *The log-likelihood function for the model* (2.2), (2.3) *under the observed spatio-temporal data* $\{X_t : \ t = 1, 2, \dots, T\}$ *takes the additive form:*

$$l(\theta) = \sum_{s=1}^{n} l_s(\theta_s),$$

(3.9)

$$l_s(\theta_s) = \sum_{t=1}^{T} \left(x_{s,t}\theta_s' Y_t - N \ln\left(1 + \exp\left\{\theta_s' Y_t\right\}\right) + \ln C_N^{x_{s,t}} \right).$$

P r o o f. Using the generalized formula for multiplying probabilities and properties of the Markov chain defined in Theorem 2.1 we construct the likelihood function:

$$L(\theta) = \mathbf{P}\left\{X_1, \ \dots, \ X_T\right\} = \mathbf{P}\{X_1\} \prod_{t=2}^{T} \mathbf{P}\left\{X_t \middle| \ X_{t-1}\right\}.$$

Expressions for $\mathbf{P}\{X_1\}$ and $\mathbf{P}\left\{X_t \middle| X_{t-1}\right\}$ are found in Theorem 2.1:

$$\mathbf{P}\{X_1\} = \prod_{s=1}^{n} C_N^{x_{s,1}} \left(\exp\left\{b_s' Z_1\right\}\right)^{x_{s,1}} \left(1 + \exp\left\{b_s' Z_1\right\}\right)^{-N},$$

$$\mathbf{P}\left\{X_t \middle| X_{t-1}\right\} = \prod_{s=1}^{n} C_N^{x_{s,t}} \left(\exp\left\{\theta_s' Y_t\right\}\right)^{x_{s,t}} \left(1 + \exp\left\{\theta_s' Y_t\right\}\right)^{-N}, \quad t \geq 2.$$

Then we get

$$L(\theta) = \prod_{t=1}^{T} \prod_{s=1}^{n} C_N^{x_{s,t}} \left(\exp\left\{\theta_s' Y_t\right\}\right)^{x_{s,t}} \left(1 + \exp\left\{\theta_s' Y_t\right\}\right)^{-N} =$$

$$= \prod_{s=1}^{n} \prod_{t=1}^{T} C_N^{x_{s,t}} \left(\exp\left\{\theta_s' Y_t\right\}\right)^{x_{s,t}} \left(1 + \exp\left\{\theta_s' Y_t\right\}\right)^{-N}.$$

Find the log-likelihood function:

$$l(\theta) = \ln L(\theta) = \ln \left(\prod_{s=1}^{n} \prod_{t=1}^{T} C_N^{x_{s,t}} \left(\exp\left\{\theta_s' Y_t\right\}\right)^{x_{s,t}} \left(1 + \exp\left\{\theta_s' Y_t\right\}\right)^{-N} \right) =$$

$$= \sum_{s=1}^{n} \sum_{t=1}^{T} \left(x_{s,t}\left(\theta_s' Y_t\right) - N \ln\left(1 + \exp\left\{\theta_s' Y_t\right\}\right) + \ln C_N^{x_{s,t}} \right) = \sum_{s=1}^{n} l_s(\theta_s),$$

that coincides with (3.9). □

To find the maximum likelihood estimators (MLE) $\{\hat{\theta_s}\}$ of the parameters of the model we need to maximize the log-likelihood function (3.9):

$$l(\theta) \to \max_{\theta \in R^{n(n+m)}} . \tag{3.10}$$

Theorem 3.2. *In case of the model* (2.2), (2.3), *if* $m = 1$, $z_{1,t} = z \neq 0$ *does not depend on* t *and Markov chain* $X_t \in L$ *is stationary, then for any finite coefficients values* $\{\theta_s\}$ *and finite* $z \in R^1$ *the Fisher information matrix is nonsingular block-diagonal matrix (here* $Y_t = (X'_{t-1}, z)'$*):*

$$G = N diag \left\{ E \left\{ Y_t Y_t' p_i(X_{t-1}, z)(1 - p_i(X_{t-1}, z)) \right\} \right\}, i = 1, \ldots, n. \tag{3.11}$$

Theorem 3.3. *Under Theorem* 3.2 *conditions, if* $T \to +\infty$ *the constructed by* (3.10) *maximum likelihood estimators* $\{\hat{\theta_s}\}$ *are consistent and asymptotically normally distributed:*

$$L\left\{ \sqrt{T}(\hat{\theta} - \theta^0) \right\} \to N_{n(n+1)} \left(0, G^{-1} \right). \tag{3.12}$$

where G is determined by (3.11).

Proofs of Theorems 3.2, 3.3 are presented in [6].

4. Statistical hypotheses testing

Theorems 3.1–3.3 are used to construct statistical tests for testing of hypotheses on the values of true unknown parameters $\{\theta_s^0\}$:

$$H_0 : \theta^0 = \theta^*;$$
$$H_1 = \overline{H_0} : \theta^0 \neq \theta^*.$$

where $\theta^* \in R^{n(n+1)}$ is some fixed (hypothetical) value of parameters. Let's consider the statistic:

$$g_T = g\left(X_1, \ldots, X_T\right) ::= T(\widehat{\theta} - \theta^*)' G (\widehat{\theta} - \theta^*) \geq 0, \tag{4.13}$$

where $\widehat{\theta}$ is the maximum likelihood estimator for model parameters, G is the matrix determined by (3.11).

Theorem 4.1. *Under Theorem* 3.2 *conditions, if hypothesis* H_0 *is true and* $T \to \infty$, *then statistic* g_T *is asymptotically chi-square distributed with* $n(n+1)$ *degrees of freedom:*

$$L_{H_0}\{g_T\} \to \chi^2_{n(n+1)}. \tag{4.14}$$

P r o o f. For proving the theorem we use the the asymptotic property (3.12) from Theorem 3.3. Let us consider the sequence of random vectors:

$$\xi_T = (\xi_{Tk}) = \sqrt{T} G^{1/2} \left(\hat{\theta} - \theta^* \right) \in R^{n(n+1)}. \tag{4.15}$$

If hypothesis H_0 is true, then by (3.12), (4.15) we get the convergence:

$$L_{H_0} \{\xi_T\} \to N_{n(n+1)}(0, I_{n(n+1)}),$$

that means

$$\xi_T \xrightarrow{D} \eta = (\eta_k) \in R^{n(n+1)}, \ L\{\eta\} = N_{n(n+1)}(0, I_{n(n+1)}), \tag{4.16}$$

where I_m is the identity $(m \times m)$-matrix. By (4.13), (4.15) the statistic g_T has the equivalent form:

$$g_T = T \left(\hat{\theta} - \theta^* \right)' G \left(\hat{\theta} - \theta^* \right) = \xi'_T \xi_T,$$

from which by (4.15) we have for $T \to \infty$:

$$g_T \xrightarrow{D} \eta' \eta = \sum_{k=1}^{n(n+1)} \eta_k^2.$$

Therefore, using the definition of the central χ^2-distribution, we come to the statement (4.14) of the theorem. □

Consider a family of decision rules based on statistic g_T and Theorem 4.1:

$$d = d\left(X_1, \ldots, X_T\right) = \begin{cases} 0, g_T < \Delta; \\ 1, g_T \geq \Delta, \end{cases} \tag{4.17}$$

where $\Delta > 0$ is some threshold value.

Corollary 4.1. *If threshold value Δ in (4.17) has the form:*

$$\Delta = F^{-1}_{\chi^2_{n(n+1)}} \left(1 - \alpha\right),$$

where $F_{\chi^2_{n(n+1)}}(\cdot)$ is the distribution function of χ^2 probability distribution with $n\left(n+1\right)$ degrees of freedom, then for $T \to \infty$ the asymptotic size of the test (4.17) is equal to the given significance level $\alpha \in (0, 1)$.

Theorem 4.2. *Under Theorem 4.1 conditions for the sequence of contiguous hypotheses $H_{1T} = \left\{\theta^0 = \theta^* + T^{-1/2} a\right\}$, where a is some nonzero vector from $R^{n(n+1)}$, if $T \to \infty$, then the test statistic g_T is asymptotically noncentral chi-square distributed with $n(n+1)$ degrees of freedom and noncentrality parameter $\Delta^2 = a' G a$:*

$$L_{H_{1T}}\{g_T\} \to \chi^2_{\Delta^2, n(n+1)}, \tag{4.18}$$

and the power of the test satisfies the asymptotics:

$$w_T = \mathbf{P}_{H_{1T}}\{d=1\} \to w^* = 1 - F_{\chi^2_{\Delta^2, n(n+1)}}\left(F^{-1}_{\chi^2_{n(n+1)}}(1-\alpha)\right). \tag{4.19}$$

P r o o f. If hypothesis H_{1T} is true and if $T \to \infty$, then we get from (3.12):

$$L_{H_{1T}}\left\{\sqrt{T}(\hat{\theta} - \theta^* - \frac{1}{\sqrt{T}}a)\right\} = L\left\{\sqrt{T}(\hat{\theta} - \theta^*) - a\right\} \to N_{n(n+1)}(0, G^{-1}). \tag{4.20}$$

Let us consider random vector ξ_T defined in (4.15). By (4.20) we have

$$L_{H_{1T}}\{\xi_T\} \to N_{n(n+1)}(G^{1/2}a, I_{n(n+1)}),$$

that means

$$\xi_T = (\xi_{Tk}) \xrightarrow{D} \eta = (\eta_k) \in R^{n(n+1)},\ L\{\eta\} = N_{n(n+1)}(G^{1/2}a, I_{n(n+1)}). \tag{4.21}$$

Then the statistic g_T from the test (4.17) is represented as the sum of the squares:

$$g_T = T\left(\hat{\theta} - \theta^*\right)' G\left(\hat{\theta} - \theta^*\right) = \xi_T'\xi_T \to \eta'\eta = \sum_{k=1}^{n(n+1)} \eta_k^2, \tag{4.22}$$

and $\{\eta_k\}$ are asymptotically normally distributed according to (4.21). By the definition of noncentral χ^2-distribution [1] we obtain the first statement (4.18) of the Theorem. Calculate the asymptotic power of the test using (4.22) and (4.18):

$$w_T := \mathbf{P}\left\{d = 1 \middle| H_{1T}\right\} = \mathbf{P}_{H_{1T}}\{g_T \geq \Delta\} =$$

$$= 1 - \mathbf{P}_{H_{1T}}\{g_T < \Delta\} \to 1 - F_{\chi^2_{\Delta^2, n(n+1)}}(\Delta) =$$

$$= 1 - F_{\chi^2_{\Delta^2, n(n+1)}}\left(F^{-1}_{\chi^2_{n(n+1)}}(1-\alpha)\right). \qquad \square$$

5. Optimal forecasting statistic

Consider now the problem of forecasting of the future state $X_{T+\tau}$ in $\tau \geq 1$ steps ahead based on observations until the time point $t = T$ inclusively: $X_1, \dots, X_T$. Denote some forecasting statistic

$$\hat{X}_{T+\tau} = g_\tau\left(X_1, \dots, X_T; \theta\right),$$

where θ is the vector of true values of the model parameters. Let us characterize the error of forecasting of the future state $X_{T+\tau}$ in $\tau \geq 1$ steps ahead based on statistic $g_\tau(X_1, \ldots, X_T; \theta)$ by the matrix mean square (MS) risk of forecasting:

$$r(\tau) = E\left\{\left(\hat{X}_{T+\tau} - X_{T+\tau}\right)\left(\hat{X}_{T+\tau} - X_{T+\tau}\right)'\right\} \in R^{n\times n}.$$

To construct forecasting statistic $\hat{X}_{T+\tau} = g_\tau(X_1, \ldots, X_T; \theta)$ we will calculate also the conditional matrix mean square risk for the forecast under fixed prehistory $\{X_1, X_2, \ldots, X_T\}$:

$$r_c(\tau, X_1, X_2, \ldots, X_T) ::= \tag{5.23}$$
$$= E\left\{\left(\hat{X}_{T+\tau} - X_{T+\tau}\right)\left(\hat{X}_{T+\tau} - X_{T+\tau}\right)' \middle| X_1, X_2, \ldots, X_T\right\} \in R^{n\times n}.$$

To find matrix mean square risk by conditional matrix mean square risk we use the formula:

$$r(\tau) = E\{r_c(\tau, X_1, X_2, \ldots, X_T)\}$$

Theorem 5.1. *Under known parameters the MS-optimal (in terms of minimum of the mean square risk $r(\tau)$) forecasting statistic $\hat{X}_{T+\tau} = g_\tau(X_T; \theta)$, by T previous observations $\{X_t, t = 1, 2, \ldots, T\}$ for the model* (2.2), (2.3) *in $\tau \geq 1$ steps ahead has the form:*

$$\hat{X}_{T+\tau} = \sum_{J\in L} J h_{X_T,J}(T, T+\tau), \tag{5.24}$$

where $h_{I,J}(t_1, t_2)$ is defined in (2.6).

P r o o f. MS-optimal forecasting statistic is defined by conditional mathematical expectation [3]:

$$\hat{X}_{T+\tau} = E\left\{X_{T+\tau} \middle| X_1, X_2, \ldots, X_T\right\}.$$

Since dependence in $\{X_t\}$ is determined by Markov chain of first order, the forecasting statistic depends only on X_T:

$$\hat{X}_{T+\tau} = E\left\{X_{T+\tau} \middle| X_T\right\}.$$

By definition of conditional mathematical expectation of discrete random variable we obtain:

$$\hat{X}_{T+\tau} = \sum_{J\in L} J\mathbf{P}\left\{X_{T+\tau} = J \middle| X_T\right\} = \sum_{J\in L} J h_{X_T,J}(T, T+\tau),$$

that coincides with (5.24). □

Corollary 5.1. *Conditional mean-square risk for forecasting statistic* (5.24) *under fixed prehistory* $\{X_1, X_2, \ldots, X_T\}$ *is*

$$r_c(\tau, X_T) = \sum_{J_1 \in L} \sum_{J_2 \in L} J_1 (J_1 - J_2)' h_{X_T, J_1}(T, T+\tau) h_{X_T, J_2}(T, T+\tau), \tag{5.25}$$

where $h_{I,J}(T, T+\tau)$ *is determined by* (2.6).

P r o o f. Calculate conditional mean-square risk according to (5.23):

$$r_c(\tau, X_T) = E\left\{\left(\hat{X}_{T+\tau} - X_{T+\tau}\right)\left(\hat{X}_{T+\tau} - X_{T+\tau}\right)' \Big| X_T\right\} = \hat{X}_{T+\tau}\hat{X}'_{T+\tau} +$$

$$+ E\left\{X_{T+\tau}X'_{T+\tau} \Big| X_T\right\} - \hat{X}_{T+\tau} E\left\{X'_{T+\tau} \Big| X_T\right\} - E\left\{X_{T+\tau} \Big| X_T\right\} \hat{X}'_{T+\tau} =$$

$$= E\left\{X_{T+\tau}X'_{T+\tau} \Big| X_T\right\} - \hat{X}_{T+\tau}\hat{X}'_{T+\tau}. \tag{5.26}$$

According to (2.6) we have:

$$E\left\{X_{T+\tau}X'_{T+\tau} \Big| X_T\right\} = \sum_{J \in L} JJ' \mathbf{P}\left\{X_{T+\tau} = J \Big| X_T\right\}$$

$$= \sum_{J \in L} JJ' h_{X_T, J}(T, T+\tau). \tag{5.27}$$

Substituting (5.24), (5.27) into (5.26), we obtain:

$$r_c(\tau,\ X_T) = \sum_{J \in L} JJ' h_{X_T, J}(T, T+\tau) -$$

$$- \left(\sum_{J \in L} J h_{X_T, J}(T, T+\tau)\right)\left(\sum_{J \in L} J h_{X_T, J}(T, T+\tau)\right)' =$$

$$= \sum_{J_1 \in L} \sum_{J_2 \in L} J_1 (J_1 - J_2)' h_{X_T, J_1}(T, T+\tau) h_{X_T, J_2}(T, T+\tau),$$

that coincides with (5.25). □

In case of parametric prior uncertainty we construct the forecasting statistic using "the plug-in" principle [3]:

$$\tilde{X}_{T+\tau} = \sum_{J \in L} J \tilde{h}_{X_T, J}(T, T+\tau), \tag{5.28}$$

where $\tilde{h}_{X_T,J}(T, T+\tau) = h_{X_T,J}(\hat{\theta}; T, T+\tau)$ uses the maximum likelihood estimator $\hat{\theta}$ instead of the true value θ.

6. Results of computer experiments

To illustrate performance of forecasting we present results of computer experiments on simulated and real statistical data.

6.1. Experiments on simulated data

We consider the model (2.2), (2.3) with the following values of parameters:

$$\begin{aligned} & n = 3, N = 2, A = \{0, 1, 2\}, S = \{1, 2, 3\}, z = 2, \\ & \theta_1 = (-0.2, -0.18, -0.15, 0.2)', \theta_2 = (-0.18, 0.24, -0.05, -0.1)', \\ & \theta_3 = (0.13, -0.13, -0.29, 0.3)'. \end{aligned} \tag{6.29}$$

Estimates of the MS-forecast error (5.23) calculated by $M = 500$ Monte-Carlo replications for different values of T, τ are presented in the Table 1.

Table 1: Monte-Carlo estimation of the MS-forecast error

s	$T = 16, \tau = 3$	$T = 17, \tau = 2$	$T = 18, \tau = 1$
1	0.509467	0.381694	0.450632
2	0.344748	0.498119	0.500252
3	0.417257	0.4541	0.481359

Figure 1 illustrates simulated data and computed "plug-in" forecasts by $T = 16$ observations in $\tau = 3$ steps ahead for the site number $s = 1$.

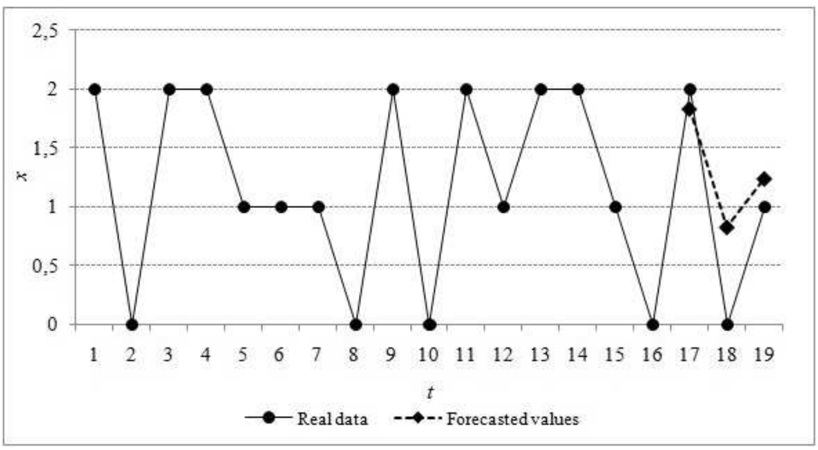

Figure 1: Illustration of forecast of simulated data for the site $s = 1$

6.2. Experiments on real data

We use real data (`http://www.gks.ru/`) that describes the level of criminality $x_{s,t}$ determined by number of registered crimes in 3 regions ($n = 3$) of the Russian Federation for 19 years ($T = 19$); exogenous variable $z_{1,t}$ is the average salary for these three regions. We consider the model (2.2), (2.3) with the following models characterizations: $N = 3$, $m = 1$, $A = \{0, 1, 2\}$, $n = 3$, $S = \{1, 2, 3\}$. Total number of parameters for the model is $n * (n + m) = 12$. Statistical estimators of the parameters obtained by these real data are

$$\hat{\theta}_1 = (-0.015, -0.15, 0.84, -0.91)', \hat{\theta}_2 = (0.17, 0.93, 1.6, -3.02)',$$

$$\hat{\theta}_3 = (-0.86, 7.62, 2.65, -10.14)'.$$

To illustrate performance of the forecasting statistic (5.28), we use the MS-forecast error for the region s:

$$\hat{r}_s = \frac{1}{\tau} \sum_{t=1}^{\tau} (\hat{x}_{s,t} - x_{s,t})^2, \quad (6.30)$$

where $\{x_{s,t} : t = 1, \ldots, T; s \in S\}$ is the observed data, which is used to construct the forecasting statistic, $\{x_{s,t} : t = T + 1, \ldots, T + \tau; s \in S\}$ are true future values which are needed to be forecasted, $\{\hat{x}_{s,t} : t = T + 1, \ldots, T + \tau; s \in S\}$ are estimates calculated by the forecasting statistic (5.28).

To study the performance of the constructed forecasts for the considered model, experiments were conducted on real data for different sizes of base period $T = 17; 18; 19$. The values of the MS-forecast error (6.30) for different values of T, τ are presented in the Table 2.

Table 2: Values of the MS-forecast error

s	$T = 16, \tau = 3$	$T = 17, \tau = 2$	$T = 18, \tau = 1$
1	0.49	0.18	0.29
2	0.006	0.005	0.003
3	0.0004	0.0001	0.0001

Figure 2 illustrates real data and computed "plug-in" forecasts in $\tau = 3$ steps ahead for the region number $s = 1$.

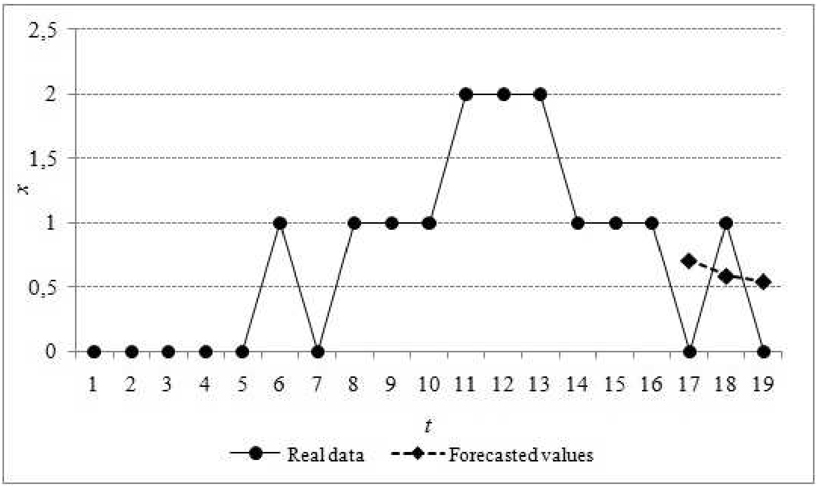

Figure 2: Illustration of forecast of real data for the region number $s = 1$

7. Conclusion

The Binomial conditional autoregressive model of spatio-temporal data is developed. It is proved that under this model the observed process is the nonhomogeneous vector Markov chain with finite space of states. Probabilistic properties of this model are studied. An algorithm for computing the maximum likelihood estimators for model parameters is developed; asymptotic properties of estimators are studied. Decision rule for statistical hypotheses testing is built and an asymptotic expression of the power of the test is obtained for a family of contiguous alternatives. The forecasting statistic that minimizes mean-square risk of the forecast error is built. The forecast error is calculated; "plug-in" forecasting statistic is constructed in the case of unknown model parameters. The computer experiments are carried out on simulated and real data.

REFERENCES

[1] A. Borovkov. Mathematical Statistics. Amsterdam, Gordon and Breach Science Publishers, 1998.

[2] S. Y. Kang, J. McGreed, P. Baade, K. Mengersen. a case study for modelling cancer incidence using Bayesian spatio-temporal models. *Australian and New Zealand Journal of Statistics*, **57**, No 3 (2015), 325–345.

[3] Yu. S. Kharin. Robustness in statistical forecasting. Robustness and complex data structures, 225-242, Springer, Heidelberg, 2013.

[4] Yu. S. Kharin, M. K. Zhurak. Statistical analysis of spatio-temporal data based on Poisson conditional autoregressive model. *Informatica (Vilnius)*, **26**, No 1 (2015), 67–87.

[5] Yu.S. Kharin, M. K. Zhurak. The binomial conditional autoregressive model of the spatio-temporal data and its probabilistic and statistical analysis. *Doklady NAN Belarusi*, **59**, No 6 (2015), 5–12. (in Russian)

[6] Yu.S. Kharin, M. K. Zhurak. Asymptotic analysis of maximum likelihood estimators for parameters of binomial conditionally autoregressive model of spatio-temporal data. *Izvestiya NAN Belarusi. Seria Phiz.-mat. Nauk,* No 1 (2016), 36–45. (in Russian)

[7] L. Mariella, M. Tarantino Spatial temporal conditional Auto-Regressive Model: A New Autoregressive Matrix. *Austrian Journal of Statistics*, **39**, No 3 (2010), 223–244.

[8] J. Moller, M. Ghorbani, E. Rubak Mechanistic spatio-temporal point process models for marked point processes, with a view to forest stand data. *Biometrics*, **72**, No 3 (2016), 687–696.

[9] O. E. Barndorff-Nielsen, F. E. Benth, A. E. D. Veraart. Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. Advances in mathematics of finance. Selected papers of the 6th general AMaMeF and Banach Center Conference, Warsaw, Poland, June 10–15, 2013. Banach Center Publications, Vol. **104**, 2015, 25–60.

[10] N. Umlauf, G. Mayr, J. Messner, A. Zeileis Why Does It Rain on Me? A Spatio-Temporal Analysis of Precipitation in Austria. *Austrian Journal of Statistics*, **41**, No 1, (2012), 81–92.

Yuriy Kharin
Research Institute for Applied Problems of Mathematics and Informatics
Belarusian State University
4 Nezavisimosty, Minsk, Belarus
e-mail: kharin@bsu.by

Marina Zhurak
Faculty of Applied Mathematics and Informatics
Belarusian State University
4 Nezavisimosty, Minsk, Belarus
e-mail: mzhurak@gmail.com