# Statistical Analysis of Poisson Conditionally Nonlinear Autoregressive Time Series by Frequencies-Based Estimators

**Yu. Kharin[a],* and M. Kislach[a],****

[a] *Research Institute for Applied Problems of Mathematics and Informatics, Belarusian State University, Minsk, 220030 Belarus*

** e-mail: kharin@bsu.by*

*** e-mail: kislachm@gmail.com*

**Abstract**—Poisson conditionally nonlinear autoregressive model is proposed for integer-valued time series. Frequencies-based estimators (FBE) for model parameters, statistical forecasting statistics, and statistical tests for this model are constructed; their performance is analyzed theoretically and by computer experiments on simulated and real data.

*Keywords:* statistical forecasting, integer-valued time series, Poisson distribution, Markov chain, frequencies-based estimators

**DOI:** 10.1134/S1054661820010083

## 1. INTRODUCTION

In forecasting (prediction) and recognition of data $\{x_t\}$ that are dependent on time $t$ and have stochastic nature the time series models are used. The theory of time series analysis is deep developed for "continuous" data when the observation space $A$ is some Euclidean space or its subspace of nonzero Lebesgue measure.

In practice, however, (because of "digitalization" of our real world) the researchers need to use discrete-valued models of time series [1–10] when the observation space $A$ is some discrete set with cardinality $|A| \leq +\infty$. Give some applied areas where discrete-valued time series models are extremely helpful: bioinformatics for analysis of genetic sequences; information systems for information protection; meteorology for weather prediction; social science for modelling of dynamics of social behavior; public health and personalized medicine; prediction of environmental processes; financial engineering; telecommunications; alarm systems. In statistical analysis (estimation of model parameters, hypotheses testing, forecasting, pattern recognition) of integer-valued time series two main approaches are used: (1) the approach based on GLM-models developed by Fokianos and Kedem [11]; (2) the approach based on thinning operators [12].

In this paper we develop a new approach [8] based on high-order Markov chains.

Received October 3, 2019; revised October 14, 2019; accepted October 21, 2019

## 2. MATHEMATICAL MODEL AND ITS PROPERTIES

Let $(\Omega, F, P)$ be the complete probability space. In this space we observe count time series $x_t \in A$ with discrete time $t \in Z = \{\dots,-1, 0, 1, \dots\}$ and denumerable state space $A = N_0 = \{0, 1, \dots\}$. Let us say that this time series satisfies the Poisson conditionally nonlinear autoregressive model of order $s \in N = \{1, 2, \dots\}$ (abbreviation $\Pi CNAR(s)$) if the conditional probability distribution of $x_t$ under its prehistory is Poisson distribution:

$$L\{x_t \mid x_{t-1}, x_{t-2}, \cdots\} = \Pi(\lambda),$$

that is

$$P\{x_t = k | x_{t-1}, x_{t-2}, \dots\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in A, \qquad \lambda = \lambda(X_{t-s}^{t-1}), \tag{1}$$

where

$$\lambda(X_{t-s}^{t-1}) = \exp(\theta' \Psi(X_{t-s}^{t-1})) = \exp\left(\sum_{i=1}^{m} \theta_i \psi_i(X_{t-s}^{t-1})\right), \tag{2}$$

$\theta = (\theta_i) \in R^m$ is the column vector of unknown parameters,

$$X_{t-s}^{t-1} = (x_{t-1}, x_{t-2}, \dots x_{t-s})' \in A^s$$

is the column vector of prehistory of the depth $s$,

$$\Psi(u) = (\psi_1(u), \dots \psi_m(u))' : A^s \to R^m$$

is the column vector of $m$ predefined base functions; $Q'$ means transposition of matrix $Q$.

**Lemma 1.** *For the model* (1), (2) *the first and the second order moments of the conditional probability distribution are*

$$\mu(J_1^s) ::= E\{x_t | X_{t-s}^{t-1} = J_1^s\} = \lambda(J_1^s),$$
$$\sigma^2(J_1^s) ::= V\{x_t | X_{t-s}^{t-1} = J_1^s\} = \lambda(J_1^s), \quad J_1^s \in A^s. \tag{3}$$

**Proof.** Expressions (3) follow from the properties of the Poisson probability distribution [13].

**Lemma 2.** *For the model* (1), (2) *the following equations hold*:

$$\ln(\lambda(J_1^s)) = \theta' \Psi(J_1^s), \quad J_1^s \in A^S. \tag{4}$$

**Proof.** Expression (4) follows from (2).

**Theorem 1.** *Count time series determined by the model* (1), (2) *is the denumerable homogeneous Markov chain of order s with state space A and the one-step transitions probabilities*:

$$p(j_{s+1}; J_1^s) ::= P\{x_t = j_{s+1} | X_{t-s}^{t-1} = J_1^s\} = \lambda^{j_{s+1}} e^{-\lambda} / j_{s+1}!,$$
$$\lambda = \lambda(J_1^s) = \exp(\theta' \Psi(J_1^s)), \quad J_1^s \in A^s, \quad j_{s+1} \in A. \tag{5}$$

**Proof.** Expression (5) follows from (1), (2) and definition of the high-order Markov chain [14]:

$$P\{x_t = j_t | x_{t-1} = j_{t-1}, \ldots, x_1 = j_1\}$$
$$= P\{x_t = j_t | x_{t-1} = j_{t-1}, \ldots, x_{t-s} = j_{t-s}\},$$
$$j_1, \ldots, j_t \in A, \quad t > s.$$

## 3. FREQUENCIES-BASED ESTIMATORS AND THEIR PROPERTIES

To construct a statistical estimator for vector of parameters $\theta \in R^m$ based on the observed realization $X_1^T = (x_1, x_2, \ldots, x_T)' \in A^T$ of length $T$ let us use the approach proposed in [8] that exploits the so-called frequencies-based estimators. Introduce the notation: $I\{C\} = \{1, \text{if } C \text{ is true}, 0 \text{ else}\}$ is the indicator function of the event $C$;

$$\nu(J_1^s) = \sum_{t=s+1}^{T} I(X_{t-s}^{t-1} = J_1^s), \quad J_1^s \in A^s,$$

$$B(X_1^T) = \{J_1^s \in A^s : \nu(J_1^s) > 0\} = \{J_1^{s,(1)}, \ldots, J_1^{s,(K)}\},$$

where $K \le T - s$, and $\nu(J_1^{s,(i)}) \ge \nu(J_1^{s,(j)})$ for all $i < j,\ i,\ j = 1, \ldots, K$.

Define a function

$$K_0 = K_0(m, T, s) : N^3 \to N, \quad m \le K_0(m, T, s) \le K,$$

nondecreasing w.r.t. $m$. Also define the subset

$$B_0 = \{J_1^{s,(1)}, J_1^{s,(2)}, \ldots, J_1^{s,(K_0)}\} \subset B(X_1^T)$$

with the cardinality $|B_0| = K_0$.

**Theorem 2.** *For the model* (1), (2), *under the observed realization* $X_1^T = (x_1, x_2, \ldots, x_T)' \in A^T$ as $T \to +\infty$ *and* $\nu(J_1^s) \to +\infty$ *the statistical estimator*

$$\hat{\mu}(J_1^s) = \sum_{t=s+1}^{T} x_t I(X_{t-s}^{t-1} = J_1^s) / \nu(J_1^s) \tag{6}$$

*is strongly consistent and asymptotically normal estimator of the conditional mean* $\mu(J_1^s) = \lambda(J_1^s)$ *determined by Lemma* 1:

$$\sqrt{T}(\hat{\mu}(J_1^s) - \lambda(J_1^s)) \xrightarrow{D} N(0, \lambda(J_1^s)). \tag{7}$$

**Proof.** According to (6) estimator $\hat{\mu}(J_1^s)$ is the sample mean for subsample $\{x_t$, if $X_{t-s}^{t-1} = J_1^s\}$, so using Lemma 1 and properties of the sample mean [14] we come to the conclusion.

Using Lemma 2 consider a system of $K_0 \ge m$ equations w.r.t. $m$ unknown parameters $\theta = (\theta_1, \theta_2, \ldots \theta_m)'$:

$$\theta' \Psi(J_1^s) = \ln(\hat{\mu}(J_1^s)), \quad J_1^s \in B_0. \tag{8}$$

Introduce the notation:

$$W(\theta) = \sum_{J_1^s \in B_0} (\theta' \Psi(J_1^s) - \ln(\hat{\mu}(J_1^s)))^2,$$
$$D = \sum_{J_1^s \in B_0} \Psi(J_1^s) \Psi'(J_1^s), \tag{9}$$
$$C = \sum_{J_1^s \in B_0} \ln(\hat{\mu}(J_1^s)) \Psi(J_1^s).$$

To solve the system (8) of $K_0$ equations w.r.t. $m$ parameters we will minimize the quadratic loss function:

$$W(\theta) \to \min_\theta. \tag{10}$$

**Theorem 3.** *For the model* (1), (2), *under the observed realization* $X_1^T = (x_1, x_2, \ldots, x_T)' \in A^T$, *if* $|D| \ne 0$ *and* $T \to +\infty$, *then the FBE-estimator of* $\theta$:

$$\hat{\theta} = D^{-1} C \tag{11}$$

*is the unique solution of the minimization problem* (10), *where* $\hat{\mu}(J_1^s)$ is *determined in* (6).

**Proof.** The equation

$$\nabla_\theta W(\theta) = 2D\theta - 2C = 0_m,$$

where $0_m$ means the vector with $m$ zero components, has only one stationary point

$$\theta = D^{-1} C.$$

Second derivative $\nabla_\theta^2 W(\theta) = 2D$ is positive definite matrix, because of definition (9) and Theorem 3 condition: $|D| \neq 0$. Therefore, $\hat{\theta} = D^{-1}C$ is the unique minimum in (10).

**Theorem 4.** *Under Theorem* 3 *conditions the FBE-estimator* (11) *is strongly consistent estimator of* θ:

$$\hat{\theta} \xrightarrow{a.s.} \theta, \quad T \to +\infty.$$

**Proof.** From Theorem 2 and from properties of strongly consistent estimator [14] we get the following almost sure convergence:

$$\sum_{J_1^s \in B_0} \ln(\hat{\mu}(J_1^s))\Psi(J_1^s) \xrightarrow{a.s.} \sum_{J_1^s \in B_0} \ln(\mu(J_1^s))\Psi(J_1^s). \quad (12)$$

Multiplying expression (4) by $\Psi'(J_1^s)$, transposing and summing by $J_1^s \in B_0$ we get

$$\sum_{J_1^s \in B_0} \ln(\mu(J_1^s))\Psi(J_1^s) = \left( \sum_{J_1^s \in B_0} \Psi(J_1^s)\Psi'(J_1^s) \right)\theta = D\theta.$$

Using (12) we come to the expression

$$\sum_{J_1^s \in B_0} \ln(\hat{\mu}(J_1^s))\Psi(J_1^s) \xrightarrow{a.s.} \sum_{J_1^s \in B_0} \ln(\mu(J_1^s))\Psi(J_1^s) = \sum_{J_1^s \in B_0} \Psi(J_1^s)\Psi'(J_1^s)\theta,$$

thus $D\hat{\theta} = C \xrightarrow{a.s.} D\theta$. From the last expression we get $\hat{\theta} \xrightarrow{a.s.} \theta$.

## 4. FORECASTING OF TIME SERIES

Let us come to construction of forecasting statistics to predict future values $x_{T+\tau} \in A$, $\tau \geq 1$, in τ step ahead, based on the observed realization $X_1^T = (x_1, x_2, \ldots, x_T)' \in A^T$ of length $T$.

**Theorem 5.** *Under Theorem* 3 *conditions the optimal forecasting statistic for the future state* $x_{T+1} \in A(\tau = 1)$, *that minimizes the probability of error* $P\{\hat{x}_{T+1} \neq x_{T+1}\}$ *is*

$$\hat{x}_{T+1} = \exp(\hat{\theta}\Psi(X_{T+1-s}^T)), \quad (13)$$

*where* $y$ *means the floor function of* $y$, *and* $\hat{\theta}$ *is determined by* (11).

**Proof.** Using Theorem 1 on forecasting of homogeneous Markov chains from [7] we have:

$$\hat{x}_{T+1} = \arg\max_{j \in A} p(j; X_{T+1-s}^T).$$

According to [13] the mode of the Poisson distribution is $[\lambda]$. Using (5) and putting $\hat{\theta}$ from (11) instead of the true value θ we get (13).

Construction of forecasting statistics $\hat{x}_{T+\tau}$ for $\tau \geq 2$ steps ahead is made iteratively by using (5) and putting $\hat{x}_{T+\tau-1}, \ldots, \hat{x}_{T+\tau-s}$ instead of their unknown values.

## 5. HYPOTHESES TESTING

In statistical forecasting we usually need to test some hypotheses on true values of parameter θ.

Let us construct statistical test under $\Pi CNAR(s)$ model for two hypotheses:

$$H_0 = \{\theta = \theta^*\}, \quad H_1 = \{\theta \neq \theta^*\} = \overline{H_0}, \quad (14)$$

where $\theta^* \in R^m$ is some known a priori hypothetical value.

**Theorem 6.** *Under Theorem* 3 *conditions the FBE-estimator* (11) *is asymptotically normal estimator of* θ:

$$\sqrt{T}\,(\hat{\theta} - \theta^*) \xrightarrow{D} N(0_m, J), \quad (15)$$

*where* $J$ *is the asymptotic covariance matrix.*

**Proof.** Expression (15) follows from Theorem 2 and from theorem [14] on functional transformation of asymptotically normal sequences.

Using Theorem 6 we construct the decision rule:

$$d = d_0(X_1^T) = \begin{cases} 0 & \Delta_T \leq \Delta \\ 1 & \Delta_T > \Delta, \end{cases} \quad (16)$$

where $d = i$ is the decision $\{H_i$ is true$\}$, $\Delta = F_{\chi_m^2}^{-1}(1-\tau)$ is the $(1-\tau)$-quantile for the $\chi_m^2$ distribution with $m$ degrees of freedom, $\tau \in (0,1)$ is the fixed significance level,

$$\Delta_T = T(\hat{\theta} - \theta^*)\hat{J}^{-1}(\hat{\theta} - \theta^*),$$

$$\hat{J} = T\sum_{k=1}^{K_0} \frac{1}{\hat{\mu}(J_1^{s,(k)})\nu(J_1^{s,(k)})} D^{-1}\Psi(J_1^{s,(k)})\Psi'(J_1^{s,(k)})D^{-1}.$$

## 6. RESULTS OF COMPUTER EXPERIMENTS

Experiments were performed in R computer language on simulated and real data.

### *6.1. Simulated Data*

For the model (1), (2) with the following parameters

$$s = 3, \quad m = 10, \quad u = (u_1, u_2, u_3), \quad \psi_1(u) = 1,$$

$$\psi_2(u) = u_1, \quad \psi_3(u) = u_2, \quad \psi_4(u) = u_3, \quad \psi_4(u) = u_4,$$

$$\psi_5(u) = u_1^2, \quad \psi_6(u) = u_2^2, \quad \psi_7(u) = u_3^2,$$

$$\psi_8(u) = u_1u_2, \quad \psi_9(u) = u_1u_3, \quad \psi_{10}(u) = u_2u_3,$$

$$\theta = (-0.04, -0.017, -0.034, 0.134, 0.004, -0.029, -0.083, 0.03, -0.015, -0.038).$$

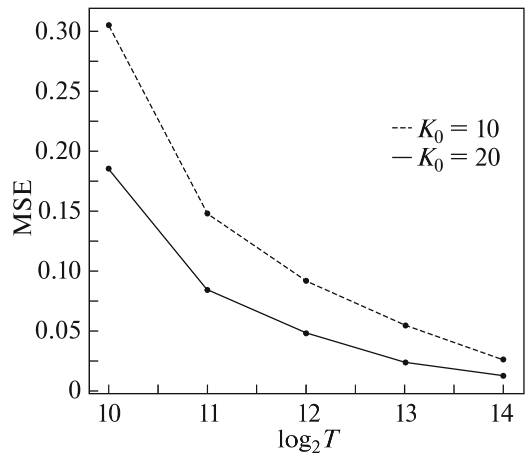


**Fig. 1.** Dependence of the mean square error estimate from $\log_2 T$.

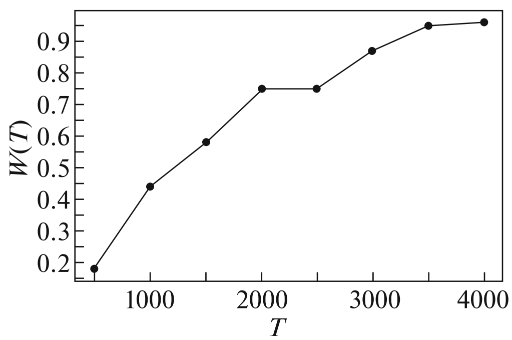


**Fig. 2.** Dependence of test power $w$ from length of time series $T$.

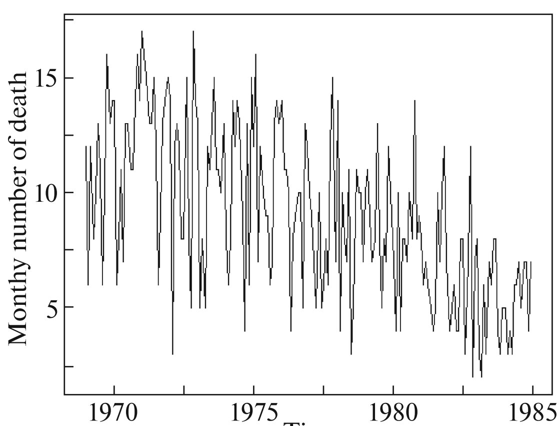


**Fig. 3.** Monthly number of killed drivers in Great Britain.

Figure 1 presents dependence of the Monte-Carlo estimate of the mean square error (MSE) for the constructed statistic (11)

$$\hat{E}\{\|\hat{\theta} - \theta\|^2\} = \sum_{i=1}^{M} (\hat{\theta}^{(i)} - \theta^{(i)})'(\hat{\theta}^{(i)} - \theta^{(i)})/M.$$

From $\log_2 T$, where $M = 100$ is the number of Monte-Carlo replications. In experiments we use two values of $K_0$: (a) $K_0 = m = 10$; (b) $K_0 = 2m = 20$.

It is seen that increasing of $K_0$ leads to decreasing of the empirical MSE.

Figure 2 presents dependence of the power $w$ of the test (16) on $T$ with the following parameters:

$$s = 2, \quad m = 4, \quad u = (u_1, u_2), \quad \psi(u) = 1, \quad \psi_2(u) = u_1,$$

$$\psi_3(u) = u_2, \quad \psi_4(u) = u_1 u_2,$$

$$\theta^* = (-0.03,\ 0.1,\ -0.7,\ 0.1),$$

$$H_1 = \{\theta = (-0.07,\ 0.1,\ -0.5,\ 0.1)\},$$

$$\varepsilon = 0.05, \quad K_0 = 4.$$

### *6.2. Real Data*

We compare forecasting statistic (13) based on the $\Pi CNAR$ model (1), (2) with forecasting statistic based on Fokianos model [15] on the time series of the monthly number of killed drivers of light goods vehicles in Great Britain between January 1969 and December 1984 [16] illustrated in Fig. 3.

For estimation of parameter $\theta$ we use data between January 1969 and December 1983 ($T = 180$) and for prediction – values between January 1984 and December 1984 ($\tau = 12$). Figure 4 presents results of prediction in $\tau = 12$ steps ahead by the model (1), (2), (11), (13) with

$$s = 4, \quad m = 5, \quad K_0 = K = 176,$$

$$u = (u_1, u_2, u_3, u_4), \quad \psi_1(u) = u_1,$$

$$\psi_2(u) = u_2, \quad \psi_3(u) = u_3, \quad \psi_4(u) = u_4,$$

$$\psi_5(u) = u_1 u_4.$$

The estimate for $\theta$ computed by (11) is

$$\hat{\theta} = (0.148,\ 0.02,\ 0.019,\ 0.173,\ -0.014).$$

As it is seen from Fig. 4 that the constructed in this paper forecasting statistic is more precise for $\tau \le 6$.

## 7. CONCLUSIONS

The following results are obtained in this paper.

(1) The Poisson conditionally nonlinear autoregressive model of order $s$ ($\Pi CNAR(s)$) is proposed.

(2) The strongly consistent frequencies-based estimator for $\Pi CNAR(s)$ is constructed.

(3) The asymptotic properties of the FBE are analyzed.

(4) Algorithm of forecasting for $\tau$ steps ahead is proposed.

(5) The performed computer experiments on real and simulated data are in agreement with the theoretical results.

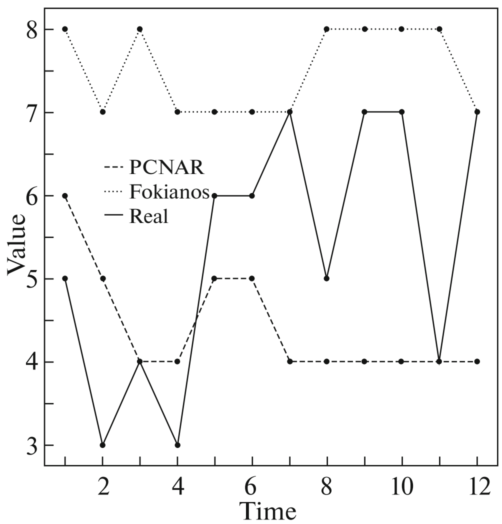


**Fig. 4.** Comparison of results of prediction.

The developed $\Pi CNAR(s)$ model can be used in robust statistical analysis [17].

## CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

## REFERENCES


1. Y. Kawahata and T. Koyama, "Application of an integer-valued autoregressive model to hit phenomena," in *Proc. 2016 IEEE Int. Conf. on Big Data* (*Big Data' 2016*) (Washington, DC, 2016), pp. 2513–2517.
2. Z. Liu, J. Almhana, V. Choulakian, and R. McGorman, "Internet traffic modeling using integer-valued time series," in *Proc. ACS/IEEE Int. Conf. on Computer Systems and Applications* (*AICCSA' 2006*) (Dubai/Sharjah, UAE, 2006), pp. 110–117.
3. Yu. S. Kharin and E. V. Vecherko, "Statistical estimation of parameters for binary Markov chain models with embeddings," Discrete Math. Appl. **23** (2), 153–169 (2013).
4. Yu. Kharin and E. Zhuk, "Filtering of multivariate samples containing "outliers" for clustering," Pattern Recogn. Lett. **19** (12), 1077–1085 (1998).
5. Yu. Kharin, "Robustness of the mean square risk in forecasting of regression time series," Commun. Stat. Theory Meth. **40** (16), 2893–2906 (2011).
6. Yu. Kharin, "Robustness of clustering under outliers," in *Advances in Intelligent Data Analysis*: *Reasoning about Data, IDA 1997*, Ed. by X. Liu, P. Cohen, and M. Berthold, Lecture Notes in Computer Science (Springer, Berlin, Heidelberg, 1997), Vol. 1280, pp. 501-511.
7. Yu. Kharin, *Robustness in Statistical Forecasting* (Springer, New York, 2013).
8. Yu. S. Kharin, V. A. Voloshko, and E. A. Medved, "Statistical estimation of parameters for binary conditionally nonlinear autoregressive time series," Math. Meth. Stat. **27** (2), 103–118 (2018).
9. M. K. Zhurak and Yu. S. Kharin, "Statistical hypotheses testing for parameters of binomial conditionally autoregressive model of spatio-temporal data," J. Belarus. State Univ. Math. Inf., No. 1, 16–22 (2017), in Russian.
10. M. K. Dauhaliova and Yu. S. Kharin, "Asymptotic analysis of statistical estimators of parameters for binomial conditionally autoregressive model for spatio-temporal data," J. Belarus. State Univ. Math. Inf., No. 2, 47–57 (2018) [in Russian].
11. B. Kedem and K. Fokianos, *Regression Models for Time Series Analysis* (Wiley, Hoboken, 2002).
12. M. G. Scotto, C. H. Weiss, and S. Gouveia, "Thinning-based models in the analysis of integer-valued time series: A review," Stat. Modell. **15** (6), 590–618 (2015).
13. N. A. J. Hastings and J. B. Peacock, *Statistical Distributions* (Wiley, New York, 1979).
14. A. A. Borovkov, *Mathematical Statistics* (Gordon and Breach, Amsterdam, 1998).
15. K. Fokianos and D. Tjøstheim, "Nonlinear Poisson autoregression," Ann. Ins. Stat. Math. **64** (6), 1205–1225 (2012).
16. T. Liboschik, K. Fokianos, and R. Fried, "tscount: An R package for analysis of count time series following generalized linear models," J. Stat. Software **82** (5), 1–51 (2017).
17. A. Kharin, "Performance and robustness evaluation in sequential hypotheses testing," Commun. Stat. Theory Meth. **45** (6), 1693–1709 (2016).


**Yuriy Kharin**, 1949, graduated from Tomsk State University in 1971, PhD in 1974, Doctor of Sciences in 1986, Research Institute for Applied Problems of Mathematics and Informatics, Belarusian State University, Director. Research area: statistical data analysis, pattern recognition. More than 350 monographs and journal articles. Correspondent member of the National Academy of Sciences of Belarus, member of the American Mathematical Society, International Institute of Mathematical Statistics, Bernoulli Society, ACM, member of editorial board of 7 journals, Laureate of Belarus State Prize for Science.

**Kislach Mikhail,** born in 1996. Received B.Sc. in actuarial mathematics from Belarusian State University, Belarus 2017, M.Sc. in applied computer data analysis from Belarusian State University 2019. Currently is a junior researcher at the Research Institute for Applied problems of Mathematics and Informatics. Field of scientific interests: statistical analysis of discrete data, time series analysis.