52. IWK

Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium

PROCEEDINGS | 10 - 13 September 2007

FACULTY OF COMPUTER SCIENCE AND AUTOMATION

COMPUTER SCIENCE MEETS AUTOMATION

VOLUME I

TECHNISCHE UNIVERSITÄT ILMENAU

52[nd] Internationales Wissenschaftliches Kolloquium
Technische Universität Ilmenau
10 – 13 September 2007

A. Avshalumov / G. Filaretov

Detection and Analysis of Impulse Point Sequences on Correlated Disturbance Phone

INTRODUCTION

The problem of low-powerful impulse point sequences analysis on phone of relatively intensive stochastic disturbance is considered. Subsequent consideration will be realized under next conditions:

a) The observable discrete process z_t can be represented as sum $z_t = x_t + i_t$; t.= 1, 2,..., N, where process x_t (disturbance) is the discrete correlated Gaussian process with variance σ_X^2 and standard autocorrelation function $\rho_{XX}(k)$, k = 0, 1, 2,....

b) Impulse component of observable process z_t is the Poisson impulse sequence; distribution function of intervals between impulses - $\varphi(t_u) = \lambda \cdot \exp(-\lambda t_u)$, where λ - unknown intensity of impulse point sequence, which necessary to find with help of observable realization.

в) Distribution function of impulses amplitudes $f(A_u) = f(A_u;\theta_1,\theta_2,...,\theta_k)$ is known up to parameters, which also have to estimate by experimental dates.

г) Impulses number M is too small in comparison with common discrete observations N (not more, than 10 – 15 %).

METHOD OF DETECTION AND ANALYSIS

Proposed method for detection and parameters estimation of Poisson impulse sequence includes two stages.

Stage 1 intended for detection and position localization of impulse sequence points. This detection can be effectively realized with the help of one of the algorithms proposed by authors [1]. The main purpose of this algorithm is to detect anomalous observations. Here, we understand anomalous observations to be those which essentially breaks the smoothness of observable sequence trajectory.

This algorithm is based on detection of statistically significant deviation observable value z_t from point z_t^*, which is found by linear interpolation in two neighbour points z_{t-1} and z_{t+1}.

Define stochastic value y_t :

$$y_t = z_t - z_t^* = z_t - \frac{z_{t+1} + z_{t-1}}{2} \tag{1}$$

This value, if anomalous observations are absent, has Gaussian distribution with zero mean and variance σ_y^2. Then presence of anomalous observation in point z_t can be established by detection of significant deviation value y_t from zero, that is if

$$|y_t| > u_{1-P} \cdot \sigma_y \tag{2}$$

Here u_{1-P} - Gaussian distribution quantile, appropriated to confidence probability P.

Really for detection of anomalous observation it is more conveniently the other formula. For its finding we present y_t in another form:

$$y_t = z_t - \frac{z_{t+1} + z_{t-1}}{2} = \frac{(z_t - z_{t+1}) + (z_t - z_{t-1})}{2} = \frac{-\nabla_{t+1} + \nabla_t}{2} = \frac{-\nabla_{t+1}^{(2)}}{2}, \tag{3}$$

where $\nabla_{t+1}^{(2)}$ - the second order difference for time moment $t+1$. Then instead of (2) we can write the equivalent formula, using the second order differences:

$$\left|\nabla_{t+1}^{(2)}\right| > u_{1-P} \cdot \sigma_{\nabla^{(2)}} \tag{4}$$

In common form thus algorithm of anomalous observations detection includes next sequence of operations:

- Computing of the first order $\nabla_t = z_t - z_{t-1}$, $t = \overline{2,N}$ and the second order differences $\nabla_t^{(2)} = \nabla_t - \nabla_{t-1}$, $t = \overline{3,N}$ for time series $z_1, z_2, ..., z_N$.
- Estimating of variance $\sigma_{\nabla^{(2)}}^2$ for time series $\nabla_t^{(2)}$.
- Choice of confidence probability P (usually P = 0,90 - 0,99).
- Fixing of anomalous observations: point z_t is related to the category of anomalous observations, if inequality (4) takes place.
- Elimination of fixed anomalous observations from time series $z_1, z_2, ..., z_N$: all these points are replaced by new values, which are computed with help of linear interpolation in two neighbour points.

In practice we need to organize iterative regime of this algorithm work. This ch necessity is connected with distinguishes of statistical properties z_t and x_t, when there are anomalous points. In process of these points eliminating time series z_t, corrected by this way, will be approximated more and more to process x_t. The iterative process lasts until new anomalous points have not been found during the latest iteration. As a rule two iterations are enough for this. As a result the position of all found anomalous points on discrete time scale $t_1^*, t_2^*, \ldots, t_{m-1}^*$ is fixed.

Stage 2 provides for realization of the next steps:

➢ For every found anomalous point the impulse amplitude is computed as deviation of observation in this point $z_{t_j^*}$ from value, which is found by linear interpolation in two neighbour points

$$A(t_j^*) = z_{t_j^*} - \frac{z_{t_j^*+1} + z_{t_j^*-1}}{2}; j = 1, 2, \ldots, m. \quad (5)$$

After this the amplitudes distribution function $f(A_u)$ histogram with ordinates m_i; $i = 1, 2, \ldots$ is built.

➢ Extracting of subset *I*, which contains substantively warped values of histogram ordinates on account of limited sensitivity of the detection anomalous observations algorithm. This extraction can be made or visual with help of histogram, or by formal exception of histogram intervals from zone $\pm u_{1-p} \cdot \sigma_{\nabla(z)}$; value $\sigma_{\nabla(z)}$ is computed for corrected z_t after last iteration.

➢ Estimation of distribution function $f(A_u)$ parameters $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ using values m_i, which not belong subset *I*.

➢ Determination of ordinates $\hat{m}_i$ for histogram intervals, which belong subset *I*, using function $f(A_u; \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$; building of corrected histogram.

➢ Computing of loss coefficient K_l in fixation of anomalous points, connected with limited sensitivity of the detection anomalous observations algorithm $K_l = \dfrac{m}{\sum_{i \notin I} \hat{m}_i + \sum_{i \in I} \hat{m}_i}$. This coefficient defines the relativity decrease of impulses numbering comparison with its real value.

- Determination of time intervals between neighbor detection points $t^{*}_{uj} = t^{*}_{j+1} - t^{*}_{j}$; j = 1, 2,..., m. As impulses with amplitudes near zero cannot be detected, selected point process differs from real: it is more rare. However on account of random location of nondetected points, this process remains Poisson impulse process [2], but with another intensity $\lambda^{*} < \lambda$.
- Estimating of intensity parameter λ^{*} with help of maximum likelihood method:

$$\hat{\lambda}^{*} = m \cdot \left(\sum_{j=1}^{m} t^{*}_{uj} \right)^{-1} .$$

- Determination of corrected estimation for intensity parameter $\hat{\lambda} = \hat{\lambda}^{*} / K_{I}$.

Proposed algorithm works out the formulated above problem in full volume. It is necessary to underline that quality of end result depends on peculiarities of concrete applied task (what kind are characteristics of process x_t, impulse component, function $f(A_u)$ and so on).

MODEL EXAMPLE

Potential possibilities of proposed algorithm are illustrated by the next model example. Process x_t is formed by quadruple passing of discrete white noise through inertia element (constant time – 10 discrete time units). The observed realization length - N= 20020 discrete volumes. Poisson impulse point process includes 917 points. Model (empirical) volume of intensity $\hat{\lambda}_{M} = 0{,}046$. Distribution function of impulse amplitudes – exponential - $f(A_u;\theta_1) = \theta_1 \exp(-\theta_1 A)$; $\theta_1 = 250$. Histograms, characterizing model distributions of amplitudes and intervals between impulses, are represented on fig. 1a) and fig. 2a).

Stage 1: for the first iteration 565 points were detected (critical value u_{1-P} =1,28); for the second iteration 158 points were detected (critical value u_{1-P} = 3,09); the third iteration was not fulfilled. In total 723 points from 917 were detected.

Stage 2: with the formula (5) amplitudes $A(t^{*}_{j})$ are estimated and the appropriate distribution histogram is built (fig. 1b). As the subset I we choose points, belonging to the first histogram interval (number of these point is equal 164). Using the all other histogram intervals with help of nonlinear estimation method we find approximation for dependence of histogram ordinates from interval centers and also unknown parameter of exponential (under condition) amplitude distribution:

$$m = b \cdot \exp(-\hat{\theta}_{\text{I}} A). \qquad (6)$$

The next estimations were got: $b = 461{,}87$; $\hat{\theta}_{\text{I}} = 250{,}8$. With help of formula (6) we define the estimation for points number in the first grouping interval or another words – for subset *I* : $\hat{m}_{\text{I}}$ = 359. Corrected histogram with using of $\hat{m}_{\text{I}}$ practically coincides with histogram of fig. 1a). After this the loss coefficient K_l in fixation of anomalous points was computed: $K_l = 0{,}788$.

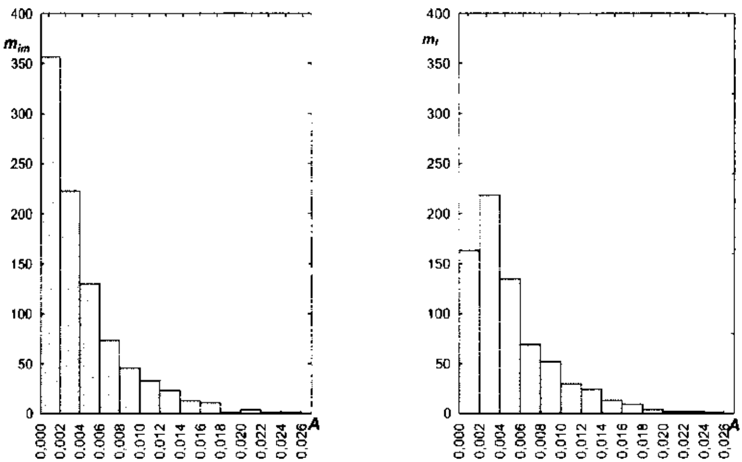

Fig. 1

Furthermore, time intervals between detected points were established. This intervals distribution histogram is shown on fig. 2b). How it was waited, we have distribution of exponential type. The intensity parameter estimation for this distribution was found: $\hat{\lambda}^*$ = 0,0361. And, at last, corrected estimation of intensity parameter is computed: $\hat{\lambda} = {}^{0{,}0361}\!/_{0{,}788} = 0{,}0458$.

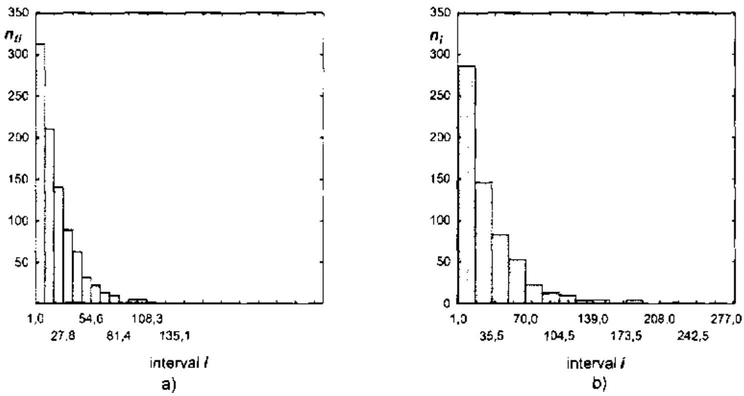

Fig. 2

Summing up, it is possible to say, that proposed method of detection and analysis is very effective. The relative error for amplitude mean is estimated to be less than 1%, and for the intensity parameter of Poisson impulse point process – about 2%. It is substantively that impulse point process capacity is less 0,1% from capacity of correlated stochastic process, on phone which this point process is detected and analyzed.

CONCLUSION

Areas of possible application proposed method can be various. In particular, it was used for aims of medical diagnostic as means of heart rhythm infringements.

References:

[1] Avshalumov A.Sh., Filaretov G.F. Algorithm of anomalous observations detection in correlated stochastic sequences.//Vestnik MEI. – 2007, №3.

[2] Cox D.R., Lewis P.A.V. The statistical analysis of series of events. New York; john Wiley, 1966.

Authors:

Dr. Gennadiy Filaretov
Moscow Power Engineering Institute,
Krasnokazarmennaya st., 14, 111250, Moscow,
RUSSIA
Phone: 007(095)517 63 19
e-mail: gefefi@yandex.ru

Alexander Avshalumov
Moscow Research Institute of Cybernetic Medicine,
Mikluho-Maklaya st., 16/10, 117871, Moscow,
RUSSIA
Phone: 007(095)7632660
e-mail: avs@cybermed.ru