DISTURBANCE OF STATISTICAL STABILITY (PART II)

Igor Gorban

Abstract: *The revue of results obtained by the author this year in the area of disturbance of statistical stability of physical quantities and processes is presented. Two new sensitive parameters characterizing their statistical instability in finite observation interval are proposed. For known and new parameters of statistical instability unit measures that give possibility to characterize the level of disturbance are introduced. It is shown by modeling that STD corridors of unit measures weakly depend from the type of the distribution. It is confirmed that important role in disturbance of statistical stability plays specific fluctuation of expectation of the process that generates changes of average expectation. Dependence of parameters of statistical instability from normalized sample variance of expectation average is found. It is shown by modeling that positive correlation calls increasing of statistical instability and negative correlation – decreasing of it. Under any type correlation increasing of sample size calls increasing of statistical stability and the process trends to statistically stable state. It is found that changes of variance influence to statistical stability of the process. Under some relations of parameters unpredictable component of the process fades-out and under other ones – fades-in. It is demonstrated high sensitivity of new parameter of statistical instability. It is shown that fluctuations of maximum day temperature and minimum day temperature in two cities (Moscow and Kiev) are instable and fluctuation of day precipitation is near stable. Disturbance of statistical stability of temperature begins from some observation weeks.*

Keywords: *statistical instability, theory of hyper-random phenomena, uncertainty, probability.*

ACM Classification Keywords: *G.3 Probability and Statistics*

1. Introduction

In the former conference [Gorban, 2010 (2)], the survey of author's researches in the theory of hyper-random phenomena, in particular in the area devoted to disturbance of statistical stability of physical quantities and processes was introduced. In the current article, the revue of the results obtained by the author in this area during last year is presented.

In the article, the main consideration was given to three directions:

- unit measures for parameters of statistical instability,
- new sensitive parameters of statistical instability,
- statistical instability of real processes.

2. Unit measure for parameters of statistical instability

Measuring of any physical magnitude beginning from it’s comparing with some unit measure.

In [Gorban, 2010 (2)] and [Gorban, 2011 (1)] two parameters of statistical instability of a sequence of N physical units $X_1, X_2, \ldots, X_N$ were introduced ad learned. They are

$$\gamma_N = \frac{\mathrm{M}\left[\bar{D}_{Y_N}\right]}{N D_{y_N}} \qquad (1)$$

and

$$\mu_N = \sqrt{\frac{\gamma_N}{1+\gamma_N}} \qquad (2)$$

where $\mathrm{M}[\cdot]$ is expectation operator,

$$\bar{D}_{Y_N} = \frac{1}{N-1}\sum_{n=1}^{N}(Y_n - \bar{m}_{Y_N})^2 \qquad (3)$$

is a sample variance of the average

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (n = \overline{1,N}), \qquad (4)$$

$$\bar{m}_{Y_N} = \frac{1}{N}\sum_{n=1}^{N} Y_n \qquad (5)$$

is it’s sample mean,

$$D_{y_N} = \frac{1}{N^2}\sum_{n=1}^{N} D_{x_n}$$

is a variance of the average, and D_{x_n} is a variance of the unit X_n .

For parameter γ_N the role of unit measure may play the variable γ_{0N} calculated according to formulae (1), (3)–(5) for sequence of sample uncorrelated units with constant variance $D_{x_n} = D_x$, zero expectation, and the same number of units; for parameter μ_N it may play the variable $\mu_{0N} = \sqrt{\gamma_{0N}/(1+\gamma_{0N})}$.

Mark, that in this approach distribution low is not essential and therefore is not stipulated.

Variable γ_{0N} for pointed sample sequence was calculated analytically in the following form [Gorban, 2011 (2)]:

$$\gamma_{0N} = \frac{N+1}{(N-1)N} C_N - \frac{2}{N-1}, \qquad (6)$$

where $C_N = \sum_{n=1}^{N} \frac{1}{n}$.

It was impossible to obtain analytically STD for the variable $\tilde{\gamma}_{0N} = \bar{D}_{Y_N} / ND_{y_N}$ without any addition assumptions. For Gaussian sample sequence it was calculated in the following form:

$$\sigma_{\tilde{\gamma}_{0N}} = \frac{1}{N-1} \sqrt{\frac{2C_N^2}{N^2} + \frac{4(N+1)C_N}{N} + A_N\left(\frac{4}{N} - 2\right) + \frac{8B_N}{N} - 12}, \qquad (7)$$

where $A_N = \sum_{n=1}^{N} \frac{1}{n^2}$, $B_N = \sum_{n=1}^{N} \frac{C_{n-1}}{n}$.

It was shown [Gorban, 2011 (2)] by modeling that the corridors $\gamma_{0N}^{\pm} = \gamma_{0N} \pm \sigma_{\tilde{\gamma}_{0N}}$ and $\mu_{0N}^{\pm} = \sqrt{\gamma_{0N}^{\pm}/(1+\gamma_{0N}^{\pm})}$ are practically identical for Gaussian and uniform distributions. Therefore it is possible to assume that deviation of the variables $\tilde{\gamma}_{0N}$ and $\tilde{\mu}_{0N}$ weakly depend from the type of the distribution.

3. New parameters of statistical instability

By using unit measure γ_{0N} it is possible to introduce relative variable

$$h_N = \gamma_N / \gamma_{0N} \qquad (8)$$

that characterizes the absolute level of statistical instability in the units γ_{0N}.

Relative level of statistical instability characterizes not only parameter μ_N but also parameter

$$I_N = \frac{\gamma_N - \gamma_{0N}}{\gamma_N} = \frac{h_N - 1}{h_N} \qquad (9)$$

that is ratio of the parameter of statistical instability calculated for unpredictable part of the process to the parameter of statistical instability calculated for whole process.

Parameter I_N is determined by the parameter μ_N and the unit measure γ_{0N}:

$$I_N = (1+\gamma_{0N}) - \frac{\gamma_{0N}}{\mu_N^2}.$$

Turndown of the parameter h_N is $[0,\infty)$ and of the parameter I_N is $(-\infty,1]$.

Dependencies of parameter γ_{0N} and its corridors $h_{0N}^{\pm} = \gamma_{0N}^{\pm} / \gamma_{0N}$ from sample size N are presented in fig. 1 by solid and dotted lines accordingly. Dependencies of parameter μ_{0N} and its corridors $I_{0N}^{\pm} = (\gamma_{0N}^{\pm} - \gamma_{0N}) / \gamma_{0N}^{\pm}$ from sample size N are presented in fig. 2 by solid and dotted lines too.

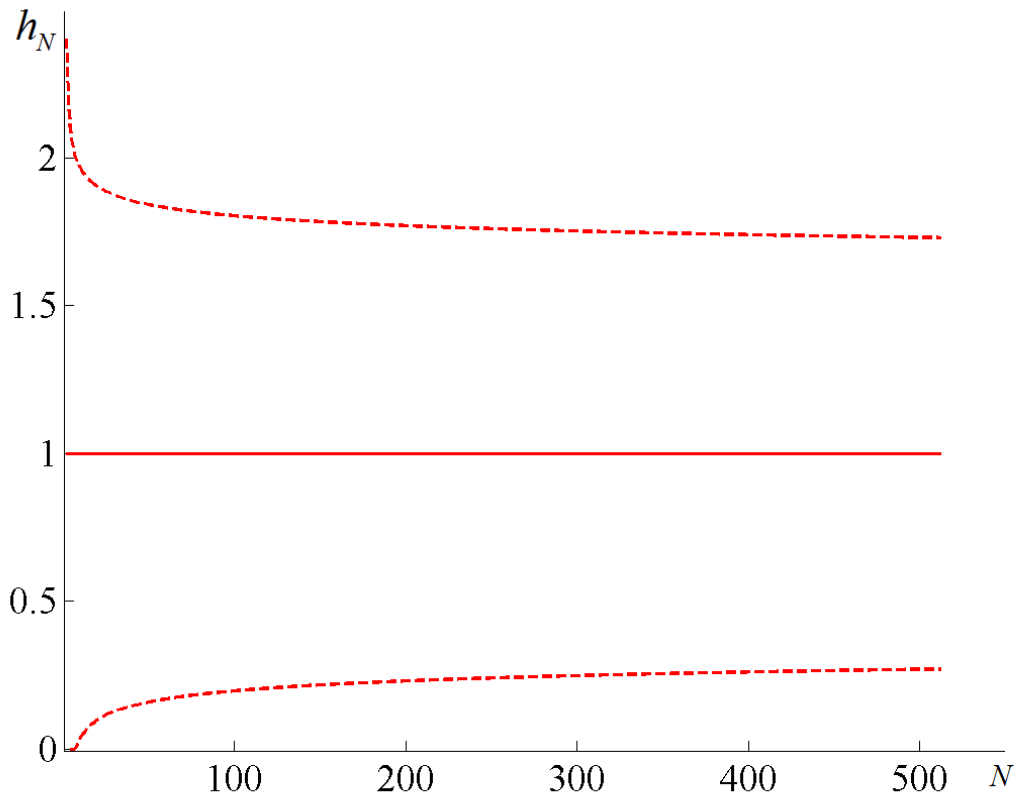

Fig. 1

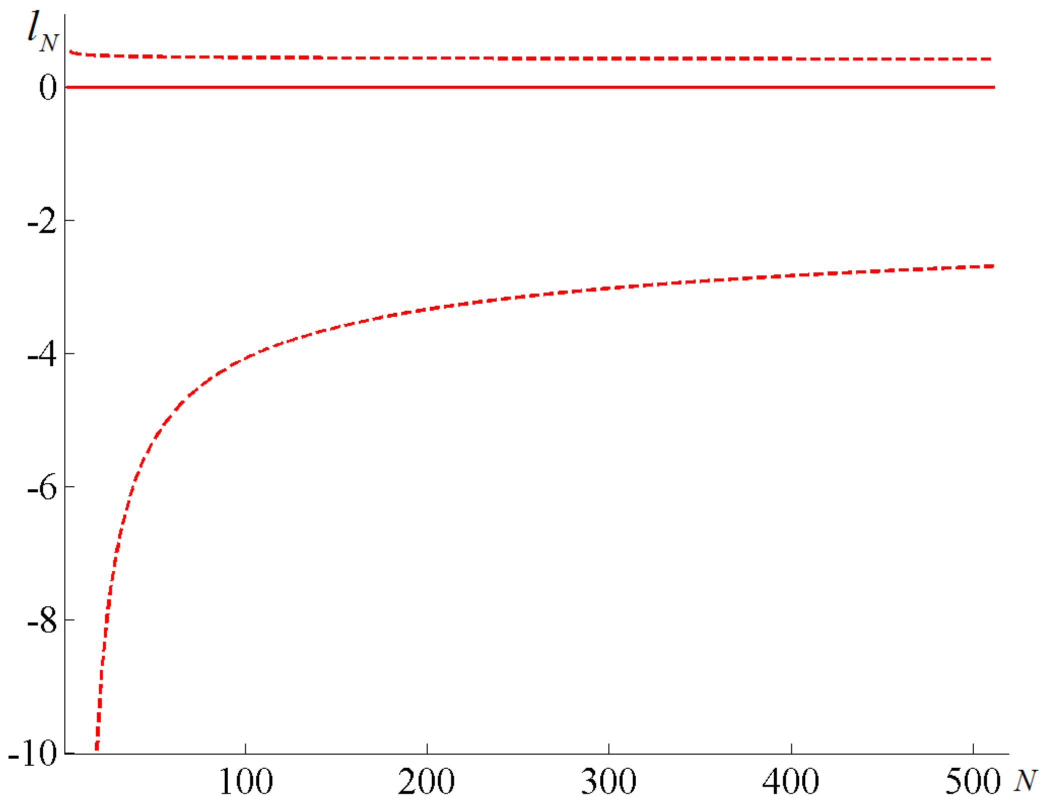

Fig. 2

It follows from the figures that with rising of sample size N the corridors $h_{0N}^{\pm}$ and $l_{0N}^{\pm}$ are converged.

4. Particularities of statistical instability parameters

Statistical instability parameters γ_N, h_N, μ_N, and l_N are physical magnitudes that characterize processes. In contrast to a lot of different other unit measures, correspondent units γ_{0N}, h_{0N}, μ_{0N}, and l_{0N} not request physical standard samples because are mathematical functions defined by sample size N. For fixed N they can be calculated without any error.

In physics, little physical constants such as light velocity, gravity constant, and others [Fundamental physical constants] are defined on definition with zero error. Statistical instability parameters γ_N, h_N, μ_N, and l_N have zero error too. In this case, zero errors is a result of that the parameters are mathematical functions.

Possibility using mathematical functions as unit measures of physical magnitudes is a result of that parameters γ_N, h_N, μ_N, and I_N have not dimension (are relative variables).

In [Gorban, 2010 (1)], [Gorban, 2010 (2)], and [Gorban, 2011 (1)] reasons which led to disturbance of statistical stability was researched. It was marked that important role plays fluctuations of some types of expectation. Periodical fluctuation components of expectation with low period and short-time steps do not disturb stability but fluctuation components with period that comparable with observation interval and also prolonged aperiodic processes lead to disturbance of stability.

Some results of more detail researches of reasons [Gorban, 2011 (2)] led to disturbance of statistical stability are presented downstream.

5. Fluctuation of expectation

Let the process described by sequence $X_1, X_2, \dots$ has constant variance D_x. Such process may be presented by sum of centered process described by sequence $\overset{\circ}{X}_1, \overset{\circ}{X}_2, \dots$ and deterministic process described by sequence of expectations $m_{x_1}, m_{x_2}, \dots$.

So average $Y_n = \bar{m}_{y_n} + \overset{\circ}{Y}_n$, where $\bar{m}_{y_n} = \frac{1}{n}\sum_{i=1}^{n} m_{x_i}$ is current average of expectations and $\overset{\circ}{Y}_n = \frac{1}{n}\sum_{i=1}^{n} \overset{\circ}{X}_i$ is average of sequence $\overset{\circ}{X}_1, \overset{\circ}{X}_2, \dots$.

Statistical instability parameter γ_N may be obtained in the following form:

$$\gamma_N = q_N + \gamma_{0N}, \tag{10}$$

where $q_N = \bar{D}_{\bar{m}_{yN}} / D_x$ is normalized sample variance of expectation average, $\bar{D}_{\bar{m}_{yN}}$ is sample variance of expectation average:

$$\bar{D}_{\bar{m}_{yN}} = \frac{1}{N}\sum_{n=1}^{N} (\bar{m}_{y_n} - \bar{m}_{\bar{m}_{yN}})^2,$$

$\bar{m}_{\bar{m}_{yN}} = \frac{1}{N}\sum_{n=1}^{N} \bar{m}_{y_n}$ is average of averages of expectations of the sequence $X_1, X_2, \dots$.

In these designations, statistical instability parameter

$$\mu_N = \sqrt{\frac{q_N + \gamma_{0N}}{q_N + (1+\gamma_{0N})}}\,, \tag{11}$$

parameter

$$h_N = \frac{q_N}{\gamma_{0N}} + 1\,, \tag{12}$$

and parameter

$$l_N = \frac{q_N}{q_N + \gamma_{0N}}\,. \tag{13}$$

It is followed from expressions (10)–(13) that changes of average of expectations $\bar{m}_{y_n}$ lead to increasing of statistical instability parameters. Parameters γ_N, h_N, μ_N, and l_N rise with rising of variance of these changes.

It must be marked that not any fluctuation of expectation of initial process leads to disturbance of statistical stability but only special one that calls fluctuation of $\bar{m}_{y_n}$.

Image about dependence of parameters μ_N and l_N from normalized variance of fluctuations of average of expectations q_N gives fig. 3. Curves 1 – 3 accords to parameter μ_N and $1' - 3'$ – to parameter l_N. Curves $1,\ 1'$ are obtained for $N = 16$, curves $2,\ 2'$ – for $N = 256$ and $3,\ 3'$ – for $N = 4096$.

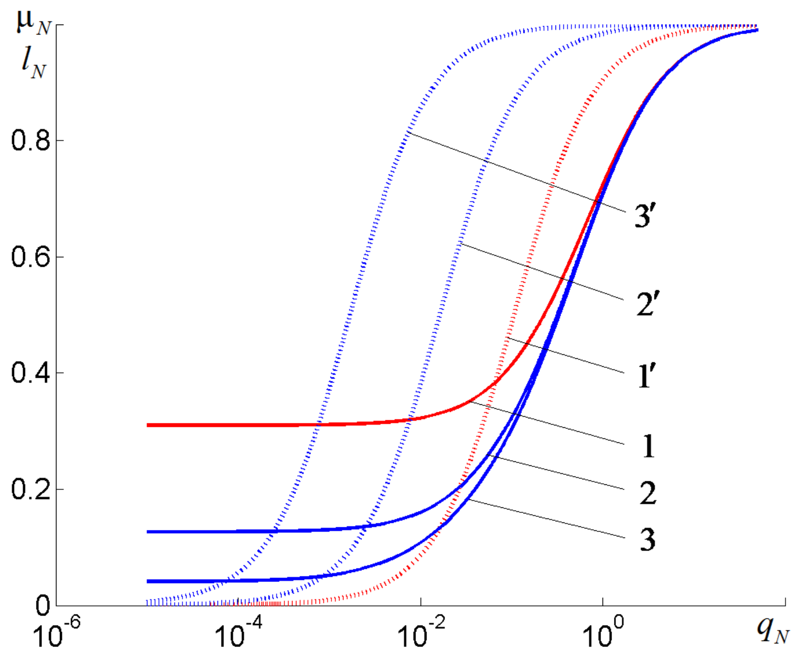

Fig. 3

Apparently from figure, when $q_N < 1$ more sensitive to changes of q_N is parameter I_N and when $q_N > 1$ – parameter μ_N. So under low disturbances of statistical stability more preferable use parameter I_N and under high disturbances – parameter μ_N.

With rising of sample size N parameter μ_N decreases and parameter I_N, contrary, increases.

When there is low disturbances of statistical stability ($q_N \ll \gamma_{0N}$) parameter I_N depends from q_N practically in line low (see expression (13)).

Not only variances of expectation may lead to change of statistical stability but another deviations from sample conditions, in particular correlation of sample units and variations of variance.

6. Correlation

Parameters of statistical instability calculated by modeling for different correlation conditions are presented in fig. 4 and 5. Solid bold curves correspond to sample conditions (uncorrelated Gaussian sequence with constant variance and zero expectation), solid extra bold curves – to sequence with positive correlated units, and dotted extra bold curves – to sequence with negative correlated ones.

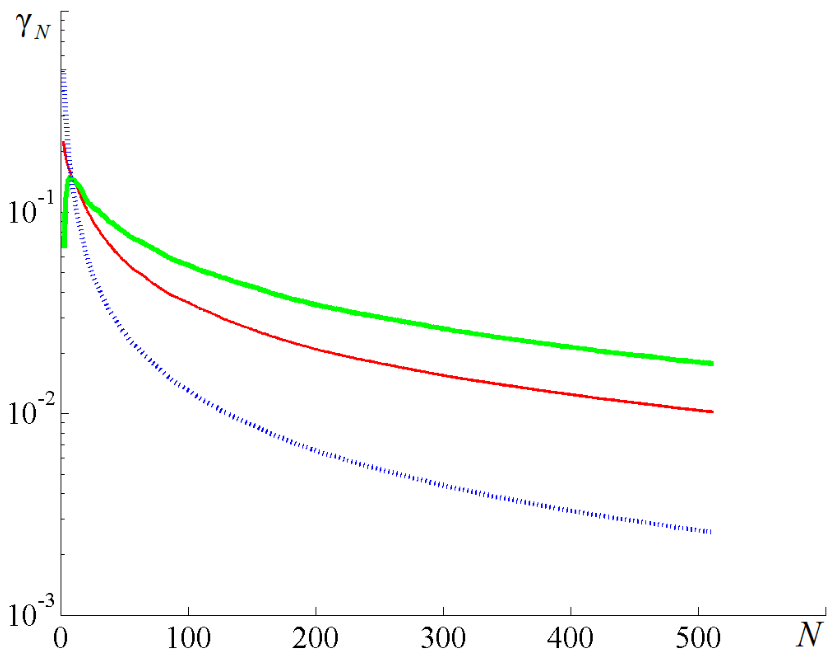

Fig. 4

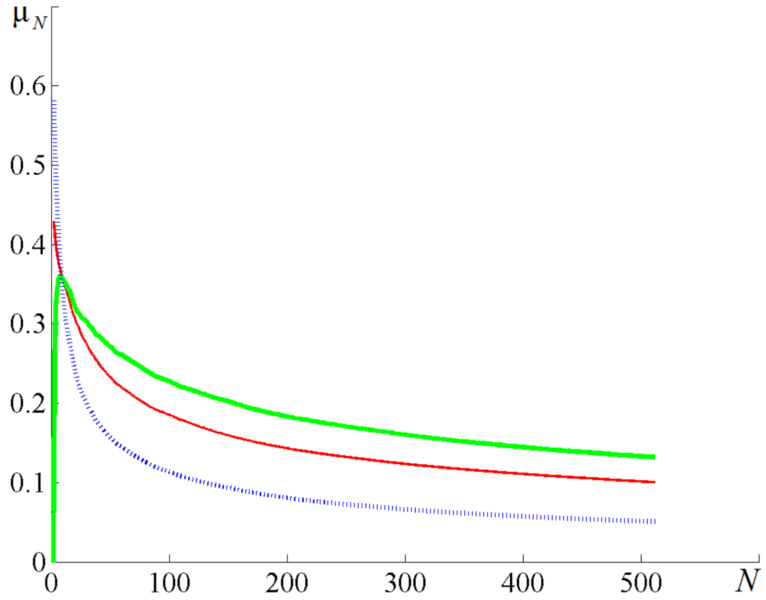

Fig. 5

It follows from figures that positive correlation calls increasing of statistical instability and negative correlation – decreasing of it.

Under any type correlation increasing of sample size N calls increasing of statistical stability and the process tends to statistically stable state.

7. Fluctuation of variance

Fluctuation of variance can call statistical instability.

Let us research sequence $X_1, X_2, \ldots$ with uncorrelated units, variance D_{x_n}, and zero expectation that can be presented by sum of two uncorrelated sequences, one of which $(\hat{X}_1, \hat{X}_2, \ldots)$ has constant variance $D_{\hat{x}} = \min_n D_{x_n}$ and zero expectation and another one ($\breve{X}_1, \breve{X}_2, \ldots$) – fluent variance $D_{\breve{x}_n}$ and zero expectation.

In this case, average $Y_n = \hat{Y}_n + \breve{Y}_n$, variance of average

$$D_{y_N} = D_{\hat{y}_N} + D_{\breve{y}_N} + 2R_{\hat{y}_N \breve{y}_N}, \qquad (14)$$

and sample variance of average

$$\bar{D}_{Y_N} = \bar{D}_{\hat{Y}_N} + \bar{D}_{\breve{Y}_N} + 2\bar{R}_{\hat{Y}\breve{Y}_N}\,, \tag{15}$$

were $\hat{Y}_n$ and $\breve{Y}_n$ are averages of sequences $\hat{X}_1, \hat{X}_2, \ldots$ and $\breve{X}_1, \breve{X}_2, \ldots$ accordingly, $D_{\hat{y}_N} = \frac{1}{N^2}\sum_{n=1}^{N} D_{\hat{x}_n}$ and $D_{\breve{y}_N} = \frac{1}{N^2}\sum_{n=1}^{N} D_{\breve{x}_n}$ are variances of $\hat{Y}_n$ and $\breve{Y}_n$, $R_{\hat{y}_N\breve{y}_N} = \frac{1}{N^2}\sum_{n=1}^{N} R_{\hat{x}_n\breve{x}_n}$ is their correlated moment, $D_{\hat{x}_n}$, $D_{\breve{x}_n}$ are variances of sequences $\hat{X}_1, \hat{X}_2, \ldots$ and $\breve{X}_1, \breve{X}_2, \ldots$, $R_{\hat{x}_n\breve{x}_m}$ is correlated moment of $\hat{X}_n$ and $\breve{X}_m$, $\bar{D}_{\hat{Y}_N}$ and $\bar{D}_{\breve{Y}_N}$ are sample variances of averages $\hat{Y}_n$ and $\breve{Y}_n$: $\bar{D}_{\hat{Y}_N} = \frac{1}{N-1}\sum_{n=1}^{N} \hat{Y}_n^2$ and $\bar{D}_{\breve{Y}_N} = \frac{1}{N-1}\sum_{n=1}^{N} \breve{Y}_n^2$, $\bar{R}_{\hat{Y}\breve{Y}_N} = \frac{1}{N-1}\sum_{n=1}^{N} \hat{Y}_n \breve{Y}_n$ is their sample correlated moment.

Taking into account that variables $\hat{X}_n$, $\breve{X}_m$ are uncorrelated ($R_{\hat{x}_n\breve{x}_m} = 0 \;\; \forall n = \overline{1,N}$ и $\forall m = \overline{1,N}$) and there are relations (14), (15), expression for parameter γ_N can be obtained in the following form:

$$\gamma_N = \frac{\gamma_{0N} + \dfrac{M[\bar{D}_{\breve{Y}_N}]}{N D_{\hat{y}_N}}}{1 + \dfrac{D_{\breve{y}_N}}{D_{\hat{y}_N}}}\,. \tag{16}$$

It follows from this expression that if expectation of sample variance $\bar{D}_{\breve{Y}_N}$ of the sequence $\breve{X}_1, \breve{X}_2, \ldots$ normalized to variance of average $D_{\breve{y}_N}$ and sample size N is more then sample variable γ_{0N}:

$$\left(\frac{M[\bar{D}_{\breve{Y}_N}]}{N D_{\breve{y}_N}} > \gamma_{0N} \right), \tag{17}$$

parameter of statistical instability $\gamma_N > \gamma_{0N}$. If inequality (17) is not true then $\gamma_N \le \gamma_{0N}$.

It follows from expression (16) that if variable $\mathrm{M}[\bar{D}_{\tilde{y}_N}]/ND_{\tilde{y}_N}$ tends to zero the process is statistically stable and in contrary case is unstable.

Results of computer modeling illustrating this resume are presented in fig. 6 and 7. Bold curves accord to sample sequence, extra bold ones – to sequence that consist of sum of sample sequence and sequence in which variance increases in line low with increasing of sample size N, and dotted extra bold ones – to sequence that consist of sum of sample sequence and sequence in which variance decreases in line low with increasing of sample size N.

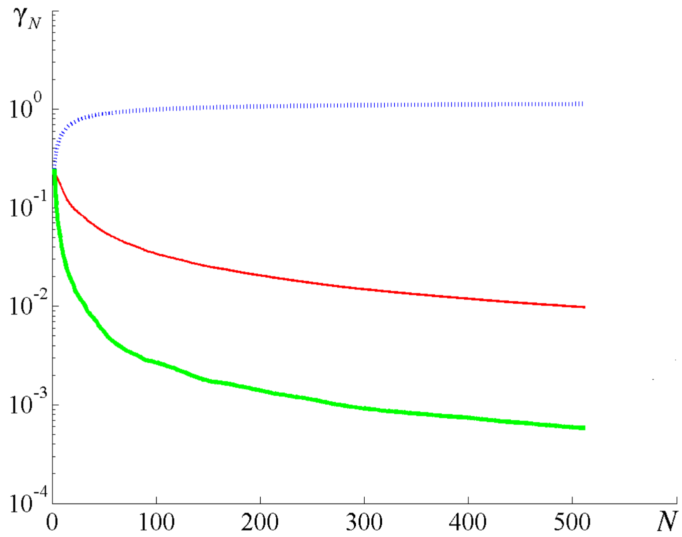

Fig. 6

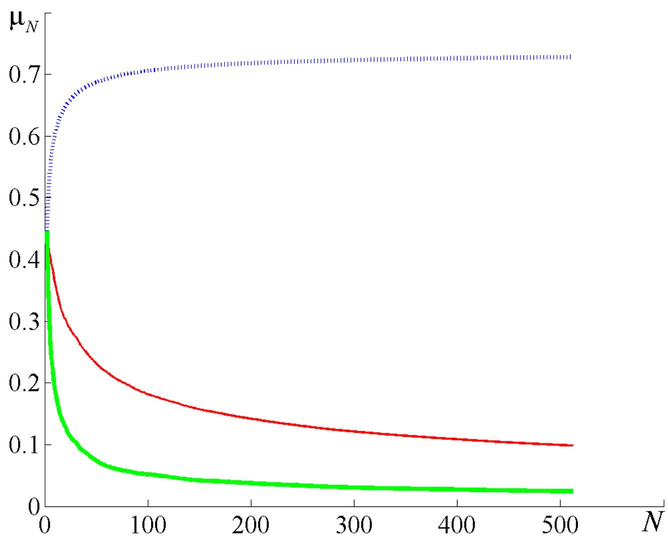

Fig. 7

It follows from figures that in the second case the process more stable and in the third case less stable than in the first one. In the second case the process explicitly tends to statistically stable state and in the third one – to statistically instable state.

So changes of variance influence to statistical stability of the process. Under some relations of parameters unpredictable component of the process fades-out and under other ones – fades-in.

8. Example of statistically instable process

Statistical stability of maximum day temperature, minimum day temperature, and day precipitation was researched with using developed methodology. Initial data was received from [Weather] for Moscow (1949 – 1992) and for Kiev (1881 – 1992).

Dependences of parameters h_N and μ_N from time t calculated according to this methodology are presented in fig. 8 for Moscow and in fig. 9 – for Kiev.

Curves in fig. 8a, 8b, 9a, 9b are obtained without ensemble averaging of sample variance $\bar{D}_{Y_N}$ and in fig. 8c, 8d and 9c, 9d – with averaging for 43 and 112 years correspondently. Overhead solid curves accord to maximum day temperature and minimum day temperature. Bottom solid curves correspond to day precipitation. For comparing, unit measures of parameters of statistical instability are presented by dash lines and abnormality from them on STD (STD corridors) – by dash-dot lines.

Curves were calculated with preliminary seasonal correction data realized by subtraction of expectation estimates and then normalizing obtained magnitudes on according STD estimates.

By analyzing parameter h_N (see fig. 8b, 8d, 9b, 9d) it is impossible to answer on the question: are researched processes statistically stable or not? Although it is clear from curves in these figures that fluctuations of maximum day temperature and minimum day temperature are less stable than fluctuation of day precipitation.

Answer on pointed question gives curves describing parameter μ_N (see fig. 8a, 8c, 9a, 9c). It follows from the figures that fluctuations of maximum day temperature and minimum day temperature are instable and fluctuation of day precipitation is near stable. Disturbance of statistical stability of temperature begins from some observation weeks.

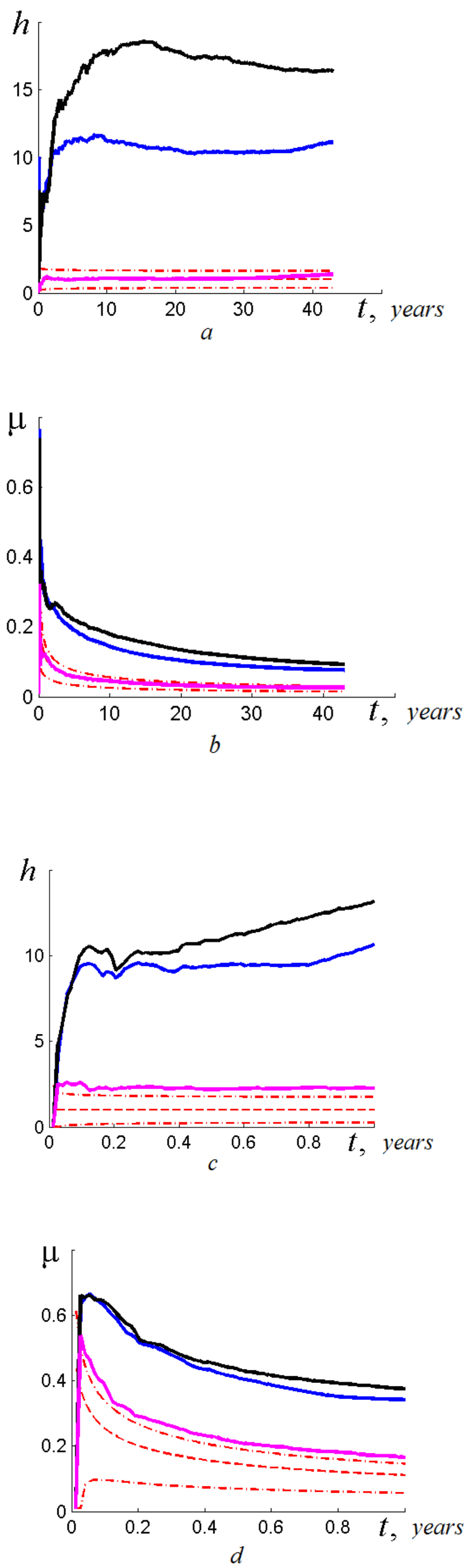
h
15
10
5
0
0 10 20 30 40
t, years
a
μ
0.6
0.4
0.2
0
0 10 20 30 40
t, years
b
h
10
5
0
0 0.2 0.4 0.6 0.8
t, years
c
μ
0.6
0.4
0.2
0
0 0.2 0.4 0.6 0.8
t, years
d

Fig. 8

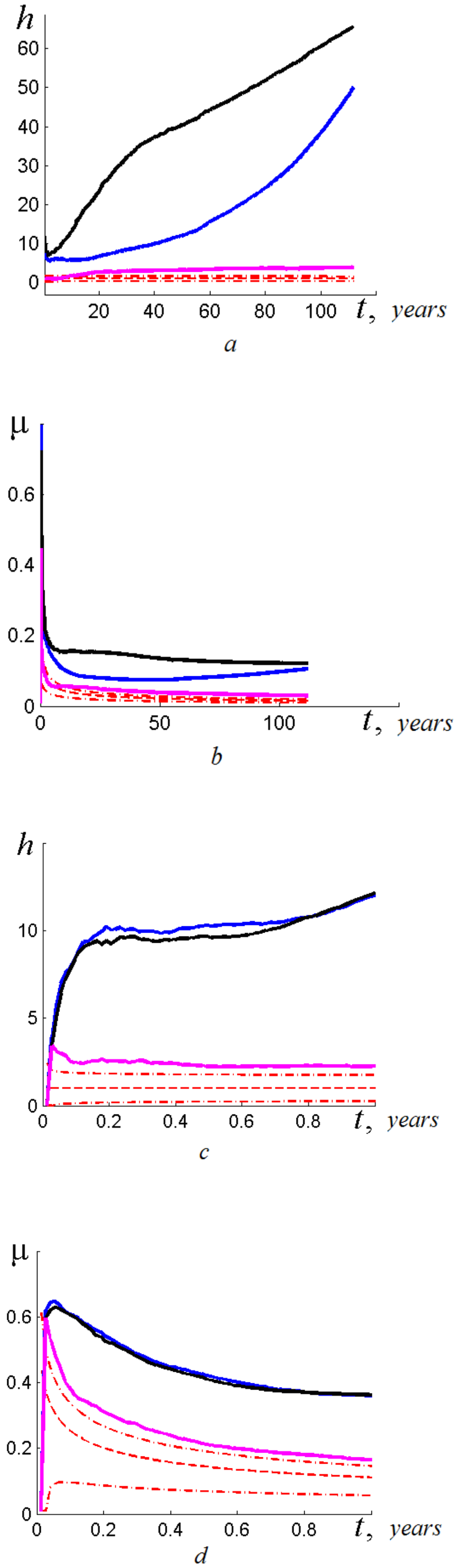
h
60
50
40
30
20
10
0
20 40 60 80 100
t, years
a
μ
0.6
0.4
0.2
0
0 50 100
t, years
b
h
10
5
0
0 0.2 0.4 0.6 0.8
t, years
c
μ
0.6
0.4
0.2
0
0 0.2 0.4 0.6 0.8
t, years
d

Fig.9

9. Conclusion

1. To describe disturbance of statistical stability on the finite observation interval two sensitive parameters (h_N and l_N) of statistical instability has been proposed.

2. For known (γ_N , μ_N) and new (h_N , l_N) parameters of statistical instability unit measures that give possibility to characterize the level of disturbance in the finite observation interval have been introduced.

3. Modeling has shown that STD corridors of unit measures weakly depend from the type of the distribution.

4. It has been confirmed that important role in disturbance of statistical stability plays specific fluctuation of expectation of the process that generates changes of average expectation. Dependence of parameters of statistical instability from normalized sample variance of expectation average has been found.

5. It has been shown by modeling that positive correlation calls increasing of statistical instability and negative correlation – decreasing of it. Under any type correlation increasing of sample size calls increasing of statistical stability and the process trends to statistically stable state.

6. It has been found that changes of variance influence to statistical stability of the process. Under some relations of parameters unpredictable component of the process fades-out and under other ones – fades-in.

7. It has been shown that fluctuations of maximum day temperature and minimum day temperature in two cities (Moscow and Kiev) are instable and fluctuation of day precipitation is near stable. Disturbance of statistical stability of temperature begins from some observation weeks.

Bibliography

[Gorban, 2010 (1)] I.I. Gorban. Distraction of statistical stability // Mathematical Machines and Systems. – №1, 2010. – P. 171 – 184. (In Russion).

[Gorban, 2010 (2)] I.I. Gorban. Disturbance of statistical stability // Information Models of Knowledge 2010. – 2010. P. 398 – 410.

[Gorban, 2011 (1)] I.I. Gorban. Theory of hyper-random phenomenon. Kiev: Naukova dumka. – 2011. – 318 p. – ISBN 978-966-00-1093-2. – http://www.immsp.kiev.ua/perspages/gorban_i_i/index.html (In Russion).

[Gorban, 2011 (2)] I.I. Gorban. Statistical instability of physical processes // Radioelectronics and Communications Systems. – 2011. (In printing).

[Gorban, 2011 (3)] I.I. Gorban. Statistical stability of temperature and precipitation in the region of Moscow // Mathematical Machines and Systems. – №3, 2011. (In Russion).

[Fundamental physical constants] Fundamental physical constants. – http://physics.nist.gov/constants.

[Weather] Weather history for CIS towns. – http://thermo.karelia.ru/weather/w_history.php.

Authors' Information

Igor Gorban – *Principal scientist of the Institute of Mathematical Machines and Systems Problem, National Academy of Sciences of Ukraine, Glushkov ave., 42, Kiev, 03187, Ukraine; e-mail: igor.gorban@yahoo.com*

Major Fields of Scientific Research: Theory of hyper-random phenomena, Space-time signal processing in complicated dynamic conditions, Fast multichannel space-time signal processing