UDC 519.246

S.O. DOVGYI, O.I. YURIKOV, M.O. ZOZYUK

ABOUT ONE STATISTICAL MODEL OF ERROR RATE IN THE STREAM OF PACKET DATA TRANSMISSION THROUGH COMMUNICATION CHANNELS

Abstract. A statistical model of the frequency of errors in the packet data transmission through communication channels is proposed. This is a stochastic sequence defined as the averaged proportion of erroneous data packets. A diffusion approximation of such a sequence is used: discrete Markov diffusion, which is defined by a difference stochastic equation. The parameters of such a model are estimated using covariance statistics on the trajectories of the stochastic sequence of signal transmission errors.

Keywords: statistical model, difference stochastic equation, stationary process, equilibrium, covariance statistics, parameters estimation along trajectories.

BASIC DEFINITIONS

Our goal is to build a model of error inclusions in packet data transmission over communication channels. In the data transmission protocols, a packet integrity check (CRC code, parity etc.) is built-in, which restores the original data at the reception point due to the redundancy of the transmission code [1, 2]. This redundancy, however, reduces the data rate and requires additional digital signal processing.

The task of determining the statistical parameters and their estimates of the fraction of the “corrupted” data packets with respect to their total volume is determined by equilibrium state of the frequency of erroneous packets. This equilibrium state is an invariant point of the regression function or, equivalently, zero point of the regression function of increments of evolutionary process [3].

Consider a string for receiving N data packets $\{1, 2, \dots, N\}$, containing both holistic and corrupted blocks arranged in random order, as illustrated in Fig. 1.

Define the following binary random variables as packets integrity indicators

$$\delta_n(k) = \begin{cases} 0, & \text{if the package } n \text{ is holistic,} \\ 1, & \text{if the package } n \text{ is corrupted,} \end{cases} \quad k \ge 0,\ 1 \le n \le N.$$

Then the normalized sum of random indicators

$$S_N(k) = \frac{1}{N} \sum_{n=1}^{N} \delta_r(k),\ \ k \ge 0, \tag{1}$$

determines the proportion of packet errors in the process of receiving and transmitting a signal over communication channels. The discrete-time random process $S_N(k)$, $k \ge 0$, is called statistical experiment [3] and takes values in the interval [0, 1]. The extreme value 0 corresponds to the situation when all received packets are fake, and the extreme value 1 corresponds to the situation when all received packets are complete and there are no errors.

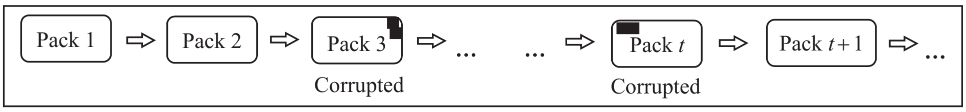

Fig. 1. Typical packet flow of data transmission with errors

It is assumed that the dynamic process $S_N(k)$, $k \geq 0$, has the following properties.

1. Stationarity, in wide sense, of the random function $S_N(k)$ with respect to time k: the mean values $E[S_N(k+1)]$, $E[(S_N(k+1))^2]$ are independent of k, as well as the covariance $E[S_N(k)S_N(i)] = R(|k-i|)$, which depends only on the difference in time instants.

2. Statistical equilibrium [4] relative to the ergodic state ρ:

$$\Delta P_{k+1} = -V[P_k - \rho],\ 0 < V < 1;\ \Delta P_{k+1} := P_{k+1} - P_k, \tag{2}$$

$$\rho = \lim_{k \to \infty} P_k,\ P_k := E[E[S_N(k) \mid S_N(k-1)]].$$

It is known that the errors dynamic process of data transmitting errors $S_N(k)$, $k \geq 0$, is approximated by normal autoregression [5], determined by a process of discrete Markov diffusion α_t, $t \geq 0$, with increments

$$\Delta\alpha_{t+1} = \alpha_{t+1} - \alpha_t,\ t \geq 0,\ \alpha_0 \text{ is given.}$$

The diffusion α_t is a solution of stochastic difference equation

$$\Delta\alpha_{t+1} = -V\alpha_t + \sigma\Delta W_{t+1},\ t \geq 0,\quad 0 < V < 1, \tag{3}$$

where ΔW_{t+1}, $t \geq 0$, is a standard Wiener process (Brownian motion) with mathematical expectation 0 and standard deviation 1.

STEADY REGIME OF ERROR RATE OF DATA RECEIPTION

As noted above, the dynamic error process $S_N(k)$, $k \geq 0$, is statistically equivalent to a process of discrete Markov diffusion α_t, $t \geq 0$, which is a solution of the stochastic difference equation (3).

The assumption of stationarity, in wide sense, of discrete Markov diffusion α_t, $t \geq 0$, implies the following numerical ratios:

$$E\alpha_0 = 0,\ E\alpha_0^2 = \sigma_0^2,\ E\alpha_t\alpha_{t+s} = B(s), \tag{4}$$

where E be the mathematical expectation.

The stationarity of diffusion α_t allows us to build a statistical model with the possibility of estimating its parameters V, σ^2 along the trajectory of observations.

An essential property of stationarity, in wide sense, of discrete Markov diffusion α_t, $t \geq 0$, are the following relations [6]:

$$\sigma^2 = \sigma^2 / \mathcal{E},\ \mathcal{E} := 2V - V^2. \tag{5}$$

The value $\mathcal{E}$, generated by the drift parameter V, is called the coefficient of stationarity.

In addition, the following representations of the covariance take place:

$$\text{cov}\,(\alpha_t, \alpha_t) = E(\alpha_t)^2 = \sigma^2,\ E\alpha_t\alpha_{t+s} = q^s\sigma^2,\ q(1-V). \tag{6}$$

For covariance analysis of statistics, we consider a two-component process $(\alpha_t, \Delta\alpha_{t+1})$, $t \geq 0$. It is known that its covariance matrix has the following representations [6]:

$$\text{cov}\,(\alpha_t, \alpha_t) = \sigma^2,\ \text{cov}\,(\alpha_t, \Delta\alpha_{t+1}) = -V\sigma^2, \tag{7}$$

$$\text{cov}\,(\Delta\alpha_{t+1}, \Delta\alpha_{t+1}) = 2V\sigma^2,\quad t \geq 0.$$

The covariance relations (7) form the basis of statistical estimates of the parameters (V, σ) of the main equation (3).

Consider the matrix of empirical covariance statistics

$$\sigma_{0T}^2 = \frac{1}{T}\sum_{t=1}^{T}\alpha_t^2, \ \sigma_{\Delta T}^2 = \frac{1}{T}\sum_{t=1}^{T}(\Delta\alpha_t)^2, \ \Delta_T = \frac{1}{T}\sum_{t=1}^{T}\alpha_t\Delta\alpha_t. \tag{8}$$

By virtue of relations (7), the mean square convergence of empirical statistics takes place:

$$\sigma_{0T}^2 \xrightarrow{L^2} \sigma^2, \ \sigma_{\Delta T}^2 \xrightarrow{L^2} -V\sigma^2, \ \Delta_T \xrightarrow{L^2} 2V\sigma^2, \ T \to \infty. \tag{9}$$

STATISTICAL PARAMETER ESTIMATION OF ERROR RATE PROCESS OF DATA TRANSMISSION

As was shown in the previous section, the stationarity conditions (4), (5) imply the covariance relations (7), and therefore the possibility to evaluate the fundamental parameters of the process of data reception errors $\alpha_t, \ t \geq 0$, which is a solution of the difference stochastic equation (3).

Namely, due to relations (7), there are two estimates of drift parameter V [7, Ch. 3]:

$$V \approx V_T^{\Delta} = \sigma_{\Delta T}^2 / \sigma_{0T}^2 = \sum_{t=1}^{T}(\Delta\alpha_t)^2 / \sum_{t=1}^{T}\alpha_t^2, \tag{10}$$

$$V \approx V_T^0 = -\Delta_T / \sigma_{0T}^2 = -\sum_{t=1}^{T}\alpha_t\Delta\alpha_t / \sum_{t=1}^{T}\alpha_t^2. \tag{11}$$

The quadratic diffusion characteristic $\alpha_t, \ t \geq 0$, according to formula (7), is estimated by the following empirical formula [7, Ch. 3]:

$$\sigma^2 \approx \sigma_T^2 = \frac{1}{T}\sum_{t=1}^{T}(\alpha_t)^2. \tag{12}$$

The stationarity coefficient $\mathcal{E}$ is estimated by the quadratic form of the drift coefficient

$$\mathcal{E} \approx \mathcal{E}_T = 2V_T^0 - (V_T^0)^2, \tag{13}$$

and the diffusion coefficient σ of equation (3) has the following statistical estimate

$$\sigma^2 \approx \sigma_T^2 = \mathcal{E}_T\sigma_T^2. \tag{14}$$

It is noteworthy that there are two different statistical estimates of the drift parameter V: (10) and (11). Such redundancy can be used to verify the adequacy of the proposed model of the process of data reception errors $\alpha_t, \ t \geq 0$.

NUMERICAL MODELING AND ANALYSIS OF THE MODEL OF THE DYNAMICS OF ERRORS OF DATA TRANSMISSION

The binary stream of the integrity/error marker of data packets consisting of zeros and ones is numerically modeled. That is, we emulate the sequence $\delta_n, \ 1 \leq n \leq N$, where N is a fixed size of the generated sample, in our case $N = 1525$. The sample δ_n is constructed as follows: N binary numbers 0 or 1 with randomly generated "average distance" $\lambda = 1/m$ of the location of zeros in the sample corresponding to the relative frequency of packet errors in the process of receiving and transmitting a signal over communication channels. Moreover, the appearance of the value "0" (error) is equally probable for all members of the sample. That is

$$ES_N = E\left[\frac{1}{N}\sum_{n=1}^{N}\delta_n\right] = m. \tag{15}$$

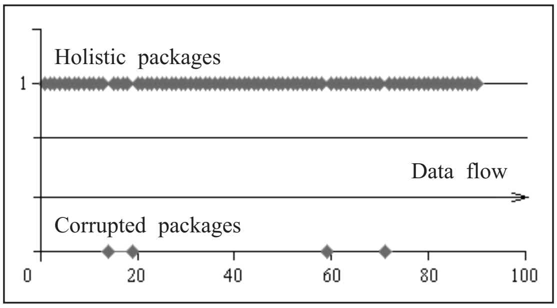

Fig. 2. The random flow generation with the error rate $m = 0.05$

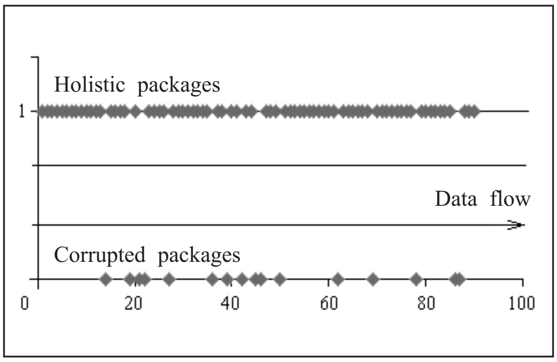

Fig. 3. The random flow generation with the error rate $m = 0.2$

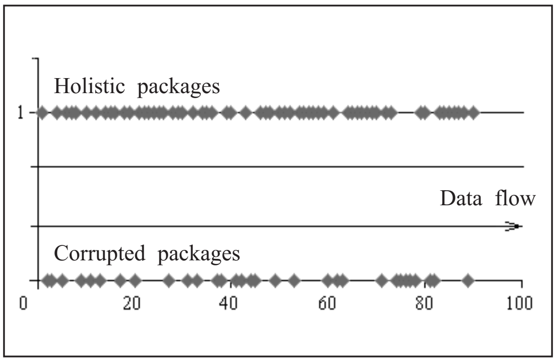

Fig. 4. The random flow generation with the error rate $m = 0.35$

And the binary zeros in the sample correspond to the arrival instants of damaged data packets for processing. The average distance of arrival time of damaged data packets determines the intensity parameter $\lambda = 1/m$.

The above numerical simulation of the binary stream of the integrity/error marker of data packets is illustrated in Figs. 2–4 graphical representations of data streams of the first 90 sample values (for clarity) for different values of error rate m $(0 \le m \le 1)$.

So, we can say that the "zero cycle" is simulated and the initial sample is obtained. To simulate the dynamics of the process of random errors of packet data transmission, we use the relations (2).

The process of modeling dynamics is described as follows.

Initial settings determination:

- N (sample size), K (number of stages);
- $\rho = 1 - m$ (equilibrium value).

The values $S_N(k)$, $k \ge 0$, are calculated according to the rule described at the beginning of this section, that is

- generate a sample u_n, $0 \le n \le N$, of uniformly distributed random variables on $[0, 1]$;
- generate a data flow simulation as follows:

$$\delta_n = \begin{cases} 0, & \text{if } u_n \le m, \\ 1 & \text{otherwise,} \end{cases} \quad 1 \le n \le N;$$

- calculate for k from 0 to K

$$S_N(k) = \frac{1}{N} \sum_{n=1}^{N} \delta_n(k);$$

- calculate the randomized dynamic flow by the formula

$$S_N(k+1) = S_N(k) - V[S_N(k) - \rho] = S_N(k)(1-V) + Vm. \quad (16)$$

A numerical implementation of the above algorithm gives the following results.

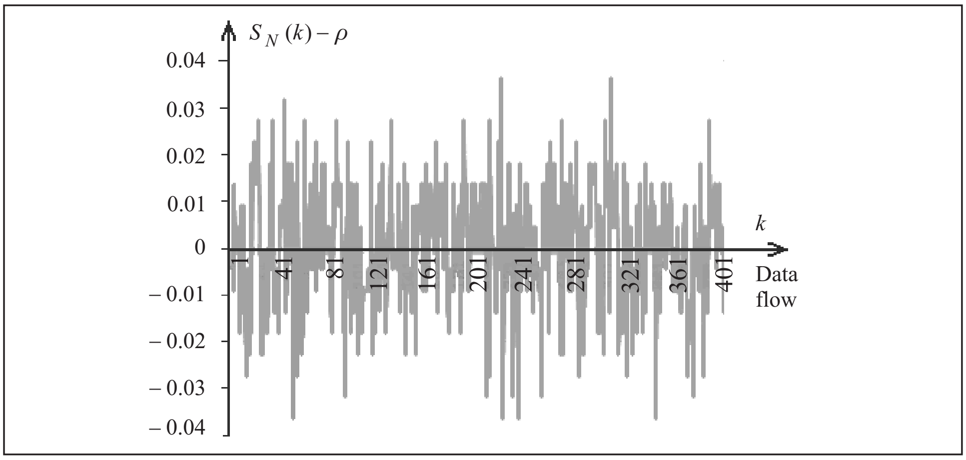

Fig. 5. The centered proportion of packet errors in the process of signal transmission (initial data: $N = 220$, $K = 400$, $m = 0.05$)

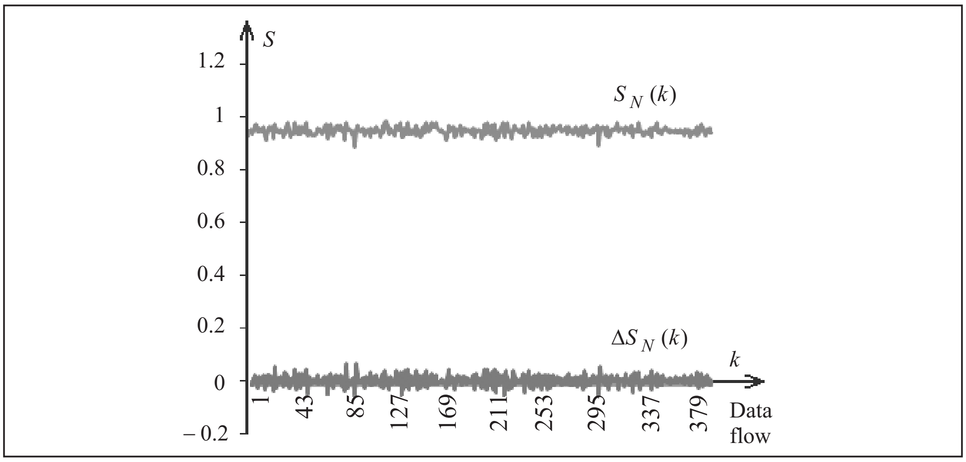

Fig. 6. Dynamics of random frequency of packet errors and its increments in the process of signal transmission (initial data: $N = 220$, $K = 400$, $m = 0.05$)

Table 1

Estimation of model parameters at			
$m = 0.05$		$m = 0.15$	
Initial data	Parameter estimates	Initial data	Parameter estimates
$N = 220$ $K = 400$	$V \approx 0.9739342$ $\sigma^2 \approx 0.0847564$ $\mathcal{E} \approx 0.9993206$	$N = 220$ $K = 400$	$V \approx 1.0253241$ $\sigma^2 \approx 0.2405275$ $\mathcal{E} \approx 0.9993587$

Fig. 5 illustrates the “diffusion type” of the behavior of a centered process of errors in the reception and transmission of data $[S_N(k) - \rho]$.

Fig. 6 represents the frequency paths of the packet errors $S_N(k)$ and their increments $\Delta S_N(k)$.

The obtained time series allow calculate the statistical parameters V, σ^2, $\mathcal{E}$ by (10)–(14). For $m = 0.05$ and $m = 0.15$ we have the following results (Table 1).

Thus, having a time series of the process of data transmitting errors, we can consider these series within the scheme of discrete Markov diffusion α_t, $t \geq 0$, which is a solution of stochastic difference equation (3) with parameters determined by the statistical estimates (11), (12) of parameters V, σ^2.

Using classification techniques [8], it is possible to test hypotheses about different type of limit frequencies (attracting states, repelling states, absorbing states). In connection with the classification theorems, it is of interest to consider the unsteady flow [9] of transmission and reception errors, as well as the Lyapunov stability problem [7].

It should be noted that a significant practical potential lies in the development of the proposed models of the dynamic centered process of errors that occur during the transmission of data presented in the form of multidimensional Markov random evolution [10, 11], as well as in pulsed processes with semi-Markov switchings [12].

STATISTICAL FEATURES OF THE ERROR FLOW DYNAMICS MODEL

Our proposed simulation of the dynamics of errors in packet data reception over the communication channels has a lag N of averaging the error frequency. The error rate dynamics $S_N(k)$ is studied by a discrete time parameter $k \geq 0$.

Stationarity and statistical equilibrium (2) determine the fundamental principle of the dynamics of "stimulation–deterrence": the error frequency process deviation $S_N(k+1) - \rho$ from the stationary value ρ at each stage $k+1$ decreases in proportion to $S_N(k)$ at this stage k with the coefficient of proportionality, set by the drift parameter V.

The use of statistical estimates (11), (12) of parameters V, σ^2 allows us to move from the generic model (1) to the model of discrete Markov diffusion (3).

As shown in [7, Ch. 3, 4], discrete Markov diffusion (3) is an effective tool for mathematical modeling and the corresponding numerical analysis of the process, determined by the normalized sum of random indicators of an attribute, in particular, the error process of packet data reception and transmission over communication channels.

REFERENCES

1. Довгий С.А. Современные телекоммуникации. Москва: Эко-Трендз, 2003. 320 с.
2. Довгий С.О., Воробієнко П.П., Гуляєв К.Д. Сучасні телекомунікації. Мережі, технології, безпека, економіка, регулювання. Київ: Азимут-Україна, 2013. 608 с.
3. Korolіouk D.V. Two component binary statistical experiments with persistent linear regression. *Theory of Probability and Mathematical Statistics*. 2015. N 90. P. 103-114.
4. Koroliouk D., Koroliuk V.S., Rosato N. Equilibrium processes in biomedical data analysis: the Wright–Fisher model. *Cybernetics and Systems Analysis.* 2014. Vol. 50, N 6. P. 80–86.
5. Korolyuk V.S., Koroliouk D. Diffusion approximation of stochastic Markov models with persistent regression. *Ukrainian Mathematical Journal.* 1995. Vol. 47, Iss. 7. P. 1065–1073.
6. Koroliouk D.V. Stationary statistical experiments and the optimal estimator for a predictable component. *Journal of Mathematical Sciences.* 2016. Vol. 214, N 2. P. 220–228.
7. Koroliouk D. Dynamics of Statistical Experiments. London: ISTE-Wiley, 2020. 224 p.
8. Bertotti M.L., Dovgyi S.O., Koroliouk D. Dynamics of ternary statistical experiments with equilibrium state. *Journ. of Computat. and Applied Math*. 2015. N 2 (119). P. 3–7.
9. Dovgii S.A., Shekhovtsov A.V. An improved Vortex Lattice Method for nonstationary problems. *Journal of Mathematical Sciences.* 2001. Vol. 104, Iss. 6. P. 1615–1627.
10. Samoilenko I.V. Asymptotic expansion for the functional of Markovian evolution in R^d in the circuit of diffusion approximation. *Journal of Applied Mathematics and Stochastic Analysis*. 2005. N 3. P. 247–257.
11. Moklyachuk M.P., Masyutka A.Yu. Extrapolation of multidimensional stationary processes. *Random Oper. and Stoch. Equ*. 2006. Vol. 14, N 3. P. 233–244.

12. Koroliuk V.S., Limnios N., Samoilenko I.V. Poisson approximation of impulsive recurrent process with semi-Markov switching. *Stochastic Analysis and Applications.* 2011. Vol. 29, Iss. 5. P. 769–778.

Надійшла до редакції 06.05.2020

С.О. Довгий, О.І. Юріков, М.О. Зозюк

ПРО ОДНУ СТАТИСТИЧНУ МОДЕЛЬ ЧАСТОТИ ПОМИЛОК У ПОТОЦІ ПАКЕТНОЇ ПЕРЕДАЧІ ДАНИХ КАНАЛАМИ ЗВ'ЯЗКУ

Анотація. Запропоновано статистичну модель частоти помилок у потоці пакетної передачі сигналу каналами зв'язку — стохастичну послідовність, що визначається як усереднена сума індикаторів помилкових пакетів даних. Застосовано дифузійне наближення такої послідовності — дискретну марковську дифузію, що визначається різницевим стохастичним рівнянням. Оцінювання параметрів моделі здійснено з використанням коваріаційних статистик за траєкторіями стохастичної послідовності помилок передачі сигналу.

Ключові слова: статистична модель, різницеве стохастичне рівняння, стаціонарний процес, рівновага, коваріаційна статистика, оцінка параметрів за траєкторіями.

С.А. Довгий, А.И. Юриков, М.О. Зозюк

ОБ ОДНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ ЧАСТОТЫ ОШИБОК В ПОТОКЕ ПАКЕТНОЙ ПЕРЕДАЧИ ДАННЫХ ПО КАНАЛАМ СВЯЗИ

Аннотация. Предложена статистическая модель частоты ошибок при передаче пакетных данных по каналам связи — стохастическая последовательность, определяемая как усредненная доля ошибочных пакетов данных. Использовано диффузионное приближение такой последовательности — дискретная марковская диффузия, которая определяется разностным стохастическим уравнением. Оценка параметров модели выполнена с использованием ковариационной статистики по траекториям стохастической последовательности ошибок передачи сигнала.

Ключевые слова: статистическая модель, разностное стохастическое уравнение, стационарный процесс, равновесие, ковариационная статистика, оценка параметров по траекториям.

Dovgyi Stanislav Oleksiyovich,
Dr. of Sciences, Academician of NAS of Ukraine, Dpt. Director of the Institute of Telecommunications and Global Information Space of NAS of Ukraine, Kyiv, e-mail: pryjmalnya@gmail.com.

Yurikov Olexiy Ivanovich,
Ph.D. Student, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, e-mail: yurikov.alex03@gmail.com.

Zozyuk Maxim Olegovich,
Ph.D. Student, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, e-mail: maksym.zoziuk@gmail.com.