The Simulation of the Trend of the Time Series in the form of the Spline of Third-Order With a Random Number of Data at the Moments of Measurement

I Ustinova, E Pakchomova

Tomsk Polytechnic University, Tomsk, Russia

E-mail: igu@tpu.ru, peg@tpu.ru

Abstract. The possibility of simulation of trend of a time series as a spline of third-order with a random number of data at the moments of measurement is discussed. Estimations of coefficients of the spline are obtained in the explicit form. The statistical characteristics of the received estimations are studied in details.

1. Introduction

One of the crucial problems of the analysis of the functioning of complicated technical systems is the simulation of the trend of telemetric parameters. Telemetric parameters are characterized the state of the system in some discrete moments of time [1]. The choice of a mathematical model and the definition of a scheme of an observation of the process are the main aspect of solve of this task. Mathematical model should describe the trend (a systematic component) of the observed values of a random process adequately. Here are the general ways of the simulation of the trend of the time series:

1. In the form of the polynomial p-degree of time t. Taking p large enough, one can get a fairly accurate representation of the trend of the time series. The value of the parameter p is selected separately in each case.
2. The curves which reflect "the net growth" can be described by exponential functions.
3. The curves of the population growth can be described by the logistic-curve in the form $f(t) = k/(1 + e^{-\lambda t})$ [2].
4. As a linear combination of known functions.
5. In the form of splines.

In the classical theory of the time series the measurements are made thru equal time intervals and only one measurement is made at each moment time [2, 3]. However, there is another scheme of measurements in which at each time is produced a random number of measurements [4, 5]. Note that the simulation of the trend in a polynomial form whose order is higher than four is impractical due to the big error in the estimation of coefficients of the polynomial. On the other hand if the number of observations is large, the polynomial of the low-order can describe the true trend unsatisfactorily. The solution is in the estimation in the form of the spline of the trend of the time series. Generally, the spline is determined as follows [6-8]. The function $S(t)$ defined and continuous on the interval $[a, b]$

called a polynomial spline of the order m with nodes $x_j \in \{a = x_0 < x_1 < x_2 < \ldots < x_n = b\}$ if at each of the intervals $[x_{j-1}, x_j)$, $j = \overline{1,n}$, $S(t)$ is an algebraic polynomial of degree m.

2. The statement of the problem

Let there is the time series $y(t) = f(t) + \xi(t)$ representing the sum of some determined function $f(t)$ which is the trend of the process $y(t)$ and $\xi(t)$ is the stochastic function which emergence is defined by errors of measurements, hindrances etc. We will believe that hindrances of measurements $\xi_i = \xi(t_i)$, i=1, 2,…, are the independent, equally distributed random variables with the mathematical waiting $M[\xi_i] = 0$ and dispersion $D[\xi_i] = \sigma^2$. We know the sequence of values $y_1, y_2, ..., y_N$, where $y_i = \dfrac{y_{i1} + y_{i2} + ... + y_{in_i}}{n_i}$. Here n_i are independent a random expression, distributed by the Poisson law with an argument λ. We assume that the time points t_i are known to us exactly. Concerning the trend it is supposed that it represents the spline of the third order. In this case, the entire period of the observation $[\,0; T\,]$ is divided into parts of equal length: $[0; T_0]$, $[T_0; 2T_0]$, ..., $[(n-1)T_0; nT_0]$. On each such interval, the trend is estimated in the form of the polynomial of the third degree. On the borders of the segments, these polynomials are sewn together so as to obtain a continuous curve. Such the piecewise polynomial curve is called by the spline [5-9].

3. The solution of the problem

Let on an interval number k the trend can be presented in the form of the polynomial of the third order $f_k(t) = a_k + b_k(t/T_0) + c_k(t/T_0)^2 + d_k(t/T_0)^3$. We will believe that on each time interval a counting of time is conducted from the beginning of this interval therefore for a sewing together of polynomials on the ends of intervals the condition $f_{k-1}(T_0) = f_k(0)$ should be satisfied, that is $a_k = a_{k-1} + b_{k-1} + c_{k-1} + d_{k-1}$. By results of measurements of values $y_i^{(k)}$ on the interval number k it is necessary to construct estimates of coefficients a_k, b_k, c_k and d_k, moreover, there is a condition:

$$\hat{a}_k = \hat{a}_{k-1} + \hat{b}_{k-1} + \hat{c}_{k-1} + \hat{d}_{k-1} \quad . \tag{1}$$

Estimates of missed coefficients for the interval number k we find by the method of least squares, based on the conditions:

$$Q_k = \sum_{i=1}^{N} n_i \left[y_i^{(k)} - \left(\hat{a}_k + b_k (t_i/T_0) + c_k (t_i/T_0)^2 + d_k (t_i/T_0)^3 \right) \right]^2 \Rightarrow \min_{b_k, c_k, d_k}, \tag{2}$$

N is the number of measurements at the interval of number k. Differentiating the equation (2) on variables b_k, c_k, d_k, we will receive system of the equations

$$\begin{cases} (b_k/T_0)\sum_{i=1}^{N} n_i t_i^2 + (c_k/T_0^2)\sum_{i=1}^{N} n_i t_i^3 + (d_k/T_0^3)\sum_{i=1}^{N} n_i t_i^4 = \sum_{i=1}^{N} n_i \left(y_i^{(k)} - \hat{a}_k\right) t_i, \\ (b_k/T_0^2)\sum_{i=1}^{N} n_i t_i^3 + (c_k/T_0^3)\sum_{i=1}^{N} n_i t_i^4 + (d_k/T_0^4)\sum_{i=1}^{N} n_i t_i^5 = \dfrac{1}{T_0}\sum_{i=1}^{N} n_i \left(y_i^{(k)} - \hat{a}_k\right) t_i^2, \\ (b_k/T_0^3)\sum_{i=1}^{N} n_i t_i^4 + (c_k/T_0^4)\sum_{i=1}^{N} n_i t_i^5 + (d_k/T_0^5)\sum_{i=1}^{N} n_i t_i^6 = \dfrac{1}{T_0^2}\sum_{i=1}^{N} n_i \left(y_i^{(k)} - \hat{a}_k\right) t_i^3. \end{cases}$$

from which we find the explicit form of the estimates of the missing parameters:

$$\hat{b}_k = \frac{30T_0}{\lambda N(N^2-1)(N^2-4)(N+3)(N^2+N+3)} \cdot \left\{ \frac{5(4N^6+12N^5+4N^4-12N^3-11N^2-3N+2)}{2N+1} \cdot S_1 - \right.$$
$$-10N(3N^3+6N^2-N-4)\cdot S_2 + \frac{7(6N^4+12N^3+3N^2-3N+2)}{2N+1}\cdot S_3 -$$
$$-\left[\frac{5(4N^6+12N^5+4N^4-12N^3-11N^2-3N+2)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i - 10N(3N^3+6N^2-N-4)\cdot\sum_{i=1}^{N} n_i t_i^2 + \right.$$
$$\left. \left. + \frac{7(6N^4+12N^3+3N^2-3N+2)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i^3 \right]\hat{a}_k \right\},$$

$$\hat{c}_k = \frac{60T_0^2}{\lambda N(N^2-1)(N^2-4)(N+3)(N^2+N+3)} \cdot \{-5N(N+1)(3N^2+3N-4)\cdot S_1 + 6(8N^3+12N^2-2N-3)\cdot S_2$$
$$-35N(N+1)\cdot S_3 - \left[-5N(N+1)(3N^2+3N-4)\cdot\sum_{i=1}^{N} n_i t_i + 6(8N^3+12N^2-2N-3)\cdot\sum_{i=1}^{N} n_i t_i^2 - 35N(N+1)\cdot \right.$$
$$\left. \cdot\sum_{i=1}^{N} n_i t_i^3 \right]\hat{a}_k \},$$
(3)

$$\hat{d}_k = \frac{210T_0^3}{\lambda N(N^2-1)(N^2-4)(N+3)(N^2+N+3)} \cdot \left\{ \frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot S_1 - 10N(N+1)\cdot S_2 + \right.$$
$$+\frac{5(3N^2+3N+2)}{2N+1}\cdot S_3 - \left[\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i - 10N(N+1)\cdot\sum_{i=1}^{N} n_i t_i^2 + \right.$$
$$\left. \left. +\frac{5(3N^2+3N+2)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i^3 \right]\hat{a}_k \right\},$$

where $S_1 = \sum_{i=1}^{N} n_i y_i^{(k)} t_i$, $S_2 = \sum_{i=1}^{N} n_i y_i^{(k)} t_i^2$, $S_3 = \sum_{i=1}^{N} n_i y_i^{(k)} t_i^3$. Replacing in (3) evaluation a_k by the expression (1), we obtain the system for determining the coefficients of the spline on the interval number *k*. That is $\hat{a}_k = \hat{a}_{k-1} + \hat{b}_{k-1} + \hat{c}_{k-1} + \hat{d}_{k-1}$,

$$\hat{b}_k = \frac{30T_0}{\lambda N(N^2-1)(N^2-4)(N+3)(N^2+N+3)} \cdot \left\{ \frac{5(4N^6+12N^5+4N^4-12N^3-11N^2-3N+2)}{2N+1} \cdot S_1 - \right.$$
$$-10N(3N^3+6N^2-N-4)\cdot S_2 + \frac{7(6N^4+12N^3+3N^2-3N+2)}{2N+1}\cdot S_3 -$$
$$-\left[\frac{5(4N^6+12N^5+4N^4-12N^3-11N^2-3N+2)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i - 10N(3N^3+6N^2-N-4)\cdot\sum_{i=1}^{N} n_i t_i^2 + \right.$$
$$\left. \left. + \frac{7(6N^4+12N^3+3N^2-3N+2)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i^3 \right]\left(\hat{a}_{k-1} + \hat{b}_{k-1} + \hat{c}_{k-1} + \hat{d}_{k-1}\right) \right\},$$
(4)

IOP Conf. Series: Materials Science and Engineering **142** (2016) 012036 doi:10.1088/1757-899X/142/1/012036

$$\hat{c}_k = \frac{60T_0^2}{\lambda N\left(N^2-1\right)\left(N^2-4\right)\left(N+3\right)\left(N^2+N+3\right)} \cdot \Big\{-5N\left(N+1\right)\left(3N^2+3N-4\right)\cdot S_1 + 6\left(8N^3+12N^2-2N-3\right)\cdot S_2$$
$$-35N\left(N+1\right)\cdot S_3 - \Big[-5N\left(N+1\right)\left(3N^2+3N-4\right)\cdot\sum_{i=1}^{N} n_i t_i + 6\left(8N^3+12N^2-2N-3\right)\cdot\sum_{i=1}^{N} n_i t_i^2 - 35N\left(N+1\right)\cdot$$
$$\cdot\sum_{i=1}^{N} n_i t_i^3\Big]\left(\hat{a}_{k-1}+\hat{b}_{k-1}+\hat{c}_{k-1}+\hat{d}_{k-1}\right)\Big\},$$

$$\hat{\mathrm{d}}_k = \frac{210T_0^3}{\lambda N\left(N^2-1\right)\left(N^2-4\right)\left(N+3\right)\left(N^2+N+3\right)} \cdot \left\{\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot S_1 - 10N\left(N+1\right)\cdot S_2 + \right.$$
$$+\frac{5\left(3N^2+3N+2\right)}{2N+1}\cdot S_3 - \left[\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i - 10N\left(N+1\right)\cdot\sum_{i=1}^{N} n_i t_i^2 + \right.$$
$$\left.\left.+\frac{5\left(3N^2+3N+2\right)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i^3\right]\left(\hat{a}_{k-1}+\hat{b}_{k-1}+\hat{c}_{k-1}+\hat{d}_{k-1}\right)\right\}.$$

$$\hat{\mathrm{d}}_k = \frac{210T_0^3}{\lambda N\left(N^2-1\right)\left(N^2-4\right)\left(N+3\right)\left(N^2+N+3\right)} \cdot \left\{\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot S_1 - 10N\left(N+1\right)\cdot S_2 + \right.$$
$$+\frac{5\left(3N^2+3N+2\right)}{2N+1}\cdot S_3 - \left[\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i - 10N\left(N+1\right)\cdot\sum_{i=1}^{N} n_i t_i^2 + \right.$$
$$\left.\left.+\frac{5\left(3N^2+3N+2\right)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i^3\right]\left(\hat{a}_{k-1}+\hat{b}_{k-1}+\hat{c}_{k-1}+\hat{d}_{k-1}\right)\right\}.$$

$$\left\{\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot S_1 - 10N\left(N+1\right)\cdot S_2 + \right.$$
$$+\frac{5\left(3N^2+3N+2\right)}{2N+1}\cdot S_3 - \left[\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i - 10N\left(N+1\right)\cdot\sum_{i=1}^{N} n_i t_i^2 + \right.$$
$$\left.\left.+\frac{5\left(3N^2+3N+2\right)}{2N+1}\cdot\sum_{i=1}^{N} n_i t_i^3\right]\left(\hat{a}_{k-1}+\hat{b}_{k-1}+\hat{c}_{k-1}+\hat{d}_{k-1}\right)\right\}.$$

Now we investigate statistical characteristics of the received estimates, namely, we will prove their asymptotic unbiasedness and sustainability. Because $E[y_i^{(k)}] = a_k + b_k(t/T_0) + c_k\left(t/T_0\right)^2 + d_k\left(t/T_0\right)^3$ ($E[x]$ is the mathematical expectation of a random variable x)

$$E[S_1|n_i] = \frac{\lambda N(N+1)}{2}\left\{a_k + \frac{2N+1}{3T_0}b_k + \frac{N(N+1)}{2T_0^2}c_k + \frac{(2N+1)(3N^2+3N-1)}{15T_0^3}d_k\right\}$$

$$E[S_2|n_i] = \frac{\lambda N(N+1)}{2}\left\{\frac{2N+1}{3}a_k + \frac{N(N+1)}{2T_0}b_k + \frac{(2N+1)(3N^2+3N-1)}{15T_0^2}c_k + \right.$$
$$\left. + \frac{N(N+1)(2N^2+2N-1)}{6T_0^3}d_k\right\},$$

$$E[S_3|n_i]=\frac{\lambda N(N+1)}{2}\left\{\frac{N(N+1)}{2}a_k+\frac{(2N+1)(3N^2+3N-1)}{15T_0}b_k+\frac{N(N+1)(2N^2+2N-1)}{6T_0^2}c_k+\right.$$
$$\left.+\frac{(2N+1)(3N^4+6N^3-3N+1)}{21T_0^3}d_k\right\},$$

$$b_k=\frac{15T_0}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{\frac{10(4N^6+12N^5+4N^4-12N^3-11N^2-3N+2)}{\lambda N(N+1)(2N+1)}M[S_1|n_i]+\right.$$
$$+\frac{20(3N^2+3N-4)}{\lambda}M[S_2|n_i]+\frac{14(6N^4+12N^3+3N^2-3N+2)}{\lambda}M[S_3|n_i]-$$
$$\left.-\frac{486N^6+1458N^5+895N^4-640N^3-691N^2-128N+60}{6(2N+1)}a_k\right\},$$

$$c_k=\frac{15T_0^2}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{\frac{20(3N^2+3N-4)}{\lambda}M[S_1|n_i]+\frac{24(4N^2-1)(2N+3)}{\lambda N(N+1)}M[S_2|n_i]-\right.$$
$$\left.-\frac{140}{\lambda}M[S_3|n_i]-\left(59N^4+118N^3-13N^2-72N-12\right)a_k\right\},$$

$$d_k=\frac{15T_0^3}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{\frac{14(6N^4+12N^3+3N^2-3N+2)}{\lambda N(2N^2+3N+1)}M[S_1|n_i]-\right.$$
$$\left.-\frac{140}{\lambda}M[S_2|n_i]+\frac{70(3N^2+3N+2)}{\lambda N(N+1)(2N+1)}M[S_3|n_i]-\frac{7(N-1)(N^2-4)(N+3)}{6(2N+1)}a_k\right\}.$$

We will designate: $\Delta a_k=\hat{a}_k-a_k$, $\Delta b_k=\hat{b}_k-b_k$, $\Delta c_k=\hat{c}_k-c_k$, $\Delta d_k=\hat{d}_k-d_k$ By averaging over the random variables n_i we get the following system:

$$\Delta\overline{a}_k=\Delta\overline{a}_{k-1}+\Delta\overline{b}_{k-1}+\Delta\overline{c}_{k-1}+\Delta\overline{d}_{k-1},$$

$$\Delta\overline{b}_k=\frac{-15T_0}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{\frac{5(4N^6+12N^5+4N^4-12N^3-11N^2-3N+2)}{(2N+1)}-\right.$$
$$\left.-\frac{10N(3N^3+6N^2-N-4)}{3}(2N+1)+\frac{7(6N^4+12N^3+3N^2-3N+2)N(N+1)}{2N+1}\right\}\cdot$$
$$\cdot\left(\Delta\overline{a}_{k-1}+\Delta\overline{b}_{k-1}+\Delta\overline{c}_{k-1}+\Delta\overline{d}_{k-1}\right),$$

$$\Delta\overline{c}_k=\frac{-30T_0^2}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{-5N(N+1)(3N^2+3N-4)+2(8N^3+12N^2-2N-3)\cdot\right.$$
$$\left.\cdot(2N+1)-\frac{35N^2(N+1)^2}{2}\right\}\cdot\left(\Delta\overline{a}_{k-1}+\Delta\overline{b}_{k-1}+\Delta\overline{c}_{k-1}+\Delta\overline{d}_{k-1}\right),$$

$$\Delta\overline{d}_k=\frac{-105T_0^3}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{\frac{6N^4+12N^3+3N^2-3N+2}{2N+1}-\right.$$
$$\left.-\frac{10N(N+1)(2N+1)}{3}+\frac{5(3N^2+3N+2)N(N+1)}{2(2N+1)}\right\}\cdot\left(\Delta\overline{a}_{k-1}+\Delta\overline{b}_{k-1}+\Delta\overline{c}_{k-1}+\Delta\overline{d}_{k-1}\right).$$

For asymptotically unbiased of the estimates it is necessary that this system was stable. This is possible if the roots of its characteristic equation (5) will be on the module less units.

$$\begin{vmatrix} 1-\lambda & 1 & 1 & 1 \\ a & a-\lambda & a & a \\ b & b & b-\lambda & b \\ c & c & c & c-\lambda \end{vmatrix} = 0 \tag{5}$$

In this equation the following designations are entered:

$$a = \frac{-15T_0}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{\frac{5(4N^6+12N^5+4N^4-12N^3-11N^2-3N+2)}{(2N+1)} - \right.$$
$$\left. -\frac{10N(3N^3+6N^2-N-4)}{3}(2N+1)+\frac{7(6N^4+12N^3+3N^2-3N+2)N(N+1)}{2(2N+1)}\right\},$$

$$b = \frac{-30T_0^2}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{-5N(N+1)(3N^2+3N-4)+2(8N^3+12N^2-2N-3)\cdot\right.$$
$$\left.\cdot(2N+1)-\frac{35N^2(N+1)^2}{2}\right\},$$

$$c = \frac{-105T_0^3}{(N-1)(N^2-4)(N+3)(N^2+N+3)}\left\{\frac{6N^4+12N^3+3N^2-3N+2}{2N+1} - \right.$$
$$\left. -\frac{10N(N+1)(2N+1)}{3}+\frac{5(3N^2+3N+2)N(N+1)}{2(2N+1)}\right\}.$$

Three roots $\lambda_1, \lambda_2, \lambda_3$ of the equation (5) are equal to zero and $\lambda_4 = 1+a+b+c$, that is $\lambda_4 = 1 - \frac{5T_0}{2(N^2+N+3)}\left\{\frac{7T_0^2}{2N+1} - 6T_0 + \frac{6N^2+6N+5}{2N+1}\right\}$. Its value is determined by the expression N/T_0, in which N is the number of points in the interval $[(k-1)T_0; kT_0]$. For recursive estimation of the coefficients of the spline of the third order requires at least three points on each partial interval. By solving the inequality $|\lambda_4| < 1$, we see that it is true for any $N/T_0 > 0$ (figure 1). Thus, the recursive estimates (4) of parameters of the spline are sustainable, and therefore asymptotically unbiased.

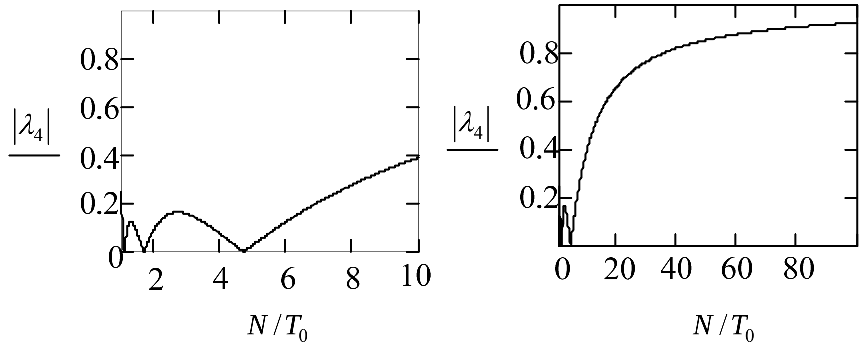

Figure 1. The dependence of the fourth root of the characteristic equation from N/T_0.

4. Conclusion

As a result of the conducted research are obtained by the recurrent evaluation of the coefficients of the spline of the third order in the modelling of the trend of the time series with random number of data at

IOP Conf. Series: Materials Science and Engineering **142** (2016) 012036 doi:10.1088/1757-899X/142/1/012036

the moments of measurement. The stability of the estimates of the parameters of the spline and their asymptotic unbiased are installed.

References

[1] Trivozhenko B E 1989 *Identification of Trends of Time Series and Flows of Events* (Tomsk: Tomsk state University).
[2] Kendall M G, Stuart A 1966 *The advanced theory of statistics. Design and analysis and time-series* (London: Published by Griffin & Company).
[3] Anderson T W 1971 *Statistical analysis of time series* (New York: Wiley and Sons).
[4] Idrisov F F, Konstantinova I G 1999 Identification of trends of time series for a random number of measurements *Russian Physics Journal* **42** 379 384.
[5] Ustinova I G , Pakhomova E G 2015 Spline estimate of the time series trend for a random number of data at measurement instants *Vestn. Tomsk. Gos. Univ. Mat. Mekh.***1(33)** 20 36.
[6] Bakhvalov N S 1975 *Numerical methods* (Moscow: Nauka).
[7] Korneychuk N P 1984 *Splines in approximation theory* (Moscow: Nauka).
[8] Shevelitskiy I V 2003 Interpolation splines in problems of digital signal processing *J. Exponenta Pro. Matematika v prilozheniyakh* **4** 42 53.
[9] March L C, Cormier D R 2002 *Spline Regression Models* (London: SAGE PUBLICATION Ltd).