Colloquium Biometricum 40
2010, 135–145

ON INFLUENCE OF VARIABILITY IN CONCOMITANT VARIABLES VALUES ON ESTIMATION OF POLYNOMIAL COEFFICIENTS IN GROWTH CURVES MODELS WITH CONCOMITANT VARIABLES CHANGING IN TIME AND THE SAME VALUES FOR ALL EXPERIMENTAL UNITS

Andrzej Bochniak, Mirosława Wesołowska-Janczarek

Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin
Akademicka 13, 20-950 Lublin
e-mail: andrzej.bochniak@up.lublin.pl, miroslawa.wesolowska@up.lublin.pl

Summary

Some concomitant variables with values changing in time can have significant influence on estimation of polynomials which describe course of changes for a studied feature in a given time interval. The situation where values of concomitant variables in successive time points are the same for all experimental unit is considered. A difference between values of these variables in time points can be small or large. The question is if exactness of curve estimate is depending on magnitude of these differences. The investigation is carried out on the data obtained by computer simulation.

Key words and phrases: growth curve methods, concomitant variables, linear mathematical models, parameter estimation methods, computer simulation

Classification AMS 2010: 62J12, 68U20

1. Introduction

The course of change of a studied feature in time for different groups of units can be described using the known growth curve method given by Potthoff and Roy (1964). Frequently, time is not only one reason of studied feature values changes. There are different variables, called concomitant variables, whose values can be measured too. For example: if a studied feature is the growth of plants in time, then concomitant variables can be the rainfall value and the air temperature in suitable time points. In the case considered here all values of those variables are the same for all plants.

The model that is suitable in this case, was provided in the work by Wesołowska-Janczarek and Fus (1996). It also presents an iterative method of parameter estimation in that model.

It is interesting to formulate the questions whether to estimate the course of change of feature influence of concomitant variables ought to be ever taken into consideration or else if the influence of concomitant variable can be omitted, then appointed curve be good or bad estimator of real curve.

The second question is whether the degree of variation of concomitant variables values in time is influencing the conformity of estimated function to true one.

The research presented in this paper is the introduction to solving these problems.

2. Iterative method of parameters estimation

One of the growth curve models with concomitant variables, when the values of *s* concomitant variables are the same for all experimental units, but each of the variables values are different in observed *p* time points, is considered here. This one was given by Wesołowska-Janczarek and Fus (1996) in the following form:

$$Y = ABT + J_n \gamma' X + E \tag{2.1}$$

where Y is $(n \times p)$ random matrix of observation of studied feature on n experimental units in p time points, A is $(n \times a)$ - matrix, which divides n units on a groups, B is $(a \times q)$ matrix of unknown coefficients in searched polynomial growth curves of q-1 degree, T is $(q \times p)$ - matrix that include the successive powers of time points form 0 to q-1 (it is Vandermond's matrix) and define internal structure of observations, X is $(s \times p)$ matrix of values of these s variables in successive p time points, γ is a vector of s regression coefficients at concomitant variables, J_n is a vector of n ones and E is a $(n \times p)$ matrix of random errors.

Under the assumption of matrix variate normal distribution of Y denoted by $Y \sim N_{n,p}(ABT + J_n\gamma' X, I_n \otimes \Sigma)$ and $\Sigma(p \times p) > 0$ (see Gupta and Nagar (1999), p. 55) assumptions estimators of parameters in this model obtained by maximum likelihood method were given in following form:

$$\begin{aligned}
n\hat{\Sigma} &= (Y - A\hat{B}T - J_n\hat{\gamma}'X)'(Y - A\hat{B}T - J_n\hat{\gamma}'X) \\
\hat{B}_{\hat{\Sigma}} &= (A'A)^{-}A'(Y - J_n\hat{\gamma}'X)\hat{\Sigma}^{-1}T'(T\hat{\Sigma}^{-1}T')^{-1} \\
\hat{\gamma}'_{\hat{\Sigma}} &= [J'_nY - J'_nA(A'A)^{-}A'Y\hat{\Sigma}^{-1}T'(T\hat{\Sigma}^{-1}T')^{-1}T]\hat{\Sigma}^{-1}X'R_{\hat{\Sigma}} \\
R_{\hat{\Sigma}} &= [nX\hat{\Sigma}^{-1}X' - J'_nA(A'A)^{-}A'J_nX\hat{\Sigma}^{-1}T'(T\hat{\Sigma}^{-1}T')^{-1}T\hat{\Sigma}^{-1}X']^{-1}.
\end{aligned} \tag{2.2}$$

The values of these estimators can be calculated by the iterative method where in the first step the following form of a matrix will be taken $n\hat{\Sigma} = Y'[I_n - A(A'A)^{-1}A']Y$.

We are interested in the answer to two questions. The first one is: how a changeability in concomitant variables values in observable time points is influencing the correctness of growth curve estimation and the second one is whether the concomitant variables omission in the model gives the estimator of true curve describing changes of feature in time that are better or worse fitting.

The investigations presented here were undertaken using the data obtained by a computer simulation. The method of simulation is presented in next part of this paper.

3. Computer simulation method

The computer simulation was conducted in *Matlab* programme. The values of observations obtained in the true experiment were taken as the basis for the computer simulation. In the experiment described by Wesołowska-Janczarek and Fus (1996) the yielding of 16 raspberry varieties was examined where values of meteorological elements such as air temperature, sunshine and rainfall were taken as concomitant variables.

The values of parameters (such as covariance matrix Σ, matrix B of polynomials coefficient, averages and standard deviations of concomitant variables given by matrix X and vector γ of regression coefficient at concomitant variables) which appear in the model given by equation (2.1) were estimated from original data.

In the second step on the basis of estimators calculated in the first step new values were generated for parameters: positive defined matrix Σ, coefficients matrix B for polynomials with shapes given by averaged values for all estimated polynomials and vector γ of regression coefficients. It was assumed during the computer simulation that the influence of concomitant variables should not exceed 30 % of the examined feature value.

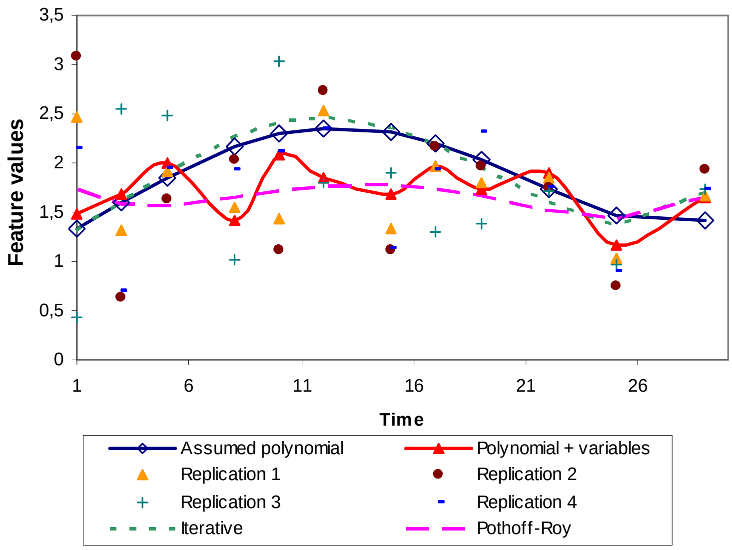


Fig. 1. Generated observations and estimations of assumed curve by iterative and Potthoff-Roy's methods for example simulation

Next, for such established parameters, values of matrix X containing information about concomitant variables (on the basis of assumed means and standard deviations) were repeatedly (1000 times) generated and observation matrix Y by rows from distribution $N_{n,p}(ABT + J_n\gamma' X;\ I_n \otimes \Sigma)$. Estimators of polynomials coefficients were next calculated for each matrix Y and compared with polynomials established in previous step.

Different variation of concomitant variables values in time were obtained by multiplying the standard deviations for experimental data by values 0.1; 0.5; 1; 2; 5. The same polynomials were taken while changing standard deviations of concomitant variables values so easier influence of variability of them can be observed.

Arrangement of time points and matrix *A* dividing experimental units into groups were kept unchanged in comparison to the conducted experiment.

Fig. 1 presents an exemplary generated polynomial, curve with considered concomitant variables, generated observations for 4 replication and finally estimation of polynomial growth curve calculated by Potthoff-Roy's and iterative methods.

4. Results

The results obtained in all conducted simulations are similar to those presented in the examples in further part of paper. The iterative estimation method described in second part was used to the generated data. As an example of the obtained results one group, named variety 1, was chosen. Fig. 2 shows estimated polynomials calculated by iterative method with different variability of concomitant variables values i.e. the elements of matrix *X*. Standard deviations of generated data x_{ij} was decreased or increased in comparison to the ones calculated for original experimental data by multiplication by following values: a) 0.1 – the least diversity; b) 0.5; c) 1 – diversity as in the original data; d) 2; e) 5 – the greatest diversity of concomitant variables values.

On presented charts the assumed polynomial (basis for generation of data in matrix *Y*) for variety 1 is marked as the dark line. Lighter lines show the first 50 estimations of a given polynomial from all 1000 repetitions made for generation of matrix *X* of concomitant variables values and matrix of observations *Y*. The remaining parameter values: matrix *B* of polynomials coefficients, covariance matrix Σ and regression vector γ were the same in any case.

In addition, the same scale on vertical axes was taken in all charts. This helps easier observation of dependence of curve estimation on diversity of values of concomitant variables, but it can suggest little variability in shape of assumed polynomial. It can be seen as an almost straight line due to large values on first chart. The exact shapes of all 16 polynomials assumed in generations are shown on fig. 3.

It can be easily seen that the lesser variability in concomitant variables values in time, the less exact estimation of the assumed curve in the iterative method with regard to the fixed regression dependence on concomitant variables i.e. is fixed elements in vector γ.

Numerical values showing averaged sums of modulus of relative errors in subsequent time points for each variety (i.e. averaged differences between assumed polynomials used to generate observation and estimated polynomials by iterative method) are presented in table 1. One can also see (from fig. 3 and table 1) that shape of polynomials used in simulations has also influence on the estimation of the curve. If the shape is more diverse (variety 1) then the accuracy of estimation is better and for the one that is the least diverse (variety 7) estimated curves are the worst fitted to assumed polynomial.

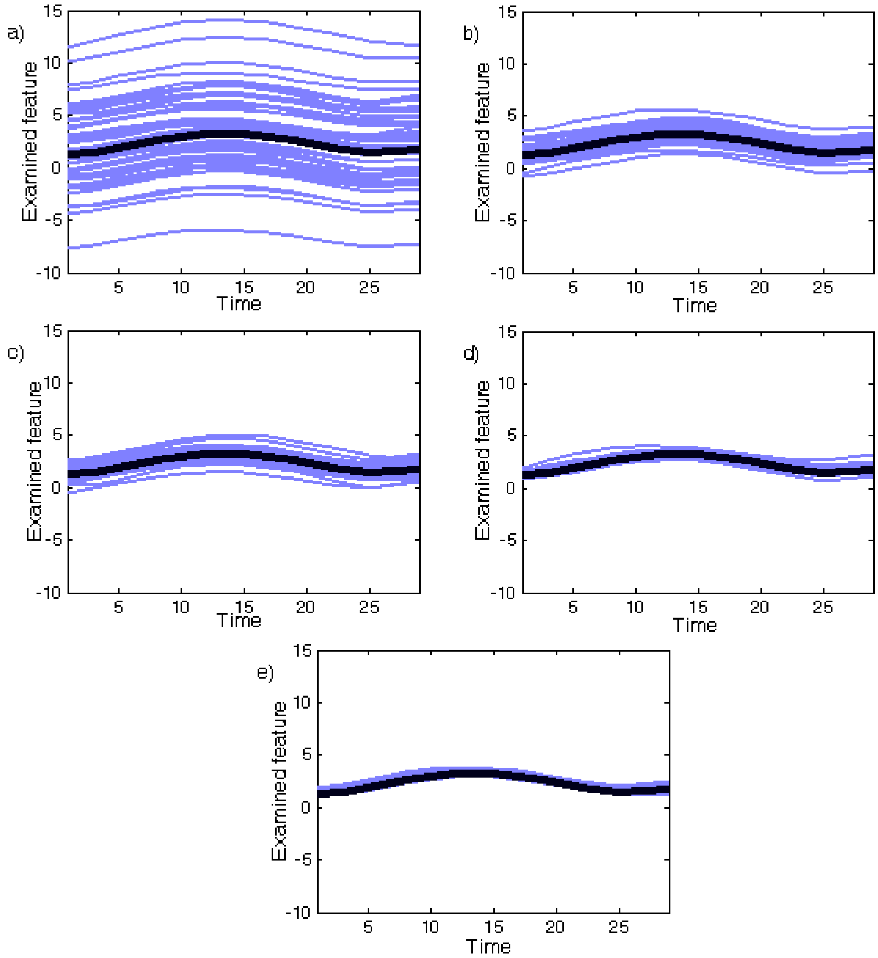


Fig. 2. Estimation of assumed growth curve by iterative method with different variability of standard deviations of concomitant variables values. Standard deviations from experiment was multiplied by a) 0.1 b) 0.5 c) 1 d) 2 e) 5 for generated data.

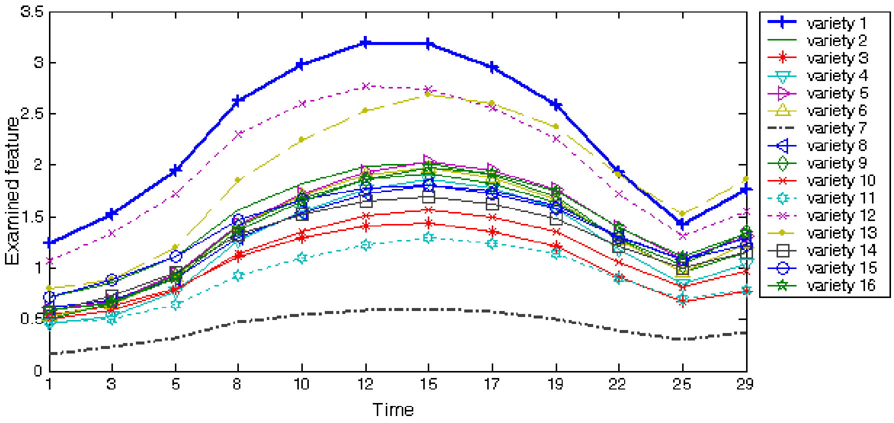
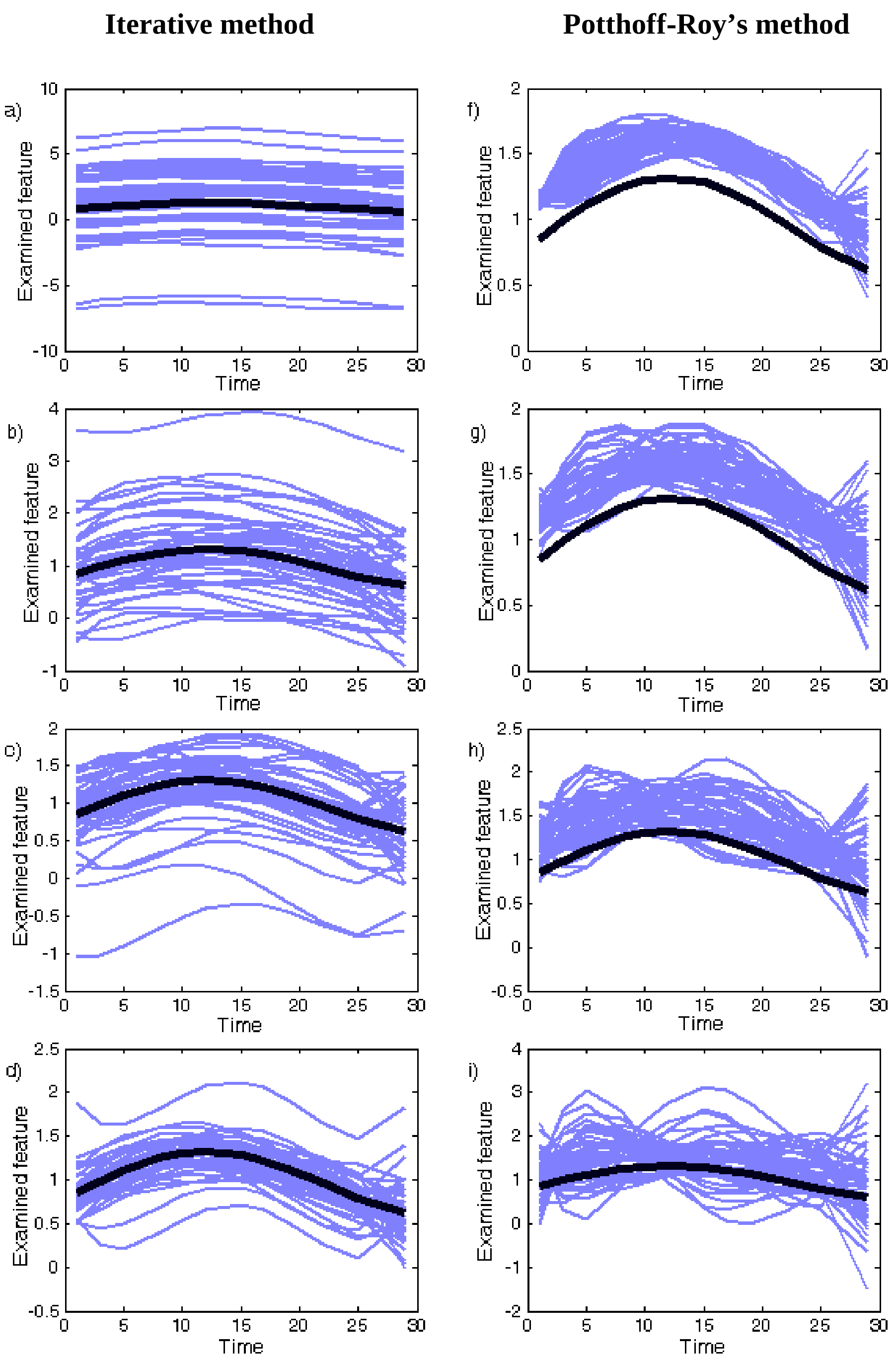


Fig. 3. Exemplary polynomials assumed in simulations

Table 1. Averaged sums of modulus of relative error for estimated curves. σ – vector of standard deviations for concomitant variables values in experimental data

	Standard deviations assumed during generation of concomitant variables values				
	0.1·σ	0.5·σ	1·σ	2·σ	5·σ
Variety 1	19.43	3.60	2.05	1.15	0.73
Variety 2	31.58	5.86	3.33	1.87	1.17
Variety 3	43.94	8.15	4.64	2.56	1.65
Variety 4	40.28	7.45	4.22	2.34	1.48
Variety 5	33.45	6.21	3.51	1.95	1.24
Variety 6	35.33	6.53	3.70	2.05	1.30
Variety 7	110.04	20.43	11.54	6.47	4.08
Variety 8	34.62	6.41	3.65	2.03	1.26
Variety 9	35.45	6.56	3.72	2.06	1.31
Variety 10	41.65	7.72	4.40	2.45	1.53
Variety 11	49.19	9.11	5.17	2.89	1.81
Variety 12	22.10	4.10	2.33	1.30	0.83
Variety 13	24.99	4.63	2.62	1.45	0.92
Variety 14	35.89	6.66	3.77	2.11	1.33
Variety 15	31.87	5.90	3.36	1.86	1.19
Variety 16	33.84	6.26	3.54	1.97	1.25


Iterative method
Potthoff-Roy's method
a)
b)
c)
d)
f)
g)
h)
i)
Examined feature
Time

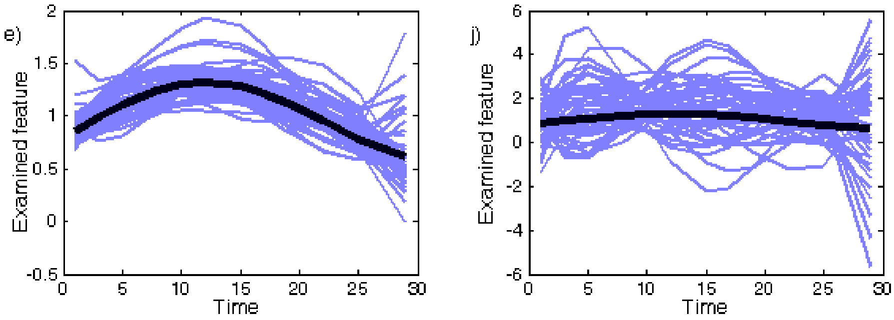


Fig. 4. Comparison of estimation of assumed curve in iterative (a-e) and Potthoff-Roy's (f-j) methods depending on different variability of concomitant variables values. Vector of standard deviations for concomitant variables from experiment is multiplied by 0.1; 0.5; 1; 2; 5 (from top to bottom)

At the same time, besides of iterative method, growing curves were also estimated by Potthoff-Roy's method (1964) which ignores the influence of concomitant variables. In this case the results are shown on the example of another variety (variety 8) from a different simulation. The data were generated in the same way as previously with varying variability of concomitant variables values by multiplying vector of standard deviations by values from the original experiment in succession by values 0.1; 0.5; 1; 2 and 5.

Fig. 4 shows charts estimated curves by iterative method (the left side) and by Pothoff-Roy's method (the right side) for exactly the same data. Each chart has an individual scale on vertical axes to ensure better presentation of estimated curves shapes.

One can see that the best estimation of curve in the iterative method is, again, in the situation where variability of concomitant variables values is the greatest (fig. 4e). In the situation where influence of concomitant variables is ignored i.e. an application of Potthoff-Roy's method, the best estimation can be obtained in the case of the least variation of concomitant variables (fig. 4f). If the variation is the greatest (fig. 4j) then estimators (light lines) visibly differ in shape in comparison of assumed polynomial (black line).

It is also interesting that most of estimated curves lie above or below assumed polynomial depending on the regression coefficients in vector γ. It is worth noticing that the iterative method gives bad estimators of vector γ in the first iteration when variability of concomitant variables is small. The consequences of it are visible in subsequent steps of calculations.

5. Conclusions

The paper presents the results obtained in preliminary studies which can bring near the solution of the problem formulated in following questions: should values of concomitant variables be always considered and what does the precision of growth curve estimation depend on.

The conclusions obtained based on the studies that have been carried out so far are following:

1. The exactness of growth curves estimation using iterative method is increased together with enlarged difference between concomitant variables values in successive time points.
2. If polynomial values in successive time points are strongly differentiated, then exactness of curve estimation obtained by iterative method is better.
3. Omitting the influence of concomitant variables (Potthoff-Roy's method) in the analysis gives the smaller exactness of estimation of the growth curve the more distinct differentiation of concomitant variables values in time points.
4. If concomitant variables influence is omitted (Potthoff-Roy's method) then estimated growth curves preserve the shape of a polynomial if the variability of concomitant variables values is small. If variability enlarges then the shape of the curve differs from the given curve (Fig 4).
5. Further studies are necessary to determine the minimum value of variance of concomitant variables values in time, and possibly to omit these variables in growth curve analysis.

References

Gupta A.K., Nagar D.K. (1999), *Matrix Variate Distributions*, Chapman&Hall/CRS.

Potthoff R.F., Roy S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problem. *Biometrika* 51, 313-326.

Wesołowska-Janczarek M., Fus L. (1996). Parameters estimation in the growth curves model with time-changing concomitant variables. In Polish. *XXVI Coll. Biometr.* 263-277.

O WPŁYWIE ZRÓŻNICOWANIA WARTOŚCI ZMIENNYCH TOWARZYSZĄCYCH NA ESTYMACJĘ WSPÓŁCZYNNIKÓW WIELOMIANÓW W MODELU KRZYWYCH WZROSTU ZE ZMIENNYMI TOWARZYSZĄCYMI O WARTOŚCIACH ZMIENIAJĄCYCH SIĘ W CZASIE I JEDNAKOWYCH DLA WSZYSTKICH JEDNOSTEK EKSPERYMENTALNYCH

Streszczenie

Zmieniające się w czasie wartości zmiennych towarzyszących mają znaczący wpływ na oszacowane wielomiany opisujące przebieg zmian badanej cechy w określonym przedziale czasowym. W pracy bierze się pod uwagę sytuację, gdy wartości zmiennych towarzyszących w kolejnych punktach czasowych są takie same dla wszystkich jednostek. Rozważania dotyczą badania wpływu różnego zróżnicowania wartości zmiennych towarzyszących w czasie na dokładność oceny poszukiwanej krzywej. Wnioski oparto o wyniki uzyskane drogą symulacji komputerowej.

Słowa kluczowe: krzywe wzrostu, zmienne towarzyszące, liniowe modele matematyczne, estymacja parametrów, symulacja komputerowa

Klasyfikacja AMS 2010: 62J12, 68U20