Twin Research (2000) 3, 92–98

www.nature.com/tr

# On the application of queuing theory for analysis of twin data

Lev S Kuravsky[1] and Sergey B Malykh[1,2]

[1]Moscow City Institute of Psychology and Education
[2]Psychological Institute of Russian Academy of Education, Moscow, Russia

A mathematical model based on queuing theory is used to study the dynamics of environmental influence on twin pairs. The model takes into consideration genetic factors and effects of non-shared environment. Histograms are exploited as base analysed characteristics, with the method of minimum chi-square yielding estimated characteristics. The proposed technique was applied to analysis of longitudinal data for MZ and DZ twins. It was shown that the same environment impact may yield different contributions to final variances of the IQ parameters under investigation. Magnitudes of these contributions depend on the genetic factor represented by distributions of an analysed parameter at the point of birth. Twin Research (2000) 3, 92–98.

Keywords: twins, twin development, longitudinal research, genetics, environment, queuing theory

## Introduction

We present a technique intended for studying dynamics of individual characteristics dependent on heredity and also environmental influence. Using twin measurements at a series of time points, this technique makes it possible to observe how environmental factors change initial personal differences resulting from genetic factors during the selected time period and to estimate parameters of interest quantitatively. Parameters describing the effect of both types of factors are discussed. In contrast to traditional confirmatory factor analysis,[1,2] histograms are used as the basis for analysis instead of covariance matrices. The approach may be used in longitudinal research together with the analysis of simplex models[3,4] and is illustrated by an application to the study of IQ.

## Materials and methods

The quantity to be analysed is the absolute difference between the members of a twin pair for the trait of interest. The range of this quantity is divided into several intervals, with each interval considered as a separate state in which a twin pair has some probability to find itself. Transitions between the states are possible over time.

Correspondence: S Malykh, Psychological Institute of Russian Academy of Education, 9 'B' Mokhovaya Street, 103009 Moscow, Russia. Tel: +7 (095) 349 1886; +7 (095) 202 9363; Fax: +7 (095) 349 1886; E-mail: lev.pvt@elc.igpp.msk.su
Received 26 July 1999; accepted 5 October 1999

Queuing theory yields a convenient mathematical model that may be used to describe the dynamics of these transitions. The model is represented by a graph (an example is presented in Figure 1) in which nodes (depicted as rectangles) correspond to the states, and branches (depicted as arrows) correspond to transitions. The process of development of differences between twins may be imagined as a random walk along the graph from one state to another following the arrows. Time is taken as continuous. The initial distribution of state probabilities at the moment of birth reflects genetic differences between twins. As a result of environmental influences, this initial distribution is transformed during the time of observation by instantaneous state-to-state transitions which take place at random time points. It is important to note that only the environmental influence differing for the members of a pair (non-shared environment) is taken into account in the current model.

It is assumed that state-to-state transitions (corresponding to each branch of the graph) meet the following two properties of Poisson's flows of events:

- ordinariness (a flow is ordinary if the probability of appearance of more than one event during a small time interval is much less than the probability of appearance of one event for the same time);
- absence of contagion (the numbers of events falling into any two disjoint intervals are independent).

It may be proved[5] that the number of events $X$ in these flows falling into any interval of the length $\tau$

https://doi.org/10.1375/twin.3.2.92 Published online by Cambridge University Press

adjoining to the time point t is distributed according to the law of small numbers:

$$P_{t,\tau}(X = m) = \frac{(a(t,\tau))^m}{m!} e^{-a(t,\tau)},$$

where $P_{t,\tau}(X = m)$ is the probability of appearance of m events during the considered interval, $a(t,\tau)$ is the mean number of events falling into an interval of the length $\tau$ adjoining to time point t. Only stationary flows (where $a(t,\tau) = \eta\tau$) will be considered here, where parameter $\eta$ is a constant representing the density of a stationary flow and is equal to mean number of events per unit time interval. Mean time interval between two adjacent events in such a flow is therefore $1/\eta$.

The system shown in Figure 1 is a finite chain of n + 1 states where transitions from the state $x_k (k \neq 0, k \neq n)$ are possible only to the preceding state $x_{k-1}$ or to the next state $x_{k+1}$. From the states $x_0$ and $x_n$ the only available states are $x_1$ and $x_{n-1}$ respectively. Processes described by means of such graphs are called 'the processes of death and propagation' since first they were applied in biology to analyse dynamics of population growth. Flow densities from state k are denoted as $\lambda_k$ and $\mu_k$. The parameter $\lambda_k$ represents environmental influences making twins less alike (divergence), whilst $\mu_k$ represents environmental processes where the environment makes twins more alike (convergence).

If the total admissible range of absolute differences between twins is denoted as D, the state $x_0$ corresponds to the pair difference interval from 0 to $D/(n + 1)$, the state $x_1$ to the interval from $D/(n + 1)$ to $2D/(n + 1)$, and so on. The following set of ordinary differential equations[5] may be drawn to describe the time history of state probabilities:

$$\frac{dp_0(t)}{dt} = -\lambda_0 p_0(t) + \mu_1 p_1(t);$$

$$\frac{dp_k(t)}{dt} = -(\lambda_k + \mu_k) p_k(t) + \lambda_{k-1} p_{k-1}(t) + \mu_{k+1} p_{k+1}(t),$$

where (k = 1,2,...,n–1);

$$\frac{dp_n(t)}{dt} = -\mu_n p_n(t) + \lambda_{n-1} p_{n-1}(t)$$

where $p_k(t)$ is the probability to be within the state $x_k$ at the time point t. To simplify the problem, flow densities are here assumed to be constant with regard to the index k: $\lambda_0 = \lambda_1 = \ldots = \lambda_{n-1} = \lambda$ and $\mu_1 = \mu_2 = \ldots = \mu_n = \mu$. To integrate these equations, one has to assign initial conditions

$$p_0(0), p_1(0), \ldots, p_n(0); \sum_{k=0}^{n} p_k(0) = 1.$$

The normalisation condition $\sum_{k=0}^{n} p_k(t) = 1$ is valid at any time point.

Applying this model to twins, it is postulated that t = 0 is the moment of their birth. For MZ twins, $p_0(0) = 1$ and $p_k(0) = 0$ (k = 1,...,n) since there are no genetic differences between them and no environment effects at this time point. Starting differences for DZ twins and unrelated pairs may be described by a normal distribution with zero mean and some standard deviations $\sigma_{dz}$ and $\sigma_{un}$ which are used to characterise genetic differences. Because of distribution symmetry,

$$p_0(0) = \frac{2}{\sqrt{2\pi}\sigma} \int_0^{D/(n+1)} e^{-x^2/2\sigma^2} dx,\ p_1(0) = \frac{2}{\sqrt{2\pi}\sigma} \int_{D/(n+1)}^{2D/(n+1)} e^{-x^2/2\sigma^2} dx, \ldots,$$

and so on ($\sigma = \sigma_{dz}$ for DZ twins and $\sigma = \sigma_{um}$ for unrelated pairs). Genetic and environmental effects are estimated in terms of the standard deviations $\sigma_{dz}$ and $\sigma_{un}$ and flow densities $\lambda$ and $\mu$.

Expected state probabilities are calculated by integrating the above set of differential equations. The expected frequency of being at the kth state in the gth group (g = 1,2,...,G) is equal to $p_{kg}N_g$, where $p_{kg}$ is the probability of being at the kth state in the gth group and $N_g$ is the number of cases in the group. The corresponding observed frequencies $F_{kg}$ result from measurements obtained during a longitudinal twin study. Under some conditions, the following statistic is distributed asymptotically according to a chi-square distribution:

$$\sum_{g=1}^{G} \sum_{k=0}^{n} \frac{(F_{kg} - p_{kg}N_g)^2}{p_{kg}N_g}.$$

This sum should be regarded as a goodness-of-fit measure, with the number of degrees of freedom equal to Gn-l where l is the total number of independent parameters. The technique that finds

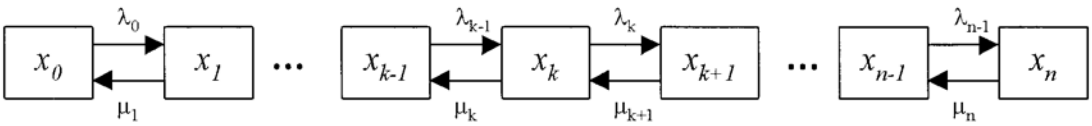


Figure 1 The model used to describe dynamics of twin development, where the $x_i$ (i = 0,1,...,n) are states, and the $\lambda_i$ (i = 0,1,...,n-l) and $\mu_i$ (i = 1,2,...2,n) are flow densities



https://doi.org/10.1375/twin.3.2.92 Published online by Cambridge University Press

independent parameters as the parameters minimising the aforementioned statistic is called the method of minimum chi-square. For the problems under consideration, it usually yields estimations which are close to ones of the maximum likelihood method. The statistic is minimised at the specified time points in which observed data are available.

Computation of parameters consists of two stages. In the preparatory stage, numerical integration of the differential equations is required to calculate all $p_k$. In our case this was done using the Microsoft® Excel 97 spreadsheet. These probability functions are computed with some specified time step h from the initial zero time point to the given specified upper time bound. Runge-Kutta methods[6] (or their equivalents) proved to be sufficient to obtain acceptable accuracy of solution. For example, the following integration scheme of the second order (the modified Euler method adapted to the equations in question) may be used:

$$p_k((m+1)h) = p_k(mh) + \tfrac{1}{2}h\{f_k[p_{k-1}(mh), p_k(mh), p_{k+1}(mh)] + f_k[p_{k-1}(mh) + hp'_{k-1}(mh), p_k(mh) + hp'_k(mh), p_{k+1}(mh) + hp'_{k+1}(mh)]\},$$

where k=0,1,...,n (if k=0, $p_{k-1}$ is to be dropped; if k=n, $p_{k+1}$ is to be dropped); m=0,1,...,∞, prime denotes $\frac{d}{dt}$; and $f_k$ [...] indicates $\frac{dp_k(t)}{dt}$ in the right-hand side of a corresponding equation.

In the second stage, an optimisation procedure to obtain estimates of the free parameters is run. The procedure of numerical non-linear optimisation called Generalized Reduced Gradient (GRG2), developed by Leon Lasdon (University of Texas at Austin) and Allan Waren (Cleveland State University), was used in this study.

Table 1 Results of fitting models to observed general IQ distributions

| Age | Degrees of freedom | Chi-square statistics | P-value |
|---|---|---|---|
| 6 years | 29 | 19.36 | 0.91 |
| 14 years | 31 | 25.09 | 0.76 |
| Total | 60 | 44.44 | 0.93 |

Table 2 Estimates of the independent parameters

| Parameter | Value |
|---|---|
| Genetic effects | |
| Initial standard deviation for DZ twins | 7.90 |
| Initial standard deviation for unrelated pairs | 18.56 |
| Environment effects | |
| Divergence flow density (0–6 years) | 0.128 |
| Convergence flow density (0–6 years) | 0.206 |
| Divergence flow density (6–14 years) | 0.194 |
| Convergence flow density (6–14 years) | 0.176 |

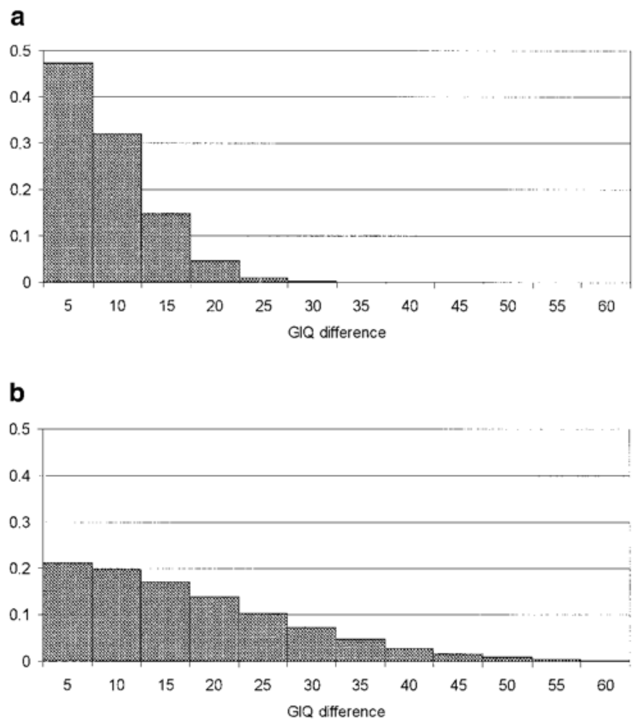


Figure 2 Distributions of expected state probabilities for the initial time point: (a) DZ twins, (b) unrelated pairs

## Results

The technique described above was applied to analysis of longitudinal measures of general IQ (GIQ) for 6- and 14-year-old pairs of MZ and DZ twins. Three groups of pairs took part in the analysis: MZ twins (47 6-year-old pairs, 37 14-year-old pairs), DZ twins (47 6-year-old pairs, 33 14-year-old pairs), and unrelated pairs (41 6-year-old pairs, 45 14-year-old pairs). The sum of goodness-of-fit measures at the time points of 6 and 14 years was minimised by estimation of the following free parameters:

- flow densities $\lambda_{0-6}$ and $\mu_{0-6}$ which were assumed to be common for the range 0–6 years,
- flow densities $\lambda_{6-14}$ and $\mu_{6-14}$ which were assumed to be common for the range 6–14 years,
- standard deviations $\sigma_{dz}$ and $\sigma_{un}$ characterising initial GIQ distributions at the zero time point.

One year was used as the unit of time measurement, and the modified Euler method was coded as a numerical integration scheme with a time step from 0.006 to 0.008 years. The model to be fitted was



https://doi.org/10.1375/twin.3.2.92 Published online by Cambridge University Press

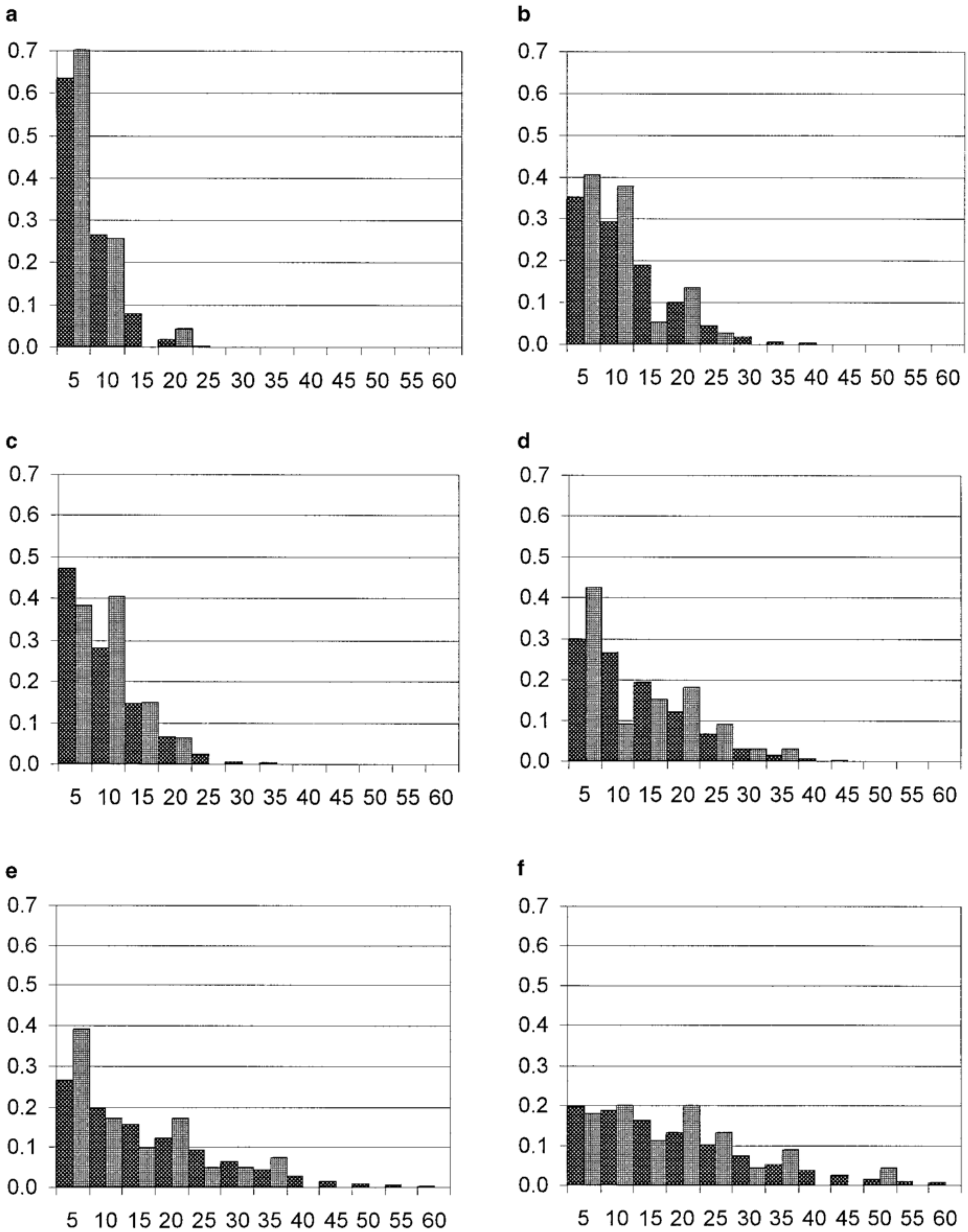


Figure 3 Observed and expected state probabilities vs GIQ difference: (a) MZ twins–6 years, (b) MZ twins–14 years, (c) DZ twins–6 years, (d) DZ twins–14 years, (e) unrelated pairs–6 years, (f) unrelated pairs–14 years

represented by a chain containing 12 states as shown in Figure 1, and the interval of GIQ differences within pairs (from 0 to 60 units) was divided into 12 equal parts. Fitting the model to the longitudinal



https://doi.org/10.1375/twin.3.2.92 Published online by Cambridge University Press

GIQ data yielded the results presented in Table 1, with the chi-square statistics indicating a very good fit of the model to the observed data.

Table 2 shows the estimates obtained for the parameters of interest. Distributions of expected state probabilities for the initial time point for DZ twins and unrelated pairs are given in Figure 2, whilst observed and expected state probabilities for 6-year and 14-year pairs are presented in Figure 3. As an illustration, estimations of functions $p_k(t)$ for MZ twins within the interval from 0 to 6 years are shown in Figure 4.

To determine how the standard deviations of GIQ differences change in the time domain, normal distributions with zero mean were selected with the aid of the optimisation procedure to obtain the best chi-square fit with regard to the distributions of expected probabilities. Comparisons of estimated parameters for 6- and 14-year ages are presented in Table 3.

Final values of state probabilities $p_k$ determined by the aforementioned numerical integration scheme at the specified time points may be considered as functions of independent parameters. These values agree with an exact solution to the precision of the numerical integration, which may be set arbitrarily small. Within some neighbourhood of the solution found by the optimisation procedure, it may be shown that:

- a set of values of the independent parameters minimising the goodness-of-fit measure exists and is unique;
- this set of independent parameters converges in probability to the set of values which yields true probabilities $p_k$ when the number of tests approaches infinity;
- the goodness-of-fit measure is distributed asymptotically according to a chi-square distribution with a number of degrees of freedom that is equal to the difference between the number of independent observed statistics and the number of independent parameters.

Rigorous formulation and proof of the corresponding theorem may be found, for example, in the monograph by H Cramer.[7] The truths of some assumptions of this theorem for the specific problem under study were proved with the aid of a special numerical procedure coded in the spreadsheet. Since parameters of the optimisation procedure in use were tuned for running one of the quasi-Newton algorithm variants, finding strict local minima, if any, was guaranteed within the specified accuracy range. As the procedure finds a point in which the gradient equals to zero, such point is unique in some neighbourhood of the solution that has been found (up to the corresponding numerical method error).

## Discussion

The results presented in Table 3 demonstrate clear dependence of initial GIQ distributions on the relationship proximity: the greater degree of this proximity the less corresponding standard deviation

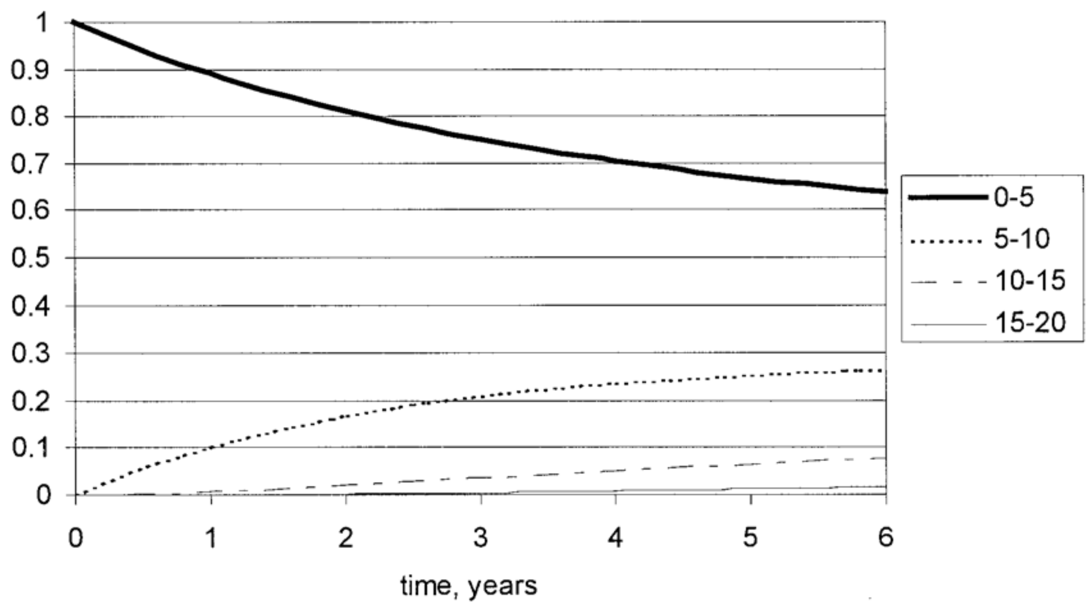


Figure 4 Functions $p_f(t)$ for MZ twins from 0 to 6 years (state intervals are shown in IQ units)



https://doi.org/10.1375/twin.3.2.92 Published online by Cambridge University Press

($\sigma_{un} > \sigma_{dz} > \sigma_{mz} = 0$). This fact agrees with the usual expectations concerning the effect of genetic factor. The estimate of the initial standard deviation for DZ twins equals 43% of this quantity for unrelated pairs (in other investigations it was located within the interval 40–60%). It would be interesting to clarify the interdependence of this parameter and pair correlation coefficient in further research. As a rule, environmental impact results in 'the washing out' of initial distributions: the greater the elapsed time, the greater the standard deviation. The only exception (for the case of 6-year-old unrelated pairs) may be explained by sampling errors.

Estimations of flow densities (see Table 2) show that before school (until the age of 6 years) the environment promotes convergence in GIQ more than divergence. In school (after the age of 6 years) the situation changes: the environment promotes divergence in GIQ more than convergence. It may be also noted that after entering school divergence density increases but convergence density decreases.

Table 3 Standard deviations of normally distributed general IQ differences in different time points

| Types of pair | Age, years | Standard deviation, GIQ units |
|---|---|---|
| MZ twins | 0 | 0 |
| | 6 | 6.39 |
| | 14 | 10.97 |
| DZ twins | 0 | 7.90 |
| | 6 | 9.09 |
| | 14 | 12.93 |
| Unrelated pairs | 0 | 18.56 |
| | 6 | 18.04 |
| | 14 | 20.47 |

Table 4 Comparison of estimations of independent parameters for the models with 5- and 10-unit state intervals (14-year-old pairs)

| Parameter | 5-unit state interval (model 1) | 10-unit state interval (model 2) | Percentage |
|---|---|---|---|
| Initial standard deviation for DZ twins | 10.85 | 12.71 | 85 |
| Initial standard deviation for unrelated pairs | 21.28 | 22.57 | 94 |
| Final standard deviation for MZ twins | 9.70 | 9.86 | 98 |
| Final standard deviation for DZ twins | 13.82 | 15.09 | 92 |
| Final standard deviation for unrelated pairs | 22.89 | 22.94 | 99.8 |
| Divergence flow density | 0.027 | 0.151 | 18 |
| Convergence flow density | 0.029 | 0.185 | 15 |

Table 5 Initial variance as proportion of final variance for different pair types

| Types of pair | Initial variance as % of final |
|---|---|
| MZ twins | 0 |
| DZ twins | 37 |
| Unrelated pairs | 82 |

To clarify how the independent parameters estimated may be changed in the case of other length of state intervals, two estimates for 14-year-old twins were obtained in another GIQ study (119 MZ pairs, 90 DZ pairs). Their results corresponding to 5- and 10-unit intervals are shown in Table 4. One can see that the difference in initial standard deviations is not greater than 6–15%. Taking into account the sizes of state intervals this may be regarded as a good fit. Comparison of flow densities indicates the dependencies between state interval size and the flow densities are non-linear.

Table 5 shows initial variance as a proportion of final variance for different pair types. Within the framework of the model under consideration, these data make it possible to draw an important conclusion: the same environment impact (represented by flow densities) may yield different contributions to final variances. Magnitudes of these contributions depend on the genetic factor (represented by distributions of an analysed parameter at the point of birth).

It would be interesting to compare results obtained by the approach presented and confirmatory factor analysis. However, since analysed absolute differences in a pair differ from parameters which are studied traditionally in twin confirmatory factor analysis, direct comparison of both approaches is impossible. Such estimated characteristics as flow densities are quite different from quantities that are used in factor analysis and cannot be calculated with its aid. Since they are of interest in twin research, the approach in question may be considered as a new source of information which supplements traditional ones.

## Acknowledgements

This work is supported by grant 99-06-80161 from the Russian Foundation for Basic Research.

## References

1 Jöreskog KG. Factor analysis by least squares and maximum-likelihood method. In: Enslein K, Ralston A, Wilf HS (eds). Statistical Methods for Digital Computers. Wiley: New York, 1977; pp 125–153.
2 Martin N, Eaves L. The genetical analysis of covariance structures. Heredity 1977; 38: 79–95.
3 Boomsma DI, Martin NG, Molenaar PCM. Factor and simplex models for repeated measures: Application to two psychomotor measures of alcohol sensitivity in twins. Behav Genet 1989; 19: 79–96.

4 Neale MC, Cardon LR. Methodology for Genetic Studies of Twins and Families. Kluwer Academic Publishers: Dordrecht, 1992.
5 Ovcharov LA. Prikladnye zadachi teorii massovogo obsluzhivaniya (Applied Problems of the Queuing Theory). Mashinostroenie: Moscow, 1969 (in Russian), pp 31–34, 53–66.
6 Bakhvalov NS. Chislennye metody (Numerical Methods). Nauka: Moscow, 1975 (in Russian), pp 447–459.
7 Cramer H. Mathematical Methods of Statistics. Princeton University Press: Princeton, 1946.