PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **19** (2009), 69–82

PLISKA
STUDIA MATHEMATICA
BULGARICA

OFFSPRING MEAN ESTIMATORS IN BRANCHING PROCESSES WITH IMMIGRATION*

Dimitar Atanasov, Vessela Stoimenova, Nikolay Yanev

In the present paper we consider the discrete time branching process with immigration and its relationship to the Bienayme-Galton-Watson process with a random number of ancestors. Several estimators of the offspring mean are considered – the Harris estimator, the conditional least squares estimator of Heyde-Seneta, the conditional weighted least squares estimator of Wei-Winnicki and the estimator of Dion and Yanev. Their properties are compared using computational results based on simulations of the entire immigration family trees. The asymptotic normality of the estimator of Dion and Yanev is combined with the general idea of the trimmed and weighted maximum likelihood. As a result, robust modifications of the offspring mean estimator is proposed.

1. Introduction

The statistical inference for branching processes with immigration (BGWI processes) is considered in a huge number of papers like those of Heyde and Leslie (1971), Heyde and Seneta (1972), Heyde (1974), Yanev and Tchoukova-Dancheva (1980), Winnicki (1988), Wei and Winnicki (1989, 1990) and others. In these papers the asymptotic properties of the BGWI process are studied and the nonparametric

*The paper is partially supported by the National Science Fund of Bulgaria, Grant No VU-MI-105/2005.

2000 *Mathematics Subject Classification:* 60J80.

Key words: branching processes with immigration, offspring mean estimatimators, simulation.

maximum likelihood, the conditional least squares and weighted conditional least squares estimation are introduced.

In the present paper we consider the estimation of the offspring mean in BGWI processes from another point of view using the relationship between the process with immigration and the process with an increasing random number of ancestors, whose statistical estimation is proposed by Yanev (1975) and studied in the nonparametric situation by Dion and Yanev (1991, 1992, 1994, 1997). The advantages of this approach and its extention for the purpose of robustification in the sense of the trimmed and weighted likelihood are studied in Atanasov et al (2007), where the estimation of the immigration component is considered.

We remind that the BGWI process is a process with two types of particles: the so-called natives and immigrants. They are characterized by the fact that each particle reproduces independently of each other. Each native particle gives rise only to natives, according to an offspring distribution $\{p_k\}$ with mean m and variance σ^2. The immigrant always produces just one immigrant as well as a random number of "natives" according to an immigration distribution $\{q_k\}$, whose mean and variance are λ and b^2 respectively. This is an example of a decomposable singular multitype branching process. As noted in Dion(1993), because of its importance for applications, it is treated separately and is considered in a different way from the multitype model.

Let $\{Y_n\}_{n=0}^{\infty}$ be a BGWI process defined by the recursive formula

$$(1) \qquad Y_n = \begin{cases} \sum\limits_{j=1}^{Y_{n-1}} X_{nj} + I_n, & if \quad Y_{n-1} > 0, \\ I_n, & if \quad Y_{n-1} = 0. \end{cases}$$

Here $\{X_{nj}\}$ and $\{I_n\}$ are independent sequences of i.i.d. nonnegative, integer valued random variables with distribution $\{p_k\}$ and $\{q_k\}$ respectively. The r.v. Y_0 is nonnegative and integer valued , which is also independent of $\{X_{nj}\}$ and $\{I_n\}$. Without any loss of generality further on one can assume that $Y_0 = 0$.

Assume further that the offspring and immigration distributions are nondegenerate and the offspring mean $m < \infty$.

As usually the process is called *subcritical* if $m < 1$, *critical* if $m = 1$ and *supercritical* if $m > 1$.

1.1. Classical nonparametric estimation

Results about the nonparametric estimation in the supercritical BGWI process are announced in Heyde (1974), Heyde and Seneta (1972), Heyde and Leslie (1971) and others.

It is well known that the branching process with immigration does not become extinct and therefore one would expect to be able to estimate consistently the mean and variance of the individual distribution and the immigration component.

In the supercritical situation many of the estimators for the classical Bienayme-Galton-Watson process for the individual characteristics can be applied for the BGWI process. Concerning the individual mean m, the Harris estimator

$$\widehat{m}_t = \frac{\sum_{j=1}^{t} Y_j}{U_t}, \tag{2}$$

where $U_t = \sum_{j=0}^{t-1} Y_j$ is the total number of particles in the first $t-1$ generations, is completely applicalble. In the subcritical case, however, the estimator $\widehat{m}$ is not consistent (and as we have already mentioned, for the BGWI process it is reasonable to consider the consistency property of the estimators when $m < 1$), therefore other estimators are needed.

The conditional least squares estimators of Heyde-Seneta for m and λ are

$$\overline{m}_n = \frac{n \sum_{k=1}^{n} Y_k Y_{k-1} - \sum_{k=1}^{n} Y_k \sum_{k=1}^{n} Y_{k-1}}{n \sum_{k=1}^{n} Y_{k-1}^2 - \left(\sum_{k=1}^{n} Y_{k-1} \right)^2} \tag{3}$$

and

$$\overline{\lambda}_n = \frac{n \sum_{k=1}^{n} Y_k \sum_{k=1}^{n} Y_{k-1}^2 - \sum_{k=1}^{n} Y_{k-1} \sum_{k=1}^{n} Y_k Y_{k-1}}{n \sum_{k=1}^{n} Y_{k-1}^2 - \left(\sum_{k=1}^{n} Y_{k-1} \right)^2}.$$

In Winnicki (1988) it is noted that the conditional least squares estimators are not satisfactory for the following reasons:

1. The estimator $\overline{m}_n$ has a larger asymptotic variance than the m.l.e. of Harris $\sum_{i=1}^{n} Y_i / \sum_{i=1}^{n} Y_{i-1}$ in the supercritical case.

2. The estimator $\overline{\lambda}_n$ is not a consistent estimator for λ in the supercritical case.

To avoid these disadvantages Wei and Winnicki (1989) proposed to use the weighted conditional least squared estimators, i.e. estimators, obtained by minimizing

$$\sum_{k=1}^{n}\left(\frac{Y_k - E(Y_n|\Im_{k-1}}{\sqrt{Var(Y_k|\Im_{k-1}))}}\right)^2.$$

The weighted conditional least squares estimators for the offspring and immigration mean are

$$\widetilde{m}_n = \frac{\sum_{k=1}^{n} Y_k \sum_{k=1}^{n} \frac{1}{Y_{k-1}+1} - n\sum_{k=1}^{n}\frac{Y_k}{Y_{k-1}+1}}{\sum_{k=1}^{n}(Y_{k-1}+1)\sum_{k=1}^{n}\frac{1}{Y_{k-1}+1} - n^2} \tag{4}$$

and

$$\widetilde{\lambda}_n = \frac{\sum_{k=1}^{n} Y_{k-1} \sum_{k=1}^{n} \frac{Y_k}{Y_{k-1}+1} - \sum_{k=1}^{n} Y_k \sum_{k=1}^{n}\frac{Y_{k-1}}{Y_{k-1}+1}}{\sum_{k=1}^{n}(Y_{k-1}+1)\sum_{k=1}^{n}\frac{1}{Y_{k-1}+1} - n^2}.$$

In Winnicki (1988) it is noted that in the supercritical case $\widetilde{m}_n$ is a more efficient estimator than $\overline{m}_n$ in the sense of achieving a lower asymptotic variance. However the estimator $\widetilde{\lambda}_n$ is not consistent. The weighted conditional least squares method offers a substantial improvement over the ordinary conditional least squares when the supercritical case is considered, but it does not solve the problem of estimating the parameters λ and b^2 of the immigration distribution. In fact it is proved that these parameters do not have consistent estimators in the supercritical case. The proof of this statement is provided by Wei and Winnicki (1987).

1.2. The relationship between BGWI and BGWR

As noted in Dion (1993), traditionally branching processes with or without immigration have been treated separately. However the estimation theory for the offspring parameters in a Bienayme-Galton-Watson process having a random number of initial ancestors $Z_0(n)$ (BGWR process) can be transferred to a process with immigration without taking account of the criticality of the processes.

Yakovlev and Yanev (1989) noted that branching processes with a large and often random number of ancestors occur naturally in the study of cell proliferation and in applications to nuclear chain reactions. Results about the classical nonparametric estimation of the offspring mean m and variance σ^2 in the BGWR process are announced in Dion and Yanev (1991, 1992, 1994, 1997), robustified versions (in the sense of the weighted and trimmed likelihood) of the classical estimators are proposed in Stoimenova, Atanasov, Yanev (2004 a,b, 2005) and Stoimenova, Atanasov (2006) . Some aspects of the classical parametric estimation are considered in Stoimenova, Yanev (2005) and of the robust parametric estimation - in Stoimenova (2005).

One may consider the partial tree in (1) underlying $\{Y_0, \dots, Y_t, \dots, Y_{n+t}\}$ and define the r.v. $Z_t(n)$ as the number of individuals among generations $t, t+1, t+2, t+n$, whose ancestors immigrated exactly t generations ago, $n, t = 0, 1, 2, \dots$.

That way $Z_0(n) = \sum_{j=0}^{n} I_j$ is the total number of immigrants from time 0 to time n, $Z_1(n)$ is the total number of their offspring, etc. Hence $\{Z_t(n)\}$ is a BGW process having a random number of ancestors and can be presented as follows:

$$Z_t(n) = \begin{cases} \sum_{i=1}^{Z_{t-1}(n)} \xi_i(t,n) & \text{if} \quad Z_{t-1}(n) > 0, \ t = 1, 2, \dots \\ 0, & \text{otherwise,} \end{cases}$$

where the r.v. $\{\xi_i(t,n)\}$ are i.i.d. with values in the set of nonnegative integers $N = \{0, 1, 2, \dots\}$; $\xi_i(t,n), i \in N$ are independent of $Z_0(n)$. For each $n = 1, 2, \dots$ $Z(n) = \{Z_t(n), t = 0, 1, \dots\}$ is a Bienayme- Galton-Watson process having a random number of ancestors $Z_0(n) \geq 1$. Such a process is denoted BGWR.

The Dion-Yanev estimator of the offspring mean m, obtained on the basis of the transformed into a BGWR process immigration process, is the Harris type estimator

$$\widehat{\widehat{m}}_t(n) = \frac{\sum_{i=1}^{t} Z_i(n)}{\sum_{i=0}^{t-1} Z_i(n)}. \tag{5}$$

The estimators of the immigration component λ of Dion and Yanev are

$$\widehat{\widehat{\lambda}}_t(n) = \frac{Z_t(n)}{nm^t}$$

if the offspring mean m is known and

$$\widehat{\widehat{\lambda}}_t(n) = \frac{Z_t(n)}{n(\widehat{m}_t(n))^t}$$

if m is unknown.

2. Comparison of the behaviour of the offspring mean estimators

In this section we quantitatively compare the offspring mean estimators, described in the previous section.

500 family trees with up to 20 generations are generated for each combination of the parameter values:

- expected number of descendants: $m = 0.9,\ 1.0,\ 1.0$;
- expected number of immigrants in each generation: $\lambda = 0.8,\ 1.0,\ 1.2$;

For each of these family trees the estimators of Wei-Winnicki $\widetilde{m}_n$, Heyde-Seneta $\overline{m}_t$ and Harris $\widehat{m}_t$ for the initial immigration trees, as well as the Dion-Yanev estimators $\widehat{\widehat{m}}_t(n)$ for the transformed trees are calculated.

We use the following formula for the relative difference of the estimator m^* to the true value of the offspring mean m:

$$r \quad = \quad \frac{m - m^*}{m},$$

which gives a measure for the normalized bias of the estimator.

On the following Figure 1 the Box-plots for the relative diferrence r for the four types of estimators over all generated trees are shown. It can be seen that the Dion-Yanev estimator exhibits the smallest deviation. The estimator of Wei-Winnicki is the most unreliable one. As usually, the lower and upper lines of the "box" are the 25th and 75th percentiles of the sample of estimators. The distance between the top and bottom of the box is the interquartile range and the line in the middle of the box is the sample median. The maximum of the sample is the top of the upper "whisker", the minimum of the sample is the bottom of the lower "whisker".

The estimators of Heyde-Seneta overestimate the real parameter values in the subcritical case and underestimates them in the supercritical case. This can be seen more precisely on Figure 2, where the estimators of Heyde-Seneta are compared for different values of m.

Let us consider the relative difference of this estimator for different values of the offspring mean m. It estimates the individual mean more precisely as m grows.

On Figures 4 and 5 the behaviour of the estimator of Wei-Winnicki and its relative difference can be seen. This estimator is not appropriate for the subcritical case and for the supercritical case the estimation is very unstable.

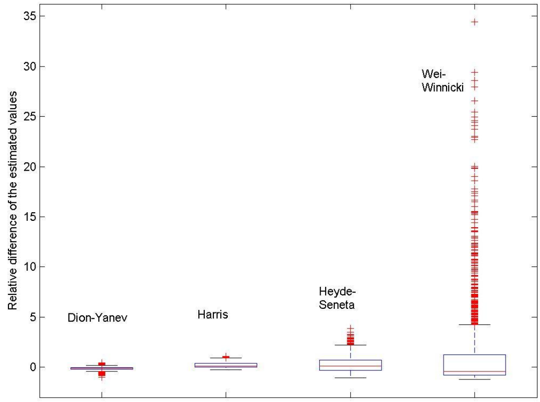

Figure 1: Comparison of the offspring mean estimators

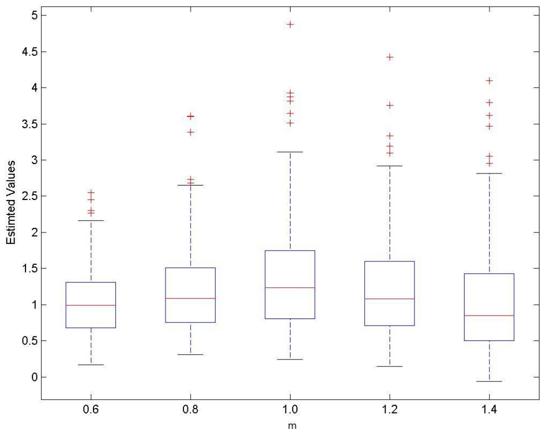

Figure 2: Behaviour of the Heyde-Seneta estimator for different values of the offspring mean

The Harris estimator has a good behaviour except for the subcritical case, where it overestimates the real value of m (as already mentioned, in the subctitical situation this estimator is not consistent).

The Dion-Yanev estimator for the transformed immigration process into a BGWR process is the most stable among all estimators we have just considered.

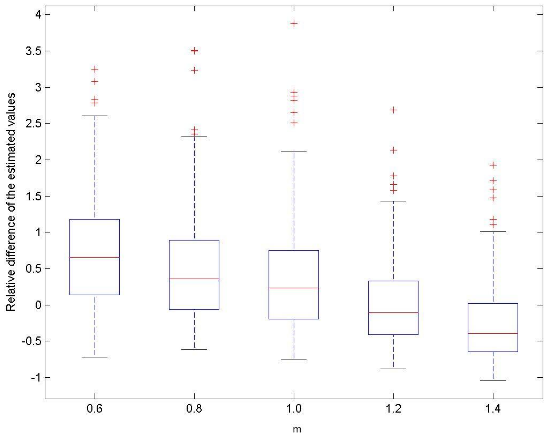

Figure 3: Relative difference of the Heyde-Seneta estimator for different values of the offspring mean

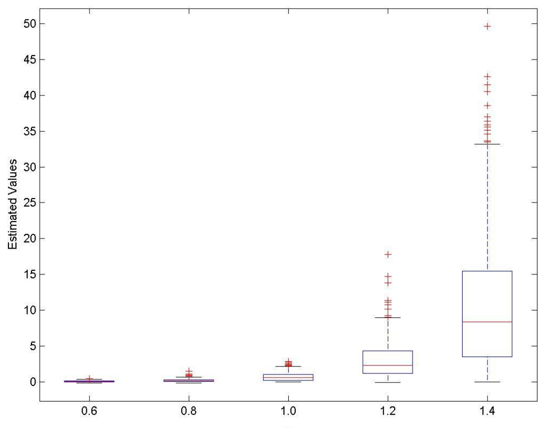

Figure 4: Behaviour of the Wei-Winnicki estimator for different values of the offspring mean

Its relative difference is small and centered around zero. This stability holds for the subcritical, critical and supercritical situation.

One should notice, however, that this stability can be explained additionally

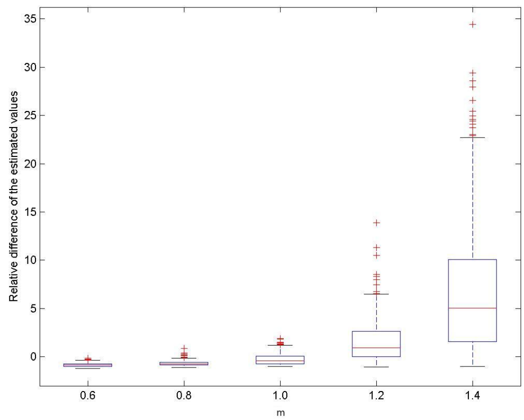

Figure 5: Relative difference of the Wei-Winichi estimator for different values of the offspring mean

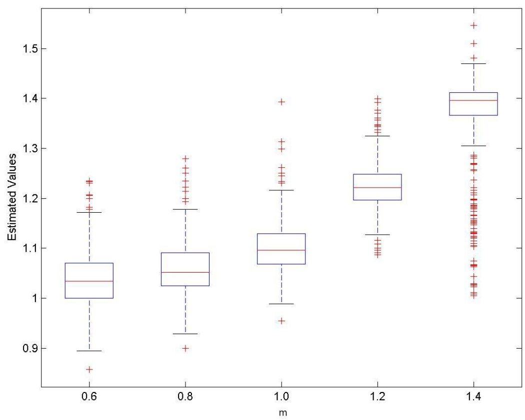

Figure 6: Behaviour of the Harris estimator for different values of the offspring mean

by the fact, that the transformation BGWI-BGWR requires more information about the immigration process (not only the generation sizes). It is noted in Dion and Yanev (1994) that in general the knowledge of $\{Z_0(n), \ldots, Z_t(n)\}$ would seem to be asymptotically equivalent to

$$\left\{ [Y_k]_0^{t+n}, \sum_{k=0}^{n} I_k \right\}$$

as $n,\ t \longrightarrow \infty$ on the set of the nonextinction.

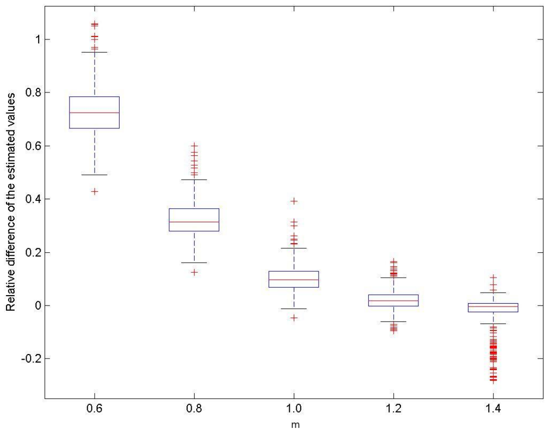

Figure 7: Relative difference of the Harris estimator for different values of the offspring mean

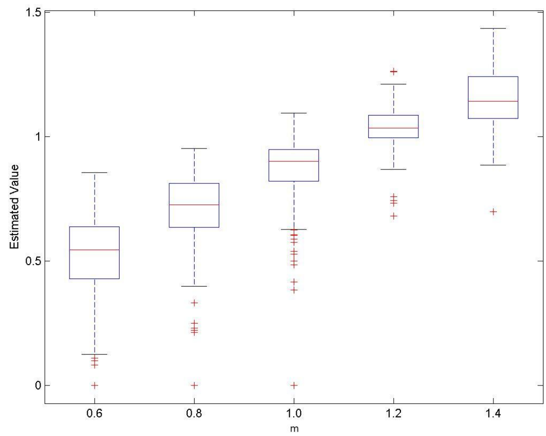

Figure 8: Behaviour of the Dion-Yanev estimator for different values of the offspring mean

3. Robust estimation of offspring mean

Stoimenova et. al. (2004a) proved that there exist a robust estimator $\bar{M}$, based on $WLTE(k)$ (see Vandev and Neykov, 1998) for the Dion-Yanev estimator in BGWR processes, and studied its breakdown properties. It is defined as follows:

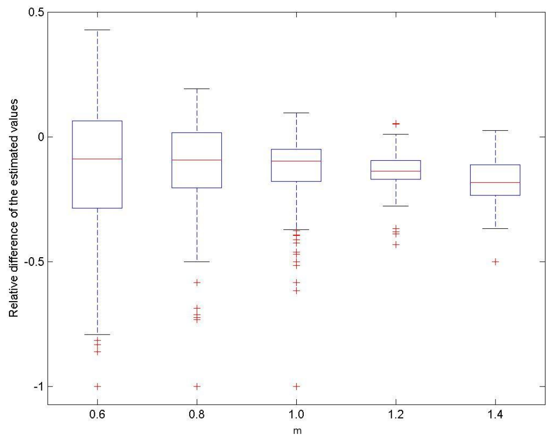

Figure 9: Relative difference of the Dion-Yanev estimator for different values of the offspring mean

$$\bar{M} = \underset{m \in R}{\operatorname{argmin}} \sum_{i=1}^{T} -w_i f(Est(S_{\nu(i)}, m)),$$

where $S = \{S_1, \cdots, S_n\}$ is a set of n independent realizations of a BGWR process with offspring mean m, w_i are nonnegative weights, such that at least T of them are not zero, T is a properly choosen trimming factor, $f(x)$ is the log-density of the asymptotically normal distribution of the Dion-Yanev estimators, ν is a permutation of the indices, such that

$$f(Est(S_{\nu(1)}, m)) \geq \cdots \geq f(Est(S_{\nu(n)}, m)),$$

and

$$Est(S_i, m) = \frac{\sqrt{Z_0^i + \cdots + Z_{t-1}^i}}{\sigma} (\hat{\hat{m}}_t(n)^i - m).$$

Here we have denoted by $\hat{\hat{m}}_t(n)^i$ the Dion-Yanev estimator of the offspring mean for the i-th samplepath, $i = 1, 2, \ldots, n$.

In this section we compare the Dion-Yanev estimator $\hat{\hat{m}}_t(n)$ for the BGWI process with its robust modification $\bar{M}$. The computational results are shown in Table 1. The column m gives the real offspring mean values, $\bar{M}$ is the column, containing the robust estimates using all simulated trees with the corresponding

m and trimming factor set to 70% of the samplepaths; $SE\bar{M}$ gives their standard errors, *Mean* and *Median* are respectively the mean and median values of all Dion-Yanev estimators, obtained for the value m.

m	$\bar{M}$	SE $\bar{M}$	Mean	Median
0.6	0.5529	0.0068	0.5203	0.5469
0.8	0.7503	0.0052	0.7087	0.7260
1.0	0.9153	0.0034	0.8730	0.9021
1.2	1.0305	0.0017	1.0387	1.0352
1.4	1.1220	0.0007	1.1601	1.1438

Table 1: Dion-Yanev estimators and their robust modifications

It can be seen that when there are no outlier trees, the Dion-Yanev estimators and their robust modifications behave in the same manner. This is due to the stable behaviour of the Dion-Yanev estimator, discussed in the previous Section.

Remark. All calculations are made under MATLAB with “BP Engine Rev. 2” package, available at *http://www.fmi.uni-sofia.bg/fmi/statist/projects/bp.*

REFERENCES

[1] D. Atanasov, V. Stoimenova, N. M. Yanev. Estimators in Branching Processes with Immigration. *Pliska Stud. Math. Bulgar.* **18** (2007), 19–40.

[2] J. P. Dion. Statistical Inference for Discrete Time Branching Processes. Lecture Notes. 7th International Summer School on Probability Theory and Mathematical Statistics (Eds A. Obretenov, V. T. Stefanov), 1993, 60–121.

[3] J. P. Dion, N. M. Yanev. Limiting Distributions of a Galton-Watson Branching Process with a Random Number of Ancestors. *C. R. Acad. bulg. Sci.* **44 (3)** (1991), 23–26.

[4] J. P. Dion, N. M. Yanev. Estimation Theory for the Variance in a Branching Process with an Increasing Random Number of Ancestors *C. R. Acad. bulg. Sci.* **45(11)** (1992), 27–30.

[5] J. P. Dion, N. M. Yanev. Statistical Inference for Branching Processes with an Increasing Number of Ancestors. *J. Statistical Planning and Inference* **39** (1994), 329–359.

[6] J. P. Dion, N. M. Yanev. Limit Theorems and Estimation Theory for Branching Processes with an Increasing Random Number of Ancestors. *J. Appl. Prob.* **34** (1997), 309–327.

[7] C. C. Heyde. On Estimating the Variance of the Offspring Distribution in a Simple Branching Process. *Adv. Appl. Prob.* **3** (1974), 421–433.

[8] C. C. Heyde, J. R. Leslie. Improved Classical Limit Analogues for Galton-Watson Process with or without Immigration. *Bull. Aus tral. Math. Soc.* **5** (1971), 145–156.

[9] C. C. Heyde, E. Seneta. Estimation Theory for Growth and Immigration Rates in a Multiplicative Process. *J. Appl. Prob.* **9** (1972), 235–256.

[10] V. Stoimenova. Robust Parametric Estimation of Branching Processes with Random Number of Ancestors. *Serdica Math. J.* **31**, No 3 (2005) 243–262.

[11] V. Stoimenova, D. Atanasov, N. Yanev. Robust Estimation and Simulation of Branching Processes. *C. R. Acad. bulg. Sci.* **57**, No 5 (2004) 19–22.

[12] V. Stoimenova, D. Atanasov, N. Yanev. Simulation and Robust Modification of Estimates in Branching Processes. *Pliska Stud. Math. Bulgar.* **16** (2004), 259–271.

[13] V. Stoimenova, D. Atanasov, N. Yanev. Algorithms for Generation and Robust Estimation of Branching Processes with Random Number of Ancestors. *Math. and Education in Math.* **34** (2005), 196–201.

[14] V. Stoimenova, D. Atanasov. Nonparametric Robust Estimation of the Individual Distribution in Branching Processes with a Random Number of Ancestors. *Math. and Education in Math.* **35** (2006), 302–308.

[15] V. Stoimenova, N. Yanev. Parametric Estimation in Branching Processes with an Increasing Random Number of Ancestors. *Pliska Stud. Math. Bulgar.* **17** (2005), 295–312.

[16] D. L. Vandev, N. M. Neykov. About Regression Estimators with High Breakdown Point *Statistics* (1998), 111–129.

[17] C. Z. Wei, J. Winnicki. Some Asymptotic Results of Branching Processes with Immigration. Ann. Statist. **31** (1989), 261–282.

[18] C. Z. Wei, J. Winnicki. Estimations of the Means in the Branching Processes with Immigration. *Ann. Statist.* **18** (1990), 1757–1773.

[19] J. Winnicki. Estimation Theory of the Branching Processes with Immigration. *Contemporary Mathematics* **80** (1988), 301–323.

[20] A. Yu. Yakovlev, N. M. Yanev. Transient Processes in Cell Proliferation Kinetics. Springer Verlag, Berlin, 1989.

[21] N. M. Yanev. On the Statistics of Branching Processes. *Theor. Prob. Appl.* **20** (1975), 623–633.

[22] N. M. Yanev. Statistical Inference for Branching Processes. In: Records and Branching Processes (Eds M. Ahsanullah, G. P. Yanev) NOVA Science Publishers, Inc., New York, 2008, 143–168.

[23] N. M. Yanev, S. Tchoukova-Dancheva. On the Statistics of Branching Processes with Immigration. *C. R. Acad. Bulg. Sci.* **33**, No 4 (1980), 469–571.

Dimitar Atanasov
New Bulgarian University
21, Montevideo Str.
Sofia, Bulgaria
e-mail: `datanasov@nbu.bg`

Nikolay Yanev
Institute of Mathematics and Informatics
Bulgarian Academy of Science
Acad. G. Bontchev Str., Bl. 8
1113 Sofia, Bulgaria
e-mail: `yanev@math.bas.bg`

Vessela Stoimenova
Sofia University
Faculty of Mathematics and Informatics
5, J. Boucher Str.
1164 Sofia, Bulgaria
e-mail: `stoimenova@fmi.uni-sofia.bg`
and
Institute of Mathematics and Informatics
Bulgarian Academy of Science
Acad. G. Bontchev Str., Bl. 8
1113 Sofia, Bulgaria